Probabilités et observations économiques

C. Tuleau-Malot

Université de Nice - Sophia Antipolis

Plan

Vocabulaire

Etude statistique :

- on étudie en général ce que l'on appelle un **caractère** noté généralement X

Vocabulaire

Etude statistique :

- on étudie en général ce que l'on appelle un **caractère** noté généralement X
- ce caractère est associée à une **population** qui est l'ensemble des individus statistiques concernés par l'étude

Vocabulaire

Etude statistique :

- on étudie en général ce que l'on appelle un **caractère** noté généralement X
- ce caractère est associée à une **population** qui est l'ensemble des individus statistiques concernés par l'étude
- dans la pratique, on ne dispose que d'un **échantillon** dont la taille est généralement notée n

Vocabulaire

Etude statistique :

- on étudie en général ce que l'on appelle un **caractère** noté généralement X
- ce caractère est associée à une **population** qui est l'ensemble des individus statistiques concernés par l'étude
- dans la pratique, on ne dispose que d'un **échantillon** dont la taille est généralement notée n

Exemple :

On veut connaître le % de fumeurs chez les plus de 18 ans.
La population est l'ensemble des personnes de plus de 18 ans.
L'échantillon est l'ensemble des personnes sondées.

Vocabulaire (2)

A cette population, on associe une loi de probabilité P :

- si on prend un individu tiré au hasard dans la population, alors le caractère X étudié est une variable aléatoire suivant cette loi P

Vocabulaire (2)

A cette population, on associe une loi de probabilité P :

- si on prend un individu tiré au hasard dans la population, alors le caractère X étudié est une variable aléatoire suivant cette loi P

- pour l'individu i, on note X_i cette variable aléatoire

Vocabulaire (2)

A cette population, on associe une loi de probabilité P :

- si on prend un individu tiré au hasard dans la population, alors le caractère X étudié est une variable aléatoire suivant cette loi P
- pour l'individu i, on note X_i cette variable aléatoire
- l'échantillon de taille n est alors l'ensemble des variables $X_1, \ldots, X_n$ qui sont supposées indépendantes et de même loi. On parle de n variables aléatoires iid (indépendantes et identiquement distribuées)

Vocabulaire (2)

A cette population, on associe une loi de probabilité P :

- si on prend un individu tiré au hasard dans la population, alors le caractère X étudié est une variable aléatoire suivant cette loi P
- pour l'individu i, on note X_i cette variable aléatoire
- l'échantillon de taille n est alors l'ensemble des variables $X_1, \ldots, X_n$ qui sont supposées indépendantes et de même loi. On parle de n variables aléatoires iid (indépendantes et identiquement distribuées)

Exemple :

Dans l'étude sur le % de fumeurs chez les plus de 18 ans, X_i est alors le fait que la personne sondée numéro i soit fumeur ou non

Estimation

En général, cette loi est inconnue.

Estimation

En général, cette loi est inconnue.

Objectifs de l'estimation :

- déterminer certaines caractéristiques de cette loi, et ce à partir de données sur l'échantillon

Estimation

En général, cette loi est inconnue.

Objectifs de l'estimation :

- déterminer certaines caractéristiques de cette loi, et ce à partir de données sur l'échantillon
- paramètres classiques : espérance et variance

Estimation (2)

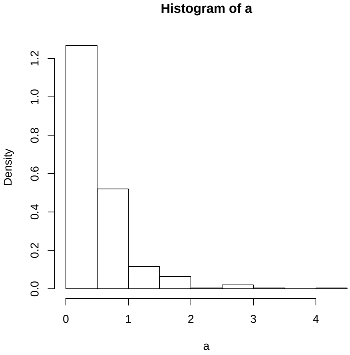
Exemples :

- si on reprend l'exemple du % de fumeurs, les variables X_i suivent une loi de Bernoulli p

$$\Rightarrow \text{comment estimer } p?$$

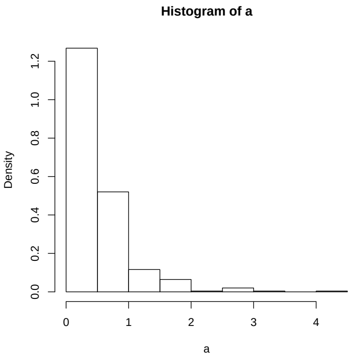
Estimation (2)

- On s'intéresse au temps d'arrivée d'un client dans une file d'attente. On peut obtenir des observations en se rendant à un guichet et en relevant le temps entre l'arrivée de deux clients successifs.



