

CTU, Master Enseignement des
Mathématiques
Statistique Inférentielle

Jean-Yves DAUXOIS

Université de Franche-Comté
Année scolaire 2011-2012

Ce polycopié contient le cours, les sujets d'exercice et leurs corrigés ainsi que les sujets des devoirs proposés.

Les énoncés des exercices sont donnés en fin de chapitre auxquelles ils font référence. Il est vivement conseillé d'essayer de faire sérieusement les exercices, sans aller trop rapidement voir leurs corrections détaillées en fin de polycopié. On sait en effet que, pour qu'une correction soit efficace, il faut qu'elle vienne après une période de recherche personnelle de la solution.

Les devoirs, quant à eux, ne sont pas des exercices supplémentaires (ces derniers accompagnés de leurs corrections sont déjà assez nombreux !). Pour qu'ils apportent réellement autre chose que les exercices, ils doivent être faits dans les conditions d'un devoir surveillé ou d'un examen. En conséquence, il vous est vivement conseillé de faire les devoirs et de m'envoyer votre copie (éventuellement les unes après les autres). En retour vous recevrez votre copie corrigée et également une correction type du devoir. Le premier des devoirs peut être résolu dès que l'on est parvenu à la fin de la seconde section du Chapitre 5. Le second est lui réalisable après avoir travaillé l'ensemble du Chapitre 5. Les trois autres, même s'ils peuvent être "attaqués" plus tôt, ne seront réalisables qu'une fois assimilé l'ensemble des notions. Ils peuvent fournir de bons exercices de révision en perspective de l'examen.

Enfin, ce polycopié contient certainement de nombreuses coquilles et mérite encore d'être amélioré. Merci d'avance aux lecteurs attentifs de transmettre leur remarques, suggestions ou indications sur la localisation des coquilles. Un petit mail à l'adresse jean-yves.dauxois@univ-fcomte.fr et l'amélioration est prise en compte...

Bon courage !

Table des matières

Partie 1. Introduction et Modèle Statistique	5
Chapitre 1. Introduction	7
Chapitre 2. Modèle Statistique	11
1. Définition	11
2. Modèle d'échantillonnage	15
3. Vraisemblance	15
4. Familles Exponentielles	16
5. Modèle position-échelle	17
6. Exercices	18
Partie 2. Estimation ponctuelle	21
Chapitre 3. Statistique et Estimateur	23
Chapitre 4. Construction d'estimateurs	27
1. Estimateurs empiriques (des moments)	27
2. Méthode de substitution	29
3. Méthode des moments	29
4. Maximum de vraisemblance	30
5. Exercices	33
Chapitre 5. Qualité d'un estimateur	37
1. Estimateur convergent	37
2. Estimateur sans biais	39
3. Risque d'un estimateur	40
4. Information de Fisher	43
5. Borne de Cramer-Rao (ou Fréchet-Darmonis-Cramer-Rao)	46
6. Exercices	48
Chapitre 6. Amélioration d'estimateurs	51
1. Statistique exhaustive	51
2. Statistique exhaustive minimale	54
3. Théorème de Rao-Blackwell	54
4. Théorème de Lehmann-Scheffé	56
5. Cas des familles exponentielles	57
6. Exercices	57

Chapitre 7. Comportement asymptotique d'un estimateur	59
1. Normalité asymptotique	59
2. Estimateurs empiriques des moments	60
3. Estimateur du maximum de vraisemblance	60
4. La δ -méthode ou l'étude asymptotique d'un estimateur obtenu par la méthode de substitution	61
5. Estimateurs par la méthode des moments	62
6. Exercices	63
Partie 3. Intervalles de confiance	65
Chapitre 8. Intervalles de confiance exacts	67
Chapitre 9. Intervalles de confiance asymptotiques	71
Chapitre 10. Exercices sur les intervalles de confiance exacts et asymptotiques	73
Partie 4. Correction des exercices	75
Correction des exercices du Chapitre 2	77
Correction des exercices du Chapitre 4	85
Correction des exercices du Chapitre 5	99
Correction des exercices du Chapitre 6	119
Correction des exercices du Chapitre 8	129
Partie 5. Devoirs	135

Partie 1

Introduction et Modèle Statistique

CHAPITRE 1

Introduction

Considérons un problème de Fiabilité où l'on étudie la durée de vie X d'un matériel. Il est raisonnable d'admettre que celle-ci est aléatoire et X est alors une variable aléatoire (v.a.) de fonction de répartition (f.d.r.) F . Supposons que l'on soit précisément intéressé par l'évaluation de la probabilité que le matériel soit en marche après un temps t_0 de fonctionnement, c'est à dire évaluer

$$\bar{F}(t_0) = P(X > t_0) = 1 - F(t_0).$$

Pour cela on observe le fonctionnement n matériels similaires et on relève leurs temps de panne respectifs: x_1, \dots, x_n . On note $K_n = \sum_{i=1}^n \mathbb{1}_{x_i \leq t_0}$ le nombre de matériels tombées en panne au temps t_0 . Il en reste donc $n - K_n$ encore en marche à cet instant.

Il est assez naturel d'estimer la probabilité $\bar{F}(t_0)$ par :

$$\hat{\bar{F}}(t_0) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{n - K_n}{n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i > t_0\}}.$$

Posons maintenant une hypothèse supplémentaire. On suppose (on sait ou on a pu vérifier) que la loi de X est une loi exponentielle $\mathcal{E}(\lambda)$, mais dont on ignore le paramètre λ .

Calculons l'espérance de X . On a

$$\mathbb{E}(X) = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^{+\infty} u e^{-u} du = \frac{\Gamma(2)}{\lambda},$$

où

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

est la fonction Gamma. On sait que $\Gamma(n) = (n-1)!$, ce qui nous donne ici $\mathbb{E}(X) = 1/\lambda$.

Il est assez naturel d'estimer l'espérance de X par la moyenne empirique des temps observés, i.e. par

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Ainsi λ peut être estimé par :

$$\hat{\lambda} = \frac{1}{\bar{x}} = \frac{n}{\sum_{i=1}^n x_i}.$$

Un calcul simple montre que

$$\bar{F}(t_0) = \int_{t_0}^{+\infty} \lambda e^{-\lambda x} dx = \exp(-\lambda t_0)$$

et on peut donc estimer la probabilité que le matériel fonctionne durant le temps t_0 par :

$$\tilde{F}(t_0) = \exp(-\hat{\lambda} t_0).$$

Les estimations précédentes sont appelées estimations ponctuelles. On constate en particulier que plusieurs estimateurs ont été proposés pour $\bar{F}(t_0)$. Ils conduisent à des estimations différentes de la même quantité pour un seul lot de matériel testé. Mais on remarque également qu'un même estimateur peut mener à différentes estimations si on considère plusieurs lots de matériels. Les valeurs observées x_1, \dots, x_n n'ont en effet aucune raison d'être les mêmes.

Ainsi on se pose naturellement les questions suivantes. Comment peut-on comparer différents estimateurs ? Quelle(s) définition(s) donner de la qualité d'un estimateur ? Comment mesurer l'erreur commise par un estimateur (puisqu'en particulier elle varie d'une observation à l'autre) ? Toutes ces questions seront abordées dans la Partie 2 de ce cours.

Ce qui précède montre que l'estimation ponctuelle a un inconvénient majeur, celui de se tromper presque toujours. Au moins dans le cas de v.a. absolument continues, ce qui était le cas précédemment, il apparaît clairement que l'on est presque sûr de ne pas "tomber" sur la valeur théorique que l'on cherche à estimer. C'est pourquoi on préfère parfois donner un intervalle plutôt qu'une valeur. On parle d'intervalle de Confiance ou parfois de fourchette d'estimation. Bien sûr il reste une erreur possible. On donnera alors l'intervalle en fonction de l'erreur que l'on s'autorise (ou que l'on nous autorise). Plus on souhaitera que la probabilité d'erreur soit petite, plus grand sera l'intervalle. Et inversement plus la probabilité d'erreur que l'on s'autorise est grande, plus on pourra donner un intervalle étroit. L'estimation par intervalles de confiance fait l'objet de la Partie 3 de cours.

Il reste un troisième axe fondamental de la Statistique Inférentielle que nous n'aborderons pas dans ce cours. Il est de nature assez différente des deux précédents et consiste à pouvoir se donner des outils statistiques pour décider entre deux hypothèses différentes. Ainsi, si l'on considère à nouveau l'exemple précédent sur la fiabilité d'un matériel, on peut être assez rapidement amené à répondre à des questions comme les suivantes. La fiabilité du matériel $\bar{F}(t_0)$ en un instant t_0 fixé (par exemple 2000h) est-elle supérieure ou pas à 0,99 ? Appartient-elle à l'intervalle $[0.975, 0.985]$ (il ne s'agit pas ici du même problème que celui du paragraphe précédent sur la notion d'intervalle de confiance comme nous le verrons en étudiant plus en détails ces notions) ? L'hypothèse de loi exponentielle pour la durée de vie X du matériel est-elle raisonnable ou pas ? Ou encore si l'on dispose de deux versions du matériel : l'un est-il plus fiable que l'autre en un instant t_0 ? Autrement dit, en notant respectivement F_1 et F_2 les fonctions de répartitions de la durée de vie de chaque matériel, a-t-on $F_1(t_0) \leq F_2(t_0)$ ou le contraire ?

La théorie des tests d'hypothèses permet de répondre, entre autres, à toutes ces questions. Dans ce domaine les erreurs sont également possibles : celles de choisir l'une des deux hypothèses alors que c'est l'autre qui est vraie. L'objectif est alors naturellement de chercher à réduire au maximum ces deux erreurs mais nous verrons rapidement que cela n'est pas possible conjointement. Ici aussi se posera également la question de l'optimalité (dans un sens à définir) de la procédure de test choisi.

D'une manière générale.

Statisticien confronté à des données : brutes (résultat du contrôle qualité d'un produit, taille d'individus, âge de la mère à la naissance du premier enfant, concentration en ozone de l'atmosphère etc...) ou résultats d'expériences (expériences biologiques, pharmaceutiques, agronomiques etc...).

Travail du statisticien. Extraire de l'information (résumée et pertinente) de ces données (comme par exemple la taille moyenne des individus). Modéliser la part d'aléa (par exemple déterminer la loi de la durée de vie X du matériel). Tirer des conclusions sur la population totale à partir d'observations sur un échantillon).

Mais il peut aussi avoir à (donner les moyens pour) prendre des décisions (comme par exemple l'activation du plan antipollution en raison d'une trop grande concentration d'ozone). Effectuer des prévisions (prévision du temps en météorologie, prévision du cours d'une action en finance).

CHAPITRE 2

Modèle Statistique

L'objet de ce chapitre est de présenter le socle sur lequel vont s'appuyer toutes les techniques statistiques présentées dans les parties ou chapitres suivants. Ainsi nous présenterons la notion fondamentale de modèle statistique et en donnerons quelques cas particuliers importants que nous retrouverons dans les développements ultérieurs. Nous présenterons aussi une notion très liée à la notion de modèle statistique : la vraisemblance. Elle est également très importante en statistique.

1. Définition

EXEMPLE 2.1. *Un problème de Fiabilité et modèle de Bernoulli*

Revenons à notre problème introductif de Fiabilité du Chapitre précédent et à sa première partie sur l'estimation ponctuelle. On a cherché à connaître la vraie valeur de la fonction de répartition $\bar{F}(t_0)$ de la durée de vie du matériel en un instant t_0 . Il est intéressant de décrire ce problème d'une autre manière.

Utilisons une v.a. Y à valeurs $\{0, 1\}$ pour modéliser l'état du matériel au temps t_0 . On note $\{Y = 1\}$ si le matériel est en marche et $\{Y = 0\}$ s'il est en panne. On a $p_0 = P(Y = 1) = \bar{F}(t_0)$ et $P(Y = 0) = 1 - p_0$. La v.a. Y est de loi de Bernoulli de paramètre p_0 , où p_0 a une valeur inconnue dans $[0, 1]$.

On a donc fait comme si l'on avait une infinité de lois possibles pour Y : toutes les lois de Bernoulli $B(1, p)$, avec p dans $[0, 1]$. Et le problème était alors de trouver la vraie valeur p_0 , à partir des "résultats" observés pour les n machines testées, notés y_1, \dots, y_n . On a estimé p_0 par $(\sum y_i)/n$. On parle de modèle et estimation paramétriques : restait seulement à estimer un paramètre. C'est essentiellement le cadre considéré par ce cours dans sa partie estimation ponctuelle.

Notons la présence des ensembles suivants :

- E =espace des observations possibles= $\{0, 1\}$;
- \mathcal{E} =tribu des événements sur $E=\mathcal{P}(E)$, ensemble des parties de E ;
- Une famille de Probabilités constituée par toutes les lois de Bernoulli,

$$\mathcal{P} = \{B(1, p) : p \in [0, 1]\}.$$

Nous verrons qu'ils définissent un modèle paramétrique qui dans le cas présent est appelé modèle de Bernoulli.

En revanche, si l'on s'intéresse à l'estimation de $\bar{F}(t)$ pour tout t dans \mathbb{R}^+ , il faudrait estimer une infinité de paramètres : toutes les valeurs prises par la fonction \bar{F} . On parle alors d'estimation non-paramétrique. C'est un sujet que nous ne ferons qu'aborder, essentiellement quand nous traiterons le sujet des tests non-paramétriques. \diamond

Nous constatons une différence avec un modèle probabiliste (E, \mathcal{E}, P) . Dans modèle probabiliste il y a une seule probabilité et les seules questions qui se posent sont de l'ordre du calcul (que l'on sait ou ne sait pas faire). Avec un modèle statistique $(E, \mathcal{E}, \mathcal{P})$, ces mêmes questions peuvent éventuellement apparaître dans un deuxième temps, mais avant il faut gérer la présence d'un ensemble de probabilités. Autrement dit la probabilité sous jacente au phénomène est pas connue ou pas entièrement (c'est surtout ce cas là que l'on traite dans ce cours). Le Statisticien cherchera à la déterminer, l'estimer.

Ce modèle

$$(E, \mathcal{E}, \mathcal{P}) = (E, \mathcal{E}, \{B(1, p) : p \in [0, 1]\})$$

peut être utilisé pour modéliser d'autres phénomènes, situations.

Exemples.

1) Jeu de pile ou face. Le problème est de connaître la probabilité p d'obtenir pile (par exemple), ce qui revient à admettre que le dé peut être pipé. On note $Y = 1$ si on obtient pile, $Y = 0$ sinon on obtient une face. Dire que la pièce peut être pipée, revient à dire que le résultat d'un lancer Y est de loi de Bernoulli $B(1, p)$ avec p inconnu dans $[0, 1]$. On fait n lancers, résultats notés y_1, \dots, y_n et on cherchera à estimer p .

2) Sondage d'intention de vote au second tour des élections présidentielles. On suppose que seulement deux candidats A et B se présentent à une élection. On note p la proportion de votant pour le candidat A et $1 - p$ pour B . En notant $\{Y = 1\}$ l'événement l'électeur vote pour A , et $\{Y = 0\}$ s'il vote pour B , le vote peut être modélisé par une v.a. Y de loi de Bernoulli $B(1, p)$, avec encore une fois p qui peut prendre n'importe quelle valeur dans $[0, 1]$. On sonde n électeurs sur leurs intentions, résultats notés y_1, \dots, y_n et on cherche à estimer p .

DÉFINITION 2.1. On appelle **modèle statistique**, la donnée d'un espace des observations E , d'une tribu \mathcal{E} d'événements sur E et d'une famille de probabilités \mathcal{P} sur l'espace probabilisable (E, \mathcal{E}) . On le note $(E, \mathcal{E}, \mathcal{P})$ ou, quand il n'y a pas de risque de confusion, plus simplement \mathcal{P} .

On supposera que la vraie loi sous-jacente au phénomène que l'on étudie appartient au modèle statistique que l'on s'est donné. Il existe des outils pour vérifier si cette hypothèse est raisonnable ou pas. Mais nous ne les présenterons pas dans le cadre de ce cours, car ils font appels à la théorie des tests qui n'est pas au programme de cet enseignement.

On note X la v.a. qui modélise le phénomène aléatoire que l'on étudie. Autrement dit la v.a. X engendre les observations dont on dispose. Elle est à valeurs dans (E, \mathcal{E}) et sa loi de probabilité P inconnue est dans la famille \mathcal{P} . On appellera parfois X **v.a. générique** du modèle statistique.

DÉFINITION 2.2. On dit qu'un modèle statistique est **paramétrique** s'il existe un entier d et un sous ensemble Θ de \mathbb{R}^d tels que la famille de probabilités \mathcal{P} puisse être paramétrée par Θ , i.e. tels que l'application :

$$\begin{aligned} \Theta &\rightarrow \mathcal{P} \\ \theta &\mapsto P_\theta \end{aligned}$$

est surjective.

On note $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$.

Dans le cas contraire on parle de modèle **non-paramétrique**.

Le modèle de Bernoulli utilisé dans la modélisation du fonctionnement du matériel au temps t_0 , pour le lancer de la pièce de monnaie ou encore le sondage d'intention de vote au second tour est un exemple de modèle paramétrique.

Le paramétrage n'est pas forcément unique. Dans exemple précédent de Bernoulli, on peut paramétrer par la probabilité que le matériel soit en panne au temps t_0 , c'est à dire $1 - p$, ou bien encore par toute fonction (bijective) de p . Comme par exemple par $\eta = \ln(p/(1 - p))$, ce qui veut dire que $p = e^\eta/(1 + e^\eta)$. Dans ce dernier cas le modèle statistique s'écrit :

$$(E, \mathcal{E}, \mathcal{P}) = (E, \mathcal{E}, \{B(1, e^\eta/(1 + e^\eta)) : \eta \in \mathbb{R}\})$$

Nous verrons un peu plus loin (dans la partie sur les familles exponentielles) que cette paramétrisation n'est pas aussi farfelue qu'on aurait pu le penser de prime abord.

Remarquons que l'on peut toujours paramétrer la famille \mathcal{P} , ne serait-ce qu'en prenant $\Theta = \mathcal{P}$ et donc l'application identité entre les deux espaces. Pour que l'on parle de modèle paramétrique, il faut que l'espace Θ soit de dimension finie, d'où l'hypothèse qu'il soit inclus dans un \mathbb{R}^d .

EXEMPLE 2.2. *Un problème de contrôle de la Qualité.*

Considérons une entreprise de fabrication de vis. On constate que les mesures du diamètre X d'une vis varient d'une pièce à l'autre. Cet aléa peut être dû au procédé de fabrication et/ou aux éventuelles erreurs de mesure. Supposons que l'on ne connaisse pas la valeur moyenne (rigoureusement l'espérance) du diamètre μ . Cherchons à préciser un modèle statistique adapté à une telle situation.

Il est souvent raisonnable d'admettre que la loi de X est normale. En effet de manière non rigoureuse on peut supposer que l'aléa est "symétrique et décroissant autour de la moyenne". On modélise donc souvent cette variation sous la forme :

$$X = \mu + \varepsilon,$$

où ε est de loi $N(0, \sigma^2)$. Autrement dit, on a

$$X \sim N(\mu, \sigma^2).$$

On suppose dans un premier temps σ^2 connu.

Pour modéliser cette situation on a donc recours au modèle statistique :

$$(E = \mathbb{R}, \mathcal{E} = \mathcal{B}_{\mathbb{R}}, \mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}\}).$$

Dans ce cas, on $\Theta = \mathbb{R}$ et $\theta = \mu$.

Si σ^2 est lui aussi inconnu, alors le modèle devient

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\})$$

et l'on a : $\Theta = \mathbb{R} \times \mathbb{R}_*^+$ et $\theta = (\mu, \sigma^2)$. Le paramètre est dit bi-dimensionnel. On peut aussi construire un modèle où l'espérance est connue et c'est la variance qui est inconnue. \diamond

DÉFINITION 2.3. *Un modèle paramétrique $(E, \mathcal{E}, \mathcal{P})$ est dit **identifiable** si la fonction $\theta \mapsto P_\theta$ de la Définition 2.2 est de plus injective, i.e. si*

$$\theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}.$$

Dans la plupart des cas le modèle est identifiable, quitte à prendre une autre paramétrisation. On supposera dans la suite que le modèle statistique est identifiable.

Abus de langage et de notation. Si la v.a. X est absolument continue, la densité de P_θ est notée f_θ . C'est une fonction intégrable de \mathbb{R} (ou une partie de \mathbb{R}) vers \mathbb{R}^+ . Si la v.a. X est discrète, on appellera également densité la fonction f_θ définie en tout x de l'espace E , où la X prend ses valeurs, par : $f_\theta(x) = P_\theta(X = x)$. On peut en effet montrer grâce à la théorie de la mesure, que dans ce dernier cas la loi de X est absolument continue par rapport à la mesure de comptage sur E . Les intégrales de la forme $\int_x \cdots dx$ utilisées dans le cas de v.a. absolument continues seront alors remplacées par des sommes de la forme $\sum_x \cdots$. Ainsi, par exemple, l'espérance s'écrit dans le cas continu $\int x f_\theta(x) dx$ et dans le cas discret $\sum_x x f_\theta(x) = \sum_x x P_\theta(X = x)$.

DÉFINITION 2.4. *On appelle **support** de la loi P_θ l'ensemble :*

$$\text{supp}(P_\theta) = \{x \in E : f_\theta(x) > 0\}.$$

On constate qu'il est dénombrable dans le cas de v.a. discrètes et infini non dénombrable dans le cas de v.a. absolument continues. Ce support peut dépendre de θ . Il en est ainsi par exemple dans le cas du modèle uniforme $\{\mathcal{U}_{[0,\theta]}; \theta > 0\}$

EXEMPLE 2.3.

Dans le cas de l'Exemple 2.1, on a :

$$f_\theta(x) = p^x (1-p)^{1-x},$$

pour tout $x \in \text{supp}(P_\theta) = \{0, 1\}$.

Dans le cas de l'Exemple 2.2, on a :

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

pour x dans $\text{supp}(P_\theta) = \mathbb{R}$. \diamond

2. Modèle d'échantillonnage

Pour étudier un phénomène aléatoire, on a souvent intérêt à observer plusieurs réalisations indépendantes de celui-ci. C'est ce que l'on a fait dans l'exemple du premier chapitre. On parle alors d'échantillon ou d'échantillonnage.

DÉFINITION 2.5. *On appelle n -échantillon de la loi P_θ , la donnée d'un vecteur $\underline{X} = (X_1, \dots, X_n)$ constitué de n v.a. indépendantes et identiquement distribuées (i.i.d.) de loi P_θ .*

*On appelle **modèle d'échantillonnage**, le modèle*

$$(E^n, \mathcal{E}^{\otimes n}, \mathcal{P}^n = \{P_\theta^{\otimes n} : \theta \in \Theta\}),$$

où $\mathcal{E}^{\otimes n}$ est la tribu produit (engendrée par les pavés) sur E^n et $P_\theta^{\otimes n} = P_\theta \otimes \dots \otimes P_\theta$ est la probabilité produit sur $(E^n, \mathcal{E}^{\otimes n})$ qui est la loi du vecteur $\underline{X} = (X_1, \dots, X_n)$ (Cf. cours de Probabilités).

Toutes les v.a. ont même loi, donc même valeur de θ . Un échantillon est un vecteur aléatoire. Sa réalisation, fruit de n observations indépendantes du même phénomène, est notée $\underline{x} = (x_1, \dots, x_n)$. On fera toujours cette distinction entre v.a. et sa réalisation en utilisant majuscules ou minuscules.

Un modèle d'échantillonnage est donc un modèle statistique particulier, où l'espace des observations est de la forme E^n , muni de sa tribu produit classique et de probabilités de la forme $P_\theta^{\otimes n}$. Aussi parfois on parlera dans ce cas simplement de modèle statistique. L'important est de bien avoir en tête quelle est la nature des observations : par exemple v.a.r., vecteur aléatoire (mais avec composantes non nécessairement indépendantes, ni de même loi) ou encore échantillon...

Grâce à l'indépendance et l'identique distribution, la densité de l'échantillon sous la loi P_θ est alors :

$$\underline{x} = (x_1, \dots, x_n) \mapsto \prod_{i=1}^n f_\theta(x_i),$$

pour tout \underline{x} de E^n . Si on considère le produit de droite non plus comme une fonction de x mais comme une fonction du paramètre θ , pour un $\underline{x} = (x_1, \dots, x_n)$ fixé, on parle de vraisemblance.

3. Vraisemblance

DÉFINITION 2.6. *Dans un modèle statistique paramétrique $(E, \mathcal{E}, \mathcal{P})$, on appelle **vraisemblance** de l'observation x la fonction*

$$\begin{aligned} L(x; \cdot) : \Theta &\rightarrow \mathbb{R}^+ \\ \theta &\mapsto L(x; \theta) = f_\theta(x). \end{aligned}$$

Bien sûr, dans le cas d'un modèle d'échantillonnage, la vraisemblance de l'échantillon observé $\underline{x} = (x_1, \dots, x_n)$ s'écrit sous la forme

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_{\theta}(x_i).$$

C'est donc la loi conjointe du n -échantillon évaluée aux valeurs observées et considérée comme fonction du paramètre θ .

4. Familles Exponentielles

Un modèle paramétrique important en Statistique est celui des familles exponentielles. Il recouvre de nombreux modèles paramétriques classiques : normal, binomial, poisson, gamma etc...

DÉFINITION 2.7. *Un modèle statistique $(E, \mathcal{E}, \mathcal{P})$ sur un espace des observations E est dit **famille exponentielle générale** s'il existe un entier p , des fonctions η , T , C et h tels que les densités puissent s'écrire, pour tout θ de Θ , sous la forme :*

$$f_{\theta}(x) = e^{\langle \eta(\theta), T(x) \rangle} C(\theta) h(x),$$

avec les contraintes que

- T soit une fonction mesurable à valeurs dans \mathbb{R}^p ;
- η soit une fonction à valeurs dans \mathbb{R}^p ;
- C soit une fonction réelle positive qui ne dépend pas de x ;
- h soit une fonction borélienne positive qui ne dépend pas de θ .

Le vecteur aléatoire $T(X)$ est appelé **statistique canonique** du modèle. Si la fonction T est l'identité, la famille exponentielle est dite **naturelle**.

On parle de **forme canonique** d'une famille exponentielle générale quand les densités de probabilités ont la forme

$$f_{\theta}(x) = e^{\langle \theta, T(x) \rangle} C(\theta) h(x),$$

pour tout θ de Θ , ce qu'il est toujours possible d'obtenir quitte à reparamétriser la famille par $\theta' = \eta(\theta)$. Dans ce cas le paramètre θ de la famille exponentielle est appelé **paramètre canonique**.

EXEMPLE 2.4.

Revenons sur le modèle de Bernoulli. La densité s'écrit :

$$\begin{aligned} f_p(x) &= p^x (1-p)^{1-x} = \left(\frac{p}{1-p} \right)^x (1-p) = \exp \left(x \ln \left(\frac{p}{1-p} \right) \right) (1-p) \\ &= \exp(\langle \eta(p), T(x) \rangle) C(p) h(x), \end{aligned}$$

avec

$$\eta(p) = \ln \left(\frac{p}{1-p} \right), \quad T(x) = x, \quad C(p) = (1-p) \text{ et } h(x) = 1.$$

Le modèle de Bernoulli est donc une famille exponentielle naturelle puisque $T = Id$. De plus, le modèle Bernoulli paramétré en fonction de η

$$(E, \mathcal{E}, \mathcal{P}) = (E, \mathcal{E}, \{B(1, e^\eta/(1 + e^\eta)) : \eta \in \mathbb{R}\})$$

est sous forme canonique.

Modèle échantillonnage construit à partir d'une famille exponentielle générale canonique reste une famille exponentielle générale canonique.

En effet si $\underline{X} = (X_1, \dots, X_n)$ est un échantillon de loi de densité

$$f_\theta(x) = e^{\langle \theta, T(x) \rangle} C(\theta) h(x),$$

alors le vecteur aléatoire \underline{X} a pour densité

$$f_\theta(x_1, \dots, x_n) = e^{\langle \theta, \sum_{i=1}^n T(x_i) \rangle} C^n(\theta) \prod_{i=1}^n h(x_i)$$

et $\sum_{i=1}^n T(X_i)$ est la statistique canonique du nouveau modèle.

On en déduit l'expression de la vraisemblance pour un échantillon $\underline{x} = (x_1, \dots, x_n)$ d'une famille exponentielle générale.

PROPOSITION 2.8. *La vraisemblance pour un échantillon $\underline{x} = (x_1, \dots, x_n)$ d'une famille exponentielle générale canonique est la fonction :*

$$\theta \mapsto L(x_1, \dots, x_n; \theta) = e^{\langle \theta, \sum_{i=1}^n T(x_i) \rangle} C^n(\theta) \prod_{i=1}^n h(x_i).$$

5. Modèle position-échelle

DÉFINITION 2.9. *Considérons un vecteur aléatoire X de loi P connue sur $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ et A un sous espace de \mathbb{R}^n . Pour tout a dans A et tout b dans \mathbb{R}^+ , on note $P_{a,b}$ la loi du vecteur $Y = a + bX$.*

Le modèle paramétrique

$$\mathcal{P}_{A,b} = \{P_{a,b} : a \in A, b \in \mathbb{R}^+\}$$

*est appelé **modèle position-échelle** engendré par P (ou par X). Le paramètre a est appelé paramètre de position et b paramètre d'échelle.*

Si b est fixé (par exemple à 1) on parle de modèle de position. Dans le cas où A ne contient que le vecteur nul de \mathbb{R}^n , on parle de modèle échelle.

EXEMPLE 2.5. *Le Modèle gaussien unidimensionnel*

Reprise de l'Exemple 2.2. Le modèle

$$\mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}\}$$

est un modèle position engendré par la loi $N(0, \sigma^2)$. Il correspond aux différentes lois du modèle pour le diamètre X de la vis. Rappelons que $X = \mu + \varepsilon$, où μ varie dans \mathbb{R} et ε est de loi $N(0, \sigma^2)$.

Le modèle

$$\mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$$

est un modèle position-échelle engendré par la loi $N(0, 1)$. Le diamètre X de la vis peut en effet s'écrire $X = \mu + \sigma\varepsilon$, où ε est de loi $N(0, 1)$. \diamond

6. Exercices

Exercice 1 (*Familles Exponentielles*)

On considère les modèles suivants :

- Modèle Binomial $\{B(m, p) : p \in [0, 1]\}$;
- Modèle de Poisson $\{\mathcal{P}(\lambda) : \lambda > 0\}$;
- Modèle gaussien à variance fixée $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}\}$;
- Modèle gaussien à paramètre bi-dimensionnel $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$;
- Modèle Gamma $\{G(\alpha, \beta) : \alpha > 0, \beta > 0\} = \{f_{\alpha, \beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbf{1}_{\mathbb{R}^+}(x) : \alpha > 0, \beta > 0\}$;
- Modèle uniforme $\{\mathcal{U}_{[0, \theta]} : \theta > 0\}$;
- Modèle de Cauchy $\{f_\theta(x) = \frac{1}{\pi(1+(x-\theta)^2)} : \theta \in \mathbb{R}\}$;
- Modèle Multinomial $\{\mathcal{M}(n, p_1, \dots, p_k) : 0 < p_i < 1, \forall i = 1, \dots, k \text{ et } \sum_{i=1}^k p_i = 1\}$.

Pour tous ces modèles, répondre aux questions suivantes.

- 1) Quelle est l'expression de la densité $f_\theta(x)$?
- 2) Le modèle constitue-t-il une famille exponentielle générale ? Naturelle ? Quel est le paramètre canonique du modèle ?
- 3) Quelle est la vraisemblance d'un échantillon $x = (x_1, \dots, x_n)$?

Exercice 2 (*Modèles position-échelle*)

1) Construire un modèle position-échelle à partir de la loi exponentielle $\mathcal{E}(1)$. Préciser la forme des f.d.r. des lois de ce modèle ainsi que leurs densités.

2) Montrer que le modèle uniforme $\{\mathcal{U}_{[a, b]} : -\infty < a < b < +\infty\}$ est un modèle position-échelle.

Exercice 3 (*Statistiques d'ordre*)

Soit X_1, \dots, X_n des v.a.r. définies sur un même espace probabilisé $(\Omega, \mathfrak{A}, P)$, indépendantes et de même loi absolument continue par rapport à la mesure de Lebesgue de densité f . Pour tout ω dans Ω , on peut ordonner les réels $X_1(\omega), \dots, X_i(\omega), \dots, X_n(\omega)$ sous la forme

$$X_{(1)}(\omega) \leq X_{(2)}(\omega) \leq \dots \leq X_{(i)}(\omega) \leq \dots \leq X_{(n)}(\omega).$$

L'application

$$X_{(i)} : \omega \in \Omega \rightarrow X_{(i)}(\omega)$$

ainsi définie pour chaque i est une v.a.r. dite $i^{\text{ème}}$ statistique d'ordre.

- 1) Calculer la loi de $X_{(n)} = \sup\{X_1, \dots, X_n\}$ (f.d.r. et densité).
- 2) Calculer la loi de $X_{(1)} = \inf\{X_1, \dots, X_n\}$ (f.d.r. et densité).
- 3) Calculer la loi du couple $(X_{(1)}, X_{(n)})$. En déduire celle de l'étendue $R = X_{(n)} - X_{(1)}$ (on donnera sa f.d.r et sa densité en fonction de F et f).
- 4) Soit N_y le nombre de X_i inférieurs à y . Quelle est la loi de N_y ? Que dire des événements $\{N_y \geq k\}$ et $\{X_{(k)} \leq y\}$? En déduire la f.d.r. de $X_{(k)}$.

5) On pourrait du résultat précédent tirer la densité de la v.a. $X_{(k)}$. Mais c'est fastidieux. Il y a bien plus simple en attaquant le problème directement, ce que l'on propose de faire maintenant. On pourra utiliser le résultat suivant : Si f est continue sur un intervalle $[a, b]$, alors, pour tout x dans cet intervalle, on a :

$$f(x) = \lim_{h \rightarrow 0^+} \frac{P(X \in]x, x + h])}{h}$$

Calculer la densité de $X_{(k)}$.

- 6) Montrer que si $\mathbb{E}(X)$ existe alors $\mathbb{E}(X_{(k)})$ aussi.
 - 7) Calculer la densité du vecteur $(X_{(1)}, \dots, X_{(n)})$.
- (Ind. on pourra calculer $P((X_{(1)}, \dots, X_{(n)}) \in B)$, pour tout borélien B de $\mathcal{B}_{\mathbb{R}^n}$).

Partie 2

Estimation ponctuelle

Reprenons l'exemple inaugural, vu au Chapitre 1, sur estimation de $\bar{F}(t_0)$, à partir d'un échantillon x_1, \dots, x_n de temps observés. On est parti de cet échantillon, avec ces n temps relevés, pour finalement utiliser seulement le nombre de temps observés supérieurs à t_0 . Ce fut également le cas dans l'écriture de ce modèle avec les y en fin de ce Chapitre. On a donc naturellement réduit l'information apportée par l'échantillon, pour ne garder que ce qui nous semblait utile dans l'objectif d'estimer $\bar{F}(t_0)$, où encore p_0 dans la seconde écriture de ce problème. C'est là la notion de Statistique : réduire l'information apportée par un échantillon. On parlera naturellement d'estimateur quand elle sera utilisée pour estimer le paramètre inconnu. Notons enfin que, toujours dans cet exemple, nous avons proposé plusieurs estimateurs et que naturellement se pose la question de la qualité et des propriétés d'un estimateur, de savoir comment comparer des estimateurs entre eux ou encore savoir comment améliorer un estimateur. C'est là le programme de cette partie.

CHAPITRE 3

Statistique et Estimateur

Comme nous l'avons dit une Statistique est une réduction de l'information apportée par un échantillon. Plus précisément voici sa définition.

DÉFINITION 3.1. Soit $(E^n, \mathcal{E}^{\otimes n}, \mathcal{P}^n = \{P_\theta^{\otimes n} : \theta \in \Theta\})$ un modèle d'échantillonnage. On appelle **statistique** la v.a. $T(\underline{X}) = T(X_1, \dots, X_n)$ où T est une fonction mesurable connue de $(E^n, \mathcal{E}^{\otimes n}, \mathcal{P}^n = \{P_\theta^{\otimes n} : \theta \in \Theta\})$ vers un espace probabilisable (F, \mathcal{F}) :

$$T : \begin{cases} E^n & \rightarrow F \\ \underline{x} = (x_1, \dots, x_n) & \mapsto T(x_1, \dots, x_n) \end{cases} .$$

Insistons bien sur le fait qu'une statistique est une v.a. Les valeurs qu'elle prendra dépendront des valeurs prises par l'échantillon. Si le modèle statistique est non trivial (i.e. non réduit à une seule probabilité) alors la loi de la statistique $T(\underline{X})$, où $\underline{X} = (X_1, \dots, X_n)$, est inconnue. Cela vient de la non connaissance de la loi de l'échantillon. En revanche la fonction T est, elle, connue. Reprenons l'exemple de l'estimation de la fiabilité vu au début du chapitre précédent. La fonction

$$T : (x_1, \dots, x_n) \rightarrow \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[t_0, +\infty[}(x_i)$$

est parfaitement connue. En revanche, la loi de la statistique $T(X_1, \dots, X_n)$ n'est pas entièrement connue puisque l'on sait seulement que la loi de $nT(X_1, \dots, X_n)$ est une Binomiale $B(n, \bar{F}(t_0))$, où la valeur de $\bar{F}(t_0)$ est inconnue.

Souvent l'espace d'arrivée de T est de dimension inférieure et plus simple que E^n (signe d'une effective réduction de l'information). Dire que la statistique est connue, signifie en particulier que la fonction connue T ne doit pas dépendre du paramètre (inconnu) θ (ou de la loi P dans \mathcal{P}). En revanche, la loi de la statistique $T(\underline{X})$ dépendra en général du paramètre du modèle.

Une statistique dépend de la taille n de l'échantillon, et on notera parfois $T_n(\underline{X})$ pour le souligner. Par abus de langage on appellera également statistique la suite $T(\underline{X}) = (T_n(\underline{X}))_{n \in \mathbb{N}}$ de statistiques quand la taille de l'échantillon augmente.

On a déjà vu dans l'exemple inaugural qu'un problème en statistique est d'estimer un paramètre θ . On peut aussi vouloir estimer $g(\theta)$, l'image de θ par une fonction g . Pour rester le plus général, on considérera dans la suite le cas général de l'estimation de $g(\theta)$. Pour ce faire, on utilise alors une statistique qui peut alors porter le nom d'estimateur.

DÉFINITION 3.2. On appelle estimateur de $g(\theta)$, toute statistique $T(\underline{X})$ de $(E^n, \mathcal{E}^{\otimes n})$ à valeurs dans $g(\Theta)$.

La seule contrainte apportée est donc que la statistique prenne ses valeurs dans $g(\Theta)$. Pour un même problème d'estimation, on pourra considérer de nombreux estimateurs. Cela dit, tous les estimateurs ne sont pas forcément judicieux.

Notation. Quand il s'agit d'estimer le paramètre θ on note souvent $\hat{\theta}$ son estimateur et $\hat{\theta}_n$ quand on souhaite préciser la taille n de l'échantillon. Pour l'estimation de $g(\theta)$ on utilise parfois aussi la notation $\widehat{g(\theta)}$.

