Exercice 1
On dispose de n observations $x_1, ..., x_n$ issues d'un échantillon $(X_1, ..., X_n)$ de loi normale $\mathcal{N}\left(0, \sigma^2\right)$.

1. Déterminer suivant les valeurs de σ_0^2 et de σ_1^2, la statistique $T\left(X_1, ..., X_n\right)$ et la région critique du test de Neyman-Pearson associé au problème suivant :
$$\left\{\begin{array}{l} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 = \sigma_1^2 \end{array}\right.$$
Commenter la forme de la statistique de test. Dans la suite de ce problème, on supposera $\sigma_1 > \sigma_0$.

2. On admettra que $T = \frac{1}{\sigma^2} \sum_{i=1}^n X_i^2$ suit une loi du χ_n^2. Déterminer la valeur des risques de première et seconde espèce α et β en fonction du seuil du test noté S, de la fonction de réparition d'une loi du χ_n^2 notée F_n et de σ_0 et σ_1.

3. Déterminer les courbes COR (caractéristiques opérationnelles du récepteur) associées au problème précédent. Comment la puissance du test de Neyman Pearson dépend-elle de σ_0 et de σ_1 ? Représenter la forme approximative des courbes COR pour différentes valeurs de ces paramètres.

4. Quelle est la loi asymptotique de la statistique de $T\left(X_1, ..., X_n\right)$ sous les deux hypothèses H_0 et H_1 issue de l'application du théorème de la limite centrale ? En utilisant cette loi asymptotique, déterminer les courbes COR du problème et montrer que leur expression en fonction de σ_0^2 et σ_1^2 est en accord avec la question précédente.

Exercice 2 : Test Statistique

On considère n variables aléatoires $Y_1, ..., Y_n$ indépendantes de densités
$$f\left(y_i; \theta\right) = \frac{1}{\sqrt{2\pi\theta y_i}} \exp\left(-\frac{y_i}{2\theta}\right) \mathbb{I}_{\mathbb{R}^+}\left(y_i\right)$$
avec $\theta > 0$ et où $\mathbb{I}_{\mathbb{R}^+}$ indique la fonction indicatrice sur $\mathbb{R}^+$. On veut à l'aide d'observations $y_1, ..., y_n$ de ces variables aléatoires tester les deux hypothèses suivantes
$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1 \qquad \text{avec } \theta_1 > \theta_0.$$

1. À l'aide du théorème de Neyman-Pearson, calculer la statistique T_n du test le plus puissant et indiquer la région critique de ce test (comme d'habitude, on retiendra pour T_n la fonction seule des observations).

2. Montrer que $\frac{Y_i}{\theta}$ possède une loi du χ_1^2. En déduire que $\frac{T_n}{\theta}$ suit une loi du χ_n^2.

3. En déduire une expression des risques de première et seconde espèce α et β du test en fonction des paramètres θ_0 et θ_1 et de la fonction de répartition d'une loi du χ_n^2 notée F_n.

4. Déterminer les courbes COR associées à ce test et expliquer comment les performances de ce test dépendent des valeurs de θ_0 et θ_1.

5. On désire à partir d'un ensemble d'observations $(y_1, ..., y_n)$ déterminer s'il est raisonnable de supposer que ces observation sont issues de la densité $f\left(y_i; \theta\right)$. Expliquer le principe du test de Kolmogorov pour résoudre ce problème: on précisera avec soin 1) comment calculer la statistique de ce test notée D_n, 2) la région critique du test, 3) l'expression du seuil du test en fonction du risque de première espèce α et de la loi asymptotique de $\sqrt{n}D_n$ et 4) la manière de prendre en compte le fait que θ soit inconnu.

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 $P(\lambda)$. O*n 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

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

     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.

Exercice 4.
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 hypothèses $H_0 : \theta = \theta_0 = 0.52$ 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 5.

On lance un dé 60 fois et on relève les nombres de fois où on a observé les différentes faces

x_i	1	2	3	4	5	6
n_i	15	7	4	11	6	17

On se demande si ce dé est truqué (Hypothèse H_1) ou non (Hypothèse H_0).

1. Déterminer la statistique du test du chi-deux noté ϕ associée à ce problème.

2. Quelle est la loi de cette statistique de test sous l'hypothèse H_0 ?

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 $S_{0.05} = 11.07$ et pour $\alpha = 0.01$, on a $S_{0.01} = 15.09$. Que conclut-on ?

