Probabilités et statistique

Notes de cours pour MAT 1978

André Giroux
Département de mathématiques et statistique
Université de Montréal
2007

Table des matières

Table des figures

1 Le modèle probabiliste

Dans ce chapitre, nous introduisons le modèle mathématique utilisé pour étudier une expérience aléatoire, apprenons à calculer la probabilité d'un évènement complexe à partir de probabilités d'évènements plus simples et étudions la façon de modifier les probabilités des évènements étant données de nouvelles informations sur l'expérience considérée (probabilités conditionnelles).

1.1 Le modèle réduit

Considérons d'abord une expérience aléatoire admettant un ensemble fini Ω d'issues possibles ω ayant toutes des chances égales de se produire (par symétrie, par exemple). Un **évènement** E est un sous-ensemble de Ω et sa **probabilité** est, par définition,

$$P\{E\} = \frac{|E|}{|\Omega|}$$

(où $|A|$ désigne le nombre d'éléments d'un ensemble fini A).

Les propriétés suivantes sont alors évidentes :

1. Pour tout E, $P\{E\} \geq 0$;

2. $P\{\Omega\} = 1$;

3. si les évènements E_1 et E_2 sont incompatibles (c'est-à-dire si leur intersection $E_1E_2 = E_1 \cap E_2$ est vide), alors

$$P\{E_1 \cup E_2\} = P\{E_1\} + P\{E_2\}.$$

Ces propriétés ont pour conséquences directes les relations suivantes :

4. Si $E^c = \Omega \setminus E$ désigne l'évènement complémentaire de E,

$$P\{E^c\} = 1 - P\{E\};$$

5. Si $E \subseteq F$,

$$P\{E\} \leq P\{F\};$$

6. Quelques soient E_1 et E_2,

$$P\{E_1 \cup E_2\} = P\{E_1\} + P\{E_2\} - P\{E_1E_2\}.$$

Exemple. Lors du jet de deux dés, il y a $36 = |\Omega|$ issues équiprobables $\omega = (i, j)$ où $1 \leq i, j \leq 6$. On a

$$\begin{gathered} P\{i+j \text{ impaire ou } ij \text{ impair }\} \\ = P\{i+j \text{ impaire }\} + P\{ij \text{ impair }\} = \frac{18}{36} + \frac{9}{36} = 0,75 \end{gathered}$$

et

$$\begin{gathered} P\{|i-j| = 2 \text{ ou } ij \text{ impair }\} \\ = P\{|i-j| = 2\} + P\{ij \text{ impair }\} - P\{|i-j| = 2 \text{ et } ij \text{ impair }\} \\ = \frac{8}{36} + \frac{9}{36} - \frac{4}{36} = 0,361. \end{gathered}$$

Remarquons que si les dés sont identiques, l'oeil ne peut distinguer que 21 cas ((i, j) avec $i \leq j$) mais que ces 21 cas ne sont pas équiprobables. On a

$$P\{i+j = 7\} = \frac{6}{36} = 0,167 \neq \frac{3}{21} = 0,143.$$

(En cas de doute sur l'équiprobabilité des cas, l'expérimentation ou la simulation permettent toujours de trancher.)

Exemple. La probabilité d'obtenir au moins un « 6 » en 4 jets d'un dé est ($|\Omega| = 6^4$)

$$1 - \left(\frac{5}{6}\right)^4 = 0,518$$

et celle d'obtenir au moins une paire « 66 » en 24 jets de deux dés est ($|\Omega| = 36^{24}$)

$$1 - \left(\frac{35}{36}\right)^{24} = 0,491.$$

1.2 Éléments de combinatoire

Pour dénombrer les éléments d'un ensemble fini, on s'appuie sur le **principe fondamental** suivant :

> Si une première opération de dénombrement a m_1 résultats possibles et si une deuxième en admet m_2, alors l'opération consistant à effectuer successivement ces deux opérations a $m_1 m_2$ résultats possibles.

1.2.1 Arrangements

Arranger k objets choisis parmi n, c'est les choisir en tenant compte de l'ordre, autrement dit, c'est former un sous-ensemble ordonné avec ces k objets. En vertu du principe fondamental, le nombre d'arrangements de k objets choisis parmi n est

$$A(n,k) = (n)_k = \frac{n!}{(n-k)!}$$

(où $n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot n$ et $0! = 1$).

$A(n,k)$ est donc le nombre de fonctions injectives de $\{1, 2, 3, \ldots, k\}$ dans $\{1, 2, 3, \ldots, n\}$. Le nombre total de fonctions de $\{1, 2, 3, \ldots, k\}$ dans $\{1, 2, 3, \ldots, n\}$ est

$$n^k$$

et est égal au nombre d'arrangements avec répétitions de k objets choisis parmi n.

Lorsque $k = n$, un arrangement (sans répétition) s'appelle **permutation** des objets.

Exemple. Il y a $2\,(n!)^2$ façons de disposer n hommes et n femmes en rangée de telle sorte que femmes et hommes alternent. Si on les dispose autour d'une table ronde et que l'on ne distingue pas deux dispositions obtenues l'une de l'autre par une rotation, il n'y a plus que $(n-1)!\,n!$ façons de faire.

Exemple. Dans une classe de k étudiants, la probabilité qu'au moins deux d'entre eux aient leur anniversaire de naissance le même jour est

$$p_k = 1 - \frac{(365)_k}{365^k}.$$

Ces probabilités p_k croissent avec k et l'on a

$$p_{23} = 0,507.$$

1.2.2 Combinaisons

Combiner k objets choisis parmi n, c'est les choisir sans tenir compte de l'ordre, autrement dit, c'est former un sous-ensemble non ordonné avec ces k objets. Encore en vertu du principe fondamental, le nombre de combinaisons de k objets choisis parmi n est

$$C(n,k) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

puisque l'on a $A(n,k) = C(n,k)\,k!$.

Puisque $C(n,k)$ est le nombre de sous-ensembles de taille k de $\{1,2,3,\ldots,n\}$,

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

Cette relation est un cas particulier du **théorème du binôme** :

$$\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = (x+y)^n.$$

Exemple. On effectue n tirages d'une urne contenant B boules dont R sont rouges. Si les tirages ont lieu avec remise,

$$P\{\text{obtenir } r \text{ boules rouges}\} = \binom{n}{r} \frac{R^r (B-R)^{n-r}}{B^n} = \binom{n}{r} \left(\frac{R}{B}\right)^r \left(1 - \frac{R}{B}\right)^{n-r}$$

et s'ils ont lieu sans remise,

$$P\{\text{obtenir } r \text{ boules rouges}\} = \binom{n}{r} \frac{(R)_r (B-R)_{n-r}}{(B)_n} = \frac{\binom{R}{r}\binom{B-R}{n-r}}{\binom{B}{n}}.$$

Ce modèle des tirages d'une urne s'applique à beaucoup de situations. Par exemple, si l'on génère n entiers au hasard entre 0 et 999 999,

$$P\{\text{obtenir } r \text{ nombres premiers}\} = \binom{n}{r} 0,078498^r \; 0,921502^{n-r}$$

(tirages avec remise, $B = 1\,000\,000$ et $R = 78\,498$). À la loterie « 6/49 »,

$$P\{\text{obtenir au moins 4 bons numéros}\} = \frac{\binom{6}{4}\binom{43}{2} + \binom{6}{5}\binom{43}{1} + \binom{6}{6}\binom{43}{0}}{\binom{49}{6}}$$

$$= 0,001$$

(6 tirages sans remise, $B = 49$ et $R = 6$).

Toujours en vertu du principe fondamental, le nombre de façons de partager un ensemble de taille n en m sous-ensembles de taille respective $n_1, n_2, \ldots, n_m$ $(n_1 + n_2 + \cdots + n_m = n)$ est

$$M(n_1, n_2, \ldots, n_m) = \binom{n}{n_1\ n_2\ \cdots\ n_m} = \frac{n!}{n_1!n_2!\cdots n_m!}$$

puisque

$$M(n_1, n_2, \ldots, n_m) = \binom{n}{n_1}\binom{n - n_1}{n_2}\cdots\binom{n - n_1 - n_2 - \cdots - n_{m-1}}{n_m}.$$

Le **théorème multinomial** est une généralisation du théorème binomial :

$$\sum_{n_1+n_2+\cdots+n_m=n} \binom{n}{n_1\ n_2\ \cdots\ n_m} x_1^{n_1} x_2^{n_2} \cdots x_m^{n_m} = (x_1 + x_2 + \cdots + x_m)^n.$$

Le nombre de termes dans la somme du théorème multinomial est égal au nombre de solutions entières positives $n_1, n_2, \ldots, n_m$ de l'équation

$$n_1 + n_2 + \cdots + n_m = n$$

c'est-à-dire à

$$\binom{n + m - 1}{n}$$

car compter le nombre de façons de placer n objets dans m cellules revient à compter le nombre de façons de disposer n "O" et m "I" sur une ligne de telle sorte que le dernier item à droite soit toujours un "I".

Exemple. Dans une main de poker (5 cartes parmi 52 cartes = 4 couleurs x 13 valeurs),

$$P\{\text{une paire}\} = \frac{\binom{13}{1\ 3\ 9}\binom{4}{2}\binom{4}{1}^3}{\binom{52}{5}} = 0,423\ ,$$

$$P\{\text{deux paires}\} = \frac{\binom{13}{2\ 1\ 10}\binom{4}{2}^2\binom{4}{1}}{\binom{52}{5}} = 0,048$$

et

$$P\{\text{« full »}\} = \frac{\binom{13}{1\ 1\ 11}\binom{4}{3}\binom{4}{2}}{\binom{52}{5}} = 0,001.$$

Exemple. En générant cinq chiffres au hasard,

$$P\{\text{une paire}\} = \frac{\binom{10}{1\ 3\ 6}\binom{5}{2\ 1\ 1\ 1}}{10^5} = 0,504\ ,$$

$$P\{\text{deux paires}\} = \frac{\binom{10}{2\ 1\ 7}\binom{5}{2\ 2\ 1}}{10^5} = 0,108$$

et

$$P\{\text{« full »}\} = \frac{\binom{10}{1\ 1\ 8}\binom{5}{2}}{10^5} = 0,009.$$

Exemple. Le nombre de solutions entières strictement positives $n_1, n_2, \ldots, n_m$ de l'équation

$$n_1 + n_2 + \cdots + n_m = n$$

est

$$\binom{n-1}{n-m}$$

et le nombre de solutions entières positives $n_1, n_2, \ldots, n_m$ de l'inégalité

$$n_1 + n_2 + \cdots + n_m \leq n$$

est

$$\binom{n+m}{n}.$$

Exemple. Si n étudiants montent dans l'ascenseur du pavillon AA, le nombre de façons dont ils peuvent en descendre entre les cinq étages $2, 3, 4, 5$

et 6 est égal à 5^n si on distingue les individus et $\binom{n+4}{n}$ si on ne les distingue pas. Le premier nombre est la valeur de la somme

$$\sum_{n_2+n_3+\cdots+n_6=n} \binom{n}{n_2\ n_3\ \cdots\ n_6}$$

et le second est le nombre de termes qu'elle contient. En l'absence d'information supplémentaires, il faudrait considérer les 5^n cas comme équiprobables. Ainsi, pour $n = 4$,

$$P\{2 \text{ au } 3,\ 1 \text{ au } 4 \text{ et } 1 \text{ au } 6\} = \frac{\binom{4}{2\ 1\ 1}}{5^4} = \frac{12}{625}.$$

1.3 Le modèle complet

Ce modèle est dû à A.N.Kolmogorov (co-inventeur avec G.J. Chaitin et R.J. Solomonoff de la notion de complexité descriptionnelle en informatique) et date de 1933. Il comporte trois éléments : l'ensemble Ω des issues possibles de l'expérience aléatoire, la famille $\mathcal{E}$ des évènements considérés et la fonction P leur attribuant des probabilités (la « mesure » de probabilité).

L'ensemble Ω peut être fini, infini dénombrable ou infini non dénombrable et la famille $\mathcal{E}$ des évènements n'est plus nécessairement égale à l'ensemble de tous les sous-ensembles de Ω. Il suffit qu'elle contienne Ω et soit stable sous les opérations ensemblistes de complémentation et de réunion finie ou dénombrable. Quant aux probabilités, quelle que soit la façon dont elles sont attribuées, il faut et il suffit que la fonction

$$P : \mathcal{E} \to [0, +\infty[$$

satisfasse les deux conditions

P1 $P\{\Omega\} = 1$;

P2 si les évènements $E_1, E_2, E_3, \ldots$ sont deux à deux incompatibles, alors

$$P\left\{\bigcup_{k=1}^{+\infty} E_k\right\} = \sum_{k=1}^{+\infty} P\{E_k\}.$$

En particulier, en vertu de **P2**, si l'évènement E est discret (c'est-à-dire fini ou dénombrablement infini), on aura

$$P\{E\} = \sum_{\omega \in E} P\{\omega\}.$$

Attribuer les probabilités tout en respectant ces contraintes, c'est modéliser le problème considéré — et le faire de façon réaliste peut être quelquefois difficile.

Les axiomes **P1** et **P2** admettent les propriétés 3., 4., 5. et 6. comme conséquences. Cette dernière propriété est d'ailleurs un cas particulier de la formule de Poincaré (aussi appelée **principe d'inclusion-exclusion**) : quels que soient $E_1, E_2, \ldots, E_n$,

$$P\left\{\bigcup_{k=1}^{n} E_k\right\} = \sum_{k=1}^{n} P\{E_k\} - \sum_{1 \le i < j \le n} P\{E_i E_j\}$$
$$+ \sum_{1 \le i < j < k \le n} P\{E_i E_j E_k\} - \cdots + (-1)^{n-1} P\{E_1 E_2 \cdots E_n\}$$

que l'on déduit de 6. par récurrence sur n.

Exemple. Dans l'espace $\Omega = \{0,1\}^*$ des suites binaires finies, on peut poser

$$P\{\omega\} = \frac{1}{2^{2l(\omega)+1}}$$

où $l(\omega)$ désigne la longueur de la suite ω. (Remarquer que l'on attribue ainsi la probabilité $1/2$ ła suite vide ε.) Alors on a bien

$$P\{\Omega\} = \sum_{\omega \in \Omega} P\{\omega\} = \sum_{k=0}^{+\infty} \sum_{l(\omega)=k} \frac{1}{2^{2l(\omega)+1}} = \sum_{k=0}^{+\infty} \frac{1}{2^{k+1}} = 1.$$

$\mathcal{E}$ est ici l'ensemble de tous les sous-ensembles E de Ω et

$$P\{E\} = \sum_{\omega \in E} P\{\omega\}.$$

Dans certaines situations, à la mesure de probabilité que l'on vient de considérer (la mesure uniforme), il est préférable de substituer la mesure de Levin, pour laquelle

$$P\{\omega\} = \frac{1}{2^{K(\omega)}}$$

où $K(\omega)$ est la complexité de Kolmogorov de ω, c'est-à-dire la longueur du plus court programme auto-délimité qui engendre ω.

Exemple. Dans l'espace $\Omega = \{0,1\}^\infty = [0,1]$ des suites binaires infinies, on désigne par Γ_x l'ensemble des nombres réels dont le développement binaire commence par la suite finie x — c'est un intervalle — et on pose

$$P\{\Gamma_x\} = \frac{1}{2^{l(x)}}.$$

Alors P peut être prolongée de façon à ce que la probabilité de n'importe quel intervalle soit égale à sa longueur. La mesure de probabilité ainsi obtenue s'appelle mesure de Lebesgue.

Exemple. Le problème des rencontres.

La probabilité p_n pour qu'une permutation de n objets choisie au hasard, $f : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$, ne contienne aucun point fixe k (aucun point tel que $f(k) = k$) peut être calculée avec la formule de Poincaré. Désignant par E_k l'évènement « $f(k) = k$ », on a, si $n \geq 2$,

$$\begin{aligned}
p_n &= P\{E_1^c E_2^2 \cdots E_n^c\} \\
&= P\{(E_1 \cup E_2 \cup \cdots \cup E_n)^c\} = 1 - P\{E_1 \cup E_2 \cup \cdots \cup E_n\} \\
&= 1 - \left(\sum_{k=1}^n \frac{(n-1)!}{n!} - \sum_{1 \leq i < j \leq n} \frac{(n-2)!}{n!} + \sum_{1 \leq i < j < k \leq n} \frac{(n-3)!}{n!} - \cdots + \frac{(-1)^{n-1}}{n!} \right) \\
&= 1 - \left(\binom{n}{1} \frac{(n-1)!}{n!} - \binom{n}{2} \frac{(n-2)!}{n!} + \binom{n}{3} \frac{(n-3)!}{n!} + \cdots + \frac{(-1)^{n-1}}{n!} \right) \\
&= \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + \frac{(-1)^n}{n!}
\end{aligned}$$

et $p_1 = 0$. On a $p_n \approx e^{-1} = 0,368$ dès que $n \geq 10$.

