utbm
université de technologie Belfort-Montbéliard

Probabilités Statistiques

Samedi 21 juin 2003

Examen final Printemps 03

Une rédaction complète sera appréciée. Tous les résultats doivent être soigneusement justifiés.
Comme dans tout document écrit, la présentation doit être irréprochable.

Matériel autorisé: Calculatrice Tables de Statistiques Feuille aide-mémoire A4 recto

Partie I (*barème indicatif* 11 points)

1°) On s'intéresse au prix d'une nuitée (2 personnes, une voiture et une tente) dans les terrains de camping en prenant un échantillon de 100 terrains de même catégorie. La collecte des résultats en 2001 a donné les résultats suivants :

Pris d'une nuitée en €	8	9	10	11	12	13	14	
nombre de terrains	1	3	8	20	31	26	8	3

a) Montrer que les estimations de la moyenne et de la variance sont 11,5 et 1,8.

b) Montrer par un test de niveau 0,95, qu'on peut considérer que le prix d'une nuitée suit une loi normale .

c) Un vacancier part un mois (30 jours) en vacances en camping avec un budget de 360 € pour le logement. Calculer la probabilité de l'événement « le budget est suffisant » dans les cas suivants :

- Il passe toutes ses vacances dans le même terrain
- Il change de terrain de camping tous les jours.

2°) On effectue une enquête dans deux pays d'Europe sur les prix en € des terrains de camping. On suppose que ces prix suivent des lois $N(M_1, \sigma_1)$ et $N(M_2, \sigma_2)$. Deux échantillons ont été prélevés, et ils ont donné les résultats suivants :

	Taille de l'échantillon	Estimation de M	Estimation de σ = s
Pays 1	$n_1 = 20$	$m_1 = 10{,}8$	$s_1 = 1{,}4$
Pays 2	$n_2 = 30$	$m_2 = 13{,}2$	$s_2 = 1{,}8$

a) Le Pays 1 annonce dans ses brochures que le prix moyen des terrains est de 10 €. Une association de consommateurs prétend que ce prix est de 11,5 €. Tester $M_1 = 10$ contre $M_1 = 11{,}5$, et déterminer qui a raison, et le risque β, en prenant $\alpha = 0{,}05$.

b) Tester l'égalité des variances au niveau 0,95.

c) Les prix des terrains dans les deux pays sont-ils égaux (en moyenne) ?

Partie II (*barème indicatif* : 13 points)

3°) Deux enquêtes portant sur 1000 français partant en vacances, en 2001 et en 1991, ont donné les résultats suivants : En 2001, 170 personnes interrogées partaient à l'étranger, alors qu'en 1991 elles étaient 130.

a) Soit p_2 la proportion des français partant à l'étranger en 2001. Pour quelle valeur de n, taille de l'échantillon, l'intervalle de confiance sur p_2 serait de la forme $\left]F_2 - 0{,}012,\ F_2 + 0{,}012\right[$?

b) Peut-on dire, au niveau 0,95, que les habitudes ont changé au cours de ces 10 années ?

4°) La durée de séjour T, en jours, dans un hôtel suit une loi de Poisson de paramètre λ, et le gérant, ancien étudiant de SQ 20, cherche à estimer le paramètre λ pour l'année 2001

a) Soit un échantillon $E_n = \{T_1, \ldots, T_n\}$ d'observation $e_n = \{t_1, \ldots, t_n\}$ de variables indépendantes de même loi que T. Déterminer la fonction de vraisemblance de l'observation e_n. En déduire les équations de vraisemblance, puis un estimateur de λ.

b) Une observation sur 20 clients (supposés indépendants) a donné les résultats : (une durée de séjour de 0 signifie que le client a annulé sa réservation).

k = nombre de jours	0	1	2	3	4	5	6
n_k =nombre de clients	2	3	5	4	3	2	1

Déterminer une estimation ponctuelle de λ

c) L'hôtelier affirme que la durée de séjour dans son hôtel a pour paramètre $\lambda = 3{,}5$. Peut-être !

Pour en être sûr à 95% , on considère la variable de test $Y = \sum_{k=1}^{20} T_k$. Etudier la loi de Y et déterminer son espérance et sa variance.

d) Déterminer deux constantes y_1 et y_2 telles que $p(Y \le y_2) \approx 0{,}975$ et $p(Y < y_1) \approx 0{,}025$

Pour pallier la limitation des tables de Poisson, on pourra utiliser librement le résultat suivant :

Si N suit une loi de Poisson de paramètre λ, alors $p(N \le n) = p(\chi^2_{2n+2} > 2\lambda)$

En déduire le domaine d'acceptation de l'hypothèse H_0 , puis la décision.

