Guía de Problemas # 7.

Andrés Iturriaga Jofré
Depto. Ingeniería Matemática
Universidad de Chile
Santiago, Chile
aiturriaga@dim.uchile.cl

Héctor Olivero Quinteros
Depto. Ingeniería Matemática
Universidad de Chile
Santiago, Chile
holivero@dim.uchile.cl

27 de Junio de 2009

Resumen

Esta guía de problemas trata acerca de "test de hipótesis". Los tópicos específicos son: Test de Neyman-Pearson, Test de Razón de Verosimilitudes y Test de Chi-Cuadrado.

Problema 1.-
Sean $X_1, \ldots, X_n$ v.a. i.i.d. con función de distribución $F(x) = \mathbb{P}(X \leq x) = x^\theta$, para $x \in (0,1)$ y $\Omega = \{\theta_0, \theta_1\}$.

(a) Determine la función de densidad de probabilidad de la variable aleatoria X.

(b) Para un nivel de significación α, demuestre que la región crítica del test de Neyman-Pearson para $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, con $\theta_0 < \theta_1$, está dada por:

$$R = \{x = (x_1, \ldots, x_n) : -\sum_{i=1}^{n} \ln(x_i) \leq k_\alpha\}, \tag{1}$$

donde k_α es una constante a determinar.

(c) Muestre que las v.a. $Y_i = -\ln(X_i)$, $i = 1, \ldots, n$, son i.i.d. con ley exponencial de parámetro θ. ¿Qué puede decir de la distribución bajo H_0 de $-\sum_{i=1}^{n} \ln(X_i)$?

(d) Para una muestra de tamaño 10 y un nivel de significación $\alpha = 5\,\%$, determine la constante k_α usando las tablas estadísticas. Considere que $\theta_0 = 1$ y $\theta_1 = 2$.
Indicación: Si $-\ln(T_i) \sim \exp(1)$ $i = 1, \ldots, n$, independientes entre si, entonces

$$-2\sum_{i=1}^{n} \ln(T_i) \sim \chi^2_{2n}. \tag{2}$$

(e) Suponga que se cuenta con la siguiente muestra:

Obs.	1	2	3	4	5	6	7	8	9	10
X	0.66	0.26	0.57	0.23	0.40	0.80	0.20	0.40	0.45	0.97
$\ln(X)$	-0.42	-1.33	-0.57	-1.49	-0.91	-0.23	-1.59	-0.91	-0.79	-0.03

Tabla 1: Muestra de Datos.

Usando lo anterior, decida el siguiente test:

$$H_0 : \theta_0 = 1$$
$$H_1 : \theta_1 = 2$$

(f) Afirme o refute la siguiente afirmación: `el p-valor del test anterior es menor que` 0.05. Fundamente su respuesta.

(g) Muestre el error de tipo II y calcúlelo aproximadamente.
<u>Indicación</u>: Use que bajo H_1, $-4\sum_{i=1}^{n} \ln(X_i) \sim \chi^2_{2n}$.

<u>Problema</u> 2.-
Sea X una v.a. con valores en $[0,1]$ y densidad $p_\theta(x)$, $\theta \in \Theta = \{\theta_0, \theta_1\}$. Se sabe que

$$\begin{aligned} p_{\theta_0}(x) &= \mathbb{1}_{[0,1]}(x), \\ p_{\theta_1}(x) &= 4x^3 \mathbb{1}_{[0,1]}(x) \end{aligned}$$

(a) Desarrolle el test de Neyman-Pearson de nivel $\alpha = 0{,}1$ para

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta = \theta_1$$

en base a una muestra de tamaño $n = 1$.

(b) Calcule la potencia del test resultante.

(c) Decida el test usando que la muestra consiste en el valor $x_1 = 0.92$.

<u>Problema</u> 3.-
Sean $X_1, \ldots, X_n$ una M.A.S. de una distribución normal de media μ conocida y varianza σ^2 desconocida. Considere el siguiente test de hipótesis:

$$H_0 : \sigma^2 = 2$$
$$H_1 : \sigma^2 = 3$$

(a) Use Neyman-Pearson para demostrar que entre todas las reglas de decisión tales que la probabilidad de error de tipo I es menor o igual a 0.05, la probabilidad de error de tipo II es minimizada cuando consideramos la región de rechazo dada por

$$R = \{x = (x_1, \ldots, x_n) : \sum_{i=1}^{n} (x_i - \mu)^2 \geq k_\alpha\} \tag{3}$$

(b) Determine k_α cuando $n = 8$.
<u>Indicación</u>: Use que si $Z \to N(0,1)$, entonces $Z^2 \to \chi^2_1$.

