STT-20694: Probabilités pour ingénieurs

SOLUTION DE L'EXAMEN NUMÉRO 2

Lundi le 15 décembre 2003

NOM DE FAMILLE: ....

PRÉNOM: ....

MATRICULE: ....

SIGNATURE: ....

- Cet examen comprends 20 questions.
- Chaque question vaut 2.5 points.
- Numéros 1 à 10: ces numéros sont des questions à choix multiples. Aucune justification n'est requise. **Pour chaque numéro il y a un seul bon choix. Pour chacun de ces numéros, vous recevrez ou bien 0 points, ou bien 2.5 points**.
- Numéros 11 à 20: **pour avoir 2.5 points il faut écrire la bonne réponse dans la case réservée à cette fin et il faut présenter vos raisonnements et vos calculs dans l'espace réservé à cette fin**. Il est possible d'obtenir une partie des points pour une solution partielle contenant suffisamment de bons éléments.
- Les notes finales seront affichées sur le site web du cours d'ici quelques jours.

Partie	Valeur	Vos points
I	25	
II	25	
Total	50	

C. Bélisle
Décembre 2003

NUMÉRO 1. Les variables aléatoires X et Y sont des exponentielles indépendantes l'une de l'autre, toutes les deux avec des moyennes égales à 1/2. Calculez la variance de $X+Y$.

- ☐ 1/4
- ■ 1/2
- ☐ 1
- ☐ 2
- ☐ 4
- ☐ 8

Explication: les variables X et Y sont des exponentielles indépendantes l'une de l'autre, avec paramètre $\lambda = 2$. On obtient donc:

$\mathbb{V}\mathrm{ar}[X+Y] = \mathbb{V}\mathrm{ar}[X] + \mathbb{V}\mathrm{ar}[Y] = 1/4 + 1/4 = 1/2.$

NUMÉRO 2. Entendu à la radio:

Un sondage effectué par l'Agence CROP révèle que 45% des Canadiens agés entre 40 et 60 ans ont avoué avoir fumé de la marijuana dans les années 70. L'erreur est au plus 2%, 19 fois sur 20.

Quelle était la taille de l'échantillon utilisé dans ce sondage?

- ☐ Environ 50
- ☐ Environ 100
- ☐ Environ 200
- ■ Environ 2 400
- ☐ Environ 4 800
- ☐ Environ 9 600

Explication:
On considère l'équation

$0.02 = (1.96)/2\sqrt{n}$

et on résoud pour n.
On obtient $n = 2\,401$

NUMÉRO 3. Un statisticien veut estimer le paramètre θ de la loi de Poisson suivante:

$$p_\theta(k) = \begin{cases} \dfrac{e^{-\theta}\theta^k}{k!} & \text{si } k \in \{0,1,2,3,...\} \\ 0 & \text{si } k \notin \{0,1,2,3,...\} \end{cases}$$

Il obtient les quatre observations suivantes: 4, 2, 7 et 5. Puisque l'estimateur de θ obtenu à partir de la méthode du maximum de la vraisemblance est simplement la moyenne échantillonnale, le statisticien obtient l'estimation suivante:

$$\hat{\theta} = \frac{4+2+7+5}{4} = \frac{18}{4} = 4.50$$

Lequel des énoncés suivants est le plus précis?

- ☐ Lorsqu'on observe les données $4, 2, 7$ et 5, c'est la valeur 4.50 qui est la valeur la plus probable du paramètre θ.
- ☐ Les valeurs $4, 2, 7$ et 5 sont les valeurs les plus probables de la distribution de Poisson avec paramètre $\theta = 4.5$.
- ■ Parmi toutes les lois de Poisson, c'est avec la loi de Poisson avec $\theta = 4.5$ qu'on a la plus grande probabilité d'observer les valeurs $4, 2, 7$ et 5 lorsqu'on obtient un échantillon aléatoire de taille 4.
- ☐ Nous n'avons pas assez d'observations pour estimer le paramètre θ.
- ☐ Aucun des quatre premiers énoncés n'est valide.
- ☐ Les quatre premiers énoncés sont tous valides.

