

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique



Université de Saïda - Dr Moulay Tahar.
Faculté des Sciences.
Département de Mathématiques.



Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Filière : MATHÉMATIQUES

Spécialité : Analyse stochastique et statistique des processus

par

Bouzid Bouchra¹

Sous la direction de

Dr K. Djerfi

Thème :

Sur la symétrie des lois de probabilité : une approche géométrique

Soutenu le 14/06/2022 devant le jury composé de

S. Rahmani	Université de Saïda Dr. Moulay Tahar	Président
K. Djerfi	Université de Saïda Dr. Moulay Tahar	Encadreur
T. Djebbouri	Université de Saïda Dr. Moulay Tahar	Examineur

Année univ. : 2021/2022

1. e-mail : bbouzid1999@gmail.com

Remerciements

En première lieu, je commence par remercier notre Dieu qui nous a donné de la force et le courage pour effectuer le présent travail. En second lieu, je remercie vivement mon encadreur Dr.K.Djerfi docteur à l'université de saïda de m'avoir proposé le sujet de ce mémoire, pour son aide, sa patience et ses conseils qui m'ont été d'un grand apport, pendant toute la durée de l'élaboration de ce travail, et qu'il a consacré pour mener à bien ce travail.

Je tiens aussi à remercier :

- Le président de jury, d'avoir eu l'aimabilité d'accepter la présidence du jury de soutenance de ce mémoire.
- Les membres de jury pour leur participation et leur dévouement.

〈Merci à vous tous〉

Dédicace

Je dédie ce travail

À ma chère mère et à mon cher père qui n'ont jamais cessé de me
supporter, me soutenir et m'encourager durant mes années
d'étude. Qu'ils trouvent ici le témoignage de ma profonde
gratitude et reconnaissance.

À mes frères, mes grands-parents et ma famille qui me donnent
de l'amour et de la vivacité.

À tous ceux qui m'ont aidé de près ou de loin et ceux qui ont
partagé avec moi les moments d'émotion lors de la réalisation de
ce travail et qui m'ont chaleureusement supporté et encouragé
tout au long de mon parcours.

À tous mes amis et à mon amie le plus distingué **Zouad Fatima** m'ont
toujours encouragé, et à qui je
souhaite plus de succès

Merci!

Table des matières

Tableau des index et notation	6
1 Rappels sur Les lois de probabilités usuelles discrètes	9
1.1 Définitions et Propriétés	9
1.2 Les lois discrètes	11
2 Rappels sur les lois de probabilités usuelles continues	30
2.1 Définitions et Propriétés	30
2.2 Les lois continues	31
3 Sur les lois à symétrie elliptique	56
3.1 Introduction	56
3.2 Définitions et propriétés	56
3.3 Exemples de mesures et de lois radiales	62
3.4 Propriétés élémentaires des lois radiales	66
3.5 Propriétés caractéristiques des lois radiales	70
3.6 Cas de la normalité	79
4 Application à la théorie de l'estimation	81

4.1	Introduction	81
4.2	Modèles	82
4.3	Admissibilité	83
4.4	Contraction de Stein et estimateur de James-Stein	84
4.5	Contraction de Stein	86
4.6	Phénomène de Stein	88
	Bibliographie	90

Tableau des index et notation

\mathbb{P}	: La loi de probabilité de la variable aléatoire discrét ou continue
p et q	: La valeur de la probabilité
E^*	: La dual de l'espace E
P	: La loi de probabilité radiale sur E
$\langle . \rangle$: Produit scalaire dans \mathbb{R}^d
H	: Un sous espace vectoriel de E
v	: Un produit scalaire sur E^*
$v_{(H)}$: Produit scalaire sur H^*
π	: La projection v^{-1} – <i>orthogonale</i> de E sur H
$B_{v,r}$: La boule de centre 0_E et rayon $r \geq 0$
(e_1, \dots, e_n)	: est une base se E
$S_{v,r}$: La sphère de centre 0_E et rayon $r \geq 0$
(e_1^*, \dots, e_n^*)	: est une base dual se E^*
$ $: Sachant
$\ \cdot\ $: la norme eculidienne de \mathbb{R}^d
$\hat{\theta}$: l'estimateur de paramètre θ
$\hat{\theta}_{JS}$: l'estimateur de James -Stein

Introduction générale

Les premières tentatives de découvrir l'interaction entre la géométrie différentielle et la théorie des probabilités et statistiques sont due à Rao à l'époque des années quarante du vingtième siècle. Les chercheurs ont essayé d'appliquer les méthodes de la géométrie différentielle et surtout riemanniennes à l'inférence statistique. Dans le cadre paramétrique par exemple l'ensemble des lois de probabilités qui gère un modèle statistique est présentée par une variété différentiable dont le système de coordonnées (cartes) est exprimé par le paramètre de loi en question. L'exemple le plus courant de cette approche est la famille normale, qui est généralement exprimée comme une famille de densités.

$$P(\mu, \sigma) = N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right)$$

Le paramètre d'intérêt dans ce cas le couple (μ, σ) qui varie sur l'ouvert de \mathbb{R}^2 déterminé par $\mu > 0$. L'espace échantillon est \mathbb{R} et les densités sont par rapport à la mesure de Lebesgue sur \mathbb{R} .

En effet, si la famille des lois qui gère un phénomène aléatoire est présentée par une variété dont le système de coordonnées est celui du paramètre de la loi en question, alors comme toute variété riemannienne, il existe un groupe

non trivial (inclus dans le groupe des isométries) qui agit sur la variété. La reconnaissance et l'étude de ce groupe de symétrie a une grande importance sur l'étude du phénomène aléatoire.

Un des exemples des résultats de cette approche en statistique est le phénomène de Stein exprimé par le lemme dû à ce dernier. On montre dans le chapitre 4 de ce mémoire que ce phénomène est réellement une conséquence de la symétrie elliptique de la loi normale.

Ce mémoire est composé de quatre chapitres. Dans les deux premiers chapitres on rappelle les principales notions de probabilités notamment les propriétés des lois de probabilités discrètes et continues. Le troisième chapitre traite la notion de symétrie sphérique de la loi normale et les propriétés correspondantes.

Le quatrième chapitre est consacré à une application en statistique. On explique l'impact de la reconnaissance du groupe de symétrie d'une loi sur l'estimation du paramètre de la loi (moyenne), et comment cette approche géométrique a produit une classe d'estimateurs qui dominent l'estimateur naturel de maximum de vraisemblance.

Chapitre 1

Rappels sur Les lois de probabilités usuelles discrètes

1.1 Définitions et Propriétés

Définition 1.1.

Les lois de probabilités sont des objets mathématiques qui permettent aux statisticiens de fabriquer des modèles pour décrire des phénomènes ou le hasard intervient.

Une loi de probabilité est une distribution théorique de fréquence.

Soit Ω une ensemble muni d'une probabilité \mathbb{P} .

Une variable aléatoire X est une application définie sur Ω dans \mathfrak{R} .

X permet de transporter la loi \mathbb{P} en loi \mathbb{P}' définie sur $\Omega' = X(\Omega)$.

on a : $\mathbb{P}'(x_j) = \mathbb{P}(X = x_j)$. La loi \mathbb{P}' appelée loi de X .

Une variable aléatoire est une application de l'univers Ω dans \mathfrak{R} .

$$X : \Omega \rightarrow \mathfrak{R}$$

$$W \mapsto X(W)$$

La variable aléatoire est dit discrète si l'ensemble $X(\Omega)$ est discret.

c'est à dire qui ne prend que des valeurs ponctuelles, isolées, typiquement N ou Z .

on définit :

– m la moyenne, ou $E(X)$ l'espérance de X par :

$$m = E(X) = \sum_{i=1}^n x_i \mathbb{P}(X = x_i)$$

– u_2 : le moment d'ordre 2 par :

$$u_2 = E(X^2) = \sum_{i=1}^n x_i^2 \mathbb{P}(X = x_i)$$

– $Var(X)$: la variance de X par :

$$Var(X) = \sum_{i=1}^n (x_i - m)^2 \mathbb{P}(X = x_i) = E(X^2) - E(X)^2$$

– σ : l'écart type par :

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

1.2 Les lois discrètes

Loi uniforme

Définition 1.2.

X est une variable aléatoire qui prend les valeurs x_1, x_2, \dots, x_n .

si pour tout i on a : $\mathbb{P}_i = \frac{1}{n}$.

alors la distribution de la variable aléatoire X est dit uniforme.

Soit $(a, b) \in \mathbb{Z}^2$ tel que $a \leq b$. $X \rightsquigarrow U([a, b] \setminus \mathbb{Z})$ si et seulement si :

$X(\Omega) = [a, b]$ et $\forall K \in X(\Omega)$, $\mathbb{P}(X = K) = \frac{1}{(b-a+1)}$.

Propriétés 1.1.

Loi d'une variable aléatoire X prenant ses valeurs dans $\{1, \dots, n\}$ avec la même probabilité :

$$\mathbb{P}(X = x) = \frac{1}{n}; \forall x \in \{1, 2, \dots, n\}$$

Moment :

$$E(X) = \frac{(n+1)}{2}$$

$$V(X) = \frac{(n^2-1)}{12}$$

L'éléments de calcul pour l'espérance et la variance :

$$\sum_{i=1}^n i = \frac{n(n+1)}{2}$$

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$$

Exemple 1.1.

X : résultat d'un jet de dé à six faces non-pipé.

Les $n = 6$ modalités possibles :

$$x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5, x_6 = 6$$

ont toutes pour probabilité élémentaire $\frac{1}{6}$:

$$K = 1, \dots, 6 \quad ; \quad \mathbb{P}([X = K]) = \frac{1}{6}$$

et on peut que calculer : $E(X) = \frac{7}{2}$, $V(X) = \frac{35}{12}$.

Loi Bernoulli

Définition 1.3.

Soit une expérience dont le résultat est aléatoire et soit A un évènement défini sur cette expérience.

Soit X la variable aléatoire prenant la valeur 1 quand A est réalisé et 0 quand \bar{A} est réalisé .

On dit que X est une variable aléatoire de Bernoulli s'il existe p et q dans \mathfrak{R} vérifiant :

$$\mathbb{P}(X = 1) = p$$

$$\mathbb{P}(X = 0) = q$$

$$p + q = 1$$

on dit également que X suit une loi de Bernoulli $B(0, 1, p)$.

Propriétés 1.2.

Loi de probabilité :

$$\mathbb{P}(X = 1) = p$$

$$\mathbb{P}(X = 0) = q = 1 - p$$

Moyenne :

$$\begin{aligned} m_x &= E(X) = \sum_{i=1}^n p_i x_i \\ &= p \times 1 + q \times 0 = p \end{aligned}$$

Variance :

$$\begin{aligned} V_x &= E(X - m_x)^2 = E(X^2) - (E(X))^2 \\ &= \sum_{i=1}^n p_i x_i^2 - m_x^2 \\ &= p - p^2 \\ &= p(1 - q) \\ &= pq \end{aligned}$$

L'écart type :

$$\sigma(X) = \sqrt{pq}$$

Exemple 1.2.

Lancer d'une pièce de monnaie :

A : pile

\bar{A} : face

$$p = q = \frac{1}{2}$$

Suite de lancers de dé :

$$A = 1, 6$$

$$\bar{A} = 2, 3, 4, 5$$

$$p = \frac{1}{3} \quad \text{et} \quad q = \frac{2}{3}$$

Loi Binomiale

Définition 1.4.

Considérons un schéma de Bernoulli de probabilité p .

Le nombre de succès au cours d'une série de n épreuves est une variable aléatoire X .

La loi de probabilité de cette variable aléatoire est par définition la loi binomiale $B(0, n, p)$.

Propriétés 1.3.

Loi de probabilité :

$$\mathbb{P}(X = x) = \binom{n}{x} p^x$$

Moyenne :

$$m_x = E(X) = np$$

Variance :

$$V_x = E(X - m_x)^2 = E(X^2) - (E(X))^2 = npq$$

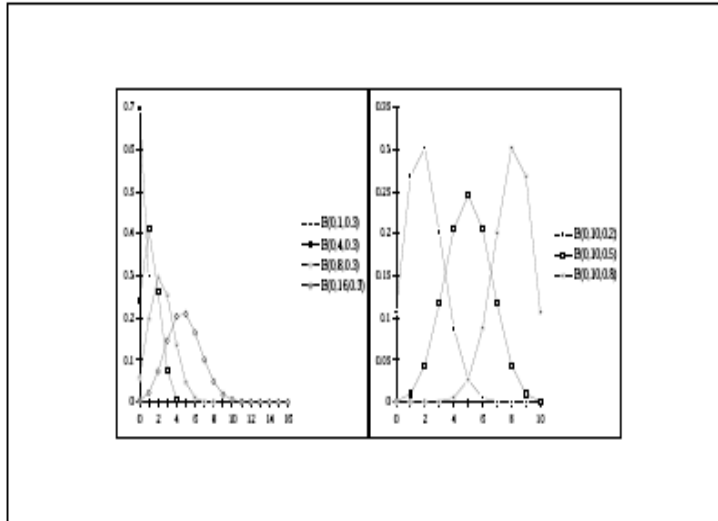


FIGURE 1.1 – Exemple de loi binomiale pour différentes valeurs des paramètres n et p .

où $n = \frac{n!}{x!(n-x)!}$ est le coefficient binomiale, c'est dire le nombre de combinaisons de x éléments choisis parmi n .

Remarque :

Si : $n > 50$ et $p < 0.1 \Rightarrow X \approx P(np)$

Si : $n > 50$ et $p > 0.1 \Rightarrow X \approx N(np, \sqrt{npq})$

Si :

$$\begin{cases} X_1 \sim B(n_1, p) \\ X_2 \sim B(n_2, p) \\ X_1 \amalg X_2 \end{cases} \Rightarrow$$

$$X_1 + X_2 \sim B(n_1 + n_2, p)$$

Translation d'une loi Binomiale :

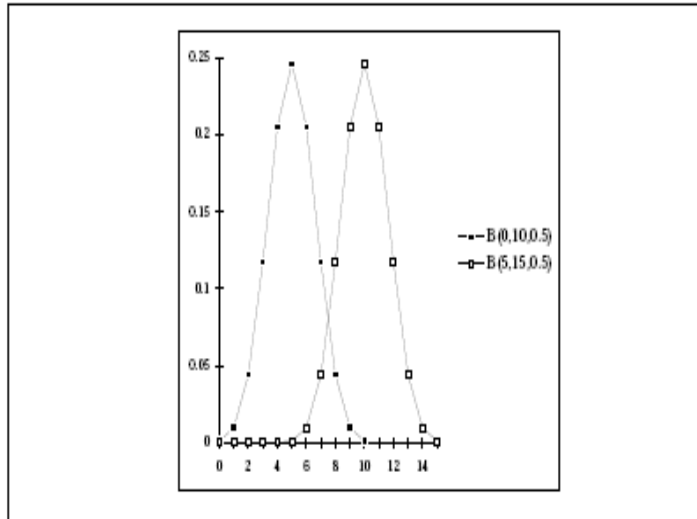


FIGURE 1.2 – Exemple de loi binomiale pour différentes valeurs du paramètre de translation d .

Si l'on ajoute un paramètre entier positif d de translation tel que $d < n$ on définit la loi binomiale $B(d, n, p)$:

$$\mathbb{P}(X = x) = (n - d)p^{(x-d)}q^{(n-x)}, \quad \forall x \in d, d + 1, \dots, d + n$$

$$m_x = d + (n - d)p$$

$$V_x = (n - d)pq$$

Exemple 1.3.

Une laboratoire doit analyser N échantillons E_1, E_2, \dots, E_N pour déterminer ceux qui contiennent un corps C .

C est présent dans l'un quelconque des échantillons E_1 , avec la probabilité p . Pour faire un nombre d'analyses inférieur à N , on utilise le protocole suivant :

On répartit les N échantillons en k groupes de n ($N = kn$) ; pour un groupe on réalise un prélèvement dans chacun des échantillons.

Les prélèvements sont mélangés, on crée ainsi un nouvel échantillon qu'on soumet à l'analyse (cet échantillon "résume le groupe").

Si le résultat est positif on analyse séparément chacun des échantillon du groupe.

Question : Combien d'analyses sont effectuées en moyenne ?

solution :

X désigne la variable aléatoire : "Nombre d'analyses effectuées"; pour chaque groupe, A est l'évènement : "Un, au moins des échantillons contient C ".

