Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2020/2021

Corrigé de l'examen de remplacement (1h)

Exercice 1. (08pts)

Donner la région critique du:

1. test, unilatéral à droite, de la moyenne d'une population gaussienne de variance connue? (2pts)
2. test, bilatéral, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances connues? (2pts)
3. test, unilatéral à gauche, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues? (2pts)
4. test, unilatéral à droite, de comparaisons de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues supposées égales? (2pts)

* * * * * * * * * * * * * * * * * *

Exercice 2. (12 pts)

Soit $(X_1, ..., X_{12})$ un échantillon d'une population X normale centrée de variance σ^2. On s'intéresse au test suivant:

$$\begin{cases} H_0: & \sigma^2 \leq 2 \\ H_1: & \sigma^2 > 2 \end{cases}.$$

1. Quel test s'agit-il: unilatéral ou bilatéral? $(1pt)$
2. Donner la statistique de test à utiliser. $(2pts)$
3. Au niveau de signification $\alpha = 0.05$, donner la région critique de ce test. $(3pts)$
4. Donner la fonction puissance de ce test. $(3pts)$
5. Tracer soigneusement le graphe de la fonction puissance. $(3pts)$

* *

1 Solution de l'exercice 1

1. Le test, unilatéral à droite, de la moyenne d'une population gaussienne de variance connue σ, est

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0.$$

La région critique associeé est:

$$W = \left\{ (x_1, ..., x_n) \in \mathbb{R}^n \mid \sqrt{n}\frac{\overline{x} - \mu_0}{\sigma} \geq z_{1-\alpha} \right\},$$

où $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ est le quantile d'ordre $1-\alpha$ de la loi normale centrée-réduite.

2. Test, bilatéral, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances connues, est:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

avec σ_1 est σ_2 connues. La région critique associée est:

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(n_1)}; x_2^{(1)}, ..., x_2^{(n_2)} \right) \in \mathbb{R}^{n_1+n_2} \mid \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \geq z_{1-\alpha/2} \right\},$$

où $z_{1-\alpha/2}$ (la valeur critique) est le quantile d'ordre $1-\alpha/2$ de $\mathcal{N}(0,1)$, c'est à dire $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$.

3. Test, unilatéral à gauche, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues, est (le test de Welch)

$$H_0 : \mu_1 \geq \mu_2$$
$$H_1 : \mu_1 < \mu_2$$

La région critique associée est

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(n_1)}; x_2^{(1)}, ..., x_2^{(n_2)} \right) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\widetilde{s}_1^2/n_1 + \widetilde{s}_2^2/n_2}} \leq t_\alpha(\nu) \right\},$$

où $t_\alpha(\nu)$ (la valeur critique) est le quantile d'ordre α de $t(\nu)$ (student), où

$$\nu := \frac{\left(\widetilde{s}_1^2/n_1 + \widetilde{s}_2^2/n_2\right)^2}{\widetilde{s}_1^4/\left(n_1^2(n_1-1)\right) + \widetilde{s}_2^4/\left(n_2^2(n_2-1)\right)}.$$

4. Test, unilatéral à droite, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues supposé égales, est

$$H_0 : \mu_1 \leq \mu_2$$
$$H_1 : \mu_1 > \mu_2.$$

avec σ_1 est σ_2 in connues. La région critique associée est:

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(n_1)}; x_2^{(1)}, ..., x_2^{(n_2)} \right) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{x}_1 - \overline{x}_2}{\widetilde{s}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{1-\alpha} \right\},$$

où $t_{1-\alpha}$ (la valeur critique) est le quantile d'ordre $1-\alpha$ de $t(n_1+n_2-2)$.

2 Solution de l'exercice 2

C'est un exercice résolut de TD.

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2020/2021

Corrigé de l'examen de remplacement (1h)

Exercice 1. (08pts)

Donner la région critique du:

1. test, unilatéral à droite, de la moyenne d'une population gaussienne de variance connue? (2pts)
2. test, bilatéral, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances connues? (2pts)
3. test, unilatéral à gauche, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues? (2pts)
4. test, unilatéral à droite, de comparaisons de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues supposées égales? (2pts)

* * * * * * * * * * * * * * * * * *

Exercice 2. (12 pts)

Soit $(X_1, ..., X_{12})$ un échantillon d'une population X normale centrée de variance σ^2. On s'intéresse au test suivant:

$$\begin{cases} H_0: & \sigma^2 \leq 2 \\ H_1: & \sigma^2 > 2 \end{cases}.$$

1. Quel test s'agit-il: unilatéral ou bilatéral? $(1pt)$
2. Donner la statistique de test à utiliser. $(2pts)$
3. Au niveau de signification $\alpha = 0.05$, donner la région critique de ce test. $(3pts)$
4. Donner la fonction puissance de ce test. $(3pts)$
5. Tracer soigneusement le graphe de la fonction puissance. $(3pts)$

* *

1 Solution de l'exercice 1

1. Le test, unilatéral à droite, de la moyenne d'une population gaussienne de variance connue σ, est

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0.$$

La région critique associeé est:

$$W = \left\{ (x_1, ..., x_n) \in \mathbb{R}^n \mid \sqrt{n}\frac{\overline{x} - \mu_0}{\sigma} \geq z_{1-\alpha} \right\},$$

où $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ est le quantile d'ordre $1-\alpha$ de la loi normale centrée-réduite.

2. Test, bilatéral, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances connues, est:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

avec σ_1 est σ_2 connues. La région critique associée est:

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(n_1)}; x_2^{(1)}, ..., x_2^{(n_2)} \right) \in \mathbb{R}^{n_1+n_2} \mid \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \geq z_{1-\alpha/2} \right\},$$

où $z_{1-\alpha/2}$ (la valeur critique) est le quantile d'ordre $1-\alpha/2$ de $\mathcal{N}(0,1)$, c'est à dire $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$.

3. Test, unilatéral à gauche, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues, est (le test de Welch)

$$H_0 : \mu_1 \geq \mu_2$$
$$H_1 : \mu_1 < \mu_2$$

La région critique associée est

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(n_1)}; x_2^{(1)}, ..., x_2^{(n_2)} \right) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\widetilde{s}_1^2/n_1 + \widetilde{s}_2^2/n_2}} \leq t_\alpha(\nu) \right\},$$

où $t_\alpha(\nu)$ (la valeur critique) est le quantile d'ordre α de $t(\nu)$ (student), où

$$\nu := \frac{\left(\widetilde{s}_1^2/n_1 + \widetilde{s}_2^2/n_2\right)^2}{\widetilde{s}_1^4/\left(n_1^2(n_1-1)\right) + \widetilde{s}_2^4/\left(n_2^2(n_2-1)\right)}.$$

4. Test, unilatéral à droite, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues supposé égales, est

$$H_0 : \mu_1 \leq \mu_2$$
$$H_1 : \mu_1 > \mu_2.$$

avec σ_1 est σ_2 in connues. La région critique associée est:

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(n_1)}; x_2^{(1)}, ..., x_2^{(n_2)} \right) \in \mathbb{R}^{n_1+n_2} \mid \frac{\overline{x}_1 - \overline{x}_2}{\widetilde{s}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{1-\alpha} \right\},$$

où $t_{1-\alpha}$ (la valeur critique) est le quantile d'ordre $1-\alpha$ de $t(n_1+n_2-2)$.

2 Solution de l'exercice 2

C'est un exercice résolut de TD.

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2020/2021

Corrigé-type de l'examen

Exercice 1. (12pts)

La statistique de test à utiliser, sous l'hypothèse nulle, pour:

1. tester la variance d'une population gaussienne de moyenne inconnue, est: (2pts)

$$\frac{n\widetilde{S}^2}{\sigma_0^2} \rightsquigarrow \chi^2_{n-1},$$

où $\widetilde{S}^2 := n^{-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$ et χ^2_{n-1} désigne la loi de qui-deux à $(n-1)$ degré de liberté (ddl).

2. comparer deux moyennes de deux populations gaussiennes indépendantes de variances connues, est: (2pts)

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}\left(0,1\right) \text{ (loi normale centrée-réduite)},$$

où $\overline{X}_j := n_j^{-1}\sum_{i=1}^{n_j} X_j^{(i)}$, $j = 1, 2$.

3. comparer deux moyennes de deux populations gaussiennes indépendantes de variances inconnues, est: (2pts)

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\widetilde{S}_1^2/n_1 + \widetilde{S}_2^2/n_2}} \simeq t\left(\nu\right),$$

où $t\left(\nu\right)$ est la loi de Student à

$$\nu := \frac{\left(\widetilde{s}_1^2/n_1 + \widetilde{s}_2^2/n_2\right)^2}{\widetilde{s}_1^4/\left(n_1^2\left(n_1-1\right)\right) + \widetilde{s}_2^4/\left(n_2^2\left(n_2-1\right)\right)}, \tag{1}$$

ddl, et $\widetilde{s}_i^2$ désigne la valeur observée de $\widetilde{S}_i^2$.

4. comparer deux moyennes de deux populations gaussiennes indépendantes de variances inconnues supposées égales, est: (2pts)

$$\frac{\overline{X}_1 - \overline{X}_2}{\widetilde{S}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow t\left(n_1 + n_2 - 2\right),$$

où

$$\widetilde{S}^2 = \frac{n_1\widetilde{S}_1^2 + n_2\widetilde{S}_2^2}{n_1 + n_2 - 2},$$

est un estimateur sans biais de $\sigma_1^2 = \sigma_2^2$ et $t\left(n_1+n_2-2\right)$ est la loi de Student à n_1+n_2-2 dll.

5. comparer deux variances de deux populations gaussiennes indépendantes, est: (2pts)

$$\frac{\frac{n_1\widetilde{S}_1^2/\sigma_1^2}{n_1-1}}{\frac{n_2\widetilde{S}_2^2/\sigma_2^2}{n_2-1}} \rightsquigarrow F\left(n_1-1, n_2-1\right), \tag{2}$$

où $F\left(n_1-1, n_2-1\right)$ désigne la loi de Fisher à $\left(n_1-1, n_2-1\right)$ dll.

6. comparer deux proportions de deux populations indépendantes, est: (2pts)

$$\frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}\left(1-\widehat{p}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \rightsquigarrow \mathcal{N}\left(0,1\right),$$

où

$$\widehat{p} := \frac{n_1 \widehat{p}_1 + n_2 \widehat{p}_2}{n_1 + n_2},$$

et l'estimateur commun de $p_1$et p_2, à condition que

$$n_1 + n_2 > 30,\ n_i p_i \geq 5 \text{ et } n_i (1 - p_i) \geq 5,\ \text{pour } i = 1, 2.$$

Exercice 2. (08pts)

Ici**,** la variable aléatoire X prend une valeur 1 (qui correspond à un article défectueux) avec une probabilité p et elle prend la valeur 0 (qui correspond à un article non défectueux). Donc il s'agit d'une v.a qui suit la loi de Bernoulli de paramètre p. La fonction masse de cette v.a discrète X et définie comme suit:

$$\mathbf{P}(X = 1) = p \text{ et } \mathbf{P}(X = 0) = 1 - p.$$

Celle-ci peut être reformuler comme suit:

$$\mathbf{P}(X = x) = p^x (1 - p)^{1-x},\ x = 0, 1.$$

Nous allons appliquer la méthode du test de *rapport de vraisemblance monotone***.** Soit $p_1 > p_2$ et écrivons

$$\frac{L_{p_1}}{L_{p_2}} = \frac{\prod_{i=1}^{16} p_1^{x_i} (1 - p_1)^{1-x_i}}{\prod_{i=1}^{16} p_2^{x_i} (1 - p_2)^{1-x_i}} = \frac{p_1^t (1 - p_1)^{16-t}}{p_2^t (1 - p_2)^{16-t}},\ \text{avec } t := \sum_{i=1}^{16} x_i.$$

Celle-ci peut être réécrite comme suit:

$$\frac{L_{p_1}}{L_{p_2}} = \left(\frac{1 - p_1}{1 - p_2}\right)^{16} \left(\frac{p_1 (1 - p_2)}{p_2 (1 - p_1)}\right)^t.$$

On pose $b := \left(\frac{1 - p_1}{1 - p_2}\right)^{16} > 0$, et $a := \frac{p_1(1-p_2)}{p_2(1-p_1)} > 1$ (car $p_1 > p_2$), ainsi $t \to \frac{L_{p_1}}{L_{p_2}}(t) = ba^t$ est une fonction croissante en t (car $a > 1$). Nous allons alors appliquer la proposition 3 (voir le cours), pour avoir le test upp :

$$\delta = \begin{cases} 1 & \text{si } \sum_{i=1}^{16} x_i > c \\ \gamma & \text{si } \sum_{i=1}^{16} x_i = c \\ 0 & \text{si } \sum_{i=1}^{16} x_i < c \end{cases}$$

où c et $0 < \gamma < 1$ sont telles que

$$\mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16} X_i > c\right) + \gamma \mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16} X_i = c\right) = \alpha = 0.02.$$

En d'autres termes

$$1 - \mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16} X_i \leq c\right) + \gamma \mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16} X_i = c\right) = 0.02,$$

ce qui implique que

$$\mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16} X_i \leq c\right) = \gamma \mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16} X_i = c\right) + 0.98 > 0.98.$$

On note que $Y := \sum_{i=1}^{16} X_i \rightsquigarrow$Binomial$(16, 0.2)$ (la loi binomiale de paramètre $n = 16$ et $p = 0.2$). De la table statistique, de la loi binomiale, la plus petite c telle que $\mathbf{P}(Y \leq c) > 0.98$ est $c = 7$, pour la quelle $\mathbf{P}(Y \leq 7) = 0.993$. Par conséquent

$$\begin{aligned}\gamma &= \frac{\mathbf{P}(Y \leq 7) - 0.98}{\mathbf{P}(Y = 7)} \\ &= \frac{0.993 - 0.98}{\mathbf{P}(Y \leq 7) - \mathbf{P}(Y \leq 6)} \\ &= \frac{0.993 - 0.98}{0.993 - \mathbf{P}(Y \leq 6)}.\end{aligned}$$

De la table statistique, de la loi binomiale on trouve $\mathbf{P}(Y \leq 6) = 0.973$, donc

$$\gamma = \frac{0.993 - 0.98}{0.993 - 0.973} = 0.65.$$

Ainsi le test optimal upp est

$$\delta = \delta(X_1, ..., X_{16}) = \begin{cases} 1 & \text{si } \sum_{i=1}^{16} x_i > 7 \\ 0.65 & \text{si } \sum_{i=1}^{16} x_i = 7 \quad (4pts) \\ 0 & \text{si } \sum_{i=1}^{16} x_i < 7 \end{cases}$$

La fonction puissance du test est

$$\begin{aligned}\pi(p) &= \mathbf{E}_p[\delta], \ 0 < p < 1 \\ &= 1 \times \mathbf{P}(\delta = 1) + 0.65 \times \mathbf{P}(\delta = 0.65) + 0 \times \mathbf{P}(\delta = 0).\end{aligned}$$

En d'autres termes

$$\pi(p) = \mathbf{P}(T > 7) + 0.65\mathbf{P}(T = 7),$$

où $T := \sum_{i=1}^{16} X_i$ est une v.a binomiale de parametrer $(16, p)$, avec $0 < p < 1$. Il est claire que:

$$\begin{aligned}\pi(p) &= 1 - \mathbf{P}(T \leq 7) + 0.65(\mathbf{P}(T \leq 7) - \mathbf{P}(T \leq 6)) \\ &= 1 - 0.35\mathbf{P}(T \leq 7) - 0.65\mathbf{P}(T \leq 6) \\ &= 1 - 0.35F_p(7) - 0.65F_p(6),\end{aligned}$$

où $F_p(x)$ désigne la fonction de répartition de la loi binomiale de paramètrer $(16, p)$. Explicitement la fonction puissance est

$$\pi(p) = 1 - 0.35\sum_{k=0}^{7} C_{16}^k p^k (1-p)^{16-k} - 0.65\sum_{k=0}^{6} C_{16}^k p^k (1-p)^{16-k}, \ \ (4pts)$$

pour $0 < p < 1$.

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2020/2021

Corrigé-type de l'examen

Exercice 1. (12pts)

La statistique de test à utiliser, sous l'hypothèse nulle, pour:

1. tester la variance d'une population gaussienne de moyenne inconnue, est: (2pts)

$$\frac{n\widetilde{S}^2}{\sigma_0^2} \rightsquigarrow \chi^2_{n-1},$$

où $\widetilde{S}^2 := n^{-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$ et χ^2_{n-1} désigne la loi de qui-deux à $(n-1)$ degré de liberté (ddl).

