

Fundamentos de Inferencia Estadística

Tema 1: Muestreo y distribuciones muestrales

Francisco Javier Mercader Martínez

1) Sea X variable aleatoria con distribución Bernoulli, de parámetro $p(X \sim b(p))$ y sea X_1, X_2, X_3 una muestra aleatoria simple (m.a.s) de X . Se pide:

a) Estudiar la distribución del vector (X_1, X_2, X_3) .

Dado que X_1, X_2, X_3 son una muestra aleatoria simple de $X \sim \text{Bernoulli}(p)$, entonces las siguientes propiedades son ciertas:

- 1) X_1, X_2, X_3 son independientes e idénticamente distribuidas.
- 2) Cada $X_i \sim \text{Bernoulli}(p)$, es decir, la probabilidad de éxito $P(X_i = 1) = p$ y $P(X_i = 0) = 1 - p$.

El vector (X_1, X_2, X_3) tiene una **distribución multinomial** con 2 posibles resultados (0 o 1) para cada componente. La distribución conjunta es:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = p^{x_1+x_2+x_3} (1-p)^{3-(x_1+x_2+x_3)},$$

donde $x_1, x_2, x_3 \in \{0, 1\}$.

Esto corresponde a la **distribución conjunta** de 3 variables Bernoulli independientes.

b) Estudiar la distribución en el muestreo del estadístico $\frac{X_1 + X_2 + X_3}{3}$.

Es estadístico $\bar{X} = \frac{X_1 + X_2 + X_3}{3}$ es el **promedio muestral** de las 3 variables. Para analizar su distribución:

- 1) $S = X_1 + X_2 + X_3$ sigue una **distribución binomial** porque es la suma de $n = 3$ variables Bernoulli independientes:

$$S \sim B(n = 3, p).$$

La función de probabilidad de S es:

$$P(S = k) = \binom{3}{k} p^k (1-p)^{3-k}, \quad k = 0, 1, 2, 3.$$

- 2) El estadístico $\bar{X} = \frac{S}{3}$ simplemente escala los valores posibles de S dividiéndolos por 3. Los valores posibles de \bar{X} son:

$$\bar{X} \in \left\{ 0, \frac{1}{3}, \frac{2}{3}, 1 \right\}.$$

- 3) La probabilidad de cada valor de \bar{X} es proporcional a la probabilidad de los valores correspondientes de S :

$$P\left(\bar{X} = \frac{k}{3}\right) = P(S = k) = \binom{3}{k} p^k (1-p)^{3-k}, \quad k = 0, 1, 2, 3.$$

Por lo tanto, la distribución de \bar{X} es discreta y está determinada por la distribución binomial de S .

- 2) Sea X variable aleatoria con distribución $\text{Exp}(\lambda)$. Sea (X_1, \dots, X_n) una m.a.s de X , estudiar la distribución en el muestreo de $S = \sum_{j=1}^n X_j$.

Dado que $X \sim \text{Exp}(\lambda)$ (exponencial con parámetro $\lambda > 0$), y que (X_1, \dots, X_n) es una muestra aleatoria simple (m.a.s) de X , podemos analizar la distribución del estadístico $S = \sum_{j=1}^n X_j$.

Propiedades relevantes:

- 1) **Distribución de la suma de variables exponenciales independientes:** Si X_1, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas ($X_i \sim \text{Exp}(\lambda)$), entonces la suma:

$$S = \sum_{j=1}^n X_j$$

sigue una distribución **Gamma** con parámetros n y λ . Esto se denota como:

$$S \sim \text{Gamma}(n, \lambda),$$

donde:

- n es el parámetro de forma.
- λ es el parámetro de escala.

Distribución Gamma:

La función de densidad de probabilidad de una variables aleatoria $S \sim \text{Gamma}(n, \lambda)$ está dada por:

$$f_S(s) = \begin{cases} \frac{\lambda^n}{\Gamma(n)} s^{n-1} e^{-\lambda s} & s > 0 \\ 0 & s \leq 0 \end{cases}$$

donde:

- $\Gamma(n)$ es la función gamma (para $n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$).
- s^{n-1} y $e^{-\lambda s}$ controlan la forma y decaimiento de la densidad.

