

Partie II

Probabilités

4. Notions de calcul des probabilités
5. Applications du calcul des probabilités
6. Variables aléatoires
7. Modèles de distributions fréquemment utilisés

Chapitre 4

Notions de calcul des probabilités

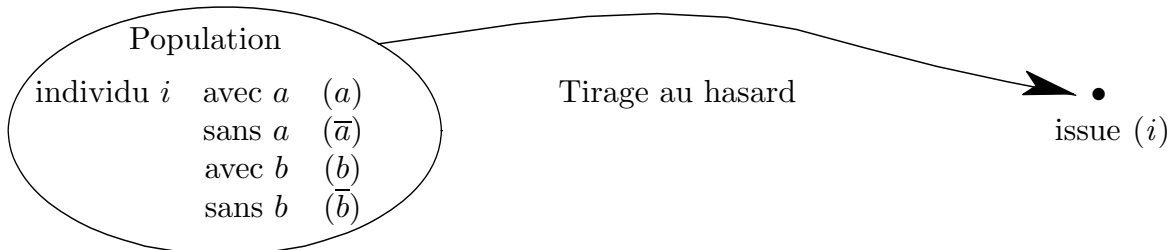
Le calcul des probabilités s'occupe d'*épreuves aléatoires* et de *phénomènes aléatoires* c'est-à-dire d'expériences ou de phénomènes naturels qui, dans des conditions déterminées et stables, ne mènent pas toujours à la même *issue*. On observe, cependant, une certaine régularité statistique. L'étude de cette régularité fait l'objet d'une théorie mathématique. Nous nous limitons à la présentation des sujets nécessaires aux applications statistiques traitées dans les chapitres ultérieurs. Ce chapitre, en particulier, est très succinct. Le lecteur trouvera facilement des textes d'introduction au calcul des probabilités (par exemple, celui de F. Taillard (1987) destiné aux élèves des écoles secondaires).

4.1 Univers

Dans le calcul des probabilités, on définit l'*univers* comme l'ensemble (noté U) de toutes les *issues* possibles (*modalités*), incompatibles deux à deux, qui peuvent être observées au cours d'une épreuve ou d'un phénomène aléatoire. Parfois, pour la même épreuve, il est possible de définir l'univers de plusieurs façons.

Exemple

Supposons que chaque individu d'une population ait ou n'ait pas les caractères a (par exemple, il fume) et b (par exemple, il a un cancer). Si on observe un individu avec a on dira qu'on observe a ; si on observe un individu sans a , on dira qu'on observe \bar{a} , etc. Soient i_1, i_2, \dots, i_N les individus.



Des exemples d'épreuve aléatoire (sondages de taille $n = 1$) sont:

S_1 : On tire au hasard un individu et on observe a ou \bar{a} . Un univers possible est la population entière:

$$U_0 = \{i_1, i_2, \dots, i_N\}.$$

Comme on s'intéresse seulement à un certain aspect de chaque individu, il convient généralement de travailler avec l'univers

$$U_1 = \{a, \bar{a}\}.$$

S_2 : On tire un individu i et on observe a ou \bar{a} et b ou \bar{b} ;

$$U_2 = \{ab, a\bar{b}, \bar{a}b, \bar{a}\bar{b}\}.$$

S_3 : On tire un individu i et on observe son poids;

$$U_3 = [1\text{Kg}, 300\text{Kg}].$$

4.2 Événements

On appelle *événement* tout sous-ensemble A de l'univers U . On dit que l'événement A a lieu ou se réalise si, lors du déroulement de l'épreuve, se présente l'issue s et que $s \in A$.

Exemples

- Il y a $4 = 2^2$ événements pour le sondage S_1 :
ce sont les sous-ensembles $\{a\}$, $\{\bar{a}\}$, $\{a, \bar{a}\} = U$ et $\emptyset = \{ \}$ (l'ensemble vide).
- Il y a $16 = 2^4$ événements pour S_2 ; en voici quelques-uns:
 $\emptyset = \{ \}$, $\{ab\}$, $\{a\bar{b}\}$, $\{\bar{a}b\}$, ..., $\{ab, a\bar{b}\}$, ..., $\{ab, a\bar{b}, \bar{a}b, \bar{a}\bar{b}\} = U$.
- L'événement $A = \text{“il fume”} = \{ab, a\bar{b}\}$ a lieu en S_2 si l'issue ab ou l'issue $a\bar{b}$ se réalise.
- Si ab se présente en S_2 (l'individu fume et a un cancer), tous les événements suivants ont lieu:
 $\{ab\}$
 $\{ab, a\bar{b}\}$, $\{ab, \bar{a}b\}$, $\{ab, \bar{a}\bar{b}\}$
 $\{ab, a\bar{b}, \bar{a}b\}$, $\{ab, a\bar{b}, \bar{a}\bar{b}\}$, $\{ab, \bar{a}b, \bar{a}\bar{b}\}$
 $\{ab, a\bar{b}, \bar{a}b, \bar{a}\bar{b}\}$

On dit que U est l'*événement certain* et que \emptyset est l'*événement impossible*. En outre, on utilise le calcul des ensembles pour définir de nouveaux événements. Par exemple,

- l'événement “non A ”, noté \bar{A} , est l'événement qui a lieu si et seulement si A n'a pas lieu; c'est donc le complément de A par rapport à U ;
- l'événement $A \cup B$ (A ou B) a lieu si A ou B (ou les deux) se réalisent;
- l'événement $A \cap B$ (A et B) a lieu si A et B ont lieu.

Si $A \cap B = \emptyset$, on dit que A et B sont *incompatibles*. A et B ne peuvent pas avoir lieu simultanément.

4.3 Probabilités

Evaluation de probabilités élémentaires

La possibilité d'identifier des *issues équiprobables* dans une épreuve aléatoire est un chemin fréquemment emprunté pour donner une valeur à la probabilité de certains événements élémentaires. Par exemple, dans le sondage S_1 , on peut admettre que tous les individus sont identiques en ce qui concerne le tirage aléatoire. Les chances de tirer un certain individu i sont donc égales à $1/N$ pour tout i , N étant la taille de la population. De même, lors du jet d'un dé on évalue la probabilité d'obtenir “un”, “deux”, ..., ou “six” à $1/6$ car on peut assimiler un dé à un cube avec 6 faces identiques et il n'y a pas de raisons qu'une face se manifeste plus souvent qu'une autre. Il est alors raisonnable de dire que la probabilité de tirer un nombre pair est de $1/2$, car il y a 3 possibilités sur 6, favorables à ce résultat. De façon similaire, dans le sondage S_1 , il est raisonnable de définir la probabilité de $A = \{a\} = \text{“il fume”}$ par

$$P(A) = N_a/N,$$

où N_a le nombre de fumeurs dans la population (voir Exemple 1 ci-dessous). Sur la base du même principe on posera $P(B) = N_b/N$ pour $B = \{b\}$. A l'aide des règles du calcul des probabilités que nous rencontrerons dans ce chapitre on pourra ensuite calculer la probabilité d'événements plus complexes comme $P(A \cap B)$ ou $P(A \cup B)$.

Malheureusement, pour la plupart des phénomènes il n'est pas possible d'identifier des issues équiprobables. Par exemple, si on souhaite évaluer les probabilités de “beau”, “nuageux”, “pluvieux”, “orageux” pour le temps qu'il fera demain à Lausanne, comment trouver des issues équiprobables ? Dans ce cas, il faut recourir à des *statistiques* météorologiques pour attribuer une valeur à $P(\text{“beau”})$ ou à $P(\text{“nuageux”})$.

Parfois même l'approche statistique n'est pas adaptée. Par exemple, supposons qu'un géologue doit évaluer la probabilité de trouver du pétrole à Lausanne. Cet expert utilisera ses connaissances scientifiques et son flair pour exprimer son degré de certitude. Il exprimera une *probabilité subjective* de trouver du pétrole.

On remarque heureusement que certaines propriétés mathématiques fondamentales des probabilités évaluées à l'aide de différentes approches ne dépendent pas de l'approche. On peut alors définir “la probabilité” par le biais de quelques axiomes et dériver ses propriétés à partir de ces axiomes.

Définition axiomatique de probabilité

Soit U un univers. Pour tout événement A , on définit $P(A)$, la *probabilité de A* , comme étant un nombre réel appartenant à l'intervalle $[0, 1]$ tel que:

1. $P(A) \geq 0$ pour tout événement A ;
2. $P(U) = 1$;
3. Si $A \cap B = \emptyset$, alors $P(A \cup B) = P(A) + P(B)$.

Conséquences élémentaires

Le lecteur pourra effectuer les démonstrations des propriétés suivantes:

- $P(\emptyset) = 0$
- $P(\bar{A}) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Si $A \subseteq B$, alors $P(A) \leq P(B)$.

Exemple 1

Considérons le sondage S_1 avec l'univers U_0 et supposons que $P(i_k) = 1/N$ pour tout individu i_k (chaque individu est équiprobable). L'événement "observer un fumeur" est le sous-ensemble $A = \{j_1, j_2, \dots, j_{N_a}\}$ de U_0 formé des fumeurs et N_a est le nombre de fumeurs. Comme on ne peut pas observer simultanément deux individus différents et

$$A = \{j_1\} \cup \{j_2\} \cup \dots \cup \{j_{N_a}\},$$

la propriété 3 de la définition axiomatique nous permet d'affirmer que

$$P(A) = \frac{N_a}{N}.$$

Par analogie,

$$P(B) = \frac{N_b}{N}$$

et

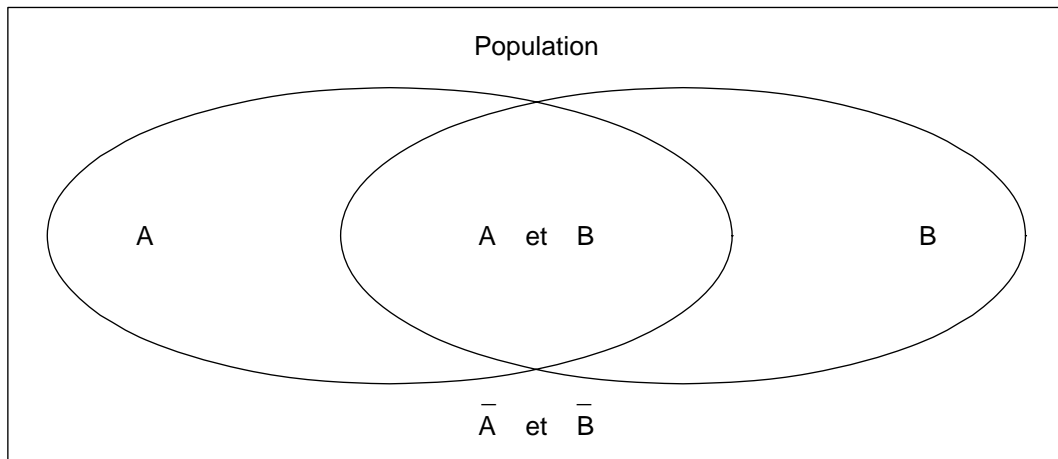
$$P(A \cap B) = \frac{N_{ab}}{N},$$

où N_{ab} est le nombre de fumeurs avec cancer.

Exemple 2

Considérons toujours le sondage S_1 . On obtient facilement

$$\begin{aligned} P(A \cup B) &= (N_{a\bar{b}} + N_{ab} + N_{\bar{a}b})/N \\ &= [(N_{a\bar{b}} + N_{ab}) + (N_{\bar{a}b} + N_{ab}) - N_{ab}]/N \\ &= N_a/N + N_b/N - N_{ab}/N = P(A) + P(B) - P(A \cap B) \end{aligned}$$

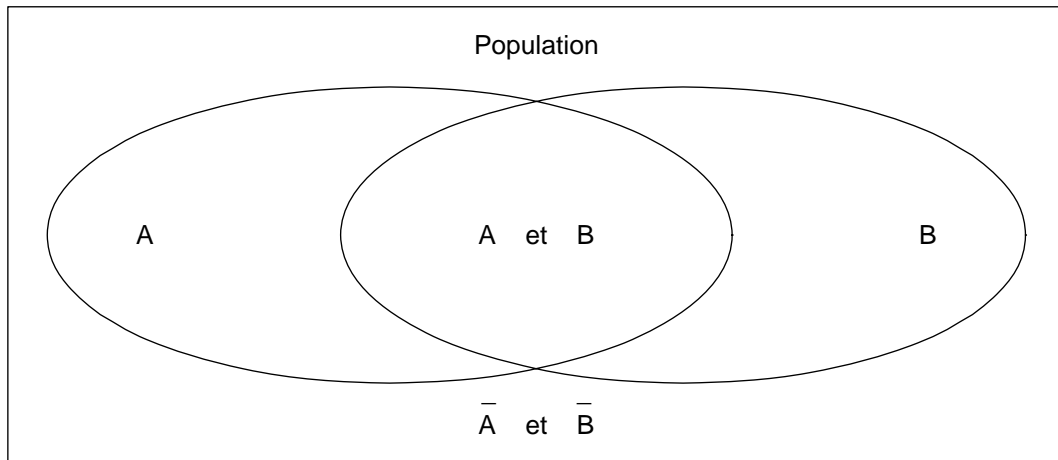
**4.4 Probabilité conjointe**

La probabilité $P(A \cap B)$ est appelée la *probabilité conjointe de A et B*.

4.5 Probabilité conditionnelle

Vision intuitive

Reprenons le sondage S_2 et soient $A = \{a\}$, $B = \{b\}$.



Notons:

N_a = nombre d'individus avec a , N_b = nombre d'individus avec b ,

N_{ab} = nombre d'individus avec a et b ; N = taille de la population.

On a alors $P(A) = N_a/N$, $P(B) = N_b/N$, $P(A \cap B) = N_{ab}/N$. La probabilité de cancer pour un fumeur peut être définie comme:

$$\frac{N_{ab}}{N_a} = \frac{N_{ab}/N}{N_a/N} = \frac{P(A \cap B)}{P(A)}.$$

Il s'agit de la proportion d'individus avec cancer dans la sous-population de fumeurs. On dit que cette proportion est la probabilité conditionnelle de cancer pour un fumeur. Pour son interprétation on peut penser au tirage d'un individu pris au hasard parmi les fumeurs: la probabilité d'observer un cancer est N_{ab}/N_a . Dans cette épreuve l'univers n'est pas U mais la sous-population de fumeurs.

Définition générale

Soient A et B deux événements d'un univers U . Si $P(A) \neq 0$, on appelle *probabilité conditionnelle de B pour A donné* le nombre noté

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

Parfois, on écrit $P_A(B)$ à la place de $P(B|A)$.

$P(B|A)$ est une probabilité dans le sens de la définition axiomatique. L'univers, toutefois, n'est pas U : pour $P_A(B)$, l'univers est A . On peut réduire ultérieurement l'univers A en introduisant une deuxième condition C (c'est-à-dire la réalisation d'un événement C) et parler de $P(B|A \cap C)$, etc.

Propriétés élémentaires de la probabilité conditionnelle

De la définition de $P(B|A)$, on déduit

$$P(A \cap B) = P(A) \cdot P(B|A),$$

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B).$$

4.6 Événements indépendants

Vision intuitive

Si $P(B|A) = P(B)$, la probabilité de B ne dépend pas de la condition A . Par exemple, si dans le sondage S_2 , $P(B|A) = P(B)$, la probabilité d'observer un cancer ne dépend pas du fait que l'individu fume ou ne fume pas. Donc $P(B|\bar{A}) = P(B)$. En outre, par la définition de probabilité conditionnelle,

$$P(A \cap B) = P(A) \cdot P(B).$$

Cette dernière formule est valable même dans le cas où $P(A) = 0$ ou $P(B) = 0$. On l'utilise donc pour la définition générale.

Définition

Deux événements A et B sont *indépendants* si $P(A \cap B) = P(A) \cdot P(B)$.

On peut étendre la définition à trois ou plus événements. Par exemple, A, B, C sont *indépendants* si

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B), \\ P(A \cap C) &= P(A) \cdot P(C), \\ P(B \cap C) &= P(B) \cdot P(C) \text{ et} \\ P(A \cap B \cap C) &= P(A) \cdot P(B) \cdot P(C). \end{aligned}$$

4.7 Formule de Bayes

La *formule de Bayes* permet de calculer $P(B|A)$ lorsque $P(A|B)$, $P(A)$ et $P(B)$ sont connus:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}.$$

Pour calculer le dénominateur on utilise souvent la *formule de la probabilité totale*. Supposons que U puisse être partagé en deux parties disjointes B et \bar{B} , et donc $U = B \cup \bar{B}$, on a

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A|B) \cdot P(B) + P(A|\bar{B}) \cdot P(\bar{B}).$$

En général, si $B_1 \cup B_2 \cup \dots \cup B_k = U$ et $B_i \cap B_j = \emptyset$ pour $i \neq j$, on a

$$P(A) = \sum_{i=1}^k P(A \cap B_i) = \sum_{i=1}^k P(A|B_i) \cdot P(B_i).$$

Exemple

Un certain type d'objet est fabriqué dans trois usines notée 1, 2 et 3. L'usine 1 produit le double des objets produits par la 2 et la 2 produit le même nombre d'objets que la 3. Le 2% des objets produits par les usines 1 et 2 sont défectueux; le 4% des objets produits par la 3 sont défectueux. Tous les objets produits par les trois usines sont placés dans un magasin et on en prend un au hasard. Quelle est la probabilité qu'il soit défectueux ?

Considérons les événements $A =$ "l'objet est défectueux", $B_1 =$ "l'objet provient de 1", $B_2 =$ "l'objet provient de 2", $B_3 =$ "l'objet provient de 3". On obtient

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) \\ &= 0.02 \times 0.5 + 0.02 \times 0.25 + 0.04 \times 0.25 = 0.025. \end{aligned}$$

Supposons qu'un objet pris au hasard soit défectueux. Quelle est la probabilité qu'il ait été fabriqué dans l'usine 1 ?