2. comparer deux moyennes de deux populations gaussiennes indépendantes de variances connues, est: (2pts)

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0,1) \text{ (loi normale centrée-réduite)},$$

où $\overline{X}_j := n_j^{-1}\sum_{i=1}^{n_j} X_j^{(i)}$, $j = 1, 2$.

3. comparer deux moyennes de deux populations gaussiennes indépendantes de variances inconnues, est: (2pts)

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\widetilde{S}_1^2/n_1 + \widetilde{S}_2^2/n_2}} \simeq t(\nu),$$

où $t(\nu)$ est la loi de Student à

$$\nu := \frac{\left(\widetilde{s}_1^2/n_1 + \widetilde{s}_2^2/n_2\right)^2}{\widetilde{s}_1^4/\left(n_1^2(n_1-1)\right) + \widetilde{s}_2^4/\left(n_2^2(n_2-1)\right)}, \tag{1}$$

ddl, et $\widetilde{s}_i^2$ désigne la valeur observée de $\widetilde{S}_i^2$.

4. comparer deux moyennes de deux populations gaussiennes indépendantes de variances inconnues supposées égales, est: (2pts)

$$\frac{\overline{X}_1 - \overline{X}_2}{\widetilde{S}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow t(n_1 + n_2 - 2),$$

où

$$\widetilde{S}^2 = \frac{n_1\widetilde{S}_1^2 + n_2\widetilde{S}_2^2}{n_1 + n_2 - 2},$$

est un estimateur sans biais de $\sigma_1^2 = \sigma_2^2$ et $t(n_1 + n_2 - 2)$ est la loi de Student à $n_1 + n_2 - 2$ dll.

5. comparer deux variances de deux populations gaussiennes indépendantes, est: (2pts)

$$\frac{\frac{n_1\widetilde{S}_1^2/\sigma_1^2}{n_1-1}}{\frac{n_2\widetilde{S}_2^2/\sigma_2^2}{n_2-1}} \rightsquigarrow F(n_1 - 1, n_2 - 1), \tag{2}$$

où $F(n_1 - 1, n_2 - 1)$ désigne la loi de Fisher à $(n_1 - 1, n_2 - 1)$ dll.

6. comparer deux proportions de deux populations indépendantes, est: (2pts)

$$\frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \rightsquigarrow \mathcal{N}(0,1),$$

où

$$\widehat{p} := \frac{n_1\widehat{p}_1 + n_2\widehat{p}_2}{n_1+n_2},$$

et l'estimateur commun de $p_1$et p_2, à condition que

$$n_1+n_2>30,\ n_ip_i \geq 5 \text{ et } n_i(1-p_i)\geq 5,\ \text{pour } i=1,2.$$

Exercice 2. (08pts)

Ici**,** la variable aléatoire X prend une valeur 1 (qui correspond à un article défectueux) avec une probabilité p et elle prend la valeur 0 (qui correspond à un article non défectueux). Donc il s'agit d'une v.a qui suit la loi de Bernoulli de paramètre p. La fonction masse de cette v.a discrète X et définie comme suit:

$$\mathbf{P}(X=1)=p \text{ et } \mathbf{P}(X=0)=1-p.$$

Celle-ci peut être reformuler comme suit:

$$\mathbf{P}(X=x)=p^x(1-p)^{1-x},\ x=0,1.$$

Nous allons appliquer la méthode du test de *rapport de vraisemblance monotone***.** Soit $p_1>p_2$ et écrivons

$$\frac{L_{p_1}}{L_{p_2}} = \frac{\prod_{i=1}^{16} p_1^{x_i}(1-p_1)^{1-x_i}}{\prod_{i=1}^{16} p_2^{x_i}(1-p_2)^{1-x_i}} = \frac{p_1^t(1-p_1)^{16-t}}{p_2^t(1-p_2)^{16-t}},\ \text{avec } t:=\sum_{i=1}^{16}x_i.$$

Celle-ci peut être réécrite comme suit:

$$\frac{L_{p_1}}{L_{p_2}} = \left(\frac{1-p_1}{1-p_2}\right)^{16}\left(\frac{p_1(1-p_2)}{p_2(1-p_1)}\right)^t.$$

On pose $b:=\left(\dfrac{1-p_1}{1-p_2}\right)^{16}>0$, et $a:=\frac{p_1(1-p_2)}{p_2(1-p_1)}>1$ (car $p_1>p_2$), ainsi $t\to\frac{L_{p_1}}{L_{p_2}}(t)=ba^t$ est une fonction croissante en t (car $a>1$). Nous allons alors appliquer la proposition 3 (voir le cours), pour avoir le test upp :

$$\delta = \begin{cases} 1 & \text{si } \sum_{i=1}^{16}x_i > c \\ \gamma & \text{si } \sum_{i=1}^{16}x_i = c \\ 0 & \text{si } \sum_{i=1}^{16}x_i < c \end{cases}$$

où c et $0<\gamma<1$ sont telles que

$$\mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16}X_i>c\right)+\gamma\mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16}X_i=c\right)=\alpha=0.02.$$

En d'autres termes

$$1-\mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16}X_i\leq c\right)+\gamma\mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16}X_i=c\right)=0.02,$$

ce qui implique que

$$\mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16}X_i\leq c\right)=\gamma\mathbf{P}_{p=0.2}\left(\sum_{i=1}^{16}X_i=c\right)+0.98>0.98.$$

On note que $Y := \sum_{i=1}^{16} X_i \rightsquigarrow$Binomial$(16, 0.2)$ (la loi binomiale de paramètre $n = 16$ et $p = 0.2$). De la table statistique, de la loi binomiale, la plus petite c telle que $\mathbf{P}(Y \leq c) > 0.98$ est $c = 7$, pour la quelle $\mathbf{P}(Y \leq 7) = 0.993$. Par conséquent

$$\begin{aligned}\gamma &= \frac{\mathbf{P}(Y \leq 7) - 0.98}{\mathbf{P}(Y = 7)} \\ &= \frac{0.993 - 0.98}{\mathbf{P}(Y \leq 7) - \mathbf{P}(Y \leq 6)} \\ &= \frac{0.993 - 0.98}{0.993 - \mathbf{P}(Y \leq 6)}.\end{aligned}$$

De la table statistique, de la loi binomiale on trouve $\mathbf{P}(Y \leq 6) = 0.973$, donc

$$\gamma = \frac{0.993 - 0.98}{0.993 - 0.973} = 0.65.$$

Ainsi le test optimal upp est

$$\delta = \delta(X_1, ..., X_{16}) = \begin{cases} 1 & \text{si } \sum_{i=1}^{16} x_i > 7 \\ 0.65 & \text{si } \sum_{i=1}^{16} x_i = 7 \quad (4pts) \\ 0 & \text{si } \sum_{i=1}^{16} x_i < 7 \end{cases}$$

La fonction puissance du test est

$$\begin{aligned}\pi(p) &= \mathbf{E}_p[\delta], \ 0 < p < 1 \\ &= 1 \times \mathbf{P}(\delta = 1) + 0.65 \times \mathbf{P}(\delta = 0.65) + 0 \times \mathbf{P}(\delta = 0).\end{aligned}$$

En d'autres termes

$$\pi(p) = \mathbf{P}(T > 7) + 0.65\mathbf{P}(T = 7),$$

où $T := \sum_{i=1}^{16} X_i$ est une v.a binomiale de parametrer $(16, p)$, avec $0 < p < 1$. Il est claire que:

$$\begin{aligned}\pi(p) &= 1 - \mathbf{P}(T \leq 7) + 0.65(\mathbf{P}(T \leq 7) - \mathbf{P}(T \leq 6)) \\ &= 1 - 0.35\mathbf{P}(T \leq 7) - 0.65\mathbf{P}(T \leq 6) \\ &= 1 - 0.35F_p(7) - 0.65F_p(6),\end{aligned}$$

où $F_p(x)$ désigne la fonction de répartition de la loi binomiale de paramètrer $(16, p)$. Explicitement la fonction puissance est

$$\pi(p) = 1 - 0.35 \sum_{k=0}^{7} C_{16}^k p^k (1-p)^{16-k} - 0.65 \sum_{k=0}^{6} C_{16}^k p^k (1-p)^{16-k}, \ \ (4pts)$$

pour $0 < p < 1$.

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2021/2022
Module: Tests Statistiques

Corrigé-type de l'examen

Solution de l'exercice 1. ($10pts$)

1) Donner la statistique du test ($1pt$) est sa loi de probabilité ($1pt$), pour:
a. un test de la moyenne d'une population gaussienne de variance connue:

$$T := \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \overset{\mathcal{D}}{=} \mathcal{N}(0,1) \text{ (sous } \mu = \mu_0). \ (2pts)$$

b. un test de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances connues:

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \overset{\mathcal{D}}{=} \mathcal{N}(0,1) \text{ (sous } \mu_1 = \mu_2). \ (2pts)$$

c. un test de comparaison de deux moyennes de deux populations gaussiennes indépendantes telles que les tailles des deux échantillons égales à $n_1 = 25 > 20$ et $n_2 = 27 > 0$ respectivement:

$$\frac{\overline{X}_1 - \overline{X}_2}{\widetilde{S}\sqrt{\frac{1}{25} + \frac{1}{27}}} \rightsquigarrow t(25 + 27 - 2) \text{ (Student à 50 ddl).} \ (2pts)$$

Remarque: Dans le cours, j'ai dit que quand les deux variances sont les supérieures à 20, on peut supposer que les variances sont égales.

d. un test de comparaisons de deux variances de deux populations gaussiennes indépendantes:

$$\frac{\frac{n_1\widetilde{S}_1^2/\sigma_1^2}{n_1-1}}{\frac{n_2\widetilde{S}_2^2/\sigma_2^2}{n_2-1}} \rightsquigarrow F(n_1 - 1, n_2 - 1), \ \left(\text{sous } H_0 : \sigma_1^2 = \sigma_2^2\right), \ (2pts)$$

loi de Fisher à $(n_1 - 1, n_2 - 1)$ ddl.

2. Donner la définition du rapport de vraisemblance maximale ($1pt$) et dans quel cas on utilise ce dernier ($1pt$):

$$R = R(x_1, ..., x_n) = \frac{\sup_{\theta\in\Theta_1} L(x_1, ..., x_n; \theta)}{\sup_{\theta\in\Theta_0} L(x_1, ..., x_n; \theta)}. \ (2pts)$$

Ce test est utile surtout là où les méthodes précédentes sont échoué. Il s'applique aussi le cas où le paramètre θ est vectoriel: $\theta = (\theta_1, ..., \theta_p) \in \Theta \subset \mathbb{R}^p$.

* *

Solution de l'exercice 2. ($10pts$)

1) Il s'agit ici d'un test bilatéral. ($1pt$)

2) L'hypothèse nulle H_0 et l'hyppthèse alternative H_1 sont

$$\begin{cases} H_0 : & \sigma_1^2 = \sigma_2^2 \\ H_1 : & \sigma_1^2 \neq \sigma_2^2 \end{cases} \ (2pts)$$

Nous avons $\alpha = 0.1$, $n_1 = 21$, $n_2 = 9$, $\widetilde{s}_{1,obs}^2 = 5.20$, $\widetilde{s}_{2,obs}^2 = 2.30$.

3) La statistique de test (variable de décision) à utiliser est:

$$\frac{\frac{21\widetilde{S}_1^2/\sigma_1^2}{20}}{\frac{9\widetilde{S}_2^2/\sigma_2^2}{8}} \rightsquigarrow F(20, 8), \ \left(\text{sous } H_0 : \sigma_1^2 = \sigma_2^2\right), \ (2pts)$$

où $F(20, 8)$ désigne la loi de Fisher à $(20, 8)$ degrés de liberté.

4) A un seuil de signification $\alpha = 10\%$, la région critique associée est

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(21)}; x_2^{(1)}, ..., x_2^{(9)} \right) \in \mathbb{R}^{30} \mid \frac{\widetilde{s}_1^2}{\widetilde{s}_2^2} \geq \frac{9 \times 20}{21 \times 8} c_1 \text{ ou } \frac{\widetilde{s}_1^2}{\widetilde{s}_2^2} \leq \frac{9 \times 20}{21 \times 8} c_2 \right\},$$

où c_1 et c_2 sont deux constantes telles que $\mathbf{P}\left(F\left(20,8\right) \geq c_1\right) = \mathbf{P}\left(F\left(20,8\right) \leq c_2\right) = 0.1/2 = 0.05$. De la table statistique de la loi Fisher on obtient $c_1 = 3.15$. Pour trouver la valeur de c_2, on utilise la formule suivante:

$$F\left(\nu_1, \nu_2\right) \overset{\mathcal{D}}{=} \frac{1}{F\left(\nu_2, \nu_1\right)}.$$

Donc $\mathbf{P}\left(F\left(20,8\right) \leq c_2\right) = 0.05 \Longleftrightarrow \mathbf{P}\left(F\left(8,20\right) \geq 1/c_2\right) = 0.025$. De la table statistique on obtient $1/c_2 = 2.45$ ce qui donne $c_2 = 0.408$. La région critique est donc

$$W = \left\{ \left(x_1^{(1)}, ..., x_1^{(21)}; x_2^{(1)}, ..., x_2^{(9)} \right) \in \mathbb{R}^{30} \mid \frac{\widetilde{s}_1^2}{\widetilde{s}_2^2} \geq 3.375 \text{ ou } \frac{\widetilde{s}_1^2}{\widetilde{s}_2^2} \leq 0.437 \right\}. \ (4pts)$$

5) Nous avons $\widetilde{s}_{1,obs}^2 / \widetilde{s}_{2,obs}^2 = 5.20/2.30 = 2.2609$. Cette valeur est ni ≥ 3.375 ni ≤ 0.437, donc on accepte l'égalité des deux variances, c'est à dire $\sigma_1^2 = \sigma_2^2$. $(1pt)$

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2022/2023
Module: tests statistiques

Corrigé-type de l'examen

Exercice 1. (12pts)

Soit X une variable aléatoire (v.a) qui suit une loi de probabilité de densité

$$f_\theta(x) = \frac{1}{\theta} x^{1/\theta - 1} \mathbb{I}_{\{0<x<1\}},\ \theta > 0,$$

où $\mathbb{I}_A$ désigne l'indicatrice de l'ensemble A.

1. Montrer que $-\log X$ est une v.a. exponentielle de paramètre θ. (**2pt**)
2. Etant donné $X_1, ..., X_n$ un échantillon aléatoire de X, déterminer la loi de probabilité de $T := -\sum_{i=1}^n \log X_i$. (**1pt**)
3. Déterminer l'estimateur du maximum de vraisemblance de θ, noté $\widehat{\theta}_{EMV}$, basé sur $X_1, ..., X_n$. (**2pts**)
4. Considérons le test suivant: $H_0 : \theta = 0.5$ contre $H_1 : \theta \neq 0.5$. Montrer que le rapport de vraisemblance généralisé associé à ce test égal à

$$R_1 = \left\{ \frac{1}{2\widehat{\theta}_{EMV}} \exp\left(\frac{2}{\widehat{\theta}_{EMV}} - \frac{1}{\widehat{\theta}^2_{EMV}} \right) \right\}^n. \text{ (\textbf{2pts})}$$

   Aide: utiliser la relation $a \times b = \exp(\log a + \log b)$, pour $a, b > 0$.
5. Au niveau de signification $\alpha = 0.05$ et pour un échantillon de taille $n = 10$, déterminer la région critique correspondante. (**3pts**)
6. Etant donné l'échantillon observé suivant:

   0.845; 0.612; 0.771; 0.977; 0.926; 0.859; 0.558; 0.910; 0.822; 0.740.
7. Peut-on dire que X suit la loi de probabilité de densité $f_{0.5}$? (**2pt**)

* * * * * * * * * **

Exercice 2. (8pts)

Pour tester, au niveau $\alpha = 0.05$, l'efficacité d'un traitement destiné à augmenter le rythme cardiaque, on a mesuré sur 5 individus ce rythme (en moyenne), avant et après l'administration du traitement:

avant	80	90	70.3	85	63
après	84	95.5	73.5	86	62

Est-il efficace ?

On suppose que le rythme cardiaque se répartit de façon quasi gaussienne pour la population considérée (avant comme après traitement) et que les variances des deux populations sont égales.

(Aide: test de comparaison de deux moyennes $\mu_1 \leq \mu_2$)

Solution

Solution de l'exercice 1. (**12pts**)

1) On montre facilement que la fonction de répartition de X est $F_\theta(x) = x^{1/\theta} \mathbb{I}_{\{0<x<1\}}$, $\theta > 0$. La v.a X est définie sur $]0, 1[$ donc $-\log X$ est définie sur $]0, +\infty[$ et nulle ailleurs. Sa fonction de répartition de cette v.a est

$$\mathbf{P}(-\log X \leq t) = \mathbf{P}\left(X \geq e^{-t}\right) = 1 - \left(e^{-t}\right)^{1/\theta} \mathbb{I}_{\{t>0\}} = 1 - e^{-\frac{t}{\theta}} \mathbb{I}_{\{t>0\}}.$$

Cette fonction de répartition correspond, en effet, à celle de la loi exponentielle de paramètre θ.

