

Statistique décisionnelle

Eugen Ursu

Université Bordeaux IV



Chapitre 2 : Variables aléatoires réelles

- 1 Introduction
- 2 Loi d'une variable aléatoire discrète
- 3 Espérance
- 4 Variance
- 5 Exemples des lois discrètes
 - Loi de Bernoulli
 - Loi Binomiale
 - Loi Géométrique
 - Loi de Poisson



Variable aléatoire réelle

Définition

Une fonction définie sur l'ensemble des résultats possibles d'une expérience aléatoire

$$X : \Omega \rightarrow X(\Omega) \subset \mathbb{R}$$

*s'appelle **variable aléatoire réelle**.*

Exemple

- On jette deux dés, et on considère X = la somme des deux chiffres ;
- On joue à pile ou face 10 fois, et on considère X = le nombre de pile ;
- On jette un dé jusqu'à ce qu'on gagne (le 6 sort), et on considère X = nombre de fois où le dé est jeté ;
- La durée de vie d'une batterie (en heures) est une variable aléatoire.

Variable aléatoire réelle

Deux types de variable aléatoire :

- **Discrète** quand $X(\Omega) = \{x_1, x_2, \dots\}$ est fini ou dénombrable.
- **Continue** quand $X(\Omega)$ est un intervalle de $\mathbb{R}$.



Variable aléatoire discrète

Une variable aléatoire est caractérisée par l'ensemble des valeurs qu'elle peut prendre et par les probabilités de ces valeurs. Cette expression s'appelle la **loi de probabilité** (ou distribution de probabilité) de la variable aléatoire.

La loi de probabilité d'une variable aléatoire discrète est entièrement déterminée par les probabilités des événements $\{X = x_i\}$, x_i parcourant l'univers image $X(\Omega)$. La loi de probabilité est donnée par les $(x_i, \mathbb{P}(X = x_i))_i$.

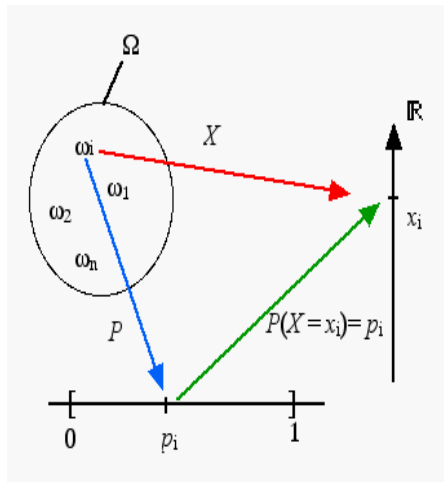
Remarque

- ❶ *Afin de simplifier l'écriture, nous noterons pour la suite du cours :*

$$\mathbb{P}(X = x_i) = p_i.$$

- ❷ *Une loi de probabilité n'est établie que si $p_i \geq 0$ et $\sum_i p_i = 1$, la somme étant étendue à tous les indices i .*

Variable aléatoire discrète



Variable aléatoire discrète : exemple

Exemple

Exemple avec un nombre fini des valeurs possibles

On jette deux dés, et on considère X = la somme des deux chiffres.
 $\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$. On considère que les deux dés sont non truqués (équiprobabilité).

x_j	2	3	4	5	6	7	8	9	10	11	12
p_j	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36



Variable aléatoire discrète : exemple

Exemple

Exemple avec une infinité des valeurs possibles

On lance une pièce jusqu'à obtenir pile pour la première fois. On a $\Omega = \{p, (f, p), (f, f, p), (f, f, f, p) \dots\}$. Soit la v.a. X "le nombre de lancers nécessaires". La distribution de X est donnée par le tableau suivant

x_i	1	2	3	4	...
p_i	1/2	1/4	1/8	1/16	...



Fonction de répartition

Définition

La **fonction de répartition** d'une variable aléatoire X (pas nécessairement discrète) est définie pour tout nombre réel $x \in \mathbb{R}$ par

$$F(x) = \mathbb{P}(X \leq x).$$



Fonction de répartition : propriétés

Proposition

La fonction de répartition F d'une variable aléatoire X jouit des propriétés suivantes :

- ❶ $0 \leq F(x) \leq 1$;
- ❷ F est non-décroissante, c'est-à-dire si $x < y$, alors $F(x) \leq F(y)$;

❸

$$\lim_{x \rightarrow \infty} F(x) = 1 \text{ et } \lim_{x \rightarrow -\infty} F(x) = 0;$$

- ❹ F est continue à droite, c'est-à-dire

$$\lim_{x \searrow y} F(x) = F(y).$$

Remarque

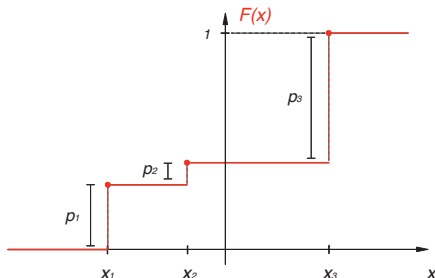
Toute fonction F satisfaisant ces quatre propriétés est en fait la fonction de répartition d'une v.a.

