

Première partie

Statistique descriptive

1 statistiques à une variable

1.1 vocabulaire, représentation

La statistique est l'étude des **populations**, dont les éléments sont des **individus** ; le plus souvent on n'étudie pas toute la population, mais seulement un **échantillon** de celle-ci. L'**effectif** d'un échantillon est le nombre d'individus qui le composent.

Plus précisément, on étudie certains **caractères** des individus, caractères qui peuvent être **qualitatifs** (par exemple le prénom, la nationalité, ...) ou **quantitatifs** (l'âge, la taille, les revenus mensuels...). Les caractères quantitatifs peuvent être **discrets** (la pointure de chaussures, le nombre de personnes au foyer, ...) ou **continus** (la taille, la superficie d'une région, ...).

Pour faciliter l'étude, en particulier des caractères continus, on peut regrouper les valeurs en **classes**, c'est à dire en intervalles deux à deux disjoints. La longueur d'un tel intervalle est appelé **amplitude** de la classe.

Par exemple, pour décrire la taille d'un adulte, on pourra considérer les intervalles $[0; 100[$, $[100, 110[$, ..., $[190, 200[$, $[200, +\infty[$, la première classe est d'amplitude 100, la dernière d'amplitude infinie alors que toutes les autres sont d'amplitude 10.

Une **série statistique** est un ensemble de couples (x_i, n_i) , où les x_i sont les valeurs prises du caractère et les n_i le nombre de fois où la valeur x_i apparaît.

L'effectif total de l'échantillon est donc $n = \sum_i n_i$.

On appelle **fréquence** d'apparition de x_i le nombre $f_i = n_i/n$.

exemple 1 : sur un échantillon de mille pièces tirées de la production journalière d'une usine, on compte le nombre de défauts constatés :

nombre de défauts	0	1	2	3	4
effectifs	570	215	140	60	15
fréquences	0.57	0.215	0.140	0.06	0.015

Ici les valeurs sont donc $x_0 = 0, x_1 = 1, \dots, x_4 = 4$ d'effectifs respectifs $n_0 = 570, n_1 = 215, n_2 = 140, n_3 = 60, n_4 = 15$.

L'effectif total est de 1000, ce qui permet de calculer facilement les fréquences situées sur la dernière ligne du tableau.

On peut imaginer de multiples représentations graphiques pour une série statistique : diagramme en batons, camemberts... Une seule présente une petite difficulté : l'histogramme, utilisé pour représenter par une suite de rectangle des résultats regroupés en classes.

exemple 2 : un technicien mesurant des tiges métalliques obtient les valeurs suivantes :

longueur (mm)	[330; 340[[340; 343[[343; 345[[345; 350[[350; 360[
effectifs	57	195	204	30	14
fréquences	$\frac{57}{500} \simeq 0.11$	$\frac{195}{500} \simeq 0.39$	$\frac{204}{500} \simeq 0.41$	$\frac{30}{500} \simeq 0.06$	$\frac{14}{500} \simeq 0.03$

Pour tracer l'histogramme on place en abscisse les différentes classes, ici $[330; 340[$, $[340; 343[$, $[343; 345[$, $[345; 350[$ et $[350; 360[$.

Pour chaque classe on calcule alors la hauteur du rectangle correspondant : c'est l'effectif divisé par l'amplitude de la classe. Ici, on trouve donc respectivement 5.7, 65, 102, 6 et 14.

Alors l'aire de chaque rectangle est proportionnelle à l'effectif de chaque classe. Attention, c'est bien l'**aire**, et non la hauteur, qui est proportionnelle à l'effectif !

1.2 caractéristiques de position

le mode est la valeur la plus fréquente d'une série statistique ; pour une série répartie en classe on parle de **classe modale**. Le mode n'est pas forcément unique.

Dans l'exemple 1, le mode est 0 ; dans l'exemple 2, la classe modale est l'intervalle $[343; 345[$.

la médiane est la valeur Me telle que la moitié des individus de la série ont un caractère inférieur ou égal à Me et l'autre moitié un caractère supérieur ou égal.

Quand les données sont regroupées en classes on parle de classe modale.

Dans l'exemple 1, la médiane est 0 ; dans l'exemple 2, la classe médiane est [340; 343].

la **moyenne** d'une série statistique (x_i, n_i) est le nombre

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n} = f_1x_1 + f_2x_2 + \dots + f_px_p$$

Dans l'exemple 1, la moyenne est donc $0.215 \times 1 + 0.14 \times 2 + 0.06 \times 3 + 0.015 \times 4 = 0.735$.

Dans le cas où les valeurs sont regroupées en classes, on peut déterminer une valeur approchée de la moyenne en remplaçant pour le calcul chaque classe par son milieu : dans l'exemple 2, la moyenne est $0.114 \times 335 + 0.39 \times 341.5 + 0.408 \times 344 + 0.06 \times 347.5 + 0.028 \times 355 = 342.517$. Ce calcul n'est satisfaisant que si les classes sont « bien choisies », i.e d'amplitudes et d'effectifs comparables.

1.3 caractéristiques de dispersion

On souhaite estimer si les valeurs d'une série statistique sont regroupées ou non autour de la valeur moyenne.

La caractéristique de dispersion la plus élémentaire et la plus facile à calculer est l'**étendue**, différence entre la plus grande et la plus petite des valeurs. On peut aussi considérer la moyenne des écarts à la moyenne de chaque valeur.

Mais on préfère utiliser la **variance** et l'**écart-type** qui pour chaque valeur :

la variance de la série statistique est

$$\frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{n} =$$

$$f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \dots + f_p(x_p - \bar{x})^2.$$

L'écart-type de la série statistique, noté σ' , est la racine carrée de sa variance.

Ainsi, la variance est simplement « la moyenne des carrés des écarts à la moyenne ».

Dans l'exemple 1, on trouve une étendue de 4, une variance égale à $0.57 \times (0.735)^2 + 0.215 \times (1 - 0.735)^2 + 0.14 \times (2 - 0.735)^2 + 0.06 \times (3 - 0.735)^2 + 0.015 \times (4 - 0.735)^2 \simeq 1.015$ et donc $\sigma' \simeq 1.01$.

remarque 1 : on peut développer l'expression donnant la variance et obtenir après calcul la formule $\sigma'^2 = \overline{x^2} - \bar{x}^2$.

remarque 2 : la formule donnant la variance fait bien intervenir un n , et pas un $n - 1$, cf 10.2 pour plus de précisions.

2 statistiques à deux variables

Une **série statistique double** est une série de n mesures de deux quantités x et y : x_1, x_2, \dots, x_n et y_1, y_2, \dots, y_n .

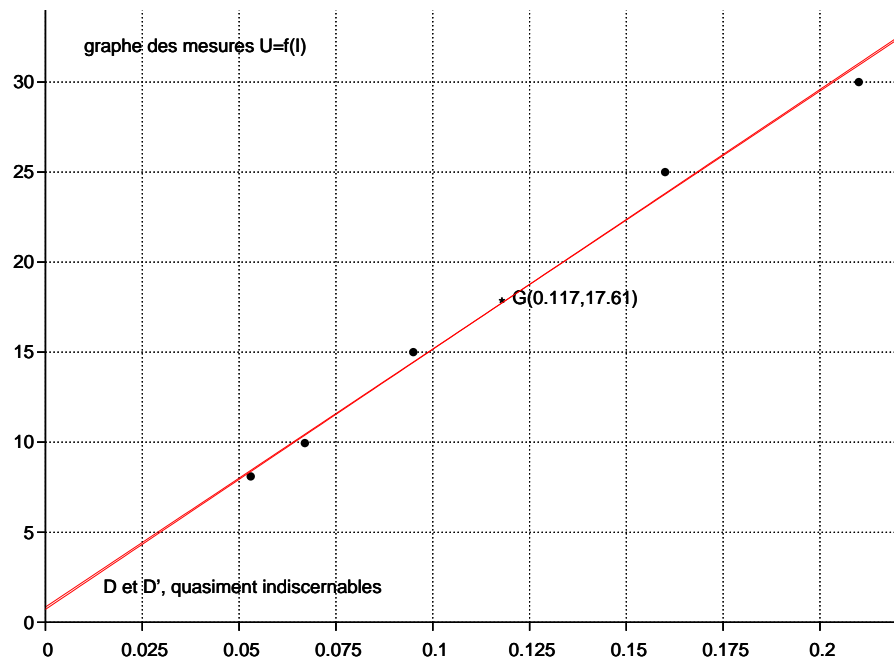
On s'intéresse surtout à la question de savoir si les mesures des deux valeurs sont indépendantes ou non. Si elles ne sont pas indépendantes, comment sont-elles reliées ?

2.1 Droite de régression linéaire

exemple 1 : On mesure simultanément le courant et l'intensité aux bornes d'une résistance. On obtient les valeurs :

intensité en ampères	0.053	0.067	0.095	0.16	0.21
tension en volts	8.1	9.95	15	25	30

On peut représenter ces mesures par un **nuage de points** $M_i(x_i, y_i)$; $G(\bar{x}, \bar{y})$ est appelé **point moyen**. Ici, on trouve pour point moyen $G(\bar{x} = 0.117, \bar{y} = 17.61)$.



Sur cet exemple, que constate-t-on ? Les mesures semblent indiquer qu'il y a une relation linéaire entre les valeurs x du courant et y de la tension, i.e que l'on peut écrire de manière « presque » exacte $y = ax + b$. Mais comment choisir les « meilleurs » a et b ?

On peut bien sûr tracer à la main une droite qui passe au plus près des points, puis déterminer par lecture graphique son coefficient directeur a et son ordonnée à l'origine b . Cette méthode est tout à fait valable, en particulier pour des valeurs obtenues en TP et tracées à la main !

Mais nous allons voir une méthode calculatoire plus systématique (et plus utilisable lors d'un traitement informatique des données) : la **méthode des moindres carrés**.

On commence par déterminer les caractéristiques de chacune des séries : les moyennes $\bar{x} = 0.117$ et $\bar{y} = 17.61$ déjà calculées, et les écarts-types $\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = 0.0593$ et $\sigma_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}} = 8.53$.

On définit alors la **covariance** de la série des (x_i, y_i) par la formule

$$\sigma_{x,y} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

On peut développer l'expression définissant la covariance $\sigma_{x,y} = \frac{1}{n} \sum_{i=1}^n (x_i y_i - \bar{x} y_i - \bar{y} x_i + \bar{x} \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \times \frac{1}{n} \sum_{i=1}^n y_i - \bar{y} \times \frac{1}{n} \sum_{i=1}^n x_i + \bar{x} \bar{y} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} + \bar{x} \bar{y}$, et obtenir ainsi une autre expression de la covariance comme moyenne des produits moins produit des moyennes :

$$\sigma_{x,y} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}.$$

En utilisant l'une ou l'autre de ces formules, on trouve ici $\sigma_{x,y} = 0.503$.

On appelle alors **droite de régression de y en x** la droite $D : y = ax + b$ passant par G et de coefficient directeur

$$a = \frac{\sigma_{x,y}}{\sigma_x^2}.$$

C'est la droite D pour laquelle la somme des $M_i P_i^2$ est minimale, les P_i étant les points de D d'abscisse x_i .

Ici, $D : y = 143.27x - 0.847$.

De même la **droite de régression de x en y** $D' : x = a'y + b'$ passant par G et de coefficient $a' = \frac{\sigma_{x,y}}{\sigma_y^2}$ minimise la somme $\sum_i M_i Q_i^2$ où les Q_i sont les points de D' d'ordonnée y_i .

Ici $D' : x = 6.927 \times 10^{-3} y - 0.00498$, soit $y = 144.36x + 0.719$.

On constate sur cet exemple que les deux droites sont quasiment indiscernables, et la loi théorique $U = RI$ (soit ici $y = Rx$) semble a peu près vérifiée, avec une valeur de R proche de 143 ou 144 Ω .

exemple 2 : dans un SAV, on note pour chaque appareil défectueux l'heure d'arrivée et le temps d'atelier nécessaire à la réparation. Dans ce cas, il est probable que le graphique ressemble à un nuage de points d'apparence aléatoire, car les deux caractéristiques n'ont probablement aucun lien entre elles. Les droites D et D' ne coïncideront pas du tout.

Un outil numérique permet d'estimer si deux variables sont liées ou pas par une relation linéaire :

le **coefficient de corrélation** $r = \frac{\sigma_{x,y}}{\sigma_x \sigma_y}$.

r est toujours compris entre -1 et 1 . S'il vaut ± 1 , les droites D et D' sont confondues, et plus il est proche de ± 1 , plus les points (x_i, y_i) semblent alignés : on dit qu'il y a une bonne corrélation linéaire entre les quantités x et y .

Ici, $r = 0.994$. En pratique, on commence à considérer une valeur $|r| > 0.7$ comme significative d'une corrélation linéaire.

remarque 1 : le fait que r soit proche de 0 n'indique pas qu'il n'y a aucune corrélation entre les variables, mais seulement qu'il n'y a pas de corrélation **linéaire**. En effet, on rencontre souvent des relations du type $y = 1/(ax + b)$, $y = kx^x$, $y = kx^c$, ... et tout ce qui précède est inadaptable pour traiter ces corrélations non linéaires.

Mais il suffit d'un changement de variable pour se ramener au cas linéaire : voir les exercices.

remarque 2 : une bonne corrélation ($|r|$ proche de 1) ne signifie pas qu'il existe une relation de cause à effet entre deux phénomènes ; une étude physique plus approfondie sera nécessaire pour le savoir : r n'est qu'un indice pour le technicien.

On peut illustrer cela par un troisième exemple : si à Grenoble on note, chaque jour de l'hiver, la hauteur de neige tombée et la température de l'air, on observera une corrélation : il neige très peu les jours très froids. Peut-on en déduire que le froid empêche la neige de tomber ? En fait, il n'y a pas de relation physique directe : en altitude, ou dans d'autres régions du globe, il peut neiger avec des températures très froides. Mais en France, ce sont les anticyclones sibériens qui amènent le froid vif...et l'air sec. Il y a bien corrélation entre les phénomènes, mais pas de lien de cause à effet.

Complément : démonstration de la formule des moindres carrés

On considère un nuage de points $M_i(x_i, y_i)$ et une droite $D : y = ax + b$.

Une manière d'exprimer le fait que la droite D passe au plus près des points M_i est de demander que le produit des carrés des écarts d'ordonnée $y_i - (ax_i + b)$ soit le plus petit possible : on souhaite trouver a et b tels que la quantité $\sum_{i=1}^n ((ax_i + b) - y_i)^2$ soit minimale.

Il s'agit d'une application de deux variables (a, b) positive et à valeurs réelles. De plus cette application est dérivable : si elle admet un minimum, on sait qu'en celui-ci les dérivées partielles doivent s'annuler. On peut calculer ces dérivées partielles :

$$\frac{\partial(\sum_{i=1}^n (ax_i + b - y_i)^2)}{\partial a} = 2 \sum_{i=1}^n x_i(ax_i + b - y_i) \text{ et } \frac{\partial(\sum_{i=1}^n (ax_i + b - y_i)^2)}{\partial b} = 2 \sum_{i=1}^n (ax_i + b - y_i).$$

Ainsi la condition nécessaire de minimum s'exprime par les deux équations $\sum_{i=1}^n x_i(ax_i + b - y_i) = 0$ et $\sum_{i=1}^n (ax_i + b - y_i) = 0$, soit encore $a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i - \sum_{i=1}^n x_i y_i = 0$ et $a \sum_{i=1}^n x_i + nb - \sum_{i=1}^n y_i = 0$. En divisant par n les deux équations on obtient le système :

$$\begin{cases} a \frac{1}{n} \sum_{i=1}^n x_i^2 + b \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i y_i \\ a \bar{x} + b &= \bar{y} \end{cases}$$

que l'on résoud en enlevant \bar{x} fois la deuxième ligne à la première :

$$\begin{cases} a(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2) &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} \\ a \bar{x} + b &= \bar{y} \end{cases}$$

Mais on reconnaît dans la première équations les expressions de la variance $\overline{x^2} - \bar{x}^2$ et de la covariance déjà étudiées. Ainsi, $a = \frac{\sigma_{x,y}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2} = \frac{\sigma_{x,y}}{\sigma_x^2}$, et la deuxième équation exprime bien le fait que la droite D passe par le point moyen du nuage de points.

Des considérations intuitives montrent (la démonstration rigoureuse dépasse le niveau de ce cours) qu'il doit bien exister une droite réalisant ce minimum, et on vient de prouver que ce n'est possible que pour les valeurs de a et b définies plus haut.

2.2 Régression linéaire passant par l'origine

Dans le cas de la résistance, si l'on connaît préalablement à l'expérience la loi d'Ohm, on sait que la relation à chercher est du type $U = RI$, et l'on peut souhaiter simplement déterminer le meilleur coefficient R correspondant aux mesures, avec une relation sans terme constant.

Si l'on recherche une relation de la forme $y = ax$, on prendra alors la valeur

$$a = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\overline{xy}}{\overline{x^2}}.$$

Dans cet exemple, $\sum_i x_i^2 = 0.086023$ et $\sum_i x_i y_i = 12.82095$, d'où $a \simeq 149.04 \Omega$.

Le choix entre ces deux méthodes dépendra des circonstances : dans cette expérience, le fait qu'une régression linéaire « classique » ne donne pas $b = 0$ peut s'expliquer par le fait que la résistance n'est pas une résistance « pure » (le terme b ayant alors une signification physique) ou bien par le fait qu'une ou plusieurs mesures sont peu précises.

Deuxième partie

Probabilités

3 Combinatoire

Avant de commencer l'étude des probabilités, il est nécessaire d'apprendre à compter le nombre d'éléments des ensembles finis qui seront les événements étudiés : c'est le **dénombrement**.

Et un moyen pour y arriver est l'utilisation de quelques méthodes combinatoires, que nous détaillons ici.

