Probabilités et statistiques - Cours 1

October 28, 2021

- Généralités sur le module -

- MCC.
- Répartition des groupes de TD :

  Imagine + ICo : M. Puech : 36.204.

  IASD + Algo : Mme Baert/Mme Chateau : 36.208

  GL : M. Bessy : 36.209

Plan

Introduction

Bioinformatique ⇒ Des données, des données, des données. . .

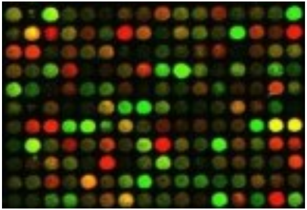
Données ⇒ Traitement des données pour inférer des comportements, propriétés, structures, pour valider des modèles

Traitement informatique ⇒ Outils mathématiques pour valider ces traitements

Introduction

Si l'informatique permet d'accéder aux données, les statistiques permettent d'en décrire le contenu. . .

Combiner les deux ? On parle de science de données et de machine learning: comprendre les données, faire des prédictions, extraire et analyser les motifs...



Plan

Définitions

Definition (Probabilités)

Étude des phénomènes aléatoires ou supposés comme tels.

Definition (Expérience aléatoire)

Observation de l'un de ces phénomènes.

L'ensemble des résultats d'une expérience aléatoire est appelé *ensemble aléatoire* ou encore *univers*, noté Ω.

Exemple

L'expérience aléatoire d'un lancer de dé à 6 faces.
$\Omega = \{1, 2, 3, 4, 5, 6\}$.

- Ω est un ensemble. La terminologie en probabilité est particulière, un peu historique...

- Autres exemples :

Ω = l'état d'une file d'attente à la poste ($= \mathbb{N}$ humanité $^{\text{ieu}}$)

= les températures possibles à Montpellier à 8h00 ($= [-10, 40]$)

= l'état possible de mon compte en banque ($= \mathbb{R}$).

- Définition : variable aléatoire -

Une fonction X de Ω dans $\mathbb{R}$.

$\rightsquigarrow$ Une variable aléatoire n'est ni une variable, ni aléatoire...

- **Exemples :**

- $\Omega = \{$  $\}$

  $X : \Omega \to \{1, 2, 3, 4, 5, 6\}$ la valeur du dé.

  $X : \Omega \to \{0, 1\}$ 0 si la valeur du dé est paire, 1 sinon

- $\Omega =$ Humanité ou (la file d'attente)

  $X : \Omega \to \mathbb{N}$ le nombre de personne dans la file

  $X : \Omega \to \mathbb{N}$ le nombre de personne ayant un chapeau dans la file.

- $\Omega = [-10, 60]$ (la température à 8h00)

  $X : \Omega \to \mathbb{R}^+$ la note du chauffage.

Définitions

Definition (Variable aléatoire)

Une variable aléatoire, ou v.a., est la mesure d'un phénomène aléatoire dont le résultat est numérique (dans $\mathbb{R}$). Elle peut être :

- discrête : l'ensemble Ω est *dénombrable*, par exemple $\mathbb{N}$ ou un ensemble fini
- continue : l'ensemble Ω est $\mathbb{R}$ ou un intervalle de $\mathbb{R}$.

Exemple

la v.a. représentant un tirage aléatoire de dés est discrête, la v.a. représentant la durée de vie d'une ampoule est continue.

Définitions

Évènement = sous-ensemble de Ω

Definition (Évènement)

On appelle évènement un ensemble de résultats.

Exemple

- La longueur de la protéine est 40nm;
- Le lancer de la pièce donne "pile";
- "X=2"; . . .
- "le nbre de personne dans file est pair" ...

Définitions

Definition (Mesure de probabilité)

Une mesure de probabilité est une mesure associant à chaque évènement d'un univers Ω une valeur entre 0 et 1, avec les propriétés suivantes :

1. $\forall E \in \Omega\ P(E) \in [0,1]$ et $P(\Omega) = 1$;
2. $\forall E \in \Omega\ P(\overline{E}) = 1 - P(E)$, où $\overline{E}$ est l'évènement complémentaire de E;
3. $\forall E_1, E_2 \in \Omega\ P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$;
4. $\forall E_1, E_2 \in \Omega\ E_1 \subseteq E_2 \Rightarrow P(E_1) \leq P(E_2)$.

