

## Chapitre 1: Espaces probabilisés

### 1 Le langage des probabilités

Pour décrire une expérience on se donne un ensemble fondamental  $\Omega$  dont les éléments sont appelés les épreuves et qui se compose de tous les résultats possibles de l'expérience.

On s'intéresse à la réalisation de certains sous-ensembles de  $\Omega$  appelés événements; un événement correspond à un fait attaché à l'expérience susceptible de se produire lors de sa réalisation.

Dans les cas élémentaires, la classe  $\mathcal{A}$  des événements, qu'on appelle aussi famille d'observables, est simplement égale à l'ensemble  $\mathcal{P}(\Omega)$  de toutes les parties de  $\Omega$ . A un stade plus élaboré, on rencontrera des familles d'observables strictement incluses dans  $\mathcal{P}(\Omega)$ . On note la correspondance suivante entre le vocabulaire de la théorie des probabilités et les notations de la théorie des ensembles.

<i>Événement</i>	<i>Ensemble</i>
événement impossible	$\emptyset$
événement certain	$\Omega$
$A$ implique $B$	$A \subset B$
$A$ et $B$ équivalents	$A = B$
réalisation de $A$ et $B$	$A \cap B$
réalisation de $A$ ou $B$	$A \cup B$
contraire de $A$	$A^c$
incompatibilité de $A$ et de $B$	$A \cap B = \emptyset$

### 2 Suite infinie d'événements

Soit  $(A_n)$  une suite infinie d'événements. On introduit les deux ensembles:

$$\begin{aligned} \liminf A_n &= \bigcup_{n=1}^{\infty} (\bigcap_{m=n}^{\infty} A_m) \\ \limsup A_n &= \bigcap_{n=1}^{\infty} (\bigcup_{m=n}^{\infty} A_m) \end{aligned}$$

On vérifie facilement que:

$$\begin{aligned} \liminf A_n &\subset \limsup A_n \\ (\limsup A_n)^c &= \liminf A_n^c \\ (\liminf A_n)^c &= \limsup A_n^c \end{aligned}$$

Si  $\mathbf{1}_A$  désigne l'indicatrice de l'événement  $A$ , alors

$$\begin{aligned} \mathbf{1}_{\limsup A_n} &= \limsup \mathbf{1}_{A_n} \\ \mathbf{1}_{\liminf A_n} &= \liminf \mathbf{1}_{A_n} \end{aligned}$$

### 3 Espaces probabilisés

On rencontrera très souvent des suites d'événements et il est nécessaire de structurer l'ensemble de ces événements sous forme de *tribu*.

**Définition 1** Une tribu ou  $\sigma$ -algèbre sur un espace  $\Omega$  est un ensemble  $\mathcal{A}$  de parties de  $\Omega$  telle que

- la partie vide est un élément de  $\mathcal{A}$
- le complémentaire d'un élément de  $\mathcal{A}$  est dans  $\mathcal{A}$
- la réunion d'une infinité dénombrable d'éléments de  $\mathcal{A}$  est dans  $\mathcal{A}$ .

On en déduit qu'une tribu contient l'ensemble entier  $\Omega$  et est stable par intersection dénombrable.

Une intersection de tribus est encore une tribu.

Il existe une plus petite tribu contenant un ensemble de parties  $\mathcal{E}$  de  $\Omega$ , on l'appelle la *tribu engendrée* par  $\mathcal{E}$ .

Un espace *mesurable* ou *probabilisable* est constitué par un ensemble  $\Omega$  non vide et une tribu  $\mathcal{A}$  de parties de  $\Omega$ .

**Définition 2** Soit  $(\Omega, \mathcal{A})$  un espace probabilisable. On appelle probabilité sur  $(\Omega, \mathcal{A})$  une application  $\mathbb{P}$  de  $\mathcal{A}$  dans  $[0, 1]$  qui vérifie

- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$  pour toute suite d'événements deux à deux incompatibles

**Définition 3** On appelle espace probabilisé un triplet  $(\Omega, \mathcal{A}, \mathbb{P})$  où  $(\Omega, \mathcal{A})$  est un espace probabilisable et  $\mathbb{P}$  une probabilité sur  $(\Omega, \mathcal{A})$ .

**Proposition 1 (Formule de Poincaré)** . Soient  $n \geq 2$  et  $(A_i)_{i=1, \dots, n}$  une suite d'événements. Alors

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})$$

### 4 Probabilités discrètes et dénombrements

Soit un point  $\omega_0$  dans un espace mesurable  $(\Omega, \mathcal{A})$ . On peut considérer la mesure de probabilité

$$\begin{aligned} \delta_{\omega_0}(A) &= 1 && \text{si } \omega_0 \in A \\ &= 0 && \text{sinon} \end{aligned}$$

On l'appelle la *mesure de Dirac* en  $\omega_0$ .

Soit  $N$  un entier strictement positif et  $\Omega = \{\omega_1, \dots, \omega_N\}$  un ensemble fini. L'*équipartition*, ou *loi uniforme* sur  $\Omega$  est la mesure de probabilité

$$P = \frac{1}{N} \sum_{n=1}^N \delta_{\omega_n} .$$

Pour cette probabilité, si  $A$  est un sous-ensemble de  $\Omega$  et si  $\text{card}(A)$  désigne le cardinal de  $A$ , on a

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$$

Le calcul des probabilités se ramène dans ce cas à des calculs de dénombrement. On rappelle la formule de la *somme*:

- si  $A$  et  $B$  sont deux ensembles finis disjoints, alors  $A \cup B$  est fini et

$$\text{card}(A \cup B) = \text{card}(A) + \text{card}(B)$$

ainsi que la formule du *produit*:

- si  $A$  et  $B$  sont deux ensembles finis, leur produit cartésien est fini et

$$\text{card}(A \times B) = \text{card}(A) \times \text{card}(B) .$$

On rappelle aussi quelques formules classiques de dénombrement:

- il y a  $n^p$  suites de longueur  $p$  à éléments choisis dans un ensemble de cardinal  $n$ ; en particulier il y a  $2^p$  parties dans un ensemble à  $p$  éléments

- il y a

$$I(n, p) = \frac{n!}{(n-p)!} = n(n-1) \dots (n-p+1)$$

injections (arrangements) d'un ensemble à  $p$  éléments dans un ensemble à  $n$  éléments

- il y a

$$C_n^p = \binom{n}{p} = \frac{n!}{p!(n-p)!}$$

parties à  $p$  éléments dans un ensemble à  $n$  éléments

- (coefficients multinomiaux) il y a

$$C_n^{n_1, \dots, n_k} = \frac{n!}{n_1! \dots n_k!}$$

suites de longueur  $n$  contenant  $n_1$  fois 1,  $\dots$ ,  $n_k$  fois  $k$  avec

$$n = n_1 + \dots + n_k .$$

## 5 Événements indépendants

**Définition 4** On dit que deux événements  $A$  et  $B$  sont indépendants si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Il est facile de vérifier que  $A$  et  $B$  sont indépendants si et seulement si  $A$  et  $B^c$  le sont;

**Définition 5** Soit  $(A_i, i \in I)$  une famille d'événements de la tribu  $\mathcal{A}$ . On dit que la famille  $(A_i, i \in I)$  est indépendante si pour toute partie finie  $J$  de  $I$ , on a

$$\mathbb{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j)$$

**Définition 6** Une famille de sous-tribus  $(\mathcal{B}_i, i \in I)$  de  $\mathcal{A}$  est dite indépendante si pour toute partie finie  $J$  de  $I$ , pour tout  $j \in J$  et tout  $C_j \in \mathcal{B}_j$ , on a

$$\mathbb{P}(\cap_{j \in J} C_j) = \prod_{j \in J} \mathbb{P}(C_j) .$$

## 6 Probabilités conditionnelles

Soit  $B$  un événement de probabilité  $\mathbb{P}(B) > 0$ . On définit la probabilité conditionnelle de l'événement  $A$  sachant  $B$  par

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

Notons que si  $A$  et  $B$  sont indépendants,  $\mathbb{P}(A|B) = \mathbb{P}(A)$ .