(c) Suponga que $\mu = 1.5$ y decida el test, usando la siguiente muestra aleatoria simple:

Obs.	1	2	3	4	5	6	7	8
X	-2.1	-1.7	2.3	0.2	-0.6	1.8	1.3	0.1

Tabla 2: Muestra de Datos.

Problema 4.-
Sea X una variable aleatoria cuya función de probabilidad puntual bajo H_0 y H_1 está dado por:

X	1	2	3	4	5	6	7
$f(X\|H_0)$	0.01	0.01	0.01	0.01	0.01	0.01	0.94
$f(X\|H_1)$	0.06	0.05	0.04	0.03	0.02	0.01	0.79

Tabla 3: Funciones de Probabilidad Puntual.

(a) Use el lema de Neyman-Pearson para encontrar la forma de la región crítica que permita decidir el test H_0 versus H_1.

(b) Explicite la región de rechazo imponiendo que el test obtenido tenga tamaño $\alpha = 0,04$. ¿El test obtenido es más potente? Argumente.

(c) Calcule la probabilidad de error de tipo II para este test.

(d) Suponiendo que se cuenta con una realización de la variable aleatoria X, cuyo valor es igual a $x = 3$. Decida el test.

Problema 5.-
La Asociación de Consumidores Organizados (ASCO) ha recibido reclamos respecto de un detergente líquido cuyo contenido real, según los reclamantes, no corresponde a lo declarado en el envase. Los consumidores argumentan que el contenido real es menos de $800cc$ mientras que la compañía indica que es más de 1 litro, porque ellos reconocen que puede haber pequeñas pérdidas por evaporación y por ello envasan más de un litro.
Se realiza una investigación en la que se mide el contenido de n envases seleccionados al azar en diversos supermercados, obteniéndose los valores $x_1, \ldots, x_n$ en litros. Suponemos que los datos observados son realizaciones de las variables aleatorias $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}(\theta, \sigma_0^2)$, donde σ_0 es conocido y toma el valor 0.2.
Usted actúa como asesor del equipo legal de ASCO y propone testear

$$H_0 : \theta < \theta_0$$
$$H_1 : \theta > \theta_1$$

donde $\theta_0 < \theta_1$ son los valores conocidos (en nuestro caso $\theta_0 = 0.8$ y $\theta_1 = 1$). Para ello desarrollará el T.R.V. (test de razón de verosimilitud) de nivel α y lo aplicará a los datos recogidos en el estudio.

(a) Indique cuáles son los conjuntos Ω_0 y Ω_1 correspondientes a H_0 y H_1.

(b) Muestre que $\sup_{\theta\in\Omega_0} f(x|\theta) = f(x|\hat{\theta}_0)$, donde $\hat{\theta}_0 = \bar{x}\mathbb{1}_{\{\bar{x}\leq\theta_0\}} + \theta_0\mathbb{1}_{\{\bar{x}>\theta_0\}}$.

(c) Muestre que $\sup_{\theta\in\Omega_1} f(x|\theta) = f(x|\hat{\theta}_1)$, donde $\hat{\theta}_1 = \bar{x}\mathbb{1}_{\{\bar{x}>\theta_1\}} + \theta_1\mathbb{1}_{\{\bar{x}\leq\theta_1\}}$.

(d) Verifique que el estadístico $\lambda(x)$ (del T.R.V.) es función creciente de $h(\bar{x}) = (\bar{x} - \hat{\theta}_0)^2 - (\bar{x} - \hat{\theta}_1)^2$ y compruebe que $h(\bar{x}) = (\bar{x} - \theta_0)^2\mathbb{1}_{\{\bar{x}>\theta_0\}} - (\bar{x} - \theta_1)^2\mathbb{1}_{\{\bar{x}<\theta_1\}}$.

(e) Verifique que $h(\bar{x})$ es función creciente de $\bar{x}$ y concluya que la región crítica del T.R.V. tiene la forma $R = \{x = (x_1, \ldots, x_n)|\bar{x} \geq k_\alpha\}$ donde k_α se determina en función del nivel de significación α del test.

(f) Compruebe que $\mathbb{P}_\theta(R) = \mathbb{P}_\theta(\bar{X} \geq k_\alpha) = 1 - \Phi\left(\frac{k_\alpha - \theta}{\sigma_0/\sqrt{n}}\right)$ y obtenga que

$$\sup_{\theta\in\Omega_0} \mathbb{P}_\theta(W) = 1 - \Phi\left(\frac{k_\alpha - \theta_0}{\sigma_0/\sqrt{n}}\right),$$

donde Φ es la función de distribución de la $\mathcal{N}(0, 1)$.

