Statistique Appliquée

Probabilités

Luc Deneire – deneire@unice.fr

Xidian University, Polytech Nice Sophia

Octobre 2020

Introduction

Informations

- Cours sur : http://jalon.unice.fr/public/pqg729/
- Questions par e-mail : deneire@unice.fr

La statistique pour modéliser *le monde*

Permet de répondre à des questions du type

- Quelle est l'efficacité d'une usine ?
- Quelle est l'influence d'un changement dans un process ?

EN

- Recueillant des données
- En en déduisant un modèle
- En appliquant les modifications à ce modèle

L'outil du modèle : les probabilités

Les probabilités sont un outil pour *apprivoiser* l'aléatoire

Notions de bases :

- Evénements
- Probabilités associées à un événement
- Espace des événements et calcul de probabilités
- Inférences

Modèle probabiliste

Pour modéliser :

- Un hasard “simple” (jet de pièces, naissances d’enfants, ...)
- Temps d’attente à un guichet
- Probabilité de réussir un examen
- Moyenne du nombre de pièces défectueuses ...

Plan du cours - Partie Probabilités

Expérience aléatoire

Une expérience aléatoire est une expérience dont l'issue est incertaine (on ne peut savoir avec certitude quelle sera le résultat de l'expérience).

Exemples : jet de dés – Météo

Issue – Issue élémentaire

Quand une expérience a été *effectuée*, elle a une *issue*. Dans un cas simple, on définit toutes les *issues élémentaires* (dés : $\{"1","2","3","4","5","6"\}$)

Univers

L'univers est l'ensemble des issues de cette expérence aléatoire. On le note Ω (dés : $\Omega = \{"1","2","3","4","5","6"\}$)

Événement

Un événement **A** est un ensemble d'issues élémentaires liées à l'expérience aléatoire et est donc un sous-ensemble de l'Univers Ω.

Exemple : **A** = "*Le dé est pair*" = $\{"2","4","6"\}$

Espace probabilisé

Espace d'événements – Tribu

L'espace d'événements est un ensemble qui contient "tous les événements d'intérêt", c'est-à-dire toutes les compositions de sous-ensembles de l'univers. Cet ensemble doit avoir une structure d'espace algébrique (σ-algèbre)
On le note $\mathcal{A}$.

Probabilité

La probabilité est un nombre, compris entre 0 et 1, associé à chaque événement (à chaque élément de la tribu). On notera la probabilité de l'événement A : $\mathbb{P}(A)$

Espace Probabilisé

Le triplet $(\Omega, \mathcal{A}, \mathbb{P}())$ est appelé un espace probabilisé

Exemple d'espace probabilisé

Exemple

Lancer de deux dés

- Les issues élémentaires : $\omega_1 = (1,1)$, $\omega_2 = (3,4)$, $\omega_3 = (4,3)$,
- L'univers : $\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(6,6)\}$
- Un événement : $\mathbf{A} = \{$la somme est égale à 6$\}$
- Un événement : $\mathbf{B} = \{$le 1$^{\text{er}}$ est entre 3 et 5; le 2$^{\text{nd}}$ entre 2 et 4$\}$
- La tribu $\mathcal{A} = \{$ tous les sous-ensembles de $\Omega\}$.
- Les probabilités $\mathbb{P}$: si on a des dés non pipés et que les lancer sont indépendants, $\mathbb{P}$ est caractérisé par $\mathbb{P}(\omega_i) = 1/36$.

$\triangleleft$

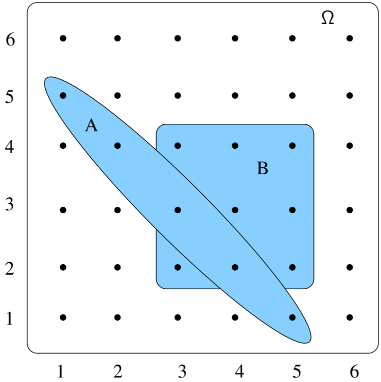
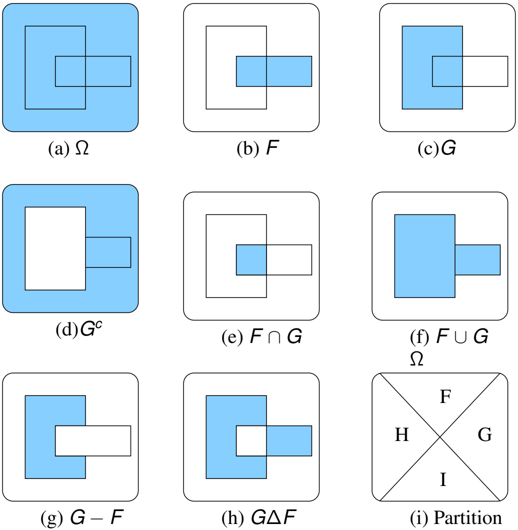


Figure: Exemple: lancer deux dés

exemples d'univers

- $\Omega = \emptyset$
- $\Omega = \{0, 1\}$; $\Omega = \{''\text{pile}'', ''\text{face}''\}$
- t-uples de longueur k ($k = 2$: $\Omega = \{\{0, 0\}; \{0, 1\}; \{1, 1\}; \{1, 0\}\}$
- : télécommunications : séquences infinies de “0” et de “1”
- nombres réels r entre $-V$ et V : $\Omega = \{r : -V \leq r \leq V\}$

Opérations de base des ensembles



(a) Ω (b) F (c) G

(d) G^c (e) $F \cap G$ (f) $F \cup G$

(g) $G - F$ (h) $G \Delta F$ (i) Partition

- **a)** L'univers Ω comprend toutes les issues possibles d'une expérience (si on tire un nombre réel au hasard, $\Omega = \mathcal{R} = \{\omega : -\infty \leq \omega \leq \infty\}$).
- **b)** Un sous-ensemble de l'univers est un événement, p.ex. $F = \{\omega : -2 \leq \omega \leq 4\}$
- **c)** Le complément à F, noté F^c est défini comme étant l'ensemble des éléments n'appartenant pas à F

$$F^c = \{\omega : \omega \notin F\}$$

- **d)** L'intersection : $F \cap G = \{\omega : \omega \in F \text{ et } \omega \in G\}$
- **e)** L'union : $F \cup G = \{\omega : \omega \in F \text{ ou } \omega \in G\}$
- **f)** La différence : $G - F = \{\omega : \omega \in G \text{ et } \omega \notin F\} = G \cap F^c$
- **g)** La différence symétrique :

$$G \Delta F = \{\omega : \omega \in G \text{ ou exclusif } \omega \in F\} = (F \cup G) - (F \cap G)$$

- **h)** La partition :

$$\Omega = F \cup G \cup H \cup I \text{ et}$$

$$\forall X, Y \in \{F, G, H, I\} \text{ et } X \neq Y : X \cap Y = \emptyset$$

- **i)** Première loi de De Morgan : $(F \cap G)^c = (F^c \cup G^c)$
- **j)** Deuxième loi de De Morgan : $(F \cup G)^c = (F^c \cap G^c)$

Soit le triplet $(\Omega, \mathcal{A}, P)$,

1. Définir l'univers Ω.
2. Définir une tribu $\mathcal{A}$ (un exemple simple de tribu peut être $\mathcal{A}$ = tous les sous-ensembles de Ω).
3. Attribuer un nombre $\mathbb{P}(A) \in [0, 1]$ à un événement A.
   - Définition classique (Laplace)
     $$\mathbb{P}(A) = \frac{\text{nombre de cas équiprobables favorables}}{\text{nombre de cas équiprobables possibles}}$$
   - Définition intuitive (fréquence relative)
     $$\mathbb{P}(A) = \lim_{n\to\infty} \frac{N_n(A)}{n}$$
   - Définition axiomatique (Kolmogorov)

Axiome

$$\mathbb{P}(A) \geq 0 \quad \textit{pour chaque événement } A \in \mathcal{A}.$$

Axiome

$$\mathbb{P}(\Omega) = 1$$

Axiome

$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ *pour A et B disjoints, qui se généralise à :*
soit les événements $A_i, i = 1, 2, ..., n$ *disjoints, alors*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)\,.$$

Une généralisation plus forte encore est : soit les événements $A_i, i = 1, 2, ...$ *disjoints, alors*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)\,.$$

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
   dém.: $\mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) \overset{A \cap A^c = \emptyset}{=} \mathbb{P}(A) + \mathbb{P}(A^c) = 1$
2. $\mathbb{P}(\emptyset) = 0 = \mathbb{P}(\Omega^c)$
3. Toutes les probabilités sont comprises entre 0 et 1 (démonstration : utiliser les deuxième et troisième axiomes ainsi que la définition de complément).
4. **Partition** : si $\{A_i\}$ est une partition (finie ou infiniment dénombrable) de Ω, alors, $\mathbb{P}(B) = \sum_i \mathbb{P}(B \cap A_i)$, pour tout événement B.
5. Si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$
6. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
7. $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$
8. $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B) + \mathbb{P}(A^c \cap B^c \cap C)$

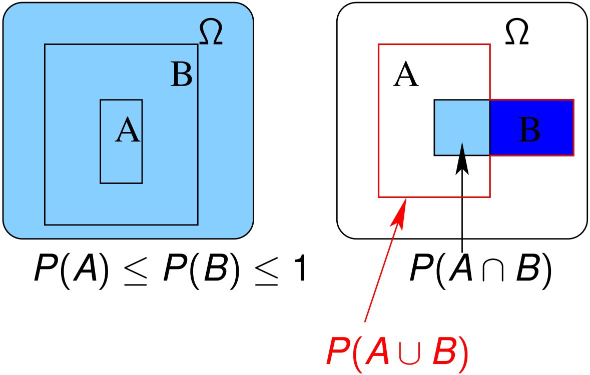


Figure: Exemples simples de la relation entre probabilités et ensembles.

Si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$

Exemple

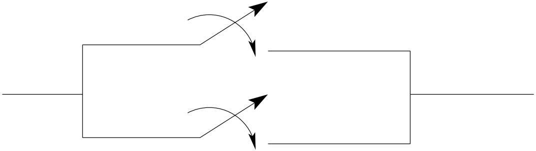
Interrupteurs en série

Soit $\mathbb{P}$(Les deux interrupteurs sont fermés) $= 1/4$ et $\mathbb{P}$(Un interrupeur est fermé) $= 1/2$. On définit les événements : $A = \{$interrupteur 1 et 2 fermés$\}$ et $B = \{$interrupteur 1 fermé$\}$ (avec $A \subset B$) . La probabilité que le circuit fonctionne est alors $\mathbb{P}(A) = 1/4 < \mathbb{P}(B)$=1/2.
$\triangleleft$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Exemple

Interrupteurs en parallèle



Soit $A = \{\text{interrupteur 1 fermé}\}$ et $B = \{\text{interrupteur 2 fermé}\}$, avec $\mathbb{P}(A) = \mathbb{P}(B) = 1/2$ et $\mathbb{P}(A \cap B) = 1/4$.
Alors, la probabilité que le circuit fonctionne est :

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4} \end{aligned} \tag{1}$$

$\triangleleft$

La probabilité qu'il pleuve sachant que je suis à Bruxelles est plus grande que si je suis à Nice ...

Probabilité conditionnelle

$\mathbb{P}(A|B) \in [0, 1]$: associée à A, *sachant que* l'événement B ($\mathbb{P}(B) \neq 0$) a été réalisé.

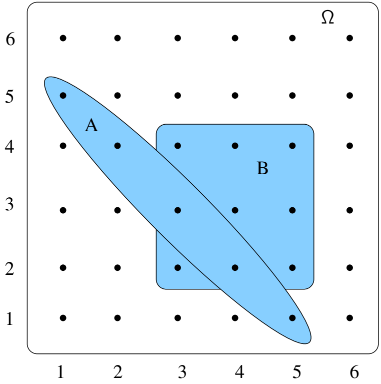


Figure: Exemple: lancer deux dés

Exemple

Lancer de deux dés

Dans l'exemple de la figure **??**, en supposant des dés non pipés, les événements A et B ont les probabilités indiquées ci-dessous.

Toutes les issues ω_i ($i = 1, \ldots, 36$) sont équiprobables ($\mathbb{P}(\omega_i) = \frac{1}{36}$).