Estimation (2)

- On s'intéresse au temps d'arrivée d'un client dans une file d'attente. On peut obtenir des observations en se rendant à un guichet et en relevant le temps entre l'arrivée de deux clients successifs.



$\Rightarrow$ comment estimer le paramètre de la loi exponentielle?

Estimateur

Estimer une quantité a revient à déterminer un **estimateur** de a à savoir une quantité construite à partir des données et qui approchera a.
Cette quantité est généralement notée :

$$\hat{A}_n := \hat{A}(X_1, \ldots, X_n)$$

Comme $X_1, \ldots, X_n$ sont des variables aléatoires, $\hat{A}_n$ est encore une variable aléatoire.

Estimateur (2)

Différentes questions surviennent :

1. Comment construire un estimateur?

Estimateur (2)

Différentes questions surviennent :

1. Comment construire un estimateur?

2. Quelle est la précision de cet estimateur?

Estimation de l'espérance

Loi des Grands Nombres :
Soit $X_1, \dots, X_n$ des variables aléatoires iid dont on note m l'espérance.
Avec probabilité 1, on a :

$$\frac{X_1 + \dots + X_n}{n} \underset{n \longrightarrow +\infty}{\longrightarrow} m$$

Estimation de l'espérance

Loi des Grands Nombres :
Soit $X_1, \ldots, X_n$ des variables aléatoires iid dont on note m l'espérance.
Avec probabilité 1, on a :

$$\frac{X_1 + \ldots + X_n}{n} \underset{n \longrightarrow +\infty}{\longrightarrow} m$$

Ainsi, une façon d'estimer m est de considérer $\frac{X_1 + \ldots + X_n}{n}$.

Estimation de l'espérance

Loi des Grands Nombres :
Soit $X_1, \ldots, X_n$ des variables aléatoires iid dont on note m l'espérance.
Avec probabilité 1, on a :

$$\frac{X_1 + \ldots + X_n}{n} \underset{n \longrightarrow +\infty}{\longrightarrow} m$$

Ainsi, une façon d'estimer m est de considérer $\frac{X_1+\ldots+X_n}{n}$.

$\frac{X_1+\ldots+X_n}{n}$: variable aléatoire
$\Rightarrow$ on appelle cela un **estimateur**
$\Rightarrow$ une observation d'un estimateur est appelé **estimation**

Estimation de l'espérance (2)

Exemple :

On reprend l'exemple des fumeurs. Nous avons vu que pour chaque individu i, on considère une variable X_i qui n'est autre qu'une variable de Bernoulli de paramètre p.

Estimation de l'espérance (2)

Exemple :

On reprend l'exemple des fumeurs. Nous avons vu que pour chaque individu i, on considère une variable X_i qui n'est autre qu'une variable de Bernoulli de paramètre p.

Or :

$$p = \mathbb{E}(X_1) = \ldots = \mathbb{E}(X_n)$$

Estimation de l'espérance (2)

Exemple :

On reprend l'exemple des fumeurs. Nous avons vu que pour chaque individu i, on considère une variable X_i qui n'est autre qu'une variable de Bernoulli de paramètre p.

Or :

$$p = \mathbb{E}(X_1) = \ldots = \mathbb{E}(X_n)$$

Un estimateur $\hat{P}_n$ est $\frac{X_1+\ldots+X_n}{n}$.

Si l'on note $x_1, \ldots, x_n$ un jeu de données associé à $X_1, \ldots, X_n$, une estimation est alors $\frac{x_1+\ldots+x_n}{n}$.

Estimation de l'espérance (2)

Exemple :

On reprend l'exemple des fumeurs. Nous avons vu que pour chaque individu i, on considère une variable X_i qui n'est autre qu'une variable de Bernoulli de paramètre p.

Or :

$$p = \mathbb{E}(X_1) = \ldots = \mathbb{E}(X_n)$$

Un estimateur $\hat{p}_n$ est $\frac{X_1 + \ldots + X_n}{n}$.

Si l'on note $x_1, \ldots, x_n$ un jeu de données associé à $X_1, \ldots, X_n$, une estimation est alors $\frac{x_1 + \ldots + x_n}{n}$.

