République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider - BISKRA
Faculté des Sciences Exactes
et des Sciences de la Nature et de la Vie
Département de Mathématiques
Spécialités : Statistique Non paramétrique

Concours d'accès à la formation de troisième cycle (Doctorat LMD)
Date : 01/04/2021
**Durée 02H, 00**

# Epreuve: Statistique Non paramétrique

**Exercice 1** *(6pts)*
Soit $X$ une variable aléatoire issue d'une loi uniforme sur $[\theta_1; \theta_2]$, tel que $0 < \theta_1 < \theta_2$.

1. Donner l'expression de la fonction de densité et de la fonction de répartition de $X$.
2. Donner l'expression de de l'espérance et de la variance de $X$ en fonction de $\theta_1$ et $\theta_2$.
3. En se basant sur l'observation d'un échantillon $X_1; ...; X_n$ du même loi que $X$, en utilisant la méthode des moments, donner un estimateur du paramètre $\theta = (\theta_1; \theta_2)$.
4. Vérifier que l'estimateur de $\hat{\theta}_2$ obtenu en (3) est asymptotiquement sans biais.

**Exercice 2** *(7pts)*
On souhaite étudier le temps $X$ (en mois) mis par un étudiant (fille ou garçon, sans distinction) diplômé de Master pour obtenir un emploi à durée indéterminée. On relève ce temps pour $n$ jeunes diplomés.
On souhaite tester l'hypothèse $(\mathbf{H}_0)$ "$X$ suit la loi exponentielle de paramètre 1/6" sur la base de ces observations.

1. Donner la statistique de test de Kolmogorov-Smirnov pour ce problème de test, ainsi qu'une expression de cette statistique facile à utiliser en pratique.
2. Expliquer pourquoi la loi de cette statistique sous l'hypothèse $(\mathbf{H}_0)$ peut être tabulée pour toute valeur de $n$.
3. Quelle est la loi asymptotique de cette statistique sous l'hypothèse $(\mathbf{H}_0)$ ?
4. Pour $n = 10$, on a relevé les résultats suivants.

| Temps d'insertion | 5.5 | 10.2 | 30.2 | 16 | 7.5 | 4.1 | 10.3 | 15.3 | 4.2 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|

Quelle est la conclusion du test de Kolmogorov-Smirnov de niveau 5% ?

***Table de Kolmogorov-Smirnov :*** *pour différentes valeurs de $n$, on donne $q_{0.95}$ tel que*

$$P\left(\sup_{x\in[0,1]} |F_{U,n}(x) - x| \leq q_{0.95}\right) = 0.95$$

*(sans la racine carrée $\sqrt{n}$), lorsque $F_{U,n}$ est la fonction de répartition empirique associée à un n-échantillon de la loi uniforme sur [0, 1].*

| $n$ | 10 | 20 | 30 | $> 100$ |
|---|---|---|---|---|
| $q_{0.95}$ | 0.409 | 0.294 | 0.242 | $1.358/\sqrt{n}$ |

**Exercice 3** *(7pts)*
On mesure le diamètre de 7 pièces produites dans une usine dans deux températures différentes $(t_1, t_2)$, les mesures sont enregistrées dans le tableau suivant :

| Sujet | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Diamètre dans $t_1$ | 2.13 | 1.77 | 1.68 | 2.4 | 2.12 | 1.92 | 2.08 |
| Diamètre dans $t_2$ | 1.19 | 1.55 | 1.62 | 1.89 | 2.01 | 1.91 | 2.10 |

Testez, au seuil de signification $\alpha = 0.025$, si la température à un impact sur les mesures de diamètre.
**On donne:** La valeurs critique du test à un seuil de risque $\alpha = 5\%$ est 3.

Choisi Sujet 1

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mohamed Khider - BISKRA
Faculté des Sciences Exactes
et des Sciences de la Nature et de la Vie
Département de Mathématiques
Spécialités : Statistique Non paramétrique

**Concours d'accès à la formation de troisième cycle (Doctorat LMD)**
Date : 01/04/2021
**Durée 02H, 00**

**Epreuve: Statistique Non paramétrique**

# Solution

**Solution de l'Exercice N° 1** *(Estimation par la méthode des moments et MLE)*

1. La fonction de densité $f$ et de la fonction de répartition $F$ de $X$ sont données respectivement par:

$$f(x)=\begin{cases}\frac{1}{\theta_2-\theta_1}, & \text{si } x\in[\theta_1;\theta_2];\\ 0, & \text{sinon.}\end{cases} \quad \text{(0.75pt)} \quad \text{et} \quad F(x)=\begin{cases}0, & \text{si } x<\theta_1;\\ \frac{x-\theta_1}{\theta_2-\theta_1}, & \text{si } x\in[\theta_1;\theta_2]\\ 1, & \text{si } x>\theta_2.\end{cases} \quad \text{(0.75pt)}$$

