3 Variables aléatoires et distributions des probabilités

3.1 Variables aléatoires

Après avoir réalisé une expérience, il arrive bien souvent qu'on s'intéresse plus à une fonction de résultat qu'au résultat lui-même.

Supposons que nous affections une valeur à chaque point de l'espace d'échantillonnage. Nous aurons, alors, défini une fonction sur cet espace. Cette fonction est appelée *variable aléatoire* (ou *variable stochastique*). Une variable aléatoire possède, en général, une signification donnée, physique, géometrique ou autre.

Du fait que la valeur d'une variable aléatoire est déterminée par le résultat de l'expérience, il est possible d'attribuer une probabilité aux différentes valeurs que la variable aléatoire peut prendre.

Une variable aléatoire qui peut prendre un nombre fini ou dénombrable infini de valeurs est dite *variable aléatoire discrète*; si elle peut prendre un nombre infini non dénombrable de valeurs, elle est dite *variable aléatoire continue.*

Exemples:

1. *D'une urne contenant 20 boules numérotées de 1 à 20, on tire sans remplacement 3 des boules. Quelqu'un parie qu'au moins une des boules tirées portera un numéro égal ou supèrieur à 17. Quelle est la probabilité qu'il gagne ?*

   Réponse:

   L'ensemble fondamental est alors formé de $\binom{20}{3}$ événements élémentaires. Disons que ξ représente le plus grand numéro tiré. Alors ξ est une variable aléatoire pouvant prendre les valeurs $3, 4, \ldots, 20$. En supposant que les $\binom{20}{3}$ tirages sont tous équiprobables, on a :

   $$P(\xi = i) = \frac{\binom{i-1}{2}}{\binom{20}{3}} \qquad \text{où} \qquad i = 3, \ldots, 20.$$

   La question est $P(\xi \geq 17)$. $\{\xi \geq 17\}$ cet un événement composé et ces événements élémentaires sont mutuellement exclusifs alors

   $$P(\xi \geq 17) = P(\xi = 17) + P(\xi = 18) + P(\xi = 19) + P(\xi = 20).$$

La probabilité de gagner le pari est donc

$$P(\xi \geq 17) = \frac{\binom{19}{2} + \binom{18}{2} + \binom{17}{2} + \binom{16}{2}}{\binom{20}{3}} = 0.508.$$

2. *Lors de la production en série d'un certain type de dispositifs électriques, on prélève au hasard un article de la production journalière et l'on mesure sa durée de bon fonctionnement.*

   L'ensemble fondamental Ω *est alors formé de tous les dispositifs produits pendant la même journée, tandis que* $\xi(w)$ *égal à la durée de bon fonctionnement du dispositif* w.

3. *Sur une route, on observe le flux des véhicules circulant dans une direction, et l'on s'intéresse à leur appartenance à l'une des catégories suivantes : voiture automobile légère, moto, bus, poids lourd. En vue d'un dépouillement statistique, on associe les valeurs* $0, 1, 2, 3$ *à ces quatre catégories, transformant ainsi un ensemble de caractères qualitatifs en un ensemble de caractères quantitatifs.*

   Remarques:

   – Par tradition, les variables sont notées ξ, η ... pour les distinguer des valeurs qu'elles sont susceptibles de prendre $(x, y, \ldots)$. D'après la définition une "variable aléatoire" n'est en réalité ni variable ni aléatoire, mais une fonction réelle définie sur l'ensemble d'échantillonnage Ω constitué par tous les résultats possibles d'une expérience stochastique.

   – Une variable aléatoire définie sur un ensemble fondamental fini ou dénombrable est évidemment toujours discrète, tandis qu'une variable aléatoire définie sur ensemble infini non dénombrable peut être soit discrète soit continue.

3.2 Fonction de distribution de variable aléatoire

3.2.1 Définition de la fonction de distribution

La *fonction de distribution* ou *la fonction de répartition* F d'une variable aléatoire ξ est définie pour tout nombre réel x, $-\infty < x < \infty$, par

$$F(x) = P(\xi < x).$$

En d'autres termes, $F(x)$ est la probabilité que la variable aléatoire ξ prenne une valeur inférieure à x.

