

Cours de Probabilités, L3S5, UE 14A

(Licence de Physique de l'Université Louis Pasteur)

J. FRANCHI

Premier semestre 2009-2010. 28 h de cours et 28 h de T.D..

Contenu

I. Fondements du calcul des probabilités	page 2
1) Probabilité ; conditionnement	page 2
2) Variables aléatoires et leur lois	page 9
3) Lois usuelles	page 13
4) Lemmes de Borel-Cantelli	page 20
5) Variables aléatoires indépendantes	page 22
6) Transformées de Laplace et de Fourier	page 23
7) Convergences des variables aléatoires	page 26
8) Loi des Grands Nombres	page 28
9) Théorème Central Limite	page 29
10) Exemple des marches aléatoires	page 30
11) Processus de Poisson	page 34
II. Eléments de statistique mathématique	page 38
12) Caractères quantitatifs monodimensionnels	page 38
13) Régression linéaire	page 44
14) Vraisemblance et estimation	page 48
15) Tests	page 59

Les probabilités et la statistique interviennent naturellement de façon importante dans plusieurs grands domaines de la physique : mécanique quantique, percolation, mécanique statistique (transitions de phases, croissance de polymères, verres de spin, etc...).

Elles sont essentielles également en épidémiologie (depuis I. Semmelweiss, voir la thèse de L.F. Céline), source de la *grande majorité* des progrès médicaux depuis 30 ans (selon P. Corvol, médecin épidémiologiste du Collège de France), et dans la pratique des sondages.

Enfin elles sont importantes aussi en économie, en étude de fiabilité, et dans bien d'autres domaines (si le juge de l'affaire d'Outreau avait appris à penser tant soit peu en termes de calcul des probabilités élémentaires, il eut probablement évité de croire vraisemblable le regroupement de tous les pédophiles de la ville dans une même cage d'escalier HLM, où les attributions sont a priori aléatoires...).

I. Fondements de la théorie des probabilités

1 Probabilité; conditionnement

Une probabilité est d'abord une *fonction* qui à un *événement* associe un nombre réel compris entre 0 et 1. Cela implique de préciser ce qu'est un événement. Or cela n'a de sens que dans le cadre d'un ensemble d'*épreuves aléatoires* ou *tirages*, qu'on note généralement Ω . Il peut s'agir par exemple de lancers de dés ou de pièces de monnaie, de tirages d'urne, de durées de vie (d'atomes ou d'individus), de tirs sur une cible, etc... Ces premiers exemples familiers montrent déjà que l'ensemble Ω peut être fini, dénombrable (ce qui signifie : infini indexé par \mathbb{N} ou \mathbb{N}^*), ou continu. Ce sera donc a priori un ensemble non vide quelconque.

Lorsque Ω est fini ou dénombrable, toutes ses parties seront des événements. Tandis qu'en général il est nécessaire de se restreindre à un sous-ensemble de parties de Ω : $\mathcal{T} \subset \mathcal{P}(\Omega)$, qu'on nomme *tribu* (ou σ -*algèbre*). On a naturellement besoin de pouvoir considérer la réunion et la conjonction (\equiv intersection) de 2 événements, de même que le complémentaire (\equiv contraire) d'un événement ; en outre il faut aussi pouvoir considérer une réunion dénombrable d'événements. Enfin il est naturel de considérer l'événement impossible (\equiv vide : \emptyset) et l'événement certain Ω . D'où la définition suivante.

Définition 1.0.1 Une *tribu* (ou σ -*algèbre*) est une partie \mathcal{T} de $\mathcal{P}(\Omega)$ stable par réunion dénombrable et par passage au complémentaire, et contenant l'ensemble vide \emptyset .

Le couple (Ω, \mathcal{T}) est appelé *espace probabilisable*. Un *événement* est un élément de \mathcal{T} .

$\mathcal{P}(\Omega)$ désigne l'ensemble de toutes les parties de Ω ; \mathcal{T} est donc un ensemble de parties de Ω . La stabilité par réunion dénombrable s'écrit formellement : pour toute suite $\{E_n \mid n \in \mathbb{N}\} \subset \mathcal{T}$ d'événements, leur réunion est aussi un événement : $\bigcup_{n \in \mathbb{N}} E_n \in \mathcal{T}$.

La stabilité par passage au complémentaire s'écrit formellement : le complémentaire $E^c := \Omega \setminus E = \{\omega \in \Omega \mid \omega \notin E\}$ de tout événement E est aussi un événement : $E^c \in \mathcal{T}$.

Nota Bene Sur Ω fini ou dénombrable, on choisira toujours par défaut la tribu $\mathcal{P}(\Omega)$.

Rappel : $\left[\bigcap_n A_n\right]^c = \bigcup_n A_n^c$; $\left[\bigcup_n A_n\right]^c = \bigcap_n A_n^c$. On vérifie aussitôt les propriétés suivantes :

Proposition 1.0.2 Ω est un événement (certain). La réunion et l'intersection d'un nombre fini d'événements sont des événements. La différence $A \setminus B := A \cap B^c$ et la différence symétrique $A \Delta B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$ de deux événements A et B sont des événements. Toute intersection dénombrable d'événements est un événement.

Nous pouvons maintenant définir rigoureusement ce qu'est une probabilité; dans la pratique, ce sera le plus souvent soit une somme (soit finie, soit infinie, c'est-à-dire une série) ou une intégrale (soit propre, dite de Riemann, soit impropre). (Pour rassembler toutes ces possibilités, et d'autres encore, on parle de *mesure*.)

Définition 1.0.3 Une probabilité sur l'espace probabilisable (Ω, \mathcal{T}) est une fonction \mathbb{P} de \mathcal{T} dans $[0, 1]$ qui vérifie: $\mathbb{P}(\Omega) = 1$, et la propriété d'additivité dénombrable:

$\mathbb{P}\left(\bigsqcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$ pour tout suite $\{A_n | n \in \mathbb{N}\}$ d'événements deux à deux disjoints. Le triplet $(\Omega, \mathcal{T}, \mathbb{P})$ est appelé espace probabilisé ou espace de probabilité. Les événements de probabilité nulle sont dits négligeables. Les événements de probabilité 1 sont dits presque sûrs.

C'est toujours dans le cadre d'un espace de probabilité, plus ou moins bien précisé, que peut avoir lieu un calcul de probabilité. Il est généralement préférable de bien le préciser. En effet c'est la non-précision de l'espace considéré qui est à l'origine de paradoxes ou d'erreurs courantes sur les probabilités.

On vérifie aisément les propriétés suivantes:

Proposition 1.0.4 (i) L'événement impossible est négligeable: $\mathbb{P}(\emptyset) = 0$.

(ii) \mathbb{P} est croissante: $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$, pour tous événements A et B .

(iii) $\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$ pour tout suite $\{A_n | n \in \mathbb{N}\}$ d'événements.

(iv) $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$, lorsque A et B sont deux événements tels que $B \subset A$. En particulier, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ pour tout événement A .

(v) Toute intersection dénombrable d'événements presque sûrs est presque sûre.

(vi) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, pour tous événements A et B .

(vii) \mathbb{P} est continue: $\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$, pour toute suite croissante d'événements $\{A_n | n \in \mathbb{N}\}$ (id est $A_n \subset A_{n+1} (\forall n)$); et de même $\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n)$, pour toute suite décroissante d'événements $\{B_n | n \in \mathbb{N}\}$ (id est $B_n \supset B_{n+1} (\forall n)$).

1.1 Exemples

1. Probabilité discrète sur un ensemble $\Omega = \{\omega_1, \dots, \omega_N\}$ fini:

elle est clairement définie par la liste des probabilités des singletons: $p_j := \mathbb{P}(\{\omega_j\})$.

Nous avons en effet $\mathbb{P}(A) = \sum_{\omega_j \in A} p_j$ pour toute partie $A \subset \Omega$.

Réciproquement, toute liste $\{p_1, \dots, p_N\}$ de réels $p_j \geq 0$ tels que $\sum_{j=1}^N p_j = 1$ définit bien (par la même formule) une probabilité unique sur Ω .

Exemples concrets : lancers de dés, de pièces de monnaie, tirages de cartes à jouer, tirages de boules dans des urnes, loteries, etc. . .

2. Probabilité discrète sur \mathbb{N} (ou sur n'importe quel autre ensemble dénombrable) :

elle est encore définie par la liste des probabilités des singletons : $p_j := \mathbb{P}(\{\omega_j\})$.

Nous avons en effet $\mathbb{P}(A) = \sum_{\omega_j \in A} p_j$ pour toute partie $A \subset \mathbb{N}$. La seule différence avec le cas précédent est que la somme peut être une série (\equiv comporter une infinité de termes).

Réciproquement, toute suite $\{p_j \mid j \in \mathbb{N}\}$ de réels $p_j \geq 0$ tels que $\sum_{j \geq 1} p_j = 1$ définit bien (par la même formule, forcément convergente) une probabilité unique sur \mathbb{N} .

3. *Cordes.* On tire une corde au hasard dans un disque de rayon R . Quelle est la probabilité que la longueur ℓ de la corde soit $\geq R$?

a. ℓ varie continûment dans $[0, 2R]$, de sorte que la probabilité cherchée vaut $1/2$.

b. ℓ est déterminée par la distance d de la corde au centre du disque ; d varie continûment dans $[0, R]$, et $\ell = 2\sqrt{R^2 - d^2} \geq R \Leftrightarrow d \leq R\sqrt{3}/2$, de sorte que la probabilité cherchée vaut $\sqrt{3}/2$.

c. ℓ est déterminée par le milieu M de la corde, qui varie continûment dans le disque ; et $\ell \geq R$ a lieu ssi M est dans le disque concentrique de rayon $\sqrt{3}/2$, de sorte que la probabilité cherchée vaut $3/4$.

Explication : la probabilité choisie est très insuffisamment précisée par l'expression "tirage au hasard". Ici on a considéré successivement la probabilité uniforme sur l'ensemble : des longueurs, des distances au centre, des milieux. Ce sont trois probabilités différentes !

4. Jeu de pile ou face illimité ; première apparition de "pile", ou d'une séquence donnée.

Exercice n° 1.1 Est-il plus probable d'obtenir au moins une fois 6 en lançant 4 dés usuels, ou bien d'obtenir au moins une fois un double 6 en lançant 24 fois 2 dés usuels ?

Exercice n° 1.2 On lance n fois de suite 3 dés normaux. Pour quelles valeurs de n la probabilité d'avoir obtenu au moins un 421 dépasse-t-elle $\frac{1}{2}$?

Exercice n° 1.3 On lance 5 pièces de monnaie. Calculer les probabilités des événements suivant : "la 1ère pièce donne face" ; "face sort exactement 2 fois" ; "face sort au plus 3 fois".

Exercice n° 1.4 On lance 10 dés usuels. Calculer les probabilités des événements suivant : "6 ne sort pas" ; "6 sort 1 fois exactement" ; "6 sort 3 fois exactement" ; "6 sort 2 fois au moins" ; "6 sort 3 fois au moins".

Exercice n° 1.5 Une armoire contient 10 paires de chaussures, parmi lesquelles on prélève au hasard 8 chaussures. Quelle est la probabilité d'avoir ainsi k paires de chaussures exactement ?

Exercice n° 1.6 Une urne contient n boules noires et b boules blanches. Deux joueurs X et Y tirent avec remise une boule dans l'urne, tour à tour, X tirant le premier. Quelle est la probabilité que X soit le premier à tirer une boule noire ? Même question sans remise.

Exercice n° 1.7 Une loterie comporte 100 billets, dont les seuls billets gagnants suivant : 1 billet gagne 50 euros, 5 billets gagnent chacun 30 euros, 10 billets gagnent chacun 10 euros. Quelle est la probabilité qu'un acheteur de 3 billets gagne 30 euros (au moins, puis exactement) ?

Exercice n° 1.8 Un joueur X lance 2 dés usuels, et obtient ainsi la somme S .

- Calculer la probabilité que $S > n$, en fonction des différentes valeurs de l'entier n .
- Un joueur Y relance les 2 dés et obtient une somme T . Quelles sont les probabilités que $S = T$, que $S > T$, que $S \geq T$?

Proposition 1.1.1 (*Formule de crible, de Poincaré*) Pour tout espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$, tout entier $n \in \mathbb{N}^*$, et tous événements A_1, \dots, A_n , on a :

$$\mathbb{P}[A_1 \cup \dots \cup A_n] = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}[A_{i_1} \cap \dots \cap A_{i_k}].$$

Preuve Par récurrence sur n : exercice.

Exercice n° 1.9 Un sac contient n jetons numérotés de 1 à n , qu'on tire tous 1 à 1, sans remise. *i)* Calculer $p_n := \mathbb{P}$ (au moins un jeton sorte au rang indiqué par son numéro), sa limite p_∞ , et majorer $|p_n - p_\infty|$.

ii) Soit $p_n(k) := \mathbb{P}$ (exactement k jetons sortent au rang indiqué par leur numéros), pour $k \in \{0, \dots, n\}$. Dédurre de *(i)* une formule pour $p_n(k)$, puis $\lim_{n \rightarrow \infty} p_n(k)$ (pour tout $k \in \mathbb{N}$).

1.2 Probabilités conditionnelles

Définition 1.2.1 (Probabilité conditionnelle) Fixons un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$, et un événement $C \in \mathcal{T}$, non négligeable. La probabilité conditionnelle relative à C (ou "sachant C ") est définie par : $\mathbb{P}(A/C) := \mathbb{P}(A \cap C) / \mathbb{P}(C)$.

On vérifie immédiatement qu'il s'agit encore d'une probabilité sur (Ω, \mathcal{T}) .

Exercice n° 1.2.1 Lancer de 2 dés usuels : $\Omega = \{1, \dots, 6\}^2$. \mathbb{P} uniforme. Soient X_1 le chiffre indiqué par le premier dé, S la somme des chiffres indiqués par les 2 dés, et $C = \{S = 5\}$. Dresser le tableau des valeurs de $\mathbb{P}(\cdot/C)$, puis de $\mathbb{P}(X_1 = \cdot/C)$.

Exercice n° 1.2.2 Vous allez chez des gens dont vous savez qu'ils ont 2 enfants, dont au moins une fille. a) Quelle est la probabilité que l'autre enfant soit aussi une fille ?

b) En l'absence de l'information qu'ils ont au moins une fille (pour cette question seulement), mais en voyant une fille ouvrir la porte, quelle est la probabilité que l'autre enfant soit aussi une fille ? c) Une fille vous ouvre la porte ; quelle est la probabilité que l'autre enfant soit aussi une fille ?

Exercice n° 1.2.3 Vous attendez un ami de Vancouver, qui voyage jusqu'à Strasbourg avec changement d'avion à New York, Londres et Francfort. La probabilité d'attentat est estimée à p pour chacun des 4 vols, avec indépendance entre les 4 vols. Votre ami n'arrivant pas, quelle est la probabilité que l'attentat ait eu lieu : a) dans le 1er avion ? b) dans le 2ème avion ? c) dans le 3ème avion ? c) dans le 4ème avion ?

Pour effectuer un calcul, il est très souvent indispensable de pouvoir "distinguer des cas". Cela s'exprime par la formule suivante, très élémentaire et très utile à la fois.

Proposition 1.2.2 (Formule des probabilités totales) *Fixons un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ et une partition de Ω en événements non négligeables : $\Omega = \bigsqcup_{j=1}^N C_j$. Alors nous*

avons $\mathbb{P}(A) = \sum_{j=1}^N \mathbb{P}(A/C_j)\mathbb{P}(C_j)$, pour tout événement A .

Exercice n° 1.2.4 Une urne contient b boules blanches et n boules noires. Quand une boule est tirée, on le remet dans l'urne, avec ℓ boules de la même couleur. On effectue ainsi 3 tirages au hasard. a) Quelle est la probabilité que la 1ère boule tirée soit noire sachant que la seconde est blanche ? b) Quelle est la probabilité que la 3ème boule soit noire ?

On a souvent à inverser un conditionnement. Cela se fait simplement, au moyen de la formule élémentaire suivante, très utile aussi, quoiqu'également de preuve immédiate.

Proposition 1.2.3 (Formule de Bayes) *Fixons un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$, et une partition de Ω en événements non négligeables : $\Omega = \bigsqcup_{j=1}^N C_j$. Alors nous avons pour tout événement non négligeable A et tout $k \in \{1, \dots, N\}$:*

$$\mathbb{P}(C_k/A) = \mathbb{P}(A/C_k)\mathbb{P}(C_k) / \sum_{j=1}^N \mathbb{P}(A/C_j)\mathbb{P}(C_j).$$

Exemple : Les candidats à un examen proviennent de 4 lycées K,L,V,W, à raison de 20% pour K, de 25% pour L, de 15% pour V, et de 40% pour W. K enregistre 35% de succès, L 30%, V 50%, W 45%. Alors la probabilité qu'un candidat reçu provienne de K est $\frac{2}{13}$, de L est $\frac{5}{26}$, de K est $\frac{5}{26}$, de K est $\frac{6}{13}$.

Exercice n° 1.2.5 Trois machines U, V, W produisent une même pièce dans une usine. U assure 40% de la production, V 35%, et W le reste. U produit 20% de pièces défectueuses, V 15%, et W 10%.

- a) Quelle est la probabilité qu'une pièce prise au hasard soit défectueuse ?
 b) Quelle est la probabilité qu'une pièce défectueuse prise au hasard provienne de U ?

Exercice n° 1.2.6 Trois condamnés X, Y, Z sont informés que l'un d'eux, choisi au hasard, va être exécuté, et que les 2 autres vont être libérés. Mais ils ne doivent pas encore savoir qui le hasard a désigné. X demande au geôlier de lui nommer l'un de ses 2 codétenus devant être libéré, arguant que cette information serait innocente, puisqu'il sait que l'un des 2 au moins doit l'être. Le geôlier refuse, arguant que cette information modifierait réellement l'estimation que X peut faire de ses chances. Qui a raison ?

Exercice n° 1.2.7 12% des individus d'une population sont porteurs d'une certaine maladie. Un test relatif à cette maladie est fiable à 95%, dans le cas d'un malade comme dans le cas d'un sujet sain. a) Quelle est la probabilité qu'un individu présentant un test positif soit effectivement malade ? b) Quelle est la probabilité qu'un individu présentant un test négatif soit effectivement sain ?

Exercice n° 1.2.8 Émile possède 5 pièces de monnaie, dont 2 sont normales, 2 ont 2 côtés "face", et une a 2 côtés "pile".

- a) Il prend une pièce au hasard et la lance ; quelle est la probabilité qu'il voie "face" ?
 b) Il voit "face" ; quelle est la probabilité que l'autre côté de la pièce soit aussi "face" ?

Il relance la même pièce.

- c) Quelle est la probabilité que le côté caché de la pièce soit "face" ?
 d) Il voit de nouveau "face" ; quelle est la probabilité que l'autre côté de la pièce soit aussi "face" ? Il choisit ensuite au hasard une des autres pièces et la lance.
 e) Quelle est la probabilité de voir de nouveau "face" (pour la troisième fois) ?

Exercice n° 1.2.9 Un livre a une probabilité $p > 0$ de se trouver dans une commode comportant k tiroirs, et des chances égales de se trouver dans chacun des tiroirs.

i) On ouvre les $(k - 1)$ premiers tiroirs, sans le trouver ; quelle est la probabilité de le trouver dans le dernier tiroir ?

ii) Soit $j \in \{2, \dots, k - 1\}$. On ouvre les $(k - j)$ premiers tiroirs, sans le trouver ; quelle est la probabilité de le trouver dans le dernier tiroir ? dans l'un des j derniers tiroirs ?

Exercice n° 1.2.10 Le quart d'une population est vacciné contre le choléra. Au cours d'une épidémie, on constate qu'il y a parmi les malades un vacciné pour 4 non-vaccinés, et qu'il y a un malade sur 12 parmi les vaccinés. Quelle est la probabilité qu'un non-vacciné tombe malade ? Le vaccin est-il efficace ?

1.3 Événements indépendants

L'indépendance est au cœur de la théorie des probabilités. On l'aborde en douceur, par le cas des événements. Le cas général des variables aléatoires viendra ensuite (section 5).

Définition 1.3.1 Fixons un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$. Deux événements A et B sont dits indépendants lorsque $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

- Exemples : 1) “tirer un roi” et “tirer un trèfle”, dans un jeu de bridge, ou de belote.
 2) “obtenir un chiffre pair avec le 1er dé” et “obtenir un 6 avec le 2ème dé” $\equiv \{X_2 = 6\}$, lors du lancer de 2 dés : tirages indépendants (au sens usuel) !
 3) “obtenir une somme paire” et “obtenir un 5 avec le 2ème dé”, lors du lancer de 2 dés.
 4) “obtenir une somme S égale à 7” et “obtenir un produit égal à 12”, lors du lancer de 2 dés (non indépendants !)
 5) $\{S = 7\}$ et $\{X_2 = 2\}$, ou bien $\{S = 6\}$ et $\{X_2 = 2\}$, lors du lancer de 2 dés.
 6) “obtenir un produit pair” et “obtenir un 5 avec le 2ème dé”, lors du lancer de 2 dés.

Exercice n° 1.3.1 Montrer que 2 événements A et B sont indépendants ssi A et B^c le sont, et ssi A^c et B^c le sont, ou bien encore (lorsqu'ils sont non négligeables) ssi $\mathbb{P}(A/B) = \mathbb{P}(A)$, et ssi $\mathbb{P}(B/A) = \mathbb{P}(B)$.

Exercice n° 1.3.2 Une urne contient des jetons numérotés, rouges ou noirs. Lors du tirage d'un jeton la probabilité d'en tirer un rouge est $3/5$; d'en tirer un de numéro impair est $2/3$; d'en tirer un rouge et pair est p . Que vaut la probabilité d'en tirer un noir impair ? Pour quelles valeurs de p les événements “noir” et “impair” sont-ils indépendants ?

Définition 1.3.2 Fixons un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$. Des événements A_1, \dots, A_n sont dits indépendants lorsque $\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k})$ pour tous $1 \leq i_1 < \dots < i_k \leq n$. Les événements d'une suite sont dits indépendants lorsque toute sous-suite finie est constituée d'événements indépendants.

Proposition 1.3.3 Les événements A_1, \dots, A_n sont indépendants ssi

$\mathbb{P}(A_1^{\varepsilon_1} \cap A_2^{\varepsilon_2} \cap \dots \cap A_n^{\varepsilon_n}) = \mathbb{P}(A_1^{\varepsilon_1})\mathbb{P}(A_2^{\varepsilon_2}) \dots \mathbb{P}(A_n^{\varepsilon_n})$ pour tout choix de $\varepsilon_1, \dots, \varepsilon_n$, avec soit $A_j^{\varepsilon_j} = A_j$ soit $A_j^{\varepsilon_j} = A_j^c$.

Preuve Pour le sens direct, on vérifie par récurrence sur $k \in \{0, \dots, n\}$ que pour tout choix de $\varepsilon_1, \dots, \varepsilon_k$, tout $\ell \in \{k+1, \dots, n\}$ et tous $i_{k+1}, \dots, i_\ell \in \{k+1, \dots, n\}$, on a :

$$P[A_1^{\varepsilon_1} \cap \dots \cap A_k^{\varepsilon_k} \cap A_{i_{k+1}} \cap \dots \cap A_{i_\ell}] = P[A_1^{\varepsilon_1}] \times \dots \times \mathbb{P}[A_k^{\varepsilon_k}] \times \mathbb{P}[A_{i_{k+1}}] \times \dots \times \mathbb{P}[A_{i_\ell}].$$

Pour la réciproque, si i_1, \dots, i_k sont fixés dans $\{1, \dots, n\}$, on note $\{j_1, \dots, j_{n-k}\} := \{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$, et on applique la formule des probabilités totales avec la partition $\{A_{j_1}^{\varepsilon_{j_1}} \cap \dots \cap A_{j_{n-k}}^{\varepsilon_{j_{n-k}}} \mid \varepsilon_{j_1}, \dots, \varepsilon_{j_{n-k}} = \text{ ou } c\}$ (c'est bien une partition de Ω , car l'ensemble des intersections des éléments de 2 partitions en forme encore une). \diamond

Exercice n° 1.3.3 On jette un dé normal n (≥ 3) fois de suite. Montrer que les événements {les lancers j et k donnent le même résultat} sont deux à deux indépendants, mais non indépendants.

Exercice n° 1.3.4 Sur $\Omega := \{a, b, c, d\}$ on définit \mathbb{P} par : $\mathbb{P}(\{a\}) = \alpha$, $\mathbb{P}(\{b\}) = \beta$, $\mathbb{P}(\{c\}) = \gamma$, $\mathbb{P}(\{d\}) = \delta$. Trouver les valeurs de $\alpha, \beta, \gamma, \delta$ telles que les événements $A := \{b, c\}$, $B := \{c, a\}$, $C := \{a, b\}$ soient 2 à 2 indépendants mais non indépendants.

Exercice n° 1.3.5 Pour ridiculiser un astrologue (ou un “voyant”, ou n’importe quelle autre sorte de charlatan), on le défie de prédire le résultat de 12 lancers successifs d’une pièce de monnaie usuelle. Quelle est la probabilité (en fonction de $n \in \{0, 1, \dots, 5\}$) qu’il se trompe au plus n fois ?

2 Variables aléatoires et leur lois

Définition 2.0.4 Une *variable aléatoire* (“v.a.”) est une fonction V définie sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ et à valeurs dans \mathbb{R}^d , telle que $\{V \in E\} = V^{-1}(E) \in \mathcal{T}$, pour tout E pavé (ou ouvert ou fermé) de \mathbb{R}^d . Sa *loi* est la probabilité $\mathbb{P}_V := \mathbb{P} \circ V^{-1}$ sur \mathbb{R}^d , définie par : $\mathbb{P}_V(E) := \mathbb{P}(V^{-1}(E)) = \mathbb{P}(V \in E)$. Lorsque $d = 1$, on parle de *variable aléatoire réelle*, “v.a.r.”.

On vérifie immédiatement que la loi d’une variable aléatoire est bien une probabilité (sur $V(\Omega) \subset \mathbb{R}^d$, ou directement sur \mathbb{R}^d , la tribu étant celle des “boréliens”, engendrée par les pavés ou par les ouverts ou par les fermés de \mathbb{R}^d). Notons que dans le cas où Ω est discret (fini ou dénombrable), dans la définition ci-dessus la condition de mesurabilité sur V est vide, c’est-à-dire qu’elle est forcément vérifiée par n’importe quelle fonction V .

On vérifie les propriétés suivantes, qui autorisent toutes les opérations usuelles sur les variables aléatoires.

Proposition 2.0.5 Une fonction de Ω dans \mathbb{R}^d est une v.a. ssi ses coordonnées (dans n’importe quelle base de \mathbb{R}^d) sont des v.a.r.. Une combinaison linéaire de v.a. est encore une v.a.. Un produit de v.a.r. est une v.a.r.. La composée d’une v.a. par une fonction continue de \mathbb{R}^d dans \mathbb{R}^d est encore une v.a..

Exemples : La somme et le produit des chiffres indiqués par 2 dés ; les durées de vie de particules fissiles ; le nombre de fois qu’une suite de N lancers d’une pièce de monnaie donne pile ; les temps qu’il faut attendre pour tirer dans une urne, lors de tirages successifs, une boule rouge, une boule verte ; ou bien lors de lancers illimités d’une pièce, pour obtenir une séquence donnée ; etc...

La notion d’*espérance* est fondamentale dans la théorie des probabilités. Il s’agit en fait simplement d’une somme ou d’une série ou d’une intégrale, qui est aussi une moyenne.

Définition 2.0.6 L’*espérance* d’une variable aléatoire V définie sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ est son intégrale par rapport à la probabilité \mathbb{P} :

$$\mathbb{E}(V) := \int_{\Omega} V d\mathbb{P}.$$

Elle existe lorsque (la norme de) V est intégrable, c'est-à-dire lorsque $\mathbb{E}(\|V\|) < \infty$.

Exemple Lorsque $\Omega = \{\omega_1, \dots, \omega_N\}$ est fini, \mathbb{P} étant donnée par la liste des probabilités des singletons : $p_j := \mathbb{P}(\{\omega_j\})$, on a $\mathbb{E}(V) = \sum_{j=1}^N p_j V(\omega_j)$: c'est une moyenne pondérée.

Lorsque $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$ est dénombrable, on a $\mathbb{E}(V) = \sum_{j \geq 1} p_j V(\omega_j)$.

