

Initiation aux mathématiques du hasard

Ecole Polytechnique Universitaire - Génie Mécanique 4

Année 2008-09

L. Mazliak

10 septembre 2008

Les notes de cours qui suivent sont à comprendre comme une initiation à la théorie des probabilités que le temps imparti ne permet que d'affleurer. Le but principal que l'on s'est fixé est de donner au lecteur une idée des notions les plus importantes de cette théorie (en tout premier lieu celle de loi d'une variable aléatoire) et des premières applications à la statistique paramétrique.

1 Probabilités élémentaires

1.1 Rappels sur la combinatoire

Soit E un ensemble de cardinal (nombre d'éléments de E , noté $|E|$) égal à n . On note par ailleurs $\mathcal{P}(E)$ l'ensemble des parties (ou sous-ensembles) de E .

a. Le cardinal de $\mathcal{P}(E)$ est égal à 2^n .

La démonstration se fait par récurrence. La formule est évidemment vraie pour $n = 1$ (la seule partie de l'ensemble vide est l'ensemble vide). On la suppose vraie pour $n \geq 0$. Soit alors F un ensemble à $n + 1$ éléments, et fixons un élément arbitraire de F noté a . Les parties de F se répartissent alors en deux catégories : celles qui ne contiennent pas a , qui sont donc des parties de l'ensemble à n éléments $G = F - \{a\}$: il y en a 2^n . Celles qui contiennent a , qu'on peut voir comme la réunion du singleton $\{a\}$ avec une partie de G : il y en a donc également 2^n . Donc dans F il y a $2^n + 2^n = 2^{n+1}$ parties d'où la propriété au rang $n + 1$.

b. Le nombre de n -uplets (permutations) qu'on peut former avec des éléments distincts de E est

$$n! = n(n-1)(n-2)\dots 3.2.1$$

En effet, si on considère un n -uplet (a_1, \dots, a_n) formé d'éléments distincts de E , on a : n possibilités pour a_1 , $n - 1$ possibilités pour a_2, \dots

c. Le nombre de k -uplets qu'on peut former avec des éléments distincts de E est

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

Même raisonnement que précédemment mais avec un k -uplet (a_1, \dots, a_k) .
d. Le nombre de parties à k éléments dans E est

$$C_n^k = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\dots(n-k+1)}{k(k-1)\dots 1}$$

Il suffit de remarquer qu'à une partie à k éléments correspondent $k!$ k -uplets (permutations) possibles. De ce fait, $C_n^k = \frac{1}{k!} A_n^k$.

1.2 Espaces de probabilités finis

Les différentes versions possibles du hasard sont représentées par les éléments d'un ensemble (ici, de cardinal fini) Ω . Une partie de Ω (qui représente donc une sélection de hasards possibles) s'appelle un *événement*. Ω est l'événement certain, \emptyset l'événement impossible.

On pose alors la définition suivante :

Definition 1.1 Une probabilité sur Ω est une application $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ telle que $P(\Omega) = 1$ et, si $A, B \subset \Omega$ tels que $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$.

Le couple (Ω, P) est dit un espace de probabilités.

Une conséquence immédiate de la définition sont les trois propriétés suivantes.

- $P(A^c) = 1 - P(A)$ (où $A^c = \Omega - A$).
En effet, $\Omega = A \cup A^c$ et c'est une union disjointe.
- Si $A \subset B$, $P(A) \leq P(B)$.
En effet, $B = A \cup (B - A)$, c'est une réunion disjointe et $P(B - A) \geq 0$.
- Si A et B sont des événements,

$$P(A \cup B) \leq P(A) + P(B).$$

En effet, on peut écrire $A \cup B = A \cup (B \cap A^c)$, donc $P(A \cup B) = P(A) + P(B \cap A^c)$ et $P(B \cap A^c) \leq P(B)$.

On déduit de cette définition une manière pratique de définir P sur Ω (fini!). Si $\Omega = \{\omega_1, \dots, \omega_n\}$, et si on pose $P(\{\omega_i\}) = p_i$, on remarque que $0 \leq p_i \leq 1$, et

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^n \{\omega_i\}\right) = \sum_{i=1}^n P(\{\omega_i\}) = \sum_{i=1}^n p_i.$$

De ce fait, se donner une probabilité sur Ω revient à se donner $p_1, \dots, p_n \in [0, 1]$ tels que $\sum_{i=1}^n p_i = 1$ et à poser $P(\{\omega_i\}) = p_i$. Pour tout $A \subset \Omega$, on a alors

$$P(A) = \sum_{i, \omega_i \in A} p_i.$$

Il faut bien comprendre qu'à chaque choix différent de (p_1, \dots, p_n) correspond une probabilité sur Ω .