Pour toute suite finie  $(A_i, i = 1, \dots, n)$  on a (si les expressions sont toutes définies)

$$\mathbb{P}(\cap_{i=1}^n A_i) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1) \dots \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}) .$$

Une suite finie ou infinie d'événements  $(A_i, i \in I)$  est appelée un système complet d'événements si

- $A_i \cap A_j = \emptyset$  pour  $i \neq j$ ;
- $\Omega = \cup_{i \in I} A_i$  .

Pour tout système complet d'événements, on a pour tout  $A \in \mathcal{A}$  (formule des probabilités totales):

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|A_i)\mathbb{P}(A_i)$$

Rappelons pour finir la *formule de Bayes*, ou formule de probabilité des causes: pour tout système complet d'événements, pour tout  $i \in I$  et tout  $A \in \mathcal{A}$ ,

$$\mathbb{P}(A_i|A) = \frac{\mathbb{P}(A|A_i)\mathbb{P}(A_i)}{\sum_{j \in I} \mathbb{P}(A|A_j)\mathbb{P}(A_j)} .$$

## Chapitre 2: Variables aléatoires

### 7 Définitions générales

**Définition 7** Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé et  $(E, \mathcal{B})$  un espace probabilisable. Une variable aléatoire définie sur  $\Omega$  et à valeurs dans  $E$  est une application mesurable  $X$  de  $(\Omega, \mathcal{A})$  dans  $(E, \mathcal{B})$ .

*Remarque.* Si  $E = \mathbb{R}$  (resp.  $\mathbb{R}^d$ ) et  $\mathcal{B}$  la tribu des boréliens, on dit simplement que  $X$  est une variable aléatoire réelle ou en abrégé v.a.r. (resp. un vecteur aléatoire).

**Définition 8** Si  $X$  est une variable aléatoire définie sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $(E, \mathcal{B})$ , on appelle loi de  $X$  la probabilité image de  $\mathbb{P}$  par  $X$ . On la notera  $P_X$ .

Ainsi, pour tout  $B \in \mathcal{B}$ , on a  $P_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{X \in B\})$ .

Deux variables aléatoires à valeurs dans le même espace probabilisable sont dites *identiquement distribuées* si elles ont même loi.

**Définition 9** Si  $(X_i)_{i \in I}$  est une famille de variables aléatoires définies sur le même espace  $(\Omega, \mathcal{A}, \mathbb{P})$  et à valeurs dans divers espaces probabilisables  $(E_i, \mathcal{B}_i)$ , on appelle tribu engendrée par les  $(X_i, i \in I)$  la plus petite tribu sur  $\Omega$  rendant chaque  $X_i$  mesurable.

### 8 Variables aléatoires réelles

Un outil précieux en probabilité est le théorème de théorie de la mesure suivant:

**Théorème 1 (de la classe monotone)** Soit  $(\Omega, \mathcal{A})$  un espace probabilisable et  $\mathcal{F} \subset \mathcal{A}$ . On suppose que

- $\Omega \in \mathcal{F}$
- si  $B_1 \in \mathcal{F}$ ,  $B_2 \in \mathcal{F}$  et  $B_1 \subset B_2$ , alors  $B_2 \setminus B_1 \in \mathcal{F}$
- si  $(B_n, n \geq 1)$  est une suite qui vérifie pour tout  $n \geq 1$   $B_n \in \mathcal{F}$  et  $B_n \subset B_{n+1}$ , alors  $\cup_n B_n \in \mathcal{F}$
- $\mathcal{F}$  contient une famille  $\mathcal{C}$  stable par intersection finie.

Alors  $\mathcal{F}$  contient la plus petite tribu  $\sigma(\mathcal{C})$  engendrée par  $\mathcal{C}$ .

**Définition 10** Soit  $X$  une v.a.r. On appelle fonction de répartition de  $X$  la fonction  $F_X$  définie sur  $\mathbb{R}$  par

$$F_X(x) = \mathbb{P}(\{X \leq x\}) = P_X([-\infty, x])$$

Cette fonction ne dépend donc que de la loi de  $X$ . Elle

- est croissante
- est continue à droite
- tend vers 0 quand  $x \rightarrow -\infty$
- tend vers 1 quand  $x \rightarrow +\infty$ .

Inversement, une fonction réelle présentant ces quatre propriétés est la fonction de répartition d'une variable aléatoire sur  $\mathbb{R}$  dont la loi est entièrement déterminée par la fonction, grâce au théorème de la classe monotone.

### Exemples de lois de probabilité discrètes

C'est une probabilité sur  $\mathbb{R}$  définie par  $\sum_{n \geq 0} p_n \delta_{x_n}$  où les  $x_n$  sont des réels distincts et où les  $p_n$  sont des réels dans  $[0, 1]$  avec  $\sum_n p_n = 1$ . Si une v.a.r. possède une telle loi, on dit qu'elle est *discrète* et on a alors

$$p_n = \mathbb{P}(X = x_n)$$

1. Loi de Bernoulli de paramètre  $p$ , avec  $p \in [0, 1]$

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p$$

2. Loi binomiale  $\mathcal{B}(n, p)$  de paramètres  $n \in \mathbb{N}$  et  $p \in [0, 1]$

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k} \quad 0 \leq k \leq n$$

3. Loi hypergéométrique de paramètres  $n, r, r_1$  avec  $\max(n, r_1) \leq r$

$$\mathbb{P}(X = k) = \frac{C_{r_1}^k C_{r-r_1}^{n-k}}{C_r^n} \quad 0 \leq k \leq r_1, 0 \leq n-k \leq r-r_1$$

4. Loi géométrique sur  $\mathbb{N}^*$  de paramètre  $p$  avec  $p \in [0, 1]$

$$\mathbb{P}(X = k) = p(1-p)^{k-1} \quad k \in \mathbb{N}^*$$

5. Loi de Poisson  $\mathcal{P}(\lambda)$  de paramètre  $\lambda > 0$

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k \in \mathbb{N}$$

### Exemples de lois de probabilité absolument continues

Une probabilité  $P$  sur  $\mathbb{R}$  est absolument continue (par rapport à la mesure de Lebesgue) de densité  $f$  si pour tout borélien  $B$  de  $\mathbb{R}$ ,

$$P(B) = \int_B f(x) d\lambda(x).$$

En application du théorème de la classe monotone, il suffit en fait que pour tous  $a$  et  $b$  avec  $a < b$ ,

$$P([a, b]) = P(]a, b]) = \int_a^b f(x) d\lambda(x)$$

1. Loi uniforme sur un intervalle  $I$  borné

$$f(x) = \frac{1}{|I|} \mathbf{1}_I(x)$$

2. Loi de Gauss ou normale de paramètres  $(m, \sigma)$ ,  $m \in \mathbb{R}$ ,  $\sigma > 0$

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

3. Loi de Cauchy  $C(a)$  de paramètre d'échelle  $a$  avec  $a > 0$

$$f(x) = \frac{a}{\pi(x^2 + a^2)}$$

4. Loi exponentielle de paramètre  $\lambda > 0$

$$f(x) = \lambda \exp\{-\lambda x\} \mathbf{1}_{]0, \infty[}(x)$$

## 9 Moments de variables aléatoires réelles

**Définition 11** Si  $X$  est une v.a.r. sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , intégrable pour  $\mathbb{P}$ , on appelle espérance de  $X$  son intégrale

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega).$$

D'après le théorème d'intégration par rapport à une mesure image, appelé aussi *théorème de transfert*, on a également de façon plus commode pour les calculs effectifs

$$\mathbb{E}[X] = \int_{\mathbb{R}} x dP_X(x)$$

L'espérance de  $X$  s'appelle parfois la *moyenne* de  $X$ .

