Plan du chapitre 4

Motivation

Vous soupçonnez qu'on sert à la Jefke des bières de moins de 25 cl en moyenne. Scandalisé, vous décidez de vérifier si c'est effectivement le cas (avant de contacter, le cas échéant, les autorités).

Comment vérifier ?

Ce problème est un problème de décision statistique. Il relève donc de la 2de partie du cours, mais on l'utilisera ici pour motiver les résultats de cette section.

Notons d'abord qu'on peut considérer la capacité (en cl) d'une bière servie à la Jefke comme une v.a. X (de nombreux paramètres physiques influençant la quantité de mousse, donc celle de bière...)

En langage probabiliste, vous voulez déterminer si $\mathrm{E}[X] < 25$(cl) ou pas.

Motivation

Puisque X est typiquement continue, la question considérée est

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx \overset{?}{<} 25,$$

où f est la fonction de densité de X. Mais f est inconnue...

L'idée naturelle consiste alors à considérer un échantillon $(X_1, \ldots, X_n)$, associé à n bières servies. On dira qu'il s'agit d'un échantillon aléatoire simple si ces v.a. sont indépendantes et identiquement distribuées (i.i.d.), autrement dit, si ces v.a.

- sont mutuellement indépendantes, et
- partagent toutes la même distribution (elles admettent la même densité f).

L'échantillon observé sera désigné par $(x_1, \ldots, x_n)$

Toutes les observations étant de densité f, elles portent toutes de l'information sur f, et donc sur $\mathrm{E}[X]$. Comment extraire cette information ?

Il est naturel de considérer la moyenne empirique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ et de fonder la décision sur la valeur $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ qu'elle prend sur l'échantillon observé $(x_1, \ldots, x_n)$.

Des questions naturelles dans ce cadre sont :

- En quel sens $\bar{X}$ fournit-elle une information sur $\mathrm{E}[X]$?
- Comment tenir compte de la variabilité intrinsèque de $\bar{X}$ pour se convaincre raisonnablement que $\mathrm{E}[X] < 25$ (si c'est possible) ?

La seconde question est justifiée par le fait que $\bar{X}$ étant une fonction des v.a. $X_1, \ldots, X_n$, elle est elle-même une v.a., avec sa propre distribution : on parlera de distribution échantillonnée.

La loi des grands nombres

Soit X une v.a. de moyenne $\mu = \mathrm{E}[X]$ et de variance $\sigma^2 = \mathrm{Var}[X] < \infty$.
Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire simple associé.

On a toujours

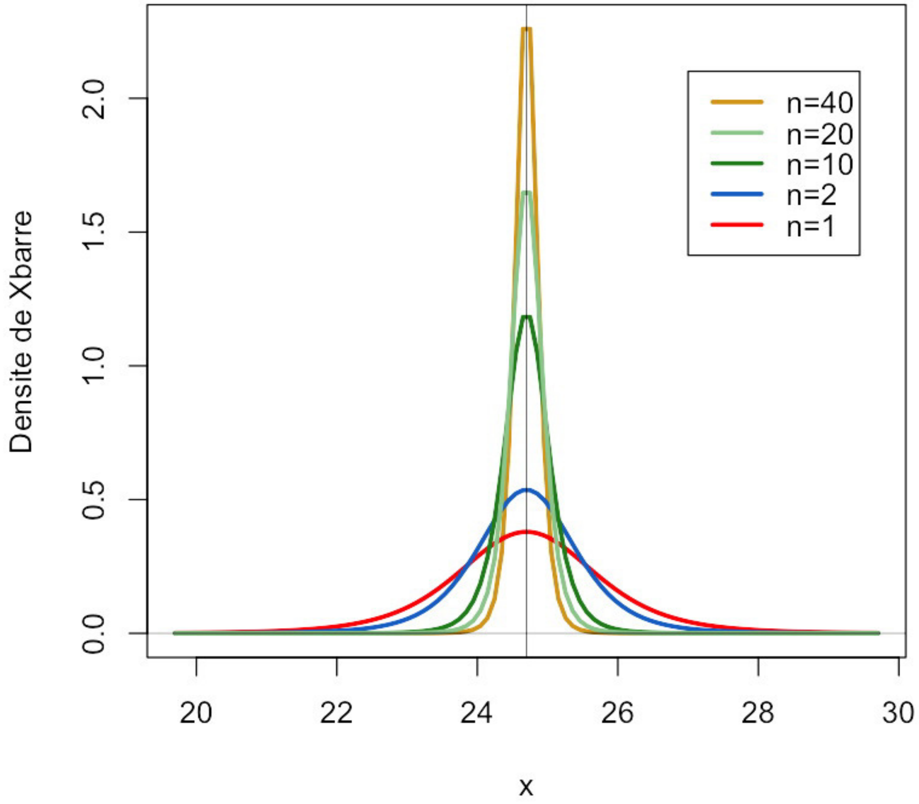
$$(i)\quad \mathrm{E}[\bar{X}^{(n)}] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\mathrm{E}[X_i]}_{=\mu} = \frac{1}{n}(n\mu) = \mu$$

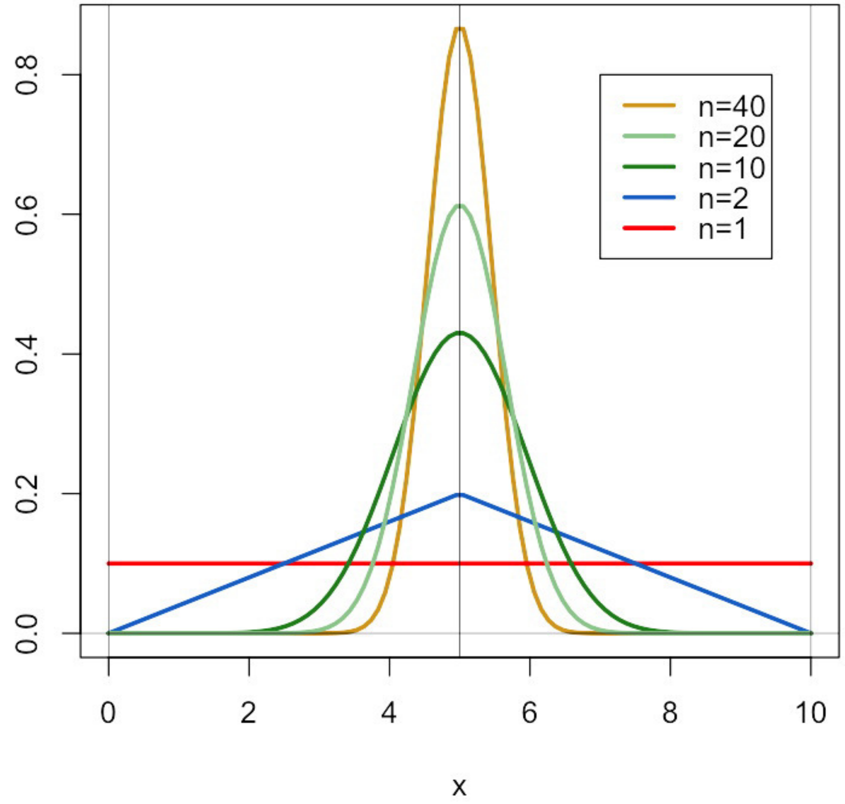
et

$$(ii)\quad \mathrm{Var}[\bar{X}^{(n)}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{\mathrm{Var}[X_i]}_{=\sigma^2} = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n},$$

où on a utilisé l'indépendance mutuelle des X_i.