1.2.1 Cas particulier important

On peut en particulier supposer que

$$p_1 = p_2 = \dots = p_n = \frac{1}{n}.$$

Ceci définit une probabilité P sur Ω dite *probabilité uniforme*. Pour celle ci, on a

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

Soit (Ω, P) un espace de probabilités comme défini précédemment.

Definition 1.2 Une variable aléatoire à valeurs dans un ensemble fini $E = \{e_1, \dots, e_n\}$ est une application $X : \Omega \rightarrow E$. La loi de X est une probabilité sur E , notée μ_X , définie par

$$\mu_X(\{e_i\}) = P(\{\omega \in \Omega, X(\omega) = e_i\}) = P(X = e_i).$$

Il faut bien comprendre que l'écriture $P(X = e_i)$ est un abus de notation que l'on fait car il est très parlant.

Exemple 1.3 Soit $A \subset \Omega$ tel que $p = P(A)$. On note $\mathbb{1}_A$ la variable aléatoire telle que $\mathbb{1}_A(\omega) = 1$ si $\omega \in A$ et 0 sinon. On vérifie immédiatement que la loi de $\mathbb{1}_A$ est la probabilité sur $\{0, 1\}$ donnée par

$$\mu_X(\{0\}) = 1 - p, \mu_X(\{1\}) = p.$$

Une telle loi est dite loi de Bernoulli de paramètre p , notée $\mathcal{B}(p)$.

Exercice 1.4 On lance 2 dés et on pose X la valeur la plus grande affichée par les deux dés. Déterminer la loi de X .

Le cas le plus important de variable aléatoire est celui où $E \subset \mathbb{R}$, car on peut alors faire des calculs algébriques avec les éléments de E . On a alors la définition suivante.

Definition 1.5 Soit X une variable aléatoire sur $\Omega = \{\omega_1, \dots, \omega_n\}$ à valeurs dans $E = \{e_1, \dots, e_n\} \subset \mathbb{R}$.

a) On appelle espérance de X le réel

$$E(X) = \sum_{\omega \in \Omega} P(\{\omega\})X(\omega)$$

b) On appelle variance de X le réel

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

Conséquences immédiates

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

- $E(\lambda X) = \lambda E(X)$
- si $X \leq Y$ (au sens où $X(\omega) \leq Y(\omega)$ pour tout $\omega \in \Omega$), $E(X) \leq E(Y)$.
En particulier, si $X \geq 0$, $E(X) \geq 0$.
- $\text{Var}(X) \geq 0$

Pour le calcul de $E(X)$, on a le résultat essentiel suivant :

Theorème 1.6 *Soit X une variable aléatoire à valeurs dans $E = \{e_1, \dots, e_k\} \subset \mathbb{R}$. On note μ_X sa loi. Alors*

$$E(X) = \sum_{i=1}^k \mu_X(\{e_i\})e_i.$$

Démonstration : On a naturellement

$$\Omega = \bigcup_{i=1}^k (X = e_i) = \bigcup_{i=1}^k \{\omega \in \Omega, X(\omega) = e_i\}$$

Donc,

$$\begin{aligned} E(X) &= \sum_{\omega \in \Omega} P(\{\omega\})X(\omega) = \sum_{i=1}^k \sum_{\omega \in (X=e_i)} P(\{\omega\})X(\omega) \\ &= \sum_{i=1}^k \left(\sum_{\omega \in (X=e_i)} P(\{\omega\})e_i \right) = \sum_{i=1}^k P(X = e_i)e_i. \quad \square \end{aligned}$$

Exemple 1.7 *Si X est une variable constante égale à λ , on a immédiatement $E(X) = \lambda$. Si $X = \mathbb{1}_A$, on trouve $E(X) = p.1 + (1-p).0 = p$.*

Exercice 1.8 *On lance deux dés. Calculer la loi et l'espérance de la somme des deux dés, et du maximum des deux dés.*

1.3 Inégalités de Markov et Bienaymé-Tchebitcheff

1.3.1 Inégalité de Markov

Soit X une variable aléatoire positive. Soit $\alpha > 0$. Alors

$$P(X > \alpha) \leq \frac{E(X)}{\alpha}.$$

Démonstration : Il suffit d'écrire

$$X = X\mathbb{1}_{X \leq \alpha} + X\mathbb{1}_{X > \alpha} \geq X\mathbb{1}_{X > \alpha} \geq \alpha\mathbb{1}_{X > \alpha}$$

d'où

$$E(X) \geq \alpha E(\mathbb{1}_{X > \alpha}) = \alpha P(X > \alpha). \quad \square$$

