Institut National de la Statistique et des Etudes Economiques

# ÉCOLE NATIONALE DE LA STATISTIQUE ET DE L'ADMINISTRATION ÉCONOMIQUE

## ELEMENTS DE STATISTIQUE MATHEMATIQUE

par

Ph. TASSI

DIRECTION GÉNÉRALE

3, Avenue Pierre Larousse - 92240 MALAKOFF Téléphone : 540-12-12

# ELEMENTS DE STATISTIQUE MATHEMATIQUE

par

Ph. TASSI

1981.

# TABLE DES MATIERES

# CHAPITRE I

## Rappels sur les lois de probabilités

------

Ce premier chapitre n'est fait que pour vous familiariser une nouvelle fois avec les lois de probabilité usuelles, que vous rencontrerez fréquemment en Statistique Mathématique. On notera $L_X$ la loi de la v.a. x, $f_X$ sa densité (prise soit par rapport à la mesure de comptage, soit par rapport à la mesure de Lebesgue sur $\mathbb{R}$), $F_X$ sa fonction de répartition.

### 1 - Loi de Bernoulli

$L_X = B(1, p)$

C'est une loi à valeurs sur $\{0,1\}$

$$\begin{cases} P(X=0) = 1 - p = q \\ P(X=1) = p \end{cases}$$

$P(X = x) = p^x (1-p)^{1-x}$ $\qquad$ x=o ou 1

$E(X) = p$

$V(X) = pq$

### 2 - Loi de Poisson

$L_X = \mathcal{P}(\lambda)$ $\qquad$ $\lambda \in \mathbb{R}^+$ est le paramètre de la loi

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

C'est une v.a. entière

$E(X) = \lambda$

$V(X) = \lambda$

**Théorème**

Soient X et Y deux v.a. indépendantes suivant des lois de Poisson de paramètres respectifs $\lambda$ et $\mu$ . Alors la v.a. $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$ .

## 3 - Loi binomiale

$L_X = B(n,p)$

C'est la somme de n v.a. de Bernoulli indépendantes.

$P(X=x) = C_n^x \; p^x \; (1-p)^{n-x} \qquad x \in \{0,1,\ldots,n\}$

$E(X) = np$

$V(X) = npq$

**Théorème 1 :**

Si X suit une loi $B(n,p)$ et Y une loi $B(m,p)$ et si X et Y sont indépendantes, alors $X+Y$ suit une loi $B(n+m,p)$.

**Théorème 2 :**

$L_X = B(n,p)$

Si $n \to +\infty$ , $p \to 0$, $np \to \lambda$, $\lambda$ fini, alors X converge en loi vers la loi de Poisson de paramètre $\lambda$.

.../...

Théorème 3 :

$L_X = B\ (n,p)$

Si $n \to +\infty$, $\frac{X - np}{\sqrt{npq}}$ converge en loi vers la loi normale centrée réduite N (0,1).

4 - Loi normale

$L_X = N\ (m,\sigma)$

Sa densité est :

$$f_X\ (x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad x \in \mathbb{R}$$

$E\ (X) = m$

$V\ (X) = \sigma^2$

Théorème

Soient X et Y deux v.a. indépendantes normales de lois $N\ (m_1, \sigma_1)$ et $N\ (m_2, \sigma_2)$. Alors X+Y est une loi normale $N\ (m_1+m_2, \sqrt{\sigma_1^2+\sigma_2^2}\ )$.

Application :

$X_1, \ldots, X_n$ échantillon indépendant d'une variable normale $N\ (m,\sigma)$.

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

$L_{\bar{X}} = N\ (m, \frac{\sigma}{\sqrt{n}})$

.../...

## 5 - Loi du chi-deux

**Définition** : X suit une loi $N(0,1)$. Alors $Y = X^2$ suit une loi du $\chi^2$ à 1 degré de liberté.

**Densité** : $$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} y^{-\frac{1}{2}}$$

**Définition** : Soient $X_1, \ldots, X_n$ n v.a. indépendantes suivant une loi $N(0,1)$, $Z = \sum_{i=1}^{n} X_i^2$ suit une loi du $\chi^2$ à n degrés de liberté.