Corollaire 2.0.7 Soient V une v.a. et f une fonction continue bornée de \mathbb{R}^d dans $\mathbb{R}^{d'}$. Alors

$$\mathbb{E}(f \circ V) = \int_{\Omega} f \circ V d\mathbb{P} = \int_{\mathbb{R}^d} f d\mathbb{P}_V.$$

Lorsque la loi de V est discrète, c'est-à-dire lorsque presque sûrement V ne prend qu'un ensemble fini ou dénombrable de valeurs, alors on a

$$\mathbb{E}(f \circ V) = \sum_v f(v) \mathbb{P}(V = v),$$

cette série convergeant dès que $f \circ V$ est intégrable (ce qui a lieu par exemple dès que V ne prend qu'un nombre fini de valeurs, ou encore, dès que f est bornée).

Justification Pour toute v.a. discrète $V = \sum_j v_j 1_{A_j}$ (avec les v_j distincts 2 à 2), on a :

$$\int_{\Omega} f \circ V d\mathbb{P} = \sum_j f(v_j) \int 1_{A_j} d\mathbb{P} = \sum_j f(v_j) \mathbb{P}(A_j) = \sum_j f(v_j) \mathbb{P}_V(\{v_j\}) = \int_{\mathbb{R}^d} f d\mathbb{P}_V.$$

Exercice n° 2.0 Lors du lancer de 2 dés usuels, calculer :

- (i) l'espérance du résultat du 2ème dé, puis du carré de ce résultat ;
- (ii) l'espérance de la somme des 2 dés.

Proposition 2.0.8 L'espérance est linéaire : $\mathbb{E}\left(\sum_{j=1}^N \lambda_j V_j\right) = \sum_{j=1}^N \lambda_j \mathbb{E}(V_j)$, pour toutes v.a. intégrables V_1, \dots, V_N et tous réels $\lambda_1, \dots, \lambda_N$ (c'est une propriété générale, fondamentale, de toute intégrale). En particulier, l'espérance d'une v.a. vectorielle intégrable, dans n'importe quelle base (déterministe), a pour coordonnées les espérances des coordonnées.

Exercice n° 2.1 On tire au hasard et sans remise toutes les boules d'une urne remplie de r boules rouges et n boules noires.

1) Montrer que que la probabilité de tirer une boule rouge au k -ième tirage ne dépend pas de k . (Il y a au moins 2 solutions bien différentes : soit par récurrence, soit en considérant directement l'ensemble de tous les tirages exhaustifs possibles.)

2) Quelle est l'espérance du rang de la k -ième boule rouge ? (Considérer le vecteur $(X_1, X_2 - X_1, \dots, X_r - X_{r-1}, r + n + 1 - X_r)$, X_j désignant le rang de la j -ième boule rouge).

Exercice n° 2.2 Une urne U contient initialement 2 boules blanches, et une urne U' 2 boules noires. A chaque instant, on tire au hasard une boule par urne, et on les interchange. Notons X_k le nombre de boules blanches dans U au k -ième instant, et V_k le vecteur-colonne donnant la loi de X_k .

a) Quelle est la relation entre V_{k+1} et V_k ? b) Calculer $\lim_{k \rightarrow \infty} \mathbb{P}(X_k = 2)$.

Soient T le premier instant où U contient deux boules noires, $p_k := \mathbb{P}(T \geq k, X_k = 1)$, et $q_k := \mathbb{P}(T \geq k, X_k = 2)$.

c) Exprimer (p_{k+1}, q_{k+1}) en fonction de (p_k, q_k) , puis p_{k+1} en fonction de (p_k, p_{k-1}) .

d) Dédurre la valeur de p_k , puis la loi de T . Que vaut $\mathbb{P}(T = \infty)$?

Exercice n° 2.3 Un marchand de journaux a X clients par jour, X étant une v.a. entière intégrable, de loi supposée connue. Il gagne a par journal vendu, perd b par journal invendu, et perd c par client insatisfait. Quel est le nombre n de journaux qu'il doit commander par jour pour optimiser son gain moyen ?

Exercice n° 2.4 Montrer que pour toute v.a.r. $Z \geq 0$, on a $\mathbb{E}(Z) = \int_0^\infty \mathbb{P}(Z > s) ds$. Particulariser au cas où Z prend ses valeurs dans \mathbb{N} .

Définition 2.0.9 Une variable aléatoire V définie sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$ (et à valeurs dans \mathbb{R}^d) admet la densité h lorsque sa loi est donnée par

$$\mathbb{E}(f \circ V) := \int_{\mathbb{R}^d} f(v) h(v) dv, \quad \text{pour toute fonction continue bornée } f : \mathbb{R}^d \rightarrow \mathbb{R}.$$

La densité h est forcément une fonction intégrable de \mathbb{R}^d dans \mathbb{R}_+ , d'intégrale égale à 1.

Attention, la plupart des variables aléatoires ne sont ni discrètes ni à densité ! Il suffit de songer à une v.a. V valant une v.a. discrète X avec probabilité $1/2$ et une v.a. à densité Y avec probabilité $1/2$: $\mathbb{P}_V = (\mathbb{P}_X + \mathbb{P}_Y)/2$. Imaginer par exemple un tir sur cible, avec probabilité $1/2$ de rater la cible, auquel cas la v.a. prend la valeur disons -1 , et probabilité $1/2$ d'atteindre la cible, auquel cas la v.a. prend ses valeurs dans le disque formé par la cible, avec par exemple la loi uniforme.

Cela dit, les lois usuelles, celles qu'on rencontre le plus souvent, sont discrètes ou à densité (on dit aussi : absolument continues).

Définition 2.0.10 Une v.a. V (définie sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$) est dite de carré intégrable ou dans L^2 lorsque $\mathbb{E}(\|V\|^2) < \infty$. Elle est alors nécessairement intégrable.

La variance d'une v.a.r. V de carré intégrable est :

$$\text{Var}(V) := \mathbb{E}[|V - \mathbb{E}(V)|^2] = \mathbb{E}(V^2) - \mathbb{E}(V)^2.$$

La covariance de deux v.a.r. V, V' de carré intégrable est :

$$\text{Cov}(V, V') := \mathbb{E}[(V - \mathbb{E}(V)) \times (V' - \mathbb{E}(V'))] = \mathbb{E}(VV') - \mathbb{E}(V)\mathbb{E}(V').$$

La matrice de covariance (ou de dispersion) d'une v.a. $V = (V_1, \dots, V_d) \in \mathbb{R}^d$ de carré intégrable est :

$$K_V := \mathbb{E}[(V - \mathbb{E}(V)) \times (V - \mathbb{E}(V))] = \mathbb{E}({}^t V V) - \mathbb{E}({}^t V)\mathbb{E}(V) = \left((\text{Cov}(V_i, V_j)) \right)_{1 \leq i, j \leq d}.$$

Proposition 2.0.11 *La covariance est bilinéaire, symétrique, positive, et $\text{Var}(V)$ est nulle ssi V est p.s. constante. En outre $\text{Var}(\lambda V + V') = \lambda^2 \text{Var}(V) + 2\lambda \text{Cov}(V, V') + \text{Var}(V')$.*

Preuve : La symétrie : $\text{Cov}(V, V') = \text{Cov}(V', V)$ est évidente. La linéarité à gauche : $\text{Cov}(a_1 V_1 + a_2 V_2, V') = a_1 \text{Cov}(V_1, V') + a_2 \text{Cov}(V_2, V')$ découle immédiatement de la linéarité de l'espérance. Par symétrie elle entraîne la linéarité à droite, et donc la bilinéarité. La positivité : $\text{Cov}(V, V) = \text{Var}(V) \geq 0$ est claire. Ensuite, $\text{Var}(\text{constante}) = 0$ est clair, et réciproquement, si $\mathbb{P}[V \neq \mathbb{E}(V)] > 0$, alors il existe $\varepsilon > 0$ tel que $\mathbb{P}(|V - \mathbb{E}(V)| > \varepsilon) > \varepsilon$, de sorte qu'alors $\text{Var}(V) = \mathbb{E}(|V - \mathbb{E}(V)|^2) > \varepsilon \times \varepsilon^2 > 0$. Enfin, la dernière formule découle très simplement du développement du carré d'une somme et de la linéarité de l'espérance (détails à rédiger en exercice). \diamond

Constatant que le trinôme en λ de cette proposition doit toujours être positif ou nul, et donc doit avoir son discriminant négatif ou nul, on obtient immédiatement l'inégalité de Schwarz (dite aussi de Cauchy-Schwarz) :

Corollaire 2.0.12 (*Inégalité de Schwarz*) *La covariance est majorée par le produit des écart-types : $|\text{Cov}(V, V')| \leq \sigma(V)\sigma(V')$, où l'écart-type de la v.a.r. V est $\sigma(V) := \sqrt{\text{Var}(V)}$. De sorte que le coefficient de corrélation linéaire $\rho(V, V') := \frac{\text{Cov}(V, V')}{\sigma(V)\sigma(V')}$*

(défini pour V et V' v.a.r. de carré intégrable et non p.s. constantes) appartient à $[-1, 1]$.

Il vaut ± 1 ssi $V = aV' + b$ (presque sûrement, pour a, b réels fixes).

Exercice n° 2.6 a) Vérifier que $K_V = \mathbb{E}({}^t V V) - {}^t \mathbb{E}(V)\mathbb{E}(V) = \left((\text{Cov}(V_i, V_j)) \right)_{1 \leq i, j \leq d}$.

b) Vérifier que pour tout $u \in \mathbb{R}^d$ on a $\text{Var}(u^t V) = u K_V u = \sum_{1 \leq i, j \leq d} u_i u_j \text{Cov}(V_i, V_j)$.

c) Montrer que K_V est une matrice symétrique, et positive : $v K_V v \geq 0$ ($\forall v \in \mathbb{R}^d$).

d) Montrer que K_V est inversible ssi il n'existe pas d'hyperplan de \mathbb{R}^d contenant p.s. V .

e) Soit M une matrice de $\mathcal{M}_{n,d}(\mathbb{R})$. Calculer $\mathbb{E}(MV)$ et K_{MV} . Même question pour AV , si A est une application affine de \mathbb{R}^d dans \mathbb{R}^n .

Exercice n° 2.7 Inégalité de Markov, ou de Bienaymé-Tchebitchev :

Vérifier que $\mathbb{P}[|V| \geq v] \leq \mathbb{E}[|V|^k]/v^k$, pour tous $v > 0$ et $k \geq 0$.

Exercice n° 2.8 Vous écrivez chaque jour avec probabilité 1 si vous n'avez pas écrit la veille, et avec probabilité 1/2 sinon. Montrez que vous écrivez ainsi en moyenne 243 lettres par an. (Considérer pour chaque jour la variable indicatrice de l'événement "écrire".)

L'inégalité suivante est une importante inégalité de convexité.

Proposition 2.0.13 (*Inégalité de Jensen*) Pour toute v.a.r. intégrable V et toute fonction convexe ψ de \mathbb{R} dans \mathbb{R} , telle que $\psi \circ V$ soit positive ou intégrable, on a

$$\mathbb{E}[\psi \circ V] \geq \psi(\mathbb{E}[V]).$$

Exemples: On a $|x_1 + \dots + x_n|^p \leq n^{p-1} (|x_1|^p + \dots + |x_n|^p)$ pour tous $n \in \mathbb{N}^*$, $p \in [1, \infty[$, et $x_1, \dots, x_n \in \mathbb{R}$. Et $(\mathbb{E}[|V|^p])^{1/p} \geq (\mathbb{E}[|V|^q])^{1/q}$ si $p \geq q > 0$.

3 Lois usuelles

3.1 Lois usuelles discrètes

Définition 3.1.1 La loi uniforme sur un espace de probabilité fini est celle qui attribue la même valeur à tous les singletons.

Corollaire 3.1.2 Si \mathbb{P} est uniforme sur Ω fini, alors $\mathbb{P}(A) = \text{Card}(A)/\text{Card}(\Omega)$ pour toute partie A de Ω .

Exercice n° 3.1.1 Calculer l'espérance et la variance d'une v.a. uniforme sur un intervalle de \mathbb{Z} .

Définition 3.1.3 La loi de Bernoulli de paramètre p , notée $\mathcal{B}(p)$, est sur l'espace $\{0, 1\}$ la loi (de probabilité) qui attribue la valeur p au singleton $\{1\}$. Ici $0 \leq p \leq 1$.

Définition 3.1.4 La loi binômiale de paramètres n et p , notée $\mathcal{B}(n, p)$, est sur l'espace $\{0, \dots, n\}$ la loi (de probabilité) qui attribue la valeur $C_n^k p^k (1-p)^{n-k}$ au singleton $\{k\}$. Ici $n \in \mathbb{N}^*$ et $0 \leq p \leq 1$.

Remarque 3.1.5 C'est la loi de la somme de n v.a. indépendantes de même loi $\mathcal{B}(p)$. La formule du binôme est exactement équivalente au fait que la somme de ces probabilités $C_n^k p^k (1-p)^{n-k}$ vaut bien 1.

Exercice n° 3.1.2 Montrer que l'espérance et la variance d'une v.a. de loi $\mathcal{B}(n, p)$ valent respectivement np et $np(1-p)$.

Définition 3.1.6 La loi hypergéométrique de paramètre N, n, p , notée $\mathcal{H}(N, n, p)$, est sur l'espace $\{0, \dots, n\}$ la loi (de probabilité) qui attribue la valeur $\frac{C_{Np}^k C_{N(1-p)}^{n-k}}{C_N^n}$ au singleton $\{k\}$. Ici $N, Np, n \in \mathbb{N}$, $n \leq N$, et $0 \leq p \leq 1$.

Remarque 3.1.7 C'est la loi du nombre d'éléments possédant un caractère donné K dans un échantillon de taille n , prélevé au hasard, uniformément parmi les parties de cardinal n , dans un ensemble de cardinal N , dont une proportion p possède le caractère K .

Cela s'applique aussi bien à des pièces défectueuses dans une production industrielle, qu'aux individus malades dans une population, etc. . .

Exercice n° 3.1.3 a) Montrer que l'espérance d'une v.a. de loi $\mathcal{H}(N, n, p)$ vaut np .
b) Calculer sa variance. c) Vérifier que $\lim_{N \rightarrow \infty} \mathcal{H}(N, n, p)(k) = \mathcal{B}(n, p)(k)$, pour n, p, k fixés.

Définition 3.1.8 La loi géométrique de paramètre p , notée $\mathcal{G}(p)$, est sur l'espace \mathbb{N}^* la loi (de probabilité) qui attribue la valeur $(1-p)^{n-1} p$ au singleton $\{n\}$. Ici $0 < p < 1$.

Remarque 3.1.9 C'est la loi du nombre N de tentatives nécessaires pour obtenir un résultat positif (succès), lors d'une suite de tentatives identiques et indépendantes, p étant la probabilité de succès à chaque tentative. On a $\mathbb{P}(N > n) = (1-p)^n$ pour tout $n \in \mathbb{N}$.

Exercice n° 3.1.4 a) Les lois géométriques vérifient la propriété de non-vieillessement :
$$\mathbb{P}(N > n + m | N > m) = \mathbb{P}(N > n)$$
 pour tous $n, m \in \mathbb{N}$.

b) Y a-t-il d'autres lois sur \mathbb{N}^* qui vérifient cette propriété ?

Exercice n° 3.1.5 Montrer que l'esp. et la variance d'une v.a. de loi $\mathcal{G}(p)$ valent $\frac{1}{p}$ et $\frac{1-p}{p^2}$.

Exercice n° 3.1.6 Au cours d'un jeu illimité de pile ou face avec $\mathbb{P}(\text{pile}) = p$, on note X_k le rang de la k -ième apparition de "pile". Calculer la loi de X_k , son espérance et sa variance. Pour $n \in \mathbb{N}^*$, calculer $\mathbb{P}((\exists k \in \mathbb{N}^*) X_k = n)$.

Définition 3.1.10 La loi de Poisson de paramètre λ , notée $\mathcal{P}(\lambda)$, est sur l'espace \mathbb{N} la loi (probabilité) qui attribue la valeur $e^{-\lambda} \frac{\lambda^n}{n!}$ au singleton $\{n\}$. Ici $\lambda > 0$.

Exercice n° 3.1.7 a) Montrer que l'esp. et la variance d'une v.a. de loi $\mathcal{P}(\lambda)$ valent λ .

b) Vérifier que $\lim_{np \rightarrow \lambda} \mathcal{B}(n, p)(\{k\}) = \mathcal{P}(\lambda)(\{k\})$, pour $\lambda > 0$, $k \in \mathbb{N}$ fixés et $n \rightarrow \infty$.

Exercice n° 3.1.8 Quelle est la valeur la plus probable pour une variable aléatoire poissonnienne de paramètre λ ?

Exercice n° 3.1.9 Un trousseau de n clefs contient une seule clef ouvrant une serrure donnée. On les essaie l'une après l'autre au hasard. Calculer la loi, l'espérance et la variance du nombre d'essais nécessaires. Même question si on réessaie à chaque fois une clef au hasard sans avoir écarté la précédente.

3.2 Lois usuelles à densité

Définition 3.2.1 La loi uniforme sur un ouvert (ou un fermé) O de \mathbb{R}^d , notée $\mathcal{U}(O)$, est la loi admettant une densité constante (par rapport à la mesure de Lebesgue) sur O .

Exercice n° 3.2.1 Calculer l'esp. et la variance d'une v.a. uniforme sur un intervalle de \mathbb{R} .

Exercice n° 3.2.2 (Aiguille de Buffon) Sur un sol plat sont tracées 2 droites parallèles D et D' , distantes de $L \geq 1$. On laisse tomber entre D et D' une aiguille de longueur 1, puis on note x la distance du centre de l'aiguille à D , et θ l'angle que fait l'aiguille avec D . On suppose que la v.a. (x, θ) est uniforme sur $[0, L] \times [0, \pi/2]$. Vérifier que la probabilité que l'aiguille intersecte D ou D' vaut $2/(\pi L)$. Qu'en est-il si $0 < L < 1$?

Définition 3.2.2 La loi exponentielle de paramètre λ , notée $\mathcal{E}(\lambda)$, est la loi admettant sur \mathbb{R}_+ la densité $t \mapsto \lambda e^{-\lambda t}$. Ici $\lambda > 0$.

Exercice n° 3.2.3 a) Les lois exponentielles vérifient la propriété de non-vieillessement : si la loi de Y est $\mathcal{E}(\lambda)$, alors $\mathbb{P}(Y > s + t | Y > s) = \mathbb{P}(Y > t)$ pour tous $s, t > 0$.

b) Y a-t-il d'autres lois à densité sur \mathbb{R}_+ qui vérifient cette propriété ?

c) En déduire la loi de la durée de vie d'un atome fissile, en fonction de sa demi-vie.

Exercice n° 3.2.4 Montrer que l'espérance et l'écart-type d'une v.a. de loi $\mathcal{E}(\lambda)$ valent $\frac{1}{\lambda}$.

Définition 3.2.3 Une variable aléatoire réelle est dite gaussienne centrée réduite ou normale centrée réduite ou gaussienne standard lorsqu'elle admet (sur \mathbb{R}) la densité :

$t \mapsto \exp(-t^2/2)/\sqrt{2\pi}$. Une variable aléatoire réelle X est dite gaussienne ou normale lorsqu'il existe $m \in \mathbb{R}$ et $\sigma > 0$ tels que $(X - m)/\sigma$ soit normale centrée réduite. On dit alors que la loi de X est $\mathcal{N}(m, \sigma^2)$.

Exercice n° 3.2.5 a) Vérifier que $m = \mathbb{E}(X)$ et $\sigma^2 = \text{Var}(X)$, et que $X \in \bigcap_{p < \infty} L^p$.

b) Vérifier que $\frac{A^2 e^{-x^2/2}}{(A^2+1)x} \leq \int_x^\infty e^{-t^2/2} dt \leq \frac{e^{-x^2/2}}{x}$ sur $[A, \infty[$, pour tout $A > 0$. Déduire un équivalent de $\mathbb{P}(X > x)$ quand $x \rightarrow +\infty$.

c) Montrer que la densité de X est $t \mapsto \exp[-(t - m)^2/2\sigma^2]/\sigma\sqrt{2\pi}$.

Nota Bene (i) La courbe de la densité gaussienne est la célèbre "courbe en cloche".

(ii) Valeurs numériques à connaître :

$$\mathcal{N}(0, 1)(] - \infty, 1,65]) \approx 0,95, \quad \text{et} \quad \mathcal{N}(0, 1)(] - \infty, 1,96]) \approx 0,975.$$

Exercice n° 3.2.6 Pour être en cours à 8h, un étudiant en voiture a le choix entre un trajet sur petite route, dont la durée (en minutes) X suit la loi normale de moyenne 35,2 et de variance 25, et un trajet sur autoroute, dont la durée Y suit la loi normale de moyenne 40 et de variance 4. Il désire arriver à l'heure. Quel trajet doit-il préférer s'il part à 7h15 ? Et s'il part à 7h30 ?

Définition 3.2.4 Une variable aléatoire V à valeurs dans \mathbb{R}^d est dite gaussienne ou normale lorsque $u \cdot V = {}^t u V$ est une v.a.r. normale ou p.s. constante pour tout $u \in \mathbb{R}^d$.

On note $\mathcal{N}(m, K)$ la loi d'un vecteur gaussien d'espérance m et de matrice de covariance K . Une probabilité (resp. une densité) est dite gaussienne lorsqu'elle est la loi (resp. la densité) d'un vecteur gaussien.

Exercice n° 3.2.7 a) Vérifier que si V est un vecteur gaussien de \mathbb{R}^d et si A est une application affine de \mathbb{R}^d dans \mathbb{R}^n , alors AV est un vecteur gaussien.

b) Montrer que si V est un vecteur gaussien, alors ses coordonnées dans n'importe quelle base sont gaussiennes (ou p.s. constantes).

Proposition 3.2.5 Soit K une matrice réelle symétrique positive, de format $d \times d$ et de rang r , et soit M une matrice réelle de format $d \times r$ telle que $K = M \times {}^t M$. Soit $m \in \mathbb{R}^d$. Soit V un vecteur de \mathbb{R}^r formé de coordonnées gaussiennes standard indépendantes. Alors $m + MV$ est un vecteur gaussien de loi $\mathcal{N}(m, K)$.

Corollaire 3.2.6 Si la matrice de covariance d'un vecteur gaussien d -dimensionnel de loi $\mathcal{N}(m, K)$ est inversible, alors ce vecteur admet la densité sur \mathbb{R}^d :

$$v \longmapsto (2\pi)^{-d/2} (\det K)^{-1/2} \exp\left(-\frac{1}{2} {}^t (v - m) K^{-1} (v - m)\right).$$

Preuve Soit V un vecteur de \mathbb{R}^d formé de coordonnées gaussiennes standard indépendantes. Soit M une matrice réelle de format $d \times d$ telle que $K = M \times {}^t M$. Alors M est inversible, et nous avons pour toute fonction-test f , par changement de variable :

$$\begin{aligned} \mathbb{E}(f(m + MV)) &= \int_{\mathbb{R}^d} f(m + Mx) \prod_{j=1}^d (2\pi)^{-1/2} e^{-x_j^2/2} dx = \int_{\mathbb{R}^d} f(m + Mx) (2\pi)^{-d/2} e^{-{}^t x x / 2} dx \\ &= \int_{\mathbb{R}^d} f(v) (2\pi)^{-d/2} e^{-{}^t (M^{-1}(v-m)) \times (M^{-1}(v-m)) / 2} |\det(M^{-1})| dv \\ &= \int_{\mathbb{R}^d} f(v) (2\pi)^{-d/2} (\det K)^{-1/2} \exp\left(-\frac{1}{2} {}^t (v - m) K^{-1} (v - m)\right) dv . \diamond \end{aligned}$$

Exercice n° 3.2.8: Simulation Soit F une fonction de répartition sur \mathbb{R} . Pour tout $p \in [0, 1]$, posons $G(p) := \inf\{x \in \mathbb{R} \mid F(x) \geq p\}$. (Nota Bene: $\inf \mathbb{R} = -\infty$ et $\inf \emptyset = +\infty$.)

a) Justifier l'existence dans \mathbb{R} de $G(p)$ si $p \in]0, 1[$. b) Montrer que $(\forall x \in \mathbb{R}) G(F(x)) \leq x$.
c) Montrer que si $G(p) \in \mathbb{R}$, alors $F(G(p)) \geq p$. d) Montrer que $G(p) \leq x \Leftrightarrow F(x) \geq p$.
e) Montrer que si U est une variable aléatoire de loi uniforme sur $[0, 1]$, alors $G \circ U$ admet F pour fonction de répartition. Nota Bene: Ceci est utilisé pour simuler des variables aléatoires.

f) Que vaut G lorsque F est bijective de \mathbb{R} sur $]0, 1[$? Comment simuler la loi $\mathcal{E}(\lambda)$?

3.3 Quelques lois de la mécanique statistique

La mécanique statistique classique s'appuie sur le modèle statistique de Maxwell, dans lequel les N particules considérées sont réparties parmi n états, dont n_j ont un niveau d'énergie e_j , de sorte que $n = n_1 + \dots + n_d$. Soit N_j le nombre de particules ayant le niveau d'énergie e_j , de sorte que $N = N_1 + \dots + N_d$. Quelle est la loi du vecteur (N_1, \dots, N_d) ?

La réponse est la loi multinômiale (généralisant la loi binômiale) :

$$\mathbb{P}\left((N_1, \dots, N_d) = (k_1, \dots, k_d)\right) = \frac{N!}{k_1! \times \dots \times k_d!} \times \frac{n_1^{k_1} \times \dots \times n_d^{k_d}}{n^N},$$

pour tout $(k_1, \dots, k_d) \in \mathbb{N}^d$ tel que $k_1 + \dots + k_d = N$. En effet, le coefficient multinômial $\frac{N!}{k_1! \times \dots \times k_d!}$ (c'est un coefficient binômial lorsque $d = 2$) est le nombre de partitions de l'ensemble des N particules en d classes de cardinaux respectifs k_1, \dots, k_d ; et d'autre part $n_1^{k_1} \times \dots \times n_d^{k_d}$ est le nombre de façons de répartir k_1 particules d'énergie e_1 parmi n_1 états, \dots , k_d particules d'énergie e_d parmi n_d états ; tandis que n^N est le nombre total de répartitions des N particules parmi les n états.

Ici \mathbb{P} est uniforme sur l'ensemble Ω des fonctions de $\{1, \dots, N\}$ dans $\{1, \dots, n\}$.

Notons que cela montre du même coup la généralisation de la formule du binôme :

$$\sum_{\{(k_1, \dots, k_d) \in \mathbb{N}^d \mid k_1 + \dots + k_d = N\}} \frac{N!}{k_1! \dots k_d!} \times z_1^{k_1} \dots z_d^{k_d} = (z_1 + \dots + z_d)^N, \text{ pour tous } z_1, \dots, z_d \in \mathbb{C},$$

et en particulier :

$$\sum_{\{(k_1, \dots, k_d) \in \mathbb{N}^d \mid k_1 + \dots + k_d = N\}} \frac{N!}{k_1! \dots k_d!} = d^N \text{ pour tous } N, d \in \mathbb{N}^*.$$

Ce modèle de Maxwell n'est en fait pas vraiment réaliste, du fait qu'il suppose les particules distinguables, ce qui n'est pas le cas en mécanique quantique. Supposant les particules indistinguables, on a le choix entre deux options : on peut les répartir entre les différents états soit sans répétition, soit avec répétitions. L'option sans répétition est la statistique de Fermi-Dirac, relative aux particules de spin demi-entier (électrons, neutrons, protons,..), qui obéissent au principe d'exclusion de Pauli, et l'option avec répétitions est la statistique de Bose-Einstein, relative aux particules de spin entier (photons, mésons,..).

Pour la statistique de Fermi-Dirac, la loi généralise la loi hypergéométrique :

$$\mathbb{P}\left((N_1, \dots, N_d) = (k_1, \dots, k_d)\right) = \frac{C_{n_1}^{k_1} \times \dots \times C_{n_d}^{k_d}}{C_n^N},$$

pour tout $(k_1, \dots, k_d) \in \mathbb{N}^d$ tel que $k_1 + \dots + k_d = N$. En effet, il faut simplement répartir k_1 particules parmi n_1 états d'énergie e_1 , \dots , k_d particules parmi n_d états d'énergie e_d , cependant qu'il y a en tout N particules à répartir parmi n états d'énergie ; avec à chaque fois une seule particule au plus par état, ce qui revient exactement à choisir le sous-ensemble des états occupés parmi les états disponibles, d'où les coefficients binômiaux.

Ici \mathbb{P} est uniforme sur l'ensemble Ω' des parties de cardinal N dans $\{1, \dots, n\}$.

Pour la statistique de Bose-Einstein, la loi est

$$\mathbb{P}\left((N_1, \dots, N_d) = (k_1, \dots, k_d)\right) = \frac{C_{n_1+k_1-1}^{k_1} \times \dots \times C_{n_d+k_d-1}^{k_d}}{C_{n+N-1}^N},$$

pour tout $(k_1, \dots, k_d) \in \mathbb{N}^d$ tel que $k_1 + \dots + k_d = N$. Le décompte est en effet le même que pour la statistique de Fermi-Dirac, les combinaisons devant être remplacées par des combinaisons avec répétitions. Or C_{n+N-1}^N est le nombre des combinaisons avec répétitions de N particules placées dans n états ; ce qui se voit en juxtaposant sur un segment les N particules, et en considérant la bijection qui à chaque ensemble de $(n-1)$ signes délimitant n sous-segments (figurant les états) associe la combinaison avec répétitions ainsi constituée.