Y désigne la variable aléatoire : "Nombre de réalisation de A dans les k groupes".

On a : $X = k + ny$, Y suit la loi binomiale $B(k, p_A) \Rightarrow$

Déterminer p_A et on déduire les caractéristiques de Y puis de X .

La probabilité qu'aucun des n échantillons d'une groupe ne contienne le corps C est :

$$\mathbb{P}_{\bar{A}} = q^n$$

D'où :

$$\mathbb{P}_A = 1 - q^n$$

Alors :

$$\mathbb{P}(y = r) = \binom{K}{r} (1 - q^n)^r (1 - q^{n(k-r)})$$

Comme $X = k + ny$, on a $\mathbb{P}(X = k + nr) = \mathbb{P}(y = r)$, d'où :

$$\begin{aligned} E(X) &= \sum_{i=k}^{k+N} i\mathbb{P}(X = i) \\ &= \sum_{r=0}^k (k + nr)\mathbb{P}(X = k + nr) \\ &= \sum_{r=0}^k (k + nr)\mathbb{P}(y = r) \\ &= k \sum_{r=0}^k \mathbb{P}(y = r) + \frac{n}{\sum_{r=0}^k r\mathbb{P}(y=r)} \\ &= k + nE(y) \\ &= k + nk(1 - q^n) \end{aligned}$$

Prenons par exemple : $N = 200$, $n = 4$, $k = 50$ et $p = 0.05$

$$E(X) = 50(1 + 4(1 - 0.95)^4) = 87.14$$

Avec $n = 5$, on a :

$$E(X) = 40(1 + 5(1 - 0.95)^5) = 85.28$$

Le gain de travail, grace à ce protocole est, en moyenne, très important lorsque p est petit.

Loi binomiale négative

Définition 1.5.

Considérons un schéma de Bernoulli de probabilité p .

Le nombre d'épreuves pour obtenir r succès est une variable aléatoire X . La loi de probabilité de cette variable aléatoire et par définition la loi binomiale négative $NB(r, r, p)$.

Si r est un entier strictement positif, $NB(0, r, p)$, est la loi du nombre d'échecs avant d'avoir r .

Pour d quelconque, on note la loi binomiale négative $NB(d, r, p)$.

Propriétés 1.4.

Loi de probabilité :

$$P(X = x) = (x - d + r - 1)p^r q^{(x-d)}, \quad \forall x \in d, d + 1.$$

Moyenne :

$$m_x = d + \frac{rq}{p}$$

Variance :

$$V_x = \frac{rq}{p^2}$$

Une loi NB(d, r, p) est toujours étalée à droite et le coefficient d'asymétrie Y_1 est donc toujours positif.

Exemple 1.4.

Soit X_k la v.a. "nombre d'acides aminés lus jusqu'à la K - ième rencontre avec une molécule d'alanine".

Question 1 :

Déterminer les valeurs prises par les variables aléatoires X_k , en déduire les valeurs moyennes $\overline{X_k}$, pour les différentes valeurs de k (on prendra $k \leq 7$).

Question 2 :

On suppose que la synthèse du lisozyme s'effectue selon un schéma de Bernoulli.

On suppose tout d'abord que la probabilité de tirer un acide aminé est la même pour les 20 acides aminés.

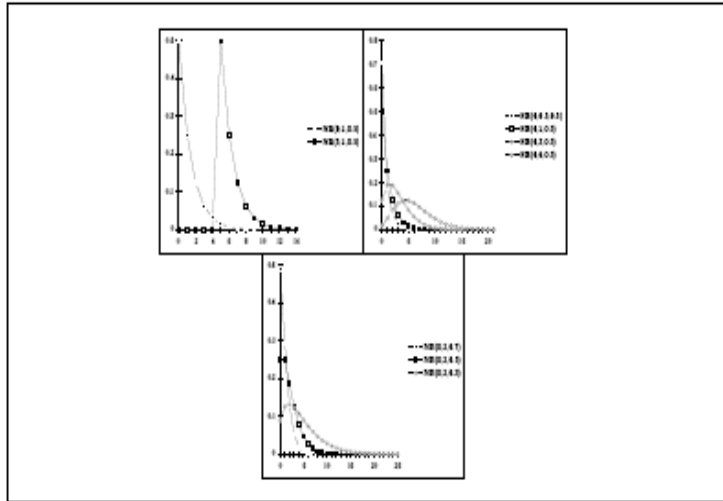


FIGURE 1.3 – Exemple de loi binomiale négative pour différentes valeurs du paramètre de translation d , du paramètre r et de la probabilité p .

Déterminer les valeurs moyennes prises par les variables aléatoires X_k théoriques. Que peut on en déduire ?

On suppose maintenant que les acides aminés sont utilisés proportionnellement à leur fréquence dans le sang.

On se pose les mêmes questions, avec $P(A) = 0.16$

Solution :

Valeurs observées de X_k pour $k \leq 7$:

$$K = 1 \quad X_1 = \{9, 17, 6, 10, 5, 26, 3, 7, 7, 2, 2, 2, 11, 3\}$$

$$K = 2 \quad X_2 = \{26, 16, 31, 10, 9, 4, 14\}$$

$$K = 3 \quad X_3 = \{32, 41, 17, 6\}$$

$$K = 4 \quad X_4 = \{42, 41, 13\}$$

$$K = 5 \quad X_5 = \{47, 45\}$$

$$K = 6 \quad X_6 = \{73, 23\}$$

$$K = 7 \quad X_7 = \{76, 34\}$$

Question 1 :

X_k suit une loi binomiale négative $NB(k, k, p)$, avec $p = \frac{1}{20}$.

k	1	2	3	4	5	6	7
$E(X_K)$	20	40	60	80	100	120	140
\overline{X}_k	7.86	15.72	24.00	32.00	46.00	48.00	55.00

\Rightarrow Désaccord entre observation et modèle. On rejète au moins une des hypothèse permettant de calculer $E(X_k)$.

Question 2 :

X_k suit une loi binomiale négative $NB(k, k, p)$, avec $p = 0.16$

k	1	2	3	4	5	6	7
$E(X_k)$	6.25	12.50	18.75	25.00	31.25	37.50	43.75
\overline{X}_k	7.86	15.72	24.00	32.00	46.00	48.00	55.00

Déffirences entre loi binomiale et loi binomiale négative :

Dans le cadre d'une suite d'épreuves Bernoulli indépendantes :

la loi binomiale $B(0, n, p)$ joue donc le rôle de loi de comptage.

la loi binomiale négative $NB(r, r, p)$ joue le rôle de loi d'intervalle (de temps si l'on assimile une épreuve de Bernoulli à une unité de temps).

Importance prise par ces notions de loi de comptage et de loi d'intervalle dans les processus stochastiques.

Loi hypergéométrique

Définition 1.6.

Soit une urne contenant N boules avec une proportion de p boules blanches et de $q = 1 - p$ boules noires.

On tire sans remise n boules. La variable aléatoire X égale au nombre de boules blanches obtenus suit une loi hypergéométrique $H(N, n, p)$.

Propriétés 1.5.

Loi de probabilité :

$$\mathbb{P}(X = x) = N_p \frac{N_q}{N}$$

Moyenne :

$$m_x = E(X) = np$$

Variance :

$$V_x = E(X - m_x)^2 = E(X^2) - (E(X))^2 = npq \frac{N - n}{N - 1}$$

Remarque :

$$\text{Si : } N \gg n (N > 10n) \Rightarrow X \approx \mathbf{B}(n, p)$$

Exemple 1.5.

Dans une population de N individus, on prélève à individus que l'on marque et relache dans la population. On effectue une seconde capture de n individus. Comment faire pour déterminer N ?

Solution :

Soit X le nombre d'individus marqués parmi les n capturés.

X suit une loi hypergéométrique.

Idée, estimée N à partir des valeurs expérimentales de X .

La méthode est appelée méthode de dilution ou de capture-recapture.

X suit une loi hypergéométrique de paramètre N, n, p d'où :

$$\mathbb{P}_N(X = k) = \frac{\binom{N_p}{K} \binom{N_q}{n-K}}{\binom{N}{n}}$$

Expérimentalement on relève la valeur $a = Np$ (ou $b = N - a$) d'où :

$$\mathbb{P}_N(X = k) = \frac{\binom{a}{K} \binom{b}{n-K}}{\binom{N}{n}}$$

On choisit N tel que $\mathbb{P}_N(X = k)$ soit maximum (méthode du maximum de vraisemblance).

On considère le rapport :

$$\frac{\mathbb{P}_N(X = k)}{\mathbb{P}_{N-1}(X = k)}$$

En supposant qu'on a relevé a , alors pour calculer $\mathbb{P}_{N,b} = N - a$ et pour calculer $\mathbb{P}_{N-1,b} = N - 1 - a$:

$$\begin{aligned} \frac{\mathbb{P}_N(X=k)}{\mathbb{P}_{N-1}(X=k)} &= \frac{(a)(N-a)}{N} \times \frac{(N-1)(a)}{N-1-a} \\ &= \frac{N-a}{N-a-n+k} \times \frac{N-n}{N} \\ &= \frac{N^2 - an - nN + an}{N^2 - aN - nN + kN} \end{aligned}$$

Remarque :

Le rapport $\frac{\mathbb{P}_N(X=k)}{\mathbb{P}_{N-1}(X=k)}$ est > 1 si $N < \frac{na}{k}$.

Le rapport $\frac{\mathbb{P}_N(X=k)}{\mathbb{P}_{N-1}(X=k)}$ est < 1 si $N > \frac{na}{k}$.

d'où $\mathbb{P}_N(X = k)$ est maximal pour $N = \frac{na}{k}$.

Application Au cours d'une première capture on prélève $a = 160$ vairons que l'on marque et relache. Au cour de la seconde capture, $n = 150$ vairons sont prélevés, dont 15 marqués. On estime donc à :

$$N = 160 \times \frac{150}{15} = 1600, \text{ Le nombre total de vairons.}$$

Loi de poisson

Définition 1.7.

Soit X une variable aléatoire pouvant prendre toutes les valeurs entières $0, 1, \dots, n, \dots$ vérifiant :

$$\mathbb{P}(X = x) = \frac{\exp(-\lambda)\lambda^x}{x!}$$

On dit que X suit une loi de poisson $\mathbb{P}(\lambda)$ de paramètre λ .

Propriétés 1.6.

Loi de probabilité :

$$\mathbb{P}(X = x) = \frac{\exp(-\lambda)\lambda^x}{x!}$$

Moyenne :

$$m_x = E(X) = \lambda$$

Variance :

$$V_x = E(X - m_x)^2 = E(X^2) - (E(X))^2 = \lambda$$

Remarque :

Si λ grand $\Rightarrow X \approx N(\lambda, \sqrt{\lambda})$

Si :

$$\left\{ \begin{array}{l} X_1 \sim \mathbb{P}(\lambda_1) \\ X_2 \sim \mathbb{P}(\lambda_2) \\ X_1 \amalg X_2 \end{array} \right. \Rightarrow$$

$$X_1 + X_2 \sim \mathbb{P}(\lambda_1 + \lambda_2)$$

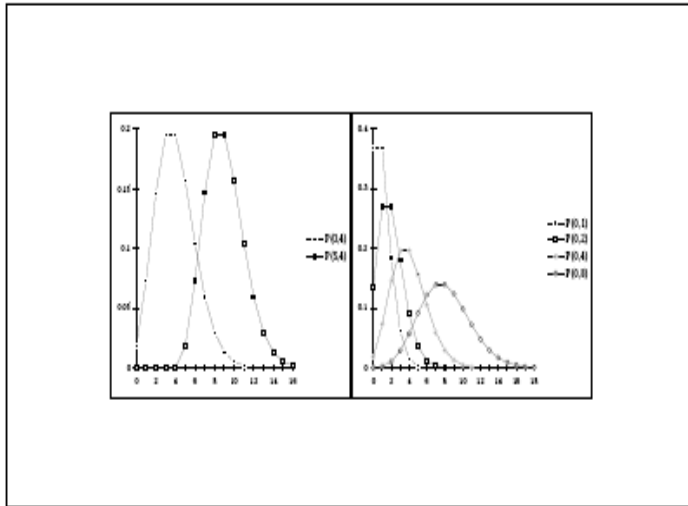


FIGURE 1.4 – Exemple de lois poisson pour différentes valeurs du paramètre d et λ .

Translation d'une loi de poisson :

Si l'on ajoute un paramètre entier positif d de translation, on définit la loi de poisson $\mathbb{P}(d, \lambda)$:

$$\mathbb{P}(X = x) = \frac{\exp(-\lambda)\lambda^{x-d}}{(x-d)!}; \quad \forall x \in d, d+1..$$

$$m_x = d + \lambda$$

$$V_x = \lambda$$

Pour une loi $\mathbb{P}(d, \lambda)$, le signe du coefficient d'asymétrie γ_1 est toujours positif.

Une loi $\mathbb{P}(d, \lambda)$ est donc : toujours étalée à droite.

Illustration du modèle :

Nombre de véhicules franchissant un poste de péage pendant une période de durée T .

Nombre de défauts dont est affecté un objet qui est fabriqué en série.

Nombre d'oeufs pondus au cours d'une ponte par certaines espèces animales.

Exemple 1.6.

On suppose que le nombre d'oeufs pondus par un insecte suit une loi de poisson de paramètre λ .

On suppose également que les oeufs pondus sont mutuellement indépendants et que la probabilité de développement d'un oeuf est P . Quelle est la probabilité pour que, parmi les oeufs pondus, il y ait exactement k survivants ?

Solution :

Il y a k survivants si et seulement si :

Le nombre d'oeufs pondus est supérieur (ou égal) à k .

On suppose que l'insecte pond n oeufs, avec $n \geq k$.

La probabilité pour que l'insecte pond n oeufs est :

$$\mathbb{P}(X = n) = \frac{\exp(-\lambda)\lambda^n}{n!}$$

Soit Y la variable aléatoire : nombre d'oeufs qui se développent parmi n oeufs pondus.

Soit A l'évènement : "l'oeuf considéré se développe".

Y suit une loi binomiale $B(n, p)$, d'où :

$$\mathbb{P}(Y = K) = \binom{n}{K} p^K q^{(n-K)}$$

La probabilité pour que l'insecte pondre n oeufs et que parmi ces n oeufs k se développe est :

$$\exp(-\lambda) \frac{\lambda^n}{n!} \times \binom{n}{k} p^k q^{(n-k)}$$

On considère l'évènement : " k oeufs survivent".

Cet évènement peut se réaliser d'une infinité de manières :

l'insecte pond $n = k$ oeufs et tous survivent.

l'insecte pond $n = k + 1$ oeufs et tous survivent sauf un.

l'insecte pond $n = k + 2$ oeufs et tous survivent sauf deux.

etc.

$$\begin{aligned} \mathbb{P}(k \text{ survivants}) &= \sum_{n=k}^{\infty} \exp(-\lambda) \frac{\lambda^n}{n!} \times \binom{n}{k} p^k q^{n-k} \\ &= \exp(-\lambda) p^k \sum_{n=k}^{\infty} \frac{\lambda^n}{n!} \frac{n!}{k!(n-k)!} q^{n-k} \\ &= \exp(-\lambda) p^k \frac{\lambda^k}{k!} \sum_{n=k}^{\infty} \frac{\lambda^{n-k}}{(n-k)!} q^{n-k} \\ &= \exp(-\lambda) p^k \frac{\lambda^k}{k!} \sum_{j=0}^{\infty} \frac{q \lambda^j}{j!} \\ &= \exp(-\lambda) p^k \frac{\lambda^k}{k!} \exp(\lambda q) \\ &= \exp(-\lambda + \lambda q) p^k \frac{\lambda^k}{k!} \\ &= \exp(-\lambda p) p^k \frac{\lambda^k}{k!} \end{aligned}$$

Il s'agit donc d'une loi de poisson de paramètre λp :

$$\mathbb{P}(k \text{ survivants}) = \exp(-\lambda p) p^k \frac{\lambda^k}{k!}$$

Loi géométrique

Définition 1.8.

On dit d'une variable aléatoire X suit une loi géométrique de paramètre p $G(0,p)$, si on a :

$$\mathbb{P}(X = k) = pq^{k-1}$$

Propriétés 1.7.

