

0. --- TABLE DES MATIÈRES

1	Introduction	4
1.1	Lien entre l'expérience aléatoire et son modèle	5
1.2	Opérations sur les ensembles	5
2	Probabilités Élémentaires	6
2.1	Axiomes fondamentales de la probabilité	6
2.2	Principe d'équiprobabilité	7
2.2.1	Ensemble fini	7
2.2.2	Ensemble infini dénombrable	7
2.2.3	Ensemble infini non-dénombrable	7
2.3	Diagramme en arbre	8
2.4	Permutations	8
2.5	Combinaisons	9
2.5.1	Triangle de Pascal	9
2.5.2	Binôme de Newton	10
2.6	Permutation d'objets semblables	10
2.7	Équivalence	11
2.8	Réccurence	12
2.9	Probabilité conditionnelles	13
2.10	Probabilités totales	13
2.11	Notion d'indépendance	14
3	Variables Aléatoires	16
3.1	Fonction de répartition	16
3.2	Fonction de masse	17
3.3	Fonction de densité	18
3.4	Règles de calcul fondamentale	18
3.5	Liens entre les différentes fonctions	19
3.6	Fonction conditionnelle	20
3.7	Médiane et quantile	21
3.8	Lois de probabilités discrètes	22
3.8.1	Loi de Bernoulli	22
3.8.2	Loi binomiale	22
3.8.3	Loi géométrique	23
3.8.4	Loi de Poisson	25
3.8.5	Approximation par une loi de Poisson	25
3.9	Loi des probabilités continues	25
3.9.1	Loi uniforme continue	25

3.9.2	Loi exponentielle	26
3.9.3	Loi gamma	26
3.9.4	Loi normale	27
3.9.5	Loi normale centrée réduite	27
3.10	Fonction d'une variable aléatoire	28
3.10.1	X et Y sont des variables aléatoires discrètes	28
3.10.2	X et Y sont des variables aléatoires mixtes	29
3.10.3	X et Y sont des variables aléatoires continues	29
3.11	Espérance mathématique	29
3.12	Espérance conditionnelle	31
3.13	Variance	31
3.14	Inégalité de Markov	32
3.15	Inégalité de Bienaymé-Tchebychev	32
3.16	Fonction caractéristique	33
3.17	Fiabilité	34
3.17.1	Durée de vie moyenne	35
4	Vecteurs aléatoires	36
4.1	Vecteur aléatoire discret	37
4.1.1	Fonction de masse conjointe	37
4.1.2	Fonction de masse marginale	37
4.1.3	Fonction de masse conditionnelle	37
4.2	Vecteur aléatoire continu	37
4.2.1	Fonction de densité conjointe	37
4.2.2	Fonction de densité marginale	38
4.2.3	Fonction de densité conditionnelle	38
4.3	Fonction de répartition conjointe	38
4.4	Fonction de répartition marginale	39
4.5	Indépendance dans un vecteur	39
4.6	Espérance conditionnelle	40
4.7	Variance conditionnelle	41
4.8	Espérance d'une transformation	41
4.9	Corrélation et covariance	42
4.10	Loi binormale	43
4.11	Estimation	43
4.11.1	Estimateur constant	44
4.11.2	Estimateur linéaire	44
4.11.3	Estimateur non linéaire	44
4.11.4	Estimateur d'une binormale	44
4.12	Combinaison linéaire	45
4.13	Variables aléatoires indépendantes et identiquement distribuées	46
4.14	Loi faible des grands nombres	47
4.15	Loi forte des grands nombres	47
4.16	Théorème central limite	47

5	Processus stochastique	49
5.1	Moyenne d'un processus stochastique	50
5.2	Fonction d'autocorrélation	50
5.3	Fonction d'autocovariance	50
5.4	Processus stochastique stationnaire au sens large	51
5.5	Chaîne de Markov	51
5.5.1	Représentation graphique	51
5.5.2	Représentation matricielle	52
5.5.3	Probabilité de transition en n étapes	53
5.6	Processus de comptage	53
5.7	Processus de Poisson	53
5.7.1	Temps d'arrivée	53
5.8	Marche aléatoire	54
5.9	Processus de Wiener	55
5.9.1	Fonction d'autocovariance	56
6	Statistique descriptive	57
6.1	Représentation numérique	57
6.1.1	Moyenne de l'échantillon	57
6.1.2	Médiane de l'échantillon	57
6.1.3	Variance de l'échantillon	57
6.1.4	Étendue de l'échantillon	58
7	Inférence statistique	59
7.1	Biais d'un estimateur	60
7.2	Erreur quadratique moyenne	60
7.3	Recherche d'un estimateur	62
7.3.1	Méthode du maximum de vraisemblance	62
7.3.2	Méthode des moments	63
7.4	Intervalle de confiance	64
7.5	moyenne μ si $\mathcal{N}(\mu, \sigma^2)$ et σ^2 inconnue	65
7.6	Test d'ajustement du χ^2	66
7.7	Tests d'hypothèses	67

1. INTRODUCTION

Définition. Une expérience est *aléatoire* si un observateur peut la répéter dans les mêmes conditions, mais sans pouvoir en prédire le résultat.

Définition. Un *espace échantillon* est un ensemble S des résultats possibles.

Remarque. Un espace d'échantillon peut être *qualitatif* ou *quantitatif*, ainsi que *discret*, *continu* ou *mixte*. Il peut aussi être *dénombrable* ou *non-dénombrable*.

Définition. Un *évènement* A est un sous-ensemble de S d'intérêt à l'observateur.

Définition. Un *évènement élémentaire* A est un résultat particulier, c'est-à-dire, un élément de S .

Exemple 1.1. On observe le résultat du lancer de deux pièces de monnaie. On note P comme un lancer pile et F comme un lancer face. L'ensemble est donc

$$S = \{PP, FF, PF, FP\}$$

avec chaque résultat ayant 25 % d'arriver. L'ensemble S est qualitatif, soit pile ou face, et discret.

Exemple 1.2. On observe la somme obtenue lors du lancer de deux dés à 6 faces. L'ensemble de résultat possible est

$$S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

Il est quantitatif et discret.

Exemple 1.3. On compte le nombre de lancers d'une pièce pour obtenir une première fois un pile. L'espace échantillon est

$$S = \{1, 2, 3, 4, 5, \dots\},$$

car il est possible qu'une grande quantité de lancer est effectuée avant d'obtenir un pile. L'ensemble S est quantitatif, discret et infini dénombrable.

Exemple 1.4. On mesure le temps d'attente à l'arrêt d'autobus. L'espace échantillon est

$$S = [0, \infty[.$$

L'ensemble S est quantitatif, continu et infini non-dénombrable.

Exemple 1.5. On mesure le temps d'attente à l'arrêt d'autobus ainsi que le nombre de personnes en file à l'arrivée de l'autobus. L'espace échantillon est

$$S = T \times U,$$

avec

$$T = [0, \infty[$$

et

$$U = \{1, 2, 3, 4, 5, \dots\}.$$

L'ensemble S est quantitatif et mixte. Un exemple d'évènement élémentaire peut être un couple tel que $(4.25 \text{ s}, 4)$.

1.1 Lien entre l'expérience aléatoire et son modèle

Définition. La *fréquence relative* f_A d'un évènement A est le rapport entre le nombre d'observations n_A de l'évènement et le nombre n de répétition de l'expérience, c'est-à-dire

$$f_A = \frac{n_A}{n}.$$

Définition. La limite lorsque l'expérience est répétée infiniment est la *probabilité* de l'évènement A , dénotée

$$\mathbb{P}[A] = \lim_{n \rightarrow \infty} f_A.$$

1.2 Opérations sur les ensembles

Soit deux ensembles A et B tel que $A, B \subset S$. La figure 1a montre une intersection tandis que la figure 1b montre une union entre A et B . La figure 1c montre le complément de A et la figure 1d montre l'exclusion de deux ensembles.

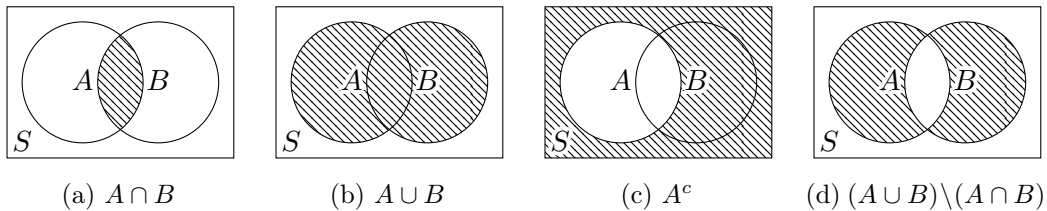


FIGURE 1 – Opérations d'ensembles démontrées sur des diagrammes de Venn

[2. _____] PROBABILITÉS ÉLÉMENTAIRES

2.1 Axiomes fondamentales de la probabilité

Axiome 1. La probabilité d'un évènement A est plus grand ou égal à 0, c'est-à-dire

$$\mathbb{P}[A] \geq 0,$$

pour tout $A \in S$.

Axiome 2. La probabilité de l'espace d'échantillon S est 1, c'est-à-dire

$$\mathbb{P}[S] = 1.$$

Axiome 3. La probabilité d'un évènement A ou d'un évènement B est équivalent à la somme de leur probabilité, c'est-à-dire

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B],$$

si $A \cap B = \emptyset$.

En général,

$$\mathbb{P}\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} \mathbb{P}[A_k],$$

si $A_i \cap A_j = \emptyset, \forall i, j$.

Théoreme 2.1. $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$.

Démonstration. On sait que $A \cup A^c = S$ et $A \cap A^c = \emptyset$. Hors, $\mathbb{P}[A \cup A^c] = \mathbb{P}[S] \Leftrightarrow \mathbb{P}[A] + \mathbb{P}[A^c] = 1$, car $A \cap A^c = \emptyset$. En réarrangeant, on obtient que $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$. \square

Théoreme 2.2. $\mathbb{P}[A] \leq 1$.

Démonstration. On sait que $\mathbb{P}[A^c] \geq 0$ et $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$. En réarrangeant, on obtient que $\mathbb{P}[A] \leq 1$. \square

Théoreme 2.3. $\mathbb{P}[\emptyset] = 0$.

Démonstration. On sait que $S^c = \emptyset$. Par conséquent, $\mathbb{P}[\emptyset] = 1 - \mathbb{P}[S] = 1 - 1 = 0$. \square

Théoreme 2.4. $A \subset B \Rightarrow \mathbb{P}[A] \leq \mathbb{P}[B]$.

Démonstration. La différence entre A et B est $A^c \cap B$ de sorte qu'on peut écrire $B = A \cup (A^c \cap B)$. Par conséquent, $\mathbb{P}[B] = \mathbb{P}[A] + \mathbb{P}[A^c \cap B] \Leftrightarrow \mathbb{P}[B] - \mathbb{P}[A] = \mathbb{P}[A^c \cap B] \geq 0$, car $A \cup (A^c \cap B) = B$. En réarrangeant, on obtient $\mathbb{P}[A] \leq \mathbb{P}[B]$. \square

Théoreme 2.5. $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$.

2.2 Principe d'équiprobabilité

Définition. Une expérience aléatoire est dite *équiprobable* si $\mathbb{P}[e_i] = c$, où $0 \leq c \leq 1$, pour tout $e_i \in S$.

2.2.1 Ensemble fini

Théoreme 2.6. Soit $S = \{e_1, e_2, \dots, e_n\}$ un ensemble d'événements fini et équiprobable, alors

$$\mathbb{P}[e_1] = \mathbb{P}[e_2] = \dots = \mathbb{P}[e_n] = \frac{1}{n}.$$

Corollaire. Soit $A \subset S$, alors $\mathbb{P}[A] = \frac{n_A}{n}$.

2.2.2 Ensemble infini dénombrable

Théoreme 2.7. Soit $S = \{e_1, e_2, \dots\}$ un ensemble d'événements infini et dénombrable. Alors, les résultats d'une expérience aléatoire ne peuvent pas être équiprobables.

Démonstration. On suppose que $\mathbb{P}[e_i] = \epsilon$ pour tout $i = \{1, 2, \dots\}$, où $0 < \epsilon < 1$. Par conséquent, $\mathbb{P}[S] = \mathbb{P}[e_1] + \mathbb{P}[e_2] + \dots = \infty$, ce qui est en contradiction avec les axiomes de la probabilité. D'une manière similaire, si $\mathbb{P}[e_i] = 0$ pour tout $i = \{1, 2, \dots\}$, alors $\mathbb{P}[S] = 0$. Ce qui est aussi en contradiction avec les axiomes de la probabilités. Par conséquent, un ensemble infini et dénombrable ne peut pas être équiprobable. \square

2.2.3 Ensemble infini non-dénombrable

Théoreme 2.8. Soit $S = [a, b]$, où $a < b$, un ensemble d'événements infini, non-dénombrable, et équiprobable, alors

$$\mathbb{P}[A] = \frac{d - c}{b - a},$$

où $A = [c, d] \in S$.

Corollaire. Soit $A \subset S$ un événement élémentaire, alors $\mathbb{P}[A] = 0$.

Exemple 2.1. On calcule la somme de deux dés lancés. Quelle est la probabilité d'obtenir chaque somme possible ?

L'espace échantillon est $S = L \times L$, où $L = \{1, 2, 3, 4, 5, 6\}$ est le résultat possible d'un dé. On peut écrire $S = \{(1, 1), (1, 2), \dots, (6, 6)\}$. On suppose qu'il y a équiprobabilité de sorte que la probabilité d'obtenir un événement élémentaire est $1/36$.

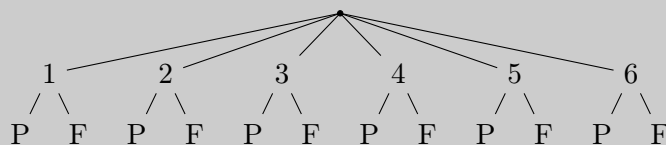
Hors, certaines des sommes sont dupliquées de sorte que la probabilité d'obtenir une somme particulière est donnée par la table suivante.

A	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}[A]$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

2.3 Diagramme en arbre

Exemple 2.2. On lance un dé, puis une pièce de monnaie. Combien de résultats est-il possible ?

On peut représenté la situation par l'arbre suivant.



Par le principe de multiplication, il y a $6 \cdot 2 = 12$ possibilités.

2.4 Permutations

Définition. Une *permutation* correspond à un choix de k objets parmi n objets distincts. Le choix se fait sans remise et dans un ordre spécifique.

La table 1 résume le principe d'une permutation à l'aide d'une pige d'objet. À chaque pige, la quantité d'objets diminue de 1.

TABLE 1 – Pige d'objets

Choix d'objet	1	2	...	k
Objets restants	n	$n - 1$...	$n - (k - 1)$

À l'aide du principe de multiplication, il y a $n \cdot (n - 1) \cdots (n - k + 1)$ combinaisons. On dénote le nombre de permutations sans remise de n éléments par

$$\mathcal{P}_n^k = \frac{n!}{(n - k)!}.$$

Lorsqu'il y a remise, le nombre de permutations de n éléments est n^k .

Exemple 2.3. On dispose de 10 composantes dont 4 défectueuses. On pige 3 composantes au hasard et sans remise. Quelle est la probabilité d'obtenir uniquement des composantes non défectueuses ?

On suppose qu'il y a équiprobabilité du choix des 3 composantes. La probabilité d'obtenir 3 composantes non défectueuses F est

$$\mathbb{P}[F] = \frac{6 \cdot 5 \cdot 4}{10 \cdot 9 \cdot 8} = \frac{\mathcal{P}_6^3}{\mathcal{P}_{10}^3} = \frac{1}{6}.$$

2.5 Combinaisons

Définition. Une *combinaison* correspond au choix de k objets parmi n objets distincts. Le choix se fait sans remise et sans ordre spécifique.

Soit C_k^n le nombre de combinaisons. Le nombre de permutations est égal au nombre de combinaisons multiplié par le nombre d'arrangement possible $k!$, c'est-à-dire

$$\mathcal{P}_n^k = C_n^k \cdot k!,$$

et en réarrangeant, on obtient

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Exemple 2.4. Combien de codes de deux lettres peut-on former à partir du mot OUI ?
Avec ordre, il y a $\mathcal{P}_3^2 = 6$ permutations et sans ordre, il y a $C_3^2 = 3$ combinaisons.

Exemple 2.5. Quel est la probabilité de gagner le gros lot à la 6/49 ?

Sans ordre, il y a 1 seul cas favorable et C_{49}^6 cas possibles. Par conséquent, la probabilité de gagner G est

$$\mathbb{P}[G] = \frac{1}{C_{49}^6} = \frac{1}{13\,983\,816}.$$

Avec ordre, il y a $6!$ cas favorables et \mathcal{P}_{49}^6 cas possibles de sorte que la probabilité est

$$\mathbb{P}[G] = \frac{6!}{\mathcal{P}_{49}^6} = \frac{720}{10\,068\,347\,520}.$$

2.5.1 Triangle de Pascal

Une combinaison peut se représenter avec le triangle de Pascal comme le montre la figure 2.

$$\begin{array}{ccccccc}
 & & & & C_0^0 & & \\
 & & & & C_1^0 & C_1^1 & \\
 & & C_2^0 & & C_2^1 & C_2^2 & \\
 & C_3^0 & & C_3^1 & & C_3^2 & C_3^3 \\
 C_4^0 & & C_4^1 & & C_4^2 & & C_4^3 & C_4^4
 \end{array}$$

FIGURE 2 – représentation du triangle de Pascal

Théoreme 2.9. $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$

Démonstration. La démonstration est triviale et est laissée au lecteur. □

2.5.2 Binôme de Newton

La combinaison est souvent appliquée dans le cas de la puissance d'un binôme tel que

$$(a + b)^n = \sum_{k=0}^n \mathcal{C}_n^k \cdot a^k \cdot b^{n-k} = \sum_{k=0}^n \binom{n}{k} \cdot a^k \cdot b^{n-k},$$

mais elle est plus souvent utilisée avec la deuxième notation.

