STT-2920 | ***Examen 2*** | **09 décembre 2020**
Probabilités pour ingénieurs | | **15 h 30 – 17 h 20**

Directives :

— Vous avez **110 minutes** pour faire cet examen, qui compte pour **50%** de la note finale.

— L'examen compte **7 questions** réparties sur **9 pages** (en comptant celle-ci).

— Vous pouvez utiliser une calculatrice scientifique **munie de l'autocollant de la faculté des sciences et de génie** et une feuille aide-mémoire manuscrite recto-verso de format lettre (8.5"×11").

— Des tables de lois sont distribuées avec cet examen. Merci de ne pas écrire dessus !

— Donnez tous les développements et calculs. Sauf dans les questions à choix multiples, **toute** réponse doit être convenablement **justifiée** pour mériter des points.

— Éteignez et rangez tout appareil électronique (téléphone cellulaire, téléavertisseur, lecteur de musique, etc.). Toute communication est interdite, et ce tant que tous les examens n'ont pas été ramassés à la fin de la période allouée.

— Il est interdit de regarder les questions avant d'en avoir reçu la permission explicite du surveillant et il est interdit d'écrire à partir de l'instant où le surveillant donne l'instruction de déposer les crayons à la fin de l'examen. Le refus d'obtempérer à cette directive résultera en une **pénalité automatique de 25 points**.

— Veuillez téléverser vos réponses moyennant un seul fichier PDF (Nom-Prénom-Numéroétudiant.pdf) dans votre boîte de dépôt, à la fin de l'examen.

— **Seulement s'il y a une déficience technique avec le portail de cours, veuillez m'envoyer par courriel (khader.khadraoui@mat.ulaval.ca) vos réponses, à la fin de l'examen**.

— Veuillez lire et ensuite signer la déclaration d'intégrité suivante : « Je comprends que le présent examen contribue à évaluer mes apprentissages dans le cadre d'un cours visé par le Programme d'agrément universitaire du Bureau canadien d'agrément des programmes de génie. Je certifie sur l'honneur à avoir complété l'examen seul.e et en conformité avec les consignes de l'évaluation. Je m'engage à ne pas utiliser d'informations rendues disponibles par une autre personne (inscrite au cours ou non) pendant l'examen. Je m'engage à ne pas fournir pas aucune information à une autre personne inscrite au cours ou non pendant mon ou pendant son examen. Je suis conscient.e qu'une fausse déclaration m'expose à des sanctions disciplinaires en vertu du Règlement disciplinaire à l'intention des étudiants et étudiantes de l'Université Laval. »

Je suis bien l'étudiant dont le nom et le numéro de dossier sont écrits ci-dessous. J'ai lu et compris les directives et je m'engage à les respecter.	
Nom :	
Prénom :	
Matricule :	
Signature :	

À remplir par les correcteurs

Q1 (/18)	Q2 (/20)	Q3 (/18)	Q4 (/20)	Q5 (/12)	Q6 (/12)	Q7 (/ 10)	**Total (/110)**

Question 1 (18 points)

On s'intéresse à une variable aléatoire X dont la densité est donnée par

$$f_\theta(x) = \begin{cases} \frac{1}{2}, & \text{si } \theta + 1 \leq x \leq \theta + 3\,; \\ 0, & \text{ailleurs,} \end{cases}$$

et pour laquelle on vous dit que $E(X) = \theta + 2$ et $Var(X) = \frac{1}{3}$.

Supposons que l'on dispose d'un échantillon aléatoire $X_1, \ldots, X_n$ provenant de cette loi. On considère deux estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$ pour θ, et on vous dit que

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad E(\hat{\theta}_2) = \theta \quad \text{et} \quad Var(\hat{\theta}_2) = \frac{16}{n}.$$

(a) Calculez le biais, la variance et en déduire l'erreur quadratique moyenne de $\hat{\theta}_1$. (12 points)

(b) Si la taille de l'échantillon est $n \geq 4$, quel est le meilleur estimateur de θ parmi $\hat{\theta}_1$ et $\hat{\theta}_2$? (6 points)

(Cochez la bonne réponse et justifiez)

- ☐ $\hat{\theta}_1$.
- ☐ $\hat{\theta}_2$.
- ☐ Les deux sont aussi bons.

