Introduction aux statistiques : MA3-3GE

E. Bretin, Z. Smida

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

Introduction

La théorie des probabilités a pour objectif de **modéliser des expériences** où plusieurs résultats sont possibles, mais où leur réalisation n'est pas déterminée à l'avance : lancer de dés, prix d'une action, perturbations sur une ligne téléphonique, files d'attente, etc.
Ceci pour **évaluer les risques** et mettre sur pied **des stratégies** pour "faire face" aux aléas.

Définition 1 (Espace de probabilité)

L'espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ *permet de décrire le problème où*

- Ω *représente l'univers, l'espace des observables.*
- $\mathcal{P}(\Omega)$ *représente les parties de* Ω *(dans le cas discret).*
- $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$ *est une probabilité sur* Ω.

Exemples (univers)

1. Pour le jeu de pile ou face, on prendra $\Omega = \{p, f\}$ ou $\Omega = \{0, 1\}$.
2. Pour une suite de 10 lancers d'une pièce, on prendra $\Omega = \{p, f\}^{10}$, l'ensemble des 10-uplets composés de p et de f.
3. Nombre de lancers d'une pièce avant qu'elle ne tombe sur pile $\Omega = \mathbb{N}^*$ (on peut ajouter $+\infty$ à cet ensemble si on le souhaite).
4. Durée de la prochaine communication téléphonique à une cabine $\Omega = \mathbb{R}^+$.

Les ensembles 3. et 4. ne sont pas de cardinaux finis. On pourra remarquer que Ω peut très bien ne pas être un ensemble de nombres.

Vocabulaire

Définition 2

Evènement élémentaire : *un résultat possible de l'expérience aléatoire.*

Exemples

1. lancer d'un dé à 6 faces non pipé :

   $A = \{$obtenir 6 à l'issue du lancer$\}$.

2. tirage d'une carte dans un jeu de 52 cartes :

   $B = \{$tomber sur le "valet de trèfle"$\}$.

3. pile ou face :

   $C = \{$obtenir "pile" lors d'un jet d'une pièce$\}$.

Définition 3

Évènement : *sous-ensemble de Ω composé d'évènements élémentaires. Un évènement est représenté par un sous-ensemble de l'univers Ω. Si l'une des issues de l'évènement A est réalisé, on dit que A est réalisé.*

Exemples

1. lancer d'un dé à 6 faces :

   E={obtenir un nombre pair strictement supérieur à 3} ={4, 6}

2. $A = \{$le lancer de dé donne 1, 3 ou 5$\} = \{1, 3, 5\}$.

   Dans cet exemple, si le dé tombe sur 3, on dit que A est réalisé.

Probabilité (cas Ω de cardinal fini)

Définition 4 (Probabilité)

Une probabilité $\mathbb{P}$ est une fonction de $\mathcal{P}(\Omega)$ dans $[0; 1]$ telle que :

1. $\mathbb{P}(\Omega) = 1$;
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ *si A et B sont disjoints.*

Propriété 1

1. $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$;
2. $\mathbb{P}(\emptyset) = 0$;
3. $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$;
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;

Exercice 1

Démontrer ces propositions

Dénombrement et équirépartition

Définition 5

On dit que $\mathbb{P}$ est équirépartie sur Ω si pour tout élément $\omega \in \Omega$,

$$\mathbb{P}(\{\omega\}) = \frac{1}{\#\Omega},$$

où $\#\Omega$ représente le cardinal de Ω (nombre d'éléments).

Propriété 2

Si $\mathbb{P}$ est équirépartie sur Ω, alors

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega}$$

Équirépartition ?

Exercice 2

On s'intéresse au jet de deux dés non pipés, différenciables et dont la modélisation conduit à introduire l'univers suivant

$$\Omega = \{(x_1, x_2), x_i \in \{1, 2, 3, 4, 5, 6\}\}$$

- *La loi de probabilité associée est-elle équirépartie ?*
- *Combien d'éléments contient* Ω *?*
- *Déterminer la probabilité d'obtenir au moins* 1 *six avec un lancer de* 2 *dés.*

Exercice 3

Même question lorsque les 2 *dés ne sont pas différenciables avec*

$$\Omega = \{\{x_1, x_2\}, x_i \in \{1, 2, 3, 4, 5, 6\}\}$$

Partition de Ω

Définition 6

Soit $\{A_i\}_{i \in I}$ une partition de Ω si

- $\cup_{i \in I} A_i = \Omega$
- $A_i \cap A_j = \emptyset, \forall (i,j) \in I^2$

Propriété 3

Soit $\{A_i\}_{i \in I}$ une partition de Ω, alors pour tout événement $A \in \mathcal{P}(\Omega)$,

$$P(A) = \sum_{i \in I} P(A \cap A_i)$$

Exercice 4 (Grille)

Un pion se déplace sur une grille case par case suivant les deux directions Est et Nord qui sont prises avec la même probabilité. Si on note $(0, 0)$ *sa position de départ, quelle est la probabilité qu'il passe par le point de coordonnées* $(7, 3)$ *? Que devient cette probabilité si Est est deux fois plus probable que Nord ?*

Exercice 5 (Paradoxe des anniversaires)

Dans votre classe, quelle est la probabilité que deux personnes fêtent leur anniversaire le même jour ?

Exercice 6 (Poker)

Quelle est la probabilité au Poker d'obtenir un carré ? Et un full ?

Cas en dimension "infinie" : l'ensemble des parties de Ω

Définition 7 (Tribu (σ−algèbre):)

Un ensemble de partie de Ω, $\mathcal{F}$ est une tribu sur Ω s'il satisfait aux trois axiomes:

1. $\emptyset \in \mathcal{F}$.
2. *Si $A \in \mathcal{F}$, alors son complémentaire $A^c = \Omega \setminus A$ est aussi dans $\mathcal{F}$.*
3. *Si on a une suite finie ou dénombrable $A_1, \ldots, A_n, \ldots$ d'éléments de $\mathcal{F}$, alors leur réunion $\bigcup_{n\geqslant 1} A_n$ est aussi dans $\mathcal{F}$.*

Définition 8 (Probabilité)

Une probabilité $\mathbb{P}$ *est une fonction de* $\mathcal{F}$ *dans* $[0;1]$ *telle que :*

1. $\mathbb{P}(\Omega) = 1$
2. *(Axiome d'additivité dénombrable:) pour toute suite* $A_1, A_2, \ldots, A_n, \ldots$ *d'événements de* $\mathcal{F}$ *qui sont deux à deux incompatibles, alors la série* $\sum_{k=1}^{+\infty} \mathbb{P}(A_k)$ *converge et a pour somme* $\mathbb{P}\left(\bigcup_{k\geqslant 1} A_k\right)$.

Cas en dimension "infinie" : propriétés

Propriété 4

1. *Si* $A \in \mathcal{F}$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$*;*
2. $\mathbb{P}(\emptyset) = 0$*;*
3. *Si* $A, B \in \mathcal{F}$ *et si* $A \subset B$ *alors* $\mathbb{P}(A) \leqslant \mathbb{P}(B)$*;*
4. *Si* $A, B \in \mathcal{F}$*, alors* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$*;*
5. *Si* $A_1, A_2, \ldots, A_n \in \mathcal{F}$*, alors* $\mathbb{P}(A_1 \cup \cdots \cup A_n) \leqslant \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n)$*;*
6. *(Continuité croissante) Si* $(B_n)_n$ *est une suite croissante d'événements de* $\mathcal{F}$*, alors* $\lim_{n \to +\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n \geqslant 1} A_n\right)$*;*
7. *(Continuité décroissante) Si* $(B_n)_n$ *est une suite décroissante d'événements de* $\mathcal{F}$*, alors* $\lim_{n \to +\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n \geqslant 1} A_n\right)$*.*

Probabilité conditionnelle

Les probabilités conditionnelles ont pour but **d'évaluer "le changement de probabilité"** dû à **l'acquisition d'informations**.

Définition 9

Soient deux événements A et B sur $(\Omega, \mathcal{F}, \mathbb{P})$, avec $\mathbb{P}(B) > 0$. La probabilité conditionnelle de A sachant B est

$$\mathbb{P}_B(A) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Propriété 5

Soient deux événements A et B sur $(\Omega, \mathcal{F}, \mathbb{P})$, avec $\mathbb{P}(B) > 0$ et $\mathbb{P}(A) > 0$, alors

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)}{\mathbb{P}(B)}\mathbb{P}(B|A)$$

Évènements indépendants

Définition 10

Deux événements A et B sont indépendants si $\mathbb{P}(A|B) = \mathbb{P}(A)$ ou de manière équivalente lorsque $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Définition 11

Plus généralement, si $(A_n)_n$ est une suite d'événements, $(A_n)_n$ sont indépendants si pour tout ensemble fini d'indices $I \subset \mathbb{N}$, on a

$$\mathbb{P}\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} \mathbb{P}(A_i).$$

Deux propriétés très utiles

Propriété 6 (Formule des probabilités totales)

Soit $\{A_i\}_{i=1,\dots,n}$ *une partition de* Ω*, telle que pour tout* i, $\mathbb{P}(A_i) > 0$. *Pour tout* $B \in \mathcal{F}$, $\mathbb{P}(B) = \displaystyle\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$.

Propriété 7 (Formule de Bayes)

Soit $\{A_i\}_{i=1,\dots,n}$ *une partition de* Ω *telle que pour tout* i, $\mathbb{P}(A_i) > 0$. *Pour tout* $B \in \mathcal{F}$ *avec* $\mathbb{P}(B) > 0$*, on a pour tout* $k = 1, \dots, n$ *:*

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

Exercice 7 (Mémoire)

Vous vous rendez chez un ami dont vous savez qu'il a deux enfants, ***mais sans savoir s'il s'agit de fille ou de garçon****. Vous y rencontrez un seul de ces enfants qui est* ***un garçon*** *et s'appelle Jacques.*

- *Le risque d'erreur est-il plus grand si vous demandez à Jacques des nouvelles de son frère ? et de sa sœur ?*
- *Vous savez que, vu son âge, Jacques est le cadet. Cette information change-t-elle la réponse précédente ?*

Exercice 8

Soient A, B deux événements tels que $P(A) = 0.4$ et $P(B) = 0.5$. Calculer $P(\overline{A} \cup B)$, $P(A \cap B|A)$, $P(\overline{A}|B)$ dans les trois cas suivants :

1. *A et B sont disjoints ;*
2. *A et B sont indépendants ;*
3. $\mathbb{P}(A \cup B) = 0.8$.

Exercice 9

Soit Ω un ensemble constitué de m boules blanches et n boules noires. On répartit ces boules entre deux urnes U_1 et U_2. U_1 contient r boules blanches et s boules noires. On attribue la probabilité p à l'urne U_1 et $(1-p)$ à l'urne U_2. Une urne étant choisie, on tire une boule, les divers tirages étant équiprobables. Sachant qu'elle est blanche, déterminer la probabilité pour que cette boule provienne de l'urne U_1.

Exercice 10

Considérons une expérience à 3 issues incompatibles x,y *et* z *de probabilités respectives* a, b *et* $1 - (a + b)$. *L'expérience est répétée jusqu'à apparition de* x *ou de* y. *Soit* A *l'évènement "*x *sort avant* y*".*
Montrer que

$$\mathbb{P}(A) = \frac{a}{a+b}.$$

Pour ce faire, vous conditionnerez A *par le résultat de la première expérience.*

Exercice 11 (Hémophilie)

On rappelle que le genre de l'individu est déterminé par la paire de chromosomes 23 (XX pour une femme, XY pour un homme) et par conséquent ce genre est déterminé par le chromosome issu de la paire 23 du père (X ou Y). On supposera pour une paire de chromosomes donnée l'équiprobabilité des 4 combinaisons possibles.

L'hémophilie de type A *est une maladie héréditaire due à une anomalie du facteur VIII du chromosome X. Le chromosome anormal est noté* **Xh**. *Cette maladie est surtout visible chez les hommes car les symptômes sont présents chez les individus dont la paire ne comporte aucun chromosome X sain permettant la coagulation du sang. Ainsi* **les hommes XhY et les femmes XhXh sont dites malades**, *alors que* **les femmes XXh sont dites porteuses saines**.