Définitions

Definition (Modèle probabiliste)

Un modèle probabiliste est la donnée d'un univers et d'une mesure de probabilité sur cet univers.

Cas discrêt : À chaque évènement on associe une valeur pour sa mesure de probabilité.

Exemple

Lancer du dé non pipé :
$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$

Pièce truquée : P("pile") = 58% P("Face") = 42%

Définitions

Plus dur d'associer une proba à chaque évènement (tribus, ensembles mesurables ...) On se donne tjs une variable aléatoire

Cas continu : On associe des mesures à des intervalles de valeurs prises par la v.a.

Exemple

$P(X < 3)$, $P(Y \geq 5)$

$P(\text{durée de vie de l'ampoule} < 10000h) < 1/3$

(Chaque évènement du type $X = \ldots$ a une mesure de probabilité nulle.)

Ex: Ω = états d'une file d'attente $X : \Omega \to \mathbb{N}$ nbr de personne dans la file

$P(X \leq 3) = 60\%$

Fonction de répartition

Definition (Fonction de répartition)

Soit X une variable aléatoire. On appelle fonction de répartition de X, notée F_X, la fonction définie par :

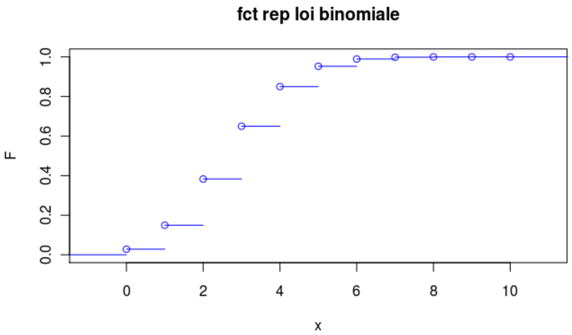
$$F_X : \begin{array}{rcl} \mathbb{R} & \rightarrow & [0,1] \\ x & \mapsto & P(X \leq x). \end{array}$$

- Ex : • Ω = file d'attente, X = nbre de personne. $F_X(3) = P(X \leq 3)$
  • $\Omega = [-10, 40)$, X = note du chauffage $F_X(50) = P(X \leq 50)$

Fonction de répartition

Exemple

Fonction de répartition de la loi binomiale $\mathcal{B}(10, 0.3)$



Fonction de répartition

Propriété

1. *La fonction de répartition est une fonction croissante;*
2. $F_X(x)$ *varie de 0 à 1 lorsque* x *varie de* $-\infty$ *à* $+\infty$*;*
3. F_X *est continue à droite et* $P(X = x_0) = F_X(x_0) - F_X(x_0^-)$, *où* $F_X(x_0^-) = \lim\limits_{x \leq x_0, x \to x_0} F_X(x)$.

Cas des variables aléatoires discrètes

La fonction de répartition reste constante entre deux valeurs x_i possibles.

Elle présente un saut de discontinuité en chaque valeur x_i.

La mesure du saut en x_i correspond à la mesure de probabilité associée à x_i. Fonction de probabilité : $p_X(x_i)$.