3.2.2 Propriétés des fonctions de répartition

- F est une fonction monotone croissante, autrement dit si $x < y$, alors $F(x) \leq F(y)$
- F est continue à gauche en tout point,
- $\lim\limits_{x \to \infty} F(x) = 1$ et $\lim\limits_{x \to -\infty} F(x) = 0$

Remarque:

- La première propriété repose sur le fait que si $x < y$, l'événement $\{\xi < x\}$ est inclus dans $\{\xi < y\}$; la probabilité du premier est donc nécessairement plus petite que celle du second.
- $P(\xi = x_0) = \lim\limits_{x \to x_0(+0)} F(x) - F(x_0)$

Exemple: *La fonction de répartition de la variable aléatoire ξ est donnée par*

$$F(x) = \begin{cases} 0 & si \quad x \leq 0 \\ \frac{x}{2} & 0 < x \leq 1 \\ \frac{2}{3} & 1 < x \leq 2 \\ \frac{11}{12} & 2 < x \leq 3 \\ 1 & 3 < x \end{cases}$$

Représenter son graph.

3.2.3 Fonction de distribution et probabilité sur ξ

Tous les calculs de probabilité concernant ξ peuvent être traités en termes de fonction de distribution.

Par exemple : Quelles sont les probabilités des événements :

$$\{\xi < a\}, \quad \{\xi \geq a\}, \quad \{a \leq \xi < b\}, \quad \{\xi = a\}, \quad \{a \leq \xi \leq b\}, \quad \{a < \xi < b\}.$$

Réponse: On peut montrer utilisant la définition que

- $P(\xi < a) = F(a)$;
- $P(\xi \geq a) = 1 - F(a)$;
- $P(a \leq \xi < b) = F(b) - F(a)$; parce que $F(b) = F(a) + P(a \leq \xi < b)$

- $P(\xi = a) = \lim\limits_{x \to a(+0)} F(x) - F(a)$
- $P(a \leq \xi \leq b) = F(b) - F(a) + P(\xi = b)$
- $P(a < \xi < b) = F(b) - F(a) - P(\xi = a)$

Remarque:
- Si ξ est une variable aléatoire discrète prenant les valeurs $x_1, x_2, x_3, \ldots$, $F(x)$ est alors la somme des $P(\xi = x_k)$ avec $x_k < x$. F est alors une fonction en escalier.

3.3 Variables aléatoires discrètes

Une variable aléatoire ne pouvant prendre qu'une quantité dénombrable de valeurs est dite *discrète.* Pour une telle variable aléatoire ξ, on définit sa *loi de probabilité* p par

$$p(x_k) = P(\xi = x_k).$$

Cette loi de probabilité ne peut être positive que pour un ensemble au plus dénombrable d'arguments. En d'autres termes, si ξ peut prendre les valeur $x_1, x_2, \ldots$ alors $p(x_i) \geq 0$ si $i = 1, 2, \ldots$

$$p(x) = 0 \text{ pour toutes les autres valeurs de } x.$$

Du fait que ξ doit bien prendre l'une de ces valeurs x_i on aura $\sum_{i=1}^{\infty} p(x_i) = 1$.

La présentation sur un graphique à 2 dimensions des valeurs de la fonction qu'est la loi de probabilité est très parlante. Les valeurs sont placées en abscisses, les images $p(x_i)$ étant en ordonnées.

Exemple: *La loi de probabilité d'une variable aléatoire ξ est donnée par $p(i) = c\lambda^i/i\,!$, $i = 0, 1, 2, \ldots$, où λ est un réel positif. Trouver*
- $P(\xi = 0)$
- $P(\xi > 2)$

Réponse: Puisque $\sum_{i=0}^{\infty} p(i) = 1$ nous avons que

$$c \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1$$

ce qui implique, puisque

$$e^x = \sum_{i=0}^{\infty} x^i/i!, \quad \text{que } ce^{\lambda} = 1 \text{ ou } c = e^{-\lambda}.$$

