3.1. Estadísticos suficientes

El uso de estadísticos muestrales en problemas de inferencia se justifica por la necesidad de resumir la información proporcionada por la muestra. Sin embargo, en este resumen puede haber pérdida de información.

Por ejemplo, si se considera el estadístico $T(X_1, \ldots, X_n) = X_1$, está claro que se pierde la información proporcionada por el resto de observaciones, $X_2, \ldots, X_n$ (salvo que la distribución teórica sea degenerada, en cuyo caso es *suficiente* considerar sólo la primera observación, ya que las demás serán repeticiones de ésta).

La propiedad de suficiencia, considerada inicialmente por Fisher (1922), se basa en la idea de que el estadístico recoja toda la información sobre la distribución de la variable (o, equivalentemente, sobre el parámetro) contenida en la muestra.

Aunque la definición puede aplicarse a familias de distribuciones arbitrarias, todo nuestro estudio se centrará en modelos paramétricos:

$$X \to \{P_\theta;\ \theta \in \Theta\}.$$

Supondremos que las distribuciones son de tipo discreto o continuo, y vendrán representadas por su función masa de probabilidad, o función de densidad, según el caso, $\{f_\theta;\ \theta \in \Theta\}$.

Estadístico suficiente

Sea $X \to \{P_\theta;\ \theta \in \Theta\}$ y $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X. Un estadístico $T(X_1, \ldots, X_n)$ es suficiente para la familia de distribuciones de X (o suficiente para el parámetro) si la distribución condicionada de $(X_1, \ldots, X_n)$ a cualquier valor de $T(X_1, \ldots, X_n)$ es independiente de θ.

Esta definición indica que, conocido el valor del estadístico, la muestra no proporciona información adicional sobre el parámetro, ya que su distribución no depende de él.

Ejemplo 3.1.1: *Sea* (X_1, X_2) *una muestra aleatoria simple de una variable* $X \to \{\mathcal{P}(\lambda); \quad \lambda \in \mathbb{R}^+\}$. *Probar que* $X_1 + X_2$ *es suficiente para esta familia de distribuciones.*

$$T = X_1 + X_2 \to \{\mathcal{P}(2\lambda); \quad \lambda \in \mathbb{R}^+\} \Rightarrow P_\lambda(T = t) = \frac{e^{-2\lambda}(2\lambda)^t}{t!}, \quad t \in \mathbb{N}.$$

Calculemos la función masa de probabilidad de (X_1, X_2) condicionada a $T = t$, siendo $t \in \mathbb{N}$ un valor arbitrario:

$$P_\lambda\left(X_1 = x_1, X_2 = x_2 \middle/ T = t\right) = \frac{P_\lambda\left(X_1 = x_1, X_2 = x_2, T = t\right)}{P_\lambda\left(T = t\right)}.$$

- $T(x_1, x_2) = x_1 + x_2 \neq t \Rightarrow P_\lambda\left(X_1 = x_1, X_2 = x_2, T = t\right) = 0.$
- $T(x_1, x_2) = x_1 + x_2 = t \Rightarrow$

$$P_\lambda\left(X_1 = x_1, X_2 = x_2, T = t\right) = P_\lambda\left(X_1 = x_1, X_2 = x_2\right) = \frac{e^{-2\lambda}\lambda^{x_1+x_2}}{x_1!x_2!}.$$

$$P_\lambda\left(X_1 = x_1, X_2 = x_2 \middle/ T = t\right) = \frac{e^{-2\lambda}\lambda^{x_1+x_2}}{x_1!x_2!} \times \frac{t!}{e^{-2\lambda}(2\lambda)^t} = \frac{t!}{x_1!x_2!2^t}.$$

$$P\left(X_1 = x_1, X_2 = x_2 \big/ T = t\right) = \begin{cases} 0, & si \;\; x_1 + x_2 \neq t \\ \dfrac{t!}{x_1! x_2! 2^t}, & si \;\; x_1 + x_2 = t. \end{cases}$$

Puesto que esta distribución no depende de λ, queda probado que $X_1 + X_2$ es suficiente.

Por ejemplo, si se pierden los datos originales pero se sabe que la suma es 2 (esto es, $T = X_1 + X_2 = 2$), los posibles datos originales serían $(2,0), (0,2)$, $(1,1)$, y conocemos la probabilidad de cada uno de ellos, a pesar de no conocer λ:

- $P\left(X_1 = 2, X_2 = 0 \big/ T = 2\right) = \dfrac{1}{4}$,
- $P\left(X_1 = 0, X_2 = 2 \big/ T = 2\right) = \dfrac{1}{4}$,
- $P\left(X_1 = 1, X_2 = 1 \big/ T = 2\right) = \dfrac{1}{2}$.

Por tanto, sabiendo que $T = 2$, el conocer el valor concreto de la muestra no proporciona ninguna información sobre λ.

La definición de estadístico suficiente no es constructiva, ya que no proporciona un método para su determinación. Además, en muchas situaciones, la definición es poco práctica porque el cálculo de las distribuciones condicionadas puede no ser simple (en el caso continuo, la distribución conjunta de la muestra y del estadístico, función de ella, es una distribución singular).

Este problema queda resuelto gracias al siguiente teorema, que proporciona una caracterización de estadísticos suficientes y un método simple para su obtención.

Teorema de factorización de Neyman-Fisher

Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \to \{P_\theta;\ \theta \in \Theta\}$. Una condición necesaria y suficiente para que el estadístico $T(X_1, \ldots, X_n)$ sea suficiente es que, para cualquier valor del parámetro, la función de densidad o masa de probabilidad de la muestra se factorice como:

$$f_\theta^n(x_1, \ldots, x_n) = h(x_1, \ldots, x_n) g_\theta(T(x_1, \ldots, x_n)), \quad \forall (x_1, \ldots, x_n) \in \chi^n$$

donde $h(x_1, \ldots, x_n)$ es independiente de θ, y $g_\theta(T(x_1, \ldots, x_n))$ depende de $(x_1, \ldots, x_n)$ sólo a través de $T(x_1, \ldots, x_n)$.

Demostración: La realizamos sólo en el caso discreto.

$\Rightarrow$) Supongamos que $T \equiv T(X_1, \ldots, X_n)$ es suficiente:

$$\begin{aligned}
P_\theta(X_1 = x_1, \ldots, X_n = x_n) &= P_\theta(X_1 = x_1, \ldots, X_n = x_n,\ T = T(x_1, \ldots, x_n)) \\
&= P_\theta\left(X_1 = x_1, \ldots, X_n = x_n \big/ T = T(x_1, \ldots, x_n)\right) P_\theta(T = T(x_1, \ldots, x_n)) \\
&= \underbrace{P\left(X_1 = x_1, \ldots, X_n = x_n \big/ T = T(x_1, \ldots, x_n)\right)}_{h(x_1, \ldots, x_n)} \underbrace{P_\theta(T = T(x_1, \ldots, x_n))}_{g_\theta(T(x_1, \ldots, x_n))}.
\end{aligned}$$

$\Leftarrow$) Para simplificar la notación, notemos la muestra por $\boldsymbol{X} = (X_1, \ldots, X_n)$ y a las realizaciones muestrales por $\mathbf{x} = (x_1, \ldots, x_n)$. Supongamos que la función masa de probabilidad de la muestra se factoriza según se indica,

$$P_\theta(\boldsymbol{X} = \mathbf{x}) = h(\mathbf{x}) g_\theta(T(\mathbf{x})), \quad \forall \mathbf{x} \in \chi^n.$$