Ici \mathbb{P} est uniforme sur l'ensemble Ω'' des combinaisons avec répétition de N éléments pris (ou placés) dans $\{1, \dots, n\}$.

Remarquons une structure de produit commune à ces trois “statistiques” : elles sont toutes trois de la forme $\mathbb{P}\left((N_1, \dots, N_d) = (k_1, \dots, k_d)\right) = c(N, n) \prod_{j=1}^d \lambda(n_j, k_j)$.

Exercice n° 3.3.1 : Montrer que lorsque le nombre n des états tend vers l'infini de façon que les proportions $n_1/n, \dots, n_d/n$ convergent vers p_1, \dots, p_d , alors les statistiques de Fermi-Dirac et de Bose-Einstein convergent vers la loi multinômiale (analogue à la statistique de Maxwell).

Table de la loi normale centrée réduite: $[10^5 \Phi(x)]$

x	,.0	,.1	,.2	,.3	,.4	,.5	,.6	,.7	,.8	,.9
0,0	50000	50399	50798	51197	51595	51994	52392	52790	53188	53586
0,1	53983	54380	54776	55172	55567	55962	56356	56749	57142	57535
0,2	57926	58317	58706	59095	59483	59871	60257	60642	61026	61409
0,3	61791	62172	62552	62930	63307	63683	64058	64431	64803	65173
0,4	65542	65910	66276	66640	67003	67364	67724	68082	68439	68793
0,5	69146	69497	69847	70194	70540	70884	71226	71566	71904	72240
0,6	72575	72907	73237	73565	73891	74215	74537	74857	75175	75490
0,7	75804	76115	76424	76730	77035	77337	77637	77935	78230	78524
0,8	78814	79103	79389	79673	79955	80234	80511	80785	81057	81327
0,9	81594	81859	82121	82381	82639	82894	83147	83398	83646	83891
1,0	84134	84375	84614	84849	85083	85314	85543	85769	85993	86214
1,1	86433	86650	86864	87076	87286	87493	87698	87900	88100	88298
1,2	88493	88686	88877	89065	89251	89435	89617	89796	89973	90147
1,3	90320	90490	90658	90824	90988	91149	91309	91466	91621	91774
1,4	91924	92073	92220	92364	92507	92647	92785	92922	93056	93189
1,5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408
1,6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449
1,7	95543	95637	95728	95818	95907	95994	96080	96164	96246	96327
1,8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062
1,9	97128	97193	97257	97320	97381	97441	97500	97558	97615	97670
2,0	97725	97778	97831	97882	97932	97982	98030	98077	98124	98169
2,1	98214	98257	98300	98341	98382	98422	98461	98500	98537	98574
2,2	98610	98645	98679	98713	98745	98778	98809	98840	98870	98899
2,3	98928	98956	98983	99010	99036	99061	99086	99111	99134	99158
2,4	99180	99202	99224	99245	99266	99286	99305	99324	99343	99361
2,5	99379	99396	99413	99430	99446	99461	99477	99492	99506	99520
2,6	99534	99547	99560	99573	99585	99598	99609	99621	99632	99643
2,7	99653	99664	99674	99683	99693	99702	99711	99720	99728	99736
2,8	99744	99752	99760	99767	99774	99781	99788	99795	99801	99807
2,9	99813	99819	99825	99831	99836	99841	99846	99851	99856	99861
3,0	99865	99869	99874	99878	99882	99886	99889	99893	99896	99900
3,1	99903	99906	99910	99913	99916	99918	99921	99924	99926	99929
3,2	99931	99934	99936	99938	99940	99942	99944	99946	99948	99950
3,3	99952	99953	99955	99957	99958	99960	99961	99962	99964	99965
3,4	99966	99968	99969	99970	99971	99972	99973	99974	99975	99976
3,5	99977	99978	99978	99979	99980	99981	99981	99982	99983	99983
3,6	99984	99985	99985	99986	99986	99987	99987	99988	99988	99989
3,7	99989	99990	99990	99990	99991	99991	99992	99992	99992	99992
3,8	99993	99993	99993	99994	99994	99994	99994	99995	99995	99995
3,9	99995	99995	99996	99996	99996	99996	99996	99996	99997	99997
4,0	99997	99997	99997	99997	99997	99997	99998	99998	99998	99998
4,1	99998	99998	99998	99998	99998	99998	99998	99998	99999	99999

La fonction de répartition Φ de la loi normale centrée réduite est définie par

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = 1 - \Phi(-x) \quad (\forall x \in \mathbb{R}).$$

La table donne les valeurs de $10^5 \Phi(x)$ pour les valeurs positives de x , de 0 jusqu'à 4,19, avec un pas de 0,01. Ces valeurs sont arrondies à l'unité la plus proche. Cette table est à double entrée, l'entrée en ligne donnant les deux premiers chiffres de x et l'entrée en colonne le chiffre des centièmes. Exemples : $\Phi(0,71) \simeq 0,76115$; $\Phi(2,48) \simeq 0,99343$.

4 Lemmes de Borel-Cantelli

Ces fameux lemmes, quoique simples, illustrent bien l'hypothèse d'additivité dénombrable faite sur les probabilités. Ils sont très utiles, et ont des conséquences qui frappent l'imagination.

Pour toute suite d'ensembles $\{A_n \mid n \in \mathbb{N}\}$, $\limsup_n A_n := \bigcap_n \bigcup_{m \geq n} A_m$ est l'ensemble des éléments qui appartiennent à une infinité d'ensembles A_n . De même, $\liminf_n A_n := \bigcup_n \bigcap_{m \geq n} A_m$ est l'ensemble des éléments qui appartiennent à tous les ensembles A_n sauf au plus un nombre fini. On a aussitôt : $(\limsup_n A_n)^c = \liminf_n A_n^c$. Et $\limsup_n A_n$ et $\liminf_n A_n$ sont des événements si les A_n en sont.

Proposition 4.0.1 (*lemmes de Borel-Cantelli*) Soit $\{A_n \mid n \in \mathbb{N}\}$ une suite d'événements.

- 1) Si $\sum_n \mathbb{P}(A_n) < \infty$, alors $\mathbb{P}(\limsup_n A_n) = 0$.
- 2) Si $\sum_n \mathbb{P}(A_n) = \infty$ et si les A_n sont indépendants, alors $\mathbb{P}(\limsup_n A_n) = 1$.

Preuve 1) $\mathbb{P}\left(\bigcap_n \bigcup_{m \geq n} A_m\right) \leq \sum_{m \geq n} \mathbb{P}(A_m) \rightarrow 0$.

$$\begin{aligned} 2) \quad 1 - \mathbb{P}\left(\bigcap_n \bigcup_{m \geq n} A_m\right) &= \mathbb{P}\left(\bigcup_n \bigcap_{m \geq n} A_m^c\right) = \lim_n \uparrow \mathbb{P}\left(\bigcap_{m \geq n} A_m^c\right) = \lim_n \prod_{m \geq n} \mathbb{P}(A_m^c) \\ &= \lim_n \prod_{m \geq n} (1 - \mathbb{P}(A_m)) \leq \lim_n \exp\left(-\sum_{m \geq n} \mathbb{P}(A_m)\right) = 0. \quad \diamond \end{aligned}$$

Exemples 1) Un singe tapant sur un clavier d'ordinateur au hasard et indéfiniment tapera presque sûrement à un certain moment les œuvres complètes de Victor Hugo sans erreur (et même une infinité de fois). Mais ce résultat ne fournit aucun majorant du temps qu'il faudra attendre pour voir ceci se réaliser une première fois !

2) Si l'univers est infini, il y a une infinité de planètes habitées par des êtres vivants.

3) Une utilisation typique du premier lemme de Borel-Cantelli : records.

Considérons une v.a.r. $X_1 \geq 0$, pour laquelle

$$(\exists c > 0) (\exists \varrho \in]0, 1[) \quad \lim_{x \rightarrow +\infty} \mathbb{P}(X_1 > x) / \varrho^x = c.$$

C'est le cas des variables exponentielles et géométriques, par exemple. Considérons une suite i.i.d. de telles v.a.r. : $\{X_1, \dots, X_n, \dots\}$, et posons $Y_n := \max\{X_1, \dots, X_n\}$, pour tout $n \in \mathbb{N}^*$. Nous voulons évaluer la valeur du "record" Y_n lorsque n est grand.

a) Fixons $\varepsilon > 0$. Pour n suffisamment grand, nous avons :

$$\mathbb{P}\left[Y_n \leq (1-\varepsilon) \frac{\log n}{\log(1/\varrho)}\right] = \mathbb{P}\left[X_1 \leq (1-\varepsilon) \frac{\log n}{\log(1/\varrho)}\right]^n = \left(1 - \mathbb{P}\left[X_1 > (1-\varepsilon) \frac{\log n}{\log(1/\varrho)}\right]\right)^n$$

$$\leq \exp\left(-n \mathbb{P}\left[X_1 > (1-\varepsilon) \frac{\log n}{\log(1/\varrho)}\right]\right) \leq \exp\left(-n \frac{c}{2} \times \varrho^{(1-\varepsilon) \frac{\log n}{\log(1/\varrho)}}\right) = \exp\left(-\frac{c}{2} n^\varepsilon\right),$$

de sorte que $\sum_n \mathbb{P}\left[Y_n \leq (1-\varepsilon) \frac{\log n}{\log(1/\varrho)}\right] < \infty$, et donc que selon le premier lemme de Borel-Cantelli, nous avons presque sûrement : $\frac{Y_n}{\log n} > \frac{1-\varepsilon}{\log(1/\varrho)}$ pour tout $n > n_1(\omega)$.

b) Pour tout $\beta > 0$, nous avons :

$$\mathbb{P}\left[Y_{n^\beta} > (1+\varepsilon) \frac{\beta \log n}{\log(1/\varrho)}\right] \leq n^\beta \mathbb{P}\left[X_1 > (1+\varepsilon) \frac{\beta \log n}{\log(1/\varrho)}\right] \leq 2c n^\beta \times n^{-\beta(1+\varepsilon)} = 2c n^{-\beta\varepsilon},$$

de sorte qu'en fixant $\beta = 2/\varepsilon$, nous obtenons : $\sum_n \mathbb{P}\left[Y_{n^\beta} > (1+\varepsilon) \frac{\log n^\beta}{\log(1/\varrho)}\right] < \infty$. Donc selon le premier lemme de Borel-Cantelli, nous avons presque sûrement : $\frac{Y_{n^\beta}}{\log n^\beta} \leq \frac{1+\varepsilon}{\log(1/\varrho)}$ pour tout $n > n_2(\omega)$. Cela entraîne que pour tout $k > n_2(\omega)^\beta$, si $n := \lceil k^{1/\beta} \rceil$, alors $n^\beta \leq k < (n+1)^\beta$, et par croissance de la suite (Y_n) , nous avons presque sûrement, pour tout entier k suffisamment grand :

$$\frac{1-\varepsilon}{\log(1/\varrho)} < \frac{Y_k}{\log k} \leq \frac{Y_{(n+1)^\beta}}{\log n^\beta} \leq \frac{1+\varepsilon}{\log(1/\varrho)} \times \frac{\log(n+1)}{\log n} < \frac{1+2\varepsilon}{\log(1/\varrho)}.$$

Ceci démontre précisément que presque sûrement : $\lim_{n \rightarrow \infty} \frac{Y_n}{\log n} = \frac{1}{\log(1/\varrho)}$.

Exercice n° 4.1 Montrer que si une suite de v.a. $\{Y_n \mid n \in \mathbb{N}\}$ est telle que $\sum_n \|Y_n\|_2^2$ converge, alors cette suite converge presque sûrement vers 0.

Exercice n° 4.2 a) Montrer que si Y est une v.a.r. ≥ 0 , alors $\mathbb{E}(Y)^2 \leq \mathbb{E}(Y^2) \times \mathbb{P}(Y > 0)$.

b) Dédurre que si $\{A_1, \dots, A_n\}$ sont des événements non tous négligeables, alors

$$\mathbb{P}\left(\bigcup_{j=1}^n A_j\right) \geq \left(\sum_{j=1}^n \mathbb{P}(A_j)\right)^2 / \left[\sum_{j=1}^n \mathbb{P}(A_j) + 2 \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j)\right].$$

c) Dédurre que pour le deuxième lemme de Borel-Cantelli il suffit de supposer l'indépendance des événements 2 à 2 (en plus de la divergence de la série).

d) Montrer que s'il existe un $c > 0$ tel que $\mathbb{P}(A_i \cap A_j) \leq c \mathbb{P}(A_i) \mathbb{P}(A_j)$ pour tous $i \neq j$ et si $\sum_j \mathbb{P}(A_j) = \infty$, alors $\mathbb{P}(\limsup_j A_j) > 0$.

e) Considérons un jeu illimité de pile ou face, avec $p := \mathbb{P}(\text{"pile"}) \in]0, 1[$.

Notons X_j la variable de Bernoulli valant 1 ssi le j -ième lancer donne pile, et posons $A_j^k := \{X_j = \dots = X_{j+k-1} = 1\}$, pour $j, k \in \mathbb{N}^*$. Montrer que $\mathbb{P}\left(\bigcap_k \bigcup_j A_j^k\right) = 1$. Interpréter.

5 Variables aléatoires indépendantes

Définition 5.0.2 Fixons un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$. Des v.a. (définies sur (Ω, \mathcal{T})) V_1, \dots, V_n, \dots sont dites indépendantes lorsque pour tous B_1, \dots, B_n, \dots , les événements $\{V_1 \in B_1\}, \dots, \{V_n \in B_n\}, \dots$ sont indépendants (définition 1.3.2, section 1.3).

Proposition 5.0.3 Les v.a. V_1, \dots, V_n, \dots sont indépendantes ssi

$$\mathbb{E}(f_1 \circ V_1 \times \dots \times f_n \circ V_n) = \mathbb{E}(f_1 \circ V_1) \times \dots \times \mathbb{E}(f_n \circ V_n)$$

pour tous $n \in \mathbb{N}^*$ et toutes fonctions (mesurables) positives f_1, \dots, f_n .

Cela se dit aussi sous la forme : la loi de (V_1, \dots, V_n, \dots) est le produit des lois des V_n .

En particulier, lorsque les v.a.r. indépendantes V_n admettent des densités h_n sur \mathbb{R} , alors la v.a. (V_1, \dots, V_d) admet la densité $h_1 \otimes \dots \otimes h_d$ sur \mathbb{R}^d

$$\left(\text{définie par : } h_1 \otimes \dots \otimes h_d(x_1, \dots, x_d) := h_1(x_1) \times \dots \times h_d(x_d). \right)$$

Remarque 5.0.4 1) Les événements A_1, \dots, A_n, \dots sont indépendants ssi les v.a.

$1_{A_1}, \dots, 1_{A_n}, \dots$ sont indépendantes (revoir la proposition 1.3.3, section 1.3).

2) Si des v.a.r. $V_1, \dots, V_n, \dots \in L^2$ sont indépendantes, alors leurs covariances sont nulles. Mais la réciproque est fautive. Contrexemples : a) Si $G \sim \mathcal{N}(0, 1)$ et $\varepsilon \sim \mathcal{B}(\pm 1, \frac{1}{2})$ sont indépendantes, alors $\text{Cov}(G, \varepsilon G) = 0$, mais $\mathbb{P}(G + \varepsilon G = 0) = \frac{1}{2}$ contredit l'éventuelle indépendance de G et de εG . b) Cas de V_1, \dots, V_n, \dots indépendantes 2 à 2, mais non indépendantes. c) Cas de V_1, \dots, V_n, \dots indépendantes mais de carré non intégrable.

3) Les v.a. discrètes V_1, \dots, V_n, \dots sont indépendantes ssi pour tous v_1, \dots, v_n, \dots les événements $\{V_1 = v_1\}, \dots, \{V_n = v_n\}, \dots$ sont indépendants.

Exemples : les résultats de différents dés, de lancers successifs d'une pièce, de tirages successifs (roulette, ...), les durées de vie de différents atomes fissiles (en l'absence de toute réaction en chaîne), etc...

Exercice n° 5.1 a) Soient U_1, U_2, U_3 trois v.a. indépendantes, uniformes sur $[0, 1]$. Montrer qu'elles sont p.s. 2 à 2 distinctes. On les réordonne en $V_1 < V_2 < V_3$. Montrer que (V_1, V_2, V_3) admet la densité $6 \times 1_{\{0 < v_1 < v_2 < v_3 < 1\}}$, et en déduire les densités de V_1, V_2, V_3 .

b) Soient U_1, \dots, U_n des v.a. indépendantes, uniformes sur $[0, 1]$, et $J_n := \min\{U_1, \dots, U_n\}$, $M_n := \max\{U_1, \dots, U_n\}$. Calculer $\mathbb{P}(x < J_n, M_n < y)$, et en déduire la loi (conjointe) de (J_n, M_n) . c) Que vaut $\lim_{n \rightarrow \infty} \mathbb{P}(J_n < x < y < M_n)$?

Exercice n° 5.2 Soient Y_1, \dots, Y_n des v.a. indépendantes, exponentielles. Quelle est la loi de $\min\{Y_1, \dots, Y_n\}$?

Exercice n° 5.3 Soient X_1, \dots, X_n n v.a. indépendantes, X_j étant poissonnienne de paramètre λ_j . a) Quelle est la loi de $X_1 + X_2$, puis de $X_1 + \dots + X_n$?

- b) Calculer la loi, l'espérance et la variance de X_j sachant $X_1 + \dots + X_n$.
- c) Soient $Y_j, j \in \mathbb{N}$, des v.a. indépendantes à valeurs dans $\{1, \dots, n\}$, de même loi donnée par $\mathbb{P}(Y_j = k) = \alpha_k$, pour $1 \leq k \leq n$. Soient N une variable de Poisson de paramètre λ , indépendante des Y_j , et $N_k := \sum_{j=1}^N 1_{\{Y_j=k\}}$. Montrer que les variables N_1, \dots, N_k sont poissonniennes indépendantes.

Exercice n° 5.4 Soient X_1, X_2 deux v.a.r. indépendantes de lois exponentielles de paramètres λ_1, λ_2 . a) Calculer les lois de $J := \min\{X_1, X_2\}$, $M := \max\{X_1, X_2\}$.

b) Supposant que $\lambda_1 = \lambda_2$, montrer que J et $M - J$ sont des variables indépendantes.

Exercice n° 5.5 Soient X, Y deux variables aléatoires indépendantes, de densité respectivement f, g , avec $n \in \mathbb{N}^*$, $f(x) = 1_{\{x>0\}} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}$, et $g(y) = 1_{\{y>0\}} \lambda e^{-\lambda y}$.

Calculer la densité de $X + Y$, $\mathbb{E}(X + Y)$, et $\text{Var}(X + Y)$.

Exercice n° 5.6 Calculer la densité du carré d'une v.a.r. normale standard, puis de la somme de 2 tels carrés, puis du quotient de deux v.a.r. normales standards indépendantes.

Exercice n° 5.7 Soient X et Y 2 v.a.r. indépendantes, admettant toutes 2 la densité $t \mapsto t^{-2} 1_{[1, \infty[}(t)$. On pose $U := XY$ et $V := X/Y$.

a) Calculer les lois de (U, V) , de U , et de V . U et V sont-elles indépendantes ?

b) Calculer $\mathbb{E}(U^{-1/2} V^{-1})$.

Exercice n° 5.8 Trois clients A, B, C arrivent au même temps 0 à la poste, où 2 guichets sont ouverts, qu'occupent A et B tout de suite. C remplace le premier des 2 qui a terminé. On admet que les temps de service X, Y, Z requis par ces 3 clients sont des v.a.r.i.i.d. de même loi $\mathcal{E}(\lambda)$.

a) Quelle est la loi du temps d'attente T de C ? b) Calculer la probabilité que C termine (et parte) le dernier. c) Calculer la loi du temps du dernier départ.

Exercice n° 5.9 Justifier la remarque 3.1.5 (somme de v.a.i.i.d. de Bernoulli). Quand la somme de 2 variables aléatoires binômiales indépendantes est-elle binômiale ?

Exercice n° 5.10 On effectue n tirages indépendants avec remise dans une urne contenant une proportion p_j de boules marquées j , pour $1 \leq j \leq r$, $r > 1$ étant fixé. On note N_j le nombre de boules marquées j qu'on tire ainsi. Préciser la loi du vecteur $N := (N_1, \dots, N_r)$, et calculer l'espérance et la variance de N_j , la covariance de N_j et N_k , et le nombre moyen de j tels que $N_j = 0$.

6 Transformées de Laplace et de Fourier

Définition 6.0.5 Soient V_1, \dots, V_n des v.a. (défi. sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$) à valeurs dans \mathbb{R}_+ . La transformée de Laplace de la variable V_1 est la fonction de \mathbb{R}_+ dans

$]0, 1]$ définie par : $\lambda \mapsto \mathbb{E}(e^{-\lambda V_1})$. La transformée de Laplace de la variable (V_1, \dots, V_n) est la fonction de $(\mathbb{R}_+)^n$ dans $]0, 1]$ définie par : $(\lambda_1, \dots, \lambda_n) \mapsto \mathbb{E}[e^{-\lambda_1 V_1 - \dots - \lambda_n V_n}]$.

Remarque 6.0.6 Dans le cas d'une v.a. positive discrète V_1 , on écrit la transformée de Laplace sous une forme légèrement différente. C'est (modulo un changement de variable trivial) ce qu'on nomme la fonction génératrice, définie sur $]0, 1]$ par : $s \mapsto \mathbb{E}(s^{V_1})$.

Définition 6.0.7 Soient V_1, \dots, V_n des v.a.r. (sur un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$). La transformée de Fourier de la variable V_1 est la fonction de \mathbb{R} dans (le disque unité de) \mathbb{C} définie par : $t \mapsto \mathbb{E}(e^{\sqrt{-1} t V_1})$. La transformée de Fourier de la variable (V_1, \dots, V_n) est la fonction de \mathbb{R}^n dans (le disque unité de) \mathbb{C} définie par : $(t_1, \dots, t_n) \mapsto \mathbb{E}[e^{\sqrt{-1} (t_1 V_1 + \dots + t_n V_n)}]$.

Théorème 6.0.8 Les transformées de Laplace et de Fourier sont injectives : la loi d'un vecteur aléatoire (V_1, \dots, V_n) est déterminée par sa transformée de Fourier (et aussi par sa transformée de Laplace s'il est à coordonnées positives).

Remarque 6.0.9 Noter que la transformée de Fourier (ou de Laplace) d'une v.a. est en fait celle de sa loi, et que l'injectivité du théorème 6.0.8 ne vaut que pour les lois, et non pour les variables elles-mêmes. Ainsi deux v.a. ayant la même loi ont la même transformée de Fourier, sans pour autant devoir être égales. Il suffit de considérer par exemple une variable normale centrée et la variable opposée.

Corollaire 6.0.10 Des v.a. V_1, \dots, V_n sont indépendantes ssi la transformée de Fourier du vecteur aléatoire (V_1, \dots, V_n) est le produit tensoriel des transformées de Fourier des V_j :

$$\mathbb{E}[e^{\sqrt{-1} (t_1 V_1 + \dots + t_n V_n)}] = \mathbb{E}[e^{\sqrt{-1} t_1 V_1}] \dots \mathbb{E}[e^{\sqrt{-1} t_n V_n}], \quad \forall t_1 \in \mathbb{R}^{d_1}, \dots, t_n \in \mathbb{R}^{d_n}.$$

Même énoncé avec la transformation de Laplace si les V_j sont à coordonnées positives.

Preuve Dans le sens direct, c'est simplement la proposition 5.0.3, avec $f_j(v_j) := e^{\sqrt{-1} t_j v_j}$. Pour la réciproque, il suffit de considérer des v.a. V'_1, \dots, V'_n indépendantes et telle que chaque V'_j ait la même loi que V_j . Le sens direct montre que le membre de droite de la formule de l'énoncé est la transformée de Fourier de (V'_1, \dots, V'_n) . Donc le théorème 6.0.8 assure que (V_1, \dots, V_n) a la même loi que (V'_1, \dots, V'_n) , et donc que V_1, \dots, V_n sont indépendantes. \diamond

Proposition 6.0.11 La transformée de Fourier d'une v.a.r. normale N de loi $\mathcal{N}(m, \sigma^2)$ est donnée par

$$\mathbb{E}(e^{\sqrt{-1} t N}) = e^{\sqrt{-1} t m - \sigma^2 t^2 / 2}, \quad \forall t \in \mathbb{R}.$$

Preuve Posons $F(t) := \int_{\mathbb{R}} e^{\sqrt{-1} t x - x^2 / 2} \frac{dx}{\sqrt{2\pi}}$. Dérivant sous le signe \int puis intégrant par parties, on trouve : $F'(t) = -t F(t)$, équation différentielle très élémentaire. Le résultat découle immédiatement de sa résolution, par simple changement affine de variable. \diamond

Corollaire 6.0.12 La transformée de Fourier d'un vecteur gaussien V (à valeurs dans \mathbb{R}^d , identifié à une matrice-colonne, et de matrice de covariance K_V) est donnée par

$$\mathbb{E}(e^{\sqrt{-1} {}^t u V}) = \exp\left(\sqrt{-1} {}^t u \mathbb{E}(V) - \frac{1}{2} {}^t u K_V u\right), \quad \forall u \in \mathbb{R}^d.$$

Exercice n° 6.1 a) Dédurre le corollaire 6.0.12 de la définition 3.2.4 et de la proposition 6.0.11.

b) Soient N gaussienne standard, et ε indépendante de N et uniforme sur ± 1 . Montrer que εN est gaussienne standard, mais que le vecteur $(N, \varepsilon N)$ n'est pas gaussien.

c) Montrer que toute combinaison linéaire (non triviale) de v.a. normales indépendantes est normale. Donner un contreexemple simple s'il n'y a pas indépendance.

d) Montrer que si le vecteur aléatoire V a ses coordonnées gaussiennes et indépendantes, alors il est gaussien (revenir à la définition 3.2.4).

e) Montrer que les coordonnées (dans la base canonique de \mathbb{R}^d) d'un vecteur gaussien sont indépendantes ssi la matrice de covariance est diagonale, et donc ssi elles sont non corrélées. Vérifier que c'est faux si le vecteur n'est pas gaussien (même si ses coordonnées sont gaussiennes).

Exercice n° 6.2 Soit V un vecteur gaussien d -dimensionnel, et pour $1 \leq j \leq k$ soient M_j une matrice réelle de format $d_j \times d$ et $V_j := M_j V$. Montrer que V_i et V_j sont indépendants ssi $M_i K_V {}^t M_j = 0$. Généraliser à l'indépendance de V_1, \dots, V_k .

Exercice n° 6.3 Soit $V := (V_0, \dots, V_d)$ un vecteur gaussien $(d+1)$ -dimensionnel, dont les coordonnées sont $\mathcal{N}(0, 1)$ et vérifient: $\text{Cov}(V_0, V_j) = p$ pour $1 \leq j \leq d$ et $\text{Cov}(V_i, V_j) = p^2$ pour $1 \leq i \neq j \leq d$, p étant un paramètre. Posons $W_j := (1 - p^2)^{-1/2} (V_j - pV_0)$ pour $1 \leq j \leq d$. Déterminer successivement les lois de: (V_0, W_1, \dots, W_d) ; $S := \sum_{j=1}^d V_j$; S/V_0 .

Exercice n° 6.4 a) Calculer les fonctions génératrices des lois usuelles: $\mathcal{B}(n, p)$, $\mathcal{G}(p)$, $\mathcal{P}(\lambda)$.

b) Calculer les transformées de Fourier des lois usuelles: $\mathcal{U}([m, n])$, $\mathcal{B}(n, p)$, $\mathcal{G}(p)$, $\mathcal{P}(\lambda)$, $\mathcal{U}([a, b])$, $\mathcal{E}(\lambda)$.

c) Retrouver à partir de là les valeurs des espérances et variances de ces différentes lois.

Exercice n° 6.5 Étude de la transmission du nom "Chenin", porté à la génération 0 par un unique homme (= humain mâle). Notons Z_n le nombre d'hommes s'appelant "Chenin" à la génération $n \in \mathbb{N}$, F la fonction génératrice de Z_1 ($F(s) := \mathbb{E}(s^{Z_1})$), p_k la probabilité (supposée fixe) qu'un homme ait k enfants mâles, et posons $m := \sum_{k \in \mathbb{N}} k p_k$. Supposons l'indépendance entre les descendance des différents hommes.

a) Calculer par récurrence la fonction G_n , génératrice de Z_n , en fonction de F .

b) Vérifier que F est monotone et convexe sur $[0, 1]$.

c) En déduire en fonction de m le comportement asymptotique de $\alpha_n := \mathbb{P}(Z_n = 0)$. Interprétation ?

