Up
Next Previous
Nodo Raíz: 6.8 Distribuciones continuas
Siguiente: 6.8.10 Distribución de Student
Previo: 6.8.6 Distribución normal o gaussiana

Subsecciones

6.8.8 Distribución χ^2

Si consideramos una v.a. $Z \leadsto \mathbf{N}(0,1)$, la v.a. $X=Z^2$ se distribuye según una ley de probabilidad **distribución χ^2 con un grado de libertad**, lo que se representa como

$$X \leadsto \chi^2_1$$

Si tenemos n v.a. independientes $Z_i \leadsto \mathbf{N}(0,1)$, la suma de sus cuadrados respectivos es una distribución que denominaremos **ley de distribución χ^2 con n grados de libertad**, χ^2_n.

$$\boxed{\{Z_i\}_{i=1}^n \leadsto \mathbf{N}(0,1) \Longrightarrow \sum_{i=1}^n Z_i^2 \leadsto \chi^2_n}$$

La media y varianza de esta variable son respectivamente:

$$\begin{aligned} \mathbf{E}[X] &= n \\ \mathbf{Var}[X] &= 2n \end{aligned}$$

y su función de densidad es:

$$f_{\chi^2_n}(x) = \begin{cases} 0 & \text{si } x \in (-\infty, 0] \\ \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{si } x \in (0, \infty) \end{cases}$$

Los percentiles de esta distribución que aparecen con más frecuencia en la práctica los podemos encontrar en la tabla 5.

Figura: Función de densidad de χ^2_n para valores pequeños de *n*.

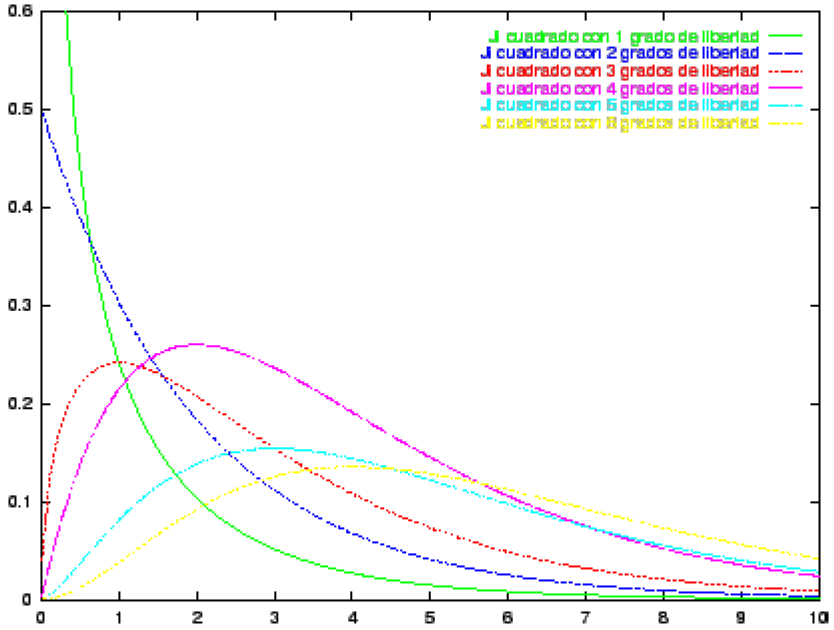
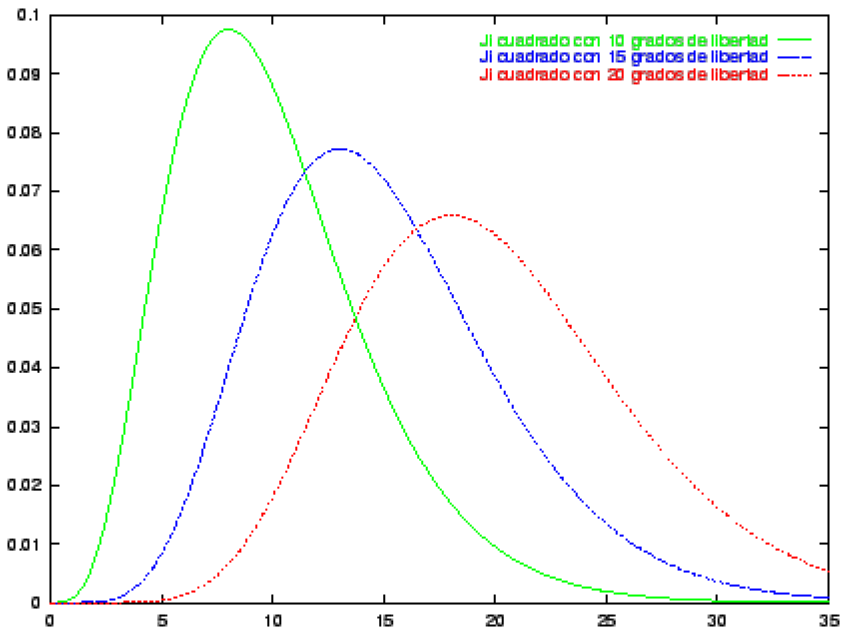


Figura: Función de densidad de χ^2_n para valores grandes de *n*.



