Université de Strasbourg
UFR de Mathématiques
et d'Informatique
Han-Ping LI

Année 2019/2020
L3
Statistique
Etude de cas

Chapitre III Intervalles de confiance

Soit $(X_1, \ldots, X_n)$ un échantillon de loi $\left\{\mathbb{P}_\theta, \theta \in \Theta\right\}$.

Le paramètre θ est estimé par $\widehat{\theta}(X_1, \ldots, X_n)$.

On peut mesurer la qualité de l'estimateur en évaluant de manière probabiliste les écarts possible entre $\widehat{\theta}$ et θ.

Un estimateur permet de calculer une valeur sur un échantillon qui devrait être proche du paramètre θ sans pour autant savoir si cette valeur est totalement fiable.

C'est pourquoi on a introduit la notion d'intervalle de confiance : c'est un intervalle dans lequel se trouve le paramètre θ avec un faible risque α ou une grande probabilité $1-\alpha$. On peut en théorie choisir α aussi proche de 0 que l'on peut, mais alors l'intervalle de confiance devient très grand et imprécis. Il faut donc trouver un compromis entre précision de l'intervalle et sûreté (risque α petit. La probabilité $1-\alpha$ est appelée niveau de confiance.

Problème : comment trouver un intervalle de confiance ? L'idée est d'utiliser une variable aléatoire U de loi connue qui relie le meilleur estimateur deθ et de θ, le paramètre à estimer.

Supposons que $X_1, ..., X_n$ suivent une loi $\mathbf{P}_\theta$. Soit $\alpha > 0$ notre marge d'erreur, il est possible de trouver deux bornes aléatoires

$$BInf(X_1, \ldots, X_n) \text{ et } BSup(X_1, \ldots, X_n) \text{ tels que}$$

$$\mathbb{P}\Big(BInf(X_1, \ldots, X_n) \leq \theta \leq BSup(X_1, \ldots, X_n)) = 1 - \alpha.$$

Critères :

- sans biais
- le plus précis

(mais pas forcément plus court)

3-1. IC de μ l' Espérance d'une loi normale $\mathcal{N}(\mu, \sigma^2)$

On sait que

1) $\overline{X}$ suit une loi $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$;

2) $(\overline{X} - \mu)$ suit une loi $\mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$;

3) $T = \frac{\sqrt{n}(\overline{X}-\mu)}{S_c}$ suit une loi suit une loi de Student à (n-1) degré de liberté.

Cette dernière est une fonction pivotale : c-à-d que la loi de $T = \dfrac{\sqrt{n}(\overline{X} - \mu)}{S_c}$ est indépendante de deux paramètres μ et σ.