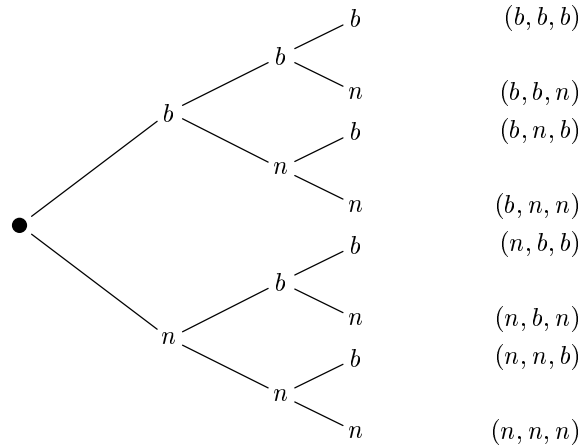
$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A)} \\ &= \frac{0.02 \times 0.5}{0.025} = 0.40. \end{aligned}$$

4.8 Technique des épreuves successives

Dans de nombreuses applications, une épreuve globale se décompose en n épreuves partielles successives. Les issues de l'univers U attaché à l'épreuve globale sont alors les n -uplets (a_1, a_2, \dots, a_n) .

Exemple

Soit une urne contenant 3 boules blanches et 4 boules noires. Supposons que l'épreuve soit "tirer 3 boules successives sans remise". Les issues non-équiprobables sont:



Considérons les événements suivants:

$A = \{(b, b, b), (b, b, n), (b, n, b), (b, n, n)\}$ noté $(b, -, -)$ (la première est blanche)

$B = (-, n, -)$ (la deuxième est noire),

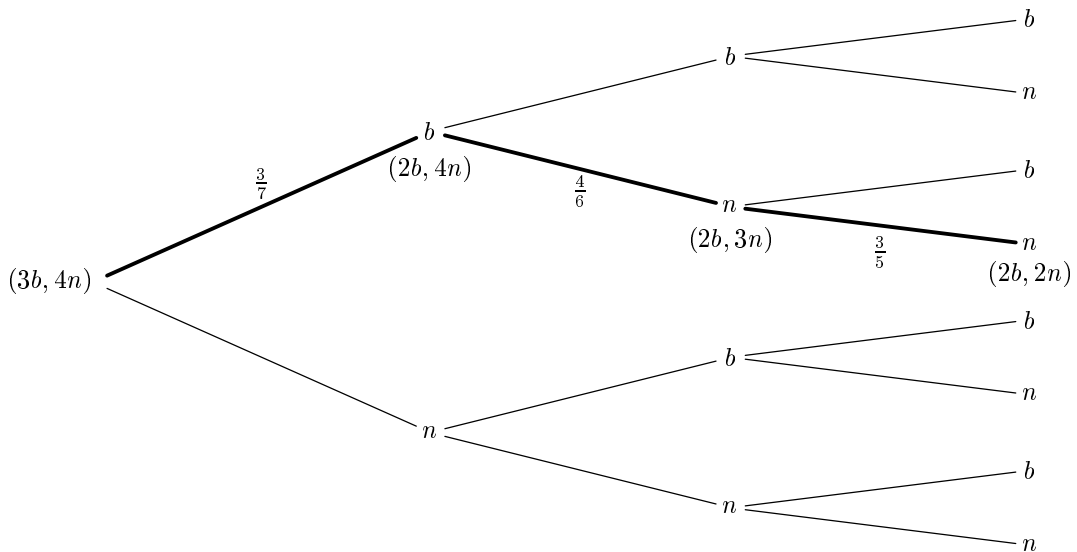
$C = (-, -, n)$ (la troisième est noire).

Alors $A \cap B = (b, n, -)$ et $A \cap B \cap C = (b, n, n)$. On a

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B),$$

et donc

$$P((b, n, n)) = P((b, -, -)) \cdot P((-, n, -)|(b, -, -)) \cdot P((-, -, n)|(b, n, -)) = \frac{3}{7} \cdot \frac{4}{6} \cdot \frac{3}{5} = \frac{6}{35}.$$



La probabilité d'un chemin est égale au produit des probabilités des branches qui forment le chemin. En outre, deux chemins différents sont toujours incompatibles. Ainsi, pour calculer la probabilité d'un événement qui est la réunion de plusieurs chemins, on additionne les probabilités de ces chemins.

Chapitre 5

Applications du calcul des probabilités

Dans ce chapitre, les notions fondamentales de calcul des probabilités (Chapitre 4) sont appliquées d'abord à un problème génétique (voir partie 5.1) puis à l'évaluation de la fiabilité des tests de dépistage et de diagnostic (dès partie 5.2). Là, les concepts de sensibilité et de spécificité sont introduits et utilisés pour calculer le risque de faux positif et la valeur prédictive du test, lesquels ont une importance capitale pour les sujets soumis au test.

5.1 Problème en génétique: transmission de caractères

Supposons que les individus d'une population aient les caractères R et r . Les individus adultes peuvent être de deux types: RR et Rr , car les rr meurent à la naissance. Nous considérons des mariages et des fils, et utilisons les notations suivantes:

	père	mère		
Mariages:	$A_1 : RR$	\sim	RR	Fils: $B_1 : RR$
	$A_2 : RR$	\sim	Rr	$B_2 : Rr$
	$A_3 : Rr$	\sim	RR	$B_3 : rr$
	$A_4 : Rr$	\sim	Rr	

Hypothèse A. La transmission des caractères suit le modèle de Mendel:

Mariage	fils	Probabilité
A_1	$\left\{ \begin{array}{l} RR \\ RR \end{array} \right\} \longrightarrow RR$	$: B_1 \quad 1$
A_2	$\left\{ \begin{array}{l} RR \\ Rr \end{array} \right\} \begin{array}{l} \nearrow \\ \searrow \end{array}$	$\begin{array}{l} RR : B_1 \quad 1/2 \\ Rr : B_2 \quad 1/2 \end{array}$
A_3	$\left\{ \begin{array}{l} Rr \\ RR \end{array} \right\} \begin{array}{l} \nearrow \\ \searrow \end{array}$	$\begin{array}{l} RR : B_1 \quad 1/2 \\ Rr : B_2 \quad 1/2 \end{array}$
A_4	$\left\{ \begin{array}{l} Rr \\ Rr \end{array} \right\} \begin{array}{l} \nearrow \\ \rightarrow \\ \searrow \end{array}$	$\begin{array}{l} RR : B_1 \quad 1/4 \\ Rr : B_2 \quad 1/2 \\ rr : B_3 \quad 1/4 \end{array}$

Hypothèse B. Les adultes se distribuent de la façon suivante:

$$\begin{aligned} P(rR) &= \lambda = \text{“fréquence de porteurs du gène } r\text{”}, \\ P(rr) &= 0 \text{ car les } rr \text{ meurent à la naissance,} \\ P(RR) &= 1 - \lambda. \end{aligned}$$

Hypothèse C. Les couples se forment au hasard et donc les génotypes du père et de la mère sont indépendants. En utilisant les abréviations p =père, m =mère,

$$\begin{aligned} P(A_1) &= P(p = RR \cap m = RR) = P(p = RR) \cdot P(m = RR|p = RR) = \\ &= P(p = RR) \cdot P(m = RR) = (1 - \lambda)^2, \\ P(A_2) &= P(A_3) = \lambda(1 - \lambda), \\ P(A_4) &= \lambda^2. \end{aligned}$$

Problème a

Sous les hypothèses A, B, C, quelle est la fréquence de porteurs du gène r dans la première génération parvenue à l'âge adulte ?

Les probabilités conditionnelles suivantes découlent de A:

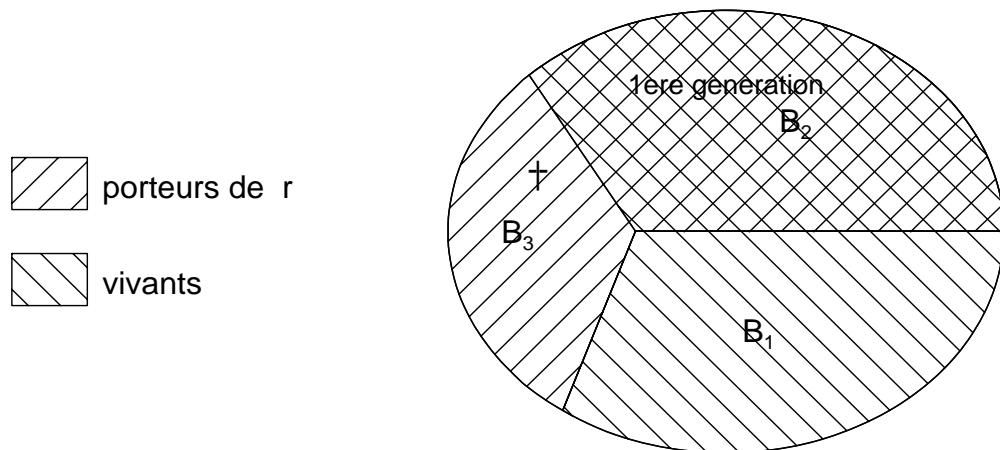
	A_1	A_2	A_3	A_4
$P(B_1 A_i)$	1	1/2	1/2	1/4
$P(B_2 A_i)$	0	1/2	1/2	1/2
$P(B_3 A_i)$	0	0	0	1/4

A l'aide de la formule de la probabilité totale on obtient:

$$P(B_1) = (1 - \lambda)^2 + \frac{1}{2}\lambda(1 - \lambda) + \frac{1}{2}\lambda(1 - \lambda) + \frac{1}{4}\lambda^2 = \dots = (1 - \frac{\lambda}{2})^2,$$

$$P(B_2) = \dots = \lambda(1 - \frac{\lambda}{2}),$$

$$P(B_3) = \dots = (\frac{\lambda}{2})^2.$$



Alors la probabilité cherchée est

$$\begin{aligned} P(\text{fils porteur de } r \mid \text{fils vivant}) &= P(B_2 \cup B_3 \mid \bar{B}_3) = P(B_2 \mid \bar{B}_3) + P(B_3 \mid \bar{B}_3) = \\ &= P(B_2 \mid \bar{B}_3) = \frac{P(B_2 \cap \bar{B}_3)}{P(\bar{B}_3)} = \frac{P(B_2)}{1 - P(B_3)} \\ &= \frac{\lambda(1 - \frac{\lambda}{2})}{1 - \frac{\lambda^2}{4}} = \frac{\lambda}{1 + \frac{\lambda}{2}} < \lambda \end{aligned}$$

ce qui signifie qu'un facteur mortel à tendance à s'éliminer naturellement et ce d'autant plus vite qu'il est répandu.

Problème b

Sous les hypothèses A, B et C, supposons que

- “mon frère est mort portant rr ”.

Quelle est la probabilité que je sois porteur du gène r ?

Comme mon frère est rr , le mariage de mes parents est le mariage noté A_4 . Ecrivons, par exemple, $P_{A_4}(B_1) = P(\text{moi} = B_1 \mid \text{mes parents} = A_4)$. Avec cette notation, on a (voir Problème a):

$$P_{A_4}(B_1) = \frac{1}{4}, \quad P_{A_4}(B_2) = \frac{1}{2}, \quad P_{A_4}(B_3) = \frac{1}{4}.$$

Donc:

$$P(\text{moi porteur} \mid \text{moi vivant}) = P_{A_4}(B_2 \cup B_3 \mid \bar{B}_3) = P_{A_4}(B_2 \mid \bar{B}_3) = \frac{\frac{1}{2}}{\frac{1}{4} + \frac{1}{2}} = \frac{2}{3}.$$

Problème c

Sous les hypothèses A, B, C supposons que

- “mon frère est mort portant rr ”.

Quelle est la probabilité que mon premier-né meurt parce qu’il porte le gène rr ?

Par le Problème b, on a que:

$$P(\text{moi} = Rr) = \frac{2}{3} \quad \text{et} \quad P(\text{moi} = RR) = \frac{1}{3}.$$

Alors:

	A_1	A_2	A_3	A_4
$P(\text{mon mariage} = A_i) :$	$\frac{1}{3}(1 - \lambda)$	$\frac{1}{3}\lambda$	$\frac{2}{3}(1 - \lambda)$	$\frac{2}{3}\lambda$
$P(\text{mon fils} = rr \mid A_i) :$	0	0	0	$\frac{1}{4}$

On en tire donc que $P(\text{mon fils} = rr) = \frac{1}{4} \cdot \frac{2}{3}\lambda = \frac{\lambda}{6}$.

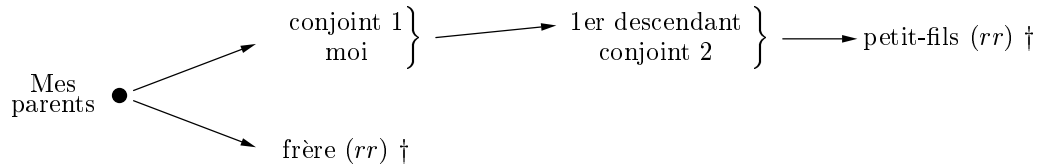
Problème d

Sous les hypothèses A, B, C, supposons que

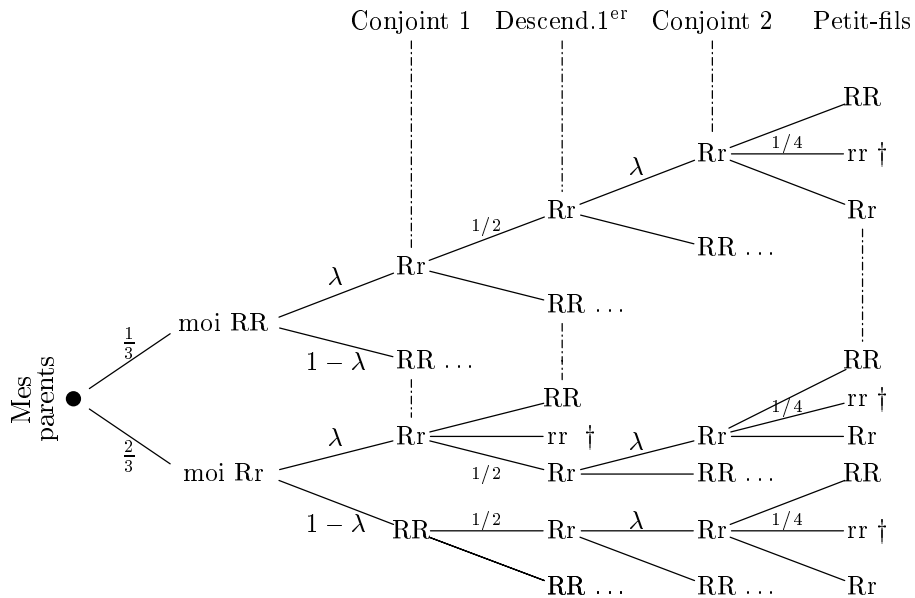
- “mon frère est mort portant rr ” et que
- “mon petit-fils meurt de rr ”.

Quelle est la probabilité que je sois porteur du gène r ?

Soit l’arbre généalogique qui nous intéresse:



Développons-le en marquant les branches qui se terminent par la mort du petit-fils:



Notons: A = “je porte r ” , B = “mon petit-fils est rr ”. Alors:

$$P(B|A) = \lambda \cdot \frac{1}{2} \cdot \lambda \cdot \frac{1}{4} + (1 - \lambda) \cdot \frac{1}{2} \cdot \lambda \cdot \frac{1}{4} = \frac{\lambda}{8} \quad (\text{branches du bas})$$

$$P(B|\bar{A}) = \lambda \cdot \frac{1}{2} \cdot \lambda \cdot \frac{1}{4} = \frac{\lambda^2}{8} \quad (\text{branches du haut})$$

Par le théorème de Bayes, on obtient la probabilité cherchée:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})} = \frac{\frac{2}{3} \cdot \frac{\lambda}{8}}{\frac{2}{3} \cdot \frac{\lambda}{8} + \frac{1}{3} \cdot \frac{\lambda^2}{8}} = \frac{2}{2 + \lambda}$$

5.2 Exemples de tests de dépistage et de diagnostic

Dans les années 1980-1990, plusieurs articles concernant les tests pour dépister les porteurs du virus HTLV-III ou pour aboutir à un diagnostic de séropositivité ont été publiés.

1. Dans JAMA (11 Janvier 1985, Vol 253; Weiss et al. 1985) on trouve que la *sensibilité* d'un certain test de type ELISA (Enzyme-Linked-Immunoabsorbent Assay) est d'environ 97.3%, tandis que sa *spécificité* est de 98.6%.
2. Un article publié dans NEJM (23 Juillet 1987, Vol 317; Meyer et Pauker 1987) observe que les *programmes de dépistage* courants utilisent des combinaisons de tests ELISA (dans la phase de dépistage) et Western blot (dans la phase de confirmation et d'établissement du *diagnostic*). La spécificité d'une *combinaison indépendante des deux tests* est d'environ 99.995%, ce qui est indice d'une très bonne performance. Cependant, l'article met en garde sur les conséquences d'une utilisation généralisée du test de dépistage, car le *risque de faux positifs* peut être important.
3. Un article du NEJM (13 Octobre 1988, Vol 319; Donald et al. 1988) observe que dans une certaine population (à faible *prévalence* de HTLV-III) de 135 000 recrues aux USA, une campagne de dépistage a fourni 15 résultats positifs au dépistage dont 14 positifs à la confirmation (1 résultat étant négatif à la confirmation). Il en découle que la proportion de résultats positifs parmi les recrues saines est approximativement $1/135000 = 0.000007 = 0.0007\%$. Cependant, la proportion de non porteurs parmi les résultats positifs est de $1/15 = 0.07 = 7\%$.

5.3 Définition de sensibilité et spécificité

Pour chaque individu examiné, nous nous intéressons aux caractères:

M = avoir la maladie

\bar{M} = ne pas avoir la maladie

T = avoir un résultat positif (+) au test

\bar{T} = avoir un résultat négatif (-) au test

Dans une phase d'évaluation, le test est appliqué à un groupe M et à un groupe \bar{M} . Les groupes sont formés à l'aide d'un *test de référence (gold standard)* dont le résultat est considéré comme sûr. On détermine les fréquences des quatre résultats possibles indiqués par n_{TM} , $n_{T\bar{M}}$, $n_{\bar{T}M}$ et $n_{\bar{T}\bar{M}}$ dans le tableau suivant

Fréquences des résultats

	M	\bar{M}	Total
T	n_{TM}	$n_{T\bar{M}}$	n_T
\bar{T}	$n_{\bar{T}M}$	$n_{\bar{T}\bar{M}}$	$n_{\bar{T}}$
Total	n_M	$n_{\bar{M}}$	n

Définitions:

$$\text{Sensibilité} = \frac{n_{TM}}{n_M} = \text{proportion de "+" parmi les malades,}$$

$$\text{Spécificité} = \frac{n_{\bar{T}\bar{M}}}{n_{\bar{M}}} = \text{proportion de "-" parmi les sains.}$$

Des valeurs élevées de sensibilité et de spécificité indiquent clairement une bonne qualité du test.