2. L'expression de l'espérance et de la variance de $X$ en fonction de $\theta_1$ et $\theta_2$ sont :
   (a) La moyenne: $E(X)=\frac{\theta_1+\theta_2}{2}$. (0.75pt)
   (b) La variance: $Var(X)=E(X^2)-E(X)^2=\frac{(\theta_2-\theta_1)^2}{12}$. (0.75pt)

3. En se basant sur l'observation d'un échantillon $X_1;...;X_n$ du même loi que $X$, donner un estimateur du paramètre $(\theta_1;\theta_2)$ par la méthode des moments est:

$$\text{(1pt)}\quad \begin{cases}\overline{X} &= \frac{\theta_1+\theta_2}{2}\\ \hat{\sigma}^2 &= \frac{(\theta_2-\theta_1)^2}{12}\end{cases} \Rightarrow \begin{cases}\hat{\theta}_1 &= \overline{X}-\sqrt{3}\hat{\sigma}. \quad \text{(0.5pt)}\\ \hat{\theta}_2 &= \overline{X}+\sqrt{3}\hat{\sigma}. \quad \text{(0.5pt)}\end{cases}$$

avec $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{X})^2$.

4. L'estimateur $\hat{\theta}_2$ Asymptotiquement sans biais?
**Première Possibilité:**

$$\begin{aligned}E(\hat{\theta}_2) &= E(\overline{X})+\sqrt{3}E(\hat{\sigma})\\ &= \frac{\theta_1+\theta_2}{2}+\frac{\sqrt{\frac{n-1}{n}}(\theta_2-\theta_1)}{2} \qquad \text{(0.5pt)}\\ &= \underbrace{\left(\frac{1-\sqrt{\frac{n-1}{n}}}{2}\right)}_{A_n}\theta_1+\underbrace{\left(\frac{1+\sqrt{\frac{n-1}{n}}}{2}\right)}_{B_n}\theta_2\end{aligned}$$

On constate que $A_n$ tend vers 0 et $B_n$ tend vers 1 lorsque $n$ tend vers l'infinie $\Rightarrow$ l'estimateur $\hat{\theta}_2$ obtenu via la méthode des moments est *asymptotiquement sans biais.* (0.5pt)

**Deuxième Possibilité:**

$$\begin{aligned}Biais(\theta_2)=\hat{E}(\hat{\theta}_2)-\theta_2 &= E(\overline{X})+\sqrt{3}E(\hat{\sigma})-\theta_2\\ &= \frac{\theta_1+\theta_2}{2}+\frac{\sqrt{\frac{n-1}{n}}(\theta_2-\theta_1)}{2}-\frac{2\theta_2}{2} \qquad \text{(0.5pt)}\\ &= \underbrace{\left(\frac{1-\sqrt{\frac{n-1}{n}}}{2}\right)}_{A_n}\theta_1+\underbrace{\left(\frac{\sqrt{\frac{n-1}{n}}}{2}-1\right)}_{B_n}\theta_2\end{aligned}$$

On constate que $A_n$ tend vers 0 et $B_n$ tend vers 0 lorsque $n$ tend vers l'infinie $\Rightarrow$ l'estimateur $\hat{\theta}_2$ obtenu via la méthode des moments est *asymptotiquement sans biais.* (0.5pt)

**Remarque:** *Une seule possibilité est suffisante comme réponse.*

Université Mohamed Khider - BISKRA
Faculté des Sciences Exactes
et des Sciences de la Nature et de la Vie
Département de Mathématiques
Spécialités : Statistique Non paramétrique

