MATH-131, *Probabilités et Statistique*

Examen blanc

Linda Mhalla

Instructions: Le temps alloué à l'examen est de 180 minutes. Seule une calculatrice non programmable est autorisée. Tout document est interdit.

Le nombre de points attribué à chaque exercice et chaque question est indiqué. Toute réponse doit être justifiée. Sauf mention contraire, tout résultat du formulaire ou du cours peut être utilisé sans preuve. En revanche, vous devez le citer de manière appropriée.

Aucune réponse rédigée sur cet énoncé ou le formulaire ne sera lue.

Prénom :

Nom de famille :

SCIPER:

Exercice	Points	Max
1		34
2		16
3		12
4		13
Total:		75

Question 1. **Applications directes du cours** .... *34 points*

(a) Soient $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ et [2]

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Expliquez pourquoi on divise par "$n-1$" et non par "n" dans l'expression de S^2.

(b) La table suivante définit-elle une fonction de masse valide pour une variable aléatoire discrète ? [2]

x_i	1	2	3	5	10	20
$f(x_i)$	0.1	0.1	0.2	0.2	0.3	0.1

(c) Soit $X \sim U(-2, 3)$. Quelle est la distribution de la variable aléatoire $Y = 2X$? [4]

(d) Si $X \sim \mathcal{N}(5, 10^2)$, calculez $\Pr(X \leq 3)$. [2]

(e) On lance cinq dés 100 fois de suite de manière indépendante. Quelle est la probabilité d'obtenir au moins une fois les cinq faces identiques ? [5]

(f) Soient $X_1, \ldots, X_{25} \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ avec μ et σ^2 inconnus et soit $x_1, \ldots, x_{25}$ une réalisation de cet échantillon. Sachant que l'on a $\bar{x} = 10$ et $s = 2$, dérivez (en indiquant toutes les étapes) un intervalle de confiance à 95% pour μ. [4]

(g) Soient A et B deux événements tels que $\Pr(B) = 0.2$, $\Pr(A \mid B) = 0.5$ et $\Pr(A \mid B^c) = 0.3$. Que vaut $\Pr(A)$? [3]

(h) Expliquez pourquoi on peut approximer une loi binomiale $\mathcal{B}(m, p)$ par une loi normale quand m est grand. [4]

(i) Soit $(X, Y)'$ un vecteur aléatoire de densité conjointe [6]

$$f_{X,Y}(x, y) = \begin{cases} 4xy, & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{sinon.} \end{cases}$$

Montrez que $f_{X,Y}$ est une densité. Les variables aléatoires X et Y sont-elles indépendantes ?

(j) On effectue un test d'adéquation du khi-deux basé sur $k = 8$ classes au niveau de significativité $\alpha = 10\%$. Sachant que nous n'avons pas besoin d'estimer de paramètres et que $t_{\text{obs}} = 14$, que pouvez-vous conclure ? [2]

Solution:

(a) Nous avons vu que

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

est un estimateur biaisé de σ^2 (1 point). En effet, $\mathrm{E}(\widehat{\sigma}^2)$ sous-estime légèrement σ^2. Diviser par $n-1$ au lieu de n permet d'obtenir un estimateur sans biais (1 point).

(b) On a, pour tout $i = 1, , \ldots, 6$, $0 \leq f(x_i) \leq 1$ (1 point). De plus, $\sum_{i=1}^{6} f(x_i) = 3 \times 0.1 + 2 \times 0.2 + 0.3 = 1$ (1 point). Ainsi, f est bien une fonction de masse valide pour une variable discrète.