Para probar la suficiencia de T calculamos la distribución condicionada de la muestra a un valor arbitrario, $T = t$.

$$P_\theta\left(\boldsymbol{X} = \mathbf{x} \big/ T = t\right) = \frac{P_\theta(\boldsymbol{X} = \mathbf{x},\ T = t)}{P_\theta(T = t)} = \begin{cases} 0, & T(\mathbf{x}) \neq t, \\ \dfrac{P_\theta(\boldsymbol{X} = \mathbf{x})}{P_\theta(T = t)}, & T(\mathbf{x}) = t. \end{cases}$$

$\boxed{T(\mathbf{x}) = t}$

- $P_\theta(\boldsymbol{X} = \mathbf{x}) = h(\mathbf{x}) g_\theta(T(\mathbf{x})) = h(\mathbf{x}) g_\theta(t)$.
- $P_\theta(T = t) = \sum_{\mathbf{x}'/T(\mathbf{x}')=t} P_\theta(\boldsymbol{X} = \mathbf{x}') = \sum_{\mathbf{x}'/T(\mathbf{x}')=t} h(\mathbf{x}') g_\theta(T(\mathbf{x}')) = g_\theta(t) \sum_{\mathbf{x}'/T(\mathbf{x}')=t} h(\mathbf{x}')$.

Por tanto, la distribución condicionada es independiente de θ, y T es suficiente:

$$P\left(\mathbf{X} = \mathbf{x} \big/ T = t\right) = \begin{cases} 0, & T(\mathbf{x}) \neq t \\ \dfrac{h(\mathbf{x})}{\sum_{\mathbf{x}'/T(\mathbf{x}')=t} h(\mathbf{x}')}, & T(\mathbf{x}) = t. \end{cases}$$

$\square$

Propiedades de los estadísticos suficientes:

- *Si $T(X_1,\dots,X_n)$ es suficiente para la familia $\{P_\theta;\ \theta\in\Theta\}$, lo es para cualquier subfamilia, $\{P_\theta;\ \theta\in\Theta'\}$ con $\Theta'\subseteq\Theta$.*

  $*$ Si la factorización es válida $\forall\theta\in\Theta$, lo es $\forall\theta\in\Theta'$. $\square$

- *Si $T(X_1,\dots,X_n)$ es suficiente para $\{P_\theta;\ \theta\in\Theta\}$ y $U(X_1,\dots,X_n)$ es otro estadístico tal que $T=f(U)$, $U(X_1,\dots,X_n)$ es también suficiente para $\{P_\theta;\ \theta\in\Theta\}$.*

  $$* \ f_\theta^n(x_1,\dots,x_n)=h(x_1,\dots,x_n)g_\theta\big(T(x_1,\dots,x_n)\big) = h(x_1,\dots,x_n)g'_\theta\big(U(x_1,\dots,x_n)\big),\ \text{con } g'_\theta=g_\theta\circ f. \quad \square$$

- *Cualquier transformación biunívoca de un estadístico suficiente es también suficiente.*

  $*$ T suficiente y $T'=h(T)\Rightarrow T=h^{-1}(T')\Rightarrow T'$ suficiente. $\square$

Nota: Estas propiedades indican que no existe un único estadístico suficiente. Sin embargo, puede probarse que dados dos estadísticos suficientes, uno de ellos es función del otro. Lo ideal es trabajar con el que proporcione mayor resumen de los datos muestrales, que se denomina *minimal suficiente*, y usualmente se obtiene del Teorema de Neyman-Fisher con una factorización adecuada.

Ejemplo 3.1.3: *Determinar un estadístico suficiente para la familia de distribuciones $\{B(1,p);\ p\in(0,1)\}$, basándose en una muestra aleatoria simple de una variable con distribución en dicha familia.*

$$(X_1,\dots,X_n) \text{ m.a.s. de } X\to\{B(1,p);\ p\in(0,1)\}\to\chi^n=(0,1)^n.$$

$\forall p\in(0,1)$:

$$P_p(X_1=x_1,\dots,X_n=x_n)=p^{\sum\limits_{i=1}^n x_i}(1-p)^{n-\sum\limits_{i=1}^n x_i},\quad \forall(x_1,\dots,x_n)\in\chi^n.$$

Para expresar esta función en la forma indicada en el *Teorema de Factorización* observamos que los dos factores dependen de p, y sólo dependen de los datos muestrales $(x_1,\dots,x_n)$ a través de $\sum\limits_{i=1}^n x_i$. Así, definiendo:

- $T(X_1,\dots,X_n)=\sum\limits_{i=1}^n X_i$,
- $h(x_1,\dots,x_n)=1;\quad g_p(t)=p^t(1-p)^{n-t}$,

se tiene, $\forall p\in(0,1)$:

$$P_p(X_1=x_1,\dots,X_n=x_n)=h(x_1,\dots,x_n)g_p\big(T(x_1,\dots,x_n)\big),\ \forall(x_1,\dots,x_n)\in\chi^n,$$

y se deduce que $\sum\limits_{i=1}^n X_i$ es suficiente para esta familia de distribuciones.

Ejemplo 3.1.4: *Calcular un estadístico suficiente para la familia de distribuciones $\{U(0,\theta);\ \theta \in \mathbb{R}^+\}$, basándose en una muestra aleatoria simple de una variable con distribución en dicha familia.*

$$X \to \{U(0,\theta);\ \theta \in \mathbb{R}^+\} \to f_\theta(x) = \begin{cases} \dfrac{1}{\theta}, & 0 < x < \theta, \\ 0, & x \geq \theta. \end{cases}$$

Por tanto, $\chi_\theta = (0,\theta) \Rightarrow \chi = \bigcup\limits_{\theta \in \mathbb{R}^+} \chi_\theta = \mathbb{R}^+ \Rightarrow \chi^n = (\mathbb{R}^+)^n$.

La función de densidad de la muestra para cada valor del parámetro, $\theta \in \mathbb{R}^+$, es:

$$f_\theta^n(x_1,\ldots,x_n) = \prod_{i=1}^n f_\theta(x_i) = \begin{cases} \dfrac{1}{\theta^n}, & 0 < \max x_i < \theta \\ 0, & \max x_i \geq \theta, \end{cases} \qquad (x_1,\ldots,x_n) \in (\mathbb{R}^+)^n.$$

Observamos que esta función depende de $\max x_i$, y para expresarla en los términos del *Teorema de Factorización*, usamos la función indicadora de $\mathbb{R}^+$:

$$I_{\mathbb{R}^+}(x) = \begin{cases} 1, & \text{si } x \in \mathbb{R}^+ \\ 0, & \text{si } x \notin \mathbb{R}^+. \end{cases}$$

$$f_\theta^n(x_1,\ldots,x_n) = \frac{1}{\theta^n} I_{\mathbb{R}^+}(\theta - \max x_i), \quad \forall (x_1,\ldots,x_n) \in \chi^n = (\mathbb{R}^+)^n.$$

Tema 3. Suficiencia y completitud A. Hermoso Carazo
└3.1. Estadísticos suficientes

$$f_\theta^n(x_1,\ldots,x_n) = \frac{1}{\theta^n} I_{\mathbb{R}^+}(\theta - \max x_i), \quad \forall (x_1,\ldots,x_n) \in \chi^n = (\mathbb{R}^+)^n.$$

De esta forma, tomando

- $T(x_1,\ldots,x_n) = \max x_i$,
- $h(x_1,\ldots,x_n) = 1$ y $g_\theta(t) = \dfrac{1}{\theta^n} I_{\mathbb{R}^+}(\theta - t)$,

tenemos la factorización:

$$f_\theta^n(x_1,\ldots,x_n) = h(x_1,\ldots,x_n) g_\theta(\max x_i), \quad \forall (x_1,\ldots,x_n) \in \chi^n = (\mathbb{R}^+)^n,$$

lo que implica que $T(X_1,\ldots,X_n) = \max X_i$ es un estadístico suficiente para la familia de distribuciones considerada.