La probabilité $p_{n,j}$ pour que la permutation admette exactement j points fixes est

$$\begin{aligned}
p_{n,j} &= \sum_{1 \leq i_1 < i_2 < \cdots < i_j \leq n} P\{E_{i_1} E_{i_2} \cdots E_{i_j} E_{i_{j+1}}^c E_{i_{j+2}}^c \cdots E_{i_n}^c\} \\
&= \binom{n}{j} P\{E_1 E_2 \cdots E_j E_{j+1}^c E_{j+2}^c \cdots E_n^c\} \\
&= \binom{n}{j} \frac{(n-j)!\, p_{n-j}}{n!} = \frac{p_{n-j}}{j!}.
\end{aligned}$$

puisque $P\{E_1 E_2 \cdots E_j E_{j+1}^c E_{j+2}^c \cdots E_n^c\} n! =$ le nombre de permutations de n objets laissant exactement les j premiers fixes $=$ le nombre de **dérangements** des $n-j$ derniers objets $= p_{n-j}(n-j)!$.

Exemple. Les fonctions surjectives.

Si $k \geq n$, le nombre $S(n,k)$ de fonctions surjectives de $\{1, 2, \ldots, k\}$ dans $\{1, 2, 3, \ldots, n\}$ peut être calculé à l'aide du principe d'inclusion-exclusion en considérant les évènements E_j « la valeur j est omise » et en raisonnant comme précédemment :

$$\frac{S(n,k)}{n^k} = P\{E_1^c E_2^2 \cdots E_n^c\}$$
$$= P\{(E_1 \cup E_2 \cup \cdots \cup E_n)^c\} = 1 - P\{E_1 \cup E_2 \cup \cdots \cup E_n\}$$

donc :

$$S(n,k) = n^k - \sum_{i=1}^{n} (n-1)^k + \sum_{1 \leq i_1 < i_2 \leq n} (n-2)^k$$
$$- \sum_{1 \leq i_1 < i_2 < i_3 \leq n} (n-3)^k + \cdots + (-1)^{n-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_{n-1} \leq n} (n-(n-1))^k$$
$$= n^k - \binom{n}{1}(n-1)^k + \binom{n}{2}(n-2)^k - \binom{n}{3}(n-3)^k + \cdots + (-1)^{n-1}\binom{n}{n-1}.$$

1.4 Probabilités conditionnelles

Si l'on sait que l'évènement F est réalisé (sans connaître exactement l'issue ω de l'expérience aléatoire), on est amené à réassigner les probabilités des évènements pour en tenir compte. La probabilité conditionnelle de E donné F (avec $P\{F\} > 0$) est, par définition,

$$P\{E/F\} = \frac{P\{EF\}}{P\{F\}}.$$

Cette nouvelle façon d'attribuer les probabilités satisfait les axiomes **P1** et **P2** et toutes les formules du calcul des probabilités s'appliquent autant aux probabilités conditionnelles qu'aux probabilités initiales.

Dans le cas d'un nombre fini d'issues équiprobables, on peut calculer $P\{E/F\}$ en comptant le nombre d'issues de F qui sont aussi dans E puisque

$$P\{E/F\} = \frac{P\{EF\}}{P\{F\}} = \frac{|EF|/|\Omega|}{|F|/|\Omega|} = \frac{|EF|}{|F|}.$$

Les probabilités conditionnelles permettent souvent de calculer facilement les probabilités d'intersections d'évènements :

$$P\{EF\} = P\{E/F\}P\{F\} = P\{F/E\}P\{E\}.$$

On dit que les évènements $F_1, F_2, \ldots, F_n$ forment une partition de Ω si ils sont deux à deux incompatibles et si

$$\Omega = \bigcup_{k=1}^{n} F_k.$$

On a alors, en vertu de **P2**, la **formule des probabilités totales** : quel que soit E,

$$P\{E\} = \sum_{k=1}^{n} P\{E/F_k\}P\{F_k\}.$$

La combinaison des relations précédentes conduit à la **formule de Bayes** : si l'on sait que l'événement E est réalisé, la probabilité à posteriori, $P\{F_j/E\}$, que ce soit la « cause » F_j qui l'ait produit peut s'obtenir des probabilités à priori, $P\{F_k\}$, des diverses « causes » à partir de l'équation

$$P\{F_j/E\} = \frac{P\{E/F_j\}P\{F_j\}}{\sum_{k=1}^{n} P\{E/F_k\}P\{F_k\}}.$$

Exemple. Lors du jet de deux dés, il y a $36 = |\Omega|$ issues équiprobables $\omega = (i, j)$ où $1 \leq i, j \leq 6$. On a

$$P\{i + j \text{ impaire } / \ ij \text{ impair }\} = \frac{0}{9} = 0$$

et

$$P\{|i - j| = 2 \ / \ ij \text{ impaire }\} = \frac{4}{9} = 0,444.$$

Exemple. La probabilité que le deuxième jet de trois dés identiques donne le même résultat que le premier jet peut être calculée à l'aide de la formule des probabilités totales, en conditionnant sur le nombre de chiffres distincts obtenus au premier jet. Désignant par k ce nombre,

$$P\{k = 1\} = \frac{\binom{6}{1}}{216} = \frac{1}{36},$$

$$P\{k=2\} = \frac{\binom{6}{1\ 1\ 4}\binom{3}{2}}{216} = \frac{15}{36}$$

et

$$P\{k=3\} = \frac{\binom{6}{3}3!}{216} = \frac{20}{36}.$$

Donc

$$P\{\text{ jet } 2 = \text{jet } 1\} = \frac{1}{216}\frac{1}{36} + \frac{\binom{3}{2}}{216}\frac{15}{36} + \frac{3!}{216}\frac{20}{36} = 0,004.$$

Exemple. On estime à 40% la proportion des courriels qui sont en fait des pourriels. Un logiciel cherche à bloquer ces derniers d'après leur contenu. Supposons que 5% des bons courriels contiennent l'une des expressions « viagra » ou « occasion exceptionnelle » et que cette proportion monte à 90% pour les pourriels. Si un message est bloqué par le logiciel, la probabilité qu'il s'agisse vraiment d'un pourriel peut être calculée à l'aide de la formule de Bayes.

$$\begin{aligned} &P\{\text{pourriel/bloqué}\} \\ &= \frac{P\{\text{bloqué/pourriel}\}P\{\text{pourriel}\}}{P\{\text{bloqué/pourriel}\}P\{\text{pourriel}\} + P\{\text{bloqué/bon}\}P\{\text{bon}\}} \\ &= \frac{0,90 \times 0,40}{0,90 \times 0,40 + 0,05 \times 0,60} = 0,545. \end{aligned}$$

Exemple. Lorsque l'on génère au hasard deux suites binaires finies (non vides), la probabilité que la plus courte ω_c soit un préfixe de la plus longue ω_l, c'est-à-dire que l'on ait $\omega_l = \omega_c\omega$ pour une suite ω appropriée — éventuellement vide, peut être calculée en conditionnant sur la longueur $l(\omega_l)$ de la plus longue. En supposant que

$$P\{\omega\} = \frac{1}{2^{2l(\omega)+1}},$$

on a

$$P\{l(\omega_l) = k\} = \sum_{l(\omega)=k} P\{\omega\} = \frac{1}{2^{k+1}}$$

et

$$P\{\text{préfixe}\} = \sum_{k=1}^{+\infty} \left(\frac{1}{2^{2+1}} + \frac{1}{2^{4+1}} + \cdots + \frac{1}{2^{2k+1}} \right) \frac{1}{2^{k+1}} = \frac{1}{14} = 0,071.$$

1.5 Indépendance stochastique

Par définition, les évènements E et F sont indépendants si

$$P\{EF\} = P\{E\}P\{F\}.$$

Cela signifie que $P\{E/F\} = P\{E\}$ et que $P\{F/E\} = P\{F\}$: le fait de savoir que l'un des deux évènements est réalisé ne change pas la probabilité que l'autre le soit.

Exemple. Lors du jet de deux dés, les évènements « $i+j$ impaire » et « ij impair » sont incompatibles, donc ne sont *pas* indépendants. Les évènements « $|i-j| = 2$ » et « ij impair » ne le sont pas non plus :

$$P\{|i-j| = 2 \ / \ ij \text{ impaire } \} = \frac{4}{9} \neq P\{|i-j| = 2\} = \frac{2}{9},$$

mais les évènements « $i = j$ » et « i est pair » le sont :

$$P\{i = j \text{ et } i \text{ pair } \} = \frac{1}{12} = \frac{1}{2}\,\frac{1}{6} = P\{i \text{ pair } \}P\{i = j\}.$$

Exemple. Si E et F sont indépendants, E et F^c le sont aussi car

$$P\{EF^c\} = P\{E\} - P\{EF\} = P\{E\}(1 - P\{F\}) = P\{E\}P\{F^c\}.$$

Lorsqu'une expérience aléatoire est répétée dans des circonstances identiques, l'ensemble des issues possibles est $\Omega^2 = \Omega \times \Omega$. Tout évènement de la forme $E_1 \times \Omega$ dont la réalisation ne dépend que de la première épreuve est indépendant de tout évènement de la forme $\Omega \times E_2$ dont la réalisation ne dépend que de la seconde. Par suite, on a

$$\begin{aligned} P\{E_1 \times E_2\} &= P\{E_1 \times \Omega \ \cap \ \Omega \times E_2\} \\ &= P\{E_1 \times \Omega\}P\{\Omega \times E_2\} = P\{E_1\}P\{E_2\}. \end{aligned}$$

Ainsi, lors de n épreuves indépendantes,

$$P\{(\omega_1, \omega_2, \ldots, \omega_n)\} = P\{\omega_1\}P\{\omega_2\} \cdots P\{\omega_n\}.$$

Exemple. Avant le premier lancer, la probabilité d'observer dix fois PILE en dix lancers d'une pièce n'est que de $2^{-10} = 0,001$. Après avoir observés neuf fois PILE, la probabilité d'observer dix fois PILE en dix lancers est de $2^{-1} = 0,5$, en vertu de l'indépendance des lancers.

2 Variables aléatoires

Une variable aléatoire est une fonction $X : \Omega \to \mathbb{R}$. Sa distribution (sa loi) est déterminée par sa **fonction de répartition** $F : \mathbb{R} \to \mathbb{R}$, définie par

$$F(x) = P\{\omega : X(\omega) \le x\}$$

et à partir de laquelle toutes les probabilités pertinentes à X peuvent être calculées. La fonction de répartition croît de 0 à 1 lorsque son argument croît de $-\infty$ à $+\infty$ et une conséquence de l'axiome **P2** est qu'elle est toujours continue par la droite. On distingue les variables aléatoires discrètes dont l'ensemble des valeurs est fini ou dénombrable des variables aléatoires continues dont l'ensemble des valeurs est infini non dénombrable.

Exemple. Lors du lancer d'un dé, soit X le nombre de PILE obtenu. Alors

$$F(x) = \begin{cases} 0 & \text{si } x < 0; \\ 1/2 & \text{si } 0 \le x < 1; \\ 1 & \text{si } 1 \le x. \end{cases}$$

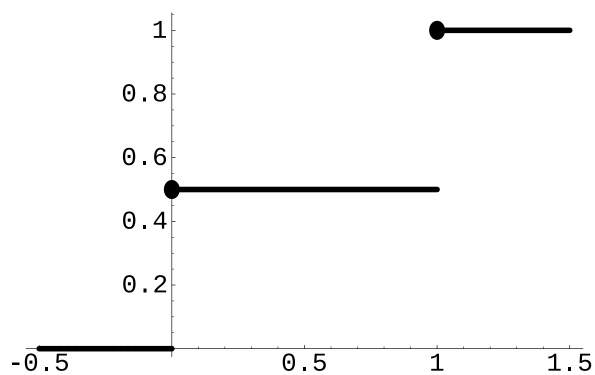

FIG. 1 – La fonction de répartition de X

Exemple. On choisit entre 0 et 1 un nombre réel Y au hasard — au sens où la probabilité que Y soit dans un intervalle $[a, b] \subseteq [0, 1]$ quelconque est égale à sa longueur $b - a$. Alors

$$F(y) = \begin{cases} 0 & \text{si } x < 0; \\ x & \text{si } 0 \le x < 1; \\ 1 & \text{si } 1 \le x. \end{cases}$$

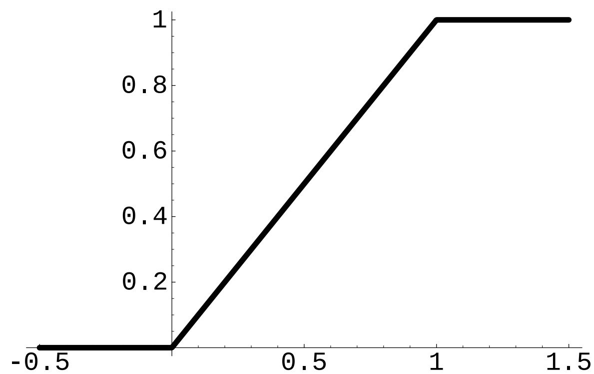

FIG. 2 – La fonction de répartition de Y

2.1 Variables aléatoires discrètes

Pour étudier une variable discrète X, prenant les valeurs x_k , on utilise principalement sa **fonction de masse** :

$$p(x_k) = P\{\omega : X(\omega) = x_k\}.$$

L'**espérance mathématique** de X est

$$\mathbb{E}(X) = \sum_k x_k p(x_k)$$

et sa **variance** est

$$\mathbb{V}(X) = \sum_k (x_k - \mathbb{E}(X))^2 p(x_k).$$

L'espérance est une mesure de positionnement de la variable et la variance est une mesure de sa dispersion. L'**écart-type**

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

a les mêmes unités que la variable.

Si $g : \mathbb{R} \to \mathbb{R}$ est une fonction, $Y = g(X)$ est une nouvelle variable aléatoire discrète dont on peut calculer l'espérance mathématique au moyen de la formule

$$\mathbb{E}(Y) = \sum_k g(x_k)p(x_k).$$

Ainsi, on peut dire que

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

On a

$$\mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b \quad \text{et} \quad \mathbb{V}(aX + b) = a^2\,\mathbb{V}(X).$$

Par conséquent, pour toute variable X, la variable

$$Y = \frac{X - \mathbb{E}(X)}{\sigma(X)}$$

a une espérance mathématique nulle et une variance unité. On dit de Y qu'elle est **centrée réduite**.

Exemple. Lors du lancer d'un dé, soit X le nombre obtenu. Alors

$$\mathbb{E}(X) = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}$$

et

$$\mathbb{V}(X) = \frac{1+4+9+16+25+36}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Exemple. Soient $E \subset \Omega$ et $X_E : \Omega \to \mathbb{R}$ la variable aléatoire définie par

$$X_E(\omega) = \begin{cases} 1 & \text{si } \omega \in E\,; \\ 0 & \text{sinon.} \end{cases}$$

Alors

$$\mathbb{E}(X_E) = P\{E\} \quad \text{et} \quad \mathbb{V}(X_E) = P\{E\}(1 - P\{E\}).$$

Exemple. Soient $\Omega = \{0,1\}^*$ et $X : \Omega \to \mathbb{R}$ la variable aléatoire définie par

$$X(\omega) = l(\omega).$$

Alors, en supposant que

$$P\{\omega\} = \frac{1}{2^{2l(\omega)+1}},$$

on a

$$p(k) = \frac{1}{2^{k+1}}$$

et

$$\mathbb{E}(X) = \sum_{k=1}^{+\infty} \frac{k}{2^{k+1}} = 1.$$

Exemple. Complexité computationnelle moyenne.

L'algorithme de fouille linéaire permet de localiser x dans une liste de longueur n au moyen de $2k+1$ comparaisons si x occupe la position k de la liste et permet de dire que x n'est pas dans la liste au moyen de $2n+2$ comparaisons si x ne figure pas dans la liste.

