Inferencia Estadística

Manuel Gijón Agudo

Octubre 2018 -

Índice

1. Tema 0: Introducción a la inferencia estadística

1.1. bla

Definiciones 1 y 2: a **source** is a finite set $\mathcal{S}$ together with a set of random variables $(X_1, X_2, ...)$ whose range is $\mathcal{S}$.

If $P(X_n = \mathcal{S}_i)$ only depends on i and not on n then we say the source is **stationary** and if the X_n are independent then it's **memoryless**.

Insert example here

Definition 2: Let $\mathcal{T}$ be a finite set called **alphabet**. A map $\mathfrak{C} : \mathbb{S} \longrightarrow \mathbb{U}_{n\geq 1} T^n$ is called a **code**.

If $|T| = r$ then $\mathfrak{C}$ is a r**-ary code**.

A code extends from $\mathbb{S}$ to $T \cup T^2 \cup ...$ to $\mathbb{S} \cup \mathbb{S}^2 \cup ...$ to $T \cup T^2 \cup ...$ in obvious way.

insert example here

Definition 3: The **average word-lenght** of a code $\mathfrak{C}$ is $L(\mathfrak{C}) := \sum_{i=1}^{n} p_i l_i$ where l_i is the length of the image of the symbol of $\mathbb{S}$, which is emitted with probability p_i.

For now, we write $\mathfrak{C}$ to be the image of $\mathfrak{C}$.

1.2. bla

Definition 4: If for any sequencies $u_1...u_n = v_1...v_m$ in $\mathfrak{C}$ implies $m = n$ and $u_i = v_i$ for $i = 1, ..., n$ then we say that $\mathfrak{C}$ is **uniquely decodeable**.

insert example here

insert example here

insert example here

Let $\mathfrak{C}_0 = \mathfrak{C}$:

- $\mathfrak{C}_n := \{\omega \in T \cup T^2 \cup ... | u\omega = v \text{ for some } u \in \mathfrak{C}_{n-1}, v \in \mathfrak{C} \text{ or } u\omega = v \text{ for some } u \in \mathfrak{C}, v \in \mathfrak{C}_{n-1}\}$
- $\mathfrak{C}_\infty := \bigcup_{k\geq 1} \mathfrak{C}_k$

Since everythig is finite either $\mathfrak{C}_m = \emptyset$ for some m and then $\mathfrak{C}_n = \emptyset$ for $n \geq m$ or it will be periodic and start repeating.

Theorem 1: $\mathfrak{C}$ is uniquely decodeable $\Longleftrightarrow \mathfrak{C} \cap \mathfrak{C}_\infty = \emptyset$.

proof: Insert proof here

insert example here

insert example here

insert example here

Definition 5: A code is a **prefix-code** if no codeword is prefix of another (ie. $\mathfrak{C}_1 = \emptyset$).

A prefix code is uniquely decodeable.

Theorem 2: (**Kraft's inequality**) $\exists$ r-ary prefix code with word lengths $l_1, l_2, .., l_q \iff$

$$\sum_{i=1}^{q} r^{-l_i} \leq 1$$

proof: Insert proof here

insert example here

Theorem 3: (**McMillan's inequality**) $\exists$ r-ary uniquely decodeable code with word lengths $l_1, l_2, .., l_q \iff$

$$\sum_{i=1}^{q} r^{-l_i} \leq 1$$

proof: Insert proof here

1.3. Extensionof sources

2. Tema 1: Muestreo

2.1. Definiciones

2.2. Métodos de muestreo

2.2.1. Muestreo aleatorio simple

2.3. Distribuciones de muestreo

3. Tema 2: Estimación de parámetros

3.1. Definiciones

Definiciones:

- Sean $X_1, ..., X_n$ una secuencia de variables aleatorias independientes idénticamente distribuidas tales que $X \sim f(x;\theta)$ $\theta \in \Theta$.
- Definimos la **estimación puntual** el parámetro θ como el proceso de seleccionar un estadístico[1] T que mejor estima el valor del parámetro para esa población.
- Llamaremos a este estadístico $T = T(X_1, ..., X_n)$ que utilizamos para estimar θ un **estimador**.

Observaciones:

- Los estimadores son variables aleatorias.
- Usaremos sus propiedades estadísticas para estudiar su calidad y comparar entre ellos varios estimadores.
- Siempre tendremos un error en la estimación, nuestro objetivo será minimizarlo.

