Chapitre 4

Estimation ponctuelle

4.1 Introduction

Une partie de la statistique consiste à estimer ce qu'on appelle les *paramètres* d'une population. En voici quelques exemples :

***Exemple* 4.1.1** *Populations et de paramètres* :

Population	Paramètre
L'ensemble des salariés québécois	μ : la moyenne des revenus des salariés québécois
Un lot de boulons	σ : l'écart-type σ des longueurs des boulons
L'ensemble des étudiants de l'UQAM	*p* : la proportion *p* des étudiants qui demeurent chez leurs parents
L'ensemble des ménages d'une petite ville	*R* : le nombre moyen de postes de radio par personne
Les accidents de Montréal	λ : le taux λ d'accidents par heure à Montréal

Considérons le premier exemple, une population de salariés, et supposons qu'on veuille estimer la moyenne μ de leurs revenus. Puisqu'il serait trop coûteux d'interroger tous les salariés de la population, on se contente d'un échantillon de *n* personnes tirées au hasard dans la population. On se servira alors de la moyenne des *n* personnes de l'échantillon pour estimer μ, une idée raisonnable au point d'être banale : par quoi estimer la moyenne d'une population sinon par la moyenne de l'échantillon? Mais les problèmes d'estimation ne sont pas toujours aussi évidents et pour ceux qui ne le sont pas (et pour celui-là aussi) nous devons élaborer une façon plus formelle et objective de valider une procédure d'estimation. La question posée dans l'exemple suivant n'a pas de réponse évidente.

***Exemple* 4.1.2** On tire un échantillon de 10 ménages afin d'estimer le nombre de téléphones par personne dans la population. Supposons qu'on obtienne les données suivantes.

Ménage (i)	1	2	3	4	5	6	7	8	9	10
Nombre de personnes (y_i)	5	4	6	8	3	5	2	3	5	2
Nombre de téléphones (x_i)	2	3	3	2	4	4	2	3	2	3

Comment estimer le paramètre « nombre de téléphones par personne » ? Une façon consiste à diviser le nombre total de téléphones par le nombre total de personnes, soit

$$\frac{\sum_{i=1}^{10} y_i}{\sum_{i=1}^{10} x_i} = \frac{2+3+3+2+4+4+2+3+2+3}{5+4+6+8+3+5+2+3+5+2} = \frac{28}{43} = 0,65\,. \tag{4.1.1}$$

Mais on pourrait également songer à calculer le nombre de téléphones par personne dans chaque ménage, et en calculer ensuite la moyenne :

$$\frac{1}{10}\sum_{i=1}^{10}\frac{y_I}{x_i} = \frac{1}{10}\left(\frac{2}{5}+\frac{3}{4}+\frac{3}{6}+\frac{2}{8}+\frac{4}{3}+\frac{4}{5}+\frac{2}{2}+\frac{3}{3}+\frac{2}{5}+\frac{3}{2}\right) = 0,79 \tag{4.1.2}$$

Laquelle des deux façons est meilleure ? ■

C'est la première méthode qui est préférable. En quel sens? C'est l'une des questions sur lesquelles nous nous penchons dans ce chapitre.

Commençons par définir les termes et établir le contexte formellement.

Population

Le terme « population » est utilisé en statistique dans le sens habituel, hormis le fait que les membres d'une population statistique ne sont pas forcément des humains ou des animaux. Mais il y a une notion supplémentaire qui s'y ajoute. Les membres d'une population seront toujours (bien que parfois de façon implicite) identifiés aux valeurs d'une variable X. Pour une population de salariés, on peut s'intéresser à la variable X = salaires annuels. L'échantillon comprendra n salaires, que nous désignerons par $X_1, \ldots, X_n$. Ce sont n variables aléatoires.

Définition *Échantillon aléatoire simple*

Une suite de n variables aléatoires $X_1; \ldots; X_n$ est appelée *échantillon aléatoire simple si* $X_1; \ldots; X_n$ sont indépendantes et de même loi.

Explication *Ces variables sont de même loi parce que les tirages sont tous faits dans la même population : si X_i est le salaire de la i^e personne choisie, alors, pour tout i, la distribution de X_i est précisément la distribution des valeurs de la population. Les variables sont indépendantes lorsque les tirages se font avec remise. (Ce qu'on ne fait pas souvent en pratique; mais si la population est grande comparée à l'échantillon, le fait de tirer avec ou sans remise ne change pas grand-chose. L'hypothèse d'indépendance est vérifiée à peu près).* ■

La population sera identifiée à la fonction de répartition F qui dépendra d'un ou plusieurs *paramètres*. Nous désignerons donc la population par $F(x|\boldsymbol{\theta})$, où $\boldsymbol{\theta}$ peut être un *vecteur* de paramètres.

***Exemple* 4.1.3** Si on tire un échantillon de pièces électroniques afin d'estimer leur durée moyenne β, et X est la durée d'une pièce tirée au hasard, alors X suit une certaine loi dont l'espérance est précisément β. Si on suppose que X est de loi exponentielle, alors la « population » est identifiée à la loi $\mathcal{E}(\beta)$ et $F(x \mid \theta) = 1\text{-}e^{-x/\beta}$.

Autre exemple : Si le but d'un échantillonnage est d'estimer le revenu moyen μ et la variance σ^2 des ménages d'une population et X est le revenu d'un ménage tiré au hasard dans la population, alors X suit une loi dont l'espérance est μ, et la variance est un deuxième paramètre inconnu σ^2. Si on suppose que la distribution des revenus est normale, alors la « population » est identifiée à la loi $\mathcal{N}(\mu\ ;\ \sigma^2)$: $F(x \mid \boldsymbol{\theta})$ représente alors la fonction de répartition d'une variable aléatoire normale de paramètres $\boldsymbol{\theta} = (\mu\ ;\ \sigma^2)$. [Dans ce dernier exemple, il est possible qu'on ne veuille estimer que la moyenneμ, alors que la variance σ^2 n'est d'aucun intérêt en soi. Nous verrons plus tard qu'il sera néanmoins nécessaire d'estimer σ^2 afin d'évaluer la précision de l'estimateur de μ.] ■

4.2 Statistiques et estimateurs

Soit $X_1, \ldots, X_n$ un échantillon aléatoire simple provenant d'une population $F(x \mid \boldsymbol{\theta})$.

Définition *Statistique*

Toute variable aléatoire $T(X_1, \ldots, X_n)$, fonction de $X_1, \ldots, X_n$, est appelée ***statistique***.

Définition *Estimateur*

Une statistique dont le but est d'estimer un paramètre est appelée ***estimateur***.

Les équations (4.1.1) et (4.1.2) sont des exemples d'estimateurs (ce sont aussi des statistiques).

***Exemple* 4.2.1** Voici quelques exemples de statistiques calculées à partir d'un échantillon aléatoire simple $X_1; \ldots; X_n$ de n salaires.

Paramètre		Estimateur	Description
μ:	Moyenne des salaires de la population	$T_1(X_1; \ldots; X_n) = \bar{X} = (1/n)\sum_{i=1}^{n} X_i$	La moyenne des n salaires $X_1; \ldots; X_n$.
σ^2:	Variance des salaires de la population	$T_2(X_1; \ldots; X_n) = \hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 / n$	L'écart-type des n salaires $X_1; \ldots; X_n$.
	Médiane des salaires de la population	$T_3(X_1; \ldots; X_n) = Méd\,(X_1; \ldots; X_n)$	La médiane des n salaires $X_1; \ldots; X_n$.
p:	Proportion des salaires inférieurs à 60 000 $ dans la population	$T_4(X_1; \ldots; X_n) = \hat{p}$	La proportion des salaires inférieurs à 60 000 $ dans l'échantillon

Le traitement théorique d'un problème d'inférence portant sur une population $F(x \mid \theta)$ consiste à choisir un estimateur (par exemple, $\bar{X}$, $\hat{\sigma}^2$, $\hat{p}$, …, etc.) et associer à chaque valeur de la statistique choisie une « décision » à propos du paramètre. La « décision » peut prendre différentes formes, trois desquelles seront traitées dans ce manuel :

1) *Estimation ponctuelle* : On peut décider que le paramètre est égal à tel ou tel nombre. Par exemple, on tire un échantillon de 100 ménages dans un quartier afin d'estimer le revenu moyen μ des ménages du quartier. Le revenu moyen des 100 ménages de l'échantillon est $\bar{X} = 42\,350$ \$. On estime alors que μ est égale à 42 350 \$. C'est une estimation *ponctuelle*, par opposition à une estimation par *intervalle de confiance*, que nous illustrons ensuite.

2) *Estimation par intervalle* : L'estimation ponctuelle dans l'exemple ci-dessus est téméraire et a peu de chance d'être correcte : c'est $\bar{x}$ qui est égale à 42 350 \$, pas μ. Une affirmation plus circonspecte, comme, par exemple, « le revenu moyen μ se situe quelque part entre 40 250 \$ et 44 450 \$ », a de meilleures chances d'être vraie. L'intervalle [40 250 \$; 44 250 \$], dont on détaillera la construction plus loin, est appelé *intervalle de confiance*.

3) *Test d'hypothèses* : Il est parfois suffisant de conclure que la valeur du paramètre est ou n'est pas égale à un nombre fixé d'avance, un nombre ayant une signification pratique. Par exemple, si μ est le poids moyen du contenu de certaines boîtes de conserves sur lesquelles on affiche « Poids net : 250 g », il est bon de s'assurer que la moyenne μ est bien égale à 250 g. Un système de contrôle de la qualité aurait pour unique objet de détecter un écart éventuel par rapport à cette moyenne afin d'y remédier.