(c) Soit $X \sim U(a, b)$ avec $a < b$. Par définition, la fonction de répartition est

$$F_X(x) = \left\{ \begin{array}{ll} 0, & x < a, \\ \int_a^x \frac{1}{b-a} \mathrm{d}u = \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x > b. \end{array} \right. \tag{1}$$

Par ailleurs, pour $y \in \mathbb{R}$,

$$F_Y(y) = \Pr(Y \leq y) = \Pr(2X \leq y) = \Pr\left(X \leq \frac{y}{2}\right) = F_X\left(\frac{y}{2}\right) \quad \textit{(2 points).}$$

Comme

$$\frac{\frac{y}{2} - a}{b - a} = \frac{y - 2a}{2b - 2a},$$

on obtient en utilisant (1)

$$F_Y(y) = \begin{cases} 0, & y < 2a, \\ \frac{y-2a}{2b-2a}, & 2a \le y \le 2b, \\ 1, & y > 2b, \end{cases} \quad \textit{(1 point)}$$

ce qui correspond à la fonction de répartition de $U(2a, 2b)$ *(1 point). Ainsi, conformément à l'intuition, si* $X \sim U(a, b)$, $2X \sim U(2a, 2b)$. *Donc si* $X \sim U(-2, 3)$, $2X \sim U(-4, 6)$.

(d) *Soit* $X \sim \mathcal{N}(5, 10^2)$. *On a*

$$\Pr(X \le 3) = \Pr\left(\frac{X-5}{10} \le \frac{3-5}{10}\right) = \Pr(Z \le -0.2) \quad \textit{(1 point)},$$

où $Z \sim \mathcal{N}(0, 1)$. *Ainsi, en utilisant la symétrie de la fonction de répartition de la loi normale centrée réduite* Φ *et en lisant dans la table, on obtient*

$$\Pr(X \le 3) = \Phi(-0.2) = 1 - \Phi(0.2) = 1 - 0.57926 = 0.42074 \quad \textit{(1 point)}.$$

(e) *Si on lance* 5 *dés, la probabilité d'obtenir les cinq faces identiques est*

$$p = \frac{\textit{nombre de cas favorables}}{\textit{nombre de cas possibles}} = \frac{6}{6^5} = \frac{1}{6^4} = \frac{1}{1296} \quad \textit{(2 points)}.$$

Soit X *le nombre de fois où on observe les cinq faces identiques lorsqu'on lance les cinq dés de manière indépendante* $m = 100$ *fois de suite. On a* $X \sim \mathcal{B}(m, p)$ *(1 point). Ainsi,*

$$\Pr(X \ge 1) = 1 - \Pr(X = 0) = 1 - (1-p)^m \approx 7.43\% \quad \textit{(2 points)}.$$

(f) *Comme* μ *et* σ *sont inconnus, on sait d'après le cours et le formulaire que*

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1},$$

où

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

En utilisant la symétrie de la loi de Student t, *on obtient donc*

$$\Pr\left(-t_{n-1,1-\alpha/2} \le \frac{\bar{X} - \mu}{S/\sqrt{n}} \le t_{n-1,1-\alpha/2}\right) = 1 - \alpha \quad \textit{(1 point)},$$

i.e.,

$$\Pr\left(-\bar{X} - t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n}} \le -\mu \le -\bar{X} + t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha,$$

i.e.,

$$\Pr\left(\bar{X} - t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

On en déduit qu'un intervalle au niveau de confiance $1 - \alpha$ *est*

$$\left[\bar{X} - t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n}}\right] \quad \textit{(1 point)}.$$

Ici, $\alpha = 5\%$ *et on lit dans la table que* $t_{n-1,1-\alpha/2} = t_{24,0.975} = 2.064$ *(1 point). La réalisation de cet IC à partir de nos données est donc*

$$\left[\bar{x} - t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}\right] = \left[10 - 2.064\frac{2}{\sqrt{25}}, 10 + 2.064\frac{2}{\sqrt{25}}\right] \approx [9.17, 10.83] \quad \textit{(1 point)}.$$

(g) *La formule des probabilités totales nous donne que*

$$\Pr(A) = \Pr(A \mid B)\Pr(B) + \Pr(A \mid B^c)\Pr(B^c) \quad \textit{(1 point).}$$

