

Résumé du cours de *Probabilité et Statistiques*
du Prof. Charles-Edouard Pfister

jean-eloi.lombard@epfl.ch

26 janvier 2009

Table des matières

1	Axiomes et définitions de bases	3
1.1	Analyse Combinatoire	3
1.2	Axiomes de Kolmogorov	4
1.2.1	Concepts fondamentaux	4
1.2.2	Axiomes	5
1.3	Espace de probabilité discret	5
1.4	Partitions, Question et Algèbre de Boole	6
1.5	Probabilités conditionnelle	6
1.5.1	Probabilité conditionnelle	6
1.5.2	Événement indépendant	6
1.6	Automate fini	7
1.7	Indépendance	7
1.8	Espace de probabilité continu : mesure de probabilité sur $\mathbb{R}$	8
2	Variable aléatoire	10
2.1	Fonctions de répartition	10
2.1.1	Distributions continues	11
2.1.2	Distributions discrètes	12
2.2	Mesure de probabilité sur $\mathbb{R}^k$	13
2.3	Plusieurs variables aléatoires réelles	13
2.4	Nouvelles variables aléatoires à partir de $X_1, \dots, X_k$	15
2.5	Espérance	17
2.6	Variance	18
2.7	Inégalité de Markov et Chebyshev	19
3	Somme de variables aléatoires	20
3.1	Limite des événements rares	20
3.2	Loi des grands nombres	20
3.3	Principe de la méthode de Monte-Carlo	21
3.3.1	But	21
3.3.2	Méthode	21
3.4	Inégalité de Hoeffding	22
3.5	Marche aléatoire sur $\mathbb{Z}$	23
3.6	L'échelle macroscopique $\alpha = 1$	24
3.7	L'échelle $a_n = \sqrt{n}$ et le Théorème de Moivre-Laplace	26
3.8	Théorème de la Limite Centrale (TLC)	27
3.9	Convergence faible	27

4	Statistiques	29
4.1	Exemples de Modèles Statistiques	29
4.1.1	Modèle de l'urne	29
4.1.2	Mesure d'une grandeur scalaire	29
4.2	Estimation Statistique : Modèle de Gauss-Laplace	30
4.3	Intervalle de confiance	32
4.4	Notion de test	34

Chapitre 1

Axiomes et définitions de bases

Lemme 1.1 (Approximation de Stirling)

$$n! \sim \sqrt{2\pi n} n^n e^{-n}$$

Lemme 1.2 (Binôme de Newton)

$$(a+b)^n = \sum_{k \geq 0} \binom{n}{k} a^{n-k} b^k \quad \forall |x| < 1, r \in \mathbb{R}$$

avec $\binom{n}{k} = \frac{[n]_k}{k!} = \frac{n!}{k!(n-k)!}$.

1.1 Analyse Combinatoire

Lemme 1.3 (Permutations) Le nombre de permutation de n objets est $n!$.

Lemme 1.4 (Tirage ordonné avec remise) Tirer r boules avec remise parmi n donne n^r possibilités.

Lemme 1.5 (Tirage ordonné sans remise) 1. le nombre de tirages ordonnés de r boules pris parmi n sans remise

2. faire la liste des r -mots sans répétition de lettre dans un alphabet de n lettres

3. ranger r boules dans m boîte une boule au plus par boîte.

est $[n]_r = \frac{n!}{(n-r)!}$. Physiquement, ce cas correspond au modèle de Maxwell-Boltzmann.

Lemme 1.6 (Combinaisons, tirage non-ordonné sans remise) Le nombre de combinaisons de r objets pris parmi n , ou encore le nombre de tirages non ordonnés et sans remise de r objets pris parmi n est $\binom{n}{r} = \frac{n!}{(n-r)!r!}$. Ce cas correspond au modèle de fermions de Fermi et Dirac.

Lemme 1.7 (r tirages non-ordonnés avec remise) Soit Le nombre de tirages possibles de r boules parmi n non-ordonnées avec remise est $\binom{n+r-1}{r}$.
Ce cas revient au modèle des Bosons de Bose et Einstein.

Lemme 1.8 Le nombre de rangements de n boules dans m boites ordonnées est

$$[m]^n = m(m+1) \dots (m+n-1)$$

1.2 Axiomes de Kolmogorov

1.2.1 Concepts fondamentaux

Définition 1.1 (Ensemble dénombrable) Un ensemble E est dit *dénombrable* si et seulement si il existe une bijection $\mathbb{N} \leftrightarrow E$.

Définition 1.2 (Expérience) Définit les buts et la manière de réaliser l'expérience.

Définition 1.3 (Algèbre de Boole) On désigne par $\mathcal{F}$ l'ensemble de tous les évènements que l'on veut étudier, dit la famille d'évènements. $\mathcal{F}$ est une algèbre de Boole :

1. $\Omega \in \mathcal{F}$ est l'évènement certain, il est donc toujours réalisé
2. $\emptyset \in \mathcal{F}$
3. si $A \in \mathcal{F}$ alors $A^c := (\Omega \setminus A) \in \mathcal{F}$.
4. si $A_1, \dots, A_n$ est une famille finie ou dénombrable d'évènements, alors :

$$\bigcup_{n \geq 1} A_n \in \mathcal{F} \quad \bigcap_{n \geq 1} A_n \in \mathcal{F}$$

Définition 1.4 (Evenement) Exprime une propriété du résultat de l'expérience.

Loi 1.1 (de Morgan)

$$\left(\bigcup_{i=1}^n E_i \right)^c = \bigcap_{i=1}^n E_i^c$$

$$\left(\bigcap_{i=1}^n E_i \right)^c = \bigcup_{i=1}^n E_i^c$$

Proposition 1.1 1. A et A^c sont incompatibles

2. $P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$ donc $P(A^c) = 1 - P(A)$.
3. $A \subset B$, donc $B = A \cup (B \setminus A)$ ce qui implique $P(B) = P(A) + P(B \setminus A) \geq P(A)$.
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Principe 1.1 (d'inclusion-exclusion de Moivre)

$$P(A) = \sum_{k=1}^n (-1)^{k+1} \sum_{i_1 \leq \dots \leq i_k} P(A_{i_1}, \dots, A_{i_k})$$

pour $n = 2$:

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

1.2.2 Axiomes

Un espace de probabilité est un triplet $(\Omega, \mathcal{F}, P)$ tel que :

Axiome 1 L'ensemble *fondamental* est l'ensemble des issues possibles, noté Ω . Les éléments de Ω , noté souvent ω codent de manière univoque les résultats possibles de l'expérience. A chaque résultat correspond un unique ω .

Axiome 2 $\mathcal{F}$ est un ensemble de sous-ensemble de Ω tel que :

1. $\Omega \in \mathcal{F}$ est l'événement certain, il est donc toujours réalisé
2. $\emptyset \in \mathcal{F}$
3. $A \in \mathcal{F} \Rightarrow (A^c = \Omega \setminus A) \in \mathcal{F}$
4. si $A_1, \dots, A_n$ est une famille finie ou dénombrable d'événements, alors :

$$\bigcup_{n \geq 1} A_n \in \mathcal{F} \quad \bigcap_{n \geq 1} A_n \in \mathcal{F}$$

Axiome 3 (définition d'une probabilité) la "fonction probabilité" $P : \mathcal{F} \rightarrow [0; 1], A \mapsto P(A)$ vérifie :

1. $P(\Omega) = 1$
2. (σ -additivité) si $A_1, \dots, A_n$ famille dénombrable ou finie d'événements vérifiant $A_i \cap A_j = \emptyset \forall i \neq j$, alors

$$P\left(\bigcup A_n\right) = \sum P(A_n)$$

Définition 1.5 (Monotone) La fonction de probabilité P est dite *monotone* si $A \subset B$ implique $P(A) \leq P(B)$

Proposition 1.2 Supposons $\{A_1, \dots, A_n\}$ est monotone décroissante ($A_{n+1} \subset A_n$) et posons $A = \bigcap_i A_i$ alors $P(A) = \lim_{n \rightarrow \infty} P(A_i)$.

Proposition 1.3 Supposons $\{A_1, \dots, A_n\}$ est monotone croissante ($A_n \subset A_{n+1}$) et posons $A = \bigcup_i A_i$ alors $P(A) = \lim_{n \rightarrow \infty} P(A_i)$.

1.3 Espace de probabilité discret

Définition 1.6 (Probabilité discrete) $(\Omega, \mathcal{F}, P)$ est discret si :

1. Ω est dénombrable
2. $\mathcal{F}$ est la collection de tous les sous-ensembles de Ω .

Notation 1 on écrit $q(\omega)$ pour désigner $P(\{\omega\})$.

Théorème 1.1 Soit $a_1, a_2, \dots \geq 0$. Soit $\phi : \mathbb{N} \rightarrow \mathbb{N}$ une bijection avec $b_i = a_{\phi(i)}$. Alors :

1. $(\sum a_i > \infty) \Leftrightarrow (\sum b_i > \infty)$
2. $(\sum a_i < \infty) \Leftrightarrow (\sum b_i < \infty \text{ et } \sum a_i = \sum b_i)$

Donc $\Omega = \{\omega_1, \dots, \omega_n\}$ dénombrable implique qu'il est possible de définir $P(A) = \sum_{i=1, \omega_i \in A} p(\omega_i)$ indépendamment de l'ordre des ω_i et donc on peut écrire $P(A) = \sum_{\omega \in A} p(\omega)$.

Proposition 1.4 Soit Ω dénombrable, $p : \Omega \rightarrow [0; 1]$ tel que $Z := \sum_{\omega \in \Omega} p(\omega) < \infty$ et $A \subset \Omega$. Alors

$$P(A) = \frac{\sum_{\omega \in A} p(\omega)}{Z}$$

est une mesure de probabilité.

Remarque Lorsque Ω est fini ou dénombrable une mesure de probabilité est entièrement décrite par la fonction $p(\omega) = P(\{\omega\})$.

1.4 Partitions, Question et Algèbre de Boole

Définition 1.7 (Question simple) Une *question simple* est définie par une partition de Ω en deux sous-ensembles A et A^c .

Définition 1.8 (Question multiple) Par extension, une *question multiple* est définie par une partition de Ω en $A_1, \dots, A_n$.

