Capítulo 6

Principios de reducción de la data

6.1. Introducción

Un experimentador usa la información en una muestra $X_1, \cdots, X_n$ para realizar el proceso de inferencia sobre algun parámetro desconocido θ. Si el tamaño de muestra es grande, los valores observados en la muestra $x_1, \cdots, x_n$ podrían ser difíciles de interpretar de forma individual por lo que es necesario resumir la información en la muestra a través del cálculo de *estadísticas* como la media, la varianza, el máximo, el mínimo, la mediana, etc.

6.2. El principio de suficiencia

Una *estadística suficiente* para θ es una estadística que, de cierta forma, captura toda la informacion acerca del parámetro contenida en la muestra. No es posible obtener información adicional en la muestra, además del valor de la estadística suficiente. Estas consideraciones nos llevan a la técnica de reducción de datos conocida como el principio de suficiencia: *"si $T(\mathbf{X})$ es una estadística suficiente para θ, entonces el proceso de inferencia sobre θ depende de la muestra $\mathbf{X}$ solo a través del valor $T(\mathbf{X})$"*.

6.2.1. Estadística suficiente

En esta sección se mencionan algunos aspectos importantes del principio de suficiencia. Una estadística suficiente es un tipo particular de estadística que condensa la muestra de tal forma que no exista pérdida de información sobre θ. Se presenta a continuación una definición formal.

Definición 6.2.1 Una estadística $T(\mathbf{X})$ es una *estadística suficiente para* θ si la distribución de la muestra $\mathbf{X}$ dado el valor de T no depende de θ.

Ejemplo 6.2.1 Sea X_1, X_2, X_3 una muestra de tamaño tres obtenida desde la distribución $\mathcal{B}(p)$. Probar si $T(\mathbf{X}) = X_1 + X_2 + X_3$ es una estadística suficiente para p.

La definición anterior no es muy manejable ya que requiere obtener una distribución condicional que podría no ser fácil en el caso de trabajar con variables aleatorias continuas.

Teorema 6.2.1 Si $f(\mathbf{x}|\theta)$ es la función de probabilidad o densidad conjunta de $\mathbf{X}$ y $q(t|\theta)$ es la función de probabilidad o densidad de $T(\mathbf{X})$, entonces T es una estadística suficiente para θ si y solo si:

$$\frac{f(\mathbf{x}|\theta)}{q\,(t|\theta)}$$

no depende de θ para todo $\mathbf{X}$.

Ejemplo 6.2.2 Sea $X_1, \cdots, X_n$ una muestra obtenida desde la distribución $\mathcal{P}(\lambda)$. Probar que $T(\mathbf{X}) = X_1 + X_2 + \cdots + X_n$ es una estadística suficiente para λ.

Ejemplo 6.2.3 Sea $X_1, \cdots, X_n$ una muestra obtenida desde la distribución $\mathcal{N}(\mu, \sigma^2)$ donde σ^2 es conocido. Probar si $T(\mathbf{X}) = \overline{X}$ es una estadística suficiente para μ.

La aplicación de la definición 6.2.1 es tediosa y requiere un conocimiento previo acerca del estadístico que podría ser suficiente. El siguiente teorema es de aplicación directa.

Teorema 6.2.2 (Teorema de Factorización) Sea $f(\mathbf{x}|\theta)$ la función de probabilidad o densidad conjunta de la muestra $\mathbf{X}$. La estadística $T(\mathbf{X})$ es suficiente para θ si y solo si existen funciones $g(t|\theta)$ y $h(\mathbf{x})$ tales que:

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}) \tag{6.2.1}$$

Según el teorema anterior para hallar una estadística suficiente se debe factorizar la función de probabilidad o densidad conjunta de la muestra en dos partes. Una parte no depende de θ y constituye la función $h(\mathbf{x})$. La otra parte, que depende de θ, también depende de la muestra $\mathbf{x}$ solo a través de cierta función $T(\mathbf{x})$ que corresponde a la estadística suficiente para θ.

Ejemplo 6.2.4 Sea $X_1, \cdots, X_n$ una muestra obtenida desde la distribución $\mathcal{G}(\alpha, \beta)$ donde β es conocido. Hallar una estadística suficiente para α.

Ejemplo 6.2.5 Sea $X_1, \cdots, X_n$ una muestra obtenida desde la distribución $\mathcal{UC}(0, \theta)$. Hallar una estadística suficiente para θ.

