Distribución χ-cuadrado

Daniel López
David Charte
Universidad de Granada
6 de enero de 2016

Índice

1. Definición

La distribución χ-cuadrado se puede definir como la suma de cuadrados de variables aleatorias siguiendo distribuciones normales 0-1:

Definición 1.1 (Distribución χ-cuadrado)**.** *Si $X_1, X_2, \ldots X_p$ son variables aleatorias independientes con distribución $N(0,1)$, entonces a la distribución que sigue $\chi^2 = X_1^2 + X_2^2 + \ldots X_p^2$ la llamamos χ-cuadrado con p grados de libertad.*

1.1. Función de densidad

Probaremos que la distribución χ-cuadrado es un caso particular de la distribución Gamma. Recordamos la función de densidad de esta distribución:

Definición 1.2 (Distribución Gamma)**.** *Decimos que la variable aleatoria X sigue una distribución Gamma si su función de densidad es:*

$$f(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x \in [0, +\infty[, \quad \alpha, \beta > 0 \ .$$

Recordamos además el siguiente teorema:

Teorema 1.1. *Sean X e Y variables aleatorias **continuas** con funciones generatrices de momentos $M_X(t)$ y $M_Y(t)$ respectivamente. Además, supongamos que las funciones **existen** para t en un entorno de 0 y que son continuas en $t = 0$. Entonces, si $M_X(t) = M_Y(t)$ $\forall t$ las variables aleatorias tienen la misma función de densidad.*

Sea $\chi^2 = X_1^2 + X_2^2 + \ldots X_p^2$ una variable aleatoria con distribución χ-cuadrado, y sea Y una variable aleatoria siguiendo una distribución $\Gamma(\alpha, \beta)$. Calculamos las funciones generatrices de momentos. Por un lado, la de χ^2 se expresará como:

$$\begin{aligned} M_{\chi^2}(t) &= E\left[e^{t\chi^2}\right] = E\left[e^{t(X_1^2+X_2^2+\cdots+X_p^2)}\right] \\ &= E\left[e^{tX_1^2}\right] E\left[e^{tX_2^2}\right] \ldots E\left[e^{tX_p^2}\right] = E\left[e^{tX_1^2}\right]^p \end{aligned}$$

puesto que las normales son independientes. Calculamos ahora $E\left[e^{tX_1^2}\right]$:

$$\begin{aligned} E\left[e^{tX_1^2}\right] &= \int_{-\infty}^{\infty} e^{x^2 t} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \int_{-\infty}^{\infty} \frac{\exp\left(-\frac{x^2}{2(1-2t)^{-1}}\right)}{\sqrt{2\pi}} dx \\ &= \frac{1}{\sqrt{1-2t}} \int_{-\infty}^{\infty} \frac{\exp\left(-\frac{x^2}{2(1-2t)^{-1}}\right)}{\sqrt{(1-2t)^{-1}}\sqrt{2\pi}} dx \\ &\left[\textit{La integral vale 1 por ser la función de densidad de } N(0,(1-2t)^{-1}) \textit{ en su dominio}\right] \\ &= \frac{1}{\sqrt{1-2t}} \end{aligned}$$

Ahora calculamos la función generatriz de momentos de la variable Y:

$$M_Y(t) = E\left[e^{tY}\right] = \int_0^\infty e^{ty} \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-\frac{y}{\beta}} dy$$
$$= \int_0^\infty \frac{\left(\frac{1}{\beta}\right)^\alpha y^{\alpha-1} e^{-y\left(\frac{1}{\beta}-t\right)}}{\Gamma(\alpha)} dy$$
$$= \left(\frac{\frac{1}{\beta}}{\frac{1}{\beta}-t}\right)^\alpha \frac{1}{\Gamma(\alpha)} \int_0^\infty \left(\frac{1}{\beta}-t\right)\left(\left(\frac{1}{\beta}-t\right)y\right)^{\alpha-1} e^{-y\left(\frac{1}{\beta}-t\right)} dy$$
$$\left[z = \left(\frac{1}{\beta}-t\right)y\right] = \left(\frac{\frac{1}{\beta}}{\frac{1}{\beta}-t}\right)^\alpha \frac{1}{\Gamma(\alpha)} \int_0^\infty z^{\alpha-1} e^{-z} dz = \left(\frac{\frac{1}{\beta}}{\frac{1}{\beta}-t}\right)^\alpha = \frac{1}{(1-\beta t)^\alpha}$$

Si evaluamos la expresión obtenida en $\alpha = \frac{p}{2}$, $\beta = 2$ obtenemos el mismo resultado que en la función que nos da la distribución χ-cuadrado. Por tanto, la variable χ^2 se distribuye según una $\Gamma(\frac{p}{2}, 2) =: \chi^2(p)$.