2) T est la somme de n v.a. iid exponentielles de paramètre θ, donc elle suit la loi Gamma(n, θ).

3) La fonction de vraisemblance est

$$L(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} \left(\frac{1}{\theta} x_i^{1/\theta} \right) = \frac{1}{\theta^n} \left(\prod_{i=1}^{n} x_i \right)^{1/\theta}.$$

En appliquant le logarithme aux deux cotés on obtient

$$\log L(x_1, ..., x_n; \theta) = -n \log \theta + \frac{1}{\theta} \log \prod_{i=1}^{n} x_i = -n \log \theta + \frac{1}{\theta} \sum_{i=1}^{n} \log x_i.$$

Le dérivée de celle-ci, par rapport à θ, est

$$\frac{d}{d\theta} \log L(x_1, ..., x_n; \theta) = -\frac{n}{\theta} - \frac{1}{\theta^2} \sum_{i=1}^{n} \log x_i,$$

qui s'annule au point $\theta^* = \left(-\frac{1}{n} \sum_{i=1}^{n} \log x_i \right)^{-1}$. La dérivée seconde de cette dernière est

$$\frac{d^2}{d\theta^2} \log L(x_1, ..., x_n; \theta) = \frac{1}{\theta^3} \left(n\theta + 2 \sum_{i=1}^{n} \log x_i \right).$$

La valeur de celle-ci au point θ^* égale à

$$\begin{aligned} \frac{1}{\theta^{*3}} \left(n\theta^* + 2 \sum_{i=1}^{n} \log x_i \right) &= \frac{1}{\left(-\sum_{i=1}^{n} \log x_i \right)^3} \left(-n \frac{1}{n} \sum_{i=1}^{n} \log x_i + 2 \sum_{i=1}^{n} \log x_i \right) \\ &= \frac{-n^3}{\left(\sum_{i=1}^{n} \log x_i \right)^2} < 0, \end{aligned}$$

qui est strictement négative. Donc θ^* présente un maximum local pour $L(x_1, ..., x_n; \theta)$, ainsi

$$\widehat{\theta}^{EMV} := \left(-\frac{1}{n} \sum_{i=1}^{n} \log x_i \right)^{-1}$$

est l'estimateur du maximum de vraisemblance de θ.

4) Le rapport de vraisemblance généralisé associé à ce test est:

$$R_1 = \frac{\sup_{\theta > 0} L(x_1, ..., x_n; \theta)}{L(x_1, ..., x_n; \theta_0)} = \frac{L\left(x_1, ..., x_n; \widehat{\theta}^{EMV} \right)}{L(x_1, ..., x_n; \theta_0)},$$

où θ_{EMV} est l'estimateur de maximum de vraisemblance de θ et $\theta_0 = 0.5$. Nous avons

$$R_1 = \frac{\prod_{i=1}^{n} \left(\frac{1}{\theta_{EMV}} x_i^{1/\widehat{\theta}_{EMV} - 1} \right)}{\prod_{i=1}^{n} \left(\frac{1}{\theta_{EMV}} x_i^{1/\widehat{\theta}_{EMV} - 1} \right)} = \frac{\frac{1}{\theta_{EMV}^n} \left(\prod_{i=1}^{n} x_i \right)^{1/\widehat{\theta}_{EMV}}}{\frac{1}{\theta_0^n} \left(\prod_{i=1}^{n} x_i \right)^{1/\theta_0}}.$$

Nous avons $\prod_{i=1}^{n} x_i = \exp\left(\sum_{i=1}^{n} \log x_i \right)$, alors

$$R_1 = \frac{\frac{1}{\widehat{\theta}_{EMV}^n} \left(\exp\left(\sum_{i=1}^{n} \log x_i \right) \right)^{1/\widehat{\theta}_{EMV}}}{\frac{1}{\theta_0^n} \left(\exp \sum_{i=1}^{n} \log x_i \right)^{1/\theta_0}}.$$

Observons que $\sum_{i=1}^{n}\log x_i = -n/\widehat{\theta}_{EMV}$, donc

$$R_1 = \frac{\frac{1}{\widehat{\theta}_{EMV}^n}\left(\exp\left(-\frac{n}{\widehat{\theta}_{EMV}}\right)\right)^{1/\widehat{\theta}_{EMV}}}{\frac{1}{\theta_0^n}\left(\exp\left(-\frac{n}{\widehat{\theta}_{EMV}}\right)\right)^{1/\theta_0}} = \frac{\frac{1}{\widehat{\theta}_{EMV}^n}\exp -\frac{n}{\widehat{\theta}_{EMV}^2}}{\frac{1}{\theta_0^n}\exp\left(-\frac{n}{\theta_0\widehat{\theta}_{EMV}}\right)}$$
$$= \left\{\frac{\theta_0}{\widehat{\theta}_{EMV}}\exp\left(\frac{1}{\theta_0\widehat{\theta}_{EMV}} - \frac{1}{\widehat{\theta}_{EMV}^2}\right)\right\}^n.$$

Pour $\theta_0 = 0.5$, on a

$$R_1 = \left\{\frac{1}{2\widehat{\theta}_{EMV}}\exp\left(\frac{2}{\widehat{\theta}_{EMV}} - \frac{1}{\widehat{\theta}_{EMV}^2}\right)\right\}^n.$$

5. La région critique de ce test est:

$$W = \{(x_1, ..., x_n) \in (]0,1[)^n : R_1 \geq c\},$$

où $\mathbf{P}_{\theta=0.5}(R_1 \geq c) = 0.05$. Il est claire que $R_1^{1/n} = (u/2)\exp(2u - u^2)$, où $u = 1/\widehat{\theta}_{EMV}$. On montre facilement que la fonction $u \to \varphi(u) = \exp(2u - u^2)$ atteint son maximum en $u = 4/3$. Donc pour $c > 0$ si $\varphi(u) \geq c$ alors ils existent $0 < c_1 < c_2 < \infty$ telle que $c_1 \leq u \leq c_2$. Ceci implique que $R_1 \geq c$ implique que qu'ils existent $0 < k_1 < k_2 < \infty$, telles que $k_1 \leq 1/\widehat{\theta}_{EMV} \leq k_2$. En d'autres termes $k_2 \leq nT \leq k_1$, ainsi $k_2' \leq T \leq k_1'$, où $k_1' = k_1/n$ et $k_2' = k_2/n$. On sait déjà que $2T/\theta_0$ suit la loi de khi-deux à $2n$ degré de liberté, par conséquent

$$W =: \{(x_1, ..., x_n) \in (]0,1[)^n : r_1 \leq 2T/\theta_0 \leq r_2\},$$

où $\mathbf{P}(\chi_{2n}^2 \leq r_1) = \mathbf{P}(\chi_{2n}^2 \geq r_2) = \alpha/2$. Pour $\theta_0 = 0.5$, $\alpha = 0.05$ et $n = 10$,on obtient

$$W =: \left\{(x_1, ..., x_{10}) \in (]0,1[)^{10} : 9.59 \leq 4T \leq 34.16\right\}.$$

6) La somme de $-\log$ des observations est égale $T_{obs} = 2.346$ et $4T_{obs} = 9.385$ qui est ni ≥ 9.59 ni ≤ 34.16, donc en garde H_0, c'est à dire $\theta = 0.5$, en conclusion X suit la loi de probabilité de densité $f_{0.5}$.

* * * * * * * * * * * * * * *

Solution de l'exercice 2 (8pts).

Soit μ_1 la moyenne théorique du rythme cardiaque **avant** traitement et μ_2 celle **après** traitement. On se pose la question de savoir si l'on peut, à faible risque d'erreur $\alpha = 0.05$, considérer que le traitement est efficace, c'est à dire $\mu_1 \leq \mu_2$. On testera donc $H_0 : \mu_1 \leq \mu_2$ contre $H_1 : \mu_1 > \mu_2$. D'après le cours, la région critique de ce test est définie par

$$W = \left\{\left(x_1^{(1)}, ..., x_{n_1}^{(1)}; x_1^{(2)}, ..., x_{n_2}^{(2)}\right) \in \mathbb{R}_+^{n_1+n_2} : \frac{\overline{x}_1 - \overline{x}_2}{\widetilde{s}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{1-\alpha}^{(n_1+n_2-2)}\right\}, \quad \textbf{(2pts)}$$

où $t_{1-\alpha}^{(n_1+n_2-2)}$ désigne le quantile d'ordre α de la loi de student à $(n_1 + n_2 - 2)$ degré de liberté et

$$\widetilde{s} := \sqrt{\frac{n_1\widetilde{s}_1^2 + n_2\widetilde{s}_1^2}{n_1 + n_2 - 2}},$$

avec $\widetilde{s}_1^2$ (resp. $\widetilde{s}_2^2$) est la variance empirique de la première (resp. la deuxième) population. Nous avons $n_1 = n_2 = 5$, $\alpha = 0.05$, et $t_{0.95}^{(8)} \simeq 1.86$ (**1pt**) , ainsi

$$W = \left\{\left(x_1^{(1)}, ..., x_5^{(1)}; x_1^{(2)}, ..., x_5^{(2)}\right) \in \mathbb{R}_+^{10} : \frac{\overline{x}^{(1)} - \overline{x}^{(2)}}{\widetilde{s}\sqrt{\frac{1}{5} + \frac{1}{5}}} \geq 1.86\right\}. \textbf{(1pt)}$$

Nous avons $\overline{x}_{1,obs} = 77.66$, $\overline{x}_{2,obs} = 80.2$, $\widetilde{s}_{1,obs}^2 = 96.14$, $\widetilde{s}_{2,obs}^2 = 131.66$,

$$\widetilde{s}_{obs} = \sqrt{\frac{5 \times 96.14 + 5 \times 131.66}{5 + 5 - 2}} = 11.932,$$

et

$$\frac{\overline{x}_{1,obs} - \overline{x}_{2,obs}}{\widetilde{s}_{obs}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{77.66 - 80.2}{11.932 \times \sqrt{\frac{1}{5} + \frac{1}{5}}} = -0.336. \textbf{(3pts)}$$

Comme $-0.336 < 1.86$ alors on garde H_0, ce qui signifie que le rythme cardiaque **après** le traitement (μ_2) est supérieur ou égal à celui **avant** le traitement (μ_1), en conclusion **le traitement est efficace**. (**1pt**)

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2022/2023
Module: tests statistiques

Solution du devoir

Le sujet:

Soit X une variable aléatoire qui suit la loi exponentielle de paramèter $\theta > 0$, c'est-à-dire de densité

$$f_\theta(x) = \frac{1}{\theta} \exp\left(-\frac{1}{\theta}x\right) \mathbb{I}_{\{x \geq 0\}}, \ \theta > 0,$$

où $\mathbb{I}_A$ désigne l'indicatrice de l'ensemble A.

1. Etant donné $X_1, ..., X_n$ un échantillon aléatoire de X, montrer que $T := \sum_{i=1}^n X_i$ suit la loi gamma de paramètre (n, θ), noté $T \rightsquigarrow \Gamma(n, \theta)$.
2. Montrer que $2T/\theta \rightsquigarrow \chi^2_{2n}$ (Khi-deux à $2n$ degrés de liberté).
3. Considérons le test suivant: $H_0 : \theta = 1$ contre $H_0 : \theta \neq 1$. Donner la forme explicite du rapport de vraisemblance généralisé (maximal) associé à ce test.
4. Au niveau de signification $\alpha = 0.05$, déterminer la région critique correspondante.
5. Etant donné l'échantillon observé suivant:

   1.19; 2.33; 2.82; 1.04; 0.58; 0.77; 0.37; 3.28; 0.04; 0.22; 0.74; 2.11.

   Peut-on dire que X suit la loi exponentielle de paramètre $\theta = 1$?

* * * * * * * * * **

Solution:

1. On peut utiliser soit la méthode de convolution soit la méthode de la fonction caratéristique. Nous choisissons ici la deuxième méthode. La fonction caractérisique d'une variable aléatoire Y qui suit la loi Gamma de paramèter (k, θ) est donnée par la formule $\varphi_Y(t) = \mathbf{E}\left[e^{itY}\right] = (1 - i\theta t)^{-n}$, où $i^2 = -1$. Donc il suffit de montrer que $\varphi_T(t) = \varphi_Y(t) = (1 - i\theta t)^{-n}$. En effet, nous avons

$$\varphi_Y(t) = \mathbf{E}\left[e^{itT}\right] = \mathbf{E}\left[e^{it\sum_{j=1}^n X_j}\right] = \prod_{j=1}^n \mathbf{E}\left[e^{itX_j}\right].$$

Comme X_1 suit la loi exponentielle de paramètre θ, alors sa fonction caratéristique est $\varphi_{X_1}(t) = (1 - i\theta t)^{-1}$, ce qui implique que

$$\varphi_Y(t) = \prod_{j=1}^n \mathbf{E}\left[e^{itX_j}\right] = \prod_{j=1}^n (1 - i\theta t)^{-1} = (1 - i\theta t)^{-n} = \varphi_Y(t).$$

Ainsi on a montrer que T suit la loi gamma de paramèter (k, θ).

2. La fonction caractéristique d'une v.a Y quit suit la loi khi-deux à ν degré de liberté est donnée par

$$\varphi_Y(t) = (1 - 2it)^{-\nu/2}.$$

La fonction caractéristique de $W = 2T/\theta$ est

$$\begin{aligned} \varphi_Y(t) &= \mathbf{E}\left[e^{itW}\right] = \mathbf{E}\left[\exp\left(i(2T/\theta)\right]\right. = \mathbf{E}\left[\exp\left(i\left(2t/\theta\right)T\right)\right] \\ &= \left(1 - i\theta\left((2t/\theta)\right)\right)^{-n} = (1 - 2it)^{-n} = (1 - 2it)^{-\nu/2}, \text{ avec } \nu = 2n, \end{aligned}$$

donc $2T/\theta \rightsquigarrow \chi^2_{2n}$.

3. Le rapport de vraisemblance généralisé associé à ce test est:

$$\begin{aligned}
R_1 &= \frac{\sup_{\theta>0} L\left(x_1,...,x_n;\theta\right)}{L\left(x_1,...,x_n;\theta_0\right)} \\
&= \frac{L\left(x_1,...,x_n;\theta_{EMV}\right)}{L\left(x_1,...,x_n;\theta_0\right)} = \frac{L\left(x_1,...,x_n;\overline{x}\right)}{L\left(x_1,...,x_n;1\right)} \\
&= \frac{\prod_{i=1}^{n} \frac{1}{\overline{x}} \exp\left(-\frac{1}{\overline{x}} x_i\right)}{\prod_{i=1}^{n} \exp\left(-x_i\right) \mathbb{I}_{\{x\geq 0\}}} = \frac{\frac{1}{\left(\overline{x}\right)^n} \exp\left(-\frac{1}{\overline{x}} \sum_{i=1}^n x_i\right)}{\exp\left(-\sum_{i=1}^n x_i\right)}.
\end{aligned}$$

Nous avons $\sum_{i=1}^n x_i = n\overline{x}$, donc

$$R_1 = \frac{1}{\left(\overline{x}\right)^n} \frac{\exp\left(-n\right)}{\exp\left(-n\overline{x}\right)} = \left(e^{-1} \frac{1}{\overline{x}} \exp\left(\overline{x}\right)\right)^n.$$

4. La région critique de ce test est:

$$W = \left\{ \left(x_1,...,x_n\right) \in \left(\mathbb{R}_+/\left\{0\right\}\right)^n : \left(e^{-1} \frac{1}{\overline{x}} \exp\left(\overline{x}\right)\right)^n > c \right\},$$

où

$$P_{\theta=1}\left(\left(e^{-1} \frac{1}{\overline{X}} \exp\left(\overline{X}\right)\right)^n > c\right) = 0.05.$$

Ceci peut être simplifiée en

$$W = \left\{ \left(x_1,...,x_n\right) \in \left(\mathbb{R}_+/\left\{0\right\}\right)^n : \frac{1}{\overline{x}} \exp \overline{x} > k \right\},$$

où

$$P_{\theta=1}\left(\frac{1}{\overline{X}} \exp \overline{X} > k\right) = 0.05.$$

Nous avons déjà noté en cours que

$$\frac{1}{\overline{X}} \exp \overline{X} > k \Longleftrightarrow \exists \left(c_1 < c_2\right) \text{ telles que } \overline{X} > c_1 \text{ ou } \overline{X} < c_2.$$

Ce qui implique (vue la continuité de la v.a.) que

$$W = \left\{ \left(x_1,...,x_n\right) \in \left(\mathbb{R}_+/\left\{0\right\}\right)^n : \ \overline{x} \geq c_1 \text{ ou } \overline{x} \leq c_2 \right\},$$

où

$$P_{\theta=1}\left(\overline{X} \geq c_1\right) = P_{\theta=1}\left(\overline{X} \leq c_2\right) = 0.05/2 = 0.025.$$

Comme $T = \sum_{i=1}^n X_i = n\overline{X}$, alors

$$\begin{aligned}
P_{\theta=1}\left(\overline{X} \geq c_1\right) &= P_{\theta=1}\left(T \geq nc_1\right) = P_{\theta=1}\left(\frac{2T}{1} \geq 2nc_1\right) \\
&= P\left(\chi^2_{2n} \geq 2nc_1\right) = 0.025
\end{aligned}$$

et

$$\begin{aligned}
P_{\theta=1}\left(\overline{X} \leq c_2\right) &= 1 - P_{\theta=1}\left(T \geq nc_2\right) = 1 - P_{\theta=1}\left(\frac{2T}{1} \geq 2nc_2\right) \\
&= 1 - P\left(\chi^2_{2n} \geq 2nc_2\right) = 0.025.
\end{aligned}$$

5. L'échantillon observé est de taille $n = 12$, donc

$$P\left(\chi^2_{24} \geq 24c_1\right) = 0.025 \Longleftrightarrow P\left(\chi^2_{24} \geq 24c_1\right) = 0.025.$$

De la table statistique de la loi de khi-deux en obtient: $24c_1 = 39.364 \Longleftrightarrow c_1 = 1.640$. De même, nous avons $P\left(\chi^2_{24} \geq 24c_2\right) = 0.975$, ce qui donne $24c_2 = 12.40 \Longleftrightarrow c_2 = 0.516$. Finalement, la région critique est

$$W = \left\{ \left(x_1,...,x_{12}\right) \in \left(\mathbb{R}_+/\left\{0\right\}\right)^{12} : \overline{x} \geq 1.640 \text{ ou } \overline{x} \leq 0.516 \right\}.$$

La moyenne des données observées est égale à 1.290 qui est ni ≥ 1.640 ni ≤ 0.516, donc en garde H_0, c'est à dire $\theta = 1$.