Por ejemplo, en la muestra $(1,\ 2.5,\ 3,\ 1.3)$ el valor máximo, 3, resume toda la información sobre θ (si conocemos que el máximo es 3, el resto de valores no aporta información adicional sobre el parámetro, ya que este indica el máximo valor que puede tomar la variable).

3.2. Estadísticos completos

Aunque esta propiedad no tiene, en principio, relación directa con la suficiencia, el uso de estadísticos suficientes y completos proporciona, como veremos, propiedades adicionales a las inferencias basadas en estadísticos suficientes.

Estadístico completo

Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \to \{P_\theta;\ \theta \in \Theta\}$. Un estadístico $T \equiv T(X_1, \ldots, X_n)$ es completo si para cualquier función medible unidimensional, g, se verifica:

$$E_\theta\left[g(T)\right] = 0,\ \forall \theta \in \Theta \ \Rightarrow\ P_\theta\big(g(T) = 0\big) = 1,\ \forall \theta \in \Theta.$$

Ejemplo 3.2.1: *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \to \{B(1,p);\ p \in (0,1)\}$. Probar que $\sum_{i=1}^{n} X_i$ es completo.*

$$T = \sum_{i=1}^{n} X_i \to \big\{B(n,p);\ p \in (0,1)\big\}.$$

Para probar que T es completo, consideremos una función medible g tal que $g(T)$ tenga media cero bajo todas las distribuciones de la familia:

$$E_p\big[g(T)\big] = \sum_{t=0}^{n} g(t) P_p(T = t) = \sum_{t=0}^{n} g(t)\binom{n}{t} p^t (1-p)^{n-t} = 0, \quad \forall p \in (0,1).$$

$$\Downarrow\ y = \frac{p}{1-p}$$

$$\sum_{t=0}^{n} g(t)\binom{n}{t} y^t = 0, \quad \forall y \in \mathbb{R}^+.$$

Ya que esta suma es un polinomio en y de grado n que se anula $\forall y \in \mathbb{R}^+$, concluimos que todos los coeficientes son nulos, de donde $g(0) = \cdots = g(n) = 0$. Por tanto, g se anula en todos los posibles valores de T y se tiene:

$$P_p\big(g(T) = 0\big) \geq P_p\big(T = 0, \ldots, n\big) = 1,\ \forall p \in (0,1) \ \Rightarrow\ T \text{ es completo.}$$

Ejemplo 3.2.2: *Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de $X \to \{B(1,p);\ p \in (0,1)\}$. Probar que $X_1 - X_2$ no completo.*

Para probar que $T = X_1 - X_2$ no es completo, basta encontrar una función $g(T)$ que tenga media cero bajo todas las distribuciones de la familia y no sea nula bajo alguna distribución de la familia.

En este caso, el propio estadístico nos proporciona dicha función ($g(T) = T$):

- $E_p[T] = E_p[X_1] - E_p[X_2] = 0, \ \forall p \in (0,1)$.
- $P_p(T=0) = P_p(X_1 = X_2) = P_p(X_1 = 0, X_2 = 0) + P_p(X_1 = 1, X_2 = 1)$
  $= (1-p)^2 + p^2 < 1, \quad p \in (0,1)$.

Ejemplo 3.2.3: *Probar que $T(X_1, \dots, X_n) = \text{máx}\, X_i$ es completo si $(X_1, \dots, X_n)$ es una muestra aleatoria simple de $X \to \{U(0,\theta);\ \theta \in \mathbb{R}^+\}$.*

La función de densidad del estadístico T es (Ejemplo 1.4.4):

$$f_\theta^T(t) = \frac{nt^{n-1}}{\theta^n}, \quad t \in (0,\theta), \quad \forall \theta \in \mathbb{R}^+.$$

Sea g una función medible tal que $E_\theta[g(T)] = 0, \ \forall \theta \in \mathbb{R}^+$. Especificamos esta condición usando la densidad de T:

$$E_\theta[g(T)] = \int_0^\theta g(t) f_\theta^T(t)dt = \int_0^\theta g(t)\frac{nt^{n-1}}{\theta^n}dt = 0, \quad \forall \theta \in \mathbb{R}^+.$$

De la última igualdad, sacando las constantes (no nulas) de la integral se deduce que la función g debe verificar:

$$\int_0^\theta g(t)t^{n-1}dt = 0, \quad \forall \theta \in \mathbb{R}^+.$$

Ahora, aplicando el *Teorema Fundamental del Cálculo*, derivamos respecto de θ ambos miembros, obteniendo $g(\theta)\theta^{n-1} = 0, \ \forall \theta \in \mathbb{R}^+$, lo que, a su vez, equivale a que $g(\theta) = 0, \ \forall \theta \in \mathbb{R}^+$. Esto es, la función g se anula en $\mathbb{R}^+$ (conjunto donde toma valores T) y, por tanto:

$$P_\theta(g(T) = 0) \geq P_\theta(T > 0) = 1, \quad \forall \theta \in \mathbb{R}^+ \ \Rightarrow \ T = \text{máx}\, X_i \text{ es completo.}$$

Importante: no aplicar el TFC cuando el theta esté dentro de la integral.

3.3. Suficiencia y completitud en familias exponenciales

Describimos ahora un tipo de familias de distribuciones que cubre muchas de las distribuciones importantes y en las que, además, bajo condiciones muy generales, existen estadísticos suficientes y completos.

Familia exponencial k-paramétrica

Sea $\{P_\theta;\ \theta \in \Theta\}$ una familia de distribuciones de probabilidad con funciones de densidad o masa de probabilidad $\{f_\theta;\ \theta \in \Theta\}$. Se dice que $\{P_\theta;\ \theta \in \Theta\}$ es de tipo exponencial (k-paramétrica, $k \in \mathbb{N}$) si:

- *Θ es un intervalo de $\mathbb{R}^k$.*
- *$\forall \theta \in \Theta,\ \{x \ / \ f_\theta(x) > 0\} = \chi$ es independiente de θ.*
- *$\forall \theta \in \Theta,\ f_\theta(x) = \exp\left\{\sum_{h=1}^{k} Q_h(\theta)T_h(x) + S(x) + D(\theta)\right\}, \ \ \forall x \in \chi,$*

  siendo $T_1, \ldots, T_k, S$, funciones medibles de x, y $Q_1, \ldots, Q_k, D$ funciones de θ.

