Université de Cergy-Pontoise
UFR Economie et Gestion

Année 2011-2012
Caleb Andrianasitera

Analyse Statistique
Examen 1^re^ session 2012
Durée : 2h

Exercice 1 (8 points)

On dispose d'un échantillon de 10 observations extrait d'une loi normale $N(m, \sigma)$ dont les paramètres m, σ sont inconnus. Les observations $(x_1, x_2, \ldots x_{10})$ sont : 2.95, 3, 2.8, 3.15, 3.20, 3.05, 3.1, 3.29, 3.30, 2.9

1. Construire un intervalle de confiance à 95% pour m.
2. Même question pour σ^2
3. Soit le test suivant :

$$\begin{cases} H_0 : m = 3 \\ H_1 : m > 3 \end{cases}$$

   (a) Donner la variable statique Z pour construire la région critique R.
   (b) Déterminer la forme de la région critique pour un risque α. Ce test est-il UPP ?
   (c) Pour un risque $\alpha = 5\%$, définir la règle de décision et conclure.
   (d) Calculer la puissance η de ce test pour $m =$ 3,089

Exercice 2 (12 points)

On considère la population des étudiants de l'Université de Cergy-Pontoise et on désigne par p la proportion des étudiants possédant un iPhone. On tire au hasard et avec remise dans cette population un échantillon de n étudiants. On désigne par X_i $(i = 1, \ldots, n)$ la variable aléatoire de Bernoulli valant 1 si le i^e^ étudiant a un iPhone et 0 sinon.

1. Quelle est la loi de la variable aléatoire $Y = \sum_{i=1}^{n} X_i$? Calculer $E(Y)$ et $V(Y)$.
2. On note par $\overline{X}_n$ la proportion d'étudiants ayant un iPhone dans un échantillon de taille n. Déduire de (1) $E(\overline{X}_n)$ et $V(\overline{X}_n)$.
3. On suppose n grand :
   (a) Enoncer le théorème central limite en choisissant des notations correspondant à l'exercice.

(b) On suppose que $n = 100$ et $p = 0,5$.
   - Calculer la probabilité que $\overline{X}_n$ soit supérieure à $0,525$.
   - Calculer la probabilité que $\overline{X}_n$ soit inférieure à $0,4875$.
   - Déterminer a pour que $P(|\overline{X}_n - 0,5| > a) = 0,10$

4. On suppose maintenant que p est inconnu et on cherche à l'estimer.
   (a) Déterminer un estimateur de p par la méthode du maximum de vraisemblance et préciser ses propriétés.
   (b) On a trouvé dans un échantillon de taille 100 que 49.5% des étudiants ont un iPhone. Donner un intervalle de confiance pour le paramètre p au seuil 5% .

CORRIGE DE L'EXAMEN DU LA SESION 1 2012

1)

$\overline{X} = \dfrac{\sum_{i=1}^{10} X_i}{10}$ *et* $S^2 = \dfrac{\sum_{i=1}^{10}\left(X_i - \overline{X}\right)^2}{9}$ sont des estimateurs sans biais de m et de σ^2

$\dfrac{\overline{X}_n - m}{\dfrac{S}{\sqrt{10}}}$ suit une loi de Student à 9 degrés de liberté.

Etant donné α, on lit dans la table de Student la valeur de t_α, telle que

$P\left[\left|\dfrac{\overline{X} - m}{\dfrac{S}{\sqrt{10}}}\right| < t_\alpha\right] = 1-\alpha$ ce qui s'écrit $P[\overline{X} - t_\alpha \dfrac{S}{\sqrt{10}} < m < \overline{X} + t_\alpha \dfrac{S}{\sqrt{10}}] = 1-\alpha$

d'où l'intervalle de confiance : $IC(m) = [\overline{x} - t_\alpha \dfrac{s}{\sqrt{10}} \quad , \quad \overline{x} + t_\alpha \dfrac{s}{\sqrt{10}}]$

Ici 1-α=95% donc α=0.05 et par suite t_α=2.262

$\overline{x} = \dfrac{\sum_{i=1}^{10} x_i}{10} = 3.074$ et

$s^2 = \dfrac{\left(\sum_{i=1}^{10} x_i^2\right) - 10\left(\overline{x}\right)^2}{9} = \dfrac{99,7416 - 94,49476}{9} = 0,0274 \, soit \; s = 0.1656$

Ce qui donne comme intervalle de confiance :

$IC(m)$=[3.074-0.118 , 3.074+0.118]=[2,956 , 3,192]

2) On sait que $9\dfrac{S^2}{\sigma^2}$ suit une loi de chi deux à 9 degrés de liberté . Etant donné α, on lit dans la table les deux valeurs k_1 et k_2 telles que

