UNIVERSITÉ DE NANTES

Licence 3 Mathématiques

Statistique: TD

Ex 1. Soit $X_1, ..., X_n$ des v.a. réelles iid telles que $\mathrm{var}(X_i) = \sigma^2 < \infty$. On pose $\theta = \mathbb{E}(X_1)$ et $\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

1. Montrer que $\widehat{\theta}_n$ est un estimateur sans biais de θ.
2. Montrer qu'il est consistant en précisant sa vitesse de convergence.
3. Montrer qu'il converge également en moyenne quadratique.

Ex 2. Soit $X_1, ..., X_n$ iid de uniforme sur $[0, \theta]$.

1. Calculer l'estimateur $\widehat{\theta}_{MM}$ de θ par la méthode des moments.
2. Montrer que $\widetilde{\theta} := \max\{X_1, ..., X_n\}$ converge en probabilité (et même presque sûrement) vers θ.
3. Calculer la limite en loi de $n(\theta - \widetilde{\theta})$.
4. Comparer les biais et vitesses de convergence de $\widehat{\theta}_{MM}$ et $\widetilde{\theta}$. Lequel vous semble meilleur?

Ex 3. Soit $X_1, ..., X_n$ iid de densité $f_\theta(x) = \theta/x^{\theta+1}\mathbb{1}\{x > 1\}$ où $\theta \in \Theta$ est un paramètre inconnu.

1. On suppose $\Theta =]1, +\infty[$. Après avoir calculer l'espérance de X_1, proposer un estimateur de θ par la méthode des moments et montrer qu'il est consistant.
2. On suppose maintenant que $\Theta =]0, +\infty[$. L'estimateur précédent est-il toujours consistant?
3. Calculer la loi de $Y_i = 1/X_i$.
4. En utilisant l'échantillon $Y_1, ..., Y_n$ proposer un nouvel estimateur de θ par la méthode des moments.
5. Montrer qu'il est consistant pour $\Theta =]0, +\infty[$.
6. Ecrire les vraisemblances pour les échantillons $X = (X_1, ..., X_n)$ et $Y = (Y_1, ..., Y_n)$.
7. Justifier que l'estimateur du maximum de vraisemblance est le même pour les deux échantillons.
8. Montrer qu'il est consistant en précisant sa vitesse de convergence.

Ex 4. Soit $X_1, ..., X_n$ iid de loi log-normale de densité

$$f_\theta(x) = \frac{1}{x\sqrt{2\pi}} \exp\Big(-\frac{(\log(x) - \theta)^2}{2}\Big)\mathbb{1}\{x > 0\},$$

avec $\theta \in \mathbb{R}$.

1. Proposer l'estimateur $\widehat{\theta}_{MM}$ de θ par la méthode des moments, basé sur le moment d'ordre 1.
2. Montrer qu'il converge presque sûrement vers θ quand $n \to \infty$.
3. Calculer l'estimateur $\widehat{\theta}_{MV}$ du maximum de vraisemblance de θ et déterminer sa loi.

Ex 5. Soit $X_1, ..., X_n$ un échantillon de variables aléatoires réelles iid de fonction de répartition F.

1. Rappeler la définition de la fonction de répartition empirique F_n.
2. Soit $x \in \mathbb{R}$, quelle est la loi de $F_n(x)$?
3. Soient x, y deux réels, calculer la covariance $\mathrm{cov}(F_n(x), F_n(y))$.
4. Donner la loi asymptotique de $\sqrt{n}\big(F_n(x) - F(x)\big)$.

Ex 6. Soit $X_1, ..., X_n$ iid de fonction de répartition F. On veut estimer $F(x)^2$ pour $x \in \mathbb{R}$.

1. Calculer l'espérance de $F_n(x)^2$.
2. En déduire un estimateur sans biais de F^2 de la forme $\alpha_n F_n^2 + \beta_n F_n$ où α_n, β_n sont des réels.
3. Déterminer la limite en loi de $\sqrt{n}(F_n(x)^2 - F(x)^2)$ quand $n \to \infty$ (indication: écrire $F_n^2 = F^2 + 2F(F_n - F) + (F_n - F)^2$).