2.6 Permutation d'objets semblables

Exemple 2.6. Combien y a-t-il d'ordres possibles des lettres «ppfff» ?

Soit 5 cases distinctes représentant l'ordre d'une pige dans les lettres. Il faut choisir les cases où mettre les «p», soit

$$\mathcal{C}_5^2 = \frac{5!}{2!3!} = 10$$

ordres possibles.

Si on échange un «p» et un «f» dans une séquence A , on obtient une séquence B différente de A . Si on échange un «p» avec une autre «p» dans une séquence A , on retrouve la même séquence A .

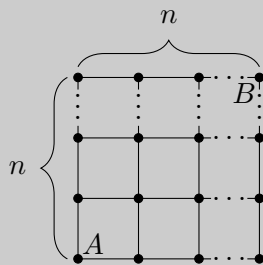
En analysant la formule, le facteur $5!$ représente le nombre d'ordres si toutes les lettres étaient différentes. Le facteur $2!$ représente les «p» s'ils étaient différents et le facteur $3!$ représente le nombre des «f» s'ils étaient différents.

En général, avec n objets objets comprenant n_1 objets de classes 1, n_2 objets de classe 2, \dots , n_k objets à la classe k , on a

$$\frac{n!}{n_1!n_2! \cdots n_k!}$$

ordres possibles.

Exemple 2.7. Soit le graphe $n \times n$ suivant.



Combien de chemins existe entre A et B suivant, s'il est seulement permis d'aller vers la droite ou vert le haut ?

Exemple 2.7 (suite). On suppose que tous les chemins possibles sont équiprobables. Peu importe le chemin, il faut avancer $n - 1$ vers la droite et $n - 1$ vers le haut pour un total de $2n - 2$ mouvement.

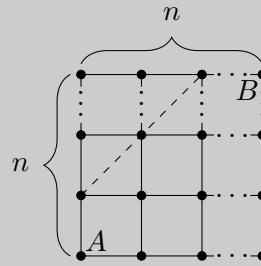
Il suffit de calculer le nombre de permutations de ces mouvements sachant qu'il y a des objets semblables. Le nombre de chemins est

$$\frac{(2n - 2)!}{(n - 1)!(n - 1)!},$$

ce qui est équivalent à C_{2n-2}^{n-1} .

2.7 Équivalence

Exemple 2.8. Soit le graphe $n \times n$ suivant.



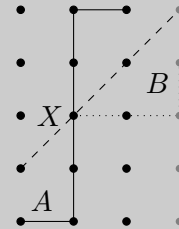
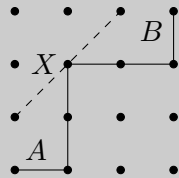
Quel est la probabilité qu'un chemin choisi au hasard, allant seulement vers la droite ou le haut, ne traverse pas les noeuds au-dessus de la diagonale entre A et B ?

Les «bons» chemins ne passent pas par la ligne critique en pointillée et les «mauvais» chemins passent par la ligne critique. La probabilité d'un bon chemin B peut s'écrire en fonction de son complément, soit

$$\mathbb{P}[B] = 1 - \mathbb{P}[M] = 1 - \frac{n_m}{C_{n-1}^{2n-2}},$$

où M est un mauvais chemin et n_m est le nombre de mauvais chemins.

Soit un mauvais chemin M . On définit le point X comme étant le premier point au-dessus de la diagonale que M atteint. On applique une transformation miroir au chemin suivant X comme la prochaine figure.



Exemple 2.8 (suite). Il s'avère que tous les mauvais chemins dans le graphe $n \times n$ correspondent à un chemin unique dans le graphe $(n-1) \times (n+1)$. C'est-à-dire la transformation est bijective.

Par conséquent, le problème est équivalent à trouver le nombre de chemins dans un graphe $(n-1) \times (n+1)$. D'une manière similaire à l'exemple 2.7, ce nombre de chemins est donné par

$$\frac{(2n-2)!}{(n-2)!n!},$$

ce qui est équivalent à C_{n-2}^{2n-2} .

Par conséquent, la probabilité d'avoir un bon chemin au hasard est

$$\mathbb{P}[B] = 1 - \frac{C_{2n-2}^{n-2}}{C_{2n-2}^{n-1}} = \frac{1}{n}.$$

2.8 Réccurrence

Exemple 2.9. Il est possible de résoudre le problème à l'exemple 2.8 à l'aide de la récursion.

En effet, le nombre de «bons» chemins $b_{i,j}$ à partir d'un noeud est la somme des «bons» chemins des noeuds à droite et en haut, c'est-à-dire

$$b_{i,j} = b_{i-1,j} + b_{i,j-1},$$

où i est le nombre de noeuds restants vers la droite et j la nombre de noeuds restants vers le haut.

On sait que $b_{0,0} = 1$, car il ne reste plus de noeud à parcourir fois rendu à B . Aussi, $b_{0,j} = 1$, car il est seulement possible de se rendre à B en allant vers le haut. Sur la diagonale, on a $b_{i,i} = b_{i-1,i}$, car on ne peut pas aller vers le haut. Par conséquent, on obtient le système d'équations à reccurrence suivant,

$$b_{i,j} = \begin{cases} 0, & \text{si } i = 0, j = 0 \\ 1, & \text{si } i = 0 \\ b_{i-1,j}, & \text{si } i = j \\ b_{i-1,j} + b_{i,j-1} & \text{sinon} \end{cases}$$

Pour une grille 4×4 , on peut calculer le nombre de «bons» chemins en développant la récurrence afin d'obtenir $b_{3,3} = 5$ chemins.

2.9 Probabilité conditionnelles

Définition. Une *probabilité conditionnelle* $\mathbb{P}[A|B]$ est la probabilité qu'un évènement A se réalise, si B s'est réalisé, soit

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Propriété 1. $\mathbb{P}[A \cap B] = \mathbb{P}[A|B] \cdot \mathbb{P}[B]$.

Propriété 2. $\mathbb{P}[A|B] = \mathbb{P}[B|A] = 0$, si $A \cap B = \emptyset$.

Propriété 3. $\mathbb{P}[A|B] \neq \mathbb{P}[B|A]$, en général.

Propriété 4. $\mathbb{P}[A|S] = \mathbb{P}[A]$, $A \in S$.

Remarque. La dernière propriété résulte que toutes probabilités peut s'exprimer sous la forme d'une probabilité conditionnelle.

Exemple 2.10. Soit le lancement d'un dé avec les évènements $A = \{5, 6\}$ et $B = \{2, 4, 6\}$, alors $\mathbb{P}[A] = 1/3$ et

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[\{6\}]}{\mathbb{P}[B]} = \frac{1}{3}.$$

Si $A = \{6\}$, alors $\mathbb{P}[A] = 1/6$ et

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[\{6\}]}{\mathbb{P}[B]} = \frac{1}{3}.$$

Exemple 2.11. On pige sans remise 3 composantes non défectueuses parmi 10 composantes, donc 4 sont défectueuses. Soit A_i le i^e composante non défectueuses.

$$\mathbb{P}[A_1 \cap A_2 \cap A_3] = \mathbb{P}[A_3|A_1 \cap A_2] \mathbb{P}[A_2|A_1] \mathbb{P}[A_1] = \frac{4}{8} \cdot \frac{5}{9} \cdot \frac{6}{10}$$

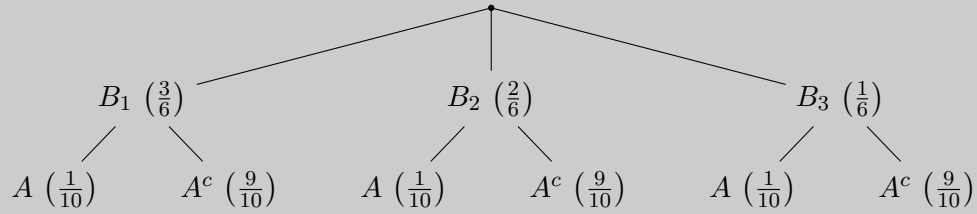
2.10 Probabilités totales

Définition. Les évènements B_1, B_2, \dots, B_n forment une *partition* si les évènements $B_i \cap B_j = \emptyset$, $\forall i \neq j$, et $B_1 \cup B_2 \cup \dots \cup B_n = S$.

Théoreme 2.10. Si $B_1, B_2, \dots, B_n \in S$ forment une partition de S , alors la probabilité d'un évènement $A \in S$ est donnée par

$$\mathbb{P}[A] = \sum_{i=1}^n \mathbb{P}[A \cap B_i] = \sum_{i=1}^n \mathbb{P}[A|B_i] \mathbb{P}[B_i].$$

Exemple 2.12. Soit la partition B_1, B_2 et B_3 représenté dans l'arbre suivant.



À partir de l'arbre, il est facile de déterminer les probabilités conditionnelles de A et A^c . Par exemple, $\mathbb{P}[A|B_1] = 3/6 \cdot 1/10 = 1/30$ et $\mathbb{P}[A^c|B_2] = 2/6 \cdot 9/10 = 3/10$.

Théorème 2.11 (Règle d'inversion). $\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\mathbb{P}[B]}{\mathbb{P}[A]}$

Théorème 2.12 (Règles de Bayes). $\mathbb{P}[B_j|A] = \frac{\mathbb{P}[A|B_j]\mathbb{P}[B_j]}{\sum_{i=1}^n \mathbb{P}[A|B_i]\mathbb{P}[B_i]}$, où B_i et B_j sont des partitions.

Exemple 2.13. On dépiste le cancer du poumon dans une clinique. On sait que

- 25 % des individus sont fumeurs
- 75 % des individus sont non fumeurs
- 10 % des fumeurs développent un cancer
- 1 % des non fumeurs développent un cancer

On détecte un cancer des poumons chez un individu sélectionné au hasard pour dépistage. Quelle est la probabilité que ça soit un fumeur ?

Soit B les fumeurs, B^c les non fumeurs, A la présence de cancer du poumon et A^c son absence. Par conséquent, on cherche $\mathbb{P}[B|A]$. Avec les données, on sait que $\mathbb{P}[A|B] = 0.10$, $\mathbb{P}[A|B^c] = 0.01$, $\mathbb{P}[B] = 0.25$ et $\mathbb{P}[B^c] = 0.75$, de sorte que le théorème de Bayes nous donne

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\mathbb{P}[B]}{\mathbb{P}[A|B]\mathbb{P}[B] + \mathbb{P}[A|B^c]\mathbb{P}[B^c]} \approx 0.769.$$

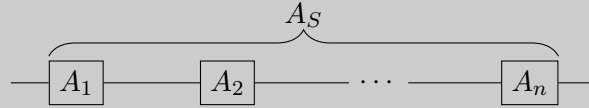
2.11 Notion d'indépendance

Définition. On dit que les événements A et B sont *indépendants* si la réalisation d'un n'affecte pas l'autre, soit

$$\mathbb{P}[A|B] = \mathbb{P}[A] \Leftrightarrow \mathbb{P}[B|A] = \mathbb{P}[B].$$

Théorème 2.13. Si A et B sont indépendants, alors $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$.

Exemple 2.14. Soit un système en série ayant n composantes comme à la figure suivante.



Un système en série fonctionne si et seulement si tous ses composantes fonctionnent. Quelle est la probabilité que le système fonctionne si chaque composante a une probabilité de 75 % de fonctionner ?

Soit l'ensemble échantillon $S = \{F, D\}$ avec F une composante qui fonctionne et D une composante défectueuse. De plus, on définit $A_i = \{F\} \in S$ comme étant la composante i qui fonctionne et $A_i^c = \{D\} \in S$ comme étant la composante i qui est défectueuse.

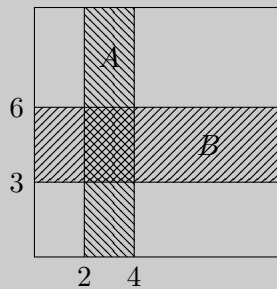
On suppose que les composantes fonctionnent et tombent en panne indépendamment les uns des autres. Le système fonctionne si

$$A_S = \bigcap_{i=1}^n A_i,$$

de sorte que

$$\mathbb{P}[A_S] = \mathbb{P}\left[\bigcap_{i=1}^n A_i\right] = \prod_{i=1}^n \mathbb{P}[A_i] = 0.75^n.$$

Exemple 2.15. On génère un point au hasard dans le carré $S = [0, 10] \times [0, 10]$. On définit A comme ayant l'abscisse du point entre 2 et 4, et B comme ayant l'ordonnée du point entre 3 et 6. Est-ce que A et B sont indépendants ?



Puisqu'il y a équiprobabilité continue, alors les probabilités sont données par le rapport entre l'aire de A ou B sur S , alors $\mathbb{P}[A] = 2/10$ et $\mathbb{P}[B] = 3/10$. De plus, $\mathbb{P}[A \cap B] = 6/100$. Puisque $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$, alors les événements A et B sont indépendants.

[3. VARIABLES ALÉATOIRES]

Définition. Une *variable aléatoire* correspond à une expérience aléatoire dont les résultats sont quantitatifs.

- On peut décrire une variable aléatoire par sa
- fonction de répartition
 - fonction de masse (variable aléatoire discrète)
 - fonction de densité (variable aléatoire continue)

Exemple 3.1. Soit X le nombre de défauts de soudure d'un transistor à 3 pattes. Quel est l'espace échantillon de X ?

L'espace échantillon est $S_X = \{0, 1, 2, 3\}$, alors X est une variable aléatoire discrète.

Exemple 3.2. Soit X un nombre réel choisie au hasard dans l'intervalle $[0, 2]$. Quel est l'espace échantillon de X ?

L'espace échantillon est $S_X = [0, 2]$, alors X est une variable aléatoire continue.

Exemple 3.3. Soit X le temps d'attente d'un client au guichet automatique. Quel est l'espace d'échantillon de X ?

L'espace échantillon est $S_X = 0 \cup]0, \infty[$. Il y a une partie discrète et continue, alors X est une variable aléatoire mixte.

3.1 Fonction de répartition

Définition. Une *fonction de répartition* $F_X(x)$ est égal à la probabilité qu'une variable aléatoire X soit plus petite qu'une valeur x , c'est-à-dire

$$F_X(x) = \mathbb{P}[X \leq x], \quad \forall x \in \mathbb{R}.$$

Propriété 1. $0 \leq F_X(x) \leq 1$.

Propriété 2. $\lim_{x \rightarrow -\infty} F_X(x) = 0$.

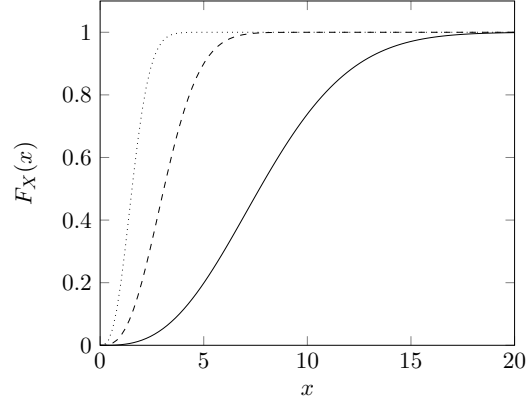
Propriété 3. $\lim_{x \rightarrow \infty} F_X(x) = 1$.

Propriété 4 (non décroissance). $x_0 < x_1 \Leftrightarrow F_X(x_0) < F_X(x_1)$.

Propriété 5 (continuité à droite). $F_X(x^+) = F_X(x)$.

Il en résulte que toutes fonctions de répartition $F_X(x)$ sont croissantes, mais peuvent contenir des discontinuités. Elles peuvent représenter des variables aléatoires discrètes, continues ou mixtes. La figure 3 est un exemple de la fonction de répartition d'une distribution de Maxwell-Boltzmann pour certains paramètres différents.

FIGURE 3 – $F_X(x)$ de certaines distributions de Maxwell-Boltzmann



3.2 Fonction de masse

Définition. Une *fonction de masse* $p_X(x_k)$ est égal à la probabilité qu'une variable aléatoire discrète X soit prenne une valeur $x_k \in S_X$, c'est-à-dire

$$p_X(x_k) = \mathbb{P}[X = x_k],$$

où $S_X = \{x_1, x_2, \dots, x_n | x_1 < x_2 < \dots < x_n\}$.

Propriété 1. $0 \leq p_X(x_k) \leq 1$, si $x_k \in S_X$.

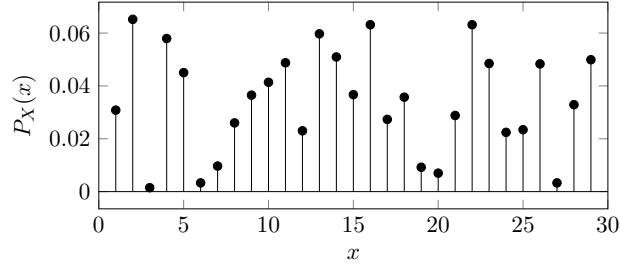
Propriété 2. $p_X(x_k) = 0$, si $x_k \notin S_X$.

Propriété 3. $\sum_{a < x_k \leq b} p_X(x_k) = \mathbb{P}[a < X \leq b]$.

Propriété 4. $\sum_{k=1}^{\infty} p_X(x_k) = 1$.

