

Notion de Variable aléatoire

Lois d'une variable aléatoire

Faculté de Médecine, UDL, SBA
1^{ère} année Médecine

23 Novembre 2016

Plan de cours

1 Variable aléatoire

- Variable aléatoire: Définition

2 Loi d'une variable aléatoire sur un espace de probabilité fini

- Loi d'une variable aléatoire discrète
- Loi d'une variable aléatoire continue
- Caractéristiques d'une Variable Aléatoire
- Fonction de répartition

Variable aléatoire: Introduction

- Une variable aléatoire désigne la grandeur mesurée lors d'une expérience aléatoire. Exemples :
 - Âge (en années)
 - Tension artérielle systolique (en mmHg)
 - Stades de gravités d'une maladie (0-1-2-3-4)
 - Sexe (Homme/Femme)
 - Couleur des yeux (marron, noir, vert, bleu, ...)
 - Cancer (Presence/Absence)
 - Nombre de malades
- Mesures qui varient d'un individu à l'autre.
- Résultats possibles de l'expérience \Rightarrow valeurs possibles de la variable aléatoire

- Types de variables aléatoires

- ◇ Si résultats numériques (variable quantitative)
 - ▷ V.a. continue : les valeurs couvrent \mathbb{R} ou un intervalle
 - ▷ V.a. discrète : les valeurs sont discrète (discontinues) (\mathbb{N})
- ◇ Sinon (variable qualitative)
 - ▷ V.a. ordinale : les valeurs sont ordonnées
 - ▷ V.a. nominale ou catégorielle : valeurs sans ordre

Variable aléatoire: Exemple de codage 1

- Si Pile, A gagne 10 DA
- Si Face, A perd 10 DA
- $\Omega = \{Pile, Face\}$, $\Pr(Pile) = \Pr(Face) = 0.5$
- G: gain de A:
 - $G = +1$, si Pile;
 - $G = -1$, si Face
- $\Pr(G = +1) = \Pr("Pile") = 0.5$ et $\Pr(G = -1) = \Pr("Face") = 0.5$
- Distribution de G : $\{(+1, 0.5), (-1, 0.5)\}$

G : variable aléatoire qui suit une certaine loi de probabilité

Variable aléatoire: Définition

- Soit E un ensemble d'événements d'ensemble fondamental Ω fini
- a un événement élémentaire de E
- **Définition:**

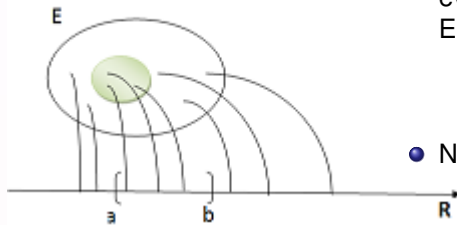
Pour tout événement a appartenant à E on fait correspondre un nombre x (variable aléatoire) selon une loi bien définie

Variable aléatoire: Définition

- Au sens strict, une **variable aléatoire** produit toujours un **résultat numérique**
- Une v.a. est un codage des résultats d'expérience conduisant à des nombres
 - Si les résultats sont déjà numériques, le **codage** peut être l'identité
- On va reprendre ce qui a été vu (événements et probabilités) pour l'adapter aux v.a.

Variable aléatoire: Définition

- A tout résultat de E , la v.a. fait correspondre un nombre
- Sur \mathbb{R} les événements sont les intervalles (et combinaisons)
- Tout intervalle $[a;b]$ définit un événement (sous-ensemble) de E
 - ◊ Si E est \mathbb{R} , certains codage ne sont pas des v.a.



