

Notions d'échantillonnage et d'estimateur

Les notions présentées dans ce chapitre permettent de formaliser d'un point de vue statistique l'estimation du paramètre d'une loi de probabilité à partir d'un relevé de population. Les propriétés élémentaires de ce qu'on appelle un estimateur sont présentées. Nous insistons sur le comportement de la moyenne empirique, estimateur de l'espérance d'une loi de probabilité.

Sommaire

4.1	Échantillonnage	34
4.1.1	Échantillon aléatoire : définition	34
4.1.2	Notion de statistique	35
4.2	Moyenne empirique	36
4.2.1	Espérance et variance	37
4.2.2	Le théorème de la limite centrale	38
4.3	Estimateur	40
4.3.1	Modèle statistique	40
4.3.2	Propriétés élémentaires d'un estimateur	41
4.3.3	\bar{X}_n est un estimateur convergent de l'espérance d'une loi	42
4.3.4	Variance empirique et variance empirique corrigée	43

Exemple introductif

Un relevé de la taille en centimètres de n individus est effectué parmi l'ensemble de la population des hommes résidant sur le territoire français, notée \mathcal{P} . On suppose que la taille est une variable aléatoire X qui suit une loi de Gauss d'espérance μ et de variance σ^2 , c'est-à-dire que $X \sim \mathcal{N}(\mu, \sigma^2)$.

Comment déterminer, à partir des données, une valeur acceptable pour le paramètre d'espérance μ , c'est-à-dire la taille moyenne de la population \mathcal{P} ?¹ Quelles sont les propriétés statistiques de la valeur affectée à μ à partir des données (on dit « estimer ») ? Comment évolue cette valeur en fonction de la quantité de données prélevées ?

¹Cette question est intuitivement très simple : vous y répondez depuis le longtemps en calculant par exemple la moyenne de toutes vos notes de français, de math, puis la moyenne générale.

Les notions abordées dans ce chapitre permettent de répondre à ces questions en formalisant d'un point de vue statistique les différentes étapes de résolution de cette famille de problèmes. À cet effet, nous allons manipuler des objets au vocabulaire un peu particulier, à savoir :

1. la variable aléatoire X observée est appelée *caractère* associé à la *population* \mathcal{P} ,
2. l'étape de collecte des données est appelée *échantillonnage*, et l'ensemble ainsi construit un *échantillon* de la v.a. X ,
3. tout indicateur numérique calculé à partir des observations est une réalisation d'une variable aléatoire appelée *statistique*²,
4. le fait de proposer une loi pour le caractère X , sans connaître pour autant ses paramètres, constitue un *modèle statistique*³,
5. le fait d'utiliser une statistique pour déterminer le paramètre d'un modèle est appelé *estimation*. La statistique est alors appelée *estimateur* de ce paramètre.

Les trois premiers points seront développés dans la première section, tandis que les deux derniers seront abordés dans la seconde.

4.1 Échantillonnage

Notions

Échantillon aléatoire, Statistique

Dans cette section, nous présentons les notions d'échantillonnage et de statistique.

4.1.1 Échantillon aléatoire : définition

Supposons que l'on s'intéresse au *caractère* X d'une population \mathcal{P} . Pour fixer les idées, on gardera à l'esprit l'exemple cité plus haut, à savoir la taille des hommes sur le territoire français.

Considérons l'expérience aléatoire qui consiste à prélever au hasard un individu de la population \mathcal{P} et à observer la valeur du caractère de cet individu (bien sûr, on suppose que chaque individu a la même chance d'être choisi). Dans notre exemple, un individu correspond à un homme résidant sur le sol français, et le caractère étudié à sa taille.