EXEMPLE 3.1.

Considérons le modèle d'échantillonnage tiré du modèle paramétrique uniforme : $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{U}_{[0,\theta]} : \theta > 0\})$. Les densités dans ce modèle sont donc de la forme :

$$f_{\theta}(x) = \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x),$$

pour $\theta \in \mathbb{R}^+$.

Supposons que l'on cherche à estimer le paramètre θ à partir d'un échantillon X_1, \dots, X_n tiré de ce modèle. Plusieurs estimations sont possibles :

- Assez naturellement on pense en premier lieu à $\hat{\theta}_1 = \max(X_1, \dots, X_n)$. On sent naturellement (et on pourra le vérifier plus tard) qu'il s'approche en croissant de θ quand la taille de l'échantillon augmente. Mais seulement, il a le désavantage d'être toujours inférieur à la "vraie" valeur θ . On peut essayer de remédier à ce problème en proposant d'autres estimateurs.
- On peut se dire que les réalisations de l'échantillon vont se répartir de manière à constituer $n + 1$ intervalles de longueurs (très) approximativement égales. D'où l'idée de considérer

$$\hat{\theta}_2 = \hat{\theta}_1 + \frac{\hat{\theta}_1}{n} = \frac{n+1}{n} \hat{\theta}_1.$$

- On peut aussi remarquer que la distance qui sépare $\theta_1 = \max(X_1, \dots, X_n)$ de θ devrait être environ égale à celle qui sépare 0 de $\min(X_1, \dots, X_n)$. On peut donc proposer l'estimateur

$$\hat{\theta}_3 = \min(X_1, \dots, X_n) + \max(X_1, \dots, X_n).$$

- On peut aussi adopter une démarche radicalement différente basée sur l'intuition que $\bar{X} = (\sum_{i=1}^n X_i)/n$ devrait être un bon estimateur du centre du support de la loi uniforme, à savoir $\theta/2$. Aussi, on peut s'intéresser à l'estimateur : $\hat{\theta}_4 = 2\bar{X}$.
- On pourrait proposer de manière assez irraisonnée (puisqu'elle n'est pas basée sur l'échantillon) $\hat{\theta}_5 = 2011$ ou toute autre valeur.
- Il y aurait bien d'autres possibilités à explorer... ◇

Deux questions se posent à la suite de cet exemple. Existe-t-il des méthodes générales pour construire de (bons) estimateurs ? Et ensuite comment les comparer

ou savoir quel est le meilleur ? Ces questions font respectivement l'objet des deux sections suivantes.

CHAPITRE 4

Construction d'estimateurs

1. Estimateurs empiriques (des moments)

On a déjà vu dans exemple introductif et l'Exemple 3.1, comment estimer l'espérance mathématique d'une v.a. Étudions davantage cet estimateur qui est à la base de nombreuses méthodes statistiques.

Soit donc X une v.a. générique d'un modèle d'échantillonnage $(E^n, \mathcal{E}^{\otimes n}, \mathcal{P}^n = \{P_\theta^{\otimes n} : \theta \in \Theta\})$. C'est à dire que X_1, \dots, X_n est un échantillon de même loi que X . Notons $\mathbb{E}_\theta(\cdot)$ et $\text{Var}_\theta(\cdot)$ respectivement les opérateurs espérance et variance sous la loi P_θ , en supposant que ces quantités sont bien définies. Pour simplifier les notations, on notera $m_\theta = \mathbb{E}_\theta(X)$ et $\sigma_\theta^2 = \text{Var}_\theta(X)$.

DÉFINITION 4.1. On appelle **moyenne empirique**, la statistique \bar{X} définie, pour une taille n d'échantillon, par :

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

Quand on peut écrire l'espérance de la v.a. générique X en fonction du paramètre du modèle, i.e. quand il existe une fonction g telle que $m_\theta = g(\theta)$ (ce qui est souvent le cas), alors on pourra donner le titre d'estimateur à \bar{X} . On dira alors qu'il estime m_θ .

PROPOSITION 4.2. La moyenne empirique est telle que

$$\begin{aligned} \mathbb{E}_\theta(\bar{X}_n) &= m_\theta \\ \text{Var}(\bar{X}_n) &= \frac{\sigma_\theta^2}{n}. \end{aligned}$$

PREUVE. Immédiate par linéarité de l'espérance et grâce à l'indépendance entre les termes pour le calcul de la variance. \square

Le premier point de la proposition montre que l'estimateur \bar{X} est, dans un certain sens, un "bon" estimateur de l'espérance m_θ puisqu'il est égal en espérance à ce qu'il cherche à estimer. On parlera d'estimateur sans biais. Nous y reviendrons au chapitre suivant.

Une généralisation évidente de ce qui précède est donnée par l'estimation empirique d'un moment de X d'ordre quelconque. Notons $m_\theta(p) = \mathbb{E}_\theta(X^p)$ le moment d'ordre p de X sous la loi P_θ , en supposant que celui-ci existe. Par analogie avec ce qui précède, on peut définir l'estimateur empirique du moment d'ordre p .

DÉFINITION 4.3. On appelle *estimateur empirique du moment d'ordre p* , la statistique

$$\hat{m}_\theta(p) = \frac{1}{n} \sum_{i=1}^n X_i^p.$$

On peut aussi s'intéresser à l'estimation de la variance σ_θ^2 . Le raisonnement est le même. On sait que l'on peut écrire :

$$\sigma_\theta^2 = \mathbb{E}_\theta(X^2) - \mathbb{E}_\theta^2(X) = m_\theta(2) - (m_\theta(1))^2.$$

D'où l'idée d'estimer σ_θ^2 par

$$S_n^2 = \hat{m}_\theta(2) - (\hat{m}_\theta(1))^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2.$$

Un calcul élémentaire montre que S_n^2 s'écrit aussi sous la forme :

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

C'est sous cette forme qu'est plus connu cet estimateur.

DÉFINITION 4.4. On appelle *estimateur de la variance empirique*, la statistique S_n^2 définie pour une taille n d'échantillon par :

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Cette méthode d'estimation empirique des moments est très générale. Elle peut, par exemple, s'appliquer pour l'estimation de la fonction de répartition. Il suffit en effet de remarquer que l'on peut écrire

$$F_\theta(x) = P_\theta(X \leq x) = \mathbb{E}_\theta(\mathbb{1}_{\{X \leq x\}}) = \mathbb{E}(Y),$$

avec $Y = \mathbb{1}_{]-\infty, x]}(X)$. On peut donc estimer $F_\theta(x)$ par

$$\hat{F}_\theta(x) = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{]-\infty, x]}(X_i)$$

et on retrouve l'estimateur de la fonction de répartition empirique.

2. Méthode de substitution

Principe de la méthode. Supposons que l'on sache estimer le paramètre θ d'un modèle statistique. On note $\hat{\theta}_n$ cet estimateur. Et supposons également que l'on soit intéressé par l'estimation de l'image $g(\theta)$ de ce paramètre par une application g (connue). La méthode de substitution (ou de "plug-in" en anglais), consiste à utiliser l'estimateur $g(\hat{\theta}_n)$. On verra ultérieurement que si la fonction g est continue on pourra aisément obtenir des informations sur la qualité de cet estimateur à partir de celles de $\hat{\theta}_n$.

Cette méthode a aussi été déjà utilisée, de manière assez naturelle, dans l'exemple introductif pour l'estimation de $\bar{F}(t_0)$ par $\exp(-\hat{\lambda}t_0)$. À partir d'un estimateur de λ nous avons obtenu un estimateur de $\bar{F}(t_0)$ qui est une fonction de λ .

Un autre exemple d'utilisation de cette méthode est le suivant. On a vu comment estimer la variance $\sigma_\theta^2 = \text{Var}_\theta(X)$ d'une v.a.r. par S_n^2 . Si l'on veut estimer son écart-type, on peut prendre

$$\hat{\sigma}_\theta = \sqrt{S_n^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}.$$

3. Méthode des moments

Principe de la méthode. Supposons qu'il existe une fonction h bijective et continue de $\Theta \subset \mathbb{R}^p$ vers $h(\Theta) \subset \mathbb{R}^p$, une fonction mesurable φ de E vers \mathbb{R}^p telle que $\mathbb{E}_\theta(\varphi(X))$ existe et toutes les deux telles que l'on ait :

$$h(\theta) = \mathbb{E}_\theta(\varphi(X)),$$

pour tout θ de Θ .

La méthode des moments consiste alors à estimer θ par

$$\hat{\theta}_n(\underline{X}) = h^{-1} \left(\frac{1}{n} \sum_{i=1}^n \varphi(X_i) \right).$$

Cette méthode a également déjà été utilisée dans l'exemple introductif où nous avons proposé un estimateur du paramètre λ du modèle exponentiel.

Cette méthode peut être vue comme un mélange des deux précédentes méthodes.

EXEMPLE 4.1. *Modèle de la loi exponentielle.*

L'équation

$$\mathbb{E}(X) = \frac{1}{\lambda}$$

s'écrit sous la forme $h(\lambda) = \mathbb{E}_\theta(\varphi(X))$ avec $h(x) = 1/x$ et φ est l'identité sur \mathbb{R}^+ . En estimant λ par

$$\hat{\lambda} = h^{-1}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{\bar{X}_n},$$

on retrouve l'estimateur utilisé dans l'exemple introductif sur la fiabilité des matériels. \diamond

EXEMPLE 4.2. *Modèles gaussiens unidimensionnels.*

Considérons les modèles statistiques introduits pour des problèmes de contrôle de la qualité dans l'Exemple 2.2. Dans le premier modèle,

$$\mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}\}$$

l'estimation par la méthode des moments redonne l'estimateur intuitif de μ par \bar{X}_n , en prenant h et φ égales aux fonctions identité puisque $\mu = \mathbb{E}(X)$.

Dans le second modèle,

$$\mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\},$$

on peut écrire la relation $h(\theta) = \mathbb{E}_\theta(\varphi(X))$ en prenant $\theta = (\mu, \sigma^2)$, $h(\theta) = (\mu, \sigma^2 + \mu^2)$ et $\varphi(x) = (x, x^2)$ application de \mathbb{R} vers $\mathbb{R} \times \mathbb{R}^+$. En effet on sait que l'on a $\mathbb{E}(X^2) = \text{Var}(X) + \mathbb{E}^2(X)$. Comme $h^{-1}(u, v) = (u, v - u^2)$, on obtient comme estimateur du paramètre multidimensionnel θ par la méthode des moments :

$$\hat{\theta}_n(\underline{X}) = h^{-1}\left(\frac{1}{n} \sum_{i=1}^n \varphi(X_i)\right) = h^{-1}\left(\frac{\bar{X}_n}{\frac{1}{n} \sum_{i=1}^n X_i^2}\right) = \left(\frac{\bar{X}_n}{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2}\right).$$

On retrouve les estimateurs vus par la méthode des moments empiriques. \diamond

On parle de *la* méthode des moments, mais il faut bien retenir que l'unicité porte sur la méthode (qui est très générale) et non sur le nombre d'estimateurs d'un même paramètre que l'on peut obtenir par cette méthode. Nous verrons dans un exercice que, par exemple, dans le modèle de la loi exponentielle (comme dans d'autres), la méthode des moments permet d'obtenir de nombreux autres estimateurs du paramètre λ . On s'inspirera en particulier des relations $\mathbb{E}(X^2) = 2/\lambda^2$ et $\bar{F}(t_0) = P(X > t_0)$ pour obtenir deux autres estimateurs par cette méthode des moments.

4. Maximum de vraisemblance

Pour introduire cette approche, considérons deux urnes contenant toutes les deux des boules bleues et rouges mais en proportion différentes : proportion $p_1 = 90\%$ de boules bleues dans la première et proportion $p_2 = 15\%$ de boules bleues dans la seconde. On tire au hasard une boule dans une des deux urnes sans savoir de laquelle il s'agit. On constate que la boule est bleue. Naturellement on parierait plutôt que la boule tirée est issue de l'urne 1. On a pris l'urne qui maximise la probabilité de l'événement que l'on a obtenu : "avoir une boule bleue". On a choisi la situation la plus vraisemblable. On va voir que c'est celle qui maximise la vraisemblance. En effet, le modèle est ici :

$$\{B(p); p \in \{0.9, 0.15\}\},$$

de v.a. générique X où $\{X = 1\}$ signifie que la boule tirée est bleue et $\{X = 0\}$ signifie qu'elle est rouge. La vraisemblance d'un tirage d'une boule bleue est donc $L(1; p) = p$ pour p dans $\{0, 9; 0, 15\}$. En prenant la valeur de $p = 0.9$ qui maximise la vraisemblance, on décide donc que la boule provient de l'urne 1. Si la boule avait été rouge, on aurait évidemment choisi l'autre urne, ce qui maximise également la vraisemblance qui est alors : $L(0, p) = 1 - p$.

Généralisons un peu. Supposons que l'on ait une infinité d'urnes avec toutes les proportions possibles p de boules bleues comprises entre 0 et 1. On effectue n tirages i.i.d. dans une même urne (inconnue) et on note respectivement X_1, \dots, X_n le résultat de chaque tirage ($x_i = 1$ si la boule tirée est bleue et 0 sinon). En s'inspirant de la méthode adoptée ci-dessus, on peut choisir d'estimer p par la valeur qui maximise la vraisemblance de l'événement observé qui est le vecteur $\underline{x} = (x_1, \dots, x_n)$. La vraisemblance est

$$L(x_1, \dots, x_n; p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}$$

et l'estimateur \hat{p} est donc défini par

$$\hat{p} = \underset{p}{\operatorname{Argmax}} L(\underline{x}; p).$$

La fonction logarithme étant croissante, on peut écrire

$$\hat{p} = \underset{p}{\operatorname{Argmax}} \left(\ln(p) \sum_{i=1}^n x_i + \left(n - \sum_{i=1}^n x_i \right) \ln(1-p) \right).$$

La fonction $p \mapsto \varphi(p) = \ln(p) \sum_{i=1}^n x_i + (n - \sum_{i=1}^n x_i) \ln(1-p)$ admet pour dérivée

$$\varphi'(p) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p}$$

et dérivée seconde

$$\varphi''(p) = -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n - \sum_{i=1}^n x_i}{(1-p)^2}$$

qui est négative puisque les x_i sont dans $\{0, 1\}$. La fonction φ est donc concave et son maximum atteint en la valeur \hat{p} qui annule la dérivée première, i.e.

$$\frac{\sum_{i=1}^n x_i}{\hat{p}} = \frac{n - \sum_{i=1}^n x_i}{1 - \hat{p}}$$

ce qui donne $\hat{p} = \bar{x} = (\sum_{i=1}^n x_i)/n$. Remarquons que l'on obtiendrait le même estimateur en utilisant la méthode des moments puisque $\mathbb{E}(X) = p$. Mais il n'y a pas aucune raison que cette méthode conduise toujours aux mêmes estimateurs.

On peut généraliser cette méthode pour un modèle statistique quelconque.

DÉFINITION 4.5. Soit $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ un modèle statistique paramétrique et X sa v.a. générique. On appelle **estimateur du maximum de vraisemblance** la

statistique $\hat{\theta}(X)$ où $\hat{\theta}$ est une application :

$$\hat{\theta} : \begin{cases} E & \rightarrow \Theta \\ x & \mapsto \hat{\theta}(x) \end{cases}$$

telle que

$$L(x; \hat{\theta}(x)) \geq L(x; \theta)$$

pour tout $\theta \in \Theta$. On note

$$\hat{\theta}(x) = \underset{\theta}{\operatorname{Argmax}} L(x; \theta).$$

Dans le cas d'un modèle d'échantillonnage la variable générique est $\underline{X} = (X_1, \dots, X_n)$ et l'Estimateur du Maximum de Vraisemblance est

$$\hat{\theta}(\underline{X}) = \underset{\theta}{\operatorname{Argmax}} L(\underline{X}; \theta).$$

Il est bien évident que d'une part l'estimateur du maximum de vraisemblance n'existe pas toujours et que, d'autre part, s'il existe rien ne garantit qu'il soit unique.

Si la fonction vraisemblance est concave, on sait que le maximum est unique et atteint en la valeur qui annule la dérivée première (cas unidimensionnel) ou le gradient (cas multidimensionnel). Insistons bien sur le fait que cette méthode ne peut être utilisée que si l'hypothèse de concavité est vérifiée. Un contre-exemple est donné par le modèle de la loi uniforme que nous traiterons en exercice.

Comme la vraisemblance est souvent sous la forme d'un produit (modèle d'échantillonnage) il est généralement plus aisé (pour les dérivations) de travailler avec la **log-vraisemblance** définie comme le logarithme népérien de la vraisemblance. La fonction \ln étant croissante, l'estimateur obtenu en maximisant la log-vraisemblance est identique à l'estimateur du maximum de vraisemblance.

Si l'on porte notre intérêt sur l'estimation de $g(\theta)$ image de θ par une fonction g connue, alors la propriété suivante peut être utile.

PROPOSITION 4.6. (*Propriété d'invariance du maximum de vraisemblance*).

Soit $\hat{\theta}(X)$ un estimateur du maximum de vraisemblance dans un modèle paramétrique $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$. Soit g une fonction bijective, mesurable et connue de Θ vers $\Theta' = g(\Theta)$. L'estimateur du maximum de vraisemblance de $\eta = g(\theta)$ dans le modèle $\{P_\eta : \eta \in \Theta'\}$, paramétré par η , est alors $\hat{\eta} = g(\hat{\theta}(X))$.

La démonstration de cette proposition est évidente (basée sur la bijectivité de la nouvelle paramétrisation).

On peut en fait montrer (c'est plus délicat) que ce résultat est vrai pour une fonction g mesurable quelconque et on posera donc comme définition que l'estimateur du maximum de vraisemblance de $g(\theta)$ est $g(\hat{\theta}(X))$.

5. Exercices

Exercice 1 (*Modèle Gamma et Méthode des moments*)

On considère le Modèle Statistique de la loi Gamma $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{G(\alpha, \beta) : \alpha > 0, \beta > 0\})$. On rappelle que la densité d'une v.a. X de loi $G(\alpha, \beta)$ est :

$$f_{\alpha, \beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\mathbb{R}^+}(x).$$

- 1) Calculer $\mathbb{E}_{\alpha, \beta}(X)$ et $Var_{\alpha, \beta}(X)$.
 - 2) Par la méthode des moments, donner un estimateur du paramètre bidimensionnel (α, β) du modèle, basé sur l'observation d'un échantillon X_1, \dots, X_n .
 - 3) Déterminer des estimateurs de α et β en utilisant conjointement des estimateurs empiriques des moments et la méthode de substitution.
-

Exercice 2 (*Modèle de la loi exponentielle et Méthode des moments*)

On a vu en cours que la méthode des moments permet d'obtenir un estimateur du paramètre λ dans un modèle de la loi exponentielle : $\hat{\lambda} = 1/\bar{X}_n$ basé sur la relation $\mathbb{E}(X) = 1/\lambda$. L'intérêt de cet exercice est de montrer que cette méthode permet la construction de plusieurs estimateurs de ce même paramètre λ .

- 1) On suppose qu'une v.a.r. X suit une loi exponentielle $\mathcal{E}(\lambda)$. Calculer $\mathbb{E}(X^2)$.
 - 2) Écrire la fiabilité $\bar{F}(t_0) = P(X > t_0)$ sous forme d'une espérance.
 - 3) On considère le modèle de la loi exponentielle $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}(\lambda) : \lambda > 0\})$. En vous inspirant des résultats des deux questions précédentes et en utilisant à chaque fois la méthode des moments, proposer deux autres estimateurs du paramètre λ .
-

Exercice 3 (*Maximum de vraisemblance pour un modèle gaussien*)

1) On considère le modèle gaussien $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}\}$. Donner l'estimateur du maximum de vraisemblance du paramètre μ basé sur une observation x_1, \dots, x_n d'un échantillon issu de ce modèle.

2) On considère maintenant le modèle gaussien avec paramètre bidimensionnel, i.e. $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$. Donner l'estimateur du maximum de vraisemblance du paramètre $\theta = (\mu, \sigma^2)$, pour le modèle d'échantillonnage associé.

Exercice 4 (*Maximum de vraisemblance pour un modèle de loi uniforme*)

On considère le modèle uniforme $\{\mathcal{U}_{[0, \theta]} : \theta > 0\}$.

1) Montrer que la vraisemblance associée à un échantillon x_1, \dots, x_n observé dans ce modèle est :

$$L(x_1, \dots, x_n; \theta) = \frac{1}{\theta^n} \mathbb{1}_{x_{(1)} \geq 0} \mathbb{1}_{x_{(n)} \leq \theta}$$

où $x_{(1)}$ et $x_{(n)}$ sont respectivement les observations des statistiques d'ordre $X_{(1)}$ et $X_{(n)}$.

2) Donner l'estimateur du maximum de vraisemblance du paramètre θ .

Exercice 5 (*Modèles de la loi exponentielle et de la loi de Poisson en Fiabilité*)*Partie 1*

On s'intéresse à la durée de vie X d'un matériel électronique. Il est raisonnable de considérer que cette durée de vie est aléatoire et que sa loi est exponentielle (il existe des méthodes statistiques, mais que nous ne verrons pas dans le cadre de ce cours, pour vérifier cette hypothèse). En revanche, on ignore la valeur du paramètre λ de cette loi.

1) Écrire le modèle statistique engendré par X . Donner également le modèle d'échantillonnage associé.

2) Donner l'estimateur du maximum de vraisemblance pour une observation x_1, \dots, x_n d'un échantillon X_1, \dots, X_n de durées de vie de ces matériels.

3) Donner une estimation par maximum de vraisemblance de la quantité $\alpha = P(X > t_0)$, où t_0 est un temps fixé.

4) Quels estimateurs de λ et de α obtient-on si on utilise la méthode des moments ?

Partie 2

Supposons maintenant que les observations de ces durées de vie soient obtenues grâce à l'expérience suivante. Au temps $t = 0$, on dispose un matériel sur un banc d'essai. Quand celui-ci tombe en panne, on remplace immédiatement (ou on ne compte pas le temps de remplacement) le matériel défectueux par un matériel identique mais neuf. Et ainsi de suite jusqu'au temps t_0 . On note alors K le nombre de pannes relevées dans l'intervalle $[0, t_0]$.

5) Calculer la probabilité que K soit nul.

6) On note T_k le temps écoulé jusqu'à la k ème panne observée. C'est à dire que $T_k = X_1 + \dots + X_k$. Montrer que la loi de la v.a.r. T_k est une Gamma $G(k, \lambda)$ (Ind. On pourra utiliser la transformée de Laplace ou la fonction caractéristique).

7) Exprimer l'événement $K = k$ en fonction d'événements liant les v.a.r. T_k et X_{k+1} . En déduire que la loi de K est une loi de Poisson, dont on déterminera la valeur du paramètre.

Partie 3

On suppose que l'on réalise n fois cette expérience et on note K_1, \dots, K_n les nombres de pannes observées dans chaque intervalle $[0, t_0]$.

8) Donner le modèle statistique associé à ces observations.

9) Donner par la méthode du maximum de vraisemblance un autre estimateur du paramètre λ , basé cette fois sur les observations k_1, \dots, k_n .

10) Qu'obtient-on comme estimateur de λ si, dans ce modèle, on utilise la méthode des moments ?

Exercice 6 (*Maximum de vraisemblance*)

Pour les modèles suivants, donner l'estimateur du maximum de vraisemblance associé à l'observation d'un échantillon X_1, \dots, X_n .

1) Modèle de la loi exponentielle décalée :

$$(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}_{t_0}(\lambda) : \lambda > 0, t_0 \in \mathbb{R}\}).$$

On rappelle que la densité de la loi exponentielle décalée $\mathcal{E}_{t_0}(\lambda)$ est :

$$f_{\lambda, t_0}(x) = \lambda \exp(-\lambda(x - t_0)) \mathbb{1}_{[t_0, +\infty[}(x).$$

2) Modèle de la loi Bêta à un seul paramètre :

$$(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{Beta(1, \theta) : \theta > 1\}).$$

On rappelle que la densité de la loi $Beta(a, b)$ est :

$$f_{a,b}(x) = \frac{1}{\beta(a,b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{[0,1]}(x),$$

où $\beta(a, b)$ est la valeur de la fonction Eulérienne Bêta prise en a et b .

Ind. On pourra montrer en premier lieu que la densité pour le modèle considéré est :

$$f_{\theta}(x) = \theta(1-x)^{\theta-1} \mathbb{1}_{[0,1]}(x).$$

CHAPITRE 5

Qualité d'un estimateur

On a vu plusieurs techniques pour construire des estimateurs. Même si la présentation n'est pas exhaustive (par manque de temps...), abordons maintenant le problème de l'évaluation de la qualité d'un estimateur et la comparaison d'estimateurs entre-eux. Le but étant bien sûr de prendre le meilleur (s'il en existe un meilleur).

On l'a vu, un estimateur $T(\underline{X})$ de $g(\theta)$ est une v.a. Pour chaque échantillon observé, l'estimateur prendra de nouvelles valeurs. Il faut donc, pour parler de la qualité d'un estimateur, tenir compte de son comportement aléatoire.

A priori donc, l'estimateur ne donnera pas toujours (en fait même rarement) la bonne valeur $g(\theta)$. Dans le cas où $T(\underline{X})$ est absolument continu, il sera même *p.s.* toujours différent de la valeur fixe $g(\theta)$. Il est à noter que la présence d'erreur n'est pas toujours la conséquence des variations aléatoires de l'estimateur. Ainsi, si l'on revient sur l'estimateur du maximum de vraisemblance dans un modèle de Bernoulli (Cf. Section 4.4), on a vu que qu'il a pour expression \bar{x} . L'estimateur ne donnera donc jamais des valeurs en dehors de l'ensemble : $\{0, 1/n, 2/n, \dots, (n-1)/n, 1\}$. Il ne donnera donc jamais la bonne valeur p si cette dernière n'est pas dans cet ensemble.

Naturellement, on voudra qu'un estimateur possède quelques unes (à défaut de toutes) des qualités suivantes.

- Quand la taille d'échantillon augmente, l'estimateur a tendance à se rapprocher (dans un sens à définir) de la valeur $g(\theta)$ qu'il estime. On parlera dans ce cas d'estimateur convergent ou consistant.
- Même si l'estimateur commet une erreur d'estimation à chaque fois, "en moyenne" (en fait en espérance) il ne se trompe pas. On dira dans un tel cas que l'estimateur est sans biais.
- L'estimateur doit être le plus précis possible : les variations de l'estimateur autour de $g(\theta)$ doivent être réduites, voire les plus petites possible. On mesurera cette précision au moyen de la notion de fonction de risque.

Il y aurait d'autres critères, mais nous n'aurons pas le temps de les étudier.

1. Estimateur convergent

Lorsque l'on augmente la taille de l'échantillon, on augmente la quantité d'information dont on dispose sur le phénomène aléatoire que l'on étudie. Aussi, il est assez naturel de souhaiter qu'un estimateur ait tendance à s'approcher de la valeur qu'il estime, lorsque la taille de l'échantillon croît.

DÉFINITION 5.1. Un estimateur $T(\underline{X}) = (T_n(\underline{X}))_{n \in \mathbb{N}}$ de $g(\theta)$ est dit (**faiblement**) **convergent** ou **consistant** si la suite $(T_n(\underline{X}))_{n \in \mathbb{N}}$ converge en probabilité (sous la loi P_θ) vers $g(\theta)$, i.e.

$$T_n(\underline{X}) \xrightarrow{P_\theta} g(\theta),$$

quand $n \rightarrow +\infty$.

Si $T(\underline{X})$ et $g(\theta)$ sont dans \mathbb{R} , la définition de la convergence de l'estimateur signifie que l'on a, pour tout $\varepsilon > 0$:

$$P(|T_n(\underline{X}) - g(\theta)| > \varepsilon) \longrightarrow 0,$$

quand $n \rightarrow +\infty$.

Si $T(\underline{X})$ et $g(\theta)$ sont dans \mathbb{R}^p , la définition de la convergence de l'estimateur s'écrit à partir de la notion précédente sous la forme :

$$\|T_n(\underline{X}) - g(\theta)\| \xrightarrow{P_\theta} 0,$$

quand $n \rightarrow +\infty$ et où $\|\cdot\|$ est une norme quelconque dans \mathbb{R}^p . On peut montrer aisément que cela est équivalent à avoir la convergence en probabilité pour chaque coordonnée.

On peut bien sûr considérer d'autres types de convergence, comme la convergence *p.s.* ou la convergence dans L^p , pour p fixé. Dans ces cas, on dira respectivement que l'estimateur est **fortement convergent** ou **consistant** ou **L^p -convergent** ou **consistant**.

EXEMPLE 5.1. *Estimateurs de la moyenne empirique et de la variance empirique.*

Soit un modèle paramétrique $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ quelconque tel que l'espérance (en supposant qu'elle existe) de sa variable générique X s'écrive sous la forme $\mathbb{E}_\theta X = g(\theta)$. C'est par exemple trivialement le cas de l'Exemple 2.2 des modèles gaussiens unidimensionnels.

On a introduit dans la Définition 4.1 la moyenne empirique \bar{X}_n et vu qu'elle est un estimateur naturel de $\mathbb{E}_\theta X$. Par la loi des grands nombres il apparaît clairement que sous de bonnes hypothèses cet estimateur est consistant et même fortement convergent. On peut également démontrer (en supposant que les moments d'ordre 2 existent) qu'il est L^2 -convergent, en remarquant que l'on a $\mathbb{E}_\theta \bar{X}_n = g(\theta)$ et $\text{Var}_\theta(\bar{X}_n) = \sigma_\theta^2/n$ qui tend vers 0 quand $n \rightarrow +\infty$.

Supposons toujours que la variance de X existe et s'écrive sous la forme $\text{Var}_\theta(X) = h(\theta)$ dans ce modèle. On peut montrer que, sous certaines conditions, l'estimateur S_n^2 de la variance empirique, vu dans la Définition 4.4, est un estimateur consistant de $\text{Var}_\theta(X)$. En effet, en utilisant l'écriture

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2.$$

et en supposant par exemple l'existence de moment d'ordre 4 pour la v.a.r. X (ce qui permet d'appliquer également la loi des grands nombres pour la moyenne empirique des X_i^2 , pour $i = 1, \dots, n$), on obtient, grâce à la loi des grands nombres et au théorème de Slutsky (transformation continue), la convergence

$$S_n^2 \xrightarrow{P_\theta} \mathbb{E}_\theta(X^2) - \mathbb{E}_\theta^2(X) = \text{Var}_\theta(X),$$

quand $n \rightarrow +\infty$. ◇

2. Estimateur sans biais

Un autre critère de qualité d'un estimateur est celui de biais. On l'a vu un estimateur est une v.a. et à ce titre il varie donc (plus ou moins) autour de sa valeur centrale que représente son espérance. Il sera naturellement préférable que cette valeur centrale soit fixée sur la valeur $g(\theta)$ qu'il estime. Intuitivement, cela revient à demander que la moyenne prise par l'estimateur sur un grand nombre d'échantillons soit égale la "cible" $g(\theta)$.

DÉFINITION 5.2. *Le **biais** d'un estimateur $T(\underline{X}) = (T_n(\underline{X}))_{n \in \mathbb{N}}$ de $g(\theta)$ est la fonction b_T définie sur Θ par :*

$$b_T(\theta) = \mathbb{E}_\theta(T(\underline{X})) - g(\theta),$$

*pour tout θ dans Θ et à condition que $\mathbb{E}_\theta(T(\underline{X}))$ existe. Il est dit **sans biais** si cette fonction est identiquement nulle, i.e. si l'on a :*

$$\mathbb{E}_\theta(T(\underline{X})) = g(\theta),$$

pour tout θ dans Θ .

Pour les cas où $T(\underline{X})$ et $g(\theta)$ sont à valeurs dans \mathbb{R}^k , on rappelle que l'espérance d'un vecteur aléatoire est le vecteur de ses espérances.

Un estimateur non biaisé ne commet pas d'erreur systématique. A l'inverse, un estimateur sera dit biaisé positivement (resp. négativement) si la fonction biais est positive (resp. négative).

Le biais est généralement une fonction de la taille n de l'échantillon et on peut, si nécessaire la noter $b_{n,T}$ dans ce cas. Aussi, si certains estimateurs se trouvent être biaisés pour toute taille finie d'échantillon, on peut espérer qu'ils soient non biaisés asymptotiquement, c'est à dire quand n tend vers $+\infty$.

DÉFINITION 5.3. *Un estimateur $T(\underline{X}) = (T_n(\underline{X}))_{n \in \mathbb{N}}$ de $g(\theta)$, où $T_n(\underline{X})$ est intégrable pour tout n , est dit **asymptotiquement sans biais** si l'on a*

$$b_{n,T}(\theta) = \mathbb{E}_\theta(T_n(\underline{X})) - g(\theta) \longrightarrow 0,$$

quand $n \rightarrow +\infty$ et ce pour tout θ dans Θ .

EXEMPLE 5.2. *Estimateurs de la moyenne empirique et de la variance empirique (suite de l'Exemple 5.1).*

On a vu que les estimateurs de la moyenne empirique et de la variance empirique sont consistants. Qu'en est-il de leurs biais ?

On a déjà vu dans la Proposition 4.2 que :

$$\mathbb{E}_\theta(\bar{X}_n) = \mathbb{E}_\theta(X).$$

La moyenne empirique est donc un estimateur sans biais de l'espérance d'une v.a.r.

Intéressons nous maintenant à l'estimateur de la variance empirique. Remarquons en premier lieu que l'on peut écrire :

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}_\theta(X) + \mathbb{E}_\theta(X) - \bar{X})^2 \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}_\theta(X))^2 + \frac{1}{n} \sum_{i=1}^n (\mathbb{E}_\theta(X) - \bar{X})^2 \\ &\quad + \frac{2(\mathbb{E}_\theta(X) - \bar{X})}{n} \sum_{i=1}^n (X_i - \mathbb{E}_\theta(X)) \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}_\theta(X))^2 + (\mathbb{E}_\theta(X) - \bar{X})^2 - 2(\mathbb{E}_\theta(X) - \bar{X})^2 \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}_\theta(X))^2 - (\bar{X} - \mathbb{E}_\theta(X))^2. \end{aligned}$$

Ainsi il vient :

$$\begin{aligned} \mathbb{E}_\theta S_n^2 &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\theta (X_i - \mathbb{E}_\theta(X))^2 - \mathbb{E}_\theta (\bar{X} - \mathbb{E}_\theta(X))^2 = \text{Var}_\theta(X) - \text{Var}_\theta(\bar{X}) \\ &= \sigma_\theta^2 - \frac{\sigma_\theta^2}{n} = \frac{n-1}{n} \sigma_\theta^2. \end{aligned}$$

L'estimateur de la variance empirique est donc un estimateur biaisé. En revanche, il est clairement asymptotiquement sans biais.

Mais on peut aisément déduire de cet estimateur un estimateur non biaisé de la variance σ_θ^2 . \diamond

DÉFINITION 5.4. On appelle estimateur de la **variance empirique modifiée** l'estimateur

$$S_n^{2*} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Cet estimateur est sans biais et consistant d'après la consistance de S_n^2 .

3. Risque d'un estimateur

Une autre manière de mesurer la qualité d'un estimateur est d'évaluer sa précision. Dans cet objectif on peut faire appel à la théorie de la décision. En effet une estimation

peut être vue comme le choix d'une valeur d (qui sera donnée par la réalisation de la statistique $T(\underline{X})$) dans l'ensemble $g(\Theta)$ pour estimer la valeur inconnue $g(\theta)$.

Pour prendre une telle décision, on se donne en premier lieu un critère mesurant et pénalisant l'écart entre la valeur choisie d et la vraie valeur $g(\theta)$. On parle de fonction de coût.

DÉFINITION 5.5. *On appelle **fonction de coût** (ou de perte) toute fonction L de $g(\Theta) \times \Theta$ vers \mathbb{R}^+ , mesurable en sa première coordonnée, qui en (d, θ) donne le coût $L(d, \theta)$ de décider d alors que la vraie valeur est $g(\theta)$.*

De nombreux choix de fonctions de coût sont possibles. Dans un premier temps, on peut penser à des fonctions de la distance entre la décision d et la vraie valeur $g(\theta)$. Ainsi si $g(\Theta)$ est dans \mathbb{R} , on peut utiliser

- le coût absolu $L(d, \theta) = |d - g(\theta)|$,
- le coût quadratique $L(d, \theta) = (d - g(\theta))^2$
- ou encore tout coût de la forme $h(|d - g(\theta)|)$ où h est une fonction positive mesurable quelconque.

Si $g(\Theta)$ est dans \mathbb{R}^k , on pourra prendre des coûts de la forme $h(\|d - g(\theta)\|)$ où h est encore une fonction positive mesurable et $\|\cdot\|$ une norme sur \mathbb{R}^k . On parlera, par exemple, de coût quadratique si $L(d, \theta) = \|d - g(\theta)\|_2^2$ où $\|\cdot\|_2$ est la norme euclidienne dans \mathbb{R}^k . On peut aussi considérer la fonction de coût définie par

$$L(d, \theta) = (d - g(\theta))(d - g(\theta))'$$

qui est à valeur dans l'espace des matrices symétriques positives. On parle dans ce cas de coût quadratique multidimensionnel.

Tous les coûts précédents ont une propriété commune importante, celle d'être symétrique. On pénalise autant une surévaluation de $g(\theta)$ qu'une sous-évaluation. On pourrait aussi considérer des fonctions de coût non symétriques, mais ce ne sera pas le cas dans le cadre de ce cours.

Nous avons déjà vu que la décision que l'on prendra est donnée par la réalisation $T(\underline{x})$ d'une statistique $T(\underline{X})$. Le coût associé à cette décision est $L(T(\underline{x}), \theta)$ et varie donc d'un échantillon à l'autre. Cette notion ne peut, à elle seule, définir la qualité d'un estimateur. Elle serait sinon dépendante du hasard lié à l'échantillon observé. C'est pourquoi on utilise une notion de coût moyen, que l'on appelle également risque.

DÉFINITION 5.6. *On appelle **risque** d'un estimateur $T(\underline{X}) = (T_n(\underline{X}))_{n \in \mathbb{N}}$ de $g(\theta)$ associé à la fonction de coût L , la fonction R de Θ vers \mathbb{R}^+ définie par*

$$R(T(\underline{X}), \theta) = \mathbb{E}_\theta(L(T(\underline{X}), \theta)),$$

pour tout θ de Θ , sous réserve que cette espérance existe.

Quand la fonction de coût est quadratique on parle de **risque quadratique**.

Terminologie. Le risque quadratique est parfois appelé erreur quadratique moyenne ou MSE pour Mean Square Error en anglais.

On considérera essentiellement le risque quadratique dans la suite.

PROPOSITION 5.7. *Soit $T(\underline{X}) = (T_n(\underline{X}))_{n \in \mathbb{N}}$ un estimateur de $g(\theta) \in \mathbb{R}$, de carré intégrable pour la loi P_θ .*

Dans le cas d'un risque quadratique on a :

$$R(T(\underline{X}), \theta) = \text{Var}_\theta(T(\underline{X})) + b_T^2(\theta).$$

Pour un estimateur sans biais, le risque quadratique est donc égal à sa variance.

