Chapitre 6

Variables aléatoires réelles admettant une densité

6.1 Introduction

La définition d'une probabilité sur $\mathbb{R}$ dans le sens le plus général est basée sur la théorie de la mesure et sera traitée dans les années ultérieures.

Nous allons nous intéresser ici au cas particulier beaucoup plus simple des probabilités qui admettent une densité.

Définition 6.1.1 Densité de probabilité sur $\mathbb{R}$

C'est toute fonction positive Riemann intégrable telle que

$$\int_{\mathbb{R}} f(x)dx = 1$$

Remarque : dans la suite de ce cours nous ne considérerons que des fonctions f continues ou continues par morceaux. Ces fonctions sont toutes Riemann intégrables.

Définition 6.1.2

On dit qu'une variable aléatoire réelle X a pour densité f si pour tout intervalle $]a, b[$ réel (a et b éventuellement infini), on a

$$P(a < X < b) = \int_a^b f(x)dx \tag{6.1}$$

Remarque : La formule 6.1 définit une fonction que nous appellerons probabilité qui va de l'ensemble des intervalles réels dans $[0, 1]$. On veut que cette fonction satisfasse la propriété de σ-additivité introduite dans le chapitre 2. Ce faisant, on peut prolonger P à une classe d'ensembles qui contient les unions dénombrables d'intervalles. Nous admettrons qu'une telle probabilité est σ-additive.

Conséquences

-Si a est un réel,

$P(X = a) = \lim_{n\to\infty} P(a-1/n < X < a+1/n) = \lim_{n\to\infty} \int_{a-1/n}^{a+1/n} f(x)dx = 0$

On a donc pour a, b finis :

$P(a < X < b) = P(a \leqslant X < b) = P(a < X \leqslant b) = P(a \leqslant X \leqslant b) = \int_a^b f(x)dx$

-La fonction de répartition de la variable X est définie par

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(x)dx.$$

La fonction F a alors les propriétés suivantes :

-Elle est continue et dérivable en tout point où f est continue.

-Si $f(a^-)$ et $f(a^+)$ existent, les dérivées à gauche et à droite de F : $F'(a^-)$ et $F'(a^+)$ existent et

$$F'(a^-) = f(a^-) \ ; \ F'(a^+) = f(a^+).$$

La fonction de répartition est donc continue alors que dans le cas discret, elle admet une discontinuité en tout point de $X(\Omega)$.

Exemples :

1) Loi uniforme :

Soient $-\infty < a < b < \infty$, on définit la fonction

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}.$$

On note cette fonction

$$f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x).$$

On vérifie facilement que cette fonction est bien une densité de probabilité. La probabilité ou loi associée est appelée loi uniforme continue sur l'intervalle $[a, b]$. On note alors $X \sim \mathcal{U}_{[a,b]}$.

2) Loi exponentielle :

Soit $\lambda > 0$, on définit la fonction

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{si } x \in [0, +\infty) \\ 0 & \text{sinon} \end{cases}$$

On note cette fonction

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

On vérifie facilement que cette fonction est bien une densité de probabilité. La probabilité ou loi associée est appelée loi exponentielle de paramètre λ. On note alors $X \sim \mathcal{E}(\lambda)$.

6.2 Espérance d'une fonction d'une variable aléatoire

Soit I un intervalle réel. On note $\mathbf{1}_I(X)$ la variable aléatoire qui vaut 1 si $X(\omega) = 1$ et 0 sinon, on a alors

$$E(\mathbf{1}_I(X)) = P(X \in I) = \int_I f(x)dx$$

Soit maintenant g une fonction en escalier,

$$g(x) = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{[a_i,b_i[}(x)$$

où les intervalles $[a_i, b_i[$ sont disjoints. Alors, puisque $g(X)$ ne prend qu'un nombre fini de valeurs, l'espérance étant linéaire

$$\begin{aligned} E(g(X)) &= \sum_{i=1}^{n} \alpha_i E(\mathbf{1}_{[a_i,b_i[}(X)) = \sum_{i=1}^{n} \alpha_i \int_{[a_i,b_i[} f(x)dx \\ &= \int_{\mathbb{R}} \sum_{i=1}^{n} \alpha_i \mathbf{1}_{[a_i,b_i[}(x) f(x)dx = \int_{\mathbb{R}} g(x)f(x)dx. \end{aligned}$$

On généralise ce résultat en admettant le

Théorème 6.2.1

Soit X de densité f, pour toute fonction g continue,

$$E(g(X)) = \int_{-\infty}^{\infty} f(x)g(x)dx, \tag{6.2}$$

à condition que l'intégrale généralisée soit absolument convergente.