Jeu de données : 0 0 0 1 0 0 1 1 1 0 0 0 1 0 0 0 1 0 0 0 0 1 0 1 0 0 0 0 1 0 0 1 0 0 0 0 1 1 0 0 0 1 0 0 0 0

Estimation de l'espérance (2)

Exemple :

On reprend l'exemple des fumeurs. Nous avons vu que pour chaque individu i, on considère une variable X_i qui n'est autre qu'une variable de Bernoulli de paramètre p.

Or :

$$p = \mathbb{E}(X_1) = \ldots = \mathbb{E}(X_n)$$

Un estimateur $\hat{p}_n$ est $\frac{X_1+\ldots+X_n}{n}$.

Si l'on note $x_1, \ldots, x_n$ un jeu de données associé à $X_1, \ldots, X_n$, une estimation est alors $\frac{x_1+\ldots+x_n}{n}$.

Jeu de données : 0 0 0 1 0 0 1 1 1 0 0 0 1 0 0 0 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 1 1 0 0 0 1 0 0 0 0

Une estimation de p relative à ce jeu de données est : $\frac{13}{50} = 0.26$

Estimation sur un exemple

Autre exemple :

Je prends le bus chaque matin et chaque matin je me demande quel temps je vais attendre au maximum.
Pour tenter de répondre à cela, je vais relever des données, à savoir, je note pour chacun des matins le temps que j'ai attendu.
Je dispose ainsi d'observations $t_1, \ldots, t_n$ des variables aléatoires $T_1, \ldots, T_n$ qui sont indépendantes et de même loi.
Je sais que T_i suit une loi uniforme sur $[0, a]$.
a est donc le temps d'attente maximum du bus et constitue donc le paramètre inconnu que je cherche à estimer.

$\Rightarrow$ Comment?

Estimation sur un exemple (2)

Méthode 1 :

- a n'est pas le paramètre d'espérance

Estimation sur un exemple (2)

Méthode 1 :

- a n'est pas le paramètre d'espérance
- $\mathbb{E}(T_i) = \frac{a}{2}$

Estimation sur un exemple (2)

Méthode 1 :

- a n'est pas le paramètre d'espérance
- $\mathbb{E}(T_i) = \frac{a}{2}$
- $\frac{T_1 + \ldots + T_n}{n}$ est un estimateur de $\frac{a}{2}$

Estimation sur un exemple (2)

Méthode 1 :

- a n'est pas le paramètre d'espérance
- $\mathbb{E}(T_i) = \frac{a}{2}$
- $\frac{T_1+\ldots+T_n}{n}$ est un estimateur de $\frac{a}{2}$
- un estimateur de a est donc $\hat{A}_{n,1} = 2.\frac{T_1+\ldots+T_n}{n}$

Estimation sur un exemple (3)

Méthode 2 :

- a représente le temps d'attente maximum

Estimation sur un exemple (3)

Méthode 2 :

- a représente le temps d'attente maximum
- on dispose de $T_1, \dots, T_n$

Estimation sur un exemple (3)

Méthode 2 :

- a représente le temps d'attente maximum
- on dispose de $T_1, \ldots, T_n$
- $\hat{A}_{n,2} = max(T_1, \ldots, T_n)$

Comparaison d'estimateur

Question : Comment dire si $\hat{A}_{n,1}$ est meilleur ou non que $\hat{A}_{n,2}$?

Comparaison d'estimateur

Question : Comment dire si $\hat{A}_{n,1}$ est meilleur ou non que $\hat{A}_{n,2}$?

Réponse : On calcule l'erreur quadratique de $\hat{A}_{n,1}$ et de $\hat{A}_{n,2}$ et l'estimateur qui a la plus petite erreur quadratique est déclaré être le meilleur.

Comparaison d'estimateur

Question : Comment dire si $\hat{A}_{n,1}$ est meilleur ou non que $\hat{A}_{n,2}$?

Réponse : On calcule l'erreur quadratique de $\hat{A}_{n,1}$ et de $\hat{A}_{n,1}$ et l'estimateur qui a la plus petite erreur quadratique est déclaré être le meilleur.

Question : Qu'est ce que l'erreur quadratique d'un estimateur?

Comparaison d'estimateur

Question : Comment dire si $\hat{A}_{n,1}$ est meilleur ou non que $\hat{A}_{n,2}$?