Exercice n° 6.6 Un joueur va au casino avec une fortune initiale $a \in \mathbb{N}^*$. À chaque partie, il gagne 1 avec probabilité p et perd 1 avec probabilité $q = 1 - p$. Les parties sont supposées indépendantes.

1) Fixons un entier $b > a$, et notons $P_b(a)$ la probabilité qu'a le joueur d'atteindre la fortune b avant d'être ruiné.

a) Montrer que $P_b(a) = p P_b(a+1) + q P_b(a-1)$. b) Déduire la valeur de $P_b(a)$.

2) Autorisons le joueur à s'endetter, notons T le premier instant où sa fortune vaut $a+1$, puis, pour $n \in \mathbb{N}$, $g_n := \mathbb{P}(T = n)$, et enfin g la fonction génératrice de T .

a) Montrer que $g_{n+2} = q(g_1 g_n + \dots + g_n g_1)$. b) Déduire que $g(s) - ps = qs g^2(s)$.
c) Calculer g , $\mathbb{P}(T < \infty)$, et $\mathbb{E}(T)$.

7 Convergences des variables aléatoires

Soient X et X_n , $n \in \mathbb{N}$, des v.a. à valeurs dans \mathbb{R}^d , définies sur un même espace de probabilité. Notons $\|X\| := [\omega \mapsto \|X(\omega)\|]$ une norme de X dans \mathbb{R}^d .

Définition 7.0.13 On dit que la suite $\{X_n, n \in \mathbb{N}\}$ converge vers X

i) dans L^p (où $p \in [1, \infty]$) lorsque $\mathbb{E}(\|X_n - X\|^p) \rightarrow 0$; la convergence dans L^1 est aussi appelée convergence en moyenne, et la convergence dans L^2 convergence en moyenne quadratique;

ii) presque sûrement lorsque $\mathbb{P}(X_n \text{ converge vers } X) = 1$;

iii) en probabilité lorsque pour tout $\varepsilon > 0$ $\mathbb{P}(\|X_n - X\| > \varepsilon)$ tend vers 0.

Théorème 7.0.14 1) Convergence monotone: Si (V_n) est une suite p.s. croissante de v.a.r. ≥ 0 , alors $\lim_n \uparrow \mathbb{E}(V_n) = \mathbb{E}(\lim_n \uparrow V_n)$.

2) Convergence dominée (de Lebesgue): Si (X_n) est une suite de v.a. qui converge p.s. vers une v.a. X , et si elle est dominée par une v.a.r. fixe intégrable $Y : \|X_n\| \leq Y \in L^1$, alors $X \in L^1$ et X_n converge vers X dans L^1 .

Exercice n° 7.1 a) Montrer que la convergence presque sûre et la convergence dans L^p entraînent (chacune) la convergence en probabilité.

b) Montrer que la convergence dans L^p entraîne la convergence dans L^q si $p > q$.

c) Supposons que X_n converge en probabilité vers X et que Y_n converge en probabilité vers Y . Soient a et b réels. Montrer que $aX_n + bY_n$ et $X_n \cdot Y_n$ convergent en probabilité.

Exercice n° 7.2 Trouver (sur $(\Omega = [0, 1], \mathbb{P} \equiv dx)$)

a) une suite de v.a.r. qui converge dans L^q mais pas dans L^p (pour $p > q$ fixés);

b) une suite de v.a.r. qui converge dans L^p mais pas presque sûrement (pour $p < \infty$ fixé);

c) une suite de v.a.r. qui converge presque sûrement mais dans aucun L^p .

Exercice n° 7.3 Soit $\{X_n | n \in \mathbb{N}^*\}$ une suite de v.a.r. indépendantes, telles que $\mathbb{P}(X_n = n) = 1/n$ et $\mathbb{P}(X_n = 0) = 1 - 1/n$ pour chaque $n \in \mathbb{N}^*$. Montrer que cette suite converge en probabilité, mais ni presque sûrement ni dans aucun L^p .

Exercice n° 7.4 a) Montrer qu'une suite de v.a. qui converge en probabilité admet une sous-suite qui converge presque sûrement. (Utiliser le premier lemme de Borel-Cantelli, proposition 4.0.1 section 4.)

b) Montrer qu'une suite convergente en probabilité et dominée par une variable de L^p converge dans L^p .

Exercice n° 7.5 a) Montrer que $\{X_n | n \in \mathbb{N}\}$ converge en probabilité vers X ssi $\mathbb{E}(\|X_n - X\| \wedge 1)$ tend vers 0.

b) Vérifier que $(X, Y) \mapsto \mathbb{E}(\|X - Y\| \wedge 1)$ est une distance sur $L^0(\Omega, \mathcal{T}, \mathbb{P}; \mathbb{R}^d)$.

Définition 7.0.15 On dit que la suite $\{X_n | n \in \mathbb{N}\}$ de variables aléatoires à valeurs dans \mathbb{R}^d converge en loi vers X (à valeurs dans \mathbb{R}^d) lorsque la loi de X_n converge vaguement (ou étroitement) vers celle de X . Id est: lorsque $\mathbb{E}(f \circ X_n)$ tend vers $\mathbb{E}(f \circ X)$, pour toute fonction f continue à support compact (ou continue bornée) de \mathbb{R}^d dans \mathbb{R} .

Proposition 7.0.16 La convergence en probabilité entraîne la convergence en loi.

Exercice n° 7.6 a) Prouver que lorsque X est p.s. constante, les convergences vers X en probabilité et en loi sont équivalentes.

b) Montrer que ces deux convergences sont généralement non équivalentes.

Exercice n° 7.7 Montrer que si X_n converge en loi vers X et si Y_n converge en probabilité vers 0, alors $X_n \cdot Y_n$ converge en probabilité vers 0. (Distinguer entre $\{|Y_n| > A\}$ et $\{|Y_n| \leq A\}$).

Remarque 7.0.17 a) Si les lois de X_n et de X ont un même support dénombrable discret S (en général $S \subset \mathbb{Z}^d$), alors la convergence en loi de X_n vers X équivaut à:

$$\mathbb{P}(X_n = s) \text{ tend vers } \mathbb{P}(X = s) \text{ pour tout } s \in S.$$

b) Si X_n converge en loi vers X et si g est continue, alors $g \circ X_n$ converge en loi vers $g \circ X$.

Proposition 7.0.18 La convergence en loi équivaut à la convergence simple des transformées de Fourier: $X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X \iff \left(\mathbb{E}[e^{\sqrt{-1} t \cdot X_n}] \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[e^{\sqrt{-1} t \cdot X}] \quad \forall t \right)$.

Proposition 7.0.19 La convergence en loi d'une suite de v.a.r. $\{X_n | n \in \mathbb{N}^*\}$ vers une v.a.r. X équivaut à la convergence simple de F_n vers F en chaque point de continuité de F ; F_n désignant la fonction de répartition de X_n , et F celle de X .

- Exemples a) La loi $B(n, p_n)$ converge vers la loi $\mathcal{P}(\lambda)$ si $n \rightarrow \infty$ et $np_n \rightarrow \lambda > 0$.
 b) La loi $H(N, n, p_N)$ converge vers la loi $B(n, p)$ si $N \rightarrow \infty$ et $p_N \rightarrow p \in]0, 1[$.
 c) Soient B une variable de Bernoulli ($\mathbb{P}(B = 1) = \mathbb{P}(B = -1) = 1/2$), et pour tout $n \in \mathbb{N} : X_n = B = X'_{2n}$ et $X'_{2n+1} = -B$. Alors les 2 suites X et X' convergent en loi, mais pas la suite $X + X'$.

Exercice n° 7.8 Notons $\{X_n | n \in \mathbb{N}^*\}$ une suite de v.a.r. indépendantes gaussiennes standard. Étudier la convergence en loi de $\frac{nX_1 + (n-1)X_2 + \dots + X_n}{n\sqrt{n}}$.

Exercice n° 7.9 Soit $\{X_n | n \in \mathbb{N}\}$ une suite de v.a.r. de fonction de répartition commune F , telle que $\lim_{x \rightarrow \infty} x(1-F(x)) = \lim_{x \rightarrow \infty} xF(-x) = 0$. Posons $M_n := \max\{X_1, \dots, X_n\}$, et $m_n := \min\{X_1, \dots, X_n\}$. Montrer que $\frac{M_n}{n}$ et $\frac{m_n}{n}$ convergent en probabilité vers 0.

Exercice n° 7.10 Soit $\{X_j | j \in \mathbb{N}^*\}$ une suite de v.a.i.i.d. $\mathcal{N}(0, 1)$. Soit $\{a_j | j \in \mathbb{N}\}$ une suite de réels tels que $a_j a_{j+1} = 0$ pour tout j et tels que $\sum_j a_j^2 < \infty$. Soit $Y_n := \sum_{j=1}^n a_{n-j} X_j$, pour tout $n \in \mathbb{N}^*$.

- a) Étudier la convergence en loi de la suite (Y_n) . b) Les v.a. Y_n et Y_{n+1} sont-elles indépendantes ? c) Étudier la convergence dans L^2 de la suite (Y_n) .

8 Loi des Grands Nombres

Théorème 8.0.20 Soit $\{X_n | n \in \mathbb{N}^*\}$ une suite de v.a. indépendantes et intégrables, de même loi. Alors la suite des moyennes de Césaro $\frac{X_1 + \dots + X_n}{n}$ converge p.s. vers $\mathbb{E}(X_1)$.

Remarque 8.0.21 La réciproque suivante de la loi des grands nombres est vraie :
 Si $\{X_n | n \in \mathbb{N}^*\}$ est une suite de v.a.r indépendantes et de même loi telle que la suite des moyennes de Césaro $\frac{X_1 + \dots + X_n}{n}$ converge p.s., alors X_1 est intégrable.

Preuve Remarquons qu'alors $\frac{X_n}{n} = \frac{X_1 + \dots + X_n}{n} - \frac{X_1 + \dots + X_{n-1}}{n-1} \times \frac{n-1}{n}$ doit converger p.s. vers 0. Donc $\mathbb{E}(|X_1|) \leq \sum_n \mathbb{P}(|X_n| > n)$ est finie d'après le deuxième lemme de Borel-Cantelli.
 ◇

Remarque 8.0.22 1) Le théorème 8.0.20 ci-dessus constitue la loi forte des grands nombres. On appelle loi faible des grands nombres le résultat (strictement plus faible) qui énonce la convergence en probabilité des moyennes, au lieu de leur convergence presque sûre.

2) On montre que la convergence dans la loi forte des grands nombres a lieu également dans L^1 .

Exercice n° 8.1 Prouver la formulation suivante de la loi faible des grands nombres :

Si les v.a.r. X_n sont non-corrélées 2 à 2 et ont la même loi admettant un second moment, alors leurs moyennes de Césaro convergent dans L^2 (et donc en probabilité) vers $\mathbb{E}(X_1)$.

Exercice n° 8.2 Soit $\{X_n | n \in \mathbb{N}^*\}$ une suite de v.a.r.i.i.d. telles que $\mathbb{E}(X_1^+) = \infty > \mathbb{E}(X_1^-)$. Montrer que p.s. $\frac{X_1 + \dots + X_n}{n} \rightarrow +\infty$. (Utiliser $X_n^k := X_n \wedge k$).

Exercice n° 8.3 a) Soient $p \in [0, 1]$ et f une fonction continue de $[0, 1]$ dans \mathbb{R} . Étudier la convergence de $n \mapsto \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} f(k/n)$. (Penser à la remarque 3.1.5).

b) Soient $\lambda > 0$ et g une fonction continue bornée de \mathbb{R}_+ dans \mathbb{R} . Étudier la convergence de $n \mapsto e^{-\lambda n} \sum_{k \geq 0} \frac{(\lambda n)^k}{k!} g(k/n)$; puis de $n \mapsto e^{-\lambda n} \sum_{k=0}^{[\lambda n]} \frac{(\lambda n)^k}{k!}$ (utiliser le théorème suivant; interpréter en terme de médiane).

9 Théorème Central Limite

Théorème 9.0.23 Soit $\{X_n | n \in \mathbb{N}^*\}$ une suite de v.a. à valeurs dans \mathbb{R}^d , indépendantes, et de même loi admettant un second moment. Alors $\frac{X_1 + \dots + X_n - n\mathbb{E}(X_1)}{\sqrt{n}}$ converge en loi, vers la loi gaussienne $\mathcal{N}(0, K_{X_1})$, où $K_{X_1} := \mathbb{E}({}^t X_1 X_1) - {}^t \mathbb{E}(X_1) \mathbb{E}(X_1)$ désigne la matrice de covariance de X_1 .

Preuve L'énoncé étant clairement invariant par une translation sur la suite X , il suffit de considérer le cas où X_1 est centrée. Notons $\phi(v) := \mathbb{E}(e^{\sqrt{-1}v \cdot X_1})$ la transformée de Fourier de X_1 , qui est de classe C^2 , puisque X_1 est supposée de carré intégrable. La formule de Taylor-Young à l'origine s'écrit alors : $\phi(v) = 1 - \frac{1}{2} {}^t v K(X_1) v + o(\|v\|^2)$. Par indépendance, la transformée de Fourier de $\frac{X_1 + \dots + X_n}{\sqrt{n}}$ est égale à $\phi(v/\sqrt{n})^n$, qui vaut $\left[1 - \frac{{}^t v K(X_1) v + n o(\|v\|^2/n)}{2n}\right]^n$, et donc qui converge vers $\exp[-\frac{1}{2} {}^t v K(X_1) v]$, du fait de la convergence de $(1 + \frac{z}{n})^n$ vers e^z uniformément sur les compacts de \mathbb{C} .

En effet, si $|z| \leq A$, nous avons :

$$\begin{aligned} \left| e^z - \left(1 + \frac{z}{n}\right)^n \right| &\leq \sum_{k=0}^n \left(\frac{1}{k!} - \frac{C_n^k}{n^k} \right) |z|^k + \sum_{k=n+1}^{\infty} \frac{|z|^k}{k!} \leq \sum_{k=0}^n \left(\frac{1}{k!} - \frac{C_n^k}{n^k} \right) A^k + \sum_{k=n+1}^{\infty} \frac{A^k}{k!} \\ &= e^A - \left(1 + \frac{A}{n}\right)^n = e^A - e^{n \log(1 + \frac{A}{n})} \rightarrow 0. \end{aligned}$$

Finalement, nous venons de montrer la convergence simple de la transformée de Fourier de $\frac{X_1 + \dots + X_n}{\sqrt{n}}$ vers celle de $\mathcal{N}(0, K_{X_1})$, ce qui suffit d'après la proposition 7.0.18. \diamond

Nota Bene Le TCL évalue les fluctuations d'une somme de v.a.i.i.d. (de carré intégrable) autour du terme (déterministe) dominant, donné par la LGN; en dimension 1 (pour simplifier) on peut en effet le réécrire sous forme d'égalité en loi :

$$X_1 + \dots + X_n \stackrel{\text{loi}}{=} n \mathbb{E}(X_1) + \sqrt{n} \sigma(X_1) \mathcal{N}(0, 1) + o(\sqrt{n}).$$

Exercice n° 9.1 (Surlocation) Une agence de voyage dispose de 160 places à louer pour une destination donnée. Elle sait que les locations sont honorées par ses clients avec une probabilité fixe p . Elle vend $N = 160\alpha > 160$ places. Pour quelles valeurs de α la probabilité de ne pas louer trop de places vaut-elle 0,95 ; 0,975 ?

Rappel Valeurs numériques à connaître :

$$\mathcal{N}(0, 1)(] - \infty, 1,65]) \approx 0,95, \quad \text{et} \quad \mathcal{N}(0, 1)(] - \infty, 1,96]) \approx 0,975.$$

Exercice n° 9.2 Le prix S_n d'une action au jour n est modélisé ainsi : $S_0 = s > 0$ est fixe, et $S_{n+1} = (1 + r + \sigma\varepsilon_{n+1})S_n$, où $r > 0$ est un taux fixe, $\sigma \in]0, 1 + r[$ est une volatilité fixe, et $\{\varepsilon_n | n \in \mathbb{N}\}$ est une suite i.i.d. de loi de Bernoulli $\mathcal{B}(\pm 1, 1/2)$.

- Étudier le comportement des suites $(\log S_n)/n$ et S_n .
- Étudier le comportement de la suite $(\log S_n)/\sqrt{n}$.
- Étudier le comportement de la suite $[(1 + r)^2 - \sigma^2]^{(-\sqrt{n}/2)} \times S_n^{[1/\sqrt{n}]}$.

Exercice n° 9.3 Notons $\{X_n, Y_n, Z_n | n \in \mathbb{N}^*\}$ une famille de v.a.i.i.d. de loi commune $\mathcal{B}(\pm 1, 1/2)$. Posons pour tout $n \in \mathbb{N}$:

$$R_n := \sum_{k=1}^n X_k Y_k, \quad S_n := \sum_{k=1}^n Y_k Z_k, \quad T_n := \sum_{k=1}^n Z_k X_k, \quad \text{puis} \quad V_n := (R_n, S_n, T_n).$$

- R_n, S_n, T_n sont-elles indépendantes 2 à 2 ? indépendantes ?
- Quelle est la loi de R_n ? c) Calculer la transformée de Fourier de V_n .
- Étudier la convergence de V_n/n , puis de V_n/\sqrt{n} .

Exercice n° 9.4 3016 mathématiciens sont invités à un colloque ; en moyenne un sur 4 répondra favorablement, et les réponses sont supposées indépendantes les unes des autres. Combien les organisateurs doivent-ils prévoir de places, afin que la probabilité de ne pas en manquer soit $\geq 0,99$?

Exercice n° 9.5 Une compagnie assure 10000 clients sur la vie, qui payent chacun une prime annuelle de A euros. On estime que chaque client a une probabilité de décès au cours d'une année égale à $6/1000$, indépendamment les uns des autres. La prime de décès est de B euros.

- Quelle est la loi du nombre annuel des décès ? Comment peut-on l'approcher ?
- Si $B = 1000$, pour quels A la compagnie a-t-elle une probabilité $< 1/100$ d'être en déficit ? c) Si $A = 15$, pour quels B la compagnie a-t-elle une probabilité $> 0,7$ de faire un bénéfice annuel > 50000 euros ?

Exercice n° 9.6 Soit $\{X_n | n \in \mathbb{N}\}$ une suite de v.a.i.i.d. de loi commune $\mathcal{B}(p)$. Posons $Y_n := X_n X_{n+1}$ et $S_n := (Y_1 + \dots + Y_n)/n$, pour tout $n \in \mathbb{N}$.

Calculer la loi et l'espérance de Y_n , puis la covariance de Y_n et de Y_{n+k} , l'espérance de S_n^2 , et enfin montrer que S_n converge en probabilité vers p^2 .

10 Exemple des marches aléatoires

Définition 10.0.24 Soit $\{X_n | n \in \mathbb{N}^*\}$ une suite de v.a. à valeurs dans \mathbb{Z}^d , indépendantes, et de même loi μ . On appelle marche aléatoire (sur \mathbb{Z}^d) le processus $\{S_n | n \in \mathbb{N}\}$ à valeurs dans \mathbb{Z}^d défini par: $S_0 := 0$, et pour tout $n \in \mathbb{N}^*$: $S_n := X_1 + \dots + X_n$. La marche est dite simple lorsque $\mu(\{e_j\}) = \mu(\{-e_j\}) = 1/(2d)$, où (e_1, \dots, e_d) désigne la base canonique de \mathbb{Z}^d .

Définition 10.0.25 Soit pour tout $n \in \mathbb{N}^*$: $\mathcal{F}_n := \sigma(X_1, \dots, X_n) = \sigma(S_1, \dots, S_n)$ la tribu représentant l'information connue au temps n . On note aussi \mathcal{F}_0 la tribu triviale, et \mathcal{F}_∞ la tribu engendrée par $\bigcup_{n \in \mathbb{N}} \mathcal{F}_n$.

On appelle temps d'arrêt toute v.a. N à valeurs dans $\overline{\mathbb{N}}$ telle que $\{N \leq n\} \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}$. On pose alors: $\mathcal{F}_N := \{A \in \mathcal{F}_\infty | (\forall n \in \mathbb{N}) A \cap \{N \leq n\} \in \mathcal{F}_n\}$.

L'exemple fondamental de temps d'arrêt est le temps d'atteinte par le processus d'une certaine partie E de \mathbb{Z}^d : $\min\{n \in \mathbb{N} | S_n \in E\}$ est bien un temps d'arrêt.

Un autre exemple, trivial, est fourni par les temps constants.

Exercice n° 10.1 Montrer que N est un temps d'arrêt ssi $\{N = n\} \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}$, que $\mathcal{F}_N := \{A \in \mathcal{F}_\infty | (\forall n \in \mathbb{N}) A \cap \{N = n\} \in \mathcal{F}_n\}$, et que \mathcal{F}_N est une tribu.

Exercice n° 10.2 Soient N et N' deux temps d'arrêt.

a) Vérifier que $\min\{N, N'\}$ et $\max\{N, N'\}$ sont aussi deux temps d'arrêt.

b) Montrer que les événements $\{N < N'\}$, $\{N \leq N'\}$, $\{N = N'\}$ appartiennent à $\mathcal{F}_{\min\{N, N'\}}$, puis que $\mathcal{F}_{\min\{N, N'\}} = \mathcal{F}_N \cap \mathcal{F}_{N'}$.

Exercice n° 10.3 Soit N un temps d'arrêt. Supposons que $\mathbb{E}(\|X_1\|)$ et $\mathbb{E}(N)$ sont finies. a) Montrer que $\mathbb{E}(S_N) = \mathbb{E}(X_1) \times \mathbb{E}(N)$. (Ecrire $S_N = \sum_j X_j 1_{\{N < j\}^c}$).

b) Dédurre (pour la marche simple, $d = 1$) que $\min\{n | S_n = 1\}$ et T ne sont pas intégrables.

Lemme 10.0.26 Soit N un temps d'arrêt. Conditionnellement à $\{N < \infty\}$, la suite $\{X_{N+n} | n \in \mathbb{N}^*\}$ est indépendante de \mathcal{F}_N , et a la même loi que $\{X_n | n \in \mathbb{N}^*\}$.

Preuve Fixons $A \in \mathcal{F}_N$, $k \in \mathbb{N}^*$, et $v_1, \dots, v_k \in \mathbb{Z}^d$. Nous avons

$$\mathbb{P}(A, N < \infty, X_{N+1} = v_1, \dots, X_{N+k} = v_k) = \sum_{n \in \mathbb{N}} \mathbb{P}(A, N = n, X_{n+1} = v_1, \dots, X_{n+k} = v_k)$$

$$= \sum_{n \in \mathbb{N}} \mathbb{P}(A, N = n) \mathbb{P}(X_{n+1} = v_1, \dots, X_{n+k} = v_k) = \mathbb{P}(A, N < \infty) \prod_{j=1}^k \mu(v_j). \diamond$$

Remarque 10.0.27 Quitte à prendre l'image de Ω par X , on peut toujours se placer sur l'espace canonique, i.e. considérer que $\Omega = (\mathbb{Z}^d)^{\mathbb{N}^*}$, $\mathcal{F} = \mathcal{F}_\infty$, de sorte que X devient la suite des coordonnées: $X_n(\omega) = \omega_n, \forall n \in \mathbb{N}$. \mathbb{P} demeure la loi de la marche S , déduite de $(X, \mu^{\otimes \mathbb{N}^*})$.

10.1 Temps de retour en 0

Soient $T_0 := 0$, $T_1 = T := \min\{n \in \mathbb{N}^* \mid S_n = 0\}$ le premier temps de retour en 0 (qui vaut ∞ s'il n'y en a pas), puis pour $k \in \mathbb{N}^*$: $T_k := \min\{n > T_{k-1} \mid S_n = 0\}$ le k -ième temps de retour en 0.

Exercice n° 10.4 a) Montrer que $\mathbb{P}(T_k < \infty) = \mathbb{P}(T < \infty)^k$. (Utiliser le lemme 10.0.26).

b) Pour tout $v \in \mathbb{Z}^d, v \neq 0$, notons $T_k(v)$ le k -ième temps de passage en v . Montrer que $\mathbb{P}(T_k(v) < \infty) = \mathbb{P}(T_1(v) < \infty) \times \mathbb{P}(T < \infty)^{k-1}$.

Proposition 10.1.1 a) Les énoncés suivant sont équivalents :

(i) $\mathbb{P}(T < \infty) = 1$. (ii) $\mathbb{P}(\limsup_n \{S_n = 0\}) = 1$. (iii) $\sum_n \mathbb{P}(S_n = 0) = \infty$.

b) Si $\mathbb{P}(T < \infty) < 1$, alors $\mathbb{P}(\limsup_n \{S_n = v\}) = 0$ pour tout $v \in \mathbb{Z}^d$.

Preuve a) L'exercice 10.4.a montre que (i) entraîne $\mathbb{P}(T_n < \infty (\forall n)) = 1$, c'est-à-dire (ii). Le premier lemme de Borel-Cantelli montre que (ii) \Rightarrow (iii). Ensuite nous avons $\sum_n 1_{\{S_n=0\}} = \sum_n 1_{\{T_n < \infty\}}$ presque sûrement, et donc en intégrant : $\sum_{n \in \mathbb{N}} \mathbb{P}(\{S_n = 0\}) = \sum_{n \in \mathbb{N}} \mathbb{P}(\{T_n < \infty\}) = \sum_n \mathbb{P}(\{T < \infty\})^n = [1 - \mathbb{P}(\{T < \infty\})]^{-1}$ d'après l'exercice 10.4.a à nouveau, d'où (iii) \Rightarrow (i).

b) $\mathbb{E}[\sum_n 1_{\{S_n=v\}}] = \sum_n \mathbb{P}(\{T_n(v) < \infty\}) = \mathbb{P}(T_1(v) < \infty) / \mathbb{P}(T = \infty) < \infty$ d'après l'exercice 10.4.b, et donc $\mathbb{P}(\limsup_n \{S_n = v\}) = \mathbb{P}[\sum_n 1_{\{S_n=v\}} = \infty] = 0$. \diamond

Proposition 10.1.2 Soient $\mathcal{R} := \{v \in \mathbb{Z}^d \mid \limsup_n \{S_n = v\} = \Omega \text{ p.s.}\}$ l'ensemble des points presque sûrement infiniment visités par la marche aléatoire, et $\mathcal{E} := \{v \in \mathbb{Z}^d \mid (\exists n \in \mathbb{N}) \mathbb{P}(S_n = v) > 0\}$ l'ensemble des points pouvant être visités par la marche aléatoire.

Alors soit \mathcal{R} est vide, soit c'est un sous-groupe de \mathbb{Z}^d et il est égal à \mathcal{E} .

Preuve Soient $v \in \mathcal{E}$, et $w \notin v + \mathcal{R}$. Fixons $k \in \mathbb{N}$ tel que $\mathbb{P}(S_k = v) > 0$, et $m \in \mathbb{N}^*$ tel que $\mathbb{P}((\forall n \geq m) S_n \neq w - v) > 0$. Alors

$$\begin{aligned} \mathbb{P}((\forall n \geq m+k) S_n \neq w) &\geq \mathbb{P}((\forall n \geq m+k) S_n - S_k \neq w - v, S_k = v) \\ &= \mathbb{P}((\forall n \geq m+k) S_n - S_k \neq w - v) \mathbb{P}(S_k = v) = \mathbb{P}((\forall n \geq m) S_n \neq w - v) \mathbb{P}(S_k = v) > 0. \end{aligned}$$

Donc nous venons de prouver que si $w - v \notin \mathcal{R}$ et $v \in \mathcal{E}$, alors $w \notin \mathcal{R}$. Autrement dit : $v \in \mathcal{E}$ et $w \in \mathcal{R}$ impliquent $w - v \in \mathcal{R}$.

Remarquons que $\mathcal{R} \subset \mathcal{E}$. Donc prenant v et w dans \mathcal{R} , nous voyons que \mathcal{R} est bien un sous-groupe, s'il est non vide. Enfin, dans ce cas toujours, prenant $w = 0$, nous obtenons $\mathcal{E} \subset \mathcal{R}$. \diamond

10.2 Récurrence des marches aléatoires simples

Dans le cas de la marche simple sur \mathbb{Z}^d (voir la définition 10.0.24), il est clair que $\mathcal{E} = \mathbb{Z}^d$.

Les propositions 10.1.1 et 10.1.2 nous assurent donc que soit la marche simple repasse indéfiniment par chaque point de \mathbb{Z}^d , soit chaque point n'est visité qu'un nombre fini de fois par elle, et donc elle tend vers l'infini, tout ceci presque sûrement. Dans le premier cas, on dit que la marche est récurrente, et dans le second, qu'elle est transitoire (ou transiente).

