Corrigé de l'examen de *Probabilités et statistique* (MATH 046)

Série 1

09 juin 2005

Question 1

Pour chacune des 5 questions ci-dessous, cocher la case de **la** *bonne réponse. On ne demande pas de justification.*

1. Considérons deux événements notés A et B.

   ☐ Si ces deux événements supposés non vides sont indépendants, alors ils ne peuvent pas être incompatibles.

2. Soit X une variable aléatoire de loi de Poisson de paramètre λ. La fonction de répartition de X au point m ($m \in \mathbb{N}$) vaut

   ☐ $\sum_{k=0}^{m} \frac{\lambda^k e^{-\lambda}}{k!}$

3. Soit un test d'une hypothèse nulle H_0 par rapport à une contre-hypothèse H_1. Le risque de deuxième espèce de ce test est

   ☐ $\mathbb{P}(\text{ne pas rejeter } H_0 \mid H_1 \text{ vraie})$

4. Soit $z_{0,10}$ le quantile d'ordre $0,90$ de la loi $N(0,1)$. L'intervalle de confiance au niveau 90% pour une proportion p, dans le cas où la taille n de l'échantillon est assez grande et où p est inconnue, est donné par

   ☐ $\left[\hat{p} - z_{0,05}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{0,05}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$

5. Soient $X_1, \ldots, X_n$ des variables indépendantes et identiquement distribuées, de moyenne μ et de variance σ^2 finie. On définit

   $$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

   Alors,

   ☐ $\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{\text{loi}} N(0,1)$ lorsque $n \to \infty$

Question 2

Soit la fonction $f : \mathbb{R} \to \mathbb{R}$ donnée par

$$f(x) = \begin{cases} Cx^2 & \text{si } -\lambda \leq x \leq 2\lambda \\ 0 & \text{sinon} \end{cases}$$

où λ est un paramètre strictement positif.

(a) Déterminer C en fonction de λ pour que f représente la fonction de densité d'une variable aléatoire continue X.

(b) Sous quelle condition dit-on que $\widehat{\theta}$ est un estimateur non biaisé pour θ?

(c) Trouver la valeur de a dans $\mathbb{R}$ tel que $T \overset{\text{déf}}{=} \dfrac{a}{X}$ soit un estimateur non biaisé pour $\dfrac{1}{\lambda}$ (avec X défini comme au point (a)).

Réponse

(a) Pour que f représente une fonction de densité il faut que (condition nécessaire):

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

On calcule

$$\int_{-\infty}^{+\infty} f(x)dx = C\int_{-\lambda}^{2\lambda} x^2 dx = 3C\lambda^3$$

et il suit que

$$C = \frac{1}{3\lambda^3}$$

(b) $\widehat{\theta}$ est un estimateur non biaisé pour θ ssi $E[\,\widehat{\theta}\,] = \theta$ pour tout θ.
(c) On calcule l'espérance mathématique de T

$$\begin{aligned} E[T] = aE\left[\frac{1}{X}\right] &= a\int_{-\infty}^{+\infty} \frac{1}{x} f(x)dx \\ &= aC\int_{-\lambda}^{2\lambda} x dx = aC\frac{3\lambda^2}{2} \\ &= \frac{a}{2\lambda} \end{aligned}$$

la dernière égalité résultant de l'expression pour C déterminée au point (a). Pour que T soit un estimateur non biaisé pour $1/\lambda$ il faut que $E[T] = 1/\lambda$, c'est-à-dire

$$\frac{a}{2\lambda} = \frac{1}{\lambda}$$

et on voit que $a = 2$.