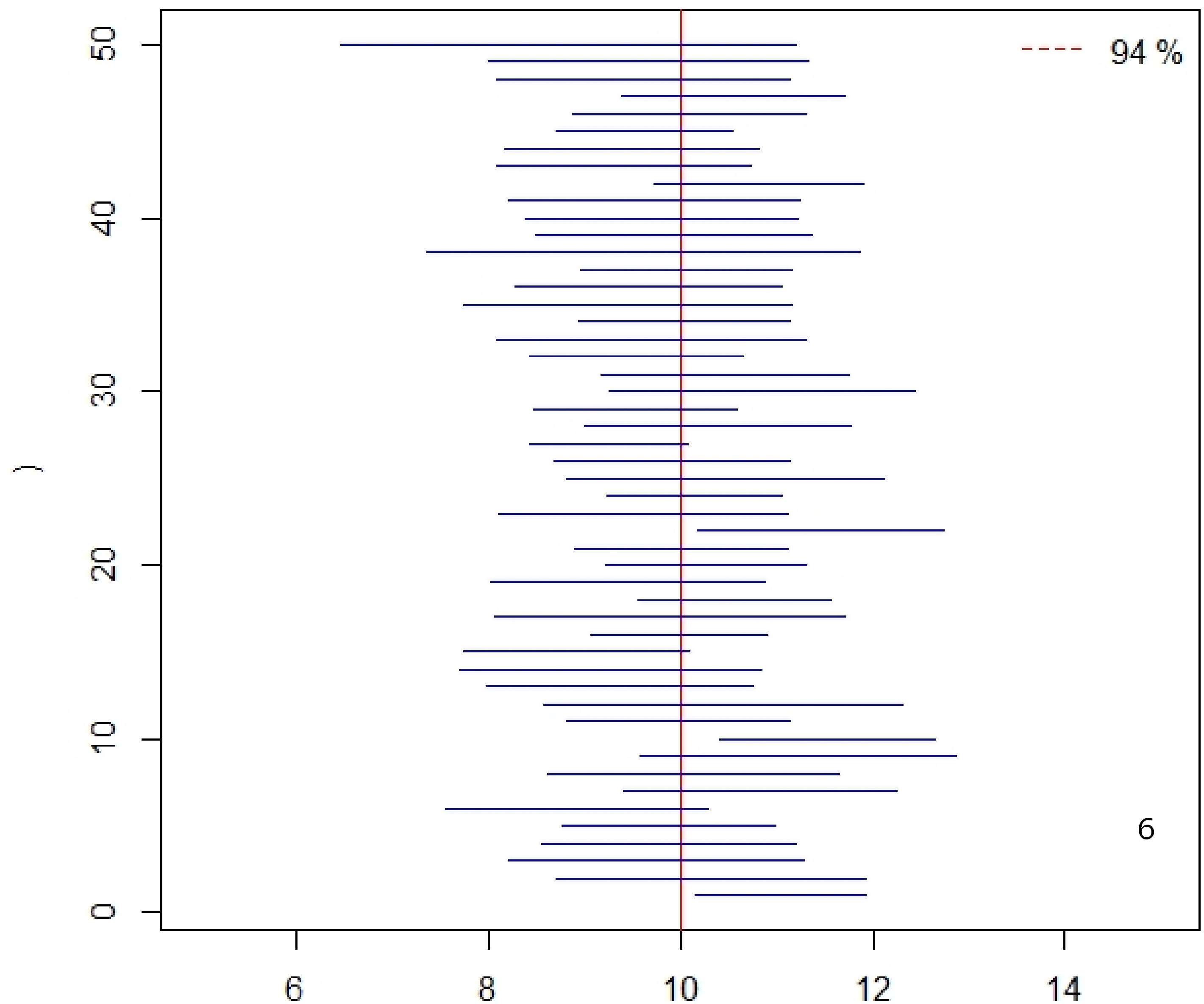
Se basant sur $\overline{X}$ le meilleur estimateur de μ, on utilise la fonction pivotale :

$$\frac{\sqrt{n}(\overline{X}-\mu)}{S_c}$$

avec une loi de Student à (n-1) degrés de liberté, indépendant de deux paramètres μ et σ, on peut déterminer un quantile C_q tel que $\mathbb{P}\left(\left|\frac{\sqrt{n}(\overline{X}-\mu)}{S_c}\right| \leq C_q\right) = 1-\alpha$. Autrement dit,

$C_1 = -C_q$, et $C_2 = C_q$:

$$\begin{aligned} BInf &= \overline{X} - C_q\frac{S_c}{\sqrt{n}}, \\ BSup &= \overline{X} + C_q\frac{S_c}{\sqrt{n}}. \end{aligned}$$



94 %
6

3-2. IC de σ^2 la Variance d'une loi normale $\mathcal{N}(\mu, \sigma^2)$

Se basant sur S_c^2 est le meilleur estimateur de σ^2, on utilise la fonction pivotale :

$$\frac{(n-1)S_c^2}{\sigma^2}$$

avec une loi de Khi-deux à (n-1) degrés de liberté, indépendant de deux paramètres μ et σ. On détermine deux quantile C_1 et C_2 tels que

$$\mathbb{P}\Big(C_1 \leq \frac{(n-1)S_c^2}{\sigma^2} \leq C_2\Big) = 1 - \alpha.$$

Ainsi,

$$\begin{aligned} BInf &= \frac{(n-1)S_c^2}{C_2}, \\ BSup &= \frac{(n-1)S_c^2}{C_1}. \end{aligned}$$

Ayant un niveau de confiance donnée,

$$\mathbb{P}\Big(\frac{(n-1)S_c^2}{\sigma^2} < C_1\Big) = \mathbb{P}\Big(\frac{(n-1)S_c^2}{\sigma^2} > C_2\Big) = \frac{\alpha}{2}.$$

fournissent un choix simple à calculer, et proche du choix optimal.