Le célèbre résultat suivant peut se traduire en langage courant par la phrase : un homme ivre finira par retrouver sa demeure, alors qu'un oiseau ivre peut se perdre définitivement.

Théorème 10.2.1 *La marche aléatoire simple sur \mathbb{Z}^d est récurrente pour $d = 1$ ou $d = 2$, mais est transitoire pour $d \geq 3$.*

Preuve La marche est récurrente ssi 0 est récurrent, i.e. ssi la marche repasse presque sûrement indéfiniment par 0, et donc, d'après la proposition 10.1.1, ssi $\sum_n \mathbb{P}(S_n = 0) = \infty$.

Remarquons que $P(S_{2n+1} = 0) = 0$ pour tout n . Nous avons donc récurrence ssi $\sum_n \mathbb{P}(S_{2n} = 0) = \infty$.

Pour $d = 1$, nous avons aussitôt $\mathbb{P}(S_{2n} = 0) = 2^{-2n} C_{2n}^n \sim (\pi n)^{-1/2}$, puisqu'il faut autant de pas vers la gauche que vers la droite, et en utilisant la formule de Stirling.

Pour $d = 2$, distinguant selon le nombre m de déplacements de la marche dans le sens horizontal, un dénombrement simple donne : $\mathbb{P}(S_{2n} = 0) = 4^{-2n} \sum_{m=0}^n \frac{(2n)!}{m!m!(n-m)!(n-m)!}$, d'où $\mathbb{P}(S_{2n} = 0) = 2^{-4n} C_{2n}^n \sum_{m=0}^n C_n^m C_n^{n-m} = (2^{-2n} C_{2n}^n)^2 \sim (\pi n)^{-1}$.

Pour $d = 3$, le même genre de considérations donne :

$$\begin{aligned} \mathbb{P}(S_{2n} = 0) &= 6^{-2n} \sum_{0 \leq l+m \leq n} \frac{(2n)!}{[l!m!(n-l-m)!]^2} = 2^{-2n} C_{2n}^n \sum_{0 \leq l+m \leq n} \left(3^{-n} \frac{n!}{l!m!(n-l-m)!} \right)^2 \\ &\leq 2^{-2n} C_{2n}^n \max_{0 \leq l+m \leq n} \left\{ 3^{-n} \frac{n!}{[l!m!(n-l-m)!]^2} \right\} \sum_{0 \leq l+m \leq n} \left(3^{-n} \frac{n!}{l!m!(n-l-m)!} \right) \\ &= 2^{-2n} C_{2n}^n \max_{n_1+n_2+n_3=n} \left\{ 3^{-n} \frac{n!}{n_1!n_2!n_3!} \right\} = \frac{2^{-2n} C_{2n}^n \times 3^{-n} \times n!}{\min \{ n_1!n_2!n_3! \mid n_1+n_2+n_3=n, |n_i-n_j| \leq 1 \}}, \end{aligned}$$

puisque si par exemple $n_1 - n_2 \geq 2$ alors $n_1!n_2!n_3! > (n_1 - 1)!(n_2 + 1)!n_3!$; cela impose $n_1, n_2, n_3 \geq (n - 2)/3$, d'où par la formule de Stirling la conclusion pour $d = 3$:

$$\mathbb{P}(S_{2n} = 0) = \mathcal{O} \left(\frac{n^{-1/2} \times 3^{-n} \times (n/e)^n \sqrt{n}}{(n_1/e)^{n_1} (n_2/e)^{n_2} (n_3/e)^{n_3} \sqrt{n_1 n_2 n_3}} \right) = \mathcal{O} \left(\frac{3^{-n} \times n^n}{\left(\frac{n-2}{3}\right)^{n+3/2}} \right) = \mathcal{O}(n^{-3/2}).$$

Finalement, si $d > 3$, considérons la projection S'_n de la marche S_n sur les trois premières coordonnées de \mathbb{Z}^d , puis $T_0 := 0$, et pour $n \in \mathbb{N}^*$:

$T_n := \min\{m > T_{n-1} \mid S'_m \neq S'_{T_{n-1}}\}$. Alors T_n est p.s. fini, S'_{T_n} est la marche simple sur \mathbb{Z}^3 , et donc ne passe plus par 0 après un certain temps aléatoire presque sûrement fini N . Comme S'_n est constante entre deux instants T_{n-1} et T_n , nous avons a fortiori $S'_n \neq 0$ (et donc $S_n \neq 0$) pour $n > T_N$, d'où la transience annoncée. \diamond

11 Processus de Poisson

Définition 11.0.2 On appelle processus de comptage tout processus $(N_t \mid t \in \mathbb{R}_+)$ donné par $N_t := \text{Card}(\mathcal{E} \cap [0, t])$, où la donnée \mathcal{E} est une partie discrète aléatoire de \mathbb{R}_+ .

Le processus de comptage N est dit à accroissements indépendants lorsque pour tout n dans \mathbb{N}^* et tous $s_1 < t_1 \leq \dots \leq s_n < t_n$ les variables $(N_{t_1} - N_{s_1}), \dots, (N_{t_n} - N_{s_n})$ sont indépendantes. Cela signifie l'indépendance des nombres de points de \mathcal{E} situés dans des intervalles disjoints deux à deux.

Remarques 1) Un processus de comptage est nécessairement croissant, à valeurs dans $\overline{\mathbb{N}}$, et presque sûrement continu à droite en tout temps (y compris aléatoire) :

$$\mathbb{P}(\inf N_{S_n} \neq N_S) \leq \mathbb{P}(N_{S_n} - N_S \geq 1) = \mathbb{P}(\mathcal{E} \cap]S, S_n] \neq \emptyset) \searrow 0 \text{ si } S_n \searrow S \text{ p.s..}$$

2) Les points de \mathcal{E} sont généralement des instants successifs de survenue d'événements (d'un type donné a priori : arrivées dans une file, passages de bus, désintégrations d'atomes, etc.), qu'on observe au cours du temps t . Pour tous $0 \leq s < t$, la variable aléatoire $N_t - N_s$ décompte le nombre de ces événements survenus dans l'intervalle de temps $]s, t]$.

Définition 11.0.3 On appelle processus de Poisson (homogène) d'intensité λ (quelconque dans \mathbb{R}_+^*) tout processus de comptage N à accroissements indépendants nul en 0 et tel que pour tous $s, t \in \mathbb{R}_+$ la loi de $(N_{s+t} - N_s)$ est $\mathcal{P}(\lambda t)$, la loi de Poisson de paramètre λt .

Fixons dorénavant un processus de Poisson N , de filtration (\mathcal{F}_t) et d'intensité λ .

Exercice n° 11.1 a) Un processus de Poisson est à accroissements stationnaires : la loi de tout vecteur $(N_{a+t_1} - N_{a+s_1}, \dots, N_{a+t_n} - N_{a+s_n})$ est indépendante de $a \in \mathbb{R}_+$.

b) Un processus de Poisson est p.s. fini à tout instant : $\mathbb{P}(\bigcap_t \{N_t < \infty\}) = 1$: les événements qu'il décompte ne peuvent pas s'accumuler en un instant fini.

c) En temps t petit, on a : $\mathbb{P}(N_t = 1) = \lambda t + o(t)$ et $\mathbb{P}(N_t \geq 1) = \lambda t + o(t)$.

d) On a presque sûrement : $\lim_{t \rightarrow \infty} N_t/t = \lambda = \mathbb{E}(N_t/t)$.

Lemme 11.0.4 On a presque sûrement $N_{S_1} = 1$ (où $S_1 := \inf\{t > 0 \mid N_t > 0\}$).

Corollaire 11.0.5 Les points de \mathcal{E} s'ordonnent en une suite strictement croissante de variables aléatoires tendant vers $+\infty$: $S_0 := 0 < S_1 < \dots < S_n < \dots$. Pour tout $n \in \mathbb{N}$, S_n est le n -ième point de \mathcal{E} .

Preuve Définissons par récurrence $S_0 := 0$ et $S_{n+1} := \inf\{t > 0 \mid N_t > N_{S_n}\}$. Le lemme 11.0.4 permet de voir que ce sont des temps p.s. finis, et que $\mathbb{P}(N_{S_{n+1}} - N_{S_n} = 1) = 1$. Ceci montre que les points de \mathcal{E} surviennent bien lors d'une suite strictement croissante, et que S_n est le n -ième point de \mathcal{E} . L'exercice (11.1.b) montre que cette suite tend vers l'infini presque sûrement. \diamond

Théorème 11.0.6 *Pour tout $n \in \mathbb{N}^*$, notons $T_n := S_n - S_{n-1}$ l'intervalle de temps entre les $(n-1)$ -ième et n -ième points de \mathcal{E} .*

La suite $(T_n, n \in \mathbb{N}^)$ est i.i.d., de loi commune la loi exponentielle de paramètre λ .*

Exercice 11.2 a) Montrer que pour tout $n \in \mathbb{N}^*$ la loi de S_n est la loi Gamma de paramètres n et λ , i.e. la loi de densité sur \mathbb{R}_+ : $s \longrightarrow \frac{(\lambda s)^{n-1}}{(n-1)!} \times \lambda e^{-\lambda s}$.

b) Pour $t > 0$ et $f \in L^1([0, t])$, calculer $\mathbb{E}\left(\sum_{n \in \mathbb{N}^*} f(S_n) 1_{\{S_n \leq t\}}\right)$.

c) Fixons $t > 0$ et $n \in \mathbb{N}^*$. Montrer que conditionnellement par rapport à $\{N_t = n\}$, le vecteur (S_1, \dots, S_n) admet la densité de Dirichlet: $(s_1, \dots, s_n) \longrightarrow \frac{n!}{t^n} 1_{\{0 < s_1 < \dots < s_n < t\}}$.

d) Montrer que (pour $t > 0$ et $n \in \mathbb{N}^*$) si U_1, \dots, U_n sont n variables i.i.d. uniformes sur $[0, t]$, et si (V_1, \dots, V_n) est le vecteur obtenu en ordonnant U_1, \dots, U_n dans le sens croissant, alors la "statistique d'ordre" (V_1, \dots, V_n) admet aussi la densité de Dirichlet ci-dessus.

e) Dédurre que pour $0 < s < t$, conditionnellement par rapport à $\{N_t = n\}$, la loi de N_s est $\mathcal{B}(n, s/t)$.

Il est frappant que la réciproque du théorème 11.0.6 soit exacte, permettant par exemple de simuler un processus de Poisson.

Théorème 11.0.7 *Soit $\{T_n; n \in \mathbb{N}^*\}$ une suite i.i.d. de variables exponentielles de paramètre $\lambda > 0$. Soit (pour tout $t \in \mathbb{R}_+$) $N_t := \sum_{n \in \mathbb{N}^*} 1_{\{T_1 + \dots + T_n \leq t\}}$. Alors N est un processus de Poisson d'intensité λ .*

Noter que cela signifie en particulier que le processus enregistré par un compteur Geiger est un processus de Poisson, dont l'intensité ($\lambda = \frac{\log 2}{D}$ si D est la demi-vie du matériau radioactif, voir l'exercice 3.2.3.c) est celle de la radioactivité mesurée.

Exercice 11.3 a) Soient N et N' deux processus de Poisson indépendants, de paramètres λ et λ' . Montrer que $N + N'$ définit un processus de Poisson de paramètre $\lambda + \lambda'$.

b) Soit $\{X_n \mid n \in \mathbb{N}\}$ une suite i.i.d. indépendante de N . Le processus défini par $\Sigma_t := \sum_{n=1}^{N_t} X_n$ est appelé "processus de Poisson composé". (Exemple: N peut décrire le nombre de clients d'un magasin et X_n la dépense du n -ième client.) Calculer la transformée de Fourier de Σ_t en fonction de celle des X_n et de la fonction génératrice de N_t , puis l'espérance et la variance de Σ_t .

Exercice 11.4 a) Soit N un processus de Poisson d'intensité λ comptant les passages successifs de voitures sur une route. Un piéton a besoin d'une durée $d > 0$ entre deux voitures pour traverser la route. Calculer la transformée de Laplace du temps Y qu'il doit attendre avant de pouvoir traverser. (On trouve $a \mapsto \frac{a+\lambda}{ae^{\lambda d} + \lambda e^{-ad}}$.) Calculer l'espérance de Y . (On trouve $\frac{e^{\lambda d} - 1 - \lambda d}{\lambda}$.)

b) Considérons une route à deux voies, les voitures étant comptées par deux processus de Poisson indépendants, d'intensité λ dans un sens et λ' dans l'autre, et nécessitant des durées respectives d et d' pour être traversées. Soient Z le temps que le piéton doit attendre pour traverser les deux voies d'un coup, et Z' le temps qu'il doit attendre pour traverser les deux voies sachant qu'il peut les traverser séparément du fait de la présence d'un refuge au milieu de la route. Donner les transformées de Laplace de Z et de Z' , leur espérances, et discuter l'intérêt du refuge.

Proposition 11.0.8 (*Processus de Poisson marqué*) Fixons un entier $r \geq 2$ et une probabilité $p = (p_1, \dots, p_r)$ sur l'ensemble $\{1, \dots, r\}$. Soit $\{Y_n | n \in \mathbb{N}^*\}$ une suite i.i.d. indépendante de N , de loi commune p . Accolons à chaque instant S_n décompté par N le type Y_n , et notons N^j le processus de comptage des instants de type j . Alors chaque N^j est un processus de Poisson d'intensité λp_j , et les processus N^j sont indépendants.

Preuve Conditionnellement à $\{N_t = k\}$, les instants S_1, \dots, S_k se répartissent dans les différents types suivant la loi multinômiale de paramètres k et p_1, \dots, p_r , par définition de la procédure. Nous avons donc pour tous $k_1, \dots, k_r \in \mathbb{N}$, notant $k := k_1 + \dots + k_r$:

$$\mathbb{P}(N_t^1 = k_1, \dots, N_t^r = k_r) = \frac{k!}{k_1! \dots k_r!} (p_1)^{k_1} \dots (p_r)^{k_r} \times e^{-\lambda t} \frac{(\lambda t)^k}{k!} = \prod_{j=1}^r e^{-\lambda p_j t} \frac{(\lambda p_j t)^{k_j}}{k_j!}.$$

Ceci prouve que pour chaque t les variables N_t^1, \dots, N_t^r sont indépendantes et ont chacune la loi de Poisson annoncée. Enfin la nullité en 0, la stationnarité, et l'indépendance des accroissements sont immédiats. \diamond

Exemple Le processus (N_t) est celui des arrivées des clients dans un grand magasin, et le type est l'étage (supposé ici unique) visité. Ainsi (N_t^j) décompte les arrivées des clients au j -ième étage du magasin.

Exercice 11.5 Le flux (poissonnien) des trams passant aux halles est en moyenne de 12 trams par heure et comporte une proportion $p = 40\%$ de trams de la ligne D (et de $1 - p$ pour la ligne A).

- Quelle est la probabilité qu'au moins 2 trams D passent en un quart d'heure donné ?
- Sachant que 3 trams D sont passés en 20 minutes, quel est le nombre moyen de trams passés dans le même temps ?
- Sachant que 10 trams sont passés en 45 minutes, quelle est la probabilité que la moitié soient des trams D ?

Exercice 11.6 Pour $t > 0$ fixé, vérifier que $S_{N_t} \leq t < S_{1+N_t}$. S_{N_t} et S_{1+N_t} sont les points de \mathcal{E} précédant et suivant directement l'instant t . Autrement dit, on peut les voir comme les instants de naissance et de décès de l'intervalle "vivant" à l'instant t . Notons $T'_t := t - S_{N_t}$ et $T''_t := S_{1+N_t} - t$ respectivement l'âge de cet intervalle vivant et le temps qu'il lui reste à vivre.

a) Montrer que pour $a, b > 0$ on a $\mathbb{P}(T'_t > a, T''_t > b) = e^{-\lambda(a+b)} 1_{\{t > a\}}$.

(Indication : utiliser l'exercice 11.2.a, après avoir partitionné suivant les valeurs de N_t ...)

b) Dédurre que T'_t et T''_t sont indépendants et donner leur lois.

c) Comparer la durée de vie moyenne de l'intervalle vivant à l'instant t avec celle du n -ième intervalle de \mathcal{E} ; explication ?

Exercice 11.7 On veut estimer λ par le maximum de vraisemblance $\hat{\lambda}$ dans les deux cas suivant. a) On observe jusqu'à l'instant $t > 0$ fixé, et on connaît donc N_t, S_1, \dots, S_{N_t} . Calculer $\hat{\lambda}$, et vérifier qu'il est sans biais. (Utiliser la densité sachant N_t .)

b) On observe jusqu'à l'instant S_n , pour n fixé, et on connaît donc S_1, \dots, S_n . Calculer $\hat{\lambda}$, et vérifier qu'il est biaisé.

II. Éléments de statistique mathématique

Le mot statistique désigne à la fois un ensemble de données et l'ensemble des activités consistant à collecter ces données, à les traiter et à les interpréter. Les statistiques interviennent donc dans pratiquement tous les domaines d'activité : gestion financière (états, banques, assurances, entreprises . . .), démographie, contrôles de qualité, études de marché, recherche médicale, autres sciences expérimentales . . .

Le traitement des données, pour en dégager un certain nombre de renseignements qualitatifs ou quantitatifs à des fins de comparaison, s'appelle la *statistique descriptive*. Elle sera l'objet des sections 12 et 13 ci-dessous.

L'estimation à l'aide des données d'un paramètre caché régissant le phénomène observé, pour déduire des observations, autant que possible, la nature précise de ce phénomène, par exemple à des fins de prédiction, s'appelle la *statistique inférentielle*. Elle sera l'objet des sections 14 et 15 ci-dessous.

L'ensemble sur lequel porte l'activité statistique s'appelle la *population*. Elle est généralement notée Ω . Ses éléments sont les *individus*. Ces individus peuvent être de natures très diverses : personnes, mois d'une année, pièces produites par une usine, résultats d'expériences répétées un certain nombre de fois . . . Les caractéristiques étudiées sur les individus d'une population sont appelées les *caractères*.

Un *caractère* est donc une application χ d'un ensemble fini Ω (la population) dans un ensemble E (*l'ensemble des valeurs du caractère*), qui associe à chaque individu $\omega \in \Omega$ la valeur $\chi(\omega)$ que prend ce caractère sur ω . La suite des valeurs $\chi(\omega)$ prises par χ constitue les *données brutes*. C'est une suite finie $(\chi_1, \chi_2, \dots, \chi_N) \in E^N$.

On peut considérer des caractères qualitatifs (comme : profession, adresse, situation de famille, sexe, etc . . .) et des caractères quantitatifs, dont la détermination produit un nombre ou une suite de nombres. Cela dit, les caractères qualitatifs peuvent toujours être transformés en caractères quantitatifs par simple codage. C'est ce qui se fait le plus souvent, quoiqu'un tel codage soit purement conventionnel et n'ait pas de vrai sens quantitatif.

12 Caractères quantitatifs monodimensionnels

Si χ est un caractère quantitatif monodimensionnel (c'est-à-dire à valeurs réelles ; on dit aussi *simple*) l'ensemble $\chi(\Omega) = \{\chi_1, \chi_2, \dots, \chi_N\}$ des valeurs atteintes par le caractère (ou données brutes) est un ensemble fini $\{x_1, \dots, x_n\}$. Nous supposons que ces valeurs sont ordonnées : $x_1 < x_2 < \dots < x_n$. Le fait que telle valeur soit relative à tel individu est un renseignement qui n'intéresse pas le statisticien. Seul l'ensemble des valeurs atteintes et le nombre de fois que chacune d'elle est atteinte est utile. On définit successivement les paramètres suivant :

- l'effectif de la valeur x_i est le nombre n_i de fois que la valeur x_i est prise par χ , c'est-à-dire le cardinal de l'ensemble $\chi^{-1}(\{x_i\})$;

- la fréquence de la valeur x_i est le rapport $f_i := \frac{n_i}{N}$ de l'effectif de x_i à l'effectif total $N = \sum_{i=1}^n n_i$ (ou cardinal) de la population ;
- l'effectif cumulé en x_i est la somme partielle cumulée $\sum_{j=1}^i n_j$ des effectifs ;
- la fréquence cumulée en x_i est la somme partielle cumulée $\sum_{j=1}^i f_j$ des fréquences.

Ces distributions statistiques sont qualifiées de *discrètes*. Lorsque le nombre des valeurs atteintes est important, on préfère regrouper les valeurs en classes pour rendre la statistique plus lisible. On partage alors l'ensemble C des valeurs du caractère en classes $]a_i, a_{i+1}]$ telles que $a_i < a_{i+1}$; les deux classes extrêmes pouvant s'étendre jusqu'à l'infini.

On parle alors de statistique groupée ou *continue*, et on nomme successivement :

- effectif de la classe $]a_i, a_{i+1}]$: le nombre n_i des valeurs prises par χ qui se trouvent dans la classe $]a_i, a_{i+1}]$, c'est-à-dire $n_i := \text{Card}[\chi^{-1}(]a_i, a_{i+1}])]$;
- effectif cumulé en a_i : le nombre de valeurs prises par χ qui se trouvent dans la classe $] - \infty, a_i]$;
- fréquence de la classe $]a_i, a_{i+1}]$: le rapport $f_i := \frac{n_i}{N}$;
- fréquence cumulée en a_i : la somme partielle cumulée $\sum_{j=1}^i f_j$ des fréquences.

La famille $(x_i, n_i)_{i=1, \dots, n}$ ou $(x_i, f_i)_{i=1, \dots, n}$ est encore appelée distribution statistique discrète. De même, la famille $(]a_i, a_{i+1}], n_i)_{i=1, \dots, n}$ ou $(]a_i, a_{i+1}], f_i)_{i=1, \dots, n}$ est encore appelée distribution statistique groupée (ou continue).

12.1 Représentations graphiques

12.1.1 Diagrammes en bâtons d'une distribution statistique discrète

On représente la distribution par une suite de segments verticaux d'abscisses x_i dont la longueur est proportionnelle à l'effectif ou la fréquence de x_i . Par exemple, la distribution

$$(1, 1); (2, 3); (3, 4); (4, 2); (5, 5); (6, 6); (7, 2); (8, 3); (9, 1); (10, 1)$$

est représentée par le diagramme en bâtons de la figure 1.

12.1.2 Polygone des fréquences d'une distribution statistique discrète

Le polygone des fréquences (resp. des effectifs) est obtenu à partir du diagramme en bâtons des fréquences (resp. des effectifs) en joignant par un segment les sommets des bâtons voisins. Le graphique de la figure 2 superpose le polygone des effectifs et le diagramme en bâtons des effectifs de l'exemple précédent.

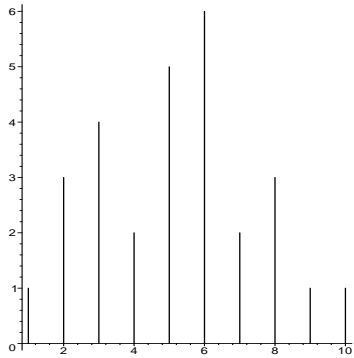


Figure 1: Diagramme en bâtons

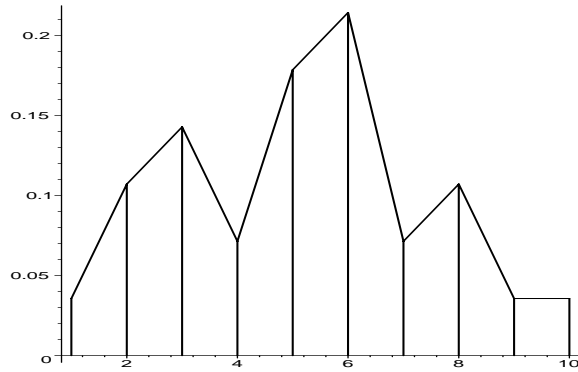


Figure 2: Diagramme en bâtons et polygone des fréquences

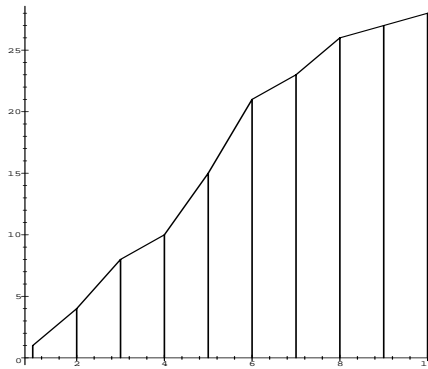


Figure 3: Diagramme en bâtons et polygone des effectifs cumulés

En remplaçant les fréquences (resp. les effectifs) par les fréquences cumulées (resp. les effectifs cumulés) on obtient le diagramme en bâtons et le polygone des fréquences cumulées (resp. des effectifs cumulés). La figure 3 donne le diagramme en bâtons et le polygone des effectifs cumulés de l'exemple précédent.

12.1.3 Histogramme d'une distribution statistique groupée

On représente la distribution par des rectangles dont les cotés horizontaux sont les classes et dont les aires sont proportionnelles aux effectifs ou aux fréquences.

En figure 4 nous donnons l'histogramme de la distribution suivante

$$([1, 3], 4); ([3, 4], 8); ([4, \frac{11}{2}], 10); ([\frac{11}{2}, 6], 14); ([6, 8], 20); ([8, 10], 12); ([10, 11], 9); ([11, \frac{25}{2}], 3).$$

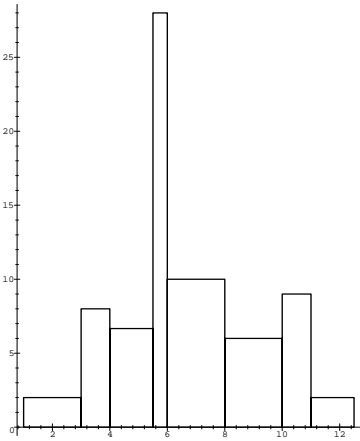


Figure 4: Histogramme

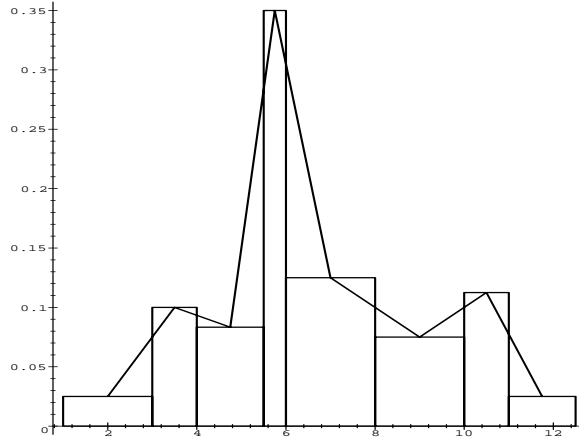


Figure 5: Histogramme et polygone des fréquences

12.1.4 Polygone des effectifs et des fréquences d'une distribution groupée

On obtient ce polygone en joignant dans l'histogramme de cette distribution les milieux des côtés horizontaux supérieurs adjacents. La figure 5 superpose l'histogramme des fréquences de l'exemple précédent et son polygone des fréquences.

12.1.5 Polygone des fréquences cumulées d'une distribution statistique groupée

C'est la représentation graphique de la fonction définie par

$$x \mapsto \sum_{j=1}^{i-1} f_j + \frac{x - a_i}{a_{i+1} - a_i} f_i$$

sur l'intervalle $]a_i, a_{i+1}[$. En particulier $a_i \mapsto \sum_{j=1}^{i-1} f_j$ (ou bien $a_{i+1} \mapsto \sum_{j=1}^i f_j$), et le polygone est déduit de ces valeurs particulières par simple interpolation affine. La fonction ainsi définie est proportionnelle à la primitive nulle en a_0 de la fonction en escalier représentée par l'histogramme de la figure 4. Pour l'exemple précédent nous obtenons le graphique de la figure 6.

12.2 Caractéristiques de position

12.2.1 Valeurs modales

On nomme valeur modale ou mode d'une distribution statistique discrète $(x_i, f_i)_{i=1, \dots, p}$ toute valeur x_1, x_2, \dots, x_p dont la fréquence est maximale. La distribution est unimodale

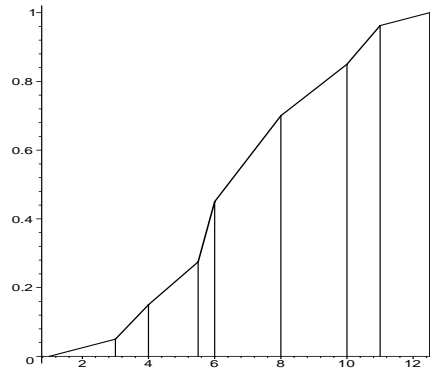


Figure 6: Polygone des fréquences cumulées d'une statistique groupée

si elle a un seul mode, plurimodale (bimodale, trimodale, ...) sinon. On visualise aisément les modes sur les représentations graphiques. Pour une distribution statistique groupée, on nomme de manière analogue classe modale toute classe de fréquence maximale.