- Notations
 - ◊ $X = \text{v.a.} : X(r) \in \mathbb{R}$
 - ◊ $[a \leq X(r) \leq b] = [a \leq X \leq b]$
= résultats r de E dont le code est entre a et b (événement)

Variable aléatoire: Exemple

- Jeu de dé : $\Omega = \{f_1, f_2, f_3, f_4, f_5, f_6\}$
- V.a. : $X(f_1) = 10; X(f_2) = 20, \dots, X(f_6) = 60$
- $[20 \leq X \leq 35] = \{f_2, f_3\}$
- $[X < 35] = \{f_1, f_2, f_3\}$
- $[X = 40] = \{f_4\}$
- $[X = 35] = \emptyset$

Variable aléatoire: Exemple 2

- Si Ω est l'ensemble des élèves d'une classe, on peut à chaque élève ω associer le nombre $X(\omega)$ de ses frères et soeurs. X est alors une fonction à valeur dans un ensemble fini: c'est une variable aléatoire discrète.
- Somme des résultats de deux dés :

$$\begin{array}{ccc} X : \Omega & \longrightarrow & E \\ (m, n) & \longmapsto & m + n . \end{array}$$

Et on a par exemple $\{X = 3\} = X^{-1}(3) = \{(1, 2), (2, 1)\}$.

- Soit (Ω, \mathbf{Pr}) donné et $X : \Omega \rightarrow E$ une variable aléatoire.
- On pose $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ et $E = \{x_1, x_2, \dots, x_k\}$.
- Définissons:

$$p_i = \mathbf{Pr}(X = x_i) = \mathbf{Pr}(\{\omega \in \Omega / X(\omega) = x_i\})$$

- On remarque que:

$$\begin{aligned}\sum_{i=1}^n p_i &= \sum_{i=1}^n \Pr(X = x_i) = \sum_{i=1}^n \Pr(\{\omega \in \Omega / X(\omega) = x_i\}) \\ &= \Pr(\cup_{i=1}^n \{\omega \in \Omega / X(\omega) = x_i\}) \\ &= \Pr(\Omega) = 1 ,\end{aligned}$$

(puisque ce sont des ensemble disjoints).

- Donc (p_1, p_2, \dots, p_k) définit une probabilité sur E .

Définition

On appelle loi de la variable aléatoire discrète X la donnée des probabilité $\Pr(X = x_i)$ (la probabilité \Pr_X sur E telle que: $p_i = \Pr(X = x_i)$).

Loi de variable aléatoire discrète: Exemple 1

- On lance deux pièces de monnaie. L'ensemble fondamental comprend 4 événements élémentaires notés PP; PF; FP; FF, de probabilité chacun $1/4$.
- On note X la variable aléatoire qui compte le nombre de piles obtenus.
- X prend les valeurs 0, 1, 2.
- $\Pr(X = 0) = 1/4$, $\Pr(X = 1) = 1/4$ et $\Pr(X = 2) = 1/4$.
- On représente souvent la loi de probabilité à l'aide d'un tableau:

X	0	1	2
$\Pr(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Loi d'une variable aléatoire discrète: Exemple2

- Lancer successif de 2 dés.
- Soit X la variable "somme".
- Sa loi sera une probabilité sur l'ensemble $\{2, 3, \dots, 12\}$.
Par exemple:

$$\mu_X(\{3\}) = \mathbf{Pr}(X^{-1}(\{3\})) = \mathbf{Pr}(\{(1, 2), (2, 1)\}) = \frac{2}{36} = \frac{1}{18}.$$

Loi d'une variable aléatoire discrète: Exemple 3

- Extraire 4 boules d'un sac contenant 4 boules blanches et 6 boules rouges. Quelle est la loi de nombre de boules blanches ?

Les événements distincts: "obtenir 0 boule blanche (b.b.)", "obtenir 1 b.b.", "obtenir 2 b.b.", "obtenir 3 b.b." et "obtenir 4 b.b.".

