

Muestras aleatorias y distribuciones de muestreo

Definición 1.1 Si las variables aleatorias X_1, X_2, \dots, X_n tiene la misma función (densidad) de probabilidad que la de la distribución de la población y su función (distribución) conjunta de probabilidad es igual al producto de las marginales, entonces X_1, X_2, \dots, X_n forman un conjunto de n variables aleatorias independientes e idénticamente distribuidas (IID) que constituyen una muestra aleatoria de la población.

En el contexto de la definición 7.1, la función (densidad) conjunta de probabilidad de X_1, x_2, \dots, X_n es la función de verosimilitud de la muestra dada por

$$L(x; \theta) = \prod_{i=1}^n f(x_i; \theta),$$

en donde $x = \{x_1, x_2, \dots, x_n\}$ denota los datos muestreados. Cuando las realizaciones x se conocen, $L(x; \theta)$ es una función del parámetro desconocido θ .

Ejemplo 1.1 Se ilustrará el concepto de muestra aleatoria dada en la definición 7.1 mediante lo siguiente: Sea X_1, X_2, \dots, X_n una muestra aleatoria de n variables aleatorias IID de una población cuya distribución de probabilidad es exponencial con densidad

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}, \quad 0 < x < \infty.$$

Cuando se observa X_1 y se registra su realización x_1 ,

$$f(x_1; \theta) = \frac{1}{\theta} e^{-x_1/\theta}, \quad 0 < x_1 < \infty.$$

Ahora se observa X_2 y se registra su realización x_2 . Dado que X_1 y X_2 son estadísticamente independientes y tienen las mismas densidades marginales,

$$f(x_2|x_1) = f(x_2; \theta) = \frac{1}{\theta} e^{-x_2/\theta}, \quad 0 < x_2 < \infty.$$

La función de densidad conjunta de X_1 y X_2 es

$$f(x_1, x_2; \theta) = f(x_1; \theta) f(x_2; \theta) = \frac{1}{\theta^2} e^{-(x_1+x_2)/\theta}, \quad 0 < x_i < \infty, i = 1, 2.$$

Por lo tanto, se desprende que para una muestra aleatoria de tamaño n

$$L(x_1, x_2, \dots, x_n; \theta) = \frac{1}{\theta} e^{-(x_1 + x_2 + \dots + x_n)/\theta}, \quad 0 < x_i < \infty, i = 1, 2, \dots, n.$$

■

1.3 Distribuciones de muestre de estadísticas

Definición 1.2 Un parámetro es una caracterización numérica de las distribuciones de la población de manera que describe, parcial o completamente, la función de densidad de probabilidad de la característica de interés. Por ejemplo, cuando se especifica el valor del parámetro de escala exponencial θ , se describe de manera completa la función de probabilidad

$$f(x; \theta) = \frac{1}{\theta} e^{(-x/\theta)}.$$

La oración describe de manera completa, sugiere que una vez que se conoce el valor de θ entonces, puede formularse cualquier proposición probabilística de interés.

Dado que los parámetros son prácticamente inherentes a todos los modelos de probabilidad, es imposible calcular las probabilidades deseadas sin un conocimiento del valor de estos. Es por esta razón que la noción de una estadística y su distribución de muestre es muy importante en inferencia estadística. Antes de dar la definición de una estadística, debe notarse que desde un punto de vista clásico (no bayesiano), un parámetro se considera como una constante fija cuyo valor se desconoce. Desde una perspectiva bayesiana un parámetro siempre es una variable aleatoria con algún tipo de distribución de probabilidad.

Definición 1.3 Una estadística es cualquier función de las variables aleatorias que se observaron en la muestra de manera que esta función no contiene cantidades desconocidas.

De manera general, denótese una estadística por $T = u(X)$. Dado que T es una función de variables aleatorias, es en sí misma una variable aleatoria, y su valor específico $t = u(x)$ puede determinarse cuando se conozcan las realizaciones x de X . Si se emplea una estadística T para estimar un parámetro desconocido θ , entonces T recibe el nombre de **estimador** de θ , y el valor específico de t como un resultado de los datos muestrales recibe el nombre de **estimación** de θ . Esto es, un estimador es una estadística que identifica al mecanismo funcional por medio del cual, una vez que las observaciones en la muestra se realizan, se obtiene una estimación.