Ex 7. Soit X une variable aléatoire de fonction de répartition F quelconque.

1. Montrer que $F^-(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}$ est bien un quantile d'ordre α.
2. Donner un exemple de fonction de répartition pour laquelle la médiane $q_{0.5}$ n'est pas unique.
3. Donner un exemple pour lequel le premier quartile $q_{0.25}$ vérifie simultanément les inégalités strictes

$$\mathbb{P}(X \leq q_{0.25}) > 0.25 \quad \text{et} \quad \mathbb{P}(X \geq q_{0.25}) > 0.75.$$

Ex 8. Soit $U_1, ..., U_n$ un échantillon de variables aléatoires iid de loi uniforme sur $]0, 1[$. On s'intéresse au comportement asymptotique du quantile empirique d'ordre $\alpha \in]0, 1[$, $U_{(n_\alpha)}$ où on note $n_\alpha = \lceil n\alpha \rceil$.

1. Rappeler la densité de la k-ième statistique d'ordre $U_{(k)}$.
2. Soit $X_1, ..., X_{n+1}$ des variables aléatoires indépendantes de loi exponentielle de paramètre 1, montrer que pour tout $k = 1, ..., n+1$, $S_k = \sum_{i=1}^k X_i$ suit une loi Gamma $\Gamma(k, 1)$ de densité

$$\gamma_k(x) = \frac{x^{k-1}}{(k-1)!} e^{-x} \ , \ x > 0.$$

3. On admet le résultat suivant: Soient S, T deux v.a. positives indépendantes de densités f_S, f_T respectivement, la variable $X = T/S$ a pour densité

$$f_X(x) = \int_0^{+\infty} f_S(y) f_T(xy) y dy \ , \ x > 0.$$

   Montrer que S_k/S_{n+1} a même loi que $U_{(k)}$ (indication: poser $S = S_k$ et $T = S_{n+1} - S_k$).
4. Déterminer la limite en loi de $\sqrt{n_\alpha}\big(S_{n_\alpha}/n_\alpha - 1\big)$.
5. En remarquant que

$$\lim_{n\to\infty} \frac{n_\alpha}{n} = \alpha \quad \text{et} \quad \lim_{n\to\infty} \sqrt{n}\Big(\frac{n_\alpha}{n} - \alpha\Big) = 0,$$

   montrer d'après la question précédente que

$$\sqrt{n}\Big(\frac{S_{n_\alpha}}{n} - \alpha\Big) \xrightarrow[n\to\infty]{loi} \mathcal{N}(0, \alpha) \ \text{ et } \ \sqrt{n}\Big(\frac{S_{n+1} - S_{n_\alpha}}{n} - (1-\alpha)\Big) \xrightarrow[n\to\infty]{loi} \mathcal{N}(0, 1-\alpha).$$

6. En déduire que

$$\sqrt{n}\Big(\alpha \ \frac{S_{n+1} - S_{n_\alpha}}{n} - (1-\alpha) \ \frac{S_{n_\alpha}}{n}\Big) \xrightarrow[n\to\infty]{loi} \mathcal{N}\big(0, \alpha(1-\alpha)\big),$$

   puis que

$$Y_n := \sqrt{n}\Big(\alpha \ \frac{S_{n+1} - S_{n_\alpha}}{S_{n_\alpha}} - (1-\alpha)\Big) \xrightarrow[n\to\infty]{loi} \mathcal{N}\Big(0, \frac{1-\alpha}{\alpha}\Big).$$

7. En remarquant que $U_{(n_\alpha)} \overset{loi}{=} \alpha/(1+Y_n/\sqrt{n})$ (cf. question 3), montrer qu'il existe ξ_n compris entre 0 et $Y_n/\sqrt{n}$ tel que
$$U_{(n_\alpha)} = \alpha - \frac{\alpha}{(1+\xi_n)^2}\frac{Y_n}{\sqrt{n}}.$$