$E(Z) = n$

$V(Z) = 2n$

**Théorème**

Soient X et Y deux v.a. indépendantes suivant respectivement deux lois $\chi^2(n)$ et $\chi^2(m)$, la variable $X+Y$ suit une loi $\chi^2(n+m)$.

## 6 - Loi gamma

Une variable aléatoire réelle X suit une loi $\gamma(\alpha,\theta)$ si sa densité est donnée par l'expression :

$$\boxed{f(x) = \frac{\theta^{\alpha}}{\Gamma(\alpha)} e^{-x\theta} x^{\alpha-1} \qquad x>0, \quad \alpha>0, \quad \theta>0}$$

.../...

$\alpha$ et $\theta$ sont deux paramètres de $\mathbb{R}^+$. $\theta$ s'appelle paramètre d'échelle.

Lorsque $\theta = 1$, on notera $\gamma(\alpha)$ au lieu de $\gamma(\alpha,1)$. $\Gamma(\alpha)$ est l'intégrale eulérienne de seconde espèce :

$$\boxed{\Gamma(\alpha) = \int_0^{+\infty} e^{-u}\, u^{\alpha-1}\, du}$$

Propriétés de $\Gamma(\alpha)$ :

$$\forall \alpha > 0 \qquad \Gamma(\alpha+1) = \alpha\, \Gamma(\alpha)$$

$$\forall n \in \mathbb{N} \qquad \Gamma(n+1) = n\,!$$

$$\Gamma(p) \sim \frac{1}{p} \quad \text{si} \quad p \to 0$$

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma\left(\tfrac{3}{2}\right) = \frac{\sqrt{\pi}}{2}$$

Caractéristiques de la loi $\gamma(\alpha)$ :

$$\boxed{\begin{aligned} E(X) &= \alpha \\ V(X) &= \alpha \\ E(X^r) &= \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} \end{aligned}}$$

Théorème 1 :

Soient X et Y deux variables aléatoires indépendantes suivant respectivement les lois $\gamma(\alpha)$ et $\gamma(\beta)$. Alors la variable X+Y suit la loi $\gamma(\alpha+\beta)$.

.../...

Théorème 2 :

Soit X une variable suivant une loi du chi-deux à n degrés de libertés.

La variable $\frac{1}{2}$ X suit une loi $\gamma(\frac{n}{2})$.

Ce théorème est très utilisé pour retrouver la densité d'une loi du chi-deux.

## 7 - Loi bêta

1°) Une variable aléatoire suit une loi bêta du type I si sa densité est donnée par :

$$\boxed{f(x) = \frac{1}{B(p,q)} x^{p-1} (1-x)^{q-1} \qquad o \leqslant x \leqslant 1 \ , \quad p > o \ , \quad q > o}$$

B (p,q) est l'intégrale eulérienne de première espèce de type I.

$$\boxed{B(p,q) = \int_o^1 x^{p-1} (1-x)^{q-1} \, dx}$$

Propriétés de B (p,q)

$$B(p,q) = B(q,p)$$

$$B(p,q) = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}$$

$$B(\tfrac{1}{2}, \tfrac{1}{2}) = \pi$$

$$B(\tfrac{3}{2}, \tfrac{1}{2}) = \frac{\pi}{2}$$

.../...

