Índice

Tema 5

Introducción a la Inferencia Estadística. Muestreo en poblaciones normales

5.1 Introducción

El tipo de razonamiento seguido en Cálculo de Probabilidades para el estudio de los fenómenos aleatorios es deductivo: establecidas ciertas hipótesis sobre el mecanismo que genera los datos (modelo de distribución), se deducen propiedades sobre el fenómeno en cuestión (hallábamos la media, varianza, la probabilidad de que la variable tome valores en un intervalo, ...).

La **Inferencia Estadística** nos va a proporcionar la metodología para realizar el proceso inverso: a partir de un conjunto de datos experimentales, infiere, induce o estima características o propiedades del fenómeno bajo estudio.

En Estadística se denomina **población** al conjunto de entes reales o potenciales sobre los que se desea obtener información. Usualmente el estadístico o investigador no puede recabar información sobre todos los elementos que componen la población, bien por el elevado coste que esto supondría, porque la toma de información lleve consigo un proceso destructivo (p.e. controles de tiempo hasta que se funde un transistor), o por otros diversos motivos. De ahí que haya que realizar el estudio sobre unos cuantos elementos de la población denominado **muestra**.

Al proceso de selección de los individuos que componen la muestra se le llama muestreo. Existen varios tipos de muestreo dependiendo de múltiples factores. Nosotros sólo veremos en este curso el muestreo aleatorio simple.

Así con las técnicas de **Inferencia Estadística** nuestro **objetivo** va a ser: extraer conclusiones y generalizaciones sobre la población basándonos en la información suministrada por la muestra.

Pasamos a concretar e **introducir la terminología adecuada** para abordar el estudio de estas técnicas.

Se supone que **la propiedad que se desea estudiar en la población puede describirse en términos de una variable aleatoria** X que tendrá una función de distribución F. En cuanto a las hipótesis que se establezcan sobre F distinguimos:

- **Técnicas paramétricas:** si la distribución de F está especificada salvo algún parámetro. Por ejemplo, que por cómo se realiza el experimento o algún otro tipo de estudio previo podamos suponer que la característica (variable aleatoria) que nos interesa sigue una distribución de Poisson, $X \sim P(\lambda)$, con parámetro λ desconocido.

- **Técnicas no paramétricas:** la distribución de F sea desconocida (no sabemos casi nada o se tiene un conocimiento muy vago de ella).

Dado cualquier experimento aleatorio, sea X una variable aleatoria que cuantifica los resultados del mismo y denotemos por F a su función de distribución. Se desea obtener información sobre F, para ello se realizan observaciones de la variable aleatoria X en la muestra formada por n individuos. A los datos que obtenemos los denotamos por $x_1, x_2, ..., x_n$. Cada dato x_i se puede considerar como una realización de una variable aleatoria X_i que se distribuye como X. Así pues, los datos observados $(x_1, x_2, ..., x_n)$ se consideran una realización de un vector aleatorio $(X_1, X_2, ..., X_n)$, cuyas componentes tienen igual distribución que X.

Decimos que tenemos n variables aleatorias $X_1, ..., X_n$ con igual distribución que la variable aleatoria X de partida, o equivalentemente que $X_1, ..., X_n$ están idénticamente distribuidas (i.d.) como X. La **muestra**, es decir, esas variables $X_1, ..., X_n$ nos van a servir para obtener información sobre F.

En el caso en que las variables aleatorias además de idénticamente distribuidas sean **independientes** diremos que $X_1, .., X_n$ constituyen una **muestra aleatoria simple** de F (o de X). Lo resumimos formalmente en la siguiente definición.

Definición 5.1 *Sea X una variable aleatoria con función de distribución F, y sean $X_1, X_2, ..., X_n$ n variables aleatorias independientes e idénticamente distribuidas (i.i.d.) como X. A la colección $X_1, X_2, ..., X_n$ se le denomina* **muestra aleatoria simple** *de F (o de X). A n se le denomina* **tamaño de la muestra**.

Ejemplo 5.1 *Se desea estudiar el tiempo de funcionamiento (en años) de un tipo de transistores. Para ello un ingeniero selecciona n transistores de ese tipo, los prueba durante cierto tiempo y anota los instantes en que fallan: $x_1, \ldots, x_n$. Para tratar este problema podemos suponer que los tiempos de fallo observados, $(x_1, \ldots, x_n)$, son valores de variables aleatorias que tienen una distribución exponencial de parámetro λ, $X_1 \sim Exp(\lambda)$,..., $X_n \sim Exp(\lambda)$. Decimos entonces que estamos estudiando una población: $X \sim Exp(\lambda)$, y que $X_1, ..., X_n$ constituyen una muestra de esa población exponencial.*

Si los datos se han recogido de modo que podamos suponer además que $X_1, ..., X_n$ son independientes, tendremos entonces una muestra aleatoria simple.