Estimation ponctuelle

L'estimation ponctuelle consiste à choisir un estimateur et à déterminer ses propriétés. L'estimation ponctuelle consiste à trouver un *estimateur* d'un paramètre inconnu θ, c'est-à-dire, une statistique dont les valeurs auraient tendance, en un sens que nous devrons préciser, à s'approcher du paramètre. Par exemple, la moyenne arithmétique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ est généralement utilisée comme estimateur de l'espérance mathématique μ des variables $X_1, X_2, \ldots, X_n$, et la « variance » échantillonnale $S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ est utilisée comme estimation de leur variance.

- Face à deux estimateurs d'un même paramètre, comment décider lequel est « meilleur »? Et qu'est-ce qu'on entend par « meilleur »? Nous définirons des critères par lesquels on évaluera les qualités d'un estimateur;
- Il est souvent possible de proposer un estimateur par intuition—pour ensuite s'assurer qu'il a les propriétés voulues—, mais pas toujours. L'intuition n'est parfois d'aucun secours. Mais il existe des techniques qui permettent de découvrir des estimateurs. Nous en présenterons deux.

Estimation par intervalle de confiance
L'estimation par *intervalle de confiance* consiste à déterminer deux bornes, *LI* et *LS*, toutes deux fonctions des observations, et affirmer que le paramètre se situe entre ces deux bornes. Une telle affirmation peut, bien sûr, être erronée, mais on déterminera les bornes de façon que la probabilité d'erreur soit faible.

Test d'hypothèse

Un *test d'hypothèse* consiste à déterminer une règle pour décider quand une hypothèse H_o concernant un paramètre doit être rejetée. Par exemple,

« *Rejeter l'hypothèse* H_o *que* $\mu = 250$ *si* $\bar{X} < 240$ »

est une règle, ou un *test statistique*.

Dans ce chapitre, nous traiterons du problème d'estimation ponctuelle. La notion d'intervalle de confiance sera discutée au chapitre 5, et celle de test d'hypothèse au chapitre 6.

Dans certains cas, le choix d'un estimateur est naturel et intuitif : nous estimons la *moyenne* μ d'une population par la *moyenne* $\bar{X}$ de l'échantillon; et nous estimons une *probabilité* de succès par la *proportion* de succès dans l'échantillon. Mais comment exprimer objectivement les qualités qui nous font choisir ces estimateurs? Deux de ces qualités sont définies dans les deux prochaines sections.

4.3 Estimateurs sans biais

L'une des qualités généralement souhaitées d'un estimateur est celui d'être sans biais :

Définition *Estimateur sans biais*

Un estimateur $\hat{\theta}$ est dit ***sans biais*** pour θ si

$$E(\hat{\theta}) = \theta,$$

quelle que soit la valeur de θ.

Cette propriété est souhaitable parce qu'elle signifie que l'estimateur n'a tendance ni à sous-estimer ni à surestimer le paramètre : *en moyenne* il vise juste. Un estimateur qui ne possède pas cette propriété est dit *biaisé*.

Définition *Le biais d'un estimateur*

Le ***biais*** d'un estimateur $\hat{\theta}$ est défini par

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Si le biais n'est pas nul, on souhaite au moins qu'il soit petit et qu'il tende à disparaître lorsque l'échantillon est grand. Un tel estimateur est dit *asymptotiquement sans biais :*

Définition *Estimateur asymptotiquement sans biais*

Un estimateur $\hat{\theta}$ est dit ***asymptotiquement sans biais*** pour θ si

$$\lim_{n\to\infty} \mathrm{B}(\hat{\theta}) = 0$$

Nous montrons maintenant que les estimateurs $\bar{X}$, S^2 et $\hat{p}$ sont des estimateurs sans biais d'une moyenne μ, d'une variance σ^2, et d'une proportion p.

Estimateur sans biais d'une moyenne

Le théorème suivant présente un estimateur sans biais de la moyenne μ d'une population.

Théorème 4.3.1 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple provenant d'une population de moyenne μ. Alors

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ est un estimateur sans bais de } \mu.$$

Démonstration $\mathrm{E}(\bar{X}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$. *C'est ce qu'il fallait démontrer.* ■

Estimateur sans biais d'une variance

Le théorème suivant montre que la variance échantillonnale $\hat{\sigma}^2$ n'est pas un estimateur sans biais de la variance σ^2 d'une population, et en propose une alternative.

Théorème 4.3.2 Soit $\{X_1; X_2; \ldots; X_n\}$ un échantillon aléatoire provenant d'une population de moyenne μ et de variance σ^2. Alors

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

est un estimateur sans bais de σ^2.

Démonstration : Il suffit de montrer que $\mathrm{E}[(n-1)S^2] = (n-1)\sigma^2$. Rappelons que pour toute variable aléatoire W, $\mathrm{E}(W^2) = \mathrm{Var}[W] + [\mathrm{E}(W)]^2$. Maintenant, puisque $(n-1)S^2 = \sum_{i=1}^{n} x_i^2 - n\bar{X}^2$, nous avons $\mathrm{E}[(n-1)S^2] = \mathrm{E}\left[\sum_{i=1}^{n} x_i^2 - n\bar{X}^2\right] = \mathrm{E}\left(\sum_{i=1}^{n} x_i^2\right) - n\mathrm{E}(\bar{X}^2) = \sum_{i=1}^{n}\mathrm{E}(x_i^2) - n\mathrm{E}(\bar{X}^2) = \sum_{i=1}^{n}(\sigma^2+\mu^2) - n\mathrm{E}(\bar{X}^2) = n(\sigma^2+\mu^2) - n[\mathrm{Var}(\bar{X}) + (\mathrm{E}(\bar{X}))^2] = n(\sigma^2+\mu^2) - n[\sigma^2/n + \mu^2] = (n-1)\sigma^2$. Donc $\mathrm{E}[(n-1)S^2] = (n-1)\sigma^2 \Rightarrow \mathrm{E}(S^2) = \sigma^2$. C'est ce qu'il fallait démontrer. ■

Remarque *On peut concevoir d'une logique qui ferait préférer un estimateur biaisé à un estimateur sans biais. Supposons qu'on tire un échantillon de n ménages dans une petite ville afin d'estimer le revenu moyen* μ. *Sachant qu'une petite poignée de ménages a un revenu extrêmement élevé, et que la présence de l'un d'eux dans l'échantillon mènerait à une surestimation grossière de* μ, *on décide d'éliminer systématiquement de l'échantillon la plus grande valeur observée et d'estimer* μ *par la moyenne des* n-1 *données restantes. Cette approche aurait tendance à sous-estimer* μ, *mais a l'avantage de réduire les risques d'une grossière erreur d'estimation.* ■

Estimateur sans biais d'une proportion *p*

On prélève un échantillon de taille n d'une grande population de pièces fabriquées, afin d'estimer la proportion p de pièces défectueuses dans la population. Si X est le *nombre* de pièces défectueuses dans l'échantillon, il est naturel de considérer comme estimateur de p la proportion $\hat{p} = \frac{X}{n}$ de pièces défectueuses dans l'échantillon.

Théorème 4.3.3 Soit p la proportion des individus d'une population qui appartiennent à une certaine classe C. Soit X le nombre d'individus qui appartiennent à la classe C dans un échantillon de taille n. Alors

$$\hat{p} = \frac{X}{n} \text{ est un estimateur sans bais de } p.$$

Démonstration : Si la population est grande ou si on tire avec remise, $X \sim \mathcal{B}(n; p)$, et donc $E(X) = np$. Alors $E(\hat{p}) = np/n = p$. ■

Remarque *Le fait qu'un estimateur soit sans biais ne suffit pas à en faire un estimateur raisonnable. Supposons que $X_1, \ldots, X_{10}$ est un échantillon de taille 10 d'une population de moyenne μ. Alors $\bar{X}$, la moyenne des 10 données, est un estimateur sans biais de μ. Mais $(X_1 + X_2)/2$, la moyenne des deux premières données est aussi un estimateur sans biais. Il est évident que ce n'est pas un estimateur raisonnable. Mais, objectivement, quel est son défaut? Comment quantifier sa faiblesse par rapport à $\bar{X}$? Un deuxième critère permet de distinguer cet estimateur de la moyenne $\bar{X}$.*

4.4 Variance d'un estimateur

Le fait qu'un estimateur est sans biais, quoique rassurant, ne garantit pas une bonne précision. Si un estimateur sans biais prend *en moyenne* la valeur juste, il n'est pas dit qu'il ne peut pas s'en éloigner — de beaucoup et souvent. La moyenne $\bar{X}$ d'un échantillon aléatoire simple $X_1, \ldots, X_n$ n'est pas le seul estimateur sans biais de la moyenne μ : chacune des observations X_i, par exemple, est un estimateur sans biais de μ. Pourtant on ne songerait pas à utiliser X_1, la première observation, comme estimateur de μ : Il est intuitivement évident que $\bar{X}$ est préférable à un estimateur basé sur une seule des observations. On conçoit qu'il a une plus forte tendance à rester près de μ. C'est là une autre caractéristique souhaitable d'un estimateur sans biais: qu'il ait tendance à rester près du paramètre. Autrement dit, qu'il ait, autant que possible, une variance *petite*. La variance d'un estimateur sans biais est un indice de sa précision.

De quoi dépend la variance ? Nous le verrons dans le cas des estimateurs $\bar{X}$, $\hat{p}$ et S^2.

Variances des estimateurs $\bar{X}$, $\hat{p}$ et S^2

Théorème 4.4.1 La variance de l'estimateur $\bar{X}$ de la moyenne μ d'une population est donnée par

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Théorème 4.4.2 *Variance de* $\hat{p}$

La variance $\hat{p} = \dfrac{X}{n}$ de l'estimateur d'une proportion p est donnée par

$$\mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

Théorème 4.4.3 *Variance de* S^2

Si *la population est normale*, la variance de S^2 est donnée par

$$\mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

Ce qu'il faut remarquer ici, c'est le fait que n est au dénominateur: plus n augmente, plus la variance est petite. Ceci correspond à ce que l'on sait déjà par intuition : une estimation est d'autant meilleure que l'échantillon est grand.