5°) Question hors barème sur 4 points)

On s'intéresse au retard par rapport à la durée de voyage prévue pour un voyage de 500 km par le train. Ce retard R (exprimé en minutes) suit une loi de densité $h(t) = \dfrac{t^2\, e^{-0,5\,t}}{16} 1_{\mathbb{R}_+}(t)$.

a) Montrer que h définit effectivement une densité de probabilité. En calculer l'espérance et la variance.

b) La direction des Chemins de Fer rembourse le billet si le retard dépasse 15 minutes. Calculer la probabilité d'être remboursé.

c) Calculer la densité de la variable aléatoire R_2 = durée totale de retard sur un aller – retour.

d) Un voyageur effectue le trajet (aller ou retour) 150 fois dans l'année. On note R_{150} la durée totale du retard, et N_{150} le nombre de fois pour lesquelles le voyageur a été remboursé.

Déterminer les lois approximatives de R_{150} et de N_{150} . Calculer $p(R_{150} > 950)$ et $p(N_{150} \ge 5)$.

(N.d.l.r : Les données du problème sont réalistes, mais certaines d'entre elles sont fictives, alors que d'autres reposent sur des données réelles.)

Document joint : Table χ^2

Je réponds ordinairement, à ceux qui me demandent la raison de mes voyages, que je sais bien ce que je fuis, mais non pas ce que je cherche.

Michel de MONTAIGNE (France: 1533 – 1592) *Essais*

II. CORRIGÉ

1°) Les calculs donnent : on a donc $\bar{x} = \frac{1}{100}\sum x_i = 11{,}52 \approx 11{,}5$ et $\sigma^2_{obs} = \frac{1}{100}\sum x_i^2 - 11{,}52^2 = 1{,}8$

x_{i-1}	x_i	x'_i	n_i	$n_i x'_i$	$n_i(x'_i-11{,}8)^2$	$\frac{x_i-\bar{x}}{1{,}342}$	loi normale	proba	eff, th	cumul	chi 2
7	8	7,5	1	7,5	16,2	-2,61	0,005	0,005	0,45		
8	9	8,5	3	25,5	27,4	-1,87	0,031	0,026	2,63		
9	10	9,5	8 (12)	76	32,6	-1,13	0,130	0,099	9,90	12,98	0,07
10	11	10,5	20	210	20,8	-0,39	0,350	0,220	22,01		0,18
11	12	11,5	31	356,5	0,0	0,36	0,639	0,289	28,92		0,15
12	13	12,5	26	325	25,0	1,10	0,864	0,225	22,48		0,55
13	14	13,5	8 (11)	108	31,4	1,84	0,967	0,103	10,32	13,12	0,34
14	15	14,5	3	43,5	26,6	2,58	0,995	0,028	2,80		
Total			100	1152	180,0						1,30

a) On a donc les estimations de la moyenne 11,8 et de la variance 1,8 et de l'écart type 1,342.

b) H_0 : X est normale contre H_1 : X n'est pas normale a pour variable de test $D^2 = \sum \frac{(N_i - np_i)^2}{np_i} \approx \chi^2_4$.

Le domaine d'acceptation est donné par la table χ^2 : 0,95 → 9,49, c'est à dire D_0 =[0 ; 9,49[. et comme d² = 1,30, on accepte l'hypothèse H_0 , c'est à dire $\boxed{\text{X suit une loi normale}}$.

c) Avec toutes les vacances dans le même terrain, il a : coût total = X_1 = 30X, avec $X_1 \approx$N(30x11,5, var = 30²x1,8).

Donc $p(X_1 < 360) = p\left(\frac{X_1 - 345}{\sqrt{900\times 1{,}8}} < \frac{360-345}{\sqrt{900\times 1{,}8}}\right) = p(N(0,1) < 0{,}37) = 0{,}644$. $\boxed{\text{p(budget suffisant) = 0,644 .}}$

Avec toutes les vacances dans des terrains ≠ (ind.), il a : coût total = $X_2 = \Sigma X_k$, avec $X_2 \approx$N(30x11,5, var = 30x1,8).

