Chap 1 : Modèles et problèmes statistiques du cours -

Notations : . données n nombres $x_1, x_2, \dots, x_n$

· Réalisations d'un n-échantillon : n variables aléatoires indépendantes de même loi $X_1, X_2, \dots, X_n$.

· La loi commune appartient à une famille de lois paramétrées par un paramètre inconnue . $\theta \in \Theta$

1.1. Modèles paramétriques.

a. Bernoulli :

- sondage sur un vote à deux issues.
- qualité dans des pièces manufacturées (avec ou sans défaut).
- étude d'inondation (un fleuve va-t-il déborder ?).
- un patient est-il malade ou pas ?
- un enfant qui naît $\begin{cases} \to G \\ \to F \end{cases}$.

$n \in \mathbb{N}^*$, $x_1, \dots, x_n$. observations $\in \{0;1\}$.

$X_1, \dots, X_n$ n-échantillon $\mathcal{B}(\theta)$ $\theta \in]0;1[$.

x_i	0	1	Total
probabilités	$1-\theta$	θ	1

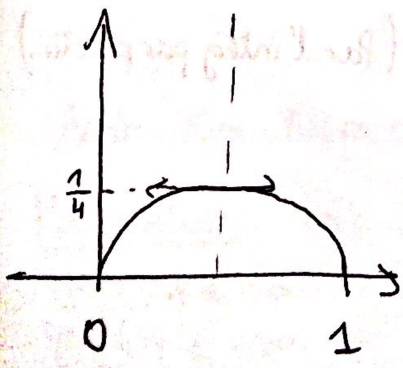
$P_\theta(X_1 = 1) = \theta$

$P_\theta(X_1 = 0) = 1-\theta$

$E_\theta(X_1) = \theta \; (= 0 \times P_\theta(X_1=0) + 1 \times P_\theta(X_1=1))$

$V_\theta(X_1) = \theta \times (1-\theta)$.

$$V_\theta(X_1) = E_\theta\left((X_1 - E_\theta(X_i))^2\right) = E_\theta(X_i^2) - (E_\theta(X_i))^2 = \theta - \theta^2 = \theta(1-\theta).$$



b) Poisson.

$X_1, \dots, X_n$. $n \in \mathcal{B}(\theta)$, $\theta \in]0;1[$.
n grand θ petit

$X_1 + \dots + X_n \sim \mathcal{B}(n, \theta)$ $\{0, \dots, n\}$. lorsque $n \to +\infty$, $\theta_n = \frac{\lambda}{n}$.

$P_\theta(X_1 + \dots + X_n = k) = \dots$

$Y \sim \mathcal{P}(\lambda)$, $P_\lambda(Y = k) = e^{-\lambda} \times \frac{\lambda^k}{k!}$

$\left(1 - \frac{\lambda}{n}\right)^n \xrightarrow[n \to +\infty]{} e^{-\lambda}$.

$E_\theta(X_1 + \dots + X_n) = n \times \theta = \lambda$. donc $E_\lambda(Y) = \lambda$.

$V_\theta(X_1 + \dots + X_n) = n \times \theta(1 - \theta)$ $V_\lambda(Y) = \lambda$.

$= \lambda \times \left(1 - \frac{\lambda}{n}\right) \xrightarrow[n \to +\infty]{} \lambda$.

$m \in \mathbb{N}^*$, $Y_1, \dots, Y_m$: m-échantillon $\mathcal{P}(\lambda)$ de $\lambda \in \mathbb{R}_+^*$.

$y_1, \dots, y_m$: observations vivant dans $\mathbb{N}$.

$P_\lambda(Y_1 = k) = e^{-\lambda} \times \frac{\lambda^k}{k!}$, $E_\lambda(Y_1) = \lambda$

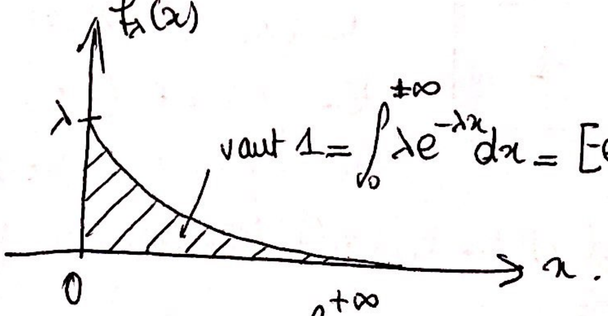
$V_\lambda(Y_1) = \lambda$.

c) Loi exponentielle :

Paramètre : $\lambda > 0$: Loi continue à densité. Valeurs possibles : $[0; +\infty[$ ou $]0; +\infty[$.

Densité : soit $x \geq 0$.