Caractéristiques de la loi $\beta$ (p,q)

$$E(X) = \frac{B\ (p+1,q)}{B\ (p,q)}$$

$$E(X^2) = \frac{B\ (p+2,q)}{B\ (p,q)}$$

$$E(X^r) = \frac{B\ (p+r,q)}{B\ (p,q)}$$

D'où les moments :

$$\boxed{E(X) = \frac{p}{p+q}} \qquad \boxed{E(X^2) = \frac{p+1}{p+q+1}} \qquad \boxed{V(X) = \frac{p\,q}{(p+q+1)(p+q)^2}}$$

2°) Loi bêta du type II

On a vu : $B(p,q) = \int_0^1 x^{p-1}\ (1-x)^{q-1}\ dx$

Procédons au changement de variable $x \to \frac{u}{1+u}$

On obtient :

$$B\ (p,q) = \int_0^{+\infty} \frac{u^{\ p-1}}{(1+u)^{p+q}}\ du$$

Sous cette forme, on obtient l'intégrale eulérienne de première espèce de type II.

Une variable aléatoire X suit une loi bêta du type II si sa densité est donnée par :

$$\boxed{f(x) = \frac{1}{B(p,q)} \quad \frac{x^{p-1}}{(1+x)^{p+q}} \qquad x > 0}$$

<u>Caractéristiques de la loi $\beta(p,q)$ II</u> :

$$\boxed{E(X) = \frac{p}{q-1}} \qquad q>1$$

$$\boxed{V(X) = \frac{p(p+q-1)}{(q-1)^2 (q-2)}} \qquad q>2$$

3°) <u>Quelques théorèmes</u>

(La démonstration en est laissée au lecteur)

<u>Théorème 1</u> :

Soient X et Y deux variables aléatoires indépendantes suivant respectivement des lois gamma $\gamma(p)$ et $\gamma(q)$. Alors la variable aléatoire $\frac{X}{Y}$ suit une loi bêta $\beta(p,q)$ du type II.

<u>Théorème 2</u> :

Soient X et Y deux variables aléatoires indépendantes suivant respectivement des lois gamma $\gamma(p)$ et $\gamma(q)$. Alors la variable aléatoire $\frac{X}{X+Y}$ suit une loi bêta du type I $\beta(p,q)$.

## 8 - Loi de Student

Soient deux variables aléatoires indépendantes X et Y, X suivant une loi normale N(0,1) et Y une loi du $\chi^2(n)$ ; alors la variable aléatoire définie par le rapport $\dfrac{X}{\sqrt{\frac{Y}{n}}}$ suit une loi de Student à n degrés de liberté, notée $T_n$.

Densité

$$L_X = T\ (n)$$

$$\boxed{f_X\ (x) = \frac{1}{\sqrt{n}\ B\ (\frac{1}{2}, \frac{n}{2})}\ (1+\frac{x^2}{n})^{-\frac{1}{2}(n+1)}}$$

## 9 - Loi de Fisher

Soient X et Y deux variables aléatoires indépendantes suivant respectivement un $\chi^2(p)$ et $\chi^2(q)$ ; alors la variable aléatoire définie par le rapport $\frac{X}{p} : \frac{Y}{q}$ suit une loi de Fisher à p et q degrés de libertée, notée F (p,q).

Densité :

$$L_X = F\ (p,q)$$

$$\boxed{f_X\ (x) = \frac{p^{p/2}\ q^{q/2}}{B\ (p/2, q/2)}\ \frac{x^{p/2\ -\ 1}}{(q+px)^{\frac{p+q}{2}}}} \quad (x>o)$$

## 10 - La famille exponentielle

Il s'agit d'une famille de lois de probabilités, englobant la plupart des lois précédentes, dont l'importance statistique est réelle. En effet, ainsi que nous le verrons par la suite, des résultats généraux existent, pour cette famille, en théorie de l'estimation, et en tests.