En utilisant le fait que $\Pr(B^c) = 1 - \Pr(B)$ *(1 point), on obtient donc*

$$\Pr(A) = 0.5 \times 0.2 + 0.3 \times 0.8 = 0.34 \quad \textit{(1 point).}$$

(h) *Soit* $X \sim \mathcal{B}(m,p)$*. On sait d'après le cours que l'on peut écrire*

$$X = \sum_{i=1}^{m} Y_i,$$

où $Y_1, \dots, Y_m \overset{\text{iid}}{\sim} \mathcal{B}(p)$ *(2 points). Le formulaire nous donne que* $\mathrm{E}(Y_1) = p$ *et* $\mathrm{Var}(Y_1) = p(1-p)$ *(1 point). Ainsi, si* m *est grand, le TCL nous donne que*

$$\sum_{i=1}^{m} Y_i \mathrel{\dot\sim} \mathcal{N}(mp, mp(1-p)), \quad \textit{i.e.,} \quad X \mathrel{\dot\sim} \mathcal{N}(mp, mp(1-p)) \quad \textit{(1 point).}$$

(i) *On a, pour tout* $x, y \in \mathbb{R}$, $f_{X,Y}(x,y) \geq 0$ *(1 point). De plus,*

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x = \int_0^1\int_0^1 4xy\mathrm{d}y\mathrm{d}x = 4\int_0^1\left(\int_0^1 xy\mathrm{d}y\right)\mathrm{d}x &= 4\int_0^1 x\left(\int_0^1 y\mathrm{d}y\right)\mathrm{d}x \\
&= 4\int_0^1 x\mathrm{d}x \int_0^1 y\mathrm{d}y \\
&= 4\left[\frac{x^2}{2}\right]_0^1\left[\frac{y^2}{2}\right]_0^1 \\
&= 1 \quad \textit{(2 points).}
\end{aligned}$$

Ainsi, $f_{X,Y}$ *est bien une densité.*

Maintenant, d'après le cours, la densité de X *s'écrit, pour tout* $x \in [0,1]$,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}y = \int_0^1 4xy\mathrm{d}y = 4x\int_0^1 y\mathrm{d}y = 4x\left[\frac{y^2}{2}\right]_0^1 = 2x \quad \textit{(1 point),}$$

et, pour tout $x \notin [0,1]$, $f_X(x) = 0$*. Par symétrie, la densité de* Y *est*

$$f_Y(y) = \begin{cases} 2y, & y \in [0,1], \\ 0, & y \notin [0,1]. \end{cases} \quad \textit{(1 point)}$$

Ainsi, pour tout $x, y \in \mathbb{R}$,

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \textit{(1 point),}$$

ce qui implique que les variables aléatoires X *et* Y *sont indépendantes.*

Notons que l'on pouvait directement remarquer que X *et* Y *étaient indépendantes étant donné que, pour tout* $x, y \in [0,1]$,

$$f_{X,Y}(x,y) = (2x) \times (2y).$$

(j) *Etant donné que l'on n'a pas eu besoin d'estimer de paramètres, la statistique de test du khi-deux,* T*, vérifie*

$$T \mathrel{\dot\sim} \chi^2_{k-1}, \quad \textit{i.e.,} \quad T \mathrel{\dot\sim} \chi^2_7 \quad \textit{(1 point).}$$

Au niveau de significativité $\alpha = 10\%$*, on rejette donc* H_0 *si et seulement si* $t_{\text{obs}} > \chi^2_{7,1-\alpha}$*, i.e.,* $t_{\text{obs}} > \chi^2_{7,0.9}$*. On a* $t_{\text{obs}} = 14$ *et* $\chi^2_{7,0.9} = 12.017$*. On rejette donc* H_0 *(1 point).*

Question 2. **Déséquilibre d'une pièce** .... *16 points*

On aimerait tester si une pièce de monnaie est déséquilibrée. Pour cela on lance la pièce 1000 fois. On a obtenu 550 fois "pile". Que pouvez-vous conclure ?