Loi de probabilité :

$$\mathbb{P}(X = x) = pq^{x-1}$$

Moyenne :

$$m_x = E(X) = \frac{1}{p}$$

Variance :

$$V_x = E(X - m_x)^2 = E(X^2) - (E(X))^2 = \frac{q}{p^2}$$

Exemple 1.7.

Dans une population de vanneaux, la probabilité de décès d'un oiseau au cours d'une année est constante et égale à $\frac{1}{3}$, quel est l'âge moyen d'une oiseau ?

Solution :

La durée de vie X d'une oiseau suit une distribution géométrique. Et l'on peut écrire :

$$\mathbb{P}(X = n) = pq^{n-1}$$

Avec $q = \frac{2}{3}$ et $p = \frac{1}{3}$, ainsi la probabilité qu'un vanneau meure à deux ans est :

$$\mathbb{P}(X = n) = \frac{2}{3} \times \frac{1}{3} = \frac{2}{9}$$

La durée moyenne de vie d'un vanneau est alors :

$$m_x = E(X) = \frac{1}{p} = 3ans$$

Chapitre 2

Rappels sur les lois de probabilités usuelles continues

2.1 Définitions et Propriétés

Définition 2.1.

Une variable aléatoire continue est un variable aléatoire qui peut prendre comme valeurs tout les nombres réels d'un certain intervalle I de \mathbb{R} est dit continue.

Et la variable aléatoire est dit continue si l'ensemble $X(\Omega)$ est un intervalle (ou une réunion d'intervalles) de \mathbb{R} .

Une variable aléatoire X est absolument continue si il existe $f(x)$ telle que sa fonction de répartition $F(x)$ est égale à :

$$\mathbb{P}(X \leq x) = F(x) = \int_{-\infty}^x f(t)dt$$

la fonction $f(x)$ est appelée densité de probabilité et on a :

$$f(x) = F'(x)$$

L'espérance mathématique ou moyenne de X est :

$$E(X) = \bar{X} = \int_{-\infty}^{\infty} t * f(t) dt$$

La variance de X est :

$$V(X) = \int_{-\infty}^{\infty} (t - E(X))^2 f(t) dt$$

L'écart type de X est :

$$\sigma(X) = \sqrt{V(X)}$$

2.2 Les lois continues

Loi uniforme

Définition 2.2.

On dit que la loi de probabilité d'une variable aléatoire réelle est uniforme sur un segment $[a; b]$, avec $0 \leq a < b$.

si sa densité de probabilité f est définie par :

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } \text{pour } x \in [a; b] \\ 0 & \text{si } \text{pour } x < a \text{ ou } x > b \end{cases}$$

On note alors $X \rightsquigarrow U([a; b])$.

f admet la représentation graphique de la Figure 2.1.

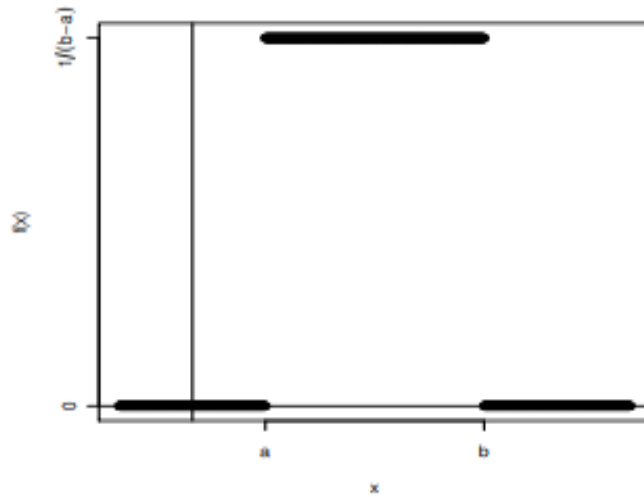


FIGURE 2.1 – La représentation graphique de la densité de probabilité d’une loi uniforme

Propriétés 2.1.

On a bien une densité de probabilité puisque :

$$f(x) \geq 0 \quad ; \quad \forall x \in \mathbb{R}.$$

f est continue sur $]-\infty; a[\cup]a; b[\cup]b; \infty[$.

$$\int_{-\infty}^{\infty} f(t)dt = \int_{-\infty}^a f(t)dt + \int_a^b f(t)dt + \int_b^{\infty} f(t)dt = 0 + 1 + 0 = 1.$$

Fonction de répartition :

On sait que : $F(x) = \mathbb{P}(\{X \leq x\}) = \int_{-\infty}^x f(t)dt$ donc si $X \rightsquigarrow U([a; b])$:

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases}$$

preuve :

On distingue trois cas :

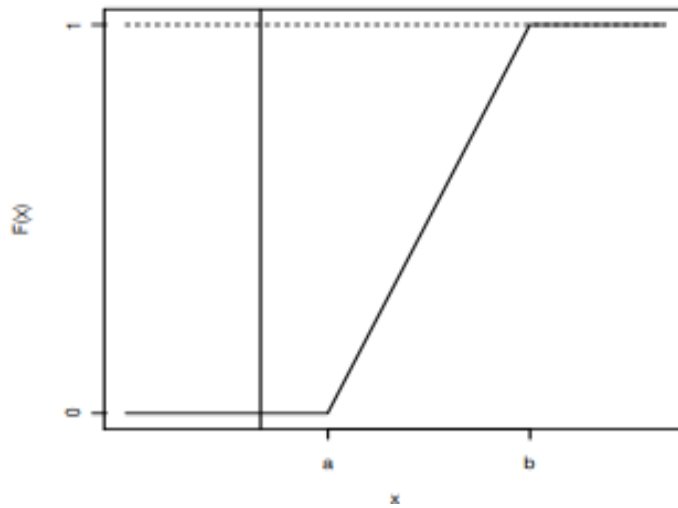


FIGURE 2.2 – La représentation graphique de la Fonction de répartition d'une loi uniforme

$$\text{Si } x < a, F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^x 0dt = 0.$$

$$\text{Si } a \leq x \leq b, F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^a 0dt + \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}.$$

$$\text{Si } x > b, F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^a 0dt + \int_a^b \frac{1}{b-a} dt + \int_b^x 0dt = 0 + 1 + 0 = 1.$$

Moments :

Si $X \rightsquigarrow U([a; b])$, alors :

$$E(X) = \int_{-\infty}^{\infty} tf(t)dt = \int_a^b \frac{t}{b-a} dt = \left[\frac{t^2}{2(b-a)} \right]_a^b = \frac{a+b}{2}$$

$$V(X) = \frac{(a^2 + ab + b^2)}{3} - \left(\frac{a+b}{2} \right)^2 = \frac{(b-a)^2}{12}$$

En effet, on a :

$$E(X^2) = \int_a^b \frac{t^2}{b-a} dt = \left[\frac{t^3}{3(b-a)} \right]_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

Exemple 2.1.

On choisit un nombre au hasard dans l'intervalle $[5; 20]$.

La variable X égale au nombre choisi, suit la loi uniforme sur $[5; 20]$.

On considère les évènements :

$A = (X > 10)$, $B = (12 \leq X \leq 16)$ et $c = (8 \leq X \leq 13)$.

Alors :

$$\mathbb{P}(A) = \frac{20-10}{20-5} \approx 0.67, \mathbb{P}(B) = \frac{16-12}{15} \approx 0.27, \mathbb{P}(C) = \frac{5}{15} \approx 0.33$$

$$\mathbb{P}(\bar{A}) = \mathbb{P}(X \leq 10) = 1 - \mathbb{P}(A) \approx 0.33$$

$$\mathbb{P}(B \cup C) = \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B \cap C) = \frac{4}{15} + \frac{5}{15} - \frac{1}{15} = \frac{8}{15}.$$

pour rappeler que les formules connues fonctionnent encore avec les lois continues.

Loi exponentielle

Définition 2.3.

Une variable aléatoire X suit une loi exponentielle de paramètre λ ($\lambda \in \mathbb{R}^{+*}$).

Si X est une variable aléatoire absolument continue dont la densité de probabilité est définie par :

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \lambda \exp(-\lambda x) & \text{si } x \geq 0 \end{cases}$$

On note alors $X \rightsquigarrow \varepsilon(\lambda)$.

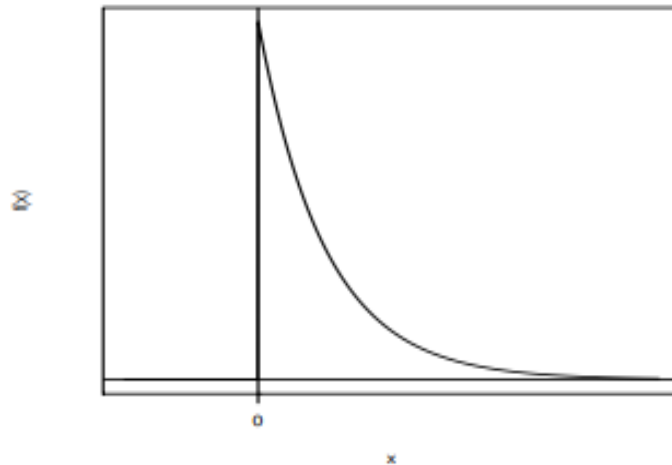


FIGURE 2.3 – La représentation graphique de la densité de probabilité d’une loi exponentielle

la fonction f admet la représentation graphique de la Figure 2.3.

Propriétés 2.2.

on a bien une densité de probabilité puisque :

$$f(x) \geq 0; \quad \forall x \in \mathbb{R}.$$

f est continue sur \mathbb{R}^+ et \mathbb{R}^{-} .*

$$\int_{-\infty}^{\infty} f(t)dt = \int_{-\infty}^0 0dt + \int_0^{+\infty} \lambda \exp(-\lambda t)dt.$$

$$\text{Or, si } A > 0, \int_0^A \lambda \exp(-\lambda t)dt = [-\exp(\lambda t)]_0^A = 1 - \exp(-\lambda A) \rightarrow 1$$

$$\text{Quand } A \rightarrow +\infty \text{ donc } \int_{-\infty}^{+\infty} f(t)dt = 1.$$

La loi exponentielle peut être considérée comme l’équivalent en continu de la loi géométrique dans le cas discret. En effet, elle modélise un temps d’attente du premier succès dans un processus de Poisson.

Fonction de répartition :

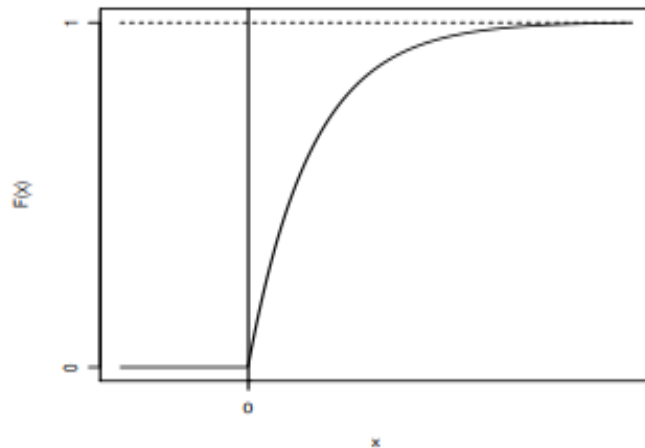


FIGURE 2.4 – La représentation graphique de la Fonction de répartition d'une loi exponentielle

Si $X \rightsquigarrow \varepsilon(\lambda)$, on a :

$$F(x) = \begin{cases} 0 & \text{si } pour x < 0 \\ 1 - \lambda \exp(-\lambda x) & \text{si } pour x \geq 0 \end{cases}$$

dont la représentation graphique est donnée à la Figure 2.4.

Moments :

Si $X \rightsquigarrow \varepsilon(\lambda)$:

$$E(X) = \int_{-\infty}^{\infty} t f(t) dt = \int_0^{\infty} \lambda t \exp(-\lambda t) dt = \frac{1}{\lambda}$$

à l'aide d'une intégration par parties :

$$V(X) = \frac{1}{\lambda^2} \quad \text{et} \quad \sigma(X) = \frac{1}{\lambda}$$

En effet, on sait que $V(X) = E(X^2) - (E(X))^2$, comme :

$$E(X^2) = \int_0^{+\infty} \lambda t^2 \exp(-\lambda t) dt$$

à l'aide de deux intégration par parties, on obtient :

$$E(X^2) = \frac{2}{\lambda^2}$$

Loi de Laplace-Gauss ou loi normale

Définition 2.4.

On appelle variable aléatoire normale ou gaussienne toute variable aléatoire absolument continue dont la densité de probabilité f est définie par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{\exp(-(x-m)^2)}{2\sigma^2}$$

m étant une constante réelle, σ une constante réelle strictement positive.

On utilise la notation suivante :

$$X \rightsquigarrow N(m, \sigma)$$

Propriétés 2.3.

On admettra que f est bien une densité de probabilité (la difficulté étant de montrer que $\int_{-\infty}^{+\infty} f(t) dt = 1$).

La courbe représentative de f est donnée par la Figure 2.5.

La courberr, dite courbe en cloche, a un axe de symétrie qui est la droite d'équation $x = m$.

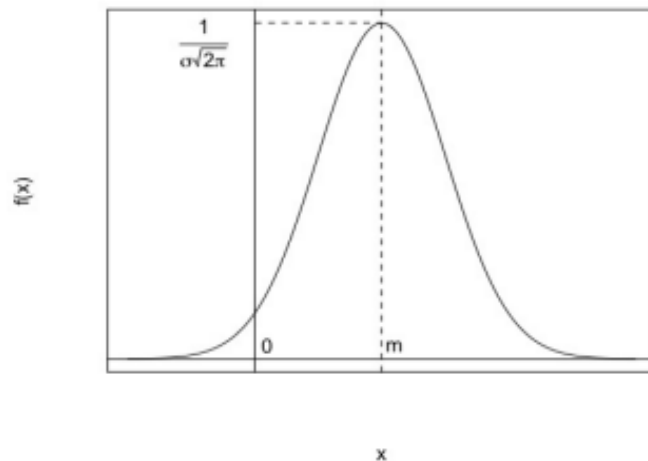


FIGURE 2.5 – La représentation graphique de la densité de probabilité d'une loi de Laplace-Gauss

La densité f a un maximum atteint pour $x = m$ valant $\frac{1}{\sigma}\sqrt{2\pi}$.

La courbe est d'autant plus pointue que σ est petit. Pour déterminer le(s) point(s) d'inflexion d'une fonction, on calcule sa dérivée seconde et on détermine le signe de cette dernière .

Si le signe change pour une abscisse particulière, la fonction y admet un point d'inflexion.

Déterminons le(s) point(s) d'inflexion de f . Le calcul de la dérivée première et Le calcul de la dérivée second donne :

par conséquent, la courbe admet deux points d'inflexion pour $x = m + \sigma$ et $x = m - \sigma$.

Soit $X \rightsquigarrow N(m, \sigma)$ et K une constante. On a les résultats suivants :

la variable KX suit une loi normale $N(Km, |K|\sigma)$.

la variable $K + X$ suit une loi normale $N(K + m, \sigma)$.

Somme de deux variables normales indépendantes :

Soient deux variables $X_1 \rightsquigarrow N(m_1, \sigma_1)$ et $X_2 \rightsquigarrow N(m_2, \sigma_2)$ indépendantes.

Alors :

$$X_1 + X_2 \rightsquigarrow N(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$

$$\text{et } X_1 - X_2 \rightsquigarrow N(m_1 - m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$

Plus généralement soient n variables aléatoires indépendantes deux à deux telles que $X_i \rightsquigarrow N(m_i, \sigma_i)$; $\forall i \in \{1, 2, \dots, n\}$, alors :

la variable $X = \sum_{i=1}^n a_i X_i$ suit une loi normale $N(m, \sigma)$ de moyenne

$$m = \sum_{i=1}^n m_i \text{ et d'écart -type } \sigma = \sqrt{(\sum_{i=1}^n a_i^2 \sigma_i^2)}$$

Si les X_i suivent la même loi $N(m, \sigma)$:

la variable $X = X_1 + X_2 + \dots + X_n$ suit une loi normale de moyenne nm et d'écart-type $\sqrt{\sigma^2 + \sigma^2 + \dots + \sigma^2} = \sqrt{\frac{n}{\sigma^2}} = \frac{\sigma}{n}$

La variable $Y = \frac{X_1 + X_2 + \dots + X_n}{n}$ suit une loi normale de moyenne

$$\frac{nm}{n} = m, \text{ d'écart-type } \frac{\sigma}{\sqrt{n}}.$$

Exemple 2.2.