Remarque. Une fonction de masse est nulle en tout point sauf aux valeurs discrètes possibles. De plus, la somme de toutes les valeurs discrètes donne 1. La figure 4 montre un exemple d'une fonction de masse.

FIGURE 4 – exemple d’une fonction de masse



3.3 Fonction de densité

Définition. Une *fonction de densité* $f_X(x)$ est égal à la probabilité de qu’une variable aléatoire X soit autour d’une valeur x , c’est-à-dire

$$f_X(x) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{P} \left[x - \frac{\epsilon}{2} \leq X \leq x + \frac{\epsilon}{2} \right],$$

où ϵ est positif.

Propriété 1. $f_X(x) \geq 0$.

Propriété 2. $\int_{-\infty}^{\infty} f_X(x) dx = 1$.

Remarque. La fonction de densité n’est pas une probabilité à cause de sa définition infinitésimale. Par conséquent, l’inégalité $0 \leq f_X(x) \leq 1$ n’est pas valide.

3.4 Règles de calcul fondamentale

Théoreme 3.1. $\mathbb{P}[a < X \leq b] = F_X(b) - F_X(a)$

Démonstration. Soit les ensembles $A = \{X \leq a\}$, $B = \{X \leq b\}$ et $C = \{a < X \leq b\}$, avec $a < b$, comme à la figure 5.

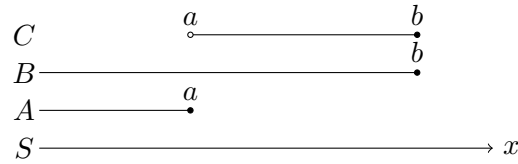


FIGURE 5 – représentation sur une droite numérique

On sait que $A \cap C = \emptyset$ et $A \cup C = B$. Par conséquent, $\mathbb{P}[B] = \mathbb{P}[A] + \mathbb{P}[C] \Leftrightarrow \mathbb{P}[C] = \mathbb{P}[B] - \mathbb{P}[A] = F_X(b) - F_X(a)$. \square

Théoreme 3.2. $\mathbb{P}[X = x] = F_X(x) - F_X(x^-)$.

Démonstration. Soit $a = x - \epsilon$ et $b = x$, où ϵ est positif. Selon le théorème 3.1, on a

$$\mathbb{P}[x - \epsilon < X \leq x] = F_X(x) - F_X(x - \epsilon).$$

En prenant la limite lorsque $\epsilon \rightarrow 0$, on obtient

$$\lim_{\epsilon \rightarrow 0} \mathbb{P}[x - \epsilon < X \leq x] = F_X(x) - \lim_{\epsilon \rightarrow 0} F_X(x - \epsilon).$$

Avec $F_X(x^-) = \lim_{\epsilon \rightarrow 0} F_X(x - \epsilon)$ et $(x - \epsilon < X \leq x) \equiv (X = x)$ lorsque $\epsilon \rightarrow 0$, on obtient

$$\mathbb{P}[X = x] = F_X(x) - F_X(x^-). \quad \square$$

3.5 Liens entre les différentes fonctions

Toutes variables aléatoires peuvent être décrites par une fonction de répartition. Lorsque la variable aléatoire est continue, alors elle peut aussi être décrite par une fonction de densité de probabilité. Si la variable aléatoire est discrète, alors elle peut être décrite par une fonction de masse.

Théorème 3.3. $f_X(x) = \frac{d}{dx}F_X(x)$ si X est continue.

Théorème 3.4. $F_X(x) = \int_{-\infty}^x f_X(t) dt$ si X est continue.

Théorème 3.5. $F_X(x) = \sum_{x_k \leq x} P_X(x_k)$ si X est discrète.

Théorème 3.6. $p_X(x_k) = \begin{cases} F_X(x_1) & \text{si } k = 1, \\ F_X(x_k) - F_X(x_{k-1}) & \text{si } k \neq 1. \end{cases}$

Exemple 3.4. Soit X un nombre réel choisi au hasard dans l'intervalle $[0, 2]$. Quelle est la fonction de répartition de X ?

On sait que $F_X(x) = \mathbb{P}[X \leq x] = x/2$ lorsque $0 \leq x \leq 2$, $F_X(x) = 0$ si $x < 0$ et $F_X(x) = 1$ si $x > 2$. On obtient donc la fonction de répartition

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0, \\ \frac{x}{2} & \text{si } 0 \leq x \leq 2, \\ 1 & \text{si } 2 \leq x. \end{cases}$$

Exemple 3.5. Soit X le temps d'attente d'un client au guichet automatique. On suppose que 10 % des visites sont sans attente. Quel est la fonction de répartition ?

La variable aléatoire est mixte. On sait que $F_X(0) = 1/10$ et que $F_X(x) = 0$ si $x < 0$. La fonction de répartition est

$$F_X(k) = \begin{cases} 0 & \text{si } x < 0, \\ \frac{x+1}{x+10} & \text{si } x \geq 0. \end{cases}$$

Exemple 3.6. Soit X le nombre de défauts de soudure d'un transistor à 3 pattes. La fonction de masse de X est donnée par

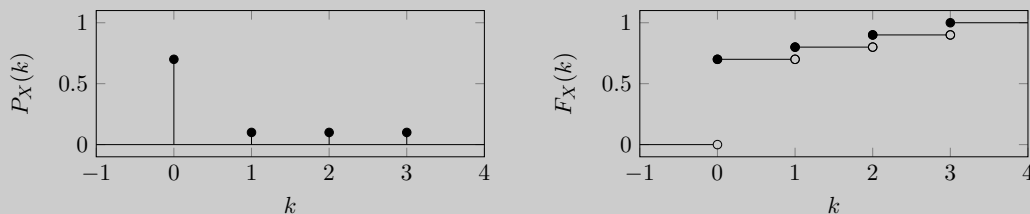
$$p_X(k) = \begin{cases} 0.7 & \text{si } k = 0, \\ 0.1 & \text{si } k = 1, \\ 0.1 & \text{si } k = 2, \\ 0.1 & \text{si } k = 3. \end{cases}$$

Quelle est la fonction de répartition de $P_X(k)$?

On applique le théorème 3.5 pour obtenir la fonction de répartition. Lorsque $x < 0$, alors $F_X(x) = 0$. Lorsque $0 \leq x < 1$, alors $F_X(x) = 0.7$. Lorsque $1 \leq x < 2$, alors $F_X(x) = 0.8$. En continuant, on obtient

$$F_X(k) = \begin{cases} 0.0 & \text{si } k < 0, \\ 0.7 & \text{si } 0 \leq k < 1, \\ 0.8 & \text{si } 1 \leq k < 2, \\ 0.9 & \text{si } 2 \leq k < 3, \\ 1.0 & \text{si } 3 \leq k. \end{cases}$$

Les fonctions de masse et de répartition sont montrées dans les figures suivantes.



3.6 Fonction conditionnelle

Définition. Une *fonction de répartition conditionnelle* $F_X(x|A)$ est égal à la probabilité qu'une variable aléatoire X soit plus petite ou égal à une valeur x sachant un certain événement A , c'est-à-dire

$$F_X(x|A) = \frac{\mathbb{P}[\{X \leq x\} \cap A]}{\mathbb{P}[A]}.$$

Définition. Une *fonction de masse conditionnelle* $p_X(x_k|A)$ est égal à la probabilité qu'une variable aléatoire discrète prenne une valeur $x_k \in S_X$ sachant un certain événement A , c'est-à-dire

$$p_X(x_k|A) = \frac{\mathbb{P}[\{X = x_k\} \cap A]}{\mathbb{P}[A]},$$

où $S_X = \{x_1, x_2, \dots, x_n | x_1 < x_2 < \dots < x_n\}$.

Définition. Une *fonction de densité conditionnelle* $f_X(x|A)$ est égal à la probabilité qu'une variable aléatoire X soit autour d'une valeur x sachant un certain événement A , c'est-à-dire

$$f_X(x|A) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{P}\left[x - \frac{\epsilon}{2} \leq X \leq x + \frac{\epsilon}{2}\right],$$

où ϵ est positif.

Théoreme 3.7. $f_X(x|A) = \frac{d}{dx} F_X(x|A)$ si X est continue.

Théoreme 3.8. $p_X(x_k|A) = \begin{cases} \frac{p_X(x_k)}{\mathbb{P}[A]} & \text{si } x_k \in A, \\ 0 & \text{si } x_k \notin A, \end{cases}$ si X est discrète.

3.7 Médiane et quantile

Définition. La médiane d'une variable aléatoire X continue est le nombre réel $x_{1/2}$ qui permet de couper l'ensemble des valeurs en deux parties égales, c'est-à-dire

$$F_X(x_{1/2}) = 1/2.$$

Définition. Le *quantile* d'ordre p d'une variable aléatoire X continue est le nombre réel x_p tel que

$$F_X(x_p) = p.$$

Remarque. La médiane est le quantile d'ordre $1/2$.

Exemple 3.7. Dans une certaine population, la taille X d'un adulte choisi au hasard possède la fonction de répartition

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1.2, \\ 1.5x - 1.8 & \text{si } 1.2 \leq x < 1.7, \\ 0.5x - 0.1 & \text{si } 1.7 \leq x < 2.2, \\ 1 & \text{si } 2.2 \leq x. \end{cases}$$

Calculer la médiane et le quantile d'ordre 95.

Puisque $F_X(1.7) = 0.75$, alors le quantile d'ordre 95 est dans la tranche $1.7 \leq x < 2.2$. Il suffit de résoudre $0.95 = 0.5x_{0.95} - 0.1$ et on obtient que $x_{0.95} = 2.1$.

3.8 Lois de probabilités discrètes

3.8.1 Loi de Bernoulli

Définition. Une variable aléatoire X suivant une *loi de Bernoulli*, dénotée

$$X \sim \mathcal{B}(1, p),$$

est le résultat d'une expérience aléatoire pouvant être soit un «échec», dénotée $X = 0$, ou un «succès», dénotée $X = 1$. La probabilité du succès est donnée par p .

Propriété 1. $S_X = \{0, 1\}$.

Propriété 2. $p_X(k) = \begin{cases} q & \text{si } k = 0, \\ p & \text{si } k = 1, \\ 0 & \text{sinon.} \end{cases}$

3.8.2 Loi binomiale

Définition. Une variable aléatoire X suivant une *loi binomiale*, dénotée

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

est le nombre de succès obtenu de n expériences de Bernoulli indépendantes où la probabilité d'un succès individuel est donnée par p .

Propriété 1. $S_X = \{0, 1, \dots, n\}$.

Propriété 2. $p_X(k) = \begin{cases} C_n^k p^k q^{n-k} & \text{si } k = 0, 1, \dots, n, \\ 0 & \text{sinon.} \end{cases}$

Démonstration. Soit A_i le i -ème succès et A_i^c le i -ème échec dans une expérience aléatoire à n essais. La probabilité d'obtenir k succès est donnée par

$$\mathbb{P} \left[\underbrace{A_1 \cap A_2 \cap \dots \cap A_k}_{k \text{ fois}} \cap \underbrace{A_{k+1}^c \cap A_{k+1}^c \cap \dots \cap A_n^c}_{n-k \text{ fois}} \right],$$

où $k = 0, 1, 2, 3, \dots, n$.

Hors, la probabilité d'un succès est $\mathbb{P}[A_i] = p$ et celle d'un échec est $\mathbb{P}[A_i^c] = q$, où $q = 1 - p$. Puisque chaque essai est indépendant des autres, la probabilité peut s'écrire

$$\underbrace{\mathbb{P}[A_1] \mathbb{P}[A_2] \dots \mathbb{P}[A_k]}_{k \text{ fois}} \cdot \underbrace{\mathbb{P}[A_{k+1}^c] \mathbb{P}[A_{k+2}^c] \dots \mathbb{P}[A_n^c]}_{n-k \text{ fois}} = p^k q^{n-k}.$$

De plus, les succès peuvent être à n'importe quel essai. Par conséquent, il faut choisir k essais parmi les n essais de sorte que la probabilité d'obtenir k succès en tout est $p_X(k) = \mathbb{P}[X = k] = C_n^k p^k q^{n-k}$. \square

Proposition. $\mathbb{P}[S_X] = 1$.

Démonstration. Par définition, on a

$$\mathbb{P}[S_X] = \sum_{k=0}^n \mathcal{C}_n^k p^k q^{n-k}$$

Selon le théorème binomial, soit

$$(a + b)^n = \sum_{k=0}^n \mathcal{C}_n^k p^k q^{n-k},$$

et que $q = 1 - p$, on peut simplifier de sorte à obtenir

$$\mathbb{P}[S_X] = (p + q)^n = 1^n = 1.$$

□

Exemple 3.8. Vous achetez un billet de 6/49 à chaque semaine depuis vos 18 ans. Quelle est la probabilité de gagner le gros lot au plus tard à votre 98^e anniversaire ?

Ce problème se décrit à l'aide d'une loi binomiale, soit

$$X \sim \mathcal{B}\left(n, \frac{1}{13\,983\,816}\right),$$

où n est le nombre d'essais et X le nombre de gros lots gagnés.

Puisqu'on joue chaque semaine de 18 ans à 98 ans, un total de $n = 52 \cdot (98 - 18) = 4160$ billets sont achetés. La probabilité d'obtenir que des échecs est donnée par

$$\left(1 - \frac{1}{13\,983\,816}\right)^{4160} = \left(\frac{13\,983\,815}{13\,983\,816}\right)^{4160},$$

de sorte que la probabilité d'obtenir au moins 1 billet gagnant est donnée par le complément, soit

$$\mathbb{P}[X \geq 1] = 1 - \left(\frac{13\,983\,815}{13\,983\,816}\right)^{4160} \approx 0.000\,297.$$

3.8.3 Loi géométrique

Définition. Une variable aléatoire X suivant une *loi géométrique*, dénotée

$$X \sim \mathcal{G}(p),$$

est le nombre d'essais nécessaires avant d'obtenir un succès. La probabilité d'un succès est donnée par p .

Propriété 1. $S_X = \{1, 2, \dots\}$.

Propriété 2. $p_X(k) = \begin{cases} q^{k-1}p & \text{si } k = 1, 2, \dots, \\ 0 & \text{si } k = 0. \end{cases}$

Démonstration. Soit A_i le i -ième succès et A_i^c le i -ième échec dans une expérience aléatoire où il faut k essais pour obtenir un succès. La probabilité d'obtenir le succès au k -ième essai est donnée par

$$\mathbb{P} \left[\underbrace{A_1^c \cap A_2^c \cap \dots \cap A_{k-1}^c}_{k-1 \text{ fois}} \cap A_k \right],$$

où $k = 1, 2, \dots$

Hors, la probabilité d'un succès est $\mathbb{P}[A_i] = p$ et celle d'un échec est $\mathbb{P}[A_i^c] = q$, où $q = 1 - p$. Puisque chaque essai est indépendant des autres, la probabilité peut s'écrire

$$p_X(k) = \underbrace{\mathbb{P}[A_1^c] \mathbb{P}[A_2^c] \dots \mathbb{P}[A_{k-1}^c]}_{k-1 \text{ fois}} \cdot \mathbb{P}[A_k] = q^{k-1}p. \quad \square$$

Propriété 3. $F_X(n) = 1 - q^n$ où $n = 1, 2, \dots$

Démonstration. Selon une variante du théorème 3.5, on a

$$F_X(n) = \sum_{k=1}^n p_X(k) = \sum_{k=1}^n q^{k-1}p = p \sum_{k=1}^n q^{k-1}.$$

On remarque que la somme est une série géométrique, alors

$$\sum_{k=1}^n q^{k-1} = \frac{1 - q^n}{1 - q}.$$

Puisque $q = 1 - p \Leftrightarrow p = 1 - q$, on obtient

$$F_X(n) = (1 - q) \frac{1 - q^n}{1 - q} = 1 - q^n. \quad \square$$

Propriété 4 (absence de mémoire). $\mathbb{P}[X > k + j | X > j] = \mathbb{P}[X > k]$ où $k, j \in \mathbb{Z}$.

Démonstration. On sait que

$$\mathbb{P}[X > k + j | X > j] = \frac{\mathbb{P}[\{X > k + j\} \cap \{X > j\}]}{\mathbb{P}[X > j]}.$$

Puisque $k \in \mathbb{Z}$, alors $\{X > k + j\} \cap \{X > j\} = \{X > k + 1\}$ de sorte que

$$\mathbb{P}[X > k + j | X > j] = \frac{\mathbb{P}[X > k + j]}{\mathbb{P}[X > j]}.$$

Avec le complément de la propriété 3, on a

$$\mathbb{P}[X > k + j | X > j] = \frac{q^{k+j}}{q^j} = q^k = \mathbb{P}[X > k]. \quad \square$$

Proposition. $\mathbb{P}[S_X] = 1$.

Démonstration. En utilisant à nouveau la série géométrique, on a

$$\mathbb{P}[S_X] = \sum_{k=1}^{\infty} p_X(k) = p \sum_{k=1}^{\infty} q^{k-1} = p \frac{1}{1-q} = (1-q) \frac{1}{1-q} = 1,$$

puisque $q = 1 - p \Leftrightarrow p = 1 - q$. □

3.8.4 Loi de Poisson

Définition. Une variable aléatoire suivant une *loi de Poisson* de paramètre α , dénotée

$$X \sim \text{Poi}(\alpha),$$

est équivalente à suivre une loi Binomiale telle que

$$X \sim \lim_{n \rightarrow \infty} \mathcal{B}\left(n, \frac{\alpha}{n}\right),$$

où α est un nombre positif.

Propriété 1. $S_X = \{0, 1, \dots\}$.