Propriété

$p_X(x_i) = F_X(x_i) - F_X(x_{i-1})$
$F_X(x) = \sum_{x_i \leq x} p(x_i)$

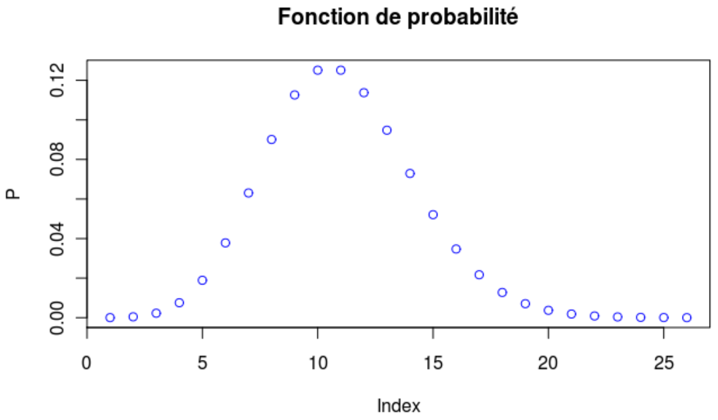
Cas des variables aléatoires discrètes

Exemple

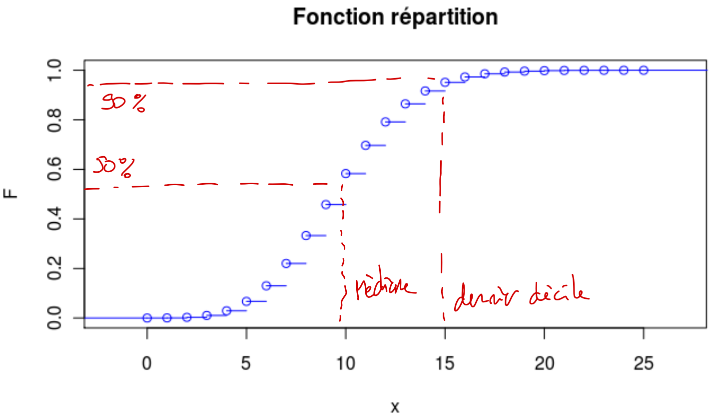
La v.a. X représente le nombre d'appels arrivant à un standard téléphonique en une minute. On associe à cette v.a. une loi de Poisson de moyenne 10. La v.a. X est définie par :

Valeur	0	1	2	...	k	...
Probabilité	e^{-10}	$10.e^{-10}$	$\frac{10^2.e^{-10}}{2}$	...	$\frac{10^k.e^{-10}}{k!}$	...

Cas des variables aléatoires discrètes



Cas des variables aléatoires discrètes



Cas des variables aléatoires continues

Quand la v.a. X est continue, F_X est continue et peut s'écrire sous la forme :

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

où f_X est la fonction de *densité de probabilité* de X.

Propriété

1. $P(X \leq a) = P(X < a)$
2. $P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(u)du$

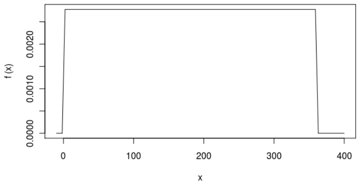
Cas des variables aléatoires continues

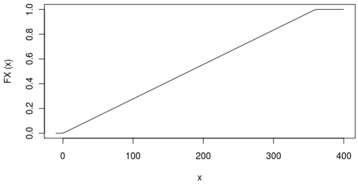
Remarque : Bien que d'un point de vue pratique c'est F_X qui est utile (et qu'on trouve dans les tables), la représentation graphique de f_X est plus parlante car elle met en évidence les zones à plus forte probabilité.

Cas des variables aléatoires continues

Exemple

La Roue de la Fortune. X est l'angle de la flêche par rapport à une origine déterminée. Pas de direction privilégiée : X suit une loi uniforme continue sur [0,360]





Notion de quantile

Definition

Le quantile d'ordre q d'une variable aléatoire X, où $q \in [0, 1]$, est la valeur x_q telle que $P(X \leq x_q) = q$ (ou de même $F_X(x_q) = q$).

Exemple

Pour $q = 0.5$, on parle de la médiane.

Intérêt en statistique : pour fixer des limites de plausibilité pour les valeurs d'une loi donnée, on considérera des quantiles $x_{0.025}$ et $x_{0.975}$ correspondant à des valeurs à l'intérieur desquelles la v.a. a une probabilité 0.95 de se trouver.

Fonction d'une variable aléatoire

Le problème du passage de la loi d'une variable aléatoire X à la loi d'une variable aléatoire exprimée en fonction de X est fréquent : $Z = g(X)$.

Pour connaître la loi de Z connaissant celle de X, on essaie de résoudre l'évènement $(Z \leq z)$ en terme d'évènement pour X.