1.3.2 Inégalité de Bienaymé-Tchebitcheff

Soit X une variable aléatoire réelle. Soit $\varepsilon > 0$. Alors,

$$P(|X - E(X)| > \varepsilon) \leq \frac{1}{\varepsilon^2} \text{Var}(X).$$

Démonstration : Posons $Y = (X - E(X))^2$. C'est une variable aléatoire réelle positive. D'après l'inégalité de Markov, on a donc

$$P((X - E(X))^2 > \varepsilon^2) \leq \frac{1}{\varepsilon^2} E[(X - E(X))^2] = \frac{\text{Var}(X)}{\varepsilon^2}. \quad \square$$

1.4 Probabilité conditionnelle et indépendance

La question qu'on se pose est comment estimer la probabilité que l'événement A se produise si on a l'information que l'événement B s'est produit. Comme les hasards élémentaires qui produisent A sont à choisir dans B (dont on sait par hypothèse qu'il s'est produit), le choix le plus naturel est de prendre pour cette estimation $\alpha P(A \cap B)$ où α est une constante. Pour fixer la valeur de cette dernière, on remarque que si $A = B$, il est naturel de dire que la probabilité que B se produise si on sait que B s'est produit est égale à 1. Ceci assure que $\alpha = \frac{1}{P(B)}$ qu'on ne peut donc considérer que si $P(B) \neq 0$. D'où la définition suivante

Definition 1.9 Soit B un événement tel que $P(B) \neq 0$. Pour tout événement A on appelle probabilité conditionnelle de A sachant B la quantité

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

Un cas particulier fondamental est celui où l'on considère que B n'a pas d'influence sur A , au sens où le fait de savoir ou non que B se produit ne change pas l'estimation faite sur A , et donc $P(A/B) = P(A)$, donc $P(A \cap B) = P(A)P(B)$. On observe que cette égalité a un sens même si $P(B) = 0$. On en tire la définition suivante :

Definition 1.10 On dit que les événements A et B sont indépendants si $P(A \cap B) = P(A)P(B)$.

Noter, et c'est une remarque essentielle, que la notion d'indépendance n'a rien à voir avec le fait que A et B sont disjoints, comme le montre l'exercice suivant.

Exercice 1.11 Soient A et B deux événements disjoints. Montrer que A et B sont indépendants si et seulement si $P(A) = 0$ ou $P(B) = 0$.

On vérifie aisément que si A et B sont indépendants, A et B^c le sont aussi. En effet,

$$P(A \cap B^c) = 1 - P([A \cap B^c]^c) = 1 - P(A^c \cup B) =$$

$$\begin{aligned}
&= 1 - P(A^c \cup (B \cap A)) = 1 - [P(A^c) + P(B \cap A)] \\
&= 1 - P(A^c) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c).
\end{aligned}$$

Passons à l'indépendance des variables aléatoires. Soient X_1, \dots, X_n n variables aléatoires à valeurs dans $E \subset \mathbb{R}$ fini.

Definition 1.12 On dit que X_1, \dots, X_n sont indépendantes si et seulement si pour tous x_1, \dots, x_n dans E , on a

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i).$$

Application : Soient X_1, \dots, X_n des variables aléatoires indépendantes de loi de Bernoulli $\mathcal{B}(p)$. On pose $X = X_1 + \dots + X_n$ et on cherche la loi de X . On vérifie alors (exercice) que X est à valeurs dans $\{0, 1, \dots, n\}$ et que

$$P(X = k) = C_n^k p^k (1-p)^{n-k}.$$

On dit que X suit la loi binomiale $\mathcal{B}(n, p)$.

Une importante propriété des variables réelles indépendantes concerne l'espérance et la variance

Proposition 1.13 Si X et Y sont deux variables aléatoires indépendantes à valeurs dans $E \subset \mathbb{R}$ fini, on a