Exercice 6.

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 {Thé} et {Café} sont équiprobables (P[Thé] $=$ P[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 ?

Corrections exercices des TDs

Exercice 1

1) Des calculs classiques permettent d'obtenir la règle de décision suivante :

$$\begin{aligned}\text{Si } \sigma_1 &> \sigma_0, \quad \text{rejet de } H_0 \text{ si } T = \sum_{i=1}^{n} X_i^2 > S \\ \text{Si } \sigma_1 &< \sigma_0, \quad \text{rejet de } H_0 \text{ si } T = \sum_{i=1}^{n} X_i^2 < S.\end{aligned}$$

La statistique du test est donc $T = T(X_1, ..., X_n) = \sum_{i=1}^{n} X_i^2$, c'est-à-dire n fois le moment d'ordre 2 empirique des variables X_i. Ceci semble naturel puisque $E\left[X_i^2\right] = \sigma^2$ varie sous les deux hypothèses H_0 et H_1.

2) Le risque de première espèce est

$$\begin{aligned}\alpha &= P\left[\text{rejeter } H_0 |\, H_0 \text{ vraie}\right] \\ &= P\left[T > S |\, \sigma = \sigma_0\right] \\ &= P\left[\frac{1}{\sigma_0^2}T > \frac{S}{\sigma_0^2}\middle|\, \sigma = \sigma_0\right]\end{aligned}$$

On sait que $\frac{1}{\sigma_0^2}T$ suit une loi du chi2 à n degrés de liberté sous l'hypothèse H_0, donc

$$\alpha = 1 - F_n\left(\frac{S}{\sigma_0^2}\right), \text{ i.e., } S = \sigma_0^2 F_n^{-1}(1-\alpha).$$

Le risque β peut se calculer de la même façon

$$\begin{aligned}\beta &= P\left[\text{rejeter } H_1 |\, H_1 \text{ vraie}\right] \\ &= P\left[T \leq S |\, \sigma = \sigma_1\right] \\ &= P\left[\frac{1}{\sigma_1^2}T \leq \frac{S}{\sigma_1^2}\middle|\, \sigma = \sigma_1\right] \\ &= F_n\left(\frac{S}{\sigma_1^2}\right).\end{aligned}$$

3) Si on appelle π la puissance du test, les courbes COR sont définies par

$$\begin{aligned}\pi &= 1 - \beta = 1 - F_n\left(\frac{S}{\sigma_1^2}\right) \\ \pi &= 1 - F_n\left(\frac{\sigma_0^2}{\sigma_1^2}F_n^{-1}(1-\alpha)\right)\end{aligned}$$

On voit donc que π est une fonction décroissante de $\frac{\sigma_0^2}{\sigma_1^2}$, c'est-à-dire une fonction croissante de $\frac{\sigma_1^2}{\sigma_0^2}$, ce qui est normal.

4) Le théorème de la limite centrale nous permet d'approcher, pour n "grand", la loi de $T(X_1, ..., X_n) = \sum_{i=1}^{n} X_i^2$ par une loi normale $\mathcal{N}\left(nE\left[X_1^2\right], n\text{var}\left[X_1^2\right]\right)$. Mais

$$\begin{aligned}E\left[X_1^2\right] &= \sigma^2 \\ \text{var}\left[X_1^2\right] &= E\left[X_1^4\right] - E\left[X_1^2\right]^2 = 3\sigma^4 - \sigma^4 = 2\sigma^4.\end{aligned}$$

Donc, on peut approcher la loi de T par une loi normale $\mathcal{N}\left(n\sigma^2, 2n\sigma^4\right)$. En notant G la fonction de répartition d'une loi normale $\mathcal{N}(0,1)$, on en déduit

$$\begin{aligned}
\alpha &= P\left[\text{rejeter } H_0 \middle| H_0 \text{ vraie}\right] \\
&= P\left[\frac{T - n\sigma^2}{\sqrt{2n}\sigma^2} > \frac{S - n\sigma^2}{\sqrt{2n}\sigma^2} \middle| \sigma = \sigma_0\right] \\
&= 1 - G\left(\frac{S - n\sigma_0^2}{\sqrt{2n}\sigma_0^2}\right)
\end{aligned}$$