Question 3

Soient X, Y, Z et T des variables aléatoires indépendantes telles que $X \sim N(1;3)$, $Y \sim N(0;1)$, $Z \sim N(0;1)$ et $T \sim \chi^2_3$.
Déterminer la loi des expressions suivantes:

$Y+Y$	$N(0;4)$
$\dfrac{\sqrt{3}Z}{\sqrt{T}}$	t_3
$X-Y$	$N(1;4)$
$Z+X$	$N(1;4)$
$T+Y^2$	χ^2_4

Attention:

- Indiquer clairement la valeur des paramètres éventuels de ces lois.
- Une loi normale de paramètres μ et σ sera notée $N(\mu, \sigma^2)$.
- On ne demande pas de justification.

Question 4

Une compagnie d'assurances répartit les gens en trois classes : personnes à bas risque, risque moyen et haut risque. Ses statistiques indiquent que la probabilité que des gens soient impliqués dans un accident sur une période d'un an est respectivement $0,05$, $0,15$ et $0,30$. On estime que 20% de la population est à bas risque, 50% à risque moyen et 30% à haut risque.

(a) Quelle proportion des gens ont eu au moins un accident au cours d'une année donnée?

(b) Si l'assuré A n'a pas eu d'accident cette année, quelle est la probabilité qu'il fasse partie de la classe à bas risque?

Réponse

(a)

$$
\begin{aligned}
&\mathbb{P}(\text{au moins 1 accident})\\
&= 1 - \mathbb{P}(0 \text{ accident})\\
&= 1 - \mathbb{P}(0 \text{ accident}|\text{bas risque})\,\mathbb{P}(\text{bas risque})\\
&\quad\quad - \mathbb{P}(0 \text{ accident}|\text{moyen risque})\,\mathbb{P}(\text{moyen risque})\\
&\quad\quad - \mathbb{P}(0 \text{ accident}|\text{haut risque})\,\mathbb{P}(\text{haut risque})\\
&= 1 - 0,95.\frac{20}{100} - 0,85.\frac{50}{100} - 0,70.\frac{30}{100}\\
&= 0,175
\end{aligned}
$$

(b)

$$
\begin{aligned}
&\mathbb{P}(\text{bas risque} \mid 0 \text{ accident})\\
&= \frac{\mathbb{P}(0 \text{ accident} \mid \text{bas risque})\mathbb{P}(\text{bas r.})}{\mathbb{P}(0 \text{ acc.}|\text{B r.})\,\mathbb{P}(\text{B r.}) + \mathbb{P}(0 \text{ acc.}|\text{M r.})\,\mathbb{P}(\text{M r.}) + \mathbb{P}(0 \text{ acc.}|\text{H r.})\,\mathbb{P}(\text{H r.})}\\
&= \frac{0,95\frac{20}{100}}{0,95\frac{20}{100} + 0,85\frac{50}{100} + 0,70\frac{30}{100}}\\
&= 0,23
\end{aligned}
$$

Question 5

La brasserie *Pivo* en Tchèquie vend de la bière en bacs de 24 bouteilles. A cause d'un problème technique un bac ne contient que 16 bouteilles de bière, les autres bouteilles sont des bouteilles à biere remplies d'eau...
L'importateur en Belgique, n'étant pas au courant du problème technique, s'étonne du nombre de plaintes reçues concernant la présence de bouteilles remplies d'eau dans les bacs de bière. Il décide d'être plus vigilant: désormais, chaque bac sera contrôlé! Voici les deux methodes qu'il envisage (les prélèvements se font toujours au hasard).

- La première méthode consiste à prélever sans remise 3 bouteilles d'un bac et à contrôler leur contenu. Si les trois bouteilles contiennent de la bière, le bac est accepté. Sinon, il est refusé.
- La seconde méthode consiste à prélever avec remise 5 bouteilles d'un bac et à les examiner. Si au moins deux de ces bouteilles sont remplies d'eau, le bac est renvoyé à la brasserie. Si ce n'est pas le cas, il est accepté.

Calculer la probabilité qu'un bac soit refusé pour *chacune* de ces deux méthodes. Quelle est donc la méthode la plus efficace pour l'importateur?

