

Suites de variables aléatoires

Probabilités et statistique pour l'informatique et la gestion

Licence 3 MIAGE

Université de Bordeaux - Collège Sciences et Technologies

Automne 2025

A. Lourme

alexandre.lourme@u-bordeaux.fr

Outline

- Couples de variables discrètes
- Théorèmes limites

Probabiliser la dépendance

Une population comporte 25% de sujets riches et 40% de jeunes. On choisit un sujet au hasard ; avec quelle probabilité cette personne est-elle riche et jeune ?

X vaut 1 si le sujet choisit est riche ($\mathbb{P}(X = 1) = 0,25$) ; 0 sinon.

Y vaut 1 si le sujet choisit est jeune ($\mathbb{P}(Y = 1) = 0,40$) ; 0 sinon.

Deux hypothèses (parmi d'autres) pour les probabilités conjointes.

		Y		
		0	1	
X	0	0,45	0,30	0,75
	1	0,15	0,10	0,25
		0,60	0,40	

$$\mathbb{P}(X = 1, Y = 1) = 0,10$$

		Y		
		0	1	
X	0	0,40	0,35	0,75
	1	0,20	0,05	0,25
		0,60	0,40	

$$\mathbb{P}(X = 1, Y = 1) = 0,05$$

Des lois conjointes différentes peuvent conduire aux mêmes lois marginales.

NB. Connaître les distributions marginales (distributions de X et de Y) ne permet pas de déterminer la loi conjointe (distribution du couple (X, Y)) ; connaître la distribution conjointe suffit à déterminer les distributions marginales.

Loi conjointe, lois marginales, conditionnelles, un exemple

(X, Y) est un couple de variables binaires.

La loi conjointe (loi du couple)

		Y	
		0	1
X	0	0,40	0,35
	1	0,20	0,05

La loi conjointe contient toute l'information sur les lois marginales et sur les lois conditionnelles.

Les lois marginales.

X	0	1
$\mathbb{P}$	0,75	0,25

La loi de X

Y	0	1
$\mathbb{P}$	0,60	0,40

La loi de Y

Les lois conditionnelles.

		Y	
		0	1
X	0	40/75	35/75
	1	20/25	5/25

Les lois de Y conditionnelles à X

		Y	
		0	1
X	0	40/60	35/40
	1	20/60	5/40

Les lois de X conditionnelles à Y

Loi conjointe, lois marginales, conditionnelles, généralités

(X, Y) est un couple de variables discrètes sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

Loi conjointe.

$$\mathcal{L}(X, Y) : \{((x, y), \mathbb{P}(X = x, Y = y)); (x, y) \in X(\Omega) \times Y(\Omega)\}$$

Lois marginales.

$$\mathcal{L}(X) : \{(x, \mathbb{P}(X = x)); x \in X(\Omega)\} \text{ et } \mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y)$$

$$\mathcal{L}(Y) : \{(y, \mathbb{P}(Y = y)); y \in Y(\Omega)\} \text{ et } \mathbb{P}(Y = y) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y)$$

Lois conditionnelles.

$$\mathcal{L}(X|Y = y) : \{(x, \mathbb{P}_{[Y=y]}(X = x)); x \in X(\Omega)\}$$

$$\mathcal{L}(Y|X = x) : \{(y, \mathbb{P}_{[X=x]}(Y = y)); y \in Y(\Omega)\}$$

avec :

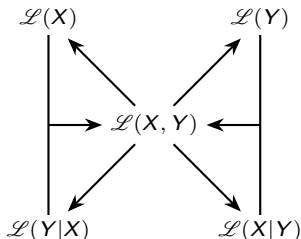
$$\mathbb{P}_{[Y=y]}(X = x) = \mathbb{P}(X = x, Y = y) / \mathbb{P}(Y = y)$$

$$\mathbb{P}_{[X=x]}(Y = y) = \mathbb{P}(X = x, Y = y) / \mathbb{P}(X = x)$$

D'une loi à l'autre (figure adjacente).

$\mathcal{L}(X, Y)$ permet de déterminer $\mathcal{L}(X)$

$\mathcal{L}(X|Y)$ et $\mathcal{L}(Y)$ permettent de d'obtenir $\mathcal{L}(X, Y)$, etc.



Indépendance

Indépendance de deux variables. Deux variables discrètes X et Y définies sur le même espace probabilisé sont indépendantes si :

$$\forall (x, y) \in X(\Omega) \times Y(\Omega), \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$$

Rq. Quand deux variables sont indépendantes, les distributions marginales suffisent à établir la loi conjointe.