- *Une femme n'est pas hémophile, mais son frère l'est. Par contre, aucun de leurs géniteurs communs n'est hémophile. Quelles peuvent être les combinaisons possibles pour les paires 23 de ses parents ? Quelle est la probabilité qu'elle soit porteuse saine ?*
- *Elle attend un fils. Quelle est la probabilité qu'il soit hémophile ? S'il ne l'est pas, quelles sont les probabilités qu'elle soit porteuse saine et que son prochain fils soit hémophile ?*

1. Introduction aux probabilités
   - Espace de probabilités et dénombrement
   - Probabilité conditionnelle et événements indépendants
   - Variables aléatoires
   - Vecteurs Aléatoires
   - Espérance et variance conditionnelle
   - Théorème central limite et approximation
2. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison
3. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Variables Aléatoires, définition

Définition 12

*On appelle **variable aléatoire** (v.a.) toute fonction de Ω dans $\mathbb{R}$.*

En général, on désigne les v.a. par des lettres majuscules.
Si la variable est notée X, alors $X(\Omega)$ désigne l'ensemble des valeurs prises par cette variable aléatoire, c'est le **support**.

Exercice 12

Dans le cas où X correspond à la somme d'un lancer de 2 dés, que peut-on dire de $X(\Omega)$?

Loi d'une variable aléatoire discrète

Définition 13

Soit X une v.a. discrète.
$\forall x \in X(\Omega)$,

$$p_X(x) = \mathbb{P}(\{w \in \Omega : X(w) = x\}) = \mathbb{P}(X^{-1}(x)) = \mathbb{P}(X = x)$$

*définit la **loi** de X.*

Propriété 8

- $\forall x \in X(\Omega),\ p_X(x) \geqslant 0.$
- $\displaystyle\sum_{x \in X(\Omega)} p_X(x) = 1.$

Exemples de variables aléatoires discrètes

Exercice 13

Dans le cas où X correspond à la somme d'un lancer de 2 dés, déterminer $\mathbb{P}(X = i)$, pour $i \in \{2, \dots, 12\}$.

Exercice 14

On jette une pièce équilibrée, jusqu'à obtenir pile et on note X le nombre de lancers effectués. X est une v.a. de type dénombrable, susceptible de prendre toutes les valeurs entières non nulles (cf loi géométrique). Montrer que pour tout $n \in \mathbb{N}^*$,

$$\mathbb{P}(X = n) = \frac{1}{2^n}.$$

Densité d'une variable aléatoire continue

Définition 14

X peut prendre toute valeur dans un certain intervalle I.
La loi de probabilité de X est donnée par sa ***densité de probabilité*** *qui est une fonction $f_X : \mathbb{R} \to \mathbb{R}$ définie par la propriété :*

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)\, dx,$$

pour tout $(a, b) \in \mathbb{R}^2$ tels que $a < b$.

Propriété 9

En particulier, il est facile de montrer que

- *f_X est positive.*
- *f_X est intégrable sur $\mathbb{R}$ et $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$.*

Fonction de répartition d'une variable aléatoire continue

Définition 15

Soit X une variable aléatoire. On note F_X la fonction de $\mathbb{R}$ dans $\mathbb{R}$ définie par :

$$\forall x \in \mathbb{R}, \qquad F_X(x) = \mathbb{P}(X \leqslant x).$$

Propriété 10

Si X est une v.a. continue de densité f_X, alors

$$\forall x \in \mathbb{R}, \qquad F_X(x) = \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f_X(t)\, dt,$$

et donc $F_X'(x) = f_X(x)$.

Exemples de variables aléatoires continues

Exercice 15

Soit X une variable aléatoire de loi exponentielle de paramètre $\lambda > 0$ de densité $f_X(x) = \lambda e^{-\lambda x}\mathbb{1}_{\mathcal{R}^+}(x)$. Montrer que X est variable aléatoire sans mémoire

$$P(X \leqslant t + t_0 \mid X \geqslant t_0) = P(X \leqslant t).$$

Exercice 16

Soit X une variable aléatoire de loi uniforme sur $[0,1]$ de densité $f_X(x) = \mathbb{1}_{[0,1]}(x)$.

- *Déterminer la fonction de répartition $F_X(x)$ pour tout $x \in \mathbb{R}$.*
- *Calculer la fonction de répartition de $Y = -\frac{1}{\lambda}\ln(X)$, $\lambda \in \mathbb{R}_+^*$.*

Espérance d'une variable aléatoire

Définition 16 (Cas d'une variable aléatoire discrète)

Soit X une variable aléatoire discrète (avec $X(\Omega) = \{x_i\}_{i\in I}$), alors l'espérance de X notée

$$\mathbb{E}[X] = \sum_{i\in I} x_i P(X = x_i).$$

Définition 17 (Cas d'une variable aléatoire continue)

Soit X une variable aléatoire continue, alors l'espérance de X notée

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\, f_X(x) dx.$$

Exemples

Exercice 17

Déterminer l'espérance d'une variable aléatoire correspondant au lancer d'un dé à 6 faces non pipé.

Exercice 18

On jette une pièce équilibrée, jusqu'à obtenir pile et on note X le nombre de lancers effectués. Déterminer le nombre moyen de lancers effectués dans cette expérience.

Exercice 19

Soit X une variable aléatoire de loi uniforme sur $[0, 1]$. Déterminer l'espérance de X.

Propriété de l'espérance

Propriété 11

Soit X une variable aléatoire et $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction quelconque, alors l'espérance de $Y = \varphi(X)$ vérifie

$$\mathbb{E}(Y) = \mathbb{E}[\varphi(X)] = \begin{cases} \displaystyle\sum_{i \in I} \varphi(x_i) P(X = x_i) & \text{si } X \text{est une v.a discrète} \\ \displaystyle\int_{\mathbb{R}} \varphi(x)\, f_X(x) dx & \text{si } X \text{est une v.a continue} \end{cases}$$

En particulier, si X est une variable aléatoire et $(a, b) \in \mathbb{R}^2$, alors

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b.$$

Variance d'une variable aléatoire

Définition 18

Soit X une variable aléatoire. La variance de X correspond à une mesure de sa dispersion autour de sa moyenne :

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Propriété 12

Soient X une variable aléatoire admettant une espérance alors

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Propriété 13

Soient X une variable aléatoire admettant une espérance et soient $(a, b) \in \mathbb{R}^2$ alors

$$\mathbb{V}[aX + b] = a^2\mathbb{V}[X]$$

Inégalité de Tchebychev

Propriété 14

Soit X une variable aléatoire d'espérance $\mu = E(X)$ et de variance finie $\sigma^2 = V(X)$. Alors pour tout réel strictement positif ε:

$$P(|X-\mu| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{\varepsilon^2}$$

Preuve

On s'intéresse au cas continue, le cas discret étant similaire :

$$\begin{aligned} P(|X-\mu| \geqslant \varepsilon) &= \int_{|X-\mu|\geqslant\varepsilon} f_X(x)dx \\ &= \int_{|X-\mu|\geqslant\varepsilon} \frac{\varepsilon^2}{\varepsilon^2} f_X(x)dx \\ &\leqslant \int_{\mathcal{R}} \frac{(X-\mu)^2}{\varepsilon^2} f_X(x)dx = \frac{\sigma^2}{\varepsilon^2} \end{aligned}$$

Exemples

Exercice 20

Déterminer la variance d'une variable aléatoire correspondant au lancer d'un dé à 6 faces non pipé.

Exercice 21

On jette une pièce équilibrée, jusqu'à obtenir pile et on note X le nombre de lancers effectués. Déterminer la variance de X.

Exercice 22

Soit X une variable aléatoire de loi uniforme sur $[0, 1]$. Déterminer la variance de X.

Exemples

Exercice 23

Soit X une variable aléatoire suivant une loi de Poisson de paramètre λ :

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

Déterminer l'espérance et la variance de X.

Exercice 24

Soit X une variable aléatoire suivant une loi Géométrique de paramètre p:

$$P(X = n) = (1 - p)^{n-1}p, \quad \forall n \in \mathbb{N}^*$$

Déterminer l'espérance et la variance de X.

Fonction caractéristique

Définition 19

La fonction caractéristique d'une variable aléatoire réelle X est la fonction à valeurs complexes définie sur $\mathbb{R}$ par

$$\varphi_X(t) = \mathbb{E}(e^{itX}).$$

Remarque 1

Dans le cas d'une variable aléatoire continue, la fonction caractéristique est la transformée de Fourier inverse de la densité

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \int_{\mathbb{R}} f_X(x) e^{itx} dx$$

Propriété 15

$$\varphi_X'(0) = i\mathbb{E}(X) \quad \text{et} \quad \varphi_X''(0) = -\mathbb{E}(X^2)$$

Exercice 25

Déterminer la fonction caractéristique d'une loi normale $X \sim \mathcal{N}(\mu,\sigma)$ *de densité*

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

Exercice 26

Déterminer la fonction caractéristique d'une loi exponentielle $X \sim \mathcal{E}(\lambda)$ *de densité*

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

1. Introduction aux probabilités
   - Espace de probabilités et dénombrement
   - Probabilité conditionnelle et événements indépendants
   - Variables aléatoires
   - Vecteurs Aléatoires
   - Espérance et variance conditionnelle
   - Théorème central limite et approximation
2. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison
3. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Couples aléatoires

Lorsque le résultat d'une expérience aléatoire est décrit par une suite de n variables aléatoires alors cette suite constitue un vecteur aléatoire de dimension n.
Pour $n = 2$, on parle de couple de variables aléatoires.

Définition 20

Un couple aléatoire (X, Y) *est une application*

$$(X,Y) : \begin{cases} \Omega & \to \mathbb{R}^2 \\ \omega & \to (X(\omega), Y(\omega)), \end{cases}$$

$\forall \omega \in \Omega$, $(X(\omega), Y(\omega))$ *est une réalisation du couple* (X, Y).

Loi d'un couple de variables aléatoires discrètes

Soient (X, Y) deux variables aléatoires discrètes telles que $X(\Omega) = \{x_i\}_{i \in I}$ et $Y(\Omega) = \{y_j\}_{j \in J}$.

Définition 21 (Loi conjointe)

*La **loi conjointe** du couple (X, Y) est définie $\forall (i, j) \in I \times J$ par*

$$p_{(X,Y)}(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j).$$

Définition 22 (Loi marginale)

*Les **lois marginales** de X et de Y sont définies $\forall i \in I$ et $\forall j \in J$ par*

$$p_X(x_i) = \sum_{j \in J} \mathbb{P}(X = x_i, Y = y_j),$$

et

$$p_Y(y_j) = \sum_{i \in I} \mathbb{P}(X = x_i, Y = y_j).$$

Exercice 27

En terminant d'effeuiller une marguerite, on compte 0 *point pour "pas du tout",* 1 *point pour "un peu",* 3 *points pour "beaucoup",* 5 *points pour "passionnément",* 10 *points pour "à la folie". On effeuille successivement deux marguerites. Soit X la v.a. égale au nombre de points obtenus avec la première marguerite. Soit Y la v.a. égale au maximum des deux nombres obtenus.*

- *Déterminer la loi du couple* (X, Y)
- *Préciser les lois marginales de X et de Y.*

Loi d'un couple de variables aléatoires continue

Soient (X, Y) deux variables aléatoires continues

Définition 23 (Fonction de densité conjointe)

La loi de densité de probabilité $f_{(X,Y)}$ est définie $\forall (A_x, A_y) \subset \mathbb{R} \times \mathbb{R}$

$$P(X \in A_x, Y \in A_y) = \int_{A_x} \int_{A_y} f_{(X,Y)}(x, y)dxdy.$$

Propriété 16

La fonction densité de couple $f_{(X,Y)}$ vérifie $f_{(X,Y)} \geqslant 0$ et

$$\int_{\mathbb{R}^2} f_{(X,Y)}(x, y)dxdy = 1.$$

Loi marginale d'un couple de variables aléatoires continue

Définition 24 (Fonction de densité marginale)

Soient (X, Y) deux variables aléatoires continues. Alors, les densités marginales de X et de Y sont définies par :

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)dy,$$

et

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)dx.$$

Exercice 28

Soit (X, Y) un couple de variable aléatoire de densité

$$f_{(X,Y)}(x,y) = \begin{cases} \dfrac{3y}{a^3} & \text{si } (x,y) \in [0,a]^2 \text{ et } y \geqslant x \\ 0 & \text{sinon} \end{cases}$$

Déterminer les lois marginales de X et de Y.