Soit $X(x)$ le nombre de comparaisons nécessaires sur l'input x. (La liste est supposée donnée). En supposant que x appartient à la liste avec probabilité p et qu'alors il y occupe la position k avec probabilité $1/n$, on a

$$\mathbb{E}(X) = \sum_{k=1}^{n} (2k+1)\frac{p}{n} + (2n+2)(1-p) = 2n+2-np.$$

Par exemple, si les éléments de la liste sont des entiers distincts entre 0 et 999 999 et si x est aussi choisi dans ce domaine, $p = n/10^6$ et

$$\mathbb{E}(X) = 2n+2-\frac{n^2}{10^6}.$$

Un **vecteur aléatoire discret** à deux dimensions est une fonction $(X,Y) : \Omega \to \mathbb{R}^2$ (on dit aussi que les variables aléatoires X et Y sont conjointement distribuées). Sa loi est déterminée par sa **fonction de masse conjointe** :

$$p(x_k, y_j) = P\{\omega : X(\omega) = x_k, Y(\omega) = y_j\}.$$

Les **fonctions de masse marginales** des composantes sont alors calculées au moyen des formules

$$p_X(x_k) = \sum_j p(x_k, y_j) \quad \text{et} \quad p_Y(y_j) = \sum_k p(x_k, y_j).$$

Les variables X et Y sont indépendantes si

$$p(x_k, y_j) = p_X(x_k) p_Y(y_j).$$

Cela signifie que tout évènement n'impliquant que X est indépendant de tout évènement n'impliquant que Y.

Lorsque les variables X et Y ne sont pas indépendantes, les probabilités qui leur sont associées peuvent être calculées au moyen des **fonctions de masse conditionnelles** :

$$p_{X/Y}(x_k/y_j) = \frac{p(x_k, y_j)}{p_Y(y_j)} \quad \text{et} \quad p_{Y/X}(y_j/x_k) = \frac{p(x_k, y_j)}{p_X(x_k)}.$$

Si $g : \mathbb{R}^2 \to \mathbb{R}$ est une fonction, $Z = g(X, Y)$ est une nouvelle variable aléatoire discrète dont on peut calculer l'espérance mathématique au moyen de la formule

$$\mathbb{E}(Z) = \sum_k \sum_j g(x_k, y_j) p(x_k, y_j).$$

Exemple. L'on choisit au hasard un entier X entre 1 et 4 puis un entier Y entre 1 et X. La loi conjointe du couple (X, Y) peut être calculée au moyen de

$$P\{X = x, Y = y\} = P\{Y = y/X = x\}P\{X = x\} = \frac{1}{x}\,\frac{1}{4}$$

pour tout $1 \leq y \leq x \leq 4$. Le tableau suivant donne les lois conjointe et marginales de X et de Y.

X \ Y	1	2	3	4	Loi de X
1	1/4	0	0	0	1/4
2	1/8	1/8	0	0	1/4
3	1/12	1/12	1/12	0	1/4
4	1/16	1/16	1/16	1/16	1/4
Loi de Y	25/48	13/48	7/48	3/48	1

Ce tableau contenant des zéros, X et Y ne sont pas indépendantes. L'espérance mathématique de Y peut être obtenue une fois que sa loi marginale a été calculée :

$$\mathbb{E}(Y) = 1\ \frac{25}{48} + 2\ \frac{13}{48} + 3\ \frac{7}{48} + 4\ \frac{3}{48} = 1,75.$$

Exemple. On choisit au hasard n entiers $X_1, X_2, \ldots, X_n$ entre 0 et 999 999 (tirages avec remise). Les variables $X_1, X_2, \ldots, X_n$ sont donc indépendantes. Si

$$X = \max\{X_1, X_2, \ldots, X_n\},$$

la fonction de masse de X peut être obtenue via sa fonction de répartition :

$$F(x) = P\{X \leq x\} = \left(\frac{x+1}{10^6}\right)^n$$

et

$$p(x) = P\{X = x\} = F(x) - F(x-1) = \frac{(x+1)^n - x^n}{10^{6n}}$$

pour $0 \leq x \leq 999\,999$.

2.2 Lois discrètes particulières

2.2.1 Loi uniforme

X suit une loi uniforme de paramètre n si

$$p(k) = \frac{1}{n} \quad \text{pour} \quad 1 \leq k \leq n.$$

On a

$$\mathbb{E}(X) = \frac{n+1}{2} \quad \text{et} \quad \mathbb{V}(X) = \frac{n(n+1)}{12}.$$

Typiquement, X est le résultat du tirage au sort d'un entier compris entre 1 et n.

Exemple. On choisit au hasard n entiers distincts (tirages sans remise) entre 0 et 999 999. Si X désigne le rang du plus grand, X suit une loi uniforme de paramètre n.

2.2.2 Loi binomiale

X suit une loi binomiale de paramètres n et p si

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{pour} \quad 0 \leq k \leq n.$$

On a

$$\mathbb{E}(X) = np \quad \text{et} \quad \mathbb{V}(X) = np(1-p).$$

Typiquement, X est le nombre de succès obtenus lors de n répétitions dans des conditions identiques d'une expérience aléatoire résultant à chaque fois en un succès avec probabilité p (**épreuves de Bernoulli**, en l'honneur de Jacques Bernoulli, auteur du premier traité de calcul des probabilités en 1713 ; lorsque $n = 1$, une variable binomiale s'appelle aussi une **variable de Bernoulli**).

Exemple. On génère au hasard une suite binaire de longueur 51. Le nombre X de changements (0 à 1 ou 1 à 0) suit une loi binomiale de paramètres $n = 50$ et $p = 1/2$.

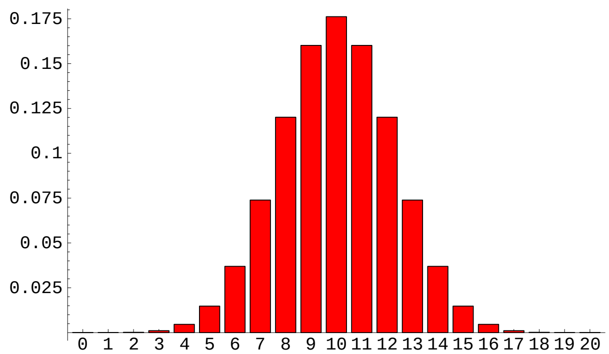

FIG. 3 – Fonction de masse de la loi binomiale, n=20 et p=0.5

2.2.3 Loi hypergéométrique

X suit une loi hypergéométrique de paramètres n, a et A si

$$p(k) = \frac{\binom{a}{k}\binom{A-a}{n-k}}{\binom{A}{n}} \quad \text{pour} \quad a + n - A \leq k \leq a.$$

On a

$$\mathbb{E}(X) = n\frac{a}{A} \quad \text{et} \quad \mathbb{V}(X) = n\frac{a}{A}\left(1 - \frac{a}{A}\right)\frac{A-n}{A-1}.$$

Typiquement, X est le nombre d'individus distingués obtenus lors de n tirages sans remise d'une population de A individus dont a sont distingués. Lorsque A est grand comparé à n, la distribution hypergéométrique est pratiquement la même que la distribution binomiale de paramètres n et $p = a/A$.

Exemple. Pour estimer la population en poissons A d'un lac, on en capture a que l'on marque puis que l'on relâche. On effectue ensuite une pêche de n poissons. L'estimateur que l'on prend pour A est alors la valeur de A

qui rende maximum la probabilité

$$f(A) = \frac{\binom{a}{k}\binom{A-a}{n-k}}{\binom{A}{n}}$$

de ce que l'on a observé (**estimateur à vraisemblance maximale**). Sa valeur est

$$\hat{A} = \left\lfloor \frac{an}{k} \right\rfloor.$$

2.2.4 Loi de Poisson

On a le résultat mathématique suivant :

$$\lim_{n\to+\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}.$$

X suit une loi de Poisson de paramètre λ si

$$p(k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{pour} \quad k \geq 0.$$

On a

$$\mathbb{E}(X) = \mathbb{V}(X) = \lambda.$$

Typiquement, X est le nombre de réalisations par unité de temps constatées lors d'un très grand nombre d'observations (de la réalisation ou non) d'un phénomène ayant une très petite probabilité de se produire. En particulier, une loi binomiale de paramètres n grand et p petit est bien approximée par une loi de Poisson de paramètre $\lambda = np$.

Exemple. Supposons que la distribution du nombre de clients à un guichet automatique varie peu durant la journée et que, pour chaque période d'une heure, ce nombre suit une loi de Poisson de moyenne 10. (Beaucoup de personnes passent devant ce guichet, très peu s'y arrêtent). La probabilité que plus de 15 clients se présentent entre midi et treize heures est

$$1 - \sum_{k=0}^{15} e^{-10}\frac{10^k}{k!} = 0,049.$$

Le probabilité que le nombre de clients entre quatorze et seize heures soit d'au plus 25 est

$$\sum_{k=0}^{25} e^{-20}\frac{20^k}{k!} = 0,888.$$

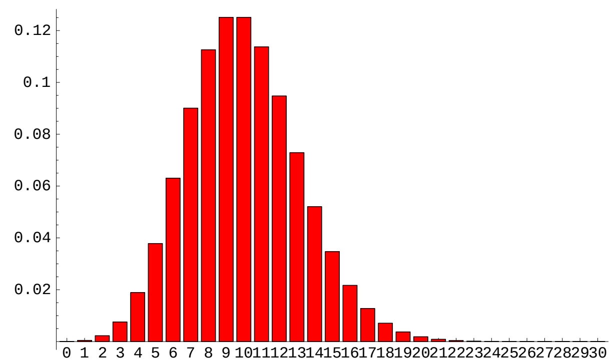

FIG. 4 – Fonction de masse de la loi de Poisson, $\lambda = 10$

2.3 Variables aléatoires continues

La distribution d'une variable aléatoire continue X peut habituellement être décrite au moyen de sa **fonction de densité de probabilité** $f : \mathbb{R} \to \mathbb{R}$, définie par la condition que, quel que soit $B \subseteq \mathbb{R}$,

$$P\{\omega : X(\omega) \in B\} = \int_B f(x)\ dx.$$

On doit donc avoir $f(x) \geq 0$ et

$$\int_{-\infty}^{+\infty} f(x)\ dx = 1.$$

Cependant, $f(x)$ n'est *pas* une probabilité et il est possible que $f(x) > 1$. Puisque, si $\Delta x > 0$ est petit, on a

$$P\left\{x - \frac{\Delta x}{2} < X \leq x + \frac{\Delta x}{2}\right\} = \int_{x-\Delta x/2}^{x+\Delta x/2} f(t)\ dt \approx f(x)\Delta x,$$

l'interprétation probabiliste de $f(x)$ est que $f(x)\Delta x$ est la probabilité que $X = x$ à « Δx près ». La fonction de densité de probabilité joue pour les variables continues un rôle analogue à celui que la fonction de masse joue pour les variables discrètes, les intégrales remplaçant les sommes.

Les relations entre fonction de répartition et fonction de densité de probabilité sont données par

$$F(x) = \int_{-\infty}^{x} f(t)\ dt \quad \text{et} \quad f(x) = \frac{d}{dx}F(x).$$

De plus,

$$P\{X = x\} = \lim_{\Delta x \to 0}(F(x + \Delta x) - F(x - \Delta x)) = F(x) - F(x-).$$

En particulier, $P\{X = x\} = 0$ pour tout x sauf aux points de discontinuité de F. Ceci n'implique pas que $P\{E\} = 0$ pour tout évènement E puisque l'axiome **P2** ne porte que sur les réunions *dénombrables* et ceci ne signifie pas non plus que le modèle probabiliste est irréaliste puisque qu'un nombre réel x est une abstraction mathématique dont on ne connaît jamais que quelques décimales.

Exemple. Soit X un nombre réel choisi au hasard entre 0 et 1. Alors quel, que soit x,

$$P\{X = x\} = 0$$

mais, en fait, c'est Γ_x qui est observé et

$$P\{X \in \Gamma_x\} = \frac{1}{2^{l(x)}}.$$

L'**espérance mathématique** de X est

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)\ dx$$

et sa **variance** est

$$\mathbb{V}(X) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f(x)\ dx.$$

L'espérance est une mesure de positionnement de la variable et la variance est une mesure de sa dispersion. L'**écart-type**

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

a les mêmes unités que la variable. Une variable d'espérance nulle et de variance unité est dite centrée réduite.

Si $g : \mathbb{R} \to \mathbb{R}$ est une fonction continue, $Y = g(X)$ est une nouvelle variable aléatoire continue dont on peut calculer l'espérance mathématique au moyen de la formule

$$\mathbb{E}(Y) = \int_{-\infty}^{+\infty} g(x)f(x)\ dx.$$

Ainsi,

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Un **vecteur aléatoire continu** à deux dimensions est une fonction $(X, Y) : \Omega \to \mathbb{R}^2$ (on dit aussi que les variables aléatoires X et Y sont conjointement distribuées). Sa loi est déterminée par sa **fonction de densité de probabilité conjointe** $f : \mathbb{R}^2 \to \mathbb{R}$, définie par la condition que quelque soit $B \subseteq \mathbb{R}^2$,

$$P\{(X, Y) \in B\} = \iint_B f(x, y)\ dxdy.$$

Les **fonctions de densité de probabilité marginales** des composantes sont alors calculées au moyen des formules

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)\ dy \quad \text{et} \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)\ dx.$$

Les variables X et Y sont indépendantes si

$$f(x, y) = f_X(x)f_Y(y).$$

Cela signifie que tout évènement dont la définition ne fait intervenir que X est indépendant de tout évènement dont la définition ne fait intervenir que Y.

Lorsque les variables X et Y ne sont pas indépendantes, les probabilités qui leur sont associées peuvent être calculées au moyen des **fonctions de densité conditionnelles** :

$$f_{X/Y}(x/y) = \frac{f(x, y)}{f_Y(y)} \quad \text{et} \quad f_{Y/X}(y/x) = \frac{f(x, y)}{f_X(x)}.$$

Si $g : \mathbb{R}^2 \to \mathbb{R}$ est une fonction continue, $Z = g(X, Y)$ est une nouvelle variable aléatoire continue dont on peut calculer l'espérance mathématique au moyen de la formule

$$\mathbb{E}(Z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y)\ dxdy.$$

Exemple. Si $g(x, y) = x + y$, on a

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f(x, y)\ dxdy$$
$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y)\ dxdy + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y)\ dxdy$$

et

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Si $g(x, y) = xy$ et *si les variables sont indépendantes*, on a

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x, y)\ dxdy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_X(x) f_Y(y)\ dxdy$$
$$= \int_{-\infty}^{+\infty} x f_X(x) dx \int_{-\infty}^{+\infty} y f_Y(y) dy$$

et

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

2.4 Lois continues particulières

2.4.1 Loi uniforme

X suit une loi uniforme de paramètres a et b si

$$f(x) = \begin{cases} \dfrac{1}{b - a} & \text{si } a < x < b; \\ & \\ 0 & \text{sinon.} \end{cases}$$

Donc, si $(c, d) \subseteq (a, b)$,

$$P\{X \in (c, d)\} = \frac{d - c}{b - a}.$$

On a

$$\mathbb{E}(X) = \frac{b+a}{2} \quad \text{et} \quad \mathbb{V}(X) = \frac{(b-a)^2}{12}.$$

Typiquement, X est un nombre réel choisi au hasard dans l'intervalle (a, b).