- $\mathbb{P}(A) = \frac{5}{36}$
- $\mathbb{P}(B) = \frac{9}{36}$
- $\mathbb{P}(A|B) = \frac{2}{9} = \frac{2/36}{9/36}$
- $$\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}}$$

◁

Exemple

Le tabac et les jeunes

Soit l'exemple suivant :

Fréquences			
	Fumeurs	Non fumeurs	Total
Hommes	340	314	654
Femmes	289	384	673
Total	629	698	1327

Probabilités (approche fréquentiste)

- $\mathbb{P}(\text{Hommes}) = 654/1327 = 0.49$
- $\mathbb{P}(\text{Femmes}) = 673/1327 = 0.51$
- $\mathbb{P}(\text{Fumeurs}) = 629/1327 = 0.47$
- $\mathbb{P}(\text{Non fumeurs}) = 698/1327 = 0.53$
- $\mathbb{P}(\text{Fumeurs} \cap \text{Hommes}) = 340/1327 = 0.26$
- $\mathbb{P}(\text{Fumeurs}|\text{Hommes}) = 340/654 = 0.53 = \boxed{0.26/0.49}$
- $\mathbb{P}(\text{Fumeurs}|\text{Femmes}) = 289/673 = 0.43 = \boxed{0.22/0.51}$

$\lhd$

Exemple

Le tablac et les jeunes (suite)

Les fréquences relatives approximent les probabilités des événements “Hommes” (H), “Femmes” (nH), “Fumeurs” (F), et “non Fumeurs” (nF).

Fréquences relatives			
	Fumeurs	Non fumeurs	Total
Hommes	0.26	0.24	0.49
Femmes	0.22	0.29	0.51
Total	0.47	0.53	1

- **Espace Probabilisé**
  $\Omega = \{(H, F); (H, nF); (nH, F); (nH, nF)\}$, $\mathcal{A} = \Omega$,
  $\mathbb{P}() = (0.26, 0.24, 0.22, 0.29)$.
- **Probabilité conjointe** (“Homme” et “Fumeur”, ...)
- **Probabilité marginale**(“Homme” ; “Fumeur”, ...)

$\triangleleft$

Les probabilités conditionnelles satisfont Kolmogorov

- La probabilité conditionnelle satisfait les trois axiomes:
  1. $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \geq 0$ pour chaque événement $A \subseteq \Omega$
  2. $\mathbb{P}(A_1 \cup A_2|B) = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B)$ pour A_1 et A_2 disjoints
  3. $\mathbb{P}(B|B) = 1$
- Les propriétés générales restent valables , p.ex., $\mathbb{P}(A \cup C|B) \leq \mathbb{P}(A|B) + \mathbb{P}(C|B)$
- On peut remplacer 3. par 3'. $\mathbb{P}(B|B) = \frac{\mathbb{P}(B \cap B)}{\mathbb{P}(B)} = 1$
- $\mathbb{P}(A|B)$: *loi* de probabilité;
- **Approche séquentielle (appelée théorème de la multiplication ou chain rule)**:
  - $\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(B)\,\mathbb{P}(A|B)}$
  - $\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1)\,\mathbb{P}(A_3|A_1 \cap A_2) \ldots \mathbb{P}\left(A_n| \bigcap_{i=1}^{n-1} A_i\right)$

Fausse alarme

Soit un radar de détection d'un avion, on cherche la probabilité d'une fausse alarme

- $\mathbb{P}(\text{Avion présent}) = 0.05$
- $\mathbb{P}(\text{ Détection d'avion s'il est présent}) = 0.99$
- Fausse détection : $\mathbb{P}(\text{Détection d'avion si pas présent}) = 0.1$

Modélisation de l'univers "système radar"

- Avion: Présent / Absent
- Radar: Détection / Non détection
- En fonction des quatre issues possibles, l'univers vaut : $\Omega = \{(P, D), (A, D), (P, N), (A, N)\}$,

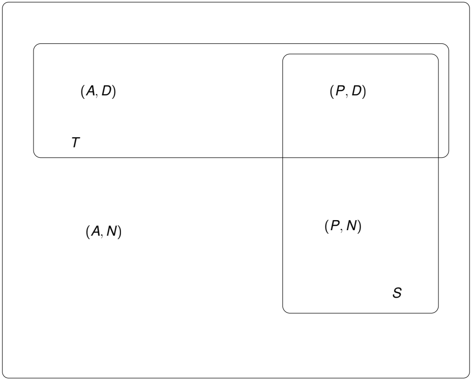


Figure: Détection de présence de l'avion : les 4 points de l'univers.

Les probabilités associées

- $S = \{\text{un avion est présent}\} = \{(P, D), (P, N)\}$
- $T = \{\text{le radar signale la présence d'un avion}\} = \{(P, D), (A, D)\}$
- $\mathbb{P}(S) = 0.05$ (présence d'un avion)
- $\mathbb{P}(T|S) = 0.99$ (détection si avion présent)
- $\mathbb{P}(T|S^c) = 0.10$ (fausse détection: «détection» si avion absent)

Les probabilités associées calculées grâce aux axiomes / propriétés

- Quelle est la probabilité d'une fausse alarme?

  $\mathbb{P}((A, D)) = \mathbb{P}(S^c \cap T) =$
  $\mathbb{P}(S^c)\,\mathbb{P}(T|S^c) = [1 - \mathbb{P}(S)]\mathbb{P}(T|S^c) = 0.95 \cdot 0.10 = 0.095$
- Quelle est la probabilité qu'un avion ne soit pas détecté?

  $\mathbb{P}((P, N)) = \mathbb{P}(S \cap T^c) =$
  $\mathbb{P}(S)\,\mathbb{P}(T^c|S) = \mathbb{P}(S)\,[1 - \mathbb{P}(T|S)] = 0.05 \cdot 0.01 = 0.0005$
- pour obtenir l'ensemble des probabilités, il faut calculer $\mathbb{P}((P, D)) = \mathbb{P}(S) - \mathbb{P}((P, N)) = 0.05 - 0.0005 = 0.0495$ et on en déduit directement $\mathbb{P}((A, N)) = 0.855$ en invoquant que la probabilité de l'univers vaut 1.

- soient $A_1, A_2, \ldots, A_n$ une partition de Ω
- soit un événement B, on peut écrire:
  $B = (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_n)$, où $B \cap A_1, B \cap A_2, \ldots, B \cap A_n$ sont disjoints ;
- $\Rightarrow \mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \ldots + \mathbb{P}(B \cap A_n)$
- et donc $\mathbb{P}(B) = \mathbb{P}(A_1)\,\mathbb{P}(B|A_1) + \mathbb{P}(A_2)\,\mathbb{P}(B|A_2) + \ldots + \mathbb{P}(A_n)\,\mathbb{P}(B|A_n)$.
- Ce résultat est connu comme étant le *théorème des probabilités totales* :

$$\boxed{\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\,\mathbb{P}(B|A_i)}$$

Bayes : vers l'inférence

Soient :

- *la probabilité a priori* : $\mathbb{P}(A) = 0.001$; (exemple : $A = \{\text{j'ai un accident}\}$)
- *l'événement connu* : B (exemple = $B = \{\text{je roule trop vite}\}$)
- *la probabilité a posteriori* : $\mathbb{P}(A|B) = 0.01$.

Donc “J'ai un accident” *Parce que* “je roule trop vite” ... Schématiquement :

- «Effet » $A \longrightarrow$ «Cause » B, $\mathbb{P}(A|B)\,,\mathbb{P}(B) \neq 0$
- À partir de $\mathbb{P}(A|B)$, calculer $\mathbb{P}(B|A)$ (cause $\longrightarrow$ effet)
- $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B|A) = \mathbb{P}(B)\,\mathbb{P}(A|B)$
- $$\boxed{\mathbb{P}(B|A) = \mathbb{P}(B)\,\frac{\mathbb{P}(A|B)}{\mathbb{P}(A)}}$$

S'il y a plusieurs causes possibles :

- $$\mathbb{P}(B_i|A) = \mathbb{P}(B_i)\,\frac{\mathbb{P}(A|B_i)}{\mathbb{P}(A)}$$
- $$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(B_i)\,\mathbb{P}(A|B_i)}{\cdots}$$

Exemple

Le tabac et les jeunes

Fréquences			
	Fumeurs	Non fumeurs	Total
Hommes	340	314	654
Femmes	289	384	673
Total	629	698	1327

- $\mathbb{P}(\text{Hommes}) = 654/1327 = 0.49$
- $\mathbb{P}(\text{Fumeurs}) = 629/1327 = 0.47$
- $\mathbb{P}(\text{Fumeurs}|\text{Hommes}) = 340/654 = 0.53 = 0.26/0.49$
- $\mathbb{P}(\text{Hommes}|\text{Fumeurs}) = 340/629 = 0.54 = 0.26/0.47$
  $\mathbb{P}(\text{Hommes}|\text{Fumeurs}) = 0.49 \cdot 0.53/0.47$
- $\mathbb{P}(\text{Fumeurs}|\text{Hommes}) > \mathbb{P}(\text{Fumeurs})$
- $\mathbb{P}(\text{Hommes}|\text{Fumeurs}) > \mathbb{P}(\text{Hommes})$

$\triangleleft$

Indépendance

Deux événements A et B sont **indépendants** si et seulement si $\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)}$.

On en déduit les propriétés suivantes :

- si $\mathbb{P}(B) \neq 0$, $\boxed{\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A)}$
- Soit deux événements indépendants A et B, conditionnés par C, $(\mathbb{P}(C) \neq 0)$:
  - $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\,\mathbb{P}(B|C)$
  - si $\mathbb{P}(B|C) \neq 0, \mathbb{P}(A|B \cap C) = \mathbb{P}(A|C)$
- Soit plusieurs événements indépendants $A_1, A_2, \ldots, A_n$:
  - $\mathbb{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbb{P}(A_i)$
    pour *chaque* S, sous-ensemble de $\{1, 2, \ldots, n\}$

- définir l'univers Ω.
- définir les probabilités associées.
  Par exemple considérer que les issues sont équiprobables:
  $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$ (Laplace).

DANGEREUX

- Utiliser – Indépendance – Bayes – Probabilités totales

Exemple

Chaîne de production

◁

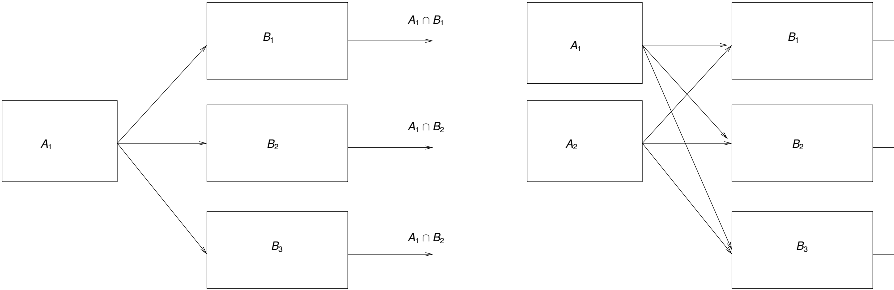


Figure: chaîne de production : risque de panne

$\mathbb{P}(A_i \text{ en panne}) = 0.2$ et $\mathbb{P}(B_i \text{ en panne}) = 0.4$

Question

Quelle est la probabilité de panne dans les deux cas ?