Loi d'une variable aléatoire discrète: Exemple 3

Soit X : nombre des boules blanches; donc:

$$X(\omega) = \{0, 1, 2, 3, 4\}.$$

Loi de probabilité de X est : $\Pr_X = p_i = \Pr(X = x_i)) = ?$

- cas possibles: $C_{10}^4 = 210$,
- cas favorables pour "0 b.b.": $C_4^0 \times C_6^6 = 15$,
- cas favorables pour "1 b.b.": $C_4^1 \times C_3^6 = 80$,
- cas favorables pour "2 b.b.": $C_4^2 \times C_2^6 = 90$,
- cas favorables pour "3 b.b.": $C_4^3 \times C_1^6 = 24$,
- cas favorables pour "4 b.b.": $C_4^4 \times C_0^6 = 1$.

Loi d'une variable aléatoire discrète: Exemple 3

Donc la loi de X est:

X	0	1	2	3	4
$\Pr(X = x_i)$	$\Pr(X=0)$ $= 15/210$	$\Pr(X=1)$ $= 80/210$	$\Pr(X=2)$ $= 90/210$	$\Pr(X=3)$ $= 24/210$	$\Pr(X=4)$ $= 1/210$

et on a $\sum_{i=1}^5 \Pr(X = x_i) = 1$.

- Soit une maladie M pour laquelle il est nécessaire de débiter un TRT avant confirmation du diagnostic. Le médicament utilisé est cependant connu pour entraîner des effets indésirables (EI).
- On sait que:
 - $\Pr(M^+) = 5\%$;
 - $\Pr(EI^+ / M^+) = 30\%$;
 - $\Pr(EI^- / M^-) = 85\%$;

Variable aléatoire: Exemple

	M^+	M^-
EI^+ ($X = 1$)	$\Pr(EI^+ \cap M^+)$ $= 0.3 * 0.05 = 1.5\%$	$\Pr(EI^+ \cap M^-) = (1 - 0.85) * (1 - 0.05) = 14.3\%$
EI^- ($X = 0$)	$\Pr(EI^- \cap M^+)$ $= (1 - 0.3) * 0.05 = 3.5\%$	$\Pr(EI^- \cap M^-) = 0.85 * (1 - 0.05) = 80.8\%$

- X est une v.a. indicatrice des EI .
- $\Pr(X = 0) = \Pr(EI^-) = \Pr(EI^- \cap M^+) + \Pr(EI^- \cap M^-) = 0,84$;
- $\Pr(X = 1) = \Pr(EI^+) = \Pr(EI^+ \cap M^+) + \Pr(EI^+ \cap M^-) = 0,16$;
- La distribution de X est : $\{(0; 0,84), (1; 0,16)\}$

Remarques:

- Les lois de distribution, ou distributions de probabilités, montrent la probabilité d'apparition de toutes les valeurs d'une variable théorique.
- Les lois de distribution servent d'abord au calcul direct de la probabilité d'apparition de certains événements, une fois que l'on connaît la loi de distribution du phénomène.
- Lors des tests statistiques paramétriques, nous nous référerons à ces mêmes lois de distribution pour trouver la probabilité que les données soient conformes à une certaine hypothèse nulle.
⇒ La compréhension des lois de distribution suppose que la notion de probabilité soit connue et comprise

Loi d'une variable aléatoire continue

- Densité de proba $f(x)$
- loi de probabilité
$$\Pr([a \leq X \leq b]) = \int_a^b f(x)dx = F(b) - F(a)$$
- Propriétés:
 - $f(x) \geq 0$
 - $\int_{-\infty}^{\infty} f(x)dx = 1$
 - La fonction de répartition: $F(x) = \int_{-\infty}^x f(t)dt$

Caractéristiques de position: Moyenne ou Espérance

- **Moyenne** au niveau de la **population**
- Notation $\mathbb{E}(X) = \mu_X = \mu$
- Calcul : somme de toutes les valeurs pondérées par leur probabilité
 - ◇ Variable discrète X
 - ★ soit X une *v.a.* prenant les valeurs x_1, x_2, \dots, x_n avec les probabilités p_1, p_2, \dots, p_n et $\sum p_i = 1, i = 1, \dots, n$:

$$\mu = \mathbf{E}(X) = \sum_{i=1}^n p_i x_i$$

- ◇ Cas d'une variable continue X
 - ★ définie par une loi de densité $f(x)$

$$\mu = \mathbf{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Soient des v.a. X et Y et des constantes a, c