En consecuencia, si tenemos $X_1, \ldots, X_n$, v.a. independientes, donde cada $X_i \leadsto \mathbf{N}\left(\mu_i, \sigma_i^2\right)$, se tiene

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2 \leadsto \chi^2_n$$

6.8.8.1 Observación

La ley de distribución χ^2 muestra su importancia cuando queremos determinar la variabilidad (sin signo) de cantidades que se distribuyen en torno a un valor central siguiendo un mecanismo normal. Como ilustración tenemos el siguiente ejemplo:

6.8.8.2 Ejemplo

Un instrumento para medir el nivel de glucemia en sangre, ofrece resultados bastantes aproximados con la realidad, aunque existe cierta cantidad de error ϵ que se distribuye de modo normal con media 0 y desviación típica $\sigma = 2$.

$$X_{\text{real}} = X_{\text{exp}} + \epsilon, \qquad \epsilon \leadsto \mathbf{N}\left(\mu = 0, \sigma^2 = 2^2\right)$$

Se realizan mediciones de los niveles de glucemia dados por el instrumento en un grupo de n=100 pacientes. Nos interesa medir la cantidad de error que se acumula en las mediciones de todos los pacientes. Podemos plantear varias estrategias para medir los errores acumulados. Entre ellas destacamos las siguientes:

1. Definimos el error acumulado en las mediciones de todos los pacientes como

$$E_1 = \sum_{i=1}^{n} \epsilon_i$$

   ¿Cuál es el valor esperado para E_1?

2. Definimos el error acumulado como la suma de los cuadrados de todos los errores (cantidades positivas):

$$E_2 = \sum_{i=1}^{n} \epsilon_i^2$$

   ¿Cuál es el valor esperado para E_2?

A la vista de los resultados, cuál de las dos cantidades, E_1 y E_2, le parece más conveniente utilizar en una estimación del error cometido por un instrumento.

Solución:

Suponiendo que todas las mediciones son independientes, se tiene que

$$E_1 = \sum_{i=1}^{n} \epsilon_i = \underbrace{\underbrace{\epsilon_1}_{N(\mu,\sigma^2)} + \underbrace{\epsilon_2}_{N(\mu,\sigma^2)} + \cdots + \underbrace{\epsilon_n}_{N(\mu,\sigma^2)}}_{N(\mu,n\cdot\sigma^2)} \Longrightarrow \mathbf{E}[E_1] = \mu = 0$$

De este modo, el valor esperado para E_1 es 0, es decir, que los errores e_i van a tender a compensarse entre unos pacientes y otros. Obsérvese que si μ no fuese conocido a priori, podríamos utilizar E_1, para obtener una aproximación de μ

$$\mu \approx \frac{E_1}{n}$$

Sin embargo, el resultado E_1 no nos indica en qué medida hay mayor o menor dispersión en los errores con respecto al 0. En cuanto a E_2 podemos afirmar lo siguiente:

$$E_2 = \sum_{i=1}^{n} \epsilon_i^2 = \sigma^2 \sum_{i=1}^{n} \left(\frac{\epsilon_i}{\sigma}\right)^2 = \sigma^2 [\underbrace{\underbrace{\left(\frac{\epsilon_1}{\sigma}\right)^2}_{\chi_1^2} + \cdots + \underbrace{\left(\frac{\epsilon_n}{\sigma}\right)^2}_{\chi_1^2}}_{\chi_n^2}] \Longrightarrow \mathbf{E}[E_2] = n\cdot\sigma^2 = 400$$

En este caso los errores no se compensan entre sí, y si σ^2 no fuese conocido, podría ser ``estimado'' de modo aproximado mediante

$$\sigma^2 \approx \frac{E_2}{n}$$

Sin embargo, no obtenemos ninguna información con respecto a μ.

En conclusión, E_1 podría ser utilizado para calcular de modo aproximado μ, y E_2 para calcular de modo aproximado σ^2. Las dos cantidades tienen interés, y ninguna lo tiene más que la otra, pues ambas formas de medir el error nos aportan información.

El siguiente resultado será de importancia más adelante. Nos afirma que la media de distribuciones normales independientes es normal pero *con menor varianza* y relaciona los grados de libertad de una v.a. con distribución χ, con los de un estadístico como la varianza (página ☐):

6.8.8.3 Teorema (Cochran)

Sean $X_1, \ldots, X_n \rightsquigarrow \mathbf{N}(\mu, \sigma^2)$ v.a. independientes. Entonces

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \leadsto \mathbf{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \leadsto \chi^{\,2}_{\,n-1}$$

$$\overline{X} \text{ y } \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \text{ son v.a. independientes.}$$

Next Up Previous

Nodo Raíz: 6.8 Distribuciones continuas
Siguiente: 6.8.10 Distribución de Student
Previo: 6.8.6 Distribución normal o gaussiana

Éste texto es la versión electrónica del manual de la Universidad de Málaga:
Bioéstadística: Métodos y Aplicaciones
U.D. Bioestadística. Facultad de Medicina. Universidad de Málaga.
ISBN: 847496-653-1