Abril 2009 N.Lacourly, J. Deride, S. Court

GUIA CONTROL 1 MA34B

1. Sea $X_1, X_2, ..., X_n$ una muestra aleatoria simple de una distribución tal que $I\!P(X_i \in [a,b]) = \theta$. Se define

$$Y_i = \begin{cases} 1 & \text{si } X_i \in [a,b] \\ 0 & \text{en caso contrario} \end{cases}$$

   a) Dé la distribución de Y_i. Deduzca el estimador de máxima verosimilitud $\hat{\theta}$ de θ.
   b) Sean la distribución a priori de θ: $\pi(\theta) = 2(1-\theta)$
      Dé el estimador de Bayes θ^* cuando se usa una función de pérdida cuadrática.
   c) Dé la esperanza y la varianza de los dos estimadores $\hat{\theta}$ y θ^* de θ. Concluye si son consistentes. Compare las dos varianzas. Concluye.
   d) Aplicación numérica: dé las soluciones a las preguntas anteriores con los valores: n=10, X_i: 1.2, 3.5, 2.4, 1.5, 6.3, 2.8, 4.2, 4.5, 3.8, 5.1 y [a,b]=[2,4]. Compare las esperanzas y varianzas si $\theta = 0{,}39$.

2. Sea una variable aleatoria X de función de densidad:

$$f(x) = \frac{1}{a^r \Gamma(r)} e^{-\frac{x}{a}} x^{r-1} \;\; si \;\; x > 0$$

en donde $a > 0$ es desconocido, $r > 0$ es dado, y $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$, $\Gamma(r+1) = r\Gamma(r)$. La función generatriz de momentos de X es igual a: $E(e^{tX}) = \psi(t) = (\frac{1}{1-ta})^r$, para $ta < 1$.

a) Se extrae al azar una muestra de tamaño m. Dé el estimador de máxima verosimilitud $\hat{a}_m$ de a.

b) Se extrae al azar una segunda muestra de tamaño n, lo que permite tener une muestra total de tamaño $m+n$. Utilizando los resultados precedentes, dar el nuevo estimador de máxima verosimilitud $\hat{a}_{m+n}$ de a. Dé la esperanza de $\hat{a}_{m+n}$. (Pueden usar la función $\psi(t)$).

c) Se considera otro estimador $\hat{a}_o$ obtenido como promedio del estimador $\hat{a}_m$ y del estimador de máxima verosimilitud $\hat{a}_n$ de a para la segunda muestra: $\hat{a}_o = \frac{\hat{a}_m + \hat{a}_n}{2}$. Calcule la esperanza y la varianza de $\hat{a}_o$. (Pueden usar la función $\psi(t)$).

d) De las esperanzas y varianzas de los dos estimadores $\hat{a}_{m+n}$ y $\hat{a}_o$ concluye sobre la consistencia de estos estimadores. Compare las varianzas de los dos estimadores $\hat{a}_{m+n}$ y $\hat{a}_o$. Concluye.

3. Sean $X_1, ..., X_n, Y_1, ..., Y_n$ v.a. independientes con $X_i \sim Exp(\frac{1}{\alpha})$ e $Y_j \sim Exp(\frac{1}{\beta})$, con $i = 1, ..., n; j = 1, ..., n$. Se define el parámetro $\theta = (\theta_1, \theta_2)$ por $\theta_1 = \alpha$ y $\theta_2 = \frac{\beta}{\alpha}$.

   (a) Determine los EMV (estimador máximo verosímil) para θ_1 y θ_2
   (b) Encuentre el sesgo y el ECM (error cuadrático medio) de $\hat{\theta}_1$

4. En una fábrica se seleccionan diariamente motores y se inspeccionan hasta encontrar el primer motor defectuoso. Sea $(X_1, \ldots, X_n)$ una m.a. de X distribuida geométricamente con p desconocido.

   a) Determine el estimador de momentos para p.

   b) Determine el estimador máximo verosímil de p.

   c) De los registros de 100 días se obtuvo la siguiente información del número de motores inspeccionados.