Espérance

Propriété 17

Soient (X, Y) *deux variables aléatoires et* $\varphi : \mathbb{R}^2 \to \mathbb{R}$ *une fonction quelconque, alors*

- *si* (X, Y) *est discret*

$$\mathbb{E}[\varphi(X, Y)] = \sum_{i \in I} \sum_{j \in J} \varphi(x_i, y_j) \mathbb{P}(x_i, y_j)$$

- *si* (X, Y) *est continu*

$$\mathbb{E}[\varphi(X, Y)] = \int_{\mathbb{R}^2} \varphi(x, y) f_{(X,Y)}(x, y) dxdy.$$

Définition 25

Soient X et Y deux variables aléatoires admettant toutes les deux une espérance mathématique. On appelle covariance de X et Y le réel défini par

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

Propriété 18

Soient X et Y deux variables aléatoires et α et β deux réels. Alors,

- $\mathbb{E}[\alpha X + \beta Y] = \alpha\, \mathbb{E}[X] + \beta\, \mathbb{E}[Y]$
- $\mathbb{V}(\alpha X + \beta Y) = \alpha^2 \mathbb{V}(X) + 2\alpha\beta Cov(X, Y) + \beta^2 \mathbb{V}(Y)$

Indépendance, définition

Définition 26

Deux variables aléatoires sont dites indépendantes si

- *si* (X, Y) *est discret*

$$p_{(X,Y)}(x_i, y_j) = p_X(x_i)p_Y(y_j), \forall (i,j) \in I \times J$$

- *si* (X, Y) *est continue*

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y), \forall (x, y)\mathbb{R}$$

Minimum de variables aléatoires indépendantes

Exercice 29

Soit (X_1, X_2) *deux variables indépendantes et équi-distribuées de loi uniforme sur* $[0, 1]$*. Déterminer la loi de la variable aléatoire* Y *définie par* $Y = \min(X_1, X_2)$*.*

Exercice 30

On considère n *variables aléatoires indépendantes* X_i*, de lois exponentielles de paramètre* $\lambda_i > 0$*.*

1) *On note* $Z = \min(X_i)$*. Calculez la fonction de répartition de* Z *et déduisez-en la loi de* Z*.*
2) *La durée de vie des composants électroniques suit une distribution exponentielle. Un dispositif comporte 3 composants de durées de vie moyenne respectives 1000h, 1200h, 1500h. Quelle est la durée de vie moyenne du dispositif ?*

Indépendance, propriétés

Propriété 19

Soint (X, Y) *deux variables indépendantes, alors pour toutes fonctions* $\phi : \mathbb{R} \to \mathbb{R}$ *et* $\psi : \mathbb{R} \to \mathbb{R}$*,*

$$\mathbb{E}[\phi(X)\psi(Y)] = \mathbb{E}[\phi(X)]\mathbb{E}[\psi(Y)].$$

Propriété 20

Si X *et* Y *sont deux variables aléatoires indépendantes, alors*

- *leur covariance est nulle et*

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

- *Sa fonction caractéristique de la somme vérifie*

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$$

- *Sa densité de la somme vérifie* $f_{X+Y}(z) = f_X * f_Y(z)$.

Exercice 31

Soient X_1 et X_2 deux variables aléatoires de loi uniforme sur $[0,1]$, indépendante. Déterminer la densité et la fonction caractéristique de $Z = X_1 + X_2$.

Exercice 32

Soit X_i une suite de lois de Bernoulli de paramètre p indépendantes. On rappelle que X suit une loi de Bernoulli si $X \in \{0,1\}$ avec $P(X=1)=p$ et $P(X=0)=1-p$.

- *Déterminer l'espérance et la variance de X.*
- *En déduire l'espérance et la variance de $Y = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ en fonction de n. Que peut-on en déduire ?*

1. Introduction aux probabilités
   - Espace de probabilités et dénombrement
   - Probabilité conditionnelle et événements indépendants
   - Variables aléatoires
   - Vecteurs Aléatoires
   - Espérance et variance conditionnelle
   - Théorème central limite et approximation
2. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison
3. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Espérance conditionnelle

Définition 27

Soit (X, Y) *un couple de variables aléatoires, soit h définie par*

- *Cas discret :*

$$h(y_j) = \mathbb{E}[X|Y = y_j] = \frac{1}{p_Y(y_i)} \sum_{i \in I} x_i p_{(X,Y)}(x_i, y_j)$$

- *Cas continue*

$$h(y) = \mathbb{E}[X|Y = y] = \frac{1}{f_Y(y)} \int_{\mathbb{R}} x f_{(X,Y)}(x, y) dx$$

Alors, l'espérance conditionné $\mathbb{E}[X|Y]$ *est la variable aléatoire définie par* $\mathbb{E}[X|Y] = h(Y)$.

Propriété 21 (Formule d'espérance totale)

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

Variance conditionnelle

Définition 28

Soit (X, Y) *un couple de variables aléatoires, soit g définie par*

$$g(y) = \mathbb{V}[X|Y = y] = \mathbb{E}[X^2|Y = y] - \mathbb{E}[X|Y = y]^2.$$

Alors Alors, la variance conditionnée $\mathbb{V}[X|Y]$ *s'identifie à la variable aléatoire définie par* $\mathbb{V}[X|Y] = g(Y)$.

Propriété 22 (Théorème de la variance totale)

$$\mathbb{V}[X] = \mathbb{E}[\mathbb{V}[X|Y]] + \mathbb{V}[\mathbb{E}[X|Y]]$$

Exercice 33 (Deux pièces ou plus)

Vous disposez de pièces (deux au départ) dont les deux faces sont numérotées 0 *et* 1*.*

- *Le premier tirage se fait avec deux pièces, son résultats vous est inconnu. Vous devez faire un pari sur la somme S_1 obtenue en additionnant les faces des deux pièces. Quel est votre pari le plus raisonnable et quel est votre risque d'erreur ?*
- *Un deuxième tirage a lieu avec $n = 2 + s_1$ pièces et on note S_2 le résultat obtenu. Calculez $P(S_1 = 1 | S_2 = 2)$*
- *Calculer $\mathbb{E}[S_2]$ et $\mathbb{V}[S_2]$*

Exercice 34 (Assurance)

Une assurance doit rembourser un nombre aléatoire N de sinistres (supposés indépendants). Le montant X d'un sinistre est aléatoire et sa distribution est supposée commune à l'ensemble des sinistres. Calculer l'espérance et l'écart type du montant total T des sinistres en fonction des moments de X et de N.

1. Introduction aux probabilités
   - Espace de probabilités et dénombrement
   - Probabilité conditionnelle et événements indépendants
   - Variables aléatoires
   - Vecteurs Aléatoires
   - Espérance et variance conditionnelle
   - Théorème central limite et approximation
2. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison
3. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Convergence presque sûr

Définition 29 (Convergence presque sûre)

On dit que (X_n) converge presque sûrement vers X si

$$P(lim_{n\to\infty}X_n = X) = 1,$$

ou de manière équivalente, s'il existe un ensemble négligeable $N \subset \Omega$ tel quel

$$\forall\omega \in \Omega \setminus N, \quad X_n(\omega) \to X(\Omega)$$

Loi des grands nombres

Théorème 1

Soit une suite (X_i) *de variables aléatoires indépendantes qui suivent la même loi en probabilité, intégrable (c'est à dire* $\mathbb{E}(|X_i|) < \infty$*). Alors*

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

converge vers $\mathbb{E}(X)$ *presque sûrement*

Convergence en loi

Définition 30 (Convergence en loi)

On dit que (X_n) *converge vers* X *en loi si pour fonction* g *à valeur réelles, continue et bornée*

$$lim_{n\to\infty}\mathbb{E}[g(X_n)] = \mathbb{E}[g(X)].$$

En particulier, dans le cas continue, on en déduit que

$$lim_{n\to\infty}\int g(x)f_{X_n}(x)dx = \int g(x)f_X(x)dx.$$

Théorème central limite

Théorème 2 (TCL)

Soit (X_i) une suite de variables aléatoires réelles, indépendantes et identiquement distribuées suivant la même loi d'espérance μ et de variance finie $\sigma^2 \neq 0$. Alors la suite de variable aléatoire U_n définie par

$$U_n = \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma \sqrt{n}}$$

converge en loi vers une variable aléatoire U de loi normale centrée réduite $\mathcal{N}(0,1)$

Théorème central limite

Preuve

L'idée est d'utiliser la fonction caractéristique φ_X d'une variable aléatoire X:

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = 1 + it\mathbb{E}[X] - \frac{t^2}{2}(\mathbb{V}[X] + \mathbb{E}(X)^2) + o(t^2).$$

Alors, en remarquant que $U_n = \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}$ où $\mathbb{E}[Y_i] = 0$ et $\mathbb{V}[Y_i] = 1$, on en déduit que

$$\begin{aligned} \varphi_{U_n}(t) &= \mathbb{E}[e^{it\sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}}] = \prod_{i=1}^{n} \varphi_{Y_i}(t/\sqrt{n}) \\ &= \left(1 - \frac{t^2}{2n} + o(t^2/n)\right)^n \to e^{-t^2/2} \end{aligned}$$

qui correspond bien à la fonction génératrice d'un loi normale centrée réduite.

Propriété de la loi normale

Propriété 23 (Fonction caractéristique)

Soit X une loi normale d'espérance μ et de variance σ^2, alors

$$\varphi_X(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$$

Propriété 24 (Somme de deux loi normale)

Soient X_1 et X_2 de variables aléatoires indépendantes de loi normale $X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ et $X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$ alors

$$X_1 + X_2 \sim \mathcal{N}\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

Exercice 35 (Dés)

On lance 30 *fois de suite un dé.*

1) *Quelle est la probabilité que la somme X obtenue soit supérieure ou égale à* 100 *?*
2) *Un autre joueur lance* 32 *fois de suite un dé. On note Y la somme obtenue. Quelle est la probabilité de l?évènement $X \leqslant Y$?*

Exercice 36 (Crible)

La taille d?une pièce extraite au hasard d'une production suit une loi normale. On passe les pièces dans deux cribles de dimension 42.00 *mm et* 40.05 *mm. On observe que* 90.3% *et* 34.7% *des pièces passent respectivement dans ces cribles. Déduisez-en μ et σ.*

Exercice 37

1) *On désire construire une voie de garage de longueur 450m dans un chantier. La longueur d'un rail est égale à* $l_n = 9m \pm 0.03m$*. Quel est l'incertitude donnée sur la longueur totale de la voie par un calcul traditionnel ?*
2) *On sait que la longueur d'un rail suit une loi de normale. Quel nouveau calcul d'incertitude peut se substituer au précédent ?*

Propriété 25

Soit $S_n \sim \mathcal{B}(n, \lambda/n)$ une suite de variable de loi Binomiale. Alors S_n converge en loi vers une loi de Poisson de paramètre λ

Preuve

Il suffit de remarquer que $S_n = \sum_{k=1}^{n} X_k$ où $X_k \sim \mathcal{B}(\lambda/n)$ est une suite de variables aléatoires indépendantes identiquement distribuées avec

$$\varphi_{S_n}(t) = \Pi_{i=1}^{n} \varphi_{X_k}(t) = \left(1 + \frac{\lambda}{n}(e^{it} - 1)\right)^n \to e^{\lambda(e^{it}-1)},$$

qui correspond bien à la fonction caractéristique d'une loi de Poisson de paramètre λ.