Or

$$\begin{aligned}
&P(\xi = 0) = e^{-\lambda} \cdot \lambda^0/0! = e^{-\lambda} \\
&P(\xi > 2) = 1 - P(\xi \leq 2) = 1 - P(\xi = 0) - P(\xi = 1) - P(\xi = 2) = \\
&\qquad\qquad = 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}.
\end{aligned}$$

On peut exprimer la fonction de répartition F d'une variable aléatoire discrète en fonction des valeurs prises par sa loi de probabilité p :

$$F(a) = \sum_{x<a} p(x).$$

Dans le cas précis où les valeurs possibles de la variable aléatoire sont $x_1, x_2, x_3, \ldots$, avec $x_1 < x_2 < x_3 < \ldots$, la fonction F de distribution est une fonction en escalier. Ses valeurs seront constantes sur les intervalles $(x_{i-1}, x_i]$ et elle aura un saut de taille $p(x_i)$ en x_i, $i = 1, 2, \ldots$.

3.4 Variables aléatoires continues

Si ξ est une variable aléatoire continue, la probabilité que ξ prenne une valeur particulière quelconque est généralement nulle. Dans ces conditions, il n'est pas possible de définir une fonction de probabilité comme pour une variable aléatoire discrète. Pour pouvoir définir une distribution de probabilités dans le cas d'une variable aléatoire continue on remarque que la probabilité pour que la valeur de ξ soit entre deux valeurs différents est une notion qui a un sens. La variable aléatoire est dite continue s'il est possible de mettre sa fonction de répartition sous la forme $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$, on dit que f est la *densité de probabilité* de la variable aléatoire ξ. Les propriétés de la fonction densité

1. $f(x) \geq 0$,
2. $\displaystyle\int_{-\infty}^{\infty} f(x) = 1$,
3. $\displaystyle P(a \leq \xi < b) = \int_a^b f(x)\,\mathrm{d}x$ puisque $P(a \leq \xi < b) = F(b) - F(a)$.

Remarque:

$$P(a < \xi < b) = P(a \leq \xi < b) = P(a < \xi \leq b) = P(a \leq \xi \leq b).$$

Exemple: *Supposons que ξ soit une variable aléatoire continue dont la densité est*

$$f(x) = \begin{cases} c(4x - 2x^2) & si \quad 0 < x < 2 \\ 0 & sinon. \end{cases}$$

Quelle est la valeur de c ? Que vaut $P(\xi > 1)$?

Réponse: Du fait que f est une densité elle vérifie la relation $\displaystyle\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$, ce qui entraîne à son tour que

$$c \int_0^2 (4x - 2x^2)\,\mathrm{d}x = 1 \qquad \text{alors} \qquad c = \frac{3}{8}.$$

Donc $P(\xi > 1) = \displaystyle\int_1^{\infty} f(x)\,\mathrm{d}x = \frac{3}{8}\int_1^2 (4x - x^2)\,\mathrm{d}x = \frac{1}{2}.$

Interprétation graphique

Si $f(x)$ est la densité de probabilité d'une variable aléatoire ξ, nous pouvons représenter $y = f(x)$ graphiquement par une courbe. Dans la mesure où $f(x) \geq 0$ la courbe ne peut-être située en-dessous de l'axe des x. L'aire totale comprise entre la courbe de l'axe des x doit être égale à 1. Géométriquement la probabilité pour que ξ soit comprise entre a et b, (c'est-à-dire $P(a < \xi < b)$), est alors représentée par l'aire.

3.5 Les caractéristiques de la variable aléatoire

Une variable aléatoire est totalement déterminée soit par sa fonction de répartition, soit par sa densité ou encore par sa fonction de probabilité. En statistiques, on se contente généralement de certaines valeurs caractéristiques qui ne décrivent pas entièrement la variable aléatoire mais qui sont en général suffisantes. On doit être alors parfaitement conscient de perdre en précision certaines informations. Les valeurs caractéristiques les plus importantes sont l'espérance mathématique et la variance.