Exemple

Le test a été administré à 1000 personnes avec M
et à 1000 personnes sans M :

	M	\bar{M}
T	950	10
\bar{T}	50	990
Total	1000	1000

Donc:

$$\text{Sensibilité} = \frac{950}{1000} = 95\%,$$

$$\text{Spécificité} = \frac{990}{1000} = 99\%.$$

5.4 Les autres mesures d'efficacité d'un test

La sensibilité et la spécificité du test ne sont pas des informations directement utiles à l'individu (patient) auquel le test est administré. En effet, supposons que la sensibilité et la spécificité d'un certain test soient:

$$\text{Sensibilité} = 95\%,$$

$$\text{Spécificité} = 99\%.$$

Supposons aussi qu'un médecin applique ce test à un patient et obtient un résultat positif. Il faut répondre à la question:

Quelle est la probabilité que le patient soit réellement malade ?

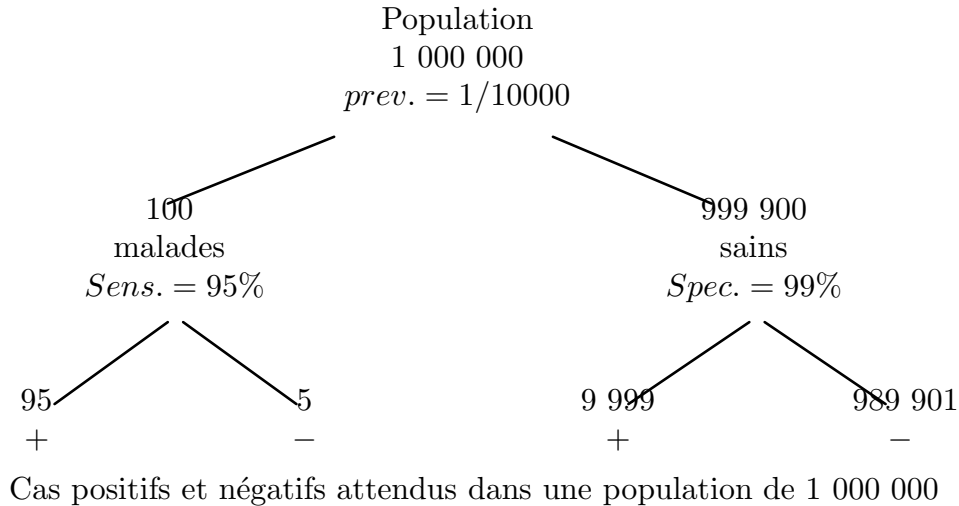
Pour répondre, il nous faut une information supplémentaire concernant la fréquence de la maladie M dans la population, c'est à dire la *prévalence* de la maladie. Supposons, par exemple, qu'il y ait 1 seul malade par 10 000 habitants. En d'autres termes:

$$\text{Prévalence de } M = \frac{1}{10\,000}.$$

Supposons en outre, que le patient soit un individu pris au hasard dans la population.

Solution intuitive

La prévalence de 1/10 000 nous permet d'affirmer que dans une population hypothétique de 1 000 000 d'individus, on peut s'attendre à 100 malades et 999 900 sains. Le test dépiste 95 cas positifs et 5 cas négatifs parmi les malades, car sa sensibilité est de 95%. Le test trouve aussi 9 999 résultats positifs et 989 901 résultats négatifs dans la partie saine de la population.



Ce schéma peut aussi être représenté dans un tableau de fréquences attendues:

Fréquences attendues

	malades	sains	Total
positifs	95	9 999	10094
négatifs	5	989 901	989 906
Total	100	999 900	1 000 000

En conclusion, la proportion de malades parmi les cas positifs est de 95/10 094, ce qui indique que les chances qu'un individu positif au test soit réellement malade sont seulement de 0.0094 ($\approx 1\%$). Assurez-vous que l'on trouve les mêmes proportions si la taille de la population est changée, par exemple 4 000 000.

Solution avec la formule de Bayes

Nous utilisons les notations du Chapitre 4, avec les interprétations de T , \bar{T} , M et \bar{M} données au début de la Section 5.3. On veut déterminer

$P(M|T)$ = Probabilité que le patient soit malade
 “donné le résultat positif du test”
 (sachant que le résultat du test est positif).

On remplace les valeurs

$$\begin{aligned} P(T|M) &= 95\% = \text{sensibilité,} \\ P(\bar{T}|\bar{M}) &= 99\% = \text{spécificité,} \\ P(M) &= 1/10\,000 = \text{prévalence,} \end{aligned}$$

et donc

$$\begin{aligned} P(T|\bar{M}) &= 1\%, \\ P(\bar{M}) &= 9\,999/10\,000, \end{aligned}$$

dans la formule de Bayes

$$P(M|T) = \frac{P(T|M)P(M)}{P(T|M)P(M) + P(T|\bar{M})P(\bar{M})}.$$

On obtient:

$$\begin{aligned} P(M|T) &= \frac{0.95 \times 0.0001}{0.95 \times 0.0001 + 0.01 \times 0.9999} \\ &= 0.0094. \end{aligned}$$

Remarque. Les valeurs de $P(T|M)$ et $P(\bar{T}|\bar{M})$ dans la population sont inconnues. Nous utilisons leurs estimations 95%, respectivement 99%, obtenues dans la phase d'évaluation à l'aide d'un échantillon de 1000 malades et 1000 sains.

Terminologie

$P(T|M)$ = *sensibilité du test*

$P(\bar{T}|\bar{M})$ = *spécificité du test*

$P(M|T)$ = *valeur prédictive positive du test*

$P(\bar{M}|\bar{T})$ = *valeur prédictive négative du test*

$P(T|\bar{M})$ = *taux de faux positifs* (= 1 – spécificité)

$P(\bar{T}|M)$ = *taux de faux négatifs* (= 1 – sensibilité)

Attention ! pour certains auteurs:

taux de faux positifs = $P(\bar{M}|T)$,

taux de faux négatifs = $P(M|\bar{T})$.

Vérifiez donc toujours la définition utilisée !

5.5 Le rôle de la prévalence

Il est souvent difficile de connaître la prévalence $P(M)$ avec précision. Il convient alors d'examiner le test pour différentes valeurs de $P(M)$. Par exemple, si $P(T|M) = 0.95$ et $P(\bar{T}|\bar{M}) = 0.99$, on obtient:

$P(M)$	$P(\bar{M} T)$	$P(M \bar{T})$
1/1 000 000	0.9999	0.00000
1/100 000	0.9991	0.00000
1/10 000	0.9906	0.00001
1/1000	0.9132	0.00005
1/500	0.8401	0.00010
1/200	0.6769	0.00025
1/100	0.5103	0.00051

Le taux $P(M|\bar{T})$ est bon: moins de 5 malades sur 10 000 (≈ 0.00051) échappent au test. Par contre le taux $P(\bar{M}|T)$ est élevé: sur 100 individus positifs plus de 50 (≈ 0.5103) sont sains. La décision de maintenir un tel test dépendra de l'importance de la maladie, des conséquences du test, des coûts des examens complémentaires et de l'éventuel traitement, des chances de succès du traitement etc. Il est important de savoir qu'on peut réduire les taux d'erreur en combinant (ou en répétant) deux ou plusieurs tests.

5.6 Combinaison de deux ou plusieurs tests

Notations:

M = le patient est malade	\bar{M} = le patient n'est pas malade
T_1 = le premier test est positif	\bar{T}_1 = le premier test est négatif
T_2 = le deuxième test est positif	\bar{T}_2 = le deuxième test est négatif

Informations disponibles:

$P(M)$	Prévalence	par exemple 10%
$P(T_1 M)$	Sensibilité du premier test	par exemple 75%
$P(\bar{T}_1 \bar{M})$	Spécificité du premier test	par exemple 80%
$P(T_2 M)$	Sensibilité du deuxième test	par exemple 90%
$P(\bar{T}_2 \bar{M})$	Spécificité du deuxième test	par exemple 95%

Hypothèse:

les résultats des tests sont indépendants et cela chez les malades et chez les sains (indépendance conditionnelle).

Valeur prédictive positive du premier test:

$$P(M|T_1) = \frac{P(T_1|M)P(M)}{P(T_1|M)P(M) + P(T_1|\bar{M})P(\bar{M})} = \frac{0.750 \times 0.100}{0.750 \times 0.100 + 0.200 \times 0.900} = 0.294.$$

Valeur prédictive positive de la combinaison

Première approche

Grâce à l'indépendance conditionnelle des résultats,

$$\begin{aligned} P(T_1 \text{ et } T_2|M) &= P(T_1|M)P(T_2|M) = 0.750 \times 0.900 = 0.675, \\ P(T_1 \text{ et } T_2|\bar{M}) &= P(T_1|\bar{M})P(T_2|\bar{M}) = 0.200 \times 0.050 = 0.010, \end{aligned}$$

et donc

$$\begin{aligned} P(M|T_1 \text{ et } T_2) &= \frac{P(T_1 \text{ et } T_2|M)P(M)}{P(T_1 \text{ et } T_2|M)P(M) + P(T_1 \text{ et } T_2|\bar{M})P(\bar{M})} \\ &= \frac{0.675 \times 0.100}{0.675 \times 0.100 + 0.010 \times 0.900} = 0.882. \end{aligned}$$

Deuxième approche

Nous notons par $P_{T_1}(\dots) = P(\dots|T_1)$ la probabilité conditionnelle (donné le résultat T_1). Après avoir obtenu le résultat T_1 , nous interprétons $P_{T_1}(M) = P(M|T_1)$ comme "prévalence de M parmi les résultats T_1 ". En outre, grâce à l'indépendance des résultats, $P_{T_1}(T_2|M) = P(T_2|M)$. La valeur prédictive du deuxième test, donné T_1 , est alors:

$$\begin{aligned} P_{T_1}(M|T_2) &= \frac{P_{T_1}(T_2|M)P_{T_1}(M)}{P_{T_1}(T_2|M)P_{T_1}(M) + P_{T_1}(T_2|\bar{M})P_{T_1}(\bar{M})} \\ &= \frac{0.900 \times 0.294}{0.900 \times 0.294 + 0.05 \times 0.706} = 0.882. \end{aligned}$$

Algorithme de mise à jour de la valeur prédictive

La deuxième approche indique un chemin général de mise à jour de la valeur prédictive lorsque plusieurs tests (conditionnellement indépendants) peuvent être réalisés. Les informations initiales sont la prévalence $P(M)$ et les caractéristiques $P(T_1|M)$ et $P(\bar{T}_1|\bar{M})$ du premier test. Nous utilisons la notation $P^{(1)} = P(M)$. Ces données, ainsi que le résultat T_1 du test, nous permettent de calculer $P(M|T_1)$:

$$\begin{array}{ccc} P^{(1)} & \text{Formule de Bayes} & \\ P(T_1|M) & \longrightarrow & P(M|T_1) \\ P(\bar{T}_1|\bar{M}) & & \end{array}$$

Nous interprétons $P(M|T_1)$ comme une nouvelle valeur $P^{(2)}$ de la "prévalence", obtenue grâce à l'information apportée par le résultat du test. Les caractéristiques et le résultat T_2 du deuxième test nous permettent d'obtenir une nouvelle valeur prédictive $P(M|T_1 \text{ et } T_2)$:

$$\begin{array}{ccc} P^{(2)} & \text{Formule de Bayes} & \\ P(T_2|M) & \longrightarrow & P(M|T_1 \text{ et } T_2) \\ P(\bar{T}_2|\bar{M}) & & \end{array}$$

La valeur $P(M|T_1 \text{ et } T_2)$ est la nouvelle valeur de prévalence $P^{(3)}$ pour la prochaine étape de l'algorithme, et ainsi de suite. L'algorithme s'arrête lorsque la valeur prédictive est suffisamment élevée ou basse pour prendre une décision définitive.

5.7 Exemples

1. Plusieurs articles ont soulevé des doutes concernant la fiabilité des tests urinaires réalisés par les laboratoires pour détecter la présence de drogues communes, comme la cocaïne, la méthadone et les amphétamines. Des chercheurs du Center of Disease Control aux USA ont conduit une étude (transversale) pour déterminer les taux d'erreur de 13 laboratoires (Hansen et al., 1985). Des échantillons d'urine ont été envoyés aux laboratoires. Certains échantillons présentaient des concentrations de drogue (ou de son métabolite) supérieures aux limites et ne nécessitant pas un rapport; cependant, aucune information concernant l'état réel des échantillons a été fournie. Les données résumant les performances des laboratoires dans leur ensemble sont collectées dans le Tableau 5.3. Calculez la spécificité et la sensibilité des tests utilisés pour la détection de chaque drogue.

Drug Class	Positive Samples		Negative Samples	
	No. Challenges	No. Correct	No. Challenges	No. Correct
Barbiturates	455	187	689	689
Amphetamines	572	177	637	618
Methadone	533	469	663	583
Cocaine	416	150	793	785
Codeine	481	216	715	708
Morphine	468	178	728	713

Tableau 5.3 Adapté et reproduit des Tables 7 et 8 de Hansen JH, Caudill SP, Boone J., *Crisis in drug testing*. JAMA 1985; 253:23982.

2. La Figure 5.2 provient de l'article du NEJM (23 Juillet 1987, Vol. 317) mentionné dans la Section 5.2. Interprétez la Figure 5.2 et comprenez pourquoi l'article met en garde sur les conséquences d'une utilisation généralisée du test.

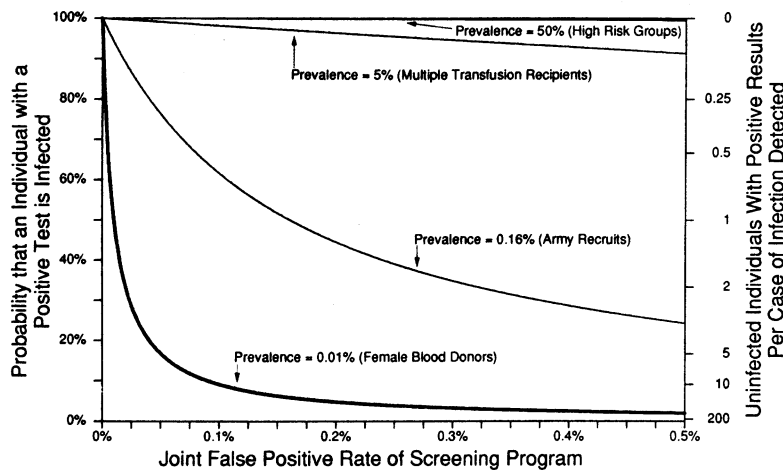


Figure 1. Meaning of Positive Screening Tests for HIV.

Figure 5.2 Meaning of Positive Screening Tests for HIV. The horizontal axis shows the joint false positive rate of the tests. The left vertical scale shows the probability that HIV infection is present in a person with positive tests. The right vertical scale shows the number of uninfected persons falsely classified as infected for every infected person correctly identified. Sensitivity is assumed to be 100 percent. The four lines correspond to four populations that might be screened, each of which has a different prevalence of HIV infection. The boldface line represents low-prevalence population such as those in which screening has recently been proposed. Source: NEJM (23 Juillet 1987, Vol. 317)

Chapitre 6

Variables aléatoires

Dans de nombreuses expériences les observations sont des nombres (taille, poids, etc.) D'autre part, lorsque le résultat d'une expérience n'est pas numérique, il convient souvent d'y associer un nombre. Cette association est une "variable aléatoire". Ce chapitre introduit les outils de calcul des probabilités avec les variables aléatoires. Ces outils se fondent sur les concepts introduits au Chapitre 4.

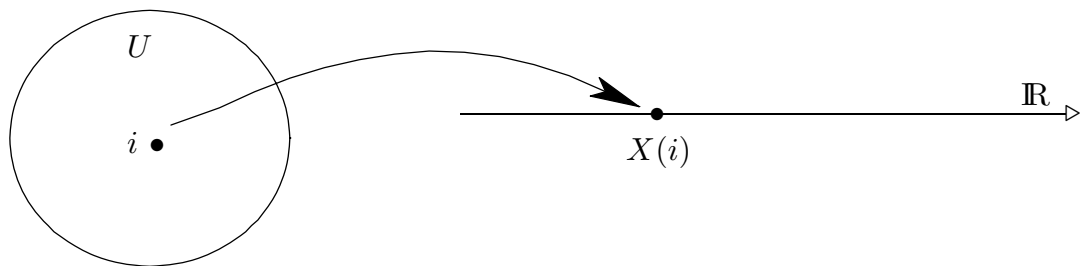
6.1 Le concept de variable aléatoire

Une variable aléatoire est un nombre réel associé au résultat d'une épreuve, donc un nombre aléatoire. Si l'épreuve est répétée, ce nombre change en général.

Exemples

- la taille d'un individu extrait au hasard d'une population;
- le nombre de "faces" dans une série de 10 jets d'une monnaie.

De façon formelle, considérons une épreuve \mathcal{E} et soit U l'ensemble de ses issues (univers). A chaque issue $i \in U$, nous attachons un nombre réel $X(i) \in \mathbb{R}$. Nous considérons donc une fonction $X : U \rightarrow \mathbb{R}$. Une telle fonction $X : U \rightarrow \mathbb{R}$ est appelée une *variable aléatoire*.



Exemples

1. Jet d'une pièce de monnaie une fois; $U = \{\text{Pile}, \text{Face}\}$. Définissons

$$M(\text{Pile}) = 0; \quad M(\text{Face}) = 1.$$

2. Jet d'un dé une fois; $U = \{\text{Face 1}, \text{Face 2}, \text{Face 3}, \text{Face 4}, \text{Face 5}, \text{Face 6}\}$. Soit

$$D(\text{Face } i) = i, \quad i = 1, 2, 3, 4, 5, 6.$$

3. On tire au hasard un individu i d'une population. $U = \{i\}$ est donc la population. Soit

$$A(i) = \text{Age de } i \text{ en années}, \quad T(i) = \text{Taille de } i.$$

4. On tire au hasard un échantillon de taille n d'une population. On s'intéresse à la présence ou à l'absence d'un caractère A chez chaque individu observé. U est donc l'ensemble des 2^n suites de n lettres de type $i = (A, A, \dots, A)$, $i = (A, A, \dots, \bar{A})$, etc. (\bar{A} représente l'absence de A). Définissons:

$$K(i) = \text{Nombre de } A \text{ dans } i.$$

Comme pour les variables considérées aux chapitres 1 à 3, on parlera de variables aléatoires discrètes et de variables aléatoires continues, selon le nombre et la nature de leurs valeurs ou modalités. Les variables aléatoires seront notées par des lettres majuscules (X, Y, \dots) et leurs modalités par des lettres minuscules ($x_1, x_2, \dots, y_1, y_2, \dots$).