Nde motores inspeccionados	1	2	3	4	5
Nde días	8	10	15	25	42

   Estime la probabilidad de que en un día cualquiera se deban inspeccionar más de dos motores para encontrar uno defectuoso.

   NOTA: Si X se distribuye geométrica de parámetro p, su función de probabilidad esta dada por:

$$P(X = x) = p(1-p)^{x-1} \qquad x = 1, 2, \ldots \qquad E(X) = \frac{1}{p}, V(X) = \frac{1}{p^2}$$

5. Sean $X_1, \ldots, X_n$ variables aleatorias *iid* con función de densidad dada por

$$f_X(x) = \begin{cases} (\alpha+1)x^{\alpha} & ; \quad 0 < x < 1 \\ 0 & ; \quad e.o.c. \end{cases}$$

   a) Encuentre el estimador de α por el método de los momentos.

   b) Encuentre el estimador de α por el método de máxima verosimilitud

   c) Evalúe los estimadores usando los siguientes datos.

x	0.1-0.3	0.3-0.6	0.6-0.7	0.7-0.9
Frecuencia	3	1	2	3

6. Sea una muestra aleatoria simple (i.i.d.) $\{y_1, y_2, ..., y_n\}$ de una v.a. discreta Y tal que $Prob(Y = 1) = \theta$, $Prob(Y = 0) = 1 - \theta$, en que $\theta \in (0, 1)$ es desconocido.

   (a) Encuentre el estimador $\hat{\theta}$ de θ por el método de los momentos.

   (b) Obtenga la función de verosimilitud de los valores muestrales.

   (c) Sea la densidad a priori de θ definida por $\pi(\theta) = 2(1-\theta)$ para $\theta \in (0, 1)$. Deduzca la distribución corregida (a posteriori) de Bayes de θ dados los valores muestrales, y muestre que estimador bayesiano puntual de θ para pérdida cuadrática es $\theta_{bayes} = \frac{(1+n\bar{y})}{(n+3)}$.

   (d) Dé la esperanza y la varianza de los dos estimadores de θ ya obtenidos. Estudie insesgamiento y consistencia.

   Considere ahora una muestra aleatoria simple de tamaño 5: $\{x_1 = 2{,}3, x_2 = 3{,}1, x_3 = 1, x_4 = 3{,}7, x_5 = 5\}$ de una v.a. X tal que $Prob(X \in [2, 4]) = \theta$ con $\theta \in (0, 1)$. Considere que la v.a. Y se define de modo que $Y = 1$ si $x \in [2, 4]$, 0 si no.

(e) Suponga que $\theta = 0{,}4$. Aplicando las partes anteriores obtenga el estimador del método de los momentos y el estimador bayesiano puntual de pérdida cuadrática para θ. ¿Cuál estimador presenta en este caso menor error cuadrático medio?

(f) Estime a partir de lo obtenido en (c) la probabilidad de que $\theta \in [0{,}35, 0{,}65]$. ¿Es éste el único intervalo en que se estima que θ pertenece con dicha probabilidad? (Note que para h entero $\Gamma(h) = (h-1)!$).

7. Sea una muestra aleatoria simple $X_1, X_2, \ldots, X_n$ de una distribución exponencial de parámetro θ con densidad $f(x) = \frac{1}{\theta}\, exp\left(-\frac{x}{\theta}\right)\ \forall x \geq 0$.