Les variables normales sont très fréquentes, par exemple la variable aléatoire réelle "poids" d'un français adulte, la variable aléatoire "quotient intellectuel" d'une population donnée.

Moments :

L'espérance et la variance d'une variable normale sont respectivement données par :

$$E(X) = m \quad \text{et} \quad V(X) = \sigma^2$$

Variable normale centrée réduite : Si $m = 0$ et $\sigma = 1$.

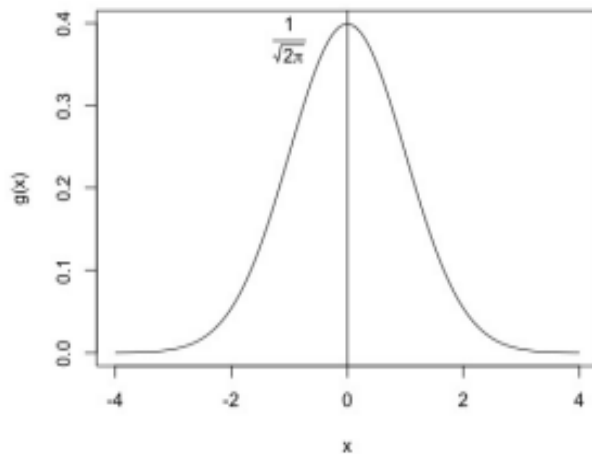


FIGURE 2.6 – La représentation graphique de la fonction g

La variable normale est appelée variable normale centrée réduite et est notée Z ou Γ on note alors $Z \rightsquigarrow N(0, 1)$.

La courbe représentative de la fonction g est donnée par la Figure 2.6.

Sa densité de probabilité est la fonction g définie par :

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

Cette fonction g est paire, la courbe a un axe de symétrie qui est la droite des ordonnées. En $x = 0$, la fonction g vaut $g(0) = \frac{1}{\sqrt{2\pi}}$.

Les points d'inflexion de la fonction g se trouvent en $x = -1$ et $x = 1$.

En général, les valeurs de $g(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$ sont données à l'aide d'une table pour $x \geq 0$.

Fonction de répartition :

Si on note π la fonction de répartition de la loi normale centrée réduite

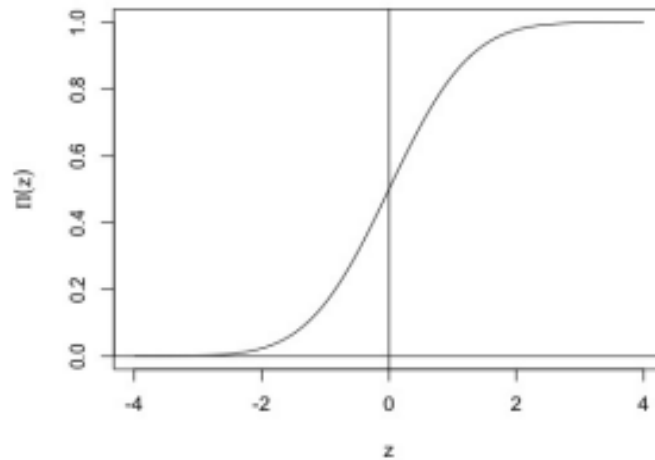


FIGURE 2.7 – La représentation graphique de la fonction de répartition π

Z associée à X :

$$\pi(z) = \int_{-\infty}^z g(x)dx = \mathbb{P}(\{Z \leq z\})$$

cette fonction est représentée graphiquement à la Figure 2.7. Il existe des tables qui donnent la valeurs de $\pi(z)$ pour $z \geq 0$. désigne l'aire du domaine plan en jaune (voir Figure 2.8).

Relation entre la fonction de répartition et la densité de probabilité des lois normales et lois normales centrées réduites :

Soit $X \rightsquigarrow N(m, \sigma)$ de densité f .

La fonction de répartition F est définie par $F(x) = \mathbb{P}(\{X \leq x\})$

et vérifie alors la relation :

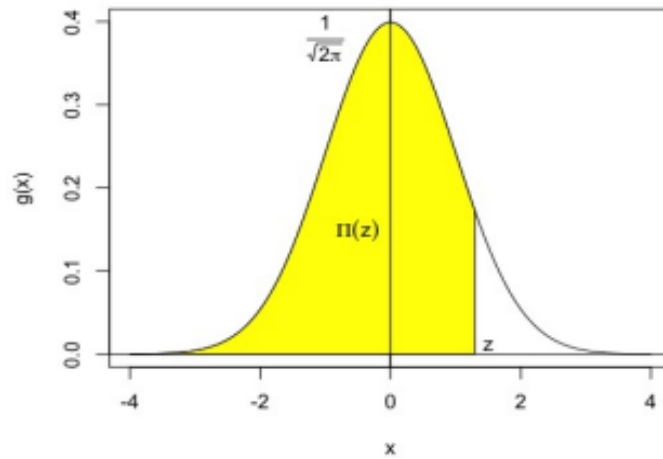


FIGURE 2.8 – La représentation graphique de la domaine de définition de la fonction de répartition π

$$F' = f$$

Soit $Z \rightsquigarrow N(0,1)$ de fonction de répartition π et de densité g alors :

$$\pi' = g$$

On a l'égalité : $\{X \leq x\} = \left\{ Z \leq \frac{(x-m)}{\sigma} \right\}$ ainsi que la relation :

$$F(x) = \pi\left(\frac{x-m}{\sigma}\right)$$

Par dérivation :

$$F'(x) = \frac{1}{\sigma} \pi\left(\frac{x-m}{\sigma}\right) \Leftrightarrow f(x) = \frac{1}{\sigma} g\left(\frac{x-m}{\sigma}\right)$$

ce qui permet l'utilisation de la table de densité de probabilité de la loi $N(0, 1)$ pour calculer les valeurs de la densité de probabilité de $N(m, \sigma)$.

Approximation d'une loi binomiale par une loi normale :

Soit une loi binomiale $B(n, p)$ de moyenne $m = np$, d'écart-type

$$\sigma = \sqrt{npq}.$$

On montre que l'on peut approximer la loi binomiale $B(n, p)$ par la loi normale $N(m = np, \sigma = \sqrt{npq})$ si $n \geq 15$, p et q étant non voisins de 0.

Dans la pratique, l'approximation est admise si $n \geq 20$, $np \geq 10$, $nq \geq 10$.

par exemple :

Soit $X \rightsquigarrow B(n = 100, p = 0.4)$ avec :

$$E(X) = 40 \text{ et } \sigma(X) = \sqrt{40 \times 0.6} = \sqrt{24}.$$

On a dans ce cas :

$$\left\{ \begin{array}{l} n = 100 \geq 20 \\ np = 40 \geq 10 \\ nq = 60 > 10 \\ npq = 24 \end{array} \right.$$

ce qui implique que $X \rightsquigarrow N(m = 40, \sigma = \sqrt{24})$. Calculons $\mathbb{P}(\{X < 38\})$.

On a :

$$\mathbb{P}(\{X = 38\})_B = \mathbb{P}(\{37.5 \leq X \leq 38.5\})_N$$

Posons :

$$Z = \frac{(X - 40)}{\sqrt{24}}$$

alors :

$$\{37.5 \leq X \leq 38.5\} = \left\{ \frac{-2.5}{\sqrt{24}} \leq Z \leq \frac{-1.5}{\sqrt{24}} \right\} = \{0.510 \leq Z \leq 0.306\}$$

Ainsi :

$$\begin{aligned} \mathbb{P}(\{X = 38\})_B &= \pi(-0.306) - \pi(-0.510) \\ &= \pi(-0.510) - \pi(-0.306) \\ &= 0.6950 - 0.6202 = 0.0748 \end{aligned}$$

Montrons que $\pi(0.306) = 0.6202$ à l'aide de l'interpolation linéaire :

on a : $\pi(0.30) = 0.6179$ et $\pi(0.31) = 0.6217$ et le tableau suivant :

0.6179	x	0.6217
0.30	0.306	0.31

$$\Leftrightarrow \frac{x - 0.6179}{0.6217 - 0.6179} = \frac{0.306 - 0.30}{0.31 - 0.30} \Leftrightarrow x = 0.6179 + 0.0038 \times \frac{0.006}{0.01} = 0.6202$$

Ensuite, $\mathbb{P}(\{X > 38\})_B = \mathbb{P}(\{X > 38.5\})_N$

$$\text{et } \{X > 38.5\} = \left\{ Z > \frac{(38.5-40)}{\sqrt{24}} \right\}$$

Par conséquent :

$$\begin{aligned} \mathbb{P}(\{X > 38\})_B &= \mathbb{P}(\{X > 38.5\})_N \\ &= 1 - \mathbb{P}(\{X \leq 38.5\})_N. \end{aligned}$$

ou encore :

$$\mathbb{P}(\{X > 38\})_B = 1 - \pi(-0.306) = \pi(0.306) = 0.6202$$

Enfin, $P(\{X \leq 38\})_B = P(\{X \leq 38.5\})_N$

$$\text{et } \{X \leq 38.5\} = \left\{ Z \leq \frac{(40-38.8)}{\sqrt{24}} \right\}.$$

Par conséquent :

$$P(\{X \leq 38\})_B = P\left(\left\{ Z \leq \frac{-1.5}{\sqrt{24}} \right\}\right)_B = \pi(-0.306)$$

Finalement :

$$P(\{X \leq 38\})_B = 1 - \pi(0.306) = 1 - 0.6202 = 0.3797$$

Loi du X^2 (Khi- deux) de Pearson

Définition 2.5.

On considère n variables indépendantes d'une loi normale centrée réduite T_1, T_2, \dots, T_n . La quantité :

$$T_1^2 + T_2^2 + \dots + T_n^2 = \sum_i^n T_i^2$$

est une variable aléatoire dont la distribution est celle d'un X^2 à n degrés de liberté.

Propriétés 2.4.

Moyenne : $E(X_n^2) = n$

Variance : $V(X_n^2) = 2n$

Lorsque n augmente, la densité f d'une loi du X^2 ressemble de plus en plus à la densité d'une loi normale (voir la Figure 2.9) : La variable X^2 est tabulée en fonction du nombre n de degrés de liberté.

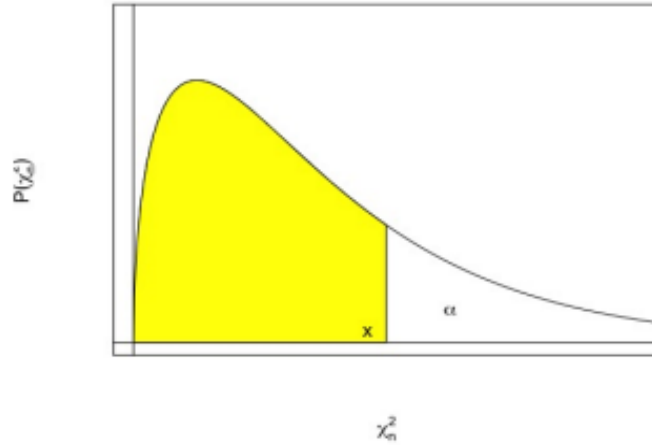


FIGURE 2.9 – La représentation graphique de la densité de probabilité d’une loi Khi- deux

La table donne pour différentes valeurs de α , la valeur de x telle que :

$$\mathbb{P}(\{X_n^2 < x\}) = 1 - \alpha$$

Remarque :

Attention, d’autres tables donnent la probabilité α , en fonction du nombre de degrés de liberté v pour qu’une variable aléatoire X suivant une loi de X_v^2 soit supérieure ou égale à une valeur donnée x :

$$\alpha = \mathbb{P}(\{X \geq x\}).$$

On a la propriété suivante :

$$X_m^2 + X_n^2 = X_{m+n}^2$$

Ce X^2 admet :

une moyenne :

$$E(X_{m+n}^2) = m + n$$

une variance :

$$\sigma^2(X_{m+n}^2) = 2(m + n)$$

et ceci par application directe du théorème sur l'addition de variables aléatoires indépendantes.

Exemple 2.3.

Calculer $\mathbb{P}(\{X_{10}^2 > 20.5\})$. On récupère à l'aide de la table, la probabilité :

$\mathbb{P}(\{X_{10}^2 < 20.5\}) = 0.975$. Par conséquent, la probabilité recherchée

$\mathbb{P}(\{X_{10}^2 > 20.5\})$ est égale à $1 - 0.975 = 0.025$.

Loi de Student -Fischer

Définition 2.6.

La loi de Student est une loi continue qui comme la loi du X^2 dépend d'un seul paramètre qu'on appellera également degré de liberté et qu'on note : $v (v \in \mathbb{N}^*)$. La variable X distribuée selon cette loi qu'on note :

$$X \rightsquigarrow t_v$$

Propriétés 2.5.

Prend toutes ses valeurs dans \mathbb{R} . Si $Y \rightsquigarrow N(0, 1)$ et $Z \rightsquigarrow X_v^2$, Y et Z étant indépendantes, la variable $X = \frac{Y}{\sqrt{\frac{Z}{v}}}$ suit une loi de Student à v degrés de liberté.

On dit qu'une variable aléatoire réelle à densité X a une loi de probabilité de Student à v degrés de liberté (n entier > 0), si et seulement si, sa densité de probabilité est donnée par la formule :

$$f_v(x) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v}\sqrt{\pi}\Gamma(\frac{v}{2})(1 + \frac{x^2}{v})^{\frac{v+1}{2}}}$$

Dans cette formule, Γ est la fonction Gamma d'Euler Définie, lorsque la partie réelle de x est positive, par :

$$\Gamma(x) = \int_0^{+\infty} \exp(-u)u^{x-1} du$$

La loi de Student à v degrés de liberté est la loi de probabilité du quotient d'une variable normale centrée réduite par la racine carrée de la somme des carrés de v variables normales centrées réduites indépendantes entre elles et indépendantes de la première variable.

Pour $v = 1$, la loi de Student s'appelle loi de Cauchy, ou loi de Lorentz. C'est la loi du rapport de deux variables normales centrées réduites indépendantes.

Courbes :

La courbe est unimodale, centrée, symétrique et plus plate que la courbe d'une loi normale. Lorsque le nombre de degrés de liberté augmente, la loi de Student tend vers la loi normale $N(0, 1)$ (voir Figure 2.10).

Moments :

Soit $X \rightsquigarrow t_v$, on a :

$$E(X) = 0 \quad \text{et} \quad V(X) = \frac{v}{v-2}, \text{ pour } v > 2.$$

Remarque :

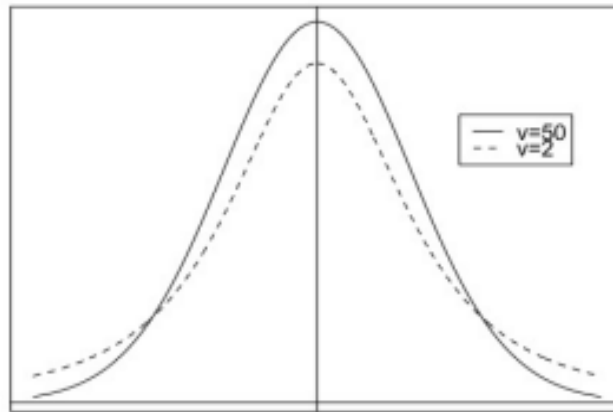


FIGURE 2.10 – La représentation graphique d'une loi student tend vers loi normale

Lorsque l'espérance existe, elle est nulle, puisque la loi est symétrique autour de 0.

Lorsque $v = 1$ ou $v = 2$, la variance n'est pas déterminée .

Lorsque v tend vers l'infini, la variance tend vers 1.

Tables :

Soit $X \rightsquigarrow t_v$. Il existe une table qui fournit les valeurs $t_{v,1-\alpha}$ pour v et α donnés, telles que :

$$\mathbb{P}(\{X < t_{v,1-\alpha}\}) = 1 - \alpha$$

Graphiquement, cette probabilité est donnée par la surface grisée de la Figure 2.11.

Exemple 2.4.