6.2 La distribution d'une variable aléatoire

Variables aléatoires discrètes

A chaque valeur de la variable M (section précédente) on peut associer une probabilité. Par exemple, si la monnaie est parfaitement symétrique, il est raisonnable de définir

$$P(M = 0) = 0.5, \quad P(M = 1) = 0.5.$$

De façon similaire, (si le dé est parfaitement équilibré) on posera

$$P(D = i) = 1/6, \quad i = 1, 2, 3, 4, 5, 6.$$

Comment associer une probabilité à chaque valeur de la variable A ? Si p_1 est la proportion d'individus tels que $A(i) = 1$, p_2 la proportion telle que $A(i) = 2$ etc., nous définissons

$$P(A = 1) = p_1, \quad P(A = 2) = p_2, \quad \dots,$$

Il faut donc estimer p_1, p_2, \dots à l'aide d'un sondage.

De manière générale, si une variable aléatoire discrète X peut prendre les valeurs x_1, x_2, \dots avec les probabilités respectives p_1, p_2, \dots , nous dirons que X a pour *distribution de probabilité* (ou *distribution*) l'ensemble des couples

$$(x_1, p_1), (x_2, p_2), \dots$$

Une distribution de probabilité a les propriétés suivantes:

- $p_i \geq 0$ pour tout i ;
- $\sum_i p_i = 1$.

On peut représenter graphiquement une distribution de probabilité à l'aide d'un diagramme en bâtons ou en colonnes (comme si c'était une distribution de fréquence).

Exemple. On jette une pièce de monnaie équilibrée cinq fois et on s'intéresse à la variable

$$K = \text{nombre de Pile.}$$

Pour calculer, par exemple, $p_2 = P(K = 2)$ on considère les issues de l'épreuve qui produisent $K = 2$:

PPFFF PFPPF PFFPF PFFFP FPPFF
FPFPF FPFFP FFPPF FFPFP FFFPP

où P indique Pile et F indique Face. On a donc

$$p_2 = P(K = 2) = 10/32,$$

car il y a 32 ($= 2^5$) issues possibles et équiprobables. En général, on trouve

$$p_k = P(K = k) = \binom{5}{k} \times \frac{1}{32} \quad \text{pour } k = 0, 1, \dots, 5$$

Ce résultat est démontré au Chapitre 7 (distribution binomiale $\mathcal{B}(n = 5, p = 1/2)$).

En pratique on souhaite calculer les probabilités d'événements plus complexes que les événements définis par des modalités simples. Par exemple, quelle est la probabilité de tirer un jeune âgé entre 20 et 22 ans ? A l'aide des règles du Chapitre 4, on trouve

$$P(A = 20 \text{ ou } A = 21 \text{ ou } A = 22) = P(A = 20) + P(A = 21) + P(A = 22).$$

En outre,

$$P(A \geq 100) = 1 - P(A < 100).$$

On applique donc les règles du calcul des probabilités à des événements complexes définis à l'aide de plusieurs modalités de A .

En général, pour un événement E défini comme un ensemble de modalités de X , on a

$$P(E) = \sum_{x_i \in E} P(X = x_i).$$

Variables aléatoires continues

L'exemple suivant montre que pour une variable aléatoire continue, la probabilité d'assumer une certaine valeur est 0, et ceci pour n'importe quelle valeur ! Toutefois, la probabilité que la valeur observée de cette variable se trouve dans un intervalle donné n'est en général pas nulle.

Exemple. Considérons une roulette avec 36 trous numérotés $1/36, 2/36, \dots, 36/36 = 1$ et soit H la variable aléatoire discrète qui donne le numéro du trou visité par la boule. Alors, $P(H = h) = 1/36$ pour $h = 1/36, \dots, 36/36$. Si le nombre de trous est 100, $P(H = h) = 1/100$ et si le nombre de trous tend à ∞ , $P(H = h) \rightarrow 0$. La variable X qui indique le nombre du trou atteint pour une roulette avec un nombre infini de trous est une variable continue et $P(X = x) = 0$ pour tout $x \in (0, 1]$. Toutefois, la probabilité que la boule visite un certain secteur de la roulette, par exemple le secteur $a < x \leq b$ ($0 < a < b \leq 1$) est égale à $b - a \neq 0$ (par exemple, $P(0 < X \leq 0.25) = 1/4$).

Evidemment, une roulette avec un nombre infini de trous n'existe pas. Ce n'est qu'une approximation mathématique d'une roulette avec un nombre de trous très élevé. Une telle description idéalisée d'un objet ou d'un phénomène réel est un *modèle mathématique*. En pratique, on utilise le concept mathématique de variable continue pour des quantités aléatoires avec un nombre très élevé de valeurs possibles (par exemple, taille, poids, revenu).

Pour calculer les probabilités afférentes à une variable continue on utilise sa densité, c'est à dire une fonction qui permet de calculer la probabilité que X soit dans un intervalle $(a, b]$. Plus précisément, la *densité de probabilité* (ou *densité*) de X est une fonction f_X telle que, $f_X(x) \geq 0$ pour tout x , $\int_{-\infty}^{+\infty} f_X(x) dx = 1$, et

$$P(a < X \leq b) = \int_a^b f_X(x) dx.$$

S'il n'y a pas de possibilité de confusion, on utilisera le symbole f à la place de f_X .

Exemples

1. La densité de la variable X dans l'exemple de la roulette avec une infinité de trous a comme densité

$$\begin{aligned} f(x) &= 0, & \text{si } x < 0 \text{ ou si } x > 1; \\ &= 1, & \text{si } 0 \leq x \leq 1. \end{aligned}$$

Si $0 < a < b \leq 1$,

$$P(a < X \leq b) = \int_a^b f(x)dx = \int_a^b 1 \cdot dx = b - a.$$

On dit que X a une *distribution uniforme* dans l'intervalle $(0, 1]$. Esquissez le graphique de $f(x)$.

2. Un arrêt de bus est desservi toutes les 10 minutes. Soit X la variable aléatoire indiquant le temps d'attente (en minutes) jusqu'au passage du prochain bus lorsqu'on se rend à cet arrêt sans tenir compte de l'horaire. Il est parfois raisonnable d'admettre que X a pour densité la fonction

$$\begin{aligned} f(x) &= 0, & \text{si } x < 0 \text{ ou si } x > 10; \\ &= \frac{1}{10}, & \text{si } 0 \leq x \leq 10. \end{aligned}$$

On a donc, par exemple,

$$\begin{aligned} P(6 < X \leq 10) &= \int_6^{10} \frac{1}{10} dx = \frac{4}{10}; \\ P(10 < X < \infty) &= 0. \end{aligned}$$

Notez que $P(6 < X \leq 10)$ est la surface sous la densité $f(x)$ pour $x \in (6, 10]$. On dit que X a une *densité uniforme* dans $(0, 10]$. Esquissez le graphique de $f(x)$.

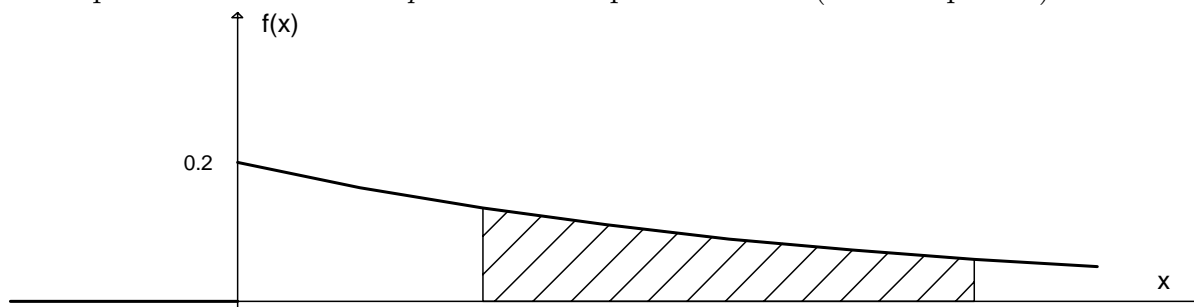
3. Soit X la variable aléatoire indiquant la "durée de vie" d'un modèle de voiture (exprimée en années). Supposons que sa densité est

$$\begin{aligned} f(x) &= 0, & \text{si } x < 0; \\ &= 0.2 \exp(-0.2x), & \text{si } x \geq 0. \end{aligned}$$

On trouve, par exemple,

$$P(2 < X \leq 6) = \int_2^6 0.2 \exp(-0.2x) dx \approx 0.369.$$

On dit que X a une *densité exponentielle* de paramètre 0.2 (voir Chapitre 7).



Densité de la durée de vie X et $P(2 < X \leq 6)$ (hachuré)

La fonction de distribution cumulative

En général, pour exprimer toutes les probabilités associées à une variable aléatoire discrète ou continue il suffit de déterminer les probabilités des intervalles de la forme $I = (-\infty, x)$, où x est un nombre réel. L'outil fondamental qui exprime ces probabilités est la fonction de distribution cumulative. La *fonction de distribution cumulative* d'une variable aléatoire X est la fonction

$$F_X(x) = P(X \leq x)$$

= probabilité de l'événement que X soit plus petit ou égal à x .

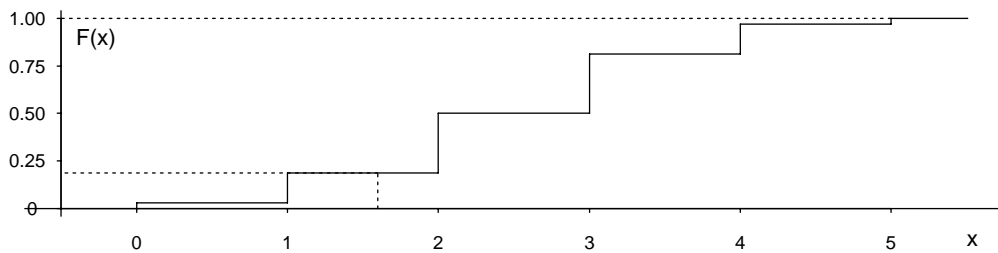
Cette fonction est définie pour tout x réel et prend des valeurs entre 0 et 1. S'il n'y a pas de confusion, on utilise le symbole F à la place de F_X .

Exemples

1. On jette une pièce de monnaie équilibrée cinq fois et on s'intéresse à la variable $K =$ nombre de Pile. Alors, $P(K = k) = \binom{5}{k}/32$ pour $k = 0, 1, \dots, 5$. La fonction de distribution cumulative est

$$F(x) = \sum_{k \leq x} P(K = k),$$

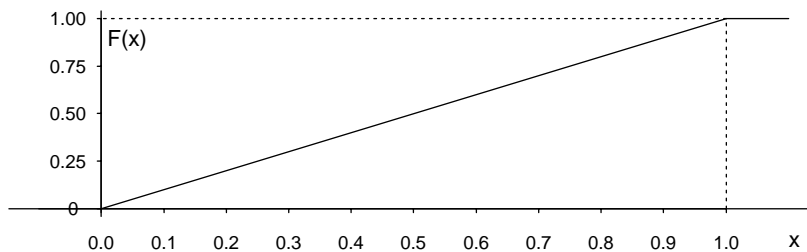
par exemple, $F(1.6) = P(K = 0) + P(K = 1)$. La figure montre cette fonction: elle est discontinue (en escalier; les traits verticaux ont une raison esthétique).



2. On choisit au hasard un nombre réel dans l'intervalle $[0, 1]$. Soit X le résultat de cette épreuve. (X est le trou visité par la boule d'une roulette avec un nombre infini de trous; voir exemple précédent.) On a:

$$F(x) = P(X \leq x) = \begin{cases} 0, & \text{si } x < 0; \\ x, & \text{si } 0 \leq x \leq 1; \\ 1, & \text{si } x > 1. \end{cases}$$

La figure montre cette fonction; elle est continue.

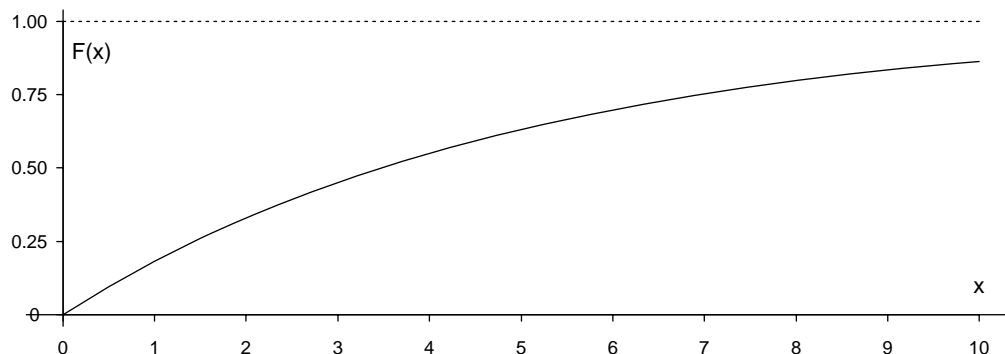


En général, une fonction de distribution cumulative quelconque a les propriétés suivantes:

- elle est non décroissante;
- elle prend des valeurs entre 0 et 1;
- elle tend vers 0 si x tend vers $-\infty$ et vers 1 si x tend vers $+\infty$.

La fonction de distribution cumulative d'une variable continue est continue. La fonction de distribution cumulative d'une variable discrète est une fonction discontinue en escalier.

Exemple. Distribution cumulative de la durée de vie:



En outre:

- A l'aide de la fonction de distribution cumulative, on peut calculer la probabilité des événements du type $\{a < X \leq b\}$:

$$P(a < X \leq b) = F(b) - F(a).$$

En effet,

$$\begin{aligned} F(b) &= P(X \leq b) = P(a < X \leq b \text{ ou } X \leq a) \\ &= P(a < X \leq b) + P(X \leq a) = P(a < X \leq b) + F(a). \end{aligned}$$

- Il y a la relation suivante entre la distribution de probabilité $(x_1, p_1), (x_2, p_2), \dots$ et la fonction de distribution cumulative F d'une variable discrète:

$$F(x) = \sum_{x_i \leq x} p_i.$$

- Il y a les relations suivantes entre la fonction de densité f et la fonction de distribution cumulative F d'une variable continue:

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t) dt, \\ f(x) &= \frac{d}{dx} F(x) \quad (\text{si } F \text{ est dérivable en } x). \end{aligned}$$

- Pour une variable aléatoire continue $P(X \leq x) = P(X < x)$.

6.3 Espérance ou moyenne d'une variable aléatoire

Souvent, il suffit d'avoir quelques nombres caractérisant la distribution au lieu de la distribution complète. Les mesures les plus fréquemment utilisées sont la moyenne (ou espérance), la variance, l'écart type et les percentiles.

Considérons d'abord une variable aléatoire discrète comme l'âge A (ans) d'un individu tiré au hasard d'une population. Supposons que $A \leq 100$. Pour un échantillon de taille n , soit n_i la fréquence absolue de $A = i$ ($i = 1, \dots, 100$). La moyenne est

$$\begin{aligned} m(A) &= \frac{1}{n} (n_1 \cdot 1 + n_2 \cdot 2 + \dots + n_{100} \cdot 100) \\ &= f_1 \cdot 1 + f_2 \cdot 2 + \dots + f_{100} \cdot 100, \end{aligned}$$

où les $f_i = n_i/n$ indiquent les fréquences relatives des âges observés. Si la taille de l'échantillon s'approche de celle de la population, les f_i s'approchent des probabilités p_i des modalités de A et $m(A)$ devient la moyenne de population

$$\mu(A) = p_1 \cdot 1 + p_2 \cdot 2 + \dots + p_{100} \cdot 100.$$

On appelle $\mu(A)$ la moyenne de population ou espérance de A .

En général, soit X une variable aléatoire discrète avec distribution (x_i, p_i) , $i = 1, 2, \dots$. L'*espérance mathématique* (ou *espérance*) ou *moyenne de population* (ou *moyenne*) de X est:

$$\mu(X) = x_1 p_1 + x_2 p_2 + \dots = \sum_i x_i p_i.$$

On utilise aussi le symbole $E(X)$ à la place de $\mu(X)$.

Considérons maintenant une variable continue comme la taille T d'un individu tiré au hasard. Partageons le domaine des valeurs de T en k intervalles de centres t_i ($i = 1, \dots, k$), bornes a_i, b_i et longueur $\Delta = b_i - a_i$. Soit f_i la fréquence relative des observations d'un échantillon de taille n dans l'intervalle i . Alors, la moyenne de T est approximativement

$$m(T) \approx \sum_{i=1}^k t_i f_i.$$

La probabilité que T soit dans l'intervalle i est $p_i = \int_{a_i}^{b_i} f(t) dt \approx f(t_i) \Delta$. Si n tend à la taille de l'univers (∞ pour une variable continue), les f_i s'approchent des p_i . Dans ce cas, si on prend des intervalles de plus en plus petits, la moyenne $m(T)$ s'approche de la moyenne de population

$$\mu(T) = \lim \sum t_i f(t_i) \Delta = \int_{-\infty}^{+\infty} t f(t) dt.$$

En général, soit X une variable aléatoire continue avec densité f . L'*espérance (mathématique)* ou *moyenne (de population)* de X est:

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

On utilise aussi le symbole $E(X)$ à la place de $\mu(X)$.

