Section : Intervalles de confiance
Précédent : Définitions
Suivant : Modèle linéaire

Echantillons gaussiens

Ce paragraphe est consacré à la construction d'intervalles de confiance de la moyenne et la variance, pour les échantillons gaussiens, autrement dit les échantillons de la loi normale $\mathcal{N}(\mu,\sigma^2)$. L'avantage de cette situation est que les estimateurs naturels de l'espérance et de la variance ont des lois explicitement calculables. Nous notons $(X_1,\dots,X_n)$ un échantillon de la loi $\mathcal{N}(\mu,\sigma^2)$, $\overline{X}$ sa moyenne empirique et S^2 sa variance empirique .

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad S^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 .$$

Nous rassemblons ci-dessous, et nous admettrons, les trois résultats permettant de calculer les intervalles de confiance de μ et σ^2.

Théorème 3.3 Si $(X_1,\dots,X_n)$ est un échantillon de la loi $\mathcal{N}(\mu,\sigma^2)$, alors :

1. $\sqrt{\frac{n}{\sigma^2}}\left(\overline{X}-\mu\right)$ suit la loi normale $\mathcal{N}(0,1)$.
2. $\sqrt{\frac{n-1}{S^2}}\left(\overline{X}-\mu\right)$ suit la loi de Student $\mathcal{T}(n-1)$.
3. $\frac{nS^2}{\sigma^2}$ suit la loi du chi-deux $\mathcal{X}^2(n-1)$.

Les deux premières affirmations servent à estimer l'espérance μ, respectivement dans le cas où la variance σ^2 est connue et dans le cas où elle est inconnue. Commençons par supposer que σ^2 est connue. Posons $z_\alpha = Q_{\mathcal{N}(0,1)}(1-\alpha/2)$. L'intervalle de dispersion optimal de niveau $1-\alpha$ pour la loi $\mathcal{N}(0,1)$ est $[-z_\alpha, z_\alpha]$. Deux valeurs de z_α sont très souvent utilisées : pour $1-\alpha=$0.95 et 0.99, z_α vaut respectivement 1.96 et 2.5758. D'après le point 1 du théorème 3.3, on a :

$$\mathbb{P}\left[\sqrt{\frac{n}{\sigma^2}}\left(\overline{X}-\mu\right) \in [-z_\alpha, z_\alpha]\right] = 1-\alpha \; .$$

Or :

$$\begin{aligned} \sqrt{\tfrac{n}{\sigma^2}}\left(\overline{X}-\mu\right) \in [-z_\alpha, z_\alpha] \iff & \quad \overline{X}-\mu \in \left[-z_\alpha\sqrt{\tfrac{\sigma^2}{n}},\, z_\alpha\sqrt{\tfrac{\sigma^2}{n}}\right] \\ \iff & \quad \mu \in \left[\overline{X}-z_\alpha\sqrt{\tfrac{\sigma^2}{n}},\, \overline{X}+z_\alpha\sqrt{\tfrac{\sigma^2}{n}}\right] \; . \end{aligned}$$

L'intervalle :

$$\left[\overline{X}-z_\alpha\sqrt{\frac{\sigma^2}{n}},\, \overline{X}+z_\alpha\sqrt{\frac{\sigma^2}{n}}\right] \; ,$$

est donc un intervalle de confiance de niveau $1-\alpha$ pour μ.

Le cas où σ^2 est inconnu se traite de la même façon, en remplaçant la loi $\mathcal{N}(0,1)$ par la loi $\mathcal{T}(n-1)$. C'est encore une loi symétrique, pour laquelle l'intervalle de dispersion optimal de niveau $1-\alpha$ est de la forme $[-t_\alpha, t_\alpha]$, où :

$$t_\alpha = Q_{\mathcal{T}(n-1)}(1-\alpha/2) \; .$$

Le même raisonnement conduit à l' intervalle de confiance suivant pour μ :

$$\left[\overline{X}-t_\alpha\sqrt{\frac{S^2}{n-1}},\, \overline{X}+t_\alpha\sqrt{\frac{S^2}{n-1}}\right] \; .$$

Passons maintenant à l'estimation de σ^2 à partir de S^2. La loi du chi-deux $\mathcal{X}^2(n-1)$ n'est pas symétrique, et l'intervalle de dispersion symétrique n'est pas optimal. Nous noterons u_α et v_α deux réels positifs tels que $[u_\alpha, v_\alpha]$ soit un intervalle de dispersion de niveau $1-\alpha$ pour la loi $\mathcal{X}^2(n-1)$. On pourra calculer l'intervalle de dispersion optimal par une procédure d'optimisation numérique, ou bien prendre l'intervalle symétrique :

$$u_\alpha = Q_{\mathcal{X}^2(n-1)}(\alpha/2) \quad \text{et} \quad v_\alpha = Q_{\mathcal{X}^2(n-1)}(1-\alpha/2)\,.$$

D'après le point 3 du théorème 3.3, on a :

$$\mathbb{P}\left[\frac{nS^2}{\sigma^2} \in [u_\alpha\,,\,v_\alpha]\right] = 1-\alpha\,.$$

Or :

$$\frac{nS^2}{\sigma^2} \in [u_\alpha, v_\alpha] \iff \sigma^2 \in \left[\frac{nS^2}{v_\alpha}, \frac{nS^2}{u_\alpha}\right]\,.$$

L'intervalle $\left[\frac{nS^2}{v_\alpha}, \frac{nS^2}{u_\alpha}\right]$ est donc un intervalle de confiance de niveau $1-\alpha$ pour σ^2.

Section : Intervalles de confiance
Précédent : Définitions
Suivant : Modèle linéaire