Initiation aux Statistiques **TS 13B**

Correction Examen du 17/02/2015

Exercice 1 Soit X de loi $U([0; a])$ et $(X_1, \dots, X_n)$ un n-échantillon de X. On cherche à estimer a.
On note $\overline{X}$ la moyenne empirique de X.

1. Soit $T_n = 2\overline{X}$.

$$E(T_n) = E\left(2\overline{X_n}\right) = 2\times\frac{1}{n}\sum_{i=1}^{n}E(X_i) = \frac{2}{n}\sum_{i=1}^{n}\frac{a}{2} = a$$

donc T_n est un estimateur sans biais de a.

$$V(T_n) = V\left(\frac{2}{n}\sum_{i=1}^{n}X_i\right) = \frac{4}{n^2}\sum_{i=1}^{n}V(X_i) = \frac{4}{n^2}\sum_{i=1}^{n}\frac{a^2}{12} = \frac{a^2}{3n}\underset{n\to\infty}{\longrightarrow}0$$

donc T_n est convergent.

2. Soit $Yn = \max(X_1, \dots, X_n)$.

La fonction de répartition de X est définie par :

$$F(X) = \begin{cases} 0 \text{ si } x<0 \\ \frac{x}{a} \text{ si } 0\le x\le a \\ 1 \text{ si } x>a \end{cases}.$$

Si G_n est la fonction de répartition de Y_n, alors :

$$G_n(x) = P(Y_n\le x) = P\left([X_1\le x]\cap\cdots\cap[X_n\le x]\right) = P(X_1\le x)\times\cdots\times P(X_n\le x)$$

car les X_i sont indépendants, et donc :

$$G_n(x) = \begin{cases} 0 \quad \text{si } x<0 \\ \left(\frac{x}{a}\right)^n \text{ si } 0\le x\le a \\ 1 \quad \text{si } x>a \end{cases}.$$

On en déduit une densité de Y_n :

$$g_n(x) = \begin{cases} 0 \quad \text{si } x<0 \\ \frac{n}{a}\left(\frac{x}{a}\right)^{n-1} \text{ si } 0\le x\le a \\ 0 \quad \text{si } x>a \end{cases}.$$

D'autre part, on a :

$$E(Y_n) = \int_0^a t\,\frac{n}{a}\left(\frac{t}{a}\right)^{n-1}dt = \frac{n}{a^n}\int_0^a t^n dt = \frac{n}{a^n}\left[\frac{t^{n+1}}{n+1}\right]_0^a = \frac{na}{n+1},$$

donc son biais $b(Y_n)$ vaut :

$$b(Y_n) = \frac{na}{n+1} - a = \frac{-a}{n+1},$$

et sa variance vaut :

$$V(Y_n) = \int_0^a t^2\frac{n}{a}\left(\frac{t}{a}\right)^{n-1}dt - \left(E(Y_n)\right)^2 = \frac{na^2}{n+2} - \frac{n^2a^2}{(n+1)^2} = \frac{na^2}{(n+2)(n+1)^2},$$

3. Soit $Z_n = \dfrac{n+1}{n} Yn$.

$$E(Z_n) = E\left(\frac{n+1}{n} Y_n\right) = \frac{n+1}{n} E(Y_n) = \frac{n+1}{n} \times \frac{na}{n+1} = a,$$

donc Z_n est un estimateur sans biais de a et :

$$V(Z_n) = V\left(\frac{n+1}{n} Y_n\right) = \left(\frac{n+1}{n}\right)^2 V(Y_n) = \left(\frac{n+1}{n}\right)^2 \times \frac{na^2}{(n+2)(n+1)^2} = \frac{a^2}{n(n+2)} \underset{n\to\infty}{\longrightarrow} 0,$$

donc il est convergent.

4. Pour de grandes valeurs de n, comme T_n et Z_n sont sans biais, ils sont meilleurs que Y_n et comme :

$$V(T_n) \underset{\infty}{\sim} \frac{a^2}{3n} \text{ et } V(Z_n) \underset{\infty}{\sim} \frac{a^2}{n^2},$$

alors Z_n est le meilleur estimateur de a.

Exercice 2 Nous observons deux caractères statistiques sur une population de 10 décathloniens. Il s'agit de la longueur maximale atteinte (en mètre) lors de cinq lancers de poids et de la longueur maximale atteinte (en mètre également) sur cinq tentatives de saut en longueur. Le tableau suivant représente les observations de ces deux caractères.

Lancer de poids	18.75	17.05	15.83	19.58	11.31	18.58	17.16	14.80	16.32	16.90
Saut en longueur	6.51	7.60	6.33	7.49	5.36	7.57	7.98	6.74	8.12	8.08

1) Calculer les paramètres de position des deux séries :
   moyenne : 16,628 / 7,178 ; médiane 16,975 / 7,53

2) Calculer les paramètres de dispersion des deux séries.
   variance : 4,99 / 0,74; écart-type : 2,23 / 0,86 ;
   écart inter-quartile : 18,58 – 15,83 = 2,75 / 7,98 – 6,51 = 1,47

3) Donner l'équation de la droite de régression. y = 0,24x + 3,19.

4) Calculer le coefficient de corrélation entre les deux séries :

$$r = \frac{\operatorname{cov}(x;y)}{s_x s_y} \simeq 0,62 \quad \text{où} \quad \operatorname{cov}(x;y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Peut-on dire qu'il y a une dépendance entre les deux caractères statistiques? Non

Exercice 3

Nous avons observé 9 ILS jusqu'à ce que chacun ait une défaillance. Le tableau suivant donne les temps, en jour, de première défaillance pour chaque ILS.

