Algorithmique et simulation avec Xcas

Renée De Graeve

20 juillet 2017

Chapitre 1

Le tableur

1.1 Généralités

1.1.1 Pour ouvrir un niveau contenant un tableur

Pour avoir un tableur, il faut utiliser le menu `Edit`, puis `Ajouter`, puis `Tableur,statistiques` ou le raccourci `Alt+t`.
On vous demande un nom, pour la sauvegarde ultérieure de ce tableur. C'est ce nom de variable qui servira à sauver la matrice définie par le tableur et c'est ce nom suivi du suffixe `.tab` qui sera de nom du fichier contenant le tableur et ses formules. Par exemple, si vous donnez comme nom `M`, la variable `M` contiendra la matrice définie par le tableur et lorsque vous appuyez sur le bouton `Save M.tab`, le fichier `M.tab` sera le fichier qui contiendra le tableur avec toutes ses formules et que l'on pourra retrouver lors de séances ultérieures.
Attention Si vous ne donnez pas de nom, le bouton `Save` ne sera pas visible et il le deviendra si vous utilisez `Table->Sauver tableur comme du texte` et vous donnez, par exemple,comme nom `M`.
Si le contenu du tableur change, la matrice `M` change sans avoir besoin de sauver, par contre le fichier `M.tab` ne changera que si l'on sauve c'est à dire si on appuie sur `Save M.tab`. On peut aussi sauver la sélection par exemple `A0:C3` vers une variable (menu `Table`), cette variable contiendra une matrice qui ne changera pas même si le tableur change.
La `Configuration generale` du menu `Configuration` permet de déterminer le nombre de lignes et de colonnes que l'on aura lors de l'ouverture du tableur.

1.1.2 Déscription d'un niveau contenant un tableur

Dans un niveau contenant un tableur nous avons :

— en haut la barre de menu de ce niveau :
`Table Edit Maths`.
On peut retrouver ces menus avec un click droit de la souris n'importe où dans le tableur et on retrouve aussi ces menus dans le menu `Tableur` du menu général.
À coté de cette barre de menus les boutons :
`eval val init 2-d 3-d` ou par exemple

`eval val init 2-d 3-d Save M.tab`,

— une ligne composée de deux cases :
  — la case de sélection qui permet soit de sélectionner une cellule (en tapant par exemple `B0`) ou un sous tableau (en tapant par exemple `B0..D3` ou `B0:D3`), ou un sous tableau avec des colonnes non contigües (en tapant par exemple `B0..3,D`), soit de savoir ce qui est sélectionné avec la souris.
  — une ligne de commande qui permet de modifier une cellule du tableur ou de savoir ce qui se trouve dans cette cellule.
  `Attention`
  Il faut savoir que si le curseur est dans cette ligne de commande, lorsqu'on clique dans une case, c'est le nom de cette case qui va s'afficher dans cette ligne, sans changer la case de sélection. Pour enlever le curseur de la ligne de commande tapez sur `Echap` ou `Escap` ou encore sur la touche d'effacement ⇐ qui effacera ce qui se trouve dans cette ligne y compris le curseur. Si le curseur n'est pas dans la ligne de commande, lorsqu'on clique dans une case, c'est la valeur de cette case qui va s'afficher dans cette ligne, en mettant son nom dans la case de sélection.

— la ligne d'état rappelant la configuration choisie : c'est aussi un bouton qui permet de configurer le tableur.
  On détermine la configuration soit en appuyant sur la ligne d'état soit on utilise le menu :
  `Edit`▶`Configuration` du tableur.
  On a par exemple dans la ligne d'état :
  `* Spreadsheet M R40C6 auto down fill`
  cela veut dire que l'on a un tableur qui a été modifié depuis la dernière sauvegarde (`*`), de nom de variable `A` qui a 40 lignes et 6 colonnes, il est réévalué automatiquement, le curseur se déplace vers le bas lorsqu'on vient de remplir une cellule et une matrice remplit plusieurs cellules.
  On peut aussi avoir par exemple :
  `- Matrix <> R4C6 manual right cell`
  cela veut dire que l'on a une matrice qui n'a pas été modifiée depuis la dernière sauvegarde (−), il ne lui correspond pas de nom de variable (<>), elle a 4 lignes et 6 colonnes, elle n'est réévaluée que si on appuie sur le bouton `eval`, le curseur se déplace vers la droite lorsqu'on vient de remplir une cellule et une matrice remplit une seule cellule.

— le tableur ou l'éditeur de matrice. Si on a coché `Graphe` dans la configuration du tableur, on aura aussi un écran de représentation graphique du tableur : cet écran se trouve soit en dessous du tableur si on a coché `Paysage`, soit à sa droite si on a décoché `Paysage`. C'est dans cet écran que s'afficheront toutes les commandes graphiques situées dans les cellules du tableur. Par exemple, on met `1` dans `A0`, `2` dans `B0`, `=cercle(A0,B0)` dans `C0` et `=cercle(B0,A0)` dans `D0`. On obtient le tracé de deux cercles et une modification de l'une des cases `A0` ou `B0` modifira ce tracé.

1.1.3 Tableur et éditeur de matrice

Le tableur est une feuille de calculs ayant la forme d'un tableau composé de lignes et de colonnes qui déterminent des cases appelées cellules. Les cellules contiennent des valeurs ou des commandes ou encore des formules qui font références aux autres cellules.
Un éditeur de matrice a aussi la forme d'un tableau composé de lignes et de colonnes qui déterminent des cases, mais ces cases ne peuvent contenir que des scalaires.
Dans le menu `Edit`▶`Configuration` du tableur, l'item `Format` permet d'avoir soit un tableur, soit un éditeur de matrice permettant d'entrer facilement des matrices quelconques ou symétriques ou etc...On a donc la possibilité lorsque l'on veut mettre dans le tableur une matrice particulière (par exemple une matrice symétrique) de choisir de le faire dans l'éditeur de matrice associé au tableur (on choisit par exemple `matrice symétrique` dans `Format`), on entre la matrice (les éléments symétriques sont mis automatiquement) puis, on repasse en mode tableur en choisissant `Tableur` dans `Format`.

Description de l'écran du tableur

Le tableur est un tableau composé de colonnes désignées par les lettres majuscules `A,B,C,...` et de lignes numérotées par `0,1,2,...`.
Les cases du tableur sont appelées cellules.
Ainsi, `A0` désigne la première cellule du tableur.

Description de l'éditeur de matrice

L'éditeur de matrice est un tableau composé de lignes et de colonnes numérotées par `0,1,2,...`
Les cases de l'éditeur de matrice sont les éléments de la matrice.
Si on sauve la matrice en lui donnant comme nom `M`, `M[0,1]` désigne la case située dans la ligne de numéro `0` et dans la colonne de numéro `1`.

1.1.4 Principe et configuration du tableur

Le menu `Edit`▶`Configuration` du tableur permet de configurer le tableur (tableur se traduit en anglais par `Spreadsheet`).
Le tableur est une feuille de calculs ayant la forme d'un tableau composé de lignes et de colonnes. Lorsqu'on a choisit `Recalculer automatiquement` dans le menu `Edit`▶`Configuration` du tableur, les cases ou cellules sont mises à jour automatiquement lorsque l'on modifie une des cases et sinon il faut utiliser le bouton `eval` ou utiliser dans le menu `Edit`▶`Configuration` du tableur, la commande `Evaluer le tableur` (à exécution directe) ou encore utiliser le raccourci clavier en appuyant sur `F9`.
L'item `Format` de ce menu configuration permet d'avoir soit un tableur, soit un éditeur de matrice permettant d'entrer facilement des matrices quelconques ou symétriques ou etc..
On peut préciser le nombre de lignes et de colonnes avec lesquelles on veut travailler : par exemple pour entrer une matrice symétrique il faut avoir le même

nombre de lignes et de colonnes, on change ce nombre avec les items `Changer le nombre de lignes` et `Changer le nombre de colonnes` dans le menu `Edit`▶ `Configuration` du tableur.
On pourra bien sûr modifier ces nombres au cours du travail, par exemple en utilisant le menu :
`Edit`▶`Configuration`▶`Ajouter/effacer` du tableur ou en utilisant la `case de sélection` (si on met `G50` dans cette case, il y aura alors création d'un nombre suffisant de lignes et de colonnes pour pouvoir sélectionner `G50`)
Toutes les fonctions (même graphiques) de `Xcas` sont utilisables dans le tableur.
Le bouton `STOP` permet d'interrompre un calcul trop long.

1.1.5 La `case de sélection`

La `case de sélection` est la case située en dessous du menu `Table`.
La `case de sélection` est une case interactive :
- elle permet de connaitre le nom de la (ou des) cellule(s) sélectionnée(s) avec la souris (si on sélectionne `A3`, `A3` se note automatiquement dans cette case, si on sélectionne `A2,A3,A4,B2,B3,B4`, `A2:B4` se note automatiquement dans cette case),
- elle permet aussi d'aller directement sur une cellule dont on spécifie le nom : en effet lorsqu'on appuie sur cette case le curseur apparait, et on peut remplacer, par exemple, `A3` par `A30` : les lignes (et les colonnes) nécessaires sont créées et la cellule `A30` se trouve sélectionnée.
- elle permet aussi de sélectionner une ou plusieurs colonnes, par exemple, en tapant dans cette case `A0:C9` ou `A0..C9` cela sélectionnera les 10 premières lignes des colonnes `A,B` et `C` ou encore en tapant dans cette case `A0..9,C` cela sélectionnera les 10 premières lignes des colonnes `A` et `C`. Grâce à cette case de sélection on peut donc sélectionner des colonnes non contigües, ce que l'on ne peut pas faire avec la souris.

1.1.6 Les différents boutons d'un tableur

Les différents boutons du tableur sont :
- `eval` pour évaluer le tableur lorsqu'on n'est pas en mode automatique. On peut aussi utiliser le menu, `Edit`▶`Configuration`▶`Recalculer automatiquement`. Cela permet de passer en mode automatique de façon à ce que le tableur soit évalué après chacune de ses modifications,
- `val` pour avoir la valeur de la cellule dans la ligne de commande à la place de la formule,
- `init` pour avoir, par exemple, une variable qui compte le nombre d'évaluation du tableur : on met par exemple :
  - `j:=0` dans la case `Init sheet` de la configuration du tableur et
  - `=(j:=j+1)` dans une cellule du tableur.
    À chaque évaluation la cellule contiendra `1,2...`. En appuyant sur `init` on réinitialise la valeur de `j` et on remet la cellule à `1`
- `Save` si vous n'avez pas donné de nom à l'ouverture, ce bouton n'existe pas. pour le créer, utiliser `Table->Sauver tableur comme du texte`. Si vous avez donné un nom à l'ouverture par exemple `toto`, le bouton

Save sauve alors à la fois le tableur et ses formules dans le fichier toto.tab (d'extension .tab) et les valeurs dans la matrice toto. La matrice toto pourra alors être reutilisée et le fichier toto.tab pourra être inséré dans un tableur lors de séances ultérieures.

— 2-d 3-d ouvre un écran de gépmétrie 2-d ou 3-d pour que l'on puisse voir les commandes graphiques du tableur

1.2 La barre de menu d'un tableur

1.2.1 Le menu Table d'un tableur

Le menu Fich est composé de :

— Si vous avez donné un nom à l'ouverture, Sauver est identique au bouton Save et sauve le tableur dans un fichier d'extension .tab. L'extension .tab est rajoutée automatiquement. Si vous n'avez pas donné de nom à l'ouverture, Sauver vous en demande un, par exemple toto et Sauver crée le bouton Save et il sauve à la fois le tableur et ses formules dans le fichier toto.tab (d'extension .tab) et les valeurs dans la matrice toto. La matrice toto pourra alors être reutilisée et le fichier toto.tab pourra être inserer dans un tableur lors de séances ultérieures.

— Sauver comme sauve le tableur sous un nom (d'extension .tab) différent de celui noté à coté du bouton Save,

— Sauver selection vers variable pour stocker dans une variable la matrice mise en surbrillance et pouvoir ainsi utiliser cette variable ailleurs (calcul formel par exemple). Par exemple si le nom de la variable est a, a[0,2] donnera dans une ligne de commandes la valeur située à la ligne 0 et à la colonne 2 de la sous-matrice selectionnée,

— Inserer pour mettre à partir de la cellule mise en surbrillance un tableur sauvé précédemment,

— Nom de variable pour donner un nom de variable au tableur différent du nom de fichier sans son suffixe .tab. Ce nom est noté dans la ligne d'état située en dessous de la ligne de commande. Par exemple si le nom de la variable est M, M[0,1] renvera la valeur située en B0,

— Imprimer pour imprimer le tableur. Vous pouvez prévisualiser avant d'imprimer.

1.2.2 Le menu Edit d'un tableur

On trouve dans ce menu :

— Evaluer le tableur F9 pour recalculer le tableur lorsqu'on n'est pas en mode automatique : pour que le recalcul soit automatique, il faut passer en mode automatique avec le menu Edit▶Configuration▶Recalculer automatiquement. Le raccourci clavier de cet item Evaluer le tableur est F9 et il a le même effet que le bouton eval : cela permet de recalculer les cellules du tableur après une modification.

— Copier la cellule (en anglais Cell copy) permet de recopier une cellule dans une autre cellule : on sélectionne à la souris la cellule que l'on veut recopier. On clique ensuite sur Copier la cellule puis,

on clique sur la cellule à remplir et on clique sur le bouton `coller` du bandeau général ou on utillise l'item `Coller` ci-après.

— `Coller` sert à copier ce qui a été auparavant sélectionné.

— `Configuration` contient les items suivants :

  — `Format` permet de choisir d'avoir un tableur ou un éditeur de matrice permettant d'éditer facilement des matrices symétriques, antisymétriques, hermitiennes, antihermitiennes,quelconques,

  — `Changer le nombre de lignes` permet de spécifier le nombre de lignes du tableur ou de la matrice,

  — `Changer le nombre de colonnes` permet de spécifier le nombre de colonnes du tableur ou de la matrice,

  — `Déplacer ->` la surbrillance ira automatiquement sur la cellule située à droite de la cellule que l'on vient de remplir,

  — `Déplacer vers le bas` : la surbrillance ira automatiquement sur la cellule située en dessous, de la cellule que l'on vient de remplir,

  — `Recalculer automatiquement` pour que le tableur soit recalculé automatiquement après chaque modification,

  — `Ne pas recalculer automatiquement` pour que le tableur ne soit pas recalculé automatiquement : le recalcul ne se fait alors que si on appuie sur le bouton `eval`,

  — `Distribuer une matrice sur plusieurs cellules` pour remplir plusieurs cellules avec une matrice : par exemple si on sélectionne `A0` et que l'on tape dans la ligne de commandes du tableur `[1,2,3]`, cela remplira 3 cellules, en mettant 1 dans `A0`, 2 dans `B0` et 3 dans `C0`, par contre si on tape dans la ligne de commandes du tableur `=[1,2,3]` cela mettra [1,2,3] dans `A0`.

  — `Conserver une matrice dans une seule cellule` permet de remplir une cellule avec une matrice : par exemple si on sélectionne `A0` et que l'on tape dans la ligne de commandes du tableur `[1,2,3]` ou `=[1,2,3]`, cela mettra [1,2,3] dans `A0`.

— `Trier` permet de trier plusieurs lignes (resp colonnes) selon l'ordre croissant (resp décroissant) d'une colonne (resp ligne).
Par exemple on a dans les colonnes `A` et `B` :
`[[3,9],[5,12],[2,14],[4,8],[1,11]]` qui represente le numéro d'une copie et sa note.
On peut alors :
- soit trier ce tableau pour ordonner le numéro des copies par ordre croissant (c'est à dire par rapport à la colonne `A`) on demande alors : `Col/crois` puis on marque `A`. On obtient alors le tableau :
`[[1,11],[2,14],[3,9],[4,8],[5,12]]`
- soit trier ce tableau pour ordonner les notes des copies par ordre décroissant (c'est à dire par rapport à la colonne `B`) on demande alors : `Col/décrois` puis on marque `B`. On obtient alors le tableau :
`[[2,14],[5,12],[1,11],[3,9],[4,8]]`
Attention
Pour trier une colonne dans une autre colonne du tableur, on ne peut pas le faire directement, car quand on écrit une formule dans le tableur, elle ne peut remplir que la case courante (sinon cela poserait trop de problèmes

pour les dépendances des cellules).
Le remplissage par une matrice n'est possible qu'en évaluation directe sans dépendances.
Si on veut trier la colonne `A` dans `B`, on crée une cellule par exemple `C0` avec `=sort(A0:A10)`, `C0` contient alors la liste `A0:A10` triée.
Puis dans `B0` on écrit `=($C$0)[Row()]` et on recopie `B0` vers le bas : on obtient alors la recopie de la liste `C0` dans `B` puisque `Row()` désigne l'indice de la cellule dans laquelle la formule est recopiée, indice qui est aussi l'indice des éléments de la liste `C0`.

— `Remplir` contient les items suivants :
  — `Remplir sélection de 0` remplit la sélection par des zéros,
  — `Copier vers la droite` recopie sur toutes les cellules situées à droite de la cellule mise en surbrillance, le contenu (ou la formule) qui s'y trouve,
  — `Copier vers le bas` recopie sur toutes les cellules situées en dessos de la cellule mise en surbrillance, le contenu (ou la formule) qui s'y trouve,
  — `Remplir la sélection avec la cellule enfoncée`, recopie le contenu (ou la formule) de la cellule qui a débuté la sélection faite avec la souris de la zone rectangulaire dans laquelle on veut faire une recopie (la cellule que l'on copie est donc un des quatre coins de la zone rectangulaire),
  — `Remplir sélection de 0` remplit la sélection avec des zéros,
  — `Remplir le tableur de 0` remplit le tableur avec des zéros,
  — `tablefunc` permet d'avoir une table numérique des valeurs d'une expression (voir 1.6.1),
  — `tableseq` permet d'avoir les valeurs numériques des termes d'une suite récurrente (voir 1.6.2),
— `Ajouter/effacer` contient les items suivants :
  — `Inserer ligne` rajoute une ligne juste avant la ligne où se trouve la cellule mise en surbrillance,
  — `Ligne+ en fin` rajoute une ligne à la fin du tableur,
  — `Inserer colonne` rajoute une colonne juste avant la colonne où se trouve la cellule mise en surbrillance,
  — `Col+ en fin` rajoute une colonne à la fin de tableur,
  — `Effacer ligne courante` supprime la ligne où se trouve la cellule mise en surbrillance,
  — `Effacer selection lignes` efface le contenu des lignes sélectionnées,
  — `Effacer col courante` supprime la colonne où se trouve la cellule mise en surbrillance,
  — `Effacer sélection cols` efface le contenu des colonnes sélectionnées,
— `Col+grande` agrandit ou diminue la taille des colonnes,
— `Col+petite` diminue la taille des colonnes,

1.2.3 Le menu `Maths` d'un tableur

Le menu `Maths`▶`stats 1-d` d'un tableur

Le menu `Statistics` ouvre pour chaque item une boite de dialogue où l'on peut préciser : la sélection, la cellule cible (c'est dans cette cellule que s'inscrira la commande choisie), si comme argument de la commande choisie, on doit considérer les lignes ou les colonnes de la sélection et si on doit mettre les valeurs ou les références de la sélection .
Voici les différents items du menu `Maths`▶`stats 1-d` :

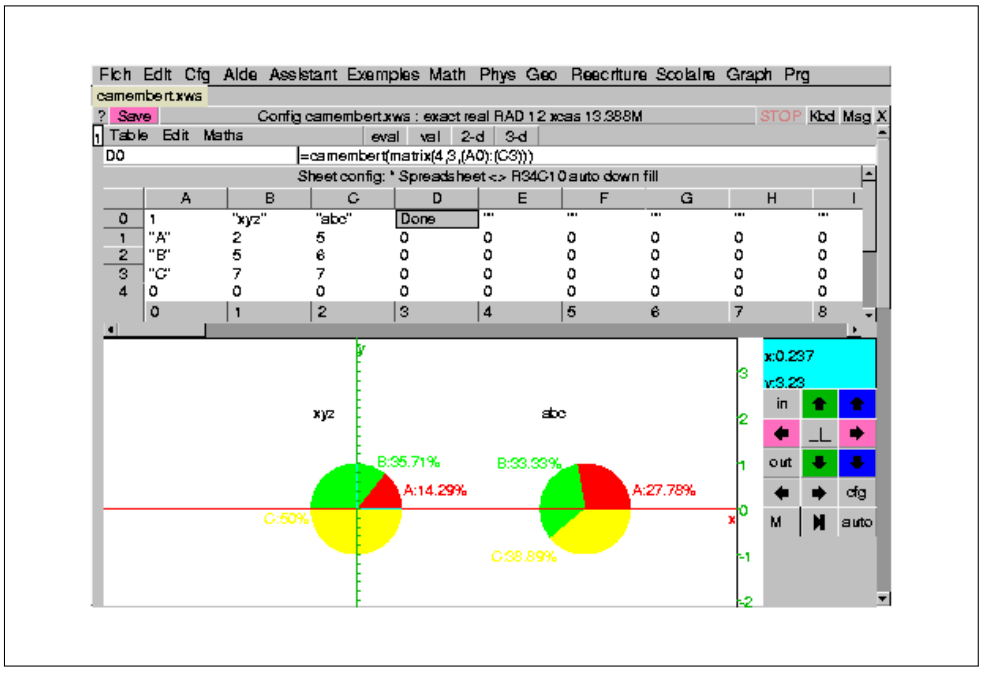
— `camembert`
On peut faire plusieurs camemberts sur le même graphique en sélectionnant toute une plage.
Exemple pour faire deux camemberts
On sélectionne la plage `A0:C3`.
On met
— dans `A0` n'importe quoi sauf une chaine vide par exemple 2
— dans `A1,A2,A3` : `"A","B","C"`

— dans `B0` le titre du premier camembert par exemple `"xyz"`,

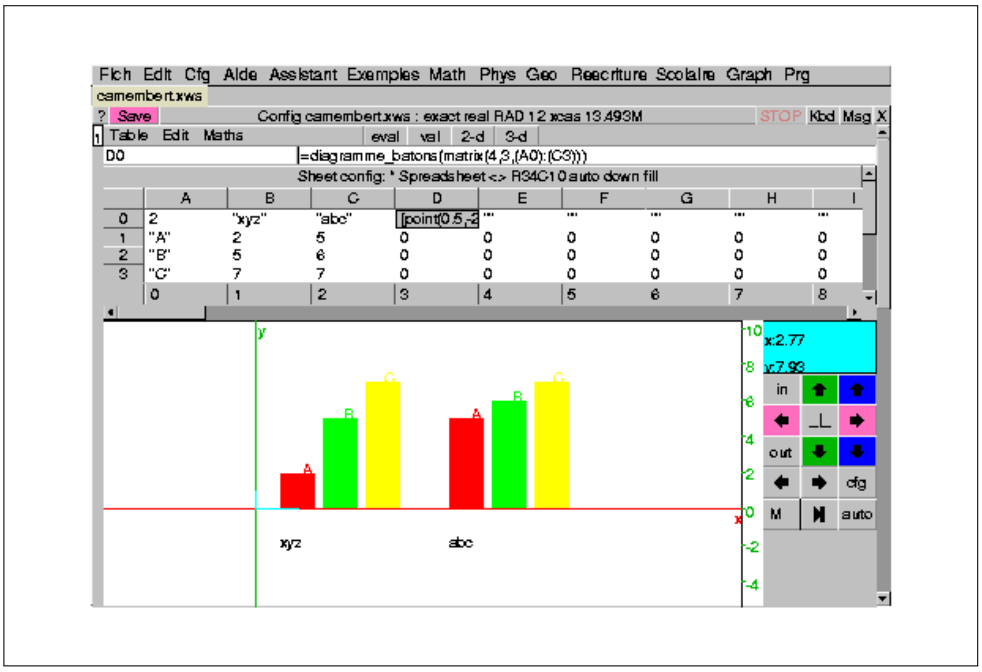
— dans `B1,B2,B3` : les valeurs du premier camembert par exemple `2,5,7`

— dans `C0` le titre du second camembert par exemple `"xyz"`,

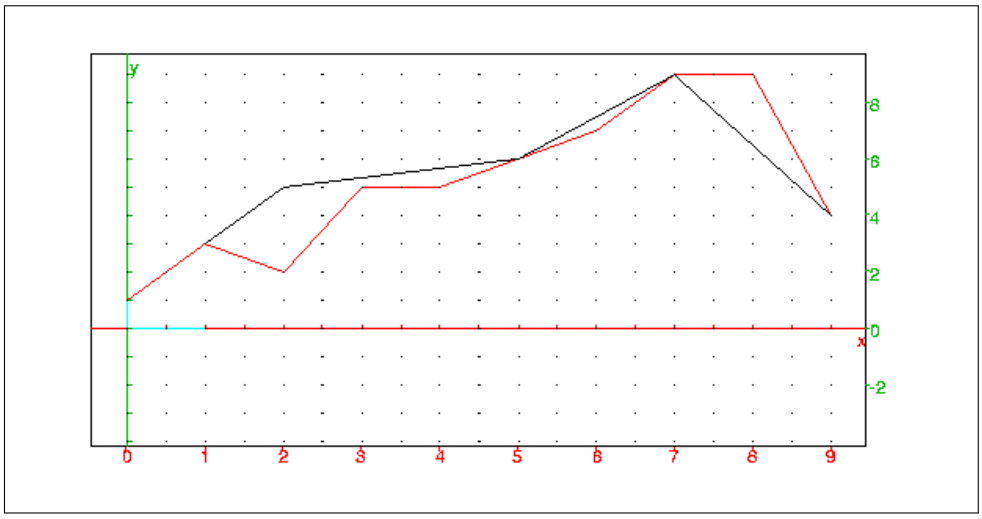
— dans `C1,C2,C3` : les valeurs du second camembert par exemple `5,6,7`
Puis on met dans `D0` : `=camembert(matrix(4,3,(A0):(C3)))` à l'aide du menu `Math`▶`Proba_stats`▶`1-d`▶`camembert`.
On obtient :



— batons
On peut faire plusieurs diagrammes en batons sur le même graphique en sélectionnant toute une plage.
Exemple pour faire deux diagrammes en batons
Avec l'exemple ci-dessus, on met dans D0 : =diagramme_batons(matrix(4,3,(A0):(C3)))
à l'aide du menu Math▶Proba_stats▶1-d▶batons.
On obtient :



— plotlist
Si plotlist a comme argument une liste L=[y1,...,yn], cela trace la ligne reliant les points d'abscisse 1,...,n et d'ordonnée L=[y1,...,yn] et si plotlist a comme argument une matrice M=[[x1,y1],...,[xn,yn]], cela trace la ligne reliant les points de coordonnées xn,yn.
Exemple On met dans A : 1,2,5,7,9 et dans B : 3,5,6,9,4 On tape dans C0 : =plotlist((A0):(B4),'affichage'=1)
On tape dans C1 ou on utilise le menu Math▶Proba_stats▶1-d▶plotlist avec pour plage A0 : B4 et pour cellule cible C1 : =plotlist(matrix(5,2,(A0):(B4)))
Attention ! Lorsqu'on met =plotlist((A0):(B4)) la plage (A0):(B4) est applatie en une liste (ici la liste 1,3,2,5,4,6,7,9,9,4) On obtient en rouge la ligne correspondant à la liste et en noir celle correspondant à la matrice :



— `Boite à moustaches` (en anglais `Boxwhiskers`)
`Boite à moustaches` permet de dessiner, dans l'écran graphique associé au tableur, les boites à moustaches des colonnes (ou des lignes si `Lignes` est cochée) des données qui ont été sélectionnées. C'est dans la cellule cible que s'inscrit la commande `moustache` avec comme argument les valeurs ou les références de la plage sélectionnée selon que l'on coche ou non `valeur`.

— `Classes (donnees ou donnees/eff)`
`Classes` permet de définir des classes : on sélectionne une colonne du tableur contenant la série que l'on veut regrouper en classes (ou si on a des effectifs, on sélectionne deux colonnes du tableur representant les données et leurs effectifs) puis, on sélectionne `Classes` dans le menu `Statistiques` sous-menu `1-d` : il faut vérifier la valeur minimum de la classe, la largeur des classes (ces deux cases sont déjà remplies avec les valeurs spécifiées dans la configuration du graphique (bouton rouge `geo`) et aussi la cellule cible qui est la première cellule à partir de laquelle on écrira les classes sur deux colonnes, Les intervalles des classes s'inscrivent dans la colonne de la cellule cible et commencent par `classe_min` et la longueur des intervalles est égale à `classe_size` : ces valeurs peuvent être spécifiées dans la configuration du graphique (bouton rouge `geo`). La colonne suivante contient les effectifs des classes obtenues dans la précédente colonne,

— `Histogramme (intervalles/eff)` (en anglais `Histogram`)
`Histogramme` permet de dessiner dans l'écran graphique associé au tableur l'histogramme de deux colonnes sélectionnées representant les intervalles de données et leurs effectifs, et inscrit la commande `histogram` correspondante dans la cellule cible.

Le menu `Maths▶stats 2-d` d'un tableur

Le menu `Maths▶stats 2-d` contient les items suivants :

— `Scatterplot` permet de tracer sur l'écran graphique les points d'abscisse la première colonne sélectionnée et d'ordonnée les autres colonnes sélectionnées si `lignes` n'est pas coché ou de tracer sur l'écran graphique

les points d'abscisse la première ligne sélectionnée et d'ordonnée les autres lignes sélectionnées si `lignes` est coché. Les couleurs seront différentes pour les points dont les ordonnées sont dans des colonnes (resp lignes) différentes.
Par exemple,

— si `lines` n'est pas coché, on remplit :
`A0,A1,A2,A3` avec `1,2,3,4` et
`B0,B1,B2,B3` avec `1,4,9,16`, puis
on sélectionne la matrice `A0:B3` et on ouvre le menu `Statistiques 2-d Scatterplot`.

— si `lines` est coché, on remplit :
`A0,B0,C0,D0` avec `1,2,3,4` et
`A1,B1,C1,D1` avec `1,4,9,16`, puis
on sélectionne la matrice `A0:D1` et on ouvre le menu `Statistiques 2-d Scatterplot`.

Une boite de dialogue s'ouvre où la plage sélectionnée est marquée (si vous n'avez rien sélectionné il faut remplir cette case) puis, il faut donner le nom de la `cellule cible` là où la commande `scatterplot` va s'inscrire. Les valeurs ou les références de la plage sélectionnée seront en argument de la commande `scatterplot` selon que l'on coche ou non `valeur`.
Supposons que `lignes` n'est pas coché.

— Si `valeur` n'est pas coché dans la boite de dialogue et si la cellule cible est `C0`, dans la ligne de commande du tableur et dans `C0` il s'inscrit alors :
`=scatterplot(matrix(4,2,(A0):(B3)))`

— Si `valeur` est coché dans la boite de dialogue et si la cellule cible est `C1`, dans la ligne de commande du tableur et dans `C1` il s'inscrit alors :
`=scatterplot([[1,1],[2,4],[3,9],[4,16]])`.

Ainsi une modification des valeurs de `A0:B3` modifira, lors d'une réévaluation du tableur,le graphique commandé par la case `C0` mais ne modifira pas le graphique commandé par la case `C1`. Autrement dit, si vous ne travaillez pas en valeur, c'est à dire avec des références, à chaque modification, on aura une modification du graphique et si vous travailler en valeur à chaque modification il faudra inscrire dans une nouvelle cellule cible la commande `scatterplot` et on aura alors plusieurs graphiques correspondant chacun aux commandes `scatterplot` des cellules cibles.
`Attention`
Lorsqu'on sélectionne une plage avec la souris on ne peut sélectionner que des colonnes contigües. On peut néanmoins remplir la plage de sélection de la boite de dialogue avec des colonnes non contigües par exemple :
`C0..5,A,D` pour dire, si `lines` n'est pas coché, que les points que l'on veut représenter ont pour abscisses la colonnes `C` et pour ordonnées la colonne `A` d'une part et pour abscisses la colonnes `C` et pour ordonnées la colonne `D` d'autre part. Mais dans ce cas `scatterplot` s'affichera comme si `valeur` etait coché (même si ce n'est pas le cas).

— `Polygonplot` permet de tracer sur l'écran graphique les points d'abscisse la première colonne sélectionnée et d'ordonnée les autres colonnes sélectionnées, en reliant entre eux les points dont les ordonnées sont dans

une même colonne.
La même boite de dialogue que pour `scatterplot` s'ouvre. Dans la ligne de commande du tableur il s'inscrit alors, par exemple, dans `D5` : `=polygonplot(matrix(3,3,A0:C2))` si ni `lignes` ni `valeur` ne sont cochés, si la cellule cible est `D5` et si la plage selectionnée est `A0:C2`.

1.3 Pour remplir le tableur

1.3.1 Comment remplir une cellule

Dans une cellule on peut mettre :

— une chaine de caractère, ou une expression numérique ou formelle : pour cela il suffit de sélectionner la cellule à remplir et de taper dans la ligne de commande ce que l'on veut mettre dans la cellule, puis de valider avec `enter`,
— une formule faisant référence aux autres cellules, dans ce cas il faut faire précéder la formule du signe = (voir la section 1.3.3 références absolues et relatives).
Attention
1/ Si le curseur ne se trouve pas dans la ligne de commande, lorsqu'on clique sur une cellule, la ligne de commande s'efface et le contenu de la cellule (valeur ou formule) s'affiche dans la ligne de commande.
2/ Si le curseur se trouve dans la ligne de commande, il faut que la ligne de commande soit vide, pour que, lorsqu'on clique sur une cellule son contenu s'affiche dans la ligne de commande.
3/ Si le curseur se trouve dans la ligne de commande, et que celle-ci contient quelque chose (par exemple =), alors lorsqu'on clique sur une cellule c'est son nom qui s'affiche dans la ligne de commande (ce qui facilite l'édition d'une formule).
4/ On enlève le curseur de la ligne de commande avec `Esc` ou `Echap`.
Ainsi, si la ligne de commande est vide et si, par exemple, on clique sur `A1` (qui contient `3`) alors le contenu de `A1` (`3`) s'affiche dans la ligne de commande :
- on peut cliquer dans une autre cellule et alors ligne de commande s'efface et son contenu s'affiche dans la ligne de commande.
- on peut cliquer dans la ligne de commande, taper quelque chose (par exemple +), puis cliquer sur une autre cellule (par exemple `A2`) et alors le nom de la cellule `A2` apparaitàla suite du contenu précédent (`3+`).
Donc quand on édite le contenu d'une case, si on clique dans le tableur alors, le nom ou la plage des noms selectionnés, s'affichent dans la ligne de commande.
Exemple
On veut remplir la case `A1` avec la formule `1+A2` :
- on efface la ligne de commande,
- on clique sur `A1`,
- on clique sur la ligne de commande et on tape `=1+`,

- on clique sur la cellule A0 puis, Enter.
Cela marque 1+A0 dans la cellule A1.
Pour annuler une modification en mode édition, il n'est donc pas possible de simplement cliquer sur une autre cellule du tableur puisque cela recopierait le nom de cette autre cellule. En ligne de commande, c'est la touche Esc qui annule l'édition.

— une liste ou une matrice à condition de faire précéder la liste ou la matrice du signe = car, si on ne met pas le signe = cela aura pour effet de remplir plusieurs cellules (voir ci-dessous).
On peut remplir d'un seul coup plusieurs cellules à partie d'une cellule lorsque l'on met dans cette cellule une liste ou une matrice.
Par exemple :
Dans A0 on met : [1,2,3], cela a pour effet de remplir A0 avec 1, B0 avec 2 et C0 avec 3.
Dans A0 on met : [[1,2],[3,4]], cela a pour effet de remplir A0 avec 1, B0 avec 2, A1 avec 3 et B1 avec 4.
Attention
Le remplissage de plusieurs cellules à l'aide d'une matrice ou d'une liste n'est possible qu'en évaluation directe sans dépendance. Par exemple :
On ne peut pas mettre [A0,B0] dans C0, mais on peut mettre =[A0,B0] dans C0. Le signe = est necessaire pour remplir **une seule** cellule avec une liste ou une matrice.
Exemple
Pour mettre une liste ou une matrice dans une case il faut mettre le signe = devant la liste ou la matrice.
On tape dans la case A0 :
=[[1,2],[3,4]]
On obtient dans la case A0 :
[[1,2],[3,4]]
Si l'on ne met pas le signe égal on peut remplir d'un seul coup plusieurs cases du tableur par les éléments de la liste ou de la matrice à condition que la formule mise ne fasse pas référence aux autres cellules.
On tape dans la case A0 :
[[1,2],[3,4]]
On obtient :
A0=1, B0=2, A1=3, B1=4
Autre exemple
Dans A0 je tape :
ranm(2,3,4)
Je remplis alors d'un seul coup les 6 cases : A0,B0,C0,A1,B1,C1 avec les éléments de la matrice [[1,3,1],[2,1,2]] (en effet ranm(2,3,4) renvoie une matrice de 2 lignes et 3 colonnes d'entiers pris au hasard de façon équirépartie dans l'ensemble des 4 nombres 0,1,2,3.
Cette matrice est définie une fois pour toute car la formule ranm(2,3,4) est évaluée puis oubliée.
Dans A0 je tape :
=ranm(2,3,4)
Cette fois, seule la cellule A0 est remplie avec la matrice :

`[[1,2,3],[3,2,2]]`.
Cette matrice changera à chaque modification du tableur sauf, si on a choisi `Ne pas recalculer automatiquement` dans la configuration du tableur (avec le menu `Edit`►`Configuration`).

1.3.2 Pour voir le contenu d'une cellule

Lorsqu'on consulte le tableur, toutes les cellules sont évaluées. Pour voir le contenu évalué d'une cellule située hors du champ de vision, on utilise, soit le curseur vertical (situé à droite du tableur) qui permet de voir les dernières lignes, soit le curseur horizontal (situé à la dernière ligne du tableur) qui permet de voir les dernières colonnes.
Remarque
Lorsque toutes les cellules sont visibles, ces deux curseurs ne sont pas présents.
Pour voir le contenu non évalué d'une cellule (c'est à dire ce que l'on a mis, initialement, dans la cellule comme formule ou comme valeur), il suffit de cliquer sur cette cellule : il apparait alors dans la ligne de commande la formule (ou la valeur) qui a été mise dans cette cellule.
Pour faire apparaitre dans la ligne de commande, le contenu évalué de la cellule, il suffit d'appuyer sur le bouton `eval`.
Lorsque le résultat est trop grand (ou trop petit) en taille on peut avoir besoin d'agrandir une colonne pour voir ce résultat entièrement (ou de diminuer la colonne pour gagner de la place). Par exemple, on veut modifier la taille de la colonne `B`, on déplace la souris sur le trait vertical qui sépare `B` de `C` : le curseur devient $\leftrightarrow$. On clique avec la souris et, sans relacher le bouton de la souris, on déplace ce trait vertical pour agrandir ou diminuer la largeur de la colonne `B`. Lorsque le résultat est trop grand, on peut aussi appuyer sur le bouton `eval` pour faire apparaitre ce résultat dans la ligne de commande.

1.3.3 Références absolues et relatives

Dans une cellule on peut mettre :
- une chaine de caractères,
- une expression algébrique,
- une formule faisant référence à d'autres cellules. Ces références peuvent être absolues ou relatives à la cellule qui contient la formule. Les références absolues sont obtenues en rajoutant `$` devant la lettre désignant la colonne ou devant le numéro de la ligne de la cellule référence.

Les références relatives permettent de désigner les cellules par rapport à une autre : ainsi `A0` mis dans la cellule `B1` désigne la cellule située dans la colonne précédente et à la ligne précédente et c'est cette information qui sera recopiée quand on recopiera la formule vers le bas ou vers la droite.
Exemples :
Dans `A0` il y a 1 et je tape dans `B1` la formule :

— `$A$0+2` : dans `B1` il y aura 3, et si je recopie cette formule vers le bas, j'obtiens des 3 dans la colonne `B` car je recopie dans toutes les cases de la colonne `B` la formule `$A$0+2`. Si je recopie cette formule vers la droite, j'obtiens aussi des 3 dans la première ligne, car je recopie dans toutes les

cases de la première ligne la formule `$A$0+2` puisque `$A$0` est la référence absolue de la case `A0`.

— `$A0+2` : dans `B1` il y aura 3, et si je recopie cette formule vers le bas cette formule deviendra `$A1+2` dans `B2`, `$A2+2` dans `B3`. La valeur de `B2` dépend donc de la valeur de `A1`, la valeur de `B3` dépend donc de la valeur de `A2` etc...
Si je recopie cette formule vers la droite, cette formule deviendra `$A0+2` dans `C1`, `$A0+2` dans `D1` ...j'obtiens donc une ligne de 3. `$A0` fait toujours référence à la colonne `A` : `A` est une référence absolue mais `0` désigne ici la ligne précédente puisque `$A0` a été mis dans `B1`.

— `A$0+2` : dans `B1` il y aura 3, et si je recopie cette formule vers le bas, j'obtiens des 3 dans la colonne `B` mais si je recopie cette formule vers la droite, cette formule deviendra `B$0+2` dans `C1`, `C$0+2` dans `D1` etc...

— `A0+2` : dans `B1` il y aura 3, et si je recopie cette formule vers le bas, cette formule deviendra `A1+2` dans `B2`, `A2+2` dans `B3` etc...si je recopie cette formule vers la droite, cette formule deviendra `B0+1` dans `C1`, `C0+1` dans `D1` etc...

1.3.4 Référence d'un sous-tableau

Il y a deux façons de désigner un morceau du tableur selon que l'on veut remplir une cellule ou le sélectionner dans la `case de sélection`.
Il sera désigné par :

— dans une cellule, par la référence de sa première case puis "deux points" (`:`), puis la référence de sa dernière case.
Les références de la première ou de la dernière case seront selon les cas, absolues ou relatives, mais **attention** cela est valable pour des colonnes contigües et représente une liste i.e. le tableau est aplati,

— dans la `case de sélection` par référence de sa première case puis "point point" (`..`), puis le numéro de la dernière ligne, puis virgule (`,`) suivi de la séquence des lettres désignant les colonnes. Cette fois cela représente une matrice est on peut facilement désigner des colonnes non contigües.

Remarque Dans la `case de sélection`, on ne peut pas utiliser "deux points" (`:`) comme séparateur entre les deux références.
Exemple
Dans une cellule `A0:B5` représente la liste des valeurs de `[A0,B0,A1,B1,..,A5,B5]` constituée par la matrice "aplatie".
Dans la `case de sélection`, `A0..B5` désigne le tableau de 6 lignes (lignes 0,1..5) et 2 colonnes (colonnes `A` et `B`).
On a aussi la possibilité de désigner dans la `case de sélection` des colonnes non consécutives, on écrira par exemple `A0..10,C,E` pour sélectionner les 11 premières lignes des colonnes `A`, `C` et `E`.
Attention ! ! !
Seule la `case de sélection` permet de définir un sous-tableau ou une matrice.
Donc dans une cellule pour désigner un sous-tableau on peut reconstituer la matrice à l'aide de la commande `list2mat` (qui transforme une liste en matrice selon le nombre de colonnes spécifié) si les colonnes sont consécutives ou en utilisant la

commande `tran` (qui transforme une matrice en sa transposée).
Dans la cellule `F0` on tape par exemple :
`=list2mat(A0:B5,2)` pour avoir une matrice avec 2 colonnes et 6 lignes Dans la cellule `F0` on tape par exemple :
`=list2mat(A0:D5,4)` pour avoir une matrice avec 4 colonnes et 6 lignes dans la cellule `F0`.
Dans la cellule `F0` on tape par exemple :
`=tran([A0:A3,C0:C3])` pour avoir une matrice avec 2 colonnes et 4 lignes dans la cellule `F0`.
Dans la cellule `F0` on tape par exemple :
`=tran([A0:A3,B0:B3,C0:C3])` pour avoir une matrice avec 3 colonnes et 4 lignes dans la cellule `F0`.
Attention ! ! !
On ne peut pas remplir plusieurs cellules d'un seul coup avec une matrice contenant des références à d'autres cellules : quand il y a une formule faisant des références à d'autres cellules on ne peut remplir qu'une seule cellule et il faut mettre le signe = devant la formule.

1.4 Pour sauver l'écran du tableur

1.4.1 Pour sauver une matrice

Vous avez rempli l'écran du tableur avec une matrice.

Pour sauver cette matrice vous pouvez utiliser le bouton `Save` de la barre de boutons cela sauvera la matrice sous le nom inscrit dans la ligne d'état après `Spreadsheet` (en général le même nom que le nom du fichier sans son extension `.tab` ou vous pouvez utiliser le menu `Fich` sous-menu `Nom de variable` en donnant un autre nom de variable.

Pour sauver une sous-matrice, il faut la sélectionner soit avec la souris, soit en utilisant la case de sélection et ensuite utiliser le menu `Fich` sous-menu `Sauver selection vers variable` puis donner le nom de la variable qui stockera la matrice sélectionnée (cf 1.1.6).
La format est celui d'une matrice. On aura donc dans la variable designée une matrice par exemple : `[[1,2,3,4,5],[1,0,2,0,1],[...]]`

1.4.2 Pour sauver un tableur

Vous avez rempli un tableur avec différentes formules. Pour sauver ce tableur il suffit d'utiliser le bouton `sauver` de la barre de boutons ou on utilise le menu `Fich` sous-menu `Sauver comme` (cf 1.1.6).
Lorsque vous sauvez ainsi, vous sauvez à la fois les formules et les valeurs. En effet, le format de sauvetage est une matrice dont chaque coefficient est une liste de trois éléments : le premier élément est la formule qui définit la cellule, le deuxième élément est la valeur prise par la cellule et le troisième est une variable détat interne. On aura par exemple `spreadsheet[[[3,3,2],[=A0+1,4,2]],[...]]`, cela vaut dire que `A0=3`, que `B0=A0+1` et que la valeur de `B0` est `4`.
Lors d'une évaluation du tableur, la troisième valeur vaut :

`0` si la cellule n'a pas encore été recalculée, `1` si la cellule est en cours de calcul,
`2` si la cellule a été calculée.

L'algorithme est le suivant :
1/ On fait toutes les cellules de la gauche vers la droite et du haut vers le bas,
2/ Si le troisième argument de la cellule vaut 2, c'est fini. Si le troisième argument de la cellule vaut 1, on envoie une erreur : "évaluation récursive",
3/ le troisième argument de la cellule vaut 0, on le met à 1 et on cherche toutes les cellules dépendant de cette cellule, on calcule leurs valeurs et on remplace puis, on met à 2 le troisième argument de la cellule.

1.5 Pour copier une partie du tableur dans une ligne d'entrée

1.5.1 Pour copier une seule cellule du tableur dans une ligne d'enrée

Il faut sélectionner la cellule à recopier avec la souris et se servir du bouton `copier` du tableur.
Puis, on met le curseur dans une ligne d'enrée, puis on appuie sur la touche `coller`, et la cellule est recopiée dans la ligne de commande.

1.5.2 Pour copier plusieurs cellules du tableur dans une ligne d'entrée

Si les cellules sont consécutives, on peut les sélectionner avec la souris, sinon on utilisera la `case de sélection`.
On tape par exemple dans la `case de sélection` :
`A0..3,D`
les quatre premières cellules des colonnes `A` et `D` sont alors sélectionnées.
Puis, on met le curseur dans une ligne de commande et on appuie sur la touche `coller`, et les quatre premières cellules des colonnes `A` et `D` sont recopiées dans la ligne de commande.

1.6 Les fonctions spécifiques du tableur

1.6.1 Tableau de valeurs de $f(x)$: `tablefunc`

On peut avoir, dans le tableur, le tableau des valeurs numériques d'une expression $f(x)$ pour $x = x0,\ x0+h,\ x0+2*h$.... en tapant dans la ligne de commande du tableur :
`tablefunc(f(x),x,x0,h)` ou `tablefunc(f(x),x)`.
Après avoir selectionné `A0`, on tape, par exemple, dans la ligne de commande du tableur :

```
tablefunc(x^2,x,-1,0.2)
```

On obtient si on a 15 lignes :

	A	B
0	x	x^2
1	0.2.0	"Tablefunc"
2	-1	1.0
3	-0.8	0.64
4	-0.6	0.36
..	..	..
14	1.4	1.96

Dans le cas où les valeurs du point de départ `x0` et du pas `h` ne sont pas précisées, ces valeurs valent par défaut `x0=X-` et `h=((X+)-(X-))/10` où `X-` et `X+` sont définis dans la configuration du graphique (bouton rouge `geo`) et valent au démarrage `X-=-5.0` et `X+=5.0` et donc `x0=-5.0` et `h=1.0`.

On peut donc avoir aussi dans le tableur, les valeurs numériques des termes d'une suite $u_n = f(n)$ pour $n = n0,\ n0+1,\ n0+2, ...$ en tapant dans la ligne de commande du tableur :
`tablefunc(f(n),n,n0).`
Après avoir selectionné `A0`, on tape, par exemple, dans la ligne de commande du tableur :

```
tablefunc(n^2,n,5,1)
```

On obtient si on a 15 lignes :

	A	B
0	n	n^2
1	1.0	"Tablefunc"
2	5	25.0
3	6.0	36.0
4	7.0	49.0
..	..	..
14	17.0	289.0

Remarque
Lorsque la colonne `A` est sélectionnée, `tablefunc(f(x),x,x0,h)` a pour effet de placer, dans la colonne `A` et à partir de la ligne `0` :
`x,h,x0,A2+A$1` et,
dans la colonne `B` et à partir de la ligne `0` :
`f(x),"Tablefunc",evalf(subst(B$0,A$0,A2))`

1.6.2 Termes d'une suite récurrente : `tableseq`

On peut avoir, dans le tableur, les valeurs numériques des termes d'une suite récurrente ($\mathtt{u_0} = \mathtt{u0},\ \mathtt{u_n} = \mathtt{f}(\mathtt{u_{n-1}})$) grâce à la commande :
`tableseq(f(n),n,u0).`
Après avoir selectionné `A0`, on tape, par exemple, dans la ligne de commande du tableur :

```
tableseq(0.5*(n+3/n),n,3)
```

On obtient, si on a 7 lignes :

	A
0	0.5*(n+3/n)
1	n
2	3
3	2
4	1.75
..	..
7	1.73205080757

Après avoir selectionné B0, on tape, par exemple, dans la ligne de commande du tableur :

```
tableseq(x+y,[x,y],[1,1])
```

On obtient, les premiers termes de la suite de Fibonacci :

	B
x+y	
1	x
2	y
3	1
4	1
5	2
..	..
7	5
..	..

Remarque
Lorsque la colonne E est sélectionnée, `tableseq(f(n),n,u0)` a pour effet de placer, dans la colonne E et à partir de la ligne 0 :
`f(n),n,u0,evalf(subst(E$0,E$1,E2))`.

1.7 Références de la cellule active : Row et Col

Row et Col sont des fonctions qui sont utilisables essentiellement dans le tableur, en dehors du tableur Row et Col sont le numéro de la ligne et de la colonne de la cellule sélectionnée du dernier tableur évalué.
Row n'a pas des paramètre et renvoie le numéro de la ligne de la cellule courrante.
On tape quand la cellule C4 est mise en surbrillance :
`Row()`
On obtient :
`4`
Col n'a pas des paramètre et renvoie le numéro de la colonne de la cellule courrante (la colonne A a pour numéro 0, la colonne B a pour numéro 1 etc...)
On tape quand la cellule C4 est mise en surbrillance :
`Col()`

On obtient :
3
Ainsi si dans la cellule A0 je mets :
=Row()+Col() puis je recopie cette formule vers le bas et j'obtiens :
dans la colonne A :
0,1,2,3,4...
puis je recopie cette formule vers la droite depuis A0 et j'obtiens :
dans la ligne 0 :
0,1,2,3,4...
puis je recopie cette formule vers la droite depuis A1 et j'obtiens :
dans la ligne 1 :
1,2,3,4,5... etc...
Attention Bien mettre le signe = car Row() et Col() font références à la ligne et à la colonne de la cellule dans laquelle se trouve la formule.

1.8 Nommer une cellule par une variable : current_sheet

current_sheet est une fonction qui est utilisable essentiellement dans le tableur, en dehors du tableur current_sheet permet d'avoir accès aux cellules du dernier tableur évalué.
current_sheet s'utilise soit avec :
- aucun paramètre : current_sheet() renvoie le tableur tout entier,
- un paramètre entier : current_sheet(j) renvoie la ligne j du tableur,
- deux paramètres entiers : current_sheet(j,k) renvoie la cellule du tableur située à la ligne j et à la colonne k.

Ainsi current_sheet(3,1) désigne la cellule B3.
Cela permet de désigner une cellule par deux variables entières, par exemple :
j:=3;k:=1;current_sheet(j,k)
Remarque
Pour avoir la colonne k du tableur dans une ligne de commande, il faut taper :
tran(current_sheet())[k] (puisque tran(current_sheet()) désigne la transposée du tableur).
On peut bien sûr utiliser current_sheet dans le tableur.
On tape :
=current_sheet(1,2)
ou encore, on peut prendre la valeur d'une case comme indice :
si A0 contient 1 et B1 contient 2 on peut taper, =current_sheet(A0,B1).
Exemple d'utilisation :
On tape la suite des nombres entiers dans la colonne A.
On tape dans A0 :
1
puis on tape dans A1 :
=A0+1
formule que l'on recopie avec avec le menu Edit du tableur, puis, Remplir et Copier vers le bas.
Dans la colonne B on met, par exemple, la suite $u_n = \sum_{k=0}^{n}(-1)^k/(k+1)$.
On tape dans B0 :

1
puis on tape dans B1 :
=B0+(-1)^A1/(A1+1), formule que l'on recopie avec le menu Edit du tableur, puis, Remplir et Copier vers le bas.
On veut extraire de cette suite, les termes d'indice pair dans la colonne C, on tape dans C0 :
=current_sheet(2*A0,1), formule que l'on recopie avec remplir et vers le bas.
On veut extraire de cette suite les termes d'indice impair dans la colonne D, on tape dans D0 :
=current_sheet(2*A0+1,1), formule que l'on recopie avec le menu Edit du tableur, puis, Remplir et Copier vers le bas.
Ou encore on utilise Row et on n'a besoin que de 3 colonnes .
On tape dans A0 :
1
puis on tape dans A1 :
=A0+(-1)^Row()/(Row()+1), formule que l'on recopie avec le menu Edit du tableur, puis, Remplir et Copier vers le bas.
On veut extraire de cette suite, les termes d'indice pair dans la colonne B, on tape dans B0 :
=current_sheet(2*Row(),0), formule que l'on recopie avec remplir et vers le bas.
On veut extraire de cette suite les termes d'indice impair dans la colonne C, on tape dans C0 :
=current_sheet(2*Row()+1,0), formule que l'on recopie avec le menu Edit du tableur, puis, Remplir et Copier vers le bas.

1.9 Compter les éléments du tableur vérifiant une propriété

On suppose que dans la colonne A il y a 1,2,3,4, dans la colonne B il y a 2,4,6,8 et dans la colonne C il y a 4,8,12,16.
Attention !
Pour désigner une plage du tableur, on tape A0:C3, mais Xcas le raplatit en une liste : dans l'exemple ci-dessus A0:C3=[1,2,2,4,3,6] **Remarque**
Pour avoir une méthode rapide pour compter, par exemple, les sommes obtenues lorsqu'on lance 101 fois deux dés. On remplit aléatoirement les colonnes A et B en tapant ranm(101,1) comme valeur pour A0 et B0. Il faut créer une cellule tampon, qui contiendra le tableau des valeurs de la zone à analyser. On place simplement dans cette cellule la definition de la plage précédée de =, par exemple si A0 à B100 contient des jets de deux dés, et que C contient la somme de la colonne A et de la colonne B, on met dans D0 := C0:C100.
Ensuite dans D2 à D12 on ecrit :
=count_eq(2,D0)... count_eq(12,D0)
Ainsi le calcul de la plage C0:C100 qui est long n'est fait qu'une fois. C'est le meme principe qu'en programmation, on utilise une variable intermediaire (la cellule tampon D0).

1.9.1 Compter les éléments d'un sous tableau vérifiant une propriété : `count`

`count` a deux ou trois paramètres : une fonction réelle `f` et une liste ou un sous-tableau éventuellement un paramètre optionnel `row` ou `col`.

Attention dans le tableur les paramètre optionnels `row` ou `col` ne servent pas, car dans une ligne du tableur un sous-tableau (par exemple `A0:C3`) désigne une liste : en effet si dans une cellule on met `=A0:C3`, la cellule contient la liste obtenue en mettant les lignes du sous tableau `A0:C3` bout à bout.
Si vous voulez utiliser le sous tableau (par exemple `A0:C3`) comme une matrice il faut mettre dans une cellule `=list2mat(A0:C3),3)` (car `3` est le nombre de colonnes de `A0:C3`). On peut aussi sauver la sélection `A0:C3` vers une variable (menu `Table`), car lors de cette affectation la matrice n'est pas aplati et donc cette variable contient une matrice.

`count` applique la fonction aux éléments de la liste ou du sous-tableau et en renvoie la somme.
Si `f` est une fonction boolénne `count` renvoie le nombre d'éléments de la liste ou du sous-tableau pour lesquels la fonction boolénne est vraie.
On tape dans une cellule :

```
=count((x)->x,A0:C3)
```

On obtient :

```
70
```

En effet, la somme des éléments de `A0:C3` vaut `(1+2+3+4)*7=70` car dans A il y a 1,2,3,4 dans B il y a 2*A soit 2,4,6,8 et dans C il y a 4*A soit 4,8,12,16 donc en tout il y a 7*A.
On tape dans une cellule :

```
=count((x)->(x<10 and x>5),A0:C3)
```

On obtient :

```
3
```

En effet, il y a `6,8,8`, soit 3 éléments qui sont entre 5 et 10.

1.9.2 Compter les éléments ayant une valeur donnée : `count_eq`

`count_eq` a deux ou trois paramètres : une nombre et une liste réelle ou un sous-tableau et éventuellement un paramètre optionnel `row` ou `col`.
Attention dans le tableur les paramètre optionnels `row` ou `col` ne servent pas, car dans le tableur car un sous-tableau est aplati en une liste.
`count_eq` renvoie le nombre d'éléments de la liste ou du sous-tableau qui sont égaux au premier argument.
On tape dans une cellule :

```
=count_eq(4,A0:C3)
```

On obtient :

```
3
```

car dans A il y a 1,2,3,4 dans B il y a 2*A soit 2,4,6,8 et dans C il y a 4*A soit 4,8,12,16.

1.9.3 Compter les éléments plus petits qu'une valeur donnée : `count_inf`

`count_inf` a deux ou deux paramètres : une nombre et une liste réelle ou un sous-tableau et éventuellement un paramètre optionnel `row` ou `col`.
Attention dans le tableur les paramètre optionnels `row` ou `col` ne servent pas, car dans le tableur car un sous-tableau est aplati en une liste.
`count_inf` renvoie le nombre d'éléments de la liste (ou du sous-tableau qui sont strictement inférieurs au premier argument.
On tape dans une cellule :

```
=count_inf(4,A0:C3)
```

On obtient :

```
4
```

car dans A il y a 1,2,3,4 dans B il y a 2*A soit 2,4,6,8 et dans C il y a 4*A soit 4,8,12,16.

1.9.4 Compter les éléments plus grands qu'une valeur donnée : `count_sup`

`count_sup` a deux ou trois paramètres : une nombre et une liste réelle ou un sous-tableau et éventuellement un paramètre optionnel `row` ou `col`.
Attention dans le tableur les paramètre optionnels `row` ou `col` ne servent pas, car dans le tableur car un sous-tableau est aplati en une liste.
`count_sup` renvoie le nombre d'éléments de la liste ou du sous-tableau qui sont strictement supérieurs au premier argument.
On tape dans une cellule :

```
=count_sup(4,A0:C3)
```

On obtient :

```
5
```

car dans A il y a 1,2,3,4 dans B il y a 2*A soit 2,4,6,8 et dans C il y a 4*A soit 4,8,12,16.

1.10 Les fonctions statistiques à une variable du tableur

1.10.1 Les fonctions graphiques

On peut utiliser directement le graphique depuis le tableur : les fonctions graphiques peuvent être utilisées dans une cellule. Pour avoir l'histogramme, la boite à moustaches, un nuage de points ou une ligne polygonale depuis le tableur on peut

se servir du menu Maths du tableur (puis stats 1-d) après avoir sélectionné dans le tableur l'argument avec la souris ou avec la case de sélection. Dans ce cas, une boite de dialogue s'ouvre, vous devez choisir la cellule cible (c'est dans cette cellule que s'inscrira la commande graphique), vous pouvez éventuellement modifier la sélection (en mettant par exemple A1..B6 dans cellules entrée), vous pouvez cocher (resp ne pas cocher) valeur pour que la commande graphique ait pour argument les valeurs (resp les références) de la sélection : donc si valeur n'est pas cochée un changement de valeur dans une cellule de la sélection aura pour conséquence un changement du graphique, si on est en mode automatique ou si on appuie sur le bouton eval.
Ou encore, on peut taper la commande correspondante (ou se servir du menu Math▶Proba_stats▶1-d) dans la ligne de commande du tableur en recopiant les arguments en se servant de la souris (ou en se servant de la case de sélection et de la touche coller lorsque les colonnes que l'on veut recopier ne sont pas consécutives), ou encore on peut sauver la sélection dans une variable en utilisant le menu du tableur Table▶Sauver sélection vers variable on donne un nom par exemple A, puis on tape moustache(A) ou... Dans ce cas on travaille avec les valeurs de la sélection.

1.10.2 Centre d'un intervalle : interval2center

interval2center a comme argument un intervalle ou une liste (resp séquence) d'intervalles (utile pour définir les centres des classes).
interval2center renvoie le centre de l'intervalle ou la liste (resp séquence) des centres de ces intervalles.
interval2center est utile pour définir les centres des classes.
On tape :

```
interval2center(3..5)
```

On obtient :

```
4
```

On tape :

```
interval2center([2..4,4..6,6..10])
```

On obtient :

```
[3,5,8]
```

1.10.3 Centre d'un intervalle : center2interval

center2interval a comme argument un vecteur de réels V d'au moins deux composantes et éventuellement un réel comme deuxième argument.
center2interval renvoie un vecteur d'intervalles ayant pour centres les composantes de l'argument V : ces intervalles sont définis en commençant le premier intervalle, par le deuxième argument ou à défaut par (3*V[0]-V[1])/2.
center2interval est utile pour définir des classes à partir de leurs centres et du minimum des classes.
On tape :

```
center2interval([3,5,8])
```

Ou on tape car la valeur par défaut du deuxième argument est 2=(3*3-5)/2 :

```
center2interval([3,5,8],2)
```

On obtient :

```
[2..4,4..6,6..10]
```

On tape :

```
center2interval([3,5,8],2.5)
```

On obtient :

```
[2.5..3.5,3.5..6.5,6.5..9.5]
```

Attention
On ne peut pas mettre n'importe quoi comme deuxème argument ! ! !
On tape :

```
center2interval([5,7,8],4)
```

Ou on tape, car la valeur par défaut du deuxième argument est 4=(3*5-7)/2 :

```
center2interval([5,7,8])
```

On obtient :

```
"center2interval Bad Argument Value"
```

La fonction suivante peut vous permettre de trouver l'intervalle dans lequel il faut choisir le deuxème argument, quand il y a une solution ! ! !
En effet on doit pouvoir trouver $a0, a1, a2...$ vérifiant :
$a0 < a1 < a2....$ et
$a0 + a1 = 2 * c0 = b0, a1 + a2 = 2 * c1 = b1, a2 + a3 = 2 * c2 = b2....$ quand $L = [c0, c1, c2...]$ avec $c0 < c1 < c2....$
On a donc :
$a1 = b0 - a0,$
$a2 = b1 - a1 = b1 - b0 + a0,$
$a3 = b2 - a2 = b2 - b1 + b0 - a0$
$a4 = b3 - a3...$
comme on doit avoir $a0 < a1$ et $a1 < a2$ (c'est à dire $a1 < c1$) il faut donc trouver $a0$ vérifiant $a0 < c0$ et $b0 - c1 < a0$ puis
$a2 < a3$ i.e. $a0 < c2 - c1 + c0$ et
$a3 < a4$ i.e.$a3 < c3$ $b0 - b1 + b2 - c3 < a0...$
On construit donc deux listes :
`l1=[c0,c2-2*c1+2*c0...]` et
`l2=[2*c0-c1,2*c0-2*c1+2*c2-c3,..]`.
La condition que doit vérifier `a0` est alors :
`max(l1)<a0<min(l2)`.

```
debut_classes(L):={
local l1,l2, n,j, a, b;
n:=size(L);
L:=sort(L);
l1:=[L[0]];
l2:=[2*L[0]-L[1]];
for (j:=1;2*j+1<n;j++) {
l1:=concat(l1,l2[j-1]-L[2*j-1]+L[2*j]);
l2:=concat(l2,l1[j]+L[2*j]-L[2*j+1]);
}
if (irem(n,2)==1) {
j:=quo(n-1,2);
l1:=concat(l1,l2[j-1]-L[2*j-1]+L[2*j]);
}
a:=max(l2);
b:=min(l1);
if (a<b) return(]a,b[); else return ("impossible");
}
```

On tape :

```
debut_classes([5,7,8])
```

On obtient :

```
]3,4[
```

On tape :

```
center2interval([5,7,8],3.5)
```

On obtient :

```
[3.5 .. 6.5,6.5 .. 7.5,7.5 .. 8.5]
```

On tape :

```
interval2center([3.5 .. 6.5,6.5 .. 7.5,7.5 .. 8.5])
```

On obtient :

```
[5.0,7.0,8.0]
```

1.10.4 Somme des cellules d'un sous-tableau : `sum`

La commande `sum` permet de calculer la somme des éléments d'une liste.
Si on a une matrice ou un sous-tableau définie dans un tableur, on sait qu'en désignant ses éléments par :
"référence de la première case de la matrice" **:** "référence de sa dernière case de la matrice", on obtient la liste des éléments de la matrice (par ex `A0:B1` est la liste `[A0,B0,A1,B1]` formée par la matrice aplatie).
Attention
Pour traiter les exemples qui suivront, on remplit par exemple la colonne `A` par

`0,1,2,..,n` et la colonne `B` par `0,1,4,..,n^2`.
Pour cela on met `0` dans `A0`, et
`=A0^2` dans `B0`, `=A0+1` dans `A1` puis on utilise le bouton `remplir` et `vers le bas` pour recopier les 2 formules dans chacune des colonnes `A` et `B`.
Après avoir selectionné `C0`, on tape, par exemple, dans la ligne de commande du tableur :

```
=sum(A0:B5)
```

On obtient dans `C0` :

```
70
```

en effet : 1+2+3+4+5+1+4+9+16+25=3*5+5*11=70
Mais dans une ligne d'entrée, si on tape :
`sum([[0,0],[1,1],[2,4],[3,9],[4,16],[5,25]])`
On obtient :
`[15,55]`
qui est la somme des colonnes de la matrice.
Après avoir selectionné `D0`, on tape, par exemple, dans la ligne de commande du tableur :

```
=sum(A0:B5)+B8
```

On obtient dans `D0` :

```
134
```

en effet : 1+2+3+4+5+1+4+9+16+25+64=3*5+5*11+64=134

1.10.5 Somme de n cellules : `sum`

On tape dans `D0` :

```
10
```

On tape dans `D1` :

```
=sum(current_sheet(j,1),j,1,D0)
```

On obtient dans `D1` la somme des cellules `B1` à `B10` :

```
385
```

En effet `current_sheet(j,1)` désigne la cellule de la colonne `B` (colonne 1) et de la ligne `j` et puisque `j` varie de `1` à `D0` qui vaut `10`, donc dans `D1` on a la somme des cellules de `B1` à `B10`.