Exercice 38

Une installation est équipée d'un éclairage continu constitué de 1000 *ampoules dont la durée de fonctionnement suit une loi exponentielle d'espérance* 1200 *heures. Un technicien fait le tour de l'installation tous les jours pour remplacer les lampes défectueuses.*

- *Combien le technicien doit-il emporter d'ampoules dans sa tournée quotidienne pour être à peu près sûr de ne pas manquer d'ampoules?*
- *Si le technicien était absent pendant 5 jours, quel serait en moyenne le pourcentage d'ampoules encore en fonctionnement ?*

1. Introduction aux probabilités
   - Espace de probabilités et dénombrement
   - Probabilité conditionnelle et événements indépendants
   - Variables aléatoires
   - Vecteurs Aléatoires
   - Espérance et variance conditionnelle
   - Théorème central limite et approximation
2. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison
3. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

1. Introduction aux probabilités
   - Espace de probabilités et dénombrement
   - Probabilité conditionnelle et événements indépendants
   - Variables aléatoires
   - Vecteurs Aléatoires
   - Espérance et variance conditionnelle
   - Théorème central limite et approximation
2. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison
3. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Échantillon

Définition 31

On appelle **échantillon de taille** *n de X une suite de* **variables aléatoires** $\{X_i\}_{1\leqslant i\leqslant n}$ *de même loi que X.*

On parle d'échantillon

- **Exhaustif** *: les variables X_i ne sont pas indépendantes les unes des autres*
- **Non exhaustif** *: les variables X_i sont 2 à 2 indépendantes. On parle alors de variables aléatoires indépendantes identiquement distribuées (v.a. i.i.d).*

On note aussi $\{x_i\}_{1\leqslant i\leqslant n}$, **la réalisation** *de cet échantillon*

Motivation

Soit X une variable aléatoire de loi $\mathcal{L}(\theta)$ paramétrée par un paramètre $\theta \in \mathrm{R}^d$.

L'objectif est d'estimer une valeur de θ à partir de la réalisation $\{x_i\}_{1 \leqslant i \leqslant n}$ d'un échantillon indépendant et identiquement distribué $\{X_i\}_{1 \leqslant i \leqslant n}$ de la loi de X,

Définition 32

Un estimateur de θ est une statistique

$$\Theta_n = h(X_1, X_2, \cdots, X_n)$$

dont une réalisation

$$\theta_n = h(x_1, x_2, \cdots, x_n)$$

permet de donner une estimation "plausible" de θ.

Propriétés d'un estimateur

Définition 33

- *Un estimateur est dit sans biais si*

$$\mathbb{E}[\Theta_n] = \theta.$$

- *Un estimateur est dit convergent si*

$$\lim_{n\to\infty} \mathbb{V}[\Theta_n] = 0.$$

Propriété 26

Si $\mathbb{E}[\Theta_n] = \theta$ *et* $\lim_{n\to\infty} \mathbb{V}[\Theta_n] = 0$, *alors*

$$\lim_{n\to\infty} P(|\Theta_n - \theta| > \varepsilon) = 0.$$

Erreur quadratique d'un estimateur

Définition 34

Soit Θ_n un estimateur de θ. L'erreur quadratique de Θ_n est définie comme

$$\mathbb{E}[(\Theta_n - \theta)^2] = (\mathbb{E}[\Theta_n] - \theta)^2 + \mathbb{V}[\Theta_n].$$

Si $\mathbb{E}[(\Theta_n - \theta)^2] \to 0$ quand $n \to \infty$, on dit que Θ_n est un estimateur de θ convergeant en moyenne quadratique.

Pour un estimateur sans biais, la convergence en moyenne quadratique équivaut à la convergence vers 0 de la variance. L'idée est alors de chercher à construire de **estimateurs sans biais** et **de plus petite variance possible**.

Exemple avec une loi de Bernoulli

Soit $\{X_i\}_{1\leqslant i\leqslant n}$ un échantillon indépendant et identiquement distribué de loi de Bernoulli $X_i \sim \mathcal{B}(p)$.

Quel choix pour l'estimateur du paramètre $p \in [0,1]$?

- Espérance et variance :

$$\mathbb{E}(X_i) = p \quad \text{et} \quad \mathbb{V}(X_i) = p(1-p).$$

- **Choix de l'estimateur:** moyenne de l'échantillon

$$\mathbf{P_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

- **Un estimateur sans biais**

$$\mathbb{E}[P_n] = \mathbb{E}[X_i] = p,$$

- **Un estimateur convergeant:**

$$\mathbb{V}[P_n] = \frac{p(1-p)}{n} \to 0 \text{ quand } n \to \infty.$$

Estimation de l'espérance et de la variance

Soit $\{X_i\}_{1 \leqslant i \leqslant n}$ un échantillon indépendant et identiquement distribué de loi d'espérance $\mathbb{E}(X_i) = \mu$ et de variance $\mathbb{V}(X_i) = \sigma^2$.

Quel choix pour l'estimateur du paramètre $\mu \in \mathbb{R}$?

- **Moyenne de l'échantillon**

$$\overline{\mathbf{X}}_{\mathbf{n}} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathbf{X}_{\mathbf{i}}$$

- **Un estimateur sans biais et convergeant:**

$$\mathbb{E}[\overline{X}_n] = \mu \text{ et } \mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}.$$

Quel choix pour l'estimateur du paramètre $\sigma^2 \in \mathbb{R}^+$?

- **Variance empirique**

$$\mathbf{S_n^2} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \left(\mathbf{X_i} - \overline{\mathbf{X}}_{\mathbf{n}}\right)^{\mathbf{2}}.$$

- **Un estimateur biaisé !**

$$\begin{aligned}
\mathbb{E}[S_n^2] &= \mathbb{E}[(X_1 - \overline{X}_n)^2] = \frac{(n-1)^2}{n^2}\mathbb{E}[(X_1 - \frac{1}{n-1}\sum_{i=2}^{n} X_i)^2] \\
&= \frac{(n-1)^2}{n^2}\mathbb{E}[((X_1-\mu) - (\frac{1}{n-1}\sum_{i=2}^{n} X_i - \mu))^2] \\
&= \frac{(n-1)^2}{n^2}\left(\mathbb{E}[(X_1-\mu)^2)] + \mathbb{E}[(\frac{1}{n-1}\sum_{i=2}^{n} X_i - \mu))^2] - 2\mathbb{E}[(X_1-\mu)(\frac{1}{n-1}\sum_{i=2}^{n} X_i - \mu)]\right) \\
&= \frac{(n-1)^2}{n^2}(\sigma^2(1+1/(n-1))) = \frac{n-1}{n}\sigma^2
\end{aligned}$$

- **Variance empirique débaisée**

$$\mathbf{S_n'^2} = \frac{\mathbf{1}}{\mathbf{n-1}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \left(\mathbf{X_i} - \overline{\mathbf{X}}_{\mathbf{n}}\right)^{\mathbf{2}} \text{ avec } \mathbb{E}[S_n'^2] = \sigma^2$$

Méthode des moments

Soit $\{X_i\}_{1 \leqslant i \leqslant n}$ un échantillon indépendant et identiquement distribué de loi $\mathcal{L}(\theta)$

- **Exprimer les moments de X_i en fonction de θ**

$$\mathbb{E}[X_i] = g(\theta), \quad \mathbb{V}[X_i] = h(\theta).$$

- **Exprimer θ en fonction des moments de X_i**

$$\theta = m(\mathbb{E}[X_i], \mathbb{V}[X_i]),$$

- **Considérer l'estimateur Θ_n suivant :**

$$\Theta_n = m(\overline{X_n}, S_n'^2).$$

Exemple avec une loi géométrique

Soit $\{X_i\}_{1 \leqslant i \leqslant n}$ un échantillon indépendant et identiquement distribué de loi géométrique $\mathcal{G}(p)$ avec $p \in [0, 1]$.

- **Loi de** X_i

$$P(X_i = k) = (1 - p)^{k-1} p$$

- **Moments** de X_i

$$\mathbb{E}[X_i] = \sum_{k=1}^{\infty} k(1 - p)^{k-1} p = \frac{1}{p}$$

- **Estimateur** de la méthode des moments pour p:

$$P_n = \frac{1}{\overline{X}_n}$$

- **Etude des propriétés** de ces 2 estimateurs ?

Exemple avec une loi uniforme

Soit $\{X_i\}_{1 \leqslant i \leqslant n}$ un échantillon indépendant et identiquement distribué de loi uniforme $\mathcal{U}([a,b])$ avec $b > a$.

- Estimation de ces 2 premiers moments:

$$\mathbb{E}[X_i] = \frac{a+b}{2} \text{ et } \mathbb{V}[X_i] = (b-a)^2/12$$

- Estimation de a et b en fonction de $\mathbb{E}[X_i]$ et $\mathbb{V}[X_i]$:

$$a = \mathbb{E}[X_i] - \sqrt{3\mathbb{V}[X_i]} \text{ et } b = \mathbb{E}[X_i] + \sqrt{3\mathbb{V}[X_i]}$$

- Exemple d'estimateur par la méthode des moments

$$A_n = \overline{X}_n - \sqrt{3S_n'^2} \text{ et } B_n = \overline{X}_n + \sqrt{3S_n'^2}.$$

- Etude des propriétés de ces 2 estimateurs ?

Fonction de vraisemblance

Soit $\{X_i\}_{1 \leqslant i \leqslant n}$ un échantillon indépendant et identiquement distribué d'une loi $\mathcal{L}(\theta)$.

Définition 35

On appelle ***la vraisemblance*** *de la réalisation* $\{x_i\}_{1 \leqslant i \leqslant n}$ *d'un échantillon* $\{x_i\}_{1 \leqslant i \leqslant n}$ *la quantité*

$$\mathcal{L}(\theta, \{x_1, x_2, \cdots x_n\}) = \begin{cases} \Pi_{i=1}^{n} P_\theta(X_i = x_i) & \textit{si la loi est discrète} \\ \Pi_{i=1}^{n} f_\theta(x_i) & \textit{si la loi est continue} \end{cases}$$

Estimateur de maximum de vraisemblance

Définition 36

La méthode du maximum de vraisemblance *consiste alors à déterminer la réalisation* $\hat{\theta}_n$ *de l'estimateur* $\hat{\Theta}_n$ *qui maximise la vraisemblance:*

$$\hat{\theta}_n = \arg\max_{\theta} \{\mathcal{L}(\theta, \{x_1, x_2, \cdots x_n\})\}$$

Propriété 27

Si Θ_n *est l'estimateur du maximum de vraisemblance pour le paramètre* θ, *alors* $h(\Theta_n)$ *est l'estimateur du maximum de vraisemblance pour le paramètre* $h(\theta)$.

Exemple avec une loi géométrique

Soit $\{X_i\}_{1 \leqslant i \leqslant n}$ un échantillon indépendant et identiquement distribué de loi géométrique $\mathcal{G}(p)$ avec $p \in [0, 1]$.

- **Expression de la vraisemblance $\mathcal{L}$**

$$\mathcal{L}(p, \{x_1, x_2, \cdots x_n\}) = \Pi_{i=1}^{n}(1-p)^{x_i - 1}p = \frac{p^n}{(1-p)^n}(1-p)^{\sum_{i=1}^{n} x_i}$$

- **Calcul du log de $\mathcal{L}$:**

$$log(\mathcal{L}(p, \{x_1, x_2, \cdots x_n\})) = n\ log(p) - n\ log(1-p) + \sum_{i=1}^{n} x_i\ log(1-p)$$

- **Dérivation par rapport à p**

$$\partial_p[log(\mathcal{L}(p, \{x_1, x_2, \cdots x_n\}))] = n/p + n/(1-p) + \sum_{i=1}^{n} x_i/(1-p)$$

Exemple avec une loi géométrique

- **Résolution de l'équation** $\partial_p[log(\mathcal{L}(p, \{x_1, x_2, \cdots x_n\}))] = 0,$

$$n/\hat{p}_n + n/(1 - \hat{p}_n) + \sum_{i=1}^{n} x_i/(1 - \hat{p}_n) = 0 \Leftrightarrow \hat{p}_n = \frac{n}{\sum_{i=1}^{n} x_i}$$

- **Déduction de l'estimateur de maximum de vraisemblance**

$$\hat{P}_n = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}_n}$$

Exemple avec une loi uniforme

Soit $\{X_i\}_{1 \leqslant i \leqslant n}$ un échantillon indépendant et identiquement distribué de loi uniforme $\mathcal{U}([a,b])$ avec $b > a$.

