Exercices, 4

Exercice 1 (Loi de Student). — Soit X une variable aléatoire gaussienne centrée réduite, indépendante de la suite de variables aléatoires (Y_n) où Y_n suit la loi $\chi^2(n)$. On pose $T_n = \frac{X}{\sqrt{\frac{Y_n}{n}}}$.

1. Montrer que la densité t_n de T_n est donnée par
$$t_n(x) = \frac{1}{\sqrt{\pi}} \frac{\Gamma((n+1)/2)}{\sqrt{n}\,\Gamma(n/2)} \frac{1}{(1+x^2/n)^{(n+1)/2}}.$$
2. Montrer que pour tout $x \in \mathbb{R}$,
$$\lim_{n\to\infty} t_n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$
On pourra tout d'abord montrer que $\Gamma(a+1/2)/\Gamma(a) \sim \sqrt{a}$ quand $a \to \infty$ en remarquant que $\Gamma(a+1/2)/\Gamma(a) = \mathbb{E}[\sqrt{U}]$ où U suit une loi $\Gamma(a,1)$.
3. Montrer que (T_n) converge en loi et identifier sa limite.

Exercice 2. — Soit $(X_1,\ldots,X_n)$ un n-échantillon de la loi $\mathcal{N}(m,\sigma^2)$. Dans chacun des cas suivants, proposer des estimateurs et déterminer des intervalles de confiance de niveau α pour la moyenne et/ou la variance : (1) la variance seule est connue ; (2) la moyenne seule est connue ; (3) moyenne et variance sont connues.

Exercice 3. — On s'intéresse au taux de chloramines (en mg.L^{-1}) dans l'eau d'une piscine. En première approximation, on considère que cette quantité suit une loi normale d'espérance μ et de variance σ^2. On effectue n prélèvements, conduisant aux valeurs de $X_1, X_2, \ldots, X_n$. Sur un échantillon de taille $n = 10$, on observe que $\overline{x}_{10} = 0,5$ et $s_{10}^2 = 10^{-2}$, où $s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \overline{x}_{10})^2$.

1. Quel est l'intervalle de confiance de niveau de confiance 95% du taux moyen μ de chloramines si la variance σ^2 est supposée connue (on la prendra égale à 10^{-2}).
2. En réalité, on ne connaît pas la variance σ^2. Quel est le nouvel intervalle de confiance obtenu ?
3. Pour éviter tout risque pour la santé, le taux de chloramines doit être inférieur à $0,6$ mg.L^{-1}. Que penser de la qualité de l'eau de la piscine ?
4. Construire un intervalle de confiance pour σ^2 de niveau de confiance 95%.

Exercice 4. — Soit $(X_1,\ldots,X_n)$ un n-échantillon de la loi $\mathcal{N}(m,\sigma^2)$. Les paramètres m et $\sigma^2 > 0$ sont inconnus.

1. Construire un intervalle de confiance de σ^2 au niveau de risque $\alpha \in]0,1[$.
2. Donner un intervalle de confiance asymptotique au niveau α et calculer sa longueur moyenne ℓ_n.
3. Montrer que ℓ_n est équivalent à $C_\alpha \sigma^2/\sqrt{n}$ quand $n \to +\infty$.

Exercice 5. — Soient Y_1, Y_2, Y_3, Y_4 des variables aléatoires gaussiennes indépendantes telles que $\mathbb{E}[Y_1] = \mathbb{E}[Y_2] = \mu_1$, $\mathbb{E}[Y_3] = \mu_2$, $\mathbb{E}[Y_4] = \mu_3$, $\mathrm{Var}(Y_1) = \mathrm{Var}(Y_3) = \sigma^2$, $\mathrm{Var}(Y_2) = \sigma^2/3$, $\mathrm{Var}(Y_4) = \sigma^2/2$; μ_1, μ_2, μ_3 et σ^2 sont inconnus et $\sigma^2 > 0$.