Obtenemos de esta forma la siguiente definición equivalente de la distribución χ-cuadrado:

Definición 1.3 (Distribución χ-cuadrado)**.** *Decimos que la variable aleatoria X sigue una distribución χ-cuadrado con p grados de libertad si su función de densidad es:*

$$f(x \mid p) = \frac{1}{\Gamma(\frac{p}{2})2^{\frac{p}{2}}} x^{\frac{p}{2}-1} e^{-\frac{x}{2}}, \quad x \in [0, +\infty[, \quad p \in \{1, 2, \dots\}\ .$$

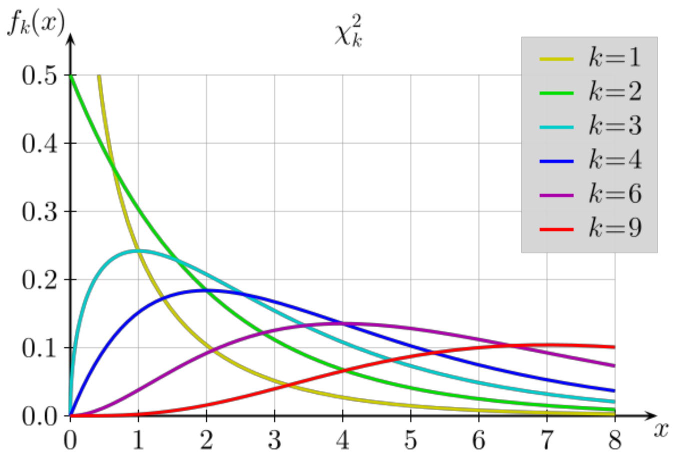

Figura 1: Función de densidad de la distribución χ-cuadrado. Imagen de Wikipedia (CC BY).

1.2. Propiedades

Propiedad 1.1. *La suma de variables aleatorias con distribución χ-cuadrado sigue una distribución χ-cuadrado.*

Propiedad 1.2. *La suma de los cuadrados de variables aleatorias con distribución normal 0-1 sigue una distribución χ-cuadrado.*

2. Función generatriz de momentos

$$M_X(t) = \left(\frac{1}{1-2t}\right)^{\frac{p}{2}}, \quad t < \frac{1}{2}$$

2.1. Esperanza

$$\mathrm{E}[X] = p$$

2.2. Varianza

$$\mathrm{Var}[X] = 2p$$

3. Estimador máximo verosimil

3.1. Insesgadez

Definición 3.1 (Insesgadez)**.** *Se denomina sesgo de un estimador a la diferencia entre la esperanza del estimador y el verdadero valor del parámetro a estimar. Un estimador es insesgado si su sesgo es nulo por ser su esperanza igual al parámetro que se desea estimar.*

3.2. Eficiencia

Definición 3.2 (Eficiencia)**.** *Un estimador $\hat{\theta}_1$ se dice que es más eficiente que otro estimador $\hat{\theta}_2$, si la varianza del primero es menor que la del segundo $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$.*

3.3. Consistencia

Teorema 3.1. *Una condición suficiente para que $\hat{\theta}$ sea un estimador consistente es que dicho estimador tiene que verificar:*

$E[\hat{\theta}] \to \theta$

$Var(\hat{\theta}) \to 0$

Con $n \to \infty$.

3.4. Suficiencia

Teorema 3.2. *Sean $X_1, X_2, ..., X_n$ variables aleatorias independientes con una función distribución de distribución conjunta $f(x_1, x_2, ..., x_n|\theta)$ que depende del parámetro θ.*
Entonces se dice que el estadístico $u(x_1, ..., x_n)$ es suficiente para θ si y solamente si $f(x_1, x_2, ..., x_n|\theta)$ se puede factorizar de la siguiente forma:

$$f(x_1, ..., x_n|\theta) = \Phi(u(x_1, ..., x_n)|\theta) \cdot h(x_1, ..., x_n)$$

4. Referencias

1. Notes on the Chi-squared Distribution - Georgia Tech Institute: http://people.math.gatech.edu/~ecroot/3225/chisquare.pdf
2. Probability. An Introduction - Geoffrey Grimmet
3. Statistical Inference - George Casella, Roger Berger