REVISTA DE LA FACULTAD DE INGENIERIA
www.ingenieria.uda.cl
23 (2009) 46-51

UNA NUEVA DISTRIBUCIÓN DE PROBABILIDAD BASADA EN LA DISTRIBUCIÓN NORMAL

David Elal, Salim Elal

Departamento de Matemáticas, Universidad de Atacama, Copiapó, Chile.
delal@mat.uda.cl ; selal@mat.uda.cl

RESUMEN

Se presenta una nueva distribución de probabilidad basada en la distribución normal con la interesante propiedad de otorgar mayor peso a las colas y en consecuencia capturar información que se puede encontrar alejada de la media. Su curtosis negativa permite clasificarla dentro de las distribuciones platicúrtica cuya principal propiedad es de descongestionar la información alrededor de la media. Debido a esta caracterización se la ha definido como Distribución Normal Platicúrtica. Se estudian sus principales propiedades y se muestra una representación estocástica la que además de entregar una elegante fórmula para cálculo de momentos de orden n, permite generar números aleatorios. La Distribución Normal Platicúrtica es simétrica y tiene un comportamiento muy similar a la Distribución Normal por lo que resulta ser una atractiva alternativa de aplicación frente a situaciones reales.

Palabras claves: distribución normal, representación estocástica, simetría y curtosis.

ABSTRACT

The following article discusses a new probability distribution based on the normal distribution with the property of giving more weigth to the tails, thereby consequentially capturing information that can be present but removed from the mean. Its negative Kurtosis allow it to be located within the platycurtic distributions whose main property is that of decongesting the information around the mean. For this reason these new distribution has been defined as Normal Platycurtic Distribution. A study of the principal properties and a stochastic representation is performed providing an effective formula for the calculation of n^{th} – moment and for the generation of random number. The Normal Platycurtic Distribution is symmetrical and behaves very similar to the Normal Distribution, which make it an attractive alternative for application in real situations.

Keywords: normal distribution, stochastic representation, symmetry and kurtosis.

1. INTRODUCCIÓN

Presentamos una nueva distribución de probabilidad basada en la distribución normal que se caracteriza por su simplicidad en el manejo algebraico y con interesantes propiedades que iremos desarrollando, quizás destacar su representación estocástica la que permite generar números aleatorios para fines de simulación. Mostramos un histograma con números aleatorios generados por dicha representación y ajustamos el modelo a dichos números. Usamos la notación $\phi(x)$ y $\Phi(t)$ para denotar la densidad normal estándar y la función de distribución normal acumulada respectivamente, es decir:

$$\phi(\mathrm{x})=\frac{1}{\sqrt{\frac{2}{\pi}}}\exp\{-\mathrm{x}^2/2)\} \quad \text{y} \quad \Phi(\mathrm{t})=\int_{-\infty}^{\mathrm{t}}\phi(\mathrm{x})\mathrm{dx}$$

Denotamos por Z la variable aleatoria cuya función de densidad es $\phi(x)$ lo que resumimos en forma compacta escribiendo $Z \sim N(0,1)$, es bien sabido que para todo $k \in \mathbb{N}$:

$$\mathrm{E}[\mathrm{Z}^{2\mathrm{k}-1}]=0 \quad \text{y} \quad \mathrm{E}[\mathrm{Z}^{2\mathrm{k}}]=\prod_{\mathrm{j}=1}^{\mathrm{k}}(2\mathrm{j}-1)$$

2. DISTRIBUCIÓN PLATICÚRTICA

Proposición 2.1 La función

$$f(x)=\left(\frac{3+\mathrm{x}^2}{4}\right)\phi(\mathrm{x})$$

es una función de densidad

Demostración: $\left(\frac{3+\mathrm{x}^2}{4}\right)\geq 0$ y $\phi(\mathrm{x})\geq 0$ para todo $x \in \mathbb{R}$ por lo que $f(x)\geq 0$, por otra parte

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty}\left(\tfrac{3+\mathrm{x}^2}{4}\right)\phi(\mathrm{x})\mathrm{dx}$$
$$= \tfrac{3}{4}\int_{-\infty}^{\infty}\phi(\mathrm{x})\mathrm{dx} + \tfrac{1}{4}\int_{-\infty}^{\infty}\mathrm{x}^2\phi(\mathrm{x})\mathrm{dx}$$
$$= \tfrac{3}{4}+\tfrac{1}{4}$$
$$= 1$$

por lo tanto $f(x)$ es una función de densidad

Definición 2.1 Si la variable aleatoria X tiene como densidad a:

$$f(x) = \left(\frac{3+\mathrm{x}^2}{4}\right)\phi(\mathrm{x})$$

decimos que se distribuye según una normal platicúrtica y lo denotamos como: $X \sim NP$

