Tema:Familias Exponenciales y Aplicaciones Estadísticas

Cirilo Alvarez Rojas-UNI-FIEECS

La clase de familias de distribuciones que se introduce en este capítulo, primero fue descubierto en estadística en forma independiente por Koopman, Pitman y Darmois a través de la investigación de sus propiedades de esta familia. Subsecuentemente muchas otras características comunes de estas familias fueron descubiertas y han resultado importantes en muchos aspectos de la teoría de estadística moderna.

Los modelos de probabilidad con las características exponenciales incluye las distribuciones Normal, Binomial, Poisson, Gamma, Beta, y modelos de regresión multinomial usados para relacionar una variable respuesta Y a un conjunto de variables predicadoras. Más generalmente, estas familias forman la base de una importante clase de modelos llamados modelos lineales generalizados.

1 Modelo Exponencial de un Parámetro

La familias de distribuciones de un modelo $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$, se dice que es *una familia exponencial de un parámetro*, si existen funciones reales $\eta(\theta)$,$B(\theta)$ en Θ; funciones de valor real T y h en $\mathbb{R}^n$, tal que la función de densidad (función de cuantía o de probabilidad) $f(x;\theta)$ de $\mathscr{P}$ puede ser escrito como

$$f(x;\theta) = h(x)e^{\eta(\theta)T(x)-B(\theta)} \tag{1.1}$$

donde $x \in \mathscr{X} \subset \mathbb{R}^n$ y $\mathscr{X}$ no depende de θ. Tenga en cuenta que las funciones η,B, y T no son únicos.

En una familia exponencial de un parámetro la variable aleatoria $T(X)$ es suficiente para θ. Este es evidente debido a que solamente necesitamos identificar $e^{\eta(\theta)T(x)-B(\theta)}$ con $g(T(x);\theta)$ y $h(x)$ consigo mismo en el teorema de factorización. Denominaremos a T como un *estadístico suficiente natural* de la familia.

Damos a continuación algunos ejemplos.

Ejemplo 1. [**La familia Poisson**] Sea $\mathscr{P} = \{f_\theta, \theta \in \Theta\}$ la familia de funciones de probabilidad de Poisson con parámetro θ desconocida. Entonces, para $x \in \{0,1,2,\ldots\}$, se tiene

$$\mathbf{P}(x;\theta) = \frac{\theta^x e^{-\theta}}{x!} = \frac{1}{x!}e^{\ln(\theta)x-\theta},\ \theta > 0 \tag{1.2}$$

Por consiguiente, $\mathscr{P}$ forma una familia exponencial de un parámetro con

$$\eta(\theta) = \ln\theta, B(\theta) = \theta, T(x) = x, h(x) = \frac{1}{x!}.$$

y $T(x) = x$ es un *estadístico suficiente natural de la familia.*

Ejemplo 2. [**La familia Binomial**] Consideremos que $\mathscr{P} = \{Bin(\mathbf{n},\theta) : \theta \in (0,1) = \Theta\}$ sea una familia de funciones de probabilidades Binomiales. Entonces, para $x \in \{0,1,2,\ldots,n\}$

$$\begin{aligned} P(x;\theta) &= \binom{\mathbf{n}}{x}\theta^x(1-\theta)^{\mathbf{n}-x} \\ &= \binom{n}{x}e^{\ln\left(\frac{\theta}{1-\theta}\right)x+n\ln(1-\theta)}. \end{aligned} \tag{1.3}$$

Luego, la familia de funciones de probabilidades Binomiales es una familia exponencial de un parámetro con

$$\eta(\theta) = \ln\left(\frac{\theta}{1-\theta}\right), B(\theta) = -n\ln(1-\theta), T(x) = x, h(x) = \binom{n}{x}.$$

y $T(x) = x$ es el estadístico suficiente natural de la familia.

Ejemplo 3. [**Distribución Normal con Media Conocida**] Sea $\mathscr{P} = \{f_\theta : \theta = \sigma^2, x \in \mathbb{R}\}$, la familia de densidades normales tal que $X \sim \mathcal{N}(0,\sigma^2)$ con función de densidad

$$f(x;\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}},\ x \in \mathbb{R}.$$

Esta densidad está parametrizado por un solo parámetro σ. Escribiendo $\eta(\sigma) = -\frac{1}{2\sigma^2}$, $T(x) = x^2$, $B(\theta) = \ln\sigma$, $h(x) = 1/\sqrt{2\pi}$, se puede representar la densidad en la forma

$$f(x;\sigma) = e^{\eta(\sigma)T(x)-B(\sigma)}h(x), \forall\sigma \in \mathbb{R}^+.$$

Por consiguiente, la función de densidad f_θ pertenece a la familia exponencial de un parámetro.

Ejemplo 4. [**Distribución Normal con Varianza Conocida, σ_0^2**] Ahora sea $\mathscr{P} = \{f_\theta : \theta = \mu, X \sim \mathscr{N}(\mu, \sigma_0^2)\}$, la familia de densidades normales con función de densidad

$$\begin{aligned} f(x;\mu) &= \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma_0^2}}, \; x \in \mathbb{R}. \\ &= e^{\frac{\mu}{\sigma_0^2}x - \frac{\mu^2}{2\sigma_0^2}} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_0^2} + \ln(2\pi\sigma_0^2)\right)} \end{aligned}$$

Esta densidad está parametrizado por un solo parámetro μ ya que la varianza es conocida. Escribiendo $\eta(\mu) = \frac{\mu}{\sigma_0^2}$,
$T(x) = x, \; B(\mu) = \frac{\mu^2}{2\sigma_0^2}, \; h(x) = e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_0^2} + \ln(2\pi\sigma_0^2)\right)}$, se puede representar la densidad en la forma

$$f(x;\sigma) = e^{\eta(\mu)T(x) - B(\mu)} h(x), \forall \mu \in \mathbb{R}.$$

Por consiguiente, la función de densidad f_μ pertenece a la familia exponencial de un parámetro.

Ejemplo 5. [**Distribución Gamma de un Parámetro**] Consideremos ahora $\mathscr{P} = \{f_\theta : \theta > 0, x \sim \Gamma(\theta, 1)\}$ familia de densidades gamma con parámetro de forma θ y parámetro de escala conocida igual a 1. Entonces la función de densidad es

$$\begin{aligned} f(x;\theta) &= \frac{1}{\Gamma(\theta)} x^{\theta-1} e^{-x}, \; x > 0. \\ &= e^{\ln(\Gamma(\theta))^{-1}} e^{(\theta-1)\ln x} e^{-x} \\ &= e^{(\theta-1)\ln x - \ln\Gamma(\theta)} e^{-x} \end{aligned}$$

Escribiendo $T(x) = \ln x, \; \eta(\theta) = \theta - 1, \; B(\theta) = \ln\Gamma(\theta)$ y $h(x) = e^{-x}$, la función de densidad se puede representar como sigue:

$$f(x;\theta) = e^{\eta(\theta)T(x) - B(\theta)} h(x)$$

y por consiguiente f_θ pertenece a la familia exponencial de un parámetro. Damos a continuación un ejemplo para $n = 2$.

Ejemplo 6. Suponga que $X = (Z,Y)^t$ es un vector aleatorio donde $Y = Z + \theta W, \; \theta > 0 \; Z$ y W son variables aleatorias que se distribuyen en forma normal estándar, $\mathscr{N}(0,1)$ y son independientes. Entonces

$$\begin{aligned} f(x;\theta) &= f(z,y;\theta) = f(z)f_\theta(y|z) = \Phi(z)\Phi((y-z)\theta^{-1}) \\ &= (2\pi\theta)^{-1}e^{\left\{-\frac{1}{2}[z^2+(y-z)^2\theta^{-2}]\right\}} \\ &= (2\pi)^{-1}e^{-\frac{1}{2}z^2}e^{-\frac{1}{2}\theta^{-2}(y-z)^2-\ln\theta}. \end{aligned}$$

(Pruebe que $Y|Z \sim \mathcal{N}(Z,\theta^2)$)

Esta es una distribución de una familia exponencial de un parámetro con

$$\begin{aligned} \eta(\theta) &= -\frac{1}{2}\theta^{-2},\ B(\theta) = \ln\theta \\ T(x) &= (y-z)^2,\ h(x) = (2\pi)^{-1}e^{-\frac{1}{2}z^2}. \end{aligned}$$

2 Distribuciones Muestrales de Familias Exponenciales

Las familias de distribuciones obtenidas por muestreo de familias exponenciales de un parámetro son asimismo familias exponenciales de un parámetro. Específicamente, Suponga que $X_1, X_2, \ldots, X_m$ son independientes y idénticamente distribuidas con con función de densidad común f_θ, donde f_θ pertenece a una familia exponencial de un parámetro tal como se ha definido en la relación (1.1). Si $\{P_\theta^{(m)}\ ;\theta\in\Theta\}$, es la familia de distribuciones del vector aleatorio **m** dimensional, $X_1, X_2, \ldots, X_m$ en $\mathbb{R}^{\mathbf{m}}$ y $f(\mathbf{x};\theta)$ son las correspondientes funciones de densidades (función frecuencia o función de probabilidad) resulta que,

$$\begin{aligned} f(\mathbf{x};\theta) &= \prod_{i=1}^{m} h(x_i)e^{[\eta(\theta)T(x_i)-B(\theta)]} \\ &= e^{\left[\eta(\theta)\sum_{i=1}^{m}T(x_i)-mB(\theta)\right]}\left[\prod_{i=1}^{m}h(x_i)\right] \end{aligned} \tag{1.4}$$

donde $x = (x_1, x_2, \ldots, x_m)$. Por tanto, $f_\theta^{(\mathbf{m})}$ forma una familia exponencial de un parámetro. Si usamos superíndice m para denotar la correspondiente T, η, B y h, entonces, obtenemos

$$\eta^{(m)}(\theta) = \eta(\theta) \quad T^{(m)} = \sum_{i=1}^{\mathbf{m}} T(x_i)$$
$$B^{(\mathbf{m})}(\theta) = \mathbf{m}B(\theta) \qquad h^{(m)}(\mathbf{x}) = \prod_{i=1}^{m} h(x_i)$$

Tenga en cuenta que el estadístico suficiente natural $T^{(m)}$ es un vector de dimensión 1 cualquiera que sea m.

Por ejemplo, si $\mathbf{X} = X_1, X_2, \ldots, X_m$ es un vector de variables aleatorias independientes e idénticamente distribuidas tal que $X_i \sim \mathscr{N}(\mu, \sigma_0^2)$ con función de densidad común $f_\theta(x) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(x-\mu)^2}$, $x \in \mathbb{R}$, entonces

$$\begin{aligned}
f_\theta^{(\mathbf{m})} &= f(\mathbf{x};\mu) = f(x_1, x_2, \ldots, x_m; \mu) \\
&= \prod_{i=1}^{\mathbf{m}} \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(x-\mu)^2} \\
&= \prod_{i=1}^{\mathbf{m}} \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(x_i^2 - 2\mu x + \mu^2)} \\
&= (2\pi\sigma_0^2)^{-\frac{\mathbf{m}}{2}} e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{\mathbf{m}}(x_i^2 - 2x_i\mu + \mu^2)} \\
&= (2\pi\sigma_0^2)^{-\frac{\mathbf{m}}{2}} e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{\mathbf{m}} x_i^2} e^{\frac{\mu}{\sigma_0^2}\sum_{i=1}^{\mathbf{m}} x_i - m\frac{u^2}{2\sigma_0^2}} \\
&= e^{\frac{\mu}{\sigma^2}\sum_{i=1}^{\mathbf{m}} x_i - m\frac{u^2}{2\sigma_0^2}} e^{-\frac{1}{2}\left(\frac{\sum_{i=1}^{\mathbf{m}} x_i^2}{\sigma_0^2} + \mathbf{m}\ln(2\pi\sigma_0^2)\right)}
\end{aligned}$$

usando el superíndice **m** tenemos,

$$\eta^{(m)}(\mu) = \eta(\mu) = \frac{\mu}{\sigma_0^2}, \quad T^{(m)} = \sum_{i=1}^{m} x_i$$
$$B^m(\mu) = mB(\mu) = \mathbf{m}\frac{\mu^2}{\sigma_0^2}$$
$$h^{(m)}(\mathbf{x}) = e^{-\frac{1}{2}\left(\frac{\sum_{i=1}^{\mathbf{m}} x_i^2}{\sigma_0^2} + \mathbf{m}\ln(2\pi\sigma_0^2)\right)}$$

que es la familia de distribuciones de **X**, entonces la distribución $f_\theta^{(m)}$ forma una familia exponencial de un parámetro con estadístico suficiente natural $T^{(m)}(\mathbf{X}) = \sum_{i=1}^{m} X_i$.

El estadístico $T^{(m)}(X_1, X_2, \ldots, X_m)$ correspondiente a la familia exponencial de un parámetro de las distribuciones de una muestra de cualesquiera de los ejemplos anteriores es precisamente $\sum_{i=1}^{m} T(X_i)$.