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2020/2021

Examen (1h)

Exercice 1. (08pts)

Donner la région critique du:

1. test, unilatéral à droite, de la moyenne d'une population gaussienne de variance connue? (2pts)
2. test, bilatéral, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances connues? (2pts)
3. test, unilatéral à gauche, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues? (2pts)
4. test, unilatéral à droite, de comparaisons de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues supposées égales? (2pts)

* * * * * * * * * * * * * * * * * *

Exercice 2. (12 pts)

Soit $(X_1, ..., X_{12})$ un échantillon d'une population X normale centrée de variance σ^2. On s'intéresse au test suivant:

$$\begin{cases} H_0: & \sigma^2 \leq 2 \\ H_1: & \sigma^2 > 2 \end{cases}.$$

1. Quel test s'agit-il: unilatéral ou bilatéral? $(1pt)$
2. Donner la statistique de test à utiliser. $(2pts)$
3. Au niveau de signification $\alpha = 0.05$, donner la région critique de ce test. $(3pts)$
4. Donner la fonction puissance de ce test. $(3pts)$
5. Tracer soigneusement le graphe de la fonction puissance. $(3pts)$

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2021/2022

Examen (1h)

Exercice 1 ***(10pts)***
1. Donner la statistique du test est sa loi de probabilité, pour:
a. un test de la moyenne d'une population gaussienne de variance connue? (2pts)
b. un test de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances connues? (2pts)
c. un test de comparaison de deux moyennes de deux populations gaussiennes indépendantes telles que les tailles des deux échantillons égales à $n_1 = 25$ *et* $n_2 = 27$ *respectivement? (2pts)*
d. un test de comparaisons de deux variances de deux populations gaussiennes indépendantes? (2pts)
2. Donner la définition du rapport de vraisemblance maximale et dans quel cas on utilise ce dernier? (2pts)

* * * * * * * * * * * * * * * * * **

Exercice 2 ***(10 pts)****. Au niveau de signification* $\alpha = 0.1$*, on veut tester l'égalité des variances de deux populations gaussiennes* X_1 *et* X_2*. Un échantillon de taille* 21 *de* X_1 *donne une variance empirique* $s_1^2 = 5.20$ *et un échantillon de taille* 9 *de* X_2 *donne une variance empirique* $s_2^2 = 2.30$.

1. *S'agit-il d'un test unilatéral ou bilatéral?*(1*pt*)
2. *Etablir l'hypothèse nulle et l'hypothèse alternative.*(2*pts*)
3. *Quelle est la statistique de test à utiliser et quelle est sa loi de probabilité?*(2*pts*)
4. *Déterminer la région critique de ce test.*(4*pts*)
5. *Conclusion.*(1*pts*)

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2021/2022
Module: Tests statistiques 18 avril 2022

Examen de rattrapage (1h)

Exercice 1. (08pts)

Donner la région critique du:

1. test, unilatéral à droite, de la moyenne d'une population gaussienne de variance connue? (2pts)
2. test, bilatéral, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances connues? (2pts)
3. test, unilatéral à gauche, de comparaison de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues? (2pts)
4. test, unilatéral à droite, de comparaisons de deux moyennes de deux populations gaussiennes indépendantes de variances inconnues supposées égales? (2pts)

* * * * * * * * * * * * * * * * * **

Exercice 2. (12 pts)

Soit $(X_1, ..., X_{12})$ un échantillon d'une population X normale centrée de variance σ^2. On s'intéresse au test suivant:

$$\begin{cases} H_0 : & \sigma^2 \leq 1 \\ H_1 : & \sigma^2 > 1 \end{cases} .$$

1. Quel test s'agit-il: unilatéral ou bilatéral? $(2pt)$
2. Donner la statistique de test à utiliser. $(3pts)$
3. Au niveau de signification $\alpha = 0.05$, donner la région critique de ce test. $(4pts)$
4. Donner la fonction puissance de ce test. $(3pts)$

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2021/2022

Examen de rattrapge (1h)

Exercice 1 ***(10pts)***
Une étude de sociologie porte sur le temps passé par des enfants, âgés de 8 *à* 16 *ans, sur des jeux électroniques. La question est de savoir si le temps moyen par jour est de* 7 *heures. On a demandé à* 15 *enfants leurs nombres d'heures de jeu par jour et les réponses sont les suivantes:*

5 9 5 8 7 6 7 9 7 9 6 9 10 9 8

1. *En supposant que ce temps est normalement distribué, avec une variance égale à* 3, *que conclut-on au niveau de signification* 5% ?
2. *Répondre à la même question si la variance était inconnue.*

* * * * * * * * * * * * * * * * * *

Exercice 2 ***(10 pts)****. On veut tester l'égalité des variances de deux populations* $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ *et* $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ *au niveau de signification* $\alpha = 0.01$. *Un échantillon de taille* 16 *de* X_1 *donne une variance empirique* $s_1^2 = 7.62$ *et un échantillon de taille* 12 *de* X_2 *donne une variance empirique* $s_2^2 = 3.96$. *Conclure.*

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2021/2022

Examen de remplacement (1h)

Exercice 1 ***(10pts)***
1. Remplir le tableau suivant par les riques associés:(2*pts*)

		Vérité	
		H_0	H_1
Décision	H_0	$\times$	$\times$
	H_1	$\times$	$\times$

2. Donner la définition d'un test sans biais et d'un test consistent. (2*pts*)

3. Répondre:(2*pts*).
a) La puissance du test égale $\mathbf{P}\left[\text{accepter } H_1 \mid H_0 \text{ est fausse}\right]$*. Vrai ou faux?*
b) Etant donné une région critique W *. On a* $\mathbf{P}\left(W \mid H_1 \text{ est fausse}\right) = \alpha$*. Vrai ou faux?*
4. Quelle est la relation entre le seuil de singnification et la fonction puisance?(02*pt*)
5. Que représente la variable de décision?

* * * * * * * * * * * * * * * * * **

Exercice 2 ***(10 pts)****. Dans une production il y a une proportion* p *d'articles non-défectueux. On prélève* 15 *articles pour réaliser le test*

$$\begin{cases} H_0 : p \geq 0.9 \\ H_1 : p < 0.9 \end{cases},$$

au niveau de signification 0.05.

1. *Quelle est la loi de la v.a. qui représente les articles non-défectueux ?*(01*pt*)
2. *Quelle est la statistique de test à utiliser et quelle est sa loi de probabilité?*(2*pts*)
3. *Déterminer la fonction test* δ*.* (4*pts*)
4. *Déteminer la fonction puissance et tracer son graphe.* (3*pts*)

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2022/2023
Module: tests statistiques

Corrigé-type de l'examen

Exercice 1. (12pts)

Soit X une variable aléatoire (v.a) qui suit une loi de probabilité de densité

$$f_\theta(x) = \frac{1}{\theta} x^{1/\theta - 1} \mathbb{I}_{\{0<x<1\}}, \ \theta > 0,$$

où $\mathbb{I}_A$ désigne l'indicatrice de l'ensemble A.

1. Montrer que $-\log X$ est une v.a. exponentielle de paramètre θ. (**2pt**)
2. Etant donné $X_1, ..., X_n$ un échantillon aléatoire de X, déterminer la loi de probabilité de $T := -\sum_{i=1}^n \log X_i$. (**1pt**)
3. Déterminer l'estimateur du maximum de vraisemblance de θ, noté $\widehat{\theta}_{EMV}$, basé sur $X_1, ..., X_n$. (**2pts**)
4. Considérons le test suivant: $H_0 : \theta = 0.5$ contre $H_1 : \theta \neq 0.5$. Montrer que le rapport de vraisemblance généralisé associé à ce test égal à

$$R_1 = \left\{ \frac{1}{2\widehat{\theta}_{EMV}} \exp\left(\frac{2}{\widehat{\theta}_{EMV}} - \frac{1}{\widehat{\theta}^2_{EMV}} \right) \right\}^n . \ (\mathbf{2pts})$$

   Aide: utiliser la relation $a \times b = \exp(\log a + \log b)$, pour $a, b > 0$.

5. Au niveau de signification $\alpha = 0.05$ et pour un échantillon de taille $n = 10$, déterminer la région critique correspondante. (**3pts**)
6. Etant donné l'échantillon observé suivant:

   0.845; 0.612; 0.771; 0.977; 0.926; 0.859; 0.558; 0.910; 0.822; 0.740.

7. Peut-on dire que X suit la loi de probabilité de densité $f_{0.5}$? (**2pt**)

* * * * * * * * * **

Exercice 2. (8pts)

Pour tester, au niveau $\alpha = 0.05$, l'efficacité d'un traitement destiné à augmenter le rythme cardiaque, on a mesuré sur 5 individus ce rythme (en moyenne), avant et après l'administration du traitement:

avant	80	90	70.3	85	63
après	84	95.5	73.5	86	62

Est-il efficace ?

On suppose que le rythme cardiaque se répartit de façon quasi gaussienne pour la population considérée (avant comme après traitement) et que les variances des deux populations sont égales.

(Aide: test de comparaison de deux moyennes $\mu_1 \leq \mu_2$)

Solution

Solution de l'exercice 1. (12pts)

1) On montre facilement que la fonction de répartition de X est $F_\theta(x) = x^{1/\theta} \mathbb{I}_{\{0<x<1\}}$, $\theta > 0$. La v.a X est définie sur $]0, 1[$ donc $-\log X$ est définie sur $]0, +\infty[$ et nulle ailleurs. Sa fonction de répartition de cette v.a est

$$\mathbf{P}(-\log X \leq t) = \mathbf{P}\left(X \geq e^{-t}\right) = 1 - \left(e^{-t}\right)^{1/\theta} \mathbb{I}_{\{t>0\}} = 1 - e^{-\frac{t}{\theta}} \mathbb{I}_{\{t>0\}}.$$

Cette fonction de répartition correspond, en effet, à celle de la loi exponentielle de paramètre θ.

2) T est la somme de n v.a. iid exponentielles de paramètre θ, donc elle suit la loi Gamma(n,θ).

3) La fonction de vraisemblance est

$$L(x_1,...,x_n;\theta) = \prod_{i=1}^{n}\left(\frac{1}{\theta}x_i^{1/\theta}\right) = \frac{1}{\theta^n}\left(\prod_{i=1}^{n}x_i\right)^{1/\theta}.$$

En appliquant le logarithme aux deux cotés on obtient

$$\log L(x_1,...,x_n;\theta) = -n\log\theta + \frac{1}{\theta}\log\prod_{i=1}^{n}x_i = -n\log\theta + \frac{1}{\theta}\sum_{i=1}^{n}\log x_i.$$

Le dérivée de celle-ci, par rapport à θ, est

$$\frac{d}{d\theta}\log L(x_1,...,x_n;\theta) = -\frac{n}{\theta} - \frac{1}{\theta^2}\sum_{i=1}^{n}\log x_i,$$

qui s'annule au point $\theta^* = \left(-\frac{1}{n}\sum_{i=1}^{n}\log x_i\right)^{-1}$. La dérivée seconde de cette dernière est

$$\frac{d^2}{d\theta^2}\log L(x_1,...,x_n;\theta) = \frac{1}{\theta^3}\left(n\theta + 2\sum_{i=1}^{n}\log x_i\right).$$

La valeur de celle-ci au point θ^* égale à

$$\begin{aligned}\frac{1}{\theta^{*3}}\left(n\theta^* + 2\sum_{i=1}^{n}\log x_i\right) &= \frac{1}{\left(-\sum_{i=1}^{n}\log x_i\right)^3}\left(-n\frac{1}{n}\sum_{i=1}^{n}\log x_i + 2\sum_{i=1}^{n}\log x_i\right)\\ &= \frac{-n^3}{\left(\sum_{i=1}^{n}\log x_i\right)^2} < 0,\end{aligned}$$

qui est strictement négative. Donc θ^* présente un maximum local pour $L(x_1,...,x_n;\theta)$, ainsi

$$\widehat{\theta}^{EMV} := \left(-\frac{1}{n}\sum_{i=1}^{n}\log x_i\right)^{-1}$$

est l'estimateur du maximum de vraisemblance de θ.

4) Le rapport de vraisemblance généralisé associé à ce test est:

$$R_1 = \frac{\sup_{\theta>0}L(x_1,...,x_n;\theta)}{L(x_1,...,x_n;\theta_0)} = \frac{L\left(x_1,...,x_n;\widehat{\theta}^{EMV}\right)}{L(x_1,...,x_n;\theta_0)},$$

où θ_{EMV} est l'estimateur de maximum de vraisemblance de θ et $\theta_0 = 0.5$. Nous avons

$$R_1 = \frac{\prod_{i=1}^{n}\left(\frac{1}{\theta_{EMV}}x_i^{1/\widehat{\theta}_{EMV}-1}\right)}{\prod_{i=1}^{n}\left(\frac{1}{\theta_{EMV}}x_i^{1/\widehat{\theta}_{EMV}-1}\right)} = \frac{\frac{1}{\theta_{EMV}^n}\left(\prod_{i=1}^{n}x_i\right)^{1/\widehat{\theta}_{EMV}}}{\frac{1}{\theta_0^n}\left(\prod_{i=1}^{n}x_i\right)^{1/\theta_0}}.$$

Nous avons $\prod_{i=1}^{n}x_i = \exp\left(\sum_{i=1}^{n}\log x_i\right)$, alors

$$R_1 = \frac{\frac{1}{\widehat{\theta}_{EMV}^n}\left(\exp\left(\sum_{i=1}^{n}\log x_i\right)\right)^{1/\widehat{\theta}_{EMV}}}{\frac{1}{\theta_0^n}\left(\exp\sum_{i=1}^{n}\log x_i\right)^{1/\theta_0}}.$$

Observons que $\sum_{i=1}^{n}\log x_i=-n/\widehat{\theta}_{EMV}$, donc

$$R_1=\frac{\frac{1}{\widehat{\theta}_{EMV}^n}\left(\exp\left(-\frac{n}{\widehat{\theta}_{EMV}}\right)\right)^{1/\widehat{\theta}_{EMV}}}{\frac{1}{\theta_0^n}\left(\exp\left(-\frac{n}{\widehat{\theta}_{EMV}}\right)\right)^{1/\theta_0}}=\frac{\frac{1}{\widehat{\theta}_{EMV}^n}\exp-\frac{n}{\widehat{\theta}_{EMV}^2}}{\frac{1}{\theta_0^n}\exp\left(-\frac{n}{\theta_0\widehat{\theta}_{EMV}}\right)}$$
$$=\left\{\frac{\theta_0}{\widehat{\theta}_{EMV}}\exp\left(\frac{1}{\theta_0\widehat{\theta}_{EMV}}-\frac{1}{\widehat{\theta}_{EMV}^2}\right)\right\}^n.$$

Pour $\theta_0=0.5$, on a
$$R_1=\left\{\frac{1}{2\widehat{\theta}_{EMV}}\exp\left(\frac{2}{\widehat{\theta}_{EMV}}-\frac{1}{\widehat{\theta}_{EMV}^2}\right)\right\}^n.$$

5. La région critique de ce test est:
$$W=\{(x_1,...,x_n)\in(]0,1[)^n:R_1\geq c\},$$
où $\mathbf{P}_{\theta=0.5}(R_1\geq c)=0.05$. Il est claire que $R_1^{1/n}=(u/2)\exp\left(2u-u^2\right)$, où $u=1/\widehat{\theta}_{EMV}$. On montre facilement que la fonction $u\to\varphi(u)=\exp\left(2u-u^2\right)$ atteint son maximum en $u=4/3$. Donc pour $c>0$ si $\varphi(u)\geq c$ alors ils existent $0<c_1<c_2<\infty$ telle que $c_1\leq u\leq c_2$. Ceci implique que $R_1\geq c$ implique que qu'ils existent $0<k_1<k_2<\infty$, telles que $k_1\leq 1/\widehat{\theta}_{EMV}\leq k_2$. En d'autres termes $k_2\leq nT\leq k_1$, ainsi $k_2'\leq T\leq k_1'$, où $k_1'=k_1/n$ et $k_2'=k_2/n$. On sait déjà que $2T/\theta_0$ suit la loi de khi-deux à $2n$ degré de liberté, par conséquent
$$W=:\{(x_1,...,x_n)\in(]0,1[)^n:r_1\leq 2T/\theta_0\leq r_2\},$$
où $\mathbf{P}\left(\chi_{2n}^2\leq r_1\right)=\mathbf{P}\left(\chi_{2n}^2\geq r_2\right)=\alpha/2$. Pour $\theta_0=0.5$, $\alpha=0.05$ et $n=10$,on obtient
$$W=:\left\{(x_1,...,x_{10})\in(]0,1[)^{10}:9.59\leq 4T\leq 34.16\right\}.$$

6) La somme de $-\log$ des observations est égale $T_{obs}=2.346$ et $4T_{obs}=9.385$ qui est ni ≥ 9.59 ni ≤ 34.16, donc en garde H_0, c'est à dire $\theta=0.5$, en conclusion X suit la loi de probabilité de densité $f_{0.5}$.