Supposons maintenant que l'on répète n fois cette expérience, *dans les mêmes conditions et de manière indépendante* : cette procédure est appelée *échantillonnage aléatoire*. Nous obtenons n valeurs notées x_1, x_2, \dots, x_n , comme autant de valeurs observées de la taille du 1^{er} individu tiré au hasard, du 2^e individu, etc. jusqu'au n^e . Imaginons que l'on renouvelle cette procédure : l'ensemble de valeurs x_1, x_2, \dots, x_n recueilli ne serait évidemment pas la même ! Il y a très peu de chance pour que l'on sélectionne à nouveau les mêmes personnes, qui plus est dans le même ordre ! On peut donc considérer que ces valeurs sont des *réalisations* d'un ensemble X_1, X_2, \dots, X_n de n variables aléatoires décrivant la taille du 1^{er} individu choisi, puis du 2^e etc. jusqu'au

²Par exemple, la taille moyenne calculée sur les observations disponibles est une *réalisation* de la v.a. *moyenne empirique*, statistique parmi d'autres.

³Dans l'exemple de la taille d'un individu, on a choisi la loi de Gauss, mais d'autres choix sont possibles !

n^e . Pourvu que l'on respecte le protocole expérimental, c'est-à-dire en renouvelant les n tirages de l'expérience dans les mêmes conditions et de manière indépendante, les v.a. X_1, X_2, \dots, X_n sont *i.i.d.*, pour « indépendantes et identiquement distribuées » : elles ont toutes la même loi que la v.a. X , appelée *variable aléatoire parente*. Dans notre exemple, c'est la v.a. décrivant la taille d'un homme en France, dans toute la population \mathcal{P} .

L'ensemble de v.a. X_1, X_2, \dots, X_n est appelé *échantillon aléatoire*, ou plus simplement *échantillon*. Une suite d'observations x_1, x_2, \dots, x_n de ces n variables aléatoires est appelée *réalisation* de l'échantillon aléatoire. La figure 4.1 illustre le protocole décrit ci-dessus. Dans les faits cependant, le nombre d'échantillon de taille n est presque systématiquement réduit à 1 : le fait d'en considérer K permet de bien comprendre pourquoi l'échantillon X_1, \dots, X_n est aléatoire, la sélection des n individus dans la population n'étant jamais la même.

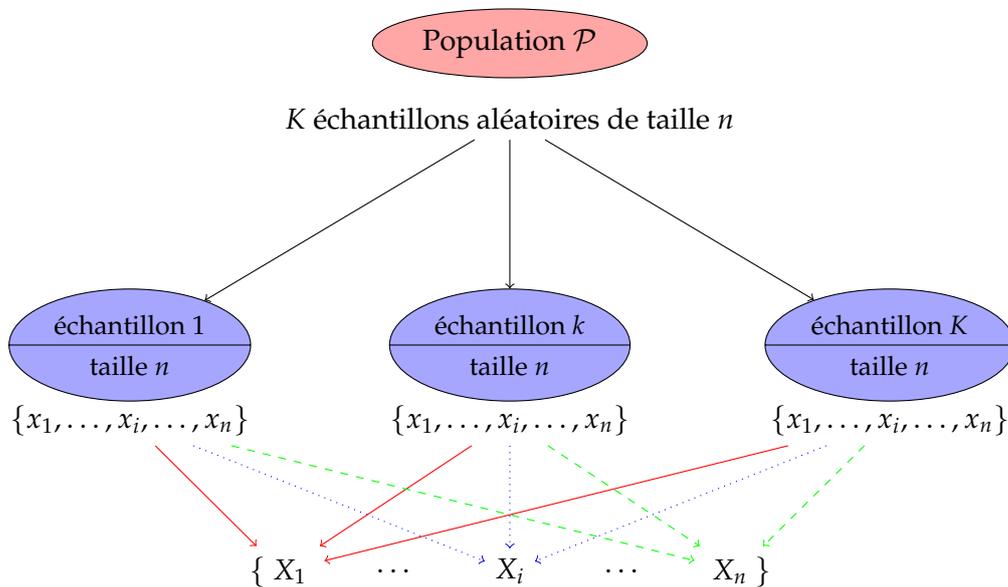


FIG. 4.1 – Principe de l'échantillonnage aléatoire

Remarque 4.1. Dans tout ce qui suit, on note en majuscule une variable aléatoire, qui peut prendre tout un intervalle de valeur. On note en minuscule une observation de cette variable, qui est une valeur réelle. Faites attention à cette distinction.