- **Expression de la vraisemblance $\mathcal{L}$**

$$\mathcal{L}(a, b, \{x_1, x_2, \cdots x_n\}) = \frac{1}{(b-a)^n} \mathbb{1}_{[a,b]}(x_i)$$

  En particulier

$$\mathcal{L}(a, b, \{x_1, x_2, \cdots x_n\}) = \begin{cases} 0 & \text{si } a > \min x_i \text{ ou } b < \max x_i \\ \dfrac{1}{(b-a)^n} & \text{si } a \leqslant \min x_i \text{ et } b \geqslant \max x_i \end{cases}$$

  On en déduit que $\hat{a}_n = \min\limits_{i=1:n} x_i$ et $\hat{b}_n = \max\limits_{i=1:n} x_i$

- **Estimateur de maximum de vraisemblance** de a et b

$$\hat{A}_n = \min_{i=1:n} X_i \text{ et } \hat{B}_n = \max_{i=1:n} X_i$$

Exercice 39

Soit $\{X_i\}_{1\leqslant i\leqslant n}$ *un échantillon indépendant et identiquement distribué d'une loi normale* $\mathcal{N}(\mu,\sigma^2)$.

1) *Calculer l'estimateur du maximum de vraisemblance pour* μ.
2) *Calculez l'estimateur du maximum de vraisemblance pour vraisemblance pour* σ^2. *Quel est l'estimation du maximum de vraisemblance pour* σ *?*

Exercice 40

Soit $\{X_i\}_{1 \leqslant i \leqslant n}$ *un échantillon indépendant et identiquement distribué d'une loi uniforme* $\mathcal{U}([0, a])$ *avec* $a > 0$.

1) *Montrer que la méthode des moments conduit à un estimateur* $\overline{A}_n = 2\overline{X_n}$ *qui est non biaisé et de variance* $\mathbb{V}[\overline{A}_n] = a^2/(3n)$

2) *Montrer que la méthode de maximum de vraisemblance conduit à l'estimateur* $\hat{A}_n = \max_{i=1:n} X_i$ *dont les moments s'identifient à*

$$\mathbb{E}[\hat{A}_n] = a\frac{n}{n+1}, \text{ et } \mathbb{V}[\hat{A}_n] = a^2\frac{n}{(n+1)^2(n+2)}.$$

3) *En déduire le meilleur estimateur du paramètre* a *?*

Exercice 41

On sait qu'une durée T vérifie qu'il existe $d > 0$ tel que $X = T - d$ suit une loi exponentielle $\mathcal{E}(\lambda)$. On considère un échantillon $\{T_i\}_{1\leqslant i\leqslant n}$ de taille n de la variable aléatoire T.

1) *Soit $A_n = \min\limits_{1\leqslant i\leqslant n}(T_i)$. Donner le domaine de A_n, sa densité, son espérance et sa variance.*
2) *On note $B_n = \overline{T}_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} T_i$. Donner une approximation de la loi de B_n pour n grand. Calculer son espérance et son écart type.*
3) *Déterminer la vraisemblance de l'échantillon en fonction de λ et d.*
4) *Si λ est fixé, pour quelle valeur $\hat{d}_n$ de d cette vraisemblance est-elle maximale ?*
5) *Calculer en fonction de d l'estimation du maximum de vraisemblance pour le paramètre λ.*
6) *En déduire l'estimation du maximum de vraisemblance pour le couple (λ, d) et déterminer le biais de l'estimateur pour d.*

1. Introduction aux probabilités
   - Espace de probabilités et dénombrement
   - Probabilité conditionnelle et événements indépendants
   - Variables aléatoires
   - Vecteurs Aléatoires
   - Espérance et variance conditionnelle
   - Théorème central limite et approximation
2. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison
3. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Motivation

Soit $\{X_i\}_{1 \leqslant i \leqslant n}$ un échantillon indépendant et identiquement distribué de loi $\mathcal{L}(\theta)$.

Nous avons vu précédemment comment obtenir un **estimateur** Θ_n du paramètre θ dont la réalisation à n fixé permet d'obtenir une valeur plausible du paramètre θ avec **une marge d'erreur non contôlée**. L'objectif de cette partie est de construire un **intervalle** dans lequel le paramètre θ a une **forte probabilité de se trouver**.

Définition 37

Soit θ un paramètre donné. On appelle **intervalle de confiance** *de niveau β un intervalle aléatoire $[T_1, T_2]$ tel que*

$$P(\theta \in [T_1, T_2]) = \beta.$$

Exemple pour une proportion

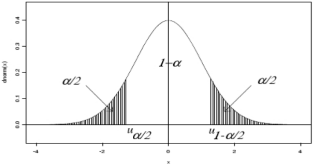
Soit $\{X_i\}_{1\leqslant i\leqslant n}$ un échantillon indépendant et identiquement distribué de loi de Bernoulli $X_i \sim \mathcal{B}(p)$.

- **Estimateur ponctuel** de p: $P_n = \overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.
- **Théorème central limite**

$$\frac{\sum_{i=1}^{n} X_i - n\mathbb{E}[X]}{\sqrt{n\mathbb{V}[X]}} = \frac{\overline{X}_n - p}{\sqrt{p(1-p)/n}} \to U \sim \mathcal{N}(0,1)$$

- **Quantile de la loi normale:**

$$P(U \leqslant u_\alpha) = \alpha$$



- **Propriété de symétrie** : $u_\alpha = -u_{1-\alpha}$.

Table de loi :

Quantile de loi normale $\alpha \mapsto u_\alpha$.

	[0.000]	[0.005]	[0.010]	[0.015]	[0.020]	[0.025]	[0.030]	[0.035]	[0.040]	[0.045]
[0.50]	0.0000	0.0125	0.0251	0.0376	0.0502	0.0627	0.0753	0.0878	0.1004	0.1130
[0.55]	0.1257	0.1383	0.1510	0.1637	0.1764	0.1891	0.2019	0.2147	0.2275	0.2404
[0.60]	0.2533	0.2663	0.2793	0.2924	0.3055	0.3186	0.3319	0.3451	0.3585	0.3719
[0.65]	0.3853	0.3989	0.4125	0.4261	0.4399	0.4538	0.4677	0.4817	0.4959	0.5101
[0.70]	0.5244	0.5388	0.5534	0.5681	0.5828	0.5978	0.6128	0.6280	0.6433	0.6588
[0.75]	0.6745	0.6903	0.7063	0.7225	0.7388	0.7554	0.7722	0.7892	0.8064	0.8239
[0.80]	0.8416	0.8596	0.8779	0.8965	0.9154	0.9346	0.9542	0.9741	0.9945	1.0152
[0.85]	1.0364	1.0581	1.0803	1.1031	1.1264	1.1503	1.1750	1.2004	1.2265	1.2536
[0.90]	1.2816	1.3106	1.3408	1.3722	1.4051	1.4395	1.4758	1.5141	1.5548	1.5982
[0.95]	1.6449	1.6954	1.7507	1.8119	1.8808	1.9600	2.0537	2.1701	2.3263	2.5758

Exemple pour une proportion

- **Intervalle centrée en** 0 **pour** U:

$$P\left(U^2 \leqslant u^2_{\alpha/2}\right) = 1-\alpha,\ \text{avec } U \simeq \frac{\overline{X_n}-p}{\sqrt{p(1-p)/n}}$$

- **équivalence**

$$\begin{aligned} U^2 \leqslant u^2_{\alpha/2} \quad & \Leftrightarrow \quad (\overline{X_n}-p)^2 \leqslant \frac{p(1-p)}{n} u^2_{\alpha/2} \\ & \Leftrightarrow \quad (1+\frac{u^2_{\alpha/2}}{n})p^2 - (2\overline{X_n}+\frac{u^2_{\alpha/2}}{n})p + \overline{X_n}^2 \leqslant 0 \\ & \Leftrightarrow \quad p \in [\Lambda_n^-, \Lambda_n^+] \text{ où } \Lambda_n^{\pm} = \frac{-b \pm \sqrt{b^2-4ac}}{2a} \end{aligned}$$

- **Intervalle de confiance pour** p**:**

$$P(p \in [\Lambda_n^-, \Lambda_n^+]) \simeq 1-\alpha$$

Exemple pour une proportion

- **Intervalle de confiance** en utilisant l'approximation $p \simeq \overline{X_n}$:

$$P(p \in [\Lambda_n^-, \Lambda_n^+]) \simeq 1 - \alpha$$

avec

$$\Lambda_n^{\pm} = \overline{X_n} \pm \frac{|u_{\alpha/2}|}{\sqrt{n}} \sqrt{\overline{X_n}(1 - \overline{X_n})}$$

Remarque 2

La taille caractéristique de cet intervalle est de l'ordre de $|u_{\alpha/2}|/\sqrt{n}$.

- *L'intervalle s'agrandit lorsque α décroît (et donc que la probabilité d'erreur diminue)*
- *L'intervalle se rétrécit lorsque n augmente avec une taille caractéristique en $1/\sqrt{n}$*

Exercice 42

Une entreprise hésite à changer son logo et lance un sondage parmi ses salariés pour savoir quel logo plaît le plus. Le taux de participation au sondage est de 67%*. Parmi les* 1011 *réponses,* 51% *des personnes interrogées déclaraient préférer le deuxième logo.*

1) *Donner un intervalle de confiance asymptotique de niveau* 95% *de la proportion de salariés qui préfèrent le second logo. Commenter.*
2) *Combien de personnes faudrait-il pour avoir une largeur d'intervalle* 0.02 *avec des données similaires.*
3) *Reprendre la question* 1 *pour un intervalle de confiance de niveau* 99%*.*

Cas de la loi normale

Soit $\{X_i\}_{1\leqslant i\leqslant n}$ un échantillon indépendant et identiquement distribué de loi normale $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

1er cas: intervalle de confiance pour μ avec σ connu

- **Moyenne empirique**

$$\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n) \Rightarrow \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = U \sim \mathcal{N}(0,1).$$

- **Intervalle pour une loi normale centrée réduite**

$$P(U \in [u_{\alpha/2}, u_{1-\alpha/2}]) = 1 - \alpha$$

- **Intervalle de confiance pour μ**

$$U \in [u_{\alpha/2}, u_{1-\alpha/2}] \Leftrightarrow \mu \in [\overline{X}_n + \frac{u_{\alpha/2}}{\sqrt{n}}\sigma, \overline{X}_n + \frac{u_{1-\alpha/2}}{\sqrt{n}}\sigma]$$

Cas de la loi normale

2ème cas: intervalle de confiance pour σ^2 avec σ inconnu

- **Moyenne empirique**

$$\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n) \Rightarrow \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = U \sim \mathcal{N}(0,1).$$

- **Utiliser la variance empirique** $S_n'^2$ à la place de σ^2

$$S_n'^2 = \frac{1}{n-1} \sum (X_i - \overline{X}_n)^2 \simeq \sigma^2$$

- **Quelle loi est la loi de** T_{n-1} **?**

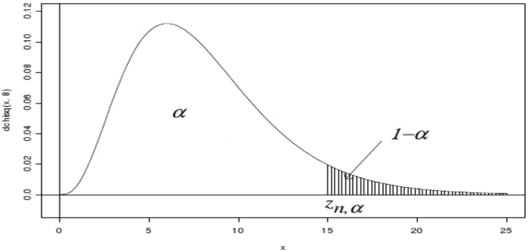
$$T_{n-1} = \frac{\overline{X}_n - \mu}{\sqrt{S_n'^2/n}}$$

Définition 38

Soient $X_1, \cdots, X_n$ indépendantes et de même loi $\mathcal{N}(\mu, \sigma^2)$, alors
$Z = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$ *suit une loi du χ^2 à n degré de liberté.*

Quantile de la loi du χ^2:

$$P(Z \leqslant z_{n,\alpha}) = \alpha$$

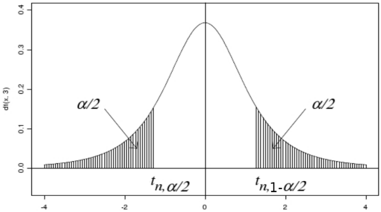


Définition 39

Si $U \sim \mathcal{N}(0,1)$ et $Z \sim \chi^2_n$, et si U et Z sont indépendants, alors $T = \dfrac{U}{\sqrt{Z/n}} \sim St(n)$ suit une loi de Student à à n degré de liberté.

Quantile de la loi du $St(n)$:

$$P(T \leqslant t_{n,\alpha}) = \alpha$$



Propriété de symétrie: $t_{n,\alpha} = t_{n,1-\alpha}$

Théorème de Fisher

Propriété 28

Soient $X_1, \cdots, X_n$ indépendantes et de même loi $\mathcal{N}(\mu, \sigma^2)$.