Donc $p(X_2 < 360) = p\left(\frac{X_2 - 345}{\sqrt{30\times 1{,}8}} < \frac{360-345}{\sqrt{30\times 1{,}8}}\right) = p(N(0,1) < 2{,}04) = 0{,}98$. $\boxed{\text{p(budget suffisant) = 0,98 .}}$

2°) Observations : voir énoncé :

a) Test : $H_0 : M_1 = 20$ contre $H_1 : M_1 = 11{,}5$, variable de test $Y = \frac{\overline{X}_1 - 10}{S_1/\sqrt{20}} \approx T_{19}$. Table : 0,95 → 1,729

Donc D_0 =]-∞ ; 10,58[avec l'observation s = 1,5. Obs. moyenne = 10,8 ∉ D_0, donc $\boxed{: H_1 : M_1 = 11{,}5}$.
$\beta = p(\overline{X}_1 < 10{,}58 | M_1 = 11{,}5) = p(T_{19} < -2{,}74) \approx 0{,}003$ donc $\boxed{\text{risque } \beta = 0{,}003}$

b) Test $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$ de variable $Z = \frac{S_2^2}{S_1^2} \approx F(29,19)$. Table F(29, 19) : 2,075. Donc le domaine D_0 = [1 ; 2,075[. Et comme l'observation $\frac{s_2^2}{s_1^2} = \left(\frac{1{,}8}{1{,}5}\right)^2 = 1{,}44 \in D_0$ on accepte $\boxed{H_0 : \text{variances égales}}$.

Estimateur de σ² = $S^2 = \frac{19S_1^2 + 29S_2^2}{48}$ d'observation s² = 2,848

c) Egalité des moyennes : : $H_0 : M_1 = M_2$ contre $H_1 : M_1 \neq M_2$, de variable $Y_m = \frac{\overline{X}_1 - \overline{X}_2}{S\sqrt{\frac{1}{20}+\frac{1}{30}}} \approx T_{48}$.

La table T_{48} donne 0,975 → 2 et donc D_0 =]-0,98 ; 0,98[avec s=1,7. Comme l'observation des moyennes donne 10,8 – 13,2 = -2,4 ∉ D_0 On doit rejeter H_0 et donc $\boxed{\text{il existe une différence entre les deux pays}}$.

3°) $p_{1 \text{ ou } 2}$ = proportion étranger / total et 2 échantillons (X_1, … , X_{1000}) avec $X_k \approx B(1,p_{1 \text{ ou } 2})$.

a) L'intervalle de confiance de p_2 s'écrit $I = \left]F_2 - 1{,}96\sqrt{\frac{p_2(1-p_2)}{n}}, F_2 + 1{,}96\sqrt{\frac{p_2(1-p_2)}{n}}\right[$ et pour l'observation f_2=0,17, on a $p_2(1-p_2) \approx 0{,}17\times 0{,}83$ et donc $1{,}96\sqrt{\frac{0{,}14}{n}} = 0{,}012$ ce qui donne $\boxed{\text{taille n = 3735}}$.

b) Test $H_0 : p_1 = p_2 = p$ contre $H_1 : p_1 \neq p_2$, avec f = estimation de p et la variable de test X':

$f = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = 0{,}15$ et $X' = \dfrac{F_1 - F_2 - (p_1 - p_2)}{\sqrt{p(1-p)\left(\frac{1}{1000} + \frac{1}{1000}\right)}} \approx \dfrac{F_1 - F_2 - (p_1 - p_2)}{\sqrt{0{,}15(1-0{,}15)\left(\frac{1}{1000} + \frac{1}{1000}\right)}} \approx N(0,1)$.

D'après la table N(0,1) on a $p\left(-1{,}96 \times 0{,}016 < F_1 - F_2 < 1{,}96 \times 0{,}016\right) = 0{,}95 \Rightarrow D_0 = \left]-0{,}03, +0{,}03\right[$.

L'observation $f_1 - f_2 = 0{,}04 \notin (D_0)$, donc on décide $\boxed{H_1 \text{ : les habitudes ont changé entre 91 et 01 .}}$

4°) X suit une loi de Poisson de paramètre λ

a) Pour $(X_1,\dots,X_n)$, d'observation $(x_1,\dots,x_n)$ on a $L(x_1,\dots,x_n,\lambda) = \prod_{1\le k\le n} \dfrac{\lambda^{x_k} e^{-\lambda}}{x_k!} = e^{-\lambda n} \dfrac{\lambda^{\sum x_k}}{x_1!\dots x_n!}$.