Réponse : On calcule l'erreur quadratique de $\hat{A}_{n,1}$ et de $\hat{A}_{n,1}$ et l'estimateur qui a la plus petite erreur quadratique est déclaré être le meilleur.

Question : Qu'est ce que l'erreur quadratique d'un estimateur?

Réponse : L'erreur quadratique d'un estimateur $\hat{A}_n$ d'un paramètre a est défini par :

$$EQ(\hat{A}_n) = \mathbb{E}\left[(\hat{A}_n - a)^2\right]$$

Comparaison d'estimateur (2)

Propriété de l'erreur quadratique :

$$EQ(\hat{A}_n) = V[\hat{A}_n] + (b(\hat{A}_n))^2$$

où :

$b(\hat{A}_n) = \mathbb{E}[\hat{A}_n] - a$: biais de $\hat{A}_n$

Comparaison d'estimateur (3)

Retour sur notre exemple du temps d'attente du bus.

Comparaison d'estimateur (3)

Retour sur notre exemple du temps d'attente du bus.

Calculons $EQ(\hat{A}_{n,1})$ et $EQ(\hat{A}_{n,2})$.

Comparaison d'estimateur (3)

$$\mathbb{E}[\hat{A}_{n,1}] \quad = \quad \mathbb{E}\left[2.\frac{T_1 + \ldots + T_n}{n}\right]$$

Comparaison d'estimateur (3)

$$\begin{aligned}\mathbb{E}[\hat{A}_{n,1}] &= \mathbb{E}\left[2.\frac{T_1+\ldots+T_n}{n}\right] \\ &= 2.\mathbb{E}\left[\frac{T_1+\ldots+T_n}{n}\right] \text{ linéarité de l'espérance}\end{aligned}$$

Comparaison d'estimateur (3)

$$\begin{aligned}
\mathbb{E}[\hat{A}_{n,1}] &= \mathbb{E}\left[2.\frac{T_1 + \ldots + T_n}{n}\right] \\
&= 2.\mathbb{E}\left[\frac{T_1 + \ldots + T_n}{n}\right] \\
&= 2.\frac{\mathbb{E}[T_1] + \ldots + \mathbb{E}[T_n]}{n}
\end{aligned}$$

Comparaison d'estimateur (3)

$$
\begin{aligned}
\mathbb{E}[\hat{A}_{n,1}] &= \mathbb{E}\left[2.\frac{T_1 + \ldots + T_n}{n}\right] \\
&= 2.\mathbb{E}\left[\frac{T_1 + \ldots + T_n}{n}\right] \\
&= 2.\frac{\mathbb{E}[T_1] + \ldots + \mathbb{E}[T_n]}{n} \\
&= 2.\frac{\mathbb{E}[T_1] + \ldots + \mathbb{E}[T_1]}{n} \text{ car de même loi}
\end{aligned}
$$

Comparaison d'estimateur (3)

$$
\begin{aligned}
\mathbb{E}[\hat{A}_{n,1}] &= \mathbb{E}\left[2.\frac{T_1 + \ldots + T_n}{n}\right] \\
&= 2.\mathbb{E}\left[\frac{T_1 + \ldots + T_n}{n}\right] \\
&= 2.\frac{\mathbb{E}[T_1] + \ldots + \mathbb{E}[T_n]}{n} \\
&= 2.\frac{n.\mathbb{E}[T_1]}{n}
\end{aligned}
$$

Comparaison d'estimateur (3)

$$\begin{aligned}
\mathbb{E}[\hat{A}_{n,1}] &= \mathbb{E}\left[2.\frac{T_1 + \ldots + T_n}{n}\right] \\
&= 2.\mathbb{E}\left[\frac{T_1 + \ldots + T_n}{n}\right] \\
&= 2.\frac{\mathbb{E}[T_1] + \ldots + \mathbb{E}[T_n]}{n} \\
&= 2\mathbb{E}[T_1]
\end{aligned}$$

Comparaison d'estimateur (3)

$$\begin{aligned}
\mathbb{E}[\hat{A}_{n,1}] &= \mathbb{E}\left[2.\frac{T_1 + \ldots + T_n}{n}\right] \\
&= 2.\mathbb{E}\left[\frac{T_1 + \ldots + T_n}{n}\right] \\
&= 2.\frac{\mathbb{E}[T_1] + \ldots + \mathbb{E}[T_n]}{n} \\
&= 2\mathbb{E}[T_1] \\
&= a
\end{aligned}$$