Definición: Decimos que un estimador $T_n = T(X_1, ..., X_n)$ para el parámetro θ es **consistente** cuando $\forall \epsilon > 0$:

$$\lim_{n \to \infty} P(|T_n - \theta| \geq \epsilon) = 0$$

Teorema: si T_n es una secuencia de estimadores tales que $E(T_n) \longleftrightarrow \theta$ y $V(T_n) \longleftrightarrow 0$ cuando $n \to \infty$ entonces T_n es consistente para el parámetro θ.

Definiciones:

- Definimos la **desviación** de un estimador T como:

  $$bias(T) = E(T) - \theta$$

- Sea T un estimador para θ. Decimos que el estimador es **no desviado** si $\forall \theta \in \Theta$:

  $$E(T) = \theta$$

  En caso contrario decimos que es **desviado**. Es obvio que en este caso $bias(T) \neq 0$.a

Para introducir el siguiente concepto usaremos un ejemplo concreto. Sean $X_1, ..., X_n$ una muestra aleatoria de una variable tal que $E(X) = \mu$ y $V(X) = \sigma^2$. Probar que:

$$E(\overline{X_n}) = \mu$$

[1]**Estadístico:** es una función medible que tiene como espacio de salida $(X_1, ..., X_n)$ una muestra estadística de valores.

$$E(S^2) = \frac{n-1}{n}\sigma^2$$

ENCONTRAR ESTA MIERDA Y CONTINUAR A PARTIR DE AQUÍ, MUHAHH-HAHHHAHHHAHHHA

3.2. Propiedades de los estimadores

3.3. Métodos para la obtención de estimadores

3.4. Métodos de remuestreo

Apéndice

3.5. Básicos de probabilidad

Caso discreto Probability density function:

Caso continuo

3.6. Distribuciones de probabilidad básicas

3.6.1. Distribuciones discretas

3.6.2. Distribuciones continuas

3.7. Ejercicios detallados

3.8. De probabilidad e introductorios

3.9. Tema 1

3.9.1. Métodos de muestreo

3.9.2. Distribucinoes de muestreo

3.10. Tema 2

3.10.1. Definiciones

3.10.2. Propiedades de los estimadores

3.10.3. Métodos para la obtención de estimadores

Encontrar un estadístico suficiente por el método de la máxima verosimilitud para θ para la distribución con la siguiente función de densidad bajo las condiciones $\theta > 0$ y $0 < x < 1$:

$$f_\theta(x) = \theta x^{\theta-1}$$

Primero calculamos la función de densidad conjunta (**función de verosimilitud**) que, asumiendo independencia, es el producto de las funciones de densidad para cada x_i.

$$\begin{aligned} l(\theta; x_1, ..., x_n) &= \prod_{i=1}^{n} f_\theta(x_i) \\ &= \prod_{i=1}^{n} \theta x_i^{\theta-1} \\ &= \theta^n \prod_{i=1}^{n} x_i^{\theta-1} \end{aligned} \tag{1}$$

Calculamos la el logaritmo de la función de verosimilitud para hacernos más sencillo calcular el **estimador de máxima verosimilitud (MLE)** $\widehat{\theta}$:

$$\begin{aligned} L(\theta; x_1, ..., x_n) &= \ln(l)(\theta; x_1, ..., x_n) \\ &= \ln\left(\theta^n \prod_{i=1}^{n} x_i^{\theta-1}\right) \\ &= n\ln(\theta) + (\theta - 1)\sum_{i=1}^{n} \ln(x_i) \end{aligned} \tag{2}$$

Hallamos el mínimo de la función para encontrar el MLE y comprobamos que es mínimo (pasos obviados aquí).

$$L_\theta(\theta; x_1, ..., x_n) = \frac{d}{d\theta} L(\theta; x_1, ..., x_n) = \frac{n}{\theta} + \sum_{i=1}^{n} \ln x_i$$

$$\begin{aligned} L_\theta(\theta; x_1, ..., x_n) &= \frac{d}{d\theta} L(\theta; x_1, ..., x_n) = 0 \\ \Rightarrow \widehat{\theta} &= \frac{-n}{\sum_{i=1}^{n} \ln x_i} \end{aligned} \tag{3}$$

Por el Teorema de Fisher–Neyman sabemos que nuestro estimador será suficiente sí y solo sí:

$$f_\theta(x) = h(x) g_\theta(T(x))$$

Donde T es el estimador. Observemos que aquí tenemos

3.10.4. Métodos de remuestreo

3.11. Tema 3