

Chapitre 2

Estimation de paramètres

Considérons des variables aléatoires X_1, \dots, X_n dont les mesures lors d'une expérience sont données par x_1, \dots, x_n . En général, on désire que X_1, \dots, X_n suivent une même loi qu'une variable aléatoire X d'espérance μ et de variance σ^2 .

Si on reprend l'exemple de nos 60 lancers d'un dé (à la section 1.1.2), X_i est la variable aléatoire qui donne le résultat du dé au i -ième jet, et x_i le résultat obtenu¹ lors du i -ième jet en réalisant l'expérience. On avait par exemple, $x_1 = 2$ et $x_{60} = 1$. Quant à X , il s'agit de la variable aléatoire qui représente le nombre montré par le jet de ce dé; son espérance vaut $\mu = 3.5$ et sa variance σ^2 vaut environ 2.92 (voir section 1.1.3).

2.1 Paramètres de position

Ces paramètres permettent de situer nos mesures x_1, \dots, x_n .

1. La moyenne arithmétique

La *moyenne arithmétique*, notée \bar{x} , est définie par

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \left(= \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_{n-1} + x_n}{n} \right)$$

C'est le paramètre de position qui est utilisé pour estimer la note semestrielle des élèves du point de vue de l'évaluation dans le bulletin scolaire.

La variable aléatoire correspondante à la moyenne arithmétique est notée ainsi

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

\bar{X} est un *estimateur* de l'espérance théorique μ , et \bar{x} est son *estimation*.

Remarque

C'est avec cet estimateur qu'on estime votre niveau dans chaque discipline lors de chaque semestre.

1. On voit mieux pourquoi on parle de variable aléatoire : la variable aléatoire X_i n'a rien d'aléatoire, mais sa réalisation x_i est aléatoire.

2. La médiane

Pour calculer la *médiane*, il faut classer les mesures de la plus petite à la plus grande. La médiane est la valeur du milieu lorsque le nombre des mesures est impair. C'est la moyenne arithmétique des deux valeurs autour du milieu lorsque le nombre des mesures est pair.

Par exemple

Mesures	Médiane
1 2 3 4 5	3
1 2 3 4 5 6	3.5
1 1 1 3 3 3	2

3. Le mode

Le *mode* est la valeur la plus fréquemment atteinte. Il est possible qu'une série de mesures ait plusieurs modes. Pour cette raison, le mode est un sous-ensemble des valeurs possibles.

Par exemple

Mesures	Mode
1 1 2 2 2 4 5	{2}
1 1 1 3 3	{1}
1 1 1 3 3 3	{1,3}

2.2 Paramètres de dispersion

Ces paramètres permettent de mesurer l'écartement des mesures x_1, \dots, x_n .

1. L'étendue

L'*étendue* est la longueur du *domaine de variation* qui est le plus petit intervalle contenant toutes les valeurs.

Par exemple

Mesures	Domaine de variation	étendue
1 2 2 4 5	[1, 5]	4
1 1 3 3 3 3	[1, 3]	2
1 1 3 3 7	[1, 7]	6
2 2 2 3 3 7	[2, 7]	5

2. La variance

La *variance*, notée s_n^2 , est la moyenne des carrés des écarts par rapport à \bar{x} .

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \left(= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} \right)$$

Par exemple, si les mesures sont $x_1 = 1, x_2 = 2, x_3 = 2, x_4 = 4$ et $x_5 = 6$. Alors la moyenne arithmétique vaut $\bar{x} = 3$ et on peut calculer la variance ainsi

$$\begin{aligned} s_n^2 &= \frac{(1-3)^2 + (2-3)^2 + (2-3)^2 + (4-3)^2 + (6-3)^2}{5} \\ &= \frac{4 + 1 + 1 + 1 + 9}{5} = \frac{16}{5} = 3.2 \end{aligned}$$

La variable aléatoire correspondante à la variance est notée ainsi

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

S_n^2 est un *estimateur* de la variance théorique σ^2 , et s_n^2 est son *estimation*.