- Encuentre el estimador de Máxima Verosimilitud $\hat{\theta}$ de θ.
- Deduzca el estimador de Máxima Verosimilitud $\hat{\beta}$ de $e^{-\lambda\theta}$ donde $\lambda > 0$
- Calcule la esperanza y la varianza de $\hat{\beta}$ (se recomienda usar la función generatriz de los momentos de X).
- Se considera una distribución de Poisson de parámetro λ, ie, $Prob(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$, en particular $E(X) = Var(X) = \lambda$, y una distribución a priori para λ correspondiente a $Gamma(2,5)$, ie, $g(\lambda) = \frac{\lambda^{r-1}e^{-\lambda/s}}{s^r\Gamma(r)}$ $\lambda > 0$ con $r = 2$ y $s = 5$, recuerde que con esta distribución tenemos $E(\lambda) = rs$ y $Var(\lambda) = rs^2$. Encuentre el estimador de Bayes de λ para una función de perdida cuadrática. ¿Es asintoticamente insesgado?¿Es consistente?

8. El voltaje X de un determinado circuito se distribuye uniformemente en un intervalo $[\theta, \theta+1]$, donde θ se desconoce. Se obtiene una muestra aleatoria simple del voltaje $\{x_1, x_2, ..., x_n\}$.

(a) Muestre que $\bar{x}$ es un estimador sesgado de θ. Calcule su sesgo. >Es asintóticamente insesgado?.

(b) Obtenga a partir de $\bar{x}$ un estimador $\hat{\theta}$ insesgado de θ. Compare las varianzas de $\bar{x}$ y $\hat{\theta}$.

(c) Deduzca los Errores Cuadráticos Medio de $\bar{x}$ y $\hat{\theta}$. Compare y concluya.

9. La duración de una bombilla sigue una distribución $X \sim \mathcal{N}(\mu, \sigma^2)$ donde $\mu = 1200[hrs]$ y $\sigma^2 = 100[hrs^2]$.

(a) Calcule el tamaño mínimo de una muestra aleatoria simple para que la media empírica $\bar{x}$ no difiera del verdadero valor μ en más de $0{,}7[hrs]$ con una probabilidad de 95 %. (Hint: Use la distribución de $\bar{x}$ y $P(|\bar{x}-\mu| \leq 0{,}7) = 0{,}95$ y $P(|z| \leq 1{,}96) = 0{,}95$ si $z \sim \mathcal{N}(0,1)$).

(b) Se tomo una muestra de 20 bombillas cuyas duraciones son:

1	2	3	4	5	6	7	8	9	10
1188	1178	1210	1195	1203	1202	1200	1190	1191	1196
11	12	13	14	15	16	17	18	19	20
1188	1189	1215	1201	1188	1200	1189	1187	1197	1210

Construya y dé un gráfico de la función de distribución empírica de los valores muestrales.

(c) Para estudiar la dispersión de la distribución teórica, se define los siguientes estadísticos (cuartiles):
• Q_1 el valor de X tal que $P(X \leq Q_1) = 1/4$
• Q_2 el valor de X tal que $P(X \leq Q_2) = 1/2$
• Q_3 el valor de X tal que $P(X \leq Q_3) = 3/4$
Encuentre estos valores en la muestra. Interprete en particular a Q_2.

10. En el estudio de los defectos de una impresora a color, para cada de hoja impresa, se observa tres alternativas:

   - opción 0 si la hoja no tiene defecto
   - opción 1 si la hoja salió con algún defecto
   - opción 2 si la hoja no salió.

   Llamamos $p_i = P(\text{que una hoja tome la opción } i)$ $(i = 0, 1, 2)$.

   Se observa una muestra aleatoria simple de n hojas a imprimir. Se llama y_i al número de veces que salió la opción i en la muestra.

   La función de probabilidad conjunta de (y_0, y_1, y_2) es:

$$P(Y = (y_0, y_1, y_2)|n; p_0, p_1, p_2) = \frac{n!}{y_0!y_1!y_2!} p_0^{y_0} p_1^{y_1} p_2^{y_2} \quad ; \sum_i y_i = n \quad ; \sum_i p_i = 1$$

   (a) Justifique la función de probabilidad utilizando la combinatoria.

   (b) Una muestra de 190 hojas dio los siguientes resultados:

opción	0	1	2
y_i observados	10	68	112

   Obtenga el estimador de Máxima Verosimilitud de los parámetros p_0, p_1 y p_2. (OJO: Observe que $p_0 + p_1 + p_2 = 1$).

