Documents autorisés : cours, TD, notes manuscrites, calculatrice. Barème indicatif : 4 + 6 + 4 + 6 Durée : 1h 30.

Les résultats sont présentés avec trois chiffres significatifs.

Exercice 1

Calcul intégral
On note f la fonction définie sur $\mathbb{R}$ par $f(x) = xe^{-x}$ sur $[0, +\infty[$ et $f(x) = 0$ sinon.
f est une densité de probabilité (démontré en TD).
On note X une variable aléatoire absolument continue de densité f.
On souhaite calculer $E(X) = \displaystyle\int_{-\infty}^{+\infty} xf(x)\,dx$.

1. En utilisant une intégration par parties, exprimer $F(x) = \displaystyle\int_0^x t^2e^{-t}\,dt$ en fonction de $\displaystyle\int_0^x te^{-t}\,dt$.
2. En déduire $\displaystyle\int_0^{+\infty} x^2e^{-x}\,dx$ puis $E(X)$.
   Indications
   (a) $\lim\limits_{x\to+\infty} x^2e^{-x} = 0$,
   (b) $\displaystyle\int_0^{+\infty} xe^{-x}\,dx = 1$.
3. La médiane de X est le nombre M vérifiant $P(x \leq M) = P(x \geq M)$.
   Déterminer graphiquement une valeur approchée de M.
   Indication
   On a représenté la fonction f ainsi que la fonction F en annexe à rendre : préciser les traits de construction.

Exercice 2

Loi binomiale et loi normale
Un producteur de pommes a fait une étude statistique sur le diamètre de l'ensemble des pommes produites. Cette étude conduit au résultat suivant : le diamètre en centimètres d'une pomme est une variable aléatoire X suivant la loi normale de moyenne $\mu = 7$ et d'écart-type $\sigma = 0,5$: $X \sim \mathcal{N}(7;\, 0,5)$.
Par ailleurs, on dit qu'une pomme est de catégorie $\mathcal{C}$ si son diamètre est compris entre $6,3$ cm et $7,7$ cm.

1. Calculer $p = P(6,3 \leq X \leq 7,7)$, probabilité qu'une pomme prise au hasard dans la production soit de calibre $\mathcal{C}$.
2. Le producteur conditionne ses pommes par cageots de 36 pommes. On note Y la variable aléatoire égale au nombre de pommes de catégorie $\mathcal{C}$ dans un cageot. Y suit la loi binomiale de paramètres $n = 36$ et p : $Y \sim \mathcal{B}(n, p)$.

(a) Quel est le nombre moyen m de pommes de catégorie $\mathcal{C}$ dans un cageot ?
Indication : $m = E(Y)$.

(b) On souhaite calculer la probabilité d'avoir au moins 28 pommes dans un cageot $P(Y \geq 28)$. Pour simplifier le calcul on envisage d'utiliser une autre loi.

i. Est-ce qu'il est envisageable d'utiliser une loi de Poisson ?
Indication
L'approximation de la loi binomiale $\mathcal{B}(n,p)$ par la loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda = np$ est envisageable pour n assez grand ($n \geq 30$), $p \leq 0,1$ et $np \leq 10$.

ii. Est-ce qu'il est envisageable d'utiliser une loi de normale ?
Indication
D'après le théorème central limite, l'approximation de la loi binomiale $\mathcal{B}(n,p)$ par la loi normale $\mathcal{N}(m,\sigma)$ avec $m = np$ et $\sigma^2 = np(1-p)$ est envisageable pour n assez grand ($n \geq 30$), $np \geq 5$ et $n(1-p) \geq 5$.

(c) Pour calculer $P(Y \geq 28)$ on décide d'utiliser la loi normale de paramètres $m = np$ et $\sigma = \sqrt{np(1-p)}$. On considère alors une variable aléatoire $Y' \sim \mathcal{N}(m,\sigma)$.

i. Préciser m et σ.

ii. Calculer $P(Y \geq 28) \approx P(Y' \geq 27,5)$.

Exercice 3

Intervalle de confiance pour une moyenne ; pour un écart-type
À la réception de colis, un responsable doute de l'exactitude des masses affichées sur les boîtes. Il prélève au hasard $n = 25$ boîtes qu'il pèse. On note x_i la réalisation de la variable aléatoire X_i, masse de la boîte i en kg ($1 \leq i \leq 25$). Les variables X_i sont indépendantes selon une loi $\mathcal{N}(\mu,\sigma)$.
Il obtient $\displaystyle\sum_{i=1}^{25} x_i = 49,5$ kg et $\displaystyle\sum_{i=1}^{25} x_i^2 = 98,3$ kg^2.
Ayant choisi a priori un seuil de $\alpha = 5$ %, il s'agit de construire des intervalles de confiance à $1 - \alpha = 95$ % pour μ.

1. Préciser une estimation ponctuelle de μ, puis une estimation de σ^2.
Indication
Les estimateurs de la moyenne μ et de la variance σ^2 sont respectivement $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ et
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right), \text{ avec } n = 25.$$

2. *Intervalle de confiance I pour μ de niveau de confiance* 95 %

(a) Donner un intervalle de confiance I pour μ de niveau de confiance 95 %.
Indication
La variance n'étant pas connue, elle doit être estimée. On utilise l'estimation s^2 trouvée ci-dessus et $\dfrac{\overline{X} - \mu}{S/\sqrt{n}} \sim \mathcal{T}_{n-1}$, loi de Student à $n-1$ degrés de liberté ($n = 25$).
On peut alors écrire $P\left(\left|\dfrac{\overline{X} - \mu}{S/\sqrt{n}}\right| \leq t_{n-1;1-\alpha/2}\right) = 1 - \alpha.$

(b) Sachant que la masse affichée sur chaque boîte est de 2 kg, les doutes du responsable sont-ils justifiés au risque de 5 % ?
Indication
On précisera si $2 \in I$ et on conclura.