- a. $E(XY) = E(X)E(Y)$
- b. $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

a)

$$\begin{aligned}
E(XY) &= \sum_{\omega \in \Omega} X(\omega)Y(\omega)P(\{\omega\}) = \sum_{(x,y) \in E \times E} \sum_{X(\omega)=x, Y(\omega)=y} X(\omega)Y(\omega)P(\{\omega\}) \\
&= \sum_{(x,y) \in E \times E} xyP(X=x, Y=y) \\
&= \sum_{(x,y) \in E \times E} xyP(X=x)P(Y=y) \text{ par l'indépendance} \\
&= \left(\sum_{x \in E} xP(X=x) \right) \left(\sum_{y \in E} yP(Y=y) \right) \\
&= E(X)E(Y).
\end{aligned}$$

b)

$$\begin{aligned}
\text{Var}(X+Y) &= E((X+Y - E(X+Y))^2) = \\
&= E(X^2 + Y^2 + 2XY - E(X)^2 - E(Y)^2 - 2E(X)E(Y)) \\
&= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 + 2[E(XY) - E(X)E(Y)] \\
&= \text{Var}(X) + \text{Var}(Y). \quad \square
\end{aligned}$$

2 Espaces de probabilités généraux

Dans cette section, nous allons étendre la définition d'une probabilité, de façon à pouvoir considérer des situations plus générales que celles que nous avons regardées jusqu'à présent. Cette extension va être faite *a minima* pour éviter de trop rentrer dans des détails techniques.

Commençons par étendre la définition de probabilité.

Definition 2.1 Soit Ω un ensemble. Une probabilité est une application $P : \Omega \rightarrow [0, 1]$ telle que pour toute suite d'événements $(A_n)_{n \geq 0}$ disjoints deux à deux,

$$P\left(\bigcup_{n \geq 0} A_n\right) = \sum_{n \geq 0} P(A_n).$$

De cette définition, on tire les conséquences suivantes, que nous admettrons :

Pour toute suite d'événements $(A_n)_{n \geq 0}$,

- $P\left(\bigcup_{n \geq 0} A_n\right) \leq \sum_{n \geq 0} P(A_n)$
- si $A_n \subset A_{n+1}$, $P(A_n) \rightarrow P\left(\bigcup_{n \geq 0} A_n\right)$
- si $B_n \subset B_{n+1}$, $P(B_n) \rightarrow P\left(\bigcup_{n \geq 0} B_n\right)$

2.1 Variables aléatoires entières

Par analogie avec le cas fini, on appellera *variable aléatoire à valeurs dans \mathbb{N}* une application X de Ω dans \mathbb{N} . La loi de X est donnée par la suite $(p_k)_{k \geq 0}$ où

$$p_k = P(X = k) = P(\{\omega \in \Omega, X(\omega) = k\}).$$

Comme dans le cas fini, on posera alors

$$E(X) = \sum_{k \geq 0} k p_k$$

mais **seulement si cette série converge**.

2.1.1 Cas particuliers importants

On reprend les notations précédentes : X est une variable aléatoire à valeurs dans \mathbb{N} et $p_k = P(X = k)$.

- Soit $0 < \alpha < 1$ fixé. Si $p_k = (1 - \alpha)\alpha^k, k \geq 0$, on dit que X suit la loi géométrique de paramètre α et on note $X \sim \mathcal{G}(\alpha)$

Exercice 2.2 Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre p . On pose

$$T = \inf\{n \geq 0, X_n = 1\}.$$

Déterminer la loi de T .

Si $X \sim \mathcal{G}(\alpha)$,

$$E(X) = (1 - \alpha) \sum_{k \geq 0} k \alpha^k = \frac{\alpha}{1 - \alpha}.$$

- Soit $\lambda > 0$ fixé. Si $p_k = e^{-\lambda} \frac{\lambda^k}{k!}, k \geq 0$ on dit que X suit la loi de Poisson de paramètre λ et on note $X \sim \mathcal{P}(\lambda)$.
Si $X \sim \mathcal{P}(\lambda)$,

$$E(X) = \sum_{k \geq 0} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda.$$

Exercice 2.3 Soit $(p_n)_{n \geq 0}$ une suite de réels positifs telle que $np_n \rightarrow \lambda$. Soient X_n des variables aléatoires de loi binomiale $\mathcal{B}(n, p_n)$. Montrer que

$$\forall k \geq 0, P(X_n = k) \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

La loi de Poisson apparaît ainsi comme la *limite* d'une loi binomiale quand pour n grand le produit des paramètres est proche d'une constante λ .

- Exercice 2.4**
- Soient X et Y deux variables aléatoires de loi de Poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$. On suppose que X et Y sont indépendantes. Déterminer la loi de $X + Y$.
 - Soient X et Y deux variables aléatoires de lois géométriques $\mathcal{G}(\alpha)$ et $\mathcal{G}(\beta)$. On pose $Z = \min(X, Y)$. Déterminer la loi de Z .

2.2 Variables à densité

Pour généraliser encore le modèle, nous allons considérer un type de variables aléatoires prenant cette fois des valeurs quelconques dans \mathbb{R} .