Como las componentes de la muestra aleatoria simple (m.a.s.) son independientes e idénticamente distribuidas se tendrá que

$$P[X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n] = \prod_{i=1}^{n} P[X_i \leq x_i] = \prod_{i=1}^{n} F(x_i).$$

• Si X es **discreta** con función de probabilidad P, entonces la función de probabilidad conjunta de la muestra $P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n]$ viene dada por

$$P[X_1 = x_1, X_2 = x_2, ..., X_n = x_n] = \prod_{i=1}^{n} P[X_i = x_i] \ .$$

• Si X es **continua** con función de densidad f, entonces la función de densidad conjunta de la muestra $f(x_1, x_2, ..., x_n)$ viene dada por

$$f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i).$$

Las relaciones anteriores nos permiten contestar a preguntas formuladas sobre la muestra. Lo ilustramos en el siguiente ejemplo.

Ejemplo 5.2 *En el contexto del Ejemplo 5.1, suponemos que tenemos una m.a.s. y queremos calcular la probabilidad de que todos los transistores seleccionados en la muestra funcionen menos de 2 años:*

$P[$*"todos los transistores seleccionados en la muestra funcionen menos de 2 años"*$] =$

$= P[X_1 \leq 2, \ldots, X_n \leq 2] = \{independencia\} = P[X_1 \leq 2] \times \ldots \times P[X_n \leq 2] =$

$= \{idéntica\ distribución\} = \left(1 - e^{-2\lambda}\right)^n$.

5.2 Estadísticos y momentos muestrales

Puesto que toda la información sobre la población está contenida en la muestra (en las observaciones), nos planteamos en primer lugar cómo resumirla adecuadamente, con el objeto de facilitar su interpretación y reducir los datos. Esto lo haremos a través del concepto de estadístico.

Definición 5.2 *Dada $X_1, X_2, ..., X_n$ una m.a. de una variable aleatoria X, una función de la muestra $T(X_1, X_2, ..., X_n)$, con $T : \mathbb{R}^n \to \mathbb{R}$, se denomina* **estadístico** *siempre que no sea función de parámetros desconocidos.*

Ejemplo 5.3 *Sea $X \sim Be(p)$ con p desconocido. Consideramos una muestra de esta población $X_1, X_2, ..., X_n$. Sean:*

• $T_1(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} X_i$.

• $T_2(X_1, X_2, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$, *(media muestral).*

• $T_3(X_1, X_2, ..., X_n) = \bar{X} - p$.

$T_1(X_1, X_2, ..., X_n)$ y $T_2(X_1, X_2, ..., X_n)$ son estadísticos. En cambio, $T_3(X_1, X_2, ..., X_n)$ no es un estadístico porque es también función del parámetro p que es desconocido.

Pasamos a considerar algunos estadísticos importantes.

Definición 5.3 *Dada una muestra $X_1, X_2, ..., X_n$, se define el momento muestral de orden r, $r \geq 1$, como*

$$m_r = \frac{\sum_{i=1}^{n} X_i^r}{n} .$$

Si $r = 1$ se tiene la ***media muestral***

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} .$$

Definición 5.4 *Dada una muestra $X_1, X_2, ..., X_n$, se define el momento muestral centrado de orden r, $r \geq 1$, como*

$$m'_r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^r}{n} .$$

Si $r = 2$ se tiene la ***varianza muestral***

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n} . \tag{5.1}$$

Proposición 5.1 *La* ***varianza muestral*** *definida en (5.1) puede expresarse como*

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}^2 .$$

Nótese que un estadístico $T = T(X_1, X_2, ..., X_n)$ es una variable aleatoria por ser una función de variables aleatorias, y por tanto tendrá una distribución, esperanza y varianza.

5.2.1 Media muestral. Propiedades

Teorema 5.1 *Consideremos una población descrita por una variable aleatoria X con media μ y varianza σ^2 finitas. Sea $X_1, \ldots, X_n$ una m.a.s. de dicha población. Entonces*

1. $E[\bar{X}] = \mu$.
2. $Var[\bar{X}] = \frac{\sigma^2}{n}$.