$f_\lambda(x) = \lambda \times e^{-\lambda x}$

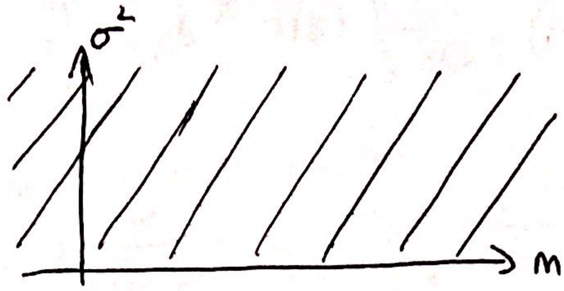


vaut $1 = \int_0^{+\infty} \lambda e^{-\lambda x} dx = \left[e^{-\lambda x}\right]_0^{+\infty}$.

$X_1, \dots, X_n$ n échantillon $\mathcal{E}(\lambda)$. $E_\lambda(X_1) = \int_0^{+\infty} x \times f_\lambda(x) dx = \frac{1}{\lambda}$ (Par l'intég par partie).

$V_\lambda(X_1) = \frac{1}{\lambda^2}$

d) Le modèle gaussien.

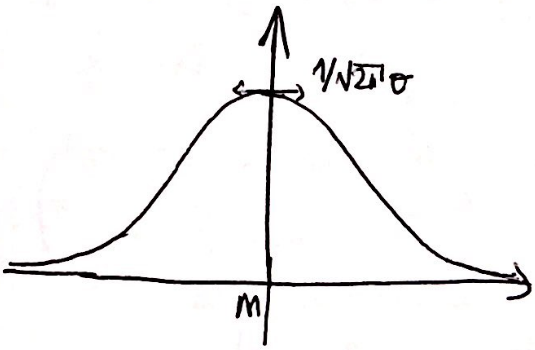


Modèle continue à densité.

$X \sim \mathcal{N}(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_+^*$.

Densité :

$x \in \mathbb{R}$, $\theta = (m, \sigma^2)$: $f_\theta(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-m)^2\right)$.



$E(|X|) < +\infty$.

$\Downarrow$

$E_\theta(X) = m$.

$V_\theta(X) = \sigma^2$.

Propriétés :

- Soit $X \sim \mathcal{N}(m, \sigma^2)$. Alors $Y \equiv \frac{X-m}{\sigma} \sim \mathcal{N}(0;1)$.
- Soit $Y \sim \mathcal{N}(0,1)$. Alors, pour $\sigma > 0$, $m \in \mathbb{R}$, $X \equiv \sigma Y + m \sim \mathcal{N}(m, \sigma^2)$.
- Soit $X \sim \mathcal{N}(m, \sigma^2)$. Soit $Y = \frac{x-m}{\sigma}$. Soit $t \in \mathbb{R}$,

$P(Y \leq t) = \cancel{P\left(\frac{X-m}{\sigma}\right)}\ P\left(\frac{X-m}{\sigma} \leq t\right) = P(X \leq t \times \sigma + m)$.

$= \int_{-\infty}^{m+t\sigma} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-m)^2\right) dx$.

Posons $y = \frac{x-m}{\sigma}$. $dy = \frac{dx}{\sigma}$

$P(Y \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy = P(\mathcal{N}(0,1) \leq t)$.

Soit $X \sim \mathcal{N}(m, \sigma^2)$ alors $Y = \frac{X-m}{\sigma} \sim \mathcal{N}(0,1)$. (et vice-versa).

Proposition : soit $Y \sim \mathcal{N}(0,1)$.

$E(|Y|) < +\infty$ $\qquad$ $E(Y) = 0$

$E(Y^2) < +\infty$ $\qquad$ $V(Y) = 1$

$$E(|Y|) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |y| e^{-\frac{y^2}{2}} dy = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} \underbrace{y e^{-\frac{y^2}{2}}}_{-y^2/2} dy = \frac{2}{\sqrt{2\pi}} \left[-e^{-\frac{y^2}{2}} \right]_0^{+\infty} = \frac{2}{\sqrt{2\pi}} < +\infty$$

$u = e^{-y^2/2}$
$u' = -y e^{-y^2/2}$ $\quad y \times y e^{-\frac{y^2}{2}}$.

$$E(Y^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}} dy = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} y^2 e^{-\frac{y^2}{2}} dt$$