Propriété 2. $p_X(k) = \begin{cases} \frac{\alpha^k}{k!} e^{-\alpha} & \text{si } k = 0, 1, \dots, \\ 0 & \text{sinon.} \end{cases}$

Proposition. $\mathbb{P}[S_X] = 1$.

Démonstration. En utilisant le développement en série de la fonction exponentielle, on a

$$\mathbb{P}[S_X] = \sum_{k=0}^{\infty} p_X(k) = e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = e^{-\alpha} e^{\alpha} = 1. \quad \square$$

3.8.5 Approximation par une loi de Poisson

Théoreme 3.9. Soit une variable aléatoire $X \sim \mathcal{B}(n, p)$. Si p est près de 0, alors

$$X \approx \text{Poi}(np).$$

Remarque. En général, l'approximation est bonne si $n \geq 30$ et $p \leq 0.05$.

3.9 Loi des probabilités continues

3.9.1 Loi uniforme continue

Définition. Une variable aléatoire X suivant une loi uniforme continue, dénotée

$$X \sim \mathcal{U}(a, b),$$

est un nombre réel choisie avec équiprobabilité dans l'intervalle $[a, b]$.

Propriété 1. $S_X = [a, b]$.

Propriété 2. $f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b, \\ 0 & \text{sinon.} \end{cases}$

Propriété 3. $F_X(x) = \begin{cases} 0 & \text{si } x < a, \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b, \\ 1 & \text{si } b < x. \end{cases}$

3.9.2 Loi exponentielle

Définition. Une variable aléatoire X suivant une *loi exponentielle* de paramètre $\lambda > 0$, dénotée

$$X \sim \text{Exp}(\lambda),$$

décrit le temps entre les événements d'un *processus de Poisson*.

Propriété 1. $S_X = [0, \infty[$.

Propriété 2. $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$

Propriété 3. $F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$

Propriété 4 (absence de mémoire). $\mathbb{P}[X > s + t | X > t] = \mathbb{P}[X > t]$ où $s, t \in [0, \infty[$.

3.9.3 Loi gamma

Définition. Une variable aléatoire X suivant une *loi gamma*, dénotée

$$X \sim \text{Gam}(\alpha, \lambda),$$

où $\alpha > 0$ et $\lambda > 0$, est continue et positive.

Propriété 1. $S_X = [0, \infty[$.

Propriété 2. $f_X(x) = \begin{cases} \frac{(\lambda x)^{\alpha-1} \lambda e^{-\lambda x}}{\Gamma(\alpha)} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$

Propriété 3. $F_X(x) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k e^{-\lambda x}}{(n-1)!}$ si $\alpha = n = 1, 2, \dots$.

Proposition. $\mathbb{P}[S_X] = 1$.

Démonstration. Il suffit d'intégrer la fonction de densité de probabilité, soit

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_0^{\infty} \frac{(\lambda x)^{\alpha-1} \lambda e^{-\lambda x}}{\Gamma(\alpha)} dx = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} (\lambda x)^{\alpha-1} \lambda e^{-\lambda x} dx,$$

et en posant $y = \lambda x$, on obtient

$$\frac{1}{\Gamma(\alpha)} \int_0^{\infty} y^{\alpha-1} e^{-y} dy = \frac{1}{\Gamma(\alpha)} \Gamma(\alpha) = 1.$$

□

3.9.4 Loi normale

Définition. Une variable aléatoire suivant une *loi normale* X de moyenne μ et de variance σ^2 , dénotée

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

suit une distribution décrite par une fonction gaussienne.

Propriété 1. $S_X = \mathbb{R}$.

Propriété 2. $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}$.

Propriété 3. $F_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} dx$.

3.9.5 Loi normale centrée réduite

Définition. Une variable aléatoire X suivant une *loi normale centrée réduite*, est une variable aléatoire suivant une loi normale de moyenne $\mu = 0$ et variance $\sigma^2 = 1$, c'est-à-dire

$$X \sim \mathcal{N}(0, 1).$$

Propriété 1. $S_X = \mathbb{R}$.

Propriété 2. $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}$.

Propriété 3. $F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{1}{2}x^2\right\} dx$.

Proposition. $\mathbb{P}[S_X] = 1$.

Démonstration. Par définition, on a

$$I = \mathbb{P}[S_X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}x^2\right\} dx,$$

donc

$$I^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}x^2\right\} dx \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}y^2\right\} dy$$

Puisque x et y sont des variables différentes, on a

$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(x^2 + y^2)\right\} dx dy,$$

et coordonnées polaires,

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} \exp\left\{-\frac{1}{2}r^2\right\} r dr d\theta,$$

de sorte qu'en posant $u = r^2$, on peut obtenir

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \exp \left\{ -\frac{1}{2} e^{-u} \right\} \Big|_0^\infty d\theta = \frac{1}{2\pi} \int_0^{2\pi} d\theta = 1$$

Par conséquent, on a $I = 1$. □

Théoreme 3.10. Si $Y = \frac{X - \mu}{\sigma^2}$, où $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $Y \sim \mathcal{N}(0, 1)$.

3.10 Fonction d'une variable aléatoire

Définition. Une fonction d'une variable aléatoire X est une transformation $g(X)$ sur toutes les valeurs de X .

Exemple 3.9. Soit X est la valeur d'amplitude d'un signal au temps t . Le signal numérisé Y peut s'écrire

$$Y = \underbrace{\text{signe}(X) \cdot \Delta \cdot \text{part} \left(\frac{|X|}{\Delta} + \frac{1}{2} \right)}_{g(X)},$$

où Δ est le pas de quantification.

3.10.1 X et Y sont des variables aléatoires discrètes

Exemple 3.10. Soit $X \sim \mathcal{B}(2, 1/4)$ et $Y = g(X)$, où $g(x) = (x - 1)^2$. Quelle est la fonction de masse de Y ?

L'espace échantillon de X est

$$S_X = \{0, 1, 2\}$$

de sorte qu'en appliquant $g(x)$ sur tout $x \in S_X$, on obtient

$$S_Y = \{1, 0, 1\}.$$

À la lumière de ce résultat, on peut facilement définir la fonction de masse de Y en partie, soit

$$f_Y(k) = \begin{cases} f_X(1) & \text{si } k = 0, \\ f_X(0) + f_X(2) & \text{si } k = 1. \end{cases}$$

3.10.2 X et Y sont des variables aléatoires mixtes

Exemple 3.11. Soit $X \sim \mathcal{N}(0, 1)$ et

$$Y = \begin{cases} -1 & \text{si } X < -1/2, \\ 0 & \text{si } -1/2 \leq X \leq 1/2, \\ 1 & \text{si } X > 1/2. \end{cases}$$

Quelle est la fonction de masse de Y ?

Il suffit de calculer les probabilités des conditions de la fonction par partie, c'est-à-dire

$$p_X(k) = \begin{cases} F_X(-1/2) & \text{si } k = -1, \\ F_X(1/2) - F_X(-1/2) & \text{si } k = 0, \\ 1 - F_X(1/2) & \text{si } k = 1, \\ 0 & \text{sinon.} \end{cases}$$

3.10.3 X et Y sont des variables aléatoires continues

Exemple 3.12. Soit $X \sim \mathcal{U}(-1, 2)$ et $Y = X^2$. Quelle est la fonction de répartition de Y ?

Par définition, on a

$$F_Y(Y) = \mathbb{P}[Y \leq y] = \mathbb{P}[X^2 \leq y] = \mathbb{P}[-\sqrt{y} \leq X \leq \sqrt{y}] = F_X(\sqrt{y}) - F_X(-\sqrt{y}),$$

où $0 \leq \sqrt{y} \leq 2$. Puisque l'uniforme n'est pas symétrique, on a

$$F_Y(y) = \begin{cases} 0 & \text{si } y < 0, \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}) & \text{si } 0 \leq y \leq 1, \\ F_X(\sqrt{y}) - F_X(1) & \text{si } 1 \leq y \leq 4, \\ 1 & \text{si } 4 < y. \end{cases}$$

3.11 Espérance mathématique

Définition. L'*espérance* d'une variable aléatoire X , dénotée $\mathbb{E}[X]$, est la somme des valeurs possibles de X pondérées par leur probabilité, c'est-à-dire

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} x_k p_X(x_k)$$

si X est discrète, et

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

si X est continue.

Exemple 3.13. Quelle est l'espérance d'un lancer d'un dé ?

On calcule l'espérance d'une variable discrète, soit

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

Exemple 3.14. Soit $X \sim \text{Poi}(\alpha)$. Quelle est l'espérance de X ?

Par définition, on calcule l'espérance avec

$$E[X] = \sum_{k=0}^{\infty} k \cdot p_X(k) = \sum_{k=0}^{\infty} k \cdot \frac{e^{-\alpha} \alpha^k}{k!} = \sum_{k=1}^{\infty} k \cdot \frac{e^{-\alpha} \alpha^k}{k!},$$

car le premier terme à $k = 0$ est nul. Par conséquent,

$$E[X] = \sum_{k=1}^{\infty} \frac{e^{-\alpha} \alpha^k}{(k-1)!} = e^{-\alpha} \sum_{k=1}^{\infty} \frac{\alpha^k}{(k-1)!} = e^{-\alpha} \sum_{i=0}^{\infty} \frac{\alpha^{i+1}}{i!} = e^{-\alpha} \alpha \sum_{i=0}^{\infty} \frac{\alpha^i}{i!}.$$

Hors, la somme est le développement en séries de la fonction exponentielle, alors

$$E[X] = e^{-\alpha} \alpha e^{\alpha} = \alpha.$$

Exemple 3.15. Soit $X \sim \text{Exp}(\lambda)$. Quelle est l'espérance de X ?

Par définition, on calcule l'espérance avec

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx.$$

En posant $u = x$ et $dv = \lambda e^{-\lambda x} dx$, on obtient

$$E[X] = -x e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^{\infty} \lambda e^{-\lambda x} dx = \frac{1}{\lambda},$$

car l'aire sous la fonction de densité de probabilités est égal à 1.

Propriété 1. $E[c] = c$, où c est une constante.

Propriété 2. $E[aX + b] = aE[X] + b$, où a et b des constantes.

Théoreme 3.11. Si $Y = g(X)$, où X est une variable aléatoire et $g(X)$ une fonction, alors

$$E[Y] = \sum_{k=1}^{\infty} g(x_k) p_X(x_k)$$

si X est discrète, et

$$E[Y] = \int_{-\infty}^{\infty} g(x)f_X(x) dx$$

si X est continue.

3.12 Espérance conditionnelle

Définition. L'*espérance conditionnelle* d'une variable aléatoire X sachant un événement A , dénotée $E[X|A]$, est la somme des valeurs possibles de X pondérées par leur probabilité conditionnelle, c'est-à-dire

$$E[X|A] = \sum_{k=1}^{\infty} x_k p_X(x_k|A)$$

si X est discrète, et

$$E[X|A] = \int_{-\infty}^{\infty} x f_X(x|A) dx$$

si X est continue.

Théoreme 3.12. Soit X une variable aléatoire. Si B_1, B_2, \dots, B_n forment une partition de S_X , alors

$$E[X] = \sum_{i=1}^n E[X|B_i] \mathbb{P}[B_i].$$

Exemple 3.16. Soit X une variable aléatoire mixte. Quelle est la forme de l'espérance de X ?

On définit C comme l'événement où X prend une valeur continue et D lorsque X prend une valeur discrète. Par conséquent, on a une partition de S_X de sorte à avoir

$$E[X] = \underbrace{E[X|C]}_f \mathbb{P}[C] + \underbrace{E[X|D]}_{\Sigma} \mathbb{P}[D].$$

3.13 Variance

Définition. La *variance* d'une variable aléatoire X , dénotée $\text{Var}[X]$, est une mesure de la dispersion des valeurs de X par rapport à la moyenne. Elle est définie comme

$$\text{Var}[X] = E[(X - E[X])^2].$$

Propriété 1. $\text{Var}[c] = 0$, où c est une constante.

Propriété 2. $\text{Var}[aX + b] = a^2 \text{Var}[X]$, où a et b des constantes.

Propriété 3. $\text{Std}[X] = \sqrt{\text{Var}[X]}$.

Propriété 4. $\text{Var}[X|A] = \text{E}[X^2|A] - \text{E}[X|A]^2$.

Théoreme 3.13. $\text{Var}[X] = \text{E}[X]^2 - \text{E}[X^2]$.

Démonstration. Par définition, on a

$$\text{Var}[X] = \text{E}\left[(X - \text{E}[X])^2\right] = \text{E}\left[X^2 - 2X\text{E}[X] + \text{E}[X]^2\right].$$

Puisque l'espérance est un opérateur linéaire, on a

$$\text{Var}[X] = \text{E}[X^2] - 2\text{E}[X\text{E}[X]] + \text{E}[\text{E}[X]^2].$$

Hors, $\text{E}[X]$ et $\text{E}[X]^2$ sont des constantes, alors

$$\text{Var}[X] = \text{E}[X^2] - 2\text{E}[X]\text{E}[X] + \text{E}[X]^2 = \text{E}[X]^2 - \text{E}[X^2]. \quad \square$$

Exemple 3.17. Soit $X \sim \mathcal{B}(1, p)$. Quelle est la variance de X ?

On sait que l'espérance de X est

$$\text{E}[X] = 0 \cdot (1 - p) + 1 \cdot p = p,$$

et celle de X^2 est

$$\text{E}[X^2] = 0^2 \cdot (1 - p) + 1^2 \cdot p = p.$$

Par conséquent, la variance est

$$\text{Var}[X] = p - p^2 = p(1 - p).$$

3.14 Inégalité de Markov

Théoreme 3.14. Soit X une variable aléatoire prenant des valeurs non négatives, alors

$$\mathbb{P}[X \geq a] \leq \frac{\text{E}[X]}{a},$$

pour tout $a > 0$.

3.15 Inégalité de Bienaymé-Tchebychev

Théoreme 3.15. Soit X une variable aléatoire de moyenne $\mu = \text{E}[X]$ et variance $\sigma^2 = \text{Var}[X]$ définies, alors

$$\mathbb{P}[|X - \mu| \geq a] \leq \frac{\sigma^2}{a^2},$$

pour tout $a > 0$.

Exemple 3.18. Soit le lancer d'une pièce de monnaie avec $X \sim \mathcal{B}(n, 1/2)$. On calcule la moyenne avec

$$\mathbb{E} \left[\frac{X}{n} \right] = \frac{1}{n} \mathbb{E}[X] = \frac{1}{n} \cdot n \cdot \frac{1}{2} = \frac{1}{2}$$

et la variance avec

$$\text{Var} \left[\frac{X}{n} \right] = \frac{1}{n^2} \text{Var}[X] = \frac{1}{n^2} \cdot n \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4n}.$$

Selon l'inégalité de Bienaymé-Tchebychev, on a

$$\mathbb{P} \left[\left| \frac{X}{n} - \frac{1}{2} \right| \geq 0.01 \right] \leq \frac{1/4n}{(0.01)^2} = \frac{10000}{4n} = \frac{2500}{n}.$$

3.16 Fonction caractéristique

Définition. Une *fonction caractéristique*, dénoté $\phi_X(\omega)$, est l'espérance d'une variable aléatoire X tel que

$$\phi_X(\omega) = \mathbb{E} [e^{j\omega X}],$$

où j est le nombre imaginaire tel que $j^2 = -1$.

Propriété 1. $\phi_X(0) = 1$.

Propriété 2. $\mathbb{E}[X^n] = (-j)^n \left[\frac{d^n}{d\omega^n} \phi_X(\omega) \right]_{\omega=0}$.

Exemple 3.19. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$. On peut montrer que

$$\phi_X(\omega) = \exp \left(j\omega\mu - \frac{1}{2}\omega^2\sigma^2 \right).$$

Soit $Y = aX + b$. Quelle est la fonction caractéristique de Y ? Par définition, on a

$$\phi_Y(\omega) = \mathbb{E} [e^{j\omega Y}] = \mathbb{E} [e^{j\omega(aX+b)}] = e^{j\omega b} \mathbb{E} [e^{j\omega aX}] = e^{j\omega b} \cdot \exp \left(j\omega a\mu + \frac{1}{2}\omega^2 a^2 \sigma^2 \right)$$

de sorte à obtenir

$$\phi_Y(\omega) = \exp \left[j\omega \underbrace{(a\mu + b)}_{\mu_Y} - \frac{1}{2}\omega^2 \underbrace{(a^2\sigma^2)}_{\sigma_Y^2} \right].$$

La fonction $\phi_Y(\omega)$ est de la même forme que $\phi_X(\omega)$, alors on peut en déduire que Y suit aussi une loi normale.

Exemple 3.20. Soit $X \sim \mathcal{B}(n, p)$. Quelle est la fonction caractéristique de X ?

La fonction caractéristique de la loi binomiale est donnée par

$$\phi_X(\omega) = \sum_{k=0}^n e^{j\omega k} \cdot \mathcal{C}_k^n p^k q^{n-k} = \sum_{k=0}^n \mathcal{C}_k^n (pe^{j\omega})^k q^{n-k} = (pe^{j\omega} + q)^n,$$

selon le binôme de Newton.