* * * * * * * * * * * * * * *

Solution de l'exercice 2 (**8pts**).

Soit μ_1 la moyenne théorique du rythme cardiaque **avant** traitement et μ_2 celle **après** traitement. On se pose la question de savoir si l'on peut, à faible risque d'erreur $\alpha=0.05$, considérer que le traitement est efficace, c'est à dire $\mu_1\leq\mu_2$. On testera donc $H_0:\mu_1\leq\mu_2$ contre $H_1:\mu_1>\mu_2$. D'après le cours, la région critique de ce test est définie par
$$W=\left\{\left(x_1^{(1)},...,x_{n_1}^{(1)};x_1^{(2)},...,x_{n_2}^{(2)}\right)\in\mathbb{R}_+^{n_1+n_2}:\frac{\overline{x}_1-\overline{x}_2}{\widetilde{s}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\geq t_{1-\alpha}^{(n_1+n_2-2)}\right\},\ (\mathbf{2pts})$$
où $t_{1-\alpha}^{(n_1+n_2-2)}$ désigne le quantile d'ordre α de la loi de student à (n_1+n_2-2) degré de liberté et
$$\widetilde{s}:=\sqrt{\frac{n_1\widetilde{s}_1^2+n_2\widetilde{s}_1^2}{n_1+n_2-2}},$$
avec $\widetilde{s}_1^2$ (resp. $\widetilde{s}_2^2$) est la variance empirique de la première (resp. la deuxième) population. Nous avons $n_1=n_2=5$, $\alpha=0.05$, et $t_{0.95}^{(8)}\simeq 1.86$ (**1pt**) , ainsi
$$W=\left\{\left(x_1^{(1)},...,x_5^{(1)};x_1^{(2)},...,x_5^{(2)}\right)\in\mathbb{R}_+^{10}:\frac{\overline{x}^{(1)}-\overline{x}^{(2)}}{\widetilde{s}\sqrt{\frac{1}{5}+\frac{1}{5}}}\geq 1.86\right\}.\ (\mathbf{1pt})$$

Nous avons $\overline{x}_{1,obs}=77.66$, $\overline{x}_{2,obs}=80.2$, $\widetilde{s}_{1,obs}^2=96.14$, $\widetilde{s}_{2,obs}^2=131.66$,
$$\widetilde{s}_{obs}=\sqrt{\frac{5\times 96.14+5\times 131.66}{5+5-2}}=11.932,$$
et
$$\frac{\overline{x}_{1,obs}-\overline{x}_{2,obs}}{\widetilde{s}_{obs}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{77.66-80.2}{11.932\times\sqrt{\frac{1}{5}+\frac{1}{5}}}=-0.336.\ (\mathbf{3pts})$$

Comme $-0.336<1.86$ alors on garde H_0, ce qui signifie que le rythme cardiaque **après** le traitement (μ_2) est supérieur ou égal à celui **avant** le traitement (μ_1), en conclusion **le traitement est efficace**. (**1pt**)

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2022/2023
Module: tests statistiques

Examen de rattrapage

Exercice 1. (10pts)

Soit X une variable aléatoire de loi binomiale $\mathcal{B}(4, p)$ définie par sa fonction de masse:

$$\mathbf{P}(X = x) = C_4^x p^x (1-p)^{4-x}, \ 0 \leq x \leq 4.$$

1. Etant donné un échantillon aléatoire $X_1, ..., X_{10}$ de X, déterminer l'estimateur du maximum de vraisemblance $\widehat{p}_{EMV}$ de p associé au modèle ci-dessus. (**2pt**)
2. Déterminer la loi de probabilité exacte de $10\widehat{p}_{EMV}$. (**1pt**)
3. En utilisant le théorème central limite, déterminer la loi de probabilité asymptotique de $\widehat{p}_{EMV}$. (**1pt**)
4. Considérons le test suivant: $H_0 : p = 0.3$ contre $H_1 : p \neq 0.3$. Déterminer le rapport de vraisemblance généralisé associé à ce test. (**2pts**)
5. Au niveau de signification $\alpha = 0.05$ et pour un échantillon de taille $n = 10$, déterminer la région critique asymptotique, correspondante. (**3pts**)
6. Etant donné l'échantillon, observé de X, suivant:

$$\{2, \ 2, \ 3, \ 2, \ 1, \ 2, \ 1, \ 2, \ 1, \ 4\}.$$

Peut-on dire que X suit la loi binomiale $\mathcal{B}(4, 0.3)$? (**1pt**)

$**********$

Exercice 2. (10pts)

Le point de fusion de 16 échantillons d'une matière grasse a été déterminé et on a obtenu $\overline{x} = 34.5$ °C. On suppose que la distribution de ce point de fusion suit la loi normale avec une espérance μ et écart-type $\sigma = 1.2$.

1. Tester $\mu = 34.5$ contre $\mu < 34.5$ avec un seuil $\alpha = 0.01$. (**5pts**)
2. Si la vraie valeur de $\mu = 34.5$, qu'elle la puissance de ce test? (**3pts**)
3. Quelle est la taille de échantillon nécessaire pour assurer que le risque de deuxième espèce soit égale à 0.01.? (**2pts**)

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2022/2023
Module: tests statistiques

Solution de l'examen de remplacement

Exercice 1. (10pts)

Soit X une variable aléatoire qui suit une loi de probabilité de densité

$$f_\theta(x) = \exp(-x+\theta)\,\mathbb{I}_{\{x>\theta\}},\ \theta > 0,$$

où $\mathbb{I}_A$ désigne l'indicatrice de l'ensemble A.

1. Etant donné $X_1, ..., X_n$ un échantillon aléatoire de X. Déterminer la loi de probabilité du $\min(X_1, ..., X_n)$. (**2pt**)

2. Montrer que $\widehat{\theta}_{EMV} := \min(X_1, ..., X_n)$ est l'estimateur du maximum de vraisemblance de θ associé au modèle ci-dessus. (**2pts**)

3. Considérons le test suivant: $H_0 : \theta = 0.1$ contre $H_1 : \theta \neq 0.1$. Déterminer le rapport de vraisemblance généralisé associé à ce test. (**2pts**)

4. Au niveau de signification $\alpha = 0.05$ et pour un échantillon de taille $n = 8$, déterminer la région critique correspondante. (**3pts**)

5. Etant donné l'échantillon observé suivant:

   $$0.61;\ 0.50;\ 1.34;\ 0.37;\ 1.39;\ 0.15;\ 1.30;\ 0.35$$

   Peut-on dire que X suit la loi de probabilité de densité $f_{0.1}$? (**1pt**)

* * * * * * * * * **

Exercice 2. (10pts)

Un producteur de pneus envisage de changer la méthode de fabrication. La distribution de la durée de vie de ses pneus traditionnels est connue: moyenne 64000 km, écart-type 8000 km; elle est pratiquement gaussienne. Dix pneus sont fabriqués avec la nouvelle méthode et une moyenne de 67300 km est constatée.

1. En supposant que la nouvelle fabrication donnerait une distribution à peu près gaussienne et de même variance, testez l'efficacité de la nouvelle méthode au niveau $\alpha = 0.05$. (**5pts**)
2. Déterminer la fonction puissance de ce test. (**5pts**)

(Aide: test de $H_0 : \mu \leq \mu_0$)

* * * * * * * **

Solution de l'exercice 1.

1) On montre facilement que la fonction de répartition de X est $F_\theta(x) = 1 - \exp(-x+\theta)\,\mathbb{I}_{\{x>\theta\}}$, $\theta > 0$. La loi de $\min(X_1, ..., X_n)$ est:

$$\begin{aligned}\mathbf{P}(\min(X_1, ..., X_n) \leq x) &= 1 - \mathbf{P}(\min(X_1, ..., X_n) \geq x) = 1 - (1 - F_\theta(x))^n \\ &= 1 - \exp(-nx + n\theta)\,\mathbb{I}_{\{x>\theta\}}\ (\text{noté } G_n(x;\theta)).\end{aligned}$$

2) La fonction de vraisemblance est

$$L(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} \exp(-x_i+\theta)\,\mathbb{I}_{\{x_i>\theta\}} = \exp(-nx_i + n\theta)\,\mathbb{I}_{\{m_n>\theta\}},$$

où $m_n := \min(x_1, .., x_n)$. La fonction $\theta \to \exp(-nx_i + n\theta)\,\mathbb{I}_{\{m_n>\theta>0\}}$ atteint son maximum en $\theta = m_n$, donc $\widehat{\theta}_{EMV} := \min(X_1, ..., X_n)$ est l'estimateur du maximum de vraisemblance de θ.

3) Le rapport de vraisemblance généralisé associé à ce test est:

$$R_1 = \frac{\sup_{\theta>0} L(x_1,...,x_n;\theta)}{L(x_1,...,x_n;\theta_0)} = \frac{L\left(x_1,...,x_n;\widehat{\theta}^{EMV}\right)}{L(x_1,...,x_n;\theta_0)},$$

où θ_{EMV} est l'estimateur de maximum de vraisemblance de θ et $\theta_0 = 0.1$. Nous avons

$$R_1 = \frac{\prod_{i=1}^n \exp\left(-x_i + \widehat{\theta}^{EMV}\right) \mathbb{I}_{\{x_i > \widehat{\theta}^{EMV}\}}}{\prod_{i=1}^n \exp(-x_i+\theta_0)\, \mathbb{I}_{\{x_i>\theta_0\}}} = \frac{\prod_{i=1}^n \exp\left(-x_i + \widehat{\theta}^{EMV}\right) \mathbb{I}_{\{x_i > m_n\}}}{\prod_{i=1}^n \exp(-x_i+\theta_0)\, \mathbb{I}_{\{x_i>\theta_0\}}}.$$

On note que $x_i > m_n$, $i = 1,...,n$, et que sous $H_0 : \theta = \theta_0$, on a aussi $x_i > \theta_0$, $i = 1,...,n$, alors

$$\mathbb{I}_{\{x_i>m_n\}} = \mathbb{I}_{\{x_i>\theta_0\}} = 1, \; i = 1,...,n,$$

ainsi $R_1 = \left\{\exp\left(\widehat{\theta}^{EMV} - \theta_0\right)\right\}^n$.

4. La région critique de ce test est $W = \{(x_1,...,x_n) \in (]\theta_0,\infty[)^n : R_1 \geq c\}$, où $\mathbf{P}_{\theta=0.1}(R_1 \geq c) = 0.05$. Il est claire que

$$R_1 \geq c \Longleftrightarrow \left\{\exp\left(\widehat{\theta}^{EMV} - \theta_0\right)\right\}^n \geq c \Longleftrightarrow \widehat{\theta}^{EMV} \geq k,$$

par conséquent $W = \left\{(x_1,...,x_n) \in (]\theta_0,\infty[)^n : \widehat{\theta}^{EMV} \geq k\right\}$, où $\mathbf{P}_{\theta=0.1}\left(\widehat{\theta}^{EMV} \geq k\right) = 0.05$. Nous avons

$$\begin{aligned}\mathbf{P}_{\theta=0.1}\left(\widehat{\theta}^{EMV} \geq k\right) &= \mathbf{P}_{\theta=0.1}(\min(X_1,...X_n) \geq k) \\ &= 1 - \mathbf{P}_{\theta=0.1}(\min(X_1,...X_n) \leq k) \\ &= 1 - G_n(k;0.1) = \{\exp(-k+0.1)\}^n.\end{aligned}$$

Il est évident que $\{\exp(-k+0.1)\}^n = 0.05 \Longleftrightarrow n(-k+0.1) = \log 0.05$. Donc pour $n = 8$ on obtient

$$k \simeq 0.1 + \frac{3}{8} = 0.475.$$

Donc la forme finale de la région critique est

$$W = \left\{(x_1,...,x_8) \in (]0.1,\infty[)^8 : \widehat{\theta}^{EMV} \geq 0.475\right\}.$$

5) Nous avons

$$\widehat{\theta}^{EMV}_{obs} = m_n = \min\{0.61;\ 0.50;\ 1.34;\ 0.37;\ 1.39;\ 0.15;\ 1.30;\ 0.35\} = 0.15,$$

qui est inferieur à 0.475, donc on garde $H_0 : \theta_0 = 0.1$, ce qui signifie que X suit la loi de probabilité de densité $f_{0.1}$.

* * * * * * * * * * * * * * *

Solution de l'exercice 2.

1) Il s'agit de tester les deux hypthèses : $H_0 : \mu \leq 64000$ contre $\mu > 64000$. Nous avons affaire à un échantillon de taille $n = 10$ provenant d'une v.a. $\mathcal{N}\left(\mu, 8000^2\right)$. D'après le cours, la forme de la région critique est

$$W = \left\{(x_1,...,x_{10}) \in \mathbb{R}^{10} \mid \frac{\overline{x} - 64000}{8000/\sqrt{10}} \geq z_{1-0.05}\right\},$$

où $z_{1-0.05} = \Phi^{-1}(0.95) = 1.64$. Nous avons $\overline{x}_{obs} = 67300$, ainsi

$$\frac{\overline{x}_{obs} - 64000}{8000/\sqrt{10}} = \frac{67300 - 64000}{8000/\sqrt{10}} = 1.30 < 1.64,$$

donc en garde $H_0 : \mu \leq 64000$, ce qui signifie que la nouvelle méthode n'est pas efficace.