En el ejemplo 4 el estadístico suficiente $T^{\mathbf{m}}(X_1,X_2,\ldots,X_m)=\sum_{i=1}^{\mathbf{m}} X_i$ se distribuye en forma normal, $\mathcal{N}(\mathbf{m}\mu,\mathbf{m}\sigma_0^2)$. La misma agrupación utilizada en el ejemplo 3 muestra que esta familia de distribución normal es exponencial de un parámetro cualquiera que sea $\mathbf{m}$.

Definición 1. Sea $\mathbf{X}=(X_1,X_2,\ldots,X_m)$ un *vector aleatorio muestral de dimensión* $\mathbf{m}$ con una distribución $P_\theta:\theta\in\Theta\subseteq\mathbb{R}$.

Suponga que $X_1,X_2,\ldots,X_m$ son conjuntamente continuas. La familia de distribuciones $\mathscr{P}=\{P_\theta:\theta\in\Theta\}$ se dice que pertenece a la familia exponencial de un parámetro si la función de densidad conjunta del vector muestral $\mathbf{X}=(X_1,X_2,\ldots,X_m)$ puede ser expresado en la forma

$$f_\theta(\mathbf{x})=f(\mathbf{x};\theta)=e^{\eta(\theta)T(\mathbf{x})-B(\theta)}h(\mathbf{x}),$$

si existen funciones reales $\eta(\theta),B(\theta)$ definidas en Θ, funciones reales $T(\mathbf{x})$ y $h(\mathbf{x})\geq 0$.

Si $X_1,X_2,\ldots,X_m$ son conjuntamente discretas, entonces $\{P_\theta:\theta\in\Theta\}$ si dice que pertenece a la familia exponencial de un parámetro si la función de probabilidad conjunta (función frecuencia o función cuantía) puede ser escrita en la forma

$$\begin{aligned}\mathbf{p}_\theta(\mathbf{x})=\mathbf{p}(\mathbf{x};\theta)&=\mathbf{p}_\theta(X_1=x_1,X_2=x_2,\ldots,X_{\mathbf{m}}=x_{\mathbf{m}})\\&=e^{\eta(\theta)T(\mathbf{x})-B(\theta)}h(\mathbf{x})\end{aligned}$$

si existen funciones reales $\eta(\theta),B(\theta)$ definidas en Θ, funciones reales $T(\mathbf{x})$ y $h(\mathbf{x})\geq 0$.

Debe tenerse en cuenta que las funciones η,T,B y h no son únicos.

Definición 2. Suponga que el vector muestral, $\mathbf{X}=(X_1,X_2,\ldots,X_m)$ tiene una distribución $P_\theta:\theta\in\Theta$, pertenece a la familia exponencial de un parámetro. Entonces el estadístico $T(\mathbf{X})$ se llama *estadístico suficiente natural* para la familia $\mathscr{P}=\{P_\theta:\theta\in\Theta\}$.

La noción de un estadístico suficiente es uno de los fundamentos en la teoría estadística y sus aplicaciones. El concepto de suficiencia fue introducido en la literatura estadística por Sir Ronald A. Fisher (Fisher (1922)). La suficiencia intenta formalizar la noción de *no pérdida de información*. Un estadístico suficiente es supuesto a que contiene por si mismo toda la información acerca de los parámetros desconocidos de la distribución subyacente que la muestra completa podría haber proporcionado. En ese sentido, no hay ninguna pérdida al restringir la atención solo a una estadística suficiente en el proceso de inferencia que uno realiza. Sin embargo, la forma de un estadístico suficiente es muy dependiente tanto de la elección de una distribución particular P_θ para modelar la variable aleatoria observable X. Sin embargo, la reducción a la estadística suficiente en modelos ampliamente utilizados por lo general hace que el sentido común sea simple. Vamos a volver a la cuestión de la suficiencia de nuevo más adelante en este capítulo.

Ahora veamos algunos ejemplos de distribuciones más comunes que pertenecen a la familia exponencial de un parámetro para un vector muestral.

Ejemplo 7. [**Distribución binomial**] Sea $X \sim Bin(\mathbf{n}, \theta)$ con $\mathbf{n} \geq 1$ considerado conocido, y $0 < \theta < 1$ es el parámetro proporción poblacional. Sea $X_1, X_2, \ldots, X_m$ una muestra aleatoria simple de esta distribución, entonces del ejemplo 2 tenemos

$$\mathbf{p}(x;\theta) = \binom{\mathbf{n}}{x} e^{\ln\left(\frac{\theta}{1-\theta}\right)x + \mathbf{n}\ln(1-\theta)}.$$

luego la función de densidad de la muestra será

$$\begin{aligned}
\mathbf{p}(x_1, x_2, \ldots, x_m; \theta) &= \prod_{i=1}^{\mathbf{m}} \mathbf{p}(x_i;\theta) \\
&= \prod_{i=1}^{\mathbf{m}} \binom{\mathbf{n}}{x_i} e^{\ln\left(\frac{\theta}{1-\theta}\right)x_i + \mathbf{n}\ln(1-\theta)} \\
&= e^{\ln\left(\frac{\theta}{1-\theta}\right)\sum_{i=1}^{\mathbf{n}} x_i + \mathbf{mn}\ln(1-\theta)} \prod_{i=1}^{\mathbf{m}} \binom{\mathbf{n}}{x_i}
\end{aligned}$$

Escribiendo

$$\begin{aligned}
\eta(\theta) &= \ln\left(\frac{\theta}{1-\theta}\right), & T(\mathbf{x}) &= \sum_{i=1}^{\mathbf{m}} x_i \\
B(\theta) &= -\mathbf{mn}\ln(1-\theta), & h(\mathbf{x}) &= \prod_{i=1}^{\mathbf{m}} \binom{\mathbf{n}}{x_i}
\end{aligned}$$

luego la función frecuencia se puede representa en la forma

$$\mathbf{p}(\mathbf{x};\theta) = e^{\eta(\theta)T(\mathbf{x}) - B(\theta)} h(\mathbf{x})$$

se observa que $\mathbf{p}(\mathbf{x};\theta)$ pertenece a la familia exponencial de un parámetro.

Ejemplo 8. [**Distribución Poisson**] Sea $X \sim Poisson(\theta)$ entonces del ejemplo 1 sabemos que la función de densidad de X pertenece a la familia exponencial de un parámetro cuya densidad es

$$\mathbf{p}(x;\theta) = \frac{1}{x!} e^{\ln(\theta)x - \theta},\ \theta > 0$$

luego la función de densidad de la muestra será

$$\begin{aligned}\mathbf{p}(\mathbf{x};\theta) &= \prod_{i=1}^{\mathbf{m}} \frac{1}{x_i!} e^{\ln(\theta)x_i-\theta} \\ &= e^{\ln(\theta)\sum_{i=1}^{\mathbf{m}} x_i - \mathbf{m}\theta} \prod_{i=1}^{\mathbf{m}} \frac{1}{x_i!}.\end{aligned}$$

Otra vez escribiendo,

$$\begin{aligned}\eta(\theta) &= \ln\theta, \quad T(\mathbf{x}) = \sum_{i=1}^{\mathbf{m}} x_i \\ B(\theta) &= \mathbf{m}\theta, \quad h(\mathbf{x}) = \prod_{i=1}^{\mathbf{m}} \frac{1}{x_i!}\end{aligned}$$

luego la función frecuencia se puede representa en la forma

$$\mathbf{p}(\mathbf{x};\theta) = e^{\eta(\theta)T(\mathbf{x})-B(\theta)} h(\mathbf{x})$$

se observa que $\mathbf{p}(\mathbf{x};\theta)$ pertenece a la familia exponencial de un parámetro.

Ejemplo 9. [**Distribución Normal con varianza Conocida,**σ_0^2] Si $X \sim \mathcal{N}(\mu,\sigma_0^2)$, entonces la función de densidad de X es

$$f(x;\mu) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(x-\mu)^2}$$

que expresando en forma exponecial resulta

$$f_\mu(x) = e^{\frac{\mu}{\sigma_0^2}x - \frac{\mu^2}{2\sigma_0^2}} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_0^2}+\ln(2\pi\sigma_0^2)\right)}$$

escribiendo la función de densidad conjunta para una muestra aleatoria simple $X_1, X_2, \ldots, X_m$, tenemos

$$f(\mathbf{x};\theta) = e^{\frac{\mu}{\sigma_0^2}\sum_{i=1}^{\mathbf{m}} x_i - \mathbf{m}\frac{\mu^2}{2\sigma_0^2}} \prod_{i=1}^{\mathbf{m}} e^{-\frac{1}{2}\left[\frac{x_i^2}{\sigma_0^2}+\ln(2\pi\sigma_0^2)\right]}$$

haciendo

$$\eta(\mu) = \frac{\mu}{\sigma_0^2}, B(\theta) = \mathbf{m}\frac{\mu^2}{2\sigma_0^2} \text{ y } h(\mathbf{x}) = \prod_{i=1}^{\mathbf{m}} e^{-\frac{1}{2}\left[\frac{x_i^2}{\sigma_0^2}+\ln(2\pi\sigma_0^2)\right]}$$

se observa que pertenece a la familia exponencial de un parámetro otra vez.

Ejemplo 10. [**Distribución** $\Gamma(\alpha,1)$] Expresamos la función de densidad conjunta para la muestra aleatoria $X_1, X_2, \ldots, X_m$ que proviene de una población Gamma con densidad

$$f(x;\alpha) = \frac{1}{\Gamma(\theta)} x^{\alpha-1} e^{-x},\ x \in \mathbb{R}.$$

Del ejemplo 5se sabe que $f(x;\alpha)$ pertenece a la familia exponencial de un parámetro cuya densidad es

$$f(x;\alpha) = e^{(\alpha-1)\ln x - \ln\Gamma(\alpha)} e^{-x}.$$

La función de densidad conjunta para la muestra aleatoria $X_1, X_2, \ldots, X_m$ resulta

$$\begin{aligned} f(\mathbf{x};\alpha) &= \prod_{i=1}^{\mathbf{m}} e^{(\alpha-1)\ln x_i - \ln\Gamma(\alpha)} e^{-x_i} \\ &= e^{(\alpha-1)\sum_{i=1}^{\mathbf{m}} \ln x_i - m\ln\Gamma(\alpha)} e^{-\sum_{i=1}^{\mathbf{m}} x_i}. \end{aligned}$$

Escribiendo

$$\begin{aligned} \eta(\alpha) &= (\alpha-1) & T(\mathbf{x}) &= \sum_{i=1}^{\mathbf{m}} \ln x_i \\ B(\alpha) &= \mathbf{m}\ln\Gamma(\alpha) & h(\mathbf{x}) &= e^{-\sum_{i=1}^{\mathbf{m}} x_i} \end{aligned}$$

se observa que $f(\mathbf{x};\alpha)$ pertenece a la familia exponencial de un parámetro.

Algunas Distribuciones que no son de la Familia Exponencial

Se deduce de la definición de una familia exponencial de un parámetro que si una determinada familia de distribuciones $P_\theta : \theta \in \Theta$ pertenece a la familia exponencial de un parámetro, entonces cada P_θ tiene exactamente el mismo soporte. Precisamente, para cualquier θ fijo, se cumple que $P_\theta(A) > 0$ si y solo si $\int_A h(x)dx > 0$, y en el caso discreto, $P_\theta(A) > 0$ si y solo si $A \cap \mathscr{X} \neq \emptyset$, donde $\mathscr{X}$ es el conjunto numerable $\mathscr{X} = \{x : h(x) > 0\}$. Como consecuencia de este hecho de soporte común, las denominadas distribuciones irregulares cuyo soporte depende del parámetro no pueden ser miembros de la familia exponencial. Ejemplos de estas familias son las familias de distribuciones $U[0,\theta]$, $U[\theta, 2\theta]$, etc. Del mismo modo, la densidad exponencial $f(x;\theta) = e^{\theta - x} I_{x>\theta}$ no puede estar en la familia exponencial.

Algunas otras distribuciones comunes también no son de la familia exponencial, pero por otras razones. Un ejemplo importante es la familia de distribuciones de Cauchy dada por el parámetro de centralización $f(x|\mu) = \frac{1}{1+\pi((x-\mu)^2)} I_{x \in \mathbb{R}}$. Supongamos que se trata esta densidad. Entonces, podemos encontrar funciones $\eta(\mu)$, $T(x)$ tal que para todo $x, \mathbf{m}$,

$$
\begin{aligned}
e^{\eta(\mu)T(x)} = \frac{1}{1+(x-\mu)^2} &\Longrightarrow \eta(\mu)T(x) - \ln(1+(x-\mu)^2) \\
&\Longrightarrow \eta(0)T(x) = -\ln(1+x^2) \\
&\Longrightarrow T(x) = c\ln(1+x^2)
\end{aligned}
$$

para alguna constante c.