**Concours d'accès à la formation de troisième cycle (Doctorat LMD)**
Date : 01/04/2021
**Durée 02H, 00**

## Epreuve: Statistique Non paramétrique

**Solution de l'Exercice N° 2** *(Test d'ajustement K.S.)*

Soit $Z = (X_1, ..., X_n)$ un $n$-échantillon de la loi $P$ de la v.a. $X$ (absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$), modélisant le temps d'insertion d'un jeune diplomé de Master. Soit $z = (x_1, ..., x_n)$ l'observation de cet échantillon. En utilisant le test de KolmogorovSmirnov d'adéquation, on veut tester :

$$(\mathbf{H}_0)\ P \text{ est une loi } Exp(1/6) \text{ contre } (\mathbf{H}_1)\ P \text{ n'est pas une loi } Exp(1/6).$$

1. Statistique de test :
$$D_n(Z) = \sqrt{n} \sup_{x \in \mathbb{R}_+} |F_n(x) - F(x)|, \qquad \text{(1pt)}$$
où $F_n$ est la fonction de répartition associée à l'échantillon $Z$ et $F$ est la fonction de répartition de la loi $E(1/6)$ :
$$F(x) = 1 - e^{-x/6}, x \geq 0. \qquad \text{(0.5pt)}$$
Cette statistique s'exprime également sous la forme :
$$D_n(z) = \sqrt{n} \max_{1 \leq i \leq n} \left\{ \left| F\left(x_{(i)}\right) - \frac{i}{n} \right|, \left| F\left(x_{(i)}\right) - \frac{i-1}{n} \right| \right\}.$$

2. On sait que lorsque $P = Exp(1/6)$, pour tout $n$, $D_n(Z)$ suit la même loi que (0.5pt)
$$U_n = \sqrt{n} \sup_{x \in [0,1]} |F_{U,n}(x) - x|, \qquad \text{(0.5pt)}$$
où $F_{U,n}$ est la fonction de répartition associée à un $n$-échantillon de la loi uniforme sur $[0, 1]$. Cette loi peut donc être tabulée pour toute valeur de $n$. (0.5pt)

3. D'après le théorème de Kolmogorov, $D_n(Z)$ converge en loi vers une loi de fonction de répartiton $H$ définie par
$$H(x) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 x^2}. \qquad \text{(1pt)}$$
Cette loi est également tabulée.

4. On sait que lorsque $P$ n'est pas la loi $Exp(1/6)$, $D_n(Z) \to \infty$ p.s. donc la fonction de test est de la forme $\phi(z) = 1_{\{D_n(Z) \geq s\}}$.
Calcul de la constante $s$ : on a vu que lorsque $P = Exp(1/6)$, $D_n(Z)$ suit la même loi que
$$U_n = \sqrt{n} \sup_{x \in [0,1]} |F_{U,n}(x) - x|,$$
(0.5pt) (0.5pt)
On a $n = 10$ et le sup $s = 0.409$, donc pour un niveau 5% on choisit, $S = 0.409 \times \sqrt{10} = 1.293$.
Conclusion : Pour calculer la valeur de $D_n(z)$, on utilise le tableau suivant.

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_{(i)}$ | 4.1 | 4.2 | 5.5 | 7 | 7.5 | 10.2 | 10.3 | 15.3 | 16 | 30.2 |
| $F\left(x_{(i)}\right)$ | 0.495 | 0.503 | 0.6 | 0.689 | 0.713 | 0.817 | 0.82 | 0.922 | 0.931 | 0.993 |

(1pt)

D'après les résultats précédent on: $D_n(z) = 1.566 > S = 1.293$, donc on rejette l'hypothèse $(\mathbf{H}_0)$ au niveau 5%.
(0.5pt) (0.5pt)

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider - BISKRA
Faculté des Sciences Exactes
et des Sciences de la Nature et de la Vie
Département de Mathématiques
Spécialités : Statistique Non paramétrique

**Concours d'accès à la formation de troisième cycle (Doctorat LMD)**
Date : 01/04/2021
**Durée 02H, 00**

# Epreuve: Statistique Non paramétrique

**Solution de l'Exercice N° 3** *(Test d'ajustement Wilcoxon)*

- Justification du choix du test: (1pt)
  Ici, on observe deux échantillons issus de la même population et on ne sait pas la loi de ces deux échantillons, donc pour étudier s'il y a une différence entre eux, donc on utilise le test de Wilcoxon alors
- Hypothèse Nulle du test:

$$H_0 : \mu_{t1} = \mu_{t2} \quad \text{(1pt)}$$

- Tableaux des calculs intermidiares de la statistique du test:

