Statistique Mathématique

Partiel — 27 février 2023

Durée : 2 heures. Documents et calculatrices interdits. Détaillez et justifiez chaque réponse. Les questions marquées d'un astérisque ($\star$) sont plus difficiles, il est normal qu'elles prennent plus de temps. Bonne chance !

Intervalles de confiance pour une Bernoulli

Dans ce problème, on considère des réalisations i.i.d. $X_1, \ldots, X_n$ d'une variable aléatoire X. Nous allons faire l'hypothèse que X suit la loi de Bernoulli de paramètre inconnu $p \in]0,1[$, ce qu'on note $X \sim \mathcal{B}(p)$. En d'autres termes, $X = 1$ avec probabilité p et $X = 0$ avec probabilité $1-p$.

Partie I : Wald ou l'approximation gaussienne

1. Quelle est l'espérance de X ? Sa variance ?
2. En utilisant la méthode des moments, proposez un estimateur $\hat{p}$ de p.
3. Montrez que $\hat{p}$ est un estimateur fortement consistant de p.
4. Montrez que $\hat{p}$ est asymptotiquement normal.
5. En utilisant les questions précédentes, montrez que

$$\sqrt{\frac{n}{\hat{p}(1-\hat{p})}}(\hat{p}-p) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)\,.$$

   Justifiez soigneusement votre réponse.

$\sqrt{n}(\hat{p}-p) \to \mathcal{N}(0,$

6. Soit $\alpha \in]0,1[$. En utilisant la question précédente, montrez que l'intervalle de confiance

$$C_A(n,\alpha) := \left[\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

   est asymptotiquement de niveau α, où $z = z_{1-\alpha/2}$ est le quantile de niveau $1-\alpha/2$ de la gaussienne centrée réduite.

On appelle traditionnellement C_A l'*intervalle de confiance de Wald*.

7. A l'aide de la figure 1, donnez au moins deux inconvénients de l'utilisation de C_W en pratique.

Partie II : une alternative, Wilson

Une tentative pour proposer un meilleur intervalle de confiance est de ne pas remplacer directement la variance asymptotique par son estimée (question I.5).

1. En utilisant la question I.4. et la définition de z (question I.6.), montrez que

$$\mathbb{P}\left(\frac{n}{p(1-p)}(\hat{p}-p)^2 \leq z^2\right) \to 1-\alpha\,. \tag{1}$$

2. En traitant p comme une inconnue dans l'équation (1), montrez que dans le cas limite on aboutit à une équation du second degré en p.
3. Résoudre cette équation et en déduire l'intervalle de confiance asymptotique suivant :

$$C_B(n,\alpha) := \left[\frac{1}{1+\frac{z^2}{n}}\left(\hat{p}+\frac{z^2}{2n}\right) \pm \frac{z}{1+\frac{z^2}{n}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\frac{z^2}{4n^2}}\right].$$

4. Montrez que, au premier ordre, C_B coïncide avec C_A lorsque n est grand.
5. En comparant les figures 1 et 2, quel intervalle de confiance préférez-vous utiliser pour $n \in [25, 100]$? Quel pourrait-être un inconvénient de ce choix ?

$X \rightsquigarrow B(p)$

Partie I :

1) Espérence de x

$X_i = \begin{cases} 1 \\ 0 \end{cases}$, $\mathbb{P}(X_i = 1) = P$ et $P(X_i = 0) = 1 - P$

$\Rightarrow \mathbb{E}[X] = (X_1 = 1)\mathbb{P}(X_i = 1) + (X_i = 0)\mathbb{P}(X_i = 0)$

$= P - 0 \times (1 - P)$

$$\boxed{\mathbb{E}[X] = P}$$

$\mathbb{E}[X^2] = 1 \times P + 0(1 - P) = P$

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

$$\boxed{\mathbb{V}(X) = P(1 - P)}$$

2-) Méthode des moment pour estimé P

moments théorique de x

$m_1 = \mathbb{E}[X] = P \Rightarrow P = m_1$

moment empirique de x

$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i \Rightarrow \hat{P} = \hat{m}_1 \Rightarrow \hat{P} = \frac{1}{n} \sum_{i=1}^{n} X_i$

3-) En utilisant la loi forte des grand nombre on a :

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[n \to +\infty]{P.S} \mathbb{E}[X_1] = P$$

$\Rightarrow \hat{P} \xrightarrow[n \to +\infty]{P.S} P$ donc $\hat{P}$ est un estimateur fortement consistant de P