Teorema de suficiencia y completitud en familias exponenciales

Sea X una variable aleatoria con distribución en una familia exponencial k-paramétrica, con funciones de densidad, o funciones masa de probabilidad, de la forma

$$f_\theta(x) = \exp\left\{\sum_{h=1}^{k} Q_h(\theta)T_h(x) + S(x) + D(\theta)\right\}, \ \ \forall x \in \chi, \quad \theta \in \Theta.$$

Si $(X_1, \ldots, X_n)$ es una muestra aleatoria simple de X, se tiene:

$i)$ La familia de distribuciones de $(X_1, \ldots, X_n)$ es también exponencial k-paramétrica.

$ii)$ El estadístico $\left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i)\right)$ es suficiente para θ.

$iii)$ Si $k \leq n$ y el conjunto imagen de $Q(\theta) = (Q_1(\theta), \ldots, Q_k(\theta))$ contiene a un abierto de $\mathbb{R}^k$, el estadístico $\left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i)\right)$ es también completo.

Demostración:

$i)$ El espacio paramétrico correspondiente a la familia de distribuciones de la muestra es Θ (intervalo de $\mathbb{R}^k$). El conjunto de valores de $(X_1, \ldots, X_n)$ es χ^n (independiente de θ, por serlo χ). Finalmente, la función de densidad (o masa de probabilidad) de la muestra, para cada valor de θ, es:

$$f_\theta^n(x_1, \ldots, x_n) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \exp\Big\{ \sum_{h=1}^k Q_h(\theta) T_h(x_i) + S(x_i) + D(\theta) \Big\}$$

$$= \exp\Big\{ \sum_{h=1}^k Q_h(\theta) \Big(\sum_{i=1}^n T_h(x_i) \Big) + \sum_{i=1}^n S(x_i) + nD(\theta) \Big\},\ \forall (x_1, \ldots, x_n) \in \chi^n$$

Así, tomando $T_h(x_1, \ldots, x_n) = \sum_{i=1}^n T_h(x_i)$ y $S_h(x_1, \ldots, x_n) = \sum_{i=1}^n S_h(x_i)$, se concluye que la familia de distribuciones de la muestra es exponencial k-paramétrica.

$ii)$ La suficiencia es inmediata por el Teorema de Factorización:

$$f_\theta^n(x_1, \ldots, x_n) = \exp\Big\{ \sum_{i=1}^n S(x_i) \Big\} \times \exp\Big\{ \sum_{h=1}^k Q_h(\theta) \Big(\sum_{i=1}^n T_h(x_i) \Big) + nD(\theta) \Big\},$$

observándose que el primer factor es independiente de θ, y el segundo depende de $(x_1, \ldots, x_n)$ sólo a través de $\Big(\sum_{i=1}^n T_1(X_i), \ldots, \sum_{i=1}^n T_k(X_i) \Big)$.

$iii)$ Ver Rohatgi, p. 347. □

Ejemplo 3.3.1: *Probar que la familia $\{\mathcal{N}(\mu, \sigma_0^2);\ \mu \in \mathbb{R}\}$ es exponencial, y encontrar un estadístico suficiente y completo basado en una muestra aleatoria simple de una variable X con distribución en dicha familia.*

$$\forall \mu \in \mathbb{R},\quad f_\mu(x) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp\Big\{ -\frac{(x-\mu)^2}{2\sigma_0^2} \Big\},\quad x \in \mathbb{R}.$$

- $\mu \in \mathbb{R} = (-\infty, +\infty)$, intervalo de $\mathbb{R}$.
- $\{x \ /\ f_\mu(x) > 0\} = \mathbb{R}$, independiente de μ.
- $f_\mu(x) = \exp\Big\{ -\ln\big(\sigma_0\sqrt{2\pi}\big) - \frac{x^2}{2\sigma_0^2} - \frac{\mu^2}{2\sigma_0^2} + \frac{x\mu}{\sigma_0^2} \Big\},\quad x \in \mathbb{R}.$

$$Q(\mu) = \frac{\mu}{\sigma_0^2},\quad T(x) = x,\quad S(x) = -\frac{x^2}{2\sigma_0^2},\quad D(\mu) = -\ln\Big(\sigma_0\sqrt{2\pi}\Big) - \frac{\mu^2}{2\sigma_0^2}.$$

Por lo tanto, la familia es exponencial uniparamétrica, y si $(X_1, \ldots, X_n)$ es una m.a.s. tenemos:

- El estadístico $\sum_{i=1}^n T(X_i) = \sum_{i=1}^n X_i$ es suficiente.
- $Q(\mu) = \frac{\mu}{\sigma_0^2},\quad \mu \in \mathbb{R} \to Q(\mathbb{R}) = \mathbb{R}$, abierto de $\mathbb{R}$.

  Así, si $n \geq 1$ (siempre), $\sum_{i=1}^n X_i$ es, además de suficiente, completo.

Ejemplo 3.3.2: *Probar que la familia $\{\mathcal{N}(\mu_0, \sigma^2);\ \sigma^2 \in \mathbb{R}^+\}$ es exponencial uniparamétrica, y encontrar un estadístico suficiente y completo basado en una muestra aleatoria simple de una variable X con distribución en dicha familia.*

$$\forall \sigma^2 \in \mathbb{R}^+, \quad f_{\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu_0)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}.$$

- $\sigma^2 \in \mathbb{R}^+ = (0, +\infty)$, intervalo de $\mathbb{R}$.
- $\{x \ / \ f_{\sigma^2}(x) > 0\} = \mathbb{R}$, independiente de σ^2.
- $f_{\sigma^2}(x) = \exp\left\{-\ln\left(\sigma\sqrt{2\pi}\right) - \dfrac{(x-\mu_0)^2}{2\sigma^2}\right\}.$

$$Q(\sigma^2) = -\frac{1}{2\sigma^2}, \quad T(x) = (x-\mu_0)^2, \quad S(x) = 0, \quad D(\sigma^2) = -\ln\left(\sigma\sqrt{2\pi}\right).$$

Por lo tanto, la familia es exponencial uniparamétrica, y si $(X_1, \ldots, X_n)$ es una m.a.s. tenemos:

- El estadístico $\sum\limits_{i=1}^{n} T(X_i) = \sum\limits_{i=1}^{n} (X_i - \mu_0)^2$ es suficiente.
- $Q(\sigma^2) = -\dfrac{1}{2\sigma^2}$, $\sigma^2 \in \mathbb{R}^+ \to Q(\mathbb{R}^+) = \mathbb{R}^-$, abierto de $\mathbb{R}$.

Así, si $n \geq 1$ (siempre), el estadístico $\sum\limits_{i=1}^{n} (X_i - \mu_0)^2$ es también completo.