d'où

$$S = n\sigma_0^2 + \sqrt{2n}\sigma_0^2 G^{-1}\left(1-\alpha\right).$$

et

$$\begin{aligned}
\beta &= P\left[\text{rejeter } H_1 \middle| H_1 \text{ vraie}\right] \\
&= P\left[T \leq S \middle| \sigma = \sigma_1\right] \\
&= P\left[\frac{T - n\sigma_1^2}{\sqrt{2n}\sigma_1^2} \leq \frac{S - n\sigma_1^2}{\sqrt{2n}\sigma_1^2} \middle| \sigma = \sigma_1\right] \\
&= G\left(\frac{S - n\sigma_1^2}{\sqrt{2n}\sigma_1^2}\right).
\end{aligned}$$

La puissance du test s'écrit donc

$$\pi = 1 - \beta = 1 - G\left(\frac{S - n\sigma_1^2}{\sqrt{2n}\sigma_1^2}\right).$$

Après avoir remplacé S par son expression trouvée ci-dessus, on obtient

$$\pi = 1 - \beta = 1 - G\left(-\sqrt{\frac{n}{2}} + \frac{\sigma_0^2}{\sigma_1^2}\left[\sqrt{\frac{n}{2}} + G^{-1}(1-\alpha)\right]\right).$$

qui est une fonction croissante de $\frac{\sigma_1^2}{\sigma_0^2}$, comme dans la question précédente.

On a donc $\boxed{\beta(n) = \frac{n}{n+2}}$

$\beta(n) \underset{n \to +\infty}{\longrightarrow} 1$ donc $\hat{\sigma}^2_{MMSE}$ se comporte comme $\tilde{\sigma}^2_{MV}$ quand $n \to +\infty$ (on fait confiance aux données)

$\beta(n) \underset{n \to 0}{\longrightarrow} 0$ donc $\hat{\sigma}^2_{MMSE}$ se comporte comme σ_0^2 qui est la moyenne de la loi a priori (on fait confiance à l'information a priori)

Exercice 2

① Le test de Neyman - Pearson s'écrit

Rejet de H_0 si $\frac{L(y_1, \dots, y_n ; \theta_1)}{L(y_1, \dots, y_n ; \theta_0)} > \text{seuil}$

soit
$$\frac{\frac{1}{(2\pi\theta_1)^{n/2}} \frac{1}{\prod \sqrt{y_i}} \exp\left(-\frac{1}{2\theta_1} \sum_{i=1}^n y_i\right)}{\frac{1}{(2\pi\theta_0)^{n/2}} \frac{1}{\prod \sqrt{y_i}} \exp\left(-\frac{1}{2\theta_0} \sum_{i=1}^n y_i\right)} > \text{seuil}$$

Un test équivalent est donc (on simplifie et on calcule le logarithme des deux membres de l'inégalité)

Rejet de H_0 si $\underbrace{\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)}_{>0 \text{ car } \theta_1 > \theta_0} \sum_{i=1}^n y_i > \text{seuil}$

Soit $\boxed{\text{Rejet de } H_0 \text{ si } \sum_{i=1}^n y_i > S_\alpha}$

La statistique de test est donc $T_n = \sum_{i=1}^n y_i$ et la région critique du test est $\left\{(y_1, \dots, y_n) / \sum_{i=1}^n y_i > S_\alpha\right\}$

2) on pose $Z_i = \frac{Y_i}{\theta}$ qui est un changement de variable bijectif de $\mathbb{R}^+$ dans $\mathbb{R}^+$ de Jacobien $\left|\frac{dy_i}{dz_i}\right| = \theta$ - On en déduit la densité de Z_i (5)

$$q(z_i) = \frac{1}{\sqrt{2\pi z_i \theta^2}} \exp\left(-\frac{z_i}{2}\right)|\theta| \quad z_i > 0$$

Soit

$$\boxed{q(z_i) = \frac{1}{\sqrt{2\pi z_i}} e^{-z_i/2} \mathbb{1}_{\mathbb{R}^+}(z_i)}$$