Donc la distribution de $\bar{X}^{(n)}$ (i) reste de moyenne μ pour tout n
(ii) devient de plus en plus concentrée quand n grandit.





Graphes de la densité de $\bar{X}^{(n)}$ pour différents n, dans le cas où $X_1, \ldots, X_n$ sont i.i.d. de loi $24.7 + t_5$ (gauche) et de loi $\mathrm{Unif}(0, 10)$ (droite).

$\mathrm{E}[\bar{X}^{(n)}]$ reste en μ ($= 24.7$ à gauche et $= 5$ à droite) et $\mathrm{Var}[\bar{X}^{(n)}] \searrow 0$ quand $n \to \infty$

La loi des grands nombres

Ceci suggère le résultat suivant (qui ne demande en fait que des moments d'ordre 1).

Théorème (Loi faible des grands nombres)

Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire simple et posons $\bar{X}^{(n)} = \frac{1}{n}\sum_{i=1}^n X_i$.
Supposons que $\sigma^2 = \mathrm{Var}[X_i] < \infty$ et notons $\mu = \mathrm{E}[X_i]$. Alors

$$\forall \varepsilon > 0, \quad P[|\bar{X}^{(n)} - \mu| > \varepsilon] \to 0$$

quand $n \to \infty$.

Interprétation : même si on prend $\varepsilon > 0$ minuscule, alors

$$\begin{aligned} P\big[\mu - \varepsilon \leq \bar{X}^{(n)} \leq \mu + \varepsilon\big] &= P\big[|\bar{X}^{(n)} - \mu| \leq \varepsilon\big] \\ &= 1 - \underbrace{P\big[|\bar{X}^{(n)} - \mu| > \varepsilon\big]}_{\to 0} \\ &\to 1 \end{aligned}$$

quand $n \to \infty$!

Preuve : fixons $\varepsilon > 0$. Pour tout $a > 0$, l'inégalité de Tchebychev fournit

$$0 \leq P\left[|\bar{X}^{(n)} - \mu| > a\sqrt{\frac{\sigma^2}{n}}\right] = P\left[\,|\bar{X}^{(n)} - \mathrm{E}[\bar{X}^{(n)}]| > a\sqrt{\mathrm{Var}[\bar{X}^{(n)}]}\,\right] \leq \frac{1}{a^2}.$$

Pour $a = \varepsilon\sqrt{\frac{n}{\sigma^2}}$, ceci donne

$$0 \leq P[|\bar{X}^{(n)} - \mu| > \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2 n}.$$

Le théorème de l'étau garantit donc que

$$P[|\bar{X}^{(n)} - \mu| > \varepsilon] \to 0$$

quand $n \to \infty$. $\square$

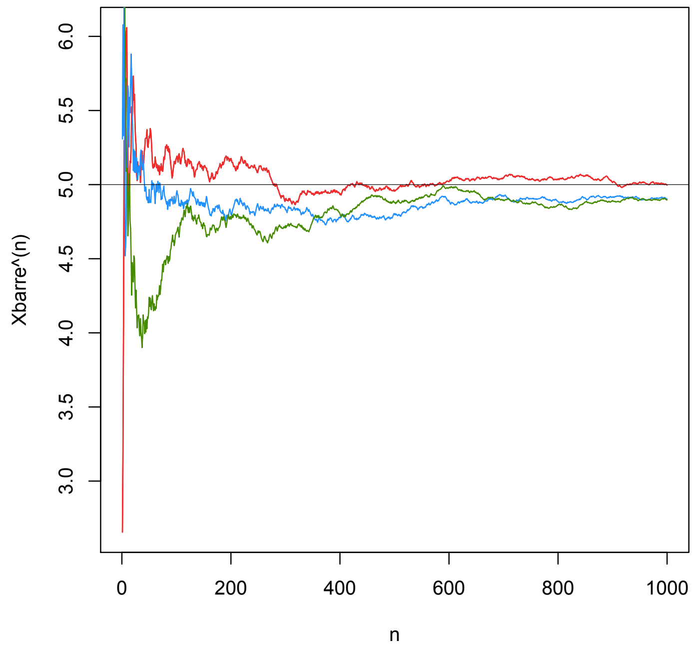
Définition

Soient $(X^{(n)})$ *une suite de v.a. et* X *une autre v.a.*
$X^{(n)} \to X$ *en probabilité (not. :* $X^{(n)} \xrightarrow{P} X$*)* $\Leftrightarrow \forall \varepsilon > 0, P[|X^{(n)} - X| > \varepsilon] \to 0$
$X^{(n)} \to X$ *presque sûrement (not. :* $X^{(n)} \xrightarrow{p.s.} X$*)* $\Leftrightarrow P[\{\omega \in \Omega : X^{(n)}(\omega) \to X(\omega)\}] = 1$

On a $\bar{X}^{(n)} \xrightarrow{P} \mu$ (c'est la loi faible des grands nombres ci-dessus)
En fait, on a aussi $\bar{X}^{(n)} \xrightarrow{p.s.} \mu$ (c'est la loi "forte" des grands nombres).

La convergence $\xrightarrow{p.s.}$ signifie
qu'il est certain (avec probabilité 1)
que la convergence sera observée

A droite :
$X_1, \dots, X_n$ i.i.d. de loi $\mathrm{Unif}(0, 10)$
$\bar{X}^{(n)} \xrightarrow{p.s.} \mu = \mathrm{E}[X_1] = 5$



La loi des grands nombres

On a vu (ch.3-p.93) que $\mathrm{E}[(U-V)^2]$ était une distance entre les v.a. U et V.
Ceci mène au concept de convergence stochastique suivant.

Définition

$X^{(n)} \to X$ en L_2 (not. : $X^{(n)} \xrightarrow{L_2} X$) $\Leftrightarrow \mathrm{E}[(X^{(n)}-X)^2] \to 0$

On parle aussi de convergence en moyenne quadratique.