Un parámetro es una constante pero una estadística es una variable aleatoria. Además, un valor del parámetro descrito describe de manera completa un modelo de probabilidad; ningún valor de la estadística puede desempeñar tal papel si cada uno de estos depende del valor de las observaciones de las muestras.

Definición 1.4 La distribución de muestre de una estadística T es la distribución de probabilidad de T que puede obtenerse como resultado de un número infinito de muestras aleatorias independientes, cada una de tamaño n , provenientes de la población de interés.

La distribución de muestra de una estadística hace posible este tipo de análisis de probabilidad, esencial para valorar el riesgo inherente cuando se formulan inferencias.

Teorema 1.1 Sea X_1, X_2, \dots, X_n un conjunto de n variables aleatorias independientes cada una con función generadora de momentos $m_{X_1}(t), m_{X_2}(t), \dots, m_{X_n}(t)$. Si

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n,$$

en donde a_1, a_2, \dots, a_n son constantes, entonces:

$$m_Y(t) = m_{X_1}(a_1 t) m_{X_2}(a_2 t) \dots m_{X_n}(a_n t).$$

Demostración.- Mediante el empleo de la definición y la hipótesis de independencia, se tiene

$$\begin{aligned} m_Y(t) &= E \left\{ e^{[t(a_1 X_1 + a_2 X_2 + \dots + a_n X_n)]} \right\} \\ &= E \left\{ e^{[t(a_1 X_1)]} e^{[t(a_2 X_2)]} \dots e^{[t(a_n X_n)]} \right\} \\ &= E \left\{ e^{[t(a_1 X_1)]} \right\} E \left\{ e^{[t(a_2 X_2)]} \right\} \dots E \left\{ e^{[t(a_n X_n)]} \right\} \\ &= m_{X_1}(a_1 t) m_{X_2}(a_2 t) \dots m_{X_n}(a_n t). \end{aligned}$$

■

De esta forma, la función generadora de momentos de una combinación lineal de n variables aleatorias independientes es el producto de las correspondientes funciones generadoras de momentos con argumentos iguales a las constantes de tiempo t .

Teorema 1.2 Sea X_1, X_2, \dots, X_n un conjunto de variables aleatorias normalmente distribuidas con medias $E(X_i) = \mu_i$ y varianzas $Var(X_i) = \sigma_i^2$ para $i = 1, 2, \dots, n$. Si

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n,$$

en donde a_1, a_2, \dots, a_n son constantes, entonces Y es una variable aleatoria con distribución normal y media

$$E(Y) = a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n$$

y con varianza

$$Var(Y) = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2.$$

Demostración.- Dado que X_i se encuentra normalmente distribuida, su función generadora de momentos es

$$m_{X_i}(t) = e^{\left(\mu_i t + \frac{\sigma_i^2 t^2}{2} \right)}.$$

De acuerdo con el teorema 7.1, la función generadora de momentos de Y es

$$\begin{aligned} m_Y(t) &= m_{X_1}(a_1 t) m_{X_2}(a_2 t) \dots m_{X_n}(a_n t) \\ &= e^{\left(\mu_1 a_1 t + \frac{a_1^2 \sigma_1^2 t^2}{2} \right)} e^{\left(\mu_2 a_2 t + \frac{a_2^2 \sigma_2^2 t^2}{2} \right)} \dots e^{\left(\mu_n a_n t + \frac{a_n^2 \sigma_n^2 t^2}{2} \right)} \\ &= e^{\left[t \sum_{i=1}^n a_i \mu_i + \frac{t^2 \sum_{i=1}^n a_i^2 \sigma_i^2}{2} \right]}. \end{aligned}$$

Por lo tanto, Y se encuentra normalmente distribuida con media $\sum_{i=1}^n a_i \mu_i$ y varianza $\sum_{i=1}^n a_i^2 \sigma_i^2$. ■

Del teorema 7.2 se desprende que si $a_i = 1$ para $i = 1, 2, \dots, n$, entonces la suma de variables aleatorias independientes normalmente distribuidas también posee una distribución normal con media y varianza igual a la suma de las medias y las varianzas de cada una de las variables aleatorias. El resultado anterior se conoce como la propiedad aditiva de la distribución normal. Debe notarse que la hipótesis de normalidad no es necesaria para obtener las fórmulas de la media y la varianza de Y en el teorema 7.2. De hecho, con base en el teorema 6.1, si X_1, X_2, \dots, X_n es un conjunto de n variables aleatorias IID con medias $E(X_i) = \mu_i$ y varianzas $Var(X_i) = \sigma_i^2$, $i = 1, 2, \dots, n$ entonces para $Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$,

$$E(Y) = \sum_{i=1}^n a_i \mu_i$$

y

$$Var(Y) = \sum_{i=1}^n a_i^2 \sigma_i^2.$$

en donde a_1, a_2, \dots, a_n son constantes.

