UNIVERSIDAD CARLOS III DE MADRID
Ingeniería Técnica en Informática de Gestión
22 de Septiembre de 2008
ESTADÍSTICA
Duración: 2 horas

Nota: Utiliza siempre $\alpha = 0,05$, salvo que se indique lo contrario. Justifica todas tus respuestas. Responde cada pregunta en hojas diferentes.

1. Decir si son verdaderas o falsas las siguientes cuestiones, razonando la respuesta
 - a) El 30 % de los accesos al servidor de una empresa tiene como objetivo la descarga de documentos. El 10 % de este tipo de descargas resulta fallida. A partir de esta información podemos asegurar que el porcentaje total de conexiones fallidas al servidor independientemente de su objetivo es $0,3 \times 0,1 = 0,03$.
 - b) Después de analizar todos los registros de fallos del servidor anterior vemos que los únicos errores de transmisión observados se han producido durante la descarga de documentos. Por tanto el porcentaje total de conexiones fallidas al servidor independientemente de su objetivo es del 10 %.
 - c) Seguimos con el servidor anterior. Cuando un usuario que intenta descargarse un documento obtiene un error, lo intenta de nuevo. En cada reintento, la probabilidad de éxito es la misma que en el primer intento. En esta condiciones, el número de intentos hasta que se consigue descargar el documento es una variable aleatoria de Poisson.
 - d) En otro servidor de esa misma empresa se observa que también el 10 % de las descargas resultan fallidas. Sin embargo, en operaciones distintas a las descargas se observan fallos el 5 % de las veces. Si a la hora de descargarse documentos los usuarios usan indistintamente un servidor u otro, el porcentaje total de descargas fallidas será del 9.25 %.

SOLUCIÓN:

- a) Falso. Si llamamos D =acceso para descarga de ficheros, tenemos que $P(D) = 0,3$. Si llamamos F =acceso fallido, tenemos que $P(F|D) = 0,10$. Por tanto

$$P(F) = P(F|D)P(D) + P(F|\bar{D})(1 - P(D)) = 0,1 \times 0,3 + P(F|\bar{D}) \times 0,7.$$

Como no conocemos $P(F|\bar{D})$ no podemos calcular $P(F)$.

- b) Falso. Lo que se afirma ahora es que $P(F|\bar{D}) = 0$. Por tanto.

$$P(F) = P(F|D)P(D) + P(F|\bar{D})(1 - P(D)) = 0,1 \times 0,3 + 0 \times 0,7 = 0,03.$$

- c) Falso. En cada intento, el tener éxito es una variable aleatoria de Bernoulli de probabilidad $p = 0,9$. Por tanto Y =número de intentos hasta que se observa el primer éxito es una variable aleatoria Geométrica.
- d) Falso. Si llamamos S_1 al suceso: utilizar el servidor 1, y S_2 al suceso: utilizar el servidor 2, tendremos que $P(S_1) = P(S_2) = 0,5$. Si llamamos F_D al suceso: fallo al descargarse un documento, se tiene que $P(F_D|S_1) = P(F_D|S_2) = 0,10$. Por tanto $P(F_D) = P(F_D|S_1)P(S_1) + P(F_D|S_2)P(S_2) = 0,10$.

2. La empresa B2C ofrece servicios de búsquedas y consultas de información en internet a sus usuarios. El tiempo x en horas que navega cada usuario puede modelizarse como una variable aleatoria de función de densidad

$$f(x) = K2x,$$

definida entre 0 y un tiempo máximo M . Se pide:

- Encuentra la relación que debe existir entre K y M para que $f(x)$ sea una función de densidad.
- Sabiendo que el tiempo medio de utilización del servicio es $4/3$ horas. ¿Cuál es el tiempo máximo de conexión M ?
- ¿Es cierto que menos del 25% de los internautas navega más de $4/3$ horas? Justifica numéricamente la respuesta.

SOLUCIÓN:

- Para que $f(x)$ sea función de densidad se debe cumplir que

$$\int_0^M f(x)dx = 1.$$

Operando se tiene

$$\int_0^M f(x) = \int_0^M K2xdx = K2 \left[\frac{x^2}{2} \right]_0^M = KM^2 = 1 \Rightarrow K = 1/M^2.$$

- Usando que $E(x) = 4/3$ se tiene que

$$\int_0^M xf(x)dx = \frac{4}{3}.$$

Operando, y usando que $K = 1/M^2$ se tiene

$$\int_0^M \frac{2x^2}{M^2}dx = \frac{2}{M^2} \left[\frac{x^3}{3} \right]_0^M = \frac{2}{3}M = \frac{4}{3} \Rightarrow M = 2 \text{ horas}$$