Fonction de répartition

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

$$F(x) = \sum_{x_i \leq x} \mathbb{P}(X = x_i) = \sum_{x_i \leq x} p_i.$$

Dans le cas où les valeurs possibles de la variable aléatoire sont $x_1, x_2, x_3, \dots$ avec $x_1 < x_2 < x_3 < \dots$, la fonction de répartition F est une **fonction en escalier** (voir figure ci-dessous).



Fonction de répartition

Remarque

La fonction de répartition détermine uniquement la loi de probabilité d'une variable aléatoire, car la variable aléatoire prend pour valeurs les abscisses des points de saut de la fonction F avec des probabilités égales aux hauteurs des sauts.



Fonction de répartition : exercice

Exercice

On considère l'événement A = "lancer des 3 pièces". On introduit une variable aléatoire X par le nombre de piles de l'événement A . La loi de probabilité de X est donnée par :

X	0	1	2	3
p_i	1/8	3/8	3/8	1/8

Trouver la fonction de répartition et tracer son graphe.



Remarque

Pour tout $a \in \mathbb{R}$ on a

$$\mathbb{P}(X = a) = F(a) - \lim_{x \nearrow a} F(x).$$

Dans tout point a de continuité de la fonction de répartition F on a $\mathbb{P}(X = a) = 0$ car en un tel point $\lim_{x \nearrow a} F(x) = F(a)$.

Exemple

Soit X une variable aléatoire, et soit

$$F(x) = \begin{cases} 0 & , \quad x < 3 \\ 1/3 & , \quad 3 \leq x < 5 \\ 2/3 & , \quad 5 \leq x < 8 \\ 1 & , \quad x \geq 8 \end{cases}$$

sa fonction de répartition. Trouver la loi de probabilité de la v.a. X

Fonction de répartition : propriétés

Remarque

Tous les calculs de probabilité concernant X peuvent être traités en termes de fonction de répartition. D'abord,

$$\begin{aligned}\mathbb{P}(X < b) &= \mathbb{P}(X \leq b) - \mathbb{P}(X = b) \\ &= F(b) - \left(F(b) - \lim_{x \nearrow b} F(x) \right) = \lim_{x \nearrow b} F(x)\end{aligned}$$

et par suite

$$\begin{aligned}\mathbb{P}(a < X < b) &= \lim_{x \nearrow b} F(x) - F(a) \\ \mathbb{P}(a \leq X < b) &= \lim_{x \nearrow b} F(x) - \lim_{x \nearrow a} F(x) \\ \mathbb{P}(a < X \leq b) &= F(b) - F(a) \\ \mathbb{P}(a \leq X \leq b) &= F(b) - \lim_{x \nearrow a} F(x)\end{aligned}$$

Fonction de répartition : exercice

Exercice

La fonction de répartition de la v.a. X est donnée par

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

Calculer $\mathbb{P}(X < 3)$, $\mathbb{P}(X > 1/2)$, $\mathbb{P}(X = 1)$ et $\mathbb{P}(2 < X \leq 4)$.



Définition

Pour une variable aléatoire discrète X , on définit l'**espérance** de X (ou la moyenne de X), notée $E[X]$, par l'expression :

$$E[X] = \sum_i x_i \mathbb{P}(X = x_i) = \sum_i x_i p_i.$$

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



Espérance : exemple

Exemple

Revenons à l'exemple avec la somme des deux dés. On jette deux dés, et on considère X = la somme des deux chiffres.
 $\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$. On considère que les deux dés sont non truqués (équiprobabilité).

x_j	2	3	4	5	6	7	8	9	10	11	12
p_i	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Il est facile à voir que

$$E[X] = \sum_{i=1}^{11} x_i p_i = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + \dots + 12 \times \frac{1}{36} = 7.$$



Espérance : propriétés

Proposition

L'espérance d'une variable aléatoire possède les propriétés suivantes : Soient X, Y deux variables aléatoires et $a, b \in \mathbb{R}$.

