Variables aléatoires, Fonctions de répartition, Densités de probabilité

Nicolas Baskiotis
nicolas.baskiotis@lip6.fr

Université Pierre et Marie Curie (UPMC)
Laboratoire d'Informatique de Paris 6 (LIP6)

Probabilités sur des ensembles dénombrables

Notions fondamentales

- Univers : un ensemble dénombrable (fini ou infini) Ω,
- Événement : un sous-ensemble de l'univers Ω,
- Mesure de probabilité :
  une fonction qui associe à chaque événement une valeur entre 0 et 1, la probabilité de Ω est 1, et la probabilité d'une union dénombrable d'événements incompatibles (ensembles disjoints) est la somme de leurs probabiilités.
- *espace probabilisé* :
  un couple (Ω, P), où P est une mesure de probabilité sur Ω,
- Variable aléatoire $\longleftarrow$ cours d'aujourd'hui.

Intuition

Variable aléatoire

Lorsqu'on est face à une expérience aléatoire, on s'intéresse plus souvent à une *valeur* attribuée au résultat qu'au résultat lui-même.

Exemples

- Lorsque l'on joue à un jeu de hasard on s'intéresse plus au gain que l'on peut obtenir qu'au résultat du jeu.
- Nombre de pannes dans un ensemble de systèmes plutôt que l'état exact des systèmes

- problème : souvent trop long à énumerer, peu informatif.
- solution : “traduire” l'univers en évènements “compréhensibles“.

$\Rightarrow$ variable aléatoire : application de l'univers vers un ensemble discret (ou plus vaste).

Exemple

Exemple du lancer de dé

On lance un dé après avoir misé 1€. Si le résultat est un 5 ou un 6 on double la mise, sinon perd la mise. Dans ce cas:

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Card $\Omega = 6$, Card $\mathcal{P}(\Omega) = 2^6$, et $\forall e \in \Omega, P(e) = \frac{1}{6}$
- Soit X la $v.a.$ qui associe à tout résultat du dé un *gain*:
  $X(1) = X(2) = X(3) = X(4) = -1$ et $X(5) = X(6) = (2 - 1) = 1$
  X est à valeur dans l'ensemble noté $\mathcal{X} = \{-1, 1\} \subset \mathbb{R}$, $X : \Omega \to \mathcal{X}$
- Question: Comment calculer la probabilité de gagner 1€?
- Réponse: Définir une probabilité sur $\mathcal{X}$, notée $\mathbb{P}$, en retournant dans l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$
  i.e. utiliser $P($*Le résultat du dé est* 5 *ou* $6)$ pour estimer $\mathbb{P}(\{1\})$.

Variable aléatoire à valeurs discrètes

Définition

Soit Ω un ensemble dénombrable, et P une mesure de probabilité sur Ω.
Soit Ω', un ensemble discret.
Une variable aléatoire est une fonction X de Ω muni de la mesure P vers Ω'.

Exemples

- Lancer d'un dé :
  Soit $\Omega = \{1, ..., 6\}$ muni de la probabilité uniforme P.
  $X : i \mapsto \begin{cases} 1 \text{ si } i \text{ est pair} \\ 0 \text{ sinon} \end{cases}$
  est une variable aléatoire de (Ω, P) vers $\Omega' = \{0, 1\}$.
- Lancer de deux dés :
  Soit $\Omega = \{1, ..., 6\}^2$ muni de la probabilité uniforme P.
  $X : (i,j) \mapsto i + j$
  est une variable aléatoire de (Ω, P) vers $\Omega' = \{2, ..., 12\}$

Loi de probabilité

Définitions

Soit (Ω, P) un espace probabilisé où Ω est dénombrable.
Soit Ω' un ensemble discret, et X une v.a. de (Ω, P) vers Ω'.

- X définit une mesure de probabilité sur Ω', notée P_X, par:
  pour tout sous-ensemble E' de Ω' :

$$P_X(E') = P\big(X^{-1}(E')\big)$$

  avec $X^{-1}(E') = \{\omega \in \Omega | X(\omega) \in E'\}$
- L'ensemble des valeurs $P_X\big(\{\omega'\}\big)$ pour $\omega' \in \Omega'$ s'appelle la *loi de probabilité* de X.