Comparaison d'estimateur (3)

$$V[\hat{A}_{n,1}] = V\left[2.\frac{T_1 + \ldots + T_n}{n}\right]$$

Comparaison d'estimateur (3)

$$
\begin{aligned}
V[\hat{A}_{n,1}] &= V\left[2.\frac{T_1 + \ldots + T_n}{n}\right] \\
&= \frac{4}{n^2}.V[T_1 + \ldots + T_n]
\end{aligned}
$$

Comparaison d'estimateur (3)

$$
\begin{aligned}
V[\hat{A}_{n,1}] &= V\left[2.\frac{T_1 + \ldots + T_n}{n}\right] \\
&= \frac{4}{n^2}.V[T_1 + \ldots + T_n] \\
&= \frac{4}{n^2}.(V[T_1] + \ldots + V[T_n]) \text{ car indépendance}
\end{aligned}
$$

Comparaison d'estimateur (3)

$$
\begin{aligned}
V[\hat{A}_{n,1}] &= V\left[2.\frac{T_1 + \ldots + T_n}{n}\right] \\
&= \frac{4}{n^2}.V[T_1 + \ldots + T_n] \\
&= \frac{4}{n^2}.(V[T_1] + \ldots + V[T_n]) \\
&= \frac{4}{n^2}.(V[T_1] + \ldots + V[T_1]) \text{ car de même loi}
\end{aligned}
$$

Comparaison d'estimateur (3)

$$
\begin{aligned}
V[\hat{A}_{n,1}] &= V\left[2.\frac{T_1 + \ldots + T_n}{n}\right] \\
&= \frac{4}{n^2}.V[T_1 + \ldots + T_n] \\
&= \frac{4}{n^2}.(V[T_1] + \ldots + V[T_n]) \\
&= \frac{4}{n^2}.(V[T_1] + \ldots + V[T_1]) \\
&= \frac{4}{n^2}.n.V[T_1]
\end{aligned}
$$

Comparaison d'estimateur (3)

$$
\begin{aligned}
V[\hat{A}_{n,1}] &= V\left[2.\frac{T_1 + \ldots + T_n}{n}\right] \\
&= \frac{4}{n^2}.V[T_1 + \ldots + T_n] \\
&= \frac{4}{n^2}.(V[T_1] + \ldots + V[T_n]) \\
&= \frac{4}{n^2}.(V[T_1] + \ldots + V[T_1]) \\
&= \frac{4}{n^2}.n.V[T_1] \\
&= \frac{4}{n}.\frac{a^2}{12}
\end{aligned}
$$

Comparaison d'estimateur (3)

$$
\begin{aligned}
V[\hat{A}_{n,1}] &= V\left[2.\frac{T_1+\ldots+T_n}{n}\right] \\
&= \frac{4}{n^2}.V[T_1+\ldots+T_n] \\
&= \frac{4}{n^2}.(V[T_1]+\ldots+V[T_n]) \\
&= \frac{4}{n^2}.(V[T_1]+\ldots+V[T_1]) \\
&= \frac{4}{n^2}.n.V[T_1] \\
&= \frac{a^2}{3.n}
\end{aligned}
$$

$$\Rightarrow EQ(\hat{A}_{n,1}) = \frac{a^2}{3.n}$$

Comparaison d'estimateur (3)

Pour calculer $\mathbb{E}[\hat{A}_{n,2}]$ et $V[\hat{A}_{n,2})]$, nous devons commencer par déterminer la loi de $\hat{A}_{n,2}$.

Comparaison d'estimateur (3)

Pour calculer $\mathbb{E}[\hat{A}_{n,2}]$ et $V[\hat{A}_{n,2}]$, nous devons commencer par déterminer la loi de $\hat{A}_{n,2}$.

Pour déterminer cette loi, nous allons regarder la fonction de répartition.

Comparaison d'estimateur (3)

Pour calculer $\mathbb{E}[\hat{A}_{n,2}]$ et $V[\hat{A}_{n,2}]$, nous devons commencer par déterminer la loi de $\hat{A}_{n,2}$.

Pour déterminer cette loi, nous allons regarder la fonction de répartition.

