L2Mass Stat -- 2018-2019

Questions types pour la seconde interrogation.

Paramètre d'une loi, Estimateurs, tests d'hypothèse

Question de cours. Donner une liste de paramètres numériques de la loi uniforme sur un segment.

De combien de paramètres à t-on besoin pour caractériser une telle loi ?

Question de cours. Une expérience aléatoire donne comme résultat une valeur X dont la loi dépend d'un paramètre θ.

Qu'est ce qu'un estimateur $\hat{\theta}$ de θ ? Quel usage en fait on ?

En quoi la valeur prise par $\hat{\theta}$ en un échantillon $(x_1, \ldots, x_n)$ est elle aléatoire ? Cette valeur dépend elle de θ ? En quoi $\hat{\theta}$ dépend il de θ ?

Question de cours. On teste une hypothèse $\mathcal{H}$ par une "statistique de test" T en calculant la valeur de T en un échantillon $(x_1, \ldots, x_n)$. A quelle condition sur cette valeur rejettera t-on $\mathcal{H}$ au seuil 10% ?

Exercice. On veut estimer la proportion p d'opinions favorables (sur quelque chose) dans une population en interrogeant 100 personnes choisies au hasard dans la population. On obtient un ensemble de réponses qu'on code par un 100-uplet $(x_1, \ldots, x_{100})$ à valeurs dans $\{0, 1\}$. 52 des x_i vallent 1 (autrement dit 52 personnes dans l'échantillon ont une opinion favorable). On espère pouvoir conclure que $p > \frac{1}{2}$

Proposer une statistique de test pour l'hypothèse $p = \frac{1}{2}$. A quel seuil peut on rejeter cette hypothèse ?

Qu'aurait du être la taille de l'échantillon pour pouvoir rejeter l'hypothèse $p = \frac{1}{2}$ au seuil 1% (à supposer que la proportion d'opinions favorables dans l'échantillon soit toujours 52%) ? Rejeterait on alors l'hypothèse $p \leq \frac{1}{2}$?

Réponse. On prend comme statistique de test $\hat{p} = \text{proportion d'opinion favorable dans l'échantillon} = \frac{1}{n} \times \sum_i x_i$.

La valeur de $\hat{p}$ en la valeur observée est 0.52. On rejette l'hypothèse $p = \frac{1}{2}$ au seuil $1 - \mathrm{F}_{\hat{p}|p=\frac{1}{2}}(0.52)$ (la probabilité sous l'hypothèse $p = \frac{1}{2}$ qu'on ait $\hat{p} > 0.52$ puisque la valeur centrale de $\hat{p}$ est 0.5).

Sous l'hypothèse $p = \frac{1}{2}$ la variable $100\hat{p}$ suit la loi binomiale $\mathcal{B}(100, \frac{1}{2})$.

Comme $100 \times \frac{1}{2}$ et $100 \times (1 - \frac{1}{2})$ sont suffisamment grands on peut approximer la loi de $\frac{\hat{p}-\mathrm{E}(\hat{p})}{\sigma(\hat{p})}$ par la loi gaussienne $\mathcal{N}(0,1)$ et alors

$$\mathrm{F}_{\hat{p}}(0.52) = \mathrm{F}_{\frac{\hat{p}-\mathrm{E}(\hat{p})}{\sigma(\hat{p})}}\big((0.52 - \mathrm{E}(\hat{p}))/\sigma(\hat{p})\big) \approx \mathrm{F}_{\mathcal{N}(0,1)}\big((0.52 - \mathrm{E}(\hat{p}))/\sigma(\hat{p})\big)$$

On a :

$\mathrm{E}(100\hat{p}) = 100 \times \frac{1}{2} = 50$ donc $\mathrm{E}(\hat{p}) = .5$

$\mathrm{Var}(100\hat{p}) = 100\frac{1}{2}(1 - \frac{1}{2}) = 25$ donc $\mathrm{Var}(\hat{p}) = 25/100^2 = 0.05^2$

$\mathrm{F}_{\mathcal{N}(0,1)}\big((0.52 - \mathrm{E}(\hat{p}))/\sigma(\hat{p})\big) = \mathrm{F}_{\mathcal{N}(0,1)}(0.02/0.05) = \mathrm{F}_{\mathcal{N}(0,1)}(0.4) \approx 0.58$.

On rejetterais donc $p = \frac{1}{2}$ au seuil 0.42 ce qui n'est pas raisonnable : ce seuil est beaucoup trop grand.

Le 99-ème centile de la loi $\mathcal{N}(0,1)$ vaut approximativement 2.33. Pour rejeter $p = \frac{1}{2}$ au seuil 0.01 il faut (après approximation gaussienne) $\frac{0.02}{\sigma(\hat{p})} \geq 2.33$, or $\sigma(\hat{p}) = \frac{1}{2\sqrt{n}}$ d'où la condition $\sqrt{n} \geq \frac{2.33}{0.04} = 58.25$, $n \geq 3393 \approx 3400$.

On observe que $1 - \mathrm{F}_{\hat{p}}(0.52)$ est une fonction croissante de p pour $p \leq 0.52$ donc si on rejette $p = \frac{1}{2}$ on rejette a fortiori toute valeur de p inférieure à $\frac{1}{2}$.

Exercice. Une machine-outil produit des tiges de longueur m mm avec un écart aléatoire ϵ qui suit une loi normale centrée en 0 d'écart type 5.

La moyenne des longueurs observées sur un échantillon de 50 tiges produites par la machine est de 102 mm.

Cette observation permet elle de conclure $m > 100$? (A quel seuil ?)

Réponse. Après approximation gaussienne (TCL) la moyenne des longueurs observées suit la loi $\mathcal{N}(m, \frac{5^2}{50})$ (la loi de la moyenne de 50 va indépendantes de loi $\mathcal{N}(m, 5^2)$). La probabilité qu'elle soit supérieure à 102 vaut $1 - \mathrm{F}_{\mathcal{N}(m,\frac{1}{2})}(102) = 1 - \mathrm{F}_{\mathcal{N}(0,1)}(\frac{102-m}{1/\sqrt{2}})$.

Pour $m = 100$ cela vaut $1 - \mathrm{F}_{\mathcal{N}(0,1)}(2.83) \approx 0.002$. On peut donc rejeter $m = 100$ au seuil $2/1000$, a forciori on peut rejeter $m < 100$ au même seuil.

Exercice. Une machine-outil produit des tiges de longueur m mm avec un écart aléatoire ϵ qui suit une loi centrée en 0.

La machine produit 10 premières tiges dont les longueurs sont

$$99.7, 100.1, 99.5, 99, 99.6, 99.2, 100.2, 99.7, 99.4, 99.8$$

La proportion de longueurs inférieures à 100 permet elle de conclure $m < 100$? (A quel seuil ?)

* Aurait on pu conclure en estimant la moyenne des longueurs des tiges ?

Réponse. Puisque la loi est centrée le nombre de tiges de longueur $\leq m$ suit la loi binomiale $\mathcal{B}(10, \frac{1}{2})$. Le nombre 8 est atypique : la probabilité que le nombre soit au moins 8 vaut $\left(\binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right) \times \left(\frac{1}{4}\right)^{10} = \frac{45+10+1}{2^10} = \frac{7}{2^7} \approx 0.055$. Au seuil 5.5% on peut donc rejeter $m = 100$, a forciori $m > 100$.