12.2.2 Médiane et quantiles

La médiane d'une distribution statistique discrète $(x_i, f_i)_{i=1, \dots, p}$ est la plus petite des valeurs x_i pour laquelle la fréquence cumulée est supérieure ou égale à 0,5. On la note $Q_{0.5}$. On peut la déterminer aisément à partir du diagramme en bâtons des fréquences cumulées. En gros, la moitié de la population présente une valeur du caractère inférieure à la médiane, l'autre moitié une valeur supérieure.

Pour une distribution statistique groupée, la médiane est définie comme le nombre réel x pour lequel le polygone des fréquences cumulées atteint la valeur 0,5.

Plus généralement, si α est un nombre réel strictement compris entre 0 et 1, le quantile d'ordre α , noté Q_α , est la plus petite des valeurs x_i pour laquelle la fréquence cumulée est supérieure à α . En dehors de la médiane, les quantiles les plus souvent utilisés sont les quartiles ($\alpha = 0,25; 0,5; 0,75$) et les déciles ($\alpha = 0,1; 0,2; \dots; 0,9$).

12.2.3 Moyenne

La moyenne d'une distribution statistique discrète $\chi = (x_i, f_i)_{i=1, \dots, p}$ est le nombre réel m défini par

$$m = m(\chi) := \frac{1}{N} \sum_{i=1}^p n_i x_i = \sum_{i=1}^p f_i x_i.$$

où N est l'effectif total de la population. On peut aussi la calculer directement à partir des données brutes par

$$m(\chi) = \frac{1}{N} \sum_{j=1}^N \chi_j,$$

c'est-à-dire en calculant le rapport entre la somme de toutes les valeurs relevées (avec répétitions éventuelles) et l'effectif total de la population.

Pour une statistique groupée $(]a_i, a_{i+1}], f_i)_{i=1, \dots, p}$ la moyenne se calcule par

$$m := \sum_{i=1}^p f_i \frac{a_i + a_{i+1}}{2}.$$

Cela revient à faire une hypothèse d'homogénéité en considérant les valeurs équidistribuées à l'intérieur d'une classe ou, au contraire, à supposer que toute la fréquence est concentrée au centre de la classe (ce qui revient au même : on remplace la distribution à l'intérieur de la classe par son barycentre).

12.2.4 Caractéristiques de dispersion

Ce sont des paramètres qui mesurent de différentes façons la dispersion d'une distribution statistique discrète $\chi = (x_i, f_i)_{i=1, \dots, p}$.

On nomme successivement :

- étendue la différence $(x_p - x_1)$ entre la plus grande et la plus petite des valeurs prises ;
- intervalle interquartile la différence $(Q_{0,75} - Q_{0,25})$;
- variance la quantité $\text{Var}(\chi) := \sum_{i=1}^p f_i (x_i - m)^2 = \sum_{i=1}^p f_i x_i^2 - m^2$;
- écart-type la quantité $\sigma(\chi) := \sqrt{\text{Var}(\chi)}$;
- moment d'ordre k (pour $k \in \mathbb{N}^*$) la quantité $m_k(\chi) := \sum_{i=1}^p f_i x_i^k = \frac{1}{N} \sum_{j=1}^N \chi_j^k$;
- moment centré d'ordre k (pour $k \in \mathbb{N}^*$) la quantité $m'_k(\chi) := \sum_{i=1}^p f_i (x_i - m)^k$.

La variance est ainsi le moment centré d'ordre 2. On la note souvent σ^2 .

Remarque 12.2.5 On a les relations plus ou moins immédiates (exercice !) suivantes, valables pour tous réels a, λ , et totalement analogues avec ce qui se passe dans le cadre probabiliste :

$$m(\lambda(\chi - a)) = \lambda(m(\chi) - a);$$

$$\text{Var}(\lambda(\chi - a)) = \lambda^2 \text{Var}(\chi); \quad \sigma(\lambda(\chi - a)) = |\lambda| \sigma(\chi).$$

Remarque 12.2.6 (*Inégalité de Bienaymé-Tchébychev*) $\sum_{\{i \mid |x_i - m| \geq \varepsilon\}} f_i \leq \sigma^2 / \varepsilon^2$.

En particulier pour $\varepsilon = 2\sigma$, on a $\sum_{i \in \{i \mid |x_i - m| \geq \varepsilon\}} f_i \leq 1/4$, ce qui montre que les 3/4 de l'effectif global sont contenus dans l'intervalle $[m - 2\sigma, m + 2\sigma]$. Ainsi, plus σ est petit, plus la distribution est concentrée autour de sa moyenne.

Preuve Sur l'ensemble $\{i \mid |x_i - m| \geq \varepsilon\}$, le nombre positif $(x_i - m)^2$ est minoré par ε^2 ; on a donc

$$\sigma^2 = \sum_{i=1}^p (x_i - m)^2 f_i \geq \sum_{\{i \mid |x_i - m| \geq \varepsilon\}} (x_i - m)^2 f_i \geq \varepsilon^2 \sum_{\{i \mid |x_i - m| \geq \varepsilon\}} f_i,$$

d'où l'inégalité cherchée en divisant par ε^2 . \diamond

Remarque 12.2.7 Pour les distributions statistiques groupées, on définit les mêmes notions en supposant que toutes les valeurs d'une classe sont concentrées au centre de la classe :

$$\text{Var} := \sum_{i=1}^p f_i \left(\frac{a_{i+1} + a_i}{2} - m \right)^2; \quad \sigma = \sqrt{\text{Var}}; \quad m_k := \sum_{i=1}^p f_i \left(\frac{a_{i+1} + a_i}{2} \right)^k.$$

En remplaçant x_i par $\frac{a_{i+1} + a_i}{2}$, les remarques 12.2.5 et 12.2.6 restent valables pour ces distributions.

13 Régression linéaire

13.1 Problématique

Etant donnée une statistique double ou statistique bivariée (de données réelles) $(X = (X_1, \dots, X_n), Y = (Y_1, \dots, Y_n))$, on cherche une relation linéaire : $Y = aX + b$. Bien entendu elle ne peut être exacte sauf cas exceptionnel. Aussi cherche-t-on les coefficients (a, b) de façon que cette relation soit le plus près possible d'être vérifiée. La notion de proximité qu'on retient est celle qui conduit au calcul le plus simple : on cherche à minimiser la somme des carrés. (a, b) sera donc le couple de réels minimisant $\sum_{j=1}^n (Y_j - aX_j - b)^2$. La droite $D_{Y/X}$ d'équation $Y = aX + b$ correspondante est la droite de régression au sens des moindres carrés (ou droite de régression) de Y par rapport à X .

13.2 Moments

Exactement analogues aux quantités intervenant dans le cadre probabiliste sont :

les moyennes empiriques $\bar{X} := \frac{1}{n} \sum_{j=1}^n X_j$, $\bar{Y} := \frac{1}{n} \sum_{j=1}^n Y_j$, puis les variances empiriques $V_x := \frac{1}{n} \sum_{j=1}^n X_j^2 - \bar{X}^2$, $V_y := \frac{1}{n} \sum_{j=1}^n Y_j^2 - \bar{Y}^2$, et la covariance empirique $\kappa := \frac{1}{n} \sum_{j=1}^n X_j Y_j - \bar{X} \bar{Y}$. Les écarts-type sont $\sigma_x := \sqrt{V_x}$ et $\sigma_y := \sqrt{V_y}$, et le coefficient de corrélation linéaire est $\rho := \kappa / (\sigma_x \sigma_y)$.

13.3 Solution

Nous avons simplement :

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n (Y_j - aX_j - b)^2 &= \frac{1}{n} \sum_{j=1}^n \left((Y_j - \bar{Y}) - a(X_j - \bar{X}) + (\bar{Y} - a\bar{X} - b) \right)^2 \\ &= (\bar{Y} - a\bar{X} - b)^2 + V_x a^2 - 2\kappa a + V_y, \end{aligned}$$

de sorte qu'apparaît aussitôt (pour peu que la statistique X ne soit pas constante) la solution unique suivante :

$$a = \kappa / V_x, \quad b = \bar{Y} - a\bar{X}.$$

L'erreur quadratique totale vaut alors

$$\sigma^2 := \frac{1}{n} \sum_{j=1}^n (Y_j - aX_j - b)^2 = V_y - \kappa^2 / V_x = (1 - \rho^2) V_y.$$

Elle est nulle lorsque qu'existe une relation linéaire entre les statistiques X et Y , et faible en cas de relation quasi-linéaire. Notons que la droite de régression passe par le centre de gravité (\bar{X}, \bar{Y}) du nuage formé par les n points (X_j, Y_j) .

De même qu'on a la droite $D_{Y/X}$ de régression de Y par rapport à X , on a la droite $D_{X/Y}$ de régression de X par rapport à Y , d'équation $X = a'Y + b'$, avec

$$a' = \kappa / V_y, \quad b' = \bar{X} - a'\bar{Y}.$$

Les deux droites $D_{Y/X}$ et $D_{X/Y}$ passant toutes deux par le centre de gravité (\bar{X}, \bar{Y}) du nuage, elles sont confondues si et seulement si $aa' = 1$, id est si et seulement si $\kappa^2 = V_x V_y$, ou encore si et seulement si $\rho = \pm 1$, et d'autant plus proches l'une de l'autre que $|\rho|$ est proche de 1. Cela explique la dénomination de ρ .

13.4 Utilisation

Il est naturel d'utiliser la droite $D_{Y/X}$ pour prédire une valeur supplémentaire Y_{n+1} de la statistique Y , connaissant une valeur supplémentaire X_{n+1} de la statistique X : $Y_{n+1} = aX_{n+1} + b$.

La qualité d'une telle prédiction (qui a priori n'a vraiment de sens que pour X_{n+1} proche de l'intervalle $[\min X, \max X]$) dépend de la valeur de $\sigma^2 := V_y - \kappa^2/V_x$, qui est la variance empirique de la statistique $Y - aX - b$. On peut déterminer une bande de confiance (on parle plutôt d'intervalle de confiance, voir la section 14.5 ci-dessous) par sa largeur autour de la droite de régression, de la façon suivante : pour $p \in]0, 1[$ fixé, soit r un réel > 0 tel que $p \approx \frac{1}{n} \sum_{j=1}^n 1_{\{|Y_j - aX_j - b| \leq r\}}$. Pensant au théorème central limite, il est naturel d'approcher la loi empirique des $(Y_j - aX_j - b)$ par une gaussienne, nécessairement centrée et de variance σ^2 . De sorte qu'on doit avoir $p \approx \mathbb{P}(|G| \leq r/\sigma)$, pour G gaussienne centrée réduite.

On peut ainsi estimer que pour X_{n+1} proche de l'intervalle $[\min X, \max X]$, on doit avoir par exemple 95% de chance de trouver Y_{n+1} dans l'intervalle (de confiance à 95%)

$$[aX_{n+1} + b - (1,96)\sigma, aX_{n+1} + b + (1,96)\sigma].$$

13.5 Autres ajustements

Bien entendu, on ne cherche pas toujours une relation linéaire entre les statistiques X et Y . Mais on peut bien souvent s'y ramener, par un changement de variable élémentaire, dès qu'on cherche une relation simple comportant 2 paramètres réels.

Exemples : si l'on espère que $Y \approx aX^b$, il suffit de considérer le nuage formé (ou la statistique double formée) par les $(\log X_j, \log Y_j)$;

si l'on espère que $Y \approx \log(ae^{X^2} + b)$, il suffit de considérer le nuage formé par les (e^{X_j}, e^{Y_j}) ;

si l'on soupçonne que $Y \approx \frac{aX}{X^2+b}$, il suffit de considérer le nuage formé (ou la statistique double formée) par les $(X_j^2, X_j/Y_j)$.

Si l'on recherche une relation fonctionnelle comportant plus de deux paramètres, on peut de nouveau minimiser la somme des carrés des différences (comme avec 2 paramètres), cette fois en commençant par annuler les dérivées partielles. Par exemple dans le cas

d'une relation polynômiale $Y \approx \sum_{k=0}^d \alpha_k X^k$: on cherchera la (les) valeur(s) du paramètre

$$(\alpha_0, \dots, \alpha_d) \in \mathbb{R}^{d+1} \text{ qui minimise(nt) } \delta = \delta(\alpha_0, \dots, \alpha_d) := \sum_{j=1}^n \left[Y_j - \sum_{k=0}^d \alpha_k X_j^k \right]^2.$$

Exercice n° 13.1 Soient $X = (1, 3, 5, 7, 8, 9, 11, 12, 15, 18)$ et

$Y = (-33, -97, -134, -141, -130, -102, -32, 37, 343, 860)$. Représenter le nuage, et tester s'il y a une relation $Y = \alpha X^3 + \beta X$, ou $Y = \text{ch}(aX - b) - 150$, ou $Y = \alpha(X - \beta)^2 - 150$.

Déterminer un intervalle de confiance à 95% pour la valeur y associée à $x = 14$.

13.6 Variations expliquée et inexpliquée

La régression (on dit aussi l'ajustement) de Y par rapport à X rend compte de ce qui dans les variations de la série Y est explicable par la série X , qu'on voit ainsi comme un facteur agissant (plus ou moins fortement) sur Y .

Ayant déterminé la droite de régression $D_{Y/X}$, il est naturel de comparer la série Y avec son explication par X , à savoir : $\hat{Y} := aX + b = (\hat{Y}_1 = aX_1 + b, \dots, \hat{Y}_n = aX_n + b)$.

D'une part, selon le calcul du paragraphe 13.3, nous avons :

$$\frac{1}{n} \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 = (1 - \varrho^2)V_y;$$

et d'autre part :

$$\frac{1}{n} \sum_{j=1}^n (\hat{Y}_j - \bar{Y})^2 = \frac{1}{n} \sum_{j=1}^n (a[X_j - \bar{X}])^2 = a^2 V_x = \varrho^2 V_y.$$

Répartissant la variance de Y , qu'on peut voir comme sa variation totale, en

$V_y = (1 - \varrho^2)V_y + \varrho^2 V_y$, nous déduisons :

$$\frac{1}{n} \sum_{j=1}^n (Y_j - \bar{Y})^2 = \frac{1}{n} \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 + \frac{1}{n} \sum_{j=1}^n (\hat{Y}_j - \bar{Y})^2,$$

relation de Pythagore que nous interprétons comme :

Variation totale de Y = variation de Y inexpliquée + variation de Y expliquée par X .

13.7 Distribution conjointe

La donnée du paragraphe 13.1, sous forme de statistique double (ou bivariée) brute $(X = (X_1, \dots, X_n), Y = (Y_1, \dots, Y_n))$ peut aussi bien être présentée sous la forme alternative de nuage de points pondérés : $((x_i, y_j); n_{ij})_{1 \leq i \leq p, 1 \leq j \leq q}$, où n_{ij} figure l'effectif du point de coordonnées (x_i, y_j) , id est le nombre de fois qu'il figure dans la statistique double considérée (X, Y) .

Noter que l'on peut (ou non) considérer que (x_1, \dots, x_p) est la liste strictement croissante des différentes valeurs de X , et que (y_1, \dots, y_q) est la liste strictement croissante des différentes valeurs de Y , quitte à admettre des effectif éventuellement nuls. Mais cela n'importe guère.

L'effectif total est alors bien sûr $n = \sum_{i,j} n_{ij}$. La fréquence du point de coordonnées (x_i, y_j) est $f_{ij} := n_{ij}/n$.

Les distributions marginales, et fréquences marginales, exactement analogues aux lois marginales probabilistes, sont données par :

$$n_{i.} := \sum_{j=1}^q n_{ij}; \quad n_{.j} := \sum_{i=1}^p n_{ij}; \quad f_{i.} := \sum_{j=1}^q f_{ij} = n_{i.}/n; \quad f_{.j} := \sum_{i=1}^p f_{ij} = n_{.j}/n.$$

Notons que l'indépendance probabiliste trouve ici son pendant dans la relation suivante, qu'on peut nommer "indépendance statistique" : $f_{ij} = f_{i.} \times f_{.j}$ ($\forall 1 \leq i \leq p, 1 \leq j \leq q$).

Les moments du paragraphe 13.2 se calculent évidemment aussi bien sous cette forme de donnée que sous la forme précédente. Nous avons en effet :

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^p n_{i.} x_i = \frac{1}{n} \sum_{ij} n_{ij} x_i; & \bar{Y} &= \frac{1}{n} \sum_{j=1}^q n_{.j} y_j = \frac{1}{n} \sum_{ij} n_{ij} y_j; \\ V_x &= \frac{1}{n} \sum_{i=1}^p n_{i.} x_i^2 - \bar{X}^2; & V_y &= \frac{1}{n} \sum_{j=1}^q n_{.j} y_j^2 - \bar{Y}^2; & \kappa &= \frac{1}{n} \sum_{ij} n_{ij} x_i y_j - \bar{X} \bar{Y}. \end{aligned}$$

Exercice n° 13.2 Vérifier les formules ci-dessus, et en donner d'autres formes alternatives.

14 Vraisemblance et estimation

14.1 Maximum de vraisemblance

Il s'agit d'estimer un paramètre inconnu θ d'une loi de probabilité $\mu = \mathbb{P}_\theta$, sur la base d'une suite de réalisations, ou d'observations, effectuées suivant cette loi. Le paramètre θ varie a priori dans un certain ensemble $\Theta \subset \mathbb{R}^d$ (dont le choix peut appartenir au statisticien).

Notons X une v.a. dont la loi est μ , et x_1, \dots, x_n une suite d'observations, nommée échantillon ou statistique (ou échantillon statistique), représentant des réalisations de X , qu'on suppose indépendantes. On considère les 2 cas les plus usuels : soit μ est discrète, soit elle admet une densité $g = g_\theta$. La fonction de vraisemblance associée à θ et à x_1, \dots, x_n est :

$$f(x_1, \dots, x_n | \theta) := \prod_{j=1}^n \mathbb{P}_\theta(X = x_j) = \prod_{j=1}^n \mathbb{P}_\theta(\{x_j\}) \quad \text{lorsque } \mu \text{ est discrète, et}$$

$$f(x_1, \dots, x_n | \theta) := \prod_{j=1}^n g_\theta(x_j) \quad \text{lorsque } \mathbb{P}_\theta \text{ admet la densité } g_\theta.$$

L'estimateur du maximum de vraisemblance est par définition la valeur (supposée unique) de θ maximisant la fonction de vraisemblance :

$$\hat{\theta} := \operatorname{argmax} f(x_1, \dots, x_n | \theta) = \operatorname{argmax} \log f(x_1, \dots, x_n | \theta).$$

Exemples 1) μ est exponentielle de moyenne θ , id est $\mathcal{E}(1/\theta)$: alors

$\log f(x_1, \dots, x_n | \theta) = -n \log \theta - (x_1 + \dots + x_n)/\theta$, de sorte que $\hat{\theta} = (x_1 + \dots + x_n)/n$.

2) $\mu = \mathcal{B}(m, p)$, avec m supposé connu et $\theta = p$: alors

$\log f(x_1, \dots, x_n | p) = \sum_{j=1}^n \log C_m^{x_j} + \sum_{j=1}^n x_j \log p + (mn - \sum_{j=1}^n x_j) \log(1 - p)$, de sorte que $\hat{p} = (x_1 + \dots + x_n)/(mn)$.

3) $\mu = \mathcal{P}(\lambda)$: alors $\log f(x_1, \dots, x_n | \lambda) = \sum_{j=1}^n x_j \log \lambda - n\lambda - \log(x_1! \dots x_n!)$, de sorte que $\hat{\lambda} = (x_1 + \dots + x_n)/n$.

4) $\mu = \mathcal{N}(m, \sigma^2)$, avec $\theta = (m, \sigma)$: alors

$\log f(x_1, \dots, x_n | m, \sigma) = -(n/2) \log(2\pi) - n \log \sigma - \sum_{j=1}^n (x_j - m)^2 / (2\sigma^2)$, de sorte que

$\hat{m} = \sum_{j=1}^n x_j / n$ et $\hat{\sigma} = \left(\sum_{j=1}^n (x_j - \hat{m})^2 / n \right)^{1/2}$ (ce sont les seules valeurs qui annulent à la fois les deux dérivées partielles de la fonction de vraisemblance).

5) $\mu = \mathcal{H}(N, n, p)$, avec n connu, N très grand, et $p \in]0, 1[$ à estimer, à l'aide d'un unique tirage, qui produit la valeur k de la variable. (Exemple: pour évaluer la population piscicole d'un lac, on remet ℓ poissons dans ce lac après les avoir marqués, de sorte que $p = \ell/N$, et on pêche ensuite n poissons.) La valeur optimale p doit résoudre l'équation:

$$\begin{aligned} 0 &= \frac{1}{N} \frac{d}{dp} \log \left[C_{Np}^k C_{N(1-p)}^{n-k} \right] = \frac{1}{N} \sum_{j=0}^{k-1} \frac{1}{p - \frac{j}{N}} - \frac{1}{N} \sum_{j=0}^{n-k-1} \frac{1}{1 - p - \frac{j}{N}} \\ &\approx \int_0^{\frac{k}{N}} \frac{dt}{p - t} - \int_0^{\frac{n-k}{N}} \frac{dt}{1 - p - t} = \log \left[\frac{p}{p - \frac{k}{N}} \right] - \log \left[\frac{1 - p}{1 - p - \frac{n-k}{N}} \right], \end{aligned}$$

dont la nullité équivaut à:

$$\frac{p}{p - \frac{k}{N}} = \frac{1 - p}{1 - p - \frac{n-k}{N}} \Leftrightarrow p = \frac{k}{n}.$$

Notons que c'est la valeur qu'on obtient trivialement en imposant, au lieu du maximum de vraisemblance, simplement que la moyenne np soit égale à la valeur k obtenue au tirage.

Exercice n° 14.1.1 Supposons qu'un processus de Poisson de paramètre inconnu λ est observé: a) jusqu'à un temps fixe T ; b) jusqu'à la survenue du n -ième événement. Dans chacun de ces deux cas, quel est l'estimateur du maximum de vraisemblance pour λ ?

14.2 Biais, risque

Un estimateur $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ est dit sans biais lorsque sa moyenne pour la loi \mathbb{P}_θ est θ , cela pour tout $\theta \in \Theta$:

$$\mathbb{E}[\hat{\theta}(X_1, \dots, X_n) | \theta] \equiv \int \hat{\theta}(x_1, \dots, x_n) d\mathbb{P}_\theta(x_1) \dots d\mathbb{P}_\theta(x_n) = \theta, \quad (\forall \theta \in \Theta).$$

La qualité d'un estimateur est évaluée par son risque (quadratique) :

$$R(\hat{\theta}, \theta) := \mathbb{E} \left[[\hat{\theta}(X_1, \dots, X_n) - \theta]^2 \middle| \theta \right],$$

qui est égal à sa variance s'il n'est pas biaisé. Il n'y a généralement pas d'estimateur dont le risque soit minimal uniformément (c'est-à-dire pour tous les θ dans Θ), même parmi les estimateurs sans biais. Il peut y avoir des estimateurs biaisés meilleurs (c'est-à-dire de risque moindre pour tout $\theta \in \Theta$) que les estimateurs sans biais.

Exemples 1) Soit A une sous-population en proportion p dans une population Ω . L'estimateur "proportion empirique" $\hat{p} = \hat{p}(x_1, \dots, x_n) := \frac{1}{n} \sum_{j=1}^n 1_A(x_j)$ de p est sans biais, de risque $R(\hat{p}, p) = p(1-p)/n$. Ici $\mathbb{P}_p = \mathcal{B}(p)$.

2) $\mathbb{P}_\theta = \mathcal{E}(1/\theta)$, comme dans l'exemple 1 de la section 14.1 ci-dessus : notons $\hat{\theta}_1(x) := x_1$, qui est sans biais, de même que l'estimateur $\hat{\theta} = (x_1 + \dots + x_n)/n$ du maximum de vraisemblance. On a aussitôt : $R(\hat{\theta}_1, \theta) = \theta^2 > R(\hat{\theta}, \theta) = \theta^2/n$.

De même, si on considère une moyenne quelconque : $\hat{\theta}_\lambda = \lambda_1 x_1 + \dots + \lambda_n x_n$, avec $\lambda_1, \dots, \lambda_n \geq 0$ tels que $\lambda_1 + \dots + \lambda_n = 1$, on trouve encore un estimateur sans biais, de risque $R(\hat{\theta}_\lambda, \theta) = \lambda_1^2 + \dots + \lambda_n^2$, qui est minimal pour $\hat{\theta}_\lambda = \hat{\theta}$.

On a plus généralement le résultat suivant, valable dès que les v.a.r.i.i.d. X_j sont de carré intégrable.

Proposition 14.2.1 1) Parmi tous les estimateurs sans biais de $\mathbb{E}(X_1)$ qui sont des fonctions linéaires de (X_1, \dots, X_n) , la moyenne empirique $\bar{X}_n := \frac{1}{n} \sum_{j=1}^n X_j$ est de risque minimal.

2) Parmi tous les estimateurs sans biais de $\text{Var}(X_1)$ qui sont des fonctions quadratiques de (X_1, \dots, X_n) , la déviatoin standard $V_n := \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2$ est de risque minimal.

Preuve 1) Un estimateur $\hat{\theta} = h(X_1, \dots, X_n) = \sum_{j=1}^n \lambda_j X_j$ de $m = \mathbb{E}(X_1)$ est sans biais ssi $\sum_{j=1}^n \lambda_j = 1$. Notons $\bar{\theta} := \bar{h}(X_1, \dots, X_n) := \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} h(X_{\pi_1}, \dots, X_{\pi_n})$ sa fonction symétrisée, qui ici vaut clairement \bar{X}_n . La loi de (X_1, \dots, X_n) étant invariante par les permutations $\pi \in \mathcal{S}_n$, nous avons pour toute fonction symétrique $Y_n = \bar{f}(X_1, \dots, X_n)$:

$$\mathbb{E}[\hat{\theta} Y_n] = \int h(X_1, \dots, X_n) Y_n d\mathbb{P} = \int h(X_{\pi_1}, \dots, X_{\pi_n}) Y_n d\mathbb{P} = \int \bar{h}(X_1, \dots, X_n) Y_n d\mathbb{P} = \mathbb{E}[\bar{\theta} Y_n].$$

Cela entraîne $\mathbb{E}[\hat{\theta}] = \mathbb{E}[\bar{\theta}]$ et $\mathbb{E}[\hat{\theta}\bar{\theta}] = \mathbb{E}[\bar{\theta}^2]$, et donc :

$$\mathbb{E}[(m - \hat{\theta})^2] - \mathbb{E}[(m - \bar{\theta})^2] = \mathbb{E}[\hat{\theta}^2] - \mathbb{E}[\bar{\theta}^2] = \mathbb{E}[(\hat{\theta} - \bar{\theta})^2] + 2 \mathbb{E}[(\hat{\theta} - \bar{\theta}) \bar{\theta}] = \mathbb{E}[(\hat{\theta} - \bar{\theta})^2],$$

d'où en effet $\mathbb{E}[(\mathbb{E}(X_1) - \hat{\theta})^2] \geq \mathbb{E}[(\mathbb{E}(X_1) - \bar{\theta})^2] = \mathbb{E}[(\mathbb{E}(X_1) - \bar{X}_n)^2]$.

2) De même, un estimateur $\hat{\theta} = g(X_1, \dots, X_n) = \sum_{1 \leq j, k \leq n} \lambda_{jk} X_j X_k$ de $v = \text{Var}(X_1)$ est sans biais ssi $\sum_j \lambda_{jj} = 1 = - \sum_{j \neq k} \lambda_{jk}$, puisque $\mathbb{E}[\hat{\theta}] = v \sum_j \lambda_{jj} + m^2 \left(\sum_j \lambda_{jj} + \sum_{j \neq k} \lambda_{jk} \right)$. Notons de nouveau $\bar{\theta} := \bar{g}(X_1, \dots, X_n)$ sa fonction symétrisée, qui est une forme quadratique dont les coefficients doivent vérifier les 2 équations ci-dessus, et être symétriques, donc égaux sur la diagonale, et égaux hors de la diagonale. Donc $\bar{\theta} = \frac{1}{n} \sum_j X_j^2 - \frac{2}{n(n-1)} \sum_{j < k} X_j X_k$.

Or on obtient la même formule en développant V_n , de sorte que $\bar{\theta} = V_n$. La fin de la preuve est exactement la même qu'en (1) ci-dessus pour l'espérance. \diamond

Bien entendu, cet énoncé ne traite pas tous les cas.