Ejemplo 3.3.3: *Probar que la familia $\{\mathcal{N}(\mu,\sigma^2);\ \mu\in\mathbb{R},\ \sigma^2\in\mathbb{R}^+\}$ es exponencial biparamétrica, y encontrar un estadístico suficiente y completo basado en una muestra aleatoria simple de una variable X con distribución en dicha familia.*

Se trata de una familia biparamétrica, cuyas funciones de densidad son

$$f_{(\mu,\sigma^2)}(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},\quad x\in\mathbb{R},\ (\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}^+.$$

- $(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}^+=(-\infty,+\infty)\times(0,+\infty)$, intervalo de $\mathbb{R}^2$.
- $\{x\ /\ f_{(\mu,\sigma^2)}(x)>0\}=\mathbb{R}$, independiente de (μ,σ^2).
- $f_{(\mu,\sigma^2)}(x)=\exp\left\{-\ln\sqrt{2\pi}-\ln\sigma-\dfrac{x^2}{2\sigma^2}-\dfrac{\mu^2}{2\sigma^2}+\dfrac{x\mu}{\sigma^2}\right\}$

$$Q_1(\mu,\sigma^2)=\frac{\mu}{\sigma^2},\quad T_1(x)=x,\quad Q_2(\mu,\sigma^2)=-\frac{1}{2\sigma^2},\quad T_2(x)=x^2$$

$$S(x)=-\ln\sqrt{2\pi},\quad D(\mu,\sigma^2)=-\ln\sigma-\frac{\mu^2}{2\sigma^2}.$$

- El estadístico $\left(\sum\limits_{i=1}^n T_1(X_i),\ \sum\limits_{i=1}^n T_2(X_i)\right)=\left(\sum\limits_{i=1}^n X_i,\ \sum\limits_{i=1}^n X_i^2\right)$ es suficiente.
- $Q(\mu,\sigma^2)=\left(Q_1(\mu,\sigma^2),\ Q_2(\mu,\sigma^2)\right)=\left(\dfrac{\mu}{\sigma^2},\ -\dfrac{1}{2\sigma^2}\right),\quad \mu\in\mathbb{R},\ \sigma^2\in\mathbb{R}^+.$
  Por tanto, $Q(\mathbb{R}\times\mathbb{R}^+)=\mathbb{R}\times\mathbb{R}^-$, que es un abierto de $\mathbb{R}^2$.

Así, si $n\geq 2$ (muestra de dos elementos como mínimo), $\left(\sum\limits_{i=1}^n X_i,\ \sum\limits_{i=1}^n X_i^2\right)$ es, además de suficiente, completo.

Problema 1

Sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de una variable aleatoria $X \to \{B(k,p);\ p\in 0,1\}$ y sea $T=\sum_{i=1}^n X_i$. Probar, $a)$ usando la definición y $b)$ aplicando el teorema de factorización, que T es suficiente para p.

∗ Conjunto de valores de $X \to \chi = \{0,1,\dots,k\}$.

∗ Espacio muestral $\to \chi^n = \{0,1,\dots,k\}^n$.

∗ $T \to \{B(nk,p);\ p\in 0,1\}$.

$a)$ *Definición*: Sea $t=0,\dots,nk$, un valor arbitrario de $T=\sum_{i=1}^n X_i$. Para cualquier realización muestral, $(x_1,\dots,x_n)\in\{0,1,\dots,k\}^n$, se tiene:

$$P_p\left(X_1=x_1,\dots,X_n=x_n \big/ T=t\right) = \frac{P_p\left(X_1=x_1,\dots,X_n=x_n,\ T=t\right)}{P_p\left(T=t\right)}\ (^*)$$

∗ $\sum_{i=1}^n x_i \neq t \Rightarrow P_p\left(X_1=x_1,\dots,X_n=x_n,\ T=t\right)=0.$

∗ $\sum_{i=1}^n x_i = t \Rightarrow P_p\left(X_1=x_1,\dots,X_n=x_n,\ T=t\right)=P_p\left(X_1=x_1,\dots,X_n=x_n\right).$

$$\overset{*}{=}\begin{cases} 0, & \sum_{i=1}^n x_i \neq t \\ \dfrac{P_p\left(X_1=x_1,\dots,X_n=x_n\right)}{P_p\left(T=t\right)} = \dfrac{\binom{k}{x_1}\cdots\binom{k}{x_n}p^{\sum_{i=1}^n x_i}(1-p)^{nk-\sum_{i=1}^n x_i}}{\binom{nk}{t}p^t(1-p)^{nk-t}}, & \sum_{i=1}^n x_i = t \end{cases}$$

$$=\begin{cases} 0, & \sum_{i=1}^n x_i \neq t \\ \dfrac{\binom{k}{x_1}\cdots\binom{k}{x_n}}{\binom{nk}{t}}, & \sum_{i=1}^n x_i = t. \end{cases} \longrightarrow \text{independiente de } p \ \Rightarrow\ T \text{ es suficiente.}$$

b) Teorema de factorización:

$\forall p \in (0,1)$:

$$P_p(X_1 = x_1, \ldots, X_n = x_n) = \underbrace{\binom{k}{x_1} \cdots \binom{k}{x_n}}_{h(x_1,\ldots,x_n)} \underbrace{p^{\sum\limits_{i=1}^{n} x_i} (1-p)^{nk - \sum\limits_{i=1}^{n} x_i}}_{g_p\left(\sum\limits_{i=1}^{n} x_i\right)}, \quad (x_1, \ldots, x_n) \in \chi^n.$$

$$\Downarrow$$

$$T = \sum_{i=1}^{n} X_i \ \text{ suficiente.}$$

a) $X \to \{U(-\theta/2, \theta/2); \quad \theta > 0\}$

$$f_\theta(x) = \frac{1}{\theta}, \quad -\theta/2 < x < \theta/2 \quad \left(\Leftrightarrow |x| < \theta/2 \right).$$

$$\chi_\theta = (-\theta/2,\ \theta/2) \ \to \ \chi = \bigcup_{\theta > 0} (-\theta/2,\ \theta/2) = \mathbb{R} \to \ \chi^n = \mathbb{R}^n.$$

$$f_\theta^n(x_1, \ldots, x_n) = \begin{cases} \dfrac{1}{\theta^n}, & \text{máx}\,|x_i| < \theta/2 \\ 0, & \text{máx}\,|x_i| \geq \theta/2, \end{cases} \qquad (x_1, \ldots, x_n) \in \mathbb{R}^n$$

$$= \underbrace{\frac{1}{\theta^n} I_{\mathbb{R}^+}\left(\theta/2 - \text{máx}\,|x_i|\right)}_{g_\theta\left(\text{máx}\,|x_i|\right)}, \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

$$h(x_1, \ldots, x_n) = 1$$

$$\Downarrow$$

$$\text{máx}\,|X_i| \ \text{ suficiente.}$$

Nota: $\text{máx}\,|X_i| = \text{máx}\left(\text{máx}\,X_i,\ -\text{mín}\,X_i\right)$. Esto indica que el estadístico bidimensional $\left(\text{máx}\,X_i,\ \text{mín}\,X_i\right)$ es también suficiente (es el que hubiésemos obtenido de no considerar el valor absoluto en la expresión de f_θ).

b) $X \to \{\Gamma(p,a) \,;\, p>0, a>0\}$ $\longleftrightarrow$ $f_{(p,a)}(x) = \dfrac{a^p}{\Gamma(p)} x^{p-1} e^{-ax}$, $x \in \mathbb{R}^+$

$\chi_{(p,a)} = \mathbb{R}^+$, $\chi = \bigcup \chi_{(p,a)} = \mathbb{R}^+$ considero $(X_1, \dots, X_n)$ m.a.s en $(\mathbb{R}^+)^n$

$f^n_{(p,a)}(x_1, \dots, x_n) = \prod^n f_{(p,a)}(x_i) = \left(\dfrac{a^p}{\Gamma(p)}\right)^n e^{-a\sum^n x_i} \prod^n x_i^{p-1}$ Aquí considero

$h(x_1, \dots, x_n) = 1$, $g_{(a,p)}(t(x_1, \dots, x_n)) = \left(\dfrac{a^p}{\Gamma(p)}\right)^n e^{-a t_1} t_2^{p-1} \Rightarrow T(X_1, \dots, X_n) =$

$\left(\sum^n X_i, \prod^n X_i\right)$

$f^n_p(x_1, \dots, x_n) = e^{-a\sum^n x_i} \left(\dfrac{a^p}{\Gamma(p)}\right)^n \left(\prod^n x_i\right)^{p-1}$ En este caso considero

$h(x_1, \dots, x_n) = e^{-a\sum^n x_i}$, $g_p(t(x_1, \dots, x_n)) = \left(\dfrac{a^p}{\Gamma(p)}\right)^n t^{p-1} \Rightarrow$ el est. es

$T(X_1, \dots, X_n) = \prod^n X_i$.

