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 :

 $Y_i = 1$ si le client i est satisfait

 $Y_i = 0$ si le client i n'est pas satisfait

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

$$P[Y_i = 0] = \theta$$

$$P[Y_i = 1] = 1 - \theta$$

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 :

 $H_0: m=m_0; \sigma^2$ quelconque

 $H_1: m \neq m_0; \sigma^2$ quelconque

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

$$U = \frac{\overline{X} - m_0}{\sigma / \sqrt{n}}$$
 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 .

1

Exercice 3.

On considère n variables aléatoires $X_1, ..., X_n$ indépendantes suivant la même loi géométrique à valeurs dans l'ensemble des entiers non nuls $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$ définie par

$$P[X_i = x_i; \theta] = \theta (1 - \theta)^{x_i - 1}, \qquad x_i \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}.$$

avec $\theta \in]0,1[$. On rappelle que la moyenne et la variance d'une loi géométrique de paramètre θ sont définies par $E\left[X_i\right]=\frac{1}{\theta}$ et $\text{var}(X_i)=\frac{1-\theta}{\theta^2}$. On considère le test d'hypothèses simples

$$H_0: \theta = \theta_0, \quad H_1: \theta = \theta_1.$$

- 1. (2pts) Déterminer la statistique du test du test de Neyman Pearson notée T_n et indiquer la région critique de ce test pour $\theta_1 > \theta_0$ et $\theta_1 < \theta_0$. Dans la suite de cet exercice, on supposera $\theta_1 > \theta_0$. La décision prise à l'aide du test de Neyman Pearson est-elle en accord avec la moyenne d'une loi géométrique définie par $E[X_i] = \frac{1}{\theta}$?
- 2. (1pt) Déterminer la loi approchée de la statistique T_n résultant de l'application du théorème de la limite centrale.
- 3. (1pt) On note

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

la fonction de répartition d'une loi normale $\mathcal{N}(0,1)$ et F^{-1} son inverse. Déterminer la valeur du seuil du test de Neyman Pearson notée K_{α} en fonction de n, θ_0 , du risque de première espèce α et de F^{-1} .

- 4. (1pt) Déterminer la puissance du test en fonction du seuil K_{α} , de n, θ_1 et de F.
- 5. (2pts) Déterminer les courbes COR du test étudié dans cet exercice et tracer la forme de ces courbes pour différentes valeurs de n.

Exercice 4.

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

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

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. 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.
- Déterminer les courbes COR (caractéristiques opérationnelles du récepteur) découlant de cette loi approchée. On posera

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

et on notera $F^{-1}(x)$ 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.

Exercice 5.

Un statisticien pose la question suivante à un échantillon de 30 participants : "Préférez-vous boire du thé ou du café ?". Parmi cet échantillon, 10 préfèrent le thé et 20 préfèrent le café. Il désire effectuer un test du chi-deux pour déterminer s'il y a une véritable préférence pour le café dans cet échantillon. Pour cela, il définit l'hypothèse H_0 par "la probabilité de boire du thé est égale à la probabilité de boire du café", i.e., les deux classes $\{\text{Thé}\}$ et $\{\text{Café}\}$ sont équiprobables $(P[\text{Thé}] = P[\text{Café}] = \frac{1}{2})$.

- 1. Déterminer la statistique du test du chi-deux noté ϕ associée à ce problème.
- 2. Rappeler la loi de ϕ sous l'hypothèse H_0 (définie par "Il n'y a pas de préférence ni pour le thé, ni pour le café").
- 3. Expliquer comment déterminer le seuil de décision S_{α} du test du chi-deux à l'aide de la fonction de répartition de la loi déterminée à la question précédente et du risque α de ce test. Pour $\alpha=0.05$, on trouve $S_{0.05}=3.84$. Que conclut-on ?

