I. Exercice 1

Les kératoses actiniques sont de petites lésions cutanées qui servent de précurseurs au cancer de la peau. Selon certaines théories, les adultes résidant dans des villes américaines proches de l'équateur sont plus susceptibles de développer des kératoses actiniques, et donc d'avoir un risque plus élevé de cancer de la peau, que les adultes résidant dans des villes américaines proches de l'équateur. risque de cancer de la peau, que les adultes résidant dans des villes américaines éloignées de l'équateur. Pour tester cette théorie, supposons une collecte de données comme suit

- les dossiers dermatologiques d'un échantillon aléatoire de n_1 résidents adultes d'une ville américaine particulière (disons, Ville 1) proche de l'équateur sont examinés pour déterminer le nombre de kératoses actiniques que chacun de ces n_1 adultes a développé.
- Les dossiers de dermatologie d'un échantillon aléatoire de n_0 adultes résidant dans une ville ville américaine particulière (disons, Ville 0) éloignée de l'équateur sont examinés pour déterminer le nombre de kératoses actiniques que chacun de ces adultes a développé

Comme modèle statistique pour évaluer cette théorie, pour un résident adulte j $(j = 1, \ldots, n_i)$ dans la ville i $(i = 0, 1)$, supposons que la variable aléatoire $Y_{ij} \sim \mathcal{P}\big(L_{ij}\lambda_i\big)$, où L_{ij} est la durée (en années) pendant laquelle l'adulte j a résidé dans la Ville i et où λ_i est le taux de développement des kératoses actiniques par an (c'est-à-dire le nombre attendu de kératoses actiniques qui se développent par an) pour un adulte résidant dans la ville i. Ainsi, la paire $\big(L_{ij}, y_{ij}\big)$ constitue l'information observée pour un résident adulte j dans la ville i.

1. Développer un intervalle de confiance à $100(1-\alpha)\%$ basé sur la normalité asymptotique de l'estimateur par maximum de vraisemblance pour le logarithme du rapport de taux $\ln \psi = \ln \frac{\lambda_1}{\lambda_0}$.
2. Si $n_1 = 30$ et $n_0 = 30$ et $\sum_{j=1}^{n_1} y_{1j} = 40$, $\sum_{j=1}^{n_1} L_{1j} = 350$, $\sum_{j=1}^{n_0} y_{0j} = 35$, et $\sum_{j=1}^{n_0} L_{0j} = 400$, calculer un IC à 95% pour le rapport de taux ψ. Commenter !

II. Exercice 2

Pour $\alpha, \beta > 0$, on pose, $\theta = (\alpha, \beta)$ et on définit la densité de probabilité f_θ par

$$f_\theta(x) = \alpha\beta^{-\alpha}x^{\alpha-1}\mathbb{1}_{[0,\beta]}(x).$$

On donne

$$\mathbb{E}\big[X\big] = \frac{\alpha\beta}{\alpha+1}, \quad \mathrm{Var}\big(X\big) = \frac{\alpha\beta^2}{(\alpha+1)^2(\alpha+2)}.$$

On pose $U = -\alpha \log\big(\frac{X}{\beta}\big)$. La variable aléatoire U suit une la loi $\Gamma(1,1)$ (i.e exponentielle de paramètre 1). On observe un échantillon i.i.d $X_1, X_2, \ldots, X_n$ de loi de densité f_θ.

1. Suggérer un estimateur de α. Indication : on posera $c = \dfrac{\Big(\mathbb{E}\big[X\big]\Big)^2}{\mathrm{Var}\Big(X\Big)}$.
2. **On suppose maintenant que β est connu**. Déterminer l'estimateur par maximum de vraisemblance $\widehat{\alpha}_n$ de α.
3. Montrer sans calcul que l'estimateur $\widehat{\alpha}_n$ est biaisé. Pour cela, on exprimera $\widehat{\alpha}_n$ en fonction de la moyenne empirique des variables $U_i = -\alpha \log\left(\frac{X_i}{\beta}\right)$ et on fera appel à l'inégalité de Jensen qu'on peut énoncer comme suit. Soit g une fonction convexe et X une variable aléatoire telle

que $\mathbb{E}\big[g(X)\big]$ existe : alors, $g\Big(\mathbb{E}[X]\Big) \leq \mathbb{E}\big[g(X)\big]$. L'inégalité précédente est stricte lorsque la fonction g est non linéaire. Dans notre cas, il suffit de poser $g(x) = \frac{1}{x}$ et utiliser l'inégalité $\mathbb{E}\big[\frac{1}{X}\big] > \frac{1}{\mathbb{E}[X]}$.

4. Montrer que l'estimateur $\tilde{\alpha}_n = \frac{n-1}{n}\widehat{\alpha}_n$ est sans biais. Indication : on sait que $\sum_{i=1}^n U_i$ suit une loi gamma $\Gamma(n,1)$ [1].

5. **Supposons maintenant que β est inconnu**. Déterminer l'estimateur par maximum de vraisemblance $\widehat{\beta}_n$ de β. Indications : vérifier à l'aide d'un graphique que la vraisemblance est nulle si $\beta \in \big[0, \max\limits_{1\leq i\leq n} X_i\big[$ et décroissante sur $\big[\max\limits_{1\leq i\leq n} X_i, +\infty\big[$.
6. L'estimateur $\widehat{\beta}_n$ est-il biaisé ?. Indication : examiner la probabilité $\mathbb{P}\big(\widehat{\beta}_n < \beta\big)$.

1. Rappelons la définition de la fonction gamma d'Euleur $\Gamma(a) = \int_0^{+\infty} x^{a-1}e^{-x}\mathrm{d}x$, pour tout $a > 0$.