Justification :

Question 2 (20 points)

Les périodes de retour des inondations dans 36 villes sont des variables aléatoires distribuées selon une même loi dont la moyenne est de 3 ans et dont l'écart-type est de 3 ans.

(a) Avec l'aide du théorème central limite, calculez une approximation pour la probabilité $\mathbb{P}(\bar{R} > 2.75)$ où $\bar{R}$ est la moyenne échantillonnale des 36 périodes de retour donnée par $\bar{R} = \frac{1}{36}\sum_{i=1}^{36} R_i$, où R_i est la période de retour (en années) des inondations dans la ville i. (8 points)

(b) Un ingénieur doit construire un pont sur une rivière qui traverse l'une de ces 36 villes. Le niveau maximal des eaux (mesuré par rapport à un certain point de référence) de la rivière est une variable aléatoire qui suit la loi normale de moyenne 3 mètres et d'écart-type 1/3 mètre. Quelle doit être la hauteur du pont (par rapport au point de référence) pour qu'il soit inondé en moyenne une année sur 40 ? (12 points)

Question 3 (18 points)

On considère $X_1, X_2, X_3, \dots$ des variables aléatoires indépendantes de même loi Bernoulli $\mathcal{B}(1,p)$ avec $0 < p < 1$. Pour tout $n \in \mathbb{N}^*$, on pose :

$$Z_n = X_{n+1} + X_n.$$

(a) Déterminez la loi de Z_n. Ensuite, calculez $\mathbb{E}[Z_n]$ et $\mathbb{V}[Z_n]$. Les variables Z_n sont-elles indépendantes ? (9 points)

(b) Soit $T_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} Z_i$. Calculez $\mathbb{E}[T_n]$ et $\mathbb{V}[T_n]$. (9 points)

Question 4 (20 points)

On s'intéresse à une variable aléatoire X dont la densité est donnée par

$$f_\theta(x) = \begin{cases} \dfrac{\Gamma\left(\theta + \dfrac{3}{2}\right)}{5\sqrt{\pi}\Gamma(\theta)} \dfrac{x^{\theta-1}\sqrt{(100-x)}}{100^\theta}, & \text{si } 0 \leq x \leq 100, \\ \\ 0, & \text{ailleurs,} \end{cases}$$

où $\Gamma(\cdot)$ est la fonction gamma de Euler. Dans ce qui suit, on suppose qu'on a un échantillon aléatoire $X_1, X_2, \ldots, X_n$ issu de cette distribution.

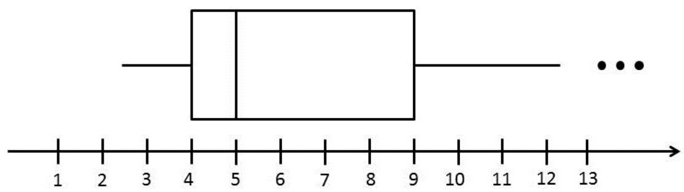
(a) Trouvez l'estimateur $\hat{\theta}_{\text{MOM}}$ par la méthode des moments pour θ. (10 points)

(b) Obtenez la fonction log-vraisemblance de l'échantillon observé $x_1, \ldots, x_n$. Que peut-on dire sur l'estimateur du maximum de vraisemblance $\hat{\theta}_{\text{MAX}}$ ici ? (10 points)

Question 5 (12 points)

Les parties (A), (B) et (C) sont indépendantes.

A) Vous recueillez un échantillon de taille 100 et produisez le diagramme en boîte ci-dessous.