Exemple. Soit

$$\begin{aligned} f(x) &= 0, & \text{si } x < 0; \\ &= 1, & \text{si } 0 \leq x \leq 1; \\ &= 0, & \text{si } 1 < x. \end{aligned}$$

Alors

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 xdx = 1/2.$$

Propriétés

- Une propriété de grande importance concerne l'espérance de la somme de deux variables aléatoires. Soient X et Y ces deux variables; alors

$$E(X + Y) = E(X) + E(Y).$$

Plus généralement, si a et b sont des constantes,

$$\mu(aX + bY) = a \mu(X) + b \mu(Y).$$

- Un autre propriété très utile concerne l'espérance d'une transformation d'une variable aléatoire. Soit g une fonction réelle quelconque (par exemple, $g(x) = \ln(x)$). La variable $Y = g(X)$ est une *transformation* de Y . Son espérance est

$$E(Y) = E(g(X)) = \sum_i g(x_i)P(X = x_i)$$

dans le cas discret et

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

dans le cas continu. Le cas particulier $g(t) = a + bt$, où a et b sont des constantes réelles, donne

$$E(a + bX) = a + bE(X).$$

6.4 La variance et l'écart type d'une variable aléatoire

L'espérance d'une variable donne une idée de la valeur moyenne de cette variable mais ne prend pas en compte d'autres aspects importantes. Par exemple, les variables

$$\begin{aligned} X_1 & \text{ avec distribution uniforme dans l'intervalle } [-1, 1], \\ X_2 & \text{ avec distribution uniforme dans l'intervalle } [-1000, 1000], \end{aligned}$$

ont toutes deux une espérance égale à 0 mais une variabilité très différente. Pour mesurer cet aspect on utilise la variance et l'écart type.

Soit X une variable aléatoire. La *variance (de population)* de X est définie comme:

$$\sigma^2(X) = E([X - \mu(X)]^2).$$

Donc,

$$\begin{aligned} \sigma^2(X) &= \sum_i [x_i - \mu(X)]^2 P(X = x_i) && \text{si } X \text{ est discrète,} \\ \sigma^2(X) &= \int_{-\infty}^{+\infty} [x - \mu(X)]^2 f_X(x) dx && \text{si } X \text{ est continue.} \end{aligned}$$

L'*écart type (de population)* de X est défini comme

$$\sigma(X) = \sqrt{\sigma^2(X)}.$$

Exemples

1. Soit X égale au résultat du jet d'un dé. Il s'agit d'une variable discrète. Le calcul de la variance peut s'effectuer par le tableau suivant:

x_i	$p_i = P(X = x_i)$	$x_i p_i$	$[x_i - \mu(X)]$	$[x_i - \mu(X)]^2 p_i$
1	1/6	1/6	-2.5	6.25/6
2	1/6	2/6	-1.5	2.25/6
3	1/6	3/6	-0.5	0.25/6
4	1/6	4/6	+0.5	0.25/6
5	1/6	5/6	+1.5	2.25/6
6	1/6	6/6	+2.5	6.25/6
Σ		21/6		17.5/6

On obtient $\mu(X) = 21/6$ et $\sigma^2(X) = 17.5/6$.

2. Si X a une distribution uniforme dans $[0, 1]$ on a $\mu(X) = 0.5$ et

$$\begin{aligned} \sigma^2(X) &= \int_0^1 (x - 0.5)^2 f(x) dx = \int_0^1 (x^2 - x + 1/4) dx \\ &= [x^3/3 - x^2/2 + x/4]_0^1 = 1/3 - 1/2 + 1/4 = 1/12. \end{aligned}$$

Propriétés

- La variance est un nombre positif. Elle n'est égale à zéro que si X est constante.
- Soit $Y = a + bX$. On démontre facilement que:

$$\sigma^2(Y) = \sigma^2(a + bX) = b^2\sigma^2(X).$$

Remarques

1. Le calcul de la variance se simplifie parfois à l'aide de la formule suivante:

$$\sigma^2(X) = E(X^2) - (E(X))^2.$$

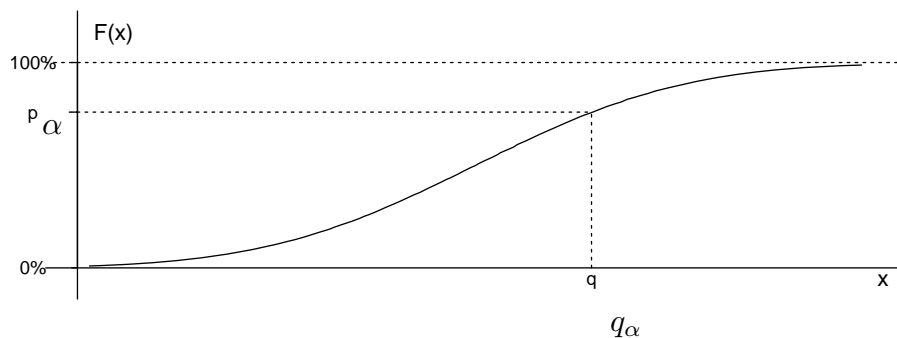
2. $E(X)$ est aussi appelé le *premier moment de X* ; $E(X^2)$ est appelé le *deuxième moment de X*.

6.5 Quantiles d'une variable aléatoire

Le *quantile (d'ordre) α* ($0 < \alpha < 1$) d'une variable aléatoire continue X avec fonction de distribution cumulative F est le nombre q_α tel que

$$F(q_\alpha) = \alpha, \quad \text{c'est-à-dire,} \quad q_\alpha = F^{-1}(\alpha).$$

Analoguement on définit des *percentiles*, des *quartiles* et des *déciles* de population. Le percentile 50% est la *médiane (de population)*, etc. Pour une variables discrète on procède comme dans le cas d'une distribution empirique (Chapitre 2).

*Exemple*

La médiane d'une distribution uniforme dans l'intervalle $[0, 1]$ vaut $1/2$.

6.6 Variables aléatoires bidimensionnelles et leur distributions

Dans beaucoup de situations nous nous intéressons à deux (ou à plusieurs) caractéristiques numériques des unités d'observation; ces caractéristiques sont observées simultanément.

Exemple. Pour un individu tiré au hasard d'une population nous nous intéressons à son âge A codé en trois classes (1 = jeune, 2 = âge moyen, 3 = senior) ainsi qu'à son état social S codé en trois classes (1 = bas, 2 = moyen, 3 = haut). A et S sont deux variables aléatoires et, à l'aide d'un sondage, nous pouvons déterminer les probabilités conjointes

$$P(A = i \text{ et } S = j) \quad \text{pour } i = 1, 2, 3 \text{ et } j = 1, 2, 3.$$

Supposons qu'on ait obtenu la table suivante:

	$S = 1$	$S = 2$	$S = 3$	Total
$A = 1$	0.10	0.20	0.20	0.50
$A = 2$	0.04	0.08	0.08	0.20
$A = 3$	0.06	0.12	0.12	0.30
Total	0.20	0.40	0.40	1.00

Nous dirons que (A, S) est une variable aléatoire bidimensionnelle discrète et que les nombres dans la table définissent la distribution de probabilité conjointe de (A, S) .

En général, une *variable aléatoire à deux dimensions* ou *bidimensionnelle* (X, Y) est un couple de variables aléatoires monodimensionnelles X et Y . Si l'ensemble des valeurs possibles de (X, Y) est fini ou dénombrable, on dit que (X, Y) est une *variable aléatoire discrète à deux dimensions*. Si (X, Y) peut assumer toutes les valeurs d'un rectangle $\{(x, y) \mid a < x \leq b, c < y \leq d\}$ dans \mathbb{R}^2 on dit que (X, Y) est une *variable aléatoire continue à deux dimensions*. Il est aussi possible de combiner une variable discrète avec une variable continue, mais nous nous limitons à considérer les cas mentionnés ci-dessus.

Variables aléatoires bidimensionnelles discrètes

Soient (X, Y) une variable aléatoire discrète avec valeurs (x_i, y_j) ($i = 1, 2, \dots, j = 1, 2, \dots$) et $p_{ij} = P(X = x_i \text{ et } Y = y_j)$ leurs probabilités. L'ensemble des triades (x_i, y_j, p_{ij}) est la *distribution (de probabilité) conjointe* de (X, Y) . Elle satisfait les conditions suivantes:

- $p_{ij} \geq 0$ pour tous les (x_i, y_i) ,
- $\sum_j \sum_i p_{ij} = 1$,
- Pour tout événement E constitué de couples de modalités de (X, Y) :

$$P(E) = \sum_{(x_i, y_j) \in E} p_{ij}.$$

Exemple

Considérons les variables (A, S) et soit $E = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$. Alors

$$P(E) = 0.10 + 0.04 + 0.20 + 0.08 = 0.42.$$

Nous écrirons $P(X = x_i, Y = x_j)$ à la place de $P(X = x_i \text{ et } Y = x_j)$.

Soit (X, Y) une variable aléatoire discrète et (x_i, y_j, p_{ij}) sa distribution. Soit

$$\begin{aligned} g_i &= P(X = x_i) = P(X = x_i, \{Y = y_1 \text{ ou } Y = y_2, \text{ ou } \dots\}) \\ &= P(X = x_i, Y = y_1) + P(X = x_i, Y = y_2) + \dots \\ &= \sum_j p_{ij}, \quad i = 1, 2, \dots \end{aligned}$$

Les paires $(x_1, g_1), (x_2, g_2), \dots$, définissent la *distribution (de probabilité) marginale* de X . Par analogie on définit aussi la distribution marginale h_j ($j = 1, 2, \dots$) de Y .

Exemple

La distribution marginale de A est

$$P(A = 1) = 0.50, \quad P(A = 2) = 0.20, \quad P(A = 3) = 0.30.$$

La *distribution (de probabilité) conditionnelle* de X pour $Y = y_j$ donné est définie par

$$g_{i|j} = P(X = x_i | Y = y_j) = \frac{p_{ij}}{h_j}, \quad i = 1, 2, \dots, \quad j \text{ fixé,}$$

tandis que la distribution conditionnelle de Y pour $X = x_i$ donné est définie par

$$h_{j|i} = P(Y = y_j | X = x_i) = \frac{p_{ij}}{g_i}, \quad j = 1, 2, \dots, \quad i \text{ fixé.}$$

Exemple

La distribution conditionnelle de S pour $A = 2$ donné est

$$P(S = 1 | A = 2) = 1/5 \quad P(S = 2 | A = 2) = 2/5 \quad P(S = 3 | A = 2) = 2/5.$$

La fonction

$$F(x, y) = P(X \leq x, Y \leq y)$$

est appelée *fonction de distribution cumulative* de (X, Y) .

Exemple

$$\begin{aligned} F(0.5, 0.7) &= P(A \leq 0.5 \text{ et } \leq 0.7) = 0.0, \\ F(1.5, 1.2) &= P(A \leq 1.5 \text{ et } \leq 1.2) = 0.1, \\ F(2.5, 3.4) &= P(A \leq 2.5 \text{ et } \leq 3.4) = 0.7, \quad \text{etc.} \end{aligned}$$

Variables aléatoires bidimensionnelles continues

Soit (X, Y) une variable aléatoire continue. La *fonction de densité conjointe* de X et Y est une fonction $f(x, y)$ qui satisfait les conditions suivantes:

- $f(x, y) \geq 0$ pour tous les $(x, y) \in \mathbb{R}^2$,
- $\int \int_{\mathbb{R}^2} f(x, y) dx dy = 1$,
- Pour tout événement représenté par un rectangle $E = \{(x, y) \mid a < x \leq b, c < y \leq d\}$ en \mathbb{R}^2 ,

$$P(E) = \int \int_E f(x, y) dx dy.$$

Soit (X, Y) une variable aléatoire continue et $f(x, y)$ sa densité. Les fonctions

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

sont les *densités marginales* de X et de Y .

La *densité conditionnelle* de X pour $Y = y$ donné (et tel que $h(y) > 0$) est définie par

$$g(x \mid y) = \frac{f(x, y)}{h(y)}.$$

La densité conditionnelle de Y pour $X = x$ donné (et tel que $g(x) > 0$) est définie par

$$h(y \mid x) = \frac{f(x, y)}{g(x)}, \quad g(x) > 0.$$

La fonction

$$F(x, y) = P(X \leq x, Y \leq y)$$

est appelée *fonction de distribution cumulative* de (X, Y) . On a:

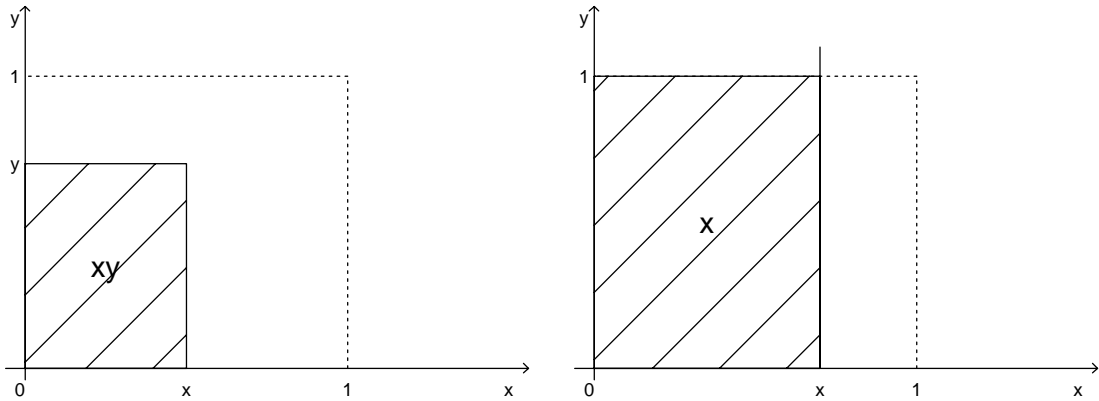
$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)$$

en tout point (x, y) où F est dérivable et

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) ds dt.$$

Exemple

On sélectionne aléatoirement un point dans le carré $[0, 1] \times [0, 1] \subset \mathbb{R}^2$. Soient X et Y les coordonnées du point.



Si $0 \leq x \leq 1$ et $0 \leq y \leq 1$, la probabilité que le point (X, Y) satisfait $X \leq x$ et $Y \leq y$ est égale à xy . Si $0 \leq x \leq 1$ et $1 < y$, la probabilité vaut x . Les autres cas sont traités de manière analogue et on trouve la fonction de distribution cumulative

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

La densité conjointe vaut donc:

$$f(x, y) = \begin{cases} 1, & \text{si } 0 \leq x \leq 1 \text{ et } 0 \leq y \leq 1; \\ 0, & \text{sinon.} \end{cases}$$

La densité marginale de X se trouve de la façon suivante:

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 f(x, y) dy = \int_0^1 1 dy = 1 \quad \text{si } 0 \leq x \leq 1,$$

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 f(x, y) dy = \int_0^1 0 dy = 0 \quad \text{sinon.}$$

De manière analogue on trouve la densité marginale de Y .

Pour $0 \leq x \leq 1$, la densité conditionnelle de Y pour x donné est

$$h(y | x) = \begin{cases} 1, & \text{si } 0 \leq y \leq 1; \\ 0, & \text{sinon.} \end{cases}$$

6.7 Variables aléatoires indépendantes

Soit (X, Y) une variable aléatoire discrète à deux dimensions. Soit (x_i, g_i) la distribution marginale de X , (y_j, h_j) la distribution marginale de Y et (x_i, y_j, p_{ij}) la distribution conjointe de (X, Y) . On dit que X et Y sont des *variables aléatoires indépendantes* si et seulement si toutes les probabilités conjointes p_{ij} sont le produit des probabilités marginales g_i et h_j correspondantes. En d'autres termes, $p_{ij} = g_i h_j$ pour toutes les paires (x_i, y_j) .

Exemple

Les variables A et S sont indépendantes car

$$P(A = 1, S = 1) = 0.5 \cdot 0.2, \quad P(A = 1, S = 2) = 0.4 \cdot 0.5, \dots, \quad P(A = 3, S = 3) = 0.4 \cdot 0.3.$$

Soit (X, Y) une variable aléatoire continue à deux dimensions avec densité conjointe $f(x, y)$ et densités marginales $g(x)$ et $h(y)$. On dit que X et Y sont des *variables aléatoires indépendantes* si (et seulement si) $f(x, y) = g(x)h(y)$ pour toutes les paires (x, y) .

Exemple

Les coordonnées X et Y du point sélectionné aléatoirement dans le carré $[0, 1] \times [0, 1] \subset \mathbb{R}^2$ sont indépendantes.

On démontre que:

Deux variables X et Y discrètes sont indépendantes si et seulement si la distribution conditionnelle de X donné $Y = y_j$ ne dépend pas de y_j et donc $g_{i|j} = g_i$ pour toutes les paires (x_i, y_j) . (De façon équivalente, $h_{j|i} = h_j$ pour toutes les paires (x_i, y_j)).

Deux variables X et Y continues sont indépendantes si et seulement si $g(x|y) = g(x)$ pour tout (x, y) . (De façon équivalente, $h(y|x) = h(y)$ pour tout (x, y)).

6.8 Covariance, corrélation et variance de la somme de deux variables

La *covariance* $V(X, Y)$ de X et Y est définie par

$$V(X, Y) = E([X - \mu(X)][Y - \mu(Y)]).$$

On utilise aussi le symbole $\sigma(X, Y)$ à la place de $V(X, Y)$.

La *corrélation* $\rho(X, Y)$ de X et Y est définie par

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

La covariance $\sigma(X, Y)$ et la corrélation $\rho(X, Y)$ ont des propriétés similaires aux coefficients de covariance et de corrélation empiriques (des données) $s(X, Y)$ et $r(X, Y)$ introduits au Chapitre 3. En particulier on remarquera que, si a et b sont des constantes,

$$\sigma^2(aX + bY) = a^2\sigma^2(X) + b^2\sigma^2(Y) + 2ab\sigma(X, Y).$$

En outre, si X et Y sont indépendantes $\sigma(X, Y) = 0$. Donc, la variance de la somme de deux variables indépendantes est la somme de leurs variances.

Pour le calcul il est utile de savoir que

$$V(X, Y) = E(XY) - E(X)E(Y).$$

Exemple

Considérons deux variables A et S quantitatives avec la même distribution que les variables de nature qualitative A et S introduites à la Section 6.6. On obtient:

$$\mu(A) = 0.50 \cdot 1 + 0.20 \cdot 2 + 0.30 \cdot 3 = 1.80,$$

$$\mu(S) = 0.20 \cdot 1 + 0.40 \cdot 2 + 0.40 \cdot 3 = 2.20.$$

La table ci-dessous donne les produits $(i - \mu(A))(j - \mu(S))$ pour $i = 1, 2, 3$, $j = 1, 2, 3$:

	$1 - \mu(S) = -1.2$	$2 - \mu(S) = -0.2$	$3 - \mu(S) = +0.8$
$1 - \mu(A) = -0.8$	+0.96	+0.16	-0.64
$2 - \mu(A) = +0.2$	-0.24	-0.04	+0.16
$3 - \mu(A) = +1.2$	-1.44	-0.24	+0.96

Les produits $p_{ij}(i - \mu(A))(j - \mu(S))$ correspondants sont

$$\begin{array}{lll} +0.0960 & +0.0320 & -0.1280 \\ -0.0096 & -0.0032 & +0.0128 \\ -0.0864 & -0.0288 & +0.1152 \end{array}$$

et leur somme est $\sigma(A, S) = 0$. Ce résultat ne nous surprend pas, car nous avons déjà remarqué que A et S sont indépendantes.