1.10.6 Moyenne des cellules d'un sous-tableau : `mean`

Si `mean` a comme argument une liste, `mean` calcule la moyenne des éléments de cette liste.
Si `mean` a comme argument deux listes, `mean` calcule la moyenne des éléments de la première listes, pondérés par les éléments de la seconde liste.
Pour traiter les exemples, on remplit la colonne `A` par `0,1,2,..,n` et la colonne `B` par `0,1,4,..,n^2` (cf 1.10.4)
Remarque
Une cellule remplie avec une chaine de caractères vide n'est pas prise en compte : on peut ainsi faire les calculs sur les réponses effectives à un questionnaire, par exemple, et ainsi de ne pas tenir compte des questionnaires non complètement remplis.

Moyenne des cellules d'un sous-tableau d'effectif 1

La commande `mean` permet de calculer la moyenne de plusieurs cellules situées dans un sous-tableau.
Après avoir selectionné `C0`, on tape, par exemple, dans la ligne de commande du tableur :

```
=mean(A0:B5)
```

On obtient dans `C0` :

```
35/6
```

en effet : `A0:B5` désigne la liste `[0,0,1,1,2,4...,5,25]` `mean(A0:B5)` renvoie donc la valeur de :
(1+2+3+4+5+1+4+9+16+25)/12=70/12=35/6=5.83333333333.
Mais dans une ligne d'entrée, si on tape :
`mean([[0,0],[1,1],[2,4],[3,9],[4,16],[5,25]])`
On obtient :
`[5/2,55/6]`

Moyenne des cellules d'un sous-tableau avec effectifs

La commande `mean` permet de calculer la moyenne des valeurs de cellules situées dans un sous-tableau pondérée par un autre sous-tableau.
Après avoir selectionné `D0`, on tape, par exemple, dans la ligne de commande du tableur :

```
=mean(A3:B5,A0:B2)
```

On obtient dans `D0` :

```
65/4
```

en effet, `mean(A3:B5,A0:B2)` calcule :
(4*1+5*2+16*1+25*4)/(1+2+1+4)=130/8=65/4. Mais dans une ligne d'entrée, si on tape :
`mean([[3,9],[4,16],[5,25]],[[0,0],[1,1],[2,4]])`

On obtient :
`[14/3,116/5]`
En effet :
`mean([3,4,5],[0,1,2])=14/3`
`mean([9,16,25],[0,1,4])=116/5`
ce sont les moyennes des colonnes du premier argument pondérée par les colonnes du deuxième argument.
Après avoir selectionné `D0`, on tape, par exemple, dans la ligne de commande du tableur :

```
=mean(A0:A5,B0:B5)
```

On obtient dans `D0` :

```
45/11
```

en effet, `mean(A0:A5,B0:B5)` calcule :
(1*1+2*4+3*9+4*16+5*25)/(1+4+9+16+25)=45/11.

1.10.7 Écart-type des cellules d'un sous-tableau : `stddev`

Il y a deux cas :
Si `stddev` a comme argument une liste, `stddev` calcule l'écart-type des éléments de ce sous-tableau.
Si `stddev` a comme argument deux listes, `stddev` calcule l'écart-type des éléments de la première liste, pondérés par les éléments de la seconde liste.
Pour traiter les exemples, on remplit la colonne `A` par `0,1,2,..,n` et la colonne `B` par `0,1,4,..,n^2` (cf 1.10.4)

Écart-type des cellules d'un sous-tableau d'effectif 1

La commande `stddev` permet de calculer l'écart type des valeurs de cellules situées dans un sous-tableau.
Après avoir selectionné `C1`, on tape, par exemple, dans la ligne de commande du tableur :

```
=stddev(A0:B5)
```

On obtient dans `C1` :

```
sqrt(1877/36)
```

Mais dans une ligne d'entrée, si on tape :
`stddev([[0,0],[1,1],[2,4],[3,9],[4,16],[5,25]])`
On obtient :
`[sqrt(35/12),sqrt(2849/36)]`

Écart-type des cellules d'un sous-tableau avec effectifs

La commande `stddev` permet de calculer l'écart-type des valeurs de cellules situées dans un sous-tableau pondérées par un autre sous-tableau.
Après avoir selectionné `D1`, on tape, par exemple, dans la ligne de commande du tableur :

```
=stddev(A3:B5,A0:B2)
```

On obtient dans `D1` :

```
sqrt(1419/16)
```

Mais dans une ligne d'entrée, si on tape :
`stddev([3,9],[4,16],[5,25]],[[0,0],[1,1],[2,4]])`
On obtient :
`[sqrt(2/9),sqrt(324/25)]`

Estimation de l'écart-type de la population mère : `stdDev`

`stdDev` a les mêmes arguments que `stddev`. Si le premier argument a comme dimension n, on a la relation :
`n:=size(L);stdDev(L)=stddev(L)*sqrt(n/(n-1))` La commande `stdDev` permet de calculer une estimation de l'écart-type de la population mère à partir d'un échantillon d'ordre n et dont les valeurs sont mises en argument.
Après avoir selectionné `C2`, on tape, par exemple, dans la ligne de commande du tableur :

```
=stdDev(A0:B5)
```

On obtient dans `C2` :

```
sqrt(1877/33)
```

ici $n = 6 * 2 = 12$ et $12/11 * 1877/36 = 1877/33$ Après avoir selectionné `D2`, on tape, par exemple, dans la ligne de commande du tableur :

```
=stdDev(A3:B5,A0:B2)
```

On obtient dans `D2` :

```
sqrt(1419/14)
```

ici $n = 8$ et $8/7 * 1419/16 = 1419/14$

1.10.8 Variance des cellules d'un sous-tableau : `variance`

Il y a deux cas :
Si `variance` a comme argument une liste, `variance` calcule la variance des éléments de cette liste.
Si `variance` a comme argument deux listes, `variance` calcule la variance des éléments de la première liste, pondérés par les éléments de la seconde liste. La variance est le carré de l'écart-type.
Pour traiter les exemples, on remplit la colonne `A` par `0,1,2,..,n` et la colonne `B` par `0,1,4,..,n^2` (cf 1.10.4).

Variance des cellules d'un sous-tableau d'effectif 1

La commande `variance` permet de calculer la variance des valeurs de cellules situées dans un sous-tableau.
Après avoir selectionné `C2`, on tape, par exemple, dans la ligne de commande du tableur :

```
=variance(A0:B5)
```

On obtient dans `C2` :

```
187/36
```

Variance des cellules d'un sous-tableau avec effectifs

La commande `variance` permet de calculer la variance des valeurs de cellules situées dans un sous-tableau pondérée par un autre sous-tableau.
Après avoir selectionné `D2`, on tape, par exemple, dans la ligne de commande du tableur :

```
=variance(A3:B5,A0:B2)
```

On obtient dans `C6` :

```
1419/16
```

1.10.9 La médiane : `median`

Il y a deux cas :
Si `median` a comme argument une liste, `median` calcule la médiane des éléments de cette liste.
Si `median` a comme argument deux listes, `median` calcule la médiane des éléments de la première liste, pondérée par les éléments de la seconde liste.
La médiane est l'élement M_e de la série à partir du lequel la fréquence cumulée de M_e égale ou dépasse 0.5 (on appelle fréquence cumulée d'une valeur `a` la somme des fréquences de `t` pour toutes les valeurs `t` $\leq$ `a`)
Un sous-tableau est transformé en une liste quand on le désigne dans un tableur par :
"référence de sa première case" `:` "référence de sa dernière case".
Pour traiter les exemples, on remplit la colonne `A` par `0,1,2,..,n` et la colonne `B` par `0,1,4,..,n^2` (cf 1.10.4).

Médiane des cellules d'un sous-tableau d'effectif 1

La commande `median` permet de calculer la médiane des valeurs de cellules situées dans un sous-tableau.
Après avoir selectionné `C3`, on tape, par exemple, dans la ligne de commande du tableur :

```
=median(A0:A10)
```

On obtient dans `C3` :

```
5.0
```

Médiane des cellules d'un sous-tableau avec effectifs

La commande `median` permet de calculer la médiane des valeurs de cellules situées dans un sous-tableau pondérée par un autre sous-tableau.
Après avoir selectionné `D3`, on tape, par exemple, dans la ligne de commande du tableur :

```
=median(A0:A10,B0:B10)
```

On obtient dans `C7` :

```
8
```

1.10.10 Le premier quartile : `quartile1`

Il y a deux cas :
Si `quartile1` a comme argument un sous tableau, `quartile1` calcule le premier quartile des éléments de ce sous-tableau.
Si `quartile1` a comme argument deux sous tableaux, `quartile1` calcule le premier quartile des éléments du premier sous-tableau, pondérés par les éléments du second sous-tableau.
Le premier quartile est l'élement Q_1 de la série à partir du lequel la fréquence cumulée de Q_1 égale ou dépasse 0.25.
Pour traiter les exemples, on remplit la colonne `A` par `0,1,2,..,n` et la colonne `B` par `0,1,4,..,n^2` (cf 1.10.4).

Le premier quartile des cellules d'un sous-tableau d'effectif 1

La commande `quartile1` permet de calculer le premier quartile des valeurs de cellules situées dans un sous-tableau.
Après avoir selectionné `C4`, on tape, par exemple, dans la ligne de commande du tableur :

```
=quartile1(A0:A10)
```

On obtient dans `C4` :

```
2.0
```

Le premier quartile des cellules d'un sous-tableau avec effectifs

La commande `quartile1` permet de calculer le premier quartile des valeurs de cellules situées dans un sous-tableau pondérée par un autre sous-tableau.
Après avoir selectionné `D4`, on tape, par exemple, dans la ligne de commande du tableur :

```
=quartile1(A0:A10,B0:B10)
```

On obtient dans `D4` :

```
7
```

1.10.11 Le troisième quartile : `quartile3`

Il y a deux cas :
Si `quartile3` a comme argument un sous tableau, `quartile3` calcule le troisième quartile des éléments de ce sous-tableau.
Si `quartile3` a comme argument deux sous tableaux, `quartile3` calcule le troisième quartile des éléments du premier sous-tableau, pondérés par les éléments du second sous-tableau. Le troisième quartile est l'élement Q_3 de la série à partir du lequel la fréquence cumulée de Q_3 égale ou dépasse 0.75.
Pour traiter les exemples, on remplit la colonne `A` par `0,1,2,..,n` et la colonne `B` par `0,1,4,..,n^2` (cf 1.10.4).

Le troisième quartile des cellules d'un sous-tableau d'effectif 1

La commande `quartile3` permet de calculer le troisième quartile des valeurs de cellules situées dans un sous-tableau.
Après avoir selectionné `C5`, on tape, par exemple, dans la ligne de commande du tableur :

```
=quartile3(A0:A10)
```

On obtient dans `C5` :

```
8.0
```

Le troisième quartile des cellules d'un sous-tableau avec effectifs

La commande `quartile3` permet de calculer le troisième quartile des valeurs de cellules situées dans un sous-tableau pondérées par un autre sous-tableau.
Après avoir selectionné `D5`, on tape, par exemple, dans la ligne de commande du tableur :

```
=quartile3(A0:A10,B0:B10)
```

On obtient dans `D5` :

```
10
```

1.10.12 Les valeurs indiquant la répartition : `quartiles`

Il y a deux cas :
Si `quartiles` a comme argument un sous tableau, `quartiles` calcule la matrice colonne contenant le minimum, le premier quartile, la médiane, le troisième quartile et le maximum des éléments de ce sous-tableau.
Si `quartiles` a comme argument deux sous tableaux, `quartiles` calcule la matrice colonne contenant le minimum, le premier quartile, la médiane, le troisième quartile et le maximum des éléments du premier sous-tableau, pondérés par les éléments du second sous-tableau.
Pour traiter les exemples, on remplit la colonne `A` par `0,1,2,..,n` et la colonne `B` par `0,1,4,..,n^2` (cf 1.10.4).

Valeurs indiquant la répartition des cellules d'un sous-tableau d'effectif 1

La commande `quartiles` permet de calculer la matrice colonne contenant le minimum, le premier quartile, la médiane, le troisième quartile et le maximum des valeurs de cellules situées dans un sous-tableau.
Après avoir selectionné `C6`, on tape, par exemple, dans la ligne de commande du tableur :

```
=quartiles(A0:A10)
```

On obtient dans `C6` :

```
[[0.0],[2.0],[5.0],[8.0],[10.0]]
```

Valeurs indiquant la répartition des cellules d'un sous-tableau avec effectifs

La commande `quartiles` permet de calculer la matrice colonne contenant le minimum, le premier quartile, la médiane, le troisième quartile et le maximum des valeurs de cellules situées dans un sous-tableau pondérées par un autre sous-tableau.
Après avoir selectionné `D6`, on tape, par exemple, dans la ligne de commande du tableur :

```
=quartiles(A0:A10,B0:B10)
```

On obtient dans `D6` :

```
[[1.0],[7.0],[8.0],[10.0],[10.0]]
```

1.11 Les fonctions statistiques à deux variables du tableur

1.11.1 Les fonctions graphiques

On peut utiliser directement le graphique depuis le tableur : toutes les fonctions graphiques peuvent être utilisées dans une cellule. Pour avoir un nuage de points on utilise `scatterplot` et pour tracer une ligne polygonal on utilise `polygonplot`.
Depuis le tableur, on peut se servir du menu du tableur `Statistiques`▶`2-d` et remplir la boite de dialogue correspondant à l'item choisi : vous devez choisir la cellule cible (c'est dans cette cellule que s'inscrira la commande graphique), vous pouvez éventuellement modifier la sélection (en mettant par exemple `A1..B6` dans `cellules entrée`), vous pouvez cocher (resp ne pas cocher) `valeur` pour que la commande graphique ait pour argument les valeurs (resp les références) de la sélection : donc si `valeur` n'est pas cochée un changement de valeur dans une cellule de la sélection aura pour conséquence un changement du graphique, si on est en mode automatique ou si on appuie sur le bouton `eval`.
Ou bien, dans la ligne de commande du tableur ou dans une ligne de commande, on utilise `scatterplot` et `polygonplot` du menu `Math`▶`Stats`▶`2-d` et on recopie les arguments en les sélectionnant, soit en se servant de la souris (ou de la `case de sélection` et de la touche `coller` lorsque les colonnes que l'on veut recopier ne sont pas consécutives), soit on sauve cette sélection avec le menu

du tableur Fich▶Sauver sélection vers variable, on donne un nom par exemple A, puis on tape polygonplot(A) ou scatterplot(A) : dans ce cas on travaille avec les valeurs de la sélection.

1.11.2 La covariance avec effectif 1 : covariance

covariance calcule la covariance numérique de plusieurs cellules situées dans deux sous-tableaux de même dimension.
Si T= t_j est le premier argument et B= b_j le deuxième argument, la covariance covariance(T,B) est alors définie par :
$cov(T,B) = \frac{1}{N}\sum_j (t_j - m_T)(b_j - m_B)$
où m_T (resp m_B) est la moyenne des éléments t_j de T (resp b_j de B) et N le nombre d'éléments de T.
Pour traiter les exemples, on remplit la colonne A par 0,1,2,..,n et la colonne B par 0,1,4,..,n^2 (cf 1.10.4).
On tape lorsque C0 est en surbrillance :

```
=covariance(A1:A4,B1:B4)
```

On obtient dans C0 :

```
25/4
```

Dans une ligne d'entrée, on tape :

```
covariance([1,2,3,4],[1,4,9,16])
```

Ou on tape :

```
covariance([[1,1],[2,4],[3,9],[4,16]])
```

On obtient :

```
25/4
```

1.11.3 La corrélation linéaire avec effectif 1 : correlation

correlation calcule la corrélation linéaire numérique de plusieurs cellules situées dans deux sous-tableaux de même dimension.
Si T= t_j est le premier argument et B= b_j le deuxième argument, la corrélation correlation(T,B) est alors : $\frac{cov(T,B)}{\sigma(T)\sigma(B)}$
où $\sigma(T)$ (resp $\sigma(B)$) est l'écart-type des éléments de T (resp B) et $cov(T,B)$ est la covariance de T et de B.

Dans le tableur

Pour traiter les exemples, on remplit la colonne A par 0,1,2,..,n et la colonne B par 0,1,4,..,n^2 (cf 1.10.4).
On tape lorsque C1 est en surbrillance :

```
=correlation(A1:A4,B1:B4)
```

On obtient dans `C1` :

```
25/sqrt(645)
```

on a en effet :
`covariance(A1:A4,B1:B4)=25/4`,
`stddev(A1:A4)=sqrt(5)/2` et,
`stddev(B1:B4)=sqrt(129)/2`
et `129*5=645`

Dans une ligne d'entrée
On tape :

```
correlation([1,2,3,4],[1,4,9,16])
```

Ou on tape :

```
correlation([[1,1],[2,4],[3,9],[4,16]])
```

On obtient :

```
25/sqrt(645)
```

1.11.4 La covariance et la corrélation linéaire avec effectifs : `covariance` et `correlation`

— Si les couples $a[j], b[j]$ ont pour effectif $n[j]$ ($j = 0..p-1$), `covariance` (resp `correlation`) a pour argument trois listes a, b, n de même longueur p, ou une matrice composée trois colonnes a, b, n et de p lignes $[a[j], b[j], n[j]]$.
`covariance` (resp `correlation`) calcule la covariance (resp corrélation) numérique des deux premières listes pondérées par la liste donnée comme dernier argument ou des deux colonnes de cette matrice pondérées par la troisiéme colonne.
Dans une ligne d'entrée
On tape dans une ligne d'entrée :

```
covariance([1,2,3,4],[1,4,9,16],[3,1,5,2])
```

Ou on tape :

```
covariance([[1,1,3],[2,4,1],[3,9,5],[4,16,2]])
```

On obtient :

```
662/121
```

On tape dans une ligne d'entrée :

```
correlation([1,2,3,4],[1,4,9,16],[3,1,5,2])
```

Ou on tape :

```
correlation([[1,1,3],[2,4,1],[3,9,5],[4,16,2]])
```

On obtient :

```
662/(180*sqrt(14))
```

Dans le tableur
On remplit les colonnes `A`, `B`, `C`.
On tape dans `A0` :

```
[[1,1,3],[2,4,1],[3,9,5],[4,16,2]]
```

On tape dans `D0` :

```
=covariance(list2mat(A0:C3,3))
```

On obtient dans `D0` :

```
662/121
```

On tape dans `D1` :

```
=correlation(list2mat(A0:C3,3))
```

On obtient dans `E1` :

```
662/(180*sqrt(14))
```

— Si les couples $a[j], b[k]$ ont pour effectif $N[j,k]$ lorsque $j = 0..p-1$ et $k = 0..q-1$, `covariance` (resp `correlation`) a pour argument deux listes a, b de longueurs respectives p et q et une matrice N de p lignes et q colonnes.
`covariance` (resp `correlation`) calcule la covariance (resp corrélation) numérique des éléments de deux listes pondérés par un tableau donné comme troisième argument.

Exercice
Soient `X=[1,2]`, `Y=[11,13,14]` et `N=[[3,4,5],[12,1,2]]`.
Calculer la covariance et la corrélation de `X`, `Y` sachant qu'il y a `N=` $N_{j,k}$ couples X_j, Y_k.
Dans une ligne d'entrée
On tape :

```
covariance([1,2],[11,13,14],[[3,4,5],[12,1,2]])
```

On obtient :

```
-83/243
```

On tape :

```
correlation([1,2],[11,13,14],[[3,4,5],[12,1,2]])
```

On obtient :

```
-83/160
```

On a :
`simplify(stddev([1,2],[12,15]))=2*sqrt(5)/9`
`simplify(stddev([11,13,14],[15,5,7]))=16*sqrt(5)/27`
et on a bien :
`-83/243=-83/160*(sqrt(20)/9)*(16*sqrt(5)/2)`
Dans le tableur
On peut disposer les données selon un tableau à double entrée à condition de rajouter `-1` comme dernier argument aux fonctions `covariance` et `correlation`.
On tape :
dans `A0` :
"X\ Y" (c'est pour l'esthétique)
dans `A1` :
1
dans `A2` :
2

dans B0,C0,D0 :
11,13,14
dans B1,C1,D1 :
3,4,5
dans B2,C2,D2 :
12,1,2
Calcul de la covariance ou de la corrélation dans le tableur :
On tape dans E0 :

```
=covariance(list2mat(A0:D2,4),-1)
```

On obtient dans E0 :

```
-83/243
```

On tape dans E1 :

```
=correlation(list2mat(A0:D2,4),-1)
```

On obtient dans E1 :

```
-83/160
```

Remarque
On peut bien sûr faire le même calcul dans une ligne d'entrée :
On sélectionne avec la souris A0..2,B,C,D, puis on tape :
covariance(,-1)
puis, on met le curseur à l'endroit de l'argument manquant, puis on appuie sur coller, ou on tape :

```
covariance([["x\y",11,13,14],
  [1,3,4,5],[2,12,1,2]],-1)
```

On obtient :

```
-83/243
```

On tape :

```
correlation([["x\y",11,13,14],
  [1,3,4,5],[2,12,1,2]],-1)
```

On obtient :

```
-83/160
```

Attention
Dans une cellule du tableur on ne peut pas désigner un sous-tableau ou une matrice par :
"référence de la première case de la matrice" **:** "référence de sa dernière case de la matrice".
en effet si dans la cellule C0 on tape :
=A0:B4 et on obtient dans la cellule C0 la liste :
[A0,B0,A1,B1,A2,B2] c'est à dire la matrice "aplatie".

Pourtant, si les couples A= $a[j]$ et B= $b[j]$ ont pour effectif N= $n[j]$ ($j = 0..p-1$), la covariance (resp la corrélation) de A,B avec effectifs N peuvent se calculer dans une ligne d'entrée, mais aussi dans le tableur même si les données ne figurent pas dans des colonnes consécutives.
Lorsque les colonnes sont consécutives, on peut reconstituer la matrice en utilisant `list2mat`, par exemple si on met les valeurs de A dans la colonne A, les valeurs de B dans la colonne B et les effectifs N dans la colonne C on tape dans la cellule E1 `=list2mat(A0:C5,6,3)` et on obtient dans la cellule E1 la matrice cherchée ayant 6 lignes et 3 colonnes A,B,C.
Lorsque les colonnes ne sont pas consécutives par exemple si on met les effectifs N= $n[j]$ dans la colonne D on tape alors pour avoir une matrice dans la cellule E1 : `=tran([A0:A5,B0:B5,D0:D5])` et on obtient dans la cellule E1 la matrice cherchée ayant 6 lignes et 3 colonnes A,B,D.

1.11.5 La régression linéaire : `linear_regression`

Pour approcher les données par la droite des moindres carrés qui a pour équation $y = mx + b$, on utilise `linear_regression`.
`linear_regression` a les mêmes arguments que `covariance`.
Si `linear_regression` a comme argument la liste X des x_j et la liste Y des y_j, `linear_regression` renvoie (m, b) tel que $y \simeq m * x + b$.
Pour traiter une régression linéaire à 2 ou plusieurs variables on se reportera à la section 1.11.8.

Avec le tableur Pour traiter les exemples, on remplit la colonne A par `0,1,2,..,n` et la colonne B par `0,1,4,..,n^2` (cf 1.10.4).
Puis, on tape dans la case C0 :

```
linear_regression(A1:A4,B1:B4)
```

On obtient dans la case C0 :

```
5,-5
```

Dans une ligne d'entrée
On tape :

```
linear_regression([[1,1],[2,4],[3,9],[4,16]])
```

Ou on tape

```
linear_regression([1,2,3,4],[1,4,9,16])
```

On obtient :

```
(5,-5)
```

ce qui veut dire que $y = 5x - 5$ est la droite qui approche au mieux les points de coordonnées : $(1, 1), (2, 4), (3, 9), (4, 16)$.
Remarques

— La deuxième droite de régression
Si on tape (on échange X et Y) :

```
linear_regression([1,4,9,16],[1,2,3,4])
```

Ou on tape :

```
linear_regression([[1,1],[4,2],[9,3],[16,4]])
```

On obtient la deuxième droite de régression :

```
25/129,45/43
```

On tape :

```
evalf(25/129,45/43)
```

On obtient :

```
(0.193798449612,1.04651162791)
```

— Ajustement linéaire et corrélation linéaire
Si $R2$ est le carré du coefficient de corrélation linéaire de X et de Y si $m1$ (resp $m2$) est la pente de la première (resp deuxième) droite de régression linéaire on a :

```
R2=m1*m2
```

On tape :

```
normal(correlation([1,2,3,4],[1,4,9,16])^2)
```

On obtient :

```
125/129
```

On tape :

```
5* 25/129
```

On obtient :

```
125/129
```

On tape :

```
evalf(125/129)
```

On obtient :

```
0.968992248062
```

Autre exemples On suppose que l'on a les couples (x_j, y_j) avec :
$x = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]$ et
$y = [7.3, 9.53, 12.47, 16.3, 21.24, 27.73, 36.22, 47.31, 61.78, 80.68, 105]$
On tape :

```
X:=[0,1,2,3,4,5,6,7,8,9,10]

Y:=[7.3,9.53,12.47,16.3,21.24,27.73,36.22,

47.31,61.78,80.68,105]

Z:=log(Y)

linear_regression(X,Z)
```

On obtient :

```
(0.266729219953,1.98904252589)
```

c'est donc la fonction linéaire d'équation $z = \ln(y) = 0.267 * x + 1.99$ qui approche au mieux les données.

On suppose qu'il y a n_j couples (x_j, y_j) avec :
$x = [1, 2, 3, 4]$, $y = [1, 4, 9, 16]$, et $n = [3, 1, 5, 2]$
On tape :

```
linear_regression([1,2,3,4],[1,4,9,16],[3,1,5,2])
```

On obtient :

```
(331/70, -22/5)
```

c'est donc la fonction linéaire d'équation $y = 331 * x/70 - 22/5$ qui approche au mieux les données.

On suppose qu'il y a $n_{j,k}$ couples (x_j, y_k) avec :
$x = [1, 2]$, $y = [11, 13, 14]$, et $n = [[3, 4, 5], [12, 1, 2]]$
On tape :

```
linear_regression([1,2],[11,13,14],[[3,4,5],[12,1,2]])
```

Ou on tape :

```
linear_regression([["x\y",11,13,14],[1,3,4,5],
                   [2,12,1,2]],-1)
```

On obtient :

```
(-83/60,143/10)
```

c'est donc la fonction linéaire d'équation $y = -83 * x/60 + 143/10$ qui approche au mieux les données. On calcule le coefficient de corrélation On tape :

```
normal(correlation([1,2],[11,13,14],[[3,4,5],[12,1,2]])2)
```

On obtient :

```
6889/25600
```

Donc $R2 \simeq 0.2691015625$ ce qui ne justifie pas un ajustement linéaire.

1.11.6 Ajustement linéaire et corrélation linéaire

Si $R2$ est le carré du coefficient de corrélation linéaire de X et de Y si $m1$ (resp $m2$) est la pente de la première (resp deuxième) droite de régression linéaire on a :

```
R2=m1*m2
```

On sait que le coefficient de correlation est un réel entre -1 et 1 donc $0 \leq R2 \leq 1$. La valeur de $R2$ va nous dire si la forme du nuage de points justifie un ajustement lineaire. Il y a une forte corrélation linéaire lorsque $\sqrt{1 - R2} \leq 0.5$ i.e. lorsque $R2 \geq 0.75$.
On a :

- Si $R2 = 0$, l'ajustement lineaire n'est pas justifié
  cela n'exclut pas une dépendance entre X et Y l'ensemble des points $M - j, k$ peut être voisin d'une courbe, mais l'ensemble des points ne peut pas être ajusté par une droite,
- Si $R2 < 0.75$, l'ajustement lineaire est bon
- Si $R2 \geq 0.75$, l'ajustement lineaire n'est pas bon
- Si $R2 = 1$ les points sont alignés et les deux droites de régression sont confondues.

1.11.7 Le graphe de la régression linéaire : `linear_regression_plot`

Pour dessiner la droite des moindres carrés : la droite qui approche au mieux les données et qui a pour équation $y = mx+b$, on utilise `linear_regression_plot`. `linear_regression_plot` a les mêmes arguments que `covariance`.
On tape :

```
linear_regression_plot([1,2,3,4],[1,4,9,16],[3,1,5,2])
```

On obtient :

Le graphe de la droite d'équation $y = 331 * x/70 - 22/5$ ou $y = 4.3 * x - 4.4$ et $R2 = 0.966$

c'est donc la fonction linéaire d'équation $y = 331 * x/70 - 22/5$ qui approche au mieux les données.

1.11.8 La régression linéaire à 2 ou plusieurs variables

Le principe

Supposons que l'on observe 3 variables `(X,Y,Z)`, et que l'on veut savoir comment `Z` dépend linéairement de `X` et de `Y`.
On a par exemple observé n triplés x_j, y_j, z_j pour $j = 0..n-1$. On cherche c, a, b pour que le plan $z = a * x + b * y + c$ approche au mieux les données.
Posons $E = \sum_{j=0}^{n-1}(z_j - a * x_j - b * y_j - c)^2$.
On cherche c, a, b pour que E soit minimum c'est à dire pour que :
$\frac{\partial E}{\partial a} = -2 * \sum_{j=0}^{n-1} x_j * (z_j - a * x_j - b * y_j - c) = 0$
$\frac{\partial E}{\partial b} = -2 * \sum_{j=0}^{n-1} y_j * (z_j - a * x_j - b * y_j - c) = 0$
$\frac{\partial E}{\partial c} = -2 * \sum_{j=0}^{n-1} (z_j - a * x_j - b * y_j - c) = 0$
On a donc à résoudre un système de 3 équations à 3 inconnues c, a, b.
Soit U la matrice de n lignes et 3 colonnes ayant comme ligne j : $[1, x_j, y_j]$ avec $j = 0..n-1$.
Le système à résoudre est :

$$\begin{bmatrix} n & \sum x_j & \sum y_j \\ \sum x_j & \sum x_j^2 & \sum x_j y_j \\ \sum y_j & \sum x_j y_j & \sum y_j^2 \end{bmatrix} \begin{bmatrix} c \\ a \\ b \end{bmatrix} = \begin{bmatrix} \sum z_j \\ \sum x_j z_j \\ \sum y_j z_j \end{bmatrix} = \texttt{tran(U)} \begin{bmatrix} z_0 \\ .... \\ z_{n-1} \end{bmatrix}$$

On remarque que la matrice associée au système précédent s'écrit :

$$\texttt{A} = \texttt{tran(U)} * \texttt{U} = \begin{bmatrix} n & \sum x_j & \sum y_j \\ \sum x_j & \sum x_j^2 & \sum x_j y_j \\ \sum y_j & \sum x_j y_j & \sum y_j^2 \end{bmatrix}$$

La solution c, a, b du système est donc : `inv(A)*tran(U)*Z`

Avec le tableur

Suppopsons que l'on a mis les données, dans le tableur (par exemple $n = 192$) comme ceci :
- en `A` une colonne remplit avec des `1`,
- en `B` une colonne remplit avec les x_j et représentant `X`,

- en C une colonne remplit avec les y_j et représentant Y,
- en D une colonne remplit avec les z_j et représentant Z.
Dans la case de sélection on marque :
A0..A191,B,C
On utilise le menu du tableur Fich*blacktriangleright*Sauver selection vers variable et on tape U comme nom de variable. Cela définira la matrice U égale à la sélection.
Puis, on met dans la case de sélection D0..D191 et on utilise à nouveau le menu du tableur Fich*blacktriangleright*Sauver selection vers variable et on tape Z comme nom de variable. Cela définira le vecteur Z égale à la sélection.
On définit A en tapant : A:=tran(U)*U
Il ne reste plus qu'à taper dans une ligne de commande :
(c,a,b):=col(inv(A)*tran(U)*Z,0)
pour définir c,a,b.
Remarque
Bien sûr il faut que la matrice A=tran(U)*U soit inversible ! ! ! !
On peut aussi taper :
B:=border(A,op(-tran(Z)*U)) puis
C:=rref(B), puis résoudre C*[[c],[a],[b]]=0.

1.11.9 La régression exponentielle : exponential_regression

Pour approcher les données par une fonction exponnentielle qui a pour équation $y = b * \exp(m * x) = b * a^x$, on utilise exponential_regression.
exponential_regression a les mêmes arguments que covariance.
Si exponential_regression a comme argument la liste X des x_j et la liste Y des y_j, exponential_regression renvoie (a, b) tel que $y \simeq ba^x$.
On tape dans la case C1 :

```
evalf(exponential_regression(A1:A4,B1:B4))
```

ou on tape dans une ligne d'entrée :

```
exponential_regression([[1.0,1],[2,4],[3,9],[4,16]])
```

On obtient :

```
2.49146187923,0.5
```

On tape dans une ligne d'entrée :

```
exponential_regression([0,1,2,3,4,5,6,7,8,9,10],[7.3,
9.53,12.47,16.3,21.24,27.73,36.22,47.31,61.78,80.68,105])
```

On obtient :

```
(1.30568684451, 7.30853268031)
```

c'est donc la fonction expopnentielle d'équation :
$y = 7.30853268031 * (1.30568684451)^x$ qui approche au mieux les données.
Remarque
On a :
$\exp(0.266729219953, 1.989042525894) = (1.30568684451, 7.30853268031)$

1.11.10 Le graphe de la régression exponentielle : `exponential_regression_plot`

Pour dessiner la fonction exponnentielle qui a pour équation $y = b * \exp(m * x) = b{*}a^x$, et qui approche au mieux les données, on utilise `exponential_regression_plot`.
`exponential_regression_plot` a les mêmes arguments que `covariance`.
On tape dans l'écran `geo` :

```
exponential_regression_plot([0,1,2,3,4,5,6,7,8,9,10],[7.3,
9.53,12.47,16.3,21.24,27.73,36.22,47.31,61.78,80.68,105])
```

On obtient :

Le graphe de la fonction expopnentielle d'équation
$y = 7.30853268031 * (1.30568684451)^x$

car c'est la fonction expopnentielle d'équation :
$y = 7.30853268031 * (1.30568684451)^x$ qui approche au mieux les données.

1.11.11 La régression logarithmique : `logarithmic_regression`

Pour approcher les données par une fonction logarithmique qui a pour équation $y = m \ln x + b$, on utilise `logarithmic_regression`.
`logarithmic_regression` a les mêmes arguments que `covariance`.
Si `logarithmic_regression` a comme argument la liste `X` des x_j et la liste `Y` des y_j, `logarithmic_regression` renvoie (m, b) tel que $y \simeq m * \ln x + b$.
On tape dans la case `C2` :

```
evalf(logarithmic_regression(A1:A4,B1:B4))
```

ou on tape dans une ligne d'entrée :

```
evalf(logarithmic_regression([[1,1],[2,4],[3,9],[4,16]]))
```

ou on tape dans une ligne d'entrée :

```
logarithmic_regression([[1.0,1],[2,4],[3,9],[4,16]])
```

On obtient :

```
10.1506450002,-0.564824055818
```

c'est donc la fonction logarithme d'équation :
$y = 10.1506450002 \ln(x) - 0.564824055818$
qui approche au mieux les données.

1.11.12 Le graphe de la régression logarithmique : `logarithmic_regression_plot`

Pour dessiner le graphe de la fonction logarithmique qui a pour équation $y = m \ln x + b$, et qui approche au mieux les données, on utilise `logarithmic_regression_plot`.
`logarithmic_regression_plot` a les mêmes arguments que `covariance`.
On tape dans l'écran `geo` :

```
logarithmic_regression_plot([[1.0,1],[2,4],[3,9],[4,16]])
```

On obtient :

Le graphe de la fonction logarithme d'équation
$y = 10.1506450002 \ln(x) - 0.564824055818$

car c'est la fonction logarithme d'équation :
$y = 10.1506450002 \ln(x) - 0.564824055818$
qui approche au mieux les données.

1.11.13 La régression polynomiale : `polynomial_regression`

Pour approcher les données par une fonction polynomiale d'équation $y = a_0x^n + .. + a_n$, on utilise `polynomial_regression`.
`polynomial_regression` a les mêmes arguments que `covariance`.
Si `polynomial_regression` a comme arguments la liste des x_j, la liste des y_j et le degré n du polynôme, `polynomial_regression` renvoie $[a_n, ..., a_0]$) tel que $y \simeq a_n * x^n + ... + a_0$.
On tape dans le tableur :

```
evalf(polynomial_regression(A1:A4,B1:B4,3))
```

ou on tape dans une ligne d'entrée :

```
evalf(polynomial_regression([[1,1],[2,4],[3,9],[4,16]],3))
```

ou on tape dans une ligne d'entrée :

```
polynomial_regression([[1.0,1],[2,4],[3,9],[4,16]],3)
```

On obtient :

```
[0.0,1.0,0.0,0.0]
```

c'est donc le polynôme d'équation $y = x^2$ qui approche au mieux les données.

1.11.14 La régression puissance : `power_regression`

Pour approcher les données par une fonction puissance d'équation $y = bx^m$, on utilise `power_regression`.
`power_regression` a les mêmes arguments que `covariance`.
Si `power_regression` a comme argument la liste des x_j et la liste des y_j, `power_regression` renvoie (m, b) tel que $y \simeq bx^m$.
On tape dans le tableur :

```
evalf(power_regression(A1:A4,B1:B4))
```

ou on tape ddans une ligne d'entrée :

```
evalf(power_regression([[1,1],[2,4],[3,9],[4,16]]))
```

ou on tape dans une ligne d'entrée :

```
power_regression([[1.0,1],[2,4],[3,9],[4,16]])
```

On obtient :

```
[2.0,1.0]
```

c'est donc la fonction puissace d'équation $y = 1 * x^2$ qui approche au mieux les données.

1.11.15 Le graphe de la régression puissance : `power_regression_plot`

Pour approcher les données par une fonction puissance d'équation $y = bx^m$, on utilise `power_regression_plot`.
`power_regression_plot` a les mêmes arguments que `covariance`.
On tape dans l'écran `geo` :

```
power_regression_plot([[1.0,1],[2,4],[3,9],[4,16]])
```

On obtient :

Le graphe de la fonction puissace d'équation $y = 1 * x^2$

car c'est la fonction puissace d'équation $y = 1 * x^2$ qui approche au mieux les données.

1.12 Définition de fonctions de `Xcas`

1.12.1 Définition de fonction de répartition

Définition
On appelle fonction de répartition d'une variable aléatoire x sur l'espace probabilisé Ω la fonction F définie pour tout x réel par :

$$F(x) = Prob(X \leq x)$$

Propriétés

- F est croissante,
- $\lim_{x->-\infty} = 0$,
- $\lim_{x->+\infty} = 1$,
- $Prob(a < X \leq b) = F(b) - F(a)$.

Définition
On dit qu'une variable aléatoire X est absolument continue si et seulement si il existe une fonction f, appelée densité de probabilité de X, telle que :

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

1.12.2 Les fonctions de répartition et de répartition inverse

Règle

Le nom de la fonction de répartition d'une loi est le nom de la loi, suivi par _cdf, et pour la fonction de répartition inverse par _icdf : cdf =cumulated distribution function = fonction de répartition.

Les premiers paramètres sont les paramètres de la loi et le dernier paramètre le nom de la variable.

On définit les fonctions suivantes :

normald(t) par $\frac{\exp(-(t^2)/2)}{\sqrt{2*pi}}$

normald(μ,σ,t) par $\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{1}{2}(\frac{(t-\mu)}{\sigma})^2)$

normal_cdf(x)= $Proba(X \leq x)$ avec $X \in \mathcal{N}(0,1)$: c'est la fonction de répartition de la loi normale centrée réduite.

normal_icdf(t) =h équivaut à $Proba(X \leq h) = t$ avec $X \in \mathcal{N}(0,1)$: c'est l'inverse de la fonction de répartition de la loi normale centrée réduite.

normal_cdf(μ,σ,x)=$Proba(X \leq x)$ avec $X \in \mathcal{N}(\mu,\sigma)$: c'est la fonction de répartition de la loi normale de moyenne μ et d'écart-type σ.

normal_icdf(μ,σ,t) =h équivaut à $Proba(X \leq h) = t$ avec $X \in \mathcal{N}(\mu,\sigma)$: c'est l'inverse de la fonction de répartition de la loi normale de moyenne μ et d'écart-type σ.

normal_cdf(a,b)=normal_cdf(b)-normal_cdf(a)

normal_cdf(μ,σ,a,b)=normal_cdf(μ,σ,b)-normal_cdf(μ,σ,a)

binomial(n,k,p)=comb(n,k)*p^k*(1-p)^n-k

binomial_cdf(n,p,x)=$Proba(X \leq x)$ avec $X \in \mathcal{B}(n,p)$: c'est la fonction de répartition de la loi binomiale de paramètre n,p c'est à dire de moyenne np et d'écart-type $\sqrt{np(1-p)}$.

binomial_icdf(n,p,t)= h équivaut à $Proba(X \leq h) = t$ avec $X \in \mathcal{B}(n,p)$: c'est l'inverse de la fonction de répartition de la loi binomiale de paramètres n et p c'est à dire de moyenne np et d'écart-type $\sqrt{np(1-p)}$.

poisson(m,k)= exp(-m)*m^k/k!

poisson_cdf(μ,x)=$Proba(X \leq x)$ avec $X \in \mathcal{P}(\mu)$: c'est la fonction de répartition de la loi de Poisson de paramètre μ, c'est à dire de moyenne μ et d'écart-type μ.

poisson_cdf(μ,x1,x2)=poisson_cdf(μ,x2)-poisson_cdf(μ,x1)

poisson_icdf(μ,t)= h équivaut à $Proba(X \leq h) = t$ avec $X \in \mathcal{P}(\mu)$: c'est l'inverse de la fonction de répartition de la loi de Poisson de paramètre μ, c'est à dire de moyenne μ et d'écart-type μ.

student_cdf(n,x)=$Proba(X \leq x)$ avec $X \in \mathcal{T}(n)$: c'est la fonction de répartition de la loi de Student ayant n degrés de liberté.

student_icdf(n,t)=h équivaut à $Proba(X \leq h) = t$ avec $X \in \mathcal{T}(n)$: c'est l'inverse de la fonction de répartition de la loi de Student ayant n degrés de liberté.

chisquare_cdf(n,x)=$Proba(X \leq x)$ avec $X \in \chi^2(n)$: c'est la fonction de répartition de la loi du χ^2 ayant n degrés de liberté.

chisquare_icdf(n,t)=h équivaut à $Proba(X \leq h) = t$ avec $X \in \chi^2(n)$: c'est l'inverse de la fonction de répartition de la loi du χ^2 ayant n degrés de liberté.

fisher_cdf(n,k,x)=snedecor_cdf(n,k,x)=$Proba(X \leq x)$ lorsque

$X \in \mathcal{F}(n,k)$: c'est la fonction de répartition de la loi de Fisher ayant n, k degrés de liberté.
`fisher_icdf(n,k,t)`=`snedecor_icdf(n,k,t)`=h ce qui veut dire que $Proba(X \leq h) = t$ avec $X \in \mathcal{F}(n,k)$: c'est l'inverse de la fonction de répartition de la loi de Fisher ayant n, k degrés de liberté.
`UTPC`,`UTPF`,`UTPN`,`UTPT` avec `C` pour χ^2, `F` pour Fisher, `N` pour Normale et `S` pour Student représentent le complément à 1 de la fonction de répartition correspondante.
Par exemple :
`UTPN(x)=1-normal_cdf(x)`
Mais attention : `UTPN(`μ, σ^2`,x)=1-normal_cdf(`μ, σ`,x)`

Chapitre 2

Résumé de probabilité

2.1 Rappel des différentes lois de probabilités

Les lois de probabilités sont des objets mathématiques qui permettent aux statisticiens de fabriquer des modéles pour décrire des phénomènes où le hasard intervient.
Une loi de probabilité est une distribution théorique de fréquences.
Soit Ω un ensemble muni d'une probabilité P. Une variable aléatoire X est une application définie sur Ω dans $\mathbb{R}$. X permet de transporter la loi P en la loi P' définie sur $\Omega' = X(\Omega)$: on a $P'(x_j) = P(X^{-1}(x_j)) = P(X = x_j)$. La loi P' est appelée loi de X.

2.2 Variable aléatoire discréte

Une variable aléatoire discréte X est une application dont la valeur est la valeur du caractère étudié, c'est à dire le résultat d'une épreuve.
Si X prend n valeurs $x_1, ...x - n$, on définit :

- m la moyenne ou, $E(X)$ l'espérance de X par :
  $m = E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$
- μ_2 le moment d'ordre 2 par :
  $\mu_2 = E(X^2) = \sum_{i=1}^{n} x_i^2 P(X = x_i))$
- $var(X)$ la variance de X par :
  $var(X) = \sum_{i=1}^{n} (x_i - m)^2 P(X = x_i) = E(X^2) - E(X)^2$
- σ l'écart type par : $\sigma = \sqrt{var(X)}$

Les n valeurs observées du caractère forment un échantillon de X d'ordre n : on dira que ces n valeurs sont les valeurs de n variables aléatoires $X_1, X_2, ..., X_n$ qui suivent la même loi que X. Par exemple, lorsqu'on lance un dé, on peut définir la variable aléatoire X qui est égale à la valeur de la face visible, donc X vaut 1 ou 2 ou ... 6.
Il y a trop de paramètres en jeu pour pouvoir déterminer le résultat du lancer d'un dé, mais à chaque lancer la valeur de X est définie.
Attention
Ce n'est pas parce que deux variables aléatoires suivent la même loi qu'elles sont égales. Par exemple, je lance deux dés, un rouge et un vert : la variable X_1 égale à

la face visible du dé rouge et la variable X_2 égale à la face visible du dé vert suivent toutes les deux une loi équirépartie de probabilité $p = 1/6$ sur $\{1, 2, 3, 4, 5, 6\}$.

La loi équirépartie

La loi équirépartie P sur un ensemble Ω à k éléments $\omega_0, \omega_1, \omega_{k-1}$ est définie par : $P(\omega_j) = 1/k$ pour tout $j = 0 \,..\, k-1$.
Choisir un élément de Ω selon la loi équirépartie P, c'est choisir au hasard un élément de Ω.
La variable aléatoire X suit une loi équirépartie si : X a pour valeurs x_0, x_1, x_{k-1} et si $P(X = x_j) = 1/k$ pour tout $j = 0 \,..\, (k-1)$. On a :
$\mu = E(X) = \frac{1}{k} \sum_{j=0}^{k-1} x_j$
$\sigma^2 = \sigma^2(X) = \frac{1}{k} \sum_{j=0}^{k-1} x_j^2 - \mu^2$

La loi de Bernouilli

La variable aléatoire X suit une loi de Bernouilli de probabilité p, si X vaut 1 ou 0 avec les probabilités respectives p et $1-p$.
On a alors :
$\mathtt{E(X)} = \mathtt{p}$,
$\mathtt{E(X^2)} = \mathtt{p}$,
$\sigma(\mathtt{X}) = \sqrt{\mathtt{p(1-p)}}$.

La loi binomiale

Si la variable aléatoire X suit une loi binomiale $\mathcal{B}(n, p)$, cela veut dire que X est égale au nombre de succès obtenus dans une série de n épreuves de Bernouilli de probabilité p. La variable aléatoire X peut donc prendre $n+1$ valeurs : $0, 1, ..., n$.
La loi binomiale $\mathcal{B}(n, p)$ est la somme de n variables de Bernouilli indépendantes.
On a :
$\mathtt{Proba(X = k)} = \mathtt{C_n^k p^k (1-p)^{n-k}}$, pour $\mathtt{0 \leq k \leq n}$,
$\mathtt{E(X)} = \mathtt{np}$,
$\sigma(\mathtt{X}) = \sqrt{\mathtt{np(1-p)}}$. **Exercice**
On fabrique des pièces et on suppose que la probabilité pour qu'une pièce soit défectueuse est $p = 0.05$ et donc il y a un contrôle de ces pièces.
Soit X la variable alèatoire égale au nombre de défectueuses trouvées lors d'un contrôle de $n = 1000$ pièces Déterminer la loi de X ainsi que son espérance et son écart-type.
Ici X suit une loi binomiale $\mathcal{B}(n, p)$ de probabilité p.
$E(X) = np = 50$
$\sigma(X) = \sqrt{np(1-p)} = 6.89202437604$

La loi des fréquences

Si la variable aléatoire Y suit la loi, dite loi des fréquences, cela veut dire que Y est égale à la fréquence des succès obtenus dans une série de n épreuves de Bernouilli de probabilité p. La variable aléatoire Y peut donc prendre $n+1$ valeurs : $0, 1/n, 2/n..., n/n$ avec les probabilités : $p_0 = (1-p)^n, p-1 = comb(n, 1)p(1-$

$p)^{n-1}, p-2 = comb(n,2)p^2(1-p)^{n-2}, ...p^n$.
On a :
$\texttt{Proba}(\texttt{Y} = \texttt{k/n}) = \texttt{C}_\texttt{n}^\texttt{k}\texttt{p}^\texttt{k}(\texttt{1} - \texttt{p})^{\texttt{n}-\texttt{k}}$, pour $\texttt{0} \leq \texttt{k} \leq \texttt{n}$,
$\texttt{E(Y)} = \texttt{p}$,
$\sigma(\texttt{Y}) = \sqrt{\texttt{p(1} - \texttt{p)/n}}$.

La loi géométrique

On dit que la variable aléatoire X suit une loi géométrique de probabilité p, si X est égale au nombre de tirages à effectuer pour avoir un succès dans une série d'épreuves de Bernouilli de probabilité p. La variable aléatoire X peut donc prendre toutes les valeurs entières : $1, .., n, ...$
On a donc :
Proba$(X = 1) = p$
Proba$(X = 2) = (1-p)p$
.....
Proba$(X = n) = (1-p)^{n-1}p$
.....
On vérifie que l'on a bien :$\sum_{j=1}^{+\infty}(1-p)^{k-1}p = 1$
Donc :
$F(1) = p$
$F(n) = \sum_{k=1}^{n}(1-p)^{k-1}p = 1 - (1-p)^n$
Espérance

$$E(X) = \sum_{n=1}^{+\infty} n(1-p)^{n-1}p = \frac{1}{p}$$

Variance et Ecart type

$$V(X) = \sum_{n=1}^{+\infty}(n - 1/p)^2(1-p)^{n-1}p = \frac{1-p}{p^2}$$

$$\sigma(X) = \frac{\sqrt{1-p}}{p}$$

Exercice
On fabrique des pièces et on suppose que la probabilité pour qu'une pièce soit défectueuse est $p = 0.05$ et donc il y a un contrôle de ces pièces.
Soit X la variable alèatoire égale à la valeur du nombre de contrôles effectués pour trouver une pièce défectueuse. Déterminer la loi de X ainsi que son espérance et son écart-type.
Ici X suit une loi géométrique de probabilité $p = 0.05$.
$E(X) = \frac{1}{p} = 20$
$\sigma(X) = \frac{\sqrt{1-p}}{p} = 19.4935886896$

La loi negbinomiale

La loi binomiale négative est une distribution de probabilité discrète. Elle dépend de 2 paramètres : un entier n (le nombre de succès attendus) et un réel p de $]0, 1[$ (la probabilité d'un succés).

On la note $\mathcal{N}eg\mathcal{B}in(n,p)$.
Elle permet de décrire la situation suivante : on fait une suite de tirages indépendants (avec pour chaque tirage, la probabilité p d'avoir un succès) jusqu'à obtenir n succès. La variable aléatoire représentant le nombre d'échecs qu'il a fallut avant d'avoir n succès, suit alors une loi binomiale négative.
Si on définit `comb(n,k)` pour $n<0$ par `comb(n,k)=n*(n-1)*..*(n-k-1)/k!`,
alors Si $X \in \mathcal{N}eg\mathcal{B}in(n,p)$ ($n \in \mathbb{N}$ et $p \in]0;1[$) alors $\texttt{Proba(X=k)} = \texttt{p}^{\texttt{n}} * (\texttt{p}-\texttt{1})^{\texttt{k}} * \texttt{comb}(-\texttt{n},\texttt{k})$
ce qui justifie le nom de loi binomiale négative et qui facilite le calcul de l'espérance (égale à $n(1-p)/p$) et de la variance (égale à $n(1-p)/p^2$).
On tape :

```
negbinomial(10,12,0.4)
```

On obtient :

```
0.0670901607617
```

La loi de Poisson

La variable aléatoire X suit une loi de Poisson $\mathcal{P}(\lambda)$ de paramètre λ ($\lambda \geq 0$) si :
- X a pour valeurs les entiers naturels,
- Prob$(X=k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$, pour $k \in \mathbb{N}$.

On a :
$\texttt{E(X)} = \lambda$
$\sigma(\texttt{X}) = \sqrt{\lambda}$

Exercice 1
Soit une variable aléatoire X qui vérifie pour $\lambda \geq 0$ et pour tout entier $n \geq 1$:
Prob$(X=n) = \dfrac{\lambda}{n}$Prob$(X=n-1)$
Montrer que X suit une loi de Poisson.

On cherche pour k entier strictement positif :
Prob$(X=k) = \dfrac{\lambda}{k}$Prob$(X=k-1) = ...\dfrac{\lambda^k}{(k)!}$Prob$(X=0)$.
Donc :

$$\text{Prob}(X=k) = \frac{\lambda^k}{(k)!}\text{Prob}(X=0)$$

On tape :
`sum(lambda^k/k!,k=0..inf)`
On obtient :
`exp(lambda)`
On doit avoir :
$\sum_{k=0}^{+\infty}$ Prob$(X=k) = 1$
Donc on a le relation :
Prob$(X=0) * \sum_{k=0}^{+\infty} \lambda^k/k! = \exp(\lambda) *$ Prob$(X=0) = 1$
c'est à dire :
Prob$(X=0) = \exp(-\lambda)$

Donc on a bien :
Prob$(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, pour $k \in \mathbb{N}$

Exercice 2
Soient X et Y deux variables aléatoires indépendantes qui suivent une loi de Poisson :
X suit une loi de Poisson $\mathcal{P}(\lambda_1)$ de paramètre λ_1 $(\lambda_1 \geq 0)$ et
Y suit une loi de Poisson $\mathcal{P}(\lambda_2)$ de paramètre λ_2 $(\lambda_2 \geq 0)$.
Déterminer la loi de la variable aléatoire $Z = X + Y$.

On a :
Prob$(X = k) = e^{-\lambda_1}\frac{{\lambda_1}^k}{k!}$, pour $k \in \mathbb{N}$
Prob$(X = k) = e^{-\lambda_2}\frac{{\lambda_2}^k}{k!}$, pour $k \in \mathbb{N}$
Donc :
Prob$(Z = n) =$ Prob$(X + Y = n) = \sum_{k=0}^{n}$ Prob$(X = k) *$ Prob$(Y = n - k) =$
$\sum_{k=0}^{n} e^{-\lambda_1}\frac{{\lambda_1}^k}{k!} * e^{-\lambda_2}\frac{{\lambda_2}^{n-k}}{(n-k)!}$
On sait que : $(\lambda_1 + \lambda_2)^n = \sum_{k=0}^{n} {\lambda_1}^k {\lambda_2}^{n-k}\frac{n!}{k!(n-k)!}$
Donc :
Prob$(Z = n) = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}$
Donc Z suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$.

2.3 Variable aléatoire absolument continue

Définitions

Variable aléatoire absolument continue
Une variable aléatoire X est absolument continue si il existe $f(x)$ telle que sa fonction de répartition $F(x)$ est égale à :

$$\text{Prob}(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)dt$$

Densité de probabilité
La fonction $f(x)$ est appelée densite de probabilité et on a :

$$f(x) = F'(x)$$

Espérance mathématique
L'espérance mathématique ou moyenne de x est :

$$E(X) = \overline{X} = \int_{-\infty}^{+\infty} t * f(t)dt$$

Variance et Ecart type
La variance de X est :

$$V(X) = \int_{-\infty}^{+\infty} (t - E(X))^2 f(t)dt$$

L'ecart type de X est :

$$\sigma(X) = \sqrt{V(X)}$$

La loi uniforme

Définition

On dit que la variable aléatoire X suit une loi uniforme sur un segment $[a, b]$ si sa densité de probabilité $f(x)$ est une constante C sur $[a, b]$ et est nulle en dehors du segment $[a, b]$.

On a donc :

$C = \dfrac{1}{b-a}$ puisque $\displaystyle\int_a^b C dt = 1$

$F(x) = 0$ pour $x < a$

$F(x) = \dfrac{x-a}{b-a}$ pour $a \leq x < b$

$F(x) = 1$ pour $x \geq b$

Espérance

$$E(X) = \frac{1}{b-a}\int_a^b t dt = \frac{a+b}{2}$$

Variance et Ecart type

$$V(X) = \frac{1}{2b-2a}\int_a^b (2t-a-b)^2 dt = \frac{(b-a)^2}{12}$$

$$\sigma(X) = \frac{(b-a)}{2\sqrt{3}}$$

Exercices

1/ Soient n variables aléatoires indépendantes U_k qui suivent une loi uniforme sur $[0, 1]$.

On considère les variables $X = \text{Max}(U_k)$ et $Y = \text{Min}(U_k)$.

Déterminer les fonctions de répartition de X et Y

Calculer $E(X)$, $E(Y)$, $V(X)$, $V(Y)$.

On a :

Proba$(X \leq x)$=Proba$(U_1 \leq x)$ et Proba$(U_2 \leq x)$...et Proba$(U_n \leq x)$

Donc puisque les U_k sont indépendantes :

Proba$(X \leq x) = \Pi_{k=1}^n$Proba$(U_k \leq x)$ Soit : $F_X(x) = 0$ pour $x < 0$

$F_X(x) = x^n$ pour $x \in [0, 1]$ $F_X(x) = 1$ pour $x > 1$ La densité de probabilités vaut $f_X(x) = nx^{n-1}$ sur $[0, 1]$ donc :

$E(X) = n\displaystyle\int_0^1 x^n dx = \frac{n}{n+1}$

$V(X) = n\displaystyle\int_0^1 (x - n/(n+1))^2 x^{n-1} dx = \frac{n}{n^3 + 4*n^2 + 5*n + 2}$

En effet, on tape :

```
assume(n>=1)
n*int(1(x-n/(n+1))^2*x^(n-1),x=0..1)
```

On obtient :

```
n/(n^3+4*n^2+5*n+2)
```

On a :

Proba$(Y \leq x)$=Proba$(U_1 \leq x)$ ou Proba$(U_2 \leq x)$...ou Proba$(U_n \leq x)$

On sait que :
Proba$(Y < x) = 1-$Proba$(Y > x)$ et
Proba$(U_k < x) = 0$ et Proba$(U_k > x) = 1$ si $x < 0$
Proba$(U_k < x) = x$ et Proba$(U_k > x) = 1 - x$ si $0 < x < 1$
Proba$(U_k < x) = 1$ et Proba$(U_k > x) = 0$ si $x > 1$
On calcule :
Proba$(Y > x)$=Proba$(U_1 > x)$ et Proba$(U_2 > x)$ et....Proba$(U_n > x)$
Comme les U_k sont independants :
Proba$(Y > x) = 1$ si $x < 0$
Proba$(Y > x) = (1 - x)^n$ si $0 < x < 1$
Proba$(Y > x) = 0$ si $1 < x$
Donc :
Proba$(Y < x) = 1 - 1 = 0$ si $x < 0$
Proba$(Y < x) = 1 - (1 - x)^n$ si $0 < x < 1$
Proba$(Y < x) = 1 - 0 = 1$ si $x > 1$
La densité de probabilité est donc :
$f_Y(x) = n(1 - x)^{n-1}$ sur [0 ;1] et 0 en dehors de [0 ;1].
donc :

$$E(Y) = n\int_0^1 x(1-x)^{n-1}dx = \frac{1}{n+1}$$

$$V(Y) = n\int_0^1 (x - 1/(n+1))^2(1-x)^{n-1}dx = \frac{n}{n^3 + 4 * n^2 + 5 * n + 2}$$

En effet, on tape :

```
assume(n>=1)
n*int(x*(1-x)^(n-1),x=0..1)
```

On obtient :

```
1/(n+1)
```

On tape :

```
assume(n>=1)
n*int((x-1/(n+1))^2*(1-x)^(n-1),x=0..1)
```

On obtient :

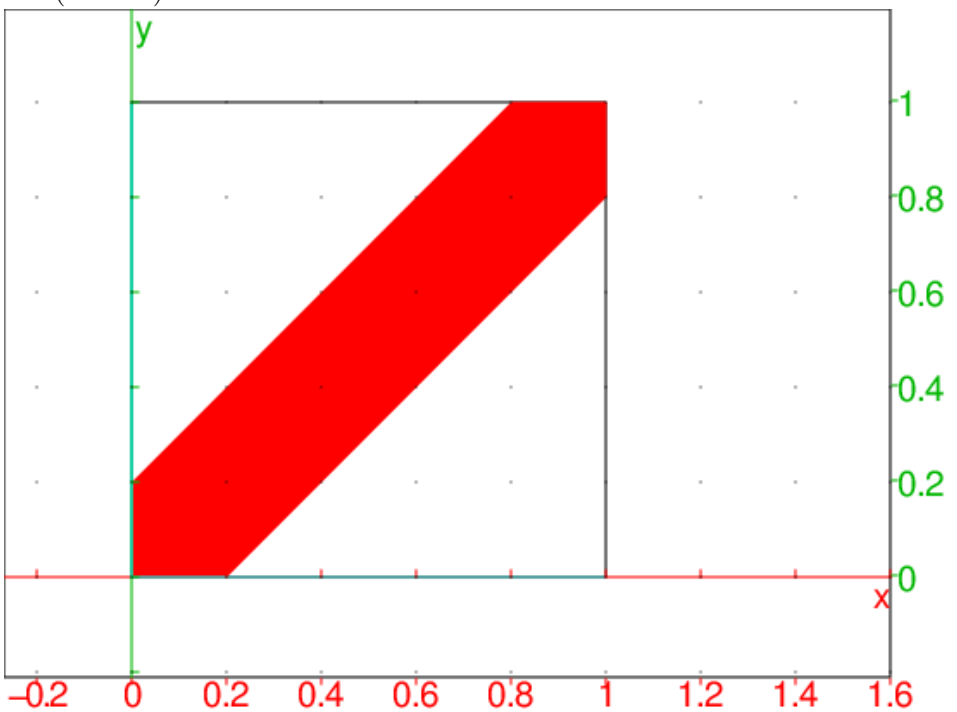
```
n/(n^3+4*n^2+5*n+2)
```

2/ Deux personnes A et B se donnent rendez-vous entre 12h et 13h. Les instants d'arrivée de A et B sont respectivement 2 variables aléatoires X et Y continues, indépendantes et uniforme sur [0,1].
Soit Z la variable aléatoire associé au temps d'attente de la première personne arrivée.
Calculer la fonction de répartition de Z.
En déduire la densité de probabilité de Z.
Calculer l'espérance de Z.
Calculer la variance et l'ecart-type de Z.
A et B conviennent qu'elles n'attendront pas plus d'une demie-heure. Quelle est la probabilité pour que le rendez-vous ait lieu ?
Si les personnes veulent que le rendez-vous ait lieu avec une probabilité de 0.96 quel est le temps d'attente maximum que A et B doivent se fixer ?
Solution
Le temps d'attente Z est égal à $|X - Y|$.
En effet, supposons que $X = x$ et $Y = y$:

si $y < x$, cela signifie que la personne B va attendre A pendant $(x - y)$ h,
si $x < y$, cela signifie que la personne A va attendre B pendant $(y - x)$ h,
Le temps d'attente est donc égal à $|x - y|$ (en heures).
Si $Z = z$, le temps d'attente $z = |x - y|$, donc Z est égal à $|X - Y|$.

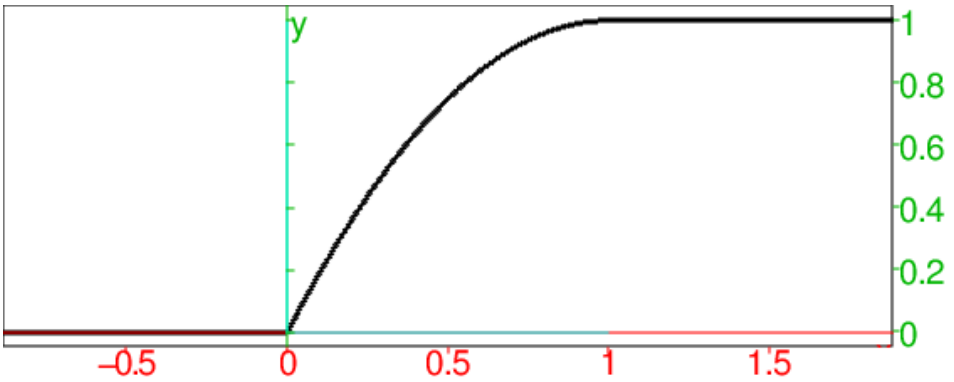
Fonction de répartition de Z

Si $t < 0$ $F(t) = Proba(Z < t) = Proba(|X-Y| < t) = 0$ puisque $|X-Y| \geq 0$
Si $t > 1$ $F(t) = Proba(Z < t) = Proba(|X-Y| < t) = 1$ puisque $|X-Y| \leq 1$
Si $t \in [0, 1]$ $F(t) = Proba(Z < t) = Proba(|X - Y| < t) = 0$
Cette probabilité est représentée pour $t = 0.2$ par l'aire rouge ci-dessous qui vaut $1 - (1 - 0.2)^2 = 0.36$:



En effet lorsque $X = x$ et $Y = y$, si $t = 0.2$, les points de coordonnées (x, y) qui vérifient $|x - y| < 0.2$ sont les points du carré $[0, 1] \times [0, 1]$ tels que $x - 0.2 < y < x + 0.2$.
Cette aire en fonction de t est : $1 - (1 - t)^2 = 2t - t^2$.
Donc $F(t) = 2t - t^2$ lorsque $t \in [0, 1]$
On tape :

```
plotfunc(2t-t^2,t=0..1,affichage=epaisseur_ligne_3);
segment(1+i,2+i,affichage=epaisseur_ligne_3);
segment(-1,0,affichage=epaisseur_ligne_3)
```



Densité de probabilité de Z
On a $F(t) = 2t - t^2$ si $t \in [0,1]$ et sinon $F(t) = 0$
Donc, si $t \in [0,1]$ alors $f(t) = F'(t) = 2 - 2t$ et sinon $f(t) = 0$
Espérance de Z
$E(Z) = \int_0^1 tf(t)dt = \int_0^1 (2t - 2t^2)dt - 1 - 2/3 = 1/3$
Donc $E(Z) = 1/3$ $h = 20$ mn
Variance et l'ecart-type de Z
$V(Z) = \int_0^1 (t - E(Z))^2 f(t)dt = \int_0^1 (t - 1/3)^2(2 - 2t)dt$
ou bien :
$V(Z) = \int_0^1 t^2 f(t)dt - E(Z)^2 = \int_0^1 2t^2 - 2t^3 dt - 1/9 = 2/3 - 1/2 - 1/9 = 1/18$
On vérifie, on tape :
`int((t-1/3)^2*(2-2t),t=0..1)`
On obtient :
`1/18`
Donc la variance de Z vaut 1/18 et
l'écart-type vaut $\sqrt{2}/6 \simeq 0.235702260396$
Probabilité pour que le rendez-vous ait lieu
Si A et B conviennent qu'elles n'attendront pas plus que 30 mn, la probabilité pour que le rendez-vous ait lieu est :
$F(1/2) = 2 * 1/2 - 1/4 = 3/4 = 0.75$
Temps d'attente si on veut $F(t) = 0.96$
Si on veut que $F(t) = 0.96$, il faut résoudre :
$2t - t^2 = 0.96$ avec $0 < t < 1$
$t^2 - 2t + 0.96 = 0$ a pour solutions :
$t_1 = 1 - \sqrt{0.04} = 0.8$ et $t_2 = 1 + \sqrt{0.04} = 1.2 > 1$
Puisque $t_2 > 1$, on élimine t_2 et on en déduit que $F(0.8) = 0.96$.
On a :0.8 $h = 48$ mn.
Si A et B conviennent qu'elles n'attendront pas plus que 48 mn, la probabilité pour que le rendez-vous ait lieu est 0.96.

