Probabilidad y estadística

MAT 041, Primer semestre

Francisco Cuevas Pacheco

14 de noviembre de 2022

Recuerdo

Definición

*Una **estadística muestral** T es cualquier función de las variables que forman la muestra aleaoria. Se anota, $T = T(\underline{X}) = T(X_1, X_2, \ldots, X_n)$.*

Algunas estadísticas muestrales son:

Total muestral	$t = \sum_{i=1}^{n} X_i$
Media muestral	$\bar{X} = \sum_{i=1}^{n} X_i/n$
Varianza muestral	$S_{n-1}^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$
Máximo muestral	$X_{(n)} = \text{máx}\{X_1, X_2, ..., X_n\}$
Mínimo muestral	$X_{(1)} = \text{mín}\{X_1, X_2, ..., X_n\}$

Definición

*Una **estadística muestral** T es cualquier función de las variables que forman la muestra aleaoria. Se anota, $T = T(\underline{X}) = T(X_1, X_2, \ldots, X_n)$.*

Algunas estadísticas muestrales son:

Total muestral	$t = \sum_{i=1}^{n} X_i$
Media muestral	$\bar{X} = \sum_{i=1}^{n} X_i/n$
Varianza muestral	$S^2_{n-1} = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$
Máximo muestral	$X_{(n)} = \text{máx}\{X_1, X_2, ..., X_n\}$
Mínimo muestral	$X_{(1)} = \text{mín}\{X_1, X_2, ..., X_n\}$

Observación: Al ser $T(\underline{X})$ una función de las variables $X_1, \ldots, X_n$ entonces T es una variable aleatoria y por lo tanto tiene una densiad asociada. Cuando se reemplazan los datos $\underline{x} = (x_1, x_2, \ldots, x_n)$ en $T(\underline{X})$ se obtiene una **realización** de esa estadística que se anota $T = T(\underline{x})$.

Distribuciones muestrales

Como vimos anteriormente, para evaluar las propiedades del estimador de θ, $\hat{\theta}$, es necesario calcular el valor esperado y su varianza.

Para esto, es necesario estudiar la distribución de algunas cantidades estadísticas que se encuentran con frecuencia en los problemas de la inferencia estadística.

Ejemplos

1. Sea $X_1, \dots, X_n$ una muestra independiente e identicamente distribuida con distribución $U[0, \theta]$. Se puede demostrar (tarea) que $\hat{\theta} = \text{máx}\{X_1, \dots, X_n\}$.
2. Sea $X_1, \dots, X_n$ una muestra independiente e identicamente distribuida con distribución $N(\mu, \sigma^2)$. Se sabe que el estimador máximo verosimil para $\hat{\mu} = \bar{X}$ y $\hat{\sigma^2} = S^2_{n-1}$.

Ejemplos

1. Sea $X_1, \ldots, X_n$ una muestra independiente e identicamente distribuida con distribución $U[0, \theta]$. Se puede demostrar (tarea) que $\hat{\theta} = \text{máx}\{X_1, \ldots, X_n\}$.
2. Sea $X_1, \ldots, X_n$ una muestra independiente e identicamente distribuida con distribución $N(\mu, \sigma^2)$. Se sabe que el estimador máximo verosimil para $\hat{\mu} = \bar{X}$ y $\hat{\sigma^2} = S_{n-1}^2$.
    1) ¿Cuál es el valor esperado de $\hat{\sigma^2}$?

Ejemplos

1. Sea $X_1, \ldots, X_n$ una muestra independiente e identicamente distribuida con distribución $U[0, \theta]$. Se puede demostrar (tarea) que $\hat{\theta} = \text{máx}\{X_1, \ldots, X_n\}$.
2. Sea $X_1, \ldots, X_n$ una muestra independiente e identicamente distribuida con distribución $N(\mu, \sigma^2)$. Se sabe que el estimador máximo verosimil para $\hat{\mu} = \bar{X}$ y $\hat{\sigma^2} = S^2_{n-1}$.
    1) ¿Cuál es el valor esperado de $\hat{\sigma^2}$?
    2) ¿Cuál es la distribución de $\frac{\hat{\mu}}{\hat{\sigma^2}}$?