   (c) Supongamos ahora que los defectos aumentan con el tiempo de uso de la impresora. Esto se traduce en una variable Z que sigue una distribución exponencial ($f(z) = \lambda e^{-\lambda z}$, $z > 0$) tal que

$$p_0 = P(Z \leq 1), \quad p_1 = P(1 < Z < 2), \quad p_2 = P(Z \geq 2)$$

   Introduciendo estos nuevos elementos en la función de probabilidad estime el parámetro λ. Determine la estimación de λ en base a los datos observados. Interprete.

11. Se desea encontrar un intervalo con 95 % de confianza para el parámetro p de la distribución de Y. Para ello se ha tomado una muestra aleatoria simple en que se mide la cantidad de intentos de 200 usuarios que logran conectarse a Internet. Se obtiene que $\bar{Y} = 3{,}5$.

   - Con ayuda del Teorema del límite central aproxime un intervalo de confianza de 95 % del tipo $[-a, a]$ para $Z = \frac{\sqrt{n}(p\bar{Y}-1)}{\sqrt{1-p}}$.
   - Deduzca la aproximación de un intervalo de confianza al 95 % para p.

Indicaciones

- Desigualdad de Chebyshev: Si $E(X) < \infty$ y $\sigma^2 = Var(X) < \infty$ entonces $prob(|X - \mu| > k\sigma) \leq \frac{1}{k^2} \quad \forall k > 0$.
- $\bar(X)$ y $\tilde{S}^2$ son independientes si $X \sim N(\mu, \sigma^2)$.
- Teorema del límite central: Si $X \sim F$ y $n \to \infty$ entonces $\bar{X} \sim N(E(X), \frac{Var(X)}{n})$ cuando $E(X) < \infty$ y $Var(X) < \infty$.

- Si $X \sim Geom(p)$ entonces $prob(X=k)=(1-p)^{k-1}p$, $E(X)=\frac{1}{p}$ y $Var(X)=\frac{1-p}{p^2}$.

12. Para cada pregunta responda Verdadero o Falso, **justificando** su respuesta en cada caso:

    - En un intervalo de confianza de largo mínimo para la media μ de una distribución Normal, si aumento el nivel de confianza, entonces aumenta el largo del intervalo.
    - En una encuesta a nivel nacional sobre una muestra de 1200 personas, el 65 % consideró que la cumbre APEC no fue beneficiosa para el país. Con un nivel de confianza del 95 % estos resultados permiten concluir que la diferencia absoluta entre la proporción estimada y la proporción poblacional es menor al 3 %. **Hint** Utilice el Teorema Central del Limite definiendo una V.A apropiada.

13. La fábrica LUPINA S.A. produce ampolletas de uso residencial. El jefe de la sección de produción lo ha contratado a Ud. para que lo ayude a estimar la tasa de vida útil de su producto estrella, la ampolleta LOBO. De acuerdo a sus conocimientos y experiencia, Usted sabe que la vida útil de una ampolleta X (en años) se puede modelar según la distribución exponencial siguiente:

    $$f(x|\lambda)=\lambda e^{-\lambda x} \; con \; x>0$$

    El jefe ha tomado previamente una muestra de 30 ampolletas y encuentra que la duración promedio de ellas es de 5 años. Se define la variable

    $$Y=\sum_{i=1}^{30} X_i$$

    - Proponga un intervalo de confianza al 95 % para la tasa de vida útil λ (**HINT:**Use el Teorema central del Límite para la variable Y, resolviendo la ecuación cuadrática asociada).
    - Muestre que la variable $T=2\lambda Y$ sigue una distribución Chi-cuadrado con $2n$ grados de libertad.(**HINT:** Utilice la función generatriz de los momentos).
    - Encuentre L y U tales que $Prob(U<T<L)=0{,}95$ y luego deduzca un intervalo de confianza al 95 % para λ usando la información del punto anterior.