3.1 cardinal

Si E est un ensemble fini, on appelle **cardinal** de E , et on note $\text{card}(E)$, le nombre de ses éléments.

exemple :

- Si $E = \{ \text{pique, trèfle, coeur, carreau} \}$, son cardinal est $\text{card}(E) = 4$.
- Si $E = \{1, 2, 3, 4, 5, 6\}$, $\text{card}(E) = 6$.
- Si E est l'ensemble des entiers, $\text{card}(E)$ est infini.
- Si E est l'ensemble des manières de placer Amandine, Bertrand et Cécile sur un banc, quel est le cardinal de E ?

On peut énumérer les possibilités, en indiquant l'ordre de droite à gauche : A, B, C , ou bien A, C, B , ou bien B, A, C ou bien B, C, A , ou bien C, A, B , ou bien C, B, A . Et donc $\text{card}(E) = 6$.

3.2 factorielle

Le dernier exemple est encore traitable directement ; mais si le nombre de personnes à classer augmente, on a intérêt à réfléchir un peu plus...par exemple, de combien de manières peut-on classer 80 étudiants de DUT Mesures Physiques ?

Il suffit de choisir, parmi les 80 étudiants, celui qui sera premier : on a 80 possibilités. Puis de choisir le second parmi les 79 restants : 79 possibilités. Puis le troisième, le quatrième, et ainsi de suite. On a donc $80 \times 79 \times 78 \times \dots \times 2 \times 1$ possibilités.

Cette démonstration se généralise, et

le nombre de façons de classer n éléments est $n! = n \times (n - 1) \times (n - 2) \times \dots \times 2 \times 1$.

Le nombre $n!$ se lit « **factorielle** n ».

Autrement dit, $n!$ est le nombre de permutations d'un ensemble à n éléments.

3.3 listes

Un autre problème classique de dénombrement est celui du tirage avec remise : typiquement, on dispose d'une urne avec 10 jetons, numérotés de 1 à 10, et on tire 3 fois de suite un jeton dont on note le numéro avant de le remettre dans l'urne. On a donc $10 \times 10 \times 10 = 10^3$ triplets de résultats possibles. Plus généralement, on démontre de même que

le nombre de manières de fabriquer une liste (ordonnée) de n -éléments tous compris entre 1 et p est p^n .

3.4 arrangements

Combien de podiums sont possibles pour une épreuve olympique avec 10 participants ? Il faut choisir la médaille d'or parmi les 10, puis la médaille d'argent parmi les 9 restants, puis la médaille de bronze parmi les 8 restants, soit $10 \times 9 \times 8 = 720$.

Plus généralement,

le nombre de manières de classer p personnes choisies parmi n est

$$A_n^p = n \times (n - 1) \times \dots \times (n - p + 1) = \frac{n!}{(n - p)!}.$$

3.5 combinaisons

A un concours de recrutement, 10 candidats se présentent pour trois postes. Combien de possibilités de recrutement ?

Ici, contrairement à l'exemple précédent, l'ordre importe peu : seul compte le fait d'être, ou non, recruté. On peut donc commencer par compter le nombre de manières de classer trois candidats parmi 10 : 720. Puis, ensuite, diviser ce nombre par le nombre de classements de ces trois recrutés entre eux, soit 3!. Et le nombre de recrutements possibles est donc de 120.

Plus généralement,

le nombre de manières de choisir p éléments parmi n est

$$\binom{n}{p} = \frac{n \times (n - 1) \times \dots \times (n - p + 1)}{p \times (p - 1) \times \dots \times 1} = \frac{n!}{p!(n - p)!}.$$

$\binom{n}{p}$ se prononce « p parmi n ». Les $\binom{n}{p}$ sont les coefficients binômiaux ; l'écriture largement utilisée dans le secondaire en France au siècle précédent est C_n^p (lire « C n p ») (attention à l'ordre : n est en haut dans $\binom{n}{p}$, en bas dans C_n^p)

On peut aussi voir $\binom{n}{p}$ comme le nombre de parties à p -éléments dans un ensemble à n éléments. Avec cette définition, il est alors clair que

$$\binom{n}{p} = \binom{n}{n-p}$$

car déterminer un ensemble à p -éléments revient exactement à déterminer son complémentaire, à $n-p$ éléments.

On a aussi :

$$\text{pour } n, p \text{ supérieurs ou égaux à } 1, \quad \binom{n}{p} = \binom{n-1}{p} + \binom{n-1}{p-1}.$$

En effet, si E est un ensemble à n éléments et $a \in E$ fixé, il y a $\binom{n-1}{p}$ parties de E à p éléments qui ne contiennent pas a , et $\binom{n-1}{p-1}$ qui contiennent a . Cette propriété est à la base de la construction du triangle de Pascal : le tableau suivant, où chaque nombre à partir de la ligne 1 est la somme des deux de la ligne du dessus sur la même colonne et la colonne de gauche,

	col.0	col.1	col.2	col.3	col.4	col.5	col.6	col.7	col.8	col.9	...
ligne 0 :	1										
ligne 1 :	1	1									
ligne 2 :	1	2	1								
ligne 3 :	1	3	3	1							
ligne 4 :	1	4	6	4	1						
ligne 5 :	1	5	10	10	5	1					
ligne 6 :	1	6	15	20	15	6	1				
ligne 7 :	1	7	21	35	35	21	7	1			
ligne 8 :	1	8	28	56	70	56	28	8	1		
ligne 9 :	1	9	36	84	126	126	84	36	9	1	
...											

contient, à l'intersection de la ligne n et de la colonne p la valeur de $\binom{n}{p}$.

C'est l'occasion de rappeler l'utilisation des $\binom{n}{p}$ pour le développement d'expressions algébriques :

pour tous a, b complexes,

$$(a + b)^n = \binom{n}{0} a^n b^0 + \binom{n}{1} a^{n-1} b^1 + \dots + \binom{n}{n-1} a^1 b^{n-1} + \binom{n}{n} a^0 b^n$$

et en particulier : $(a + b)^2 = a^2 + 2ab + b^2$, $(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$.

Citons enfin une dernière propriété des $\binom{n}{k}$ qui nous sera utile plus tard dans l'étude de la loi binomiale : si $n, k \geq 1$, on a $\frac{kn!}{k!(n-k)!} = \frac{n(n-1)!}{(k-1)!(n-k)!} = n \frac{(n-1)!}{(k-1)!(n-1-(k-1)!}$, donc

pour k, n entiers tels que $0 \leq k \leq n$, $k \binom{n}{k} = n \binom{n-1}{k-1}$.

4 Probabilités - définitions élémentaires

4.1 expériences aléatoires

On appelle **expérience aléatoire** une expérience dont les issues (les résultats) ne sont pas déterminés à l'avance.

L'ensemble, souvent noté Ω , de toutes les issues possibles est appelé **univers** ou **espace d'échantillonnage** de l'expérience.

exemples :

- On jette un dé à six faces, il y a six issues possibles : $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Un fabricant contrôle les produits sortis de ses chaînes : il y a deux issues possibles, ou bien le produit est sans défaut et peut être vendu, ou bien le produit présente des défauts et va être jeté : $\Omega = \{\text{conforme}, \text{non conforme}\}$.
- On choisit un nombre entier positif : $\Omega = \mathbb{N}$. A la différence des exemples précédents, Ω est ici infini. On parle là d'infini discret (les valeurs possibles sont toutes isolées).
- On choisit un point dans le plan. Là, $\Omega = \mathbb{R}^2$, et l'univers est aussi infini, mais cette fois-ci on parle d'infini continu.

4.2 événements

Un sous-ensemble, ou partie, de Ω est appelé un **événement**. L'ensemble des événements est donc l'ensemble noté $\mathcal{P}(\Omega)$ des parties de Ω .

En particulier Ω et \emptyset sont appelés **événement certain** et **événement impossible**.

Un ensemble qui ne contient qu'une seule issue est un **événement élémentaire**.

exemple : dans l'expérience du dé, « on obtient 1 » est un événement élémentaire, « on obtient un nombre impair » ou « on obtient un nombre inférieur ou égal à 4 » sont deux événements (non élémentaires).

4.3 opérations sur les événements

A et B sont deux événements. Alors :

L'événement contraire de A est son complémentaire dans Ω , noté \bar{A} ou $\Omega - A$, et se comprend « A n'est pas réalisé ».

La réunion de A et B est $A \cup B$ et se comprend « A ou B (ou les deux) sont réalisés ».

L'intersection de A et B est $A \cap B$ et se comprend « A et B sont réalisés simultanément ».

exemple : Dans l'expérience du dé, si $A = \{1, 3, 5\} =$ « on obtient un nombre impair » et $B = \{1, 2, 3, 4\} =$ « on obtient un nombre inférieur ou égal à 4 », alors $\bar{A} = \{2, 4, 6\} =$ « on obtient un nombre pair », $\bar{B} = \{5, 6\} =$ « on obtient un nombre strictement supérieur à 4 », $A \cup B = \{1, 2, 3, 4, 5\} =$ « on obtient un nombre impair ou

un nombre inférieur ou égal à 4 », et $A \cap B = \{1, 3\} =$ « on obtient un nombre impair inférieur ou égal à 4 ».

Deux événements sont **incompatibles** s'ils ne peuvent se produire simultanément, i.e si leur intersection $A \cap B$ est vide.

Bien sûr, un événement et son contraire sont toujours incompatibles.

4.4 loi de probabilité

On peut associer à une expérience aléatoire et à son univers une **probabilité** qui permet de quantifier le fait qu'un événement est « probable » ou est « peu probable ».

Une probabilité est une application p de $\mathcal{P}(\Omega)$ dans $[0; 1]$ telle que $p(\Omega) = 1$ et telle que si A et B sont deux événements incompatibles, $p(A \cup B) = p(A) + p(B)$.

Comme conséquence de cette définition, on a donc les propriétés suivantes :

$$0 \leq p(A) \leq 1 \text{ pour tout événement } A$$

$$p(\emptyset) = 0, \quad p(\Omega) = 1$$

$$p(\bar{A}) = 1 - p(A) \text{ pour tout événement } A$$

$$p(A \cup B) + p(A \cap B) = p(A) + p(B) \text{ pour tous événements } A, B.$$

remarque : pour chaque univers, on peut bien sûr imaginer plusieurs lois de probabilités différentes. Dans l'expérience de la fabrication d'objets, on peut imaginer qu'une chaîne de fabrication fonctionnant bien ait pour probabilité $p(\text{« conforme »}) = 0,95$ et $p(\text{« non conforme »}) = 0,05$, alors qu'une chaîne moins efficace ait des probabilités associées $p(\text{« conforme »}) = 0,75$ et $p(\text{« non conforme »}) = 0,25$.

On doit donc définir non seulement l'univers Ω , mais aussi la loi de probabilité p dont on le munit. Pour être rigoureux on parle donc en toute rigueur d'un **espace probabilisé** (Ω, p) .

4.5 le cas particulier des univers finis

Pour étudier un phénomène à l'aide des probabilités, on a besoin de connaître la loi de probabilité p , qui est une fonction de $\mathcal{P}(\Omega)$ dans $[0; 1]$, donc a priori on a besoin de connaître sa valeur sur chaque sous-ensemble de Ω . Mais en fait, quand Ω est fini, la connaissance de p sur chaque événement élémentaire suffit : si $A \subset \Omega$ est un événement quelconque, A est fini et on peut écrire $A = \{a_1, a_2, \dots, a_k\}$, donc $p(A) = p(\{a_1\}) + p(\{a_2\}) + \dots + p(\{a_k\})$.

Un cas encore plus particulier mais fondamental est le cas de l'équiprobabilité : sur un univers fini, on dit que la loi est **équiprobable** si tous les événements élémentaires ont la même probabilité. Dans ce cas, la probabilité de chaque événement élémentaire est simplement $1/\text{card}(\Omega)$, et la probabilité d'un événement A est :

$$p(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{\text{nombre d'éléments de } A}{\text{nombre d'éléments de } \Omega}.$$

exemples : dans des expériences de tirage au sort (pile ou face, dé, jeu de cartes, ...), sans précisions supplémentaires on supposera que le jeu n'est pas truqué, ce qui revient à dire que la loi est équiprobable : tous les événements élémentaires ont la même probabilité (une chance sur deux de faire pile, une chance sur deux de faire face ; une chance sur six de tirer 1, une chance sur six de tirer 2, etc... ; une chance sur 32 de tirer chacune des cartes du paquet).

4.6 le cas particulier des probabilités infinies discrètes

Ω infini est dit **discret** si on peut énumérer ses éléments, i.e si on peut écrire $\Omega = \{x_1, x_2, \dots\}$. Typiquement, cela correspond à des expériences dont le résultat est un entier naturel. Comme dans le cas précédent, on obtient la probabilité d'un événement quelconque comme somme (éventuellement infinie) des événements élémentaires qui le composent.

exemple : On considère la probabilité de désintégration des atomes d'un composé radioactif durant un intervalle de temps de longueur t fixé. Ici, $\Omega = \mathbb{N}$, et on montrera en exercice que $p_n = \frac{\Lambda^n t^n}{n!} e^{-\Lambda t}$. Quelle est la probabilité d'obtenir moins de 5 désintégrations ?

remarque : on n'a jamais équiprobabilité sur un ensemble infini discret. En effet, supposons les probabilités élémentaires p_i ($i \in \mathbb{N}$) toutes égales à un même α . Alors $p(\Omega) = \sum_{i=0}^{+\infty} \alpha$ ne peut valoir que 0 (si $\alpha = 0$) ou $+\infty$ (si $\alpha > 0$), mais en aucun cas 1.

4.7 le cas particulier des probabilités continues

Pour étudier les probabilités sur des univers continus infinis (par exemple : choix d'un nombre au hasard dans $[0; 1]$; durée de vie d'une voiture dans $[0; +\infty[$, ...) on va comme dans le cas fini partir d'« événements de base » qui permettent de reconstituer tous les événements, donc de calculer toutes les propriétés. Mais ici le problème est un peu plus délicat. En effet, en général, avec un univers continu la probabilité de chaque événement élémentaire est nulle, et cette information ne permet pas de déterminer la valeur de $p(A)$ pour tout événement A .

Pour ce qui suit on prend pour Ω un intervalle de \mathbb{R} (par exemple, $[0; 1]$, ou $[0; +\infty[$, ou \mathbb{R} lui-même...).

Ces événements de base vont ici être les segments $[a, b]$. Dans la plupart des cas, les événements qui nous intéressent pourront être décrits comme réunion, intersection, complémentaires, ... de segments et on pourra donc déduire ainsi leur probabilité de celles de ces segments grâce aux règles de calcul sur les probabilités.

Donc dans le cas des probabilités continues on associe à chaque probabilité une **densité** de probabilité qui est une fonction intégrable et positive, telle que $\int_{\Omega} f = 1$. Et la probabilité p est caractérisée par le fait que pour tout événement A , $p(A) = \int_A f$. En particulier,

$$p([a; b]) = \int_a^b f \text{ pour tout segment } [a; b].$$

exemple : le cas le plus simple est celui de la probabilité uniforme sur $[0; 1]$, qui correspond à l'expérience « on choisit au hasard un nombre compris entre 0 et 1, sans privilégier aucune valeur ».

Alors la densité correspondante est $f = 1$, et la probabilité d'obtenir un nombre entre a et b (pour $0 \leq a \leq b \leq 1$) est égale à $p([a, b]) = \int_a^b 1 = b - a$.

Ainsi, avec $a = 0$ et $b = 1$, la probabilité est 1 : le choix d'un nombre entre 0 et 1 donne à coup sûr un nombre entre 0 et 1 !

Au contraire, si $a = b$, on constate que la probabilité de choisir un nombre a donné à l'avance est nulle.

Si $a = 0,25$ et $b = 0,75$: on a une chance sur deux que le nombre choisi soit dans l'intervalle $[a; b]$ de longueur $1/2$.

4.8 probabilités conditionnelles

Soit (Ω, p) un espace probabilisé, et A un événement de probabilité non nulle.

On appelle « probabilité que B soit réalisé sachant que A l'est », ou plus simplement « probabilité de B sachant A », la quantité

$$p(B|A) = \frac{p(A \cap B)}{p(A)}.$$

Ainsi $p(\Omega|A)$ est toujours égal à 1.

exemple : on lance deux dés bien équilibrés. Quelle est la probabilité que la somme des résultats soit strictement supérieure à 10 sachant que l'un des dés a donné 6.

"somme > 10" = {(6, 6); (6, 5); (5, 6)}, "l'un des dés donne 6" est de cardinal 11, et l'intersection est de cardinal 3, donc la probabilité est 3/11.

Connaissant $p(A|B)$, on aimerait parfois connaître $p(B|A)$. C'est souvent possible en écrivant de deux manières différentes $p(A \cap B)$ à l'aide des définitions de $p(A|B)$ et de $p(B|A)$:

$$p(A \cap B) = p(A)p(B|A) = p(B)p(A|B).$$

exemple : 48 des 53 étudiants de *TI* ont eu la moyenne en automatique, et 14 des 26 étudiants *MCPC*. Quelle est la probabilité qu'un étudiant ayant eu la moyenne soit en *TI* ?

On a $p(S|MCPC) = 14/26$ et $p(S|TI) = 48/53$. De plus $p(S) = 62/79$ et $p(TI \cap S) = 48/79$. Donc $p(TI|S) = (48/79)/p(S) = (48/79) \times (79/62) = 48/62 = 24/31$.

La formule ci-dessus peut s'exprimer sous la forme plus directement utilisable suivante :

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}.$$

Dans les cas un peu plus compliqués, on peut avoir besoin de la formule de Bayes.

Considérons donc des événements incompatibles A_1, A_2, \dots, A_n , et un événement B qui ne peut se produire que si l'un des A_i se produit, les $p(B|A_i)$ étant connus. On cherche la probabilité pour que, B s'étant produit, A_k en soit la cause.