Para responder estas preguntas, presentaremos algunas transformaciones útiles

Distribución Chi Cuadrado

Se dice que X sigue una distribución de Chi Cuadrado con k grados de libertad si:

$$f_X(y) = \frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} y^{\frac{k}{2}-1} e^{-\frac{y}{2}}$$

La función generadora de momentos está dada por

$$\psi_X(t) = (1-2t)^{-k/2}, \quad \text{si } t < 1/2$$

Esta propiedad la denotamos por $X \sim \chi^2(k)$

Propiedad 1

Distribución normal, recuerdo

Sea $X_1, \dots, X_n$ una muestra aleatoria $N(\mu, \sigma^2)$. Entonces se tiene que

$$\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \sim N(0,1)$$

Propiedad 1

Distribución normal, recuerdo

Sea $X_1, \ldots, X_n$ una muestra aleatoria $N(\mu, \sigma^2)$. Entonces se tiene que

$$\sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) \sim N(0,1)$$

Demostración: usando la Función generadora de momentos.

Forma Cuadrática

Sea $Y \sim N(0,1)$. Entonces $Y^2 \sim \chi^2(1)$, es decir,

$$f_{Y^2}(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}$$

Demostración

Denotemos por $\Phi(\cdot)$ a la función de distribución de una variable aleatoria normal estandar. La demostración se basa en usar el teorema de transformación, separando en dos casos:

1. Caso 1: $y < 0$. Luego $\mathbb{P}(Y^2 \leq y) = 0$
2. Caso 2: $y \geq 0$. Entonces se tiene que

$$\begin{aligned} F_{Y^2}(y) = \mathbb{P}(Y^2 \leq y) & \\ &= \mathbb{P}(-y \leq |Y| \leq y) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \\ &= \Phi(\sqrt{y}) - (1 - \Phi(\sqrt{y})) \\ &= 2\Phi(\sqrt{y}) - 1. \end{aligned}$$

Por otro lado, sabemos que $\frac{dF(y)}{dy} = f(y)$. Luego

$$\frac{dF(y)}{dy} = 2\frac{d}{dy}\int_{-\infty}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}\, ds = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}$$

Propiedad 2

Propiedad: Sumas de Chi cuadrado

Sea $X_1, \dots, X_n$ una muestra aleatoria independiente proveniente de $N(0,1)$.
Entonces $Z = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$.

Propiedad 2

Propiedad: Sumas de Chi cuadrado

Sea $X_1, \ldots, X_n$ una muestra aleatoria independiente proveniente de $N(0,1)$. Entonces $Z = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$.

Demostración:

Procederemos mediante la función generadora de momentos. De hecho,

$$\psi_Z(t) = \psi_{\sum_{i=1}^{n} X_i^2}(t) = \prod_{i=1}^{n} \psi_{X_i^2}(t) = \prod_{i=1}^{n} (1-2t)^{-1/2} = (1-2t)^{-n/2}.$$

Propiedad

Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución $N(\mu, \sigma^2)$, y sea $S^2_{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$. Se tiene entonces que

Forma Cuadrática

$$D = \frac{(n-1)S^2_{n-1}}{\sigma} \sim \chi^2(n-1).$$

Demostración

Demostración:

Consideramos el resultado anterior:

$$Q = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

Demostración

Demostración:

Consideramos el resultado anterior:

$$Q = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

Y consideramos la siguiente expansion

$$\begin{aligned}
Q &= \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \\
&= \sum_{i=1}^{n} \left(\frac{X_i - \bar{X} + \bar{X} - \mu}{\sigma} \right)^2 \\
&= \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 + \sum_{i=1}^{n} \left(\frac{\bar{X} - \mu}{\sigma} \right)^2 + 2 \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right) \left(\frac{\bar{X} - \mu}{\sigma} \right) \\
&= \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 + \sum_{i=1}^{n} \left(\frac{\bar{X} - \mu}{\sigma} \right)^2 \\
&= D + L
\end{aligned}$$

Demostracion (cont.)