Puisque

$$\begin{aligned}
\mathrm{E}[(\bar{X}^{(n)}-\mu)^2] &= \mathrm{E}[(\bar{X}^{(n)}-\mathrm{E}[\bar{X}^{(n)}])^2] \\
&= \mathrm{Var}[\bar{X}^{(n)}] \\
&= \frac{\sigma^2}{n} \\
&\to 0,
\end{aligned}$$

on a que $\bar{X}^{(n)} \xrightarrow{L_2} \mu$.

Il existe des liens entre les divers concepts de convergence.

Théorème

(i) $X^{(n)} \xrightarrow{p.s.} X \Rightarrow X^{(n)} \xrightarrow{P} X$

(ii) $X^{(n)} \xrightarrow{L_2} X \Rightarrow X^{(n)} \xrightarrow{P} X$

Les implications dans l'autre sens ne sont pas vraies en général
(nous renvoyons au livre de référence pour des contre-exemples).

Le point (i) explique qu'on parle de loi forte et de loi faible des grands nombres.

On peut appliquer la loi forte des grands nombres à toute suite de variables aléatoires i.i.d. admettant des moments d'ordre 1. Deux exemples...

---

1) Si les X_i sont i.i.d. et admettent des moments d'ordre 2, alors $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \overset{p.s.}{\to} \mathrm{E}[X^2]$.

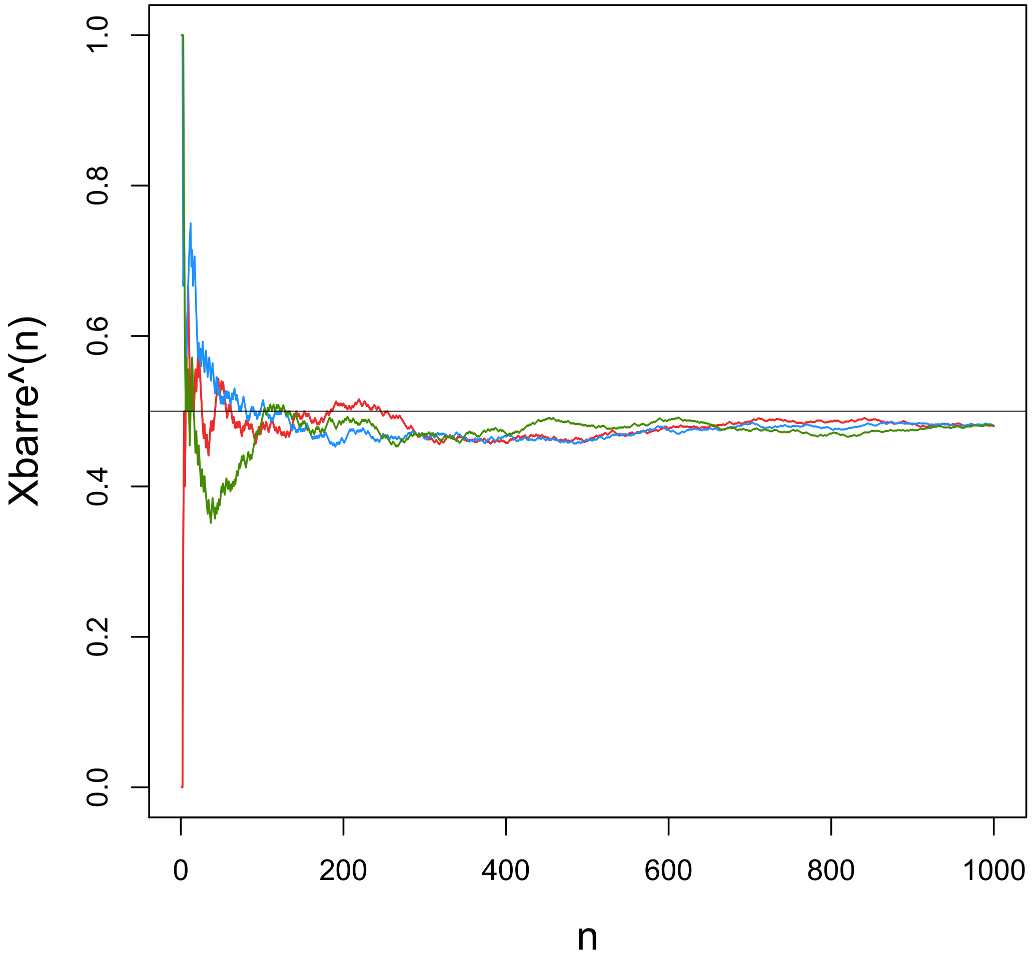
2) Soient E une expérience aléatoire et $(\Omega, \mathcal{A}, P)$ un espace probabilisé associé. Soit X la v.a. qui vaut 1 si un événement fixé $A(\in \mathcal{A})$ se produit et 0 sinon. Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire simple correspondant (// n répétitions de E).

Alors la loi forte des grands nombres affirme que

$$\frac{\text{nombre de réalisations de } A \text{ parmi les } n \text{ répétitions}}{n} = \frac{X_1 + \ldots + X_n}{n} = \bar{X}^{(n)} \overset{p.s.}{\to} \mu = P[A]$$

La définition fréquentielle des probabilités découle de la définition axiomatique !

La loi des grands nombres



E = lancer une pièce de monnaire équilibrée et A = "obtenir face"

Le résultat suivant est souvent utile.

Théorème

Soient $(X^{(n)})$, $(Y^{(n)})$ *deux suites de v.a. et* X, Y *deux autres v.a. Alors*
(i) $X^{(n)} \xrightarrow{p.s.} X$ *et* $Y^{(n)} \xrightarrow{p.s.} Y \Rightarrow X^{(n)} + Y^{(n)} \xrightarrow{p.s.} X + Y$ *et* $X^{(n)}Y^{(n)} \xrightarrow{p.s.} XY$
(ii) $X^{(n)} \xrightarrow{P} X$ *et* $Y^{(n)} \xrightarrow{P} Y \Rightarrow X^{(n)} + Y^{(n)} \xrightarrow{P} X + Y$ *et* $X^{(n)}Y^{(n)} \xrightarrow{P} XY$
(iii) $X^{(n)} \xrightarrow{L_2} X$ *et* $Y^{(n)} \xrightarrow{L_2} Y \Rightarrow X^{(n)} + Y^{(n)} \xrightarrow{L_2} X + Y$

Soit X une v.a. de moyenne $\mu = \mathrm{E}[X]$ et de variance $\sigma^2 = \mathrm{Var}[X] < \infty$.
Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire simple associé.
On peut estimer σ^2 par la variance empirique