Correction exercice 1

1) La vraisemblance de ce problème est

$$L(y_1, ..., y_n; \theta) = \prod_{i=1}^n P[Y_i = y_i]$$
$$= \prod_{i=1}^n \theta^{1-y_i} (1-\theta)^{y_i}$$
$$= \theta^{n-n\overline{y}} (1-\theta)^{n\overline{y}}$$

avec

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

On rejette donc H_0 si

$$\frac{L(y_1,...,y_n;\theta_1)}{L(y_1,...,y_n;\theta_0)} > K_\alpha \Longleftrightarrow \overline{y} \ln \left(\frac{\theta_0}{\theta_1} \frac{1-\theta_1}{1-\theta_0} \right) > S_\alpha$$

Pour $\theta_0=0.52$ et $\theta_1=0.48$, on a

$$\frac{\theta_0}{\theta_1} \frac{1 - \theta_1}{1 - \theta_0} = \left(\frac{0.52}{0.48}\right)^2 > 1$$

donc on rejette H_0 si

$$\overline{y} > \nu_{\alpha}$$

où ν_{α} est un seuil dépendant du risque de première espèce α . Pour déterminer ce seuil, on se fixe une valeur de α

$$\alpha = P \left[\text{Rejeter } H_0 | H_0 \text{ vraie} \right]$$

$$= P \left[\overline{Y} > \nu_{\alpha} | \theta = \theta_0 \right]$$

En utilisant le théorème de la limite centrale, on peut approcher la loi de \overline{Y} comme suit

$$\overline{Y} \sim \mathcal{N}\left(1 - \theta, \frac{\theta\left(1 - \theta\right)}{n}\right)$$

Donc

$$\alpha = P \left[U = \frac{\overline{Y} - (1 - \theta_0)}{\sqrt{\frac{\theta_0(1 - \theta_0)}{n}}} > \frac{\nu_\alpha - (1 - \theta_0)}{\sqrt{\frac{\theta_0(1 - \theta_0)}{n}}} \right] U \sim \mathcal{N}(0, 1)$$

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

On en déduit

$$\frac{\nu_{\alpha} - (1 - \theta_0)}{\sqrt{\frac{\theta_0(1 - \theta_0)}{n}}} = F^{-1} (1 - \alpha)$$

où F est la fonction de répartition d'une loi normale $\mathcal{N}(0,1)$, d'où

$$\nu_{\alpha} = \sqrt{\frac{\theta_0 (1 - \theta_0)}{n}} F^{-1} (1 - \alpha) + (1 - \theta_0)$$

2) La puissance du test est

$$\pi = P \left[\text{Rejeter } H_0 \middle| H_1 \text{ vraie} \right]$$

$$= P \left[\overline{Y} > \nu_\alpha \middle| \theta = \theta_1 \right]$$

$$= 1 - F \left(\frac{\nu_\alpha - (1 - \theta_1)}{\sqrt{\frac{\theta_1 (1 - \theta_1)}{n}}} \right)$$

Correction exercice 2

1)
$$\tilde{\sigma}_{MV}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m_0)^2$$

2)
$$\hat{m}_{MV} = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad \hat{\sigma}_{MV}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2$$

3) Le test GLR est défini par

Rejet de
$$H_0$$
 si $\frac{L(X_1,...,X_n;H_1)}{L(X_1,...,X_n;H_1)} > S_{\alpha}$

c'est-à-dire

Rejet de
$$H_0$$
 si
$$\frac{\left(2\pi\hat{\sigma}_{MV}^2\right)^{-n/2}\exp\left[-\frac{1}{2\hat{\sigma}_{MV}^2}\sum\left(X_i-\overline{X}\right)^2\right]}{\left(2\pi\tilde{\sigma}_{MV}^2\right)^{-n/2}\exp\left[-\frac{1}{2\tilde{\sigma}_{MV}^2}\sum\left(X_i-m_0\right)^2\right]} > K_{\alpha}$$

c'est-à-dire

Rejet de
$$H_0$$
 si $\frac{\tilde{\sigma}_{MV}^2}{\hat{\sigma}_{MV}^2} > S_{\alpha} \Leftrightarrow \frac{\sum (X_i - m_0)^2}{\sum_{i=1}^n (X_i - \overline{X})^2} > S_{\alpha}$