(g) Calcule k_α suponiendo que el test tiene nivel α y aplíquelo en las siguientes condiciones: $\alpha = 0.05$, $n = 49$ y $\bar{x} = 0.887$.

(h) Si se rechaza la hipótesis de los consumidores H_0, ¿diría usted que la compañia tiene la razón?

Problema 6.-
Sean $X_1, \ldots, X_n$ una muestra aleatoria simple proveniente de una población cuya función de densidad de probabilidad depende de un parámetro desconocido $\theta > -1$. Suponga que la función de densidad de X_i, para $i = 1, \ldots, n$, está dada por

$$(1+\theta)x^\theta \mathbb{1}_{(0,1)}(x). \tag{4}$$

Desarrolle el test de razón de verosimilitud para

$$\begin{aligned} H_0 : \theta &\leq 1 \\ H_1 : \theta &> 1, \end{aligned}$$

en base a una muestra de tamaño $n > 1$. Muestre que la región crítica de este test está basada en el estadístico $-\sum_{i=1}^n \ln(X_i)$.

Problema 7.-
Estudios relacionados con el comportamiento de ciertos bichos indican que éstos tienden a organizarse al azar, linealmente, en un intervalo de longitud θ, a la derecha de un punto donde se ubica una feromona. Existe una controversia sobre el valor de θ entre dos equipos de científicos: en el laboratorio A se sostiene que $\theta \leq 1$ mientras que en el laboratorio B se sostiene lo contrario, es decir, $\theta > 1$. Para resolver la controversia se diseña un experimento que consiste en medir las posiciones $X_1, \ldots, X_n$ de n bichos, con respecto a la feromona.

(a) Defina el modelo paramétrico correspondiente, indicando los supuestos que están implícitos.

(b) Muestre que el test de razón de verosimilitud de nivel α para $H_0 : \theta \leq 1$ vs $H_1 : \theta > 1$ tiene una región de rechazo dada por

$$R = \{x \in \mathcal{X} | \widehat{\theta}_n(x) \geq k_\alpha\}, \tag{5}$$

donde $\widehat{\theta}_n(x) = \text{máx}\{x_1, \ldots, x_n\}$ y k_α se calcula imponiendo que el test tenga un nivel de significación α.

(c) Compruebe que $k_\alpha = (1-\alpha)^{1/n}$.

Problema 8 (Test de Comparación de Medias)**.-**
Sean $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ dos muestras aleatorias simples de los modelos $\mathcal{N}(\mu, \sigma^2)$ y $\mathcal{N}(\nu, \tau^2)$ respectivamente, con $n \neq m$. Desarrolle el Test de Razón de Verosimilitud de nivel α para

$$H_0 : \mu = \nu \quad \text{versus} \quad H_1 : \mu \neq \nu, \tag{6}$$

donde σ^2 y τ^2 son valores conocidos.

Problema 9 (Problema del Dado Cargado)**.-**
Un dado se lanza 102 veces. Los resultados se muestran en la tabla siguiente:

Cara obtenida	1	2	3	4	5	6
Frecuencia	12	11	22	20	16	21

Tabla 4: Datos.

Testee si el dado está cargado. Escriba la hipótesis nula y decida usando la aproximación chi cuadrado, para un nivel de significación del 5 %.

Problema 10 (Test de Ajuste para una Distribución Discreta)**.-**
Los datos siguientes muestran las frecuencias de conteo para 400 observaciones acerca del número de colonias bacterianas por campo en un microscopio utilizando muestras de una capa de leche.

Número por campo	0	1	2	3	4	5	6	7	8	9	>9
Frecuecia	56	104	80	62	42	27	9	9	5	3	3

Tabla 5: Datos.

Testee la hipótesis de que los datos provengan de una variable aleatoria de Poisson, para un nivel de significación del 5 %.

Problema 11 (Test de Ajuste para una Distribución Continua)**.-**
Testee que los datos de la tabla siguiente provienen de una variable aleatoria $\mathcal{N}(0,1)$:

-2.46	-2.11	-1.23	-0.99	-0.42
-0.39	-0.21	-0.15	-0.1	-0.07
-0.02	0.27	0.40	0.42	0.44
0.70	0.81	0.88	1.07	1.39
1.40	1.47	1.62	1.64	1.76

Tabla 6: Muestra.

Problema 12 (Test de Independencia)**.-**
Se quiere probar si hay una relación entre la aprobación del curso MA3403 y el sexo de los alumnos. Para esto se cuenta con la siguiente tabla de datos:

	No aprueba	Aprueba	Total
Hombre	45	34	79
Mujer	9	7	16
Total	54	41	95

Tabla 7: Datos.

Analice la independencia entre las variables “sexo” y “aprobación”, para un nivel de significación del 5 %.