Erreur quadratique moyenne

Pour un estimateur sans biais, la variance est utile comme mesure de précision car elle mesure la dispersion de l'estimation par rapport au paramètre:

$$\mathrm{Var}(\hat{\theta}) = \mathrm{E}\{[\hat{\theta} - \mathrm{E}(\hat{\theta})]^2\} = \mathrm{E}[(\hat{\theta} - \theta)^2] \ .$$

Mais lorsque l'estimateur est biaisé, la dispersion par rapport au paramètre est mesurée par *l'erreur quadratique moyenne*.

Définition *Erreur quadratique moyenne*

L*'erreur quadratique moyenne* EQM($\hat{\theta}$) d'un estimateur $\hat{\theta}$ est définie par

$$\mathrm{EQM}(\hat{\theta}) = \mathrm{E}[(\hat{\theta} - \theta)^2]$$

Le théorème suivant décompose l'erreur quadratique moyenne en deux parties: la variance et le carré du biais.

Théorème 4.4.1 *Erreur quadratique moyenne et variance*

$$\mathrm{EQM}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + [\mathrm{B}(\hat{\theta})]^2$$

Démonstration : $\mathrm{EQM}(\hat{\theta}) = \mathrm{E}(\hat{\theta} - \theta)^2 = \mathrm{E}\left\{[(\hat{\theta} - \mathrm{E}(\hat{\theta})) + (\mathrm{E}(\hat{\theta}) - \theta)]^2\right\} = \mathrm{E}[(\hat{\theta} - \mathrm{E}(\hat{\theta}))^2] + \mathrm{E}[(\mathrm{E}(\hat{\theta}) - \theta)^2] +$ $2\mathrm{E}[(\hat{\theta} - \mathrm{E}(\hat{\theta}))(\mathrm{E}(\hat{\theta}) - \theta)] = \mathrm{Var}(\hat{\theta}) + [\mathrm{B}(\hat{\theta})]^2 + 2\,[\mathrm{E}(\hat{\theta}) - \theta]\mathrm{E}[\hat{\theta} - \mathrm{E}(\hat{\theta})] = \mathrm{Var}(\hat{\theta}) + [\mathrm{B}(\hat{\theta})]^2$, étant donné que $\mathrm{E}[\hat{\theta} - \mathrm{E}(\hat{\theta})] = 0$. ■

Remarque *Quels sont les facteurs qui contribuent à la précision d'un estimateur ?*

Nous avons dit qu'un estimateur sans biais est précis lorsque sa variance est faible. Quels sont donc les facteurs qui contribuent à sa précision ? Notons d'abord que pour les trois estimateurs discutés ici, la variance décroît lorsque n croît : $\mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$; $\mathrm{Var}(\hat{p}) = \dfrac{p(1-p)}{n}$; $\mathrm{Var}(S^2) = \dfrac{2\sigma^4}{n-1}$.

Outre l'effet de n, un autre facteur intervient :

Dans le cas de $\bar{X}$ *et de* S^2*, l'écart-type σ de la population est un facteur important: la précision de l'estimateur est d'autant meilleure que les données de la population sont peu dispersées. On compense pour un σ trop grand en augmentant la taille de l'échantillon.*

Dans le cas de $\hat{p}$*, c'est le produit p(1-p) qui influence la précision de l'estimateur.* $\bar{X}$ *Quant à* $\hat{p}$*, sa précision dépend du produit p(1-p), qui est faible lorsque p est proche de 0 ou de 1 et atteint son maximum de 1/4 lorsque p = 1/2.* ■

Coefficient de variation

La qualité d'un estimateur $\hat{\theta}$ est normalement mesurée non pas par son écart-type $\sqrt{\mathrm{Var}(\hat{\theta})}$, mais par son coefficient de variation $\mathrm{cv}(\hat{\theta})$, défini par

$$\mathrm{cv}(\hat{\theta}) = \frac{\sqrt{\mathrm{Var}(\hat{\theta})}}{\theta}.$$

L'écart-type peut être grand ou petit dépendant de l'unité de mesure alors que le coefficient de variation est invariable.

Jusqu'ici, les estimateurs dont nous avons examiné les propriétés ont été proposés par intuition : ils viennent naturellement à l'esprit en vertu de leur ressemblance au paramètre. Mais l'intuition seule ne réussit pas toujours à identifier un estimateur potentiel. Supposons que vous disposez d'un échantillon d'une population de loi gamma de paramètres α et β. Comment estime α ? Comment estimer β ? Dans ce cas, il est nécessaire de disposer d'un outil qui mène à un estimateur. Nous présentons ici deux méthodes : la ***méthode des moments*** et la ***méthode du maximum de vraisemblance***.

4.5 La méthode des moments

La méthode des moments s'appuie sur deux principes élémentaires :

- il est raisonnable d'estimer la moyenne de la population par la moyenne de l'échantillon;
- s'il est raisonnable d'estimer θ par $\hat{\theta}$, alors il est raisonnable d'estimer une fonction $g(\theta)$ par $g(\hat{\theta})$.

Ainsi donc, si les observations $X_1, X_2, \ldots, X_n$ sont de moyenne μ, il est raisonnable d'estimer μ par la moyenne échantillonnale $\bar{X}$. Et si c'est μ^2 qu'on veut estimer, on l'estimera par $\bar{X}^2$. On ne prétend pas que cette façon de déterminer des estimateurs est optimale—il y en aura souvent de meilleures—mais elle est « raisonnable » et ne donne pas de trop mauvais résultats. Sa vertu principale est son attrait intuitif et sa simplicité.

***Exemple* 4.5.1** Soit $X_1, X_2, \ldots, X_n$ un échantillon aléatoire simple d'une population dont la fonction de densité est $f(x) = \theta e^{-\theta x}$, $x > 0$. Comment estimer θ ?

Solution On sait comment estimer une moyenne $\mu = E(X)$: l'estimateur recommandé dans la section précédente est la moyenne échantillonnale $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Or $E(X) = 1/\theta$, et donc $\theta = 1/\mu$. Donc pour estimer 1/μ, on replace μ par $\bar{X}$, ce qui donne l'estimateur $\hat{\theta} = 1/\bar{X}$. Cet estimateur n'est pas sans biais. ■

Remarque *Le dernier exemple montre que rien ne garantit que cette méthode mène à un estimateur sans biais.* $\bar{X}^2$ *n'est pas sans biais pour* μ^2 ■

L'approche décrite dans l'exemple illustre la méthode dite *des moments*. Soit une population dont la distribution dépend d'un vecteur de paramètres $\boldsymbol{\theta}$; et un échantillon aléatoire simple X_1, X_2, … , X_n tiré de cette population. On part du principe qu'on peut toujours estimer, sans biais, une espérance mathématique par une moyenne arithmétique. En particulier, l'espérance μ d'une variable aléatoire X peut être estimée par $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$; et l'espérance de X^2 par $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$. En général, on peut estimer le k^e moment $E(X^k)$ par $\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$. Or les moments sont fonctions des paramètres. Si, comme dans l'exemple, θ est scalaire, et $E(X) = a(\theta)$, alors on pose $\bar{X} = a(\theta)$ et la solution de cette équation est l'estimateur $\hat{\theta}$ de θ : $\hat{\theta} = a^{-1}(\bar{X})$.

Si $\boldsymbol{\theta} = [\theta_1; \theta_2]$ est de dimension 2, et les deux premiers moments sont $E(X) = a(\boldsymbol{\theta})$ et $E(X^2) = b(\boldsymbol{\theta})$, alors $\hat{\boldsymbol{\theta}}$ est la solution des équations

$$\bar{X} = a(\boldsymbol{\theta}) \text{ et } \overline{X^2} = b(\boldsymbol{\theta})$$

En général, si $\boldsymbol{\theta}$ est de dimension m, et si le j^e moment est $E(X^j) = a_j(\boldsymbol{\theta})$, alors $\hat{\boldsymbol{\theta}}$ est la solution des équations

$$\bar{X} = a_1(\boldsymbol{\theta}) \; ; \; \overline{X^2} = a_2(\boldsymbol{\theta}) \; ;...; \; \overline{X^m} = a_m(\boldsymbol{\theta})$$

***Exemple* 4.5.2** Soit X_1; X_2; …; X_n un échantillon aléatoire simple d'une population normale de moyenne μ et de variance σ^2. Déterminer les estimateurs par les moments de μ et de σ^2.

Solutions On sait qu'alors $E(X_i) = \mu$ et $E(X_i^2) = \mathrm{Var}(X_i) + \mu^2 = \sigma^2 + \mu^2$. On pose alors $\bar{X} = \mu$ et $\overline{X^2} = \sigma^2 + \mu^2$, ce qui mène aux estimateurs suivants : $\hat{\mu} = \bar{X}$ et $\hat{\sigma}^2 = \overline{X^2} - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$. On remarque que cet estimateur est légèrement biaisé.■

4.6 La méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance estime le paramètre par la valeur pour laquelle l'échantillon observé est le plus probable. Le prochain exemple illustre la méthode dans le cas où l'échantillon a été résumé par une seule variable.

***Exemple* 4.6.1** Supposons que la valeur observée d'une variable X de loi binomiale de paramètres $n = 20$ et p inconnu est $X = 7$. Certaines valeurs extrêmes de p peuvent être écartées d'emblée. Par exemple, les hypothèses $p = 0{,}01$ ou $p = 0{,}99$. Pourquoi ? Parce que l'observation $X = 7$ est trop peu probable sous ces hypothèses. Nous ne retenons pour p que les valeurs qui rendent l'observation relativement probables. En fait, l'estimateur $\hat{p}$ de p est la valeur qui maximise la probabilité $\varphi(p) = P(X = 7 \mid p)$. On évalue cette probabilité pour quelques valeurs de p :

p	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8
$\varphi(p)$	0,00197	0,05455	0,16426	0,16588	0,07393	0,01456	0,00102	0,00001

La fonction $\varphi(p)$ est maximisée pour une valeur de p se situant entre 0,3 et 0,4. On peut la déterminer exactement : le maximum de $\varphi(p) = \binom{20}{7} p^7(1-p)^{13}$ est atteint lorsque $\varphi'(p) = 0$.