3.5.1 L'espérance mathématique

L'espérance mathématique, notée $E(\xi)$, d'une variable aléatoire ξ, appelée également *moyenne*, si elle existe. Dans le cas d'une variable aléatoire discrète ξ, dont les valeurs possibles sont $x_1, x_2, \ldots$ *l'espérance* de ξ est définie par l'expression

$$E(\xi) = \sum_{x_i \,:\, p(x_i)>0} x_i p(x_i), \qquad \text{si elle existe.}$$

En termes concrets, l'espérance de ξ est la moyenne pondérée des valeurs que ξ peut prendre, les poids étant les probabilités que ces valeurs soient prises.

Exemple: *On cherche l'espérance $E(\xi)$ de la variable ξ, résultat du lancer d'un dé équilibré.*

Réponse: Comme $p(1) = p(2) = \ldots = p(6) = \dfrac{1}{6}$, on aura

$$E(\xi) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Pour le cas continu, dont la densité de la probabilité est $f(x)$, l'espérance de ξ est définie :

$$E(\xi) = \int_{-\infty}^{\infty} x f(x)\, \mathrm{d}x \qquad \text{si elle existe.}$$

Analogie avec une notion de mécanique.

Le concept d'espérance est à rapprocher de la notion de *centre de gravité* d'un groupe de masses, au sens de la mécanique. Considérons en effet une variable ξ loi de probabilité $P(x_i)$, $i \geq 1$. On sait que si des masses $P(x_i)$, $i \geq 1$ sont réparties sur une barre sans poids aux abscisses x_i, $i \geq 1$, le point sur lequel la barre pourra être posée et rester en équilibre est appelé centre de gravité.

Remarque:

- L'espérance n'est pas toujours définie. Considérons le cas où ξ est une variable aléatoire pouvant prendre une infinité de valeurs distinctes, $E(\xi) = \sum_{i=1}^{\infty} p(x_i) \cdot x_i$. Lorsque la somme a un sens et ne dépend pas de l'ordre des termes, et dans ce cas seulement on considère que ξ a une espérance mathématique.
- L'espérance d'une variable aléatoire ξ n'appartient pas obligatoirement aux valeurs possibles de ξ. Ainsi l'espérance mathématique des valeurs prises par un dé ideal est 3.5 alors que le dé ne peut faire apparaître que des valeurs entières. $E(\xi)$ est un nombre autour duquel se répartissent les valeurs possibles de ξ.

3.5.2 Quelques théorèmes relatifs à l'espérance mathématique

– Si c est une constante quelconque,

$$E(c\xi) = cE(\xi).$$

– Si ξ et η sont des variables aléatoires quelconques,

$$E(\xi + \eta) = E(\xi) + E(\eta).$$

Ces propriétés permettent d'affirmer que l'espérance est un opérateur linéaire.

– Si ξ et η sont des variables aléatoires indépendantes,

$$E(\xi\eta) = E(\xi) + E(\eta).$$

La généralisation de ces théorèmes est aisée.

3.5.3 Variance et écart-type

Si $E(\xi)$ nous donne une moyenne pondérée des valeurs possibles de ξ, elle ne nous dit rien des variations de ξ autour de l'espérance. On peut s'en rendre compte grâce aux exemples suivants. Soit les variables

$$\xi_1 = 0 \text{ avec probabilité } 1$$

$$\xi_2 = \begin{cases} -1 & \text{avec probabilité } \dfrac{1}{2} \\ 1 & \text{avec probabilité } \dfrac{1}{2} \end{cases}$$

$$\xi_3 = \begin{cases} -100 & \text{avec probabilité } \dfrac{1}{2} \\ 100 & \text{avec probabilité } \dfrac{1}{2} \end{cases}$$

Si toutes ont la même espérance, il y a de bien plus grands écarts entre les différentes valeurs de ξ_2 qu'entre celle de ξ_1 et de plus grands écarts entre celles de ξ_3 qu'entre celles de ξ_2.

La variance est une mesure de la dispersion. On s'intéresse à la grandeur $E(|\xi - E(\xi)|)$. Techniquement, il n'est pas facile de manipuler cette quantité. Ainsi on appelle *variance* de ξ, que l'on note Var (ξ), la quantité

$$\text{Var}\,(\xi) = E[(\xi - E(\xi))^2], \qquad \text{si elle existe.}$$

On peut établir une autre formule pour le calcul de Var (ξ) :

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

Dans l'analogie des masses, c'est le moment d'inertie par rapport au centre de gravité. La variance de ξ est d'autant plus grande que ξ peut prendre des valeurs plus diverses.