### Définition

Une loi de probabilité $P_\theta$, de densité $f(x,\theta)$, est dite appartenir à la famille exponentielle s'il existe des fonctions $h(x)$, $c(\theta)$, $Q_j(\theta)$ et $T_j(x)$ telles que :

$$f(x, \theta) = c(\theta)\ h(x)\ \exp\left[\sum_{j=1}^{r} Q_j(\theta)\ T_j(x)\right]$$

ou : $$\text{Log } f(x, \theta) = \beta(\theta) + b(x) + \sum_{j=1}^{r} Q_j(\theta)\ T_j(x)$$

### Remarque

On peut mettre la famille exponentielle sous forme naturelle par le nouveau paramétrage :

$$\lambda_j = Q_j(\theta)$$

Exemples :

. Loi binomiale :

$$f(x,p) = C_n^x \, p^x \, (1-p)^{n-x}$$

$$f(x,p) = C_n^x \, (1-p)^n \, \left(\frac{p}{1-p}\right)^x = C_n^x \, (1-p)^n \, e^{x \,\mathrm{Log}\, \frac{p}{1-p}}$$

d'où

$$T(x) = x$$

$$h(x) = C_n^x$$

$$c(p) = (1-p)^n \qquad T(x) = x$$

$$Q(p) = \mathrm{Log}\, \frac{p}{1-p} \qquad h(x) = \frac{1}{x!}$$

. Loi de Poisson :

$$f(x,\lambda) = e^{-\lambda} \, \frac{\lambda^x}{x!}$$

$$c(\lambda) = e^{-\lambda} \qquad T(x) = x$$

$$Q(\lambda) = \mathrm{Log}\ \lambda \qquad h(x) = \frac{1}{x!}$$

. Loi normale :

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

$$= \frac{1}{\sqrt{2\pi}} \, \sigma^{-1} \, e^{-\frac{m^2}{2\sigma^2}} \, e^{-\frac{x^2}{2\sigma^2} + \frac{mx}{\sigma^2}}.$$

...

$T_1(x) = x^2$

$T_2(x) = x$

$h(x) = 1$

$c(m,\sigma) = \sigma^{-1} \; e^{-m^2/2\sigma^2}$

$Q_1(m,\sigma) = -\dfrac{1}{2\sigma^2}$

$Q_2(m,\sigma) = \dfrac{m}{\sigma^2}$

. <u>Loi exponentielle</u> :

$$f(x,\theta,p) = \frac{1}{\Gamma(p)} \theta^p \, e^{-\theta x} \, x^{p-1} \, \mathbb{1}_{R^+}(x)$$

$T_1(x) = x$

$T_2(x) = \text{Log } x$

$h(x) = \mathbb{1}_{R^+}(x)$

$c(\theta,p) = \dfrac{\theta^p}{\Gamma(p)}$

$Q_1(\theta,p) = -\theta$

$Q_2(\theta,p) = p-1$

. <u>Loi uniforme</u> :

$$f(x,\theta) = \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x)$$

Ne se met pas sous la forme exponentielle.

. <u>Loi de Cauchy</u> :

$$f(x, \theta) = \frac{1}{\pi} \quad \frac{1}{1+(x-\theta)^2}$$

Ne se met pas sous la forme exponentielle.

# ANNEXE

## LES CARACTERISTIQUES DE SYMETRIE ET D'APLATISSEMENTS

Soit X une variable aléatoire.

On notera respectivement $m_k$ et $\mu_k$ les moments non centrés et centrés de X :

$$m_k = E(X^k)$$

$$\mu_k = E(X - m)^k$$

### 1 - COEFFICIENTS DE PEARSON

$$\boxed{\begin{aligned} \beta_1 &= \frac{\mu_3^2}{\mu_2^3} \\ \beta_2 &= \frac{\mu_4}{\mu_2^2} \end{aligned}}$$

Lorsqu'une distribution est symétrique, les moments centrés d'ordre impair sont nuls. $\beta_1$ joue donc le rôle d'une mesure du degré de symétrie ; $\beta_2$ représente le degré d'aplatissement de la loi étudiée, en un sens précisé plus loin.

Pratiquement, ces coefficients ne sont guère utilisés, et on leur préfère :

### 2 - COEFFICIENTS DE FISHER

$$\boxed{\begin{aligned} \gamma_1 &= \frac{\mu_3}{\mu_2^{3/2}} \\ \gamma_2 &= \beta_2 - 3 \end{aligned}}$$