3.17 Fiabilité

Définition. La *fonction de fiabilité* d'un système, dénotée $R(t)$, est la probabilité que le temps de vie T du système, une variable aléatoire continue et positive, soit supérieur à une valeur t , c'est-à-dire

$$R(t) = \mathbb{P}[T > t] = 1 - F_T(t).$$

Définition. Le *taux de défaillance* d'un système, dénotée $r(t)$, est la probabilité qu'un système tombe subitement en panne, sachant qu'il fonctionnait au moment précédent, c'est-à-dire

$$r(t) = \lim_{s \rightarrow t} f_T(s|T > t) = \frac{f_T(t)}{1 - F_T(t)} = -\frac{R'(t)}{R(t)}.$$

Théoreme 3.16. $R(t) = \exp \left\{ - \int_0^t r(s) ds \right\}.$

Démonstration. Il suffit de résoudre l'équation différentielle

$$r(t) = -\frac{R'(t)}{R(t)},$$

avec $R(0) = 1$, pour obtenir

$$R(t) = \exp \left\{ - \int_0^t r(s) ds \right\}.$$

□

Exemple 3.21. Soit $T \sim \text{Exp}(\lambda)$. Quel est le taux de défaillance ?

On calcule la fonction de fiabilité avec

$$R(t) = 1 - F_T(t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}$$

de sorte à obtenir le taux de défaillance

$$r(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

3.17.1 Durée de vie moyenne

Définition. La *durée de vie moyenne* d'un système est l'espérance de T , le temps de vie du système.

Théoreme 3.17. $E[T] = \int_0^\infty R(t) dt.$

Démonstration. Par définition, on a

$$E[T] = \int_0^\infty t f_T(t) dt = \int_0^\infty \left(\int_0^t ds \right) f_T(t) dt = \int_0^\infty \int_0^t f_T(t) ds dt.$$

En inversant l'ordre d'intégration, on obtient

$$E[T] = \int_0^\infty \underbrace{\int_s^\infty f_T(t) dt}_{\mathbb{P}[T > s]} ds = \int_0^\infty R(s) ds. \quad \square$$

Exemple 3.22. Soit un système en série à n composantes indépendantes. Quelle est la fonction de fiabilité du système ?

La probabilité que le système fonctionne après un temps t est équivalent à la probabilité que tous les composantes fonctionnent après un temps t , soit

$$\mathbb{P}[T > t] = \mathbb{P}[\{T_1 > t\} \cap \cdots \cap \{T_n > t\}] = \mathbb{P}[T_1 > t] \cdots \mathbb{P}[T_n > t],$$

car les événements sont indépendants. Par conséquent, on obtient

$$R(t) = \prod_{k=1}^n R_k(t).$$

Exemple 3.23. Soit un système en parallèle à n composantes indépendantes. Quelle est la fonction de fiabilité du système ?

La probabilité que le système soit en panne est équivalent à la probabilité que tous les composantes soit en panne après un temps t , soit

$$\mathbb{P}[T \leq t] = \mathbb{P}[\{T_1 \leq t\} \cap \cdots \cap \{T_n \leq t\}] = \mathbb{P}[T_1 \leq t] \cdots \mathbb{P}[T_n \leq t],$$

car les événements sont indépendants. Par conséquent, on obtient

$$R(t) = 1 - \prod_{k=1}^n [1 - R_k(t)].$$

4. VECTEURS ALÉATOIRES

Définition. Un *vecteur aléatoire* $\vec{X} = [X_1 \ X_2 \ \cdots \ X_n]^T$ est une généralisation à n dimensions d'une variable aléatoire. Il est formé de plusieurs variables aléatoires observées lors d'une même expérience.

Exemple 4.1. Soit le lancer de 2 dés. On pose X comme étant le résultat du premier dé et Y comme étant le résultat du deuxième dé. Quelle est la probabilité d'avoir $(X, Y) = (j, k)$?

L'espace échantillon est donné par

$$S_{X,Y} = \left\{ \begin{array}{cccc} (1,1), & (2,1), & \cdots & (6,1), \\ (1,2), & (2,2), & \cdots & (6,2), \\ \vdots & \vdots & \ddots & \vdots \\ (1,6), & (2,6), & \cdots & (6,6) \end{array} \right\},$$

de sorte la probabilité d'avoir l'événement voulu est

$$\mathbb{P}[\{X = j\} \cap \{Y = k\}] = \frac{1}{36}, \quad \forall (j, k) \in \{1, \dots, 6\} \times \{1, \dots, 6\}.$$

Exemple 4.2. On génère un point au hasard dans le triangle T de coordonnées $(0, 0)$, $(1, 1)$ et $(0, 1)$. Soit X l'abscisse du point et Y l'ordonnée. Sachant qu'il y a équiprobabilité, quelle est la fonction de densité conjointe ?

Puisqu'il y a équiprobabilité, tous les points dans le triangle ont la même probabilité d'être généré. Par conséquent, on peut définir la fonction de densité comme

$$f_{X,Y}(x, y) = \begin{cases} c & \text{si } (x, y) \text{ est dans le triangle} \\ 0 & \text{sinon} \end{cases},$$

où la constante c est donnée par

$$\iint_T f_{X,Y}(x, y) \, dA = 1 \Leftrightarrow c = 2.$$

4.1 Vecteur aléatoire discret

4.1.1 Fonction de masse conjointe

Définition. La *fonction de masse conjointe*, dénotée $p_{\vec{X}}(\vec{x})$, est la probabilité qu'un vecteur aléatoire discret \vec{X} prenne une certaine valeur, c'est-à-dire

$$p_{\vec{X}}(\vec{x}) = \mathbb{P} \left[\vec{X} = \vec{x} \right] = \mathbb{P} \left[\bigcap_{k=1}^n \{X_k = x_k\} \right].$$

Propriété 1. $\sum_{\forall \vec{x} \in S_{\vec{X}}} p_{\vec{X}}(\vec{x}) = 1.$

Notation. Soit un vecteur aléatoire \vec{X} à n composantes. Alors la notation de la fonction du vecteur aléatoire $g_{\vec{X}}(\vec{x})$ est équivalente à $g_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$.

4.1.2 Fonction de masse marginale

Définition. Soit \vec{X} , \vec{Y} et \vec{Z} des vecteurs aléatoires discrets tels que $\vec{X} = \begin{bmatrix} \vec{Y} & \vec{Z} \end{bmatrix}^T$. La *fonction de masse marginale*, dénotée $p_{\vec{Z}}(\vec{z})$, est la probabilité que \vec{Z} prend une valeur \vec{z} . On a alors

$$p_{\vec{Z}}(\vec{z}) = \sum_{\forall \vec{y} \in S_{\vec{Y}}} p_{\vec{X}} \left(\begin{bmatrix} \vec{y} & \vec{z} \end{bmatrix}^T \right).$$

4.1.3 Fonction de masse conditionnelle

Définition. Soit \vec{X} , \vec{Y} et \vec{Z} des vecteurs aléatoires discrets tels que $\vec{X} = \begin{bmatrix} \vec{Y} & \vec{Z} \end{bmatrix}^T$. La *fonction de masse conditionnelle*, dénotée $p_{\vec{Z}|\vec{Y}}(\vec{z}|\vec{y})$, est la probabilité que \vec{Z} prend une valeur \vec{z} sachant que \vec{Y} prend une valeur \vec{y} . On a alors

$$p_{\vec{Z}|\vec{Y}}(\vec{z}|\vec{y}) = \frac{p_{\vec{X}} \left(\begin{bmatrix} \vec{y} & \vec{z} \end{bmatrix}^T \right)}{p_{\vec{Y}}(\vec{y})}.$$

4.2 Vecteur aléatoire continu

Remarque. Les fonctions de densité ne sont pas des probabilités de prendre une valeur précise, mais bien une valeur au alentour.

4.2.1 Fonction de densité conjointe

Définition. La *fonction de densité conjointe*, dénotée $f_{\vec{X}}(\vec{x})$, est la probabilité qu'un vecteur aléatoire continu \vec{X} prenne une certaine valeur \vec{x} , c'est-à-dire

$$f_{\vec{X}}(\vec{x}) = \frac{1}{\epsilon_1 \epsilon_2 \dots \epsilon_n} \mathbb{P} \left[\bigcap_{k=1}^n \left\{ x - \frac{\epsilon_k}{2} \leq X_k \leq x + \frac{\epsilon_k}{2} \right\} \right],$$

où $\epsilon_i \rightarrow 0^+$ pour tout $i \in \{1, 2, \dots, n\}$.

Propriété 1. $\int \cdots \int_{\vec{x} \in S_{\vec{X}}} f_{\vec{X}}(\vec{x}) d\vec{x} = 1.$

4.2.2 Fonction de densité marginale

Définition. Soit \vec{X} , \vec{Y} et \vec{Z} des vecteurs aléatoires continus tels que $\vec{X} = \begin{bmatrix} \vec{Y} & \vec{Z} \end{bmatrix}^T$. La *fonction de densité marginale*, dénotée $f_{\vec{Z}}(\vec{z})$, est la probabilité que \vec{Z} prend une valeur \vec{z} . On a alors

$$f_{\vec{Z}}(\vec{z}) = \int \cdots \int_{\vec{y} \in S_{\vec{Y}}} f_{\vec{X}} \left(\begin{bmatrix} \vec{y} & \vec{z} \end{bmatrix}^T \right) d\vec{y}.$$

4.2.3 Fonction de densité conditionnelle

Définition. Soit \vec{X} , \vec{Y} et \vec{Z} des vecteurs aléatoires continus tels que $\vec{X} = \begin{bmatrix} \vec{Y} & \vec{Z} \end{bmatrix}^T$. La *fonction de densité conditionnelle*, dénotée $f_{\vec{Z}|\vec{Y}}(\vec{z}|\vec{y})$, est la probabilité que \vec{Z} prend une valeur \vec{z} sachant que \vec{Y} prend une valeur \vec{y} . On a alors

$$f_{\vec{Z}|\vec{Y}}(\vec{z}|\vec{y}) = \frac{f_{\vec{X}} \left(\begin{bmatrix} \vec{y} & \vec{z} \end{bmatrix}^T \right)}{f_{\vec{Y}}(\vec{y})}.$$

4.3 Fonction de répartition conjointe

Définition. La *fonction de répartition conjointe*, dénotée $F_{\vec{X}}(\vec{x})$, est la probabilité que les composantes d'un vecteur \vec{X} soient plus petites ou égales à celles de \vec{x} , c'est-à-dire

$$F_{\vec{X}}(\vec{x}) = \mathbb{P} \left[\bigcap_{k=1}^n \{X_k \leq x_k\} \right].$$

Théoreme 4.1. $F_{\vec{X}}(\vec{x}) = \sum_{\forall \vec{x}' \leq \vec{x}} p_{\vec{X}}(\vec{x}')$ dans le cas discret.

Théoreme 4.2. $F_{\vec{X}}(\vec{x}) = \int \cdots \int_{\forall \vec{x}' \leq \vec{x}} f_{\vec{X}}(\vec{x}') d\vec{x}'$ dans le cas continu.

Théoreme 4.3. $f_{\vec{X}}(\vec{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F_{\vec{X}}(\vec{x})$ dans le cas continu.

4.4 Fonction de répartition marginale

Définition. Soit \vec{X} , \vec{Y} et \vec{Z} des vecteurs aléatoires tels que $\vec{X} = \begin{bmatrix} \vec{Y} & \vec{Z} \end{bmatrix}^T$. La *fonction de répartition marginale*, dénotée $F_{\vec{Z}}(\vec{z})$, est la probabilité que \vec{Z} prend une valeur \vec{z} . On a alors

$$F_{\vec{Z}}(\vec{z}) = \lim_{\vec{y} \rightarrow \infty} F_{\vec{X}} \left(\begin{bmatrix} \vec{y} & \vec{z} \end{bmatrix}^T \right).$$

Notation. Soit un vecteur $\vec{x} \in \mathbb{R}^n$. La limite lorsque $\vec{x} \rightarrow \infty$ dénote la limite lorsque $x_i \rightarrow \infty$ pour tout $i \in \{1, 2, \dots, n\}$.

Exemple 4.3. Quelle est la probabilité que $\begin{bmatrix} X & Y \end{bmatrix}^T$ soit dans un rectangle $R = [a, b] \times [c, d]$? On suppose que la fonction de répartition conjointe est

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

En utilisant seulement la fonction de répartition conjointe, on peut montrer que

$$\mathbb{P}[R] = F_{X,Y}(b, d) - F_{X,Y}(b, c) - F_{X,Y}(a, d) + 4F_{X,Y}(a, c).$$

4.5 Indépendance dans un vecteur

Définition. Soit les variables aléatoires X et Y . Ces variables sont *indépendantes* si et seulement si

1. $S_{X,Y} = S_X \times S_Y$
2. $F_{X,Y}(x, y) = F_X(x)F_Y(y)$
 $\Rightarrow p_{X,Y}(x_j, y_k) = p_X(x_j)p_Y(y_k)$
 $\Rightarrow f_{X,Y}(x, y) = f_X(x)f_Y(y)$

Exemple 4.4. La veille d'un examen, un professeur estime qu'il recevra X questions par courriel, où $X \sim \text{Poi}(\alpha)$. Le professeur répond à chaque question avec une probabilité p , et ce indépendamment d'une question à l'autre. Soit Y le nombre de réponses du professeur. Déterminer $p_Y(y)$.

L'espace échantillon du vecteur aléatoire discrète (X, Y) est

$$S_{X,Y} = \left\{ \begin{pmatrix} (0, 0), & (1, 0), & \cdots & (j, 0), & \cdots \\ & (1, 1), & \cdots & (j, 1), & \cdots \\ & & \ddots & \vdots & \\ & & & (j, j), & \cdots \\ & & & & \ddots \end{pmatrix} \right\}.$$

Exemple 4.4 (suite). On peut montrer que

$$p_{X,Y}(j, k) = \underbrace{p_{Y|X}(k, j)}_{\mathcal{B}(j,p)} \underbrace{p_X(j)}_{\text{Poi}(\alpha)} = C_j^k p^k q^{j-k} \cdot \frac{e^{-\alpha}}{j!} \alpha^j,$$

de sorte que

$$p_Y(k) = \sum_{j=k}^{\infty} p_{X,Y}(j, k) = \sum_{j=k}^{\infty} \frac{j!}{k!(j-k)!} p^k q^{j-k} \frac{e^{-\alpha}}{j!} \alpha^j = \frac{p^k}{k!} e^{-\alpha} \sum_{j=k}^{\infty} \frac{1}{(j-k)!} q^{j-k} \alpha^j.$$

Exemple 4.4 (suite). En appliquant une translation à la série, on obtient le développement en série de l'exponentielle, soit

$$p_Y(k) = \frac{p^k}{k!} e^{-\alpha} \sum_{j=0}^{\infty} \frac{(q\alpha)^j}{j!} \alpha^k = \frac{p^k}{k!} e^{-\alpha} e^{q\alpha} \alpha^k = \frac{(p\alpha)^k}{k!} e^{-(p\alpha)}.$$

Par conséquent, $Y \sim \text{Poi}(p\alpha)$.

4.6 Espérance conditionnelle

Théoreme 4.4. Soit X et Y des variables aléatoires, alors

$$\mathbb{E}[Y|X = x_j] = \sum_{k=1}^{\infty} y_k p_{Y|X}(x_k, y_k)$$

si les variables sont discrètes, et

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

si X est continue.

Remarque. On peut définir l'espérance comme une fonction, c'est-à-dire $g(x) = \mathbb{E}[Y|X = x]$ de sorte que $g(X) = \mathbb{E}[Y|X]$ est une variable aléatoire.

Théoreme 4.5. $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$.

Démonstration (cas continu). On sait que

$$\mathbb{E}[\mathbb{E}[Y|X]] = \int_{-\infty}^{\infty} \mathbb{E}[Y|X = x] f_X(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy \cdot f_X(x) dx.$$

On remarque que

$$\mathbb{E}[\mathbb{E}[Y|X]] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \cdot \underbrace{f_{Y|X}(y|x) f_X(x)}_{f_{X,Y}(x,y)} dy dx = \int_{-\infty}^{\infty} y \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x, y) dx}_{f_Y(y)} dy$$

de sorte que

$$\mathbb{E}[\mathbb{E}[Y|X]] = \int_{-\infty}^{\infty} y f_Y(y) dy = \mathbb{E}[Y]. \quad \square$$

4.7 Variance conditionnelle

Théoreme 4.6. $\text{Var}[Y|X] = \mathbb{E}[Y^2|X] - (\mathbb{E}[Y|X])^2$.

Théoreme 4.7. $\text{Var}[Y] = \mathbb{E}[\text{Var}[Y|X]] + \text{Var}[\mathbb{E}[Y|X]]$.

4.8 Espérance d'une transformation

Théoreme 4.8. Soit \vec{X} un vecteur aléatoire et $g(\vec{x})$ une transformation, alors

$$\mathbb{E}[g(\vec{X})] = \sum_{\forall \vec{x} \in S_{\vec{X}}} g(\vec{x}) p_{\vec{X}}(\vec{x})$$

dans le cas discret, et

$$\mathbb{E}[g(\vec{X})] = \int \cdots \int g(\vec{x}) f_{\vec{X}}(\vec{x}) d\vec{x}$$

dans le cas continu.

Exemple 4.5. Soit X et Y des variables aléatoires dont la fonction de masse conjointe est donnée par la table suivante.

Y \ X	X		
	0	1	2
0	1/6	1/6	1/6
1	0	1/6	1/6
2	0	0	1/6

Calculer $\mathbb{E}[X^2 Y]$.

Il suffit d'utiliser le théorème précédent, soit

$$\mathbb{E}[X^2 Y] = \sum_{j=0}^2 \sum_{k=0}^2 (j^2 k) p_{X,Y}(j, k) = \frac{13}{6}.$$

Exemple 4.6. Soit $X \sim \mathcal{N}(1, 1)$ et $Y \sim \mathcal{N}(2, 1)$ des variables aléatoires indépendantes. Calculer $E[X + Y]$ et $E[XY]$.