De façon semblable, on dit que le vecteur aléatoire (X, Y) suit une loi uniforme dans le domaine $D \subseteq \mathbb{R}^2$ si

$$f(x,y) = \begin{cases} \dfrac{1}{\text{aire de } D} & \text{si } (x,y) \in D; \\ \\ 0 & \text{sinon.} \end{cases}$$

Cela correspond à choisir un point au hasard dans D. Pour tout domaine $B \subseteq D$,

$$P\{(X,Y) \in B\} = \frac{\text{aire de } B}{\text{aire de } D}.$$

Exemple. Supposons que la fréquence des trains sur une ligne de métro soit d'un toutes les 7 minutes. Alors le temps d'attente X de la prochaine rame pour un usager sera une variable aléatoire uniforme sur l'intervalle $(0, 7)$ et il attendra en moyenne $3,5$ minutes. S'il attend depuis 2 minutes, la probabilité qu'il attende encore 2 minutes est

$$P\{X > 4/X > 2\} = \frac{P\{X > 4\}}{P\{X > 2\}} = \frac{3/7}{5/7} = 0,6.$$

Exemple. Les racines de l'équation quadratique $ax^2 + bx + c = 0$ sont

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Supposons que B soit un nombre choisi au hasard entre -10 et 10. Alors la probabilité que les racines de l'équation $x^2 + Bx + 1 = 0$ soient réelles est

$$P\{B^2 - 4 > 0\} = P\{B > 2\} + P\{B < -2\} = 2\,\frac{8}{20} = 0,8.$$

Exemple. Les valeurs propres de la matrice symétrique

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

sont

$$\frac{a + c \pm \sqrt{(a-c)^2 + 4b^2}}{2}.$$

Supposons que A et C soient des variables aléatoires uniformes et indépendantes sur l'intervalle $(0, 2)$. Alors la fonction de densité de probabilité du vecteur aléatoire (A, C) sera

$$f(a,c) = \begin{cases} 1/4 & \text{si } 0 < x, y < 2; \\ 0 & \text{sinon.} \end{cases}$$

et la probabilité que les valeurs propres de la matrice

$$\begin{pmatrix} A & 1 \\ 1 & C \end{pmatrix}$$

soient positives est (figure(5))

$$\begin{aligned} P\{AC > 1\} &= \iint_{ac>1} f(a,c)\, dadc = \frac{1}{4}\int_{1/2}^{2}\left(2 - \frac{1}{a}\right)\, da \\ &= \frac{3 - 2\ln 2}{4} = 0,403. \end{aligned}$$

Exemple. La méthode de Monte-Carlo.
Si $0 \leq f(x) \leq 1$ lorsque $0 \leq x \leq 1$, on peut écrire que

$$\int_0^1 f(x)\ dx = P\{Y \leq f(X)\}$$

où (X, Y) est un vecteur aléatoire uniforme dans le carré $[0, 1] \times [0, 1]$ puisque l'intégrale et la probabilité sont toutes deux égales à l'aire sous la courbe $y = f(x)$. On peut donc estimer l'intégrale en générant un grand nombre de points au hasard dans le carré et en prenant pour valeur de l'intégrale la proportion de ceux qui satisfont l'inégalité. Ainsi

$$\int_0^1 e^{-x^2/2}\ dx = 0,856.$$

Exemple. Si (X, Y) est un vecteur uniforme dans le disque $x^2 + y^2 < 1$, la fonction de densité marginale de X est

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\ dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\ dy = \frac{2}{\pi}\sqrt{1-x^2} \text{ si } |x| < 1$$

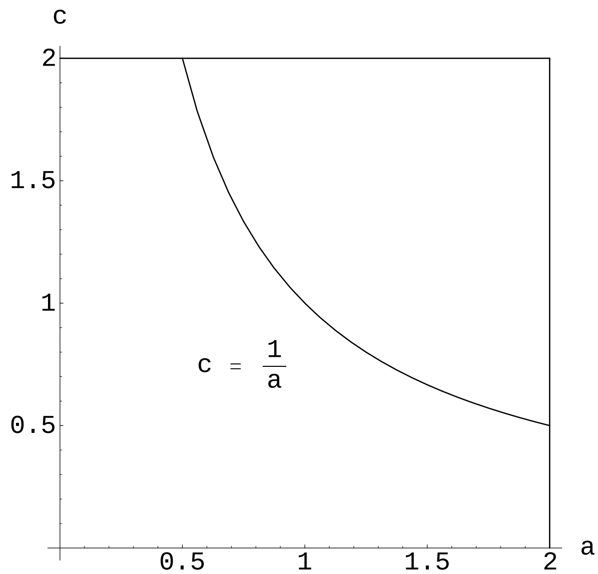

FIG. 5 – Le domaine d'intégration pour les valeurs propres

et l'expression pour $f_Y(y)$ est semblable. Puisque

$$f_X(x)f_Y(y) \neq f(x,y),$$

X et Y ne sont pas indépendantes. La fonction de densité conditionnelle de X donnée Y est

$$f_{X/Y}(x/y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{2\sqrt{1-y^2}} \text{ si } |x| < \sqrt{1-y^2};$$

la valeur y de Y donnée, X est uniforme sur l'intervalle $(-\sqrt{1-y^2}, \sqrt{1-y^2})$.

2.4.2 Loi exponentielle

X suit une loi exponentielle de paramètre λ si

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0; \\ 0 & \text{sinon.} \end{cases}$$

Typiquement, X est le temps d'attente de la prochaine réalisation d'un phénomène gouverné par une loi de Poisson de paramètre λ. En effet, si Y

est le nombre de réalisations du phénomène durant un temps x, Y suit une loi de Poisson de paramètre λx et, si $x > 0$,

$$1 - F(x) = P\{X > x\} = P\{Y = 0\} = e^{-\lambda x}$$

d'où

$$-f(x) = -\lambda e^{-\lambda x}.$$

On a

$$\mathbb{E}(X) = \frac{1}{\lambda} \quad \text{et} \quad \mathbb{V}(X) = \frac{1}{\lambda^2}.$$

Le moment à partir duquel on commence à mesurer le temps d'attente n'a aucune importance. En effet, une variable exponentielle est « sans mémoire ». On a

$$P\{X > x + t \;/\; X > t\} = \frac{P\{X > x + t\}}{P\{X > t\}} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x} = P\{X > x\}.$$

Exemple. Le nombre annuel de pannes du système d'exploitation « Vitrine 99 » suit une loi de Poisson de paramètre $\lambda = 2,5$. Soit X le temps d'attente (en années) de la prochaine panne. Mesurant le temps à partir du premier janvier, la probabilité qu'il n'y ait aucune panne avant le premier avril est

$$P\{X > 90/365\} = \int_{90/365}^{+\infty} (2,5)e^{-2,5x}\ dx = 0,540.$$

En démarrant le système le premier juillet, la probabilité qu'il n'y ait aucune panne avant la fin de l'année est

$$P\{X > 181/365\} = \int_{181/365}^{+\infty} (2,5)e^{-2,5x}\ dx = 0,289.$$

L'espérance mathématique de X est 146 jours.

2.4.3 Loi normale

Si Y suit une loi binomiale de paramètres n et p, les valeurs de la variable centrée réduite

$$\frac{Y - np}{\sqrt{np(1-p)}}$$

vont de $-\sqrt{np/(1-p)}$ à $\sqrt{n(1-p)/p}$ et ne diffèrent l'une de l'autre que de $1/\sqrt{np(1-p)}$. Lorsque n tend vers l'infini, cette variable s'approche donc d'une variable continue prenant toutes les valeurs réelles. On peut montrer que, quels que soient a et b,

$$\lim_{n \to +\infty} P\left\{a < \frac{Y - np}{\sqrt{np(1-p)}} \leq b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx$$

(théorème de de Moivre-Laplace).

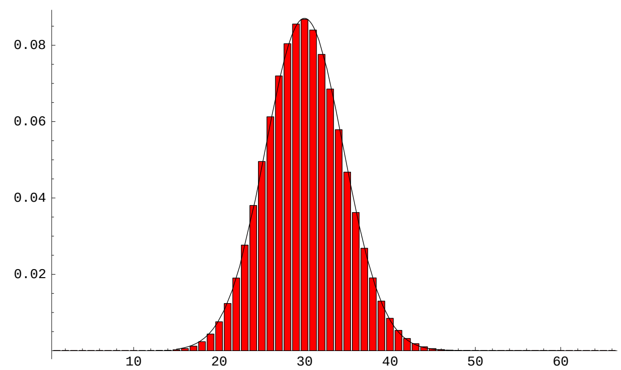

FIG. 6 – Le théorème de deMoivre-Laplace, $n = 100$ et $p = 0,3$

Z suit une loi normale standard ($Z \sim N(0,1)$) si

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}.$$

On a

$$\mathbb{E}(Z) = 0\ ,\ \mathbb{V}(Z) = 1.$$

La fonction de répartition d'une normale standard est aussi dénotée par un symbole particulier :

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

Par symétrie,

$$\Phi(-x) + \Phi(x) = 1$$

de telle sorte que les tables de $\Phi(x)$ ne sont nécessaires que pour les valeurs positives de x. Les **quantiles** z_α sont définis pour $0 < \alpha < 1$ par

$$P\{Z > z_\alpha\} = 1 - \Phi(z_\alpha) = \alpha.$$

On a

$$z_{0,05} = 1,645 \ , \ z_{0,025} = 1,96 \ , \ z_{0,01} = 2,33 \ \text{ et } \ z_{0,005} = 2,58.$$

X suit une loi normale de paramètres μ et σ^2 ($X \sim N(\mu, \sigma^2)$) si

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

On a

$$\mathbb{E}(X) = \mu \ , \ \mathbb{V}(X) = \sigma^2$$

car $X \sim N(\mu, \sigma^2)$ si et seulement si

$$X = \sigma Z + \mu$$

où $Z \sim N(0, 1)$. Toute probabilité concernant X peut s'obtenir à partir de la fonction Φ :

$$P\{a < X \leq b\} = P\left\{\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right\} = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Exemple. Si $X \sim N(1, 4)$,

$$P\{0 < X \leq 2\} = 2\,\Phi(0,5) - 1 = 0,383$$

et

$$P\{X > 3\} = 1 - \Phi(1) = 0,159.$$

Le théorème central limite (dont le théorème de de Moivre-Laplace est un cas particulier) explique l'importance extrême de la loi normale (aussi appelée loi de Laplace-Gauss). C'est la distribution de base de la statistique, avec certaines distributions apparentées.

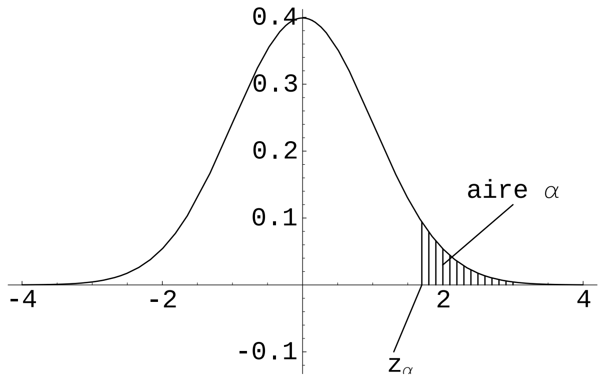

FIG. 7 – La loi normale standard

2.4.4 Loi du khi-deux

Si $Z_1, Z_2, \ldots, Z_n$ sont des normales standards indépendantes, la variable

$$\chi_n^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

suit une loi du khi-deux à n degrés de liberté. On a

$$\mathbb{E}(\chi_n^2) = n \quad \text{et} \quad \mathbb{V}(\chi_n^2) = 2n.$$

Les quantiles d'une loi χ_n^2 sont notés $\chi_{\alpha,n}^2$:

$$P\{\chi_n^2 > \chi_{\alpha,n}^2\} = \alpha.$$

Exemple. Lorsque $n = 1$, la fonction de densité de probabilité f peut facilement être obtenue en dérivant la fonction de répartition F. On a

$$F(x) = P\{Z^2 \leq x\} = 2P\{0 \leq Z \leq \sqrt{x}\} = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-t^2/2}\, dt$$

donc

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-x/2}.$$

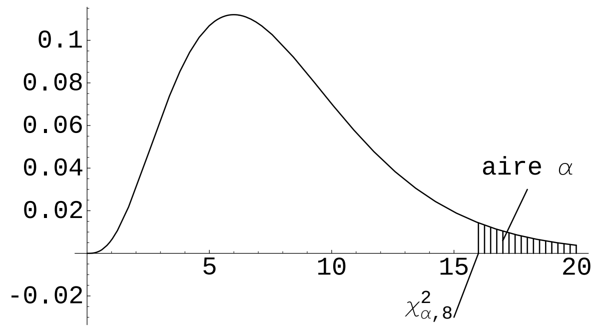

FIG. 8 – La loi du khi-deux, $n = 8$

2.4.5 Loi de Student

Si Z est une normale standard et χ^2_n est une khi-deux à n degrés de liberté et si ces variables sont indépendantes, la variable

$$T_n = \frac{Z}{\sqrt{\chi^2_n/n}}$$

suit une loi de Student à n degrés de liberté. On a

$$\mathbb{E}(T_n) = 0 \quad \text{et} \quad \mathbb{V}(T_n) = \frac{n}{n-2} \quad \text{si} \quad n > 2.$$

Les quantiles d'une loi T_n sont notés $t_{\alpha,n}$:

$$P\{T_n > t_{\alpha,n}\} = \alpha.$$

(Student (l'étudiant) est un pseudonyme pour W.S.Gosset qui a développé cette loi). Le graphe de la loi de Student a la même allure générale que celui de la loi normale seulement plus aplatie.

2.4.6 Loi de Fisher-Snedecor

Si les variables χ^2_n et χ^2_m sont indépendantes, la variable

$$F_{n,m} = \frac{\chi^2_n/n}{\chi^2_m/m}$$

suit une loi de Fisher-Snedecor à n et m degrés de liberté. On a

$$\mathbb{E}(F_{n,m}) = \frac{m}{m-2} \quad \text{et} \quad \mathbb{V}(F_{n,m}) = \frac{2m^2(n+m-2)}{(m-2)^2 n(m-4)}.$$

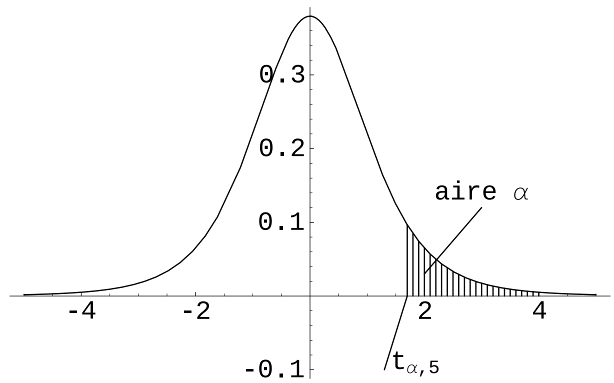

FIG. 9 – La loi de Student, $n = 5$

Les quantiles d'une loi $F_{n,m}$ sont notés $F_{\alpha,n,m}$:

$$P\{F_{n,m} > F_{\alpha,n,m}\} = \alpha.$$

Le graphe de la loi de Fisher-Snedecor a la même allure générale que celui de la loi du khi-deux.

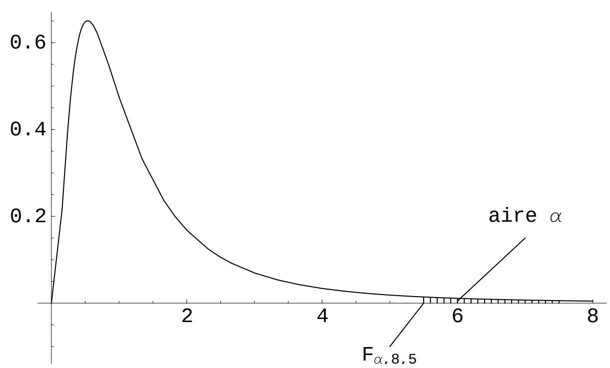

FIG. 10 – La loi de Fisher-Snedecor, $n = 8, m = 5$

3 Sommes de variables aléatoires

Dans ce chapitre, nous obtenons certaines propriétés des sommes de variables aléatoires indépendantes et nous étudions le comportement de ces sommes lorsque le nombre de termes tend vers l'infini.

3.1 Espérance d'une somme

Si X et Y sont deux variables discrètes conjointement distribuées, on a

$$\begin{aligned}\mathbb{E}(X+Y) &= \sum_k \sum_j (x_k + y_j) p(x_k, y_j) \\ &= \sum_k \sum_j x_k p(x_k, y_j) + \sum_k \sum_j y_j p(x_k, y_j) \\ &= \sum_k x_k p_X(x_k) + \sum_j y_j p_Y(y_j) = \mathbb{E}(X) + \mathbb{E}(Y).\end{aligned}$$

En général, si $X_1, X_2, \ldots, X_n$ sont des variables aléatoires conjointement distribuées quelconques, on a

$$\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n).$$

Si X et Y sont deux variables discrètes conjointement distribuées *indépendantes*, on a

$$\begin{aligned}\mathbb{E}(XY) &= \sum_k \sum_j x_k y_j p(x_k, y_j) = \sum_k \sum_j x_k y_j p_X(x_k) p_Y(y_j) \\ &= \sum_k x_k p_X(x_k) \sum_j y_j p_Y(y_j) = \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

Cette relation reste valable si les variables sont continues comme on l'a vu mais peut ne pas être vraie si les variables ne sont pas indépendantes.

Exemple. On effectue n tirages d'une urne contenant B boules dont R sont rouges. Soit X le nombre de boules rouges obtenues.