4- Montrons que $\hat{P}$ est asymptotiquem normale

$$E[X] = P = \mu \text{ et } Var(X) = P(1-P) = \sigma^2$$

d'après TCL

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow[n \to +\infty]{Loi} \mathcal{N}(0,1)$$

$$\Rightarrow \sqrt{n}\left(\frac{\hat{P} - P}{\sqrt{P(1-P)}}\right) \xrightarrow[n \to +\infty]{Loi} \mathcal{N}(0,1)$$

donc $\hat{P}$ est est asymptotiquement normal

5-) D'après la question 4 :

$$\frac{\sqrt{n}}{\sqrt{P(1-P}}(\hat{P} - P) \xrightarrow[n \to +\infty]{Loi} \mathcal{N}(0,1)$$

$$\sqrt{\frac{n}{P(1-P)}}(\hat{P} - P) \xrightarrow[n \to +\infty]{Loi} \mathcal{N}(0,1)$$

de plus $\hat{P} \xrightarrow[n \to +\infty]{P.S} P$ donc $\hat{P} \xrightarrow[n \to +\infty]{Loi} P$ ie $\hat{P} = P$ lorsq $n \to +\infty$ en loi

$$\Rightarrow \sqrt{\frac{n}{\hat{P}(1-\hat{P})}}(\hat{P} - P) \xrightarrow[n \to +\infty]{Loi} \mathcal{N}(0,1)$$

6) Montrons que l'intervalle de confiance de niveau $1-\alpha$ est $C_A(n,\alpha)$

On sait que si $T \sim \mathcal{N}(0,1)$

alors $\mathbb{P}(-z \leq T \leq z) \geq 1-\alpha$ avec $z = z_{1-\frac{\alpha}{2}}$

$$\Rightarrow \mathbb{P}\left(-z \leq \sqrt{\frac{n}{\hat{p}(1-\hat{p})}}(\hat{p}-p) \leq z\right) \geq 1-\alpha$$

$$\Rightarrow \mathbb{P}\left(-\frac{1}{\sqrt{\frac{n}{\hat{p}(1-\hat{p})}}} z \leq \hat{p}-p \leq \frac{1}{\sqrt{\frac{n}{\hat{p}(1-\hat{p})}}} z\right) \geq 1-\alpha$$

$$\Rightarrow \mathbb{P}\left(p \in \left[\hat{p} - \frac{1}{\sqrt{\frac{n}{\hat{p}(1-\hat{p})}}} z \,,\, \hat{p} + \frac{1}{\sqrt{\frac{n}{\hat{p}(1-\hat{p})}}} z\right]\right) \geq 1-\alpha$$

$$\Rightarrow C_A(n,\alpha) = \left[\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \,,\, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

7)

Partie 2

1/ Appliquons le Théorème de Porte-manteau sur la question 4

On peut le faire puisque :

$$T_n = \sqrt{\frac{n}{p(1-p)}}(\hat{p}-p) \xrightarrow[n\to+\infty]{\mathcal{L}oi} T \sim \mathcal{N}(0,1)$$

$$\Rightarrow \mathbb{P}\left(\sqrt{\frac{n}{p(1-p)}}(\hat{p}-p) \in [-z, z]\right) \xrightarrow[n\to+\infty]{\mathcal{L}oi} \mathbb{P}(-z \leq T \leq z)$$

$$\mathbb{P}(-z \leq T \leq z) = \mathbb{P}(T \leq z) - \mathbb{P}(T \leq -z)$$

$$= 1 - \frac{\alpha}{2} - \left(\frac{\alpha}{2}\right) \quad \text{car } -z_{1-\frac{\alpha}{2}} = z_{\frac{\alpha}{2}}$$

$$= 1 - \alpha$$

$$\Rightarrow \mathbb{P}\left(\sqrt{\frac{n}{p(1-p)}}(\hat{p}-p) \in [-z, z]\right) \xrightarrow[n\to+\infty]{\mathcal{L}oi} 1-\alpha$$

$$\Rightarrow \text{or } \sqrt{\frac{n}{p(1-p)}}(\hat{p}-p) \in [-z, z] \Leftrightarrow \left(\sqrt{\frac{n}{p(1-p)}}(\hat{p}-p)\right)^2 \leq z^2$$

$$\Rightarrow \mathbb{P}\left(\frac{n}{p(1-p)}(\hat{p}-p)^2 \leq z^2\right) \xrightarrow[n\to+\infty]{\mathcal{L}oi} 1-\alpha$$

2)

$$\frac{n\left(\hat{p}^2 - 2\hat{p}p + p^2\right)}{p - p^2} \leq z^2$$

$$n\hat{p}^2 - 2n\hat{p}p + np^2 \leq pz^2 - p^2z^2$$

$$\left(n+z^2\right)p^2 + \left(z^2 - 2n\hat{p}\right)p + n\hat{p}^2 \leq 0$$

$$\Delta = \left(z^2 - 2n\hat{p}\right)^2 - 4\left(n+z^2\right)n\hat{p}^2$$

$$= z^4 - 4nz^2\hat{p} + \cancel{4n^2\hat{p}^2} - \cancel{4n^2\hat{p}^2} - 4z^2n\hat{p}^2$$

$$= z^4 - 4nz^2\left(\hat{p} + \hat{p}^2\right)$$

$$z^2\left(z^2 - 4n\left(\hat{p} + \hat{p}^2\right)\right)$$