Proposición 2.2 Sea $X \sim NP$ y sea $f(x) = \left(\frac{3+\mathrm{x}^2}{4}\right)\phi(\mathrm{x})$ función de densidad de X

entonces:

a) $f(x)$ es simétrica alrededor del cero

b) $f(x)$ es unimodal

Demostración Claramente la función $f(x)$ es simétrica dado que $f(x) = f(-x)$ para todo $x \in \mathbb{R}$, por potra parte

$$f^{'}(x) = -\tfrac{1}{4}x\phi(x) - \tfrac{1}{4}x^3\phi(x)$$

siendo $f''(x) < 0$ por lo que $f(x)$ presenta un solo máximo en $x = 0$ lo que la hace unimodal

Proposición 2.3 Sea $X \sim NP$. Si $F(t)$ y $\Gamma(t)$ representan las funciones de distribución acumulada y generatriz de momentos de la variable aleatoria X respectivamente, entonces:

$$F(t) = \Phi(t) - \tfrac{1}{4} t\phi(t)$$
$$\Gamma(t) = \exp\left(\frac{t^2}{2}\right)\left[1 + \frac{t^2}{4}\right]$$

Demostración.

$$F(t) = \int_{-\infty}^{t} \left(\tfrac{3+x^2}{4}\right)\phi(x)dx$$
$$= \tfrac{3}{4}\Phi(t) + \tfrac{1}{4}\int_{-\infty}^{t} x^2\phi(x)dx$$
$$= \tfrac{3}{4}\Phi(t) + \tfrac{1}{4}[-t\phi(t) + \Phi(t)]$$
$$= \Phi(t) - \tfrac{1}{4}t\phi(t)$$

por otra parte

$$\Gamma(t) = E[\exp(tX)]$$
$$= \int_{-\infty}^{\infty} \exp(tx) \left(\tfrac{3+x^2}{4}\right)\phi(x)dx$$
$$= \tfrac{3}{4}\exp\left(\tfrac{t^2}{2}\right) + \tfrac{1}{4}\int_{-\infty}^{\infty} \exp(tx)x^2\phi(x)dx$$
$$= \tfrac{3}{4}\exp\left(\tfrac{t^2}{2}\right) + \tfrac{1}{4}\left(t^2\exp\left(\tfrac{t^2}{2}\right) + \exp\left(\tfrac{t^2}{2}\right)\right)$$
$$= \exp\left(\tfrac{t^2}{2}\right)\left[1 + \tfrac{t^2}{4}\right]$$

Proposición 2.4 Si $X \sim NP$, para todo $k \in \mathbb{N}$, se cumple que

$$E[X^{2k-1}] = 0 \quad \text{y}$$
$$E[Z^{2k}] = \tfrac{k+2}{2}\prod_{j=1}^{k}(2j-1)$$

Demostración.

$$E[X^{2k-1}] = \int_{-\infty}^{\infty} x^{2k-1}\left(\tfrac{3+x^2}{4}\right)\phi(x)dx$$
$$= \tfrac{3}{4}E[Z^{2k-1}] + \tfrac{1}{4}E[Z^{2k+1}]$$
$$= 0$$

por otra parte

$$E[X^{2k}] = \int_{-\infty}^{\infty} x^{2k}\left(\tfrac{3+x^2}{4}\right)\phi(x)dx$$
$$= \tfrac{3}{4}E\left[Z^{2k}\right] + \tfrac{1}{4}E\left[Z^{2(k+1)}\right]$$
$$= \tfrac{3}{4}\prod_{j=1}^{k}(2j-1) + \tfrac{1}{4}\prod_{j=1}^{k+1}(2j-1)$$
$$= \tfrac{k+2}{2}\prod_{j=1}^{k}(2j-1)$$

El siguiente resultado justifica el nombre de Distribución Normal Platicúrtico debido a que dicho modelo presenta una curtosis negativa lo que caracteriza a las funciones de densidad platicúrtica. Pasamos a recordar la fórmula que permite el cálculo de la curtosis que denotamos por la letra K.

$$K = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^2}{(\mu_2 - \mu_1)^2} - 3$$

donde μ_1, μ_2, μ_3 y μ_4 son los momentos de orden 1,2,3 y 4 respectivamente.