NUMÉRO 4. La loi normale $N(\mu, \sigma^2)$ est un bon modèle pour décrire la distribution des poids des rats de laboratoire à la naissance. On a mesuré les poids de 49 rats de laboratoire à leurs naissances. La moyenne de ces 49 poids était 85.7 grammes. L'écart-type de ces 49 poids était 7.70 grammes. Notre estimation de la moyenne théorique μ est la moyenne échantillonnale 85.7 grammes. Quelle est l'erreur-type associée à cette estimation?

- ☐ 0.157 gramme.
- ■ 1.100 grammes.
- ☐ 1.210 grammes.
- ☐ 7.700 grammes.
- ☐ 8.470 grammes.
- ☐ 11.130 grammes.

Explication:

erreur type $= \frac{s}{\sqrt{n}} = \frac{7.70}{\sqrt{49}} = \frac{7.70}{7} = 1.10$

NUMÉRO 5. Les durées de vie de 225 guizmos ont été mesurées. La moyenne de ces 225 durées de vie est de 2.37 années et l'écart-type est de 1.62 années. Parmi les modèles suivants, quel est celui qui vous semble le plus approprié?

- ☐ La loi exponentielle. [C'est une loi avec moyenne = écart-type!]
- ■ La loi gamma.
- ☐ La loi géométrique. [C'est une loi discrète!]
- ☐ La loi normale. [Si $T \sim N(2.37, (1.62)^2$ alors $\mathbb{P}[T < 0] \approx 0.0721$]
- ☐ La loi de Poisson. [C'est une loi discrète!]
- ☐ La loi uniforme. [Ce n'est pas réaliste!]

NUMÉRO 6. On a testé 400 items choisis au hasard à partir d'une certaine population et on a déterminé que 38 d'entre eux étaient défectueux. On estime donc à 9.5 le pourcentage d'items défectueux dans cette population. Quelle est l'erreur type associée à cette estimation?

- ☐ Environ 1.0%
- ■ Environ 1.5%
- ☐ Environ 2.0%
- ☐ Environ 2.5%
- ☐ Environ 3.0%
- ☐ Environ 3.5%

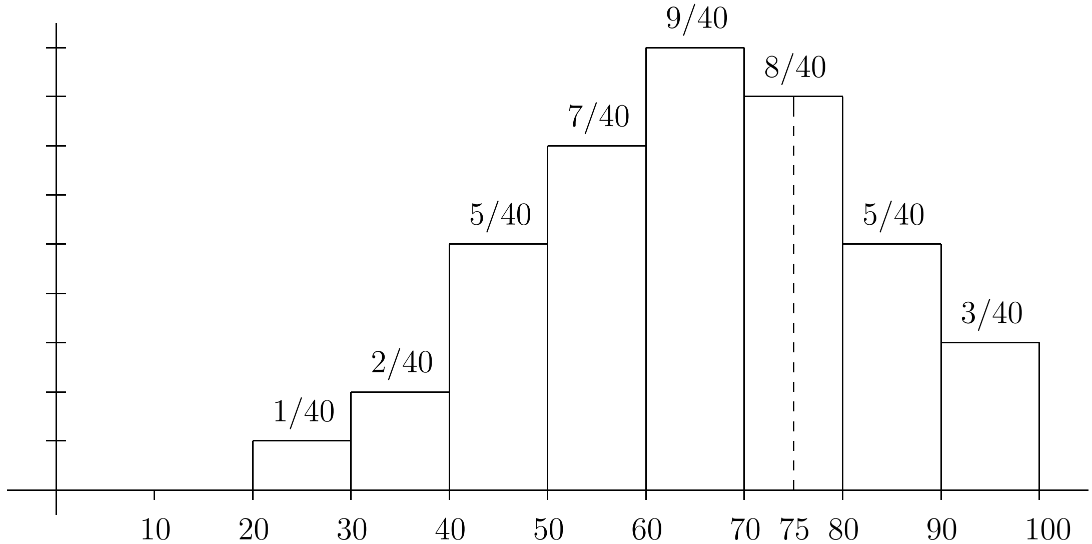
Explication:
On utilise la formule

$$\text{erreur type } = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

avec $n = 400$ et $\hat{p} = 0.095$
et on obtient

$$\text{erreur type } \approx 0.015 = 1.5\%.$$

Information pour les numéros 7 et 8. L'histogramme suivant représente la distribution des notes finales dans le cours STT-10400 à l'hiver 1999.