$\varphi'(p) = \binom{20}{7}[7p^6(1-p)^{13} - 13p^7(1-p)^{12}]$ et donc $\varphi'(p) = 0 \Leftrightarrow 7(1-p) = 13p \Leftrightarrow p = 7/20$. ■

Il se trouve que dans cet exemple, l'observation a été réduite à une seule valeur X et c'est la probabilité de la valeur observée que nous devions maximiser par rapport au paramètre. Normalement, un échantillon est une suite de valeurs X_1; X_2; …; X_n et la fonction à maximiser est la *fonction de vraisemblance*, dont la définition suit.

Définition *Fonction de vraisemblance*

La *fonction de vraisemblance* d'un échantillon aléatoire simple $X_1, X_2, \ldots, X_n$ est définie par

$$L(\boldsymbol{\theta}) = L(x_1;x_2; \ldots, x_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i \mid \boldsymbol{\theta})$$

dans la cas continu (f étant la densité de chaque X_i); ou

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} p(x_i \mid \boldsymbol{\theta})$$

dans le cas discret, p étant la fonction de probabilité de chaque X_i.

Comme dans le dernier exemple, nous allons maximiser la fonction de vraisemblance par rapport à $\boldsymbol{\theta}$:

Définition L'*estimateur du maximum de vraisemblance* d'un paramètre $\boldsymbol{\theta}$ est la valeur $\hat{\theta}$ qui maximise la fonction de vraisemblance $L(\boldsymbol{\theta})$.

***Exemple* 4.6.2** Soit X_1 ; X_2 ; …; X_n un échantillon aléatoire simple d'une population de loi exponentielle de paramètre β ($f(x) = \frac{1}{\beta} e^{-x/\beta}$, $x > 0$). Déterminer l'estimateur du maximum de vraisemblance.

Solution La fonction de vraisemblance est $L(\beta) = \prod_{i=1}^{n} \frac{1}{\beta} e^{-x_i/\beta} = \frac{1}{\beta^n} e^{-\sum x_i/\beta} = \frac{1}{\beta^n} e^{-n\bar{x}/\beta}$. La valeur de β qui maximise $L(\beta)$ maximise également le logarithme de la fonction de vraisemblance, $\ell(\beta) = \ln L(\beta)$. Nous allons donc maximiser $\ell(\beta) = -n \ln \beta - n\bar{x}/\beta$. La dérivée de $\ell(\beta)$ est $\ell'(\beta) = -n/\beta + n\bar{x}/\beta^2$ et $\ell'(\beta) = 0 \Leftrightarrow \beta = \bar{x}$.

L'estimateur du maximum de vraisemblance est donc $\hat{\beta} = \bar{X}$. ■

Paramètres d'une loi binomiale

On montre que la proportion échantillonnale $\hat{p}$ est l'estimateur du maximum de vraisemblance de p.

Théorème 4.6.1 *Estimation d'une proportion*

Soit X une observation de loi $\mathcal{B}(n\,;p)$. L'estimateur du maximum de vraisemblance de p est $\hat{p} = \frac{X}{n}$.

Démonstration

Il s'agit de maximiser la fonction de vraisemblance $L(p) = \binom{n}{x} p^x (1-p)^{n-x}$ par rapport à p. Il est plus simple de maximiser le logarithme $\ell(p) = \ln L(p) = \ln\binom{n}{x} + x \ln p + (n-x)\ln(1-p)$, ce qui revient au même étant donné que la fonction logarithmique est croissante. On dérive par rapport à p: $\ell'(p) = \frac{x}{p} - \frac{(n-x)}{(1-p)} = \frac{x(1-p)-(n-x)p}{p(1-p)} = \frac{x-np}{p(1-p)}$. Alors $\ell'(p) = 0 \Rightarrow \frac{x-np}{p(1-p)} = 0 \Rightarrow X - np = 0 \Rightarrow p = \frac{x}{n}$. ■

Paramètres d'une population normale

On montre ici que les estimateurs du maximum de vraisemblance de μ et de σ sont précisément ceux obtenus par la méthode des moments.

Théorème 4.6.2 *Paramètres d'une population normale*

Soit X_1, X_2, …, X_n un échantillon aléatoire simple d'une population de loi $\mathcal{N}(\mu;\ \sigma^2)$. Les estimateurs du maximum de vraisemblance de μ et de σ^2 sont

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ \text{ et } \ \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}.$$

Démonstration

Nous commençons par démontrer que $\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2$.

$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}[(X_i-\bar{X}+\bar{X}-\mu)]^2 = \sum_{i=1}^{n}(X_i-\bar{X})^2 + \sum_{i=1}^{n}(\bar{X}-\mu)^2 + 2\sum_{i=1}^{n}(X_i-\bar{X})(\bar{X}-\mu) =$
$\sum_{i=1}^{n}(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2 + 2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\bar{X}) = \sum_{i=1}^{n}(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2$, car $\sum_{i=1}^{n}(X_i-\bar{X}) = 0$.

Maintenant, nous maximisons le logarithme de la fonction de vraisemblance

$$L = \prod \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i-\mu)^2\, 2\sigma^2} . = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\Sigma(x_i-\mu)^{2/2\sigma^2}} .$$

$$\Rightarrow \ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 - \frac{n(\bar{x}-\mu)^2}{2\sigma^2} .$$

Pour tout σ^2 fixe, ln L est maximisée lorsque le dernier terme est nul, donc lorsque $\mu = \bar{x}$, ce qui donne l'estimateur de μ : $\hat{\mu} = \bar{X}$.

Il reste à maximiser $\varphi(\sigma^2) = \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2$ par rapport à σ^2, le premier terme, $-\frac{n}{2}\ln 2\pi$, étant indépendant de σ^2.

La dérivée de $\varphi(\sigma^2)$ est $-\frac{n}{2\sigma^2} + \frac{\Sigma(x_i-\bar{x})^2}{2\sigma^4}$. La solution de $\varphi'(\sigma^2) = 0$ est bien $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$.

On vérifie aisément que la dérivée seconde de φ est négative à $\hat{\sigma}^2$, confirmant ainsi qu'il s'agit bien d'un maximum. On voit ici que l'estimateur du maximum de vraisemblance n'est pas nécessairement sans biais. ■

4.7 L'avantage d'un modèle

On illustre ici ce que l'utilisation d'un modèle, généralement sous la forme d'une loi génératrice des observations, peut apporter comme avantage: une cohérence et une parcimonie certaines dans la description des résultats d'une étude.

L'estimation de plusieurs proportions —comme, par exemple, lorsqu'on souhaite estimer des probabilités du genre $P(a_i < X \leq b_i)$ pour plusieurs intervalles $]a_i\ ;\ b_i]$—donne parfois des estimations peu crédibles et peu cohérentes.

Supposons qu'un fabricant de chaussures, voulant exploiter un nouveau marché, doit estimer la distribution des longueurs de pied de la population visée. Il s'agirait d'estimer les fréquences des classes suivantes (en centimètres), correspondant aux 11 différentes pointures prévues:

≤ 30; 30-35; 35-40; 40-45; 45-50; 50-55; 55-60; 60-65; 65-70; 70-75; >75

Afin d'estimer ces effectifs, on obtient les longueurs de pieds d'un échantillon de 100 personnes que voici :

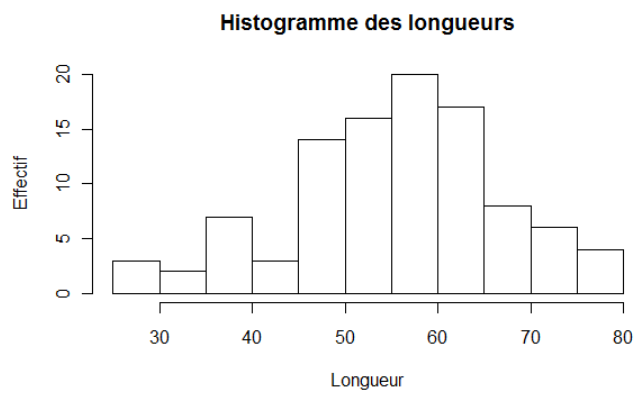
29,0	35,1	38,1	45,5	46,7	48,8	50,5	52,3	53,6	55,1	56,6	58,4	59,2	60,2	61,7	63,2	64,8	67,3	70,4	74,4
29,0	35,6	38,4	45,7	47,8	49,0	50,5	52,6	53,8	55,4	56,6	58,4	59,2	60,2	61,7	63,5	65,0	68,1	71,6	78,7
30,0	36,1	41,7	45,7	48,3	49,5	50,8	53,1	54,6	55,6	57,1	58,9	59,7	60,5	61,7	63,8	65,3	68,1	72,9	79,0
31,0	36,6	42,4	46,5	48,5	49,5	51,8	53,3	54,9	56,1	57,7	58,9	59,7	60,7	62,5	64,8	66,3	68,3	73,2	80,0
34,8	37,1	43,7	46,7	48,8	50,5	51,8	53,6	54,9	56,4	58,2	59,2	59,9	61,0	63,0	64,8	66,5	68,8	73,4	80,0

Estimations immédiates : les fréquences observées

On peut estimer les fréquences des classes par les fréquences observées dans l'échantillon. On obtient les estimations suivantes :

Intervalle	Fréquence	Intervalle	Fréquence	Intervalle	Fréquence
≤ 30	0,03	45 - 50	0,14	65 - 70	0,8
30 - 35	0,02	50 - 55	0,16	70 - 75	0,06
35 - 40	0,07	55 - 60	0,2	> 75	0,04
40 - 45	0,03	60 - 65	0,17		

Voici donc une représentation par histogramme de la distribution estimée :



L'inconvénient de cette approche, c'est le grand nombre d'estimations qu'elle exige, autant d'occasions de commettre des erreurs. On note en particulier, dans cet histogramme, certains « creux » (entre 30 et 35 et entre 40 et 45) auxquels on ne peut vraiment croire, puisqu'il est presque certain que la distribution réelle est unimodale et probablement symétrique, à peu de choses près. On admet même facilement qu'une courbe normale *pourrait* assez fidèlement approcher la vraie distribution. Si on accepte cette hypothèse, nous n'avons plus qu'à estimer les deux paramètres, μ et σ^2 : de là on déduit les fréquences des classes.