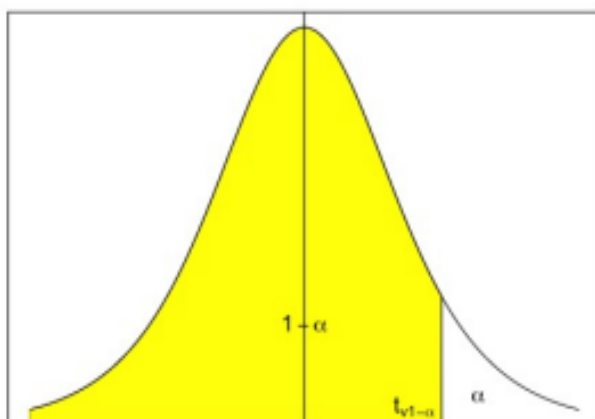


FIGURE 2.11 – La représentation graphique de la probabilité d'une loi student pour les valeurs v et α

$v = 10, \alpha = 0.1$ et $X \rightsquigarrow t_{10}$ donc $\mathbb{P}(\{X < 1.372\}) = 0.9$

et $t_{10;0.9} = 1.372$

$v = 20, \alpha = 0.05$ et $X \rightsquigarrow t_{20}$ donc $\mathbb{P}(\{X < 1.725\}) = 0.95$

et $t_{20;0.95} = 1.725$

Il existe une autre table qui fournit pour v et α donnés la valeur $t_{v,\alpha}$ telle que :

$$\mathbb{P}(\{-t_{v,\alpha} < X < t_{v,\alpha}\}) = 1 - \alpha$$

.

On remarque alors que :

$$\mathbb{P}(\{X < -t_{v,\alpha}\}) = \mathbb{P}(\{X > t_{v,\alpha}\}) = 1 - (1 - \frac{\alpha}{2}) = \frac{\alpha}{2}$$

.

Loi de Fisher-Snedecor

Définition 2.7.

La loi de Fisher-Snedecor est une loi continue dépendant de deux paramètres notés v_1 et v_2 , entiers naturels non nuls. La variable X distribué selon cette loi prend toutes ses valeurs dans \mathbb{R}^{+*} ou dans \mathbb{R}^+ .

Si $Y \rightsquigarrow X_{v_1}^2$ et $Z \rightsquigarrow X_{v_2}^2$, Y et Z étant indépendantes, la variable $X = \frac{Y}{Z}$ suit une loi de Fisher-Snedecor. On note :

$$X \rightsquigarrow F(v_1, v_2)$$

Propriétés 2.6.

La loi F de Fisher-Snedecor à (v_1, v_2) degrés de liberté est la loi de probabilité du rapport de deux variables de Khi-deux indépendantes divisées par leurs nombres de degrés de liberté (v_1 pour le numérateur, v_2 pour le dénominateur).

Pour $v_1 = 1$, la loi F de Fisher-Snedecor à $(1, v_2)$ degrés de liberté est la loi de probabilité du carré d'une variable de Student à v_2 degrés de liberté.

La densité de probabilité est, par définition :

$$f_{v_1, v_2}(x) = V_1^{\frac{v_1}{2}} V_2^{\frac{v_2}{2}} \frac{\Gamma(\frac{V_1+V_2}{2})}{\Gamma(\frac{V_1}{2})\Gamma(\frac{V_2}{2})} \frac{x^{\frac{v_1}{2}-1}}{(v_1x + v_2)^{\frac{v_1+v_2}{2}}}.$$

pour $x > 0$, v_1 et $v_2 \in \mathbb{N}^*$.

Dans cette formule, Γ est la fonction Gamma d'Euler définie, lorsque la partie réelle de x est positive, par :

$$\Gamma(x) = \int_0^{+\infty} \exp(-u)u^{x-1} du.$$

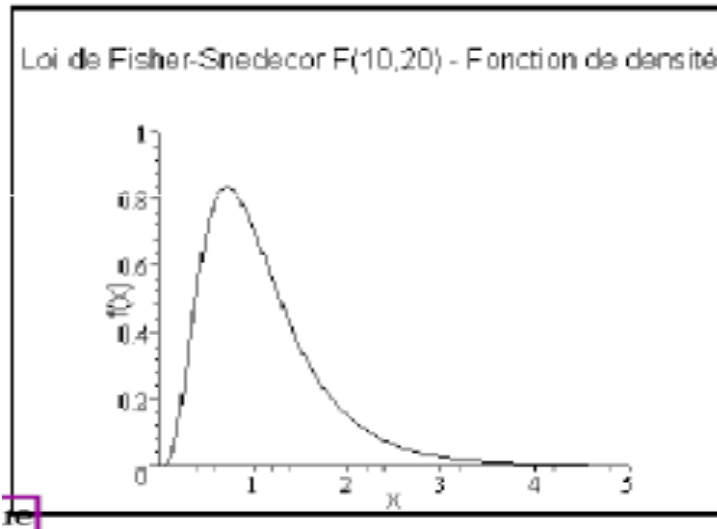


FIGURE 2.12 – La représentation graphique de la densité de probabilité d’une loi de Fischer-Snedecor

La fonction $f(v_1, v_2)$ est bien une densité de probabilité sur $]0; +\infty[$, car :
ses valeurs sont positives.

la fonction est intégrable et son intégrale est donnée par :

$$\int_0^{\infty} f_{v_1, v_2}(x) = V_1^{\frac{v_1}{2}} V_2^{\frac{v_2}{2}} \frac{\Gamma(\frac{V_1+V_2}{2})}{\Gamma(\frac{V_1}{2})\Gamma(\frac{V_2}{2})} \int_0^{\infty} \frac{x^{\frac{v_1}{2}-1}}{(v_1x + v_2)^{\frac{v_1+v_2}{2}}} dx$$

Pour calculer l’intégrale :

$$I = \int_0^{\infty} \frac{x^{\frac{v_1}{2}-1}}{(v_1x + v_2)^{\frac{v_1+v_2}{2}}} dx$$

on pose :

$$t = \frac{v_1x}{v_1x + v_2} \Rightarrow dx = \frac{n_2}{n_1} \frac{dt}{1-t^2}.$$

De plus :

$$v_1 x + v_2 = v_2 \times \frac{1}{1-t}$$

ce qui implique lorsque $x = 0$, $t = 0$ et lorsque x tend vers l'infinie, t tend vers 1. Par conséquent :

$$I = \int_0^1 \left(\frac{v_2}{v_1} \frac{t}{1-t}\right)^{\frac{v_1}{2}-1} \left(\frac{1-t}{v_2}\right)^{\frac{v_1+v_2}{2}} \frac{v_2}{v_1} \frac{dt}{1-t^2} = V_1^{-\frac{v_1}{2}} V_2^{-\frac{v_2}{2}} \int_0^1 t^{\frac{v_1}{2}-1} (1-t)^{\frac{v_2}{2}-1}$$

Dans l'intégrale, on reconnaît la fonction Beta d'Euler définie, lorsque des parties réelles de x et de y sont positives, par :

$$B(x, y) = \int_0^1 u^{x-1} (1-u)^{y-1} du = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

donc $\int_0^1 t^{\frac{v_1}{2}-1} (1-t)^{\frac{v_2}{2}-1} dt = B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)$ ce qui implique que :

$$\int_0^{+\infty} f_{(v_1, v_2)}(x) = V_1^{\frac{v_1}{2}} V_2^{\frac{v_2}{2}} \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} V_1^{-\frac{v_1}{2}} V_2^{-\frac{v_2}{2}} \frac{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)}{\Gamma\left(\frac{v_1+v_2}{2}\right)} = 1$$

L'intégrale de $f_{(v_1, v_2)}$ est bien égale à 1, ce qui montre que $f_{(v_1, v_2)}$ est bien une densité de probabilité.

Si $X \rightsquigarrow F(v_1, v_2)$ la variable $\frac{1}{X} \rightsquigarrow F(v_2, v_1)$ donc :

$$F(v_1, v_2, 1-\alpha) = \frac{1}{F(v_2, v_1, \alpha)}$$

Courbes :

On a représenté ci-dessus (Figure 2.13) la loi F de Fischer-Snedecor pour diverses valeurs de v_1 et de v_2 .

Moments :

Soit $X \rightsquigarrow F(v_1, v_2)$.

Pour $v_2 > 2$, l'espérance est définie par :

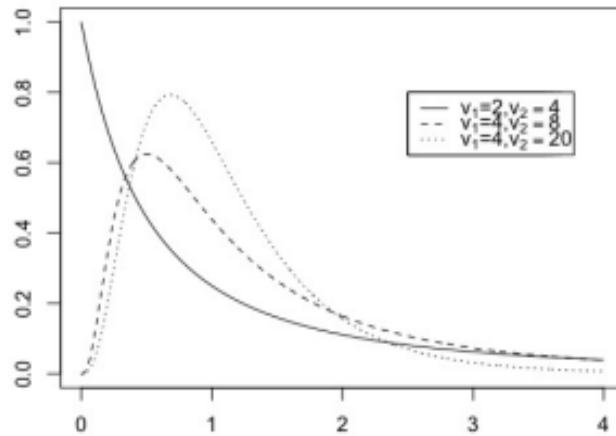


FIGURE 2.13 – La représentation graphique de loi F de de Fischer-Snedecor pour diverses valeurs de v_1 et de v_2

$$E(X) = \frac{v_2}{v_2 - 2}$$

Remarque :

Pour $v_2 \leq 2$, l'espérance n'est pas déterminée.

Pour $v_2 > 4$, la variance est définie par :

$$V(x) = \frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2 - 2)^2(v_2 - 4)}$$

Remarque :

Pour $v_2 \leq 4$, la variance n'est pas déterminée.

Tables :

Soit $X \rightsquigarrow F(v_1, v_2)$. La table fournit, pour $\alpha = 0.025$, pour v_1 et v_2 donnés, les valeurs $F_{v_1, v_2, 1-\alpha}$ telles que $\mathbb{P}(\{X \leq F_{v_1, v_2, 1-\alpha}\}) = 1 - \alpha$.

Cette table sert à la comparaison des variances de deux populations à partir de deux échantillons.

Chapitre 3

Sur les lois à symétrie elliptique

3.1 Introduction

Beaucoup de résultats en analyse statistique multidimensionnelle sont obtenus sous des hypothèses de normalité. or il s'avère que, pour certains d'entre eux, la propriété fondamentale qui intervient dans leur démonstration est l'invariance de la loi normale par rotation (ou plus généralement par transformation orthogonale).

Philoche([5]) montre que le classique test F du modèle linéaire reste valide dans le cas de lois invariantes par transformation orthogonale.

De nombreuses propriétés des lois invariantes par rotation ont été obtenues par Eaton dans ([3]).

3.2 Définitions et propriétés

Lois radiales et Lois à symétrie elliptique

Définition 3.1.

Soit v un produit scalaire sur E^* . Une mesure (resp. une probabilité) sur E est dite radiale (resp. loi radiale) de paramètre de dispersion v , si elle est invariante par toute transformation v^{-1} – orthogonale.

On suppose le paramètre de dispersion est un produit scalaire sur E^* se justifie par le fait, que nous démontrerons plus loin, que si une loi radiale admet un moment d'ordre 2 ce dernier est un paramètre de dispersion.

Définition 3.2.

Une mesure (resp. une probabilité) sur E est une mesure (resp. une loi) à symétrie elliptique sur E , de paramètre de position $\lambda \in E$ et de paramètre de dispersion v si elle est l'image, par la translation de vecteur λ , d'une mesure (resp. loi) radiale sur E de paramètre de dispersion v .

Remarque 3.1.

Si une mesure à symétrie elliptique sur E admet v pour paramètre de dispersion, elle admet aussi pour paramètre de dispersion αv pour tout $\alpha \in \mathbb{R}_+^*$.

On verra plus loin qu'en fait le paramètre de dispersion est définie à un facteur multiplicatif près.

Etant donné le lien entre mesures radiales et mesures à symétrie elliptique, nous nous limiterons par la suite à l'étude des propriétés des mesures radiales. Celles des mesures à symétrie elliptiques s'en déduisent facilement.

Proposition 3.1.

1. Une loi de probabilité P sur E est une loi radiale de paramètre de dispersion v si et seulement si sa fonction caractéristique Φ_p factorise à travers $\|\cdot\|^2$. Il

existe donc alors une fonction Ψ_p de \mathbb{R}_+ dans \mathbb{C} telle que :

$$\forall t \in E \quad \varphi_p(t) = \Psi_p(\|t\|^2)$$

2. Dans ce cas, les lois images de P par toutes les formes linéaires sur E de norme 1 sont les mêmes et leur fonction caractéristique est l'application :

$$a \rightsquigarrow \Psi_p(a^2).$$

Démonstration :

1. a) **Condition nécessaire**

Soient $t \in E$ et $t' \in E$ tels que $\|t\| = \|t'\|$. Il existe alors une transformation v^{-1} -orthogonale g de E , telle que $g(t') = t$. Nous pouvons écrire :

$$\begin{aligned} \varphi_p(t) &= \mathbb{E}_p[\exp(i \langle t, \cdot \rangle)] \\ &= \mathbb{E}_p[\exp(i \langle g(t'), \cdot \rangle)] \\ &= \mathbb{E}_p[\exp(i \langle t', g^{-1}(\cdot) \rangle)] \\ &= \mathbb{E}_{p_{g^{-1}}}[\exp(i \langle t', \cdot \rangle)] \\ &= \mathbb{E}_p[\exp(i \langle t', \cdot \rangle)] \\ &= \varphi_p(t'). \end{aligned}$$

La condition nécessaire en découle.

b) **Condition suffisante**

<Supposons qu'il existe Ψ_p de \mathbb{R}_+ dans \mathbb{C} telle que :

$$\forall t \in E \quad \varphi_p(t) = \Psi_p(\|t\|^2)$$

Soit g une transformation v^{-1} – *orthogonale*. Nous avons, pour tout $t \in E$:

$$\varphi_{p_g}(t) = \varphi_p(g^{-1}(t)) = \Psi_p(\|g^{-1}(t)\|^2) = \Psi_p(\|t\|^2) = \varphi_p(t)$$

donc $P_g = P$.

2. Soit $f \in E^*$ telle que $\|f\| = 1$. On désigne par t le vecteur de E tel que $f = \langle t, \cdot \rangle$.

Soient F le sous-espace de E engendré par t , π le projecteur v^{-1} – *orthogonal* de E sur F et ξ l'isomorphisme de \mathbb{R} sur F ($\xi(a) = at$).

Il est clair que $f = \xi^{-1} \circ \pi$. Pour tout $s \in E$, on a :

$$\varphi_{p_\pi}(s) = \varphi_p(\pi(s))$$

en vertu de la remarque et du fait que π est autoadjoint.

En particulier, pour tout $a \in \mathbb{R}_+$ et $s = at$:

$$\varphi_{p_\pi}(at) = \varphi_p(at) = \Psi_p(\|at\|^2) = \Psi_p(a^2).$$

Pour tout $a \in \mathbb{R}$, on a alors :

$$\begin{aligned} \varphi_{p_f}(a) &= \mathbb{E}_{p_f}[\exp(ia \cdot)] \\ &= \mathbb{E}_{p_\pi}[\exp(ia\xi^{-1}(\cdot))] \\ &= \mathbb{E}_{p_\pi}[\exp(i \langle at, \xi^{-1}(\cdot) \rangle)] \\ &= \mathbb{E}_{p_\pi}[\exp(i \langle at, \cdot \rangle)] \\ &= \varphi_{p_\pi}(at) \\ &= \Psi_p(a^2). \end{aligned}$$

Proposition 3.2.

Soit P une loi de probabilité radiale sur E .

Sauf si P est la loi de Dirac en 0_E , le paramètre de dispersion de P est défini à un facteur multiplicatif près.

Démonstration

Soient v_1 et v_2 deux paramètres de dispersion de P . Notons $\|\cdot\|_1$ et $\|\cdot\|_2$ les normes associées.

Il existe une base de E , (e_1, \dots, e_n) , qui est à la fois v_1^{-1} – orthonormale et v_2^{-1} – orthogonale.

Soient $\lambda_1, \dots, \lambda_n$ les coefficients diagonaux de la matrice diagonale de v_2^{-1} dans cette base (pour $1 \leq i \leq n$ on a $\lambda_i > 0$). Soient :

$$m = \min_{1 \leq k \leq n} \lambda_k \text{ et } i \text{ un indice tel que } \lambda_i = m$$

et

$$M = \max_{1 \leq k \leq n} \lambda_k \text{ et } j \text{ un indice tel que } \lambda_j = M$$

Il suffit de démontrer que si $m < M$ alors P est la mesure de Dirac en 0_E .