$u = y \quad v' = y e^{-y^2/2}$ Integ par parties.
$u' = 1 \quad v = -e^{-y^2/2}$

$$= \frac{2}{\sqrt{2\pi}} \left([uv]_0^{+\infty} - \int_0^{+\infty} u' v \right)$$

$$= \frac{2}{\sqrt{2\pi}} \left(\left[-y e^{-\frac{y^2}{2}} \right]_0^{+\infty} + \int_0^{+\infty} e^{-\frac{y^2}{2}} dy \right)$$

$$= \frac{2}{\sqrt{2\pi}} \left(0 + \int_0^{+\infty} e^{-\frac{y^2}{2}} dy \right)$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} e^{-\frac{y^2}{2}} dy$$

$$E(Y^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy = 1 < +\infty$$

Exercice : $E(Y^4) = 3$

$$E(Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y e^{-\frac{y^2}{2}} dy = 0.$$

$$V(Y) = E\left((Y - E(Y))^2 \right) = E(Y^2) = 1.$$

Exercice: mq $E(Y^4) = 3$.

$$E(Y^4) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^4 e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^3 \cdot y e^{-y^2/2} dy.$$

Intégration par parties : $u' = y e^{-y^2/2} \quad u = -e^{-y^2/2}$
$v = y^3 \quad v' = 3y^2$.

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^3 \cdot y e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} \left(-\left[y^3 e^{-y^2/2} \right]_{-\infty}^{+\infty} + 3 \int_{-\infty}^{+\infty} y^2 e^{-y^2/2} dy \right).$$

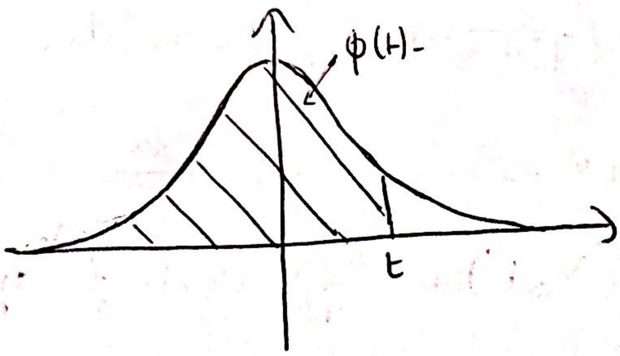
$$= \frac{1}{\sqrt{2\pi}} \left(3 \int_{-\infty}^{+\infty} y^2 e^{-y^2/2} dy \right)$$

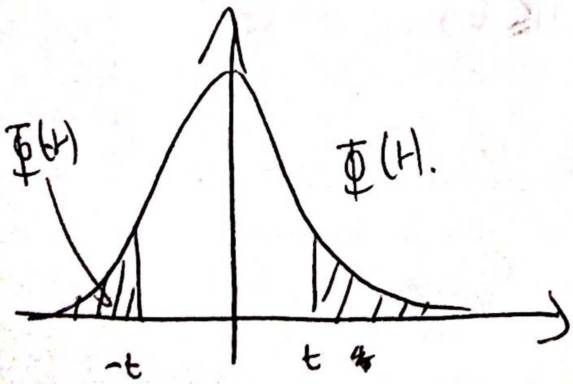
$$= \frac{1}{\sqrt{2\pi}} \left(3 \int_{-\infty}^{+\infty} e^{-y^2/2} dy \right) \quad (\text{cf} : E(Y^2))$$

$$= 3 < +\infty$$

Formule de répartition de Y.

$$\phi(t) = P(Y \leq t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{y^2}{2}} dy.$$





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

$P(a \leq Y \leq b) = P(a \leq Y < b) = \Phi(b) - \Phi(a)$.

1.2. Décision statistique, coût, risque.

$X_1, \dots, X_n$ n-échantillon f_θ, $x_1, \dots, x_n$ les n observations

Je me pose une question qui me concerne θ → estimation ponctuelle
→ estimation par intervalle de θ.
→ t est $(\theta) > 1/2$ ou pas ?

Je veux y répondre à partir de ce que j'observe et que je connais.

d : décision $d(x_1, \dots, x_n)$

Fonction de coût qui me quantifie chaque erreur.

Ex : Estimation

$$L(d(x_1, \dots, x_n), \theta) \geq 0$$
$$L_1(d(x_1, \dots, x_n), \theta) = |d(x_1, \dots, x_n) - \theta|.$$
$$L_2(d(x_1, \dots, x_n), \theta) = (d(x_1, \dots, x_n) - \theta)^2$$

$$E_\theta\left(L(d(X_1, \dots, X_n), \theta)\right) = \underbrace{R(d, \theta)}_{\text{risque}}$$

Ex : Bernoulli $\mathcal{B}(\theta)$.

$x_1, \dots, x_n$ $\quad x_1 \begin{cases} 1 \text{ si Face} \\ 0 \text{ si Pile.} \end{cases}$

$$\frac{x_1 + \dots + x_n}{n} = \bar{x}_n = d(x_1, \dots, x_n).$$

$$R(\bar{x}_n, \theta) = E_\theta\left[(\bar{X}_n - \theta)^2\right].$$