$f^n_a(x_1, \dots, x_n) = \left(\dfrac{1}{\Gamma(p)}\right)^n \left(\prod^n x_i\right)^{p-1} e^{-a\sum^n x_i} a^{pn}$. Aquí, $h(x_1, \dots, x_n) =$

$\left(\dfrac{1}{\Gamma(p)}\right)^n \left(\prod^n x_i\right)^{p-1}$ y $g_a(t(x_1, \dots, x_n)) = e^{-at} a^{pn} \Rightarrow T(X_1, \dots, X_n) = \sum^n X_i$

c) $X \to \{\beta(p,q);\ \ p>0,\ q>0\}$

$$f_{p,q}(x) = \frac{1}{\beta(p,q)} x^{p-1}(1-x)^{q-1}, \ \ 0<x<1 \ \ \longrightarrow \ \ \chi^n = (0,1)^n.$$

$$f^n_{p,q}(x_1, \dots x_n) = \underbrace{\frac{1}{(\beta(p,q))^n} \Big(\prod_{i=1}^n x_i\Big)^{p-1} \Big(\prod_{i=1}^n (1-x_i)\Big)^{q-1}}_{g_{p,q}\left(\prod_{i=1}^n x_i, \prod_{i=1}^n (1-x_i)\right)}, \ \ (x_1, \dots, x_n) \in (0,1)^n.$$

$$h(x_1, \dots, x_n) = 1,$$

$$\Downarrow$$

$$\Big(\prod_{i=1}^n X_i, \ \prod_{i=1}^n (1-X_i)\Big) \ \textbf{suficiente para} \ \ (p,q).$$

c1) $X \to \{\beta(p_0, q);\ q > 0\} \longrightarrow \chi^n = (0,1)^n$

$$f_q^n(x_1, \dots x_n) = \frac{1}{\big(\beta(p_0,q)\big)^n}\Big(\prod_{i=1}^n x_i\Big)^{p_0-1}\Big(\prod_{i=1}^n (1-x_i)\Big)^{q-1}, \quad (x_1,\dots,x_n) \in (0,1)^n.$$

$$h(x_1,\dots,x_n) = \Big(\prod_{i=1}^n x_i\Big)^{p_0-1}, \qquad g_q\Big(\prod_{i=1}^n (1-x_i)\Big) = \frac{1}{\big(\beta(p_0,q)\big)^n}\Big(\prod_{i=1}^n (1-x_i)\Big)^{q-1}.$$

$$\Downarrow$$

$\prod_{i=1}^n (1-X_i)$ suficiente para q.

c2) $X \to \{\beta(p, q_0);\ p > 0\} \longrightarrow \chi^n = (0,1)^n$

$$f_p^n(x_1, \dots x_n) = \frac{1}{\big(\beta(p,q_0)\big)^n}\Big(\prod_{i=1}^n x_i\Big)^{p-1}\Big(\prod_{i=1}^n (1-x_i)\Big)^{q_0-1}, \quad (x_1,\dots,x_n) \in (0,1)^n.$$

$$h(x_1,\dots,x_n) = \Big(\prod_{i=1}^n (1-x_i)\Big)^{q_0-1}, \qquad g_p\Big(\prod_{i=1}^n x_i\Big) = \frac{1}{\big(\beta(p,q_0)\big)^n}\Big(\prod_{i=1}^n x_i\Big)^{p-1}.$$

$$\Downarrow$$

$\prod_{i=1}^n X_i$ suficiente para p.

---

d) $X \to \{P_{N_1,N_2};\ N_1, N_2 \in \mathbb{N},\ N_1 \le N_2\}$.

$$P_{N_1,N_2}(X = x) = \frac{1}{N_2 - N_1 + 1}, \quad x = N_1, \dots, N_2.$$

$$\chi_{N_1,N_2} = \{x \in \mathbb{N} \,/ x = N_1, \dots, N_2\} \;\to\; \chi = \bigcup_{N_1 \le N_2} \chi_{N_1,N_2} = \mathbb{N} \to \;\chi^n = \mathbb{N}^n.$$

$$P_{N_1,N_2}(X_1 = x_1, \dots, X_n = x_n)$$

$$= \underbrace{\frac{1}{(N_2 - N_1 + 1)^n} I_{[0,+\infty)}\big(\min x_i - N_1\big) I_{[0,+\infty)}\big(N_2 - \max x_i\big)}, \quad (x_1,\dots,x_n) \in \mathbb{N}^n.$$

$g_{N_1,N_2}(\min x_i, \max x_i)$.
$h(x_1,\dots,x_n) = 1$.

$$\Downarrow$$

$\big(\min X_i,\ \max X_i\big)$ suficiente para (N_1, N_2).

$d1)\ X \to \{P_{N_2};\ N_2 \in \mathbb{N},\ N_2 \geq N_1^0\}$

$$P_{N_2}(X = x) = \frac{1}{N_2 - N_1^0 + 1}, \quad x = N_1^0, \ldots, N_2.$$

$$\chi_{N_2} = \{x \in \mathbb{N} \ / \ N_1^0 \leq x \leq N_2\} \to \chi = \bigcup_{N_2 \geq N_1^0} \chi_{N_2} = \left\{x \in \mathbb{N} \ / \ x \geq N_1^0\right\}$$

$$\longrightarrow \chi^n = \left\{x \in \mathbb{N} \ / \ x \geq N_1^0\right\}^n.$$

$$P_{N_2}(X_1 = x_1, \ldots, X_n = x_n) = \underbrace{\frac{1}{(N_2 - N_1^0 + 1)^n} I_{[0,+\infty)}(N_2 - \text{máx}\, x_i)}_{g_{N_2}(\text{máx}\, x_i)}, \quad (x_1, \ldots, x_n) \in \chi^n.$$

$$h(x_1, \ldots, x_n) = 1$$

$$\Downarrow$$

$\text{máx}\, X_i$ suficiente para N_2.

Problema 6

Basándose en una muestra de tamaño arbitrario, obtener un estadístico suficiente y completo para la familia de distribuciones definidas por todas las densidades de la forma

$$f_\theta(x) = e^{\theta - x}, \quad x > \theta.$$

Ya que no se especifica el espacio paramétrico, tomamos el más general posible, $\Theta = \mathbb{R}$.

$$\chi_\theta = (\theta, +\infty) \Rightarrow \chi = \bigcup_{\theta \in \mathbb{R}} (\theta, +\infty) = \mathbb{R} \longrightarrow \chi^n = \mathbb{R}^n.$$

Estadístico suficiente:

$$f_\theta^n(x_1, \ldots, x_n) = e^{n\theta - \sum_{i=1}^n x_i} I_{(0,+\infty)}\left(\text{mín}\, x_i - \theta\right)$$

$$= \underbrace{e^{-\sum_{i=1}^n x_i}}_{h(x_1, \ldots, x_n)} \times \underbrace{e^{n\theta} I_{(0,+\infty)}(\text{mín}\, x_i - \theta)}_{g_\theta(\text{mín}\, x_i)}, \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

$$\Downarrow$$

$T = \text{mín}\, X_i$ suficiente.

