

Distribución χ -cuadrado

DANIEL LÓPEZ
DAVID CHARTE
Universidad de Granada
19 de enero de 2016

Índice

1. Definición	2
1.1. Función de densidad	2
2. Función generatriz de momentos	4
2.1. Esperanza	4
2.2. Varianza	4
3. Inferencia	4
3.1. Información de Fisher	4
4. Referencias	5

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, \dots, X_p son variables aleatorias independientes con distribución $N(0, 1)$, entonces a la distribución que sigue $\chi^2 = X_1^2 + X_2^2 + \dots + X_p^2$ 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:

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

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 $\chi^2 = X_1^2 + X_2^2 + \dots + X_p^2$ 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:

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

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

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

Por tanto, $M_{\chi^2}(t) = \frac{1}{(\sqrt{1-2t})^p}$.

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

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

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:*

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

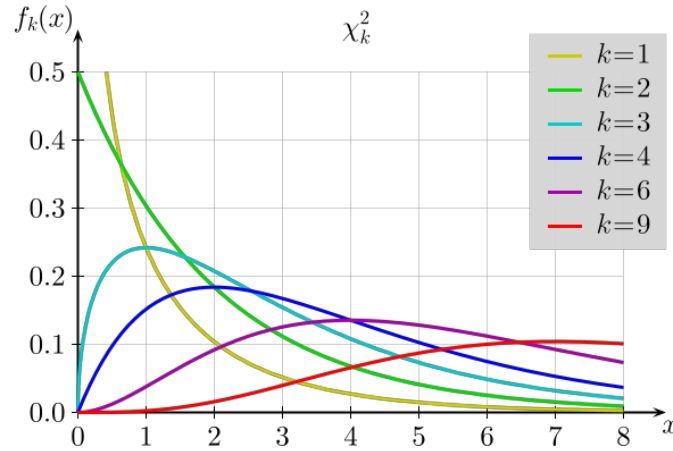


Figura 1: Función de densidad de la distribución χ -cuadrado. Imagen de Wikipedia (CC BY).

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

Para el cálculo de la esperanza derivamos la función generatriz de momentos y la evaluamos en $t=0$:

$$\frac{\partial \phi}{\partial t} = \frac{p}{2} \left(\frac{1}{1-2t} \right)^{\frac{p}{2}-1} \frac{2}{(1-2t)^2} = p \left(\frac{1}{1-2t} \right)^{\frac{p}{2}+1}$$

Una vez hallada la derivada, evaluando en $t=0$ tenemos que:

$$E[X] = p$$

2.2. Varianza

La varianza se puede expresar como $\sigma^2 = E[X^2] - E[X]^2$ por lo que es tan solo calcular el momento de orden 2 usando la función generatriz de momentos, por lo que volvemos a derivar la expresión anterior:

$$\frac{\partial^2 \phi}{\partial t^2} = p \left(\frac{p}{2} + 1 \right) \left(\frac{1}{1-2t} \right)^{\frac{p}{2}} \left(\frac{2}{(1-2t)^2} \right) = (p^2 + 2p) \left(\frac{1}{1-2t} \right)^{\frac{p}{2}+2}$$

Evaluamos la expresión en $t=0$ y tenemos que:

$$E[X^2] = (p^2 + 2p)$$

Y por lo tanto la expresión de la varianza es la siguiente:

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

3. Inferencia

3.1. Información de Fisher

Definición 3.1 (Información de Fisher). *Para una variable X cuya distribución tiene función de densidad $f(x | \theta)$, la información de Fisher se define como*

$$I_X(\theta) = E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \log f(x | \theta) \right)^2 \right]$$

Recordamos que, bajo ciertas circunstancias, la información de Fisher se puede calcular mediante la siguiente expresión:

$$I_X(\theta) = E_{\theta} \left[-\frac{\partial^2}{\partial \theta^2} \log f(x | \theta) \right]$$

En el caso de la distribución χ -cuadrado, calculamos primero la derivada:

$$\begin{aligned} \frac{\partial}{\partial p} \log f(x | p) &= \frac{\partial}{\partial p} \log \left(\frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}} \right) = \frac{\partial}{\partial p} \left(\left(\frac{p}{2} - 1 \right) \log x - \frac{x}{2} - \log \Gamma \left(\frac{p}{2} \right) - \frac{p}{2} \log 2 \right) \\ &= \frac{1}{2} \log x - \frac{1}{2} \psi \left(\frac{p}{2} \right) - \frac{1}{2} \log 2 = \frac{1}{2} \left(\log \frac{x}{2} - \psi \left(\frac{p}{2} \right) \right) \end{aligned}$$

$$\frac{\partial^2}{\partial p^2} \log f(x | p) = \frac{\partial}{\partial p} \frac{1}{2} \left(\log \frac{x}{2} - \psi \left(\frac{p}{2} \right) \right) = -\frac{1}{4} \psi' \left(\frac{p}{2} \right)$$

Por tanto, la información de Fisher nos queda:

$$\begin{aligned} I_X(p) &= E_p \left[-\frac{\partial^2}{\partial p^2} \log f(x | p) \right] = \int \left(-\frac{\partial^2}{\partial p^2} \log f(x | p) \right) f(x | p) \, dx \\ &= \int \frac{1}{4} \psi' \left(\frac{p}{2} \right) f(x | p) \, dx = \frac{1}{4} \psi' \left(\frac{p}{2} \right) \int f(x | p) \, dx = \frac{1}{4} \psi' \left(\frac{p}{2} \right) \end{aligned}$$

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