Tomando este resultado de nuevo, obtenemos para todo $x.\mu$

$$
\begin{aligned}
c\eta(\mu)\ln(1+x^2) &= \ln(1+(x-\mu)^2) \\
&\Longrightarrow \eta(\mu) = \frac{1}{c}\frac{\ln(1+(x-\mu)^2)}{\ln(1+x^2)}.
\end{aligned}
$$

Este significa que $\frac{\ln(1+(x-\mu)^2)}{\ln(1+x^2)}$ debe ser una función constante de x, que es una contradicción. La elección de $\mu = 0$ como el valor especial de μ no es importante.

3 Forma Canónica y Propiedades Básicas

Suponga que $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ sea una familia exponencial de un parámetro, con función densidad (o función frecuencia) de la forma $f(x|\theta) = e^{\eta(\theta)T(x)-B(\theta)}h(x)$. Si $\eta(\theta)$ es una función uno a uno de θ, entonces podemos dejar θ por completo, y parametrizar la distribución en términos de η mismo. Si hacemos eso, obtenemos una densidad g reparametrizada en la forma $e^{\eta T(x)-B^*(\eta)}h(x)$. Por un ligero abuso de notación, de nuevo usaremos la notación f para g y B para B^*.

Definición 3. Sea $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ un vector aleatorio muestral de dimensión $\mathbf{m}$ que se distribuye según $P_\eta : \eta \in \tau \subseteq \mathbb{R}$. La familia de distribuciones $\mathscr{P} = \{P_\eta : \eta \in \tau\}$ se dice que pertenece a la *familia exponencial canónico de un parámetro* si la densidad (función cuantía) de P_η puede ser escrita en la forma

$$
f(x;\eta) = e^{\eta T(\mathbf{x})-B(\eta)}h(x)
$$

donde

$$\eta \in \tau = \{\eta : e^{B(\eta)} = \int_{\mathbb{R}^{\mathbf{m}}} e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x} < \infty\}$$

en el caso continuo, y

$$\tau = \{\eta : e^{B(\eta)} = \sum_{\mathbf{x} \in \mathscr{X}} e^{\eta T(\mathbf{x})} h(\mathbf{x}) < \infty\}$$

en caso discreto, siendo $\mathscr{X}$ un conjunto contable en la cual $h(\mathbf{x}) > 0$.

Para una distribución en una familia exponencial canónica de un parámetro, el parámetro η se llama *parámetro natural*, y τ se llama *espacio paramétrico natural*. Tenga en cuenta que τ describe el conjunto más grande de los valores de η para el cual la función de densidad (función frecuencia) se puede definir. En una aplicación particular, se puede tener conocimientos extraños de que η pertenece a algún subconjunto propio de τ. De esta manera, el conjunto $\{P_\eta\}$ tal que $\eta \in \tau$ se llama familia exponencial canónico completo de un parámetro. En general, nos referimos a la familia completa, a menos que se indique lo contrario.

La familia exponencial canónica se llama regular si τ es un conjunto abierto en $\mathbb{R}$, y se llama no singular si $\mathbf{Var}_\eta(T(\mathbf{X})) > 0$ para todo $\eta \in \tau^0$, el interior del espacio paramétrico natural.

Analíticamente es conveniente trabajar con una distribución exponencial de la familia en su forma canónica. Una vez que un resultado se ha derivado de la forma canónica, si se desea se puede volver a escribir la respuesta en términos del parámetro original θ. Hacer esta retransformación al final es algebraicamente y notacionalmente más simple que llevar la función original $\eta(\theta)$ y a menudo sus derivadas de orden superior nos conduce por medio de unos cálculos. La mayoría de las fórmulas y teoremas siguientes se dará respecto a la forma canónica.

Ejemplo 11. [**Distribución Binomial en forma Canónica**] Sea $X \sim Bin(n, \theta)$ con función de probabilidad dada por

$$P_\theta(X = x) = \binom{\mathbf{n}}{x} \theta^x (1-\theta)^{\mathbf{n}-x} \quad x \in \{0, 1, 2, \ldots, \mathbf{n}\}$$

que, expresado en forma exponencial en el ejemplo 2 resula

$$P_\theta(X=x) = e^{\ln\left(\frac{\theta}{1-\theta}\right)x+\mathbf{n}\ln(1-\theta)}\binom{\mathbf{n}}{x}.$$

$$\text{si } \ln\frac{\theta}{1-\theta} = \eta \Rightarrow \frac{\theta}{1-\theta} = e^\eta \Rightarrow \theta = \frac{e^\eta}{1+e^\eta} \text{ y } 1-\theta = \frac{1}{1+e^\eta}.$$

Por tanto, la forma de la familia exponencial canónica de la distribución binomial es

$$P_\theta(X=x) = e^{\eta x-\mathbf{n}\ln(1+e^\eta)}\binom{\mathbf{n}}{x}, x\in\{0,1,2,3,\ldots,\mathbf{n}\}.$$

y el espacio paramétrico natural es $\tau = \mathbb{R}$.

Propiedades de Convexidad

Cuando se escribe en su forma canónica, una densidad (o función frecuencia) en una la familia exponencial tiene algunas propiedades de convexidad. Estas propiedades de convexidad son útiles en las operaciones con momentos y otras funciones de $T(\mathbf{X})$, los estadísticos suficientes naturales que aparecen en las expresiones de la densidad de la distribución.

Teorema 1. El espacio paramétrico natural τ es convexo, y $B(\eta)$ es una función convexa en τ,

Prueba. Consideremos solo el caso continuo, ya que el caso discreto admite básicamente la misma prueba. Sean η_1 y η_2 dos elementos del espacio paramétrico τ, y sea $o<\alpha<1$. Necesitamos probar que $\alpha\eta_1+(1-\alpha)\eta_2$ pertenezca a τ; es decir,

$$\int_{\mathbb{R}^d} e^{(\alpha\eta_1+(1-\alpha)\eta_2)T(\mathbf{x})}h(\mathbf{x})d\mathbf{x} < \infty$$

pero,

$$\begin{aligned}
\int_{\mathbb{R}^d} e^{(\alpha\eta_1+(1-\alpha))T(\mathbf{x})}h(\mathbf{x})d\mathbf{x} &= \int_{\mathbb{R}^d} e^{(\alpha\eta_1)T(\mathbf{x})}e^{(1-\alpha)\eta_2T(\mathbf{x})}h(\mathbf{x})d\mathbf{x} \\
&= \int_{\mathbb{R}^d}\left(e^{\eta_1T(\mathbf{x})}\right)^\alpha\left(e^{\eta_2T(\mathbf{x})}\right)^{(1-\alpha)}h(\mathbf{x})d\mathbf{x} \\
&\leq \left(\int_{\mathbb{R}^d} e^{\eta_1T(\mathbf{x})}h(\mathbf{x})d\mathbf{x}\right)^\alpha\left(\int_{\mathbb{R}^d} e^{\eta_2T(\mathbf{x})}h(\mathbf{x})d\mathbf{x}\right)^{(1-\alpha)}
\end{aligned}$$

por la desigualdad de Hotder, resulta

$$\leq \left(\int_{\mathbb{R}^d} e^{\eta_1 T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}\right)^{\alpha} \left(\int_{\mathbb{R}^d} e^{\eta_2 T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}\right)^{(1-\alpha)} < \infty$$

debido a que, por hipótesis $\eta_1, \eta_2 \in \tau$, y en consecuencia

$$\int_{\mathbb{R}^d} e^{\eta_1 T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x} \text{ y } \int_{\mathbb{R}^d} e^{\eta_2 T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}$$

ambos son finitos.

Observe que en este argumento, en realidad se ha demostrado la desigualdad

$$e^{B(\alpha\eta_1+(1-\alpha)\eta_2)} \leq e^{\alpha B(\eta_1)+(1-\alpha)B(\eta_2)}.$$

Pero, este es lo mismo afirmar que

$$B(\alpha\eta_1 + (1-\alpha)\eta_2) \leq \alpha B(\eta_1) + (1-\alpha)B(\eta_2)$$

es decir, $B(\eta)$ es una función convexa en τ

$\square$

4 Momentos y función Característica

El siguiente resultado es un hecho muy especial acerca de la familia exponencial canónica, y es el origen de un gran número de fórmulas válidas de forma cerrada para toda la familia exponencial canónica. El hecho en sí es realmente un hecho en el análisis matemático. Debido a la forma especial de densidades de la familia exponencial hecho en análisis se traduce en resultados para la familia, una instancia de interacción entre las matemáticas y la probabilidad.

Teorema 2. (a) La función $e^{B(\eta)}$ es infinitamente diferenciable en todo $\eta \in \tau^0$. Además, en el caso continuo, $e^{B(\eta)} = \int_{\mathbb{R}^d} e^{\eta_1 T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}$ puede ser diferenciado cualquier número de veces dentro del signo de la integral, y en el caso discreto, $e^{B(\eta)} = \sum_{\mathbf{x} \in \mathscr{X}} e^{\eta T(\mathbf{x})} h(\mathbf{x})$ puede ser diferenciado cualquier número de veces dentro del signo de sumatoria.

(b) En el caso continuo, para cualquier $k \geq 1$

$$\frac{d^k}{d\eta^k} e^{B(\eta)} = \int_{\mathbb{R}^d} [T(\mathbf{x})]^k e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}$$

y en el caso discreto,

$$\frac{d^k}{d\eta^k} e^{B(\eta)} = \sum_{\mathbf{x} \in \mathscr{X}} [T(\mathbf{x})]^k e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}$$

Prueba. Consideremos $k = 1$. Entonces por la definición de la derivada de una función, $\frac{d}{d\eta} e^{B(\eta)}$ existe si y solo si $\lim\limits_{\delta \longrightarrow 0} \frac{e^{B(\eta+\delta)} - e^{B(\eta)}}{\delta}$ existe. Pero,

$$\lim_{\delta \longrightarrow 0} \frac{e^{B(\eta+\delta)} - e^{B(\eta)}}{\delta} = \int_{\mathbb{R}^d} \frac{e^{B(\eta+\delta)T(\mathbf{x})} - e^{B(\eta)T(\mathbf{x})}}{\delta} h(\mathbf{x}) d\mathbf{x},$$

y por la aplicación del teorema de la convergencia dominada $\int_{\mathbb{R}^d} \frac{e^{B(\eta+\delta)T(\mathbf{x})} - e^{B(\eta)T(\mathbf{x})}}{\delta} h(\mathbf{x}) d\mathbf{x}$, existe,y el límite puede ser llevado acabo dentro de la integral para obtener

$$\begin{aligned} &\lim_{\delta \longrightarrow 0} \int_{\mathbb{R}^d} \frac{e^{B(\eta+\delta)T(\mathbf{x})} - e^{B(\eta)T(\mathbf{x})}}{\delta} h(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^d} \lim_{\delta \longrightarrow 0} \frac{e^{B(\eta+\delta)T(\mathbf{x})} - e^{B(\eta)T(\mathbf{x})}}{\delta} h(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^d} \frac{d}{d\eta} e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^d} T(\mathbf{x}) e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x} \end{aligned}$$

En seguida hacemos inducción sobre k utilizando el teorema de la convergencia dominada otra vez.

$\square$

Esta fórmula compacta para una derivada arbitraria de $e^{B(\eta)}$ conduce a las siguientes fórmulas importantes de momentos.

Teorema 3. En cualquier $\eta \in \tau$

(a) $E_\eta[T(\mathbf{X})] = B'(\eta)$; $Var_\eta(T(X)) = B''(\eta)$

(b) Los coeficientes de asimetría y curtosis de $T(\mathbf{X})$ son iguales a

$$\beta(\eta) = \frac{B^{(3)}(\eta)}{[B''(\eta)]^{3/2}} \qquad \gamma(\eta) = \frac{B^{(4)}(\eta)}{[B''(\eta)]^2}$$

(c) En cualquier $\eta + t \in \mathscr{T}$, la función característica existe de $T(\mathbf{X})$ existe y es igual a

$$\Phi_\eta(t) = e^{B(it+\eta)-B(\eta)}$$

Prueba. (a) Como antes tratemos el caso continuo. Consideremos el resultado del teorema anterior que para cualquier $k \geq 1$ se cumple:

$$\frac{d^k}{d\eta^k} e^{B(\eta)} = \int_{\mathbb{R}^d} [T(\mathbf{x})]^k e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}.$$

usando para $k = 1$, se obtiene

$$B'(\eta) e^{B(\eta)} = \int_{\mathbb{R}^d} T(\mathbf{x}) e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}$$

multiplicando ambos miembros de la ecuación anterior por $e^{-B(\eta)}$ se obtiene

$$B'(\eta) = \int_{\mathbb{R}^d} T(\mathbf{x}) e^{\eta T(\mathbf{x}) - B(\eta)} h(\mathbf{x}) d\mathbf{x}$$

que da el resultado

$$E_\eta[T(\mathbf{X})] = B'(\eta)$$

De la misma manera se tiene

$$\begin{aligned}
\frac{d^2}{d\eta^2} e^{B(\eta)} &= \int_{\mathbb{R}^d} [T(\mathbf{x})]^2 e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}. \\
\left(B'(\eta) e^{B(\eta)}\right)' &= \int_{\mathbb{R}^d} [T(\mathbf{x})]^2 e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}. \\
\left(B''(\eta) + \{B(\eta)\}^2\right) e^{B(\eta)} &= \int_{\mathbb{R}^d} [T(\mathbf{x})]^2 e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

de donde se obtiene

$$B''(\eta) + \{B(\eta)\}^2 = \int_{\mathbb{R}^d} [T(\mathbf{x})]^2 e^{\eta T(\mathbf{x}) - B(\eta)} h(\mathbf{x}) d\mathbf{x}.$$

la cual da

$$E_\eta[T(\mathbf{X})]^2 = B''(\eta) + \{B(\eta)\}^2$$

combinando este resultado con lo ya obtenido $E_\eta[T(\mathbf{X})] = B'(\eta)$ se obtiene

$$Var_\eta[T(\mathbf{X})] = E_\eta[T(\mathbf{X})]^2 - \{E_\eta[T(\mathbf{X})]\}^2 = B''(\eta)$$