4) On décompose $\sum (X_i - m_0)^2$ comme suit

$$\sum (X_i - m_0)^2 = \sum (X_i - \overline{X} + \overline{X} - m_0)^2$$
$$= \sum (X_i - \overline{X})^2 + n(\overline{X} - m_0)^2$$

donc le test GLR est défini par

Rejet de
$$H_0$$
 si $\frac{\sum \left(X_i - \overline{X}\right)^2 + n\left(\overline{X} - m_0\right)^2}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2} > S_\alpha \Leftrightarrow T_n^2 > \mu_\alpha$
 $\Leftrightarrow T_n \in]-\infty, -\mu_\alpha[\cup]\mu_\alpha, \infty[$

5) La statistique T_n s'écrit sous la forme suivante :

$$T_n = \frac{\sigma}{\sqrt{n}} \frac{U}{\sigma \sqrt{V}} = \frac{1}{\sqrt{n(n-1)}} \frac{U}{\sqrt{\frac{V}{n-1}}}$$

où

$$W_n = \frac{U}{\sqrt{\frac{V}{n-1}}} \sim t_{n-1}$$

On en déduit

Rejet de
$$H_0$$
 si $W_n \in]-\infty, -c_{\alpha}[\ \cup\]c_{\alpha}, \infty[$

et

$$\begin{array}{rcl} 1 - \alpha & = & 1 - P \left[\text{rejeter } H_0 \middle| H_0 \text{ vraie} \right] \\ & = & P \left[\text{accepter } H_0 \middle| H_0 \text{ vraie} \right] \\ & = & P \left[\left| W_n \middle| < c_\alpha \middle| H_0 \text{ vraie} \right] = 0.95 \end{array}$$

Les tables de la loi de Student donnent la valeur de c_{α} .

Correction exercice 3

d'où
$$E\left(\frac{1}{x}\right) = \theta + \frac{1-\theta}{n\theta^2} \times \theta^3 - \frac{\theta^4}{h^2} \frac{\nu^3}{h^2} + \cdots$$

$$= 9 + \frac{9(4-9)}{h} - \frac{1}{4^{2}} + 3(\frac{1}{4^{2}})$$

Leci permet de montre que 1 rr en astimateur asymptotiquement sans brais de 0 et que sa variance asymptotique en P(1-10), qui et le some de Cramer-Rac pour le paremètre 0

3 Exeriaz

c'Al-à dire

Reget de to hi
$$\frac{\theta_1}{\theta_0} \frac{(1-\theta_1)^{\frac{5}{2}} \frac{2^{\frac{1}{4}} - 4}{(1-\theta_0)^{\frac{5}{2}} \frac{2^{\frac{1}{4}} - 4}{(1-\theta_0)}}{\sum_{i=1}^{4} \frac{(1-\theta_1)^{\frac{5}{2}} \frac{2^{\frac{1}{4}} - 4}{(1-\theta_0)}}{\sum_{i=1}^{4} \frac{(1-\theta_1)^{\frac{5}{2}} \frac{2^{\frac{1}{4}} - 4}{(1-\theta_0)}} > S_{\infty}$$

51 81>80, also 1-0, <1-0, done 1-01 <1 done (1-0) <0 done

2) Dapris le théorème de la limite centrale, pour n grand, on peut approcher le lui de Th = \(\frac{5}{2}\) Kr par une loi normale de nE[xi] = n et de varance nvarxi = n/n-01

morgani
$$n \in [x_i] = \frac{n}{\theta}$$
 et de variance $n \vee a \vee x_i = \frac{n(n-\theta)}{\theta^2}$

done
$$\left[\text{Tr} \sim N\left(\frac{n}{\theta}, \frac{n(n-\theta)}{\theta^2} \right) \right]$$