ILS	1	2	3	4	5	6	7	8	9
Temps	130	20	348	100	14	212	64	50	135

1. Déterminer une estimation de R(t).

Méthode des rangs médians car n = 9 < 20 : $R(t) = \dfrac{n + 0,7 - i}{n + 0,4}$

Temps	14	20	50	64	100	130	135	212	348
R(t)	0,93	0,82	0,71	0,61	0,5	0,39	0,29	0,18	0,07

2. En supposant que le temps de bon fonctionnement suit une loi exponentielle de paramètre $\lambda = 0{,}006$ panne/jour, déterminer à quel instant t_0, la fiabilité est égale à 85%.

$$R(t) = 80\% = e^{-\lambda t_0} \Rightarrow t_0 = -\frac{\ln(0,8)}{0,006} \simeq 27 \text{ jours}$$

3. Calculer la probabilité qu'un ILS n'ait aucune défaillance sur une période d'un an.
IP(T > 365) = R(365) = $e^{-0,006X365}$ = 0,112

4. Donner une estimation du MTTF et du paramètre λ en considérant que l'on peut modéliser le temps de bon fonctionnement des éléments étudiés par une loi exponentielle.

$R\left(\dfrac{1}{\lambda}\right) = e^{-1} \simeq 0,368$. On peut déterminer graphiquement l'antécédent de 0,368 ce qui donne directement

MTTF = $\dfrac{1}{\lambda} \simeq 140$ d'où : $\lambda \simeq 0,0071$.

Exercice 4

Un parc de 13 photocopieurs a été testé et nous avons observé leurs temps de première défaillance, et ce jusqu'à une période d'un an maximum.

Photocopieurs	1	2	3	4	5	6	7	8	9	10	11	12	13
Temps (jours)	149	360	365**	365**	40	66	15	365**	260	150*	150*	166	275*

Les temps suivis d'une astérisque correspondent à des photocopieurs ayant été retirés de l'étude avant la fin des tests. Ces photocopieurs doivent être considérés comme des éléments suspendus. Les temps suivis de deux astérisques correspondent à des photocopieurs n'ayant eu aucune défaillance.

1. Donner une estimation de R(t) en utilisant la méthode de Kaplan-Meier :

$$R^*(t_i) = \prod_{\substack{j=1 \\ t_j \le t_i}}^{13} \left(1 - \frac{d(t_j)}{v(t_i - \varepsilon)}\right) \Rightarrow ;$$

15	40	66	149	150*	150*	166	260	275*	365**	365**	365**	365**
0,92	0,85	0,77	0,69	0,69	0,69	0,59	0,49	0,49	0,37	0,37	0,37	0,37

2. Donner une estimation du MTTF et du paramètre λ, en considérant que l'on peut modéliser le temps de bon fonctionnement des éléments étudiés par une loi exponentielle. .

Estimation du MTTF $\simeq 360$ et du paramètre $\lambda = \dfrac{1}{MTTF} \simeq 0,003$.

Exercice 5

Un fabricant de voitures affirme que la consommation moyenne de ses voitures est de 4,7 litre de carburant pour 100 km (en ville). Un organisme de défense des consommateurs prélève au hasard un échantillon de n = 100 voitures et observe une consommation moyenne de 4,83 litre pour 100 km avec un écart-type $\sigma = 0{,}3$ litre.

1. Déterminer un intervalle de confiance de niveau 0,99 pour la consommation moyenne m.

n> 30, on peut utiliser une loi normale :

$$u_{1-\frac{\alpha}{2}} = 2{,}57 \Rightarrow I_{1\%} = \left[4{,}83 - 2{,}57\frac{0{,}3}{\sqrt{100}}; 4{,}83 + 2{,}57\frac{0{,}3}{\sqrt{100}}\right] = [4{,}753; 4{,}907]$$

2. Peut-on accuser ce fabricant de publicité mensongère ? OUI : $170 \notin [150{,}29; 165{,}71]$

Exercice 6

On désire savoir si l'âge des candidats au permis de conduire influe sur leur réussite. A cette fin, on les classe en trois catégories d'âge et on construit le tableau ci-dessous à partir des résultats obtenus au premier passage à l'examen par 656 candidats.

Age = a	$18 \le a \le 25$	$26 \le a \le 55$	$55 \le a$	Total
Réussite	153	151	28	332
Echec	185	91	48	324
Total	338	242	76	656

Quelle est la conclusion ?

$$\mathrm{t}_n = \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{\left(x_{ij} - \frac{x_{i\bullet}x_{\bullet j}}{n}\right)^2}{\frac{x_{i\bullet}x_{\bullet j}}{n}} = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{3}\frac{\left(nx_{ij} - x_{i\bullet}x_{\bullet j}\right)^2}{x_{i\bullet}x_{\bullet j}} \Rightarrow t_{656} \simeq 23{,}07$$

Or $(l-1)(c-1) = (2-1)(3-1) = 2$ et $\chi_{2,1-\alpha} \in \{4{,}6;\ 5{,}99;\ 9{,}2;\ 13{,}8\} \Rightarrow \chi^2_{2,1-\alpha} << \mathrm{t}_n, \forall\alpha$ donc il semble que l'âge influe sur la réussite.