Pour rappel, la fonction de répartition se définit par :

$$\text{pour tout t réel,}\quad F(t) = P(\hat{A}_{n,2} \leq t)$$

Comparaison d'estimateur (3)

$$
\begin{aligned}
F(t) &= P(\hat{A}_{n,2} \leq t) \\
&= P(max(T_1, \ldots, T_n) \leq t)
\end{aligned}
$$

Comparaison d'estimateur (3)

$$\begin{aligned} F(t) &= P(\hat{A}_{n,2} \leq t) \\ &= P(max(T_1, \dots, T_n) \leq t) \\ &= P(T_1 \leq t, \dots, T_n \leq t) \end{aligned}$$

Comparaison d'estimateur (3)

$$
\begin{aligned}
F(t) &= P(\hat{A}_{n,2} \leq t) \\
&= P(max(T_1, \dots, T_n) \leq t) \\
&= P(T_1 \leq t, \dots, T_n \leq t) \\
&= P(T_1 \leq t) \dots . P(T_n \leq t) \text{ car indépendance}
\end{aligned}
$$

Comparaison d'estimateur (3)

$$\begin{aligned}
F(t) &= P(\hat{A}_{n,2} \leq t) \\
&= P(max(T_1, \ldots, T_n) \leq t) \\
&= P(T_1 \leq t, \ldots, T_n \leq t) \\
&= P(T_1 \leq t). \ldots .P(T_n \leq t) \\
&= P(T_1 \leq t). \ldots .P(T_1 \leq t) \text{ car de même loi} \\
&= (P(T_1 \leq t))^n
\end{aligned}$$

Comparaison d'estimateur (3)

Or la fonction de répartition pour une loi uniforme sur $[0; a]$ est :

$$P(T_1 \leq t) = \begin{cases} 0 & \text{si } t \leq 0 \\ \frac{t}{a} & \text{si } 0 \leq t \leq a \\ 1 & \text{si } t \geq a \end{cases}$$

Comparaison d'estimateur (3)

Donc :

$$F(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ \frac{t^n}{a^n} & \text{si } 0 \leq t \leq a \\ 1 & \text{si } t \geq a \end{cases}$$

Comparaison d'estimateur (3)

D'où : La fonction de densité de $\hat{A}_{n,2}$ est :

$$f(t) = \begin{cases} \frac{n.t^{n-1}}{a^n} & \text{si } 0 \leq t \leq a \\ 0 & \text{sinon} \end{cases}$$

Comparaison d'estimateur (3)

On en déduit :

$$\begin{aligned}
\mathbb{E}[\hat{A}_{n,2}] &= \int_{-\infty}^{+\infty} t.f(t).dt \\
&= \int_0^a t.\frac{n.t^{n-1}}{a^n}.dt \\
&= \int_0^a .\frac{n.t^n}{a^n}.dt \\
&= \frac{1}{a^n}[\frac{n}{n+1}.t^{n+1}]_0^a \\
&= \frac{n}{n+1}a \\
&= a - \frac{1}{n+1}a
\end{aligned}$$

Comparaison d'estimateur (3)

Par ailleurs :

$$V[\hat{A}_{n,2}] = \mathbb{E}[\hat{A}_{n,2}^2] - (\mathbb{E}[\hat{A}_{n,2}])^2$$

$$\begin{aligned}
\mathbb{E}[\hat{A}_{n,2}^2] &= \int_{-\infty}^{+\infty} t^2.f(t).dt \\
&= \int_0^a t^2.\frac{n.t^{n-1}}{a^n}.dt \\
&= \int_0^a .\frac{n.t^{n+1}}{a^n}.dt \\
&= \frac{1}{a^n}[\frac{n}{n+2}.t^{n+2}]_0^a \\
&= \frac{n}{n+2}.a^2
\end{aligned}$$

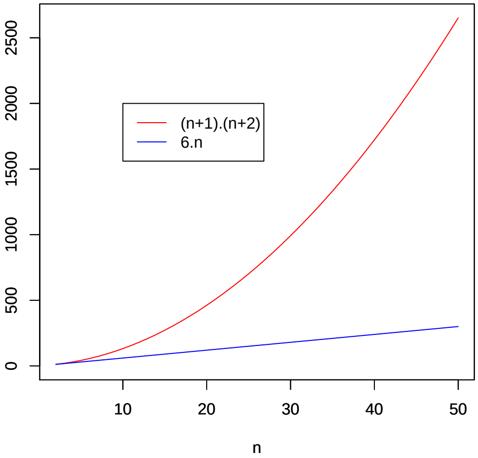
Comparaison d'estimateur (3)

Au final :

$$\begin{aligned} EQ(\hat{A}_{n,2}) &= \frac{n}{n+2}a^2 - \left(\frac{n}{n+1}a\right)^2 + \left(\frac{1}{n+1}a\right)^2 \\ &= \frac{2}{(n+2).(n+1)}a^2 \end{aligned}$$

On doit donc comparer $\frac{2}{(n+2).(n+1)}.a^2$ à $\frac{2}{6n}.a^2$.