3. L'écart type

L'*écart type* est égal à la racine carrée de la variance. Il est logiquement noté s_n .

$$s_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad \left(= \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}} \right)$$

Dans l'exemple précédent, on avait la variance $s_n^2 = 3.2$. Ainsi l'écart type correspondant vaut $s_n = \sqrt{3.2} \cong 1.79$.

De même, la variable aléatoire correspondante à l'écart type est notée ainsi

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

S_n est un *estimateur* de l'écart type théorique σ , et s_n est son *estimation*.

L'estimateur sans biais de la variance et de l'écart type

L'estimateur de l'espérance μ donné par

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

est un bon estimateur : on dit qu'il est sans biais (voir section 1).

Par contre, l'estimateur de la variance σ^2 donné par

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

a un biais, il a en effet tendance à sous-estimer σ^2 .

On corrige ce biais en utilisant l'estimateur, noté S_{n-1} ou plus simplement S , défini par

$$S^2 = S_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \left(= \frac{n}{n-1} S_n^2 \right)$$

Pour les mêmes raisons, l'estimateur sans biais de l'écart type σ est

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

2.3 Les boîtes à moustaches

Inventées par John Tukey et John Wilder en 1977, les boîtes à moustaches, aussi appelées diagrammes en boîte ou boîtes de Tukey) sont très utiles pour donner une représentation simple à comprendre des données mesurées (un peu comme un histogramme).

2.3.1 Les quartiles de Tukey et les quartiles de Freund et Perles

Les *quartiles* permettent de séparer les données en quatre parties égales ; c'est une généralisation de la médiane qui sépare les données en deux parties égales.

Le quartile inférieur sépare le premier quart des données ; le deuxième quartile est la médiane : il sépare la moitié des données ; le quartile supérieur sépare le dernier quart des données.

Voici deux manières de calculer les quartiles à partir d'une série de données.

1. Les quartiles utilisés pour faire les boîtes à moustaches.

Cette méthode a été établie par John Tukey en 1983.

On met les n données dans l'ordre et on coupe les données en deux ensembles sur lesquels on calcule la médiane.

- (a) Si n est impair, il y a une valeur centrale (la médiane), et on coupe les données en deux ensembles en mettant la médiane dans chacun des deux ensembles. Le quartile inférieur est la médiane du premier ensemble ; le quartile supérieur est la médiane du deuxième ensemble.
- (b) Si n est pair, il y a deux valeurs centrales (la médiane est la moyenne arithmétique de ces deux valeurs), et on coupe en deux ensembles en mettant dans chaque ensemble la valeur centrale correspondante.

Voici des exemples pour cette première méthode.

Mesures	Les ensembles	Quartile inférieur	Quartile supérieur
1 2 4 5	{ 1, 2 } { 4, 5 }	1.5	4.5
1 3 5 5 7	{ 1, 3, 5 } { 5, 5, 7 }	3	5
1 3 4 6 7 9	{ 1, 3, 4 } { 6, 7, 9 }	3	7
1 3 5 6 7 9 15	{ 1, 3, 5, 6 } { 6, 7, 9, 15 }	4	8

Sous «R», on trouve ces quartiles en utilisant la commande `fivenum` (qui donne le minimum, le quartile inférieur, la médiane, le quartile supérieur, le maximum).

2. Les quartiles utilisés par les logiciels «Excel», «OpenOffice» et «R».

Cette méthode a été établie par John Freund et Benjamin Perles en 1987.

On met les n données dans l'ordre. Le quartile inférieur est la donnée numéro $\frac{1n+3}{4}$, le quartile supérieur est la donnée numéro $\frac{1+3n}{4}$. Si ces numéros ne sont pas entiers, on utilise une interpolation linéaire pour trouver les quartiles inférieur et supérieur (voir page suivante).

Lorsque n est impair, on retrouve les quartiles de Tukey.

Concernant l'interpolation linéaire

Si on cherche la donnée numéro 3.25, on calcule la moyenne pondérée $\frac{3x_3+x_4}{4}$ (on voit l'analogie avec les notes : 3.25 est la moyenne des notes 3, 3, 3 et 4).