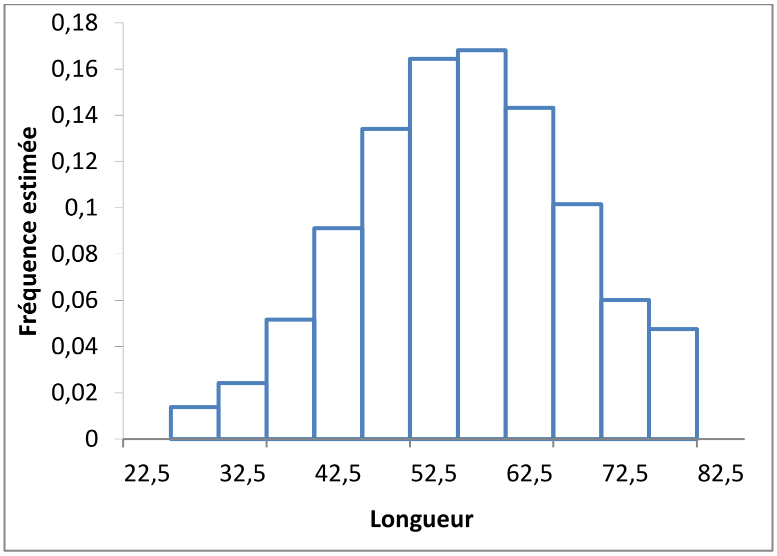
Estimation à l'aide d'un modèle normal

On estime μ par la moyenne de l'échantillon, qui est 55,599; et on estime σ par l'écart-type de l'échantillon S = 11,61715.

Selon la loi normale, les probabilités de ces intervalles (et donc les fréquences) sont les suivantes :

Intervalle	Fréquence	Intervalle	Fréquence	Intervalle	Fréquence
≤ 30	0,0138	45 - 50	0,1341	65 - 70	0,1016
30 - 35	0,0243	50 - 55	0,1645	70 - 75	0,0601
35 - 40	0,0516	55 - 60	0,1682	> 75	0,0475
40 - 45	0,0911	60 - 65	0,1432		

Les fréquences estimées ont maintenant une allure plus crédible :



■

Résumé

1 *Estimateur sans biais* Un estimateur $\hat{\theta}$ est dit ***sans biais*** pour θ si E($\hat{\theta}$) = θ quelle que soit la valeur de θ.

Paramètre θ	Estimateur $\hat{\theta}$	Var($\hat{\theta}$)
μ	$\bar{X}$	$\mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$
σ^2	S^2	$\mathrm{Var}(S^2) = \dfrac{2\sigma^4}{n-1}$ (si la population est normale)
p	$\hat{p} = \dfrac{X}{n}$	$\mathrm{Var}(\hat{p}) = \dfrac{\hat{p}(1-\hat{p})}{n}$

EXERCICES

4.1 Soit $X_1, \ldots, X_n$ un échantillon aléatoire tiré d'une population de moyenne μ et de variance σ^2. Considérons l'ensemble des estimateurs de la forme $\hat{\mu} = \sum_{i=1}^{n} a_i X_i$ où $a_1, \ldots, a_n$ sont des constantes.

a) Quelles conditions les a_i devraient-elles satisfaire pour que $\hat{\mu}$ soit sans biais ?

b) Parmi tous les estimateurs sans biais de la forme $\hat{\mu} = \sum_{i=1}^{n} a_i X_i$, montrer que celui qui a la plus petite variance est $\bar{X}$. [*Suggestion* : Déduire de l'inégalité $\sum_{i=1}^{n}\left(a_i - \frac{1}{n}\right)^2 \geq 0$ que $\sum a_i^2 \geq 1/n$.]

4.2 On doit prélever deux échantillons de deux populations de moyennes μ_1 et μ_2 et variances connues σ_1^2 et σ_2^2. On a l'intention d'estimer $\mu_1 - \mu_2$ par $\bar{X}_1 - \bar{X}_2$. On dispose d'un budget d'échantillonnage limité, de sorte que les tailles n_1 et n_2 des deux échantillons doivent satisfaire $n_1 + n_2 = 100$. Comment doit-on choisir n_1 et n_2 ?

4.3 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de moyenne μ. Montrez que les estimateurs $T_1 = X_1$, $T_2 = 2X_1 - X_2$ et $T_3 = 2\sum_{i=1}^{n} \frac{i}{n(n+1)} X_i$ sont tous sans biais pour μ. Calculez leur variance, et dites pourquoi $\bar{X}$ est préférable à ces trois.

4.4 Soit $X_1; \ldots; X_n$ un échantillon aléatoire tiré d'une population de densité

$$f(x) = \frac{x}{\theta^2} e^{-x/\theta}; \quad x > 0, \theta > 0.$$

Vous cherchez un estimateur sans biais de θ et vous considérez $\bar{X}$ comme un premier essai. Montrez que $\bar{X}$ est biaisé, et montrez comment modifier cet estimateur pour en obtenir un qui est sans biais.

4.5 Soit $X_1; \ldots; X_n$ un échantillon aléatoire tiré d'une population de densité

$$f(x) = \frac{1}{\Gamma(\theta)2^{\theta}} x^{\theta-1} e^{-x/2}; \; x > 0, \theta > 0.$$

Vous cherchez un estimateur sans biais de θ et vous considérez $\bar{X}$ comme un premier essai. Montrez que $\bar{X}$ est biaisé, et montrez comment modifier cet estimateur pour en obtenir un qui est sans biais.

4.6 Soit $X_1; \ldots; X_n$ un échantillon aléatoire d'une population de loi $\mathcal{P}(\lambda)$.

a) Montrer que $\bar{X}$ et S^2 sont tous deux des estimateurs sans biais de λ.

b) Montrer que la loi conditionnelle du vecteur $\mathbf{X} = [X_1; X_2; \ldots; X_n]$ étant donné $\bar{X} = x/n$ est multinomiale de paramètres x et $[1/n; 1/n; \ldots; 1/n]$.

c) Montrer que $E(X_i \mid \bar{X}) = \bar{X}$, $\mathrm{Var}(X_i \mid \bar{X}) = \frac{(n-1)\bar{X}}{n}$, et que $E(X_i^2 \mid \bar{X}) = \frac{(n-1)\bar{X}}{n} + \bar{X}^2$.

d) Montrer que $E(S^2 \mid \bar{X}) = \bar{X}$.

e) Montrer que $\mathrm{Var}(S^2) \geq \mathrm{Var}(\bar{X})$
[*Suggestion* : Utiliser le fait que $\mathrm{Var}(S^2) = E\{\mathrm{Var}(S^2 \mid \bar{X})\} + \mathrm{Var}\{E(S^2 \mid \bar{X})\}$].

4.7 Soit $X_1; \ldots; X_n$ un échantillon aléatoire d'une population de loi de $\mathcal{G}(p)$.

a) Montrer que $\bar{X}$ est un estimateur sans biais de $1/p$.

b) Montrer que $\dfrac{1-1/n}{\overline{X}-1/n}$ est sans biais pour p. *Suggestion* : Se servir de développement de Taylor $(1\text{-}t)^r = \sum_{x=0}^{\infty}\binom{r-1+x}{r-1}t^x$.

4.8 Soit $\overline{X}_1$; …; $\overline{X}_k$ les moyennes de k échantillons indépendants de tailles n_1; …; n_k, tirés tous de la même population de moyenne μ et de variance σ^2.

a) Pour estimer μ on considère l'estimateur linéaire $\hat{\mu} = a_1\overline{X}_1 + a_2\overline{X}_2 + \ldots + a_k\overline{X}_k$, où a_1, …, a_k sont des constantes. Quelle(s) condition(s) les a_i doivent-elles satisfaire pour que $\hat{\mu}$ soit sans biais ?

b) Parmi tous les estimateurs linéaires sans biais, montrer que celui qui a la plus petite variance est celui où $a_i = n_i/n$ où $n = \sum_{i=1}^{n} n_i$.

[*Indice* : *Montrer que* $\sum_{i=1}^{n}\dfrac{a_i^2}{n_i} \geq \dfrac{1}{n}$ *à partir de l'énoncé* $\sum_{i=1}^{n} n_i\left(\dfrac{a_i}{n_i}-\dfrac{1}{n}\right)^2 \geq 0$.]

4.9 Si $X \sim \mathcal{B}(n; p)$ et $\hat{p} = \dfrac{X}{n}$, montrez que Var($\hat{p}$) ≤ $1/4n$.

4.10 Soit X_1; …;, X_n un échantillon d'une population $\mathcal{N}(\mu; \sigma^2)$. Dans cet exercice, on suppose μ connu, et on compare l'estimateur S^2 à l'estimateur $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$.

a) Montrez que $\dfrac{n\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2$ donc que $\mathrm{E}\left[\dfrac{n\hat{\sigma}^2}{\sigma^2}\right] = n$ et par conséquent $\hat{\sigma}^2$ est un estimateur sans biais.

b) Montrer que Var[$n\hat{\sigma}^2/\sigma^2$] = $2n$, et par conséquent Var[$\hat{\sigma}^2$] = $2\sigma^4/n$.

c) Montrer que Var[S^2] = $2\sigma^4/(n-1)$.

d) Quel est le meilleur estimateur, $\hat{\sigma}^2$ ou S^2 ?