Proposition 1.5 Si $A \subset \Omega$ est une algèbre de Boole finie alors il existe une unique partition $A_1, \dots, A_n$ tel que $A = \mathcal{F}_{A_1, \dots, A_n}$. $A_1, \dots, A_n$ sont dits les *atomes* de l'algèbre.

1.5 Probabilités conditionnelle

1.5.1 Probabilité conditionnelle

Définition 1.9 (Probabilité conditionnelle) La probabilité conditionnelle de A sachant B (avec $P(B) > 0$) est définie par :

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

Proposition 1.6 (Règle de multiplication)

$$P(B_1 \dots B_n) = P(B_1)P(B_2|B_1) \dots P(B_n|(B_1 \dots B_{n-1}))$$

Proposition 1.7 (Formule des probabilités totales) Soit $A_1, \dots, A_n$ une partition de Ω et $B \in \mathcal{F}$, alors :

$$P(B) = \sum_{i=1}^n P(BA_i) = \sum_{i=1}^n P(B|A_i)P(A_i) \quad \forall B \in \mathcal{F}$$

Théorème 1.2 (de Bayes) Soit $A_1, \dots, A_n$ une partition de Ω . La probabilité qu'un des A_j soit réalisé sachant que B est réalisé est :

$$P(A_j|B) = \frac{P(B \cap A_j)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_i P(B|A_i)P(A_i)}$$

1.5.2 Événement indépendant

Définition 1.10 (Totalemtent indépendant) Soit $A_1, \dots, A_n$ n événements. Ils sont dit totalement indépendants si pour tout sous-ensemble $A'_1, \dots, A'_r$

$$P(A'_1 \dots A'_r) = P(A'_1) \dots P(A'_r)$$

1.6 Automate fini

Définition 1.11 (Automate fini) Système qui peut se trouver dans un nombre fini d'états vérifiant :

1. $\mathbb{A} = \{1, \dots, r\}$ r états possibles
2. l'automate change d'état à chaque pas de temps
3. on observe l'automate pendant un temps t_n
4. $\Omega_n = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \mathbb{A} \forall i = 1, \dots, n\} = \mathbb{A}^n$ où ω_i est l'état de l'automate au temps i .

Il existe des applications $X_k : \Omega \rightarrow A$, $\omega \mapsto X(\omega) = \omega_k$ pour $k = 1, \dots, r$. La mesure de probabilité de P est donnée par

$$P(\{\omega\}) = q(\omega) = P(X_1 = \omega_1, \dots, X_n = \omega_n)$$

Définition 1.12 (Séquence) Une séquence est une suite de même résultats.

Définition 1.13 (Matrice stochastique) Une matrice $M = M_{kl}$ de dimensions $m \times n$ est une *matrice stochastique* si et seulement si

1. $M_{kl} > 0$
2. $\sum_{l=1}^n M_{kl} = 1$ (la somme des éléments d'une ligne vaut 1).

Proposition 1.8 (de Markov) L'automate à la propriété de Markov si :

$$P(X_j = \omega_j | X_1 = \omega_1, \dots, X_{j-1} = \omega_{j-1}) = P(X_j = \omega_j | X_{j-1} = \omega_{j-1})$$

c'est-à-dire que seul l'état précédent influence le nouvel état.

en plus on va supposer :

$$P(X_j = \omega_j | X_{j-1} = \omega_{j-1}) = M_{\omega_{j-1}, \omega_j}$$

donc il est possible d'écrire :

$$P(X_1 = \omega_1, \dots, X_n = \omega_n) = P(X_1 = \omega_1) M_{\omega_1, \omega_2} \dots M_{\omega_{n-1}, \omega_n}$$

car $P(X_3 = \omega_3 | X_1 = \omega_1, X_2 = \omega_2) = P(X_3 = \omega_3 | X_2 = \omega_2)$. L'automate est donc décrit par M et $P(X_1 = \omega_1)$

1.7 Indépendance

Définition 1.14 (Événements indépendants) Deux événements sont indépendants pour la mesure de probabilité P si et seulement si :

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \\ P(B|A) &= P(B) \end{aligned}$$

Remarque L'indépendance de A et B est toujours définie par rapport à une mesure de probabilité P .

Remarque Si $P(A) = 0$ (ou $P(A) = 1$) alors pour tout $B \in \mathcal{F}$ alors $P(A \cap B) = P(A)P(B)$. A et B sont indépendants sous P .

Remarque Si $A \cap B = \emptyset$ (A et B incompatibles) avec $P(A), P(B) \neq 0, 1$ alors $0 = P(A \cap B) \neq P(A)P(B)$ (A et B ne sont pas indépendants).

Remarque Si $P(A \cap B) = P(A)P(B)$ alors $P(A \cap B^c) = P(A)P(B^c)$.

Définition 1.15 (Algèbres de Boole indépendantes) n σ -algèbre de Boole $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont indépendantes pour la mesure de probabilité P si et seulement si pour tout $C_1 \in \mathcal{F}_1, \dots, C_n \in \mathcal{F}_n$ on a $P(C_1 \dots C_n) = P(C_1) \dots P(C_n)$.

Proposition 1.9 Supposons :

1. l'algèbre $\mathcal{A}$ engendrée par la partition $A_1, \dots, A_l$
2. l'algèbre $\mathcal{B}$ engendrée par la partition $B_1, \dots, B_m$
3. l'algèbre $\mathcal{C}$ engendrée par la partition $C_1, \dots, C_n$

Alors $\mathcal{A}$, $\mathcal{B}$ et $\mathcal{C}$ sont indépendants sous P si et seulement si

$$P(A_i B_j C_k) = P(A_i)P(B_j)P(C_k)$$

pour $\forall i = 1, \dots, l, j = 1, \dots, m, \text{ et } k = 1, \dots, n$.

Définition 1.16 (Indépendance) n événements $B_1, \dots, B_n$ sont indépendants sous P si les n algèbres données par les partitions $B_1, B_1^c, \dots, B_n, B_n^c$ sont indépendants sous P .

Proposition 1.10 Soit $J = \{1, \dots, n\}$. n événements $B_1, \dots, B_n$ sont indépendants sous P si et seulement si

$$P\left(\prod_{j \in J} B_j\right) = \prod_{j \in J} P(B_j)$$

1.8 Espace de probabilité continu : mesure de probabilité sur $\mathbb{R}$

Définition 1.17 (Générateur de nombre aléatoire) Un *générateur de nombres aléatoires* (GNA) est un procédé qui choisit de manière équiprobable un nombre réel de $[0, 1]$.

Soit $\Omega = \mathbb{R}$ et $\mathcal{F} = \mathcal{B}(\mathbb{R})$ (une σ -algèbre de Boole pour $\Omega = \mathbb{R}$ contenant tout les intervalles de $\mathbb{R}$). Supposons P définie sur $\mathcal{F}$ et que pour tout $t \in \mathbb{R},]-\infty, t[\in \mathcal{F}$.

Définition 1.18 (Fonction de répartition F) Soit P une mesure de probabilité sur $\mathbb{R}$, elle est caractérisée par sa *fonction de répartition* $F(t) = P(]-\infty, t])$ vérifiant les propriétés suivantes :

1. $0 \leq F(t) \leq 1$ monotone croissante
2. $\lim_{t \rightarrow -\infty} F(t) = 0, \lim_{t \rightarrow \infty} F(t) = 1$
3. F est continue à droite
4. F à un saut en $t \Leftrightarrow P(\{t\}) > 0$.

Théorème 1.3 Soit F une fonction de repartition, alors F définit univoquement une mesure de probabilité P tel que :

$$P([a; b]) = F(b) - F(a)$$

Définition 1.19 (Densité de probabilité f) Si F est dérivable on peut définir la densité de probabilité de P par

$$f(t) = \frac{dF(t)}{dt}$$

et calculer la probabilité de l'événement $(a, b]$ par

$$P((a, b]) = \int_a^b f(t)dt \Leftrightarrow P(A) = \int_A f(s)ds$$

Proposition 1.11 f a donc les propriétés :

1. $f(t) \geq 0$
2. $\int_{-\infty}^{\infty} f(t)dt = 1$

Définition 1.20 (Ensemble discret) $D \subset \mathbb{R}$ est *discret* si et seulement si pour tout intervalle I borné

$$\text{card}(I \cap D) < \infty$$

D est donc fini ou dénombrable.

Proposition 1.12 Si l'ensemble des discontinuités de F est un ensemble discret D la probabilité de l'événement A se calcul par

$$P(A) = \sum_{x \in A \cap D} q(x)$$

Chapitre 2

Variable aléatoire

Les lettres majuscules $X, Y, \dots$ correspondent aux variables aléatoires (à des applications) alors que les lettres minuscules $x, y, \dots$ correspondent aux valeurs que prennent ces variables aléatoires (réels).

Définition 2.1 (Variable aléatoire) Soit $(\Omega, \mathcal{F}, P)$. Une variable aléatoire réelle X est une application $X : \Omega \rightarrow \mathbb{R}$ telle que pour tout intervalle $A \in B(\mathbb{R})$:

$$X^{-1}(A) \in \mathcal{F}$$

Remarque La condition supplémentaire assure que $P(A)$ est définie pour tout $A \in B(\mathbb{R})$.

Définition 2.2 (Mesure de probabilité discrète) X est une variable aléatoire discrète si elle prend toutes les valeurs d'un sous-ensemble discret de $\mathbb{R}$. Une telle variable aléatoire à une loi de probabilité μ_x qui est une mesure de probabilité discrète sur $\mathbb{R}$

2.1 Fonctions de répartition

Définition 2.3 (Fonction de répartition) La condition de la définition d'une variable aléatoire (Def. 2.1) permet d'introduire la notion de fonction de répartition. Soit X une variable aléatoire. La fonction de répartition F_X de X est définie sur $\mathbb{R}$ par :

$$F_X(t) := P(X \leq t) \quad \forall t \in \mathbb{R}$$

Définition 2.4 (Loi de X) Chaque variable aléatoire définit ainsi une mesure de probabilité sur $\mathbb{R}$ notée μ définie par sa fonction de répartition.