Numéro 7. Pour obtenir la cote B+ ou mieux, il fallait avoir une note finale de 75 ou plus. Quel pourcentage d'étudiants ont obtenu une cote B+ ou mieux?

- ☐ Environ 10%
- ☐ Environ 15%
- ☐ Environ 20%
- ☐ Environ 25%
- ■ Environ 30%
- ☐ Environ 35%

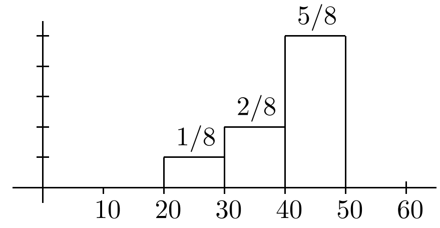
Explication:
Sur l'histogramme, j'ai indiqué les proportions (c'est-à-dire les surfaces) associées à chacune des boîtes. On veut la surface à droite de 75. On obtient

$$\frac{8/40}{2} + 5/40 + 3/40 = \frac{12}{40} = \frac{30}{100}$$

Numéro 8. La note de passage était 50. D'après cet histogramme, quelle était la note finale moyenne des étudiants qui ont échoué le cours?

- ☐ Environ 35
- ■ Environ 40
- ☐ Environ 42
- ☐ Environ 44
- ☐ Environ 46
- ☐ Aucune de ces réponses.

Explication:
Si on se restreint aux étudiants qui ont échoué le cours, on obteint l'histogramme suivant. Sa moyenne est 40.



Numéro 9. L'écart inter quartile de la loi t de Student avec 10 degrés de liberté est

- ■ 7/5
- □ 5/4
- □ 2
- □ 2.228
- □ 4.456
- □ 9

Explication:
La table de la loi t de Student nous donne dit que la surface à droite de 0.700 est 0.25. Par symétrie, la surface à gauche de -0.700 est également 0.25. Donc le premier quartile est -0.700 et le troisième quartile est 0.700. Lécart inter quartile est donc

$0.700 - (-0.700) = 1.4 = \frac{14}{10} = \frac{7}{5}$.

Numéro 10. L'étudiant Joe Smith est en train de faire cet examen. Pour chacun des numéros 1 à 10, il lance un dé bien balancé et il choisit sa réponse selon le résultat du lancer du dé. Les numéros 1 à 10 valent chacun 2.5 points. Si T dénote le total des points que Joe Smith obtiendra pour les numéros 1 à 10, alors l'écart-type de T est environ

- □ 1.0
- □ 1.5
- □ 2.0
- □ 2.5
- ■ 3.0
- □ Aucune de ces réponses.

Explication:
Si N dénote le nombre de bonne réponses que Joe Smith obtiendra, alors on a $T = (2.5)N$ et $N \sim$ binomiale(10,1/6). On obtient donc

$\sigma_T = (2.5) \times \sigma_N = (2.5) \times \sqrt{10(1/6)(5/6)} \approx 3.0$

NUMÉRO 11. La distribution suivante est un bon modèle pour une certaine population:

$$f(x) = \begin{cases} \theta(\theta+1)\ x(1-x)^{\theta-1} & \text{si } 0 \le x \le 1 \\ 0 & \text{sinon.} \end{cases}$$

Sa moyenne est $2/(2+\theta)$ et sa variance est $2\theta/\{(2+\theta)^2(3+\theta)\}$. On obtient un échantillon aléatoire de taille 4 à partir de cette population. Voici les 4 valeurs observées: 0.57, 0.22, 0.45, 0.36. On estime le paramètre θ à l'aide de la méthode des moments. Quelle estimation obtient-on?