4.11 Soit n observations d'une population de densité $f(x) = \dfrac{1}{\beta}e^{-x/\beta}$; $x > 0$, $\theta > 0$.

a) Montrer que $\overline{X}$ est sans biais pour β, et déterminer sa variance.

b) Soit $\theta = \beta^2$. Montrer que $\hat{\theta}_1 = \dfrac{1}{2n}\sum_{i=1}^{n}X_i^2$, $\hat{\theta}_2 = S^2 = \dfrac{\sum(X_i-\overline{X})^2}{n-1}$ et $\hat{\theta}_3 = \dfrac{n}{n+1}\overline{X}^2$ sont tous trois des estimateurs sans biais de θ. [*Rappel* : Var(X_i) = β^2]

c) Déterminer la variance de $\hat{\theta}_1$. [*Suggestion* : utiliser le fait que pour une variable X de loi exponentielle, $\mathrm{E}(X^k) = k!\beta^k$].

d) Déterminer la variance de $\hat{\theta}_3$.

e) Montrer que $\hat{\theta}_3$ est un meilleur estimateur que $\hat{\theta}_1$.

f) Montrer que $\dfrac{n-1}{n\overline{X}}$ est sans biais pour $1/\theta$.

4.12 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi gamma de paramètres α et β.

a) Montrer que la méthode des moments donne les estimateurs suivants : $\hat{\alpha} = \dfrac{n\bar{X}^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$ et $\hat{\beta} = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n\bar{X}}$.

b) On choisit 12 jours au hasard et on note la consommation d'énergie (en millions de kilowattheures) ces jours-là :

6,4; 4,5; 10,8; 7,2; 6,8; 4,9; 3,5; 16,3; 4,8; 7,0; 8,8; 5,5.

Supposons que la consommation suit une loi gamma de paramètres α et β. Estimez ces paramètres par la méthode des moments.

4.13 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de loi uniforme sur $[0; \beta]$. Déterminer par la méthode des moments l'estimateur de β.

4.14 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de loi uniforme sur $[\alpha; 1]$. Déterminer par la méthode des moments l'estimateur de α.

4.15 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de densité

$$f(x) = \begin{cases} \dfrac{2}{\beta}(1 - x/\beta) & 0 < x < \beta \\ 0 & \text{ailleurs} \end{cases}.$$

Déterminer par la méthode des moments l'estimateur de β.

4.16 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de densité

$$f(x) = \begin{cases} \dfrac{2}{\beta}(1 - x/\beta) & 0 < x < \beta \\ 0 & \text{ailleurs} \end{cases}.$$

Déterminer l'estimateur de β par la méthode du maximum de vraisemblance [Il suffira d'indiquer quelle équation doit être résolue. La solution n'est pas facile à trouver].

4.17 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de loi gamma de paramètres $\alpha = 2$ (connu) et β inconnu. Déterminer par la méthode des moments l'estimateur de β.

4.18 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de loi géométrique de paramètre p. Déterminer par la méthode des moments l'estimateur de p.

4.19 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de densité

$$f(x) = \begin{cases} \dfrac{1}{\beta} e^{-(x-\alpha)/\beta} & x > \alpha \\ 0 & \text{ailleurs} \end{cases}.$$

Déterminer par la méthode des moments kes estimateurs de α et de β.

4.20 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de loi géométrique de paramètre p. Déterminer par la méthode du maximum de vraisemblance l'estimateur de p.

4.21 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de loi uniforme sur $[0; \beta]$. Déterminer par la méthode du maximum de vraisemblance l'estimateur de β. [rép. $\hat{\beta} = \max\{X_1; \ldots; X_n\}$]

4.22 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi uniforme sur [α; 1]. Déterminer par la méthode du maximum de vraisemblance l'estimateur de α. . [rép. $\hat{\alpha} = \min\{X_1; \ldots; X_n\}$]

4.23 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi uniforme sur [α ; β]. Déterminer par la méthode du maximum de vraisemblance les estimateurs de α et de β . [rép. $\hat{\alpha} = \min\{X_1; \ldots; X_n\}$; $\hat{\beta} = \max\{X_1; \ldots; X_n\}$].

4.24 Soit X un échantillon de taille 1 d'une population de fonction de densité $f(x) = 2x/\theta^2$ pour $0 < x < \theta$ et zéro ailleurs.

a) Déterminer un estimateur sans biais de la forme $\hat{\theta} = cX$. Déterminer la valeur de $\hat{\theta}$ lorsque $X = 2$.

b) On tire un échantillon de taille 4 de cette population. Les données de l'échantillon sont 3 ; 4 ; 6 ; 12.

Estimer θ par la méthode du maximum de vraisemblance

4.25 On tire un échantillon de taille 4 d'une population de densité $f(x) = (1/\beta^2)xe^{-x/\beta}$, pour $x > 0$ et zéro ailleurs. Les données d'un échantillon de taille 4 sont $X_1 = 3$; $X_2 = 4$; $X_3 = 6$; $X_4 = 12$. Estimer β par la méthode du maximum de vraisemblance

4.26 Le nombre de fautes de frappe dans une page de livre est une variable de loi de Poisson de paramètre γ. Deux réviseurs vérifient l'un 10 pages, l'autre 20 pages. Soit X le nombre total de fautes détectées par le premier (sur 10 pages); et Y le nombre total de fautes détectées par le deuxième (20 pages).

a) Déterminer la condition que doivent satisfaire a et b pour que l'estimateur $\hat{\gamma} = aX + bY$ soit sans biais pour γ [c'est-à-dire, exprimer a en fonction de b].

b) De tous les estimateurs sans biais de la forme $\hat{\gamma} = aX + bY$, quel est le meilleur? Évaluez $\hat{\gamma}$ si $X = 3$ et $Y = 5$.

4.27 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi exponentielle à deux paramètres, de densité $f(x) = \begin{cases} \dfrac{1}{\beta} e^{-(x-\alpha)/\beta} & x > \alpha \\ 0 & \text{ailleurs} \end{cases}$.

a) Déterminer par la méthode du maximum de vraisemblance les estimateurs de α et de β.

b) Supposons que les durées des ampoules d'un certain type sont de loi exponentielle à deux paramètres. Estimez α et β à partir des données suivantes, qui représentent les durées d'un échantillons d'ampoules :

415 ; 433 ; 489 ; 531 ; 466 ; 410 ; 479 ; 403 ; 562 ; 422 ; 475 et 439 heures.

4.28 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi uniforme sur [0; β]. Considérez les deux estimateurs suivants : $\hat{\beta}_1 = 2\overline{X}$ et $\hat{\beta}_2 = \max\{X_1; \ldots; X_n\}$.

a) Montrer que $\hat{\beta}_1$ est sans biais.

b) Montrer que $\hat{\beta}_2$ est biaisé, déterminez-en le biais, et déterminer un facteur c tel que $\hat{\beta}_3 = c\hat{\beta}_2$ soit sans biais.

c) Déterminer la variance de $\hat{\beta}_1$ et la variance de $\hat{\beta}_3$ et dites lequel des deux estimateurs est meilleur.

4.29 Un inspecteur doit estimer la proportion p de ménages munis d'un détecteur de fumée. Il décide donc d'inspecter des ménages et de continuer jusqu'au moment où il obtient un **10^{e}** ménage avec détecteur de fumée. Sachant que ce 10^{e} ménage est le **50^{e}** inspecté, déterminer l'estimateur de p par la méthode des moments et estimer p.

4.30 La teneur X en mercure des poissons d'un étang très pollué est une variable de densité

$$f(x) = \begin{cases} (\theta+2)x^{\theta+1} & 0 < x < 1 \\ 0 & \text{ailleurs} \end{cases}, \qquad \theta > 0\,.$$

Soit X_1, X_2, … , X_n les teneurs en mercure de n poissons tirés dans l'étang. Déterminer l'estimateur de θ par la méthode du maximum de vraisemblance.

4.31 Supposons que le nombre de taches dans une plaque d'émail est une variable de loi de Poisson de paramètre λ. Dans un échantillon de n plaques vous comptez le nombre de taches dans chacune des plaques.

a) Supposons que vous ne retenez que l'information suivante : Le nombre X de plaques n'ayant *aucune* tache. Déterminer l'estimateur par les moments de λ.

b) Supposons maintenant qu'en fait vous avez retenu toute l'information disponible, soit les valeurs X_1, X_2, … , X_n, X_i étant le nombre taches dans la plaque i. Déterminer l'estimateur par les moments de λ.

c) Estimer λ à l'aide de l'estimateur en b) sachant que 25 plaques ont été échantillonnées avec les résultats suivants : 18 des plaques n'avaient aucune tache, 5 en avaient une, et 2 en avaient 4.

4.32 Cinq pièces identiques font partie d'un appareil électronique. La durée de vie (en nombre d'heures) de chacune est une variable de loi exponentielle de moyenne β, et les 5 durées de vie sont indépendantes. L'appareil tombe en panne dès que l'une des 5 pièces cesse de fonctionner. Vous observez que l'appareil cesse de fonctionner après exactement 8 heures. Utiliser cette seule information pour déterminer l'estimateur du maximum de vraisemblance de β. Évaluer l'estimateur numériquement.

4.33 Afin d'estimer le nombre N de poissons dans une mare, vous en pêchez 5, vous les marquez, puis les remettez dans la mare. [Vous savez donc que la proportion de poissons marqués dans la marre est $p = 5/N$]. Ensuite vous repêchez 6 poissons, *avec* remise. Vous trouvez que parmi ces 6, un seul était marqué. Déterminer l'estimateur du maximum de vraisemblance de N [*Suggestion* : commencez par estimer p].