Notamos entonces que Q contiene la expresión que estamos buscando. Para encontrar la distribución de D, recordamos que

$$L = n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 \sim \chi^2(1).$$

Intuitivamente, tenemos la siguiente igualdad

Intuición[1]

$$\chi^2(n) = \chi^2(n-1) + \chi^2(1).$$

[1] Estamos abusando de notación, ya que se están sumando cantidades que poseen las distribuciones que son parte de la suma.

Demotración (Final.)

Para finalizar la demostración, se procede mediante la función generadora de momentos, nuevamente:

$$\psi_Q(t) = \psi_D(t)\psi_L(t)$$
$$\psi_D(t) = \psi_Q(t)/\psi_L(t)$$
$$\psi_D(t) = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}}$$
$$\psi_D(t) = (1-2t)^{-(n-1)/2}$$

Finalizando la demostración

Demotración (Final.)

Para finalizar la demostración, se procede mediante la función generadora de momentos, nuevamente:

$$\psi_Q(t) = \psi_D(t)\psi_L(t)$$
$$\psi_D(t) = \psi_Q(t)/\psi_L(t)$$
$$\psi_D(t) = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}}$$
$$\psi_D(t) = (1-2t)^{-(n-1)/2}$$

Finalizando la demostración

Encuentre el detalle que hace que esto deba ser realizado con más cuidado.

Definición

Definición: t-student

Sea Z una variable aleatoria con distribución normal estandar y sea V una variable aleatoria con distribución Chi Cuadrado con n grados de libertad.
Se dice que X sigue una distribución de t-student, con n grados de libertad, si $X = \frac{Z}{\sqrt{V/n}}$.

La densidad de probabilidad está dada por

$$f_X(x|n) = \frac{\Gamma(n/2 + 1/2)}{\Gamma(n/2)\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \qquad x \in \mathbb{R}.$$

Definición

Definición: F de Snedecor

Sean X e Y dos variables aleatorias independientes con distribución Chi cuadrádo con n_1 y n_2 grados de libertad respectivamente.
Se dice que Z sigue una distribución F de Snedecor con n_1 y n_2 grados de libertad, si

$$Z = \frac{X/n_1}{Y/n_2}$$

La densidad de probabilidad está dada por

$$f_Z(z|n_1, n_2) = \frac{\Gamma\left(\frac{(n_1+n_2)}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{-\frac{n_1}{2}} z^{\frac{n_1}{2}-1} \left(1 + \frac{n_1}{n_2} z\right)^{-\frac{(n_1+n_2)}{2}}, \qquad z \in \mathbb{R}.$$

Distribución de la media

Recordamos el siguiente resultado

Recuerdo:

$$\sqrt{n}\frac{(\bar{X}-\mu)}{\sigma} \sim N(0,1)$$

En la práctica, σ es desconocido y se suele estimar por S_n. Luego:

Distribución de la media

Recordamos el siguiente resultado

Recuerdo:

$$\sqrt{n}\frac{(\bar{X}-\mu)}{\sigma} \sim N(0,1)$$

En la práctica, σ es desconocido y se suele estimar por S_n. Luego:

Resultado:

$$\sqrt{n}\frac{(\bar{X}-\mu)}{S_n} \sim t(n-1)$$

Distribución de la media

Recordamos el siguiente resultado

Recuerdo:

$$\sqrt{n}\frac{(\bar{X}-\mu)}{\sigma} \sim N(0,1)$$

En la práctica, σ es desconocido y se suele estimar por S_n. Luego:

Resultado:

$$\sqrt{n}\frac{(\bar{X}-\mu)}{S_n} \sim t(n-1)$$

La demostración es directa, ya que:

$$\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \sim N(0,1) \qquad \text{y} \qquad \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

Luego, se cumple la definición.