5.2.2 Varianza muestral. Propiedades

Teorema 5.2 *Consideremos una población descrita por una variable aleatoria X con media μ y varianza σ^2 finitas. Sea $X_1, \ldots, X_n$ una m.a.s. de dicha población. Entonces*

$$E[S^2] = \sigma^2 \left(\frac{n-1}{n} \right) . \tag{5.2}$$

Definición 5.5 *(Cuasivarianza muestral)*

$$S_c^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} .$$

De (5.2) y puesto que $nS^2 = (n-1)S_c^2$, se tiene que

$$E[S_c^2] = \sigma^2 . \tag{5.3}$$

Ya indicamos que un estadístico $T = T(X_1, X_2, ..., X_n)$ es una variable aleatoria y hemos calculado la media y varianza de algunos de ellos. Otro aspecto importante a destacar es que al ser una variable aleatoria también tendrá una distribución. A la distribución de T se le denomina **distribución muestral** de T o **distribución en el muestreo** de T. Esta distribución dependerá

- de la expresión de T, y
- de la distribución de las X_i.

Ya conocemos la distribución de algunos estadísticos:
• Si $X_1, X_2, ..., X_n$ es una m.a.s. de una población de Bernoulli, $Be(p)$, entonces

$$T(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} X_i \sim B(n, p)$$

• Si $X_1, X_2, ..., X_n$ es una m.a.s. de una población de Poisson $\mathcal{P}(\lambda)$, entonces

$$T(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} X_i \sim \mathcal{P}(n\lambda)$$

• Si $X_1, X_2, ..., X_n$ es una m.a.s. de una población normal $N(\mu, \sigma^2)$, entonces

$$T(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} X_i \sim N(n\mu\ ,\ n\sigma^2)$$

Por su importancia en Estadística dedicamos la siguiente sección al estudio de la distribución de algunos estadísticos cuando se tienen muestras de poblaciones normales.

5.3 Distribuciones en el muestreo de poblaciones normales

5.3.1 Distribución de la media y diferencia de medias

Teorema 5.3 *Sea $X_1, X_2, ..., X_n$ una m.a.s. procedente de una población $N(\mu, \sigma^2)$. Entonces*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right) .$$

Teorema 5.4 *Sean $X_1, X_2, ..., X_n$ una m.a.s. de tamaño n procedente de una población $N(\mu_X, \sigma_X^2)$, e $Y_1, Y_2, ..., Y_m$ una m.a.s. de tamaño m independiente de la anterior y procedente de una población $N(\mu_Y, \sigma_Y^2)$. Entonces*

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)$$

Definición 5.6 (Punto crítico de la distribución $N(0,1)$) *Sea $\alpha \in (0,1)$, mediante z_α representaremos a aquel $x \in \mathbb{R}$ tal que*

$$z_\alpha = x\ /\ \Phi(x) = \alpha,$$

donde $\Phi(\cdot)$ denota a la función de distribución de la $N(0,1)$.

z_α se denomina **punto crítico a nivel** *α de la distribución $N(0,1)$.*

Además, como la distribución $N(0,1)$ es simétrica respecto del origen, se verifica que $z_\alpha = -z_{1-\alpha}$.

La función de distribución de la $N(0,1)$ está tabulada. Es fácil por tanto calcular puntos críticos.

Ejemplo 5.4 *Los puntos críticos a nivel $\alpha = 0.95$, y $\alpha = 0.05$ en la $N(0,1)$ son $z_{0.95} = 1.645$, $z_{0.05} = -z_{0.95} = -1.645$.*

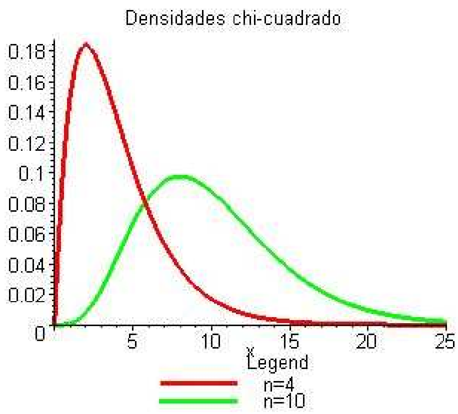
5.3.2 Distribución chi-cuadrado

Definición 5.7 *Sean $Z_1, Z_2, ..., Z_n$ n variables aleatorias i.i.d. según una distribución $N(0,1)$. Consideremos la variable aleatoria*

$$V = Z_1^2 + Z_2^2 + ... + Z_n^2 \ .$$

A la distribución de la variable aleatoria V se le denomina ***distribución chi-cuadrado con n grados de libertad****, lo que se representa como $V \sim \chi_n^2$*

Algunos gráficos de densidades chi-cuadrado:



Observamos que sólo toma valores positivos y es asimétrica.