Exemple

$Z = 2X + 3$
$F_Z(z) = P(Z \leq z) = P(2X + 3 \leq z) = P(2X \leq z - 3) = P(X \leq \frac{z-3}{2}) = F_X(\frac{z-3}{2})$

Fonction d'une variable aléatoire

Exemple

$T = X^2$

$F_T(t) = P(X^2 \leq t) =$

$$\begin{cases} 0 \text{ si } t \leq 0 \\ P(-\sqrt{t} \leq X \leq \sqrt{t}) = F_X(\sqrt{t}) - F_X(-\sqrt{t}) \end{cases}$$

Exemple

$U = \frac{c}{X}$ où $c > 0$ et X à valeurs dans $]0, +\infty[$

Pour $u > 0$ on a $F_U(u) = P(U \leq u) = P(\frac{c}{X} \leq u) = P(\frac{c}{X} \leq u) =$
$P(X \geq \frac{c}{u}) = 1 - P(X < \frac{c}{u}) = 1 - F_X(\frac{c}{u})$.

Plan

Définitions

L'espérance mathématique correspond à la notion de moyenne pour une distribution empirique.

Exemple

Le temps de fabrication d'un produit connaît des variations aléatoires selon une loi supposée connue. L'espérance mathématique va indiquer quel est "en moyenne" le temps de fabrication du produit.

Définitions

Definition (Espérance)

On appelle *espérance mathématique* d'une variable aléatoire X, si elle existe, la valeur notée $E(X)$ telle que :

- $E(X) = \sum_i x_i p_X(x_i)$ dans le cas discrêt
- $E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$ dans le cas continu.

• Ex : X : valeur d'un dé à 6 faces équilibré :

$$E[X] = \sum_{i=1}^{6} x_i p_X(x_i) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \dots + 6 \times \frac{1}{6} = \frac{21}{6} = 3.5$$

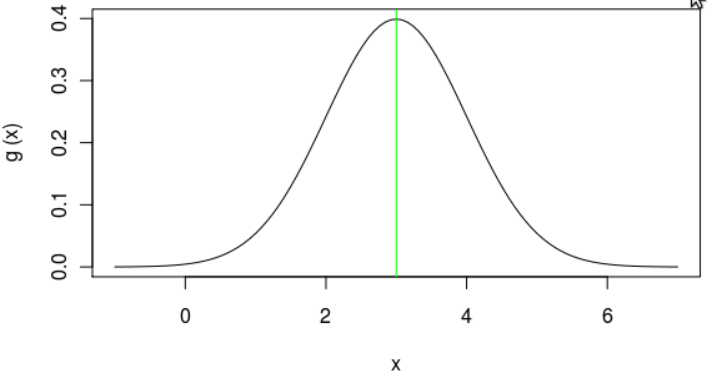
Définitions

Du point de vue du graphe de f_X (resp. p_X) l'espérance correspond au centre de gravité de la surface sous la courbe (resp. des bâtonnets représentant les probabilités des points).

En particulier, s'il y a un axe de symétrie, elle se situe au niveau de cet axe.

$E(X)$ s'appelle aussi la moyenne théorique de la loi.

Définitions



Espérance d'une fonction d'une variable aléatoire

$Y = g(X)$

$E(Y) = \sum_i g(x_i) p_X(x_i)$ dans le cas discrêt.

$E(Y) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$ dans le cas continu.

Ex: Le carré de la valeur d'un dé à 6 face équilibré.

$Y = X^2 \quad E(Y) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + \dots + 6^2 \times \frac{1}{6} = \frac{91}{6} \sim 15$

Linéarité de l'espérance, moments, variance

$\boxed{E(X+Y) = E(X) + E(Y) \qquad E(aX+b) = aE(X)+b.}$

Propriété

Soit X une variable aléatoire, et $Y = ag(X) + bh(X)$, où a et b sont deux constantes et g et h deux fonctions. Alors on a :

$$E(Y) = aE(g(X)) + bE(h(X)).$$

Exemple

$X \simeq \mathcal{N}(1,1)$ et $Y = 3X + 1$. Alors $E(Y) = 3 \times 1 + 1 = 4$.

Linéarité de l'espérance, moments, variance

Definition (Moment simple)

On appelle *moment simple d'ordre r* de la variable aléatoire X, où r est un entier > 0, la valeur : $\mu_r = E(X^r)$.

μ_1, aussi notée μ est la moyenne théorique de X.

Les puissances supérieures fournissent diverses caractéristiques de la forme de la distribution, surtour si on s'intéresse aux moments centrés.

Linéarité de l'espérance, moments, variance

Definition (Moments centrés)

On appelle *moment centré d'ordre r* de la variable aléatoire X, où r est un entier > 0, la valeur : $\mu'_r = E((X-\mu)^r)$.