4.34 Un appareil électrique fonctionne tant que les n piles qui y sont installées fonctionnent. Supposons que la durée de fonctionnement (en nombre d'heures) d'une pile est de loi exponentielle de moyenne β. Votre appareil a cessé de fonctionner exactement 12 heures après que vous avez installé n piles neuves (donc une des piles a cessé de fonctionner). À partir de cette seule information, estimez β par la méthode du maximum de vraisemblance. L'estimateur est-il sans biais?

4.35 Un chercheur en médecine a besoin d'un sujet possédant un gène particulier, très rare. Afin d'en trouver un, il tire au hasard des individus dans une certaine (très grande) population jusqu'au moment où il tombe sur une personne possédant ce gène. Il lui a fallu tirer **200** personnes dans la population pour obtenir ce sujet. Déterminer l'estimateur du maximum de vraisemblance $\hat{p}$ de p, la proportion d'individus dans la population qui possèdent le gène en question. L'estimateur est-il sans biais?

4.36 Considérons une population de loi de Pareto, $f(x) = \alpha x^{-\alpha-1}$, $x > 1$.

a) Déterminer l'estimateur du maximum de vraisemblance de α étant donné un échantillon de taille n d'une population de densité f.

b) Supposons que les salaires (en dizaines de milliers de dollars) des professeurs du secondaire sont de loi de Pareto de paramètre α. Un échantillon de 15 professeurs donne les salaires suivants :

31 332 \$; 23 038 \$; 10 395 \$; 29 432 \$; 59 329 \$; 45 139 \$; 13 581 \$; 15 016 \$;
10 582 \$; 17 295 \$; 22 154 \$; 23 280 \$; 14 013 \$; 20 799 \$; 17 000 \$.

Supposez que les salaires, exprimés en dizaines de milliers de dollars, sont de loi de Pareto de paramètre α. Estimez α. L'estimateur est-il sans biais?

4.37 Voici la distribution observée du nombre d'accidents faits en une année par un grand groupe des clients d'une compagnie d'assurance :

Nombre d'accidents	Nombre d'assurés
0	81714
1	11306
2	1618
3	250
4	40
5	7

a) Déterminer l'estimation du maximum de vraisemblance de p sous la supposition que le nombre d'accidents est de loi $\mathcal{B}(5; p)$

b) Déterminer l'estimation du maximum de vraisemblance de λ sous la supposition que le nombre d'accidents est de loi $\mathcal{P}(\lambda)$.

c) Déterminer l'estimation de la probabilité qu'un client n'ait aucun accident dans chacun des modèles ci-dessus. [Vous pouvez tenir pour acquis que si $\hat{p}$ est l'estimateur du maximum de vraisemblance de p, alors $(1-\hat{p})^n$ est l'estimateur du maximum de vraisemblance de $(1-p)^n$; de même, si $\hat{\lambda}$ est l'estimateur du maximum de vraisemblance de λ, alors $e^{-\hat{\lambda}}$ est l'estimateur du maximum de vraisemblance de $e^{-\lambda}$]

d) Déterminer l'estimation de la probabilité qu'un client n'ait aucun accident sans faire de supposition sur la distribution du nombre d'accidents.

4.38 On tire deux échantillons aléatoires simples, de taille n_1 et n_2, d'une même population normale de moyenne μ et de variance σ^2. Les seules informations retenues sont les moyennes $\overline{X}_1$ et $\overline{X}_2$. Montrez que les estimateurs $\hat{\mu}$ et $\hat{\sigma}^2$ du maximum de vraisemblance μ et σ^2, respectivement, sont $\hat{\mu} = \dfrac{n_1\overline{X}_1 + n_2\overline{X}_2}{n}$, où $n = n_1+n_2$; et $\hat{\sigma}^2 = \dfrac{n_1(\overline{X}_1 - \hat{\mu})^2 + n_2(\overline{X}_2 - \hat{\mu})^2}{2}$.

4.39 Soit un échantillon aléatoire simple de taille n d'une population $\mathcal{N}(0; \sigma^2)$.

a) Montrer que l'estimateur du maximum de vraisemblance de σ^2 est $\dfrac{\sum_{i=1}^{n} X_i^2}{n}$. Montrer que cet estimateur est sans biais et déterminer sa variance.

b) On sait que l'estimateur $S^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$ aussi est sans biais. Mais montrer qu'il est moins bon que celui présenté en a).

[Il sera utile de se rappeler que l'espérance d'une variable aléatoire de loi χ^2_ν est ν et que sa variance est 2ν].

4.40 Supposons que la durée de fonctionnement (en nombre d'heures) d'une *pile* est de loi exponentielle de moyenne β. Vous installez 4 piles dans un appareil électrique et deux autres piles dans un autre appareil. Les deux appareils ne fonctionnent que lorsque toutes les piles (4 dans le premier, 2 dans le deuxième) fonctionnent. Le premier appareil a cessé de fonctionner exactement 12 heures après que vous ayez installé 4 piles neuves (donc une des piles a cessé de fonctionner); le deuxième a cessé de fonctionner exactement 7 heures après l'installation des 2 piles. Estimer β par la méthode du maximum de vraisemblance.

4.41 Le nombre d'arrivées à un poste de péage dans l'intervalle de temps $[0; t]$ est une variable aléatoire de loi de Poisson de paramètre $\lambda = \alpha t$. Considérer les deux procédures suivantes pour estimer α.

Procédure 1 : Vous comptez le nombre d'arrivées X dans l'intervalle $[0; t]$ et vous estimez α par X/t.

Procédure 2 : Vous observez n arrivées successives et vous estimez α par $n/\Sigma Y_i$, Y_i étant le temps d'attente entre la $(i\text{-}1)^e$ et la i^e arrivées, $i = 2,\ldots, n$, et Y_1 étant le temps d'attente avant la première arrivée.

Montrez que les deux estimateurs sont sans biais, et que leurs variances sont, respectivement, α/t et $(3n\text{-}2)\alpha^2/[(n\text{-}1)(n\text{-}2)]$. Déterminer une condition, en termes de n et t sous laquelle le 2^e estimateur est préférable au premier.

4.42 *Estimation du paramètre d'une loi hypergéométrique* L'acheteur potentiel d'un immeuble de $N = 100$ appartements tire un échantillon de $n = 30$ appartements.

Selon le vendeur, 35 appartements ont des cuisines rénovées. Mais dans l'échantillon, 5 seulement des 30 appartements le sont.

a) Soit N_1 le nombre d'appartements avec cuisine rénovée dans l'immeuble. Si l'affirmation du vendeur (que $N_1 = 35$) est vraie, quelle est l'espérance du nombre de cuisines rénovées dans l'échantillon ?

b) L'affirmation du vendeur (que $N_1 = 35$) est-elle crédible (à la lumière du résultat de l'échantillon) ?

c) Quelles valeurs de N_1 sont crédibles ?

d) Mais est-ce que toutes les valeurs de N_1 inférieures à 30 sont crédibles ? Certaines seraient-elles trop petites ?

4.43 Vous tirez un échantillon de 20 pièces fabriquées d'un processus dont le taux de défectuosités (le pourcentage de pièces défectueuses) est censé être de 2 %. Vous observez 4 pièces défectueuses.

a) Sous l'hypothèse que le taux de défectuosités est bien de 2 %, quelle est l'espérance du nombre de pièces défectueuses dans votre échantillon ? Si X est le nombre de pièces

b) Sous l'hypothèse que le taux de défectuosités est bien de 2 %, quelle est la probabilité d'avoir 4 pièces ou plus dans votre échantillon ?

c) À la lumière de la probabilité calculée en b), que concluez-vous à propos de l'hypothèse que $p = 2$ % ?

d) Est-ce que l'hypothèse $p = 0{,}05$ vous semble plus vraisemblable ?

e) Lesquelles des valeurs suivantes de p sont vraisemblables, compte tenu de l'observation $X = 4$: 0,05; 0,06; 0,07; 0,08; 0,09; 0,10 ?

f) Supposons qu'en fait le lot est juste acceptable (p = 0,10). Quelle est la probabilité de le rejeter ?

g) On change de critère car on trouve la probabilité calculée en i) trop élevée : on rejettera le lot seulement si $X \geq 5$. Quelle est la probabilité de rejeter un lot dans lequel 10 % des pièces sont défectueuses ?

h) La probabilité calculée en ii) pourrait être considérée suffisamment faible ou pas, ce sont les personnes impliquées qui en décident. On aura normalement fixé un seuil α (la probabilité maximale de rejet d'un lot acceptable.) Supposons qu'on pose α = 0,01. À partir de quelle valeur de X devrait-on rejeter le lot ?

4.44 Supposons que le nombre X de rongeurs qui passent dans votre cours en une nuit est une variable de loi de Poisson de paramètre λ.

a) Si vous observez $X = 2$, la valeur $\lambda = 5$ est-elle vraisemblable ?

b) Parmi les valeurs suivantes, lesquelles considérez-vous « vraisemblables » ?