Si les tirages ont lieu avec remise, posons

$$X = \sum_{k=1}^{n} X_k$$

où

$$X_k = \begin{cases} 1 & \text{si on obtient une boule rouge au } k^{\text{ième}} \text{ tirage} \\ 0 & \text{sinon.} \end{cases}$$

Alors

$$\mathbb{E}(X_k) = \frac{R}{B}$$

et

$$\mathbb{E}(X) = \sum_{k=1}^{n} \mathbb{E}(X_k) = n\frac{R}{B}.$$

Si les tirages ont lieu sans remise, posons

$$X = \sum_{j=1}^{R} Y_j$$

où

$$Y_j = \begin{cases} 1 & \text{si la } j^{\text{ième}} \text{ boule rouge est tirée} \\ 0 & \text{sinon.} \end{cases}$$

Alors

$$\mathbb{E}(Y_j) = \frac{\binom{1}{1}\binom{B-1}{n-1}}{\binom{B}{n}} = \frac{n}{B}$$

et on a encore

$$\mathbb{E}(X) = \sum_{j=1}^{R} \mathbb{E}(Y_j) = n\frac{R}{B}.$$

3.2 Variance d'une somme

Si X et Y sont deux variables conjointement distribuées, on a

$$\begin{aligned} \mathbb{V}(X+Y) &= \mathbb{E}((X+Y)^2) - \mathbb{E}(X+Y)^2 \\ &= \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 + 2\left(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\right). \end{aligned}$$

La **covariance** de X et Y est

$$\mathrm{COV}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

et leur **coefficient de corrélation** est

$$\rho(X, Y) = \frac{\mathrm{COV}(X, Y)}{\sigma(X)\sigma(Y)}.$$

En vertu de l'inégalité de Cauchy-Schwarz, on a toujours que

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

avec égalité si et seulement si il existe une relation linéaire entre X et Y. En général, une corrélation positive entre X et Y signifie que ces variables ont tendance à varier dans le même sens, une corrélation négative, dans des sens opposés.

Exemple. Soient X et Y deux variables ne prenant chacune que les valeurs 0 ou 1. Alors

$$\begin{aligned} \mathrm{COV}(X, Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= P\{X = 1, Y = 1\} - P\{X = 1\}P\{Y = 1\} \\ &= (P\{X = 1/Y = 1\} - P\{X = 1\})\, P\{Y = 1\} > 0 \end{aligned}$$

si et seulement si $P\{X = 1/Y = 1\} > P\{X = 1\}$. On voit clairement ici qu'une covariance positive indique une tendance pour les deux variables à varier dans le même sens.

Si X et Y sont indépendantes, alors $\mathrm{COV}(X, Y) = 0$ mais la réciproque n'est pas vraie.

Exemple. Si (X, Y) est un vecteur uniforme dans le disque $x^2 + y^2 < 1$, on a vu que les variables X et Y ne sont pas indépendantes. Cependant,

$$\mathrm{COV}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$$

(les intégrales sont toutes nulle par symétrie).

Pour deux variables quelconques, on a donc

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\mathrm{COV}(X, Y)$$

et, si X et Y ne sont pas corrélées,

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y).$$

En général, si $X_1, X_2, \ldots, X_n$ sont des variables aléatoires conjointement distribuées, on a

$$\mathbb{V}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathbb{V}(X_k) + 2 \sum_{1 \leq i < j \leq n} \mathrm{COV}(X_i, X_j)$$

et si, de plus, ces variables sont deux à deux non corrélées,

$$\mathbb{V}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathbb{V}(X_k).$$

Exemple. On effectue n tirages d'une urne contenant B boules dont R sont rouges. Soit X le nombre de boules rouges obtenues.

Si les tirages ont lieu avec remise, posons

$$X = \sum_{k=1}^{n} X_k$$

où

$$X_k = \begin{cases} 1 & \text{si on obtient une boule rouge au } k^{\text{ième}} \text{ tirage} \\ 0 & \text{sinon.} \end{cases}$$

Les variables X_k sont indépendantes et

$$\mathbb{V}(X_k) = \frac{R}{B}\left(1 - \frac{R}{B}\right)$$

donc

$$\mathbb{V}(X) = \sum_{k=1}^{n} \mathbb{V}(X_k) = n\,\frac{R}{B}\left(1 - \frac{R}{B}\right).$$

Si les tirages ont lieu sans remise, posons

$$X = \sum_{j=1}^{R} Y_j$$

où

$$Y_j = \begin{cases} 1 & \text{si la } j^{\text{ième}} \text{ boule rouge est tirée} \\ 0 & \text{sinon.} \end{cases}$$

Les variables Y_j ne sont pas indépendantes. On a

$$\mathbb{V}(Y_j) = \frac{n}{B}\left(1 - \frac{n}{B}\right)$$

et

$$\begin{aligned}\mathrm{COV}(Y_i, Y_j) &= P\{Y_j = 1/Y_i = 1\}P\{Y_i = 1\} - P\{Y_i = 1\}P\{Y_j = 1\} \\ &= \frac{n-1}{B-1}\,\frac{n}{B} - \left(\frac{n}{B}\right)^2 = -\frac{n}{B}\,\frac{B-n}{B(B-1)}\end{aligned}$$

donc

$$\begin{aligned}\mathbb{V}(X) &= \sum_{j=1}^{R} \mathbb{V}(Y_j) + 2 \sum_{1 \le i < j \le R} \mathrm{COV}(Y_i, Y_j) \\ &= R\frac{n}{B}\left(1 - \frac{n}{B}\right) - 2\binom{R}{2}\frac{n}{B}\,\frac{B-n}{B(B-1)} \\ &= n\,\frac{R}{B}\left(1 - \frac{R}{B}\right)\frac{B-n}{B-1}.\end{aligned}$$

3.3 Quelques inégalités générales

L'inégalité de Markov.
Soient X une variable aléatoire positive et $a > 0$. Alors

$$P\{X \ge a\} \le \frac{\mathbb{E}(X)}{a}.$$

En effet, si, par exemple, X est discrète,

$$\mathbb{E}(X) = \sum_k x_k p(x_k) \ge \sum_{x_k \ge a} x_k p(x_k) \ge a \sum_{x_k \ge a} p(x_k) = aP\{X \ge a\}.$$

Si l'on connaît la distribution impliquée, on peut généralement faire beaucoup mieux que l'inégalité de Markov. Par exemple, $\chi^2_{0{,}05,\ 8} = 15{,}507$ donc

$$P\{\chi^2_8 \ge 15{,}507\} = 0{,}05$$

alors que l'inégalité de Markov ne donne que

$$P\{\chi^2_8 \ge 15{,}507\} \le \frac{8}{15{,}507} = 0{,}516.$$

L'inégalité de Tchebychev.
Soient X une variable aléatoire et $a > 0$. Alors

$$P\{|X - \mathbb{E}(X)| \ge a\} \le \frac{\mathbb{V}(X)}{a^2}.$$

En effet, en vertu de l'inégalité de Markov,

$$\begin{aligned} P\{|X-\mathbb{E}(X)| \geq a\} &= P\{(X-\mathbb{E}(X))^2 \geq a^2\} \\ &\leq \frac{\mathbb{E}((X-\mathbb{E}(X))^2)}{a^2} = \frac{\mathbb{V}(X)}{a^2}. \end{aligned}$$

L'inégalité de Tchebychev est souvent présentée sous la forme

$$P\{|X-\mathbb{E}(X)| \geq k\ \sigma(X)\} \leq \frac{1}{k^2}.$$

Pour la loi normale, les probabilités correspondant à $k = 1, 2, 3$ sont

$$2\Phi(1) - 1 = 0,683\ ,\ \ 2\Phi(2) - 1 = 0,955\ \text{ et }\ 2\Phi(3) - 1 = 0,997.$$

3.4 La loi des grands nombres

Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes et de même loi, admettant μ pour espérance mathématique commune. Alors, quel que soit $\epsilon > 0$, on a

$$\lim_{n\to+\infty} P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\} = 0.$$

Cette loi des grands nombres peut être déduite de l'inégalité de Tchebychev. On a

$$\mathbb{E}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \mu$$

et, désignant par σ^2 la variance commune des variables X_k,

$$\mathbb{V}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

en vertu de leur indépendance. Par suite,

$$P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2}$$

d'où le résultat.

La loi des grands nombres justifie d'attribuer pour probabilité à un évènement E la limite de la fréquence relative d'apparition de E en n épreuves indépendantes lorsque n tend vers l'infini : c'est-à-dire, désignant par $n(E)$ le nombre de fois où E s'est produit lors des n premières épreuves, de poser

$$P\{E\} = \lim_{n\to+\infty} \frac{n(E)}{n}.$$

En effet, en désignant par X_k la variable de Bernoulli qui vaut 1 si l'évènement E est observé lors de la $k^{\text{ième}}$ épreuve, on a, pour tout $\epsilon > 0$,

$$P\left\{\left|\frac{n(E)}{n} - P\{E\}\right| \geq \epsilon\right\} = P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - P\{E\}\right| \geq \epsilon\right\}$$
$$\leq \frac{P\{E\}(1-P\{E\})}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}.$$

En choisissant par exemple $\epsilon = 0,05$ et $n = 2000$, on voit que 19 fois sur 20 la marge d'erreur est d'au plus 5%.

3.5 La fonction génératrice des moments

Le $k^{\text{ième}}$ moment de la variable aléatoire X est

$$\mu_k = \mathbb{E}(X^k)$$

et la fonction

$$\varphi_X(t) = \mathbb{E}\left(e^{tX}\right) = \sum_{k=0}^{+\infty} \mu_k \frac{t^k}{k!}$$

s'appelle fonction génératrice des moments car :

$$\mu_k = \frac{d^k}{dt^k}\varphi_X(t)\Big|_{t=0}.$$

Ainsi

$$\mathbb{E}(X) = \varphi_X'(0) \ \text{ et } \ \mathbb{V}(X) = \varphi_X''(0) - \varphi_X'(0)^2.$$

Elle possède trois propriétés fondamentales :

1. $\varphi_{aX+b}(t) = \varphi_X(at)e^{bt}$;
2. si X et Y sont indépendantes, $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$;
3. φ_X détermine complètement la distribution de X.

Si $X \sim N(\mu, \sigma^2)$, on peut écrire que $X = \sigma Z + \mu$ où $Z \sim N(0,1)$. Comme

$$\varphi_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(x-t)^2/2}\, dx\; e^{t^2/2} = e^{t^2/2},$$

on a, en vertu de la première propriété,

$$\varphi_X(t) = e^{\mu t + (\sigma t)^2/2}.$$

En utilisant les deux autres propriétés de la fonction génératrice des moments, on en tire une propriété remarquable de la loi normale : *une somme de variables aléatoires normales indépendantes est encore une variable aléatoire normale.*

3.6 Le théorème central limite

Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes et de même loi (quelconque), admettant μ pour espérance mathématique et σ^2 pour variance communes. Alors, quel que soit x

$$\lim_{n\to+\infty} P\left\{\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt.$$

Autrement dit, pour n grand, la variable aléatoire

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

est pratiquement une variable normale standard. (Cette approximation est généralement bonne dès que $n \geq 30$.)

Ce théorème fondamental se démontre aisément à l'aide des fonctions génératrices des moments — la vraie difficulté consiste à démontrer la propriété 3. de cette dernière fonction.

Considérons les variables centrées réduites $Y_k = (X_k - \mu)/\sigma$ de telle sorte que

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{Y_1 + Y_2 + \cdots + Y_n}{\sqrt{n}}$$

et montrons que la fonction génératrice des moments de cette dernière variable tend vers la fonction génératrice des moments d'une variable normale standard. Si

$$\varphi(t) = 1 + \frac{1}{2}t^2 + \cdots$$

est la fonction génératrice des moments commune des variables Y_k, celle de $(Y_1 + Y_2 + \cdots + Y_n)/\sqrt{n}$ est

$$\varphi\left(\frac{t}{\sqrt{n}}\right)^n$$

et il faut vérifier que

$$\lim_{n\to+\infty} \varphi\left(\frac{t}{\sqrt{n}}\right)^n = e^{t^2/2}$$

c'est-à-dire que

$$\lim_{n\to+\infty} \frac{\ln \varphi\left(\frac{t}{\sqrt{n}}\right)}{1/n} = \frac{t^2}{2}.$$

On y arrive en posant $x = 1/\sqrt{n}$ et en appliquant deux fois la règle de l'Hospital à l'origine.

Exemple. Soient $X_1, X_2, \ldots, X_{100}$ des variables aléatoires uniformes sur $(0,1)$ indépendantes. Alors

$$\begin{aligned} &P\{X_1 + X_2 + \cdots + X_{100} < 45\} \\ &= P\left\{\frac{X_1 + X_2 + \cdots + X_{100} - 50}{10/\sqrt{12}} < \frac{45-50}{10/\sqrt{12}}\right\} \\ &\approx P\{Z < -1,732\} = 1 - \Phi(1,732) = 0,042. \end{aligned}$$

Exemple. Supposons que lorsqu'un individu monte dans l'ascenseur du pavillon AA, il choisisse l'un des étages $2, 3, 4, 5$ et 6 avec probabilités respectives $1/3, 1/4, 1/6, 1/6$ et $1/12$. Soit X l'énergie nécessaire pour le transporter, supposée proportionnelle au nombre d'étages montés. Alors

$$\mathbb{E}(X) = \left(\frac{1}{3} + \frac{2}{4} + \frac{3}{6} + \frac{4}{6} + \frac{5}{12}\right) = \frac{29}{12}$$

et

$$\mathbb{V}(X) = \left(\frac{1}{3} + \frac{4}{4} + \frac{9}{6} + \frac{16}{6} + \frac{25}{12}\right) - \left(\frac{29}{12}\right)^2 = \frac{1283}{144}$$

(en unités d'énergie appropriées). Si, une journée, 500 individus utilisent l'ascenseur, la probabilité que la quantité d'énergie consommée excède 1000 unités est

$$P\left\{\sum_{k=1}^{500} X_k \geq 1001\right\} = P\left\{\sum_{k=1}^{500} X_k > 1000,5\right\}$$

puisque la variable $\sum_{k=1}^{500} X_k$ est à valeurs entières. Cette *correction de continuité* améliore la précision de l'approximation. On a

$$P\left\{\sum_{k=1}^{500} X_k > 1000,5\right\} = P\left\{\frac{\sum_{k=1}^{500} X_k - 500\ 29/12}{\sqrt{500}\sqrt{1283/144}} > \frac{1000,5 - 500\ 29/12}{\sqrt{500}\sqrt{1283/144}}\right\},$$

ce qui est pratiquement égale à

$$P\left\{Z > \frac{1000,5 - 500\ 29/12}{\sqrt{500}\sqrt{1283/144}}\right\} = 1 - \Phi(3,114) = 0,999.$$

4 L'échantillon statistique

Supposons que l'on s'intéresse à une caractéristique numérique X des individus ω d'une population Ω. Cette caractéristique X est considérée comme une variable aléatoire de loi $\mathcal{L}$ inconnue. Si on prélève suivant les règles de l'art un échantillon de taille n de cette population, on obtiendra n variables aléatoires indépendantes et de même loi $\mathcal{L}$, $X_1, X_2, \ldots, X_n$. Une **statistique** est une variable aléatoire calculée à partir de cet échantillon. Les statistiques les plus usitées sont la moyenne $\overline{X}$ et la variance S^2 de l'échantillon, définies respectivement par

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

et par

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2.$$

En désignant par μ l'espérance mathématique et par σ^2 la variance de la loi $\mathcal{L}$, on a

$$\mathbb{E}(\overline{X}) = \mu \quad \text{et} \quad \mathbb{V}(\overline{X}) = \frac{\sigma^2}{n}.$$

De plus

$$\begin{aligned}
\mathbb{E}(S^2) &= \frac{1}{n-1}\sum_{k=1}^{n}\mathbb{E}\left(X_k^2 - 2X_k\overline{X} + \overline{X}^2\right) \\
&= \frac{1}{n-1}\sum_{k=1}^{n}\left(\mu^2 + \sigma^2 - \frac{2}{n}\sum_{j=1}^{n}\mathbb{E}(X_kX_j) + \mu^2 + \frac{\sigma^2}{n}\right) \\
&= \frac{1}{n-1}\sum_{k=1}^{n}\left(\mu^2 + \sigma^2 - \frac{2}{n}((n-1)\mu^2 + \mu^2 + \sigma^2) + \mu^2 + \frac{\sigma^2}{n}\right) = \sigma^2.
\end{aligned}$$

Quelle que soit la loi $\mathcal{L}$, lorsque n est grand, la distribution de

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

est pratiquement normale standard.

Dans le cas (usuel en statistique) où l'on suppose que $\mathcal{L}$ est une loi normale, pour tout n,

1. la variable
$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$
suit une loi normale standard ;
2. la variable
$$\chi^2_{n-1} = (n-1)\,\frac{S^2}{\sigma^2}$$
suit une loi du khi-deux à $n-1$ degrés de liberté ;
3. les variables $\overline{X}$ et S^2 sont indépendantes ;
4. la variable
$$T_{n-1} = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$
suit une loi de Student à $n-1$ degrés de liberté.