Remarque : $\mu'_1 = 0$, d'où la notion de centrage.

Linéarité de l'espérance, moments, variance

Definition (Variance)

On appelle *variance* de la variable aléatoire X la valeur :

$$V(X) = E((X-\mu)^2) = \sigma^2 \ \ (= \mu_2').$$

Definition (Ecart-type)

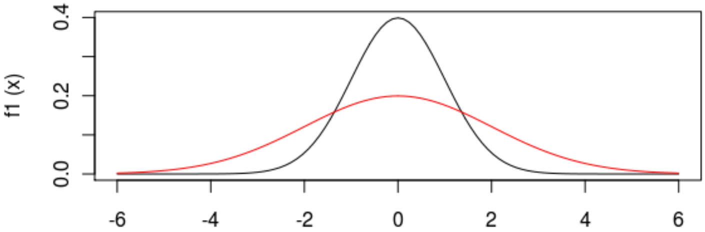
On appelle *écart-type* de la variable aléatoire X la valeur $\sigma = \sqrt{V(X)}$.

Linéarité de l'espérance, moments, variance

La variance représente la dispersion de la distribution.

Exemple

La loi normale centrée de variance 1 est plus "resserrée" que la loi normale de variance 4



Linéarité de l'espérance, moments, variance

Propriété (Formule de décentrage de la variance)

$$\mu'_2 = E(X^2) - \mu^2.$$

Propriété

$$V(aX + b) = a^2 V(X).$$

Tirage aléatoire dans une population finie

On considère une population de N individus sur lesquels s'observe un certain caractère quantitatif $\mathcal{X}$ (par exemple l'âge arrondi en années).

Supposons qu'il y ait r valeurs distinctes atteintes par $\mathcal{X}$ sur cette population (avec $2 \leq r \leq N$), notées $x_1, \ldots, x_r$.

Les valeurs $x_1, \ldots, x_r$ s'observent avec des fréquences relatives (fréquences divisées par N) $p_1, \ldots, p_r$. La moyenne observée dans la population est $\sum_{i=1}^{r} x_i p_i$.

Tirage aléatoire dans une population finie

Si on considère maintenant la variable aléatoire X : "La valeur d'un individu tiré au hasard dans cette population" (chaque individu a une probabilité $1/N$ d'être tiré).

Alors $P(X = x_i) = p_i$.

Il y a identité entre la distribution empirique $\mathcal{X}$ et la distribution de la variable aléatoire discrête X.

Plan

Couples de v.a.

On considère dans la suite un couple de variables aléatoires (X, Y) à valeurs dans $\mathbb{R}^2$.

Les évènements considérés sont des parties de $\mathbb{R}^2$.

On suppose que les deux variables aléatoires sont de même nature (discrètes toutes les deux ou continues toutes les deux).

Couples de v.a.

Definition

Soit (X, Y) un couple de v.a., on appelle *fonction de répartition conjointe* de (X, Y), notée $F_{X,Y}$, la fonction définie sur $\mathbb{R}^2$ par : $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$.

Definition (Cas discrêt)

Soit (X, Y) un couple de v.a. discrêtes pouvant prendre les couples de valeurs $(x_i, y_j), i = 1, 2, \ldots, j = 1, 2 \ldots\}$. On appelle *fonction de probabilité conjointe* la fonction, notée $p_{X,Y}$ qui donne les probabilités associées associées à ces couples de valeurs : $\forall i \forall j \ : \ p_{X,Y}(x_i, y_j) = P(X = x_i, Y = y_j)$.

Couples de v.a.

Definition (Cas continu)

Soit (X, Y) un couple de v.a. continues, on appelle *fonction de densité de probabilité conjointe* la fonction non négative sur $\mathbb{R}^2$, notée $f_{X,Y}$ telle que : $F_{X,Y}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v) du dv$.

Couples de v.a.

Propriété (Lois marginales)

$F_X(x) = P(X \leq x) = P(X \leq x, Y \in \mathbb{R}) = F_{X,Y}(x, +\infty)$.
De même $F_Y(y) = F_{X,Y}(+\infty, y)$.

Cas discrêt : $p_X(x_i) = \sum_j p_{X,Y}(x_i, y_j)$.