(b) El coeficiente de asimetría se define como

$$\beta_\eta = \frac{E[T(\mathbf{X}) - E(T(\mathbf{X}))]^3}{(Var[T(\mathbf{X})])^{3/2}}$$

donde se tiene la siguiente igualdad

$$\begin{aligned} E[T(\mathbf{X}) - E(T(\mathbf{X}))]^3 = E[T(\mathbf{X})]^3 - 3E[T(\mathbf{X})]^2 E[T(\mathbf{X})] \\ +2\{E[T(\mathbf{X})]\}^3 \end{aligned} \tag{1.5}$$

utilizando la identidad

$$\frac{d^3}{d\eta^3} e^{B(\eta)} = \int_{\mathbb{R}^d} [T(\mathbf{x})]^3 e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}.$$

se obtiene la tercera derivada puesto que la segunda derivada de $e^{B(\eta)}$ ya se conoce al calcular la varianza de $T(\mathbf{X})$, y que resulta como sigue:

$$\left[\left(B''(\eta) + \{B(\eta)\}^2\right) e^{B(\eta)}\right]' = \int_{\mathbb{R}^d} [T(\mathbf{x})]^3 e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}.$$

y derivando el lado izquierdo de la expresión anterior se obtiene

$$\left[B^{(3)}(\eta) + 3B'(\eta)B''(\eta) + \{B'(\eta)\}^3\right] e^{B(\eta)} = \int_{\mathbb{R}^d} [T(\mathbf{x})]^3 e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}.$$

multiplicando ambos miembros de en la expresión anterior resulta

$$B^{(3)}(\eta) + 3B'(\eta)B''(\eta) + \{B'(\eta)\}^3 = \int_{\mathbb{R}^d} [T(\mathbf{x})]^3 e^{\eta T(\mathbf{x}) - B(\eta)} h(\mathbf{x}) d\mathbf{x};$$

es decir,

$$E[T(\mathbf{X})]^3 = B^{(3)}(\eta) + 3B'(\eta)B''(\eta) + \{B'(\eta)\}^3$$

Luego reemplazando los valores respectivos de los momentos se obtiene que

$$E[T(\mathbf{X}) - E(T(\mathbf{X}))]^3 = B^{(3)}(\eta)$$

y la fórmula de la asimetría resulta lo establecido en el ítem (b).

La fórmula del coeficiente curtosis se obtiene procediendo de la misma manera, usando $k = 4$ en la identidad de la derivada

$$\frac{d^4}{d\eta^4} e^{B(\eta)} = \int_{\mathbb{R}^d} [T(\mathbf{x})]^4 e^{\eta T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}.$$

(c) La función generatriz de momento se obtiene como sigue:

$$\begin{aligned}\Phi_\eta(t) &= E(e^{itT(\mathbf{X})}) = \int_{\mathbb{R}^d} e^{itT(X)} e^{\eta T(\mathbf{x})-B(\eta)} h(\mathbf{x}) d\mathbf{x}.\\ &= e^{-B(\eta)} \int_{\mathbb{R}^d} e^{(it+\eta)T(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}.\\ &= e^{-B(\eta)} e^{B(it+\eta)} \int_{\mathbb{R}^d} e^{(it+\eta)T(\mathbf{x})-B(it+\eta)} h(\mathbf{x}) d\mathbf{x}.\\ \Phi_\eta(t) &= e^{B(it+\eta)-B(\eta)}\end{aligned}$$

Una consecuencia importante de las fórmulas de la media y varianza es el siguiente resultado de la monotonicidad □

Corolario 1. Para una familia exponencial no singular, $E_\eta[T(\mathbf{X})]$ es estrictamente creciente en interior $\eta \in \tau^0$

Prueba. De la parte (a) del teorema **3**, la varianza de $T(\mathbf{X})$ es la derivada de la esperanza de $T(\mathbf{X})$, y por ser no singular la varianza es estrictamente creciente. Esto implica que la esperanza es estrictamente creciente. □

Como una consecuencia de esta monotonicidad creciente de la esperanza de $T(\mathbf{X})$ en el parámetro natural, de las familias exponenciales canónicas no -singulares puede ser re - parametrizado usando la misma media de $T(\mathbf{X})$como parámetro. esto es útil para algunos propósitos.

Ejemplo 12. [**Distribución Binomial**] Sabemos del ejemplo 11, en la representación canónica de la distribución binomial, $B(\eta) = \mathbf{n}\ln(1+e^\eta)$. Por diferenciación directa se obtiene

$$B'(\eta) = \frac{\mathbf{n}e^\eta}{1+e^\eta}; \qquad B''(\eta) = \frac{\mathbf{n}e^\eta}{(1+e^\eta)^2}$$

$$B^{(3)}(\eta) = -\frac{\mathbf{n}e^\eta(e^\eta-1)}{(1+e^\eta)^3}; \quad B^{(4)}(\eta) = \frac{\mathbf{n}e^\eta(e^{2\eta}-4e^\eta+1)}{(1+e^\eta)^4}$$

Teniendo en cuenta el ejemplo 11 que el parámetro θ y el parámetro natural η están relacionados como $\theta = \frac{e^\eta}{1+e^\eta}$. Usando este y las fórmulas del Teorema **3**, podemos reescribir la media, varianza, los coeficientes de asimetría y de curtosis de X como

$$E(X) = \mathbf{n}\theta,\ Var(X) = \mathbf{n}\theta(1-\theta),\ \beta_\theta = \frac{1-2\theta}{\sqrt{\mathbf{n}\theta(1-\theta)}} \text{ y } \gamma_\theta = \frac{\frac{1}{\theta(1-\theta)}-6}{\mathbf{n}}$$

Para completar, es útil tener las fórmulas de la media y de la varianza en una parámetros originales, que se exponen a continuación. La prueba se sigue de una aplicación del Teorema **3** y de la regla de la cadena.

Teorema 4. Sea $\{P_\theta : \theta \in \Theta\}$ una familia de distribuciones de una familia de distribuciones exponenciales de un parámetro con densidad (función frecuencia)

$$f(x|\theta) = e^{\eta(\theta)T(x)-B(\eta)}h(x)$$

Entonces, para cualquier θ en la que $\eta'(\theta) \neq 0$

$$E_\theta[T(X)] = \frac{B'(\theta)}{\eta'(\theta)}; \quad Var_\theta(X) = \frac{B''(\theta)}{[\eta'(\theta)]^2} - \frac{B'(\theta)\eta''(\theta)}{[\eta'(\theta)]^3}$$

5 Propiedad de la Clausura

La familia exponencial cumple una serie de propiedades importantes de clausura. Por ejemplo, si un vector aleatorio de dimensión m, $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ se distribuye según una familia exponencial, entonces la distribución condicional de cualquier sub -vector dado el resto también se distribuye como una familia exponencial. Existen un número de tales propiedades de clausura, de los cuales solo vamos a discutir cuatro.

Primero, si $\mathbf{X} = X_1, X_2, \ldots, X_m$ se distribuye según la familia exponencial, entonces el estadístico suficiente natural, $T(\mathbf{X})$ también se distribuye según la familia exponencial. La verificación de este en general no puede realizarse sin utilizar la teoría de la medida. Sin embargo, fácilmente podemos demostrar esto en algunos casos particulares. Consideremos el caso continuo con $\mathbf{m} = 1$ y supongamos que $T(\mathbf{X})$ es una función uno a uno diferenciable de $\mathbf{X}$. Entonces por la fórmula del Jacobiano, $T(\mathbf{X})$ función de densidad

$$f_T(t|\eta) = e^{\eta t - B(\eta)} \frac{h\left(T^{-1}(t)\right)}{|T'\left(T^{-1}(t)\right)|}$$

Esto es una vez más está en la forma de la familia exponencial un parámetro, con el mismo estadístico suficiente natural T, y la función B sin cambios. La función h ha cambiado a una nueva función

$$h^*(t) = \frac{h\left(T^{-1}(t)\right)}{|T'\left(T^{-1}(t)\right)|}$$

Similarmente, en el caso discreto, la función de probabilidad (fp) de $T(\mathbf{X})$ está dado por

$$\mathbf{p}_\eta(T(\mathbf{X}) = t) = \sum_{\mathbf{x}:g(\mathbf{x})=t} e^{\eta t - B(\eta)}h(\mathbf{x}) = e^{\eta t - B(\eta)}h^*(t),$$

donde $h^*(t) = \sum_{\mathbf{x}:g(\mathbf{x})=t} h(\mathbf{x})$.

Supongamos en seguida que, el vector aleatorio $\mathbf{X} = X_1, X_2, \ldots, X_m$ tiene una densidad (fd) $f(\mathbf{x}|\eta)$ en la familia exponencial y $\mathbf{Y} = Y_1, Y_2, \ldots, Y_n$ son $\mathbf{n}$ observaciones iid de esta densidad $f(\mathbf{x}|\eta)$. Tenga en cuenta que cada observación individual Y_i es un vector de dimensión $\mathbf{m}$. La función de densidad conjunta del vector $\mathbf{Y} = Y_1, Y_2, \ldots, Y_n$ es

$$\begin{aligned} f(\mathbf{y}|\eta) &= \prod_{i=1}^{\mathbf{n}} f(y_i|\eta) = \prod_{i=1}^{\mathbf{n}} e^{\eta T(y_i) - B(\eta)} h(y_i) \\ &= e^{\eta \sum_{i=1}^{\mathbf{n}} T(y_i) - nB(\eta)} \prod_{i=1}^{\mathbf{n}} h(y_i). \end{aligned}$$

Reconocemos que esto está otra vez la forma de una familia exponencial de parámetro, con los estadísticos suficientes naturales $\sum_{i=1}^{\mathbf{n}} T(Y : i)$, la nueva función B como $\mathbf{n}B$, y la nueva función h como $\prod_{i=1}^{\mathbf{n}} h(y_i)$. La función de densidad conjunta $\prod_{i=1}^{\mathbf{n}} f(y_i|\eta)$ es conocida como la *función de verosimilitud* en las estadísticas. Así funciones de *verosimilitud* obtenidos a partir de una muestra iid de una distribución de la familia exponencial un parámetro también son miembros de la familia exponencial de un parámetro.

Las propiedades de clausura descritos en lo anterior se encuentran formalmente establecidas en el siguiente teorema.

Teorema 5. Considere $\mathbf{X} = X_1, X_2, \ldots, X_m$ un vector aleatorio cuya distribución pertenece a la familia exponencial de un parámetro con estadístico suficiente natural $T(\mathbf{X})$.

(a) El estadístico, $T = T(X)$ también tiene una distribución que pertenece a la familia exponencial de un parámetro.

(b) Sea $\mathbf{Y} = \mathbf{A}\mathbf{X} + u$ es una transformación no singular de $\mathbf{X}$. Entonces $\mathbf{Y}$ también tiene una distribución que pertenece a la familia exponencial de un parámetro.

(c) Sea I_0 cualquier subconjunto propio de $I\{1, 2, \ldots, \mathbf{m}\}$. Entonces la distribución conjunta condicional de $X_i, i \in I_o$ dado $X_j, j \in I - I_0$ también pertenece a la familia exponencial de un parámetro.

(d) Para un $\mathbf{n} \geq 1$ dado, suponga que $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias iid con la misma distribución de $\mathbf{X}$. Entonces la distribución conjunta de $(Y_1, Y_2, \ldots, Y_n)$ también pertenece a la familia exponencial de un parámetro.

6 Familia Exponencial Multiparamétrico

Similar al caso de distribuciones con un solo parámetro, varias distribuciones comunes con múltiples parámetros también pertenecen a una familia exponencial multiparamétrico general. Un ejemplo es la distribución normal en $\mathbb{R}$ con ambos parámetros desconocidos. Otro ejemplo es el de una distribución normal multivariada. Propiedades técnicas y analíticos de familias exponenciales multiparamétricos son muy similares a los de la familia exponencial de un parámetro. Por esa razón, la mayor parte de nuestra presentación en esta sección se centra en ejemplos.

