Variable aléatoire discrète

Abdallah Ben Abdallah

Université de Sfax
Institut Supérieur d'Informatique et de Multimédia de Sfax

Octobre 2020

Plan du Chapitre

1. Généralités

Exemple 1. Un système composé de $n = 2$ composants est dit parallèle s'il fonctionne si l'un des composants fonctionne.

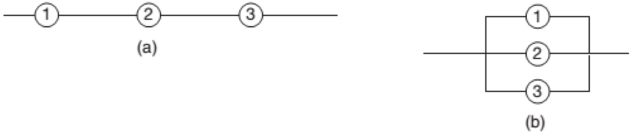


Figure: (a) système en série et (b) système parallèle

La probabilité qu'un composant fonctionne est 0,8. On note

$$X = \textit{Le nombre des composants qui fonctionnent}.$$

- X prend les valeurs 0, 1 et 2. X est dite **variable aléatoire discrète**.
- Vérifier que $\mathbb{P}(X = 0) = 0.04, \mathbb{P}(X = 1) = 0.32$ et $\mathbb{P}(X = 2) = 0.64$.

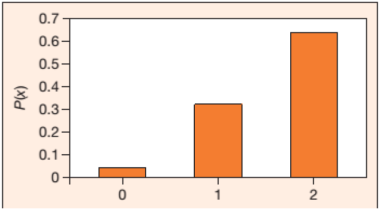


Figure: La densité de probabilité de X

On considère une expérience aléatoire d'univers Ω et de mesure de probabilité $\mathbb{P}$.

Définition 1.

Une **variable aléatoire discrète** X sur l'univers Ω est une application :

$$X : \Omega \longrightarrow S = \{x_i,\ i \in I\}$$

où I est une partie non vide (fini ou infini) de $\mathbb{N}$.

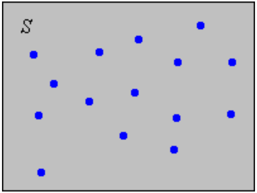


Figure: S est un **ensemble discret**

Définition 2.

La **densité** (ou la **loi**) **de probabilité** d'une variable aléatoire discrète $X : \Omega \to S = \{x_i \; i \in I\}$ est une application $p(.)$ définie sur S par:

$$\forall x_i \in S, \; p(x_i) = \mathbb{P}(X = x_i).$$

Exercice 1: On lance trois fois successives une pièce de monnaie. On marque 1 si on obtient pile et 0 si on obtient face. Soit X la variable aléatoire qui compte la somme obtenue. Déterminer la loi de probabilité de X.

Propriétés de la loi de probabilité

Soit $p(.)$ une loi de probabilité d'une variable aléatoire discrète $X : \Omega \to S = \{x_i,\ i \in I\}$.

(P1) Pour tout $x_i \in S$, on a $p(x_i) \in [0, 1]$.

(P2) On a :

$$\sum_{i \in I} p(x_i) = 1$$

Une fonction $p : S \to \mathbb{R}$ est une loi de probabilité d'une variable aléatoire discrète si et seulement si elle vérifie les deux conditions suivantes:

1. $p(x_i) \geq 0, \forall x_i \in S$.
2. $\sum_{i \in I} p(x_i) = 1$.

Définition 3.

La **fonction de répartition** d'une variable aléatoire discrète $X : \Omega \to S$, est définie par :

$$\forall x \in \mathbb{R},\ F(x) = \mathbb{P}(X \leq x)$$

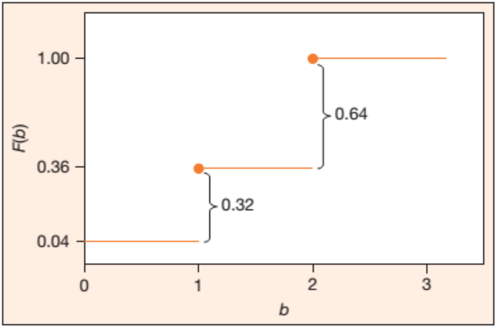


Figure: Fonction de répartition de la variable X de l'exemple 1.

Propriétés : Soit F une fonction de répartition d'une variable aléatoire discrète $X : \Omega \to S$. On a :

(P1) Pour tout $x \in \mathbb{R}$, on a $F(x) \in [0, 1]$.

(P2) F est croissante: si $x \leq y$, alors $F(x) \leq F(y)$.

(P3) F est continue à droite sur $\mathbb{R}$.

