

Estadística I

Guillermo Julián Moreno

13/14 C1

Índice

1	Estadística descriptiva de datos univariantes	2
1.1	Estadísticos de tendencia central	2
1.2	Estadísticos de dispersión	2
1.3	Representación gráfica de datos	3
1.3.1	Estimadores núcleo o kernel	4
2	Estadística descriptiva de datos bivariantes	6
2.1	Representación gráfica	6
2.2	Regresión	6
3	Muestreo aleatorio	9
3.1	Conceptos de probabilidad	9
3.1.1	Distribuciones aleatorias	10
3.1.2	Criterios de convergencia	10
3.1.3	Desigualdades básicas	11
A	Distribuciones notables	11
B	Ejercicios	17
B.1	Tema 1 - Estadística descriptiva	17

1. Estadística descriptiva de datos univariantes

La estadística descriptiva es el conjunto de técnicas para resumir la información proporcionada por una gran masa de datos. El primer objetivo natural es resumir la información que proporcionan esos datos.

1.1. Estadísticos de tendencia central

Definición 1.1 Media.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Es la medida de tendencia central más utilizada. Es bastante sensible a los valores atípicos (*outliers*), observaciones anormalmente grandes que aparecen en el conjunto de datos por errores de transcripción o medición.

Definición 1.2 Mediana. Es el valor que divide a los datos en dos mitades, de tal forma que la mitad son menores y la otra mitad mayores que la mediana.

La mediana se calcula de la siguiente forma: dado un conjunto de datos $\{x_1, \dots, x_n\}$, la mediana es $x_{\frac{n+1}{2}}$ si n es impar y el promedio entre $x_{\frac{n}{2}}$ y $x_{\frac{n}{2}+1}$.

1.2. Estadísticos de dispersión

Definición 1.3 Varianza.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n n(x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n nx_i^2 - \bar{x}^2$$

Definición 1.4 Desviación típica.

$$\sigma = \sqrt{\sigma^2}$$

La desviación típica es la raíz de la varianza.

Definición 1.5 Cuantil. Para $p \in (0, 1)$ se llama cuantil p o q_p al valor que deja el $100p\%$ de los datos a la izquierda.

Definición 1.6 Cuartil. Los cuartiles son los tres datos que dejan a la izquierda el 25, 50 y 75 por ciento de los datos respectivamente. Es decir:

- $Q_1 = q_{0,25}$

- $Q_2 = q_{0,5}$. El cuartil dos es la mediana.
- $Q_3 = q_{0,75}$

Hay varios métodos para el cálculo de cuantiles. Para hacerlo a mano, podemos usar el siguiente método.

Si el dato en la posición $p(n+1)$ no es un número entero, entonces se interpola entre las observaciones ordenadas que están en la posición $\lfloor p(n+1) \rfloor$ y $\lfloor p(n+1) \rfloor + 1$ de la siguiente forma: sea j la parte entera de $p(n+1)$ y m la parte decimal. Entonces,

$$q_p = (1 - m)x_j + mx_{j+1}$$

Definición 1.7 Coeficiente de asimetría. El tercer momento con respecto a la media se define como

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$$

que, en su versión adimensional dividimos por σ^3 .

Al ser una función cúbica, los valores que se alejen mucho de la media tendrán un valor muy alto en valor absoluto (positivo o negativo según se aleje por la derecha o izquierda, respectivamente). Si la distribución de datos es muy asimétrica, los valores más altos no se cancelan con los valores altos del otro lado (porque no hay) y saldrá un valor más alejado de cero.¹

1.3. Representación gráfica de datos

Definición 1.8 Box-plot. El diagrama de caja o *box-plot* (imagen 1) nos permite visualizar las medidas de dispersión respecto a la mediana. Hay que añadir una nueva medida, el **rango intercuartílico**, la diferencia entre el primer y el tercer cuartil:

$$RI = Q_3 - Q_1$$

Definición 1.9 Histograma. El histograma se trata de una aproximación discreta a la función de densidad continua $f(t)$ de la variable que estamos midiendo. Es un diagrama de frecuencias que *mantiene la forma* de esa función de densidad.

Definimos una serie, las marcas de intervalos a_1^n, \dots, a_n^n , donde n es el número de intervalos y la longitud de cada intervalo es $h_n = a_{j+1}^n - a_j^n$. Sea el conjunto $\{x_i\}_{i=0, \dots, m}$ los datos de nuestra muestra. Entonces, el estimador, la función \hat{f}_n , se define de la siguiente forma:

$$\hat{f}_n(t) = \frac{\#(i \mid x_i \in (a_j^n, a_{j+1}^n])}{nh_n} = \frac{\sum_{i=1}^m \mathbb{1}_{(a_j^n, a_{j+1}^n]}(x_i)}{nh_n}$$

¹Está explicado como el p. culo, ya.