Definición 4. Sea $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ tiene una distribución en la familia de distribuciones $\mathscr{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^k\}$. La familia de distribuciones $\mathscr{P}$ se dice que pertenece a la familia exponencial de k parámetros si su función de densidad (fp) puede ser representado en la forma

$$f(\mathbf{x}|\theta) = e^{\sum_{i=1}^{k} \eta_i(\theta)T_i(x) - B(\theta)} h(x).$$

De nuevo, obviamente, la elección de las funciones pertinentes a η_i, T_i, h no son únicos. Como en el caso de la familia exponencial de un parámetro, el vector de estadísticas $(T_1, T_2, \ldots, T_k)$ se denomina *estadístico suficiente natural*, y si los reparametrizamos utilizando $\eta_i = \eta_(\theta)$, $i = 1, 2, \ldots, k$, la familia es llamada *familia exponencial canónico k-paramétrico*.

Hay una suposición implícita en esta definición que el número de $\theta's$ que varían libremente es igual al número de $\eta's$ que varían libremente, y que ambos son iguales a número k en el contexto específico. La manera formal de decir esto es asumir lo siguiente:

Suposiciones La dimensión de Θ así como la dimensión de la imagen de Θ según la aplicación

$$(\theta_1, \theta_2, \ldots, \theta_k) \longrightarrow (\eta_1(\theta_1, \theta_2, \ldots, \theta_k), \eta_2(\theta_1, \theta_2, \ldots, \theta_k), \ldots, \eta_k(\theta_1, \theta_2, \ldots, \theta_k))$$

es igual a k.

Hay algunos ejemplos importantes donde esta hipótesis no se cumple. No se tomarán en cuenta como miembros de la familia exponencial de k -parámetros. El nombre familia exponencial curva es comúnmente usado para ellos, y esto será discutido en la última sección.

Los términos *forma canónica*, *parámetro natural*, y *espacio paramétrico natural* significarán las mismas cosas como en el caso de la familia exponencial de un parámetro. De esta manera, si reparametrizamos las distribuciones usando $\eta_1, \eta_2, \ldots, \eta_k$ como k parámetros, entonces el vector $\eta = (\eta_1, \eta_2, \ldots, \eta_k)$ se llama vector paramétrico natural, y la parametrización $f(x|\eta) = e^{\sum_{i=1}^{k} \eta_i T_i(x) - B(\eta)} h(x)$ es llamada forma canónica, y el conjunto de todos los vectores η para la cual $f(x|\eta)$ es una densidad válida (fp) se el espacio paramétrico natural. Los teoremas principales para el caso $k = 1$ se cumplen para una k general.

Teorema 6. Los resultados del teoremas 1 y 5 se cumplen para la familia de k -parámetros.

Las pruebas son casi literalmente las mismas. Las formulas de los momentos difieren ligeramente debido a la presencia de un parámetro en el contexto actual.

Teorema 7. Suponga que el vector aleatorio $\mathbf{X} = X_1, X_2, \ldots, X_m$ tiene una distribución $P_\eta, \eta \in \mathscr{T}$, pertenece a la familia exponencial canónico de k -parámetros, con una densidad (fp)

$$f(x|\eta) = e^{\sum_{i=1}^{k} \eta_i T_i(x) - B(\eta)} h(x),$$

donde

$$\mathscr{T} = \{\eta \in \mathbb{R}^k : \int_{\mathbb{R}^k} e^{\sum_{i=1}^k \eta_i T_i(x)} h(x) dx < \infty\}$$

(siendo la integral reemplazada por la sumatoria en el caso discreto).

(a) En cualquier $\eta \in \mathscr{T}^0$,

$$e^{B(\eta)} = \int_{\mathbb{R}^k} e^{\sum_{i=1}^k \eta_i T_i(x)} h(x) dx$$

es parcialmente diferenciable infinitamente con respecto a cada η_i, y las derivadas parciales de cualquier orden pueden ser obtenidas diferenciando dentro del signo de la integral.

(b)

$$E_\eta(T_i(\mathbf{X})) = \frac{\partial}{\partial \eta_i} B(\eta).$$

y

$$Cov_\eta\left(T_i(\mathbf{X}), T_j(\mathbf{X})\right) = \frac{\partial^2}{\partial \eta_i \partial \eta_j} B(\eta),\ 1 \leq i, j \leq k.$$

(c) Si η, t son tales que $\eta, \eta + t \in \mathscr{T}$, entonces la función característica conjunta de $(T_1(\mathbf{X}), T_2(\mathbf{X}), \ldots, T_k(\mathbf{X}))$ existe y es igual a

$$\Phi(t) = e^{B(\eta + it) - B(\eta)}.$$

Otra nueva terminología importante es de un *rango completo.*

Definición 5. Una familia de distribuciones $\{P_\eta : \eta \in \tau\}$ pertenece a la familia exponencial canónico de k -parámetros se llama familia exponencial de *rango completo* si en todo $\eta \in \tau^0$, la matriz covarianza de orden $k \times k$, definida por

$$\Sigma_\eta = \left[\left(\frac{\partial^2}{\partial \eta_i \partial \eta_j} B(\eta)\right)\right]$$

es no singular.

Definición 6. [**Matriz de Información de Fisher**] Suponga una familia de distribuciones en la familia de exponencial canónico de k -parámetros es no singular. Entonces para $\eta \in \mathcal{T}^0$ la matriz Σ_η se llama matriz de información de Fisher (en η).

La matriz de información de Fisher es de suma importancia en teoría estadística paramétrica y reside en el corazón de la teoría de optimización tanto en muestras finita como en muestras grandes dentro de problemas de inferencia estadística para familias paramétricas regulares.

Ahora vamos a ver algunos ejemplos de distribuciones en familias exponenciales de k parámetros donde $k > 1$.

Ejemplo 13. [**Distribución normal de Dos Parámetros**] Sea la familia normal de dos parámetros,

$$\mathscr{P} = \{\mathscr{N}(\mu, \sigma^2) : \Theta = \{\theta = (\mu, \sigma^2) : (-\infty, \infty) \times (0, \infty)\}\}.$$

Entonces, una representación de la función de densidad para esta familia es

$$f(x|\theta) = f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} I(x \in \mathbb{R}).$$

Si denotamos $(\mu, \sigma) = (\theta_1, \theta_2) = \theta$, entonces parametrizado por θ, la función de densidad de X es

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{\frac{\theta_1}{\theta_2^2}x - \frac{1}{2\theta_2^2}x^2 - \frac{\theta_1^2}{2\theta_2^2} - \ln\theta_2} I(x \in \mathbb{R}).$$

la cual corresponde a la familia exponencial de dos parámetros con,

$$\eta_1(\theta) = \frac{\theta_1}{\theta_2^2}, \quad T_1(x) = x$$

$$\eta_2(\theta) = -\frac{1}{2\theta_2^2}, \quad T_2(x) = x^2$$

$$B(\theta) = \frac{\theta_1^2}{2\theta_2^2} + \ln\theta_2, \quad h(x) = \frac{1}{\sqrt{2\pi}} I(x \in \mathbb{R}).$$

El espacio paramétrico en la parametrización θ es $\Theta = (-\infty, \infty) \times (0, \infty)$.

Si deseamos representar la función de densidad en su forma canónica, consideramos la siguiente parametrización,

$$\eta_1 = \frac{\theta_1}{\theta_2^2}, \quad T_1(x) = x$$

$$\eta_2 = -\frac{1}{2\theta_2^2}, \quad T_2(x) = x^2$$

$$B(\eta) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\ln(-2\eta_2)$$

Resulta que, $f(x|\eta) = e^{\eta_1 T_1(x) + \eta_2 T_2(x) - B(\eta)} h(x)$ donde el espacio paramétrico natural para $\eta = (\eta_1, \eta_2)$ es,

$\eta = \mathbb{R} \times \mathbb{R}^{-1} = \{(\eta_1, \eta_2) : -\infty < \eta_1 < \infty, -\infty < \eta_2 < 0\}$.

Para una muestra de tamaño **n** de una población exponencial de forma canónica tenemos,

$$\begin{aligned} f(\mathbf{x}|\eta) &= \prod_{I=1}^{\mathbf{n}} e^{\eta_1 x_i + \eta_2 x_i^2 - B(\eta)} \\ &= e^{\eta_1 \sum_{i=1}^{\mathbf{n}} x_i + \eta_2 \sum_{i=1}^{\mathbf{n}} x_i^2 - \mathbf{n} B(\eta)} \prod_{i=1}^{\mathbf{n}} h(x_i) \end{aligned}$$

Ejemplo 14. **[Distribución Gamma con dos parámetros]** Se ha visto en el ejemplo 5 que si uno fija uno de los parámetros de una distribución Gamma, entonces un miembro de la familia exponencial de un parámetro. En este ejemplo se muestra que la distribución Gamma es un miembro de la familia exponencial de 2 -parámetros. Para mostrar esto, solo observa eso con $\theta = (\alpha, \beta) = \theta_1, \theta_2$, la función de densidad resulta

$$f_\theta(x) = f(x|\theta) = e^{\theta_1 \ln x - \frac{x}{\theta_2} - \theta_1 \ln \theta_2 - \ln \Gamma(\theta_1)} \frac{1}{x} I(x > 0)$$

Esta densidad está en la familia exponencial de 2 -parámetros con $\eta_1 = \theta_1$, $T_1(x) = \ln x$, $\eta_2 = -\frac{1}{\theta_2}$, $T_2(x) = x$, $B(\theta) = \theta_1 \ln \theta_2 + \ln \Gamma(\theta_1)$, y $h(x) = \frac{1}{x} I(x > 0)$.

El espacio paramétrico en la parametrización -θ es $\Theta = (0, \infty) \times (0, \infty) = \{(\theta_1, \theta_2) : 0 < \theta_1 < \infty, 0 < \theta_2 < \infty\}$. Para la forma canónica, use $\eta_1 = \theta 1$, $\theta_2 = -\frac{1}{\theta_2}$, y luego, el espacio de parámetro natural es $\eta = (0, \infty) \times (-\infty, 0) = \{(\eta_1, \eta_2) : 0 < \eta < \infty, -\infty < \eta_2, < 0\}$. El estadístico suficiente natural es $(T_1(X), T_2(X) = (X, \ln X))$

Ejemplo 15. **[Distribución Normal Multivariada General]** Asuma que $\mathbf{X} \sim \mathcal{N}_n(\mu, \Sigma)$ donde μ es arbitrario y Σ es una matriz definida -positiva (y por su puesto, simétrica). Escribiendo $\theta = (\mu, \Sigma)$, se puede pensar de θ como un subconjunto en un espacio Euclidiano de dimensión

$$k = n + n\frac{n^2 - n}{2} = n + \frac{n(n+1)}{2} = \frac{n(n+3)}{2}$$

La densidad de **X** es

$$
\begin{aligned}
f(\mathbf{x}|\theta) &= Ce^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}I(\mathbf{x}\in\mathbb{R}^n).\\
&= Ce^{-\frac{1}{2}\mathbf{x}'\Sigma^{-1}\mathbf{x}+\mu'\Sigma^{-1}\mathbf{x}-\frac{1}{2}\mu'\Sigma^{-1}\mu}I(\mathbf{x}\in\mathbb{R}^n)\\
&= Ce^{-\frac{1}{2}\sum_i\sum_j\sigma^{ij}x_ix_j+\sum_{i=1}\left(\sum_k\sigma^{ki}\mu_k\right)x_i-\frac{1}{2}\mu'\Sigma^{-1}\mu}I(\mathbf{x}\in\mathbb{R}^n).\\
&= Ce^{-\frac{1}{2}\sum_i\sigma^{ij}x_i^2-\sum\sum_{i<j}\sigma^{ij}x_ix_j+\sum_i\left(\sum_k\sigma^{ki}\mu_k\right)x_i-\frac{1}{2}\mu'\Sigma^{-1}\mu}I(\mathbf{x}\in\mathbb{R}^n),
\end{aligned}
$$

donde la contante C es igual a $C=[(2\pi)^n|\Sigma|]^{-\frac{1}{2}}$.