Réciproquement si la relation (6.2) est vraie pour toute fonction g continue bornée, alors f est la densité de X.

Définition 6.3.1

Soit X une variable aléatoire de densité f. En prenant $g(x) = x$ et en supposant que $\int_{-\infty}^{+\infty} |x|f(x)\,dx < \infty$, on définit l'espérance de X par

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)\,dx$$

Définition 6.3.2

Soit X une variable aléatoire de densité f. En prenant $g(x) = (x - \mathbb{E}(X))^2$ et en supposant que $\int_{-\infty}^{+\infty}(x - \mathbb{E}(X))^2 f(x)\,dx < \infty$, on définit la variance de X par

$$\mathbb{V}(X) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f(x)\,dx = \int_{-\infty}^{+\infty} x^2 f(x)\,dx - (\mathbb{E}(X))^2.$$

Exemples (Espérances et variances des lois uniforme et exponentielle) :

1) Loi uniforme :

Soient $-\infty < a < b < \infty$ et soit $X \sim \mathcal{U}_{[a,b]}$, alors

$$\mathbb{E}(X) = \frac{a+b}{2} \quad \mathbb{V}(X) = \frac{(b-a)^2}{12}.$$

2) Loi exponentielle :

Soit $\lambda > 0$ et soit $X \sim \mathcal{E}(\lambda)$, alors

$$\mathbb{E}(X) = \frac{1}{\lambda} \quad \mathbb{V}(X) = \frac{1}{\lambda^2}.$$

6.3 Changement de variables

Considérons le problème suivant : X est une variable aléatoire réelle de densité f, on considère $Y = h(X)$ où h est une certaine fonction. Y a-t-elle une densité et que vaut elle ?

On suppose qu'il existe $-\infty \leqslant a < b \leqslant \infty$ tels que

-f est nulle en dehors de $]a, b[$

-h est dérivable strictement croissante de $]a, b[$ vers $]c, d[$, $(h(a) = c; h(b) = d)$.

Nous appliquons le théorème 6.2 sur les densités. Soit g continue bornée,

$$E(g(Y)) = E(g(h(X))) = E(g \circ h(X)) = \int_a^b g \circ h(x) f(x) dx.$$

Dans l'intégrale on fait le changement de variable

$$y = h(x); \; x = h^{-1}(y); \; dx = (h^{-1})'(y) dy.$$

On obtient donc

$$E(g(Y)) = \int_{h(a)}^{h(b)} g(y) f(h^{-1}(y))(h^{-1})'(y) dy = \int_c^d g(y)(f(h^{-1}(y))(h^{-1})'(y)) dy$$

Par le théorème des densités (6.2), la densité de Y est donc

$$u(y) = f(h^{-1}(y))(h^{-1})'(y).$$

Quand h est décroissante, on obtient le même résultat avec un changement de signe et $h(a) = d$; $h(b) = c$. Au total on a bien montré

Proposition 6.3.1

Soit X une variable aléatoire de densité f nulle en dehors de $]a, b[$, $-\infty \leqslant a < b \leqslant \infty$ et soit h une application dérivable, strictement monotone de $]a, b[$ vers $]c, d[$, alors $Y = h(X)$ est une variable aléatoire réelle de densité

$$u(y) = f(h^{-1}(y)) \mid (h^{-1})'(y) \mid 1_{]c,d[}(y).$$

Exemple :

Si X est de densité f quelle est la densité de $Y = exp(X)$?

* $h = exp$ est bijective de $]-\infty, \infty[$ vers $]0, +\infty[$

* $h^{-1} = log$; $(h^{-1})'(y) = \frac{1}{y}$.