La *loi de X* est la mesure de probabilité μ_x sur $\mathbb{R}$ qui est définie par la fonction de répartition F_X

$$P(X \in A) = \mu_X(A) \quad \forall A \in B(\mathbb{R})$$

Remarque Il est utile de penser à une variable aléatoire comme un couple formé de :

1. une application $X : \Omega \rightarrow \mathbb{R}$
2. μ_x mesure de probabilité sur $\mathbb{R}$

Proposition 2.1 Soit X une variable aléatoire réelle, alors :

1. $0 \leq F_X \leq 1$. F_X est montone non-décroissante et $\lim_{t \rightarrow \infty} F_X(t) = 1$, $\lim_{t \rightarrow -\infty} F_X(t) = 0$.
2. F_X est continue à droite. F_X à un saut en t si et seulement si $P(X = t) > 0$. La "hauteur du saut" est donnée par $F_X(t_+) - F_X(t_i)$.
3. (utile en exo.) la loi de X , μ_X est univoquement déterminé par F_X : $\mu_X([a, b]) = F_X(b) - F_X(a)$ (conséquence directe du théorème fondamental du calcul intégral).

Définition 2.5 (Variable aléatoire continue) Une variable aléatoire X est dite *continue* s'il existe une fonction non-négative f , dite densité de probabilité de X tel que pour un ensemble B arbitraire :

$$P(X \in B) = \int_B f(x) dx$$

La fonction de répartition F est différentiable

$$\frac{dF(x)}{dx} = f(x)$$

et

$$F_X(t) = \int_{-\infty}^t f(x) dx$$

Définition 2.6 (Variable aléatoire discrète) Soit X une variable aléatoire discrète telle que $P(X = x_1) = q(x_1)$. $\{x_n, n \geq 1\}$ est un sous-ensemble discret E de $\mathbb{R}$ vérifiant $\sum_{x_n \in E} q(x_n) = 1$.

$$P(X \in]-\infty, t]) = F_X(t) = \sum_{x \in E \cap]-\infty, t]} q(x)$$

donc $F_X(t)$ est constante par morceaux.

Définition 2.7 (Cas général) Soit $F_X \in C^1$ continue par morceaux avec des sauts en $x \in E$. Pour tout $t \notin E$ on peut calculer

$$f(t) = \frac{dF_X(t)}{dt}$$

et donc :

$$P(X \in A) = \int_A f(t) dt + \sum_{x \in E \cap A} q(x) = \int_A dF_X$$

2.1.1 Distributions continues

Remarque Soit $(\Omega = \mathbb{R}, \mathcal{F}, P)$. Pour définir P on utilise une fonction auxiliaire f la densité de probabilité.

Loi 2.1 (Exponentielle) Une v.a. de densité

$$f(x) = \lambda e^{-\lambda x} \quad x \geq 0$$

est dite *exponentielle de paramètre λ* . Sa moyenne et sa variance sont :

$$\mathbb{E}(X) = \frac{1}{\lambda} \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

Loi 2.2 (Gaussienne (ou Normale)) $\Omega = \mathbb{R}$, $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ avec $\mu \in \mathbb{R}$ l'espérance et σ^2 la variance. On définit $P(A) = \int_A f(x)dx$. La mesure de probabilité correspondant à f de paramètres μ et σ est $N(\mu, \sigma^2)$ dite probabilité gaussienne. Pour $N(0, 1)$ la loi est dite normale centrée réduite.

Proposition 2.2 Pour une variable aléatoire gaussienne X normale centrée réduite :

1. $P(-1 \leq X \leq 1) \leq 0,7$
2. $P(-2 \leq X \leq 2) \leq 0,95$
3. $P(-3 \leq X \leq 3) \leq 0,99$
4. $P(X \notin [-4, 4]) \approx 0$

Loi 2.3 (de Cauchy) Soit $a \in \mathbb{R}^+$. La *loi de Cauchy* de paramètre a est définie par :

$$\gamma_a(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}$$

Cette loi n'a pas d'espérance car il est impossible d'intégrer $\frac{|x|a}{\pi(a^2 + x^2)}$ sur $]-\infty, \infty[$.

2.1.2 Distributions discrètes

Loi 2.4 (de Poisson) $E \subset \mathbb{R}$ est discret si pour tout intervalle $I \subset \mathbb{R}$, $|E \cap I| \leq \infty$. Soit E un ensemble discret dénombrable et soit $q : \mathbb{R} \rightarrow \mathbb{R}^+$ tel que

$$\begin{cases} 0 & \text{si } x \notin E \\ q(x) \geq 0 & \text{si } x \in E \end{cases}$$

et $\sum_{x \in E} q(x) = 1$. À l'aide de q on définit une mesure de probabilité sur $\mathbb{R}$ $P(A) = \sum_{x \in A \cap E} q(x)$. Avec $\Omega = \mathbb{R}$, $E = \{0, 1, \dots\}$ et

$$q(x) = \begin{cases} 0 & \text{si } x \notin E \\ \Pi_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda} & \text{si } x, \lambda \in E \end{cases}$$

on définit une probabilité de Poisson Π_k de paramètre λ .

Proposition 2.3

$$\mathbb{E}(X) = \lambda \quad \text{Var}(X) = \lambda$$

Loi 2.5 (Binomiale) Supposons n répétitions indépendantes d'une même épreuve de probabilité de succès p . La variable aléatoire X qui compte le nombre de succès au bout de n essais est dite "loi binomiale de paramètres (n, p) ". La loi de probabilité d'une variable aléatoire binomiale de paramètre (n, p) est :

$$p(i) = \binom{n}{i} p^i (1-p)^{n-i} \quad i = 0, 1, \dots, n$$

Proposition 2.4

$$\mathbb{E}(X) = np \quad \text{Var}(X) = np(1-p)$$

Loi 2.6 (Multinomiale) Un vecteur aléatoire $(X_1, \dots, X_r)$ suit une loi multinomiale de paramètres $n, p_1, \dots, p_r$ avec $0 < P_i < 1$ et $p_1 + \dots + p_r = 1$ si :

1. $X_1(\omega) = \dots = X_r(\omega) = \{0, \dots, n\}$

- 2.

$$P\left((X_1 = k_1) \dots (X_r = k_r)\right) = \frac{n!}{k_1! \dots k_r!} p_1^{k_1} \dots p_r^{k_r}$$

et si $k_1 + k_r = n$ alors

$$P\left((X_1 = k_1) \dots (X_r = k_r)\right) = 0$$

Loi 2.7 (Multinomiale) Cette loi s'applique dans le cas de tirages sans remises. Considérons une urne avec N boules dont R rouges (les autres sont blanches) et une expérience aléatoire au cours de laquelle on tire n boules ($\Omega = \{0, 1, \dots, n\}$). La probabilité de tirer r boules rouges est donnée par la loi Hypergéométrique :

$$\mathcal{H}(n, R, N)(r) = \frac{\binom{N-R}{n-r} \binom{R}{r}}{\binom{N}{n}}$$

2.2 Mesure de probabilité sur $\mathbb{R}^k$

Définition 2.8 (Mesure discrete de probabilité sur $\mathbb{R}^k$) $D \subset \mathbb{R}^k$ si pour toute boule $B(t, r) = \{x, |t - x| \leq r\}$, $\text{card}(D \cap B(t, r)) < \infty$.

Proposition 2.5 La mesure discrete de probabilité est spécifiée par :

1. D ensemble discret de $\mathbb{R}^k$
2. pour tout $a \in D$, $p(a) \geq 0$ et $\sum_{a \in D} p(a) = 1$
3. pour tout $A \in \mathbb{R}^k$ $\mu_A = \sum_{a \in D \cap A} p(a)$

Définition 2.9 (Mesure de probabilité continue sur $\mathbb{R}^k$) Il existe une fonction $f : \mathbb{R}^k \rightarrow \mathbb{R}$ tel que $x = (x_1, \dots, x_k) \mapsto f(x_1, \dots, x_k) \geq 0$ vérifiant :

$$\int_{\mathbb{R}^k} f(x_1, \dots, x_k) dx_1, \dots, dx_k = 1$$

Pour tout $A \subset \mathbb{R}^k$:

$$\mu_A = \int_A f(x_1, \dots, x_k) dx_1, \dots, dx_k$$

2.3 Plusieurs variables aléatoires réelles

Définition 2.10 (Loi jointe) Soit $X_1, \dots, X_k$ des variables aléatoires définies sur $(\Omega, \mathcal{F}, P)$ avec $\mathbf{X} = (X_1, \dots, X_k) \in \mathbb{R}^k$ La loi jointe de $\mathbf{X}$ ou la loi de $\mathbf{X}$ est la mesure de probabilité :

$$\mu_{\mathbf{X}}(B) = P(\mathbf{X} \in B) \quad \forall B \in \mathbb{R}^k$$

Définition 2.11 (Loi marginale) La loi d'un X_i est une *loi marginale* de μ_X . Prenons par exemple $X_1, X_2, X_3 \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, alors :

$$\begin{aligned} P(X_2 \in A) &= P(X_1 \in \mathbb{R}, X_2 \in A, X_3 \in \mathbb{R}) \\ &= P(\underline{X} \in \mathbb{R} \times A \times \mathbb{R}) \\ &= \mu_{\underline{X}}(\mathbb{R} \times A \times \mathbb{R}) \end{aligned}$$

Cas discret

$$p(X_2) = \sum_{x_1 \in D_1} \sum_{x_3 \in D_3} p_{\underline{X}}(x_1, x_2, x_3)$$

car $\sum_{x_2 \in A} p(x_2) = 1$.

Cas continu

$$\begin{aligned} p(x_2 \in A) &= \int_A \left(\int_{\mathbb{R}} \int_{\mathbb{R}} f(x_1, x_2, x_3) dx_1 \right) dx_3 \, dx_2 \\ &= \int_A g(x_2) dx_2 \end{aligned}$$

avec g la densité de la loi de X_2 .