$P\left(\chi_9^2 > k_2\right) = \dfrac{\alpha}{2}$ et $P\left(\chi_9^2 < k_1\right) = \dfrac{\alpha}{2}$ d'où $P\left[k_1 < \left(\dfrac{9S^2}{\sigma^2}\right) < k_2\right] = 1-\alpha$ ce qui s'écrit encore :

$P\left[\dfrac{9S^2}{k_2} < \sigma^2 < \dfrac{9S^2}{k_1}\right] = 1-\alpha$ d'où l'intervalle de confiance pour σ^2

$IC(\sigma^2) = \left[\dfrac{9s^2}{k_2} \quad , \quad \dfrac{9s^2}{k_1}\right] = \left[\dfrac{0,2466}{k_2} \quad , \quad \dfrac{0,2466}{k_1}\right]$

Ici α=0.05 donc $P\left(\chi_9^2 > k_2\right) = \frac{\alpha}{2} = 0.025 \Rightarrow k_2 = 19.0$ et

$P\left(\chi_9^2 > k_1\right) = 1 - \frac{\alpha}{2} = 0.975 \Rightarrow k_1 = 2{,}70$ ce qui donne comme intervalle à 95% pour σ^2 $IC(\sigma^2) =$ [0,013 , 0,0913]

3)

a)C'est un test qui porte sur la moyenne d'une variable suivant une loi normale $N(m,\sigma)$ (variance inconnue)donc on va s'appuyer sur la statistique

$Z = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ qui est un estimateur sans biais de m.

la région critique sera définie par la statistique $\overline{X}_n$. qui suit la la loi normale N $\left(3, \frac{\sigma}{\sqrt{n}}\right)$ sous l'hypothèse H_0 mais σ est inconnu. Donc on va remplacer estimer σ^2 par son estimateur sans biais

$S_n^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2}{n-1}$. On sait alors que sous l'hypothèse H_0 la variable

$T_0 = \frac{\overline{X}_n - 3}{\frac{S_n}{\sqrt{10}}}$ suit une loi de Student à 9 ddl.

b)La région critique est de la forme : $\overline{X}_n \geq k$ et k est tel que

$\alpha = P(\overline{X}_n \geq k / H_0) = P\left(T_0 \geq t_\alpha / T_0 \sim St_9\right)$

On trouve dans une table de Student la valeur t_α et on en déduit :

$k = m_0 + t_\alpha \frac{s}{\sqrt{n}}$ et la région critique (R) est définie par $\overline{X}_n \geq m_0 + t_\alpha \frac{s}{\sqrt{n}}$

Cette région critique est indépendante des valeurs de m de l'hypothèse H_1 .
donc ce test est UPP

c) Pour $\alpha = 5\%$ on trouve dans une table de Student la valeur $t_\alpha = 1{,}833$ et on en déduit :

$$k = m_0 + t_\alpha \frac{s}{\sqrt{n}} = 3 + 1{,}833 \frac{0{,}1656}{\sqrt{10}} = 3{,}0953$$

et la région critique (R) est définie par $\overline{X}_n \geq 3{,}0953$

D'où la règle de décision suivante : Si $\overline{x} \geq 3{,}0953$ on rejette *l'hypothèse* $m = 3$

Décision : $\overline{x} = 3{,}074 \notin (R)$ *donc on accepte l'hypothèse* $m = 3$

d) **attention :on a changé m$_1$=4 en m$_1$=3.089**

Pour $m_1 = 3{,}089$

$$\eta = P(\overline{X}_n \geq k / m = 3{,}089) = P\left(T_1 \geq \frac{3{,}0953 - 3{,}089}{\frac{s_n}{\sqrt{10}}} / T_1 \sim St_9 \right)$$

$$= P(T_1 \geq 0{,}120 / T_1 \sim St_9) \approx \frac{0{,}90}{2} = 0{,}45$$

Exercice 2 :

1)Les X_i étant n variables de Bernoulli indépendantes de paramètres p la variable $Y = \sum_{i=1}^{n} X_i$ suit une loi binomiale de paramètre n et p B(n,p)

2)On sait que E(Y)=np et V(Y)=np(1-p)

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{Y}{n} \Rightarrow E(\overline{X}_n) = E\left(\frac{Y}{n}\right) = \frac{E(Y)}{n} = \frac{np}{n} = p$$

$$V(\overline{X}_n) = V\left(\frac{\sum_{i=1}^{n} X_i}{n} \right) = \frac{V(Y)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

3) $\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}}$ converge en loi vers la variable normale centrée réduite U ou encore $\frac{\overline{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}}$ converge en loi vers la variable normale centrée réduite U quand n tend vers +∞

