Índice

3.3. Estadístico de Student. Distribución t de Student

Saber que $\bar{X}$ tiene distribución en el muestreo $N(\mu, \sigma/\sqrt{n})$, o equivalentemente que

$$\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma} \quad \text{tiene distribución } N(0,1)$$

puede resultar de poca utilidad si la varianza poblacional σ^2 es un parámetro desconocido de la distribución teórica, ya que no se puede, entonces, usar esta conclusión para hacer previsiones acerca de la diferencia $\bar{X}-\mu$. En tal caso, cabe pensar que el resultado no será muy distinto si se sustituye σ por el valor de la desviación típica muestral, s, puesto que, al menos para muestras grandes, σ^2 y s^2 tendrán valores similares. Tal idea llevó a Student a considerar el estadístico:

$$t = \sqrt{n-1}\,\frac{\bar{X}-\mu}{s}\,.$$

La sustitución de $\sqrt{n}$ por $\sqrt{n-1}$ quedará explicada un poco más adelante, aunque de momento, si se quiere reforzar la analogía, se puede expresar

$$t = \sqrt{n}\,\frac{\bar{X}-\mu}{S}\,,$$

donde S^2 es la cuasivarianza muestral ($S^2 = ns^2/(n-1)$) que, a efectos de dar una aproximación de σ^2, tiene tantas ventajas, si no más, que s^2.

En cualquier caso, para poder usar el estadístico t en sustitución de $\sqrt{n}(\bar{X}-\mu)/\sigma$, la cuestión se centra en determinar su distribución en el muestreo (confiando en que el resultado no dependa de σ^2, pues, de lo contrario, no se habría adelantado nada). Para ello, parece natural prescindir de μ y adoptar el siguiente planteamiento:

Sean $X,\ X_1, X_2, \ldots, X_n$, $n+1$ variables aleatorias independientes, con distribución $N(0,\sigma)$; sea $Y = \sum_{i=1}^n X_i^2$, y $U = \sqrt{Y/n}$. Se pretende hallar la distribución de $t = X/U$.

La distribución de Y/σ^2 es χ^2_n, así que la densidad de Y será

$$\frac{1}{2^{n/2}\Gamma(\frac{n}{2})\sigma^n}\, y^{\frac{n}{2}-1}e^{-y/(2\sigma^2)} \quad \text{para } y > 0,$$

con lo cual la densidad de U es

$$\frac{n^{n/2}}{2^{(n-2)/2}\Gamma(\frac{n}{2})\sigma^n}\, u^{n-1}e^{-nu^2/(2\sigma^2)} \quad \text{para } u > 0.$$

Como X y U son independientes, su densidad conjunta será

$$Ke^{-x^2/(2\sigma^2)}u^{n-1}e^{-nu^2/(2\sigma^2)} = Ku^{n-1}e^{-(x^2+nu^2)/(2\sigma^2)}$$

para $x \in I\!R$ y $u > 0$, siendo $K = \dfrac{1}{\sqrt{2\pi}} \dfrac{n^{n/2}}{2^{(n-2)/2}\Gamma(\frac{n}{2})\sigma^{n+1}}$.

El cambio bidimensional $z = U$, $t = X/U$, da como densidad de (z, t)

$$Kz^n e^{-z^2(t^2+n)/(2\sigma^2)} \qquad \text{para } t \in I\!R \text{ y } z > 0;$$

con lo cual la densidad marginal de t resulta ser

$$K\int_0^\infty z^n e^{-z^2(t^2+n)/(2\sigma^2)}\, dz = \frac{K}{2}\int_0^\infty v^{(n-1)/2} e^{-v(t^2+n)/(2\sigma^2)}\, dv$$

$$= \frac{K}{2}\left(\frac{t^2+n}{2\sigma^2}\right)^{-(n+1)/2} \Gamma\left(\frac{n+1}{2}\right) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}$$

para $t \in I\!R$ (después de sustituir la constante K y simplificar).