3. *Intervalle de confiance I' pour σ de niveau de confiance* 95 %
Donner un intervalle de confiance I' pour σ de niveau de confiance 95 %.
Indication
La moyenne μ n'étant pas connue, $\dfrac{n-1}{\sigma^2}S^2 \sim \chi^2_{n-1}$ $(n = 25)$.
On utilise alors $P\left(\chi_{n-1;\alpha/2} \leq (n-1)\dfrac{S^2}{\sigma^2} \leq \chi_{n-1;1-\alpha/2}\right) = 1 - \alpha$.

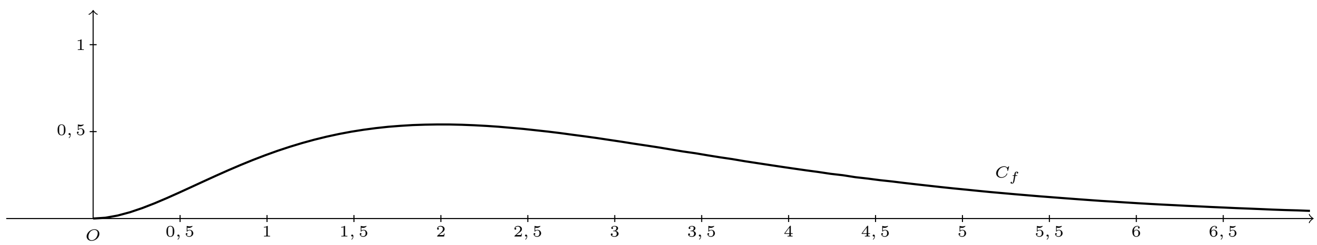
Exercice 4

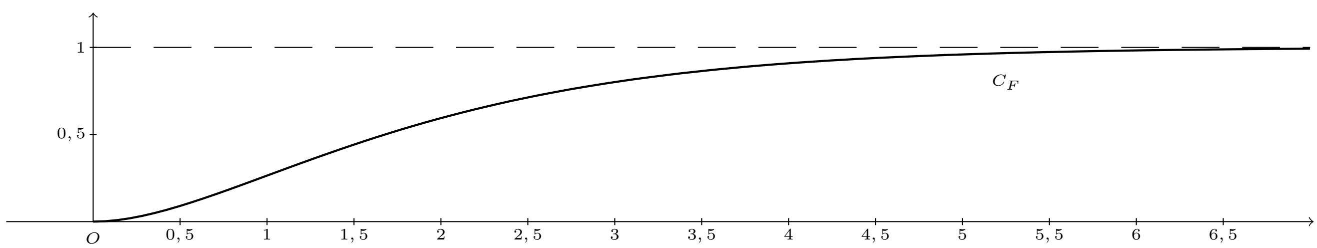
Test pour une moyenne (σ connu)
Dans une usine, un procédé de fabrication produit des équipements informatiques dont la *durée de vie* X en heures est une variable aléatoire qui suit la loi normale $\mathcal{N}(\mu_0,\, \sigma)$, avec $\mu_0 = 10\,000$ heures et $\sigma = 1\,200$ heures.
Un nouveau procédé est mis en place sur une autre chaîne de production. On admet que la *durée de vie* suit aussi une loi normale. Sur un échantillon de $n = 100$ équipements issus de cette nouvelle chaîne, la durée de vie moyenne est de $\overline{x} = 10\,300$ heures.
On se demande si on peut affirmer au niveau 5 % que ce nouveau procédé est différent de l'ancien en terme de durée de vie de l'équipement informatique. On veut donc tester l'hypothèse nulle $H_0 : \mu = \mu_0$, contre l'hypothèse alternative $H_1 : \mu \neq \mu_0$ au niveau 5 %.
En notant X_i la *durée de vie* d'un équipement informatique donné i, on peut supposer que $X_i \sim \mathcal{N}(\mu, \sigma)$ et les X_i indépendantes.
Sous l'hypothèse nulle $H_0 : \mu = \mu_0$, $T = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0;\, 1)$ avec $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.

1. Préciser le nombre u vérifiant $P(-u \leq T \leq u) = 0,95$ avec $T \sim \mathcal{N}(0, 1)$.
2. En déduire $P(|T| \geq u)$, puis la région de rejet.
3. Est-ce que $\overline{x}$ appartient à la région de rejet ?
4. En déduire la décision. Expliquer.
5. Préciser la p-value $P_c(\overline{x}) = P_{H_0}\left(\left|\overline{X} - \mu_0\right| \geq |\overline{x} - \mu_0|\right)$.
6. En déduire le degré de signification du test (test *significatif*, *très significatif*, *hautement significatif*).
Indication

Degré de signification	p-value	Notation
Test significatif	$0,01 < P_c \leq 0,05$	*
Test très significatif	$0,001 < P_c \leq 0,01$	**
Test hautement significatif	$P_c \leq 0,001$	***

Annexe 1 à rendre





Nom : _