La loi exponentielle

Définition

On dit que la variable aléatoire X suit une loi exponentielle si sa densité de probabilité vaut pour $a > 0$:
$f(x) = a\exp(-ax)$ pour $x \geq 0$ et
$f(x) = 0$ pour $x < 0$ On a donc :
$F(x)$ =Proba$(X \leq x) = a\int_0^x \exp(-at)dt = 1 - \exp(-ax)$

Espérance

$$E(X) = a\int_0^{+\infty} t\exp(-at)dt = \frac{1}{a}$$

Variance et Ecart type

$$V(X) = a\int_0^{+\infty} (t-1/a)^2\exp(-at)dt = \frac{1}{a^2}$$

$$\sigma(X) = \frac{1}{a}$$

Exercices

1/ Un pont levant permet le passage des voitures en position basse et des bateaux en position haute.
Son temps de montée est de 2 minutes (mn), son temps de descente est de 2 mn et il reste en position haute 8 mn.
A/ a/ Un atomobiliste arrive lorsque le pont est en position haute.
Combien attend-il au minimum ? au maximum ?
b/ On suppose que ce temps d'attente en minutes est une variable aléatoire D Qui suit une loi uniforme sur $[2, 10]$. Calculer $E(D)$ et interpréter le résultat.
c/ Calculer la probabilité que le temps d'attente de l'automobiliste ne dépasse pas 5 mn. B/ L'intervalle de temps en heure (h) entre 2 passages de bateaux est une variable aléatoire T qui suit une loi exponentielle de paramètre $\lambda = 0.05$: T est le temps de latence.
a/ Soit $f(x) = \exp(-x/20)/20$ pour $x \in [0, +\infty[$.
Calculer une primitive de $f(x)$. En déduire $P(T \leq t)$
b/ Montrer que $F(x) = (-20 - x) * \exp(-x/20)$ est une primitive de $x * f(x)$.
En déduire $E(T)$ et interpréter le résultat.
c/ Calculer :
$P(T \leq 12)$, $P(T > 24)$ et $P(12 \leq T \leq 24)$

Solution

A/ a/ L'automobiliste qui arrive lorsque le pont est en position haute attend au minimum 2mn et au maximum 8+2=10 mn.
b/ La densité de probabilité de D est une constante c telle que :
$\int_2^{10} cdx = 1$ donc $c = 1/8$.
On a : $E(D) = \int_2^{10} x/8dx = 100^2/16 - 4/16 = 6 = (2+10)/2$ mn.
L'automobiliste qui arrive lorsque le pont est en position haute attend en moyenne 6 mn.
c/ $P(D \leq 5) = \int_2^5 1/8dx = (5-2)/8 = 3/8 = 0.375$
B/ a/ La densité de probabilité de T est égale à :
$f(x) = \exp(-x/20)/20$ puisque $0.05 = 1/20$.
une primitive de $f(x)$ est : $-\exp(-x/20)$ donc :
$P(T \leq t) = \int_0^t f(x)dx = 1 - \exp(-t/20)$.
b/ Calculons $F'(x) = -\exp(-x/20) - (-20 - x) * \exp(-x/20)/20$ ou encore :
$F'(x) = -\exp(-x/20)/20 = xf(x)$
On a $F(0) = (-20 - 0) * \exp(-/20) = 20$ donc
$E(T) = \int_0^{+\infty} xf(x)dx = 0 - F(0) = 20$ Le temps de latence est en moyenne de 20 h.
c/ Calcul de :

$P(T \leq 12) = 1 - \exp(-12/20) = 1 - \exp(-0.6) \simeq 0.451188363906$,
$P(T > 24) = 1 - P(T \leq 24) = \exp(-1.2) \simeq 0.301194211912$ et
$P(12 \leq T \leq 24) = 1 - P(T \leq 12) - P(T > 24) = \exp(-0.6) - \exp(-1.2) = 0.247617424182$

La loi Normale ou loi de Gauss

La variable aléatoire X suit une loi Normale ou loi de Gauss de paramètres $\mu, \sigma(\sigma > 0)$ si :
- X a pour valeurs tous les réels,
- $Prob(a \leq X < b) = \int_a^b f(t)dt$ où $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ (f est la densité de probabilité et a comme représentation graphique une courbe en cloche).
On note cette loi $\mathcal{N}(\mu, \sigma)$.
On a :
$\mathtt{E(X)} = \mu$
$\sigma(\mathtt{X}) = \sigma$.
On dit que $\mathcal{N}(0, 1)$ est la loi normale centrée réduite. Si X suit la loi $\mathcal{N}(0, 1)$ alors :
$Prob(a \leq X < b) = \int_a^b = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$
Si X suit la loi $\mathcal{N}(\mu, \sigma)$ alors $\dfrac{X - \mu}{\sigma}$ suit la loi $\mathcal{N}(0, 1)$.
On a des tables où on peut lire que :
$P(\dfrac{|X - \mu|}{\sigma} > 1.96) = 0.05$,
$P(\dfrac{|X - \mu|}{\sigma} > 2.58) = 0.01$,
$P(\dfrac{|X - \mu|}{\sigma} > 3.1) = 0.001$,
et on a $P(\dfrac{|X - \mu|}{\sigma} > t) = 1 - 2\displaystyle\int_0^t f(x)dx$.

Théorèmes

On montre qu'une loi binomiale $\mathcal{B}(n, p)$ peut être approchée :

— par la loi normale $\mathcal{N}(np, \sqrt{np(1-p)})$ si $np > 15$ et $n(1-p) > 15$.
Cela veut dire que pour tout entier k :
$C_n^k p^k (1-p)^{n-k}$ est proche de $\frac{1}{\sqrt{2np(1-p)\pi}} e^{-\frac{(k-np)^2}{2np(1-p)}}$ quand n est grand.
Exemple :
On a $p(k) = C_{100}^k 0.2^k 0.8^{100-k}$ est proche de $f(k) = \frac{1}{4\sqrt{2\pi}} e^{-\frac{(k-20)^2}{32}}$.
Ainsi si X suit la loi $\mathcal{B}(n, p)$ et Y suit la loi $\mathcal{N}(np, \sqrt{np(1-p)})$ alors :
$Proba(X = x)$ sera approché par $Proba(x - 0.5 < Y < x + 0.5)$
$Proba(X < x)$ sera approché par $Proba(Y < x - 0.5)$
$Proba(X \leq x)$ sera approché par $Proba(Y < x + 0.5)$

— par la loi de Poisson $\mathcal{P}(np)$ si $np \leq 10$, $p \leq 0.1$ et $n \geq 15$.
Exemple :
On a $p(k) = C_{50}^k 0.04^k 0.96^{50-k}$ est proche de $e^{-2}\frac{2^k}{k!}$.

2.3.1 Probabilités et fréquences

La distribution des fréquences issues de la répétition d'expériences identiques et indépendantes varient alors qu'une loi de probabilité associée à la réalisation d'une expérience est un invariant.
La fréquence d'un élément est calculée à partir de données expérimentales alors que sa probabilité est calculée mathématiquement selon le modéle choisi. Les calculs de la statistique consistent à nous aider à choisir le bon modèle.
Exemple :
On lance 100 fois, une pièce bien équilibrée, et on obtient 48 faces : la probabilité de tomber sur "face" est 0.5 alors que la fréquence d'apparition des "faces" est ici 0.48.
En statistique on étudie la valeur d'un caractère et en en langage probabiliste on étudie la valeur d'une variable aléatoire.
En statistique on parle de fréquences et en langage probabiliste on parle de probabilité.
En statistique on parle de moyennes et en en langage probabiliste on parle d'espérance.
Dans le monde théorique défini par une loi de probabilité P sur un ensemble Ω, les fréquences des éléments de Ω dans une suite de n expériences identiques et indépendantes tendent vers leur probabilité quand n augmente indéfiniment.

2.4 Probabilités conditionnelles

Définition
Soit, deux événements A et B dans un espace de probabilisé.
On dit que les événements A et B sont indépendants si et seulement si :
$\text{Prob}(A \cap B) = \text{Prob}(A) * \text{Prob}(B)$ **Définition**
La probabilité de B sachant A notée $\text{Prob}_A(B)$ est la probabilité de B quand A est réalisé.
Théorème
On a :

$$\text{Prob}_A(B) = \frac{\text{Prob}(A \cap B)}{\text{Prob}(A)}$$

Exercice Dans une fabrique de pièces il y a 5% de pièces défectueuses.
Lors des contrôles :
- les pièces bonnes sont refusées avec une probabilité de 4%.
- les pièces défectueuses sont refusées avec une probabilité de 98%.
Calculer pour une pièce :

— La probabilité de commettre une erreur.
— La probabilité d'être bonne sachant qu'elle a été refusée
— La probabilité d'être défectueuse sachant qu'elle a été acceptée

On a les événements :
A la pièce est acceptée,
R la pièce est refusée,
B la pièce est bonne,
D la pièce est défectueuse.
C la pièce est jugée conforme.

On a les probabilités :
$\text{Prob}_B(R) = 0.04$
$\text{Prob}_D(R) = 0.98$
$\text{Prob}_D(A) = 0.02$
$\text{Prob}(B) = 0.95$
$\text{Prob}(D) = 0.05$

— Une pièce est jugée non conforme si elle est bonne et refusée ou si elle est défectueuse et acceptée donc :
$\text{Prob}(R) = \text{Prob}(R \cap B) + \text{Prob}(A \cap D)$
On a :
$\text{Prob}(R \cap B) = \text{Prob}_B(R) * \text{Prob}(B) = 0.04 * 0.95 = 0.038$
$\text{Prob}(A \cap D) = \text{Prob}_D(A) * \text{Prob}(D) = 0.02 * 0.05 = 0.001$
Donc la probabilité d'être refusée est : $\text{Prob}(R) = 0.038 + 0.049 = 0.913$
Donc la probabilité d'être acceptée est :
$\text{Prob}(A) = 1 - 0.913 = 0.087$

— La probabilité d'être bonne sachant qu'elle a été refusée
On a :
$\text{Prob}_R(B) = \frac{\text{Prob}(R \cap B)}{\text{Prob}(R)} = 0.038 * 0.087 = 0.003306$
— La probabilité d'être défectueuse sachant qu'elle a été acceptée.
La probabilité d'être défectueuse et acceptée est :
$\text{Prob}(A \cap D) = 1 - \text{Prob}(R \cap D) = 1 - 0.049 = 0.951$ $\text{Prob}_A(D) = \frac{\text{Prob}(A \cap D)}{\text{Prob}(A)} = 0.001 * 0.087 = 8.7e - 5$

2.5 Variables aléatoires

Une variable aléatoire permet de faire correspondre à un espace probabilisable Ω, un espace probabilisable d'univers $\mathbb{R}$.

Définitions

Soit (Ω, τ) un espace probabilisable.
On appelle **variable aléatoire** sur cet espace toute application X de Ω vers $\mathbb{R}$ telle que toute réunion dénombrables d'intervalles B de $\mathbb{R}$, $X^{-1} \in B$.

On appelle **fonction de répartition** d'une variable aléatoire X la fonction F de $\mathbb{R}$ dans [0,1] par :
$F(x) = \text{Prob}(X \geq x)$.

Une variable aléatoire peut être :
discréte et finie (si $X(\Omega) = \{x_1, x_2....x_n\}$) ou,
discréte et infinie (si $X(\Omega) = \{x_1, x_2....x_j...\}$) ou,
absolument continue (si sa fonction de répartition $F(x) = \int_{-\infty}^{x} f(t)dt$).
La fonction f est alors appelée **densité de probabilité** de la variable aléatoire X.

On appelle **espérance mathématique** (ou moyenne) de la variable aléatoire X, le nombre :

— X est **discréte et finie** :
$E(X) = \sum_{j=1}^{n} x_j \text{Prob}(X = x_j)$
— X est **discréte et infinie** :
$E(X) = \sum_{j=1}^{+\infty} x_j \text{Prob}(X = x_j)$

— X est **absolument continue** :
$E(X) = \int_{-\infty}^{+\infty} xf(x)dx$

On appelle **variance** de la variable aléatoire X, le nombre :

— X est **discréte et finie** :
$V(X) = \sum_{j=1}^{n}(x_j - E(X))^2 \text{Prob}(X = x_j)$

— X est **discréte et infinie** :
$V(X) = \sum_{j=1}^{+\infty}(x_j - E(X))^2 \text{Prob}(X = x_j)$

— X est **absolument continue** :
$V(X) = \int_{-\infty}^{+\infty}(x - E(X))^2 f(x)dx$

On a :
$V(X) = E(X^2) - E(X)^2$.

On appelle **écart type** de la variable aléatoire X, le nombre :
$\sigma(X) = \sqrt{V(X)}$

2.6 Le processus de Poisson

2.6.1 Définitions

Le processus de Poisson gére les événements qui se produisent aléatoirement. On dit que des événements aléatoires suivent un processus de Poisson de paramètre λ si le nombre d'événements produit dans l'intervalle t suit une loi de Poisson de paramètre λt.
Si les intervalles de temps sont disjoints le nombre d'événements produit dans ces intervalles sont indépendants.
On a donc si $N(t)$ est le nombre d'événements produit dans l'intervalle t : Proba$(N(t) = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$ et donc :
Proba$(N(t) = 0) = e^{-\lambda t}$ On considère les variables aléatoires $T_1,\ T_2,\ T_n$ qui représentent l'intervalle d'attente entre la production de 2 évenements. $T_1,\ T_2,\ T_n$ sont indépendantes et suivent une loi exponnentielle de paramètre λ, en effet :
Proba$(T_1(t) \leq t) = 1-$Proba$(T_1(t) \geq t) = 1-$Proba$(N(t) = 0) = 1 - e^{-\lambda t}$.
Soit S_n la variable aléatoire qui représente le temps d'attente de la production du n-ième évenement. On a donc :
$S_n(t) = T_1(t) + T_2(t) + ... + T_n(t)$ et

Proba$(S_n(t) \leq t) =$Proba$(N(t) \geq n) = 1 - e^{-\lambda t}\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!}$.

La densité de probabilité $f_{S_n}(t)$ de S_n est donc égale à :

$$f_{S_n}(t) = -e^{-\lambda t}(-\lambda\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!} + \lambda\sum_{k=1}^{n-1}\frac{(\lambda t)^{k-1}}{(k-1)!})$$

Donc :

$$f_{S_n}(t) = \lambda e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}$$

S_n suit donc une loi gamma ou loi de Erlang de paramètres n et λ.
Puisque $S_n(t) = T_1(t) + T_2(t) + ... + T_n(t)$ et que les T_j sont indépendantes et suivent la même loi, on a :

$$E(S_n) = n * E(T_1) = \frac{n}{\lambda} \text{ et}$$

$$V(S_n) = n * V(T_1) = \frac{n}{\lambda^2}$$

2.6.2 Exercices

— Des événements aléatoires suivent un processus de Poisson de paramètre 3 par heure. Quelle est la probabilité qu'aucun événement survienne entre 12h et 14h ?
Calculer la moyenne du temps d'attente pour la venue du 5-ième événement.

On a :
Proba$(N(2) = 0) = e^{-3*2} = e^{-6}$
la probabilité qu'aucun événement survienne entre 12h et 14h est donc de $e^{-6} \simeq 0.00247875217667$ $E(S_5) = 5/3$h=1h40mn
La venue du 5-ième événement aura lieu en moyenne au bout de 1h40mn.

— Le nombre de clients qui entrent dans un magasin suit un processus de Poisson de paramètre λ par heure.
Sachant que 2 clients sont venus entre 14h et 15h :
Quelle est la probabilité qu'ils soient venus tous les deux entre 14h et 14h20 ?
Quelle est la probabilité pour qu'au moins 1 client soit venus entre 14h et 14h30 ?

On a :
Proba$(N(1) = 2) = e^{-\lambda}\lambda^2/2$
Proba$(N(1/3) = 2)|(N(1) = 2) =$
Proba$((N(1/3) = 2) \cap (N(1) = 2))/$Proba$(N(1) = 2)$.
Proba$((N(1/3) = 2) \cap (N(1) = 2)) =$
Proba$((N(1/3) = 2) \cap (N(2/3) = 0)) =$
Proba$(N(1/3) = 2*$Proba$(N(2/3) = 0)) = e^{-\lambda/3} * (\lambda/3)^2/2 * e^{-2\lambda/3} = e^{-\lambda}(\lambda/3)^2$
Donc Proba$(N(1/3) = 2|(N(1) = 2)) = 1/9 \simeq 0.111111111111$ est la probabilité que 2 clients soient venus tous les deux entre 14h et 14h20 sachant que 2 clients sont venus entre 14h et 15h
On cherche :
Proba$(N(1/2) \geq 1)|(N(1) = 2)) =$
1-Proba$((N(1/2) = 0)|(N(1) = 2)) =$
1-Proba$((N(1/2) = 0) \cap (N(1) = 2))/$Proba$(N(1) = 2) =$
1-Proba$(N(1/2) = 0))*$Proba$(N(1/2) = 2))/$Proba$(N(1) = 2) =$
$1 - e^{-\lambda/2} * e^{-\lambda/2} * (\lambda/2)^2/2/(e^{-\lambda}\lambda^2/2) = 1 - 1/4 = 3/4$
La probabilité pour qu'au moins 1 client soit venu entre 14h et 14h30 sachant que 2 clients sont venus entre 14h et 15h est donc de 0.75.

2.7 Couple de variables aléatoires discrètes

2.7.1 Définitions

On appelle **fonction de répartition** du couple de variables aléatoires discrètes (X, Y) sur le même espace probabilisé, la fonction F définie par :
$F(x, y) =$ Prob$((X \geq x) \cap (Y \geq y))$.

On appelle **loi conjointe** du couple de variables aléatoires discrètes (X, Y), l'application p définie par :
$p(x_i, y_j) = \text{Prob}((X = x_i) \cap (Y = y_j))$.
On a donc si $x_{p_x} \leq x < x_{p_x+1}$ et $y_{q_y} \leq y < y_{q_y+1}$:
$F(x, y) = \sum_{i=1}^{p_x} \sum_{i=1}^{q_y} p(x_i, y_j)$.

On appelle **loi de probabilités marginales** du couple de variables aléatoires discrètes (X, Y), les lois de probabilité de X et de Y i.e. l'application p_X définie par :
$p(x_i) = \text{Prob}(X = x_i)$ et
l'application p_Y définie par :
$p(y_j) = \text{Prob}(Y = y_j)$.

On appelle **répartitions marginales** du couple de variables aléatoires discrètes (X, Y), les fonctions de répartition F_X pour X et F_Y pour Y définies par :
$F_X(x) = \text{Prob}(X \leq x)$ et
$F_Y(y) = \text{Prob}(Y \leq y)$.

On appelle **loi de probabilité conditionnelle** de Y sachant que $X = x_i$ l'application qui a y_j fait correspondre :

$$\frac{\text{Prob}((X = x_i) \cap (Y = y_j))}{\text{Prob}(X = x_i)}$$

2.7.2 Exercices

1. Soient deux variables aléatoires X et Y valant 0 ou 1 et telles que :
   $\text{Prob}((X = 0) \cap (Y = 0)) = 0.4$
   $\text{Prob}((X = 0) \cap (Y = 1)) = 0.2$
   $\text{Prob}((X = 1) \cap (Y = 0)) = 0.1$
   $\text{Prob}((X = 1) \cap (Y = 1)) = 0.3$
   Calculer la probabilité de X sachant que $Y = 1$.
2. Soient deux variables aléatoires X et Y indépendantes qui suivent une loi de Poisson de paramètres λ_1 et λ_2. Calculer la probabilité de X sachant que $X + Y = n$ avec $n \in \mathbb{N}$.

1. On a :
   $\text{Prob}(Y = 1) = \text{Prob}((X = 0) \cap (Y = 1)) + \text{Prob}((X = 0) \cap (Y = 1)) = 0.2 + 0.3 = 0.5$ et
   $$\text{Prob}((X = 0)/(Y = 1)) = \frac{\text{Prob}((X = 0) \cap (Y = 1))}{\text{Prob}(Y = 1)}$$
   $$\text{Prob}((X = 1)/(Y = 1)) = \frac{\text{Prob}((X = 1) \cap (Y = 1))}{\text{Prob}(Y = 1)}$$
   Donc :
   $$\text{Prob}((X = 0)/(Y = 1)) = \frac{0.2}{0.5} = 0.4 \text{ et}$$
   $$\text{Prob}((X = 1)/(Y = 1)) = \frac{0.3}{0.5} = 0.6$$
2. On a :
   $$\text{Prob}(X = k) = \frac{\lambda_1^k e^{-\lambda_1}}{k!} \text{ et}$$
   $$\text{Prob}((X = k)/(X + Y = n)) = \frac{\text{Prob}((X = k) \cap (Y = n - k))}{\text{Prob}(X + Y = n)}$$
   On a montré (cf 2.2) que $Z = X + Y$ suit une loi de poisson de paramètre

$\lambda_1 + \lambda_2$ donc
Prob$(X + Y = n) = \dfrac{(\lambda_1 + \lambda_2)^n e^{-(\lambda_1+\lambda_2)}}{n!}$
Les variables aléatoires X et Y sont indépendantes donc pour $0 \leq k \leq n$ on a :
Prob$((X = k) \cap (Y = n - k))$ = Prob$(X = k) *$ Prob$(Y = n - k)$
Donc :
Prob$((X = k)/(X + Y = n)) = \dfrac{{\lambda_1}^k e^{-\lambda_1} {\lambda_2}^{n-k} e^{-\lambda_2} n!}{k!(n-k)!(\lambda_1 + \lambda_2)^n e^{-(\lambda_1+\lambda_2)}}$
Prob$((X = k)/(X + Y = n)) = \dfrac{{\lambda_1}^k {\lambda_2}^{n-k} n!}{(\lambda_1 + \lambda_2)^n k!(n-k)!} =$
$\dfrac{n!}{k!(n-k)!}(\dfrac{\lambda_1}{\lambda_1 + \lambda_2})^k (\dfrac{\lambda_2}{\lambda_1 + \lambda_2})^{n-k}$.
Donc la probabilité de X sachant que $X + Y = n$ avec $n \in \mathbb{N}$ est la loi binomiale $B(n, \dfrac{\lambda_1}{\lambda_1 + \lambda_2}, \dfrac{\lambda_2}{\lambda_1 + \lambda_2})$

2.8 Couple de variables aléatoires continues

2.8.1 Définitions

On appelle **fonction de répartition** du couple de variables aléatoires continues (X, Y) sur le même espace probabilisé, la fonction $F_{(X,Y)}$ définie par :
$F_{(X,Y)}(x, y) = \int_{-\infty}^{x} (\int_{-\infty}^{y} f_{(X,Y)}(u, v)dv)du.$
La fonction $f_{(X,Y)}$ est appelée **densité de probabilité** du couple de variables aléatoires continues (X, Y).

La **probabilité relative à un pavé** $D = \{(x, y) \in \mathbb{R}^2 a \leq x \leq b, c \leq y \leq d\}$ est :
Prob$((X = xetY = yet(x, y) \in D)) = \int_c^d (\int_a^b f(x, y)dx)dy.$

La **probabilité relative à un domaine de Borel** D (D est la reunion (ou intersection) dénombrable d'une suite de pavés) est :
Prob$((X = xetY = yet(x, y) \in D)) = \int \int_D f(x, y)dxdy.$

On appelle **répartitions marginales** du couple de variables aléatoires continues (X, Y), les fonctions de répartition F_X pour X et F_Y pour Y définie par :
$F_X(x) = \int_{-\infty}^{x} (\int_{-\infty}^{+\infty} f_{(X,Y)}(u, v)dv)du$ et
$F_Y = \int_{-\infty}^{y} (\int_{-\infty}^{+\infty} f_{(X,Y)}(u, v)du)dv.$
On appelle **densités marginales** du couple de variables aléatoires continues (X, Y), les fonctions f_X pour X et f_Y pour Y définies par :
$f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y)dy$ et
$f_Y(y) == \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y)dx.$
On a donc : $f_{(X,Y)}(x, y) = \dfrac{\partial^2 F_{(X,Y)}}{\partial x \partial y}$.

On appelle **probabilité conditionnelle** de Y sachant que $x \leq X \leq x + dx$

l'application qui a y fait correspondre :

$$\frac{\int_{-\infty}^{y} f_{(X,Y)}(x,v)dv}{\int_{-\infty}^{+\infty} f_{(X,Y)}(x,v)dv}$$

On appelle **densité conditionnelle** de Y sachant que $x \leq X \leq x+dx$ l'application $f_{Y/X}$ définie par :

$$f_{Y/X}(y/x) = \frac{f(x,y)}{f_X(x)}$$

Si X et Y sont indépendantes alors $f_{(X,Y)}(x,y) = f_X(x) * f_Y(y)$

2.8.2 Exercices

1. Soit f est la fonction **densité de probabilité** du couple de variables aléatoires continues (X,Y) définie par :
   $f(x,y) = \frac{12}{5}x(2x-y)$ si $x \in [0,1]$ et $y \in [0,1]$ et $f(x,y) = 0$ sinon.
   Calculer la probabilité conditionnelle de X sachant que $y \leq Y \leq y+dy$.

   On vérifie :
   $\int_0^1 \int_0^1 \frac{12}{5}x(2-x-y)dxdy = \int_0^1 \frac{12}{5}(1-\frac{1}{3}-\frac{y}{2})dy = \frac{12}{5}(\frac{2}{3}-\frac{1}{4}) = 1$
   Ou on tape avec `Xcas` :
   `int(int(12/5*x*(2-x-y),x=0..1),y=0..1)`
   On obtient : `1`
   On calcule $f_Y(y)$ et on tape :
   `int(12/5*x*(2-x-y),x=0..1)`
   On obtient : `-6/5*y+8/5`
   On calcule $f_{X/Y}(x/y)$ et on tape :
   `factor((12/5*x*(2-x-y))/(-6/5*y+8/5))`
   On obtient : `(6*x*(y+x-2))/(3*y-4)`
   Donc $f_{X/Y}(x/y)$ est égale à :

   $$\frac{6x(y+x-2)}{3*y-4}$$

2. Soit f est la fonction **densité de probabilité** du couple de variables aléatoires continues (X,Y) définie par :
   $f(x,y) = \frac{6}{7}(x^2 + \frac{xy}{2})$ si $x \in [0,1]$ et $y \in [0,2]$ et $f(x,y) = 0$ sinon.
   Calculer :
   (a) la densité de probabilité de X.
   (b) $P(X > Y)$.
   (c) $P(Y > 1/2 | X < 1/2)$.

   On vérifie que :
   $\int_0^1 (\int_0^2 f(x,y)dy)dx = 1$
   On tape :

```
6/7*int(int(x^2+x*y/2,y=0..2),x=0..1)
```
On obtient : `1`

(a) la densité de probabilité de X est égale à :
$f_X(x) = \int_0^2 f(x,y)dy$
On tape :
```
6/7*(int(x^2+x*y/2,y=0..2)
```
On obtient : `(6*(2*x^2+x))/7`
la fonction de répartition de X est donc égale à :
$F_X(x) = \int_0^1 f_X(u)du$
On tape :
```
6/7*(int(2*u^2+u,u=0..x)
```
On obtient : `(6*(2/3*x^3+1/2*x^2))/7`

(b) $P(X > Y)$.
On a si $T = \{((x,y) \in [0,1] \times [0,2];\ x > y\}$:
$P(X > Y) = \int \int_T f(x,y)dxdy$ T est le triangle rectangle isocèle de sommets $(0,0),(1,0),(1,1)$ donc :
$P(X > Y) = \int_0^1 \int_0^x f(x,y)dxdy$
On tape :
```
6/7*int(int(x^2+x*y/2,y=0..x),x=0..1)
```
On obtient : `15/56`

(c) $P(Y > 1/2|X < 1/2)$. On a :
$$P(Y > 1/2|X < 1/2) = \frac{P((Y > 1/2) \cap (X < 1/2))}{P(X < 1/2)}$$
On a déja calculer $P(X < 1/2)$:
$P(X < 1/2) = F_X(1/2)$
On tape :
```
subst((6*(2/3*x^3+1/2*x^2))/7,x=1/2)
```
On obtient : `5/28` On calcule $P((Y > 1/2) \cap (X < 1/2))$:
$P((Y > 1/2) \cap (X < 1/2)) = \int_0^{1/2}(\int_{1/2}^2 f(x,y)dy)dx$
On tape :
```
6/7*int(int(x^2+x*y/2,y=1/2..2),x=0..1/2)
```
On obtient : `69/448`
Donc :
$P(Y > 1/2|X < 1/2) = (69/448)/(5/28) = 69/80$

3. Soit f est la fonction **densité de probabilité** du couple de variables aléatoires continues (X,Y) définie par :
$f(x,y) = e^{-(x+y)}$ si $x \in]0,+\infty[$ et $y \in]0,+\infty[$ et $f(x,y) = 0$ sinon.
Calculer la densité de probabilité de Y.
Calculer $P(X < Y)$.

On vérifie tout d'abord que :
$\int_0^{+\infty}(\int_0^{+\infty} e^{-(x+y)}dx)dy = 1$
On tape :
```
int(int(exp(-(x+y)),y=0..inf),x=0..inf)
```
On obtient :
`1`

On a :
$f_Y(y) = \int_0^{+\infty} e^{-(x+y)}dx$
On tape :
`int(exp(-(x+y)),x=0..inf)`
On obtient :
`exp(-y)`
On a si D est la portion de plan $\{x > 0, y > x\}$:
$P(X < Y) = \int \int_D e^{-(x+y)}dx)dy$
On tape :
`int(int(exp(-(x+y)),y=x..inf),x=0..inf)`
On obtient :
`1/2`

4. Soit f est la fonction **densité de probabilité** du couple de variables aléatoires continues (X, Y) définie par :
$f(x, y) = \dfrac{e^{-x/y}e^{-y}}{y}$ si $x \in]0, +\infty[$ et $y \in]0, +\infty[$ et $f(x, y) = 0$ sinon.
Calculer la densité de probabilité de Y.
Calculer la probabilité de l'événement $X > 1$ sachant que $Y = y$.

On vérifie tout d'abord que :
$\int_0^{+\infty}(\int_0^{+\infty} \frac{e^{-x/y}dx)e^{-y}}{y}dy = 1$
On tape :
`assume(y>0)`
`integrate(integrate(exp(-x/y),x=0..inf)*exp(-y)/y,y=0..inf)`
On obtient :
`1`
La densité de probabilité de Y
$f_Y(y) = \int_0^{+\infty} f(x, y)dx = \int_0^{+\infty} \frac{e^{-x/y}e^{-y}}{y}dx$
On tape :
`integrate(exp(-x/y)*exp(-y)/y,x=0..inf)`
On obtient :
`exp(-y)`
La densité de probabilité de X sachant que $Y = y$ est :
$\dfrac{f(x, y)}{f_Y(y)} = \dfrac{\frac{e^{-x/y}e^{-y}}{y}}{e^{-y}} = \dfrac{e^{-x/y}}{y}$
La probabilité de l'événement $X > 1$ sachant que $Y = y$ est :
Prob$(X > 1/Y = y) = 1 -$ Prob$(X < 1/Y = y) = 1 - \int_0^1 \frac{e^{-x/y}}{y}dx$
On tape :
`integrate(exp(-x/y)/y,x=0..1)`
On obtient :
-exp(-(1/y))+1
Donc Prob$(X > 1/Y = y) = e^{-1/y}$

5. Soit f est la fonction densité de probabilité du couple de variables aléatoires continues (X, Y) définie par :
$f(x, y) = \dfrac{6}{7}(x^2 + \dfrac{xy}{2})$ si $x \in [0, 1]$ et $y \in [0, 2]$ et $f(x, y) = 0$ sinon.

Calculer la densité de probabilité de X.
Calculer la probabilité de l'événement $X > Y$.
Calculer la probabilité conditionnelle de X sachant que $y \leq Y \leq y + dy$.

On vérifie tout d'abord que :
$\frac{6}{7}\int_0^1(\int_0^2(x^2 + \frac{xy}{2}dy)dx = 1$
On tape :
`integrate(integrate(6/7*(x^2+x*y/2),y=0..2),x=0..1)`
On obtient :
1
La densité de probabilité de X vaut :
$f_X(x) = \int_0^2 f(x,y)dy = \frac{6}{7}\int_0^2 x^2 + \frac{xy}{2}dy = 2x^2 + x$
La probabilité de l'événement $X > Y$ est :
Prob$(X > Y) = \int\int_T dxdy$ où T est l'intersection du rectangle $R = [0,1] \times [0,2]$ et du demi-plan $x > y$ c'est à dire le triangle T rectangle isocèle de côtés 1.
Prob$(X > Y) = \frac{6}{7}\int_0^1(\int_0^x x^2 + \frac{xy}{2}dy)dx$
On tape :
`integrate(integrate(6/7*(x^2+x*y/2),y=0..x),x=0..1)`
On obtient :
15/56

6. Soient X et Y deux variables aléatoires uniformes et indépendantes sur [0,1].
Calculer la fonction de répartition, la moyenne,la variance et l'écart type des variables aléatoires $Z1 = X + Y$, $Z2 = |X - Y|$ et $Z3 = X/Y$.

On cherche les fonctions de répartition de $Z1$, $Z2$ et de $Z3$ c'est à dire :
$F_{Z1}(z) =$ Prob$(Z1 \leq z) =$ Prob$(X + Y \leq z)$ et
$F_{Z2}(z) =$ Prob$(Z2 \leq z) =$ Prob$(|X - Y| \leq z)$.
$F_{Z3}(z) =$ Prob$(Z3 \leq z) =$ Prob$(X/Y \leq z)$.
X et Y sont deux variables aléatoires indépendantes qui suivent une loi uniforme sur [0,1] donc :
Prob$(X \leq x) = 0$ si $x \leq 0$
Prob$(X \leq x) = x$ si $x \in [0,1]$
Prob$(X \leq x) = 1$ si $x \geq 1$
et
Prob$(Y \leq y) = 0$ si $y \leq 0$
Prob$(Y \leq y) = y$ si $y \in [0,1]$
Prob$(Y \leq y) = 1$ si $y \geq 1$
Donc puisque X et Y sont deux variables aléatoires indépendantes :
la densité de probabilité du couple (X,Y) est le produit de la densité de probabilité de X par la densité de probabilité de Y :
$f_{(X,Y)}(x,y) = f_X(x) * f_Y(y)$ c'est à dire :
$f_{(X,Y)}(x,y) = 0$ si $(x,y) \notin [0,1] \times [0,1]$
$f_{(X,Y)}(x,y) = xy$ si $(x,y) \in [0,1] \times [0,1]$

(a) **Étude de** $Z1$

Prob$(X+Y \leq z) = \int\int_D dxdy$ où D est l'intersection du carré C ($C = [0,1] \times [0,1]$) et de $\{(x,y);\ y+x \leq z\}$.
On va considérer 4 cas :

— si $z \leq 0$, l'intersection de $\{(x,y);\ y+x \leq z\}$ et du carré C est vide donc $F_{Z1}(z) = \text{Prob}(Z1 \leq z) = 0$ et $f_{Z1}(z) = 0$

— si $0 < z \leq 1$, l'intersection D de $\{(x,y);\ y+x \leq z\}$ et du carré C est un le triangle rectangle isocèle de côtés z donc :
$F_{Z1}(z) = \text{Prob}(Z1 \leq z) = \int\int_D dxdy = z^2/2$ et $f_{Z1}(z) = z$

— si $1 < z < 2$, l'intersection D de $\{(x,y);\ y+x \leq z\}$ et du carré C est un le carré privé d'un triangle rectangle isocèle T de côtés $2-z$ donc on a :
$F_{Z1}(z) = \text{Prob}(Z1 \leq z) = \int\int_D dxdy = 1-(2-z)^2/2$ et $f_{Z1}(z) = 2-z$

— si $z \geq 2$, l'intersection de $\{(x,y);\ y+x \leq z\}$ et du carré C est C donc $F_{Z1}(z) = \text{Prob}(Z1 \leq z) = 1$ et $f_{Z1}(z) = 0$

Donc :

$$f_{Z1}(z) = \begin{cases} z & \text{si} \quad 0 \leq z \leq 1 \\ 2-z & \text{si} \quad 1 \leq z \leq 2 \\ 0 & \text{si} \quad z \leq 0 \text{ ou } z \geq 2 \end{cases}$$

Calcul de $E(Z1)$
On a :
$E(Z1) = \int_{-\infty}^{+\infty} z f_{Z1}(z)dz = \int_0^1 z^2 dz + \int_1^2 z(2-z)dz =$
$1/3 + 3 - 8/3 + 1/3 = 1$
Ou on tape :

```
int(z^2,z=0..1)+int(z*(2-z),z=1..2)
```

On obtient : `1`
Donc $E(Z1) = 1$
Calcul de $V(Z1)$ **et de** $\sigma(Z1)$
On a :
$V(Z1) = \int_{-\infty}^{+\infty} (z-1)^2 f_{Z1}(z)dz =$
$\int_0^1 z(z-1)^2 dz + \int_1^2 (z-1)^2(2-z)dz$
On tape :

```
int(z*(z-1)^2,z=0..1)+int((z-1)^2*(2-z),z=1..2)
```

On obtient : `1/6`
Donc $V(Z1) = 1/6$ et $\sigma(Z1) = \sqrt{6}/6$.

(b) **Étude de** $Z2$
Prob$(|X-Y| \leq z) = \int\int_D dxdy$ où D est l'intersection de $\{(x,y)/|x-y| \leq z\}$ et du carré C =$[0,1] \times [0,1]$.
On va considérer 3 cas :

— si $z \leq 0$, l'intersection de $\{(x,y)/|x-y| \leq z\}$ et du carré C est vide puisque $\{(x,y)/|x-y| \leq z\}$ est vide donc $F_{Z2}(z) = \text{Prob}(Z2 \leq z) = 0$ et $f_{Z2}(z) = 0$

— si $0 < z \leq 1$, l'intersection B de $\{|x-y| \leq z\}$ et du carré C est une portion du carré limité par les droites $y = x+z$ et $y = x-z$ c'est à dire C privé de deux triangles rectangles isocèles de côtés

$1-z$ donc :
$F_{22}(z) = \text{Prob}(Z2 \leq z) = \int\int_B dxdy = 1 - 2*(1-z)^2/2 = 2z - z^2$ et $f_{Z2}(z) = 2 - 2z$

— si $z \geq 1$, l'intersection de $\{(x,y)/y + x \leq z\}$ et du carré C est C donc $F_{Z2}(z) = \text{Prob}(Z2 \leq z) = 1$ et $f_{Z2}(z) = 0$

Donc :

$$f_{Z2}(z) = \begin{cases} 2-2z & \text{si} \quad 0 \leq z \leq 1 \\ 0 & \text{si} \quad z \leq 0 \text{ ou } z \geq 1 \end{cases}$$

Calcul de $E(Z2)$

On a :

$$E(Z2) = \int_{-\infty}^{+\infty} z f_{Z2}(z)dz = \int_0^1 z(2-2z)dz = 1 - 2/3 = 1/3$$

Ou on tape :

```
int(z*(2-2z),z=0..1)
```

On obtient : `1/3`

Donc $E(Z2) = 1/3$

Calcul de $V(Z2)$ **et de** $\sigma(Z2)$

On a :

$$V(Z2) = \int_{-\infty}^{+\infty} (z-1)^2 f_{Z1}(z)dz = \int_0^1 (2-2z)(z-1/3)^2 dz$$

On tape :

```
int((2-2z)*(z-1/3)^2,z=0..1)
```

On obtient : `1/18`

Donc $V(Z2) = 1/18$ et $\sigma(Z2) = \sqrt{2}/6$

(c) **Étude de** $Z3$

$\text{Prob}(X/Y \leq z) = \int\int_D dxdy$ où D est l'intersection du carré C ($C = [0,1] \times [0,1]$) et de $\{(x,y);\ x/y \leq z\}$.

On va considérer 3 cas :

— si $z \leq 0$, l'intersection de $\{(x,y);\ x/y \leq z\}$ et du carré C est vide donc $F_{Z3}(z) = \text{Prob}(Z3 \leq z) = 0$ et $f_{Z3}(z) = 0$

— si $0 < z \leq 1$, l'intersection T de $\{(x,y);\ x/y \leq z\}$ et du carré C est un le triangle rectangle de côtés 1 et z donc :
$F_{Z3}(z) = \text{Prob}(Z3 \leq z) = \int\int_T dxdy = z/2$ et $f_{Z3}(z) = 1/2$

— si $1 < z$, l'intersection D de $\{(x,y);\ x/y \leq z\}$ et du carré C est un le carré privé d'un triangle rectangle T de côtés 1 et $1/z$ donc on a :
$F_{Z3}(z) = \text{Prob}(Z3 \leq z) = \int\int_D dxdy = 1 - 1/(2*z)$ et $f_{Z1}(z) = 1/(2*z^2)$

Donc :

$$f_{Z3}(z) = \begin{cases} 0 & \text{si} \quad z \leq 0 \\ 1/2 & \text{si} \quad 0 \leq z \leq 1 \\ 1/(2*z^2) & \text{si} \quad z \geq 1 \end{cases}$$

Calcul de $E(Z3)$

On ne peut pas calculer $E(Z3)$ car :

$+\int_1^{+\infty} z/(2*z^2)dz$ n'est pas convergente.

7. Soient X une variable aléatoire uniforme sur $[0,1]$ et Y une variable aléatoire qui suit une loi exponentielle de paramètre $\lambda = 1$.œn suppose X et Y

indépendantes.
Calculer la fonction de répartition, la moyenne,la variance et l'écart type des variables aléatoires $Z1 = X + Y$, $Z2 = |X - Y|$ et $Z3 = X/Y$.

On a :
$f_X(x) = 1$ pour $0 \leq x \leq 1$ et
$f_Y(y) = 0$ pour $x < 0$ ou $x > 1$ donc :
Prob$(X \leq x) = 0$ si $x \leq 0$
Prob$(X \leq x) = x$ si $x \in [0, 1]$
Prob$(X \leq x) = 1$ si $x \geq 1$
On a :
$f_Y(y) = \exp(-y)$ pour $y \geq 0$ et
$f_Y(y) = 0$ pour $y < 0$ donc :
Prob$(Y \leq y) = 0$ si $y \leq 0$
Prob$(Y \leq y) = 1 - \exp(-y)$ si $y \in [0, \infty[$

On cherche les fonctions de répartition de $Z1$, $Z2$ et de $Z3$ c'est à dire :
$F_{Z1}(z) =$ Prob$(Z1 \leq z) =$ Prob$(X + Y \leq z)$ et
$F_{Z2}(z) =$ Prob$(Z2 \leq z) =$ Prob$(|X - Y| \leq z)$.
$F_{Z3}(z) =$ Prob$(Z3 \leq z) =$ Prob$(X/Y \leq z)$.

Donc puisque X et Y sont deux variables aléatoires indépendantes :
la densité de probabilité du couple (X, Y) est le produit de la densité de probabilité de X par la densité de probabilité de Y :
$f_{(X,Y)}(x, y) = f_X(x) * f_Y(y)$ c'est à dire :
Si $D = D = [0, 1] \times [0, +\infty[$ on a $f_{(X,Y)}(x, y) = 0$ si $(x, y) \notin D$
$f_{(X,Y)}(x, y) = \exp(-y)$ si $(x, y) \in D$

(a) **Étude de** $Z1$
Prob$(X + Y \leq z) = \int \int_K \exp(-y)dxdy$ où K est l'intersection de D ($D = [0, 1] \times [0, +\infty[$) et de $\{(x, y);\ y + x \leq z\}$.
On va considérer 3 cas :
— si $z \leq 0$, l'intersection de $\{(x, y);\ y + x \leq z\}$ et de D est vide donc $F_{Z1}(z) =$ Prob$(Z1 \leq z) = 0$ et $f_{Z1}(z) = 0$
— si $0 < z \leq 1$, l'intersection K de $\{(x, y);\ y + x \leq z\}$ et de D est un le triangle rectangle isocèle de côtés z donc :
$F_{Z1}(z) =$ Prob$(Z1 \leq z) = \int_0^z (\int_0^x \exp(-y)dy)dx$
On tape :
`int(int(exp(-y),y=0..x),x=0..z)`
On obtient : `z+exp(-z)-1`
donc $F_{Z1}(z) =$ Prob$(Z1 \leq z) = z + \exp(-z) - 1$ et $f_{Z1}(z) = 1 - \exp(-z)$
— si $1 < z < +\infty$, l'intersection K de $\{(x, y);\ y + x \leq z\}$ et de D est un le trapèze rectangle de hauteur 1 et de côtés z et $z - 1$ donc on a :
$F_{Z1}(z) =$ Prob$(Z1 \leq z) = \int_0^1 (\int_0^{z-x} \exp(-y)dy)dx$.
On tape :

`int(int(exp(-y),y=0..z-x),x=0..1)`
On obtient : `-exp(-z+1)+1+exp(-z)`
donc $F_{Z1}(z) = \text{Prob}(Z1 \leq z) = -\exp(-z+1)+1+\exp(-z)-1$
et $f_{Z1}(z) = \exp(-z+1) - \exp(-z)$
Donc :

$$f_{Z1}(z) = \begin{cases} 0 & \text{si} \quad z \leq 0 \\ 1-\exp(-z) & \text{si} \quad 0 \leq z \leq 1 \\ \exp(-z+1)-\exp(-z) & \text{si} \quad 1 \leq z \end{cases}$$

Calcul de $E(Z1)$
On a :
$E(Z1) = \int_{-\infty}^{+\infty} z f_{Z1}(z)dz =$
$\int_0^1 z - z\exp(-z)dz + \int_1^{+\infty} z(\exp(-z+1)-\exp(-z))dz$
On tape :
`normal(int(z-z*exp(-z),z=0..1)+`
`int(z*exp(-z+1)-z*exp(-z),z=1..inf))`
On obtient : `3/2`
Donc $E(Z1) = 3/2$
Calcul de $V(Z1)$ **et de** $\sigma(Z1)$
On a :
$V(Z1) = \int_{-\infty}^{+\infty} (z-3/2)^2 f_{Z1}(z)dz = \int_0^1 (z-3/2)^2(1-\exp(-z))dz+$
$\int_1^{+\infty} (z-3/2)^2(\exp(-z+1)-\exp(-z))dz$
On tape :
`int((z-3/2)^2*(1-exp(-z)),z=0..1)+`
`int((z-3/2)^2*(exp(-z+1)-exp(-z)),z=1..inf)`
On obtient : `13/12`
Donc $V(Z1) = 13/12$ et $\sigma(Z1) = \sqrt{1}3/(2\sqrt{3})$.

(b) **Étude de** $Z2$
Prob$(|X-Y| \leq z) = \int\int_D dxdy$ où D est l'intersection de $\{(x,y)/|x-y| \leq z\}$ et de D =$[0,1] \times [0,+\infty[$.
On va considérer 3 cas :

— si $z \leq 0$, l'intersection de $\{(x,y)/|x-y| \leq z\}$ et de D est vide puisque $\{(x,y)/|x-y| \leq z\}$ est vide donc $F_{Z2}(z) = \text{Prob}(Z2 \leq z) = 0$ et $f_{Z2}(z) = 0$

— si $0 < z \leq 1$, l'intersection B de $\{|x-y| \leq z\}$ et de D est une portion de D limitée par les droites $y = x+z$ et $y = x-z$ donc :
$F_{22}(z) = \text{Prob}(Z2 \leq z) = \int\int_D f(x,y)dxdy$
$F_{22}(z) = \int_0^z (\int_0^{x+z} \exp(-y)dy)dx + \int_z^1 (\int_{x-z}^{x+z} \exp(-y)dy)dx$
On tape :
`normal(int(int(exp(-y),y=0..x+z),x=0..z)+`
`int(int(exp(-y),y=x-z..x+z),x=z..1))`
On obtient : `z+exp(-(z+1))-exp(-z)-exp(z-1)+1`

Donc $F_{22}(z) = z + \exp(-(z+1)) - \exp(-z) - \exp(z-1) + 1$
et $f_{Z2}(z) = 1 - \exp(-(z+1)) + \exp(-z) - \exp(z-1)$

— si $z \geq 1$, l'intersection de $\{(x,y)/y+x \leq z\}$ et de D est la portion de D située sous la droite $y = x + z$ donc $F_{Z2}(z) = \text{Prob}(Z2 \leq z) = \int_0^1 (\int_0^{x+z} \exp(-y)dy)dx$ On tape :
`int(int(exp(-y),y=0..x+z),x=0..1)`
On obtient : `exp(-1-z)+1-exp(-z)`
Donc $F_{22}(z) = \exp(-1-z)+1-\exp(-z)$ et $f_{Z2}(z) = -\exp(-1-z) + \exp(-z)$

Donc :

$$f_{Z2}(z) = \begin{cases} 0 & \text{si} \quad z \leq 0 \\ 1 - \exp(-(z+1)) + \exp(-z) - \exp(z-1) & \text{si} \quad 0 \leq z \leq 1 \\ -\exp(-1-z) + \exp(-z) & \text{si} \quad z \geq 1 \end{cases}$$

Calcul de $E(Z2)$
On a :
$E(Z2) = \int_{-\infty}^{+\infty} z f_{Z2}(z)dz$
On tape :
`normal(int(z*(1-exp(-(z+1))+exp(-z)-exp(z-1)),z=0..1)+`
`int(z*(-exp(-1-z)+exp(-z)),z=1..inf))`
On obtient : `-2*exp(-1)+3/2`
Donc $E(Z2) = -2\exp(-1) + 3/2$
Calcul de $V(Z2)$ **et de** $\sigma(Z2)$
On a :
$V(Z2) = \int_{-\infty}^{+\infty} (z + 2\exp(-1) - 3/2)^2 f_{Z1}(z)dz$
On tape :
`normal(int((z+2*exp(-1)-3/2)^2*(1-exp(-(z+1))+exp(-z)-exp(z-1)`
`int((z+2*exp(-1)-3/2)^2*(-exp(-1-z)+exp(-z)),z=1..inf))`
On obtient : `6*exp(-1)-4*exp(-2)+(-11)/12`
Donc $V(Z2) = 6\exp(-1) - 4\exp(-2) + (-11)/12$ et $\sigma(Z2) = \sqrt{V(Z2)}$

(c) **Étude de** $Z3$
$\text{Prob}(X/Y \leq z) = \int\int_K dxdy$ où K est l'intersection de D ($D = [0,1] \times [0,+\infty[$) et de $\{(x,y);\ x/y \leq z\}$.
On va considérer 2 cas :

— si $z \leq 0$, l'intersection de $\{(x,y);\ x/y \leq z\}$ et de D est vide donc $F_{Z3}(z) = \text{Prob}(Z3 \leq z) = 0$ et $f_{Z3}(z) = 0$

— si $0 < z$, l'intersection K de $\{(x,y);\ x/y \leq z\}$ et de D est égal à D privé du triangle rectangle de côtés 1 et $1/z$ donc :
$F_{Z3}(z) = \text{Prob}(Z3 \leq z) = \int\int_D f(x,y)dxdy = z/2$
On tape :
`int(int(exp(-y),y=x/z..inf),x=0..1)`
On obtient : `-z*exp(-(1/z))+z`
Donc :
$F_{Z3}(z) = -z\exp(-(1/z)) + z$ et $f_{Z3}(z) = (-z * exp(-(1/z)) + z - exp(-(1/z)))/z$

Donc :

$$f_{Z3}(z) = \begin{cases} 0 & \text{si} \quad z \leq 0 \\ (-z\exp(-(1/z)) + z - \exp(-(1/z)))/z & \text{si} \quad 0 \leq z \end{cases}$$

Calcul de $E(Z3)$
On a :
$E(Z3) = \int_{-\infty}^{+\infty} z f_{Z2}(z)dz$
On tape :
`romberg((-z*exp(-1/z)+z-exp(-1/z)),z=0..1e20)`
On obtient : `-2*exp(-1)+3/2`
Donc $E(Z2) \sim 7.8$
Calcul de $V(Z3)$ **et de** $\sigma(Z3)$
On a :
$V(Z3) = E(Z3^2) - (E(Z3))^2$
$E(Z3^2) = \int_0^{+\infty} z * (-z * exp(-1/z) + z - exp(-1/z))dz$
On tape :
`limit(z*(-z*exp(-1/z)+z-exp(-1/z)),z=inf)`
On obtient : `1/2`
Donc l'intégrale calculant $E(Z3^2)$ diverge donc on ne peut pas calculer la variance de $Z3$.

8. Soit f est la fonction **densité de probabilité** du couple de variables aléatoires continues (X, Y) définie pour $c \in \mathbb{R}$ par :
$f(x, y) = c(x^2 - y^2)\exp(-x)$ si $x \in]0, +\infty[$ et $y \in [-x, x]$ et$f(x, y) = 0$ sinon.
Calculer :

   (a) la valeur de c,

   (b) la probabilité conditionnelle de Y sachant que $x \leq X \leq x + dx$.

   (a) On doit avoir :
   $I = c\int_0^{+\infty}(\int_{-x}^{x}(x^2 - y^2)\exp(-x)dy)dx = 1$
   On a :
   $I = c\int_0^{+\infty} 2(x^3 - x^3/3)\exp(-x)dx = 4c/3\int_0^{+\infty} x^3\exp(-x)dx$
   La primitive de $x^3\exp(-x)$ est de la forme $(ax^3+bx^2+cx+d)\exp(-x)$ donc on a pour tout x :
   $ax^3 + (3a - b)x^2 + (2b - c)x + c - d = x^3$ soit :
   $a = 1, b = 3, c = 6$ $d = 6$ c'est à dire :
   $I = 4c/3\int_0^{+\infty} x^3\exp(-x)dx = 4c/3 * 6 = 1$ Donc $8c = 1$ d'où $c = \frac{1}{8}$.
   Ou on tape avec `Xcas` :
   `int(int((x^2-y^2)*exp(-x),y=-x..x),x=0..inf)`
   On obtient : `8`
   Donc $8c = 1$ d'où

$$c = \frac{1}{8}$$

(b) la probabilité conditionnelle de Y sachant que $x \leq X \leq x + dx$ est :

$$\text{Prob}(Y < y0|X = x) = \int_{-x}^{y0} \frac{f(x,y)}{f_X(x)} dy$$

Calcul de $f_X(x)$:
$f_X(x) = \int_{-x}^{x} f(x,y)dy = c\int_{-x}^{x}(x^2-y^2)\exp(-x)dy = 4c/3x^3\exp(-x)$
donc :
Prob$(Y < y0|X = x) = \int_{-x}^{y0} \frac{3(x^2-y^2)}{4x^3}dy = \int_{-x}^{y0} \frac{3}{4x} - \frac{3y^2)}{4x^3}dy =$
$\frac{3y0}{4x} + \frac{3}{4} - \frac{y0^3)}{4x^3} - \frac{1}{4}$
donc
Prob$(Y < y|X = x) = \frac{1}{2} + \frac{3y}{4x} - \frac{y^3)}{4x^3}$
Ou on tape avec `Xcas` :

```
int((x^2-y^2)*exp(-x)/int((x^2-y^2)*exp(-x),y=-x..x),y=-x..y)
```

On obtient : `(3*x^2*y-y^3)/(4*x^3)-(1/-2)`

Chapitre 3

Résumé de statistique descriptive

3.1 Généralités

La statistique a pour objet de recueillir des observations portant sur des sujets présentant une certaine propriété et de traduire ces observations par des nombres qui permettent d'avoir des renseignements sur cette propriété.
Le but de la statistique descriptive est de structurer et de représenter l'information contenue dans les données.
La **population** est l'ensemble des sujets observés.
Le **caractère** est la propriété étudiée sur ces sujets.

3.2 Statistique à 1 variable

3.2.1 Série statistique qualitative

Lorsque le caractère étudié est qualitatif, chaque caractère sera indexé et pour chaque variété du caractère, on indiquera le nombres des membres de la population ayant cette variété : c'est une **série statistique qualitative**.
Exemple :
On considère comme population 100 nouveau-nés et le caractère est le sexe.
On indexe les garçons par `G` et les filles par `F`.
La série sera par exemple :
`G:63, F:37`

3.2.2 Série statistique quantitative

Lorsque le caractère étudié est exprimable directement par un nombre, l'énumération des nombres exprimant la valeur de ce caractère pour chaque membre de la population étudiée est une **série statistique quantitative**.
Exemple :
On considère comme population 20 adolescents et le caractère est la taille exprimée en centimètres.
La série est obtenue par simple énumération :
`155,147,153,154,155,148,151,162,144,159,156,156,161,154,153,171,165,159,154,155`
On obtient une **série statistique d'effectifs égaux à 1**.

La série sera plus lisible si on note pour chaque valeur du caractère le nombre de personnes présentant ce caractère : on obtient une **série statistique avec effectifs**.

"taille"	"effectif"
144	1
147	1
148	1
151	1
153	2
154	3
155	3
156	2
159	2
161	1
162	1
165	1
171	1

Une présentation de ce type s'impose quand la population est grande.
On peut aussi, puisque le caractère n'est discret que par convention, utiliser des **classes** par exemple d'**étendue** 1 cm ou 2 cm pour avoir une représentation plus globale.
On a alors en utilisant des classes d'étendue 2cm :

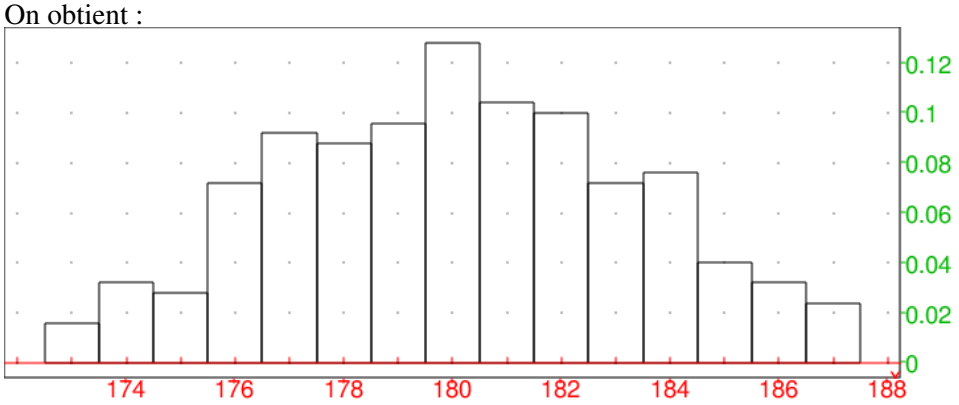
"taille x"	"effectif"
$143.5 \leq x < 145.5$	1
$145.5 \leq x < 147.5$	1
$147.5 \leq x < 149.5$	1
$149.5 \leq x < 151.5$	1
$151.5 \leq x < 153.5$	2
$153.5 \leq x < 155.5$	6
$155.5 \leq x < 157.5$	2
$157.5 \leq x < 159.5$	2
$159.5 \leq x < 161.5$	1
$161.5 \leq x < 163.5$	1
$163.5 \leq x < 165.5$	1
$165.5 \leq x < 167.5$	0
$167.5 \leq x < 169.5$	0
$169.5 \leq x < 171.5$	1

Ainsi, la fréquence de 155 est 3/20,
et la fréquence cumulée de 155 est :(1+1+1+1+2+3+3)/20=12/20. **Exercice** Le but de l'activité est l'étude de la taille (en cm) portant sur 250 individus jouant au basket.

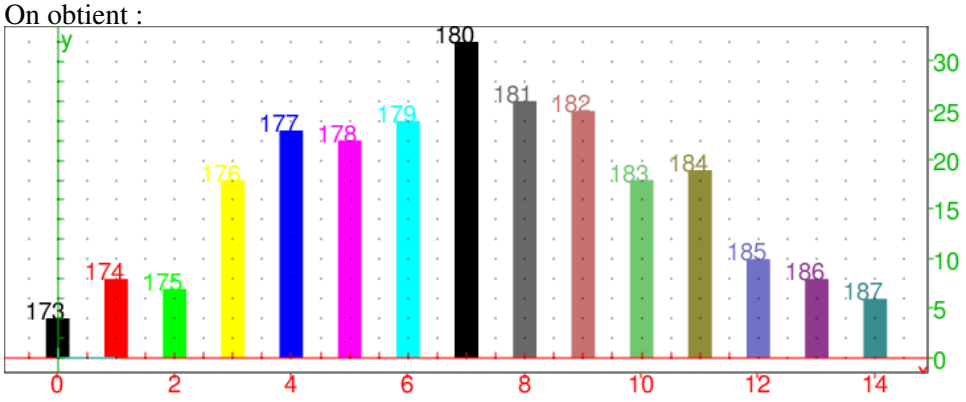
Taille	173	174	175	176	177	178	179	180	181	182	183	184	185	186	18
Effectif	4	8	7	18	23	22	24	32	26	25	18	19	10	8	6

L'activité commence par un calcul des paramètres statistiques puis se poursuit avec des représentations graphiques : diagrammes en bâtons, diagramme en boîte et polygone des fréquences cumulées croissantes. On tape :

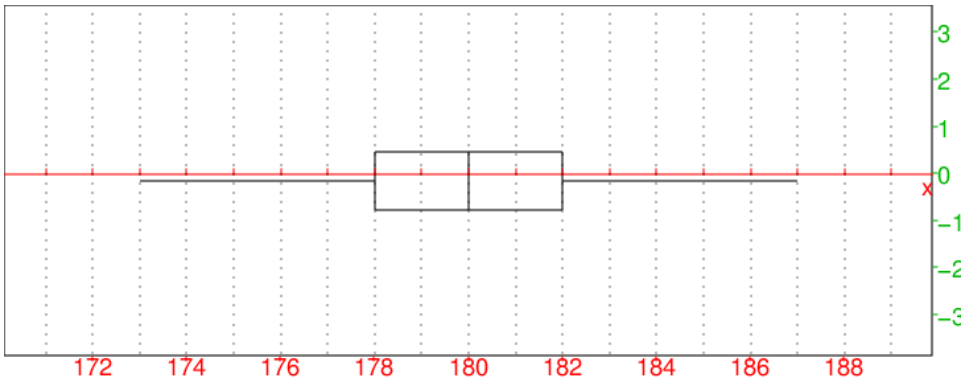
```
T:=(173+j)$(j=0..14)
```
On obtient :
```
173,174,175,176,177,178,179,180,181,182,183,184,185,186,187
```
On tape :
```
Ef:=(4 ,8 ,7, 18, 23, 22, 24, 32, 26, 25, 18, 19, 10, 8,
6)
sum(Ef)
```
On obtient :
```
250
```
On tape :
```
histogramme(tran([[T],[Ef]]))
```
On obtient :



On tape :
```
diagramme_batons([[T],[Ef]])
```
On obtient :



On tape :
```
moustache([T],[Ef])
```
On obtient :



On tape :

```
Efc:=(sum(Ef[j],j=0..k)/250.)$(k=0..size(Ef)-1)
```

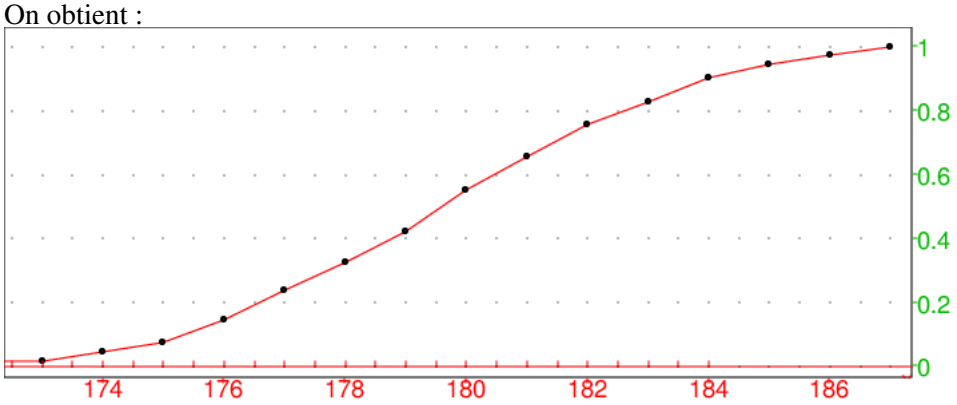
On obtient :

```
0.016,0.048,0.076,0.148,0.24,0.328,0.424,0.552,0.656,0.756,0.828,0.90
```

On tape :

```
affichage(nuage_points([[T],[Efc]])), point_point+epaisseur_point_2),
```

On obtient :



On tape :

```
approx(mean([T],[Ef]))
```

On obtient :

```
180.104
```

On tape :

```
approx(sttdev([T],[Ef]))
```

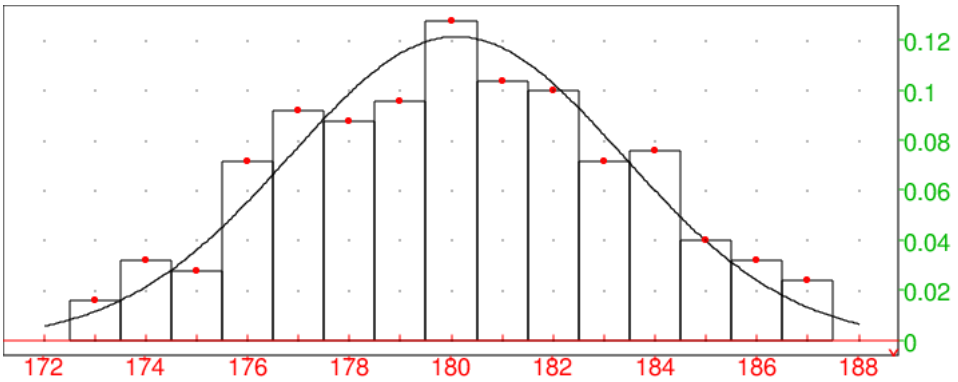
On obtient :

```
3.27859482096
```

On tape :

```
histogramme(tran([[T],[Ef]])),
plotfunc(loi_normale(180.104,3.27859482096,x),x=172..188),
nuage_points(tran([[T],[Ef]/250.]),affichage=1+point_point+epaisseur_
```

On obtient :



3.2.3 Vocabulaire des séries quantitatives à 1 variable

Soit une série quantitative à 1 variable `L`.
La différence entre la plus grande valeur et la plus petite valeur du caractère effectivement obtenue est l'**étendue** de la série `L`.
Le nombre de membres de la population étudiée est l'**effectif total**.
Si le caractère est discret, il est commode d'indiquer pour chaque valeur du caractère, le nombre des membres de la population ayant cette valeur : c'est l'**effectif de cette valeur**.
Si le caractère est continu, on partage l'intervalle sur lequel s'étendent ces valeurs en intervalles (en général égaux) que l'on appelle **classe**. Le nombre des membres de la population ayant leur valeur dans une classe est l'**effectif de cette classe**.
La valeur moyenne des bornes d'une classe est le **centre** de cette classe.
L'**effectif cumulé** d'une valeur (ou d'une classe) est la somme de l'effectif de cette valeur (ou de cette classe) et de tous les effectifs des valeurs (ou des classes) qui précèdent.
La **fréquence** d'une valeur (ou d'une classe) est le rapport de l'effectif de cette valeur (ou de cette classe) par l'effectif total.
Avec `Xcas` on tape par exemple :

```
frequences([1,2,1,1,2,1,2,4,3,3])
```

On obtient ;

```
[[1,0.4],[2,0.3],[3,0.2],[4,0.1]]
```

La **fréquence cumulée** d'une liste de valeurs (ou d'une classe) est la somme de la fréquence de cette valeur (ou de cette classe) et de toutes les fréquences des valeurs (ou des classes) qui la précèdent.
Avec `Xcas` on tape par exemple :

```
frequences_cumulee([[0.75,30],[1.75,50],[2.75,20]])
```

ou

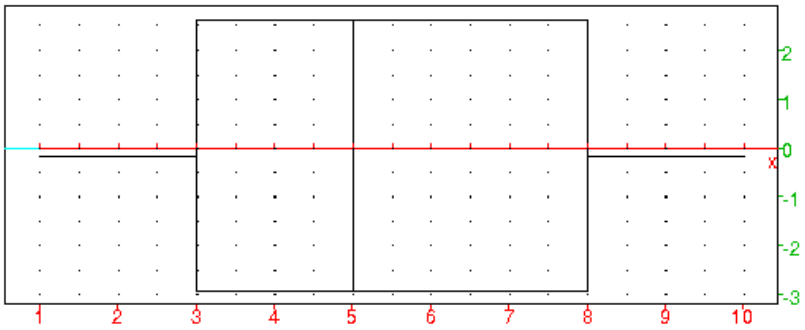
```
frequences_cumulee([[0.25..1.25,30],[1.25..2.25,50],
[2.25..3.25,20]])
```

On obtient le diagramme des fréquences cumulées.
L'**histogramme** des effectifs (resp fréquences) d'un caractère discret ou continu est le graphique qui permet de visualiser l'effectif (resp fréquences) des différentes valeurs du caractère : on met en abscisse les différentes valeurs du caractère (ou le centre des différentes classes), puis on forme des rectangles accolés deux à deux, ses rectangles ont deux cotés parralléles à l'axe des ordonnées, le coté porté par l'axe des abscisses a pour longueur l'amplitude de la classe, et l'autre est tel que

l'aire du rectangle est égale à l'effectif (resp fréquences) de la valeur considérée.
L'**histogramme des fréquences** permet de visualiser les fréquences des différentes classes au moyen de la surface de rectangles : chaque rectangle correspond à une classe et a pour surface la fréquence de cette classe.
Avec `Xcas` on tape par exemple :
`histogramme([[0.75,30],[1.75,50],[2.75,20]])`
ou
`histogramme([[0.25..1.25,30],[1.25..2.25,50],`
`[2.25..3.25,20]])`
On obtient un histogramme des fréquences.
La **fonction de répartition des fréquences** est égale pour chaque valeur du caractère à la fréquence cumulée de cette valeur.
Le **mode** est la valeur du caractère dont l'effectif est le plus grand.
Le **maximum** est la plus grande valeur du caractère effectivement obtenue.
Le **minimum** est la plus petite valeur du caractère effectivement obtenue.
La **médiane** partage la série statistique en deux groupes de même effectif. C'est une valeur du caractère à partir de laquelle l'effectif des valeurs qui lui sont inférieures est superieur ou ègal à l'effectif des valeurs qui lui sont supérieures (par exemple la médiane de [140,145,146,147] est 146 et la médiane de [140,145,146] est 145). La médiane est donc la valeur du caractère à partir de laquelle la fréquence cumulée atteint ou dépasse 0.5.
Les **quartiles** sont trois valeurs du caractère qui partage la série statistique en quatre groupes de même effectif :
- le **1-ier quartile** est la valeur du caractère à partir de laquelle la fréquence cumulée atteint ou dépasse 0.25.
- le **2-ième quartile** est confondu avec la médiane.
- le **3-ième quartile** est la valeur du caractère à partir de laquelle la fréquence cumulée atteint ou dépasse 0.75.