4.1.2 Notion de statistique

Une fois construit un jeu de données, c'est-à-dire dans notre jargon un échantillon aléatoire, divers calculs peuvent être fait sur ces observations. Il faut cependant faire attention : l'échantillon étant aléatoire, les calculs portent eux-mêmes sur des variables aléatoires. C'est là qu'intervient la notion de *statistique*.

Définition 4.1. Soit X_1, \dots, X_n un échantillon aléatoire. Alors, toute variable aléatoire définie comme une fonction g des variables X_1, X_2, \dots, X_n est appelée *statistique*.

Ainsi,

$$T_n = g(X_1, X_2, \dots, X_n),$$

est une statistique relative à l'échantillon X_1, \dots, X_n . Si l'on dispose d'une réalisation x_1, \dots, x_n de l'échantillon, on peut calculer une réalisation $t_n = g(x_1, \dots, x_n)$ de la statistique T_n .

Remarque 4.2. Le « n » en indice dans la notation T_n est là pour vous rappeler que toute statistique dépend de la taille de l'échantillon, mais il n'est pas obligatoire et on l'omettra si cela permet de rendre le discours plus clair.

On peut définir une infinité de statistiques, beaucoup d'entre elles n'étant que très peu intéressantes pour l'analyse de nos données ! Les statistiques suivantes apparaissent de manière systématique dans l'analyse de données, et il vous faudra les retenir :

1. la somme des v.a. de l'échantillon

$$S_n = \sum_{i=1}^n X_i,$$

2. la moyenne empirique, très liée à S_n ,

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

3. la variance empirique

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

4. la variance empirique corrigée

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

Exemple 4.1. Supposons que l'on ait relevé les tailles de 5 individus et que l'on obtienne 172cm, 182cm, 178cm, 175cm, 169cm. Alors les v.a. suivantes sont des statistiques :

- La somme des valeurs $s_n = 876$ est une réalisation de la statistique $S_n = \sum_{i=1}^5 X_i$ où X_i est une v.a. aléatoire décrivant la taille de l'individu i .
- La moyenne des valeurs $\bar{x} = 175.2$ est une réalisation de la moyenne empirique \bar{X}_n .
- La variance empirique $s^2 = 20.56$ et la variance empirique corrigée $s^{*2} = 25.7$ sont des réalisations des v.a S^2 et S^{*2} .
- Le produit des valeurs $p_n = 164795212400$ est une réalisation de la statistique définie par $P_n = \prod_{i=1}^5 X_i$, statistique peu intéressante en ce qui nous concerne.

4.2 Moyenne empirique

Notions

Moyenne empirique, Statistique S_n , Théorème de la Limite Centrale

Dans cette partie, nous étudions de manière précise les propriétés de la moyenne empirique, dont nous rappelons la définition :

Définition 4.2. Soit X_1, \dots, X_n un échantillon aléatoire. La v.a. \bar{X}_n définie par $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est appelée *moyenne empirique* de l'échantillon. Si x_1, \dots, x_n est une réalisation de cet échantillon, alors $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ est une réalisation de \bar{X}_n .

C'est une statistique fondamentale que vous utilisez (malgré vous?) depuis bien longtemps. Nous étudierons également le comportement de la statistique $S_n = \sum_{i=1}^n X_i$, qui est très comparable à celui de la moyenne empirique. Nous verrons donc tout d'abord espérance et variance de \bar{X}_n . Puis, nous verrons ce que l'on peut dire sur sa loi de probabilité, selon que l'on connaisse ou pas la loi de X , v.a. parente dont est issu l'échantillon.