Si on cherche la donnée numéro 3.5, on calcule la moyenne pondérée $\frac{2x_3+2x_4}{4} = \frac{x_3+x_4}{2}$.
Si on cherche la donnée numéro 3.75, on calcule la moyenne pondérée $\frac{x_3+3x_4}{4}$.

Reprenons les mêmes exemples que pour la première méthode.

Mesures	$\frac{1n+3}{4}$	$\frac{3n+1}{4}$	Quartile inférieur	Quartile supérieur
1 2 4 5	1.75	3.25	1.75	4.25
1 3 5 5 7	2	4	3	5
1 3 4 6 7 9	2.25	4.75	3.25	6.75
1 3 5 6 7 9 15	2.5	5.5	4	8

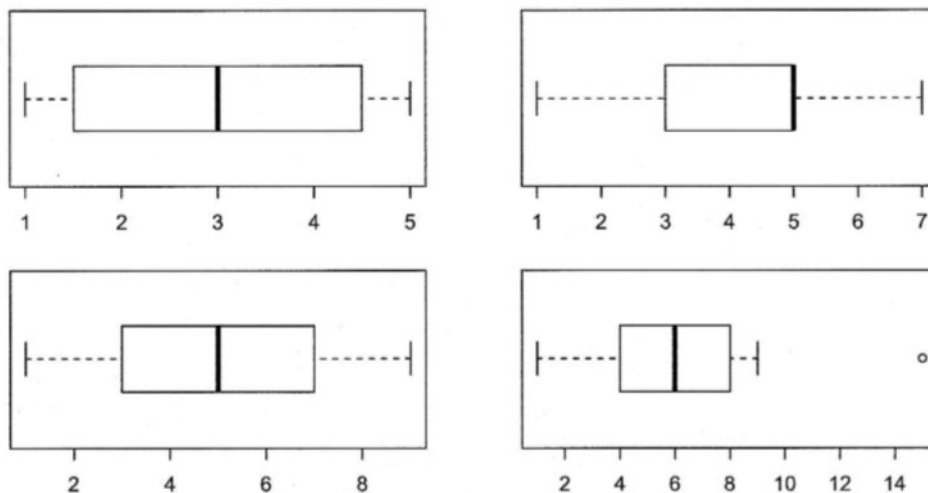
Sous «R», on trouve ces quartiles en utilisant la commande `quantile` (qui donne le minimum, le quartile inférieur, la médiane, le quartile supérieur, le maximum).

Sous «OpenOffice», on les trouve à l'aide de la commande `quantile(zone;nombre)` (qui donne le minimum pour `nombre = 0`, le quartile inférieur pour `nombre = 1`, la médiane pour `nombre = 2`, le quartile supérieur pour `nombre = 3`, le maximum pour `nombre = 4`).

2.3.2 La boîte à moustaches de Tukey

C'est une boîte dont les côtés correspondent aux quartiles inférieur et supérieur; on indique la médiane par un trait à l'intérieur de cette boîte. Des moustaches sortent de la boîte pour indiquer les valeurs minimale et maximale. Ces moustaches² sont au plus 50% plus longues que la boîte, sinon on prend les valeurs minimale ou maximale suivantes et on indique les valeurs extrêmes par des points : ce sont les valeurs atypiques.

Voici les boîtes à moustaches qui correspondent aux exemples ci-dessus (avec la première méthode). Sur le dernier exemple, la longueur de la boîte vaut 4 ($= 8 - 4$) et la dernière valeur est atypique (car elle est à distance 7 (> 6) du quartile supérieur).



Sous «R», on peut dessiner ces boîtes à moustaches en utilisant la commande `boxplot`.

2. Selon les contextes, les moustaches peuvent être construite de manière plus théoriques.

2.4 Statistique, estimateur et estimation

Une pièce de monnaie est jetée six fois de suite. Imaginons que l'on obtienne la suite $(P; P; F; P; F; P)$ où P signifie que la pièce est tombée sur le côté pile et F sur le côté face.

