INTRODUCCIÓN A LA INFERENCIA

Dr. Cristian Carvajal Muquillaza

Universidad de Playa Ancha

Universidad de Playa Ancha
Facultad de Ciencias Naturales y Exactas
DEPARTAMENTO DISCIPLINARIO DE MATEMÁTICA Y ESTADÍSTICA

Contenido

Universidad de Playa Ancha Facultad de Ciencias Naturales y Exactas

DEPARTAMENTO DISCIPLINARIO DE MATEMÁTICA Y ESTADÍSTICA

Familia de distribuciones

Definición (Familia de distribuciones)

Una ***familia de distiribuciones*** *es un conjunto de la forma:*

$$\mathscr{F} = \{f(x;\vec{\theta}) : \vec{\theta} \in \Theta\}$$

Donde:

1. $f(x;\vec{\theta})$ *es una función de masa o densidad,*
2. $\vec{\theta}$ *es un* ***vector de parámetros****,*
3. Θ *es el conjunto de posibles valores de* $\vec{\theta}$ *llamado* ***Espacio Paramétrico****.*

Obs: *Si* $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$*, entonces* $\vec{\theta} \subseteq \mathbb{R}^k$

Familia de distribuciones

Ejemplo:

$$\mathscr{F} = \left\{ f(x;\vec{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\dfrac{-(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}; \vec{\theta} = (\mu, \sigma^2); \Theta = \mathbb{R} \times \mathbb{R}^+ \right\}$$

Familia de distribuciones

Ejemplo: Continuación

Es la Familia llamada de Gauss, más conocida como la **Familia Normal**

- El parámetro $\vec{\theta}$ es **bidimensional**,
- μ es llamado **parámetro de centro** y tiene como espacio paramétrico a $\mathbb{R}$,
- σ^2 es llamado **parámetro de forma** y su espacio paramétrico es $\mathbb{R}^+$.

Zoom

1. $\vec{\theta}$ por lo general es desconocido,
2. La Inferencia Estadística permite estimar los valores de $\vec{\theta}$.

Familia de distribuciones

Ejemplo:

La población de estaturas de los habitantes de una pais se asume con distribución Normal, es decir si X es la variable aleatoria definida como "**Estatura de los habitantes de la población**",entoces:

$$X \sim N(\mu; \sigma^2)$$

El problema consiste en obtener información de los parámetros μ y σ^2, la técnica básica consiste en reiterar observaciones de los individuos de la población bajo las mismas condiciones. El concepto que se define a continuación formaliza esta idea.

Muestras Aleatorias

Muestras Aleatorias

Las variables aleatorias $\underline{X} = (X_1, X_2, ..., X_n)$ son llamadas **MUESTRAS ALEATORIAS** de tamaño n desde la población $f(x)$ si $X_1, X_2, ..., X_n$ son variables aleatorias mutuamente independientes y las funciones marginales de densidad o masa de cada X_i tiene la misma función $f(x)$, lo que normalmente se llama **variables aleatorias independiente e idénticamente distribuida**, lo que se abrevia variables **iid**.

Zoom

$\underline{X} = (X_1, X_2, ..., X_n)$ $ma(n)$ de X significa

1. X_i independiente de $X_j, \forall i \neq j$ (Independencia)
2. $X_i \sim X, \forall i$ (Equidistribución).

Muestras Aleatorias

Definición (Muestra Aleatoria)

Sea $\underline{X} = (X_1, X_2, \ldots, X_n)$ *una m.a (n), desde una población* $f(x)$, *entonces la conjunta* $f_{\underline{X}}(\underline{x})$ *es igual al producto de las marginales, es decir:*

$$
\begin{aligned}
f_{\underline{X}}(\underline{x}) &= f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \ldots \cdot f_{X_n}(x_n) \\
&= \prod_{i=1}^{n} f(x_i)
\end{aligned}
$$

Muestras Aleatorias

Ejemplo:

Sean $\underline{X} = (X_1, X_2, ..., X_n)$ una m.a (n), desde una población con distribución Poisson de parámetro λ, es decir $X \sim P(\lambda)$, con función de masa

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Resp:

$$\begin{aligned} f_{\underline{X}}(\underline{x};\lambda) &= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \\ &= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \end{aligned}$$

Muestras Aleatorias

Zoom

La función $f_{\underline{X}}(\underline{x};\lambda)$ es llamada **Función de $\underline{X}$ Verosimilitud**, y algunas veces se denota

$$L(\underline{X};\vec{\theta}) \quad \vee \quad L(\vec{\theta};\underline{X})$$

Estadísticos

Definición (Estadístico)

Sean $\underline{X} = (X_1, X_2, ..., X_n)$ *una muestra aleatoria de tamaño n, desde una población y sea* $T\left(\underline{X}\right)$ *una función a valor real o a valor vectorial cuyo dominio está incluido en el espacio muestral de* $\underline{X}$ *, entonces la variable aleatoria o vector aleatorio* $Y = T(\underline{X}) = T\,(X_1, X_2, ..., X_n)$ *es llamado* ***Estadístico****, si* $T(\underline{X})$ *sólo depende de los componentes de* $\underline{X}$.