- La probabilidad de navegar más de $4/3$ horas es

$$\int_{4/3}^2 \frac{1}{4}2xdx = \frac{1}{2} \left[\frac{x^2}{2} \right]_{4/3}^2 = 0,56 > 0,25$$

por lo que la afirmación no es cierta

- El tiempo T de espera para ser atendido telefónicamente en la Campaña de la Renta"de la Agencia Tributaria se modeliza según un modelo exponencial de media 2 minutos. Se pide:

- Calcula la probabilidad de tener que esperar más de 1 minuto
- Calcula la probabilidad de que, si llevamos esperando 1 minuto, tengamos que esperar al menos 3 minutos más.
- Después de atenderse a 30 personas, calcula la probabilidad de que el tiempo total de espera de las 30 personas sea menor a 40 minutos.

SOLUCIÓN:

- El tiempo de espera es $T \sim Exp(\lambda = 1/2)$

$$P(T > 1) = e^{-\lambda \times 1} = e^{-1/2} = 0,6065.$$

- Nos piden $P(T > 4|T > 1)$. Al ser la exponencial un modelo sin memoria, se tiene que

$$P(T > 4|T > 1) = P(T > 3) = e^{-3/2} = 0,223.$$

- c) El tiempo total será $T = \sum_{i=1}^{30} t_i$ y por el teorema central del límite será, aproximadamente, $T \sim N(E(T), \text{Var}(T))$. Como $E(t_i) = 2$, $\text{Var}(t_i) = 1/\lambda^2 = 4$

$$E(T) = \sum_{i=1}^{30} E(t_i) = 30 \times 2 = 60,$$
$$\text{Var}(T) = \sum_{i=1}^{30} \text{Var}(t_i) = 30 \times 4 = 120$$

y por tanto $T \sim N(60, 120)$. Por tanto

$$P(T < 40) = P\left(z < \frac{40 - 60}{\sqrt{120}}\right) = P(z < -1,83) = 0,034$$

Soluciones del Examen A

1. Elige la respuesta correcta

Un contraste de hipótesis bilateral, $H_0 : \mu = \mu_0$ y $H_1 : \mu \neq \mu_0$, con nivel de significación 5% utilizando una muestra de tamaño 50^2 con media $\bar{x} = 30$ y cuasivarianza $\hat{\sigma}^2 = 15^2$ es equivalente a comprobar que μ_0 pertenezca a un intervalo de confianza:

- (A) IC95%: $[30 - 0.3z_{0.025}, 30 + 0.3z_{0.025}]$
- (B) IC95%: $[30 - 0.3z_{0.05}, 30 + 0.3z_{0.05}]$
- (C) IC95%: $[30 - 0.006z_{0.025}, 30 + 0.006z_{0.025}]$
- (D) IC5%: $[30 - 0.3z_{0.025}, 30 + 0.3z_{0.025}]$

2. Elige la respuesta correcta

Un intervalo de confianza al 95% para la media de una población con varianza σ^2 conocida construido a partir de una muestra de $n_1 = 200$ observaciones tiene amplitud 20. Si tomo una nueva muestra de tamaño $n_2 = 800$ y construyo un intervalo de confianza al 95% para la media, ¿cuál será la amplitud del nuevo intervalo?

- (A) 80
- (B) 5
- (C) 10
- (D) 40

3. Verdadero/Falso

¿Cuáles de las siguientes conclusiones son posibles tras realizar un contraste de hipótesis?

- True Rechazar la hipótesis nula.
- False Rechazar la hipótesis nula y rechazar la hipótesis alternativa.
- True Falta de evidencia para rechazar la hipótesis nula.
- (A) t_{10}
- (B) $N(0, 1)$
- (C) t_9
- (D) Ninguna de estas.

- True Rechazar la hipótesis nula.
- False Rechazar la hipótesis nula y rechazar la hipótesis alternativa.
- True Falta de evidencia para rechazar la hipótesis nula.

4. Elige la respuesta correcta

Hemos rechazado la hipótesis nula de un contraste de hipótesis con nivel de significación $\alpha = 0.05$, ¿cómo será su p-valor?

- (A) Igual a 0.05
- (B) Mayor que $1 - 0.05$
- (C) Menor que 0.05
- (D) Igual a $1 - 0.05$
- (E) Mayor que 0.05

5. Elige la respuesta correcta

Se quiere comprobar que en una población exponencial la media $\mu = 1/\lambda$ sea igual a 4. Se construye el contraste de hipótesis $H_0 : \mu = 4$ y $H_1 : \mu \neq 4$. Se toman $n = 10$ elementos y se calcula el estadístico $\bar{X}\sqrt{n}/\hat{\sigma}$ ¿Qué distribución tendrá?

- (A) t_{10}
- (B) $N(0, 1)$
- (C) t_9
- (D) Ninguna de estas.