Réponse

Méthode 1
On a

$$\begin{aligned}\mathbb{P}(\text{bac refusé}) &= 1 - \mathbb{P}(\text{bac accepté}) \\ &= 1 - \mathbb{P}(\text{prélever 3 bouteilles de bière}) \\ &= 1 - \frac{16}{24} \cdot \frac{15}{23} \cdot \frac{14}{22} \\ &= 1 - \frac{140}{506} = 0,723\end{aligned}$$

Méthode 2
Soit X le nombre de bouteilles remplies de l'eau parmi les 5 bouteilles prélevées d'un bac. Alors, X suit un loi binomiale de paramètres 5 et 1/3. Il vient

$$\begin{aligned}\mathbb{P}(\text{bac refusé}) &= \mathbb{P}(X \geq 2) \\ &= 1 - (\mathbb{P}(X = 0) + \mathbb{P}(X = 1)) \\ &= 1 - \binom{5}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^5 - \binom{5}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^4 \\ &= 1 - \frac{32}{81} = 0,395\end{aligned}$$

Conclusion
La première méthode est la plus efficace pour l'importateur.

Question 6

La largeur (en cm) d'une fente entaillée dans une pièce fabriquée en aluminium est distribuée selon une loi normale de paramètres $\mu = 2$ et $\sigma = 0,007$. Les limites de tolérance sont données comme étant $1,988$ et $2,012$.

(a) Quel sera le pourcentage de pièces défectueuses?

(b) Quelle est la valeur maximale que peut prendre σ afin que le pourcentage de pièces défectueuses ne dépasse pas 1%, si la largeur des fentes suit une distribution normale de paramètres $\mu = 2$ et σ?

Réponse

Notons $Z \sim N(0,1)$.
(a)

$$\begin{aligned}
\mathbb{P}(\text{pièce défectueuse}) \\
&= 1 - \mathbb{P}(\text{bonne pièce}) \\
&= 1 - \mathbb{P}(1,988 \leq X \leq 2,012) \\
&= 1 - \mathbb{P}(X \leq 2,012) + \mathbb{P}(X \leq 1,988) \\
&= 1 - \mathbb{P}\left(\frac{X-2}{0,007} \leq \frac{2,012-2}{0,007}\right) + \mathbb{P}\left(\frac{X-2}{0,007} \leq \frac{1,988-2}{0,007}\right) \\
&= 1 - \mathbb{P}(Z \leq 1,71) + \mathbb{P}(Z \leq -1,71) \\
&= 1 - \mathbb{P}(Z \leq 1,71) + 1 - \mathbb{P}(Z \leq 1,71) \\
&= 2 - 2 \cdot 0,9564 \\
&= 0,0872
\end{aligned}$$

(b)

$$\begin{aligned}
&1 - \mathbb{P}(X \leq 2,012) + \mathbb{P}(X \leq 1,988) \leq 0,01 \\
&ssi\ 1 - \mathbb{P}\left(\frac{X-2}{\sigma} \leq \frac{2,012-2}{\sigma}\right) + \mathbb{P}\left(\frac{X-2}{\sigma} \leq \frac{1,988-2}{\sigma}\right) \leq 0,01 \\
&ssi\ 1 - \mathbb{P}\left(Z \leq \frac{0,012}{\sigma}\right) + \mathbb{P}\left(Z \leq \frac{-0,012}{\sigma}\right) \leq 0,01 \\
&ssi\ 1 - \mathbb{P}\left(Z \leq \frac{0,012}{\sigma}\right) + 1 - \mathbb{P}\left(Z \leq \frac{0,012}{\sigma}\right) \leq 0,01 \\
&ssi\ \mathbb{P}\left(Z \leq \frac{0,012}{\sigma}\right) > \frac{1,99}{2} \\
&ssi\ \frac{0,012}{\sigma} > 2,58 \\
&ssi\ \sigma < 4,65.10^{-3}
\end{aligned}$$

Question 7

Dans ses expériences sur les petits pois, Gregor Mendel avait observé que 315 étaient ronds et jaunes, 108 ronds et verts, 101 étaient ridés et jaunes, et 32 étaient ridés et verts. D'après sa théorie sur l'hérédité, les pois devraient être répartis dans les proportions 9 : 3 : 3 : 1. Peut-on mettre en doute sa théorie à la suite de ces observations au niveau de probabilité 5%? Justifiez votre réponse.