* La densité de Y est donc $u(y) = f(log(y))\dfrac{1}{y}\mathbf{1}_{]0,+\infty[}(x)$.

6.4 Lois gaussiennes $\mathcal{N}(m, \sigma^2)$

On introduit la loi normale à partir de la loi binomiale, on a en effet le résultat suivant dû à Moivre-Laplace :

Théorème 6.4.1 (De Moivre-Laplace)

Soit S_n une variable aléatoire de loi binomiale de paramètres n et p. On pose $S_n^* = \dfrac{S_n - np}{\sqrt{np(1-p)}}$ la variable aléatoire réelle centrée réduite associée. Pour tous réels $a < b$ fixés :

$$\lim_{n \to +\infty} P(a < S_n^* < b) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)dt. \qquad (6.3)$$

Le résultat théorique qui justifie cette approximation est le théorème de De Moivre-Laplace qui est lui même un cas particulier du théorème central limite. Ce dernier est, avec la loi des grands nombres, certainement le plus important théorème du calcul des probabilités. L'approximation qui nous intéresse fait intervenir une famille de fonctions appelées densités gaussiennes (ou normales) liées à la célèbre courbe en cloche de Gauss.

Définition 6.4.1

Soit $m \in \mathbb{R}$ et $\sigma \in]0, +\infty[$. On appelle densité gaussienne ou normale $f_{m,\sigma}$ sur $\mathbb{R}$ la fonction :

$$f_{m,\sigma} : \mathbb{R} \to \mathbb{R}^+ \qquad x \mapsto \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-m)^2}{2\sigma^2}) \tag{6.4}$$

La fonction $f_{0,1}$ est appelée densité normale standard (ou centrée réduite).

Une propriété importante de $C_{0,1}$ est que l'aire qu'elle délimite avec l'axe des abscisses vaut 1 :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)dt = 1 \tag{6.5}$$

On montre aussi, en faisant le changement de variable $u = \dfrac{t-m}{\sigma}$, que $\int_{-\infty}^{+\infty} f_{m,\sigma}(t)dt = 1$.

L'aire délimitée par $C_{m,\sigma}$, l'axe des abscisses et les droites d'équation $t = a$, $t = b$ joue un rôle important dans l'approximation des probabilités binomiales $P(a < S_n < b)$. Par changement de variable, le calcul de cette aire se ramène à celui de l'aire correspondante pour $C_{0,1}$ avec a^* et b^* à la place de a et b. Elle peut donc s'écrire sous la forme $\Phi(b^*) - \Phi(a^*)$ où la fonction Φ est définie par :

$$\forall x \in \mathbb{R}, \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2)dt \tag{6.6}$$

Cette intégrale ne peut s'exprimer à l'aide des fonctions usuelles. On peut en calculer une valeur approchée avec toute précision souhaitée .

La table en annexe donne les valeurs de $\Phi(x)$ par pas de 0.01 pour x compris entre 0 et 3 et quelques valeurs pour x compris entre 3 et 4.5. Pour x négatif, la parité de la densité entraîne la relation $\Phi(x) = 1 - \Phi(-x)$.

Approximation de la loi binomiale par la loi normale

D'après le théorème de De Moivre-Laplace, la loi binomiale $\mathcal{B}(n,p)$ peut être approchée par la loi normale $\mathcal{N}(np, np(1-p))$. En pratique, on utilise cette approximation lorsque $n \geqslant 30, np \geqslant 15$ et $np(1-p) > 5$.

Correction de continuité

Si la variable aléatoire X suit la loi binomiale $\mathcal{B}(n,p)$, alors la variable aléatoire X prend des valeurs entières positives entre 0 et n. Remplacer la loi binomiale $\mathcal{B}(n,p)$ par la loi normale $\mathcal{N}(np, np(1-p))$ revient à considérer la variable aléatoire X comme une variable qui prend donc toutes les valeurs réelles. L'intervalle $[k-0,5; k+0,5]$ est l'ensemble des nombres réels qui s'arrondissent à k, c'est-à-dire pour $k \in \{1, 2, .., n-1\}$, nous remplacerons $P(X = k)$ par $P(k - 0,5 \leqslant X < k + 0,5)$.