De esta manera, hemos representado la densidad de **X** en la forma de familia exponencial de k -parámetros con la estadística suficiente natural k -dimensional

$$
\begin{aligned}
&\mathbf{T}(\mathbf{X}) = (T_1(\mathbf{X}),T_2(\mathbf{X}),T_3(\mathbf{X})) \text{ donde}\\
&T_1(\mathbf{X}) = X_1,X_2,\ldots,X_n\\
&T_2(\mathbf{X}) = X_1^2,X_2^2,\ldots,X_n^2\\
&T_3(\mathbf{X}) = X_1X_2,\ldots,X_{n-1}X_n
\end{aligned}
$$

y los parámetros naturales definidos por

$$\sum_k\sigma^{k1}\mu_k,\ldots,\sum_k\sigma^{kn}\mu_k,-\frac{1}{2}\sigma^{11},\ldots-\frac{1}{2}\sigma^{nn},-\sigma^{12},\ldots,-\sigma^{n-1,n}.$$

Ejemplo 16. [**Distribución multinomial**] Considere $k+1$ celdas de distribuciones multinomiales $\theta_1,\theta_2,\ldots,\theta_k,\theta_{k+1}=1-\sum_{i=1}^k\theta_i$. Escribiendo $\theta=(\theta_1,\theta_2,\ldots,\theta_k)$, la función de probabilidad conjunta (o función mas de probabilidad conjunta) de $\mathbf{X}=(X_1,X_2,\ldots,X_k)$, las frecuencias de celdas de las primeras k celdas es

$$
\begin{aligned}
f(\mathbf{x}|\theta) &= \prod_{i=1}^k\theta_i^{x_i}\left(1-\sum_{i=1}^k\theta_i\right)^{n-\sum_{i=1}^k x_i}\frac{n!}{\left(\prod_{i=1}^k x_i!\right)\left(n-\sum_{i=1}^k x_i\right)!}\\
&\qquad\times I_{x_1,x_2,\ldots,x_k\geq 0,\sum_{i=1}^k x_i\leq n}\\
&= e^{\sum_{i=1}^k x_i(\ln\theta_i)-\left(\sum_{i=1}^k x_i\right)\ln\left(1-\sum_{i=1}^k\theta_i\right)+n\ln\left(1-\sum_{i=1}^k\theta_i\right)}\\
&\qquad\times\frac{n!}{\left(\prod_{i=1}^k x_i!\right)\left(n-\sum_{i=1}^k x_i\right)!}\times I_{x_1,x_2,\ldots,x_k\geq 0,\sum_{i=1}^k x_i\leq n}\\
&= e^{\sum_{i=1}^k x_i\left(\ln\frac{\theta_i}{1-\sum_{i=1}^k\theta_i}\right)+n\ln\left(1-\sum_{i=1}^k\theta_i\right)}\frac{n!}{\left(\prod_{i=1}^k x_i!\right)\left(n-\sum_{i=1}^k x_i\right)!}\\
&\qquad I_{x_1,x_2,\ldots,x_k\geq 0,\sum_{i=1}^k x_i\leq n}
\end{aligned}
$$

Esto está en la forma de familia exponencial de k -parámetros con la estadística natural suficiente y parámetros naturales

$$\mathbf{T}(\mathbf{X}) = (X_1, X_2, \ldots, X_k), \quad \eta_i = \ln \frac{\theta_i}{\left(1-\sum_{i=1}^{k}\theta_i\right)}, 1 \leq i \leq k$$

y

$$h(\mathbf{x}) = \frac{n!}{\left(\prod_{i=1}^{k} x_i!\right)\left(n-\sum_{i=1}^{k} x_i\right)!} I_{x_1, x_2, \ldots, x_k \geq 0, \sum_{i=1}^{k} x_i \leq n}$$

Ejemplo 17. [**Distribución gausiana inversa de dos parámetros**] En el Teorema 11.5 se demostró que para el simple paseo aleatorio simétrico en $\mathbb{R}$, el tiempo de la r . ésimo retorno al origen v_r satisface el resultado débil de convergencia

$$\mathbf{P}\left(\frac{v_r}{r^2} \leq x\right) \longrightarrow 2\left[1-\Phi\left(\frac{1}{\sqrt{x}}\right)\right], x > 0$$

cuando $r \longrightarrow \infty$. La función de densidad de este límite función de distribución es $f(x) = \frac{e^{-\frac{1}{2x}} x^{-\frac{3}{2}}}{\sqrt{2\pi}} I(x>0)$. Esta es una distribución gausiana inversa especial. La distribución gausiana inversa general tiene la densidad

$$f(x|\theta_1, \theta_2) = \left(\frac{\theta_2}{\pi x^3}\right)^{1/2} e^{-\theta_1 x - \frac{\theta_2}{x} + 2\sqrt{\theta_1\theta_2}} I(x>0).$$

El espacio paramétrico para $\theta = (\theta_1, \theta_2)$ es $[0, \infty) \times [0, \infty)$

Tenga en cuenta que la densidad gausiana inversa especial asignada a lo anterior corresponde a $\theta_1 = 0$, $\theta_2 = \frac{1}{2}$. La densidad Gausiana inversa general K es la densidad de la primera vez que un proceso Wiener p (comenzando en cero) golpea la línea recta con la ecuación $y = \sqrt{2\theta_2} - \sqrt{2\theta_1} t$; $t > 0$.

De la fórmula de $f(x|\theta_1, \theta_2)$ se desprende que es un miembro de la familia exponencial de dos parámetros con la estadística natural suficiente $T(\mathbf{X}) = \left(X, \frac{1}{X}\right)$ y el espacio de parámetro natural $\mathscr{T} = (-\infty, 0] \times (-\infty, 0]$. Tenga en cuenta que el espacio de parámetro natural no es abierto.

7 Estadísticos Suficientes y Completos en la Familia Exponencial

Las familias exponenciales bajo condiciones suaves en el espacio de parámetros tienen la propiedad de que si una función $g(\mathbf{T})$ del estadístico suficiente natural $\mathbf{T} = T(X)$ tiene valor esperado cero según cada $\theta \in \Theta$, entonces $g(\mathbf{T})$ sí mismo debe ser esencialmente idénticamente igual a cero. Una familia de distribuciones que tiene esta propiedad se llama familia completa. La propiedad de completud, particularmente en conjunción con la propiedad de suficiencia, ha tenido un papel históricamente importante en la inferencia estadística. Lehmann (1959), Lehmann y Casella (1998) y Brown (1986) dan muchas aplicaciones. Sin embargo, nuestra motivación para estudiar la completitud de una familia exponencial de rango completo es principalmente para presentar un teorema bien conocido en estadística, que en realidad también es una herramienta muy efectiva y eficiente para probabilistas. Este teorema, conocido como *el teorema de Basu* (Basu (1955)), es una herramienta eficiente para los probabilistas en la minimización de cálculos de distribución torpes. Se requiere la completitud para establecer el teorema de Basu.

Definición 7. Una familia de distribuciones $\{\mathbf{P}_\theta : \theta \in \Theta\}$ en un espacio muestral $\mathbb{X}$ se llama *completo* si $E_{\mathbf{P}_\theta[g(X)]} = 0$ para todo $\theta \in \Theta$, implica que $\mathbf{P}_\theta\,(g(X) = 0) = 1$ para todo $\theta \in \Theta$.

Es útil ver primero un ejemplo de una familia que no es completa.

Ejemplo 18. Asuma que $X \sim \text{Bin}(2, \theta)$, y el parámetro θ es igual a $\frac{1}{4}$ o $\frac{3}{4}$. En la notación de la definición de completitud, Θ es el conjunto de dos punto $\{\frac{1}{4}, \frac{3}{4}\}$. Considere la función g definida por

$$g(0) = g(2) = 3, \quad g(1) = -5.$$

Entonces,

$$\begin{aligned}
\mathbf{E}_{\mathbf{P}_\theta}[g(\theta)] &= \sum_{x=0}^{2} g(x) \binom{n}{x} \theta^x (1-\theta)^{2-x} \\
&= 16\theta^2 - 16\theta + 3 = 0, \text{ si } \theta = \frac{1}{4} \text{ o } \frac{3}{4}
\end{aligned}$$

Por lo tanto, hemos exhibido una función g que infringe la condición de completitud de esta familia de distribuciones.

Por lo tanto, la completitud de una familia de distribuciones no es universalmente cierta. El problema con el parámetro de dos puntos establecido en el ejemplo anterior es que es demasiado pequeño. Si el espacio de parámetros es más rico, la familia de distribuciones binomiales para cualquier n fija es, de hecho, completa. De hecho, cualquier distribución en la familia exponencial general de k -parámetros como un todo es una familia completa, siempre que el conjunto de valores de los parámetros no sea demasiado delgado. Aquí hay un teorema general.

Teorema 8. Supongamos una familia de distribuciones $\mathscr{P} = \{\mathbf{P}_\theta : \theta \in \Theta\}$ pertenece a una familia exponencial de k -parámetros, y que el conjunto Θ del cual se conoce que el parámetro pertenece al interior del conjunto no vacío. Entonces la familia $\mathscr{B}$ es completa.

La prueba de esto requiere el uso de propiedades de funciones que son analíticas en un dominio en C^k, donde C es el plano complejo. No probamos el teorema aquí; ver Brown $(1986, p.43)$ para una prueba. La suposición interior no vacía nos protege del conjunto, ya que es demasiado pequeño.

Ejemplo 19. Supongamos que $X \sim \text{Bin}(n, \theta)$, donde n es fijo, y el conjunto de valores posibles para θ contiene un intervalo (por pequeño que sea). Entonces, en la terminología del teorema anterior, Θ tiene un interior no vacío. Por tanto, tal familia de distribuciones binomiales es completa. La única función $g(X)$ que satisface $\mathbf{E}_{\mathbf{P}_\theta}[g(\mathbf{X})] = 0$ para todo θ en un conjunto Θ que contiene un intervalo, es la función cero $g(x) = 0$ para todo $x = 0, 1; , \ldots, n$. Contraste esto con el Ejemplo 18.

Requerimos una definición antes de poder establecer el teorema de Basu.

Definición 8. Supongamos que X tiene una distribución $\mathbf{P}_\theta$ que pertenece a una familia $\mathscr{P} = \{\mathbf{P}_\theta : \theta \in \Theta\}$. Una estadística $S(X)$ se llama $\mathscr{P}$ -auxiliar (o, simplemente, auxiliar (ancillary)), si para cualquier conjunto A, $\mathbf{P}_\theta(S(X) \in A)$ no depende de $\theta \in \Theta$ es decir, si $S(X)$ tiene la misma distribución debajo de cada $\mathbf{P}_\theta \in \mathscr{P}$

Ejemplo 20. Supongamos que x_1, x_2 son iid $\mathscr{N}(\mu, 1)$ y μ pertenece a algún subconjunto Θ de la línea real. Considere $S(X_1, X_2) = X_1 - X_2$. Entonces, según cualquier $\mathbf{P}_\theta \mu$, $S(X_1, X_2) \sim \mathscr{N}(0, 1)$, una distribución fija que no depende de μ. Por lo tanto, $S(X_1, X_2) = X_1 - X_2$ es auxiliar, sea cual que sea el conjunto de valores de μ.

Ejemplo 21. Supongamos que X_1, X_2 son iid $U[0, \theta]$, y θ pertenece a algún subconjunto Θ de $(0, \infty)$, Sea $S(X_1, X_2) = \frac{X_1}{X_2}$. Podemos escribir $S(X_1, X_2)$ como

$$S(X_1, X_2) \stackrel{\mathscr{L}}{=} \frac{\theta U_1}{\theta U_2}$$

donde U_1, U_2 son iid $U[0,1]$. Por lo tanto, según cualquier $\mathbf{P}_\theta, S(X_1,X_2)$ se distribuye como la relación de dos variables $U[0,1]$ independientes. Esta es una distribución fija que no depende de θ. Por lo tanto,$S(X_1,X_2) = \frac{X_1}{X_2}$ es auxiliar, sea cual sea el conjunto de valores de θ.

Ejemplo 22. Supongamos que $X_1, X_2, \ldots, X_n$ son iid $\mathscr{N}(\mu, 1)$ y μ pertenece a algún subconjunto de Θ de la línea real. Sea $S(X_1, X_2, \ldots, X_n = \sum_{i=1}^n (X_i - \overline{X})^2)$. Podemos escribir $S(X_1, X_2, \ldots, X_n)$ como

$$S(X_1, X_2, \ldots, X_n) \stackrel{\mathscr{L}}{=} \sum_{i=1}^n [(\mu + Z_i) - (\mu + \overline{Z})]^2 = \sum_{i=1}^n (Z_i - \overline{Z})^2,$$

donde $Z_1, Z_2, \ldots, Z_n$ son iid $\mathscr{N}(0,1)$. Por consiguiente , según $\mathbf{P}_\mu$,$S(X_1, X_2, \ldots, X_n)$ tiene una distribución fija, es decir la distribución de $\sum_{i=1}^n (Z_i - \overline{Z})^2$ (en realidad, es una distribución $\chi^2(n-1)$). Así $S(X_1, X_2, \ldots, X_n = \sum_{i=1}^n (X_i - \overline{X})^2)$ es auxiliar sea cual sea el conjunto de valores de μ.

Teorema 9. **[El teorema de Basu para la familia exponencial]**. En cualquier familia $\mathscr{P}$ exponencial de k -parámetros, con un espacio de paramétrico θ que tiene un interior no vacío, el estadístico suficiente natural de la familia de estadísticos $\mathbf{T}(X)$ y cualquier estadística $S(X)$ $\mathscr{P}$ -auxiliar se distribuyen independientemente según cada $\theta \in \Theta$.

A continuación mostramos aplicaciones de este resultado siguiendo la siguiente sección.

8 Factorización de Neyman-Fisher y Teorema de Basu

Existe una versión más general del teorema de Basu que se aplica a familias de distribuciones paramétricas arbitrarias. La intuición es la misma que en el caso de una familia exponencial, es decir, una estadística suficiente, que contiene toda la información, y una estadística auxiliar, que no contiene información, deben ser independientes. Para esto, necesitamos definir qué significa una estadística suficiente para una familia paramétrica general. Aquí está la definición original de Fisher (Fisher (1922)).