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

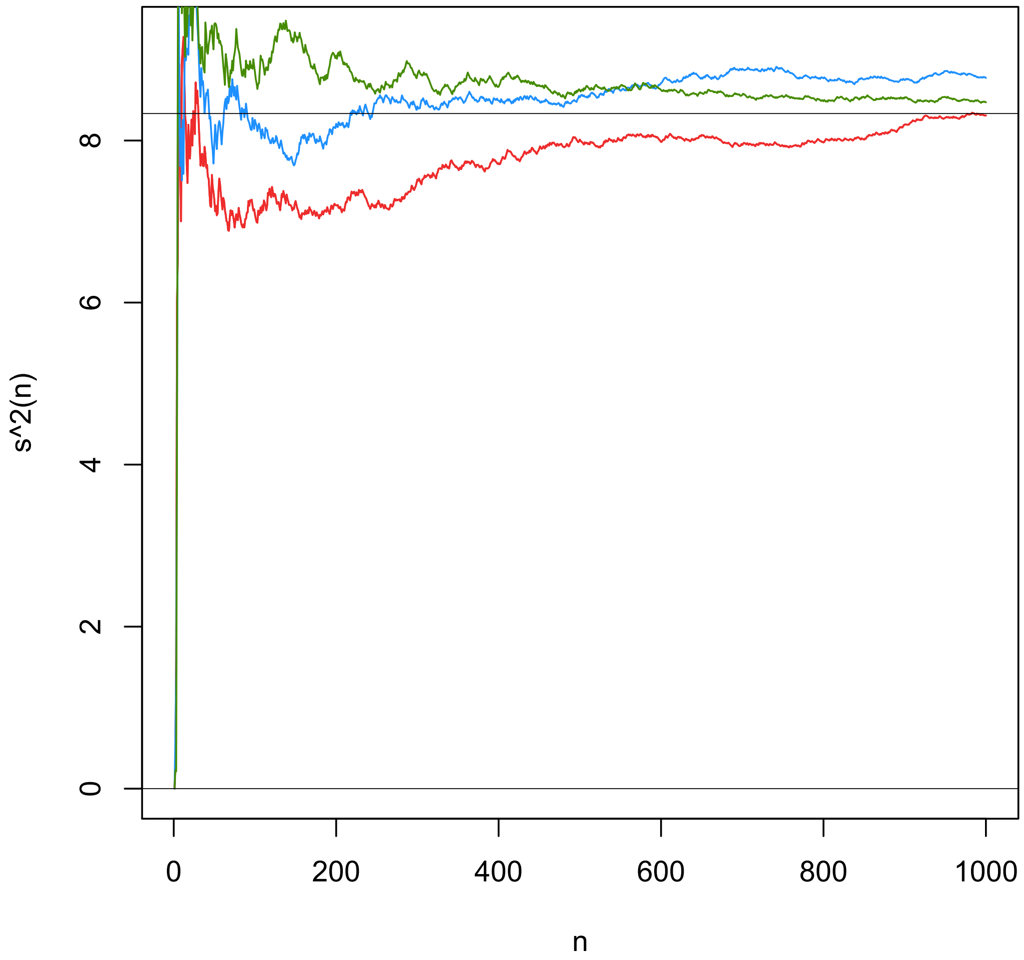
Mais la loi des grands nombres ne permet pas d'affirmer que $s^2 \xrightarrow{p.s.} \sigma^2$ (pourquoi ?)

Mais puisque

$$\begin{aligned} s^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 && \left(// \ \sigma^2 = \mathrm{E}[(X-\mu)^2]\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i^2 + \bar{X}^2 - 2X_i\bar{X}) \\ &= \left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right) + \bar{X}^2 - 2\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)\bar{X} \\ &= \left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right) - \bar{X}^2 && \left(// \ \sigma^2 = \mathrm{E}[X^2] - \mu^2\right), \end{aligned}$$

le théorème précédent fournit

$$s^2 = \left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right) - \bar{X}\times\bar{X} \overset{p.s.}{\to} \mathrm{E}[X^2] - \mu\times\mu = \sigma^2$$



Pour $X_1, \ldots, X_n$ i.i.d. de loi $\mathrm{Unif}(0, 10)$, on a $s^2 \overset{p.s.}{\to} \sigma^2 = \frac{(10-0)^2}{12} \approx 8.33$

Il est naturel d'estimer l'écart-type $\sigma = \sqrt{\sigma^2}$ par

$$s = \sqrt{s^2}.$$

On aura alors

$$s = \sqrt{s^2} \stackrel{p.s.}{\to} \sqrt{\sigma^2} = \sigma,$$

ce qui est une conséquence du théorème suivant.

Théorème

Soient $(X^{(n)})$ une suite de v.a. et X une autre v.a.. Soit $g : C \to \mathbb{R}$ continue, où C contient l'ensemble des valeurs possibles de X. Alors

(i) $X^{(n)} \stackrel{p.s.}{\to} X \Rightarrow g(X^{(n)}) \stackrel{p.s.}{\to} g(X)$

(ii) $X^{(n)} \stackrel{P}{\to} X \Rightarrow g(X^{(n)}) \stackrel{P}{\to} g(X)$

Plan du chapitre 4

Le théorème central-limite

Soit X une v.a. de moyenne $\mu = \mathrm{E}[X]$ et de variance $\sigma^2 = \mathrm{Var}[X] < \infty$.
Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire simple associé. On a vu que

(i) $\mathrm{E}[\bar{X}^{(n)}] = \mu$

(ii) $\mathrm{Var}[\bar{X}^{(n)}] = \sigma^2/n$

(iii) Si $n \to \infty$, alors $\bar{X}^{(n)} \to \mu$ (p.s., en probabilité, en L_2)

Ces résultats livrent un moyen d'estimer μ sur la base de $X_1, \ldots, X_n$ (par $\bar{X}^{(n)}$), d'une manière telle que si $n \to \infty$, l'estimation se fait finalement sans erreur.

$\rightsquigarrow$ Si on sert une infinité de bières, alors on pourra décider sans se tromper si $\mu = \mathrm{E}[X] < 25$ cl ou pas...

Bien entendu, on ne prendra jamais qu'un échantillon de taille n fixée.
Quel que soit n, une erreur sera commise dans l'estimation de μ par $\bar{X}^{(n)}$.
Il est capital de pouvoir quantifier cette erreur.

Le théorème central-limite

Pour quantifier l'erreur, on doit typiquement calculer

$$P[|\bar{X}^{(n)} - \mu| > \varepsilon],$$

ce qui requiert de connaître la distribution de $\bar{X}^{(n)}$.

Mais celle-ci dépend de la distribution des X_i, qui est le plus souvent inconnue :

- $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2) \Rightarrow \bar{X} = (X_1 + X_2 + \ldots + X_n)/n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
- $X_1, \ldots, X_n$ i.i.d. $\mathrm{Bern}(p) = \mathrm{Bin}(1, p) \Rightarrow \bar{X} \sim \frac{1}{n}\mathrm{Bin}(n, p)$
- $X_1, \ldots, X_n$ i.i.d. $\mathrm{Poi}(\lambda) \Rightarrow \bar{X} \sim \frac{1}{n}\mathrm{Poi}(n\lambda)$
- $\ldots$

(voir les propriétés d'additivité en ch.3-p.75).