Exemple: *On veut comparer les distributions des trois variables aléatoires,* ξ_1*,* ξ_2*,* ξ_3

x_k	0	1	2	3	4
$P(\xi_1 = x_k)$	0.2	0.2	0.2	0.2	0.2
y_k	8	9	10	11	12
$P(\xi_2 = y_k)$	0.2	0.2	0.2	0.2	0.2
z_k	-2	0	4	6	
$P(\xi_3 = z_k)$	0.3	0.2	0.2	0.3	

On vérifie que ξ_1 et ξ_2 ont la même variance, mais se distinguent par leurs moyennes, tandis que ξ_1 et ξ_3 varient autour de la même moyenne, mais avec des dispersions différentes.

Quelques théorèmes relatifs à la variance

– Si c est une constante quelconque,

$$\text{Var}\ (c\xi) = c^2 \text{Var}\ (\xi).$$

– $E[(\xi - a)^2]$ est minimum quand $a = E(\xi)$.

– Si ξ et η sont des variables aléatoires indépendantes,

$$\text{Var}\ (\xi + \eta) = \text{Var}\ (\xi) + \text{Var}\ (\eta)$$
$$\text{Var}\ (\xi - \eta) = \text{Var}\ (\xi) + \text{Var}\ (\eta).$$

Les généralisations à plus de deux variables sont aisées.

– Si a et b sont deux constantes,

$$\text{Var}\ (a\xi + b) = a^2 \text{Var}\ (\xi).$$

La variance est un nombre non négatif, sa racine carrée est appelé *écart-type* de la variable aléatoire.

Remarque:

- Notons qu'une variable aléatoire, peut très bien ne pas avoir d'espérance mathématique et de variance.
- Si ξ a certaines dimensions et s'exprime donc en certaines unités, des cm par exemple, la variance de ξ s'exprimera en cm^2 et l'écart-type avec les mêmes unités que ξ. C'est la raison pour laquelle l'écart-type est d'une usage fréquent.

Exemple: *On est parfois amené à associer à une variable aléatoire donnée ξ, de caractéristiques $E(\xi) = \mu$ et $Var\,(\xi) = \sigma^2$, une variable aléatoire centrée réduite $\xi^* = a\xi + b$ satisfaisant aux conditions*

$$E(\xi^*) = 0 \quad et \quad Var\,(\xi^*) = 1.$$

Réponse: A l'aide des relations ci-dessus, on voit immédiatement que $a = 1/\sigma$ et $b = -\mu/\sigma$, c'est-à-dire que $\xi^* = \dfrac{\xi - \mu}{\sigma}$.

3.6 Moments

Le *moments d'ordre* r d'une variable aléatoire ξ relatif à la moyenne μ, appelé aussi *moment central d'ordre* r, est défini par l'expression :

$$\mu_r = E[(\xi - \mu)^r] \text{ où } r = 0, 1, 2, \ldots; \text{ il en résulte } \mu_0 = 1,\ \mu_1 = 0,\ \mu_2 = \sigma^2.$$

Alors

$$\mu_r = \sum (x_j - \mu)^r \cdot p(x_j) \qquad \text{(variable discrète)}$$
$$\mu_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x)\,\mathrm{d}x \qquad \text{(variable continue).}$$

3.7 Fonctions génératrices des moments

On définit pour tout réel t, la *fonction génératrice des moments* φ de la variable aléatoire ξ par

$$\varphi(t) = \begin{cases} \displaystyle\sum_x e^{tx} p(x) & \text{si } \xi \text{ est discrète, de la loi de probabilité } p \\ \displaystyle\int_{-\infty}^{\infty} e^{tx} f(x)\,\mathrm{d}x & \text{si } \xi \text{ est continue, de densité } f. \end{cases}$$

Cette fonction φ est appelée fonction génératrice des moments du fait que tous les moments d'ordre n de ξ peuvent être calculés en dérivant n fois φ puis en évaluant la dérivée à $t = 0$.