1.
   - Soit A_i l'événement " la chaîne A_i fonctionne correctement" et de même pour B_i. On a $\mathbb{P}(A_i) = 0.8$ et $\mathbb{P}(B_i) = 0.6$
   - Soit l'événement F : "la chaîne fonctionne". Selon le premier schéma, on a $F = (A_1 \cap B_1) \cup (A_1 \cap B_2) \cup (A_1 \cap B_3) = A_1 \cap (B_1 \cup B_2 \cup B_3)$.
   - A_1 est indépendant des B_i et donc $\mathbb{P}(F) = \mathbb{P}(A_1).\mathbb{P}(B_1 \cup B_2 \cup B_3)$.
   - De plus, par les lois des ensembles, $\mathbb{P}(B_1 \cup B_2 \cup B_3) = 1 - \mathbb{P}(B_1 \cup B_2 \cup B_3)^c = 1 - \mathbb{P}\left(B_1^c \cap B_2^c \cap B_3^c\right)$, et par indépendance de B_i, on a indépendance des B_i^c et donc $\mathbb{P}\left(B_1^c \cap B_2^c \cap B_3^c\right) = \mathbb{P}\left(B_1^c\right).\mathbb{P}\left(B_2^c\right).\mathbb{P}\left(B_3^c\right)$.
   - Globalement on a alors $\mathbb{P}(F) = \mathbb{P}(A_1).\mathbb{P}(B_1 \cup B_2 \cup B_3) = \mathbb{P}(A_1).(1 - \mathbb{P}\left(B_1^c\right).\mathbb{P}\left(B_2^c\right).\mathbb{P}\left(B_3^c\right)) = 0.8(1 - (0.4)^3) = 0.75$, soit une probabilité de panne de 25 % .
2. Sur le deuxième schéma, $F = (A_1 \cup A_2) \cap \mathbb{P}(B_1 \cup B_2 \cup B_3)$, soit, toujours par indépendance,$\mathbb{P}(F) = \mathbb{P}(A_1 \cup A_2).\mathbb{P}(B_1 \cup B_2 \cup B_3)$. Avec le même raisonnement que ci-dessus, on obtient $\mathbb{P}(A_1 \cup A_2) = 1 - \mathbb{P}(A_1^c)\,\mathbb{P}(A_2^c) = 0.96$ et donc $\mathbb{P}(F) = 0.96 \times 0.94 = 0.9$, soit une probabilité de panne de 10 %.

Associer un nombre à un événement

$\Omega = \{''\mathit{faux}'', ''\mathit{vrai}''\}$: compliqué à manipuler
$\Rightarrow$ Associer un nombre :

"*faux*"	$\rightarrow 0$
"*vrai*"	$\rightarrow 1$

Définition : Variable aléatoire

Une *Variable aléatoire* X est une fonction de Ω dans $\mathbb{R}$ telle qu'à tout ω correspond une valeur $X(\omega) = x$.

Définition : Domaine de variation

Le *domaine de variation* de X est l'ensemble $R_X \subset \mathbb{R}$ que peut prendre la variable aléatoire X.

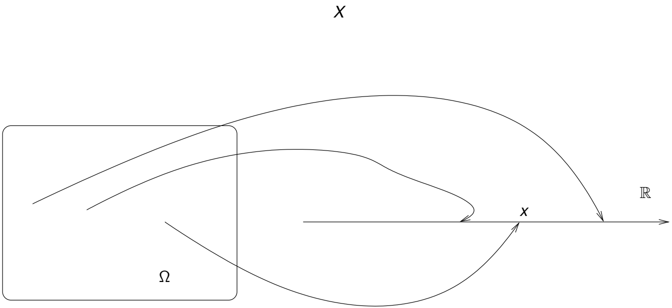


Figure: La variable aléatoire : une fonction de l'univers dans l'espace des réels

Définition : Réalisation

On appelle x une *réalisation* de la variable aléatoire liée à **l'événement** ω.

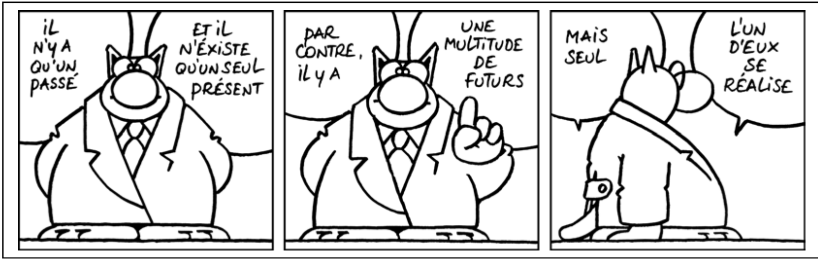


Figure: Définition d'une réalisation par le chat de Geluck

Exemple de variable aléatoire

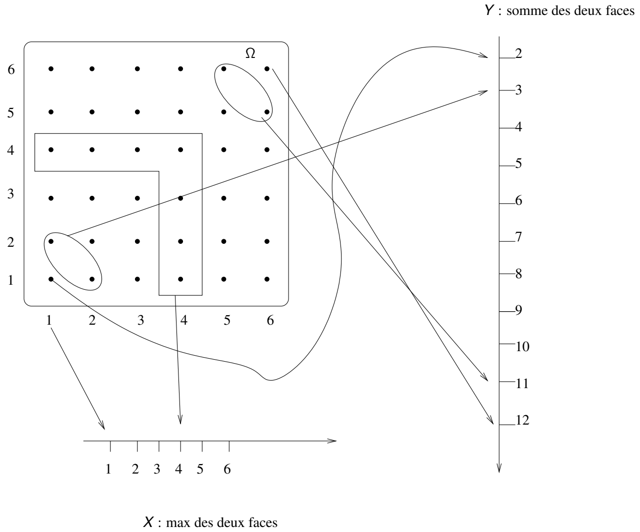


Figure: Exemple de deux variables aléatoires liées à la même expérience aléatoire

Variable aléatoire discrète

La *Variable aléatoire discrète* X prend ses valeurs dans un ensemble *fini* de valeurs *dénombrable* ou *non dénombrable*.

exemple Les domaines de variation de X et Y sont respectivement $R_X = \{1, 2, 3, 4, 5, 6\}$ et $R_Y = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$.

une v.a est définie par sa *fonction de probabilité* (*masse de probabilité* pour les v.a. discrètes).

Masse de probabilité

La fonction de probabilité est la fonction $p_X(x)$ définie par :

$$p_X(x) = \begin{cases} \mathbb{P}\left(\underbrace{\{X = x\}}_{\text{événement } \in \Omega}\right) \overset{\text{simpl.}}{=} \mathbb{P}(X = x) & \text{si } x \in R_X \\ 0 & \text{si } x \notin R_X \end{cases}$$

Propriétés

- $p_X(x) \geq 0 \qquad \forall x \in \mathbb{R}.$
- $\displaystyle\sum_{i=1}^{n} p_X(x_i) = 1 \qquad \text{si } R_X = \{x_1, x_2, \cdots x_n\}.$

Exemple : Détermination d'une masse de probabilité

Une classe de n élèves présente un examen. La masse de probabilité X du nombre d'élèves ayant la note x est une masse de probabilité triangulaire entre $x = 2$ et $x = 18$ sur 20, avec $p_X(2) = 0, p_X(10) = 8.a, p_X(18) = 0$.
On a donc

$$p_X(x) = \begin{cases} 0 & x < 3 \\ a * (x - 2) & 2 < x < 11 \\ a * (8 - (x - 10)) & 10 < x < 18 \\ 0 & x > 17 \end{cases}$$

Trouver a tel que $\sum p_X(x_i) = 1$.
$\Rightarrow \sum p_X(x_i) = 64.a$ et donc $a = 1/64$.

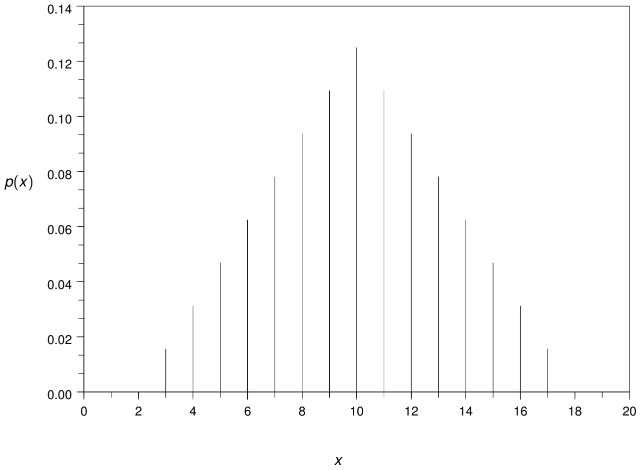


Figure: Fonction ou masse de probabilité des notes d'une classe

Definition : Fonction de répartition

$F(x) = F_X(x) \triangleq \mathbb{P}(\{X \leq x\}) = \sum_{x_i \leq x} p_X(x_i)$

- Si on classe les éléments de R_X par ordre: $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$,

$$F_X(x_{(k)}) = \mathbb{P}\big(\{X \leq x_{(k)}\}\big) = \sum_{i=1}^{k} p_X(x_{(i)})$$

- $0 \leq F(x) \leq 1, \qquad \forall x \in \mathbb{R}$
- $F(x_{(k)}) - F(x_{(k)}^-) = p_X\big(x_{(k)}\big)$
- La fonction $F(x)$ est monotone croissante (au sens large) :

$$\forall x_i < x_j \ , \ F(x_i) \leq F(x_j)$$

- La fonction $F(x)$ “démarre” en 0 et “termine” en 1
  - $\lim_{x \to -\infty} F_X(x) = 0$
  - $\lim_{x \to +\infty} F_X(x) = 1$
- $\forall x_i \in R_X, \qquad F(x_i) - F_X(x_{i-1}) = \mathbb{P}(\{x_{i-1} < X \leq x_i\})$.

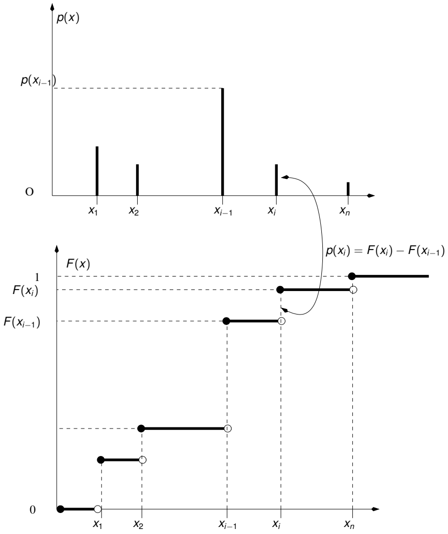


Figure: La fonction de répartition : lien avec la masse de probabilité.

Définition : Variable aléatoire continue

Une *variable aléatoire continue* est une variable aléatoire définie sur un domaine de variation continu.

Fonction de répartition

$F_X(x) = \mathbb{P}(X \leq x)$

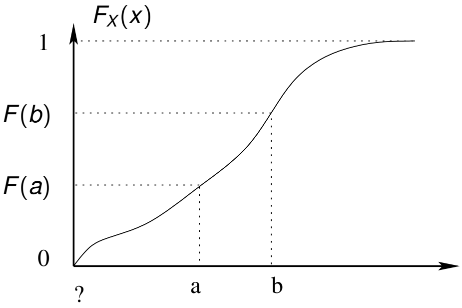


Figure: Fonction de répartition d'une v.a. continue.

Propriétés de la fonction de répartition

$$\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a)$$

car $(a \leq X \leq b)$ = $(X \leq b) \backslash (X \leq a)$ et donc
$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a)$.

Probabilité et Fonction de répartition

Dans le cas continu, la probabilité $\mathbb{P}(X = x) = 0$ et donc
$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X < x)$.