Commençons par remarquer que $p(B) = p(A_1 \cap B) + \dots + p(A_n \cap B)$; comme $p(A_k \cap B) = p(A_k)p(B|A_k)$, on obtient la formule des **probabilités totales** :

$$p(B) = p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + \dots + p(B|A_n)p(A_n).$$

Alors en écrivant $p(A_k|B) = p(A_k \cap B)/p(B) = p(B|A_k)p(A_k)/p(B)$, et en remplaçant $p(B)$ par la formule précédente, on obtient la **formule de Bayes** :

$$p(A_k|B) = \frac{p(A_k)p(B|A_k)}{\sum_{i=1}^n p(A_i)p(B|A_i)}$$

exemple : Un test de dépistage d'une maladie rare touchant une personne sur 10000 semble efficace : il détecte 99% des personnes infectées, avec seulement 0,5% de « faux positifs ». Quelle est la probabilité qu'une personne dont le test est positif soit effectivement malade ?
 $p(M|P) = p(P|M)p(M)/(p(P|M)p(M) + p(P|\bar{M})p(\bar{M})) \simeq 1,94\%$.

4.9 événements indépendants

On dit que deux événements A et B sont **indépendants** quand l'un des deux est de probabilité nulle, ou bien, quand les deux sont de probabilité non nulle, si le fait de savoir que l'un est réalisé n'influe pas sur la probabilité que l'autre le soit. Autrement dit deux événements de probabilité non nulle sont indépendants quand $p(B|A) = p(B)$ (ou de manière équivalente quand $p(A|B) = p(A)$).

Comme $p(B|A) = \frac{p(A \cap B)}{p(A)}$, cela équivaut à la

proposition : deux événements sont indépendants si et seulement si $p(A \cap B) = p(A)p(B)$.

remarque : ne pas confondre les deux notions d'événements indépendants et d'événements incompatibles ! Deux événements incompatibles ne sont **jamais** indépendants (sauf si les deux sont de probabilités nulle). En effet, si A et B sont incompatibles et que l'on sait que A est réalisé, justement, B ne peut pas se produire...il n'y a donc pas indépendance.

5 variables aléatoires

5.1 définition

Une **variable aléatoire** sur un espace probabilisé (Ω, p) est une application $X : \Omega \rightarrow \mathbb{R}$.

exemples : le résultat obtenu en jetant un dé, la mesure d'une tige métallique, le nombre de pannes quotidiennes d'une machine, ... sont des variables aléatoires.

Chaque variable aléatoire définit une probabilité p_X sur \mathbb{R} de la manière suivante :

pour tout $A \subset \mathbb{R}$, $p_X(A) = p(X^{-1}(A)) = p(X \in A) = p(\{x \in \Omega \mid X(x) \in A\})$.

$p_X(A)$ est donc la probabilité pour que la variable aléatoire X prenne ses valeurs dans un ensemble A fixé. Et la **loi** d'une variable aléatoire X est la donnée des valeurs $p_X(A)$ pour tous les $A \subset \mathbb{R}$.

La **fonction de répartition** F_X d'une variable aléatoire X est la fonction définie par $F_X(x) = p(X^{-1}(]-\infty; x[)) = p(X < x)$.

F_X est une fonction croissante, à valeurs dans $[0; 1]$. De plus, comme $p(\emptyset) = 0$ et $p(\Omega) = 1$, la limite de F_X en $-\infty$ est 0, sa limite en $+\infty$ est 1, et donc $p(X \geq x) = 1 - F_X(x)$.

Bien sûr, donner la loi d'une variable aléatoire en donnant chaque $p_X(A)$ est fastidieux, et souvent impossible si $X(\Omega)$ est infini. Mais nous allons voir que l'on peut être plus efficace en décrivant la loi seulement pour certains événements « simples » : les événements élémentaires (pour une variable discrète) ou les événements du type $] -\infty; x[$ (pour une variable continue), ce qui revient à donner seulement la fonction de répartition.

5.2 cas des univers finis ou infinis discrets

5.2.1 variables aléatoires

Si les valeurs prises par X forment un ensemble discret $X(\Omega) = \{x_1, x_2, \dots, x_n, \dots\}$ on peut décrire la loi de la variable X en donnant seulement les $p_X(\{x_i\})$, probabilité que les éléments de Ω « prennent » la valeur x_i . En effet pour tout événement A , $p_X(A)$ sera la somme des $p_X(\{a\})$ pour $a \in A$.

remarque : la notation $p_X(\{x_i\})$ est lourde et abrège en $p_X(x_i)$ ou $p(X = x_i)$.

Ainsi la probabilité de chaque événement élémentaire x_i est un nombre $p(X = x_i) = p_i$ tel que $\sum_{i=1}^{\text{card } \Omega} p_i = 1$, et décrire la loi de X revient à donner l'ensemble des valeurs p_i .

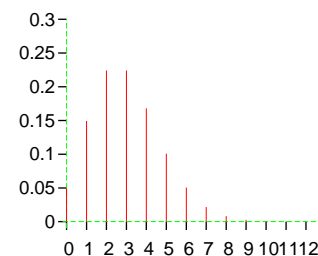
exemple 1 : on joue à pile ou face, et on appelle X la variable aléatoire qui vaut 0 en cas de résultat pile et 1 en cas de résultat face. Alors $\Omega = \{\text{pile}, \text{face}\}$, et $X(\Omega) = \{0; 1\}$.

Si la pièce est bien équilibrée on a donc $p_X(0) = p(\text{pile}) = 1/2$, $p_X(1) = p(\text{face}) = 1/2$.

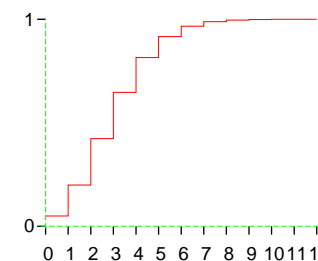
exemple 2 : on étudie une expérience de désintégration nucléaire. Soit X la variable aléatoire représentant le nombre d'atomes désintégrés en 1 seconde. Alors on a déjà vu que $p(X = k)$ vaut $\frac{\lambda^k}{k!} e^{-\lambda}$ où λ est une constante qui dépend du composé radioactif.

On peut représenter graphiquement les p_k et la fonction de répartition (avec ici $\lambda = 3$) :

valeurs des $p_k = p(X = k)$



fonction de répartition



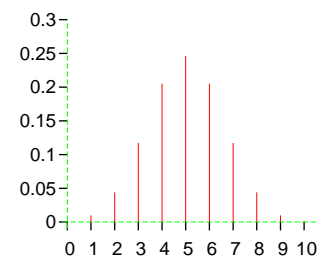
On peut déduire de ces probabilités élémentaires les probabilités des autres événements. Ainsi, $p(\text{le nombre d'atomes désintégrés est compris entre 2 et 4}) = p(X \in [2; 4]) = p(X = 2) + p(X = 3) + p(X = 4) = F_X(5) - F_X(2)$ (car $F_X(5) = p(X < 5) = p(X \leq 4)$, $F_X(2) = p(X < 2)$). De même, $p(\text{deux atomes au moins se sont désintégrés}) = p(X \in [2; +\infty[) = 1 - p(X < 2) = 1 - F_X(2)$.

exemple 3 : on joue dix fois de suite à pile ou face, et on compte le nombre de « face ». Ω est ici constitué des listes de 10 résultats successifs « pile » ou « face ». Que vaut $X(\Omega)$? C'est $\{0; 1; 2; \dots; 10\}$: on peut obtenir face 0, 1, 2, ..., 9 ou 10 fois.

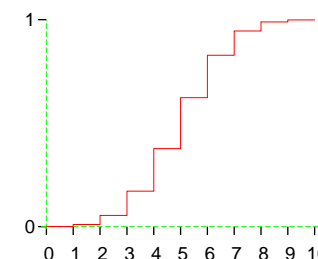
Et si k désigne un nombre entre 0 et 10, $p_X(k) = \frac{\text{nombre de listes avec } k \text{ fois « face »}}{\text{nombre total de listes}}$; mais le nombre total de listes de 10 résultats pile ou face est 2^{10} , alors que le nombre de listes avec exactement k résultats face est $\binom{10}{k}$. Ainsi, $p(X = k) = \frac{\binom{10}{k}}{2^{10}}$.

On obtient les représentations graphiques :

valeurs des $p_k = p(X = k)$



fonction de répartition



On en déduit de même ici la probabilité d'autres événements, par exemple $p(\text{le nombre de résultats pile est impair}) = p(X \in \{1, 3, 5, 7, 9\}) = p(X = 1) + p(X = 3) + p(X = 5) + p(X = 7) + p(X = 9)$, ou bien $p(X > 3) = 1 - p(X \leq 3) = 1 - F_X(4)$.

5.2.2 espérance, variance, écart-type

On définit pour une variable aléatoire discrète X qui prend les valeurs $\{x_1, x_2, \dots, x_n, \dots\}$ avec les probabilités $\{p_1, p_2, \dots, p_n, \dots\}$ (où $p_i = p(X = x_i)$) les trois quantités suivantes :

$$\text{l'espérance de } X \text{ est } E(X) = \sum_{i=1}^n p_i x_i = p_1 x_1 + p_2 x_2 + \dots,$$

la **variance** de X est

$$\text{Var}(X) = \sum_{i=1}^n p_i (x_i - E(X))^2 = p_1 (x_1 - E(X))^2 + p_2 (x_2 - E(X))^2 + \dots,$$

(à chaque fois les sommes ci-dessus peuvent être finies ou infinies)

$$\text{l'écart-type de } X \text{ est } \sigma(X) = \sqrt{\text{Var}(X)}.$$

Le terme espérance vient de la théorie des jeux, pour laquelle ont été développées les probabilités. Il s'agit en quelque sorte du « gain moyen » au cours du jeu : dans une loterie, avec une chance sur 10 de gagner 20€ et neuf chances sur dix de perdre la mise de 3€, l'espérance de gain, ou gain moyen sur un grand nombre de parties, sera bien de $\frac{1}{10}20 + \frac{9}{10}(-3) = -\frac{7}{10}$: l'organisateur du jeu est favorisé !

La variance (et l'écart-type) mesurent, eux, la dispersion des valeurs autour de cette espérance : ils sont d'autant plus élevés que les probabilités des valeurs éloignées de l'espérance (valeur moyenne) sont importantes.

Citons quelques-unes des propriétés remarquables de ces quantités :

L'espérance est linéaire : $E(X + Y) = E(X) + E(Y)$ et $E(aX) = aE(X)$

$$E(X + b) = E(X) + b$$

$$\text{Var}(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2$$

$$\text{Var}(aX) = a^2 \text{Var}(X) \text{ et } \sigma(aX) = |a| \sigma(X)$$

$$\text{Var}(X + b) = \text{Var}(X) \text{ et } \sigma(X + b) = \sigma(X)$$

Une variable aléatoire X est dite **centrée** si $E(X) = 0$, et **réduite** si $\text{Var}(X) = 1$. Des propriétés précédentes on déduit que

si X est une variable aléatoire quelconque,

$$Z = \frac{X - E(X)}{\sigma(X)} \text{ est une variable aléatoire centrée réduite.}$$

5.3 variables aléatoires continues

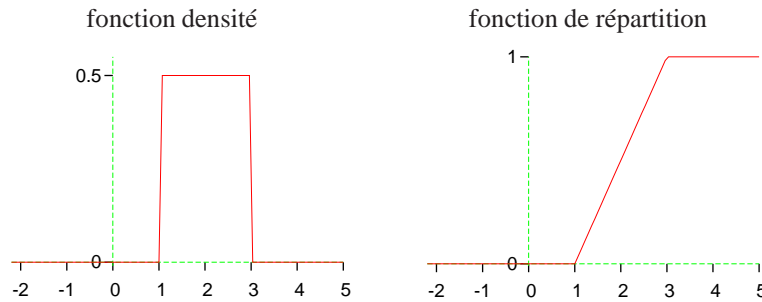
On dit qu'une variable aléatoire X admet une densité f_X si pour tout $x \in \mathbb{R}$ $F_X(x) = p(X < x) = \int_{-\infty}^x f_X(t) dt$: la densité est la dérivée de la fonction de répartition.

On a en particulier $p(a \leq X < b) = F_X(b) - F_X(a)$, et aussi $p(X = a) = \int_a^a f_X(t) dt = 0$: la probabilité d'obtenir une valeur donnée doit être considérée comme nulle, et ce sont les probabilités d'obtenir un résultat dans un segment $[a, b]$ qui nous intéresseront, et permettront de calculer les probabilités de tous les événements.

exemple : si X est une variable aléatoire uniformément répartie sur $[a; b]$, on détermine

$$f_X(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{si } x > b \end{cases} \text{ et } F_X(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases}.$$

que l'on peut représenter graphiquement, pour $a = 1, b = 3$:



remarque : nous ne présentons ici qu'un cas particulier, suffisant pour une première approche : en fait une variable aléatoire continue n'admet pas forcément de densité et on parle, si cette densité existe, de variables aléatoires absolument continues. Toutes les variables continues que nous rencontrerons seront en fait absolument continues.

On définit alors pour une variable aléatoire continue X , par analogie avec le cas discret :

$$\text{l'espérance de } X : E(X) = \int_I x f_X(x) dx$$

$$\text{la variance de } X : \text{Var}(X) = \int_I f_X(x) (x - E(X))^2 dx$$

$$\text{l'écart-type de } X : \sigma(X) = \sqrt{\text{Var}(X)}.$$

Ces formules sont bien entendu analogues aux formules vues dans le cas discret.

Plus précisément, les probabilités élémentaires $p_i = p(X = x_i)$ sont remplacées par la fonction densité $f_X(x)$, que l'on peut interpréter en disant que $f_X(x)dx$ est la probabilité de l'événement infinitésimal $p(x \leq X \leq x + dx)$.

Dans le cas discret, on obtient la probabilité de l'événement A par la formule $p(A) = \sum_{a \in A} p(X = a)$ (la probabilité d'un événement est la somme des probabilités des événements élémentaires qui le composent). Dans le cas continu, la somme est remplacée par une intégrale, et on a la formule $p(A) = \int_A f_X(x)dx$.

De même, l'espérance $E(X) = \sum_i x_i p_i$, qui est la somme des produits de la forme (valeur de la variable aléatoire X) \times (probabilité d'obtenir cette valeur), devient $\int_{x \in \mathbb{R}} x f(x)dx$: on somme ici encore le produit de la valeur x par la probabilité $f(x)dx$ que X prenne une valeur infiniment proche de x .

5.4 l'inégalité de Tchébycheff

Citons ce résultat qui justifiera, plus tard, la loi des grands nombres faisant le lien entre probabilités et statistiques :

Si X est une variable aléatoire quelconque, on a pour tout $\lambda > 0$:

$$p(E(X) - \lambda\sigma(X) < X < E(X) + \lambda\sigma(X)) \geq 1 - 1/\lambda^2,$$

soit encore :

$$p(\frac{|X - E(X)|}{\sigma(X)} \geq \lambda) \leq 1/\lambda^2.$$

Autrement dit, si λ tend vers l'infini, la probabilité que X prenne une valeur dans $[E(X) - \lambda\sigma(X); E(X) + \lambda\sigma(X)]$ devient proche de 1.

Pour $\lambda = 2$ par exemple, la probabilité que X prenne une valeur dans $[E(X) - 2\sigma(X); E(X) + 2\sigma(X)]$ est au moins de $3/4$. Pour $\lambda = 3$, la probabilité que X prenne une valeur dans $[E(X) - 3\sigma(X); E(X) + 3\sigma(X)]$ est au moins de $8/9$.

Ces valeurs sont, la plupart du temps, loin d'être optimales. Mais elles ont l'avantage d'être vérifiées sans aucune hypothèse sur la loi de X . Quand cette loi est connue, nous verrons plus loin comment améliorer ces résultats.

6 Couples de variables aléatoires

On peut vouloir étudier plusieurs variables définies sur une même population, et les liens entre ces variables. Par exemple : l'âge X , la taille Y , le revenu mensuel Z , ... Dans ce but nous allons introduire la notion de couple de variables aléatoires.

On considère donc deux variables X et Y définies sur un même univers Ω . (X, Y) est une nouvelle variable aléatoire, à deux dimensions : c'est une application de $\mathcal{P}(\Omega)$ dans \mathbb{R}^2 .

6.1 loi d'un couple de variable aléatoires discrètes

Si Ω est discret, pour connaître la **loi conjointe**, la loi de (X, Y) , il suffit de la connaître sur les événements élémentaires (x, y) : on doit donc connaître les $p_{xy} = p(X = x \text{ et } Y = y)$. Alors si A est une partie quelconque de \mathbb{R}^2 , on a $p((X, Y) \in A) = \sum_{(x,y) \in A} p((X, Y) = (x, y)) = \sum_{(x,y) \in A} p_{xy}$. En particulier, les p_{xy} sont tous positifs, et leur somme vaut 1.

La loi d'un couple de variables aléatoires (X, Y) étant donnée, on définit la **loi marginale** de X par $p(X = x) = \sum_y p(X = x, Y = y) = \sum_y p_{xy}$; et de même la **loi marginale**

de Y est donnée par $p(Y = y) = \sum_x p(X = x, Y = y) = \sum_x p_{xy}$.

On représente en général ces informations sous forme d'un tableau : les cases centrales donnent les probabilités des événements élémentaires, la somme de chaque ligne ou chaque colonne fournissant les probabilités marginales.

exemple : dans une urne contenant 4 boules indiscernables au toucher marquées 1 à 4, on en tire simultanément deux.

On appelle X le plus petit des deux numéros sortis, et Y le plus grand. Donner la loi conjointe et les lois marginales de (X, Y) .

$X Y$	2	3	4	loi marginale en X
1	1/6	1/6	1/6	1/2
2	0	1/6	1/6	1/3
3	0	0	1/6	1/6
loi marginale en Y	1/6	1/3	1/2	

Indépendance : on dit que deux variables discrètes X et Y sont indépendantes si tous les événements $X = x$ et $Y = y$ sont indépendants, autrement dit si l'on a $p(X = x, Y = y) = p(X = x) \times p(Y = y)$ pour tous x, y .