Indépendance de n variables. n variables discrètes : $X_1, \dots, X_n$ définies sur le même espace probabilisé sont mutuellement indépendantes si :

$$\forall (x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega), \mathbb{P}(\cap_{i=1}^n [X_i = x_i]) = \prod_{i=1}^n \mathbb{P}(X_i = x_i)$$

Suite de variables indépendantes. $(X_n)_{n \in \mathbb{N}}$ est une suite de variables discrètes indépendantes si pour tout $n \in \mathbb{N}$ les variables $X_0, \dots, X_n$ sont indépendantes.

Exercices

Exercice 1. détermination d'une loi conjointe

Chaque jour, quand Bob arrive au bureau, sa machine a 20% de chances d'être en panne (et 80% de chances de fonctionner). X désigne le premier jour de fonctionnement, Y le premier jour de panne. Quelle est la loi du couple (X, Y) ?

Exercice 2. indépendance de deux variables de Bernoulli

Déterminer la loi du couple (X, Y) où X et Y sont deux variables de Bernoulli indépendantes de paramètre $p = 1/4$.

Exercice 3. indépendance mutuelle

En Front Office un trader exerce : un an dans 50% des cas, deux ans dans 30% des cas, trois ans dans 20% des cas. Une banque dispose de trois traders aux durées d'exercice indépendantes et pouvant traiter n'importe laquelle de ses opérations. Déterminer la durée de vie de cette banque en salle de marché si elle ne peut pas renouveler les traders défaillants.

Covariance et corrélation linéaire

Covariance. X et Y sont deux variables aléatoires discrètes d'espérance finie. La covariance du couple (X, Y) est : $cov(X, Y) = E([X - E(X)] \times [Y - E(Y)])$.

La covariance mesure la dépendance de X et Y ; son ordre de grandeur dépend des unités de mesure des deux variables.

Corrélation. X et Y sont deux variables aléatoires discrètes de variance finie et non nulle. Le coefficient de corrélation linéaire du couple (X, Y) est :

$$cor(X, Y) = cov(X, Y) / \sqrt{V(X) \times V(Y)}.$$

Le coefficient de corrélation linéaire mesure aussi la dépendance de X et Y ; mais sa valeur ne dépend pas des unités de mesure des deux variables.

Propriétés. X, Y, Z sont trois variables discrètes de variance finie ; α est un réel.

(i) $cov(X, Y) = cov(Y, X)$

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

(iii) $cov(\alpha X + Y, Z) = \alpha cov(X, Z) + cov(Y, Z)$

(iv) $V(\alpha X + Y) = \alpha^2 V(X) + V(Y) + 2\alpha cov(X, Y)$

(v) $cor(X, Y)$ est un réel compris entre -1 et 1

(vi) Si X et Y sont indépendantes : (a) $E(f(X) \times g(Y)) = E(f(X)) \times E(g(Y))$ pour toutes fonctions f et g continues (b) $cov(X, Y) = 0$ (c) $V(X + Y) = V(X) + V(Y)$

Exercices

Exercice 4. (X, Y) est un couple de variables uniformément réparti sur $\{(1, 0), (-1, 0), (0, 1), (0, -1)\}$.

(a) Calculer la covariance de (X, Y) . (b) X et Y sont-elles indépendantes ?

Exercice 5. Dix signaux envoyés dans un circuit électronique doivent traverser un hub à trois sorties. Les signaux passent indépendamment les uns des autres par l'une des trois sorties. On note X_i le nombre de signaux passés par la sortie i .

- (a) Quelle est la loi de X_1 ?
(b) Déterminer les variances de X_1 , X_2 , $X_1 + X_2$.
(c) Déterminer la covariance de (X_1, X_2) .

Exercice 6. S'il travaille avec Jean un jour, le lendemain Marc reste avec Jean dans 20% des cas ou travaille avec Paul sinon ; si Marc travaille avec Paul un jour, le lendemain il reste avec Paul dans 30% des cas ou travaille avec Jean sinon. Au jour 1, quand il arrive au bureau, Marc a 40% de chances de travailler avec Paul. La variable X_i vaut 1 si Marc travaille avec Paul au jour i ; 0 sinon.

- (a) Déterminer la loi du couple (X_1, X_2) .
(b) Déterminer $cov(X_1, X_2)$ et $cor(X_1, X_2)$.
(c) Simuler sous R mille réalisations de (X_1, X_2) .
(d) Estimer $cor(X_1, X_2)$ empiriquement grâce aux mille réalisations de (X_1, X_2) .

Outline

- Couples de variables discrètes
- Théorèmes limites

Théorème central limit (TCL)

Théorème Central Limit. (X_n) est une suite de variables aléatoires indépendantes de même loi, de variance finie σ^2 et d'espérance μ . On note $\bar{X}_n = \sum_{k=1}^n X_k/n$.

