BONJOUR

Méthodes Statistiques pour l'Informatique

Cours de N. SAHEB

Table des Matières

Références

- R. GRAHAM, D. KNUTH, O. PATASHNIK, *Mathématiques Concrètes,* International Thomson Publishing, 1998.
- F. DRESS, *Probabilités Statistique,* Dunod, 1997.
- J. ISTAS, *Probabilités et Statistiques,* Elipses, 1999.

Probabilités Discrètes

Exemple. On lance deux dés.

$$\Omega = D \times D = \left\{ ⚀⚀\ ,\ ⚀⚁\ , \ldots,\ ⚅⚅ \right\}$$

avec

$$D = \left\{ ⚀\ ,\ ⚁\ ,\ ⚂\ ,\ ⚃\ ,\ ⚄\ ,\ ⚅ \right\}$$

L'espace Ω contient $6 \times 6 = 36$ éléments.

Distribution uniforme :

tous les éléments de Ω sont de même probabilité $\frac{1}{36}$

$$\Longleftrightarrow$$

tous les éléments de D sont de même probabilité $\frac{1}{6}$.

****

Définitions. Un *espace probabilisé discret* est un couple (Ω, Pr), où Ω est un ensemble non vide au plus dénombrable et Pr une application de Ω dans $[0,1]$ telle que :

$$\sum_{\omega \in \Omega} Pr(\omega) = 1.$$

- Ω : *espace des événements élémentaires*
- $A \subset \Omega$: *événement*
- Pr : une *loi* (ou *distribution*) *de probabilité sur* Ω
- On prolonge Pr sur $\mathcal{P}(\Omega)$ par :

$$Pr(A) = \sum_{\omega \in A} Pr(\omega), \quad \forall A \subset \Omega$$

$Pr(A)$: *probabilité de* A

Proposition

- $Pr(\emptyset) = 0$
- $Pr(\bar{A}) = 1 - Pr(A)$
- $Pr(\cup_{i \in I} A_i) = \sum_{i \in I} Pr(A_i)$, pour toute famille au plus dénombrable $A_i, i \in I$ d'éléments de $\mathcal{P}(\Omega)$ 2-à-2 disjoints.

Exemple. On peut adopter une hypothèse non uniforme en posant :

$$Pr_2(⚀) = Pr_2(⚅) = \frac{1}{4}$$

$$Pr_2(⚁) = Pr_2(⚂) = Pr_2(⚃) = Pr_2(⚄) = \frac{1}{8}$$

Sur $\Omega = D \times D$, nous pouvons définir la distribution produite :

$$Pr_{22}(dd') = Pr_2(d)Pr_2(d')$$

Dans le cas où les deux dés sont pipés, nous avons

$$Pr_{22}(⚅⚂) = \frac{1}{4}.\frac{1}{8} = \frac{1}{32}.$$

On démontrera facilement que Pr_{22} est bien une distribution de probabilité sur Ω.

Définition. Soit (Ω, Pr) un espace probabilisé discret et soit Ω' un ensemble non vide au plus dénombrable. Une *variable aléatoire* (v.a.) X à valeurs dans Ω' est une application de Ω dans Ω'. Nous prenons souvent pour Ω' un sous-ensemble de $\mathbb{N}$ ou de $\mathbb{R}$.
On pourra munir Ω' d'une loi de probabilité Pr_X en posant, pour tout $\omega' \in \Omega'$:

$$Pr_X(\omega') = Pr(X^{-1}(\{\omega'\}))$$

Proposition. Pr_X est une loi de probabilité sur Ω', i.e.

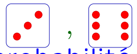
$$\sum_{\omega' \in \Omega'} Pr_X(\omega') = 1.$$

Nous avons de plus, pour tout $A' \subset \Omega'$:

$$Pr_X(A') = Pr(X^{-1}(A'))$$

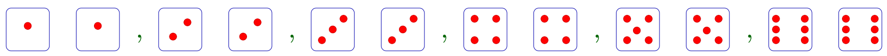
Lorsqu'il n'y a pas le danger de confusion, Pr_X sera désignée par Pr.

Exemple.

Soit $S(\omega)$ la v.a. qui est la somme des points des deux dés, par exemple $S($  $) = 3 + 6 = 9$. Le tableau ci-dessous caractérise la loi de probabilité pour la v.a. S par rapport à chacune des distributions (uniforme et biaisée) :

s	2	3	4	5	6	7	8	9	10	11	12
$Pr_{11}(S = s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$Pr_{22}(S = s)$	$\frac{4}{64}$	$\frac{4}{64}$	$\frac{5}{64}$	$\frac{6}{64}$	$\frac{7}{64}$	$\frac{12}{64}$	$\frac{7}{64}$	$\frac{6}{64}$	$\frac{5}{64}$	$\frac{4}{64}$	$\frac{4}{64}$

A titre d'exemple, on peut définir l'événement "double" par :

$A = \{$  $\}$

Or nous avons par définition :

$$Pr(A) = \sum_{\omega \in A} Pr(\omega)$$

Pour la distribution uniforme, cette probabilité vaut

$$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}.$$

Alors que, pour Pr_{22}, elle vaut

$$\frac{1}{16} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{16} = \frac{3}{16}.$$

Espérance Mathématique et Variance

Une variable aléatoire admet un certain nombre de valeurs typiques. Nous considérons dans la suite les v.a. à valeurs dans $\mathbb{R}$.