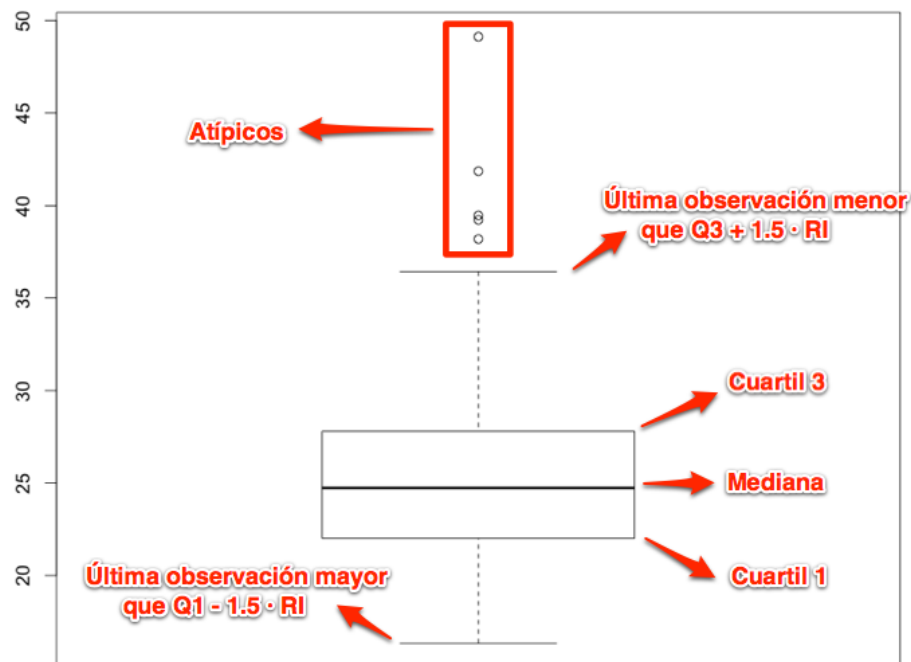


Figura 1: Diagrama de caja

Recordemos que

$$\mathbb{1}_A(n) = \begin{cases} 1 & n \in A \\ 0 & n \notin A \end{cases}$$

A grandes rasgos, lo que hace en una función es definir un número de intervalos fijos de ancho h_n . Al evaluar $\hat{f}^n(t)$ buscamos en qué intervalo cae t y contamos cuántas de nuestras mediciones caen también en ese intervalo.

1.3.1. Estimadores núcleo o kernel

Definición 1.10 Método de ventana móvil. El método de ventana móvil nos da una estimación de la función de densidad en un punto t midiendo los x_i que están en el intervalo de radio h_n centrado en t . Matemáticamente:

$$\hat{f}_n(t) = \frac{1}{n2h_n} \sum_{i=1}^n \mathbb{1}_{[t-h_n, t+h_n]}(x_i) = \frac{1}{n2h_n} \sum_{i=1}^n \mathbb{1}_{[-1,1]} \left(\frac{t-x_i}{h_n} \right)$$

Podemos reemplazar la función $\frac{1}{2} \mathbb{1}_{[-1,1]}$ por otra, llamada la función de densidad K , kernel o núcleo:

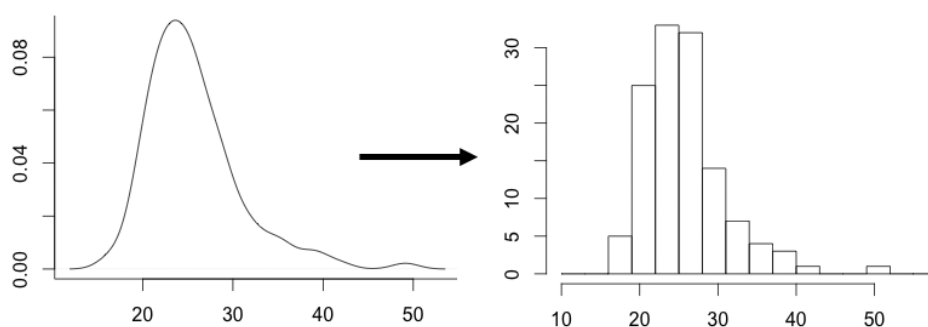


Figura 2: El histograma es una aproximación de la función de densidad real en base a la muestra que hemos obtenido.