2) La fonction puissance du test est définie par:

$$\begin{aligned}\pi(\mu) &= \mathbf{P}_\mu\left(\frac{\overline{X}-64000}{8000/\sqrt{10}} \geq 1.64\right), \ \mu \in \mathbb{R} \\ &= \mathbf{P}_\mu\left(\frac{\overline{X}-\mu+\mu+64000}{8000/\sqrt{10}} \geq 1.64\right) \\ &= \mathbf{P}_\mu\left(\frac{\overline{X}-\mu}{8000/\sqrt{10}} \geq 1.64 - \frac{\mu+64000}{8000/\sqrt{10}}\right) \simeq \mathbf{P}\left(Z \geq -3.95 \times 10^{-4}\mu - 23.65\right) \\ &= 1-\Phi\left(-3.95 \times 10^{-4}\mu - 23.65\right).\end{aligned}$$

Le fait que $1-\Phi(-z) = \Phi(z)$, implique que

$$\pi(\mu) = \Phi\left(3.95 \times 10^{-4}\mu + 23.65\right), \ \mu \in \mathbb{R}.$$

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2020/2021

Interrogation: Tests Statistiques

Examen noté sur 10

Soit $(X_1, ..., X_n)$ un échantillon, de taille $n \geq 1$, d'une population normale X de moyenne μ et d'écart-type 2. A partir d'un échantillon de taille 12, on veut tester, au niveau de signification $\alpha = 0.02$, les deux hypothèses suivantes:

$$\begin{cases} H_0 : & \mu \leq 1 \\ H_1 : & \mu > 1 \end{cases}$$

1. Quelle est la statistique du test associée au test uniformément le plus puissant? $(1.5pts)$
2. Quelle est la région critique associée à ce test? $(2pts)$
3. Donner la forme du test uniformément le plus puissant, noté δ. $(1/2pt)$
4. Déduire des questions précédentes, la valeur de $\mathbf{E}_{\mu=1}[\delta]$ et quelle est, dans ce cas, la loi de probabilité de δ? $(1pts)$
5. Déterminer l'expression de $\mathbf{E}_{\mu}[\delta]$, pour $\mu \in \mathbb{R}$. Que représente cette quantité? $(2pts)$
6. Déterminer le risque de première espèce. $(1/2pt)$
7. Quelle est la relation entre le risque de première espèce et le seuil de signification α, en justifiant la réponse? $(1pt)$
8. Tracer "soigneusement" le graphe de la fonction puissance. $(1.5pts)$

Remarque importante: Une copie bien soignée = Une copie bien considérée.

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2020/2021

Interrogation N°2 (sur 07 points)

Il s'agit de tester, au niveau de signification 5%, l'égalité de deux moyennes:

$$\begin{cases} H_0: & \mu_1 = \mu_2 \\ H_1: & \mu_1 \neq \mu_2 \end{cases}$$

qui correspondent aux deux populations gaussiennes indépendantes de même variance. On a tiré un échantillon de chacune de ces deux dernières de tailles $n_1 = 13$ et $n_2 = 18$ ayant les moyennes 0.96 et 1.13; et les variances 1.98 et 1.76 respectivement.

1. Donner la statistique de test. (1pt)
2. \\Déterminer la région critique. (3pts)
3. Que conclut-on? (1pt)
4. Déterminer la p-valeur du test. Que concluez-vous? (2pts)

———

Vous aurez besoin une des valeurs des probabilités suivantes:

$$\begin{array}{lll} \mathbf{P}\left(F_{(13,18)} \leq 0.217\right) = 0.003; & \mathbf{P}\left(F_{(18,13)} \leq 1.131\right) = 0.581; & \mathbf{P}\left(F_{(12,17)} \leq 1.765\right) = 0.861, \\ \mathbf{P}\left(T_{18} \leq 0.217\right) = 0.584; & \mathbf{P}\left(T_{29} \leq 0.331\right) = 0.628; & \mathbf{P}\left(T_{13} \leq 1.765\right) = 0.949, \\ \mathbf{P}\left(\chi^2_{18} \leq 15\right) = 0.338; & \mathbf{P}\left(\chi^2_{13} \leq 12.5\right) = 0.512; & \mathbf{P}\left(\chi^2_{29} \leq 11.66\right) = 0.001, \end{array}$$

où $F_{(\nu_1,\nu_2)}, T_\nu$ et χ^2_ν désignent, respectivement, les lois de Fisher, Student et Qui-deux.

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2021/2022

Interrogations de remplacement (bis) 1 et 2 (1h30)

Interrogation 1. (07pts)

Soit $(X_1, ..., X_{10})$ un échantillon d'une population X normale centrée de variance σ^2. Considérons le test de la variance

$$\begin{cases} H_0: & \sigma^2 = 1 \\ H_1: & \sigma^2 < 1 \end{cases}.$$

1. Au niveau de signification 5%, construire le test uniformément le plus puissant. $(4pts)$
2. Tracer le graphe de la fonction puissance de ce test. $(3pts)$

* * * * * * * * * * * * * * * * **

Interrogation 2. (10 pts)

Dans une production il y a une proportion p d'articles non-défectueux. On prélève 12 articles pour réaliser le test

$$\begin{cases} H_0 : p \geq 0.9 \\ H_1 : p < 0.9 \end{cases},$$

au niveau de signification 5%.

1. Déterminer la région critique de ce test.$(4pts)$
2. Déterminer les risques de première et de deuxième espèces.$(3pts)$
3. Tracer le graphe de sa fonction puissance.$(3pts)$

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2020/2021

Interrogation 1

Exercice 1 (10pts)

Soit X une population normale d'espérance inconnue μ et de variance 4. On prélève un échantillon de taille 16 pour réaliser le test

$$\begin{cases} H_0 : \mu = 10 \\ H_1 : \mu > 10 \end{cases}.$$

1. Quel type de test s'agit-il: unilatéral ou blilatéral?. $(1pts)$
2. Construire le test uniformément le plus puissant, δ, au niveau $\alpha = 0.05.$ $(3pts)$
3. Déterminer la fonction puissance de δ, notée $\pi(\mu)$. $(2pts)$
4. Tracerl e graphe de la fonction $\pi(\mu)$. $(2pts)$
5. Si $\mu = 12$, calculer le risque de deuxième espèce. $(2pts)$

Inerrogation 2

Exercice 1 (07 pts)

Soient $X_1 \rightsquigarrow \mathcal{N}(\mu_1, \sigma_1^2)$ et $X_2 \rightsquigarrow \mathcal{N}(\mu_2, \sigma_2^2)$ deux v.a gaussiennes de deux populations indépendantes. On prélève un échantillon de taille 16 de X_1 ayant une variance empirique $\widetilde{s}^2_{1,obs} = 7.62$ et un échantillon de taille 12 de X_2 ayant une variance empirique $\widetilde{s}^2_{2,obs} = 3.96$. On veut tester l'égalité des variances de deux populations au niveau de signification $\alpha = 0.01$.

1. Quelle est la statistique adéquate à se test?. $(1pts)$
2. Donner la région critique de ce test. $(3pts)$
3. Doner la la forme de ce test noté $\delta.$ $(1pts)$
4. Quelle est la loi de probabilité de δ? $(1pts)$
5. Que conclut-on? $(1pt)$

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2020/2021

Interrogation: Tests Statistiques

Examen noté sur 10

Soit $(X_1, ..., X_n)$ un échantillon, de taille $n \geq 1$, d'une population normale X de moyenne μ et d'écart-type 2. A partir d'un échantillon de taille 12, on veut tester, au niveau de signification $\alpha = 0.02$, les deux hypothèses suivantes:

$$\begin{cases} H_0: & \mu \leq 1 \\ H_1: & \mu > 1 \end{cases}$$

1. Quelle est la statistique du test associée au test uniformément le plus puissant? $(1.5pts)$
2. Quelle est la région critique associée à ce test? $(2pts)$
3. Donner la forme du test uniformément le plus puissant, noté δ. $(1/2pt)$
4. Déduire des questions précédentes, la valeur de $\mathbf{E}_{\mu=1}[\delta]$ et quelle est, dans ce cas, la loi de probabilité de δ? $(1pts)$
5. Déterminer l'expression de $\mathbf{E}_{\mu}[\delta]$, pour $\mu \in \mathbb{R}$. Que représente cette quantité? $(2pts)$
6. Déterminer le risque de première espèce. $(1/2pt)$
7. Quelle est la relation entre le risque de première espèce et le seuil de signification α, en justifiant la réponse? $(1pt)$
8. Tracer "soigneusement" le graphe de la fonction puissance. $(1.5pts)$

* *

Solution

1) Nous allons appliquer le principe du rapport de vraisemblance monotone. Soient $\mu_2 > \mu_1$ on a

$$\frac{L_{\mu_2}}{L_{\mu_1}} = \frac{L_{\mu_2}(x_1, ..., x_{12})}{L_{\mu_1}(x_1, ..., x_{12})} = \frac{\prod_{i=1}^{12} \frac{1}{\sqrt{2\pi \times 2^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu_2}{2}\right)^2\right\}}{\prod_{i=1}^{12} \frac{1}{\sqrt{2\pi \times 2^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu_1}{2}\right)^2\right\}}.$$

Après la simplification on trouve

$$\frac{L_{\mu_2}}{L_{\mu_1}} = \exp\left\{\frac{1}{4}(\mu_2 - \mu_1)t + \frac{3}{2}\left(\mu_1^2 - \mu_2^2\right)\right\},$$

où $t = t(x_1, ..., x_{12}) = \sum_{i=1}^{12} x_i$. Il est clair que, puisque $\mu_2 - \mu_1 > 0$, la fonction

$$t \to \exp\left\{\frac{1}{4}(\mu_2 - \mu_1)(t - 6\mu_1 - 6\mu_2)\right\}$$

est croissante. Donc l'hypothèse du rapport de vraisemblance monotone est vérifiée. On en déduit que, la statistique associée au test uniformément le plus puissant est

$$T = t(X_1, ..., X_{12}) = \sum_{i=1}^{12} X_i.$$

2) La région critique associée à ce test est

$$W=\left\{(x_1,...,x_{12})\in\mathbb{R}^{12}\mid\sum_{i=1}^{12}x_i\geq k\right\},$$

où k est une constante telle que

$$P_{\mu=1}\left(\sum_{i=1}^{12}X_i\geq k\right)=0.02.$$

Cette probabilité peut être réécrite sous la forme

$$P_{\mu=1}\left(\frac{\overline{X}-1}{\sqrt{2^2/12}}\leq c\right)=1-0.02=0.98,$$

où

$$c=\sqrt{3}\,(k/12-1)\,. \tag{1}$$

Comme $Z:=\frac{\overline{X}-1}{\sqrt{2^2/12}}\rightsquigarrow\mathcal{N}(0,1)$, donc

$$c=\Phi^{-1}(0.98)=2.05. \tag{2}$$

Ce qui implique de l'équation (1) que $k=26.20$, et par conséquent

$$W=\left\{(x_1,...,x_{12})\in\mathbb{R}^{12}\mid\sum_{i=1}^{12}x_i\geq 26.20\right\}.$$

3) Le test uniformément le plus puissant associé à W est

$$\delta=\delta(x_1,...,x_n)=\begin{cases}1\text{ si }\sum_{i=1}^{12}x_i\geq 26.20\\ 0\text{ si }\sum_{i=1}^{12}x_i<26.20\end{cases}$$

4) Nous avons $\mathbf{E}_{\mu=1}[\delta]=\mathbf{P}_{\mu=1}(W)=0.02$. Il est clair que la v.a δ suit la loi de Bernoulli de paramètre 0.02. En effet la v.a. $\delta=\delta(X_1,...X_n)$ prend deux valeurs $\{0,1\}$, tel que

$$\mathbf{P}_{\mu=1}(\delta=1)=\mathbf{P}_{\mu=1}(W)=0.02$$
$$\mathbf{P}_{\mu=1}(\delta=0)=\mathbf{P}_{\mu=1}(\overline{W})=1-0.02=0.98.$$

5) Pour $\mu\in\mathbb{R}$, nous avons

$$\mathbf{E}_\mu[\delta]=\mathbf{P}_\mu(W)=\mathbf{P}_\mu\left(\sum_{i=1}^{12}X_i\geq 26.20\right),$$

ce qui égale à

$$\mathbf{P}_\mu\left(\frac{\overline{X}-\mu}{\sqrt{2^2/12}}\geq\frac{26.20/12-\mu}{\sqrt{2^2/12}}\right).$$

Comme $Z^*:=\frac{\overline{X}-\mu}{\sqrt{2^2/12}}\rightsquigarrow\mathcal{N}(0,1)$, alors cette dernière est équivalente à

$$\mathbf{P}_\mu(Z^*\geq 3.78-1.73\mu)=1-\mathbf{P}_\mu(Z^*\leq 3.78-1.73\mu),$$

par conséquent

$$\mathbf{E}_\mu[\delta]=1-\Phi(3.78-1.73\mu),\ \mu\in\mathbb{R},$$

elle n'est autre que la fonction puissance du test δ, notée $\pi(\mu;\delta)$.

6) Le risque de première espèce est définit par $\alpha(\mu)=\pi(\mu;\delta)$, pour $\mu\leq 1$, c'est à dire

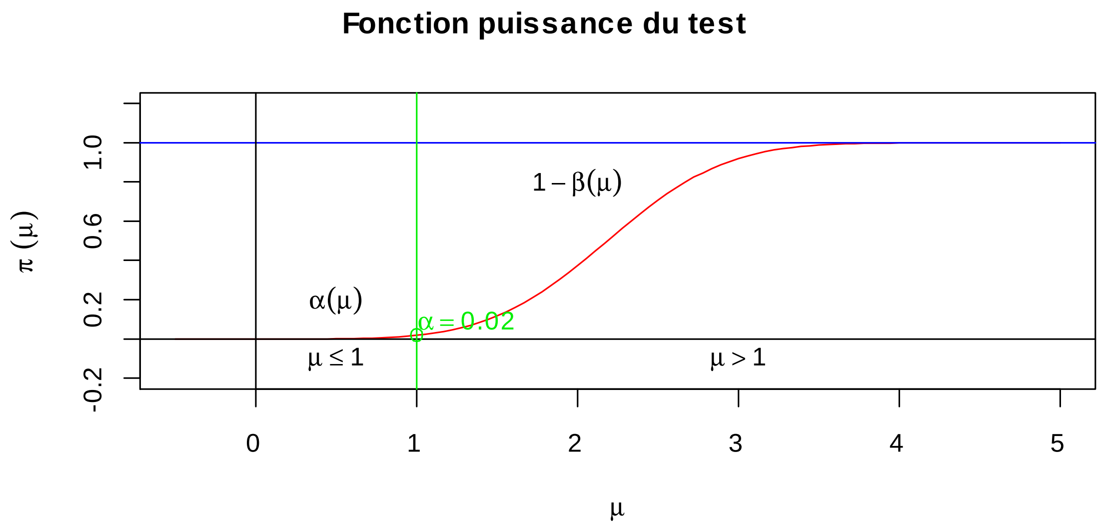
$$\alpha(\mu)=1-\Phi(3.78-1.73\mu),\ \text{pour }\mu\leq 1.$$

7) La relation entre le risque de première espèce $\alpha(\mu)$ est le seuil de signification $\alpha=0.02$, est telle que

$$\alpha=\sup_{\mu\leq 1}\alpha(\mu).$$

En effet, comme la fonction $\mu \to 1-\Phi(3.78-1.73\mu)$ est croissante alors le supremum de $\alpha(\mu)$ atteint la borne la valeur $\mu=1$ dans l'intervalle $\mu \leq 1$, en $1-\Phi(3.78-1.73)=1-\Phi(2.05)$. D'après l'équation (2), on sait déjà que $\Phi(2.05)=0.98$, donc $1-0.98=0.02$ ce qui correspond exactement à la valeur du seuil de signification α.

8) Graphe de la fonction puissance du



Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2022/2023

Solution de l'interrogation de remplacement 1 (1h30)

Interrogation 1. (10pts)

Soit X une variable aléatoire uniformément distribuée sur l'intervalle $[0, \theta]$, $\theta > 0$. Etant donné un échantillon $(x_1, ..., x_n)$ de X, on souhaite tester les deux hypothèses

$$\begin{cases} H_0 : & \theta \geq 1, \\ H_1 : & \theta < 1, \end{cases}$$

au niveau de signification $\alpha = 0.10$.