On est sauvé par LE théorème le plus important en probabilité et statistique...

Le théorème central-limite

Théorème ("Théorème central-limite" ou "TCL")

Soit $(X_1, \dots, X_n)$ un échantillon aléatoire simple. Supposons que $\sigma^2 = \mathrm{Var}[X_i] < \infty$. Posons $\mu = \mathrm{E}[X_i]$ et $\bar{X}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Alors, pour tout x,

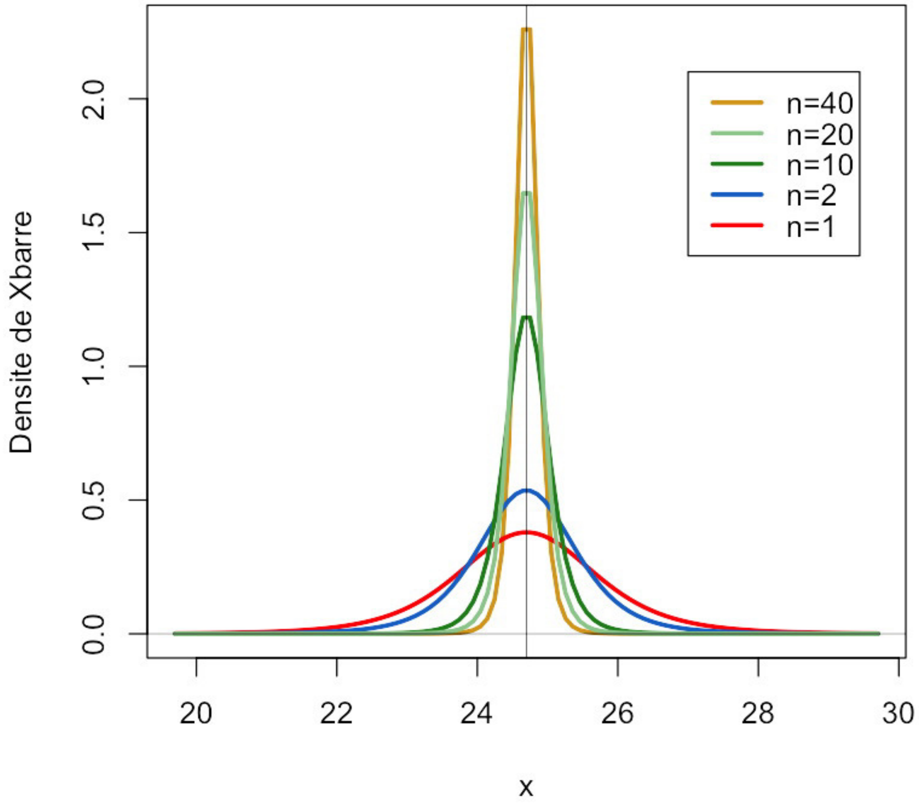
$$P\left[\frac{\bar{X}^{(n)} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq x\right] \to \Phi(x),$$

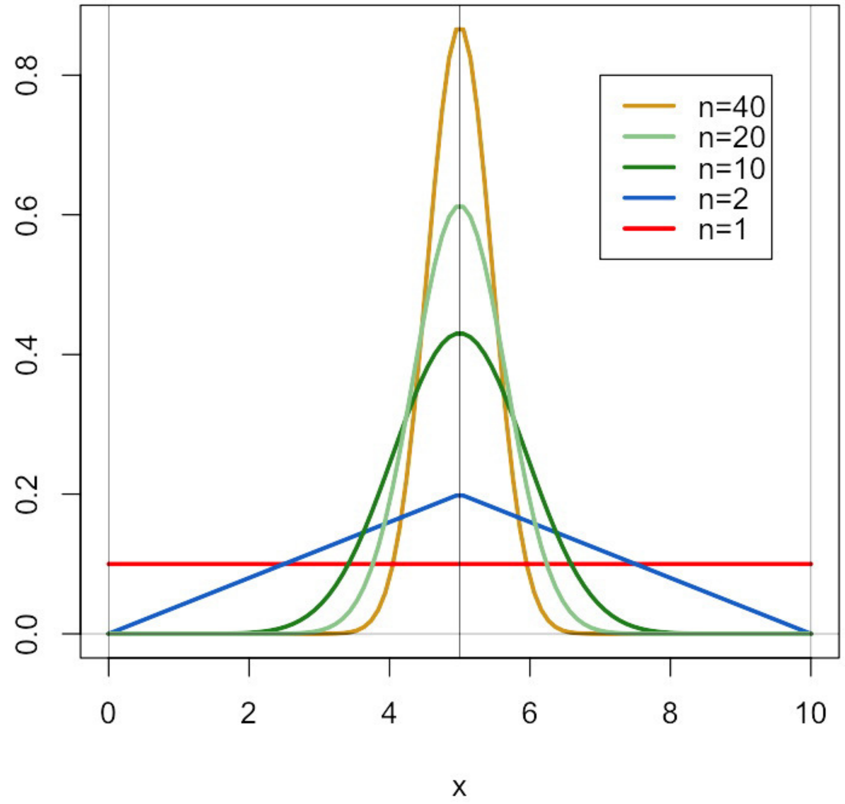
où $x \mapsto \Phi(x) = P[\mathcal{N}(0,1) \leq x]$ est la fonction de répartition de la loi $\mathcal{N}(0,1)$.

Autrement dit : pour n grand,

$$Z^{(n)} = \frac{\bar{X}^{(n)} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}(\bar{X}^{(n)} - \mu)}{\sigma}$$

est presque de loi $\mathcal{N}(0,1)$.





Graphes de la densité de $\bar{X}^{(n)}$ pour différents n, dans le cas où $X_1, \ldots, X_n$ sont i.i.d. de loi $24.7 + t_5$ (gauche) et de loi $\mathrm{Unif}(0, 10)$ (droite).

La distribution de $\bar{X}^{(n)}$ ressemble de plus en plus à une loi normale quand n grandit.

Définition

Soient $(X^{(n)})$ une suite de v.a. et X une autre v.a. Notons $F^{(n)}$ et F les fonctions de répartition correspondantes. $X^{(n)} \to X$ en loi (not : $X^{(n)} \xrightarrow{\mathcal{L}} X$) $\Leftrightarrow F^{(n)}(x) \to F(x)$ en tout point x où F est continue.

Le TCL affirme que $Z^{(n)} \xrightarrow{\mathcal{L}} Z$, où $Z \sim \mathcal{N}(0,1)$ (on écrit parfois $Z^{(n)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$).

Prouver le TCL requiert le résultat suivant (parallèle au théorème en ch.2-p. 95).

