Inferencia Estadística

Tema 1: Introducción

C. Carleos, N.Corral, M.T. López

Departamento de Estadística Universidad de Oviedo

23 de septiembre de 2021

Inferencia Estadística

Objetivo: extraer conclusiones sobre el comportamiento en la población de una o varias variables con la información suministrada por una muestra.

- No se conocen los parámetros, o valores poblacionales, o aspectos importantes para el comportamiento de variables (p, μ, σ, independencia de dos variables)
- Dispone de la información suministrada por una muestra $(X_1, \ldots, X_n)$.

Probabilidad

A toda característica de la población(variable) se le asocia el conjunto de posibles resultados o espacio muestral, Ω, y una colección de sucesos $\mathcal{A}$ con estructura de σ-álgebra.
Sobre estos sucesos se define la probabilidad como una medida, entre 0 y 1, de la posibilidad de que ocurra un suceso

$$P : \mathcal{A} \longrightarrow [0, 1]$$

Axiomas de Kolmogorov

- $P(A) \geq 0 \quad A \in \mathcal{A}$
- $P(\Omega) = 1$
- Si $A_1, A_2, \ldots$ son sucesos mutuamente excluyentes (incompatibles dos a dos, o de intersección vacía), entonces: $P(A_1 \cup A_2 \cup \cdots) = \sum P(A_i)$

Probabilidad

Propiedades que se deducen de los axiomas:

- $P(\varnothing) = 0$
- Si $A \subseteq B$ entonces $P(A) \leq P(B)$
- $P(A) \leq 1$
- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probabilidad

- **Independencia** Dados dos sucesos $A, B \in \mathcal{A}$ se dice que son independientes si $P(A \cap B) = P(A)\,P(B)$.
- **Probabilidad condicionada** Dados dos sucesos $A, B \in \mathcal{A}$ con $P(B) > 0$ entonces: $P(A|B) = P(A \cap B)/P(B)$ ().
- **Teorema de la probabilidad total:** Sean $A_1, A_2, ..., A_n$ una partición del espacio muestral, tal que $P(A_i) > 0$ y sea B un suceso cualquiera con probabilidades condicionadas $P(B|A_i)$, entonces: $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$
- **Teorema de Bayes** Sea $\{A_1, A_2, ..., A_n\}$ una partición del espacio muestral, tal que $P(A_i) > 0$. Sea B un suceso con $P(B) > 0$ con probabilidades condicionadas $P(B|A_i)$, entonces: $P(A_i|B) = \sum_{i=1}^{n} \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$

Variable aleatoria

Una variable aleatoria (v.a.) X es una cuantificación de los resultados de un experimento aleatorio para trabajar con modelos probabilísticos sobre $\mathbb{R}$, $X : \Omega \to \mathbb{R}$

La definición formal involucra concepto de medibilidad: dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ $\quad X$ es una variable aleatoria si es una aplicación medible ($X^{-1}(B) \in \mathcal{A}, B \in \mathcal{B}$ con $\mathcal{B}$ la σ-álgebra de Borel en $\mathbb{R}$)

La variable aleatoria permite el "paso" de la probabilidad, P, del espacio inicial a otra probabilidad sobre subconjuntos de R, $\mathcal{B}$, llamada probabilidad inducida: $P(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}$

El comportamiento de la probabilidad inducida por X se puede caracterizar por su función de distribución $F : \mathbb{R} \to [0, 1]$, definida por

$$F(a) = P(\omega / X(\omega) \leq a), \; a \in \mathbb{R}$$

Las variables aleatorias pueden ser discretas o continuas

Función de distribución

Propiedades :

- $F(-\infty) = 0$.
- $F(\infty) = 1$.
- F es continua por la derecha.
- F es no decreciente.
- $P(X = a) = F(a) - F(a^-)$.

Según la forma de la función de distribución las variables aleatorias pueden ser discretas (función de distribución escalonada), continuas (función de distribución absolutamente continua) y mixtas.

Variable aleatoria discreta

Toma un número finito o numerable de valores x_i con probabilidades $p_i > 0$ verificando que $\sum p_i = 1$ (distribución de probabilidad)

- Su función de distribución es $F(a) = \sum_{x_i \leq a} p(x_i)$
- Su esperanza, si existe, es $\mu = E(X) = \sum_i x_i\, p(x_i)$ (representa un centro de la distribución)
- Su varianza, si existe, es $\sigma^2 = Var(X) = \sum_i (x_i - \mu)^2\, p(x_i)$ (mide dispersión respecto a la media)

Variable aleatoria continua

Puede tomar valores en un continuo, intervalo, queda caracterizada por su función de densidad $f(x) \geq 0$ y con $\int_R f(x)dx = 1$

- Su función de distribución es $F(a) = \int_{-\infty}^{a} f(x)dx$
- Su esperanza, si existe, es $\mu = E(X) = \int_R x\, f(x)dx$
- Su varianza, si existe, es $\sigma^2 = Var(X) = \int_R (x-\mu)^2 f(x)dx = E(X^2) - E^2(X)$

Distribuciones discretas

- Bernoulli, B(p), $\mu = p$, $\sigma^2 = p(1-p) = pq$
- Binomial, B(n,p), $\mu = np$, $\sigma^2 = np(1-p) = npq$
- Geométrica, G(p) (X=1,2,.....) $\mu = 1/p$, $\sigma^2 = q/p^2$
- Hipergeométrica, H(N,D,n);
  $\mu = nD/N$, $\sigma^2 = nD/N\,(N-D)/N\,(N-n)/(N-1)$
- Poisson, P(λ), $\mu = \lambda$, $\sigma^2 = \lambda$