- 1 $E[aX + bY] = aE[X] + bE[Y]$, où a, b sont des constantes ;
- 2 $E[g(X)] = \sum_i g(x_i)p_i$, pour toute fonction réelle g .
- 3 $|E[XY]| \leq \sqrt{E[X^2]}\sqrt{E[Y^2]}$ (inégalité de Cauchy-Schwarz), et l'égalité a lieu si et seulement si $Y = cX$ ou $X = cY$ pour un certain $c \in \mathbb{R}$.



Espérance : interprétation

Remarque

- *L'espérance (la valeur moyenne) d'une variable aléatoire X peut s'interpréter comme une valeur autour de laquelle se groupent les valeurs de la variable aléatoire X . On dit que $E[X]$ est un paramètre de localisation.*
- *Si $X(\Omega)$ est infini, on n'est pas sûr que l'espérance existe.*

Remarque

*L'espérance d'une variable aléatoire X est parfois nommée **premier moment**. De manière générale, la quantité $E[X^n]$ est appelée **n -ième moment de X** et*

$$E[X^n] = \sum_i x_i^n p_i.$$

Variance

Définition

On appelle **variance** de X , que l'on note $\text{Var}(X)$ la quantité

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

Une autre formule pour le calcul de la variance est :

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

Remarque

La variance d'une variable aléatoire X peut s'interpréter comme une mesure du degré de dispersion des valeurs de la variable aléatoire X par rapport à sa valeur moyenne. Si la variance est petite alors les valeurs de la variable aléatoire X sont groupées dans un petit intervalle autour de la valeur moyenne. Si par contre, la variance est grande, les valeurs de la variable aléatoire X sont fortement dispersées dans un grand intervalle autour de la valeur moyenne.

Variance : propriétés

Proposition

La variance d'une variable aléatoire possède les propriétés suivantes :

- ❶ $Var(a) = 0$, où a est une constante ;
- ❷ $Var(aX + b) = a^2 Var(X)$, où a, b sont des constantes ;
- ❸ Si $Var(X) = 0$, alors X est une constante.

Dans les applications, comme mesure pour la dispersion des valeurs de la variable aléatoire X autour de sa valeur moyenne, il est plus commode d'utiliser le nombre

$$\sigma(X) = \sqrt{Var(X)}$$

appelé l'**écart type** de la variable aléatoire X .



Exercice

Soit X une variable aléatoire qui prend une des trois valeurs $-1, 0, 1$ avec les probabilités suivantes

$$\mathbb{P}(X = -1) = 0,2, \quad \mathbb{P}(X = 0) = 0,5, \quad \mathbb{P}(X = 1) = 0,3.$$

Calculer $E[X]$, $E[X^2]$ et $\text{Var}(X)$.



Loi de Bernoulli

On réalise une expérience dont le résultat sera interprété soit comme un succès soit comme un échec. L'espace fondamental est $\Omega = \{S, E\}$ sur lequel on construit une variable aléatoire discrète X en lui donnant la valeur 1 lors d'un succès et 0 lors d'un échec.

On appelle **variable de Bernoulli** (ou variable indicatrice) la variable aléatoire X telle que $X : \Omega \rightarrow \mathbb{R}$ et $X(\Omega) = \{0, 1\}$.

La loi de probabilité associée à la variable de Bernoulli X telle que

$$\begin{aligned}\mathbb{P}(X = 1) &= p \\ \mathbb{P}(X = 0) &= 1 - p\end{aligned}$$

est appelée **loi de Bernoulli** noté $\mathcal{B}(1, p)$.



Loi de Bernoulli : propriétés

Proposition

Si $X \sim \mathcal{B}(1, p)$ alors $E[X] = p$ et $Var(X) = p(1 - p)$.

Démonstration.

$$E[X] = \sum_{i=1}^2 x_i p_i = 1 \times p + 0 \times (1 - p) = p$$

$$E[X^2] = \sum_{i=1}^2 x_i^2 p_i = 1^2 \times p + 0^2 \times (1 - p) = p$$

$$Var(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p)$$



Loi Binomiale

Supposons qu'on exécute maintenant n épreuves indépendantes, chacune ayant p pour probabilité de succès et $1 - p$ pour probabilité d'échec. La variable aléatoire qui compte le nombre de succès sur l'ensemble de n épreuves est dite variable aléatoire **binomiale** de paramètres (n, p) .

La loi de probabilité d'une v.a. binomiale de paramètres (n, p) (notée $\mathcal{B}(n, p)$) est donnée par :

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Remarque

Une variable de Bernoulli n'est donc qu'une variable aléatoire binomiale de paramètres $(1, p)$.



Loi Binomiale : exercice

Exercice

On jette cinq pièces équilibrées. Les résultats sont supposés indépendants. Donner la loi de probabilité de la variable X qui compte le nombre de piles obtenus.