4.2.1 Espérance et variance

On s'attend assez naturellement à ce que \bar{X}_n se comporte *en moyenne* comme la variable parente X (c'est de cette manière que \bar{X}_n a été définie). C'est ce qu'énonce la proposition suivante, dans laquelle on en profite pour également donner la variance de \bar{X}_n . Attention, ce résultat est vrai *quelque soit la loi de probabilité de X , pas uniquement dans le cas gaussien !*

Proposition 4.1. Soit X_1, \dots, X_n un échantillon de variable parente X , d'espérance μ et de variance σ^2 mais de loi inconnue. Alors, l'espérance et la variance de la moyenne empirique sont données par

$$\mathbb{E}(\bar{X}_n) = \mu, \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Démonstration. La preuve est simple mais *fondamentale*. Elle utilise les propriétés élémentaires de l'espérance et de la variance et se fonde sur la nature même d'un échantillon : les v.a. qui le composent sont i.i.d., et en particulier

$$\begin{aligned} \mathbb{E}(X_1) &= \mathbb{E}(X_2) = \dots = \mathbb{E}(X_n) = \mathbb{E}(X) = \mu, \\ \text{et } \text{Var}(X_1) &= \text{Var}(X_2) = \dots = \text{Var}(X_n) = \text{Var}(X) = \sigma^2. \end{aligned}$$

Passons au calcul de $\mathbb{E}(\bar{X}_n)$: en utilisant la linéarité de l'espérance et comme $\mathbb{E}(X_i) = \mu$ pour tout $i = 1, \dots, n$, nous avons

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

Pour la variance de \bar{X}_n , on utilise le fait que X_1, \dots, X_n sont indépendantes : ainsi la variance de la somme est la somme des variances (propriété 1.5). On rappelle également que $\text{Var}(aX) = a^2 \text{Var}(X)$. Ainsi,

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n^2} \sum_{i=1}^n 1 = \frac{\sigma^2}{n^2} \times n = \frac{\sigma^2}{n}.$$

□

Remarque 4.3. La moyenne empirique de n v.a. X_i a donc la même espérance que la v.a. parente X , mais une variance beaucoup plus faible : ceci explique par exemple que l'on considère que la moyenne de plusieurs notes est plus fiable pour évaluer un étudiant qu'une seule note de cet étudiant !

Nous obtenons de manière similaire l'espérance et la variance de la statistique $S_n = \sum_{i=1}^n X_i = n\bar{X}_n$.

Proposition 4.2. Soit X_1, \dots, X_n un échantillon aléatoire. La v.a. \bar{X}_n définie par

$$\mathbb{E}(S_n) = n\mu, \quad \text{Var}(S_n) = n\sigma^2.$$

Démonstration. La preuve est laissée en exercice. □

4.2.2 Le théorème de la limite centrale

Nous avons obtenu l'espérance et la variance de la moyenne empirique, quelque soit la loi de probabilité de la variable parente. Ne pourrait-on pas faire encore mieux, à savoir déterminer sa loi de probabilité ? Sans connaître le loi de X , cela paraît voué à l'échec... Et pourtant ! Nous allons voir que l'on peut en donner une approximation, dont la qualité dépend de la taille de l'échantillon considéré.

Commençons par un cas simple. Supposons que l'on ait à faire à un échantillon gaussien, c'est-à-dire que la loi parente X , et donc chacune des v.a. X_1, X_2, \dots, X_n ont pour loi $\mathcal{N}(\mu, \sigma^2)$. Nous avons vu au chapitre trois que toute combinaison linéaire de v.a. gaussiennes était gaussienne elle-même : c'est précisément le cas de la moyenne empirique, en tant que somme des X_1, X_2, \dots, X_n , pondérée par $1/n$. Comme, de plus on sait calculer son espérance et sa variance, nous avons

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \text{ si } X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2).$$

De même, S_n est une combinaison linéaire des X_i avec X_1, \dots, X_n gaussien. Nous avons donc

$$S_n \sim \mathcal{N}(n\mu, n\sigma^2), \text{ si } X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2).$$

Que se passe-t-il lorsque que l'on connaît espérance et variance de la v.a. parente, mais rien sur sa loi de probabilité ? Que peut-on dire sur la loi de \bar{X}_n ? Le théorème suivant, absolument essentiel en statistique, donne un résultat théorique qui a des retombées pratiques très importantes.