Distribución de $T = \min X_i$: (R1: Problema 8)

$$F_\theta(x) = \begin{cases} 0, & x \leq \theta \\ \displaystyle\int_\theta^x e^{\theta-t}dt = -e^{\theta-t}\Big|_\theta^x = 1 - e^{\theta-x}, & x > \theta. \end{cases}$$

$$F_\theta^T(t) = 1 - \big(1 - F_\theta(t)\big)^n = \begin{cases} 0, & t \leq \theta \\ 1 - e^{n(\theta-t)}, & t > \theta \end{cases} \Rightarrow f_\theta^T(t) = ne^{n(\theta-t)}, \quad t > \theta.$$

Completitud: Sea g una función medible tal que $E_\theta\big[g(T)\big] = 0, \ \forall \theta \in \mathbb{R}$:

$$E_\theta[g(T)] = \int_\theta^{+\infty} g(t) f_\theta^T(t)dt = \int_\theta^{+\infty} g(t)ne^{n(\theta-t)}dt = 0, \quad \forall \theta \in \mathbb{R}.$$

$$\Downarrow$$

$$\int_\theta^{+\infty} g(t)e^{-nt}dt = 0, \ \forall \theta \in \mathbb{R} \overset{TFC}{\Rightarrow} g(t)e^{-nt} = 0, \ \forall t \in \mathbb{R} \Rightarrow g(t) = 0, \ \forall t \in \mathbb{R}.$$

$$\Downarrow$$

$$P_\theta\big(g(T) = 0\big) \geq P_\theta\left(T \in \mathbb{R}\right) = 1, \quad \forall \theta \in \mathbb{R}.$$

Problema 7

Comprobar que las siguientes familias de distribuciones son exponenciales uniparamétricas y, considerando una muestra aleatoria simple de una variable con distribución en dicha familia, obtener, si existe, un estadístico suficiente y completo.

$a)$ $\{B(k_0, p); \ 0 < p < 1\}$

$b)$ $\{\mathcal{P}(\lambda); \ \lambda > 0\}$

$c)$ $\{BN(k_0, p); \ 0 < p < 1\}$

$d)$ $\{exp(\lambda); \ \lambda > 0\}$

a) $\{B(k_0, p);\ 0 < p < 1\}$.

$$P_p(X = x) = \binom{k_0}{x} p^x (1-p)^{k_0 - x}, \quad x = 0, 1, \dots, k_0.$$

$*$ $p \in (0,1)$, intervalo de $\mathbb{R}$.

$*$ $\{x \ / \ P_p(X = x) > 0\} = \{0, 1, \dots, k_0\}$, independiente de p.

$$* \ P_p(X = x) = \exp\left\{ \ln\binom{k_0}{x} + x \ln p + (k_0 - x)\ln(1-p) \right\}$$

$$= \exp\left\{ \ln\binom{k_0}{x} + x \ln \frac{p}{1-p} + k_0 \ln(1-p) \right\}.$$

$$Q(p) = \ln \frac{p}{1-p}, \quad T(x) = x, \quad S(x) = \ln\binom{k_0}{x}, \quad D(p) = k_0 \ln(1-p).$$

Por lo tanto, la familia es de tipo exponencial y tenemos:

$*$ El estadístico $\sum_{i=1}^{n} T(X_i) = \sum_{i=1}^{n} X_i$ es suficiente.

$*$ $Q\big((0,1)\big) = \mathbb{R}$ es un abierto de $\mathbb{R}$. Como $n \geq 1$, el estadístico $\sum_{i=1}^{n} X_i$ es, además de suficiente, completo.

8.- Estudiar si las siguientes familias de distribuciones son exponenciales biparamétricas. Obtener un estadístico suficiente y completo.

a) $\{\Gamma(p,a) : p>0, a>0\}$

$X \sim \Gamma(p,a)$, $p,a \in \mathbb{R}^+ \Leftrightarrow f_{p,a}(x) = \dfrac{a^p}{\Gamma(p)} x^{p-1} e^{-ax}$, $x \in \mathbb{R}$

- $\Theta = \mathbb{R}^+ \times \mathbb{R}^+$ (espacio paramétrico) $\quad (p,a) \in \mathbb{R}^+ \times \mathbb{R}^+ = (0,+\infty) \times (0,+\infty)$, intervalo de $\mathbb{R}^2$
  $p>0$, $(p \in \mathbb{R}^+)$, intervalo de $\mathbb{R}$ y $a>0$ $(a \in \mathbb{R}^+)$ intervalo de $\mathbb{R}$

- $\{x / f_{p,a}(x) > 0\} = \mathbb{R}$, independientemente de (p,a)

- $f_{p,a}(x) = \exp\left\{ \ln\left(\dfrac{a^p}{\Gamma(p)} x^{p-1} e^{-ax}\right)\right\} = \exp\left\{ \ln\left(\dfrac{a^p}{\Gamma(p)}\right) + \ln(x^{p-1}) + \ln(e^{-ax}) \right\} =$

  $= \exp\left\{ \ln\left(\dfrac{a^p}{\Gamma(p)} \right.\right.$ meet.google.com está compartiendo tu pantalla. Dejar de compartir Ocultar $\left.\left.\right) + (p-1)\ln(x) - ax \right\} =$

$Q_1(a,p)$ $\quad T_1(x)$

$$= \exp\left\{\underbrace{\ln(a^p) - \ln(\Gamma(p))}_{D(a,p)} + \overbrace{p}^{Q_1(a,p)}\underbrace{\ln(x)}_{T_1(x)} \underbrace{- \ln(x)}_{S(x)} \underbrace{-a}_{Q_2(a,p)}\overbrace{x}^{T_2(x)}\right\}$$

Por lo tanto, la familia es de tipo exponencial biparamétrica y si $(X_1, \dots, X_n)$ es una m.a.s tenemos:
(por el Teorema de Suficiencia y completitud en familias exponenciales)

- El estadístico $\left(\sum_{i=1}^{n} T_1(X_i), \sum_{i=1}^{n} T_2(X_i)\right) = \left(\sum_{i=1}^{n} \ln(x_i), \sum_{i=1}^{n} x_i\right)$ es suficiente

- $Q(p,a) = (Q_1(p,a), Q_2(p,a)) = (p, -a)$ con $p, a \in \mathbb{R}^+$.

  Por tanto, $Q(\mathbb{R}^+ \times \mathbb{R}^+) = \mathbb{R}^+ \times \mathbb{R}^-$, que es un abierto de $\mathbb{R}^2$.

Así, si $n \geq 2$ (muestra de dos elementos como mínimo), $\left(\sum_{i=1}^{n} \ln(x_i), \sum_{i=1}^{n} x_i\right)$, es además de suficiente, completo.