**Définition 12** Une v.a.r. d'espérance nulle est dite centrée.

Toujours grâce au théorème de transfert, on intègre les fonctions réelles des v.a.r. par la formule suivante:

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega)) d\mathbb{P}(\omega) = \int_{\mathbb{R}} g(x) dP_X(x).$$

lorsque  $g(X)$  est intégrable pour  $\mathbb{P}$ . Cette formule est essentiellement utilisée dans les deux cas suivants:

- si  $X$  a ses valeurs dans  $\mathbb{N}$ ,

$$\mathbb{E}[g(X)] = \sum_{n=0}^{\infty} g(n) \mathbb{P}(X = n)$$

- si  $X$  admet une densité  $f_X$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ ,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x) dx.$$

**Définition 13** Lorsque  $X$  est de puissance  $k$ -ième intégrable ( $k \in \mathbb{N}^*$ ), on appelle

moment d'ordre $k$ de $X$	$\mathbb{E}[X^k]$
moment absolu d'ordre $k$ de $X$	$\mathbb{E}[ X ^k]$
moment centré d'ordre $k$ de $X$	$\mathbb{E}[(X - \mathbb{E}[X])^k]$

Le moment centré d'ordre deux s'appelle la *variance*, notée  $Var(X)$ . Sa racine carrée positive s'appelle l'*écart-type*, noté  $\sigma_X$ . Une expression équivalente de la variance est donnée par la formule

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

### Inégalité de Markov

Si  $X$  est une variable aléatoire positive intégrable, alors pour tout  $a > 0$

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}[X]}{a} .$$

### Inégalité de Bienaymé-Tchebicheff

Si  $X$  est une v.a.r. de carré intégrable, pour tout  $a > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| > a) \leq \frac{Var(X)}{a^2} .$$

**Définition 14** Si  $X$  et  $Y$  sont deux v.a.r. de carré intégrable définies sur la même espace probabilisé, on appelle *covariance* de  $X$  et  $Y$  l'expression

$$\begin{aligned} Cov(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] . \end{aligned}$$

Si de plus  $X$  et  $Y$  ne sont pas dégénérées (ne sont pas p.s. constantes), on pose

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

## 10 Vecteurs aléatoires

**Définition 15** Soit  $X = (X_1, \dots, X_d)$  un vecteur aléatoire dont les composantes sont intégrables. On appelle *espérance* de  $X$  et on note  $\mathbb{E}[X]$  le vecteur des espérances.

**Définition 16** Si les composantes du vecteur aléatoire  $X$  sont de carré intégrable, on appelle *matrice de covariance* de  $X$  la matrice carrée

$$K_X = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^t]$$

On vérifie que pour  $1 \leq i, j \leq d$ ,  $K_X(i, j) = Cov(X_i, X_j)$ .

**Proposition 2** Soit  $X$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ ,  $v \in \mathbb{R}^d$  et  $A$  une application linéaire de  $\mathbb{R}^d$  dans  $\mathbb{R}^{d'}$ .

1. Si  $X$  a ses composantes intégrables,

$$\mathbb{E}[v^t X] = v^t \mathbb{E}[X] \quad \text{et} \quad \mathbb{E}[AX] = A\mathbb{E}[X] .$$

2. Si les composantes de  $X$  ont leurs carrés intégrables,

$$Var(v^t X) = v^t K_X v \quad \text{et} \quad K_{AX} = AK_X A^t .$$



**Définition 17** Si  $X$  est un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$  on appelle lois marginales de  $X$  les lois des composantes  $(X_i; i = 1, \dots, d)$ , c'est-à-dire les images la loi  $P_X$  par les projections  $(\pi_i; i = 1, \dots, d)$  de  $\mathbb{R}^d$  dans  $\mathbb{R}$ .

**Proposition 3** Si  $P_X$  admet la densité  $f_X$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$ , alors la  $i$ -ième loi marginale  $P_{X_i}$  admet la densité

$$f_{X_i}(x) = \int \dots \int_{\mathbb{R}^{d-1}} f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_d$$

**Proposition 4 (Formule de changement de variables)** Soit  $X$  un vecteur aléatoire de dimension  $d$  dont la loi  $P_X$  possède la densité  $f_X$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$ . Soit  $G$  un ouvert de  $\mathbb{R}^d$  tel que  $P_X(G^c) = 0$ , et soit  $T$  un difféomorphisme de  $G$  sur  $T(G) \subset \mathbb{R}^d$ . Alors la loi de  $U=T(X)$  admet la densité

$$\begin{aligned} f_U(u) &= |J_{T^{-1}}(u)| f_X(T^{-1}(u)) & \text{si } u \in T(G) \\ &= 0 & \text{si } u \notin T(G) \end{aligned}$$

où  $J_{T^{-1}}(u)$  désigne le déterminant jacobien de l'application  $T^{-1}$  calculé au point  $u$ .

**Exemple.** Soit  $(X, Y)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^2$  dont la loi admet la densité  $f$ . Alors

1. la loi de  $(U, V) = (X, X + Y)$  a pour densité sur  $\mathbb{R}^2$

$$g(u, v) = f(u, v - u)$$

2. la loi de  $(X, Y - X)$  a pour densité sur  $\mathbb{R}^2$

$$h(u, v) = f(u, u + v)$$

3. la loi de  $(U, V) = (X, XY)$  a pour densité sur  $\mathbb{R}^* \times \mathbb{R}$

$$j(u, v) = \frac{1}{|u|} f(u, \frac{v}{u})$$

4. la loi de  $(X, \frac{Y}{X})$  a pour densité sur  $\mathbb{R}^* \times \mathbb{R}$

$$k(u, v) = |u| f(u, uv)$$

On déduit aisément de ces diverses formules les densités de  $X+Y$ ,  $Y-X$ ,  $XY$  et  $\frac{Y}{X}$  en intégrant par rapport à  $u$  sur  $\mathbb{R}$  (ou  $\mathbb{R}^*$ ).

Pour le calcul intégral des fonctions réelles des vecteurs aléatoires, on fera comme dans le cas unidimensionnel grand usage des expressions suivantes:

- si  $X$  a ses valeurs dans  $\mathbb{N}^d$ ,

$$\mathbb{E}[g(X)] = \sum_{n_1, \dots, n_d} g(n_1, \dots, n_d) \mathbb{P}(X = (n_1, \dots, n_d))$$

- si  $X$  admet une densité  $f_X$  par rapport à la mesure de Lebesgue  $\lambda$  sur  $\mathbb{R}^d$ ,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^d} g(x) f_X(x) d\lambda(x).$$

## 11 Variables aléatoires indépendantes

**Définition 18** Une famille de sous-tribus  $(\mathcal{B}_i, i \in I)$  de  $\mathcal{A}$  est dite indépendante si  $\forall J$  fini,  $J \subset I$ ,  $J = \{j_1, \dots, j_n\}$ , pour tout choix de  $C_{j_k} \in \mathcal{B}_{j_k}$ , on a

$$\mathbb{P}(C_{j_1} \cap \dots \cap C_{j_n}) = \prod_{k=1}^n \mathbb{P}(C_{j_k})$$

**Définition 19** Une famille de variables aléatoires à valeurs dans les espaces  $(E_i, \mathcal{C}_i)$  est dite indépendante si la famille des tribus  $(X_i^{-1}(\mathcal{C}_i), i \in I)$  l'est.