Definición 1.11 Estimador núcleo. Dada una función de densidad K simétrica, no necesariamente positiva, definimos el estimador kernel como:

$$\hat{f}_n(t) = \frac{1}{n} \sum_{i=1}^n K_h(t - x_i) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{t - x_i}{h_n}\right)$$

con $K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$.

La elección del núcleo K no afecta especialmente a lo bien aproximada que esté la función de densidad. Sin embargo, sí que influye la selección de la ventana h_n (figura 3), también llamada *bandwith* en inglés. Si escogemos una ventana muy pequeña, damos demasiado peso a los datos de nuestra muestra. Si elegimos una ventana muy grande, nuestra muestra pierde importancia y podemos perder información importante.

La elección del h_n más habitual es el que minimiza la distancia L^2 entre \hat{f} y f , es decir, el parámetro que minimice $\int (\hat{f}_h - f)^2$. Sin embargo, hay un problema: no sabemos qué es f . Hay trucos que imagino que veremos más tarde.

Las funciones kernel más usadas son la uniforme, $\frac{1}{2} \mathbb{1}_{[-1,1]}$, la gaussiana $\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ y la de Epanechnikov, que matemáticamente es la que mejor aproxima f .

El estimador kernel $\hat{f}_n(t)$ es la función de densidad de una medida de probabilidad que es la convolución² de dos medidas de probabilidad: una, $K_h(x)$ (el kernel reescalado) y otra que da probabilidad $\frac{1}{n}$ a cada punto de la muestra $\{x_i\}$ (distribución o medida empírica).

Generación de datos del estimador kernel Supongamos que K es el núcleo gaussiano. Podemos generar datos artificiales de la densidad así:

$$x_i^0 = x_i^* + h_n Z_i, \quad i = 1, \dots, k$$

donde x_i^* es una observación elegida al azar entre los datos originales y Z_i una observación

²Ya aprenderemos en algún momento de nuestra vida qué narices es una convolución

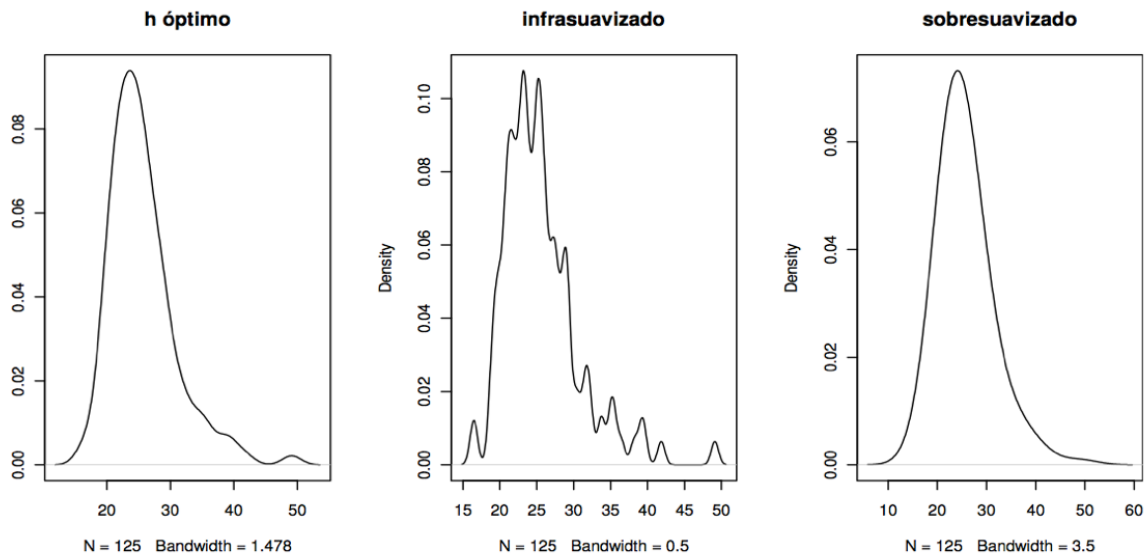


Figura 3: Los efectos que causa elegir una ventana más grande o más pequeña en el estimador

aleatoria con probabilidad $N(0, 1)$. Es decir, lo que hacemos es añadir un dato aleatorio de la muestra y sumamos una pequeña perturbación aleatoria.

2. Estadística descriptiva de datos bivariantes

En esta sección estudiaremos dos variables (X, Y) para explorar la relación entre ambas y tratar de inferir si existe una relación funcional para predecir los valores de una variable en función de los de la otra.