Réponse: $\hat{\theta} \;=\; 3.00$

JUSTIFICATION:

On doit résoudre l'équation

$$\text{moyenne échantillonnale} \;=\; \text{moyenne théorique},$$

c'est-à-dire l'équation

$$\overline{x} = \frac{2}{2+\theta}.$$

Lorsqu'on résoud pour θ, on obtient

$$\hat{\theta} = \frac{2(1-\overline{x})}{\overline{x}}.$$

La moyenne échantillonnale est donnée par

$$\overline{x} = \frac{0.57 + 0.22 + 0.45 + 0.36}{4} = \frac{1.60}{4} = 0.40.$$

On obtient donc l'estimation

$$\hat{\theta} = \;=\; = 3.00.$$

Numéro 12. La distribution suivante est un bon modèle pour une certaine population:

$$f(x) = \begin{cases} \dfrac{\theta\ x^{\theta-1}}{3^\theta} & \text{si } 0 \le x \le 3 \\ 0 & \text{sinon.} \end{cases}$$

Sa moyenne est $3\theta/(\theta+1)$ et sa variance est $9\theta/\{(\theta+1)^2(\theta+2)\}$. On obtient un échantillon aléatoire de taille 4 à partir de cette population. Voici les 4 valeurs observées: 1.84, 2.94, 1.27 et 2.38. On estime le paramètre θ à l'aide de la méthode du maximum de la vraisemblance. Quelle estimation obtient-on?

Réponse: $\hat{\theta} \ = \ 2.50$

Justification:

La fonction de vraisemblance est

$$L(\theta) = \frac{\theta\, x_1^{\theta-1}}{3^\theta}\ \frac{\theta\, x_2^{\theta-1}}{3^\theta}\ \frac{\theta\, x_3^{\theta-1}}{3^\theta}\ \frac{\theta\, x_4^{\theta-1}}{3^\theta} = \frac{\theta^4 (\prod_{i=1}^4 x_i)^{\theta-1}}{3^{4\theta}}.$$

La fonction de log-vraisemblance est

$$\mathcal{L}(\theta) = 4\ \ln(\theta) - 4\theta\ \ln(3) + (\theta-1)\sum_{i=1}^4 \ln(x_i).$$

La dérivée de la fonction de log-vraisemblance est

$$\frac{d}{d\theta}\mathcal{L}(\theta) = \frac{4}{\theta} - 4\ \ln(3) + \sum_{i=1}^4 \ln(x_i).$$

On pose ça égal à 0 et on résoud pour θ. On obtient la solution suivante:

$$\hat{\theta} \quad = \quad \frac{4}{4\ \ln(3) - \displaystyle\sum_{i=1}^4 \ln(x_i)} \quad = \quad \frac{1}{\dfrac{1}{4}\displaystyle\sum_{i=1}^4 \ln(3/x_i)}$$

Avec $(x_1, x_2, x_3, x_4) = (1.84, 2.94, 1.27, 2.38)$, on obtient $\hat{\theta} = 2.50$.

INFORMATION POUR LES NUMÉROS 13 ET 14. Treize piles produites par la machine A ont été testées. La tension moyenne de ces 13 piles est 12.75 volts et l'écart-type est 0.80 volts. Dix-sept piles produites par la machine B ont été testées. La tension moyenne de ces 17 piles est 11.45 volts et l'écart-type est 0.60 volts. Les histogrammes suggèrent qu'il est raisonnable de supposer des lois normales de même variance pour décrire les distributions de tensions des piles produites par chacune de ces machines.

NUMÉRO 13. Si μ_A dénote la tension moyenne théorique des piles produites par la machine A et μ_B dénote la tension moyenne théorique des piles produites par la machine B, alors notre estimation de $\mu_A - \mu_B$ est $12.75 - 11.45 = 1.30$ volts. Quelle est l'erreur type associée à cette estimation?