Le choix optimal n'est pas celui qui préconise l'intervalle le plus court. Voir un exemple.

3-3. Intervalle de confiance d'une proportion p inconnue

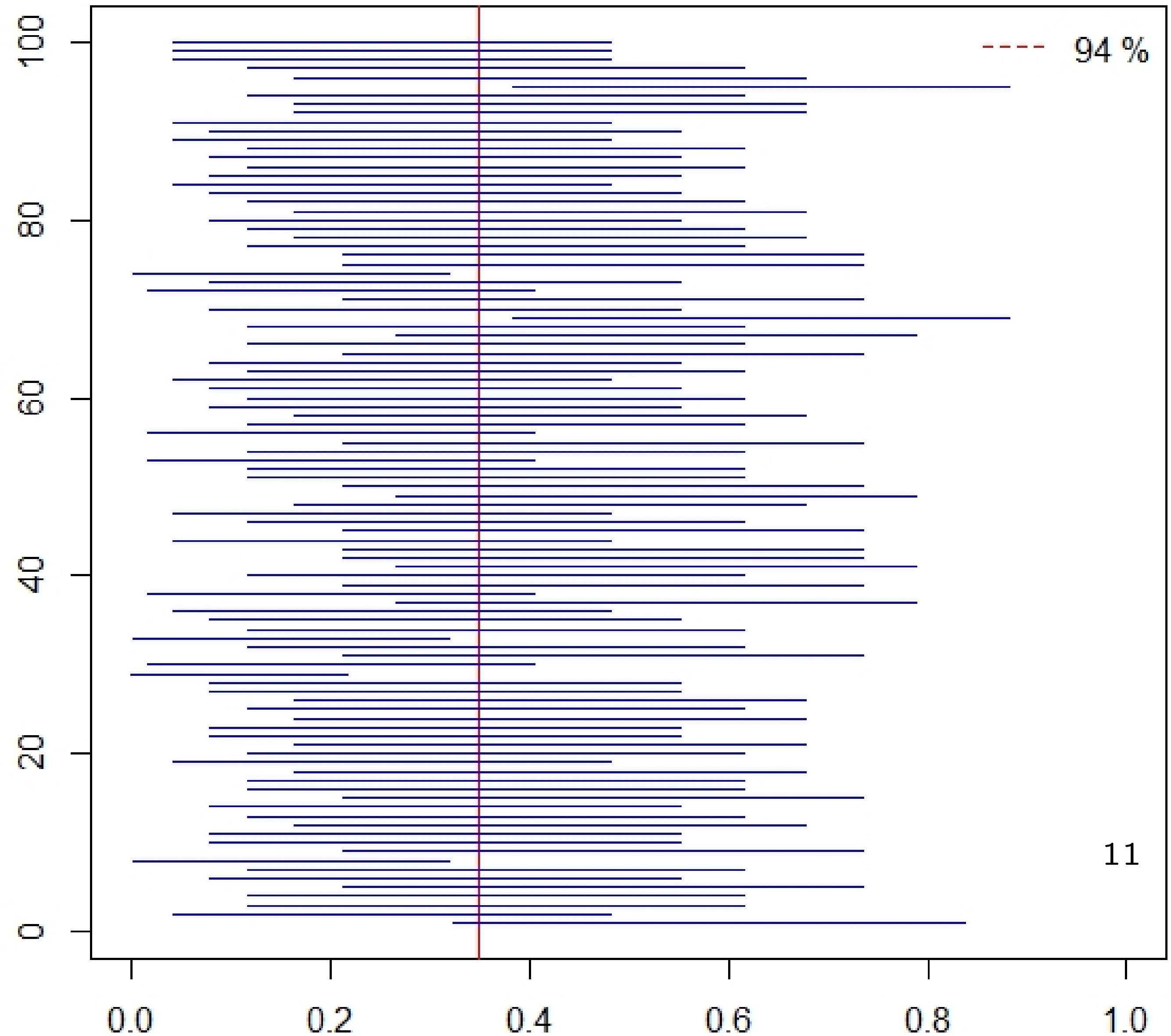
On s'intéresse à la proportion p d'individus possédant une certaine caractéristique dans une population. On prélève d'un échantillon de taille n avec $X_j = 1$ ou 0 selon. $\overline{X}$ représente la proportion calculée sur l'échantillon.