Du point de vue des probabilités, à chaque fois qu'on lance une pièce de monnaie, on a une probabilité $p \in [0, 1]$ d'obtenir le côté pile et $q = 1 - p$ d'obtenir le côté face. Si $p = 1$, cela signifie que la pièce est truquée et qu'elle a deux côtés pile. Si $p = \frac{2}{3}$, cela signifie que la pièce n'est pas bien équilibrée et qu'elle tombe deux fois plus souvent sur pile que face. Si $p = \frac{1}{2}$ cela signifie que la pièce est bien équilibrée. Il n'existe probablement aucune pièce de monnaie sur terre qui soit parfaitement bien équilibrée, néanmoins on se permet la plupart du temps de supposer que $p = \frac{1}{2}$!

À une telle expérience, on fait correspondre une famille de variables aléatoires X_i qui indiquent 1 si le i -ième jet tombe sur pile et 0 s'il tombe sur face. Ainsi, la suite obtenue se traduit par $(x_1, x_2, x_3, x_4, x_5, x_6) = (1, 1, 0, 1, 0, 1)$ où les x_i sont les données observées. Comme la pièce de monnaie reste à la même à chaque nouveau lancé, on peut raisonnablement supposer que chacune de ces variables aléatoires suit la même loi (c'est une loi de Bernoulli comme vue en section 1.1.6) donnée par

$$\mathbf{P}(X_i = 1) = p \quad \text{et} \quad \mathbf{P}(X_i = 0) = 1 - p$$

Dans ce cas, il était facile de trouver la distribution (comme la plupart du temps). Il est déjà plus difficile d'estimer son paramètre, ici p !

Définition

Toute fonction T des variables aléatoires X_1, X_2, \dots, X_n est appelée *statistique*. Lorsqu'une telle fonction est utilisée pour estimer des paramètres d'une même distribution des variables aléatoires X_1, \dots, X_n , on dit que $T(X_1, \dots, X_n)$ est un *estimateur*.

Notations

1. Si $T = T(X_1, \dots, X_n)$ est l'estimateur du paramètre inconnu, $t = T(x_1, \dots, x_n)$ est la *réalisation de cet estimateur*, aussi appelée *estimation*, à l'aide des données observées.
2. Lorsqu'on estime un paramètre θ , on a l'habitude d'écrire son estimateur avec un chapeau : $\hat{\theta}$.

Remarque

Une statistique T est une variable aléatoire !

Exemple

Dans notre exemple, il est naturel d'estimer p à l'aide de l'estimateur suivant :

$$\hat{P} = \hat{P}(X_1, \dots, X_6) = \frac{\text{nombre de pile}}{\text{nombre total de jets}} = \frac{\sum_{i=1}^6 X_i}{6} = \bar{X}$$

La réalisation de cet estimateur est dans notre cas l'estimation $\hat{p} = \frac{\sum_{i=1}^6 x_i}{6} = \frac{4}{6} = \frac{2}{3}$.

2.5 La méthode du maximum de vraisemblance

Une des techniques parmi les plus populaires pour trouver des estimateurs est celle dite du *maximum de vraisemblance*. Si X_1, \dots, X_n est une famille de variables aléatoires indépendantes issues de la même densité $f(x; \theta)$ (où θ est le paramètre de la densité f). La fonction de vraisemblance³ est définie par :

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

Justification de cette formule dans le cas discret

Il faut penser à $L(\theta; x_1, \dots, x_n)$ comme étant la probabilité que l'échantillon x_1, \dots, x_n se produise. Dans le cas discret on a

$$L(\theta; x_1, \dots, x_n) = \mathbf{P}(X_1 = x_1 \text{ et } X_2 = x_2 \text{ et } \dots \text{ et } X_n = x_n)$$

Comme les X_i sont indépendants, on a $L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n \mathbf{P}(X_i = x_i) = \prod_{i=1}^n f(x_i; \theta)$.