Le théorème de la classe monotone permet de montrer les trois résultats suivants:

**Proposition 5** Les v.a.r.  $X_1, \dots, X_n$  sont indépendantes si et seulement si pour tout  $(a_1, \dots, a_n)$  dans  $\mathbb{R}^n$ ,

$$\mathbb{P}(\cap_{i=1}^n \{X_i \leq a_i\}) = \prod_{i=1}^n \mathbb{P}(X_i \leq a_i)$$

**Proposition 6** Soient  $X_1, \dots, X_n, Y_1, \dots, Y_p$  des v.a.r. indépendantes. Si  $f$  et  $g$  sont des fonctions boréliennes de  $\mathbb{R}^n$  dans  $\mathbb{R}$  et de  $\mathbb{R}^p$  dans  $\mathbb{R}$ , les v.a.r.  $f(X_1, \dots, X_n)$  et  $g(Y_1, \dots, Y_p)$  sont indépendantes.

**Proposition 7** Soient  $X_1, \dots, X_n$  des v.a. à valeurs dans les espaces  $(E_1, \mathcal{B}_1), \dots, (E_n, \mathcal{B}_n)$ . Les v.a.  $X_1, \dots, X_n$  sont indépendantes si et seulement si sur l'espace probabilisable  $(E_1 \times \dots \times E_n, \mathcal{B}_1 \otimes \dots \otimes \mathcal{B}_n)$  on a l'égalité des probabilités

$$P_{(X_1, \dots, X_n)} = P_{X_1} \otimes \dots \otimes P_{X_n}$$

On aura donc dans ce cas, dès que les fonctions réelles mesurables  $f_1, \dots, f_n$  sont telles que  $f_i(X_i)$  soit intégrable pour tout  $i = 1, \dots, n$ ,

$$\mathbb{E}[\prod_{i=1}^n f_i(X_i)] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)].$$

**Exemple.** Un vecteur aléatoire  $X = (X_1, \dots, X_n)$  de densité  $f$  a ses composantes indépendantes si et seulement si

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$$

où  $f_i$  est la densité de la  $i$ -ème marginale.

**Proposition 8** Soient  $X_1$  et  $X_2$  deux v.a.r. indépendantes de densités respectives  $f_1$  et  $f_2$ . Leur somme  $Y = X_1 + X_2$  admet pour densité la convolution des fonctions  $f_1$  et  $f_2$  définie par

$$h(y) = \int_{\mathbb{R}} f_1(x) f_2(y - x) dx$$

**Proposition 9** La covariance de deux v.a.r. indépendantes est nulle et la variance de leur somme est égale à la somme de leurs variances.

**Remarque.** On prendra bien garde que la réciproque de la proposition précédente est fautive. Par exemple si  $(X_1, X_2)$  est un couple de v.a.r. dont la loi est donnée par

$$\begin{aligned} \frac{1}{4} &= \mathbb{P}(X_1 = 0, X_2 = 1) \\ &= \mathbb{P}(X_1 = 0, X_2 = -1) \\ &= \mathbb{P}(X_1 = 1, X_2 = 0) \\ &= \mathbb{P}(X_1 = -1, X_2 = 0), \end{aligned}$$

on calcule que  $Cov(X_1, X_2) = 0$  mais on n'a pas

$$P_{(X_1, X_2)} = P_{X_1} \otimes P_{X_2}.$$

## 12 Fonctions génératrices

**Définition 20** Soit  $X$  une v.a. à valeurs dans  $\mathbb{N}$ . Pour  $s \in [0, 1]$ , on pose

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k).$$

On appelle  $G_X$  la fonction génératrice de  $X$ .

Une fonction génératrice est continue sur  $[0, 1]$ ,  $C^\infty$  sur  $[0, 1[$ . Elle caractérise la loi  $P_X$  de  $X$  car

$$\mathbb{P}(X = n) = \frac{G_X^{(n)}(0)}{n!}$$

**Calculs de fonctions génératrices:**

1. Loi binomiale  $\mathcal{B}(n, p)$

$$G_X(s) = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} s^k = (ps + (1-p))^n$$

2. Loi de Poisson  $\mathcal{P}(\lambda)$

$$G_X(s) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} s^k = e^{-\lambda(1-s)}$$

3. Loi géométrique sur  $\mathbb{N}^*$  de paramètre  $p$

$$G_X(s) = \sum_{k=1}^{\infty} p(1-p)^{k-1} s^k = \frac{ps}{1-s+ps}$$

**Proposition 10** Soient  $X$  et  $Y$  deux v.a. indépendantes à valeurs dans  $\mathbb{N}$ , de fonctions génératrices  $G_X$  et  $G_Y$ . Alors

$$G_{X+Y}(s) = G_X(s)G_Y(s).$$

**Corollaire 1** La somme de  $n$  v.a. de Bernoulli indépendantes de même paramètre  $p$  est une  $\mathcal{B}(n, p)$ .

La somme de deux v.a.  $\mathcal{B}(n, p)$  et  $\mathcal{B}(m, p)$  indépendantes est une  $\mathcal{B}(n+m, p)$ .

La somme de deux v.a. de Poisson indépendantes de paramètres  $\lambda$  et  $\mu$  est une v.a. de Poisson de paramètre  $\lambda + \mu$ .

**Proposition 11** Soit  $X$  une v.a. à valeurs dans  $\mathbb{N}$ . Alors  $X$  est intégrable si et seulement si  $G_X'(1-) < \infty$  et alors  $\mathbb{E}[X] = G_X'(1-)$ .

### 13 Lois conditionnelles

**Définition 21** Soient  $X$  et  $Y$  deux v.a. à valeurs dans  $\mathbb{N}$ . Pour tout  $k \in \mathbb{N}$  tel que  $P_X(\{k\}) = \mathbb{P}(X = k) > 0$ , on appelle loi conditionnelle de  $Y$  sachant que  $X$  vaut  $k$  la probabilité sur  $\mathbb{N}$  définie par

$$P_{Y|X=k}(\{l\}) = \mathbb{P}(Y = l|X = k) = \frac{\mathbb{P}(X = k, Y = l)}{\mathbb{P}(X = k)}.$$

**Définition 22** Soit  $(X, Y)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^2$  de densité  $f_{(X,Y)}$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}^2$ . On appelle densité conditionnelle de  $Y$  sachant que  $X$  vaut  $x$  la densité

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

définie en tout point  $x$  tel que  $f_X(x) > 0$ .

Dans les deux cas, on appellera *espérance conditionnelle* de  $Y$  sachant que  $X$  vaut  $x$  l'espérance par rapport à la loi conditionnelle.

Cas discret:

$$\mathbb{E}[Y|X = x] = \sum_y y P_{Y|X=x}(\{y\})$$

Cas à densité:

$$\mathbb{E}[Y|X = x] = \int y f_{Y|X=x}(y) dy = \frac{\int y f_{(X,Y)}(x, y) dy}{\int f_{(X,Y)}(x, y) dy}.$$

On vérifie facilement que si les variables aléatoires  $X$  et  $Y$  sont indépendantes, la loi conditionnelle de  $Y$  sachant que  $X$  vaut  $x$  est égale à la loi de  $Y$  quel que soit  $x$ . Dans ce cas, la connaissance de la valeur de  $X$  n'apporte aucune information sur la répartition des valeurs de  $Y$ .

### 14 Vecteurs gaussiens

**Définition 23** On dit qu'une v.a.r. est gaussienne ou normale si elle admet la densité

$$f_{(m,\sigma)}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

où  $m \in \mathbb{R}$ ,  $\sigma > 0$ , ou encore si elle est constante égale à  $m$ . On notera cette loi  $\mathcal{N}(m, \sigma^2)$ , avec  $\sigma = 0$  dans le cas constant. On appelle loi normale centrée réduite la loi  $\mathcal{N}(0, 1)$ .