- $\mathbb{E}(c) = c$
- $\mathbb{E}(X + c) = \mathbb{E}(X) + c$
- Si $c = -\mathbb{E}(X) \Rightarrow \mathbb{E}(X - \mathbb{E}(X)) = \mathbb{E}(X) - \mathbb{E}(X) = 0$
Une v.a. d'espérance nulle est dite **centrée**
- $\mathbb{E}(aX) = a\mathbb{E}(X)$
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

- Exemple 1 : $\mu = (p \times 1) + (q \times 0) = p$
- Exemple 2 : $\mu = 1/6 + 2/6 + 3/6 + 4/6 + 5/6 + 6/6 = 3,5$
- Exemple 3 :

$$\mu = \mathbb{E}(X) = \int_{0.5}^{2.5} xf(x)dx = \int_{0.5}^{2.5} xexp(x/2)dx = 2.21$$

Caractéristiques de dispersion: Variance, Écart-type

- **Variance** = mesure de la variabilité autour de l'espérance
- Notation $\text{var}(X) = \sigma_X^2 = \sigma^2$
- Définition: $\text{var}(x) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$
- Calcul
 - Cas d'une variable discrète X

$$\sigma^2 = \sum_{i=1}^n p_i [x_i - \mu]^2 = \sum_{i=1}^n p_i^2 x_i^2 - (\mathbb{E}(X))^2$$

- Cas d'une variable continue X

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - (\mathbb{E}(X))^2$$

- σ^2 = Variance , σ = Écart-type

- $Var(X) \geq 0$ (somme de carrés)
 - Variance nulle pour une constante
 - Variance faible pour une variable peu dispersée
- Si X possède une unité
 - ◇ $\mathbb{E}(X)$ et σ ont la même unité
 - ◇ $Var(X)$ a cette unité au carré
- Si c est une constante
 - ◇ $Var(c) = 0$
 - ◇ $Var(X + c) = Var(X)$
 - ◇ $Var(cX) = c^2 Var(X)$
- $Var(X + Y) = ?$

Caractéristiques de dispersion: Exemple

- Exemple 1 : $\sigma^2 = p \times (1 - p)^2 + q \times (0 - p)^2 = p \times q$

- Exemple 2 :

$$\sigma^2 = 1/6[(1 - 3,5)^2 + (2 - 3,5)^2 + \dots + (6 - 3,5)^2] = 2.9$$

- Exemple 3 : $\sigma^2 = \int_{0.5}^{2.5} (x - 2.21)^2 \exp(x/2) =$

$$\int_{0.5}^{2.5} x^2 \exp(x/2)^2 - 2.21^2 = 0.68$$

Fonction de répartition

- Soit X une v.a. quantitative
- On cherche une fonction définissant la probabilité de tout intervalle $[a; b]$
- Soit l'événement $[X \leq x]$ où x est un nombre, i.e., ensemble des résultats d'expérience dont le codage est inférieur ou égal à x .
- $\Pr([X \leq x])$ dépend de la valeur x
- $F_X(x) = F(x) = \Pr([X \leq x]) =$ **fonction de répartition de X**