- $\overline{X_n} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ *et* $S_n'^2 = \dfrac{1}{n-1}\sum (X_i - \overline{X_n})^2$ *sont indépendants*
- $U = \dfrac{\overline{X_n} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0,1)$
- $Z_n = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2_n$
- $Z_{n-1} = \dfrac{1}{\sigma^2}(n-1)S_n'^2 \sim \chi^2_{n-1}$
- $\dfrac{\overline{X_n} - \mu}{\sqrt{S_n'^2 / n}} = \dfrac{U}{\sqrt{Z_{n-1}/(n-1)}} \sim St(n-1)$

Cas de la loi normale

Retour au 2ème cas: intervalle de confiance pour σ^2 avec σ inconnu

- **Moyenne empirique:**

$$\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n) \Rightarrow \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = U \sim \mathcal{N}(0,1).$$

- **Quelle loi est la loi de T_{n-1} ?**

$$T_{n-1} = \frac{\overline{X}_n - \mu}{\sqrt{S_n'^2/n}} \sim St(n-1)$$

- **Intervalle** pour T_{n-1}

$$P(T_{n-1} \in [t_{n-1,\alpha/2}, t_{n-1,1-\alpha/2}]) = 1 - \alpha$$

- **Intervalle de confiance pour μ**

$$T_{n-1} \in [t_{n-1,\alpha/2}, t_{n-1,1-\alpha/2}] \Leftrightarrow \mu \in [\overline{X}_n + \frac{t_{n-1,\alpha/2}}{\sqrt{n}} S_n', \overline{X}_n + \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}} S_n']$$

Cas de la loi normale

3ème cas: intervalle de confiance pour σ^2 avec μ inconnu

- **Variance empirique**

$$S_n'^2 = \frac{1}{n-1}\sum (X_i - \overline{X_n})^2 \text{ et } \frac{1}{\sigma^2}(n-1)S_n'^2 = Z_{n-1} \sim \chi^2_{n-1}$$

- **Intervalle** pour Z_{n-1}

$$P(Z_{n-1} \in [z_{n-1,\alpha/2}, z_{n-1,1-\alpha/2}]) = 1-\alpha$$

- **Intervalle de confiance pour** σ^2

$$P\left(\sigma^2 \in \left[\frac{S_n'^2}{z_{n-1,1-\alpha/2}/(n-1)}, \frac{S_n'^2}{z_{n-1,\alpha/2}/(n-1)}\right]\right) = 1-\alpha$$

Exercice 43

Une chaîne de fabrication est dédiée à la production de yaourts. Une doseuse remplit les pots et la masse ainsi versée suit une loi normale d'espérance µ et d'écart type σ. Chaque heure, un contrôle portant sur 16 yaourts est effectué par le fabricant pour contrôler la valeur de μ. Les valeurs trouvées sont les suivantes :

$$124.5, 121.6, 125.5, 125.3, 125.6, 124.6, 116.8, 124.3,$$
$$127.9, 124.8, 128.2, 118.9, 130.0, 121.2, 123.8, 130.6$$

Au vu de cet échantillon.

1) *Calculer un intervalle de confiance pour μ de niveau $\alpha = 10\%$*
2) *Calculer un intervalle de confiance pour σ^2 de niveau $\alpha = 10\%$*

Exercice 44

Deux étudiants désirent estimer le paramètre a d'une loi uniforme sur l'intervalle $[0, a]$. *Ils disposent pour cela d'un échantillon* $\{X_1, X_2, \cdot X_n\}$ *de taille* $n = 50$.

- *Le premier propose d'utiliser l'estimateur* $A_n^1 = 2\overline{X}_n$
- *Le second propose d'utiliser l'estimateur* $A_n^2 = \max_{i=1:n} X_i$

1) *Donner l'expression des deux intervalles de confiances sur a associés à chacune de ces statistique pour un niveau de confiance* $1 - \alpha$.
2) *En déduire les 2 intervalles correspondants pour* $\alpha = 10\%$ *et dans le cas où* $\overline{x}_n = 18.7$ *et* $m_n = 35.0$.

1. Introduction aux probabilités
   - Espace de probabilités et dénombrement
   - Probabilité conditionnelle et événements indépendants
   - Variables aléatoires
   - Vecteurs Aléatoires
   - Espérance et variance conditionnelle
   - Théorème central limite et approximation
2. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison
3. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Objectif et principe

Un test d'hypothèse a pour **objectif** de décider sur la base d'un **échantillon** si la loi mère vérifie ou non une **hypothèse** sur un de ses **paramètres** caractéristiques.

Démarche:

- Choix de **l'hypothèse** H_0 à tester (et de l'hypothèse alternative H_1)
- Choix d'une **statistique** sur un échantillon qui permet de tester H_0
- Choix d'un niveau de signification : **le risque** α de rejeter H_0 alors qu'elle est vraie
- Définition de **la zone d'acceptation** de H_0 et de rejet de H_0 (intervalle unilatéral ou bilatéral)

Hypothèses H_0, H_1 et risques

		Réalité	
		H_0 vraie	H_0 fausse
Conclusion du test	Non rejet de H_0	Ok	Risque β
	Rejet de H_0	Risque α	Ok

- **Risque α:** risque de rejeter à tort H_0. On rejette H_0 si l'échantillon a une très faible probabilité de se produire
- **Risque β:** risque de ne pas rejeter H_0 alors que H_1 est vrai.
- **la puissance du test** $1-\beta$: pour un risque α donné, on essaie de choisir un test avec un risque β le plus petit possible en jouant sur **la statistique de test** ou encore sur la forme de **l'intervalle d'acceptation** de H_0.

Test d'hypothèse sur une proportion

Soit $\{X_i\}_{1 \leqslant i \leqslant n}$ un échantillon indépendant et identiquement distribué de loi de Bernoulli $X_i \sim \mathcal{B}(p)$.

- **Exemple de choix d'hypothèses**:

$$H_0 = "p = p_0" \text{ contre } H_1 = "p \neq p_0"$$

- **Choix d'une statistique** de test: $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.
- En supposant H_0 vraie, **une zone d'admissibilité** correspond à un $[\ell_\alpha^-, \ell_\alpha^+]$ telle que

$$P(\overline{X}_n \in [\ell_\alpha^-, \ell_\alpha^+]) = 1 - \alpha.$$

En particulier, si H_0 est vraie, et pour n suffisamment grand,

$$\frac{\overline{X}_n - p_0}{\sqrt{p_0(1 - p_0)/n}} \simeq U \sim \mathcal{N}(0, 1).$$

- Utilisation **des quantiles de la loi normale**

$$P(u_{\alpha/2} \leqslant U \leqslant u_{1-\alpha/2}) = 1 - \alpha \Leftrightarrow P(\overline{X}_n \in [\ell_\alpha^-, \ell_\alpha^+]) \simeq 1 - \alpha,$$

avec

$$\ell_\alpha^- = p_0 + u_{\alpha/2}\sqrt{p_0(1-p_0)/n} \text{ et } \ell_\alpha^- = p_0 + u_{1-\alpha/2}\sqrt{p_0(1-p_0)/n}$$

- **La construction du test est terminé**. Il suffit maintenant de vérifier si la réalisation $\overline{x}_n$ de la statistique $\overline{X}_n$ appartient ou pas à l'intervalle $[\ell_\alpha^-, \ell_\alpha^+]$ pour ne pas rejeter ou rejeter H_0.

Test d'hypothèse sur une proportion

Soit $\{X_i\}_{1\leqslant i\leqslant n}$ un échantillon indépendant et identiquement distribué de loi de Bernoulli $X_i \sim \mathcal{B}(p)$.

- Dans certaine application, **il peut aussi être intéressant de tester**

$$H_0 = "p = p_0" \text{ contre } H_1 = "p = p_1 > p_0"$$

- On peut alors garder la statistique $\overline{X}_n$.
- Il est dans ce cas plus intéressant de considérer une **zone d'admissibilité** de la forme $[-\infty, \ell_\alpha^+]$ telle que

$$P(\overline{X}_n \leqslant \ell_\alpha^+) = 1 - \alpha.$$

Comme précédemment, sous l'hypothèseH_0, et pour n suffisamment grand,

$$\frac{\overline{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}} \simeq U \sim \mathcal{N}(0,1).$$

- Utilisation des **quantiles de la loi normale**

$$P(U \leqslant u_{1-\alpha}) = 1-\alpha \Leftrightarrow P(\overline{X}_n \leqslant \ell_\alpha^+) \simeq 1-\alpha,$$

avec

$$\ell_\alpha^+ = p_0 + u_{1-\alpha}\sqrt{p_0(1-p_0)/n}$$

- Il est alors possible de déterminer **le risque** β. En effet, sous l'hypothèse H_1,

$$\frac{\overline{X}_n - p_1}{\sqrt{p_1(1-p_1)/n}} \simeq U \sim \mathcal{N}(0,1),$$

et alors

$$\begin{aligned}\beta &= P(\overline{X}_n \leqslant p_0 + u_{1-\alpha}\sqrt{p_0(1-p_0)/n})\\ &\simeq P\left(U \leqslant \frac{p_0 - p_1 + u_{1-\alpha}\sqrt{p_0(1-p_0)/n)}}{\sqrt{p_1(1-p_1)/n}}\right)\end{aligned}$$

Exercice 45

Pièce équilibré

On désire vérifier qu'une pièce n'avantage pas face avec un nombre minimum de lancers. Pour ce faire, on souhaite réaliser un test avec :

- $H_0 = "p = 0.5"$ *contre* $H_1 = "p = 0.51"$
- $\alpha = 0.2$ *et* $\beta = 0.4$

Construire un test simple répondant à ces critères.

Pièce défectueuse

On désire vérifier la proportion p de pièces défectueuses d'une production. Pour ce faire, on souhaite réaliser un test avec :

- $H_0 = "p = 0.1"$ *contre* $H_1 = "p = 0.2"$
- $\alpha = 0.2$ *et* $\beta = 0.4$

Construire un test simple répondant à ces critères et calculez β*. Répondre aux mêmes questions si* $n = 60$*.*

Test sur l'espérance d'une loi normale

Soit $\{X_i\}_{1 \leqslant i \leqslant n}$ un échantillon indépendant et identiquement distribué de loi normale $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

- Exemple de **choix d'hypothèses:**

$$H_0 = "\mu = \mu_0" \text{ contre } H_1 = "\mu \neq \mu_0"$$

- Choix **d'une statistique de test**: $\overline{X}_n$ et $S_n'^2$.
- Si H_0 est vraie,

$$\frac{\overline{X}_n - p_0}{\sqrt{S_n'^2 / n}} = T_{n-1} \sim St(n-1).$$

- **Quantile pour la loi de Student**

$$P(t_{n-1,\alpha/2} \leqslant T_{n-1} \leqslant t_{n-1,1-\alpha/2}) = 1 - \alpha.$$

- **Zone d'admissibilité** de H_0 :

$$\frac{\overline{X}_n - p_0}{\sqrt{S_n'^2 / n}} \in [t_{n-1,\alpha/2}, t_{n-1,1-\alpha/2}].$$

Test sur la variance d'une loi normale

Soit $\{X_i\}_{1\leqslant i\leqslant n}$ un échantillon indépendant et identiquement distribué de loi normale $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

- Exemple **de choix d'hypothèses**:

$$H_0 = "\sigma = \sigma_0" \text{ contre } H_1 = "\sigma \neq \sigma_0"$$

- **Choix d'une statistique** de test: $S_n'^2$.
- Si H_0 est vraie,

$$\frac{(n-1)S_n'^2}{\sigma_0^2} = Z_{n-1} \sim \chi^2_{n-1}.$$

- **Quantile pour la loi de Chi-deux:**

$$P(z_{n-1,\alpha/2} \leqslant Z_{n-1} \leqslant z_{n-1,1-\alpha/2}) = 1 - \alpha.$$

- **Zone d'admissibilité** de H_0:

$$S_n'^2 \in \left[\frac{\sigma_0^2\ t_{n-1,\alpha/2}}{n-1}, \frac{\sigma_0^2\ t_{n-1,1-\alpha/2}}{n-1}\right].$$

Réussite du test et *p*-value

Comme nous l'avons vu précédemment, plus le risque α est supposé faible, plus la zone d'admissibilité de H_0 est grande. Il peut être intéressant d'introduire aussi **une notion de niveau de réussite du test**.