On vérifie que

$$\int x f_{m,\sigma}(x) dx = m \quad \text{et} \quad \int (x - m)^2 f_{m,\sigma}(x) dx = \sigma^2$$

donc pour une variable  $X$  de loi  $\mathcal{N}(m, \sigma^2)$ ,

$$\mathbb{E}[X] = m \quad \text{et} \quad \text{Var}(X) = \sigma^2.$$

**Définition 24** On dit qu'un vecteur aléatoire  $d$ -dimensionnel est gaussien si toute combinaison linéaire de ses composantes est une v.a.r. gaussienne.

Par la formule de changement de variables, il est facile de voir que si  $X$  est une v.a.r. gaussienne de paramètres  $(m, \sigma)$ , alors  $aX + b$  est aussi une v.a.r. gaussienne de paramètres  $(am + b, |a|\sigma)$ .

**Proposition 12** *La somme de deux v.a.r. gaussiennes indépendantes est une v.a.r. gaussienne. Un vecteur aléatoire dont les composantes sont des v.a.r. gaussiennes indépendantes est un vecteur gaussien.*

**Définition 25** *Si  $a$  et  $\lambda$  sont deux réels  $> 0$ , on appelle loi  $\Gamma(a, \lambda)$  la probabilité sur  $\mathbb{R}$  de densité*

$$\gamma(a, \lambda)(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \mathbf{1}_{]0, \infty[}(x),$$

où  $\Gamma(a) = \int_0^\infty u^{a-1} e^{-u} du$ .

**Proposition 13** *Soit  $X$  une v.a.r. normale centrée réduite. Alors  $X^2$  suit la loi  $\Gamma(\frac{1}{2}, \frac{1}{2})$ .*

**Proposition 14** *Soient  $X$  et  $Y$  deux v.a.r. indépendantes de lois  $\Gamma(a, \lambda)$  et  $\Gamma(b, \lambda)$ . Alors  $X + Y$  a pour loi  $\Gamma(a + b, \lambda)$ .*

**Proposition 15** *Soient  $X_1, \dots, X_n$  des v.a.r. indépendantes normales centrées réduites. Alors*

$$Z = \sum_{i=1}^n X_i^2$$

*suit la loi  $\Gamma(\frac{n}{2}, \frac{1}{2})$  appelée loi du khi-deux à  $n$  degrés de liberté ( $\chi^2(n)$ ).*

**Proposition 16** *Soient  $U$  et  $V$  deux v.a.r. indépendantes, avec  $U$  normale centrée réduite et  $V$  suivant la loi du khi-deux à  $n$  degrés de liberté. Alors*

$$T = \sqrt{n} \frac{U}{\sqrt{V}}$$

*a pour densité*

$$t_n(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \frac{1}{(1 + \frac{x^2}{n})^{\frac{n+1}{2}}}.$$

*On dit que  $T$  suit la loi de Student à  $n$  degrés de liberté.*

**Théorème 2 (de Fisher)** *Soient  $n \geq 2$ ,  $X_1, \dots, X_n$   $n$  v.a.r. indépendantes de même loi  $\mathcal{N}(m, \sigma^2)$ , où  $m$  est un réel et  $\sigma$  un réel  $> 0$ . On pose*

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i \\ Q &= \sum_{i=1}^n (X_i - \bar{X})^2 \\ S &= \sqrt{\frac{Q}{n-1}}. \end{aligned}$$

*Alors,*

1.  $\bar{X}$  suit la loi  $\mathcal{N}(m, \frac{\sigma^2}{n})$
2.  $\frac{Q}{\sigma^2}$  suit la loi du khi-deux à  $n - 1$  degrés de liberté
3.  $\bar{X}$  et  $Q$  sont indépendantes
4.  $T = \sqrt{n} \frac{\bar{X} - m}{S}$  suit la loi de Student à  $n - 1$  degrés de liberté.

## Chapitre 3: Limites de suites de variables aléatoires

### 15 Approximation de lois.

#### 15.1 Approximation d'une loi hypergéométrique par une loi binomiale.

La loi hypergéométrique de paramètres  $n, r, r^1$  correspond au tirage de  $n$  objets parmi  $r$  lorsque la population est partagée en une catégorie d'effectif  $r^1$  et une catégorie d'effectif  $r - r^1$ .

**Théorème 3** Pour tout  $j \in \mathbb{N}^*$ , soit  $X_j$  une variable aléatoire de loi hypergéométrique de paramètres  $n, r_j, r_j^1$ . Si  $\lim_{j \rightarrow \infty} r_j = \infty$  et  $\lim_{j \rightarrow \infty} \frac{r_j^1}{r_j} = p$ , où  $0 < p < 1$ , alors

$$\lim_{j \rightarrow \infty} \mathbb{P}(X_j = k) = C_n^k p^k (1-p)^{n-k}$$

pour  $0 \leq k \leq n$ .

#### 15.2 Approximation d'une loi binomiale par une loi de Poisson.

**Théorème 4** Soit  $(p_n, n \geq 1)$  une suite de réels de  $]0, 1[$  telle que  $\lim_{n \rightarrow \infty} np_n = \lambda$ , où  $\lambda$  est un réel  $> 0$ . Pour tout entier  $n \geq 1$ , soit  $S_n$  une variable aléatoire de loi  $\mathcal{B}(n, p_n)$ . Alors, pour tout entier  $k \in \mathbb{N}$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(S_n = k) = e^{-\lambda} \frac{\lambda^k}{k!} .$$

Ce théorème donne une approximation de la loi binomiale lorsque  $n$  est grand et  $p$  petit.

#### 15.3 Approximation de la loi binomiale par la loi de Gauss.

**Théorème 5** Soit  $0 < p < 1$ . Pour tout  $n \geq 1$ , soit  $S_n$  une variable aléatoire de loi  $\mathcal{B}(n, p)$  et soit

$$Z_n = \frac{S_n - np}{\sqrt{np(1-p)}} .$$

Alors, quand  $n \rightarrow \infty$ ,

$$\sup_{-\infty \leq a < b \leq +\infty} \left| \mathbb{P}(a < Z_n < b) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \right| \rightarrow 0 .$$

**Corollaire 2** Soit  $0 < p < 1$ . Pour tout  $n \geq 1$ , soit  $S_n$  une variable aléatoire de loi  $\mathcal{B}(n, p)$ . Pour tous réels  $a < b$ ,

$$\lim_{n \rightarrow \infty} \left[ \mathbb{P}(a < S_n < b) - \frac{1}{\sqrt{2\pi}} \int_{\frac{a-np}{\sqrt{np(1-p)}}}^{\frac{b-np}{\sqrt{np(1-p)}}} e^{-\frac{x^2}{2}} dx \right] = 0 .$$

## 16 Convergence en loi.

**Définition 26** On dit qu'une suite de variables aléatoires réelles  $(X_n)$  converge en loi vers une variable aléatoire  $X$  si en tout point  $x$  où la fonction de répartition  $F_X$  de  $X$  est continue, on a

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

**Théorème 6 (limite central)** . Soit  $(X_n)$  une suite de variables aléatoires réelles indépendantes et de même loi admettant un moment d'ordre deux, de moyenne  $m$  et d'écart-type  $\sigma > 0$ . Si on pose

$$Z_n = \frac{\sum_{j=1}^n X_j - nm}{\sqrt{n}\sigma}$$

alors la suite  $(Z_n)$  converge en loi vers une variable aléatoire de loi  $\mathcal{N}(0,1)$ .

Notons

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du .$$

Alors pour tout couple  $a < b$ , on a donc

$$\mathbb{P}(a < Z_n \leq b) \rightarrow \Phi(b) - \Phi(a) .$$

De la table de la loi normale, on retiendra notamment que

$$\Phi(1.96) - \Phi(-1.96) = 0.95 .$$

## 17 Loi des grands nombres.

**Définition 27** On dit qu'une suite  $(X_n)$  de variables aléatoires réelles converge p.s. vers une variable aléatoire réelle  $X$  si

$$\mathbb{P}(\{\omega : \lim_n X_n(\omega) = X(\omega)\}) = 1$$

**Proposition 17** Soit  $X$  une variable aléatoire réelle et soit  $(X_n)$  une suite de v.a.r. telle qu'il existe une suite  $\varepsilon_n$  tendant vers 0 et vérifiant

$$\sum_n \mathbb{P}(|X_n - X| > \varepsilon_n) < \infty .$$

Alors  $(X_n)$  converge p.s. vers  $X$ .

