TD 4 STATISTIQUE - 1SN

Exercice 1.
Afin de tester la satisfaction des clients à service donné, on effectue un sondage et on définit une variable aléatoire Y_i de la façon suivante :

$$\begin{aligned} Y_i &= 1 \text{ si le client } i \text{ est satisfait} \\ Y_i &= 0 \text{ si le client } i \text{ n'est pas satisfait} \end{aligned}$$

A l'aide d'un échantillon $(Y_1, ..., Y_n)$ de même loi de Bernoulli

$$\begin{aligned} P[Y_i = 0] &= \theta \\ P[Y_i = 1] &= 1 - \theta \end{aligned}$$

on désire tester les hypothèese $H_0 : \theta = \theta_0 = 0.52$ et $H_1 : \theta = \theta_1 = 0.48$.

1. Construire la vraisemblance des observations $y_1, ..., y_n$ et expliciter la région de rejet de H_0 du test de Neyman-Pearson (pour l'application numérique, on choisira un risque de première espèce $\alpha = 0.1$).

2. Déterminer la puissance de ce test.

Exercice 2. Soit $X_1, ..., X_n$ un échantillon d'une loi normale de moyenne m et de variance σ^2. On veut faire le test d'hypothèses binaires suivant :

$$\begin{aligned} H_0 &: m = m_0; \sigma^2 \text{ quelconque} \\ H_1 &: m \neq m_0; \sigma^2 \text{ quelconque} \end{aligned}$$

Pour construire le test, on retient le test du rapport des vraisemblances maximales ou test GLR (Generalized Likelihood Ratio).

1. On suppose $m = m_0$ connu. Rappeler l'estimateur du maximum de vraisemblance (EMV) de σ^2.

2. Lorsque m et σ^2 sont inconnus, rappeler leurs estimateurs du maximum de vraisemblance.

3. Donner la forme du test GLR.

4. En décomposant $\sum_{i=1}^n (x_i - m_0)^2$, montrer que l'on peut définir un test équivalent à l'aide de la statistique
$$T_n = \frac{\overline{X} - m_0}{\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2}}$$

5. On rappelle que sous l'hypothèse H_0, les deux variables aléatoires
$$U = \frac{\overline{X} - m_0}{\sigma/\sqrt{n}} \text{ et } V = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{\sigma^2}$$
ont des lois connues $U \sim \mathcal{N}(0, 1)$ et $V \sim \chi^2_{n-1}$. En déduire la loi de T_n. Soit $\alpha = 5\%$ le risque de première espèce. Donner la région critique du test effectué à l'aide de T_n.

Exercice 3.

On considère les observations $x_i, i = 1, ..., n$ (avec $n = 10$) définies par

$x_1 = 1$	$x_2 = 0$	$x_3 = 1$	$x_4 = 1$	$x_5 = 1$	$x_6 = 1$	$x_7 = 1$	$x_8 = 2$	$x_9 = 0$	$x_{10} = 0$

On suppose que les variables aléatoires associées à ces observations sont indépendantes et issues de la même loi de Poisson $P(\lambda)$. *On rappelle que si X suit une une loi de Poisson de paramètre λ, on a* $E[X] = \text{var}[X] = \lambda$ *et* $\varphi_X(t) = E\left[e^{itX}\right] = \exp\left[\lambda\left(e^{it} - 1\right)\right]$. On désire tester les deux hypothèses

$$\begin{cases} H_0 : \lambda = \lambda_0 \text{ (absence de planète)} \\ H_1 : \lambda = \lambda_1 \text{ (présence de planète)} \end{cases}.$$

avec $\lambda_1 < \lambda_0$.

1. Vérifier que la statistique du test de Neyman-Pearson peut s'écrire $T = \sum_{i=1}^{n} X_i$ et déterminer la région critique associée.

2. Déterminer la fonction caractéristique de T et en déduire que T suit une loi de Poisson que l'on précisera sous chaque hypothèse.

3. Préciser le test de puissance maximale tel que le risque de première espèce α vérifie $\alpha \leq 0.05$. On précisera le risque maximal α, la décision prise au vu des données x_i, $i = 1, ..., 10$ et la puissance de ce test. Pour les applications numériques, on prendra $\lambda_0 = 1$ et $\lambda_1 = 0.1$.

4. On suppose que n est suffisamment grand pour pouvoir utiliser les résultats du théorème de la limite centrale.

   - Donner la loi approchée de T issue de ce théorème.
   - Quelle est la valeur du seuil obtenue lorsqu'on confond la loi de T avec son approximation. En comparant avec la valeur obtenue précédemment, dire ce que vous pensez de cette approximation pour $n = 10$.
   - Déterminer les courbes COR (caractéristiques opérationnelles du récepteur) découlant de cette loi approchée. On posera

   $$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

   et on notera $\Phi^{-1}(x)$ son inverse. En supposant que n est suffisamment grand pour faire les approximations nécessaires, déterminer les paramètres qui influent sur la performance asymptotique ($n \to \infty$) du test. De ces deux cas

   Premier Cas : $n = 100, \lambda_0 = 1, \lambda_1 = 0.1$

   Deuxième Cas : $n = 100, \lambda_0 = 2, \lambda_1 = 1.1$

   indiquer celui qui engendre la meilleure performance.