Preuve. Elle est aisée. Il suffit de développer

$$R(T(\underline{X}), \theta) = \mathbb{E}_\theta(T(\underline{X}) - g(\theta))^2 = \mathbb{E}_\theta(T(\underline{X}) - \mathbb{E}_\theta(T(\underline{X})) + \mathbb{E}_\theta(T(\underline{X})) - g(\theta))^2$$

qui donne la somme de la variance avec le carré du biais, le terme double produit s'annulant. \square

EXEMPLE 5.3. *Estimateurs de la moyenne empirique (suite des exemples 5.1 et 5.2).*

On a vu que, dans un modèle paramétrique où la variable générique X est telle que $\mathbb{E}_\theta(X) = g(\theta)$, l'estimateur de la moyenne empirique \bar{X}_n est un estimateur sans biais de $\mathbb{E}_\theta(X)$. Son risque quadratique, qui est donc égal à sa variance, est :

$$R(\bar{X}_n, \theta) = \text{Var}_\theta(\bar{X}_n) = \frac{\sigma_\theta^2}{n},$$

d'après le résultat vu dans la Proposition 4.2. \diamond

On peut maintenant comparer la précision de deux estimateurs.

DÉFINITION 5.8. *Soient $S(\underline{X})$ et $T(\underline{X})$ deux estimateurs de $g(\theta)$. On dit que $T(\underline{X})$ est **préférable** à $S(\underline{X})$ si l'on a :*

$$R(T(\underline{X}), \theta) \leq R(S(\underline{X}), \theta),$$

pour tout θ de Θ et avec une inégalité stricte pour au moins un θ de Θ .

La fonction de risque étant une fonction de θ , il n'est pas toujours possible de dire quel est l'estimateur préférable entre deux estimateurs donnés : leurs fonctions de risque peuvent se croiser. La notion précédente introduit donc seulement un ordre partiel.

Quand les estimateurs sont sans biais, l'estimateur $T(\underline{X})$ est préférable à $S(\underline{X})$ si sa variance est inférieure à celle de $S(\underline{X})$.

DÉFINITION 5.9. *Un estimateur $T(\underline{X})$ de $g(\theta)$ est dit **admissible** (resp. **\mathcal{T} -admissible**) s'il n'existe pas d'estimateur de $g(\theta)$ qui lui soit préférable (resp. dans une classe d'estimateurs \mathcal{T} de $g(\theta)$).*

Ainsi on peut, par exemple, s'intéresser aux estimateurs admissibles dans la classe des estimateurs sans biais de $g(\theta)$.

Par ailleurs, un estimateur non admissible n'a aucun intérêt. On préférera utiliser l'estimateur qui lui est préférable et qui est admissible.

DÉFINITION 5.10. *Un estimateur $T(\underline{X})$ de $g(\theta)$ est dit **optimal** (resp. **\mathcal{T} -optimal**) s'il est préférable à tous les estimateurs de $g(\theta)$ (resp. dans une classe d'estimateurs \mathcal{T} de $g(\theta)$).*

Sauf dans les cas triviaux, il n'existe pas d'estimateur optimal dans la classe de tous les estimateurs possibles. En revanche, si l'on restreint la classe des estimateurs que l'on considère (par exemple les estimateurs sans biais, linéaires etc...) on peut parfois trouver des estimateurs optimaux intéressants.

4. Information de Fisher

Nous venons de voir la notion de risque qui mesure la qualité d'un estimateur de $g(\theta)$. Elle mesure la qualité d'un estimateur. Naturellement se pose la question de l'existence d'une borne inférieure pour ce risque. Autrement dit, existe-t-il une fonction $B_{inf}(\theta)$, éventuellement dépendante de la taille n de l'échantillon observé, telle que l'on ait pour tout estimateur $T(X)$ de $g(\theta)$:

$$R(T(X), \theta) \geq B_{inf}(\theta),$$

pour tout θ de Θ ? On verra que, sous certaines conditions, une telle borne existe et est appelée borne de Cramer-Rao. Elle fait intervenir la notion d'information apportée par un modèle, appelée information de Fisher.

Pour simplifier les notations, on suppose dans ce paragraphe que le paramètre θ est dans \mathbb{R} . Les résultats resteront vrais avec les adaptations nécessaires au cas où θ est multidimensionnel (ces adaptations seront évoquées en fin de paragraphe). On note $L'(x; \theta)$ (resp. $L''(x; \theta)$) la dérivée première (resp. seconde) en θ de la fonction vraisemblance $L(x; \theta)$, pour la valeur x de l'observation dans le modèle paramétrique considéré.

Dans la suite on supposera que le modèle paramétrique $(E, \mathcal{E}, \{P_\theta : \theta \in \Theta\})$, de v.a. générique X , vérifie les hypothèses suivantes.

- H1** L'espace des paramètres Θ est un ouvert.
- H2** Les lois P_θ ont toutes même support, qui ne dépend donc pas de θ .
- H3** Les dérivées premières et secondes $L'(x; \theta)$ et $L''(x; \theta)$ de la vraisemblance existent pour tout x dans E .
- H4** Les fonctions $L'(x; \theta)$ et $L''(x; \theta)$, vues cette fois-ci comme fonction de la variable x (c'est à dire les densités), sont intégrables pour tout θ de Θ et on peut

toujours intervertir intégrale et dérivation :

$$\begin{aligned}\frac{\partial}{\partial\theta} \int_A L(x; \theta) dx &= \int_A L'(x; \theta) dx, \\ \frac{\partial^2}{\partial\theta^2} \int_A L(x; \theta) dx &= \int_A L''(x; \theta) dx,\end{aligned}$$

pour tout A dans \mathcal{E} .

Considérons la v.a.

$$S(X, \theta) = \frac{\partial}{\partial\theta} \ln L(X; \theta) = \frac{L'(X, \theta)}{L(X, \theta)},$$

qui, en tant que fonction de θ , est parfois appelée **fonction score**. Sous les hypothèses précédentes, cette v.a. est centrée. On a en effet :

$$\mathbb{E}_\theta(S(X, \theta)) = \int_E \frac{\partial}{\partial\theta} \ln L(x; \theta) L(x; \theta) dx = \int_E L'(x; \theta) dx = \frac{\partial}{\partial\theta} \int_E L(x; \theta) dx = 0,$$

puisque la dernière intégrale vaut 1 par définition d'une densité.

Posons maintenant l'hypothèse supplémentaire :

H5 la fonction score est de carré intégrable.

DÉFINITION 5.11. On appelle **information de Fisher** la variance du score, i.e.

$$I(\theta) = \text{Var}_\theta(S(X, \theta)) = \mathbb{E}_\theta \left(\left(\frac{\partial}{\partial\theta} \ln L(X; \theta) \right)^2 \right).$$

On peut établir une autre écriture de l'information de Fisher.

PROPOSITION 5.12. L'information de Fisher est aussi égale à

$$I(\theta) = -\mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} S(X, \theta) \right) = -\mathbb{E}_\theta \left(\frac{\partial^2}{\partial\theta^2} \ln L(X; \theta) \right).$$

Preuve. On remarque que l'on peut écrire :

$$\frac{\partial}{\partial\theta} S(x, \theta) = \frac{\partial}{\partial\theta} \frac{L'(x, \theta)}{L(x, \theta)} = \frac{L''(x, \theta)}{L(x, \theta)} - \frac{(L'(x, \theta))^2}{(L(x, \theta))^2} = \frac{L''(x, \theta)}{L(x, \theta)} - S(x, \theta)^2$$

Ainsi, on a :

$$\mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} S(X, \theta) \right) = \mathbb{E}_\theta \left(\frac{L''(X, \theta)}{L(X, \theta)} \right) - I(\theta).$$

En remarquant que

$$\mathbb{E}_\theta \left(\frac{L''(X, \theta)}{L(X, \theta)} \right) = \int_E L''(x, \theta) dx = \frac{\partial^2}{\partial\theta^2} \int_E L(x, \theta) dx = 0,$$

justifiée par l'hypothèse **H4**, on a bien le résultat annoncé. \diamond

EXEMPLE 5.4. *Information de Fisher dans le cas d'un modèle gaussien réel avec variance σ^2 connue.*

On considère le modèle :

$$\mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}\},$$

où σ^2 est supposé connu.

La log-vraisemblance pour l'observation x est :

$$\ln L(x; \mu) = -\ln(\sigma\sqrt{2\pi}) - \frac{(x - \mu)^2}{2\sigma^2}.$$

D'où :

$$S(x, \mu) = \frac{x - \mu}{\sigma^2} \text{ et } \frac{\partial}{\partial \mu} S(x, \theta) = -\frac{1}{\sigma^2}.$$

Cette dernière fonction étant constante en x , l'information de Fisher dans ce modèle est donc

$$I(\mu) = \frac{1}{\sigma^2}.$$

On trouve bien sûr le même résultat en écrivant

$$\mathbb{E}_\mu (S^2(X, \mu)) = \mathbb{E}_\mu \left(\left(\frac{X - \mu}{\sigma^2} \right)^2 \right) = \frac{\sigma^2}{\sigma^4}.$$

On remarque que l'information de Fisher est d'autant plus grande que σ^2 est petit. \diamond

Considérons maintenant l'information de Fisher dans un modèle d'échantillonnage, où l'on a donc observé un échantillon X_1, \dots, X_n de v.a. i.i.d. de même loi que X . On a alors

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n L(x_i; \theta) \text{ et } \ln L(x_1, \dots, x_n; \theta) = \sum_{i=1}^n \ln L(x_i; \theta).$$

En différenciant deux fois par rapport à θ , on obtient :

$$\frac{\partial}{\partial \theta} S(X_1, \dots, X_n; \theta) = \frac{\partial^2}{\partial \theta^2} \ln L(X_1, \dots, X_n; \theta) = \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \ln L(X_i; \theta) = \sum_{i=1}^n \frac{\partial}{\partial \theta} S(X_i; \theta),$$

ce qui prouve la proposition suivante.

PROPOSITION 5.13. *L'information de Fisher pour un modèle d'échantillonnage, i.e. pour l'échantillon X_1, \dots, X_n est n fois celle de la variable générique X de cet échantillon. C'est à dire que l'on a :*

$$I_n(\theta) = nI(\theta)$$

où $I_n(\theta)$ est l'information de Fisher de l'échantillon X_1, \dots, X_n et $I(\theta)$ celle de X .

EXEMPLE 5.5. *Information de Fisher pour un échantillon dans le cas d'un modèle gaussien réel avec variance σ^2 connue (suite de l'Exemple 5.4).*

D'après le résultat de l'Exemple 5.4 et la proposition précédente, l'information de Fisher pour un échantillon dans ce modèle est

$$I_n(\theta) = nI(\theta) = \frac{n}{\sigma^2}.$$

◇

Remarque. Dans le cas où le paramètre θ est multidimensionnel.

- Les fonctions $L'(x; \theta)$ et $L''(x; \theta)$ sont en fait le gradient et la matrice Hessienne de la fonction vraisemblance $L(x; \theta)$.
- Le score est un vecteur aléatoire, composition du gradient de la log-vraisemblance et de la v.a. générique X . On a $S(X, \theta) = \nabla_{\theta} \ln L(X, \theta)$.
- L'information de Fisher est une matrice et correspond à la matrice de covariance du score $S(X, \theta)$. C'est également l'opposé de l'espérance de la matrice Hessienne de la log-vraisemblance en X . On a

$$I(\theta) = \Sigma_{S(X, \theta)} = -\mathbb{E}_{\theta} \nabla^2 \ln L(X, \theta).$$

5. Borne de Cramer-Rao (ou Fréchet-Darmonis-Cramer-Rao)

THÉORÈME 5.14. *Soit un modèle paramétrique $(E, \mathcal{E}, \{P_{\theta}, \theta \in \Theta\})$, où Θ est dans \mathbb{R} , de v.a. générique X , et vérifiant les hypothèses **H1-H4** du paragraphe précédent (on peut ajouter l'hypothèse **H5** si l'on souhaite pouvoir utiliser l'autre expression de l'information de Fisher). Soit dans ce modèle un estimateur $T(X)$ sans biais et de carré intégrable de $g(\theta) \in \mathbb{R}$.*

Supposons que la fonction $x \mapsto T(x)L'(x; \theta)$ soit intégrable sur E et que l'on puisse intervertir dérivation et intégrale, i.e. :

$$\frac{\partial}{\partial \theta} \int_E T(x)L'(x; \theta) dx = \int_E T(x)L''(x; \theta) dx.$$

Supposons enfin que l'information de Fisher $I(\theta)$ soit strictement positive pour tout θ de Θ .

Alors la fonction g est dérivable et l'on a pour tout θ dans Θ :

$$\text{Var}(T(X)) \geq \frac{(g'(\theta))^2}{I(\theta)}.$$

*La borne $(g'(\theta))^2/I(\theta)$ est appelée **borne de Cramer-Rao** (ou Fréchet-Darmonis-Cramer-Rao).*

Remarque. L'hypothèse d'intégrabilité de la fonction $T(\cdot)L'(\cdot; \theta)$ et d'inversion possible entre dérivation et intégrale, est assurée dès qu'il existe une fonction h intégrable qui majore $T(\cdot)L'(\cdot; \theta)$, i.e. telle que :

$$|T(x)L'(x; \theta)| \leq h(x), \quad \forall x \in E \text{ et } \int_E h(x) dx < +\infty.$$

Preuve (du Théorème). L'estimateur $T(X)$ étant intégrable et sans biais, on a :

$$\mathbb{E}_\theta(T(X)) = \int_E T(x)L(x; \theta)dx = g(\theta).$$

D'après l'hypothèse du Théorème on peut dériver cette égalité et obtenir

$$g'(\theta) = \int_E T(x)L'(x; \theta)dx$$

et l'intégrale de droite est bien définie.

Maintenant, on peut écrire :

$$g'(\theta) = \int_E T(x)S(x, \theta)L(x; \theta)dx = \mathbb{E}_\theta(T(X)S(X, \theta)).$$

Ayant montré que $S(X, \theta)$ est une v.a. centrée, il vient :

$$g'(\theta) = \mathbb{E}_\theta(T(X)S(X, \theta)) - \mathbb{E}_\theta(T(X))\mathbb{E}_\theta(S(X, \theta)) = \text{Cov}_\theta(T(X), S(X, \theta)).$$

L'inégalité de Cauchy-Schwarz (ou Hölder) donne alors

$$(g'(\theta))^2 = (\text{Cov}_\theta(T(X), S(X, \theta)))^2 \leq \text{Var}_\theta(T(X))\text{Var}_\theta(S(X, \theta))$$

ce qui est équivalent à

$$\text{Var}_\theta(T(X)) \geq \frac{(g'(\theta))^2}{I(\theta)},$$

en se souvenant que la variance de $S(X, \theta)$ est par définition $I(\theta)$. □

Remarques.

- Si on considère un modèle d'échantillonnage de taille n , la borne de Cramer-Rao devient naturellement $(g'(\theta))^2/(I_n(\theta))$.
- La borne de Cramer-Rao ne dépend que du modèle paramétrique, de ce que l'on veut estimer ($g(\theta)$) et de la taille de l'échantillon. Plus la taille de l'échantillon augmente, plus cette borne diminue (ce qui est intuitif et heureux !).
- Rien ne dit que le minimum est atteint.
- Dans le cas où le paramètre θ est multidimensionnel, à valeurs dans $\Theta \subset \mathbb{R}^p$, la fonction g à valeurs dans \mathbb{R}^k et la matrice d'information de Fisher inversible, la borne de Cramer-Rao pour la matrice de covariance $\Sigma_{T(X)}$ est :

$$\Sigma_{T(X)} \geq \nabla'_\theta g(\theta)I^{-1}(\theta)\nabla_\theta g(\theta),$$

où $\nabla_\theta g(\theta)$ est le gradient en θ de la fonction g .

DÉFINITION 5.15. *Un estimateur sans biais atteignant la borne de Cramer-Rao est dit **efficace**.*

*Il est dit **asymptotiquement efficace** si*

$$\lim_{n \rightarrow +\infty} \frac{(g'(\theta))^2}{I_n(\theta)\text{Var}_\theta(T_n(X))} = 1,$$

où $T_n(X)$ est l'expression de l'estimateur pour une taille n d'échantillon.

Notons qu'un estimateur sans biais efficace est forcément optimal, dans la classe des estimateurs sans biais. La réciproque est fautive, la borne n'étant pas nécessairement atteinte.

EXEMPLE 5.6. *Information de Fisher pour un échantillon dans le cas d'un modèle gaussien réel avec variance σ^2 (suite des exemples 5.4 et 5.5).*

On a vu que la variance de l'estimateur de la moyenne empirique \bar{X}_n est de variance σ^2/n . Cette quantité est égale à la borne de Cramer-Rao dans un modèle gaussien réel avec la variance connue (seule l'espérance μ est inconnue). L'estimateur \bar{X}_n étant sans biais, il est donc efficace dans ce modèle pour estimer le paramètre μ . \diamond

6. Exercices

Exercice 1 (*Qualité des estimateurs dans les modèles de Poisson et de la loi exponentielle*)

On considère deux modèles :

- celui de la loi de Poisson $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathcal{P}(\lambda) : \lambda > 0\})$, où $\mathcal{P}(\lambda)$ désigne la loi de Poisson de paramètre λ ;
- celui de la loi de exponentielle $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}(\lambda) : \lambda > 0\})$, où $\mathcal{E}(\lambda)$ désigne la loi exponentielle de paramètre λ .

On a vu que ces modèles sont en particulier utiles pour modéliser des problèmes de Fiabilité.

Pour chacun de ces modèles, répondre à l'ensemble des questions suivantes. On considérera à chaque fois l'observation d'un échantillon X_1, \dots, X_n .

- 1) Rappeler l'expression de l'estimateur du maximum de vraisemblance dans ce modèle (on a vu qu'il est également estimateur par la méthode des moments).
- 2) Étudier la consistance, le biais et le risque quadratique de cet estimateur.
- 3) Si cet estimateur est biaisé, est-il asymptotiquement sans biais ? Donner, si nécessaire, un estimateur sans biais. L'estimateur sans biais (l'initial ou le second) est-il efficace ? Sinon l'est-il asymptotiquement ? Est-il consistant ?

Exercice 2 (*Fiabilité et fonction de répartition empirique*)

Un matériel a une durée de vie modélisée par une v.a. X de f.d.r. F . Un étudiant en Licence de Mathématiques sait qu'il devra l'utiliser pendant un temps x_0 . Il souhaite naturellement qu'il n'y ait pas de panne durant cette période.

Cet étudiant, ayant suivi le module de Statistique Inférentielle, cherche en premier lieu à estimer la loi (en fait la f.d.r.) de cette durée de vie, c'est à dire estimer $F(x)$ pour tout x de \mathbb{R}^+ . Il a alors l'idée de faire fonctionner, sur banc d'essai, n machines identiques à celle qu'il utilisera dans l'avenir. Il note x_1, \dots, x_n les n temps de panne observés, qui sont donc les réalisations des v.a. X_1, \dots, X_n i.i.d. de même loi que X .

- 1) Par la méthode des moments il propose un estimateur de $F(x)$, pour tout x dans \mathbb{R}^+ . Pouvez-vous en faire autant ?

2) Son estimateur est-il consistant ? Que dire de son biais et de son risque quadratique ?

3) Se souvenant de ses cours, il sait que, pour être précis, il aurait dû, au préalable, introduire un modèle paramétrique. Quel(s) modèle(s) pourrait-il proposer ? Que sont les observations sous ce(s) modèle(s) ? Une estimation par maximum de vraisemblance nous donnerait-elle quelque chose de différent dans ce modèle ?

4) Que dire alors de l'efficacité de l'estimateur proposé dans la première question ?

Exercice 3 (*L'Agriculteur et la Statistique*)

Un agriculteur possède un champ carré dont il veut estimer la superficie. Quand il mesure un côté de son champ, il sait (un statisticien de passage lui a confirmé), ou il suppose, que l'erreur expérimentale de la mesure est une variable aléatoire de loi normale centrée et de variance σ^2 . Il réalise une première mesure de ce côté et trouve une valeur $x_1 = 510$ mètres. Il en déduit une superficie de $s_1 = 26.01$ hectares. Il réalise une deuxième mesure et trouve alors $x_2 = 490$, d'où une valeur de la superficie $s_2 = 24.01$. Il abandonne ses mesures et réfléchit pour savoir quelle est la bonne façon de procéder. Doit-il adopter comme estimation de la surface s_1 , s_2 , ou une estimation combinant les deux mesures, telle que :

$$\begin{aligned} s_3 &= x_1 x_2 = 24.99, \\ s_4 &= \frac{s_1 + s_2}{2} = 25.01, \\ s_5 &= \left(\frac{x_1 + x_2}{2} \right)^2 = 25 ? \end{aligned}$$

Faut-il recommencer ses mesures jusqu'à ce qu'il trouve deux résultats identiques, ou bien combiner intelligemment n mesures pour construire des estimations du type s_4 ou s_5 (généralisées à ces n mesures) ?

1) On se propose d'aider l'agriculteur à résoudre son problème. Préciser le modèle considéré ainsi que la fonction $q(\theta)$ que l'on cherche à estimer. Étudier les cinq estimateurs proposés. On calculera notamment leurs biais, variances et risques quadratiques moyens. (Ind. si $X \sim \mathcal{N}(m, \sigma^2)$ alors $\text{Var}(X^2) = 2(\sigma^4 + 2m^2\sigma^2)$).

A l'aide de ces calculs, aider l'agriculteur à choisir l'estimateur qui vous semble préférable aux autres.

2) Donner les estimateurs qui généralisent s_4 et s_5 au cas où l'agriculteur a pu faire n mesures du côté de son champ. Effectuer la même étude qu'à la question 1) pour ces estimateurs. Étudier également leurs consistances. Que dire de leur L^2 -consistance ? Conclure.

3) Donner l'estimateur du maximum de vraisemblance et l'étudier s'il est différent de ceux considérés précédemment.

Exercice 4 (*Comparaison d'estimateurs dans un modèle uniforme*)

On considère le modèle uniforme $\{\mathcal{U}_{[0,\theta]} : \theta > 0\}$. On considère un échantillon X_1, \dots, X_n et on note $X_{(1)}$ et $X_{(n)}$ respectivement la première et la dernière statistique d'ordre.

On a vu en cours que l'on pouvait proposer les estimateurs suivants pour le paramètre θ .

$$\begin{aligned}\hat{\theta}_1 &= X_{(n)} \\ \hat{\theta}_2 &= \frac{n+1}{n} X_{(n)} \\ \hat{\theta}_3 &= X_{(1)} + X_{(n)} \\ \hat{\theta}_4 &= 2\bar{X}_n,\end{aligned}$$

où \bar{X}_n est l'estimateur de la moyenne empirique.

1) Rappeler brièvement l'idée à la base de la proposition de chacun de ces estimateurs.

2) Pour chacun d'entre eux, étudier la consistance, le biais et donner l'expression de son risque quadratique.

3) Comparer les fonctions de risque quadratique. Qu'en conclure ?

Exercice 5 (*Optimalité pour les estimateurs linéaires de l'espérance mathématique*)

Soit $\mathcal{P} = (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{P_{\theta} : \theta \in \Theta\})$, de v.a. générique X , un modèle paramétrique tel que l'on ait $\mathbb{E}_{\theta}(X) = g(\theta)$. Pour simplifier les notations on notera μ_{θ} cette espérance et σ_{θ}^2 la variance de X . On a vu que \bar{X}_n est un estimateur sans biais de μ_{θ} de risque quadratique $R(\bar{X}_n, \theta) = \sigma_{\theta}^2/n$. Cet estimateur s'exprime bien évidemment comme une combinaison linéaire des v.a.r. de l'échantillon X_1, \dots, X_n . On dit qu'il est linéaire.

1) Montrer qu'un estimateur linéaire et sans biais est forcément une combinaison linéaire des v.a. X_1, \dots, X_n ayant pour somme des coefficients 1. Calculer le risque quadratique d'un tel estimateur.

2) En utilisant l'inégalité de Cauchy-Schwarz pour la somme de réels, montrer que la moyenne empirique est un estimateur optimal dans la classe des estimateur linéaires et sans biais de μ_{θ} .

3) On considère maintenant la classe des estimateurs linéaires, mais pas nécessairement sans biais. Ces estimateur sont de la forme

$$S(\underline{X}) = \sum_{i=1}^n c_i X_i,$$

où les c_1, \dots, c_n sont des réels donnés. Calculer le risque d'un estimateur dans cette classe (en fonction des c_1, \dots, c_n). On cherche l'estimateur optimal dans cette classe. En admettant que la fonction à minimiser est convexe, montrer que le minimum est atteint pour les c_i tous égaux à $\mu_{\theta}^2/(\sigma_{\theta}^2 + n\mu_{\theta}^2)$. En déduire qu'il n'existe pas d'estimateur optimal dans cette classe.

CHAPITRE 6

Amélioration d'estimateurs

On a signalé, à maintes reprises, qu'une statistique constitue souvent une réduction de l'information de manière à ne retenir que ce qui nous paraissait utile et nécessaire dans l'estimation du paramètre θ du modèle (ou d'une fonction $g(\theta)$ de celui-ci). De plus, nous avons introduit quelques critères pour étudier et comparer la qualité des estimateurs.

Des questions se posent alors naturellement.

- Comment s'assurer que nous n'avons pas perdu une partie importante de l'information apportée dans l'échantillon dans notre problème d'estimation ? Nous n'avons en effet pas évalué la qualité de cette réduction d'information.
- Comment améliorer un estimateur obtenu par les méthodes précédentes de manière à ce qu'il ne prenne en compte que le strict nécessaire dans l'information apportée par l'échantillon ?
- Existe-t-il une(des) méthode(s) pour construire un estimateur optimal ?

Telles sont les questions auxquelles nous allons apporter quelques réponses dans les paragraphes qui suivent.

1. Statistique exhaustive

Dans la plupart des exemples considérés précédemment, la statistique utilisée était une v.a. $T(X_1, \dots, X_n)$ avec T fonction de E^n à valeurs dans \mathbb{R} ou \mathbb{R}^2 . On a ainsi réduit l'information apportée par l'échantillon sous forme d'un vecteur (x_1, \dots, x_n) en un scalaire ou un vecteur de dimension bien inférieure. On a donc naturellement perdu de l'information. Le problème est de savoir si cette information perdue aurait pu ou pas être utile à l'estimation du paramètre $g(\theta)$. Si l'information perdue s'avère inutile dans l'optique de l'estimation de $g(\theta)$, on dira alors que la statistique est exhaustive, qu'elle préserve donc toute l'information exploitable pour résoudre le problème posé.

EXEMPLE 6.1. *Exhaustivité dans un modèle de Bernoulli pour le contrôle statistique de la qualité.*

Un industriel voudrait connaître la proportion p de pièces défectueuses qu'il fabrique dans une journée. Pour cela il prélève n pièces aléatoirement dans l'ensemble de sa production de la journée. Il suppose que la qualité de sa production n'a pas évolué au cours de la journée, autrement dit que cette proportion n'a pas varié. Il note alors le nombre k de pièces défectueuses observées dans cet échantillon et estime la qualité de sa production par k/n . Il néglige donc toute une partie de l'information apportée

par l'échantillon de son contrôle, comme celle de savoir quelles pièces se sont révélées défectueuses. Est-ce handicapant pour la qualité de son estimation ? Intuitivement on sent bien que non. Montrons le rigoureusement...

En premier lieu, notons que le modèle statistique pour une telle expérience est le modèle d'échantillonnage associé au modèle de la loi de Bernoulli : $(\{0, 1\}, \mathcal{P}(\{0, 1\}), \{\mathcal{B}(p) : p \in [0, 1]\})$ de v.a. générique X . De l'observation $\underline{x} = (x_1, \dots, x_n)$ d'un échantillon $\underline{X} = (X_1, \dots, X_n)$ dans ce modèle, l'industriel a donc retenu seulement l'information apportée par la statistique

$$T(\underline{X}) = \sum_{i=1}^n X_i,$$

qui vaut k dans cet exemple.

Or calculons la loi de \underline{X} conditionnellement à $T(\underline{X}) = k$. On a :

$$P(\underline{X} = \underline{x} | T(\underline{X}) = k) = \begin{cases} 0 & \text{si } \sum_{i=1}^n x_i \neq k \\ \frac{p^k(1-p)^{n-k}}{C_n^k p^k(1-p)^{n-k}} = \frac{1}{C_n^k} & \text{si } \sum_{i=1}^n x_i = k. \end{cases}$$

On constate que cette loi ne dépend pas du paramètre p que l'on cherche à estimer. Il s'agit de la loi uniforme sur $\{\underline{x} \in \{0, 1\}^n : \sum_{i=1}^n x_i = k\}$. Ainsi, toute l'information sur le paramètre p contenue dans l'échantillon \underline{X} est en fait contenue dans la statistique $T(\underline{X})$. On dit que cette statistique est exhaustive. \diamond

DÉFINITION 6.1. Soit $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ un modèle paramétrique et $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans ce modèle. Une statistique $T(\underline{X})$ est dite **exhaustive** pour le paramètre θ si la loi de $\underline{X} = (X_1, \dots, X_n)$ conditionnelle à $T(\underline{X})$ est indépendante du paramètre θ .

On remarque que, dans le cas d'une statistique exhaustive, la loi de l'échantillon conditionnelle à T est entièrement connue. Ainsi, une fois que l'on a calculé $T(\underline{x})$ sur l'échantillon observé \underline{x} , ce dernier ne nous apporte plus aucune information sur le paramètre θ et peut être oublié.

Le calcul de la loi conditionnelle n'étant pas toujours aussi facile que dans l'exemple précédent, on utilisera souvent le théorème suivant qui donne un moyen plus aisé pour prouver l'exhaustivité d'une statistique.

THÉORÈME 6.2. (Théorème de factorisation) Soit $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ un modèle paramétrique et $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans ce modèle. Une statistique $T(\underline{X})$ à valeurs dans un espace probabilisable (F, \mathcal{F}) est exhaustive si, et seulement si, il existe une fonction mesurable h de E^n vers \mathbb{R}^+ , ne dépendant pas de θ , et une fonction mesurable g_θ de F vers \mathbb{R}^+ telles que la vraisemblance puisse s'écrire sous la forme :

$$L(x_1, \dots, x_n; \theta) = g_\theta(T(\underline{x}))h(\underline{x}),$$

pour tout $\underline{x} = (x_1, \dots, x_n)$ dans E^n et tout θ dans Θ .

La démonstration de ce théorème est admise.

EXEMPLE 6.2. *Statistiques exhaustives dans le modèle de Bernoulli (Suite de l'Exemple 6.1.)*

Nous avons vu, au début du paragraphe 4, que la vraisemblance d'un échantillon X_1, \dots, X_n dans un tel modèle est :

$$L(x_1, \dots, x_n; p) = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}.$$

On peut écrire :

$$L(x_1, \dots, x_n; p) = g_p(T(\underline{x}))h(\underline{x}),$$

avec $g_p(x) = p^x(1-p)^{n-x}$ et h égale à 1. Grâce au théorème de factorisation on retrouve que la Statistique

$$T(\underline{X}) = \sum_{i=1}^n X_i$$

est bien exhaustive pour le paramètre p dans ce modèle. On constate aisément que la moyenne empirique

$$\bar{X}_n = \frac{T(\underline{X})}{n}$$

est également une statistique exhaustive dans ce modèle. De nombreuses autres statistiques sont exhaustives comme en particulier les statistiques

$$\begin{aligned} T_1(\underline{X}) &= (X_1, \dots, X_n), \\ T_2(\underline{X}) &= (X_1 + X_2, X_3 + \dots + X_n), \\ T_3(\underline{X}) &= (X_1, X_2, X_3 + \dots + X_n). \end{aligned}$$

Mais l'on voit bien qu'elles réduisent moins l'information que les deux statistiques précédentes $T(\underline{X})$ et \bar{X}_n . Elles seront donc moins intéressantes. \diamond

PROPOSITION 6.3. *S'il existe une statistique exhaustive, l'estimateur du maximum de vraisemblance en est une fonction.*

Preuve. Notons $T(\underline{X})$ cette statistique exhaustive. D'après le théorème de factorisation la vraisemblance peut s'écrire sous la forme :

$$L(\underline{x}; \theta) = g_\theta(T(\underline{x}))h(\underline{x}).$$

On a donc

$$\begin{aligned} \hat{\theta} &= \underset{\theta}{\text{Argmax}} L(\underline{x}; \theta) \\ &= \underset{\theta}{\text{Argmax}} g_\theta(T(\underline{x})) \end{aligned}$$

et $\hat{\theta}$ est donc bien une fonction de $T(\underline{x})$. \square

2. Statistique exhaustive minimale

Comme on a pu le constater dans l'exemple précédent, il n'y a pas unicité de la statistique exhaustive. D'une manière générale, toute fonction mesurable bijective d'une statistique exhaustive est exhaustive, i.e. si $T(\underline{X})$ est une statistique exhaustive et si h est une fonction mesurable bijective de (F, \mathcal{F}) vers (F', \mathcal{F}') alors $S(\underline{X}) = h(T(\underline{X}))$ est aussi exhaustive. C'est évident d'après le théorème de factorisation.

La statistique $T(\underline{X}) = \underline{X}$ est toujours une statistique exhaustive. Mais elle n'est pas d'un grand intérêt et ne réduit absolument pas l'information. Il ne s'agit donc pas seulement de trouver une statistique exhaustive mais plutôt de trouver parmi les statistiques exhaustives celle(s) qui réduit(ent) au maximum l'information.

En d'autres termes, le problème est de trouver une statistique exhaustive qui soit minimale, c'est à dire qui ait supprimé le maximum d'information ne concernant pas θ tout en préservant toute l'information sur θ . D'après ce que l'on vient de voir, des transformations bijectives de statistiques exhaustives ne changeront rien à ce problème de réduction maximale de l'information. On dit que ces statistiques sont équivalentes¹. On cherchera davantage des transformations non bijectives, qui réduisent en particulier la dimension de la statistique.

DÉFINITION 6.4. *On dit qu'une **statistique exhaustive est minimale**², si elle est une fonction mesurable de toutes les autres statistiques exhaustives.*

Autrement dit, la statistique T est minimale si pour toute statistique exhaustive S il existe une fonction h telle que $T = h(S)$. S'il existe une fonction h' telle que $S = h'(T)$ soit exhaustive mais avec h' non bijective (c'est à dire que l'on ne puisse pas écrire T comme une fonction de S), alors T n'est pas une fonction de toutes les statistiques exhaustives et n'est donc pas minimale.

Une statistique exhaustive minimale réduit donc au maximum l'information apportée par l'échantillon. En effet, soit T une telle statistique. Si on pouvait réduire davantage l'information tout en préservant l'exhaustivité, il existerait une fonction h mesurable et non bijective (sinon ça ne changerait rien) telle que la statistique $S = h(T)$ soit encore une statistique exhaustive. Ce qui est en contradiction avec le fait que T soit minimale, puisqu'elle n'est pas fonction de la statistique S qui est pourtant exhaustive.

3. Théorème de Rao-Blackwell

Nous avons vu dans les parties précédentes comment construire des estimateurs de $g(\theta)$ dans un modèle paramétrique $\{P_\theta : \theta \in \Theta\}$. Nous savons également étudier

¹On dira que deux statistiques sont équivalentes s'il existe une relation bijective mesurable entre les deux. Si deux statistiques S et T sont équivalentes, alors la première est exhaustive si, et seulement si, la seconde l'est. On pourra donc raisonner sur les classes d'équivalence des statistiques exhaustives, c'est à dire sur les ensembles des statistiques équivalentes. Aussi, quand on parlera de statistique exhaustive, cela pourra être à une bijection près

²La minimalité est rigoureusement définie sur l'ensemble des classes d'équivalence avec l'ordre \prec défini, pour deux statistiques non équivalentes S et T , par $S \prec T$ s'il existe une fonction mesurable non bijective telle que $S = h(T)$

leur qualité. Nous avons vu empiriquement (en cours et en TD) comment déduire d'un estimateur biaisé un estimateur sans biais. Mais nous n'avons pas encore vu de méthode permettant d'améliorer un estimateur en diminuant son risque.

C'est l'intérêt du théorème de Rao-Blackwell qui est basé sur l'utilisation d'une statistique exhaustive pour le paramètre θ .

THÉORÈME 6.5. (Théorème de Rao-Blackwell) Soit $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ un modèle paramétrique et $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans ce modèle. Soit $T(\underline{X})$ un estimateur de $g(\theta)$ de carré intégrable.

Si le modèle possède une statistique exhaustive $S(\underline{X})$ pour le paramètre θ , alors l'estimateur

$$\tilde{T}(\underline{X}) = \mathbb{E}_\theta(T(\underline{X})|S(\underline{X}))$$

de $g(\theta)$ a un risque quadratique inférieur à $T(\underline{X})$, c'est à dire que l'on a :

$$R(\tilde{T}(\underline{X}), \theta) \leq R(T(\underline{X}), \theta),$$

pour tout θ dans Θ .

De plus cette inégalité est stricte pour au moins un θ de Θ , i.e. $\tilde{T}(\underline{X})$ est préférable à $T(\underline{X})$, sauf si $T(\underline{X})$ est sans biais et une fonction de la statistique exhaustive $S(\underline{X})$.

Si $T(\underline{X})$ est un estimateur sans biais de $g(\theta)$ alors $\tilde{T}(\underline{X})$ est également sans biais pour $g(\theta)$ et l'inégalité sur les risques quadratiques se traduit également sur les variances.

Rappel.

Par manque de maturité mathématique, en particulier en théorie de la mesure, on ne peut ici définir proprement la notion d'espérance conditionnelle. En conséquence, on se contentera de la définition un peu approximative suivante de $\mathbb{E}(Y/X)$, suivant que les v.a. X et Y sont discrètes ou continues.

* Soient X et Y deux v.a. discrètes dont les lois P_X et P_Y sont concentrées respectivement sur I et J . Pour tout x_i dans I , on définit :

$$\begin{aligned} \mathbb{E}(Y/X = x_i) &= \sum_{y_j \in J} y_j P_Y^{X=x_i}(y_j) \\ &= \sum_{y_j \in J} y_j P(Y = y_j/X = x_i). \end{aligned}$$

* Soient X et Y deux v.a. de loi conjointe absolument continue et notons $f_Y^{X=x}$ la densité conditionnelle de Y conditionnellement à $X = x$. On définit

$$\mathbb{E}(Y/X = x) = \int_{\mathbb{R}} y f_Y^{X=x}(y) dy.$$

Dans les deux cas, la fonction $e : x \rightarrow \mathbb{E}(Y/X = x)$ est une fonction réelle d'une variable réelle. On peut montrer qu'elle est mesurable et on peut considérer sa composition avec la variable aléatoire X , i.e. considérer $e \circ X$. Celle-ci définit une variable aléatoire réelle que l'on appelle espérance conditionnelle de Y sachant X , notée $\mathbb{E}(Y/X)$.

Ce théorème est admis bien que non difficile à démontrer, au moins quand on domine assez bien l'outil espérance conditionnelle. Remarquons cependant que la statistique obtenue par ce conditionnement est bien un estimateur puisque la loi de \underline{X} (et donc de $T(\underline{X})$) conditionnelle à $S(\underline{X})$ ne dépend pas de θ , par définition d'une statistique exhaustive.

