Université de Cergy-Pontoise **2007-2008**
Mathématiques **L2 PCS4**

Corrigé de l'Examen du 5 Mai 2008

Exercice 1 : $f(x) = \mathbb{1}_{[0,1]}(x)$.

1. Pour avoir une loi de probabilité, il faut vérifier que $\int_{\mathbb{R}} f(t)\,dt = 1$, c'est à dire $\int_0^1 6t(t-1)\,dt = 1$. On a : $\int_0^1 6t(t-1)\,dt = [3x^2 - 2x^3]_0^1 = 1$.
2. La fonction de répartition est définie par $F(x) = \int_0^x f(t)\,dt$. Si $x \leq 0$, on a $F(x) = 0$. Pour $0 < x < 1$, on a $F(x) = \int_0^x 6t(t-1)\,dt = 3x^2 - 2x^3$ et si $x \geq 1$, on a $F(x) = 1$.
3. $P(X \leq \frac{1}{2}) = F(\frac{1}{2}) = \frac{1}{2}$, $P(\frac{1}{3} \leq X \leq \frac{2}{3}) = F(\frac{2}{3}) - F(\frac{1}{3}) = \frac{13}{27}$ et $P(X \geq \frac{1}{6}) = 1 - F(\frac{1}{6}) = \frac{25}{27}$.
4. $E(X) = \int_0^{+\infty} tf(t)\,dt = \int_0^1 6t^2(t-1)\,dt = [2x^3 - \frac{3}{2}x^3]_0^1 = \frac{1}{2}$.
   $V(X) = E(X^2) - (E(X))^2$. Or $E(X^2) = \int_0^{+\infty} t^2 f(t)\,dt = \int_0^1 6t^3(t-1)\,dt = [\frac{3}{2}x^3 - \frac{6}{5}x^3]_0^1 = \frac{3}{10}$. Finalement, on obtient $V(X) = \frac{3}{10} - \frac{1}{2} = \frac{1}{20}$.

Exercice 2 : Soit X est la variable aléatoire qui compte le nombre d'abonnés connectés à un instant donné.

1. X suit une loi binomiale $\mathcal{B}(n,p)$ de paramètre $n = 5000$ et $p = 0,2$. On a $E(X) = np = 1000$, $V(X) = np(1-p) = 800$ et $\sigma(X) = \sqrt{V(X)} \simeq 28,28$.
2. Puisque $n > 30$, $np > 15$ et $np(1-p) > 15$, on peut appliquer le théorème de la limite centrale et on sait alors que $\mathcal{N}(np, np(1-p))$ approxime $\mathcal{B}(n,p)$. Ceci implique que $Y = \frac{X-1000}{\sqrt{800}}$ suit une loi normale centrée réduite $\mathcal{N}(0,1)$.
3. Si N est le nombre d'abonnés connectés à partir duquel il y a saturation, il faut que l'on ait $P(X \geq N) \leq 0,025$. On a donc la condition $1 - P(X < N) \leq 0,025$, c'est à dire $P(X < N) \geq 0,975$. Puisqu'on approxime une loi discrète par une loi continue, il faut faire une correction de continuité. On cherche donc N tel $\phi(\frac{N-0,5-1000}{\sqrt{800}}) \geq 0,975$ où ϕ désigne la fonction de répartition de la loi normale centrée réduite. Or $0,975 = \phi(1,96)$. La condition sur N est $\frac{N-0,5-1000}{\sqrt{800}} \geq 1,96$. On obtient $N \geq 1056,38$. Il faut donc pouvoir gérer simultanément 1057 connections simultanément.

Exercice 3 : 1. La moyenne empirique est donnée par

$$m = \frac{36 \times 2,5 + 22 \times 7,5 + 15 \times 12,5 + 9 \times 17,5 + 6 \times 22,5 + 5 \times 27,5 + 7 \times 32,5}{100} = 11$$

La variance empirique qui est un estimateur avec biais est donnée par :

$$S^2 = \frac{36 \times 2,5^2 + 22 \times 7,5^2 + 15 \times 12,5^2 + 9 \times 17,5^2 + 6 \times 22,5^2 + 5 \times 27,5^2 + 7 \times 32,5^2}{100} - m^2 = 86,75$$

Pour obtenir une estimation ponctuelle de la variance sans biais, on calcule $\tilde{S}^2 = \frac{n}{n-1}S^2 = \frac{100}{99}86,75 = 87,62$.