Réponse

Il s'agit d'un test χ^2 d'homogénéité.
Le nombre total de pois est 556. Comme les nombres théoriquement attendus sont dans les proportions 9 : 3 : 3 : 1 (et $9+3+3+1=16$), on s'attend à

$$\begin{aligned}
\frac{9}{16}\cdot 556 &= 312.75 \qquad \textit{ronds et jaunes} \\
\frac{3}{16}\cdot 556 &= 104.25 \qquad \textit{ridés et jaunes} \\
\frac{3}{16}\cdot 556 &= 104.25 \qquad \textit{ronds et verts} \\
\frac{1}{16}\cdot 556 &= 34.75 \qquad \textit{ridés et verts}
\end{aligned}$$

Ainsi

$$T=\frac{(315-312.75)^2}{312.75}+\frac{(108-104.25)^2}{104.25}+\frac{(101-104.25)^2}{104.25}+\frac{(32-34.75)^2}{34.75}=0.47$$

Il y a 4 catégories, donc le nombre de degrés de liberté est $\nu=3$.
Puisque $Q>\chi^2_{3;0.95}=7.81$, nous ne pouvons pas rejeter la théorie au niveau 0.05.

On conclut que l'expérience s'accorde avec la théorie.

Question 8

Un cultivateur de moules désire étudier la qualité des moules de son élevage. Il est connu que la qualité d'une moule est directement proportionnelle à son poids. Donc, il choisit 10 moules au hasard et il note leur poids en grammes comme suit:

$$21; 23; 33; 40; 37; 29; 24; 32; 24; 17$$

(a) Construisez un intervalle de confiance à 95% pour le poids moyen des moules, en supposant que le poids des moules soit une variable normale de moyenne et écart-type inconnus.

(b) A la suite d'un contrôle de qualité effectué par les experts de l'*Association MouleEnLiberté*, on lui propose de cultiver une nouvelle variété de moules provenant du Quichaha. Sur la base d'une étude effectuée sur un échantillon de 15 moules du Quichaha cultivé dans un autre élevage, on observe un poids moyen de $\overline{x} = 35.3$ g et écart-type $s = 3.1$ g. Conseilleriez-vous au cultivateur de passer à la nouvelle variété de moules pour augmenter la qualité de son élevage? Construisez le test approprié au niveau de probabilité 0.05.

[Spécifiez les hypothèses qui vous permettent de conduire le test de votre choix.]

Réponse

(a) Il s'agit de construire un intervalle de confiance pour une moyenne, sous l'hypothèse de population normale, avec écart-type inconnu. L'intervalle de confiance sera tel que

$$\mathbb{P}\left(\overline{x} - \frac{s}{\sqrt{n-1}} t_{n-1;1-\frac{\alpha}{2}} \leq \mu \leq \overline{x} + \frac{s}{\sqrt{n-1}} t_{n-1;1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$

où $\overline{x} = 28$ et $s = 7.41$ sont respectivement la moyenne et l'écart-type calculés sur l'échantillon. Etant données les valeurs de $n - 1 = 9$ et $\alpha = 0.05$ (le niveau de confiance est $1 - \alpha = 0.95$), on obtient $t_{9;0.975} = 2.26$. Ainsi,

$$\mathbb{P}(L^- \leq \mu \leq L^+) = 0.95$$

c'est l'intervalle de confiance à 95% qu'on demande, oú $L^- = 22.42$ et $L^+ = 33.58$.