Nous définissons les endomorphismes f et g de E par :

$$\forall k \quad (1 \leq k \leq n, k \neq i, k \neq j) \quad f(e_k) = g(e_k) = e_k$$

$$f(e_i) = e_j, f(e_j) = e_i, \quad g(e_i) = \sqrt{\frac{m}{M}}e_j \quad \text{et} \quad g(e_j) = \sqrt{\frac{M}{m}}e_i$$

Il est clair que f est v_1^{-1} – orthogonale et que g est v_2^{-1} – orthogonale.

Supposons $m < M$. Soit $t \neq 0_E$ dans E et soit x_0 le vecteur de E dont les composantes dans la base (e_1, \dots, e_n) sont $(\|t\|_1 \delta_{kj})_{1 \leq k \leq n}$.

Nous avons $\|x_0\|_1 = \|t\|_1$. Définissons alors la suite $(x_n)_{n \in \mathbb{N}}$ de vecteurs de E par :

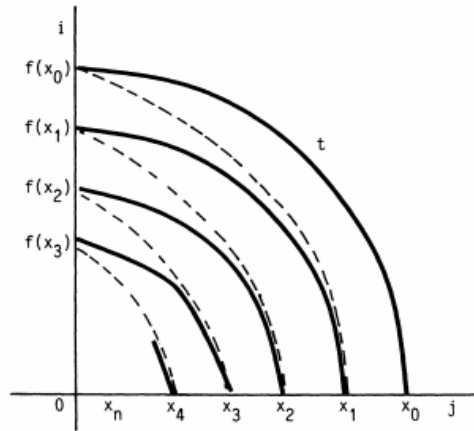


FIGURE 3.1 – La représentation graphique de la suite x_n

$$\forall n \in \mathbb{N}^* \quad ; \quad x_n = g \circ f(x_{n-1})$$

présenté par le figure 3.1

Alors, pour tout $n \in \mathbb{N}^*$:

$$X_n = \left(\frac{m}{M}\right)^{\frac{n}{2}} \quad ; \quad X_{n-1} = \left(\frac{m}{M}\right)^{\frac{n}{2}} X_0$$

et par conséquent :

$$\lim_{n \rightarrow +\infty} X_n = 0_E$$

En outre :

$$\begin{aligned} \varphi_p(X_n) &= \varphi_p(g \circ f(X_{n-1})) \\ &= \varphi_p(f(X_{n-1})) \end{aligned}$$

car g est v_2^{-1} – orthogonale et P est v_2 – radiale.

Alors :

$$\varphi_p(X_n) = \varphi_p(X_{n-1})$$

car f est v_1^{-1} – orthogonale et P est v_1 – radiale.

Il en résulte que, pour tout $n \in \mathbb{N}$:

$$\varphi_p(X_n) = \varphi_p(X_0) = \varphi_p(t)$$

et donc que la suite $(\varphi_p(X_n))_{n \in \mathbb{N}}$ est constante.

Par conséquent :

$$1 = \varphi_p(0_E) = \lim_{n \rightarrow +\infty} \varphi_p(X_n) = \varphi_p(t).$$

L'égalité précédente ayant lieu pour tout $t \in E$, P est la loi de Dirac en 0_E .

En définitive nous avons montré que si P n'est pas la loi de Dirac en 0_E alors $m = M$ et donc $v_2^{-1} = mv_1^{-1}$.

3.3 Exemples de mesures et de lois radiales

La mesure de Lebesgue sur E

Définition 3.3.

Soit v un produit scalaire sur E^* . Soit (e_1, \dots, e_n) une base v_1^{-1} –orthonormale de E et ξ l'isomorphisme naturel de \mathbb{R}^n dans E défini par :

$$\forall (X_1, \dots, X_n) \in \mathbb{R}^n \quad \xi(X_1, \dots, X_n) = \sum_{i=1}^n X_i e_i.$$

Il est clair que ξ est une isométrie bimesurable de \mathbb{R}^n sur E .

On définit alors la mesure de Lebesgue sur E , notée λ_v , comme étant la mesure image par ξ de la mesure de Lebesgue sur \mathbb{R}^n .

L'invariance de la mesure de Lebesgue sur \mathbb{R}^n par transformation orthogonale implique d'une part que λ_v est une mesure radiale sur E admettant v pour paramètre de dispersion et d'autre part que λ_v ne dépend pas de la base v_1^{-1} – orthonormale choisie.

On vérifie aisément que, si v_1 et v_2 sont deux produits scalaires sur E^* alors λ_{v_1} et λ_{v_2} sont deux mesures équivalentes.

Loi normale n-dimensionnelle

Définition 3.4.

On appelle loi normale sur E toute loi de probabilité P sur E dont l'image par toute forme linéaire sur E est une loi normale sur \mathbb{R} . Une telle loi P admet un moment d'ordre 1 et un moment d'ordre 2.

Supposons P centrée et calculons sa fonction caractéristique :

$$\begin{aligned}\forall t \in E^* \quad \varphi_p(t) &= \varphi_{p_t}(1) \\ &= \exp\left(-\frac{1}{2} \int_{\mathbb{R}} X^2 P_t(dX)\right) \\ &= \exp\left(-\frac{1}{2} \mathbb{E}_p(t^2)\right) \\ &= \exp\left(-\frac{1}{2} v(t, t)\right)\end{aligned}$$

où v désigne la variance de P .

Si v est définie positive, alors d'après la proposition 3.1, la loi P est radiale de paramètre de dispersion v . Dans le cas où P n'est pas centrée, si m désigne la moyenne, alors la loi P est une loi à symétrie elliptique notée $N_E(m, v)$.

Loi de Cauchy n-dimensionnelle

Définition 3.5.

Soit v un produit scalaire sur E^* . Une probabilité P sur E est une loi de Cauchy de paramètre v si pour toute forme linéaire t sur E , la loi image P_t est une loi de Cauchy sur \mathbb{R} de paramètre d'échelle $\sqrt{v(t, t)}$.

Calculons la fonction caractéristique d'une telle loi. Pour tout $t \in E^*$:

$$\varphi_p(t) = \varphi_{p_t}(1) = \exp(-\sqrt{v(t, t)})$$

Par conséquent P est loi radiale admettant v comme paramètre de dispersion. Cette loi admet pour densité relativement à la mesure de Lebesgue λ_v sur E l'application $y \rightsquigarrow \frac{K}{(1+\|y\|^2)^{\frac{n+1}{2}}}$.

où $n = \dim E$ et K est une constante de normalisation.

Loi uniforme sur une sphère n-dimensionnelle

Soit v un produit scalaire sur E^* .

Proposition 3.3.

Il existe sur S_v une unique loi radiale de E de paramètre de dispersion v appelée loi uniforme sur S_v et notée U_v .

Démonstration

Nous adaptons celle donnée par J.L. Philoche([5]) dans le cas $E = \mathbb{R}^n$.

1. Existence :

Soit λ_v la mesure de Lebesgue sur E associée à v . Soit N l'application de $B_v - \{0_E\}$ dans S_v définie par :

$$\forall X \in B_v - \{0_E\} \quad N(X) = \frac{1}{\|X\|} \cdot X$$

Soit U_v la mesure de probabilité sur E définie par :

$$\forall A \in \mathbb{B}(E) \quad U_v(A) = \frac{1}{\lambda_v(B_v)} \lambda_v(N^{-1}(A \cap S_v))$$

autrement dit U_v est la normalisée de la mesure image par N de la trace de λ_v sur B_v .

Montrons que U_v est radiale de paramètre de dispersion v .

Soit g une transformation v_1^{-1} – orthogonale de E . Il est clair que

$Nog = goN$. Pour tout $A \in \mathbb{B}(E)$:

$$\begin{aligned} U_v(g^{-1}(A)) &= \frac{1}{\lambda_v(B_v)} \lambda_v[N^{-1}(g^{-1}(A) \cap S_v)] \\ &= \frac{1}{\lambda_v(B_v)} \lambda_v[N^{-1}(g^{-1}(A) \cap g^{-1}(S_v))] \\ &= \frac{1}{\lambda_v(B_v)} \lambda_v[N^{-1}(g^{-1}(A \cap S_v))] \\ &= \frac{1}{\lambda_v(B_v)} \lambda_v[g^{-1}(N^{-1}(A \cap S_v))] \\ &= \frac{1}{\lambda_v(B_v)} \lambda_v(N^{-1}(A \cap S_v)) \\ &= U_v(A) \end{aligned}$$

Par conséquent U_v est radiale de paramètre de dispersion v .

2. Unicité :

Sa démonstration fait appel à la théorie de la mesure de Haar.

Le groupe O_v des transformations v_1^{-1} – orthogonales de E est un groupe topologique compact. Il en découle qu'il existe une unique probabilité v invariante par les translations à gauche et à droite (voir Nachbin). v est appelée la mesure de Haar de O_v .

Soit $\mathcal{C}(S_v)$ l'ensemble des fonctions continues sur S_v à valeurs réelles. Pour tout $f \in \mathcal{C}(S_v)$, pour tout $g \in O_v$ et tout $x \in S_v$, on définit $f_x(g)$ et $f_g(x)$ par :

$$f_x(g) = f_g(x) = f(g^{-1}(x))$$

Puisque O_v opère transitivement sur S_v , pour tout $f \in \mathcal{C}(S_v)$ l'intégrale $\int_{O_v} f_x(g)dv(g)$ ne dépend pas de $x \in S_v$. On peut donc définir sur S_v une loi de probabilité Q par :

$$\forall f \in \mathcal{C}(S_v) \quad \int_{S_v} f dQ = \int_{O_v} f_x dv.$$

Soient P une loi v – radiale sur S_v et $f \in \mathcal{C}(S_v)$. Alors :

$$\begin{aligned} \int_{S_v} f(x) dP(x) &= \int_{O_v} (\int_{S_v} f(x) dP(x)) dv(g) \\ &= \int_{O_v} (\int_{S_v} f_g(x) dP(x)) dv(g) \\ &= \int_{S_v} (\int_{O_v} f_x(g) dv(g)) dP(x) \\ &= \int_{O_v} f_x(g) dv(g) \\ &= \int_{S_v} f dQ. \end{aligned}$$

Par conséquent $P = Q$ ce qui établit l'unicité.

Définition 3.6.

Soit $r \in \mathbb{R}_+$. On appelle loi uniforme sur $S_{v,r}$ la loi image, par l'homothétie de rapport r , de la loi uniforme sur S_v . On la note $U_{v,r}$.

3.4 Propriétés élémentaires des lois radiales

Propriété 3.1.

Tout mélange de lois radiales sur E est une loi radiale sur E .

Propriété 3.2.

Soit P une loi radiale sur E . Si P admet un atome en $a \in E$, alors $a = 0_E$.

Démonstration :

Supposons $a \neq 0_E$. Considérons la sphère $S(0_E, \|a\|)$ de centre 0_E et de rayon $\|a\|$. Pour tout $x \in S(0_E, \|a\|)$, on a $P(\{x\}) = P(\{a\}) > 0$ et ainsi $P(S(0_E, \|a\|)) = +\infty$ ce qui est absurde.

Proposition 3.4.

Soit P une loi radiale sur E . Si P est une probabilité d'ordre 2 alors P admet son moment d'ordre 2 pour paramètre de dispersion.

Démonstration :

Soit v un paramètre de dispersion de P . D'après la proposition 3.1 il existe une fonction Ψ_p de \mathbb{R}_+ dans \mathbb{C} telle que :

$$\forall t \in E \quad \varphi_p(t) = \Psi_p \circ g(t)$$

où $g(t) = \|t^2\| = v^{-1}(t, t)$. La fonction g est différentiable sur E et vérifie pour tout $(t, z) \in E \times E$:

$$g'(t).z = 2v^{-1}(t, z) \quad \text{et} \quad g''(t) = g'$$

et donc, pour tout $(x, y) \in E \times E$:

$$g'(0_E) = 0 \quad \text{et} \quad g''(0_E).(x, y) = 2v^{-1}(x, y).$$

Soit m_2 le moment d'ordre 2 de P , et φ_p est de classe \mathcal{C}^2 sur E et vérifie :

$$\varphi_p''(0_E) = -m_2^{-1}$$

Pour tout $t \in E$, $\varphi_p''(t)$ vérifie, pour tout $(x, y) \in E \times E$:

$$\varphi_p''(t).(x, y) = \Psi_p'(g(t)).(g''(t).(x, y)) + \Psi_p''(g(t))(g'(t).x, g'(t).y)$$

Par conséquent, pour $t = 0_E$:

$$\begin{aligned} \varphi_p''(0_E).(x, y) &= \Psi_p'(g(0_E)).(g''(0_E).(x, y)) + \Psi_p''(g(0_E))(g'(0_E).x, g'(0_E).y) \\ &= 2 \Psi_p'(0) \quad v^{-1}(x, y). \end{aligned}$$

Autrement dit, pour tout $(x, y) \in E \times E$:

$$m_2^{-1}(x, y) = -2\Psi_p'(0) \quad v^{-1}(x, y).$$

En conséquence, le moment d'ordre 2 est proportionnel au paramètre de dispersion, c'est donc aussi un paramètre de dispersion de P .

Proposition 3.5. (*Image d'une loi radiale par une application linéaire*)

Soit P une loi radiale sur E de paramètre de dispersion v .

Soit f une application linéaire surjective de E dans F .

Alors la loi image P_f de P par f est radiale sur F et admet pour paramètre de dispersion $v({}^t f(\cdot), {}^t f(\cdot))$.

Démonstration :

1. Soient H un sous-espace vectoriel de E et π la projection v^{-1} – orthogonale de E sur H .

Montrons que P_π est radiale de paramètre de dispersion $v_{(H)} = v({}^t\pi, {}^t\pi)$ autrement dit, que P_π est invariante par toute transformation $(v^{-1})_H$ – orthogonale. Soit γ une transformation $(v^{-1})_H$ – orthogonale sur H . La transformation δ sur E définie par :

$$\forall x \in E \quad \delta(x) = \gamma(\pi(x)) + x - \pi(x)$$

est v^{-1} – orthogonale et admet pour réciproque δ^{-1} donnée par :

$$\forall y \in E \quad \delta^{-1}(y) = \gamma^{-1}(\pi(y)) + y - \pi(y).$$

Pour tout borélien B de $\mathbb{B}(F)$, comme $\gamma \circ \pi = \pi \circ \gamma$ nous avons :

$$\begin{aligned} P_\pi(\gamma^{-1}(B)) &= P((\gamma \circ \pi)^{-1}(B)) \\ &= P((\pi \circ \delta)^{-1}(B)) \\ &= P(\delta^{-1} \circ \pi^{-1}(B)) \\ &= P(\pi^{-1}(B)) \\ &= P_\pi(B) \end{aligned}$$

ce qui constitue le résultat cherché.

2. Désignons par H le sous-espace vectoriel v^{-1} – orthogonal de $\text{Ker } f$ et par π la projection v^{-1} – orthogonale sur H . La fonction f factorise alors à travers $\pi : f = g \circ \pi$ où g est un isomorphisme de H sur F . Il est alors clair que l'image par g d'une loi radiale sur H de paramètre de dispersion $v_{(H)}$ est radiale sur F de paramètre de dispersion $v_{(H)}({}^tg, {}^tg)$.

Puisque $P_f = (P_\pi)_g$, P_f est radiale de paramètre de dispersion :

$$v_{(H)}({}^t g, {}^t g) = v({}^t \pi \circ {}^t g, {}^t \pi \circ {}^t g) = v({}^t f, {}^t f).$$

3.5 Propriétés caractéristiques des lois radiales

Nous développons ici deux types de caractérisation des lois radiales.

D'une part (cf. prop.3.6 et corol.3.1), toute loi radiale est présentée classiquement comme un mélange de lois uniformes sur des sphères. Dans ce cas le rayon et le vecteur normalisé sont indépendants.

D'autre part (cf. prop.3.7), à la suite de Eaton ([3]), nous caractérisons une loi radiale par la loi conditionnelle de toute forme linéaire relativement à toute autre forme linéaire orthogonale.

Proposition 3.6.

Soient P une loi de probabilité sur E et v un produit scalaire sur E^ .*

Alors les deux conditions suivantes sont équivalentes :

- i) P est radiale de paramètre de dispersion v .*
- ii) P est un mélange de lois uniformes sur les sphères de E de centre 0_E .*

Dans ce cas une version régulière de la loi conditionnelle de P sachant $\|\cdot\| = r$ est $U_{v,r}$.