On reconnait d'après les tables la densité d'une loi du χ_1^2 donc

$$\boxed{Z_i = \frac{Y_i}{\theta} \sim \chi_1^2}$$

$U_n = \frac{T_n}{\theta} = \sum_{i=1}^{n} \frac{Y_i}{\theta} = \sum_{i=1}^{n} Z_i$ qui admet la fonction caractéristique

$$\phi_{U_n}(t) = E[e^{itU_n}] \underset{\text{les variables } Z_i \text{ sont indépendants}}{=} \prod_{k=1}^{n} \underbrace{E[e^{itZ_k}]}_{\frac{1}{(1-2it)^{1/2}}} = \frac{1}{(1-2it)^{n/2}}$$

C'est la fonction caractéristique d'une loi du χ_n^2 donc

$$\boxed{U_n \sim \chi_n^2}$$

3) $\alpha = P[\text{Rejeter } H_0 \mid H_0 \text{ vraie}] = P\left[\sum_{i=1}^{n} Y_i > s_\alpha \mid \theta_0 = \theta_0\right]$

$$= P\left[U_n = \frac{T_n}{\theta_0} > \frac{s_\alpha}{\theta_0} \mid \theta = \theta_0\right]$$

donc $\boxed{1 - F_n\left(\frac{s_\alpha}{\theta_0}\right) = \alpha}$ d'où $\boxed{s_\alpha = \theta_0 F_n^{-1}(1-\alpha)}$

$\beta = P[\text{Rejeter } H_1 \mid H_1 \text{ vraie}] = P\left[\sum_{i=1}^{n} Y_i \leq s_\alpha \mid \theta = \theta_1\right]$

$$= P\left[\frac{1}{\theta_1}\sum_{i=1}^{n} Y_i \leq \frac{s_\alpha}{\theta_1} \mid \theta = \theta_1\right]$$

$$\boxed{\beta = F_n\left(\frac{s_\alpha}{\theta_1}\right)}$$

4) Les courbes COR représentent $\pi = 1-\beta$ en fonction de α.
Elles sont définies par

$$\pi = 1 - F_n\left(\frac{s_\alpha}{\theta_1}\right) = \boxed{1 - F_n\left[\frac{\theta_0}{\theta_1} F_n^{-1}(1-\alpha)\right]}$$

Quand $\frac{\theta_1}{\theta_0} \nearrow$ (θ_1 s'éloigne de θ_0), $\frac{\theta_0}{\theta_1} \searrow$ donc $F_n\left[\frac{\theta_0}{\theta_1} F_n^{-1}(1-\alpha)\right] \searrow$

donc $\pi \nearrow$

le test est donc d'autant plus puissant que $\frac{\theta_1}{\theta_0}$ est grand, ce qui est naturel

5) la statistique du test de Kolmogorov se calcule comme suit

$$\boxed{D_n = \sup_{x \in \mathbb{R}} |\hat{F}(x) - F_0(x)| = \max_{i \in \{1,\dots,n\}} \{E_i^+, E_i^-\}}$$

avec $E_i^+ = \left| \frac{i}{n} - F_0(x_i^*) \right|$ et $E_i^- = \left| \frac{i-1}{n} - F_0(x_i^*) \right|$

où F_0 est la fonction de répartition de loi de densité $f(y;\theta)$

La région critique du test est $\boxed{\{(y_1,\dots,y_n) / D_n > s_\alpha\}}$

On a d'après le cours

$$P[\sqrt{n}\, D_n < y] \xrightarrow[n\to+\infty]{} K(y) = \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2 y)$$

et $\boxed{s_\alpha \simeq \frac{1}{\sqrt{n}} K^{-1}(1-\alpha)}$

Quand θ est inconnu, on peut l'estimer à l'aide de la méthode du maximum de vraisemblance et tester si $(y_1,\dots,y_n)$ suit une loi de densité $f(y;\hat{\theta}_{MV})$ où $\hat{\theta}_{MV}$ est l'estimateur du maximum de vraisemblance de θ.

Correction exercice 3

1) Des calculs élémentaires donnent

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

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 $T \sim P(n\lambda_0)$ et sous H_1, on a $T \sim P(n\lambda_1)$.
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 $K_\alpha = n\lambda_0 + \sqrt{n\lambda_0}\Phi^{-1}(\alpha)$.
c) Un calcul simple conduit à

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

c'est-à-dire asymptotiquement

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

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.