(P4) $\lim_{x \to +\infty} F(x) = 1$ et $\lim_{x \to -\infty} F(x) = 0$.

Exercice 2 : Quatre micro-puces doivent être placées dans un ordinateur. Deux des quatre puces sont choisis au hasard pour inspection avant l'assemblage de l'ordinateur. Soit X le nombre de puces défectueuses trouvées entre les deux puces inspectés. Trouvez la loi de probabilité de X dans les cas suivants

1. Deux de ces puces étaient défectueux.
2. Un des quatre puces était défectueux.
3. Aucun des puces était défectueux.

2. Espérance et Variance

On considère une expérience aléatoire d'univers Ω et de mesure de probabilité $\mathbb{P}$.

L'**espérance** d'une variable aléatoire discrète $X : \Omega \to S \subset \mathbb{R}$ est définie par

$$\mathbb{E}(X) = \sum_{k \in S} k\mathbb{P}(X = k).$$

Remarque : Si nous voyons la variable aléatoire X comme une distribution de masse (avec masse totale 1), alors l'**espérance** est le **centre de gravité** comme c'est défini en physique.

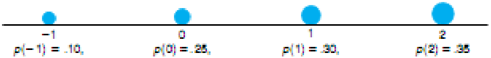


Figure: L'epérance = le centre de gravité=0.9

La **variance** d'une variable discrète $X : \Omega \longrightarrow S = \{x_i, i \in I\}$ est définie par

$$\mathbb{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \sum_{i \in I} (x_i - \mathbb{E}(X))^2 \mathbb{P}(X = x_i).$$

L'**écart type** de X est donné par :

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

L'écart type est une **mesure de la dispersion** de la variable autour de l'espérance.

Exercice 3 : Calculer l'écart type de la variable X définie dans l'exemple 1.

Exercice 4 : Le gérant d'un magasin dans une usine connait que la demande quotidienne X, (nombre d'utilisations par jour) d'un certain article est une variable aléatoire de loi de probabilité suivante:

Demande	0	1	2
Probabilité	0.1	0.5	0.4

Trouvez la demande quotidienne prévue sur l'article ainsi que la variance.

Inégalité de Tchebychev :

Soit $X : \Omega \longrightarrow S \subset \mathbb{R}$ une variable aléatoire discrète d'ésperance μ et d'écart type σ. Alors,

$$\mathbb{P}(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}.$$

L'inégalité de Tchebychev est équivalente à

$$\mathbb{P}(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

Par exemple pour $k = 2$,

$$\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \geq 3/4 = 0,75.$$

Exercice 5 : La production quotidienne de moteurs électriques à une certaine usine est de moyenne 120 moteurs et avec un écart type de 10 moteurs.

1. Quelle est le minimum des probabiltés pour que la production quotidienne soit entre 100 et 140 moteurs?
2. Trouvez l'intervalle le plus court des niveaux de production quotidiens avec au moins 90% de certitude.

Théorème 1 :

Soit $X : \Omega \longrightarrow S \subset \mathbb{R}$ une variable aléatoire discrète et $h : S \longrightarrow \mathbb{R}$ une application. Alors, pour tout réel positif k,

$$\mathbb{E}(h(X)) = \sum_{k \in S} h(k)\mathbb{P}(X = k).$$

Propriétés Soit X et Y deux variables aléatoires discrète et $a, b \in \mathbb{R}$.

1. $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.
2. $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.
3. $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$.
4. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.

Généralisation de propiété 4. $X_1, \ldots, X_n$ des variables aléatoires discrètes et $a_1, \ldots, a_n$ des réels :

$$\mathbb{E}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i).$$

Exercice 6 : Soit $X_1, X_2 \ldots, X_n$ des variables aléatoires discrètes à valeurs réelles de même loi de probabilité et d'espérance μ. On considère la variable aléatoire

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Montrer que $\mathbb{E}(\overline{X}) = \mu$.

Exercice 7 : Soit X une variable aléatoire d'espérance $\mathbb{E}(X) = 5$ et de variance $\mathbb{V}(X) = 4$. Déterminer $\mathbb{V}(3X - 2)$ et $\mathbb{E}(X^2)$.

Exercice 8 : Soit X une variable aléatoire d'espérance $\mathbb{E}(X) = 2$ et $\mathbb{E}(X(X-1)) = 4$. Déterminer $\mathbb{E}(X^2)$ et $\mathbb{V}(X)$.

3. Variables discrètes usuelles

2.1 Variable de Bernoulli:

On considère une expérience aléatoire dont les résultats sont :

- soit un **succès**,
- soit un **échec**.