2.1. Representación gráfica

Definición 2.1 Diagrama de dispersión. El diagrama de dispersión representa cada variable en función de la otra para que podamos ver la posible relación entre ambas. Ver figura 4.

2.2. Regresión

Definición 2.2 Recta de regresión.

La recta de regresión de y sobre x es la recta de forma $\hat{y} = \hat{a} + \hat{b}x$ que más se aproxima a los datos, minimizando los cuadrados de la distancia:

$$(\hat{a}, \hat{b}) = \arg \min_{a, b} \sum_{i=1}^n (y_i - a - bx_i)^2$$

La recta de regresión se calcula obteniendo primero \hat{b} :

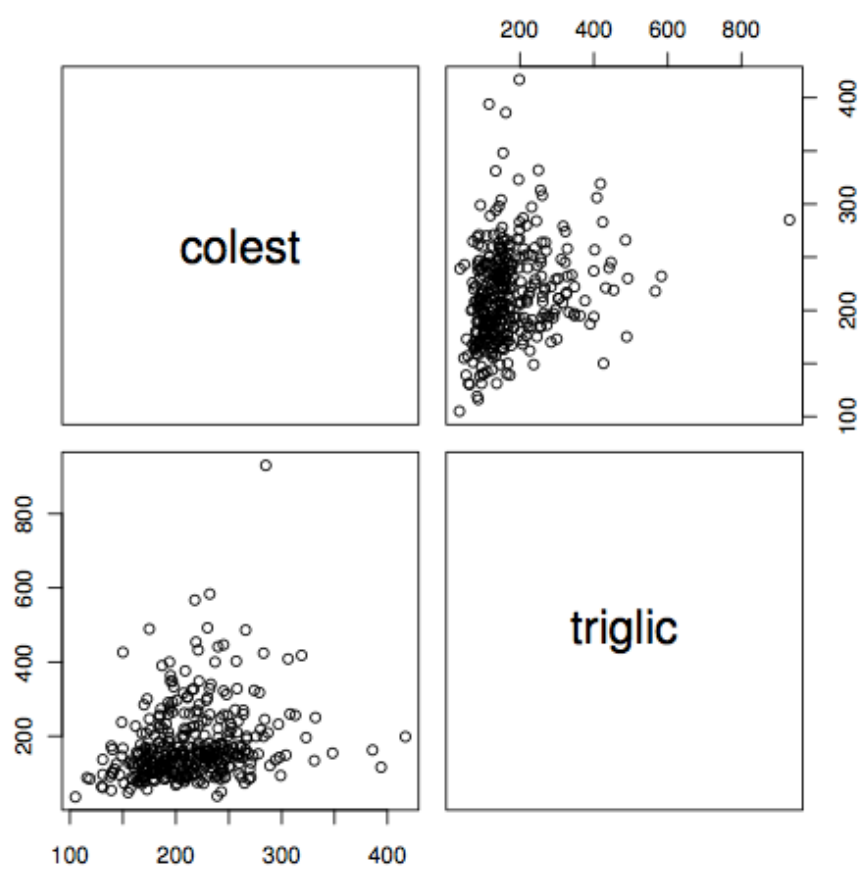


Figura 4: Diagrama de dispersión

$$\hat{b} = \frac{\sigma_{x,y}}{\sigma_x^2}$$

donde

$$\sigma_{x,y} = \frac{1}{n} \left(\sum_{i=1}^n x_i y_i \right) - \overline{xy}$$

y después, sabiendo que la recta pasa por el punto (\bar{x}, \bar{y}) , obtenemos \hat{a}

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

El valor b se denomina **coeficiente de regresión lineal** o parámetro de la regresión. Cada valor $e_i = y_i - \hat{y}_i$ se denomina **residuo**. Hay que notar que

$$\begin{aligned} \sum_{i=1}^n e_i &= \sum_{i=1}^n (y_i - \hat{a} - \hat{b}x_i) = \sum_{i=1}^n (y_i - (\bar{y} - \hat{b}\bar{x}) - \hat{b}x_i) = \\ &= \sum_{i=1}^n (y_i - \hat{b}x_i) - n\bar{y} + n\hat{b}\bar{x} = n\bar{y} - n\hat{b}\bar{x} - n\bar{y} + n\hat{b}\bar{x} = 0 \end{aligned}$$

Esta ecuación ($\sum_{i=1}^n e_i = 0$) junto con

$$\sum_{i=1}^n x_i e_i = 0$$

son las dos restricciones entre los residuos que nos dan la recta.