Lequel/lesquels des énoncés suivants est/sont vrai(s) ? (2 points)

I. L'échantillon semble provenir d'une distribution symétrique.

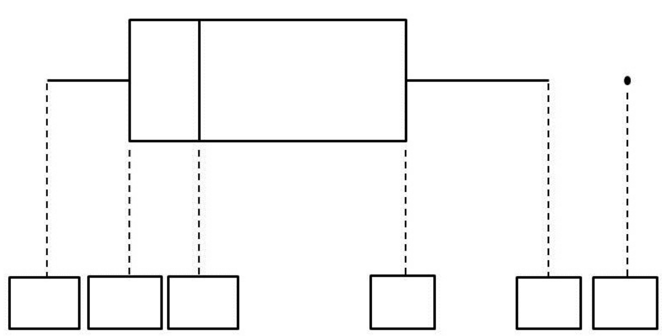
II. Il y a 4 fois plus d'observations entre 5 et 9 qu'entre 4 et 5.

- ☐ Aucun des énoncés n'est vrai.
- ☐ Seul l'énoncé **I** est vrai.
- ☐ Seul l'énoncé **II** est vrai.
- ☐ Les deux énoncés sont vrais.

B) On vous donne un échantillon de taille 10, qui vous est présenté (ordonné) ci-dessous, dont la médiane est 20 et les premier et troisième quartiles sont 16 et 28, respectivement :

$$8,\ 11,\ 16,\ 18,\ 19,\ 21,\ 22,\ 28,\ 40,\ 54.$$

Inscrivez les valeurs correpondantes dans les cases sous le diagramme en boîte ci-dessous. (6 points)



La question 5 se poursuit à la page suivante ...

Question 5 (suite)

C) Un échantillon formé par 15 mesures de températures (en ^oC) enregistrées dans une certaine région :

6 7.5 9 10 11 11.8 12 13 13.5 14.5
15 16 17 18 20.

Pour ces données on vous dit que les premier, second et troisième quartiles sont, respectivement, 10.75, 13 et 15.75.

(a) Y a-t-il une (ou des) donnée(s) aberrante(s) parmi les températures enregistrées ? (2 points)

Oui : □ Non : □

Justification :

(b) On s'est aperçu qu'il y a eu une erreur de lecture lors de la prise des mesures et que la température 17^oC est en réalité doit être remplacée par une valeur comprise entre 18^oC et 20^oC. Lequel/lesquels des énoncés ci-dessous est/sont vrai(s) ? (2 points)

□ La moyenne et la médiane sont affectées.

□ Seule la moyenne est affectée.

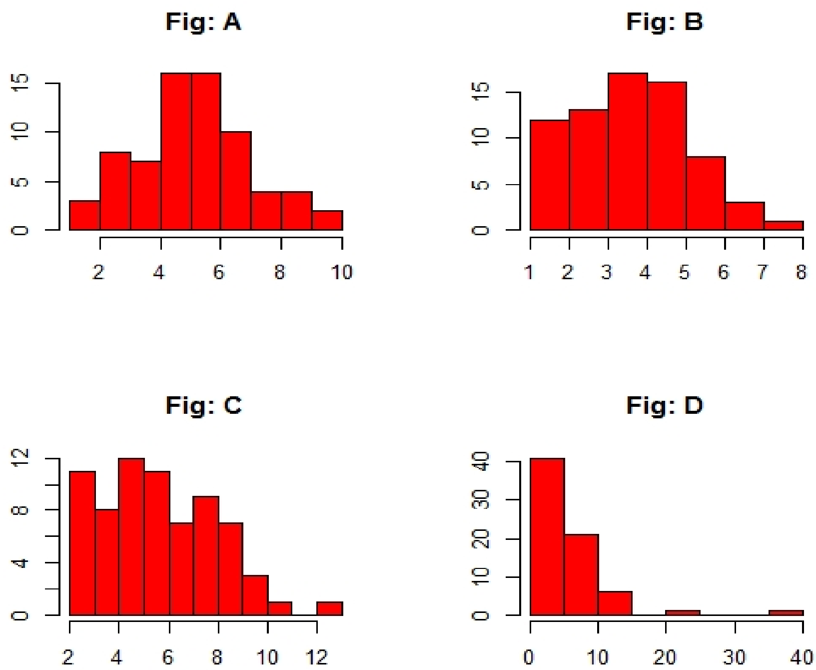
□ Seule la médiane est affectée.

