#### **TEMA 1:**

#### Introducción a la inferencia estadística

Estimación L

Grado en Estadística Aplicada Curso 2019-2020

Cartagena99

<del>ELAMES PENTILS WHARTSAFF! TOBS (4)</del>S BNLLNER WIT AT SAFFE BONGS FAR POC

www.cartagena99.com no se hace responsable de la información contenida en el mormación contenida en el mormación contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información contenida en el lidocume no estado de la información de la lidocume no estado de la lid

#### INFERENCIA ESTADÍSTICA

En el **cálculo de probabilidades** se estudian varios aspectos de las distribuciones de probabilidad, asumiendo que la distribución considerada es **conocida**.

No obstante, en la práctica, la **distribución de una variable** de interés puede no ser conocida.

#### DEFINICIÓN: INFERENCIA ESTADÍSTICA

La inferencia estadística (o estadística matemática) es un área de la estadística compuesta por una serie de técnicas (basados en el cálculo de probabilidades) que permiten **obtener información** acerca de la ley de probabilidad de un fenómeno aleatorio, denominado población, **mediante la observación** del mismo.

Para obtener información sobre dicho fenómeno aleatorio, se llevan a cabo **repeticiones** del mismo o se seleccionan **individuos** de la población. El conjunto de dichas repeticiones/individuos recibe el nombre de **muestra**.

Cartagena99

ZŁAME SPŁATIC WHARTSA FLYTORS 14 ZNLLIVER WYATE AFF. 88943 FAP 8

www.cartagena99.com no se hace responsable de la información contenida en el similar mación contenida en el si El la miormación contenida en el la commenida en el la información contenida en el la contenida en el la conte

#### DEFINICIÓN: MUESTREO ALEATORIO SIMPLE

El muestreo aleatorio simple es un procedimiento para seleccionar muestras en el que todos los individuos de la población tienen la **misma probabilidad** de ser elegidos.

Matemáticamente, si X es la variable aleatoria población, una muestra aleatoria simple (m.a.s.) de tamaño n son n variables aleatorias independientes e identicamente distribuidas (v.a.i.i.d.)  $X_1, X_2, \ldots, X_n$  con la misma distribución que X.

 $X_i$  representa la **aleatoriedad** del *i*-ésimo individuo elegible en la muestra.

Sea  $f_X()$  la función de masa o de densidad (según corresponda) de la v.a. X. Entonces, la función de masa/densidad conjunta de la m.a.s. viene dada por:

$$f(x_1,\ldots,x_n)=f_X(x_1)\cdots f_X(x_n)=\prod_{i=1}^n f_X(x_i),$$

Cartagena99

PALLY STRATE AFF 88945 F4976

#### TIPOS DE INFERENCIA

- Inferencia paramétrica.- En este caso, se asume que la distribución de probabilidad del fenómeno de interés pertenece a una familia paramétrica, como pueden ser la distribución binomial o la normal, pero se desconoce el (o los) parámetros que rigen dicha distribución.
  - Así el objetivo de la inferencia paramétrica es obtener información sobre el valor de dicho(s) parámetro(s).
- Inferencia no paramétrica.- En este caso se desconoce la distribución de probabilidad del fenómeno de interés, por lo que se intentará obtener información sobre la misma u otras cuestiones como, por ejemplo, la independencia entre variables aleatorias.

Cartagena99

<del>ᢓĿAME SPENTIK WHANESEPLITOBS 14:</del> ONLLIVER WHATELFF 88945 FAPS O

www.cartagena99.com no se hace responsable de la información contenida en e

#### INFERENCIA PARAMÉTRICA: TÉCNICAS

Dentro de la inferencia paramétrica, podemos encontrar 3 tipos de técnicas:

- Estimación puntual.- Consiste en hacer un pronóstico (técnicamente una estimación) sobre el valor del parámetro desconocido. Se busca que el valor proporcionado sea lo más cercano posible al verdadero valor del parámetro.
- Estimación por intervalos.- En este caso, en lugar de ofrecer un único valor, se proporciona un rango de valores en el que existe una probabilidad alta (normalmente el 95 %) de encontrar el valor del parámetro desconocido.
- O Constrastes de hipótesis.- Consiste en proporcionar una regla de decisión para elegir entre posibles conjuntos de valores para el parámetro desconocido (por ejemplo, si es igual a 0 cm).