L'intérêt d'avoir des variables indépendantes est là : la connaissance des lois marginales suffit pour connaître la loi conjointe.

exemple : étudier l'indépendance des variables X et Y de l'exercice précédent.

6.2 loi d'un couple de variables aléatoires continues : densité

Si X et Y sont des variables continues, on appelle **densité conjointe** de X et Y une fonction $f(x, y)$ positive telle que $\int \int f(x, y) \, dx dy = 1$ et $p((a < X < b) \cap (c < Y < d)) = \int_a^b \int_c^d f(x, y) \, dx dy$.

Les **densités marginales** de X et Y sont respectivement $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) \, dy$ et $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) \, dx$.

Indépendance : on dit que deux variables continues X et Y sont indépendantes si tous les événements $a < X < b$ et $c < Y < d$ sont indépendants, autrement dit si l'on a $p(a < X < b, c < Y < d) = p(a < X < b) \times p(c < Y < d)$ pour tous les intervalles $[a, b]$ et $[c, d]$. Cela équivaut à demander que la densité conjointe soit égale au produit des densités marginales : $f(x, y) = f_X(x)f_Y(y)$.

exemple : on peut voir un couple de variables aléatoires (X, Y) comme la détermination d'une répartition de masse totale de 1 sur \mathbb{R}^2 . Dans le cas discret, il s'agit de répartir des masses ponctuelles. Dans le cas continu, des masses continues, la densité de probabilité correspond à une densité « physique ».

Par exemple quelle est la densité associée à la répartition de la masse uniformément sur un triangle de sommets $(0, 0)$, $(0, 1)$, $(1, 1)$?

Le triangle T est déterminé par les équations $y \leq 1$, $x \geq 0$ et $y \geq x$. On cherche une fonction nulle en dehors de T et constante sur T .

Sa valeur k doit donc vérifier $\int_D k = 1$, soit $\int_0^1 \int_x^1 k \, dx dy = 1$, soit encore $\int_0^1 (1-x)k = k/2 = 1$, donc $k = 2$.

La densité marginale en x vaut 0 si $x < 0$ ou si $x > 1$, et sinon $f_X(x) = \int_{y=x}^1 2 \, dy = 2(1-x)$; elle est, logiquement, plus importante quand x est proche de 0 que de 1.

De même la densité marginale en y vaut 0 en dehors de $[0; 1]$ et sur $[0; 1]$: $f_Y(y) = \int_{x=0}^y 2 \, dx = 2y$.

6.3 propriétés

On peut citer les propriétés suivantes des couples de variables aléatoires, qu'elles soient discrètes ou continues :

pour deux variables aléatoires X et Y quelconques :
$$E(X + Y) = E(X) + E(Y)$$

et si les variables X et Y sont indépendantes, alors :
$$E(XY) = E(X)E(Y)$$

et
$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

(attention, il ne suffit pas que ces relations soient vérifiées pour que les variables aléatoires soient indépendantes !!)

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

Si X et Y sont indépendantes, $\text{cov}(X, Y) = \rho(X, Y) = 0$

$$|\text{cov}(X, Y)| \leq \sigma(X) \sigma(Y)$$

$$-1 \leq \rho(X, Y) \leq 1$$

Pour tous a, b réels, $\text{Var}(aX + bY) = a^2\text{Var}(X) + 2abcov(X, Y) + b^2\text{Var}(Y)$

remarque : en toute généralité, la formule donnant $E(XY)$ est

dans le cas discret,
$$E(XY) = \sum_{x,y} p(X = x, Y = y)xy$$

dans le cas continu,
$$E(XY) = \int \int f(x, y)xy \, dx dy,$$

dont le calcul sera souvent laborieux et parfois délicat : si les variables sont indépendantes, la formule $E(XY) = E(X)E(Y)$ est bien plus efficace !

6.4 covariances

On appelle **covariance** de X et Y le nombre $\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$
et **coefficient de corrélation** de X et Y le nombre $\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$.

On a alors les propriétés suivantes :

Ces deux nombres mesurent l'importance de la dépendance entre les variables X et Y ; $\rho(X, Y)$ a l'avantage d'être un nombre sans dimension, toujours compris entre 0 et 1.

remarque : $\rho(X, Y) = 0$ n'implique pas que X et Y sont indépendantes !

Si $\rho(X, Y) = \pm 1$, X et Y sont liées par une relation de dépendance linéaire : Y est de la forme $Y = aX + b$.

exemple : en reprenant l'exemple précédent du tirage de deux boules, X étant la plus petite des deux valeurs et Y la plus grande, on trouve $E(X) = 1/2 \times 1 + 1/3 \times 2 + 1/6 \times 3 = 5/3$ et $E(Y) = 1/6 \times 2 + 1/3 \times 3 + 1/2 \times 4 = 10/3$.

Comme de plus $E(XY) = 1/6 \times 1 \times 2 + 1/6 \times 1 \times 3 + 1/6 \times 1 \times 4 + 1/6 \times 2 \times 3 + 1/6 \times 2 \times 4 + 1/6 \times 3 \times 4 = 35/6$, la covariance $\text{cov}(X, Y) = E(XY) - E(X)E(Y) = 5/18$ est bien non nulle : on retrouve le fait que les variables X et Y ne sont pas indépendantes.

7 Quelques lois de probabilités classiques

7.1 probabilités discrètes

7.1.1 la loi de Bernoulli ou loi 0-1

Une épreuve de Bernoulli est une expérience aléatoire n'ayant que deux issues (« succès - échec »), de probabilités p et $q = 1 - p$. Par exemple, le jeu de pile ou face est une épreuve de Bernoulli avec $p = q = 1/2$.

On définit alors la variable aléatoire X qui vaut 0 en cas d'échec et 1 en cas de succès.

La loi de probabilité de la variable X est appelée **loi de Bernoulli de paramètre p** . On parle aussi de **loi 0-1**.

Son espérance vaut $E(X) = q \times 0 + p \times 1 = p$, et sa variance $\text{Var}(X) = p(1 - p)^2 + q(0 - p)^2 = pq^2 + qp^2 = pq(p + q) = pq$.

Ainsi,

$$\begin{aligned} &\text{si } X \text{ suit une loi de Bernoulli de paramètre } p, \\ & p(X = 0) = q, \quad p(X = 1) = p, \\ & \text{et } E(X) = p, \quad \text{Var}(X) = p(1 - p), \quad \sigma(X) = \sqrt{p(1 - p)}. \end{aligned}$$

7.1.2 la loi binomiale

Un schéma de Bernoulli consiste en la répétition de n épreuves de Bernoulli identiques et indépendantes.

On définit alors la variable aléatoire « nombre total de succès » X , à valeurs dans $\{0, 1, \dots, n\}$.

La loi de probabilité de X est appelée **loi binomiale de paramètres n et p** .

Décrivons plus précisément X : X est caractérisée par le fait que pour tout k compris entre 0 et n , $p(X = k) = \binom{n}{k} p^k q^{n-k}$.

On peut alors calculer l'espérance de la loi binomiale :

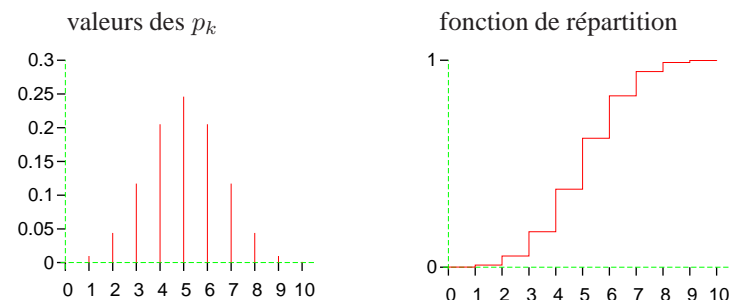
$$\begin{aligned} E(X) &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=1}^n n \binom{n-1}{k-1} p^k q^{n-k} \\ &= \sum_{k=0}^{n-1} n \binom{n-1}{k} p p^k q^{n-1-k} \\ E(X) &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} = np(p + q)^{n-1} \\ &= np. \end{aligned}$$

et de même on peut montrer que la variance vaut $\text{Var}(X) = npq$.

Ainsi,

$$\begin{aligned} &\text{si } X \text{ suit une loi binomiale de paramètres } n, p, \\ & \text{pour tout } k \in \{0; 1; \dots, n\}, \quad p(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \\ & \text{et } E(X) = np, \quad \text{Var}(X) = np(1 - p), \quad \sigma(X) = \sqrt{np(1 - p)}. \end{aligned}$$

On a déjà vu les représentations graphiques des p_k et de la fonction de répartition pour $n = 10, p = 1/2$:



exemple : Un candidat a 5% de chances d'être admis à un concours. Quelle est l'espérance du nombre de candidats admis dans une des salles d'examen de 40 places ?

Le nombre de candidats admis dans la salle suit une loi binomiale de paramètres $n = 40$ et $p = 0,05$. L'espérance est $E = 2$, il y aura donc « en moyenne » deux admis dans chaque salle de 40 candidats.

La variance $V = 2 \times 0,95 = 1,90$ et l'écart-type $\sigma \simeq 1,38$ mesurent la dispersion des valeurs autour de cette valeur moyenne quand on regarde le nombre d'admis sur un grand nombre de salles d'examen. Ce point sera développé en statistiques plus loin.

7.1.3 la loi de Poisson

Il s'agit d'une loi qui permet de modéliser des événements rares, comme un phénomène de désintégration radioactive. On admet que la décomposition d'un échantillon radioactif vérifie les propriétés suivantes :

1. le nombre moyen de désintégrations durant l'intervalle Δt est $\Lambda \Delta t$, avec Λ ne dépendant ni de t , ni de Δt .
2. deux désintégrations ne se produisent jamais simultanément
3. les désintégrations successives sont indépendantes

L'hypothèse 1 est réaliste tant que l'échantillon étudié comporte un grand nombre d'atomes susceptibles de se désintégrer, de sorte que les premières désintégrations n'influent pas sur le nombre d'atomes restant à désintégrer. L'hypothèse 2 est réaliste tant que l'activité radioactive n'est pas trop intense : on peut supposer que toutes les désintégrations sont espacées dans le temps. L'hypothèse 3 est physiquement vérifiée.

On peut alors vérifier que la probabilité α qu'un atome se désintègre durant un intervalle de temps infiniment petit $[t, t + dt]$ est Λdt . En effet, durant cet intervalle de temps, le nombre moyen d'atomes désintégrés est, par l'hypothèse 1, Λdt . Et comme les désintégrations ne sont pas simultanées (hypothèse 2), sur un intervalle de temps infinitésimal, on a soit 1 désintégration (avec probabilité α) soit 0 désintégration (avec probabilité $1 - \alpha$). Ainsi on a bien $\Lambda dt = 1 \times \alpha + 0 \times (1 - \alpha) = \alpha$, égalité que nous allons réutiliser.

On souhaite maintenant déterminer la probabilité $p_k(t)$ qu'exactly k atomes se soient désintégrés pendant un intervalle de temps $[0, t]$.

L'événement « aucun atome ne s'est décomposé pendant l'intervalle $[0, t + dt]$ » est l'intersection des deux événements indépendants « aucun atome ne s'est décomposé durant $[0, t]$ » (de probabilité $p_0(t)$) et « aucun atome ne s'est décomposé durant $[t, t + dt]$ » (de probabilité $1 - \Lambda dt$).

$$\text{Ainsi, } p_0(t + dt) = p_0(t)(1 - \Lambda dt).$$

L'événement « exactement k atomes se sont décomposés durant l'intervalle $[0, t + dt]$ » est la réunion des deux événements incompatibles « exactement k atomes se sont décomposés durant l'intervalle $[0, t]$ et aucun durant $[t, t + dt]$ » (de probabilité $p_k(t)(1 - \Lambda dt)$) et « exactement $k - 1$ atomes se sont décomposés durant l'intervalle $[0, t]$ et un durant $[t, t + dt]$ » (de probabilité $p_{k-1}(t)\Lambda dt$).

Ainsi,

$$p_k(t + dt) = p_k(t)(1 - \Lambda dt) + p_{k-1}(t)\Lambda dt.$$

La première relation donne l'équation différentielle $p'_0(t) = -\Lambda p_0(t)$, donc $p_0(t) = e^{-\Lambda t}$ (car à l'évidence $p_0(0) = 1$).

La deuxième donne les équations $p'_k(t) = -\Lambda p_k(t) + \Lambda p_{k-1}(t)$. $p_k(t)$ est donc de la forme $K_k(t)e^{-\Lambda t}$, avec $K'_k(t) = \Lambda e^{\Lambda t} p_{k-1}(t)$.

Comme $p_0(t) = e^{-\Lambda t}$, $K'_1(t) = \Lambda$, donc $K_1(t) = \Lambda t$, et $p_1(t) = \Lambda t e^{-\Lambda t}$. On montre ensuite par récurrence que $K_k(t) = \frac{\Lambda^k t^k}{k!}$, et donc que $p_k(t) = \frac{\Lambda^k t^k}{k!} e^{-\Lambda t}$.

On peut vérifier que l'on a bien $\sum_{k=0}^{+\infty} p_k(t) = 1$, quelles que soient les valeurs de t et Λ .

Plus généralement : une variable aléatoire suit une **loi de Poisson de paramètre** $\lambda > 0$ si pour $k \geq 0$, $p(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$.

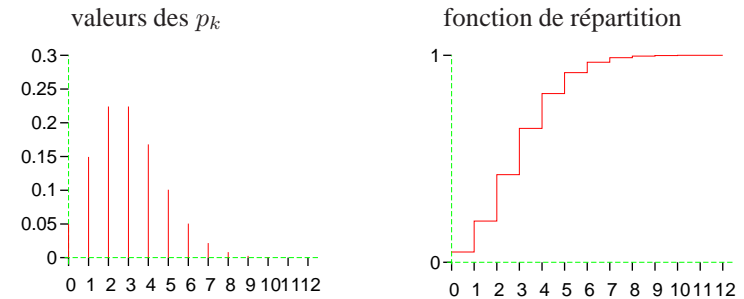
Ainsi, à chaque fois que l'on cherchera à modéliser des événements rares, dont le nombre moyen par unité de temps est λ connu et qui se succèdent de manière indépendante sans que deux événements ne soient jamais simultanés, le nombre X d'événements qui ont lieu durant l'unité de temps fixée suivra une loi de Poisson de paramètre λ .

Alors :

$$\text{si } X \text{ suit une loi de Poisson paramètre } \lambda, \quad \text{pour tout } k \in \mathbb{N}, p(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

$$\text{et on a :} \quad E(X) = \lambda, \quad \text{Var}(X) = \lambda, \quad \sigma(X) = \sqrt{\lambda}.$$

On a déjà vu pour $\lambda = 3$ la représentation des p_k et de la fonction de répartition :



exemples : un standard téléphonique reçoit en moyenne 10 appels par heure.

On peut (et on doit, pour répondre aux questions) supposer que les appels sont indépendants les uns des autres, ne sont jamais simultanés, et que le nombre moyen d'appels sur une durée Δt est proportionnel à Δt : ici, 10 par heure, donc aussi $10 \times 5/60$ par période de 5 minutes, $10/4$ par période de 15 minutes, etc. . .

- quelle est la probabilité pour qu'en une heure il reçoive exactement 2 appels ?
Le nombre X d'appels sur une période d'une heure suit une loi de Poisson de paramètre 10, et donc $p(X = 2) = \frac{e^{-10} 10^2}{2!} \simeq 0, 23\%$.
- quelle est la probabilité pour qu'en 5 minutes il reçoive exactement 3 appels ?
Le nombre X d'appels suit une loi de Poisson de paramètre $\frac{10 \times 5}{60} = \frac{5}{6}$, et donc $p(X = 3) = \frac{e^{-5/6} (\frac{5}{6})^3}{3!} \simeq 4, 2\%$.
- quelle est la probabilité pour qu'en 15 minutes, il reçoive au moins 2 appels ?
Le nombre X d'appels suit une loi de Poisson de paramètre $\frac{10 \times 15}{60} = \frac{5}{2}$, donc $p(X \geq 2) = 1 - p(X < 2) = 1 - p(X = 0) - p(X = 1) = 1 - e^{-5/2} - e^{-5/2} (\frac{5}{2}) \simeq 71, 3\%$.
- quelle est la probabilité pour qu'en une heure, il reçoive au plus 8 appels ?
Le nombre X d'appels suit une loi de Poisson de paramètre 10, donc $p(X \leq 8) = p(X = 0) + p(X = 1) + p(X = 2) + p(X = 3) + p(X = 4) + p(X = 5) + p(X = 6) + p(X = 7) + p(X = 8)$.

On peut bien entendu faire les calculs à l'aide des formules utilisées ci-dessus. Mais il est plus rapide et aussi plus précis d'utiliser les tables de valeur de la loi. Ici, on obtient sans aucun calcul à l'aide de la table située en fin d'ouvrage la valeur $p(X \leq 8) \simeq 33,28\%$.

remarque : si deux variables aléatoires X et Y **indépendantes** suivent des lois de Poisson de paramètres respectifs λ_1 et λ_2 , alors $X + Y$ suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$.