**Définition 28** On dit qu'une suite  $(X_n)$  de variables aléatoires réelles converge en probabilité vers une variable aléatoire réelle  $X$  si pour tout  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0 .$$

**Théorème 7 (Loi faible des grands nombres)** Soit  $(X_n)$  une suite de variables aléatoires réelles définies sur le même espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ , indépendantes, de même moyenne  $m$  et de même écart-type  $\sigma$ . Alors la suite des variables aléatoires

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$$

converge en probabilité vers  $m$ .

## Chapitre 4 : Introduction à la statistique

### 18 La décision statistique.

Un statisticien observe un phénomène dont la loi de probabilité n'est pas connue au départ. Il doit au vu de l'observation d'un vecteur aléatoire prendre une décision qui souvent entraînera un coût.

Considérons par exemple une livraison de produits manufacturés formée de  $N$  objets. Un certain nombre de ces objets sont défectueux. Soit  $N\theta$  ce nombre inconnu, où  $\theta$  est une proportion. On voudrait avoir une idée de la valeur de  $\theta$ . Pour cela, on tire un échantillon de  $n$  objets sans remise, et on les vérifie. L'observation est donc, pour chaque objet  $i$ , une variable binaire indiquant s'il est défectueux ( $X_i = 1$ ) ou non ( $X_i = 0$ ). On note  $Y = \sum_{i=1}^n X_i$  le nombre d'objets défectueux dans l'échantillon. La loi de  $Y$  est hypergéométrique de paramètres  $(n, N, N\theta)$  et l'espérance de  $Y$  est  $n\theta$ . Ainsi une valeur raisonnable à proposer pour  $\theta$  est  $Y/n$ . On a traité ici un problème d'*estimation* du paramètre  $\theta$ .

Prenons un autre exemple. Un premier somnifère procure une durée de sommeil aléatoire  $X$  et un second une durée également aléatoire  $Y$ . On se demande lequel des deux est le plus efficace. Pour cela, on administre le premier à un groupe de  $m$  patients, le second à un autre groupe de  $n$  patients. Au vu des résultats observés, peut-on penser que le premier est plus efficace que le second? Ceci est un problème de *test* d'hypothèses.

### 19 L'estimation.

On suppose que l'espace des événements  $(E, \mathcal{E})$  associés à une expérience est muni d'une famille de probabilités  $(Q_\theta, \theta \in \Theta)$ , où  $\Theta$  s'appelle l'ensemble des états de la nature. Soit  $g$  un caractère numérique défini sur  $\Theta$ : lorsque la nature est dans l'état  $\theta$ , le caractère prend la valeur  $\lambda = g(\theta)$ . On se propose d'estimer  $\lambda$ , c'est-à-dire d'indiquer un nombre  $\hat{\lambda}$  que l'on pense être une bonne approximation du nombre inconnu  $\lambda$ . Le nombre  $\hat{\lambda}$  est alors la valeur estimée, ou l'*estimation* de  $\lambda$ . Pour faire cette estimation, on utilise le résultat de l'expérience, ou, mieux, on répète  $n$  fois l'expérience de façon indépendante, ce qui fait qu'on observe le vecteur  $(X_1, \dots, X_n)$  à valeurs dans  $\mathbb{R}^n$  défini sur  $(\Omega, \mathcal{A}) = (E, \mathcal{E})^{\otimes n}$  muni de la probabilité produit  $\mathbb{P}_\theta = Q_\theta^{\otimes n}$ . L'estimation  $\hat{\lambda}$  sera une fonction des observations  $(X_1, \dots, X_n)$ .

*Estimateurs sans biais.*

**Définition 29** On dit que  $\hat{\lambda} = f(X_1, \dots, X_n)$  est un estimateur sans biais de  $\lambda$  si pour tout  $\theta \in \Theta$ ,

$$\mathbb{E}_\theta[\hat{\lambda}] = \lambda.$$

Il est intéressant d'obtenir des estimateurs sans biais pour que la valeur à estimer soit une valeur centrale pour les observations. De plus on souhaitera que la loi de l'estimateur soit la plus concentrée possible autour de la valeur cible.

**Définition 30** Le risque quadratique d'un estimateur  $\hat{\lambda}$  de  $g(\theta)$  est

$$R_\theta(\hat{\lambda}) = \mathbb{E}_\theta[(\hat{\lambda} - g(\theta))^2].$$



Le risque quadratique d'un estimateur sans biais est donc sa variance.

On peut espérer avoir de meilleurs résultats lorsque la taille de l'échantillon observé augmente. On est alors amené à la définition suivante.

**Définition 31** On dit qu'une suite d'estimateurs  $(\hat{\lambda}_n)$  est consistante si elle converge en probabilité  $\mathbb{P}_\theta$  vers  $g(\theta)$  pour toute valeur de  $\theta$  dans  $\Theta$ .

### Estimateurs empiriques

Une première méthode de construction d'estimateurs repose sur l'emploi de la moyenne empirique et de la loi des grands nombres.

En effet supposons que  $\theta$  soit un paramètre d'espérance, c'est-à-dire que l'on observe un  $n$ -échantillon  $(X_1, \dots, X_n)$  d'une loi  $\mathbb{P}_\theta$  pour laquelle  $\mathbb{E}_\theta[X] = \theta$ , et admettant un moment d'ordre deux. Si l'on estime  $\theta$  par la *moyenne empirique*  $\bar{X}_n$ , on a alors

- $\mathbb{E}_\theta[\bar{X}_n] = \theta$ , i.e.  $\bar{X}_n$  est un estimateur sans biais de  $\theta$ ;
- $R_\theta(\bar{X}_n) = \text{var}_\theta(X)/n$  tend vers 0 quand  $n \rightarrow \infty$ ;
- $(\bar{X}_n)$  est un estimateur consistant de  $\theta$ .

De plus, la précision de l'estimateur moyenne empirique peut être évaluée à l'aide du théorème limite central; par exemple

$$\bar{X}_n - \theta \in \left[ -\frac{1.96\sigma_\theta(X)}{\sqrt{n}}; +\frac{1.96\sigma_\theta(X)}{\sqrt{n}} \right]$$

avec une probabilité proche de 95%.

Si l'on veut estimer  $g(\theta)$ , avec  $g$  fonction continue, on peut utiliser le lemme suivant.

**Lemme 1** Si  $(Z_n)$  converge en probabilité vers une constante  $a$ , et si  $g$  est une fonction continue en  $a$ , alors

$$g(Z_n) \xrightarrow{\mathbb{P}} g(a) \quad \text{quand } n \rightarrow \infty.$$

On peut également s'intéresser à l'estimation de la variance de l'observation. On a alors recours à la proposition suivante.

**Proposition 18** Soit  $(X_1, \dots, X_n)$  un échantillon de la loi de  $X$  qui admet un moment d'ordre 4 pour tout  $\theta$ . On note  $m_\theta = \mathbb{E}_\theta[X]$  et  $\sigma_\theta^2 = \text{var}_\theta(X)$ . On appelle variance empirique de l'échantillon la v.a.

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Alors  $S_n^2$  est un estimateur sans biais et consistant de  $\sigma_\theta^2$ .