Cartagena99

Cartagena99

Cartagena99

Cartagena99

www.cartagena99.com no se hace responsable de la información contenida en el la información en el la informació

En esta asignatura estudiaremos **inferencia paramétrica**, resolviendo los problemas de estimación puntual y por intervalos.

Por ello, vamos a asumir que la v.a. población X se distribuye según una **familia** parámetrica de distribuciones conocida  $\mathcal{F}$ , pero cuyo parámetro (o vector de parámetros)  $\theta$  es **desconocido**:  $\mathcal{F} = \{f_{X,\theta}(x), \theta \in \Theta\}$ .

 $\Theta$  representa el conjunto de posibles valores de  $\theta$  y recibe el nombre de **espacio** paramétrico.

Para reforzar la idea de que los cálculos dependen del valor del parámetro, la función de verosimilitud se suele denotar de la siguiente manera:

$$f_{ heta}(x_1,\ldots,x_n)=f_{X, heta}(x_1)\cdots f_{X, heta}(x_n)=\prod^n f_{X, heta}(x_i),$$

Cartagena99

ELAMES PENTILS WHARES FILTORS (4)
ENLINER WYARES FARS

www.cartagena99.com no se hace responsable de la información contenida en e Si la miormación contenida en eliciolomenidas inclusionas la información contenida en el

Obtén la función de verosimilitud muestral si  $X_1, X_2, \ldots, X_n$  es una m.a.s. de X, que se distribuye según: a) B(1,p); b) U(a,b) y c)  $\Gamma(a,p)$ .



## ELAME SPENTIL WHARES FOR 1685 45 EARLING PRINTING WHARES FOR 1685 45 EARLING PRINTING WHARES FOR SOME FOR SOME

www.cartagena99.com no se hace responsable de la información contenida en el la información en el la informació

#### EJERCICIO 1

Obtén la función de verosimilitud muestral si  $X_1, X_2, \ldots, X_n$  es una m.a.s. de X, que se distribuye según: a) B(m,p); b)  $P(\lambda)$  y c)  $N(\mu,\sigma)$ .

Cartagena99

CLAMES PENTILS WHARTS APP. 1884 45 COLUMN FROM THE APP. SOLVES FAPS OF COLUMN FROM THE APP. SOLVES FARS OF COLUMN FROM THE APP. SOLVES FARS OF COLUMN FROM THE APP. SOLVES FARS OF COLUMN FROM THE APP. SOLVES FAR

www.cartagena99.com no se hace responsable de la información contenida en el la comenda en el la comenda en el la información contenida en el la comenda en el la información contenida en el la información en el la información contenida en el la información en el la informa

Como acabamos de ver, es relativamente sencillo obtener información sobre la muestra completa a través de su función de verosimilitud.

No obstante, generalmente no necesitamos información sobre toda la muestra si no solo sobre algún aspecto concreto de la misma.

#### DEFINICIÓN: ESTADÍSTICO MUESTRAL

Sea  $X_1, \ldots, X_n$  una m.a.s. de X. Llamamos **estadístico muestral** (o estimador) a cualquier función de la muestra T, que no dependa de parámetros desconocidos:

$$T:\Omega^n o \mathbb{R}^k$$

Generalmente, k=1, en cuyo caso podemos encontrar, por ejemplo, la media, el mínimo o el máximo.

Nótese que, dado que  $X_1, X_2, \ldots, X_n$  son v.a.,  $T(X_1, X_2, \ldots, X_n)$  es también una v.a. y, por tanto, tendrá una distribución.