Definition 2.5 Soit $f : \mathbb{R} \rightarrow \mathbb{R}^+$ une fonction continue par morceaux. On dit qu'une fonction $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire admettant la densité f si pour tous $-\infty \leq a \leq b \leq +\infty$,

$$P(a \leq X \leq b) = \int_a^b f(t) dt.$$

Notons qu'alors $\int_{-\infty}^{+\infty} f(t) dt = 1$ (il suffit de prendre $a = -\infty$ et $b = +\infty$).

En quelque sorte, la définition précédente dit que pour chaque t $P(X = t)$ est un infiniment petit qui vaut $f(t) dt$. Naturellement, en toute rigueur, $P(X = t) = 0$ (on prend $a = b = t$ dans la définition).

L'espérance de X est alors donnée quand l'intégrale existe par

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

On peut voir cette formule comme une extension de la définition donnée dans le cas discret. Symboliquement

$$E(X) = \sum_t t.P(X = t) = \sum_t t.f(t) dt$$

Plus généralement, on a la définition suivante

Definition 2.6 Soit X une variable aléatoire de loi de densité f . Soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ continue par morceaux telle que $\int_{-\infty}^{+\infty} |\varphi(t)| f(t) dt$ converge. Alors

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

En particulier, quand on peut définir $E(X^2)$ (c'est-à-dire quand $\int_{-\infty}^{+\infty} x^2 f(x) dx$ converge), on pose $\text{Var}(X) = E(X^2) - E(X)^2$.

2.3 Loi gaussienne

L'exemple de loi à densité qui est le plus important est celui des gaussiennes.

Definition 2.7 On appelle densité gaussienne (ou normale) centrée réduite la fonction

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

On a le résultat suivant

Proposition 2.8 Soit X une variable aléatoire de loi gaussienne centrée réduite. Pour tout $k \geq 1$, on a

$$E(X^{2k+1}) = 0, E(X^{2k}) = 1.3.5 \dots (2k-1)$$

Démonstration :

$$E(X^{2k+1}) = \int_{-\infty}^{+\infty} x^{2k+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$$

car l'intégrale est convergente et la fonction intégrée est impaire.

$$\begin{aligned} E(X^{2k}) &= \int_{-\infty}^{+\infty} x^{2k} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \\ &= \frac{1}{\sqrt{2\pi}} \left(\left[-\frac{x^{2k+1}}{2k+1} e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \frac{x^{2k+2}}{2k+1} e^{-\frac{x^2}{2}} dx \right). \end{aligned}$$

D'où $E(X^{2k+2}) = (2k+1)E(X^{2k})$. Comme $E(X^0) = 1$, on obtient le résultat cherché. \square

3 Théorèmes limites

Les théorèmes limites en probabilités concernent le comportement asymptotique des moyennes de variables aléatoires indépendantes. Dans cette section, $(X_n)_{n \geq 0}$ représente **une suite de variables aléatoires indépendantes, de même loi admettant une espérance m et une variance σ^2 .**

3.1 Loi des Grands nombres

Le premier résultat concerne l'évolution des moyennes elles mêmes.

Theorème 3.1 Avec les hypothèses de cette section, pour tout $\varepsilon > 0$, on a

$$\lim_{n \rightarrow +\infty} P\left(\left| \frac{X_1 + \dots + X_n}{n} - m \right| > \varepsilon\right) = 0.$$

Autrement dit, pour n grand, la moyenne empirique $\frac{X_1 + \dots + X_n}{n}$ est très proche de m .

Démonstration : Comme les X_i sont indépendantes et de même loi, on a $E(X_1 + \dots + X_n) = nm$ et $\text{Var}(X_1 + \dots + X_n) = n\sigma^2$.

Alors, par Bienaymé-Tchebitcheff,

$$\begin{aligned} P\left(\left| \frac{X_1 + \dots + X_n}{n} - m \right| > \varepsilon\right) &= P(|X_1 + \dots + X_n - nm| > n\varepsilon) \leq \\ &\leq \frac{1}{n^2\varepsilon^2} \text{Var}(X_1 + \dots + X_n) = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0. \quad \square \end{aligned}$$

La loi des grands nombres modélise le fait que la fréquence empirique est une bonne approximation de la probabilité. Si on lance 1000 fois une pièce, on obtiendra à peu près 500 fois pile et 500 fois face. Le théorème modélise cet à peu près.