6.9 Variables aléatoires à plusieurs dimensions

Par analogie, il est possible de définir des variables aléatoires à 3, à 4 ... à n dimensions. Les définitions de distribution conjointe, marginale et conditionnelle peuvent être étendues, ainsi que celles de variables aléatoires indépendantes. Par exemple, les composantes d'un vecteur de variables aléatoires continues (X_1, X_2, \dots, X_n) avec densité conjointe $f(x_1, x_2, \dots, x_n)$ sont indépendantes si et seulement si f peut être factorisée,

$$f(x_1, x_2, \dots, x_n) = g_1(x_1)g_2(x_2) \cdot \dots \cdot g_n(x_n),$$

où g_1, g_2, \dots, g_n sont les densités marginales.

Si X_1, X_2, \dots, X_n sont indépendantes,

$$\sigma^2(X_1 + X_2 + \dots + X_n) = \sigma^2(X_1) + \sigma^2(x_2) + \dots + \sigma^2(X_n).$$

Compléments

1. Distribution d'une transformation d'une variable aléatoire

Il arrive souvent qu'une variable X soit facile à définir dans un certain contexte, mais que la variable importante soit une transformation de X .

Exemple

Les coûts des séjours hospitaliers pour un certain traitement ont souvent une distribution asymétrique tandis que leurs logarithmes ont une distribution symétrique. En statistique, il est généralement plus facile de travailler avec des distributions symétriques qu'avec des distributions asymétriques. Pour cette raison on utilise parfois le modèle mathématique symétrique

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

(qui s'appelle *densité de Gauss*; esquissez son graphique) pour décrire de façon approximative mais simple la distribution du logarithme du coût. [Les paramètres μ et σ assument des valeurs spécifiques en relation avec les données traitées: voir chapitres suivants.] Quelle est le modèle correspondant pour la distribution du coût ? Pour répondre à cette question supposons que X soit la variable aléatoire, avec densité $f_X(x)$, qui donne le logarithme du coût d'un séjour pris de façon aléatoire. Le coût est alors une variable aléatoire Y telle que $Y = \exp(X)$. On trouve d'abord la fonction de distribution cumulative $F_Y(y)$ de Y :

$$F_Y(y) = P(Y \leq y) = P(\exp(X) \leq y) = P(X \leq \ln(y)) = F_X(\ln(y)) = \int_{-\infty}^{\ln(y)} f_X(x) dx.$$

Pour trouver la densité $f_Y(y)$ de Y on dérive F_Y :

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(\ln(y)) \frac{1}{y} = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(\frac{-(\ln(y) - \mu)^2}{2\sigma^2}\right)$$

qui s'appelle *densité lognormale*. (Esquissez son graphique).

En général, soit $g(x)$ une fonction dérivable et croissante (par exemple, $g(x) = \exp(x)$) et soit X une variable aléatoire. On veut calculer la distribution de la *transformation* ou *transformée* $Y = g(X)$ à partir de celle de X . On trouve d'abord la fonction de distribution cumulative

$$F_Y(y) = P(Y \leq y).$$

Par définition de Y on a

$$F_Y(y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

La densité de la variable Y est donc:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y).$$

Ainsi on trouve le résultat

$$F_Y(y) = F_X(g^{-1}(y)), \text{ et} \\ f_Y(y) = f_X(g^{-1}(y)) / (g'(g^{-1}(y))).$$

Un calcul analogue montre que si g est décroissante:

$$F_Y(y) = 1 - F_X(g^{-1}(y)), \text{ et}$$

$$f_Y(y) = f_X(g^{-1}(y)) / (-g'(g^{-1}(y))).$$

Exemples

1. Soit $Y = g(X) = X + a$ une *translation* Translation (ou *transformation de position*) par la constante a . Les formules donnent

$$F_Y(y) = F_X(g^{-1}(y)) = F_X(y - a), \text{ et}$$

$$f_Y(y) = f_X(y - a)$$

car $y = g(g^{-1}(y)) = g^{-1}(y) + a$, c'est-à-dire, $g^{-1}(y) = y - a$.

2. Soit $Y = g(X) = bX$ avec $b > 0$ une constante positive (*transformation d'échelle*). Alors

$$F_Y(y) = F_X(g^{-1}(y)) = F_X\left(\frac{y}{b}\right), \text{ et}$$

$$f_Y(y) = \frac{1}{b} f_X\left(\frac{y}{b}\right),$$

car $y = g(g^{-1}(y)) = bg^{-1}(y)$, c'est-à-dire, $g^{-1}(y) = y/b$.

3. Soit X une variable avec densité

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

et soit $Y = \mu + \sigma X$ avec μ réel et σ positif et réel. Alors, pour Y :

$$F_Y(y) = F_X(g^{-1}(y)) = F_X\left(\frac{y - \mu}{\sigma}\right)$$

$$f_Y(y) = \frac{1}{\sigma} f_X\left(\frac{y - \mu}{\sigma}\right) = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y - \mu}{\sigma}\right)^2\right).$$

Une formule approximative pour l'espérance et la variance d'une transformée

Soit $Y = g(X)$ une transformée d'une variable aléatoire X . Nous savons déjà que

$$E(Y) = E(g(X)) = \int g(x) f_X(x) dx.$$

Mais, cette intégrale peut être difficile à évaluer. Le développement de Taylor d'ordre 1 donne une formule approximative et utile dans de nombreuses applications:

$$\begin{aligned} Y = g(X) &= g(E(X) + [X - E(X)]) \\ &\approx g(E(X)) + g'(E(X))(X - E(X)). \end{aligned}$$

En utilisant cette approximation on trouve

$$\begin{aligned} \mu(Y) &\approx g(\mu(X)), \\ \sigma^2(Y) &\approx [g'(\mu(X))]^2 \cdot \sigma^2(X). \end{aligned}$$

La première formule est une conséquence de $E(X - E(X)) = E(X) - E(X) = 0$ et la seconde découle de $V(a + bX) = b^2V(X)$ (a, b constantes).

Exemple

Soit X une variable aléatoire ayant comme densité

$$f_X(x) = 30x^2(1-x)^2, \quad (0 < x < 1).$$

Il est aisé de calculer l'espérance et la variance de X . On obtient $\mu(X) = 1/2$, $\sigma^2(X) = 1/28$. Soit $Y = \ln(X)$. Alors la densité de Y vaut

$$f_Y(y) = 30e^{3y}(1-e^y)^2 \quad (y < 0).$$

La moyenne et la variance de Y sont moins faciles à trouver. Par intégration numérique, on obtient $\mu(Y) = -0.7833$ et $\sigma^2(Y) = 0.2136$. Les approximations fournissent $\mu(Y) \approx \ln(1/2) = -0.69$ et $\sigma^2(Y) \approx 2^2 \cdot \sigma^2(X) = 0.14$.

Remarque. Ces approximations qui se basent sur un développement de Taylor d'ordre 1 sont bonnes dans le cas où la variance de X est petite. Si, par exemple, X est la moyenne de n variables aléatoires X_1, \dots, X_n indépendantes et identiquement distribuées, alors $\sigma^2(X) = \sigma^2(X_1 + \dots + X_n)/n = \sigma^2(X_1)/n$. Pour de grandes valeurs de n , cette variance est proche de 0 et les approximations sont bonnes.

2. Distribution de la somme de deux variables aléatoires

Soient X et Y deux variables discrètes prenant les valeurs $0, 1, \dots, n$. Nous cherchons la distribution de la somme $Z = X + Y$. Par définition nous avons:

$$\begin{aligned} p_z &= P(Z = z) = P(X + Y = z) \\ &= P(X = z \text{ et } Y = 0) + P(X = z - 1 \text{ et } Y = 1) + \dots + P(X = 0 \text{ et } Y = z) \end{aligned}$$

(pour $z = 0, 1, \dots, 2n$) car les événements $(X = z - k, Y = k)$ sont incompatibles pour différentes valeurs de k . On peut simplifier cette formule si on admet que les deux variables sont indépendantes. Dans ce cas, on obtient:

$$p_z = P(X = z)P(Y = 0) + P(X = z - 1)P(Y = 1) + \dots + P(X = 0)P(Y = z).$$

Exemple

Soient X et Y deux variables avec distributions binomiales $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$ (voir Chapitre 7). Quelle est la distribution de $X + Y$? Etant donné que X et Y sont indépendantes, on obtient pour $z = 0, \dots, n + m$:

$$\begin{aligned} P(X + Y = z) &= \binom{n}{z} p^z (1 - p)^{n-z} \cdot \binom{m}{0} p^0 (1 - p)^m + \\ &\quad \binom{n}{z-1} p^{z-1} (1 - p)^{n-(z-1)} \cdot \binom{m}{1} p^1 (1 - p)^{m-1} + \\ &\quad + \dots + \binom{n}{0} p^0 (1 - p)^n \binom{m}{z} p^1 (1 - p)^{m-z}. \end{aligned}$$

On peut simplifier:

$$\begin{aligned} P(X + Y = z) &= p^z (1 - p)^{n+m-z} \left(\binom{n}{z} \binom{m}{0} + \binom{n}{z-1} \binom{m}{1} + \dots + \binom{n}{0} \binom{m}{z} \right) \\ &= p^z (1 - p)^{n+m-z} \binom{n+m}{z}. \end{aligned}$$

Il s'ensuit que $X + Y$ a une distribution binomiale $\mathcal{B}(n + m, p)$.

Dans le cas où X et Y sont deux variables indépendantes et continues ayant une densité conjointe $f_{X,Y}(x, y)$ on obtient le résultat suivant pour la distribution de la somme $Z = X + Y$:

$$\begin{aligned} F_Z(z) &= P(X + Y \leq z) = \int_{-\infty}^{\infty} P(X + Y \leq z \mid Y = y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} P(X \leq z - y \mid Y = y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{X|Y}(x \mid y) dx f_Y(y) dy, \end{aligned}$$

où $f_{X|Y}$ est la densité conditionnelle. La densité est donc

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X|Y}(z - y \mid y) f_Y(y) dy.$$

Dans le cas de l'indépendance de X et Y , ces formules se simplifient:

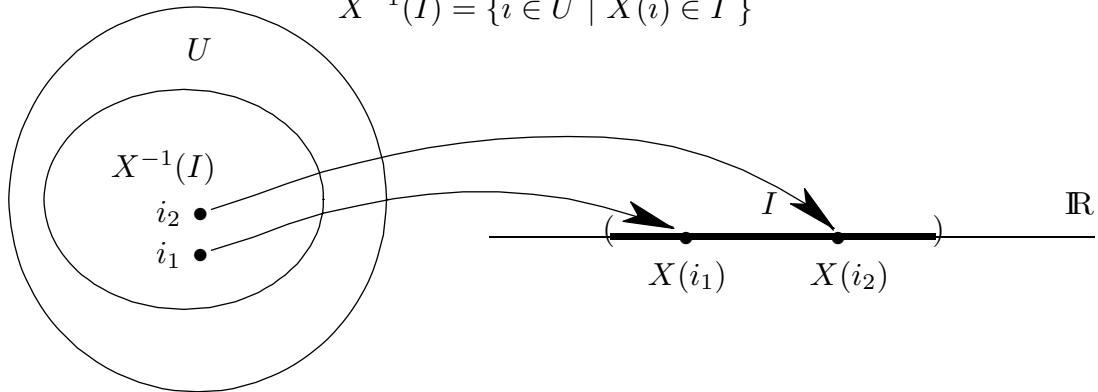
$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y) dy,$$

$$F_{X+Y}(z) = \int_{-\infty}^{\infty} F_X(z-y)f_Y(y) dy.$$

3. Probabilité induite

Si une probabilité P_U est définie pour les événements d'un univers U et X est une variable aléatoire, il est possible d'en déduire des probabilités pour des valeurs individuelles de X ou pour des ensembles de valeurs de X . Il convient de considérer les intervalles réels. Si I est un tel intervalle, on définit l'*image réciproque de I selon X* , comme l'ensemble

$$X^{-1}(I) = \{i \in U \mid X(i) \in I\}$$



Alors, si $X^{-1}(I)$ est un événement pour lequel une probabilité P_U est définie, il est naturel de définir la probabilité $P(I)$ comme

$$P(I) = P_U(X^{-1}(I)).$$

On dit que P est la *probabilité de $I \subset \mathbb{R}$ induite par X* .

Remarque. Si U est fini, l'ensemble des valeurs de X est fini. Un intervalle I contient un nombre fini de valeurs de X et $X^{-1}(I)$ est l'ensemble fini des issues i telles que $X(i) \in I$.

Chapitre 7

Modèles de distributions fréquemment utilisés

Ce chapitre définit brièvement les modèles de distributions univariées les plus fréquemment utilisés comme descriptions approximatives de distributions réelles. Comme ces modèles dépendent de *paramètres* qui doivent être déterminés à l'aide des données que l'on souhaite décrire (Chapitre 8) on les appelle des *modèles paramétriques*. Parmi ces modèles, celui de la distribution binomiale et celui de la distribution de Gauss sont fondamentales pour le développement des techniques d'inférence présentés dans les prochains chapitres.

7.1 La distribution binomiale

Nous considérons n répétitions indépendantes d'une épreuve \mathcal{E} qui ne peut avoir que deux issues, A et \bar{A} (non A). L'épreuve \mathcal{E} est parfois appelée une *épreuve de Bernoulli* dont les issues sont "succès" et "échec". Pour $n = 1$, l'espace des issues est $U = \{A, \bar{A}\}$. Soit p la probabilité de A : $p = P(A)$. Considérons la variable aléatoire X définie par:

$$X(A) = 1; \quad X(\bar{A}) = 0.$$

Sa distribution est

$$P(X = 0) = 1 - p; \quad P(X = 1) = p.$$

On dit que X a (ou "suit") une distribution binomiale de paramètres n et p , avec $n = 1$, et on écrit de façon abrégée: $X \sim \mathcal{B}(1, p)$. Si $n = 3$, l'espace des issues est

$$U = \{\bar{A}\bar{A}\bar{A}, A\bar{A}\bar{A}, \bar{A}A\bar{A}, \bar{A}\bar{A}A, AA\bar{A}, A\bar{A}A, \bar{A}AA, AAA\}.$$

Nous considérons la variable X définie par

$$\begin{aligned} X = 0 & \quad \text{si } \{\bar{A}\bar{A}\bar{A}\} \text{ a lieu;} \\ X = 1 & \quad \text{si } \{A\bar{A}\bar{A}, \bar{A}A\bar{A}, \bar{A}\bar{A}A\} \text{ a lieu;} \\ X = 2 & \quad \text{si } \{AA\bar{A}, A\bar{A}A, \bar{A}AA\} \text{ a lieu;} \\ X = 3 & \quad \text{si } \{AAA\} \text{ a lieu.} \end{aligned}$$

X est donc le nombre de A dans 3 répétitions de \mathcal{E} . Sa distribution est

$$P(X = 0) = (1-p)^3; \quad P(X = 1) = 3(1-p)^2p; \quad P(X = 2) = 3(1-p)p^2; \quad P(X = 3) = p^3.$$

Dans ce cas, X a une distribution binomiale $\mathcal{B}(n, p)$ avec $n = 3$ et on écrit $X \sim \mathcal{B}(3, p)$. En général, l'espace des issues $U = \{i\}$ est l'ensemble des 2^n séquences i du type

$$i = AA \dots AA, \quad i = AA \dots A\bar{A}, \quad i = AA \dots \bar{A}A, \dots$$

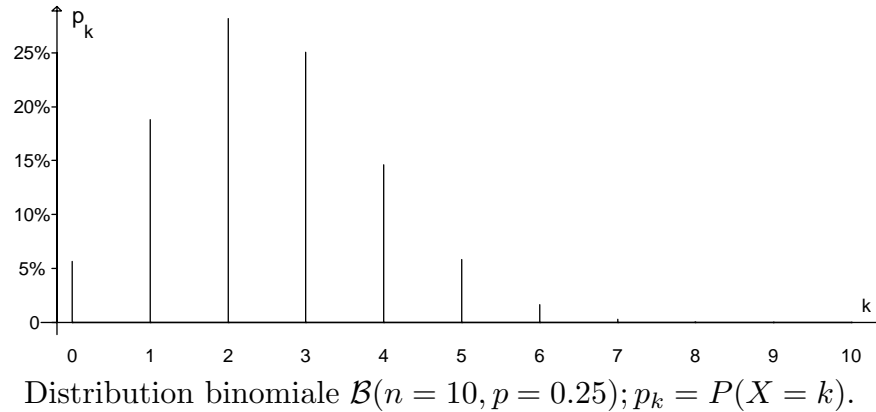
Chaque issue est une suite de n éléments et chaque élément est égal à A ou à \bar{A} . Soit

$$X(i) = \text{Nombre de } A \text{ dans l'issue } i.$$

Alors X a (ou "suit") une *distribution binomiale de paramètres n et p* , on écrit $X \sim \mathcal{B}(n, p)$, et on trouve que

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n. \quad (1)$$

Le symbole $\binom{n}{k}$ (" n sur k ") est un "coefficient binomial". Il représente le nombre de choix de k places (celle des A) parmi n places disponibles (les n éléments de la séquence). A l'aide du calcul combinatoire on trouve que $\binom{n}{k} = n! / [(n-k)!k!]$, où $k! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot k$ ("factoriel de k "). Par exemple, $\binom{3}{1} = 3! / [(3-1)!1!] = 3$. Par convention, $0! = 1$ et $\binom{n}{0} = 1$ pour tout $n > 0$.



Exemple 1. La proportion d'individus infectés par un virus dans une colonie d'insectes de taille $N = 1'000'000$ est $p = 10\%$. On prend un échantillon de taille $n = 10$. Quelle est la probabilité d'y trouver moins de 4 insectes infectés ?

Soit

$X =$ Nombre d'insectes infectés dans l'échantillon.