Propiedades del estadístico S :

- 1) **Esperanza ($E[S]$):**

Si $S \sim \text{Gamma}(n, \lambda)$, entonces:

$$E[S] = \frac{n}{\lambda}.$$

- 2) **Varianza ($\text{Var}(S)$):**

La varianza de S está dada por:

$$\text{Var}(S) = \frac{n}{\lambda^2}.$$

- 3) **Caso especial ($n = 1$):**

Cuando $n = 1$, la distribución Gamma coincide con la distribución exponencial. Es decir:

$$\text{Gamma}(1, \lambda) = \text{Exp}(\lambda).$$

- 3) Sea X variable aleatoria con distribución $N(\mu, \sigma^2)$. Sea (X_1, \dots, X_n) una m.a.s de X , estudiar la distribución en el muestreo de $S = \sum_{j=1}^n X_j$.

Dado que $X \sim N(\mu, \sigma^2)$, y que (X_1, \dots, X_n) es una muestra aleatoria simple (m.a.s) de X , las X_i son independientes e idénticamente distribuidas con $X_i \sim N(\mu, \sigma^2)$. Queremos analizar la distribución en el muestreo de $S = \sum_{j=1}^n X_j$.

Propiedades relevantes:

- 1) **Suma de variables normales independientes:** Si X_1, \dots, X_n son independientes y $X_i \sim N(\mu, \sigma^2)$, entonces la suma:

$$S = \sum_{j=1}^n X_j$$

sigue una distribución normal:

$$S \sim N(n\mu, n\sigma^2).$$

Derivación:

- 1) **Esperanza ($E[S]$):** La esperanza de S es la suma de las esperanzas de las X_i :

$$E[S] = E\left[\sum_{j=1}^n X_j\right] = \sum_{j=1}^n E[X_j] = \sum_{j=1}^n \mu = n\mu.$$

- 2) **Varianza ($\text{Var}(S)$):** La varianza de S es la suma de las varianzas de las X_i , ya que son independientes:

$$\text{Var}(S) = \text{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \text{Var}(X_j) = \sum_{j=1}^n \sigma^2 = n\sigma^2.$$

- 3) **Distribución:** Dado que una combinación lineal de variables normales independientes también sigue una distribución normal, se concluye que:

$$S \sim N(n\mu, n\sigma^2).$$

- 4) Sea X una variable aleatoria con función de densidad:

$$f(x, \theta) = \frac{2x}{\theta} \exp\left(-\frac{x^2}{\theta}\right) \chi_{(0, +\infty)}(x).$$

Obtener la distribución en el muestreo estadístico:

$$T(X_1, X_2, \dots, X_n) = \sum_{j=1}^n X_j^2.$$

Obtener su media y su varianza.

Paso 1: Verificar la distribución de X

La función de densidad de X es:

$$f(x; \theta) = \frac{2x}{\theta} \exp\left(-\frac{x^2}{\theta}\right) \chi_{(0, \infty)}(x).$$

Esta densidad corresponde a una **distribución Rayleigh generalizada** con parámetro de escala θ . Para esta distribución:

- X^2 sigue una distribución exponencial con parámetro $\lambda = \frac{1}{\theta}$.

Entonces:

$$Y = X^2 \sim \text{Exp}\left(\lambda = \frac{1}{\theta}\right).$$

Paso 2: Distribución del estadístico $T = \sum_{j=1}^n X_j^2$

Dado que $Y_j = X_j^2 \sim \text{Exp}\left(\frac{1}{\theta}\right)$, y las Y_j son independientes, la suma de n variables exponenciales independientes sigue una distribución **Gamma**.

Por lo tanto, el estadístico:

$$T = \sum_{j=1}^n X_j^2 = \sum_{j=1}^n Y_j$$

sigue la distribución:

$$T \sim \text{Gamma}\left(n, \lambda = \frac{1}{\theta}\right),$$

donde:

- n es el parámetro de forma.
- $\lambda = \frac{1}{\theta}$ es el parámetro de escala.

La densidad de la distribución Gamma es:

$$f_T(t; n, \lambda) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{\Gamma(n)}, \quad t > 0.$$

Paso 3: Esperanza y Varianza del estadístico T

Para una distribución Gamma con parámetros (n, λ) , las propiedades son:

1) Esperanza:

$$E[T] = \frac{n}{\lambda}.$$

2) Varianza:

$$\text{Var}(T) = \frac{n}{\lambda^2}.$$

En este caso, como $\lambda = \frac{1}{\theta}$:

$$E[T] = \frac{n}{\frac{1}{\theta}} = n\theta,$$

$$\text{Var}(T) = \frac{n}{\left(\frac{1}{\theta}\right)^2} = n\theta^2.$$

5) Sea X una variable aleatoria con función de densidad:

$$f(x, \theta) = \frac{\theta}{(1+x)^{1+\theta}} \chi_{(0,+\infty)}(x).$$

Obtener la distribución en el muestreo del estadístico:

$$T(X_1, X_2, \dots, X_n) = \frac{\sum_{j=1}^n \ln(1 + X_j)}{n}.$$

Obtener su media y su varianza.