Démonstration

i) \Rightarrow ii) Il existe une application Ψ_p de \mathbb{R}_+ dans \mathbb{C} telle que :

$$\forall t \in E \quad \varphi_p(t) = \Psi_p(\|t\|^2).$$

Soit t fixé dans E . Pour tout u de E tel que $\|u\| = 1$, nous avons :

$$\varphi_p(t) = \Psi_p(\|t\|^2) = \Psi_p(\|t\|^2\|u\|^2) = \varphi_p(\|t\|u)$$

Alors :

$$\begin{aligned} \varphi_p(t) &= \int_{S_v} \varphi_p(\|t\|u) U_v(du) \\ &= \int_{S_v} [\int_E \exp(i \langle y, \|t\|u \rangle) P(dy)] U_v(du) \\ &= \int_E [\int_{S_v} \exp(i \langle y, u \rangle) U_v(du)] P(dy) \\ &= \int_E \varphi_{U_v}(\|t\|y) P(dy) \\ &= \int_E \Psi_{U_v}(\|t\|^2\|y\|^2) P(dy) \\ &= \int_{\mathbb{R}_+} \Psi_{U_v}(\|t\|^2 r^2) P_{\|\cdot\|}(dr) \\ &= \int_{\mathbb{R}_+} \varphi_{U_v}(rt) P_{\|\cdot\|}(dr) \\ &= \int_{\mathbb{R}_+} \varphi_{U_{v,r}}(t) P_{\|\cdot\|}(dr) \end{aligned}$$

D'où il vient, pour tout borélien $B \in \mathbb{B}(E)$:

$$P(B) = \int_{\mathbb{R}_+} U_{v,r}(B) P_{\|\cdot\|}(dr)$$

ce qui termine cette partie de la démonstration.

ii) \Rightarrow i) évident d'après la propriété 3.1

Remarque 3.2.

De la proposition précédente on déduit qu'une loi de probabilité sur \mathbb{R}_+ caractérise une loi radiale sur E comme étant la loi de son rayon.

corollaire 3.1.

Soient P une loi de probabilité sur E sans atome en 0_E et v un produit scalaire sur E^* . Soit N l'application de $E - \{0_E\}$ dans E définie par :

$$\forall x \in E - \{0_E\} \quad N(X) = \frac{1}{\|X\|} \cdot X$$

Alors les deux conditions sont équivalentes :

1. P est radiale de paramètre de dispersion v .
2. La loi P_N est la loi uniforme sur la sphère S_v et les variables aléatoires N et $\|\cdot\|$ sont indépendantes.

Démonstration :

1) \Rightarrow 2) Supposons que P soit v – radiale. Calculons la loi P_N de N . Soit g une transformation v^{-1} – orthogonale de E et soit f une fonction numérique positive $\mathbb{B}(E)$ – mesurable. Alors :

$$\begin{aligned} \mathbb{E}_{P_N}(f \circ g) &= \mathbb{E}_P(f \circ g \circ N) \\ &= \mathbb{E}_P(f \circ N \circ g) \\ &= \mathbb{E}_P(f \circ N) \\ &= \mathbb{E}_{P_N}(f) \end{aligned}$$

Donc P_N est v – radiale. Or P_N est portée par S_v par conséquent

$$P_N = U_v.$$

Calculons la loi conditionnelle de N sachant $\|\cdot\|$ notée $P_N^{\|\cdot\|}$.

Pour $P_{\|\cdot\|}$ -presque tout $r \in \mathbb{R}_+$, d'après la proposition précédente 3.6.

$$P_N^{\|\cdot\|=r} = (U_{v,r})_N = U_v = P_N$$

Par conséquent N et $\|\cdot\|$ sont indépendantes.

2) \Rightarrow 1) Pour tout $x \in E - \{0_E\}$, on a $x = \|x\|.N(x)$. Des hypothèses de 2) on déduit facilement que la loi conditionnelle de P sachant $\|\cdot\| = r$ est $U_{v,r}$ et donc que P est v -radiale.

Remarque 3.3.

Du corollaire précédent on déduit qu'une loi v -radiale sur E est caractérisée par le couple $(N, \|\cdot\|)$ dès que ces deux variables sont indépendantes et que la loi de N est U_v .

Proposition 3.7.

Soient P une probabilité sur E et v un produit scalaire sur E^* . Alors les deux conditions suivantes sont équivalentes :

- i* - P est radiale de paramètre de dispersion v .
- ii* - Pour tout couple (f, g) de formes linéaires non nulles v -orthogonales sur E , la loi conditionnelle de g sachant f est symétrique sur \mathbb{R} .

Démonstration :

i) \Rightarrow *ii*) Soient f et g deux formes linéaires non nulles v -orthogonales sur E . Notons $H_f = (Ker f)^\perp$ et $H_g = (Ker g)^\perp$.

L'application linéaire (f, g) de E sur \mathbb{R}^2 est surjective. En effet, pour tout $(\alpha, \beta) \in \mathbb{R}^2$, il existe $x_\alpha \in H_f$ et $x_\beta \in H_g$ tels que $f(x_\alpha) = \alpha$ et $g(x_\beta) = \beta$. Comme $v(f, g) = 0$, on a $H_g \subset Ker f$ et $H_f \subset Ker g$ et donc :

$$(f, g)(x_\alpha + x_\beta) = (\alpha, \beta).$$

D'après la proposition 3.5, $P_{(f,g)}$ est radiale sur \mathbb{R}^2 . Comme P_f et P_g sont elles-mêmes radiales sur \mathbb{R} (i.e. symétriques), on en déduit facilement que la

loi conditionnelle de g relativement à f est symétrique sur \mathbb{R} puisque $|g|$ et $\text{sgn}(g)$ sont indépendantes (cf. corollaire 3.1).

ii) \Rightarrow i) La démonstration s'appuie sur le lemme suivant :

Lemme 3.1.

Soient f et g deux formes linéaires sur E . Si la loi conditionnelle de g sachant f est symétrique sur \mathbb{R} , alors pour tout couple de nombres réels (a, b) on a :

$$\varphi_{p_{af+bg}} = \varphi_{p_{af-bg}}$$

Démonstration du lemme :

$$\begin{aligned} \forall t \in \mathbb{R} \quad \varphi_{p_{af+bg}}(t) &= \mathbb{E}_p(e^{it(af+bg)}) \\ &= \mathbb{E}_p(\mathbb{E}_p(e^{it(af+bg)} \mid f)) \\ &= \mathbb{E}_p(e^{ita f} \mathbb{E}_p(e^{itbg} \mid f)) \\ &= \mathbb{E}_p(e^{ita f} \mathbb{E}_p(e^{-itbg} \mid f)) \\ &= \varphi_{p_{af-bg}}(t). \end{aligned}$$

Démontrons maintenant que ii) \Rightarrow i)

Soit γ une transformation v^{-1} -orthogonale sur E . Pour tout $t \in E$ soient :

$$f = \langle \frac{1}{2}(t + \gamma(t)), \cdot \rangle \quad \text{et} \quad g = \langle \frac{1}{2}(t - \gamma(t)), \cdot \rangle.$$

On remarque que $v(f, g) = 0$, $f + g = \langle t, \cdot \rangle$ et $f - g = \langle \gamma(t), \cdot \rangle$.

En vertu de l'hypothèse ii) et du lemme précédent, on a $\varphi_{p_{f+g}} = \varphi_{p_{f-g}}$ d'où $\varphi_{p_{f+g}}(1) = \varphi_{p_{f-g}}(1)$ et donc $\varphi_p(t) = \varphi_p(\gamma(t))$. Par conséquent P est radiale de paramètre de dispersion v .

Proposition 3.8. (Cas de lois à densité)

Soient v un produit scalaire sur E^* , P une probabilité sur E absolument continue par rapport à la mesure de Lebesgue λ_v sur E . Alors les deux conditions suivantes sont équivalentes :

- i) P est radiale et admet v pour paramètre de dispersion
- ii) P admet une densité f_p de la forme :

$$\forall y \in E \quad f_p(y) = \xi_p(\|y\|^2)$$

où ξ_p est une application mesurable de \mathbb{R}_+ dans \mathbb{R}_+ .

Démonstration :

Soit f_p une densité de P relativement à la mesure de Lebesgue λ_v .

- i) \Rightarrow ii) soit g une transformation v^{-1} – orthogonale. Pour tout $A \in \mathbb{B}(E)$, on a :

$$\begin{aligned} \int_A f_p \, d\lambda_v &= P(A) \\ &= P(g^{-1}(A)) \\ &= \int_{g^{-1}(A)} f_p \, d\lambda_v \\ &= \int_A f_p \circ g^{-1} \, d(\lambda_v)_g \\ &= \int_A f_p \circ g^{-1} \, d\lambda_v \end{aligned}$$

Il en résulte que $f_p \circ g^{-1} = f_p \lambda_v$ -presque sûrement. D'où le résultat.

- ii) \Rightarrow i) Soit g une transformation v^{-1} – orthogonale. Pour tout $A \in \mathbb{B}(E)$, on a :

$$\begin{aligned}
P(g^{-1}(A)) &= \int_{g^{-1}(A)} f_p(x) \, d\lambda_v(x) \\
&= \int_{g^{-1}(A)} \xi_p(\|x\|^2) \, d\lambda_v(x) \\
&= \int_{g^{-1}(A)} \xi_p(\|g(x)\|^2) \, d\lambda_v(x) \\
&= \int_{g^{-1}(A)} (f_p \circ g)(x) \, d\lambda_v(x) \\
&= \int_A f_p(y) \, d(\lambda_v)_g(y) \\
&= \int_A f_p(y) \, d\lambda_v(y) \\
&= P(A)
\end{aligned}$$

d'où le résultat.

La proposition 3.9 suivante et le théorème fondamental 3.1 mettent en évidence une propriété particulièrement précieuse des lois radiales.

Proposition 3.9.

Soit v un produit scalaire sur E^ . Soient H un hyperplan de E et π la projection v^{-1} – orthogonale sur H .*

Pour tout $r > 0$, la loi image par π de la loi uniforme $U_{v,r}$ sur la sphère $S_{v,r}$ est absolument continue par rapport à la mesure de Lebesgue $\lambda_{v(H)}$ sur H .

Démonstration :

On considère une probabilité P sur E , v – radiale et absolument continue par rapport à la mesure de Lebesgue λ_v sur E . Une telle loi existe toujours (il suffit de considérer la loi normale $N_E(0_E, v)$).

Alors d'après la proposition 3.8, P admet une densité f_p de la forme $f_p(x) = \xi_p(\|x\|^2)$ où ξ_p est une application mesurable de \mathbb{R}_+ dans \mathbb{R}_+ .

Pour toute application numérique mesurable positive φ de H dans \mathbb{R}_+ on a :

$$\begin{aligned}
\int_H \varphi(y) d(U_{v,r})_\pi(y) &= \int_E (\varphi \circ \pi)(x) U_{v,r}(dx) \\
&= \mathbb{E}_p(\varphi \circ \pi \mid \|\cdot\| = r)
\end{aligned} \tag{3.1}$$

en vertu de proposition 3.6.

Mais pour toute fonction mesurable Ψ de \mathbb{R}_+ dans \mathbb{R}_+ on a :

$$\mathbb{E}_{p_{\|\cdot\|}}[\Psi \cdot \mathbb{E}_p(\varphi \circ \pi \mid \|\cdot\|)]$$

$$\begin{aligned}
&= \int_E \Psi(\|x\|) (\varphi \circ \pi)(x) dP(x) \\
&= \int_E \Psi(\|x\|) (\varphi \circ \pi)(x) \xi_p(\|x\|^2) d\lambda_v(x) \\
&= \int_{H \times H^\perp} \Psi(\|y+z\|) \varphi(y) \xi_p(\|y+z\|^2) d\lambda_{v(H)}(y) d\lambda_{v(H^\perp)}(z) \\
&= 2 \int_K \Psi(r) \varphi(y) \xi_p(r^2) g(y, r) d\lambda_{v(H)}(y) d\lambda_{\mathbb{R}_+}(r)
\end{aligned}$$

en effectuant le changement de variable $(y, z) \rightsquigarrow (y, \|y+z\|)$

où $K = \{(y, r) \in H \times \mathbb{R}_+^* / \|y\| < r\}$; compte tenu du fait que $\dim H^\perp = 1$ le jacobien $g(y, r)$ de cette transformation vaut $r(r^2 - \|y\|^2)^{\frac{-1}{2}}$.

On remarque que $\lambda_{\mathbb{R}_+}$ est équivalente à $(\lambda_v)_{\|\cdot\|}$ et qu'une densité de $\lambda_{\mathbb{R}_+}$ relativement à $(\lambda_v)_{\|\cdot\|}$ est l'application :

$$K : r \rightarrow \frac{\Gamma(\frac{n}{2} + 1)}{n\pi^{\frac{n}{2}}} r^{1-n}$$

où n est la dimension de E .

Il vient alors :

$$\begin{aligned}
&\mathbb{E}_{p_{\|\cdot\|}}[\Psi \cdot \mathbb{E}_p(\varphi \circ \pi \mid \|\cdot\|)] \\
&= 2 \int_K \Psi(r) \varphi(y) \xi_p(r^2) g(y, r) k(r) d\lambda_{v(H)}(y) d(\lambda_v)_{\|\cdot\|}(r) \\
&= 2 \int_k \Psi(r) \varphi(y) g(y, r) k(r) d\lambda_{v(H)}(y) dP_{\|\cdot\|}(r) \\
&= \int_{\mathbb{R}_+^*} \Psi(r) (2 \int_{\{\|y\| < r\}} \varphi(y) g(y, r) k(r) d\lambda_{v(H)}(y)) dP_{\|\cdot\|}(r).
\end{aligned}$$

On en déduit que, pour tout $r > 0$:

$$\mathbb{E}_P(\varphi \circ \pi \mid \|\cdot\| = r) = 2 \int_H \varphi(y) g(y, r) k(r) \mathbf{1}_{\{\|y\| < r\}}(y) d\lambda_{v(H)}(y).$$

Ceci étant vrai pour toute fonction φ mesurable de H dans \mathbb{R}_+ , en reportant l'égalité précédente dans (3.1), on en déduit que pour tout $r > 0$, $(U_{v,r})_\pi$ est absolument continue par rapport à la mesure de Lebesgue $\lambda_{v(H)}$ sur H est portée par $B_{v(H),r}$ et admet pour densité relativement à $\lambda_{v(H)}$ sur cet ensemble :

$$2k(r)g(\cdot, r) = \frac{\Gamma(\frac{n}{2})}{\pi^{\frac{n}{2}}} \cdot \frac{r^{2-n}}{(r^2 - \|\cdot\|^2)^{\frac{1}{2}}}.$$

Théorème 3.1.

Soient v un produit scalaire sur E^ et P une loi radiale sur E de paramètre de dispersion v , n'admettant pas d'atome en 0_E .*

Alors la projection v^{-1} – orthogonale de P sur tout sous-espace vectoriel propre H de E est absolument continue relativement à la mesure de Lebesgue $\lambda_{v(H)}$ sur H .

Démonstration

1. Supposons d'abord que $\dim H = (\dim E) - 1$. Si n désigne la projection v^{-1} – orthogonale sur H , pour tout $r > 0$, $\lambda_{v(H)}$ est absolument continue par rapport à et admet une densité h_r , d'après la proposition précédente.

Comme P est v – radiale, on a, d'après la proposition 3.6, pour toute fonction φ mesurable de H dans R_+ .

$$\begin{aligned}
E_{p_\pi}(\varphi) &= E_p(\varphi \circ \pi) \\
&= \int_{R_+^*} \left(\int_E (\varphi \circ \pi)(x) dU_{v,r}(x) \right) dP_{\|\cdot\|}(r) \\
&= \int_{R_+^*} \left(\int_H \varphi(y) d(U_{v,r})_\pi(y) \right) dP_{\|\cdot\|}(r) \\
&= \int_H \varphi(y) \left(\int_{R_+^*} h_r(y) dP_{\|\cdot\|}(r) \right) d\lambda_{(H)}(y).
\end{aligned}$$

Donc P_π est absolument continue par rapport à $\lambda_{v(H)}$ et admet pour densité le mélange des densités des projections des lois uniformes

$$y \rightarrow \int_{R_+^*} h_r(y) dP_{\|\cdot\|}(r).$$

2. Dans le cas où H est un sous-espace vectoriel propre de E , la propriété se déduit facilement de 1 en se rappelant que la projection d'une loi radiale est encore radiale.