Cas continu : $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dy$.

Indépendance de deux v.a.

Definition

Deux variables aléatoires X et Y sont dites *indépendantes* si, étant donnés deux évènements quelconques $(X \in A)$ et $(Y \in B)$, on a :

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Exemple

Deux lancers de dés successifs sont indépendants.

Indépendance de deux v.a.

Proposition

X et Y sont deux variables aléatoires indépendantes si et seulement si $\forall(x,y) \in \mathbb{R}^2$, on a $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.

Cas discrêt : X et Y discrêtes sont indépendantes si et seulement si $\forall i \forall j \ : \ p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j)$.

Cas continu : X et Y continues sont indépendantes si et seulement si $\forall(x,y) \in \mathbb{R}^2 \ : \ f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

<u>Corollaire</u> : Si X et Y sont indépendantes, alors :

- $\mathbb{E}(XY) = \mathbb{E}(X) \times \mathbb{E}(Y)$
- $var(X+Y) = var(X) + var(Y)$.

<u>Ex</u> : Deux paquets de copie : moyenne : 10 variance : 1

On tire une copie dans chaque paquet : X : note paquet 1, Y : note paquet 2

Indépendance ⇝ $\mathbb{E}(X+Y) = 10+10 = 20$ $\quad var(X+Y) = 1+1 = 2$

On tire une copie dans le paquet 1 et on double la note : $X+X$

Pas indépendance ⇝ $\mathbb{E}(X+X) = 20$ $\quad var(X+X) = var(2X) = 4\,var(X) = 4$

X

Indépendance de deux v.a.

Proposition

Si X et Y sont indépendantes, alors pour toutes fonctions g et h, les variables aléatoires $g(X)$ et $h(Y)$ sont également indépendantes.

Espérance mathématique, covariance, corrélation

Cas discrêt : $E(g(X,Y)) = \sum_i \sum_j g(x_i, y_j) p_{X,Y}(x_i, y_j)$.

Cas continu : $E(g(X,Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f_{X,Y}(x,y) dxdy$.

Proposition (Linéarité de l'espérance mathématique)

$E(aX + bY) = aE(X) + bE(Y)$.

Espérance mathématique, covariance, corrélation

Definition (Moment simple croisé d'ordre (p, q))

$$E(X^p Y^q)$$

Definition (Moment centré croisé d'ordre (p, q))

$$E([X - E(X)]^p [Y - E(Y)]^q)$$

Espérance mathématique, covariance, corrélation

Definition (Covariance)

On appelle covariance de X et de Y, notée $cov(X, Y)$, la valeur :

$$cov(X, Y) = E([X - E(X)][Y - E(Y)]).$$

Remarque : $cov(X, Y) = cov(Y, X)$.

Espérance mathématique, covariance, corrélation

Proposition (Formule de décentrage de la covariance)

$cov(X, Y) = E(XY) - E(X)E(Y)$.

Proposition

Si X et Y sont deux v.a. indépendantes, alors $cov(X, Y) = 0$.

Espérance mathématique, covariance, corrélation

Remarque : Deux variables aléatoires peuvent avoir une covariance nulle sans pour autant être indépendantes.

Exemple

X et Y discrètes, chacune pouvant prendre les valeurs 0,1, ou 2. Les probabilités conjointes sont données dans le tableau ci-dessous.

	0	1	2	Y
0	0	4/9	0	4/9
1	2/9	0	2/9	4/9
2	0	1/9	0	1/9
Y	2/9	5/9	2/9	1

Espérance mathématique, covariance, corrélation

Exemple

On a $E(X) = \frac{4}{9} + 2 \times \frac{1}{9} = \frac{2}{3}$ et $E(Y) = \frac{5}{9} + 2 \times \frac{2}{9} = 1$.
De même $E(XY) = 1 \times 2 \times \frac{2}{9} + 2 \times 1 \times 19 = \frac{2}{3}$.
D'où $cov(XY) = E(XY) - E(X)E(Y) = 0$.
Or X et Y ne sont pas indépendantes puisque, par exemple,
$P(X = 0, Y = 0)$ est nulle alors que
$P(X = 0)P(Y = 0) = \frac{2}{9} \times \frac{4}{9} \neq 0$.