Exemple Si X, Y sont défini pour $a < x, y < b$ avec comme loi jointe $f_{X,Y}(x, y)$ alors les lois marginales sont données par :

$$f_X(x) = \int_a^b f_{X,Y}(x, y) dy \quad f_Y(y) = \int_a^b f_{X,Y}(x, y) dx$$

Remarque La loi jointe contient plus d'information que toutes les lois marginales. On peut trouver les lois marginales avec la loi jointe, mais pas le contraire. Ce n'est le cas que pour des variables aléatoires indépendantes dans quel cas :

$$\begin{aligned} P(X = k, Y = l) &= P(X = k)P(Y = l) \quad \text{cas discret} \\ f_{X,Y}(x, y) &= f_X(x)f_Y(y) \quad \text{cas continu} \end{aligned}$$

Théorème 2.1 Les variables aléatoires continues (ou discrètes) X et Y sont indépendantes si et seulement si la densité de la loi jointe peut-être exprimée sous la forme :

$$f_{X,Y}(x, y) = h(x)g(y) \quad -\infty < x, y < \infty$$

Exemple Si la loi jointe de X et Y est donnée par :

$$f(x, y) = \begin{cases} 2e^{-2x}e^{-3y} & \text{si } 0 < x, y < \infty \\ 0 & \text{sinon} \end{cases}$$

alors clairement X et Y sont indépendantes. Toutefois dans le cas où la loi jointe de X et Y est :

$$f(x, y) = \begin{cases} 24xy & \text{si } 0 < x, y < \infty \\ 0 & \text{sinon} \end{cases}$$

alors en posant :

$$I(x, y) = \begin{cases} 1 & \text{si } 0 < x, y < 1, 0 < x + y < 1 \\ 0 & \text{sinon} \end{cases}$$

alors la loi jointe prends la forme :

$$f(x, y) = 24xyI(x, y)$$

qui ne peut pas être factorisée en un terme qui ne dépends que de x et un autre qui ne dépends que de y . En d'autres termes le domaine où $f(x, y) \neq 0$ n'est pas du type $A \times B$ avec $A, B \subset \mathbb{R}$.

Définition 2.12 (Modèle minimal) Utilise uniquement l'information a disposition :

1. code le résultat d'une expérience aléatoire avec des valeurs $\underline{X} = (X_1, \dots, X_n)$ des n $X_1, \dots, X_n$.
2. la mesure de probabilité est donnée par la loi jointe $\mu_{\underline{X}}$.

Définition 2.13 (Variables aléatoires indépendantes (Important !)) Les variables $X_1, \dots, X_k$ sont indépendantes si :

$$\mu_{X_1, \dots, X_k}(B_1 \dots B_k) = \mu_{X_1}(B_1) \dots \mu_{X_k}(B_k)$$

Proposition 2.6 Les variables aléatoires $X_1, \dots, X_k$ sont indépendantes si :

$$P(X_1 < t_1, \dots, X_k < t_k) = P(X_1 < t_1) \dots P(X_k < t_k)$$

2.4 Nouvelles variables aléatoires à partir de $X_1, \dots, X_k$

Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ et X une variable aléatoire réelle. Définissons une nouvelle variable aléatoire $Y = \phi(X)$.

Proposition 2.7 (Cas discret) Supposons :

1. X discrète
2. $D \subset \mathbb{R}$ tel que :
 - (a) $q(x) \geq 0$ avec $q(x)$ une notation pour $P(X = x)$
 - (b) $\sum_{x \in D} q(x) = 1$

alors

$$P(Y = y) = \sum_{x \in D \setminus \{\phi(x)=y\}} P(X = x)$$

Proposition 2.8 (Cas continu) Supposons ϕ strictement monotone de classe C^1 et que la loi de probabilité de X est donnée par la densité de probabilité f_X , alors :

$$f_Y(y) = f_X(\phi^{-1}(y)) \left| \frac{d\phi^{-1}(y)}{dy} \right|$$

Proposition 2.9 Soit k fonctions $\phi_i : \mathbb{R} \rightarrow \mathbb{R}$ et k variables aléatoires X_i . Posons $Y_i = \phi_i(X_i)$, alors les k variables aléatoires Y_i sont aussi indépendantes.

Exemple Connaissant la loi jointe de X et Y

$$f_{X,Y}(x,y) = \begin{cases} e^{-(x+y)} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{sinon} \end{cases}$$

il est possible de calculer par exemple la densité de $Z = \frac{X}{Y}$. Calculons la fonction de répartition de Z :

$$\begin{aligned} F_Z(t) &= P\left(\frac{X}{Y} \leq t\right) = \iint_{\frac{x}{y} \leq t} f(x,y) dx dy \\ &= \int_0^\infty dy \int_0^{yt} f(x,y) dx \\ &= \int_0^\infty dy \int_0^{yt} e^{-(x+y)} dx \\ &= \int_0^\infty e^{-y} dy \int_0^{yt} e^{-x} dx \\ &= \int_0^\infty e^{-y} [-e^{-x}]_0^{yt} dy \\ &= \int_0^\infty e^{-y} (1 - e^{-yt}) dy \\ &= \left[-e^{-y} + \frac{1}{1+t} e^{-y(1+t)} \right]_0^\infty \\ &= 1 - \frac{1}{1+t} \end{aligned}$$

donc la densité de X/Y est :

$$f_{X/Y}(t) = \frac{dF_{X/Y}(t)}{dt} = \frac{1}{(1+t)^2}$$

Définition 2.14 (Convolution) La densité d'une somme de v.a. indépendantes est obtenue par *convolution*. La *convolution* de :

$$f_{X+Y}(a) = \int_{-\infty}^\infty f_X(a-y) f_Y(y) dy$$

Exemple Supposons X et Y indépendantes et calculons la densité de $X + Y$.

$$\begin{aligned} F_{X+Y}(t) &= P(X+Y \leq t) = \iint_{\{(x,y) \in \mathbb{R}^2 : x+y \leq t\}} f_{X+Y}(x,y) dx dy \\ &= \\ &\stackrel{\text{ind.}}{=} \int_{-\infty}^\infty dx \int_{-\infty}^{t-x} f_X(x) f_Y(y) dy \\ &= \int_{-\infty}^\infty f_X(x) dx \int_{-\infty}^{t-x} f_Y(y) dy \end{aligned}$$

d'où

$$f_{X+Y}(t) = \frac{dF_{X+Y}(t)}{dt} = \int_{-\infty}^\infty f_X(x) dx f_Y(t-x)$$

Remarque Dans le cas discret (loi de Poisson, loi Binomiale), on calcule :

$$P(X + Y = k) = \sum_{i=0}^n P(X = i, Y = k - i) = \sum_{i=0}^n P(Y = k - i) = \dots$$

2.5 Espérance

L'espérance et la variance (Section 2.6) sont deux paramètres réels qui permettent de caractériser la loi d'une variable aléatoire.

Définition 2.15 (Espérance) Intuitivement l'espérance est la moyenne de X , c'est un paramètre de position. Si $\sum_{x/p(x)>0} xp(x) < \infty$ alors l'espérance de la variable aléatoire X , notée $\mathbb{E}[X]$ est définie par :

$$\mathbb{E}[X] = \sum_x xp(x)$$

ou dans le cas continu, si $\int xp(x)dx < \infty$ alors

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

et dans le cas général, en supposant $X \in C^1$ continue par morceaux vérifiant $\sum_{x/p(x)>0} xp(x) < \infty$ et $\int |g(x)|p(x)dx < \infty$ l'espérance est définie par :

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

Remarque L'espérance n'est pas toujours définie. C'est le cas pour une distribution de Cauchy 2.3

Théorème 2.2 Soit

1. X une variable aléatoire,
2. $\phi : \mathbb{R} \rightarrow \mathbb{R}$
3. soit f_X la fonction de répartition de X .
4. $\int \phi(x)f_X(x)dx < \infty$

alors l'espérance de $Y := \phi(X)$ est donnée par :

$$\mathbb{E}[Y] = \int_{\mathbb{R}} \phi(x)f_X(x)dx$$

Proposition 2.10 Soit X et Y deux variables aléatoires telles que $\mathbb{E}(X)$ et $\mathbb{E}(Y)$ existent, alors :

1. $\mathbb{E}(X)$ existe si et seulement si $\mathbb{E}(|X|)$ existe.
2. linéarité de l'espérance : pour tout $c \in \mathbb{R}$, $\mathbb{E}[cX] = c\mathbb{E}[X]$ et $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
3. si $X > 0$ alors $\mathbb{E}[X] > 0$
4. $X \leq Y$ alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$
5. $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
6. soit X et Y deux variables aléatoires indépendantes, alors $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

Théorème 2.3 Soit X et Y deux variables aléatoires réelles et $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ une application. Définissons une nouvelle variable aléatoire $g(X, Y)$ avec :

dans le cas discret $\sum_{x \in D_X} \sum_{y \in D_Y} |g(x, y)| P(X = x, Y = y) < \infty$ alors

$$\mathbb{E}(g(X, Y)) = \sum_{x \in D_X} \sum_{y \in D_Y} g(x, y) P(X = x, Y = y)$$

et dans le cas continu $\iint_{\mathbb{R}^2} |g(x, y)| f_{X,Y}(x, y) dx dy < \infty$ alors

$$\mathbb{E}(g(X, Y)) = \iint_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y) dx dy$$

Exemple Soit X et Y deux variables aléatoires continues, $Z = g(X, Y) = XY$, alors

$$\mathbb{E}(Z) = \mathbb{E}(XY) = \iint_{\mathbb{R}^2} xy f_{X,Y}(x, y) dx dy$$

et si X et Y sont indépendantes :

$$\mathbb{E}(XY) = \int_{\mathbb{R}} x f_X(x) dx \int_{\mathbb{R}} y f_Y(y) dy = \mathbb{E}(X) \mathbb{E}(Y)$$

Remarque Si les variables aléatoires sont indépendantes alors $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ implique $\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y)$. L'implication inverse est fausse.

2.6 Variance

Définition 2.16 (Variance) Soit X une variable aléatoire. Si $Y = (X - \mathbb{E}(X))^2$ possède une espérance, alors la variance de X , notée $Var(X)$ est définie par ;

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

Définition 2.17 (Covariance) Soit X_1 et X_2 deux variables aléatoires. La covariance de X_1 et X_2 est :

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

Si $Cov(X_1, X_2) = 0$ on dit que X_1 et X_2 sont non-corrélées.

Remarque La variance est un paramètre de dispersion qui n'est pas toujours défini.