Remarque :

Pour que la somme des valeurs approchées des $P(X = k)$, k variant de 0 à n, soit égale à 1, nous remplacerons $P(X = 0)$ par $P(X < 0,5)$ et $P(X = n)$ par $P(n - 0,5 \leqslant X)$.

Exemple 6.1

On lance 3600 fois un dé. Évaluer la probabilité que le nombre d'apparitions du 1 soit compris strictement entre 540 et 660.

Soit S le nombre d'apparitions du 1. Cette variable aléatoire suit la loi binomiale $\mathcal{B}(3600, 1/6)$. On a $\mathbb{E}S = 600$ et $VarS = 500$. En notant $S^* = (S - \mathbb{E}S)/\sigma(S)$ la variable centrée réduite associée à S :

$$P(540 < S < 660) = P(\frac{540-600}{\sqrt{500}} < S^* < \frac{660-600}{\sqrt{500}}).$$

Comme n est grand on peut utiliser l'approximation liée au théorème de De Moivre-Laplace :

$$P(\frac{-60}{\sqrt{500}} < S^* < \frac{60}{\sqrt{500}}) \simeq \Phi(\frac{6}{\sqrt{5}}) - \Phi(\frac{-6}{\sqrt{5}})$$

En utilisant la parité de la densité $f_{0,1}$, on a pour tout $a > 0$, $\Phi(a) - \Phi(-a) = 2\Phi(a) - 1$. En utilisant la table des valeurs de Φ on obtient donc :

$P(540 < S < 660) \simeq 2\Phi(\frac{6}{\sqrt{5}}) = 2\Phi(2.68) - 1 \simeq 0.9926 \simeq 0.99.$

Chapitre 7

Statistique

La statistique poursuit deux buts :

-résumer de manière synthétique un corpus de données, on parle alors de statistique descriptive ou exploratoire,

-confronter ces données à un modèle probabiliste,on parle alors de statistique inférentielle.

Nous allons décrire brièvement le premier aspect. Le second sera traité ultérieurement en troisième année.

7.1 Présentation des données statistiques

Série statistique
Soit $\mathbf{x}$ un *caractère* défini comme une application d'un ensemble fini $\Omega = \{\omega_1, ...\omega_n\}$ dans $\mathbb{R}$. La série statistique associée à $\mathbf{x}$ est le vecteur $X = \{x_1, ...x_n\}$ où $x_i = \mathbf{x}(\omega_i)$.

Distribution statistique
On appelle distribution statistique de X l'ensemble des couples (u_i, n_i), $i \in \{1, 2, ..., p\}$, où les u_i sont les valeurs de $\mathbf{x}$ et les n_i les effectifs correspondants. On a :

$$\sum_{i=1}^{p} n_i = n$$

On dit que n est la *taille* de la série X ou de sa distribution statistique.
On appelle fréquence de u_i le réel $f_i = \frac{n_i}{n}$.

Fonction de répartition

On appelle fonction de répartition de X l'application F de $\mathbb{R}$ dans $\mathbb{R}$ définie par

$$F(x) = \sum_{u_i \leqslant x} f_i$$

F est une fonction en escalier, croissante de 0 à 1, continue à droite.

Distribution classée

On appelle distribution classée de X la donnée d'une suite finie d'intervalles jointifs, appelés *classes*, $[a_{i-1}, a_i[, i \in 1, 2, .., k$, tels que $x_i \in [a_0, a_k[, \forall i$, et des effectifs v_i correspondants.
Soit $\varphi_i = \frac{v_i}{n}$ la fréquence de la i-ième classe, soit $\alpha_i = a_i - a_{i-1}$ l'amplitude de cette classe et soit $h_i = \frac{\varphi_i}{\alpha_i}$.
Admettant l'hypothèse d'une répartition homogène à l'intérieur de chaque classe, on appelle *densité de fréquence* l'application h de $\mathbb{R}$ dans $\mathbb{R}$ définie par

$$h(x) = \begin{cases} h_i & \text{si } x \in [a_{i-1}, a_i[\\ 0 & \text{si } \; x < a_0 \text{ ou } x > a_k \end{cases}$$

On a

$$\int_{-\infty}^{+\infty} h(x)\, dx = 1$$

La représentation graphique de h est appelé *l'histogramme*.
On appelle fonction de répartition de la distribution classée de X l'application de $\mathbb{R}$ dans $\mathbb{R}$ définie par

$$\Phi(x) = \int_{-\infty}^{x} h(t)\, dt.$$

On peut voir que Φ est croissante de 0 à 1, affine sur chaque intervalle $[a_{i-1}, a_i[$, continue et telle que $\Phi(a_i) = \sum_{j \leqslant i} \varphi_j$.