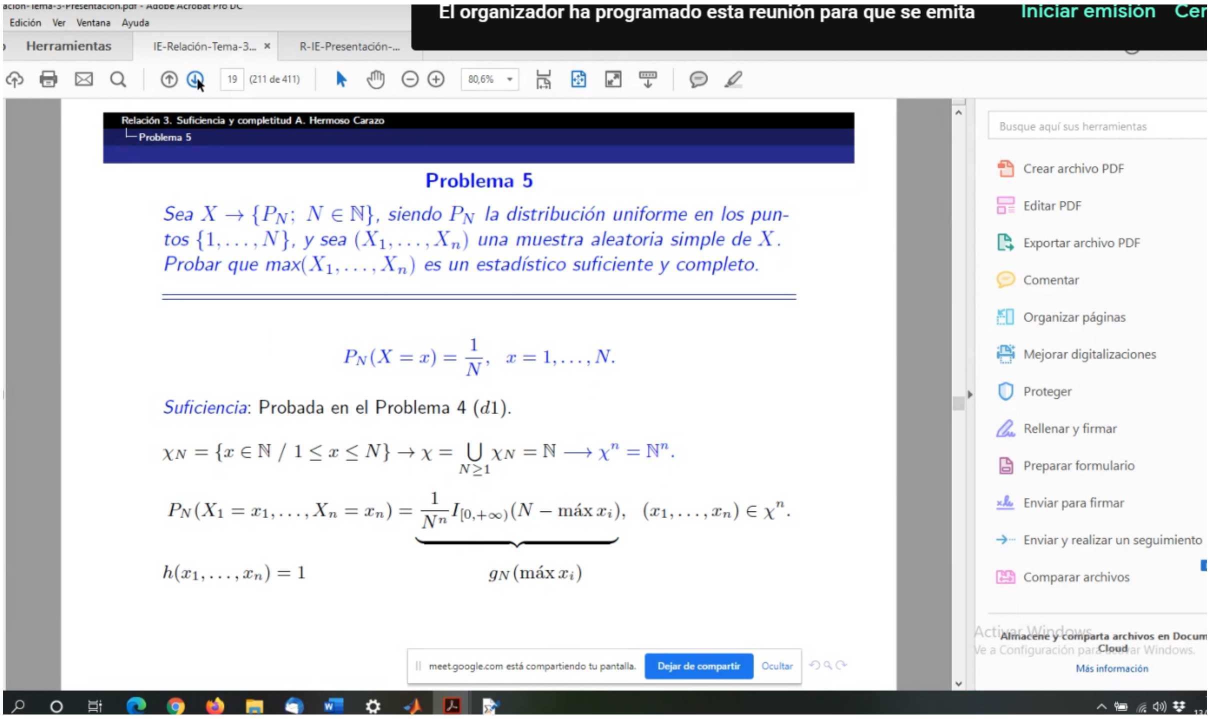
Por lo tanto, la familia es de tipo exponencial biparamétrica y si $(x_1, \dots, x_n)$ es una m.a.s tenemos:

- El estadístico $\left(\sum_{i=1}^{n} T_1(x_i), \sum_{i=1}^{n} T_2(x_i)\right) = \left(\sum_{i=1}^{n} \ln(x_i), \sum_{i=1}^{n} \ln(1-x_i)\right)$ es suficiente

- $Q(p,q) = (Q_1(p,q), Q_2(p,q)) = (p, q)$ con $p, q \in \mathbb{R}^+$

  Por tanto, $Q(\mathbb{R}^+ \times \mathbb{R}^+) = \mathbb{R}^+ \times \mathbb{R}^+$, que es un abierto de $\mathbb{R}^2$.

Así, si $n \geq 2$ (muestra de dos elementos como mínimo), $\left(\sum_{i=1}^{n} \ln(x_i), \sum_{i=1}^{n} \ln(1-x_i)\right)$, es además de suficiente, completo.



Problema 5

Sea $X \to \{P_N;\ N \in \mathbb{N}\}$, siendo P_N la distribución uniforme en los puntos $\{1, \dots, N\}$, y sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de X. Probar que $\max(X_1, \dots, X_n)$ es un estadístico suficiente y completo.

$$P_N(X = x) = \frac{1}{N}, \quad x = 1, \dots, N.$$

Suficiencia: Probada en el Problema 4 (d1).

$$\chi_N = \{x \in \mathbb{N} \ / \ 1 \leq x \leq N\} \to \chi = \bigcup_{N \geq 1} \chi_N = \mathbb{N} \longrightarrow \chi^n = \mathbb{N}^n.$$

$$P_N(X_1 = x_1, \dots, X_n = x_n) = \underbrace{\frac{1}{N^n} I_{[0,+\infty)}(N - \max x_i)}_{g_N(\max x_i)}, \quad (x_1, \dots, x_n) \in \chi^n.$$

$$h(x_1, \dots, x_n) = 1$$

Completitud: Sea g una función medible tal que $E_N\left[g(T)\right] = 0, \ \forall N \in \mathbb{N}$:

$$E_N\left[g(T)\right] = \sum_{t=1}^{N} g(t) P_N(T = t) = \sum_{t=1}^{N} g(t) \frac{t^n - (t-1)^n}{N^n} = 0, \quad \forall N \in \mathbb{N}.$$

$$\Updownarrow$$

$$G(N) = \sum_{t=1}^{N} g(t)\big(t^n - (t-1)^n\big) = 0, \quad \forall N \in \mathbb{N}.$$

$*$ $N = 1 \to g(1) = 0.$

$*$ $N \geq 2 \to G(N) - G(N-1) = g(N)\big(N^n - (N-1)^n\big) = 0 \Leftrightarrow g(N) = 0.$

Por tanto, $g(N) = 0, \ \forall N \in \mathbb{N}$ y se tiene:

$$P_N(g(T) = 0) \geq P_N(T \in \mathbb{N}) = 1, \quad \forall N \in \mathbb{N}.$$

Nota: *Si se suprime un elemento de la familia, el estadístico sigue siendo suficiente, pero deja de ser completo.* $X \to \{P_N; \ N \in \mathbb{N} - N_0\}$.

Sea g una función medible tal que $E_N\left[g(T)\right] = 0, \ \forall N \in \mathbb{N} - N_0$:

$$\Updownarrow$$

$$G(N) = \sum_{t=1}^{N} g(t)\big(t^n - (t-1)^n\big) = 0, \quad \forall N \in \mathbb{N} - N_0.$$

$*$ $N = 1 \to g(1) = 0.$

$*$ $2 \leq N \leq N_0 - 1 \to G(N) - G(N-1) = g(N)\big(N^n - (N-1)^n\big) = 0 \Rightarrow g(N) = 0.$

$*$ $N = N_0 + 1 \to G(N_0+1) = \sum_{t=N_0}^{N_0+1} g(t)\big(t^n - (t-1)^n\big) = 0.$

$*$ $N > N_0 + 1 \to G(N) - G(N-1) = g(N)\big(N^n - (N-1)^n\big) = 0 \Rightarrow g(N) = 0.$

$$\text{Sea } c \neq 0 \text{ y } g(t) = \begin{cases} c, & t = N_0 \\ -c\dfrac{N_0^n - (N_0-1)^n}{(N_0+1)^n - N_0^n}, & t = N_0 + 1 \\ 0, & N \neq N_0, N_0 + 1. \end{cases}$$

- $E_N\left[g(T)\right] = 0, \ \forall N \in \mathbb{N} - N_0.$
- $N > N_0 \Rightarrow P_N(g(T) \neq 0) = P_N(T = N_0, N_0 + 1) > 0 \Rightarrow P_N(g(T) = 0) < 1.$

Por tanto, T no es completo.