Exercice 3.2 Soit $(X_i)_{i \geq 0}$ une suite de variables aléatoires indépendantes de même loi admettant une densité f sur \mathbb{R}^+ , i.e. $\int_0^{+\infty} f(t)dt = 1$ et une espérance $E(X) = \int_0^{+\infty} tf(t)dt$. Montrer que pour tout $A > 0$, on a $\lim_{n \rightarrow +\infty} P(S_n > A) = 1$.

3.2 Théorème central de la limite

La loi des grands nombres que nous venons de montrer décrit le comportement asymptotique de la moyenne empirique. Par contre, elle ne dit pas comment estimer la différence entre la moyenne empirique et l'espérance. Ceci est réalisé par le théorème suivant que nous admettrons.

Theorème 3.3 Avec les hypothèses de cette section, si $-\infty \leq a < b \leq +\infty$ sont deux nombres (éventuellement infinis), on a

$$P\left(a < \frac{(X_1 + \dots + X_n) - nm}{\sqrt{n}\sigma} < b\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

Autrement dit, pour n assez grand, $\frac{(X_1 + \dots + X_n) - nm}{\sqrt{n}\sigma}$ suit une loi qui ressemble beaucoup à la loi gaussienne.

Ce théorème sera fondamental dans de nombreuses applications statistiques.

4 Introduction aux statistiques

Jusqu'à présent, la situation que nous avons considérée était la suivante : la probabilité gouvernant une expérience aléatoire étant donnée (ou résultant d'une modélisation de l'expérience), il s'agissait de déterminer la probabilité d'événements résultant de cette expérience aléatoire.

Nous allons maintenant nous pencher sur une situation qui est en quelque sorte dans la direction inverse : on observe l'expérience aléatoire, et on essaye de déterminer la loi qui la gouverne.

4.1 Statistique paramétrique

Nous allons en fait nous limiter à une situation particulièrement favorable, celle de la *statistique paramétrique*. Formellement, elle se présente de la façon suivante : une famille de lois $(p_\theta)_{\theta \in \Theta}$ est donnée où Θ est un ensemble où le paramètre θ prend ses valeurs. On suppose que pour l'une (et une seule) valeur du paramètre θ notée θ_0 , la loi p_{θ_0} est celle qui gouverne effectivement l'expérience aléatoire. Mais θ_0 est inconnue et le problème est justement de le trouver en regardant les résultats de l'expérience aléatoire.

Considérons un exemple pratique qui nous accompagnera tout au long de notre démarche. Une usine produit un très grand nombre de pièces (qu'on approximera comme si ce nombre était infini), dont une certaine proportion $0 \leq \theta \leq 1$ inconnue est défectueuse. Quand le nombre total de pièces est infini, le fait d'en tirer une ne modifie pas la situation. De ce fait, on peut considérer que la probabilité de tirer une pièce défectueuse est θ et p_θ , probabilité qui gouverne le tirage d'une pièce défectueuse, est donc la loi de Bernoulli de paramètre θ avec $\theta \in \Theta = [0, 1]$.

Comment préciser le 'bon' θ_0 ? La façon naturelle de procéder est de prélever n pièces parmi la production, de les observer et d'en tirer des conclusions.

On pose la définition suivante :

Definition 4.1 *On appelle n -échantillon pour la loi μ un n -uplet de variables aléatoires (X_1, \dots, X_n) indépendantes et de loi μ .*

Pour réaliser notre recherche du 'bon' θ_0 , l'idée va être de réaliser n fois l'expérience aléatoire dans des conditions indépendantes, ce qui va fournir X_1, \dots, X_n variables aléatoires indépendantes de loi p_θ .

Pour estimer θ , il va falloir bricoler une valeur à partir de celles de X_1, \dots, X_n , sous la forme d'une fonction $\varphi(X_1, \dots, X_n)$. Une telle fonction sera appelée un *estimateur de la loi μ* . Si on observe $X_1 = x_1, \dots, X_n = x_n$, on estimera θ par $\varphi(x_1, \dots, x_n)$.

4.2 Estimateurs sans biais

Naturellement, si on prend une telle fonction sans aucune restriction, il n'y a aucune raison pour qu'elle donne une bonne estimation de θ . C'est pourquoi on introduit la notion de *biais*.

Definition 4.2 *Un estimateur sans biais est un estimateur X tel que pour toute valeur $\theta \in \Theta$, si θ est la vraie valeur du paramètre, l'espérance de X est égale à θ .*

X est donc un estimateur sans biais si quelle que soit la vraie situation, on a la garantie que, en moyenne, X l'approxime raisonnablement.