Ces faits deviennent plausibles si l'on remarque les relations suivantes et si l'on se rappelle la définition d'une variable khi-deux à n degrés de liberté et d'une variable de Student à n degrés de liberté. En vertu de l'identité
$$\sum_{k=1}^{n} (y_k - \overline{y})^2 = \sum_{k=1}^{n} y_k^2 - n\overline{y}^2,$$
on a en effet
$$(n-1)\,\frac{S^2}{\sigma^2} = \sum_{k=1}^{n} \left(\frac{X_k - \mu}{\sigma}\right)^2 - \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)^2$$
et
$$\frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{Z}{\sqrt{\chi^2_{n-1}/(n-1)}}.$$

Remarque. On peut toujours supposer que $\Omega = \mathbb{R}$ et que $X(\omega) = \omega$. C'est la fonction
$$F(x) = P\{X \le x\}$$
qui définit la distribution et qu'il faut étudier.

5 Estimation des paramètres d'une distribution

Pour estimer un paramètre θ d'une distribution, on calcule d'abord une valeur à partir de l'échantillon obtenu et on construit ensuite autour de cette valeur un intervalle dans lequel on a une certaine confiance de le trouver.

5.1 Estimation ponctuelle

Un bon estimateur E pour le paramètre θ de la distribution de la variable X est une statistique telle que $\mathbb{E}(E - \theta) = 0$ et que $\mathbb{V}(E)$ est la plus petite possible. Si $f(x|\theta)$ est la fonction de densité de probabilité de X, la fonction de densité conjointe de l'échantillon $X_1, X_2, \ldots, X_n$ est

$$f(x_1, x_2, \ldots, x_n|\theta) = f(x_1|\theta)f(x_2|\theta) \cdots f(x_n|\theta).$$

Lorsque les valeurs $x_1, x_2, \ldots, x_n$ des variables $X_1, X_2, \ldots, X_n$ ont été observées, l'expression

$$f(x_1, x_2, \ldots, x_n|\theta)$$

est une fonction de θ que l'on appelle la **fonction de vraisemblance** et la méthode de l'**estimateur à vraisemblance maximale** consiste à choisir la valeur $\hat{\theta}$ du paramètre θ qui maximise la « probabilité » de ce qui a été observé.

5.1.1 Moyenne d'une variable de Bernoulli

La fonction de vraisemblance est

$$p^{x_1+x_2+\cdots+x_n}\ (1-p)^{n-x_1-x_2-\cdots-x_n}$$

et son logarithme est

$$(x_1 + x_2 + \cdots + x_n)\ \log p + (n - x_1 - x_2 - \cdots - x_n)\ \log(1-p).$$

Cette expression est maximale lorsque $p = \sum_{k=1}^{n} x_k/n$ comme on le voit en annulant sa dérivée par rapport à p. Autrement dit

$$\hat{p} = \overline{X}.$$

Puisque

$$\mathbb{E}(\hat{p}) = \mu,$$

$\hat{p}$ est un **estimateur non biaisé** pour μ.

5.1.2 Moyenne et variance d'une variable normale

La fonction de vraisemblance est

$$(2\pi)^{-n/2}\sigma^{-n}\exp -\left\{\frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}\right\}$$

et son logarithme est

$$-\frac{n}{2}\log 2\pi-\frac{n}{2}\log\sigma^2-\frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}.$$

En annulant les dérivées partielles par rapport à μ et par rapport à σ^2 de cette expression, on obtient les estimateurs à vraisemblance maximale

$$\hat{\mu}=\overline{X}$$

et

$$\widehat{\sigma^2}=\frac{n-1}{n}\,S^2.$$

On remarque que l'estimateur pour la variance est biaisé.

5.2 Intervalles de confiance

5.2.1 Moyenne et variance d'une variable normale

a) Lorsque σ^2 est connue, la variable

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

suit une loi normale standard. Par conséquent, on a

$$P\left\{\overline{X}-z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right\}=1-\alpha,$$

$$P\left\{\overline{X}-z_{\alpha}\,\frac{\sigma}{\sqrt{n}}<\mu\right\}=1-\alpha$$

et

$$P\left\{\mu<\overline{X}+z_{\alpha}\,\frac{\sigma}{\sqrt{n}}\right\}=1-\alpha.$$

Lorsque les valeurs $x_1,x_2,\ldots,x_n$ des variables $X_1,X_2,\ldots,X_n$ ont été observées et que leur moyenne $\overline{x}$ a été calculée, on exprime les probabilités précédentes en disant que

$$I_{1-\alpha}=\left(\overline{x}-z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x}+z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right)$$

est un **intervalle de confiance de seuil** $1-\alpha$ pour μ et que

$$D_{1-\alpha} = \left(\overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}}\ ,\ +\infty\right)$$

et

$$G_{1-\alpha} = \left(-\infty\ ,\ \overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

sont des intervalles de confiance de seuil $1-\alpha$ à droite et à gauche respectivement pour μ.

Exemple. Les saucisses.

Lorsqu'un certain appareil mesure le nombre de calories contenues dans une saucisse à hot dog, ce nombre peut être considéré comme une variable aléatoire normalement distribuée, de moyenne inconnue μ (la vraie teneur en calories de la saucisse) et de variance $\sigma^2 = 9$ (caractéristique de l'appareil de mesure). On a mesuré à $n = 16$ reprises une saucisse (supposée homogène et coupée en seize morceaux) et trouvé les valeurs suivantes pour son contenu en calories :

$$153, 156, 154, 147, 148, 152, 152, 153$$
$$140, 150, 147, 144, 150, 149, 142, 151.$$

L'estimateur à vraisemblance maximale pour μ est

$$\hat{\mu} = \overline{x} = 149,25$$

et un intervalle de confiance de seuil 95% pour μ est

$$I_{0,95} = \left(\overline{x} - z_{0,025}\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x} + z_{0,025}\frac{\sigma}{\sqrt{n}}\right) = (147,78\ ,\ 150,72).$$

Un intervalle de confiance de seuil 90% pour μ serait

$$I_{0,90} = \left(\overline{x} - z_{0,05}\frac{\sigma}{\sqrt{n}}\ ,\ \overline{x} + z_{0,05}\frac{\sigma}{\sqrt{n}}\right) = (148,016\ ,\ 150,484).$$

(http ://www.calorieking.com — les données sont simulées).

b) Lorsque σ^2 est inconnue, la variable

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}$$

suit une loi de Student à $n-1$ degrés de liberté. On en déduit comme précédemment que, lorsque les valeurs $x_1, x_2, \ldots, x_n$ des variables $X_1, X_2, \ldots, X_n$ ont été observées et que leur moyenne $\overline{x}$ et leur variance s^2 ont été calculées,

$$I_{1-\alpha} = \left(\overline{x} - t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} \ , \ \overline{x} + t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} \right)$$

est un intervalle de confiance de seuil $1-\alpha$ pour μ et que

$$D_{1-\alpha} = \left(\overline{x} - t_{\alpha,n-1} \frac{s}{\sqrt{n}} \ , \ +\infty \right)$$

et

$$G_{1-\alpha} = \left(-\infty \ , \ \overline{x} + t_{\alpha,n-1} \frac{s}{\sqrt{n}} \right)$$

sont des intervalles de confiance de seuil $1-\alpha$ à droite et à gauche respectivement pour μ.

Exemple. Les saucisses (suite).

Le nombre de calories contenues dans une saucisse à hot dog peut être considéré comme une variable aléatoire distribuée normalement, de moyenne μ et de variance σ^2 inconnues. Un échantillon de $n = 16$ saucisses a produit les valeurs suivantes :

$$153, 156, 154, 147, 148, 152, 152, 153$$
$$140, 150, 147, 144, 150, 149, 142, 151.$$

L'estimateur à vraisemblance maximale pour μ est

$$\hat{\mu} = \overline{x} = 149,25$$

et un estimateur pour σ^2 est

$$s^2 = 19,5333.$$

Un intervalle de confiance de seuil 95% pour μ est

$$\begin{aligned} I_{0,95} &= \left(\overline{x} - t_{0,025,\,n-1} \frac{s}{\sqrt{n}} \ , \ \overline{x} + t_{0,025,\,n-1} \frac{s}{\sqrt{n}} \right) \\ &= (146,937 \ , \ 151,563) \end{aligned}$$

et un intervalle de confiance de seuil 90% pour μ serait

$$\begin{aligned} I_{0,90} &= \left(\overline{x} - t_{0,05,\,n-1}\,\frac{s}{\sqrt{n}}\ ,\ \overline{x} + t_{0,05,\,n-1}\,\frac{s}{\sqrt{n}}\right) \\ &= (147,339\ ,\ 151,161). \end{aligned}$$

c) La variable

$$(n-1)\frac{S^2}{\sigma^2}$$

suit une loi du khi-deux à $n-1$ degrés de liberté. Donc

$$1-\alpha = P\left\{\sigma^2 < (n-1)\frac{S^2}{\chi^2_{1-\alpha,n-1}}\right\} = P\left\{(n-1)\frac{S^2}{\chi^2_{\alpha,n-1}} < \sigma^2\right\}$$

et

$$1-\alpha = P\left\{(n-1)\frac{S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < (n-1)\frac{S^2}{\chi^2_{1-\alpha/2,n-1}}\right\}.$$

Lorsque les valeurs $x_1, x_2, \ldots, x_n$ des variables $X_1, X_2, \ldots, X_n$ ont été observées et que leur variance s^2 a été calculée, on exprime les probabilités précédentes en disant que

$$I_{1-\alpha} = \left((n-1)\frac{s^2}{\chi^2_{\alpha/2,n-1}}\ ,\ (n-1)\frac{s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$$

est un intervalle de confiance de seuil $1-\alpha$ pour σ^2 et que

$$D_{1-\alpha} = \left((n-1)\frac{s^2}{\chi^2_{\alpha,n-1}}\ ,\ +\infty\right)$$

et

$$G_{1-\alpha} = \left(0\ ,\ (n-1)\frac{S^2}{\chi^2_{1-\alpha,n-1}}\right)$$

sont des intervalles de confiance de seuil $1-\alpha$ à droite et à gauche respectivement pour σ^2.

Exemple. Les saucisses (fin).

Un intervalle de confiance de seuil 95% pour la variance inconnue σ^2 de la teneur en calories des saucisses à hot dog calculé à partir de l'échantillon de taille $n = 16$ précédent est

$$I_{0,95} = \left((n-1)\frac{s^2}{\chi^2_{0,025,n-1}} \ , \ (n-1)\frac{s^2}{\chi^2_{0,975,n-1}} \right)$$
$$= (8,91869 \ , \ 32,8972).$$

5.2.2 Différence des moyennes de deux variables normales

On considère deux échantillons indépendants $X_1, X_2, \ldots, X_n$ ainsi que $Y_1, Y_2, \ldots, Y_m$ associés à deux variables X et Y de lois normales $N(\mu_x, \sigma_x^2)$ et $N(\mu_y, \sigma_y^2)$ respectivement et l'on cherche à estimer la différence $\mu_x - \mu_y$.

a) Lorsque σ_x^2 et σ_y^2 sont connues, la variable

$$\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

suit une loi normale standard. Par conséquent, les intervalles de confiance de seuil $1-\alpha$ pour la différence $\mu_x - \mu_y$ sont, les variables $X_1, X_2, \ldots, X_n$ et $Y_1, Y_2, \ldots, Y_m$ ayant été observées et les moyennes $\overline{x}$ et $\overline{y}$ ayant été calculées,

$$I_{1-\alpha} = \left(\overline{x} - \overline{y} - z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \ , \ \overline{x} - \overline{y} + z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \right),$$

$$D_{1-\alpha} = \left(\overline{x} - \overline{y} - z_{\alpha}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \ , \ +\infty \right)$$

et

$$G_{1-\alpha} = \left(-\infty \ , \ \overline{x} - \overline{y} + z_{\alpha}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \right).$$

Exemple. Les beignes.

Le pourcentage en sucre des beignes produits à la succursale A d'une chaîne de franchises est une variable aléatoire X distribuée normalement avec variance $\sigma_x^2 = 18$ et celui Y de la succursale B est une variable de même type avec variance $\sigma_y^2 = 16$. On peut calculer à partir des échantillons suivants

X	27	30	26	27	22	22	33	21	27	32		
Y	20	17	18	17	18	17	19	26	24	21	22	22

un intervalle de confiance à gauche de seuil 90% pour $\mu_x - \mu_y$. C'est

$$G_{0,90} = \left(-\infty\ , \overline{x} - \overline{y} + z_{0,10}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}\right)$$

$$= \left(-\infty\ ,\ 26,7 - 20,0833 + 1,28155\sqrt{\frac{18}{10} + \frac{16}{12}}\right) = (-\infty\ , 8,8852).$$

(http ://www.nutritiondata.com — les données sont simulées).

b) Lorsque σ_x^2 et σ_y^2 sont inconnues mais égales (à σ^2), la variable

$$(n + m - 2)\,\frac{S_p^2}{\sigma^2}$$

où

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n + m - 2}$$

suit une loi du khi-deux à $n + m - 2$ degrés de liberté, donc la variable

$$\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

suit une loi de Student à $n + m - 2$ degrés de liberté. Par conséquent, les intervalles de confiance de seuil $1 - \alpha$ pour la différence $\mu_x - \mu_y$ sont, les variables $X_1, X_2, \ldots, X_n$ et $Y_1, Y_2, \ldots, Y_m$ ayant été observées et les moyennes $\overline{x}$ et $\overline{y}$ ainsi que les variances s_x^2 et s_y^2 ayant été calculées,

$$I_{1-\alpha} = \left(\overline{x} - \overline{y} - t_{\alpha/2,n+m-2}\, s_p\sqrt{\frac{1}{n} + \frac{1}{m}}\ ,\ \overline{x} - \overline{y} + t_{\alpha/2,n+m-2}\, s_p\sqrt{\frac{1}{n} + \frac{1}{m}}\right),$$

$$D_{1-\alpha} = \left(\overline{x} - \overline{y} - t_{\alpha,n+m-2}\, s_p\sqrt{\frac{1}{n} + \frac{1}{m}}\ ,\ +\infty\right)$$

et

$$G_{1-\alpha} = \left(-\infty\ ,\ \overline{x} - \overline{y} + t_{\alpha,n+m-2}\, s_p\sqrt{\frac{1}{n} + \frac{1}{m}}\right).$$

Exemple. Les beignes (fin).

Le pourcentage en sucre des beignes produits dans une franchise d'une chaîne est une variable aléatoire normale. Deux échantillons indépendants (pris deux jours différents) ont donné les résultats suivants :

X	27	30	26	27	22	22	33	21	27	32		
Y	20	17	18	17	18	17	19	26	24	21	22	22

En considérant que ces échantillons sont issues de variables aléatoires normales X et Y de moyennes inconnues mais de même variance, on peut en déduire un intervalle de confiance à gauche de seuil 90% pour la différence $\mu_x - \mu_y$ de ces moyennes. C'est

$$G_{0,90} = \left(-\infty\ ,\ \overline{x} - \overline{y} + t_{0,05,\,n+m-2}\, s_p \sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$
$$= \left(-\infty\ ,\ 26,7 - 20,0833 + 1,32534\ 3,5568 \sqrt{\frac{1}{10} + \frac{1}{12}}\right)$$
$$= (-\infty\ ,\ 4,2684).$$

puisque

$$s_p = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}} = \sqrt{\frac{9\ 17,3444 + 11\ 8,81061}{20}} = 3,5568.$$

5.2.3 Moyenne d'une variable de Bernoulli

Si X suit une loi de Bernoulli de paramètre p, la variable

$$\frac{\overline{X} - p}{\sqrt{p(1-p)}/\sqrt{n}}$$

est approximativement normale (pour n raisonnablement grand) de telle sorte que

$$1 - \alpha \approx P\left\{p < \overline{X} + z_\alpha \sqrt{\frac{p(1-p)}{n}}\right\},$$
$$1 - \alpha \approx P\left\{\overline{X} - z_\alpha \sqrt{\frac{p(1-p)}{n}} < p\right\}$$

et

$$1-\alpha \approx P\left\{\overline{X} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < p < \overline{X} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right\}.$$

En remplaçant p par son estimateur à vraisemblance maximum certaines de ses apparitions dans ces relations, on obtient

$$1-\alpha \approx P\left\{p < \overline{X} + z_\alpha\sqrt{\overline{X}(1-\overline{X})\frac{1}{n}}\right\},$$

$$1-\alpha \approx P\left\{\overline{X} - z_\alpha\sqrt{\overline{X}(1-\overline{X})\frac{1)}{n}} < p\right\}$$

et

$$1-\alpha \approx P\left\{\overline{X} - z_{\alpha/2}\sqrt{\overline{X}(1-\overline{X})\frac{1}{n}} < p < \overline{X} + z_{\alpha/2}\sqrt{\overline{X}(1-\overline{X})\frac{1}{n}}\right\}.$$

Lorsque les valeurs $x_1, x_2, \ldots, x_n$ des variables $X_1, X_2, \ldots, X_n$ ont été observées et que leur moyenne $\overline{x}$ a été calculée, on exprime les probabilités précédentes en disant que les intervalles

$$G_{1-\alpha} = \left(-\infty\ ,\ \overline{x} + z_\alpha\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\right),$$

$$D_{1-\alpha} = \left(\overline{x} - z_\alpha\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\ ,\ +\infty\right)$$

et

$$I_{1-\alpha} = \left(\overline{x} - z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\ ,\ \overline{x} + z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\right)$$

sont des intervalles de confiance approximatifs de seuil $1-\alpha$ pour p. La longueur l de ce dernier intervalle étant

$$l = 2z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}} \leq \frac{z_{\alpha/2}}{\sqrt{n}},$$

on obtiendra certainement un intervalle de longueur l plus petite qu'une longueur prescrite b si l'on prend

$$n \geq \left(\frac{z_{\alpha/2}}{b}\right)^2.$$

De façon semblable, les points

$$\overline{x} - \overline{y} \pm z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n} + \frac{\overline{y}(1-\overline{y})}{n}}$$

sont les extrémités d'un intervalle de confiance approximatif de seuil $1 - \alpha$ pour la différence des paramètres $p_x - p_y$ de deux variables de Bernoulli.