1. Montrer que la fonction de vraisemblance, associée à l'echantillon $(x_1, ..., x_n)$, est

$$L_\theta(x_1, ..., x_n) = \begin{cases} \frac{1}{\theta^n} & 0 \leq \max x_i \leq \theta, \\ 0 & \text{sinon.} \end{cases} \quad (2pts)$$

2. On définit le rapport

$$R := \frac{L_{\theta_1}(x_1, ..., x_n)}{L_{\theta_2}(x_1, ..., x_n)}, \text{ pour } \theta_1 > \theta_2 > 0.$$

Montrer que R est une fonction croissante par rapport à une fonction $t = t(x_1, ..., x_n)$. $(2pts)$
3. En se basant sur la statistique $T = t(X_1, ..., X_n)$, déterminer la région critique associée au test statistique basé sur R. $(2pts)$
4. Déterminer la fonction puissance de ce test.$(2pts)$
5. Quelle est la puissance du test associée à la valeur $\theta = 0.3$. $(1pt)$
6. Etant donné un échantillon $\{0.976, 0.932, 0.602, 0.331, 0.394, 0.882, 0.183, 0.408, 0.751, 0.585\}$ de la variable X. Que peut-on conclure?$(1pt)$

****

Solution. La densité de la loi uniforme sur $[0, \theta]$ est définie par

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} & \text{si } 0 \leq x \leq \theta \\ 0 & \text{sinon} \end{cases},$$

qui peut être réécrtie sous la forme $f_\theta(x) = \theta^{-1}\mathbf{1}_{\{0 \leq x \leq \theta\}}$, où $\mathbf{1}_A$ désigne la fonction indicatrice associée à l'ensemble A. Ainsi

$$L_\theta(x_1, ..., x_n) = \frac{1}{\theta^n} \prod_{i=1}^n \mathbf{1}_{\{0 \leq x_i \leq \theta\}}.$$

Rappelons que $\mathbf{1}_A \times \mathbf{1}_B = \mathbf{1}_{A \cap B}$, donc $\prod\limits_{i=1}^n \mathbf{1}_{\{0 \leq x_i \leq \theta\}} = \mathbf{1}_{\cap_{i=1}^n \{0 \leq x_i \leq \theta\}} = \mathbf{1}_{\{0 \leq \max x_i \leq \theta\}}$, par conséquent,

$$L_\theta(x_1, ..., x_n) = \frac{1}{\theta^n}\mathbf{1}_{\{0 \leq \max x_i \leq \theta\}}.$$

Soient $\theta_1 > \theta_2 > 0$

$$R = \frac{L_{\theta_1}(x_1, ..., x_n)}{L_{\theta_2}(x_1, ..., x_n)} = \left(\frac{\theta_2}{\theta_1}\right)^n \frac{\mathbf{1}_{\{0 \leq \max x_i \leq \theta_1\}}}{\mathbf{1}_{\{0 \leq \max x_i \leq \theta_2\}}}.$$

Observons que si $0 \leq \max x_i \leq \theta_2$ alors $0 \leq \max x_i \leq \theta_1$ car $\theta_1 > \theta_2$, ce qui implique $\mathbf{1}_{\{0 \leq \max x_i \leq \theta_1\}} = \mathbf{1}_{\{0 \leq \max x_i \leq \theta_1\}} = 1$. Donc $R = \left(\frac{\theta_2}{\theta_1}\right)^n$ pour $0 \leq \max x_i \leq \theta_2$. Maintenant, si $\theta_2 \leq \max x_i \leq \theta_1$, alors $\mathbf{1}_{\{0 \leq \max x_i \leq \theta_1\}} = 1$ et $\mathbf{1}_{\{0 \leq \max x_i \leq \theta_2\}} = 0$, ce qui signifie que $R = \left(\frac{\theta_2}{\theta_1}\right)^n \frac{1}{0} = \infty$. En conclusion

$$R = \begin{cases} \left(\frac{\theta_2}{\theta_1}\right)^n & \text{si } 0 \leq \max x_i \leq \theta_2 \\ \infty & \theta_2 \leq \max x_i \leq \theta_1 \end{cases}.$$

Il est claire que la fonction

$$R = \begin{cases} \left(\frac{\theta_2}{\theta_1}\right)^n & \text{si } 0 \leq t \leq \theta_2 \\ \infty & \theta_2 \leq t \leq \theta_1 \end{cases}, \text{ avec } \theta_1 > \theta_2 > 0$$

est croissante en $t = \max x_i$. La région critique est $W = \{(x_1, ..., x_n) \in \mathbb{R}^n : \max x_i \leq k\}$, où

$$\mathbf{P}_{\theta=1} (\max X_i \leq k) = 0.1.$$

Rappelons que la fonction de répartition de la loi uniforme sur $[0, \theta]$ est d́éfinie par

$$F_\theta(x) = \mathbf{P}_\theta(X \leq x) = \begin{cases} 0 & x < 0 \\ \frac{x}{\theta} & \text{si } 0 \leq x < \theta \\ 1 & x \geq \theta \end{cases}.$$

On sait que $\mathbf{P}_\theta(\max X_i \leq x) = [F_\theta(x)]^n$, alors

$$\mathbf{P}_\theta(\max X_i \leq k) = \begin{cases} 0 & k < 0 \\ \left(\frac{k}{\theta}\right)^n & \text{si } 0 \leq k < \theta \\ 1 & k \geq \theta \end{cases}.$$

Au point $\theta = 1$, on a

$$\mathbf{P}_{\theta=1}(\max X_i \leq k) = \begin{cases} 0 & k < 0 \\ k^n & \text{si } 0 \leq k < 1 \\ 1 & k \geq 1. \end{cases}.$$

Donc $\mathbf{P}_{\theta=1}(\max X_i \leq k) = 0.1$, implique que $k^n = 0.1$, ainsi $k = (0.1)^{1/n}$, ainsi

$$W = \left\{(x_1, ..., x_n) \in \mathbb{R}^n : \max x_i \leq (0.1)^{1/n}\right\}.$$

La fonction puissance du test est

$$\begin{aligned} \pi(\theta) &= \mathbf{P}_\theta\left(\max X_i \leq (0.1)^{1/n}\right), \ \theta > 0 \\ &= \left(\frac{(0.1)^{1/n}}{\theta}\right)^n = \frac{0.1}{\theta^n}, \ \theta > 0. \end{aligned}$$

La puissance du test au point $\theta = 0.3$ est $1 - \beta(0.3) = \pi(0.3) = 0.1/(0.3)^n$. La taille de l'échantillon observé égale à $n = 10$, donc

$$W = \left\{(x_1, ..., x_{10}) \in \mathbb{R}^{10} : \max x_i \leq (0.1)^{1/10} = 0.794\,33\right\}.$$

Nous avons aussi

$$\max\{0.976, 0.932, 0.602, 0.331, 0.394, 0.882, 0.183, 0.408, 0.751, 0.585\} = 0.976.$$

Cette valeur est supperieur à $0.794\,33$, donc on garde H_0, c'est à dire $\theta \geq 1$.

$$\left(\frac{2}{\theta}\right)^n \prod_{i=1}^n (\theta - x_i), \ 0 \leq \max_i x_i \leq \theta$$

$$\prod_{i=1}^n \left(1 - \frac{x_i}{\theta}\right)$$

$$\sum_{i=1}^n \log\left(1 - \frac{x_i}{x_{n:n}}\right) > k$$

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2021/2022
Module: Tests Statistiques

Corrigé-type de l'interrogation N°**1**

Exercice 1 (07pts)

Soit X une population normale d'espérance inconnue μ et de variance 2. On prélève un échantillon de taille 18 pour réaliser le test

$$\begin{cases} H_0 : \mu = 5 \\ H_1 : \mu < 5 \end{cases}.$$

1. Quel type de test s'agit-il: unilatéral ou bilatéral? $(0.5pt)$
2. Construire le test uniformément le plus puissant, δ, au niveau $\alpha = 0.05.\,(3pts)$
3. Déterminer la fonction puissance de δ, notée $\pi(\mu)\,.\,(2pts)$
4. Tracer le graphe de la fonction $\pi(\mu)\,.$ $(1pt)$
5. Si $\mu = 7$, calculer le risque de deuxième espèce. $(0.5pt)$

* * * * * * * * * * * * * * * * * *

Solution.

1) Il s'agit d'un test unilatéral à gauche.

2) Nous allons appliquer la méthode du test de *rapport de vraisemblance croissant (monotone)***.** Soit $\mu_1 > \mu_2$ et écrivons

$$\begin{aligned}\frac{L_{\mu_1}}{L_{\mu_2}} &= \frac{\prod_{i=1}^{18} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_i - \mu_1)^2\right\}}{\prod_{i=1}^{18} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_i - \mu_2)^2\right\}} \\ &= \exp\left(9\left(\mu_2^2 - \mu_1^2\right)\right) \times \exp\left\{(\mu_1 - \mu_2)\sum_{i=1}^{18} x_i\right\}.\end{aligned}$$

On pose $t := \sum_{i=1}^{18} x_i$, $b := \mu_1 - \mu_2$ et $d := \exp\left(9\left(\mu_2^2 - \mu_1^2\right)\right) > 0$, ainsi on a une fonction $t \to \frac{L_{\mu_1}}{L_{\mu_2}}(t) = d \exp bt$, croissante en t, car $\mu_1 > \mu_2$ implique $b > 0$. Alors la loi de X possède un rapport de vraisemblance croissant en $t = \sum_{i=1}^{18} x_i$. Donc en appliquant la proposition 3 et la remarque 3 du cours, on obtient le test upp:

$$\delta = \begin{cases} 1 & \text{si } \sum_{i=1}^{18} x_i \leq c \\ 0 & \text{si } \sum_{i=1}^{18} x_i > c \end{cases}$$

avec $\mathbf{P}_{\mu=5}\left(\sum_{i=1}^{18} X_i \leq c\right) = \alpha = 0.05$. Celle-ci peut être réecrite comme suit:

$$\mathbf{P}_{\mu=5}\left(\frac{\sum_{i=1}^{18} X_i - 18 \times 5}{\sqrt{18 \times 2}} \leq \frac{c - 18 \times 5}{\sqrt{18 \times 2}}\right) = \alpha = 0.05.$$

En d'autres terms $\mathbf{P}(Z \leq c/6 - 15) = 0.05$, ce qui implique que

$$c/6 - 15 = \Phi^{-1}(0.05) = -\Phi^{-1}(1 - 0.05) = -\Phi^{-1}(0.95).$$

Daprès la table statistique des quantiles de la loi de Gauss on a $\Phi^{-1}(0.95) = 1.64$, donc $c/6 - 15 = -1.64$, ainsi $c = 80.16$. La forme explicite de la fonction test upp est

$$\delta(x_1, ..., x_{18}) = \begin{cases} 1 & \text{si } \sum_{i=1}^{18} x_i \leq 80.16 \\ 0 & \text{si } \sum_{i=1}^{18} x_i > 80.16 \end{cases}.$$

Ceci peux être réécrit, en termes de la moyenne $\overline{x} = \dfrac{1}{18}\sum_{i=1}^{18} x_i$, comme suit:

$$\delta(x_1, ..., x_{18}) = \begin{cases} 1 & \text{si } \overline{x} \leq \frac{80.16}{18} = 4.45 \\ 0 & \text{sinon} \end{cases}.$$

Remarque: utiliser $\sum_{i=1}^{18} x_i$ ou $\overline{x}$ donne le même résultat.

3) La fonction puissance du test est

$$\pi(\mu) = \mathbf{P}_\mu\left(\sum_{i=1}^{18} X_i \leq 80.16\right) = \mathbf{P}_\mu\left(\overline{X} \leq 4.45\right), \ \mu \leq 5,$$

En d'autres termes

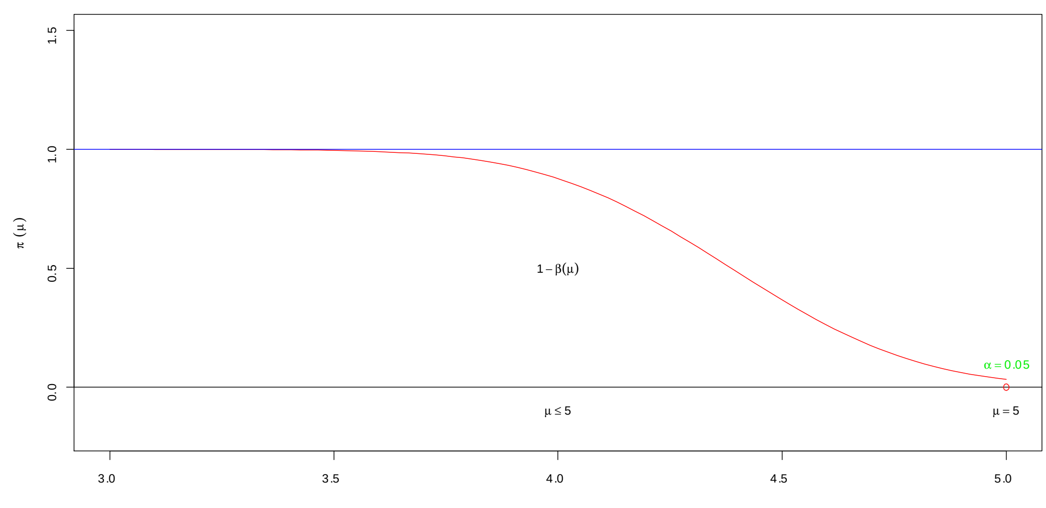
$$\begin{aligned}\pi(\mu) &= \mathbf{P}_\mu\left(\frac{\overline{X} - \mu}{\sqrt{2/18}} \leq \frac{4.45 - \mu}{\sqrt{2/18}}\right) \\ &= \mathbf{P}_\mu\left(Z \leq \frac{4.45 - \mu}{1/3}\right) = \Phi(13.35 - 3\mu) = 1 - \Phi(3\mu - 13.35), \ \mu \leq 5.\end{aligned}$$

En utlisant le fait que $\Phi(-x) = 1 - \Phi(x)$, on ecrit

$$\pi(\mu) = 1 - \Phi(3\mu - 13.35), \ \mu \leq 5.$$

On peut vérifier que $\pi(5) = 0.049 \sim 0.05 = \alpha$.

4) Le graphe de la fonction puissance est donné par la figure suivante:



5) La valeur $\mu = 7$ n'appartient pas au domaine $\mu \leq 5$ du test, donc celle-ci ne peut être considérée.

1. Faculté des Sc. Exactes et Sc. de la Nature et la Vie

Département de Mathématiques
Master 1: 2022/2023
Module: tests statistiques

Interrogation 2 (01h30)

Exercice 1 (10pts)

Dans une production il y a une proportion p d'articles défectueux. On prélève 25 articles pour réaliser le test

$$\begin{cases} H_0 : p \geq 0.4 \\ H_1 : p < 0.4 \end{cases},$$

au niveau de signification 0.05.

1. Construire le test uniformément le plus puissant, noté $\delta = \delta(x_1, ..., x_{25})$. $(2pts)$
2. Quelle est la loi de probabilité de la v.a. $\delta(X_1, ..., X_{25})$? $(1/2pt)$
3. Calculer $\mathbf{E}_{p=0.4}[\delta(X_1, ..., X_{25})]$. $(1/2pt)$
4. Déterminer la fonction puissance de δ, notée $\pi(p)$. $(2pts)$
5. Déduire les expressions des risques de première espèce $\alpha(p)$ et deuxieme espèce $\beta(p)$. $(1pt)$
6. Déduire: $\sup_{p \geq 0.4} \alpha(p)$ et $\sup_{p<0.4} \beta(p)$. $(1pt)$
7. Calculer le puissance du test au valeur $p = 0.3$. $(1pt)$
8. Tracer le graphe de la fonction $\pi(p)$. $(1.5pt)$
9. Le test δ est-il uniformément sans bias? justifier vore réponce.$(1/2pt)$

* * * * * * * * * * * * * * * * * *

Solution.