Cartagena99

<u>ELAME STEARTIC WHARFS APH T889 44</u> <u>ONLLVER WIARTS APPS 889 45 FAPPS 0</u>

www.cartagena99.com no se hace responsable de la información contenida en e Si a lliformación contenida en silicocumento está circulas cual eleces o de ecro

Algunos de los estadísticos muestrales más utilizados son:

- Media muestral:  $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
- Varianza muestral:  $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i \bar{X})^2$
- Cuasivarianza muestral:  $S_c^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i \bar{X})^2 = \frac{n}{n-1} S^2$
- ullet Momento muestral respecto al origen de orden r:  $A_r = rac{1}{n} \sum_{i=1}^n X_i^r$
- Desviación típica muestral:  $S=\sqrt{S^2}$

$$ullet$$
 Cuasidesviación típica muestral:  $S_c=\sqrt{S_c^2}$ 

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 = A_2 - A_1^2$$

### Cartagena99

<del>PLAME SPENTIK WHATSA FELTBOBYAS</del> PALLINGR RATE AFFE 88945 F4978 C

www.cartagena99.com no se hace responsable de la información contenida en e Si la información contenida en elipcommenidas instancias sonalistenes o de echo

Generalmente, la distribución en el muestreo de los estadísticos muestrales depende de la **distribución concreta de** X. No obstante, en algunos casos concretos es posible obtener expresiones para su **esperanza** y su **varianza**.

#### PROPIEDADES DE LA MEDIA Y LA VARIANZA MUESTRAL

Sea  $X_1,\ldots,X_n$  una m.a.s. de X, tal que  $E[X]=\mu$  y  $Var[X]=\sigma^2$ . Entonces,

- $\circ$   $E[\bar{X}] = \mu$  (la esperanza de la media muestral es la media poblacional).
- $Var[\bar{X}] = \frac{\sigma^2}{n}$  (la varianza de la media muestral es la varianza poblacional entre el tamaño muestral por lo que disminuye cuando este último aumenta).
- $E[S^2] = \frac{n-1}{n} \sigma^2.$
- ${\bf \circ}\; E[S_c^2] = \sigma^2$  (la esperanza de la cuasivarianza muestral es la varianza poblacional).

Cartagena99

<del>ELAMES PENTILS WHARTS AFF! TOOP 145</del> ONLING PRIVATE AFF. 809 45 F49 PSC

www.cartagena99.com no se hace responsable de la información contenida en el la información en el la informació

#### EJERCICIO 2

Sea  $X_1, \ldots, X_n$  una m.a.s. de X, con  $\alpha_r = E[X^r]$ . Demuestra que, en ese caso:

$$E[A_r] = lpha_r \ Var[A_r] = rac{lpha_{2r} - lpha_r^2}{n}$$

Cartagena99

ELAMA O ENVIA WHATSAPP. 685 45 ENLL VER WITATE AFFERSOVES FARRO

www.cartagena99.com no se hace responsable de la información contenida en e

Como consecuencia de los resultados anteriores, tenemos que

• Si 
$$X_1, \ldots, X_n$$
 es m.a.s. de  $B(1, p)$ ,

$$ullet E[ar{X}] = p$$
  $Var[ar{X}] = rac{p(1-p)}{n}$ 

• Si 
$$X_1,\ldots,X_n$$
 es m.a.s. de  $B(m,p),$   
•  $E[ar{X}]=mp$   $Var[ar{X}]=rac{mp(1-p)}{a}$ 

• Si 
$$X_1, \ldots, X_n$$
 es m.a.s. de  $P(\lambda)$ ,

$$ullet E[ar{X}] = \lambda \qquad Var[ar{X}] = rac{\lambda}{n}$$

• Si 
$$X_1,\dots,X_n$$
 es m.a.s. de  $U(a,b),$   
•  $E[ar{X}]=rac{a+b}{2}$   $Var[ar{X}]=rac{(b-a)^2}{12\pi}$ 

• Si 
$$X_1, \ldots, X_n$$
 es m.a.s. de  $\Gamma(\alpha, \lambda)$ ,

# • $E[\bar{X}] = \frac{\alpha}{\lambda}$ $Var[\bar{X}] = \frac{\alpha}{\alpha \lambda^2}$

www.cartagena99.com.no.se.hace.responsable.de la información contenida en el

De una población con distribución de Poisson de parámetro  $\lambda$ , se obtiene una m.a.s.  $(X_1, \ldots, X_n)$ . Determina la distribución en el muestreo de la media muestral y comprueba que su esperanza coincide con la esperanza poblacional.