7.2 probabilités continues

7.2.1 la probabilité uniforme

Si une variable aléatoire à valeur dans un segment $[a; b]$ a pour densité la fonction constante définie par $f(x) = \frac{1}{b-a}$ sur $[a, b]$, 0 ailleurs, on dit qu'elle suit une loi de probabilité uniforme.

$$\text{Alors } E(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{a+b}{2}.$$

$$\text{De même } \text{Var}(X) = E(X^2) - E(X)^2 = \frac{1}{b-a} \int_a^b x^2 dx - E(X)^2 = \frac{1}{b-a} \frac{b^3 - a^3}{3} - \frac{a^2 + b^2 + 2ab}{4} = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + b^2 + 2ab}{4} = \frac{b^2 + a^2 - 2ab}{12} = \frac{(b-a)^2}{12}.$$

Ainsi,

$$\text{si } X \text{ suit une loi uniforme sur } [a; b], \quad f_X(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{1}{b-a} & \text{si } x \in [a; b] \\ 0 & \text{si } x > b \end{cases},$$

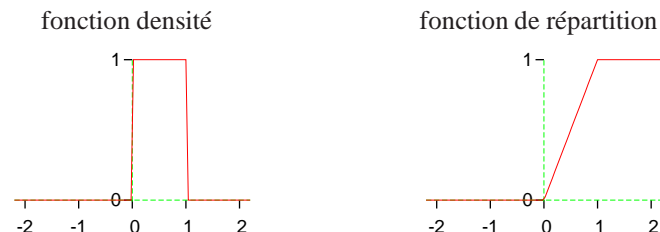
$$F_X(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } x \in [a; b] \\ 1 & \text{si } x > b \end{cases},$$

et on a :

$$E(X) = \frac{a+b}{2},$$

$$\text{Var}(X) = \frac{(b-a)^2}{12}, \quad \sigma(X) = \frac{b-a}{2\sqrt{3}}.$$

D'où la représentation de la **loi uniforme sur $[0, 1]$** :



remarque : il n'existe pas de loi uniforme sur tout \mathbb{R} : si la densité est une constante c , on devrait avoir $\int_{\mathbb{R}} c = 1$: mais l'intégrale n'est pas définie si $c \neq 0$, et vaut 0 si $c = 0$.

7.2.2 la loi exponentielle

Soit X la durée de vie d'un composant « sans vieillissement », dont la durée de vie est indépendante du fonctionnement passé. Plus précisément, on veut que pour tous $T, t \geq 0$:

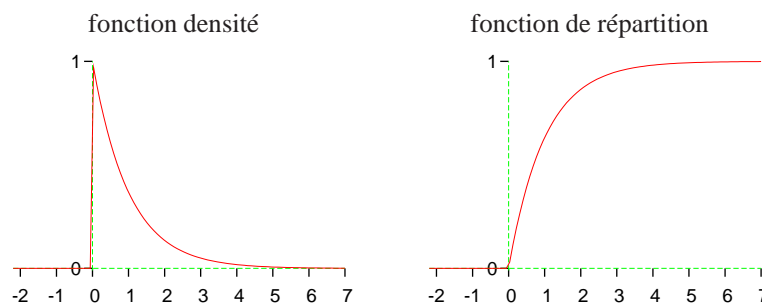
$$p(X \geq t+T \mid X \geq T) = p(X \geq t).$$

$$\text{Alors } \frac{1 - F_X(T+t)}{1 - F_X(T)} = 1 - F_X(t), \text{ donc si } G_X = 1 - F_X, \text{ pour tous } T, t \geq 0,$$

$$G_X(T+t) = G_X(T)G_X(t).$$

Cette équation fonctionnelle implique que G_X est une exponentielle de la forme $G_X(t) = e^{-kt}$, et donc pour $t \geq 0$, $F_X(t) = 1 - e^{-kt}$ (et $F_X(t) = 0$ si $t < 0$). La densité de la variable aléatoire X est alors la dérivée de F_X , soit $f_X(t) = ke^{-kt}$ sur $[0, +\infty[$, 0 sur $] -\infty, 0]$.

On dit que X suit la **loi exponentielle de paramètre k** , et on peut tracer l'allure du graphe de la fonction densité et de la fonction de répartition :



Résumons les valeurs à connaître pour une loi exponentielle :

si X suit la loi exponentielle de paramètre k ,

$$f_X(t) = \begin{cases} 0 & \text{si } t < 0 \\ ke^{-kt} & \text{si } t \geq 0 \end{cases},$$

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - e^{-kt} & \text{si } t \geq 0 \end{cases},$$

et on a :

$$E(X) = 1/k,$$

$$\text{Var}(X) = 1/k^2, \quad \sigma(X) = 1/k.$$

application pratique : la loi exponentielle modélise les pannes d'un composant sans vieillissement : si la durée de vie moyenne d'un tel composant est $1/k$, la variable « durée de vie du composant » suit une loi exponentielle de paramètre k .

On appelle usuellement la durée de vie moyenne « temps moyen entre deux défaillances » (MTBF, soit : « mean time between failures » en anglais, et « moyenne des temps de bon fonctionnement » en français).

Le **taux moyen de défaillance** du composant est alors k .

exemple : le taux moyen de défaillance d'un capteur est de 0.2 par an.

Alors p («la durée du vie du capteur est inférieure à t ») = $1 - e^{-0.2t}$, donc p («le capteur fonctionne à l'instant t ») = $e^{-0.2t}$.

Le MTBF, ou durée de vie moyenne, est l'espérance : $1/0.2 = 5$ ans.

7.2.3 la loi normale

C'est la loi la plus utilisée car en pratique elle « colle bien » à la réalité, pour décrire la répartition de valeurs aléatoires autour d'une valeur moyenne. Par exemple, la répartition des notes à un examen, la répartition des tailles dans une population.

Si une variable aléatoire à valeurs dans \mathbb{R} a pour densité la fonction définie par $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, on dit qu'elle suit la loi normale de paramètres μ et σ , souvent notée $N(\mu, \sigma^2)$. On peut démontrer facilement les propriétés suivantes :

si X suit la loi normale $N(\mu, \sigma^2)$,

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

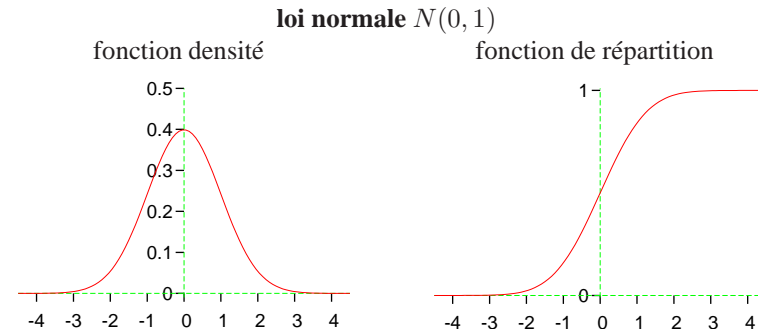
$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt,$$

et on a :

$$E(X) = \mu,$$

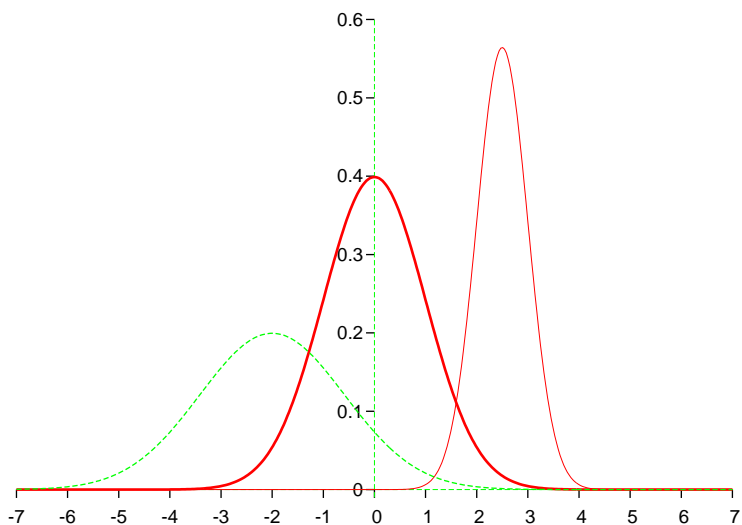
$$\text{Var}(X) = \sigma^2, \quad \sigma(X) = \sigma.$$

Traçons l'allure du graphe de la fonction densité et de la fonction de répartition :



Les valeurs de F_X ne sont pas directement accessibles, car on ne sait pas exprimer à l'aide des fonctions usuelles les primitives de f_X . On doit utiliser des tables ou des ordinateurs qui donnent des valeurs approchées.

Pour voir l'effet des paramètres μ et σ , on peut représenter trois exemples de densités ; de gauche à droite, $(\mu = -2, \sigma = 2)$, $(\mu = 0, \sigma = 1)$ et $(\mu = \frac{5}{2}, \sigma = \frac{1}{2})$:



On voit que l'espérance μ correspond à l'abscisse de l'axe vertical de symétrie de la courbe, alors que la valeur de σ correspond à la largeur de la courbe : plus σ est faible et plus les valeurs sont concentrées autour de la valeur moyenne.

Il serait impossible de donner une table pour toutes les valeurs de σ et μ . Mais on sait que si X suit une loi normale $N(\mu, \sigma^2)$, d'espérance μ et de variance σ^2 , la loi $\frac{X - \mu}{\sigma}$ suit la loi normale centrée réduite $N(0, 1)$ (dont la densité vaut $\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$). Et c'est en utilisant la table de celle-ci que l'on détermine la loi de X . En effet, si $Z = \frac{X - \mu}{\sigma}$, on sait que Z suit une loi normale centrée réduite, et

$$p(a \leq X \leq b) = p(a - \mu \leq X - \mu \leq b - \mu) = p\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = p\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) = p(Z \leq \frac{b - \mu}{\sigma}) - p(Z \leq \frac{a - \mu}{\sigma}), \text{ donc}$$

$$p(a \leq X \leq b) = p(a - \mu \leq X - \mu \leq b - \mu) = p(Z \leq \frac{b - \mu}{\sigma}) - p(Z \leq \frac{a - \mu}{\sigma})$$

et ces deux dernières valeurs, qui sont des valeurs de la fonction de répartition de Z , se trouvent dans la table de la loi normale centrée réduite $N(0; 1)$.

exemple : X est une variable aléatoire de loi normale $N(1, 4)$. Déterminer $p(X \leq 3)$, $p(X > 2)$ et $p(-2.5 \leq X \leq 3)$.

On applique ce qui précède (avec $\mu = 1, \sigma = 2$) : $p(X \leq 3) = p(Z \leq 1)$ soit 0.8413.

De même on a $p(X > 2) = p(Z > 0,5) = 1 - p(Z \leq 0,5) = 1 - 0.6915 = 0.3085$,

Enfin, $p(-2.5 \leq X \leq 3) = p(-1.75 \leq Z \leq 1) = p(Z \leq 1) - (1 - p(Z \leq 1.75)) = 0.8413 - 1 + 0.9599 = 0.8012$.

remarque 1 : interprétation de l'écart-type

L'inégalité de Tchébycheff ($p(E(X) - \lambda\sigma(X) < X < E(X) + \lambda\sigma(X)) \geq 1 - 1/\lambda^2$) peut être améliorée pour une loi normale. En effet,

$$\begin{aligned} p(E(X) - \lambda\sigma(X) < X < E(X) + \lambda\sigma(X)) &= p(\mu - \lambda\sigma < X < \mu + \lambda\sigma) \\ &= p(-\lambda < \frac{X - \mu}{\sigma} < \lambda) \\ &= p\left(\frac{X - \mu}{\sigma} < \lambda\right) - p\left(\frac{X - \mu}{\sigma} < -\lambda\right) \end{aligned}$$

Mais la densité de la loi normale centrée réduite étant paire, on a $p\left(\frac{X - \mu}{\sigma} < -\lambda\right) = 1 - p\left(\frac{X - \mu}{\sigma} < \lambda\right)$, donc finalement, $p(E(X) - \lambda\sigma(X) < X < E(X) + \lambda\sigma(X)) = 2p(Z < \lambda) - 1$, avec $Z \sim N(0, 1)$.

Pour $\lambda = 1, \lambda = 2, \lambda = 3$, les valeurs 0, 3/4 et 8/9 obtenues par l'inégalité de Tchébycheff sont donc largement améliorées :

si X suit la loi normale $N(\mu, \sigma^2)$,

$$\begin{aligned} p(\mu - \sigma < X < \mu + \sigma) &\simeq 0.68 = 68\% \\ p(\mu - 2\sigma < X < \mu + 2\sigma) &\simeq 0.954 = 95.4\% \\ p(\mu - 3\sigma < X < \mu + 3\sigma) &\simeq 0.997 = 99.7\% \end{aligned}$$

remarque 2 : somme de deux lois normales

Si X et Y sont deux variables aléatoires **indépendantes** qui suivent les lois normales $N(\mu_1, \sigma_1^2)$ et $N(\mu_2, \sigma_2^2)$, alors $X + Y$ suit une loi normale $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Une conséquence qui sera fondamentale en statistiques : si X_1, X_2, \dots, X_n sont n variables aléatoires indépendantes de même loi $N(\mu, \sigma^2)$, leur moyenne $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ suit la loi normale $N(\mu, \frac{\sigma^2}{n})$.

7.2.4 autres lois

Nous utiliserons plus tard, sans les étudier préalablement, d'autres lois qui apparaissent naturellement en statistiques : loi du χ^2 , loi de Student.

Troisième partie

Statistiques

8 Approximation d'une loi par une autre

8.1 principe

Les applications statistiques des probabilités nécessitent de connaître les valeurs des probabilités et fonctions de répartition des variables aléatoires dont les lois ont été étudiées précédemment.

Dans le cas des lois discrètes (en particulier, loi binomiale et loi de Poisson) les valeurs des $p(X = k)$ ont déjà été établies, et sont calculables par des formules relativement simples. En revanche, déterminer une valeur de fonction de répartition $F_X(k) = p(X \leq k)$ nécessite de faire la somme des $p(X = j)$ pour tous les j strictement inférieurs à k , ce qui peut être très fastidieux dès que k dépasse quelques unités.

Pour la loi de Poisson $P(\lambda)$, les tables font déjà figurer, pour chaque valeur du paramètre λ , la valeur des $p(X \leq k)$. En revanche, pour la loi binomiale $B(n, p)$, la difficulté tient à ce que la loi est déterminée par 2 paramètres, n et p , et qu'il faudrait donc disposer d'une table pour chaque valeur du couple (n, p) .

On va donc essayer, quand n est grand, de contourner le problème en remplaçant les valeurs de la loi $B(n, p)$ par les valeurs correspondantes pour une loi de Poisson ou une loi normale de mêmes espérance et variance.

8.2 approximation d'une loi binomiale par une loi de Poisson

Quand n est grand et p petit, le succès lors d'un tirage de Bernoulli est un événement rare. De plus, dans ce cas, l'espérance np et la variance $npq = np(1-p)$ sont peu différents.

Cela incite à utiliser, en lieu et place de la loi binomiale $B(n, p)$, une loi de Poisson $P(np)$ de paramètre $\lambda = np$. Et en pratique, on peut vérifier que l'approximation ainsi obtenue n'est pas trop mauvaise si $n \geq 30$, $p \leq 0,1$ et $np \leq 5$.

exemple : on reprend l'exemple des 40 candidats d'une salle de concours qui ont chacun 5% de chances d'être admis. $n = 40 > 30$, $p = 0,05 < 0,1$, et on a bien $np = 2 \leq 5$.

Avec la loi exacte qui est une loi binomiale, $p(X = k) = \binom{40}{k} (0,05)^k (0,95)^{40-k}$. Avec la loi de Poisson de paramètre 2 on trouve $p(X = k) = e^{-2} 2^k / k!$.

Calculons quelques valeurs approchées à 4 chiffres significatifs :

k	Loi binomiale $n = 40, p = 0,05$ valeur de $p(X = k)$	Loi de Poisson $\lambda = np = 2$ valeur de $p(X = k)$
0	0.1353	0.1285
1	0.2707	0.2706
2	0.2707	0.2777
3	0.1804	0.1851
4	0.0902	0.0901
5	0.0361	0.0342
6	0.0120	0.0105
7	0.0034	0.0027
8	0.0009	0.0006
9	0.0002	0.0001
10+	0.0000	0.0000

On constate effectivement que l'approximation n'est pas trop mauvaise et permet de déterminer, par simple lecture de la table de la loi de Poisson, des valeurs approchées du type $p(X \leq 5) \simeq 0.9834$, la valeur exacte, que l'on devrait calculer explicitement par $p(X = 0) + p(X = 1) + \dots + p(X = 5) = \sum_{k=0}^5 \binom{40}{k} 0.05^k 0.95^{40-k}$ étant proche de 0.986.

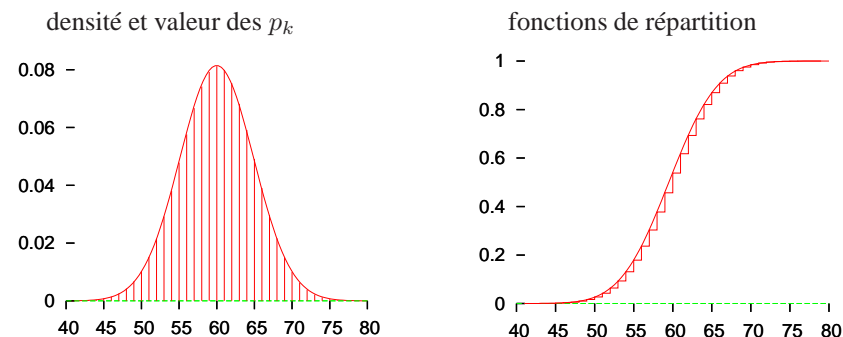
8.3 approximation d'une loi binomiale par une loi normale

Si n est grand mais si p n'est pas petit, l'approximation par une loi de Poisson n'est pas utilisable, et on préférera utiliser la loi normale $N(np, np(1-p))$.

Mais les $p(X = k)$ sont nuls pour une loi normale (continue) : on va approcher $p(X = k)$ par $p(k - 0.5 < X < k + 0.5)$. En effet, sachant que les valeurs de la variable X sont entières, si on utilise une loi normale on peut assimiler chaque valeur à l'entier le plus proche.

exemple : on souhaite comparer $B(100, 0.6)$ avec $N(60, 24)$.

Pour cela on trace sur le même graphique, d'une part les $p(X = k)$ et le graphe de la densité, d'autre part les $p(X \leq k)$ et le graphe de la fonction de répartition $p(X < k + 0.5)$.