Estadísticos

Ejemplo:

Supona que se está muestreando una variable aleatoria $X \sim N\left(\mu; \sigma^2\right)$. Determinar si las siguientes funciones son Estadísticos:

1. $T(\underline{X}) = \frac{1}{n} \sum_{i=1}^{n} x_i$
2. $T(\underline{X}) = 2x_1 + \mu x_2$
3. $T(\underline{X}) = \sum_{i=1}^{n} x_i$
4. $T(\underline{X}) = \prod_{i=1}^{n} x_i$
5. $T(\underline{X}) = \frac{\sigma^2}{n-1} \sum_{i=1}^{n} x_i$

Estadísticos

Algunos Estadísticos Importantes:

Media Muestral

$$T(\underline{X}) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Varianza Muestral

$$T(\underline{X}) = S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Estadísticos

Algunos Estadísticos Importantes:

Desviación Estándar Muestral

$$T(\underline{X}) = S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Mínimo

$$T(\underline{X}) = x_{(1)}$$

Máximo

$$T(\underline{X}) = x_{(n)}$$

Estadísticos

Teorema

Sean $X_1, X_2, .., X_n$ una m.a (n), y $\bar{X}$ la media muestral, entonces:

1. $\text{mín}_a \sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{X})^2$
2. $(n-1)S^2 = \sum_{i=1}^{n}(x_i - \bar{X})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{X}^2$

Estadísticos

Lema

Sean $X_1, X_2, .., X_n$ una m.a (n) desde una población y sea $g(x)$ una función tal que $E(g(X_1))$ y $Var(g(X_1))$ existen, entonces:

1. $E\left(\sum_{i=1}^{n} g(X_i)\right) = nE(g(X_1))$
2. $Var\left(\sum_{i=1}^{n} g(X_i)\right) = nVar(g(X_1))$

Universidad de Playa Ancha Facultad de Ciencias Naturales y Exactas | DEPARTAMENTO DISCIPLINARIO DE MATEMÁTICA Y ESTADÍSTICA

Estadísticos

Teorema

Sean $X_1, X_2, .., X_n$ una m.a (n), desde una población con media μ y varianza $\sigma^2 < \infty$, entonces:

1. $E(\bar{X}) = \mu$
2. $Var(\bar{X}) = \dfrac{\sigma^2}{n}$
3. $E(S^2) = \sigma^2$

Estadísticos de Orden

Definición (Estadísticos de Orden)

Los estadísticos de orden de una muestra aleatoria $(X_1, X_2, \ldots, X_n)$ *son los valores muestrales ordenados de forma creciente, los cuales son denotados por* $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, *donde:*

$$
\begin{aligned}
X_{(1)} &= \min_{1 \leq i \leq n} X_i \\
X_{(2)} &= \text{Segundo más pequeño} X_i \\
&\vdots \\
X_{(n)} &= \max_{1 \leq i \leq n} X_i
\end{aligned}
$$

Estadísticos de Orden

Teorema (Estadísticos de Orden)

Sean $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ estadísticos de orden de una muestra aleatoria $(X_1, X_2, \ldots, X_n)$, desde una población con cdf $F_X(x)$ y pdf $f_X(x)$, entonces la pdf de $X_{(j)}$ es:

$$f_{x_{(j)}}(x_{(j)}) = \frac{n!}{(j-1)!(n-j)!} f_x(x)[F_x(x)]^{j-1}[1 - F_x(x)]^{n-j}$$

Estadísticos de Orden

Teorema (Estadísticos de Orden)

Sean $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ estadísticos de orden de una muestra aleatoria $(X_1, X_2, \ldots, X_n)$, desde una población con cdf $F_X(x)$ y pdf $f_X(x)$, entonces la pdf conjunta de $X_{(i)}$ y $X_{(j)}$; $1 \leq i < j \leq n$ es:

$$f(x_{(i)}, x_{(j)}) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f(u) f(v) [F(u)]^{i-1} \times [F(v) - F(u)]^{j-1-i} [1 - F(v)]^{n-j}$$