Teorema 5.5 (Propiedades de la distribución chi-cuadrado)

- *Si $V \sim \chi_n^2$ entonces $E[V] = n$ y $Var[V] = 2n$.*
- *Si $U \sim \chi_n^2$ y $V \sim \chi_m^2$ con U y V independientes, entonces $U + V \sim \chi_{n+m}^2$.*

Teorema 5.6 (Teorema de Fisher) *Sea $X_1, X_2, ..., X_n$ una m.a.s. de tamaño n procedente de una población $N(\mu, \sigma^2)$. Entonces los estadísticos $\bar{X}$ y S_c^2 son independientes y además*

$$\frac{(n-1)S_c^2}{\sigma^2} \sim \chi_{n-1}^2$$

Definición 5.8 (Puntos críticos de la distribución chi-cuadrado) *Sea $\alpha \in (0,1)$, mediante $\chi_{n,\alpha}^2$ representaremos a aquel $x \in \mathbb{R}$ tal que*

$$\chi_{n,\alpha}^2 = x \ / \ P[\chi_n^2 \leq x] = \alpha.$$

$\chi_{n,\alpha}^2$ se denomina **punto crítico a nivel** *α de la distribución χ_n^2.*

Ejemplo 5.5 *Los puntos críticos a nivel $\alpha = 0.10$, y $\alpha = 0.90$ en la distribución chi-cuadrado con 15 grados de libertad son $\chi_{15,0.10}^2 = 8.546756$, y $\chi_{15,0.90}^2 = 22.307130$.*

5.3.3 Distribución t de Student

Definición 5.9 *Sean Z y V dos variables aleatorias independientes de modo que $Z \sim N(0,1)$ y $V \sim \chi_n^2$. Sea*

$$T = \frac{Z}{\sqrt{V/n}}$$

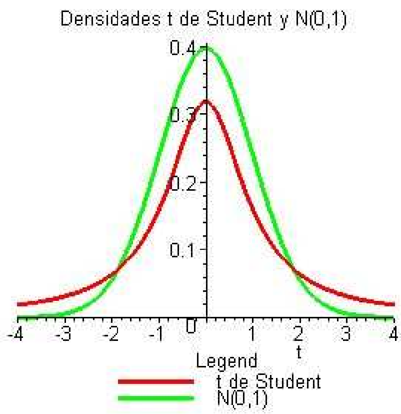
A la distribución de la variable aleatoria T se le denomina **distribución t de Student con n grados de libertad**, *y se representa $T \sim t_n$*

Teorema 5.7 (Propiedades de la distribución t de Student)

- *La distribución t de Student toma valores en todo $\mathbb{R}$ y es simétrica respecto de cero.*
- *Por ser simétrica respecto del origen, se tiene que si $T \sim t_n$ entonces*

$$P[T \leq -x] = 1 - P[T \leq x]$$

- *El gráfico de la función de densidad de la t de Student es similar a la de la $N(0,1)$, sólo que hay más área en las colas de la distribución. En el siguiente gráfico se comparan.*



- *Sea $X_1, X_2, ..., X_n$ una m.a.s. de tamaño n procedente de una población $N(\mu, \sigma^2)$. Entonces*

$$\sqrt{n}\frac{\bar{X} - \mu}{S_c} \sim t_{n-1}$$

Definición 5.10 (Puntos críticos de la distribución t de Student) *Sea $\alpha \in (0,1)$, mediante $t_{n,\alpha}$ representaremos a aquel $x \in \mathbb{R}$ tal que*

$$t_{n,\alpha} = x \ \ / \ \ P[t_n \leq x] = \alpha.$$

$t_{n,\alpha}$ se denomina **punto crítico a nivel** *α de la distribución t_n.*

Por la simetría respecto del origen se verifica que $t_{n,\alpha} = -t_{n,1-\alpha}$.