Definición 2.3 Varianza residual. La varianza residual s_R^2 o $\hat{\sigma}_e^2$ mide, aproximadamente el *error cuadrático* cometido en la aproximación dada por la recta de regresión:

$$s_R^2 = \hat{\sigma}_e^2 = \frac{1}{n} \sum_{i=1}^n e_i^2$$

Definición 2.4 Coeficiente de correlación lineal. El coeficiente de correlación lineal o coeficiente de Pearson

$$r = \frac{\hat{\sigma}_{x,y}}{\hat{\sigma}_x \hat{\sigma}_y}$$

que cumple las siguientes condiciones:

$$\begin{aligned} 0 &\leq r^2 \leq 1 \\ \hat{\sigma}_e^2 &= \hat{\sigma}_y^2 (1 - r^2) \\ r &= \hat{b} \frac{\hat{\sigma}_x}{\hat{\sigma}_y} \end{aligned}$$

nos indica el grado de ajuste lineal entre las dos variables. Un valor absoluto más cercano a 1 indica una correlación más fuerte. Un valor absoluto cercano a cero indica una correlación débil. El signo, positivo o negativo, indica si la correlación es creciente o decreciente.

3. Muestreo aleatorio

La muestra aleatoria de una cierta v.a. X se denomina como la **muestra aleatoria** o simplemente **muestra**.

Durante este tema, usaremos conceptos de Probabilidad, que repasaré aquí brevemente porque no me apetece escribir demasiado.

3.1. Conceptos de probabilidad

Definición 3.1 Distribución de una v.a..

$$P_X(B) = P(X \in B)$$

Definición 3.2 Función de distribución.

$$F(t) = P(X \leq t)$$

Definición 3.3 Media de una distribución. También llamada esperanza de X :

$$E(X) = \int_{-\infty}^{\infty} F(t) dt$$

Teorema 3.4 (Teorema de cambio de espacio de integración). *Sea g una función real medible tal que $E(g(X))$ es finita, entonces*

$$E(g(X)) = \int_{\mathbb{R}} g(x) dF(x) = \int_{\mathbb{R}} g(x) dP(x)$$

En particular

$$\mu = \int_{\mathbb{R}} x dF(x)$$

y

$$\sigma^2 = \int_{\mathbb{R}} (x - \mu)^2 dF(x)$$

Definición 3.5 Momentos. El momento μ_k es la esperanza de X elevado a una potencia de orden k . Es el valor esperado de la distancia de orden k con respecto a la media

$$\mu_k = E((X - \mu)^k)$$

3.1.1. Distribuciones aleatorias

Ver apéndice A (página 11).

3.1.2. Criterios de convergencia

Queremos buscar convergencias entre variables aleatorias.

Definición 3.6 Convergencia en distribución. Se dice que X_n converge débilmente o en distribución a X si la función de distribución de X_n $F_n(x)$ tiende a $F(x)$ para todo x punto de continuidad de F , donde F y F_n son las funciones de distribución de X y X_n respectivamente.

Esto es equivalente a decir que

$$\lim_{n \rightarrow \infty} P(X_n \in (-\infty, x]) = P(X \in (-\infty, x])$$

Notación:

$$X_n \xrightarrow[n \rightarrow \infty]{d} X \text{ ó } X_n \xrightarrow[n \rightarrow \infty]{w} X$$

Definición 3.7 Convergencia en probabilidad. Se dice que X_n converge en probabilidad a X si $\forall \epsilon > 0$ se tiene que

$$P(|X_n - X| > \epsilon) \xrightarrow[n \rightarrow \infty]{} 0$$

. Es decir, que para cualquier error que tomemos el error cometido en la aproximación va a tender a cero siempre que tome un X_n suficientemente grande.

Notación:

$$X_n \xrightarrow[n \rightarrow \infty]{P} X$$

Definición 3.8 Convergencia casi segura. También denotada c.s o a.s en inglés, convergencia en casi todo punto (c.t.p) o convergencia con probabilidad 1. Se dice que X_n converge a X casi seguro si el conjunto de puntos que no son convergentes tiende a ser vacío. Es decir

$$\mathbb{P} \left\{ X_n \xrightarrow[n \rightarrow \infty]{} X \right\} = 1$$

Más estrictamente, la condición se expresa como

$$\mathbb{P} \left\{ \omega \in \Omega \mid X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega) \right\} = 1$$

Notación

$$X_n \xrightarrow[n \rightarrow \infty]{c.s.} X$$

Teorema 3.9. *Se puede probar que si $\{X_n\}$ es una sucesión de variables aleatorias y X es variable aleatoria,*

$$X_n \xrightarrow[n \rightarrow \infty]{c.s.} X \Rightarrow X_n \xrightarrow[n \rightarrow \infty]{P} X \Rightarrow X_n \xrightarrow[n \rightarrow \infty]{d} X$$

Al contrario no tiene por qué darse.