Les graphiques sont confirmés par le tableau suivant listant quelques valeurs approchées à 4 chiffres significatifs :

	$B(n = 100, p = 0.6)$	$N(\mu = np = 60, \sigma^2 = 24)$	$N(\mu = np = 60, \sigma^2 = 24)$
k	$p(X = k)$	$F(k + \frac{1}{2}) - F(k - \frac{1}{2})$	$F(k + 1) - F(k)$
40-	0.0	0.0	0.0
41	0.0001	0.0	0.0001
42	0.0001	0.0001	0.0001
43	0.0002	0.0002	0.0003
44	0.0004	0.0004	0.0006
45	0.0008	0.0008	0.001
46	0.0015	0.0014	0.0018
47	0.0026	0.0024	0.0032
48	0.0042	0.0041	0.0052
49	0.0068	0.0066	0.0082
50	0.0103	0.0102	0.0125
51	0.0152	0.0151	0.0181
52	0.0215	0.0215	0.0253
53	0.0292	0.0294	0.0338
54	0.0381	0.0385	0.0434
55	0.0478	0.0484	0.0534
56	0.0576	0.0583	0.063
57	0.0667	0.0674	0.0714
58	0.0742	0.0748	0.0776
59	0.0792	0.0796	0.0809
60	0.0812	0.0813	0.0809
61	0.0799	0.0796	0.0776
62	0.0754	0.0748	0.0714
63	0.0682	0.0674	0.063
64	0.0591	0.0583	0.0534
65	0.0491	0.0484	0.0434
66	0.0391	0.0385	0.0338
67	0.0297	0.0294	0.0253
68	0.0217	0.0215	0.0181
69	0.0151	0.0151	0.0125
70	0.01	0.0102	0.0082
71	0.0063	0.0066	0.0052
72	0.0038	0.0041	0.0032
73	0.0022	0.0024	0.0018
74	0.0012	0.0014	0.001
75	0.0006	0.0008	0.0006
76	0.0003	0.0004	0.0003
77	0.0001	0.0002	0.0001
78	0.0001	0.0001	0.0001
79+	0.0	0.0	0.0

Alors si F désigne la fonction de répartition de la loi normale : la valeur $p(X = k)$ est très bien approchée par $F(k + \frac{1}{2}) - F(k - \frac{1}{2})$, la valeur $p(X \leq k)$ est très bien approchée par $F(k + \frac{1}{2})$ (de plus on vérifie sans surprise que l'approximation de $p(X = k)$ par $F(k + \frac{1}{2}) - F(k - \frac{1}{2})$ est meilleure que celle par $F(k + 1) - F(k)$).

remarque : domaine de validité

On retiendra que pour $n \geq 50$ et $np(1-p) \geq 18$, on peut approcher la loi binomiale de paramètres n et p par la loi normale $N(np, np(1-p))$.

8.4 approximation d'une loi de Poisson par une loi normale

Précisons enfin que, si le paramètre λ est supérieur à 10, on peut approcher la loi de Poisson $P(\lambda)$ par une loi normale de mêmes espérance et variance $N(\lambda, \lambda)$, en remplaçant comme précédemment $p(X = k)$ par $p(k - 0.5 < X < k + 0.5)$.

Cela permet de pallier l'absence des valeurs de la loi pour λ supérieur à 10 dans la table située plus loin.

remarque : nous venons déjà d'utiliser deux fois la loi normale pour approcher une autre loi. Il s'agit en fait de cas particuliers d'un résultat bien plus général, qui veut que la loi normale apparaisse souvent comme limite d'autres lois.

Voir en particulier dans la suite de ce cours :

- le théorème central limite : quelles que soit les lois de variables X_1, \dots, X_n identiques et indépendantes, leur moyenne \bar{X} suit asymptotiquement une loi normale si n tend vers l'infini
- quand le nombre de degré de liberté tend vers l'infini, la loi de Student tend vers la loi normale centrée réduite.

9 Echantillonnage

9.1 la loi des grands nombres

Le jeu de « pile ou face » avec une pièce bien équilibrée a déjà été modélisé dans ce qui précède, en disant que la probabilité d'obtenir le résultat « pile » est de $1/2$, et celle d'obtenir « face » aussi.

Mais que signifie en pratique ce terme « probabilité » ? En jetant une fois la pièce, on obtient soit « pile », soit « face », et non un improbable « $1/2$ pile, $1/2$ face ». En jouant deux fois de suite, on n'obtient pas forcément toujours exactement un résultat « pile » et un résultat « face »...c'est en termes statistiques que l'on peut interpréter ce $1/2$: la limite de la fréquence d'obtention d'un résultat tend vers la probabilité de ce résultat quand on répète à l'infini l'expérience.

En effet, supposons que l'on jette n fois la pièce, et que l'on note S_n le nombre de « pile » ; alors S_n/n est la fréquence d'apparition de « pile ». On sait que S_n suit une loi binomiale de paramètres n et $1/2$, avec $E(S_n) = n/2$, $\text{Var}(S_n) = n/4$ et $\sigma(S_n) = \sqrt{n}/2$; alors d'après l'inégalité de Tchébycheff :

$$p \left(\left| \frac{S_n}{n} - \frac{1}{2} \right| \leq \frac{\lambda}{2\sqrt{n}} \right) \geq 1 - \frac{1}{\lambda^2}, \text{ pour tout } \lambda > 0.$$

Autrement dit, avec une probabilité aussi proche de 1 que l'on veut, on peut rendre S_n/n aussi proche que l'on veut de $1/2$ si n est assez grand.

Ce résultat se généralise par la

loi faible des grands nombres

dans une expérience aléatoire, on considère un événement de probabilité p .

Alors,

si on répète l'expérience n fois en notant f_n la fréquence d'apparition du résultat,

$$f_n \rightarrow p \text{ quand } n \rightarrow +\infty.$$

Ainsi il suffit de répéter un nombre suffisant de fois une expérience pour que la fréquence observée d'un résultat se rapproche de la probabilité théorique.

Nous allons par la suite préciser ce résultat en lui donnant une portée pratique.

remarque : obtenir 100 fois de suite « face » ne signifie pas que la probabilité d'obtenir « pile » au tirage suivant est plus forte ! **Le hasard n'a pas de mémoire.**

9.2 le problème de l'échantillonnage

Connaissant les propriétés d'une population, on souhaite déterminer celles d'un échantillon pris au hasard de la population.

On décrit un caractère de la population par une variable aléatoire de loi **connue** X , dont l'espérance est μ et la variance σ^2 .

Puis on modélise le choix d'un échantillon d'effectif n par des variables aléatoires indépendantes X_1, X_2, \dots, X_n , qui suivent chacune la loi de X . Et on s'intéresse à leur moyenne $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$, dont l'espérance est μ et la variance σ^2/n . En particulier on souhaite déterminer avec quelle probabilité la moyenne calculée sur un échantillon sera proche ou non de la « moyenne théorique » μ .

remarque : pour effectuer cette modélisation par des variables aléatoires indépendantes, il faut supposer que l'effectif total est largement supérieur à celui des échantillons prélevés, de sorte que le choix d'un individu dans la population ne modifie pas les propriétés de celle-ci. On assimile donc le choix d'un échantillon à des tirages avec remise.

En pratique, on procède ainsi dès que l'on prélève moins de 5 à 10% de l'effectif total.

9.3 le cas des lois normales

Puisque les X_i sont indépendantes, si chaque $X_i \sim N(\mu, \sigma^2)$, alors on a déjà vu que $X_1 + X_2 + \dots + X_n \sim N(n\mu, n\sigma^2)$, et donc $\bar{X} \sim N(n\mu/n, n\sigma^2/n^2)$. Ainsi,

si X_1, X_2, \dots, X_n sont des variables aléatoires indépendantes de même loi $N(\mu, \sigma^2)$, alors $\bar{X} \sim N(\mu, \sigma^2/n)$.

On peut donc facilement évaluer la valeur moyenne des résultats de l'expérience.

exemple : une machine fabrique des disques dont les diamètres suivent une loi normale d'espérance 15 cm et d'écart-type 0.25 cm. On s'intéresse à un échantillon de 10 disques choisis au hasard dans la production : ici X_1, \dots, X_{10} sont les diamètres des disques, et la valeur moyenne des diamètres de l'échantillon est \bar{X} .

On s'attend, d'après la loi des grands nombres, à obtenir un diamètre moyen $E(\bar{X})$ proche de 15 cm.

Mais on peut préciser cela : comme $\bar{X} \sim N(15, 0.00625)$, on peut par exemple affirmer que dans 95.4% des cas la moyenne de l'échantillon sera dans l'intervalle $[15 - 2 \times 0.25/\sqrt{10}, 15 + 2 \times 0.25/\sqrt{10}] = [14.84, 15.16]$.

9.4 le cas général

Dans l'exemple qui précède, on a utilisé le fait que $X \sim N(\mu, \sigma^2)$, $\bar{X} \sim N(\mu, \sigma^2/n)$. Mais ce résultat reste en fait vrai quel que soit la loi de X pourvu que n soit assez grand :

théorème central limite

si n est assez grand, si X_1, \dots, X_n sont des variables aléatoires indépendantes qui suivent une même loi d'espérance μ et de variance σ^2 , leur moyenne \bar{X} suit approximativement une loi normale de paramètre $N(\mu, \sigma^2/n)$.

Si n est assez grand (et en pratique, si n dépasse 30) il n'est donc plus nécessaire de supposer que la loi de X est une loi normale pour considérer que celle de \bar{X} l'est.

exemple : une entreprise fabrique des piles ; l'espérance de la tension est de $1.6V$ et l'écart-type de $0.1V$. Sur un échantillon de 50 piles, quelle sera la tension moyenne ?

La loi de répartition des tensions n'est pas supposée normale, mais la taille de l'échantillon est suffisante pour considérer que la moyenne suit, elle, une loi normale d'espérance 1.6 et de variance $\sigma^2 = 0.1^2/50$, donc d'écart-type $\sigma = 0.0141$. Ainsi on a, par exemple, 95.4% de chances pour que la tension moyenne soit dans l'intervalle $[1.572, 1.628]$, ...

9.5 application aux pourcentages

On peut appliquer ce qui précède aux études de pourcentages : on suppose qu'une proportion p des éléments d'une population possèdent une certaine propriété, et on s'intéresse à la proportion correspondante dans un échantillon fixé de taille n de la population.

Si on note X le nombre d'éléments de l'échantillon choisi qui possèdent la propriété considérée, on sait que la variable aléatoire X suit une loi binomiale de paramètres n et p , d'espérance np et de variance $np(1-p)$. De plus si n est assez grand, on sait que l'on peut approcher la loi de X par la loi normale de paramètres np et $np(1-p)$. Et la variable aléatoire $F = X/n$ qui correspond à la proportion d'individus de l'échantillon possédant la propriété suit donc approximativement une loi normale de paramètres p et $p(1-p)/n$.

Si la proportion p d'une population possède une propriété, la fréquence d'apparition de cette propriété dans un échantillon aléatoire de taille n suffisamment grande suit la loi normale $N\left(p, \frac{p(1-p)}{n}\right)$.

exemple : la proportion de gauchers dans la population est de 10%.

Par conséquent la probabilité d'observer plus de 15% de gauchers dans un groupe de 30 personnes est $p(F > 0.15) = 1 - p(F \leq 0.15)$ pour $F \sim N(0.1, 0.09/30)$, soit environ 19%.

10 Estimation

10.1 le problème

Il s'agit du problème inverse de l'échantillonnage, et qui en pratique se rencontre bien plus souvent.

On dispose de n mesures x_i d'une quantité X , et on souhaite en déduire une estimation de $E(X) = \mu$ et de $\text{Var}(X) = \sigma^2$.

exemple 1 : on mesure le diamètre d'un échantillon de 300 tiges métalliques. On obtient les résultats suivants :

diamètre d en mm	42	43	44	45	46	47	48
effectif	13	26	63	149	30	15	4

exemple 2 : on répète 5 mesures d'une tension u . On obtient les résultats suivants, en volts : 11.83, 11.91, 12.01, 12.03, 12.1.

10.2 estimations ponctuelles

Disposant d'un échantillon de taille n , il semble naturel de commencer par déterminer sa moyenne et sa variance :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \text{et} \quad \sigma'^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}.$$

La moyenne d'échantillon \bar{x} est alors prise comme **estimation ponctuelle** de l'espérance de X :

$$E(X) = \mu \simeq \bar{x}$$

On pourrait croire que la variance d'échantillon σ'^2 est une bonne estimation de $\text{Var}(X) = \sigma^2$. Mais procéder ainsi conduit à sous-estimer σ^2 pour les petites valeurs de n , et on utilisera plutôt $s^2 = \frac{n}{n-1} \sigma'^2$ pour **estimation ponctuelle** de σ^2 :

$$\text{Var}(X) = \sigma^2 \simeq s^2 = \frac{n}{n-1} \sigma'^2$$

On dit que la variance d'échantillon σ'^2 est un **estimateur biaisé** de la variance σ^2 , alors que s^2 est appelé **estimateur sans biais**. Ce terme s'explique par le fait que la valeur moyenne de σ'^2 (valeur moyenne déterminée sur un grand nombre d'échantillons de même taille n) ne tend pas vers σ^2 , contrairement à la valeur moyenne de s^2 . C'est seulement dans le cas où n est grand, autrement dit pour des effectifs importants, que l'on peut assimiler les deux variances.

Pour détailler un peu ce point, fixons un entier n et considérons des variables aléatoires indépendantes X_1, \dots, X_n de même loi que X , et leur moyenne $\bar{X} = \frac{X_1 + \dots + X_n}{n}$. Nous souhaitons déterminer l'espérance de la variable aléatoire $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, qui correspond à σ'^2 .

Pour i fixé, calculons l'espérance de la quantité $(X_i - \bar{X})^2 = (X_i - \mu + \mu - \bar{X})^2 = (X_i - \mu)^2 + (\mu - \bar{X})^2 + 2(X_i - \mu)(\mu - \bar{X})$.

Comme X_i et \bar{X} sont d'espérance μ , on a $E((X_i - \mu)^2) = \text{Var}(X_i - \mu) = \sigma^2$ et $E((\mu - \bar{X})^2) = \text{Var}(\bar{X}) = \sigma^2/n$.

De plus, $E((X_i - \mu)(\mu - \bar{X})) = E((X_i - \mu)(\frac{1}{n} \sum_j (\mu - X_j))) = \frac{1}{n} \sum_j E((X_i - \mu)(\mu - X_j)) = -\frac{1}{n} \sum_j \text{cov}(X_i - \mu, X_j - \mu)$. Si $i \neq j$, $X_i - \mu$ et $X_j - \mu$ sont indépendantes donc $E((X_i - \mu)(\mu - \bar{X})) = -\frac{1}{n} \text{Var}(X_i - \mu) = -\frac{\sigma^2}{n}$. Ainsi : $E((X_i - \bar{X})^2) = \sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2$.

Donc on a aussi $E(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2) = \frac{n-1}{n} \sigma^2 < \sigma^2$. Cela explique que l'on préfère prendre $\frac{n}{n-1} \sigma'^2$ comme estimation ponctuelle de la variance : si l'on répète sur un grand nombre d'échantillons l'expérience, la moyenne des variances des échantillons se rapprochera bien de σ^2 et non de $\frac{n-1}{n} \sigma^2$.

Dans les exemples, on trouve respectivement :

exemple 1 : pour le diamètre des tiges, une moyenne d'échantillon de $\bar{d} = 44.73$, une variance d'échantillon de $\sigma'^2 = 1.292$ et donc $s^2 = 1.296$. Donc $\sigma' = 1.136$ et $s = 1.138$.

L'estimation ponctuelle de l'espérance de d est donc 44.73 et de la variance 1.296

exemple 2 : pour la tension, une moyenne d'échantillon de $\bar{u} = 11.97$, une variance d'échantillon de $\sigma'^2 = 0.009$ et donc $s^2 = 0.01128$. Donc $\sigma' = 0.094$ et $s = 0.106$.

L'estimation ponctuelle de l'espérance de u est donc 11.97 et de la variance 0.011.

10.3 estimation par intervalles de confiance

Le défaut des estimations ponctuelles vues plus haut est que si un autre échantillon est choisi, les valeurs estimées pour l'espérance et la variance seront différentes. On préfère donc estimer moyenne et écart-type non pas par une simple valeur ponctuelle, mais par des intervalles qui contiennent μ et σ avec une certaine probabilité connue à l'avance.

On obtiendra donc un intervalle centré sur l'estimation ponctuelle, et de largeur d'autant plus grande que l'on souhaite avoir une forte probabilité que la valeur exacte soit à

l'intérieur.

10.3.1 estimation de la moyenne : un premier cas idéal

Traisons tout d'abord dans le cas, peu probable en pratique mais plus simple à étudier, où la variance σ^2 est connue, et où l'on cherche seulement à estimer l'espérance μ . On connaît aussi bien sûr la moyenne d'échantillon \bar{x} , estimation ponctuelle de μ .

On va refaire, à l'envers, un raisonnement déjà vu pour l'échantillonnage : X_1, X_2, \dots, X_n sont les variables indépendantes correspondant aux mesures de notre échantillon de taille n .

Alors $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ est assimilé à une loi normale $N(\mu, \sigma^2/n)$: c'est exact si X suit une loi normale, et c'est une approximation valable pour $n \geq 30$ si X suit une loi quelconque.

Mais alors $p(\mu - t \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + t \frac{\sigma}{\sqrt{n}}) = p(-t \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq t)$, et comme $\sqrt{n} \frac{\bar{X} - \mu}{\sigma}$ suit une loi normale centrée réduite, $p(-t \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq t) = 2F(t) - 1$, F étant la fonction de répartition de la loi normale centrée réduite.

Ecrivons « à l'envers » cette inégalité : $p(-t \leq \sqrt{n} \frac{\mu - \bar{X}}{\sigma} \leq t) = 2F(t) - 1 = p(\bar{X} - t \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + t \frac{\sigma}{\sqrt{n}})$.

Donc μ se trouve, avec une probabilité $2F(t) - 1$, dans l'intervalle $[\bar{x} - t \frac{\sigma}{\sqrt{n}}, \bar{x} + t \frac{\sigma}{\sqrt{n}}]$, appelé **intervalle de confiance** de la moyenne avec la probabilité $2F(t) - 1$ (ou avec le risque $2 - 2F(t)$).

