

Lois classiques et estimation

Loi multinomiale.

Exercice 1. Soient $Y_1, \dots, Y_n$ des variables aléatoires i.i.d. à valeurs dans un ensemble $\mathcal{Y}$ muni d'une tribu $\mathcal{A}$. On considère une partition $(A_1, A_2, \dots, A_K)$ de $\mathcal{Y}$ où $A_k \in \mathcal{A}$ pour $k = 1, \dots, K$. On pose pour $k = 1, \dots, K$:

$$p_k = P(Y_1 \in A_k), \quad N_k = \sum_{i=1}^n 1_{\{Y_i \in A_k\}}$$

- (a) Calculer $\sum_{k=1}^K p_k$ et $\sum_{k=1}^K N_k$.
- (b) Quelle est la loi de N_k ?
- (c) Calculer $\text{Cov}(1_{\{Y_1 \in A_k\}}, 1_{\{Y_1 \in A_\ell\}})$. En déduire $\text{Cov}(N_k, N_\ell)$.
- (d) On considère K boîtes numérotées de 1 à K et n boules numérotées de 1 à n . Soit $(n_1, \dots, n_K) \in \mathbb{N}^K$ tel que $n_1 + \dots + n_K = n$. Combien il y a-t-il de façon de répartir les n boules dans les K boîtes de sorte que pour $k = 1, \dots, K$ la boîte numéro k contienne n_k boules exactement ?
- (e) Soit $(n_1, \dots, n_K) \in \mathbb{N}^K$ tel que $n_1 + \dots + n_K = n$. On note $\mathcal{F}$ l'ensemble des fonctions $\varphi : \{1, \dots, n\} \rightarrow \{1, \dots, K\}$ telles que $\text{card}(\varphi^{-1}(\{k\})) = n_k$ pour $k = 1, \dots, K$. Montrer que

$$\{N_1 = n_1, \dots, N_K = n_K\} = \bigcup_{\varphi \in \mathcal{F}} \bigcap_{i=1}^n \{X_i \in A_{\varphi(i)}\}.$$

En déduire que

$$P(N_1 = n_1, \dots, N_K = n_K) = \frac{n!}{n_1! \dots n_K!} p_1^{n_1} \dots p_K^{n_K}.$$

- (f) Dans le cas où $(p_1, \dots, p_K)$ est inconnu et que l'on observe $Y_1, \dots, Y_n$, montrer que l'estimateur du maximum de vraisemblance de $(p_1, \dots, p_K)$ est $(\hat{p}_1, \dots, \hat{p}_K)$ où $\hat{p}_k = N_k/n$.

Loi de Poisson.

Exercice 2. La qualité de l'eau d'une pisciculture est considérée mauvaise si la concentration θ en bactéries *E. coli* est plus grande que 2000 pour 100 millilitres. Pour l'évaluer, on effectue n prélèvements de 100 millilitres et pour chacun des prélèvements on évalue le nombre de bactéries qu'il contient. Soit $X_1, \dots, X_n$ les mesures obtenues. On adopte le modèle suivant : $X_1, \dots, X_n$ est un échantillon de la loi de Poisson de paramètre θ .

- (a) Déterminer l'estimateur du maximum de vraisemblance de θ . On le note $\hat{\theta}_n$ dans la suite.
- (b) Calculer l'espérance et la variance de $\hat{\theta}_n$.
- (c) Montre que pour tout $\varepsilon > 0$ on a

$$P(|\hat{\theta}_n - \theta| > \varepsilon \sqrt{\theta}) \leq \frac{1}{n\varepsilon^2}.$$

En déduire un intervalle de confiance pour θ de niveau d'erreur α pour $\alpha \in]0, 1[$.

- (d) Déterminer une statistique φ à valeurs dans $\{0, 1\}$ telle que $P(\varphi = 1) \leq \alpha$ si la qualité de l'eau est mauvaise et $P(\varphi = 1)$ soit le plus grand possible si la qualité de l'eau est bonne.

Loi normale.

Exercice 3. Dans l'atmosphère, le taux d'un gaz nocif est égal à μ . Pour l'estimer, on effectue n mesures conduisant aux valeurs $X_1, X_2, \dots, X_n$. On adopte le modèle suivant :

$$X_i = \mu + \varepsilon_i, \quad 1 \leq i \leq n$$

où $\varepsilon_1, \dots, \varepsilon_n$ sont des variables aléatoires i.i.d. de loi $\mathcal{N}(0, \sigma^2)$.

- (a) Quelle est la loi de $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$?
- (b) Construire un intervalle de confiance pour μ de niveau de confiance 95 % si la variance σ^2 est connue.
- (c) Construire un intervalle de confiance asymptotique pour μ de niveau de confiance 95 % si la variance σ^2 est inconnue.

Exercice 4. Soient X et Z deux variables aléatoires indépendantes de même loi $N(0, 1)$. On considère la variable aléatoire

$$Y = \rho X + \sqrt{1 - \rho^2} Z$$

où $\rho \in [-1, 1]$ est un paramètre inconnu. On observe n couples aléatoires $(X_1, Y_1), \dots, (X_n, Y_n)$ i.i.d. et tous de même loi que (X, Y) .

- (a) Calculer $E(XY)$ et $\text{Var}(XY)$. (On rappelle que $E(Z^4) = 3$.)
- (b) Soit T_n l'estimateur de ρ défini par

$$T_n = \frac{1}{n} \sum_{i=1}^n X_i Y_i.$$

Déterminer le risque quadratique de T_n .

- (c) On suppose que le paramètre inconnu ρ appartient à l'intervalle $[2 - \sqrt{3}, 1]$.
Calculer $E[(X - Y)^2]$ et $\text{Var}((X - Y)^2)$.
En déduire la construction d'un estimateur S_n de ρ tel que S_n soit strictement meilleur que T_n au sens du risque quadratique.