Définition : Densité de probabilité

La *densité de probabilité* $f_X(x)$ est définie par :

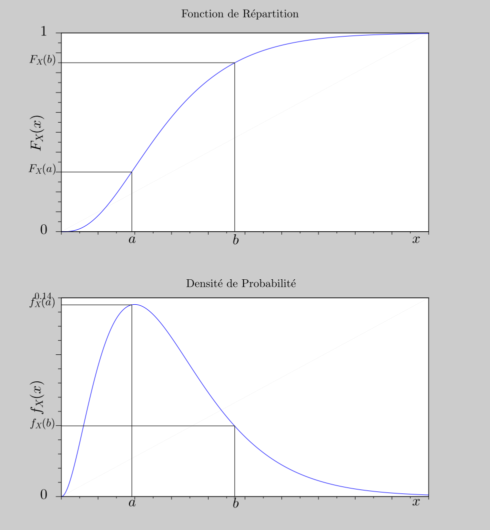
$$f_X(x) = \lim_{\epsilon \to 0} \frac{\mathbb{P}(x < X \leq x + \epsilon)}{\epsilon}$$

$\mathbb{P}(x < X \leq x + \epsilon) = F_X(x + \epsilon) - F_X(x)$ et donc, en supposant la fonction de répartition dérivable, on a la définition suivante :

Definition : Densité de probabilité

La *densité de probabilité* $f_X(x)$ est définie par :

$$f_X(x) = \frac{dF_X(x)}{dx}$$


Fonction de Répartition
1
$F_X(b)$
$F_X(x)$
$F_X(a)$
0
a
b
x
Densité de Probabilité
0.14
$f_X(a)$
$f_X(x)$
$f_X(b)$
0
a
b
x

Soit une v.a. X et un événement **A**,
que devient X **si A ?**

DEFINITION : Fonction de répartition conditionnelle

Soit **A** et $\mathbb{P}(A) \neq 0$ la *fonction de répartition conditionnelle* $F_{X|\mathbf{A}}(x|\mathbf{A})$ est telle que :

$$F_{X|\mathbf{A}}(x|\mathbf{A}) = \mathbb{P}((X \leq x)|\mathbf{A}) = \frac{\mathbb{P}((X \leq x) \cap \mathbf{A})}{\mathbb{P}(\mathbf{A})}.$$

DEFINITION :Masse de probabilité conditionnelle

Soit **A** et $\mathbb{P}(A) \neq 0$ la *masse de probabilité conditionnelle* $p_{X|\mathbf{A}}(x|\mathbf{A})$ est telle que :

$$p_{X|\mathbf{A}}(x|\mathbf{A}) = \mathbb{P}((X = x)|\mathbf{A}) = \frac{\mathbb{P}((X = x) \cap \mathbf{A})}{\mathbb{P}(\mathbf{A})}.$$

DEFINITION :Densité de probabilité conditionnée sur un événement

Soit $\mathbf{A}$ et $\mathbb{P}(A) \neq 0$ la *densité de probabilité conditionnelle* $f_{X|\mathbf{A}}(x|\mathbf{A})$ telle que :

$$f_{X|\mathbf{A}}(x|\mathbf{A}) = \frac{dF_{X|\mathbf{A}}(x|\mathbf{A})}{dx}$$

DEFINITION :masse de probabilité jointe

Soient deux v.a. discrètes X et Y, la **masse de probabilité jointe** est :

$$p_{XY}(x, y) = \mathbb{P}((X = x) \cap (Y = y))$$

En particulier, $p_{XY}(x, y) \geq 0$ et $\sum_x \sum_y p_{XY}(x, y) = 1$

DEFINITION :masse de probabilité marginale

oient deux v.a. discrètes X et Y, les **masses de probabilité marginales** cont :

$$p_X(x) = \mathbb{P}((X = x)) = \sum_y p_{XY}(x, y)$$

$$p_Y(y) = \mathbb{P}((Y = y)) = \sum_x p_{XY}(x, y)$$

variable aléatoire conditionnelle discrète

Soient les v.a. X et Y,
Que devient X, si $Y = y_i$ est connu ?
De façon simple, il s'agit, $\forall x_i$ de connaitre $\mathbb{P}(X = x_i | Y = y_j)$. Il s'agit donc de la variable aléatoire conditionnée sur l'événement $(Y = y_j)$
ICI : $Y = y_i$ est un événement de proba non nulle

DEFINITION :Fonction de répartition jointe

Soient deux v.a. continues X et Y, la **Fonction de répartition jointe** est :

$$F_{XY}(x, y) = \mathbb{P}((X \leq x) \cap (Y \leq y))$$

DEFINITION :Densité de probabilité jointe

Soient deux v.a. continues X et Y, la **Densité de probabilité jointe** est :

$$f_{XY}(x, y) = \frac{dF_{XY}(x, y)}{dxdy}$$

En particulier, $f_{XY}(x, y) \geq 0$ et $\int_x \int_y f_{XY}(x, y) = 1$ et
Pour toute région R de l'espace engendré par x et y :

$$\mathbb{P}([X, Y] \in R) = \int \int_R f_{XY}(x, y)\, dxdy$$

DEFINITION :Densité de probabilité marginale

Soient deux v.a. discrètes X et Y, les **masses de probabilité marginales** cont :

$$f_X(x) = \int_{y=-\infty}^{y=+\infty} f_{XY}(x, y)\, dy$$

$$f_Y(y) = \int_{x=-\infty}^{x=+\infty} f_{XY}(x, y)\, dx$$

variable aléatoire conditionnelle continue

Soient les v.a. X et Y,
Que devient X, si $Y = y$ est connu ?

On se rappelera que $\mathbb{P}(x < X \leq x + \Delta_x) = f_X(x)\,\Delta_x$, pour Δ_x petit.
On notera que $\mathbb{P}((x < X \leq x + \Delta_x)|\mathbf{A}) = f_{X|A}(x)\,\Delta_x$ car la dépendance sur **A**
Donc

$$\mathbb{P}((x < X \leq x + \Delta_x)|(y < Y \leq x + \Delta_y)) = \frac{\mathbb{P}((x < X \leq x + \Delta_x),(y < Y \leq y + \Delta_y))}{\mathbb{P}(y < Y \leq y + \Delta_y)}$$

On en déduit directement que

$$f_{X|Y}(x|y)\Delta_x = \frac{f_{XY}(xy)\Delta_x\Delta_y}{f_Y(y)\Delta_y}$$

DEFINITION :Densité de probabilité conditionnelle

Soient deux variable aléatoires continues X et Y, de densité conjointe $f_{XY}(xy)$ et de densité $f_Y(y)$ non nulle sur le support de Y, alors la densité conditionnelle, dite de X conditionnée sur Y est donnée par :

$$f_{X|Y}(x|y) = \frac{f_{XY}(xy)}{f_Y(y)}.$$

On exprime également la densité conditionnelle de la manière suivante, en explicitant la valeur de Y sur laquelle on conditionne :

$$f_{X|Y}(x|Y = y_o) = \frac{f_{XY}(xy)}{f_Y(y_o)}.$$

Le Mode d'une v.a. X est la valeur x_m : $x_m = \arg\max_x f_X(x)$. Valeur la plus *vraisemblable*.

La Médiane d'une v.a. X est la valeur $x_{\frac{1}{2}}$: $\mathbb{P}\left(X \leq x_{\frac{1}{2}}\right) = 0.5$.

Les Quantiles d'une v.a. X, plus précisément le *p-quantile* est d'une v.a. X est la valeur x_p : $\mathbb{P}(X \leq x_p) = p$.

L'espérance mathématique d'une v.a. X est la moyenne de cette v.a., pondérée par sa densité de probabilité. Intuitivement, c'est la valeur qu'on s'attend à observer en moyenne (que l'on "espère" observer).

La variance d'une v.a. X est une mesure (du carré) de la variation qu'on peut observer autour de la moyenne. L'espérance et la variance donnent une bonne idée du domaine de variation de la variable aléatoire X.

Les moments d'une v.a. X sont l'espérance des puissances de la v.a. (donc la moyenne est le moment d'ordre 1, puisque c'est l'espérance de la X à la puissance 1, le moment d'ordre 2 est lié à la variance, etc.).

DEFINITION :Mode

Le mode d'une variable aléatoire X est la valeur x_m telle que :

$$\forall x \in R_X, x \neq x_m, \qquad p_X(x_m) > p_X(x) \text{ pour une v.a. discrète ;}$$
$$\forall x \in R_X, x \neq x_m, \qquad f_X(x_m) > f_X(x) \text{ pour une v.a. continue.}$$

On impose une inégalité stricte ($f_X(x_m) > f_X(x)$), donc :

- le mode n'est pas toujours défini (exemple de l'uniforme) ;
- à strictement parler, il n'y a qu'un seul mode. Cependant, on parle souvent de *v.a. multimodale* s'il y a plusieurs maxima locaux ; dans le cas contraire, on parle de *v.a. unimodale*.

v.a. multimodale : notes d'un D.S.

Un bon examen a souvent un résultat bi-modal :

- Ceux qui ont travaillé : notes autour de 14
- Ceux qui n'on pas travaillé : notes autour de 8
- soit N la v.a. représentant les notes et l'événement **P** : **P** $\equiv$ "l'étudiant a préparé".
    1. $N|P \sim N(14, 4)$
    2. $N|P^c \sim N(8, 4)$

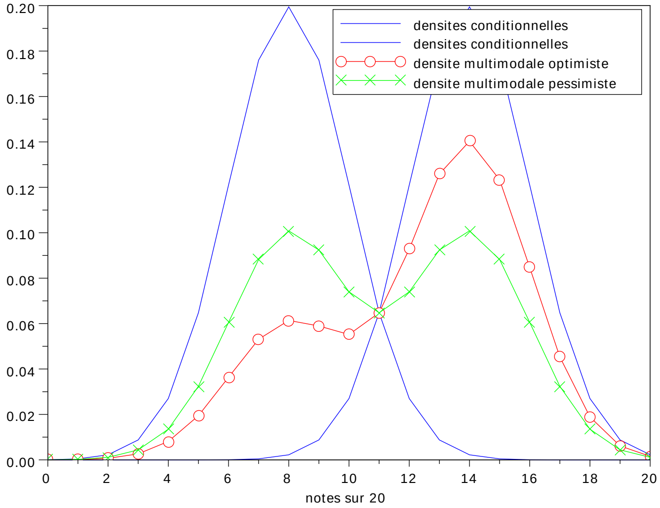


Figure: Densité de probabilité bimodale

DEFINITION :Médiane

La médiane d'une variable aléatoire X est la valeur $x_{\frac{1}{2}}$ telle que :

$$\mathbb{P}\left(X \leq x_{\frac{1}{2}}\right) \triangleq F(x_{\frac{1}{2}}) = \frac{1}{2}$$

N'existe pas toujours pour une v.a. discrète.

DEFINITION :$p-$quantile

Le $p-$quantile d'une variable aléatoire X est la valeur x_p telle que :

$$\mathbb{P}(X \leq x_p) \triangleq F(x_p) = p, \qquad p \in [0, 1].$$

On distingue en particulier

- **La médiane :** pour $p = 1/2$, qui “divise” en deux le domaine de variation de la v.a.
- **Les quartiles :** pour $p = 1/4$ (le premier quartile), $p = 1/2$ et $p = 3/4$ (le troisième quartile. Les quartiles “divisent” le domaine de variation de la v.a. en quatre parties “égales” (c’est-à-dire dont la surface sous la densité de probabilité est divisée en quatre parties égales). On a donc que la probabilité de se trouver entre deux quartiles successifs vaut 1/4.
- **Les déciles :** pour $p = k/10$, $k = 1, 2, ..., 9$. $x_{0.1}$ est le premier décile, etc. On a donc que la probabilité de se trouver entre deux déciles successifs vaut 1/10.
- **Les centiles :** pour $p = k/100, k = 1, 2, , 99$. L’utilité est plutôt pour les grands et petits centiles, par exemple, a probabilité d’obtenir une valeur supérieur au 99$^{\text{ème}}$ centile est de 1 pourcent.

DEFINITION :Intervalle de confiance

Un intervalle de confiance (bilatéral) $[a, b]$ au niveau p est tel que

$$\mathbb{P}(a \leq X \leq b) = p.$$

Un intervalle de confiance (unilatéral) $[-\infty, b]$ au niveau p est tel que :

$$\mathbb{P}(X \leq b) = p.$$

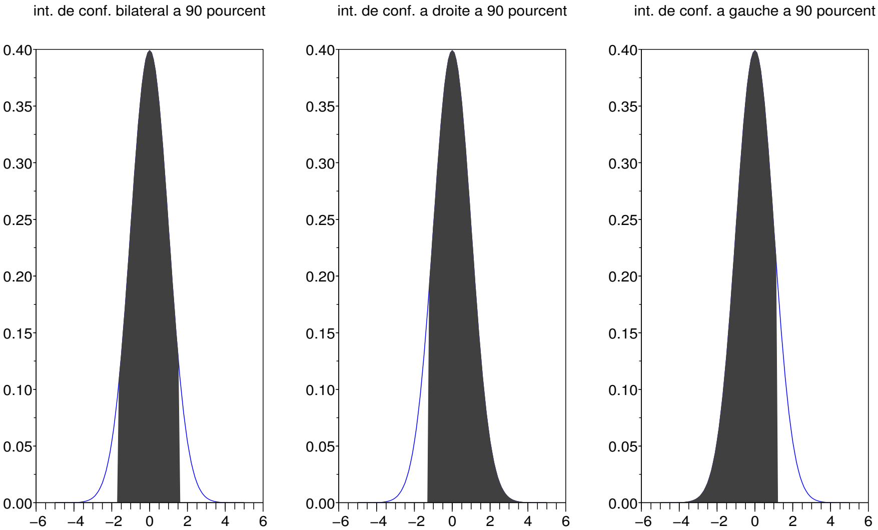
Un intervalle de confiance (unilatéral) $[a, \infty]$ au niveau p est tel que :

$$\mathbb{P}(a \leq X) = p.$$

Dans le cas bilatéral $\mathbb{P}(X < a) = \mathbb{P}(X > b) = (1 - p)/2 = \alpha/2$, où $\alpha = 1 - p$.