Solution:

Première possibilité :

Nous considérons le test statistique ayant pour hypothèses:
H_0*: "la pièce est équilibrée" contre* H_1*: "la pièce est déséquilibrée" (2 points).*

Afin de décider entre H_0 *et* H_1*, nous allons utiliser le nombre de "pile" obtenus après ces* 1000 *lancers. Soit* N *la variable aléatoire donnant le nombre de "pile" obtenus en* 1000 *lancers. Sous* H_0*, on a* $N \sim \mathcal{B}(m = 1000, p = 1/2)$ *(2 points).*

Nous cherchons une statistique de test qui prenne des valeurs faibles sous H_0 *et élevées sous* H_1*. Un choix naturel est donc d'utiliser comme statistique de test*

$$T = N - \frac{1}{2} \times 1000 = N - 500 \quad \textit{(4 points).}$$

Rappelons que N *peut s'écrire* $N = \sum_{i=1}^{m} Y_i$*, où* $Y_1, \ldots, Y_m \overset{\text{iid}}{\sim} \mathcal{B}(p)$*. Ainsi, après utilisation du formulaire, le TCL nous donne* $N \dot{\sim} \mathcal{N}(mp, mp(1-p))$ *(2 points). Notons que l'approximation devrait être bonne puisque* $m = 1000$ *est grand. Ainsi,* $N \dot{\sim} \mathcal{N}(500, 1000 \times 1/2 \times 1/2)$*, i.e.,* $N \dot{\sim} \mathcal{N}(500, 250)$*. On obtient donc*

$$T \dot{\sim} \mathcal{N}(0, 250) \quad \textit{(2 points).} \tag{2}$$

Nous choisissons comme niveau de significativité $\alpha = 5\%$ *(1 point). Nous rejetons* H_0 *si et seulement si* $|t_{\text{obs}}| > \nu^*$*, où* ν^* *vérifie* $\Pr(|T| > \nu^*) = \alpha$*, i.e.,*

$$\Pr\left(-\frac{\nu^*}{\sqrt{250}} \le \frac{T}{\sqrt{250}} \le \frac{\nu^*}{\sqrt{250}}\right) = \alpha.$$

Ainsi, en utilisant (2)*, on déduit que* $\nu^* = z_{1-\alpha/2} \times \sqrt{250} = z_{0.975} \times \sqrt{250} = 1.96 \times \sqrt{250} = 30.99$ *(2 points). Comme* $t_{\text{obs}} = 550 - 500 = 50$*, on rejette* H_0 *et on conclut que la pièce est déséquilibrée (1 point).*

Une autre approche pour prendre notre décision consiste à utiliser la p-valeur. On a

$$\begin{aligned}
p_{\text{obs}} = \Pr(|T| \ge t_{\text{obs}}) = \Pr(|T| \ge 50) &= 1 - \Pr(-50 \le T \le 50) \\
&= 1 - \Pr\left(-\frac{50}{\sqrt{250}} \le \frac{T}{\sqrt{250}} \le \frac{50}{\sqrt{250}}\right) \\
&\approx 1 - [\Phi(3.16) - \Phi(-3.16)] \\
&= 1 - [\Phi(3.16) - (1 - \Phi(3.16))] \\
&= 2\,(1 - \Phi(3.16)) \\
&\approx 1.6 \times 10^{-3}.
\end{aligned}$$

On a $p_{\text{obs}} < 0.05$ *donc on rejette* H_0*. Compte tenu de la faible valeur de* p_{obs}*, l'évidence contre* H_0 *est assez forte. Notons que pour utiliser l'approche p-valeur ici, vous avez besoin de la valeur de* $\Phi(3.16)$ *que vous ne pouvez pas trouver dans la table.*

Deuxième possibilité :

Soit X *le résultat du lancer de la pièce. Si la pièce est équilibrée, la loi de la variable* X *est donnée par*

$$\Pr(X = \textit{"pile"}) = \Pr(X = \textit{"face"}) = \frac{1}{2}, \quad \textit{(2 points).}$$

On peut donc tester si la pièce est déséquilibrée à l'aide d'un test d'adéquation du khi-deux avec $k = 2$ *classes : la classe des "pile" et celle des "face" (2 points). L'hypothèse nulle considérée est* H_0 *: "les observations proviennent de la loi théorique spécifiée" (1 point). Soient* o_1, o_2 *(réalisations de variables aléatoires notées* O_1, O_2*) les fréquences observées dans chacune des deux classes et soient*

E_1, E_2 les fréquences théoriques correspondantes sous H_0. La statistique de test T est la statistique du khi-deux (ou statistique de Pearson), donnée par

$$T = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \quad \textit{(2 points).}$$