On sait que

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

Puisque l'espérance est un opérateur linéaire, on a

$$E[X + Y] = \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x, y) \, dx dy}_{E[X]} + \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x, y) \, dx dy}_{E[Y]}$$

de sorte à obtenir que $E[X + Y] = 3$.

Exemple 4.6 (suite). On sait que

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (xy) f_{X,Y}(x, y) \, dx dy.$$

Puisque les variables aléatoires sont indépendantes, on a

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

de sorte à obtenir $E[XY] = 2$.

Théoreme 4.9. Soit un vecteur aléatoire \vec{X} à n variables indépendantes et une transformation $g(\vec{x})$, alors

$$E[g(\vec{X})] = E[g_1(X_1)] E[g_2(X_2)] \cdots E[g_n(X_n)],$$

si $g(X) = g_1(X_1)g_2(X_2) \cdots g_n(X_n)$.

4.9 Corrélation et covariance

Définition. La *corrélation* de deux variables aléatoires X et Y est l'espérance de leur produit, soit $E[XY]$.

Remarque. Lorsque $E[XY] = 0$, on dit que X et Y sont *orthogonales*.

Définition. La *covariance* de deux variables aléatoires X et Y , dénotée $\text{Cov}[X, Y]$, est une mesure permettant de quantifier leurs écarts conjoints par rapport à leurs espérances respectives, c'est-à-dire

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])].$$

Propriété 1. $\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$.

Propriété 2. $\text{Cov}[X, X] = \text{Var}[X]$.

Définition. Le *coefficient de corrélation* $\rho_{X,Y}$ est une mesure de corrélation linéaire entre deux variables aléatoires X et Y . Il est défini comme

$$\rho_{X,Y} = \frac{\text{Cov}[X, Y]}{\text{Std}[X] \text{Std}[Y]}.$$

Propriété 1. $-1 \leq \rho_{X,Y} \leq 1$.

Propriété 2. $\rho_{X,Y} = 1 \Leftrightarrow Y = aX + b$ où $a > 0$.

Propriété 3. $\rho_{X,Y} = -1 \Leftrightarrow Y = aX + b$ où $a < 0$.

Propriété 4. $\rho_{X,Y} = 0$ si X et Y sont indépendantes.

Remarque. Si $\rho_{X,Y} = 0$, alors X et Y ne sont pas nécessairement indépendantes. On dit que ce sont des *variables non-corrélées*.

Théoreme 4.10. Soit $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ des variables aléatoires telles que $\rho_{X,Y} = 0$, alors X et Y sont indépendantes.

4.10 Loi binormale

Définition. Soit un vecteur aléatoire $[X \ Y]^T$. Si $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ et $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ alors on dit que le vecteur suit une *loi binormale*, dénotée

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho),$$

où $\rho \equiv \rho_{X,Y}$. On peut alors montrer que

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 + 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right\}.$$

Propriété 1. $X | \{Y = y\} \sim \mathcal{N}\left(\mu_X + \rho \frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1 - \rho^2)\right)$.

4.11 Estimation

Définition. Un *estimateur* $g(X)$ est une fonction cherchant à prévoir la valeur de Y à partir de X . Elle est définie comme étant la solution minimisant l'erreur quadratique moyenne, c'est-à-dire

$$\min E \left[(Y - g(X))^2 \right].$$

4.11.1 Estimateur constant

Théoreme 4.11. *Si $g(X)$ est un estimateur constant de Y , alors $g(X) = E[Y]$.*

Démonstration. Dans un cas où $g(X) = c$, le problème se résume à

$$\min E \left[(Y - c)^2 \right] \equiv \min E \left[Y^2 \right] - 2cE[Y] + c^2.$$

En dérivant, on obtient la constante optimale

$$\frac{d}{dc} (E[Y^2] - 2cE[Y] + c^2) = -2E[Y] + 2c = 0 \Leftrightarrow c = E[Y]$$

de sorte que

$$g(X) = E[Y]. \quad \square$$

Corollaire. $E \left[(Y - g(X))^2 \right] = \text{Var}[Y]$.

4.11.2 Estimateur linéaire

Théoreme 4.12. *Si $g(X)$ est un estimateur linéaire de Y , alors $g(X) = \hat{a}X + \hat{b}$, où*

$$\hat{a} = \frac{\text{Std}[Y]}{\text{Std}[X]} \rho_{X,Y} \quad \text{et} \quad \hat{b} = E[Y] - \hat{a}E[X].$$

Corollaire. $E \left[(Y - g(X))^2 \right] = \text{Var}[Y] (1 - \rho_{X,Y}^2)$.

4.11.3 Estimateur non linéaire

Théoreme 4.13. *Si $g(X)$ est un estimateur non-linéaire de Y , alors $g(X) = E[Y|X]$.*

Démonstration (cas discret). Selon le théorème 4.5, on a

$$\min E \left[(Y - g(X))^2 \right] \equiv \min E \left[E \left[(Y - g(X))^2 | X \right] \right],$$

de sorte que problème peut s'écrire sous la forme

$$\min \sum_{j=1}^{\infty} \underbrace{E[Y - g(X)|X = x_j]}_{\text{fonction positive à min. } \forall x_j} p_X(x_j).$$

On retrouve alors plusieurs cas d'estimateurs constants, mais conditionnés pour $X = x_j$. Par conséquent,

$$g(X) = E[Y|X]. \quad \square$$

4.11.4 Estimateur d'une binormale

Théoreme 4.14. *Soit un vecteur aléatoire $[X \ Y]^T \sim \mathcal{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, alors le meilleur estimateur $g(X)$ de Y est un estimateur linéaire.*

Exemple 4.7. Soit $\begin{bmatrix} X & Y \end{bmatrix}^T \sim \mathcal{N}(\mu_X = 3, \mu_Y = 1, \sigma_X^2 = 4, \sigma_Y^2 = 9, \rho = 1/4)$. Quel est le meilleur estimateur de Y en fonction de X ?

Avec le théorème précédent, on a que le meilleur estimateur est linéaire, alors

$$\hat{a} = \frac{\text{Std}[Y]}{\text{Std}[X]} \rho = \frac{3}{8}$$

et

$$\hat{b} = E[Y] - \hat{a}E[X] = -\frac{1}{8},$$

de sorte que

$$g(X) = \frac{3}{8}X - \frac{1}{8}.$$

4.12 Combinaison linéaire

Définition. Soit un vecteur aléatoire $\begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}^T$, alors la *combinaison linéaire* Z des variables aléatoires du vecteur est

$$Z = a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n = a_0 + \sum_{i=1}^n a_iX_i,$$

où a_0, a_1, \dots, a_n sont des constantes.

Propriété 1. $E[Z] = a_0 + \sum_{i=1}^n a_iE[X_i]$.

Propriété 2. $\text{Var}[Z] = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}[X_i, X_j]$.

Démonstration. Soit $Y_i = a_iX_i$. Par définition, on a

$$\text{Var}[Z] = \text{Var}\left[a_0 + \sum_{i=1}^n Y_i\right] = \text{Var}\left[\sum_{i=1}^n Y_i\right]$$

et

$$\text{Var}[Z] = E\left[\left(\sum_{i=1}^n Y_i\right)^2\right] - E\left[\sum_{i=1}^n Y_i\right]^2 = E\left[\left(\sum_{i=1}^n Y_i\right)^2\right] - \left(\sum_{i=1}^n E[Y_i]\right)^2.$$

En écrivant explicitement les carrés, on a

$$\text{Var}[Z] = E\left[\left(\sum_{i=1}^n Y_i\right)\left(\sum_{j=1}^n Y_j\right)\right] - \left(\sum_{i=1}^n E[Y_i]\right)\left(\sum_{j=1}^n E[Y_j]\right)$$

et

$$\text{Var}[Z] = \mathbb{E} \left[\sum_{i=1}^n \sum_{j=1}^n Y_i Y_j \right] - \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[Y_i] \mathbb{E}[Y_j]$$

On peut écrire les sommes comme

$$\text{Var}[Z] = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[Y_i Y_j] - \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[Y_i] \mathbb{E}[Y_j] = \sum_{i=1}^n \sum_{j=1}^n \left(\mathbb{E}[Y_i Y_j] - \mathbb{E}[Y_i] \mathbb{E}[Y_j] \right).$$

Avec la définition de la covariance, on obtient

$$\text{Var}[Z] = \sum_{i=1}^n \sum_{j=1}^n \text{Cov}[Y_i, Y_j] = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}[X_i, X_j] \quad \square$$

4.13 Variables aléatoires indépendantes et identiquement distribuées

Définition. Les variables aléatoires X_1, X_2, \dots, X_n sont indépendantes et identiquement distribuées si elles sont tous indépendantes entre elles et qu'elles suivent tous la même distribution.

Notation. On dénote $S_n = X_1 + X_2 + \dots + X_n$ la somme de n variables aléatoires indépendantes et identiquement distribuées.

Propriété 1. $\mathbb{E}[S_n] = n\mathbb{E}[X]$.

Propriété 2. $\text{Var}[S_n] = n\text{Var}[X]$.

Exemple 4.8. Si $X_1 \sim \mathcal{B}(n_1, p)$ et $X_2 \sim \mathcal{B}(n_2, p)$ sont des variables aléatoires indépendantes, alors on a

$$X_1 + X_2 \sim \mathcal{B}(n_1 + n_2, p).$$

Exemple 4.9. Si $X_i \sim \text{Poi}(\alpha_i)$ sont des variables aléatoires indépendantes, alors on a

$$X_1 + X_2 + \dots + X_n \sim \text{Poi}(\alpha),$$

où $\alpha = \alpha_1 + \alpha_2 + \dots + \alpha_n$.

Exemple 4.10. Si $X_i \sim \text{Exp}(\lambda)$ sont des variables aléatoires indépendantes, alors on a

$$X_1 + X_2 + \dots + X_n \sim \text{Gam}(n, \lambda).$$

Exemple 4.11. Si $X_i \sim \text{Gam}(\alpha_i, \lambda)$ sont des variables aléatoires indépendantes, alors on a

$$X_1 + X_2 + \cdots + X_n \sim \text{Gam}(\alpha, \lambda),$$

où $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_n$.

Exemple 4.12. Si $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ sont des variables aléatoires indépendantes, alors on a

$$a_0 + \sum_{i=1}^n a_i X_i \sim \mathcal{N}(\mu, \sigma^2),$$

où

$$\mu = a_0 + \sum_{i=1}^n a_i \mu_i \quad \text{et} \quad \sigma^2 = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}[X_i, X_j].$$

4.14 Loi faible des grands nombres

Théoreme 4.15. Soit une variable aléatoire $S_n = X_1 + X_2 + \cdots + X_n$, où X_1, X_2, \dots, X_n sont des variables aléatoires indépendantes et identiquement distribuées de moyenne μ , alors

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\left| \frac{S_n}{n} - \mu \right| < c \right] = 1,$$

pour tout $c > 0$.

4.15 Loi forte des grands nombres

Théoreme 4.16. Soit une variable aléatoire $S_n = X_1 + X_2 + \cdots + X_n$, où X_1, X_2, \dots, X_n sont des variables aléatoires indépendantes et identiquement distribuées de moyenne μ et de variance σ^2 , alors

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu \right] = 1.$$

4.16 Théorème central limite

Théoreme 4.17. Soit une variable aléatoire $S_n = X_1 + X_2 + \cdots + X_n$, où X_1, X_2, \dots, X_n sont des variables aléatoires indépendantes et identiquement distribuées de moyenne μ et de variance σ^2 . Si n est grand, alors

$$S_n \approx \mathcal{N}(n\mu, n\sigma^2) \Leftrightarrow \frac{S_n}{n} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \approx \mathcal{N}(0, 1).$$

Remarque. En général, le théorème central limite est applicable si $n \geq 30$.

Exemple 4.13. Soit $X \sim \mathcal{B}(2000, 1/2)$. Donnez une approximation de X .

Une loi binomiale est équivalente à une somme de lois de Bernoulli indépendantes et identiquement distribuées de paramètre p . On approxime donc la loi binomiale à l'aide du théorème central limite, soit

$$X \approx \mathcal{N}(np, npq).$$

On a alors que

$$\mathbb{P}[X = k] \approx \frac{1}{\sqrt{2\pi \cdot npq}} \exp \left\{ -\frac{(k - np)^2}{2npq} \right\}$$

Puisqu'on approxime une loi discrète par une loi continue, on applique une *correction de continuité*, c'est-à-dire

$$\mathbb{P}[a \leq X \leq b] \approx \Phi \left(\frac{b + 1/2 - np}{\sqrt{npq}} \right) - \Phi \left(\frac{a - 1/2 - np}{\sqrt{npq}} \right)$$

et

$$\mathbb{P}[a < X \leq b] \approx \Phi \left(\frac{b + 1/2 - np}{\sqrt{npq}} \right) - \Phi \left(\frac{a + 1/2 - np}{\sqrt{npq}} \right).$$

Exemple 4.14. Soit les variables aléatoires X et Y avec la fonction densité conjointe suivante

$$f_{X,Y}(x, y) = \frac{3}{4},$$

si $0 < y < 1$ et $x^2 < y$. Calculez $f_X(x)$ et $f_Y(y)$, ainsi que $\mathbb{P}[Y \geq X]$.

Pour un x , on a

$$f_X(x) = \int_{x^2}^1 f_{X,Y}(x, y) dy = \frac{3}{4}(1 - x^2),$$

si $-1 < x < 1$. Pour un y , on a

$$f_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} f_{X,Y}(x, y) dx = \frac{3}{2}\sqrt{y},$$

si $0 < y < 1$.

Par définition, on a

$$\mathbb{P}[Y \geq X] = \int_0^1 \int_{-\sqrt{y}}^{\sqrt{y}} f_{X,Y}(x, y) dx dy = \frac{7}{8}.$$

5. PROCESSUS STOCHASTIQUE

Définition. Un *processus stochastique* est une expérience aléatoire qui se déroule dans le temps. Il est défini par son *espace des états* $\{X(t) : t \in T\}$, où $X(t)$ est la variable aléatoire correspondant à l'observation du processus au temps $t \leq 0$.

Définition. Une *trajectoire* d'un processus stochastique est une événement élémentaire de l'espace des états.

Il existe 4 types de processus stochastique :

1. processus stochastique à temps discret et état discret (PSTDDE)

Exemple. Le nombre de personnes en classe à chaque jour.

2. processus stochastique à temps discret et état continu (PSTDEC)

Exemple. La valeur d'une action à la fermeture de la bourse.

3. processus stochastique à temps continu et état discret (PSTCED)

Exemple. La longueur de la file d'attente à la cafétéria.

4. processus stochastique à temps discret et état discret (PSTCEC)

Exemple. La valeur d'une action en temps réel.

Définition. Soit $\{X(t) : t \in T\}$, où $T = [0, \infty[$, un processus stochastique à temps continu et état continu. La *fonction de répartition d'ordre n* est la probabilité que n états soient plus petite ou égal à n valeurs, c'est-à-dire

$$F(x_1, \dots, x_n; t_1, \dots, t_n) = \mathbb{P}[\{X(t_1) \leq x_1\} \cap \dots \cap \{X(t_n) \leq x_n\}].$$

Définition. Soit $\{X(t) : t \in T\}$, où $T = [0, \infty[$, un processus stochastique à temps continu et état continu. La *fonction de densité d'ordre n* est la probabilité que n états soient autour de n valeurs, c'est-à-dire

$$f(x_1, \dots, x_n; t_1, \dots, t_n) = \mathbb{P}[\{X(t_1) = x_1\} \cap \dots \cap \{X(t_n) = x_n\}].$$

Théoreme 5.1. $f(x_1, \dots, x_n; t_1, \dots, t_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F(x_1, \dots, x_n; t_1, \dots, t_n).$

Définition. Un *accroissement* $X(t_1, t_2)$, où $t_1 < t_2$, est la différence des états à t_1 et t_2 , c'est-à-dire

$$X(t_1, t_2) = X(t_2) - X(t_1).$$

Exemple 5.1. Soit un patient qui recoit une dose de médicament Y . Au temps $t = 0$, on a $Y \sim \mathcal{U}(0, 1)$. La quantité active X après t unités de temps est définie par $X(t) = Ye^{-t}$. Quel est la fonction de répartition du premier ordre, soit $F(x; t)$?

Par définition, on a $F(x; t) = \mathbb{P}[X(t) \leq x] = \mathbb{P}[Y] \leq xe^t$. Pour une loi uniforme, on a

$$F(x; t) = \begin{cases} 0 & \text{si } xe^t < 0 \\ xe^t & \text{si } 0 \leq x \leq e^{-t} \\ 1 & \text{si } e^{-t} < x \end{cases}.$$

5.1 Moyenne d'un processus stochastique

Définition. La *moyenne* d'un processus stochastique, dénotée $m_X(t)$, est donnée par

$$m_X(t) = \mathbb{E}[X(t)] = \int_{-\infty}^{\infty} xf(x; t) dx.$$

Exemple 5.2. Quel est la moyenne de $X(t) = Ye^{-t}$, où $Y \sim \mathcal{U}(0, 1)$?