Correction exercice 4

1) La vraisemblance de ce problème est

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

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

$$\begin{aligned}\alpha &= P\left[\text{Rejeter } H_0 \middle| H_0 \text{ vraie}\right] \\ &= P\left[\overline{Y} > \nu_\alpha \middle| \theta = \theta_0\right]\end{aligned}$$

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

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

Donc

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

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

$$\begin{aligned}\pi &= P\left[\text{Rejeter } H_0 \middle| H_1 \text{ vraie}\right] \\ &= P\left[\overline{Y} > \nu_\alpha \middle| \theta = \theta_1\right] \\ &= 1 - F\left(\frac{\nu_\alpha-(1-\theta_1)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n}}}\right)\end{aligned}$$

Exercice 5

1) La statistique du test du χ^2 est

$$\phi = \sum_{k=1}^{K} \frac{(n_k - np_k)^2}{np_k}$$

avec $K = 6$ (nombre de classes), $p_k = \frac{1}{6}$ (probabilité de chaque classe sous H_0) et n_k est le nombre d'éléments de la classe k. On a donc

$$\phi = \frac{(15-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(4-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(17-10)^2}{10}.$$

On a donc

$$\phi = \frac{25+9+36+1+16+49}{10} = 13.6.$$

2) La loi de la statistique de test ϕ sous l'hypothèse H_0 est une une loi du chi-deux à $K-1=5$ degrés de liberté.
3) On a

$$\alpha = P[\phi > S_\alpha | \phi \sim \chi^2_5] = 1 - F_{\chi^2_5}(S_\alpha), \text{ soit } S_\alpha = F^{-1}_{\chi^2_5}(1-\alpha).$$

On a $\phi > S_{0.05} = 11.07$ donc on rejette l'hypothèse H_0 (le dé n'est pas truqué) avec le risque $\alpha = 0.05$. Par contre, $\phi < S_{0.01} = 15.09$ donc on accepte l'hypothèse H_0 avec le risque $\alpha = 0.01$.

Correction exercice 6

1. La statistique du test du chi-deux noté ϕ est définie par

$$\phi = \sum_{k=1}^{2} \frac{(Z_k - np_k)^2}{np_k}$$

avec $p_1 = p_2 = 0.5$, $n = 30$, $Z_1 = 10$ et $Z_2 = 20$. Une application numérique donne

$$\phi = \frac{(10-15)^2}{15} + \frac{(20-15)^2}{15} = \frac{10}{3}.$$

2. Sous l'hypothèse H_0, ϕ suit une loi du χ^2 à 1 degré de liberté, i.e., $\phi \sim \chi^2_1$.

3. On a

$$\alpha = P[\text{Rejeter} H_0 | H_0 \text{ vraie}] = P[\phi > S_\alpha | \phi \sim \chi^2_1] = 1 - F_{\chi^2_1}(S_\alpha).$$

donc

$$S_\alpha = F^{-1}_{\chi^2_1}(1-\alpha).$$

Pour $\alpha = 0.05$, on trouve $S_{0.05} = 3.84 > \phi$ donc on accepte H_0 (pas de préférence pour le thé ou le café) avec $\alpha = 0.05$.

Exercice 1 **de révision**
(corrections sur page perso, examen MFEE année 2022-2023)

On considère un ensemble de variables aléatoires $(X_1, ..., X_n)$ indépendantes et de même loi beta $B(1, \theta)$ de densité

$$f(x_i; \theta) = \begin{cases} \theta(1 - x_i)^{\theta - 1} \text{ si } x_i \in]0, 1[\\ 0 \quad \text{sinon} \end{cases}$$

avec $\theta > 0$. On désire tester les deux hypothèses

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

1. À l'aide du théorème de Neyman-Pearson, montrer que la statistique du test le plus puissant est $T_n = -\sum_{i=1}^n \ln(1 - X_i)$ et indiquer la région critique de ce test.

2. On admet que si X_i suit une loi de densité $f(x_i; \theta)$, alors $Y_i = -\ln(1 - X_i)$ suit une loi gamma $\mathcal{G}(1, \theta)$. En utilisant le théorème central limite, déterminer la loi asymptotique de T_n sous les deux hypothèses.

3. En utilisant la loi asymptotique de la question précédente, déterminer les risques de première et seconde espèce α et β du test en fonction des paramètres θ_0 et θ_1 et de la fonction de répartition d'une loi normale $\mathcal{N}(0, 1)$ notée F.