2. Puique l'échantillon est de taille supérieure à 30, $\bar{X}$ suit une loi normale $\mathcal{N}(\mu, \frac{\tilde{S}^2}{n})$ où μ est la moyenne de la durée de vie des composants. Comme l'écart-type est inconnu, on utilise $\frac{\tilde{S}^2}{n}$ pour la variance. On en déduit alors que $\frac{\bar{X}-\mu}{\frac{\tilde{S}}{\sqrt{n}}}$ suit une loi normale centrée réduite.
3. Pour déterminer un intervalle de confiance au risque 5%, on cherche t tel que

$$P(|\frac{\bar{X}-\mu}{\frac{\tilde{S}}{\sqrt{n}}}| \leq t) = 1 - 0,05 = 0,95$$

Si ϕ désigne la fonction de répartition de la loi normale centrée réduite, on cherche dans les tables t tel que $2\phi(t) - 1 = 0,95$, c'est à dire $\phi(t) = 0,975$. On obtient $t = 1,96$. De plus $\tilde{S} = 9,36$. On obtient alors $m - 1,96 \times \frac{9,36}{10} \leq \mu \leq m + 1,96 \times \frac{9,36}{10}$. L'intervalle de confiance au risque 5% est donné par $[9,17; 12,83]$

Exercice 4 : 1. Une estimation de la proportion de la population qui considère la télévision comme source principale d'information est donnée par $f = \frac{110}{200}$ à l'aide de l'échantillon.

2. Puisque l'échantillon est de taille supérieure à 100, on peut supposer que F suit une loi normale $\mathcal{N}(p, \frac{f(1-f)}{n-1})$ avec $\frac{f(1-f)}{n-1} = \frac{0{,}55\times0{,}45}{199} = 0,0012$ et $\frac{F-p}{\sqrt{\frac{f(1-f)}{n-1}}}$ suit une loi normale centrée réduite avec $\sqrt{\frac{f(1-f)}{n-1}} = 0,035$.

3. Pour déterminer un intervalle de confiance au risque 5%, on cherche t tel que $\phi(|\frac{F-p}{0{,}035}| \leq t) = 0,95$. A l'aide des tables, on obtient $t = 1,96$ puis l'intervalle de confiance :

$$[0,55 - 1,96 \times 0,035; 0,55 + 1,96 \times 0,035] = [0,48; 0,62]$$

Pour déterminer un intervalle de confiance au risque 1%, on cherche t tel que $\phi(|\frac{F-p}{0{,}035}| \leq t) = 0,99$. A l'aide des tables, on obtient $t = 2,57$ puis l'intervalle de confiance :

$$[0,55 - 2,57 \times 0,035; 0,55 + 2,57 \times 0,035] = [0,46; 0,64]$$

4. Dans cette question, l'échantillon est inférieure à 100 et on va donc utiliser les abaques pour l'estimation de p. Ici, $f = \frac{27}{40} = 0,56$. La lecture des abaques donne l'intervalle : $[0,42; 0,72]$.

Exercice 5 : 1. On est dans le cas d'une proportion et on a $p_0 < P_1$. On veut déterminer K_α tel que $P(F \geq K_\alpha) = \alpha$ où F désigne la variable aléatoire qui associe à chaque échantillon de taille 100 la proportion des individus déclarant une allergie. On peut considérer que F suit une loi normale $\mathcal{N}(p_0, \frac{p_0(1-p_0)}{n})$ et donc $Z = \frac{F-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ suit une loi normale centrée réduite. On cherche donc K_α tel que $P(\frac{F-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \geq \frac{K_\alpha-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}) = 1-\alpha$, c'est à dire

$$1 - P(Z < \frac{K_\alpha - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}) = \alpha$$

ou encore

$$\phi(\frac{K_\alpha - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}) = 1 - \alpha$$

Pour $\alpha = 5\%$, à l'aide des tables, on obtient $\frac{K_\alpha-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = 1,65$ et donc

$$K_\alpha = p_0 + 1,65 \times 0,049 = 0,48$$

Pour $\alpha = 10\%$, à l'aide des tables, on obtient $\frac{K_\alpha-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = 1,29$ et donc

$$K_\alpha = p_0 + 1,29 \times 0,049 = 0,46$$

2. Puisque la proportion de l'échantillon étudié est 0,47, dans le cas où $\alpha = 5\%$, on a $0,47 \leq K_\alpha$ et on accepte l'hypothèse H_0. Dans le cas où $\alpha = 10\%$, on a $0,47 > K_\alpha$ et on rejette l'hypothèse H_0 et on accepte l'hypothèse H_1.

3. Le risque de deuxième espèce est déterminé par $\beta = P(\text{accepter } H_0/H_1 \text{ vraie}) = P(F \leq K_\alpha/H_1 \text{ vraie})$. Cette probabilité est donnée par

$$P(\frac{F - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} \leq \frac{K_\alpha - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}}) = \phi(\frac{K_\alpha - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}})$$

On obtient $\beta = \phi(-0,408) = 1 - \phi(0,408) = 0,35$.