Avec l'exemple précédent, on obtient que la fonction de densité de probabilité est donnée par

$$f(x; t) = \frac{\partial}{\partial x} F(x; t) = e^t,$$

pour $0 \leq x \leq e^{-t}$. La moyenne est donc

$$m_X(t) = \int_0^{e^{-t}} xe^t dx = \frac{e^{-t}}{2}.$$

5.2 Fonction d'autocorrélation

Définition. La *fonction d'autocorrélation*, dénotée $R_X(t_1, t_2)$, est la corrélation d'un processus stochastique $X(t)$ avec lui-même à deux temps différents. Par conséquent, on a

$$R_X(t_1, t_2) = \mathbb{E}[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2) dx_1 dx_2.$$

5.3 Fonction d'autocovariance

Définition. La *fonction d'autocovariance*, dénotée $C_X(t_1, t_2)$, est la covariance d'un processus stochastique $X(t)$ avec lui-même à deux temps différents. Par conséquent, on a

$$C_X(t_1, t_2) = R_X(t_1, t_2) - m_X(t_1)m_X(t_2).$$

5.4 Processus stochastique stationnaire au sens large

Définition. Si $m_X(t) = c$, où c est une constante, et si $R_X(t_1, t_2) = h(t_2 - t_1)$, où h est une fonction, alors on dit que le processus stochastique est *stationnaire au sens large* (SSL).

Exemple 5.3. Soit un signal aléatoire $X(t) = Y \sin(t + Z)$, où Y et Z sont des variables aléatoires indépendantes et $Z \sim \mathcal{U}(0, 2\pi)$. Montrer que ce signal est SSL.

Par définition, la moyenne est donnée par

$$m_X(t) = E[Y \sin(t + Z)] = E[Y] \underbrace{E[\sin(t + Z)]}_{\int_0^{2\pi} \sin(t + z) \frac{1}{2\pi} dz = 0} = 0,$$

ce qui est une constante. De plus, la fonction d'autocorrélation est donnée par

$$R_X(t_1, t_2) = E[Y \sin(t_1 + Z) Y \sin(t_2 + Z)] = E[Y^2] E[\sin(t_1 + Z) \sin(t_2 + Z)],$$

et avec une identité trigonométrique, on obtient

$$\begin{aligned} R_X(t_1, t_2) &= \frac{1}{2} E[Y^2] \left(E[\cos(t_1 - t_2)] - \underbrace{E[\cos(t_1 - t_2 + 2Z)]}_0 \right) \\ &= \frac{1}{2} E[Y^2] \cos(t_1 - t_2) = h(t_2 - t_1). \end{aligned}$$

Par conséquent, ce signal est stationnaire au sens large.

5.5 Chaîne de Markov

Définition. Une *chaîne de Markov* est un processus stochastique sans mémoire à temps discret et état discret dont l'espace des états $\{X_n : n = 0, 1, \dots\}$ est constitué de nombres entiers.

Notation. On dénote $p_{i,j}$ la probabilité de passer de l'état i vers l'état j , où i et j sont des nombres entiers.

Théoreme 5.2. $p_{i,j} = \mathbb{P}[X_{n+1} = j | X_n = i]$.

5.5.1 Représentation graphique

Une chaîne de Markov peut se représenter comme un graphe, où les noeuds sont les états et les segments sont les probabilités $p_{i,j}$. La figure 6 représente le graphe d'une chaîne de Markov ayant 4 états.

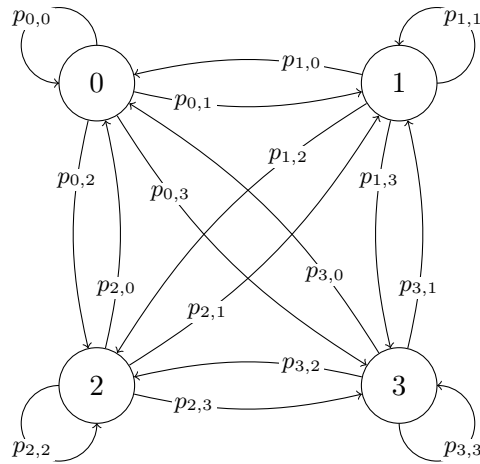


FIGURE 6 – Chaîne de Markov à 4 états

5.5.2 Représentation matricielle

Définition. Soit P la *matrice des probabilités de transition* en une étape où chaque cellule $p_{i,j}$ sont les probabilités de passer de l'état i vers l'état j telles que

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,n} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,0} & p_{n,1} & \cdots & p_{n,n} \end{bmatrix}.$$

Exemple 5.4. On établie des prévisions météorologiques. On défini 3 états suivants : une journée ensoleillée (état 0), une journée nuageuse (état 1) et une journée pluvieuse (état 2). Les probabilités de transition sont définies par la matrice suivante :

$$P = \begin{bmatrix} 0.40 & 0.35 & 0.25 \\ 0.30 & 0.20 & 0.50 \\ 0.75 & 0.10 & 0.15 \end{bmatrix}$$

Quelle est la probabilité qu'il pleuve le lendemain d'une journée ensoleillée ? Quelle est la probabilité que trois journées ensoleillées suivent une journée nuageuse ? Quelle est la probabilité qu'il ne pleuve pas pendant les deux jours suivant une journée pluvieuse ?

On cherche $p_{0,2}$ dans la matrice, alors la probabilité est de 25%.

On cherche $p_{1,0}p_{0,0}p_{0,0}$, alors la probabilité est de 4.8%.

Il faut calculer toutes les trajectoires possibles. Il y a 4 cas possibles On cherche $p_{2,0}p_{0,0} + p_{2,0}p_{0,1} + p_{2,1}p_{1,0} + p_{2,1}p_{1,1}$, alors la probabilité est 61.25%.

5.5.3 Probabilité de transition en n étapes

Notation. On dénote $p_{i,j}^n$ la probabilité de passer de l'état i vers l'état j en n étapes, où i et j sont des nombres entiers.

Théoreme 5.3. $p_{i,j}^n = \mathbb{P}[X_{m+n} = j | X_m = i]$.

Théoreme 5.4. Si $P^n = \underbrace{P \cdot P \cdots P}_{n \text{ fois}}$, alors $p_{i,j}^n = P_{i,j}^n$.

5.6 Processus de comptage

Définition. Un *processus de comptage* est un processus stochastique à temps continu et état discret dont l'espace des états est $\{N(t) : t \geq 0\}$ avec $N(t)$ le nombre d'événements jusqu'à un temps t tel que

1. $N(t) \geq 0$ pour tout $t \geq 0$
2. $N(t) \in \mathbb{N}$ pour tout $t \geq 0$
3. $N(t_1) \leq N(t_2)$ si $t_1 \leq t_2$

5.7 Processus de Poisson

Définition. Un *processus de Poisson* est un processus de comptage où

1. $N(0) = 0$
2. $N(t_1, t_2)$ et $N(t_3, t_4)$ sont indépendants si $t_1 < t_2 \leq t_3 < t_4$
3. $N(\tau, \tau + t) \sim \text{Poi}(\lambda t)$

Théoreme 5.5. $C_N(t_1, t_2) = \lambda \cdot \min(t_1, t_2)$.

5.7.1 Temps d'arrivée

Notation. On dénote T_n le temps entre le n -ième et celui qui le précède.

Théoreme 5.6. $T_n \sim \text{Exp}(\lambda)$.

Démonstration. Soit $M_n(t) = N(T_{n-1} + t) - N(T_{n-1})$ le nombre d'événements compris entre l'événement $n - 1$ et un temps t plus tard. On peut alors définir que

$$\mathbb{P}[T_n > t] = \mathbb{P}[M_n(t) = 0] = \mathbb{P}[N(T_{n-1} + t) - N(T_{n-1}) = 0]$$

Par définition, on a que $N(T_{n-1}) = n - 1$ de sorte que

$$\mathbb{P}[T_n > t] = \mathbb{P}[N(T_{n-1} + t) = n - 1] = \mathbb{P}[\{N(T_{n-1}) = n - 1\} \cap \{N(t) = 0\}]$$

Puisque les accroissements dans un processus de Poisson sont indépendants, on a

$$\mathbb{P}[T_n > t] = \underbrace{\mathbb{P}[N(T_{n-1}) = n - 1]}_1 \mathbb{P}[N(t) = 0] = \mathbb{P}[N(t) = 0] = e^{-\lambda t},$$

car $N(t) = N(t) - N(0) \sim \text{Poi}(\lambda t)$.

Par conséquent, on a que

$$\mathbb{P}[T_n \leq t] = 1 - e^{-\lambda t},$$

ce qui est la fonction de répartition d'une distribution exponentielle de paramètre λ . \square

Définition. Le *temps d'arrivée* S_n d'un événement n est la somme de la différence de tous les événements consécutifs antérieurs, c'est-à-dire

$$S_n = \sum_{k=1}^n T_k.$$

Théoreme 5.7. $S_n \sim \text{Gam}(n, \lambda)$.

Exemple 5.5. Soit un processus de Poisson de taux $\lambda = 2$ avec $t_1 = 3$, $t_2 = 7$, $t_3 = 8$ et $t_4 = 10$. Quelle est la probabilité d'avoir 1 événement dans l'intervalle $]t_1, t_2]$ et plus de 1 événement dans l'intervalle $]t_3, t_4]$?

Soit $X = N(t_1, t_2) \sim \text{Poi}(\lambda)$ et $Y = N(t_3, t_4) \sim \text{Poi}(\lambda)$. On cherche

$$\mathbb{P}[\{X = 1\} \cap \{Y \geq 1\}] = \mathbb{P}[X = 1] \mathbb{P}[Y \geq 1] = 8e^{-8} (1 - e^{-4}).$$

5.8 Marche aléatoire

Définition. Une *marche aléatoire* est un processus stochastique sans mémoire à temps discret et état discret dont l'espace des états est $\{X_n : n \in \mathbb{N}\}$ avec X_n la position du processus tel que

1. $|X_n - X_{n-1}| = 1$ pour tout $n \in \mathbb{N}^*$
2. $X_n \in \mathbb{Z}$ pour tout $n \in \mathbb{N}$

Remarque. On définit souvent ce processus comme le mouvement discret d'une particule en une dimension.

Exemple 5.6. Soit une particule se déplaçant aléatoirement en une dimension. La probabilité qu'elle se dirige vers la droite peut importe sa position est $1/2$ tandis que celle qu'elle se dirige vers la gauche est $1/2$.

On définit les variables aléatoires D et G comme étant le nombre de déplacements vers la droite et la gauche respectivement. On peut montrer que

$$D \sim \mathcal{B}(n, 1/2) \quad \text{et} \quad G \sim \mathcal{B}(n, 1/2),$$

où n est le nombre d'étapes.

Exemple 5.6 (suite). Si un déplacement correspond à une unité entière sur \mathbb{Z} , alors la position X_n à l'étape n est

$$X_n = D - G = D - (n - D) = 2D - n,$$

si $X_0 = 0$, de sorte que

$$p_{0,i} = \mathbb{P}[X_n = i | X_0 = 0] = \mathbb{P}[2D - n = i] = \mathbb{P}\left[D = \frac{n+i}{2}\right],$$

soit la probabilité d'être à la position i en n étapes à partir de 0.

5.9 Processus de Wiener

Définition. Un *processus de Wiener* est un processus stochastique à temps continu et état continu dont l'espace des états est $\{W(t) : t \geq 0\}$, où

1. $W(0) = 0$
2. $W(t_1, t_2)$ et $W(t_3, t_4)$ sont indépendants si $t_1 < t_2 \leq t_3 < t_4$
3. $W(t_1, t_2)$ et $W(t_1 + \tau, t_2 + \tau)$ suivent la même loi pour tout τ et $t_1 < t_2$
4. $W(t) \sim \mathcal{N}(0, \sigma^2 t)$

Remarque. On peut construire ce processus à partir d'une marche aléatoire de pas $\pm\epsilon$ et d'unité de temps δ . On pose alors $\epsilon = \sigma\sqrt{\delta}$ afin d'obtenir

$$W(t) = \lim_{\delta \rightarrow 0^+} X(t).$$

Remarque. Les trajectoires de ce processus sont des fonctions continues, mais dérivables nulle part. La *dérivée généralisée* de ce processus s'appelle le *bruit blanc gaussien*.

Définition. Un *mouvement Brownien standard* est un processus de Wiener où le *coefficient de diffusion* est $\sigma^2 = 1$.

Exemple 5.7. Calculer la fonction de densité du 2-ième ordre $f(w_1, w_2; t_1, t_2)$ du processus de Wiener. On suppose que $t_1 < t_2$.

Par définition, on a

$$f(w_1, w_2; t_1, t_2) = \mathbb{P}[\{W(t_1) = w_1\} \cap \{W(t_2) = w_2\}].$$

En écrivant l'expression en terme d'accroissements indépendants, on a

$$\begin{aligned} f(w_1, w_2; t_1, t_2) &= \mathbb{P}[\{W(t_1) = w_1\} \cap \{W(t_2) - W(t_1) = w_2 - w_1\}] \\ &= \mathbb{P}[W(t_1) = w_1] \mathbb{P}[W(t_2) - W(t_1) = w_2 - w_1]. \end{aligned}$$

Exemple 5.7 (suite). Hors, $X = W(t_1) \sim \mathcal{N}(0, \sigma^2 t_1)$ et $Y = W(t_2) - W(t_1) \sim \mathcal{N}(0, \sigma^2(t_2 - t_1))$ de sorte que

$$f(w_1, w_2; t_1, t_2) = f_X(w_1) f_Y(w_2 - w_1)$$

et

$$f(w_1, w_2; t_1, t_2) = \frac{1}{\sqrt{2\pi\sigma^2 t_1}} \exp\left\{-\frac{(w_1^2)}{2\sigma^2 t_1}\right\} \cdot \frac{1}{\sqrt{2\pi\sigma^2(t_2 - t_1)}} \exp\left\{-\frac{(w_2 - w_1)^2}{2\sigma^2(t_2 - t_1)}\right\}.$$

5.9.1 Fonction d'autocovariance

Théoreme 5.8. $C_w(t_1, t_2) = \sigma^2 \cdot \min(t_1, t_2)$.

[6. STATISTIQUE DESCRIPTIVE]

Définition. Un *échantillon* est un ensemble de données, dénoté x_1, x_2, \dots, x_n , représentant les résultats d'une expérience répétée n fois.

Définition. Un *échantillon ordonné* est un ensemble de données, dénoté $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, tel que $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, représentant les résultats d'une expérience répétée n fois.

Définition. Une *population* est l'ensemble des résultats possibles, ou encore un modèle théorique, de l'expérience.

Définition. Une *classe* est un sous-ensemble de la population.

Définition. L'*effectif* est le nombre de résultats de l'échantillon observés dans chaque classe.

6.1 Représentation numérique

Définition. Une *statistique* est une fonction de l'échantillon, soit $g(x_1, x_2, \dots, x_n)$.

6.1.1 Moyenne de l'échantillon

Définition. La *moyenne* \bar{x} de l'échantillon x_1, x_2, \dots, x_n est une statistique écrivant la valeur moyenne de l'échantillon. Elle est définie par

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k.$$

6.1.2 Médiane de l'échantillon

Définition. La *médiane* \tilde{x} d'un échantillon ordonné $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ est une statistique séparant l'échantillon ordonné en deux sous-échantillons ordonnés de même cardinalité. Elle est définie par

$$\tilde{x} = \begin{cases} x_{(n+1/2)} & \text{si } n \text{ est impair,} \\ \frac{x_{(n/2)} + x_{(n/2+1)}}{2} & \text{si } n \text{ est pair.} \end{cases}$$

6.1.3 Variance de l'échantillon

Définition. La *variance* s^2 d'un échantillon x_1, x_2, \dots, x_n est une statistique décrivant la dispersion des données. Elle est définie par

$$s^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{k=1}^n x_k^2 - n\bar{x}^2 \right).$$

6.1.4 Étendue de l'échantillon

Définition. L'*étendue* R d'un échantillon ordonnée $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ est une statistique décrivant l'étendue des données. Elle est définie par

$$R = x_{(n)} - x_{(1)}.$$

7. INFÉRENCE STATISTIQUE

On suppose que la population est décrite par la variable aléatoire X idéale. On a alors un échantillon aléatoire X_1, X_2, \dots, X_n telle que X_k sont des variables aléatoires indépendantes et identiquement distribuées.

On cherche à *estimer* les paramètres inconnus $\theta_1, \theta_2, \dots, \theta_m$ de la loi de X . La statistique $T = g(X_1, X_2, \dots, x_n)$ est aussi une variable aléatoire. En général, sa loi dépend de $\theta_1, \theta_2, \dots, \theta_m$.

Exemple 7.1. On suppose que X suit une loi quelconque de moyenne $\mu = E[X]$ et $\sigma^2 = \text{Var}[X]$. Les paramètres de la loi sont μ et σ^2 .

On cherche un estimateur de μ tel que

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k.$$

Par le théorème central limite, on a que

$$\bar{X} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Pour un échantillon particulier x_1, x_2, \dots, x_n , \bar{x} est une estimation partielle de μ .

La loi $\mathcal{N}(\mu, \sigma^2/n)$ pour \bar{X} est dite exacte si on suppose que $X \sim \mathcal{N}(\mu, \sigma^2/n)$.

Exemple 7.2. Soit $X \sim \mathcal{U}(0, \theta)$.

On peut estimer θ avec la statistique $X_{(n)} = \max X_1, X_2, \dots, X_n$.

$$\mathbb{P}[X_{(n)} \leq x] = \mathbb{P}[\{X_1 \leq x\} \cap \{X_2 \leq x\} \cap \dots \cap \{X_n \leq x\}] = (\mathbb{P}[X \leq x])^n$$

$$\mathbb{P}[X_{(n)} \leq x] = \begin{cases} 0 & \text{si } x < 0, \\ \left(\frac{x}{\theta}\right)^n & \text{si } 0 \leq x \leq \theta, \\ 1 & \text{si } \theta < x. \end{cases}$$

Définition. Un *estimateur* d'un paramètre inconnu θ est une statistique qui correspond à θ .

