Université Paris-Saclay-Faculté de médecine
M2R santé publique
Démarche statistique

Session du 10 janvier 2022 à 09h30.

I. Modèle statistique

Pour un certain village africain, les données disponibles suggèrent fortement que le nombre attendu de nouveaux cas de SIDA se développant au cours d'une année donnée est directement proportionnel au nombre attendu de nouveaux cas de SIDA qui se sont développés au cours de l'année précédente. Un objectif statistique important consiste à estimer la valeur de cette constante de proportionnalité inconnue $\theta(\theta > 1)$ qui est supposée ne pas varier d'une année sur l'autre, et ensuite de calculer un intervalle de confiance à 95%.

Pour atteindre cet objectif, il convient d'utiliser le modèle statistique suivant : Pour $j = 0, 1, \ldots, n$ années consécutives de données, Y_j est la variable aléatoire désignant le nombre de nouveaux cas de SIDA apparus au cours de l'année j. Supposons en outre que les $(n + 1)$ variables aléatoires $Y_0, Y_1, \ldots, Y_n$ sont telles que la loi conditionnelle de Y_{j+1} sachant toutes les années précédentes $Y_k = y_k$ pour $k = 0, 1, \ldots, j$, **ne dépend que** de y_j est une loi de Poisson avec $\mathbb{E}\big(Y_{j+1} \mid Y_j = y_j\big) = \theta y_j, j = 0, 1, \ldots, (n - 1)$. De plus, la distribution de la variable aléatoire Y_0 est une loi de Poisson avec $\mathbb{E}\big(Y_0\big) = \theta$, où $\theta > 1$.

On note $\mathcal{L}_n \equiv \mathcal{L}\big(\mathbf{y}; \theta\big)$ la vraisemblance associée au jeu de données observées $\mathbf{y} = (y_0, y_1, \ldots, y_n)$.

1. **(1 point)** À l'aide de la formule de Bayes, écrire la vraisemblance $\mathcal{L}\big(\mathbf{y}; \theta\big)$ en fonction de la fonction de masse de Y_0 qu'on notera $p_{Y_0}(y_0, \theta)$ et les fonctions de masse conditionnelles $p_{Y_{j+1}}\big(y_{j+1} \mid Y_j = y_j, Y_{j-1} = y_{j-1}, \ldots, Y_1 = y_1, Y_0 = y_0\big), \quad j = 0, 1, \ldots n - 1$.

2. **(1 point)** Déduire que

$$\mathcal{L}\big(\mathbf{y}; \theta\big) = \left(\frac{\theta^{y_0} e^{-\theta}}{y_0!}\right) \prod_{j=0}^{n-1} \frac{(\theta y_j)^{y_{j+1}} e^{-\theta y_j}}{y_{j+1}!}.$$

3. **(10 points)** Calculer l'estimateur par maximum de vraisemblance de θ qu'on notera $\widehat{\theta}_n$.

4. **(3 points)** Sachant que $\mathbb{E}\big(Y_{j+1}\big) = \mathbb{E}_{y_j}\Big[\mathbb{E}\big(Y_{j+1} \mid Y_j = y_j\big)\Big]$, déduire que

$$-\mathbb{E}\left(\frac{\partial^2 \ln\left(\mathcal{L}_n\right)}{\partial \theta^2}\right) = \theta^{-1}\left(\frac{1 - \theta^{(n+1)}}{1 - \theta}\right).$$

5. **(3 points)** Si $n = 25$ et $\widehat{\theta}_n = 1.2$, proposer un intervalle de confiance à 95% de θ.

II. Calculs classiques

Soient $X_1, X_2, \ldots, X_n$ un **échantillon** de variables aléatoires d'une loi uniforme de densité

$$f(x) = 1, \quad 0 < x < 1.$$

On note $G_n = \left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}}$ la moyenne dite géométrique des $X_1, \ldots, X_n$.

1. (1 point) Calculer $\mathbb{E}\big(X^r\big)$ et déduire les expressions explicites de $\mathbb{E}\big(G_n\big)$ et $\mathbb{E}\big(G_n^2\big)$ respectivement.

2. (1 point) Examiner la convergence en probabilité de G_n.