TD 4.1
Stat

Statistiques

Exercice 1 :

1/ $\mathcal{L}(Y_1, \dots Y_n, \theta) = \prod_{i=1}^{n} p(y_i) = \prod_{i=1}^{n} \theta^{1-y_i}(1-\theta)^{y_i}$

$= \theta^{n-\sum_{i=1}^{n} y_i} (1-\theta)^{\sum_{i=1}^{n} y_i}$

Théorème de Neymann Pearson

$$\frac{\mathcal{L}(Y_1, \dots Y_n, \theta_1)}{\mathcal{L}(Y_1, \dots Y_n, \theta_0)} > S_\alpha \iff \left(\frac{\theta_1}{\theta_0}\right)^{n-\sum_{i=1}^{n} y_i} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{\sum_{i=1}^{n} y_i} > S_\alpha$$

$$\iff n \ln\left(\frac{\theta_1}{\theta_0}\right) + \sum_{i=1}^{n} y_i \ln\left(\frac{1-\theta_1}{1-\theta_0} \times \frac{\theta_0}{\theta_1}\right) > S_\alpha'$$

$$\iff \sum_{i=1}^{n} y_i \ln\left(\frac{\theta_0^2}{\theta_1^2}\right) > S_\alpha'$$

$$\iff \sum_{i=1}^{n} y_i > \mu_\alpha$$

$$T(X) = \sum_{i=1}^{n} y_i$$

$E(T) = E\left(\sum_{i=1}^{n} y_i\right) = n\,E(Y_i) = n(1-\theta) \qquad Y_i \hookrightarrow \mathcal{B}(1-\theta)$

$Var(T) = Var\left(\sum_{i=n}^{n} y_i\right) = n\,Var(Y_i) = n\theta(1-\theta)$

$T(X) \hookrightarrow \mathcal{N}(n(1-\theta), n\theta(1-\theta))$ On approxime la loi de Bernouilli par une loi normale avec le TCL.

2/ $\alpha = \mathbb{P}(H_0 \text{ rej} \mid H_0 \text{ vraie})$

$= \mathbb{P}\left(\sum y_i > \mu_\alpha \mid \sum y_i \hookrightarrow \mathcal{N}(n(1-\theta_0), n\theta_0(1-\theta_0))\right)$

$= \mathbb{P}\left(\frac{\sum x_i - n(1-\theta_0)}{\sqrt{n\theta_0(1-\theta_0)}} > \frac{\mu_\alpha - n(1-\theta_0)}{\sqrt{n\theta_0(1-\theta_0)}} \;\middle|\; \frac{\sum x_i - n(1-\theta_0)}{\sqrt{n\theta_0(1-\theta_0)}} \hookrightarrow \mathcal{N}(0,1)\right)$

ϕ : Fonction de répartition pour la loi normale.

$\alpha = 1 - \phi\left(\frac{\mu_\alpha - n(1-\theta_0)}{\sqrt{n\theta_0(1-\theta_0)}}\right)$

$$\boxed{\mu_\alpha = \phi^{-1}(1-\alpha)\; n\theta_0(1-\theta_0) + n(1-\theta_0)}$$

$\beta = \mathbb{P}(H_1 \text{ rej} \mid H_1 \text{ vraie}) \qquad \Pi = 1 - \beta$

$\Pi = \mathbb{P}(H_0 \text{ rej} \mid H_1 \text{ vraie})$

$= \mathbb{P}\left(\frac{\sum y_i - n(1-\theta_1)}{\sqrt{n\theta_1(1-\theta_1)}} > \frac{\mu_\alpha - n(1-\theta_1)}{\sqrt{n\theta_1(1-\theta_1)}} \;\middle|\; \frac{\sum x_i - n(1-\theta_1)}{\sqrt{n\theta_1(1-\theta_1)}} \hookrightarrow \mathcal{N}(0,1)\right)$

$$\Pi = 1 - \phi\left(\frac{\mu_\alpha - n(1-\theta_1)}{\sqrt{n\theta_1(1-\theta_1)}}\right) = f(\alpha)$$

$$= 1 - \phi$$

$$= 1 - \phi\left(\frac{\phi^{-1}(1-\alpha)\sqrt{n\theta_0(1-\theta_0)} + n(1-\theta_0) - n(1-\theta_1)}{\sqrt{n\theta_1(1-\theta_1)}}\right)$$

Exercice 3:

1/ $\mathcal{L}(x_1, \dots x_n \mid \lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}$

