Formulaire de statistiques

Jeffery Petit

Préambule Définitions et propriétés utiles à destination des étudiants de L1 Psycho à l'université de Nantes. Ce formulaire contient uniquement l'essentiel : pas de graphique, peu de blabla. Le formalisme utilisé, parfois plus complexe que le document de cours officiel, n'est pas indispensable. Ce document ne sera pas autorisé à l'examen.

Formules usuelles de probabilité

Les formules qui vont suivre s'appliquent à une variable $Z \sim \mathcal{N}(\mu = 0, \sigma = 1)$.

Propriété 1 (Symétrie de la loi normale) *On suppose que a est un nombre réel quelconque :*

$$\mathbb{P}(Z > a) = 1 - \mathbb{P}(Z < a) \tag{1}$$

$$\mathbb{P}(Z < -a) = \mathbb{P}(Z > a) \tag{2}$$

Propriété 2 (Probabilités des intervalles) *Supposons que b et c soient des nombres réels quelconques tels que $b < c$:*

$$\mathbb{P}(b < Z < c) = \mathbb{P}(Z < c) - \mathbb{P}(Z < b) \tag{3}$$

Par conséquent, si l'on suppose que d est un nombre réel positif :

$$\mathbb{P}(-d < Z < d) = 2 \times \mathbb{P}(Z < d) - 1 \tag{4}$$

Normalisation

Propriété 3 (Normalisation) *Le calcul de probabilté et de quantile nécessite de toujours se ramener au cas d'une varaible suivant une loi normale dite* ***centrée, réduite****, i.e. une loi $\mathcal{N}(0,1)$. Pour ce faire, on utilisera fréquemment la propriété suivante :*

$$X \sim \mathcal{N}(\mu, \sigma) \Rightarrow Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1) \tag{5}$$

Quantiles

Définition 1 (Notation de quantiles) *Soit X une variable aléatoire de loi continue, on appelle quantile d'ordre α, et on note q_α la valeur vérifiant :*

$$\mathbb{P}(X < q_\alpha) = \alpha \iff \mathbb{P}(X > q_\alpha) = 1 - \alpha \tag{6}$$

Par convention, lorsqu'on travaille sur une loi normale centrée réduite, e.g. $Z \sim \mathcal{N}(0,1)$, les quantiles sont notés z_α, i.e. $\mathbb{P}(Z < z_\alpha) = \alpha$

Propriété 4 (Symétrie des quantile pour la loi $\mathcal{N}(0,1)$) *Supposons que z_α soit le quantile d'ordre α de la loi $\mathcal{N}(0,1)$*

$$z_\alpha = -z_{1-\alpha} \iff z_{1-\alpha} = -z_\alpha \tag{7}$$

Propriété 5 (Calcul des quantiles pour une loi quelconque $\mathcal{N}(\mu,\sigma)$) *Supposons que q_α soit le quantile d'ordre α de la loi $\mathcal{N}(\mu,\sigma)$ et que z_α soit le quantile d'ordre α de la loi $\mathcal{N}(0,1)$. On a alors :*

$$q_\alpha = \mu + \sigma \times z_\alpha \tag{8}$$

Intervalles de confiance

Définition 2 (Définition d'un IC pour une moyenne μ ou une proportion p_0) *On appelle « Intervalle de Confiance (IC) de niveau $1-\alpha$ pour la moyenne μ » et on note $IC_{1-\alpha}(\mu)$ (resp. $IC_{1-\alpha}(p_0)$), un intervalle qui a une probabilité $1-\alpha$ de contenir la vraie valeur de μ (resp. p_0).*

$\longrightarrow$ *α est souvent appelé **seuil** ou **niveau de risque** de l'IC.*

De manière générale, un IC bilatéral pour une moyenne (resp. une proportion) sera définit ainsi $IC_{1-\alpha}(\mu) = [\bar{x} \pm \varepsilon_\alpha]$ (resp. $IC_{1-\alpha}(p_0) = [p \pm \varepsilon_\alpha]$) où ε_α sera la marge d'erreur à calculer pour un risque donné.

Propriété 6 (Formule des écart-types)

$$s = \sqrt{\frac{N}{N-1}} \times \sigma \tag{9}$$

IC bilatéral pour une moyenne μ

Dans ce qui suit on suppose que l'on dispose d'un échantillon dont la distribution est une loi normale. On note N l'effectif de cet échantillon, $\bar{x}$ la moyenne observée sur l'échantillon, σ l'écart-type d'observation et s l'écart-type d'échantillon.