Théorème 4.1 (Théorème de la limite centrale). Soit X_1, X_2, \dots, X_n un échantillon de loi parente X , d'espérance μ , de variance σ^2 et de loi inconnue. Alors, la moyenne empirique \bar{X}_n vérifie

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

De manière équivalente, la v.a. $S_n = n\bar{X}_n = \sum_{i=1}^n X_i$ vérifie

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Remarque 4.4. Le \mathcal{L} signifie « convergence en loi », c'est-à-dire que \bar{X}_n a la même loi qu'une $\mathcal{N}(\mu, \sigma^2)$ lorsque $n = \infty$.

Évidemment, la taille n de l'échantillon peut se rapprocher de l'infini, mais elle ne l'atteint jamais. D'un point de vue pratique, ce théorème veut dire que, *pour n suffisamment grand*, la moyenne et la somme de « beaucoup » de v.a. indépendantes de même loi ont tendance à se comporter comme une v.a. gaussienne. On dit que l'on *approche*

- la loi de \bar{X}_n par $\mathcal{N}(\mu, \sigma^2/n)$,
- la loi de S_n par $\mathcal{N}(n\mu, n\sigma^2)$.

On propose souvent la frontière $n \geq 30$ pour pouvoir appliquer le TLC⁴. Bien sûr, ça n'est pas une limite stricte, puisque l'approximation dépend en grande partie de la nature de la loi des X_i : si elles sont proches de la loi gaussienne, l'approximation sera valable pour de faibles valeurs de n ; au contraire, il faudra n grand si les X_i sont très loin d'être gaussiennes.

Remarque 4.5. *Attention, ce théorème ne veut certainement pas dire que les X_i deviennent des v.a. gaussiennes lorsqu'on en observe beaucoup ! Il s'agit d'un résultat portant sur le comportement globale (la somme, la moyenne), et pas sur le comportement individuel de chaque variable.*

Exemple 4.2. *La quantité de nicotine contenue dans une marque M de cigarettes est une v.a. d'espérance $\mu = 0.8\text{mg}$ et d'écart-type $\sigma = 0.1\text{mg}$.*

Une fumeur consomme 5 paquets de 20 cigarettes de la marque M par semaine. Quelle est la probabilité pour que la quantité de nicotine consommée soit supérieure à 82mg ?

On pose X_1, X_2, \dots, X_n un échantillon aléatoire décrivant la quantité de nicotine dans chacune des n cigarettes. Ici, $n = 5 \times 20 = 100$. Chaque X_i a pour espérance et variance $\mathbb{E}(X_i) = \mu = 0.8$ et $\text{Var}(X_i) = \sigma^2 = 0.01$. On pose $S_n = \sum_{i=1}^n X_i$, la v.a. décrivant la quantité de nicotine consommée en une semaine. On cherche

$$\mathbb{P}(S_n > 82).$$

On ne connaît pas la loi des X_i , ni celle de S_n . Mais d'après le TLC la loi de S_n est proche de $\mathcal{N}(n\mu, n\sigma^2)$. Ainsi la v.a. centrée et réduite $(S_n - n\mu) / (\sqrt{n\sigma^2})$ est proche d'une $\mathcal{N}(0, 1)$. D'où

$$\begin{aligned} \mathbb{P}(S_n > 82) &= \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} > \frac{82 - n\mu}{\sigma\sqrt{n}}\right) \\ &= 1 - \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \frac{82 - n\mu}{\sigma\sqrt{n}}\right) \\ &\approx 1 - \Phi\left(\frac{82 - 80\mu}{0.1\sqrt{100}}\right) = 1 - \Phi(2) = 0.0228. \end{aligned}$$

Exemple 4.3 (Application à la loi binomiale).

