INSA **Examen de Probabilités et Statistiques** **2024/2025**
4ème année GEA

Durée de l'épreuve : 2 heures.
Sont autorisés : les notes manuscrites prises en cours et en TD, les documents mis à disposition par les enseignants sur Moodle (autorisés seulement sous forme papier), à l'exception des examens corrigés, ainsi qu'une calculatrice scientifique.
Le barème indiqué est provisoire.

Exercice 1 : Probabilités ($\sim$ 10 pts)

Partie I : Questionnaire à choix multiples

On considère un questionnaire comportant cinq questions. Pour chacune des cinq questions posées, trois propositions de réponses sont faites (A, B et C), une seule d'entre elles étant exacte.

Un candidat répond à toutes les questions posées en écrivant un mot-réponse de cinq lettres. Par exemple, le mot "BBAAC" signifie que le candidat a répondu B à la première et à la deuxième question, A à la troisième et à la quatrième question, et C à la cinquième.

1. Combien y a-t-il de mots-réponses possibles à ce questionnaire ?
2. On suppose que le candidat répond au hasard à chacune des cinq questions du questionnaire.
   a) Quelle est la probabilité de l'événement E : « le candidat a exactement une seule bonne réponse » ?
   b) Quelle est la probabilité de l'évènement F : « le candidat n'a aucune réponse exacte » ?
   c) Quelle est la probabilité de l'évènement G : « le mot-réponse du candidat est un palindrome » ? (On précise qu'un palindrome est un mot pouvant se lire indifféremment de gauche à droite ou de droite à gauche : par exemple, « ABCBA » est un palindrome.)
3. Chaque question vaut 2 points. Si le candidat répond au hasard à chacune des questions, quelle est la probabilité qu'il obtienne la moyenne à l'examen ?

Partie II : Influences

Soit F et R deux variables aléatoires à valeurs entières, dont la loi conjointe est donnée par le tableau ci-dessous :

$F \backslash R$	0	1	2	3
0	0.1	0.1	0.1	0.1
1	0.1	0.1	0.1	0
2	0.1	0.1	0	0
3	0.1	0	0	0

1. Etude d'une variable.
   a) Déterminer la loi de F.
   b) Déterminer l'espérance et la variance de F.
2. Etude de l'influence de F sur R.
   a) Quelles valeurs peut prendre la variable aléatoire $\mathbb{E}(R|F)$? Répondre en recopiant le tableau suivant :

F	$\mathbb{E}(R\|F)$
0	
1	
2	
3	

b) Donner la loi de $\mathbb{E}(R|F)$.

c) Calculer l'espérance de $\mathbb{E}(R|F)$. Est-ce qu'on pouvait s'attendre à ce résultat ?

d) Calculer la covariance du couple (F, R).

3. On considère maintenant une suite de $n \in \mathbb{N}^*$ variables aléatoires $(F_i, R_i)_{1\leq i\leq n}$, indépendantes et de même loi que le couple (F, R). On s'intéresse à la moyenne empirique :

$$\overline{R}_n = \frac{1}{n}\sum_{i=1}^{n} R_i.$$

En utilisant le théorème central limite, donner une approximation de la probabilité $\mathbb{P}(\overline{R}_{100} \leq 1.2)$.

Indication : on pourra utiliser l'un des résultats suivants : si $Z \sim \mathcal{N}(0, 1)$, on a :

$$\mathbb{P}(Z \leq 2.0) = 0.9772, \quad \mathbb{P}(Z \leq 1.5) = 0.9332 \quad \mathbb{P}(Z \leq 1) = 0.8412.$$

Exercice 2 : Statistique inférentielle ($\sim$ 10 pts)

Les trois parties suivantes sont indépendantes. Vous pouvez les traiter dans l'ordre de votre choix.

Partie I : Estimation ponctuelle

Soit X une variable aléatoire de densité

$$f_\theta(x) = \frac{2}{\theta^2} x\, \mathbb{1}_{[0,\theta]}(x), \quad \text{où } \theta > 0,$$

et $\mathbb{1}_{[0,\theta]}(x)$ désigne la fonction indicatrice de l'intervalle $[0, \theta]$.
On suppose que l'on observe un échantillon $\{X_1, \ldots, X_n\}$ de variables aléatoires indépendantes et identiquement distribuées suivant la loi de densité f_θ.