7.2 Caractéristiques et moments

Soit $X = \{x_1, x_2, ..., x_n\}$ une série statistique de distribution (u_i, n_i), $i \in \{1, 2, ..., p\}$.

7.2.1 Caractéristiques de position

Mode

On appelle mode toute valeur u_i telle que n_i soit maximum.
Dans le cas d'une distribution classée de X, on appelle mode toute classe telle que la densité de fréquence h_i soit maximum.

Médiane

On appelle médiane tout réel μ tel que

$$\sum_{u_i \leqslant \mu} n_i \leqslant \frac{n}{2} \quad \text{et} \quad \sum_{u_i \geqslant \mu} n_i \leqslant \frac{n}{2}.$$

Soit F la fonction de répartition de X. S'il existe u_k tel que $F(u_k) = \frac{1}{2}$, alors tout $\mu \in [u_k, u_{k+1}]$ est une médiane. Dans l'autre cas, la médiane, unique est

$$\mu = \inf\{u_k : F(u_k) \geqslant \frac{1}{2}\}.$$

Dans le cas d'une distribution classée de X dont les effectifs v_i sont tous strictement positifs, la fonction de répartition Φ de la distribution classée est continue et strictement croissante. On définit alors la médiane μ de la distribution classée par

$$\Phi(\mu) = \frac{1}{2}.$$

Moyenne

Définition :

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{p} n_i u_i = \sum_{i=1}^{p} f_i u_i.$$

Propriété de linéarité. Soient X et Y deux séries statistiques de taille n et soit $Z = \lambda X + \mu Y, \lambda$ et μ réels. On a $m(Z) = \lambda m(X) + \mu m(Y)$. Posons $x'_i = x_i - m$. La série X' est dite *série centrée.* On a $m(X') = 0$.

7.2.2 Caractéristiques de dispersion

Étendue

$$E = \max_i x_i - \min_i x_i.$$

Intervalle interquartile

En présence d'une série classée de X telle que $v_i > 0 \ \forall i$, on définit les quartiles q_1 et q_3 par $\Phi(q_1) = 0,25$ et $\Phi(q_3) = 0.75$. L'intervalle interquartile est $I = q_3 - q_1$.

Écart-moyen

$$e_m = \frac{1}{n}\sum_{i=1}^{n}|x_i - m| = \frac{1}{n}\sum_{i=1}^{p} n_i|u_i - m| = \sum_{i=1}^{p} f_i|u_i - m|$$

Variance et écart-type

$$V = \frac{1}{n}\sum_{i=1}^{n}(x_i - m^2 = \frac{1}{n}\sum_{i=1}^{p} n_i(u_i - m)^2 = \sum_{i=1}^{p} f_i(u_i - m)^2$$

L'écart-type de X est $\sigma = \sqrt{V}$.

Propriétés

a) Soit X une série statistique. Le changement de variable $y = \lambda x + \mu$ conduit à

$$\sigma(Y) = |\lambda|\sigma(X)$$

b)

$$V = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - m^2 = \frac{1}{n}\sum_{i=1}^{p} n_i u_i^2 - m^2 = \sum_{i=1}^{p} f_i u_i^2 - m^2$$

7.2.3 Moments

Moments d'ordre $k, k \in \mathbb{N}$

$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k = \frac{1}{n}\sum_{i=1}^{p} n_i u_i^k = \sum_{i=1}^{p} f_i u_i^k.$$

Moments centrés d'ordre $k, k \in \mathbb{N}$

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^k = \frac{1}{n}\sum_{i=1}^{p} n_i(u_i - m)^k = \sum_{i=1}^{p} f_i(u_i - m)^k.$$