Pontificia Universidad Católica de Chile
Facultad de Matemáticas
Departamento de Estadística
Primer semestre de 2019
Ayudante: Hernán Robledo (harobledo@uc.cl)

Inferencia Estadística / Métodos Estadísticos - EYP2114/EYP2405
Ayudantía 1

1. Determine el Modelo Estadístico para cada ejemplo, determine si corresponde a la clase Paramétrica o No-Paramétrica, y finalmente concluya determinando si el modelo es Identificado o no.

   a) Se obtienen 10 muestras del peso de una población de hombres. Asumamos que el peso de cada persona distribuye Normal(μ, σ) y las muestras son independientes entre sí.

   b) Se desea investigar si realmente existe una diferencia significativa entre el los costos entregados por los medidores de luz antiguos versus los medidores 'inteligentes'. Se toma una muestra aleatoria de n casas y se recogen datos correspondientes a la diferencia de consumo entre ambos períodos. Asumamos que esta diferencia de consumo distribuye t-student(η).

   c) Se tiene una muestra aleatoria $(x_1, x_2, ..., x_n)$, en que cada x_i una distribuye Q, donde Q es una distribución de probabilidad.

   d) Se realiza un experimento en el cual se toma una muestra aleatoria $(x_1, x_2, ..., x_n)$, donde cada Variable distribuye independientemente Normal($\alpha + \beta$,1), con α y β números reales positivos.

2. Reducción de Datos

   a) Sean $X_1, X_2, ..., X_n$ una muestra proveniente de una distribución Normal(μ, σ^2), con σ^2 conocido. Defina $T(X) = \frac{\sum_{i=1}^{n} X_i}{n}$. Determine por definición si $T(X)$ es un estadístico suficiente para μ.

b) Sean $X_1, X_2, ..., X_n$ una muestra proveniente de una distribución $\exp(\lambda)$, y sea $T(X) = \sum_{i=1}^n X_i$. Determine por definición si $T(X)$ es un estadístico suficiente para λ.

SOLUCIÓN: *Se nos pide usar la definición para resolver este problema, la cual corresponde a determinar si la razón*

$$\frac{p(x|\theta)}{q(t|\theta)}$$

es constante como función del parámetro desconocido θ. *En este caso el parámetro corresponde a* λ, *y* $p(x|\theta)$ *corresponde a la distribución de probabilidad de la muestra, esto es:*

$$p(x|\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \exp\Big(- \lambda \sum_{i=1}^{n} x_i \Big)$$

Por otra parte, $q(t|\lambda)$ *corresponde a la distribución de probabilidad del estadístico. En este caso, el estadístico* $T(X) = \sum_{i=1}^n X_i$ *corresponde a una suma de variables aleatorias exponenciales de parámetro* λ, *la cual distribuye Gamma*(n, λ). *De esta forma:*

$$q(t|\lambda) = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t}$$

Entonces, la razón queda:

$$\frac{p(x|\theta)}{q(t|\theta)} = \frac{\lambda^n \exp\Big(- \lambda \sum_{i=1}^n x_i \Big)}{\frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t}} = \frac{\Gamma(n)}{t^{n-1}}$$

Este último término no depende de λ, *así que concluimos que el estadístico es suficiente.* □

c) Sean $X_1, X_2, ..., X_n$ una muestra proveniente de una distribución $\exp(\theta)$. Determine si $T(X) = \text{mín}_i(X_1, X_2, ...X_n)$ es un estadístico suficiente para θ.

SOLUCIÓN: *Bajo la misma idea de desarrollo del ejercicio anterior, tenemos que* $p(x|\lambda)$ *es la misma función, pero en este caso, el estadístico* $T(X) = \text{mín}_i(X_1, X_2, ...X_n)$ *distribuye exp*$(n\lambda)$, *de esta forma:*

$$\frac{p(x|\theta)}{q(t|\theta)} = \frac{\lambda^n \exp\Big(- \lambda \sum_{i=1}^n x_i \Big)}{n\lambda \exp -n\lambda t} = \frac{1}{n} \lambda^{n-1} \exp\Big(- \lambda \sum_{i=1}^{n} x_i + n\lambda t \Big)$$

Y notamos que no podemos reducir esta expresión para que no dependa de λ. *Así, concluimos que el estadístico* $T(X) = \text{mín}_i(X_1, X_2, ...X_n)$ *no es suficiente para el parámetro* λ, *lo cual se puede interpretar como que el valor mínimo de la muestra no dice nada acerca del valor de* λ, *ni contiene información acerca de este.* □

d) Sean $X_1, X_2, ..., X_n$ una muestra proveniente de una distribución Poisson(λ). Determine vía definición y mediante el Teorema de Factorización si $T(X) = \sum_{i=1}^{n} X_i$ es un estadístico suficiente.

SOLUCIÓN: *Repita el procedimiento de los problemas anteriores sabiendo que* $T(X) = \sum_{i=1}^{n} X_i$ *distribuye Poisson(*$n\lambda$*).*

e) Sea $X_1, X_2, ..., X_n$ una muestra proveniente de una distribución Uniforme$(0, \theta)$. Encuentre un estadístico suficiente para θ.

f) Sea $X_1, X_2, ..., X_n$ una muestra de una población con densidad $f(x|\theta)$ dada por:

$$f(x|\theta) = \frac{1}{\sigma} \exp\{-(\frac{x-\mu}{\sigma})\}$$

donde $x \geqslant \mu, \mu \in \mathbb{R}, \sigma > 0$.