Proposition 2.11

$$\begin{aligned} Var(aX + b) &= a^2 Var(X) \quad \forall a, b \in \mathbb{R} \\ Var(X + Y) &= Var(X) + Var(Y) + 2cov(X, Y) \end{aligned}$$

Remarque Pour deux v.a. indépendantes la variance de la somme est la somme des variances car l'indépendance donne $\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y)$ donc $Cov(X, Y) = 0$ alors $Var(X + Y) = Var(X) + Var(Y)$

Définition 2.18 (Écart type) L'écart-type, noté $\sigma(x)$ est défini par :

$$\sigma(x) = \sqrt{Var(X)}$$

2.7 Inégalité de Markov et Chebyshev

Théorème 2.4 (Inégalité de Markov) Soit X une variable aléatoire non-négative, alors pour tout $a > 0$,

$$P(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}$$

Exemple Le nombre de pièces fabriquées dans une usine décrit une variable aléatoire X d'espérance $\mathbb{E}(X) = 50$. Quelle est la probabilité que la production dépasse 75 ?

$$P(X > 75) \leq \frac{\mathbb{E}(X)}{75} = \frac{2}{3}$$

Théorème 2.5 (Inégalité de Chebyshev)

$$P((X - \mathbb{E}[X]) \geq a) \leq \frac{Var(X)}{a^2}$$

Exemple Le nombre de pièces fabriquées dans une usine décrit une variable aléatoire X de variance $Var(X) = 25$ et d'espérance $\mathbb{E}(X) = 50$. Quelle est la probabilité que la production se situe entre 40 et 60 ?

$$P(|X - \mathbb{E}(X)| \geq 10) \leq \frac{25}{10^2} = \frac{1}{4}$$

Chapitre 3

Somme de variables aléatoires

3.1 Limite des événements rares

Théorème 3.1 Soit $A_1, \dots, A_n$ n événements indépendants avec $P(A_i) = p_i$, $0 < p_i < 1$ (plus les p_i sont petit plus le théorème est puissant). Posons X_i l'indicatrice de A_i ($X_i = 1$ si A_i est réalisé, 0 sinon) et $S_n = X_1 + \dots + X_n$ qui compte le nombre d'événements réalisés, supposons $X_1, \dots, X_n$ indépendants.

Soit $Z(n)$ une variable aléatoire de Poisson de paramètres $\lambda = p_1 + \dots + p_n$. Alors pour tout $j \in \{1, \dots, n\}$. On peut estimer le nombre de A_i qui se sont réalisés avec :

$$|P(S_n) - P(Z(n))| \leq \sum_{i=1}^n p_i^2$$

Exemple Utilisons le théorème 3.1 pour quantifier l'erreur commise lors de l'approximation d'une loi binomiale par une loi de Poisson de paramètre $\lambda = np = q$ ($Z \sim \Pi_{np}$).

Prenons $p_i = \frac{q}{n}$ pour tout $i \in J = \{1, \dots, n\}$ alors :

$$\sum_{i=1}^n p_i^2 = np^2 = n \frac{q^2}{n^2} = \frac{q^2}{n}$$

alors :

$$|P(S_n \in J) - P(Z \in J)| \leq \frac{q^2}{n}$$

Définition 3.1 (Indépendantes identiquement distribuées(iid)) Soit $X_1, \dots, X_n$ n variables aléatoires. Elles sont dites *indépendantes identiquement distribuées* si les X_i sont indépendantes et ont la même loi.

3.2 Loi des grands nombres

Dans la suite on s'intéresse à la situation suivante : soit $a_1 < \dots < a_n$ une suite de nombre réels tels que $\lim_{n \rightarrow \infty} a_n = \infty$ et une variable aléatoire définie

comme :

$$\frac{X_1 + X_2 + \dots + X_n}{a_n}$$

avec $a_n = n$ pour l'échelle macroscopique et $a_n = \sqrt{n}$ pour l'échelle microscopique.

Théorème 3.2 (Loi Faible des Grands Nombres) Correspond au cas $a_n = n$. Soit $X_1, X_2, \dots$ une suite de v.a. i.i.d. alors pour tout $\epsilon > 0$:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_i) \right| > \epsilon \right) \rightarrow 0$$

Exemple Si X_i sont i.i.d. v.a. de Bernoulli de paramètre p ($p(x_i = 1) = p$ et $p(x_i = 0) = 1 - p$). Soit A tel que $P(A) = p$ et $X_i = J_A$ où i désigne le i^{e} tour de l'expérience, alors :

$$\begin{aligned} \frac{\sum_{i=1}^n x_i}{n} &= \text{fréquence de } A \\ &= \frac{\text{nombre de réalisations de } A \text{ pendant } n \text{ répétitions}}{n} \end{aligned}$$

donc "la fréquence d'un événement est proche de sa probabilité lorsque n est grand."

Théorème 3.3 (Loi Forte des Grands Nombres) Soit $X_1, \dots, X_n$ i.i.d. d'espérance μ alors :

$$P \left(\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mu \right) = 1$$

3.3 Principe de la méthode de Monte-Carlo

3.3.1 But

Soit $f \in C^0 : [0, 1]^d \rightarrow \mathbb{R}$ et c une constante telle que $\sup_k |f(t_k)| = c$. Le but est de calculer numériquement :

$$I = \int_{[0,1]^d} f(t_1, \dots, t_d) dt_1 \dots dt_d$$

3.3.2 Méthode

1. On divise $[0, 1]^d$ en r^d cubes de volumes $\frac{1}{r^d}$
2. On numérote chaque cube $A = \{1, \dots, r^d\}$
3. Soit S_k le centre du cube k .
4. On définit la somme de Riemann $I = \sum_k f(S_k) \frac{1}{r^d}$
5. l'idée est d'interpréter I comme l'espérance d'une variable aléatoire Y

$$\begin{aligned} Y : \quad A &\rightarrow \mathbb{R} \\ k &\mapsto f(S_k) \end{aligned}$$

$$\text{soit } I = \sum_k Y(k) P(k) = \sum_k f(S_k) \frac{1}{r^d} = \mathbb{E}(Y).$$

6. On choisit au hasard un cube et on évalue Y en k .
7. On répète alors n fois l'expérience de manière indépendante et on obtient $\omega = (\omega_1, \dots, \omega_n)$.
8. On considère chaque expérience comme une variable aléatoire indépendantes des autres et identiquement distribuées en posant $X_j(\omega) = Y(\omega_j)$ d'où :

$$\frac{1}{n}(X_1(\omega) + \dots + X_n(\omega)) = \frac{1}{n}(Y(\omega_1) + \dots + Y(\omega_n))$$

9. la Loi des Grands Nombres donne :

$$\lim_{n \rightarrow \infty} P \left(\left[\frac{1}{n}(X_1 + \dots + X_n) - I \right] \geq \epsilon \right) = 0$$

10. finalement on fait une estimation de l'erreur à l'aide de l'inégalité de Chebyshev. Si $Var(X_i)$ existe alors

$$\begin{aligned} Var \left(\frac{1}{n}(X_1 + \dots + X_n) \right) &= \frac{1}{n^2} n Var(X_i) \\ &= \frac{Var(X_i)}{n} \end{aligned}$$

or $Var(X_i) = Var(Y_i) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \leq c^2$ donc d'après l'inégalité de Chebychev :

$$P \left(\left[\frac{1}{n}(X_1 + \dots + X_n) - I \right] \geq \epsilon \right) \leq \frac{1}{n} \frac{Var(Y)}{\epsilon^2} \leq \frac{c^2}{n\epsilon^2}$$

3.4 Inégalité de Hoeffding

L'inégalité de Hoeffding permet d'estimer la probabilité qu'une somme de v.a. indépendante prenne une valeur qui s'écarte de la somme des espérance. C'est une généralisation de la Loi des Grands Nombres (puisque les v.a. n'ont pas besoin d'être i.i.d.).

Théorème 3.4 (Inégalité de Hoeffding) Soit $X_1, X_2, \dots$ des variables aléatoires. Pour tout $i \in \mathbb{N}$, il existe $a_i < b_i$ tel que $a_i < X_i - \mathbb{E}(X) < b_i$ alors pour tout $c > 0$:

$$P \left(\left[\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}(X_i) \right] \geq c \right) \leq \exp \left(\frac{-2c^2}{\sum_{i=1}^n (b_i - a_i)^2} \right)$$

Remarque Si :

$$P \left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}(X_i) \leq -c \right) \Leftrightarrow P \left(\sum_{i=1}^n (-X_i) - \sum_{i=1}^n \mathbb{E}(-X_i) \geq c \right)$$

alors :

$$P \left(\left| \sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}(X_i) \right| \geq c \right) \leq 2 \exp \left(\frac{-2c^2}{\sum_{i=1}^n (b_i - a_i)^2} \right)$$

Remarque Soit $X_1, X_2, \dots$ des v.a. i.i.d. . Alors il est possible de donner une meilleure approximation de la Loi des Grands Nombres (Théorème 3.3), par :

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n(X_i - \mathbb{E}(X_i))\right| \geq \epsilon\right) \leq 2\exp\left(\frac{-2\epsilon^2 n}{(b-a)^2}\right)$$

Définition 3.2 (Mediane) Soit X une variable aléatoire avec F_X sa fonction de répartition supposée continue. La *médiane* est la valeur m pour laquelle $F_X(m) = \frac{1}{2}$. Le point m vérifie donc :

$$P(x \leq m) = P(x \geq m) = \frac{1}{2}$$

3.5 Marche aléatoire sur $\mathbb{Z}$

Le marcheur part de l'origine et à chaque pas de temps on lance un pièce de monnaie (équilibrée : $P(\text{"pile"}) = P(\text{"face"}) = \frac{1}{2}$). Pour "pile" il fait un pas à droite et pour face un à gauche. Introduisons la variable aléatoire X_k vérifiant $P(X_k = 1) = P(X_k = -1) = \frac{1}{2}$. Supposons $X_1, \dots, X_n$ indépendantes et identiquement distribuées. Donc l'espérance $\mathbb{E}(X) = 0$. Posons $S_n = \sum_{i=1}^n X_i$ la position du marcheur après n pas. Définissons l'ensemble fondamental par :

$$\begin{aligned}\Omega_n &= \{\omega = (\omega_1, \dots, \omega_n), \omega_n \text{ la position du marcheur au temps } n\} \\ &= \{\text{les marches possibles après } n \text{ lancés si } \omega_0 = 0\}\end{aligned}$$

Sur cet espace de probabilité $X_k(\omega) = \omega(k) - \omega(k-1)$ pour $k = 1, \dots, n$. Supposons toute les marches équiprobables

$$P(\{\omega\}) = \frac{1}{2^n} \quad \omega \in \Omega_n$$

et donc l'inégalité de Hoeffding (Théorème 3.4) donne :

$$P(S_n \geq c) \leq \exp\left(\frac{-2c^2}{4n}\right)$$

la probabilité de retrouver le marcheur à droite de c au temps n .