On améliore donc un estimateur en prenant son espérance conditionnelle par rapport à une statistique exhaustive. Le nouvel estimateur obtenu est alors une fonction de la statistique exhaustive (propriété de l'espérance conditionnelle). Un bon estimateur doit ainsi être fonction de toutes les statistiques exhaustives du paramètre θ du modèle et donc, si elle existe, de la statistique exhaustive minimale.

4. Théorème de Lehmann-Scheffé

Le théorème précédent nous permet déjà d'améliorer la qualité d'un estimateur. Mais il ne nous assure pas de tomber sur un estimateur optimal. L'obtention directe d'un estimateur optimal sera possible grâce au Théorème de Lehmann-Scheffé donné ci-dessous. Mais il nous faut auparavant introduire la notion de statistique complète qu'il utilise.

DÉFINITION 6.6. Soit $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ un modèle paramétrique et $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans ce modèle. Une statistique $T(\underline{X})$ est dite **complète** (ou totale) si toute fonction borélienne f vérifiant $\mathbb{E}_\theta |f(T(\underline{X}))| < +\infty$ et $\mathbb{E}_\theta(f(T(\underline{X}))) = 0$ pour tout θ de Θ est nécessairement telle que

$$f(T(\underline{X})) = 0, \quad P_\theta - p.s.$$

pour tout θ de Θ .

THÉORÈME 6.7. Toute statistique exhaustive et complète est minimale.

THÉORÈME 6.8. (Théorème de Lehmann-Scheffé) Soit $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ un modèle paramétrique et $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans ce modèle. Soit $T(\underline{X})$ un estimateur sans biais de $g(\theta)$ de carré intégrable et $S(\underline{X})$ une statistique exhaustive et complète de θ .

Alors l'estimateur amélioré de Rao-Blackwell $\tilde{T}(\underline{X}) = \mathbb{E}_\theta(T(\underline{X})|S(\underline{X}))$ est optimal dans la classe des estimateurs sans biais de $g(\theta)$.

5. Cas des familles exponentielles

La structure de famille exponentielle a l'avantage de nous assurer de l'existence d'une statistique exhaustive et complète comme l'établit la proposition suivante.

PROPOSITION 6.9. *La statistique canonique $T(X)$ dans une famille exponentielle générale est une statistique exhaustive et complète, donc minimale. Pour le modèle d'échantillonnage associé la statistique canonique est $\sum_{i=1}^n T(X_i)$.*

La preuve sera (en partie) effectuée dans un exercice.

EXEMPLE 6.3. *Modèle gaussien réel avec variance σ^2 connue.*

On a vu en Feuille de T.D. n°1 que le modèle $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu, \sigma^2) : \mu \in \mathbb{R}\})$ constitue une famille exponentielle naturelle dont la statistique canonique dans le modèle d'échantillonnage associé est $T(\underline{X}) = \sum_{i=1}^n X_i$. On sait également que \bar{X}_n est un estimateur sans biais de μ . La proposition précédente et le théorème de Lehmann-Scheffé assurent que \bar{X}_n est optimal dans la classe des estimateurs sans biais de μ . \diamond

6. Exercices

Exercice 1 (Statistiques exhaustives)

On considère les modèles déjà largement étudiés dans les feuilles d'exercices précédentes :

- modèle de Poisson $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathcal{P}(\lambda) : \lambda > 0\})$;
- modèle de la loi de exponentielle $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}(\lambda) : \lambda > 0\})$.
- modèle gaussien avec σ^2 positif connu : $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu, \sigma^2) : \mu \in \mathbb{R}\})$;
- modèle gaussien avec μ dans \mathbb{R} connu : $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu, \sigma^2) : \sigma^2 > 0\})$;
- modèle gaussien général : $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\})$.

1) Pour chacun de ces modèles donner l'expression d'une statistique exhaustive (éventuellement vectorielle).

2) Retrouver le résultat pour le modèle de Poisson en utilisant une autre méthode.

Exercice 2 (Statistique exhaustive et Famille Exponentielle Générale)

On considère une famille exponentielle générale de statistique canonique $T(X)$ où X est la variable générique dans ce modèle.

1) Montrer que $\sum_{i=1}^n T(X_i)$ est une statistique exhaustive pour le modèle d'échantillonnage associé.

2) En utilisant un résultat obtenu dans l'Exercice 1 du chapitre 2, montrer que la moyenne empirique \bar{X}_n est une statistique exhaustive dans un modèle d'échantillonnage de la loi Binomiale.

Exercice 3 (Estimation optimale dans le modèle de Poisson)

Il est courant de constater que le nombre d'appels reçus en une heure par un standard téléphonique suit une loi de Poisson. On s'intéresse au problème de l'estimation de la probabilité qu'il n'y ait pas d'appel en une heure. Pour cela, on considère le modèle statistique $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathcal{P}(\lambda) : \lambda > 0\})$, de v.a. générique X . On note $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans un modèle et on cherche donc à estimer $g(\lambda) = P_\lambda(X = 0) = \exp(-\lambda)$.

- 1) Proposer un estimateur $W(\underline{X})$ de $g(\lambda)$ fonction des v.a. $\mathbb{1}_{\{X_i=0\}}$, pour $i = 1, \dots, n$.
- 2) Donner son biais, son risque quadratique et sa loi.
- 3) L'estimateur proposé $W(\underline{X})$ est-il fonction de la statistique exhaustive :

$$T(\underline{X}) = \sum_{i=1}^n X_i?$$

Sinon, que proposeriez-vous pour améliorer l'estimation ?

- 4) Calculer la loi de chaque X_i , pour $i = 1, \dots, n$, conditionnelle à $\{T(\underline{X}) = t\}$.
- 5) On note

$$Y_i = \mathbb{1}_{\{X_i=0\}}.$$

Calculer l'espérance conditionnelle $\mathbb{E}_\lambda(Y_i|T(\underline{X}) = t)$ et $\mathbb{E}_\lambda(Y_i|T(\underline{X}))$.

6) En déduire l'expression, en fonction de T , de l'estimateur $W^*(\underline{X})$, amélioration de l'estimateur $W(\underline{X})$ par le théorème de Rao-Blackwell. Que dire du biais de $W^*(\underline{X})$?

7) Calculer $\mathbb{E}_\lambda(z^{T(\underline{X})})$ puis $\text{Var}_\lambda(z^{T(\underline{X})})$. En déduire le risque quadratique de l'estimateur $W^*(\underline{X})$.

- 8) Montrer que la statistique $T(\underline{X})$ est également complète. Conclure.

Exercice 4 (*Estimation optimale dans le modèle uniforme*)

On considère le modèle de la loi uniforme $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{U}_{[0,\theta]} : \theta > 0\})$ et un échantillon X_1, \dots, X_n dans ce modèle. On se propose d'améliorer, si possible, l'estimateur $\hat{\theta}(\underline{X}) = X_{(1)} + X_{(n)}$ vu en cours et dans l'exercice 4 du chapitre 5 .

- 1) Donner une statistique exhaustive dans ce modèle pour le paramètre θ .
- 2) Calculer la densité de la loi de $X_{(1)}$ conditionnelle à $\{X_{(n)} = x_n\}$. En déduire l'expression de $\mathbb{E}_\theta(X_{(1)}|X_{(n)} = x_n)$ puis de $\mathbb{E}_\theta(X_{(1)}|X_{(n)})$.
- 3) Déterminer alors $\tilde{\theta}(\underline{X})$, estimateur amélioré de $\hat{\theta}(\underline{X})$ par le théorème de Rao-Blackwell.
- 4) La statistique $X_{(n)}$ est-elle complète ? Conclure.

Comportement asymptotique d'un estimateur

On s'intéresse ici au comportement des estimateurs quand la taille n de l'échantillon X_1, \dots, X_n augmente, c'est à dire quand on augmente l'information que l'on a sur le modèle paramétrique.

On a déjà abordé ce sujet quand nous avons introduit la notion de consistance d'un estimateur qui est une qualité appréciable pour un estimateur.

Il est courant que l'on cherche également à prouver l'asymptotique normalité d'un estimateur. Ceci pour au moins deux "bonnes" raisons : d'une part c'est souvent le cas et, d'autre part, c'est très utile, en particulier pour la construction d'intervalle de confiance, d'autant plus quand la loi de l'estimateur pour une taille finie d'échantillon n'est pas connue.

1. Normalité asymptotique

DÉFINITION 7.1. *Soient :*

- $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ un modèle statistique paramétrique ;
- pour chaque n dans \mathbb{N} , un échantillon $\underline{X} = (X_1, \dots, X_n)$ dans ce modèle ;
- et $T(\underline{X}) = (T_n(\underline{X}))_{n \in \mathbb{N}}$ un estimateur de $g(\theta)$ dans \mathbb{R}^p .

On dit que l'estimateur $T(\underline{X})$ est **asymptotiquement gaussien** si l'on a la convergence en loi

$$\sqrt{n}(T_n(\underline{X}) - g(\theta)) \xrightarrow{\mathcal{L}} N_d(0, \Sigma(\theta)),$$

quand n tend vers $+\infty$. La matrice de covariance $\Sigma(\theta)$ est appelée matrice de covariance asymptotique.

Quand $g(\theta)$ est un réel, on dit que l'estimateur est **asymptotiquement normal** et la convergence en loi s'écrit :

$$\sqrt{n}(T_n(\underline{X}) - g(\theta)) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2),$$

quand n tend vers $+\infty$. La variance σ_θ^2 est appelée variance asymptotique.

Remarquons qu'un estimateur de $g(\theta)$ asymptotiquement gaussien est forcément consistant. On a en effet, quand n tend vers $+\infty$:

$$T_n(\underline{X}) - g(\theta) = \frac{1}{\sqrt{n}} \sqrt{n}(T_n(\underline{X}) - g(\theta)) \xrightarrow{P} 0,$$

grâce au théorème de Slutsky.

2. Estimateurs empiriques des moments

En reprenant les notations de la Section 1, on a les résultats suivants, bien connus depuis le cours de Probabilités. Ils sont fondamentaux en Statistique.

THÉORÈME 7.2. (*Loi forte des grands nombres et Théorème de la limite centrale*) *Si la variable générique du modèle est dans L^2 , on a, quand $n \rightarrow +\infty$:*

$$\begin{aligned}\bar{X}_n &\xrightarrow{p.s.} m_\theta, \\ \sqrt{n} \frac{\bar{X}_n - m_\theta}{\sigma_\theta} &\xrightarrow{\mathcal{L}} N(0, 1).\end{aligned}$$

On a le même type de résultats pour les estimateurs empiriques des moments. Il s'obtiennent également par la loi forte des grands nombres et le théorème de la limite centrale. On a, quand $n \rightarrow +\infty$:

$$\begin{aligned}\hat{m}_\theta(p) &\xrightarrow{p.s.} m_\theta(p), \\ \sqrt{n} \frac{\hat{m}_\theta(p) - m_\theta(p)}{\sqrt{\text{Var}_\theta(X^p)}} &\xrightarrow{\mathcal{L}} N(0, 1).\end{aligned}$$

Rappelons également la version multidimensionnelle du théorème de la limite centrale. Notons pour cela $z_{n,1}, \dots, z_{n,p}$ les p -coordonnées d'un vecteur z_n de \mathbb{R}^p et \bar{z}_n le vecteur des moyennes des composantes des n premiers vecteurs de la suite (z_n) , i.e.

$$\bar{z}_n = \begin{pmatrix} \frac{1}{n} \sum_{j=1}^n z_{j,1} \\ \vdots \\ \frac{1}{n} \sum_{j=1}^n z_{j,p} \end{pmatrix}.$$

THÉORÈME 7.3. (*Théorème de la limite centrale multidimensionnel*)

Soit (Z_n) une suite de vecteurs aléatoires dans $(\mathbb{R}^p, B_{\mathbb{R}^p})$, indépendants, de même loi d'espérance μ et de matrice de covariance Σ . On a alors :

$$\sqrt{n}(\bar{Z}_n - \mu) \xrightarrow{\mathcal{L}} N_p(0, \Sigma),$$

quand $n \rightarrow +\infty$.

3. Estimateur du maximum de vraisemblance

À plusieurs reprises, nous avons pu noter que l'estimateur du maximum de vraisemblance ne possède pas nécessairement de bonnes propriétés à taille finie. Il peut être biaisé et non admissible. Tel fut le cas par exemple pour l'estimateur du maximum de vraisemblance $X_{(n)}$ du paramètre θ dans un modèle uniforme $\{\mathcal{U}_{[0,\theta]} : \theta > 0\}$.

C'est en fait en raison de ses propriétés asymptotiques, détaillées dans le théorème suivant (que nous ne démontrerons pas), que cet estimateur du maximum de vraisemblance est très célèbre et souvent utilisé.

THÉORÈME 7.4. Soit $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}^p\})$ un modèle statistique paramétrique (identifiable) tel que les hypothèses du paragraphe **H1-H5** de la section 4 du Chapitre 5 soient vérifiées. On note $I(\theta)$ l'information de Fisher pour la v.a. générique X dans ce modèle et $\hat{\theta}_n(\underline{X})$ l'estimateur du maximum de vraisemblance associé à l'observation d'un échantillon X_1, \dots, X_n .

On a, quand n tend vers $+\infty$:

$$\hat{\theta}_n(\underline{X}) \xrightarrow{p.s.} \theta,$$

où θ est la vraie valeur du paramètre.

Sous les hypothèses supplémentaires que $I(\theta)$ soit inversible pour tout θ de Θ (ou strictement positive dans le cas où θ est réel) et que $L''(x; \theta)$ soit continue (en x et θ), on a alors la convergence :

$$\sqrt{n}(\hat{\theta}_n(\underline{X}) - \theta) \xrightarrow{\mathcal{L}} N_p(0, I^{-1}(\theta)),$$

quand n tend vers $+\infty$. L'estimateur du maximum de vraisemblance est donc asymptotiquement gaussien et efficace.

On peut donc dire qu'asymptotiquement, c'est à dire pour un échantillon de grande taille n , l'estimateur du maximum de vraisemblance est approximativement de loi normale centrée sur le paramètre θ et de matrice de covariance l'inverse de l'information de Fisher $I_n(\theta)$ associée au modèle d'échantillonnage, i.e.

$$\hat{\theta}_n(\underline{X}) \underset{n \text{ grand}}{\approx} N_p(\theta, I_n^{-1}(\theta)).$$

4. La δ -méthode ou l'étude asymptotique d'un estimateur obtenu par la méthode de substitution

Il est courant que l'on cherche à estimer $g(\theta)$ alors que l'on dispose déjà d'un estimateur $\hat{\theta}_n$ de θ , obtenu par exemple par la méthode des moments ou du maximum de vraisemblance. Nous avons déjà rencontré cette situation à plusieurs reprises dans ce cours ou en T.D. (Cf. e.g. l'exemple introductif en Section 1 du Chapitre 1).

La méthode de substitution, vue au paragraphe 2, nous propose d'estimer $g(\theta)$ par $g(\hat{\theta}_n)$. Si la fonction g est continue, le théorème de Slutsky nous permet de déduire la consistance de $g(\hat{\theta}_n)$ à partir de celle de $\hat{\theta}_n$.

La δ -méthode permet, elle, de prouver aisément l'asymptotique normalité de $g(\hat{\theta}_n)$ en se basant sur celle de $\hat{\theta}_n$ que l'on a pu obtenir par un calcul direct ou, plus souvent, à l'aide des théorèmes précédents.

THÉORÈME 7.5. Soit $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}\})$ un modèle paramétrique et $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans ce modèle. Supposons que l'on dispose d'un estimateur $\hat{\theta}_n(\underline{X})$ de θ asymptotiquement normal, i.e. tel que

$$\sqrt{n}(\hat{\theta}_n(\underline{X}) - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2),$$

quand n tend vers $+\infty$.

Soit g une fonction dérivable de Θ dans $\Theta' \subset \mathbb{R}$ telle que $g(\hat{\theta}_n(\underline{X}))$ soit de carré intégrable.

Alors l'estimateur $g(\hat{\theta}_n(\underline{X}))$ de $g(\theta)$ est également asymptotiquement normal. Plus précisément, on a :

$$\sqrt{n} \left(g(\hat{\theta}_n(\underline{X})) - g(\theta) \right) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2 (g'(\theta))^2),$$

quand n tend vers $+\infty$.

Preuve. L'idée principale est d'effectuer un développement de Taylor de $g(\hat{\theta}_n(\underline{X}))$ autour de $g(\theta)$. On a en effet :

$$g(\theta_n) - g(\theta) = (\theta_n - \theta) (g'(\theta) + \varepsilon(\theta_n - \theta)),$$

avec $\varepsilon(\theta_n - \theta)$ qui tend vers 0 quand $\theta_n - \theta$ tend vers 0.

On peut donc écrire ici :

$$\sqrt{n} \left(g(\hat{\theta}_n(\underline{X})) - g(\theta) \right) = \sqrt{n} (\hat{\theta}_n(\underline{X}) - \theta) \left(g'(\theta) + \varepsilon(\hat{\theta}_n(\underline{X}) - \theta) \right).$$

L'asymptotiquement normalité de $\hat{\theta}_n(\underline{X})$ nous donne, d'une part, la convergence de $\varepsilon(\hat{\theta}_n(\underline{X}) - \theta)$ en probabilité vers 0 et, d'autre part, la convergence en loi du premier facteur à droite de l'égalité vers une loi normale $N(0, \sigma_\theta^2)$. Le théorème de Slutsky permet de terminer la démonstration. \square

Remarque. La δ -méthode est également vraie dans le cas multidimensionnel. Supposons en effet que Θ soit dans \mathbb{R}^p et que g soit une fonction de Θ vers $\Theta' \subset \mathbb{R}^q$. Si l'on a

$$\sqrt{n} (\hat{\theta}_n(\underline{X}) - \theta) \xrightarrow{\mathcal{L}} N_p(0, \Sigma(\theta)),$$

alors

$$\sqrt{n} \left(g(\hat{\theta}_n(\underline{X})) - g(\theta) \right) \xrightarrow{\mathcal{L}} N_q(0, J_g(\theta) \Sigma(\theta) J_g'(\theta)),$$

où

$$J_g(\theta) = \left(\frac{\partial g_i}{\partial \theta_j}(\theta) \right)_{i=1, \dots, q; j=1, \dots, p}$$

est la matrice jacobienne de la fonction g . \diamond

5. Estimateurs par la méthode des moments

Les résultats précédents nous permettent aisément d'obtenir le comportement asymptotique des estimateurs obtenus par la méthode des moments. Rappelons que ces estimateurs sont de la forme

$$\hat{\theta}_n(\underline{X}) = h^{-1} \left(\frac{1}{n} \sum_{i=1}^n \varphi(X_i) \right)$$

avec h fonction bijective et continue de $\Theta \subset \mathbb{R}^p$ vers $h(\Theta) \subset \mathbb{R}^p$ et φ une fonction mesurable de E vers \mathbb{R}^p telle que $\mathbb{E}_\theta(\varphi(X))$ existe et toutes les deux telles que l'on ait :

$$h(\theta) = \mathbb{E}_\theta(\varphi(X)),$$

pour tout θ de Θ .

PROPOSITION 7.6. *Supposons que la fonction réciproque h^{-1} soit continue, dérivable et telle que l'estimateur $\hat{\theta}_n(\underline{X})$ par la méthode des moments soit de carré intégrable.*

Alors $\hat{\theta}_n(\underline{X})$ est fortement consistant et asymptotiquement gaussien. Plus précisément on a, quand $n \rightarrow +\infty$, d'une part la convergence p.s.

$$\hat{\theta}_n(\underline{X}) \xrightarrow{p.s.} \theta,$$

et, d'autre part,

$$\sqrt{n}(\hat{\theta}_n(\underline{X}) - \theta) \xrightarrow{\mathcal{L}} N_p(0, J_{h^{-1}} \Sigma_{\varphi(X)}(\theta) J_{h^{-1}}'),$$

où $\Sigma_{\varphi(X)}(\theta)$ est la matrice de covariance de $\varphi(X)$ et $J_{h^{-1}}$ la matrice Jacobienne de la fonction h^{-1} . En dimension 1, cette convergence s'écrit :

$$\sqrt{n}(\hat{\theta}_n(\underline{X}) - \theta) \xrightarrow{\mathcal{L}} N\left(0, \text{Var}_\theta(\varphi(X)) \left((h^{-1})'(\theta)\right)^2\right).$$

Preuve. Par la loi forte des grands nombres et le théorème de Slutsky, on a la convergence

$$\hat{\theta}_n(\underline{X}) \xrightarrow{p.s.} h^{-1}(\mathbb{E}_\theta(\varphi(X))) = \theta,$$

quand $n \rightarrow +\infty$.

Le théorème de la limite centrale appliquée à $\frac{1}{n} \sum_{i=1}^n \varphi(X_i)$ et la δ -méthode permettent d'obtenir le caractère asymptotiquement gaussien. \square

6. Exercices

Les exercices concernant cette partie constituent souvent les premières questions des exercices sur les intervalles de confiance.

Partie 3

Intervalles de confiance

Nous avons vu dans la partie précédente comment construire un estimateur de $g(\theta)$ dans un modèle paramétrique. Nous avons également introduit des critères d'évaluation de la qualité, à taille finie ou asymptotiquement, de cet estimateur. Nous avons enfin étudié la possibilité d'améliorer un estimateur et de construire directement des estimateurs optimaux.

Malgré tout cela, cette estimation de $g(\theta)$ par un estimateur $T(\underline{X})$ conserve un inconvénient majeur : celui de donner une estimation ponctuelle. Même si l'on prend toutes les précautions pour qu'elle soit proche de la vraie valeur, il n'en reste pas moins vraie qu'elle est souvent (voire toujours) différente de la valeur cible. En particulier, si l'estimateur proposé $T(\underline{X})$ est absolument continu, on a :

$$P(T(\underline{X}) = g(\theta)) = 0$$

et l'on est donc presque sûr de se tromper.

EXEMPLE 7.1. *Modèle gaussien réel avec variance σ^2 connue.*

Même si, dans ce modèle, l'estimateur \bar{X}_n est optimal dans la classe des estimateurs sans biais du paramètre μ (Cf. Exemple 6.3), on a

$$P(\bar{X}_n = \mu) = 0$$

et cet estimateur se trompe donc presque sûrement toujours. ◇

C'est pourquoi on cherchera aussi parfois à donner un intervalle (resp. une région dans le cas où $g(\theta)$ est un vecteur) de valeurs possibles pour $g(\theta)$. On parle d'intervalle ou de région de confiance. Le terme fourchette d'estimation est également utilisé.

On construira cet intervalle à partir des observations d'un échantillon et on souhaitera qu'il contienne la vraie valeur $g(\theta)$, avec une probabilité suffisamment faible de se tromper.

CHAPITRE 8

Intervalles de confiance exacts

Considérons un modèle paramétrique $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ et $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans ce modèle.

DÉFINITION 8.1. Soit α un réel dans $[0, 1]$. On appelle **région de confiance** $1 - \alpha$ pour $g(\theta)$ la fonction C de E^n à valeurs dans $g(\Theta)$ telle que

$$P(C(\underline{X}) \ni g(\theta)) = 1 - \alpha.$$

Si $g(\Theta)$ est inclus dans \mathbb{R} , on parle d'**intervalle de confiance** et l'on a $C(\underline{X}) = [L(\underline{X}), U(\underline{X})]$, où $L(\underline{X})$ et $U(\underline{X})$ sont respectivement les bornes inférieures et supérieures de l'intervalle.

Bien sûr l'expression d'une région de confiance ou d'un intervalle de confiance ne doit pas dépendre du paramètre (inconnu) θ . Il doit être uniquement déterminé par des statistiques.

Dans le cadre de ce cours nous considérerons essentiellement des intervalles de confiance. Aussi dans la suite de ce chapitre nous utiliserons seulement la terminologie d'intervalle de confiance. Mais le raisonnement serait tout à fait similaire pour les régions de confiance.

L'intervalle de confiance introduit dans la définition précédente est parfois appelé intervalle de confiance bilatéral. On peut naturellement définir des intervalles unilatéraux. Ainsi, on parlera d'intervalle unilatéral gauche (resp. droit) tout intervalle de la forme $] - \infty, U(\underline{X})]$ (resp. $[L(\underline{X}), +\infty[$).

Les bornes de l'intervalle de confiance sont des v.a. puisqu'elles sont fonctions de l'échantillon \underline{X} . D'un échantillon observé à l'autre elles donneront donc un autre intervalle. C'est pourquoi on donne une probabilité que l'intervalle contienne la vraie valeur $g(\theta)$. Par abus de notation, il peut arriver que l'on note

$$P(g(\theta) \in [L(\underline{X}), U(\underline{X})]) = 1 - \alpha$$

ou bien encore

$$P(L(\underline{X}) \leq g(\theta) \leq U(\underline{X})) = 1 - \alpha.$$

et que l'on parle de la probabilité que $g(\theta)$ soit dans l'intervalle $C(\underline{X})$. Mais il ne faut pas oublier que l'aléa porte sur les bornes et non sur $g(\theta)$.

La confiance de l'intervalle est notée $1 - \alpha$. Il y a donc une probabilité α que cet intervalle ne contienne pas la vraie valeur. On dit parfois que l'intervalle est de niveau α .

EXEMPLE 8.1. *Modèle gaussien réel avec variance σ^2 connue.*

Si $\underline{X} = (X_1, \dots, X_n)$ est un échantillon dans le modèle gaussien $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}\}$, on sait que la moyenne empirique \bar{X}_n est de loi $N(\mu, \sigma^2/n)$. Ainsi,

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

En notant z_α le quantile d'ordre α de la loi $N(0, 1)$, i.e. le réel tel que

$$P(Z \leq z_\alpha) = \alpha,$$

on peut écrire

$$P(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha$$

si l'on veut (ce sera souvent le cas) répartir la probabilité d'erreur équitablement au dessus et au dessous de l'intervalle.

Or, on a :

$$\begin{aligned} -z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2} \\ \iff \bar{X}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}. \end{aligned}$$

Ainsi, l'intervalle

$$\left[\bar{X}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

est un intervalle de confiance $1 - \alpha$ pour l'espérance μ du modèle gaussien. On constate clairement dans cet exemple que les bornes de l'intervalle de confiance sont aléatoires. \diamond

Il est évident que, plus on augmente la confiance de l'intervalle (i.e. plus on diminue α), plus l'intervalle sera grand, comme le souligne l'application numérique à suivre.

EXEMPLE 8.2. *Modèle gaussien réel avec variance σ^2 connue (A.N. de l'Exemple 8.1).*

À l'aide d'un logiciel statistique (par exemple *R*) on simule un échantillon de taille $n = 20$ d'une loi normale $N(0, 1)$. La statistique exhaustive $T(\underline{X}) = \sum_{i=1}^n X_i$ pour ce modèle nous donne sur cet échantillon $T(\underline{x}) = 8.2314$. En prenant une valeur $\alpha = 5\%$, on trouve (par exemple à l'aide de tables statistiques ou avec le même logiciel) que le quantile d'ordre $1 - \alpha/2$ est : $z_{1-\alpha/2} = 1.96$. Ainsi l'intervalle de confiance $1 - \alpha = 95\%$ est : $[-0.0267, 0.8498]$.

Comme nous l'avons dit plus haut, si l'on augmente la confiance de l'intervalle, l'intervalle de confiance sera plus grand. Ainsi, si l'on souhaite un intervalle de confiance 99% (i.e. $\alpha = 1\%$), le quantile d'ordre $1 - \alpha/2$ est : $z_{1-\alpha/2} = 2.5758$ (les tables donnent 2.58). Avec les valeurs précédentes de l'échantillon, l'intervalle de confiance 99% est alors : $[-0.1644, 0.9875]$. L'intervalle obtenu est contient donc le précédent.

Supposons maintenant que l'on simule dix nouvelles observations dans ce même modèle (on dispose maintenant de 30 observations) et que la nouvelle valeur de la statistique exhaustive (avec $n = 30$) soit $T(\underline{x}) = 6.2539$. L'intervalle de confiance 95% devient alors : $[-0.1494, 0.5663]$, ce qui illustre bien que, quand n augmente, l'intervalle a tendance à se rétrécir et se centrer sur la valeur théorique de μ (ici 0). \diamond

On a vu dans les exemples précédents comment construire un intervalle de confiance pour le paramètre μ dans un modèle gaussien avec σ^2 connu. Un examen attentif de la technique utilisée fait apparaître le rôle central qu'a joué la v.a. Z . On remarque d'une part qu'il s'agit d'une v.a. qui s'exprime comme une fonction des observations (via \bar{X}_n) et du paramètre μ . Il ne s'agit donc pas d'une statistique ! D'autre part cette v.a. Z est de loi entièrement connue (ici une $N(0, 1)$). On dit que Z est une variable pivotale. Il est souvent utile d'utiliser de telles variables pour construire des intervalles de confiance.

DÉFINITION 8.2. Soit $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ un modèle paramétrique et $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans ce modèle.

On appelle **variable pivotale** pour $g(\theta)$ toute v.a. $\pi(\underline{X}, g(\theta))$, fonction de l'échantillon \underline{X} et du paramètre $g(\theta)$, dont la loi ne dépend pas de θ .

Bien sûr une telle variable pivotale n'a d'intérêt que si l'on connaît sa loi (il n'est pas suffisant de savoir qu'elle ne dépend pas de θ).

EXEMPLE 8.3. *Modèle gaussien réel avec variance σ^2 connue (suite de l'Exemple 8.1).*

La v.a.

$$Z = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$$

est bien une fonction de \underline{X} et du paramètre μ . Elle est de loi connue $N(0, 1)$, clairement indépendante du paramètre μ . Il s'agit donc une variable pivotale pour μ dans le modèle gaussien réel avec variance connue. \diamond

CHAPITRE 9

Intervalles de confiance asymptotiques

Parfois il n'est pas possible, ou il est peu aisé, de trouver une variable pivotale. En revanche, grâce aux théorèmes asymptotiques vus en fin de chapitre précédent, il est souvent possible de déterminer une variable asymptotiquement pivotale.

DÉFINITION 9.1. Soit $(E, \mathcal{E}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ un modèle paramétrique et, pour chaque n un échantillon $\underline{X}_n = (X_1, \dots, X_n)$ dans ce modèle.

On appelle **variable asymptotiquement pivotale** pour $g(\theta)$ toute suite de v.a. $(\pi(\underline{X}_n, g(\theta)))_{n \in \mathbb{N}}$, fonctions de l'échantillon \underline{X}_n et du paramètre $g(\theta)$, convergeant en loi vers une v.a. de loi ne dépendant pas de θ . C'est à dire que l'on a, quand $n \rightarrow +\infty$:

$$\pi(\underline{X}_n, g(\theta)) \xrightarrow{\mathcal{L}} Z,$$

où Z est une v.a. de loi ne dépendant pas de θ .

EXEMPLE 9.1. Variable asymptotiquement pivotale pour l'espérance dans un modèle statistique paramétrique.

Considérons un modèle paramétrique $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ tel que la v.a. générique X soit de carré intégrable, d'espérance qui s'écrive sous la forme $g(\theta)$ et de variance σ^2 . Soit enfin, pour tout n dans \mathbb{N} , un échantillon \underline{X}_n dans ce modèle.

D'après le théorème de la limite centrale, on sait que

$$\pi(\underline{X}_n, g(\theta)) = \sqrt{n} \frac{\bar{X}_n - g(\theta)}{\sqrt{\sigma^2}} \xrightarrow{\mathcal{L}} N(0, 1),$$

quand $n \rightarrow +\infty$. Ainsi, la suite de v.a. $(\pi(\underline{X}_n, g(\theta)))_{n \in \mathbb{N}}$ est asymptotiquement pivotale. \diamond

D'une manière générale, les résultats d'asymptotique normalité présentés pour les estimateurs de la moyenne empirique, du maximum de vraisemblance ou obtenus par la méthode des moments, permettent de construire des suites de v.a. asymptotiquement pivotale.

Une v.a. asymptotiquement pivotale permet naturellement de construire des intervalles de confiance asymptotiques.

DÉFINITION 9.2. Soit α un réel dans $[0, 1]$. On appelle **intervalle de confiance asymptotique** $1-\alpha$ pour $g(\theta)$ toute suite $(C_n(\underline{X}_n))_{n \in \mathbb{N}} = ([L_n(\underline{X}_n), U(\underline{X}_n)])_{n \in \mathbb{N}}$ d'intervalles tels que

$$\lim_{n \rightarrow +\infty} P(C_n(\underline{X}_n) \ni g(\theta)) = 1 - \alpha.$$

EXEMPLE 9.2. *Intervalle de confiance asymptotique pour l'espérance dans un modèle statistique paramétrique (suite de l'Exemple 9.1) quand la variance σ^2 est connue.*

En s'inspirant de la construction de l'intervalle de confiance exact pour l'espérance d'une loi normale (Cf. Exemple 8.1), et en utilisant la suite de v.a. asymptotiquement pivotale $\pi(\underline{X}_n, g(\theta))$ vue dans l'Exemple 9.1, on montre aisément qu'un intervalle de confiance asymptotique $1 - \alpha$ pour $g(\theta)$ est donné par :

$$\left[\bar{X}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right].$$

Insistons bien sur le fait que ce résultat n'est valable que si la variance σ^2 est connue. \diamond

CHAPITRE 10

Exercices sur les intervalles de confiance exacts et asymptotiques

Exercice 1 (*Etude asymptotique du modèle de Bernoulli*)

On considère le modèle de Bernoulli $(\{0, 1\}, \mathcal{P}(\{0, 1\}), \{\mathcal{B}(p) : p \in]0, 1[\})$. On a vu, à maintes reprises (en particulier dans l'exercice 2 du chapitre 5 dont on gardera les notations), que ce modèle pouvait être, entre autres, utile dans un problème de modélisation en Fiabilité.

1) Montrer, de deux manières différentes, que l'estimateur par maximum de vraisemblance \hat{p}_n du paramètre p de ce modèle est asymptotiquement normal. Donner une approximation de la loi de \hat{p}_n quand la taille n de l'échantillon est grande.

2) En utilisant les résultats de l'exercice 2 du chapitre 5, montrer que l'estimateur de la fonction de répartition empirique en $\hat{F}_n(x)$ est également asymptotiquement normal.

3) Construire un intervalle de confiance asymptotique $1 - \alpha$ pour le paramètre p du modèle de Bernoulli.

4) En déduire un intervalle de confiance asymptotique $1 - \alpha$ pour $F(x)$, avec x fixé.

Exercice 2 (*Etude asymptotique et Intervalles de confiance (exacts et asymptotiques) dans le modèle de la loi exponentielle*)

On considère le modèle de la loi exponentielle $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}(\lambda) : \lambda > 0\})$ et $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans ce modèle. On rappelle que l'estimateur du maximum de vraisemblance du paramètre λ basé sur l'observation d'un tel échantillon est

$$\hat{\lambda}_n = \frac{1}{\bar{X}_n}.$$

1) En utilisant la propriété vue en cours sur l'estimateur du maximum de vraisemblance, montrer que l'estimateur $\hat{\lambda}_n$ est asymptotiquement normal (on précisera bien la convergence en loi obtenue).

2) Retrouver le résultat de la question précédente en utilisant en particulier la δ -méthode.

3) Déduire, de ce comportement asymptotiquement normal de $\hat{\lambda}_n$, un intervalle de confiance asymptotique $1 - \alpha$ pour λ .

4) Montrer que si Y est une v.a. de loi Gamma $G(\alpha, \beta)$, alors la v.a. βY est de loi $G(\alpha, 1)$.

5) En utilisant le résultat de la question précédente et celui vu dans l'exercice 5 (partie 2) du chapitre 4, montrer que l'intervalle

$$\left[\frac{\chi_{\alpha/2}^2(2n)}{2n\bar{X}_n}, \frac{\chi_{1-\alpha/2}^2(2n)}{2n\bar{X}_n} \right]$$

est un intervalle de confiance $1 - \alpha$ exact pour λ , où $\chi_{\alpha}^2(n)$ est le quantile d'ordre α d'une loi $\chi^2(n)$. (Ind. On rappelle qu'une loi $\chi^2(n)$ est une loi $G(n/2, 1/2)$)

Partie 4

Correction des exercices

Correction des exercices du Chapitre 2

Exercice 1 (*Familles Exponentielles*)

On considère les modèles suivants :

- Modèle Binomial $\{B(m, p) : p \in [0, 1]\}$;
- Modèle de Poisson $\{\mathcal{P}(\lambda) : \lambda > 0\}$;
- Modèle gaussien à variance fixée $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}\}$;
- Modèle gaussien à paramètre bi-dimensionnel $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$;
- Modèle Gamma $\{G(\alpha, \beta) : \alpha > 0, \beta > 0\} = \{f_{\alpha, \beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\mathbb{R}^+}(x) : \alpha > 0, \beta > 0\}$;
- Modèle uniforme $\{\mathcal{U}_{[0, \theta]} : \theta > 0\}$;
- Modèle de Cauchy $\{f_\theta(x) = \frac{1}{\pi(1+(x-\theta)^2)} : \theta \in \mathbb{R}\}$;
- Modèle Multinomial $\{\mathcal{M}(n, p_1, \dots, p_k) : 0 < p_i < 1, \forall i = 1, \dots, k \text{ et } \sum_{i=1}^k p_i = 1\}$.

Pour tous ces modèles, répondre aux questions suivantes.

- 1) Quelle est l'expression de la densité $f_\theta(x)$?
- 2) Le modèle constitue-t-il une famille exponentielle générale ? Naturelle ? Quel est le paramètre canonique du modèle ?
- 3) Quelle est la vraisemblance d'un échantillon $x = (x_1, \dots, x_n)$?