En muchos casos, una estadística suficiente es un vector:

$$T(\mathbf{X}) = (T_1(\mathbf{X}), \cdots, T_r(\mathbf{X}))$$

que ocurre cuando el parámetro es también un vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_s)$ siendo usual que $r = s$. El teorema de factorización puede ser usado para encontrar una estadistica suficiente en la situación anterior.

Ejemplo 6.2.6 Sea $X_1, \cdots, X_n$ una muestra obtenida desde la distribución $\mathcal{N}(\mu, \sigma^2)$ donde ambos parámetros son desconocidos, es decir $\boldsymbol{\theta} = (\mu, \sigma^2)$. Según el teorema de factorización:

$$\begin{aligned} f(\mathbf{x}|\mu, \sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right)\right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(n(\bar{x} - \mu)^2 + (n-1)s^2\right)\right\} \end{aligned}$$

Si $T_1(\mathbf{x}) = \bar{x}$ y $T_2(\mathbf{x}) = s^2$, entonces:

$$\begin{aligned} f(\mathbf{x}|\mu, \sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(n(t_1 - \mu)^2 + (n-1)t_2\right)\right\} \\ &= g(T_1, T_2|\mu, \sigma^2)h(\mathbf{x}) \end{aligned}$$

luego $T(\mathbf{X}) = (T_1, T_2) = (\overline{X}, S^2)$ es una estadística suficiente para el modelo normal.

Las distribuciones que pertenecen a una familia exponencial cumplen con las condiciones vistas en el teorema de factorización por lo que es posible obtener las estadísticas suficientes de forma bastante directa.

Teorema 6.2.3 Sea $X_1, \cdots, X_n$ una muestra obtenida desde una población cuya función de probabilidad o densidad $f(x|\boldsymbol{\theta})$ pertenece a una familia exponencial dada por:

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right\}$$

donde $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_d)$, $d \leq k$. Entonces:

$$\mathbf{T}(\mathbf{X}) = (T_1, \cdots, T_k) = \left(\sum_{j=1}^{n} t_1(X_j), \cdots, \sum_{j=1}^{n} t_k(X_j)\right)$$

es una estadística suficiente para $\boldsymbol{\theta}$. También se dice que $T_1, \cdots, T_k$ son estadísticas conjuntamente suficientes

Ejemplo 6.2.7 Sea $X_1, \cdots, X_n$ una muestra obtenida desde la distribución $\mathcal{BE}(\alpha, \beta)$. Hallar una estadística conjuntamente suficiente para (α, β).

Teorema 6.2.4 Sean $T_1(\mathbf{X}), \cdots, T_k(\mathbf{X})$ estadísticas conjuntamente suficientes. Entonces cualquier conjunto de transformaciones uno a uno de $T_1, \cdots, T_k$ es también conjuntamente suficiente.

6.2.2. Estadística minimal suficiente

Ya que es posible encontrar muchas estadísticas suficientes en un mismo problema sera necesario establecer cuál es la mejor. Recordar que el propósito de una estadística suficiente es lograr resumir la data sin pérdida de información acerca del parámetro θ, es decir que se debe buscar aquella estadística que logre la mayor reducción de data reteniendo aún toda la información sobre θ.

Definición 6.2.2 Una estadística suficiente $T(\mathbf{X})$ es llamada *estadística minimal suficiente* si, para cualquier otra estadística suficiente $T^{'}(\mathbf{X})$, T es función de $T^{'}$.

Teorema 6.2.5 Sea $f(\mathbf{x}|\theta)$ la función de probabilidad o densidad de una muestra $\mathbf{X}$. Suponga que existe una función $T(\mathbf{X})$ tal que para dos puntos muestrales $\mathbf{x}_1$ y $\mathbf{x}_2$ el ratio:

$$\frac{f(\mathbf{x}_1|\theta)}{f(\mathbf{x}_2|\theta)}$$

no depende de θ si y solo si $T(\mathbf{x}_1) = T(\mathbf{x}_2)$. Entonces T es una estadística minimal suficiente para θ.

Ejemplo 6.2.8 Sea $X_1, \cdots, X_n$ una muestra obtenida desde la distribución $\mathcal{N}(\mu, \sigma^2)$ ambos parámetros desconocidos. Probar que $(\overline{X}, S^2)$ es una estadística minimal suficiente para (μ, σ^2).

Una estadística minimal suficiente no es única. Cualquier función uno a uno de una estadística minimal suficiente es también una estadística minimal suficiente. Luego $T^{'}(\mathbf{X}) = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ es también una estadística minimal suficiente en el ejemplo 6.2.8.