On sait comment déterminer l'intervalle de confiance du paramètre p optimal (la méthode "exacte") par une procédure relativement complexe.

Si p est très proche de 0 ou 1 ou si n n'est pas assez grand, alors on utilise R directement pour obtenir l'intervalle de confiance :

```
install.packages("binom"); library("binom")
```

```
binom.confint(sum(echant),n, conf.level=0.95,method="exact")
```



94 %
11
0
20
40
60
80
100
0.0
0.2
0.4
0.6
0.8
1.0

Si n est assez grand, et p n'est pas trop proche de 0 ou 1, alors la loi de $\frac{\sqrt{n}(\overline{X}-p)}{\sqrt{p(1-p)}}$ est approximativement $N(0,1)$ car $\mu = \mathbf{E}(X_i) = p$ et $\sigma^2 = \mathbf{Var}(X_i) = p(1-p)$. Plus précisément

1) Si $0.1 \leq p \leq 0.9$, il faut

$$n \geq 30, \; np \geq 5, \text{ et } \; n(1-p) \geq 5.$$

2) Si $p < 0.1$ ou $p > 0.9$, il faut alors

$$n \geq 50, \; np \geq 10, \text{ et } \; n(1-p) \geq 10.$$

$$\mathcal{L}\Big(\frac{\sqrt{n}(\overline{X}-p)}{\sqrt{p(1-p)}}\Big) \text{ est approximativement } \mathcal{N}(0,1).$$

On choisit une constante $C_q = qnorm(1-\frac{\alpha}{2})$ telle que

$$\mathbf{P}\Big(\mathcal{N}(0,1) \leq C_q\Big) = 1-\frac{\alpha}{2} \Leftrightarrow \mathbf{P}(\Big|\mathcal{N}(0,1)\Big| \leq C_q) = 1-\alpha$$

Ainsi avec une probabilité approximativement $1-\alpha$, on a

$$\left|\frac{\sqrt{n}(\overline{X}-p)}{\sqrt{\overline{X}(1-\overline{X})}}\right| \leq C_q$$

On en déduit qu'avec une probabilité approximativement $1-\alpha$, (méthode de Wald)

$$\begin{aligned} BInf &= \overline{X} - C_q\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}, \\ BSup &= \overline{X} + C_q\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}. \end{aligned}$$

3-4. Intervalle de confiance de la loi de Poisson (n est assez grand

On s'intéresse au paramètre λ de la loi de Poisson. On prélève d'un échantillon de taille n. $\overline{X}$ représente la moyenne d'échantillon.

Si n est assez grand, et $n\lambda$ n'est pas petit, alors

la loi de $\frac{\sqrt{n}(\overline{X}-\lambda)}{\sqrt{\lambda}}$ est approximativement $N(0,1)$ car $\mu = \mathbf{E}(X_i) = \lambda$ et $\mu = \mathbf{Var}(X) = \lambda$. Plus précisément

Si $n\lambda \geq 15$,

$$\mathcal{L}\Big(\frac{\sqrt{n}(\overline{X}-\lambda)}{\sqrt{\lambda}}\Big) \text{ est approximativement } \mathcal{N}(0,1).$$

Ainsi avec une probabilité approximativement $1-\alpha$, on a

$$\left|\frac{\sqrt{n}(\overline{X}-\lambda)}{\sqrt{\lambda}}\right| \leq C_q$$

Or, Comme $\lim_{n\to\infty} \overline{X} = \lambda$ Ainsi avec une probabilité 100%, on a aussi

$$\mathcal{L}\Big(\frac{\sqrt{n}(\overline{X}-\lambda)}{\sqrt{\overline{X}}}\Big) \text{ est approximativement } \mathcal{N}(0,1).$$

Ainsi avec une probabilité approximativement $1-\alpha$, on a

$$\left|\frac{\sqrt{n}(\overline{X}-p)}{\sqrt{\overline{X}(1-\overline{X})}}\right| \leq C_q$$

De cette deuxième approximation, on en déduit qu'avec une probabilité approximativement $1-\alpha$,

$$\begin{aligned} BInf &= \overline{X} - C_q\sqrt{\frac{\overline{X}}{n}}, \\ BSup &= \overline{X} + C_q\sqrt{\frac{\overline{X}}{n}}. \end{aligned}$$