Proposition 3.1 (Différentes échelles) Avec le choix de c , il est possible de considérer différentes échelles en posant $c = tn^\alpha$ avec $t \in [0, 1]$:

1. $\alpha \in \mathbb{R}^+$ (cas général)

$$P(S_n \geq t_n) \leq e^{-\frac{t^2 n^{2\alpha-1}}{2}}$$

2. $\alpha = 1$ (l'échelle macroscopique)

$$P(S_n \geq t_n) \leq e^{-\frac{2t^2 n^2}{4n}} = e^{-\frac{nt^2}{2}}$$

donc macroscopiquement le marcheur ne bouge pas.

3. $\alpha = 1/2$ i.e. $c = \sqrt{nt}$

$$P(S_n \geq t\sqrt{n}) \leq e^{-\frac{t^2}{2}}$$

3.6 L'échelle macroscopique $\alpha = 1$

Définition 3.3 (Retour à l'origine) Le retour à l'origine au temps $2n$ est l'événement $\{S_{2n} = 0\}$. Posons

$$U_{2n} := P(S_{2n} = 0) = \binom{2n}{n} \frac{1}{2^{2n}} \quad (3.1)$$

or la formule de Stirling approxime $n!$ par :

$$n! \approx \sqrt{2\pi n} n^n e^{-n}$$

donc l'Eq. 3.1 devient :

$$U_{2n} \approx \frac{\sqrt{2\pi 2n} (2n)^{2n} e^{2n}}{(\sqrt{2\pi n} n^n e^{-n})^2} = \frac{\sqrt{2\pi 2n}}{2\pi n} = \frac{1}{\sqrt{\pi n}} \xrightarrow{n \rightarrow \infty} 0$$

donc pour n grands, le retour à l'origine est peu probable.

Définition 3.4 (Premier retour à l'origine) Le premier retour à l'origine au temps $2n$ est l'événement $\{S_2 \neq 0, S_4 \neq 0, \dots, S_{2n} = 0\}$. On peut montrer que $P(S_2 \neq 0, \dots, S_{2n} \neq 0) = U_{2n}$. Ainsi on trouve :

$$\begin{aligned} f_{2n} &= P(S_2 \neq 0, \dots, S_{2n} = 0) \\ &= P(S_2 \neq 0, \dots, S_{2n} \neq 0) + P(S_2 \neq 0, \dots, S_{2n} = 0) - P(S_2 \neq 0, \dots, S_{2n} \neq 0) \\ &= P(S_2 \neq 0, \dots, S_{2n-2} \neq 0) - P(S_2 \neq 0, \dots, S_{2n} \neq 0) \\ &= U_{2n-2} - U_{2n} \\ &= \frac{(2n-2)!}{(n-1)!(n-1)!} \frac{1}{2^{2n-2}} - U_{2n} \\ &= \left(\frac{2n}{2n-1} - 1 \right) U_{2n} \\ &= \frac{U_{2n}}{2n-1} \\ &\approx \frac{1}{n^{3/2}} \end{aligned}$$

donc le nombre de cas total augmente plus vite que le nombre de cas favorables.

Proposition 3.2 (Probabilité du retour à l'origine) La marche dans $\mathbb{Z}$ est récurrente car la probabilité que le marcheur revienne à l'origine est de 1 :

$$\begin{aligned} P(\text{"marcheur retour à l'origine"}) &= \sum_{k \geq 1} f_{2k} \\ &= \sum_{k \geq 1} (U_{2k-2} - U_{2k}) \\ &= U_0 \\ &= 1 \end{aligned}$$

Donc la marche aléatoire est récurrente sur $\mathbb{Z}$.

Proposition 3.3 (Temps du premier retour à l'origine) Définissons la variable aléatoire T par :

$$\begin{aligned} T(w) &:= \text{“temps du premier retour à l'origine”} \\ &= \min(k \leq 2n | S_k(w) = 0) \end{aligned}$$

Si le nombre de pas est impair, la probabilité d'un retour à l'origine est nulle :

$$P(T = 2k + 1) = 0$$

mais si le nombre de pas est pair alors la probabilité du premier retour à l'origine est :

$$P(T = 2k) = f_{2k}$$

T n'as pas d'espérance car :

$$\sum_{k=1}^{\infty} 2k P(T = 2k) = \infty$$

Proposition 3.4 (Temps du dernier retour à l'origine) Le temps du dernier retour à l'origine est :

$$Z_{2n}(w) := \max\{k \leq 2n | S_k(w) = 0\}$$

La probabilité du dernier retour à l'origine est donnée par :

$$\begin{aligned} P(Z_{2n} = 2k) &= P(S_{2k}=0, S_{2k+1} \neq 0, \dots, S_{2n} \neq 0) \\ &= P(S_{2k+1} \neq 0, \dots, S_{2n} \neq 0 | S_{2k} = 0) P(S_{2k} = 0) \\ &= U_{2k} U_{2n-2k} \\ &= \frac{1}{\sqrt{\pi k}} \frac{1}{\sqrt{\pi(n-k)}} \\ &= \frac{1}{\pi \sqrt{k(n-k)}} \\ &= \frac{1}{\pi n} \frac{1}{\sqrt{\frac{k}{n} (1 - \frac{k}{n})}} \end{aligned}$$

Pour avoir une meilleur idée de la probabilité de revenir à l'origine en fonction du temps, introduisons la densité de probabilité sur $[0, 1]$:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}} & 0 \leq x \leq 1 \\ 1 & x \geq 1 \end{cases}$$

Par intégration on obtient la fonction de répartition :

$$F(t) = \int_0^t f(s) ds = \frac{2}{\pi} \arcsin(\sqrt{t})$$

Exemple Par exemple, pour $t = \frac{1}{40}$ la probabilité d'un retour à l'origine est $F(\frac{1}{40}) = \frac{1}{10}$.

Lemme 3.1 (Principe du miroir) Soit a et b deux entiers strictement positifs et $m < n$ deux entiers. Alors le nombre de chemins de (m, a) à (n, b) et touchant l'axe des abscisses est exactement le nombre de chemin de (m, a) à $(n, -b)$.

Lemme 3.2

$$P(S_k \neq 0, k = 1, 2, \dots, 2n) = U_{2n}$$

3.7 L'échelle $a_n = \sqrt{n}$ et le Théorème de Moivre-Laplace

Théorème 3.5 (Moivre-Laplace) Soit X_k des v.a. de Bernoulli i.i.d. de paramètre p , $S_n = \sum_i X_i$ Pour tout $-\infty < a < b < \infty$:

$$\lim_{n \rightarrow \infty} P\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} < b\right) = \Phi(b) - \Phi(a) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Remarque Le théorème de Moivre-Laplace n'est qu'un cas particulier du théorème de la limite Centrale

Exemple Considérons 100 jets d'un piece de monnaie équilibrée avec X_i la variable aléatoire associée.

$$\mathbb{E}(X_i) = \frac{1}{2} \quad \text{Var}(X_i) = \frac{1}{4}$$

et pour 100 jets l'espérance et la racine de la variance de la somme sont donnés par :

$$\mathbb{E}(S_{100}) = 100 \cdot \frac{1}{2} = 50 \quad \sqrt{\text{Var}(S_{100})} = \sqrt{100 \cdot \frac{1}{4}} = 5$$

et donc la probabilité que la somme se trouve dans l'intervalle $[50 - 5; 50 + 5]$ est :

$$p([50 - 5; 50 + 5]) = 0.68$$

et la probabilité que la somme se trouve dans l'intervalle $[50 - 3.5; 50 + 3.5]$ est

$$p([50 - 15; 50 + 15]) = 0.997$$

Proposition 3.5 (Théorème de Moivre et marche aléatoire) Posons $Y_i = hX_i$ h est juste un facteur d'échelle, au lieu de faire des pas de 1, ils sont de h . Définissons $S_n = X_1 + \dots + X_n$, $\tilde{S}_n = Y_1 + \dots + Y_n$ la position du marcheur à l'instant $t = n\tau$ avec τ le pas de temps. Alors $\mathbb{E}(\tilde{S}_n) = 0$ et $\mathbb{E}(\tilde{S}_n^2) = \mathbb{E}((\text{"dist. à l'origine"})^2) = n\mathbb{E}(Y_i^2) = nh^2$. Imposons une limite à h :

$$nh^2 = Dt$$

avec $D = \frac{h^2}{\tau}$. Alors lorsque n tends à l'infini, pour t et D fixé :

$$\begin{aligned} \tau &\rightarrow 0 \\ \Rightarrow h &= \sqrt{\frac{Dt}{n}} \rightarrow 0 \\ &\Rightarrow \frac{\alpha}{\tau} \rightarrow \infty \end{aligned}$$

et le théorème de Moivre-Laplace donne la probabilité que le marcheur se situe à une distance inférieure à η de l'origine :

$$\begin{aligned} P(\tilde{S}_n \leq \eta) &= P\left(\frac{S_n}{\sqrt{n}} \leq \frac{\eta}{\sqrt{Dt}}\right) \\ &= \int_{-\infty}^{\frac{\eta}{\sqrt{Dt}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \\ &= \frac{1}{\sqrt{2\pi t D}} \int_{-\infty}^{\eta} e^{-\frac{u^2}{2tD}} du \end{aligned}$$

3.8 Théorème de la Limite Centrale (TLC)

Théorème 3.6 (Limite Centrale) La somme d'un grand nombre de v.a. i.i.d. tends vers une loi normale.