Ici on a $o_1 = 550$ et $o_2 = 1000 - o_1 = 450$ (1 point). Par ailleurs,

$$E_1 = \frac{1}{2} \times 1000 = 500 \quad \textit{et} \quad E_2 = \frac{1}{2} \times 1000 = 500 \quad \textit{(1 point).}$$

On en déduit que $t_{\text{obs}} = 10$ (1 point). Etant donné que l'on n'a pas eu besoin d'estimer de paramètres pour calculer E_1 et E_2 (1 point), on sait que, sous H_0, $T \dot\sim \chi^2_{k-1}$, i.e., $T \dot\sim \chi^2_1$ (1 point). Nous choisissons comme niveau de significativité $\alpha = 5\%$ (1 point). Ainsi, on rejette H_0 si et seulement si $t_{\text{obs}} > \chi^2_{1,1-\alpha}$, i.e., $t_{\text{obs}} > \chi^2_{1,0.95}$, i.e., $t_{\text{obs}} > 3.841$ (2 points). Compte tenu de la valeur de H_0, on rejette donc H_0 et on conclut que la pièce est déséquilibrée (1 point).

Question 3. **Test médical** .... *12 points*

Un étudiant vivant en Suisse va chez le médecin et il est alors choisi au hasard pour un test de la grippe porcine. On sait que cette maladie concerne actuellement une personne sur 10000 en Suisse. La probabilité que le test soit positif alors que la personne testée n'a pas la grippe porcine est de 1%. Si la personne est malade, le test est toujours positif.

(a) Si l'étudiant est testé positif, quelle est la probabilité qu'il soit effectivement malade ? [10]

(b) On sait maintenant que l'étudiant est revenu du Mexique il y a quelques jours et que 1 personne sur 200 ayant visité le Mexique récemment a contracté cette maladie là-bas. La probabilité que l'étudiant ait la grippe porcine change-t-elle avec cette information ? Si oui, calculez la nouvelle probabilité. [2]

Solution:

(a) Nous considérons les événements (4 points)

$$\begin{aligned} M &= \textit{la personne est malade,} \\ S &= \textit{la personne est saine,} \\ T_M &= \textit{la personne est déclarée malade par le test,} \\ T_S &= \textit{la personne est déclarée saine par le test.} \end{aligned}$$

Nous recherchons la probabilité $\Pr(M|T_M)$. La formule de Bayes nous donne que

$$\Pr(M|T_M) = \frac{\Pr(T_M|M)\Pr(M)}{\Pr(T_M)} = \frac{\Pr(T_M|M)\Pr(M)}{\Pr(T_M|M)\Pr(M) + \Pr(T_M|S)\Pr(S)} \quad \textit{(3 points).} \tag{3}$$

Nous savons d'après l'énoncé que

$$\Pr(M) = 0.0001 \quad \textit{et donc} \quad \Pr(S) = 1 - \Pr(M) = 0.9999 \quad \textit{(1 point),} \tag{4}$$

et que

$$\Pr(T_M|M) = 1 \quad \textit{et} \quad \Pr(T_M|S) = 0.01 \quad \textit{(1 point).} \tag{5}$$

Ainsi, en combinant (3), (4) et (5), on obtient

$$\Pr(M|T_M) \approx 0.0099 \quad \textit{(1 point).}$$

(b) Parmi les grandeurs intervenant dans (3), les seules modifications par rapport à la question précédente concernent $\Pr(M)$ et $\Pr(S)$ (1 point). On a désormais $\Pr(M) \approx 1/200 = 0.005$ et donc $\Pr(S) = 1 - \Pr(M) \approx 0.995$. En utilisant (3), on obtient alors

$$\Pr(M|T_M) \approx 0.3344 \quad \textit{(1 point).}$$

Question 4. **Loi géométrique et maximum de vraisemblance** .... *13 points*

Une variable aléatoire X suit la loi géométrique de paramètre $p \in (0,1)$ si

$$\Pr(X = i) = p(1-p)^{i-1}, \quad i \in \mathbb{N}\backslash\{0\}.$$

Soit $X_1, \ldots, X_n$ un échantillon tiré de la loi de X. Nous cherchons à estimer le paramètre p.