Théorème de Neyman-Pearson:

$$\frac{\prod_{i=1}^{n} e^{-\lambda_1} \frac{\lambda_1^{x_i}}{x_i!}}{\prod_{i=1}^{n} e^{-\lambda_0} \frac{\lambda_0^{x_i}}{x_i!}} > S_\alpha \iff \left(\frac{e^{-\lambda_1}}{e^{-\lambda_0}}\right)^n \times \frac{\lambda_1^{\sum x_i}}{\lambda_0^{\sum x_i}} > S_\alpha$$

$$\iff \sum_{i=1}^{n} x_i \underbrace{\ln\left(\frac{\lambda_1}{\lambda_0}\right)}_{<0} > S_\alpha'$$

$$\iff \sum_{i=1}^{n} x_i < \mu_\alpha$$

$$T(x) = \sum_{i=1}^{n} X_i$$

2/ F.C de T $\varphi_T(t) = E(e^{jtT}) = E\left(\prod_{i=1}^{n} e^{jtx_i}\right)$ comme X_i iid

$$\varphi_T(t) = E(e^{jtx_i})^n = e^{n\lambda(e^{jt}-1)}$$

Donc $T \hookrightarrow \mathcal{P}(n\lambda)$. $H_0: T \hookrightarrow P(10)$
$T \hookrightarrow P(1)$

3/ $\alpha = P(\text{rejet } H_0 \mid H_0 \text{ vraie})$

$$= P\left(\sum_{i=1}^{n} x_i < \mu_\alpha \mid T \hookrightarrow P(n\lambda_0)\right)$$

$$= \Phi_{T_{10}}(\mu_\alpha)$$

$\mu_\alpha = \Phi_{T_{10}}^{-1}(\alpha)$ $\longrightarrow \mu_\alpha = 5 \quad \alpha = 0{,}0293$

$\beta = P(\text{rejet } H_1 \mid H_1 \text{ vraie})$

$$= 1 - \Phi_{T_1}(\mu_\alpha) = 1 - \Phi_{T_1}\left(\Phi_{T_{10}}^{-1}(\alpha)\right)$$

$\Pi = \Phi_{T_1}\left(\Phi_{T_{10}}^{-1}(\alpha)\right) = 0{,}9963 = 1 - \beta = P[\text{rej } H_0 \mid H_1 \text{ vraie}]$

TD 4.2
Stat

4/ $T \hookrightarrow \mathcal{N}(n\lambda, n\lambda)$

$$\alpha = \mathbb{P}\left[\sum x_i < k_\alpha \mid \sum x_i \hookrightarrow \mathcal{N}(n\lambda_0, n\lambda_0) \right]$$

$$= \mathbb{P}\left[\frac{\sum x_i - n\lambda_0}{\sqrt{n\lambda_0}} < \frac{k_\alpha - n\lambda_0}{\sqrt{n\lambda_0}} \;\middle|\; \sum x_i \hookrightarrow \mathcal{N}(0,1) \right].$$

$$\alpha = \Phi_{\mathcal{N}}\left(\frac{k_\alpha - n\lambda_0}{\sqrt{n\lambda_0}} \right)$$

$$k_\alpha = \sqrt{n\lambda_0}\, \Phi_{\mathcal{N}}^{-1}(\alpha) + n\lambda_0.$$

$$\alpha = \Phi(x) \qquad x = \Phi^{-1}(\alpha).$$
$$(1-\alpha) = 1 - \Phi(x)$$
$$(1-\alpha) = \Phi(-x)$$
$$\Phi^{-1}(1-\alpha) = -x.$$
$$x = -\Phi^{-1}(1-\alpha) \iff \Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha)$$

$k_\alpha = 4,8$ pour cette application là, on peut se permettre d'approximer T par la loi normale car $k_\alpha \simeq \mu_\alpha$.

$$\Pi = \Phi\left(\frac{k_\alpha - n\lambda_1}{\sqrt{n\lambda_1}} \right) = \Phi\left(\frac{\Phi^{-1}(\alpha)\sqrt{n\lambda_0} + n\lambda_0 - n\lambda_1}{\sqrt{n\lambda_1}} \right)$$

$$= \underset{PD}{\Phi}\left(\underset{PFA}{\Phi^{-1}(\alpha)} \sqrt{\frac{\lambda_0}{\lambda_1}} + \frac{\sqrt{n}(\lambda_0 - \lambda_1)}{\sqrt{\lambda_1}} \right)$$

$$\underset{n \to +\infty}{\sim} \Phi\left(\frac{\sqrt{n}(\lambda_0 - \lambda_1)}{\sqrt{\lambda_1}} \right).$$

Come $\lambda_{1_1} < \lambda_{1_2}$.

Donc 1^{er} cas est le meilleur ($+ \lambda_1 \leq 0$ $+ \Phi \uparrow$).