Soit $X_1, X_2, \dots$ des v.a. i.i.d. d'espérance μ et de variance σ^2 avec $\sigma^2 < \infty$, alors :

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \leq a\right) = \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{t^2}{2}} dt$$

Remarque Le Théorème de la Limite Centrale caractérise “ce qui se passe autour de l'espérance de la somme” alors que la Loi des Grands Nombres caractérise “ce qui se passe à des distances supérieurs à $n\epsilon$ de l'espérance de la somme”

Théorème 3.7 (Estimation de Berry-Esseen) Si, de plus, $\mathbb{E}(|X_i - \mu|^3) < \infty$, alors l'estimation de Berry-Esseen permet de quantifier la qualité de l'approximation du T.L.C. :

$$\sup_{t \in \mathbb{R}} \left| \lim_{n \rightarrow \infty} P\left(\frac{1}{\sqrt{\pi\sigma^2}} \sum_{i=1}^n (X_i - \mu) \leq a\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{t^2}{2}} dt \right| \leq \frac{\mathbb{E}(|X_i - \mu|^3)}{\sigma^3 \sqrt{n}}$$

3.9 Convergence faible

Définition 3.5 (Point de continuité) t est un point de continuité de F si et seulement si il n'y a pas de “sauts” en t :

$$F(t_-) = F_t$$

la continuité “à droite” est garantie par la définition de la fonction de répartition.

Définition 3.6 (Convergence faible d'une suite de fonctions de répartitions)

Une suite de fonctions de répartitions F_n converge faiblement vers la fonction de répartition F si pour tout point de continuité t on a :

$$\lim_{n \rightarrow \infty} F_n(t) = F(t)$$

Proposition 3.6 Soit X une variable aléatoire constante $P(X = a) = 1$. Alors les 2 affirmations suivantes sont équivalentes :

$$\forall \epsilon > 0 \quad \lim_{n \rightarrow \infty} P(|X_n - a| \geq \epsilon) = 0 \Leftrightarrow X_n \xrightarrow{\mathcal{L}} X$$

avec $\xrightarrow{\mathcal{L}}$ signifiant “converge en loi vers”.

Remarque On peut formuler la Loi des Grands Nombres ainsi : $X_1, X_2, \dots$ i.i.d. avec $\mathbb{E}(X_i) = \mu$, alors :

$$\frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{\mathcal{L}} \mu$$

Exemple Soit $X_n \sim F_n$. Supposons $\mathbb{E}(X_n) = \mu_n$ et $Var(X_n) = \sigma^2$ tels que

1. $\lim_{n \rightarrow \infty} \mu_n = \mu$
2. $\lim_{n \rightarrow \infty} \sigma_n^2 = 0$

et un v.a. Z constante tel que $P(Z = \mu) = 1$ alors :

$$X_n \xrightarrow{\mathcal{L}} Z$$

et

$$\lim_{n \rightarrow \infty} P(|X_n - \mu| \geq \epsilon) = 0$$

et s'il existe $\phi \in C^0 : \mathbb{R} \rightarrow \mathbb{R}$ bornée alors :

$$\lim_{n \rightarrow \infty} \mathbb{E}(\phi(X_n)) = \mathbb{E}(\phi(Z)) = \phi(\mu)$$

Chapitre 4

Statistiques

4.1 Exemples de Modèles Statistiques

4.1.1 Modèle de l'urne

Considérons une urne avec N boules dont R rouges (les autres sont blanches) et une expérience aléatoire au cours de laquelle on tire n boules ($\Omega = \{0, 1, \dots, n\}$). La probabilité de tirer r boules rouges est donnée par la loi Hypergéométrique :

$$\mathcal{H}(n, R, N)(r) = \frac{\binom{N-R}{n-r} \binom{R}{r}}{\binom{N}{n}}$$

Considérons deux cas :

1. On connaît N mais pas R et le but est d'estimer R à l'aide d'une expérience aléatoire. Soit $\theta = R$ le paramètre inconnu avec $\theta \in \Theta = \{0, 1, \dots, n\}$. On considère donc la famille de mesure de probabilité $H(n, \theta, N) = P_\theta$ et l'expérience aléatoire est de tirer n boules ($1 \leq n \leq N$) ce qui donne une estimation du nombre de boules rouges :

$$R_{\text{estimé}} = \left\lceil \frac{Nr}{n} \right\rceil$$

2. Jeu inverse : on connaît la taille de la sous-population R et on veut estimer N à l'aide d'une expérience aléatoire. On a donc $\theta = N$ avec $\theta \in \Theta = \{R, R+1, R+2, \dots\}$ et $P_\theta(r) = H(n, R, \theta)(r)$ d'où :

$$N_{\text{estimé}} = \left\lceil \frac{nR}{r} \right\rceil$$

4.1.2 Mesure d'une grandeur scalaire

Définition 4.1 (Modèle statistique) Chaque observation x_i d'une expérience est considérée comme la réalisation d'une variable aléatoire X_i dont la distribution est partiellement inconnue. La variable aléatoire X_i peut être écrite comme

$$X_i = \mu + Z_i$$

avec μ la vraie valeur de la mesure et Z_i une variable aléatoire qui modélise l'erreur.

Définition 4.2 (Erreur systématiques et accidentelles) Il est possible de distinguer deux types d'erreurs :

1. systématiques : l'erreur est toujours la même pour les mêmes conditions expérimentales. En l'absence d'erreur systématiques $\mathbb{E}(Z_i) = 0$
2. accidentelles : erreur aléatoire. En absence d'erreur aléatoire $Var(Z_i) = \sigma^2 = 0$.

Exemple Si pour une expérience on jette une pièce et on regarde si le résultat est pile, alors on modélise les jets par une loi de Bernoulli.

Remarque Une fois le choix du modèle réalisé il reste à évaluer les différents paramètres manquant pour connaître la distribution de la variable aléatoire.

Proposition 4.1 (Méthode des moindres carrés) (le but est d'estimer m et σ^2). Soit $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ le vecteur observation contenant l'ensemble des mesures d'une grandeur scalaire m . Si les mesures effectuées étaient parfaite, le vecteur observation aurait la forme

$$\underline{m} = (m^*, \dots, m^*) = m^*(1, \dots, 1) := m^* \underline{d}$$

On estime m en cherchant le vecteur $\underline{m} = (m, \dots, m)$ tel que $dist(\underline{m}, \underline{x})$ est minimale ce qui revient à minimiser

$$\sum_{i=1}^n (x_i - m)^2$$

On projette donc $\underline{x}$ sur $\underline{d}$:

$$(m, \dots, m) = \langle \underline{x}, \frac{\underline{d}}{\sqrt{n}} \rangle \frac{\underline{d}}{\sqrt{n}} = \frac{1}{n} \sum_i x_i \underline{d} = \hat{\mathbf{m}}(\underline{x}) = \bar{x}$$

Donc dans le cas de v.a. i.i.d. non-corrélées d'espérance m et de variance σ^2

$$\hat{X} = \frac{1}{n}(X_1, \dots, X_n) \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{X})^2$$

estiment respectivement m et σ^2 .

4.2 Estimation Statistique : Modèle de Gauss-Laplace

Supposons $X_i = \mu + Z_i$ avec $Z_i \sim N(0, 1)$ et les X_i indépendants et identiquement distribués, μ est inconnu.

Définition 4.3 (Modèle statistique) Un modèle statistique est un quadruplet $(\Omega, \mathcal{F}, P_\theta, \Theta)$ ou P_θ est la famille de mesures de probabilités avec θ comme paramètre et Θ est l'ensemble de définition de θ .

Définition 4.4 (Modèle de Gauss) On utilise le modèle minimal, c'est -à-dire $\Omega = \Sigma$ l'espace des échantillons. A chaque θ est associé une mesure de probabilité P_θ , la loi jointe des $X_1, \dots, X_n$, dont la densité est :

$$f_\theta(x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \right)$$

Définition 4.5 (Estimateur ou statistique) Un estimateur de θ est une application T_n de E^n dans F qui à un échantillon $(X_1, \dots, X_n)$ de loi P_θ associe une variable aléatoire réelle dont on peut déterminer la loi de probabilité.

Exemple Si la pièce est biaisée $p = \frac{2}{3}$ et que le résultat de 5 lancers est $\{1, 1, 0, 1, 0\}$ alors l'estimateur du nombre de "pile" est $\hat{p} = \frac{3}{5}$. Donc $\hat{p} \approx p$ mais ils ne sont pas égaux.

Proposition 4.2 La valeur d'un estimateur est aléatoire.

Définition 4.6 (Distribution d'échantillonnage) La loi de $\hat{\theta}$ est la *distribution d'échantillonnage*.

Remarque Pour chaque valeur du paramètre il existe une distribution d'échantillonnage associée.

Définition 4.7 (Biais d'un estimateur) Le *bias d'un estimateur* $\hat{\theta}$ est défini par :

$$b(\theta) = \mathbb{E}_\theta(\hat{\theta}) - \theta$$

avec $\mathbb{E}_\theta$ l'espérance calculée en supposant θ la vraie valeur de $\hat{\theta}$.

Proposition 4.3 Un estimateur est dit non-biaisé si et seulement si $b(\theta) = 0$.

Définition 4.8 (Carré moyen d'erreur) Le carré de l'erreur d'un estimateur est défini par $(\theta - \hat{\theta})^2$. L'espérance du carré de l'erreur est le *carré moyen d'erreur*, CME, défini par :

$$CME_\theta(\hat{\theta}) = \mathbb{E}_\theta((\theta - \hat{\theta})^2)$$

Proposition 4.4

$$\mathbb{E}_\theta((\theta - \hat{\theta})^2) = \text{Var}_\theta(\hat{\theta}) + b^2(\theta)$$

Définition 4.9 (Meilleur estimateur) Si $\hat{\theta}_1$ et $\hat{\theta}_2$ sont deux estimateurs non-biaisés pour un même paramètre θ alors on dit que $\hat{\theta}_1$ est un *meilleur estimateur* que $\hat{\theta}_2$ si

$$\text{Var}_\theta(\hat{\theta}_1) < \text{Var}_\theta(\hat{\theta}_2)$$

Exemple Dans le modèle de Gauss $\bar{x} = \frac{1}{n}(X_1 + \dots + X_n)$ est un estimateur non biaisé de m . La loi de $\bar{X} = \hat{m}$ (i.e. la distribution d'échantillonnage) est $N\left(m, \frac{\sigma^2}{n}\right)$.