Distribuciones continuas

- Uniforme, U(a,b), $\mu = (a+b)/2$, $\sigma^2 = (b-a)^2/12$
- Exponencial, E(a), $\mu = 1/a$, $\sigma^2 = 1/a^2$
- Gamma, $\gamma(p,a)$, $\mu = p/a$, $\sigma^2 = p/a^2$
- Normal, $N(\mu,\sigma)$
- Beta, $\beta(p,q)$
- Asociadas al muestreo en poblaciones normales: chi-cuadrado, t y F

Muestra aleatoria simple

Se obtiene por repetidas observaciones independientes y en las mismas condiciones de una variable aleatoria (v.a.) X
El vector aleatorio $(X_1, \ldots, X_n)$ es una **muestra aleatoria simple** de tamaño n de X si todas sus componentes son independientes y con la misma distribución

El comportamiento de la variable aleatoria determina el comportamiento de la muestra:

$P(x_1, \ldots, x_n) = \prod_i P_X(x_i)$ (caso discreto)
$f(x_1, \ldots, x_n) = \prod_i f_X(x_i)$ (caso continuo)

Estadístico

Un estadístico es cualquier aplicación medible del conjunto de posibles resultados muestrales en $\mathbb{R}^p$

$$T : (X_1, \ldots, X_n) \longrightarrow \mathbb{R}^p, (\quad \text{en general} \quad p \ll n).$$

- Con el manejo de estadísticos se pretende simplificar la información contenida en la muestra .
- Los estadísticos permiten trasladar la distribución de probabilidades de $\mathbb{R}^n$ al espacio $\mathbb{R}^p$ $(p \ll n)$.
- Estadísticos mas utilizados: $\overline{X}, \widehat{S}^2$

Cambio de variable

Teorema (de cambio de variable): Sea $f(x)$ la función de densidad de una v. a. X continua con soporte $S = (a, b)$ y g una función diferenciable en S con derivada distinta de cero (es decir inyectiva) entonces la función de densidad de la variable $Y = g(X)$ viene dada por:

$$f_Y(y) = f(g^{-1}(y))|(g^{-1})'(y)| \quad y \in g(a, b)$$

El teorema puede extenderse al caso de una función g no inyectiva en S pero tal que exista una partición finta o numerable del soporte $S = \cup_{i \in I} A_i$ donde g verifique esas condiciones en cada A_i. En este caso, la función de densidad de Y = g(X) es:

$$f_Y(y) = \sum_{i \in I} f_X(g_i^{-1}(y))|(g_i^{-1})'(y)| \quad y \in g(a, b)$$

Transformaciones de variable

- Sea $X \equiv U(0,1)$ e $Y = -log(X) \Leftrightarrow Y \equiv exp(1)$.
- Sea $X \equiv N(0,1)$ e $Y = X^2$, considerando las funciones $g_1(x) = x^2,\ x > 0$ y $g_2(x) = x^2,\ x \leq 0$, se tiene

$$f_Y(y) = \frac{f_{N(0,1)}(\sqrt{y}) + f_{N(0,1)}(-\sqrt{y})}{2\sqrt{y}} \quad y \in (0,\infty)$$

$$Y \equiv \gamma(\frac{1}{2},\frac{1}{2}) \equiv \chi^2_1.$$

- **variable tipificada** de X, v.a. con media, μ y varianza, σ^2 finitas es $Z_x = \frac{X-\mu}{\sigma}$ es adimensional.
- Si X es v.a. absolutamente continua con función de distribución F(x) $\Rightarrow$ Y=F(X) tiene distribución U(0,1).

Momentos muestrales

Dada una muestra aleatoria simple $(X_1, \ldots, X_n)$ de una variable aleatoria X se define los estadísticos:

1. **momento muestral de orden r** como $\overline{X^r} = \sum X_i^r / n$
   la media muestral es el momento muestral de orden 1:
$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$
2. **momento muestral centrado de orden r** como $\overline{(X - \overline{X})^r} = \sum (X_i - \overline{X})^r / n$
   la varianza muestral es el momento muestral centrado de orden 2
$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}$$

Estadístico $\overline{X}$

El estadístico **media muestral** $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ verifica

1. $\mathrm{E}(\overline{X}) = \mu \quad \mathrm{Var}(\bar{X}) = \sigma^2/n$
2. Si $X \equiv N(\mu, \sigma)$ entonces $\overline{X} \equiv N(\mu, \sigma/\sqrt{n})$.
3. X no normal se verifica el TCL: $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$
4. Otros resultados asintóticos: $\overline{X} \xrightarrow{P} \mu; \quad \overline{X} \xrightarrow{c.s.} \mu$

Estadístico $\widehat{S}^2$

El estadístico **cuasivarianza muestral**

$$\widehat{S}^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

verifica

1. $E(\widehat{S}^2) = \sigma^2$.
2. Si $X \equiv N(\mu, \sigma)$ entonces $\frac{(n-1)\widehat{S}^2}{\sigma^2} \equiv \chi^2_{n-1} = \gamma((n-1)/2, 1/2)$.
3. Si $X \equiv N(\mu, \sigma)$ entonces $\bar{X}$ y S^2 (o $\widehat{S}^2$) son independientes.

Otros estadísticos importantes

1. **coeficiente de asimetría** (de Fisher) $A_F = \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{S} \right)^3$
2. **coeficiente de curtosis** $K = \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{S} \right)^4 - 3.$