Théorème (théorème de continuité)

Supposons que $X^{(n)}$ et X admettent les fonctions génératrices des moments $M_{X^{(n)}}(t)$ et $M_X(t)$ et qu'il existe $\varepsilon > 0$ tel que $M_{X^{(n)}}(t) \to M_X(t)\ \forall t \in]-\varepsilon, \varepsilon[$. Alors $X^{(n)} \xrightarrow{\mathcal{L}} X$.

Preuve du TCL : par définition de la fonction génératrice des moments,

$$M_{Z^{(n)}}(t) = \mathrm{E}\left[e^{tZ^{(n)}}\right] = \mathrm{E}\left[e^{\frac{t\sqrt{n}}{\sigma}(\bar{X}^{(n)}-\mu)}\right] = \mathrm{E}\left[e^{\frac{t\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right)}\right]$$

$$= \mathrm{E}\left[e^{\frac{t}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu)}\right] = \mathrm{E}\left[e^{\frac{t}{\sigma\sqrt{n}}(X_1-\mu)} \times \ldots \times e^{\frac{t}{\sigma\sqrt{n}}(X_n-\mu)}\right].$$

Puisque les X_i sont i.i.d., ceci livre

$$M_{Z^{(n)}}(t) = \mathrm{E}\left[e^{\frac{t}{\sigma\sqrt{n}}(X_1-\mu)}\right] \times \ldots \times \mathrm{E}\left[e^{\frac{t}{\sigma\sqrt{n}}(X_n-\mu)}\right]$$

$$= \left(\mathrm{E}\left[e^{\frac{t}{\sigma\sqrt{n}}(X_1-\mu)}\right]\right)^n = \left(M_{X_1-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n.$$

Puisque $\frac{t}{\sigma\sqrt{n}} \to 0$ quand $n \to \infty$, un développement de Taylor centré en 0 se justifie :

$$\begin{aligned}
M_{Z^{(n)}}(t) &= \left(M_{X_1-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n \\
&= \left(M_{X_1-\mu}(0) + \frac{t}{\sigma\sqrt{n}} M'_{X_1-\mu}(0) + \left(\frac{t}{\sigma\sqrt{n}}\right)^2 \frac{M''_{X_1-\mu}(0)}{2} + \dots\right)^n \\
&= \left(1 + \frac{t}{\sigma\sqrt{n}} \mathrm{E}[X_1 - \mu] + \frac{t^2}{2\sigma^2 n} \mathrm{E}[(X_1-\mu)^2] + \dots\right)^n \\
&= \left(1 + 0 + \frac{t^2}{2n} + \dots\right)^n \to e^{t^2/2} = M_{\mathcal{N}(0,1)}(t)
\end{aligned}$$

quand $n \to \infty$ (en écrivant $f(n) = e^{\ln f(n)}$ et en utilisant L'Hospital).

Le TCL découle donc du théorème de continuité. $\square$

Illustration du TCL

Soient E une expérience aléatoire et $(\Omega, \mathcal{A}, P)$ un espace probabilisé associé.
Soit X la v.a. qui vaut 1 si un événement fixé $A(\in \mathcal{A})$ se produit et 0 sinon.
Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire simple correspondant (// n répétitions de E).

Alors on a $\mu = \mathrm{E}[X] = P[A]$ et $\sigma^2 = \mathrm{Var}[X] = P[A](1 - P[A])$ (ch.2-p.51)

Donc le TCL affirme que, pour tout x,

$$P\left[\frac{\bar{X}^{(n)} - P[A]}{\sqrt{\frac{P[A](1-P[A])}{n}}} \leq x \right] \to \Phi(x).$$

Ce cas particulier porte le nom de théorème de de Moivre–Laplace.

Si $P[A] = P[\text{obtenir "face" en lançant une pièce}] = \frac{1}{2}$ et $n = 100$, ceci donne

$$\frac{\bar{X}^{(n)} - P[A]}{\sqrt{\frac{P[A](1-P[A])}{n}}} = 20\left(\bar{X}^{(n)} - \frac{1}{2}\right) \approx \mathcal{N}(0,1).$$

En particulier,

$$\begin{aligned}
P[0.4 < \bar{X}^{(n)} \leq 0.6] &= P\left[-0.1 < \bar{X}^{(n)} - \frac{1}{2} \leq 0.1\right] \\
&= P\left[-2 < 20\left(\bar{X}^{(n)} - \frac{1}{2}\right) \leq 2\right] \\
&\approx P\left[-2 < \mathcal{N}(0,1) \leq 2\right] \\
&= \Phi(2) - \Phi(-2) \\
&\approx 0.9545
\end{aligned}$$

Il y a donc $\approx$ 95.45% de chance qu'on ait entre 40% et 60% de "face" en 100 lancers

Ici, on pourrait obtenir le résultat exact :

$$
\begin{aligned}
P[0.4 < \bar{X}^{(n)} \leq 0.6] &= P[40 < 100\bar{X}^{(n)} \leq 60] \\
&= P[40 < \underbrace{X_1 + \ldots + X_{100}}_{\sim \mathrm{Bin}\left(100, \frac{1}{2}\right)} \leq 60] \\
&= P[\mathrm{Bin}(100, \tfrac{1}{2}) = 41] + P[\mathrm{Bin}(100, \tfrac{1}{2}) = 42] \\
&\qquad + \ldots + P[\mathrm{Bin}(100, \tfrac{1}{2}) = 60] \\
&= \sum_{k=41}^{60} \tbinom{100}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{100-k} \\
&= 0.953956\ldots
\end{aligned}
$$

On voit que l'approximation fournie par le TCL est très bonne.

Et il y a des avantages numériques évidents à utiliser le TCL pour n grand.

Le théorème central-limite

Le caractère miraculeux du TCL est qu'il permet de calculer des probabilités aussi efficacement que ci-dessus dans les situations où on ignore la distribution des X_i.

Pour illustrer ceci, nous aurons besoin des deux résultats suivants.

Théorème

$X^{(n)} \xrightarrow{P} X \Rightarrow X^{(n)} \xrightarrow{\mathcal{L}} X$

Théorème (Lemme de Slutzky)

Soient $(X^{(n)})$, $(Y^{(n)})$ deux suites de v.a., X une autre v.a., et a une constante. Alors
(i) $X^{(n)} \xrightarrow{\mathcal{L}} X$ *et* $Y^{(n)} \xrightarrow{\mathcal{L}} a \Rightarrow X^{(n)} + Y^{(n)} \xrightarrow{\mathcal{L}} X + a$
(ii) $X^{(n)} \xrightarrow{\mathcal{L}} X$ *et* $Y^{(n)} \xrightarrow{\mathcal{L}} a \Rightarrow X^{(n)} Y^{(n)} \xrightarrow{\mathcal{L}} Xa$
(iii) $X^{(n)} \xrightarrow{\mathcal{L}} X$ *et* $Y^{(n)} \xrightarrow{\mathcal{L}} a (\neq 0) \Rightarrow X^{(n)} / Y^{(n)} \xrightarrow{\mathcal{L}} X / a$

Le théorème central-limite

Un exemple type d'application de ce théorème est le suivant.