6.2.3. Estadística ancillar

En las secciones anteriores se consideraron las estadísticas suficientes que contienen toda la información sobre θ en la muestra. En esta sección se introduce un tipo diferente de estadística que tiene un propósito complementario.

Definición 6.2.3 Una estadística $S(\mathbf{X})$ cuya distribución no depende del paramétro θ es llamada *estadística ancillar*.

Ejemplo 6.2.9 Sea $X_1, \cdots, X_n$ una muestra obtenida desde la distribución $\mathcal{U}(\theta, \theta + 1)$. Sean $X_{(1)} < \cdots < X_{(n)}$ las estadísticas de orden de la muestra. Probar que $R = X_{(n)} - X_{(1)}$ es una estadística ancillar.

6.2.4. Estadística suficiente y completa

Otro concepto útil en la construcción de una estadística es la *completez*, que carece de una interpretación clara y de fácil entendimiento pero que resulta de mucha utilidad para poder encontrar una estadística óptima.

Definición 6.2.4 Sea $f(t|\theta)$ una familia con función de probabilidad o densidad para una estadística $T(\mathbf{X})$. La familia es llamada *completa* si $\mathrm{E}_\theta[g(T)] =$

0 implica necesariamente que $\Pr_\theta[g(T)=0]=1$. Luego, $T(\mathbf{X})$ es llamada una *estadística completa*.

Ejemplo 6.2.10 Sea $X_1,\cdots,X_n$ una muestra obtenida desde la distribución $\mathcal{B}(p)$. Sea $T(\mathbf{X})=X_1+\cdots+X_n\sim\mathcal{BI}(n,p)$. Sea g una función tal que:

$$\mathrm{E}_p[g(T)]=\sum_{t=0}^{n}g(t)\binom{n}{t}p^t(1-p)^{n-t}=(1-p)^n\sum_{t=0}^{n}g(t)\binom{n}{t}\left(\frac{p}{1-p}\right)^t=0$$

entonces $g(t)=0$ para $t=0,1,2,\cdots,n$ y $\Pr_p(g(T)=0)=1$, para todo p. Luego T es una estadística completa.

Ejemplo 6.2.11 Sea $X_1,\cdots,X_n$ una muestra obtenida desde la distribución $\mathcal{UC}(0,\theta)$. Se tiene que $T(\mathbf{X})=X_{(n)}$ es una estadística suficiente y que su función de densidad es:

$$f(t|\theta)=\begin{cases}nt^{n-1}\theta & 0<t<\theta\\ 0 & \text{de otro modo}\end{cases}$$

Suponga $g(t)$ es una función que satisface $\mathrm{E}_\theta[g(T)]=0$ para todo θ, entonces se tiene:

$$0=\frac{d}{d\theta}\mathrm{E}_\theta[g(T)]=\frac{d}{d\theta}\int_0^\theta g(t)nt^{n-1}\theta^{-n}dt$$

Para toda función Riemman integrable $\frac{d}{d\theta}\int_0^\theta g(t)dt=g(\theta)$. Luego:

$$\begin{aligned}
&= \frac{d}{d\theta}\left[\theta^{-n}\int_0^\theta g(t)nt^{n-1}\right]\\
&= \theta^{-n}\frac{d}{d\theta}\int_0^\theta g(t)nt^{n-1}dt+\frac{d}{d\theta}(\theta^{-n})\int_0^\theta ng(t)t^{n-1}dt\\
&= (\theta^{-n})g(\theta)n\theta^{n-1}+0\\
&= \frac{1}{\theta}g(\theta)n=0
\end{aligned}$$

entonces $g(\theta)=0$, luego T es una estadística completa.

Teorema 6.2.6 Sea $X_1, \cdots, X_n$ una muestra obtenida desde una población cuya función de probabilidad o densidad $f(x|\boldsymbol{\theta})$ pertenece a una familia exponencial dada por:

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left\{\sum_{j=1}^{k} w_j(\boldsymbol{\theta})t_j(x)\right\}$$

donde $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)$, entonces la estadística:

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \cdots, \sum_{i=1}^{n} t_k(X_i)\right)$$

es *suficiente* y *completa*.

Ejemplo 6.2.12 Sea $X_1, \cdots, X_n$ una muestra obtenida desde la distribución $\mathcal{BE}(\alpha, \beta)$. Hallar estadísticas suficientes y completas para los parámetros.