Nous posons

- $X = 1$ si le résultat est un succès,
- $X = 0$ si c'est un échec.

La loi de probabilité de la variable X est donnée par

$$\mathbb{P}(X = 1) = p, \;\; \mathbb{P}(X = 0) = 1 - p,$$

où $p \in]0, 1[$ est la probabilité du succès.
La variable aléatoire X est appelé **variable de Bernoulli de paramètre p.**

Si X est une variable de Bernoulli de paramètre p, alors

$$\mathbb{E}(X) = p, \;\; \mathbb{V}(X) = p(1-p).$$

1.2 Variable binomiale:

On considère une expérience aléatoire qui consiste à répéter n expériences indépendantes de Bernoulli chacune avec un succès de probabilité p.

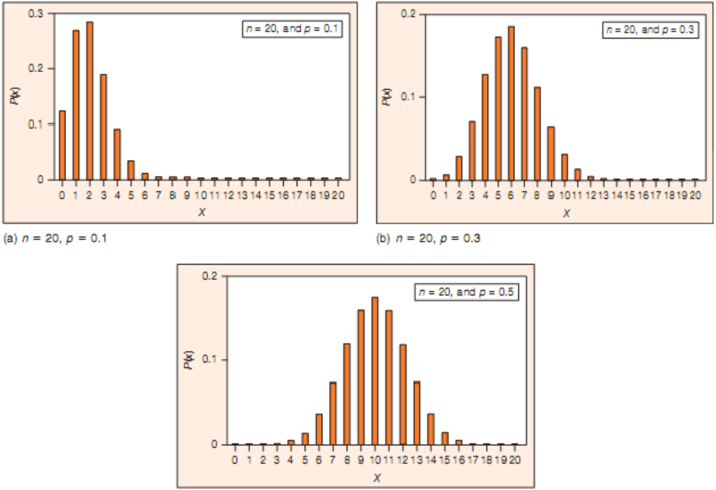
Soit X le **nombre de succès** qui se produisent dans les n expériences.

X est une variable **binomiale** de paramètres (n, p) et on note

$$X \rightsquigarrow \mathcal{B}(n, p).$$

La loi de probabilité de X est donnée par:

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

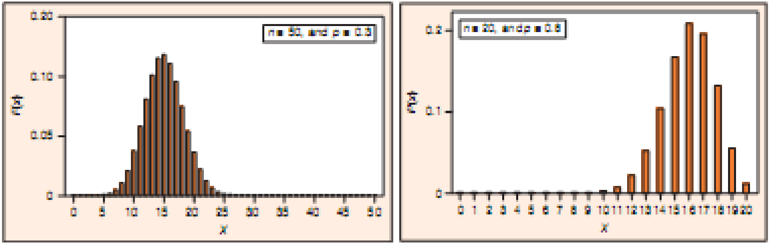


(a) $n = 20$, $p = 0.1$

(b) $n = 20$, $p = 0.3$

(c) $n = 20$, $p = 0.5$

Figure: La loi binomiale $n = 20$.



(d) $n = 50$, $p = 0.3$ (e) $n = 20$, $p = 0.8$

Figure: La loi binomiale $n = 20$ et $n = 50$.

Si X est une variable de binomiale de paramètre (n, p), alors

$$\mathbb{E}(X) = np, \quad \mathbb{V}(X) = np(1-p).$$

Exercice 9 : Un système de protection anti-missile est composé d'un ensemble à n radars fonctionnant de manière indépendante, chacun avec une probabilité de $0,9$ à détecter un missile pénètrant une zone spécifiée. (Tous les radars couvrent la même zone.)
Si un missile pénètre la zone, trouver la probabilité qu'il sera détecté si

1. $n = 2$.
2. $n = 4$.

Exercice 9 : Il est connu qu'un disque CD fabriqué par une certaine compagnie sera défectueux avec probabilité 0,01. La compagnie vend les disques CD dans des paquets de 10 et elle offre un remboursement garanti s'il y a plus qu'un disque défectueux.

1. Quelle est la proportion de paquets rendus?
2. Si quelqu'un achète trois paquets, quelle est la probabilité pour que exactement un d'eux sera rendu?