Soit X une v.a. de moyenne $\mu = \mathrm{E}[X]$ et de variance $\sigma^2 = \mathrm{Var}[X] < \infty$.
Soit $(X_1, \ldots, X_n)$ un échantillon aléatoire simple associé.
Le TCL affirme que

$$\sqrt{n}\,\frac{\bar{X}^{(n)} - \mu}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

D'autre part, on a vu que $s \xrightarrow{p.s.} \sigma$, donc aussi $s \xrightarrow{P} \sigma$ et $s \xrightarrow{\mathcal{L}} \sigma$.

Le lemme de Slutzky livre donc que

$$\sqrt{n}\,\frac{\bar{X}^{(n)} - \mu}{s} = \left(\sqrt{n}\,\frac{\bar{X}^{(n)} - \mu}{\sigma}\right) \times \left(\frac{\sigma}{s}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \times 1 = \mathcal{N}(0,1).$$

Ce qui rend ce résultat si important est qu'il tient quelle que soit la distribution des X_i (et est donc applicable même quand on ne connaît pas celle-ci).

En posant $z_\beta = \Phi^{-1}(1-\beta)$, le résultat ci-dessus permet d'écrire que

$$P\Big[\bar{X}^{(n)} - z_{\alpha/2}\,\frac{s}{\sqrt{n}} \le \mu \le \bar{X}^{(n)} + z_{\alpha/2}\,\frac{s}{\sqrt{n}}\Big] \;=\; P\Big[-z_{\alpha/2} \le \sqrt{n}\,\frac{\bar{X}^{(n)}-\mu}{s} \le z_{\alpha/2}\Big]$$

$$\to P[-z_{\alpha/2} \le \mathcal{N}(0,1) \le z_{\alpha/2}] = 1-\alpha.$$

Pour $\alpha = 5\%$, ceci implique que, si n est grand,

$$P\Big[\bar{X}^{(n)} - 1.96\,\frac{s}{\sqrt{n}} \le \mu \le \bar{X}^{(n)} + 1.96\,\frac{s}{\sqrt{n}}\Big] \approx 0.95 :$$

l'intervalle aléatoire $[\bar{X}^{(n)} - 1.96\,\frac{s}{\sqrt{n}}, \bar{X}^{(n)} + 1.96\,\frac{s}{\sqrt{n}}]$ constitue une "fourchette" ayant la propriété de contenir la valeur inconnue de μ avec une probabilité d'environ 95% !

La longueur de cet intervalle, pour un niveau d'erreur α fixé, donne une information importante sur l'incertitude qui règne sur μ

Si un échantillon $X_1, \ldots, X_{100}$ de $n = 100$ bières de la Jefke a donné

$$\bar{X}^{(n)} = 24.7 \quad \text{et} \quad s = 1.04,$$

on a

$$\left[\bar{X}^{(n)} - 1.96\,\frac{s}{\sqrt{n}}, \bar{X}^{(n)} + 1.96\,\frac{s}{\sqrt{n}}\right] = [24.5, 24.9],$$

ce qui tend à indiquer que $\mu < 25$!

Plan du chapitre 4

Le lemme de Fisher

La bière n'est pas chère à la Jefke, et on peut donc faire en sorte de fonder la décision sur un échantillon aléatoire simple de taille n très grande.

Mais il arrive qu'on soit amené à travailler avec de petits échantillons. Cela peut s'expliquer par la rareté des observations, par le coût énorme pour obtenir des observations supplémentaires (c'est le cas en génétique, par exemple), etc.

Si n est trop petit ($n < 30$?), l'approximation fournie le TCL est trop peu précise pour donner des résultats satisfaisants en pratique.

Que peut-on faire dans ce cas ?

Comme on va le montrer, on peut encore procéder aux mêmes types d'analyse que ci-dessus, sous l'hypothèse que la distribution des X_i est normale.

Le résultat fondamental est le suivant.

Théorème (lemme de Fisher)

Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. $\mathcal{N}(\mu, \sigma^2)$ ($n \geq 2$). Alors

(i) $\bar{X}^{(n)} \sim \mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$

(ii) $\dfrac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$

(iii) $\bar{X} \perp\!\!\!\perp s^2$.

Contrairement à la loi des grands nombres et au TCL (qui, puisque $n \to \infty$ dans ces théorèmes, seront dits "asymptotiques"), le résultat ci-dessus sera qualifié d'exact.

Ceci traduit le fait que le résultat tient pour tout n fixé, et pourra donc être utilisé même pour n petit.

Le point (i), énoncé en ch.4-p.18, est laissé comme exercice.
Pour montrer les points (ii)-(iii), nous utiliserons le lemme suivant.

Lemme

Soit A une matrice $(k \times k)$ symétrique ($A^T = A$), idempotente ($A^2 = A$), et de trace $r (\in \mathbb{N}_0)$. Soit B une matrice $(s \times k)$ et telle que $BA = 0$. Soit $Z \sim \mathcal{N}_k(0, I_k)$. Alors (i) $Z^T A Z \sim \chi^2_r$ et (ii) $Z^T A Z \perp\!\!\!\perp BZ$.

Preuve du lemme : (i) puisque A est symétrique, elle admet la décomposition

$$A = U \Lambda U^T,$$

où U est une matrice $(k \times k)$ orthogonale ($U^T U = I_k = U U^T$) et

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_k \end{pmatrix}.$$

Le lemme de Fisher

Si on pose $(\tilde{Z}_1, \ldots, \tilde{Z}_k)^T := \tilde{Z} := U^T Z (\sim \mathcal{N}_k(U^T 0, U^T I_k U) = \mathcal{N}_k(0, I_k))$, on a donc

$$Z^T A Z = Z^T (U \Lambda U^T) Z = (U^T Z)^T \Lambda (U^T Z) = \tilde{Z}^T \Lambda \tilde{Z} = \sum_{\ell=1}^{k} \lambda_\ell \tilde{Z}_\ell^2.$$

L'idempotence de A implique que $U\Lambda U^T = A = A^2 = (U\Lambda U^T)(U\Lambda U^T) = U\Lambda^2 U^T$, ce qui indique que $\Lambda = \Lambda^2$. Autrement dit, $\lambda_\ell = \lambda_\ell^2 \; \forall \ell$, c'est-à-dire $\lambda_\ell = 0$ ou $1 \; \forall \ell$.

Le nombre de λ_ℓ égaux à 1 vaut $\mathrm{trace}[\Lambda] = \mathrm{trace}[\Lambda U^T U] = \mathrm{trace}[U \Lambda U^T] = \mathrm{trace}[A] = r$; au prix d'une permutation des colonnes de U, on peut faire en sorte que $\lambda_1 = \ldots = \lambda_r = 1$ et $\lambda_{r+1} = \ldots = \lambda_k = 0$.