Espérance mathématique, covariance, corrélation

Propriété (Propriétés de la covariance)

1. $cov(aX + b, cY + d) = ac.cov(X, Y)$
2. $cov(X + Y, Z) = cov(X, Z) + cov(Y, Z)$

Espérance mathématique, covariance, corrélation

Definition (Corrélation linéaire)

On appelle *coefficient de corrélation linéaire* de X et Y, notée $corr(X, Y)$, la valeur :

$$corr(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y},$$

où σ_X et σ_Y sont respectivement les écart-types de X et Y.

Espérance mathématique, covariance, corrélation

Propriété

1. $corr(X, Y) = corr(Y, X)$
2. $corr(aX + b, cY + d) = corr(X, Y)$

Proposition

Pour tout couple de v.a. (X, Y), on a : $-1 \leq corr(X, Y) \leq 1$.

Remarque : La corrélation s'annule si et seulement si la covariance s'annule, et donc une corrélation nulle **n'implique pas l'indépendance.**

Somme de deux v.a.

On sait déjà que $E(X+Y) = E(X) + E(Y)$.

Proposition

$$V(X+Y) = V(X) + V(Y) + 2cov(X, Y)$$

Remarque : Si X et Y sont indépendantes, alors $V(X+Y) = V(X) + V(Y)$.

Les n-uplets de v.a., somme de n v.a.

On généralise les notions précédentes à n variables aléatoires : $(X_1, \dots, X_n)$ à valeurs dans $\mathbb{R}^n$.

Definition (Fonction de répartition conjointe)

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

La notion d'indépendance se généralise.

Les n-uplets de v.a., somme de n v.a.

On se limite maintenant au cas où les v.a. sont indépendantes et identiquement distribuées : v.a.i.i.d.

Par exemple, le n-uplet représente n observations successives d'un même phénomène aléatoire ($\Rightarrow$ échantillons aléatoires).

Proposition (Proposition fondamentale)

Soit $X_1, \ldots, X_n$ une suite de v.a.i.i.d de loi de probabilité ayant pour moyenne théorique μ et de variance théorique σ^2. On a, pour la somme :

$$S_n = X_1 + \ldots + X_n \quad E(S_n) = n\mu \quad V(S_n) = n\sigma^2$$

Les n-uplets de v.a., somme de n v.a.

Proposition (Proposition fondamentale, suite)

Soit $X_1, \ldots, X_n$ une suite de v.a.i.i.d de loi de probabilité ayant pour moyenne théorique μ et de variance théorique σ^2. On a, pour la moyenne :

$$\overline{X_n} = \frac{S_n}{n} \quad E(\overline{X_n}) = \mu \quad V(\overline{X_n}) = \frac{\sigma^2}{n}.$$

Sondage aléatoire dans une population et v.a.i.i.d.

On considère une population de N individus, et n tirages successifs parmi cette population.

Sondage aléatoire simple : On sélectionne un premier individu avec équiprobabilité parmi les N individus, puis un deuxième individu avec équiprobabilité sur les $N-1$ individus restants, et ainsi de suite.

Pour une variable quantitative d'interêt sur les individus, on note X_1 l'observation aléatoire du premier tirage, X_2 celle du deuxième tirage, etc., X_n celle du $n$ième tirage.

Ce sondage **n'est pas** une situation de v.a.i.i.d.

Sondage aléatoire dans une population et v.a.i.i.d.

Une façon de contourner le problème = tirage avec remise, qui lui donne des v.a.i.i.d.

Mais jamais appliqué en pratique...

Si le *taux de sondage* $\frac{n}{N}$ est faible, le sondage sans remise est très proche du sondage avec remise.

Exemple

Un échantillon de taille 1000 dans la population française des individus âgés de 15 ans et plus.

Sondage aléatoire dans une population et v.a.i.i.d.

Cela justifie qu'en pratique, on utilise les résultats de la théorie statistique classique développée dans ce cours, dans les situations de sondage.

Si $\frac{n}{N} < 0.1$, on a des approximations correctes, d'autant plus que d'autres approximations du même ordre de grandeur sont souvent inévitables dans la théorie des sondages elle-même.

Remarque : n tirages successifs sans remise sont équivalents à un tirage simultané de n individus parmi N.