Définition 40

La *p*-valeur *d'un test correspond au seuil critique α_c pour laquelle le conclusion du test change:*

$$\begin{cases} \text{si } \alpha \leqslant \alpha_c, & \text{on ne rejette pas } H_0 \\ \text{si } \alpha > \alpha_c, & \text{on rejette } H_0 \end{cases}$$

En général, on cherchera donc à avoir **une *p*-valeur petite**, de sorte à pouvoir valider l'hypothèse (H_1) avec un faible risque de se tromper.

Exercice 46

Une chaîne de fabrication est dédiée à la production de yaourts. Une doseuse remplit les pots et la masse ainsi versée suit une loi normale $\mathcal{N}(\mu, \sigma^2)$ où la précision est a priori donnée par $\sigma = 3g$.

1) *Chaque jour, le fabricant contrôle σ en prélevant un échantillon de 30 pots. Construire un test correspondant. Vérifier que la réalisation $s'_n = 3.13g$ permet bien d'accepter $\sigma = 3g$.*

2) *On suppose ici que $\sigma = 3g$. Chaque heure, un contrôle portant sur 16 yaourts est effectué par le fabricant pour contrôler la valeur de μ. Les valeurs trouvées sont les suivantes :*

$$124.5, 121.6, 125.5, 125.3, 125.6, 124.6, 116.8, 124.3,$$
$$127.9, 124.8, 128.2, 118.9, 130.0, 121.2, 123.8, 130.6$$

Que pensez-vous de l'hypothèse $H_0 = "\mu = 125"$ contre l'hypothèse $H_1 = "\mu < 125"$ pour $\alpha = 0.05$ au vu de cet échantillon ?

3) *Même question si σ est inconnu ?*

Exercice 47

On désire tester la durée de vie de composants censée suivre une loi exponentielle d'espérance $1/\lambda$ au moins égale à 800h. Pour cela sont testés en simultané $n = 100$ composants. L'hypothèse alternative $H_1 = "1/\lambda < 800h"$ et le risque de première espèce est fixé à $\alpha = 0.1$.

1) *Construire le test vérifiant que l'espérance n'est pas inférieure à 800h.*
2) *Calculer l'espérance de la durée du test de la question précédente. Qu'en pensez-vous ?*

Pour accélérer le test, les composants sont regroupés par dispositifs de 10 composants mis en série. Pour chaque série de 10 composants, une horloge mémorise l'instant I de la première panne d'un composant de la série.

3) *Quelle est la loi de I ? Calculer l'espérance de la durée du test accéléré.*
4) *En utilisant les résultats sur les lois de chi-deux et les lois d'Erlang, construire le nouveau test.*

Exercice 48

Pour analyser l'accueil téléphonique d'un service après-vente, une centaine d'appels ont été analysés. La durée de la phase d'orientation de l'appel semble suivre une loi exponentielle. Une précédente étude avait estimé cette durée à $d = 150$ *secondes. Certains pensent que cette durée a augmenté, ce que l'étude doit infirmer ou confirmer. La moyenne des durées observées étant égale à* $\overline{x}_n = 175.1$*, l'hypothèse d'une durée de* 150 *secondes est-elle acceptable ?*

1. Introduction aux probabilités
   - Espace de probabilités et dénombrement
   - Probabilité conditionnelle et événements indépendants
   - Variables aléatoires
   - Vecteurs Aléatoires
   - Espérance et variance conditionnelle
   - Théorème central limite et approximation
2. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison
3. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Objectif

L'objectif d'un test d'ajustement est de savoir si une **variable aléatoire** X suit ou non **une loi donnée**. Il s'agit d'un test où les hypothèses sont de la forme :

$$H_0 = "X \text{ suit la loi } \cdots" \quad \text{contre} \quad H_1 = "X \text{ ne suit pas la loi }"$$

L'idée est donc de prélever **un échantillon**, et d'essayer de **mesurer une "distance"** entre les valeurs observées, et les valeurs attendues si H_0 est vérifiée.

Exemple avec dé:

On effectue 72 lancers d'un dé à 6 faces. Voici les résultats obtenus.

i	1	2	3	4	5	6
n_i	4	10	16	10	12	20

On souhaite tester l'hypothèse

$$H_0 = \text{"le dé n'est pas pipé"}$$

contre l'hypothèse

$$H_1 = \text{"le dé est pipé"}$$

Question: sous l'hypothèse H_0, quelle est la loi vérifiée par ces lancers de dé ?

Loi multinomiale

Définition 41 (Loi multinomiale)

La loi multinomiale *généralise la loi binomiale en considérant m variables aléatoires N_i couplées de loi binomiale $\mathcal{B}(p_i, n)$ pour $i \in \{1, 2, \cdots m\}$:*

$$P(N_i = n_i) = \frac{n!}{n_i!(n-n_i!)} p_i^{n_i}(1-p_i)^{n-n_i},$$

suivant la loi

$$P(N_1 = n_1, N_2 = n_2, \cdots N_m = n_m) = \frac{n!}{n_1!n_2!\cdots n_m!} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$$

Propriété 29

Si $(N_1, N_2, \cdots, N_m)$ suit une loi multinomiale de paramètres $(p_1, p_2, \cdot p_m, n)$, alors

$$\lim_{n\to\infty} \sum_{i=1}^{m} \frac{(N_i - np_i)^2}{np_i} = Z_{m-1} \sim \chi^2_{m-1}.$$

Retour à l'exemple avec un dé:

On suppose que l'hypothèse H_0 est vérifiée.

- Le vecteur aléatoire $(N_1, N_2, N_3, N_4, N_5, N_6)$ suit une **loi multinomiale** de paramètres $(1/6, 1/6, 1/6, 1/6, 1/6, 1/6, 72)$.
- **Statistique de test:**

$$\Delta^2 = \sum_{i=1}^{6} \frac{(N_i - \frac{72}{6})^2}{\frac{72}{6}} \simeq Z_5 \sim \chi^2_5$$

- **Zone d'acceptation**

$$I = [0, z_{5,1-\alpha}].$$

- **Application** avec $\alpha = 0.05$:

$$\delta^2 = \sum_{i=1}^{6} \frac{(n_i - \frac{72}{6})^2}{\frac{72}{6}} = 12.67 > 11.0705.$$

$\Rightarrow$ on ne peut pas accepter H_0.

Test du chi-deux, cas variables discrètes

Soient $X_1, X_2, \cdots X_n$, n répétitions indépendantes d'une variables aléatoire X, à valeurs dans $e_1, e_2, \cdots e_k$.

Nous souhaitons tester **l'hypothèse**

$H_0 = "X$ suit la loi discrète $P(X = e_j) = p_j"$ contre

$H_1 = "X$ ne suit pas la loi discrète $P(X = e_j) = p_j"$

Soient N_j la variable aléatoire correspondant au nombre de variables X_i ayant pour réalisation e_j.

- **Sous l'hypothèse** H_0:
  le vecteur $(N_1, \cdots, N_k)$ suit une loi binomiale de paramètre $(p_1, \cdots, p_k, n)"$
- **Statistique de test**

$$\Delta^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \simeq Z_{k-1} \sim \chi^2_{k-1}$$

- **Zone d'acceptation de** H_0

$$I = [0, z_{k-1,1-\alpha}]$$

Test du chi-deux, cas variables continues

Soient $X_1, X_2, \cdots X_n$, n répétitions indépendantes d'une variables aléatoire X continues dans un ensemble E.
Nous souhaitons tester **l'hypothèse**
$H_0 =$ " X suit la loi continue de densité f" contre
$H_1 =$ " X ne suit pas la loi continue de densité f ".

L'idée consiste alors à

- **Introduire une partition** de l'ensemble E en k **intervalles**

$$E = \cup_{j=1}^{k} e_j \text{ avec } e_j =]a_{j-1}, a_j]$$

- **Se ramener au cas discret** avec
  1) la loi
  $$p_j = P(X \in e_j) = \int_{a_{j-1}}^{a_j} f(x)dx.$$
  2) La variable aléatoire N_j correspondant au nombre de variables X_i dans l'intervalle e_j.

Exercice 49

Nous désirons tester l'équiprobabilité en France des naissances de filles et de garçons dans les familles de 6 enfants. Pour cela, les données correspondant à un échantillon de $n = 1000$ familles ont été collectées et l'observation du nombre de garçons a donné :

0	1	2	3	4	5	6
17	90	231	278	255	104	24

On admet que les naissances des filles et des garçons sont équiprobables. Peut-on conclure de cette étude que les sexes des différents enfants d'une même famille sont indépendants ?

Exercice 50

Une compagnie ferroviaire cherche à analyser la rentabilité de ses lignes locales. Elle étudie donc X_i, le nombre moyen de passagers par jour de $i = 1, ..., n$ lignes de train.

nb moyen de voya.	100 – 150	150 – 200	200 – 250	250 – 300	300 – 350	350 – 400	400 – 450	450 – 500
nb de lignes	462	192	141	72	56	39	22	16

On souhaite tester une modélisation avec une loi de Pareto $\mathcal{P}(a, r)$ avec $a > r = 100$ de densité

$$f(x) = \begin{cases} \dfrac{ar^a}{x^{1+a}} \text{ si } x > r \\ 0 \text{ sinon} \end{cases}$$

1) *En supposant que $r = 100$, appliquer la méthode des moments pour trouver l'estimateur $\overline{A}_n^1 = \dfrac{\overline{X}_n}{\overline{X}_n - r}$ du paramètre a.*
2) *Réaliser le test d'adéquation $H_0 =$ "X_i suit une loi de Pareto $\mathcal{P}(2, 100)$" contre $H_1 =$ "X_i ne suit pas une loi de Pareto $\mathcal{P}(2, 100)$"*

Motivation

Nous souhaitons maintenant comparer un paramètre non pas à une valeur donnée mais à **une valeur inconnue** que nous avons estimée. Par exemple, nous souhaitons comparer l'effet de deux traitements médicaux sur des patients

Nous nous focalisons par la suite uniquement dans le cas **d'échantillons Gaussiens indépendants**:

- Soient $X_1, \cdots X_{n_X}$ i.i.d. selon une loi $\mathcal{N}(\mu_X, \sigma_X^2)$
- Soient $Y_1, \cdots Y_{n_Y}$ i.i.d selon une loi $\mathcal{N}(\mu_Y, \sigma_Y^2)$.

Nous supposons que les (X_i), (Y_j) sont également mutuellement indépendants. L'idée est alors de faire un test d'égalité :

$$H_0 = "\sigma_X^2 = \sigma_Y^2 \text{ et } \mu_X = \mu_Y"$$

Loi de Fisher-Snedecor

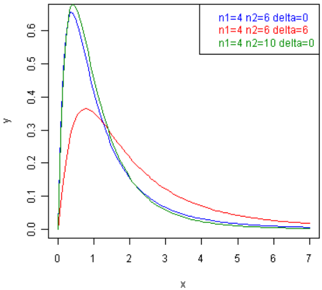
Définition 42

Si $X \sim \chi^2_n$ et $Y \sim \chi^2_m$, et si X et Y sont indépendantes, alors $F_{n,m} = \dfrac{X/n}{Y/m} \sim \mathcal{F}(n,m)$ suit une loi de Fisher de paramètres n et m.