| *pièces* | *diamètre en $t_1$* | *diamètre en $t_2$* | $D_i$ | $\|D_i\|$ | classement | $W_+$ | $W_-$ |
|---|---|---|---|---|---|---|---|
| 1 | 2.13 | 1.19 | 0.94 | 0.94 | 7 | 7 | |
| 2 | 1.77 | 1.55 | 0.22 | 0.22 | 6 | 6 | |
| 3 | 1.68 | 1.62 | 0.06 | 0.06 | 3 | 3 | |
| 4 | 2.4 | 1.89 | 0.51 | 0.51 | 5 | 5 | |
| 5 | 2.12 | 2.01 | 0.11 | 0.11 | 4 | 4 | |
| 6 | 1.92 | 1.91 | 0.01 | 0.01 | 1 | 1 | |
| 7 | 2.08 | 2.10 | −0.02 | 0.02 | 2 | | 2 |

(2.5pt)

- La statistique du test: D'après le tableau on déduit que:
  - $\sum W_+ = 26$ (0.5pt)
  - $\sum W_- = 2$ (0.5pt)
  - $W = \min(\sum W_+, \sum W_-) = 2$ (0.5pt)
- La decision sur $H_0$:

(0.5pt)

On a $W = 2 < W_{0.05}(8) = 3$ au seuil 5% et on décide de rejeter $H_0$ donc il ya une déffirence significative entre les deux mesures. (0.5pt)

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider - BISKRA
Faculté des Sciences Exactes
et des Sciences de la Nature et de la Vie
Département de Mathématiques
Spécialités : Statistique Non paramétrique

**Concours d'accès à la formation de troisième cycle (Doctorat LMD)**
Date : 01/04/2021
**Durée 02H, 00**

# Epreuve: Statistique Non paramétrique

**Exercice 1** *(6pts)* Supposons qu'on s'interesse à l'analyse descriptive des notes moyennes obtenues par les 60 candidats participants dans le concours d'Accès à la Formation de 3ème Cycle (Doctorat LMD) dans le département "*mathématiques*". Un recueil de données au niveau de ce département nous a fournis le tableau suivant :

| Note | [0 ; 4[ | [4 ; 6[ | [6 ; 8[ | [8 ; 10[ | [10;14[ |
|---|---|---|---|---|---|
| Effectifs ($n_i$) | 24 | $n_2$ | $n_3$ | 5 | 4 |

1. Sachant que $\overline{X} = 5$ alors déterminer $n_2$ et $n_3$.
2. Calculer les quartiles $Q_1$, $Q_2$ et $Q_3$ et le mode $Mo$ de la série..
3. Si un certain fractale $Q_p = 12$, alors déterminer la valeur de $p$.

**Exercice 2** *(8pts)* Soit $\hat{f}_n$ l'estimateur de Parzen-Rosenblatt de la densité $f$ dont la forme est:

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right),$$

avec $h$ est le pramètre de lissage et $K$ est la fonction noyau sur $\mathbb{R}$.

1. Montrez que si $f$ est de classe $C^2$ alors pour tout $x \in \mathbb{R}$ et $h \to 0$ on a:
$$Biais\{f_h(x)\} = \frac{h^2}{2}f''(x)\mu_2(K) + o(h^2); \quad \text{où } \mu_2(K) = \int_{\mathbb{R}} u^2K(u)du.$$
2. Montrez que pour $K \in \mathbf{L}^2$, on a pour tout $x \in \mathbb{R}$ et $nh \to \infty$:
$$Var\{f_h(x)\} = \frac{1}{nh}f(x)\int_{\mathbb{R}} K^2(u)\,du + o\left(\frac{1}{nh}\right);$$
3. En déduire une expression approchée pour l'Erreur Quadratique Moyenne Integrée (EQMI).
4. Trouvez la fenêtre optimale minimisant l'EQMI approchée.
5. Montrez que si $f \sim N(\mu;\sigma^2)$ alors
$$\int f''(x)^2dx = \frac{3}{8\sigma^5\sqrt{\pi}},$$
(*Indication: on utilisera le fait que pour* $Y \sim N(\mu;\sigma^2)$, $\mathbf{E}\{(X-\mu)^{2k}\} = (2k)!\sigma^{2k}/(2^k k!)$) et si de plus $K$ est gaussien, alors la fenêtre optimale minimisant l'EQMI approchée est:
$$h^* = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{1/5}, \ \hat{\sigma} \text{ est l'estimateur corrigé de l'écart-type.}$$
6. Expliquez l'intuition derrière la formule suivante
$$h_* = A\left(\frac{4}{3n}\right)^{1/5} \quad avec \ \ A = \min\left(\hat{\sigma}, \frac{X_{[3n/4]} - X_{[n/4]}}{1.349}\right),$$
où $X_{[n/4]}$ et $X_{[3n/4]}$ représentent réspectivement le premier et le troisième quartile