Exemple. Si 450 des 800 personnes interwievées ont exprimé leur soutien au candidat A, un intervalle de confiance approximatif de seuil 95% pour la proportion p des électeurs qui voteront pour A est

$$\begin{aligned} I_{0,95} &= \left(\overline{x} - z_{0,025}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\ ,\ \overline{x} + z_{0,025}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}\right) \\ &= (0,533651\ ,\ 0,591349) \end{aligned}$$

(un appui de 56% avec une marge d'erreur de 3% dix-neuf fois sur vingt). Si l'on avait observé la même proportion d'appuis avec un échantillon de 1600 personnes, cet intervalle serait réduit à

$$I_{0,95} = (0,542101\ ,\ 0,582899)$$

et la marge d'erreur ne serait plus que de 2%.

6 Tests d'hypothèses sur les paramètres d'une distribution

Un test d'hypothèse au seuil de signification α d'une hypothèse H_0 concernant un paramètre θ d'une distribution doit être tel que

$$P\{\text{rejeter } H_0 \text{ lorsqu'elle est vraie}\} \leq \alpha.$$

Pour le faire, il faut déterminer la statistique STT et la région critique C_α appropriées en supposant que H_0 est vraie, calculer la valeur stt de la statistique pour l'échantillon et rejeter H_0 si $stt \in C_\alpha$.

L'hypothèse H_0 est celle du statu quo et elle n'est rejetée au profit de l'hypothèses alternative H_1 que si les données obtenues sont hautement improbables sous H_0, autrement elle demeure acceptée (ce qui ne signifie *pas* prouvée).

6.1 Moyenne d'une variable normale

6.1.1 Variance connue

a) Pour tester l'hypothèse

$$H_0 : \mu = \mu_0,$$

contre l'hypothèse alternative

$$H_1 : \mu \neq \mu_0,$$

on utilise la statistique

$$STT = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Lorsque les valeurs $x_1, x_2, \ldots, x_n$ des variables $X_1, X_2, \ldots, X_n$ ont été observées et que

$$stt = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

a été calculée, on rejette H_0 au seuil de signification α si

$$|stt| > z_{\alpha/2}.$$

Alternativement, on peut d'abord calculer la **p-valeur** du test,

$$\text{p-valeur} = P\{|Z| > |stt|\},$$

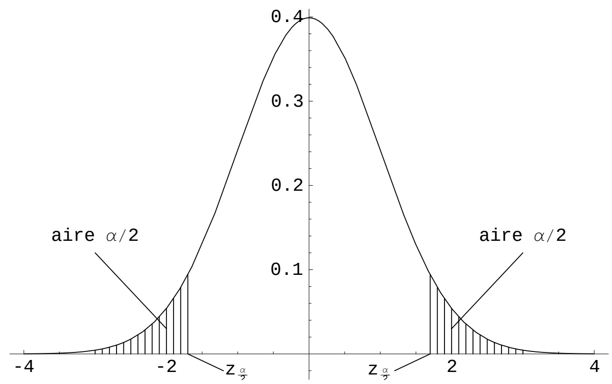

FIG. 11 – Zone de rejet pour $\mu = \mu_0$ contre $\mu \neq \mu_0$

et rejeter H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

b) Pour tester l'hypothèse

$$H_0 : \mu = \mu_0$$

contre l'alternative

$$H_1 : \mu > \mu_0,$$

on utilise la même statistique et on rejette si

$$stt > z_\alpha.$$

La p-valeur du test est

$$P\{Z > stt\}.$$

c) Pour tester l'hypothèse

$$H_0 : \mu = \mu_0$$

contre l'alternative

$$H_1 : \mu < \mu_0,$$

on utilise la même statistique et on rejette si

$$stt < -z_\alpha.$$

La p-valeur du test est

$$P\{Z < stt\}.$$

Exemple. La productivité.

La productivité en 2004 d'une usine d'un certain groupe industriel est considérée comme une variable aléatoire distribuée normalement avec moyenne 4600 et écart-type 500. En 2005, on a observé les productivités suivantes dans les neuf usines du groupe :

$$5452, 5355, 5844, 4245, 6647, 4545, 5450, 5548, 5658$$

Pour tester l'hypothèse $H_0 : \mu = 4600$ contre l'hypothèse alternative $H_1 : \mu > 4600$ en supposant que la variance de la distribution est restée la même, on utilise la statistique

$$STT = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

dont la valeur est ici

$$stt = \frac{5416 - 4600}{500/\sqrt{9}} = 4,896.$$

La p-valeur du test est

$$P\{Z > 4,896\} = 4.89035 \times 10^{-7}$$

et il faut rejeter H_0. (Données simulées).

6.1.2 Variance inconnue

a) Pour tester l'hypothèse

$$H_0 : \mu = \mu_0,$$

contre l'hypothèse alternative

$$H_1 : \mu \neq \mu_0,$$

on utilise la statistique

$$STT = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}.$$

Lorsque les valeurs $x_1, x_2, \ldots, x_n$ des variables $X_1, X_2, \ldots, X_n$ ont été observées et que

$$stt = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

a été calculée, on rejette H_0 si

$$|stt| > t_{\alpha/2,n-1}.$$

Alternativement, on peut d'abord calculer la p-valeur du test,

$$\text{p-valeur} = P\{|T_{n-1}| > |stt|\},$$

et rejeter H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

b) Pour tester l'hypothèse

$$H_0 : \mu = \mu_0$$

contre l'alternative

$$H_1 : \mu > \mu_0,$$

on utilise la même statistique et on rejette si

$$stt > t_{\alpha,n-1}.$$

La p-valeur du test est

$$P\{T_{n-1} > stt\}.$$

c) Pour tester l'hypothèse

$$H_0 : \mu = \mu_0$$

contre l'alternative

$$H_1 : \mu < \mu_0,$$

on utilise la même statistique et on rejette si

$$stt < -t_{\alpha,n-1}.$$

La p-valeur du test est

$$P\{T_{n-1} < stt\}.$$

Exemple. La productivité (fin).

En ne faisant plus l'hypothèses que la variance est restée la même, pour tester $H_0 : \mu = 4600$ contre $H_1 : \mu > 4600$, on utilise la statistique

$$STT = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

dont la valeur est ici

$$stt = \frac{5416 - 4600}{699,159/\sqrt{9}} = 3,50135.$$

La p-valeur du test est

$$P\{T_8 > 3,50135\} = 0,00403156$$

et il faut encore rejeter H_0.

6.2 Égalité des moyennes de variables normales

6.2.1 Variances connues

Pour tester l'hypothèse

$$H_0 : \mu_x = \mu_y$$

contre l'alternative

$$H_1 : \mu_x \neq \mu_y,$$

on utilise la statistique

$$STT = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}.$$

Lorsque les valeurs $x_1, x_2, \ldots, x_n$ et $y_1, y_2, \ldots, y_m$ des variables $X_1, X_2, \ldots, X_n$ et $Y_1, Y_2, \ldots, Y_m$ ont été observées et que

$$stt = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

a été calculée, on rejette H_0 si

$$|stt| > z_{\alpha/2}.$$

Alternativement, on peut d'abord calculer la p-valeur du test,

$$\text{p-valeur} = P\{|Z| > |stt|\},$$

et rejeter H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

6.2.2 Variances inconnues égales

Pour tester l'hypothèse

$$H_0 : \mu_x = \mu_y$$

contre l'alternative

$$H_1 : \mu_x \neq \mu_y,$$

on utilise la statistique

$$STT = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

Lorsque les valeurs $x_1, x_2, \ldots, x_n$ et $y_1, y_2, \ldots, y_m$ des variables $X_1, X_2, \ldots, X_n$ et $Y_1, Y_2, \ldots, Y_m$ ont été observées et que

$$stt = \frac{\overline{x} - \overline{y}}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

a été calculée, on rejette H_0 si

$$|stt| > t_{\alpha/2,n+m-2}.$$

Alternativement, on peut d'abord calculer la p-valeur du test,

$$\text{p-valeur} = P\{|T_{n+m-2}| > |stt|\},$$

et rejeter H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

6.2.3 Variances inconnues, grands échantillons

Pour tester l'hypothèse

$$H_0 : \mu_x = \mu_y$$

contre l'alternative

$$H_1 : \mu_x \neq \mu_y,$$

on utilise la statistique

$$STT = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$$

qui est, sous H_0 et pour de grandes valeurs de n et m, pratiquement distribuée normalement. Lorsque les valeurs $x_1, x_2, \ldots, x_n$ et $y_1, y_2, \ldots, y_m$ des variables $X_1, X_2, \ldots, X_n$ et $Y_1, Y_2, \ldots, Y_m$ ont été observées et que

$$stt = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

a été calculée, on rejette H_0 si

$$|stt| > z_{\alpha/2}.$$

Alternativement, on peut d'abord calculer la p-valeur du test,

$$\text{p-valeur} = P\{|Z| > |stt|\},$$

et rejeter H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

Exemple. La production.

Pour vérifier si une certaine caractéristique est la même sur les pièces produites par deux machines différentes, on prélève deux échantillons de taille $n = 200$ et $m = 300$ respectivement. Pour le premier échantillon, on obtient $\overline{x} = 10,5$ et $s_x^2 = 4,5$ et pour le second, $\overline{y} = 9,9$ et $s_y^2 = 6,1$ pour la mesure de cette caractéristique. Utilisant la statistique

$$\frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$$

pour tester $H_0 : \mu_x = \mu_y$ contre $H_1 : \mu_x \neq \mu_y$, on obtient ici

$$stt = 1,93272$$

et la p-valeur

$$\text{p-valeur} = P\{|Z| > |stt|\} = 0,0532707.$$

On doit donc accepter H_0 au seuil $0,05$ et rejeter H_0 au seuil $0,10$.

6.2.4 Variables jumelées

Pour tester l'hypothèse

$$H_0 : \mu_x = \mu_y$$

contre l'alternative

$$H_1 : \mu_x \neq \mu_y,$$

lorsque les variables X_k et les variables Y_k sont jumelées, on considère les variables $W_k = X_k - Y_k$ et on teste l'hypothèse

$$H_0' : \mu_w = 0$$

contre

$$H_1' : \mu_w \neq 0$$

en considérant la statistique

$$STT = \frac{\overline{W}}{S_w/\sqrt{n}}.$$

Lorsque les valeurs $w_1, w_2, \ldots, w_n$ des variables $W_1, W_2, \ldots, W_n$ ont été observées et que

$$stt = \frac{\overline{w}}{s_w/\sqrt{n}}$$

a été calculée, on rejette H_0 si

$$|stt| > t_{\alpha/2,n-1}.$$

Alternativement, on peut d'abord calculer la p-valeur du test,

$$\text{p-valeur} = P\{|T_{n-1}| > |stt|\},$$

et rejeter H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

Exemple. La production (suite).

Pour vérifier si une nouvelle technique de production réduit la valeur d'une certaine caractéristique indésirable du produit, on l'essaie sur dix machines. On obtient les échantillons suivants où X_i désigne la valeur de la caractéristique obtenue avec la vieille technique sur la machine i et Y_i la valeur avec la nouvelle.

X	10,6	9,8	11,6	11,0	13,0	10,0	9,8	9,8	10,6	10,0
Y	9,4	10,8	11,0	9,2	8,2	10,6	10,2	10,0	10,6	10,2

Pour tester l'hypothèse $H_0 : \mu_x = \mu_y$ contre l'hypothèse $H_1 : \mu_x > \mu_y$, on considère les variables $W_k = X_k - Y_k$ et la statistique

$$STT = \frac{\overline{W}}{S_w/\sqrt{n}}.$$

On obtient ici

$$stt = 1,1146$$

et la p-valeur

$$\text{p-valeur} = P\{T_9 > stt\} = 0,14695.$$

Il faut donc accepter H_0. (Données simulées).

6.3 Variance d'une variable normale

Pour tester l'hypothèse

$$H_0 : \sigma = \sigma_0,$$

contre l'hypothèse alternative

$$H_1 : \sigma \neq \sigma_0,$$

on utilise la statistique

$$STT = \frac{(n-1)S^2}{\sigma_0^2}.$$

Lorsque les valeurs $x_1, x_2, \ldots, x_n$ des variables $X_1, X_2, \ldots, X_n$ ont été observées et que

$$stt = \frac{(n-1)s^2}{\sigma_0^2}$$

a été calculée, on rejette H_0 si

$$stt \notin \left(\chi^2_{1-\alpha/2,n-1}, \chi^2_{\alpha/2,n-1}\right).$$

Alternativement, on peut d'abord calculer la p-valeur du test,

$$\text{p-valeur} = 2\min\left(P\{\chi^2_{n-1} < stt\}, 1 - P\{\chi^2_{n-1} < stt\}\right)$$

et rejeter H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

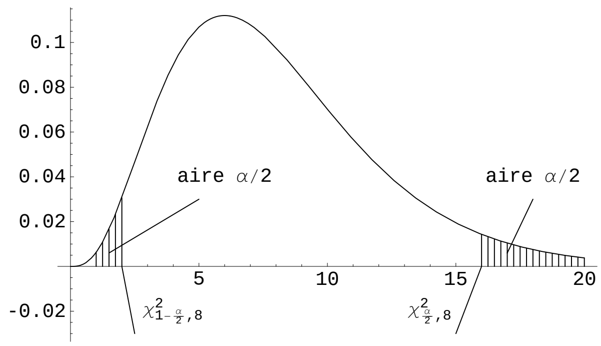

FIG. 12 – Zone de rejet pour $\sigma = \sigma_0$
contre $\sigma \neq \sigma_0$

6.4 Égalité des variances de variables normales

Pour tester l'hypothèse

$$H_0 : \sigma_x = \sigma_y,$$

contre l'hypothèse alternative

$$H_1 : \sigma_x \neq \sigma_y,$$

on utilise la statistique

$$STT = \frac{S_x^2}{S_y^2}.$$

Lorsque les valeurs $x_1, x_2, \ldots, x_n$ et $y_1, y_2, \ldots, y_m$ des variables $X_1, X_2, \ldots, X_n$ et $Y_1, Y_2, \ldots, Y_m$ ont été observées et que

$$stt = \frac{s_x^2}{s_y^2}$$

a été calculée, on rejette H_0 si

$$stt \notin \left(F_{1-\alpha/2,n-1,m-1}, F_{\alpha/2,n-1,m-1}\right).$$

Alternativement, on peut d'abord calculer la p-valeur du test,

$$\text{p-valeur} = 2\min\left(P\{F_{n-1,m-1} < stt\}, 1 - P\{F_{n-1,m-1} < stt\}\right)$$

et rejeter H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

Exemple. La production (fin).

On teste aussi l'hypothèses $H_0 : \sigma_x = \sigma_y$ versus l'alternative $H_1 : \sigma_x > \sigma_y$ pour la nouvelle technique de production à l'aide de la statistique

$$STT = \frac{S_x^2}{S_y^2}.$$

On obtient ici

$$stt = 1,4314$$

et la p-valeur

$$\text{p-valeur} = P\{F_{9,9} > 1,4314\} = 0,300863.$$

Il faut donc accepter H_0.