Confiance et erreur

α représente alors la probabilité qu'on a de se tromper si on fait l'hypothèse que la réalisation x de la v.a. X est dans l'intervalle $[a, b]$, et on a que $a = x_{\alpha/2}$ et $b = x_{1-\alpha/2}$.


int. de conf. bilateral a 90 pourcent
int. de conf. a droite a 90 pourcent
int. de conf. a gauche a 90 pourcent
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
–6 –4 –2 0 2 4 6

DEFINITION :Espérance mathématique

L'espérance mathématique μ ou $\mathrm{E}[X]$ est définie par :

$$\mu = \mathrm{E}[X] = \begin{cases} \displaystyle\sum_{i=1}^{n} x_i p_X(x_i) & \text{cas discret} \\ \displaystyle\int_{-\infty}^{\infty} x f_X(x)\, dx & \text{cas continu} \end{cases}$$

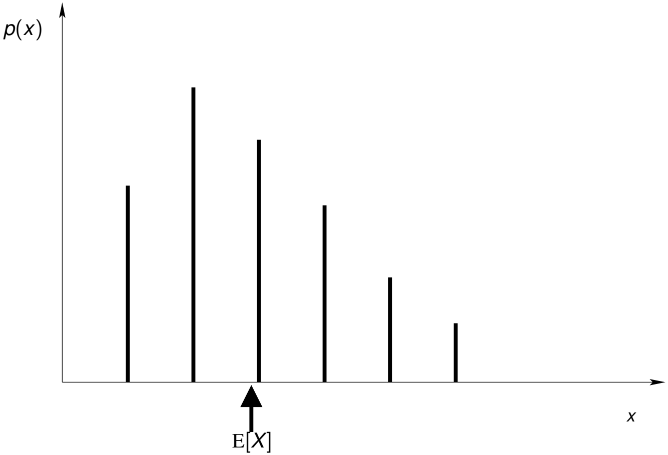


Figure: Interprétation d'un espérance comme étant un centre de gravité

L'espérance d'une fonction d'une variable aléatoire, contrairement à la loi de probabilité de cette fonction, s'obtient très facilement :

DEFINITION :Espérance d'une fonction d'une v.a.

Si $Y = g(X)$, où $g(.)$ est une fonction, alors :

$$\mathrm{E}[Y] = \mathrm{E}[g(X)] = \begin{cases} \displaystyle\sum_{i=1}^{n} g(x_i)p_X(x_i) & \text{pour une v.a. discrète} \\ \displaystyle\int_{-\infty}^{\infty} g(x)f_X(x)\,dx & \text{pour une v.a. continue} \end{cases}$$

Si $Y = aX + b$, alors, l'application directe de la linéarité de la somme et de l'intégrale donne la propriété suivante :

Propriété

Linearité de l'espérance

Si $Y = aX + b$ alors :

$$\mathrm{E}[Y] = a\mathrm{E}[X] + b$$

DEFINITION :v.a. de Bernoulli

x	$p_X(x)$
1	p
0	1-p
$x \neq \{0,1\}$	0

On dira que $X \sim Be(p)$: la variable aléatoire X est distribuée selon la loi de Bernoulli

DEFINITION :Variable aléatoire binomiale

Soit $Y_1 \sim \cdots \sim Y_n \sim Be(p)$ avec les Y_i mutuellement indépendantes, alors :

$$X = Y_1 + \cdots + Y_n \sim Bi(n,p)$$

est une variable aléatoire binomiale.

Exemple

v.a. de Bernoulli et Binômiale

Une variable aléatoire de Bernoulli aura un espérance égale à (pour $X \sim Be(p)$) :

$$\mathrm{E}[X] = 0.(1-p) + 1.p = p.$$

Une variable aléatoire Binômiale $X \sim Bi(n,p)$ aura une espérance :

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{i}^{n} X_i\right] = \sum_{i}^{n} \mathrm{E}[X_i] = np.$$

$\triangleleft$

DEFINITION :Variance

La variance d'une variable aléatoire X est donnée par :

$$\mathrm{var}[X] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \int_{R_X} (X - \mathrm{E}[X])^2 f_X(x)\, dx.$$

Bernoulli

Si $X \sim Be(p)$, alors

$$\begin{aligned} \mathrm{E}[X] &= \textstyle\sum_{k=0}^{1} k p_X(k) \\ &= 0(1-p) + 1.p \\ &= p. \end{aligned} \tag{2}$$

$$\begin{aligned} \mathrm{var}[X] &= \textstyle\sum_i (x_i - \mathrm{E}[X])^2 p_X(x_i) \\ &= \textstyle\sum_{k=0}^{1} (k-p)^2 P_X(k) \\ &= (0-p)^2(1-p) + (1-p)^2 p \\ &= p(1-p) \end{aligned} \tag{3}$$

DEFINITION :Variable aléatoire binomiale

Une variable aléatoire est binomiale $X \sim Bi(n,p)$ si et seulement si :

$$p_X(x) = \begin{cases} C_n^x p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, \cdots, n \text{ avec } C_n^x = \dfrac{n!}{x!(n-x)!} \\ 0 & \text{sinon} \end{cases}$$

Binomiale

si $X \sim Bi(n,p)$, alors

$$\begin{aligned} \mathrm{E}[X] &= \textstyle\sum_{k=0}^{n} k p_X(k) \\ &= \textstyle\sum_{k=0}^{n} = C_n^k p^k (1-p)^{n-k} \\ &= np. \end{aligned} \tag{4}$$

Espérance d'une fonction de v.a. multiples

$$\mathrm{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\, dxdy$$

Cas particulier :

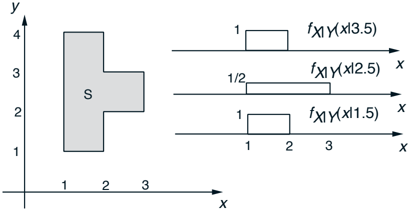
$$\mathrm{E}[aX + bY] = a\mathrm{E}[X] + b\mathrm{E}[Y]$$

Rappel de densité conditionnelle

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

et

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)\, dx = 1$$


y
4
3
2
1
S
1
2
3
x
1
f X|Y(x|3.5)
x
1/2
f X|Y(x|2.5)
x
1
f X|Y(x|1.5)
1
2
3
x

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\text{area of the circle}} & \text{if } (x,y) \text{ is in the circle,} \\ 0 & \text{otherwise,} \end{cases}$$
$$= \begin{cases} \dfrac{1}{\pi r^2} & \text{if } x^2 + y^2 \leq r^2, \\ 0 & \text{otherwise.} \end{cases}$$

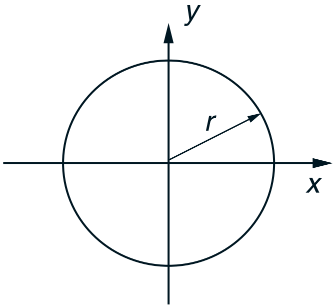


Figure 3.19: Circular target for Example 3.16.

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx \\
&= \frac{1}{\pi r^2} \int_{x^2+y^2} \leq r^2 dx \\
&= \frac{1}{\pi r^2} \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} dx \\
&= \frac{2}{\pi r^2} \sqrt{r^2 - y^2}
\end{aligned}$$

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{f_{XY}(x,y)}{f_Y(y)} \\
&= \frac{\frac{1}{\pi r^2}}{\frac{2}{\pi r^2}\sqrt{r^2-y^2}} \\
&= \frac{1}{2\sqrt{r^2-y^2}}
\end{aligned}$$

$$\mathrm{E}[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$$

$$\begin{aligned}
\mathrm{E}[X|Y=y] &= \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} \frac{x}{2\sqrt{r^2-y^2}} dx \\
&= \left. \frac{x^2}{4\sqrt{r^2-y^2}} \right]_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} \\
&= 0
\end{aligned}$$

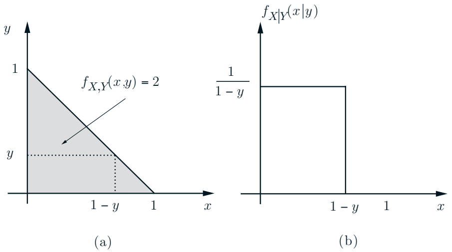


Figure 4.6: (a) The joint PDF in Example 4.15. (b) The conditional density of X.

We have

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^{1-y} 2\,dx = 2(1-y), \qquad 0 \le y \le 1,$$

and

$$f_{X|Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{1-y}, \qquad 0 \le x \le 1-y.$$

On a : $\mathrm{E}[X|Y=y] = \frac{1-y}{2}$
Si on considère que y varie, on peut le considérer comme une v.a. (Y)
On écrit alors $\mathrm{E}[X|Y] = \frac{1-Y}{2}$ (!!!!! c'est une v.a. !!!!!)
On peut donc en prendre l'espérance :

$$\begin{aligned} \mathsf{E}_Y(\mathsf{E}_X(X|Y)) = \mathrm{E}[EX|Y] = & \int_{-\infty}^{\infty} \mathrm{E}[X|Y=y]\, f_Y(y)\, dy \\ & = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, f_Y(y)\, dxdy \\ & = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)\, dxdy \\ & = \mathrm{E}[X] \end{aligned}$$

Espérances itérées

$$\mathsf{E}_Y(\mathsf{E}_X(X|Y)) = \mathrm{E}[\mathrm{E}[X|Y]] = \mathrm{E}[X]$$

Exemple

- $\mathrm{E}[X|Y] = \frac{1-Y}{2}$
- $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]] = \frac{1-\mathrm{E}[Y]}{2}$
- par symétrie $\mathrm{E}[X] = \mathrm{E}[Y]$
- Donc $\mathrm{E}[X] = 1/3$

DEFINITION :Variable aléatoire de Bernoulli

X est une *variable de Bernoulli*

x	$p_X(x)$
1	p
0	1-p
$x \neq \{0, 1\}$	0

On dira que $X \sim Be(p)$:

Exemple

Encore la famille et garçon/fille

Soit $\mathbf{A}$ = "le cadet est un garçon" et $\mathbb{P}(\mathbf{A}) = 1/2$. On a donc $X \sim Be(1/2)$.
Soit $\mathbf{A}$ = "les trois premiers enfants sont des garçons", $\mathbb{P}(\mathbf{A}) = 1/8$ et donc $X = Be(1/8)$. $\triangleleft$

DEFINITION :Variable aléatoire binomiale

Soit $Y_1 \sim \cdots \sim Y_n \sim Be(p)$ avec les Y_i mutuellement indépendantes, alors :

$$X = Y_1 + \cdots + Y_n \sim Bi(n, p)$$

est une variable aléatoire binomiale.

Exemple

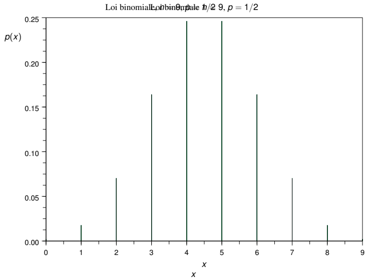
Jet de n pièces de monnaie

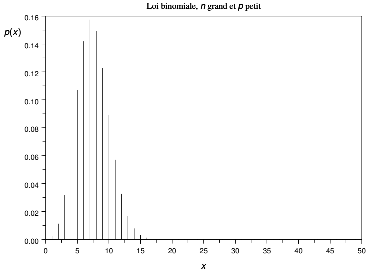
Supposons que l'on jette une pièce n fois.
$\mathbb{P}(\text{pile}) = p$ indépendamment d'un jet à l'autre.

Le succès est ici “pile” et donc $Y_i = 1 \Rightarrow X = \sum_{i=1}^{n} Y_i$.