Exemple Pour $\mu = \mathcal{N}(m, \sigma^2)$, la médiane \tilde{m}_n de l'échantillon (ou médiane empirique) est un estimateur sans biais de la moyenne, mais non linéaire. Un calcul délicat fournit $\text{Var}(\tilde{m}_n) \approx 1,57 \times \sigma^2/n = 1,57 \text{Var}(\bar{m}_n)$, ce qui montre que dans ce cas la moyenne empirique est meilleure que la médiane empirique. Le théorème 14.4.2 de Rao-Blackwell (voir la section 14.4 ci-dessous) montre que dans ce cas la moyenne empirique est meilleure que tout estimateur sans biais.

Définition 14.2.2 *Un estimateur $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ de θ est dit convergent lorsque $\lim_{n \rightarrow \infty} \hat{\theta}(X_1, \dots, X_n) = \theta$, cela en probabilité relativement à \mathbb{P}_θ , pour tout $\theta \in \Theta$.*

Exemples Les estimateurs des exemples de la section 14.1 sont convergents.

Exercice n° 14.2.1 Soient X_1, \dots, X_n, \dots des v.a.r.i.i.d. de loi $\mathcal{B}(p)$.

- Si (pour n fixé) $X_1 + \dots + X_n = k$, quel est l'estimateur du maximum de vraisemblance pour p ? Est-il sans biais? Converge-t-il?
- Supposons qu'on s'arrête au premier n tel que $X_1 + \dots + X_n = k$, pour k fixé; quel est alors l'estimateur du maximum de vraisemblance pour p ? Est-il sans biais?

14.3 Cas gaussien

Dans le cas $\mu = \mu_{(m, \sigma)} = \mathcal{N}(m, \sigma^2)$, il n'est pas très difficile de pousser les calculs plus loin. On peut établir ainsi en particulier :

- Proposition 14.3.1** (i) *La moyenne empirique \bar{X}_n a pour loi $\mathcal{N}(m, \sigma^2/n)$;*
(ii) $\sigma^{-2} \sum_{j=1}^n (X_j - \bar{X}_n)^2$ ($= (n-1)\sigma^{-2}$ fois la déviation standard V_n de la proposition 14.2.1) *suit la loi χ_{n-1}^2 (du χ^2 à $(n-1)$ degrés de liberté, voir l'exercice n° 5.6);*
(iii) *La moyenne empirique et la déviation standard sont indépendantes.*

Le corollaire suivant est immédiat, et utile, puisqu'il met en évidence une loi indépendante de l'écart-type σ (souvent inconnu dans la pratique).

Corollaire 14.3.2 La variable $Z_{n-1} := (\bar{X}_n - m)\sqrt{n/V_n}$ suit la loi dite de Student à $(n-1)$ degrés de liberté, qui est par définition la loi du rapport d'une gaussienne standard et de $\sqrt{\chi_{n-1}^2/(n-1)}$, où χ_{n-1}^2 désigne une v.a. ayant cette loi, indépendante de la gaussienne.

14.4 Statistique exhaustive

Définition 14.4.1 Une statistique exhaustive des v.a.i.i.d. X_1, \dots, X_n est une fonction (déterministe et indépendante de θ) $\tau_n = \tau(X_1, \dots, X_n)$ qui permet de factoriser la fonction de vraisemblance sous la forme : $f(X_1, \dots, X_n | \theta) = h_n(\tau_n, \theta) \times k_n(X_1, \dots, X_n)$, pour une fonction k_n indépendante de θ .

Dans les exemples 1,2,3 ci-dessus, $\tau(X_1, \dots, X_n) := X_1 + \dots + X_n$ est une statistique exhaustive. Dans l'exemple 4 ci-dessus,

$\tau(X_1, \dots, X_n) = ((X_1 + \dots + X_n), (X_1^2 + \dots + X_n^2))$ est une statistique exhaustive ; donc $(\bar{X}, V) := (\frac{X_1 + \dots + X_n}{n}, (X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2)$ en est une autre.

Théorème 14.4.2 (Rao-Blackwell) Si une statistique exhaustive $\tau = \tau(X_1, \dots, X_n)$ existe et si $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ est un estimateur sans biais de θ , alors il existe une certaine fonction de τ qui est un estimateur sans biais de θ , de risque $\leq R(\hat{\theta}, \theta)$. (Cette fonction est égale à l'espérance conditionnelle : $\mathbb{E}[\hat{\theta} | \tau]$, relativement à $k_n(X_1, \dots, X_n) dX_1 \dots dX_n$.)

Preuve (dans le cas discret) Puisque $\hat{\theta}$ est sans biais, nous avons :

$$\begin{aligned} \theta &= \sum_x \hat{\theta}(x) h_n(\tau(x), \theta) k_n(x) = \sum_t \sum_{x \in \tau^{-1}(t)} \hat{\theta}(x) k_n(x) h_n(t, \theta) \\ &= \sum_t \left[\frac{\sum_{x \in \tau^{-1}(t)} \hat{\theta}(x) k_n(x)}{\sum_{x \in \tau^{-1}(t)} k_n(x)} \right] \times \left[\sum_{x \in \tau^{-1}(t)} k_n(x) \right] h_n(t, \theta) = \sum_t g(t) \sum_{x \in \tau^{-1}(t)} k_n(x) h_n(t, \theta) \\ &= \sum_x g(\tau(x)) h_n(\tau(x), \theta) k_n(x), \end{aligned}$$

avec $g(t) := \frac{\sum_{x \in \tau^{-1}(t)} \hat{\theta}(x) k_n(x)}{\sum_{x \in \tau^{-1}(t)} k_n(x)}$, c'est-à-dire $g \circ \tau = \mathbb{E}[\hat{\theta} | \tau]$; l'égalité ci-dessus montre que $\mathbb{E}[\hat{\theta} | \tau]$ est également sans biais. Ensuite, nous avons par inégalité de Schwarz :

$$\left[\sum_{x \in \tau^{-1}(t)} \hat{\theta}(x) k_n(x) \right]^2 \leq \sum_{x \in \tau^{-1}(t)} \hat{\theta}(x)^2 k_n(x) \times \sum_{x \in \tau^{-1}(t)} k_n(x),$$

et donc $g(t)^2 \leq \frac{\sum_{x \in \tau^{-1}(t)} \hat{\theta}(x)^2 k_n(x)}{\sum_{x \in \tau^{-1}(t)} k_n(x)}$ (c'est-à-dire $\mathbb{E}[\hat{\theta} | \tau]^2 \leq \mathbb{E}[\hat{\theta}^2 | \tau]$), de sorte que :

$$R(\mathbb{E}[\hat{\theta} | \tau], \theta) = \sum_x |g(\tau(x)) - \theta|^2 h_n(\tau(x), \theta) k_n(x)$$

$$\begin{aligned}
&= \sum_x g(\tau(x))^2 h_n(\tau(x), \theta) k_n(x) - 2\theta \sum_x g(\tau(x)) h_n(\tau(x), \theta) k_n(x) + \theta^2 \\
&= \sum_x g(\tau(x))^2 h_n(\tau(x), \theta) k_n(x) - \theta^2 = \sum_t g(t)^2 \sum_{x \in \tau^{-1}(t)} k_n(x) h_n(t, \theta) - \theta^2 \\
&\leq \sum_t \left[\frac{\sum_{x \in \tau^{-1}(t)} \hat{\theta}(x)^2 k_n(x)}{\sum_{x \in \tau^{-1}(t)} k_n(x)} \right] \times \left[\sum_{x \in \tau^{-1}(t)} k_n(x) \right] h_n(t, \theta) - \theta^2 \\
&= \sum_x \hat{\theta}(x)^2 h_n(\tau(x), \theta) k_n(x) - \theta^2 = \sum_x |\hat{\theta}(x) - \theta|^2 h_n(\tau(x), \theta) k_n(x) = R(\hat{\theta}(x), \theta). \quad \diamond
\end{aligned}$$

Corollaire 14.4.3 *En présence d'une statistique exhaustive τ , les estimateurs sans biais ne sont à rechercher que parmi les fonctions de τ .*

Lorsqu'il n'existe qu'un seul estimateur sans biais fonction d'une statistique exhaustive, alors il est optimal parmi les estimateurs sans biais.

Exemple Reprenons l'exemple 3 vu plus haut : $\mu = \mathcal{P}(\lambda)$. La statistique $\bar{X} := \sum_{j=1}^n X_j/n$ est exhaustive, et optimale pour l'estimation (sans biais) de λ . En effet, $f(\bar{X})$ est sans biais ssi $\lambda e^{\lambda n} = \sum_{x \in \mathcal{N}} \frac{(\lambda n)^x}{x!} f\left(\frac{x}{n}\right)$ pour tout $\lambda > 0$, ce qui, par injectivité de la transformation de Laplace, impose que $f = Id$ est la seule solution.

Cherchons un estimateur optimal pour $\theta = e^{-\lambda}$, en observant que l'estimateur élémentaire $1_{\{X_1=0\}}$ est sans biais. La statistique \bar{X} est encore exhaustive, et donc $\mathbb{P}(X_1 = 0 \mid \bar{X})$ est un bon estimateur, en fait optimal. Or (voir l'exercice 5.3) la loi de X_1 sachant $n\bar{X}$ est $\mathcal{B}(n\bar{X}, 1/n)$, de sorte que $\mathbb{P}(X_1 = 0 \mid \bar{X}) = (1 - 1/n)^{n\bar{X}} \neq e^{-\bar{X}}$.

Exercice n° 14.4.1 a) Soient X_1, \dots, X_n des v.a.r.i.i.d. uniformes sur un intervalle $[0, \theta]$. Montrer que $\max\{X_1, \dots, X_n\}$ est une statistique exhaustive pour θ , et que c'est aussi l'estimateur du maximum de vraisemblance pour θ . Est-il biaisé ?

b) Soient X_1, \dots, X_n des v.a.r.i.i.d. de loi $\mathcal{U}([\theta, \theta'])$. Trouver une statistique exhaustive pour (θ, θ') , et l'estimateur du maximum de vraisemblance pour (θ, θ') .

Exercice n° 14.4.2 Soient X_1, \dots, X_n des v.a.r.i.i.d. de loi $\mathcal{E}(\lambda)$. Montrer que l'estimateur sans biais de risque minimum de $e^{-2\lambda}$ est $\hat{\theta} := 1_{\{S_n > 2\}} \left(\frac{S_n - 2}{S_n}\right)^{n-1}$. (On a noté $S_n := X_1 + \dots + X_n = n\bar{X}$.)

(Indications : Noter qu'imposer que $\mathbb{E}[\varphi(S_n)] = \int_0^\infty \varphi(s) e^{-\lambda s} \frac{\lambda^n s^{n-1}}{(n-1)!} ds$ vaille $e^{-2\lambda}$ pour tout $\lambda > 0$ détermine φ , et que selon le théorème de Rao-Blackwell $\mathbb{P}[X_1 > 2 \mid S_n]$ est l'estimateur sans biais optimal. Ensuite, la loi de X_1 sachant S_n est celle du minimum de $(n-1)$ v.a.i.i.d. uniformes sur $[0, S_n]$, ce qui se justifie par :

$$\mathbb{E}[f(X_1) \varphi(S_n)] = \int_0^\infty \int_0^\infty f(x) \varphi(x+y) e^{-\lambda(x+y)} \frac{\lambda^n y^{n-2}}{(n-2)!} dx dy$$

$$= \int_0^\infty \left[\int_0^z f(x) \frac{n-1}{z^{n-1}} (z-x)^{n-2} dx \right] \varphi(z) e^{-\lambda z} \frac{\lambda^n z^{n-1}}{(n-1)!} dz.$$

En effet, ceci signifie que $\mathbb{E}[f(X_1) | S_n] = \int_0^{S_n} f(x) \frac{n-1}{(S_n)^{n-1}} (S_n-x)^{n-2} dx$. Or si M_n est le minimum de $(n-1)$ v.a.i.i.d. uniformes U_1, \dots, U_{n-1} sur $[0, S_n]$, on a :

$$\mathbb{P}[M_n > x] = 1_{\{S_n > x\}} \left(\frac{S_n-x}{S_n} \right)^{n-1} \implies \mathbb{E}[f(M_n)] = \int_0^{S_n} f(x) \frac{n-1}{(S_n)^{n-1}} (S_n-x)^{n-2} dx.$$

Nota Bene La propriété $[\mathbb{E}[f(X_1) \varphi(S_n)] = \mathbb{E}[g(S_n) \varphi(S_n)]]$ pour toute fonction-test φ équivaut à l'assertion : $\mathbb{E}[f(X_1) | S_n] = g(S_n)$. C'est clair dans le cas discret : particulier à $\varphi = 1_{\{k\}}$ (pour toute valeur k prise par S_n), la propriété s'écrit aussitôt $\mathbb{E}[f(X_1) 1_{\{S_n=k\}}] = g(k) \mathbb{P}[S_n = k]$, ou encore : $\mathbb{E}[f(X_1) | S_n = k] = g(k)$.

Exercice n° 14.4.3 Soient $X_1, \dots, X_n, Y_1, \dots, Y_n$ des v.a.r. indépendantes, les X_j de loi $\mathcal{E}(\lambda)$, et les Y_j de loi $\mathcal{E}(\lambda')$. Trouver une statistique exhaustive pour (λ, λ') , et les estimateurs du maximum de vraisemblance pour λ et λ' . Ont-ils un biais ?

14.5 Intervalles de confiance

On veut ici quantifier la proximité d'un estimateur $\hat{\theta}_n$ (fonction d'un échantillon x_1, \dots, x_n) avec le paramètre θ , supposé réel, de la loi inconnue \mathbb{P}_θ . Précisément, on cherche un intervalle $J_n(\hat{\theta}_n) =]b_n(\hat{\theta}_n), b'_n(\hat{\theta}_n)[$, dont on puisse affirmer qu'il contient θ avec une grande probabilité.

Définition 14.5.1 Fixons $0 < \alpha < 1$. Un intervalle de confiance de niveau $1 - \alpha$ est un intervalle $J_n(\hat{\theta}_n) =]b_n(\hat{\theta}_n), b'_n(\hat{\theta}_n)[$, fonction d'un estimateur $\hat{\theta}_n$ et de n , tel que

$$\mathbb{P}_\theta(b_n(\hat{\theta}_n) < \theta < b'_n(\hat{\theta}_n)) \geq 1 - \alpha.$$

On peut souvent obtenir un tel intervalle de confiance sans trop de mal, du moins lorsque la loi de l'estimateur $\hat{\theta}_n$ est accessible : on détermine pour cela d'abord un intervalle symétrique (la symétrie étant facultative) $[\theta - a_n(\theta), \theta + a_n(\theta)]$ tel que pour tout $\theta \in \Theta$

$$\mathbb{P}_\theta(\theta - a_n(\theta) < \hat{\theta}_n < \theta + a_n(\theta)) = 1 - \alpha;$$

puis on résoud autant que possible : $\theta - a_n(\theta) < \hat{\theta}_n < \theta + a_n(\theta) \iff b_n(\hat{\theta}_n) < \theta < b'_n(\hat{\theta}_n)$. C'est souvent aisé, du moins pour les exemples simples, car les bornes $\theta \pm a_n(\theta)$ se trouvent souvent être des fonctions continues strictement croissantes de θ .

La qualité (précision) d'un intervalle de confiance sera d'autant meilleure que l'estimateur utilisé sera bon et que la taille n de l'échantillon sera grande.

14.5.2 Exemple 1 : Estimation de la moyenne d'une loi normale $\mathcal{N}(m, \sigma^2)$

Nous avons vu que dans ce cas la moyenne empirique est le meilleur des estimateurs (sans biais) : c'est donc \bar{x}_n qu'il faut utiliser. Sa loi est $\mathcal{N}(m, \sigma^2/n)$. Nous avons alors aussitôt $a_n(m) = q_{\alpha/2} \times \sigma/\sqrt{n}$, notant q_p le quantile d'ordre p de la loi $\mathcal{N}(0, 1)$. En effet, notant Φ la fonction de répartition gaussienne standard, nous avons : d'une part

$$1 - \alpha = \mathbb{P}_\theta \left(m - a_n(m) < \bar{x}_n < m + a_n(m) \right) = \mathbb{P}_\theta \left(-\frac{\sqrt{n}}{\sigma} a_n(m) < \frac{\sqrt{n}}{\sigma} (\bar{x}_n - m) < \frac{\sqrt{n}}{\sigma} a_n(m) \right) \\ = 2 \mathcal{N}(0, 1) \left(\left[0, \frac{\sqrt{n}}{\sigma} a_n(m) \right] \right) = 2 \Phi \left(\frac{\sqrt{n}}{\sigma} a_n(m) \right) - 1,$$

et d'autre part $2 \Phi(q_{\alpha/2}) - 1 = 2(1 - \alpha/2) - 1 = 1 - \alpha$.

Donc un intervalle de confiance de niveau $1 - \alpha$ pour la moyenne m de $\mathcal{N}(m, \sigma^2)$ est

$$J = J_n(\bar{x}_n) = \left] \bar{x}_n - q_{\alpha/2} \times \sigma/\sqrt{n}, \bar{x}_n + q_{\alpha/2} \times \sigma/\sqrt{n} \right[.$$

En particulier : $\mathcal{N}(m, \sigma^2) \left(\bar{x}_n - (1,96) \sigma/\sqrt{n} < m < \bar{x}_n + (1,96) \sigma/\sqrt{n} \right) \approx 0,95$,

et $\mathcal{N}(m, \sigma^2) \left(\bar{x}_n - (1,65) \sigma/\sqrt{n} < m < \bar{x}_n + (1,65) \sigma/\sqrt{n} \right) \approx 0,90$.

14.5.3 Exemple 2 : Estimation de la moyenne d'une loi normale $\mathcal{N}(m, ?)$

La faiblesse du traitement de l'exemple précédent est qu'il suppose implicitement l'écart-type σ connu, ce qui est rarement le cas. Un bon moyen de contourner cette difficulté est de recourir au corollaire 14.3.2. Nous avons en effet (la loi de Student étant symétrique comme la loi gaussienne standard) :

$$\mathbb{P} \left(-z_{n-1, \alpha/2} < (\bar{x}_n - m) \sqrt{n/V_n} < z_{n-1, \alpha/2} \right) = 1 - \alpha,$$

où $z_{n-1, p}$ désigne le quantile d'ordre p de la loi de Student à $(n - 1)$ degrés de liberté.

Donc un intervalle de confiance (indépendant de σ !) de niveau $1 - \alpha$ pour la moyenne m de $\mathcal{N}(m, \sigma^2)$ est

$$J' = J'_n(\bar{x}_n) = \left] \bar{x}_n - z_{n-1, \alpha/2} \times \sqrt{V_n/n}, \bar{x}_n + z_{n-1, \alpha/2} \times \sqrt{V_n/n} \right[.$$

14.5.4 Exemple 3 : Estimation de la variance d'une loi normale $\mathcal{N}(?, \sigma^2)$

Utilisant le bon estimateur V_n (de déviation standard, voir la proposition 14.2.1) et la proposition 14.3.1, nous avons :

$$\mathbb{P} \left(k_{n-1, \alpha} < (n - 1) \sigma^{-2} V_n < k'_{n-1, \alpha} \right) = 1 - \alpha,$$

si $\chi_{n-1}^2 \left([k_{n-1, \alpha}, k'_{n-1, \alpha}] \right) = 1 - \alpha$ (χ_{n-1}^2 désignant la loi du χ^2 à $(n - 1)$ degrés de liberté).

Donc un intervalle de confiance (indépendant de m !) de niveau $1 - \alpha$ pour la variance σ^2 de $\mathcal{N}(m, \sigma^2)$ est

$$J'' = J''_n(V_n) = \left] \frac{(n-1)V_n}{k'_{n-1, \alpha}}, \frac{(n-1)V_n}{k_{n-1, \alpha}} \right[.$$

14.5.5 Exemple 4 : Estimation de la moyenne d'une loi quelconque

Supposons ici seulement que n soit assez grand, et que la loi μ_m , de moyenne inconnue m , admette un second moment. Notons σ son écart-type. Nous savons que la moyenne empirique \bar{x}_n est un bon estimateur de m (et qu'il converge p.s. vers m).

Appliquons le théorème central limite :

$(m - \bar{x}_n)\sqrt{n}/\sigma = (nm - x_1 - \dots - x_n)/(\sigma\sqrt{n})$ doit être proche d'une v.a. G de loi $\mathcal{N}(0, 1)$.

Posant $\alpha := (a - \bar{x}_n)\sqrt{n}/\sigma$ et $\beta := (b - \bar{x}_n)\sqrt{n}/\sigma$, ce qui équivaut à $a = \bar{x}_n + \alpha\sigma/\sqrt{n}$ et $b = \bar{x}_n + \beta\sigma/\sqrt{n}$, on obtient :

$$\mathbb{P}(m \in J \equiv [a, b]) = \mathbb{P}(\alpha \leq (m - \bar{x}_n)\sqrt{n}/\sigma \leq \beta) \approx \mathbb{P}(G \in [\alpha, \beta]).$$

On choisit ordinairement (mais ce n'est pas une obligation) l'intervalle $[\alpha, \beta]$ centré, et de sorte que $\mathbb{P}(G \in [\alpha, \beta])$ vaille (le plus souvent) 0,9 ou 0,95. On obtient par exemple :

$$\mathbb{P}(\bar{x}_n - (1,96)\sigma/\sqrt{n} \leq m \leq \bar{x}_n + (1,96)\sigma/\sqrt{n}) \approx 0,95.$$

Donc $J = [\bar{x}_n - (1,96)\sigma/\sqrt{n}, \bar{x}_n + (1,96)\sigma/\sqrt{n}]$ est (presque !) un intervalle de confiance à 95% pour m .

Il est généralement irréaliste de supposer l'écart-type σ connu. Comme dans l'exemple 2 ci-dessus, on peut le remplacer par l'estimateur (voir la proposition 14.2.1 ci-dessus) :

$$\sigma \approx \hat{\sigma} = \hat{\sigma}(x_1, \dots, x_n) = \sqrt{V_n}.$$

L'intervalle $J = [\bar{x}_n + \alpha\sqrt{V_n/n}, \bar{x}_n + \beta\sqrt{V_n/n}]$ est alors réellement fonction seulement de l'échantillon (x_1, \dots, x_n) et de n . Mais son niveau n'est plus nécessairement 95%.

Exercice n° 14.5.1 Une caractéristique électrique d'un composant électronique varie, du fait des dispersions de fabrication, suivant une loi normale $\mathcal{N}(\mu, \mu^2)$. On se propose d'estimer son espérance μ par un intervalle de confiance. Un prélèvement de 20 composants effectué sur un lot de série a donné les résultats suivant : 10,1 ; 10,5 ; 9,4 ; 10,2 ; 9,5 ; 9,8 ; 10,1 ; 10,3 ; 10,6 ; 9,7 ; 10,2 ; 10,4 ; 10,3 ; 9,6 ; 9,9 ; 9,8 ; 10,1 ; 10,3 ; 10,5 ; 9,8. ($\bar{x} = 10,055$).

a) Construire un intervalle de confiance pour μ de niveau de confiance 90%. Est-il le plus court parmi tous les intervalles de confiance de niveau de confiance 90% ?

b) Construire un intervalle de confiance pour μ de niveau de confiance 95%.

Exercice n° 14.5.2 La fiabilité dans le temps d'un composant électrique, c'est-à-dire sa durée de vie X , est supposée de type exponentiel: $\mathbb{P}(X \geq t) = \exp(-t/\lambda), t \geq 0$. On cherche à estimer à l'aide d'un intervalle de confiance le paramètre λ . Si la moyenne observée d'un échantillon de taille n est \bar{x} , donner un intervalle de confiance de niveau de confiance 95% du paramètre λ . Est-il le plus court parmi tous les intervalles de confiance à 95% (lorsque n est grand) ?

Exercice n° 14.5.3 Ayant lu dans le magazine ELLE du 19.8.02 que 55% des Français utilisaient des bains moussants, vous décidez de vérifier vous-même l'affirmation du magazine ELLE par un mini-sondage, en posant autour de vous la question :

“Utilisez-vous des bains moussants ? OUI, NON”

et en construisant un intervalle de confiance à 90% sur la proportion considérée.

- a) Vous envisagez de ne considérer l'affirmation comme confirmée que si cet intervalle à 90% basé sur le groupe des personnes interrogées est inclus dans [54%; 56%]. Est-ce raisonnable ?
- b) Découragé(e) par la réponse obtenue en a), vous vous contentez d'un échantillon de 50 personnes. Si la proportion de oui est de 53% parmi les 50 réponses, déterminez un intervalle de confiance à 90%.

14.6 Estimation bayésienne

Le problème général de l'estimation est de trouver θ en fonction de l'observation $X = (X_1, \dots, X_n)$, tandis qu'on connaît a priori plutôt la loi de X en fonction de θ , en particulier via la fonction de vraisemblance $f(X|\theta)$. Il s'agit donc de renverser un conditionnement, d'une certaine façon. Ce qui fait songer à la formule de Bayes.

Suivant cette observation, on se donne une loi a priori sur l'ensemble Θ des valeurs possibles de θ , disons ν , de moyenne m , et au lieu de considérer la famille $\{\mathbb{P}_\theta | \theta \in \Theta\}$ de lois sur $X(\Omega)$, on considère sur $X(\Omega) \times \Theta$ la seule probabilité de référence: $\mathcal{Q} := \mathbb{P}_\theta^{\otimes n} d\nu(\theta)$. Ce qui signifie que la loi sous \mathcal{Q} de l'échantillon $X = (X_1, \dots, X_n)$ s'écrit (pour toute fonction-test φ):

$$\mathbb{E}_{\mathcal{Q}}[\varphi \circ X] = \int_{X(\Omega) \times \Theta} \varphi(x_1, \dots, x_n) d\mathbb{P}_\theta(x_1) \dots d\mathbb{P}_\theta(x_n) d\nu(\theta) = \int_{\Theta} \mathbb{E}_\theta[\varphi \circ X] d\nu(\theta).$$

Mais on a aussi davantage, puisque le paramètre θ est ainsi lui-même devenu une variable aléatoire, de loi ν . On a de fait ainsi défini une loi conjointe pour le vecteur $(X, \theta) \in X(\Omega) \times \Theta$, qui s'exprime (pour toute fonction-test F) par :

$$\mathbb{E}_{\mathcal{Q}}[F(X, \theta)] = \int_{X(\Omega) \times \Theta} F(x_1, \dots, x_n, \theta) d\mathbb{P}_\theta(x_1) \dots d\mathbb{P}_\theta(x_n) d\nu(\theta).$$

En particulier, pour toute fonctions-test φ sur $X(\Omega)$ et g sur Θ :

$$\mathbb{E}_{\mathcal{Q}}[\varphi(X) \times g(\theta)] = \int_{\Theta} \mathbb{E}_\theta[\varphi(X)] g(\theta) d\nu(\theta).$$

Cela permet de définir l'estimateur bayésien de θ comme suit.

Définition 14.6.1 *Étant donnée une loi de probabilité ν sur Θ , de moyenne $m \in \Theta$, l'estimateur bayésien $\hat{\theta}_B$ de m est défini comme: $\hat{\theta}_B := \mathbb{E}_{\mathcal{Q}}[\theta | X]$, l'espérance étant calculée par rapport à la loi de référence $\mathcal{Q} := \mathbb{P}_{\theta}^{\otimes n} d\nu(\theta)$.*

Nota Bene Selon le Nota Bene de la section 14.4, cela signifie que $\hat{\theta}_B$ est caractérisé par la condition: $\mathbb{E}_{\mathcal{Q}}[\hat{\theta}_B(X) \varphi(X)] = \mathbb{E}_{\mathcal{Q}}[\theta \varphi(X)]$ pour toute fonction test φ .

Proposition 14.6.2 *L'estimateur bayésien $\hat{\theta}_B$ est de risque minimal par rapport à \mathcal{Q} .*

Preuve Prenant $\varphi \equiv 1$ dans la condition ci-dessus, on a $\mathbb{E}_{\mathcal{Q}}[\hat{\theta}_B(X)] = \mathbb{E}_{\mathcal{Q}}[\theta] = \int \theta d\nu(\theta) = m$ d'une part, ce qui signifie que $\hat{\theta}_B$ est "sans biais par rapport à \mathcal{Q} ", et d'autre part, pour tout autre estimateur $\hat{\theta} = \hat{\theta}(X)$, prenant $\varphi = \hat{\theta}_B - \hat{\theta}$ on a:

$$\mathbb{E}_{\mathcal{Q}}[(\hat{\theta}_B - m)(\hat{\theta}_B - \hat{\theta})] = 0; \text{ de sorte que}$$

$$\mathbb{E}_{\mathcal{Q}}[(\hat{\theta} - m)^2] = \mathbb{E}_{\mathcal{Q}}[(\hat{\theta} - \hat{\theta}_B)^2] + \mathbb{E}_{\mathcal{Q}}[(\hat{\theta}_B - m)^2] + 2\mathbb{E}_{\mathcal{Q}}[(\hat{\theta}_B - m)(\hat{\theta}_B - \hat{\theta})] \geq \mathbb{E}_{\mathcal{Q}}[(\hat{\theta}_B - m)^2]. \diamond$$

La relation avec la formule de Bayes est directe, au moins dans le cas de lois à densité ou discrètes. Comme d'habitude, $f(X | \theta)$ désigne la fonction de vraisemblance.