□ Ni la moyenne, ni la médiane est affectée.

Question 6 (12 points)

On a simulé quatre échantillons aléatoires provenant de différentes lois théoriques avec le logiciel Rcommander (dont un issu d'une loi normale). Pour chacun, on a créé un histogramme pour illustrer les caractéristiques de sa distribution.

- 4 points pour une bonne réponse,
- 0 point pour l'absence de réponse,
- -2 point pour une mauvaise réponse.



(a) Parmi ces quatre figures, lequel qui a été construit avec l'échantillon simulé à partir d'une loi normale ?

☐ Fig : A ☐ Fig : B ☐ Fig : C ☐ Fig : D

Justification :

(b) La moyenne de l'échantillon de la Fig : A vaut environ...

☐ 4 ☐ 5 ☐ 4.25 ☐ 6 ☐ 5.75

Justification :

(c) La médiane de l'échantillon de la figure : A vaut environ...

☐ 4 ☐ 5 ☐ 4.25 ☐ 6 ☐ 5.75

Justification :

<u>Question 7</u> (10 points)

Déterminez si les énoncés suivants sont vrais ou faux.

- 1 point pour une bonne réponse,
- 0 point pour l'absence de réponse,
- -1 point pour une mauvaise réponse.

V F Si $Cov(H, G) = 0$, alors $Cov(2H + 3, 5G) = 0$.

V F Si le profit qu'on tire de la vente d'un item est de 3 $, alors la corrélation entre le nombre d'items vendus et le revenu net est égale à 1.

V F On suppose que les personnes atteints de Covid-19 et ayant besoin de soins intensifs arrivent à un hôpital comme un processus de Poisson avec une intensité de 5 malades par jour. La probabilité que 5 malades dans ce cas arriveront durant la prochaine journée est égale à la probabilité que 4 malades dans ce cas arriveront durant la même prochaine journée.

V F On suppose que les personnes atteintes de Covid-19 et ayant besoin de soins intensifs arrivent à un hôpital comme un processus de Poisson avec une intensité de 5 malades par jour. Le temps moyen entre les arrivées de ces malades est 4.8 heures plus ou moins 0.98 heures.

V F Selon les données du mois de décembre 2020 de l'Institut National de Santé Publique du Québec (INSPQ) la probabilité d'apparition du Covid-19 parmi la population de la province est 0.0177 (sur une population totale de 8 572 054 habitants). Si N dénote le nombre de personnes atteintes du Covid-19 dans la province, alors $N \sim$ binomiale(8 572 054, 0.0177).

V F Selon les données du mois de décembre 2020 de l'Institut National de Santé Publique du Québec (INSPQ) la probabilité du décès à cause du Covid-19 parmi les personnes atteintes dans la province est 0.048. On choisit 3250 personnes au hasard parmi les 151939 cas confirmés de Covid-19. Si N dénote le nombre de malades décédés parmi les 3250 cas confirmés choisis, alors $N \sim$ hypergéométrique(3250, 7293, 144646).

V F On suppose que la loi normale avec moyenne 100 et avec écart-type 16 est un bon modèle pour la distribution des quotients intellectuels dans la population adulte au Canada. On choisit 25 individus au hasard dans cette population. La probabilité que la moyenne des quotients intellectuels de ces 25 individus soit supérieure à 108 est 0.0062.

V F Si l'écart-type d'un échantillon est nul, c'est que toutes les observations valent 0.

V F L'écart-type, la variance et la covariance sont des mesures permettant d'évaluer la dispersion des valeurs dans une distribution.

V F Les valeurs minimale et maximale d'un jeu de données peuvent toujours être identifiées sur un diagramme en boîte.