Notations

- L'événement $X \in]-\infty, a]$ sera noté par $X \leq a$
- L'événement $X \in]a, b]$ sera noté par $a < X \leq b$
- L'événement $X \in \{a\}$ sera noté par $X = a$
- On a donc $P_X(B) = P(X^{-1}(B)) = P(X \in B)$

Loi d'une variable aléatoire

Propriété

Une variable aléatoire est totalement définie par sa loi de probabilité, caractérisé par :

- son domaine de définition : l'ensemble des valeur qu'elle peut prendre,
- les probabilités attribuées à chacune de ses valeurs $P(X = x)$.

Questions :

soit Ω un ensemble de cardinal n,

- quel est le plus grand cardinal de l'ensemble des valeurs d'une application de Ω ?
- combien d'applications de Ω vers $\{1, ..., n\}$ différentes existe-t-il ?

Exemple

Jeux de hasard

- On lance un dé après avoir misé 3 euros. Si le résultat est $1, 2, 3$ ou 4, on perd la mise. Sinon, on triple la mise.
  Quelle est la probabilité de gagner 6 euros ? de perdre sa mise ?
  $\Omega = \{1, ..., 6\}$ muni de la probabilité uniforme. On note X la v.a. qui représente le gain:
  $$P(X = 6) = P\big(\{5, 6\}\big) = \frac{1}{3}$$
  $$P(X = -3) = \frac{2}{3}$$
- Le joueur décide de jouer deux fois de suite. Quelle est la loi de probabilité du gain sur l'ensemble des deux lancers ?
  L'univers est $\{1, ..., 6\}^2$, muni de la loi de probabilité uniforme.
  Le gain total G suit la loi :
  $$P(G = -6) = P\big(\{1, ..., 4\}^2\big) = \frac{16}{36}$$
  $$P(G = 3) = ?$$
  $$P(G = 12) = ?$$

Fonction de répartition

Définition

Soit X une $v.a.r.$
on appelle **fonction de répartition** de X, notée F la fonction:

$$\begin{array}{rcl} F : \mathbb{R} & \to & [0,1] \\ x & \mapsto & F(x) = P(X \leq x) \end{array}$$

Propriétés

- $P(a < X \leq b) = F(b) - F(a)$
- F est croissante bornée:
  - $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0$
  - $F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1$

Fonction de répartition (suite)

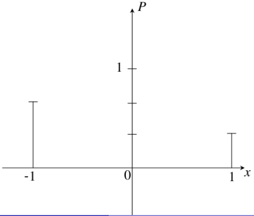
Exemple: Lancer de dé

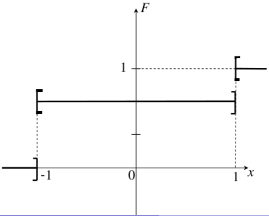
Avec l'exemple du lancer de dé précédent nous avons $\Omega = \{1,2,3,4,5,6\}$, et $X : \Omega \to \mathbb{R}$ définie par:

$$X(e) = -1 \text{ si } e \in \{1,2,3,4\}, \text{ et } X(e) = +1 \text{ si } e \in \{5,6\}$$

L'ensemble des valeurs possibles est $\mathcal{X} = \{-1, 1\}$. On peut alors caractériser la loi de X par sa fonction de répartition $F : F(x) = P(X \leq x)$

- si $x < -1, \mathcal{X} \cap (]-\infty, x]) = \emptyset \Rightarrow F(x) = 0$
- si $x \in [-1, 1[, \mathcal{X} \cap (]-\infty, x]) = \{-1\} \Rightarrow F(x) = \frac{2}{3}$
- si $x \in [1, \infty[, \mathcal{X} \cap (]-\infty, x]) = \{-1, 1\} \Rightarrow F(x) = \frac{2}{3} + \frac{1}{3} = 1$





Variables aléatoires indépendantes

Définitions

Soit (Ω, P) un espace probabilisé.

- Soient X et X' deux v.a. de Ω vers Ω'.
  Les variables X et X' sont indépendantes si :

$$\forall A \subset \Omega', \forall B \subset \Omega', P(X \in A \cap X' \in B) = P(X \in A)P(X' \in B)$$

- Soient X_1, ..., X_n, n v.a. de Ω vers Ω'.
  $X_1, ..., X_n$ sont mutuellement indépendantes si, pour tous sous-ensembles $E_1, ..., E_n$ de Ω', on a:

$$P\left(\bigcap_{i\in\{1,\cdots,n\}} X_i \in E_i\right) = \prod_{i\in I} P(X_i \in E_i)$$

Retour à l'exemple précédent

Si on lance deux fois un dé, avec à chaque fois un gain si le résultat est 5 ou 6, les gains obtenus à chacun des lancers sont indépendants.