6,0; 6,1; 6,2; 6,3; 6,4; 6,5; 6,6; 6,7; 6,8; 6,9; 7,0

c) Supposons qu'en deux nuits, on observe en moyenne deux rongeurs par nuit. La valeur $\lambda = 5$ est-elle raisonnable ?

d) On compte le nombre de rongeurs en dix nuits. Soit $X_1, X_2, \ldots, X_n$ les n valeurs observées. Supposons qu'on estime λ par $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Supposons que $n = 10$ et $\lambda = 3$. Quelle est la probabilité d'une surestimation de plus de 10 % ?

e) Pour n assez grand, on peut considérer Y comme étant de loi approximativement normale. Montrer que la probabilité d'une surestimation de plus de 10 % est à peu près $1-\Phi[(0,1)\sqrt{n\lambda}]$

f) Supposons qu'en fait les valeurs exactes de $X_1; \ldots; X_n$ ne peuvent pas être observées : les traces laissées par le passage d'un rongeur permettent de conclure *qu'au moins* un rongeur est passé mais ne permettent pas de savoir *combien* sont passés. La seule information disponible est donc W = nombre de nuits (parmi n) durant lesquelles *aucun* rongeur n'est passé. Peut-on quand même estimer λ ? Sachant que $\hat{p} = W/n$ est une estimation raisonnable de la probabilité que $X_i = 0$, laquelle est égale à $e^{-\lambda}$, on estime $e^{-\lambda}$ par $\hat{p}$ et donc on peut estimer λ par $\hat{\lambda} = -\ln \hat{p}$. Si n est assez grand, W est à peu près de loi normale. Montrer que la probabilité d'une surestimation de plus de 10 % est alors à peu près égale à $P(\hat{\lambda} > 1{,}1\lambda) = \Phi\left[\frac{\sqrt{ne^{-\lambda}}(e^{-\lambda/10}-1)}{\sqrt{1-e^{-\lambda}}}\right]$.

g) Posez $n = 100$ et comparer $P[\overline{X} > (1{,}1)\lambda]$ à $P(\hat{\lambda} > (1{,}1)\lambda)$ pour quelques valeurs de λ.

4.45 Le nombre de pièces défectueuses dans une boîte de k pièces est une variable de loi binomiale de paramètre p. Un fabriquant vend n de ces boîtes avec la promesse de remplacer toute boîte dans laquelle au moins une pièce est défectueuse. Supposons que sur n boîtes vendues, Y n'ont aucune pièce défectueuse alors que les n-Y en ont au moins une (on ne connaît rien de leur nombre, sinon qu'il est non nul). Peut-on quand même estimer p ? On sait que $\hat{\varphi} = Y/n$ est une estimation raisonnable de la probabilité qu'une boîte ne contienne aucune pièce défectueuse, donc une estimation de $\varphi = (1-p)^k$. Une estimation raisonnable de p est alors $\hat{p} = 1- \hat{\varphi}^{1/k}$.

a) Montrer que la probabilité de surestimer p par plus de 10 % est à peu près $\Phi\left[\sqrt{n}\,\frac{(1-1{,}1p)^k-(1-p)^k}{\sqrt{(1-p)^k[1-(1-p)^k]}}\right]$.

b) Évaluer cette probabilité pour $k = 4$, $n = 100$ et quelques valeurs de p.

c) Supposons qu'en fait on connaît le nombre de pièces défectueuses dans chaque boîte prélevée. On estime alors p par la proportion de pièces défectueuses parmi les nk pièces de l'échantillon. Déterminer approximativement la probabilité de surestimer p par plus de 10 % pour $k = 4$, $n = 100$ et les mêmes valeurs de p qu'en b).

4.46 Supposons que la durée de vie (en milliers d'heures) d'une pièce électronique est une variable de fonction de densité $f(x)=\frac{1}{\beta}e^{-x/\beta}$ ($x>0$). La fonction de répartition est donc $F(x) = 1-e^{-x/\beta}$, $x \geq 0$.

a) On tire une pièce au hasard. Elle dure 5000 heures ($X = 5$). La valeur $\beta = 3$ est-elle vraisemblable ?

b) Quelle serait la plus grande valeur acceptable de β ?

c) Un appareil est muni de 4 de ces pièces et tombe en panne dès que l'une d'elle cesse de fonctionner. Supposons que l'appareil tombe en panne après 5000 heures. La valeur $\beta = 3$ est-elle vraisemblable ?

4.47 Supposons que le nombre de fautes typographiques dans une page est une variable X de loi de Poisson de paramètre λ. Un vérificateur peut cependant ne pas découvrir toutes les fautes dans une page.

a) Soit Y le nombre de fautes *découvertes* par un vérificateur dans une page donnée. Supposons que $Y \sim \mathcal{B}(n; p)$, où n est le nombre réel de fautes dans la page. Montrer que Y est de loi de Poisson de paramètre $p\lambda$.

b) On demande au vérificateur de vérifier une page tirée au hasard. Soit Y le nombre de fautes qu'il a découvertes. À une autre occasion, on tire une autre page, mais cette fois-ci on s'assure (par une vérification très minutieuse) que toutes les fautes soient découvertes. Soit X le nombre de fautes découvertes. Montrer que $\hat{\lambda} = X$ et $\hat{p} = Y/X$ sont les estimateurs par les moments ainsi que les estimateurs à vraisemblance maximale.

c) Supposons maintenant qu'aux deux occasions on note le nombre réel de fautes X_1 et X_2 ainsi de le nombre de fautes découvertes par le vérificateur, Y_1 et Y_2. Montrez que les estimateurs du maximum de vraisemblance sont $\hat{\lambda}=\frac{X_1+X_2}{2}$ et $\hat{p}=\frac{Y_1+Y_2}{X_1+X_2}$.

d) Remarquez que l'estimateur en c) n'est pas défini si $X_1 + X_2 = 0$ et que $\mathrm{E}(\hat{p}) = \infty$, à moins qu'on décide de donner une valeur à $\hat{p}$ lorsque $X_1 + X_2 = 0$. Supposons donc qu'on définisse $\hat{p}$ comme ceci : $\hat{p}=\begin{cases}\frac{Y_1+Y_2}{X_1+X_2} & \text{si } X_1+X_2>0\\ 0 & \text{sinon}\end{cases}$. Montrer que le biais de $\hat{p}$ est $-pe^{-2\lambda}$.

4.48 Le nombre d'arrivées au quai d'un métro suit une loi de Poisson au taux de 2 arrivées/minute. Vous arrivez au quai à un moment aléatoire et trouvez 10 personnes sur le quai. Estimer le temps écoulé depuis le dernier départ.

4.49 Dans une population symétrique, la médiane et la moyenne coïncident, ce qui suggère que la médiane $\tilde{Y}$ d'un échantillon pourrait servir d'estimateur de la moyenne μ. Considérons un échantillon de taille $n = 2m+1$ d'une population de fonction de répartition F et de fonction de densité f.

a) Montrer que la fonction de densité de $\tilde{Y}$ est $g(\tilde{y}) = \frac{(2m+1)!}{m!m!} f(\tilde{y})[F(\tilde{y})]^m[1-F(\tilde{y})]^m$.

b) Pour simplifier et sans perte de généralité, supposons que μ = 0, montrer que $E(\tilde{Y}) = 0$.

[Posez $h(\tilde{y}) = \tilde{y}g(\tilde{y})$ et montrez que $h(\tilde{y}) = h(-\tilde{y})$ et concluez que $E(\tilde{y}) = \int \tilde{y}g(\tilde{y})d\tilde{y} = 0$.

4.50 Soit $X_1, \ldots, X_n$ un échantillon d'une population de loi de Poisson de moyenne λ. Alors $\bar{X}$ et S^2 sont tous deux sans biais pour λ. Lequel est meilleur?

a) Soit $T = \sum_{i=1}^n X_i$. Montrer que la loi conditionnelle de $[X_1; \ldots; X_n]$ étant donnée $T = t$ est une loi multinomiale de paramètres t et probabilités toutes égales à $\frac{1}{n}$.

b) Montrer que $E(\sum_{i=1}^n X_i^2 \mid \sum_{i=1}^n X_i = t) = \frac{(n-1)t+t^2}{n}$ et donc que $E(S^2 \mid T = t) = \bar{X}$.

c) Montrer que $\mathrm{Var}(\bar{X}) \le \mathrm{Var}(S^2)$.

[*Utilisez la formule* $\mathrm{Var}(\bar{X}) = \mathrm{Var}[E(\bar{X} \mid T = t)] + E[\mathrm{Var}(\bar{X} \mid T = t)]$]

d) Montrer que $\mathrm{Cov}(\bar{X}; S^2) = 0$. [Notez que $E(\bar{X}S^2) = E[E(\bar{X}S^2 \mid T = t)]$.]

4.51 [Suite du numéro 4.11]

Considérer l'estimateur $\hat{\theta}_3 = \frac{n\bar{X}^2}{n+1} = \frac{T^2}{n(n+1)}$, où $T = n\bar{X}$. Soit X l'une quelconque des variables $X_1, \ldots, X_n$.

a) Montrer que la densité conditionnelle de X étant donné $T = t$ est $f(x \mid t) = (n-1)\left(1-\frac{x}{t}\right)^{n-1}\frac{1}{t}$, $0 < x < t$.

b) Montrer que $\mathrm{E}(X \mid t) = \frac{t}{n}$ et que $\mathrm{E}(X^2 \mid t) = \frac{2t^2}{n(n+1)}$.

c) Montrer que $E(\hat{\theta}_2 \mid \hat{\theta}_3) = \frac{t^2}{n(n+1)} = \hat{\theta}_3$, où $\hat{\theta}_2 = S^2$. [Puisque $\hat{\theta}_3$ est une fonction bijective de T, il revient au même de montrer que $E(\hat{\theta}_2 \mid T = t) = \frac{t^2}{n(n+1)}$, ce qui est moins encombrant.]

d) Montrer que $\mathrm{Var}(\hat{\theta}_2) \ge \mathrm{Var}(\hat{\theta}_3)$.

[*Utilisez la formule* $\mathrm{Var}(\hat{\theta}_2) = \mathrm{Var}[\mathrm{E}(\hat{\theta}_2 \mid \hat{\theta}_3)] + \mathrm{E}[\mathrm{Var}(\hat{\theta}_2 \mid \hat{\theta}_3)]$].

e) Montrer que $\mathrm{Cov}(\hat{\theta}_2; \hat{\theta}_3) = 0$.

f) Montrer de même que $E(\hat{\theta}_1 | \hat{\theta}_3) = \hat{\theta}_3$, et utiliser ce fait pour montrer que $\text{Var}(\hat{\theta}_1) \geq \text{Var}(\hat{\theta}_3)$ et que $\text{Cov}(\hat{\theta}_1; \hat{\theta}_3) = 0$.