8. En conclure que
$$Z_n := \sqrt{n}\big(U_{(n_\alpha)} - \alpha\big) \xrightarrow[n\to\infty]{loi} \mathcal{N}(0, \alpha(1-\alpha)).$$

Ex 9. On a mesuré la consommation quotidienne en eau dans une ville sur une année. On suppose les 365 observations $X_1, ..., X_n$ issues de variables aléatoires indépendantes et de loi normale $\mathcal{N}(m, \sigma^2)$ avec m et σ^2 inconnus. On rappelle, dans le modèle Gaussien, les estimateurs
$$\hat{m} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{m})^2$$
sont indépendants et $(n-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n-1)$. On mesure sur l'échantillon une moyenne $\overline{X}_n = 18$ (en milliers de m^3) et un écart-type $\hat{\sigma} = 2$.

1. Construire un intervalle de confiance à 95% sur σ (indication: prendre respectivement 313 et 419 comme valeurs des quantiles d'ordre 0.025 et 0.975 de la loi $\chi^2(364)$).
2. Montrer que $g(m, X) = \sqrt{n}\ (\overline{X}_n - m)/\hat{\sigma}$ est pivotale et déterminer sa loi.
3. Construire un intervalle de confiance à 95% sur m (indication: le quantile d'ordre 0.975 de la loi $\mathcal{N}(0,1)$ vaut approximativement 1.96).
4. Déduire un intervalle de confiance à 95% sur la consommation mensuelle moyenne pour un mois de 30 jours.
5. On veut évaluer la consommation maximale attendue du 1er janvier prochain, que l'on note X_{n+1}. Montrer que
$$S = \sqrt{\frac{n+1}{n}}\frac{X_{n+1} - \overline{X}_n}{\hat{\sigma}}$$
suit une loi de Student $\mathcal{T}(n-1)$.
6. Déduire une majoration de X_{n+1} au risque $\alpha = 0.05$ de se tromper.

Ex 10. On s'intéresse à la proportion p de la population touchée par le syndrôme de dépression hivernale. Sur un échantillon de $n = 400$ personnes interrogées, 16 affirment en souffrir.

1. Décrire le modèle statistique.
2. Construire un intervalle de confiance asymptotique de niveau 95% sur p (prendre 1.96 comme quantile d'ordre 0.975 de la loi $\mathcal{N}(0,1)$).

Ex 11. Soit $X_1, ..., X_n$ iid de loi de Cauchy décentrée de densité
$$f_\theta(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}, x \in \mathbb{R},$$
où $\theta \in \mathbb{R}$ est le paramètre inconnu du modèle.

1. Peut-on construire un estimateur de θ par la méthode des moments?
2. Ecrire la vraisemblance du modèle (on remarquera qu'il n'existe pas de forme analytique simple pour l'estimateur du maximum de vraisemblance).
3. Montrer que θ est l'unique médiane de f_θ. En déduire un estimateur consistant (simple) de θ.
4. En utilisant la normalité asymptotique du quantile empirique, donner un intervalle de confiance asymptotique de niveau de confiance $1-\alpha$ sur θ.
5. Déduire une procédure pour tester l'hypothèse $\mathcal{H}_0 : \theta = 0$ contre $\mathcal{H}_1 : \theta \neq 0$ au niveau α asymptotiquement.

Ex 12. Un couple de 5 enfants aimerait savoir si la probabilité p d'avoir une fille à chaque naissance est égale à 1/2. On note $\widehat{p}$ la proportion de filles parmi les 5 enfants.

1. Décrire la forme de la région critique pour tester $\mathcal{H}_0 : p = 1/2$ contre $\mathcal{H}_1 : p \neq 1/2$.
2. Les 5 enfants sont des filles. Quelle est la probabilité de cet événement sous $\mathcal{H}_0$?
3. Calculer la p-value du test. Quelle est la conclusion au niveau $\alpha = 0.05$?
4. Quelle est la forme de la région critique si l'hypothèse alternative est $\mathcal{H}_1 : p > 1/2$?
5. Calculer la p-value du test et la conclusion au niveau $\alpha = 0.05$ dans ce cas.