Conceptos de convergencia

Teorema (Convergencia en Probabilidad)

Una sucesión $\{X_n\}_{n\in\mathbb{N}}$ de variables aleatorias ***converge en probabilidad*** *a la variable aleatoria* ***X****, lo que se denota $X_n \xrightarrow{p} X$ si: $\forall \varepsilon > 0$;*

$$\lim_{n\to\infty} P\left(|X_n - X| \geq \varepsilon\right) = 0 \quad \vee \quad \lim_{n\to\infty} P\left(|X_n - X| < \varepsilon\right) = 1$$

Universidad de Playa Ancha
Facultad de Ciencias Naturales y Exactas
DEPARTAMENTO DISCIPLINARIO DE MATEMÁTICA Y ESTADÍSTICA

Supongamos que tenemos una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ donde $X_n \sim \text{Uniforme}(0, \frac{1}{n})$. Queremos verificar si esta sucesión converge en probabilidad a cero, es decir, si $X_n \xrightarrow{p} 0$ cuando $n \to \infty$.
Para demostrar la convergencia en probabilidad, usamos la definición:

$$\lim_{n\to\infty} P(|X_n - 0| \geq \varepsilon) = \lim_{n\to\infty} P(X_n \geq \varepsilon)$$

Dado que $X_n \sim \text{Uniforme}(0, \frac{1}{n})$, la probabilidad de que $X_n \geq \varepsilon$ es igual a $\frac{1}{n\varepsilon}$ para $\varepsilon < \frac{1}{n}$ y 0 en otro caso.
Por lo tanto,

$$\lim_{n\to\infty} P(X_n \geq \varepsilon) = \lim_{n\to\infty} \frac{1}{n\varepsilon} = 0$$

para cualquier $\varepsilon > 0$. Así, demostramos que $X_n \xrightarrow{p} 0$ cuando $n \to \infty$.

Universidad de Playa Ancha Facultad de Ciencias Naturales y Exactas
DEPARTAMENTO DISCIPLINARIO DE MATEMÁTICA Y ESTADÍSTICA

Supongamos que tenemos una sucesión de variables aleatorias $\{Y_n\}_{n\in\mathbb{N}}$ donde $Y_n \sim \text{Normal}(0, \frac{1}{n})$. Queremos verificar si esta sucesión converge en probabilidad a cero, es decir, si $Y_n \xrightarrow{p} 0$ cuando $n \to \infty$.
Para demostrar la convergencia en probabilidad, utilizamos la definición:

$$\lim_{n\to\infty} P(|Y_n - 0| \geq \varepsilon) = \lim_{n\to\infty} P(|Y_n| \geq \varepsilon)$$

Dado que $Y_n \sim \text{Normal}(0, \frac{1}{n})$, la probabilidad $P(|Y_n| \geq \varepsilon)$ para cualquier $\varepsilon > 0$ es:

$$P(|Y_n| \geq \varepsilon) = 2 \cdot \Phi\left(-\sqrt{n}\varepsilon\right)$$

donde $\Phi(\cdot)$ es la función de distribución acumulada de la distribución normal estándar.
Entonces, $\lim_{n\to\infty} P(|Y_n| \geq \varepsilon) = 0$ ya que $\sqrt{n}\varepsilon \to \infty$ cuando $n \to \infty$.
Por lo tanto, demostramos que $Y_n \xrightarrow{p} 0$ cuando $n \to \infty$.

Conceptos de convergencia

Ejemplo:

Sea $X_n \sim Exp(n)$, pruebe que $X_n \xrightarrow{p} 0$. Es decir, la sucesión $X_1, X_2, ..$ converge en probabilidad a la variable aleatoria $X = 0$

Respuesta:

$$\begin{aligned}
\lim_{n\to\infty} P\left(|X_n - 0| \geq \varepsilon\right) &= \lim_{n\to\infty} P\left(|X_n| \geq \varepsilon\right) \\
&= \lim_{n\to\infty} P\left(X_n \geq \varepsilon\right) \quad (\text{ya que} X_n \geq 0) \\
&= \lim_{n\to\infty} e^{-n\varepsilon} \quad (\text{ya que} X_n \sim Exp(n)) \\
&= 0
\end{aligned}$$

Conceptos de convergencia

Teorema (Ley débil de los grandes números)