Ainsi pour une population dont une caractéristique est donnée par une variable aléatoire X de variance connue σ^2 , si l'on dispose d'un échantillon d'effectif n et de moyenne \bar{x} , alors : l'espérance $E(X) = \mu$ se trouve dans l'intervalle $[\bar{x} - t \frac{\sigma}{\sqrt{n}}, \bar{x} + t \frac{\sigma}{\sqrt{n}}]$ avec une probabilité de $2F(t) - 1$. (F étant la fonction de répartition de la loi normale centrée réduite).

On ne peut pas, en toute rigueur, appliquer ce qui précède à nos exemples. En effet la variance de la variable aléatoire X n'est, comme souvent en pratique, pas connue. Nous verrons dans ce qui suit que pour l'exemple 1, n étant grand, on peut appliquer ce qui précède en remplaçant la variance σ^2 par son estimation ponctuelle s^2 . Ainsi, on trouve par exemple :

l'intervalle de confiance 95% (ou : au risque 5%) pour la moyenne est [44.6, 44.86],
l'intervalle de confiance 99.7% (ou : au risque 0.3%) pour la moyenne est [44.53, 44.92],

les centres de tous ces intervalles étant, à chaque fois, la moyenne d'échantillon $\bar{x} = 44.73$.

Ainsi, l'énoncé à retenir est le suivant :

pour une population dont une caractéristique est donnée par une variable aléatoire X , si l'on dispose d'un échantillon d'effectif n de moyenne \bar{x} et donnant une estimation ponctuelle de l'écart-type s , alors :

l'espérance $E(X) = \mu$ se trouve dans l'intervalle $[\bar{x} - t \frac{s}{\sqrt{n}}, \bar{x} + t \frac{s}{\sqrt{n}}]$ avec une probabilité de $2F(t) - 1$.
(F étant la fonction de répartition de la loi normale centrée réduite).

En particulier :

l'espérance $E(X) = \mu$ se trouve dans l'intervalle $[\bar{x} - \frac{s}{\sqrt{n}}, \bar{x} + \frac{s}{\sqrt{n}}]$ avec une probabilité de 68%,

l'espérance $E(X) = \mu$ se trouve dans l'intervalle $[\bar{x} - 1.96 \frac{s}{\sqrt{n}}, \bar{x} + 1.96 \frac{s}{\sqrt{n}}]$ avec une probabilité de 95%,

l'espérance $E(X) = \mu$ se trouve dans l'intervalle $[\bar{x} - 2 \frac{s}{\sqrt{n}}, \bar{x} + 2 \frac{s}{\sqrt{n}}]$ avec une probabilité de 95.4%,

l'espérance $E(X) = \mu$ se trouve dans l'intervalle $[\bar{x} - 3 \frac{s}{\sqrt{n}}, \bar{x} + 3 \frac{s}{\sqrt{n}}]$ avec une probabilité de 99.7%.

10.3.2 estimation de la moyenne par la loi de Student

Pour estimer la moyenne dans le cas où σ est connu, on a utilisé le fait que $\sqrt{n} \frac{\bar{X} - \mu}{\sigma}$ suit une loi normale $N(0, 1)$.

Si σ n'est pas connu, il est naturel de le remplacer par son estimation ponctuelle s ; mais la variable $\sqrt{n} \frac{\bar{X} - \mu}{s}$ suit en fait une **loi de Student à $n - 1$ degrés de liberté** ; c'est seulement pour n grand (en pratique, $n \geq 30$) que cette loi coïncide avec une loi normale centrée réduite.

Résumons donc comment déterminer l'intervalle de confiance $1 - \alpha$ (ou : au risque α) :

pour une population dont une caractéristique est donnée par une variable X , si l'on dispose d'un échantillon d'effectif n , de moyenne \bar{x} , et d'estimation ponctuelle de l'écart-type s , alors

l'espérance $E(X) = \mu$ se trouve dans l'intervalle $[\bar{x} - t(\alpha)\frac{s}{\sqrt{n}}, \bar{x} + t(\alpha)\frac{s}{\sqrt{n}}]$ avec une probabilité de $1 - \alpha$,

$t(\alpha)$ étant défini par la relation $p(|\sqrt{n} \frac{\bar{X} - \mu}{s}| \leq t(\alpha)) = 1 - \alpha$ et lu directement dans la table de la loi de Student, à l'intersection de la colonne du risque α et de la ligne du nombre $n - 1$ de degrés de liberté.

Dans le cas de l'exemple 1, comme $n = 300$ est largement supérieur à 30, la loi de Student coïncide avec la loi normale utilisée précédemment, et on retrouve donc les mêmes valeurs, par exemple si $\alpha = 0.05$, $t(\alpha) = 1.9679$, et l'intervalle de confiance à 95% est $[44.6, 44.86]$.

Mais dans l'exemple 2, que nous ne pouvons traiter directement, on trouve maintenant, si $\alpha = 0.05$ les valeurs : $n = 5$, $\bar{u} = 11.07$, $s = 0.106$, $t(0.05) = 2.7764$ d'après la table de la loi de Student à 4 degrés de liberté, donc finalement l'intervalle de confiance au risque 5% est $[11.83, 12.10]$ (et non $[11.88, 12.06]$ comme on trouverait en utilisant une loi normale).

De même l'intervalle de confiance au risque 1% est $[11.75, 12.19]$ (et non $[11.83, 12.11]$).

10.3.3 estimation d'une proportion

On considère une situation binaire : dans une population dont les membres possèdent ou non une propriété (par exemple : être gaucher ou droitier ; avoir voté pour le candidat A à un scrutin ; être défectueux ; ...), on calcule la fréquence f d'apparition de cette propriété sur un échantillon de taille n , et on souhaite estimer la proportion p d'individus, dans la population totale, qui possèdent la propriété.

L'estimation ponctuelle de p naturellement f . Mais comment calculer un intervalle de confiance ?

On admet le résultat suivant :

Soit F la variable aléatoire qui correspond à la fréquence d'une caractéristique dans un échantillon de taille n , supposée grande ($n \geq 30$ en pratique).

Alors F suit une loi normale d'espérance p et de variance $\frac{p(1-p)}{n}$.

En pratique, on remplacera la variance (qui dépend de p , que l'on cherche à évaluer) par la valeur $\frac{f(1-f)}{n}$ calculée pour l'échantillon.

(remarque : Si n est inférieur à 30, il faudrait utiliser une loi de Student.)

exemple : dans un lot de 50 pièces, on a trouvé 12 pièces défectueuses. On souhaite estimer, par un intervalle de confiance 95%, la proportion de pièces défectueuses sur l'ensemble de la production.

Pour cela, on calcule $n = 50$ et $f = 6/25 = 0.24$. On assimile alors la loi de F à une loi normale d'espérance p et d'écart-type $\sqrt{\frac{0.24(1-0.24)}{50}} \simeq 0.06$, donc celle de la variable $\frac{p-F}{0.06}$ à une loi normale centrée réduite.

Par conséquent, l'intervalle cherché est $[f - 1.96 \times 0.06, f + 1.96 \times 0.06]$ soit $[0.122, 0.358]$.

Supposons que la même proportion de pièces défectueuses est estimée, ponctuellement, sur un intervalle plus grand. Par exemple 240 pièces défectueuses sur 1000 pièces, alors l'intervalle de confiance 95% devient $[0.24 \pm 1.96\sqrt{\frac{0.24(1-0.24)}{1000}}]$ soit $[0.226, 0.254]$: on constate, en toute logique, qu'estimer une proportion sur un intervalle d'effectif plus important permet d'obtenir un intervalle de confiance plus petit.

11 Tests statistiques

On souhaite vérifier des mesures sur un échantillon sont compatibles avec une hypothèse faite sur une population (par exemple : « la durée de vie en fonctionnement d'un composant est de 1000h », « les pièces produites par les usines A et B ont la même longueur », « la résistance du composant suit une loi normale d'espérance 10Ω et d'écart-type 0.5Ω », ...).

Alors on effectue un **test d'hypothèse**, selon le principe ci-dessous :

- on note H_0 l'hypothèse dont on souhaite tester la compatibilité avec les mesures. H_0 est l'**hypothèse nulle**, et l'hypothèse alternative est notée H_1 .
- on fixe $\alpha = p(\text{"rejeter } H_0 \text{"} / \text{"} H_0 \text{ est vraie"})$, appelé **risque de première espèce** (souvent $\alpha = 1\%$ ou $\alpha = 5\%$).
- supposant H_0 vraie, on calcule l'intervalle de confiance de risque α pour la quantité étudiée.
- si la valeur obtenue pour l'échantillon est effectivement dans cet intervalle, on décide d'accepter H_0 . Sinon, on accepte H_1 .

remarque 1 : $\beta = p(\text{"accepter } H_0 \text{"} / \text{"} H_0 \text{ est fautive"})$, appelé **risque de seconde espèce**, n'est pas du tout maîtrisé ici.

remarque 2 : le principe de ces tests n'est pas de choisir entre deux hypothèses symétriques : il faut voir l'hypothèse nulle H_0 comme l'hypothèse privilégiée, l'hypothèse « par défaut », qui ne sera invalidée que si les données statistiques dont on dispose la rendent très peu probable.

A l'inverse, le fait d'accepter l'hypothèse nulle ne prouve pas que celle-ci est vraie, mais seulement que l'on n'a pas assez de données poussant à la rejeter.

11.1 Test bilatéral de la moyenne

exemple : on reprend l'exemple 1 du chapitre précédent : sur un échantillon de 300 disques le diamètre moyen est de 44.73 mm, et la variance observée 1.292 (donc l'estimation ponctuelle de la variance est 1.296).

Si le fabricant affirme qu'il vend des disques de 45mm de diamètre, peut-on, au risque de 5%, dire qu'il a raison ?

On prend pour hypothèse $H_0 : \mu = 45$, et $H_1 : \mu \neq 45$. On suppose H_0 . Alors on sait que, n étant grand, on peut considérer, pour $\sigma = s = 1.296$, que $\sqrt{n} \frac{\bar{X} - \mu}{\sigma}$ suit une loi normale centrée réduite. Par conséquent, on cherche t tel que $p(|\sqrt{n} \frac{\bar{X} - \mu}{\sigma}| \leq t) = 2F(t) - 1 = 0.95$. On a ainsi $F(t) = 0.975$, et donc, par lecture de la table, $t = 1.96$. Ainsi la moyenne d'un échantillon d'effectif 300 à 95% de chances de se trouver dans l'intervalle $[45 - 1.96s/\sqrt{300}; 45 + 1.96s/\sqrt{300}]$, soit ici $[44.87; 45.13]$.

Par conséquent, on rejette H_0 : le diamètre moyen des disques n'est pas 45mm.

11.2 Test unilatéral de la moyenne

exemple : un fabricant utilise comme argument de vente le fait que la durée de vie en fonctionnement de ses composants est de 1000h. Sur un échantillon de 50 composants, on mesure une moyenne de 998h et un écart-type de 15h. Peut-on dire, au risque de 1%, que le fabricant a raison ?

Ici H_0 est l'hypothèse $\mu \geq 1000$. $\alpha = 1\%$, et H_1 est l'hypothèse $\mu < 1000$.

On suppose H_0 . Comme n est grand, on peut supposer que la durée de vie moyenne dans l'échantillon suit une loi normale de paramètres μ et $\sigma^2 = \frac{50}{49} \times 15^2 = 229.6$. On

cherche donc t tel que $p(\mu - t \frac{\sigma}{\sqrt{50}} < \bar{X}) = 1 - \alpha$, donc $p(-t \leq \sqrt{50} \frac{\bar{X} - \mu}{\sigma}) = 1 - \alpha$.

Mais $p(-t \leq \sqrt{50} \frac{\bar{X} - \mu}{\sigma}) = 1 - F(-t) = F(t)$, on cherche donc t tel que $F(t) = 0.99$, soit $t = 2.33$.

Ainsi, on a montré que dans 99% des cas la moyenne d'échantillon sera dans $[\mu - 2.33 \frac{\sigma}{\sqrt{50}}, +\infty[$, donc a fortiori dans $[1000 - 2.33 \frac{\sigma}{\sqrt{50}}, +\infty[$ car, puisque l'on suppose H_0 , μ est au moins égal à 1000.

Ici, l'intervalle au risque 1% est donc $[995, +\infty[$: on accepte l'hypothèse.

11.3 Test de comparaison de moyennes

Il s'agit cette fois de comparer deux populations, à l'aide d'un échantillon de chacune : on considère les caractéristiques de deux populations, décrites par des variables X_1 et X_2 d'espérances μ_1 et μ_2 et de variances σ_1^2 et σ_2^2 .

Dans chaque population on prélève un échantillon d'effectif n_1 et n_2 , dont on calcule les moyennes \bar{x}_1 et \bar{x}_2 .

Si \bar{x}_1 et \bar{x}_2 sont proches, peut-on dire que $\mu_1 = \mu_2$?

Si \bar{X}_1 et \bar{X}_2 suivent une loi normale (c'est en particulier vrai si X_1 et X_2 suivent une loi normale, ou approximativement si X_1 et X_2 sont quelconques et n_1 et n_2 sont au moins égaux à 30), alors : $\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)$.

Si on suppose $H_0 : \mu_1 = \mu_2$, on a $\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1)$. Donc, un risque de première espèce α étant fixé, on peut déterminer un intervalle contenant $\bar{X}_1 - \bar{X}_2$ avec une probabilité $1 - \alpha$; s'il contient $\bar{x}_1 - \bar{x}_2$, on accepte H_0 , sinon on la refuse.

remarque 1 : en pratique on utilise souvent les estimations ponctuelles des variances s_1^2 et s_2^2 pour valeurs de σ_1^2 et σ_2^2 .

remarque 2 : dans le cas général, si les effectifs sont petits il faut utiliser une loi de Student et non une loi normale.

exemple : pour un échantillon de 40 lampes de type 1, la durée de vie moyenne est $\bar{x}_1 = 1025$ h et une estimation ponctuelle de l'écart-type $s_1 = 120$ h. Pour 40 lampes de type 2, $\bar{x}_2 = 1070$ h et $s_2 = 140$ h.

Peut-on affirmer au risque $\alpha = 0.05$ que les lampes 1 ont une durée de vie moyenne supérieure à celle des lampes 2 ?

$$p(-t \leq \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \leq t) = 0.95 \text{ pour } t = 1.96.$$

Donc l'intervalle $[-1.96\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}, 1.96\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}] = [-57.1, 57.1]$ contient $\bar{X}_1 - \bar{X}_2$ dans 95% des cas. Comme il contient $\bar{x}_1 - \bar{x}_2 = -45$: la différence des durées de vie n'est pas significative au risque 5%, on accepte H_0 .

11.4 Test du χ^2

Avec ce test, basé sur le même principe que les précédents, on étudie non pas la compatibilité des mesures avec une hypothèse sur une valeur moyenne, mais avec une hypothèse portant sur la distribution de l'ensemble des valeurs.

Ainsi on dispose d'une série statistique $(x_1, n_1), \dots, (x_k, n_k)$ qui représente n mesures d'une quantité X [les x_i pouvant éventuellement être des classes regroupant plusieurs valeurs].

Ces mesures sont-elles compatibles avec une loi pour X donnant les probabilités p_i pour chaque valeur (ou classe) x_i ?

Le test suivant permet de déterminer, avec un risque d'erreur de 5%, si la loi de X est compatible avec l'expérience :

- on choisit de tester l'hypothèse $H_0 : \ll p_1 = \frac{n_1}{n}, \dots, p_k = \frac{n_k}{n} \gg$;
- on fixe le risque α de première espèce (la probabilité de refuser l'hypothèse H_0 alors qu'elle est vraie) ;
- on définit $\nu = k - 1$ en général, et $\nu = k - 2$ si X suit une loi binomiale, une loi de Poisson ou une loi normale ;
- on calcule alors le nombre $\chi_o^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} = n \sum_{i=1}^k \frac{(f_i - p_i)^2}{p_i}$, les $f_i = n_i/n$ étant les fréquences observées ;
- on détermine, dans la table du χ^2 à ν degrés de liberté la valeur c^2 telle que $p(\chi^2 \geq c^2) = \alpha$;
- Alors : si $\chi_o^2 \leq c^2$, on accepte H_0 , sinon on la refuse.

exemple : On jette 100 fois de suite un dé à 6 faces. On obtient les résultats :

valeur d en mm	1	2	3	4	5	6
effectif	10	17	11	15	19	28

Peut-on affirmer avec un risque de 5% que le dé est équilibré ?

Si c'est le cas, on a $p_i = 1/6$ et $np_i = 100/6$, donc $\chi_o^2 = \frac{6}{100} \sum_{i=1}^6 (n_i - \frac{100}{6})^2 = \frac{6}{100} \frac{640}{3} = 12.8$.

La loi est uniforme (si le dé est équilibré, les six résultats sont équiprobables) donc on lit $c^2 = 11.07$ dans la table du χ^2 à 5 degrés de liberté. On refuse donc l'hypothèse que le dé est équilibré (au risque 5% que le dé soit en fait équilibré).

En revanche, si l'on avait pris $\alpha = 1\%$, on aurait trouvé $c^2 = 15.09$, et on aurait accepté l'hypothèse.

remarque 1 : pour que ce test soit valable, il faut que les effectifs théoriques np_i soient tous au moins égaux à 5. Dans le cas contraire, on est obligé de regrouper plusieurs valeurs pour arriver à des effectifs supérieurs ou égaux à 5.

remarque 2 : le fait que l'hypothèse H_0 soit acceptée ne signifie pas que l'on a trouvé la loi de X ...seulement que celle que l'on envisage est, avec un risque α connu, compatible avec l'expérience.