Définition 4.10 (Fonction de vraisemblance) Considérons le cas où la famille de mesures de probabilités de P_θ est donnée par $f(x, \theta)$. Pour θ fixé on définit alors la *fonction de vraisemblance* :

$$L_x(\theta) : \theta \mapsto f(x, \theta)$$

Définition 4.11 (Estimateur de maximum de vraisemblance) L'idée est de choisir, pour un $\underline{x}$ donné, le θ qui rend maximale la probabilité. $\hat{\theta}$ est un *estimateur de maximum de vraisemblance* si et seulement si :

$$L_x(\hat{\theta}) \geq L_x(\theta) \quad \forall \theta \in \Theta$$

avec Θ l'ensemble de valeurs possible de θ .

Exemple Pour le modèle de Gauss avec σ^2 connu, il reste à définir le paramètre $\theta = m$. On étudie :

$$m \mapsto \ln(L_{\underline{X}}(\theta))$$

on cherche le m qui extremise cette expression :

$$\frac{d}{dm} \ln(L_{\underline{X}}(\theta)) = 0$$

implique que l'estimateur de maximum de vraisemblance est $\hat{m}(\underline{X}) = \frac{1}{n} \sum_{i=1}^n x_i$

Définition 4.12 (Loi du χ^2) Soit $Z_1, \dots, Z_n$ n v.a. i.i.d. avec $Z_i \sim N(0, 1)$. Par définition la loi $\sum_k Z_k^2$ est celle du χ_n^2 .

Théorème 4.1 Si $X_1, \dots, X_n$ sont i.i.d. et $X_i \sim N(m, \sigma^2)$,

$$\hat{m} = \bar{X} := \frac{1}{n}(X_1, \dots, X_n) \quad S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

alors

1. $\bar{X} \sim N\left(m, \frac{\sigma^2}{n}\right)$
2. S^2 et $\bar{X}$ sont indépendants.
3. $\mathbb{E}_{(m, \sigma^2)}(S^2) = \sigma^2 \sim \chi^2$ à $n-1$ degrés de libertés, notée χ_{n-1}^2 .

4.3 Intervalle de confiance

Dans le modèle de Gauss que signifie $m = \bar{x}_n \pm \frac{s_n(x)}{\sqrt{n}}$? Pour chaque échantillon on définit un intervalle de confiance.

Définition 4.13 (Intervalle de confiance de niveau $1 - \alpha$) Soit $\alpha > 0$ petit (typiquement 0.05). Un intervalle de confiance de niveau $1 - \alpha$ est défini par deux variables aléatoires T_+ et T_- de $\Sigma \rightarrow \Theta = \mathbb{R}$ tel que pour tout $\theta \in \Theta$:

$$P_\theta(\theta \in [T_-(\underline{X}), T_+(\underline{X})]) = 1 - \alpha \quad (4.1)$$

Remarque Le paramètre m n'est pas aléatoire.

Remarque L'équation 4.1 est vraie pour tout m et σ^2 , en particulier pour leurs vraies valeurs (inconnues).

Remarque La Loi des Grands Nombres implique que si on répète n fois l'expérience 95% des intervalles $I(x)$ contiennent la vraie valeur du paramètre m .

Définition 4.14 (Loi de Student) Si $X_1, \dots, X_n$ v.a. i.i.d avec $X_i \sim N(0, 1)$ alors la v.a.

$$T_n := \frac{\sqrt{n}(\bar{X}_n - m)}{S_n}$$

à une loi de Student t_{n-1} à $n - 1$ degrés de liberté.

Théorème 4.2 Si $X_1, \dots, X_n$ v.a. i.i.d. et $\mathbb{E}(X_i) = m$, $Var(X_i) = \sigma^2$ alors la v.a. T_n converge faiblement vers une v.a. $Z \sim N(0, 1)$.

Proposition 4.5 (Construction de l'intervalle de confiance) On dispose de n mesures d'une quantité μ :

$$y_i = \mu + \epsilon_i \quad \forall i = 1, \dots, n$$

et nous supposons les ϵ_i iid $\sim N(0, 1)$. La moyenne $\bar{\mu}$ est un estimateur de μ . σ^2 est connu

$$\frac{\bar{X}_n - m}{\sqrt{Var(\bar{X}_n)}}$$

à une loi $N(0, 1)$. En posant $Z = \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$

$$\begin{aligned} P_{(m, \sigma^2)} \left(Z \in \left[q \left(\frac{\alpha}{2} \right), q \left(1 - \frac{\alpha}{2} \right) \right] \right) &= 1 - \alpha \\ &= P \left(|Z| \leq q \left(1 - \frac{\alpha}{2} \right) \right) \end{aligned}$$

si $\alpha = 0.05$, $q(1 - \frac{\alpha}{2}) = 1.96$ et $m = \bar{X}_n \pm 1.96 \frac{\sigma}{\sqrt{n}}$.

σ^2 inconnu Si σ^2 n'est pas connu on utilise les estimations :

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n) \quad S^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$$

qui permettent de définir la loi de student à $n-1$ degré de liberté $T_n = \sqrt{n} \frac{|\bar{X} - \mu|}{S}$. Ainsi il est possible de construire l'intervalle de confiance pour le paramètre $\theta = \mu$ pour α donné comme :

$$\begin{aligned} P_\theta \left(|T_n| \leq q_{t_{n-1}} \left(1 - \frac{\alpha}{2} \right) \right) &= 1 - \alpha \\ \Leftrightarrow P_\theta \left(\left\{ \underline{x} : \theta \in \left[\bar{X}(x) \pm \frac{S(x)}{\sqrt{n}} q_{t_{n-1}} \left(1 - \frac{\alpha}{2} \right) \right] \right\} \right) &= 1 - \alpha \end{aligned}$$

Exemple On a mesuré la longueur de huit cylindres, les résultats sont :

29.4 30.8 30.6 31.5 32.1 31.7 30.3 30.8

La moyenne est $\mu = 30.90$ et l'écart-type $s = 0.86$. L'intervalle est donc de confiance au niveau $1 - \alpha$ avec $\alpha = 0.025$ est :

$$I_\pm = \mu \pm \frac{s}{\sqrt{n}} q_{t_7}(0.975) = 30.90 \pm \frac{0.86}{\sqrt{8}} * 2.365$$

Donc la probabilité que l'intervalle $[30.18, 31.62]$ recouvre la vraie valeur de μ est de 0.975.

Remarque Les intervalles sont aléatoires et la probabilité que ces intervalles aléatoires couvrent la vraie valeur du paramètre lorsqu'on répète l'expérience est de 0.95 (pour $\alpha = 0.05$).

4.4 Notion de test

Les observations $(x_1, \dots, x_n)$ sont des réalisations indépendantes et identiquement distribuées selon une loi sous-jacente connue à l'exception d'un paramètre θ . L'hypothèse nulle correspond à la valeur du paramètre supposée exacte.

Définition 4.15 (p-valeur) Soit S_{obs} , la valeur observée d'un score lors d'une expérience. La probabilité

$$P_{sous H_0} = \{S \text{ plus extrême que } S_{obs}\}$$

est dite la *p-valeur* du test. On rejette H_0 si et seulement si la valeur du test est inférieur à 5%.

Marche à suivre pour un test de signification à un niveau 5%.

1. Formulation de l'hypothèse H_0 .
2. choix d'un score de test S et calcul de la valeur observée de ce score S_{obs} .
3. calcul de la p-valeur
4. rejeter l'hypothèse si la p-valeur est inférieur à 5%.

Définition 4.16 (Erreur de première et seconde espèce) Si l'on rejette H_0 car on a observé un événement $x \in \mathbb{R}$ alors que H_0 était vrai, on fait *une erreur de première espèce*.

Si l'on ne rejette pas H_0 alors que l'hypothèse était fausse, on commet *une erreur de seconde espèce*.

Théorème 4.3 (Test du χ^2 de Pearson) Le but de ce test est de vérifier que la répartition de l'échantillon correspond à une distribution théorique.

Le score observé du test de Pearson est donné par :

$$Pearson_{observe} = \sum_{i=1}^I \frac{n_i - np_i}{np_i}$$

avec n le nombre de mesures, n_i le nombre d'observation de la "face" i et p_i les fréquences théoriques, et I le nombre de "faces". L'hypothèse nulle testée est

$$H_0 = \{\text{la distrib. théorique est la distrib. sous-jacente aux données}\}$$

La distribution de la statistique de Pearson est bien approchée par une loi de χ^2_{I-1} . L'hypothèse nulle est rejetée pour un test de niveau 5% si

$$Pearson_{observe} > q\chi^2_{I-1}(95\%)$$

avec $q\chi^2_{I-1}(95\%)$ le 95% quantile d'une loi χ^2_{I-1} .

Proposition 4.6 Dans la pratique on vérifie si $np_j \geq 10$ pour tout j pour utiliser la loi de χ^2 pour T .

Exemple Supposons un dé à $k=6$ faces équilibré, c'est-à-dire $H_0 = \{p(\frac{1}{6}, \dots, \frac{1}{6})\}$ que l'on lance 120 fois et effectuons un test de niveau 5% :

1	2	3	4	5	6
26	20	16	27	15	16

La loi du $t_{\text{theorique}} = \chi_5^2(1 - 0.05) = 11.07$ alors que

$$t_{\text{observe}} = \frac{(26 - 20)^2 + \dots + (16 - 20)^2}{20} = 7.1 \leq 11.07$$

donc le dé est équilibré.

Bibliographie

- [1] Charles Pfister *Notes de cours* (Christophe Mortier et Raphael Faerber)
- [2] Sheldon M. Ross *Initiation aux probabilités* Presses Polytechniques et Universitaires Romandes, 2007
- [3] Jean-Pierre Lecoutre *Statistique et probabilités* Dunod, 2003
- [4] Eva Cantoni, Philippe Huber, Elvezio Ronchetti *Maîtriser l'aléatoire* Springer, 2006
- [5] Stephan Morgenthaler *Introduciton à la statistique*, Presses Polytechniques et Universitaires Romandes, 2000