L'estimateur du maximum de vraisemblance

L'estimateur du maximum de vraisemblance⁴ de θ , noté $\hat{\theta}_{\text{MLE}}$, est l'estimateur pour lequel la fonction de vraisemblance est la plus grande. Autrement dit, il s'agit de la valeur du paramètre pour laquelle l'échantillon observé est le plus vraisemblable.

Pour chercher cet estimateur, on doit trouver le maximum de la fonction $L(\theta; x_1, \dots, x_n)$. C'est une optimisation classique : un symptôme d'un maximum est un zéro de la dérivée.

En pratique, il peut être plus facile de trouver le maximum du logarithme de la fonction de vraisemblance (en effet, la fonction \ln est croissante et transforme les produits en somme; ainsi $\ln(L(\theta; x_1, \dots, x_n)) = \sum_{i=1}^n \ln(f(x_i; \theta))$).

Exemple : les MLEs de la loi normale

On considère la loi normale de paramètres μ et σ dont la densité est donnée en page 16. En exercice, on trouvera⁵ les estimateurs du maximum de vraisemblance suivants.

$$\hat{\mu}_{\text{MLE}} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{et} \quad \hat{\sigma}_{\text{MLE}} = \sqrt{\frac{\sum_{i=1}^n (X_i - \hat{\mu}_{\text{MLE}})^2}{n}}$$

Théorème Si f est une fonction d'un estimateur θ , alors $\widehat{f(\theta)}_{\text{MLE}} = f(\hat{\theta}_{\text{MLE}})$.

Idée de la preuve

Le MLE $\hat{\theta}_{\text{MLE}}$ est la valeur pour laquelle la fonction de vraisemblance est maximale. On nomme $\nu = f(\theta)$ un nouveau paramètre. On doit exprimer la fonction de vraisemblance en fonction de ce nouveau paramètre; on a $\theta = f^{-1}(\nu)$ (on crée une fonction réciproque de f en choisissant un domaine de définition). Ainsi $L(\theta) = L(f^{-1}(\nu))$. Cette fonction admet un maximum lorsque $f^{-1}(\hat{\nu}_{\text{MLE}}) = \hat{\theta}_{\text{MLE}}$, c'est-à-dire $\hat{\nu}_{\text{MLE}} = f(\hat{\theta}_{\text{MLE}})$.

3. En anglais, vraisemblance se dit likelihood. Ainsi la fonction de vraisemblance est notée L .

4. En anglais, estimateur du maximum de vraisemblance se dit maximum likelihood estimator (MLE).

5. Lorsqu'on cherche à optimiser une fonction à plusieurs variables, on cherche les valeurs des variables qui annulent le *gradient* (c'est le vecteur dont la i -ème composante est la dérivée de la fonction par rapport à la i -ème variable uniquement).

2.6 Qualités d'un estimateur

Le biais d'un estimateur

Soit $\hat{\theta}$ un estimateur d'un paramètre θ . Le *biais* de $\hat{\theta}$ est défini par :

$$b_{\theta}(\hat{\theta}) = E(\hat{\theta} - \theta) \stackrel{\substack{\text{linéarité de} \\ \text{l'espérance}}}{=} E(\hat{\theta}) - \theta$$

Un estimateur $\hat{\theta}$ est dit *sans biais* lorsque $b_{\theta}(\hat{\theta}) = 0$. Autrement dit, lorsqu'en moyenne l'estimateur est égal au paramètre que l'on cherche à estimer.

Exemple : les estimateurs les plus célèbres

On suppose qu'on a n variables aléatoires X_i indépendantes et suivant toutes une même loi d'espérance μ et de variance σ^2 . Alors l'espérance μ peut être estimée sans biais par l'estimateur \bar{X} et la variance σ^2 peut être estimée sans biais par l'estimateur S^2 . Ces estimateurs sont définis comme suit :

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

Dans le cas d'une loi normale $\mathcal{N}(\mu, \sigma^2)$, on remarque que \bar{X} n'est rien d'autre que l'estimateur de vraisemblance de μ , tandis que $S^2 = \frac{n}{n-1} \hat{\sigma}_{\text{MLE}}^2$.