Paso 1: Verificar la distribución de X

La función de densidad de probabilidad X es:

$$f(x; \theta) = \frac{\theta}{(1+x)^{1+\theta}} \chi_{(0, \infty)}(x).$$

Esta densidad corresponde a la **distribución de Pareto** con parámetro de forma $\theta > 0$ y parámetro de escala igual a:

- 6) Sea X una variable aleatoria con función de densidad en todo \mathbb{R} :

$$f(x, \theta) = \exp(-(x - \theta)) \exp(-\exp(-(x - \theta))).$$

Obtener la distribución en el muestreo del estadístico:

$$T(X_1, X_2, \dots, X_n) = \frac{\sum_{j=1}^n \exp(-X_j)}{n}.$$

Obtener su media y su varianza.

- 7) Sea X_1, X_2, \dots, X_n una m.a.s de una variable X con distribución $N(\mu, \sigma^2)$. Sean \bar{X} y S^2 su media y cuasi-varianzas muestrales, respectivamente. Sea X_{n+1} una nueva observación de X independiente de X_1, X_2, \dots, X_n . Obtener la distribución en el muestreo del estadístico:

$$\frac{X_{n+1} - \bar{X}}{S} \sqrt{\frac{n}{n+1}}.$$

- 8) Sean X_1, X_2, \dots, X_{n_1} e Y_1, Y_2, \dots, Y_{n_2} muestras aleatorias simples independientes de dos poblaciones $X \sim N(\mu_1, \sigma^2)$ e $Y \sim N(\mu_2, \sigma^2)$, respectivamente. Obtener la distribución en el muestreo estadístico:

$$\frac{\alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}}},$$

siendo α y β dos números reales fijos.

- 9) Una empresa de agua produce botellas que deberían contener 300ml pero que presentan en la práctica una variabilidad modelada por una distribución Normal con media $\mu = 298$ ml y desviación típica $\sigma = 3$ ml.
- a) ¿Cuál es la probabilidad de que una botella elegida al azar de la producción contenga menos de 295ml?
 - b) ¿Cuál es la probabilidad de que el contenido medio de un paquete de seis botellas sea inferior a 295ml?
- 10) Se realiza una medición de peso en un laboratorio, sabiendo que la desviación típica de las mediciones es $\sigma = 10$ mg. La medición se repite 3 veces, se calcula la media \bar{x} , y este es el resultado proporcionado como estimación del peso.
- a) ¿Cuál es la desviación típica del resultado calculado?
 - b) ¿Cuántas veces se debe repetir la medición para que la desviación típica del valor medio se reduzca a 5?
- 11) El resultado de una encuesta fue que el 59% de la población española opina que el contexto económico es bueno o muy bueno. Supongamos que, extrapolando al conjunto de la población, efectivamente la proporción de todos los españoles que piensan que la situación es buena o muy buena es del 0.59.
- a) Sabemos que las encuestas incluyen márgenes de error que son aproximadamente ± 3 puntos. ¿Cuál es la probabilidad de que una muestra aleatoria de 300 españoles presente una proporción muestral que caiga dentro del intervalo 0.59 ± 0.03 ?
 - b) Contesta a la pregunta anterior en el caso en que la muestra consta de 600 personas y cuando la muestra consta de 1200 personas. ¿Cuál es el efecto de aumentar el tamaño de la muestra?
- 12) Un aparato de medición es exacto (el valor proporcionado medio es el valor auténtico de la señal) y la desviación típica del valor medido es 0.1 unidades. La distribución del valor medido es aproximadamente normal. ¿Cuál es la probabilidad de que el valor de una medición se aleje de la señal auténtica en más de 0.1 unidades? ¿Y si se repite la medición 5 veces y se toma la media de los 5 valores obtenidos?
- 13) En condiciones normales, una máquina produce piezas con una tasa de defectuosas del 1%. Para controlar que la máquina sigue bien ajustada, se escogen al azar cada día 100 piezas en la producción y se somete a un test. ¿Cuál es la probabilidad de que, si la máquina está bien ajustada, haya, en una de esas muestras, más del 2% de piezas defectuosas?

Si un día, 3 piezas resultan defectuosas, ¿cuáles son las conclusiones que sacaríamos sobre el funcionamiento de la máquina?