Que peut-on dire de l'application du TLC lorsque les v.a. de l'échantillon sont des v.a. de Bernoulli de paramètre p , c'est-à-dire $\mathcal{B}(p)$?

Nous savons que la somme $S_n = \sum_{i=1}^n X_i$ de n v.a. de Bernoulli suit une loi binomiale $b(n, p)$. Quant à l'espérance et à la variance des X_i , on sait dans le cas de la loi de Bernoulli⁵ que

$$\mathbb{E}(X_i) = p, \quad \text{Var}(X_i) = p(1 - p), \quad \text{pour } i = 1, \dots, n.$$

⁴Pour « théorème de la limite central ». On écrit parfois TCL pour « théorème central limite ».

⁵Revoir le chapitre 1 !

Si on applique le TLC, nous avons donc pour n assez grand, que la loi de S_n peut être approchée par $\mathcal{N}(np, np(1-p))$.

En pratique, on considère que l'approximation est bonne lorsque les valeurs se situent autour de $n \geq 30, p < 0.1, np(1-p) > 5$.

4.3 Estimateur

Notions

Modèle statistique, Estimateur sans biais, Convergent, Risque quadratique, Estimateur de l'espérance, Estimateur de la variance

Nous savons dorénavant manipuler un échantillon aléatoire et calculer des grandeurs (des statistiques) relatives à cet échantillon. Reprenons alors notre problème, qui consiste à déterminer à partir d'un échantillon les paramètres d'une loi dont on suppose qu'elle décrit bien le caractère d'une population auquel on s'intéresse. Cette démarche est appelée démarche d'*estimation*.

4.3.1 Modèle statistique

Dans le cas général, on s'intéresse à un caractère X d'une population \mathcal{P} , dont la distribution de probabilité est inconnue. Notons F la fonction de répartition (inconnue) de X .

Proposer un *modèle statistique*, c'est proposer une famille de lois de probabilité qui correspond *a priori* au caractère d'intérêt, indicée par un jeu de paramètres. Si on note F_θ la fonction de répartition avec θ comme jeu de paramètres, on dispose donc d'une famille de distributions potentielles pour F , telles que

$$F \in (F_\theta)_{\theta \in \Theta},$$

où Θ est un ensemble contenant toutes les valeurs possibles de θ .

Exemple 4.4. Soit X la taille en centimètre des hommes en France. On suppose que la v.a. X est gaussienne. En faisant une telle supposition, on propose le modèle statistique suivant :

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

Si suppose, pour simplifier, que la variance σ^2 est connue et l'espérance μ inconnue, alors le jeu de paramètre θ à estimer correspond au seul paramètre μ . Les valeurs possibles de μ sont tous les réels de la droite \mathbb{R} .

Estimer le paramètre θ , c'est trouver la « meilleure » valeur de θ en fonction des données dont on dispose, c'est-à-dire en fonction de l'information apportée par un échantillon X_1, X_2, \dots, X_n du caractère auquel on s'intéresse. C'est un des problèmes que l'on se pose en *estimation statistique*.

Définition 4.3. Soit X_1, \dots, X_n un échantillon aléatoire de v.a. parente X . Soit $(F_\theta), \theta \in \Theta$ un modèle statistique pour la loi de X . On appelle *estimateur* du paramètre θ une statistique notée $\hat{\theta}(X_1, \dots, X_n)$ dont chaque réalisation peut être considérée comme une approximation de θ .

Remarque 4.6. Pour alléger les notations, on écrit souvent simplement $\hat{\theta}$ un estimateur de θ , et on dit « θ chapeau ».

Comme nous l'avons dit, il existe une infinité de statistiques possibles pour un échantillon. Donc une infinité d'estimateurs potentiels pour un paramètre ! Comment qualifier le fait qu'un estimateur soit « meilleur » qu'un autre ? Les propriétés définies dans le prochain paragraphe permettent d'apporter des éléments de réponse à cette question.