Reprenons notre exemple de production industrielle. Pour tout $\theta \in [0, 1]$, (X_1, \dots, X_n) est donc un n -échantillon de la loi de Bernoulli de paramètre $\theta \in [0, 1]$. On sait alors que $\sum_{i=1}^n X_i$ suit une loi binomiale $\mathcal{B}(n, \theta)$ et donc que $\bar{X} =$

$$\frac{1}{n} \sum_{i=1}^n X_i \text{ (moyenne empirique) suit une loi } \mathcal{B}(n, \frac{\theta}{n}).$$

De ce fait, $E_\theta(\bar{X}) = \theta$ et \bar{X} apparaît comme un estimateur sans biais de X .

La notation précédente E_θ sera employée pour stipuler qu'on fait les calculs en supposant formellement que la valeur du paramètre est θ .

Exercice 4.3 *Soit (X_1, \dots, X_n) un n -échantillon d'une loi μ d'espérance m et de variance σ^2 inconnues. Montrer que*

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}$$

et

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

sont respectivement des estimateurs sans biais de m et σ^2 (appelés respectivement la moyenne empirique et la variance empirique).

Reprenons notre exemple de production industrielle où X suit la loi binomiale $\mathcal{B}(n, \theta)$. On a montré que $\psi(X) = \frac{X}{n}$ est un estimateur sans biais de θ . En fait, montrons qu'il s'agit du seul. Soit en effet $\varphi(X)$ un autre esb. Posons $\alpha(X) = \varphi(X) - \psi(X)$.

Par hypothèse, on a pour tout $\theta \in [0, 1]$,

$$0 = E_\theta(\alpha(X)) = \sum_{k=0}^n \alpha(k) C_n^k \theta^k (1-\theta)^{n-k} = (1-\theta)^n \sum_{k=0}^n \alpha(k) C_n^k \left(\frac{\theta}{1-\theta}\right)^k.$$

Notons que si θ parcourt $]0, 1[$, $\frac{\theta}{1-\theta}$ parcourt tout \mathbb{R}_*^+ et donc on en déduit

$\sum_{k=0}^n C_n^k \alpha(k) u^k = 0, \forall u \in \mathbb{R}_*^+$. De ce fait, pour que ce polynôme en u soit nul, il faut que $\alpha(k) = 0, \forall 0 \leq k \leq n$.

On voit donc que la classe des estimateurs sans biais peut être extrêmement réduite. Ceci peut être ennuyeux car il se peut que les estimateurs en question soient moins pratiques que certains estimateurs légèrement biaisés. On peut de ce fait être amené à utiliser d'autres types d'estimateurs que les estimateurs sans biais.

4.2.1 Maximum de vraisemblance

Regardons un exemple élémentaire pour exposer le principe de la méthode.

Dans notre situation industrielle, supposons qu'on hésite uniquement entre deux valeurs possibles de la proportion de pièces défectueuses θ_0 : 1/100 ou 90/100. On tire donc une pièce au hasard suivant une loi de Bernoulli p_θ (la valeur du tirage est 1 si la pièce est défectueuse, 0 sinon). Le tirage donne 1. Il est logique alors de déduire que la valeur du paramètre est 90/100. Pourquoi ? Parce qu'on compare sous $p_{1/100}$ et $p_{90/100}$ la chance d'obtenir une pièce défectueuse (qui est ce qui a effectivement été observé) et on trouve qu'elle est plus grande si $\theta_0 = 90/100$ que si $\theta_0 = 1/100$.

Definition 4.4 On appelle fonction de vraisemblance du n -échantillon (X_1, \dots, X_n) de la loi discrète p_θ sur $E \subset \mathbb{R}$ la fonction définie pour $\theta \in \Theta$ et $x \in E^n$ par

$$\begin{aligned} L(\theta, x) &= P((X_1, \dots, X_n) = (x_1, \dots, x_n)) = P(X_1 = x_1, \dots, X_n = x_n) = \\ &= \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n p_\theta(x_i). \end{aligned}$$

$L(\theta, x)$ représente donc la probabilité d'obtenir le point x qu'on a effectivement observé au tirage de (X_1, \dots, X_n) en supposant que la loi soit donnée par p_θ .

Definition 4.5 Un estimateur φ de θ s'appelle un estimateur du maximum de vraisemblance si pour tout $x \in E$,

$$L(\varphi(x), x) = \sup_{\theta \in \Theta} L(\theta, x).$$

Exemple 4.6 Reprenons notre exemple de production où θ est la proportion inconnue. Les observations sont x_1, x_2, \dots, x_n (x_i vaut 0 ou 1). On a $p_\theta(1) = \theta$ et $p_\theta(0) = 1 - \theta$.

$$L(\theta, x) = \prod_{i=1}^n p_\theta(x_i) = \theta^k (1 - \theta)^{n-k}$$

si $k = \sum_{i=1}^n x_i$.