On peut définir les **déciles**. Il y a 9 déciles :
le **1-ier décile** est la valeur du caractère à partir de laquelle la fréquence cumulée atteint ou dépasse 0.1.
le **2-ième décile** est la valeur du caractère à partir de laquelle la fréquence cumulée atteint ou dépasse 0.2.
etc...
le **9-ième décile** est la valeur du caractère à partir de laquelle la fréquence cumulée atteint ou dépasse 0.9.
On peut aussi définir le **centile** (il y a 99 centiles) et le **quantile** d'ordre **p** :
le **1-ier centile** est la valeur du caractère à partir de laquelle la fréquence cumulée atteint ou dépasse 0.01.
etc...
le **99-ième centile** est la valeur du caractère à partir de laquelle la fréquence cumulée atteint ou dépasse 0.99.
Le **quantile** d'ordre **p** (**p** un réel de [0,1[), est la valeur du caractère à partir de laquelle la fréquence cumulée atteint ou dépasse **p**.
Le **semi-interquartile** est égal à $\frac{1}{2}(Q_3 - Q_1)$ où Q_1 et Q_3 désigne le premier et le troisième quartile. Cet indice fournit un renseignement sur l'étalement des valeurs de part et d'autre de la médiane.
L'**interquartile** est égal à $Q_3 - Q_1$ où Q_1 et Q_3 désigne le premier et le troisième

quartile. Cet indice fournit un renseignement sur l'étalement des valeurs de part et d'autre de la médiane.
L'**interdécile** est égal à $D_9 - D_1$ où D_1 et D_9 désigne le premier et le neuvième décile. Cet indice fournit un renseignement sur l'étalement des valeurs de part et d'autre de la médiane.
Exemples avec `Xcas`
On tape :
`L:=[1,2,3,4,5,6,7,8,9,10]`
`min(L)` et on obtient `1`
`quartile1(L)` et on obtient `3.0`
`median(L)` et on obtient `5.0`
`quartile3(L)` et on obtient `8.0`
`max(L)` et on obtient `10`
`quartiles(L)` pour avoir le résultat des 5 commandes précédentes et on obtient
`[[1.0],[3.0],[5.0],[8.0],[10.0]]`
`quantile(L,0.9)` et on obtient `9.0`
La **boite à moustaches** permet de visualiser ces différentes valeurs :
c'est un rectangle dont un coté est un trait allant de Q_1 à Q_3 sur lequelle un trait vertical indique la valeur de la médiane et d'où deux traits horizontaux (les moustaches) débordent : l'un va de la valeur minimum à Q_1 et l'autre de Q_3 à la valeur maximum. Sur ces deux moustaches, on trouvent quelquefois deux traits verticaux indiquant la valeur du premier et du neuvième décile.
Avec `Xcas` on tape :
`L:=[1,2,3,4,5,6,7,8,9,10]`
`moustache(L)`
Cela ouvre le graphique et dessine une boite à moustaches où on peut lire que :
Q2=médiane=5, Q1=3, Q3=8, minimum=1, maximum=10.



La **moyenne** est le quotient de la somme des valeurs du caractère (pas toujours distinctes) par l'effectif total. Si le caractère prend n valeurs distinctes x_k d'effectifs e_k pour $k = 0...(n-1)$ alors l'effectif total vaut $N = \sum_{k=0}^{n-1} e_k$ et la moyenne m est : $m = \dfrac{1}{N}\sum_{k=0}^{n-1} e_k x_k$.
La **variance** est la moyenne des carrés des écarts à la moyenne des valeurs du caractère. Si le caractère prend n valeurs distinctes x_k d'effectifs e_k ($k = 0...(n-1)$), si la moyenne vaut m et, si l'effectif total vaut N alors la variance $v = s^2$ est :

$s^2 = \frac{1}{N}\sum_{k=0}^{n-1} e_k(x_k - m)^2 = \frac{1}{N}(\sum_{k=0}^{n-1} e_k x_k^2) - m^2.$

L'**écart-type** s est la racine carrée de la variance.

Soit une série statistique quantitative d'effectif N à 1 variable, un **échantillon d'ordre n** désigne le système des n valeurs prises par le caractère au cours de n tirages indépendants. Les valeurs prises par l'échantillon sont donc les valeurs prises par n variables aléatoires $X_1, ..., X_n$ qui suivent la même loi que la variable aléatoire X égale à la valeur du caractère étudié. Par exemple, si dans une ville de N habitants, on étudie la taille (exprimée en centimètres) de ses habitants, la taille de 100 personnes prises au hasard dans cette ville est un échantillon d'ordre 100.
En général, on ignore la loi de la variable aléatoire égale à la taille des habitants de cette ville, et on veut dégager un certain nombre d'éléments caractéristiques de cette variable grâce à l'échantillon.

3.3 Série statistique quantitative à 2 variables

3.3.1 Définition

Lorsque pour une population donnée, on étudie deux caractères qui sont exprimables chacun directement par un nombre, l'énumération des couples de nombres exprimant les valeurs de ces deux caractères pour chaque membre de la population étudiée est une **série statistique quantitative à 2 variables**.
Exemple :
Dans une classe de Terminale S, si à chaque élève on associe son poids (en kilogrammes) et sa taille (en centimètres), on obtient une série statistique à 2 variables.

3.3.2 Vocabulaire des séries quantitatives à 2 variables

Soit une série statistique quantitative d'effectif N, à 2 variables, un **échantillon d'ordre n** désigne le système des n couples de valeurs prises par ces 2 variables au cours de n tirages indépendants. Par exemple, si dans une ville de N habitants on étudie la taille et le poids de ses habitants, la taille et le poids de 100 personnes prises au hasard est un échantillon de statistique à 2 variables d'ordre 100. On essaye de déterminer si ces 2 variables sont indépendantes en calculant par exemple leur coefficient de corrélation.

3.3.3 Moyennes, variances, covariances d'effectif 1

Soit une série statistique à deux variables d'ordre n pour les caractères X et Y représentée par les couples $(x_j,\ y_j)$ pour $0 \leq j \leq (n-1)$.
Ici les x_j (resp y_j) ne sont pas forcément distincts.
La moyenne de X est : $\bar{x} = \frac{1}{n}\sum_{j=0}^{n-1} x_j$.
La moyenne de Y est : $\bar{y} = \frac{1}{n}\sum_{j=0}^{n-1} y_j$.
La variance de X est : $\sigma^2(X) = \frac{1}{n}\sum_{j=0}^{n-1}(x_j - \bar{x})^2 = \frac{1}{n}\sum_{j=0}^{n-1} x_j^2 - \bar{x}^2$.
La variance de Y est : $\sigma^2(Y) = \frac{1}{n}\sum_{j=0}^{n-1}(y_j - \bar{y})^2 = \frac{1}{n}\sum_{j=0}^{n-1} y_j^2 - \bar{y}^2$.
La covariance de (X, Y) est :
$cov(X,Y) = \frac{1}{n}\sum_{j=0}^{n-1}(x_j - \bar{x})(y_j - \bar{y}) = \frac{1}{n}\sum_{j=0}^{n-1} x_j y_j - \bar{x}\bar{y}$.

3.3.4 Moyennes, variances, covariances avec effectifs

Soit une série statistique à deux variables d'ordre n pour les caractères X et Y représentée par les couples $(x_j,\ y_k)$ d'effectifs $n_{j,k}$ ($0 \leq j \leq (p-1)$ et $0 \leq k \leq (q-1)$).
Ici les x_j (resp y_k) sont distincts.
Soit $n = \sum_{j=0}^{p-1} \sum_{k=0}^{q-1} n_{j,k}$
La moyenne de X est :
$\bar{x} = \frac{1}{n} \sum_{j=0}^{p-1} (x_j * \sum_{k=0}^{q-1} n_{j,k})$.
La moyenne de Y est :
$\bar{y} = \frac{1}{n} \sum_{k=0}^{q-1} (y_k * \sum_{j=0}^{p-1} n_{j,k})$.
La variance de X est :
$\sigma^2(X) = \frac{1}{n} \sum_{j=0}^{p-1} ((x_j - \bar{x})^2 * (\sum_{k=0}^{q-1} n_{j,k})) = \frac{1}{n} \sum_{j=0}^{p-1} (x_j^2 * \sum_{k=0}^{q-1} n_{j,k}) - \bar{x}^2$.
La variance de Y est :
$\sigma^2(Y) = \frac{1}{n} \sum_{k=0}^{q-1} ((y_k - \bar{y})^2 * (\sum_{j=0}^{p-1} n_{j,k})) = \frac{1}{n} \sum_{k=0}^{q-1} (y_k^2 * \sum_{j=0}^{p-1} n_{j,k}) - \bar{y}^2$.
La covariance de (X, Y) est :
$cov(X,Y) = \frac{1}{n} \sum_{j=0}^{p-1} \sum_{k=0}^{q-1} (x_j - \bar{x})(y_k - \bar{y}) n_{j,k} = \frac{1}{n} \sum_{j=0}^{n-1} \sum_{k=0}^{q-1} x_j y_k n_{j,k} - \bar{x}\bar{y}$.

3.3.5 Corrélation statistique

Rappel : Lorsqu'on a deux variables aléatoires X et Y, de covariance $cov(X,Y)$ et d'écart-type respectif $\sigma(\mathtt{X})$ et $\sigma(\mathtt{Y})$ on définit leur coefficient de corrélation $\rho(X,Y)$ par :

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}$$

Supposons que l'on a relevé des valeurs (x_j, y_j) de `X` et `Y` au cours de n épreuves indépendantes. On définit, par analogie, un coefficient de corrélation $r(X,Y)$ de l'échantillon par :

$$r(X,Y) = \frac{cov(X,Y)}{s(X)s(Y)}$$

où $s(X)$ (resp $s(Y)$) désigne l'écart-type des valeurs de X (resp Y) pour l'échantillon. On a :

$$r(X,Y) = \frac{\frac{1}{n}\sum_{j=0}^{n-1} x_j y_j - \frac{1}{n}\sum_{j=0}^{n-1} x_j * \frac{1}{n}\sum_{j=0}^{n-1} y_j}{\sqrt{\frac{1}{n}\sum_{j=0}^{n-1}(x_j - \frac{1}{n}\sum_{k=0}^{n-1} x_k)^2} * \sqrt{\frac{1}{n}\sum_{j=0}^{n-1}(y_j - \frac{1}{n}\sum_{k=0}^{n-1} y_k)^2}}$$

Propriétés :
$-1 \leq \rho \leq +1$
si X et Y sont indépendants alors $\rho(X,Y) = 0$ mais la réciproque est fausse.

3.3.6 Les fonctions `covariance` et `correlation` de `Xcas`

On se reportera aussi aux sections 1.11.2, 1.11.3 et 1.11.4.
Avec `Xcas` on tape :

```
covariance([1,2],[11,13,14],[[3,4,5],[12,1,2]])
```

On obtient :

```
-83/243
```

Avec Xcas on tape :

```
correlation([1,2],[11,13,14],[[3,4,5],[12,1,2]])
```

On obtient :

```
-83/160
```

3.3.7 Ajustement linéaire

On se reportera aussi aux sections 1.11.5 et 1.11.9.
Une série statistique à deux variables d'ordre n fournit un nuage de n points. Ajuster linéairement cet ensemble de points consiste à trouver une droite qui approche "le mieux possible le nuage de points". Un ajustement linéaire va permettre de faire des prévisions ou d'estimer des valeurs.

Première droite des moindres carrés est définie pour que la somme des carrés des écarts en ordonnée entre les mesures et les points de cette droite soit minimale.
Soient A_j $(0 \leq j \leq n-1)$ les points de coordonnées $(x_j,\ y_j)$ formant le nuage de points. Soit D une droite d'équation $y = ax+b$ et soient B_j pour $0 \leq j \leq (n-1)$ les points de D de coordonnées $(x_j,\ ax_j+b)$.
On cherche a et b pour que :
$S = \sum_{j=0}^{n-1}(y_j - ax_j - b)^2$ soit minimum.
Pour a fixé le minimum de S est atteint lorsque la droite D passe par le point moyen G de coordonnées $(\bar{x},\ \bar{y})$ donc lorsque $b = b_0 = \bar{y} - a\bar{x}$.
On trouve ensuite que pour $b = b_0$, S est minimum pour :
$a = a_0 = \dfrac{\frac{1}{n}\sum_{j=0}^{n-1} x_j y_j - \bar{x}\bar{y}}{\frac{1}{n}\sum_{j=0}^{n-1} x_j^2 - \bar{x}^2} = \dfrac{cov(X,Y)}{\sigma^2(X)}$.
La première droite des moindres carrés est la droite d'équation $y = a_0 x + b_0$. Elle a donc pour équation $y = \bar{y} + \dfrac{cov(X,Y)}{\sigma^2(X)}(x - \bar{x})$.

Deuxième droite des moindres carrés est définie pour que la somme des carrés des écarts en abscisse entre les mesures et les points de cette droite soit minimale.
On change simplement le rôle de X et de Y. On trouve la droite Δ d'équation :
$x = \bar{x} + \dfrac{cov(X,Y)}{\sigma^2(Y)}(y - \bar{y})$.

Avec Xcas on tape dans une ligne d'entrée de géométrie, pour tracer le nuage de points :

```
scatterplot([[1,11],[1,13],[1,14],[2,11],[2,13],
[2,14]]).
```

ou on tape :

```
scatterplot([1,1,1,2,2],[11,13,14,11,13])
```

Ou dans le tableur, on sélectionne l'argument et on utilise le menu Statistiques du tableur puis 2d et Scatterplot.
On tape dans une ligne d'entrée, pour avoir l'équation de la droite des moindres

carrés :
`linear_regression([1,1,1,2,2],[11,13,14,11,13])`
On obtient :
`-2/3,40/3`

3.4 Les théorèmes des statistiques inférentielles

3.4.1 Problèmes de jugement sur échantillon

L'exploitation des données peut prendre plusieurs formes :

a/ L'inférence statistique ou "théorie de l'estimation" : connaissant un échantillon, on désire émettre une estimation sur la population totale. Dans ce cas, on n'a pas d'idée a priori sur le paramètre à estimer :
on construira **un intervalle de confiance** I_α **au seuil** α.
Cet intervalle I_α dépend de l'échantillon et contient, en général, la valeur du paramètre sauf dans $\alpha\%$ des cas c'est à dire, il y a seulement $\alpha\%$ des échantillons qui ont un I_α qui ne contient pas le paramètre (on dit qu'on a un risque d'erreur égal à α).

b/ Le test d'hypothèses permet de savoir si il y a accord entre théorie et expérience. Dans ce cas on a une idée a priori sur la valeur que doit avoir le paramètre : on construit le test d'hypothèses (deux hypothèses H_0 et H_1 seront en concurrence), puis on prélève un échantillon et on regarde si cet échantillon vérifie le test ce qui permet d'accepter ou de refuser l'hypothèse privilégiée H_0.
Par exemple : on veut contrôler qu'une fabrication correspond bien à ce qui a été décidé, pour cela on fabrique un test d'hypothèses, puis on teste l'hypothèse H_0 sur un échantillon de la production.
c/ Le test d'homogénéite permet de comparer une distribution expérimentale à une distribution théorique.
Remarque :
en a/ et en b/ on a seulement comparer ou estimer des valeurs caractéristiques comme fréquences ou moyennes, en c/ on compare deux distributions.

3.4.2 Théorème de Bienaymé-Tchebychef

Théorème La probabilité pour qu'une variable aléatoire X diffère de sa moyenne (en valeur absolue) d'au moins k fois son écart type, est au plus égale à $1/k^2$, c'est à dire si X a comme moyenne $m = E(X)$ et comme écart type σ on a : $Proba(|X - m| \geq k\sigma) \leq \frac{1}{7}k^2$ **Exemples**

— On lance 100 fois un dé et on considère comme événement :
on obtient 6 ou on n'obtient pas 6.
Par l'expérience on a obtenu n_1 fois le nombre 6 Trouver un majorant de $Proba(|n_1/100 - 1/6| \geq 1/10)$.
La probabilité d'avoir un 6 est : $p = 1/6$ et de ne pas avoir un 6 est 5/6.
Si Y est la variable aléatoire égale à la fréquence de l'événement favorable on a $E(Y) = p = 1/6 \simeq 0.166666666667$ et $\sigma(Y) = sqrt(1/6 * 5/6/100) = sqrt(5)/60 \simeq 0.037267799625$ Théorème de Bienaymé-Tchebychef nous dit que :
$Proba(|n_1/100 - 1/6| \geq k\sigma(Y)) \leq \frac{1}{7}k^2$

On cherche k pour avoir $k\sigma(Y) = ksqrt(5)/60 \leq 1/10$
on prend $k = 2.6832815732$ car $k \leq 6/sqrt(5) \simeq 2.683281573$
donc $Proba(|n_1/100 - 1/6| \geq 1/10) < 1/2.68^2 \simeq 0.139$
Cela veut dire que : $n_1/100$ se trouve dans l'intervalle
$1/6 - 1/10 \simeq 0.0666666666667; 1/6 + 1/10 \simeq 0.266666666667$ avec la probabilité $1 - 0.139 = 0.861$ ou encore que n_1 se trouve dans l'intervalle $6; 26$ avec la probabilité 0.861.

— même exercice mais cette fois on lance le dé 6000 fois.
Par l'expérience on a obtenu n_1 fois le nombre 6 Trouver un majorant de $Proba(|n_1/6000 - 1/6| \geq 1/100)$.
Si Y est la variable aléatoire égale à la fréquence de l'événement favorable on a $E(Y) = p = 1/6 \simeq 0.166666666667$ et $\sigma(Y) = sqrt(1/6 * 5/6/6000) = sqrt(5/60)/60 \simeq 0.00481125224325$ On cherche k pour avoir $k\sigma(Y) = ksqrt(5/60)/60 \leq 1/100$
on prend $k = 2.07846096908$ car $k \leq 6/sqrt(50/6) \simeq 2.07846096908$
donc $Proba(|n_1/6000 - 1/6| \geq 1/100) < 1/2.07846096908^2 \simeq 0.231481481482$
Cela veut dire que : $n_1/6000$ se trouve dans l'intervalle
$1/6 - 1/100 \simeq 0.16566666666667; 1/6 + 1/10 \simeq 0.176666666667$ avec la probabilité $1 - 0.231481481482 = 0.768518518518$ ou encore que n_1 se trouve dans l'intervalle $940; 1060$ avec la probabilité de 0.768518518518.
Remarque En approchant la loi binomiale par la loi normale de moyenne $n * p = 6000 * 1/6 = 1000$ et d'écart type $\sigma = sqrt(np(1-p) = sqrt(6000 * 1/6 * 5/6) \simeq 28.8675134595$ On a $60/28.8675134595 = 2.07846096908$ On cherche dans une table $\psi(t) = Prob(0 < T < t) = \psi(2.07846096908)$ et on trouve 0.481. Donc $Prob(-t < T < t) = 2 * 0.481 = 0.962$ ou dans une table $\Pi(t) = Prob(-\infty < T < t) = \Pi(2.07846096908)$ et on trouve 0.981. Donc $Prob(-t < T < t) = 2 * 0.981 - 1 = 0.962$ donc n_1 se trouve dans l'intervalle $940; 1060$ avec la probabilité de 0.481*2=0.962.

— On extrait 1000 fois avec remise une carte d'un jeu de 32 cartes et on considère comme événement :
on obtient un as ou on n'obtient pas un as.
Par l'expérience on a obtenu n_1 fois un as Trouver un minorant de $Proba(105 < n_1 < 145)$.
La probabilité d'avoir un as est : p=1/8 et de ne pas avoir un as est 7/8. Si Y est la variable aléatoire égale à la fréquence de l'événement favorable on a $E(Y) = p = 1/8 = 0.125$ et $\sigma(Y) = sqrt(1/8 * 7/8/1000) = sqrt(7/10)/80 \simeq 0.0104582503317$. On a $Proba(105 < n_1 < 145) = Proba(|n_1 - 125| < 20) = Proba(|n_1/1000 - 0.125| < 1/50)$ Le théorème de Bienaymé-Tchebychef nous dit que :
$Proba(|n_1/1000 - 1/8| \geq k\sigma(Y)) \leq \frac{1}{7}k^2$
On choisit $k\sigma(Y) = 1/50$ c'est à dire $k = 1/50/0.0104582503317 = 1.91236577493$ donc $1/k^2 = 0.273437500001$ Cela veut dire que $Proba(|n_1/1000 - 0.125| < 1/50) \geq 0.273437500001$ donc
$Proba(105 < n_1 < 145) \leq 1 - 0.273437500001 = 0.7265625$ **Remarque** $Proba(|n_1/1000 - 0.125| > 1/100) \geq 1/0.956182887465^2 = 1.09375000001$ ce qui ne nous apporte rien !

3.4.3 Loi des grands nombres

Notation
On note ici $\bar{X}_n = \dfrac{X_1 + X_2 + .. + X_n}{n}$ pour bien faire ressortir que $\bar{X}_n$ dépend de n, mais quelquefois dans la suite on écrira simplement : $\bar{X} = \dfrac{X_1 + X_2 + .. + X_n}{n}$ pour ne pas alourdir les notations.

Loi faible des grands nombres :
Soient X_1, X_2,.., X_n des variables aléatoires indépendantes de moyenne μ_1, μ_2,.., μ_n et d'écart-type σ_1, σ_2,.., σ_n.
Si quand n tend vers l'infini $\frac{1}{n}\sum_{j=1}^{n}\mu_j$ tend vers μ et,
si quand n tend vers l'infini $\frac{1}{n^2}\sum_{j=1}^{n}\sigma_j^2$ tend vers 0,
alors $\bar{X}_n = \dfrac{X_1 + X_2 + .. + X_n}{n}$ converge en probabilité vers μ quand n tend vers l'infini (i.e. pour tout ϵ et pour tout η il existe n_0 tel que pour tout $n > n_0$ on a $Proba(|\bar{X}_n - \mu| > \epsilon) < \eta$).
Cas des échantillons :
Si X_1,X_2,..,X_n sont un échantillon de X de moyenne μ et décart-type σ, on a $\mu_1 = \mu_2 = .. = \mu_n = \mu$ et $\sigma_1 = \sigma_2 = .. = \sigma_n = \sigma$. Donc $\frac{1}{n}\sum_{j=1}^{n}\mu_j = \mu$ et quand n tend vers l'infini $\frac{1}{n^2}\sum_{j=1}^{n}\sigma_j^2 = \frac{\sigma^2}{n}$ tend vers 0 ce qui montre que la variable aléatoire
$\bar{X}_n = \dfrac{X_1 + X_2 + .. + X_n}{n}$ converge en probabilité vers μ quand n tend vers l'infini.

Loi forte des grands nombres :
Soient X_1, X_2,.., X_n des variables aléatoires indépendantes de moyenne μ_1, μ_2,.., μ_n et d'écart-type σ_1, σ_2 ,.., σ_n.
Si quand n tend vers l'infini $\frac{1}{n}\sum_{j=1}^{n}\mu_j$ tend vers μ et,
si $\sum_{j=1}^{\infty}\frac{\sigma_j^2}{j^2}$ est convergente,
alors $\bar{X}_n = \dfrac{X_1 + X_2 + .. + X_n}{n}$ converge presque sûrement vers μ quand n tend vers l'infini (i.e. dire que Y_n converge presque sûrement vers U c'est dire que l'ensemble des points de divergene est de probabilité nulle i.e.
$Proba(\omega, \lim_{n\to+\infty}(Y_n(\omega) \neq U(\omega)) = 0)$.
Cas des échantillons :
Si X_1,X_2,..,X_n sont un échantillon de X de moyenne μ et décart-type σ, on a $\mu_1 = \mu_2 = .. = \mu_n = \mu$ et $\sigma_1 = \sigma_2 = .. = \sigma_n = \sigma$.
Donc $\frac{1}{n}\sum_{j=1}^{n}\mu_j = \mu$ et $\displaystyle\sum_{j=1}^{\infty}\frac{\sigma^2}{j^2} = \sigma^2\sum_{j=1}^{\infty}\frac{1}{j^2}$ est convergente ce qui montre que :
$\bar{X}_n = \dfrac{X_1 + X_2 + .. + X_n}{n}$ converge presque sûrement vers μ quand n tend vers l'infini.

Le théorème central-limite :
Quand n tend vers l'infini, alors
$\bar{Y}_n = \sqrt{n}\dfrac{(\bar{X}_n - \mu)}{\sigma}$ converge en loi vers U variable aléatoire qui suit la loi normale centrée réduite (dire que Y_n converge en loi vers $U \in \mathcal{N}(0,1)$ veut dire que si F est

la fonction de répartition de la loi normale centrée réduite et si F_n est la fonction de répartition de Y_n alors pour tout $x \in \mathbb{R}$, $F_n(x)$ tend vers $F(x)$ quand n tend vers l'infini).

3.4.4 Moyenne et variance empirique

Soit un échantillon d'effectif n et $(x_1, x_2, ..., x_n)$ les n valeurs observées.
La moyenne empirique est :
$m = \dfrac{x_1 + x_2 + ... + x_n}{n}$
La variance empirique est :
$s^2 = \dfrac{(x_1 - m)^2 + ... + (x_n - m)^2}{n}$

Soit un échantillon de taille n.
Les n valeurs observées $(x_1, x_2, ..., x_n)$ du caractère sont considérées comme étant les valeurs de n variables aléatoires indépendantes $X_1, X_2, .., X_n$ suivant la même loi F d'espérance μ et d'écart-type σ.

L'ensemble des moyennes d'échantillons de taille n est la variable aléatoire $\bar{X} = \dfrac{X_1 + X_2 + .. + X_n}{n}$.
Si le résultat observé est $x_1, x_2..x_n$, alors la valeur observée de $\bar{X}$ est la moyenne empirique m :
$m = \dfrac{x_1 + x_2 + .. + x_n}{n}$.

L'ensemble des variances d'échantillons de taille n est la variable aléatoire $S^2 = \dfrac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + .. + (X_n - \bar{X})^2}{n}$.
Si le résultat observé est $x_1, x_2..x_n$, alors la valeur observée de S^2 est la variance empirique s^2 :
$s^2 = \dfrac{(x_1 - m)^2 + (x_2 - m)^2 + .. + (x_n - m)^2}{n}$.

3.4.5 Étude de $\bar{X}$

Théorèmes
La variable aléatoire $\bar{X} = \dfrac{X_1 + X_2 + .. + X_n}{n}$ converge en probabilité vers μ.
De plus $\bar{X}$ a pour moyenne μ et pour variance $\dfrac{\sigma^2}{n}$.
Quand n tend vers l'infini, $\sqrt{n}\dfrac{(\bar{X} - \mu)}{\sigma}$ converge en loi vers U variable aléatoire qui suit la loi normale centrée réduite.

3.4.6 Estimateur de μ

On appelle estimateur de μ, une variable aléatoire U_n fonction d'un échantillon X_1,X_2,..,X_n qui vérifie :
$\lim_{n->\infty} E(U_n) = \mu$ et $\lim_{n->\infty} \sigma^2(U_n) = 0$
On dit que U_n est un estimateur **sans biais** de μ si c'est un estimateur de μ qui vérifie $E(U_n) = \mu$.

Théorème
$\bar{X} = \frac{(X_1 + X_2 + .. + X_n)}{n}$ est un estimateur sans biais de μ.

3.4.7 Étude de S^2

Théorème
La variable $S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + .. + (X_n - \bar{X})^2}{n}$ converge presque sûrement vers σ^2 quand n tend vers l'infini.
De plus S^2 a pour moyenne :
$E(S^2) = \frac{n-1}{n}\sigma^2$
et pour variance :
$\sigma^2(S^2) = V(S^2) = \frac{n-1}{n^3}((n-1)\mu_4 - (n-3)\sigma^4)$ où $\mu_4 = E((X-\mu)^4)$.

Théorème limite pour S^2 :
Quand n tend vers l'infini, $\sqrt{n}\frac{(S^2 - \frac{n-1}{n}\sigma^2)}{\sqrt{\mu_4 - \sigma^4}}$ converge en loi vers U variable aléatoire qui suit la loi normale centrée réduite (dire que Y_n converge en loi vers $U \in \mathcal{N}(0,1)$ veut dire que si F est la fonction de répartition de la loi normale centrée réduite et si F_n est la fonction de répartition de Y_n alors pour tout $x \in \mathbb{R}$, $F_n(x)$ tend vers $F(x)$ quand n tend vers l'infini).

3.4.8 Estimateur de σ^2

On appelle estimateur de σ^2, une variable aléatoire V_n fonction d'un échantillon X_1,X_2,..,X_n qui vérifie :
$\lim_{n->\infty} E(V_n) = \sigma^2$ et $\lim_{n->\infty} \sigma^2(V_n) = 0$
On dit que V_n est un estimateur **sans biais** de σ^2 si c'est un estimateur de σ^2 qui vérifie $E(V_n) = \sigma^2$.
Théorème
$Z^2 = \frac{(X_1 - \mu)^2 + .. + (X_n - \mu)^2}{n}$ est un estimateur sans biais de σ^2.
$S^2 = \frac{(X_1 - \bar{X})^2 + .. + (X_n - \bar{X})^2}{n}$ est un estimateur de σ^2.
$\frac{n}{n-1}S^2 = \frac{(X_1 - \bar{X})^2 + .. + (X_n - \bar{X})^2}{n-1}$ est un estimateur sans biais de σ^2.
En effet :
Pour S^2 cela découle des théorèmes précédents.
Pour Z^2 on a :
$E(Z^2) = \frac{1}{n}\sum_{j=1}^{n} E((X_j - \mu)^2) = \frac{1}{n}n\sigma^2 = \sigma^2$
et puisque $\sigma^2(X-\mu)^2 = E((X-\mu)^4) - (\sigma^2)^2 = \mu_4 - (\sigma^2)^2$ on a :
$\sigma^2(Z^2) = \frac{1}{n}(\mu_4 - (\sigma^2)^2)$ (où $\mu_4 = E((X-\mu)^4)$ est le moment centré d'ordre 4).
Remarque :
À partir des valeurs $x_1, x_2, .., x_n$ de l'échantillon, on utilisera lorsqu'on connait μ, $\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + .. + (x_n - \mu)^2}{n}$ comme estimateur de σ^2 et si μ est inconnu on utilisera comme estimateur de σ^2 $\frac{(x_1 - m)^2 + (x_2 - m)^2 + .. + (x_n - m)^2}{n-1}$

avec $m = \dfrac{x_1 + x_2 + .. + x_n}{n}$.

3.4.9 En résumé

Le problème est d'obtenir, au vu de l'échantillon empirique, des renseignements sur la population dont l'échantillon est issu (c'est à dire sur la population parente de moyenne μ et d'écart-type σ), en particulier sur la valeur de sa moyenne μ.
En général σ n'est pas connu, on prend faute de mieux, quand n est grand :
$\sigma = s\sqrt{\frac{n}{n-1}}$ où s^2 est la valeur observée de :
$S^2 = \dfrac{(X_1 - Y)^2 + (X_2 - Y)^2 + .. + (X_n - Y)^2}{n}$ qui a pour moyenne $\dfrac{n-1}{n}\sigma^2$.
Grâce au théorème central-limite, la variable $\bar{X} = \dfrac{X_1 + .. + X_n}{n}$ va nous servir à trouver une valeur de μ car :
$\bar{X}$ a pour moyenne μ et pour variance $\dfrac{\sigma^2}{n} \simeq \dfrac{s^2}{n-1}$ donc la variable aléatoire :
$\sqrt{n}\dfrac{(\bar{X} - \mu)}{\sigma} \simeq \sqrt{n-1}\dfrac{(\bar{X} - \mu)}{s}$ converge en loi vers $U \in \mathcal{N}(0, 1)$.

3.5 Les tests d'hypothèses

Concernant une variable aléatoire X, on souhaite comparer la valeur effective d'un paramètre p à une valeur attendue p_0. Il s'agit de savoir si la valeur observée sur un échantillon est vraisemblable avec $p = p_0$.
Test statistique : procédure conduisant au vu de l'échantillon à rejeter, avec un certain risque d'erreur α une hypothèse que l'on cherche à tester appelée H_0. La procédure de test est fondée sur une opposition d'hypothèses et on note H_1 l'hypothèse alternative : cela veut dire que l'on risque de rejeter à tort l'hypothèse H_0 avec une probabilité égale à α.
Test bilatéral : test pour lequel l'hypothèse H_0 est rejetée, si la statistique utilisée prend une valeur en dehors d'un intervalle.
Test unilatéral à droite : test pour lequel l'hypothèse H_0 est rejetée, si la statistique utilisée prend une valeur supérieure à une valeur.
Test unilatéral à gauche : test pour lequel l'hypothèse H_0 est rejetée, si la statistique utilisée prend une valeur inférieure à une valeur.
Construction d'un test :
- choix du seuil de risque α,
- choix des hypothèses H_0 et H_1,

par exemple on choisira un test unilatéral à droite si on sait à priori que $p \leq p_0$. On aura alors $H_0 : p = p_0$ et $H_1 : p > p_0$,
- choix d'une variable statistique S servant de variable de décision,
- détermination de la région critique au seuil α,
- énoncé de la règle de décision.

Utilisation du test :
- prélèvement d'un échantillon,
- au vu de la valeur observée s de S, rejeter ou accepter H_0.

Remarques
Le seuil de risque α est toujours petit ($\alpha < 0.1$) : si on demande un test à 95% cela veut dire que le seuil de risque est $\alpha = 0.05$.
N'oubliez pas que lorsque l'on rejette l'hypothèse H_0 cela veut dire que l'hypothèse H_0 risque d'être vraie dans moins de $100 * \alpha$ cas pour 100 cas et que lorsque l'on accepte l'hypothèse H_0 cela veut dire que l'hypothèse H_0 risque d'être vraie dans plus de $100 * \alpha$ cas pour 100 cas.

3.5.1 Étude de la fréquence p d'un caractère X

Soit une variable aléatoire X qui suit une loi de Bernouilli de paramètre p (on étudie un caractère, si ce caractère est observé alors $X = 1$ et sinon $X = 0$ et on a $Proba(X = 1) = p$). Soit $\bar{X}$ la moyenne des échantillons de taille n : ici, $\bar{X}$ est égal pour chaque échantillon de taille n à la fréquence observée F du caractère.
Si n est grand ($n \geq 30$), $\bar{X}$ suit approximativement la loi normale $\mathcal{N}(p, \sqrt{\frac{p(1-p)}{n}})$.
Si n est petit, on a $(n * \bar{X})$ suit la loi binomiale $\mathcal{B}(n, p)$.
On choisit le seuil α et selon les cas :
Test d'hypothèses bilatéral : $H_0 : p = p_0$ et $H_1 : p \neq p_0$
Test d'hypothèses unilatéral à droite (à gauche) : $H_0 : p = p_0$ et $H_1 : p > p_0$ (resp $H_0 : p = p_0$ et $H_1 : p < p_0$)
On calcule, sous l'hypothèse H_0, soit au moyen des tables de la loi normale (pour n grand, $np(1-p) > 7$), soit au moyen des tables de la loi binomiale (pour n petit), soit avec `Xcas`, les bornes de l'intervalle d'acceptation au seuil α, de l'hypothèse H_0.

— dans le cas bilatéral, on cherche les réels a_1 et a_2 vérifiant :
$Proba(a_1 < F = \bar{X} < a_2) = 1 - \alpha$
- pour n grand, on cherche dans une table de loi centrée réduite h tel que $Proba(Y < h) = 1 - \alpha/2$, on pose
$F = (Y - p_0)/\sqrt{\frac{p_0(1-p_0)}{n}}$ et on obtient :
$Proba(F < p_0 + h * \sqrt{\frac{p_0(1-p_0)}{n}}) = 1 - \alpha/2$, donc
$a_1 = p_0 - h * \sqrt{\frac{p_0(1-p_0)}{n}}$ et $a_2 = p_0 + h * \sqrt{\frac{p_0(1-p_0)}{n}}$
On peut aussi taper dans `Xcas` si $\alpha = 0.05$:
```
a1:=normal_icdf(p0,sqrt((p0*(1-p0)/n),0.025)
a2:=normal_icdf(p0,sqrt(p0*(1-p0)/n),0.975)
```
- pour n petit, $n * F$ suit la loi binomiale $\mathcal{B}(n, p_0)$, on cherche dans une table de loi binomiale $\mathcal{B}(n, p_0)$ les valeurs $n * p_1$ et $n * p_2$ tels que :
$Proba(n * p_1 < n * F < n * p_2) = 1 - \alpha$
et donc :
$Proba(p_1 < F < p_2) == 1 - \alpha$
On peut aussi taper dans `Xcas` si $\alpha = 0.05$:
```
p1:=1/n*binomial_icdf(n,p0,0.025)
p2:=1/n*binomial_icdf(n,p0,0.975)
```

— dans le cas unilatéral à droite, on cherche le réel a vérifiant :
$Proba(F < a) = 1 - \alpha$:
- pour n grand, on cherche dans une table de loi centrée réduite h tel que

$Proba(Y < h) = 1-\alpha$, on a alors $Proba(F < p_0+h\sqrt{\frac{p_0(1-p_0)}{n}}) = 1-\alpha$
donc $a = p_0 + h\sqrt{\frac{p_0(1-p_0)}{n}}$
On peut aussi taper dans `Xcas` si $\alpha = 0.05$:
`a:=normal_icdf(p0,sqrt(p0*(1-p0)/n),0.975)`
- pour n petit, on cherche $n * p_2$ tel que :
$Proba(n * F < n * p_2) = 1 - \alpha$
on a donc $Proba(F < p_2 = 1 - \alpha$
On peut aussi taper dans `Xcas` si $\alpha = 0.05$:
`p2:=1/n*binomial_icdf(n,p0,0.975)`

— dans le cas unilatéral à gauche, on cherche le réel b vérifiant :
$Proba(F < b) = \alpha$:
- pour n grand, on cherche dans une table de loi centrée réduite h tel que $Proba(Y < h) = \alpha$, on a alors $Proba(F < p_0 + h\sqrt{\frac{p_0(1-p_0)}{n}}) = \alpha$ donc $b = p_0 + h\sqrt{\frac{p_0(1-p_0)}{n}}$
On peut aussi taper dans `Xcas` si $\alpha = 0.05$:
`b:=normal_icdf(p0,sqrt(p0*(1-p0)/n),0.05)`
- pour n petit, on cherche $n * p_1$ tel que :
$Proba(n * F < n * p_1) = \alpha$
on a donc :
$Proba(F < p_1) = \alpha$
On peut aussi taper dans `Xcas` si $\alpha = 0.05$:
`p1:=1/n*binomial_icdf(n,p0,0.05)`

Règle de décision :
Soit la fréquence f d'un échantillon de taille n.
On rejette l'hypothèse H_0 au seuil α :

— dans le cas bilatéral
si $f \notin [a_1; a_2]$, (ou si $f \notin [p_1; p_2]$)
— dans le cas unilatéral à droite
si $f > a$, (ou si $f > p_2$)
— dans le cas unilatéral à gauche
si $f < b$, (ou si $f < p_1$)

sinon on accepte l'hypothèse H_0 au seuil α.
Exemple
On choisit $n = 30$, $p_0 = 0.3$ et $\alpha = 0.05$ et on compare les résultats de la loi normale et de la loi binomiale.

— dans le cas bilatéral
Pour $n = 30$ et $H_0 : p = 0.3$ $H1 : p \neq 0.3$ et $\alpha = 0.05$
on a $p_0 * (1 - p_0) = 0.3 * 0.7 = 0.21$ et on tape :
`normal_icdf(0.3,sqrt(0.21/30),0.975)=0.463982351931`
`normal_icdf(0.3,sqrt(0.21/30),0.025)=0.136017648069`
on pose $I = [0.1360; 0.464]$
ou on tape :
`1/30*binomial_icdf(30,0.3,0.025)=2/15`
`1/30*binomial_icdf(30,0.3,0.975)=7/15`

on pose $I = [0.1333; 0.46666]$
On rejette H_0 au seuil de 5%, si la fréquence f obtenue à partir d'un échantillon de taille $n = 30$ est en dehors de l'intervalle I.

— dans le cas unilatéral à droite
Pour n=30 $H_0 : p = 0.3$ $H1 : p \geq 0.3$ et $\alpha = 0.05$
on a $p_0 * (1 - p_0) = 0.3 * 0.7 = 0.21$ et on tape :
`normal_icdf(0.3,sqrt(0.21/30),0.95)=0.437618327917`
on pose $I =]-\infty; 0.464]$
ou on tape :
`1/30*binomial_icdf(30,0.3,0.95)=13/30`
on pose $I =]-\infty; 0.43334]$
On rejette H_0 au seuil de 5%, si la fréquence f obtenue à partir d'un échantillon de taille $n = 30$ est en dehors de l'intervalle I.

— dans le cas unilatéral à gauche
Pour n=30 $H_0 : p = 0.3$ $H1 : p \leq 0.3$ et $\alpha = 0.05$
on a $p_0 * (1 - p_0) = 0.3 * 0.7 = 0.21$ et on tape :
`normal_icdf(0.3,sqrt(0.21/30),0.05)=0.162381672083`
on pose $I = [0.16238; +\infty[$
ou on tape :
`1/30*binomial_icdf(30,0.3,0.05)=1/6`
on pose $I = [0.166667; +\infty[$
On rejette H_0 au seuil de 5%, si la fréquence f obtenue à partir d'un échantillon de taille $n = 30$ est en dehors de l'intervalle I.

3.5.2 Étude de la valeur moyenne μ d'un caractère X

On va faire des tests d'hypothèses sur μ c'est à dire que dans ce qui suit, on suppose que $\mu = \mu_0$, ie que l'on connait μ.

n est grand ($n > 30$)

Théorèmes :
Si $X \in \mathcal{N}(\mu, \sigma)$ alors $\bar{X} \in \mathcal{N}(\mu, \sigma/\sqrt{n})$.
Si X suit une loi quelconque et si l'échantillon est de grande taille ($n > 30$), $\bar{X}$ suit approximativement une loi $\mathcal{N}(\mu, \sigma/\sqrt{n})$.
- si l'écart-type σ est connu, on connait la loi suivie par $\bar{X}$,
- si l'écart-type σ n'est pas connu, puisque n est grand on va pouvoir estimer σ par $s\sqrt{n/(n-1)}$ où s est l'écart-type d'un échantillon de taille n et on se raméne au cas précedent (σ connu) en prenant $\sigma = s\sqrt{n/(n-1)}$. Ainsi on connait la loi suivie par $\bar{X}$: $\bar{X}$ suit approximativement une loi $N(\mu, s/\sqrt{n-1})$.
Recette quand on connait la loi $\mathcal{N}(\mu, \sigma/\sqrt{n})$ suivie par $\bar{X}$ (σ connu)
On choisit le seuil α et selon les cas :
Test d'hypothèses bilatéral : $H_0 : \mu = \mu_0$ et $H_1 : \mu \neq \mu_0$
Test d'hypothèses unilatéral à droite : $H_0 : \mu = \mu_0$ et $H_1 : \mu > \mu_0$ (resp à gauche : $H_0 : \mu = \mu_0$ et $H_1 : \mu < \mu_0$)
On calcule, au moyen des tables de loi normale (n grand, $n > 30$) les bornes de l'intervalle d'acceptation au seuil α, de l'hypothèse H_0.

— dans le cas bilatéral, on cherche les réels a_1 et a_2 vérifiant :

$Proba(a_1 < \bar{X} < a_2) = 1 - \alpha$:
on cherche dans une table de loi centrée réduite h tel que :
$Proba(Y < h) = 1 - \alpha/2$, on a $Proba(\bar{X} < \mu_0 + h * \sigma/\sqrt{n}) = 1 - \alpha/2$, donc
$a_1 = \mu_0 - h * \sigma/\sqrt{n}$ et $a_2 = \mu_0 + h * \sigma/\sqrt{n}$
Avec `Xcas`, on tape :
`a1:=normal_icdf(`$\mu_0, \frac{\sigma}{\sqrt{n}}, \alpha/2$`)`
`a2:=normal_icdf(`$\mu_0, \frac{\sigma}{\sqrt{n}}, 1 - \alpha/2$`)`

— dans le cas unilatéral à droite, on cherche le réel a vérifiant :
$Proba(\bar{X} < a) = 1 - \alpha$:
on cherche dans une table de loi centrée réduite h tel que :
$Proba(Y < h) = 1 - \alpha$, on a $Proba(\bar{X} < \mu_0 + h * \sigma/\sqrt{n}) = 1 - \alpha$, donc
$a = \mu_0 + h * \sigma/\sqrt{n}$.
Avec `Xcas`, on tape :
`a:=normal_icdf(`$\mu_0, \frac{\sigma}{\sqrt{n}}, 1 - \alpha$`)`

— dans le cas unilatéral à gauche, on cherche le réel b vérifiant :
$Proba(\bar{X} < b) = \alpha$:
on cherche dans une table de loi centrée réduite h tel que :
$Proba(Y < h) = \alpha$, on a alors $Proba(\bar{X} < \mu_0 + h * \sigma/\sqrt{n}) = \alpha$, donc
$b = \mu_0 + h * \sigma/\sqrt{n}$.
Avec `Xcas`, on tape :
`b:=normal_icdf(`$\mu_0, \frac{\sigma}{\sqrt{n}}, \alpha$`)`

Règle de décision :
Soit m la moyenne d'un échantillon de taille n.
On rejette l'hypothèse H_0 au seuil α :

— dans le cas bilatéral
si $m \notin [a_1; a_2]$,

— dans le cas unilatéral à droite
si $m > a$,

— dans le cas unilatéral à gauche
si $m < b$,

sinon on accepte l'hypothèse H_0 au seuil α.

$X \in \mathcal{N}(\mu, \sigma)$ et n est petit, ($n \leq 30$)

On a deux cas selon que l'écart-type σ est connu ou pas :
- si l'écart-type σ est connu
On sait que si $X \in \mathcal{N}(\mu, \sigma)$ alors $\bar{X} \in \mathcal{N}(\mu, \sigma/\sqrt{n})$ on se reportera à la "Recette quand on connait la loi $\mathcal{N}(\mu, \sigma/\sqrt{n})$ suivie par $\bar{X}$" écrite ci-dessus.
- si l'écart-type σ est inconnu
Lorsque n est petit, on ne peut plus approcher σ par $s\sqrt{n/(n-1)}$ où s est l'écart-type d'un échantillon de taille n.
C'est pourquoi, lorsque n est petit et que $X \in \mathcal{N}(\mu, \sigma)$, on utilise la statistique :
$T = \sqrt{n-1}(\frac{\bar{X} - \mu_0}{S})$ où $S^2 = 1/n \sum_{j=1}^{n} (X_j - \bar{X})^2$.
T suit une loi de Student à $n - 1$ degrés de liberté et T ne dépend pas de σ.
Recette quand on ne connait pas la loi suivie par $\bar{X}$
On est dans le cas où σ est inconnu, $X \in \mathcal{N}(\mu, \sigma)$ et n est petit.
On choisit le seuil α et selon les cas :

Test d'hypothèses bilatéral : $H_0 : \mu = \mu_0$ et $H_1 : \mu \neq \mu_0$
Test d'hypothèses unilatéral à droite : $H_0 : \mu = \mu_0$ et $H_1 : \mu > \mu_0$ (resp à gauche : $H_0 : \mu = \mu_0$ et $H_1 : \mu < \mu_0$).
Au moyen des tables de la loi de Student (n petit, $n \leq 30$)

- dans le cas bilatéral, on cherche le nombre réel h vérifiant :
  $Proba(-h < T_{n-1} < +h) = 1 - \alpha$
  Avec `Xcas`, on tape si $\alpha = 0.05$:
  `h:=student_icdf(n-1,0.975)`
- dans le cas unilatéral à droite, on cherche le nombre réel h_1 vérifiant :
  $Proba(T_{n-1} < h_1) = 1 - \alpha$
  Avec `Xcas`, on tape si $\alpha = 0.05$:
  `h1:=student_icdf(n-1,0.95)`
- dans le cas unilatéral à gauche, on cherche le nombre réel h_2 vérifiant :
  $Proba(T_{n-1} < h_2) = \alpha$
  Avec `Xcas`, on tape si $\alpha = 0.05$:
  `h2:=student_icdf(n-1,0.05)`

Règle de décision :
Soit t la valeur prise par T par un échantillon de taille n : $t = \sqrt{n-1}(\frac{m-\mu_0}{s})$
où m est la moyenne de l'échantillon et s son écart-type.
On rejette l'hypothèse H_0 au seuil α :

- dans le cas bilatéral
  si $t \notin [-h; +h]$,
- dans le cas unilatéral à droite
  si $t > h_1$,
- dans le cas unilatéral à gauche
  si $t < h_2$,

sinon on accepte l'hypothèse H_0 au seuil α.

X ne suit pas une loi normale et n est petit

On ne sait pas faire...

3.5.3 Étude de l'écart-type σ de $X \in \mathcal{N}(\mu, \sigma)$

On sait que si X suit une loi normale $\mathcal{N}(\mu, \sigma)$, les statistiques :

$$Z^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)^2 \text{ et}$$

$$S^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})^2$$

sont des estimateurs de σ, de plus Z^2 et $\frac{n}{n-1}S^2$ sont des estimateurs sans biais de σ (cf 3.4.8), car on a $E(Z^2) = E(\frac{n}{n-1}S^2) = \sigma$ et S^2 ne dépend pas de μ.
On sait que :
la statistique $\dfrac{nZ^2}{\sigma^2}$ suit une loi du χ^2 à n degrés de liberté et que
la statistique $\dfrac{nS^2}{\sigma^2}$ suit une loi du χ^2 à $(n-1)$ degrés de liberté.
Lorsque μ est connue, on utilisera la statistique $\dfrac{nZ^2}{\sigma^2}$ comme variable de décision,

et si μ n'est pas connue, on utilisera la statistique $\frac{nS^2}{\sigma^2}$ comme variable de décision.

Recette quand X suit une loi normale $\mathcal{N}(\mu, \sigma)$

On choisit le seuil α et selon les cas :
Test d'hypothèses bilatéral : $H_0 : \sigma = \sigma_0$ et $H_1 : \sigma \neq \sigma_0$,
Test d'hypothèses unilatéral à droite : $H_0 : \sigma = \sigma_0$ et $H_1 : \sigma > \sigma_0$ (resp à gauche : $H_0 : \sigma = \sigma_0$ et $H_1 : \sigma < \sigma_0$).
On calcule au moyen des tables de $\chi^2(n)$ les nombres réels h_1 et h_2 vérifiant :

— dans le cas bilatéral
- si la valeur moyenne μ est connue
$Proba(\chi^2_n < h_1) = 1 - \alpha/2$
$Proba(\chi^2_n < h_2) = \alpha/2$
Avec `Xcas`, on tape si $\alpha = 0.05$:
```
h1:=chisquare_icdf(n,0.975)
h2:=chisquare_icdf(n,0.025)
```
- si la valeur moyenne μ n'est pas connue
$Proba(\chi^2_{n-1} < h_1) = 1 - \alpha/2$
$Proba(\chi^2_{n-1} < h_2) = \alpha/2$
Avec `Xcas`, on tape si $\alpha = 0.05$:
```
h1:=chisquare_icdf(n-1,0.975)
h2:=chisquare_icdf(n-1,0.025)
```

— dans le cas unilatéral à droite
- si la valeur moyenne μ est connue
$Proba(\chi^2_n < h_1) = 1 - \alpha$
Avec `Xcas`, on tape si $\alpha = 0.05$:
```
h1:=chisquare_icdf(n,0.95)
```
- si la valeur moyenne μ n'est pas connue
$Proba(\chi^2_{n-1} < h_1) = 1 - \alpha$
Avec `Xcas`, on tape :
```
h1:=chisquare_icdf(n-1,0.95)
```

— dans le cas unilatéral à gauche
- si la valeur moyenne μ est connue
$Proba(\chi^2_n < h_2) = \alpha$
Avec `Xcas`, on tape si $\alpha = 0.05$:
```
h2:=chisquare_icdf(n,0.05)
```
- si la valeur moyenne μ n'est pas connue
$Proba(\chi^2_{n-1} < h_2) = \alpha$
Avec `Xcas`, on tape si $\alpha = 0.05$:
```
h2:=chisquare_icdf(n-1,0.05)
```

Règle de décision :

Soit u la valeur prise par $\frac{nZ^2}{\sigma^2}$ (ou par $\frac{nS^2}{\sigma^2}$ si μ n'est pas connue) pour un échantillon de taille n :
- si μ est connue, on calcule $u = \frac{\sum_{j=0}^{n}(x_j - \mu)^2}{\sigma_0^2}$ où les x_j sont les valeurs de l'échantillon (car selon $H_0 : \sigma = \sigma_0$).

- si μ n'est pas connue, on calcule $u = \dfrac{n * s^2}{\sigma_0^2}$ où s est l'écart-type de l'échantillon (car selon $H_0 : \sigma = \sigma_0$).
On rejette l'hypothèse $H_0 : \sigma = \sigma_0$ au seuil α :
— dans le cas bilatéral
si $u \notin [h_2; h_1]$,
— dans le cas unilatéral à droite
si $u > h_1$,
— dans le cas unilatéral à gauche
si $u < h_2$,
sinon on accepte l'hypothèse H_0 au seuil α.

3.6 Intervalle de confiance

L'estimation a pour but, à partir d'échantillons, de donner des valeurs numériques aux paramètres de la population dont ces échantillons sont issus.
Il peut s'agir d'estimation ponctuelle ou d'estimation par intervalle.
Un intervalle de confiance I_α au seuil α, pour le paramètre p_0, est un intervalle qui contient p_0 avec une confiance de $1 - \alpha$, cela veut dire que pour un grand nombre n d'échantillons environ $n * \alpha$ des I_α ne contiennent pas p_0 (en effet les intervalles de confiance I_α dépendent de l'échantillon) `Remarques` Le seuil de risque α est toujours petit ($\alpha < 0.1$) : si on vous demande un intervalle de confiance à 95% cela veut dire que le seuil de risque est $\alpha = 0.05$.
N'oubliez pas que l'estimation d'une valeur par un intervalle de confiance comporte un risque, celui de situer la valeur dans un intervalle où elle ne se trouve pas ! ! ! ! (c'est α qui determine le risque d'erreur)
Plus on demande un risque faible et plus l'intervalle de confiance est grand.

3.6.1 Valeur de la fréquence p d'un caractère X

Estimation ponctuelle

Lorsque la taille n de l'échantillon est grande, on prend comme estimation ponctuelle de p la fréquence f observée sur l'échantillon.
Remarque
Cela ne donne aucune information sur la qualité de l'estimation.

Estimation par un intervalle

Cas des échantillons de taille $n > 30$
Soit X une variable aléatoire de Bernouilli de paramètre p (X vaut 0 ou 1 et $Proba(X = 1) = p$).
Soit $\bar{X}$ la variable aléatoire égale à la moyenne des valeurs prises par X pour des échantillons de taille n. On a $\bar{X} = F$ est égal à la fréquence du nombre d'apparitions de la valeur 1 pour chaque échantillon de taille n, .
On sait que $n * F$ suit une loi binomiale $\mathcal{B}(n, p)$, cette loi est proche de la loi normale
$\mathcal{N}(np, \sqrt{np(1-p)})$ car n est grand ($n > 30$).

On peut donc considérer que F suit approximativement la loi $\mathcal{N}(p, \sqrt{\frac{p(1-p)}{n}})$.
Recette
- On choisit α (par exemple $\alpha = 0.05$),
- On cherche à l'aide d'une table de loi normale centrée réduite, h vérifiant :
$Proba(Y < h) = 1 - \alpha/2$ pour $Y \in \mathcal{N}(0, 1)$.
On a donc en posant $Y = \frac{F-p}{\sqrt{\frac{p(1-p)}{n}}}$:
$Proba(p - h\sqrt{\frac{p(1-p)}{n}} < F < p + h\sqrt{\frac{p(1-p)}{n}}) = 1 - \alpha$
- On calcule la valeur f de F pour l'échantillon
On a donc $n(f-p)^2 < h^2 p(1-p)$ c'est à dire $(h^2+n)p^2 - p(h^2+2nf) + nf^2 < 0$ donc p se trouve à l'intérieur des racines de l'équation du second degré :
$(h^2+n)x^2 - x(h^2+2nf) + nf^2 = 0$ que l'on peut résoudre (calcul du discriminant $\Delta = h^4 + (-(4 * h^2)) * n * f^2 + 4 * h^2 * n * f$ etc...)
mais il est plus simple de dire, que l'on peut estimer l'écart-type de $n * F$. On a $\sigma(n * F) = \sqrt{np(1-p)}$ que l'on peut estimer par $\sqrt{nf(1-f)}\sqrt{\frac{n}{n-1}}$.

Donc l'écart-type de $\bar{X} = F$, $\sigma(F) = \sigma(\bar{X}) = \sqrt{\frac{p(1-p)}{n}}$ peut être estimé par $\frac{1}{n}\sqrt{nf(1-f)}\sqrt{\frac{n}{n-1}} = \sqrt{\frac{f(1-f)}{n-1}}$, donc on a :
$Proba(p - h\sqrt{\frac{f(1-f)}{n-1}} \leq f \leq p + h\sqrt{\frac{f(1-f)}{n-1}}) = 1 - \alpha$
ou encore
$Proba(f - h\sqrt{\frac{f(1-f)}{n-1}} \leq p \leq f + h\sqrt{\frac{f(1-f)}{n-1}}) = 1 - \alpha$
Si $a_1 = f - h\sqrt{\frac{f(1-f)}{n-1}}$ et $a_2 = f + h\sqrt{\frac{f(1-f)}{n-1}}$ on a $a_1 \leq p \leq a_2$
Avec `Xcas`, on tape si $\alpha = 0.05$:
```
a1:=normal_icdf(f,sqrt(f*(1-f)/(n-1),0.025)
a2:=normal_icdf(f,sqrt(f*(1-f)/)n-1,0.975)
```
Résultat $I_\alpha = [a_1\ ;\ a_2]$ est un intervalle de confiance de p au seuil α.
Cas des échantillons de taille $n \leq 30$
Soit X une variable aléatoire de Bernouilli de paramètre p (X vaut 0 ou 1 et $Proba(X = 1) = p$).
Soit la variable aléatoire $F = \bar{X}$.
On sait que nF suit une loi binomiale $\mathcal{B}(n, p)$. On utilisera donc une table de la loi binomiale.
Recette
- On choisit α (par exemple $\alpha = 0.05$)
- On calcule la valeur f de F pour l'échantillon
- On approche p par f, ainsi $n * F = n * \bar{X} \in \mathcal{B}(n, f)$, on cherche $n * p_1$ et $n * p_2$ à l'aide d'une table de loi binomiale pour avoir :
$Proba(n * F < n * p_1) = 1 - \alpha/2$ et $Proba(n * F < n * p_2) = \alpha/2$
Avec `Xcas`, on tape si $\alpha = 0.05$:
```
p1:=1/n*binomial_icdf(,n,f,0.025)
p2:=1/n*binomial_icdf(n,f,0.975)
```
On a donc :
$Proba(p_2 < f < p_1) = 1 - \alpha$.
Résultat
$I_\alpha = [p_2\ ;\ p_1]$ est un intervalle de confiance de p au seuil α.

3.6.2 Valeur moyenne μ d'un caractère X

Estimation ponctuelle

Lorsque la taille n de l'échantillon est grande, on prend comme estimation ponctuelle de μ la moyenne m observée sur l'échantillon.
Remarque
Cela ne donne aucune information sur la qualité de l'estimation.

Estimation par un intervalle

Cas des échantillons de taille $n > 30$
Si n est grand ($n > 30$), on connait la loi suivie par $\bar{X}$: $\bar{X}$ suit approximativement une loi $\mathcal{N}(\mu, \sigma/\sqrt{n})$ (ou si σ n'est pas connu $\bar{X}$ suit approximativement une loi $\mathcal{N}(\mu, s/\sqrt{n-1})$).
Recette lorsque la loi $\mathcal{N}(\mu, \sigma/\sqrt{n})$ suivie par $\bar{X}$ est connue
- On choisit α (par exemple $\alpha = 0.05$).
- On calcule la valeur m de $\bar{X}$ pour l'échantillon (ie sa moyenne) et si σ n'est pas connu, l'écart-type s de l'échantillon.
- On cherche h, dans une table de loi normale centrée réduite, pour avoir :
$Proba(Y < h) = 1 - \alpha/2$ pour $Y \in \mathcal{N}(0, 1)$ on a alors :
$Proba(\mu - h * \sigma/\sqrt{n} < m < \mu + h * \sigma/\sqrt{n}) = 1 - \alpha$
on a donc :
$Proba(m - h * \sigma/\sqrt{n} < \mu < m + h * \sigma/\sqrt{n}) = 1 - \alpha$
ou si σ n'est pas connu :
$Proba(\mu - h * s/\sqrt{n-1} < \bar{X} < \mu + h * s/\sqrt{n-1}) = 1 - \alpha$
on a donc $Proba(m - h * s/\sqrt{n-1} < \mu < m + h * \sigma/\sqrt{n-1}) = 1 - \alpha$.
Si σ est connu on pose :
$a_1 = m - h * \sigma/\sqrt{n}$ et $a_2 = m + h * \sigma/\sqrt{n}$
ou si σ n'est pas connu on pose :
$a_1 = m - h * s/\sqrt{n-1}$ et $a_2 = m + h * s/\sqrt{n-1}$
on a $a_1 \leq \mu \leq a_2$
Avec `Xcas`, si σ est connu, on tape si $\alpha = 0.05$:
```
a1:=normal_icdf(m,σ/sqrt(n),0.025)
a2:=normal_icdf(m,σ/sqrt(n),0.975)
```
ou si σ n'est pas connu, on tape si $\alpha = 0.05$:
```
a1:=normal_icdf(m,s/sqrt(n-1),0.025)
a2:=normal_icdf(m,s/sqrt(n-1),0.975)
```
Résultat
$I_\alpha = [a_1 \ ; \ a_2]$ est un intervalle de confiance de μ au seuil α.

Cas des petits échantillons issus d'une loi normale
Si σ est connu, la loi $\mathcal{N}(\mu, \sigma/\sqrt{n})$ suivie par $\bar{X}$ est connue et on se reportera donc à la recette du paragraphe précédent.
Si σ n'est pas connu, on note $S^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})^2$ alors
$T = (\frac{\bar{X}-\mu}{S})\sqrt{n-1}$ suit une loi de Student à $(n-1)$ degrés de liberté.
Recette lorsque n est petit et $X \in \mathcal{N}(\mu, \sigma)$
- On choisit α (par exemple $\alpha = 0.05$).
- On calcule la valeur m de $\bar{X}$ pour l'échantillon (m est la moyenne de l'échan-

tillon) et l'écart-type s de l'échantillon (s^2 est la valeur de S^2 pour l'échantillon).
- On cherche h, dans une table de Student pour $(n-1)$ degrés de liberté, pour avoir :
$Proba(-h < T_{n-1} < h) = Proba(-h < (\frac{\bar{X}-\mu}{S})\sqrt{n-1} < h) = 1-\alpha$
Avec `Xcas`, on tape si $\alpha = 0.05$:
`h:=student_icdf(n-1,0.975)`
puisque m est la valeur de $\bar{X}$ et s la valeur de S pour l'échantillon on a :
$Proba(m - hs/\sqrt{n-1} < \mu < m + hs/\sqrt{n-1}) = 1-\alpha.$
Résultat
$I_\alpha = [m - hs/\sqrt{n-1}; m + hs/\sqrt{n-1}]$ est un intervalle de confiance de μ au seuil α.
Exemple
Pour obtenir un intervalle de confiance de μ au risque $\alpha = 0.05$ et $n-1 = 4$ on tape :
`h:=student_icdf(4,1-0.05/2)`
on obtient :
`h=2.7764451052`$\simeq$ `2.776` donc :
$m - hs/\sqrt{4} < \mu < m + hs/\sqrt{4}$.
On prend un échantillon d'effectif $n = 5$ ($4 = n-1$), pour lequel on trouve :
$m = 0.484342422505$ et $s = 0.112665383246$
On tape :
```
m:=0.484342422505
s:=0.112665383246
m+h*s/sqrt(4)
```
On obtient :
`0.64072197445`
On tape :
```
m-hs/sqrt(4)
0.32796287056.
```
donc un intervalle de confiance de μ au risque 0.05 est :
$[0.32796287056; 0.64072197445]$

3.6.3 Valeur de l'écart-type σ de $X \in \mathcal{N}(\mu, \sigma)$

Estimation ponctuelle

Lorsque la taille n de l'échantillon est grande, on prend comme estimation ponctuelle de σ, $s\sqrt{\frac{n}{n-1}}$, où s est l'écart-type de l'échantillon.
bf Remarque
Cela ne donne aucune information sur la qualité de l'estimation.