1.4 Variable de Poisson

Une variable aléatoire $X : \Omega \to S = \mathbb{N}$ est dite **variable de Poisson** de paramètre $\lambda > 0$ si sa loi de probabilité est donnée par:

$$\forall k \in \mathbb{N}\ , \ \ \mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

L'espérance et la variance sont données par

$$\mathbb{E}(X) = \lambda, \ \ \mathbb{V}(X) = \lambda.$$

Exercice 10 : Le nombre de fautes d'impression F sur une page Web suit la distribution de Poisson de paramètre $\lambda = 2,5$. Quelle est la probabilité pour qu'il y aura plus que 2 fautes sur la page Web?

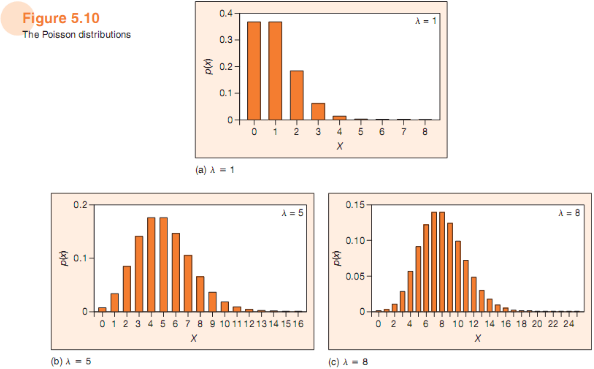


Figure: La loi Poisson $\lambda = 1, \lambda = 5$ et $\lambda = 8$.

La loi de Poisson de paramètre $\lambda = np$ approxime la loi Binomiale $\mathcal{B}(n, p)$ lorsque n est grand :

$$C_n^k p^k (1-p)^{n-k} \simeq e^{-\lambda} \frac{\lambda^k}{k!}.$$

1.5 Variable géométrique

Une variable aléatoire $X : \Omega \to S = \mathbb{N}^*$ est appelée **variable géométrique** de paramètre $p \in]0, 1[$ si sa loi de probabilité est donnée par:

$$\forall k = 1, 2, \dots , \quad \mathbb{P}(X = k) = p(1 - p)^{k-1}$$

L'espérance et la variance sont données par

$$\mathbb{E}(X) = \frac{1}{p}, \quad \mathbb{V}(X) = \frac{1 - p}{p^2}.$$

Exercice 9 : On lance successivement un dé cubique parfait numéroté de 1 à 6 et on s'arrête jusqu'à obtenir la première fois le numéro 6. On désigne par X le nombre de lancés.

1. Déterminer la loi de X.
2. Calculer la probabilité de l'événement A : "le dé est jeté au moins 3 fois".

3. Couple de variables aléatoires

Pour une expérience aléatoire donnée, nous sommes souvent intéressés à des **relations** entre deux ou plusieurs variables aléatoires.

Soit X_1 et X_2 deux variables aléatoires discrète.

La **loi du couple** de (X_1, X_2) est définie par la fonction $p(.,.)$:

$$p(x_1, x_2) = \mathbb{P}(X_1 = x_1, X_2 = x_2).$$

Les **lois marginales** de X_1 et X_2 sont définies par :

$$p(x_1) = \sum_{x_2} p(x_1, x_2) \quad \text{et} \quad p(x_2) = \sum_{x_1} p(x_1, x_2).$$

Définition 4.

On dit que X_1 et X_2 sont **indépendantes** si,

$$\mathbb{P}(X_1 = x_1, X_2 = x_2) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2).$$

Exercice 10 : On choisit aléatoirement trois batteries d'un stock de 12 batteries : 3 nouveaux, 4 utilisées, mais fonctionnent encore et 5 batteries défectueuses. Notons X le nombre de batteries neuves et Y le nombre de batteries utilisées, mais fonctionnent encore.

1. Déterminer la loi du couple (X, Y).
2. En déduire les lois marginales de X et Y.
3. Les variables X et Y sont elles indépendantes?

Théorème 3 : (Stabilité de la loi binomiale)

Soit X et Y deux variables aléatoires indépendantes de loi Binomiale $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$ respectivement. Alors, la variable aléatoire $X + Y$ suit la loi Binomiale $\mathcal{B}(n + m, p)$.

Théorème 4 : (Stabilité de la loi de Poisson)

Soit X et Y deux variables aléatoires indépendantes de loi de Poisson de paramètres λ et μ respectivement. Alors, la variable aléatoire $X + Y$ suit la loi de Poisson de paramètre $\lambda + \mu$.

Proposition :

Si X et Y sont indépendantes alors,

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

5. Covariance et Corrélation

On considère une expérience aléatoire d'univers Ω et mesure de probabilité $\mathbb{P}$.