Réponse:	0.255 volt

JUSTIFICATION:

$$n_A = 13 \qquad \overline{x}_A = 12.75 \qquad s_A = 0.80$$

$$n_B = 17 \qquad \overline{x}_A = 11.45 \qquad s_A = 0.60$$

$$\begin{aligned}
\text{erreur type} &= s_c \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \\
&= \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}} \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \\
&= \sqrt{\frac{(13-1)(0.80)^2 + (17-1)(0.60)^2}{13 + 17 - 2}} \sqrt{\frac{1}{13} + \frac{1}{17}} \\
&= \sqrt{\frac{12\ (0.80)^2 + 16(0.60)^2}{28}} \sqrt{\frac{1}{13} + \frac{1}{17}} \\
&= \sqrt{0.48000}\ \sqrt{0.13575} \\
&= 0.69282 \times 0.36844 \\
&= 0.255
\end{aligned}$$

NUMÉRO 14. [Suite du numéro 13]. Vous notez que le rapport des variances échantillonnales est

$$\frac{s_A^2}{s_B^2} = \frac{(0.80)^2}{(0.60)^2} = \frac{0.64}{0.36} \approx 1.778$$

et cela vous inquiète puisque nous avons supposé que les variances théoriques étaient égales. Pour vous rassurer, vous faites le raisonnement suivant. Lorsqu'on obtient des échantillons aléatoires indépendants, de tailles 13 et 17, à partir de lois normales de même variance, on s'attend à ce que le rapport des variances échantillonnales (celle calculée à partir de l'échantillon de taille 13 divisée par celle calculée à partir de l'échantillon de taille 17) soit environ ___1.143___ , plus ou moins environ ___0.687___ .

Complétez la phrase précédente.

JUSTIFICATION:

Avec $\sigma_A^2 = \sigma_B^2$, on obtient $\dfrac{S_A^2}{S_B^2} \sim F(n_A - 1, n_B - 2) = F(12, 16)$.

On utilise la formule

$$\text{espérance de la loi } F(k,m) \; = \frac{m}{m-2}$$

et on obtient

$$\text{espérance de } \frac{S_A^2}{S_B^2} = \frac{16}{16-2} = \frac{16}{14} = \frac{8}{7} = 1.143.$$

On utilise la formule

$$\text{écart-type de la loi } F(k,m) \; = \sqrt{\frac{2m^2(k+m-2)}{k(m-2)^2(m-4)}}$$

et on obtient

$$\text{écart-type de } \frac{S_A^2}{S_B^2} = \sqrt{\frac{2 \times 16^2 \times (12+16-2)}{12 \times (16-2)^2 \times (16-4)}} = \sqrt{\frac{13\,312}{28\,224}} = \sqrt{0.47166} = 0.687.$$

INFORMATION POUR LES NUMÉROS 15, 16 ET 17. Les clients arrivent à un centre de service comme un processus de Poisson avec une intensité de 12 clients par heure. Cinq pourcent des clients sont détenteurs d'une carte de crédit Ovation. Autrement dit, on suppose que chaque fois qu'un client arrive on a une chance sur 20 qu'il soit détenteur d'une carte Ovation, indépendamment des autres clients.

NUMÉRO 15. Calculez la probabilité qu'exactement 4 clients arriveront durant les prochaines 15 minutes.

Réponse:	0.168

JUSTIFICATION:

Si on pose

$$N = \text{ le nombre de clients qui arriveront durant les 15 prochaines minutes,}$$

alors on a $N \sim$ Poisson(3). On obtient donc

$$\mathbb{P}[N = 4] = \mathrm{e}^{-3}\,\frac{3^4}{4!} = 0.1680.$$

NUMÉRO 16. Calculez l'écart-type du nombre de clients qui arriveront durant les 3 prochaines heures.