Sean $X_1, X_2, ..$ *variables aleatorias iid con* $\mathbb{E}\left(X_i\right) = \mu$ *y* $Var\left(X_i\right) = \sigma^2 < \infty$*, y se define* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ *, entonces* $\forall \varepsilon > 0$;

$$\lim_{n\to\infty} P\left(|\bar{X}_n - \mu| < \varepsilon\right) = 1$$

Esto es que $\bar{X}_n$ ***converge en probabilidad*** *a* μ*, es decir*

$$\bar{X}_n \xrightarrow{p} \mu.$$

Conceptos de convergencia

Teorema (Ley fuerte de los grandes números)

Sean $X_1, X_2, ..$ *variables aleatorias iid con* $\mathbb{E}(X_i) = \mu$ *y* $Var(X_i) = \sigma^2 < \infty$, *y se define* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, *entonces* $\forall \varepsilon > 0$;

$$P\left(\lim_{n\to\infty} \left|\bar{X}_n - \mu\right| < \varepsilon\right) = 1$$

Esto es que $\bar{X}_n$ ***converge de forma casi segura hacia*** μ.

Universidad de Playa Ancha
Facultad de Ciencias Naturales y Exactas
DEPARTAMENTO DISCIPLINARIO DE MATEMÁTICA Y ESTADÍSTICA

Conceptos de convergencia

Definición (Convergencia en Distribución)

Una sucesión de variables aleatorias $X_1, X_2, ..$ ***converge en distribución*** *a la variable aleatoria* $\boldsymbol{X}$ *si:*

$$\lim_{n\to\infty} F_{X_n}(X) = F_X(X)$$

en todos los puntos x donde $F_X(X)$ *es continua.*

Universidad de Playa Ancha Facultad de Ciencias Naturales y Exactas
DEPARTAMENTO DISCIPLINARIO DE MATEMÁTICA Y ESTADÍSTICA

Teorema Central del Límite

Teorema (Teorema Central del Límite)

Sean $X_1, X_2, ..$ *variables aleatorias iid con* $\mathbb{E}(X_i) = \mu$ *y* $Var(X_i) = \sigma^2 < \infty$,*y se define* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. *Sea* $G_n(x)$ *la función de distribución de* $\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}$. *Entonces*

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

Esto es , $\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}$ ***converge en distribución a una Normal Estándar***.

Principios de Reducción de datos

Principio de suficiencia

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

Principios de Reducción de datos

Definición (Estadístico Suficiente)

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

Teorema (Estadístico Suficiente)

Si $f(x|\theta)$ es la función de probabilidad o densidad conjunta de X, y $q(t|\theta)$ es la función de probabilidad o densidad de $T(X)$, entonces $T(X)$ es un estadístico suficiente para θ si y solo si:

$$\frac{f(x|\theta)}{q(T(x|\theta)}$$

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

Principios de Suficiencia

Ejemplo: Estadístico Suficiente

Sea $X_1, \ldots, X_n$ una m.a (n) con distribución Bernoulli de parámetro θ, es decir $X \sim Ber(\theta)$. Pruebe que $T(x) = (X_1 + \ldots + X_n)$ es un ESTADÍSTICO SUFICIENTE para θ.

Respuesta:

Si $X \sim Ber(\theta)$, entonces la distribución conjunta se define:

$$\prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$
$$= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Principios de Suficiencia

Ejemplo: CONTINUACIÓN

Como $T(x) = (X_1 + \ldots + X_n)$, suma de Bernoulli es una binomial, por lo que

$$q(T(x|\theta)) = \binom{n}{t}\theta^t(1-\theta)^{n-t} \quad ; \quad t = \sum_{i=1}^{n} x_i$$

luego se tiene

$$\frac{f(x|\theta)}{q(T(x|\theta))} = \frac{\theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}}$$

Principios de Suficiencia

Ejemplo: CONTINUACIÓN

$$
\begin{aligned}
\frac{f(x|\theta)}{q(T(x|\theta))} &= \frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
&= \frac{1}{\binom{n}{t}} \\
&= \frac{1}{\binom{n}{\sum_{i=1}^n x_i}}
\end{aligned}
$$

Entonces, como la razón no depende de θ, $T(X)$ es un ESTADÍSTICO SUFICIENTE para θ.

Principios de Suficiencia

Teorema (Teorema de Factorización)

Sea $f(x|\theta)$ la pdf o pmf conjunta de una muestra X. Un estadístico $T(x)$ es un estadístico suficiente para θ si y sólo si existen funciones $g(t|\theta)$ y $h(x)$ tales que:

$$f(x|\theta) = g(T(X)|\theta) \cdot h(x)$$