1. Ecrire le modèle linéaire correspondant.

2. Déterminer les estimateurs des moindres carrés "standard" $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3$ et $\hat{\sigma}^2$. Quelle est la loi de $\hat{\mu}_1$? A-t-on $\hat{\mu}_1$ indépendante de $\hat{\sigma}^2$?
3. Déterminer les estimateurs du maximum de vraisemblance (ou moindres carrés généralisés) $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3$ et $\hat{\sigma}^2$. Quelle est la loi de $\hat{\mu}_1$? A-t-on $\hat{\mu}_1$ indépendante de $\hat{\sigma}^2$?
4. Déterminer un intervalle de confiance pour μ_1 de niveau α.
5. On suppose maintenant que μ_2 et μ_3 sont égaux à μ_1.
   (a) Quel est le modèle linéaire associé?
   (b) Déterminer l'estimateur $\widetilde{\mu}_1$ des moindres carrés généralisés de μ_1. Comparer les performances de $\hat{\mu}_1$ et de $\widetilde{\mu}_1$.
6. On utilise $\widetilde{\mu}_1$ pour estimer μ_1 dans le premier modèle.
   (a) Calculer son erreur quadratique moyenne.
   (b) Montrer que si $|\mu_2 + 2\mu_3 - 3\mu_1|$ est inférieur à un seuil que l'on calculera, l'erreur quadratique moyenne de $\widetilde{\mu}_1$ est inférieure à celle de $\hat{\mu}_1$.

Exercice 6 (Intervalle de prévision). — On considère le modèle de régression linéaire

$$Y_i = b_0 + b_1 x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

où les ε_i sont des variables aléatoires indépendantes $\mathcal{N}(0, \sigma^2)$ et b_0, b_1 et σ^2 sont inconnus.

1. Quels sont les estimateurs des moindres carrés ordinaires $\hat{b}_0, \hat{b}_1$ et $\hat{\sigma}^2$ de ces paramètres ? Quelle est la loi du couple $((\hat{b}_0, \hat{b}_1), \hat{\sigma}^2)$?
2. On dispose d'une observation y_0 sur une unité statistique pour laquelle la valeur de x_0 de la variable explicative est inconnue. On suppose que y_0 est la réalisation d'une variable Y_0 s'écrivant

   $$Y_0 = b_0 + b_1 x_0 + \eta,$$

   où η est une variable aléatoire $\mathcal{N}(0, \sigma^2)$ indépendante du vecteur $(\varepsilon_1, \dots, \varepsilon_n)$. On cherche un intervalle de confiance pour x_0. On fera l'hypothèse supplémentaire que

   $$|x_0 - \overline{x}| \leq 1, \text{ où } \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

   (a) Quelle est la loi de $Y_0 - \hat{b}_0 - \hat{b}_1 x_0$?
   (b) En utilisant l'estimateur $\hat{\sigma}$ de σ, déterminer un intervalle de confiance I_1 de niveau α pour x_0.
3. On dispose maintenant de m observations $y_{01}, \dots, y_{0m}$ correspondant à la valeur x_0 inconnue; ce sont des observations de m variables aléatoires telles que

   $$Y_{0j} = b_0 + b_1 x_0 + \eta_j,$$

   où $(\eta_1, \dots, \eta_m)$ et $(\varepsilon_1, \dots, \varepsilon_n)$ sont indépendantes, et η_j suit la loi $\mathcal{N}(0, \sigma^2)$.
   (a) Montrer que

   $$\widetilde{\sigma}^2 = \frac{(n-2)\hat{\sigma}^2 + \sum_{j=1}^{m}(Y_{0j} - \overline{Y}_0)^2}{n+m-3},$$

   où $\overline{Y}_0 = \frac{1}{m}\sum_{j=1}^{m} Y_{0j}$ est un autre estimateur sans biais de σ^2. Quelle est sa loi ?
   (b) Quelle est la loi de $\overline{Y}_0 - \hat{b}_0 - \hat{b}_1 x_0$?
   (c) A l'aide de $\widetilde{\sigma}^2$ et de $\overline{Y}_0$, donner un intervalle de confiance I_2 pour x_0 de niveau α.
   (d) Aurait-on pu construire un intervalle de confiance I_3 pour x_0 à l'aide de $\hat{\sigma}^2$ et de $\overline{Y}_0$?