Solution

- *Modèle statistique de la loi Binomiale* $\{B(m, p) : p \in [0, 1]\}$

La densité, pour tout x dans \mathbb{N} , est

$$\begin{aligned} f_p(x) &= \binom{m}{x} p^x (1-p)^{m-x} \\ &= \binom{m}{x} (1-p)^m \left(\frac{p}{1-p}\right)^x \\ &= \exp\left\{x \ln\left(\frac{p}{1-p}\right)\right\} (1-p)^m \binom{m}{x}. \end{aligned}$$

En posant

$$C(p) = (1-p)^m, \quad h(x) = \binom{m}{x}, \quad T(x) = x, \quad \text{et } \eta(p) = \ln\left(\frac{p}{1-p}\right),$$

on constate que le modèle de la loi Binomiale est une famille exponentielle naturelle dont le paramètre canonique est $\theta = \ln(p/(1-p))$. La vraisemblance de l'échantillon

x_1, \dots, x_n est :

$$L(x_1, \dots, x_n; p) = \prod_{i=1}^n \binom{m}{x_i} p^{x_i} (1-p)^{m-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{nm - \sum_{i=1}^n x_i} \prod_{i=1}^n \binom{m}{x_i}.$$

- *Modèle Statistique de la loi de Poisson* $\{\mathcal{P}(\lambda) : \lambda > 0\}$

La densité, en tout x de \mathbb{N} , est

$$\begin{aligned} f_\lambda(x) &= \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \exp\{x \ln \lambda\} e^{-\lambda} \frac{1}{x!}. \end{aligned}$$

En posant

$$C(\lambda) = e^{-\lambda}, \quad h(x) = \frac{1}{x!}, \quad T(x) = x \text{ et } \eta(\lambda) = \ln \lambda,$$

on vérifie que ce modèle est une famille exponentielle naturelle de paramètre canonique $\theta = \ln(\lambda)$. La vraisemblance de l'échantillon x_1, \dots, x_n est :

$$L(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}.$$

- *Modèle Statistique de la loi normale avec variance connue* $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}\}$

La densité est

$$\begin{aligned} f_\mu(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right\} \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma^2} + \frac{x\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right\} \\ &= \exp\left\{x \frac{\mu}{\sigma^2}\right\} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{\mu^2}{2\sigma^2}\right\} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}. \end{aligned}$$

En posant

$$C(\mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}, \quad h(x) = \exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \quad T(x) = x \text{ et } \eta(\mu) = \frac{\mu}{\sigma^2},$$

il apparaît que ce modèle constitue une famille exponentielle naturelle de paramètre canonique $\theta = \mu/\sigma^2$. La vraisemblance est

$$L(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma}\right)^2\right\} = \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\}.$$

- *Modèle Statistique de la loi normale à deux paramètres* $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$

Dans ce modèle la densité est :

$$\begin{aligned} f_{\mu, \sigma^2}(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\} \\ &= \exp \{ \langle \eta(\mu, \sigma^2), T(x) \rangle \} \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{\mu^2}{2\sigma^2} \right\}, \end{aligned}$$

où

$$\eta(\mu, \sigma^2) = \left(\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2} \right) \text{ et } T(x) = (x, -x^2).$$

En posant

$$C(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{\mu^2}{2\sigma^2} \right\} \text{ et } h(x) = 1,$$

on constate que ce modèle est une famille exponentielle générale de paramètre canonique $\theta = (\mu/\sigma^2, 1/(2\sigma^2))$.

- *Modèle statistique de la loi Gamma* $\{G(\alpha, \beta) : \alpha > 0, \beta > 0\}$

La densité est

$$\begin{aligned} f_{\alpha, \beta}(x) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\mathbb{R}^+}(x) \\ &= \exp \{ (\alpha - 1) \ln x - \beta x \} \frac{\beta^\alpha}{\Gamma(\alpha)} \mathbb{1}_{\mathbb{R}^+}(x). \end{aligned}$$

En posant

$$C(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}, \quad h(x) = \mathbb{1}_{\mathbb{R}^+}(x), \quad \eta(\alpha, \beta) = (\alpha - 1, \beta) \text{ et } T(x) = (\ln x, -x),$$

ce modèle s'écrit sous la forme d'une famille exponentielle générale. Le lecteur trouvera sans peine le paramètre canonique et l'expression de la vraisemblance.

- *Modèle statistique de la loi uniforme* $\{\mathcal{U}_{[0, \theta]} : \theta > 0\}$

La densité est

$$f_\theta(x) = \frac{1}{\theta} \mathbb{1}_{[0, \theta]}(x)$$

et on constate que l'on ne peut pas l'écrire sous la forme d'une famille exponentielle.

- *Modèle statistique de la loi de Cauchy* $\left\{ f_\theta(x) = \frac{1}{\pi(1 + (x - \theta)^2)} : \theta \in \mathbb{R} \right\}$

La densité est

$$\begin{aligned} f_\theta(x) &= \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2} \\ &= \frac{1}{\pi} \exp \left\{ -\ln(1 + (x - \theta)^2) \right\} \end{aligned}$$

que l'on ne peut pas écrire sous la forme $f_\theta(x) = C(\theta)h(x) \exp\{\langle \eta(\theta), T(x) \rangle\}$. Ainsi, il ne s'agit pas d'une famille exponentielle.

• *Modèles statistique de la loi Multinomiale* $\{\mathcal{M}(n, p_1, \dots, p_k) : 0 < p_i < 1, \forall i = 1, \dots, k \text{ et } \sum_{i=1}^k p_i = 1\}$

Ici le paramètre est $\theta = (p_1, \dots, p_k) \in \mathbb{R}^k$. Pour tous $x_i \in \mathbb{N}$, $i = 1, \dots, k$, tels que $\sum_{i=1}^k x_i = n$, la densité de la loi binomiale est

$$\begin{aligned} f_{p_1, \dots, p_k}(x_1, \dots, x_k) &= \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \\ &= \exp \{x_1 \ln p_1 + \cdots + x_k \ln p_k\} \frac{n!}{x_1! \cdots x_k!}. \end{aligned}$$

Soit $\Delta = \{(x_1, \dots, x_k) \in \mathbb{N}^k : \sum_{i=1}^k x_i = n\}$. En posant

$$C(p_1, \dots, p_k) = 1$$

$$h(x_1, \dots, x_k) = \frac{n!}{x_1! \cdots x_k!} \mathbb{1}_{\Delta}(x_1, \dots, x_k)$$

$$\eta(\theta) = \begin{pmatrix} \ln p_1 \\ \vdots \\ \ln p_k \end{pmatrix} \text{ et } T(x) = \begin{pmatrix} \ln x_1 \\ \vdots \\ \ln x_k \end{pmatrix},$$

il apparaît que le modèle de la loi multinomiale est une famille exponentielle. Ici aussi nous laissons le soin au lecteur de trouver le paramètre canonique et l'expression de la vraisemblance.

Exercice 2 (Modèles position-échelle)

1) Construire un modèle position-échelle à partir de la loi exponentielle $\mathcal{E}(1)$. Préciser la forme des f.d.r. des lois de ce modèle ainsi que leurs densités.

2) Montrer que le modèle uniforme $\{\mathcal{U}_{[a,b]} : -\infty < a < b < +\infty\}$ est un modèle position-échelle.

Solution

1) Soit $X \sim \mathcal{E}(1)$. Cette v.a.r. est à valeurs dans \mathbb{R}_+^* et sa f.d.r. est $F_\lambda(x) = 1 - e^{-x}$, pour x strictement positif. Posons $Y = x_0 + X/\lambda$. Cette v.a.r. est à valeurs dans $]x_0, +\infty[$ et a pour f.d.r., pour tout $y > x_0$:

$$\begin{aligned} F_\lambda(y) &= P(Y \leq y) \\ &= P\left(x_0 + \frac{1}{\lambda}X \leq y\right) \\ &= P(X \leq \lambda(y - x_0)) \\ &= 1 - e^{-\lambda(y-x_0)}. \end{aligned}$$

Sa densité est alors :

$$f_\lambda(y) = F'_\lambda(y) = \lambda e^{-\lambda(y-x_0)},$$

pour tout $y > x_0$. On obtient un modèle position-échelle de paramètre de position x_0 et de paramètre d'échelle $1/\lambda$.

2) Soit X une v.a.r. de loi uniforme $\mathcal{U}_{[0,1]}$. Sa f.d.r. est $F(x) = x$ sur $[0, 1]$. Posons $Y = a + (b - a)X$. Cette v.a.r. est clairement à valeurs dans $[a, b]$. Sa fonction de répartition est donnée, pour tout y dans $[a, b]$ par

$$\begin{aligned} P(Y \leq y) &= P(a + (b - a)X \leq y) \\ &= P\left(X \leq \frac{y - a}{b - a}\right) \\ &= \frac{y - a}{b - a}. \end{aligned}$$

Remarquons que l'on aurait pu obtenir le même résultat en déterminant la loi de Y en utilisant le théorème du changement de variable.

Le modèle considéré est donc un modèle position échelle engendré par la loi $\mathcal{U}_{[0,1]}$ de paramètre de position a et de paramètre d'échelle $c = b - a$.

Exercice 3 (*Statistiques d'ordre*)

Soit X_1, \dots, X_n des v.a.r. définies sur un même espace probabilisé $(\Omega, \mathfrak{A}, P)$, indépendantes et de même loi absolument continue par rapport à la mesure de Lebesgue de densité f . Pour tout ω dans Ω , on peut ordonner les réels $X_1(\omega), \dots, X_i(\omega), \dots, X_n(\omega)$ sous la forme

$$X_{(1)}(\omega) \leq X_{(2)}(\omega) \leq \dots \leq X_{(i)}(\omega) \leq \dots \leq X_{(n)}(\omega).$$

L'application

$$X_{(i)} : \omega \in \Omega \rightarrow X_{(i)}(\omega)$$

ainsi définie pour chaque i est une v.a.r. dite $i^{\text{ème}}$ statistique d'ordre.

- 1) Calculer cette de $X_{(n)} = \sup\{X_1, \dots, X_n\}$ (f.d.r. et densité).
- 2) Calculer la loi de $X_{(1)} = \inf\{X_1, \dots, X_n\}$ (f.d.r. et densité).
- 3) Calculer la loi du couple $(X_{(1)}, X_{(n)})$. En déduire celle de l'étendue $R = X_{(n)} - X_{(1)}$ (on donnera sa f.d.r et sa densité en fonction de F et f).
- 4) Soit N_y le nombre de X_i inférieurs à y . Quelle est la loi de N_y ? Que dire des événements $\{N_y \geq k\}$ et $\{X_{(k)} \leq y\}$? En déduire la f.d.r. de $X_{(k)}$.
- 5) On pourrait du résultat précédent tirer la densité de la v.a. $X_{(k)}$. Mais c'est fastidieux. Il y a bien plus simple en attaquant le problème directement, ce que l'on propose de faire maintenant. On pourra utiliser le résultat suivant : Si f est continue sur un intervalle $[a, b]$, alors, pour tout x dans cet intervalle, on a :

$$f(x) = \lim_{h \rightarrow 0^+} \frac{P(X \in]x, x + h])}{h}$$

- 6) Montrer que si $\mathbb{E}(X)$ existe alors $\mathbb{E}(X_{(k)})$ aussi.
 - 7) Calculer la densité du vecteur $(X_{(1)}, \dots, X_{(n)})$.
- (Ind. on pourra calculer $P((X_{(1)}, \dots, X_{(n)}) \in B)$, pour tout borélien B de $\mathcal{B}_{\mathbb{R}^n}$).

Solution

1) On a

$$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(\cap_{i=1}^n \{X_i \leq x\}) = \prod_{i=1}^n P(X_i \leq x) = F^n(x),$$

où l'avant dernière égalité est justifiée par l'indépendance entre les v.a.r. X_1, \dots, X_n . Cette fonction étant dérivable (puisque F l'est) sur \mathbb{R}^+ , la densité de X_n est :

$$f_{X_{(n)}}(x) = nF^{n-1}(x)f(x).$$

2) On a

$$P(X_{(1)} > x) = P(\cap_{i=1}^n \{X_i > x\}) = \prod_{i=1}^n P(X_i > x) = (1 - F(x))^n,$$

où l'avant dernière égalité est ici aussi justifiée par l'indépendance entre les v.a.r. X_1, \dots, X_n . D'où

$$F_{X_{(1)}}(x) = 1 - (1 - F(x))^n$$

et

$$f_{X_{(1)}}(x) = -n(1 - F(x))^{n-1}(-f(x)) = n(1 - F(x))^{n-1}f(x).$$

3) Supposons dans un premier temps que $x_1 \leq x_n$. On peut écrire :

$$\begin{aligned} P(X_{(1)} \leq x_1, X_{(n)} \leq x_n) &= P(X_{(n)} \leq x_n) - P(X_{(1)} > x_1, X_{(n)} \leq x_n) \\ &= F^n(x_n) - P(\cap_{i=1}^n \{X_i \in]x_1, x_n]\}) = F^n(x_n) - (F(x_n) - F(x_1))^n. \end{aligned}$$

En dérivant deux fois, on obtient

$$\begin{aligned} \frac{\partial F_{X_{(1)}, X_{(n)}}}{\partial x_1}(x_1, x_n) &= n(F(x_n) - F(x_1))^{n-1} f(x_1) \\ \text{et} \quad \frac{\partial^2 F_{X_{(1)}, X_{(n)}}}{\partial x_1 \partial x_n}(x_1, x_n) &= n(n-1)(F(x_n) - F(x_1))^{n-2} f(x_1)f(x_n). \end{aligned}$$

Maintenant si $x_1 > x_n$, on a

$$P(X_{(1)} \leq x_1, X_{(n)} \leq x_n) = P(X_{(n)} \leq x_n) = F^n(x_n)$$

qui en dérivant par rapport à x_1 et x_n s'annule. On a donc la densité du couple $(X_{(1)}, X_{(n)})$:

$$f_{(X_{(1)}, X_{(n)})}(x_1, x_n) = n(n-1)(F(x_n) - F(x_1))^{n-2} f(x_1)f(x_n)\mathbb{1}_{\{x_1 \leq x_n\}}.$$

Disposant de la densité du couple $(X_{(1)}, X_{(n)})$, pour trouver la densité de la v.a.r. $R = X_{(n)} - X_{(1)}$, on peut dans un premier calculer la densité du couple (Q, R) , où $Q = X_{(1)}$, et ensuite calculer la loi marginale de la seconde coordonnée de ce couple.

Le calcul de la loi du couple (Q, R) s'effectue facilement grâce à la formule du changement de variable. Prenons la fonction $\varphi(u, v) = (u, v - u)$ qui est évidemment

un \mathcal{C}^1 -difféomorphisme de fonction réciproque $\varphi^{-1}(y_1, y_2) = (y_1, y_1 + y_2)$. Le Jacobien de φ^{-1} est égal à 1. Ainsi la formule du changement de variable nous donne :

$$\begin{aligned} f_{Q,R}(q, r) &= f_{X_{(1)}, X_{(n)}}(\varphi^{-1}(q, r)) |J_{\varphi^{-1}}(q, r)| \mathbb{1}_{Im\varphi}(q, r) \\ &= n(n-1) (F(q+r) - F(q))^{n-2} f(q) f(q+r) \mathbb{1}_{\mathbb{R} \times \mathbb{R}^+}(q, r). \end{aligned}$$

La densité marginale de R est donc :

$$f_R(r) = \int_{\mathbb{R}^+} n(n-1) (F(q+r) - F(q))^{n-2} f(q) f(q+r) dq.$$

Sa f.d.r. est alors :

$$F_R(r) = \int_0^r f_R(x) dx.$$

4) On a

$$N_y = \sum_{i=1}^n \mathbb{1}_{X_i \leq y}.$$

Les v.a.r. $\mathbb{1}_{X_i \leq y}$, pour $i = 1, \dots, n$, étant i.i.d. de loi de Bernoulli de paramètre $F(y)$, la loi de N_y est une Binomiale de paramètres n et $F(y)$, i.e.

$$N_y \sim B(n, F(y)).$$

Par ailleurs, on a l'égalité entre les événements :

$$\{N_y \geq k\} = \{\text{Il y a un nombre supérieur ou égal à } k \text{ de } X_i \text{ inférieurs à } y\} = \{X_{(k)} \leq y\}.$$

Ainsi, il vient :

$$F_{X_{(k)}}(x) = P(X_{(k)} \leq x) = P(N_x \geq k) = \sum_{i=k}^n C_n^i (F(x))^i (1 - F(x))^{n-i}.$$

5) Calculons la probabilité $P(X_{(k)} \in]x, x + dx])$. Différents événements disjoints peuvent donner l'événement dont on veut calculer la probabilité :

- $k-1$ variables "tombent" dans l'intervalle $] -\infty, x]$, 1 variable dans l'intervalle $]x, x + dx]$ et $n-k$ dans l'intervalle $]x + dx, +\infty[$;
- $k-2$ variables "tombent" dans l'intervalle $] -\infty, x]$, 2 variables dans l'intervalle $]x, x + dx]$ et $n-k$ dans l'intervalle $]x + dx, +\infty[$;
- $k-1$ variables "tombent" dans l'intervalle $] -\infty, x]$, 2 variables dans l'intervalle $]x, x + dx]$ et $n-k-1$ dans l'intervalle $]x + dx, +\infty[$;
- $k-3$ variables "tombent" dans l'intervalle $] -\infty, x]$, 3 variables dans l'intervalle $]x, x + dx]$ et $n-k$ dans l'intervalle $]x + dx, +\infty[$;
- etc...

Le premier événement s'écrit :

$$\begin{aligned} &\{X_{(1)}, \dots, X_{(k-1)} \text{ sont inférieurs à } x, \\ &X_{(k)} \text{ est dans l'intervalle }]x, x + dx] \\ &\text{et } X_{(k+1)}, \dots, X_{(n)} \text{ sont supérieurs à } x + dx\} \end{aligned}$$

La probabilité $P_1(dx)$ de cet événement s'obtient aisément en remarquant que l'on est dans la situation d'un tirage d'une loi multinomiale à trois résultats possibles. Ainsi :

$$P_1(dx) = \frac{n!}{(k-1)!1!(n-k)!} (F(x))^{k-1} (F(x+dx) - F(x)) (1 - F(x+dx))^{n-k}.$$

D'où on tire :

$$\lim_{dx \rightarrow 0} \frac{P_1(dx)}{dx} = \frac{n!}{(k-1)!(n-k)!} (F(x))^{k-1} f(x) (1 - F(x))^{n-k}.$$

Regardons maintenant les probabilités des autres événements ci-dessus. Pour chacun d'entre eux, il y a au moins deux variables X_i qui se trouvent dans l'intervalle $]x, x+dx]$. La probabilité de ces événements contiendra donc un terme de la forme $(F(x+dx) - F(x))^m$ avec $2 \leq m \leq n$. Ces termes divisés par dx tendront alors vers 0 quand dx tend vers 0. Ainsi toutes les probabilités des événements autres que le premier de la liste précédente divisées par dx ont une limite qui tend vers 0 quand dx tend vers 0. On a donc :

$$f_{X_{(k)}}(x) = \lim_{dx \rightarrow 0} \frac{P(X_{(k)} \in]x, x+dx])}{dx} = \frac{n!}{(k-1)!(n-k)!} (F(x))^{k-1} f(x) (1 - F(x))^{n-k}.$$

6) Comme $F(x)$ et $1 - F(x)$ sont dans $[0, 1]$, on peut écrire :

$$\mathbb{E}|X_{(k)}| = \int_{\mathbb{R}} |x| f_{X_{(k)}}(x) dx \leq \frac{n!}{(k-1)!(n-k)!} \int_{\mathbb{R}} |x| f(x) dx = \frac{n!}{(k-1)!(n-k)!} \mathbb{E}|X|$$

dont on tire aisément le résultat voulu.

7) Notons Σ_n l'ensemble des permutations sur l'ensemble $\{1, 2, \dots, n\}$. Soit B un borélien de \mathbb{R}^n . On a :

$$\begin{aligned} P((X_{(1)}, \dots, X_{(n)}) \in B) &= \sum_{\sigma \in \Sigma_n} P(\{(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in B\} \cap \{X_{\sigma(1)} < \dots < X_{\sigma(n)}\}) \\ &= \sum_{\sigma \in \Sigma_n} \int_B \mathbb{1}_{u_1 < u_2 < \dots < u_n} \left(\prod_{i=1}^n f(u_i) \right) du_1 \cdots du_n \\ &= \int_B n! \left(\prod_{i=1}^n f(u_i) \right) \mathbb{1}_{u_1 < u_2 < \dots < u_n} du_1 \cdots du_n. \end{aligned}$$

Cette égalité étant vraie pour tout borélien B de \mathbb{R}^n , on en déduit que

$$f_{X_{(1)}, \dots, X_{(n)}}(u_1, \dots, u_n) = n! \left(\prod_{i=1}^n f(u_i) \right) \mathbb{1}_{u_1 < u_2 < \dots < u_n}.$$

Correction des exercices du Chapitre 4

Exercice 1 (Modèle Gamma et Méthode des moments)

On considère le Modèle Statistique de la loi Gamma $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{G(\alpha, \beta) : \alpha > 0, \beta > 0\})$. On rappelle que la densité d'une v.a. X de loi $G(\alpha, \beta)$ est :

$$f_{\alpha, \beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\mathbb{R}^+}(x).$$

- 1) Calculer $\mathbb{E}_{\alpha, \beta}(X)$ et $Var_{\alpha, \beta}(X)$.
- 2) Par la méthode des moments, donner un estimateur du paramètre bidimensionnel (α, β) du modèle, basé sur l'observation d'un échantillon X_1, \dots, X_n .
- 3) Déterminer des estimateurs de α et β en utilisant conjointement des estimateurs empiriques des moments et la méthode de substitution.

Solution

1) On a :

$$\mathbb{E}_{\alpha, \beta}(X) = \int_0^\infty x \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \frac{1}{\Gamma(\alpha)} \int_0^\infty \beta^\alpha x^\alpha e^{-\beta x} dx.$$

En effectuant le changement de variable $u = \beta x$ et en notant que $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$, il vient :

$$\mathbb{E}_{\alpha, \beta}(X) = \frac{1}{\Gamma(\alpha)} \int_0^\infty u^\alpha e^{-u} \frac{du}{\beta} = \frac{1}{\beta \Gamma(\alpha)} \Gamma(\alpha + 1) = \frac{\alpha}{\beta}.$$

De la même manière¹ on montre que l'on a :

$$\mathbb{E}(X^2) = \frac{(\alpha + 1)\alpha}{\beta^2}.$$

Ainsi

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{(\alpha + 1)\alpha}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}.$$

2) Les calculs de la question précédente nous ont donné

$$\mathbb{E}_{\alpha, \beta}(X) = \frac{\alpha}{\beta} \text{ et } \mathbb{E}_{\alpha, \beta}(X^2) = \frac{\alpha(\alpha + 1)}{\beta^2}.$$

Ainsi on peut écrire

$$h(\alpha, \beta) = \mathbb{E}_{\alpha, \beta}(\varphi(X))$$

¹Noter que $\Gamma(\alpha + 2) = (\alpha + 1)\alpha\Gamma(\alpha)$.

avec

$$h(\alpha, \beta) = \left(\frac{\alpha}{\beta}, \frac{\alpha(\alpha + 1)}{\beta^2} \right) \text{ et } \varphi(X) = (X, X^2).$$

Des estimateurs de α et β par la méthode des moments sont alors

$$(\hat{\alpha}, \hat{\beta}) = h^{-1} \left(\frac{1}{n} \sum_{i=1}^n \varphi(X_i) \right) = h^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i, \frac{1}{n} \sum_{i=1}^n X_i^2 \right).$$

Un calcul aisé donne l'expression de h^{-1} , la réciproque de h . On obtient :

$$(\alpha, \beta) = h^{-1}(u, v) = \left(\frac{u^2}{v - u^2}, \frac{u}{v - u^2} \right).$$

Nous obtenons alors

$$\hat{\alpha} = \frac{\left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2}{\frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2}$$

$$\hat{\beta} = \frac{\frac{1}{n} \sum_{i=1}^n X_i}{\frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2}.$$

On remarque que l'on peut écrire :

$$\begin{aligned} \frac{1}{n} \left(\sum_{i=1}^n (X_i - \bar{X})^2 \right) &= \frac{1}{n} \left(\sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \right) \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X} \frac{1}{n} \sum_{i=1}^n X_i + \frac{1}{n} \sum_{i=1}^n \bar{X}^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X}^2 + \bar{X}^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2. \end{aligned}$$

Une nouvelle expression des estimateurs $\hat{\alpha}$ et $\hat{\beta}$ est alors :

$$\hat{\alpha} = \frac{(\sum_{i=1}^n X_i)^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\hat{\beta} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n (X_i - \bar{X})^2}.$$

3) On a vu que

$$\mathbb{E}_{\alpha, \beta}(X) = \frac{\alpha}{\beta} \text{ et } \text{Var}_{\alpha, \beta}(X) = \frac{\alpha}{\beta^2}.$$

Les estimateurs empiriques des moments d'ordre 1 et 2 de X nous donnent ainsi des fonctions des estimateurs de α et β sous la forme :

$$\frac{\hat{\alpha}}{\hat{\beta}} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\frac{\hat{\alpha}}{\hat{\beta}^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

De ces équations on tire

$$\hat{\beta} = \frac{\frac{1}{n} \sum_{i=1}^n X_i}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

et

$$\hat{\alpha} = \hat{\beta} \frac{1}{n} \sum_{i=1}^n X_i = \frac{(\sum_{i=1}^n X_i)^2}{n \sum_{i=1}^n (X_i - \bar{X})^2}.$$

On retrouve donc les mêmes estimateurs que dans la question précédente.

Exercice 2 (*Modèle de la loi exponentielle et Méthode des moments*)

On a vu en cours que la méthode des moments permet d'obtenir un estimateur du paramètre λ dans un modèle de la loi exponentielle : $\hat{\lambda} = 1/(\bar{X}_n)$ basé sur la relation $\mathbb{E}(X) = 1/\lambda$. L'intérêt de cet exercice est de montrer que cette méthode permet la construction de plusieurs estimateurs de ce même paramètre λ .

- 1) On suppose qu'une v.a.r. X suit une loi exponentielle $\mathcal{E}(\lambda)$. Calculer $\mathbb{E}(X^2)$.
- 2) Soit $t_0 > 0$. Écrire la fiabilité $\bar{F}(t_0) = P(X > t_0)$ sous forme d'une espérance.
- 3) On considère le modèle de la loi exponentielle $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}(\lambda) : \lambda > 0\})$. En vous inspirant des résultats des deux questions précédentes et en utilisant à chaque fois la méthode des moments, proposer deux autres estimateurs du paramètre λ .

Solution

1) On a :

$$\mathbb{E}(X^2) = \int_0^{\infty} x^2 \cdot \lambda e^{-\lambda x} dx = \int_0^{\infty} \frac{u^2}{\lambda} e^{-u} \frac{du}{\lambda} = \frac{\Gamma(3)}{\lambda^2} = \frac{2!}{\lambda^2} = \frac{2}{\lambda^2},$$

où la deuxième égalité est obtenue par changement de variable $u = \lambda x$.

2) On peut écrire

$$\bar{F}(t_0) = \mathbb{E}(\mathbb{1}_{]t_0, +\infty[}(X)).$$

3) L'égalité

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

peut s'écrire sous la forme $\mathbb{E}(\varphi(X)) = h(\lambda)$, où $\varphi(x) = x^2$ et $h(x) = 2/x^2$. La fonction $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ étant bijective de réciproque

$$h^{-1}(y) = \sqrt{\frac{2}{y}},$$

la méthode des moments nous donne

$$\hat{\lambda} = h^{-1}\left(\frac{1}{n} \sum_{i=1}^n X_i^2\right) = \sqrt{\frac{2}{\frac{1}{n} \sum_{i=1}^n X_i^2}}$$

comme estimateur de λ .

Par ailleurs, on sait (calcul aisé !) que pour une loi exponentielle on a

$$\bar{F}(t_0) = e^{-\lambda t_0}.$$

Posons

$$\varphi(x) = \mathbb{1}_{]t_0, +\infty[}(x) \text{ et } h(\lambda) = e^{-\lambda t_0}$$

où $h : \mathbb{R}^+ \rightarrow [0, 1]$ est bijective de réciproque $h^{-1}(y) = -(\ln y)/t_0$.

La méthode des moments nous donne :

$$\hat{\lambda} = h^{-1}\left(\frac{1}{n} \sum_{i=1}^n \varphi(X_i)\right) = -\frac{1}{t_0} \ln\left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i > t_0\}}\right).$$

Exercice 3 (Maximum de vraisemblance pour un modèle gaussien)

1) On considère le modèle gaussien $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}\}$. Donner l'estimateur du maximum de vraisemblance du paramètre μ basé sur une observation x_1, \dots, x_n d'un échantillon issu de ce modèle.

2) On considère maintenant le modèle gaussien avec paramètre bidimensionnel, i.e. $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$. Donner l'estimateur du maximum de vraisemblance du paramètre $\theta = (\mu, \sigma^2)$, pour le modèle d'échantillonnage associé.

Solution

1) La vraisemblance pour un échantillon x_1, \dots, x_n est :

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_n; \mu) &= \prod_{i=1}^n f_{\mu}(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma}\right)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2\right\}. \end{aligned}$$

Le logarithme de la vraisemblance est alors :

$$\ell(x_1, \dots, x_n; \mu) = \ln \mathcal{L}(x_1, \dots, x_n; \mu) = -n \ln(\sqrt{2\pi\sigma^2}) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2.$$

Les fonctions

$$\mu \mapsto \left(\frac{x_i - \mu}{\sigma}\right)^2,$$

pour $i = 1, \dots, n$, sont convexes. Ainsi, comme fonction de μ , la log-vraisemblance est concave et son maximum est atteint en la valeur qui annule la dérivée première. On résout donc :

$$\frac{\partial}{\partial \mu} \ell(x_1, \dots, x_n; \mu) = 0.$$

On a :

$$\frac{\partial}{\partial \mu} \ell(x_1, \dots, x_n; \mu) = 0 \iff -\frac{1}{2} \sum_{i=1}^n 2 \left(\frac{x_i - \mu}{\sigma^2} \right) (-1) = 0.$$

L'estimateur du maximum de vraisemblance est donc :

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

2) La log-vraisemblance est toujours :

$$\ell(x_1, \dots, x_n; \mu, \sigma^2) = \ln \mathcal{L}(x_1, \dots, x_n; \mu, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2.$$

En vérifiant que la matrice Hessienne (matrice des dérivées secondes partielles) de la log-vraisemblance est définie négative, la fonction log-vraisemblance est concave. Le maximum est donc atteint en la valeur qui annule le gradient (vecteur des dérivées partielles premières).

On a vu que

$$\frac{\partial}{\partial \mu} \ell(\mu, \sigma^2; x_1, \dots, x_n) = \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma^2} \right)$$

et on a :

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} \ell(x_1, \dots, x_n; \mu, \sigma^2) &= \frac{\partial}{\partial \sigma^2} \left\{ -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2} \right\} \\ &= -\frac{n}{2\sigma^2} - \frac{1}{2} \sum_{i=1}^n (-1) \frac{(x_i - \mu)^2}{\sigma^4} = -\frac{n}{2\sigma^2} + \frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^4}. \end{aligned}$$

En annulant ces deux dérivées partielles, on doit résoudre en μ et σ^2 le système :

$$\begin{cases} \sum_{i=1}^n (x_i - \mu) = 0 \\ \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^4} = \frac{n}{\sigma^2} \end{cases} \iff \begin{cases} \bar{x} = \mu \\ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \sigma^2 \end{cases}.$$

Il apparaît alors que les estimateurs du maximum de vraisemblance coïncident avec les estimateurs empiriques de l'espérance et de la variance :

$$\begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = S_n^2 \end{cases}.$$

Exercice 4 (*Maximum de vraisemblance pour un modèle de loi uniforme*)

On considère le modèle uniforme $\{\mathcal{U}_{[0,\theta]} : \theta > 0\}$.

1) Montrer que la vraisemblance associée à un échantillon x_1, \dots, x_n observé dans ce modèle est :

$$L(x_1, \dots, x_n; \theta) = \frac{1}{\theta^n} \mathbb{1}_{x_{(1)} \geq 0} \mathbb{1}_{x_{(n)} \leq \theta}$$

où $x_{(1)}$ et $x_{(n)}$ sont respectivement les observations des statistiques d'ordre $X_{(1)}$ et $X_{(n)}$.

2) Donner l'estimateur du maximum de vraisemblance du paramètre θ .

Solution

1) La densité de la v.a.r. générique dans ce modèle de la loi uniforme est :

$$f_\theta(x) = \frac{1}{\theta} \mathbb{1}_{[0, \theta]}(x).$$

La vraisemblance de l'échantillon (x_1, \dots, x_n) est alors :

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n f_\theta(x_i) = \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{1}_{[0, \theta]}(x_i) \\ &= \frac{1}{\theta^n} \mathbb{1}_{[0, +\infty[}(\inf_{i=1, \dots, n} x_i) \mathbb{1}_{[0, \theta]}(\sup_{i=1, \dots, n} x_i) = \frac{1}{\theta^n} \mathbb{1}_{[0, +\infty[(x_{(1)}) \mathbb{1}_{]-\infty, \theta]}(x_{(n)}). \end{aligned}$$

2) La fonction $\theta \mapsto \mathcal{L}(x_1, \dots, x_n; \theta)$ est nulle sur l'intervalle $] -\infty, x_{(n)}[$ et coïncide avec la fonction $1/\theta^n$ sur $[x_{(n)}, +\infty[$. Cette fonction n'est pas continue en $x_{(n)}$ (et donc pas dérivable). Ainsi elle n'est pas convexe sur \mathbb{R} . On ne peut donc appliquer le raisonnement habituel (recherche du zéro de la dérivée première).

Mais il apparaît clairement que le maximum de la vraisemblance est atteint en $\theta = x_{(n)}$ puisque avant (strictement) ce point la vraisemblance est nulle, qu'en ce point elle prend la valeur

$$\mathcal{L}(x_1, \dots, x_n; x_{(n)}) = \frac{1}{(x_{(n)})^n}$$

et qu'après elle est décroissante. Ainsi l'estimateur du maximum de vraisemblance est

$$\hat{\theta}_n = X_{(n)}.$$

Exercice 5 (Modèles de la loi exponentielle et de la loi de Poisson en Fiabilité)

Partie 1

On s'intéresse à la durée de vie X d'un matériel électronique. Il est raisonnable de considérer que cette durée de vie est aléatoire et que sa loi soit exponentielle (il existe des méthodes statistiques, mais que nous ne verrons pas dans le cadre de ce cours, pour vérifier cette hypothèse). En revanche, on ignore la valeur du paramètre λ de cette loi.

1) Écrire le modèle statistique engendré par X . Donner également le modèle d'échantillonnage associé.

2) Donner l'estimateur du maximum de vraisemblance pour une observation x_1, \dots, x_n d'un échantillon X_1, \dots, X_n de durées de vie de ces matériels.

3) Donner une estimation par maximum de vraisemblance de la quantité $\alpha = P(X > t_0)$, où t_0 est un temps fixé.

4) Quels estimateurs de λ et de α obtient-on si on utilise la méthode des moments ?

Partie 2

Supposons maintenant que les observations de ces durées de vie soient obtenues grâce à l'expérience suivante. Au temps $t = 0$, on dispose un matériel sur un banc d'essai. Quand celui-ci tombe en panne, on remplace immédiatement (ou on ne compte pas le temps de remplacement) le matériel défectueux par un matériel identique mais neuf. Et ainsi de suite jusqu'au temps t_0 . On note alors K le nombre de pannes relevées dans l'intervalle $[0, t_0]$.

5) Calculer la probabilité que K soit nul.

6) On note T_k le temps écoulé jusqu'à la k ème panne observée. C'est à dire que $T_k = X_1 + \dots + X_k$. Montrer que la loi de la v.a.r. T_k est une Gamma $G(k, \lambda)$ (Ind. On pourra utiliser la transformée de Laplace ou la fonction caractéristique).

7) Exprimer l'événement $K = k$ en fonction d'événements liant les v.a.r. T_k et X_{k+1} . En déduire que la loi de K est une loi de Poisson, dont on déterminera la valeur du paramètre.

Partie 3

On suppose que l'on réalise n fois cette expérience et on note K_1, \dots, K_n les nombres de pannes observées dans chaque intervalle $[0, t_0]$.

8) Donner le modèle statistique associé à ces observations.

9) Donner par la méthode du maximum de vraisemblance un autre estimateur du paramètre λ , basé cette fois sur les observations k_1, \dots, k_n .

10) Qu'obtient-on comme estimateur de λ si, dans ce modèle, on utilise la méthode des moments ?

Solution

1) La durée de vie du matériel est modélisée par une v.a.r. X , supposée aléatoire de loi $\mathcal{E}(\lambda)$, où λ est inconnu. Ainsi X engendre le modèle paramétrique

$$(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}(\lambda); \lambda > 0\}) = \{\mathcal{E}(\lambda); \lambda > 0\} = \{f_\lambda(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x); \lambda > 0\}.$$

Le modèle d'échantillonnage associé est

$$(\mathbb{R}_+^n, \mathcal{B}_{\mathbb{R}_+^n}, \{\mathcal{E}^{\otimes n}(\lambda); \lambda > 0\})$$

et la densité de l'échantillon X_1, \dots, X_n est :

$$f_\lambda(x_1, \dots, x_n) = \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\} \mathbb{1}_{\mathbb{R}_+^n}(x_1, \dots, x_n).$$

2) D'après ce qui précède, la vraisemblance de l'échantillon observé est x_1, \dots, x_n est

$$\mathcal{L}(x_1, \dots, x_n; \lambda) = \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\} \mathbb{1}_{\mathbb{R}_+^n}(x_1, \dots, x_n)$$

et la log-vraisemblance

$$\ell(x_1, \dots, x_n; \lambda) = \ln \mathcal{L}(x_1, \dots, x_n; \lambda) = n \ln \lambda - \lambda \sum_{i=1}^n x_i + \ln \mathbb{1}_{\mathbb{R}_+^n}(x_1, \dots, x_n).$$

On a :

$$\frac{\partial}{\partial \lambda} \ell(x_1, \dots, x_n; \lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i$$

et

$$\frac{\partial^2}{\partial \lambda^2} \ell(x_1, \dots, x_n; \lambda) = -\frac{n}{\lambda^2} < 0.$$

Ainsi, la log-vraisemblance est concave et son maximum est atteint en la valeur qui annule la dérivée première, c'est à dire en λ tel que :

$$\frac{n}{\lambda} - \sum_{i=1}^n x_i = 0.$$

L'estimateur du maximum de vraisemblance est donc :

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}}.$$

3) On a

$$P(X > t_0) = e^{-\lambda t_0}.$$

On cherche donc à estimer $\alpha = g(\lambda)$, où $g(x) = e^{-t_0 x}$. D'après le cours, l'estimateur par maximum de vraisemblance de α est donné par

$$g(\hat{\lambda}) = e^{-\hat{\lambda} t_0} = \exp \left\{ -\frac{t_0}{\bar{X}} \right\}.$$

4) La relation $\mathbb{E}(X) = 1/\lambda$ suggère d'estimer λ par

$$\hat{\lambda} = \frac{1}{\bar{X}}.$$

On retrouve le même estimateur que par maximisation de la vraisemblance.

Par ailleurs on peut écrire :

$$\alpha = P(X > t_0) = \mathbb{E}(\mathbb{1}_{\{X > t_0\}}).$$

La méthode des moments suggère d'estimer α par

$$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i > t_0\}},$$

qui est un estimateur différent de celui obtenu par la méthode du maximum de vraisemblance.

5) L'événement $\{K = 0\}$ revient à dire que la première panne est intervenue après le temps t_0 . On peut donc écrire :

$$P(K = 0) = P(X > t_0) = \exp(-\lambda t_0).$$

6) Calculons la transformée de Laplace de la loi Gamma de paramètres α et β , de densité

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\mathbb{R}^+}(x).$$

Sa transformée de Laplace, quand elle existe, est :

$$\mathcal{L}_{G(\alpha,\beta)}(s) = \mathbb{E}(e^{sX}) = \int_{\mathbb{R}^+} e^{sx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\mathbb{R}^+}(x).$$

Cette intégrale converge si, et seulement si, $s < \beta$. Calculons cette intégrale, pour $s < \beta$. On a, en faisant le changement de variable $u = (\beta - s)x$:

$$\mathcal{L}_{G(\alpha,\beta)}(s) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_{\mathbb{R}^+} \frac{u^{\alpha-1}}{(\beta - s)^{\alpha-1}} e^{-u} \frac{du}{\beta - s} = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta - s)^\alpha} = \frac{1}{\left(1 - \frac{s}{\beta}\right)^\alpha},$$

pour $s < \beta$.