(a) On considère une épreuve de Bernoulli dont la probabilité de succès est p. On renouvelle cette épreuve de manière indépendante jusqu'au premier succès. On appelle X la variable aléatoire donnant le rang du premier succès. Expliquez pourquoi X suit une loi géométrique. [3]

(b) Dérivez l'expression de $\widehat{p}_{\mathrm{ML}}$, l'estimateur du maximum de vraisemblance de p. [7]

(c) Dérivez la loi approximative de $\widehat{p}_{\mathrm{ML}}$ quand n est grand. [3]

Solution:

(a) La probabilité d'un succès est p et celle d'un échec $1-p$ (1 point). Pour $i \in \mathbb{N}\backslash\{0\}$, on a $X = i$ si et seulement si le rang du premier succès est i. Cela équivaut au fait qu'il y ait eu exactement $i-1$ échecs lors des $i-1$ premières épreuves (1 point). Ainsi, du fait de l'indépendance entre les différentes épreuves (1 point), on a

$$\Pr(X = i) = p(1-p)^{i-1}, \quad i \in \mathbb{N}\backslash\{0\}.$$

(b) Soit $x_1, \ldots, x_n$ une réalisation de l'échantillon $X_1, \ldots, X_n$. La vraisemblance pour le paramètre p est donnée par

$$L(p) = \prod_{i=1}^{n} \Pr(X = x_i) = \prod_{i=1}^{n} p(1-p)^{x_i - 1} = p^n (1-p)^{\sum_{i=1}^{n}(x_i - 1)} = p^n (1-p)^{\left(\sum_{i=1}^{n} x_i\right) - n} \quad \textit{(2 points).}$$

La log-vraisemblance s'écrit donc

$$\ell(p) = \log(L(p)) = n\log(p) + \log([1-p]) \left[\left(\sum_{i=1}^{n} x_i\right) - n\right] \quad \textit{(1 point).}$$

Ainsi,

$$\ell'(p) = \frac{n}{p} - \frac{\left[\left(\sum_{i=1}^{n} x_i\right) - n\right]}{1-p} \quad \textit{(1 point).}$$

ce qui donne

$$\ell'(p) = 0 \Leftrightarrow n(1-p) = p\left[\left(\sum_{i=1}^{n} x_i\right) - n\right] \Leftrightarrow p = \frac{n}{\sum_{i=1}^{n} x_i} \quad \textit{(1 point).} \tag{6}$$

Par ailleurs, étant donné que

$$\ell''(p) = -\frac{n}{p^2} - \frac{\left[\left(\sum_{i=1}^{n} x_i\right) - n\right]}{(1-p)^2},$$

$\sum_{i=1}^{n} x_i \geq n$ et $p \in (0,1)$, on a que $\ell''(p) < 0$ pour tout $n \in \mathbb{N}\backslash\{0\}$ (2 points). Ainsi, le p obtenu en (6) correspond bien à un maximum et donc la réalisation de l'estimateur du maximum de vraisemblance de p (à partir de l'observation de l'échantillon $X_1, \ldots, X_n$) est donnée par

$$\widehat{p}_{\mathrm{ML}} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}.$$

L'estimateur du maximum de vraisemblance est donc

$$\widehat{p}_{\mathrm{ML}} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}.$$

(c) Nous savons d'après le cours/formulaire que, pour n grand, $\widehat{p}_{\mathrm{ML}} \dot\sim \mathcal{N}\left(p, J(\widehat{p}_{\mathrm{ML}})^{-1}\right)$ (1 point), où

$$J(p) = -\ell''(p) = \frac{n}{p^2} + \frac{\left[\left(\sum_{i=1}^{n} x_i\right) - n\right]}{(1-p)^2}.$$

Finalement, on obtient que $\widehat{\lambda}_{\mathrm{ML}} \dot\sim \mathcal{N}\left(p, \left[\frac{n}{\widehat{p}_{\mathrm{ML}}^2} + \frac{\left[\left(\sum_{i=1}^{n} x_i\right) - n\right]}{(1-\widehat{p}_{\mathrm{ML}})^2}\right]^{-1}\right)$ (2 points).