Exercice 1

Les kératoses actiniques sont des 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 éloignées de l'équateur. Pour tester cette théorie, supposons une collecte de données comme suit

— 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

— 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.

Comme modèle statistique pour évaluer cette théorie, pour un résident adulte j $(j = 1, \dots, n_i)$ dans la ville i $(i = 0, 1)$, nous considérons 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 des 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 !

Exercice 2

On suppose que la durée T (en mois) de la rémission des patients atteints de leucémie qui ont suivi un certain type de traitement par chimiothérapie suit une loi exponentielle

$$f_T(t;\theta) = \theta e^{-\theta t}, \quad t > 0, \theta > 0.$$

Supposons que le suivi d'un échantillon aléatoire de n patients atteints de leucémie ayant suivi ce traitement conduise aux n temps de rémission observés $t_1, t_2, \dots, t_n$.

1. Calculer l'expression explicite de la variance (lorsque n est grand) de l'estimateur par maximum de vraisemblance $\widehat{\theta}_n$ de θ.
2. Un statisticien chargé d'analyser ces données se rend compte qu'il n'est pas possible de connaître avec certitude le nombre exact de mois pendant lesquels chaque patient est en rémission après avoir terminé le traitement de chimiothérapie. Il suggère la procédure alternative suivante pour estimer θ : Après un certain temps spécifié (période en mois) de longueur t^* (une constante positive connue) après la fin du traitement de chimiothérapie, soit $Y_i = 1$ si le ième patient est toujours en rémission après t^* mois et soit $Y_i = 0$ sinon. Nous avons
$$\mathbb{P}(Y_i = 1) = \mathbb{P}(T_i > t^*), i = 1, 2, \dots, n.$$
Calculer un second estimateur $\widehat{\theta}^*_n$ par maximum de vraisemblance pour θ basé sur l'observation de l'échantillon $Y_1, Y_2, \dots, Y_n$.

3. En supposant que $t^* \geq \mathbb{E}(T)$, lequel des deux estimateurs $\widehat{\theta}_n$ et $\widehat{\theta}_n^*$ a la plus petite variance, et pourquoi cela devrait-il être le résultat attendu ? Y a-t-il des circonstances où l'estimateur avec la plus grande variance pourrait être préféré ?