Teorema 3.10 (Teorema de Slutsky). *Sean $\{X_n\}$, $\{Y_n\}$ sucesiones de variables aleatorias tales que $X_n \xrightarrow[n \rightarrow \infty]{d} X$, $Y_n \xrightarrow[n \rightarrow \infty]{P} c$ con $c \in \mathbb{R}$ constante. Entonces*

$$1. X_n + Y_n \xrightarrow[n \rightarrow \infty]{d} X + c$$

$$2. X_n \cdot Y_n \xrightarrow[n \rightarrow \infty]{d} X \cdot c$$

$$3. \frac{X_n}{Y_n} \xrightarrow[n \rightarrow \infty]{d} \frac{X}{c} \text{ si } c \neq 0.$$

3.1.3. Desigualdades básicas

Teorema 3.11 (Desigualdad de Markov). *Sea X v.a. Entonces, $\forall \epsilon > 0$,*

$$\mathbb{P}\{|X| > \epsilon\} \leq \frac{\mathbb{E}(X)}{\epsilon}$$

Teorema 3.12 (Desigualdad de Chebichev). *En las mismas condiciones del teorema anterior, se cumple que*

$$\mathbb{P}\{|X - \mathbb{E}(X)| > \epsilon\} \leq \frac{V(X)}{\epsilon^2}$$

A. Distribuciones notables

La distribución normal

Función de densidad:

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0$$

Momentos:

$$\mathbb{E}(X) = \mu, \quad \mathbb{V}(X) = \sigma^2$$

Aplicaciones: Es un modelo muy habitual para la distribución de magnitudes (en física, genética, etc.) que se pueden considerar como la suma de muchos pequeños efectos independientes (TCL). En Estadística aparece como distribución límite de muchos estadísticos que se usan para la inferencia.

La distribución exponencial

Función de densidad:

$$f(x) = \theta e^{-\theta x} \mathbb{I}_{[0, \infty)}(x) \quad (\theta > 0)$$

Momentos:

$$\mathbb{E}(X) = \frac{1}{\theta}, \quad \mathbb{V}(X) = \frac{1}{\theta^2}$$

Aplicaciones: en modelos de fiabilidad (tiempo de espera hasta que se produce una avería en un sistema).

Una propiedad interesante ("falta de memoria"): Si X sigue una distribución exponencial de parámetro θ , se tiene para $a > 0$ y $x > 0$,

$$\mathbb{P}\{X > x + a | X > x\} = e^{-\theta a}$$

(no depende de x).

La distribución gamma

Función de densidad:

$$f(x) = \frac{a^p}{\Gamma(p)} e^{-ax} x^{p-1} \mathbb{I}_{[0, \infty)}(x), \quad (a > 0, p > 0),$$

donde $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx$. Esta función verifica $\Gamma(p) = (p-1)!$ cuando $p \in \mathbb{N}$ y $\Gamma(p+1) = p\Gamma(p)$

Momentos:

$$\mathbb{E}(X) = \frac{p}{a}, \quad \mathbb{V}(X) = \frac{p}{a^2}$$

Aplicaciones: Cuando $p \in \mathbb{N}$ se llama distribución de Erlang y se usa en problemas de fiabilidad (tiempo de espera hasta p fallos), cantidad de lluvia caída, cuantía de las reclamaciones a las compañías de seguro, modelos de supervivencia,.... Para $a = 1/2$ $p = n/2$, con $n \in \mathbb{N}$, se llama distribución χ^2 con n grados de libertad y desempeña un importante papel en Estadística.

La distribución uniforme

Función de densidad:

$$f(x; a, \theta) = \frac{1}{\theta - a} \mathbb{I}_{[a, \theta]}(x), \quad (a, \theta \in \mathbb{R}, a < \theta)$$

Momentos:

$$\mathbb{E}(X) = \frac{\theta + a}{2}, \quad \mathbb{V}(X) = \frac{(\theta - a)^2}{12}$$

Aplicaciones:

La uniforme se relaciona con otras distribuciones a través de la siguiente **propiedad**: si X es v.a. con f. de dist. F continua, entonces $Y = F(X)$ tiene distribución uniforme estándar (i.e. con $a = 0, \theta = 1$). Esta propiedad se utiliza en los métodos de generación de números (pseudo-)aleatorios: se generan números de una v.a. Y uniforme estándar y se transforman con F^{-1} para obtener observaciones aleatorias con la distribución F .