3.6 Cas de la normalité

Proposition 3.10.

Soit v un produit scalaire sur E^ . Soit P une loi radiale sur E , de paramètre de dispersion v .*

Si la projection v^{-1} – orthogonale π de P sur un sous-espace vectoriel H de E ($\dim H \neq 0$) est une loi normale, alors P est une loi normale $N_E(0_E, \sigma^2 v)$.

Démonstration :

Soit f une forme linéaire sur H de norme 1. Alors $f \circ \pi$ appartient à E^* et est de norme égale aussi à 1.

D'après la proposition 3.1, pour tout $t \in E$, $\varphi_p(t) = \Psi(\|t\|^2)$ où Ψ est la fonction caractéristique de $P_{f \circ \pi} = (P_\pi)_f$.

Or par hypothèse P_π est normale, donc $P_{f \circ \pi} = (P_\pi)_f$ est normale et φ_p à la forme fonctionnelle de la fonction caractéristique de $N_E(0_E, \sigma^2 v)$.

Proposition 3.11.

Soient v un produit scalaire sur E^* et P une loi v – radiale sur E . Soit (e_1, \dots, e_n) une base v^{-1} – orthonormale de E .

Si les marges P_i de P sur les sous-espaces vectoriels E_i engendrés par $(e_i/1 \leq i \leq n)$ sont indépendantes, alors la loi P est égale à $N_E(0_E, \sigma^2 v)$.

Démonstration :

Pour $t \in E$, soit $(t_i)_{1 \leq i \leq n}$ la suite des composantes de t dans la base (e_1, \dots, e_n) . Alors, d’une part :

$$\varphi_p(t) = \prod_{i=1}^n \varphi_{p_i(t_i)} = \prod_{i=1}^n \Psi_p(t_i^2)$$

en vertu de l’indépendance et, d’autre part :

$$\varphi_p(t) = \Psi_p\left(\sum_{i=1}^n t_i^2\right)$$

Ainsi la fonction continue Ψ_p vérifie l’égalité $\Psi_p(a+b) = \Psi_p(a)\Psi_p(b)$. Par conséquent Ψ_p est une fonction exponentielle $\Psi_p(s) = e^{\alpha s}$. De plus, comme Ψ_p est bornée sur \mathbb{R}_+ , nous avons $\alpha < 0$ ce qui donne le résultat cherché.

Chapitre 4

Application à la théorie de l'estimation

4.1 Introduction

Pour la plupart des lois de probabilité usuelles, l'estimateur du maximum de vraisemblance est défini de façon unique, et se calcule explicitement.

Sur le plan théorique, il présente de nombreux avantages. Sous des hypothèses vérifiées par de nombreux modèles courants, on démontre qu'il est asymptotiquement sans biais et convergent. On démontre de plus que sa variance est minimale. La méthode du maximum de vraisemblance est donc théoriquement la meilleure des méthodes d'estimation.

Mais cette méthode d'estimation par maximum de vraisemblance a aussi ses défauts, on cite parmi ses défauts :

1. La complexité de maximisation de la fonction de vraisemblance.
2. Une technique de maximisation donne forcément des estimateurs peu

lisses.

3. L'approche du maximum de vraisemblance n'admet pas toujours de justification probabiliste et décisionnelle.

4.2 Modèles

Deux modèles seront considérés par la suite :

Modèle 1

C'est le modèle d'une suite gaussienne :

$$y_j = \theta_j + \epsilon \xi_j, \quad j = 1, \dots, d$$

avec $0 < \epsilon < 1$, les ξ_j étant des v.a.i.i.d. de loi $N(0, 1)$, Et si on note y, θ, ξ les vecteurs.

$$y = (y_1, \dots, y_d), \quad \theta = (\theta_1, \dots, \theta_d), \quad \xi = (\xi_1, \dots, \xi_d) \sim N_d(0, 1)$$

où $N_d(0, 1)$ désigne la loi normale standard en dimension d , le modèle peut s'écrire :

$$y = \theta + \epsilon \xi, \quad \xi \sim N_d(0, 1)$$

Le problème statistique consiste à estimer le paramètre inconnu $\theta \in \mathbb{R}^d$

Modèle 2

On observe les vecteurs aléatoires X_1, \dots, X_n vérifiant :

$$X_i = \theta + \eta_i, \quad i = 1, \dots, n$$

avec $\theta \in \mathbb{R}^d$, les η_i étant des vecteurs gaussiens i.i.d. de loi $N_d(0, 1)$.

Et le problème statistique consiste à estimer le paramètre $\theta \in \mathbb{R}^d$.

Le vecteur $\bar{X} = \frac{1}{n} \sum_i^n X_i$ est une statistique exhaustive dans ce modèle.

On peut écrire :

$$\bar{X} = \theta + \epsilon \xi$$

avec :

$$\epsilon = \frac{1}{\sqrt{n}} \quad \text{et} \quad \xi = \frac{1}{\sqrt{n}} \sum_i^n \eta_i \sim N_d(0, 1)$$

Dans la suite \mathbb{E}_θ désigne l'espérance relative à la loi de y dans le Modèle 1 ou à celle de \bar{X} dans le modèle 2 et $\|\cdot\|$ désignera la norme euclidienne de \mathbb{R}^d .

Remarque 4.1.

Le modèle 1 avec $\epsilon = \frac{1}{\sqrt{n}}$ est équivalent au modèle 2 au sens suivant :

Quelle que soit la fonction mesurable $\hat{\theta} : \mathbb{R}^d \rightarrow \mathbb{R}^d$, le risque quadratique $\mathbb{E}_\theta \|\hat{\theta}(y) - \theta\|^2$ de l'estimateur $\hat{\theta}(y)$ dans le modèle 1 avec $\epsilon = \frac{1}{\sqrt{n}}$ est égal au risque $\mathbb{E}_\theta \|\hat{\theta}(\bar{X}) - \theta\|^2$ de l'estimateur $\hat{\theta}(\bar{X})$ dans le modèle 2.

Remarque 4.2.

Alors que le modèle 2 est classique en statistique paramétrique, le modèle 1 sera utile dans le contexte d'applications à l'estimation non-paramétrique.

4.3 Admissibilité

Définition 4.1.

Un estimateur θ^* du paramètre θ est dit **inadmissible** sur $\Theta \subseteq \mathbb{R}^d$ par rapport au risque quadratique s'il existe un autre estimateur $\hat{\theta}$ tel que :

$$\mathbb{E}_\theta \|\hat{\theta} - \theta\|^2 \leq \mathbb{E}_\theta \|\theta^* - \theta\|^2 \quad \text{pour tout } \theta \in \Theta$$

et il existe $\theta_0 \in \Theta$ tel que :

$$\mathbb{E}_{\theta_0} \|\hat{\theta} - \theta_0\|^2 \leq \mathbb{E}_{\theta_0} \|\theta^* - \theta_0\|^2$$

Dans le cas contraire, l'estimateur θ^* est dit **admissible**.

Exemple 4.1.

Le risque quadratique de l'estimateur \bar{X} dans le modèle 2 vaut :

$$\mathbb{E}_\theta \|\bar{X} - \theta\|^2 = \frac{d}{n} = d\epsilon^2, \quad \forall \theta \in \mathbb{R}^d$$

En considérant le modèle 2, **Stein** a montré que si $d \geq 3$ l'estimateur \bar{X} est **inadmissible**. Cette propriété se nomme **phénomène de Stein**. En outre, Stein a proposé un estimateur qui est meilleur que \bar{X} partout sur \mathbb{R}^d si $d \geq 3$. La construction est basée sur une **contraction** qui rapproche les valeurs de \bar{X} vers 0 en fonction de $\|\bar{X}\|$.

4.4 Contraction de Stein et estimateur de James-Stein

Pour expliquer l'idée de la contraction de stein pour le modèle 1 (le raisonnement pour le modèle 2 étant analogue). On a besoin des lemmes suivants :

Lemme 4.1. (*Lemme de Stein, 1981*)

Supposons qu'une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ vérifie :

(i) $f(u_1, \dots, u_d)$ est absolument continue en chaque coordonnée u_i pour presque toutes les valeurs des autres coordonnées ($u_j, j \neq i$).

(ii)

$$\mathbb{E}_\theta\left[\frac{\partial f}{\partial y_i}(y)\right] < \infty, \quad i = 1, \dots, d$$

Alors :

$$\mathbb{E}_\theta[(\theta_i - y_i)f(y)] = -\epsilon^2 \mathbb{E}_\theta\left[\frac{\partial f}{\partial y_i}(y)\right], \quad i = 1, \dots, d$$

preuve :

La preuve utilise essentiellement l'intégration par parties avec une légère modification due au fait que la fonction f n'est pas différentiable au sens usuel.

Lemme 4.2. Pour $d \geq 3$, pour tout $\theta \in \mathbb{R}^d$,

$$0 < \mathbb{E}_\theta\left(\frac{1}{\|y\|^2}\right) < \infty$$

La preuve repose sur la propriété de symétrie sphérique de la loi gaussienne, dans le sens :

Pour $\xi \sim N_d(\mu, \sigma^2)$, on a :

$$\forall v, v' \in \mathbb{R}^d : \|v\| = \|v'\| \Rightarrow \|\xi + v\| = \|\xi + v'\|$$

et pour $d \geq 3$, il existe une constante $C > 0$ telle que :

$$\int_{\mathbb{R}^d} \exp\left(-\frac{\|u\|^2}{6}\right) \|u\|^{-2} du = C \int_0^\infty e^{-\frac{r^2}{6}} r^{d-3} dr < \infty$$

4.5 Contraction de Stein

Stein a introduit la classe des estimateurs de la forme $\hat{\theta} = g(y)y$, où

$g : \mathbb{R}^d \rightarrow \mathbb{R}$ est une fonction à choisir.

Les coordonnées du vecteur $\hat{\theta}$ sont de la forme :

$$\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_d), \text{ avec } \hat{\theta}_j = g(y)y_j$$

Le vecteur aléatoire y est lui même un estimateur de θ (analogue à \bar{X} dans le modèle 2). Son risque vaut :

$$\mathbb{E}_\theta \|y - \theta\|^2 = d\epsilon^2$$

Cherchons une fonction g telle que le risque $\hat{\theta} = g(y)y$ soit inférieur à celui de y .

$$\mathbb{E}_\theta \|\hat{\theta} - \theta\|^2 = \sum_{i=1}^d \mathbb{E}_\theta [g(y)y_i - \theta_i]^2 =$$

$$\sum_{i=1}^d \{ \mathbb{E}_\theta [(y_i - \theta_i)^2] + 2\mathbb{E}[(\theta_i - y_i)(1 - g(y))y_i] + \mathbb{E}_\theta [y_i^2(1 - g(y))^2] \}$$

Supposons que la fonction g est telle que les conditions du lemme 4.1 soient vérifiées pour $f_i(y) = (1 - g(y))y_i, i = 1, \dots, d$. Alors :

$$\mathbb{E}_\theta [(\theta_i - y^i)(1 - g(y))y_i] = -\epsilon^2 \mathbb{E}_\theta \left[1 - g(y) - y_i \frac{\partial g}{\partial y_i}(y) \right]$$

et

$$\mathbb{E}_\theta[(\hat{\theta}_i^2 - \theta_i^2) = \epsilon^2 - 2\epsilon^2 \mathbb{E}_\theta(1 - g(y) - y_i \frac{\partial g}{\partial y_i})] + \mathbb{E}_\theta[y_i^2(1 - g(y))^2]$$

En prenant la somme sur i , on obtient :

$$\mathbb{E}_\theta \|\hat{\theta} - \theta\|^2 = d\epsilon^2 + \mathbb{E}_\theta[W(y)]$$

avec :

$$W(y) = -2d\epsilon^2(1 - g(y)) + 2\epsilon^2 \sum_{i=1}^d y_i \frac{\partial g}{\partial y_i}(y) + \|y\|^2(1 - g(y))^2$$

Pour que le risque de $\hat{\theta}$, soit plus petit que celui de y , il suffit de choisir g telle que :

$$\mathbb{E}_\theta[W(y)] < 0$$

Afin de satisfaire à cette inégalité, Stein (1956) a proposé de chercher g parmi les fonctions de la forme :

$$g(y) = 1 - \frac{c}{\|y\|^2}$$

avec une constante $c > 0$ convenablement choisie. Dans ce cas les fonction $f_i(y) = (1 - g(y))y_i$ vérifient les conditions du lemme 4.1, on obtient donc :

$$W(y) = \frac{1}{\|y\|^2}(-2d\epsilon^2 c + 4\epsilon^2 c + c^2)$$

La valeur c fournissant le minimum de $W(y)$ vaut :

$$c_{opt} = \epsilon^2(d - 2)$$

Dans ce cas $W(y) = -\frac{\epsilon^4(d-2)^2}{\|y\|^2}$, et $-\infty < \mathbb{E}_\theta[W(y)] < 0$ dès que $d \geq 3$.

d'après le lemme 4.2. Et La fonction g et l'estimation $\hat{\theta} = g(y)y$ associés à ce choix de c valent :

$$g(y) = 1 - \frac{\epsilon^2(d-2)}{\|y\|^2} \quad \text{et} \quad \hat{\theta}_{JS} = \left(1 - \frac{\epsilon^2(d-2)}{\|y\|^2}\right)y \quad (4.1)$$

on appelle $\hat{\theta}_{JS}$ estimateur de James-Stein (James et Stein 1961). Si la norme $\|y\|$ est assez grande, la fonction g effectue une contraction de y vers 0, que l'on appelle contraction de Stein (en anglais Stein's shrinkage).

4.6 Phénomène de Stein

De façon similaire, pour le modèle 2 l'estimateur de James-Stein s'obtient si l'on remplace y par \bar{X} et ϵ par $\frac{1}{\sqrt{n}}$ dans (4.1) :

$$\hat{\theta}_{JS} = \left(1 - \frac{(d-2)}{n\|\bar{X}\|^2}\right)\bar{X}$$

On a donc le résultat suivant.

Théorème 4.1. *(Phénomène de Stein)*

Soit $d \geq 3$. L'estimateur $\hat{\theta} = y$ est inadmissible sur \mathbb{R}^d dans le modèle 1 et l'estimateur $\hat{\theta} = \bar{X}$ est inadmissible sur \mathbb{R}^d dans le modèle 2.

Conclusion et perspectives

Dans ce mémoire on explique comment la connaissance du groupe de symétrie d'une loi de probabilité sur l'étude d'un phénomène aléatoire, et en particulier sur l'estimation du paramètre d'intérêt. L'approche géométrique se base sur la symétrie gouvernée par l'action d'un groupe (groupe de Lie) sur une variété riemannienne, ce qui produit l'invariance de certaines quantités géométrique et par suite la réduction de l'étude du phénomène aléatoire à des parties réduites de l'espace des événements.

La difficulté de cette approche provient du fait qu'on ne connaît pas tout de ces groupes de symétrie, la loi normale est une exception, mais les travaux de recherches actuels et en avenir peuvent mener à des résultats qui vont changer notre vision à l'univers de l'aléatoire.

Bibliographie

- [1] Cellier, Dominique, Dominique Fourdrinier, and Christian Robert. "Robust shrinkage estimators of the location parameter for elliptically symmetric distributions." *Journal of Multivariate Analysis* 29.1 (1989) : 39-52.
- [2] Cellier, Dominique, and Dominique Fourdrinier. "Sur les lois à symétrie elliptique." *Séminaire de probabilités de Strasbourg* 24 (1990) : 300-328.
- [3] Eaton, Morris L. "A characterization of spherical distributions." *Journal of Multivariate Analysis* 20.2 (1986) : 272-276.
- [4] Neveu, Jacques. *Bases mathématiques du calcul des probabilités*. Masson, 1970.
- [5] Philoche, J. L. "Une condition de validité pour le test F." *Statistique et analyse des données* 2.1 (1977) : 37-59.
- [6] Stein, Charles. "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution." *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*. Vol. 1. No. 1. 1956.
- [7] Tassi, Philippe, and Sylvia Legait. *Théorie des probabilités en vue des applications statistiques*. Editions Technip, 1990.

[8] Tsybakov, Alexandre B. Introduction à l'estimation non paramétrique.
Vol. 41. Springer Science Business Media, 2003.