Definición 9. Sea $n \geq 1$ dado, y suponga $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ tiene una distribución conjunta $\mathbf{P}_{\theta,n}$ perteneciente a alguna familia

$$\mathscr{P}_n = \{\mathbf{P}_{\theta,n} : \theta \in \Theta\}.$$

Un estadístico $\mathbf{T}(\mathbf{X}) = \mathbf{T}(X_1, X_2, \ldots, X_n)$ que toman valores en algún espacio Euclidiano se llama *estadístico suficiente* para la familia $\mathscr{P}_n$ si la distribución condicional conjunta de $X_1, X_2, \ldots, X_n$ dado $\mathbf{T}(X_1, X_2, \ldots, X_n)$ es lo mismo según e cada $\theta \in \Theta$.

Por lo tanto, podemos interpretar el estadístico suficiente $\mathbf{T}(X_1, X_2, \ldots, X_n)$ de la siguiente manera: una vez que conocemos el valor de $\mathbf{T}$, el conjunto de valores de los datos individuales $X_1, X_2, \ldots, X_n$ no tiene nada más que transmitir. Podemos pensar en la suficiencia como una reducción de datos sin costo; podemos guardar solo $\mathbf{T}$ y descartar los valores de datos individuales sin perder ninguna información. Sin embargo, lo que es suficiente depende, a menudo de manera crucial, de la forma funcional de las distribuciones $\mathbf{P}_{\theta,n}$. Por lo tanto, la suficiencia es útil para la reducción de datos sujetos a la lealtad de la forma funcional elegida de $\mathbf{P}_{\theta,n}$.

Afortunadamente, existe una receta universal fácilmente aplicable para identificar automáticamente una estadística suficiente para una familia determinada $\mathscr{P}_n$. Este es el teorema de la factorización.

Teorema 10. [**Teorema de la factorización Neyman –Fisher**]. Sea $f(x_1, x_2, \ldots, x_n|\theta)$ la función de densidad conjunta (pmf conjunta) correspondiente a la distribución $\mathbf{P}_{\theta,n}$. Luego, un estadístico $\mathbf{T} = \mathbf{T}(X_1, X_2, \ldots, X_n)$ es suficiente para la familia $\mathscr{P}_n$ si y solo si para cualquier $\theta \in \Theta$, $f(X_1, X_2, \ldots, X_n|\theta)$ puede ser factorizado en la forma

$$f(X_1, X_2, \ldots, X_n|\theta) = g(\theta, \mathbf{T}(X_1, X_2, \ldots, X_n))h(X_1, X_2, \ldots, X_n)$$

La intuición del teorema de factorización es que la única forma en que el parámetro está vinculado a los valores de datos $X_1, X_2, \ldots, X_n$ en la función de verosimilitud

$$f(x_1, x_2, \ldots, x_n|\theta)$$

es a través del estadístico $\mathbf{T}(X_1, X_2, \ldots, X_n)$ porque no hay θ en la función $h(x_1, x_2, \ldots, x_n)$. Por lo tanto, solo deberíamos saber qué es $\mathbf{T}$, pero no los valores individuales $X_1, X_2, \ldots, X_n$.

Aquí hay un ejemplo sobre el uso del teorema de factorización.

Ejemplo 23. [**Estadístico Suficiente para una Distribución Uniforme**] Sean $X_1, X_2, \ldots, X_n$ variables aleatorias iid y distribuidas como $U[0, \theta]$ para algún $\theta > 0$. Entonces la función de densidad conjunta de la muestra es

$$f(x_1,x_2,\ldots,x_n|\theta) = \prod_{i=1}^{n}\frac{1}{\theta}I(x_i \leq \theta) = \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} I(x_i \leq \theta)$$
$$= \left(\frac{1}{\theta}\right)^n I(x_{(n)} \leq \theta)$$

donde $X_{(n)} = \max(x_1,x_2,\ldots,x_n)$. Si consideramos

$$T(X_1,X_2,\ldots,X_n) = X_{(n)}, \quad g(\theta,t) = \left(\frac{1}{\theta}\right) I(t \leq \theta), h(X_1,X_2,\ldots,X_n) = 1.$$

Luego por el teorema de factorización, el máximo muestral $X_{(n)}$ es un estadístico suficiente para la familia $U[0,\theta]$. El resultado tiene un sentido intuitivo.

Aquí está ahora la versión general del teorema de Basu.

Teorema 11. [**Teorema general de Basu**] Sea $\mathscr{P}_n = \{P_\theta : \theta \in \Theta\}$ una familia de distribuciones. Supongamos que $T(X_1,X_2,\ldots,X_n)$ es suficiente para $\mathscr{P}_n$, y $S(X_1,X_2,\ldots,X_n)$ es un estadístico auxiliar (ancillary)según $\mathscr{P}_n$. Entonces T y S se distribuyen independientemente según cada $\mathbf{P}_{\theta,n} \in \mathscr{P}_n$.

9 Aplicaciones del teorema de Basu a la probabilidad

Sabemos que el estadístico suficiente por sí mismo captura toda la información acerca de θ que el conocimiento completo de $\mathbf{X} = (X_1,X_2,\ldots,X_n)$ (muestra) podría haber proporcionado. Por otro lado, una estadística auxiliar no puede proporcionar ninguna información acerca de θ, ya que su distribución no incluye θ. El teorema de Basu dice que una estadística que proporciona toda la información, y otra que no proporciona información, deben ser independiente, siempre que se mantenga la condición adicional del interior no vacía, para asegurar la integridad de la familia $\mathscr{P}_n$. Por lo tanto, los conceptos de información , suficiencia, ancillaridad, completitud e independencia se unen en el teorema de Basu. Sin embargo, nuestro principal interés es simplemente usar el teorema de Basu como una herramienta conveniente para llegar rápidamente a algunos resultados que son puramente resultados en el dominio de la probabilidad. Damos a continuación algunos ejemplos.

Ejemplo 24. [**Independencia de la media y la varianza para una muestra normal**] Supongamos que $X_1,X_2,\ldots,X_n$ son iid $\mathscr{N}(\eta,\tau^2)$ para algunos η,τ. Es sabido que la media de la muestra $\overline{X}$ y la varianza muestral S^2 se distribuyen independientemente para cualquier tamaño de muestra, n, y sean cuales sean η y τ. Ahora lo probamos. Para esto, primero establecemos la afirmación de que si el resultado es válido para $\eta = 0, \tau = 1$, luego vale para todo η, τ. De hecho, fije cualquier η, τ y escriba $X_i = \eta + \tau Z_i$, $1 \leq i \leq n$, donde $Z_1, Z_2, \ldots, Z_n$ son iid $\mathcal{N}(0,1)$. Ahora,

$$\left(\overline{X}, \sum_{i=1}^{n}(X_i - \overline{X})^2\right) = \left(\eta + \tau Z, \sum_{i=1}^{n}(Z_i - \overline{Z})^2\right)$$

Por consiguiente, $\overline{X}$ y $\sum_{i=1}^{n}(X_i - \overline{X})^2$ son independientemente distribuidos según η, τ si y solo si $\overline{Z}$ y $\sum_{i=}^{n}(Z_i - \overline{Z})^2$ son independientemente distribuidos. Este es una etapa para deshacerse de los parámetros η, τ de la consideración.

Pero, ahora, ¡importamos un parámetro! que incruste la distribución $\mathcal{N}(0,1)$ en una familia más grande de distribuciones $\{\mathcal{N}(\mu,1), \mu \in \mathbb{R}\}$. Considere ahora una muestra ficticia $Y_1, Y_2, \ldots, Y_n$ de $\mathbf{P}_\mu = \mathcal{N}(\mu,1)$. La densidad conjunta de $Y = Y_1, Y_2, \ldots, Y_n$ es una densidad de la familia exponencial de un parámetro con la estadística suficiente natural $T(Y) = \sum_{i=1}^{n} Y_i$. Por el ejemplo 22, $\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ es auxiliar. El espacio de parámetro para μ obviamente tiene un interior no vacío, por lo tanto, todas las condiciones del teorema de Basu están satisfechas, y por lo tanto, según cada μ, $\sum_{i=1}^{n} Y_i$ y $\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ se distribuyen independientemente. En particular, se distribuyen independientemente según $\mu = 0$, es decir, cuando las muestras son iid $\mathcal{N}(0,1)$, que es lo que necesitamos probar.

Ejemplo 25. **[Un resultado de distribución exponencial]** Supongamos que $X_1, X_2, \ldots, X_n$ son variables aleatorias exponenciales iid con media λ. Luego, transformando $X_1, X_2, \ldots, X_n$ a

$$\left(\frac{X_1}{\sum_{i=1}^{n} X_i}, \cdots, \frac{X_{n-1}}{\sum_{i=1}^{n} X_i}, \sum_{i=1}^{n} X_i\right),$$

uno puede mostrar al realizar la transformación mediante el método jacobiano, que

$$\left(\frac{X_1}{\sum_{i=1}^{n} X_i}, \cdots, \frac{X_{n-1}}{\sum_{i=1}^{n} X_i}\right),$$

es independiente de $\sum_{i=1}^{n} X_i$. Podemos mostrar esto sin hacer ningún cálculo utilizando el teorema de Basu.

Para esto, una vez más, escribiendo $X_i = \lambda Z_i$, $i \leq i \leq n$ donde las Z_i son variables aleatorias iid exponenciales estándares, observe primero que

$$\left(\frac{X_1}{\sum_{i=1}^{n} X_i}, \cdots, \frac{X_{n-1}}{\sum_{i=1}^{n} X_i} \right),$$

es una estadística auxiliar (vector). A continuación, observe que la densidad conjunta de $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ es una familia exponencial de un parámetro, con el estadístico suficiente natural $T(\mathbf{X}) = \sum_{i=1}^{m} X_i$. Debido a que el espacio de parámetro $(0, \infty)$ obviamente contiene un interior no vacío, según el teorema de Basu, según de cada λ,

$$\left(\frac{X_1}{\sum_{i=1}^{n} X_i}, \cdots, \frac{X_{n-1}}{\sum_{i=1}^{n} X_i} \right) \quad \text{y} \quad \sum_{i=1}^{n} X_i$$

son independientemente distribuidos.

Ejemplo 26. **[Un cálculo de covarianza]** Supongamos que $X_1, X_2, \ldots, X_n$ son iid $\mathcal{N}(0,1)$, y que $\overline{X}$ y M_n denotan la media y la mediana del conjunto de muestras $X_1, X_2, \ldots, X_n$. Al usar nuestro viejo truco de importar un parámetro medio μ, primero observamos que la estadística de diferencia $\overline{X} - M_n$ es auxiliar. Por otro lado, la densidad conjunta de $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ es, por supuesto, una familia exponencial de un parámetro con el estadística suficiente natural $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$. Según el teorema de Basu, $\sum_{i=1}^{n} X_i$ y $\overline{X} - M_n$ son independientes en cada μ, lo que implica

$$\begin{aligned}
\mathbf{C}ov(X_1 + \cdots X_n, \overline{X} - M_n) = 0 &\Longrightarrow \mathbf{C}ov(n\overline{X}, \overline{X} - M_n) = 0 \\
&\Longrightarrow n\mathbf{C}ov(\overline{X}, \overline{X} - M_n) = 0 \\
&\Longrightarrow \mathbf{C}ov(\overline{X}, \overline{X} - M_n) = 0 \\
&\Longrightarrow \mathbf{C}ov(\overline{X}, \overline{X}) - \mathbf{C}ov(\overline{X}, M_n) = 0
\end{aligned}$$

luego,

$$\mathbf{C}ov(\overline{X}, M_n) = \mathbf{C}ov(\overline{X}, \overline{X}) = \mathbf{V}ar(\overline{X}) = \frac{1}{n}$$

Hemos logrado este resultado sin hacer ningún cálculo en absoluto. Un desarrollo directo a este problema requiere el manejo de la distribución conjunta de $(\overline{X}, M_n)$.

Ejemplo 27. [**Cálculo de una expectativa**.] Supongamos que $X_1, X_2, \ldots, X_n$ son iid $U[0,1]$, y sean $X_{(1)}, X_{(n)}$ el estadístico de orden mínimo y más el estadístico máximo de $X_1, X_2, \ldots, X_n$. Importe un parámetro $\theta > 0$ y considere la familia de distribuciones $U[0,\theta]$. Se ha demostrado que el estadístico de máximo $X_{(n)}$ es suficiente; también es completo. Por otro lado, el cociente $\frac{X_{(1)}}{X_{(n)}}$ es auxiliar. Para ver esto, de nuevo, escriba $(X_1, X_2, \ldots, X_n) \stackrel{L}{=} (\theta U_1, \theta U_2, \ldots, \theta U_n)$, donde $U_1, U_2, \ldots, U_n$ son iid $U[(0,1)]$. Como una consecuencia, $\frac{X_{(1)}}{X_{(n)}} \stackrel{L}{=} \frac{U_{(1)}}{U_{(n)}}$. Por consiguiente $\frac{X_{(1)}}{X_{(n)}}$ es auxiliar. Por la versión general del teorema de Basu, que funciona para cualquier familia de distribuciones (no solo una familia exponencial), se deduce que $X_{(n)}$ y $\frac{X_{(1)}}{X_{(n)}}$ se distribuyen de forma independiente según cada θ. Por lo tanto,

$$E[X_{(1)}] = E\left[\frac{X_{(1)}}{X_{(n)}} X_{(n)}\right] = E\left[\frac{X_{(1)}}{X_{(n)}}\right] E[X_{(n)}]$$
$$\Longrightarrow E\left[\frac{X_{(1)}}{X_{(n)}}\right] = \frac{E[X_{(1)}]}{E[X_{(n)}]} = \frac{\frac{\theta}{n+1}}{\frac{n\theta}{n+1}} = \frac{1}{n}.$$

Una vez más, podemos obtener este resultado utilizando el teorema de Basu sin realizar integraciones o cálculos.

