M1 – Statistique – 2020/2021

Devoir maison :

1 Exercice 1

Nous considérons la famille de densités sur $\mathbb{R}^+$ donnée par

$$f_{\alpha,\beta}(x) = C(\alpha,\beta)\frac{1}{x^{3+\alpha}}\mathbb{1}_{[\beta,+\infty[}(x)$$

avec $\alpha > 0$ et $\beta > 0$. Nous admettrons que

$$\mathbb{E}\left[\ln X\right] = \frac{1}{2+\alpha} + \ln\beta \text{ et } \mathbb{V}\left[\ln X\right] = \frac{1}{(2+\alpha)^2}.$$

1.1. Calculer $C(\alpha,\beta)$, $\mathbb{E}[X]$ et $\mathbb{V}[X]$.

1.2. Nous considérons deux cas :

(I) α est connu et β est inconnu.

(II) β est connu et α est inconnu.

Nous répondrons aux deux questions suivantes pour les deux cas (I) et (II).

1.2.a. Utiliser $\mathbb{E}[X]$ pour calculer l'estimateur de la méthode des moments avec n observations i.i.d.

1.2.b. Toujours avec n observations i.i.d., calculer l'estimateur du maximum de vraisemblance.

1.3. Dans le cas (I) uniquement et si nous notons $\widehat{\beta}$ l'estimateur du maximum de vraisemblance calculé dans la question 2.b (cas (I)), donner sa fonction de répartition $F_{\widehat{\beta}}$ puis celle de $n\left(\widehat{\beta}-\beta^\star\right)$ où $\beta^\star$ est le vrai paramètre. En déduire la loi limite de $n\left(\widehat{\beta}-\beta^\star\right)$ lorsque n tend vers $+\infty$.

1.4. Dans le cas (II) uniquement, étudier la normalité asymptotique des deux estimateurs de α vus précédemment. Lequel préférez-vous ?

2 Exercice 2

Une chaîne de télévision a programmé une série de K émissions hebdomadaires, stables dans la grille de programmes, c'est-à-dire passant toujours le même jour de la semaine et à la même heure.

Pour suivre l'audience, nous disposons d'un panel de téléspectateurs indépendants entre eux, de taille n supposée élevée (de l'ordre de quelques milliers d'individus ; en pratique, $n = 6000$). Ce panel permet, pour l'émission numéro $k \in \{1, \dots, K\}$ de la série, de connaître la réalisation du nombre aléatoire N_k d'individus ayant regardé cette émission k.

Nous désignons par p_k la probabilité de regarder l'émission numéro k de la série ; il est supposé constant égal à $p \in]0;1[$ sur l'ensemble du panel et pendant les K émissions de la série.

2.1. Qu'est-ce que cela implique comme hypothèse par rapport au fait de choisir des p_k potentiellement différents ?

2.2. Nous commençons par supposer que k est fixé entre 1 et K et nous désignons N_k la variable aléatoire qui compte le nombre de téléspectateurs de l'émission k. Pour les trois prochaines questions, nous supposons donc que nous n'avons qu'une seule observation N_k (par opposition aux questions de la partie 3.).

2.2.a. Quelle est la loi de N_k ? Que valent $\mathbb{E}[N_k]$ et $\mathbb{V}[N_k]$?

2.2.b. Donner l'information de Fisher du modèle.

2.2.c. Calculer $\widehat{p}_k$ l'estimateur du maximum de vraisemblance de p.

2.3. Nous supposons maintenant que nous avons K observations $N_1, \dots, N_K$ supposées indépendantes.

2.3.a. Calculer l'estimateur $\widehat{p}$ de p par la méthode du maximum de vraisemblance.

2.3.b. Exprimer $\widehat{p}$ en fonction de $\widehat{p}_1, \dots, \widehat{p}_K$.

2.3.c. Calculer l'information de Fisher de ce modèle à K observations.

2.3.d. Donner la loi asymptotique de $\widehat{p}$ lorsque n tend vers l'infini, K restant fixé. En déduire un intervalle de confiance approché, de niveau $1 - \alpha = 95\%$, pour p