Remarque. Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors \bar{X} est un estimateur de μ et S^2 est un estimateur de σ^2 .

Pour une somme de X n fois, on trouve que

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

et aussi

$$\frac{(n-1)S^2}{\sigma^2} \sim \text{Gam}\left(\frac{n-1}{2}, \frac{1}{2}\right),$$

une loi du *khi-deux*. On peut montrer que ces deux variables aléatoires sont indépendantes.

7.1 Biais d'un estimateur

Définition. Le biais d'un estimateur T , dénoté $\text{Biais}[T]$, est la différence entre l'espérance de T et la valeur θ qu'il est sensé estimer, c'est-à-dire

$$\text{Biais}[T] = \text{E}[T] - \theta.$$

Propriété 1. Si $\text{E}[T] = \theta$, alors T est sans biais.

7.2 Erreur quadratique moyenne

Définition. L'erreur quadratique moyenne d'un estimateur T de θ , dénotée $\text{EQM}[T]$ est une mesure de la précision de l'estimateur, définie par

$$\text{EQM}[T] = \text{E}\left[(T - \theta)^2\right].$$

Propriété 1. $\text{EQM}[T] = \text{Var}[T] + (\text{Biais}[T])^2$.

Définition. Soient deux estimateurs T_1 et T_2 de θ . On dit que T_1 est *meilleur* que T_2 si

$$\text{EQM}[T_1] < \text{EQM}[T_2].$$

Exemple 7.3. Soit $X \sim \mathcal{U}(0, \theta)$ et $X_{(n)}$ est un estimateur de θ . On a montré que

$$\mathbb{P}[X_{(n)} \leq x] = \left(\frac{x}{\theta}\right)^n$$

pour $0 \leq x \leq \theta$.

On calcule le biais. On a premièrement

$$f_{X_{(n)}} = \frac{nx^{n-1}}{\theta^n} \quad \text{pour } 0 \leq x \leq \theta.$$

Par conséquent,

$$\text{E}[X_{(n)}] = \int_0^\theta x \cdot \frac{nx^{n-1}}{\theta^n} dx = \frac{n}{\theta^n} \cdot \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta.$$

et

$$\text{Biais} [X_{(n)}] = \frac{n}{n+1}\theta - \theta = -\frac{1}{n+1}\theta.$$

On peut définir

$$\hat{\theta} = \frac{n+1}{\theta} X_{(n)}$$

comme étant un estimateur sans biais de θ , car $\text{Biais} [\hat{\theta}] = 0$.

Exemple 7.4. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$. Comparer l'erreur quadratique moyenne des estimateurs de σ^2 suivant

$$T_1 = S^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

et

$$T_2 = \frac{n-1}{n} S^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2.$$

On sait que

$$\frac{(n-1)S^2}{\sigma^2} \sim \text{Gam}\left(\frac{n-1}{2}, \frac{1}{2}\right).$$

Premièrement, on a

$$\mathbb{E}[S^2] = \frac{\sigma^2}{n-1} \mathbb{E}\left[\frac{(n-1)S^2}{\sigma^2}\right] = \frac{\sigma^2}{n-1} \cdot \frac{(n-1)/2}{1/2} = \sigma^2$$

et

$$\text{Var}[S^2] = \frac{\sigma^2}{(n-1)^2} \text{Var}\left[\frac{(n-1)S^2}{\sigma^2}\right] = \frac{\sigma^4}{(n-1)^2} \cdot \frac{(n-1)/2}{(1/2)^2} = \frac{2\sigma^4}{n-1},$$

de sorte que

$$\text{EQM}[T_1] = \frac{2\sigma^4}{n-1} + (\sigma^2 - \sigma^2)^2 = \frac{2\sigma^4}{n-1}.$$

Deuxièmement, on a

$$\mathbb{E}\left[\frac{n-1}{n} S^2\right] = \frac{n-1}{n} \sigma^2$$

et

$$\text{Var}\left[\frac{n-1}{n} S^2\right] = \left(\frac{n-1}{n}\right)^2 \text{Var}[S^2] = \left(\frac{n-1}{n}\right)^2 \cdot \frac{2\sigma^4}{n-1},$$

de sorte que

$$\text{EQM}[T_2] = \frac{n-1}{n} \cdot \frac{\sigma^4}{n} + \left(\frac{n-1}{n} \sigma^2 - \sigma^2 \right)^2 = \frac{2n-1}{2n} \cdot \frac{2\sigma^4}{n}.$$

On trouve que $\text{EQM}[T_2] < \text{EQM}[T_1]$.

7.3 Recherche d'un estimateur

Soit une population X avec $f_X(x_j; \theta)$ et l'échantillon X_1, X_2, \dots, X_n .

7.3.1 Méthode du maximum de vraisemblance

1. On calcul la *fonction de vraisemblance* $L(\theta)$, soit

$$L(\theta) = \prod_{k=1}^n f_X(x_k; \theta).$$

2. Pour maximiser (θ) , on utilise

$$\ln L(\theta).$$

3. On pose

$$\frac{d}{d\theta} \ln L(\theta) = 0.$$

4. On isole θ .

On note l'expression trouvée θ_{VM} comme étant l' *estimateur à vraisemblance maximale* de θ .

Exemple 7.5. Soit $X \sim \text{Exp}(\theta)$, où $\theta \equiv \lambda$.

1. $L(\theta) = \prod_{k=1}^n f_X(x_k; \theta) = \theta^n \exp \left\{ -\theta \sum_{k=1}^n x_k \right\} = \theta^n e^{-\theta n \bar{X}}.$
2. $\ln L(\theta) = n \ln \theta - \theta n \bar{X}.$
3. $\frac{d}{d\theta} \ln L(\theta) = \frac{n}{\theta} - n \bar{X} = 0$
4. $\theta_{\text{VM}} = \frac{1}{\bar{X}}$

Exemple 7.6. Soit $X \sim \text{Poi}(\theta)$ où $\theta \equiv \alpha$.

1. $L(\theta) = \prod_{k=1}^n p_X(x_k; \theta) = \frac{e^{-\theta} e^{x_1}}{x_1!} \cdot \frac{e^{-\theta} e^{x_2}}{x_2!} \cdots \frac{e^{-\theta} e^{x_n}}{x_n!} = \frac{e^{n\theta} \theta^{n\bar{X}}}{x_1! x_2! \cdots x_n!}$

2. $\ln L(\theta) = -n\theta + n\bar{X} \ln \theta - \sum_{k=1}^n \ln x_k!$
3. $\frac{d}{d\theta} \ln L(\theta) = -n + \frac{n\bar{X}}{\theta} = 0$
4. $\theta_{\text{VM}} = \bar{X}$

Exemple 7.7. Soit $X \sim \mathcal{U}(0, \theta)$.

1. $L(\theta) = \begin{cases} \frac{1}{\theta^n} & \text{si } 0 \leq X_k \leq \theta, \\ 0 & \text{sinon.} \end{cases}$
2. $\theta_{\text{VM}} = X_{(n)}$

7.3.2 Méthode des moments

1. On pose que

$$\mathbb{E}[X^m] = \frac{1}{n} \sum_{k=1}^n X_k^m.$$

2. En commençant par $m = 1$, on cherche la première équation où $\mathbb{E}[X^m] = h(\theta)$.
3. On isole θ . On note θ_N l'expression trouvée.

Exemple 7.8. Soit $X \sim \text{Poi}(\theta)$, où $\theta \equiv \alpha$.

1. $\mathbb{E}[X] = \bar{X}$
2. $\theta_{\text{M}} = \bar{X}$

Exemple 7.9. Soit $X \sim \mathcal{N}(0, \theta^2)$, où $\theta \equiv \sigma$.

1. $\mathbb{E}[X^2] = \frac{1}{n} \sum_{k=1}^n X_k^2 = \theta^2$
2. $\theta_{\text{M}} = \sqrt{\frac{1}{n} \sum_{k=1}^n X_k^2}$

Exemple 7.10. Soit $X \sim \mathcal{U}(0, \theta)$.

1. $\mathbb{E}[X] = \bar{X} = \frac{\theta}{2}$

$$2. \theta_M = 2\bar{X}$$

7.4 Intervalle de confiance

Exemple 7.11. Un premier sondage révèle que $p_1 = 45 \pm 3\%$ des personnes sont en faveur d'un événement, tandis qu'un deuxième sondage révèle $p_2 = 51 \pm 10\%$ des personnes sont en faveur.

Exemple 7.12. Soit X une population avec un paramètre inconnu θ . On définit $[\text{LI}, \text{LS}]$, où LI et LS sont les limites inférieure et supérieure respectivement, comme un intervalle de confiance à $100(1 - \alpha)\%$ pour θ si

$$\mathbb{P}[\text{LI} \leq \theta \leq \text{LS}] = 1 - \alpha$$

On a que LI et LS sont des probabilités.

On cherche l'intervalle de confiance pour une moyenne μ d'une loi $\mathcal{N}(\mu, \sigma^2)$ avec σ^2 connue. Par conséquent, on sait que

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

et

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

On définit $z_\alpha = \mathbb{P}[Z \leq z_\alpha] = \alpha$. On a alors

$$\begin{aligned} \mathbb{P}[-z_{\alpha/2} \leq Z \leq z_{\alpha/2}] &= \mathbb{P}\left[z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right] \\ &= \mathbb{P}\left[\underbrace{\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}_{\text{LI}} \leq \mu \leq \underbrace{\bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}_{\text{LS}}\right] \\ &= 1 - \alpha \end{aligned}$$

Par conséquent, la formule pour l'intervalle de confiance à $100(1 - \alpha)\%$ est

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

Remarque. Si X ne suit pas une loi normale, mais que $\text{Var}[X]$ est tout de même connue, alors la formule trouvée reste valide si n est grand (T.C.L.).

Remarque. Si n est grand et si μ est le seul paramètre de X , alors on obtient l'intervalle de confiance *approximatif* suivant

$$\bar{X} \pm z_{\alpha/2} \frac{\sqrt{h(\bar{X})}}{\sqrt{n}},$$

où $h(\bar{X}) = \text{Var}[X] = h(\mu)$.

Exemple 7.13. Soit $X \sim \mathcal{N}(\mu, 1)$ avec $x_n \in \{2.3, 1.8, 2.0, 1.7, 1.4, 2.2\}$. Calculer l'intervalle de confiance pour μ avec un niveau de confiance de 95%.

On a que $\bar{x} = 1.9$ et $\alpha = 0.025$. Par conséquent, les limites sont données par

$$1.9 \pm z_{0.025} \frac{1}{\sqrt{6}}$$

de sorte que LI = 1.0998 et LS = 2.7002.

La moyenne μ est comprise dans l'intervalle [1.0998, 2.7002] au niveau de confiance de 95%.

7.5 moyenne μ si $\mathcal{N}(\mu, \sigma^2)$ et σ^2 inconnue

Si $n \geq 30$ (grand), alors on remplace σ par S . Les bornes sont donc données par

$$\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{n}}$$

(aussi valide si X n'est pas normale)

Si $n < 30$ (petit), on a

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

(loi de Student à $n - 1$ degrés de liberté)

On définit $t_{\alpha,n}$ par $\mathbb{P}[T > t_{\alpha,n}] = \alpha$. L'intervalle de confiance à $100(a - \alpha)\%$ pour μ est

$$\bar{X} \pm t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}.$$

Exemple 7.14. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$ avec $x_n \in \{2.3, 1.8, 2.0, 1.7, 1.4, 2.2\}$. Calculer l'intervalle de confiance pour μ .

On a que $\bar{x} = 1.9$ et $S = 0.3347$. Par conséquent, l'intervalle de confiance à 95% est

$$1.9 \pm t_{0.025, 5} \frac{0.3347}{\sqrt{6}}$$

et LI = 1.5487 et LS = 2.2513.

Remarque. Un intervalle de confiance unilatérale est un intervalle de la forme $[\text{LI}, \infty[$ et $]-\infty, \text{LS}]$. On réutilise les formules de LI et LS en remplaçant $z_{\alpha/2}$ par z_{α} ou $t_{\alpha/2, n-1}$ par $t_{\alpha, n-1}$.

7.6 Test d'ajustement du χ^2

On cherche à valider ou rejeter un modèle proposé pour X . On teste l'*hypothèse nulle* $H_0 : f_X(x) = f_0(x)$, où $f_X(x)$ est la densité inconnue et $f_0(x)$ la densité proposée, versus la *contre-hypothèse* $H_1 : f_X(x) \neq f_0(x)$.

1. On divise S_X en k classes.
2. On prélève un échantillon de taille n de X .
3. On calcule la statistique

$$D^2 = \sum_{j=1}^k \frac{(n_j - m_j)^2}{m_j},$$

où n_j est l'effectif observé dans la classe j et m_j est l'effectif espéré dans la classe j en supposant H_0 comme étant vraie.

4. On rejette H_0 au seuil de signification α si et seulement si

$$D^2 > \chi_{\alpha, k-r-1}^2,$$

où r est le nombre de paramètres estimés pour $f_0(x)$. Dans le cas contraire, on dit qu'on *ne rejette pas* H_0 .

Remarque. On a $m_j = n \underbrace{\mathbb{P}[X \in C_j | H_0 \text{ vraie}]}_{\text{calcul fait avec } f_0}$, où C_j est la classe j .

Remarque. Il faut que $m_j \geq 5$ pour tout j . Si ce n'est pas le cas, on peut regrouper des classes adjacentes (k diminue).

Exemple 7.15. Soit un test de $H_0 : X \sim \mathcal{U}(0, 5)$ contre $H_1 : X \sim \mathcal{U}(0, 5)$ avec le tableau d'effectifs suivant et $n = 80$.

j	1	2	3
	$[0, 1]$	$]1, 2]$	$]2, 5]$
n_j	20	23	37

On a $m_1 = 16$, $m_2 = 16$ et $m_3 = 48$. Par conséquent, $D^2 \approx 6.58$. On compare D^2 à $\chi_{0.05, 3-0-1}^2 = \chi_{0.05, 2}^2 \approx 5.99$. Par conséquent, on a $D^2 > \chi^2$ et on rejette H_0 .

Exemple 7.16. Soit un test de $H_0 : X \sim \mathcal{B}(3, p)$ contre $H_1 : X \sim \mathcal{B}(3, p)$ au seuil $\alpha = 0.05$ à partir du tableau d'effectif suivant et $n = 100$.

j	0	1	2	3
n_j	33	36	24	7

On estime p par la méthode des moments, soit

$$E[X] = 3p = \bar{X} \Leftrightarrow p = 0.35.$$

On a $m_0 \approx 27.46$, $m_1 \approx 44.36$, $m_2 \approx 23.89$ et $m_3 \approx 4.29$. Puisque $m_3 < 5$, alors on regroupe les classes 2 et 3. On obtient alors le tableau d'effectif suivant.

j	0	1	{2, 3}
n_j	33	36	31
m_j	27.46	44.36	28.18

Par conséquent, $D^2 \approx 2.97$. On compare D^2 à $\chi_{0.05,3-1-1}^2 = \chi_{0.05,1}^2 \approx 3.84$. Puisque $D^2 < \chi_{0.05,1}^2$, alors on ne rejette pas H_0 .

7.7 Tests d'hypothèses

Soit H_0 et H_1 des hypothèses au sujet de $\mu/\mu_1/\mu_2$ ou $\sigma^2/\sigma_1^2/\sigma_2^2$. En général, on pose un énoncé que l'on croit *faux* comme H_0 .

Définition. L'*erreur de première espèce* α est la probabilité de rejeter H_0 sachant que H_0 est vrai, c'est-à-dire

$$\alpha = \mathbb{P}[\text{rejeter } H_0 | H_0 \text{ est vrai}].$$

Remarque. En général, α est une valeur fixée par l'expérimentateur.

Définition. L'*erreur de deuxième espèce* β est la probabilité d'accepter H_0 sachant que H_0 est fausse, c'est-à-dire

$$\beta = \mathbb{P}[\text{accepter } H_0 | H_0 \text{ est fausse}].$$

Remarque. En général, β dépend de α , de n et d'une contre-hypothèse particulière.

Définition. On définit la *puissance du test* comme $1 - \beta$.

Puisque α est fixée à l'avance, on dit que «rejeter H_0 » est une *conclusion forte*. D'une manière similaire, puisque β peut être élevée, on dit que «ne pas rejeter H_0 » est une *conclusion faible*.

Remarque. Dans le cas d'une seule population, on suppose que $X \sim \mathcal{N}(\mu, \sigma^2)$. Dans le cas de deux populations, on suppose que $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

Exemple 7.17. Soit le test d'une moyenne et variance inconnue avec l'échantillon de $n = 8$ notes d'examens :

$$\{75\%, 72\%, 64\%, 81\%, 79\%, 65\%, 82\%, 85\%\}$$

On suppose que $X \sim \mathcal{N}(\mu, \sigma^2)$ comme population. Peut-on affirmer que la moyenne

μ est au-dessus de 75% ? ($\alpha = 0.05$)

On remarque que l'unité de mesure est sans conséquent pour le test. Soit $H_0 : \mu = 75$ et $H_1 : \mu > 75$, une conclusion forte. On a donc

$$t_0 = \frac{\bar{X} - 75}{S/\sqrt{8}} \approx 0.135$$

et

$$t_{\alpha, n-1} = t_{0.05, 7} \approx 1.835.$$

Par conséquent, on rejette si $t_0 > t_{\alpha, n-1}$. Puisque $0.135 \leq 1.835$, on ne rejette pas H_0 , car on ne peut pas affirmer que $\mu > 75$.