Ejemplo 5.6 *Los puntos críticos a nivel $\alpha = 0.975$, y $\alpha = 0.025$ en la distribución t de Student con 10 grados de libertad son $t_{10,0.975} = 2.228139$, y $t_{10,0.025} = -t_{10,0.975} = -2.228139$.*

5.3.4 Distribución F de Snedecor

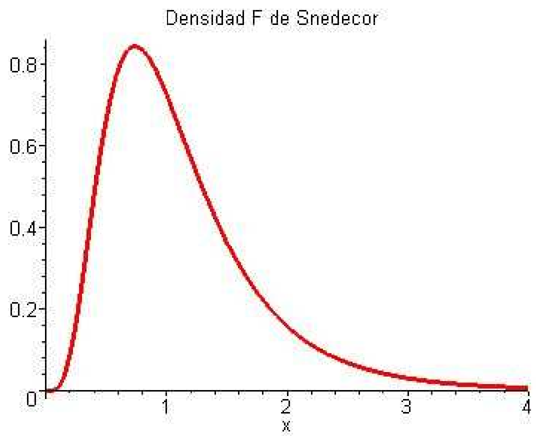
Definición 5.11 *Sean U y V dos variables aleatorias independientes de modo que $U \sim \chi^2_n$ y $V \sim \chi^2_m$. Sea*

$$F = \frac{U/n}{V/m}$$

A la distribución de la variable aleatoria F se le denomina **distribución $\mathcal{F}$ de Snedecor con n y m grados de libertad**, *y se representa $F \sim \mathcal{F}_{n,m}$.*

n: grados de libertad del numerador, m: grados de libertad del denominador.

Ejemplo de gráfico de una densidad F de Snedecor:



Destacamos que sólo toma valores positivos y es asimétrica.

Teorema 5.8 (Propiedades de la F de Snedecor)

- *Si $F \sim \mathcal{F}_{n,m}$ entonces $\dfrac{1}{F} \sim \mathcal{F}_{m,n}$.*
- *Sean $X_1, X_2, ..., X_n$ una m.a.s. de tamaño n procedente de una población $N(\mu_X, \sigma^2_X)$, e $Y_1, Y_2, ..., Y_m$ una m.a.s. de tamaño m independiente de la anterior y procedente de una población $N(\mu_Y, \sigma^2_Y)$. Entonces*

$$\frac{\sigma^2_Y}{\sigma^2_X}\frac{S^2_{c,X}}{S^2_{c,Y}} \sim \mathcal{F}_{n-1,m-1},$$

donde $S^2_{c,X} = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$ *y* $S^2_{c,Y}\dfrac{1}{m-1}\displaystyle\sum_{j=1}^{m}(Y_j - \bar{Y})^2$.

Definición 5.12 (Puntos críticos de la distribución F de Snedecor) *Sea $\alpha \in (0,1)$, mediante $F_{n,m,\alpha}$ representaremos a aquel $x \in \mathbb{R}$ tal que*

$$F_{n,m,\alpha} = x \ / \ P(\mathcal{F}_{n,m} \leq x) = \alpha.$$

$F_{n,m,\alpha}$ se denomina **punto crítico a nivel** *α de la distribución F de Snedecor con n y m grados de libertad.*

Se verifica que $F_{n,m,\alpha} = \dfrac{1}{F_{m,n,1-\alpha}}$.

Ejemplo 5.7 *Los puntos críticos a nivel* $\alpha = 0.90$ *y* 0.10 *para una distribución* F *de Snedecor con* $n = 5$ *y* $m = 12$ *grados de libertad son*

$$F_{5,12,0.90} = 2.394022$$
$$F_{5,12,0.10} = \frac{1}{F_{12,5,0.90}} = \frac{1}{3.268239} = 0.305975\ .$$

5.4 Otros resultados

Teorema 5.9 *(Teorema Central del Límite) Consideremos una población descrita por una v.a.* X *con media* $\mu = E[X]$ *y varianza* $\sigma^2 = Var[X]$*, ambas finitas. Sea* $X_1, \ldots, X_n$ *una m.a.s. de dicha población. Si el tamaño muestral* n *es elevado, entonces podemos aproximar la distribución de* $\bar{X}$ *por una distribución normal* $N\left(\mu, \frac{\sigma^2}{n}\right)$ *.*

Como una aplicación importante del resultado anterior podemos citar la siguiente.

Corolario 5.1 *Dada* $X_1, \ldots, X_n$ *una m.a.s de una población de Bernoulli,* $X \sim Be(p)$*. Si el tamaño de la muestra* n *es elevado, entonces podemos aproximar la distribución de* $\bar{X}$ *por la de una distribución normal* $N\left(p, \frac{p(1-p)}{n}\right)$ *, pues en una población de Bernoulli,* $X \sim Be(p)$*, se tiene que* $E[X] = p$*, y* $Var[X] = p(1-p)$ *.*

En este caso particular $\bar{X}$ *se suele denotar por* $\hat{p}$ *: "proporción de éxitos observados en la muestra", y podríamos expresar este resultado como*

$$\hat{p} \ \approx \ N\left(p\ ,\ \frac{p(1-p)}{n}\right).$$