Proposition 14.6.3 *La loi conditionnelle (relativement à \mathcal{Q}) de θ sachant $X = (X_1, \dots, X_n)$ possède par rapport à ν la densité:*

$$f(\theta | X) := f(X | \theta) / \int f(X | \theta) d\nu(\theta).$$

Preuve Recourant de nouveau à la condition caractérisant $\hat{\theta}_B$, ou directement au Nota Bene de la section 14.4, nous avons pour toutes fonctions test g, φ :

$$\begin{aligned} \mathbb{E}_{\mathcal{Q}}[\mathbb{E}_{\mathcal{Q}}[g(\theta) | X] \varphi(X)] &= \mathbb{E}_{\mathcal{Q}}[g(\theta) \varphi(X)] = \int \mathbb{E}_{\theta}[\varphi(X)] g(\theta) d\nu(\theta) \\ &= \int \left[\int \varphi(x) f(x | \theta) dx \right] g(\theta) d\nu(\theta) = \int \left[\int f(\theta | x) g(\theta) d\nu(\theta) \right] \times \int f(x | \theta) d\nu(\theta) \times \varphi(x) dx \\ &= \iint \left[\int f(\theta | x) g(\theta) d\nu(\theta) \right] \times \varphi(x) \times f(x | \theta) dx d\nu(\theta) = \mathbb{E}_{\mathcal{Q}} \left[\int f(\theta | X) g(\theta) d\nu(\theta) \times \varphi(X) \right]. \end{aligned}$$

Cela signifie (par le Nota Bene de la section 14.4): $\mathbb{E}_{\mathcal{Q}}[g(\theta) | X] = \int g(\theta) f(\theta | X) d\nu(\theta)$. \diamond

Nota Bene Le calcul ci-dessus est rédigé dans le cas de lois $\mu_{\theta} = \mathbb{P}_{\theta}$ à densité. Dans le cas de lois μ_{θ} discètes, ce calcul reste valable, simplement en remplaçant l'intégration par rapport à dx par la sommation sur x (comme dans la preuve du théorème 14.4.2 de Rao-Blackwell). Noter que $x = (x_1, \dots, x_n)$, de sorte que $dx = dx_1 \dots dx_n$, ou $\sum_x = \sum_{x_1, \dots, x_n}$.

On en déduit en particulier le calcul de $\hat{\theta}_B$.

Corollaire 14.6.4 $\hat{\theta}_B = \int \theta f(\theta | X) d\nu(\theta) = \int f(X | \theta) \theta d\nu(\theta) / \int f(X | \theta) d\nu(\theta)$.

Et pour tous estimateurs a_n, b_n :

$$\mathbb{P}[a_n(X) < \theta < b_n(X) | X] = \int_{a_n(X)}^{b_n(X)} f(X | \theta) d\nu(\theta) / \int f(X | \theta) d\nu(\theta).$$

La seconde formule permet d'obtenir un intervalle de confiance $[a_n(x_1, \dots, x_n), b_n(x_1, \dots, x_n)]$.

Exemples 4) $\mu_\theta = \mathcal{B}(\theta)$ et $\nu = \mathcal{U}([0, 1])$: $f(\theta | x_1, \dots, x_n) = \frac{\theta^{\sum_{j=1}^n x_j} \times (1 - \theta)^{n - \sum_{j=1}^n x_j}}{\int_0^1 \theta^{\sum_{j=1}^n x_j} (1 - \theta)^{n - \sum_{j=1}^n x_j} d\theta}$,

d'où, en utilisant que $\int_0^1 \theta^k (1 - \theta)^\ell d\theta = \frac{k! \ell!}{(k + \ell + 1)!}$: $\mathbb{E}[\theta | X_1, \dots, X_n] = \frac{\sum_{j=1}^n X_j + 1}{n + 2}$.

5) $\mu_\theta = \mathcal{N}(\theta, 1)$ et $\nu = \mathcal{N}(0, 1)$: on obtient $\mathbb{E}[\theta | X_1, \dots, X_n] = \frac{\sum_{j=1}^n X_j}{n + 1}$, car on trouve

$$e^{-\theta^2/2} f(\theta | x_1, \dots, x_n) = k_n(x_1, \dots, x_n) \times \exp \left[-\frac{1}{2}(n + 1) \left(\theta - \frac{\sum_{j=1}^n x_j}{n + 1} \right)^2 \right],$$

de sorte que la loi conditionnelle de θ sachant (X_1, \dots, X_n) est $\mathcal{N}\left(\frac{\sum_{j=1}^n X_j}{n + 1}, \frac{1}{n + 1}\right)$;

un intervalle de confiance à 95% pour θ est donc $\left[\frac{\sum_{j=1}^n x_j}{n + 1} - \frac{1,96}{\sqrt{n + 1}}, \frac{\sum_{j=1}^n x_j}{n + 1} + \frac{1,96}{\sqrt{n + 1}} \right]$.

Exercice n° 14.6.1 Un processus de Poisson de paramètre inconnu λ est observé durant le temps fixe T . Soit N le nombre d'événements observés durant ce temps. On se donne pour λ la loi a priori $\mathcal{E}(\alpha)$. Calculer l'estimateur bayésien de λ .

15 Tests

Au lieu d'estimer le paramètre inconnu θ de la loi \mathbb{P}_θ , on cherche ici à savoir s'il vérifie une hypothèse statistique notée H_0 , identifiée à une partie de Θ , et nommée aussi hypothèse nulle. Pour cela on emploie un test. Par exemple $H_0 = \{\theta < 7\}$, et le test est une règle devant permettre de trancher ou de parier, en fonction d'un échantillon statistique relatif à μ_θ , si l'hypothèse nulle H_0 doit être acceptée ou bien rejetée. Dans le cas le plus général, on considère une hypothèse alternative H_1 , contre laquelle on teste H_0 , conditionnellement à $H_0 \cup H_1$. Ici on a choisi simplement $H_1 = (H_0)^c$ (le complémentaire étant entendu dans Θ).

Bien entendu, l'acceptation ou le rejet de l'hypothèse ne peut pas être une décision déterministe : il y a toujours en pratique une probabilité non nulle que la décision arrêtée soit erronée. Il faut donc décider a priori un seuil de confiance, dit aussi niveau de test, c'est-à-dire une probabilité de rejet à tort. La plupart du temps, on fixe ce niveau à 5%.

La décision d'acceptation ou de rejet de l'hypothèse nulle H_0 sera prise en fonction d'un estimateur $\hat{\theta}_n$ de θ , bien entendu de la meilleure qualité possible, et de loi calculable. Ce sera la "variable de décision".

On nomme région critique l'événement $W := \{\hat{\theta}_n \in \Theta_1\}$, Θ_1 étant l'ensemble déterministe (calculé) des valeurs pour lesquelles on décide a priori de *rejeter* H_0 .

Donc : si $\omega \in W$, on rejette H_0 , et si $\omega \in W^c$, on accepte H_0 . Et on veut que

$$\sup_{\theta \in H_0} P_\theta(W) = \sup_{\theta \in H_0} P_\theta(\text{rejeter } H_0) = 0,05.$$

Il y a clairement deux types de risque : rejeter H_0 à tort, et accepter H_0 à tort.

Exemples : 1) Évacuer Basse-Terre (Guadeloupe) alors que l'explosion attendue du volcan de la Soufrière ne se produira pas ; ou ne pas l'avoir évacuée alors que l'explosion a lieu.

2) Classer malade un individu qui ne l'est pas ; ou classer sain un individu malade.

On voit sur ces deux exemples : d'une part que la décision prise peut être très lourde de conséquences, et d'autre part que les deux types de risque ne revêtent pas le même ordre de gravité. Dans le premier exemple, on prendra pour H_0 le sous-ensemble de Θ correspondant au déclenchement de l'explosion (et donc le rejet de H_0 se traduira par la non-évacuation), et dans le second, le sous-ensemble de Θ correspondant à la maladie (et donc le rejet de H_0 se traduira par le classement parmi les individus sains et le non-traitement). Il est en effet tout indiqué de fixer un seuil bas (5%, voire encore moins) pour le risque le plus grave, soit le rejet de H_0 à tort, dit risque de première espèce.

Quant au risque de deuxième espèce, soit l'acceptation de H_0 à tort, il est contrôlé par la quantité

$$\inf_{\theta \in H_1} P_\theta(W) = \inf_{\theta \in H_1} P_\theta(\text{rejeter } H_0),$$

qu'on nomme puissance du test, et qu'on souhaite la plus grande possible (dans les 2 exemples ci-dessus, cela signifie minimiser le risque de décision d'évacuer Basse-Terre alors qu'il n'y a pas d'explosion, comme celle de C. Allègre contre l'avis de H. Tazieff en 1976, ou bien de décision de traiter un individu sain).

L'élaboration d'un test suit ainsi la démarche :

- fixation du seuil ou niveau α (5% bien souvent) ;
- choix de l'hypothèse nulle H_0 ($\subset \Theta$, et de H_1 , souvent $\Theta \setminus H_0$) ;
- choix de l'estimateur $\hat{\theta}_n$ comme variable de décision ;
- choix de l'allure de la région critique $W = \{\hat{\theta}_n \in \Theta_1\}$ (et donc de $\Theta_1 \subset \Theta$) ;
- calcul de Θ_1 en fonction du seuil α (et de la loi de $\hat{\theta}_n$) ;
- calcul éventuel de la puissance (qui doit être aussi grande que possible) ;
- calcul de la valeur expérimentale $\hat{\theta}_n(\omega)$;
- conclusion : rejet de H_0 si $\omega \in W$, et acceptation sinon.

15.1 Test de Neyman-Pearson

On se limite ici au cas très simple où $H_0 = \{\theta_0\}$ et $H_1 = \{\theta_1\}$ sont des singletons.

En terme de la fonction de vraisemblance (voir la section 14.1) $f(x|\theta) = f(x_1, \dots, x_n|\theta)$, nous avons simplement $\alpha = \mathbb{P}_{\theta_0}(W) = \int_W f(x|\theta_0) dx$, et $\mathbb{P}_{\theta_1}(W) = \int_W f(x|\theta_1) dx$, qui est la puissance et doit être la plus grande possible.

Proposition 15.1.1 (Neyman-Pearson) Lorsque $H_0 = \{\theta_0\}$ et $H_1 = \{\theta_1\}$ sont des singletons, il y a une région critique optimale : $W = \left\{ x \in \mathbb{R}^n \mid \frac{f(x|\theta_0)}{f(x|\theta_1)} \leq k_\alpha \right\}$, k_α étant déterminé par l'équation $\alpha = \mathbb{P}_{\theta_0}(W)$.

15.1.2 Exemple 1 : Test de la moyenne d'une loi normale $\mathcal{N}(m, \sigma^2)$

Nous avons dans ce cas $f(x|m) = (\sigma\sqrt{2\pi})^{-n} \exp\left[-\sum_{j=1}^n (x_j - m)^2 / (2\sigma^2)\right]$, et donc selon la proposition 15.1.1, le meilleur choix de test est :

$$W = \left\{ \sum_{j=1}^n [(x_j - m_1)^2 - (x_j - m_0)^2] \leq 2\sigma^2 \log k_\alpha \right\} = \left\{ (\bar{x}_n - \frac{m_0+m_1}{2})(m_0 - m_1) \leq \frac{\sigma^2}{n} \log k_\alpha \right\}.$$

Considérons le cas $m_0 < m_1$ (le cas contraire se traite de la même façon). Alors

$W = \left\{ \bar{x}_n \geq \frac{m_0+m_1}{2} - \frac{\sigma^2 \log k_\alpha}{n(m_1-m_0)} =: \lambda \right\}$, avec k_α (ou λ) déterminé par (notant Φ la fonction de répartition gaussienne centrée réduite) : $\alpha = \mathbb{P}_{m_0}(\bar{X}_n \geq \lambda) = 1 - \Phi\left(\frac{\lambda-m_0}{\sigma/\sqrt{n}}\right)$, id est $\lambda = m_0 + (\sigma/\sqrt{n}) q_\alpha$.

Nous aboutissons à la règle de décision du test, en fonction de la variable de décision \bar{x}_n (la moyenne empirique, qui est ici l'estimateur optimal) : on accepte la valeur m_0 pour la moyenne si $\bar{x}_n < \lambda$, et on accepte la valeur m_1 sinon.

Notons que la puissance de ce test est $\mathbb{P}_{m_1}(\bar{X}_n \geq \lambda) = 1 - \Phi\left(\frac{\lambda-m_1}{\sigma/\sqrt{n}}\right) = 1 - \Phi\left(q_\alpha - \frac{m_1-m_0}{\sigma/\sqrt{n}}\right)$.

15.1.3 Exemple 2 : Test de la moyenne d'une loi normale $\mathcal{N}(m, ?)$

Comme dans la section 14.5, il est préférable de ne pas supposer l'écart-type σ connu, et pour cela, recourant au corollaire 14.3.2, d'utiliser la loi de Student. Nous avons toujours $W = \left\{ \bar{x}_n \geq \lambda \right\}$, avec cette fois λ déterminé par :

$\alpha = \mathbb{P}_{m_0}(\bar{X}_n \geq \lambda) = \mathbb{P}\left(T_{n-1} \geq (\lambda - m_0)\sqrt{n/v_n}\right)$, où T_{n-1} suit la loi de Student à $(n-1)$ degrés de liberté, et où v_n désigne la déviation standard relative à l'échantillon (x_1, \dots, x_n) . De sorte que $\lambda = m_0 + z_{n-1, \alpha} \sqrt{v_n/n}$, où $z_{n-1, \alpha}$ désigne le quantile d'ordre α de la loi de Student à $(n-1)$ degrés de liberté.

Nous aboutissons donc à une règle de décision du test identique à celle de l'exemple précédent, pour une valeur de λ différente, et indépendante de l'écart-type σ .

La puissance de ce test est $\mathbb{P}_{m_1}(\bar{X}_n \geq \lambda) = \mathbb{P}(T_{n-1} \geq z_{n-1,\alpha} - (m_1 - m_0)\sqrt{n/v_n})$.

15.1.4 Exemple 3: Test de la variance d'une loi normale $\mathcal{N}(m, \sigma^2)$

Appliquons de nouveau la proposition 15.1.1 (de Neyman-Pearson): le meilleur choix de test (avec ici $f(x|\sigma) = (\sigma\sqrt{2\pi})^{-n} \exp\left[-\sum_{j=1}^n (x_j - m)^2/(2\sigma^2)\right]$) est:

$$W = \left\{x \mid \frac{f(x|\sigma_0)}{f(x|\sigma_1)} \leq k_\alpha\right\} = \left\{x \mid \frac{1}{2}[\sigma_1^{-2} - \sigma_0^{-2}] \sum_{j=1}^n (x_j - m)^2 \leq \log k_\alpha - n \log(\sigma_1/\sigma_0)\right\}.$$

Alors utilisant comme variable de décision la variance empirique $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{j=1}^n (x_j - m)^2$ (qui est un bon estimateur, lorsque la moyenne m est connue), et considérant le cas $\sigma_0 > \sigma_1$ (le cas contraire se traite de la même façon), nous avons:

$$W = \left\{x \mid \hat{\sigma}_n \leq 2 \left[\frac{1}{n} \log k_\alpha - \log(\sigma_1/\sigma_0) \right] / [\sigma_1^{-2} - \sigma_0^{-2}]\right\} =: \{\hat{\sigma}_n \leq s\}.$$

Puisque $n \hat{\sigma}_n^2 / \sigma_0^2$ suit sous \mathbb{P}_{σ_0} la loi χ_n^2 , la valeur k_α est déterminée par:

$$\alpha = \mathbb{P}_{\sigma_0}[W] = \mathbb{P}[\chi_n^2 \leq n s \sigma_0^{-2}] = \mathbb{P}\left[\chi_n^2 \leq \frac{2\sigma_1^2}{\sigma_0^2 - \sigma_1^2} [\log k_\alpha - n \log(\sigma_1/\sigma_0)]\right].$$

La puissance du test est $\mathbb{P}_{\sigma_1}[W] = \mathbb{P}[\chi_n^2 \leq n s \sigma_1^{-2}] = \mathbb{P}\left[\chi_n^2 \leq \frac{2\sigma_0^2}{\sigma_0^2 - \sigma_1^2} [\log k_\alpha - n \log(\frac{\sigma_1}{\sigma_0})]\right]$.

En conclusion, on accepte la valeur σ_1 si $\hat{\sigma}_n \leq s$, et on accepte la valeur σ_0 sinon.

15.1.5 Exemple 4: Test de la variance d'une loi normale $\mathcal{N}(?, \sigma^2)$

Comme dans la section 14.5, il est naturel d'utiliser ici (dans le cas d'une moyenne inconnue) comme variable de décision le bon estimateur V_n (de déviation standard, voir la proposition 14.2.1). La proposition 15.1.1 (de Neyman-Pearson) ne s'applique pas directement, mais les exemples précédents conduisent à choisir (toujours dans le cas $\sigma_0 > \sigma_1$) une région critique du même type: $W = \{V_n \leq v\}$. Puisque $(n-1)V_n/\sigma_0^2$ suit sous \mathbb{P}_{σ_0} la loi χ_{n-1}^2 , la valeur v est déterminée par:

$$\alpha = \mathbb{P}_{\sigma_0}[W] = \mathbb{P}[\chi_{n-1}^2 \leq (n-1)v\sigma_0^{-2}].$$

La puissance du test est $\mathbb{P}_{\sigma_1}[W] = \mathbb{P}[\chi_{n-1}^2 \leq (n-1)v\sigma_1^{-2}]$.

En conclusion, on accepte la valeur σ_1 si $V_n \leq v$, et on accepte la valeur σ_0 sinon.

15.2 Test de la moyenne d'une loi quelconque

On ne suppose plus ici que l'alternative soit entre deux singletons, mais seulement que n est assez grand, et que la loi \mathbb{P}_m , de moyenne inconnue m , admet un second moment.

Notons σ son écart-type. Nous savons que la moyenne empirique \bar{x}_n est un très bon estimateur de m (et qu'il converge p.s. vers m).

Si par exemple $H_0 = \{\theta \leq a\}$, on prendra une région critique de la forme $\{\bar{x}_n > a'\}$ (si la statistique observée tombe dedans, on décide de rejeter l'hypothèse nulle), avec $a' > a$ à déterminer pour que le niveau soit bien 5%, c'est-à-dire ici pour que

$$\mathbb{P}_a(\bar{X}_n > a') = 0,05 \quad (\text{si } H_0 = \{\theta \leq a\}).$$

Pour n grand, si σ est supposé connu, recourant au théorème central limite, on assimilera $\sqrt{n}(\bar{X}_n - a)/\sigma$ à une variable gaussienne centrée réduite G , de sorte que a' doit vérifier :

$$0,05 = \mathbb{P}_a(\bar{X}_n > a') = \mathbb{P}_a(\sqrt{n}(\bar{X}_n - a)/\sigma > \sqrt{n}(a' - a)/\sigma) \approx \mathbb{P}(G > \sqrt{n}(a' - a)/\sigma),$$

ce qui donne $\sqrt{n}(a' - a)/\sigma \approx 1,65$, id est $a' \approx a + (1,65)\sigma/\sqrt{n}$.

Toutefois, comme déjà dit dans la section 14.5, il est généralement irréaliste de supposer l'écart-type σ connu. Comme dans les exemples vus plus haut, on peut recourir à des tests plus sophistiqués, ou estimer la variance via la déviation standard...

Exemples 1) $\mathbb{P}_\theta = \mathcal{E}(1/\theta)$. Via le changement de variable $s_1 = x_1, \dots, s_n = x_1 + \dots + x_n$:

$$\begin{aligned} \mathbb{P}_a(\bar{X}_n < a') &= a^{-n} \int_{x_1 + \dots + x_n < na'} e^{-(x_1 + \dots + x_n)/a} dx_1 \dots dx_n = \int_{x_1 + \dots + x_n < na'/a} e^{-(x_1 + \dots + x_n)} dx_1 \dots dx_n \\ &= \int_{\{0 < s_1 < \dots < s_n < na'/a\}} e^{-s_n} ds_1 \dots ds_n = \int_0^{na'/a} \left(\int_{\{0 < s_1 < \dots < s_{n-1} < s_n\}} ds_1 \dots ds_{n-1} \right) e^{-s_n} ds_n \\ &= \int_0^{na'/a} \frac{s^{n-1}}{(n-1)!} e^{-s} ds = 1 - e^{-na'/a} \sum_{j=0}^{n-1} \frac{(na'/a)^j}{j!} = e^{-na'/a} \sum_{j \geq n} \frac{(na'/a)^j}{j!}. \end{aligned}$$

Il est assez légitime d'approcher la loi $\mathcal{P}(na'/a)$ par une loi normale, id est $\mathcal{N}(na'/a, na'/a)$. Ce qui donne (pour une variable gaussienne centrée réduite G) :

$$0,05 = \mathbb{P}_a^{\otimes n}(\bar{X}_n > a') = \mathcal{P}(na'/a)([0, n]) \approx \mathbb{P}(G > \sqrt{n}(\sqrt{a'/a} - \sqrt{a/a'})),$$

et donc $a'/a \approx 1 + (1,65)/(2\sqrt{n})$.

$$2) \mathbb{P}_\theta = \mathcal{P}(\theta). \text{ Alors } \mathbb{P}_a(\bar{X}_n > a') = \mathcal{P}(na)[\{j \mid j > na'\}] = e^{-na} \sum_{j > na'} \frac{(na)^j}{j!}.$$

Il est légitime d'approcher la loi $\mathcal{P}(na)$ par une loi normale, id est $\mathcal{N}(na, na)$. Ce qui donne (pour une variable gaussienne centrée réduite G) :

$$0,05 = \mathbb{P}_a(\bar{X}_n > a') \approx \mathbb{P}(G > \sqrt{n/a}(a' - a)),$$

et donc $a' \approx a + (1,65)\sqrt{a/n}$.

Exercice n° 15.1 Deux associés A et B se partagent les tâches d'un cabinet commun de conseil. L'associé A pense qu'il traite moins de 30% des dossiers. Pour vérifier cela, il décide de choisir au hasard un échantillon de 100 dossiers (traités soit par lui-même, soit par l'associé B), sur lesquels il prévoit d'effectuer deux tests, avec le même risque de rejet à tort de 5%, en prenant successivement comme hypothèse nulle

- i) $H_0 : p \leq 30\%$ ii) $H'_0 : p \geq 30\%$, p étant la proportion réelle des dossiers traités par A.
a) Pour combien de dossiers traités par A (au plus) l'hypothèse H_0 est-elle acceptée ?
b) Pour combien de dossiers traités par A H'_0 est-elle rejetée ? acceptée ?
c) Déterminer la probabilité d'accepter l'hypothèse H_0 si la proportion réelle de dossiers traités par A est de 40%.
d) Déterminer la probabilité d'accepter l'hypothèse H'_0 si la proportion réelle de dossiers traités par A est de 40%.

15.3 Analyse de la variance (à un facteur)

Il s'agit ici de l'une des procédures les plus utilisées dans la pratique. Il s'agit de comparer les moyennes de plusieurs variables gaussiennes de même variance fixe, et de tester en particulier leur égalité.

La donnée typique est ici une suite de variables aléatoires gaussiennes supposées indépendantes Y_{ij} , doublement indexées par $(i, j) \in \{1, \dots, I\} \times \{1, \dots, J\}$, dont on se donne a priori les lois: $\mathbb{P}_{Y_{ij}} = \mathcal{N}(m_i, \sigma^2)$, qui ne peuvent différer que par leur moyenne m_1, \dots, m_I . Le facteur (ici monodimensionnel) dont on recherche l'impact (l'incidence) éventuel(le) sur la statistique est l'indice i . Voyons précisément comment tester l'hypothèse nulle $H_0 := \{m_1 = \dots = m_I\}$ contre $H_1 := H_0^c$.

Le principe de ce test (dû à Fisher) est de comparer les dispersions des moyennes observées à la dispersion globale. Posons pour cela :

$$\bar{y}_i := \frac{1}{J} \sum_{j=1}^J y_{ij}; \quad \bar{y} := \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J y_{ij} = \frac{1}{I} \sum_{i=1}^I \bar{y}_i; \quad \sigma_i^2 := \frac{1}{J} \sum_{j=1}^J (y_{ij} - \bar{y}_i)^2; \quad \sigma^2 := \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y})^2.$$

Lemme 15.3.1 *La variance globale est la somme de la variance des moyennes (variance V_F due au facteur) et de la moyenne des variances (variance résiduelle V_R):*

$$\sigma^2 = \frac{1}{I} \sum_{i=1}^I (\bar{y}_i - \bar{y})^2 + \frac{1}{I} \sum_{i=1}^I \sigma_i^2 =: V_F + V_R.$$

Preuve Nous avons

$$\frac{1}{IJ} \sum_{i,j} y_{ij}^2 - \frac{1}{I} \sum_i \bar{y}_i^2 = \frac{1}{I} \sum_i \left[\frac{1}{J} \sum_j y_{ij}^2 - \left(\frac{1}{J} \sum_j y_{ij} \right)^2 \right] = \frac{1}{I} \sum_i \sigma_i^2,$$

d'où

$$\sigma^2 = \frac{1}{IJ} \sum_{i,j} y_{ij}^2 - \bar{y}^2 = \left(\frac{1}{I} \sum_i \bar{y}_i^2 - \bar{y}^2 \right) + \frac{1}{I} \sum_i \sigma_i^2 = V_F + V_R. \diamond$$

(Noter la ressemblance avec la section 13.6 sur les variations expliquée et inexpliquée.)

Si l'hypothèse nulle $H_0 = \{m_1 = \dots = m_I\}$ est correcte, alors V_F doit être petite par rapport à V_R . Pour les comparer précisément, on les normalise par leurs degrés de liberté, de sorte qu'on considère :

- la variance corrigée due au facteur : $S_F := \frac{I}{I-1} V_F$; et
- la variance résiduelle corrigée : $S_R := \frac{J}{J-1} V_R$;
- une région critique du type $W = \{\varphi := S_F/S_R \geq c\}$.

Dans le cas où H_0 est vraie, la loi de $\varphi := S_F/S_R$ est celle du rapport de deux variables χ^2 indépendantes, qu'on nomme loi de Fisher, et qui est tabulée. On obtient ainsi $c \approx 4,066$ pour que $\mathbb{P}_{H_0}[\varphi \geq c] = 0,05$, et on accepte H_1 si $\varphi \geq c$, et H_0 sinon.

Exemple : Études de la stabilité dans le temps de l'hydrophilie d'éponges artificielles.

éponge n ^o	après	indice d'hydrophilie
1	3 mois	43
2	3 mois	40
3	3 mois	41
4	6 mois	36
5	6 mois	40
6	6 mois	39
7	12 mois	28
8	12 mois	24
9	12 mois	33
10	24 mois	32
11	24 mois	29
12	24 mois	32

Nous avons alors successivement :

$$I = 4; J = 3; \bar{y}_1 = 41, 33; \bar{y}_2 = 38, 33; \bar{y}_3 = 28, 33; \bar{y}_4 = 31;$$

$$\sigma^2 = 32, 85; \sigma_1^2 = 1, 56; \sigma_2^2 = 2, 89; \sigma_3^2 = 13, 56; \sigma_4^2 = 2;$$

$$V_F = 27, 85; V_R = 5; S_F = 37, 13; S_R = 7, 5; \varphi \approx 4, 95 > 4, 066.$$

De sorte qu'il convient ici de rejeter l'hypothèse de stabilité dans le temps de l'hydrophilie.

BIBLIOGRAPHIE

- BREMAUD P. *Introduction aux probabilités.* Springer, Berlin 1984.
- COTTRELL M., GENON-CATALOT V., DUHAMEL C., MEYRE T.
Exercices de probabilité. Cassini, Paris 1999.
- FOATA D., FUCHS A. *Calcul des probabilités.* Dunod, Paris 1998.
- HOGG R., CRAIG A. *Introduction to mathematical statistics.*
The Macmillan Company, New York 1965.
- LESIGNE E. *Une introduction aux théorèmes limites du calcul des probabilités.*
Ellipses, Paris 1997.
- LETAC G. *Intégration et probabilités.* Masson, Paris 1997.
- REES D.G. *Essential Statistics.* Chapman & Hall, London New York 1989.
- ROSS S. *Introduction to probability models.* Academic press, 1980.
- SAPORTA G. *Probabilités, analyse des données et Statistique.* Technip, 2006.