6.5 Moyenne d'une variable de Bernoulli

Pour tester l'hypothèse

$$H_0 : p = p_0$$

contre l'alternative

$$H_1 : p \neq p_0,$$

on utilise la statistique

$$STT = X_1 + X_2 + \cdots + X_n.$$

Lorsque les valeurs $x_1, x_2, \ldots, x_n$ des variables $X_1, X_2, \ldots, X_n$ ont été observées et que

$$stt = x_1 + x_2 + \cdots + x_n$$

a été calculée, on calcule la p-valeur du test,

$$\text{p-valeur} = 2 \min \left(\sum_{k=0}^{stt} \binom{n}{k} p_0^k (1-p_0)^{n-k}, \sum_{k=stt}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k} \right)$$

et on rejette H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

7 Tests de validité

On peut aussi faire des hypothèses sur le type d'une distribution ou sur l'indépendance de deux variables aléatoires conjointement distribuées.

7.1 Tests d'ajustement

Pour tester l'hypothèse

$$H_0 : P\{Y = i\} = p_i \quad \text{pour} \quad 1 \leq i \leq k$$

à partir de l'échantillon $Y_1, Y_2, \ldots, Y_n$, on introduit d'abord de nouvelles variables $X_1, X_2, \ldots, X_k$ définies par

$$X_i = \text{ le nombre des variables } Y_1, Y_2, \ldots, Y_n \text{ égales à } i$$

(donc $\sum_{i=1}^{k} X_i = n$) et on considère la statistique

$$STT = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i}.$$

K. Pearson (l'un des fondateurs de la statistique mathématique) a montré que si n est grand, STT suit pratiquement une loi du khi-deux à $k - 1$ degrés de liberté. (Cette approximation est généralement bonne dès que $n \times \min\{p_1, p_2, \ldots, p_k\} \geq 5$.)

Les valeurs $y_1, y_2, \ldots, y_n$ des variables $Y_1, Y_2, \ldots, Y_n$ ayant été observées et

$$stt = \sum_{i=1}^{k} \frac{x_i^2}{np_i} - n$$

ayant été calculée, on rejette H_0 au seuil de signification α si

$$stt > \chi^2_{\alpha,k-1}.$$

Alternativement, on peut d'abord calculer la p-valeur du test,

$$\text{p-valeur} = P\{\chi^2_{k-1} > stt\},$$

et rejeter H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

S'il est nécessaire d'estimer m quantités pour obtenir des estimateurs $\hat{p}_i$,

$$STT = \sum_{i=1}^{k} \frac{(X_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

suit une loi du khi-deux à $k-1-m$ degrés de liberté et il faut rejeter H_0 si

$$stt = \sum_{i=1}^{k} \frac{x_i^2}{n\hat{p}_i} - n > \chi^2_{\alpha,k-1-m}.$$

La p-valeur du test est alors

$$\text{p-valeur} = P\{\chi^2_{k-1-m} > stt\}.$$

Exemple. Les fréquences d'apparition de chacun des dix chiffres 0,1,2,...,9 dans les 6 000 000 000 premières décimales du nombre π sont

$$599963005, 600033260, 599999169, 600000243, 599957439,$$

$$600017176, 600016588, 600009044, 599987038, 600017038.$$

Pour tester l'hypothèse H_0 d'équiprobabilité, on utilise la statistique

$$STT = \sum_{i=1}^{10} \frac{(X_i - 600000000)^2}{600000000}$$

qui donne ici

$$stt = 8,99554.$$

La p-valeur du test est

$$P\{\chi^2_9 > 8,99554\} = 0,437686.$$

Il faut donc accepter l'hypothèse H_0. (Source : J.-P. Delahaye, Le fascinant nombre π, Pour la science, 1997).

7.2 Tests d'indépendance

Pour tester l'hypothèse

$$H_0 : P\{X=i, Y=j\} = P\{X=i\}P\{Y=j\} \ \text{ pour } \ 1 \leq i \leq r \ \text{ et } \ 1 \leq j \leq s$$

à partir d'un échantillon de taille n où

$$N_{i,j} = \text{le nombre d'individus tels que } \ X = i \ \text{ et que } \ Y = j$$

(donc $\sum_{i=1}^{r} \sum_{j=1}^{s} N_{i,j} = n$), soient

$$N_i = \sum_{j=1}^{s} N_{i,j} \ \text{ et } \ M_j = \sum_{i=1}^{r} N_{i,j}.$$

On estime les probabilités inconnues $P\{X = i\}$ et $P\{Y = j\}$ par

$$\hat{p}_i = \frac{N_i}{n} \ \text{ et } \ \hat{q}_j = \frac{M_j}{n}$$

et on considère la statistique

$$STT = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(N_{i,j} - n\hat{p}_i\hat{q}_j)^2}{n\hat{p}_i\hat{q}_j}.$$

Les valeurs $n_{i,j}$ des variables $N_{i,j}$ ayant été observées et

$$stt = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{n_{i,j}^2}{n\hat{p}_i\hat{q}_j} - n$$

ayant été calculé, on rejette H_0 si

$$stt > \chi^2_{\alpha,(r-1)(s-1)}.$$

Alternativement, on peut d'abord calculer la p-valeur du test,

$$\text{p-valeur} = P\{\chi^2_{(r-1)(s-1)} > stt\},$$

et rejeter H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

Exemple. Pour tester l'indépendance H_0 du sexe et du statut matrimonial d'un individu, on dispose des données suivantes (en pourcentages) :

	célibataire	marié	séparé	veuf	divorcé
masculin	44	48	2	2	4
féminin	38	46	2	8	6

La valeur de la statistique

$$STT = \sum_{i=1}^{2} \sum_{j=1}^{5} \frac{(N_{i,j} - n\hat{p}_i\hat{q}_j)^2}{n\hat{p}_i\hat{q}_j}$$

est ici

$$stt = \frac{(44-41)^2}{41} + \frac{(48-47)^2}{47} + \frac{(2-2)^2}{2} + \frac{(2-5)^2}{5} + \frac{(4-5)^2}{5}$$
$$+ \frac{(38-41)^2}{41} + \frac{(46-47)^2}{47} + \frac{(2-2)^2}{2} + \frac{(8-5)^2}{5} + \frac{(6-5)^2}{5} = 4,48158$$

et la p-valeur

$$P\{\chi_4^2 > 4,48158\} = 0,344737.$$

Il faut accepter H_0. (http ://www.stat.gouv.qc.ca).

8 La régression linéaire

Pour ajuster une droite $y = a + bx$ à un nuage de données (x_k, y_k) $(1 \leq k \leq n)$, on utilise la droite des moindres carrés, celle qui minimise la somme des carrés des erreurs commises :

$$\sum_{k=1}^{n}(y_k - a - bx_k)^2.$$

En annulant les dérivées partielles de cette expression par rapport aux variables a et b, on voit que les coefficients a et b sont en donnés par les relations

$$a = \frac{\overline{y}\sum_{k=1}^{n} x_k^2 - \overline{x}\sum_{k=1}^{n} x_k y_k}{\sum_{k=1}^{n}(x_k - \overline{x})^2}$$

et

$$b = \frac{\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})}{\sum_{k=1}^{n}(x_k - \overline{x})^2}.$$

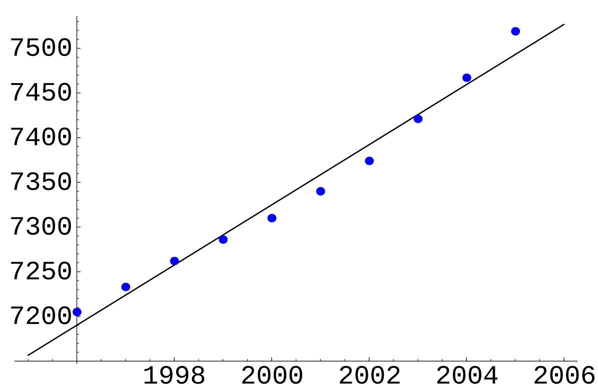

FIG. 13 – Une droite des moindres carrés

8.1 Les estimateurs de moindres carrés

On considère une variable aléatoire de la forme

$$Y = \alpha + \beta x + E$$

où $E \sim N(0, \sigma^2)$. Donc $Y \sim N(\alpha + \beta x, \sigma^2)$. Les estimateurs de moindres carrés A et B des coefficients de régression α et β, basés sur l'échantillon

$(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$, sont

$$A = \frac{\overline{Y} S_{x,x} - \overline{x} S_{x,Y}}{S_{x,x}}$$

et

$$B = \frac{S_{x,Y}}{S_{x,x}}$$

où l'on a posé

$$S_{u,v} = \sum_{k=1}^{n} (u_k - \overline{u})(v_k - \overline{v}) = \sum_{k=1}^{n} u_k v_k - n\overline{u}\overline{v}.$$

La somme des résidus est alors

$$SS_R = \sum_{k=1}^{n} (Y_k - A - Bx_k)^2 = \frac{S_{x,x} S_{Y,Y} - S_{x,Y}^2}{S_{x,x}}.$$

En écrivant

$$B = \sum_{k=1}^{n} \frac{x_k - \overline{x}}{S_{x,x}} \, Y_k,$$

on voit que

$$B \sim N\left(\beta, \frac{\sigma^2}{S_{x,x}}\right)$$

et en écrivant

$$A = \overline{Y} - B\overline{x},$$

on obtient

$$A \sim N\left(\alpha, \frac{\sigma^2 \sum_{k=1}^{n} x_k^2}{n S_{x,x}}\right).$$

Quant à SS_R, on peut montrer qu'elle est indépendante de A et de B et que

$$\frac{SS_R}{\sigma^2} \sim \chi_{n-2}^2.$$

8.2 Un test d'hypothèse

Pour tester l'hypothèse

$$H_0 : \beta = 0$$

contre l'hypothèse alternative

$$H_1 : \beta \neq 0,$$

on utilise la statistique

$$STT = \sqrt{\frac{(n-2)S_{x,x}}{SS_R}}\, B$$

car

$$\frac{S_{x,x}B}{\sigma}$$

suit une loi normale standard et la relation

$$\mathbb{E}\left(\frac{SS_R}{\sigma^2}\right) = n-2$$

permet d'estimer σ. Les valeurs $y_1, y_2, \ldots, y_n$ des variables $Y_1, Y_2, \ldots, Y_n$ ayant été observées et

$$stt = \sqrt{\frac{(n-2)S_{x,x}}{ss_R}}\, b$$

ayant été calculée, on rejette H_0 au seuil γ si

$$|stt| > t_{\alpha/2,n-2}.$$

Alternativement, on peut d'abord calculer la p-valeur du test,

$$\text{p-valeur} = P\{|T_{n-2}| > |stt|\},$$

et rejeter H_0 pour tout seuil α tel que

$$\text{p-valeur} \leq \alpha.$$

Exemple. La population québécoise.

Si x désigne l'année et Y la taille de la population québécoise (en milliers), on a les données suivantes :

x	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Y	7205	7233	7262	7286	7310	7340	7374	7421	7467	7519

La droite des moindres carrés pour ces données est

$$Y = -59984,2 + 33,6545\ x$$

(figure (13)). Pour tester l'hypothèse $H_0 : \beta = 0$ contre l'alternative $H_1 : \beta \neq 0$, on utilise la statistique

$$STT = \sqrt{\frac{(n-2)S_{x,x}}{SS_R}}\, B$$

qui vaut ici

$$sst = 19,3803$$

La p-valeur est

$$P\{|T_8| > 19,3803\} = 5,2 \times 10^{-8}.$$

Il faut rejeter H_0. (http ://www.stat.gouv.qc.ca).

8.3 Intervalles de confiance

a) Un intervalle de confiance de seuil $1-\gamma$ pour β a pour extrémités

$$b \pm t_{\gamma/2,n-2}\sqrt{\frac{ss_R}{(n-2)S_{x,x}}}.$$

b) Un intervalle de confiance de seuil $1-\gamma$ pour α a pour extrémités

$$a \pm t_{\gamma/2,n-2}\sqrt{\frac{ss_R \sum_{k=1}^n x_k^2}{(n-2)nS_{x,x}}}.$$

c) Un intervalle de confiance de seuil $1-\gamma$ pour la valeur moyenne $\alpha+\beta x_0$ de $Y(x_0)$ a pour extrémités

$$a + bx_0 \pm t_{\gamma/2,n-2}\sqrt{\frac{ss_R}{n-2}\left(\frac{1}{n} + \frac{(\overline{x}-x_0)^2}{S_{x,x}}\right)}$$

puisque

$$A+Bx_0 = \sum_{k=1}^n \left(\frac{1}{n} + \frac{(\overline{x}-x_k)(\overline{x}-x_0)}{S_{x,x}}\right) Y_k \sim N\left(\alpha + \beta x_0,\ \sigma^2\left(\frac{1}{n} + \frac{(\overline{x}-x_0)^2}{S_{x,x}}\right)\right).$$

d) Un intervalle de prédiction de seuil $1-\gamma$ pour $Y(x_0)$ a pour extrémités

$$a + bx_0 \pm t_{\gamma/2,n-2}\sqrt{\frac{ss_R}{n-2}\left(1 + \frac{1}{n} + \frac{(\overline{x}-x_0)^2}{S_{x,x}}\right)}$$

puisque

$$Y - A - Bx_0 \sim N\left(0, \sigma^2\left(1 + \frac{1}{n} + \frac{(\overline{x} - x_0)^2}{S_{x,x}}\right)\right).$$

Exemple. La population québécoise (suite).

Un intervalle de confiance de seuil 5% pour la moyenne de la population québécoise en 2007, $\mathbb{E}(Y(2007))$, est

$$(7531, 92\ ,\ 7588, 84).$$

Un intervalle de prédiction de seuil 5% pour la population québécoise en 2007, $Y(2007)$, est

$$(7514, 2\ ,\ 7606, 56).$$

8.4 Évaluation du modèle

La variation

$$S_{Y,Y} = \sum_{k=1}^{n}(Y_k - \overline{Y})^2$$

est partiellement expliquée par les variations des Y_k dues aux variations des x_k et partiellement expliquée par les fluctuations aléatoires de Y_k autour des valeurs prédites par les x_k, fluctuations mesurées par

$$SS_R = \sum_{k=1}^{n}(Y_k - A - Bx_k)^2.$$

Par conséquent, le **coefficient de détermination**,

$$R^2 = \frac{S_{Y,Y} - SS_R}{S_{Y,Y}}$$

est une mesure de la proportion de la variation due aux variations de x_k. Lorsqu'il est proche de 1, cela signifie que la variation est surtout due aux variations des x_k et que le modèle de la régression linéaire est bon.

Exemple. La population québécoise (fin).

Le coefficient de détermination pour la régression de la population québécoise est

$$r^2 = 0, 979145.$$

Remarquons qu'au signe près, R est égal au coefficient de corrélation

$$r(x,y) = \frac{S_{x,y}}{\sqrt{S_{x,x}S_{y,y}}}$$

des suites $(x_1, x_2, \ldots, x_n)$ et $(y_1, y_2, \ldots, y_n)$.

8.5 Extension du modèle

On peut adapter la régression linéaire à des variables

$$Y = f(x, \theta)$$

en effectuant des changements de variables appropriés.

Exemple. La population mondiale a cru de façon exponentielle au $XX^{ième}$ siècle, suivant les données :

x	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
Y	165	175	186	207	230	252	302	370	444	527	606

où la population Y est en dizaines de milliers. En supposant une relation de la forme

$$Y = e^{\alpha + \beta x + E}$$

où $E \sim N(0, \sigma^2)$, on obtient en appliquant la méthode des moindres carrés aux données $(x, \log Y)$ la courbe

$$Y = e^{-20{,}8636 + 0{,}0135986x}.$$

(http ://www.un.org).

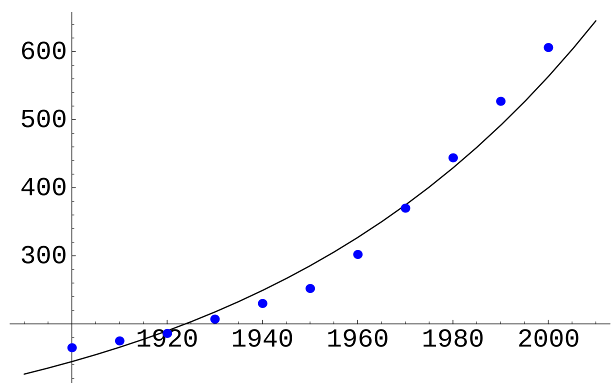

Fig. 14 – Une exponentielle des moindres carrés.

Index