1. L'estimateur de la moyenne est sans biais

En effet, on a :

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) \stackrel{\substack{\text{linéarité de} \\ \text{l'espérance}}}{=} \frac{\sum_{i=1}^n E(X_i)}{n} = \frac{\sum_{i=1}^n \mu}{n} = \frac{n\mu}{n} = \mu$$

Donc l'estimateur \bar{X} de l'espérance μ est sans biais, car $b_{\mu}(\bar{X}) = E(\bar{X}) - \mu = 0$.

2. L'estimateur de la variance est sans biais

En effet, on a :

$$\begin{aligned} E(S^2) &= E\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right) \stackrel{\text{lin.}}{=} \frac{1}{n-1} \sum_{i=1}^n E((X_i - \bar{X})^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n E(X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &\stackrel{\text{lin.}}{=} \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - 2E\left(\sum_{i=1}^n X_i \cdot \bar{X}\right) + \sum_{i=1}^n E(\bar{X}^2) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - 2nE\left(\frac{\sum_{i=1}^n X_i}{n} \cdot \bar{X}\right) + nE(\bar{X}^2) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - 2nE(\bar{X} \cdot \bar{X}) + nE(\bar{X}^2) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2) \right) \end{aligned}$$

Or, d'après la seconde formule de la variance, on a $V(X) = E(X^2) - E(X)^2$. Ainsi, on a la formule suivante pour n'importe quelle variable aléatoire X :

$$E(X^2) = V(X) + E(X)^2 \quad \star$$

Cela permet de reprendre le calcul précédent :

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2) \right) \\ &\stackrel{\star}{=} \frac{1}{n-1} \left(\sum_{i=1}^n (V(X_i) + E(X_i)^2) - n(V(\bar{X}) + E(\bar{X})^2) \right) \end{aligned}$$

Puisque l'espérance de chaque X_i est égale à μ , que la variance de chaque X_i est égale à σ^2 et que l'estimateur \bar{X} est sans biais, on a :

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} \left(\sum_{i=1}^n (\sigma^2 + \mu^2) - n(V(\bar{X}) + \mu^2) \right) \\ &= \frac{1}{n-1} (n(\sigma^2 + \mu^2) - n(V(\bar{X}) + \mu^2)) = \frac{1}{n-1} (n\sigma^2 - nV(\bar{X})) \end{aligned}$$

Or, grâce au théorème de la page 24 qui dit que $V(\bar{X}) = \frac{\sigma^2}{n}$ (parce que les variables aléatoires sont indépendantes), on peut finalement montrer que S^2 est sans biais, car :

$$E(S^2) = \frac{1}{n-1} (n\sigma^2 - nV(\bar{X})) = \frac{1}{n-1} \left(n\sigma^2 - n\frac{\sigma^2}{n} \right) = \frac{1}{n-1} (n-1)\sigma^2 = \sigma^2$$

L'erreur quadratique moyenne de l'erreur d'un estimateur

Soit $\hat{\theta}$ un estimateur d'un paramètre θ . L'erreur quadratique moyenne de $\hat{\theta}$ est définie⁶ par :

$$\text{MSE}_\theta(\hat{\theta}) = E((\hat{\theta} - \theta)^2) \stackrel{\star}{=} V(\hat{\theta}) + (b_\theta(\hat{\theta}))^2$$

Où $V(\hat{\theta})$ est la *variance* de $\hat{\theta}$ (rappelons qu'un estimateur est une variable aléatoire).

Théorème (sans preuve)

On suppose qu'on a n variables aléatoires X_i indépendantes et qui suivent toutes une loi normale $\mathcal{N}(\mu, \sigma^2)$ d'espérance μ et de variance σ^2 . On a :

$$\text{MSE}_\mu(\bar{X}) = V(\bar{X}) = \frac{\sigma^2}{n} \quad \text{et} \quad \text{MSE}_{\sigma^2}(S^2) = V(S^2) = \frac{2\sigma^4}{n-1}$$

6. En anglais, on parle de **mean squared error**, abrégée **MSE**.