La distribución beta

Función de densidad:

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbb{I}_{[0,1]}(x),$$

siendo $a, b > 0$ y Γ la función gamma que aparece en la definición de la distribución del mismo nombre.

Momentos: $\mathbb{E}(X) = \frac{a}{a+b}, \quad \mathbb{V}(X) = \frac{ab}{(a+b+1)(a+b)^2}.$

Aplicaciones: Dependiendo de los valores de los parámetros la densidad beta adopta formas muy variadas. Esta distribución (o sus versiones "reescaladas" en otros intervalos diferentes a $[0,1]$) proporciona un modelo muy flexible para describir variables aleatorias reales de soporte compacto.

La distribución de Weibull

Función de densidad:

$$f(x; \theta, k) = \frac{k}{\theta} \left(\frac{x}{\theta}\right)^{k-1} e^{-(x/\theta)^k} \mathbb{I}_{[0, \infty)}(x), \quad (k > 0, \theta > 0)$$

Momentos:

$$\mathbb{E}(X) = \theta \Gamma\left(1 + \frac{1}{k}\right), \quad \mathbb{V}(X) = \theta^2 \left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right) \right]$$

Aplicaciones:

Tiempos de supervivencia, problemas de fiabilidad en ingeniería, distribuciones de velocidad del viento en ingeniería, de periodos de incubación de algunas enfermedades, etc.

La distribución de Pareto

Función de densidad:

$$f(x; a, \theta) = \theta \frac{a^\theta}{x^{\theta+1}} \mathbb{I}_{[a, \infty)}(x), \quad (a > 0, \theta > 1)$$

Momentos:

$$\mathbb{E}_\theta(X) = \frac{\theta a}{\theta - 1}, \quad \mathbb{V}_\theta(X) = \left(\frac{a}{\theta - 1} \right)^2 \frac{\theta}{\theta - 2}, \quad \text{si } \theta > 2$$

Aplicaciones:

Distribución de ingresos, de reservas de petróleo, de área quemadas en bosques, de tamaños de ficheros enviados por e-mail, de tamaños de partículas,...

La distribución de Cauchy

Función de densidad:

$$f(x; \theta, a) = \frac{1}{\pi a \left[1 + \left(\frac{x - \theta}{a} \right)^2 \right]}$$

Momentos: No tiene momentos finitos

Aplicaciones: En el estudio de emisiones de partículas. Si Z es un ángulo aleatorio distribuido uniformemente entre $-\pi/2$ y $\pi/2$, $\tan(Z)$ tiene distribución de Cauchy. El cociente de dos v.a. normales estándar independientes tiene también distribución de Cauchy.

La distribución lognormal

Función de densidad:

$$f(x; m, a) = \frac{1}{x a \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\log x - m}{a} \right)^2} \mathbb{I}_{(0, \infty)}(x), \quad (m \in \mathbb{R}, a > 0)$$

Momentos:

$$\mathbb{E}(X) = e^{m + \frac{1}{2}a^2}, \quad \mathbb{V}(X) = (e^{a^2} - 1)e^{2m + a^2}$$

Aplicaciones: Si X tiene distribución lognormal, $\log X$ tiene distribución normal. Se usa en geología (tamaño de rocas sedimentarias) y en general en aquellos casos en los que una variable puede considerarse producto de muchos factores de pequeño efecto individual.

La distribución de Bernoulli

Función de probabilidad (o “de masa”): Se dice que una v.a. X tiene distribución de Bernoulli de parámetro $p \in [0, 1]$ (y se denota $X \sim B(1, p)$ o bien $X \sim Be(p)$) si

$$\mathbb{P}(X = 1) = p, \mathbb{P}(X = 0) = 1 - p.$$

Momentos:

$$\mathbb{E}(X) = p, \mathbb{V}(X) = p(1 - p)$$

Aplicaciones: Experimentos aleatorios “binarios”, i.e. con sólo dos posibles resultados.

La distribución binomial

Función de probabilidad: Se dice que una v.a. X tiene distribución binomial de parámetro $p \in [0, 1]$ (y se denota $X \sim B(n, p)$) si

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n$$

Momentos:

$$\mathbb{E}(X) = np, \mathbb{V}(X) = np(1 - p)$$

Aplicaciones: Número de éxitos en n pruebas de Bernoulli independientes en cada una de las cuales la probabilidad de éxito es p . La suma de n v.a. independientes con distribución $B(1, p)$ es $B(n, p)$.