1. Montrer que f_θ est une densité de probabilité.
2. Calculer $\mathbb{E}[X]$ et $\mathrm{Var}(X)$.
3. Montrer que la fonction de répartition de X est donnée par :

$$F_X(t) = \begin{cases} 0 & \text{si } t < 0, \\ \left(\frac{t}{\theta}\right)^2 & \text{si } 0 \leq t \leq \theta, \\ 1 & \text{si } t > \theta. \end{cases}$$

4. **Premier estimateur :**

   (a) Par la méthode des moments, en déduire un estimateur $\hat{\Theta}_n^{(1)}$ de θ.

   (b) Calculer le biais et la variance de $\hat{\Theta}_n^{(1)}$. Que peux-on conclure sur ses propriétés ?

5. **Deuxième estimateur :**

   (a) Montrer que l'estimateur du maximum de vraisemblance pour θ est :

$$\hat{\Theta}_n^{(2)} = \max(X_1, \ldots, X_n).$$

   (b) En utilisant la question 3, en déduire la fonction de répartition de $\hat{\Theta}_n^{(2)}$.

   (c) En déduire la densité de $\hat{\Theta}_n^{(2)}$.

   (d) L'estimateur $\hat{\Theta}_n^{(2)}$ est-il sans biais ? asymptotiquement sans biais ?

   (e) Calculer l'erreur quadratique moyenne $\mathbb{E}\Big[\big(\hat{\Theta}_n^{(2)} - \theta\big)^2\Big]$.

6. Lequel des deux estimateurs choisiriez-vous pour estimer θ à partir d'un échantillon réel ? Justifiez brièvement.

Partie II : Estimation par intervalles de confiance

On effectue un sondage sur un échantillon de 400 électeurs. On relève 212 intentions de vote en faveur d'un candidat A, 188 en faveur de B.

1. Donner au niveau de 95%, un intervalle de confiance asymptotique des intentions de vote en faveur de A dans la population entière.
2. Quelle taille minimale de l'échantillon faudrait-il prendre pour que, au même niveau de 95% et pour la même proportion de votants en faveur de A, l'intervalle ne contienne pas la valeur 0.5 ?

Note : On pourra utiliser le résultat d'une des commandes `R` suivantes où $\alpha = 0.05$:
qnorm(alpha) = -1.644854 ; *qnorm(alpha/2)= -1.959964* ; *qnorm(1-alpha) = 1.644854* ;*qnorm(1-alpha/2)= 1.959964*.

Partie III : Tests statistiques d'hypothèses

Dans les années 70, les athlètes féminines de la République Démocratique Allemande (RDA) étaient soupçonnées de dopage par prise de substances hormonales virilisantes (dites androgènes). Des mesures de la quantité d'androgènes par litre de sang ont été réalisées sur 9 athlètes, donnant les résultats suivants (en unités arbitraires) :

$$3.22 \quad 3.07 \quad 3.17 \quad 2.91 \quad 3.40 \quad 3.58 \quad 3.23 \quad 3.11 \quad 3.62.$$

1. On souhaite vérifier si la moyenne de la quantité d'androgènes chez ces athlètes est différente de la moyenne connue chez une femme lambda, qui est de 3.1 unités.
   (a) Formuler les hypothèses d'un test statistique adapté.
   (b) Quelles sont les conditions à vérifier pour effectuer ce test ?
   (c) Effectuer le test au risque $\alpha = 5\%$.

   Note : On pourra utiliser le résultat d'une des commandes `R` suivantes où $\alpha = 0.05$ et $n = 9$:
   qt(alpha, df=n-1) = -1.859548 ; *qt(alpha/2, df=n-1)= -2.306004* ; *qt(1-alpha, df=n-1) = 1.859548* ;*qt(1-alpha/2, df=n-1)= 2.306004*.
2. On souhaite maintenant tester si la moyenne de la quantité d'androgènes chez ces athlètes est supérieure à la moyenne connue chez une femme lambda, qui est de 3.1 unités. Voici la sortie du test réalisé avec `R` :

```
 One Sample t-test

data:  androgenes
t = 1.996, df = 8, p-value = 0.041
alternative hypothesis: true mean is greater than 3.1
95 percent confidence interval:
 3.152553      Inf
sample estimates:
mean of x
 3.231111
```

   Interpréter ces résultats au risque $\alpha = 5\%$, en explicitant toutes les étapes de construction du test statistique utilisé.