Si n= 100 et p=0,5 $\sqrt{\frac{p(1-p)}{100}} = \frac{0.5}{10} = 0.05$ et $\frac{\overline{X}_n - 0.5}{0.05} \rightarrow$N(0.1)

a) $P\left[\overline{X}_n > 0.525\right] = P\left[\frac{\overline{X}_n - 0.5}{0.05} > 0{,}5\right] = 1 - \Pi(0{,}5) = 1 - 0.6915 = 0.3085$

b)

$$P\left[\overline{X}_n < 0.4875\right] = P\left[\frac{\overline{X}_n - 0.5}{0.05} < -0{,}25\right] = \Pi(-0{,}25)$$

$$= 1 - \Pi(0{,}25) = 1 - 0{,}5987 = 0.4013$$

c)

$$P\left[\left|\overline{X}_n - 0.5\right| > a\right] = P\left[\frac{\left|\overline{X}_n - 0.5\right|}{0.05} > \frac{a}{0.05}\right] = P\left[|U| > \frac{a}{0.05}\right] = 0.10$$

$$\Rightarrow \frac{a}{0.05} \approx 1.6449 \Rightarrow a \approx 0.0822$$

4)

a) $P(X_i = k_i) = p^{k_i}(1-p)^{k_i} \quad pour\ k_i = 0\ ou\ 1$

car chaque variable X_i suit une loi de Bernoulli de paramètre p

La vraisemblance de l'échantillon $L(k_1, k_2,..,k_n)=P(X_1=k_1, X_2=k_2, ,X_n=k_n)$

Les variables étant indépendantes on a :

$L(k_1, k_2,..,k_n)=P(X_1=k_1, X_2=k_2, ,X_n=k_n)=P(X_1=k_1)P(X_2=k_2,)*.. P(X_n=k_n)$

D'où $L(k_1, k_2,..,k_n)=\prod_{i=1}^{n} p^{ki}(1-p)^{1-ki} = p^{\sum_{i=1}^{n} ki}(1-p)^{\sum_{i}(1-ki)} = p^k(1-p)^{n-k} \quad où\ k = \sum_{i=1}^{n} k_i$

L'estimateur de p par la méthode du maximum de vraisemblance s'obtient en résolvant en p l'équation suivante : $\frac{\partial LogL}{\partial p} = 0$

$$LogL(k_1, k_2,..,k_n) = Log(p^k(1-p)^{n-k}) = kLogp + (n-k)Log(1-p)$$

$$\frac{\partial Logl}{\partial p} = \frac{k}{p} - \frac{(n-k)}{(1-p)} = \frac{k(1-p) - p(n-k)}{p(1-p)} = \frac{k - np}{p(1-p)} = 0 \Leftrightarrow p = \frac{k}{n}$$

Ce qui donne comme estimateur de p $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{Y}{n}$

Cet estimateur est sans biais car $E(\overline{X}_n) = p$

Elle est convergente car $E(\overline{X}_n) = p \quad et \quad V(\overline{X}_n) = \frac{p(1-p)}{n} \to 0\ quand\ n \to +\infty$

En notant par X la variable de Bernoulli générique la quantité d'information de Fisher associée à l'échantillon de taille n

$$I(p) = nV\left(\frac{\partial Log\left[p^X(1-p)^{1-X}\right]}{\partial p}\right)$$

$$Logp^X(1-p)^{1-X} = XLogp + (1-X)Log(1-p) \Rightarrow$$

$$\frac{\partial Log\left[p^X(1-p)^{1-X}\right]}{\partial p} = \frac{X}{p} - \frac{(1-X)}{1-p} = \frac{X-p}{p(1-p)} \Rightarrow$$

$$V\left(\frac{\partial Log\left[p^X(1-p)^{1-X}\right]}{\partial p}\right) = V\left(\frac{X-p}{p(1-p)}\right) = \frac{V(X)}{p^2(1-p)^2} = \frac{p(1-p)}{p^2(1-p)^2} = \frac{1}{p(1-p)}$$

Ce qui donne $I(p)=\frac{n}{p(1-p)} \Rightarrow V(\overline{X}_n)=\frac{1}{I(p)}$ donc $\overline{X}_n$ est un estimateur efficace de p

b)Déterminons dans la table de la loi normale la valeur t telle que :

$$P\left(\left|\frac{\overline{X}_n-p}{\sqrt{\frac{p(1-p)}{n}}}\right|<t\right)=0.95 \Rightarrow \Pi(t)=0.975 \Rightarrow t=1.96$$

1ère méthode : on peut majorer le p du dénominateur par ½ donc

$p \in \left(0.495-1.96\sqrt{\frac{1}{400}} \quad , \quad 0.495+1.96\sqrt{\frac{1}{400}}\right)$ Donc p se trouve dans

$IC(p)=$ [0.397 ; 0.593] avec une confiance de 95%

2èmeméthode :on remplace $\sqrt{\frac{p(1-p)}{n}}$ par $\sqrt{\frac{\overline{x}_n(1-\overline{x}_n)}{n}}$ .le calcul donne le même résultat