Université de Strasbourg
M1 - Magistère
Statistique - études de cas

Ségolen Geffray
geffray@math.unistra.fr
Année 2017/2018

TP2: illustration des propriétés d'un intervalle de confiance

Le but de ce TP est de proposer une illustration par simulations de certaines propriétés d'un intervalle de confiance pour un paramètre réel noté θ lorsque cet intervalle est déterminé à partir d'un échantillon i.i.d. noté $(X_1, ..., X_n)$. Notons

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et } S_n'^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 .$$

• Intervalle de confiance bilatéral symétrique pour $\theta = \mathbb{E}[X]$ pour un grand échantillon i.i.d. de loi admettant un moment d'ordre 2:

- L'**intervalle de confiance** asymptotique bilatéral pour θ au niveau de confiance $(1-\alpha)$ est issu de l'encadrement en probabilité suivant

$$\mathbb{P}\left[\overline{X}_n - F_N^{-1}(1-\alpha/2)\frac{\sqrt{S_n'^2}}{\sqrt{n}} \leq \theta \leq \overline{X}_n + F_N^{-1}(1-\alpha/2)\frac{\sqrt{S_n'^2}}{\sqrt{n}}\right] \approx 1-\alpha\,,$$

  l'approximation étant d'autant meilleure que n est grand.
  NB: $F_N^{-1}(1-\alpha/2)$ désigne le quantile d'ordre $(1-\alpha/2)$ de la loi gaussienne standard notée $\mathcal{N}(0,1)$.

- Il y a une probabilité égale environ à $(1-\alpha)$ pour que l'intervalle de confiance aléatoire contienne la moyenne de la population. De manière à peu près équivalente, sur 100 réalisations de l'intervalle de confiance, le paramètre inconnu sera effectivement dans la fourchette proposée dans $(1-\alpha)100\%$ des cas.

• Intervalle de confiance bilatéral symétrique pour $\theta = \mathbb{E}[X]$ pour un échantillon i.i.d. gaussien de taille quelconque:

- L'**intervalle de confiance** bilatéral pour θ au niveau de confiance $(1-\alpha)$ est issu de l'encadrement en probabilité suivant

$$\mathbb{P}\left[\overline{X}_n - F_{T(n-1)}^{-1}(1-\alpha/2)\frac{\sqrt{S_n'^2}}{\sqrt{n}} \leq \theta \leq \overline{X}_n + F_{T(n-1)}^{-1}(1-\alpha/2)\frac{\sqrt{S_n'^2}}{\sqrt{n}}\right] = 1-\alpha\,,$$

  où $F_{T(n-1)}^{-1}(1-\alpha/2)$ désigne le quantile d'ordre $(1-\alpha/2)$ de la loi de Student à $(n-1)$ degrés de liberté notée $T(n-1)$.

- Il y a une probabilité d'exactement $(1-\alpha)$ pour que l'intervalle de confiance aléatoire contienne la moyenne de la population. De manière à peu près équivalente, sur 100 réalisations de l'intervalle de confiance, le paramètre inconnu sera effectivement dans la fourchette proposée dans $(1-\alpha)100\%$ des cas.

• Intervalle de confiance bilatéral symétrique pour $\theta = \mathbb{E}[X]$ pour un grand échantillon i.i.d. de loi de Poisson:

- L'**intervalle de confiance** asymptotique bilatéral pour θ au niveau de confiance $(1-\alpha)$ est issu de l'encadrement en probabilité suivant

$$\mathbb{P}\left[\overline{X}_n - F_N^{-1}(1-\alpha/2)\frac{\sqrt{\overline{X}_n}}{\sqrt{n}} \leq \theta \leq \overline{X}_n + F_N^{-1}(1-\alpha/2)\frac{\sqrt{\overline{X}_n}}{\sqrt{n}}\right] \approx 1-\alpha\,,$$

  l'approximation étant d'autant meilleure que n est grand.
  NB: $F_N^{-1}(1-\alpha/2)$ désigne le quantile d'ordre $(1-\alpha/2)$ de la loi gaussienne standard notée $\mathcal{N}(0,1)$.

- Il y a une probabilité égale environ à $(1-\alpha)$ pour que l'intervalle de confiance aléatoire contienne la moyenne de la population. De manière à peu près équivalente, sur 100 réalisations de l'intervalle de confiance, le paramètre inconnu sera effectivement dans la fourchette proposée dans $(1-\alpha)100\%$ des cas.

• Intervalle de confiance bilatéral symétrique pour $\theta = \mathbb{E}[X]$ pour un échantillon i.i.d. de loi de Poisson de taille quelconque:

- L'**intervalle de confiance** bilatéral pour θ au niveau de confiance $(1-\alpha)$ est issu de l'encadrement en probabilité suivant

$$\mathbb{P}\left[\left(-\frac{1}{\sqrt{4n\alpha}} + \sqrt{\frac{1}{4n\alpha} + \overline{X}_n}\right)^2 \leq \theta \leq \left(\frac{1}{\sqrt{4n\alpha}} + \sqrt{\frac{1}{4n\alpha} + \overline{X}_n}\right)^2\right] \geq 1-\alpha\,.$$

- Il y a une probabilité d'au moins $(1-\alpha)$ pour que l'intervalle de confiance aléatoire contienne la moyenne de la population. De manière à peu près équivalente, sur 100 réalisations de l'intervalle de confiance, le paramètre inconnu sera effectivement dans la fourchette proposée dans $(1-\alpha)100\%$ des cas.

• Illustration numérique:
Elaborer un plan de simulations explicitant le nombre d'échantillons générés noté M, la taille requise pour l'échantillon que l'on a noté n et les différentes lois utilisées pour simuler les échantillons. Vous étudierez la probabilité de couverture (définie comme étant la proportion de fois où le paramètre inconnu à estimer appartient effectivement à l'intervalle de confiance proposé) et la largeur de l'intervalle proposé de façon à illustrer les éléments présentés ci-dessus.

On utilisera les valeurs $M = 1000$ et $n = 20, 50, 100, 200, 500$.

Le travail à effectuer peut se mettre sous la forme de l'algorithme suivant:
• choisir une loi de probabilité paramétrée par θ (et éventuellement d'autres paramètres), notée P_θ, pour une valeur fixée de θ parmi l'ensemble des valeurs admissibles,
• pour $n = 20, 50, 100, 200, 500$,

- pour $m = 1, ..., M$,
  - simuler un échantillon i.i.d. $(X_1^{\{m\}}, ..., X_n^{\{m\}})$ selon P_θ,

  - calculer les bornes inférieure $\widehat{\theta}_{n,\inf}^{\{m\}}$ et supérieure $\widehat{\theta}_{n,\sup}^{\{m\}}$ de l'intervalle de confiance considéré,
  - déterminer la largeur de l'intervalle de confiance ainsi obtenu,
  - déterminer si la vraie valeur du paramètre θ appartient ou non à l'intervalle de confiance,

- déterminer la probabilité de couverture (définie comme étant la proportion de fois où le paramètre inconnu à estimer appartient effectivement à l'intervalle de confiance proposé) et la largeur moyenne de l'intervalle.