4.3.2 Propriétés élémentaires d'un estimateur

Définition 4.4 (Biais d'un estimateur). Soit $\hat{\theta}$ un estimateur du paramètre θ . On appelle *biais* de l'estimateur la quantité définie par

$$b(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \theta).$$

En fait, le biais décrit simplement comment, *en moyenne*, on s'écarte de la vraie valeur de θ avec notre estimateur $\hat{\theta}$. Évidemment, on a tout intérêt à ce que le biais soit nul !

Propriété 4.1 (Estimateur sans biais). Un estimateur est « sans biais » si $b(\hat{\theta})$ est nul, c'est-à-dire si

$$\mathbb{E}(\hat{\theta}) = \theta.$$

Cette propriété sur l'écartement de la valeur qu'on veut estimer est intéressante... mais certainement pas suffisante ! On a par ailleurs tout intérêt à ce que la variance de l'estimateur ne soit pas trop grande : si un estimateur varie trop, on ne peut pas lui donner une très grande confiance, car il se peut que la valeur qu'on ait calculée sur les données dont on dispose soit très loin de la bonne. Au contraire, si la variance est faible, on est à peu près sûr que, même si on refaisait les calculs avec un autre échantillon, on aurait des valeurs proches.

En fait, le mieux est de pouvoir exprimer la variance d'un estimateur en fonction de la taille du jeu de données et de montrer qu'elle diminue lorsque la taille de l'échantillon augmente : *grosso modo*, cela signifie que l'on est de plus en plus sûr de notre approximation. Et plus elle diminue vite, moins on a besoin de données.

Quand on utilise un estimateur, on regarde donc avant tout son biais et sa variance pour pouvoir le caractériser. La notion de convergence résume ces bonnes propriétés :

Proposition 4.3. Si un estimateur $\hat{\theta}$ est sans biais et si $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}) = 0$, alors on dit qu'il est *convergent*.

Introduisons un dernier indicateur pour jauger de la qualité d'un estimateur.

Définition 4.5. On appelle *risque quadratique* la grandeur définie par

$$\begin{aligned} R(\hat{\theta}) &= \mathbb{E}[(\theta - \hat{\theta})^2] \\ &= \text{biais}(\hat{\theta})^2 + \text{Var}(\hat{\theta}). \end{aligned}$$

Il est en effet courant de devoir faire un compromis entre l'*exactitude* d'un estimateur (son biais) et sa *précision* (sa variance). Le risque quadratique faisant intervenir ces deux grandeurs, il est tout indiqué pour doser le compromis biais/variance.

4.3.3 \bar{X}_n est un estimateur convergent de l'espérance d'une loi

En début de chapitre, nous nous sommes posé la question de l'estimation du paramètre d'espérance μ d'une v.a. $\mathcal{N}(\mu, \sigma^2)$ lorsque l'on dispose d'un échantillon X_1, \dots, X_n . La moyenne empirique \bar{X}_n semble appropriée, mais voyons comment se comportent les propriétés de cet estimateur.

Proposition 4.4. Soit X une variable aléatoire d'espérance μ et de variance σ^2 . La moyenne empirique \bar{X}_n est un estimateur sans biais et convergent de μ . Son risque quadratique se résume à sa variance.

Démonstration. Nous avons vu que $\mathbb{E}(\bar{X}_n) = \mathbb{E}(X) = \mu$. De plus, $\text{Var}(\bar{X}_n) = \sigma^2/n$, ce qui tend évidemment vers zéro quand $n \rightarrow \infty$. Puis, en utilisant la proposition 4.3, on a bien convergence de \bar{X}_n . \square

Donc, la moyenne empirique est un bon estimateur du paramètre μ d'une loi normale. Mais pas uniquement ! Traitons un autre exemple :

Exemple 4.5. On suppose que la durée de vie d'une ampoule électrique d'une certaine marque est une v.a. X de loi exponentielle. On rappelle que la densité d'une loi exponentielle de paramètre λ est donnée par

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x).$$

L'espérance et la variance de X sont respectivement égales à $1/\lambda$ et à $1/\lambda^2$. On aimerait estimer le paramètre de cette loi. ⁶

Pour ce faire, on réalise des mesures sur 200 ampoules : on obtient une durée de vie moyenne de 10.2 mois. Comment estimer la valeur de λ ?