4. Déterminer les courbes COR associées à ce test montrer qu'elles ne dépendent que de n et de $\frac{\theta_1}{\theta_0}$. Tracer l'allure de ces courbes COR pour différentes valeurs de n dans une première figure et pour différentes valeurs de $\frac{\theta_1}{\theta_0}$ dans une seconde figure.

5. On désire vérifier que l'ensemble des observations $(x_1, ..., x_n)$ suit une loi beta $B(1, \theta)$ avec le paramètre $\theta = \frac{1}{2}$ à l'aide d'un test du χ^2. Déterminer la fonction de répartition de cette loi et en déduire que l'intervalle $]0, 1[$ est la réunion de trois intervalles équiprobables pour la loi $B\left(1, \frac{1}{2}\right)$ que l'on précisera. On compte le nombre d'observations x_i appartenant à ces trois intervalles et on trouve $K_1 = 13$, $K_2 = 8$ et $K_3 = 9$. Quelle est la valeur de la statistique du test du $\chi 2$? Exprimer le seuil de ce test noté S_α en fonction du risque α et de l'inverse de la fonction de répartition d'une loi du χ^2 dont on précisera le nombre de degrés de liberté. On donne $S_{0.05} = 5.991$. Qu'en conclut-on ?

Exercice 2 **de révision**
(corrections sur page perso, examen MFEE année 2021-2022)

On considère un triplet de variables aléatoires (N_1, N_2, N_3) de loi multinomiale définie par $P[N_1 = n_1, N_2 = n_2, N_3 = n_3] = \frac{n!}{n_1!n_2!n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}$ avec $n_1 + n_2 + n_3 = n$ et $\boldsymbol{p} = (p_1, p_2, p_3) = (\theta, \theta, 1 - 2\theta)$. On admettra que pour un tel modèle statistique, la loi marginale de N_i est une loi binomiale $B(n, p_i)$ de moyenne $E[N_i] = np_i$ et de variance $\text{var}(N_i) = np_i(1 - p_i)$. On veut à l'aide de l'observation de (N_1, N_2, N_3) tester les deux hypothèses suivantes

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1 \qquad \text{avec } \theta_1 > \theta_0.$$

1. À l'aide du théorème de Neyman-Pearson, montrer que la statistique du test le plus puissant est $T_n = N_3$ (on n'oubliera pas que $N_1 + N_2 + N_3 = n$) et indiquer la région critique de ce test.

2. En utilisant le théorème central limite, déterminer la loi asymptotique de T_n sous les deux hypothèses.

3. En utilisant la loi asymptotique de la question précédente, déterminer les risques de première et seconde espèce α et β du test en fonction des paramètres θ_0 et θ_1 et de la fonction de répartition d'une loi normale $\mathcal{N}(0, 1)$ notée F.

4. Déterminer les courbes COR associées à ce test et tracer l'allure de ces courbes pour différentes valeurs de n.

5. On désire à partir d'un ensemble de $K = 10$ observations de N_1 notées $(n_{11}, ..., n_{1K})$ et données dans le tableau ci-dessous déterminer s'il est raisonnable de supposer que N_1 suit une loi binomiale $B(K, \theta)$.

2	3	3	2	2	0	4	3	1	0

- Expliquer pourquoi on est amené à choisir $\theta = 0.2$ pour effectuer ce test.
- Les probabilités associées à une loi binomiale $B(K, \theta)$ avec $\theta = 0.2$ et $K = 10$ sont données dans le tableau ci-dessous

p_0	p_1	p_2	p_3	p_4	$\sum_{k=5}^{10} p_k$
0.11	0.27	0.30	0.20	0.09	0.03

  On décide de faire un test du χ^2 avec 3 classes notées C_1, C_2 et C_3. Justifier le choix suivant pour ces trois classes : $C_1 = \{0, 1\}$, $C_2 = \{2\}$ et $C_1 = \{3, ..., 10\}$.
- Exprimer la statistique de ce test notée ϕ en fonction des données du problème (sans chercher à la calculer) et préciser la région critique.
- Quelle est la loi asymptotique de ϕ ?
- Les seuils obtenus pour $\alpha = 0.05$ et $\alpha = 0.01$ sont $S_{0.05} = 5.99$ et $S_{0.01} = 9.21$. Expliquer pourquoi $S_{0.05} < S_{0.01}$.