Distribución del ratio de varianzas

Considere dos muestras aleatorias $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$, provenientes de una distribución normal de media 0 y varianzas σ_X^2 y σ_Y^2 respectivamente.

Resultado

El ratio

$$\frac{S_X^2}{S_Y^2} \sim F(n-1, m-1)$$

Distribución del ratio de varianzas

Considere dos muestras aleatorias $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$, provenientes de una distribución normal de media 0 y varianzas σ_X^2 y σ_Y^2 respectivamente.

Resultado

El ratio

$$\frac{S_X^2}{S_Y^2} \sim F(n-1, m-1)$$

La demostración es directa, y se deja al lector

Distribución del máximo y mínimo

Distribución del máximo

Sean $X_1, \ldots, X_n$ variables aleatorias independientes con función de distribución F_X. Sea $M = \max_k X_k$, el valor máximo. Luego se tiene que

$$F_M(x) = F_X^n(x)$$

Distribución del máximo

Sean $X_1, \dots, X_n$ variables aleatorias independientes con función de distribución F_X. Sea $M = \max_k X_k$, el valor máximo. Luego se tiene que

$$F_M(x) = F_X^n(x)$$

Demostración:
Procedemos por definición. De hecho,

$$\begin{aligned}
F_M(x) &= \mathbb{P}(\max_k X_k \leq x) \\
&= \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) \\
&= \mathbb{P}(X_1 \leq x) \dots \mathbb{P}(X_n \leq x) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) \\
&= \prod_{i=1}^{n} F_{X_i}(x) \\
&= F_X(x)^n
\end{aligned}$$

Distribución del máximo y mínimo

Distribución del máximo

Sean $X_1, \ldots, X_n$ variables aleatorias independientes con función de distribución F_X. Sea $M = \text{mín}_k X_k$, el valor mínimo. Luego se tiene que

$$F_M(x) = (1 - F_X(x))^n$$

Distribución del máximo

Sean $X_1, \ldots, X_n$ variables aleatorias independientes con función de distribución F_X. Sea $M = \min_k X_k$, el valor mínimo. Luego se tiene que

$$F_M(x) = (1 - F_X(x))^n$$

Demostración:
Procedemos por definición. De hecho,

$$\begin{aligned}
1 - F_M(x) &= \mathbb{P}(\min_k X_k \geq x) \\
&= \mathbb{P}(X_1 \geq x, \ldots, X_n \geq x) \\
&= \mathbb{P}(X_1 \geq x) \ldots \mathbb{P}(X_n \geq x) \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \geq x) \\
&= \prod_{i=1}^{n} (1 - F_{X_i}(x)) \\
&= (1 - F_X(x))^n
\end{aligned}$$

Distribución de los estadísticos de orden

Distribución del estadístico de orden

Sean $X_1, \ldots, X_n$ variables aleatorias independientes con función de distribución F_X. Sea $X_{(k)}$ el k-ésimo dato. Luego se tiene que

$$F_{X_{(k)}}(x) = \sum_{j=r}^{n} \binom{n}{j} F_X^j (1 - F_X(x))^{n-j}$$

Distribución de los estadísticos de orden

Distribución del estadístico de orden

Sean $X_1, \ldots, X_n$ variables aleatorias independientes con función de distribución F_X. Sea $X_{(k)}$ el k-ésimo dato. Luego se tiene que

$$F_{X_{(k)}}(x) = \sum_{j=r}^{n} \binom{n}{j} F_X^j (1 - F_X(x))^{n-j}$$

Demostración: Ejercicio para el lector.