Estimation par un intervalle

Cas où μ est connue

On pose $Z^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)^2$. Alors $n\dfrac{Z^2}{\sigma^2}$ suit une loi du χ^2 à n degrés de liberté.
Recette lorsque μ est connue et $X \in \mathcal{N}(\mu, \sigma)$
- On choisit α (par exemple $\alpha = 0.05$).

- On calcule la valeur $z^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu)^2$ de Z^2 pour les valeurs x_j de l'échantillon.
- On cherche t_1 et t_2, dans une table du χ^2 pour n degrés de liberté, pour avoir :
$Proba(\chi_n^2 < t_1) = Proba(n\frac{Z^2}{\sigma^2} < t_1) = 1 - \alpha/2$ et
$Proba(\chi_n^2 < t_2) = Proba(n\frac{Z^2}{\sigma^2} < t_2) = \alpha/2$
Avec `Xcas`, on tape si $\alpha = 0.05$:

```
t1=chisquare_icdf(n,0.975)
t2=chisquare_icdf(n,0.025)
```

on a donc $Proba(t_2 < \frac{nZ^2}{\sigma^2} < t_1) = 1 - \alpha$.
et puisque z^2 est la valeur de Z^2 pour l'échantillon on a :
$Proba(nz^2/t_1 < \sigma^2 < nz^2/t_2) = 1 - \alpha$.
Résultat
$I_\alpha = [\sqrt{nz^2/t_1}\ ;\ \sqrt{nz^2/t_2}]$ est un intervalle de confiance de σ au seuil α.

Cas où μ n'est pas connue
On pose $S^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})^2$.
Alors, $n\dfrac{S^2}{\sigma^2}$ suit une loi du χ^2 à $n-1$ degrés de liberté.
Recette si μ n'est pas connue et $X \in \mathcal{N}(\mu, \sigma)$
- On choisit α (par exemple $\alpha = 0.05$).
- On calcule la valeur m de $\bar{X}$ pour l'échantillon (m est la moyenne de l'échantillon) et l'écart-type s de l'échantillon (s^2 est la valeur de S^2 pour l'échantillon.
- On cherche t_1 et t_2, dans une table du χ^2 pour $(n-1)$ degrés de liberté, pour avoir :
$Proba(\chi_{n-1}^2 < t_1) = Proba(n\frac{S^2}{\sigma^2} < t_1) = 1 - \alpha/2$ et
$Proba(\chi_{n-1}^2 < t_2) = Proba(n\frac{S^2}{\sigma^2} < t_2) = \alpha/2$
Avec `Xcas`, on tape si $\alpha = 0.05$:

```
t1=chisquare_icdf(n-1,0.975)
t2=chisquare_icdf(n-1,0.025)
```

on a donc $Proba(t_2 < n\frac{S^2}{\sigma^2} < t_1) = 1 - \alpha$ et puisque s^2 est la valeur de S^2 pour l'échantillon on a :
$Proba(ns^2/t_1 < \sigma^2 < ns^2/t_2) = 1 - \alpha$.
Résultat
$I_\alpha = [s\sqrt{n/t_1}\ ;\ s\sqrt{n/t_2}]$ est un intervalle de confiance de σ au seuil α.

3.7 Un exemple

On a effectué 10 pesées indépendantes sur une balance d'une même masse μ et on a obtenu :
10.008,10.012,9.990,9.998,9.995,10.001,9.996,9.989,10.000,10.015
Avec `Xcas` on a facilement la moyenne m, l'écart-type s et la variance de l'échantillon.
On tape :

```
L:=[10.008,10.012,9.990,9.998,9.995,10.001,9.996,9.989,
10.000,10.015]
m=mean(L)=10.0004
s=stddev(L)=0.00835703296719
```

`variance(L)=6.98400000147e-05`
On a donc :
$s^2 \simeq 0,00007$

3.7.1 $\sigma = 0.01$ et μ est inconnu

On suppose que μ est inconnue mais que la balance est telle que l'erreur de mesure a un écart-type σ de 0.01.
On cherche à déterminer μ au vu de l'échantillon.

H_0 : $\mu = 10$ et H_1 : $\mu > 10$ au seuil de 5%

On veut tester les hypothèses $H_0 : \mu = 10$ et $H_1 : \mu > 10$
Règle :
On calcule la moyenne m de l'échantillon : on a trouvé $m = 10.004$.
On détermine a pour avoir $Proba(\bar{X} < a) = 0.95$.
Au seuil de 5%, on rejette l'hypothèse unilatérale à droite H_0 si $m > a$ sinon on accepte $H_0 : \mu = 10$.
Si on suppose que le résultat de la mesure est une variable aléatoire X qui suit une loi normale $\mathcal{N}(\mu, 0.01)$, alors $\bar{X}$ suit une loi normale $\mathcal{N}(\mu, 0.01/\sqrt{10})$.
Donc avec l'hypothèse $H_0 : \mu = 10$ on a
$\bar{X} \in \mathcal{N}(10, 0.00316)$ et $Y = \frac{\bar{X}-10}{0.00316} \in \mathcal{N}(0, 1)$
Avec une table de loi normale centrée réduite on cherche h pour que :
$Proba(Y < h) = 0.95$ lorsque $Y \in \mathcal{N}(0, 1)$ et on trouve $h = 1.64$.
On a donc $Proba((\bar{X} - 10)/0.00316 < 1.64) = 0.95$.
On calcule $(m - 10)/0.00316 = 0.126582278481$ et $0.126582278481 < 1.64$ donc on accepte l'hypothèse H_0 $\mu = 10$ au seuil de 5%.
Avec `Xcas` on tape :
`a:=normal_icdf(10,0.01/sqrt(10),0.95)`
On obtient :
`a=10.0051824`
Puisque $m = 10.0004 < a$ on accepte l'hypothèse $H_0 : \mu = 10$.

H_0 : $\mu = 10$ et H_1 : $\mu \neq 10$ au seuil de 5%

On veut tester les hypothèses $H_0 : \mu = 10$ et $H_1 : \mu \neq 10$.
Règle :
On calcule la moyenne m de l'échantillon : on a trouvé $m = 10.004$.
On détermine a pour avoir $Proba(a_1 < \bar{X} < a_2) = 0.95$.
Au seuil de 5%, si $a_1 < m < a_2$, on accepte l'hypothèse bilatérale $H_0 : \mu = 10$ et sinon on la rejette.
Avec une table de loi normale centrée réduite on cherche h pour que :
$Proba(Y < h) = 0.975$ lorsque $Y \in \mathcal{N}(0, 1)$ et on trouve $h = 1.96$.
On a aussi $Proba(Y < -h) = 0.025$ et donc $Proba(-h < Y < h) = 0.95$.
Si on suppose que le résultat de la mesure est une variable aléatoire X qui suit une loi normale $\mathcal{N}(\mu, 0.01)$, alors $\bar{X}$ suit une loi normale $\mathcal{N}(\mu, 0.01/\sqrt{10})$.
On a donc $Proba(|\bar{X} - 10|/0.00316 < h) = 0.95$ soit
$Proba(|\bar{X} - 10| < 1.96 * 0.00316) = 0.95$.
Puisque $1.96 * 0.00316 = 0.0061936$ et que $|m - 10| = 0.0004 < 0.0061936$ on

accepte l'hypothèse H_0 au seuil de 5%.
Avec `Xcas` on tape :
$\mathtt{a_1} := \mathtt{normal\_icdf}(\mathtt{10}, \mathtt{0.01}/\sqrt{\mathtt{10}}, \mathtt{0.025})$
$\mathtt{a_2} := \mathtt{normal\_icdf}(\mathtt{10}, \mathtt{0.01}/\sqrt{\mathtt{10}}, \mathtt{0.975})$
On obtient :
$\mathtt{a_1} = \mathtt{9.99380204968}$
$\mathtt{a_2} = \mathtt{10.0061979503}$
Puisque $a_1 < m = 10.0004 < a_2$ on accepte l'hypothèse $H_0 : \mu = 10$ au seuil de vraisemblance de 5%.

Intervalle de confiance de μ au seuil de 5%

On veut avoir une estimation de μ au seuil de 5%.
On a trouvé précédement que $\bar{X} \in \mathcal{N}(10, 0.00316)$:
$Proba(|\bar{X} - \mu| < 1.96 * 0.00316) = 0.95$.
Pour l'échantillon considéré la valeur de $\bar{X}$ est égale à m d'où,
$Proba(|m - \mu| < 1.96 * 0.00316) = 0.95$
Un intervalle de confiance de μ au seuil de 5% est donc :
$|\mu - 10.0004| < 0.0062$ c'est à dire $[9.9942; 10.0066]$ est un intervalle de confiance de μ au seuil de 5%.
Avec `Xcas` on tape :

```
a1:=normal_icdf(10,0.01/sqrt(10),0.025)
a2:=normal_icdf(10,0.01/sqrt(10),0.975)
```

On obtient :

```
a1=9.99380204968
a2=10.0061979503
```

Donc $[a_1; a_2]$ est un intervalle de confiance de μ au seuil de 5%.

3.7.2 $\mu = 10$ et σ est inconnu

Mainenant, on ne connait pas la précision de la balance mais on a une masse $\mu = 10$ et on voudrait déterminer la précision de la balance au vue de l'échantillon des 10 pesées sauvées dans `L` :

```
L:=[10.008,10.012,9.990,9.998,9.995,10.001,9.996,9.989,
10.000,10.015]
```

On sait que $\mu = 10$.
On pose $Z^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)^2$.
On calcule la valeur z^2 de Z^2, par exemple, avec `Xcas` on tape :
`L10:=makelist(10,0,9)` (`L10` est une liste de longueur 10 dont tous les éléments sont égaux à 10).
`Lc:=L-L10`
`z2:=mean(Lc^2)` on obtient `z2=0.00007`
donc $z^2 \simeq 0.00007$

H_0 : $\sigma = 0.005$ et H_1 : $\sigma > 0.005$ au seuil de 5%

On veut tester les hypothèses $H_0 : \sigma = 0.005$ et $H_1 : \sigma > 0.005$.
$10 * Z^2/\sigma^2$ suit une loi du χ^2 ayant 10 degrés de liberté.
Règle :

On accepte au seuil de 5%, l'hypothèse unilatérale à droite $\sigma = 0.005$, si $z^2 < a$ lorsque a vérifie :
$Proba(10 * Z^2/0.005^2 < 10 * a/0.005^2) = 0.95$.
D'après les tables du χ^2 on trouve :
$Proba(\chi^2_{10} > 18.307) = 0.005$ donc
$a = 18.307 * 0.005^2/10 = 0.0000457$.
Avec `Xcas` on tape :
`h:=chisquare_icdf(10,0.95)`
On obtient :
`h:=18.3070380533`
donc $h \simeq 18.307$
`a:=h*0.005^2/10`
donc $a \simeq 0.0000457$
Puisque $z^2 = 0.00007 > a = 0.0000457$, on ne peut pas accepter l'hypothèse $H_0 : \sigma = 0.005$ au seuil de 5%.

H_0 : $\sigma = 0.005$ et H_1 : $\sigma \neq 0.005$ au seuil de 5%

On veut tester les hypothèses $H_0 : \sigma = 0.005$ et $H_1 : \sigma \neq 0.005$.
$10 * Z^2/\sigma^2$ suit une loi du χ^2 ayant 10 degrés de liberté.
Règle :
On accepte à un niveau de 5%, l'hypothèse bilatérale $\sigma = 0.005$, si $b < Z^2 < a$ lorsque a et b vérifient :
$Proba(10 * b/0.005^2 < 10 * Z^2/0.005^2 < 10 * a/0.005^2) = 0.95$.
D'après les tables on trouve :
$Proba(\chi^2_{10} < 3.25) = 0.025$ et
$Proba(\chi^2_{10} > 20.5) = 0.025$
Donc $a = 20.5 * 0.005^2/10 = 0.00005125$ et $b = 3.25 * 0.005^2/10 = 8.125e-06$.
Avec `Xcas` on tape :
`h1:=chisquare_icdf(10,0.025)`
On obtient :
`h1:=3.24697278024`
donc $h_1 \simeq 3.25$
On tape :
`h2:=chisquare_icdf(10,0.975)`
On obtient :
`h2:=20.4831773508`
donc $h_2 \simeq 20.5$
On tape :
`b:=h1*0.005^2/10`
On obtient :
`8.125e-06`
On tape :
`a:=h2*0.005^2/10` On obtient :
`5.125e-05`
Puisque $z^2 = 0.00007 > a = 0.00005125$, on ne peut donc pas accepter l'hypothèse H_0 $\sigma = 0.005$ au seuil de 5%.

Intervalle de confiance de σ au seuil de 5%

On veut avoir une estimation de σ au seuil de 5%.
On sait que $10 * Z^2/\sigma^2$ suit une loi du χ^2 ayant 10 degrés de liberté.
On a vu précédemment (en 3.7.2) que $h_1 = 3.25$ et $h_2 = 20.5$ et donc que :
$Proba(3.25 < 10 * Z^2/\sigma^2 < 20.5) = 0.95$ donc,
$Proba(10 * Z^2/20.5 < \sigma^2 < 10 * Z^2/3.25) = 0.95$
On a $z^2 = 0.00007$=`z2`, donc $10 * z^2 = 0.0007$.
On a alors :
$0.0007/20.5 = 3.41463414634e{-}05 < \sigma^2 < 0.0007/3.25 = 0.000215384615385$
donc $[0.000034; 0.000216]$ est un intervalle de confiance de σ^2 au seuil de 5%,
donc $[0.0058; 0.0147]$ est un intervalle de confiance de σ au seuil de 5%.
Avec `Xcas` on tape :

```
h1:=3.25
h2:=20.5
a1:=sqrt(10*z2/h2)
```

On obtient :

```
0.0058
```

On tape :

```
a2:=sqrt(10*z2/h1)
```

On obtient :

```
0.0147
```

c'est à dire $[a1\ ;\ a2]$ est un intervalle de confiance de σ au seuil de 5%.

3.7.3 $\mu = 10$ et σ sont inconnus

Mainenant, on ne connait ni le poids μ de la masse, ni la précision σ de la balance et on voudrait determiner μ et σ au vue de l'échantillon.
La valeur de $\bar{X}$ pour l'échantillon est $m = 10.0004$ et la valeur de S pour l'échantillon est $s = 0.00835703296719$.
On peut estimer grossièrement μ par 10.0004, mais n est trop petit pour que cela soit fiable.
Lorsqu'on connait σ, on peut ici, utiliser $\bar{X}$, pour étudier μ car on sait que $\bar{X}$ suit une loi normale $\mathcal{N}(\mu, \sigma/\sqrt{n})$ car on a supposé que X suit une loi normale $\mathcal{N}(\mu, \sigma)$: on va donc essayer d'avoir des renseignements sur σ.

Intervalle de confiance de σ au seuil de 5%

On veut avoir une estimation de σ au seuil de 5%.
Un estimateur sans biais de σ^2 est $nS^2/(n-1)$ mais on ne peut pas estimer σ par $\sqrt{n * s^2/(n-1)} =$ `stdDev(L)` $= 0.00880908621914$ car n est trop petit.
Cherchons un intervalle de confiance pour σ au seuil de 5%.
On sait que la variable statistique $nS^2/\sigma^2 = 10S^2/\sigma^2$ suit une loi du χ^2 ayant 9 degrés de liberté ($9 = (n-1)$, car l'échantillon est de taille $n = 10$ et on enlève 1 ,car on utilise la moyenne de l'échantillon pour calculer S^2).
Cette variable ne dépend pas de μ.
D'après les tables du χ^2 on trouve :
$Proba(\chi^2_9 < 2.70) = 0.025$ et
$Proba(\chi^2_9 > 19.02) = 0.025$

Avec `Xcas` on tape :
`a1:=chisquare_icdf(9,0.025)`
On obtient :
`2.70038949998`
donc $a_1 \simeq 2.70$
`a2:=chisquare_icdf(9,0.975)`
On obtient :
`19.0227677986`
donc $a_2 \simeq 19.02$
Donc $Proba(2.70 < 10S^2/\sigma^2 < 19.02) = 0.95$
Pour l'échantillon $10S^2 = 10s^2 = 6.98400000147e-04$ donc
$(6.98400000147e-04)/19.02 < \sigma^2 < (6.98400000147e-04)/2.70$
$3.67192429099e-05 < \sigma^2 < 0.000258666666721$
On a :
$\sqrt{3.67192429099e-05} = 0.00605964049345$ et
$\sqrt{0.000258666666721} = 0.0160831174441.$
Donc $[0.0060\ ;\ 0.0161]$ est un intervalle de confiance pour σ au seuil de 5%.

Tests d'hypothèses pour μ

On va faire différents tests d'hypothèses pour μ.
Comme l'intervalle de confiance pour σ au seuil de 5% ne donne pas σ avec une grande précision on va utiliser la loi de Student pour avoir des renseignements sur μ.
La variable statistique $T = \sqrt{n-1}(\frac{\bar{X}-\mu}{S})$ suit une loi de Student à $(n-1)$ degrés de liberté. Cette variable ne dépend pas de σ.

- On teste $H_0 : \mu = 10$ et $H_1 : \mu > 10$ au seuil de 5%

On veut tester les hypothèses, $H_0 : \mu = 10$ et $H_1 : \mu > 10$ au seuil de 5%.
Règle :
Si la valeur t de T pour l'échantillon est telle que $t < a$ pour a défini par :
$Proba(T < a) = 0.95$
on accepte l'hypothèse unilatérale à droite H_0 ($\mu = 10$) au seuil de 5%.
On lit dans la table de Student que :
$Proba(T_9 < 1.833) = 0.95.$
Avec `Xcas` on tape :
`a:=student_icdf(9,0.95)`
On obtient :
`1.83311293265`
donc $a \simeq 1.833$
On calcule $t = \sqrt{n-1}(\frac{m-\mu}{s}) = \frac{\sqrt{9}(10.0004-10)}{0.00835703296719} = 0.143591631708$
Puisque $0.143 < 1.833$ on accepte l'hypothèse unilatérale à droite $H_0 : \mu = 10$ au seuil de 5%.

- On teste $H_0 : \mu = 10$ et $H_1 : \mu \neq 10$ au seuil de 5%

Règle :
On lit dans la table de Student que :
$Proba(|T_9| < 2.262) = 0.975.$

Avec `Xcas` on tape :
`a:=student_icdf(9,0.975)`
On obtient :
`a:=2.2621571628`
Donc $a \simeq 2.262$
On vérifie que si `b:=student_icdf(9,0.025)=-2.2621571628`
on a `b=-a`.
Donc $Proba(|T_9| < 2.262) = 0.95$.
Puisque $t = 0.143 < 2.262$ on accepte l'hypothèse bilatérale $H_0 : \mu = 10$ au seuil de 5%.

Intervalle de confiance de μ au seuil de 5%

On veut avoir une estimation de μ au seuil de 5%.
On lit dans la table de Student que :
$Proba(|T_9| < 2.262) = 0.975$.
Avec `Xcas` on tape :
`a:=student_icdf(9,0.975)`
On obtient :
`a:=2.2621571628`
Donc $a \simeq 2.262$
On a donc :
$|t| = \sqrt{n-1}(\frac{|m-\mu|}{s}) = \frac{\sqrt{9}|10.0004-\mu|}{0.00835703296719} < 2.262 = a$
donc
$9.99409879714 = m - as/\sqrt{9} < \mu < m + as/\sqrt{9} = 10.0067012029$
Donc $[9.994; 10.0067]$ est un intervalle de confiance de μ au seuil de 5%.
Remarque $\bar{X}$ suit une loi normale $\mathcal{N}(\mu, \sigma/\sqrt{n})$, si on estime σ par la moyenne des bornes de l'intervalle de confiance trouvé en 3.7.3 on obtient :
$(0.0060 + 0.0161)/2 = 0.01105$ on calcule :
$10.0004 - 1.96 * 0.01105 = 9.978742$
$10.0004 + 1.96 * 0.01105 = 10.022058$
Donc $Proba(|\bar{X} - \mu| < 1.96 * \sigma) = 0.95$ se traduit par :
$9.978742 < \mu < 10.022058$ au seuil de 5%
ce qui donne une moins bonne éstimation qu'avec l'utilisation de la loi de Student.

3.8 Les tests d'homogénéité

Face à deux séries d'observations c'est à dire à deux échantillons, le problème est de savoir si les différences observées sont dues aux fluctuations de l'échantillonnage ou au fait que les échantillons ne proviennent pas de la même population.

3.8.1 Comparaison de deux fréquences observées

Soient f_1 et f_2 les fréquences observées d'un caractère dont la fréquence théorique est p. Cette observation est faite à partir de deux échantillons de taille respéctive n_1 et n_2.
On veut savoir si les fréquences f_1 et f_2 sont significativement différentes ce qui

voudrait dire que les deux échantillons proviennent de deux populations différentes de paramètre p_1 et p_2 ou si au contraire les deux échantillons proviennent d'une même population de paramètre $p = p_1 = p_2$.
On veut donc tester l'hypothèse $H_0 : p_1 = p_2 = p$ contre $H_1 : p_1 \neq p_2$ au seuil α.
Soit F_1 (resp F_2) la variable aléatoire égale à la fréquence du caractère pour des échantillons de taille n_1 (resp n_2).
On a sous l'hypothèse H_0 :
F_1 a pour moyenne p et comme écart-type $\sqrt{\frac{p(1-p)}{n_1}}$
F_2 a pour moyenne p et comme écart-type $\sqrt{\frac{p(1-p)}{n_2}}$
Si n_1 et n_2 sont très grands on a vu que :
F_1 suit approximativement une loi $\mathcal{N}(p, \sqrt{\frac{p(1-p)}{n_1}})$ et
F_2 suit approximativement une loi $\mathcal{N}(p, \sqrt{\frac{p(1-p)}{n_2}})$
Donc
$F_1 - F_2$ suit approximativement une loi $\in \mathcal{N}(0, \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}})$
On va estimer p grâce à la reunion des deux échantillons :

$$p \simeq f = \frac{n_1 * f_1 + n_2 * f_2}{n_1 + n_2}$$

alors
F_1 a pour moyenne p et comme écart-type $\sqrt{\frac{f(1-f)}{n_1}}$
F_2 a pour moyenne p et comme écart-type $\sqrt{\frac{f(1-f)}{n_2}}$
On pose $s_{12} = \sqrt{\frac{f(1-f)}{n_1} + \frac{f(1-f)}{n_2}} = \sqrt{\frac{f(1-f)(n_1+n_2)}{n_1 n_2}}$ donc
$F = F_1 - F_2 \in \mathcal{N}(0, s_{12})$

Recette
On choisit le seuil α.
Avec une table de loi normale centrée réduite, on cherche, pour $U \in \mathcal{N}(0,1)$, h tel que :
$Proba(U \leq h) = 1 - \alpha/2$.
on a alors :
$Proba(\frac{|F_1 - F_2|}{s_{12}} < h) = 1 - \alpha$.
Avec `Xcas` on tape si $\alpha = 0.05$ et si $s12 = s_{12}$:
`a:=normal_icdf(0,s12,1-0.05/2)`
On a alors :
$Proba(|F_1 - F_2| < a) = 1 - \alpha$ avec $a = s_{12} * h$.
On calcule selon les cas :
$\frac{|f_1 - f_2|}{s_{12}}$ que l'on compare à h ou
$|f_1 - f_2|$ que l'on compare à a.
Si $\frac{|f_1 - f_2|}{s_{12}} < h$ ou $|f_1 - f_2| < a$ on admet que les deux échantillons ne sont pas significativement différents au seuil α, sinon on dira que les deux échantillons ne proviennent pas de la même population (voir aussi l'utilisation de la loi du χ^2 en 3.10.2).
Exercice (le même qu'en section 3.10.2)

Pour tester l'efficacité d'un vaccin antigrippal on soumet 300 personnes à une expérience :
- sur 100 personnes non vaccinées, 32 sont atteintes par la grippe,
- sur 200 personnes vaccinées, 50 sont atteintes par la grippe,
Ce résultat permet-il d'apprécier l'efficacité du vaccin ?
On a le tableau suivant :

	grippé	non grippé	taille
vacciné	32	68	100
non vacciné	50	150	200
total	82	218	300

On calcule les valeurs $f1$ et $f2$ qui sont les proportions des grippés des deux échantillons on tape :
`f1:=32/100`
`f2:=50/200=25/100`
On tape :
`f1-f2`
On obtient :
`7/100`
Donc $|f_1 - f_2| == 0.07$
On calcule la valeur p proportion des grippés lorsqu'on reunit les deux échantillons on tape :
`p:=82/300`
On obtient :
`41/150`
Donc $p \simeq 0.273333333333$
On calcule s_{12}, on tape :
`s12:=sqrt(p*(1-p)*(1/100+1/200))`
On obtient :
`sqrt(4469/1500000)`
Donc $s_{12} \simeq 0.0545832697201$
La variable $F = F_1 - F_2$ suit la loi normale $\mathcal{N}(0, s_{12})$ et sa valeur est $f = 0.07$.
On cherche la valeur a qui vérifie :
$Proba(|F| > a) = 0.05$ ou encore
$Proba(F \leq a) = 0.975$ et pour cela on tape :
`a:=normal_icdf(0,sqrt(4469/1500000),0.975)`
On obtient :
`0.10698124281`
Puisque `|f1-f2|=0.07<a=0.10698124281`, on en déduit que les deux échantillons ne sont pas significativement différents au seuil de 5% : on peut donc dire que le vaccin n'est pas efficace mais ce n'est pas une certitude...
Remarque
On a `h:=normal_icdf(0,1,0.975)=1.95996398454`
et `|f1-f2|=0.07<h*sqrt(4469/1500000)=0.10698124281`

3.8.2 Comparaison de deux moyennes observées

Soient m_1 et m_2 les moyennes observées d'un caractère dont la moyenne théorique est μ. Cette observation est faite à partir de deux échantillons de taille respéctive n_1 et n_2.
On veut savoir si les moyennes m_1 et m_2 sont significativement différentes ce qui voudrait dire que les deux échantillons proviennent de deux populations différentes de moyenne μ_1 et μ_2 ou si au contraire les deux échantillons proviennent d'une même population ou de populations de même moyenne $\mu = \mu_1 = \mu_2$.
Soient deux caractères normaux indépendants X et Y distribués respectivement selon les lois $\mathcal{N}(\mu_1, \sigma(X))$ et $\mathcal{N}(\mu_2, \sigma(Y))$,
On veut donc tester l'hypothèse $H_0 : \mu_1 = \mu_2 = \mu$ contre $H_1 : \mu_1 \neq \mu_2$ au seuil α.
Soient deux échantillons considérés l'un comme échantillon du caractère X et l'autre comme échantillon du caractère Y, de taille respective n_1 et n_2 de moyenne respective m_1 et m_2 et d'écart-type respectif s_1 et s_2.
Soit $\bar{X}$ (resp $\bar{Y}$) la variable aléatoire égale à la moyenne du caractère X (resp Y) pour des échantillons de taille n_1 (resp n_2).
On a :
$\bar{X}$ a pour moyenne μ_1 et comme écart-type $\frac{\sigma(X)}{\sqrt{n_1}}$
$\bar{Y}$ a pour moyenne μ_2 et comme écart-type $\frac{\sigma(Y)}{\sqrt{n_2}}$

Cas où $\sigma(X)$ et $\sigma(Y)$ sont connus
On a si $\mu_1 = \mu_2$:
$\frac{\bar{X}-\bar{Y}}{\sqrt{\sigma(X)^2/n_1+\sigma(Y)^2/n_2}}$ suit approximativement une loi $\mathcal{N}(0,1)$.

Cas où $\sigma(X)$ et $\sigma(Y)$ ne sont pas connus
On les estime :

- si n_1 et n_2 sont grands,
$\sigma(X) \simeq s_1\sqrt{\frac{n_1}{n_1-1}}$ donc $\frac{\sigma(X)^2}{n_1} \simeq \frac{s_1^2}{n_1-1}$
$\sigma(Y) \simeq s_2\sqrt{\frac{n_2}{n_2-1}}$ donc $\frac{\sigma(Y)^2}{n_2} \simeq \frac{s_2^2}{n_2-1}$
On pose :
$s_{12} = \sqrt{\sigma(X)^2/n_1 + \sigma(Y)^2/n_2} \simeq \sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}$
Donc sous l'hypothèse $H_0 : \mu_1 = \mu_2 = \mu$, on a $(\bar{X} - \bar{Y}) \in \mathcal{N}(0, s_{12})$
Recette si n_1 et n_2 sont grands
Avec `Xcas` on tape si $\alpha = 0.05$:
`a:=normal_icdf(0,s12,0.975)`
On regarde si :
$|m_1 - m_2| < a$
Si c'est le cas, on admet que $\mu_1 = \mu_2$ et que les deux échantillons ne sont pas significativement différents au seuil α, sinon on dira que $\mu_1 \neq \mu_2$ et que les deux échantillons ne proviennent pas de la même population.

- si n_1 et n_2 sont petits,
on peut estimer $\sigma(X)$ et $\sigma(Y)$ grâce à la reunion des deux échantillons et en faisant l'hypothèse $\sigma(X) = \sigma(Y)$ (pour vérifier cette hypothèse on pourra faire une étude de l'hypothèse $\sigma(X) = \sigma(Y)$ grâce au test expliqué au paragraphe suivant).
On montre qu'une bonne approximation est :

$\sigma^2 = \sigma(X)^2 = \sigma(Y)^2 \simeq s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$.
En effet, la statistique $\dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$ est un estimateur sans biais de σ^2 si σ est l'écart-type de X. La valeur de cette statistique est obtenue à partir de deux échantillons de taille respective n_1 et n_2 et d'écart-type respectif s_1 et s_2 qui sont les valeurs de S_1 et S_2 pour ces deux échantillons (avec comme notation $S^2 = \dfrac{1}{n}\sum_j (X_j - \bar{X})^2$ pour un échantillon de taille n de la variable X d'écart-type σ, on sait que $\dfrac{n}{n-1}S^2$ est un estimateur sans biais de σ^2) :
On a :
$\sigma^2 = \dfrac{n_1}{n_1 - 1}E(S_1^2) = \dfrac{n_2}{n_2 - 1}E(S_2^2)$ donc
$E(\dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}) = \dfrac{n_1 E(S_1^2) + n_2 E(S_2^2)}{n_1 + n_2 - 2} = \dfrac{(n_1 - 1)\sigma^2 + (n_2 - 1)\sigma^2}{n_1 + n_2 - 2} = \sigma^2$
donc $\sigma^2 \simeq s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$.
Alors sous l'hypothèse H_0 : $\mu_1 = \mu_2 = \mu$, et $\sigma(X) = \sigma(Y) = \sigma$, la statistique :

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma(X)^2/n_1 + \sigma(Y)^2/n_2}} \simeq \frac{(\bar{X} - \bar{Y})\sqrt{n_1 + n_2 - 2}}{\sqrt{(n_1 s_1^2 + n_2 s_2^2)(1/n_1 + 1/n_2)}}$$

suit une loi de Student à $n_1 + n_2 - 2$ degrés de liberté.
Recette si n_1 et n_2 sont petits
Avec `Xcas` on tape si $\alpha = 0.05$:
```
a:=student_icdf(n1+n2-2,0.975)
```
On regarde si :
$|m_1 - m_2| < a$
Si c'est le cas, on admet que $\mu_1 = \mu_2$ et que les deux échantillons ne sont pas significativement différents au seuil α, sinon on dira que $\mu_1 \neq \mu_2$ et que les deux échantillons ne proviennent pas de la même population.

3.8.3 Comparaison de deux écarts-types observés

Soient s_1 et s_2 les écarts-types observés d'un caractère dont l'écart-type théorique est σ. Cette observation est faite à partir de deux échantillons de taille respéctive n_1 et n_2.
On veut savoir si les écarts-types s_1 et s_2 sont significativement différents ce qui voudrait dire que les deux échantillons proviennent de deux populations différentes d'écart-type respectif σ_1 et σ_2 ou si au contraire les deux échantillons proviennent d'une même population ou de deux populations de même écart-type $\sigma = \sigma_1 = \sigma_2$.
Soient deux caractères normaux indépendants X et Y distribués respectivement selon les lois $\mathcal{N}(\mu_1, \sigma_1)$ et $\mathcal{N}(\mu_2, \sigma_2)$.
Soient deux échantillons (un échantillon pour le caractère X et l'autre pour le caractère Y) de taille respective n_1 et n_2, de moyenne respective m_1 et m_2 et d'écart-type respectif s_1 et s_2 .
Posons :
$S_1^2 = 1/n_1 \sum_{j=1}^{n_1}(X_j - \bar{X})^2$ et
$S_2^2 = 1/n_2 \sum_{j=1}^{n_2}(Y_j - \bar{Y})^2$

Lorsque $\sigma_1 = \sigma_2 = \sigma$, la statistique :
$F_{1,2} = \dfrac{n_1(n_2-1)S_1^2}{n_2(n_1-1)S_2^2}$ suit une loi de Fisher-Snedecor $\mathcal{F}(n_1-1, n_2-1)$ à (n_1-1) et à (n_2-1) degrés de liberté.
De même la statistique :
$F_{2,1} = \dfrac{n_2(n_1-1)S_2^2}{n_1(n_2-1)S_1^2}$ suit une loi de Fisher-Snedecor $\mathcal{F}(n_2-1, n_1-1)$ à (n_2-1) et à (n_1-1) degrés de liberté.

Cette statistique $F_{1,2}$ ou $F_{2,1}$ va nous permettre de tester les hypothèses :
$H_0 : \sigma_1 = \sigma_2$ et $H_1 : \sigma_1 \neq \sigma_2$.
On rejettera l'hypothèse bilatérale H_0 si la valeur de $F_{1,2}$ est trop éloignée de 1.
Attention à l'ordre n_1, n_2, car les tables ne donnent que les valeurs de $\mathcal{F}$ supérieures à 1, on sera quelquefois amené à changer l'ordre des variables (on a $F_{1,2} = 1/F_{2,1}$).
Pour avoir $Proba(a < F_{1,2} < b) = 1-\alpha$, on cherche a et b vérifiant :
$Proba(\mathcal{F}(n_1-1, n_2-1) < b) = 1-\alpha/2$ et
$Proba(\mathcal{F}(n_1-1, n_2-1) < a) = \alpha/2$
dans une table de Fisher-Snedecor $\mathcal{F}(n_1-1, n_2-1)$ à (n_1-1) et (n_2-1) degrés de liberté.
On a alors, si on échange l'ordre de n_1, n_2 :
$Proba(\mathcal{F}(n_2-1, n_1-1) < 1/a) = 1-\alpha/2$
$Proba(\mathcal{F}(n_2-1, n_1-1) < 1/b) = \alpha/2$
Recette
- Choisir le seuil α
- Prélever les échantillons de taille n_1 et n_2,
- Calculer leurs écarts-types s_1 et s_2,
- Si $n_1(n_2-1)s_1^2 > n_2(n_1-1)s_2^2$, calculer :
$f = \dfrac{n_1(n_2-1)s_1^2}{n_2(n_1-1)s_2^2}$ (cas 1)
ou sinon, calculer :
$f = \dfrac{n_2(n_1-1)s_2^2}{n_1(n_2-1)s_1^2}$ (cas 2).
- Déterminer grâce a la table de Fisher h vérifiant :
$Proba(1 < \mathcal{F}(n_1-1, n_2-1) < h) = 1-\alpha/2$ (cas 1)
ou vérifiant :
$Proba(1 < \mathcal{F}(n_2-1, n_1-1) < h) = 1-\alpha/2$ (cas 2).
Avec `Xcas` on tape si $\alpha = 0.05$ et si $n_1(n_2-1)s_1 > n_2(n_1-1)s_2$,
`h:=fisher_icdf(n1-1,n2-1,0.975)`
ou si $n_1(n_2-1)s_1 < n_2(n_1-1)s_2$,
`h:=fisher_icdf(n2-1,n1-1,0.975)`
- si $f > h$ (c'est à dire si f s'éloigne trop de 1) on rejette l'hypothèse bilatérale $H_0 : \sigma_1) = \sigma_2$ sinon on l'accepte.
Remarque
Avec `Xcas` on tape si $\alpha = 0.05$:
`h:=fisher_icdf(n1-1,n2-1,0.975)`
`k:=fisher_icdf(n2-1,n1-1,0.975)`
Alors `k=1/k` et `h` et `k` définissent les bornes en dehors d esquelles il faut rejetter l'hypothèse au seuil 0.05.

3.9 Le test du χ^2

Dans ce chapitre on cherche à savoir si deux variables sont indépendantes (test d'indépendance) et à comparer la distribution du caractère étudié à une distribution théorique (test d'adéquation).
Par exemple, certains tests ne sont valables que lorsque le phénomène étudié suit une loi normale, ou bien lorsqu'on suppose l'indépendance de deux variables : il est donc important de savoir si cela est bien le cas.

3.9.1 Adéquation d'une distribution expérimentale à une distribution théorique

Considérons un échantillon de taille n ayant une distribution $x_1, .., x_k$ d'effectifs n_1,..,n_k (avec $n_1 + ... + n_k = n$) correspondant à l'observation d'une variable aléatoire X : X est discrète ou X est continue et dans ce cas on effectue un regroupement en k classes des valeurs de X, et $x_1, .., x_k$ représentent alors le centre de ces classes.
On veut comparer cette distribution empirique à une distribution théorique d'effectifs e_1,..,e_k (si chaque valeur x_j est obtenue avec la probabilité théorique p_j on a $e_j = np_j$).
La statistique $D^2 = \sum_{j=1}^{k} \frac{(n_j - e_j)^2}{e_j}$ est une bonne mesure de l'écart entre les effectifs observés et les effectifs théoriques : plus D^2 est proche de zéro, plus la distribution de l'échantillon est conforme à la distribution théorique.
L'objectif sera donc d'estimer si D^2 est suffisamment faible pour que l'on puisse ajuster la loi théorique à la distribution observée.
On montre que si n est grand, si $e_j > 5$ pour tout j, et si les e_j ont été obtenus sans avoir eu recours à l'échantillon, la statistique D^2 suit approximativement une loi du χ^2 à $\nu = (k - 1)$ degrés de liberté où k est le nombre de classes.
Lorsque l'on a eu recours à l'échantillon pour déterminer r paramètres, le nombre de degrés de liberté est alors de $\nu = (k - r - 1)$.
On note dans la suite ν le nombre de degrés de liberté.
La statistique D^2 est alors utilisée comme variable de décision dans le test d'hypothèses :
H_0 : pour tout $j = 1...k$, $Proba(X = x_j) = p_j$
H_1 : il existe $j = 1...k$, $Proba(X = x_j) \neq p_j$
On rejettera l'hypothèse d'adéquation au modèle dès que l'écart D^2 est supérieur à ce que l'on peut attendre de simples fluctuations dues à l'échantillonnage. La région critique au seuil α (c'est la région où il faudra rejeter l'hypothèse) est la région pour laquelle : $d^2 > h\}$ quand $Proba(\chi^2_\nu < h) = 1 - \alpha$ et lorsque d^2 est la valeur de D^2 pour l'échantillon.
On remarquera que D^2 fait intervenir le nombre de classes et les effectifs de chaque classe et que D^2 ne fera intervenir les x_j que pour estimer les paramètres p_j de la loi. Pour les effectifs e_j trop petits on effectuera un regroupement de classes.
Recette
Dans une table du χ^2 on cherche h tel que :
$Proba(\chi^2_{k-1} < h) = 1 - \alpha$
Avec `Xcas` on tape pour trouver h, si on a k classes et si $\alpha = 0.05$:

`chisquare_icdf(k-1,0.975)`
On prélève un échantillon de taille n et on note sa distribution $n_1, .., n_k$ correspondant aux k classes de centre $x_1, .., x_k$.
On calcule la valeur d^2 de D^2 : $d^2 = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j} = \sum_{j=1}^{k} \frac{(n_j - e_j)^2}{e_j}$
Règle
On rejette l'hypothèse H_0 au seuil α, quand d^2 est supérieure à h.
Exemple
Dans un croisement de fleurs rouges et blanches, on a obtenu le résultat suivant sur un échantillon de 600 plants de la 2-ième génération :
141 fleurs rouges, 315 fleurs roses, 144 fleurs blanches.
Ces résultats sont-ils conformes à la distribution théorique :
25% fleurs rouges, 50% fleurs roses, 25% fleurs blanches.
On a 3 classes donc 3-1=2 degrés de liberté :
$n_1 = 141$ et $e_1 = 600 * 25/100 = 150$
$n_2 = 315$ et $e_2 = 600 * 50/100 = 300$
$n_3 = 144$ et $e_3 = 600 * 25/100 = 150$
On calcule $d^2 = \sum_{j=1}^{3} \frac{(n_j - e_j)^2}{e_j} = \frac{81}{150} + \frac{15^2}{300} + \frac{36}{150}$ On tape dans `Xcas` :
`81/150+15^2/300+36/150`
On obtient :
`=153/100`
On tape :
`chisquare_icdf(2,0.95)`
On obtient :
`5.99146454711`
Comme $1.53 < 5.992$ on ne peut pas rejeter l'hypothèse H_0 au seuil de 5%, donc on l'accepte.

3.9.2 Adéquation d'une distribution expérimentale à une distribution de Poisson

Pour pouvoir calculer les effectifs théoriques, on est souvent obligé d'estimer le paramètre μ à partir de l'échantillon (μ est estimé par la moyenne m de l'échantillon).
Règle
Soit k est le nombre de classes.
Si on s'est servi de l'échantillon pour estimer μ, alors la statistique D^2 suit une loi du χ^2 à $k-2$ degrés de liberté (cas 1),
sinon D^2 suit une loi du χ^2 à $k-1$ degrés de liberté (cas 2) .
Pour savoir si la distribution $n_1, .., n_k$ correspondant aux k classes de centre $x_1, .., x_k$ est conforme à une distribution de Poisson, on utilise le test d'hypothèses :
H_0 : pour tout j $Proba(X = x_j) = e^{-x_j}\lambda^{x_j}/x_j! = p_j$ et
H_1 : il existe $j = 1...k$, $Proba(X = x_j) \neq p_j$
On rejette l'hypothèse H_0 au seuil α, quand la valeur d^2 de D^2 est supérieure à h avec h vérifiant :
- cas 1 : $Proba(\chi^2_{k-2} \leq\ h) = 1 - \alpha$,

- cas 2 : $Proba(\chi^2_{k-1} \leq\ h) = 1 - \alpha$.

Exemple

On a effectué un échantillon de taille 100 et on a obtenu, pour les 11 valeurs entières d'une variable aléatoire X les effectifs suivants :

X	n_j
0	1
1	8
2	19
3	23
4	17
5	15
6	8
7	3
8	3
9	2
10	1

Peut-on dire que X suit une loi de Poisson ?

On suppose que cela est vrai et on estime le paramètre de la loi de Poisson par la moyenne de l'échantillon.

Ssoit on utilise le tableur, soit on tape :

```
L1:=[0,1,2,3,4,5,6,7,8,9,10]
L2:=[1,8,19,23,17,15,8,3,3,2,1]
mean(L1,L2)
```

On obtient :

```
379/100
```

On cherche les effectifs théoriques on tape ($n = 100$) :

`100*poisson(3.79,0)`, `100*poisson(3.79,1)`, etc ...

`100*poisson(3.79,9)`, `100*poisson(3.79,10)`.

On rappelle que : `poisson(3.79,k)=exp(-3.79)*(3.79^k)/k!`

ou bien on tape

```
L:=[];for(j:=0;j<11;j++) {
L:=concat(L,poisson(3.79,j)*100);}
```

ou encore on tape :

```
L:=seq(100*poisson(3.79,k),k,0,10)
```

On obtient la liste `L` des 11 valeurs de e_j pour $j = 0..10$:

```
[2.25956018511,8.56373310158,16.2282742275,20.5017197741,
19.4253794859,14.7244376503,9.30093644912,5.0357927346,
2.38570680802,1.00464764471,0.380761457345]
```

il faut changer la valeur de la dernière classe car elle doit comporter toutes les valeurs supérieures ou égales à 10 (la somme des e_j est égale à la taille de l'échantillon ici 100) :

```
L[10]:=100-sum(L[j],j,0,9)
```

On obtient :_ `0.56981193906`

Donc on obtient la liste `L` des e_j :

```
[2.25956018511,8.56373310158,16.2282742275,20.5017197741,
19.4253794859,14.7244376503,9.30093644912,5.0357927346,
```

```
2.38570680802,1.00464764471,0.56981193906]
```
On regroupe les petits effectifs pour avoir $e_j > 5$, on a donc 7 classes :
On tape :
```
L[0]+L[1]
```
On obtient :
```
10.8232932867
```
```
L[7]+L[8]+L[9]+L[10]
```
On obtient :
```
8.99595912639
```
Ou encore, on tape :
```
L:=accumulate_head_tail(L,2,4)
```
Donc on obtient la liste `L` d'effectifs théoriques e_j avec $e_j > 5$:
```
L:=[10.8232932867,16.2282742275,20.5017197741,
19.4253794859,14.7244376503,9.30093644912,8.99595912639]
```
La liste `L2` d'effectifs empiriques correspondant à ces 7 classes est :
```
L2:=[9,19,23,17,15,8,9]
```
On calcule :
```
L3:=(L2-L)^2
```
```
d2:=evalf(sum((L3[j]/L[j]),j,0,6))
```
On obtient :
```
1.57493190982
```
On sait que D^2 suit une loi du χ^2 ayant (7-2)=5 degrés de liberté car on a estimé λ par m moyenne de l'échantillon.
On tape pour connaitre la région critique au seuil de $\alpha = 0.05$:
```
chisquare_icdf(5,0.95)
```
On obtient :
```
11.0704976935
```
donc $h \simeq 11.07$:
$Proba(D^2 < 11.07) = 0.95$ ou encore $Proba(D^2 > 11.07) = 0.05$.
Cela veut dire que D^2 a des valeurs supérieure à 11.07 que dans 5% des cas c'est à dire très peu souvent ou encore que la probabilité que D^2 soit supérieur à 11.07 par le seul fait du hasard sur l'échantilonnage est 0.05, et dans ce cas il n'y aurait que 5 chances sur 100 pour que l'on ait alors une distribution de Poisson.
Donc si la valeur observée d^2 de D^2 est supérieure à 11.07 on rejetera l'hypothèse H_0 au seuil $\alpha = 0.05$.
Dans l'exemple ci-dessus, la valeur observée de D^2 est $d^2 = 1.575$, donc on estime que l'hypothèse selon laquelle la distribution est une distribution de Poisson n'est pas à rejeter au seuil de 5%.

3.9.3 Adéquation d'une distribution expérimentale à une distribution normale

Pour pouvoir calculer les effectifs théoriques, on est souvent obligé d'estimer les paramètres μ et σ à partir de l'échantillon (μ par la moyenne m de l'échantillon et σ par $s\sqrt{\frac{n}{n-1}}$ où s est l'écart-type de l'échantillon).

Règle
Soit k le nombre de classes.
Si on s'est servi de l'échantillon pour estimer μ et σ la statistique D^2 suit une loi

du χ^2 à $k-3$ degrés de liberté (cas 1),
si on s'est servi de l'échantillon pour estimer μ ou σ la statistique D^2 suit une loi du χ^2 à $k-2$ degrés de liberté (cas 2)
sinon D^2 suit une loi du χ^2 à $k-1$ degrés de liberté (cas 3) (k est le nombres de classes).
On rejette l'hypothèse H_0 au seuil α, quand la valeur d^2 de D^2 est supérieure à h avec h vérifiant :
- cas 1 : $Proba(\chi^2_{k-3} \leq h) = 1-\alpha$,
- cas 2 : $Proba(\chi^2_{k-2} \leq h) = 1-\alpha$,
- cas 3 : $Proba(\chi^2_{k-1} \leq h) = 1-\alpha$.

Exemple
On a effectué un échantillon de taille 250 et on a obtenu, pour les valeurs d'une variable aléatoire X, réparties en 10 classes, les effectifs suivants :

X	n_j
45..46	11
46..47	15
47..48	27
48..49	35
49..50	47
50..51	58
51..52	28
52..53	16
53..54	10
54..55	3

On va tout d'abord calculer la moyenne m et l'écart-type s de l'échantillon :
On tape :
```
L1:=[45..46,46..47,47..48,48..49,49..50,50..51,51..52
,52..53,53..54,54..55]
L2:=[11,15,27,35,47,58,28,16,10,3]
```
On tape :
```
m:=mean(L1,L2)
```
On obtient :
```
6207/125
```
Donc $m \simeq 49.656$
```
s:=stddev(L1,L2)
```
On obtient :
```
sqrt(249229/62500)
```
On obtient une estimation de σ en tapant :
```
s*sqrt(250/249)
```
On obtient :
```
2.00091946736
```
Donc $s \simeq 2$
On cherche les effectifs théoriques on tape :
```
normal_cdf(49.656,2,45,46)
```
On obtient :
```
0.0238187239894,
```

`normal_cdf(49.656,2,46,47)`,
etc ...
`normal_cdf(49.656,2,54,55)`.
On rappelle que :

$$\texttt{normal\_cdf}(\mu,\sigma,x_1,x_2)=\int_{x_1}^{x_2}\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(x-\mu)^2}{2*\sigma^2})dx$$

ou bien on tape

```
L:=[];
for(j:=0;j<10;j++) {
L:=concat(L,normal_cdf(49.656,2,45+j,46+j));}
```

ou encore on tape :

```
L:=seq(normal_cdf(49.656,2,45+j,46+j),j,0,9)
```

On obtient la liste `L` des p_j (p_j est la probabilité théorique pour que la valeur de X soit dans la j-ième classe) :

```
[0.0238187239894,0.0583142776342,0.11174619649,
0.167620581364,0.196825404189,0.180926916339,
0.130193320084,0.0733363670394, 0.0323343295781,
0.0111577990363]
```

Il faut modifier le premier terme et le dernier terme de `L` car la première classe est en fait $]-\infty;46[$ et la dernière $[54;+\infty[$.
On tape :

```
normal_cdf(49.656,2,-infinity,46)
```

On obtient :

```
0.0337747758231
```

On tape :

```
normal_cdf(49.656,2,54,+infinity)
```

On obtient :

```
0.0149278314584
L:=[0.0337747758231,0.0583142776342,0.11174619649,
0.167620581364,0.196825404189,0.180926916339,
0.130193320084,0.0733363670394,0.0323343295781,
0.0149278314584]
```

On obtient la liste `L` des effectifs théoriques e_j de chaque classe en tapant :

```
L:=250*L
```

On obtient :

```
[8.44369395578,14.5785694086,27.9365491225,41.
905145341,49.2063510472,45.2317290848,
32.548330021,18.3340917599,8.08358239453,
3.7319578646]
```

On regroupe les 2 dernières classes (`L[8]+L[9]=11.8155402591`),
Ou encore , on tape :
`L:=accumulate_head_tail(L,1,2)` on obtient la liste `L` des effectifs théoriques des 9 classes :

```
L:= [8.44369395578,14.5785694086,27.9365491225,
41.905145341,49.2063510472,45.2317290848,
32.548330021,18.3340917599,11.8155402591]
```

La liste `L2:=[11,15,27,35,47,58,28,16,10,3]` des effectifs de l'échantillon après un regroupement en 9 classes, on tape :

```
L2:=accumulate_head_tail(L2,1,2)
```
On obtient :
```
L2:=[11,15,27,35,47,58,28,16,13]
```
On calcule la valeur de D^2 :
```
d2:=sum(((L-L2)[j])^2/L[j],j,0,8)
```
On obtient :
```
6.71003239422
```
On calcule $Proba(\chi^2_6 < h) = 0.95$, pour cela on tape (car on a $9 - 3 = 6$ degrés de liberté) :
```
chisquare_icdf(6,0.95)
```
On obtient :
```
12.5915872437
```
donc $h \simeq 12.6$
L'hypothèse n'est pas à rejeter au seuil de 5% puisque $d2 = 6.71 < 12.6$.

3.10 Comparaison de la distribution de plusieurs échantillons

3.10.1 Cas général : on a m échantillons

Soient m échantillons, comment savoir si la distribution des fréquences de ces m échantillons sont celles d'échantillons d'une même loi ?

Notations

On suppose que les m échantillons peuvent prendre k valeurs numerotées de 1 à k.
On note à l'aide d'un indice (i) placé en haut ce qui concerne le i-ième échantillon ainsi, $n^{(i)}$ est la taille de l'échantillon i et $n_j^{(i)}$ est le nombre d'occurences de la valeur j dans la série i, donc i varie de 1 à m et j varie de 1 à k.
On a m échantillons et dans chaque échantillon il y a k classes.
On a donc :
$\sum_j n_j^{(i)} = n^{(i)}$ qui est la taille de l'échantillon (i).
On pose :
$\sum_{i,j} n_j^{(i)} = n$ et
$\sum_i n_j^{(i)} = n_j$.
Donc n est la taille de l'échantillon total constitué par les m échantillons, n_j est le nombre total d'occurences de la valeur j dans l'échantillon total.
D'après la loi des grands nombres, si on considère que les m échantillons suivent la même loi X que l'échantillon total on a $Proba(X = j) \simeq n_j/n$.
Donc on peut considérer que l'effectf théorique de la valeur j de l'échantillon (i) est :
$\nu_j^{(i)} = n^{(i)} * n_j/n$.
La variable de décision est alors :
$$D^2 = \sum_{i,j} \frac{(n_j^{(i)} - \nu_j^{(i)})^2}{\nu_j^{(i)}}$$
Cette variable suit une loi de χ^2 ayant $s = (m-1)(k-1)$ degrés de liberté.

3.10.2 Application à deux échantillons prenant deux valeurs

Soient $f_1^{(1)}$ et $f_1^{(2)}$ les fréquences observées sur deux échantillons d'un caractère dont la fréquence théorique est p. Cette observation est faite à partir de deux échantillons de taille respéctive $n^{(1)}$ et $n^{(2)}$ (même notations qu'en 3.10.1).
On veut savoir si les fréquences $f_1^{(1)}$ et $f_1^{(2)}$ sont significativement différentes ce qui voudrait dire que les deux échantillons proviennent de deux populations différentes de paramètre p_1 et p_2 ou si aucontraire les deux échantillons proviennent d'une même population de paramètre $p = p_1 = p_2$ c'est à dire que ces deux échantillons sont ceux d'une même loi (voir aussi 3.8.1).
On a :
$n^{(1)} + n^{(2)} = n$
$f_1^{(1)} n^{(1)} = n_1^{(1)}$ et $(1 - f_1^{(1)})n^{(1)} = n_2^{(1)} (= f_2^{(1)} n^{(1)})$
$f_1^{(2)} n^{(2)} = n_1^{(2)}$ et $(1 - f_1^{(2)})n^{(2)} = n_2^{(2)} (= f_2^{(2)} n^{(2)})$
$n_1 = f_1^{(1)} n^{(1)} + f_1^{(2)} n^{(2)}$
$n_2 = (1 - f_1^{(1)})n^{(1)} + (1 - f_1^{(2)})n^{(2)}$
$\nu_j^{(i)} = n^{(i)} n_j / n$ donc
$\nu_1^{(1)} - n_1^{(1)} = n^{(1)}(f_1^{(1)} n^{(1)} + f_1^{(2)} n^{(2)})/(n^{(1)} + n^{(2)}) - f_1^{(1)} n^{(1)} =$
$n^{(2)} n^{(1)} (f_1^{(2)} - f_1^{(1)})/(n^{(1)} + n^{(2)})$
$\nu_1^{(2)} - n_1^{(2)} = n^{(2)}(f_1^{(1)} n^{(1)} + f_1^{(2)} n^{(2)})/(n^{(1)} + n^{(2)}) - f_1^{(2)} n^{(2)} =$
$n^{(1)} n^{(2)} (f_1^{(1)} - f_1^{(2)})/(n^{(1)} + n^{(2)})$
$\nu_2^{(1)} - n_2^{(1)} = n^{(1)}((1-f_1^{(1)})n^{(1)} + (1-f_1^{(2)})n^{(2)})/(n^{(1)} + n^{(2)}) - (1-f_1^{(1)})n^{(1)} =$
$n^{(1)} n^{(2)} (f_1^{(1)} - f_1^{(2)})/(n^{(1)} + n^{(2)})$
$\nu_2^{(2)} - n_2^{(2)} = n^{(2)}((1-f_1^{(1)})n^{(1)} + (1-f_1^{(2)})n^{(2)})/(n^{(1)} + n^{(2)}) - (1-f_1^{(2)})n^{(2)} =$
$n^{(1)} n^{(2)} (f_1^{(2)} - f_1^{(1)})/(n^{(1)} + n^{(2)})$
$1/\nu_1^{(1)} + 1/\nu_1^{(2)} + 1/\nu_2^{(1)} + 1/\nu_2^{(2)} =$

$$(n^{(1)}+n^{(2)})\left(\frac{1}{n^{(1)}}+\frac{1}{n^{(2)}}\right)\left(\frac{1}{f_1^{(1)} n^{(1)} + f_1^{(2)} n^{(2)}}+\frac{1}{(1 - f_1^{(1)})n^{(1)} + (1 - f_1^{(2)})n^{(2)}}\right) =$$

$$\frac{(n^{(1)} + n^{(2)})^2}{n^{(1)} n^{(2)}}\left(\frac{1}{f_1^{(1)} n^{(1)} + f_1^{(2)} n^{(2)}} + \frac{1}{(1 - f_1^{(1)})n^{(1)} + (1 - f_1^{(2)})n^{(2)}}\right) =$$

$$\frac{(n^{(1)} + n^{(2)})^3}{n^{(1)} n^{(2)} (n_1 f_1^{(1)} + n_2 f_1^{(2)})(n_1(1 - f_1^{(1)}) + n_2(1 - f_1^{(2)}))}$$

La variable D^2 suit une loi du χ^2 a 1 degré de liberté : on a 2 échantillons ($m = 2$) et chaque échantillon ne prend que 2 valeurs, ($k = 2$) donc $s = (m-1)(k-1) = 1$.
La variable D^2 s'écrit alors :

$$D^2 = \frac{n^{(1)} n^{(2)} (f_1^{(1)} - f_1^{(2)})^2 (n^{(1)} + n^{(2)})}{(n^{(1)} f_1^{(1)} + n^{(2)} f_1^{(2)})(n^{(1)}(1 - f_1^{(1)}) + n^{(2)}(1 - f_1^{(2)}))}$$

ou encore

$$D^2 = \frac{n(n_1^{(1)} n_2^{(2)} - n_1^{(2)} n_2^{(1)})^2}{n^{(1)} n^{(2)} n_1 n_2}$$

D^2 suit une loi du χ^2 ayant 1 degré de liberté.
Exercice (le même qu'en section 3.8.1)
Pour tester l'efficacité d'un vaccin antigrippal on soumet 300 personnes à une expérience :
- sur 100 personnes non vaccinées, 32 sont atteintes par la grippe,

- sur 200 personnes vaccinées, 50 sont atteintes par la grippe,
Ce résultat permet-il d'apprécier l'efficacité du vaccin ?
On a le tableau suivant :

	grippé	non grippé	taille
vacciné	32	68	100
non vacciné	50	150	200
total	82	218	300

On calcule la valeur $d2$ de D^2 on tape :
`d2:=300*(150*32-68*50)^2/(100*200*82*218)`
On obtient :
`7350/4469`
donc $d^2 \simeq 1.645$
On cherche la valeur h qui vérifie :
$Proba(\chi_1^2 > h) = 0.05$ ou encore $Proba(\chi_1^2 \leq h) = 0.95$
pour cela on tape :
`chisquare_icdf(1,0.95)`
On obtient :
`3.84145882069`
donc $h \simeq 3.84$
Puisque `d2` $\simeq$ `1.645` $<$ `3.84` on en déduit que les deux échantillons ne sont pas significativement différents au seuil de 5% : on peut donc mettre en doute l'efficacité du vaccin.

3.11 Application : le test d'indépendance

C'est une application du test d'adéquation.
Considérons une variable aléatoire X valant $x_1, .., x_k$ avec une probabilité théorique p_1,..,p_k (avec $p_1 + ... + p_k = 1$) et une variable aléatoire Y valant $y_1, .., y_l$ avec une probabilité théorique q_1,..,q_l (avec $q_1 + ... + q_l = 1$).
On a un échantillon de taille n (n grand) pour lequel le nombre d'éléments présentant le caractère x_i et le caractère y_j est $n_{i,j}$ ($\sum n_{i,j} = n$).
On veut savoir, au vue de l'échantillon si les variables X et Y sont indépendantes.
On peut estimer les p_i et les q_j par :
$p_i \simeq \sum_{j=1}^{l} n_{i,j}/n$
$q_j \simeq \sum_{i=1}^{k} n_{i,j}/n$
En estimant ces valeurs, on a estimé $k-1+l-1 = k+l-2$ paramètres (car quand on a estimé $p_1, .., p_{k-1}$ on a l'estimation de p_k et quand on a estimé $q_1, .., q_{l-1}$ on a l'estimation de q_l).

Si X et Y sont indépendantes (hypothèse H_0), alors :
$Proba((X = x_i) \cap (Y = y_j)) = p_i q_j$
donc l'effectif théorique des éléments présentant le caractère x_i et y_j est :
$e_{i,j} = np_iq_j$.
La statistique $D^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{i,j} - np_iq_j)^2}{np_iq_j}$ suit approximativement une loi du χ^2 ayant $(k-1)(l-1)$ degrés de liberté (car $(k-1)(l-1) = kl - 1 - (k+l-2)$).

Règle
On calcule d^2 la valeur de D^2 pour l'échantillon et $\nu = (k-1)(l-1)$ le nombre de degrés de liberté.
On cherche dans une table la valeur de h vérifiant : $Proba(\chi^2_\nu < h) = 1 - \alpha$
Avec `Xcas` on tape si $\alpha = 0.05$:
`h:=chisquare_icdf((k-1)(l-1),0975)`
Si $d^2 < h$, on accepte l'hypothèse d'indépendance au seuil α, sinon on la rejette.

3.12 Le test de corrélation

On considère une série statistique double, c'est à dire que pour chaque individu d'une même population, on étudie deux caractères X et Y. On veut savoir si ces deux caractères ont une relation entre eux.
L'ensemble des valeurs (x_j, y_j) de (X, Y) s'appelle un nuage de points.

Rappel : Soient deux variables aléatoires X et Y, on définit le coefficient de corrélation ρ de ces deux variables par le nombre :
$\rho = \dfrac{E((X - E(X))(Y - E(Y)))}{\sigma(X)\sigma(Y)} = \dfrac{E(XY) - E(X)E(Y)}{\sigma(X)\sigma(Y)} = \dfrac{cov(X,Y)}{\sigma(X)\sigma(Y)}$
Si X et Y sont indépendantes alors $\rho = 0$.
Dans le cas où le nuage de points de coordonées $(x_j; y_j)$ est linéaire, l'équation de la droite, dite de régression, est :
$y = ax + b$ avec
$a = \dfrac{E((X - E(X))(Y - E(Y)))}{\sigma(X)^2}$ et
$b = E(Y) - aE(X)$
On a donc :
$\rho = a\sigma(X)/\sigma(Y)$
Théorème :
Au vue d'un échantillon de taille n, on peut estimer ρ par l'estimateur :
$R = \dfrac{\sum_{j=1}^n (X_j - \bar{X})(Y_j - \bar{Y})}{\sqrt{(\sum_{j=1}^n (X_j - \bar{X})^2)(\sum_{j=1}^n (Y_j - \bar{Y})^2)}}$
Lorsque X et Y suivent une loi normale les variables :
$V = \dfrac{1}{2}\ln(\dfrac{(1+R)(1-\rho)}{(1-R)(1+\rho)})$ suit une loi normale $\mathcal{N}(\dfrac{\rho}{2n-2}, \dfrac{1}{\sqrt{n-3}})$ et,
$T = \dfrac{\sqrt{n-2}R}{\sqrt{1-R^2}}$ suit une loi de Student à $n-2$ degrés de liberté.
Si $R^2 = 1$, les points de coordonées $(x_j; y_j)$ sont alignés sur la droite des moindres carrés et,
si $R^2 = 0$, cela permet de conclure à l'inadéquation du modèle linéaire.
Attention : si $R = 0$, les variables X et Y ne sont pas obligatoirement indépendantes. De même, lorsque R^2 est proche de 1, on peut penser (c'est un indice et non une preuve) qu'il y a un lien de cause à effet entre X et Y.
On peut donc tester au seuil α l'hypothèse $H_0 : \rho = 0$.
Par exemple, pour $\alpha = 0.05$, on considère que $\rho = 0$ est vraisemblable si :
$\dfrac{1}{2}\ln(\dfrac{(1+R)}{(1-R)}) < 1.96 * \dfrac{1}{\sqrt{n-3}}$
Pour estimer a et b on utilise les statistiques :

$A = \frac{\sum((X_j - \bar{X})(Y_j - \bar{Y}))}{\sum(X_j - \bar{X})^2}$ et $B = \bar{Y} - A\bar{X}$
On montre que A et B sont des estimateurs sans biais de a et b.

Chapitre 4

Résolution d'exercices de statistiques

4.1 Statistiques à 1 variable

— **Exercice 1**
Voici les notes obtenues dans une classe de terminale.
6,10,14,17,9,6,4,12,9,10,10,11,12,18,10,9,11,8,7,10
Calculer la moyenne et l'écart-type de cette série.
On tape :

```
mean([6,10,14,17,9,6,4,12,9,10,10,11,12,
      18,10,9,11,8,7,10])
```

On obtient :

```
[10.15]
```

On tape :

```
stddev([6,10,14,17,9,6,4,12,9,10,10,11,12,
      18,10,9,11,8,7,10])
```

On obtient :

```
sqrt(4451/400) ≃ [3.33579076082]
```

On tape :

```
quartiles([6,10,14,17,9,6,4,12,9,10,10,11,12,
      18,10,9,11,8,7,10])
```

On obtient :

```
[[4.0],[8.0],[10.0],[11.0],[18.0]]
```

Donc la plus mauvaise note est 4 et la meilleure note est 18.
La médiane est 10, de plus 25% des élèves ont une note inférieure à 8, et 25% des élèves ont une note supérieure à 11 ou encore 50% des élèves ont une note entre 8 et 11.

