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 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	$T_2(X_1; \ldots; X_n) = \sum_{i=1}^{n}(X_i - \bar{X})^2 / n = \hat{\sigma}^2$	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 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, …, X_n, et la « variance » échantillonnale $S^2 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ 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

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

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 *Estimateur sans biais pour* μ

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 $E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\mu = \mu.$

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 Estimateur *sans biais de* σ^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

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \tag{4.3.1}$$

est un estimateur sans bais de σ^2.

Démonstration : Il suffit de montrer que $E[(n-1)S^2] = (n-1)\sigma^2$. Rappelons que pour toute variable aléatoire W, $E(W^2) = Var[W] + [E(W)]^2$. Maintenant, puisque $(n-1)S^2 = \sum_{i=1}^{n} x_i^2 - n\bar{X}^2$, nous avons $E[(n-1)S^2] = E\left[\sum_{i=1}^{n} x_i^2 - n\bar{X}^2\right] = E\left(\sum_{i=1}^{n} x_i^2\right) - nE(\bar{X}^2) = \sum_{i=1}^{n} E(x_i^2) - nE(\bar{X}^2) = \sum_{i=1}^{n}(\sigma^2 + \mu^2) - nE(\bar{X}^2) = n(\sigma^2 + \mu^2) - n[Var(\bar{X}) + (E(\bar{X}))^2] = n(\sigma^2 + \mu^2) - n[\sigma^2/n + \mu^2] = (n-1)\sigma^2$. Donc $E[(n-1)S^2] = (n-1)\sigma^2 \Rightarrow E(S^2) = \sigma^2$. 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 *Estimateur d'une proportion*

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

$$\hat{p} = \frac{X}{n} \text{ est un estimateur sans bais de } p. \tag{4.3.2}$$

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 *Variance de* $\overline{X}$

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

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} \tag{4.4.1}$$

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

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

$$\mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n} \tag{4.4.2}$$

Théorème 4.4.3 *Variance de* S^2

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

$$\mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1} \tag{4.4.3}$$

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:

$$\mathrm{Var}(\hat{\theta}) = \mathrm{E}\{[\hat{\theta} - \mathrm{E}(\hat{\theta})]^2\} = \mathrm{E}[(\hat{\theta} - \theta)^2] \tag{4.4.4}$$

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* $\mathrm{EQM}(\hat{\theta})$ 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*

$$\mathrm{EQM}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + [\mathrm{B}(\hat{\theta})]^2 \tag{4.4.5}$$

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 :

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n} \;;\; \mathrm{Var}(\hat{p}) = \frac{p(1-p)}{n} \;;\; \mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}\ . \qquad (4.4.6)$$

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

$$\mathrm{cv}(\hat{\theta}) = \frac{\sqrt{\mathrm{Var}(\hat{\theta})}}{\theta}\ . \qquad (4.4.7)$$

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 $1/\mu$, 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 $\hat{\theta}$ 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 $\hat{\theta}$ est la solution des équations

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

Exemple 4.5.2 *Estimateurs par les moments de* μ *et de* σ^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 :

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

On remarque que l'estimateur de σ^2 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 $\varphi'(p) = 0$. On a $\varphi'(p) = \binom{20}{7}[7p^6(1-p)^{13} - 13p^7(1-p)^{12}]$ 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 $X_1; X_2; \ldots; X_n$ 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 le 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 *Estimateur du maximum de vraisemblance*

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

Exemple 4.6.2 *L'EMV du paramètre d'une population de loi exponentielle*

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'EMV de β.

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'EMV de β est donc $\hat{\beta} = \bar{X}$. ■

Paramètres d'une loi binomiale

On montre que la proportion échantillonnale $\hat{p}$ est l'EMV de p.

Théorème 4.6.1 *Estimation d'une proportion*

Soit X une observation de loi $\mathcal{B}(n\,;p)$. L'EMV de p est $\hat{p} = \dfrac{X}{n}$.

Démonstration

Il s'agit de maximiser la fonction de vraisemblance

$$L(p) = \binom{n}{x} p^x(1-p)^{n-x} \text{ par rapport à } p.$$

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

$$\ell'(p) = 0 \Rightarrow \frac{x-np}{p(1-p)} = 0 \Rightarrow X - np = 0 \Rightarrow p = \frac{x}{n}.$$

Ceci conclut la démonstration. ■

Paramètres d'une population normale

On montre ici que les EMV de μ et de σ sont précisément ceux obtenus par la méthode des moments.

Théorème 4.6.2 *Paramètres d'une population normale*

Soit $X_1, X_2, \ldots, X_n$ 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 $-\frac{n}{2\sigma^2} + \frac{\sum(x_i-\bar{x})^2}{2\sigma^4}$. 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:

$\leq$ 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 :

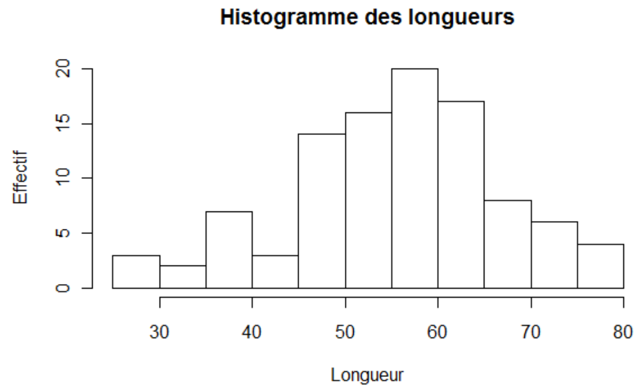
29,0	35,1	38,1	45,5	46,7	48,8	50,5	52,3	53,6	55,1	56,6	58,4	59,2	60,2	61,7	63,2	64,8	67,3	70,4	74,4
29,0	35,6	38,4	45,7	47,8	49,0	50,5	52,6	53,8	55,4	56,6	58,4	59,2	60,2	61,7	63,5	65,0	68,1	71,6	78,7
30,0	36,1	41,7	45,7	48,3	49,5	50,8	53,1	54,6	55,6	57,1	58,9	59,7	60,5	61,7	63,8	65,3	68,1	72,9	79,0
31,0	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
$\leq$ 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.

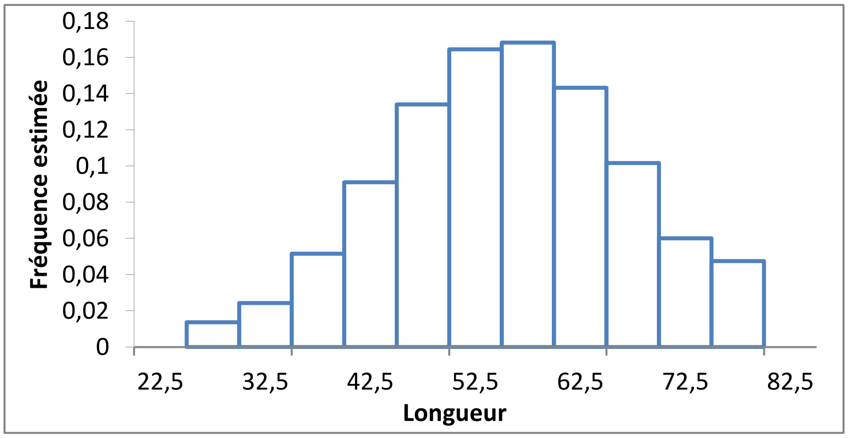
Estimation à l'aide d'un modèle normal

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 $E(\hat{\theta}) = \theta$ quelle que soit la valeur de θ.

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