Références

- [1] M.Morin **Polycopié de statistiques et probabilités**, Département de Mesures Physiques, IUT 1 de Grenoble, 1981.
- [2] S.Olympieff **Notes de cours manuscrites**, Département de Mesures Physiques, IUT 1 de Grenoble.
- [3] B.Verlant, G.Saint-Pierre **Statistiques et probabilités**, Editions Foucher, ISBN : 2-216-08885-4 : un livre de cours de BTS informatique de gestion, au programme un peu plus limité (pas de loi de Student ou du χ^2).
- [4] J.M.Bernabotto **Cours de statistiques**, <http://infiniment.free.fr/stats.html> : un cours de probas-stats d'IUT très proche de celui-ci
- [5] K.Protassov **Analyse statistique des données expérimentales**, Editions EDP Sciences, ISBN : 2-86883-590-2 : d'un niveau plus élevé que ce cours, un livre très expérimental à l'usage des physiciens.
- [6] La dernière version mise-à-jour et corrigée de ce poly se trouve à l'adresse <http://maths.tetras.org>

Quatrième partie

Annexe : tables des principales lois

On donne ici quelques tables des lois usuelles (loi de Poisson, loi normale centrée réduite, loi de Student, loi du χ^2), présentée sous une forme directement utilisable en suivant les propositions du cours qui précède.

11.5 Loi de Poisson

si X suit une loi de Poisson de paramètre λ , on lit à l'intersection de la ligne k et de la colonne λ les valeurs de $p(X = k) = e^{-\lambda} \lambda^k / k!$ et $p(X \leq k) = \sum_{i=0}^k e^{-\lambda} \lambda^i / i!$

k	$\lambda = 0.1$		$\lambda = 0.3$		$\lambda = 0.5$		$\lambda = 0.7$		$\lambda = 0.9$	
	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$
0	0.9048	0.9048	0.7408	0.7408	0.6065	0.6065	0.4966	0.4966	0.4066	0.4066
1	0.0905	0.9953	0.2222	0.9631	0.3033	0.9098	0.3476	0.8442	0.3659	0.7725
2	0.0045	0.9998	0.0333	0.9964	0.0758	0.9856	0.1217	0.9659	0.1647	0.9371
3	0.0002	1.0	0.0033	0.9997	0.0126	0.9982	0.0284	0.9942	0.0494	0.9865
4	0.0	1.0	0.0003	1.0	0.0016	0.9998	0.005	0.9992	0.0111	0.9977
5	0.0	1.0	0.0	1.0	0.0002	1.0	0.0007	0.9999	0.002	0.9997
6	0.0	1.0	0.0	1.0	0.0	1.0	0.0001	1.0	0.0003	1.0
7	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0

k	$\lambda = 1$		$\lambda = 2$		$\lambda = 3$		$\lambda = 4$		$\lambda = 5$	
	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$
0	0.3679	0.3679	0.1353	0.1353	0.0498	0.0498	0.0183	0.0183	0.0067	0.0067
1	0.3679	0.7358	0.2707	0.406	0.1494	0.1991	0.0733	0.0916	0.0337	0.0404
2	0.1839	0.9197	0.2707	0.6767	0.224	0.4232	0.1465	0.2381	0.0842	0.1247
3	0.0613	0.981	0.1804	0.8571	0.224	0.6472	0.1954	0.4335	0.1404	0.265
4	0.0153	0.9963	0.0902	0.9473	0.168	0.8153	0.1954	0.6288	0.1755	0.4405
5	0.0031	0.9994	0.0361	0.9834	0.1008	0.9161	0.1563	0.7851	0.1755	0.616
6	0.0005	0.9999	0.012	0.9955	0.0504	0.9665	0.1042	0.8893	0.1462	0.7622
7	0.0001	1.0	0.0034	0.9989	0.0216	0.9881	0.0595	0.9489	0.1044	0.8666
8	0.0	1.0	0.0009	0.9998	0.0081	0.9962	0.0298	0.9786	0.0653	0.9319
9	0.0	1.0	0.0002	1.0	0.0027	0.9989	0.0132	0.9919	0.0363	0.9682
10	0.0	1.0	0.0	1.0	0.0008	0.9997	0.0053	0.9972	0.0181	0.9863
11	0.0	1.0	0.0	1.0	0.0002	0.9999	0.0019	0.9991	0.0082	0.9945
12	0.0	1.0	0.0	1.0	0.0001	1.0	0.0006	0.9997	0.0034	0.998
13	0.0	1.0	0.0	1.0	0.0	1.0	0.0002	0.9999	0.0013	0.9993
14	0.0	1.0	0.0	1.0	0.0	1.0	0.0001	1.0	0.0005	0.9998
15	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0002	0.9999
16	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0

k	$\lambda = 6$		$\lambda = 7$		$\lambda = 8$		$\lambda = 9$		$\lambda = 10$	
	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$	$p(X = k)$	$p(X \leq k)$
0	0.0025	0.0025	0.0009	0.0009	0.0003	0.0003	0.0001	0.0001	0.0	0.0
1	0.0149	0.0174	0.0064	0.0073	0.0027	0.003	0.0011	0.0012	0.0005	0.0005
2	0.0446	0.062	0.0223	0.0296	0.0107	0.0138	0.005	0.0062	0.0023	0.0028
3	0.0892	0.1512	0.0521	0.0818	0.0286	0.0424	0.015	0.0212	0.0076	0.0103
4	0.1339	0.2851	0.0912	0.173	0.0573	0.0996	0.0337	0.055	0.0189	0.0293
5	0.1606	0.4457	0.1277	0.3007	0.0916	0.1912	0.0607	0.1157	0.0378	0.0671
6	0.1606	0.6063	0.149	0.4497	0.1221	0.3134	0.0911	0.2068	0.0631	0.1301
7	0.1377	0.744	0.149	0.5987	0.1396	0.453	0.1171	0.3239	0.0901	0.2202
8	0.1033	0.8472	0.1304	0.7291	0.1396	0.5925	0.1318	0.4557	0.1126	0.3328
9	0.0688	0.9161	0.1014	0.8305	0.1241	0.7166	0.1318	0.5874	0.1251	0.4579
10	0.0413	0.9574	0.071	0.9015	0.0993	0.8159	0.1186	0.706	0.1251	0.583
11	0.0225	0.9799	0.0452	0.9467	0.0722	0.8881	0.097	0.803	0.1137	0.6968
12	0.0113	0.9912	0.0263	0.973	0.0481	0.9362	0.0728	0.8758	0.0948	0.7916
13	0.0052	0.9964	0.0142	0.9872	0.0296	0.9658	0.0504	0.9261	0.0729	0.8645
14	0.0022	0.9986	0.0071	0.9943	0.0169	0.9827	0.0324	0.9585	0.0521	0.9165
15	0.0009	0.9995	0.0033	0.9976	0.009	0.9918	0.0194	0.978	0.0347	0.9513
16	0.0003	0.9998	0.0014	0.999	0.0045	0.9963	0.0109	0.9889	0.0217	0.973
17	0.0001	0.9999	0.0006	0.9996	0.0021	0.9984	0.0058	0.9947	0.0128	0.9857
18	0.0	1.0	0.0002	0.9999	0.0009	0.9993	0.0029	0.9976	0.0071	0.9928
19	0.0	1.0	0.0001	1.0	0.0004	0.9997	0.0014	0.9989	0.0037	0.9965
20	0.0	1.0	0.0	1.0	0.0002	0.9999	0.0006	0.9996	0.0019	0.9984
21	0.0	1.0	0.0	1.0	0.0001	1.0	0.0003	0.9998	0.0009	0.9993
22	0.0	1.0	0.0	1.0	0.0	1.0	0.0001	0.9999	0.0004	0.9997
23	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0002	0.9999
24	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0001	1.0
25	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0

Pour $\lambda > 10$, on utilise la loi normale en estimant $p_k = p(X = k)$ par $p(k - 0.5 < X' < k + 0.5)$, avec $X' \sim N(\lambda, \lambda)$ suivant une loi normale de même espérance $\mu = \lambda$ et même variance $\sigma^2 = \lambda$.

11.6 Loi normale centrée réduite

	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

le tableau donne avec une précision de 10^{-4} les valeurs $F(x)$ de la fonction de répartition de la loi normale centrée réduite,

$$F(x) = p(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Si $0 \leq x \leq 3.99$, on place ses deux premiers chiffres en ligne et son troisième en colonne : la valeur de $F(x)$ se trouve à l'intersection.

Par exemple, $F(2.14)$ est à l'intersection de la ligne "2.1" et de la colonne "0.04".

Pour x négatif, on utilise la règle $F(-x) = 1 - F(x)$.

11.7 Loi de Student

$\alpha \rightarrow$	0.9	0.7	0.5	0.2	0.1	0.05	0.02	0.01	0.002	0.001
$1 - \alpha \rightarrow$	0.1	0.3	0.5	0.8	0.9	0.95	0.98	0.99	0.998	0.999
$\downarrow n$										
1	0.1584	0.5095	1.0	3.0777	6.3138	12.7062	31.8205	63.6567	318.3088	636.6192
2	0.1421	0.4447	0.8165	1.8856	2.92	4.3027	6.9646	9.9248	22.3271	31.5991
3	0.1366	0.4242	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145	12.924
4	0.1338	0.4142	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.1322	0.4082	0.7267	1.4759	2.015	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.1311	0.4043	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.1303	0.4015	0.7111	1.4149	1.8946	2.3646	2.998	3.4995	4.7853	5.4079
8	0.1297	0.3995	0.7064	1.3968	1.8595	2.306	2.8965	3.3554	4.5008	5.0413
9	0.1293	0.3979	0.7027	1.383	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.1289	0.3966	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	0.1286	0.3956	0.6974	1.3634	1.7959	2.201	2.7181	3.1058	4.0247	4.437
12	0.1283	0.3947	0.6955	1.3562	1.7823	2.1788	2.681	3.0545	3.9296	4.3178
13	0.1281	0.394	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123	3.852	4.2208
14	0.128	0.3933	0.6924	1.345	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	0.1278	0.3928	0.6912	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	0.1277	0.3923	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.015
17	0.1276	0.3919	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	0.1274	0.3915	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	0.1274	0.3912	0.6876	1.3277	1.7291	2.093	2.5395	2.8609	3.5794	3.8834
20	0.1273	0.3909	0.687	1.3253	1.7247	2.086	2.528	2.8453	3.5518	3.8495
21	0.1272	0.3906	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	0.1271	0.3904	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188	3.505	3.7921
23	0.1271	0.3902	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073	3.485	3.7676
24	0.127	0.39	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	0.1269	0.3898	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	0.1269	0.3896	0.684	1.315	1.7056	2.0555	2.4786	2.7787	3.435	3.7066
27	0.1268	0.3894	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707	3.421	3.6896
28	0.1268	0.3893	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	0.1268	0.3892	0.683	1.3114	1.6991	2.0452	2.462	2.7564	3.3962	3.6594
30	0.1267	0.389	0.6828	1.3104	1.6973	2.0423	2.4573	2.75	3.3852	3.646
40	0.1265	0.3881	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.551
50	0.1263	0.3875	0.6794	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614	3.496
60	0.1262	0.3872	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317	3.4602
70	0.1261	0.3869	0.678	1.2938	1.6669	1.9944	2.3808	2.6479	3.2108	3.435
80	0.1261	0.3867	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387	3.1953	3.4163
90	0.126	0.3866	0.6772	1.291	1.662	1.9867	2.3685	2.6316	3.1833	3.4019
100	0.126	0.3864	0.677	1.2901	1.6602	1.984	2.3642	2.6259	3.1737	3.3905
200	0.1258	0.3859	0.6757	1.2858	1.6525	1.9719	2.3451	2.6006	3.1315	3.3398
500	0.1257	0.3855	0.675	1.2832	1.6479	1.9647	2.3338	2.5857	3.1066	3.3101
1000	0.1257	0.3854	0.6747	1.2824	1.6464	1.9623	2.3301	2.5808	3.0984	3.3003
∞	0.1257	0.3853	0.6745	1.2816	1.6449	1.96	2.3264	2.5758	3.0903	3.2906

si X est une loi de Student à n degrés de liberté,
on lit sur la ligne n et la colonne α
la valeur de $t(\alpha)$ telle que

$$p(|X| \leq t(\alpha)) = 1 - \alpha.$$

11.8 Loi du χ^2

si X est une loi du χ^2 à n degrés de liberté,
on lit sur la ligne n et la colonne α
la valeur de c^2 telle que

$$p(X^2 \geq c^2) = \alpha.$$

$\alpha \rightarrow$	0.999	0.99	0.975	0.95	0.9	0.5	0.1	0.05	0.025	0.01	0.001
$\downarrow n$											
1	0.0	0.0002	0.001	0.0039	0.0158	0.4549	2.7055	3.8415	5.0239	6.6349	10.8276
2	0.002	0.0201	0.0506	0.1026	0.2107	1.3863	4.6052	5.9915	7.3778	9.2103	13.8155
3	0.0243	0.1148	0.2158	0.3518	0.5844	2.366	6.2514	7.8147	9.3484	11.3449	16.2662
4	0.0908	0.2971	0.4844	0.7107	1.0636	3.3567	7.7794	9.4877	11.1433	13.2767	18.4668
5	0.2102	0.5543	0.8312	1.1455	1.6103	4.3515	9.2364	11.0705	12.8325	15.0863	20.515
6	0.3811	0.8721	1.2373	1.6354	2.2041	5.3481	10.6446	12.5916	14.4494	16.8119	22.4577
7	0.5985	1.239	1.6899	2.1673	2.8331	6.3458	12.017	14.0671	16.0128	18.4753	24.3219
8	0.8571	1.6465	2.1797	2.7326	3.4895	7.3441	13.3616	15.5073	17.5345	20.0902	26.1245
9	1.1519	2.0879	2.7004	3.3251	4.1682	8.3428	14.6837	16.919	19.0228	21.666	27.8772
10	1.4787	2.5582	3.247	3.9403	4.8652	9.3418	15.9872	18.307	20.4832	23.2093	29.5883
11	1.8339	3.0535	3.8157	4.5748	5.5778	10.341	17.275	19.6751	21.92	24.725	31.2641
12	2.2142	3.5706	4.4038	5.226	6.3038	11.3403	18.5493	21.0261	23.3367	26.217	32.9095
13	2.6172	4.1069	5.0088	5.8919	7.0415	12.3398	19.8119	22.362	24.7356	27.6882	34.5282
14	3.0407	4.6604	5.6287	6.5706	7.7895	13.3393	21.0641	23.6848	26.1189	29.1412	36.1233
15	3.4827	5.2293	6.2621	7.2609	8.5468	14.3389	22.3071	24.9958	27.4884	30.5779	37.6973
16	3.9416	5.8122	6.9077	7.9616	9.3122	15.3385	23.5418	26.2962	28.8454	31.9999	39.2524
17	4.4161	6.4078	7.5642	8.6718	10.0852	16.3382	24.769	27.5871	30.191	33.4087	40.7902
18	4.9048	7.0149	8.2307	9.3905	10.8649	17.3379	25.9894	28.8693	31.5264	34.8053	42.3124
19	5.4068	7.6327	8.9065	10.117	11.6509	18.3377	27.2036	30.1435	32.8523	36.1909	43.8202
20	5.921	8.2604	9.5908	10.8508	12.4426	19.3374	28.412	31.4104	34.1696	37.5662	45.3147
21	6.4467	8.8972	10.2829	11.5913	13.2396	20.3372	29.6151	32.6706	35.4789	38.9322	46.797
22	6.983	9.5425	10.9823	12.338	14.0415	21.337	30.8133	33.9244	36.7807	40.2894	48.2679
23	7.5292	10.1957	11.6886	13.0905	14.848	22.3369	32.0069	35.1725	38.0756	41.6384	49.7282
24	8.0849	10.8564	12.4012	13.8484	15.6587	23.3367	33.1962	36.415	39.3641	42.9798	51.1786
25	8.6493	11.524	13.1197	14.6114	16.4734	24.3366	34.3816	37.6525	40.6465	44.3141	52.6197
26	9.2221	12.1981	13.8439	15.3792	17.2919	25.3365	35.5632	38.8851	41.9232	45.6417	54.052
27	9.8028	12.8785	14.5734	16.1514	18.1139	26.3363	36.7412	40.1133	43.1945	46.9629	55.476
28	10.3909	13.5647	15.3079	16.9279	18.9392	27.3362	37.9159	41.3371	44.4608	48.2782	56.8923
29	10.9861	14.2565	16.0471	17.7084	19.7677	28.3361	39.0875	42.557	45.7223	49.5879	58.3012
30	11.588	14.953	16.791	18.493	20.599	29.336	40.256	43.773	46.979	50.892	59.703
40	17.916	22.164	24.433	26.509	29.051	39.335	51.805	55.758	59.342	63.691	73.402
50	24.674	29.707	32.357	34.764	37.689	49.335	63.167	67.505	71.42	76.154	86.661
60	31.738	37.485	40.482	43.188	46.459	59.335	74.397	79.082	83.298	88.379	99.607
70	39.036	45.442	48.758	51.739	55.329	69.334	85.527	90.531	95.023	100.425	112.317
80	46.52	53.54	57.153	60.391	64.278	79.334	96.578	101.879	106.629	112.329	124.839
90	54.155	61.754	65.647	69.126	73.291	89.334	107.565	113.145	118.136	124.116	137.208
100	61.918	70.065	74.222	77.929	82.358	99.334	118.498	124.342	129.561	135.807	149.449
250	186.6	200.9	208.1	214.4	221.8	249.3	279.1	287.9	295.7	304.9	324.8
500	407.9	429.4	439.9	449.1	459.9	499.3	540.9	553.1	563.9	576.5	603.4
750	636.0	662.9	676.0	687.5	700.8	749.3	800.0	814.8	827.8	843.0	875.4
1000	867.5	898.9	914.3	927.6	943.1	999.3	1057.7	1074.7	1089.5	1107.0	1143.9
1500	1336.4	1375.5	1394.6	1411.1	1430.2	1499.3	1570.6	1591.2	1609.2	1630.4	1675.0
2000	1810.2	1855.8	1877.9	1897.1	1919.4	1999.3	2081.5	2105.2	2125.8	2150.1	2201.2