Exemple

Résultats de n répétitions indépendantes d'une expérience aléatoire

Soit (Ω, P) un espace probabilisé, X une v.a. sur Ω vers Ω'.
On note $\Omega_n = \Omega^n$, et P_n la mesure produit sur Ω_n :

$$\forall \omega = (\omega_1, ..., \omega_n) \in \Omega_n, P_n(\{\omega\}) = \prod_{i=1}^{n} P(\omega_i)$$

On note $X_i : \omega \in \Omega_n \mapsto X(\omega_i)$.
Les v.a. X_i sont mutuellement indépendantes et suivent la même loi que X :

$$\forall i, \forall E' \subset \Omega', P_n(X_i \in E') = P(X \in E')$$

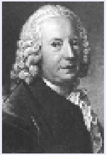
Retour à l'exemple précédent (2)

Si on lance n fois un dé, avec à chaque fois un gain si le résultat est 5 ou 6.
Ω_n est l'ensemble des réalisations possibles des n lancers.
Pour un événement élémentaire $\omega = (\omega_1, ..., \omega_n)$:

- ω_i est le résultat du i-ième lancer,
- X_i représente le gain obtenu au i-ième lancer.

Deux lois de probabilités discrètes importantes

Loi de Bernoulli



La loi de Bernoulli est la loi d'une $v.a.$ X à valeur dans $\{0, 1\}$.
$X = 1$ représente le "succès" de l'expérience, et $X = 0$ l'"échec".

$$\forall x \in \{0, 1\}, P(X = x) = p^x(1 - p)^{1-x}$$

La probabilité de succès $p = P(X = 1)$ est le paramètre de la loi.

Loi binomiale

Soit X, le nombre de succès d'une épreuve de Bernoulli de paramètre p, répétée n fois indépendamment. La loi de X est appelée la *loi binomiale* de paramètres n et p :

$$\forall k \in \{0, ..., n\}, P(X = k) = C_n^k p^k (1 - p)^{n-k}$$

Loi conjointe

Definition

Soit (Ω, P) un espace de probabilité, et soient X et Y deux $v.a.$ sur cet espace, à valeur resp. dans F et G. (X, Y) est une $v.a.$, appelée loi conjointe de X et Y; les valeurs de (X, Y) sont dans $F \times G$.

Propriétés

- la connaissance uniquement de X et de Y ne suffit pas à connaitre la loi jointe, sauf si X est indépendant de Y.
- $\forall x \in F, P(X = x) = \sum_{y \in G} P(X = x, Y = y)$

$\Rightarrow$ la connaisance de la loi jointe permet de déduire la loi de X, appelée dans ce cas *loi marginale*.

Exemple

Soit (X, Y) un couple de $v.a.$ de loi telle que $P((X, Y) = (i,j)) = 1/9$ ssi $0 \leq i \leq 2$ et $-i \leq j \leq i$.

- Quelle est la représentation graphique de la loi ?
- Quelles sont les lois marginales de X et de Y ?

Test de primalité de Fermat

répéter k fois :
    soit a un nombre aléatoire entre 1 et $n-1$
    si a^{n-1} mod n $!=$ 1 alors
        retourner "n n'est pas premier"
    finsi
fin de la boucle
retourner "n est probablement premier"

Principe

- Algorithme probabiliste :
  fait appel à un générateur de nombre aléatoire durant son exécution.
- Si n est premier, $\forall a \in \{1, ..., n-1\}$, a^{n-1} mod $n == 1$,
  si n n'est pas premier et n'est pas un nombre de Carmichael (très rares), plus de la moitié des entiers a entre 1 et $n-1$ vérifient : a^{n-1} mod n $!=$ 1.

Test de primalité de Fermat (2)

1. répéter k fois :
2. soit a un nombre aléatoire entre 1 et $n-1$
3. si a^{n-1} mod n $!=\ 1$ alors
4. retourner “n n’est pas premier”
5. finsi
6. fin de la boucle
7. retourner “n est probablement premier”

Question

Supposons que n n’est pas premier, ni un nombre de Carmichael.
Pour quelles valeurs de k l’algorithme retourne-t’il

“n n’est pas premier”

avec une probabilité plus grande que $1 - 10^{-6}$?