La distribución de Poisson

Función de probabilidad: Se dice que una v.a. X tiene distribución de Poisson de parámetro $\lambda > 0$ (y se denota $X \sim \mathcal{P}(\lambda)$) si

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Momentos:

$$\mathbb{E}(X) = \lambda, \mathbb{V}(X) = \lambda$$

Aplicaciones: Frecuentemente se utiliza como modelo probabilístico para el estudio de fenómenos como el número de “sucesos” (tales como llegadas de clientes a un servicio, llamadas telefónicas a una centralita, accidentes,...) que se producen en un periodo de tiempo prefijado. Aparece como límite de la binomial en el siguiente sentido: Si $X_n \sim B(n, p_n)$ y $np_n \rightarrow \lambda > 0$, entonces

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

La distribución binomial negativa

Función de probabilidad: Se dice que una v.a. X tiene distribución binomial negativa de parámetros $p \in [0, 1]$ y $r \in \mathbb{N}$ (y se denota $X \sim BN(r, p)$) si

$$\mathbb{P}(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, r+2, \dots$$

Momentos:

$$\mathbb{E}(X) = \frac{r}{p}, \quad \mathbb{V}(X) = r \frac{1-p}{p^2}$$

Aplicaciones: Es un modelo discreto de “tiempo de espera”: En una sucesión de experimentos de Bernoulli con probabilidad éxito p , la distribución del número de pruebas necesarias para obtener r éxitos es $BN(r, p)$. La distribución $BN(1, p)$ se denomina **geométrica**.

B. Ejercicios

B.1. Tema 1 - Estadística descriptiva

Ejercicio 2: Demostrar que

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \min_{a \in \mathbb{R}} \sum_{i=1}^n (x_i - a)^2$$

Definimos una función

$$g(a) = \sum_{i=1}^n (x_i - a)^2$$

, buscamos su derivada

$$g'(a) = -2 \sum_{i=1}^n (x_i - a)$$

e igualamos a cero:

$$\begin{aligned} -2 \sum_{i=1}^n (x_i - a) &= 0 \\ \sum_{i=1}^n x_i - \sum_{i=1}^n a &= 0 \\ n\bar{x} &= na \\ \bar{x} &= a \end{aligned}$$

Esto quiere decir que la media muestral es el valor que minimiza la distancia con cada uno de los datos de la muestra.

Ejercicio 5: Determina si es verdadero o falso:

- a) Si añadimos 7 a todos los datos de un conjunto, el primer cuartil aumenta en 7 unidades y el rango intercuartílico no cambia.
- b) Si todos los datos de un conjunto se multiplican por -2, la desviación típica se dobla.

APARTADO A) *Añadir siete a todos los datos es una traslación, así que la distribución de los datos no cambia.*

APARTADO B) *Teniendo en cuenta que si multiplicamos todos los datos del conjunto por -2 la media también se multiplica por -2 , y sustituyendo en la fórmula de la varianza:*

$$\sigma' = \sqrt{\frac{1}{n} \sum_{i=1}^n (-2x_i)^2 - (-2\bar{x})^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n 4(x_i^2 - \bar{x}^2)} = \sqrt{4\sigma^2} = 2\sigma$$

Por lo tanto, la desviación típica sí se dobla.

APARTADO C) *Usando los cálculos del apartado anterior vemos que la varianza se multiplica por cuatro.*

APARTADO D) *Efectivamente: cambiar el signo haría una reflexión de los datos sobre el eje Y y la asimetría estaría orientada hacia el lado contrario.*

Índice alfabético

Box-plot, 3

Coeficiente

de asimetría, 3

de correlación lineal, 8

de Pearson, 8

Convergencia

débil, 10

en distribución, 10

en probabilidad, 10

Convergencia casi segura, 10

Cuantil, 2

Cuartil, 2

Desigualdad

de Chebichev, 11

de Markov, 11

Desviación típica, 2

Diagrama

de dispersión, 6

Distribución, 9

función de, 9

Estimador núcleo, 4

Histograma, 3

Media, 2

de una distribución, 9

Mediana, 2

Momentos, 10

Muestra, 9

Rango

intercuartílico, 3

Recta de regresión, 6

Regresión lineal

coeficiente de, 8

Residuo, 8

Skewness, 3

Teorema

de cambio de espacio de integración, 9

de Slutsky, 11

Varianza, 2

residual, 8

Ventana móvil, 4