Comparaison d'estimateur (3)



Comparaison d'estimateur (3)

Donc $\frac{2}{(n+2).(n+1)}.a^2$ est plus petit que $\frac{2}{6n}.a^2$

$$\Rightarrow \hat{A}_{n,2} \text{ est meilleur que } \hat{A}_{n,1}.$$

Comparaison d'estimateur : illustration

Données	Jeu 1	Jeu 2	Jeu 3	Jeu 4	Jeu 5	Jeu 6
$\hat{A}_{n,1}$	9.30	9.20	9.72	9.43	11.12	10.41
$\hat{A}_{n,2}$	9.43	9.83	9.88	9.84	9.99	9.84

Données	Jeu 7	Jeu 8	Jeu 9	Jeu 10
$\hat{A}_{n,1}$	10.23	10.45	10.13	10.07
$\hat{A}_{n,2}$	9.98	9.98	9.99	9.76

estimateur	$\hat{A}_{n,1}$	$\hat{A}_{n,2}$
variance	0.36	0.03

Estimateur de la variance

Qu'en est-il pour la variance?

Estimateur de la variance

Qu'en est-il pour la variance?

On sait que :

$$V[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Estimateur de la variance

Qu'en est-il pour la variance?

On sait que :

$$V[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Donc, on peut estimer $V[X]$ par :

$$\hat{V}_{n,1} = \frac{1}{n}\left((X_1 - \bar{X}_n)^2 + \ldots + (X_n - \bar{X}_n)^2\right)$$

où $\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$.

Estimateur de la variance (2)

Il existe un second estimateur de la variance qui est défini par :

$$\hat{V}_{n,2} = \frac{1}{n-1}\left((X_1 - \bar{X}_n)^2 + \ldots + (X_n - \bar{X}_n)^2\right)$$

soit

$$\hat{V}_{n,2} = \frac{n}{n-1}\hat{V}_{n,1}$$

Estimateur de la variance : illustration

Données	Jeu 1	Jeu 2	Jeu 3	Jeu 4	Jeu 5	Jeu 6
$\hat{V}_{n,1}$	9.38	8.61	8.05	9.62	6.75	8.2
$\hat{V}_{n,2}$	9.57	8.79	8.22	9.82	6.89	8.36
Données	Jeu 7	Jeu 8	Jeu 9	Jeu 10		
$\hat{V}_{n,1}$	7.26	7.97	6.45	9.2		
$\hat{V}_{n,2}$	7.41	8.13	6.58	9.4		

estimateur	$\hat{V}_{n,1}$	$\hat{V}_{n,2}$
moyenne	8.15	8.32

Estimateur : conclusion

- Un estimateur est une variable aléatoire, et donc possède une loi de probabilité
- on veut que quand n est grand il devienne précis : sa densité doit être très piquée autour de la valeur du paramètre que l'on cherche à estimer
  - l'espérance de l'estimateur doit être proche de la quantité cherchée
  - la variance de l'estimateur doit être aussi faible que possible

Introduction

Ce que l'on a pû voir, avec les 10 jeux de données, c'est que même lorsque l'on simule des observations qui suivent la même loi, les estimations changent d'un jeu de données à un autre.

$\Rightarrow$ fluctuation d'échantillonnage

Introduction

Ce que l'on a pû voir, avec les 10 jeux de données, c'est que même lorsque l'on simule des observations qui suivent la même loi, les estimations changent d'un jeu de données à un autre.

$\Rightarrow$ fluctuation d'échantillonnage

Comment quantifier cela?

Introduction

Ce que l'on a pû voir, avec les 10 jeux de données, c'est que même lorsque l'on simule des observations qui suivent la même loi, les estimations changent d'un jeu de données à un autre.

$\Rightarrow$ fluctuation d'échantillonnage

Comment quantifier cela?

En construisant un intervalle de confiance!