# ELAME SPENTIL WHATSAFF TO THE TOOL TO THE

www.cartagena99.com no se hace responsable de la información contenida en el la información contenida en el la información contenida en el la cont

#### EJERCICIO 3

Calcular la distribución en el muestreo del estadístico  $\sum_{i=1}^{n} X_i$  obtenido a partir de una m.a.s. de tamaño n de una población:

a) B(m,p)

b)  $Exp(\lambda)$ 

c)  $P(\lambda)$ 

d)  $N(\mu, \sigma)$ 

e)  $\Gamma(\alpha,\lambda)$ 

Cartagena99

ELAMES PENTILS WHATES PH TOO BY

Como acabamos de ver, existen ciertos estadísticos cuya distribución en el muestreo es relativamente sencilla de obtener, si la distribución poblacional cumple ciertas condiciones (como la reproductividad).

No obstante, este no es siempre así por lo que la distribución en el muestreo puede ser compleja de obtener. En esos casos, resulta de utilidad el resultado siguiente.

#### DISTRIBUCIÓN ASINTÓTICA DE LOS ESTADÍSTICOS MUESTRALES

Sea  $X_1, \ldots, X_n$  una m.a.s. de X, tal que  $E[X] = \mu$  y  $Var[X] = \sigma^2$ . Entonces, por la ley débil de los grandes números,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \stackrel{p}{ o} \mu$$

Así mismo, como la continuidad respeta la convergencia en probabilidad,

$$S^2 \stackrel{p}{\rightarrow} \sigma^2$$
  $S^2 \stackrel{p}{\rightarrow} \sigma^2$ 

Cartagena99

Por otro lado, por el Teorema Cantrel della Initiatica della della I ONLINED RAVATE LESSONS FORS

www.cartagepa99.com.no.se.hace.responsable.de la información contenida en e

El resultado anterior implica que, en la práctica, si el tamaño muestral n es suficientemente grande (superior a 30), independientemente de la distribución de X, se tiene que:

$$ar{X} \sim N\left(\mu, rac{\sigma}{\sqrt{n}}
ight)$$

Cartagena99

ELAMES PENTILS WHARTS APPLIES 45

ENLIVER WYARTS APPLIES FAPS

www.cartagena99.com no se hace responsable de la información contenida en el

#### DEFINICIÓN: ESTADÍSTICO DE ORDEN

Sea  $X_1, \ldots, X_n$  una m.a.s. de X. Llamaremos estadístico de orden al estadístico que ordena la muestra de menor a mayor valor y lo denotaremos como:

$$(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$$

Nótese que, aunque las  $X_1, \ldots, X_n$  son v.a.i.i.d., las variables  $X_{(1)}, X_{(2)},$  $\ldots, X_{(n)}$  no son independientes, ni identicamente distribuidas. De hecho, su función de densidad/masa conjunta se obtiene como:

$$f_{\theta}(x_{(1)}, \dots, x_{(n)}) = n! \prod^{n} f_{X, \theta}(x_{(i)}), \quad x_{(1)} < x_{(2)} < \dots < x_{(n)}$$

$$X_{(1)} = \min\{X_1, \dots, X_n\} \rightarrow F_{X_{(1)}}(y) = 1 - (1 - F_{X,\theta}(y))^n$$

 $X_{(n)} = \max\{X_1, \frac{1}{2} \mid X_1 \mid X_2 \mid X_3 \mid X_4 \mid X$  $\mu$ 

ww.cartagena99.com.no.se.hace.responsable.de la información contenida en e

#### EJERCICIO 4

Sea  $X_1,\ldots,X_n$  una m.a.s. de  $U(0,\theta)$ . Calcula la esperanza y la varianza del máximo muestral.

Cartagena99

ELAMES PENTIL WHATSAFF TO THE TOO SELLING FOR THE TOO SELLING FOR

www.cartagena99.com no se hace responsable de la información contenida en el silización de la información contenida en el la información en el la in