(b) Il s'agit de conduire un test d'hypothèse pour une comparaison de deux moyennes, sous l'hypothèse de population normale, avec écart type inconnu. Si on note μ_1 le poids moyen de la population des moules de la variété élevée par le cultivateur et μ_2 le poids moyen de la population des moules du Quichaha, le problème du test peut être résumé comme suit

$$\begin{aligned} H_0 &: \mu_1 \geq \mu_2 \\ H_1 &: \mu_1 < \mu_2. \end{aligned}$$

En considérant les hypothèses énoncées (population normale, écart-type inconnu), et en supposant que

1. les deux échantillons soient mutuellement indépendants;

2. la dispersion du poids de ces deux variétés de moules soit la même, c.à.d. $\sigma_1 = \sigma_2$;

il est clair que la statistique du test à appliquer est la suivante

$$T = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2} \frac{n_1+n_2}{n_1 n_2}}}$$

Avec les valeurs $\overline{x}_1 = 28$, $\overline{x}_2 = 35$, $s_1 = 7.41$, $s_2 = 3.1$, $n_1 = 10$ et $n_2 = 15$, sous l'hypothèse nulle on obtient $T = -3.12$.
Sous H_0 la statistique du test considéré suit la loi de Student avec $n_1 + n_2 - 2 = 23$ degrés de liberté. Puisque le test considéré est unilateral, la valeur critique est donnée par $t_{23;0.05} = -1.71$; il en suit que nous rejetons l'hypothèse nulle au niveau 0.05, car $T = -3.12 < -1.71$.

On conclut que, au niveau 5%, le cultivateur aurait intérêt à élever les moules du Quichaha.

Question 9

A la suite des réclamations de certains clients, le propriétaire d'un restaurant s'intéresse à la variabilité de la contenance des carafes du vin rouge. Il mesure le contenu de 12 carafes prises au hasard et obtient $s = 1.5$ cl. Il se demande alors si la variance de la distribution du contenu des carafes est significativement différente de 1 au niveau de probabilité 0.05.

(a) Effectuez le test approprié en supposant que la distribution du contenu des flacons est normale.

(b) Construisez un intervalle de confiance à 90% pour σ^2.

Réponse

(a) Il s'agit de construire un test d'hypothèse pour une variance, sous l'hypothèse de population normale. Le problème du test peut être résumé comme suit

$$\begin{aligned} H_0 &: \sigma^2 = 1 \\ H_1 &: \sigma^2 \neq 1. \end{aligned}$$

En considérant l'hypothèse énoncée (population normale), il est clair que la statistique du test à appliquer est la suivante

$$T = \frac{ns^2}{\sigma^2}$$

Avec les valeurs $n = 12$ et $s^2 = 2.25$, sous l'hypothèse nulle on obtient $T = 27$.
Sous H_0 la statistique du test considéré suit la loi chi-carré avec $n - 1 = 11$ degrés de liberté. Puisque le test considéré est bilatéral, on rejettera H_0 si $T \notin [\chi^2_{11;0.025}; \chi^2_{11;0.975}]$, c.à.d. si $T \notin [3.82; 21.9]$; il en suit que nous rejetons l'hypothèse nulle au niveau 0.05, car $T \notin [3.82; 21.9]$.

On conclut que, au niveau 5%, la variance du contenu des carafes n'est pas significativement différente de 1.
(b) Il s'agit de construire un intervalle de confiance pour une variance, sous l'hypothèse de population normale. L'intervalle de confiance sera tel que

$$\mathbb{P}\left(\frac{ns^2}{\chi^2_{11;0.95}} \leq \sigma^2 \leq \frac{ns^2}{\chi^2_{11;0.05}}\right) = 0.90,$$

oú $n = 12$ et $s = 1.5$. Etant données les valeurs de $\chi^2_{11;0.05} = 4.57$ et $\chi^2_{11;0.95} = 19.7$, on obtient l'intervalle de confiance

$$\mathbb{P}(L^- \leq \mu \leq L^+) = 0.90$$

oú $L^- = 1.37$ et $L^+ = 5.91$.