1.4 La distribución de muestre de \bar{X}

Sea X_1, X_2, \dots, X_n una muestra aleatoria que consiste en n variables aleatorias IID tales que $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ para toda $i = 1, 2, \dots, n$. Entonces la estadística

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

se define como la media de las n variables aleatorias IID o, sencillamente, media muestral. Nótese que una vez que se conocen las realizaciones x_1, x_2, \dots, x_n de X_1, X_2, \dots, X_n , respectivamente, realización \bar{x} de \bar{X} se obtiene promediando los datos muestrales.

Si en, $E(Y) = \sum_{i=1}^n a_i \mu_i$ y $Var(Y) = \sum_{i=1}^n a_i^2 \sigma_i^2$, $a_i = 1/n$, $i = 1, 2, \dots, n$ entonces el valor esperado y la varianza de \bar{X} son

$$E(\bar{X}) = \sum_{i=1}^n \frac{1}{n} \mu = n \left(\frac{\mu}{n} \right) = \mu.$$

y

$$Var(\bar{X}) = \sum_{i=1}^n \frac{1}{n^2} \sigma^2 = n \left(\frac{\sigma^2}{n^2} \right) = \frac{\sigma^2}{n},$$

respectivamente, en donde μ y σ^2 son la media y la varianza de la distribución de la población a partir de la cual se obtuvo la muestra. Luego de esta última ecuación de $Var(\bar{X})$ la desviación estándar de \bar{X} es

$$d.e.(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

la cual, en algunas ocasiones recibe el nombre de **error estándar de la media**. Si el tamaño de la muestra crece, la precisión de la media muestral para estimar la media poblacional aumenta.

Teorema 1.3 Sea X_1, X_2, \dots, X_n una muestra aleatoria que consiste de n variables aleatorias independientes y normalmente distribuidas con media $E(X_i) = \mu$ y varianzas $Var(X_i) = \sigma^2$, $i = 1, 2, \dots, n$. Entonces la distribución de la media muestral \bar{X} es normal con media μ y varianza σ^2/n .

Demostración.- Este teorema es un corolario del teorema 7.2. Esto es, sea $a_i = 1/n$; dado que las medias

y las varianzas son iguales, respectivamente, la función generadora de momentos de \bar{X} es

$$\begin{aligned} m_{\bar{X}}(t) &= e^{\left(t \sum_{i=1}^n \frac{1}{n} \mu + \frac{t^2 \sum_{i=1}^n \frac{1}{n^2} \sigma^2}{2} \right)} \\ &= e^{\left(t\mu + \frac{t^2 \sigma^2}{2n} \right)}, \end{aligned}$$

que es la función generadora de momentos de una variable aleatoria normalmente distribuida con media μ y varianza σ^2/n . De esta forma, la **función de densidad de probabilidad de \bar{X} cuando se muestrea una población cuya distribución es normal**, está dada por

$$f(\bar{x}, \mu, \sigma/\sqrt{n}) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{\left[-\frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right]}, \quad -\sigma < \bar{x} < \sigma.$$

■

Ejemplo 1.2 Demostrar que si X_1, X_2, \dots, X_n son n variables aleatorias independientes exponencialmente distribuidas con función de densidad de probabilidad

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta} \quad x > 0,$$

entre \bar{X} tiene una distribución gama.

Demostración.-

■

Teorema 1.4 Sean X_1, X_2, \dots, X_n variables aleatorias IID con una distribución de probabilidad no especificada y que tiene una media μ y varianza σ^2 finita. El promedio muestral $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$ tiene una distribución con media μ y varianza σ^2/n que tiende hacia una distribución normal conforme n tiende a ∞ . En otras palabras, la variable aleatoria $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ tiene como límite una distribución normal estándar.

Demostración.-

■