$p_X(x) = \mathbb{P}(X = x) = C_n^x p^x (1-p)^{n-x}$ si $x = 0, 1, 2, \cdots, n$ avec $C_n^x = \dfrac{n!}{x!(n-x)!}$

$\triangleleft$


Loi binomiale, n = 9, p = 1/2
p(x)
0.25
0.20
0.15
0.10
0.05
0.00
0
1
2
3
4
5
6
7
8
9
x


Loi binomiale, n grand et p petit
p(x)
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
25
30
35
40
45
50
x

DEFINITION :Variable aléatoire binomiale

Une variable aléatoire est binomiale $X \sim Bi(n,p)$ si et seulement si :

$$p_X(x) = \begin{cases} C_n^x p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, \cdots, n \text{ avec } C_n^x = \dfrac{n!}{x!(n-x)!} \\ 0 & \text{sinon} \end{cases}$$

DEFINITION :Variable aléatoire géométrique

Une variable aléatoire suit une loi géométrique de paramètre p (on notera $X \sim Ge(p)$) si et seulement si :

$$p_X(x) = \begin{cases} p(1-p)^{x-1} & \text{si } x = 1, 2, \ldots \\ 0 & \text{sinon} \end{cases}$$

et

$$F(x) = 1 - (1-p)^x \qquad x = 1, 2, \ldots$$

Soit une v.a. géométrique $X \sim Ge(p)$, son espérance vaut :

$$\mathrm{E}[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}.p = \frac{1}{p} \tag{5}$$

Sa variance vaut :

$$\mathrm{var}[X] = \frac{1-p}{p^2} \tag{6}$$

v.a. Pascal

Réussir une épreuve k fois.

DEFINITION :Variable aléatoire de Pascal

Une variable aléatoire est dite de Pascal $X \sim Pa(k, p)$ si et seulement si :

$$p_X(x) = \begin{cases} C_{x-1}^{k-1} p^k (1-p)^{x-k} & \text{si } x = k, k+1, \cdots \\ 0 & \text{sinon} \end{cases}$$

Exemple

Poker

Calculer les probabilités d'avoir, respectivement, une paire, un brelan et un carré ? (nombre infini de cartes : $\mathbb{P}(''As'') = 1/13$)
Dans ce cas, on obtient :

- Pour une paire : $p_X(x) = C^1_{x-1} p^2 (1-p)^{x-2}$ Soit :

x	2	3	4	5	6	$\cdots$
$p_X(x)$	0.0059	0.0109	0.0151	0.0186	0.0215	$\cdots$

Et donc la probabilité d'avoir une paire sur 5 cartes vaut $13.0.0186 = 0.242$

- Pour un brelan : $p_X(x) = C^2_{x-1} p^3 (1-p)^{x-3}$ Soit :

x	2	3	4	5	6	$\cdots$
$p_X(x)$	0.0	0.00045	0.00126	0.00232	0.00358	$\cdots$

Et donc la probabilité d'avoir un brelan sur 5 cartes vaut $13.0.00232 = 0.06$

$\triangleleft$

Loi binômiale : comportement particulier pour p petit et n grand.
Nombre d'occurences de A si $\mathbb{P}(A) = p$ pour p petite ?
: X tend vers la *loi de Poisson* de paramètre μ

$$X \sim Po(\mu)$$

.
La fonction de probabilité $p_X(x)$ de la v.a. de Poisson vaudra alors :

DEFINITION :Variable aléatoire de Poisson

Une v.a. discrète est dite de Poisson ($X \sim Po(\mu)$) si et seulement si :

$$p_X(x) = \begin{cases} \dfrac{e^{-\mu}\mu^x}{x!} & \text{si } x = 0, 1, 2, \cdots \qquad \mu > 0 \\ 0 & \text{sinon} \end{cases}$$

Soit une v.a. de Poisson $X \sim Po(\lambda)$, son espérance vaut :

$$\begin{aligned} \mathrm{E}[X] &= \sum_{i=0}^{\infty} i \frac{e^{-\lambda}\lambda^i}{i!} \\ &= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda \end{aligned} \tag{7}$$

Sa variance vaut également λ.

Soit une loi binomiale $X \sim Bi(n,p)$ de valeurs n grande et p petite et de produit np fini. Dans ce cas, on a un événement de faible probabilité, mais qui après un grand nombre d'essais, se produira np fois en moyenne. Cette loi binomiale peut être approximé par une loi de Poisson W de paramètre $w = np$. On a donc, pour une variable aléatoire binomiale :

$$\boxed{X \overset{n\to\infty}{\sim} Po(np)}$$

En effet, la masse de probabilité $p_X(x)$ de la binomiale vaut :

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} \frac{w^x}{n^x} \left(1 - \frac{w}{n}\right)^{n-x}$$

$$= \frac{n(n-1)\ldots(n-x+1)}{x!} \frac{w^x}{n^x} \left(1 - \frac{w}{n}\right)^{n-x}$$

$$= \frac{n}{n} \frac{n-1}{n} \ldots \frac{n-x+1}{n} \frac{w^x}{x!} \left(1 - \frac{w}{n}\right)^{n} \left(1 - \frac{w}{n}\right)^{-x}$$

$$\underset{n\to\infty}{\longrightarrow} 1 \cdot 1 \cdot \ldots \cdot 1 \frac{w^x}{x!} \mathrm{e}^{-w} \cdot 1 = \mathrm{e}^{-w} \frac{w^x}{x!} = p_W(x)$$

DEFINITION :Variable aléatoire uniforme

Une variable aléatoire uniforme que l'on note

$$X \sim Un(a, b)$$

est définie par sa densité de probabilité :

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

Elle est définie de façon équivalente par sa fonction de répartition :

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } x \in [a, b] \\ 1 & \text{sinon} \end{cases}$$

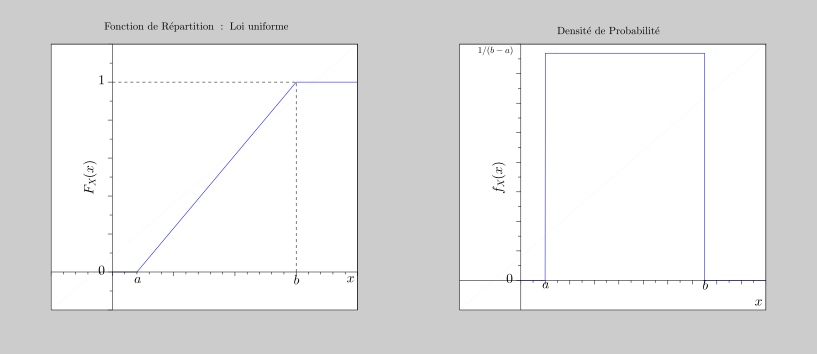


Figure: Loi de probabilité uniforme

On obtient aisément que si $X \sim Un(a, b), \mathrm{E}[X] = (a + b)/2$, et que variance vaut $\mathrm{var}[X] = (b - a)^2/12$.

DEFINITION :Variable aléatoire exponentielle

Une variable aléatoire est dite exponentielle de paramètre λ, notée $X \sim Exp(\lambda)$ si et seulement si :

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \; ; F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

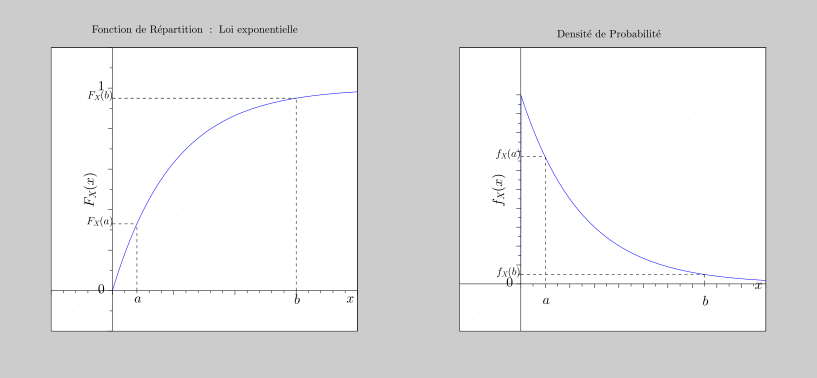


Figure: Loi de probabilité exponentielle

Exemple

Durée de vie d'un composant

Les composants suivent une loi exponentielle de paramètre $\lambda = \frac{1}{20}\text{ans}^{-1}$.
On demande de calculer :

- La probabilité que le composant fonctionne plus de (10 ; 15 ; 20 - ;25) ans.
- La demi-vie, i.e. le temps x tel que la probabilité que la durée de vie excède x soit égale à 0.5.

Solution

- On a que $\mathbb{P}(X > x) = 1 - \mathbb{P}(X < x) = 1 - F_X(x)$, ce qui donne des probabilités de (0.61 ; 0.47 ; 0.37 ; 0.29). On notera que si l'espérance de vie est de 20 ans, la probabilité d'atteindre 20 ans n'est que de 37 % !.
- On cherche x tel que $1 - F_X(x) = 0.5$, donc $F_x = e^{-x/\lambda} = 0.5$: $x = -20 * \ln 0.5 \simeq 13.8$, et la demi-vie est de presque 14 ans.

$\triangleleft$

Si $X \sim \exp(\lambda)$, alors

$$\mathrm{E}[X] = \int_0^\infty x\lambda e^{-\lambda x} dx = \left[-xe^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} \right] \Big\|_0^\infty = \frac{1}{\lambda}. \tag{8}$$

De même, $\mathrm{var}[X] = 1/\lambda^2$.

La plus aléatoire

DEFINITION :Variable aléatoire Normale

Une variable aléatoire normale (encore appelée Gaussienne) de paramètres μ (fini) et σ^2 (positif), notée

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

est définie par sa densité de probabilité :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \forall x \in \mathbb{R}$$

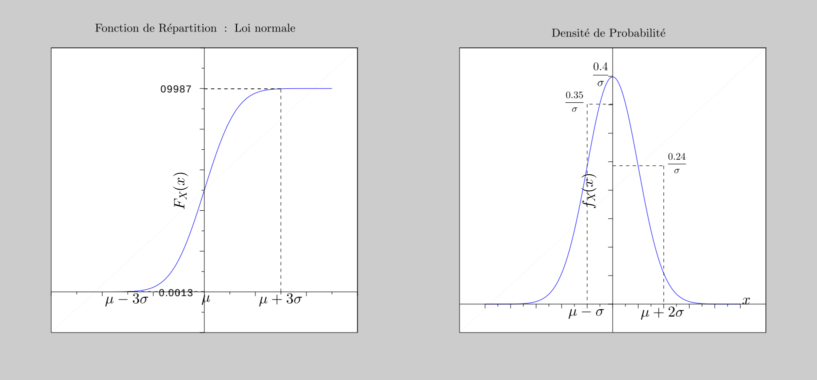


Figure: Loi de probabilité normale

Pas d'expression analytique de la primitive de $f_X(x)$ et que la fonction de répartition $F_X(x)$ est donc définie sous forme intégrale.

La variable aléatoire Laplacienne est similaire à la normale, sauf qu'elle décroit plus lentement (on dit que c'est une variable aléatoire à queue lourde).

DEFINITION :Variable Aléatoire Laplacienne

Une variable aléatoire continue X est dite Laplacienne si et seulement si sa densité de probabilité est donnée par :

$$f_X(x) = \frac{1}{\sqrt{2\sigma^2}} \exp\left(-\sqrt{\frac{2}{\sigma^2}}|x-\mu|\right) \qquad x \in \mathbb{R}. \tag{9}$$

Sa fonction de répartition est alors donnée par :

$$F_X(x) = 0.5 * (1 + \mathrm{sign}(x) * \exp\left(-\sqrt{\sigma^2/2}|x-\mu|\right) \qquad x \in \mathbb{R}. \tag{10}$$

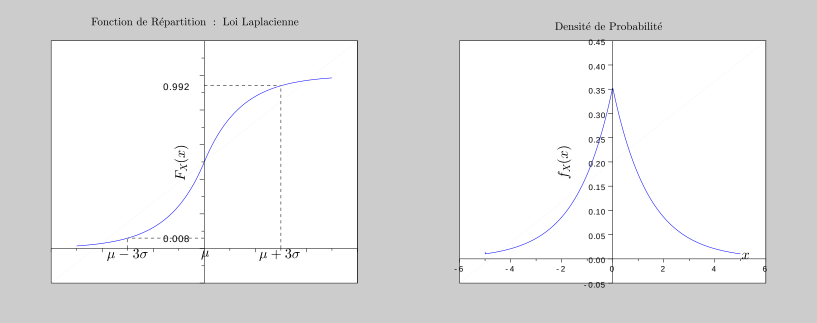


Figure: Loi de probabilité laplacienne

La variable aléatoire de Cauchy est similaire à la normale, sauf qu'elle décroit plus lentement (on dit que c'est une variable aléatoire à queue lourde.