Tout d'abord, formalisons un peu tout cela : on dispose d'un échantillon de données X_1, \dots, X_n de loi parente $X \sim \mathcal{E}(\lambda)$, où X_i représente la durée de vie de l'ampoule i . Nous avons $n = 200$ observations, pour lesquelles on nous donne la durée de vie moyenne, c'est-à-dire, une réalisation $\bar{x}_{200} = 10.2$ de la moyenne empirique \bar{X}_n ! Or, on sait que la moyenne empirique est un estimateur sans biais et convergent de l'espérance : dans le cas de la loi exponentielle, l'espérance vaut $1/\lambda$. Donc, la valeur obtenue est une bonne approximation de $1/\lambda$. On obtient une approximation de λ en inversant la valeur obtenue pour la moyenne empirique, d'où l'estimateur

$$\hat{\lambda}(X_1, \dots, X_n) = \frac{1}{\bar{X}_n},$$

ce qui fait, sur ce jeu de données, une valeur approchée de $1/10.2 = 0.0980392$.

Question subsidiaire, néanmoins facile : saurez-vous donner un estimateur de la variance de X ?

⁶On tient là un modèle statistique ! On se donne une famille de loi pour X , toutes les lois exponentielles, mais on se pose la question de la valeur du paramètre de cette loi par rapport à un jeu de données

4.3.4 Variance empirique et variance empirique corrigée

On se pose maintenant la question de l'estimation du paramètre de variance σ^2 d'une v.a. de loi quelconque. La variance empirique, définie par

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

semble être l'estimateur naturel. Cependant, en calculant l'espérance de S^2 , on s'aperçoit que c'est un estimateur biaisé de σ^2 ! Il est facile d'introduire un facteur de correction qui permette d'obtenir un estimateur *sans biais* de la variance d'une v.a. C'est ainsi que ce construit la variance empirique corrigée⁷ $S^{*2} = \frac{n}{n-1} S^2$.

Proposition 4.5. Soit X_1, \dots, X_n un échantillon de variable parente X , d'espérance μ et de variance σ^2 . Alors,

$$\mathbb{E}(S^2) = \frac{n-1}{n} \sigma^2, \quad \mathbb{E}(S^{*2}) = \sigma^2.$$

Démonstration. Le calcul de l'espérance de S^2 est assez simple : on commence par appliquer les règles de calculs de l'espérance

$$\mathbb{E}(S^2) = \mathbb{E} \left(\frac{1}{n} \sum X_i^2 - \bar{X}_n^2 \right) = \mathbb{E}(X_i^2) - \mathbb{E}(\bar{X}_n^2).$$

Puis, il suffit de « retourner » la formule $\text{Var}(X_i) = \mathbb{E}(X_i^2) - (\mathbb{E}X_i)^2$ pour isoler $\mathbb{E}(X_i^2)$, ainsi,

$$\mathbb{E}(X_i^2) = \text{Var}(X_i) + (\mathbb{E}X_i)^2 = \sigma^2 + \mu^2.$$

De même pour $\mathbb{E}(\bar{X}_n^2)$, on a

$$\mathbb{E}(\bar{X}_n^2) = \text{Var}(\bar{X}_n) + (\mathbb{E}\bar{X}_n)^2 = \frac{\sigma^2}{n} + \mu^2.$$

Finalement,

$$\mathbb{E}(S^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2.$$

L'espérance de S^{*2} est immédiate ! On a construit cette v.a. de telle sorte à ce que son espérance soit le paramètre σ^2 :

$$\mathbb{E}(S^{*2}) = \mathbb{E} \left(\frac{n}{n-1} S^2 \right) = \frac{n}{n-1} \times \frac{n-1}{n} \sigma^2 = \sigma^2.$$

□

⁷On pourrait également montrer que c'est un estimateur convergent