Réponse:	1.732 clients

JUSTIFICATION:

Si on pose

$$N = \text{ le nombre de clients qui arriveront durant les 3 prochaines heures,}$$

alors on a $N \sim$ Poisson(36). On obtient donc

$$\sigma_N = \sqrt{\mathbb{V}\mathrm{ar}[N]} = \sqrt{36} = 6.$$

NUMÉRO 17. Un client détenteur d'une carte Ovation vient d'arriver. Obtenez l'espérance du temps qui s'écoulera avant qu'un autre détenteur de carte Ovation arrive.

Réponse:	100 minutes

JUSTIFICATION:

Chaque fois qu'un client arrive, on a 1 chance sur 20 que ce soit un détenteur de carte Ovation. Si on pose

$$N = \text{ le nombre d'arrivées nécessaires pour avoir un client avec carte Ovation,}$$

alors on a $N \sim$ géométrique(1/20), et donc $\mathbb{E}[N] = 20$. Le temps moyen entre 2 arrivées successives de clients est 5 minutes. La réponse est donc

$$20 \times 5 \text{ minutes } = 100 \text{ minutes.}$$

NUMÉRO 18. On a estimé la variance d'une loi normale. Notre estimation est la variance échantillonnale $s^2 = 12.30$. L'erreur type associée à cette estimation est 4.65. Quel était la taille de l'échantillon?

Réponse:	15

JUSTIFICATION:

On utilise l'équation

$$\text{erreur type } = \sqrt{\frac{2}{n-1}}\ s^2.$$

On a donc

$$4.65 = \sqrt{\frac{2}{n-1}}\ (12.30).$$

On résoud pour n et on obtient

$$n = 2\ \left(\frac{12.30}{4.65}\right)^2\ +\ 1 = 15.$$

NUMÉRO 19. On lance une paire de dés 600 fois et on pose $X =$ le nombre de fois que le total des deux dés est égal à 7. Calculez une approximation pour $\mathbb{P}[99 \leq X \leq 105]$.

Réponse:	0.2893

JUSTIFICATION:

$$\begin{aligned}
\mathbb{P}[99 \leq X \leq 105] &\approx \Phi\left(\frac{105.5 - 100}{\sqrt{500/6}}\right) - \Phi\left(\frac{98.5 - 100}{\sqrt{500/6}}\right) \\
&\approx \Phi(0.6025) - \Phi(-0.1643) \\
&\approx \Phi(0.60) - \Phi(-0.16) \\
&\approx 0.7257 - 0.4364 \\
&= 0.2893
\end{aligned}$$

NUMÉRO 20. On a estimé l'écart-type d'une loi normale à partir d'un échantillon aléatoire de taille 81. Notre estimation est l'écart-type échantillonnal $s = 7.30$. Obtenez un intervalle de confiance de niveau 90% pour l'écart-type théorique.

Réponse:	$[6.47, 8.40]$

JUSTIFICATION:

$$\begin{aligned}
\left[\sqrt{\frac{n-1}{\chi^2_{\alpha/2,n-1}}}\ s\,,\ \sqrt{\frac{n-1}{\chi^2_{1-\alpha/2,n-1}}}\ s\right] &= \left[\sqrt{\frac{81-1}{\chi^2_{0.10/2,81-1}}}\ (7.30)\,,\ \sqrt{\frac{81-1}{\chi^2_{1-0.10/2,81-1}}}\ (7.30)\right] \\
&= \left[\sqrt{\frac{80}{\chi^2_{0.05,80}}}\ (7.30)\,,\ \sqrt{\frac{80}{\chi^2_{0.95,80}}}\ (7.30)\right] \\
&= \left[\sqrt{\frac{80}{\chi^2_{0.05,80}}}\ (7.30)\,,\ \sqrt{\frac{80}{\chi^2_{0.95,80}}}\ (7.30)\right] \\
&= \left[\sqrt{\frac{80}{101.88}}\ (7.30)\,,\ \sqrt{\frac{80}{60.39}}\ (7.30)\right] \\
&= [(0.8861) \times (7.30), (1.1510) \times (7.30)] \\
&= [6.47, 8.40]
\end{aligned}$$