— **Exercice 2**
Sur un échantillon de 100 personnes, on relève leur taille.
y désigne la taille en centimètres et n l'effectif, on obtient le tableau :

y	150<y ≤ 155	155<y ≤ 160	160<y ≤ 165	165<y ≤ 170
n	30	25	23	22

Construire l'histogramme.
Calculer la moyenne et l'écart-type de cette série.
Avec `Xcas`, on utilise un tableur que l'on obtient avec le raccourci clavier `Alt+t`.
On remplit la colonne `A` :
on met `150..155` dans `A0`, `155..160` dans `A1` etc...ou on tape dans `A1=A0+5` puis on remplit vers le bas avec le menu du tableur `Edit` puis `Remplir` puis `Copier vers le bas`.
On remplit la colonne `B` :
on met `30` dans `B0`, `25` dans `B1` etc...
On sélectionne ces 2 colonnes pour cela, soit on le fait à la souris soit on tape dans la `case de sélection` `A0..B3`.
Avec le menu du tableur `Statistiques` puis `1d` puis `histogram`, on obtient l'histogramme.
On tape dans `A4` :
`=mean(A0:A3,B0:B3)`
On obtient la moyenne : `3187/20`
On tape dans `B4` :
`=stddev(A0:A3,B0:B3)`
On obtient l'écart-type : `sqrt(12731/400)` $\simeq$ `5.64158665625`
On obtient le résultat dans la ligne de commande en appuyant sur `eval` et la valeur approchée avec la commande `evalf`.

— **Exercice 3**
Sur un échantillon de 100 personnes, on relève leur poids.
x désigne le poids en kilogrammes et n l'effectif, on obtient le tableau :

x	40<x $\leq$ 45	45<x $\leq$ 50	50<x $\leq$ 55	55<x $\leq$ 60
n	22	33	24	21

Construire l'histogramme.
Déterminer la moyenne et l'écart-type de cette série.
Avec `Xcas`, on utilise un tableur que l'on obtient avec le raccourci clavier `Alt+t`.
On remplit la colonne `C` :
on met `40..45` dans `C0`, `45..50` dans `C1` etc...
On remplit la colonne `D` :
on met `22` dans `D0`, `33` dans `D1` etc...
On sélectionne ces 2 colonnes, on peut faire cette sélection à la souris ou en tapant dans la `case de sélection` `C0..D3`.
Avec le menu du tableur `Statistiques` puis `1d` puis `histogram`, on obtient l'histogramme.
On tape dans `C4` :
`=mean(C0:C3,D0:D3)`
On obtient la moyenne : `497/10`
On tape dans `D4` :
`=stddev(C0:C3,D0:D3)`
On obtient l'écart-type : `sqrt(1383/50)` $\simeq$ `5.25927751692`
On obtient le résultat dans la ligne de commande en appuyant sur `val` et la valeur approchée avec la commande `evalf` .

4.2 Intervalle de confiance

Un exercice pour bien comprendre qu'un intervalle de confiance dépend de l'échantillon.
Une usine fabrique des pièces de diamètre μ. On suppose que la variable aléatoire X qui, à chaque pièce associe son diamètre suit une loi normale de moyenne μ et d'écart-type $\sigma = 1.1$.
On cherche à estimer μ à partir d'un échantillon d'effectif $n = 40$.
On regroupe les résultats en classes, on a obtenu :
2 pièces ont un diamètre entre 32.5 et 33.5,
7 pièces ont un diamètre entre 33.5 et 34.5,
19 pièces ont un diamètre entre 34.5 et 35.5,
8 pièces ont un diamètre entre 35.5 et 36.5,
3 pièces ont un diamètre entre 36.5 et 37.5,
1 pièce a un diamètre entre 37.5 et 38.5,
1/ Déterminer la moyenne, l'écart-type et l'histogramme de cet echantillon.
Réponse :
On tape dans la colonne `A` :
33,34,35,36,37,38 (ou 32.5..33.5 etc...mais c'est plus long ! ! !) et
dans la colonne `B` :
2,7,18,8,3,2.
puis en `A9` on tape `=mean(A0:A5,B0:B5)`,
on trouve `1409/40 =35.225` et
en `B9` on tape `=stddev(A0:A5,B0:B5)`,
on trouve `sqrt(2039/1600≃ 1.12888219049`.
Pour réaliser l'histogramme on sélectionne les colonnes `A` et `B`, on peut faire la sélection à la souris ou on tape dans la `case de sélection A0,5,B`.
Avec le menu `Statistiques` du tableur, on choisit `1d` puis `histogram` et on obtient l'histogramme.
2/ Déterminer à partir de l'échantillon : - un intervalle de confiance à 95% pour la moyenne μ, et
- un intervalle de confiance à 99% pour la moyenne μ.
Réponse :
On connait σ et on sait que la variable $\bar{\mathtt{X}}$ égale à la moyenne des échantillons de taille n suit une loi normale de moyenne μ et d'écart-type $\sigma/\sqrt{n}$.
On a $\sigma/\sqrt{n} = 1.1/\sqrt{40} = 0.173925271309$.
On sait que l'on a :
$Prob(|\bar{X} - \mu| < k * \sigma/\sqrt{n}) = 0.95$ pour $k = 1.96$ et,
$Prob(|\bar{X} - \mu| < k * \sigma/\sqrt{n}) = 0.99$ pour $k = 2.576$
donc on a :
$\bar{X} - 1.96 * \sigma/\sqrt{n} < \mu < \bar{X} + 1.96 * \sigma/\sqrt{n}$ dans 95% des cas et,
$\bar{X} - 2.576 * \sigma/\sqrt{n} < \mu < \bar{X} + 2.576 * \sigma/\sqrt{n}$ dans 99% des cas.
Au vu de l'échantillon on a $\bar{X} = 35.225$:
$a_1 = 35.225 - 1.96 * 1.1/\sqrt{40} = 34.8841064682$
$b_1 = 35.225 + 1.96 * 1.1/\sqrt{40} = 35.5658935318$
$a_2 = 35.225 - 2.576 * 1.1/\sqrt{40} = 34.7769685011$
$b_2 = 35.225 - 2.576 * 1.1/\sqrt{40} = 35.6730314989$
Avec `Xcas` on tape et on obtient :

```
a1:=normal_icdf(1409/40,1.1/sqrt(40),0.025)
=34.8841127322
b1:=normal_icdf(1409/40,1.1/sqrt(40),0.975)
=35.5658872678
a2:=normal_icdf(1409/40,1.1/sqrt(40),0.005)
=34.7769981895
b2:=normal_icdf(1409/40,1.1/sqrt(40),0.995)
=35.6730018105
```

on en déduit que :
`[34.88;35.57]` est un intervalle de confiance à 95% pour μ et que
`[34.77;35.68]` est un intervalle de confiance à 99% pour μ.

3/ On suppose encore que $\sigma = 1.1$ et qu'un échantillon de taille $n = 100$ a une moyenne de 35.225.
Déterminer à partir de cet échantillon, un intervalle de confiance à 95% pour la moyenne μ.

Réponse :
On a :
$Proba(\bar{X} - 1.96 * \sigma/\sqrt{n} < \mu < \bar{X} + 1.96 * \sigma/\sqrt{n}) = 0.95.$
Au vu de cet échantillon la valeur de $\bar{X}$ est de 35.225 on a :
$35.225 - 1.96 * 1.1/\sqrt{100} = 35.0094$
$35.225 + 1.96 * 1.1/\sqrt{100} = 35.4406$
Ou avec `Xcas` on tape :

```
normal_icdf(35.225,1.1/sqrt(100),0.025)
```

On obtient : `35.0094039617`

```
normal_icdf(35.225,1.1/sqrt(100),0.975)
```

On obtient : `35.4405960383`
On en déduit que :
`[35; 35.45]` est un intervalle de confiance pour μ au seuil de 5%.
Donc quand on augmente la taille de l'échantillon on a un intervalle de confiance de plus faible amplitude, en effet, on a une information plus précise avec un échantillon de taille plus grande.

4/ On suppose que X suit la loi normale $\mathcal{N}(35.25, 1.1)$.
Simuler la prise de 5 échantillons de taille 100 et déterminer pour chacun des échantillons un intervalle de confiance pour la moyenne μ au seuil de 5%, dans les deux cas suivant :
a/ lorsqu'on suppose que l'on connait $\sigma = 1.1$
b/ lorsqu'on estime σ à l'aide de l'échantillon.

Réponse :
On demande d'avoir 102 lignes dans le tableur en tapant `A102` dans la case de sélection.
On tape en `A0` : `=randnorm(35.25,1.1)`
puis on sélectionne `A0` et on appuie sur `remplir` et `vers le bas`.
On tape en `A100` : `=mean(A0:A99)`
On tape en `A101` : `=stddev(A0:A99)`
On tape en `A102` : `=A101*10/sqrt(99)`
puis on recopie toutes ces formules sur les colonnes `B,C,D,E` si on veut voir les 5 échantillons et on se met en mode `manual`,on sélectionne pour cela `Ne pas recalculer automatiquement` dans le sous-menu `Configuration` du

menu `Edit` du tableur. En effet en mode `auto` chaque fois que l'on valide une cellule contenant `=randnorm(35.25,1.1)`, on a un nouvel échantillon grâce au recalcul automatique.
On obtient par exemple :
La ligne 100 est la liste `m` des valeurs des moyennes des 5 échantillons :
`[35.2341469676,35.3942572081,35.0898127739,`
`35.1447916945,35.2456441276],`
La ligne 101 est la liste `s` des valeurs des écarts-types des 5 échantillons :
`[1.00342913254,1.14149481601,1.19977064554,`
`1.00252282025,1.09862748198],`
ligne 102 est la liste σ_est des valeurs estimées de l'écart-type σ :
`[1.00848422314,1.14724545601,1.20581486841,`
`1.00757334501,1.10416216427]`
On rajoute 2 lignes au tableur (dans la `case de sélection` on tape A104)
Dans la cellule `A103` on tape puisque $1.1/\sqrt{100} = 0.11$:
`normal_icdf(A100,0.11,0.025)..normal_icdf(A100,0.11,0.975)`
On recopie cette formule sur la ligne 103.
On obtient sur la ligne 103 :
`[35.0185509293..35.4497430059,35.1786611698..35.6098532464,`
`34.8742167356..35.3054088122,34.9291956562..35.3603877328,`
`35.0300480893..35.4612401659]`
d'où lorsqu'on connait $\sigma = 1.1$, les intervalles de confiance pour μ, au seuil de 5%, sont pour les 5 échantillons :
`[35.0185509293; 35.4497430059]`
`[35.1786611698; 35.6098532464]`
`[34.8742167356; 35.3054088122]`
`[34.9291956562; 35.3603877328]`
`[35.0300480893; 35.4612401659]`
Dans la cellule `A104`, on tape, puisque l'on estime $\sigma/\sqrt{100}$ par $s/\sqrt{99}$:
`normal_icdf(A100,A101/sqrt(99),0.025)..`
`normal_icdf(A100,A101/sqrt(99),0.975)`
On recopie cette formule sur la ligne 104.
On obtient sur la ligne 104 :
`[35.0364840599..35.4318098753,35.1693970987..35.6191173175,`
`34.8534730597..35.3261524881,34.9473073189..35.3422760701,`
`35.0292283434..35.4620599118]`
d'où lorsqu'on estime σ, les intervalles de confiance pour μ, au seuil de 5%, sont pour les 5 échantillons :
`[35.0364840599; 35.4318098753]`
`[35.1693970987; 35.6191173175]`
`[34.8534730597; 35.3261524881]`
`[34.9473073189; 35.3422760701]`
`[35.0292283434; 35.4620599118]`
Si on reunit ses 5 échantillons on a :
$n = 500$
$m = (m[0]+m[1]+m[2]+m[3]+m[4])/5 = 176.108652772/5 = 35.2217305544$
$s^2 = (s[0]^2 + s[1]^2 + s[2]^2 + s[3]^2 + s[4]^2)/5 = 1.19227287804$ donc
$\sigma\_est = s\sqrt{500/499} = \sqrt{1.19227287804 * 500/499} = 1.09300603953$

d'où pour cet échantillon, un intervalle de confiance pour μ, au seuil de 5%, est :
`[35.1259243509; 35.3175367579]`
car
35.2217305544-1.96*1.09300603953/sqrt(500)=35.1259243509 et
35.2217305544+1.96*1.09300603953/sqrt(500)=35.3175367579
5/ On suppose que X suit la loi normale $\mathcal{N}(35.25, 1.1)$.
Simuler la prise de 5 échantillons de taille 40 et déterminer pour chacun des échantillons un intervalle de confiance pour la moyenne μ, au seuil de 5%, dans les deux cas suivant :
a/ lorsqu'on suppose que l'on connait $\sigma = 1.1$
b/ lorsqu'on estime σ à l'aide de l'écart type de l'échantillon.
Réponse :
On considére un échantillon de taille $n = 40$.
Il a pour moyenne $m = 35.531073986$ et pour écart type $s = 1.00296897139$
Pour les quatres autres échantillons de taille 40 on trouve par exemple :
$m = 35.6360091101$ et $s = 1.29301963917$
$m = 35.0684414822$ et $s = 0.951157103863$
$m = 35.4535840905$ et $s = 0.917989271482$
$m = 35.0910551678$ et $s = 1.05109677585$
a/ $\bar{X}$ suit une loi normale de moyenne μ et d'écart-type :
$\sigma/\sqrt{40} = 1.1/\sqrt{40} = 0.173925271309$.
On a :
$\bar{X} - 1.96 * \sigma/\sqrt{n} < \mu < \bar{X} + 1.96 * \sigma/\sqrt{n}$ dans 95% des cas.
On a, pour le premier échantillon :
$m - 1.96 * 1.1/\sqrt{40} = 35.1901804542$ et,
$m + 1.96 * 1.1/\sqrt{40} = 35.8719675178$,
d'où un intervalle de confiance de `[35.19;35.88]` pour μ, au seuil de 5%.
b/ On suppose que l'on ne connait pas σ. Ici, n est trop petit pour évaluer σ à l'aide de l'écart type s de l'échantillon.
On considère alors, $\mathtt{T} = \sqrt{\mathtt{n}-1}\dfrac{\bar{\mathtt{X}} - \mu}{\mathtt{S}}$ avec :
$\mathtt{S}^2 = \frac{1}{\mathtt{n}}\sum_{\mathtt{j}=1}^{\mathtt{n}}(\mathtt{X_j} - \bar{\mathtt{X}})^2$
$\mathtt{X_j}$ est la variable aléatoire qui au $\mathtt{j}^{\mathtt{ime}}$ tirage associe son résultat et
$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$.
Alors T suit une loi de Student à $(n-1)$ degrés de liberté.
Ici $n = 40$ et on lit sur la table de Student que lorsque il y a $\nu = 39$ degrés de liberté, $Proba(-t < T < t) = 0.95$ pour $t = 2.023$.
Ou bien avec Xcas on tape :
`student_icdf(39,0.025)`
On obtient :
`-2.02269092002`
`student_icdf(39,0.975)`
On obtient :
`2.02269092002`
Donc $\bar{X} - t * S/\sqrt{39} < \mu < \bar{X} + t * S/\sqrt{39}$.
Pour le premier échantillon on trouve : $m = 35.531073986$
$s = 1.00296897139$
$m - 2.023 * s/\sqrt{39} = 35.2061729645$

$m + 2.023 * s/\sqrt{39} = 35.8559750075$
d'où un intervalle de confiance de `[35.2;35.86]` pour μ au seuil de 5%.
Pour les 4 autres échantillons, on trouve :
`[35.2171492913;36.0548689289]`
`[34.7603243584;35.376558606]`
`[35.1562113297;35.7509568513]`
`[34.7505636611;35.4315466745]`
En estimant σ par $s\sqrt{40/39}$ on aurait obtenu pour le premier échantillon :
`[35.2162967736;35.8458511984]`
En effet avec `Xcas` on tape :
`normal_icdf(35.531073986,1.00296897139/sqrt(39),0.025)`
On obtient :
`35.2162967736`
On tape :
`normal_icdf(35.531073986,1.00296897139/sqrt(39),0.975)`
On obtient :
`35.8458511984`

4.3 Intervalle de confiance d'une fréquence

— **Exercice 1**
Sur un registre d'état civil, on a relevé 552 naissances dont 289 garçons.
a/ Estimer la fréquence `p` de naissance d'un garçon.
b/ Donner un intervalle de confiance pour cette estimation.
Réponse
a/ on a $f = 289/552 \simeq 0.523$, on estime donc `p` à 0.523
b/ Soit p_{est} un nombre entre 0 et 1 et on fait l'hypothése bilatérale :
$H_0 : p = p_{est}$ ($H_1 : p \neq p_{est}$).
Peut-on accepter ou rejeter l'hypothèse $p = p_{est}$ au seuil $\alpha = 0.05$?
Soit X la variable aléatoire égale à 1 pour la naissance d'un garçon et égale à 0 pour la naissance d'une fille. Soient $X_1 ... X_{552}$ les variables X correspondant aux 552 naissances et soit Y la variable aléatoire égale à $\frac{\sum X_i}{n}$ = moyenne du nombre de garçons obtenus pour un échantillon d'effectif 552.
La distribution Y est voisine d'une distribution normale de moyenne p_{est} et d'écart type $\sigma = \sqrt{\frac{p_{est}(1-p_{est})}{552}}$.
On a donc $Proba(|Y - p_{est}| < 1.96\sigma) = 0.95$.
La valeur de Y lors de l'expérimentation est $289/552$.
Soit A l'évènement $|Y - p_{est}| \geq 1.96\sigma$.
On rejette l'hypothèse bilatérale $p = p_{est}$ si $Proba(A) \leq 0.05$ et on l'accepte sinon.
Pour quelles valeurs de p_{est} a-t-on $P(A) \leq \alpha = 0.05$?
On a :
$Proba(|Y - p_{est}| = |289/552 - p_{est}| \leq 1.96\sqrt{\frac{p_{est}*(1-p_{est})}{552}}) = 0.95$
Résolvons l'inéquation où l'inconnue est p_{est} :
$(289/552 - p_{est})^2 - 1.96^2 * \frac{p_{est}*(1-p_{est})}{552} < 0$
ce qui veut dire que p_{est} est à l'intérieur des racines de l'équation en x :
$(289 - 552x)^2 - 1.96^2 * 552x * (1 - x) = 0$.

On tape :
`solve((289-552*x)^2-1.96^2*552*x*(1-x)=0)`
On obtient :
`[0.481866594885,0.564909321131]`
Ou on tape :
`solve((289-552*x)^2-1.96^2*552*x*(1-x)<0)`
On obtient :
`[(x>0.481866594885) && (x<0.564909321131)]`
Conclusion : on a $p \in [0.481\ ;\ 0.565]$ avec une probabilité de 0.95.
Remarque : on peut faire un calcul plus rapide car on peut estimer σ par : $\sqrt{\frac{f*(1-f)}{551}} = \sqrt{289*(552-289)/552^2/551} \simeq 0.0212770747076$.
Donc $k = 1.96 * 0.0213 < 0.042$ ce qui donne comme intervalle de confiance de p au seuil de 5% égal à :
$[0.523 - 0.042 = 0.481\ ;\ 0.523 + 0.042 = 0.565]$
Avec `Xcas`, on tape :
`normal_icdf(289/552,0.0212770747076,0.025)`
On obtient :
`0.481848424514`
`normal_icdf(289/552,0.0212770747076,0.975)`
On obtient :
`0.565253024761`
ce qui donne $[0.4818\ ;\ 0.5653]$ comme intervalle de confiance de p au seuil de 5%.

— **Exercice 2**
Dans un hôpital sur un échantillon de 458 malades admis pendant un trimestre il y a eu 141 décès. Estimer le pourcentage de décès par un intervalle de confiance au seuil de 0.01.
Réponse
On a :
$n = 458$,
$f = 141/458 \simeq 0.307860262009$,
$\sigma \simeq \sqrt{f(1-f)/457} = 0.0215931304967$
`normal_icdf(0,1,0.995)=2.57582930355 ≃ 2,58`
On obtient, si Y est la variable aléatoire moyenne du nombre de décès pour des échantillons de taille 458 :
$P(|Y-p| \geq k) = 0.01$ pour $k = 0.258 * \sigma \simeq 0.056$ donc un intervalle de confiance du pourcentage de décès, au seuil de 0.01 égal à : $[0.252;\ 0.364]$
Avec `Xcas`, on tape :
`normal_icdf(141/458,0.0215931304967,0.005)`
On obtient :
`0.25224004372`
`normal_icdf(141/458,0.0215931304967,0.995)`
On obtient :
`0.363480480297`
ce qui donne $[0.252\ ;\ 0.364]$, comme intervalle de confiance de p au seuil de 5%.

4.4 Statistiques à 2 variables

— Exercice 1
Sur un échantillon de 100 personnes, on relève leur poids et leur taille.
x désigne le poids en kilogrammes et y désigne la taille en centimètres on obtient le tableau :

"y x"	$40 < x \leq 45$	$45 < x \leq 50$	$50 < x \leq 55$	$55 < x \leq 60$
$150 < y \leq 155$	20	9	1	0
$155 < y \leq 160$	2	18	4	1
$160 < y \leq 165$	0	5	12	6
$165 < y \leq 170$	0	1	7	14

Représenter graphiquement le nuage de points.
Calculer le coefficient de corrélation.

Réponse
On remplit le tableur (colonnes A,B,C,D,E) avec le tableau ci-dessus en remplacant chaque classe par son centre, pou cela, on se place sur A0 et on met :
```
[["y x",42.5,47.5,52.5,57.5],[152.5,20,9,1,0],
[157.5,2,18,4,1],[162.5,0,5,12,6],[167.5,0,1,7,14]]
```
dans la ligne de commande, et sur une seule ligne, puis on valide (ne pas mettre = pour remplir plusieurs cellules).
On tape alors dans F0 :
```
=covariance(list2mat(A0:E4,5),-1)
```
ou on tape dans une ligne d'entrée :
```
covariance([["y\x",42.5,47.5,52.5,57.5],
[152.5,20,9,1,0],[157.5,2,18,4,1],[162.5,0,5,12,6],
[167.5,0,1,7,14]],-1)
```
On obtient dans F0 ou en réponse :
```
24.18
```
On tape dans F1 :
```
correlation(list2mat(A0:E4,5),-1)
```
ou on tape dans une ligne d'entrée :
```
correlation([["y x",42.5,47.5,52.5,57.5],[152.5,20,9,1,0],
[157.5,2,18,4,1],[162.5,0,5,12,6],[167.5,0,1,7,14]],-1)
```
On obtient dans F1 ou en réponse :
```
0.814946211639
```

— Exercice 2
Lors d'une épidémie on a relevé, à intervalles réguliers, le nombre de cas déclarés :

numéro du relevé : x_j	1	2	3	4
nombre de cas : y_j	94	221	446	1050

1/ Représenter les points (x_j, y_j) dans un repère convenable.
Un ajustement affine parait-il justifié ?
2/ On pose $z_j = \ln(y_j)$ où ln désigne le logarithme népérien.
Détérminer des valeurs décimales approchées à 0.01 près de z_1, z_2, z_3, z_4.

Représenter les points $(x-j, z_j)$ dans un repère convenable.
3/ Détérminer l'équation de la droite de régression pour la série statistique (x_j, z_j).
4/ Tracer cette droite sur le graphique de la question 2/.
Utiliser cet ajustement affine pour obtenir une formiule donnant approximativement y_j en fonction de x_j.
En déduire le nombre de cas prévisibles au 5-ième relevé.

Réponse
Avec `Xcas` on tape :
`evalf(log([94,221,446,1050]))`
On obtient :
`[4.54329478227,5.39816270152,6.10031895202,`
`6.95654544315]`
et si on demande un calcul avec 3 chiffres significatifs (bouton rouge `cas` et on met `Chiffres` à 3 puis `OK`) on obtient :
`[4.54,5.4,6.1,6.96]`
On tape :
`linear_regression([1,2,3,4],[4.54,5.4,6.1,6.96])`
On obtient :
`0.794190823313,3.76410341145`
donc $z = 0.794\,x + 3.76$
On tape :
`exponential_regression([1,2,3,4],[94.0,221,446,1050])`
On obtient :
`2.21264984755,43.1250231194`
donc $y = 43.1 * 2.21^x$
Vérifions :
$y = \exp(z) = \exp(0.794 * x + 3.76) = 42.9 * 2.21^x)$
Pour $x = 5$ on a $z(5) = 0.794 * 5 + 3.76 = 7.73$ donc
$y(5) = \exp(z(5)) = \exp(7.73) \simeq 2275$
ou encore $y(5) = 43.1 * 2.21^5 \simeq 2272$

4.5 Comparaison de deux échantillons

4.5.1 Test de comparaison de deux moyennes

— **Exercice 1**
Pour une même épreuve, voici les notes obtenues dans une classe de terminale du lycée A.
6,10,14,17,9,6,4,12,9,10,10,11,12,18,10,9,11,8,7,10.
et les notes obtenues dans une classe de terminale du lycée B.
2,10,14,13,9,6,1,12,9,10,10,10,12,15,19,9,11,8,9,10
1/ Analyser les résultats de chaque groupe.
2/ Peut-on considérer que les 2 groupes sont issus d'une même population ?
Réponse
1/ On tape :
`mean([6,10,14,17,9,6,4,12,9,10,10,11,12,18,10,`

```
                    9,11,8,7,10])
```

On obtient :

```
                       [10.15]
```

On tape :

```
 stddev([6,10,14,17,9,6,4,12,9,10,10,11,12,18,10,
                    9,11,8,7,10])
```

On obtient :

```
          sqrt(4451/400) ≃ [3.33579076082]
```

Le lycée A a une moyenne de 10.15 et un écart type d'environ 3.34.
On tape :

```
  mean([2,10,14,13,9,6,1,12,9,10,10,10,12,15,19,
                    9,11,8,9,10])
```

On obtient :

```
                         9.95
```

On tape :

```
 stddev([2,10,14,13,9,6,1,12,9,10,10,10,12,15,19,
                    9,11,8,9,10])
```

On obtient :

```
           sqrt(6179/400) ≃ 3.93033077488
```

Le lycée B a une moyenne de 9.95 et un écart type d'environ 3.93.
On tape pour le lycée A :

```
quartiles([6,10,14,17,9,6,4,12,9,10,10,11,12,18,10,
                    9,11,8,7,10])
```

On obtient :

```
         [[4.0],[8.0],[10.0],[11.0],[18.0]]
```

On tape pour le lycée B :

```
quartiles([2,10,14,13,9,6,1,12,9,10,10,10,12,15,19,
                    9,11,8,9,10])
```

On obtient :

```
         [[1.0],[9.0],[10.0],[12.0],[19.0]]
```

On voit que dans le lycée B la moitie des élèves ont entre 9 et 12 alors que dans le lycée A la moitie des élèves ont entre 8 et 11. Donc bien que la moyenne du lycée B soit inférieure à la moyenne du lycée A il semble que la classe du lycée B soit meilleure que celle du lycée A. 2/ Si on considère que les deux classes constituent deux échantillons pris au hasard dans une population où la note de l'épreuve est une variable aléatoire X de moyenne μ et d'écart-type σ.
La reunion des 2 échantillons donne un échantillon de taille $n = 40$.
de moyenne $\mu \simeq (9.95 + 10.15)/2 = 10.05$ et d'écart-type s.
On tape :

```
stddev([6,10,14,17,9,6,4,12,9,10,10,11,12,18,10,9,
11,8,7,10, 2,10,14,13,9,6,1,12,9,10,10,10,12,15,19,
9,11,8,9,10])
```

On obtient :

```
s=sqrt(5319/400) ≃ 3.64657373434.
```

La variable aléatoire $\bar{X}_{40}$, égale à la moyenne des échantillons de taille 40

a donc pour moyenne :
$\mu = 10.05$ et pour écart-type :
$\sigma/\sqrt{40} \simeq s/\sqrt{39} = 0.583919119794$.
Donc $\sigma \simeq 3.69302877574$
Remarque : On sait que la statistique $displaystyle\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}$ est un estimateur sans biais de σ^2 si σ est l'écart-type de X (cf 3.8.2). La valeur de cette statistique est obtenue à partir de deux échantillons de taille respective n_1 et n_2 et d'écart-type respectif s_1 et s_2 qui sont les valeurs de S_1 et de S_2 (avec comme notation $S^2 = \frac{1}{n}\sum_j (X_j - \bar{X})^2$ pour un échantillon de taille n de la variable X).
On tape (ici $n_1 = n_2 = 20$, $n_1 + n_2 - 2 = 38$, $s_1^2 = \frac{4451}{400}$, $s_2^2 = \frac{6179}{400}$) :
`sqrt(20/38*(6179/400+4451/400))`
On obtient alors comme approximation de σ : `3.7398986758`
On pose comme hypothèse H_0 : $\mu_1 = \mu_2 = \mu$ et pour hypothèse alternative H_1 : $\mu_1 \neq \mu_2$ et on teste ces hypothèses au seuil de 0.05.
$\bar{A} - \bar{B}$ suit une loi normale de moyenne 0 et d'écart type $\sigma\sqrt{1/20 + 1/20}$.
On a :
$\sigma = s\sqrt{40/39}$
$s = \sqrt{5319/400}$
$\sigma\sqrt{1/20 + 1/20} \simeq s\sqrt{40/39/10} \simeq \sqrt{5319/100/39} \simeq 1.16783823959$
Avec `Xcas`, on tape :
`normal_icdf(0,1.16783823959,0.975)`
On obtient :
`2.28892088937`
`normal_icdf(0,1.16783823959,0.025)`
On obtient :
`-2.28892088937`
On a $m_1 - m_2 = 10.15 - 9.95 = 0.2$ et comme
$-2.28892088937 < 0.2 < 2.28892088937$,
on accepte l'hypothèse $\mu_1 = \mu_2$ au seuil de 5%.
Autre méthode
On peut aussi utiliser la loi de Student :
On considére que $\mu_1 = \mu_2 = \mu$.
Alors $T = \dfrac{(\bar{A} - \bar{B})\sqrt{38}}{\sqrt{(20s_1^2 + 20s_2^2)(1/20 + 1/20)}}$ suit une loi de Student à 38 degrés de liberté.
On calcule la valeur t de T pour l'échantillon on tape :
`t:=(10.15-9.95)*sqrt(38)/sqrt(2*3.34^2+3.93^2)`
On obtient : `0.200644948434` On tape :
`student_icdf(38,0.975)`
On obtient : `2.02439416391`
Puisque $-2.02439416391 < 0.2 < 2.02439416391$, on accepte l'hypothèse $\mu_1 = \mu_2 = \mu$ au seuil de 5%.

— **Exercice 2**

Deux entreprises A et B livrent des pièces dans des paquets de 100 pièces.
On note X_1 (resp X_2) la variable aléatoire égale au nombre de pièces défectueuses par paquet provenant de A (resp B).
On note $\overline{X}_1$ (resp $\overline{X}_2$) la variable aléatoire égale au nombre moyen de

pièces défectueuses par paquet pour des échantillons aléatoires de 49 paquets (resp 64 paquets) provenant de A (resp B).
I) Sur un échantillon de 49 paquets provenant de A on compte le nombre de pièces défectueuses dans chaque paquet et on trouve :
7, 5, 5, 4, 4, 4, 9, 7, 9, 2, 7, 8, 7, 8, 4, 4, 9, 10,
5, 10, 6, 4, 5, 6, 1, 2, 5, 7, 8, 0, 6, 0, 1, 5, 2, 0,
5, 2, 3, 3, 4, 1, 3, 10, 1, 0, 10, 2, 7
1/ Calculer la moyenne m_1 et l'écart-type s_1 de cet échantillon.
2/ Donner une estimation de la moyenne μ_1 et de l'écart-type σ_1 de X_1.
3/ Donner une estimation de la moyenne et de l'écart-type de $\overline{X}_1$
Réponse :
1/ On met les données dans la colonne A du tableur ou on donne un nom à la liste du nombre de pièces défectueuses dans chaque paquet.

- Dans le tableur on tape les données dans les cellules `A0..A48`, puis en `A49` on tape :
`mean((A0):(A48))`
On obtient :
`237/49 ≃ 4.83673469388.`
en `A50` on tape :
`stddev((A0):(A48))`
On obtient :
`sqrt(21006/2401 ≃ 2.9578462847.`
- Dans une ligne d'entrée, on tape :
```
L:=[7,5,5,4,4,4,9,7,9,2,7,8,7,8,4,4,9,10,
5,10,6,4,5,6,1,2,5,7,8,0,6,0,1,5,2,0,
5,2,3,3,4,1,3,10,1,0,10,2,7]
```
puis on tape : `mean(L)`
On obtient m_1 :
`237/49 ≃ 4.83673469388 ≃ 4.84.`
puis on tape : `stddev(A)`
On obtient s_1 :
`sqrt(21006/2401) ≃ 2.9578462847.`
2/ On a un échantillon de grande taille ($n_1 = 49$) donc d'après la loi des grands nombres, on estime μ_1 par $m_1 = 237/49 \simeq 4.84$ et σ_1 par :
$s_1\sqrt{\frac{n_1}{n_1-1}} = \sqrt{21006/2401}\sqrt{49/48} \simeq 2.98849836021 \simeq 2.99.$
Donc X_1 suit à peu près une loi normale de moyenne $m_1 = 4.84$ et d'écart-type $\sigma_1 = 2.99$.
3/ La variable aléatoire $\overline{X}_1$ égale à la moyenne des échantillons de taille 49 suit à peu près une loi normale de moyenne $\mu_1 = 4.84$ et d'écart-type $\sigma_1/\sqrt{49} \simeq s_1/\sqrt{48} \simeq 0.426928337173 \simeq 0.427$.
II) Sur un échantillon de 64 paquets provenant de l'entreprise B on trouve une moyenne $m_2 = 3.88$ et un écart-type $s_2 = 1.45$:
1/ Donner une estimation de la moyenne μ_2 et de l'écart-type σ_2 de X_2.
2/ Donner une estimation de la moyenne et de l'écart-type de $\overline{X}_2$.
Réponse :
1/ On a un échantillon de grande taille ($n_2 = 64$) donc d'après la loi des

grands nombres, on estime μ_2 par $m_2 = 3.88$ et σ_2 par :
$s_2\sqrt{\frac{n_2}{n_2-1}} = 1.45\sqrt{64/63} \simeq 1.46146262897 \simeq 1.46$
On a donc X_2 suit une loi de moyenne $m_2 = 3.88$ et d'écart-type $\sigma_1 = 1.46$.
2/ La variable aléatoire $\overline{X}_2$ égale à la moyenne des échantillons de taille $n_2 = 64$ suit à peu près une loi normale de moyenne $\mu_2 = 3.88$ et d'écart-type $\sigma_2/\sqrt{64} \simeq s_2/\sqrt{63} \simeq 0.182682828621 \simeq 0.183$
III) On note D la variable aléatoire $\overline{X}_1 - \overline{X}_2$.
1/ Quelle est la loi de probabilité de D ? Déterminer la moyenne et l'écart-type de D.
2/ On pose pour hypothèse nulle $H_0 : \mu_1 = \mu_2$ et pour hypothèse alternative $H_1 : \mu_1 \neq \mu_2$. Calculer sous H_0, les nombres h et k tels que :
$Proba(-h < D < h) = 0.99$ et $Proba(-k < D < k) = 0.95$.
3/ Peut-on conclure après examen des échantillons donnés en I et II que la différence des moyennes observées est significative au seuil de risque de 1% ? au seuil de risque de 5% ?
Réponse :
1/ D suit à peu près une loi normale de moyenne :
$\mu_1 - \mu_2 \simeq 4.84 - 3.88 = 0.96$
et d'écart-type :
$\sqrt{\sigma_1^2/49 + \sigma_2^2/64} \simeq \sqrt{s_1^2/48 + s_2^2/63} =$
$\sqrt{21006/(2401 * 48) + 1.45^2/63} \simeq 0.453082418658 \simeq 0.46$
car la variance de D est la somme des variances de $\overline{X}_1$ et de $\overline{X}_2$.
2/ Sous l'hypothèse H_0, D suit à peu près une loi normale de moyenne 0 et d'écart-type $\sigma(D) = 0.46$.
h et k sont tels que :
$Proba(-h < D < h) = 0.99$ et $Proba(-k < D < k) = 0.95$.
Donc d'après les tables de la loi normale on a :
$h = 2.58 * \sigma_D = 2.58 * 0.46 = 1.1868 \simeq 1.19$ et
$k = 1.96 * \sigma_D = 1.96 * 0.46 = 0.9016 \simeq 0.9$.
Ou bien avec `Xcas` on a :
`h:=normal_icdf(0,0.46,0.995)=1.18488147963`
`k:=normal_icdf(0,0.46,0.975)=0.901583432888`
3/ Puisque la valeur de D pour l'échantillon est égale à 0.96, on conclut qu'au seuil de 5% on rejette l'hypothèse H_0 (car 0.96>k) mais par contre, au seuil de 1% on accepte l'hypothèse H_0 (car 0.96<h).
Bien comprendre :
Au seuil de 5%, on rejette H_0 et on se trompe dans moins de 5% des cas, c'est à dire que l'on rejette H_0 à tort dans moins de 5% des cas, mais si on ne veut se tromper que dans 1% des cas, on ne peut pas rejeter H_0 et donc on l'accepte...
En fait, on peut dire que l'on rejette H_0 au seuil de 4% (i.e. on risque de se tromper en rejettant H_0 dans 4% des cas), car :
`0.96>normal_icdf(0,0.46,0.98)=0.944724498891` ou encore
`normal_cdf(0,0.46,0.96)=0.981553967548>0.98`

— **Exercice 3**
On a administré un somnifère A à 50 personnes choisies au hasard et on a

observé une moyenne de sommeil de 8h22 avec un écart-type de 0h24.
On a administré un somnifère B à 100 personnes choisies au hasard et on a observé une moyenne de sommeil de 7h15 avec un écart-type de 0h30.
Ces deux somnifères ont-ils une efficacité signicativement différente ? de combien ?

Réponse

Soit X_1 la variable aléatoire égale au nombre de minutes de sommeil lorsque l'on a pris le somnifère A et soit X_2 la variable aléatoire égale au nombre de minutes de sommeil lorsque l'on a pris le somnifère B.
On note $\overline{X}_1$ (resp $\overline{X}_2$) la variable aléatoire égale à la moyenne du nombre de minutes de sommeil pour des échantillons de taille 50 (resp 100) lorsque l'on a pris le somnifère A (resp le somnifère B).
Au vu de l'échantillon d'effectif $n_1 = 50$, $\overline{X}_1$ a comme moyenne m_1 de 8h22 soit de 502 minutes et comme écart-type σ_1.
On a donc :
$m_1 = 8 * 60 + 22 = 502$ et,
$\sigma_1 \simeq 24/\sqrt{49}$.
Au vu de l'échantillon d'effectif $n_2 = 100$, $\overline{X}_2$ a comme moyenne m_2 de 7h15 soit de 435 minutes et comme écart-type σ_2.
On a donc :
$m_2 = 7 * 60 + 15 = 435$ et,
$\sigma_2 \simeq 30/\sqrt{99}$.
On en déduit que $\overline{X}_1 - \overline{X}_2$ suit approximativement une loi normale $\mathcal{N}(\mu, \sigma)$ avec comme écart-type :
$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{24^2/49 + 30^2/99} = 4.56574321789 \simeq 4.566$.
On cherche un intervalle de confiance pour μ au seuil de 5% et au seuil de 1%. On sait que l'on a :
$Proba(|\overline{X}_1 - \overline{X}_2| \leq \mu + 1.96\sigma) = 0.95$
$Proba(|\overline{X}_1 - \overline{X}_2| \leq \mu + 2.58\sigma) = 0.99$
$\overline{X}_1 - \overline{X}_2$ a comme valeur $m_1 - m_2 = 67$ donc,
$Proba(67 - 1.96 * 4.566 \leq \mu \leq 67 + 1.96 * 4.566) = 0.95$ et,
$Proba(67 - 2.58 * 4.566 \leq \mu \leq 67 + 2.58 * 4.566) = 0.99$
On a donc :
- Un intervalle de confiance pour μ au seuil de 5% est :
l'intervalle $[58; 76]$ (car $58 \simeq 67 - 1.96 * 4.566$ et $76 \simeq 67 + 1.96 * 4.566$),
- Un intervalle de confiance pour μ au seuil de 1% est :
l'intervalle $[55; 79]$ (car $55 \simeq 67 - 2.58 * 4.566$ et $79 \simeq 67 + 2.58 * 4.566$).

Avec `Xcas` on tape :

```
normal_icdf(67,sqrt(24^2/49+30^2/99),0.975)
```

On obtient :

```
75.9486922697 ≃ 76
```

On tape :

```
normal_icdf(67,sqrt(24^2/49+30^2/99),0.025)
```

On obtient :

```
58.0513077303 ≃ 58
```

On tape :

```
normal_icdf(67,sqrt(24^2/49+30^2/99),0.995)
```

On obtient :

`78.7605751731 ≃ 79`
On tape :
`normal_icdf(67,sqrt(24^2/49+30^2/99),0.005)`
On obtient :
`55.239424827 ≃ 55`
Donc :
- μ est dans l'intervalle $[58; 76]$ avec un risque d'erreur de 5% et ,
- μ est dans l'intervalle $[55; 79]$ avec un risque d'erreur de 1%.
Le somnifère A allonge la durée du sommeil d'environ 67mn=1h07 par rapport au somnifère B, avec une incertidude de (76-58)/2=9mn (resp (79-55)/2=12mn) pour un seuil de 5% (resp 1%).
Remarque : dans l'exercice précédent on a considéré deux groupes de patients indépendants. On aurait pu faire l'experience sur un même groupe (après un certain temps). On aurait eu affaire alors a des échantillons non indépendants mais appariés (pour un exemple voir l'exercice suivant).

— **Exercice 4** : Échantillons appariés
On a fait faire une double correction de 30 copies par deux examinateurs A et B afin de comparer leur notation. Les copies sont numerotées de 0 à 29.
On a obtenu pour A :
`13,15,12,15,8,7,11,10,9,13,3,18,17,5,9,10,`
`11,14,12,10,9,8,13,6,8,16,14,11,12,10`
On a obtenu pour B :
`12,13,12,15,7,5,12,10,8,13,4,17,16,4,9,11,10,`
`13,13,9,10,7,14,8,7,15,13,10,13,10`
Réponse
On tape :
`A:=[13,15,12,15,8,7,11,10,9,13,3,18,17,5,9,`
`10,11,14,12,10,9,8,13,6,8,16,14,11,12,10]`
`mean(A)`
On obtient :
`329/30≃ 10.9666666667`
On tape :
`stddev(A)`
On obtient :
`sqrt(10769/900) ≃ 3.45912641509`
On tape :
`B:=[12,17,11,16,7,7,10,10,8,13,1,18,16,4,8,`
`10,10,14,11,10,8,8,12,6,7,16,13,11,12,9]`
On tape :
`mean(B)`
On obtient :
`21/2=10.5`
On tape :
`stddev(B)`
On obtient :
`sqrt(508/45) ≃ 3.35989417823`

Les deux examinateurs n'ont pas obtenus la même moyenne et la différence

des moyennes entre l'examinateur A et l'examinateur B est $m = 0.3$.
On tape : A-B
On obtient :
[1,-2,1,-1,1,0,1,0,1,0,2,0,1,1,1,0,1,0,1,0,1,0,1,
0,1,0,1,0,0,1] ou encore on définit le caractère différence D avec effectifs :
D:=([-2,-1,0,1,2],[1,1,12,15,1])
On tape :
mean(A-B)
ou mean(D)
On obtient :
m=7/15=0.466666666667
On tape :
stddev(A-B)
ou stddev(D)
On obtient :
s=sqrt(131/225) $\simeq$ 0.763034876151
Rappel : Si au sein d'une production répartie selon une loi normale de moyenne μ et d'écart type σ, on prélève au hasard un échantillon de petite taille n, la variable aléatoire $\bar{X}$ qui à un échantillon de taille n fait correspondre la moyenne de cet échantillon suit une loi normale de moyenne μ et d'écart type $\sigma/\sqrt{n}$.
Cette loi n'est pas connue car σ est inconnu. On ne peut pas estimer σ par $s\sqrt{\frac{n}{n-1}}$ (où s désigne l'écart type de l'échantillon) car n est petit et cela ne fournit pas une bonne estimation de σ.
La variable $\mathtt{T} = \sqrt{\mathtt{n}-1}(\frac{\bar{\mathtt{X}}-\mu}{\mathtt{S}})$ (où $\mathtt{S}^2 = \frac{1}{\mathtt{n}}\sum_1^{\mathtt{n}}(\mathtt{X_j} - \bar{\mathtt{X}})^2$ et $\mathtt{X_j}$ la variable aléatoire qui au $\mathtt{j}^{\mathtt{ime}}$ tirage associe son résultat) suit une loi qui ne dépend pas de σ : cette loi est la loi de Student à $\mathtt{n}-1$ degrés de liberté.

On suppose que le caractère différence D est distribué selon une loi normale $\mathcal{N}(\mu, \sigma(D))$ (il faudrait le vérifié avec le test du χ^2).
On pose comme hypothèse $H_0 : \mu = 0$ et comme hypothèse alternative $H_1 : \mu \neq 0$.
On considère la variable $\mathtt{T} = \sqrt{\mathtt{n}-1}(\frac{\bar{\mathtt{D}}-0}{\mathtt{S}})$, T a pour valeur :
$t = \sqrt{n-1} * \frac{m-0}{s}$
On tape :
t:=sqrt(29)*7/15/sqrt(131/225) donc $t \simeq= 3.29352823645$
Au seuil de 5% et avec 29 degrées de liberté on lit dans la table de Student, la valeur critique h que t ne doit pas dépasser.
On trouve $h = 2.05$.
Ou bien on tape avec Xcas :
h:=student_icdf(29,0.975)
On obtient :
h=2.04522964213 $\simeq$ 2.05
Puisque $t = 3.29352823645 > h$ on conclut qu'au seuil de 5%, les 2 moyennes sont significativement différentes.
Peut-on dire que D suit une loi normale $\mathcal{N}(0.5, 1)$?

On tape :
`e1:=30*normal_cdf(0.5,1,-infinity,-0.5)`
On trouve :
`4.75965761794`
On tape :
`e2:=30*normal_cdf(0.5,1,-0.5,0.5)`
On trouve :
`10.2403423821`
On tape :
`d2:=((e1-2)^2)/e1+((e2-12)^2)/e2+((e2-15)^2)/e2+((e1-1)^2)/e1`
On trouve :
`7.084447157`
On a 4 classes donc 3 degrés de liberté, on tape :
`chisquare_icdf(3,0.95)=7.81472790325`
On a `d2<7.81472790325` donc, au seuil de 5% on ne peut pas rejeter l'hypothèse que `D` suit une loi normale $\mathcal{N}(0.5, 1)$

4.6 Jet d'un dé et test du χ^2

4.6.1 On jette un dé 90 fois

On a obtenu :
1 a été obtenu 11 fois,
2 a été obtenu 16 fois,
3 a été obtenu 17 fois,
4 a été obtenu 22 fois,
5 a été obtenu 14 fois,
6 a été obtenu 10 fois.
Peut-on admettre au vu de cette expérience que le dé est régulier ?
Il y a 6 classes et le degré de liberté est égal à 5 puisque l'effectif de la dernière classe est imposé lorsque l'on a l'effectif des 5 premières. Pour chaque classe l'effectif théorique de l'échantillon est 90*1/6=15 (chaque face ayant une probabilité théorique égale à 1/6 de sortir si le dé est équilibré).
On calcule l'écart quadradique réduit, c'est la valeur de :
$\chi^2 = \sum_{j=1}^{6} \frac{(X_j - 90/6)^2}{90/6}$ pour l'échantillon considéré.
On obtient ici :
$\frac{1}{15}((11-15)^2+(16-15)^2+(17-15)^2+(22-15)^2+(14-15)^2+(10-15)^2) = 6.4$
Dans une table du χ^2 on lit qu'au seuil 0.05 et pour un degré de liberté 5 la valeur limite de χ^2 est égale à 11.1.
Avec `Xcas`, on tape :
`chisquare_icdf(5,0.95)`
On obtient :
`11.0704976935` $\simeq$ `11.1`.
Or on a $6.4 < 11,1$, donc au seuil 0.05, on ne rejette pas l'hypothèse : "le dé est régulier" car si on dit que le dé n'est pas régulier on se trompe dans plus de 5% des cas.

4.6.2 On jette un dé 180 fois

On a obtenu :
1 a été obtenu 22 fois,
2 a été obtenu 32 fois,
3 a été obtenu 34 fois,
4 a été obtenu 44 fois,
5 a été obtenu 28 fois,
6 a été obtenu 20 fois.
Peut-on admettre au vu de cette expérience que le dé est régulier ?
Par rapport à l'exercice précédent on a doublé le nombre de lancers et on a aussi doublé les effectifs de chaque classe.
On calcule l'écart quadradique réduit, c'est la valeur de :
$\chi^2 = \sum_{j=1}^{6} \frac{(X_j - 180/6)^2}{180/6}$ pour l'échantillon considéré.
On obtient ici :
$\frac{2}{15}((11-15)^2+(16-15)^2+(17-15)^2+(22-15)^2+(14-15)^2+(10-15)^2) = 2*6.4 = 12.8$
Dans une table du χ^2 on lit qu'au seuil 0.05 et pour un degré de liberté 5 la valeur limite de χ^2 est 11.1.
Avec `Xcas`, on tape :
```
chisquare_icdf(5,0.95)
```
On obtient :
`11.0704976935`.
Or on a 12.8>11,1, donc on rejette l'hypothèse : "le dé est régulier" au seuil de 5% ce qui veut dire "le dé n'est pas régulier" dans plus de 95% des cas.

4.6.3 Intervalle de confiance

Donner un intervalle de confiance du nombre d'apparitions de la face 4, au seuil de 5% lorsqu'on lance le dé 90 fois de suite.
On pose $n = 90$ et $p = 1/6$.
On considère la variable aléatoire X égale au nombre de fois que le 4 est obtenu. X suit une loi binomiale $\mathcal{B}(90, 1/6)$, de moyenne $\mu = 90 * 1/6 = 15$ et d'écart-type $\sigma = \sqrt{npq} = \sqrt{91 * 1/6 * 5/6} = \sqrt{12.5} = 3.53553390593$.
En effet, on obtient un 4 avec la probabilité théorique de $p = 1/6$ et on obtient une face différente de 4 avec la probabilité théorique de $q = 5/6$.
On sait, d'apres la loi binomiale, que dans un échantillon d'effectif n, on a une probabilité de $C_n^k p^k q^{n-k}$ d'avoir k apparitions d'un caractère de probabilité p (ici le caractère est d'obtenir un 4 et, on a $p = 1/6$, $q = 5/6$ et $n = 90$).
On peut approcher la loi binomiale par la loi normale de même moyenne et de même écart-type car $n > 30$ et on a :
$\mu = np = 15$ et $\sigma = \sqrt{npq} = \sqrt{12.5}$
$Prob(|X - \mu|/\sigma < 1.96) = 0.95$ donc
$Prob(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 0.95$ donc
$Prob(8.07035354438 < X < 21.9296464556) = 0.95$
Avec `Xcas` on tape :
```
normal_icdf(15,sqrt(12.5),0.975)
```
On obtient : `21.9295191217`

`normal_icdf(15,sqrt(12.5),0.025)`
On obtient : `8.07048087825`
Donc si $n = 90$, l'effectif des différentes classes (en particulier l'effectif de la classe 4) devraient être dans l'intervalle $[8; 22]$, au seuil de 0.05 c'est à dire avec un risque d'erreur de 5%.
Remarque
De même si $n = 180$ $\mu = 30$ et $\sigma = \sqrt{180 * 5/36 = 5}$ donc :
$Prob(20.2 = 30 - 1.96 * 5 < k < 30 + 1.96 * 5 = 39.8) = 0.95$
Avec `Xcas` on tape :
`normal_icdf(30,5,0.975)`
On obtient : `39.7998199227`
`normal_icdf(30,5,0.025)`
On obtient : `20.2001800773`
Donc si $n = 180$, les effectifs des différentes classes sont dans l'intervalle $[20; 40]$, au seuil de 0.05 c'est à dire avec un risque d'erreur de 5%.

4.7 Simulation de la loi uniforme sur [0 ;1]

Si X suit une loi uniforme sur [0 ;1], X a pour espérance $1/2$ et pour écart-type $\sqrt{1/12} \simeq 0.288675134595$.
En effet :
$E(X) = \int_0^1 xdx = 1/2$ et
$E(X^2) = \int_0^1 x^2dx = 1/3$ et donc
$\sigma(X) = \sqrt{1/3 - (1/2)^2} = \sqrt{1/12}$
Dans `Xcas` la fonction `rand()` renvoie, de façon équirpartie, un nombre aléatoire entre 0 et 2^{32} et `rand(0,1)` ou `rand(0..1)()` renvoie, de façon équirpartie, un nombre aléatoire entre 0 et 1 : on remarquera que r :=`rand(0..1)` définit une fonction `r` et que `r()` renvoie alors de façon équirpartie, un nombre aléatoire entre 0 et 1.
Exercice :
Simuler dans la colonne `A` du tableur, le tirage de 100 nombres aléatoires.
Calculer la moyenne dans `A100` et l'écart type dans `A101` de la série obtenue.
Refaire la même chose dans les colonnes `B, C, D, E`.
Refaire la même chose avec les 100 lignes ainsi crées.
Comparer avec les valeurs théoriques.
Réponse :
De 0 à 99 et sur 5 colonnes les cellules sont remplies aléatoirement : les cellules `A0:E99` contiennent `rand(0,1)`.
La ligne 100 (`A100:E100`) contient les moyennes des lignes de 0 à 99 pour chacune des colonnes `A..E`.
La ligne 101 (`A101:E101`) contient les écarts-types des lignes de 0 à 99 pour chacune des colonnes `A..E`.
La colonne `F` (`F0:F99`) va servir à faire la moyenne des colonnes de `A` à `E` pour chacune des lignes de 0 à 99.
La colonne `G` (`G0:G99`) va servir à mettre les écarts-types des colonnes de `A` à `E` pour chacune des lignes de 0 à 99.
On remplit ensuite `F100, F101, G100, G101` :

`F100=[mean(A100:E100),mean(F0:F99),mean(A0:E99)],F100` est la moyenne de la ligne 100 (moyenne des moyennes de 5 échantillons d'effectif 100), suivi de la moyenne de la colonne F (moyenne des moyennes de 100 échantillons d'effectif 5), suivi de la moyenne totale (moyenne d'un échantillon d'effectif 500). Évidemment ces 3 moyennes sont les mêmes !
`F101=[mean(A101:E101),stddev(F0:F99)]`
`F101`est la moyenne de la ligne 101 (moyenne des écarts-types de 5 échantillons d'effectif 100) suivi de l'écart-type de la colonne `F` (écart-type des moyennes de 100 échantillons d'effectif 5).
`G100=[stddev(A100:E100),mean(G0:G99)]`
`G100` l'écart-type de la ligne 100 (écart-type des moyennes de 5 échantillons d'effectif 100) suivi de la moyenne de la colonne `G` (moyenne des l'écarts-types de 100 échantillons d'effectif 5).
`G101=[stddev(A101:E101),stddev(G0:G99),stddev(A0:E99)]`
`G101` est l'écart-type de la ligne 101 (l'écart-type des l'écarts-types de 5 échantillons d'effectif 100), suivi de l'écart-type de la colonne G (l'écart-type des l'écarts-types de 100 échantillons d'effectif 5), suivi de l'écart-type total (l'écart-type d'un échantillon d'effectif 500).

Pour `n=500`, on trouve par exemple :
`m=mean(A0:E99)=0.484342422505` et
`s=stddev(A0:E99)=0.285946471987`
Ici, on est parti d'une loi connue : la loi uniforme sur [0 ;1] de moyenne $\mu = 0.5$ et d'écart-type $\sigma = \sqrt{1/12} \simeq 0.288675134595$.
Dans la pratique on ne connait ni μ ni σ.
D'après la théorie, si on considère tous les échantillons de taille n, la variable aléatoire :
$\bar{X} = \frac{(X_1+...+X_n)}{n}$ suit approximativement une loi normale $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$ lorsque n est grand et la variable aléatoire :
$S^2 = \frac{(X_1-\bar{X})^2+...+(X_n-\bar{X})^2}{n}$ a pour moyenne $\frac{n-1}{n}\sigma^2$
Pour la loi uniforme on a :
la moyenne de la série des moyennes des échantillons de taille n est égale à 0.5 ,
l'écart-type de la série des moyennes des échantillons de taille n est $\sqrt{\frac{1}{12n}}$,
la moyenne de la série des écarts-types des échantillons de taille n est $\sqrt{\frac{n-1}{12n}}$,
l'écart-type $\sigma(S^2)$ de la série des écarts-types des échantillons de taille n est plus petit que $K/\sqrt{n}$ où K est une constante qui ne dépend que de la loi.
Au vu d'un échantillon d'effectif 500 (`n=500`), de moyenne m, et d'écart-type s, on convient de dire que la moyenne empirique :
`m=mean((A0):(E99))=0.484342422505` (m est la valeur observée de $\bar{X}$) est l'approximation de la moyenne μ et que l'écart-type empirique :
`s=stddev(A0:E99)=0.285946471987` (s^2 est la valeur observée de S^2) est l'approximation de la moyenne de la série des écarts-types des échantillons de taille $n = 500$ et on a :
$\sqrt{E(S^2)} = \sigma * \sqrt{499/500} \simeq s$.
Lorsque l'on ne connait pas σ on en calcule une valeur approchée à partir de l'écart type d'un échantillon de grande taille ici 500 :
on a : s=`0.285946471987`
On calcule la valeur théorique estimée :

$\sigma - est = s * \sqrt{n/(n-1)}$
On tape et on obtient :
=0.285946471987*sqrt500/499=0.286232848095
au lieu de $\sigma = 0.288386314978$
De plus la théorie nous dit que la distribution des moyennes des échantillons d'effectif 500 suit sensiblement une loi normale de moyenne μ et d'écart-type :
$\frac{\sigma}{\sqrt{500}} \simeq \frac{s}{\sqrt{499}}$ =0.0128007221147.
Ceci nous permet de dire qu'au seuil de 5%, on a :
$|m - \mu| < 1.96 * 0.0127879149858 = 0.0250894153448$ soit :
$0.45925300716 = m - \frac{s}{\sqrt{499}} < \mu < m + \frac{s}{\sqrt{499}} = 0.50943183785.$
D'où un intervalle de confiance au seuil de 5% pour μ de :
$[0.4592; 0.5095]$.
Voici les résultats des lignes 100 et 101 :
- ligne 100 est la valeur de la moyenne de 5 échantillons d'effectif 100, on trouve :
0.466489640726, 0.487896819143, 0.499799806252,
0.453281438346, 0.514244408058.
Ces 5 moyennes ont pour moyenne la moyenne totale :
mean(A100:E100) =m =0.484342422505
et ces 5 moyennes ont pour écart-type :
stddev(A100:E100)=0.022041777341 $\simeq \sigma/\sqrt{100}$
- ligne 101 est la valeur de l'écart-type de 5 échantillons d'effectif 100, on trouve :
0.264640095911, 0.302416108249, 0.299622396086,
0.276154743049, 0.280843050885
ces 5 écarts-types ont pour moyenne :
mean(A101:E101) =0.284735278836
valeur approchée de $\sigma * \sqrt{99/100} \simeq$0.287228132327
et pour écart-type :
stddev(A101:E101)=0.0143305924398.
Dans la colonne F on fait la moyenne des lignes ce qui correspond à 100 échantillons de 5 tirages (n=5).
Ces 100 moyennes ont pour moyenne la moyenne totale :
mean(F0:F99)=0.484342422505
et es 100 moyennes ont pour écart-type :
stddev(F0:F99)=0.112665383246
valeur approchée de $\sigma/\sqrt{5} \simeq$ 0.129099444874.
Dans la colonne G on fait l'écart-type des lignes ce qui correspond à 100 échantillons de 5 tirages (n=5).
Ces 100 écarts-types ont pour moyenne :
mean(G0:G99)=0.252572046948
valeur approchée de $\sigma * \sqrt{4/5} \simeq$ 0.258198889747
et pour écart-type :
stddev(G0:G99)=0.0726584981978

Observations :
- Comment évoluent les moyennes :
les valeurs de la ligne 100 des moyennes de chaque colonne A..F c'est à dire la moyenne de 100 observations est assez proche de la valeur attendue 0.5, alors que

la colonne F moyenne des lignes c'est à dire de 5 observations est loin de la valeur attendue 0.5. On voit bien que l'écart-type des moyennes d'un échantillon de taille n dépend de n et que plus n est grand plus cet écart-type diminue. Les écarts-types de ces 2 séries ne sont donc pas les mêmes : d'après la théorie, l'écart-type des moyennes d'un échantillon de taille n est :
$\sigma/\sqrt{n}$ si σ est l'écart-type de la population toute entière qui est pour la loi uniforme de : $\sqrt{1/12} = 0.288675134595$.
De façon expérimentale on a :
- écart type des 5 moyennes correspondant à 5 échantillons de taille 100 :
`0.022041777341`
On calcule la valeur théorique :
`sqrt(1/12)/10=0.0288675134595`
- écart type des 100 moyennes correspondant à 100 échantillons de taille 5 :
`0.112665383246`
On calcule la valeur théorique :
`sqrt(1/12)/sqrt(5)=0.129099444874`
- Comment évoluent les écarts-types :
d'après la théorie, la moyenne des écarts-types d'un échantillon de taille n est :
$\sigma * \sqrt{n - 1/n}$ si σ est l'écart-type de la population toute entière qui est pour la loi uniforme de : $\sqrt{1/12} = 0.288675134595$.
De façon expérimentale on a :
- moyenne des 5 écarts-types correspondant à 5 échantillons de taille 100 :
`0.284735278836`
On calcule la valeur théorique :
`sqrt(1/12)*sqrt(99/100)=0.298337151271`
- moyenne des 100 écarts-types correspondant à 100 échantillons de taille 5 :
`0.252572046948`
On calcule la valeur théorique :
`sqrt(1/12)*sqrt(4/5)=0.258198889747`
Lorsque l'on ne connait pas σ on en calcule une valeur approchée à partir de l'écart type d'un échantillon de grande taille ici 500 :
on a : s=`0.285946471987`
La valeur théorique estimée σest de σ :
$\sigma est = s * \sqrt{n/(n-1)}$
On tape :
`0.285946471987*sqrt(500/499)`
On obtient :
`0.286232848095` (au lieu de σ=`0.288386314978`
Quant aux écarts-types des écarts-types des échantillons de taille n, on voit qu'il sont d'autant plus petit que n est grand : c'est pourquoi on peut approcher l'écart type par l'écart-type d'un seul échantillon de grande taille.

Chapitre 5

Exemples d'exercies utilisant le tableur

5.1 $PGCD$

5.1.1 L'algorithme d'Euclide

Voici la description de cet algorithme pour obtenir le pgcd de deux entiers a et b :
on effectue des divisions euclidiennes successives :

$$\begin{aligned} a = &\quad b \times q_1 + r_1 \quad 0 \leq r_1 < b \\ b = &\quad r_1 \times q_2 + r_2 \quad 0 \leq r_2 < r_1 \\ r_1 = &\quad r_2 \times q_3 + r_3 \quad 0 \leq r_3 < r_2 \\ &..... \end{aligned}$$

Après un nombre fini d'étapes (au plus b), il existe un entier n tel que : $r_n = 0$.
on a alors :
$PGCD(a, b) = PGCD(b, r_1) = ...PGCD(r_{n-1}, r_n) = PGCD(r_{n-1}, 0)$
donc
$PGCD(a, b) = r_{n-1}$

5.1.2 Une mise en œuvre simple

On effectue une commande en appuyant sur `enter`, si on appuie à nouveau sur `enter` `Xcas` exécute à nouveau la même commande puisque cette commande reste inscrite dans la ligne des commandes.
Les différentes commandes et réponses sont numérotées à partir de `0`.
Les réponses sont alors `ans(0), ans(1)` etc ... La dernière réponse peut aussi être désignée par `ans()` ou par `ans(-1)`, et l'avant dernière réponse peut être désignée par `ans(-2)` etc...
Pour calculer $PGCD(78, 56)$, on tape :
`78 enter` puis `56 enter` puis,
`irem(ans(-2),ans()) enter`
On obtient `22` (car 22 est le reste de la division de 78 par 56)
puis `enter enter enter` etc...et cela nous permet d'obtenir la suite des restes.

5.1.3 La suite des restes avec le tableur

Voici une mise en œuvre de l'algorithme d'Euclide avec un tableur.
À l'aide d'un tableur que l'on obtient avec le raccourci clavier `Alt+t`, on écrit la suite des restes. Si l'on veut déterminer le PGCD(78,56), on définit la suite des restes dans la colonne `A` par :
On met `78` dans `A0`
On met `56` dans `A1`
On met `=irem(A0,A1)` dans `A2`
puis on utilisee le menu du tableur `Edit` puis `Remplir` et `Copier vers le bas`, lorsque `A2` est en surbrillance et l'on a ainsi la suite des restes des divisions successives dans la colonne `A`.
Le dernier reste non nul `2` est donc le pgcd de 78 et de 56.

5.2 Identité de Bézout

L'algorithme d'Euclide permet aussi de trouver un couple u, v vérifiant :
$au + bv = PGCD(a, b)$

5.2.1 Avec le tableur

On pose $a = r_0$ et $b = r_1$, et on définit, dans la colonne `A`, "la suites des restes" r_n : pour $n \geq 2$, $r_{n-2} = q_n r_{n-1} + r_n$ avec $r_n < r_{n-1}$.
Puis on définit, dans les colonnes `B` et `C`, deux suites u_n et v_n de façon qu'à chaque étape on ait : $r_n = u_n a + v_n b$ pour $n \geq 0$.
Puisque $r_0 = a$ on a : $u_0 = 1$ et $v_0 = 0$
Puisque $r_1 = b$ on a : $u_1 = 0$ et $v_1 = 1$
Puisque $r_n = r_{n-2} - q_n r_{n-1}$ pour $n \geq 2$, u_n et v_n vérifient la même relation de récurrence pour $n \geq 2$ à savoir :
$u_n = u_{n-2} - q_n u_{n-1}$ et $v_n = v_{n-2} - q_n v_{n-1}$ avec
$q_n =$ quotient entier de r_{n-2} par r_{n-1} pour $n \geq 2$.
On a au début :
`A0` vaut $r_0 = a$ et `A1` vaut $r_1 = b$,
`B0` vaut $u_0 = 1$ et `B1` vaut $u_1 = 0$ puisque $a = 1 * a + 0 * b$,
`C0` vaut $v_0 = 0$ et `C1` vaut $v_1 = 1$ puisque $b = 0 * a + 1 * b$.
Dans la colonne `A` on veut avoir la suite des restes et dans les colonnes `B` et `C` les suites `u` et `v` qui seront les coefficients de `a` et `b` pour avoir sur chaque ligne `n` du tableur : `An=aBn+bCn`.
Pour cela on a besoin de la suite des quotients $q_n =$ (quotient entier de r_{n-2} par r_{n-1} pour $n \geq 2$) que l'on met dans la colonne `D` (`D2=iquo(A0,A1)` puis on tape sur `remplir` et `vers le bas`, lorsque `D2` est en surbrillance).
Les lignes l_n des colonnes `A, B, C` vérifient la même relation de récurrence :
$l_n = l_{n-2} - q_n l_{n-1}$
On definit donc :
`A0` par `a`
`A1` par `b`
`A2` par `=irem(A0,A1)`
puis on tape sur `remplir` et `vers le bas`, lorsque `A2` est en surbrillance et on

obtient la suite des restes des divisions successives.
`B0` par `1`
`B1` par `0`
`B2` par `=B0-D2*B1`
puis, on tape sur `remplir` et `vers le bas`, lorsque `B2` est en surbrillance et on obtiendra la suite des "u" lorsqu'on aura rempli la colonne `D`.
`C0` par `0`
`C1` par `1`
`C2` par `=C0-D2*C1`
puis, on tape sur `remplir` et `vers le bas`, lorsque `C2` est en surbrillance et on obtiendra la suite des "v" lorsqu'on aura rempli la colonne `D`.
`D0` par `0`
`D1` par `0`
`D2` par `=iquo(A0,A1)`
(en fait `D0` et `D1` ne servent pas)
puis, on tape sur `remplir` et `vers le bas`, lorsque `D2` est en surbrillance, pour avoir la suite des quotients des divisions successives.
Le tableur va alors afficher les valeurs de ces différentes suites et sur la ligne du dernier reste non nul on pourra lire les coefficients de l'identité de Bézout.
Pour l'exemple $a = 78$ et $b = 56$, on trouve :
- sur la ligne `6` : `A6=0` et,
- sur la ligne `5` : `A5=2`, `B5=-5`, `C5=7` (et `D5=1`),
ce qui signifie que :
`2=-5*78+7*56`

5.2.2 Les pas de Louis

Louis marche sur la bordure du trottoir, comptant les fois où une de ses semelles chevauche un joint entre deux blocs (si son talon est tangent il chevauche le joint, mais si la pointe est tangente il ne chevauche pas).
Le trottoir comporte 150 blocs d'un mètre (il n'y a pas de joint au debut ni à la fin du trottoir).
Son pas est de 61 cm et ses semelles mesurent 21 cm et les joints n'ont pas d'épaisseur.
Combien de fois Louis marche-t-il sur un joint ?