La loi exponentielle étant une loi Gamma particulière de paramètres 1 et λ , sa transformée de Laplace est

$$\mathcal{L}_{\mathcal{E}(\lambda)}(s) = \frac{1}{\left(1 - \frac{s}{\lambda}\right)}.$$

Calculons la transformée de Laplace de la v.a. $T_k = X_1 + \dots + X_k$, modélisant le temps écoulé jusqu'à obtenir k pannes. Puisque les v.a.r. X_1, \dots, X_n sont indépendantes et de même loi exponentielle de paramètre λ , on a :

$$\mathcal{L}_{T_k}(s) = \mathbb{E}(e^{s(X_1 + \dots + X_k)}) = \prod_{i=1}^k \mathbb{E}(e^{sX_i}) = \prod_{i=1}^k \frac{1}{\left(1 - \frac{s}{\lambda}\right)} = \frac{1}{\left(1 - \frac{s}{\lambda}\right)^k}.$$

On reconnaît la transformée de Laplace d'une loi $G(k, \lambda)$. Par la propriété de caractérisation de la loi par la transformée de Laplace, on en déduit que cette dernière est donc la loi de la v.a. T_k .

7) On a

$$\{K = k\} = \{T_k \leq t_0 < T_{k+1}\} = \{T_k \leq t_0 < T_k + X_{k+1}\}.$$

Ainsi,

$$\begin{aligned} P(K = k) &= P(T_k \leq t_0 < T_k + X_{k+1}) = \iint_{u \leq t_0 < u+v} f_{T_k, X_{k+1}}(u, v) du dv \\ &= \iint_{u \leq t_0 < u+v} \frac{\lambda^k}{\Gamma(k)} u^{k-1} e^{-\lambda u} \lambda e^{-\lambda v} du dv = \int_{u \leq t_0} \frac{\lambda^k}{\Gamma(k)} u^{k-1} e^{-\lambda u} \left(\int_{t_0-u}^{+\infty} \lambda e^{-\lambda v} dv \right) du \\ &= \int_{u \leq t_0} \frac{\lambda^k}{\Gamma(k)} u^{k-1} e^{-\lambda u} e^{-\lambda(t_0-u)} du = \frac{\lambda^k}{\Gamma(k)} e^{-\lambda t_0} \int_0^{t_0} u^{k-1} du = \frac{(\lambda t_0)^k}{k!} e^{-\lambda t_0}. \end{aligned}$$

On a vu dans la question 5) que cette formule est également vraie pour $k = 0$. Ainsi la loi de la v.a.r. K est une Poisson de paramètre λt_0 .

8) Le modèle statistique associé à nos observations est le modèle d'échantillonnage tiré du modèle

$$(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathcal{P}(\lambda t_0); \lambda > 0\}).$$

Le modèle d'échantillonnage est précisément :

$$(\mathbb{N}^n, \mathcal{P}(\mathbb{N}^n), \{\mathcal{P}^{\otimes n}(\lambda t_0); \lambda > 0\}).$$

9) La vraisemblance des observations (k_1, \dots, k_n) est :

$$\begin{aligned} \mathcal{L}(k_1, \dots, k_n; \lambda) &= \prod_{i=1}^n \frac{(\lambda t_0)^{k_i} e^{-\lambda t_0}}{k_i!} \\ &= \frac{(\lambda t_0)^{\sum_{i=1}^n k_i} e^{-n\lambda t_0}}{\prod_{i=1}^n k_i!}. \end{aligned}$$

La log-vraisemblance est :

$$\ell(k_1, \dots, k_n; \lambda) = \left(\sum_{i=1}^n k_i \right) \ln(\lambda t_0) - n\lambda t_0 - \ln \left(\prod_{i=1}^n k_i! \right).$$

Sa dérivée première est :

$$\frac{\partial}{\partial \lambda} \ell(\lambda; k_1, \dots, k_n) = \frac{(\sum_{i=1}^n k_i) t_0}{\lambda t_0} - n t_0$$

et sa dérivée seconde

$$\frac{\partial^2}{\partial \lambda^2} \ell(\lambda; k_1, \dots, k_n) = -\frac{\sum_{i=1}^n k_i}{\lambda^2},$$

qui est clairement négative. La fonction log-vraisemblance est donc concave et son maximum atteint en la valeur qui annule la dérivée première. L'estimateur du maximum de vraisemblance est donc :

$$\hat{\lambda} = \frac{\sum_{i=1}^n K_i}{n t_0}.$$

10) Effectuons le calcul de l'espérance d'une loi de Poisson de paramètre λ , même si ce résultat est connu

$$\mathbb{E}(X) = \sum_{k=0}^{+\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Dans notre cas, on a donc $\mathbb{E}(X) = \lambda t_0$. Ainsi un estimateur par la méthode des moments de λ est :

$$\hat{\lambda} = \frac{1}{t_0} \bar{K} = \frac{\sum_{k=1}^n K_i}{n t_0},$$

qui est le même que celui obtenu par la méthode du maximum de vraisemblance.

Exercice 6 (*Maximum de vraisemblance*)

Pour les modèles suivants, donner l'estimateur du maximum de vraisemblance associé à l'observation d'un échantillon X_1, \dots, X_n .

1) Modèle de la loi exponentielle décalée :

$$(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}_{t_0}(\lambda) : \lambda > 0, t_0 \in \mathbb{R}\}).$$

On rappelle que la densité de la loi exponentielle décalée $\mathcal{E}_{t_0}(\lambda)$ est :

$$f_{\lambda, t_0}(x) = \lambda \exp(-\lambda(x - t_0)) \mathbb{1}_{[t_0, +\infty[}(x).$$

2) Modèle de la loi Bêta à un seul paramètre :

$$(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\text{Beta}(1, \theta) : \theta > 1\}).$$

On rappelle que la densité de la loi Beta(a, b) est :

$$f_{a,b}(x) = \frac{1}{\beta(a,b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{[0,1]}(x),$$

où $\beta(a, b)$ est la valeur de la fonction Eulérienne Bêta prise en a et b .

Ind. On pourra montrer en premier lieu que la densité pour le modèle considéré est :

$$f_{\theta}(x) = \theta(1-x)^{\theta-1} \mathbb{1}_{[0,1]}(x).$$

Solution

1) La vraisemblance de l'échantillon observé x_1, \dots, x_n est donnée par :

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_n; \lambda, t_0) &= \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n (x_i - t_0) \right\} \mathbb{1}_{[t_0, +\infty[^n}(x_1, \dots, x_n) \\ &= \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n (x_i - t_0) \right\} \mathbb{1}_{[t_0, +\infty[}(x_{(1)}), \end{aligned}$$

où $x_{(1)} = \inf_{i=1, \dots, n} x_i$.

La fonction $t_0 \mapsto \mathcal{L}(x_1, \dots, x_n; \lambda, t_0)$ n'est pas continue, et donc pas dérivable, en $x_{(1)}$. On ne peut donc pas utiliser le critère de concavité. Cependant, en remarquant que cette fonction est nulle sur $]x_{(1)}, +\infty[$ et croissante sur l'intervalle $]-\infty, x_{(1)}]$ (puisque elle est d'expression $C \cdot e^{n\lambda t_0}$ sur cet intervalle), le maximum est atteint en $x_{(1)}$ et ceci pour n'importe quelle valeur de λ .

Il nous reste maintenant à maximiser la log-vraisemblance en fonction de λ . La fonction

$$\lambda \mapsto \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\} \exp \{ n\lambda x_{(1)} \}$$

est dérivable et concave (on le vérifiera plus loin) sur \mathbb{R}^+ . Elle atteint donc son maximum en la valeur qui annule la dérivée première par rapport à λ ou pour simplifier les calculs la dérivée de son logarithme. On a

$$\ell(\lambda, t_0; x_1, \dots, x_n) = \ln \mathcal{L}(\lambda, t_0; x_1, \dots, x_n) = n \ln \lambda - \lambda \sum_{i=1}^n x_i + n\lambda x_{(1)}.$$

Il vient :

$$\frac{\partial}{\partial \lambda} \ell(\lambda, t_0; x_1, \dots, x_n) = \frac{n}{\lambda} - \left(\sum_{i=1}^n x_i - nX_{(1)} \right).$$

La dérivée seconde est alors $-n/\lambda^2$ qui est bien négative et donc la fonction concave. L'estimateur du maximum de vraisemblance est alors :

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i - nX_{(1)}}$$

En résumé, pour ce modèle l'estimateur du maximum de vraisemblance de $\theta = (\lambda, t_0)$, est :

$$\hat{\theta} = (\hat{\lambda}, \hat{t}_0) = \left(\frac{n}{\sum_{i=1}^n X_i - nX_{(1)}}, X_{(1)} \right)$$

2) La densité d'une loi Beta(1, θ) est

$$f_{\theta}(x) = \frac{1}{\beta(1, \theta)} (1-x)^{\theta-1} \mathbb{1}_{[0,1]}(x) = \theta(1-x)^{\theta-1} \mathbb{1}_{[0,1]}(x)$$

puisque

$$\beta(1, \theta) = \frac{\Gamma(1)\Gamma(\theta)}{\Gamma(\theta+1)} = \frac{\Gamma(1)\Gamma(\theta)}{\theta\Gamma(\theta)} = \frac{1}{\theta}.$$

La vraisemblance de l'échantillon x_1, \dots, x_n est alors :

$$\mathcal{L}(x_1, \dots, x_n; \theta) = \theta^n \prod_{i=1}^n (1-x_i)^{\theta-1} \mathbb{1}_{[0,1]}(x_i)$$

et la log-vraisemblance

$$\ell(x_1, \dots, x_n; \theta) = \ln \mathcal{L}(x_1, \dots, x_n; \theta) = n \ln \theta + \sum_{i=1}^n (\theta-1) \ln(1-x_i) + \ln C,$$

où C est une constante.

On en déduit les dérivées partielles

$$\begin{aligned} \frac{\partial}{\partial \theta} \ell(x_1, \dots, x_n; \theta) &= \frac{n}{\theta} + \sum_{i=1}^n \ln(1-x_i) = 0 \\ \frac{\partial^2}{\partial \theta^2} \ell(x_1, \dots, x_n; \theta) &= -\frac{n}{\theta^2}. \end{aligned}$$

La log-vraisemblance est donc concave (puisque de dérivée seconde négative) et son maximum est donc atteint en

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^n \ln(1-X_i)}.$$

Mais rappelons ici que l'espace des paramètres pour θ est $]1, +\infty[$. Or si $\hat{\theta} \leq 1$, la fonction

$$\begin{aligned} L(x_1, \dots, x_n; \cdot) :]1, +\infty[&\rightarrow \mathbb{R}^+ \\ \theta &\mapsto L(x_1, \dots, x; \theta) \end{aligned}$$

atteint son maximum en $\theta = 1$. Ainsi l'estimateur du maximum de vraisemblance pour ce modèle est :

$$\tilde{\theta} = \max(1, \hat{\theta}) = \max\left(1, -\frac{n}{\sum_{i=1}^n \ln(1 - X_i)}\right).$$

Correction des exercices du Chapitre 5

Exercice 1 (*Qualité des estimateurs dans les modèles de Poisson et de la loi exponentielle*)

On considère deux modèles :

- celui de la loi de Poisson $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathcal{P}(\lambda) : \lambda > 0\})$, où $\mathcal{P}(\lambda)$ désigne la loi de Poisson de paramètre λ ;
- celui de la loi de exponentielle $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}(\lambda) : \lambda > 0\})$, où $\mathcal{E}(\lambda)$ désigne la loi exponentielle de paramètre λ .

On a vu que ces modèles sont en particulier utiles pour modéliser des problèmes de Fiabilité.

Pour chacun de ces modèles, répondre à l'ensemble des questions suivantes. On considérera à chaque fois l'observation d'un échantillon X_1, \dots, X_n .

1) Rappeler l'expression de l'estimateur du maximum de vraisemblance dans ce modèle (on a vu qu'il est également estimateur par la méthode des moments).

2) Étudier la consistance, le biais et le risque quadratique de cet estimateur.

3) Si cet estimateur est biaisé, est-il asymptotiquement sans biais ? Donner, si nécessaire, un estimateur sans biais. L'estimateur sans biais (l'initial ou le second) est-il efficace ? Est-il consistant ?

Solution

Modèle de Poisson

1) On a vu que l'estimateur du maximum de vraisemblance est :

$$\hat{\lambda}_n = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

2) Par la loi forte des grands nombres, on a *p.s.*

$$\hat{\lambda}_n = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mathbb{E}_\lambda(X) = \lambda, \text{ quand } n \rightarrow +\infty,$$

et $\hat{\lambda}_n$ est donc un estimateur consistant.

De plus, on a :

$$\mathbb{E}_\lambda(\hat{\lambda}_n) = \mathbb{E}_\lambda(X) = \lambda$$

et cet estimateur est également sans biais.

Calculons maintenant son risque quadratique. Comme l'estimateur est sans biais, on a :

$$R(\hat{\lambda}_n, \lambda) = \text{Var}_\lambda(\bar{X}_n) = \frac{1}{n} \text{Var}_\lambda(X)$$

On sait² que la variance d'une loi exponentielle de paramètre λ est λ . Ainsi, on a

$$R(\hat{\lambda}_n, \lambda) = \mathbb{E}_\lambda \left(\hat{\lambda}_n - \lambda \right)^2 = \frac{\lambda}{n}.$$

Remarquons au passage que $R(\hat{\lambda}_n, \lambda) \rightarrow 0$, quand $n \rightarrow +\infty$, et on a donc aussi la L^2 -convergence de $\hat{\lambda}_n$.

3) On a vu que

$$\frac{\partial^2}{\partial \lambda^2} \ell(\lambda; x_1, \dots, x_n) = -\frac{1}{\lambda^2} \sum_{i=1}^n x_i,$$

où $\ell(\lambda; x_1, \dots, x_n)$ est la log-vraisemblance du modèle de Poisson. On en déduit l'information de Fisher :

$$I_n(\lambda) = \mathbb{E}_\lambda \left(\frac{1}{\lambda^2} \sum_{i=1}^n X_i \right) = \frac{1}{\lambda^2} \sum_{i=1}^n \mathbb{E}_\lambda(X_i) = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}$$

Ainsi, la borne de Cramer-Rao est

$$BCR = \frac{\lambda}{n}.$$

Comme elle est égale à $R(\hat{\lambda}_n, \lambda)$, l'estimateur du maximum de vraisemblance est efficace.

Modèle de la loi exponentielle

1) On a vu que

$$\hat{\lambda}_n = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}_n}$$

est l'estimateur du maximum de vraisemblance.

2) Par la loi forte des grands nombres, on a :

$$\bar{X}_n \xrightarrow{\text{p.s.}} \mathbb{E}_\lambda(X) = \frac{1}{\lambda}, \text{ quand } n \rightarrow +\infty,$$

ce qui implique que

$$\hat{\lambda}_n \xrightarrow{\text{p.s.}} \lambda$$

et l'estimateur $\hat{\lambda}_n$ est donc consistant.

²Si on ne sait pas, on peut le retrouver en effectuant le calcul

$$\mathbb{E}(X(X-1)) = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2.$$

D'où

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Pour déterminer le biais de l'estimateur, rappelons que si X_1, \dots, X_n sont des v.a.r. indépendantes et de même loi exponentielle de paramètre λ , alors $\sum_{i=1}^n X_i$ est de loi $\text{Gamma}(n, \lambda)$. Ainsi,

$$\begin{aligned} \mathbb{E}_\lambda(\hat{\lambda}_n) &= \mathbb{E}_\lambda \left(\frac{n}{\sum_{i=1}^n X_i} \right) = \int_0^{+\infty} \frac{n}{x} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx = \frac{n\lambda^n}{\Gamma(n)} \int_0^{+\infty} x^{n-2} e^{-\lambda x} dx \\ &= \frac{n\lambda^n}{\Gamma(n)} \int_0^{+\infty} \left(\frac{y}{\lambda}\right)^{n-2} e^{-y} \frac{dy}{\lambda} = \frac{n\lambda^n}{\Gamma(n)\lambda^{n-1}} \int_0^{+\infty} y^{n-2} e^{-y} dy \\ &= \frac{n\lambda\Gamma(n-1)}{(n-1)\Gamma(n-1)} = \frac{n}{n-1}\lambda, \end{aligned}$$

où la quatrième égalité est obtenue par le changement de variable $y = \lambda x$. L'estimateur $\hat{\lambda}_n$ est donc un estimateur biaisé.

Évaluons maintenant le risque quadratique de notre estimateur. On a :

$$R(\hat{\lambda}_n, \lambda) = \text{Var}_\lambda(\hat{\lambda}_n) + b_{\hat{\lambda}_n}^2(\lambda)$$

où $b_{\hat{\lambda}_n}(\lambda)$ est le biais de notre estimateur. Le calcul précédent nous donne

$$b_{\hat{\lambda}_n}(\lambda) = \mathbb{E}_\lambda(\hat{\lambda}_n) - \lambda = \lambda \left(\frac{n}{n-1} - 1 \right) = \frac{\lambda}{n-1}.$$

De plus, la variance de $\hat{\lambda}_n$ est donnée par :

$$\text{Var}_\lambda(\hat{\lambda}_n) = \mathbb{E}_\lambda(\hat{\lambda}_n^2) - \left(\mathbb{E}_\lambda(\hat{\lambda}_n) \right)^2$$

On a :

$$\begin{aligned} \mathbb{E}_\lambda(\hat{\lambda}_n^2) &= \int_0^{+\infty} \frac{n^2}{x^2} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx = \frac{n^2\lambda^n}{\Gamma(n)} \int_0^{+\infty} x^{n-3} e^{-\lambda x} dx \\ &= \frac{n^2\lambda^n}{\Gamma(n)} \int_0^{+\infty} \left(\frac{y}{\lambda}\right)^{n-3} e^{-y} \frac{dy}{\lambda} = \frac{n^2\lambda^n}{\Gamma(n)\lambda^{n-2}} \int_0^{+\infty} y^{n-3} e^{-y} dy \\ &= \frac{n^2\lambda^n\Gamma(n-2)}{\lambda^{n-2}(n-1)(n-2)\Gamma(n-2)} = \frac{n^2\lambda^2}{(n-1)(n-2)}, \end{aligned}$$

où la troisième égalité est ici aussi obtenue après le changement de variable $y = \lambda x$. D'où :

$$\text{Var}_\lambda(\hat{\lambda}_n) = \frac{n^2\lambda^2}{(n-1)(n-2)} - \frac{n^2\lambda^2}{(n-1)^2}.$$

et

$$\begin{aligned} R(\hat{\lambda}_n, \lambda) &= \frac{n^2\lambda^2}{(n-1)(n-2)} - \frac{n^2\lambda^2}{(n-1)^2} + \frac{\lambda^2}{(n-1)^2} \\ &= \frac{\lambda^2}{(n-1)^2(n-2)} (n^2(n-1) - n^2(n-2) + (n-2)) \\ &= \frac{\lambda^2}{(n-1)^2(n-2)} (n^2 + n - 2) \end{aligned}$$

3) Nous avons vu que cet estimateur est biaisé. Il apparaît clairement asymptotiquement sans biais.

Un estimateur sans biais de λ est donné par :

$$\tilde{\lambda}_n = \frac{n-1}{n} \hat{\lambda}_n = \frac{n-1}{n} \frac{n}{\sum_{i=1}^n X_i} = \frac{n-1}{\sum_{i=1}^n X_i}.$$

Son risque quadratique est donné par :

$$\begin{aligned} R(\tilde{\lambda}_n, \lambda) &= \text{Var}_\lambda(\tilde{\lambda}_n) \\ &= \text{Var}_\lambda\left(\frac{n-1}{n} \hat{\lambda}_n\right) \\ &= \left(\frac{n-1}{n}\right)^2 \text{Var}_\lambda(\hat{\lambda}_n) \\ &= \left(\frac{n-1}{n}\right)^2 \lambda^2 \left(\frac{(n-1)n^2 - n^2(n-2)}{(n-1)^2(n-2)}\right) \\ &= \lambda^2 \left(\frac{n-1-n+2}{n-2}\right) \\ &= \frac{\lambda^2}{n-2} \end{aligned}$$

Par ailleurs, on a vu que :

$$\frac{\partial^2}{\partial \lambda^2} \ell(\lambda; x_1, \dots, x_n) = -\frac{n}{\lambda^2}.$$

L'information de Fisher est donc :

$$I_n(\lambda) = -\mathbb{E}_\lambda \left(\frac{\partial^2}{\partial \lambda^2} \ell(\lambda; X_1, \dots, X_n) \right) = \frac{n}{\lambda^2}.$$

La borne de Cramer-Rao est donc :

$$BCR = \frac{\lambda^2}{n}.$$

Puisque

$$R(\tilde{\lambda}_n, \lambda) = \frac{\lambda^2}{n-2} > \frac{\lambda^2}{n} = BCR,$$

notre estimateur $\tilde{\lambda}_n$ n'atteint pas la borne de Cramer-Rao et n'est donc pas efficace. En revanche comme on a :

$$\frac{BCR}{R(\tilde{\lambda}_n, \lambda)} = \frac{1}{I_n(\lambda) \text{Var}_\lambda(\tilde{\lambda}_n)} = \frac{\lambda^2 n - 2}{n \lambda^2} = \frac{n-2}{n} \rightarrow 1, \text{ quand } n \rightarrow +\infty,$$

il est asymptotiquement efficace.

Pour la consistance, on sait que

$$\hat{\lambda}_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \lambda$$

et donc

$$\tilde{\lambda}_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \lambda,$$

c'est à dire que $\tilde{\lambda}_n$ est fortement consistant.

Exercice 2 (*Fiabilité et fonction de répartition empirique*)

Un matériel a une durée de vie modélisée par une v.a. X de f.d.r. F . Un étudiant en Licence de Mathématiques sait qu'il devra l'utiliser pendant un temps x_0 . Il souhaite naturellement qu'il n'y ait pas de panne durant cette période.

Cet étudiant, ayant suivi le module de Statistique Inférentielle, cherche en premier lieu à estimer la loi (en fait la f.d.r.) de cette durée de vie, c'est à dire estimer $F(x)$ pour tout x de \mathbb{R}^+ . Il a alors l'idée de faire fonctionner, sur banc d'essai, n machines identiques à celle qu'il utilisera dans l'avenir. Il note x_1, \dots, x_n les n temps de panne observés, qui sont donc les réalisations des v.a. X_1, \dots, X_n i.i.d. de même loi que X .

1) Par la méthode des moments il propose un estimateur de $F(x)$, pour tout x dans \mathbb{R}^+ . Pouvez-vous en faire autant ?

2) Son estimateur est-il consistant ? Que dire de son biais et de son risque quadratique ?

3) Se souvenant de ses cours, il sait que, pour être précis, il aurait dû, au préalable, introduire un modèle paramétrique. Quel(s) modèle(s) pourrait-il proposer ? Que sont les observations sous ce(s) modèle(s) ? Une estimation par maximum de vraisemblance nous donnerait-elle quelque chose de différent dans ce modèle ?

4) Que dire alors de l'efficacité de l'estimateur proposé dans la première question ?

Solution

1) On a

$$F(x) = P(X \leq x) = \mathbb{E}(\mathbb{1}_{\{X \leq x\}}).$$

On estime donc $F(x)$ par :

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}$$

2) Par la loi forte des grands nombres, on a pour tout x dans \mathbb{R}^+ :

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \xrightarrow{\text{p.s.}} F(x) = \mathbb{E}(\mathbb{1}_{\{X \leq x\}}),$$

quand $n \rightarrow +\infty$ et $\hat{F}(x)$ est donc un estimateur consistant.

Par ailleurs,

$$\mathbb{E}(\hat{F}(x)) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(\mathbb{1}_{\{X_i \leq x\}}) = F(x)$$

et $\hat{F}(x)$ est donc un estimateur sans biais.

Comme l'estimateur est sans biais son risque quadratique est égal à sa variance et il est donc égal à :

$$\begin{aligned} R(\hat{F}(x), F(x)) &= \text{Var} \left(\hat{F}(x) \right) = \frac{1}{n^2} \text{Var} \left(\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var} \left(\mathbb{1}_{\{X_i \leq x\}} \right) \\ &= \frac{1}{n^2} \sum_{i=1}^n F(x) \bar{F}(x) = \frac{F(x) \bar{F}(x)}{n}, \end{aligned}$$

où $\bar{F} = 1 - F$, la troisième égalité est justifiée par indépendance des v.a.r. X_1, \dots, X_n et l'avant dernière égalité est obtenue en remarquant, par exemple, que $Y_i = \mathbb{1}_{\{X_i \leq x\}}$ est une v.a.r. de loi de Bernoulli de paramètre $F(x)$. Sa variance est donc $F(x)(1 - F(x))$.

3) Soit x fixé. On veut estimer $F(x)$ qui est une probabilité, celle de l'événement $\{X \leq x\}$. En reprenant la v.a.r. $Y = \mathbb{1}_{\{X \leq x\}}$ introduit dans la question précédente, on peut voir également le problème comme l'estimation du paramètre $F(x)$ de la loi de Y . Il s'agit donc d'un modèle de Bernoulli :

$$(\{0, 1\}, \mathcal{P}(\{0, 1\}), \{B(F(x)) : F(x) \in [0, 1]\}).$$

4) On a vu dans le cours que la dérivée seconde, par rapport à p , de la log-vraisemblance $\ell(y_1, \dots, y_n; p)$ dans un modèle de Bernoulli est donnée par

$$\frac{\partial^2}{\partial p^2} \ell(y_1, \dots, y_n; p) = -\frac{1}{p^2} \sum_{i=1}^n y_i - \frac{n - \sum_{i=1}^n y_i}{(1-p)^2}$$

L'information de Fisher est donc donnée par :

$$\begin{aligned} I_n(p) &= -\mathbb{E}_p \left(\frac{\partial^2}{\partial p^2} \ell(Y_1, \dots, Y_n; p) \right) \\ &= \frac{1}{p^2} \sum_{i=1}^n \mathbb{E}(y_i) + \frac{n - \sum_{i=1}^n \mathbb{E}(Y_i)}{(1-p)^2} \\ &= \frac{np}{p^2} + \frac{n - np}{(1-p)^2} \\ &= \frac{n}{p(1-p)}. \end{aligned}$$

Ainsi la borne de Cramer-Rao est :

$$BCR = \frac{p(1-p)}{n}.$$

Revenant à notre problématique de l'estimation de $F(x)$, le risque quadratique de notre estimateur s'obtient en prenant $p = F(x)$ et est alors égal à :

$$R(\hat{F}(x), F(x)) = \frac{F(x) \bar{F}(x)}{n} = BCR.$$

L'estimateur $\hat{F}(x)$ atteint donc la borne de Cramer-Rao et est par conséquent un estimateur efficace.

Exercice 3 (*L'Agriculteur et la Statistique*)

Un agriculteur possède un champ carré dont il veut estimer la superficie. Quand il mesure un côté de son champ, il sait (un statisticien de passage lui a confirmé), ou il suppose, que l'erreur expérimentale de la mesure est une variable aléatoire de loi normale centrée et de variance σ^2 . Il réalise une première mesure de ce côté et trouve une valeur $x_1 = 510$ mètres. Il en déduit une superficie de $s_1 = 26.01$ hectares. Il réalise une deuxième mesure et trouve alors $x_2 = 490$, d'où une valeur de la superficie $s_2 = 24.01$. Il abandonne ses mesures et réfléchit pour savoir quelle est la bonne façon de procéder. Doit-il adopter comme estimation de la surface s_1 , s_2 , ou une estimation combinant les deux mesures, telle que :

$$\begin{aligned} s_3 &= x_1 x_2 = 24.99, \\ s_4 &= \frac{s_1 + s_2}{2} = 25.01, \\ s_5 &= \left(\frac{x_1 + x_2}{2} \right)^2 = 25 ? \end{aligned}$$

Faut-il recommencer ses mesures jusqu'à ce qu'il trouve deux résultats identiques, ou bien combiner intelligemment n mesures pour construire des estimations du type s_4 ou s_5 (généralisées à ces n mesures) ?

1) On se propose d'aider l'agriculteur à résoudre son problème. Préciser le modèle considéré ainsi que la fonction $q(\theta)$ que l'on cherche à estimer. Étudier les cinq estimateurs proposés. On calculera notamment leurs biais, variances et risques quadratiques moyens. (Ind. si $X \sim \mathcal{N}(m, \sigma^2)$ alors $\text{Var}(X^2) = 2(\sigma^4 + 2m^2\sigma^2)$).

A l'aide de ces calculs, aider l'agriculteur à choisir l'estimateur qui vous semble préférable aux autres.

2) Donner les estimateurs qui généralisent s_4 et s_5 au cas où l'agriculteur a pu faire n mesures du côté de son champ. Effectuer la même étude qu'à la question 1) pour ces estimateurs. Étudier également leurs consistances. Que dire de leur L^2 -consistance ? Conclure.

3) Donner l'estimateur du maximum de vraisemblance et l'étudier s'il est différent de ceux considérés précédemment.

Solution

1) La vraie longueur (inconnue) d'un côté est μ . L'erreur expérimentale étant distribuée suivant une loi $N(0, \sigma^2)$, le modèle correspondant aux mesures effectuées par l'agriculteur est donc le modèle paramétrique gaussien de variance σ^2 connue, i.e.

$$(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{N(\mu, \sigma^2) : \mu > 0\}).$$

Notons que le paramètre μ est supposé positif car il s'agit d'une longueur !

Notre problème est donc d'estimer la surface du champ, c'est à dire $g(\mu) = \mu^2$.

* Étudions le premier estimateur $S_1 = X_1^2$. On a

$$\mathbb{E}_\mu(S_1) = \mathbb{E}_\mu X_1^2 = \text{Var}_\mu(X_1) + \mathbb{E}_\mu^2 X_1 = \sigma^2 + \mu^2.$$

L'estimateur S_1 est donc biaisé et son biais égal à σ^2 . Sa variance est

$$\text{Var}_\mu(S_1) = \text{Var}_\mu X_1^2 = 2(\sigma^4 + 2\mu^2\sigma^2).$$

Ainsi le risque quadratique de cet estimateur est :

$$R(S_1, \mu) = 2(\sigma^4 + 2\mu^2\sigma^2) + \sigma^4 = 3\sigma^4 + 4\mu^2\sigma^2.$$

* Pour le second estimateur on a bien sûr la même chose puisque les v.a.r. X_1 et X_2 ont même loi.

* Étudions maintenant le troisième estimateur $S_3 = X_1X_2$. Par indépendance, on a :

$$\mathbb{E}_\mu S_3 = \mathbb{E}_\mu(X_1X_2) = \mathbb{E}_\mu X_1 \mathbb{E}_\mu X_2 = \mu^2.$$

Cet estimateur est donc sans biais. Sa variance est :

$$\begin{aligned} \text{Var}_\mu(S_3) &= \text{Var}_\mu(X_1X_2) = \mathbb{E}_\mu(X_1^2X_2^2) - \mathbb{E}_\mu^2(X_1X_2) = \mathbb{E}_\mu X_1^2 \mathbb{E}_\mu X_2^2 - \mu^4 \\ &= (\sigma^2 + \mu^2)^2 - \mu^4 = \sigma^4 + 2\mu^2\sigma^2. \end{aligned}$$

Son risque quadratique est alors :

$$R(S_3, \mu) = \sigma^4 + 2\mu^2\sigma^2.$$

Cet estimateur S_3 est donc déjà meilleur que S_1 et S_2 en terme de biais et de risque quadratique.

* Considérons le quatrième estimateur :

$$S_4 = \frac{S_1 + S_2}{2} = \frac{X_1^2 + X_2^2}{2}.$$

On a

$$\mathbb{E}_\mu S_4 = \frac{1}{2}(\mathbb{E}_\mu X_1^2 + \mathbb{E}_\mu X_2^2) = \sigma^2 + \mu^2.$$

Cet estimateur est donc biaisé de biais σ^2 . Grâce à l'indépendance sa variance peut s'écrire :

$$\text{Var}_\mu S_4 = \frac{1}{4} \text{Var}(X_1^2 + X_2^2) = \frac{1}{4}(\text{Var}(X_1^2) + \text{Var}(X_2^2)) = \sigma^4 + 2\mu^2\sigma^2.$$

On en déduit son biais :

$$R(S_4, \mu) = \sigma^4 + 2\mu^2\sigma^2 + \sigma^4 = 2\sigma^4 + 2\mu^2\sigma^2.$$

Cet estimateur est donc moins bon que S_3 qui est sans biais et de risque inférieur, qui lui est donc préférable.

* Le cinquième estimateur est défini par :

$$S_5 = \left(\frac{X_1 + X_2}{2} \right)^2.$$

Son espérance est :

$$\begin{aligned}\mathbb{E}_\mu(S_5) &= \mathbb{E}_\mu\left(\frac{X_1 + X_2}{2}\right)^2 = \frac{1}{4}\mathbb{E}_\mu(X_1^2 + 2X_1X_2 + X_2^2) = \frac{1}{4}(2(\sigma^2 + \mu^2) + 2\mathbb{E}_\mu(X_1X_2)) \\ &= \frac{1}{4}(2(\sigma^2 + \mu^2) + 2\mu^2) = \frac{1}{2}\sigma^2 + \mu^2.\end{aligned}$$

Cet estimateur est donc biaisé, de biais : $\sigma^2/2$. Remarquons que l'on aurait pu trouver cette espérance plus rapidement en notant que la v.a.r. $Y = (X_1 + X_2)/2$ est de loi $N(\mu, \sigma^2/2)$. D'où

$$\mathbb{E}_\mu S_5 = \mathbb{E}_\mu Y^2 = \text{Var}Y + \mathbb{E}_\mu^2 Y = \frac{\sigma^2}{2} + \mu^2.$$

Cette approche est encore utile pour déterminer le risque quadratique de l'estimateur S_5 . On a en effet :

$$\text{Var}_\mu\left(\left(\frac{X_1 + X_2}{2}\right)^2\right) = \text{Var}_\mu Y^2 = 2\left(\frac{\sigma^4}{4} + 2\mu^2\frac{\sigma^2}{2}\right) = \frac{\sigma^4}{2} + 2\mu^2\sigma^2.$$

Ainsi, le risque quadratique de S_5 est :

$$R(S_5, \mu) = \text{Var}_\mu(S_5) + b_{S_5}^2(\mu) = \frac{\sigma^4}{2} + 2\mu^2\sigma^2 + \frac{\sigma^4}{4} = \frac{3\sigma^4}{4} + 2\mu^2\sigma^2.$$

Certes, comparativement à S_3 , l'estimateur S_5 est biaisé mais son risque quadratique est inférieur. Au sens du risque quadratique, il est donc préférable à S_3 .

2) On généralise les estimateurs S_4 et S_5 , pour une taille quelconque n d'échantillon, sous la forme :

$$S_4 = \frac{1}{n} \sum_{i=1}^n X_i^2 \text{ et } S_5 = \left(\frac{1}{n} \sum_{i=1}^n X_i\right)^2 = \bar{X}_n^2.$$

On a

$$\mathbb{E}_\mu S_4 = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\mu X_i^2 = \sigma^2 + \mu^2.$$

Ainsi S_4 est donc biaisé de biais σ^2 . De plus,

$$\text{Var}_\mu S_4 = \frac{1}{n^2} \sum_{i=1}^n \text{Var}_\mu X_i^2 = \frac{2}{n}(\sigma^4 + 2\mu^2\sigma^2).$$

Le risque quadratique de cet estimateur est

$$R(S_4, \mu) = \frac{2}{n}(\sigma^4 + 2\mu^2\sigma^2) + \sigma^4,$$

qui tend vers σ^4 , quand $n \rightarrow +\infty$.

Pour l'estimateur S_5 , comme on sait que \bar{X}_n est de loi $N(\mu, \sigma^2/n)$, on peut écrire :

$$\mathbb{E}_\mu S_5 = \mathbb{E}_\mu \bar{X}_n^2 = \text{Var}_\mu \bar{X}_n + \mathbb{E}_\mu^2 \bar{X}_n = \frac{\sigma^2}{n} + \mu^2.$$

Ainsi, cet estimateur S_5 est biaisé de biais σ^2/n . Il est clairement asymptotiquement sans biais.

Par ailleurs,

$$R(S_5, \mu) = \text{Var}_\mu(S_5) + b_{S_5}^2(\mu) = 2 \left(\frac{\sigma^4}{n^2} + 2\mu^2 \frac{\sigma^2}{n} \right) + \frac{\sigma^4}{n^2} = \frac{3\sigma^4}{n^2} + 4\mu^2 \frac{\sigma^2}{n}.$$

Pour déterminer quel est l'estimateur préférable entre S_4 et S_5 , comparons leurs risques. On a :

$$\begin{aligned} R(S_4, \mu) &= \frac{2+n}{n} \sigma^4 + 4\mu^2 \frac{\sigma^2}{n} \\ R(S_5, \mu) &= \frac{3}{n^2} \sigma^4 + 4\mu^2 \frac{\sigma^2}{n}. \end{aligned}$$

Or,

$$n^2(2+n) - 3n = 2n^2 + n^3 - 3n = n(n^2 + 2n - 3)$$

et le dernier terme est positif dès $n = 1$. Ainsi

$$\frac{2+n}{n} > \frac{3}{n^2}$$

et donc

$$R(S_4, \mu) > R(S_5, \mu)$$

pour $n \geq 1$. Au sens du risque quadratique, S_5 est donc toujours meilleur que S_4 .

Comparons les maintenant en terme de consistance. Par la loi forte des grands nombres nous avons *p.s.* :

$$S_4 = \frac{1}{n} \sum_{i=1}^n X_i^2 \rightarrow \mathbb{E}_\mu X^2 = \mu^2 + \sigma^2,$$

quand $n \rightarrow +\infty$. Cet estimateur n'est donc pas consistant. En revanche l'estimateur S_5 l'est puisque la loi des grands nombres nous donne la convergence *p.s.* de \bar{X}_n vers μ et donc de S_5 vers μ^2 (par le théorème de Slutsky). Cet estimateur est donc en fait fortement consistant. Notons que l'on aurait pu retrouver ce type de résultats en regardant leurs risques quadratiques. On constate en effet que

$$\begin{aligned} R(S_4, \mu) &= \mathbb{E}_\mu (S_4 - \mu^2)^2 \rightarrow \sigma^4 \\ R(S_5, \mu) &= \mathbb{E}_\mu (S_5 - \mu^2)^2 \rightarrow 0, \end{aligned}$$

quand $n \rightarrow +\infty$. Ainsi S_4 n'est pas L^2 -consistant alors que S_5 l'est.

3) On cherche à estimer $g(\mu) = \mu^2$. Or, on sait que l'estimateur du maximum de vraisemblance de μ dans ce modèle est \bar{X}_n . La propriété de l'estimation par maximum de vraisemblance vu en cours permet d'affirmer que l'estimateur du maximum de vraisemblance de μ^2 est \bar{X}_n^2 , c'est à dire S_5 .

Exercice 4 (*Comparaison d'estimateurs dans un modèle uniforme*)

On considère le modèle uniforme $\{\mathcal{U}_{[0,\theta]} : \theta > 0\}$. On considère un échantillon X_1, \dots, X_n et on note $X_{(1)}$ et $X_{(n)}$ respectivement la première et la dernière statistique d'ordre.