Nuevamente la situación merece un nombre propio:

> *Si X, $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes y con distribución $N(0, \sigma)$, la distribución de*
>
> $$\frac{X}{\sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2}},$$
>
> *de densidad*
>
> $$\frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} \qquad \textit{para } t \in I\!R,$$
>
> *se denomina* distribución t de Student con n grados de libertad.

Suele decirse que *la distribución t_n es la de una $N(0, 1)$ dividida por la raíz cuadrada de una χ_n^2 dividida por sus grados de libertad, ambas independientes entre sí*. Lo cual refleja la propiedad fundamental de la distribución de Student: su independencia de la varianza σ^2 de las variables que intervienen en su definición.

Antes de examinar otras propiedades de la distribución t, se puede ver que resuelve el problema que ha suscitado su definición.

Teorema de Student: *Si $(X_1, X_2, \ldots, X_n)$ es una muestra aleatoria simple de una población $N(\mu, \sigma)$, el estadístico de Student:*

$$\sqrt{n-1}\,\frac{\bar{X}-\mu}{s} = \sqrt{n}\,\frac{\bar{X}-\mu}{S}$$

tiene distribución t de Student con $n-1$ grados de libertad.

En efecto,

$$\sqrt{n-1}\,\frac{\bar{X}-\mu}{s} = \frac{\sqrt{n}(\bar{X}-\mu)/\sigma}{\sqrt{\frac{1}{n-1}\,ns^2/\sigma^2}}$$

y basta observar que $\sqrt{n}(\bar{X}-\mu)/\sigma$ tiene distribución $N(0,1)$, mientras que, según el teorema de Fisher, ns^2/σ^2 tiene la misma distribución que $\sum_{i=1}^{n-1} X_i'^2$, siendo $X_1', \ldots, X_{n-1}'$ variables $N(0,1)$, independientes entre sí y de la media muestral $\bar{X}$. Por tanto, $\sqrt{n-1}(\bar{X}-\mu)/s$ tiene distribución t_{n-1}.

Cuando σ^2 sea desconocida, el resultado permite juzgar que valores son previsibles para $\bar{X}-\mu$, utilizando S en lugar de σ y la distribución t_{n-1} en lugar de la $N(0,1)$. Véase el ejercicio 3.3.

Acerca de la distribución t de Student, cabe señalar que, al igual que la distribución χ^2, depende de un único parámetro n (el número de sumandos que intervienen en la suma del denominador) que nuevamente se denomina “grados de libertad” de la distribución. En cambio, la distribución de Student tiene como soporte el intervalo $(-\infty, \infty)$, siendo su densidad simétrica respecto al origen, de manera que su representación gráfica (para $n = 10$) es de la forma:

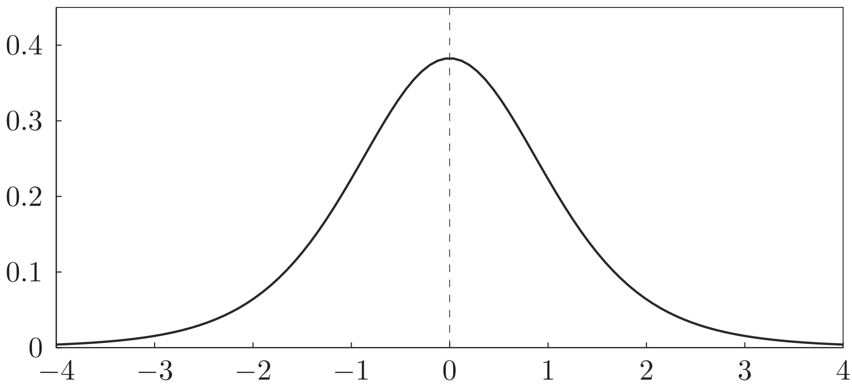

Figura 3.2: Densidad t de Student