Définition. En arithmétique usuelle la valeur moyenne de n nombres est définie comme leur somme divisée par n. En calcul des probabilités, l'*espérance* d'une v.a. est définie comme la somme des valeurs prises pondérées par les probabilités respectives, c'est-à-dire :

$$\mathbb{E}X = \sum_{x \in X(\Omega)} x.Pr(X = x)$$

lorsque cette somme converge absolument. ($X(\Omega)$ est l'ensemble des valeurs prises par la v.a. X). Sinon, on dit que X *n'admet pas d'espérance*.

Exemple. Revenons au cas de la v.a. S, désignant la somme des deux dés, et calculons son espérance pour chacune des distributions Pr_{11} et Pr_{22} :

- $\mathbb{E}_{11}S = 2.\frac{1}{36} + 3.\frac{2}{36} + ... + 12.\frac{1}{36} \;= 7.$
- $\mathbb{E}_{22}S = 2.\frac{4}{64} + 3.\frac{4}{64} + 4.\frac{5}{64} + ... + 12.\frac{4}{64} \;= 7.$

Etant donné que l'espérance de la v.a. X, désignant le point obtenu par un seul dé, vaut :

- $\mathbb{E}_1X = \frac{1}{6}.(1 + 2 + ... + 6) \;= 3.5$ (pour le dé uniforme)
- $\mathbb{E}_2X = \frac{1}{4}.(1 + 6) + \frac{1}{8}.(2 + 3 + 4 + 5) \;= 3.5$ (pour le dé pipé)

on peut se demander s'il y a un lien entre l'espérance d'une somme de v.a. et la somme des espérances.

Linéarité de l'espérance

Proposition. Soient X_1 et X_2 deux v.a. définies sur le même espace probabilisé discret (Ω, Pr) et admettant toutes deux une espérance. Soit $\alpha \in \mathbb{R}$. Alors :

$$\begin{aligned} \mathbb{E}(\alpha X_1) &= \alpha \mathbb{E}X_1 \\ \mathbb{E}(X_1 + X_2) &= \mathbb{E}X_1 + \mathbb{E}X_2 \end{aligned}$$

Que peut-on dire de l'espérance d'un produit de v.a. ?

Définitions. On peut considérer plusieurs v.a. sur le même espace probabilisé discret. Soient les v.a. X et Y définies sur (Ω, Pr). La loi (ou distribution) *conjointe* est la donnée de :

$$Pr_{XY}(X = x, Y = y) = Pr(\{\omega \in \Omega | X(\omega) = x, Y(\omega) = y\})$$

pour tout x et tout y possibles.
X et Y sont dites *indépendantes*, si pour tout x et tout y possibles, on a $Pr_{XY}(X = x, Y = y) = Pr_X(x)Pr_Y(y)$.

Proposition. Si X et Y sont deux v.a. indépendantes admettant une espérance, alors la v.a. produite XY admet une espérance et

$$\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y$$

.

Exemple. Soient X et Y les points obtenus en lançant deux dés (nous considérons les cas des dés uniformes et celui des dés pipés). Ces deux v.a. sont indépendantes (par la définition de Pr_{11} ou de Pr_{22}). Soient $S = X + Y$ et $P = XY$. Nous avons :

- $\mathbb{E}S = \frac{7}{2} + \frac{7}{2} = 7, \mathbb{E}P = \frac{7}{2} . \frac{7}{2} = \frac{49}{4}$.(un calcul direct fondé sur les probabilités des différentes valeurs de P aboutit au même résultat dans les deux cas.)
- Mais S et P ne sont pas indépendantes. Nous avons toutefois $\mathbb{E}(S + P) = \mathbb{E}S + \mathbb{E}P = 7 + \frac{49}{4}$. Un calcul des probabilités des valeurs de SP permet d'obtenir son espérance : $\mathbb{E}(SP) = \frac{637}{6}$ dans le cas des dés uniformes et 112 dans le cas des pipés ; qui sont différentes, dans les deux cas de

$$\mathbb{E}S . \mathbb{E}P = 7 . \frac{49}{4} = \frac{343}{4}.$$

Définitions. Un paramètre important qui vient tout de suite après l'espérance est la variance. Elle mesure la dispersion d'une v.a. Si X est une v.a. définie sur (Ω, Pr), sa *variance* est définie par :

$$\mathbb{V}X = \mathbb{E}((X - \mathbb{E}X)^2)$$

lorque celle-ci existe.

Un exemple. La probabilité d'un gain égal à 100 millions d'euros pour un billet de loterie est de $\frac{1}{100}$. Nous disposons de deux choix pour acheter deux billets de loterie : nous pouvons acheter soit deux billets distincts, soit deux billets quelconques.

Les espérances dans les deux alternatives sont égales et valent chacune 2 millions d'euros. Comparons-les en fonction de leur variances.

Dans la première alternative (billets distincts), on peut gagner 0 ou 100 millions d'euros avec les probabilités respectives 0.98 et 0.02. Dans la deuxième (billets quelconques), on peut gagner 0, 100 ou 200 millions d'euros avec les probabilités respectives 0.9801, 0.0198 et 0.0001.

Pour la première alternative, nous avons donc la variance suivante (en 10^{12} euros^2 !) :

$$0.98(0 - 2)^2 + 0.02(100 - 2)^2 = 196.$$

Et pour la deuxième :

$$0.9801(0 - 2)^2 + 0.0198(100 - 2)^2 + 0.0001(200 - 2)^2 = 198.$$

On en déduit que la seconde alternative est légèrement plus risquée que la première.

Définition. La racine carrée de la variance est appelée *écart-type* et est notée par σ :

$$\sigma X = \sqrt{\mathbb{V}X}.$$

Proposition. Nous avons :

$$\mathbb{V}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

Proposition. Si X et Y sont deux v.a. indépendantes admettant chacune une variance, alors la v.a. $X + Y$ admet une variance qui est la somme des deux variances.