10 Familia exponencial curvada

Hay algunos ejemplos importantes en los que la densidad (pmf) tiene la forma básica de familia expositiva $f(x|\theta) = e^{\sum_{i=1}^{n} \eta_i(\theta) T_i(x) - B(\theta)} h(x)$, pero la suposición de que las dimensiones de Θ, y la del espacio de rango de $\eta(\theta)_1, \eta(\theta)_2, \ldots, \eta(\theta)_k$ son lo mismo se viola, más precisamente, la dimensión de Θ es un entero positivo q estrictamente menos de k. Comencemos con un ejemplo.

Ejemplo 28. Suponga que $X \sim \mathcal{N}(\mu, \mu^2)$, $\mu \neq 0$. Escribiendo $\mu = \theta$, la densidad de X es

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}|\theta|} e^{\frac{1}{2\theta^2}((x-\theta)^2} I(x \in \mathbb{R})$$
$$= \frac{1}{\sqrt{2\pi}} e^{\frac{x}{\theta} - \frac{x^2}{2\theta^2} - \frac{1}{2} - \ln|\theta|} I(x \in \mathbb{R})$$

Escribiendo $\eta_1(\theta) = \frac{1}{\theta}, \eta_2 = -\frac{1}{2\theta^2}, T_1(x) = x, T_2(x) = x^2, B(\theta) = \frac{1}{2} + \ln|\theta|$ y $h(x) = \frac{1}{\sqrt{2\pi}} I(x \in \mathbb{R})$, este está en la forma $f(x|\theta) = e^{\sum_{i=1}^{n} \eta_i(\theta) T_i(x) - B(\theta)} h(X)$, con $k = 2$ a pesar de que $\theta \in \mathbb{R}$ que es solo unidimensional. Las dos funciones $\eta_1(\theta) = \frac{1}{\theta}, \eta_2 = -\frac{1}{2\theta^2}$ están relacionados entre sí por la identidad $\eta_2 = -\frac{1}{2}\eta_1$, de modo que una gráfica de (η_1, η_2) en el plano sería una curva, no una línea recta. Las distribuciones de este tipo se conocen con el nombre de *familia curva exponencial* . La dimensión de la estadística natural suficiente es más que la dimensión de Θ para tales distribuciones.

Sea $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ tenga un distribución $\mathbf{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^q$. Suponga que $\mathbf{P}_\theta$ (fmp) de la forma

$$f(x|\theta) = e^{\sum_{i=1}^{n} \eta_i(\theta) T_i(x) - B(\theta)} h(x)$$

donde $k > q$. Entonces la familia $\{\mathbf{P}_\theta, \theta \in \Theta\}$ se llama una *familia curva exponencial* .

Ejemplo 29. **[Una Normal Bivariada Específica.]** Supongamos $\mathbf{X} = (X_1, X_2)$ tiene una distribución normal bivariada con cero promedios, desviaciones estándar iguales a uno y un parámetro de correlación, $1 < \rho < 1$. La densidad de $\mathbf{X}$ es

$$\begin{aligned} f(x|\rho) &= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[x_1^2 + x_2^2 - 2\rho x_1 x_2\right]} I(x_1, x_2 \in \mathbb{R}) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x_1^2 + x_2^2}{2(1-\rho^2)} + 2\frac{\rho}{1-\rho^2} x_1 x_2} I(x_1, x_2 \in \mathbb{R}). \end{aligned}$$

Por lo tanto, aquí tenemos una familia curva exponencial con $q = 1, k = 2, \eta_1(\rho) = -\frac{1}{2(1-\rho^2)}, \eta_2(\rho) = \frac{\rho}{1-\rho^2}, T_1(x) = x_1 + x^2, T_2(x) = x_1 x_2, B(\rho) = \frac{1}{2}\ln(1-\rho^2)$, y $h(x) = \frac{1}{2\pi} I(x_1, x_2 \in \mathbb{R})$.

Ejemplo 30. **[Poissones con Covariables Aleatorias.]** Supongamos que dado $Z_i = z_i, i = 1, 2, \ldots, n, X_i$ son variables independientes $\mathrm{Poi}(\lambda z_i)$, y $Z_1, Z_2, \ldots, Z_i$ tienen algunas fmp $\mathbf{p}(z_1, z_2, \ldots, z_n)$. Se supone implícitamente que cada $Z_i > 0$ con probabilidad uno. Luego, la fmp conjunta de $(X_1, X_2, \ldots, X_n, Z_1, Z_2, \ldots, Z_n)$ es

$$\begin{aligned} f(x_1, x_2, \ldots, x_n, z_1, z_2, \ldots, z_n) &= \prod_{i=1}^{n} \frac{e^{-\lambda z_i}(\lambda z_i)^{x_i}}{x_i!} \mathbf{p}(z_1, z_2, \ldots, z_n) I_{(x_1, x_2, \ldots, x_n \in \mathbb{N}_0)} \\ &\qquad I_{(z_1, z_2, \ldots, z_n \in \mathbb{N}_1)} \\ &= e^{-\lambda \sum_{i=1}^{n} z_i + \left(\sum_{i=1}^{n} x_i\right) \ln \lambda} \prod_{i=1}^{n} \frac{z_i^{x_i}}{x_i!} \mathbf{p}(z_1, z_2, \ldots, z_n) I_{(x_1, x_2, \ldots, x_n \in \mathbb{N}_0)} I_{(z_1, z_2, \ldots, z_n \in \mathbb{N}_1)} \end{aligned}$$

donde $\mathbb{N}_0$ es el conjunto de enteros no negativos, y $\mathbb{N}_1$ es el conjunto de enteros positivos.

Esto está en la familia curva exponencial con

$$q = 1, \quad k = 2, \quad \eta_1(\lambda) = -\lambda, \quad \eta_2(\lambda) = \ln \lambda, \quad T_1(x, z) = \sum_{i=1}^{n} z_i$$

$$T_2(x, z) = \sum_{i=1}^{n} X_i, \text{ y}$$

$$h(x, z) = \prod_{i=1}^{n} \frac{z_i^{x_i}}{x_i!} \mathbf{p}(zn) I_{(x_1, x_2, \ldots, x_n \in \mathbb{N}_0)} I_{(z_1, z_2, \ldots, z_n \in \mathbb{N}_1)}$$

Si consideramos las covariables como fijas, la distribución conjunta de $(X_1, X_2, \ldots, X_n)$ se convierte en una familia exponencial normal de un parámetro.

Ejercicio 1. Demuestre que la distribución geométrica pertenece a la familia exponencial de un parámetro si $0 < \theta < 1$ y escríbala en la forma canónica y utilizando la parametrización de la media.

Ejercicio 2. Demuestre que la distribución de Poisson pertenece a la familia exponencial de un parámetro si $\lambda > 0$. Escríbala en la forma canónica y utilizando la parametrización media.

Ejercicio 3. Demuestre que la distribución binomial negativa con los parámetros r y θ pertenece a la familia exponencial de un parámetro si r se considera fijo y $0 < \theta < 1$. Escríbalo en la forma canónica y utilizando la parametrización media.

Ejercicio 4. Muestre que la distribución binomial negativa generalizada con el fmp $f(x|\theta) = \frac{\Gamma(\alpha x)}{\Gamma(\alpha)x!}\theta^{\alpha}(1-\theta)^x$, $x = 0, 1, 2, \ldots$, pertenece a la familia exponencial de un parámetro si $\alpha > 0$ se considera fijo y $0 < \theta < 1$. Muestra que la distribución binomial negativa generalizada de dos parámetros con el $f(x|\alpha, \theta) = \frac{\Gamma(\alpha x)}{\Gamma(\alpha)x!}\theta^{\alpha}(1-\theta)^x$, $x = 0, 1, 2, \ldots$ no pertenece a la familia exponencial de dos parámetros.

Ejercicio 5. Muestre que la distribución $\mathscr{N}(\mu, \mu)$ pertenece a la familia exponencial de un parámetro si $\mu > 0$. Escríbala en la forma canónica y utilizando la parametrización de medias.

Ejercicio 6. Supongamos que $X_1, X_2, \ldots, X_n$ es una muestra de una distribución normal estándar, y suponga $X_{(1)}, X_{(n)}$ son los estadísticos de orden mínimo y de oren máximo de $X_1, X_2, \ldots, X_n$, y s^2

es la varianza de la muestra. Probar, aplicando el teorema de Basu a una familia exponencial adecuada de dos parámetros, que

$$E\left[\frac{X_{(n)} - X_{(1)}}{s}\right] = 2\frac{E\left[X_{(n)}\right]}{E(s)}$$

Ejercicio 7. Supongamos que $X_1, X_2, \ldots, X_n$ es una muestra iid de una distribución normal d -dimensional $\mathcal{N}_d(0, \Sigma)$, dónde Σ es una matriz definida positiva. Supongamos que S es la matriz de covarianza de la muestra y $\overline{X}$ el vector de promedios de la muestra. La estadística $D_n^2 = n\overline{X}S^{-1}\overline{X}$ se llama estadístico -D^2 de Mahalanobis Encuentre $E(D_n^2)$ usando el teorema de Basu.

Ejercicio 8. Supongamos que X_i, $1 \leq i \leq n$ son iid $\mathcal{N}(\mu_X, \sigma_X^2)$, Y_i, $1 \leq i \leq n$ son iid $\mathcal{N}(\mu_Y, \sigma_Y^2)$ dónde $mu_x, \mu_Y \in \mathbb{R}$, y $\sigma_X^2, \sigma_Y^2 > 0$. Sean $\overline{X}, s_x^2$ denotan la media y la varianza de $X_1, X_2, \ldots, X_n$ y $\overline{X}, s_x^2$ denotan la media y la varianza del $Y_1, Y_2, \ldots, Y_n$. También sea que r denote el coeficiente de correlación de la muestra basado en los pares (X_i, Y_i), $1 \leq i \leq n$. Demostrar que $\overline{X}, \overline{Y}, s_x^2, s_y^2, r$ son mutuamente independiente en todos $\mu_X, \mu_y, \sigma_X, \sigma_Y$.

Ejercicio 9. Muestre que la distribución de la mezcla $\frac{1}{2}\mathcal{N}(\mu, 1) + \frac{1}{2}\mathcal{N}(\mu, 2)$ no pertenece a la familia exponencial de un parámetro. Generalice este resultado a mezclas más generales de distribuciones normales.

Ejercicio 10. (a) Demuestre que la distribución exponencial doble con un valor σ conocido y una media desconocida no pertenece a la familia exponencial de un parámetro, pero la distribución exponencial doble con una media conocida y una σ desconocida pertenecen a la familia exponencial de un parámetro.

(b) Demuestre que la distribución exponencial doble de dos parámetros no pertenece a la familia exponencial de dos parámetros.

Ejercicio 11. Suponga que $X \sim Bin(n,\theta), Y \sim Bin(n,\theta^2)$, y que X,Y son independientes.Demuestre que la distribución de (X,Y) es una familia curva exponencial.

Ejercicio 12. Suponga que $X_1, X_2, \ldots, X_n$ son conjuntamente multivariadas con medias μ_i, varianzas todas iguales a 1 y una correlación común de pares ρ. Demuestre que la distribución de $(X_1, X_2, \ldots, X_n)$ es una familia curva exponencial.

Ejercicio 13. Suponga $X_1, X_2, \ldots, X_n$ son iid $\mathcal{N}(\mu, \mu^2)$, $\mu \neq 0$. Sea $T(X_1, X_2, \ldots, X_n) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$. Encuentre una función $g(T)$ tal que $E_\mu[g(T)] = 0$ para todo μ, $\mathbf{P}_\mu(g(T) = 0) < 1$ para cualquier μ.

Referencias

[1] Anirban DasGupta. (2011), *Probability for Sattistics and Machine Learnig*, Springer Texts in Statistics.

[2] George Casella, Roger L. Berger. (2002), *Statistical Inference*, 2nd Edition, P. cm.

[3] Bickel, J. Bickel,Kjell A Doksum (2002), *Mathematical Statistics* 2nd Edition Printice Hall, inc

Disponible en `www.cs.columbia.edu`,

Disponible en`people.eecs.berkeley.edu`