DEFINITION :Variable Aléatoire de Cauchy

Une variable aléatoire continue X est dite de Cauchy si et seulement si sa densité de probabilité est donnée par :

$$f_X(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathbb{R}. \tag{11}$$

Sa fonction de répartition est alors donnée par :

$$F_X(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2} \qquad x \in \mathbb{R}. \tag{12}$$

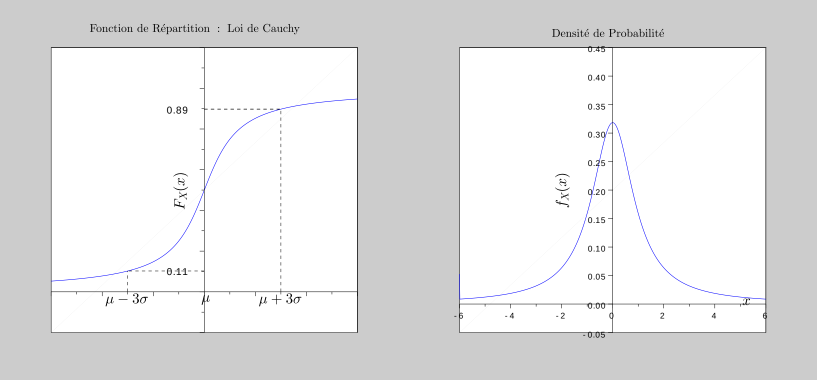


Figure: Loi de probabilité de Cauchy

DEFINITION :v.a. de Rayleigh

Une v.a. de Rayleigh est définie par sa densité de probabilité :

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) & x \geq 0 \\ 0 & x \leq 0 \end{cases} \tag{13}$$

Sa fonction de répartition est donnée par :

$$F_X(x) = \begin{cases} 1 - \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{14}$$

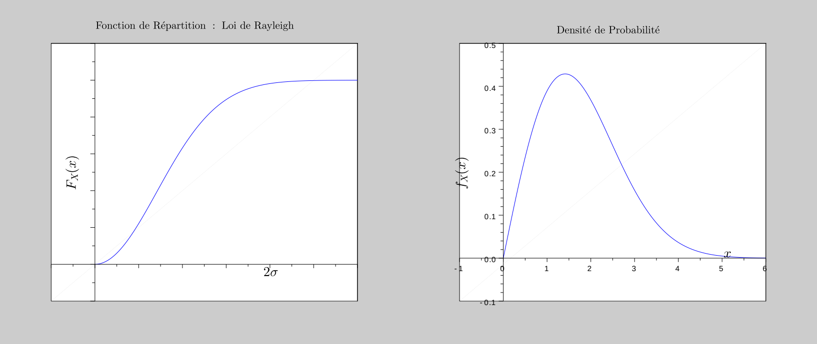


Figure: Loi de probabilité de Rayleigh (variance = 4)

La variance donne une idée de la variabilité des issues possibles autour de la moyenne.
Dans ce cadre, on peut se poser la question suivante :
“Quelle est la probabilité que la réalisation d'une variable aléatoire quelconque soit écartée de la moyenne de plus d'une quantité donnée ?”. En termes mathématiques :

$$\mathbb{P}(|X - \mathrm{E}[X]\,| > \gamma) \leq p.$$

En d'autres termes, si $\mu = 0$ et $\sigma^2 = 1$, quand je dis que la réalisation de la variable aléatoire sera comprise entre -3 et 3, quelle est la probabilité de se tromper ? (ici p).

On peut trouver une borne supérieure de p de la manière suivante :

$$
\begin{aligned}
\mathrm{var}[X] &= \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 f_X(x)\,dx \\
&= \int_{\{x:|x-\mathrm{E}[X]|>\gamma\}} (x - \mathrm{E}[X])^2 f_X(x)\,dx + \underbrace{\int_{\{x:|x-\mathrm{E}[X]|\leq\gamma\}} (x - \mathrm{E}[X])^2 f_X(x)\,dx \geq}\\
&\geq \int_{\{x:|x-\mathrm{E}[X]|>\gamma\}} (x - \mathrm{E}[X])^2 f_X(x)\,dx \\
&\geq \int_{\{x:|x-\mathrm{E}[X]|>\gamma\}} \gamma^2 f_X(x)\,dx \qquad \text{car dans ce domaine, } |x - \mathrm{E}[X]\,| > \gamma \\
&= \gamma^2 \int_{\{x:|x-\mathrm{E}[X]|>\gamma\}} f_X(x)\,dx \\
&= \gamma^2 \mathbb{P}(|X - \mathrm{E}[X]\,| > \gamma)\,,
\end{aligned}
\tag{15}
$$

Inégalité de Tchebychev

$$
\mathbb{P}(|X - \mathrm{E}[X]\,| > \gamma) \leq \frac{\mathrm{var}[X]}{\gamma^2} \tag{16}
$$

Si $\gamma = 3\sigma$, on obtient $\mathbb{P}(|X - \mathrm{E}[X]\,| > 3\sigma) \leq \frac{\sigma^2}{9\sigma^2} \simeq 0.11$.

DEFINITION :Fonction de répartition jointe

Soient deux v.a. continues X et Y, la **Fonction de répartition jointe** est :

$$F_{XY}(x, y) = \mathbb{P}((X \leq x) \cap (Y \leq y))$$

DEFINITION :Densité de probabilité jointe

Soient deux v.a. continues X et Y, la **Densité de probabilité jointe** est :

$$f_{XY}(x, y) = \frac{dF_{XY}(x, y)}{dxdy}$$

En particulier, $f_{XY}(x, y) \geq 0$ et $\int_x \int_y f_{XY}(x, y) = 1$ et
Pour toute région R de l'espace engendré par x et y :

$$\mathbb{P}([X, Y] \in R) = \int \int_R f_{XY}(x, y)\, dxdy$$

DEFINITION :Densité de probabilité marginale

Soient deux v.a. discrètes X et Y, les **masses de probabilité marginales** cont :

$$f_X(x) = \int_{y=-\infty}^{y=+\infty} f_{XY}(x, y)\, dy$$

$$f_Y(y) = \int_{x=-\infty}^{x=+\infty} f_{XY}(x, y)\, dx$$

DEFINITION :Covariance

La *Covariance* de deux v.a. X et Y, notée $\mathrm{cov}[X, Y]$ vaut :

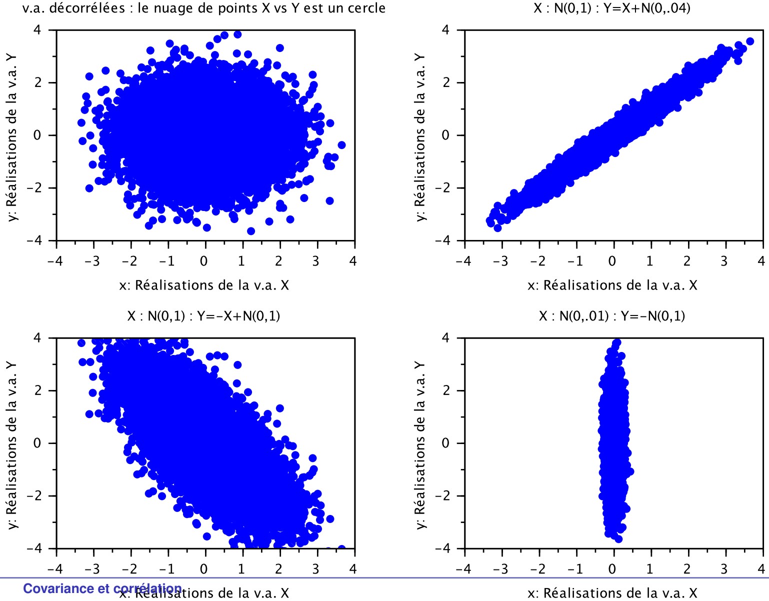
$$\mathrm{cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

Si $\mathrm{cov}[X, Y] = 0$, X et Y sont **décorrélées**.

Relation avec l'indépendance

X et Y indépendants $\Rightarrow \mathrm{cov}[X, Y] = 0$

Mais pas l'inverse !


v.a. décorrélées : le nuage de points X vs Y est un cercle
X : N(0,1) : Y=X+N(0,.04)
X : N(0,1) : Y=−X+N(0,1)
X : N(0,.01) : Y=−N(0,1)
y: Réalisations de la v.a. Y
x: Réalisations de la v.a. X
4
2
0
−2
−4
−3
−1
1
3

1. $X \simeq \mathcal{N}(0,1)$, $Y \simeq \mathcal{N}(0,1)$, X et Y Indépendants :
$$\begin{aligned}\mathrm{cov}[X,Y] = \mathrm{E}[XY] &= \int_x \int_y x.y f_{XY}(x,y)\,dxdy \\ &= \int_x x f_X(x)\,dx \int_y y f_Y(y)\,dy \\ &= \mathrm{E}[X]\,\mathrm{E}[Y] \\ &= 0\end{aligned}$$
2. $X \simeq \mathcal{N}(0,1)$, $Y = X + \mathcal{N}(0,.04)$ notons $Z \simeq \mathcal{N}(0,.04)$:
$\mathrm{cov}[X,Y] = \mathrm{E}[X.(X+Z)] = \mathrm{E}\left[X^2\right] + \mathrm{E}[X]\,\mathrm{E}[Z] = 1$
3. $X \simeq \mathcal{N}(0,1)$, $Y = -X + \mathcal{N}(0,1)$ notons $Z \simeq \mathcal{N}(0,1)$:
$\mathrm{cov}[X,Y] = \mathrm{E}[X.(-X+Z)] = -\mathrm{E}\left[X^2\right] + \mathrm{E}[X]\,\mathrm{E}[Z] = -1$
4. $X \simeq \mathcal{N}(0,.01)$, $Y = -\mathcal{N}(0,1)$:
$\mathrm{cov}[X,Y] = \mathrm{E}[X.Y] = \mathrm{E}[X]\,\mathrm{E}[Y] = 0$

La covariance ne dit pas tout

DEFINITION :Coefficient de corrélation

Le *Coefficient de corrélation* ρ de deux v.a. X et Y de variance non nulle est défini par :

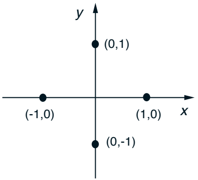
$$\rho \triangleq \frac{\mathrm{cov}[X, Y]}{\sqrt{\mathrm{var}[X]\,\mathrm{var}[Y]}}$$

1. $\rho = 0$
2. $\rho = 1/\sqrt{1.(1 + 0.04)} = 0.98$
3. $\rho = -1/\sqrt{1.(1 + 1)} = -0.71$
4. $\rho = 0$

Exemple

Covariance nulle n'implique pas indépendance

Soient les v.a. X et Y pouvant prendre 4 valeurs équiprobables comme indiqué ci-dessous :



$\mathrm{E}[X] = \mathrm{E}[Y] = 0$
$\Rightarrow \mathrm{cov}[X, Y] = \mathrm{E}[XY] = 1/4.(0 * 1 - 1 * 0 + 0 * (-1) + 1 * 0) = 0$
Or X et Y sont clairement dépendants (si X=1, alors Y=0)
$\triangleleft$

Exemple

pièces truquées

Soient n jets de pièces truquées, les jets sont indépendants.
Soit X le nombre de “pile” et Y le nombre de “face”.
On a toujours $x + y = n$ et donc (par linéarité) $\mathrm{E}[X] + \mathrm{E}[Y] = n$

$$\Rightarrow x - \mathrm{E}[X] = -(y - \mathrm{E}[Y]), \qquad \forall (x, y)$$

$$\Rightarrow \mathrm{cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] = -\mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = -\mathrm{var}[X]$$

$$\Rightarrow \rho(X, Y) = \frac{\mathrm{cov}[X, Y]}{\sqrt{\mathrm{var}[X]\,\mathrm{var}[Y]}} = \frac{-\mathrm{var}[X]}{\sqrt{\mathrm{var}[X]\,\mathrm{var}[X]}} = -1$$

$\triangleleft$

Exemple

Somme de v.a. non indépendantes

Soient n v.a. $X_1, X_2, ..., X_n$:

$$\mathrm{var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{var}[X_i] + 2 \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathrm{cov}\left[X_i, X_j\right]$$

Exemple : sinusoïdes en phase, ou avec des phases aléatoires.