Pour déterminer le maximum éventuel, on dérive $L(\theta, x)$:

$$\frac{d}{d\theta} L(\theta, x) = k\theta^{k-1}(1 - \theta)^{n-k} - \theta^k(n - k)(1 - \theta)^{n-k-1}.$$

On a

$$\frac{d}{d\theta} L(\theta, x) = 0 \iff k(1 - \theta) - (n - k)\theta = 0 \iff k - n\theta = 0 \iff \theta = \frac{k}{n}.$$

Il s'agit bien d'un maximum puisque $L(0, x) = L(1, x) = 0$ et $L(\theta, x) \geq 0$. On retrouve ainsi l'estimateur

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

Exemple 4.7 Un signal aléatoire peut à chaque instant être allumé (probabilité θ) ou éteint (probabilité $1 - \theta$) avec $0 < \theta < 1$ inconnu. A chaque instant, l'allumage est indépendant des autres instants. Pour estimer θ , on observe les intervalles de temps séparant les n premières fois où le signal est allumé. Notons X_1, X_2, \dots, X_n ces n premiers intervalles. Clairement, les X_i sont indépendantes et de même loi (le prouver en exercice) et on a déjà vu que cette loi est géométrique de paramètre $\theta : P(X_i = k) = \theta(1 - \theta)^{k-1}, k \geq 1$.

Donc (X_1, \dots, X_n) est un n -échantillon de la loi en question. On observe des intervalles k_1, \dots, k_n .

$$L(\theta, k_1, \dots, k_n) = \prod_{i=1}^n \theta(1 - \theta)^{k_i-1} = \alpha^n (1 - \theta)^{\sum_{i=1}^n k_i - n}.$$

Posons $s = \sum_{i=1}^n k_i : L(\theta, k_1, \dots, k_n) = \alpha^n (1 - \theta)^{s-n}$. Donc

$$\frac{\partial}{\partial \theta} L(\theta, k_1, \dots, k_n) = n\alpha^{n-1}(1 - \theta)^{s-n} - (s - n)\alpha^n(1 - \theta)^{s-n-1} = 0$$

$$\iff n(1 - \theta) - (s - n)\theta = 0 \iff n - s\theta = 0 \iff \theta = \frac{n}{s}.$$

Il s'agit clairement d'un maximum.

On a donc comme estimateur du maximum de vraisemblance

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

4.2.2 Utilisation du théorème central de la limite

Le théorème central de la limite peut permettre de répondre à un certain nombre de situations d'estimation.

Reprenons notre production d'objets. X_n est une variable qui vaut 1 si la n -ième pièce tirée est défectueuse, 0 sinon. On suppose les X_n indépendants, de même loi de Bernoulli $\mathcal{B}(\theta)$ (espérance θ , variance $\theta(1 - \theta)$) et on désire savoir combien de temps il faut attendre pour que la probabilité que l'on ait tiré au moins 50 pièces défectueuses soit supérieure à 0.95

On veut donc trouver n tel que $P(X_1 + \dots + X_n \geq 50) \geq 0.95$. On suppose n assez grand pour que, d'après le théorème central de la limite, $\frac{(X_1 + \dots + X_n) - n\theta}{\sqrt{n\theta(1 - \theta)}}$ soit 'raisonnablement' de loi gaussienne.

$$\begin{aligned} P(X_1 + \dots + X_n \geq 50) &= P(+\infty \geq \frac{(X_1 + \dots + X_n) - n\theta}{\sqrt{n\theta(1 - \theta)}} \geq \frac{50 - n\theta}{\sqrt{n\theta(1 - \theta)}}) \\ &\simeq \frac{1}{\sqrt{2\pi}} \int_{\frac{50 - n\theta}{\sqrt{n\theta(1 - \theta)}}}^{+\infty} e^{-\frac{x^2}{2}} dx. \end{aligned}$$

Une table de la loi gaussienne montre que $0.95 = \frac{1}{\sqrt{2\pi}} \int_{\rho}^{+\infty} e^{-\frac{x^2}{2}} dx$ avec $\rho = -1,645$. On prend donc $\frac{50 - n\theta}{\sqrt{n\theta(1 - \theta)}} \leq -1.645$ soit $n\theta - 1.645\sqrt{n\theta(1 - \theta)} \geq 50$

ce qui permet de déterminer n en fonction de θ en résolvant cette (in)équation du second degré.