Réponse
Il y a 149 joints.
Il faut tout d'abord voir qu'au p-ième pas les semelles de Louis sont entre $(p * 61)$ cm et $(p * 61 + 21)$ cm. Ses semelles couperont le n-ième joint si :
$100 * n = p_n * 61 + r_n$ avec $0 \leq r_n < 21$
On va donc dans un premier temps écrire la suite des restes r_n à l'aide du tableur.
On definit donc :
`A0` par `0`
`A1` par `=A0+1`
puis, on tape sur `remplir` et `vers le bas`, lorsque `A1` est en surbrillance, pour avoir la suite des entiers 0,1, etc...
`B0` par `0`

`B` par `1=irem(100*A1,61)`
puis on tape sur `remplir` et `vers le bas`, lorsque `B1` est en surbrillance, pour avoir la suite des restes r_n.
Bien sûr il n'est pas facile de compter les restes inférieurs à 21.
On va donc définir une colonne qui fera ce comptage.
On definit donc :
`C0` par `0`
`C1` par `=ifte(B1<21,C0+1,C0)`
La réponse au problème est donc la valeur de `C149` et on trouve `51`.
Louis marche donc 51 fois sur un joint.

On peut aussi utiliser la commande `count` (ou la commande `count_inf`) qui compte les éléments du tableur pour lesquels une fonction booléenne est vraie :
On tape, par exemple dans `D0` :
`count(x->x<21,B1:B149)`
ou on tape dans `D0` :
`count_inf(21,B1:B149)`
On prend en compte les cellules `B1:B149`, car il n'y a pas de joints au début, ni à la fin.
On obtient dans `D0` : `51`.

On peut aussi écrire un petit programme dont voici l'algorithme :

```
louis nbloc
0 -> k
pour n de 1 a nbloc-1 faire
si (n*100 mod 61 <21) alors k+1 -> k
fsi
fpour
afficher k
```

qui se traduit en langage `Xcas` par :

```
louis(nbloc):={
k:=0;
for (n:=1;n<nbloc:n++){
if (irem(n*100,61)<21) k:= k+1;
}
return k;
}
```

Puis, on tape `louis(150)`
et on obtient `51`.

5.3 Accélération de convergence vers $\pi^2/6$

On considère la série de terme général $1/n^2$, et on souhaite déterminer une valeur approchée de sa somme pour n allant de 1 à l'infini.
On rappelle que la valeur exacte de cette somme est égale à $\pi^2/6$: pour cela on développe en séries de Fourier la fonction f qui est 2π-périodique et est égale à x^2 sur $[-\pi,\ \pi]$.

On a :
$f(x) = x^2 = \dfrac{\pi^2}{3} + 4 * \sum_{n=1}^{\infty} \dfrac{(-1)^n}{n^2} \cos nx$ pour $x \in [-\pi,\ \pi]$
Pour $x = \pi$ on obtient : $\sum_{n=1}^{\infty} \dfrac{1}{n^2} = \dfrac{\pi^2}{6}$.

1. Déterminer à l'aide de `Xcas` une valeur approchée de $\pi^2/6$ puis de :

$$\sum_{j=1}^{n} \frac{1}{j^2}$$

pour $n = 10$, $n = 100$ et $n = 1000$.
Définir et observer dans le tableur la suite récurrente :
$\mathtt{u_1} = 1$
$\mathtt{u_2} = 1 + 1/4$
$\mathtt{u_n} = \mathtt{u_{n-1}} + 1/\mathtt{n}^2$ pour $n > 1$.
Réponse :
On cherche une valeur approchée de $\frac{\pi^2}{6}$ et on tape :
`evalf(pi^2/6)`
On obtient :
`1.64493406685`
On tape :
`evalf(sum(1/j^2,j,1,10))`
On obtient :
`1.54976773117`
On tape :
`evalf(sum(1/j^2,j,1,100))`
On obtient :
`1.63498390018`
On tape :
`evalf(sum(1/j^2,j,1,1000))`
On obtient :
`1.64393456668`
Puis on ouvre un tableur (avec `Alt+t`) et on tape :
dans `A0` : `0`
dans `A1` : `=A0+1`
puis, on tape sur `remplir` et `vers le bas`, lorsque `A1` est en surbrillance, pour avoir la suite des entiers 0,1, etc...
On tape :
dans `B0` : `0`
dans `B1` : `=B0+1/A1^2`
puis, on tape sur `remplir` et `vers le bas`, lorsque `B1` est en surbrillance, pour avoir la suite des valeurs exactes de u_n.
On tape :
dans `C0` : `=evalf(B0)`
puis, on tape sur `remplir` et `vers le bas`, lorsque `C0` est en surbrillance, pour avoir la suite des valeurs approchées de u_n.

2. Comparer ces valeurs à la valeur approchée de $\pi^2/6$: donner le nombre de décimales exactes pour les 3 valeurs de n.
3. On souhaite accélérer la convergence des sommes partielles. On va donc déterminer un encadrement de $\sum_{j=n+1}^{\infty} 1/j^2$ à l'aide de l'intégrale de la fonction $1/x^2$. Montrer que :

$$\int_{n+1}^{\infty} \frac{1}{x^2}\,dx \leq \sum_{j=n+1}^{\infty} \frac{1}{j^2} \leq \int_{n}^{\infty} \frac{1}{x^2}\,dx$$

4. En déduire que :

$$0 \leq \left(\sum_{j=1}^{n} \frac{1}{j^2} + \frac{1}{n}\right) - \frac{\pi^2}{6} \leq \frac{1}{n(n+1)}$$

5. Calculer $\frac{1}{n} + \sum_{j=1}^{n} \frac{1}{j^2}$ pour $n = 10, n = 100$ et $n = 1000$ et déterminer le nombre de décimales exactes pour ces 3 valeurs de n, justifier ce nombre de décimales pour $n = 10$, $n = 100$ et $n = 1000$.
   Avec le tableur, il suffira d'ajouter $\frac{1}{n}$ à la colonne où se trouve u_n.

6. Montrer que pour $k \geq 2$ on a :

$$\frac{1}{k} - \frac{1}{k+1} + \frac{1}{2k(k+1)} - \frac{1}{2(k+1)(k+2)} < \frac{1}{k^2} < \frac{1}{k} - \frac{1}{k+1} + \frac{1}{2(k-1)k} - \frac{1}{2k(k+1)}$$

   En déduire que :
   $w_n = u_n + 1/(n+1) + 1/2/(n+1)/(n+2)$ converge vers $\pi^2/6$ et que l'on a :
   $0 < \pi^2/6 - w_n < 1/n^3$

Réponse :
On tape :
```
normal(1/k^2-1/k+1/(k+1)-1/2/k/(k+1)+1/2/(k+1)/(k+2))
```
On obtient :
```
2/(k^4+3*k^3+2*k^2)
```
On tape :
```
normal(1/(k^2)-1/k+1/(k+1)-1/2/k/(k-1)+1/2/(k+1)/k)
```
On obtient :
```
-1/(k^4-k^2)
```

5.4 $\ln(2)$

Il s'agit ici d'obtenir un encadrement de $\ln(2)$ en utilisant des méthodes numériques approchées de calcul de l'intégrale : $I = \displaystyle\int_0^1 \frac{1}{1+x}\,dx$

1. Tracer le graphe de la fonction f définie par $f(x) = 1/(1+x)$ entre 0 et 1, la corde qui relie les points d'abscisses 0 et 1 et la tangente au point d'abscisse $\frac{1}{2}$.

Réponse :
On change l'initialisation de la fenêtre graphique avec le bouton rouge `geo`.
On prend :
`X-=WX-=-0.1 X+=WX+=1.1`
`Y-=WY-=0.4 Y+=WY+=1.1`
On tape :
```
f(x):=1/(1+x)
G:=plotfunc(1/(1+x),x)
segment(i*f(0),1+i*f(1))
tangent(G,1/2)
```

2. On subdivise l'intervalle $[0,1]$ en n parties et on encadre l'intégrale par la méthode des rectangles inférieurs et supérieurs.
Montrez que : $u_n < I < v_n$ avec pour $n \geq 1$:
$u_n = 1/(n+1) + 1/(n+2) + ... + 1/(n+n)$ et
$v_n = 1/n + 1/(n+1) + ... + 1/(n+n-1)$.
Calculer u_n et v_n pour n=10, 100, 1000.
Réponse :
Avec le tableur on écrit la suite récurrente :
$u_1 = 1/2$
$u_2 = 1/3 + 1/4$
$u_n = u_{n-1} + 1/((2n-1)*2n)$ pour $n > 1$ et
$v_1 = 1$
$v_1 = 1/2 + 1/3$
$v_n = u_n + 1/(2n)$ pour $n \geq 1$
pour cela on tape :
`0` dans `A0`
`=A0+1` dans `A1`
puis, on tape sur `remplir` et `vers le bas`, lorsque `A1` est en surbrillance, pour avoir la suite des entiers 0,1, etc...
On tape :
`0` dans `B0`
`1/2` dans `B1`
`=B1+1/((2*A2-1)*2*A2)` dans `B2`
puis, on tape sur `remplir` et `vers le bas`, lorsque `B2` est en surbrillance, pour avoir la suite des valeurs exactes de u_n.
On tape :
`=evalf(B0)` dans `C0`
puis, on tape sur `remplir` et `vers le bas`,lorsque `C0` est en surbrillance, pour avoir la suite des valeurs approchées de u_n.
On tape :
`0` dans `D0`
`=B1+1/(2*A1)` dans `D1`
puis, on tape sur `remplir` et `vers le bas`, lorsque `D1` est en surbrillance, pour avoir la suite des valeurs exactes de v_n.
On tape :
`=evalf(D0)` dans `E0`
puis, on tape sur `remplir` et `vers le bas`, lorsque `E0` est en surbrillance, pour avoir la suite des valeurs approchées de v_n.

3. On subdivise l'intervalle $[0,1]$ en n parties et on applique la méthode du point milieu et la méthode des trapèzes. On obtient ainsi deux suites w_n et t_n.
Montrez que : $w_n = \sum_{j=0}^{n-1} \frac{1}{n+j+1/2}$ et que $t_n = (u_n + v_n)/2$.
Calculer w_n et t_n pour $n = 10,\ 100,\ 1000$.
Réponse :
Le rectangle de base $[j/n; (j+1)/n]$ au point milieu d'abscisse $(2j+1)/2n$ a pour surface :
$\frac{1}{n} * \frac{1}{1+(2j+1)/2n} = \frac{2}{2n+2j+1} = \frac{1}{n+j+1/2}$
Le trapèze de base $[j/n; (j+1)/n]$ a pour surface :
$\frac{1}{2n} * (\frac{1}{1+j/n} + \frac{1}{1+(j+1)/n}) = \frac{1}{2}(\frac{1}{n+j} + \frac{1}{n+j+1})$

Avec le tableur, on écrit les suites récurrentes :
$w_0 = 2/3$
$w_1 = 2/5 + 2/7$
$w_n = w_{n-1} + 1/((2n-1/2) * (2n-3/2)(n-1/2))$ et
$t_0 = 3/4$
$t_1 = 1/3 + 3/8$
$t_n = (u_n + v_n)/2$ pour cela on tape :
`0` dans `F0`
`2/3` dans `F1`
`=F1+1/(2*A2-1/2)+1/(2*A2-3/2)-1/(A2-1/2)` dans`F2`
puis, on tape sur `remplir` et `vers le bas`, lorsque `F2` est en surbrillance, pour avoir la suite des valeurs exactes de w_n.
`G0` = `evalf(F0)`
puis, on tape sur `remplir` et `vers le bas`, lorsque `G0` est en surbrillance, pour avoir la suite des valeurs approchées de w_n.
On tape :
`0` dans `H0`
`=(B1+D1)/2` dans `H1`
puis, on tape sur `remplir` et `vers le bas`, lorsque `H1` est en surbrillance, pour avoir la suite des valeurs exactes de t_n.
`=evalf(H0)` dans `I0`
puis, on tape sur `remplir` et `vers le bas`, lorsque `I0` est en surbrillance, pour avoir la suite des valeurs approchées de t_n.
On trouve pour $n = 10$:
$u_{10} = 155685007/232792560 \simeq 0.668771403175$
$v_{10} = 33464927/46558512 \simeq 0.718771403175$
$w_{10} = 358143560536/516924483075 \simeq 0.69283536041$
$t_{10} = 161504821/232792560 \simeq 0.693771403175$

4. Vérifiez que la fonction f est convexe en montrant que $f'' > 0$.
On admettra les propriétés suivantes des fonctions convexes :
   — le segment reliant $(x_1, f(x_1))$ à $(x_2, f(x_2)$ $(x_1 < x_2)$ est au-dessus de la courbe représentative de f entre x_1 et x_2.
   — la tangente en un point de la courbe est en-dessous de la courbe représentative de f.

5. En comparant graphiquement des aires, montrez que :

$$t_n \geq \int_0^1 f(x)\,dx$$

.

6. On considère une subdivision $[\frac{j}{n}, \frac{j+1}{n}]$. Tracer sur cet intervalle la courbe de f et la tangente de f au point milieu de l'intervalle. Par une comparaison graphique d'aires (aire sous la tangente=aire du rectangle) montrez que :

$$w_n \leq \int_0^1 f(x)\,dx$$

7. Indiquez le nombre de décimales exactes obtenues lorsque l'on approche $\ln(2)$ à l'aide de u_n, v_n, w_n, t_n pour $n = 10,\ n = 100,\ n = 1000$.
   Montrez qu'en général u_n et v_n donnent un encadrement d'ordre $1/n$ de $\ln(2)$, et que w_n et t_n donnent un encadrement d'ordre $1/n^2$ de $\ln(2)$.

5.5 L'algorithme de Héron avec Xcas

5.5.1 Les fonctions de `Xcas` utilisées

Voici les fonctions de `Xcas` qui vous seront utiles dans ce TP.
`ans()` renvoie la dernière réponse obtenue.
`numer` renvoie le numérateur de la fraction `n/d`, ainsi `numer(n/d)` vaut `n` et `denom` renvoie le dénominateur de la fraction `n/d`, ainsi `denom(n/d)` vaut `d`.
`evalf(a)` évalue `a` à l'aide d'un nombre flottant comportant 12 chiffres significatifs (sauf si on a changé ce nombre dans la configuration du `cas` obtenu avec le bouton rouge `cas`).
`iquo(a,b)` calcule le quotient entier `q` de `a` par `b` (`a=b*q+r` avec `0<= r <b`).
`f(x):=exprx` où `exprx` représente une expression de `x` permettant de définir la fonction `f`.
`plotseq(f(x),x=a,n)` permet de visualiser les `n` premiers termes de la suite définie par $u_0 = a$ et $u_{n+1} = f(u_n)$ pour $n \geq 0$ (il faut penser à changer la fenêtre de visualisation bouton rouge `geo`).

5.5.2 Le problème : valeur approchée de $\sqrt{7}$

On considère la suite récurrente définie par :
$u_0 = 3$ et $u_{n+1} = \frac{1}{2}(u_n + \frac{7}{u_n})$ pour $n \geq 0$.
0/ Définir la fonction f pour que $u_{n+1} = f(u_n)$ pour $n \geq 0$.
1/ Calculer les 5 premiers termes de la suite u et donner une valeur approchée de u_5. On pourra utiliser `f(ans())` qui applique `f` à la dernière réponse.
2/ Visualiser les cinq premiers termes de la suite u et trouver une bonne fenêtre de visualisation.
3/ Ouvrir un tableur avec le raccourci clavier `Alt+t` et mettre :
dans la colonne `A` les valeurs de n,
dans la colonne `B` les valeurs de u_n,
dans la colonne `C` les valeurs du numérateur c_n de u_n,

dans la colonne D les valeurs du dénominateur d_n de u_n,
dans la colonne E les valeurs du quotient exact de $c_n * 10^{12}$ par d_n,
dans la colonne F les valeurs approchées de u_n données par evalf.
Que représente la colonne E ?
Observez les différentes colonnes et notez vos observations.
4/ On veut montrer que la suite u_n est convergente.
- montrer que la suite u est définie et que $u_n > \sqrt{7}$ pour tout $n \geq 0$,
- en déduire que le signe de $u_{n+1} - u_n$ est indépendant de n,
- montrer que la suite u est décroissante et converge vers $\sqrt{7}$,
5/ On pose $e_n = u_n - \sqrt{7}$. Montrer que $\sqrt{7} > 5/2$ et en déduire que $e_0 < 0.5$.
Montrer que $e_n = \dfrac{e_{n-1}^2}{2 * u_{n-1}}$ pour tout $n \geq 1$.
En déduire que $e_n < 5 * (\dfrac{e_{n-1}}{5})^2 < 5 * (\dfrac{e_0}{5})^{2^n} < 5 * (0.1)^{2^n}$ pour tout $n \geq 1$.
Quelle erreur fait-on lorsqu'on prend u_5 comme valeur approchée de $\sqrt{7}$? (on montrera que $e_5 = (u_5^2 - 7)/(u_5 + \sqrt{7}) < (u_5^2 - 7)/5$).
Donner les 20 premières décimales de $\sqrt{7}$.

5.5.3 Correction

0/ On tape : `f(x):=1/2*(x+7/x)`
1/ Dans une ligne de commande on tape : `3` puis, `f(ans())` cela va afficher la valeur exacte de u_1, on valide à nouveau `f(ans())` on obtient u_2 etc...
On obtient :
$u_1 = \dfrac{8}{3)}$,
$u_2 = \dfrac{127}{48)}$,
$u_3 = \dfrac{32257}{12192}$,
$u_4 = \dfrac{2081028097}{786554688}$.
$u_5 = \dfrac{8661355881006882817}{3273684811110137472}$ Quand on a la valeur exacte de u_5 pour avoir la valeur approchée on sélectionne la réponse avec la souris, puis utilise `evalf` du menu `Calc`▶`Numérique`.
On obtient :
`2.64575131106`
2/ On tape : `plotseq(f(x),x=3,5)` avec par exemple comme fenêtre :
$[2.5; 3] \times [2.5; 2.7]$
3/ Dans `A0` on tape `0` et dans `A1` on tape $=$ `A0` $+$ `1`, puis dans le menu `Edit`, on choisit `Remplir` et `Copier vers le bas` quand `A1` est en surbrillance.
Dans `B0` on tape 3 et dans `B1` on tape `=f(B0)`, puis dans le menu `Edit`, on choisit `Remplir` et `Copier vers le bas`, quand `B1` est en surbrillance.
Dans `C0` on tape `=numer(B0)`, puis dans le menu `Edit`, on choisit `Remplir` et `Copier vers le bas`, quand `C0` est en surbrillance.
Dans `D0` on tape `=denom(B0)`, puis dans le menu `Edit`, on choisit `Remplir` et `Copier vers le bas`, quand `D0` est en surbrillance.
Dans `E0` on tape `=iquo(C0*10^12,D0)`, puis dans le menu `Edit`, on choisit `Remplir` et `Copier vers le bas`, quand `E0` est en surbrillance.

Dans F0 on tape =evalf(B0), puis dans le menu Edit, on choisit Remplir et Copier vers le bas, quand F0 est en surbrillance.
La colonne E contient la partie entière de $10^{12} * u_n$. On est donc capable de voir la partie entière de u_n et de ses 12 premières décimales. En remplacant 12 par 22 on verra donc les 22 premières décimales de u_n.
4/ $u_0 > 0$ donc u_1 est défini et $u_1 > 0$ et par récurrence u_n est défini et $u_n > 0$ pour tout $n \geq 0$.
Si u_n converge vers l, l vérifie : $2 * l = l + 7/l$ donc $l^2 = 7$.
Donc si la limite existe ce ne peut être que $\sqrt{7}$ puisque $u_n > 0$ pour tout $n \geq 0$.
On a f croissante sur $[\sqrt{7} + \infty[$ et $f(\sqrt{7}) = \sqrt{7}$ donc puisque :
$u_0 = 3 > \sqrt{7}$ par récurrence $u_n > \sqrt{7}$ pour tout $n \geq 0$.
On a f croissante sur $[\sqrt{7} + \infty[$ et $u_n > \sqrt{7}$ pour tout $n \geq 0$ donc puisque :
$u_0 = 3 > u_1 = 8/3$ on obtient par récurrence $u_n > u_{n+1}$ pour tout $n \geq 0$.
Ou bien on forme la différence :
$u_{n+1} - u_n = 1/2(u_n - u_{n-1} - 7(u_n - u_{n-1})/(u_n * u_{n-1})) =$
$(u_n - u_{n-1}) * 1/2 * (u_n * u_{n-1} - 7)/(u_n * u_{n-1})$
$(u_n * u_{n-1} - 7) > 0$ puisque $u_n > \sqrt{7}$ pour tout $n \geq 0$ et $u_n * u_{n-1} > 0$ donc
$u_{n+1} - u_n$ a le même signe que $u_n - u_{n-1}$ pour tout $n \geq 1$
$u_1 - u_0 < 0$ donc la suite u est décroissante.
u est décroissante et minorée par $\sqrt{7}$ donc u est convergente.
On a vu que si u est convergente, sa limite est $\sqrt{7}$, donc u converge vers $\sqrt{7}$.
5/ $5^2 = 25 < 7 * 2^2 = 28$ donc $5/2 < \sqrt{7} < u_0 = 3$ et $e_0 = u_0 - \sqrt{7} < 1/2$
$e_n = u_n - \sqrt{7} = 1/2 * (u_{n-1}^2 + 7 - 2 * \sqrt{7} * u_{n-1})/u_{n-1}) = \frac{e_{n-1}^2}{2 * u_{n-1}}$.
On sait que $5/2 < \sqrt{7} < u_n$ donc $e_n < 5 * (\frac{e_{n-1}}{5})^2$ et par récurrence :
$e_n < 5 * (\frac{e_0}{5})^{2^n} < 5 * (\frac{1}{10})^{2^n}$ pour tout $n \geq 1$.
Donc $e_5 < 5 * (\frac{1}{10})^{32} = 5 * 10^{-32}$ est une majoration de l'erreur à priori.
Comme $e_5 = (u_5^2 - 7)/(u_5 + \sqrt{7}) < (u_5^2 - 7)/5$, on fait le calcul de $(u_5^2 - 7)/5$ et on trouve $e_5 < 2 * 10^{-38}$.
Puisque :
e_4<evalf(((2081028097/786554688)^2-7)/5)$< 4 * 10^{-19}$,
pour avoir un encadrement de $\sqrt{7}$ à 10^{-20} près, il faut calculer u_5 avec 21 décimales.
On obtient : $u_5 - 10^{-25} < u_5 - e_5 = \sqrt{7} < u_5$ et
$2.645751311064590590501 < u_5 < 2.645751311064590590502$ donc
$2.645751311064590590500 < \sqrt{7} < 2.645751311064590590502$

5.6 Suites adjacentes et convergence de $\sum_{k=0}^{n} \frac{(-1)^k}{2k+1}$

5.6.1 Les fonctions de Xcas utilisées

Voici les fonctions de Xcas qui vous seront utiles dans ce TP.
normal(expr) renvoie l'expression expr simplifiée.
evalf(a) évalue a à l'aide d'un nombre comportant 12 chiffres significatifs, sauf si on a changé Chiffres dans la configuration du cas (cas en rouge).
Si exprx est une expression de x,

`f(x):=exprx` définit la fonction `f` de variable `x` ($x \overset{f}{\longmapsto} exprx$),
`subst(exprx,x,2)` remplace `x` par 2 dans `exprx` et
`derive(exprx,x)` calcule la dérivée de `exprx` par rapport à `x`.

5.6.2 u et v sont deux suites ajacentes de limite $\frac{\pi}{4}$

I/ On considère les suites u et v définies par :
$u_0 = 1 - \frac{1}{3}$ et $u_n = u_{n-1} + \frac{1}{4n+1} - \frac{1}{4n+3}$ pour $n \geq 1$.
$v_n = u_n + \frac{1}{4n+3}$ pour $n \geq 0$.
1/ Calculer les 6 premiers termes de la suite u et donner une valeur approchée de u_6 (on pourra utiliser un tableur que l'on obtient avec le raccourci clavier `Alt+t`.
2/ Calculer les 6 premiers termes de la suite v et donner une valeur approchée de v_6 (on pourra utiliser le tableur).
3/ Montrer que les suites u et v sont adjacentes.

II/ On considère la suite de fonctions f_n de $[0, \frac{\pi}{2}[$ dand $\mathbb{R}$ définie par :
$f_0(x) = x - \tan(x)$ et $f_n(x) = f_{n-1}(x) - \frac{(-1)^n}{2n+1}\tan(x)^{2n+1}$ pour $n \geq 1$.
0/ Calculer $f_n(0)$ pour tout $n \geq 0$.
1/ Ouvrir le tableur et faites afficher dans les colonnes :
les valeurs de n, les valeurs de $f_n(x)$ et les valeurs de $f'_n(x)$,
En observant ces colonnes, déterminer une expression simplifiée de la dérivée de f_n. Prouver votre conjecture.
2/ En déduire que pour $p \geq 0$:
- la fonction f_{2p+1} est croissante sur $[0, \frac{\pi}{2}[$.
- la fonction f_{2p} est décroissante sur $[0, \frac{\pi}{2}[$.
3/ Calculer pour $p \geq 0$, $f_{2p}(\frac{\pi}{4})$ et $f_{2p+1}(\frac{\pi}{4})$ en fonction de u_p et de v_p (on pourra utiliser le tableur pour déterminer la formule, puis on fera une démonstration).

III/ 1/ Montrer que la limite de u et de v est égale à $\frac{\pi}{4}$ (on étudiera le signe de $f_{2p}(\frac{\pi}{4})$ et de $f_{2p+1}(\frac{\pi}{4})$).
2/ Donner un encadrement de $\frac{\pi}{4}$.
Quelle erreur fait-on lorsqu'on prend $4 * u_6$ comme valeur approchée de π ?
3/ Trouver une valeur de n pour que $4 * u_n$ et $4 * v_n$ donnent un encadrement de π de diamètre inférieur à 10^{-3}.

5.6.3 Correction

I/ Dans `A0` on tape 0 et dans `A1` on tape `=A0+1` puis dans le menu `Edit` du tableur, on choisit `Remplir` et `Copier vers le bas`, quand `A1` est en surbrillance.
Dans `B0` on tape :
`2/3` (ou `=evalf(2/3)` et,
dans `B1` on tape :
`=B0+1/(4*A1+1)-1/(4*A1+3)`
puis dans le menu `Edit`, on choisit `Remplir` et `Copier vers le bas`, quand `B1` est en surbrillance.
Dans `C0` on tape :

`=B0+1/(4*A0+3)`
puis dans le menu `Edit` du tableur, on choisit `Remplir` et `Copier vers le bas`, quand `C0` est en surbrillance.
La suite u est croissante car $u_n - u_{n-1} = \frac{1}{4n+1} - \frac{1}{4n+3} > 0$ pour $n \geq 1$.
La suite v est décroissante car $v_n - v_{n-1} = \frac{1}{4n+1} - \frac{1}{4n-1} < 0$ pour $n \geq 1$.
Donc $v_n - u_n = 1/(4 * n + 3)$ tend vers 0 quand n tend vers l'infini.
Les suites u et v sont donc adjacentes.

II/
0/ $f_n(0) = 0$ pour tout $n \geq 0$.
1/ Dans `A0` on tape `0` et dans `A1` on tape `=A0+1` puis, dans le menu `Edit` du tableur, on choisit `Remplir` et `Copier vers le bas`, quand `A1` est en surbrillance.
Dans `D0` on tape `=x-tan(x)` et
dans `D1` on tape `=D0-(-1)^(A1)*tan(x)^(2*A1+1)/(2*A1+1)` puis, dans le menu `Edit` du tableur, on choisit `Remplir` et `Copier vers le bas`, quand `B1` est en surbrillance.
Dans `E0` on tape `=normal(derive(D0,x))` puis, dans le menu `Edit`, on choisit `Remplir` et `Copier vers le bas`, quand `E0` est en surbrillance.
On remarque que l'on a $f'_n(x) = (-1)^{n+1} * \tan(x)^{2n+2}$.
On le montre en faisant le calcul :
$f_n(x) = x - \tan(x) + \frac{1}{3}\tan(x)^3 + ... - \frac{(-1)^n}{2n+1}\tan(x)^{2n+1}$
$f'_n(x) = 1 - (1 + \tan(x)^2) * (1 - \tan(x)^2 + ... + (-1)^n \tan(x)^{2n})$
On reconnait la somme d'une série géométrique de raison : $-\tan(x)^2$.
$f'_n(x) = 1 - (1 + \tan(x)^2) * \frac{1 - (-\tan(x)^2)^{n+1}}{1 + \tan(x)^2} = (-1)^{n+1}\tan(x)^{2n+2}$
2/ La fonction f_{2p} a une dérivée negative pour tout $p \geq 0$, donc f_{2p} est décroissante. En particulier $f_{2p}(0) = 0 > f_{2p}(\pi/4)$.
La fonction f_{2p+1} a une dérivée positive pour tout $p \geq 0$, donc f_{2p+1} est croissante. En particulier $f_{2p+1}(0) = 0 < f_{2p+1}(\pi/4)$.
3/ Dans `F0` on tape `=subst(D0,x,pi/4)` puis, dans le menu `Edit`, on choisit bouton `Remplir` et `Copier vers le bas`, quand `E0` est en surbrillance.
$f_{2p}(\pi/4) = \pi/4 - v_p$ et $f_{2p+1}(\pi/4) = \pi/4 - u_p$.

III/ 1/ On a donc pour tout $p \geq 0$:
$f_{2p+1}(\pi/4) = \pi/4 - u_p > 0$ et $f_{2p}(\pi/4) = \pi/4 - v_p < 0$
les suites u et v sont donc adjacentes et convergentes vers $\pi/4$.
2/ Donc pour tout $p \geq 0$: $u_p < \pi/4 < v_p$.
On a donc $4 * u_6 < \pi < 4 * v_6 = 4 * u_6 + 4/27$.
3/ Pour avoir :
$4 * u_n < \pi < 4 * v_n = 4 * u_n + 4/4 * n + 3 \leq 4 * u_n + 10^{-3}$,
il faut prendre comme n un entier qui vérifie $4 * n + 3 \geq 4000$, soit $n \geq 1000$.
Donc u_{1000} et v_{1000} donnent un encadrement de π de diamètre $\leq 10^{-3}$.

5.6.4 Accélération de convergence

On considère toujours les suites u et v définies par :
$u_0 = 1 - \frac{1}{3}$ et $u_n = u_{n-1} + \frac{1}{4n+1} - \frac{1}{4n+3}$ pour $n \geq 1$.

$v_n = u_n + \dfrac{1}{4n+3}$ pour $n \geq 0$.
1/ Montrer que, pour $p > 0$,

$$\frac{1}{2}(\frac{1}{4p+1} - \frac{1}{4p+5}) < \frac{1}{4p+1} - \frac{1}{4p+3} < \frac{1}{2}(\frac{1}{4p-1} - \frac{1}{4p+3})$$

2/ En déduire que, pour $n > p > 0$,

$$\frac{1}{2}(\frac{1}{4p+1} - \frac{1}{4n+5}) < u_n - u_{p-1} < \frac{1}{2}(\frac{1}{4p-1} - \frac{1}{4n+3})$$

et en déduire un encadrement de $u_n - u_p$

3/ En faisant tendre n vers $+\infty$, montrer que

$$\frac{1}{2(4p+5)} \leq \frac{\pi}{4} - u_p \leq \frac{1}{2(4p+3)}$$

4/ On pose $w_n = u_n + \dfrac{1}{2(4n+5)}$ et $t_n = u_n + \dfrac{1}{2(4n+3)}$.
En utilisant le tableur, montrer que w_n et t_n sont deux suites adjacentes qui convergent vers $\dfrac{\pi}{4}$ plus rapidement que u_n et v_n. Trouver une valeur de n pour que $4w_n$ et $4t_n$ donnent un encadrement de π de diamètre inférieur à 10^{-3}.

Correction et prolongement
$u_{10} = \displaystyle\sum_{k=0}^{10} \frac{1}{4k+1} - \frac{1}{4k+3} = \sum_{k=0}^{10} \frac{2}{(4k+1)(4k+3)}$
On encadre les termes de la série :
$\displaystyle\sum_{k=0}^{10} \frac{2}{(4k+1)(4k+3)}$
$\frac{1}{2}(\frac{1}{4k+1} - \frac{1}{4k+5}) < \frac{2}{(4k+1)(4k+3)} < \frac{1}{2}(\frac{1}{4k-1} - \frac{1}{4k+3})$ donc
$\frac{1}{2}(\frac{1}{4p+1} - \frac{1}{4n+5}) < u_n - u_{p-1} < \frac{1}{2}(\frac{1}{4p-1} - \frac{1}{4n+3})$.
On a :
$w_{10} = u_{10} + \dfrac{1}{2(4k+5)} = \dfrac{1}{2} + \displaystyle\sum_{k=0}^{10} \frac{4}{(4k+1)(4k+3)(4k+5)}$
On peut continuer le même processus en encadrant les termes de la série :
$\displaystyle\sum_{k=0}^{10} \frac{4}{(4k+1)(4k+3)(4k+5)}$.
On a :
$\dfrac{4}{(4k+1)(4k+5)(4k+9)} < \dfrac{4}{(4k+1)(4k+3)(4k+5)} < \dfrac{4}{(4k-3)(4k+1)(4k+5)}$
donc
$\frac{1}{2}(\frac{1}{(4k+1)(4k+5)} - \frac{1}{(4k+5)(4k+9)}) < \frac{4}{(4k+1)(4k+3)(4k+5)} <$
$\frac{1}{2}(\frac{1}{(4k-3)(4k+1)} - \frac{1}{(4k+1)(4k+5)})$
On a donc comme précédemment :
$\frac{1}{2}(\frac{1}{(4p+1)(4p+5)} - \frac{1}{(4n+5)(4n+9)}) < w_n - w_{p-1} <$

$\frac{1}{2}(\frac{1}{(4p-3)(4p+1)} - \frac{1}{(4n+1)(4n+5)})$

On pose donc :

$$s_{10} = w_{10} + \frac{1}{2(4k+5)(4k+9)}$$

$$s_{10} = \frac{3}{5} + \sum_{k=0}^{10} \frac{24}{(4k+1)(4k+3)(4k+5)(4k+9)}$$

On peut continuer le même processus en encadrant les termes de la série :

$$\sum_{k=0}^{10} \frac{24}{(4k+1)(4k+3)(4k+5)(4k+9)}$$

et on pose :

$$r_{10} = s_{10} + \frac{2}{(4k+5)(4k+9)*(4k+13)}$$

$$r_{10} = \frac{29}{45} + \sum_{k=0}^{10} \frac{240}{(4k+1)(4k+3)(4k+5)(4k+9)(4k+13)}$$

On tape pour avoir la valeur approchée de $\frac{\pi}{4}$:

```
evalf(pi/4)
```

On obtient :

```
0.785398163397
```

On tape pour avoir la valeur approchée de u_{10} :

```
sum(2.0/((4*k+1)*(4*k+3)),k,0,10)
```

On obtient :

```
0.774040381616
```

On tape pour avoir la valeur approchée de w_{10} :

```
sum(2/((4*k+1)*(4*k+3)),k,0,10)+1.0/(2*45)
```

On obtient :

```
0.785151492727
```

Ou on tape pour avoir la valeur approchée de w_{10} :

```
sum(4/((4*k+1)*(4*k+3)*(4*k+5)),k,0,10)+0.5
```

On obtient :

```
0.785151492727
```

On tape pour avoir la valeur approchée de s_{10} :

```
sum(4/((4*k+1)*(4*k+3)*(4*k+5)),k,0,10)+ 1.0/2+1/(2*45*49)
```

On obtient :

```
0.785378250097
```

Ou on tape pour avoir la valeur approchée de s_{10} :

```
sum(24/((4*k+1)*(4*k+3)*(4*k+5)*(4*k+9)),k,0,10)+0.6
```

On obtient :

```
0.785378250097
```

On tape pour avoir la valeur approchée de r_{10} :

```
sum(24/((4*k+1)*(4*k+3)*(4*k+5)*(4*k+9)),k,0,10)+
0.6+2/(45*49*53)
```

On obtient :

```
0.78539536386
```

Ou on tape pour avoir la valeur approchée de r_{10} :

```
sum(240/((4*k+1)*(4*k+3)*(4*k+5)*(4*k+9)*(4*k+13)),k,0,10)
+29.0/45
```

On obtient :
`0.78539536386`
On a pour la dernière somme 5 décimales exactes de $\pi/4$:
$|l - z_{10}| < 2(1/(41 * 45 * 49) - 1/(45 * 49 * 53)) < 5.1 * 10^{-6}$

5.6.5 Comparaison avec une intégrale

Soit pour $t \in [0; +\infty[$ la fonction $h(t) = \dfrac{1}{4t+1} - \dfrac{1}{4t+3}$.
1/ Calculer pour $k \in \mathbb{N}$

$$\int_k^{k+1} h(t)\ dt$$

2/ Montrer que pour $k \in \mathbb{N}$

$$\int_k^{k+1} h(t)\ dt < h(k) < \int_{k-1}^{k} h(t)\ dt$$

3/ En déduire que

$$\frac{1}{4}(\ln(\frac{4n+5}{4n+7}) - \ln(\frac{4p+5}{4p+7})) < \sum_{k=p+1}^{n} h(k) < \frac{1}{4}(\ln(\frac{4n+1}{4n+3}) - \ln(\frac{4p+1}{4p+3}))$$

4/ En faisant tendre n vers $+\infty$, montrer que

$$\frac{1}{4}\ln(\frac{4p+7}{4p+5}) \leq \frac{\pi}{4} - u_p \leq \frac{1}{4}\ln(\frac{4p+3}{4p+1})$$

5/ On pose $s_n = u_n + \dfrac{1}{4}\ln(\dfrac{4p+7}{4p+5})$ et $z_n = u_n + \dfrac{1}{4}\ln(\dfrac{4p+3}{4p+1})$. En utilisant le tableur, montrer que s_n et z_n sont deux suites adjacentes qui convergent vers $\dfrac{\pi}{4}$ plus rapidement que u_n et v_n. Trouver une valeur de n pour que $4s_n$ et $4z_n$ donnent un encadrement de π de diamètre inférieur à 10^{-3}.

5.6.6 Prolongement avec la formule d'Euler-Mac Laurin

La formule d'Euler-Mac Laurin à l'ordre 4 On suppose la fonction g suffisamment dérivable et on pose $I_0 = \int_0^1 g(t)\ dt$. On va intégrer cette intégrale par partie, pour cela on définit le polynôme $P_1(x)$ tel que : $P_1'(x) = 1$ et $\int_0^1 P_1(t)\ dt = 0$ donc $P_1(x) = x - \frac{1}{2}$.
$P_1(x)$ est une fonction impaire en $(x - \frac{1}{2})$ et on a $P_1(1) = -P_1(0) = \frac{1}{2}$.
On a :

$$I_0 = \int_0^1 g(t)\ P_1'(t)\ dt = P_1(1)g(1) - P_1(0)g(0) - \int_0^1 g'(t)\ P_1(t)\ dt$$

On continue, en définissant $P_2(x)$ tel que :
$P_2'(x) = P_1(x)$ et $\int_0^1 P_2(t)\ dt = 0$ donc $P_2(x) = \frac{1}{2}(x - \frac{1}{2})^2 - \frac{1}{24}$.
$P_2(x)$ est une fonction paire en $(x - \frac{1}{2})$ et on a $P_2(1) = P_2(0) = \frac{1}{12}$.
On a :

$$\int_0^1 g'(t)\ P_2'(t)\ dt = P_2(1)g'(1) - P_2(0)g'(0) - \int_0^1 g''(t)\ P_2(t)\ dt$$

$$I_0 = \frac{g(1)+g(0)}{2} - \frac{g'(1)-g'(0)}{12} + \int_0^1 g''(t)\ P_2(t)\ dt$$

On recommence en définissant :
$P_3'(x) = P_2(x)$ et $\int_0^1 P_3(t)\ dt = 0$.
On a : $P_3(x) = 1/6(x-\frac{1}{2})^3 - \frac{1}{2}4(x-\frac{1}{2})$ en effet $P_3(x)$ est une primitive de $P_2(x)$ qui est impaire en $(x-\frac{1}{2})$.
On en déduit que :
$P_3(1) = -P_3(0)$ et puisque $\int_0^1 P_2(t) = 0$ on a :
$P_3(1) = P_3(0)$ donc $P_3(1) = -P_3(0) = 0$.
Ceci se généralise dans la suite pour $k > 0$ on a :
$P_{2*k+1}(1) = P_{2*k+1}(0) = 0$.
On a :

$$\int_0^1 g''(t)\ P_3'(t)\ dt = P_3(1)g''(1) - P_3(0)g''(0) - \int_0^1 g'''(t)\ P_3(t)\ dt$$

$$I_0 = \frac{g(1)+g(0)}{2} - \frac{g'(1)-g'(0)}{12} - \int_0^1 g'''(t)\ P_3(t)\ dt$$

En définissant $P_4(x)$ tel que :
$P_4'(x) = P_3(x)$ et $\int_0^1 P_4(t)\ dt = 0$, $P_4(x)$ est donc une fonction paire en $(x-\frac{1}{2})$ et on a $P_4(x) = \frac{1}{2}4(x-\frac{1}{2})^4 - 1/48(x-\frac{1}{2})^2 + 7/5760$.
On a $P_4(1) = P_4(0) = -1/720$.
Donc la formule d'Euler-MacLaurin à l'ordre 4 est :

$$\int_0^1 g(t)\ dt = \frac{g(1)+g(0)}{2} - \frac{g'(1)-g'(0)}{12} + \frac{g'''(1)-g'''(0)}{720} + \int_0^1 g^{(4)}(t)\ P_4(t)\ dt$$

puisque $P_5(0) = P_5(1) = 0$, on a $\int_0^1 g^{(4)}(t)\ P_4(t)\ dt = -\int_0^1 g^{(5)}(t)\ P_5(t)\ dt$.
Donc on a aussi :

$$\int_0^1 g(t)\ dt = \frac{g(1)+g(0)}{2} - \frac{g'(1)-g'(0)}{12} + \frac{g'''(1)-g'''(0)}{720} - \int_0^1 g^{(5)}(t)\ P_5(t)\ dt$$

Remarque
On peut montrer que le polynôme $P_k(x) = B_k(x)/k!$ où $B_k(x)$ est le k-ième polynôme de Bernoulli, il est solution de :
$B_k(x+1) - B_k(x) = kx^{k-1}$, et il vérifie :
$\frac{te^{xt}}{e^t-1} = \sum_{n=0}^{\infty} B_n(x)\frac{t^n}{n!}$.
Le k-ième nombre de Bernouilli (c'est la commande `bernoulli(k)` de `Xcas`) est la valeur du k-ième polynôme de Bernoulli en zéro.
La formule d'Euler-Mac Laurin plus générale à l'ordre r. On a :
$\int_p^q f(t)dt = \frac{1}{2}(f(p)+f(q)) + \sum_{n=p+1}^{q-1} f(n) + \sum_{k=2}^{r} \frac{(-1)^k}{k!} bernouilli(k)(f^{(k-1)}(p) - f^{(k-1)}(q)) + \int_0^1 \sum_{n=p+1}^{q} (-1)^r f(u+n-1)\frac{B_r(u)}{r!} du$

Comparaison de $\sum_{k=p+1}^{n} f(k)$ avec $\int_p^n f(t)\ dt$

On a $\sum_{k=p+1}^{n} f(k) - \int_p^n f(t)\ dt = \sum_{k=p+1}^{n} \int_{k-1}^{k} (f(k) - f(t))\ dt$.
On se ramène à l'intervalle $[0;1]$ en posant $u+k-1 = t$:

$\int_{k-1}^{k}(f(k)-f(t))\,dt=\int_0^1(f(k)-f(u+k-1))\,du$
puis on applique la formule d'Euler-Mac Laurin à $g(u)=f(k)-f(u+k-1)$
$(g(1)=0,\ g(0)=f(k)-f(k-1),\ g'(u)=-f'(u+k-1)....)$:
$$\int_0^1(f(k)-f(u+k-1))\,du=\frac{f(k)-f(k-1)}{2}+\frac{f'(k)-f'(k-1)}{12}$$
$$-\frac{f'''(k)-f'''(k-1)}{720}+\int_0^1 f^{(4)}(u+k-1)\,P_4(u)\,du$$
donc
$$\sum_{k=p+1}^{n}\int_0^1(f(k)-f(u+k-1))\,du=\frac{f(n)-f(p)}{2}+\frac{f'(n)-f'(p)}{12}-\frac{f'''(n)-f'''(p)}{720}+$$
$$\int_0^1\sum_{k=p+1}^{n}f^{(4)}(u+k-1)\,P_4(u)\,du$$
On trouve :
$P_0=1$
$P_1=x-1/2$
$P_2=x^2/2-x/2+1/12$
$P_3=x^3/6-x^2/4+x/12$
$P_4=x^4/24-x^3/12+x^2/24-1/720$
$P_5=x^5/120-x^4/48+x^3/72-x/720$
$P_6=x^6/720-x^5/240+x^4/288-x^2/1440+1/30240$
car `int(x^6/720-x^5/240+x^4/288-x^2/1440,x,0,1)=-1/30240`

Application à $f(t)=\dfrac{1}{4t+1}-\dfrac{1}{4t+3}$

La série de terme général $u_n=f(n)=\dfrac{2}{(4n+1)(4n+3)}$ est convergente et,
$\sum_{k=0}^{\infty}f(k)=\frac{\pi}{4}=l$, on a :
$$f'tt)=\frac{-4}{(4t+1)^2}+\frac{4}{(4t+3)^2}$$
$$f''(t)=\frac{32}{(4t+1)^3}-\frac{32}{(4t+3)^3}$$
$$f'''(t)=\frac{-384}{(4t+1)^4}+\frac{384}{(4t+3)^4}$$
$$f^{(4)}(t)=\frac{6144}{(4t+1)^5}-\frac{6144}{(4t+3)^5}$$
$$f^{(5)}(t)=-\frac{6144*20}{(4t+1)^6}+\frac{6144*20}{(4t+3)^6}$$
On remarque au passage que $f^{(5)}$ est négative et croissante et que :
pour $k\geq p+1$ on a $0<|f^{(5)}(k)|<C/p^6$.
On calcule :
$$\int_p^n f(t)\,dt=1/4(\ln(4n+1)-\ln(4n+3)-(\ln(4p+1)-\ln(4p+3)))$$
en faisant tendre n vers $+\infty$ on a :
$\sum_{k=p+1}^{\infty}f(k)=l-\sum_{k=0}^{p}f(k)$ donc
$$l=\sum_{k=0}^{p}f(k)+\int_p^{\infty}f(t)\,dt-\frac{f(p)}{2}-\frac{f'(p)}{12}+\frac{f'''(p)}{720}-$$

$$\sum_{k=p+1}^{n} \int_0^1 f^{(5)}(u+k-1)\, P_5(u)\, du$$

$$l = \sum_{k=0}^{p} \frac{1}{4k+1} - \frac{1}{4k+3} + \frac{1}{4}(\ln(1+\frac{2}{4p+1}) - 5\frac{1}{(4p+1)(4p+3)} + \frac{1}{3}(\frac{1}{(4p+1)^2} - \frac{1}{(4p+3)^2}) - \frac{8}{15}(\frac{1}{(4p+1)^4} - \frac{1}{(4p+3)^4}) - \sum_{k=p+1}^{\infty} \int_0^1 f^{(5)}(u+k-1)\, P_5(u)\, du.$$

et on a :

$$|\sum_{k=p+1}^{\infty} \int_0^1 f^{(4)}(u+k-1)\, P_4(u)\, du| \leq \sum_{k=p+1}^{\infty} |f^{(5}(k-1)| \int_0^1 |P_5(u)|\, du| < \frac{Cste}{p^5}$$

En effet on a :

$\sum_{k=p+1}^{\infty} |f^{(5}(k-1)| < C \sum_{k=p+1}^{\infty} 1/p^6 < Cste/p^5$

car on compare le reste de la série $\sum_{k=p+1}^{\infty} 1/p^6$ avec l'intégrale $\int_p^{\infty} 1/t^6 dt = 1/(5p^5)$ On a donc le développement asymptotique z_p à l'ordre 5 de l :

$$l = \frac{\pi}{4} \simeq z_p = u_p + \frac{1}{4}\ln(1+\frac{2}{4*p+1}) - \frac{1}{(4*p+1)(4*p+3)} + \frac{1}{3}*(\frac{1}{(4*p+1)^2} - \frac{1}{(4*p+3)^2}) - \frac{8}{15}*(\frac{1}{(4*p+1)^4} - \frac{1}{(4*p+3)^4})$$

On peut faire un développement limité à l'ordre 5 :
$\frac{1}{4}\ln(1+\frac{2}{4*p+1})$ lorsque p tend vers $+\infty$.
En posant $x = 1/(4p+1)$, on tape :

```
series(1/4*log(1+2*x),x=0,5)
```

On obtient :

```
2/4*x+1/-2*x^2+8/3/4*x^3-x^4+32/5/4*x^5+
x^6*order_size(x)
```

On tape :

```
sum(2/((4*k+1)*(4.0*k+3)),k,0,10)+subst(2/4*x+1/-2*x^2+
8/3/4*x^3-x^4+32/5/4*x^5,x,1.0/41)+
subst(-1/((4*p+1)*(4*p+3))+ 1/3*(1/(4*p+1)^2-
1/(4*p+3)^2)-8/15*(1/(4*p+1)^4-1/(4*p+3)^4),p,10.0)
```

On obtient :

```
0.785398163726
```

On peut aussi faire un développement asymptotique (p tend vers $+\infty$) à l'ordre 5 de :
$1/4*\ln(1+2/(4p+1)) - 1/((4p+1)(4p+3)) + 1/3*(1/(4p+1)^2 - 1/(4p+3)^2) - 8/15*(1/(4p+1)^4 - 1/(4p+3)^4)$
En posant $x = 1/(4p+1)$, on tape :

```
series(1/4*log(1+2x)-x/(1/x+2)+1/3*(x^2-1/(1/x+2)^2)-

8/15*(x^4-1/(1/x+2)^4),x=0,5)
```

On obtient :

```
2/4*x+-3/2*x^2+4*x^3-9*x^4+16*x^5+x^6*order_size(x)
```

On tape :

```
sum(2/((4*k+1)*(4*k+3)),k,0,10)+
subst(2/4*x+-3/2*x^2+4*x^3-9*x^4+16*x^5,x=1.0/41)
```

On obtient :

```
0.785398168153
```

Sans faire un développement limité de $\frac{1}{4}\ln(1+\frac{2}{4*p+1})$ lorsque p tend vers $+\infty$
on tape et on obtient :

```
sum(2/((4*k+1)*(4*k+3)),k,0,10)+1.0/4*log(1+2/41)
0.785947393863

sum(2/((4*k+1)*(4*k+3)),k,0,10)+1/4*log(1+2/41)-1.0/(41*43)
0.785380178889

sum(2/((4*k+1)*(4*k+3)),k,0,10)+1/4*log(1+2/41)-1.0/(41*43)+
8*21/3/(41^2*43^2)
0.785398195927

sum(2/((4*k+1)*(4*k+3)),k,0,10)+1/4*log(1+2/41)-1.0/(41*43)+
8*21/3/(41^2*43^2)-8/15*(1/41^4-1/43^4)
0.785398163187
```

On rappelle que `evalf(pi/4)=0.785398163397`
On a donc pour la dernière somme 9 décimales exactes de $\pi/4$ ...

Chapitre 6

Probabilités et simulation

6.1 Rappels : les fonctions aléatoires de Xcas

6.1.1 Pour initialiser les nombres aléatoires : srand randseed RandSeed

`srand()` (ou `randseed` ou `RandSeed`) sert à initialiser la suite des nommbres aléatoires que l'on obtient avec `rand()` ou avec `randnorm()`.
`RandSeed` a toujours un argument entier, alors que `randseed` ou `srand` peut ne pas avoir d'arguments (dans ce cas le générateur aléatoire est intialisé avec l'horloge du système).
On tape :

```
srand()
```

On obtient par exemple :

```
1054990506
```

Ou on tape :

```
srand
```

On obtient par exemple :

```
1054990506
```

Ou on tape :

```
RandSeed(10549905061234)
```

On obtient par exemple :

```
10549905061234
```

6.1.2 Tirage équiréparti rand alea hasard

Tirage équiréparti sur $[0, 1, .., 2^{32}[$ **:** `rand() alea() hasard()`

`rand()` renvoie au hasard, de façon équiprobable, un nombre entier de $[0, 2^{32}[$ ($2^{32} = 4294967296$).
On tape :

```
rand()
```

ou on tape

```
hasard()
```

On obtient :

```
1804289383
```

Pour avoir, au hasard, de façon équiprobable, un nombre de $[0;1[$, on peut donc utiliser :

```
evalf(rand()/2^32)
```

On obtient :

```
0.391549611697
```

Mais il est plus simple de taper : `rand(0,1)` (voir le paragraphe suivant).

Tirage aléatoire équiréparti sur l'intervalle [a ;b[: `rand(a,b)` `hasard(a,b)` `rand(a..b)()` `hasard(a..b)()`

Si `a` et `b` sont des réels `rand(a,b)` désigne un nombre décimal aléatoire compris dans l'intervalle $[a;b[$.
Donc, `rand(a,b)` ou (`hasard(a,b)`) renvoie au hasard, et de façon équiprobable, un nombre décimal de $[a;b[$.
Pour avoir, au hasard et de façon équiprobable, un nombre décimal de $[0;1[$, on tape :

```
rand(0,1)
```

On obtient :

```
0.391549611697
```

Pour avoir, au hasard et de façon équiprobable, un nombre décimal de $[0;0.5[$, on tape :

```
rand(0,0.5)
```

On obtient :

```
0.303484437987
```

Pour avoir, au hasard et de façon équiprobable, un nombre décimal de $]-0.5;0]$, on tape :

```
rand(0,-0.5)
```

ou on tape :

```
rand(-0.5,0)
```

On obtient par exemple :

```
-0.20047219703
```

Si a et b sont des réels `rand(a..b)` ou `alea(a..b)` ou `hasard(a..b)` désigne une fonction qui est un générateur de nombres aléatoires compris dans l'intervalle $[a; b[$.
Donc, `rand(a..b)()` renvoie au hasard, et de façon équiprobable, un nombre décimal de $[a; b[$.
Pour avoir, au hasard et de façon équiprobable, un nombre décimal de $[0; 1[$, on tape :

```
rand(0..1)()
```

On obtient :

```
0.391549611697
```

Pour avoir, au hasard et de façon équiprobable, plusieurs nombres aléatoires décimaux compris dans l'intervalle $[1; 2[$, on tape :

```
r:=rand(1..2)
```

puis il suffit de taper `r()`.
On tape :

```
r()
```

On obtient :

```
1.14160255529
```

Tirage aléatoire d'entiers équirépartis sur [0,..,n[: `rand(n) alea(n) hasard(n)`

Si n est un entier relatif `rand(n)` ou `hasard(n)` renvoie au hasard, et de façon équiprobable, un entier de $[0, 1, .., n[$ (ou de $]n, ..1, 0]$ si n est negatif).
On tape :

```
rand(2)
```

Ou on tape :

```
hasard(2)
```

On obtient :

```
1
```

ou on obtient :

```
0
```

On tape :

```
rand(-2)
```

Ou on tape :

```
hasard(-2)
```

On obtient :

```
-1
```

ou on obtient :

```
0
```

On tape pour avoir un entier aléatoire entre 6 et 10, bornes comprises :

```
6+rand(11-6)
```

Ou on tape :

```
6+hasard(11-6)
```

On obtient par exemple :

```
8
```

6.1.3 Tirage aléatoire sans remise de p objets parmi n : `rand alea hasard`

`rand` a dans ce cas, soit 2, soit 3 arguments.
Si `rand` a 2 arguments : les arguments sont un entier p et une liste L alors `rand(p,L)` renvoie, au hasard, `p` éléments de la liste `L`.
Si `rand` a 3 arguments : les arguments sont trois entiers `p,min,max` alors `rand(p,min,max)` renvoie, au hasard, `p` entiers de `[min,..,max]` On tape :

```
rand(3,["r","r","r","r","v","v","v"])
```

On obtient :

```
["r","r","v"]
```

On tape :

```
rand(2,1,10)
```

On obtient :

```
[3,7]
```

On tape :

```
rand(2,4,10)
```

On obtient :

```
[5,7]
```

6.1.4 Tirage aléatoire selon la loi binomiale négative

6.1.5 Tirage aléatoire avec remise de n objets parmi k

Choisir au hasard n nombres selon la loi multinomiale de probabilité P. Cela veut dire qu'on effectue un tirage avec remise de `n` objets parmi `k=size(P)` objets. Pour `j=0..k-1`, l'objet `j` a la probabilité `P[j]` dêtre tiré (on doit avoir pour `sum(P)=1`). On tape :

```
randmult(n,P):={
  local k,j,l,r,X,L;
  k:=size(P);
  X:=cumsum(P);
  si X[k-1]!=1 alors return "erreur"; fsi;
    L:=[0$(j=1..k)];
  pour j de 1 jusque n faire
    r:=alea(0,1);
    //afficher(r);
    l:=0;
    tantque r>X[l] faire
      l:=l+1;
    ftantque;
  L[l]:=L[l]+1
  fpour;
  return L;
}:;
```

On tape :
`randmult(5,[1/2,1/3,1/6])`
On obtient par exemple :
`[3,1,0]`
ou bien, on utilise la commande `Xcas : randvector` avec `multinomial` comme paramètre On tape :
`randvector(5,multinomial,[1/2,1/3,1/6])`
On obtient par exemple :
`[3,1,1]`
Si on effectue un tirage avec remise de `n` objets de la liste `C`. Si `k=size(C)`, l'objet `C[j]` a la probabilité `P[j]` dêtre tiré pour (`j=0..k-1`). On doit avoir `k=size(C)=size(P)` et `sum(P)=1`.
Si `n==1`, on renvoie l'objet qui a été tiré.
Si `n!=1`, on renvoie la séquence de `k` listes constituées du nom des objets et de leur nombre d'apparition.
On tape :

```
randmultinom(n,P,C):={
  local k,j,l,r,X,L;
  k:=size(P);
  si size(C)!=k alors retourne "erreur"; fsi;
  X:=cumsum(P);
```

```
  si X[k-1]!=1 alors return "erreur"; fsi;
    L:=[C[j],0]$(j=0..k-1);
  pour j de 1 jusque n faire
    r:=alea(0,1);
    l:=0;
    tantque r>X[l] faire
      l:=l+1;
    ftantque;
  L[l,1]:=L[l,1]+1
  fpour;
  si n==1 alors return L[l,0];fsi;
 return L;
}:;
```

On tape :
`randmultinom(5,[1/2,1/3,1/6],["R","V","B"])`
On obtient par exemple :
`[["R",2],["V",2],["B",1]]`
ou bien, on utilise la commande `Xcas : randvector` avec `multinomial` comme paramètre On tape :
`randvector(5,multinomial,[1/2,1/3,1/6],["R","V","B"])`
On obtient dans ce cas la liste des 5 tirages, par exemple :
`["R","B","R","R","V"]`
On simule le tirage de 1 objet parmi 3 objets de probabilité respective :
$P = [1/2, 1/3, 1/6]$ en faisant 6000 tirages.
On tape :
`randmult(6000,[1/2,1/3,1/6])`
On obtient par exemple :
`[3026,2009,965]`
ou bien, on utilise la commande `Xcas : randvector` avec `multinomial` comme paramètre On tape :
`randvector(6000,multinomial,[1/2,1/3,1/6])`
On obtient, par exemple :
`[2947,2040,1013]`
On écrit le programme qui simule `m` fois le choix au hasard de `n` nombres selon la loi multinomiale de probabilité `P` et qui compte le nombre `r` de fois que l'on a obtenu le tirage `K` (`sum(K)=n`) et qui renvoie `r/m` c'est à dire une estimation de la probabilité d'obtenir `K`. `probmult(m,n,P,K)` renvoie donc ne estimation de `multinomiale(n,P,K)`. On tape :

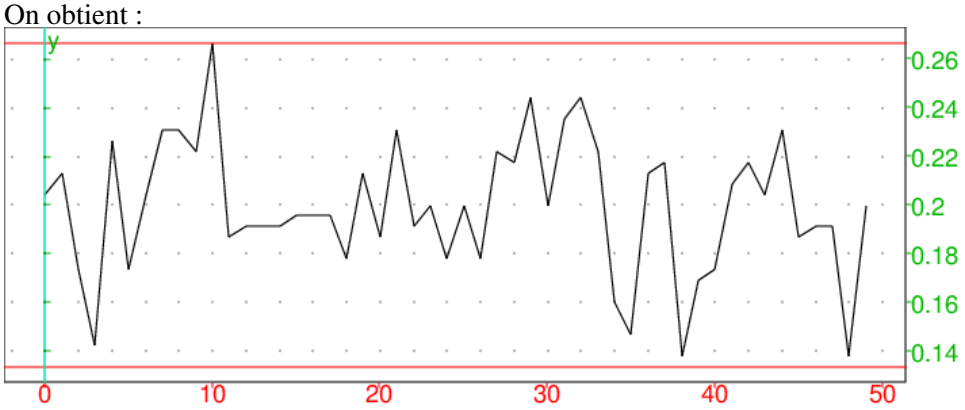
```
probmult(m,n,P,K):={
  local l,T,r;
  r:=0;
  pour l de 1 jusque m faire
    T:=randmult(n,P);
    si T==K alors r:=r+1; fsi;
  fpour;
return r/m;
}:;
```

Exercice Soit une urne ayant 3 boules noires, 2 boules rouges et 1 boule verte. On tire avec remise 2 boules de cette urne.
Quelle est la probabilité de tirer une boule rouge et une boule verte ?
On peut simuler le tirage avec remise 2 boules de cette urne avec :
```
randmult(2,[1/2,1/3,1/6])
```
qui renvoie par exemple :
```
[1,0,1]
```
```
randmultinom(2,["N","R","V"],[1/2,1/3,1/6])
```
qui renvoie par exemple :
```
[["N",1],["R",0],["V",1]]
```
On simule 6000 fois ce tirage et on tape :
```
probmult(6000,2,[1/2,1/3,1/6],[0,1,1])
```
On obtient par exemple :
```
671/6000
```
$671/6000. \simeq 0.111833333333$ On utilise la commande `multinomial` de Xcasé et on tape : `multinomial(2,[1/2,1/3,1/6],[0,1,1])`
On obtient :
```
1/9
```
$1/9. \simeq 0.111111111111$
Exercice Une urne contient 12 boules rouges et 3 boules vertes. On se propose de simuler le tirage d'une boule de l'urne puis d'observer la fluctuation d'échantillonnage sur des échantillons de taille 225. D'après le contenu de l'urne, la probabilité de tirer une boule verte est de $\frac{1}{5} = 0.2$.
Notre simulation est-elle convenable ? On tape :
```
L:=[randmultinomial([4/5,1/5],["R","V"])$(j=1..225)];
```
On obtient :
```
Done
```
On tape :
```
count_eq("V",L)
```
On obtient par exemple :
```
45
```
On analyse tout d'abord 50 échantillons de taille 225 pour voir la fluctuation.
On note `N` le nombre de fois que l'on fait une simulation (une simulation c'est 225 tirages).
`n` le nombre de fois que l'on a obtenu une boule verte,
`p` le pourcentage de boules vertes obtenues par cette simulation,
`Lp` la séquence des pourcentages obtenues.
On tape :

```
test0(N):={
  local L,p,n,k,Lp;
  Lp:=NULL;
pour k de 1 jusque N faire
  L:=[randmultinomial([4/5,1/5],["R","V"])$(j=1..225)];
  n:=count_eq("V",L)
  p:=n/225.;
  Lp:=Lp,p;
fpour;
```

```
 retourne Lp;
  }:;
```

Puis :
`plotlist(test0(50)),droite(y=2/15),droite(y=4/15)`
On obtient :

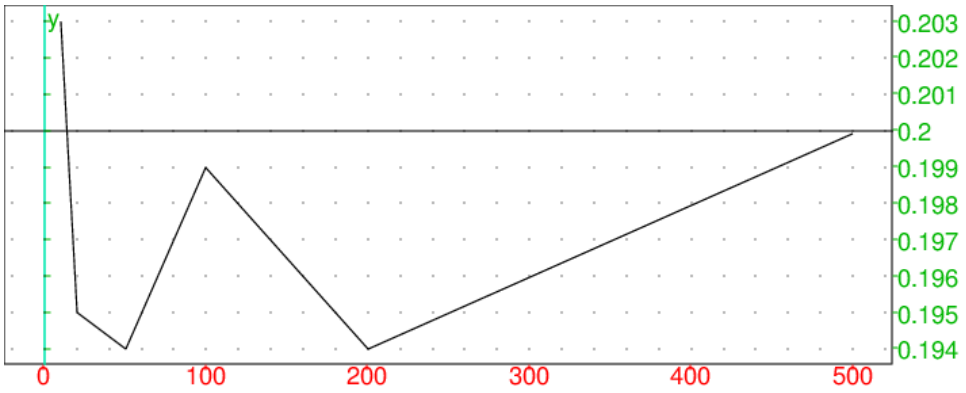


On analyse successivement t échantillons de taille $n = 225$ pour, $t \in 10, 20, 50, 100, 200, 500$.
Pour notre problème, l'intervalle de fluctuation au seuil de 95% est : $p - \frac{1}{\sqrt{n}} p + \frac{1}{\sqrt{n}}$
avec $p = \frac{1}{5}$ et $n = 225$ c'est à dire $\frac{2}{15}, \frac{4}{15}$
Pour savoir si la simulation est corecte on fait un programme pour savoir si on a bien dans 95% des cas p dans l'intervalle $\frac{2}{15}, \frac{4}{15}$
On note `N` le nombre de fois que l'on fait une simulation (une simulation c'est 225 tirages).
Pour la `k` ième simulation, (`k=1..N`) on note :
`L` la liste des 225 tirages obtenus,
`n` le nombre de fois que l'on a obtenu une boule verte,
`p` le pourcentage de boules vertes obtenues par cette simulation,
l `s` le nombre de tirages tels que `2/15<p<4/15` lorsqu'on a fait `k` simulations,
`sn` le nombre de fois que l'on a obtenu une boule verte lorsqu'on a fait `k` simulations.
`pcn` le pourcentage de boules vertes obtenues par ces `N*225` tirages est donc `sn/(225*N)` Le nombre de fois où on a `2/15<p<4/15` est `s`. En pourcentage cela fait donc `pc=s/N`.
On vérifie alors si `pc>0.95`

```
test0(N):={
  local s,L,p,n,pc,sn,pcn,k,Le;
  s:=0;sn:=0;
Le:=NULL;
pour k de 1 jusque N faire
  L:=[randmultinomial([4/5,1/5],["R","V"])$(j=1..225)];
  n:=count_eq("V",L)
  p:=n/225;
  Le:=Le,p;
fpour;
 retourne Le;
  }:;
test(N):={
```

```
  local s,L,p,n,pc,sn,pcn,k,Le;
  s:=0;sn:=0;
Le:=NULL;
pour k de 1 jusque N faire
  L:=[randmultinomial([4/5,1/5],["R","V"])$(j=1..225)];
  n:=count_eq("V",L)
  p:=n/225;
  Le:=Le,p;
  si p>2/15 and p<4/15 alors s:=s+1; fsi;
  sn:=sn+n;
fpour;
  pc:=evalf(s/N);
  pcn:=evalf(sn/N/225);
  si pc>0.95 alors retourne pcn,pc,"correcte"; sinon retourne pcn,pc,"pas cor
  }:;
```

On tape :
`test(10)`
On obtient :
`0.203111111111,1.0,"correcte"`
On tape :
`test(20)`
On obtient :
`0.194888888889,0.95,"pas correcte"`
On tape :
`test(50)`
On obtient :
`0.194311111111,0.98,"correcte"`
On tape :
`test(100)`
On obtient :
`0.198888888889,0.97,"correcte"`
On tape :
`test(200)`
On obtient :
`0.193777777778,0.99,"correcte"`
`test(500)`
On obtient :
`0.19984,0.984,"correcte"`
On tape :
`plotlist([10,20,50,100,200,500],[0.203,0.195,0.194,0.199,0.1940.1999]
), droite(y=0.2)`
On obtient :



6.1.6 Tirage selon une loi normale : `randnorm randNorm`

`randnorm(m,sigma)` ou `randNorm(m,sigma)` renvoie au hasard des nombres répartis selon la loi normale de moyenne `m` et d'écart type `sigma`.
On tape :

```
randnorm(0,1)
```

On obtient par exemple :

```
0.549605372982
```

ou on obtient par exemple :

```
-0.58946494465
```

On tape :

```
randnorm(2,1)
```

On obtient par exemple :

```
2.54178274488
```

6.1.7 Tirage selon une loi exponentielle : `randexp`

`randexp(a)` renvoie au hasard des nombres répartis selon la loi exponentielle de paramètre `a` positif.
La densité de probabilité est proportionnelle à $\exp(-a*t)$ et on a :
Proba($X \leq t$)=$a\int_0^t \exp(-a*u)du$.
On tape :

```
randexp(1)
```

On obtient par exemple :

```
0.310153677284
```

ou on obtient par exemple :

```
0.776007926195
```

6.1.8 Matrice aléatoire : ranm randmatrix randMat

ranm (ou randmatrix ou randMat) peut avoir comme 1,2 ou 3 arguments :

— avec un entier s comme argument, ranm renvoie une liste de longueur s dont les éléments sont des entiers pris au hasard de façon équiprobable dans :
[-99,-98,...,98,99].
On tape :
```
ranm(5)
```
On obtient par exemple :
```
[-40,27,4,-1,94]
```

— avec deux entiers n,p comme argument, ranm renvoie une matrice de n lignes et p colonnes dont les éléments sont des entiers pris au hasard de façon équiprobable dans :
[-99,-98,...,98,99].
On tape :
```
ranm(2,3)
```
On obtient par exemple :
```
[[-32,53,-44],[10,-4,25]]
```

— avec deux entiers n,p et un entier relatif a comme argument, ranm renvoie une matrice de n lignes et p colonnes dont les éléments sont des entiers pris au hasard de façon équiprobable dans [0;a[(ou]a;0] si a est négatif)
On tape :
```
ranm(2,3,10)
```
On obtient par exemple :
```
[[8,3,7],[7,9,1]]
```

— avec deux entiers n,p et un intervalle a..b comme argument, ranm renvoie une matrice de n lignes et p colonnes dont les éléments sont des réels pris au hasard de façon équiprobable dans [a;b[.
On tape :
```
ranm(2,3,0..1)
```
On obtient par exemple :
```
[[0.840187716763,0.394382926635,0.783099223394],
 [0.798440033104,0.911647357512,0.197551369201]]
```

— deux entiers n,p et une fonction aléatoire de Xcas qu'il faut quoter, dans ce cas ranm renvoie une matrice de n lignes et p colonnes dont les éléments sont pris au hasard selon la fonction donnée en troisième argument.
On tape :
```
ranm(3,2,'rand(3)')
```
ou
```
ranm(3,2,3)
```
On obtient par exemple :
```
[[2,1],[0,0],[1,0]]
```
On tape :
```
ranm(1,2,'randnorm(0,1)')
```
On obtient par exemple :

```
[[1.37439065645,-1.33195982697]]
```

6.2 Déplacement aléatoire

6.2.1 Déplacement sur un axe

Une tortue se déplace sur un axe gradué.
Au début de chaque parcours la tortue se trouve à l'origine.
On choisit de la faire avancer en jouant à pile ou face : pile la tortue reste sur place, face elle avance d'une unité.
Un parcours aléatoire de la tortue est constitué par 5 tirages aléatoires.
On veut simuler 30 parcours aléatoires et trouver la probabilité de l'événement : la tortue est arrivée au point d'abscisse x_i pour $x_i \in \mathbb{N}$.