La probabilité que le premier insecte échantillonné soit infecté est p . Comme N est très grand par rapport à n , on peut supposer que la probabilité que le deuxième insecte échantillonné soit infecté est encore p . Il s'agit d'une approximation car, si un premier insecte infecté est tiré de la colonie, la probabilité que le deuxième soit infecté est légèrement inférieure à p . Pour simplifier, on supposera que la probabilité de tirer un insecte infecté reste constante. En utilisant la formule (1) avec $n = 10$ et $p = 0.1$, on trouve

$$\begin{aligned} P(X = 0) &= (0.9)^{10} = 0.349, & P(X = 1) &= 10 \cdot (0.9)^9 \cdot (0.1) = 0.387, \\ P(X = 2) &= 0.194, & P(X = 3) &= 0.05 \end{aligned}$$

et donc

$$P(X \leq 3) = 0.349 + 0.387 + 0.194 + 0.057 = 0.987.$$

Il est donc très improbable de trouver plus de trois insectes infectés dans l'échantillon, car $P(X > 3) = 1 - 0.987 = 0.013$.

Exemple 2. Quelle taille doit avoir l'échantillon de l'exemple 1 pour que la probabilité d'y trouver plus de 3 insectes infectés soit supérieure à 95% ?

Il faut résoudre l'inégalité

$$1 - P(X \leq 3) = 1 - \sum_{k=0}^3 \binom{n}{k} p^k (1-p)^{n-k} > 95\%.$$

A l'aide d'un programme informatique on trouve que $n \geq 76$.

L'espérance et la variance d'une variable aléatoire X suivant une distribution binomiale $\mathcal{B}(n, p)$ sont:

$$\mu(X) = np \quad \text{et} \quad \sigma^2(X) = np(1 - p).$$

Il est facile de démontrer ces résultats dans le cas $n = 1$. En utilisant la définition de μ et de σ^2 :

$$\begin{aligned} \mu(X) &= 1 \cdot p + 0 \cdot (1 - p) = p, \\ \sigma^2(X) &= (1 - p)^2 \cdot p + (0 - p)^2 \cdot (1 - p) = p(1 - p). \end{aligned}$$

Dans le cas général, il convient d'observer que X est la somme de n variables X_1, X_2, \dots, X_n indépendantes telles que $X_i \sim \mathcal{B}(1, p)$:

$$X = X_1 + X_2 + \dots + X_n.$$

Alors,

$$\begin{aligned} \mu(X) &= E\left(\sum X_i\right) = \sum E(X_i) = np, \\ \sigma^2(X) &= \sigma^2\left(\sum X_i\right) = \sum \sigma^2(X_i) = npq. \end{aligned}$$

7.2 La distribution de Poisson

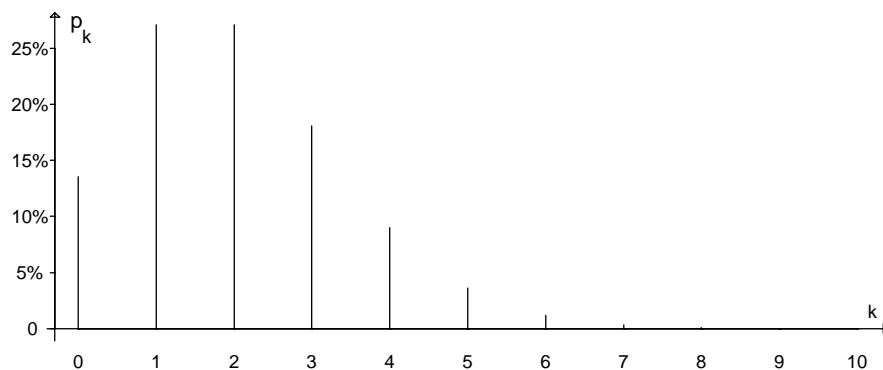
Dans les applications typiques de la distribution binomiale, le nombre n de répétitions de l'épreuve \mathcal{E} est assez petit (une centaine au maximum). Toutefois, dans certaines situations, n est très grand. Dans ces cas, si la probabilité que l'une des deux issues possibles de \mathcal{E} se réalise est très petite, le modèle de Poisson peut être adéquat.

On dit qu'une variable aléatoire discrète X a (ou suit) une *distribution de Poisson* $\mathcal{P}(\lambda)$ de paramètre λ si

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots \quad (2)$$

On obtient:

$$\mu(X) = \lambda \quad \text{et} \quad \sigma^2(X) = \lambda.$$



Distribution de Poisson $\mathcal{P}(\lambda = 2)$; $p_k = P(X = k)$.

Le résultat suivant établit la relation entre le modèle binomial et le modèle de Poisson.

Résultat. Soit X une variable aléatoire binomiale de paramètres p et n . Si $n \rightarrow \infty$ et $p \rightarrow 0$ de telle façon que $np \rightarrow \lambda$ (constante), alors $\lim P(X = k) = e^{-\lambda} \lambda^k / k!$ (la probabilité de $X = k$ selon le modèle de Poisson).

Pour démontrer ce résultat, considérons l'expression générale de la probabilité binomiale:

$$\begin{aligned} P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)(n-2) \cdots (n-k+1)}{k!} p^k (1-p)^{n-k}. \end{aligned}$$

Supposons que $np = \lambda$; donc $p = \lambda/n$ et $1-p = (n-\lambda)/n$. Alors,

$$P(X = k) = \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left[1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)\right] \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

Si $n \rightarrow \infty$ de telle façon que $np = \lambda$ reste fixé, alors on voit que

$$P(X = k) \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}.$$

Le Tableau 1 montre que ce résultat permet de calculer approximativement la distribution binomiale lorsque n est grand et p petit. Il donne la probabilité que $X = k$ pour $k = 0, 1, \dots, 6$ selon les distributions binomiales $\mathcal{B}(3, 1/3)$, $\mathcal{B}(5, 1/5)$, $\mathcal{B}(10, 1/10)$ et $\mathcal{P}(1)$. On peut observer que, pour $n > 10$, $\mathcal{P}(1)$ est une bonne approximation de $\mathcal{B}(n, 1/n)$.

Tableau 1: Approximation d'une distribution binomiale par une distribution de Poisson. L'approximation est déjà bonne pour n relativement petit.

k	$\mathcal{B}(3, 1/3)$	$\mathcal{B}(5, 1/5)$	$\mathcal{B}(10, 1/10)$	$\mathcal{P}(1)$
0	.296	.328	.3487	.3679
1	.444	.410	.3874	.3678
2	.222	.205	.1937	.1839
3	.037	.051	.0574	.0613
4	—	.006	.0111	.0153
5	—	.000	.0015	.0030
6	—	—	.0001	.0005

Exemple

La probabilité qu'une voiture qui passe par un carrefour très fréquenté ait un accident est très petite, par exemple $p = 0.0001$. Toutefois, pendant une certaine période de la journée, par exemple entre 4 heures et 6 heures de l'après-midi, un nombre très grand de voitures passent par le carrefour, par exemple, $n = 1000$. Dans ces conditions, quelle est la probabilité qu'il y ait au moins deux accidents pendant cette période ?

Pour résoudre ce problème en utilisant le modèle de Poisson, on doit supposer que:

- la probabilité p est la même pour toutes les voitures;
- pour une voiture, le fait d'avoir ou de ne pas avoir un accident est indépendant de ce qui arrive à toute autre voiture.

En d'autres termes, nous supposons que le nombre X d'accidents sur 1000 voitures qui passent par le carrefour est une variable binomiale avec paramètre $p = 0.0001$. Evidemment, les hypothèses sus-mentionnées ne sont que des approximations du phénomène réel. En outre, nous supposons de façon implicite que le nombre n de voitures qui passent par le carrefour pendant la période de temps qui nous intéresse est fixé, par exemple, $n = 1000$; en réalité ce nombre varie.

A l'aide du modèle binomial on obtient

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{1000}{k} (0.0001)^k (0.9999)^{1000-k},$$

et la probabilité demandée est

$$1 - [P(X = 0) + P(X = 1)] = 1 - [(0.9999)^{1000} + 1000(0.0001)(0.9999)^{999}].$$

L'évaluation de cette expression avec une calculatrice peut poser des difficultés (si la précision est faible). Pour éviter ces difficultés, nous utilisons la formule (2). En posant $\lambda = np = 0.1$, on trouve

$$P(X = k) \approx e^{-0.1} (0.1)^k / k!.$$

Donc

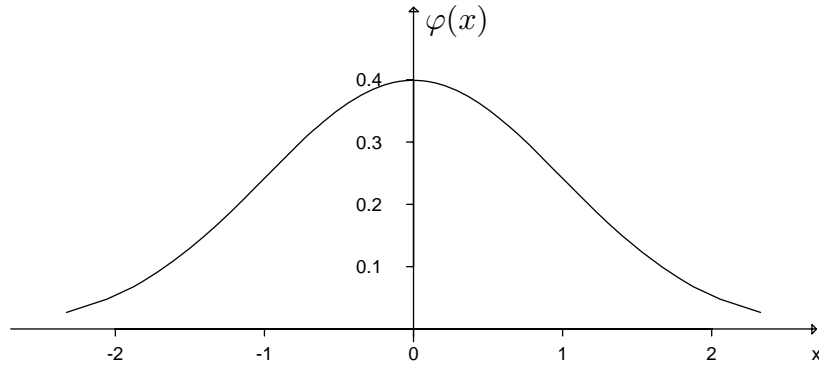
$$P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - e^{-0.1}(1 + 0.1) = 0.0045.$$

7.3 La distribution de Gauss ou normale

1. On dit que la variable aléatoire X a (ou suit) une *distribution normale standard* ou une *distribution de Gauss centrée et réduite* $\mathcal{N}(0, 1)$ si elle a pour densité

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

Le graphique de φ est une courbe “en cloche”.



2. Si X a une distribution $\mathcal{N}(0, 1)$ on obtient

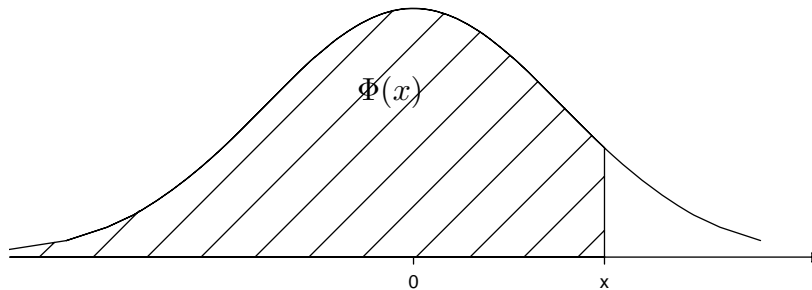
$$\mu(X) = 0 \quad \text{et} \quad \sigma^2(X) = 1.$$

La notation $\mathcal{N}(0, 1)$ rappelle que X a pour moyenne 0 et pour variance 1. (Pour cette raison on dit que X est centrée et réduite.)

3. La fonction de distribution cumulative de X est

$$\Phi(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-s^2/2) ds.$$

Pour déterminer les valeurs de $\Phi(x)$, on se réfère à des tables numériques (voir Tables de la distribution de Gauss) ou on utilise des programmes d'intégration numérique.



4. Comme la fonction $\varphi(x)$ est symétrique par rapport à l'axe vertical ($x = 0$),

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

En interprétant $\Phi(x)$ comme une aire, on obtient

$$\begin{aligned} P(X \geq x) &= 1 - P(X < x) = 1 - \Phi(x), \\ P(X > -x) &= 1 - \Phi(-x) = \Phi(x), \\ P(x_1 < X < x_2) &= \Phi(x_2) - \Phi(x_1), \\ P(-x_1 < X < x_2) &= \Phi(x_2) - \Phi(-x_1) = \Phi(x_1) + \Phi(x_2) - 1, \\ P(-x < X < x) &= 2\Phi(x) - 1. \end{aligned}$$

Exemple 1.

$$\begin{aligned} P(X > 1) &= 1 - \Phi(1) = 0.1587, \\ P(X > 2) &= 1 - \Phi(2) = 0.0228, \\ P(-2 < X < 2) &= 2 \cdot (0.9772) - 1 = 0.9544. \end{aligned}$$

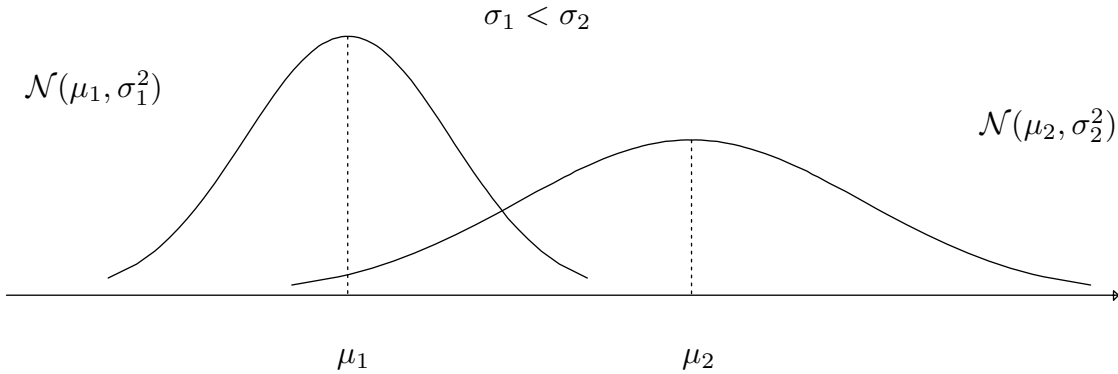
5. La variable aléatoire

$$Y = \sigma X + \mu$$

a une *distribution de Gauss de moyenne μ et variance σ^2* , notée $\mathcal{N}(\mu, \sigma^2)$. Sa densité est

$$f_{\mu, \sigma}(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-(y - \mu)^2 / (2\sigma^2)).$$

En changeant les valeurs de μ et de σ on obtient des cloches de différentes largeurs et positions.



6. La transformation précédente ($Y = \sigma X + \mu$) effectuée en sens inverse permet de passer d'une variable aléatoire Y de distribution $\mathcal{N}(\mu, \sigma^2)$ à la *variable centrée et réduite*

$$X = \frac{Y - \mu}{\sigma}$$

qui suit une distribution $\mathcal{N}(0, 1)$. Cette transformation permet de calculer des probabilités relatives à la variable Y à l'aide de la fonction cumulative et des tables de $\mathcal{N}(0, 1)$. Par exemple, soit $Y \sim \mathcal{N}(0.5, 4)$. Alors:

$$P(Y > 1.5) = P(Y - 0.5 > 1.5 - 0.5) = P((Y - 0.5)/2 > 1/2) = P(X > 0.5) = 0.3085.$$

7. Les deux remarques suivantes sont utiles en pratique.

- La probabilité qu'une variable Y de distribution $\mathcal{N}(\mu, \sigma^2)$ s'éloigne de sa moyenne à plus d'un écart type est d'environ 32%:

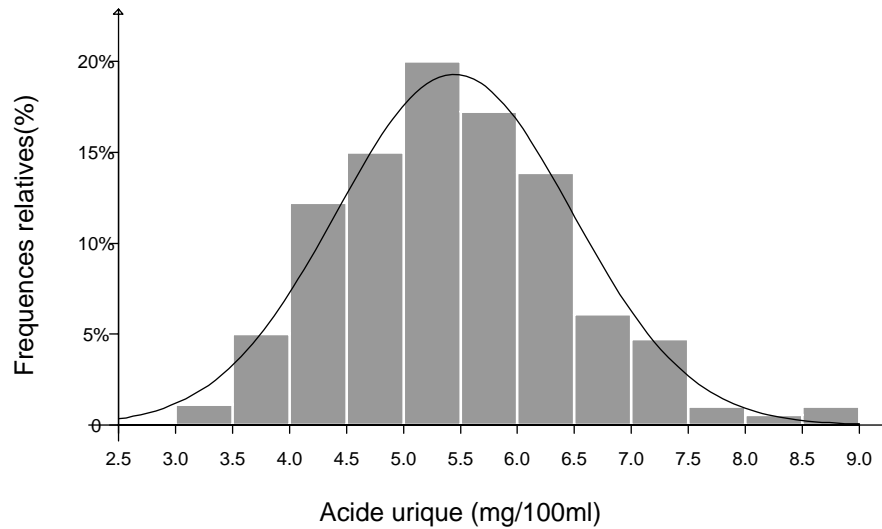
$$P(\mu - 1 \cdot \sigma < Y < \mu + 1 \cdot \sigma) \approx 68\%.$$

- La probabilité que l'éloignement soit supérieur à deux écarts type est d'environ 5%:

$$P(\mu - 2 \cdot \sigma < Y < \mu + 2 \cdot \sigma) \approx 95\%.$$

8. La distribution Gaussienne est une des distributions les plus utilisées en statistique. Beaucoup de variables biologiques ont un comportement approximativement Gaussien.

Exemple 2.



Histogramme de la concentration d'acide urique sérique chez 267 donneurs de sang de sexe masculin

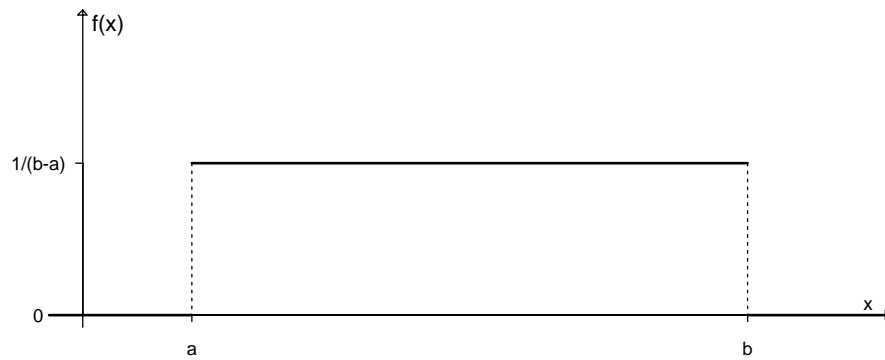
7.4 La distribution uniforme

La *distribution uniforme* dans l'intervalle $[a, b]$ est définie par sa densité

$$f(x) = 1/(b-a) \quad \text{si } x \in [a, b],$$

$$= 0 \quad \text{dans le cas contraire.}$$

La moyenne de cette distribution est $(a+b)/2$ et sa variance $(b-a)^2/12$. Deux exemples se trouvent au Chapitre 6, Section 6.2.