3) $d = P[Regeter Ho| Hovraie] = P[Th < H2 | Th ~ N(\frac{n}{\theta} \graphi \frac{M-Do)}{Do^2})$

$$= D \left[d = F \left[\frac{4a - \frac{n}{00}}{\sqrt{n(n-0.)}} \right] \right]$$

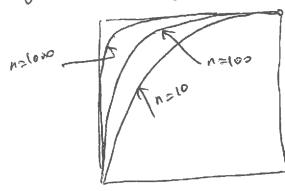
G) TI = Pl Negeter HolMi sionê]
$$= Pl Th < Va | \theta = \theta_i) = \left[F \left[\frac{Va - \frac{h}{\theta_1}}{\sqrt{\frac{h(h_1\theta_1)}{\theta_1}}} \right] \right]$$

5) Cowler COR Il suffit de remplace l'expression de to de 3) dans l'expression de T trower entry pour obtenir

Soot
$$T = F \left[\frac{n}{\theta_0} \frac{n}{\theta_1} + \frac{\theta_1}{\theta_0} \sqrt{\frac{1-\theta_0}{1-\theta_1}} + \frac{1}{\theta_0} \sqrt{\frac{1-\theta_0}{1-\theta_1}} \right]$$

$$T = F \left[\sqrt{\frac{\theta_1-\theta_1}{\theta_0}} + \frac{\theta_1}{\theta_0} \sqrt{\frac{1-\theta_0}{1-\theta_1}} + \frac{\theta_1}{\theta_0} \sqrt{\frac{1-\theta_0}{1-\theta_1}} \right]$$

Plus n'est grand, plus II regrande, ce qui donne les caviles suivantes



Exercise 3

1)
$$\phi = \sum_{i=1}^{2} \frac{(k_i - np_i)^2}{np_i}$$
 over $n = 60$ ext $p_i = 10$ ent $np_i = 10$
 $d^i c u^i \qquad \phi = \frac{1}{10} \frac{(n_0 + u^2)^2}{(n_0 + u^2)^2} + \frac{(n_0 - u^2)^2}{(n_0 - u^2)^2} + \frac{(n_0 - u^2$

Correction exercice 4

1) Des calculs élémentaires donnent

Rejet de
$$H_0$$
 si $T = \sum_{i=1}^n X_i < S_{\alpha}$

2) La fonction caractéristique de T est

$$E\left[e^{itT}\right] = \prod_{j=1}^{n} E\left[e^{itX_{j}}\right] = \exp\left[n\lambda\left(e^{it} - 1\right)\right]$$

qui est la fonction caractéristique d'une loi de Poisson de paramètre $n\lambda$ donc $T \sim P(n\lambda)$. Sous H_0 , on a $T \sim P(n\lambda_0) = P(10)$ et sous H_1 , on a $T \sim P(n\lambda_1) = P(1)$.

- 3) a) Pour n grand, l'approximation normale est $\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\lambda, n\lambda)$.
- b) On trouve $K_{\alpha} = n\lambda_0 + \sqrt{n\lambda_0}F^{-1}(\alpha) \sim 4.8$.
- c) Un calcul simple conduit à

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

c'est-à-dire asymptotiquement

$$PD = 1 - \beta \sim F\left(\sqrt{n} \frac{\lambda_0 - \lambda_1}{\sqrt{\lambda_1}}\right)$$

Le paramètre qui règle la performance asymptotique du test est donc $\sqrt{n}\frac{\lambda_0-\lambda_1}{\sqrt{\lambda_1}}$. Dans les deux cas proposés $\lambda_0-\lambda_1=0.9$ et n=100. Le premier test est meilleur car PD est une fonction décroissante de λ_1 lorsque $\lambda_0-\lambda_1$ et n sont fixés.