On a vu en cours que l'on pouvait proposer les estimateurs suivants pour le paramètre θ .

$$\begin{aligned}\hat{\theta}_1 &= X_{(n)} \\ \hat{\theta}_2 &= \frac{n+1}{n} X_{(n)} \\ \hat{\theta}_3 &= X_{(1)} + X_{(n)} \\ \hat{\theta}_4 &= 2\bar{X}_n,\end{aligned}$$

où \bar{X}_n est l'estimateur de la moyenne empirique.

1) Rappeler brièvement l'idée à la base de la proposition de chacun de ces estimateurs.

2) Pour chacun d'entre eux, étudier la consistance, le biais et donner l'expression de son risque quadratique.

3) Comparer les fonctions de risque quadratique. Qu'en conclure ?

Solution

1) Voir la fin du Chapitre 3.

2)

* Considérons le premier estimateur $\hat{\theta}_1 = \sup_{i=1, \dots, n} X_i = X_{(n)}$.

On a vu dans la correction de l'exercice 3 du Chapitre 2 que :

$$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = \prod_{i=1}^n F_X(x) = (F_X(x))^n$$

Or, si X est une v.a.r. de loi $U_{[0, \theta]}$, sa f.d.r. est

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq 0; \\ \frac{x}{\theta} & \text{if } x \in [0, \theta]; \\ 1 & \text{if } x \geq \theta. \end{cases}$$

D'où

$$F_{X_{(n)}}(x) = \begin{cases} 0 & \text{if } x \leq 0; \\ \left(\frac{x}{\theta}\right)^n & \text{if } x \in [0, \theta]; \\ 1 & \text{if } x \geq \theta. \end{cases}$$

On a alors

$$\begin{aligned}P(|X_{(n)} - \theta| > \varepsilon) &= P(X_{(n)} - \theta > \varepsilon) + P(X_{(n)} - \theta < -\varepsilon) \\ &= P(X_{(n)} > \theta + \varepsilon) + P(X_{(n)} < \theta - \varepsilon) \\ &= 1 - F_{X_{(n)}}(\theta + \varepsilon) + F_{X_{(n)}}(\theta - \varepsilon) = F_{X_{(n)}}(\theta - \varepsilon) \\ &= \left(\frac{\theta - \varepsilon}{\theta}\right)^n = \left(1 - \frac{\varepsilon}{\theta}\right)^n\end{aligned}$$

qui tend vers 0 quand $n \rightarrow +\infty$, au moins quand $0 < \varepsilon < \theta$. La probabilité précédente étant toujours nulle si $\varepsilon > \theta$ ($X_{(n)}$ et θ ne pouvant être distants de plus de θ puisque $X_{(n)}$ est dans $[0, \theta]$), On a donc

$$\hat{\theta}_1 = X_{(n)} \xrightarrow{P} \theta,$$

quand $n \rightarrow +\infty$. Ainsi l'estimateur $\hat{\theta}_1$ est consistant.

Considérons maintenant son biais. On peut écrire :

$$\begin{aligned} \mathbb{E}_\theta \hat{\theta}_1 &= \mathbb{E}_\theta X_{(n)} = \int_0^\theta xn \left(\frac{x}{\theta}\right)^{n-1} \frac{dx}{\theta} = \frac{n}{\theta^n} \int_0^\theta x^n dx \\ &= \frac{n}{\theta^n} \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1} \theta. \end{aligned}$$

L'estimateur $\hat{\theta}_1$ est donc biaisé de biais :

$$b_{\hat{\theta}_1}(\theta) = \mathbb{E}_\theta \hat{\theta}_1 - \theta = \frac{n}{n+1} \theta - \theta = -\frac{\theta}{n+1}$$

Calculons enfin le risque quadratique de $\hat{\theta}_1$. On a

$$\begin{aligned} R(\hat{\theta}_1, \theta) &= \mathbb{E}_\theta (\hat{\theta}_1 - \theta)^2 = \int_0^\theta (x - \theta)^2 n \left(\frac{x}{\theta}\right)^{n-1} \frac{dx}{\theta} = \frac{n}{\theta^n} \int_0^\theta x^{n-1} (x - \theta)^2 dx \\ &= \frac{n}{\theta^n} \int_0^1 (n\theta)^{n-1} \theta^2 (n-1)^2 \theta du = \frac{n\theta^{n+2}}{\theta^n} \int_0^1 (u\theta)^{n-1} \theta^2 (u-1)^2 \theta du \\ &= \frac{n\theta^{n+2}}{\theta^n} \int_0^1 u^{n-1} (1-u)^2 du = \frac{n\theta^{n+2}}{\theta^n} \frac{\Gamma(n)2!}{(n+2)(n+1)n\Gamma(n)} \\ &= \frac{2\theta^2}{(n+2)(n+1)}, \end{aligned}$$

en se souvenant que

$$\beta(n, 3) = \int_0^1 u^{n-1} (1-u)^2 du = \frac{\Gamma(n)\Gamma(3)}{\Gamma(n+3)},$$

où $\beta(\cdot, \cdot)$ est la fonction d'Euler de deuxième espèce et $\Gamma(\cdot)$ celle de première espèce.

* Considérons maintenant le second estimateur

$$\hat{\theta}_2 = \frac{n+1}{n} X_{(n)}.$$

On sait que

$$\hat{\theta}_1 = X_{(n)} \xrightarrow{P} \theta,$$

quand $n \rightarrow +\infty$, ce qui implique bien sûr que

$$\hat{\theta}_2 = \frac{n+1}{n} X_{(n)} \xrightarrow{P} \theta$$

et $\hat{\theta}_2$ est donc un estimateur consistant.

On a de plus

$$\mathbb{E}_\theta \hat{\theta}_2 = \frac{n+1}{n} \mathbb{E}_\theta X_{(n)} = \frac{n+1}{n} \mathbb{E}_\theta \hat{\theta}_1 = \frac{n+1}{n} \frac{n}{n+1} \theta = \theta.$$

et $\hat{\theta}_2$ est donc un estimateur sans biais.

Calculons enfin son risque quadratique. Puisque l'estimateur est sans biais, on peut écrire

$$\begin{aligned} R(\hat{\theta}_2, \theta) &= \text{Var}_\theta \left(\frac{n+1}{n} \hat{\theta}_1 \right) = \left(\frac{n+1}{n} \right)^2 \text{Var}_\theta(\hat{\theta}_1) \\ &= \left(\frac{n+1}{n} \right)^2 \left(R(\hat{\theta}_1, \theta) - b_{\hat{\theta}_1}^2(\theta) \right) = \left(\frac{n+1}{n} \right)^2 \left(\frac{2\theta^2}{(n+2)(n+1)} - \frac{\theta^2}{(n+1)^2} \right) \\ &= \theta^2 \left(\frac{n+1}{n} \right)^2 \left(\frac{2(n+1) - (n+2)}{(n+1)^2(n+2)} \right) = \frac{\theta^2}{n^2(n+2)} n = \frac{\theta^2}{n(n+2)}. \end{aligned}$$

* Le troisième estimateur est défini par $\hat{\theta}_3 = X_{(1)} + X_{(n)}$.

Remarquons en premier lieu que $X_{(1)}$ converge en probabilités vers 0, quand n tend vers $+\infty$. En effet, pour tout $\varepsilon > 0$, on peut écrire :

$$P(|X_{(1)}| > \varepsilon) = \begin{cases} 0 & \text{if } \varepsilon > \theta \\ P(X_{(1)} > \varepsilon) = \prod_{i=1}^n (1 - F_X(\varepsilon)) & \text{if } 0 < \varepsilon < \theta \end{cases}$$

Mais, quand $0 < \varepsilon < \theta$, on a :

$$\prod_{i=1}^n (1 - F_X(\varepsilon)) = \left(1 - \frac{\varepsilon}{\theta} \right)^n$$

qui tend vers 0 quand n tend vers $+\infty$. Par le théorème de Slutsky, les convergences en probabilités respectives de $X_{(1)}$ et $X_{(n)}$ vers 0, entraînent la convergence

$$\hat{\theta}_3 \xrightarrow{P} \theta,$$

quand $n \rightarrow +\infty$ et $\hat{\theta}_3$ est donc un estimateur consistant.

Calculons maintenant le biais de cet estimateur. On a :

$$\mathbb{E}_\theta \hat{\theta}_3 = \mathbb{E}_\theta X_{(1)} + \mathbb{E}_\theta X_{(n)}.$$

et on a vu que

$$\mathbb{E}_\theta X_{(n)} = \frac{n}{n+1} \theta.$$

Toujours dans l'exercice 3 du Chapitre 2, nous avons vu que la densité de $X_{(1)}$ est :

$$f_{X_{(1)}}(x) = n(1 - F(x))^{n-1} f(x).$$

On peut alors écrire (en faisant le changement de variable $u = x/\theta$ dans la première intégrale) :

$$\begin{aligned}\mathbb{E}_\theta X_{(1)} &= \int_0^\theta xn \left(1 - \frac{x}{\theta}\right)^{n-1} \frac{1}{\theta} dx = n \int_0^1 u(1-u)^{n-1} \theta du \\ &= n\theta\beta(2, n) = n\theta \frac{\Gamma(2)\Gamma(n)}{\Gamma(n+2)} = \theta \frac{\Gamma(n+1)}{(n+1)\Gamma(n+1)} = \frac{\theta}{n+1}\end{aligned}$$

Ainsi

$$\mathbb{E}_\theta \hat{\theta}_3 = \frac{n}{n+1}\theta + \frac{\theta}{n+1} = \theta$$

et $\hat{\theta}_3$ est donc un estimateur sans biais.

Cet estimateur étant sans biais, son risque est égal à sa variance et on a donc :

$$\begin{aligned}R(\hat{\theta}_3, \theta) &= \text{Var}_\theta(\hat{\theta}_3) = \text{Var}_\theta(X_{(1)} + X_{(n)}) \\ &= \text{Var}_\theta(X_{(1)}) + \text{Var}_\theta(X_{(n)}) + 2\text{Cov}_\theta(X_{(1)}, X_{(n)}).\end{aligned}$$

Remarquons en premier lieu que $X_{(1)}$ et $X_{(n)}$ ont même variance. On montre en effet aisément que si $X \sim \mathcal{U}_{[0, \theta]}$, alors la v.a.r. Y définie par $Y = \theta - X$ est également de loi uniforme sur l'intervalle $[0, \theta]$. Or,

$$Y_{(1)} = \min_{i=1, \dots, n} Y_i = \min_{i=1, \dots, n} (\theta - X_i) = \theta - \max_{i=1, \dots, n} X_i = \theta - X_{(n)}.$$

D'où, par invariance par translation de la variance, il vient

$$\text{Var}_\theta(Y_{(1)}) = \text{Var}_\theta(\theta - X_{(n)}) = \text{Var}_\theta(-X_{(n)}) = \text{Var}_\theta(X_{(n)}).$$

Comme Y et X ont même loi, on a bien

$$\text{Var}_\theta(X_{(1)}) = \text{Var}_\theta(X_{(n)}).$$

Par ailleurs, nous avons déjà vu que

$$R(\hat{\theta}_1, \theta) = \frac{2\theta^2}{(n+1)(n+2)} \text{ et } b_{\hat{\theta}_1}(\theta) = -\frac{\theta}{n+1}.$$

D'où

$$\begin{aligned}\text{Var}_\theta(X_{(n)}) &= \text{Var}_\theta(\hat{\theta}_1) = R(\hat{\theta}_1, \theta) - b_{\hat{\theta}_1}^2(\theta) = \frac{2\theta^2}{(n+1)(n+2)} - \frac{\theta^2}{(n+1)^2} \\ &= \frac{\theta^2}{(n+1)^2(n+2)} (2(n+1) - (n+2)) = \frac{n\theta^2}{(n+1)^2(n+2)}.\end{aligned}$$

Il nous faut ensuite calculer la covariance

$$\text{Cov}_\theta(X_{(1)}, X_{(n)}) = E(X_{(1)}X_{(n)}) - E(X_{(1)})E(X_{(n)}).$$

Or, nous avons vu (toujours dans le même exercice 3 du Chapitre 2) que

$$f_{X_{(1)}, X_{(n)}}(x_1, x_n) = n(n-1)(F(x_n) - F(x_1))^{n-2} f(x_1)f(x_n)\mathbb{1}_{x_1 < x_n}.$$

En l'appliquant dans le cas de la loi uniforme, on obtient :

$$\begin{aligned}
 E(X_{(1)}X_{(n)}) &= \int_{0 < u < v < \theta} uv \cdot n(n-1) \left(\frac{v}{\theta} - \frac{u}{\theta}\right)^{n-2} \frac{1}{\theta^2} dudv \\
 &= \frac{n(n-1)}{\theta^n} \int_{0 < u < v < \theta} uv(v-u)^{n-2} dudv \\
 &= \frac{n(n-1)}{\theta^n} \int_0^\theta v^n \left(\int_0^v \frac{u}{v} \left(1 - \frac{u}{v}\right)^{n-2} du \right) dv \\
 &= \frac{n(n-1)}{\theta^n} \int_0^\theta v^n \left(\int_0^1 w(1-w)^{n-2} v dw \right) dv \\
 &= \frac{n(n-1)}{\theta^n} \int_0^\theta v^{n+1} \beta(2, n-1) dv = \frac{n(n-1)}{\theta^n} \frac{\Gamma(2)\Gamma(n-1)}{\Gamma(n+1)} \left[\frac{v^{n+2}}{n+2} \right]_0^\theta \\
 &= \theta^2 \frac{n(n-1)}{n+2} \frac{\Gamma(n-1)}{n(n-1)\Gamma(n-1)} = \frac{\theta^2}{n+2}.
 \end{aligned}$$

Ainsi,

$$\text{Cov}_\theta(X_{(1)}, X_{(n)}) = \frac{\theta^2}{n+2} - \frac{\theta}{n+1} \frac{n}{n+1} \theta = \frac{\theta^2}{n+2} - \frac{n\theta^2}{(n+1)^2} = \frac{\theta^2}{(n+1)^2(n+2)}.$$

On a maintenant tout ce qu'il faut pour calculer le risque de l'estimateur $\hat{\theta}_3$. Celui-ci est donc égal à :

$$\begin{aligned}
 R(\hat{\theta}_3, \theta) &= 2\text{Var}(X_{(n)}) + 2\text{Cov}_\theta(X_{(1)}, X_{(n)}) = \frac{2n\theta^2}{(n+1)^2(n+2)} + \frac{2\theta^2}{(n+1)^2(n+2)} \\
 &= \frac{2\theta^2}{(n+1)(n+2)}.
 \end{aligned}$$

* Étudions enfin le dernier estimateur $\hat{\theta}_4$. Par la loi forte des grands nombres, on a :

$$\bar{X}_n \xrightarrow{p.s.} \frac{\theta}{2},$$

quand $n \rightarrow +\infty$, ce qui implique

$$\hat{\theta}_4 = 2\bar{X}_n \xrightarrow{p.s.} \frac{\theta}{2}$$

et $\hat{\theta}_4$ est donc un estimateur fortement consistant.

Par ailleurs, on sait que $\mathbb{E}_\theta \bar{X}_n = \theta/2$. D'où

$$\mathbb{E}_\theta \hat{\theta}_4 = 2\mathbb{E}_\theta \bar{X}_n = \theta$$

et $\hat{\theta}_4$ est donc un estimateur sans biais.

Pour le calcul de son risque quadratique, on peut alors écrire :

$$R(\hat{\theta}_4, \theta) = \text{Var}_\theta(\hat{\theta}_4) = 4\text{Var}_\theta(\bar{X}_n) = \frac{4}{n}\text{Var}(X)$$

Or,

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

et

$$\mathbb{E}_\theta X^2 = \int_0^\theta x^2 \frac{dx}{\theta} = \frac{\theta^2}{3}.$$

Il vient

$$\text{Var}(X) = \frac{\theta^2}{3} - \frac{\theta^2}{4} = \frac{\theta^2}{12}$$

dont on déduit le risque quadratique de notre estimateur

$$R(\hat{\theta}_4, \theta) = \frac{4}{n} \frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

3) En résumé les fonctions de risque quadratique des quatre estimateurs sont :

$$R(\hat{\theta}_1, \theta) = \frac{2\theta^2}{(n+2)(n+1)}$$

$$R(\hat{\theta}_2, \theta) = \frac{\theta^2}{n(n+2)}$$

$$R(\hat{\theta}_3, \theta) = \frac{2\theta^2}{(n+1)(n+2)}$$

$$R(\hat{\theta}_4, \theta) = \frac{\theta^2}{3n}.$$

On constate que $\hat{\theta}_1$ et $\hat{\theta}_3$ ont même risque. De plus,

$$\frac{R(\hat{\theta}_1, \theta)}{R(\hat{\theta}_2, \theta)} = \frac{2n(n+2)}{(n+1)(n+2)} = \frac{2n}{n+1}$$

qui est strictement plus grand que 1, dès que $n > 1$. Ainsi, $R(\hat{\theta}_2, \theta) < R(\hat{\theta}_1) = R(\hat{\theta}_3, \theta)$, pour tout $n > 1$. Enfin,

$$\frac{R(\hat{\theta}_4, \theta)}{R(\hat{\theta}_2, \theta)} = \frac{n(n+2)}{3n} = \frac{n+2}{3}$$

qui est aussi strictement plus grand que 1 pour $n > 1$. Thus, $R(\hat{\theta}_2, \theta) < R(\hat{\theta}_4, \theta)$, pour tout $n > 1$.

En conclusion $\hat{\theta}_2$ est préférable à tous les autres estimateurs proposés. Ces derniers ne sont donc pas admissibles. Mais ceci ne garantit pas que $\hat{\theta}_2$ le soit. On peut en fait montrer que l'estimateur optimal dans la classe des estimateurs de la forme $\alpha X_{(n)}$ est

$$\hat{\theta} = \frac{n+2}{n+1} X_{(n)}.$$

Exercice 5 (*Optimalité pour les estimateurs linéaires de l'espérance mathématique*)

Soit $\mathcal{P} = (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{P_{\theta} : \theta \in \Theta\})$, de v.a. générique X , un modèle paramétrique tel que l'on ait $\mathbb{E}_{\theta}(X) = g(\theta)$. Pour simplifier les notations on notera μ_{θ} cette espérance et σ_{θ}^2 la variance de X . On a vu que \bar{X}_n est un estimateur sans biais de μ_{θ} de risque quadratique $R(\bar{X}_n, \theta) = \sigma_{\theta}^2/n$. Cet estimateur s'exprime bien évidemment comme une combinaison linéaire des v.a.r. de l'échantillon X_1, \dots, X_n . On dit qu'il est linéaire.

1) Montrer qu'un estimateur linéaire et sans biais est forcément une combinaison linéaire des v.a. X_1, \dots, X_n ayant pour somme des coefficients 1. Calculer le risque quadratique d'un tel estimateur.

2) En utilisant l'inégalité de Cauchy-Schwarz pour la somme de réels, montrer que la moyenne empirique est un estimateur optimal dans la classe des estimateur linéaires et sans biais de μ_{θ} .

3) On considère maintenant la classe des estimateurs linéaires, mais pas nécessairement sans biais. Ces estimateur sont de la forme

$$S(\underline{X}) = \sum_{i=1}^n c_i X_i,$$

où les c_1, \dots, c_n sont des réels donnés. Calculer le risque d'un estimateur dans cette classe (en fonction des c_1, \dots, c_n). On cherche l'estimateur optimal dans cette classe. En admettant que la fonction à minimiser est convexe, montrer que le minimum est atteint pour les c_i tous égaux à $\mu_{\theta}^2/(\sigma_{\theta}^2 + n\mu_{\theta}^2)$. En déduire qu'il n'existe pas d'estimateur optimal dans cette classe.

Solution

1) Un estimateur linéaire est de la forme :

$$T(\underline{X}) = \sum_{i=1}^n c_i X_i,$$

où c_1, \dots, c_n sont des réels fixés. Pour qu'il soit sans biais, on doit avoir $\mathbb{E}_{\theta}(T(\underline{X})) = \mu_{\theta}$, c'est à dire :

$$\sum_{i=1}^n c_i \mathbb{E}_{\theta} X_i = \mu_{\theta}.$$

Comme toutes les variables X_i ont même espérance μ_{θ} , il faut bien que l'on ait :

$$\sum_{i=1}^n c_i = 1.$$

Comme cet estimateur est supposé sans biais, son risque s'écrit :

$$R(T(\underline{X}), \theta) = \text{Var}_{\theta} T(\underline{X}) = \text{Var}_{\theta} \sum_{i=1}^n c_i X_i = \sum_{i=1}^n c_i^2 \text{Var}_{\theta} X_i = \sigma_{\theta}^2 \sum_{i=1}^n c_i^2,$$

où l'avant dernière égalité est justifiée par l'indépendance des X_1, \dots, X_n .

2) Par Cauchy-Schwarz , on peut écrire :

$$1 = \left| \sum_{i=1}^n c_i \right| \leq \sqrt{\sum_{i=1}^n c_i^2} \sqrt{\sum_{i=1}^n 1} = \sqrt{n} \sqrt{\sum_{i=1}^n c_i^2},$$

dont on tire :

$$\frac{1}{n} \leq \sum_{i=1}^n c_i^2.$$

On a donc :

$$R(T(\underline{X}), \theta) = \sigma_\theta^2 \sum_{i=1}^n c_i^2 \geq \frac{\sigma_\theta^2}{n} = R(\bar{X}_n, \theta),$$

pour tout θ dans Θ . Ainsi l'estimateur \bar{X}_n est optimal dans la classe des estimateurs linéaires et sans biais de μ_θ .

3) Soit maintenant

$$S(\underline{X}) = \sum_{i=1}^n c_i X_i$$

un estimateur linéaire mais pas forcément sans biais de μ_θ . Son biais est :

$$b_S(\theta) = \mathbb{E}_\theta S(\underline{X}) - \mu_\theta = \sum_{i=1}^n c_i \mu_\theta - \mu_\theta = \mu_\theta \left(\sum_{i=1}^n c_i - 1 \right).$$

Ainsi, le risque quadratique de cet estimateur s'écrit :

$$R(S(\underline{X}), \theta) = \sigma_\theta^2 \sum_{i=1}^n c_i^2 + \mu_\theta^2 \left(\sum_{i=1}^n c_i - 1 \right)^2.$$

La fonction

$$(c_1, \dots, c_n) \longrightarrow f(c_1, \dots, c_n) = \sigma_\theta^2 \sum_{i=1}^n c_i^2 + \mu_\theta^2 \left(\sum_{i=1}^n c_i - 1 \right)^2$$

étant convexe, son minimum est atteint en la valeur qui annule le gradient

$$\nabla f(c_1, \dots, c_n) = \begin{pmatrix} 2\sigma_\theta^2 c_1 + 2\mu_\theta^2 (\sum_{i=1}^n c_i - 1) \\ \vdots \\ 2\sigma_\theta^2 c_n + 2\mu_\theta^2 (\sum_{i=1}^n c_i - 1) \end{pmatrix}.$$

Or,

$$(1) \quad \nabla f(c_1, \dots, c_n) = 0 \iff 2\sigma_\theta^2 c_i + 2\mu_\theta^2 \left(\sum_{i=1}^n c_i - 1 \right) = 0, \text{ pour } i = 1, \dots, n.$$

Pour que la fonction f atteigne son minimum, il faut donc déjà que c_1, \dots, c_n soient tels que :

$$\begin{aligned} \sigma_\theta^2 \sum_{i=1}^n c_i + n\mu_\theta^2 \left(\sum_{i=1}^n c_i - 1 \right) &= 0 \\ \Leftrightarrow \sum_{i=1}^n c_i (\sigma_\theta^2 + n\mu_\theta^2) &= n\mu_\theta^2 \\ \Leftrightarrow \sum_{i=1}^n c_i &= \frac{n\mu_\theta^2}{\sigma_\theta^2 + n\mu_\theta^2}. \end{aligned}$$

En utilisant cette égalité dans l'équation (1), il vient, pour $i = 1, \dots, n$:

$$\sigma_\theta^2 c_i + 2\mu_\theta^2 \left(\frac{n\mu_\theta^2 - \sigma_\theta^2 - n\mu_\theta^2}{\sigma_\theta^2 + n\mu_\theta^2} \right) = 0,$$

c'est à dire :

$$c_i = \frac{\mu_\theta^2}{\sigma_\theta^2 + n\mu_\theta^2},$$

pour $i = 1, \dots, n$. La v.a.r.

$$S^*(\underline{X}) = \frac{\mu_\theta^2}{\sigma_\theta^2 + n\mu_\theta^2} \sum_{i=1}^n X_i$$

est donc celle qui minimise le risque quadratique. Mais ce n'est pas une statistique puisqu'elle dépend du paramètre inconnu du modèle.

Correction des exercices du Chapitre 6

Exercice 1 (*Statistiques exhaustives*)

On considère les modèles déjà largement étudiés dans les feuilles d'exercices précédentes :

- modèle de Poisson $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathcal{P}(\lambda) : \lambda > 0\})$;
- modèle de la loi de exponentielle $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}(\lambda) : \lambda > 0\})$.
- modèle gaussien avec σ^2 positif connu : $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu, \sigma^2) : \mu \in \mathbb{R}\})$;
- modèle gaussien avec μ dans \mathbb{R} connu : $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu, \sigma^2) : \sigma^2 > 0\})$;
- modèle gaussien général : $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\})$.

1) Pour chacun de ces modèles donner l'expression d'une statistique exhaustive (éventuellement vectorielle).

2) Retrouver le résultat pour le modèle de Poisson en utilisant une autre méthode.

Solution

1) **Modèle de Poisson** $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathcal{P}(\lambda) : \lambda > 0\})$.

La vraisemblance des observations est dans ce modèle :

$$L(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda} \prod_{i=1}^n \frac{1}{x_i!} = g_\lambda(T(\underline{x})) h(\underline{x}),$$

en posant

$$h(\underline{x}) = \frac{1}{\prod_{i=1}^n x_i!} ; g_\lambda(u) = \lambda^u e^{-n\lambda} \text{ et } T(\underline{x}) = \sum_{i=1}^n x_i.$$

Par le théorème de factorisation la statistique $T(\underline{x}) = \sum_{i=1}^n x_i$ est exhaustive pour λ .

Modèle Exponentiel $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}(\lambda) : \lambda > 0\})$

On a :

$$L(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\} = g_\lambda(T(\underline{x})) h(\underline{x}),$$

en posant

$$h(\underline{x}) = 1 ; g_\lambda(t) = \lambda^n e^{-\lambda t} \text{ et } T(\underline{x}) = \sum_{i=1}^n x_i.$$

Par le théorème de factorisation, la statistique $T(\underline{x}) = \sum_{i=1}^n x_i$ est exhaustive pour λ .

Modèle Gaussien avec σ^2 connu : $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu, \sigma^2) : \mu \in \mathbb{R}\})$

La vraisemblance de l'échantillon s'écrit :

$$\begin{aligned} L(x_1, \dots, x_n; \mu) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right\} \\ &= \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2}\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2\right\} \\ &= \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2\right)\right\} \\ &= \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp\left\{\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right\} \\ &= g_{\mu}(T(\underline{x})) h(\underline{x}), \end{aligned}$$

en posant

$$\begin{aligned} h(\underline{x}) &= \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right\} \\ g_{\mu}(t) &= \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp\left\{\frac{\mu}{\sigma^2}t - \frac{n\mu^2}{2\sigma^2}\right\} \\ T(\underline{x}) &= \sum_{i=1}^n x_i. \end{aligned}$$

Le théorème de factorisation nous assure que $T(\underline{x}) = \sum_{i=1}^n x_i$ est exhaustive pour μ .

Modèle Gaussien avec μ connu $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu, \sigma^2) : \sigma^2 > 0\})$

On peut écrire la vraisemblance sous la forme

$$L(x_1, \dots, x_n; \sigma^2) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right\} = g_{\sigma^2}(T(\underline{x})) h(\underline{x})$$

avec

$$h(\underline{x}) = 1 ; g_{\sigma^2}(t) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2\sigma^2}t\right\} \text{ et } T(\underline{x}) = \sum_{i=1}^n (x_i - \mu)^2.$$

Le théorème de factorisation nous assure que la statistique $T(\underline{x}) = \sum_{i=1}^n (x_i - \mu)^2$ est exhaustive pour σ^2 .

Modèle Gaussien $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\})$

La vraisemblance s'écrit

$$\begin{aligned} L(\sigma^2; x_1, \dots, x_n; \mu) &= \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2}\right\} \\ &= g_{\mu, \sigma^2}(T(\underline{x})) h(\underline{x}) \end{aligned}$$

avec

$$h(\underline{x}) = 1$$

$$g_{\mu, \sigma^2}(s, t) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp \left\{ \frac{\mu}{\sigma^2} s - \frac{1}{2\sigma^2} t - \frac{n\mu^2}{2\sigma^2} \right\}$$

$$T(\underline{x}) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2 \right).$$

Le théorème de factorisation nous assure que la statistique $T(\underline{x}) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ est exhaustive pour (μ, σ^2) .

2) On va retrouver le résultat pour le modèle de Poisson en revenant à la définition d'une statistique exhaustive. Montrons en premier lieu que si X_1, \dots, X_n sont *i.i.d.* de loi de Poisson $\mathcal{P}(\lambda)$, alors $\sum_{i=1}^n X_i$ est de loi de Poisson $\mathcal{P}(n\lambda)$.

En effet, la transformée de Laplace de la loi $\mathcal{P}(\lambda)$ est :

$$\mathcal{L}_{\mathcal{P}(\lambda)}(s) = \mathbb{E}(e^{sX}) = \sum_{k=0}^{+\infty} e^{sk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda e^s)^k}{k!} = e^{-\lambda} e^{\lambda e^s} = \exp\{\lambda(e^s - 1)\}.$$

Celle de la v.a. $\sum_{i=1}^n X_i$ est alors, par indépendance :

$$\begin{aligned} \mathcal{L}_{\sum_{i=1}^n X_i}(s) &= \mathbb{E}\left(e^{s \sum_{i=1}^n X_i}\right) = \prod_{i=1}^n \mathbb{E}(e^{sX_i}) = (\exp\{\lambda(e^s - 1)\})^n \\ &= \exp\{n\lambda(e^s - 1)\} = \mathcal{L}_{\mathcal{P}(n\lambda)}(s), \end{aligned}$$

ce qui, par caractérisation de la loi par la transformée de Laplace prouve bien le résultat annoncé.

Maintenant, on peut écrire :

$$P\left(X_1 = k_1, \dots, X_n = k_n \mid \sum_{i=1}^n X_i = k\right) = \begin{cases} 0 & \text{si } \sum_{i=1}^n k_i \neq k \\ \frac{P(X_1=k_1, \dots, X_n=k_n)}{P(\sum_{i=1}^n X_i=k)} & \text{si } \sum_{i=1}^n k_i = k. \end{cases}$$

Or,

$$\begin{aligned} \frac{P(X_1 = k_1, \dots, X_n = k_n)}{P(\sum_{i=1}^n X_i = k)} &= \frac{P(X_1 = k_1) \cdots P(X_n = k_n)}{P(\sum_{i=1}^n X_i = k)} = \frac{\frac{\lambda^{k_1}}{k_1!} e^{-\lambda} \cdots \frac{\lambda^{k_n}}{k_n!} e^{-\lambda}}{P(\sum_{i=1}^n X_i = k)} \\ &= \frac{k!}{k_1! \cdots k_n!} \frac{1}{n^k}. \end{aligned}$$

Cette dernière ne dépendant pas de λ , la statistique $T(\underline{X}) = \sum_{i=1}^n X_i$ est donc, d'après la définition, une statistique exhaustive pour λ .

Exercice 2 (Statistique exhaustive et Famille Exponentielle Générale)

On considère une famille exponentielle générale de statistique canonique $T(X)$ où X est la variable générique dans ce modèle.

1) Montrer que $\sum_{i=1}^n T(X_i)$ est une statistique exhaustive pour le modèle d'échantillonnage associé.

2) En utilisant un résultat obtenu dans l'Exercice 1 du chapitre 2, montrer que la moyenne empirique \bar{X}_n est une statistique exhaustive dans un modèle d'échantillonnage de la loi Binomiale.

Solution

1) Rappelons qu'une famille exponentielle générale est un modèle paramétrique où les densités sont de la forme :

$$f_\theta(x) = \exp \{ \langle \eta(\theta), T(x) \rangle \} C(\theta) h(x)$$

où $T(x)$ est la statistique canonique.

La vraisemblance d'un échantillon x_1, \dots, x_n est alors :

$$\begin{aligned} L(x_1, \dots, x_n; \theta) &= \exp \left\{ \langle \eta(\theta), \sum_{i=1}^n T(x_i) \rangle \right\} (C(\theta))^n \prod_{i=1}^n h(x_i) \\ &= g_\theta(S(\underline{x})) h(\underline{x}) \end{aligned}$$

où

$$\begin{aligned} h(\underline{x}) &= \prod_{i=1}^n h(x_i) \\ g_\theta(t) &= \exp \{ \langle \eta(\theta), t \rangle \} (C(\theta))^n \\ S(\underline{x}) &= \sum_{i=1}^n T(x_i). \end{aligned}$$

Ainsi, d'après le théorème de factorisation, la statistique canonique du modèle d'échantillonnage $S(\underline{x}) = \sum_{i=1}^n T(x_i)$ est toujours une statistique exhaustive .

2) On a vu que le modèle binomial constitue une famille exponentielle naturelle, donc de statistique canonique $T(x) = x$. D'après le résultat de la question précédente, les statistiques $\sum_{i=1}^n X_i$ et $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ sont exhaustives pour le paramètre p du modèle binomial.

Exercice 3 (*Estimation optimale dans le modèle de Poisson*)

Il est courant de constater que le nombre d'appels reçus en une heure par un standard téléphonique suit une loi de Poisson. On s'intéresse au problème de l'estimation de la probabilité qu'il n'y ait pas d'appel en une heure. Pour cela, on considère le modèle statistique $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{P(\lambda) : \lambda > 0\})$, de v.a. générique X . On note $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans un modèle et on cherche donc à estimer $g(\lambda) = P_\lambda(X = 0) = \exp(-\lambda)$.

1) Proposer un estimateur $W(\underline{X})$ de $g(\lambda)$ fonction des v.a. $\mathbb{1}_{\{X_i=0\}}$, pour $i = 1, \dots, n$.

- 2) Donner son biais, son risque quadratique et sa loi.
 3) L'estimateur proposé $W(\underline{X})$ est-il fonction de la statistique exhaustive :

$$T(\underline{X}) = \sum_{i=1}^n X_i?$$

Sinon, que proposeriez-vous pour améliorer l'estimation ?

- 4) Calculer la loi de chaque X_i , pour $i = 1, \dots, n$, conditionnelle à $\{T(\underline{X}) = t\}$.
 5) On note

$$Y_i = \mathbb{1}_{\{X_i=0\}}.$$

Calculer l'espérance conditionnelle $\mathbb{E}_\lambda(Y_i|T(\underline{X}) = t)$ et $\mathbb{E}_\lambda(Y_i|T(\underline{X}))$.

6) En déduire l'expression, en fonction de T , de l'estimateur $W^*(\underline{X})$, amélioration de l'estimateur $W(\underline{X})$ par le théorème de Rao-Blackwell. Que dire du biais de $W^*(\underline{X})$?

7) Calculer $\mathbb{E}_\lambda(z^{T(\underline{X})})$ puis $\text{Var}_\lambda(z^{T(\underline{X})})$. En déduire le risque quadratique de l'estimateur $W^*(\underline{X})$.

- 8) Montrer que la statistique $T(\underline{X})$ est également complète. Conclure.

Solution

- 1) On a

$$g(\lambda) = P_\lambda(X = 0) = \mathbb{E}_\lambda(\mathbb{1}_{\{X=0\}}).$$

Ainsi un estimateur par la méthode des moments est :

$$W(\underline{X}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i=0\}}.$$

- 2) On a :

$$\mathbb{E}_\lambda(W(\underline{X})) = \frac{1}{n} n P(X = 0) = g(\lambda),$$

ce qui prouve que $W(\underline{X})$ est un estimateur sans biais de $g(\lambda) = P_\lambda(X = 0)$.

Par ailleurs, puisque $W(\underline{X})$ est sans biais, son risque quadratique s'écrit :

$$R(W(\underline{X}), \lambda) = \text{Var}_\lambda(W(\underline{X})) = \frac{1}{n^2} \text{Var}_\lambda\left(\sum_{i=1}^n \mathbb{1}_{\{X_i=0\}}\right)$$

On sait que

$$\sum_{i=1}^n \mathbb{1}_{\{X_i=0\}} \sim B(n, P_\lambda(X = 0)) = B(n, g(\lambda))$$

et a pour variance $ng(\lambda)(1 - g(\lambda))$.

Ainsi,

$$R(W(\underline{X}), \lambda) = \frac{1}{n^2} ng(\lambda)(1 - g(\lambda)) = \frac{e^{-\lambda}(1 - e^{-\lambda})}{n}.$$

Déterminons maintenant la loi de $W(\underline{X})$. Comme $\sum_{i=1}^n \mathbb{1}_{\{X_i=0\}}$ est de loi binomiale de paramètres n et $g(\lambda)$, la statistique $W(\underline{X})$ est une v.a. discrète à valeurs dans

$$\left\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\right\}$$

telle que :

$$\begin{aligned} P_\lambda \left(W(\underline{X}) = \frac{k}{n} \right) &= P_\lambda \left(\sum_{i=1}^n \mathbb{1}_{\{X_i=0\}} = k \right) \\ &= \binom{n}{k} (g(\lambda))^k (1 - g(\lambda))^{n-k} \\ &= \binom{n}{k} e^{-k\lambda} (1 - e^{-\lambda})^{n-k}. \end{aligned}$$

3) La statistique $W(\underline{X})$ n'est pas une fonction de la statistique exhaustive $T(\underline{X}) = \sum_{i=1}^n X_i$. Ainsi, grâce au théorème de Rao-Blackwell, l'estimateur

$$W^*(\underline{X}) = \mathbb{E}_\lambda((W(\underline{X})|T(\underline{X})))$$

a un risque quadratique inférieur à celui de $W(\underline{X})$.

4) Effectuons le calcul pour X_1 . Par symétrie le résultat sera vrai pour tout $i = 1, \dots, n$.