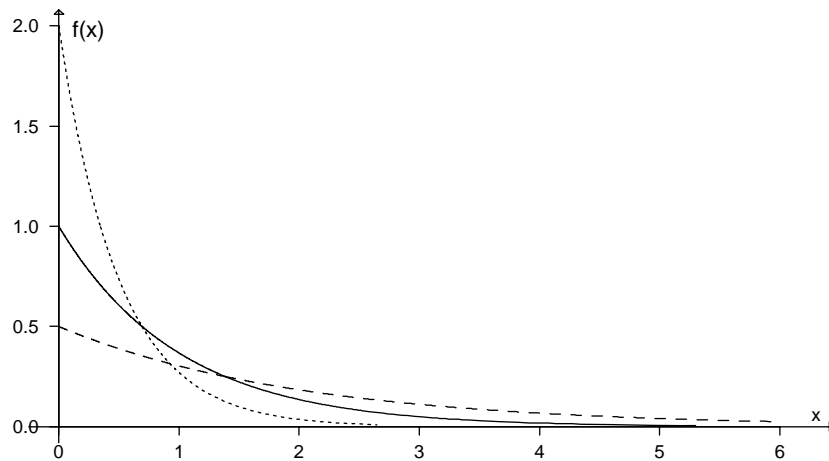


7.5 La distribution exponentielle

La *distribution exponentielle* de paramètre θ est définie par sa densité

$$f(x) = \theta \exp(-\theta x), \quad x \geq 0, \quad \theta > 0.$$

Sa moyenne est $1/\theta$ et sa variance $1/\theta^2$. Un exemple est donné au Chapitre 6, Section 6.2.



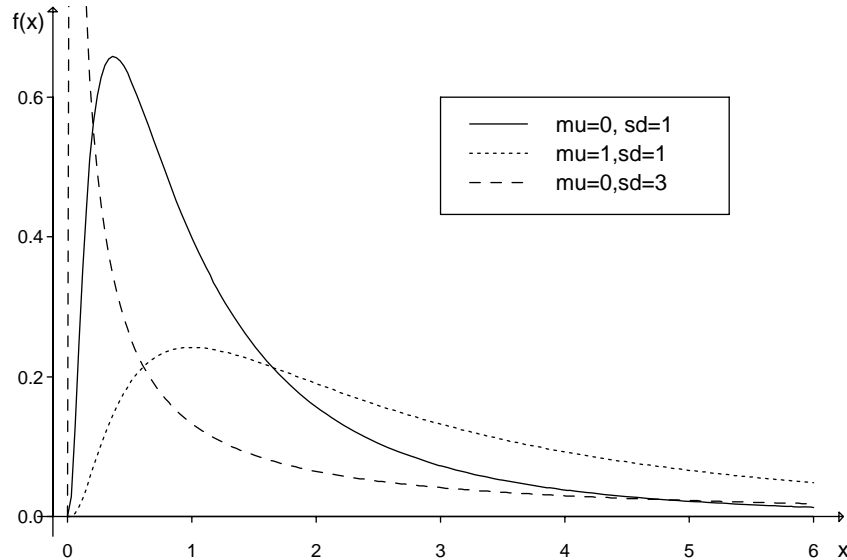
Densités exponentielles pour $\theta=2, 1$ et 0.5

7.6 La distribution lognormale

La *distribution lognormale* de paramètres μ et σ est définie par sa densité

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(\log(x) - \mu)^2\right\}, \quad x > 0, \quad \mu > 0, \quad \sigma > 0.$$

Sa moyenne est $\exp(\mu + \frac{1}{2}\sigma^2)$ et sa variance $\exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$. Un exemple est présenté au Chapitre 6, Complément 1.



7.7 La distribution χ^2

Supposons que X_1, \dots, X_n soient n variables aléatoires indépendantes et identiquement distribuées selon une distribution normale standard. On dit que la variable aléatoire

$$Z = X_1^2 + \dots + X_n^2$$

a une *distribution χ^2* (ki-carré) à n degrés de liberté. La densité de cette distribution est

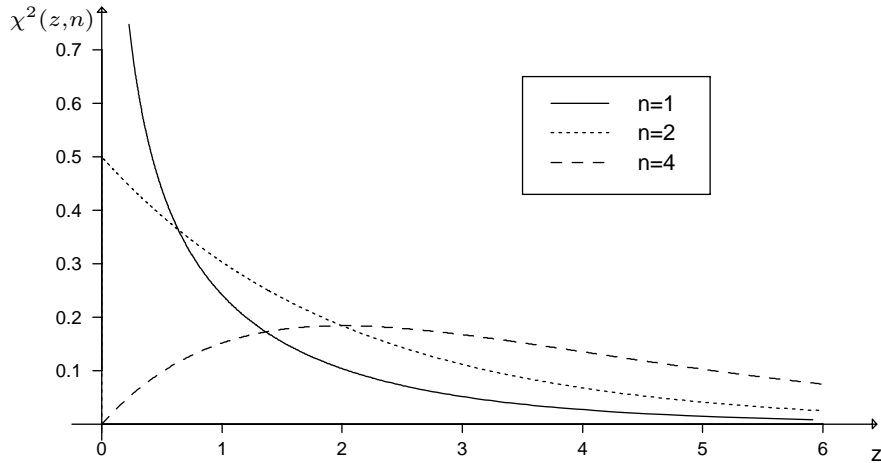
$$f(z) = \frac{1}{2^{n/2}\Gamma(n/2)} \exp(-z/2)z^{(n/2-1)}, \quad z \geq 0,$$

ou $\Gamma(\cdot)$ indique la fonction Γ , définie par

$$\Gamma(p) = \int_0^\infty x^{p-1} \exp(-x) dx, \quad p > 0.$$

La fonction Γ a les propriétés suivantes:

$$\Gamma(p) = (p-1)\Gamma(p-1), \quad \Gamma(1/2) = \sqrt{\pi}, \quad \Gamma(n) = (n-1)! \quad \text{pour } n \text{ entier.}$$



La fonction de distribution cumulative est généralement calculée à l'aide d'un programme informatique ou de "tables de la distribution χ^2 " (voir Tables). La moyenne et la variance de la distribution χ^2 sont $E(Z) = n$ et $\sigma^2(Z) = 2n$. La distribution χ^2 est utilisée dans certains procédés d'inférence (Chapitre 11 et Chapitre 14).

7.8 La distribution t

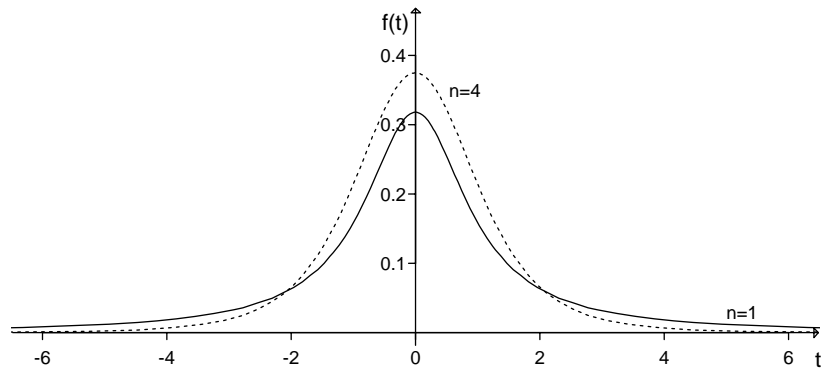
Supposons que X_0, X_1, \dots, X_n soient n variables aléatoires indépendantes et identiquement distribuées selon une distribution normale standard. On dit que la variable aléatoire

$$T = \frac{X_0}{\sqrt{\frac{1}{n}(X_1^2 + \dots + X_n^2)}}$$

a une *distribution t* (ou *distribution de Student*) à n degrés de liberté. La densité (voir figure) de cette distribution est

$$f(t) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{n\pi}} \cdot (1 + t^2/n)^{-(n+1)/2}.$$

La fonction de distribution cumulative est généralement calculée à l'aide d'un programme informatique ou de "tables de la distribution t " (voir Tables). La moyenne et la variance de la distribution t sont $E(T) = 0$ et $\sigma^2(T) = n/(n-2)$, pour $n > 2$. La distribution t est utilisée dans certains procédés d'inférence (Chapitre 12 et Chapitre 17).



Distribution t , avec $n = 1$ et $n = 4$

7.9 La distribution F

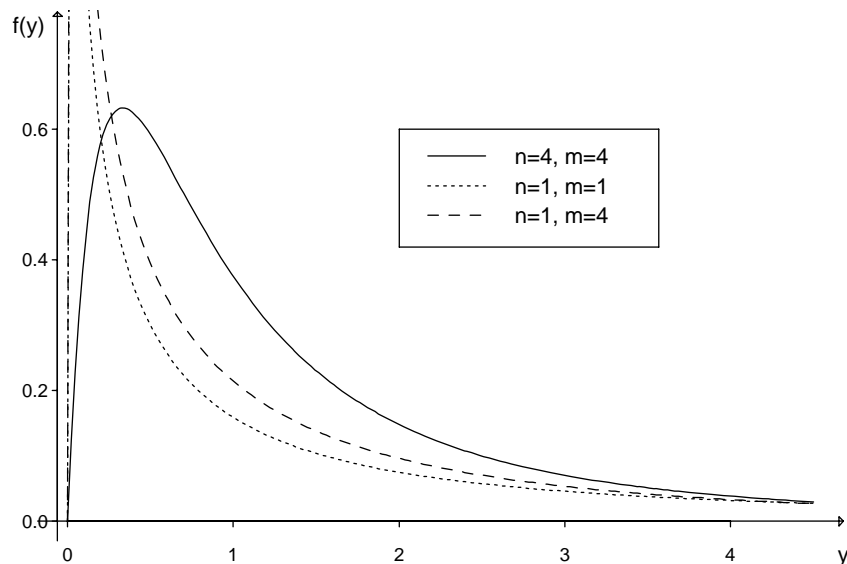
Supposons que X_1, \dots, X_{n+m} soient $n + m$ variables aléatoires indépendantes avec distribution normale standard. On dit que la variable aléatoire

$$Y = \frac{\frac{1}{n}(X_1^2 + \dots + X_n^2)}{\frac{1}{m}(X_{(n+1)}^2 + \dots + X_{m+n}^2)}$$

a une *distribution F* avec n degrés de liberté au numérateur et m degrés de liberté au dénominateur. La densité (voir figure) de cette distribution est

$$f(y) = \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} n^{n/2} m^{m/2} \cdot y^{n/2-1} (m+ny)^{-(n+m)/2}, \quad y \geq 0.$$

Les calculs concernant la distribution F sont généralement effectués à l'aide d'un programme d'ordinateur ou de "tables de la distribution F " (voir Tables). La distribution F est utilisée dans certains procédés d'inférence (Chapitre 20).



7.10 La distribution normale multivariée

Soit $X = (X_1, \dots, X_n)$ un vecteur de variables aléatoires (voir Complément) indépendantes et identiquement distribuées selon $\mathcal{N}(0, 1)$. Leur densité conjointe est donc

$$f(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp \left(- \sum_{i=1}^n x_i^2 / 2 \right).$$

On dit que f est la *densité normale multivariée standard en \mathbb{R}^n* .

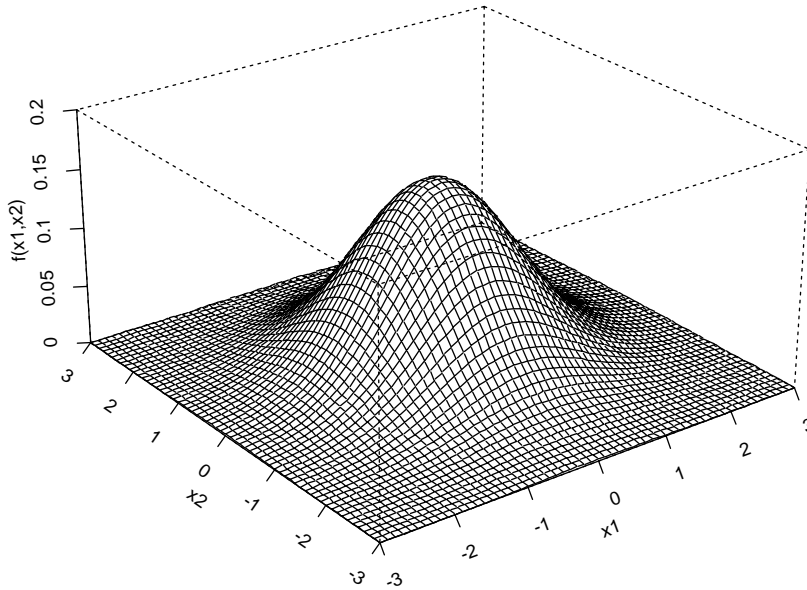
Soit $\mu = (\mu_1, \dots, \mu_m)^T$ un vecteur (colonne) de constantes, C une matrice $m \times n$ de constantes et

$$Y = \mu + CX$$

un nouveau vecteur de m variables aléatoires. La distribution conjointe des composantes de Y est alors une *distribution normale multivariée* avec moyenne μ (vecteur) et *matrice de covariance* $\Sigma = CC^T$. Cette distribution est notée par $\mathcal{N}(\mu, \Sigma)$. On a évidemment $E(Y_i) = \mu_i$ et $\sigma(Y_i, Y_j) = \Sigma_{ij}$ (élément i, j de la matrice Σ). Si C est une matrice $m \times n$ de rang m , la densité conjointe de Y_1, \dots, Y_m est

$$g(y_1, \dots, y_m) = \left(\frac{1}{\sqrt{2\pi}} \right)^m \frac{1}{\sqrt{\det(\Sigma)}} \exp \left(-(y - \mu)^T \Sigma^{-1} (y - \mu) / 2 \right),$$

où $\det(\Sigma)$ est le déterminant de Σ et Σ^{-1} l'inverse de Σ .



Densité normale bivariée standard.

Complément

1. Vecteurs et matrices de variables aléatoires

- Soit $X = (X_{ij})$ une matrice $r \times s$ de variables aléatoires:

$$X = \begin{pmatrix} X_{11} & \dots & X_{1s} \\ \vdots & \vdots & \vdots \\ X_{r1} & \dots & X_{rs} \end{pmatrix}.$$

- L'espérance $E(X)$ de X est définie comme la matrice des espérances

$$E(X) = \begin{pmatrix} E(X_{11}) & \dots & E(X_{1s}) \\ \vdots & \vdots & \vdots \\ E(X_{r1}) & \dots & E(X_{rs}) \end{pmatrix}.$$

- Soit A une matrice $q \times r$ de constantes et B une matrice $s \times t$ de constantes. Alors

$$E(AXB) = AE(X)B.$$

- Si $s = 1$, X est un vecteur de r variables aléatoires, c'est à dire une matrice $r \times 1$:

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix}.$$

La *matrice de covariance* de X est définie comme

$$\Sigma_X = \begin{pmatrix} \sigma(X_1, X_1) & \dots & \sigma(X_1, X_r) \\ \sigma(X_2, X_1) & \dots & \sigma(X_2, X_r) \\ \vdots & \vdots & \vdots \\ \sigma(X_r, X_1) & \dots & \sigma(X_r, X_r) \end{pmatrix}.$$

Donc

$$\Sigma_X = E[(X - E(X))(X - E(X))^T],$$

où $(\cdot)^T$ indique l'opération de transposition matricielle.

- Soit A une matrice $q \times r$ de constantes et

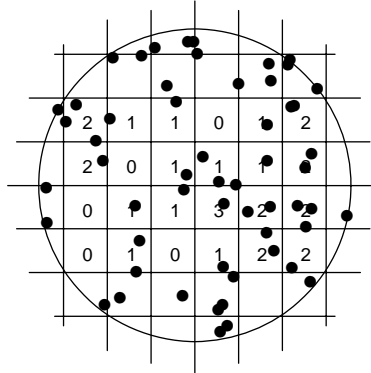
$$Z = AX.$$

Alors

$$\Sigma_Z = A\Sigma_X A^T.$$

2. Une application du modèle de Poisson

Considérons une plaque de Petri sur laquelle, après incubation, on compte les colonies bactériennes. Supposons que la loupe soit quadrillée (voir la figure) et que l'on recense les nombres de colonies par petit carreau. Combien trouvera-t-on de carreaux vides ? de carreaux contenant une seule colonie ? deux colonies, etc. ?



Plaque de Petri

Pour utiliser le modèle de Poisson, il est nécessaire de faire des suppositions concernant l'homogénéité du mélange initial. Plus précisément on suppose que ce qui se passe dans un champ (carreau) n'influence pas ce qui se passe dans un champ voisin. En outre, si on imagine un quadrillage de plus en plus fin (plus fin que celui sur lequel on va effectuer le comptage) on suppose que la probabilité de trouver plus d'une colonie dans un champ s'approche de zéro. Enfin, on suppose que la probabilité de trouver une colonie est la même pour chacun de ces champs très fins. On peut alors démontrer (voir par exemple Çinlar E., 1975) que, sous ces conditions, un modèle de Poisson est adéquat pour calculer la probabilité de trouver k colonies dans un champ.

Dans une certaine étude, 8 espèces de bactéries ont été observées. Pour chaque espèce on a recensé n_0 carreaux avec aucune colonie, n_1 carreaux avec une colonie, n_2 carreaux avec 2 colonies, etc. Dans le Tableau 2, on constate que pour les 8 espèces étudiées, les fréquences n_k observées correspondent très bien aux fréquences $nP(X = k)$ prédites par des modèles de Poisson avec $\lambda = \sum kn_k/n$ (nombre moyen de colonies observées par carreau) et $n = \sum n_k$.

Tableau 2

espèce	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$
1	5 6.1	19 18.0	26 26.7	26 26.4	21 19.6	13 11.7	8 9.5	
2	26 27.5	40 42.2	38 32.5	17 16.7	7 9.1			
3	59 55.6	86 82.2	49 60.8	30 30.0	20 15.4			
4	83 75.0	134 144.5	135 139.4	101 89.7	40 43.3	16 16.7	7 7.4	
5	8 6.8	16 16.2	18 19.2	15 15.1	9 9.0	7 6.7		
6	7 3.9	11 10.4	11 13.7	11 12.0	7 7.9	8 7.1		
7	3 2.1	7 8.2	14 15.8	21 20.2	20 19.5	19 15.0	7 9.6	9 9.6
8	60 62.6	80 75.8	45 45.8	16 18.5	9 7.3			

Dans chaque case figurent deux nombres: la fréquence observée (en haut) et la fréquence prédite (en bas).