Théorème 5 :

Soit X_1 et X_2 deux variables aléatoires discrètes de loi conjointe $p(x_1, x_2)$. Si $g(x_1, x_2)$ est une fonction quelconque, alors

$$\mathbb{E}(g(X_1, X_2)) = \sum_{x_1} \sum_{x_2} g(x_1, x_2) p(x_1, x_2).$$

La **covariance** de (X_1, X_2) est définie par

$$\text{Cov}(X_1, X_2) = \mathbb{E}\left[(X_1 - \mu_1)(X_2 - \mu_2)\right],$$

où

$$\mu_1 = \mathbb{E}(X_1), \mu_2 = \mathbb{E}(X_2).$$

Le **coefficient de corrélation** de (X_1, X_2) est définie par

$$r(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sigma(X_1)\sigma(X_2)}.$$

En pratique on utilise la formule suivante :

$$\text{Cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2).$$

Propriétés:

(P1) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ et $\text{Cov}(X, X) = \mathbb{V}(X)$.

(P2) $\text{Cov}(aX + bZ, Y) = a \text{ Cov}(X, Y) + b \text{ Cov}(Z, Y)$.

(P3) $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2 \text{ Cov}(X, Y)$.

(P4) Si $Y = aX + b$ alors

$$r(X, Y) = \begin{cases} 1 & \text{si} \quad a > 0 \\ -1 & \text{si} \quad a < 0 \end{cases}$$

6. Régression linéaire

On considère une expérience aléatoire d'univers Ω et mesure de probabilité $\mathbb{P}$.

Soit (X, Y) deux variables discrètes tel que le coefficient de corrélation $r(X, Y)$ proche de ± 1.

Ainsi, nous pouvons écrire

$$Y \simeq aX + b$$

avec a et b sont des paramètres à déterminer.

Ou bien,

$$Y = aX + b + \epsilon$$

avec ϵ est une variable aléatoire vérifiant

- $\mathbb{E}(\epsilon) = 0$,
- X et ϵ indépendant.

Théorème :

Si X et Y sont deux variables aléatoires vérifiant

$$Y = aX + b + \epsilon,$$

avec ϵ est une variable aléatoire tel que

- $\mathbb{E}(\epsilon) = 0$,
- X et ϵ indépendant.

Alors,

$$b = \frac{cov(X, Y)}{\mathbb{V}(X)}, \quad a = \mathbb{E}(Y) - b\mathbb{E}(X).$$

Maintenant, nous considérons n observations $(x_1, y_1), \cdots, (x_n, y_n)$ du couple des variables aléatoires (X, Y) tel que le coefficient de corrélation est proche de ± 1.

Alors, les points aurons la forme de dispersion suivante

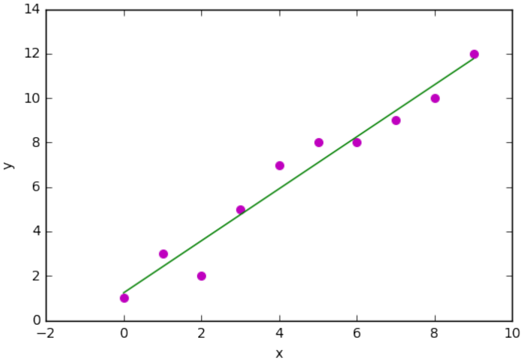


Figure: Dépendance linéaire entre X et Y.

Si le coefficient de corrélation $r(X, Y)$ est loin de ± 1, alors la relation entre X et Y est plus complexe, comme par exemple, une relation non linéaire :

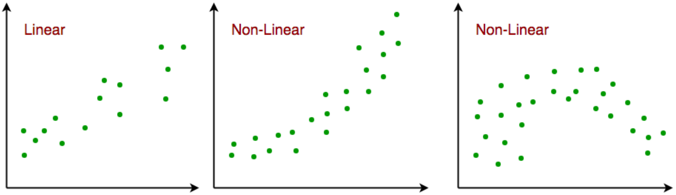


Figure: la dépendance entre X et Y peut être non linéaire.

Bibliographie

R. L. Scheaffer, M. S. Mulekar and J. T. McClave
Probability and Statistics for Engineers .
Brooks/Cole, Cengage Learning, (fifth edition) 2011.

Sheldon M. Ross
Introduction to Probability and Statistics for Engineers and Scientists.
Elsevier Academic Press, 2004.

Douglas C. Montgomery, George C. Runger
Applied Statistics and Probability for Engineers.
John Wiley & Sons, 2003.