Fonction de répartition: Définition

C'est la fonction $F : x \in E \subset \mathbb{R} \rightarrow [0, 1]$, définie par:

$$F(x) = \mathbf{Pr}(X \leq x) .$$

- F permet de calculer la probabilité pour que X soit compris entre deux valeurs a et b :

$$\mathbf{Pr}(a \leq x \leq b) = F(b) - F(a) ,$$

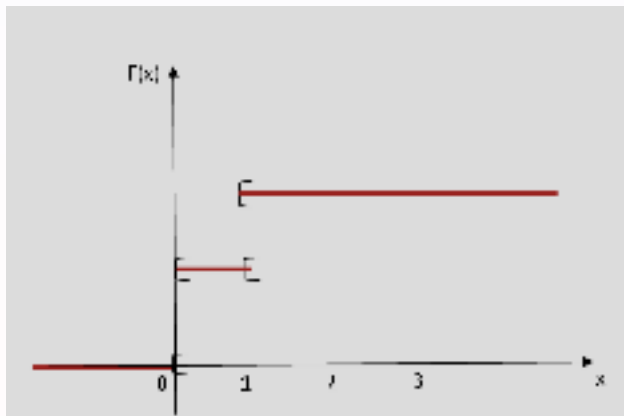
(penser à la fonction des fréquences cumulées)

Fonction de répartition: Propriétés

- ❶ F est monotone croissante;
- ❷ $0 \leq F(x) \leq 1$;
- ❸ $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$;
- ❹ F est continue à droite en tout x réel.
- ❺ $a < b \Rightarrow \mathbf{Pr}([X \leq b]) = \mathbf{Pr}([X \leq a]) + \mathbf{Pr}([a < X \leq b])$
car $[X \leq a]$ et $[a < X \leq b]$ = événements exclusifs

Fonction de répartition: exemple d'une v.a. discrète

- Jet d'une pièce : $\Omega = \{P, F\}$; $\Pr(P) = \Pr(F) = \frac{1}{2}$
- V.a. $X : X(F) = 0 ; X(P) = 1$
- Fonction de répartition

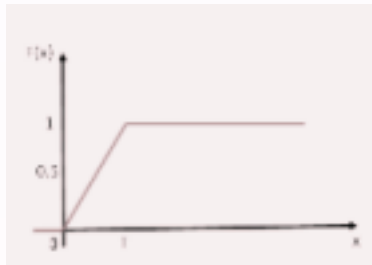


Fonction de répartition : exemple d'une v.a. continue

- Appel téléphonique dans l'intervalle $[0, T]$, $t =$ instant d'appel :

$$\Pr(t_1 \leq t \leq t_2) = (t_2 - t_1)/T, (t_1 \text{ et } t_2 \in [0, T])$$

- Fonction de répartition



- Si $x < 0$, l'appel n'a pas eu lieu avant x : $F(x) = 0$
- Si $x > T$, l'appel a eu lieu avant x : $F(x) = 1$
- Sinon $F(x) = \Pr(0 \leq t \leq x) = x/T$

Pour définir une v.a.: Résumé

	v.a. discrète	v.a. continue
Définition de la loi de probabilité	Tableau des $p_i = \mathbf{Pr}(X = x_i)$	Densité de proba $f(x)$ $\mathbf{Pr}([a \leq X \leq b]) = \int_a^b f(x)dx$ $= F(b) - F(a)$
Propriétés	$p_i \geq 0$ $\sum_{i=1}^n p_i = 1$ $F(x) = \sum_{x_i \leq x} p_i$	$f(x) \geq 0$ $\int_{-\infty}^{\infty} f(x)dx = 1$ $F(x) = \int_{-\infty}^x f(t)dt$ $f(x)dx = \mathbf{Pr}(x \leq X \leq x + dx)$ $f(x)dx \approx \mathbf{Pr}(X = x)$

- Cours Probabilité, Benchikh Tawfik,
<http://www.univ-sba.dz/lsp/s/images/pdf/cours/>
- Probabilités et Biostatistique.
www.chups.jussieu.fr/polys/biostats/CoursProba1.pdf
- FMPMC Pitié-Salpêtrière-probabilités, cours 1 et 2.
www.chups.jussieu.fr/polys/biostats/coursProba-1-2/
- Notions de Probabilités, Roch Giorgi, LERTIM, Faculté de Médecine, Université de la Méditerranée, Marseille, France <http://lertim.fr>