On a donc en fait

$$Z^T A Z = \sum_{\ell=1}^{r} \tilde{Z}_\ell^2 \sim \chi_r^2,$$

où la loi χ_r^2 suit du fait que $\tilde{Z} \sim \mathcal{N}_k(0, I_k)$ implique que les $\tilde{Z}_\ell$ sont i.i.d. $\mathcal{N}(0, 1)$.

(ii) Par hypothèse, on a $0 = BA = BU\Lambda U^T$, ou de manière équivalente,

$$0 = BU\Lambda = BU \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

ce qui implique que $BU = (0 \,|\, C)$ pour une certaine matrice $(s \times (k-r))$ C.

Donc on a que

- $Z^T A Z = \sum_{\ell=1}^{r} \tilde{Z}_\ell^2 = f_1(\tilde{Z}_1, \tilde{Z}_2, \ldots, \tilde{Z}_r)$.
- $BZ = (BU)(U^T Z) = (0 \,|\, C)\tilde{Z} = f_2(\tilde{Z}_{r+1}, \tilde{Z}_{r+2}, \ldots, \tilde{Z}_k)$.

Puisque les $\tilde{Z}_\ell$ sont mutuellement indépendants, on conclut que $Z^T A Z \perp\!\!\!\perp BZ$. $\square$

Preuve de (ii)-(iii) dans le lemme de Fisher :
Posons $Z = (Z_1, \ldots, Z_n)^T$, où $Z_i := (X_i - \mu)/\sigma$.
Soit $B = \frac{1}{n} 1_n^T$, où $1_n := (1, 1, \ldots, 1)^T \in \mathbb{R}^n$, et soit $A = I_n - nB^T B$.

On vérifie directement que

- $\bar{X} = \sigma \bar{Z} + \mu = \sigma BZ + \mu$
- $$\frac{ns^2}{\sigma^2} = \sum_{i=1}^n (Z_i - \bar{Z})^2 = \sum_{i=1}^n Z_i^2 - n\bar{Z}^2 = Z^T Z - n(BZ)^2 = Z^T Z - n(Z^T B^T)(BZ) = Z^T A Z.$$

Au vu du lemme, il suffit donc de montrer que

(a1) A est symétrique

(a2) A est idempotente

(a3) $\mathrm{trace}[A] = n - 1$

(b) $BA = 0$

(z) $Z \sim \mathcal{N}_n(0, I_n)$.

Le lemme de Fisher

Pour certains points, nous aurons besoin de l'identité $BB^T = \frac{1}{n^2} 1_n^T 1_n = \frac{1}{n^2} n = \frac{1}{n}$:

(a1) $A^T = [I_n - nB^T B]^T = (I_n)^T - n(B^T B)^T = I_n - nB^T B = A$

(a2) $A^2 = [I_n - nB^T B][I_n - nB^T B] = I_n - 2nB^T B + n^2 B^T (BB^T) B = I_n - nB^T B = A$

(a3) $\mathrm{trace}[A] = \mathrm{trace}[I_n] - n\,\mathrm{trace}[B^T B] = n - n\,\mathrm{trace}[BB^T] = n - n\,\mathrm{trace}[\frac{1}{n}] = n - 1$

(b) $BA = B[I_n - nB^T B] = B - n(BB^T)B = B - B = 0$

(z) Les X_i étant i.i.d. $\mathcal{N}(\mu, \sigma^2)$, on a que les $Z_i = (X_i - \mu)/\sigma$ sont i.i.d. $\mathcal{N}(0, 1)$. Par le point (iv), ch.3-p.129, il en découle que $Z = (Z_1, \ldots, Z_n)^T \sim \mathcal{N}_n(0, I_n)$. $\square$

Le lemme de Fisher

Théorème (lemme de Fisher)

Soient $X_1, \ldots, X_n$ des variables aléatoires i.i.d. $\mathcal{N}(\mu, \sigma^2)$ ($n \geq 2$). Alors

(i) $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

(ii) $\frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$

(iii) $\bar{X} \perp\!\!\!\perp s^2$.

Rappelons que si $Z \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2_\nu$ sont mutuellement indépendantes, alors

$$\frac{Z}{\sqrt{Y/\nu}} \sim t_\nu.$$

Il découle donc du lemme de Fisher que

$$\sqrt{n-1}\,\frac{(\bar{X}-\mu)}{s} = \frac{\frac{\sqrt{n}\,(\bar{X}-\mu)}{\sigma}}{\sqrt{\left(\frac{ns^2}{\sigma^2}\right)/\,(n-1)}} \sim t_{n-1}.$$

Le lemme de Fisher

Pour peu que la distribution des X_i soit normale, on peut donc écrire ($\forall n \geq 2$)

$$
\begin{aligned}
P\Big[\bar{X} - t_{n-1;\alpha/2}\,\frac{s}{\sqrt{n-1}} \leq \mu \leq \bar{X} + t_{n-1;\alpha/2}\,\frac{s}{\sqrt{n-1}}\Big] \\
&= P\Big[-t_{n-1;\alpha/2} \leq \sqrt{n-1}\,\frac{\bar{X}-\mu}{s} \leq t_{n-1;\alpha/2}\Big] \\
&= P[-t_{n-1;\alpha/2} \leq t_{n-1} \leq t_{n-1;\alpha/2}] = 1-\alpha,
\end{aligned}
$$

où $t_{n-1;\beta}$ désigne le quantile d'ordre $1-\beta$ de la loi t_{n-1}. Par exemple, pour $\alpha = 0.05$ et $n = 9$, ceci fournit

$$
P\Big[\bar{X} - 2.31\,\frac{s}{\sqrt{n-1}} \leq \mu \leq \bar{X} + 2.31\,\frac{s}{\sqrt{n-1}}\Big] = 0.95.
$$

Pour peu que la quantité de bière servie à la Jefke soit de loi normale, on peut donc, même si on n'a en poche que de l'argent pour 9 bières, construire des "fourchettes" remplissant le même rôle que celles en ch.4-p.29.

Références

Dehon, C., Hallin, M., Paindaveine, D., Thomas-Agnan, C., et Vermandele, C. (2020). *Probabilités et inférence statistique*. Editions de l'Université de Bruxelles, Editions Ellipses.

Devore, J.-L. (1991). *Probability and Statistics for Engineering and the Sciences*. Pacific Grove, Brooks-Cole.

Evans, M., N. Hastings et B. Peacock (2000). *Statistical Distributions*. 3rd ed., New York, Wiley Series in Probability and Statistics.

Tassi, P. (2004). *Méthodes statistiques*. 3e éd., Paris, Économica.