- **Quantile de la loi $\mathcal{F}(n,m)$:** $f_{n,m,\alpha}$ est défini par

$$P(F_{n,m} \leqslant f_{n,m,\alpha}) = \alpha$$

- **Hypothèse de symétrie:** $f_{n,m,\alpha} = f^{-1}_{m,n,1-\alpha}$



Test d'égalité des variances

- **Hypothèse** : $H_0 = "\sigma_X^2 = \sigma_Y^2"$ contre $H_1 = "\sigma_X^2 \neq \sigma_Y^2"$
- **Théorème de Fisher**

$$\frac{(n_X - 1)S_{n_X}'^2}{\sigma_X^2} \sim \chi^2_{n_X - 1} \quad \text{et} \quad \frac{(n_Y - 1)S_{n_Y}'^2}{\sigma_Y^2} \sim \chi^2_{n_Y - 1}$$

- **Statistique de test** sous l'hypothèse $H_0 = "\sigma_X^2 = \sigma_Y^2"$

$$F = S_X'^2 / S_Y'^2 \sim \mathcal{F}(n_X - 1, n_Y - 1).$$

- **Zone d'acceptation** de H_0 pour $S_X'^2 / S_Y'^2$

$$I = [f_{n_X-1, n_Y-1, \alpha/2}, f_{n_X-1, n_Y-1, 1-\alpha/2}]$$

Comparaison des espérances avec σ_X^2 et σ_Y^2 connues

- **Hypothèses:** $H_0 = "\mu_X = \mu_Y"$ contre $H_1 = "\mu_X \neq \mu_Y"$
- **Estimateurs** : $\overline{X}_{n_X} \sim \mathcal{N}(\mu_X, \sigma_X^2/n_X)$ et $\overline{Y}_{n_Y} \sim \mathcal{N}(\mu_Y, \sigma_Y^2/n_Y)$ avec

$$\overline{X}_{n_X} - \overline{X}_{n_Y} \sim \mathcal{N}(\mu_X - \mu_Y, \sigma_X^2/n_X + \sigma_Y^2/n_Y)$$

- **Statistique de test** sous l'hypothèse $H_0 = "\sigma_X^2 = \sigma_Y^2"$

$$U = \frac{\overline{X}_{n_X} - \overline{X}_{n_Y}}{\sqrt{\sigma_X^2/n_X + \sigma_Y^2/n_Y}} \sim \mathcal{N}(0,1).$$

- **Zone d'acceptation** de H_0 pour U

$$I = [u_{\alpha/2}, u_{1-\alpha/2}]$$

Test de Student : cas $\sigma_X^2 = \sigma_Y^2$ inconnues

L'objectif est d'utiliser les estimateurs $S'^2_{n_X,X}$ et $S'^2_{n_Y,Y}$ de σ_X^2 et σ_Y^2:

- **Théorème de Fisher**

$$Z = \frac{(n_X - 1)S'^2_{n_X,X}}{\sigma_X^2} + \frac{(n_Y - 1)S'^2_{n_Y,Y}}{\sigma_Y^2} \sim \chi^2_{n_X+n_Y-2}$$

  et

$$T = \sqrt{n_X + n_Y - 2}\frac{U}{\sqrt{Z}} \sim St(n_X + n_Y - 2)$$

- **Statistique de test:** sous l'hypothèse $H_0 = "\sigma_X^2 = \sigma_Y^2"$

$$T = \frac{\overline{X}_{n_X} - \overline{X}_{n_Y}}{S'_{X,Y}\sqrt{n_X^{-1} + n_Y^{-1}}} \text{ avec } S'^2_{X,Y} = \frac{(n_X - 1)S'^2_{n_X,X} + (n_Y - 1)S'^2_{n_Y,Y}}{n_X + n_Y - 2}$$

- **Zone d'acceptation** de H_0 pour T

$$I = [t_{n_X+n_Y-2,\alpha/2}, t_{n_X+n_Y-2,1-\alpha/2}]$$

Exercice 51

Une expérience a été conduite dont le but est d'évaluer l'effet de l'inoculation d'un mycorrhize sur la croissance en hauteur de plantules de Pinus kesiya. Dans l'expérience, 10 *plantules, formant le Groupe 1, ont été inoculées, et* 10 *autres (Groupe 2) ont été laissées telles quelles. Le tableau ci-dessous donne les hauteurs obtenues dans les deux groupes de plantules.*

Parcelles	*Groupe 1*	*Groupe 2*
1	23.0	8.5
2	17.4	9.6
3	17.0	7.7
4	20.5	10.1
5	20.5	10.1
6	24.0	13.2
7	22.5	10.3
8	22.7	9.1
9	19.4	10.5
10	18.8	7.4

Vérifier si l'inoculation améliore significativement la croissance en hauteur de plantules.

Loi de Bernoulli (L.D)

Définition 43

On considère une expérience qui se traduit par un ***succès*** *avec une probabilité p ou par un* ***échec*** *avec la probabilité $1 - p$. On note X la v.a. définie par :*

$$X = \begin{cases} 1 & \text{si c'est un succès} \\ 0 & \text{si c'est un échec} \end{cases}$$

On dit que X suit une ***loi de Bernoulli*** *de paramètre p et on note :*

$$X \sim \mathcal{B}(p).$$

Sa ***loi de probabilité*** *est définie par :*

$$P(X = 0) = 1 - p \text{ et } P(X = 1) = p.$$

Propriété 30

L'espérance *et la* ***variance*** *de $X \sim \mathcal{B}(p)$ sont données par :*

$$\mathbb{E}(X) = p \text{ et } \mathbb{V}(X) = p(1 - p).$$

Définition 44

On répète de façon indépendante n fois une expérience de Bernoulli dont la probabilité de succès est p. La loi de la v.a. X qui compte le ***nombre de succès*** *sur ces n expériences indépendantes est appelée* ***loi Binomiale*** *de paramètres n et p, ce que l'on note :*

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

Elle prend ses valeurs dans $\{0, 1, \ldots, n\}$ *et sa* ***loi de probabilité*** *est définie par :*

$$\forall k \in \{0, 1, \ldots, n\}, \quad P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Propriété 31

L'espérance *et la* ***variance*** *de* $X \sim \mathcal{B}in(n, p)$ *sont données par :*

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

Définition 45

*La **loi de Poisson** de paramètre λ apparaît comme la limite lorsque $n \to +\infty$, de la loi binomiale $\mathcal{B}in(n, \lambda/n)$, où λ est un réel strictement positif. Une v.a. X suivant la loi de Poisson prend ses valeurs dans $\mathbb{N}$. On dit que X suit une **loi de Poisson** de paramètre λ et l'on note :*

$$X \sim \mathcal{P}(\lambda).$$

*Sa **loi de probabilité** est définie par :*

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

.

Propriété 32

***L'espérance** et la **variance** de $X \sim \mathcal{P}(\lambda)$ sont données par :*

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

Loi géométrique (L.D)

Définition 46

*Une v.a. X qui suit une **loi géométrique** de paramètre p prend ses valeurs dans $\mathbb{N}^* = \{1, 2, \ldots, n, \ldots\}$ et on note :*

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

*Sa **loi de probabilité** est donnée par :*

$$\forall k \in \mathbb{N}, \quad P(X = k) = p(1 - p)^{k-1}.$$

C'est la loi du temps d'attente du premier succès dans une succession d'épreuves de Bernoulli indépendantes.

Propriété 33

***L'espérance** et la **variance** de $X \sim \mathcal{G}(p)$ valent respectivement*

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

1. Introduction aux probabilités
   - Espace de probabilités et dénombrement
   - Probabilité conditionnelle et événements indépendants
   - Variables aléatoires
   - Vecteurs Aléatoires
   - Espérance et variance conditionnelle
   - Théorème central limite et approximation
2. Introduction aux statistiques
   - Estimateur
   - Intervalle de confiance
   - Test d'hypothèse paramétrique
   - Autres tests
     - Test d'ajustement
     - Test de comparaison
3. Annexe : Lois de probabilité usuelles
   - Lois Discrètes (L.D)
   - Lois Continues (L.C)

Loi uniforme (L.C)

Définition 47

*Une v.a. X qui suit la **loi uniforme** sur l'intervalle $I = [a, b]$, notée*

$$X \sim U\,[a, b],$$

*prend ses valeurs dans tout l'intervalle I et sa loi de probabilité est définie par la **densité de probabilité** :*

$$f(x) = \begin{cases} \dfrac{1}{b-a} & si\ a \leqslant x \leqslant b \\ 0 & sinon \end{cases}$$

C'est la loi exacte des phénomènes continus uniformément répartis sur un intervalle.

Propriété 34

*Son **espérance** et sa **variance** sont données par :*

$$\mathbb{E}(X) = \frac{a+b}{2} \text{ et } \mathbb{V}(X) = \frac{(a-b)^2}{12}.$$

Loi exponentielle (L.C)

Définition 48

*Soit $\lambda \in \mathbb{R}_+$, une v.a. X qui suit la **loi exponentielle** de paramètre λ, notée*

$$X \sim \mathcal{E}(\lambda),$$

*est une v.a. continue à valeurs dans $\mathbb{R}_+$. Sa **densité de probabilité** est donnée par :*

$$f(x) = \begin{cases} 0 & si\ x < 0 \\ \lambda e^{-\lambda x} & si\ x \geqslant 0 \end{cases}$$

C'est la loi du temps d'attente d'un phénomène poissonien de taux λ : temps d'attente du premier évènement ou intervalle de temps entre deux évènements consécutifs, dans certaines situations on parle de durée de vie.

Propriété 35

*Son **espérance** et sa **variance** sont données par :*

$$\mathbb{E}(X) = \frac{1}{\lambda} \text{ et } \mathbb{V}(X) = \frac{1}{\lambda^2}.$$

Loi de Erlang

Définition 49

Soit $k \in \mathbb{N}^*$ et $\lambda \in \mathbb{R}_+$. Une v.a. X qui suit la ***loi de Rayleigh*** *de paramètres λ, notée*

$$X \sim Erl(k, \lambda),$$

est une v.a. continue à valeurs dans $\mathbb{R}_+$. Sa ***densité de probabilité*** *est donnée par :*

$$f(x) = \frac{\lambda^k x^{k-1} \exp(-\lambda x)}{(k-1)!} \text{ pour } x \geqslant 0.$$

Propriété 36

La distribution d'Erlang est la distribution de la somme de k variables aléatoires indépendantes et identiquement distribuées selon une loi exponentielle de paramètre λ Son ***espérance*** *et sa* ***variance*** *sont données par :*

$$\mathbb{E}(X) = k/\lambda \text{ et } \mathbb{V}(X) = k\lambda^2$$

Loi normale ou loi Gaussienne (L.C)

Définition 50

Soit $\mu \in \mathbb{R}, \sigma \in \mathbb{R}_+$. Une v.a. X qui suit la ***loi normale*** *de paramètres μ et σ, notée*

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

est une v.a. continue à valeurs dans $\mathbb{R}$. Sa ***densité de probabilité*** *est donnée par :*

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

Propriété 37

Son ***espérance*** *et sa* ***variance*** *sont données par :*

$$\mathbb{E}(X) = \mu \text{ et } \mathbb{V}(X) = \sigma^2.$$

De très nombreuses distributions naturelles sont approximées par la loi normale.

Loi de Rayleigh

Définition 51

Soit $\sigma \in \mathbb{R}_+$. Une v.a. X qui suit la ***loi de Rayleigh*** *de paramètres σ, notée*

$$X \sim \mathcal{R}(\sigma),$$

est une v.a. continue à valeurs dans $\mathbb{R}_+$. Sa ***densité de probabilité*** *est donnée par :*

$$f(x) = \frac{x}{\sigma^2} \exp\left(\frac{-x^2}{2\sigma^2}\right) \text{ pour } x \geqslant 0.$$

Propriété 38

Son ***espérance*** *et sa* ***variance*** *sont données par :*

$$\mathbb{E}(X) = \sigma\sqrt{\pi/2} \text{ et } \mathbb{V}(X) = \frac{4-\pi}{2}\sigma^2$$

Loi de Weibull

Définition 52

Soient $\beta > 0$ et $\Sigma > 0$. Une v.a. X qui suit la ***loi de Weibull*** *de paramètres β et σ, notée*

$$X \sim Weib(\beta, \sigma),$$

est une v.a. continue à valeurs dans $\mathbb{R}_+$. Sa ***densité de probabilité*** *est donnée par :*

$$f(x) = \frac{\beta}{\sigma}\left(\frac{x}{\sigma}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\sigma}\right)^{\beta}\right) \text{ pour } x \geqslant 0.$$

Propriété 39

Son ***espérance*** *et sa* ***variance*** *sont données par :*

$$\mathbb{E}(X) = \sigma\Gamma(1 + 1/\beta) \text{ et } \mathbb{V}(X) = \sigma^2(\Gamma(1 + 2/\beta) - \Gamma(1 + 1/\beta)^2)$$