Définition 3 (Marge d'erreur si $N < 30$) *La marge d'erreur pour un seuil α que l'on peut noter ε_α est la quantité suivante :*

$$\varepsilon_\alpha = \frac{s}{\sqrt{N}} \times t_{1-\frac{\alpha}{2}} \tag{10}$$

Où $t_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi de Student à $N-1$ degrés de liberté (ddl).

Propriété 7 (Condition sur N si $N < 30$) *Supposons que l'on cherche la valeur de N à partir de laquelle on obtient une marge d'erreur plus faible qu'un certain ε au seuil α. La condition que N doit alors vérifier est :*

$$N \geq (\frac{s}{\varepsilon} \times t_{1-\frac{\alpha}{2}})^2 \tag{11}$$

Définition 4 (IC si $N < 30$)

$$\begin{aligned} IC_{1-\alpha}(\mu) &= [\bar{x} \pm \varepsilon_\alpha] \\ &= [\bar{x} - \varepsilon_\alpha ; \bar{x} + \varepsilon_\alpha] \\ &= [\bar{x} - \tfrac{s}{\sqrt{N}} \times t_{1-\frac{\alpha}{2}} ; \bar{x} + \tfrac{s}{\sqrt{N}} \times t_{1-\frac{\alpha}{2}}] \end{aligned} \tag{12}$$

Définition 5 (Marge d'erreur si $N \geq 30$) *La marge d'erreur pour un seuil α que l'on peut noter ε_α est la quantité suivante :*

$$\varepsilon_\alpha = \frac{s}{\sqrt{N}} \times z_{1-\frac{\alpha}{2}} \tag{13}$$

Où $z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$.

Propriété 8 (Condition sur N si $N \geq 30$) *Supposons que l'on cherche la valeur de N à partir de laquelle on obtient une marge d'erreur plus faible qu'un certain ε au seuil α. La condition que N doit alors vérifier est :*

$$N \geq (\frac{s}{\varepsilon} \times z_{1-\frac{\alpha}{2}})^2 \tag{14}$$

Définition 6 (IC si $N \geq 30$)

$$\begin{aligned} IC_{1-\alpha}(\mu) &= [\bar{x} \pm \varepsilon_\alpha] \\ &= [\bar{x} - \varepsilon_\alpha ; \bar{x} + \varepsilon_\alpha] \\ &= [\bar{x} - \tfrac{s}{\sqrt{N}} \times z_{1-\frac{\alpha}{2}} ; \bar{x} + \tfrac{s}{\sqrt{N}} \times z_{1-\frac{\alpha}{2}}] \end{aligned} \tag{15}$$

IC bilatéral pour une proportion p_0

Dans ce qui suit on note p la proportion observée dans l'échantillon.

Conditions indispensables

$$N \geq 30$$

$$N \times p \geq 5$$

$$N \times (1-p) \geq 5$$

Définition 7 (Marge d'erreur pour une proportion) *La marge d'erreur pour un seuil α que l'on peut noter ε_α est la quantité suivante :*

$$\varepsilon_\alpha = \sqrt{\frac{p(1-p)}{N}} \times z_{1-\frac{\alpha}{2}} \tag{16}$$

Où $z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$.

Propriété 9 (Condition sur N pour une proportion) *Supposons que l'on cherche la valeur de N à partir de laquelle on obtient une marge d'erreur plus faible qu'un certain ε au seuil α. La condition que N doit alors vérifier est :*

$$N \geq (\frac{z_{1-\frac{\alpha}{2}}}{\varepsilon})^2 \times p(1-p) \tag{17}$$

Définition 8 (IC pour une proportion)

$$\begin{aligned} IC_{1-\alpha}(p_0) &= [p \pm \varepsilon_\alpha] \\ &= [p - \varepsilon_\alpha ; \bar{x} + \varepsilon_\alpha] \\ &= [p - \sqrt{\tfrac{p(1-p)}{N}} \times z_{1-\frac{\alpha}{2}} ; p + \sqrt{\tfrac{p(1-p)}{N}} \times z_{1-\frac{\alpha}{2}}] \end{aligned} \tag{18}$$

Remarque : attention, l'IC s'exprime soit en pourcentage, soit en décimal. Par exemple, pour une proportion de 52.94% soit on calcul l'IC en posant $p = 52.94$, soit en posant $p = 0.5294$ mais il ne faut pas mélanger les deux !

Jeffery Petit
Laboratoire des Sciences du Numérique de Nantes (LS2N)
École Centrale Nantes
jeffery.petit@ls2n.fr