$k \setminus \lambda$	0.8	0.9	1.0	2.0	3.0	4.0	5.0	6.0
	0.4400	0.4000	0.0070	0.1050	0.0400	0.0100	0.0007	0.000
0	0.4493	0.4066	0.3679	0.1353	0.0498	0.0183	0.0067	0.0025
1	0.3595	0.3659	0.3679	0.2707	0.1494	0.0733	0.0337	0.0149
2	0.1438	0.1647	0.1839	0.2707	0.2240	0.1465	0.0842	0.0446
3	0.0383	0.0494	0.0613	0.1804	0.2240	0.1954	0.1404	0.0892
4	0.0077	0.0111	0.0153	0.0902	0.1680	0.1954	0.1755	0.1339
5	0.0012	0.0020	0.0031	0.0361	0.1008	0.1563	0.1755	0.1606
6	0.0002	0.0003	0.0005	0.0120	0.0504	0.1042	0.1462	0.1606
7			0.0001	0.0034	0.0216	0.0595	0.1044	0.1377
8				0.0009	0.0081	0.0298	0.0653	0.1033
9				0.0002	0.0027	0.0132	0.0363	0.0688
10					0.0008	0.0053	0.0181	0.0413
11					0.0002	0.0019	0.0082	0.0225
12					0.0001	0.0006	0.0034	0.0113
13						0.0002	0.0013	0.0052
14						0.0001	0.0005	0.0022
15							0.0002	0.0009
16								0.0003
17								0.0001

$k \setminus \lambda$	7.0	8.0	9.0	10.0	11.0	12.0	13.0
II \ /\	1.0	0.0	0. 0	10.0	11.0	12.0	10.0
0	0.0009	0.0003	0.0001				
$oxed{1}$	0.0064	0.0027	0.0011	0.0005	0.0002	0.0001	
2	0.0223	0.0107	0.0050	0.0023	0.0010	0.0004	0.0002
3	0.0521	0.0286	0.0150	0.0076	0.0037	0.0018	0.0008
4	0.0912	0.0573	0.0337	0.0189	0.0102	0.0053	0.0027
5	0.1277	0.0916	0.0607	0.0378	0.0224	0.0127	0.0070
6	0.1490	0.1221	0.0911	0.0631	0.0411	0.0255	0.0152
7	0.1490	0.1396	0.1171	0.0901	0.0646	0.0437	0.0281
8	0.1304	0.1396	0.1318	0.1126	0.0888	0.0655	0.0457
9	0.1014	0.1241	0.1318	0.1251	0.1085	0.0874	0.0661
10	0.0710	0.0993	0.1186	0.1251	0.1194	0.1048	0.0859
11	0.0452	0.0722	0.0970	0.1137	0.1194	0.1144	0.1015
12	0.0263	0.0481	0.0728	0.0948	0.1094	0.1144	0.1099
13	0.0142	0.0296	0.0504	0.0729	0.0926	0.1056	0.1099
14	0.0071	0.0169	0.0324	0.0521	0.0728	0.0905	0.1021
15	0.0033	0.0090	0.0194	0.0347	0.0534	0.0724	0.0885
16	0.0014	0.0045	0.0109	0.0217	0.0367	0.0543	0.0719
17	0.0006	0.0021	0.0058	0.0128	0.0237	0.0383	0.0550
18	0.0002	0.0009	0.0029	0.0071	0.0145	0.0255	0.0397
19	0.0001	0.0004	0.0014	0.0037	0.0084	0.0161	0.0272
20		0.0002	0.0006	0.0019	0.0046	0.0097	0.0177
21		0.0001	0.0003	0.0009	0.0024	0.0055	0.0109
22			0.0001	0.0004	0.0012	0.0030	0.0065
23				0.0002	0.0006	0.0016	0.0037
24				0.0001	0.0003	0.0008	0.0020
25					0.0001	0.0004	0.0010
26						0.0002	0.0005
27						0.0001	0.0002
28							0.0001
29							0.0001
30							