Simulation d'un parcours

On note `T` l'abscisse du point d'arrivée de la tortue.
On écrit le programme `parcours`, en utilisant `rand(2)` qui renvoie de façon équiprobable 0 ou 1.
On considère que 0 correspond à pile et correspond 1 à face.

```
parcours() :={
  local T,r;
  T:=0;
  // on fait 5 tirages
  for (k:=1;k<6;k++){
    r:=rand(2);
    // la tortue avance si r==1 (tirage = face)
    if (r==1){
       T:=T+1;
    }
  }
  return(T);
};
```

Voici les résultats obtenus lorsque l'on fait 10 fois `parcours()` ,on tape :
`for (k:=1;k<11;k++) parcours()`
On obtient par exemple :
`4,2,4,1,1,4,3,2,2,1`

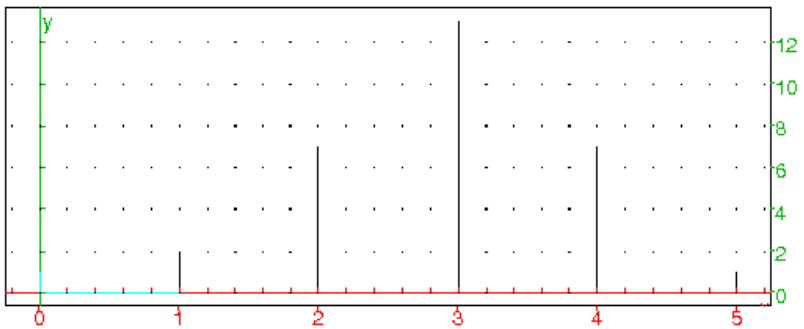
Simulation de n parcours

On note `T` l'abscisse du point d'arrivée de la tortue et `TA` le tableau des résultats : `TA[0]` represente le nombre de fois que la tortue est au point `0` à l'arrivée.
On écrit le programme suivant dans l'éditeur de programmes de `Xcas` (raccourci `Alt+p`) et on sauve ce programme dans le fichier `parsim`.

```
parcoursim(n) :={
  local T,r,TA,R,j,k;
  ClrGraph();
```

```
  TA:=[0,0,0,0,0,0];
  for (j:=1;j<n+1;j++){
    T:=0;
    for (k:=1;k<6;k++){
      r:=rand(2);
      if (r==1){
        T:=T+1;
      }
    }
  TA[T]:=TA[T]+1;
  };
  orint(TA);
  switch_axes(NULL);
  xyztrange(-0.5,5.2,-0.1,16.0,-10.0,10.0,-10.0,-10.0,
            -0.5,5.2,-0.1,16.0,1);
  R:=segment(0,i*TA[0]);
  R:=R,segment(1,1+i*TA[1]);
  R:=R,segment(2,2+i*TA[2]);
  R:=R,segment(3,3+i*TA[3]);
  R:=R,segment(4,4+i*TA[4]);
  R:=R,segment(5,5+i*TA[5]);
  return R;
};
```

Attention Ici `parcoursim` renvoie une liste de segments et écrit en bleu la valeur de `TA`. Le programme se trouve dans un éditeur `prg` de `Xcas`, on le teste avec le bouton `OK`, puis si on a obtenu `//Success compilling`, le programme est validé.
On tape dans la ligne de commande :
`parcoursim(30)`
On obtient en bleu :
`TA:[0,4,14,6,6,0]`
et le graphique :



Voici des résultats obtenus pour la liste des abscisses des points d'arrivée :
pour `parcoursim(30)` on a trouvé `[0,2,7,13,7,1]`
pour `parcoursim(300)` on a trouvé `[7,41,94,102,47,9]`
pour `parcoursim(1000)` on a trouvé `[36,172,310,306,148,28]`

pour `parcoursim(10000)` on a trouvé `[287,1575,3184,3136,1517,301]`

Analyse des résultats

Soit l'univers Ω formé par les 5 tirages successifs possibles (chacun étant équiprobable) :
$\Omega = \{\{p,p,p,p,p\},\{p,p,p,p,f\},\{p,p,p,f,p\},...,\{f,f,f,f,f,\}\}$.
Ω a $2^5 = 32$ éléments.
Soit `A` la variable aléatoire égale à l'abscisse du point d'arrivée.
On a :
$\mathtt{P(A=0)} = \dfrac{1}{2^5} = 0.03125$ car cela correspond à 5 fois "pile",
$\mathtt{P(A=1)} = \dfrac{5}{2^5} = 0.15625$ car cela correspond à 4 fois "pile" et 1 fois "face" ce qui peut se produire de 5 façons,
$\mathtt{P(A=2)} = \dfrac{10}{2^5} = 0.3125$ car cela correspond à 3 fois "pile" et 2 fois "face" ce qui peut se produire de $C_5^2 = 10$ façons,
$\mathtt{P(A=3)} = \dfrac{10}{2^5} = 0.3125$ car cela correspond à 2 fois "pile" et 3 fois "face" ce qui peut se produire de $C_5^3 = 10$ façons,
$\mathtt{P(A=4)} = \dfrac{5}{2^5} = 0.15625$ car cela correspond à 1 fois "pile" et 4 fois "face" ce qui peut se produire de 5 façons,
$\mathtt{P(A=5)} = \dfrac{1}{2^5} = 0.03125$ car cela correspond à 5 fois "face".

6.2.2 Déplacement dans deux directions

Au début de chaque parcours la tortue se trouve à l'origine.
On choisit de la faire avancer en jouant à pile ou face :
- pile, la tortue avance d'une unité selon un axe vertical,
- face, elle avance d'une unité selon un axe horizontal.
Un parcours aléatoire de la tortue est constitué par 5 tirages aléatoires.
On veut simuler `n` parcours aléatoires et trouver la probabilité de l'évènement : le parcours aléatoire de la tortue se termine au point de coordonnées $[x,y]$ pour $(x,y) \in N \times N$

Simulation d'un parcours

On note `X,Y` les coordonnées du point d'arrivée de la tortue.
On écrit :

```
parcours2() :={
  local X,Y,r;
  X:=0;
  Y:=0;
  for (k:=1;k<6;k++){
    r:=rand(2);
    if (r==1){
       X:=X+1;
    } else {
```

```
        Y:=Y+1;
      }
    }
    return([X,Y]);
};
```

Voici les résultats obtenus lorsque l'on fait 10 fois `parcours2()` :
`[2,3],[1,4],[3,2],[4,1],[3,2],[2,3],[1,4],`
`[1,4],[3,2],[2,3]`
On remarque qu'à chaque tirage soit `X`, soit `Y` est augmenté d'une unité donc à chaque tirage `X+Y` aygmente de 1. Au début `X+Y=0`, donc, au bout de 5 tirages c'est à dire à la dernière étape `X+Y=5`.
Il suffit donc de connaitre l'abscisse d'arrivée `X` pour connaitre le point d'arrivée (`point(X,5-X)`). Ce problème est donc le même que le précédent.

Simulation de n parcours

On note `XA` le tableau des résultats selon les abscisses.
On remarquera qu'ici `Y` ne sert à rien puisqu'on peut repérer le point d'arrivée seulement à l'aide de son abscisse, elle permet juste de visualiser le point d'arrivée.
On écrit :

```
parcoursim2(n) :={
  local X,Y,r,j,k,XA;
  XA:=[0,0,0,0,0,0];
  for (j:=1;j<n+1;j++){
    r:=rand(2);
    X:=0;
    Y:=0;
    for (k:=1;k<6;k++){
      if (r==1){
        X:=X+1;
      } else {
        Y:=Y+1;
      }
    r:=rand(2);
    }
  XA[X]:=XA[X]+1;
  }
  switch_axes(NULL);
  ClrGraph();
  xyztrange(-0.5,5.2,-0.1,16.0,-10.0,10.0,-10.0,-10.0,
            -0.5,5.2,-0.1,16.0,1);

  return([XA,segment(0,i*XA[0]),segment(1,1+i*XA[1]),
     segment(2,2+i*XA[2]),segment(3,3+i*XA[3]),
     segment(4,4+i*XA[4]),segment(5,5+i*XA[5])]);
};
```

Voici les résultats obtenus :
pour `parcoursim2(30)` on a trouvé :
`XA=[0,4,9,9,7,1]`
pour `parcoursim2(300)` on a trouvé :
`XA=[6,48,91,99,46,10]`
pour `parcoursim2(1000)` on a trouvé :
`XA=[26,170,313,320,148,23]`
pour `parcoursim2(10000)` on a trouvé :
`XA=[290,1498,3207,3128,1572,305]`
Attention Ici `parcoursim2` renvoie une liste de segments : ces segments seront donc dessinés dans un écran de géométrie et dans l'écran `DispG`. Il faut donc écrire `ClrGraph()` en début de programme si on veut effacer l'écran de géométrie `DispG`.

Analyse des résultats

On a donc la même analyse que dans le parcours linéaire.
Soit `A` la variable aléatoire égale aux coordonnées du point d'arrivée.
$\mathtt{P}(\mathtt{A}=[0,5]) = \dfrac{1}{2^5} = 0.03125$ car cela correspond à 5 fois "pile",
$\mathtt{P}(\mathtt{A}=[1,4]) = \dfrac{5}{2^5} = 0.15625$ car cela correspond à 4 fois "pile" et 1 fois "face" ce qui peut se produire de 5 façons,
$\mathtt{P}(\mathtt{A}=[2,3]) = \dfrac{10}{2^5} = 0.3125$ car cela correspond à 3 fois "pile" et 2 fois "face" ce qui peut se produire de $C_5^2 = 10$ façons,
$\mathtt{P}(\mathtt{A}=[3,2]) = \dfrac{10}{2^5} = 0.3125$ car cela correspond à 2 fois "pile" et 3 fois "face" ce qui peut se produire de $C_5^3 = 10$ façons,
$\mathtt{P}(\mathtt{A}=[4,1]) = \dfrac{5}{2^5} = 0.15625$ car cela correspond à 1 fois "pile" et 4 fois "face" ce qui peut se produire de 5 façons,
$\mathtt{P}(\mathtt{A}=[5,0]) = \dfrac{1}{2^5} = 0.03125$ car cela correspond à 5 fois "face".

6.3 Les trois cartes bicolores

On met dans un chapeau trois cartes : une des cartes a deux côtés rouges, une autre a un côté rouge et un côté blanc et la troisième a deux côtés blancs.
On tire une carte : le côté que nous voyons est rouge.
Quelle est la probabilité pour que l'autre côté soit blanc ?

6.3.1 Simulation de n tirages

Pour écrire le programme de simulation, on numérote les cartes par 0,1 et 2 et on numérote les faces de chaque carte par 0 et 1 : par exemple la carte blanche a le numéro `0`, la carte bicolore a le numéro `1`, la carte rouge a le numéro `2` et la face blanche de la carte bicolore a le numéro `0`.
Puis, on représente donc une carte par un vecteur qui est la couleur de ses faces : par exemple la carte bicolore sera représentée par `[B,R]`.
On peut aussi représenter le blanc par 0 et le rouge par 1 :

par exemple la carte bicolore sera représentée par `[0,1]`.
On représente ainsi les cartes par un vecteur de deux composantes 0 ou 1 (0 et 1 désigne la couleur).
La variable `C:=[[0,0],[0,1],[1,1]]` représente donc les trois cartes :
`[0,0]` est la carte avec 2 faces blanches, `[1,1]` est la carte avec 2 faces rouges et `[0,1]` est la carte bicolore (on a supposé que la face blanche a le numéro 0 et la rouge le numéro 1).
`C` est donc une matrice et la valeur de `C[a,b]` (pour `a=0,1,2` et `b=0,1)` représente la couleur de la face `b` de la carte `a`.
On tire une des cartes (`a:=rand(3);`),
puis on tire la face visible (`b:=rand(2);`).
Si `b` est la face visible, `irem(b+1,2)` c'est à dire `b+1 mod 2` est la face cachée.
On simule n tirages qui donne comme coté visible une face rouge et on compte le nombre de cartes bicolores.
On écrit pour cela le programme `cartebicolor` :

```
cartebicolor(n):={
local C,a,b,nbi;
C:=[[0,0],[0,1],[1,1]];
//nbi est le nbre de cartes bicolores obtenus
//qd la face visible est blanche
nbi:=0
for (k:=0;k<n;k++){
//on tire une carte
a:=rand(3);
// on tire une face (la face visible)
b:=rand(2);
// on refait le tirage si la face visible est blanche
while (C[a,b]==0) {
a:=rand(3);
b:=rand(2);
}
//la face visible est rouge, si la face cachee est blanche,
// nbi augmente de 1
if (C[a,irem(b+1, 2)]]==0) {
nbi:=nbi+1;
}
}
return(evalf(nbi/n));
};
```

On a obtenu :

```
cartebicolor(300)=0.34
cartebicolor(3000)=0.343666666667
cartebicolor(30000)=0.331533333333
```

6.3.2 Analyse du résultat

Etant donné qu'il y a autant de côtés rouges que de côtés blancs, le problème suivant a la même réponse :
On tire une carte : le côté que nous voyons est blanc.
Quelle est la probabilité pour que l'autre côté soit rouge ?
ou encore :
On tire une carte : nous voyons un côté de cette carte.
Quelle est la probabilité pour que l'autre côté ne soit pas de la même couleur ?
Cela revient à demander quelle est la probabilité pour que l'on ait tiré la carte bicolore. Comme il y a trois cartes dont une seule est bicolore, la probabilité cherchée est égale à $\frac{1}{3}$.
On peut aussi traiter ce problème avec les probabilités conditionnelles :
soit Ω l'ensemble des faces visibles. On repére la face visible par 2 nombres le numéro de sa carte et son numéro de face (par exemple [1,0] désigne la face 0 de la carte 1 alors que [0,1] désigne la face 1 de la carte 0) on a
$\Omega = \{[0,0],[0,1],[1,0],[1,1],[2,0],[2,1]\}$.
Les trois premiers éléments de Ω ont comme face visible une face blanche, les trois derniers éléments de Ω sont comme face visible une face rouge.
Soit A l'évènement "le coté visible est rouge",
soit B l'évènement "le coté non visible est blanc",
soit C l'évènement "le coté visible est rouge et le coté non visible est blanc" ou "le coté visible est blanc et le coté non visible est rouge" (ie la carte tirée est bicolore).
P(A)=$\frac{1}{2}$
P(B)=$\frac{1}{2}$
P(C)=$\frac{1}{3}$
P(A et B)=$\frac{1}{6}$
P(C)=P(A et B)+P(nonA et nonB)=$\frac{1}{3}$
P(B/A)=P(A et B)/P(A)=$\frac{1}{6}/\frac{1}{2}=\frac{1}{3}$
On peut aussi numéroter les faces rouges et les faces blanches et dire qu'un couple represente une carte et le premier élément du couple est la face visible.
Par exemple, (R_3, B_3) représente la carte bicolore ayant comme face visible la face rouge.
$\Omega = \{(R_1,R_2),(R_2,R_1),(R_3,B_3),(B_3,R_3),(B_1,B_2),(B_2,B_1)\}$.
Soit A l'évènement "le coté visible est rouge" :
A=$\{(R_1,R_2),(R_2,R_1),(R_3,B_3)\}$
Donc :
P(A)=$\frac{1}{2}$
soit B l'évènement "le coté non visible est blanc" :
A et B=$\{(R_3,B_3)\}$
Donc :
P(A et B)=$\frac{1}{6}$
Donc :

P(B/A)=P(A et B)/P(A)=$\frac{1}{6}/\frac{1}{2}=\frac{1}{3}$

6.4 Les quatre cartes bicolores

On met dans un chapeau quatre cartes : une des cartes a deux côtés blancs, une autre a un côté rouge et un côté blanc et les deux restantes ont deux côtés rouges.
On tire une carte : le côté que nous voyons est rouge.
Quelle est la probabilité pour que l'autre côté soit blanc ?

6.4.1 Simulation

```
cartebic4(n):={
  local C,a,b,nbi;
  C:=[[0,0],[0,1],[1,1],[1,1]];
  nbi:=0
  for (k:=0;k<n;k++){
    a:=rand(4);
    b:=rand(2);
    while (C[a,b]==0) {
      a:=rand(4);
      b:=rand(2);
    }
    if (C[a,irem(b+1, 2)]==0) {
      nbi:=nbi+1;
    }
  }
  return(evalf(nbi/n));
};
```

On a obtenu :

```
cartebic4(300)= 0.18
cartebic4(3000)= 0.203666666667
cartebic4(30000)=0.2019
```

6.4.2 Analyse du résultat

On va traiter ce problème avec les probabilités conditionnelles :
soit Ω l'ensemble des faces visibles.
On repére la face visible par 2 nombres le numéro de sa carte et son numéro de face (par exemple [1,0] désigne la face 0 de la carte 1 alors que [0,1] désigne la face 1 de la carte 0) on a
$\Omega = \{[0,0],[0,1],[1,0],[1,1],[2,0],[2,1],[3,0],,[3,1]\}$.
Onsuppose que la carte 0 a 2 faces blanches,que la face 0 de la carte 1 est blanche et que sa face 1 est rouge et que les cartes 2 et 3 ont 2 faces rouges. Donc les trois premiers éléments de Ω sont des faces blanches, les cinq derniers éléments de Ω sont des faces rouges.
Soit A l'évènement "le coté visible est rouge",

soit B l'évènement "le coté non visible est blanc",
soit C l'évènement la carte tirée est bicolore.
P(A)=$\frac{5}{8}$
P(B)=$\frac{3}{8}$
P(A et B)=$\frac{1}{8}$
P(C)=P(A et B)+P(nonA et nonB)=$\frac{1}{4}$
P(B/A)=P(A et B)/P(A)=$\frac{1}{8}/\frac{5}{8}=\frac{1}{5}$.
Donc la probabilité que la face cachée soit blanche sachant que la face visible est rouge est : $\frac{1}{5}$
Etant donné qu'il n'y a pas autant de côtés rouges que de côtés blancs, le problème posé n'est pas le même que :
On tire une carte : le côté que nous voyons est blanc.
Quelle est la probabilité pour que l'autre côté soit rouge ?
On a :
P(nonA)=$\frac{3}{8}$
P(nonB)=$\frac{5}{8}$
P(nonA et nonB)=$\frac{1}{8}$
P(nonB/nonA)=P(nonA et nonB)/P(nonA)=$\frac{\frac{1}{8}}{\frac{3}{8}}=\frac{1}{3}$.
On retrouve la même probabilité que dans le cas des trois cartes bicolores car la probabilité demandée ne tient compte que de l'ensemble des cartes qui ont un côté blanc et dans les deux problèmes cet ensemble est le même.
ce n'est pas non plus le même problème que :
On tire une carte : nous voyons un côté de cette carte.
Quelle est la probabilité pour que l'autre côté ne soit pas de la même couleur ?
On demande ici la probabilité de tirer la carte bicolore c'est à dire : P(C)=$\frac{1}{4}$
Remarque
P(C)=P(A)*P(B/A)+P(nonA)*P(nonB/nonA)=
P(A et B)+P(nonA et nonB)= $\frac{1}{4}=\frac{3}{8}*\frac{1}{3}+\frac{5}{8}*\frac{1}{5}$

6.5 La voiture et les deux chèvres

Un candidat à un jeu doit choisir entre trois portes et gagne ce qui se trouve derrière la porte choisie. Il y a une voiture derrière une porte et une chèvre derrière chacune des deux autres portes. Le candidat choisit une porte et le présentateur qui connait la porte gagnante, ouvre une des deux portes restantes derrière laquelle se trouve une chèvre, et demande au candidat si il veut changer son choix. A votre avis le candidat a-t-il plus de chances de gagner la voiture en changeant systématiquement son choix ?
Pour répondre à cette question, commençons par une simulation.

6.5.1 Simulation

Le paramètre `n` représente le nombre de jeux.
`ng1` est le nombre de fois où le candidat gagne quand il ne change jamais de choix (situation1) et
`ng2` est le nombre de fois où le candidat gagne quand il change systématiquement de choix (situation2).
La porte où l'on met la voiture est tirée au hasard (`v:=rand(3);P[v]:=1`).
Le candidat choisit une porte au hasard (`a:=rand(3)`).
Si `(a==v)` il gagne dans la situation1 (`ng1:=ng1+1`) et perd dans la situation2 (`ng2` reste inchangé).
Si `(a!=v)` il gagne dans la situation2 (`ng2:=ng2+1`) et perd dans la situation1 (`ng1` reste inchangé).
Dans ce qui suit la variable `P` ne sert à rien et permet juste de visualiser les 3 portes (si `P[n]==0`, derrière la porte de numéro `n` il y a une chèvre, et si `P[n]==1`, derrière la porte de numéro `n` il y a une voiture).
On écrit le programme `chevre` qui compte le nombre de gains dans `ng1` quand on ne change pas son choix et qui compte le nombre de gains dans `ng2` quand on change systématiquement son choix.

```
chevre(n):={
local a,v,ng1,ng2;
ng1:=0;
ng2:=0
for (k:=0;k<n;k++){
\\on choisit la porte v o\'u l'on met la voiture
v:=rand(3);
P:=[0,0,0];
P[v]:=1;
//le candidat choisit une porte a
a:=rand(3);
if (a==v){ng1:=ng1+1;}
else {ng2:=ng2+1;}
}
return ([evalf(ng1/n),evalf(ng2/n)]);
};
```

On a obtenu : `chevre(10000)= [0.3303,0.6697]`

6.5.2 Analyse du résultat

La voiture est derrière la porte `v`.
Le candidat choisit une porte `a` au hasard.
Si `(a==v)` il gagne dans la situation1 et perd dans la situation2 et,
si `(a!=v)` il gagne dans la situation2 et perd dans la situation1.
On a donc :
P(a=v)=$\frac{1}{3}$
et donc P(a !=v)=1-P(a=v)=$\frac{2}{3}$
Le candidat a donc deux fois plus de chances de gagner s'il change son choix

systématiquement !
Remarque
Pourtant malgré la simplicité de la situation, notre intuition semble en défaut ...
Si ce qui précéde ne vous a pas convaincu faites le même problème avec 100 portes (1 voiture et 99 chèvres) et le présentateur ouvre 98 portes derrière lesquelles il y a des chèvres : on comprend bien qu'en désignant une porte, la voiture a plus de chances (99 chances sur 100) d'être derrière les portes restantes, et en ouvrant les 98 portes le présentateur élimine 98 chèvres et donc derrière la porte restante il y a 99 chances sur 100 pour qu'il y ait la voiture.

6.6 Comment couper des spaghettis en trois ?

Voici l'énoncé d'un problème :
On coupe de façon aléatoire un spaghetti en trois morceaux. Quelle est la probabilité pour qu'avec les trois morceaux obtenus on puisse former un triangle ? Comment peut-on simuler cette situation ou autrement dit que veut dire "on coupe de façon aléatoire un spaghetti en trois morceaux" ?
On suppose dans ce qui suit le spaghetti de longueur 1.

- Première méthode : on choisit au hasard deux points x et y de [0,1].
- Deuxième méthode : on choisit au hasard un point x de [0,1], puis on choisit au hasard le point y dans [0,x].
- Troisième méthode : on choisit au hasard un point x de [0,1], puis on choisit au hasard l'un des segments [0,x] ou [x,1], puis on choisit au hasard le point y dans le segment choisi.
- Quatrième méthode : on choisit au hasard un point x de [0,1], puis on choisit le plus grand des segments [0,x] ou [x,1], puis on choisit au hasard le point y dans le segment choisi.

Ces différentes méthodes conduisent-elles au même résultat ?
Quelle est la méthode qui donne la plus forte probabilité ?
Pour répondre à ces questions commençons par des simulations.
Pour cela, il faut savoir répondre à la question : à quelles conditions trois segments de longueurs a, b et $c = 1 - a - b$ forment-ils un triangle ?
Une condition necessaire et suffisante est que :
$a < b + c$ et $b - c < a$ et $c - b < a$ ou encore que :
$a < 1 - a$ et $a + 2b - 1 < a$ et $1 - a - 2b < a$ ou encore que :
$a < 0.5$ et $b < 0.5$ et $0.5 < a + b$

6.6.1 Simulation première méthode

On choisit au hasard deux points d'abscisses x et y de l'intervalle [0 ;1].
On note :
`x` et `y` les abscisses des points de coupures.
`a` et `b` la longueur du premier et du deuxième morceau de spaghetti.
`t` le nombre de triangles obtenus au bout de `n` essais.

```
spag1(n):={
  local x,y,a,b,t;
  t:=0;
```

```
  for (k:=1;k<=n;k++){
     x:=evalf(rand(2^30)/2^30);
     y:=evalf(rand(2^30)/2^30);
     if (x<y) {
        a:=x;
        b:=y-x;
     } else {
        a:=y;
        b:=x-y;
     }
     if ((a<0.5) and (b<0.5) and (a+b>0.5)) {
        t:=t+1;
     }
  }
  return(evalf(t/n));
};
```

On a trouvé pour n=30000 : `0.2506`
On a trouvé pour n=300000 : `0.24965`

6.6.2 Simulation deuxième méthode

On choisit au hasard un point d'abscisse x de l'intervalle [0 ;1], puis on choisit au hasard un point d'abscisse y de l'intervalle $[0; x]$.
On note :
`x` et `y` les abscisses des points de coupures.
`a` et `b` la longueur du premier et du deuxième morceau de spaghetti.
`t` le nombre de triangles obtenus au bout de `n` essais.

```
spag2(n):={
  local x,y,a,b,t;
  t:=0;
  for (k:=1;k<=n;k++){
     x:=evalf(rand(2^30)/2^30);
     y:=evalf(rand(2^30)/2^30)*x;
     a:=y;
     b:=x-y;
     if ((a<0.5) and (b<0.5) and (a+b>0.5)) {
        t:=t+1;
     }
  }
  return(evalf(t/n));
};
```

On a trouvé pour n=30000 : `0.193266666667`
On a trouvé pour n=300000 : `0.191666666667`

6.6.3 Simulation troisième méthode

On choisit au hasard un point d'abscisse x de l'intervalle [0 ;1], puis on choisit au hasard l'intervalle $[0; x]$ ou $[x; 1]$ puis, on choisit au hasard un point d'abscisse

y dans l'intervalle choisi.
On note :
x et y les abscisses des points de coupures.
a et b la longueur du premier et du deuxième morceau de spaghetti.
t le nombre de triangles obtenus au bout de n essais.

```
spag3(n):={
  local x,y,a,b,t;
  t:=0;
  for (k:=1;k<=n;k++){
     x:=evalf(rand(2^30)/2^30);
     if (rand(2)==0){
       y:=evalf(rand(2^30)/2^30)*x;
       a:=y;
       b:=x-y;
     } else {
       y:=evalf(rand(2^30)/2^30)*(1-x)+x;
       a:=x;
       b:=y-x;
     }
     if ((a<0.5) and (b<0.5) and (a+b>0.5)) {
       t:=t+1;
     }
  }
  return(evalf(t/n));
};
```

On a trouvé pour n=30000 : 0.195533333333
On a trouvé pour n=300000 : 0.194083333333

6.6.4 Simulation quatrième méthode

On choisit au hasard un point d'abscisse x de l'intervalle [0 ;1], puis on choisit le plus grand des deux intervalles $[0; x]$ ou $[x; 1]$ puis, on choisit au hasard un point d'abscisse y dans l'intervalle choisi.
On note :
x et y les abscisses des points de coupures.
a et b la longueur du premier et du deuxième morceau de spaghetti.
t le nombre de triangles obtenus au bout de n essais.

```
spag4(n):={
  local x,y,a,b,t;
  t:=0;
  for (k:=1;k<=n;k++){
     x:=evalf(rand(2^30)/2^30);
     if (x>0.5){
       y:=evalf(rand(2^30)/2^30)*x;
       a:=y;
       b:=x-y;
```

```
      } else {
        y:=evalf(rand(2^30)/2^30)*(1-x)+x;
        a:=x;
        b:=y-x;
      }
      if ((a<0.5) and (b<0.5) and (a+b>0.5)) {
        t:=t+1;
      }
  }
  return(evalf(t/n));
};
```

On a trouvé pour n=30000 : `0.388366666667`
On a trouvé pour n=300000 : `0.385946666667`

On remarque que :
0.194083333333*2 = 0.388166666666
0.191666666667*2 = 0.383333333334
ln(2)-0.5= 0.19314718056

6.6.5 Analyse des résultats

Première méthode

Première méthode : on choisit au hasard deux points x et y de [0,1].
On sait que si l'on obtient $x < 0.5$, pour obtenir un triangle dans ce cas, il faut choisir y dans l'intervalle $[\frac{1}{2}, x + \frac{1}{2}]$ qui est un intervalle de longueur x. La probabilité d'obtenir un y qui convient est donc alors égale à x.
On sait que si l'on obtient $x > 0.5$, pour obtenir un triangle dans ce cas, il faut choisir y dans l'intervalle $[x - \frac{1}{2}, \frac{1}{2}]$ qui est un intervalle de longueur $1 - x$. La probabilité d'obtenir un y qui convient est donc alors égale à $1 - x$.
Donc la probabilité d'obtenir un triangle est :

$$\int_0^{\frac{1}{2}} x dx + \int_{\frac{1}{2}}^1 (1-x) dx = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

Deuxième méthode

Deuxième méthode : on choisit au hasard un point x de [0,1], puis on choisit au hasard le point y dans [0,x].
On sait que si l'on obtient $x < 0.5$, on a une probabilité nulle d'obtenir un triangle puisque ensuite on choisit y vérifiant $y < x$. On sait que si l'on obtient $x > 0.5$, pour obtenir un triangle dans ce cas, il faut choisir y dans l'intervalle $[x - \frac{1}{2}, \frac{1}{2}]$ qui est un intervalle de longueur $1 - x$. La probabilité d'obtenir un y qui convient est donc égale à $\frac{1-x}{x}$.
Donc la probabilité d'obtenir un triangle est :

$$\int_{\frac{1}{2}}^1 \frac{1-x}{x} dx = \ln(2) - \frac{1}{2}$$

Troisième méthode

Troisième méthode : on choisit au hasard un point x de [0,1], puis on choisit au hasard l'un des segments [0,x] ou [x,1], puis on choisit au hasard le point y dans le segment choisi.
Si on choisit avec une probabilité 0.5 l'un des deux segments $[0,x[$ ou $[x,1[$, si $x < 0.5$ pour obtenir un y qui convient il faut choisir (avec une probabilité de 0.5) l'intervalle $[x,1[$ (qui est un intervalle de longueur $1-x$), puis choisir y dans l'intervalle $[\frac{1}{2}, x+\frac{1}{2}]$ qui est un intervalle de longueur x et la probabilité d'obtenir un y qui convient est donc égale à $\frac{1}{2} * \frac{x}{1-x}$.
Si $x > 0.5$ pour obtenir un y qui convient il faut choisir (avec une probabilité de 0.5) l'intervalle $[0,x[$ (de longueur x), puis choisir y dans l'intervalle $[x-\frac{1}{2}, \frac{1}{2}]$ qui est un intervalle de longueur $1-x$ et la probabilité d'obtenir un y qui convient est donc égale à $\frac{1}{2} * \frac{1-x}{x}$.
Donc la probabilité d'obtenir un triangle est :
$$\frac{1}{2}\int_0^{\frac{1}{2}} \frac{x}{1-x}dx + \frac{1}{2}\int_{\frac{1}{2}}^1 \frac{1-x}{x}dx = \frac{1}{2}(\ln(2)-\frac{1}{2}) + \frac{1}{2}(\ln(2)-\frac{1}{2}) = \ln(2)-\frac{1}{2}$$

Quatrième méthode

Quatrième méthode : on choisit au hasard un point x de [0,1], puis on choisit le plus grand des segments [0,x] ou [x,1], puis on choisit au hasard le point y dans le segment choisi.
Si $x < 0.5$, on choisit y dans $[x,1[$ (de longueur $1-x$), puis pour obtenir un y qui convient il faut le choisir dans l'intervalle $[\frac{1}{2}, x+\frac{1}{2}]$ qui est un intervalle de longueur x et la probabilité d'obtenir un y qui convient est donc égale à $\frac{x}{1-x}$.
Si $x > 0.5$, on choisit y dans $[0,x[$ (de longueur x), puis pour obtenir un y qui convient il faut le choisir dans l'intervalle $[x-\frac{1}{2}, \frac{1}{2}]$ qui est un intervalle de longueur $1-x$ et la probabilité d'obtenir un y qui convient est donc égale à $\frac{1-x}{x}$.
Donc la probabilité d'obtenir un triangle est :
$$\int_0^{\frac{1}{2}} \frac{x}{1-x}dx + \int_{\frac{1}{2}}^1 \frac{1-x}{x}dx = \ln(2) - \frac{1}{2} + \ln(2) - \frac{1}{2} = 2*\ln(2) - 1$$

Quelques questions

Lorsque x a été choisi, on choisit de placer y soit sur $[0,x[$ soit sur $[x,1[$ avec quelle probabilité doit-on faire ce choix pour avoir les cotés d'un triangle avec une probabilité de 0.25 ?
Il faut choisir le segment $[0,x[$ avec une probabilité de x et donc choisir le segment $[x,1[$ avec une probabilité de $1-x$.
En effet la probabilité d'obtenir les 3 côtés d'un triangle est alors :
$$\int_0^{\frac{1}{2}} (1-x)*\frac{x}{1-x}dx + \int_{\frac{1}{2}}^1 x*\frac{1-x}{x}dx = \int_0^{\frac{1}{2}} xdx + \int_{\frac{1}{2}}^1 (1-x)dx = \frac{1}{8}+\frac{1}{8} = \frac{1}{4}.$$
Voici la simulation :

```
spag5(n):={
```

```
  local x,y,a,b,t;
  t:=0;
  for (k:=1;k<=n;k++){
     x:=evalf(rand(2^30)/2^30);
     if (evalf(rand(2^30)/2^30)<x){
       y:=evalf(rand(2^30)/2^30)*x;
       a:=y;
       b:=x-y;
     } else {
       y:=evalf(rand(2^30)/2^30)*(1-x)+x;
       a:=x;
       b:=y-x;
     }
     if ((a<0.5) and (b<0.5) and (a+b>0.5)) {
       t:=t+1;
     }
  }
  return(evalf(t/n));
};
```

On a trouvé pour n=30000 : `0.2502`
On a trouvé pour n=300000 : `0.251556666667`

Que se passe-t-il si on choisit le segment $[0, x[$ avec une probabilité de $1 - x$ et le segment $[x, 1[$ avec une probabilité de x ?
Voici la simulation :

```
spag6(n):={
  local x,y,a,b,t;
  t:=0;
  for (k:=1;k<=n;k++){
     x:=evalf(rand(2^30)/2^30);
     if (evalf(rand(2^30)/2^30)<1-x){
       y:=evalf(rand(2^30)/2^30)*x;
       a:=y;
       b:=x-y;
     } else {
       y:=evalf(rand(2^30)/2^30)*(1-x)+x;
       a:=x;
       b:=y-x;
     }
     if ((a<0.5) and (b<0.5) and (a+b>0.5)) {
       t:=t+1;
     }
  }
  return(evalf(t/n));
};
```

On a trouvé pour n=30000 : `0.138533333333`
On a trouvé pour n=300000 : `0.136773333333`

Exercice : Montrer que de façon théorique, on trouve : `2*ln(2)-5/4`
On vérifie : `evalf(2*log(2)-5/4)=0.13629436112`

6.6.6 Comment simuler l'expérimentation ?

première façon

Supposons que l'on fasse faire à un groupe de personnes l'expérimentation de la quatrième méthode (on recoupe le plus grand morceau). Lorsqu'une personne efféctue l'expérience la première cassure (celle qui détermine x) se fera en général entre h et $1-h$: h étant l'emplacement des doigts. On suppose ensuite que l'emplacement des doigts nécessaire pour faire la cassure est proportionnel à la longueur donc si y se trouve sur $[0, x[$ la cassure se fera sur $[hx, x-xh[$.
On écrit donc la fonction suivant dépendant de n nombre d'expériences et h l'emplacement des doigts.

```
spagex(n,h):={
local x,y,a,b,t;
t:=0;
for (k:=1;k<=n;k++){
  x:=evalf(rand(2^30)/2^30);
  x:=h+x*(1-2*h);
  if (x>0.5){
    y:=h*x+evalf(rand(2^30)/2^30)*x*(1-2*h);
    a:=y;
    b:=x-y;
  } else {
    y:=(1-x)*h+evalf(rand(2^30)/2^30)*(1-x)*(1-2*h)+x;
    a:=x;
    b:=y-x;
  }
  if ((a<0.5) and (b<0.5) and (a+b>0.5)) {
    t:=t+1;
  }
}
return(evalf(t/n));
};
```

On trouve pour $n = 30$ et $h = 0.08$: `0.6`
On trouve pour $n = 3000$ et $h = 0.08$: `0.626666666667`
On trouve pour $n = 3000$ et $h = 0.1$: `0.561`
On trouve pour $n = 300$ et $h = 0.1$: `0.535666666667`
On trouvera dans le répertoire `simulation`, les valeurs du couple $[x, y]$ trouvées lors de l'exécution de spag4(100) dans le fichier `Asim` et, les valeurs du couple $[x, y]$ trouvées lors de l'exécution de spagex(100,0.1) dans le fichier `Aex`. Bien sûr, on doit rajouter dans ces deux programmes une variable globale dans laquelle on engrange les valeurs de $[x, y]$.
Le calcul théorique de la probabilité d'obtenir un triangle est alors :

$$\frac{1}{1-2h}\left(\int_h^{\frac{1}{2}} \frac{x}{(1-x)*(1-2h)}dx + \int_{\frac{1}{2}}^{\frac{1}{2-2h}} dx + \int_{\frac{1}{2-2h}}^{1-h} \frac{1-x}{x(1-2h)}dx\right)$$

ce qui donne la formule :
$\frac{1}{(1-2h)^2}(\ln(2(1-h)^2) + \frac{-6h^2+9h-2}{2(1-h)(1-2h)^2}$.
En effet,
- quand $h < x < \frac{1}{2}$, on choisit y dans $]x+(1-x)*h; 1-(1-x)*h[$ (segment de longueur $(1-x)*(1-2*h)$) on aura un triangle si $\frac{1}{2} < y < x+\frac{1}{2}$ (segment de longueur x),
- quand $\frac{1}{2} < x < \frac{1}{2-2h}$, on choisit y dans $]h.x; x-h.x[$ on est sûr d'avoir un triangle car $y < x-h*x < \frac{1}{2}$,
- quand $\frac{1}{2-2h} < x < 1-h$, on choisit y dans $]h.x; x-h.x[$ (intervalle de longueur $x(1-2h)$) on aura un triangle si $\frac{1}{2}-x < y < \frac{1}{2}$ (intervalle de longueur $1-x$).
d'ou les trois intégrales qu'il faut diviser par $1-2h$ car on choisit x dans $]h; 1-h[$ (intervalle de longueur $1-2h$)

Une autre façon de simuler l'expérimentation

On peut aussi, par exemple, supposer que l'on coupe le spaghetti en suivant une loi de probabilité de densité $f(x)$, avec comme graphe de f une parabole, par exemple, pour un spaghetti de longueur 1, on peut choisir :
$f(x) = kx(1-x)$ pour $x \in [0,1]$.
On doit donc avoir $\int_0^1 f(t)dt = k/6 = 1$ donc $k = 6$.
On suppose donc que $f(x) = 6x(1-x)$ pour $x \in [0,1]$
On suppose que l'on recoupe le morceau le plus grand.
On a alors :
Soit $F(x) = \int_0^x f(t)dt = 3x^2 - 2x^3$.
Si U est une variable aléatoire uniforme (donné par exemple par la fonction `rand()` de `Xcas`) on a :
$X = F^{-1}(U)$ et,
$Proba(U < F(x)) = Proba(F^{-1}(U) < x) = Proba(X < x) = F(x)$.
Pour déterminer `X` selon cette loi, on cherche `x` vérifiant $\texttt{x} = \texttt{F}^{-1}(\texttt{u})$.
Dans le programme ci-dessous on note g la fonction F et on écrit :

```
g(v):=3*v^2-2*v^3;
u:=evalf(rand(2^30)/2^30);
j:=0.1;
while (x> g(j)){j:=j+0.1;}
x:=j-0.05;
```

on peut aussi écrire :

```
g(v):=3*v^2-2*v^3;
u:=evalf(rand(2^30)/2^30);
solve(g(x)=u,x)
```

Ainsi, si $\texttt{F(J}-\texttt{0.1)} < \texttt{U} < \texttt{F(J)}$ on a $\texttt{J}-\texttt{0.1} < \texttt{F}^{-1}(\texttt{U}) = \texttt{X} < \texttt{J}$.
Pour choisir `y` dans l'intervalle `[0;a]` selon cette loi, la densité de probabilité

correspondante est $f_a(t) = \frac{6t(a-t)}{a^3}$ et $F_a(t) = F(t/a)$.
Pour déterminer la loi de Y, lorsqu'on coupe un spaghetti de longueur x, selon cette loi, on écrit :

```
y:=evalf(rand(2^30)/2^30);
j:=0.1;
while (y> g(j)){j:=j+0.1;}
y:=(j-0.05)*x;
```

Si $x \in [\frac{1}{2}; 1]$, et $y \in [0; x]$, la probabilité d'avoir un triangle est que :
$y \in [x - \frac{1}{2}; \frac{1}{2}]$.
Cherchons la probabilité d'avoir :
$y \in [x - \frac{1}{2}; \frac{1}{2}]$, sachant que $x \in [\frac{1}{2}; 1]$, et $y \in [0; x]$.
On a :
$Proba(y \in [x - \frac{1}{2}; \frac{1}{2}]) = F_x(\frac{1}{2}) - F_x(x - \frac{1}{2}) = F(\frac{1}{2}x) - F((2x-1)/2x) = \frac{-2x^3+3x-1}{2x^3}$
Pour le calcul théorique de la probabilité d'avoir un triangle, on utilise la symétrie : en effet, on a soit $x \in [\frac{1}{2}; 1]$ et $y \in [0; x]$ soit, $x \in [0; \frac{1}{2}]$ et $y \in [x; 1]$ (donc on fait le calcul de cette probabilité lorsque $x \in [\frac{1}{2}; 1]$ et on multiplie par 2 cette probabilité pour avoir le résultat).
Donc la probabilité d'avoir un triangle, avec ce choix de découpage est :
$2\int_{\frac{1}{2}}^{1} f(x)(F(\frac{1}{2}x) - F((2x-1)/2x))dx$
Puisque $F(\frac{1}{2}x) - F((2x-1)/2x) = \frac{-2x^3+3x-1}{2x^3}$, et que $f(x) = 6x(1-x)$, on tape :

```
normal(6*int((1-x)*(-2*x^3+3*x-1)/(x^2),x,1/2,1))
```

On obtient :

```
-24*log(1/2)-16
```

et avec la commande evalf on obtient :

```
0.635532333439
```

C'est à dire :
$6\int_{1/2}^{1} \frac{(1-x)(-2x^3+3x-1)}{x^2}dx = (-24\ln(1/2) - 16) \simeq 0.635532333439$
Voici le programme de simulation avec Xcas

```
spagb(n):={
//integrate(6*x*(1-x)*(g(0.5/x)-g(1-0.5/x)),x,0.5,1)
local x,y,a,b,t;
t:=0;
g(u):=3*u^2-2*u^3;
//Ab:=[];
for (k:=1;k<=n;k++){
x:=evalf(rand(2^30)/2^30);

j:=0.1;while (x> g(j)){j:=j+0.1;}
x:=j-0.05;
y:=evalf(rand(2^30)/2^30);
j:=0.1;while (y> g(j)){j:=j+0.1;}
```

```
if (x>0.5){
y:=(j-0.05)*x;
a:=y;
b:=x-y;
} else {
y:=(j-0.05)*(1-x)+x;
a:=x;
b:=y-x;
}
//Ab:=append(Ab,[x,y]);
if ((a<0.5) and (b<0.5) and (a+b>0.5)) {
t:=t+1;
}
}
return(evalf(t/n));
};
```

6.7 La ration de pain

En un pays lointain, le pain était limité à 200 grammes par personne et par jour. Le boulanger ne fabriquait donc que des pains de 200 grammes pour ses 1000 clients. Chaque matin, un vieux professeur allait chez le boulanger chercher sa ration quotidienne. Un jour il dit au boulanger :
- "Vous volez vos clients, les pains que vous vendez sont 1 pour cent plus petits qu'ils ne devraient l'être et vous devez donc donner à tous vos clients un pain gratuit tous les 100 jours".
- "Mais Monsieur, dit le boulanger tous les pains ne peuvent pas tous avoir le même poids ! Certains sont quelquefois, quelques pour cent plus lourds et d'autres quelques pour cent plus légers !"
- "Depuis 100 jours, je pèse mon pain et j'ai obtenu une courbe de Gauss de moyenne un poids de 198.04 grammes, et c'est inadmissible ! Si vous ne modifiez pas le poids de vos pains, je le signalerai à la répression des fraudes"
- "Je vous promets de faire le nécessaire dès demain"
Le Boulanger ne voulait pas changer sa manière de faire et chaque matin, avant le passage du professeur, il choisissait un pain et le pesait : s'il pesait au moins 200 grammes il le mettait de côté pour le professeur, sinon il en choisissait un autre jusqu'à obtenir un pain d'au moins 200 grammes qu'il mettait de côté pour le professeur. Cent jours plus tard, le professeur dit :
- "Vous n'avez rien changé ! vous continuez à voler vos clients"
"Mais Monsieur, vous ne pouvez rien prouver car tous les pains que je vous ai donnés ces derniers mois pesaient tous au moins 200 grammes"
- "Justement si !"
Pouvez-vous trouver l'argument du professeur ?

6.7.1 Simulation avec une loi binomiale

Simulation de la fabrication des pains

Voici un programme qui fabrique `n` pains dont le poids en grammes est dans l'intervalle [192,204] et suit une loi binomiale de moyenne 198.
On utilise pour cela la loi binomiale : on peut se servir de la simulation du parcours sur un axe (cf 6.2.1) : pour fabriquer un pain on ajoute à 192 le nombre de faces obtenu quand on lance 12 fois de suite une pièce.
Contrairement, au parcours on ne classe pas les pains par leur poids, on met les poids des `n` pains fabriqués dans une liste `A`.

```
pain(n) :={
  local T,r,A,j,k;
  A:=makelist(x->192,1,n,1);
  for (j:=0;j<n;j++){
    r:=rand(2);
    T:=0;
    for (k:=0;k<12;k++){
      if (r==1){
        T:=T+1;
      }
    r:=rand(2);
    }
  A[j]:=A[j]+T;
  }
return(A);
};
```

Un exemple de fournée de 100 pains :

```
pain(100) = [197,199,199,198,197,199,196,199,199,
196,199,198,195,196,197,198,199,198,197,198,197,201,202,196,200,
201,197,195,200,200,197,198,196,199,197,196,197,201,198,198,199,
201,202,201,199,201,197,200,197,199,196,201,201,197,199,199,195,
198,199,199,198,198,200,195,198,197,199,200,200,196,195,199,197,
200,200,201,200,199,198,198,200,199,199,198,197,197,200,199,198,
195,199,198,198,198,197,200,195,198,200,196]
```

En théorie on aurait du avoir :
0 pains de poids 193 g et 0 de poids 203 g
2 pains de poids 194 g et 2 de poids 202 g
5.5 pains de poids 195 g et 5.5 de poids 201 g
12 pains de poids 196 g et 12 de poids 200 g
19 pains de poids 197 g et 19 de poids 199 g
23 pains de poids 198 g

Simulation de la pesée du professeur

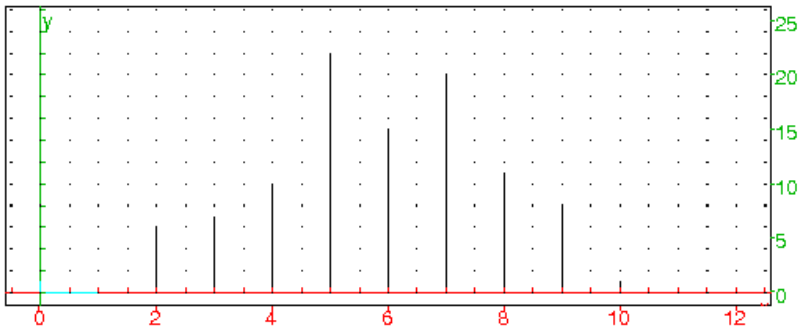
Remarque
Pour la simulation, on ne refait pas le pain tous les jours !

`A:=pain(n);`
est mis au début, et non dans la boucle (là où il est commenté), car sinon le programme est trop long à l'exécution.
`p` représente le nombre de jours pendant lesquels on effectue la pesée et `pj` représente le poids obtenu chaque jour.
On classe ces poids dans `P` : `P[0]` est égal au nombre de pains de poids 192 grammes, `P[1]` est égal au nombre de pains de poids 193 grammes... `m` est alors la moyenne des poids obtenus.

```
client(p,n):={
local pj,A,P,D,j,k,m;
P:=makelist(x->0,0,12,1);
A:=pain(n);
S:=0;
for (k:=0;k<p;k++){
    //A:=pain(n);
    j:=rand(n);
    pj:=A[j];
    S:=S+pj;
    pj:=pj-192;
    P[pj]:=P[pj]+1;
};
m:=evalf(S/p);
print(P);
print(m);
xyztrange(-0.2,12.2,-1,36,-10,10,-10,-10,-0.2,12.2,
          -1,36,1);
return segment(0,i*P[0]),segment(1,1+i*P[1]),
 segment(2,2+i*P[2]),segment(3,3+i*P[3]),
 segment(4,4+i*P[4]),segment(5,5+i*P[5]),
 segment(6,6+i*P[6]),segment(7,7+i*P[7]),
 segment(8,8+i*P[8]),segment(9,9+i*P[9]),
 segment(10,10+i*P[10]),segment(11,11+i*P[11]),
 segment(12,12+i*P[12]);
};
```

On tape :
`client(100,1000)`
On obtient écrit en bleu :
`P:[0,0,6,7,10,22,15,20,11,8,1,0,0]`
`m=197.83`
En théorie on doit avoir :
3 pains de poids 193 g et 3 de poids 203 g
16 pains de poids 194 g et 16 de poids 202 g
54 pains de poids 195 g et 54 de poids 201 g
121 pains de poids 196 g et 121 de poids 200 g
193 pains de poids 197 g et 193 de poids 199 g
225 pains de poids 198 g
Voici le "diagramme en bâtons" de la distribution des pains que le professeur a ob-

tenu :

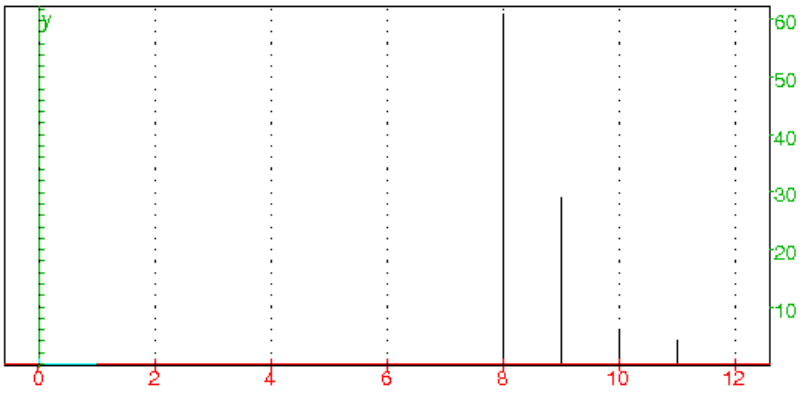


Il suffit de rajouter la ligne dans le programme précédent (au bon enfroit !) :
`while (pj<200) {j:=rand(n); pj:=A[j];}`
qui permet de choisir un pain de poids supérieur ou égal à 200 grammes.

```
chouchou(p,n):={
  local pj,A,P,S,j,k,m;
  P:=makelist(x->0,0,12,1);
  A:=pain(n);
  S:=0;
  for (k:=0;k<p;k++){
    //A:=pain(n);
    j:=rand(n);
    pj:=A[j];
    //si le poids pj<200g on prend un autre pain
    while (pj<200) {j:=rand(n); pj:=A[j];}
    S:=S+pj;
    pj:=pj-192;
    P[pj]:=P[pj]+1;
  };
  m:=evalf(S/p);
  print(P);
  print(m);
  xyztrange(-0.2,12.2,-1,62.5,-10,10,-10,-10,-0.2,
            12.2,-1,60,1);
  return segment(0,i*P[0]),segment(1,1+i*P[1]),
    segment(2,2+i*P[2]),segment(3,3+i*P[3]),
    segment(4,4+i*P[4]),segment(5,5+i*P[5]),
    segment(6,6+i*P[6]),segment(7,7+i*P[7]),
    segment(8,8+i*P[8]),segment(9,9+i*P[9]),
    segment(10,10+i*P[10]),segment(11,11+i*P[11]),
    segment(12,12+i*P[12]);
};
```

On tape : `chouchou(100,1000)`
On obtient écrit en bleu :
`P=[0,0,0,0,0,0,0,0,61,29,6,4,0]`

`m=200.53`
et le "diagramme en bâtons" de la distribution des pains que le professeur a obtenu a été le suivant :



6.7.2 Analyse du résultat

6.7.3 Simulation avec une loi gaussienne

On considère que le poids des pains du boulanger suit une loi gaussienne de moyenne 198 grammes et d'écart type σ grammes.

Simulation avec rand()

Le nombre de pains de poids x est donc approximativement de :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-198}{\sigma})^2}$$

Pour $\sigma = 2$ on a :
$f(192) = f(204) = 0.002$ $f(193) = f(203) = 0.009$ $f(194) = f(202) = 0.027$
$f(195) = f(201) = 0.065$ $f(196) = f(200) = 0.121$ $f(197) = f(199) = 0.176$
$f(198) = 0.200$

et on a bien $f(192) + f(193) + ... + f(204) = 1$

Voici une fournée de 1000 pains selon cette répartition représentée par la liste `G` des poids de ces 1000 pains.

```
//f(x):=0.5/sqrt(2*pi)*exp(-0.125*(x-198)^2)
G:=[192,204,192,204];
for(j:=0;j<9;j++){
G:=concat(G,[193,203]);
};
for(j:=0;j<27;j++){
G:=concat(G,[194,202]);
};
for(j:=0;j<65;j++){
G:=concat(G,[195,201]);
};
```

```
for(j:=0;j<121;j++){
G:=concat(G,[196,200]);
};
for(j:=0;j<176;j++){
G:=concat(G,[197,199]);
};
for(j:=0;j<200;j++){
G:=concat(G,[198]);
};
```

La fournée `G` de 10000 pains se trouve dans le fichier `painG`.
Le professeur achéte p pains (par exemple $p = 100$).

```
prof(p):={
local pj,A,P,j,m,S,k;
P:=makelist(0,0,12,1);
A:=G;
S:=0;
for (k:=0;k<p;k++){
    j:=rand(1000);
    pj:=A[j];
    S:=S+pj;
    pj:=pj-192;
    P[pj]:=P[pj]+1;
};
m:=evalf(S/p);
print(P);
print(m);
return segment(0,i*P[0]),segment(1,1+i*P[1]),
 segment(2,2+i*P[2]),segment(3,3+i*P[3]),
 segment(4,4+i*P[4]),segment(5,5+i*P[5]),
 segment(6,6+i*P[6]),segment(7,7+i*P[7]),
 segment(8,8+i*P[8]),segment(9,9+i*P[9]),
 segment(10,10+i*P[10]),segment(11,11+i*P[11]),
 segment(12,12+i*P[12]);
};
```

Le pain du professeur a un poids (en grammes) toujours supérieur ou égal à 200.

```
profchou(p):={
local pj,A,P,j,m,S,k;
P:=makelist(0,0,12,1);
A:=G;
S:=0;
for (k:=0;k<p;k++){
    j:=rand(1000);
    pj:=A[j];
    while (pj<200) {j:=rand(n); pj:=A[j];}
    S:=S+pj;
    pj:=pj-192;
```

```
    P[pj]:=P[pj]+1;
};
m:=evalf(S/p);
print(P);
print(m);
return segment(0,i*P[0]),segment(1,1+i*P[1]),
 segment(2,2+i*P[2]), segment(3,3+i*P[3]),
 segment(4,4+i*P[4]),segment(5,5+i*P[5]),
 segment(6,6+i*P[6]),segment(7,7+i*P[7]),
 segment(8,8+i*P[8]), segment(9,9+i*P[9]),
 segment(10,10+i*P[10]),segment(11,11+i*P[11]),
 segment(12,12+i*P[12]);
};
```

Simulation avec `randnorm()`

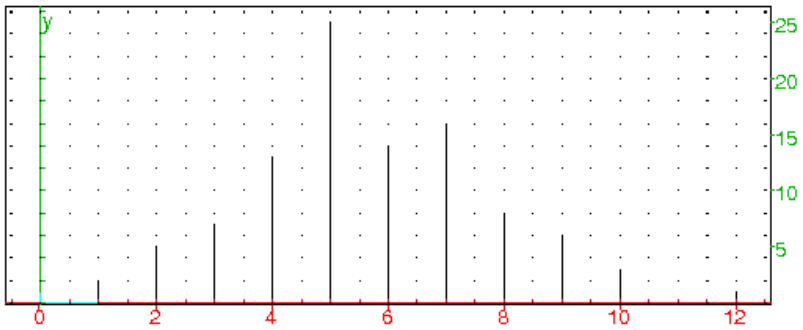
Pour simuler une fournée du boulanger, on va utiliser `randnorm(198,2)`
Le professeur achéte p pains (par exemple $p = 100$)

```
profnorm(p):={
local pj,P,j,m,S,k;
P:=makelist(0,0,12,1);
S:=0;
for (k:=0;k<p;k++){
    pj:=floor(randnorm(198,2));
    S:=S+pj;
    pj:=pj-192;
    if (pj<0) {P[0]:=P[0]+1;}
          else
          {if (pj>12) {P[12]:=P[12]+1;}
               else
               {P[pj]:=P[pj]+1;}
           }
};
m:=evalf(S/p);
return([P,m,segment(0,i*P[0]),segment(1,1+i*P[1]),
 segment(2,2+i*P[2]),segment(3,3+i*P[3]),
 segment(4,4+i*P[4]),segment(5,5+i*P[5]),
 segment(6,6+i*P[6]),segment(7,7+i*P[7]),
 segment(8,8+i*P[8]),segment(9,9+i*P[9]),
 segment(10,10+i*P[10]),segment(11,11+i*P[11]),
 segment(12,12+i*P[12])]);
};
```

Puis on tape par exemple :
`profnorm(100)` écrit en bleu :
`P=[0,0,0,0,0,0,0,0,61,29,6,4,0]`
`m=200.53`
et le "diagramme en bâtons" de la distribution des pains On obtient écrit en bleu :

```
P:[0,2,5,7,13,25,14,16,8,6,3,0,1]
m:197.66
```

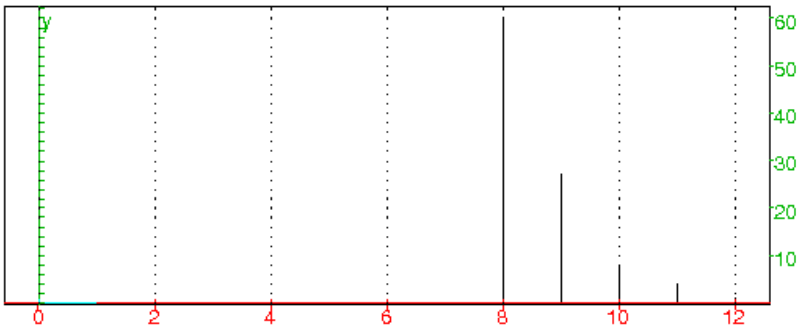
et le "diagramme en bâtons" de la distribution des pains selon la loi normale de moyenne 198 et d'écart-type 2 :



On écrit ensuite le programme `profchounorm` pour simuler le poids du pain du professeur lorsque ce pain a toujours un poids (en grammes) supérieur ou égal à 200.

```
profchounorm(p):={
  local pj,P,j,m,S,k;
  P:=makelist(0,0,12,1);
  S:=0;
  for (k:=0;k<p;k++){
    pj:=floor(randnorm(198,2));
    while (pj<200) {pj:=floor(randnorm(198,2));}
    S:=S+pj;
    pj:=pj-192;
    if (pj<0)
     {P[0]:=P[0]+1;}
    else
     {if (pj>12) {P[12]:=P[12]+1;}
      else
       {P[pj]:=P[pj]+1;}
     };
  };
  m:=evalf(S/p);
  print(P);
  print(m);
  return segment(0,i*P[0]),segment(1,1+i*P[1]),
  segment(2,2+i*P[2]),segment(3,3+i*P[3]),
  segment(4,4+i*P[4]),segment(5,5+i*P[5]),
  segment(6,6+i*P[6]),segment(7,7+i*P[7]),
  segment(8,8+i*P[8]),segment(9,9+i*P[9]),
  segment(10,10+i*P[10]),segment(11,11+i*P[11]),
  segment(12,12+i*P[12]);
};
```

On valide ce programme, puis on tape par exemple :
`profchounorm(100)`
On obtient écrit en bleu :
`P=[0,0,0,0,0,0,0,0,60,27,8,4,1]`
`m=200.59`
et le "diagramme en bâtons" de la distribution des pains selon la loi normale de moyenne 198 et d'écart-type 2, pour des poids $p >= 200$.



Index

Table des matières