Loi Binomiale : propriétés

Proposition

Si $X \sim \mathcal{B}(n, p)$ alors $E[X] = np$ et $\text{Var}(X) = np(1 - p)$.

Démonstration.

$$\begin{aligned} E[X] &= \sum_{k=0}^n k \mathbb{P}(X = k) = \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n k \frac{n(n-1)!}{k(k-1)!(n-k)!} p p^{k-1} (1-p)^{n-k} \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{k=1}^n C_{n-1}^{k-1} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{i=0}^{n-1} C_{n-1}^i p^i (1-p)^{n-1-i} = np(p + 1 - p)^{n-1} = np. \end{aligned}$$



Loi géométrique

On exécute une série d'épreuves indépendantes ayant chacune la probabilité p d'être un succès, $0 < p < 1$, jusqu'à obtenir le premier succès. Si l'on désigne le nombre d'épreuves nécessaires jusqu'à ce résultat par X on aura

$$\mathbb{P}(X = n) = (1 - p)^{n-1}p, \quad n = 1, 2, \dots$$

Une telle variable aléatoire s'appelle v.a. **géométrique**.



Loi géométrique : exercice

Exercice

Une urne contient 3 boules blanches et 4 noires. On tire des boules une par une avec remise jusqu'à l'apparition d'une noire.

- 1 *Quelle est la probabilité qu'il faille exactement 5 tirages ?*
- 2 *Quelle est la probabilité qu'il faille au moins 5 tirages ?*



Loi géométrique : propriétés

Proposition

Si X est une variable aléatoire géométrique alors $E[X] = \frac{1}{p}$ et $Var(X) = \frac{1-p}{p^2}$.

Démonstration.

$$\begin{aligned} E[X] &= \sum_{n=1}^{\infty} n \mathbb{P}(X = n) = \sum_{n=1}^{\infty} n(1-p)^{n-1}p = -p \sum_{n=0}^{\infty} ((1-p)^n)' \\ &= -p \left(\sum_{n=0}^{\infty} (1-p)^n \right)' = -p \left(\frac{1}{1-(1-p)} \right)' = \frac{1}{p}. \end{aligned}$$



Loi de Poisson

Une variable aléatoire X pouvant prendre pour valeur $0, 1, 2, \dots$ est dite de **Poisson avec paramètre** λ s'il existe un réel $\lambda > 0$ tel que

$$\mathbb{P}(X = i) = p_i = e^{-\lambda} \frac{\lambda^i}{i!} \quad i = 0, 1, 2, \dots \quad (1)$$

L'équation 1 définit bien une loi de probabilité puisque

$$\sum_{i=0}^{\infty} \mathbb{P}(X = i) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1.$$



Remarque

Les v.a. de Poisson ont un champ d'application vaste, en particulier du fait qu'on peut les utiliser pour approximer des v.a. binomiales de paramètres (n, p) pour autant que n soit grand et p assez petit pour que np soit d'ordre de grandeur moyen. En pratique, on remplace la loi binomiale par une loi de Poisson de paramètre $\lambda = np$ dès que $n > 30$ et $np < 5$.

Des variables aléatoires qui obéissent à la loi de Poisson :

- *le nombre d'individus dépassant l'âge de 100 ans dans une communauté ;*
- *le nombre de clients pénétrant dans un bureau de poste en l'espace d'un jour ;*
- *le nombre de faux numéros téléphoniques composés par jour.*

Loi de Poisson : exemple

Exemple

Soit une loi binomiale de paramètres $(100, 0.01)$, les valeurs des probabilités pour k de 0 à 5 ainsi que leur approximation à 10^{-3} avec une loi de Poisson de paramètre $(\lambda = np = 1)$ sont données dans le tableau ci-dessous :

i	0	1	2	3	4	5
p_i	0.366	0.370	0.185	0.061	0.015	0.000
approximation	0.368	0.368	0.184	0.061	0.015	0.003

Remarque

Cette approximation est correcte car $n > 30$ et $np < 5$.

Loi de Poisson : exercice

Exercice

Admettons que le nombre d'erreurs par page dans un livre suive une distribution de Poisson avec paramètre $\lambda = 0,5$. Calculer la probabilité qu'il ait au moins une erreur sur la page numéro 10.



Loi de Poisson : propriétés

Proposition

Si $X \sim \mathcal{P}(\lambda)$ alors $E[X] = \lambda$ et $Var(X) = \lambda$.

Démonstration.

$$\begin{aligned} E[X] &= \sum_{i=0}^{\infty} \frac{ie^{-\lambda}\lambda^i}{i!} = \lambda \sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned}$$