- *version 1.* $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ converge en loi¹ vers une variable normale centrée réduite.
- *version 2.* si n est grand², $\bar{X}_n$ est approximativement distribuée selon $\mathcal{N}(\mu, \sigma/\sqrt{n})$.

Exemple d'utilisation.

20% des employés d'une ESN codent en Python. Avec quelle probabilité y-a-t-il de 17% à 23% de codeurs Python parmi trente employés choisis au hasard ?

$X_i = 1$ si l'employé i code en Python ; 0 sinon. $X_1, \dots, X_{30}$ sont des variables indépendantes et distribuées selon $\mathcal{B}(0, 2)$; leur espérance est $\mu = 0,2$ et leur variance $\sigma^2 = 0,16$.

D'après TCL, $\bar{X}_{30}$ est approximativement distribuée selon $\mathcal{N}(0,2; 0,4/\sqrt{30})$.

$\mathbb{P}(0,17 \leq \bar{X}_{30} \leq 0,23) = \Phi((0,23 - 0,2)/(0,4/\sqrt{30})) - \Phi((0,17 - 0,2)/(0,4/\sqrt{30}))$
 $\approx 0,32$. Ainsi, 32% des échantillons de trente employés comporteront de 17% à 23% de codeurs Python.

¹ une suite de variables aléatoires converge en loi vers $\mathcal{L}$ si la suite de leurs fonctions de répartition converge en tout point vers la fonction de répartition de $\mathcal{L}$

² on considère généralement n grand quand $n \geq 30$

Exercices

Exercice 6.

Un serveur renvoie cinq anomalies par semaines en moyenne et la variance du nombre d'anomalies est 9. Avec quelle probabilité observera-t-on de quatre à six anomalies par semaine en moyenne sur une année ?

Exercice 7.

82% des Suisses aiment le chocolat. Avec quelle probabilité un échantillon de quarante Suisses comportera-t-il au moins 80% d'amateurs de chocolat ?

Exercice 8.

A la station Gaîté, on attend le métro trois minutes en moyenne et la variance des temps d'attente vaut 0,8. Avec quelle probabilité le temps d'attente moyen calculé sur cinquante trajets est-il inférieur à 3,1 minutes ?

Loi Faible des Grands Nombres

Loi Faible des Grands Nombres (LfGN).

(X_n) est une suite de variables aléatoires indépendantes de variance finie σ^2 , d'espérance μ et $\bar{X}_n = \sum_{k=1}^n X_k/n$; alors :

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| > \epsilon) = 0$$

Ainsi, quand on fixe un intervalle centré sur une moyenne de population (μ), la moyenne d'échantillon ($\bar{X}_n$) a peu de chances de tomber en dehors de cet intervalle quand la taille d'échantillon (n) est grande.

Exemple d'utilisation.

Pour connaître l'espérance μ de la masse d'un grain de riz, Sam calcule la masse moyenne d'un grain sur un échantillon de taille n . Pourquoi doit-il choisir n aussi grand que possible ?

On pèse une infinité de grains de riz : X_i la masse du grain i .

(X_n) est une suite de variables indépendantes d'espérance μ , de variance σ^2 .

D'après LfGN, la probabilité que la masse moyenne de n grains ($\bar{X}_n$) s'écarte de μ de plus de $\epsilon = 0,01$ peut être rendue définitivement inférieure à 0,001 (par exemple) si n est assez grand.

Illustration de LfGN

Pour $n \in \{1000, 10000, 100000\}$:

- simuler sous R, cent échantillons de taille n issus de $\mathcal{E}(0, 1)$
- calculer la moyenne empirique $\bar{x}_n$ de chacun des échantillons
- pour quelle proportion α des échantillons a-t-on : $|\bar{x}_n - 10| > 1/100$?

```
n = 100000 # taille des échantillons
bx = NULL # liste contenant les moyennes empiriques
for (i in 1:100){
  sample = rexp(n,0.1) # un échantillon de taille n issu de la loi exponentielle
  bx[i] = mean(sample)} # moyenne empirique de l'échantillon simulé
sum(abs(bx-10)>1/10)/100 # proportion des échantillon tels que |bx[i]-10|>1/10
```

Pour $n = 1000$; 10000 ; 100000 , on observe : $\alpha \approx 0,68$; $0,32$; $0,01$. Ainsi, plus la taille d'échantillon est grande, plus la probabilité est petite d'observer un échantillon dont la moyenne empirique s'écarte de 10 de plus de $1/10$.