Donc $\ell n\, L = -\lambda n + \sum \left(x_k - \ell n\, x_k!\right)$ et $\begin{cases} \dfrac{\partial \ell n L}{\partial \lambda} = 0 \\ \dfrac{\partial^2 \ell n L}{\partial \lambda^2} < 0 \end{cases} \Leftrightarrow \begin{cases} -n + \dfrac{1}{\lambda}\sum x_k = 0 \\ \dfrac{1}{\lambda^2}\sum x_k < 0 \end{cases}$ donc $\lambda = \overline{x}$. L'estimateur de λ de maximum de vraisemblance est donc $\boxed{T(X_1,\dots X_n) = \dfrac{1}{n}\sum X_k = \overline{X}}$

b) Estimation de λ : m = (2 x 0 + 1 x 3 + … + 6 x 1) / 20 = 53 / 20 = 2,65. On sait que c'est une estimation sans biais.

c) On veut tester $H_0 : \lambda = 3{,}5$ contre $H_1 : \lambda \neq 3{,}5$, et on considère la variable de test Y, somme de 20 variables de Poisson indépendantes de paramètre $\lambda = 3{,}5$ (d'après H_0). Donc Y suit une loi P(3,5x20) = P(70). Pour déterminer le domaine d'acceptation on cherche y_2 entier vérifiant $p(Y \le y_2) \approx 0{,}975 = p\left(\chi^2_{2n+2} > 140\right) \Rightarrow p\left(\chi^2_{2n+2} < 140\right) = 0{,}025$. Par lecture de la table, on trouve dans la colonne 0,025 : 2n+2 = 174 et donc $p(Y \le 86) = 0{,}975$.

Par un calcul analogue, on obtient $p(Y \le 54) = 0{,}025$ et donc $p(Y < 55) = 0{,}025$. Donc $\boxed{D_0 = \{55, 56, \dots, 86\}}$

L'observation étant $53 \notin D_0$ à partir de l'échantillon on décide $\boxed{H_1 \text{ et donc } \lambda \neq 3{,}5}$

5°) On ne s'intéresse qu'aux valeurs positives de t. On utilisera $\forall n \in N^* \ \Gamma(n) = \int_0^\infty t^{n-1} e^{-t} dt = (n-1)!$ fonction Γ.

a) h est une densité car h est continue sur R_+ et $\displaystyle\int_0^\infty \frac{t^2 e^{-0{,}5t}}{16} dt = \int_0^\infty \frac{4u^2 e^{-u}}{16} 2du = \frac{1}{2}\Gamma(3) = 1$ (avec t = 2u) .De plus $E(T) = \int_0^\infty t\, h(t)\, dt = 6$ et $Var(T) = \int_0^\infty t^2\, h(t)\, dt - 36 = 12$.

b) $p(T > 15) = \int_{15}^\infty h(t)\, dt = 0{,}02$ obtenu avec deux intégrations par parties, donc $\boxed{p(>15 \text{ min retard}) = 0{,}02}$.

c) R_2= somme de deux T indépendantes, donc densité g(t)=h*h(t) avec (* = produit de convolution). Donc on a :

$$g(t) = \int_{-\infty}^\infty h(x) h(t-x)\, dx = \int_0^t \frac{x^2 e^{-0{,}5x}}{16} \times \frac{(t-x)^2 e^{-0{,}5(t-x)}}{16} dx = \dots = \frac{t^5 e^{-0{,}5t}}{7680}$$ (intégration par rapport à x en non pas t !)

d) Nombre de voyages : n = 150 et probabilité de retard = 0,02.

Les voyages étant supposés indépendants, N_{150} suit une loi binomiale B(n=150 ; p = 0,02), qu'on approche par une loi de Poisson de paramètre $\lambda = 3$ car n assez grand et p petit . On a donc, d'après la table P(3) $p(N_{150} \le 4) = 0{,}815$ et donc $\boxed{p(N_{150} \ge 5) = 0{,}125}$.

De même R_{150} est la somme de 150 variables indépendantes de densité h, d'espérance 6 et de variance 12. D'après le théorème central limite, $\dfrac{R_{150} - 150 \times 6}{\sqrt{150 \times 12}} \approx N(0,1)$ et $p(R_{150} > 950) = p\left(N(0,1) > \dfrac{950 - 900}{\sqrt{150 \times 12}}\right) = 0{,}12$.

Remarque sur cette partie 5°) :

On peut voir que, d'après la densité R est la somme de 3 variables exponentielles indépendantes de paramètre 0,5, donc les calculs de l'espérance et de la variance sont immédiats.

De plus R_2 est la somme de 6 variables exponentielles indépendantes.

On pourrait considérer que le retard sur un voyage est la somme de trois retards sur les tiers su parcours.

Un rappel utile quand on veut utiliser cette méthode :

Si les T_k suivent des lois $E(\lambda)$ $\sum_{1\le k\le n} T_k$ a pour densité $\varphi(t) = \dfrac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} 1_{R_+}(t)$