Soutien PBS 1 — Lois à densité

Loi à densité

Une variable aléatoire est continue X lorsque F_X a une dérivée f_X. Alors

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

En particulier, $\int_{\mathbb{R}} f_X(x)dx = 1$ et $\forall a \in \mathbb{R}$ on a $P(X = a) = 0$.

On a aussi la formule

$$f_{aX+b}(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right).$$

Application

Un intervalle central bilatéral à 95% est un intervalle $[a, b]$ tel que :

$$P(X \in [a, b]) = 0,95 \quad \text{avec} \quad P(X < a) = 0,025 \quad \text{et} \quad P(X > b) = 0,025.$$

Par exemple pour $Z \sim \mathcal{N}(0, 1)$, on a $I_{95\%}(X) = [-1,96\,;\,1,96]$.

Pour $X \sim \mathcal{N}(\mu, \sigma^2)$, on se ramène à $Z \sim \mathcal{N}(0, 1)$ via $Z = \frac{X-\mu}{\sigma}$.
On utilise le résultat de Z :

$$-1,96 \leq \frac{X-\mu}{\sigma} \leq 1,96 \iff \mu - 1,96\sigma \leq X \leq \mu + 1,96\sigma$$

Et on en déduit que $I_{95\%}(X) = [\mu - 1,96\sigma\,;\,\mu + 1,96\sigma]$.

EXERCICE 1

Soit $\lambda > 0$ fixé et $X \sim \mathcal{E}(\lambda)$ suivant une loi exponentielle de densité :

$$f_X(x) = \begin{cases} k\,e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer $k \in \mathbb{R}$ pour que f_X soit bien une densité de probabilité.
2. Déterminer $I_{95\%}(X)$, c'est-à-dire $[a, b]$ tel que $P(X < a) = 0,025$ et $P(X > b) = 0,025$.

Espérance

L'espérance est la valeur que l'on peut attendre "en moyenne" sur un grand nombre d'expériences

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

Pour tout $a \in \mathbb{R}$ et toutes X et Y, même si X et Y sont dépendantes

$$\mathbb{E}[aX] = a\mathbb{E}[X] \qquad \text{et} \qquad \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Variance

La variance mesure l'écart moyen des valeurs par rapport à la moyenne

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

Si X et Y sont indépendantes

$$V[aX] = a^2 V[X] \qquad \text{et} \qquad V[X+Y] = V[X] + V[Y]$$

Ces formules sont fausses si X et Y ne sont pas indépendantes (considérer le cas extrême où $Y = X$)

EXERCICE 2

Soient $X_1, X_2, \ldots, X_n$ des variables indépendantes et identiquement distribuées.
On note μ leur espérance commune ($\mathbb{E}[X_i] = \mu$) et σ^2 leur variance commune ($V(X_i) = \sigma^2$).
On définit la variable somme S_n et la moyenne empirique $\overline{X}_n$ par :

$$S_n = X_1 + X_2 + \cdots + X_n \quad \text{et} \quad \overline{X}_n = \frac{S_n}{n}$$

1. Exprimer l'espérance $\mathbb{E}[S_n]$ et la variance $V(S_n)$ en fonction de n, μ et σ^2. Justifier.
2. En déduire l'espérance $\mathbb{E}[\overline{X}_n]$, la variance $V(\overline{X}_n)$ et l'écart-type.
3. En utilisant la question 1, retrouver les formules de l'espérance et de la variance de la loi binomiale.

Inégalité de Bienaimé-Tchebychev

Elle contrôle l'écart à la moyenne. Si la variance est petite, il est improbable de s'éloigner trop de l'espérance.
Pour tout $\varepsilon > 0$

$$P\big(|X - \mathbb{E}[X]| \geq \varepsilon\big) \leq \frac{V(X)}{\varepsilon^2}$$

Elle a une conséquence très forte : la convergence en probabilités.
On reprend les notations de l'EXERCICE 2. Pour tout $\varepsilon > 0$

$$\lim_{n \to +\infty} P\big(|\overline{X}_n - \mu| \geq \varepsilon\big) = 0$$

La probabilité d'un écart, même infime, finit donc par s'annuler totalement.

EXERCICE 3

1. Rappeler la formule de la variance σ^2 d'une loi $Ber(p)$ en fonction de p.
2. Montrer que pour tout réel $p \in [0,1]$, on a $\sigma^2 \leq \frac{1}{4}$.
3. En déduire un intervalle $[a,b]$ tel que

$$\mathbb{P}(a \leq \overline{X}_n \leq b) \ \geq \ 95\%$$

Indication : utiliser les résultats de l'EXERCICE 2.

EXERCICE 4 — Synthèse

Soit X une variable aléatoire de densité :

$$f_X(x) = \begin{cases} k\,x & \text{si } x \in [0,1] \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer $k \in \mathbb{R}$, puis donner la densité f_Y de la variable $Y = 2X + 1$.
2. Déterminer l'intervalle central $I_{95\%}(X)$.
3. Calculer l'espérance $\mathbb{E}[X]$ et la variance $V(X)$.
4. Soient $X_1, \ldots, X_n$ des copies indépendantes de X et $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Donner $\mathbb{E}[\overline{X}_n]$ et $V(\overline{X}_n)$, puis déduire de l'inégalité de Bienaymé-Tchebychev un intervalle $[a, b]$ tel que $P(a \leq \overline{X}_n \leq b) \geq 95\%$.