Rappelons nous que $\sum_{i=1}^n X_i$ est de loi de Poisson $\mathcal{P}(n\lambda)$. Conditionnellement à $\{T = t\}$, la valeur X_1 est à valeur dans $\{0, 1, \dots, t\}$ et on a :

$$\begin{aligned} P_\lambda(X_1 = k|T = t) &= \frac{P_\lambda(X_1 = k, T = t)}{P_\lambda(T = t)} = \frac{P_\lambda(X_1 = k, \sum_{i=2}^n X_i = t - k)}{P_\lambda(\sum_{i=1}^n X_i = t)} \\ &= \frac{\frac{\lambda^k}{k!} e^{-\lambda} \cdot \frac{((n-1)\lambda)^{t-k}}{(t-k)!} e^{-(n-1)\lambda}}{\frac{(n\lambda)^t}{t!} e^{-n\lambda}} = \frac{t!}{k!(t-k)!} \frac{(n-1)^{t-k}}{n^t} \\ &= \frac{t!}{k!(t-k)!} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{t-k}, \end{aligned}$$

ce qui prouve que

$$\mathcal{L}(X_1|T = t) = B\left(t, \frac{1}{n}\right)$$

5) On a :

$$\begin{aligned} \mathbb{E}_\lambda(Y_i|T = t) &= E_\lambda(\mathbb{1}_{\{X_i=0\}}|T = t) = P_\lambda(X_i = 0|T = t) \\ &= \left(1 - \frac{1}{n}\right)^t, \end{aligned}$$

grâce au résultat de la question précédente. Ainsi,

$$\mathbb{E}_\lambda(Y_i|T) = \left(1 - \frac{1}{n}\right)^T$$

6) On a :

$$\begin{aligned} W^*(\underline{X}) &= \mathbb{E}_\lambda(W(\underline{X})|T(\underline{X})) = \mathbb{E}_\lambda\left(\frac{1}{n} \sum_{i=1}^n Y_i \middle| T(\underline{X})\right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\lambda(Y_i|T(\underline{X})) = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{1}{n}\right)^T = \left(1 - \frac{1}{n}\right)^T \end{aligned}$$

Comme $W(\underline{X})$ était un estimateur sans biais, le théorème de Rao-Blackwell nous assure que $W^*(\underline{X})$ est également sans biais.

7) On a :

$$\begin{aligned} \mathbb{E}_\lambda(z^{T(\underline{X})}) &= \sum_{t=0}^{\infty} z^t P_\lambda(T(\underline{X}) = t) = \sum_{t=0}^{\infty} z^t \frac{(n\lambda)^t}{t!} e^{-n\lambda} \\ &= e^{-n\lambda} \sum_{t=0}^{\infty} \frac{(n\lambda z)^t}{t!} = \exp\{-n\lambda + n\lambda z\} = \exp\{n\lambda(z - 1)\}. \end{aligned}$$

On sait que

$$\text{Var}_\lambda(z^{T(\underline{X})}) = \mathbb{E}_\lambda\left((z^{T(\underline{X})})^2\right) - (\mathbb{E}_\lambda(z^{T(\underline{X})}))^2.$$

Or,

$$\mathbb{E}_\lambda\left((z^{T(\underline{X})})^2\right) = \mathbb{E}_\lambda\left((z^2)^{T(\underline{X})}\right) = \exp\{n\lambda(z^2 - 1)\},$$

en appliquant le résultat précédent à z^2 .

Ainsi,

$$\text{Var}_\lambda(z^{T(\underline{X})}) = \exp\{n\lambda(z^2 - 1)\} - \exp\{2n\lambda(z - 1)\}$$

Le calcul du risque quadratique de $W^*(\underline{X})$ donne alors

$$\begin{aligned} R(W^*(\underline{X}), \lambda) &= \text{Var}_\lambda(W^*(\underline{X})) = \text{Var}_\lambda\left(\left(1 - \frac{1}{n}\right)^{T(\underline{X})}\right) \\ &= \exp\left\{n\lambda \left[\left(1 - \frac{1}{n}\right)^2 - 1\right]\right\} - \exp\left\{2n\lambda \left[1 - \frac{1}{n} - 1\right]\right\} \\ &= \exp\left\{n\lambda \left(1 - \frac{2}{n} + \frac{1}{n^2} - 1\right)\right\} - \exp\{-2\lambda\} \\ &= \exp\left\{-\lambda \left(2 - \frac{1}{n}\right)\right\} - \exp\{-2\lambda\} \\ &= e^{-2\lambda} (e^{\lambda/n} - 1). \end{aligned}$$

8) Soit f une fonction borélienne telle que

$$\mathbb{E}_\lambda(f(T(\underline{X}))) = 0,$$

pour tout λ . On a :

$$\mathbb{E}_\lambda(f(T(\underline{X}))) = \sum_{t=0}^{+\infty} f(t)P(T(\underline{X}) = t) = \sum_{t=0}^{+\infty} f(t) \frac{(n\lambda)^t}{t!} e^{-n\lambda},$$

qui est une fonction de λ . Notons h la fonction définie pour tout $\lambda > 0$ par :

$$h(\lambda) = \sum_{t=0}^{+\infty} f(t) \frac{n^t}{t!} \lambda^t.$$

Il s'agit d'une série entière que l'on sait pouvoir s'écrire sous la forme :

$$h(\lambda) = \sum_{t=0}^{+\infty} \frac{h^{(t)}(0)}{t!} \lambda^t,$$

dont on tire par identification :

$$h^{(t)}(0) = f(t)n^t.$$

Par hypothèse la fonction h doit être identiquement nulle et donc nécessairement $f(t) = 0$, pour tout t dans \mathbb{N} . Ceci montre que, par définition, la statistique $T(\underline{X})$ est complète.

En résumé, la statistique $T(\underline{X})$ est exhaustive et complète et d'après le théorème de Lehmann-Scheffé l'estimateur $W^*(\underline{X})$ est optimal dans la classe des estimateurs sans biais de $g(\lambda)$.

Exercice 4 (*Estimation optimale dans le modèle uniforme*)

On considère le modèle de la loi uniforme $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{U}_{[0,\theta]} : \theta > 0\})$ et un échantillon X_1, \dots, X_n dans ce modèle. On se propose d'améliorer, si possible, l'estimateur $\hat{\theta}(\underline{X}) = X_{(1)} + X_{(n)}$ vu en cours et dans l'exercice 4 du chapitre 5.

- 1) Donner une statistique exhaustive dans ce modèle pour le paramètre θ .
- 2) Calculer la densité de la loi de $X_{(1)}$ conditionnelle à $\{X_{(n)} = x_n\}$. En déduire l'expression de $\mathbb{E}_\theta(X_{(1)}|X_{(n)} = x_n)$ puis de $\mathbb{E}_\theta(X_{(1)}|X_{(n)})$.
- 3) Déterminer alors $\tilde{\theta}(\underline{X})$, estimateur amélioré de $\hat{\theta}(\underline{X})$ par le théorème de Rao-Blackwell.
- 4) La statistique $X_{(n)}$ est-elle complète ? Conclure.

Solution

- 1) La vraisemblance est :

$$\begin{aligned} L(\theta; x_1, \dots, x_n) &= \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{1}_{[0,\theta]}(x_i) \\ &= \frac{1}{\theta^n} \mathbb{1}_{[0,\theta]}(x_{(n)}) \\ &= g_\theta(T(\underline{x})) h(\underline{x}), \end{aligned}$$

avec

$$h(\underline{x}) = 1 ; g_\theta(t) = \frac{1}{\theta^n} \mathbb{1}_{[0,\theta]}(t) \text{ et } T(\underline{x}) = x_{(n)}.$$

D'après le théorème de factorisation, la statistique $T(\underline{x}) = x_{(n)}$ est exhaustive pour le paramètre θ .

2) La loi conditionnelle de $X_{(1)}$ sachant $\{X_{(n)} = x_{(n)}\}$ est de densité

$$f_{X_{(1)}}^{X_{(n)}=x_{(n)}}(x_1) = \frac{f_{X_{(1)},X_{(n)}}(x_1, x_n)}{f_{X_{(n)}}(x_n)}.$$

Or, nous avons vu dans l'exercice 3 du chapitre 2 que la loi de la n -ième statistique d'ordre est de densité :

$$f_{X_{(n)}}(x_n) = nF^{n-1}(x_n) f(x_n)$$

et que celle du couple $(X_{(1)}, X_{(n)})$ est :

$$f_{X_{(1)},X_{(n)}}(x_1, x_n) = n(n-1)(F(x_n) - F(x_1))^{n-2} f(x_1)f(x_n) \mathbb{1}_{x_1 \leq x_n}$$

Comme dans notre cas les v.a. initiales sont de loi uniforme, on a

$$\begin{aligned} f_{X_{(n)}}(x_n) &= n \left(\frac{x_n}{\theta}\right)^{n-1} \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x_n) \\ f_{X_{(1)},X_{(n)}}(x_1, x_n) &= n(n-1) \left(\frac{x_n}{\theta} - \frac{x_1}{\theta}\right)^{n-2} \frac{1}{\theta^2} \mathbb{1}_{0 < x_1 \leq x_n < \theta} \\ &= \frac{n(n-1)}{\theta^n} (x_n - x_1)^{n-2} \mathbb{1}_{0 \leq x_1 \leq x_n \leq \theta}. \end{aligned}$$

Ainsi,

$$\begin{aligned} f_{X_{(1)}}^{X_{(n)}=x_n}(x_1) &= \frac{\frac{n(n-1)}{\theta^n} (x_n - x_1)^{n-2}}{\frac{n}{\theta^n} x_n^{n-1}} \\ &= (n-1) \frac{(x_n - x_1)^{n-2}}{x_n^{n-1}} \mathbb{1}_{[0,x_n]}(x_1) \end{aligned}$$

On a alors :

$$\begin{aligned} \mathbb{E}(X_{(1)} | X_{(n)} = x_n) &= \int_0^{x_n} x_1 (n-1) \frac{(x_n - x_1)^{n-2}}{x_n^{n-1}} dx_1 \\ &= (n-1) \int_0^{x_n} \frac{x_1}{x_n} \left(1 - \frac{x_1}{x_n}\right)^{n-2} dx_1 \\ &= (n-1) \int_0^1 u(1-u)^{n-2} x_n du \\ &= x_n (n-1) \beta(2, n-1), \end{aligned}$$

où l'avant dernière égalité est obtenue par le changement de variable $u = x_1/x_n$. Comme

$$\beta(2, n-1) = \frac{\Gamma(2)\Gamma(n-1)}{\Gamma(n+1)} = \frac{1 \cdot \Gamma(n-1)}{n(n-1)\Gamma(n-1)} = \frac{1}{n(n-1)},$$

on a

$$\mathbb{E}(X_{(1)}|X_{(n)} = x_n) = \frac{x_{(n)}}{n}$$

et finalement :

$$\mathbb{E}(X_{(1)}|X_{(n)}) = \frac{X_{(n)}}{n}.$$

3) L'estimateur $\tilde{\theta}(\underline{X})$ amélioré par le théorème de Rao-Blackwell est :

$$\begin{aligned} \tilde{\theta}(\underline{X}) &= \mathbb{E}\left(\hat{\theta}(\underline{X})|X_{(n)}\right) \\ &= \mathbb{E}\left(X_{(1)} + X_{(n)}|X_{(n)}\right) \\ &= \mathbb{E}\left(X_{(1)}|X_{(n)}\right) + X_{(n)} \\ &= \frac{X_{(n)}}{n} + X_{(n)} \\ &= \frac{n+1}{n}X_{(n)}. \end{aligned}$$

4) Pour toute fonction h mesurable telle que l'intégrale suivante existe, on peut écrire :

$$\mathbb{E}h(X_{(n)}) = \int_0^\theta nh(x_n) \frac{x_n^{n-1}}{\theta^n} dx_n.$$

Ainsi, la condition

$$\mathbb{E}h(X_{(n)}) = 0, \forall \theta > 0$$

est équivalente à

$$\int_0^\theta h(x_n)x_n^{n-1}dx_n = 0, \forall \theta > 0.$$

Si la fonction h est continue, on obtient en différentiant par rapport à θ que l'on doit avoir :

$$h(x)x^{n-1} = 0, \forall x > 0 \text{ et donc } h(x) = 0, \forall x > 0.$$

Si h n'est pas continue on pourrait montrer que le résultat reste vrai sauf sur un ensemble de P_θ mesure nulle.

La statistique $X_{(n)}$ est donc complète. Comme $\tilde{\theta}(\underline{X})$ est un estimateur sans biais, le théorème de Lehmann-Scheffé nous assure qu'il est optimal dans la classe des estimateurs sans biais de θ .

Correction des exercices du Chapitre 8

Exercice 1 (*Etude asymptotique du modèle de Bernoulli*)

On considère le modèle de Bernoulli $(\{0, 1\}, \mathcal{P}(\{0, 1\}), \{\mathcal{B}(p) : p \in]0, 1[\})$. On a vu, à maintes reprises (en particulier dans l'exercice 2 du chapitre 5 dont on gardera les notations), que ce modèle pouvait être, entre autres, utile dans un problème de modélisation en Fiabilité.

1) Montrer, de deux manières différentes, que l'estimateur par maximum de vraisemblance \hat{p}_n du paramètre p de ce modèle est asymptotiquement normal. Donner une approximation de la loi de \hat{p}_n quand la taille n de l'échantillon est grande.

2) En utilisant les résultats de l'exercice 2 du chapitre 5, montrer que l'estimateur de la fonction de répartition empirique en $\hat{F}_n(x)$ est également asymptotiquement normal.

3) Construire un intervalle de confiance asymptotique $1 - \alpha$ pour le paramètre p du modèle de Bernoulli.

4) En déduire un intervalle de confiance asymptotique $1 - \alpha$ pour $F(x)$, avec x fixé.

Solution

1) On a vu que l'estimateur du maximum de vraisemblance pour l'échantillon observé x_1, \dots, x_n est :

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n.$$

Montrons en premier lieu l'asymptotique normalité, via le théorème de la limite centrale. Les v.a.r. X_1, \dots, X_n sont i.i.d. et dans L^2 , d'espérance p et variance $p(1-p)$. Le théorème de la limite centrale nous donne

$$\sqrt{n} \left(\frac{\hat{p}_n - p}{\sqrt{p(1-p)}} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1),$$

dont on tire

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, p(1-p)).$$

On peut également obtenir ce résultat en utilisant les propriétés asymptotiques de l'estimateur du maximum de vraisemblance (Cf. Chapitre 7, Section 3). On vérifie en effet aisément que les hypothèses du Théorème 7.5 sont remplies et que l'on a donc :

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, I^{-1}(p)),$$

où $I(p)$ est la matrice d'information de Fisher de la v.a.r. X de loi de Bernoulli $B(p)$ (i.e. pour un échantillon de taille 1).

On a vu dans la question 4) de l'Exercice 2 du Chapitre 5 que :

$$I_n(p) = \frac{n}{p(1-p)}$$

et donc que pour un échantillon de taille 1 ($n = 1$), on a

$$I(p) = \frac{1}{p(1-p)} \implies I^{-1}(p) = p(1-p).$$

Ainsi on retrouve bien la convergence

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, p(1-p)).$$

On en déduit que, pour de grandes valeurs de la taille d'échantillon n , la loi de $\sqrt{n}(\hat{p}_n - p)$ est approximativement une loi $N(0, p(1-p))$ et donc, toujours pour n grand,

$$\hat{p}_n \sim N\left(p, \frac{p(1-p)}{n}\right).$$

2) On a aussi vu dans le même Exercice 2 du Chapitre 5 que, si X_1, \dots, X_n sont i.i.d. de f.d.r. F , on peut alors estimer F par la fonction de répartition empirique :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}.$$

On a également vu que, pour x fixé, cela revenait à mener une estimation paramétrique dans le modèle de Bernoulli, en utilisant les v.a.r.

$$Y_i = \mathbb{1}_{\{X_i \leq x\}}, \text{ pour } i = 1, \dots, n,$$

qui sont de loi de Bernoulli $B(F(x))$. L'estimateur de la f.d.r. empirique s'écrit alors, en x fixé, comme :

$$\hat{F}_n(x) = \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i.$$

En utilisant les résultats de la question 1 on a :

$$\sqrt{n} \left(\hat{F}_n(x) - F(x) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, F(x)(1-F(x))).$$

3) On aurait peut être envie de prendre comme variable asymptotiquement pivotale la variable

$$\pi(\underline{X}_n, p) = \sqrt{n} \left(\frac{\hat{p}_n - p}{\sqrt{p(1-p)}} \right),$$

de laquelle nous déduirions un intervalle de confiance asymptotique $1 - \alpha$ pour p . Mais on voit vite qu'il n'est pas aisé d'isoler p à partir de la double inégalité

$$-z_{1-\alpha/2} \leq \sqrt{n} \left(\frac{\hat{p}_n - p}{\sqrt{p(1-p)}} \right) \leq z_{1-\alpha/2}$$

où z_α est le α -quantile de la $N(0, 1)$.

Il est bien plus facile d'utiliser la variable asymptotiquement pivotale

$$\pi(\underline{X}_n, p) = \sqrt{n} \left(\frac{\hat{p}_n - p}{\sqrt{\hat{p}_n(1 - \hat{p}_n)}} \right)$$

qui converge, elle aussi, vers une loi $N(0, 1)$:

$$(2) \quad \pi(\underline{X}_n, p) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1).$$

On a en effet

$$\sqrt{n} \left(\frac{\hat{p}_n - p}{\sqrt{p(1 - p)}} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1)$$

et

$$\sqrt{\frac{p(1 - p)}{\hat{p}_n(1 - \hat{p}_n)}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 1,$$

la dernière convergence étant justifiée par la propriété de forte consistance de l'estimateur du maximum de vraisemblance (Cf. Théorème 7.5). On aurait pu également invoquer ici la loi des grands nombres.

On peut alors écrire :

$$\begin{aligned} -z_{1-\alpha/2} &\leq \sqrt{n} \left(\frac{\hat{p}_n - p}{\sqrt{\hat{p}_n(1 - \hat{p}_n)}} \right) \leq z_{1-\alpha/2} \\ \iff -z_{1-\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} &\leq \hat{p}_n - p \leq z_{1-\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} \\ \iff \hat{p}_n - z_{1-\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} &\leq p \leq \hat{p}_n + z_{1-\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} \end{aligned}$$

De la convergence obtenue en (2) on tire :

$$P \left(-z_{1-\alpha/2} \leq \sqrt{n} \left(\frac{\hat{p}_n - p}{\sqrt{\hat{p}_n(1 - \hat{p}_n)}} \right) \leq z_{1-\alpha/2} \right) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha.$$

Ainsi, l'intervalle

$$\left[\hat{p}_n - z_{1-\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}}, \hat{p}_n + z_{1-\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} \right]$$

est alors un intervalle de confiance asymptotique $(1 - \alpha)$ pour p .

4) En utilisant les notations de la question 2), l'intervalle

$$\left[\hat{F}_n(x) - z_{1-\alpha/2} \sqrt{\frac{\hat{F}_n(x) (1 - \hat{F}_n(x))}{n}}, \hat{F}_n(x) + z_{1-\alpha/2} \sqrt{\frac{\hat{F}_n(x) (1 - \hat{F}_n(x))}{n}} \right]$$

est un intervalle de confiance asymptotique $(1 - \alpha)$ pour $F(x)$.

Exercice 2 (*Etude asymptotique et Intervalles de confiance (exacts et asymptotiques) dans le modèle de la loi exponentielle*)

On considère le modèle de la loi exponentielle $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{\mathcal{E}(\lambda) : \lambda > 0\})$ et $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans ce modèle. On rappelle que l'estimateur du maximum de vraisemblance du paramètre λ basé sur l'observation d'un tel échantillon est

$$\hat{\lambda}_n = \frac{1}{\bar{X}_n}.$$

1) En utilisant la propriété vue en cours sur l'estimateur du maximum de vraisemblance, montrer que l'estimateur $\hat{\lambda}_n$ est asymptotiquement normal (on précisera bien la convergence en loi obtenue).

2) Retrouver le résultat de la question précédente en utilisant en particulier la δ -méthode.

3) Dédurre, de ce comportement asymptotiquement normal de $\hat{\lambda}_n$, un intervalle de confiance asymptotique $1 - \alpha$ pour λ .

4) Montrer que si Y est une v.a. de loi Gamma $G(\alpha, \beta)$, alors la v.a. βY est de loi $G(\alpha, 1)$.

5) En utilisant le résultat de la question précédente et celui vu dans l'exercice 5 (partie 2) du chapitre 4, montrer que l'intervalle

$$\left[\frac{\chi_{\alpha/2}^2(2n)}{2n\bar{X}_n}, \frac{\chi_{1-\alpha/2}^2(2n)}{2n\bar{X}_n} \right]$$

est un intervalle de confiance $1 - \alpha$ exact pour λ , où $\chi_{\alpha}^2(n)$ est le quantile d'ordre α d'une loi $\chi^2(n)$. (Ind. On rappelle qu'une loi $\chi^2(n)$ est une loi $G(n/2, 1/2)$)

Solution

1) Grâce à nouveau au Théorème 7.5 sur le comportement asymptotique de l'estimateur du maximum de vraisemblance, on a :

$$\sqrt{n} \left(\hat{\lambda}_n - \lambda \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, I^{-1}(\lambda)).$$

On a vu dans l'Exercice 1 du Chapitre 5 que l'information de Fisher pour un échantillon de taille n dans ce modèle est

$$I_n(\lambda) = \frac{n}{\lambda^2}.$$

Celle pour un échantillon de taille 1 est donc $1/\lambda^2$. Finalement on a précisément :

$$\sqrt{n} \left(\hat{\lambda}_n - \lambda \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, \lambda^2).$$

2) D'après le théorème de la limite centrale, on a :

$$\sqrt{n} \left(\frac{\bar{X}_n - \mathbb{E}X}{\sigma_X} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1)$$

donc ici :

$$\sqrt{n} \left(\frac{\bar{X}_n - \frac{1}{\lambda}}{\frac{1}{\lambda}} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1).$$

On en déduit que

$$\sqrt{n} \left(\bar{X}_n - \frac{1}{\lambda} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N\left(0, \frac{1}{\lambda^2}\right).$$

En utilisant la δ -méthode avec la fonction $g(x) = 1/x$ dérivable sur \mathbb{R}_+^* , de dérivée $g'(x) = -1/x^2$, on a

$$\sqrt{n} \left(g(\bar{X}_n) - g(1/\lambda) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N\left(0, \frac{1}{\lambda^2} \left(g' \left(\frac{1}{\lambda} \right) \right)^2 \right),$$

c'est à dire

$$\sqrt{n} \left(\hat{\lambda}_n - \lambda \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N\left(0, \frac{1}{\lambda^2} \left(-\frac{1}{1/\lambda^2} \right)^2 \right) = N(0, \lambda^2).$$

3) Du résultat précédent on tire :

$$\sqrt{n} \left(\frac{\hat{\lambda}_n - \lambda}{\hat{\lambda}_n} \right) = \sqrt{n} \left(\frac{\hat{\lambda}_n - \lambda}{\lambda} \right) \cdot \frac{\lambda}{\hat{\lambda}_n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1)$$

puisque $\hat{\lambda}_n$ est fortement consistant par propriété de l'estimateur du maximum de vraisemblance. Ainsi

$$\pi(\underline{X}_n, \lambda) = \sqrt{n} \left(\frac{\hat{\lambda}_n - \lambda}{\hat{\lambda}_n} \right)$$

est une v.a. asymptotiquement pivotale et normale.

En notant toujours z_α le α -quantile de la loi $N(0, 1)$, on a :

$$\begin{aligned} -z_{1-\alpha/2} &\leq \sqrt{n} \left(\frac{\hat{\lambda}_n - \lambda}{\hat{\lambda}_n} \right) \leq z_{1-\alpha/2} \\ \iff -\frac{\hat{\lambda}_n}{\sqrt{n}} z_{1-\alpha/2} &\leq \hat{\lambda}_n - \lambda \leq \frac{\hat{\lambda}_n}{\sqrt{n}} z_{1-\alpha/2} \\ \iff \hat{\lambda}_n - \frac{\hat{\lambda}_n}{\sqrt{n}} z_{1-\alpha/2} &\leq \lambda \leq \hat{\lambda}_n + \frac{\hat{\lambda}_n}{\sqrt{n}} z_{1-\alpha/2}. \end{aligned}$$

Comme on a

$$P \left(-z_{1-\alpha/2} \leq \sqrt{n} \left(\frac{\hat{\lambda}_n - \lambda}{\hat{\lambda}_n} \right) \leq z_{1-\alpha/2} \right) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha,$$

d'après la convergence obtenue plus haut, l'intervalle

$$\left[\hat{\lambda}_n - z_{1-\alpha/2} \frac{\hat{\lambda}_n}{\sqrt{n}}, \hat{\lambda}_n + z_{1-\alpha/2} \frac{\hat{\lambda}_n}{\sqrt{n}} \right]$$

est un intervalle de confiance asymptotique $(1 - \alpha)$ pour λ .

4) Soit Y une v.a.r. de loi $G(\alpha, \beta)$ et h une fonction mesurable bornée. On a :

$$\begin{aligned}\mathbb{E}(h(\beta Y)) &= \int_0^\infty h(\beta y) \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} dy \\ &= \int_0^\infty h(u) \frac{\beta}{\Gamma(\alpha)} u^{\alpha-1} e^{-u} \frac{du}{\beta} \\ &= \int_0^\infty h(u) \frac{1}{\Gamma(u)} u^{\alpha-1} e^{-u} du,\end{aligned}$$

en effectuant le changement de variable $u = \beta y$ dans la première intégrale pour obtenir la seconde. Ainsi, d'après le critère des fonctions positives, la v.a.r. βY est de loi $G(\alpha, 1)$.

5) On sait, d'après l'Exercice 5 du Chapitre 4, que si les v.a.r. X_1, \dots, X_n sont i.i.d. de loi $\mathcal{E}(\lambda)$, on a :

$$\sum_{i=1}^n X_i \sim G(n, \lambda).$$

En utilisant le résultat de la question précédente, on en déduit que

$$\lambda \sum_{i=1}^n X_i \sim G(n, 1) \text{ et aussi que } 2\lambda \sum_{i=1}^n X_i \sim G(n, 1/2).$$

Ainsi

$$2n\lambda\bar{X}_n \sim G\left(\frac{2n}{2}, \frac{1}{2}\right) = \chi^2(2n).$$

La v.a.r.

$$\pi(\underline{X}_n, \lambda) = 2n\lambda\bar{X}_n$$

est alors une variable pivotale.

En notant $\chi_{\alpha/2}^2(2n)$ et $\chi_{1-\alpha/2}^2(2n)$ les quantiles d'ordre respectifs $\alpha/2$ et $1 - \alpha/2$ de la loi $\chi^2(2n)$, on a alors :

$$P(\chi_{\alpha/2}^2(2n) \leq 2n\lambda\bar{X}_n \leq \chi_{1-\alpha/2}^2(2n)) = 1 - \alpha.$$

Or,

$$\begin{aligned}\chi_{\alpha/2}^2(2n) &\leq 2n\lambda\bar{X}_n \leq \chi_{1-\alpha/2}^2(2n) \\ \Leftrightarrow \frac{\chi_{\alpha/2}^2(2n)}{2n\bar{X}_n} &\leq \lambda \leq \frac{\chi_{1-\alpha/2}^2(2n)}{2n\bar{X}_n}.\end{aligned}$$

Un intervalle de confiance exacte $1 - \alpha$ pour λ est donc :

$$\left[\frac{\chi_{\alpha/2}^2(2n)}{2n\bar{X}_n}, \frac{\chi_{1-\alpha/2}^2(2n)}{2n\bar{X}_n} \right].$$

Partie 5

Devoirs

Devoir n°1

Ce devoir peut être abordé dès que les sections 1 et 2 du Chapitre 5 ont été travaillées (ainsi que les chapitres précédents bien sûr !).

Exercice 1

1) On considère le modèle statistique paramétrique des lois uniformes sur $[\theta, a]$ pour a fixé, i.e. $\mathcal{P} = \{\mathcal{U}_{[\theta, a]} : \theta < a\}$. On rappelle que la densité d'une loi $\mathcal{U}_{[\theta, a]}$ est :

$$f_{\theta}(x) = \frac{1}{a - \theta} \mathbb{1}_{[\theta, a]}(x).$$

On suppose que l'on observe un échantillon X_1, \dots, X_n dans ce modèle. Donner au moins deux estimateurs du paramètre θ . Quel est l'estimateur du maximum de vraisemblance dans ce modèle (justifier votre réponse) ?

2) Quel est l'estimateur du maximum de vraisemblance si on considère maintenant le modèle :

$$\mathcal{P} = \{\mathcal{U}_{[\theta, \theta+a]} : \theta > 0\},$$

où a est toujours un réel fixé ?

Exercice 2

On considère un modèle statistique paramétrique $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{P_a : a \in \mathbb{R}^+\})$, où P_a est une loi de probabilité continue de densité

$$f_a(x) = \frac{(k+1)x^k}{a^{k+1}} \mathbb{1}_{[0, a]}(x),$$

avec k un paramètre connu ($k > -1$).

On note X la v.a. générique dans ce modèle et X_1, \dots, X_n un échantillon dans ce modèle.

1) Montrer que f_a est bien une densité de probabilité.

2) Calculer $\mathbb{E}_a(X)$. En déduire un estimateur du paramètre a . S'agit-il d'un estimateur sans biais ? Que dire de sa consistance (on considérera la convergence *p.s* et dans L^2) ?

3) Quel est l'estimateur du maximum de vraisemblance de a (on justifiera bien qu'il s'agit d'un maximum) ?

4) Déterminer la fonction de répartition, puis la densité de la loi de la n ième statistique d'ordre $X_{(n)} = \max_{i=1, \dots, n} X_i$ pour cet échantillon. Calculer son espérance. Déduire de ceci et de la question précédente un estimateur sans biais de a .

Master Enseignement des Mathématiques,
CTU, Université de Franche-Comté,
Année 2011/2012.

Devoir n°2

Ce devoir peut être abordé dès que le Chapitre 5 a été travaillé (ainsi que les chapitres précédents bien sûr !).

Exercice 1

On considère le modèle constitué par l'ensemble des lois de Poisson de paramètre λ pour $\lambda > 0$. Déterminer l'estimateur du maximum de vraisemblance de la probabilité $P(X = 0)$. Cet estimateur est-il consistant ?

Exercice 2 On considère le modèle statistique paramétrique dont la v.a. générique X est discrète de loi définie, pour $k \in \mathbb{N}$, par :

$$P(X = k) = \frac{\theta^k}{(1 + \theta)^{k+1}},$$

où θ est un paramètre positif.

- 1) Montrer³ que $\mathbb{E}_\theta(X) = \theta$ et $\text{Var}_\theta(X) = \theta^2 + \theta$.
- 2) Donner un estimateur de θ par la méthode des moments.
- 3) Donner une statistique exhaustive pour le modèle.
- 4) Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ . On pourra admettre l'existence du maximum au zéro de la dérivée première.
- 5) Cet estimateur est-il sans biais ? Consistant ? Préciser les éventuels modes de convergence (*p.s.* ? dans L^2 ?).
- 6) Cet estimateur est-il efficace ?

³On pourra utiliser la somme de la série géométrique et pour la variance calculer $\mathbb{E}_\theta(X(X-1))$.

Master Enseignement des Mathématiques,
CTU, Université de Franche-Comté,
Année 2011/2012.

Devoir n°3

Ce devoir peut être résolu progressivement mais ne sera entièrement réalisable qu'une fois l'ensemble du cours et des exercices travaillés. Il peut servir de bon exercice de révision puisqu'il porte sur l'ensemble du programme.

Soit g une fonction de \mathbb{R} vers \mathbb{R} supposée positive, paire, intégrable et d'intégrale 1. Pour θ dans $[-1, 1]$, on désigne par P_θ la loi de probabilité sur $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ continue et de densité :

$$f_\theta(x) = g(x)[(1 - \theta)\mathbb{1}_{\mathbb{R}^-}(x) + (1 + \theta)\mathbb{1}_{\mathbb{R}^+}(x)].$$

On considère le modèle statistique $(\mathbb{R}, \mathcal{B}_\mathbb{R}, \{P_\theta : \theta \in [-1, 1]\})$ et $\underline{X} = (X_1, \dots, X_n)$ un échantillon dans ce modèle. On définit la statistique

$$T(\underline{X}) = \sum_{i=1}^n Y_i,$$

avec $Y_i = \mathbb{1}_{\mathbb{R}^+}(X_i)$, pour $i = 1, \dots, n$.

1) Que représente $T(\underline{X})$?

2) Montrer que la vraisemblance pour l'observation $\underline{x} = (x_1, \dots, x_n)$ de l'échantillon \underline{X} peut s'écrire sous la forme :

$$L(x_1, \dots, x_n; \theta) = (1 + \theta)^{T(\underline{x})} (1 - \theta)^{n - T(\underline{x})} \prod_{i=1}^n g(x_i).$$

3) Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_n(\underline{X})$ de θ (on justifiera bien qu'il s'agit d'un maximum).

4) Montrer que $\hat{\theta}_n(\underline{X})$ est un estimateur sans biais et consistant de θ .

5) L'estimateur $\hat{\theta}_n(\underline{X})$ est-il efficace ?

6) Étudier la convergence en loi de $\sqrt{n}(\hat{\theta}_n(\underline{X}) - \theta)$.

7) En déduire un intervalle de confiance asymptotique $1 - \alpha$ pour le paramètre θ .

Master Enseignement des Mathématiques,
CTU, Université de Franche-Comté,
Année 2011/2012.

Devoir n°4

Ce devoir peut être résolu progressivement mais ne sera entièrement réalisable qu'une fois l'ensemble du cours et des exercices travaillés. Il peut servir de bon exercice de révision puisqu'il porte sur l'ensemble du programme.

On considère le modèle de la loi Bêta à un seul paramètre :

$$(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{Beta(1, \theta) : \theta > 0\}).$$

On rappelle que la densité de la v.a. générique X dans ce modèle est

$$f_{\theta}(x) = \theta(1-x)^{\theta-1} \mathbb{1}_{[0,1]}(x).$$

On suppose que l'on observe un échantillon X_1, \dots, X_n dans ce modèle.

- 1) Donner un estimateur de θ par la méthode des moments
- 2) Écrire la vraisemblance de l'échantillon observé x_1, \dots, x_n .
- 3) Donner une statistique exhaustive.
- 4) Donner l'estimateur du maximum de vraisemblance de θ . On note $\hat{\theta}$ cet estimateur.
- 5) Montrer que la v.a. $Y = -\ln(1-X)$ est de loi exponentielle de paramètre θ .
- 6) L'estimateur $\hat{\theta}$ est-il consistant ?
- 7) Montrer que la v.a. $Z = \sum_{i=1}^n Y_i$, où $Y_i = -\ln(1-X_i)$ pour tout $i = 1, \dots, n$, est de loi Gamma de paramètre n et θ .
- 8) Calculer l'espérance de $\hat{\theta}$ et en déduire que

$$\tilde{\theta} = \frac{n-1}{Z}$$

est un estimateur sans biais de θ .

- 9) Ce dernier estimateur est-il consistant ?
- 10) Calculer l'information de Fisher apportée par l'échantillon.
- 11) L'estimateur $\tilde{\theta}$ est-il efficace ? Sinon, l'est-il asymptotiquement ?
- 12) On note \bar{Y} la moyenne empirique des v.a. Y_1, \dots, Y_n . Montrer que l'on a la convergence en loi

$$\sqrt{n}(\theta\bar{Y} - 1) \xrightarrow{\mathcal{L}} N(0, 1),$$

quand $n \rightarrow +\infty$.

- 13) En déduire un intervalle de confiance asymptotique $1 - \alpha$ pour θ .

Quelques rappels

- On rappelle que la fonction Bêta est définie par

$$\beta(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx,$$

pour tout a et b strictement positifs et que l'on a la relation

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

où

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx.$$

On rappelle que l'on a également : $\Gamma(a+1) = a\Gamma(a)$ pour tout a strictement positif.

- Une v.a. X est dite de loi exponentielle de paramètre λ si elle est absolument continue de densité

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x).$$

- Une v.a. X est dite de loi Gamma de paramètres α et β si elle est absolument continue de densité

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\mathbb{R}^+}(x).$$

- Une v.a. X est dite de loi de Poisson de paramètres λ si elle est discrète avec les probabilités

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

pour $k \in \mathbb{N}$.

Master Enseignement des Mathématiques,
CTU, Université de Franche-Comté,
Année 2011/2012.

Devoir n°5

Ce devoir peut être résolu progressivement mais ne sera entièrement réalisable qu'une fois l'ensemble du cours et des exercices travaillés. Il peut servir de bon exercice de révision puisqu'il porte sur l'ensemble du programme.

On dit qu'une v.a.r. X est de loi de Pareto de paramètres α (supposé strictement supérieur à 1) et θ (supposé strictement positif) si elle est absolument continue de densité

$$f_{\alpha,\theta}(x) = (\alpha - 1)\theta^{\alpha-1}x^{-\alpha}\mathbb{1}_{[\theta,+\infty[}(x).$$

Cette loi est très utilisée en gestion de la qualité, en actuariat ou bien encore en théorie des files d'attente (par exemple pour la modélisation des réseaux internet). L'objet de ce problème est d'étudier cette loi et quelques problèmes d'inférence statistique dans ce cadre.

Partie A *Quelques résultats de probabilités (utiles dans la suite du problème)*

1) Montrer que la fonction de répartition d'une telle v.a.r. est non nulle seulement si $x > \theta$ et que dans ce cas elle peut s'écrire sous la forme :

$$F(x) = 1 - \left(\frac{\theta}{x}\right)^{\alpha-1}.$$

2) Soit Y la v.a.r. définie par

$$Y = \log\left(\frac{X}{\theta}\right).$$

Montrer que la loi de Y est une loi exponentielle⁴ de paramètre $\alpha - 1$.

Partie B *Inférence sur le paramètre α , en supposant θ connu*

Dans cette partie, on suppose que le paramètre θ est connu. On considère donc le modèle paramétrique

$$(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{P_\alpha : \alpha > 1\}),$$

où P_α est la loi de Pareto de paramètre α et θ . On suppose que l'on observe un échantillon X_1, \dots, X_n dans ce modèle.

1) Donner une statistique exhaustive dans ce modèle.

⁴On rappelle qu'une v.a.r. est dite de loi exponentielle de paramètre λ si elle est absolument continue de densité :

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x).$$

2) Donner l'estimateur du maximum de vraisemblance $\hat{\alpha}$. On justifiera bien qu'il s'agit d'un maximum.

3) Montrer que cet estimateur est fortement consistant.

4) Montrer, en utilisant les résultats sur le comportement asymptotique des estimateurs du maximum de vraisemblance, que l'on a :

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{\mathcal{L}} N(0, (\alpha - 1)^2),$$

quand $n \rightarrow +\infty$.

5) Retrouver ce résultat par une autre méthode⁵.

6) Donner un intervalle de confiance asymptotique 95% pour α .

Partie C *Inférence sur le paramètre θ , en supposant α connu*

Dans cette partie, on suppose que le paramètre α est connu et que θ ne l'est plus. On considère donc le modèle paramétrique

$$(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{P_\theta : \theta > 0\}),$$

où P_θ est la loi de Pareto de paramètre α et θ . On suppose que l'on observe un échantillon X_1, \dots, X_n dans ce modèle.

1) Donner une statistique exhaustive pour le paramètre θ .

2) Donner l'estimateur du maximum de vraisemblance de θ .

3) Calculer la loi de cet estimateur.

Partie D *Inférence dans le modèle général*

On considère maintenant le modèle paramétrique

$$(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}, \{P_{\alpha, \theta} : \alpha > 1, \theta > 0\}),$$

où $P_{\alpha, \theta}$ est la loi de Pareto de paramètre α et θ . On suppose que l'on observe un échantillon X_1, \dots, X_n dans ce modèle.

1) Donner une statistique exhaustive dans ce modèle.

2) Quel est l'estimateur du maximum de vraisemblance du paramètre multidimensionnel (α, θ) ?

⁵On rappelle que la variance d'une loi exponentielle $\mathcal{E}(\lambda)$ est $1/\lambda^2$.