Ex 13. Un brasseur teste un nouveau procédé de fermentation pour produire sa bière. Sur les 25 nouvelles bouteilles produites, le degré moyen d'alcool est de 6° avec un écart-type de 1°.

1. Sachant que la bière fait normalement 5.5°, décrire la procédure de test pour savoir si le nouveau procédé de fermentation modifie le taux d'alcool. Sous quelles hypothèses le test est-il exact? asymptotiquement exact?
2. Donner une expression de la p-value du test.
3. Pour des degré moyen et écart-type identiques, à partir de combien de bouteilles peut-il conclure (au niveau $\alpha = 0.05$) que le degré d'alcool a été modifié?

Ex 14. Soit $X_1, ..., X_n$ les réponses à un référendum ($X_i \in \{0, 1\}$) supposées issues de v.a. iid. On observe une proportion $\hat{p}_n = 0.56$ de "oui" parmi 144 personnes sondées. Peut-on conclure au risque $\alpha = 0.05$ que le "oui" l'emportera? (indication: faire le test de $p = 1/2$ contre $p > 1/2$ en utilisant 1.65 comme quantile d'ordre 0.95 de la loi $\mathcal{N}(0, 1)$).

Ex 15. (TP en python) Soient $X_1, ..., X_n$ iid de de densité

$$f_\theta(x) = \theta^2 x e^{-x/\theta}, \ x > 0$$

où $\theta > 0$ est un paramètre inconnu. On admet que l'espérance de X_1 vaut $m = 2\theta$ et que la variance est finie. Cette densité correspond à la densité d'une loi Gamma de paramètres `shape`=2 et `scale`=θ.

1. Générer un échantillon iid $X_1, ..., X_n$ de taille $n = 5$ avec pour vraie valeur du paramètre $\theta = 2$.
2. Construire (à la main) la statistique de test de Student pour l'hypothèse $\mathcal{H}_0 : \theta = 2$ contre $\mathcal{H}_1 : \theta \neq 2$ (indication: réécrire l'hypothèse $\mathcal{H}_0$ comme une condition sur m).
3. Comparer avec le résultat de la fonction `scipy.stats.ttest_1samp`.
4. On veut évaluer la probabilité de rejeter $\mathcal{H}_0$ si elle est vraie (l'erreur de première espèce) lorsqu'on se fixe un niveau asymptotique $\alpha = 0.05$. Reproduire l'expérience 10000 fois et calculer la fréquence de rejet de $\mathcal{H}_0$. Commenter.
5. Reprendre la question précédente pour $n = 50$. Commenter.
6. On veut évaluer la probabilité de rejeter $\mathcal{H}_0$ si elle est fausse (la puissance du test) au niveau $\alpha = 0.05$. Générer 10000 échantillons de taille $n = 5$ de loi $\Gamma(2, 1.5)$ et calculer la fréquence de rejet de $\mathcal{H}_0$. Recommencer pour $n = 50$ et commenter.
7. Représenter graphiquement la puissance du test pour $n = 50$ et θ variant dans l'intervalle $[1, 3]$.
8. En remarquant que sous $\mathcal{H}_0$, $S := \sum_{i=1}^n X_i$ suit une loi Gamma $\Gamma(2n, \theta)$, construire un test exact de niveau $\alpha = 0.05$ pour $\mathcal{H}_0$ (indication: soit γ_β le quantile d'ordre β de la loi $\Gamma(2n, \theta)$, si $\mathcal{H}_0$ est vraie, alors $\mathbb{P}(\gamma_{\alpha/2} < S < \gamma_{1-\alpha/2}) = 1 - \alpha$).
9. Comparer la puissance avec celle du test précédent sur l'intervalle $[1, 3]$.