$\triangleleft$

Fonction d'une variable

Par exemple : Puissance $W = \text{cste}\, V^2 = V.I$

Soit $Y = g(X)$:

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx = \int_{-\infty}^{\infty} y f_Y(y)\, dy$$

Comment trouver $f_Y(y)$

Méthode générale : par la fonction de répartition

$$f_Y(x) = \frac{dF_Y(y)}{dy}(y)$$

Exemple

Calcul de la tension à partir de la puissance

Soit $W \sim \mathrm{Un}[1:10] \Rightarrow V$?
En supposant $V > 0$, on a donc $V = \sqrt{W}$, avec $f_W(w) = 1/9, 1 \leq w \leq 10$.

$$F_V(v) = \mathbb{P}(V \leq v) = \mathbb{P}\left(\sqrt{W} < v\right) = \mathbb{P}\left(W < v^2\right) = (v^2 - 1)/9, \quad 1 \leq v \leq \sqrt{10}$$

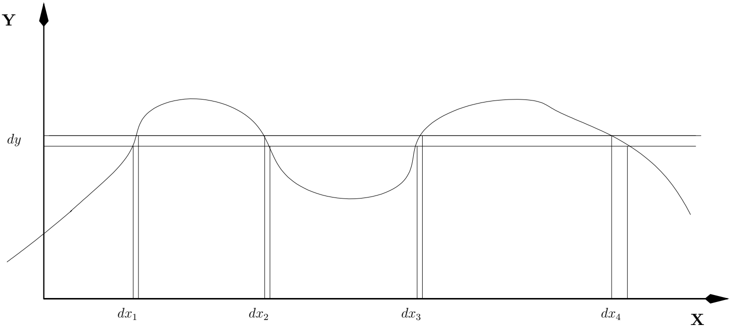
$$f_V(v) = \frac{dF_V(dv)}{dv}(v) = \frac{d(v^2-1)/9}{dv} = 2\frac{v}{9}, \quad 1 \leq v \leq \sqrt{10}$$

On vérifie que $\int_{-\infty}^{\infty} f_V(v)\, dv = 1$

$\lhd$

Cas unidimensionnel

Soit une v.a. **Y**, telle que $\mathbf{Y} = g(\mathbf{X})$. Fonction a priori non bijective



Tronçons i où la fonction est bijective ($\mathbf{Y} = g_i(\mathbf{X}), i = 1, ..n$) : on peut écrire $\mathbf{X} = g_i^{-1}(\mathbf{Y})$ où $g_i^{-1}(.)$ est la fonction inverse de $g_i(.)$.
On a $\mathbb{P}(y < Y \leq y + dy) = f_Y(y)dy = f_Y(y)|dy| = \sum_{x_i|y=g(x_i)} f_X(x_i)|dx_i|$: :

$$f_Y(y) = f_X(g_1^{-1}(y)) \left| \frac{dg_1^{-1}(y)}{dy} \right| + \cdots + f_X(g_n^{-1}(y)) \left| \frac{dg_n^{-1}(y)}{dy} \right| .$$

Exemple

Changement de variable $\mathbf{Y} = \mathbf{X}^2$

Soit $\mathbf{Y} = \mathbf{X}^2$ sur [-1,1]. Dans ce cas, on a deux tronçons :

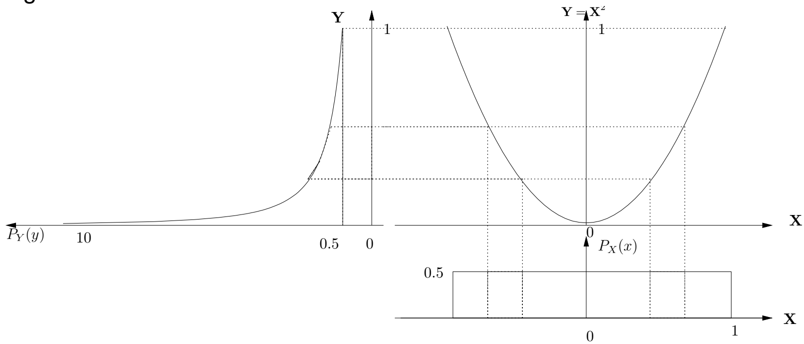
1. sur $x \in [-1, 0]$: $\mathbf{X} = -\sqrt{\mathbf{Y}}$ $(g_1^{-1}(.) = -\sqrt{(.)})$
2. sur $x \in [0, 1]$: $\mathbf{X} = \sqrt{\mathbf{Y}}$ $(g_2^{-1}(.) = \sqrt{(.)})$

On obtient alors

$$\begin{aligned} f_Y(y) &= f_X(g_1^{-1}(y)) \left| \frac{dg_1^{-1}(y)}{dy} \right| + f_X(g_2^{-1}(y)) \left| \frac{dg_2^{-1}(y)}{dy} \right| \\ &= f_X(\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}} \\ &= \frac{1}{2} \frac{1}{\sqrt{y}}. \end{aligned}$$

$\lhd$

La figure ci-dessous illustre ce résultat.



$$\begin{aligned}\mathbf{Y}_1 &= g_1(\mathbf{X}_1, \ldots, \mathbf{X}_n)\\ &\vdots\\ \mathbf{Y}_n &= g_n(\mathbf{X}_1, \ldots, \mathbf{X}_n)\end{aligned}$$

Soient $g_i(.)$ continues et différentiables

$$f_Y(y_1, \ldots, y_n) = f_X(x_1, \ldots, x_n) \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} \right|,$$

on notera que dans ce cas-ci, il est plus compliqué d'écrire l'expression en fonction de g_i^{-1}, mais on aurait plutôt des fonctions de type $\mathbf{X}_i = h_i(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$.

Précision de fabrication en micro-électronique

Dans le processus de fabrication de circuits intégrés, une des parties cruciales est la précision de la lithographie. On peut quantifier cette précision comme étant la déviation en coordonnées horizontales et verticales (x et y) par rapport à l'endroit à graver.
Dans le cas de technologies "70 nm", on peut considérer que les déviations en x et y sont des variables aléatoires indépendantes qui suivent des lois gaussiennes de moyenne nulle et de variance $\sigma^2 = 0.2\text{nm}^2$. La densité de probabilité conjointe des déviations $(\mathbf{X}, \mathbf{Y})$ est donnée par :

$$f_{XY}(x, y) = f_X(x) f_Y(y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

La caractérisation de la précision en x et en y ne répond pas à la question suivante : "quelle est la loi de probabilité de la distance entre le point désiré et le point obtenu par la lithographie ?" Pour obtenir cette loi, on passe des coordonnées cartésiennes aux coordonnées polaires selon la transformation :

$$(\mathbf{X}, \mathbf{Y}) = (\mathbf{R}\cos(\Theta), \mathbf{R}\sin(\Theta)).$$

Le jacobien de la transformation vaut r :

$$\left|\frac{\partial(x,y)}{\partial(r,\theta)}\right| = \begin{vmatrix} \cos(\theta) & \sin(\theta) \\ -r\sin(\theta) & r\cos(\theta) \end{vmatrix} = r,$$

et donc

$$\begin{aligned} f_{\mathbf{R},\Theta}(r,\theta) &= \frac{r}{2\pi\sigma^2} e^{-[(r\cos(\theta))^2+(r\sin(\theta))^2]/2\sigma^2} \\ &= \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2} \end{aligned}$$

Pour trouver les marginales, il faut intégrer sur l'autre variable aléatoire, soit :

$$\begin{aligned} f_\Theta(\theta) &= \int_{r=0}^{\infty} f_{\mathbf{R},\Theta}\, dr \\ &= \int_{r=0}^{\infty} \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2} dr \\ &= \frac{1}{2\pi} \end{aligned}$$

L'angle est donc uniformément distribué sur $[0, 2\pi]$. D'autre part, on en déduit que

$$f_{\mathbf{R}}(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}$$

On en déduit également que les variables aléatoires **R** et Θ sont indépendantes (la loi conjointe est donnée par le produit des lois marginales). La densité de probabilité suivie par la distance **R** est celle d'une variable dite de Rayleigh.

Somme de variables aléatoires indépendantes

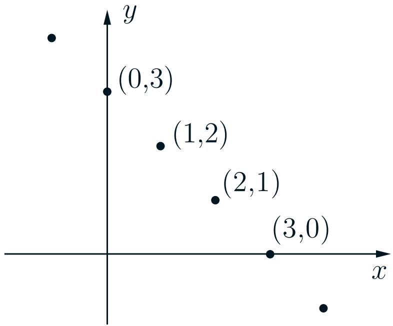
Cas Discret

Soient X et Y, deux v.a. indépendantes de masse $p_X(x)$ et $p_Y(y)$, à valeurs entières.
Soit $W = X + Y$, alors, pour tout entier w :

$$
\begin{aligned}
p_W(w) &= \mathbb{P}(X + Y = w) \\
&= \sum_{(x,y):x+y=w} \mathbb{P}(X = x \text{ et } Y = y) \\
&= \sum_x \mathbb{P}(X = x \text{ et } Y = w - x) \\
&= \sum_x p_X(x)\, p_Y(w - x) \\
&= p_X(x) * p_Y(y)
\end{aligned}
$$

où $p_X(x) * p_Y(y)$ est la convolution de $p_X(x)$ et de $p_Y(y)$

La masse $p_W(3)$ est la probabilité que $X + Y = 3$: c'est la probabilité de toutes les paires (x, y) telles que $x + y = 3$



et $p_{X,Y}(x, 3 - x) = p_X(x)\, p_Y(3 - x)$

Trois Tireurs de Penalty

Somme de v.a. continues

Soient deux v.a. X et Y continues et $W = X + Y$. Pour déterminer $f_W(w)$, nous allons dériver $F_W(w)$.

$$\begin{aligned} F_W(w) &= \mathbb{P}(W \leq w) \\ &= \mathbb{P}(X + Y \leq w) \\ &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{w-x} f_X(x)\, f_Y(y)\, dydx \\ &= \int_{x=-\infty}^{\infty} f_X(x) \left[\int_{y=-\infty}^{w-x} f_Y(y)\, dy \right] dx \\ &= \int_{x=-\infty}^{\infty} f_X(x)\, F_Y(w-x)\, dx \end{aligned}$$

On en déduit :

$$
\begin{aligned}
f_W(w) &= \frac{F_W\,(w)}{dw} \\
&= \frac{d}{dw}\int_{x=-\infty}^{\infty} f_X(x)\,F_Y\,(w-x)\,dx \\
&= \int_{x=-\infty}^{\infty} f_X(x)\,\frac{dF_Y\,(w-x)}{dw}dx \\
&= \int_{x=-\infty}^{\infty} f_X(x)\,f_Y(w-x)\,dx
\end{aligned}
$$

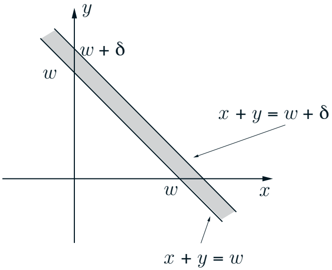


Illustration de la convolution

$$\mathbb{P}(w \leq X + Y \leq w + \delta) = f_W(w)\,.\delta$$

$$\begin{aligned} f_W(w)\,.\delta &= \int_{x=-\infty}^{\infty} f_X(x) \int_{y=w-x}^{w-x+\delta} f_Y(y)\,dydx \\ &\simeq \int_{x=-\infty}^{\infty} f_X(x)\, f_Y(w-x)\,\delta dx \end{aligned}$$

Convolution avec l'applet Java

http://pages.jh.edu/ signals/convolve/