**Exercice 3** *(6pts)* Lors d'une première compilation d'un programme demandé à des étudiants d'informatiques, on observe le nombre d'erreurs suivant:

| Nombre d'erreurs | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Effectifs | 25 | 18 | 16 | 11 | 8 | 7 | 6 | 4 |

Testez, au seuil de signification $\alpha = 5\%$, l'ajustement des données au modèle

$$Pr(X = x) = \frac{8-x}{36}, x = 0, 1, ..., 7.$$

On donne $\chi^2_{0.95}(7) = 12.017$.

Sujet 2

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mohamed Khider - BISKRA
Faculté des Sciences Exactes
et des Sciences de la Nature et de la Vie
Département de Mathématiques
Spécialités : Statistique Non paramétrique

**Concours d'accès à la formation de troisième cycle (Doctorat LMD)**

Date : 01/04/2021

**Durée 02H, 00**

# Epreuve: Statistique Non paramétrique

**Exercice 1** *(6pts)*

On s'intéresse à la modélisation de la durée de vie d'une population constituée de femmes et d'hommes. Notons que les durées de vie des femmes (respectivement, des hommes) de cette population sont modélisées par l'intermédiaire d'une variable aléatoire exponentielle de paramètre $\lambda$ (respectivement, $\mu$). De plus, les femmes représentent une proportion $\beta$ de cette population. Supposons qu'on dispose d'un $n$ réalisations de l'échantillon $X_1, ..., X_n$ et soit la variable aléatoire indicatrice $Y$ définie par:

$$Y_i = \begin{cases} 1, & \text{si } X_i \text{ est l'age d'une femme;} \\ 0, & \text{si } X_i \text{ est l'age d'un homme.} \end{cases}$$

**Note:** Utiliser la densité exponentielle définie par $f(z) = \frac{1}{\theta}e^{-\frac{z}{\theta}}, \quad z \geq 0.$

1. Déterminer la fonction de densité et de répartition de la durée de vie $X$ des individus de cette population.
2. En se basant sur l'observation d'un échantillon $(X_1, Y_1); ...; (X_n, Y_n)$, proposé un estimateur pour la triplette $(\beta, \lambda, \mu)$.
3. En se basant uniquement sur l'observation d'un échantillon $X_1; ...; X_n$, proposé un modèle mathématique pour la détermination de lestimateur pour la triplette $(\beta, \lambda, \mu)$ par la méthode des moments (sans résoudre le modèle).

**Exercice 2** *(7pts)*

On considère les notes des 83 étudiants dont 23 filles et 60 garçons, et on obtient pour chaque classe de notes les effectifs suivants.

| Notes | [0,8] | ]8,11] | ]11,13.5] | ]13.5,20] |
|---|---|---|---|---|
| Filles | 5 | 8 | 5 | 5 |
| Garçons | 14 | 18 | 18 | 10 |

1. Construire un test du $\chi^2$ d'indépendance au niveau asymptotique 5%.
2. Quelle est la conclusion de ce test ?
3. Quelle critique peut-on opposer à cette conclusion ?

***On donne:*** $\chi^2_{(3,0.05)} = 0.35$, $\chi^2_{(4,0.95)} = 0.71$, $\chi^2_{(3,0.05)} = 7.82$ *et* $\chi^2_{(4,0.95)} = 9.49$.

**Exercice 3** *(7pts)*

On mesure les conditions physiques d'un échantillon aléatoire de 8 personnes par deux méthodes numériques différentes A et B, leurs mesures sont donnés par le tableau suivant :

| Sujet | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Méthode A | 11.2 | 8.6 | 6.5 | 17.3 | 14.3 | 10.7 | 9.8 | 13.3 |
| Méthode B | 10.4 | 12.1 | 9.1 | 15.6 | 16.7 | 10.7 | 12.8 | 15.5 |

Testez, au seuil de signification $\alpha = 5\%$, s'il y a une différence entre la méthode A et la méthode B.

**On donne:** La valeurs critique du test à un seuil de risque $\alpha = 5\%$ est 2.04.

Sujet 3