2) Ici**,** la variable aléatoire X prend une valeur 1 (qui correspond à un article défectueux) avec une probabilité p et elle prend la valeur 0 (qui correspond à un article non défectueux). Donc il s'agit d'une v.a qui suit la loi de Bernoulli de paramètre p. La fonction masse de cette v.a discrète X et définie comme suit:

$$\mathbf{P}(X = 1) = p \text{ et } \mathbf{P}(X = 0) = 1 - p.$$

Celle-ci peut être reformuler comme suit:

$$\mathbf{P}(X = x) = p^x (1-p)^{1-x}, \ x = 0, 1.$$

Nous allons appliquer la méthode du test de *rapport de vraisemblance monotone***.** Soit $p_1 > p_2$ et écrivons

$$\frac{L_{p_1}}{L_{p_2}} = \frac{\prod_{i=1}^{25} p_1^{x_i} (1-p_1)^{1-x_i}}{\prod_{i=1}^{25} p_2^{x_i} (1-p_2)^{1-x_i}} = \frac{p_1^t (1-p_1)^{25-t}}{p_2^t (1-p_2)^{25-t}}, \text{ avec } t := \sum_{i=1}^{25} x_i.$$

Celle peut être réécrite comme suit:

$$\frac{L_{p_1}}{L_{p_2}} = \left(\frac{1-p_1}{1-p_2}\right)^{20} \left(\frac{p_1(1-p_2)}{p_2(1-p_1)}\right)^t.$$

On pose $b := \left(\dfrac{1-p_1}{1-p_2}\right)^{25} > 0$, et $a := \frac{p_1(1-p_2)}{p_2(1-p_1)} > 1$ (car $p_1 > p_2$), ainsi $t \to \frac{L_{p_1}}{L_{p_2}}(t) = ba^t$ est une fonction croissante en t (car $a > 1$). Nous allons alors appliquer la proposition 3, pour avoir le test upp :

$$\delta = \begin{cases} 1 & \text{si } \sum_{i=1}^{25} x_i < c \\ \gamma & \text{si } \sum_{i=1}^{25} x_i = c \\ 0 & \text{si } \sum_{i=1}^{20} x_i > c \end{cases}$$

où c et $0 < \gamma < 1$ sont telles que

$$\mathbf{P}_{p=0.4}\left(\sum_{i=1}^{25} X_i < c\right) + \gamma \mathbf{P}_{p=0.4}\left(\sum_{i=1}^{25} X_i = c\right) = \alpha = 0.05. \tag{1}$$

En d'autres termes

$$\mathbf{P}_{p=0.4}\left(\sum_{i=1}^{25} X_i \leq c\right) + (\gamma - 1)\,\mathbf{P}_{p=0.4}\left(\sum_{i=1}^{25} X_i = c\right) = 0.05,$$

ce qui implique que

$$\mathbf{P}_{p=0.4}\left(\sum_{i=1}^{25} X_i \leq c\right) = (1-\gamma)\,\mathbf{P}_{p=0.4}\left(\sum_{i=1}^{25} X_i = c\right) + 0.05 > 0.05.$$

On note que $T := \sum_{i=1}^{25} X_i \rightsquigarrow$Binomial$(25, 0.4)$ (la loi binomiale de paramètre $n = 25$ et $p = 0.4$). De la table statistique, de la loi binomiale, la plus petite c telle que $\mathbf{P}(T \leq c) > 0.05$ est $c = 6$, pour la quelle $\mathbf{P}(T \leq 7) = 0.074$. Par conséquent

$$\gamma = \frac{\mathbf{P}(T \leq 6) - 0.05}{\mathbf{P}(T = 6)} = \frac{0.074 - 0.05}{\mathbf{P}(T \leq 6) - \mathbf{P}(T \leq 5)} = \frac{0.074 - 0.05}{0.074 - \mathbf{P}(T \leq 5)}.$$

De la table statistique, de la loi binomiale on trouve $\mathbf{P}(T \leq 5) = 0.029$, donc

$$\gamma = \frac{0.074 - 0.05}{0.074 - 0.029} \simeq 0.53.$$

Ainsi le test optimal upp est

$$\delta = \delta(X_1, ..., X_{20}) = \begin{cases} 1 & \text{si } \sum_{i=1}^{20} x_i > 7 \\ 0.53 & \text{si } \sum_{i=1}^{20} x_i = 7 \\ 0 & \text{si } \sum_{i=1}^{20} x_i < 7 \end{cases}$$

3) Nous avons

$$\begin{aligned} \mathbf{E}_{p=0.2}\left[\delta(X_1, ..., X_{20})\right] &= 1 \times \mathbf{P}\left(\sum_{i=1}^{20} X_i > 7\right) + 0.327\mathbf{P}\left(\sum_{i=1}^{20} X_i = 7\right) + 0 \times \mathbf{P}\left(\sum_{i=1}^{20} X_i < 7\right) \\ &= \mathbf{P}\left(\sum_{i=1}^{20} X_i > 7\right) + 0.327\mathbf{P}\left(\sum_{i=1}^{20} X_i = 7\right). \end{aligned}$$

D'après la relation (1) cette dernière quantité égale $\alpha = 0.05$, ainsi

$$\mathbf{E}_{p=0.2}\left[\delta(X_1, ..., X_{20})\right] = 0.05.$$

4) La fonction puissance du test est

$$\begin{aligned}\pi(p) &= \mathbf{E}_p[\delta],\ 0<p<1 \\ &= 1\times\mathbf{P}(\delta=1)+0.327\times\mathbf{P}(\delta=0.327)+0\times\mathbf{P}(\delta=0).\end{aligned}$$

En d'autres termes $\pi(p)=\mathbf{P}(T>7)+0.327\mathbf{P}(T=7)$, où $T:=\sum_{i=1}^{20}X_i$ est une v.a binomiale de parametrer $(20,p)$, avec $0<p<1$. Il est clair que

$$\begin{aligned}\pi(p) &= 1-\mathbf{P}(T\leq 7)+0.327(\mathbf{P}(T\leq 7)-\mathbf{P}(T\leq 6)) \\ &= 1-0.673\mathbf{P}(T\leq 7)-0.327\mathbf{P}(T\leq 6) \\ &= 1-0.673F_p(7)-0.327F_p(6),\end{aligned}$$

où $F_p(x)$ désigne la fonction de répartition de la loi binomiale de paramètrer $(20,p)$. Explicitement la fonction puissance est

$$\pi(p)=1-0.673\sum_{k=0}^{7}C_{20}^k p^k(1-p)^{20-k}-0.327\sum_{k=0}^{6}C_{20}^k p^k(1-p)^{20-k},$$

pour $0<p<1$. Le graphe de la fonction puissance est donné par la figure Fig.1.

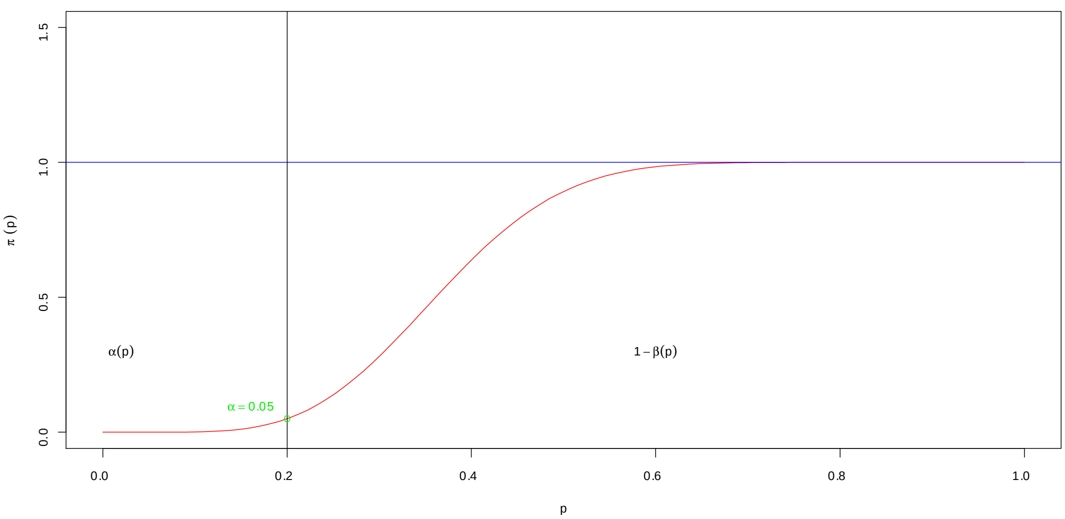


Fig.1

____

```
f<-function(x){1-0.673*pbinom(7,20,x)-0.327*pbinom(6,20,x)}
x<-seq(0,1,length=100)
plot(x,f(x),type="l",col="red",ylim=c(0,1.5),xlab=expression(p),ylab=expression(pi~(p)))
abline(h=1,col="blue")
abline(v=0.2)
points(0.2,0.05,col="green2")
text(0.16,0.1,expression(alpha==0.05),col="green2")
text(0.6,0.3,expression(1-beta(p)))
text(0.02,0.3,expression(alpha(p)))
```

Université Mohamed Khider, Biskra
Faculté des Sc. Exactes et Sc. de la Nature et la Vie
Département de Mathématiques
Master 1: 2021/2022

Corrigé-type de l'interrogation N°**2**

Exercice 1 (10pts)

Dans une production il y a une proportion p d'articles défectueux. On prélève 20 articles pour réaliser le test

$$\begin{cases} H_0 : p \leq 0.2 \\ H_1 : p > 0.2 \end{cases},$$

au niveau de signification 0.05.

1. Quel type de test s'agit-il: unilatéral ou blilatéral? $(1pt)$
2. Construire le test uniformément le plus puissant, noté $\delta = \delta(x_1, ..., x_{20})$. $(3pts)$
3. Calculer $\mathbf{E}_{p=0.2}[\delta(X_1, ..., X_{20})]$. $(2pts)$
4. Déterminer la fonction puissance de δ, notée $\pi(p)$. $(2pts)$
5. Tracer le graphe de la fonction $\pi(p)$. $(2pts)$

****

Solution.

1) Il s'agit d'un test unilatéral à droite.

2) Ici**,** la variable aléatoire X prend une valeur 1 (qui correspond à un article défectueux) avec une probabilité p et elle prend la valeur 0 (qui correspond à un article non défectueux). Donc il s'agit d'une v.a qui suit la loi de Bernoulli de paramètre p. La fonction masse de cette v.a discrète X et définie comme suit:

$$\mathbf{P}(X = 1) = p \text{ et } \mathbf{P}(X = 0) = 1 - p.$$

Celle-ci peut être reformuler comme suit:

$$\mathbf{P}(X = x) = p^x (1 - p)^{1-x}, \ x = 0, 1.$$

Nous allons appliquer la méthode du test de *rapport de vraisemblance monotone***.** Soit $p_1 > p_2$ et écrivons

$$\frac{L_{p_1}}{L_{p_2}} = \frac{\prod_{i=1}^{20} p_1^{x_i} (1 - p_1)^{1-x_i}}{\prod_{i=1}^{20} p_2^{x_i} (1 - p_2)^{1-x_i}} = \frac{p_1^t (1 - p_1)^{20-t}}{p_2^t (1 - p_2)^{20-t}}, \text{ avec } t := \sum_{i=1}^{20} x_i.$$

Celle peut être réécrite comme suit:

$$\frac{L_{p_1}}{L_{p_2}} = \left(\frac{1 - p_1}{1 - p_2}\right)^{20} \left(\frac{p_1 (1 - p_2)}{p_2 (1 - p_1)}\right)^t.$$

On pose $b := \left(\frac{1 - p_1}{1 - p_2}\right)^{20} > 0$, et $a := \frac{p_1(1-p_2)}{p_2(1-p_1)} > 1$ (car $p_1 > p_2$), ainsi $t \to \frac{L_{p_1}}{L_{p_2}}(t) = ba^t$ est une fonction croissante en t (car $a > 1$). Nous allons alors appliquer la proposition 3, pour avoir le test upp :

$$\delta = \begin{cases} 1 & \text{si } \sum_{i=1}^{20} x_i > c \\ \gamma & \text{si } \sum_{i=1}^{20} x_i = c \\ 0 & \text{si } \sum_{i=1}^{20} x_i < c \end{cases}$$

où c et $0 < \gamma < 1$ sont telles que

$$\mathbf{P}_{p=0.2}\left(\sum_{i=1}^{20} X_i > c\right) + \gamma \mathbf{P}_{p=0.2}\left(\sum_{i=1}^{20} X_i = c\right) = \alpha = 0.05. \tag{1}$$

En d'autres termes

$$1 - \mathbf{P}_{p=0.2}\left(\sum_{i=1}^{20} X_i \leq c\right) + \gamma \mathbf{P}_{p=0.2}\left(\sum_{i=1}^{20} X_i = c\right) = 0.05,$$

ce qui implique que

$$\mathbf{P}_{p=0.2}\left(\sum_{i=1}^{20} X_i \leq c\right) = \gamma \mathbf{P}_{p=0.2}\left(\sum_{i=1}^{20} X_i = c\right) + 0.95 > 0.95.$$

On note que $T := \sum_{i=1}^{20} X_i \rightsquigarrow$Binomial$(20, 0.2)$ (la loi binomiale de paramètre $n = 20$ et $p = 0.2$). De la table statistique, de la loi binomiale, la plus petite c telle que $\mathbf{P}(T \leq c) > 0.95$ est $c = 7$, pour la quelle $\mathbf{P}(T \leq 7) = 0.968$. Par conséquent

$$\gamma = \frac{\mathbf{P}(T \leq 7) - 0.95}{\mathbf{P}(T = 7)} = \frac{0.968 - 0.95}{\mathbf{P}(T \leq 7) - \mathbf{P}(T \leq 6)} = \frac{0.968 - 0.95}{0.968 - \mathbf{P}(T \leq 6)}.$$

De la table statistique, de la loi binomiale on trouve $\mathbf{P}(T \leq 6) = 0.913$, donc

$$\gamma = \frac{0.968 - 0.95}{0.968 - 0.913} = 0.327.$$

Ainsi le test optimal upp est

$$\delta = \delta(X_1, ..., X_{20}) = \begin{cases} 1 & \text{si } \sum_{i=1}^{20} x_i > 7 \\ 0.327 & \text{si } \sum_{i=1}^{20} x_i = 7 \\ 0 & \text{si } \sum_{i=1}^{20} x_i < 7 \end{cases}$$

3) Nous avons

$$\begin{aligned} \mathbf{E}_{p=0.2}[\delta(X_1, ..., X_{20})] &= 1 \times \mathbf{P}\left(\sum_{i=1}^{20} X_i > 7\right) + 0.327\mathbf{P}\left(\sum_{i=1}^{20} X_i = 7\right) + 0 \times \mathbf{P}\left(\sum_{i=1}^{20} X_i < 7\right) \\ &= \mathbf{P}\left(\sum_{i=1}^{20} X_i > 7\right) + 0.327\mathbf{P}\left(\sum_{i=1}^{20} X_i = 7\right). \end{aligned}$$

D'après la relation (1) cette dernière quantité égale $\alpha = 0.05$, ainsi

$$\mathbf{E}_{p=0.2}[\delta(X_1, ..., X_{20})] = 0.05.$$

4) La fonction puissance du test est

$$\begin{aligned} \pi(p) &= \mathbf{E}_p[\delta], \ 0 < p < 1 \\ &= 1 \times \mathbf{P}(\delta = 1) + 0.327 \times \mathbf{P}(\delta = 0.327) + 0 \times \mathbf{P}(\delta = 0). \end{aligned}$$

En d'autres termes $\pi(p) = \mathbf{P}(T > 7) + 0.327\mathbf{P}(T = 7)$, où $T := \sum_{i=1}^{20} X_i$ est une v.a binomiale de parametrer $(20, p)$, avec $0 < p < 1$. Il est clair que

$$\begin{aligned} \pi(p) &= 1 - \mathbf{P}(T \leq 7) + 0.327(\mathbf{P}(T \leq 7) - \mathbf{P}(T \leq 6)) \\ &= 1 - 0.673\mathbf{P}(T \leq 7) - 0.327\mathbf{P}(T \leq 6) \\ &= 1 - 0.673F_p(7) - 0.327F_p(6), \end{aligned}$$

où $F_p(x)$ désigne la fonction de répartition de la loi binomiale de paramètrer $(20, p)$. Explicitement la fonction puissance est

$$\pi(p) = 1 - 0.673 \sum_{k=0}^{7} C_{20}^{k} p^{k} (1-p)^{20-k} - 0.327 \sum_{k=0}^{6} C_{20}^{k} p^{k} (1-p)^{20-k},$$

pour $0 < p < 1$. Le graphe de la fonction puissance est donné par la figure Fig.1.

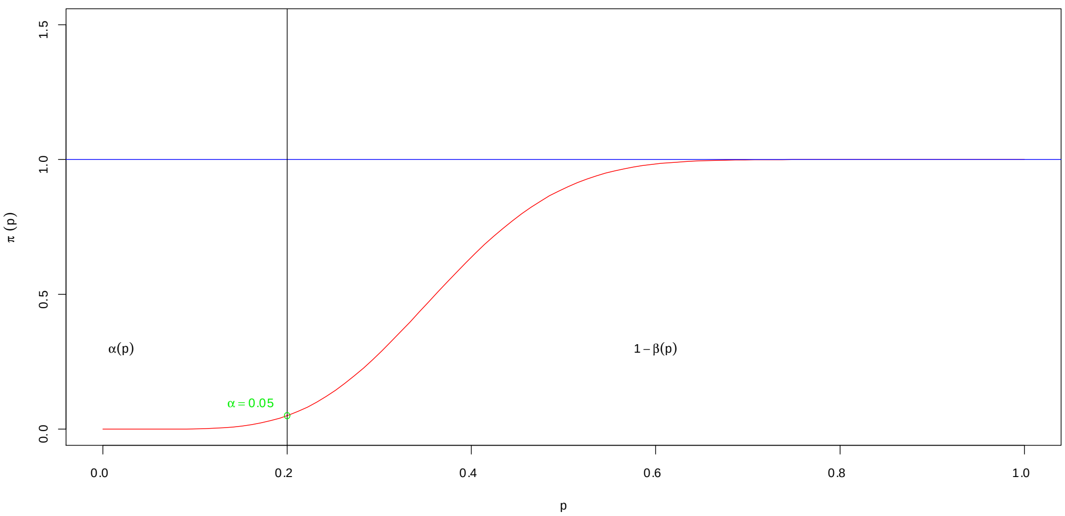


Fig.1

____

```
f<-function(x){1-0.673*pbinom(7,20,x)-0.327*pbinom(6,20,x)}
x<-seq(0,1,length=100)
plot(x,f(x),type="l",col="red",ylim=c(0,1.5),xlab=expression(p),ylab=expression(pi~(p)))
abline(h=1,col="blue")
abline(v=0.2)
points(0.2,0.05,col="green2")
text(0.16,0.1,expression(alpha==0.05),col="green2")
text(0.6,0.3,expression(1-beta(p)))
text(0.02,0.3,expression(alpha(p)))
```