- Muestre que $T(X) = \text{mín}_i(X_1, X_2, ..., X_n)$ es suficiente para μ cuando σ es conocido.
  SOLUCIÓN: *Descartamos de inmediato el uso de la definición para determinar si el estadístico es suficiente o no , ya que deconocemos la distribución de esta función. Utilicemos teorema de factorización:*

$$f(x|\theta) = \prod_{i=1}^{n} \frac{1}{\sigma} \exp\{-(\frac{x_i - \mu}{\sigma})\} \, \mathbb{1}_{x \geqslant \mu}(x_i) = \frac{1}{\sigma^n} \exp\{-\frac{1}{\sigma}(\sum_{i=1}^{n} x_i - n\mu)\} \prod_{i=1}^{n} \mathbb{1}_{[\mu,\infty]}(x_i)$$

Donde la función $\mathbb{1}(\cdot)$ *corresponde a la función indicatriz. Siempre que el soporte de nuestra distribución dependa de un parámetro debemos considerar usar la función indicatriz, ya que si no no podremos obtener el estadístico suficiente. Podemos escribir equivalentemente la última ecucación como:*

$$f(x|\mu) = \frac{1}{\sigma^n} \exp\{-\frac{1}{\sigma}(\sum_{i=1}^{n} x_i - n\mu)\} \mathbb{1}_{[\mu,\infty]}(\text{mín}_i \, x_i)$$

Finalmente podemos reordenar la función para notar identificar las funciones $g(T(X)|\theta)h(x)$*:*

$$f(x|\mu) = \mathbb{1}_{[\mu,\infty]}(\text{mín}_i \, x_i) \exp \frac{n\mu}{\sigma} \cdot \frac{1}{\sigma^n} \exp -\frac{\sum_{i=1}^{n} x_i}{\sigma}$$

Donde tenemos que

$$g(T(X)|\mu) = \mathbb{1}_{[\mu,\infty]}(\text{mín}_i \, x_i) \exp \frac{n\mu}{\sigma}$$

$$h(x) = \frac{1}{\sigma^n} \exp -\frac{\sum_{i=1}^{n} x_i}{\sigma}$$

Por lo tanto, concluimos que $T(X) = \text{mín}_i(X_1, X_2, ..., X_n)$ *es suficiente para* μ.

- Encuentre un estadístico suficiente para σ cuando μ es conocido.
SOLUCIÓN: *A partir del desarrollo anterior, podemos reordenar la función $f(x|\sigma)$ para notar el estadístico suficiente para σ:*

$$g(T(X)|\sigma) = \frac{1}{\sigma^n} \exp - \frac{\sum_{i=1}^n x_i}{\sigma} \exp \frac{n\mu}{\sigma}$$

$$h(x) = \mathbb{1}_{[\mu,\infty]}(\min_i x_i)$$

Nótese que el criterio para reordenar la función fue que ahora necesitábamos que todo lo que dependiera de σ debía estar dentro de nuestra función $g(T(X)|\sigma)$, y lo que no, fuera. Así concluimos entonces que $T(X) = \sum_{i=1}^n X_i$ es suficiente para σ.

- Determine un estadístico de dos dimensiones para $\theta = (\mu, \sigma)$ SOLUCIÓN: *Reordenamos nuevamente la función $g(T(X)|\mu, \sigma)$:*

$$g(T_1(X), T_2(X)|\mu, \sigma) = \frac{1}{\sigma^n} \exp\{-\frac{1}{\sigma}(\sum_{i=1}^n x_i - n\mu)\} \mathbb{1}_{[\mu,\infty]}(\min_i x_i)$$

$$h(x) = 1$$

En este caso en que μ y σ son desconocidos necesitamos entonces dos estadísticos suficientes para ellos, así determinamos que el estadístico T definido por:

$$T = (T_1, T_2) = (\min_i(X_1, X_2, ..., X_n), \sum_{i=1}^n X_i)$$

es un estadístico de suficiente de dos dimensiones para $\theta = (\mu, \sigma)$. $\square$

g) Sea $X_1, X_2, ..., X_n$ una muestra de una población que distribuye Gamma(k, ν). Determine un estadístico de dos dimensiones para $\theta = (k, \nu)$
SOLUCIÓN: *Utilicemos el teorema de Factorización para esto:*

$$f(x|k, v) = \prod_{i=1}^n \frac{\nu^k}{\Gamma(k)} x_i^{k-1} e^{-\nu x_i} = \frac{\nu^{nk}}{\Gamma(k)^n} \cdot (\prod_{i=1}^n x_i)^{k-1} \cdot e^{-\nu \sum_{i=1}^n x_i}$$

Si consideramos toda la función como $g(T_1(X), T_2(X)|\nu, k)$ y a $h(x) = 1$, entonces concluimos que es suficiente el estadístico T definido por:

$$T = (T_1, T_2) = (\prod_{i=1}^n x_i, \sum_{i=1}^n x_i)$$

$\square$