*Preuve.* On pose  $Q = \sum_{i=1}^n (X_i - \bar{X}_n)^2$ , on calcule

$$\begin{aligned} \mathbb{E}_\theta[Q] &= \mathbb{E}_\theta[\sum_i ((X_i - m_\theta) - (\bar{X}_n - m_\theta))^2] \\ &= \sum_i \mathbb{E}_\theta[(X_i - m_\theta)^2] - 2 \sum_i \mathbb{E}_\theta[(X_i - m_\theta)(\bar{X}_n - m_\theta)] + n\mathbb{E}_\theta[(\bar{X}_n - m_\theta)^2] \\ &= n\sigma_\theta^2 - \frac{2n}{n}\sigma_\theta^2 + n\frac{1}{n}\sigma_\theta^2 \\ &= (n-1)\sigma_\theta^2. \end{aligned}$$

Pour la convergence en probabilité on écrit

$$S_n^2 = \frac{n}{n-1} \left[ \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X}_n)^2 \right].$$

Le premier terme tend vers  $\mathbb{E}_\theta[X^2]$  par la loi des grands nombres appliquée aux  $(X_i^2)$ , et le second vers  $(\mathbb{E}_\theta[X])^2$  en vertu du lemme précédent.

*Estimateur du maximum de vraisemblance.*

Une méthode générique permettant de déterminer un estimateur consiste à choisir la valeur de  $\theta$  qui maximise la probabilité d'avoir observé les données  $(x_1, \dots, x_n)$ , autrement dit on attribue à  $\theta$  ce qu'on peut appeler la valeur la plus vraisemblable. C'est la méthode dite du maximum de vraisemblance.

Si le modèle statistique consiste en l'observation d'un  $n$ -échantillon i.i.d. d'une loi  $Q_\theta$  discrète que l'on a observé  $\omega = (x_1, \dots, x_n)$ , la *vraisemblance* associée est la fonction de  $\theta$

$$L_\omega(\theta) = \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i),$$

probabilité d'avoir observé  $\omega$ .

Dans le cas de lois avec densité, i.e. lorsque la loi  $Q_\theta$  admet une densité  $f_\theta$ , la vraisemblance est définie par

$$L_\omega(\theta) = \prod_{i=1}^n f_\theta(x_i).$$

L'estimateur du maximum de vraisemblance est défini dans les deux cas de la façon suivante.

**Définition 32** *L'estimateur du maximum de vraisemblance de  $\theta$  est la valeur*

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L_\omega(\theta).$$

Il est souvent plus commode de calculer le logarithme de la vraisemblance, appelé la log-vraisemblance, et de maximiser ce logarithme.

*Exemple.* Dans le modèle de Bernoulli, la vraisemblance s'écrit

$$L_\omega(\theta) = \theta^{s_n} (1 - \theta)^{n - s_n}, \quad \text{où } s_n = \sum_{i=1}^n x_i$$

et donc

$$\ln(L_\omega(\theta)) = s_n \ln(\theta) + (n - s_n) \ln(1 - \theta).$$

On obtient ici  $\hat{\theta} = \bar{X}_n$ .

## 20 Intervalles de confiance.

On cherche assez souvent à obtenir, plutôt qu'un estimateur ponctuel, un ensemble dans lequel on peut penser que  $\lambda$  se trouve avec une très forte probabilité. On se donne donc un coefficient de sécurité  $\beta$  proche de 1 et une région  $R$  dépendant des observations  $(X_1, \dots, X_n)$  telle que la

probabilité que  $R$  contienne  $\lambda$  soit au moins  $\beta$ . On dit que  $R$  est une *région de confiance* pour  $\lambda$  de *coefficient de sécurité*  $\beta$ .

La plupart du temps,  $\lambda$  est à valeurs dans  $\mathbb{R}$  et  $R$  sera un intervalle. Il faut donc trouver deux fonctions  $f_1$  et  $f_2$  de  $\mathbb{R}^n$  dans  $\mathbb{R}$  telles que pour tout  $\theta$  dans  $\Theta$ ,

$$\mathbb{P}_\theta(f_1(X_1, \dots, X_n) \leq \lambda \leq f_2(X_1, \dots, X_n)) \geq \beta$$

### Intervalle de confiance pour une proportion

Dans une population divisée en deux catégories A et B avec les proportions  $p$  et  $1 - p$ , on effectue  $n$  tirages indépendants avec remise. On note  $X_i = 1$  si le  $i$ -ième tirage a donné un individu de la catégorie A,  $X_i = 0$  si c'est la catégorie B. Alors

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

représente la proportion d'individus de la catégorie A tirés dans l'échantillon. Si l'échantillon est de grande taille, la *variable pivotale*

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

suit à peu près la loi  $\mathcal{N}(0, 1)$  d'après le théorème de la limite centrale. Dans ce cas, comme  $\bar{X}$  est proche de  $p$  (loi des grands nombres), on peut estimer que

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

suit également à peu près la loi  $\mathcal{N}(0, 1)$ .

Si  $c_\beta$  désigne un nombre lu dans la table tel que  $\mathcal{N}(0, 1)([-c_\beta, c_\beta]) = \beta$ , on a donc comme intervalle de confiance approximatif pour  $p$  avec coefficient de sécurité  $\beta$

$$\bar{X} - \frac{c_\beta}{\sqrt{n}} \sqrt{\bar{X}(1-\bar{X})} \leq p \leq \bar{X} + \frac{c_\beta}{\sqrt{n}} \sqrt{\bar{X}(1-\bar{X})}$$

### Intervalle de confiance pour la moyenne d'une loi gaussienne

Si  $X$  suit la loi  $\mathcal{N}(m, \sigma^2)$ , avec  $m$  et  $\sigma$  inconnus, on sait que la variable pivotale

$$T = \sqrt{n} \frac{\bar{X} - m}{S}$$

suit la loi de Student à  $n - 1$  degrés de liberté. On lit la valeur  $t_\beta$  dans la table de cette loi et on obtient l'intervalle de confiance

$$\bar{X} - \frac{t_\beta}{\sqrt{n}} S \leq m \leq \bar{X} + \frac{t_\beta}{\sqrt{n}} S.$$

### Intervalle de confiance pour la variance d'une variable gaussienne

D'après le théorème de Fischer, on sait que  $Q/\sigma^2$  suit la loi  $\chi^2(n-1)$ . On trouve alors dans la table de cette loi deux nombres  $u_\beta$  et  $u'_\beta$  tels que  $\chi^2([u_\beta, u'_\beta]) = \beta$  et on en déduit l'intervalle de confiance

$$\frac{Q}{u'_\beta} \leq \sigma^2 \leq \frac{Q}{u_\beta}.$$

## 21 Tests d'hypothèses.

Dans l'ensemble des états de la nature  $\Theta$ , on privilégie une partie  $H$  et on note  $K$  son complémentaire. Si le véritable  $\theta$  qui régit l'expérience en cours appartient à  $H$ , on dit que l'hypothèse  $H$  est vraie, sinon on dit qu'elle est fausse. A partir des observations  $X_1, \dots, X_n$ , on va décider si l'on pense que l'hypothèse est vraie ou non. Il faut donc choisir une région  $R$  de  $\Omega$  telle que

- $\mathbb{P}_\theta(R)$  soit petit si  $\theta \in H$ : c'est l'erreur de première espèce;
- $1 - \mathbb{P}_\theta(R)$  soit petit si  $\theta \notin H$ : c'est l'erreur de seconde espèce.

Si l'on se donne un *seuil*  $\varepsilon$ , on imposera à l'erreur de première espèce d'être toujours majorée par  $\varepsilon$ .

Rejeter l'hypothèse  $H$  signifie qu'on la met très fortement en doute, car on a fait des observations peu compatibles avec cette hypothèse.

## 22 Tests sur les paramètres de lois gaussiennes.

### 22.1 Test sur une loi gaussienne.

Ici la variable observée suit une loi  $\mathcal{N}(m, \sigma^2)$ . Si  $m_0$  et  $\sigma_0$  désignent des valeurs connues, on cherche à tester

1.  $H : m = m_0$  contre  $K : m \neq m_0$
2.  $H : m \leq m_0$  contre  $K : m > m_0$
3.  $H : \sigma = \sigma_0$  contre  $K : \sigma \neq \sigma_0$
4.  $H : \sigma \leq \sigma_0$  contre  $K : \sigma > \sigma_0$

1. Sous l'hypothèse  $H$ , la variable pivotale

$$T = \sqrt{n} \frac{\bar{X} - m_0}{S}$$

suit la loi  $t(n-1)$ . On lit alors dans cette table le nombre  $t_\varepsilon$  tel que  $t_{n-1}([-t_\varepsilon, t_\varepsilon]) = 1 - \varepsilon$  et on refuse  $H$  si  $|t_{obs}| > t_\varepsilon$ .

2. On note cette fois  $t_\varepsilon$  le nombre tel que  $t_{n-1}([t_\varepsilon, \infty]) = \varepsilon$  et on refuse  $H$  si  $t_{obs} > t_\varepsilon$ , où  $t_{obs}$  est le même que précédemment.
3. Lorsque  $\sigma = \sigma_0$  la variable  $U = Q/\sigma_0^2$  suit la loi  $\chi^2(n-1)$ . On cherche deux nombres  $u_\varepsilon$  et  $u'_\varepsilon$  tels que

$$\chi^2(n-1)([u_\varepsilon, u'_\varepsilon]) = 1 - \varepsilon.$$

On rejette  $H$  si  $u_{obs} < u_\varepsilon$  ou  $u_{obs} > u'_\varepsilon$ .

4. On lit dans la table le nombre  $u_\varepsilon$  tel que  $\chi^2(n-1)([u_\varepsilon, \infty]) = \varepsilon$ . On rejette  $H$  si  $u_{obs} > u_\varepsilon$ .

### 22.2 Test sur deux lois gaussiennes.

On s'intéresse à l'observation de deux variables  $X$  et  $Y$  indépendantes de lois respectives  $\mathcal{N}(m, \sigma^2)$  et  $\mathcal{N}(m', \sigma'^2)$  et on veut tester l'hypothèse  $H : m = m'$  contre  $K : m \neq m'$ . On utilise pour cela un échantillon  $(X_1, \dots, X_n)$  de  $X$  et un échantillon  $(Y_1, \dots, Y_{n'})$  de  $Y$ , ces deux échantillons étant indépendants. Sous l'hypothèse  $H$ , la variable pivotale

$$T = \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{\sum(X_i - \bar{X})^2 + \sum(Y_j - \bar{Y})^2}{n+n'-2}}} \frac{1}{\sqrt{\frac{1}{n} + \frac{1}{n'}}}$$

suit la loi  $t(n+n'-2)$ . On lit alors dans la table la valeur de  $t_\varepsilon$  telle que  $t(n+n'-2)([-t_\varepsilon, t_\varepsilon]) = 1 - \varepsilon$  et on refuse  $H$  si  $|t_{obs}| > t_\varepsilon$ .

## 23 Le test du khi-deux.

### 23.1 Test du $\chi^2$ d'adéquation à une loi discrète

Le test du khi-deux est un test d'adéquation, ou de conformité, à une loi donnée. Sous sa forme la plus simple, il teste la validité de l'hypothèse faite sur la loi d'une variable  $X$  à valeurs dans un ensemble fini. Par exemple, si une variable  $X$  peut prendre les valeurs  $x_1, \dots, x_k$  avec les probabilités  $q_1, \dots, q_k$ , et si on observe un échantillon de  $X$  de taille  $n$ , alors on peut pour tout  $j = 1, \dots, k$  compter le nombre de fois  $N_j^n$  où la valeur  $x_j$  a été observée.

L'hypothèse  $H$  signifiera que la vraie loi  $(q_1, \dots, q_k)$  de  $X$  est une loi connue  $(p_1, \dots, p_k)$ , son complémentaire signifiera que ce n'est pas vrai. Sous l'hypothèse  $H$ , d'après un théorème dû à Karl Pearson, la quantité

$$\chi_{k,n}^2 = \sum_{j=1}^k \frac{(N_j^n - np_j)^2}{np_j},$$

calculée à partir des observations, suit approximativement la loi  $\chi^2(k-1)$  lorsque  $n$  est grand.

*Exemple* L'examen de 320 familles de 5 enfants a montré la répartition suivante:

classes: nombre de garçons	0	1	2	3	4	5
effectifs: nombre de familles	13	44	85	103	57	18

Si l'on suppose les naissances indépendantes et équiprobables pour chaque sexe, alors on doit avoir pour  $i = 1, \dots, 6$

$$p_i = C_5^{i-1} 2^{-5}.$$

La valeur calculée de  $\chi_{k,n}^2$  est 11,34, on la compare à la valeur  $c_\varepsilon = 11,1$  lue dans la table du  $\chi^2(5)$  pour un seuil  $\varepsilon = 5\%$ . Comme elle est plus grande, on refuse l'hypothèse faite.

### 23.2 Test d'indépendance du $\chi^2$

Soit  $X \in \{a_1, \dots, a_k\}$  et  $T \in \{b_1, \dots, b_l\}$  deux v.a. à nombre fini de valeurs distinctes. On observe  $n$  couples de réponses  $((X_1, Y_1), \dots, (X_n, Y_n))$  et on souhaite tester l'hypothèse

$H$ :  $X$  et  $Y$  sont indépendantes contre  $K$ : c'est faux.

La loi du couple  $(X, Y)$  est

$$p = (p_{i,j}, i = 1, \dots, k, j = 1, \dots, l), \quad p_{i,j} = \mathbb{P}(X = a_i, Y = b_j),$$

et les lois marginales de  $X$  et  $Y$  sont les suites

$$p_{i,\bullet} = \mathbb{P}(X = a_i) = \sum_{j=1}^l p_{i,j}, \quad p_{\bullet,j} = \mathbb{P}(Y = b_j) = \sum_{i=1}^k p_{i,j}.$$

Sous l'hypothèse nulle, la loi du couple est la loi  $p^0$  produit des marginales

$$p_{i,j}^0 = p_{i,\bullet} p_{\bullet,j} \quad \forall i = 1, \dots, k \quad \forall j = 1, \dots, l.$$

On souhaite procéder comme pour le test d'adéquation et estimer la loi du couple par la loi empirique ( $\hat{p}$ ) :

$$\hat{p}_{i,j} = \frac{N_{i,j}}{n}, \quad N_{i,j} = \sum_{m=1}^n \mathbf{1}_{\{X_m=a_i, Y_m=b_j\}},$$

et évaluer la distance entre  $\hat{p}$  et la loi sous  $H$ . La différence avec la situation précédente vient de ce que la loi  $p^0$  sous  $H$  n'est pas connue. Il faut donc l'estimer elle aussi, ce que l'on fera en estimant les marginales:

$$p_{i,j}^0 = \frac{N_{i,\bullet}}{n} \frac{N_{\bullet,j}}{n}$$

où

$$N_{i,\bullet} = \sum_{j=1}^l N_{i,j}, \quad N_{\bullet,j} = \sum_{i=1}^k N_{i,j}.$$

On peut alors montrer que lorsque  $n$  tend vers l'infini, la variable

$$\chi_{k,l,n}^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(N_{i,j} - np_{i,j}^0)^2}{np_{i,j}^0},$$

converge en loi vers un  $\chi^2$  à  $(k-1)(l-1)$  degrés de liberté. La construction de la zone de rejet  $R$  est analogue à la construction précédente: on rejette l'hypothèse d'indépendance si le  $\chi_{k,l,n}^2$  observé est supérieur à la valeur lue dans la table.