Proposición 2.5 La curtosis en el modelo NP es de $-0.\overline{3}$

Demostración. Considerando que los momentos impares del modelo NP son cero y sabiendo que:

$$\mu_2 = \frac{3}{2} \qquad \text{y} \qquad \mu_4 = 6$$

se tiene que

$$K = -0.\overline{3}$$

La siguiente proposición es un importante resultado que deja de manifiesto que el peso de las colas del modelo normal platicúrtico supera el peso de las colas del modelo normal lo que le permite capturar mayor información distante de la media.

Proposición 2.6 Si $X \sim NP$ y $f(x)$ su función de densidad, entonces

$$f(x) = f(x) \geq \phi(\mathrm{x}) \quad \text{si} \quad |\mathrm{x}| \geq 1$$

$$f(x) = f(x) \leq \phi(\mathrm{x}) \quad \text{si} \quad |\mathrm{x}| \leq 1$$

Demostración. Suponga que $|x| \geq 1$, en tal caso se tiene que la expresión $\left(\frac{3+x^2}{4}\right) \geq 1$ y dado que $\phi(x) \geq 0$ para todo $x \in \mathbb{R}$, entonces

$$f(x) = \left(\frac{3+\mathrm{x}^2}{4}\right)\phi(x) \geq \phi(\mathrm{x})$$

por otra parte, si $|x| \leq 1$, se tiene que la expresión $\left(\frac{3+x^2}{4}\right) \leq 1$ y en consecuencia

$$f(x) = \left(\frac{3+\mathrm{x}^2}{4}\right)\phi(x) \leq \phi(\mathrm{x})$$

3. REPRESENTACIÓN ESTOCÁSTICA DEL MODELO NP

En esta sección presentamos una representación estocástica del modelo NP para generar números aleatorios lo que se logra conociendo previamente la representación estocástica del modelo normal bimodal que pasamos a desarrollar

Definición 3.1 Si la variable aleatoria W tiene como densidad a:

$$\mathrm{g(x)} = \mathrm{x}^2\phi(\mathrm{x})$$

decimos que se distribuye según una normal bimodal y lo denotamos como $W \sim NB$

La siguiente proposición muestra algunas propiedades básicas del modelo NB

Proposición 3.1 Sea $W \sim NB$ y sean $f_X(x),\ F_X(t)\ y\ \Gamma_X(t)$ las funciones de densidad, distribución acumulada y generatriz de momentos respectivamente de la variable aleatoria X, entonces

a) $f_X(x)$ es simétrica respecto del cero
b) $F_X(t) = \Phi(t) - t\phi(t)$
c) $\Gamma_X(t) = \exp\left(\frac{t^2}{2}\right)\left(1+t^2\right)$

Demostración. Las demostraciones de cada item son similares a las del modelo NP de la proposición2.3

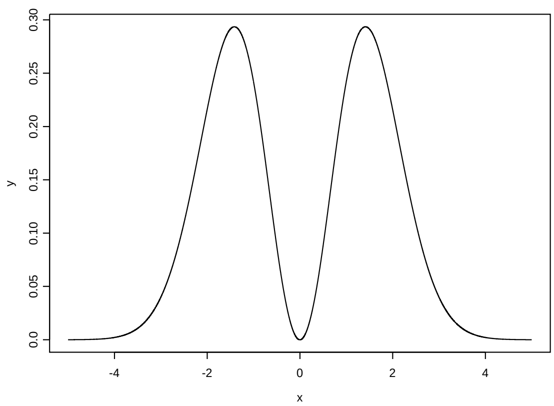

Figura 1. densidad modelo NB

Proposición 3.2 Sean $X\ y\ U$ variables aleatorias independientes donde $X \sim \chi^2_3$ (chi-cuadrado con 3 grados de libertad) y U es tal que:

$$P[U=1] = P[U=-1] = \frac{1}{2}$$

Si $W = \sqrt{XU}$ entonces $W \sim NB$

Demostración. Sea $F_W(w)$ la función de distribución acumulada de W, entonces:

$$\begin{aligned} F_W(w) &= P\left[W \le w\right] \\ &= P\left[\sqrt{X}\,U \le w\right] \\ &= \frac{1}{2}\left[\sqrt{X} \le w\right] + \frac{1}{2}\left[-\sqrt{X} \le w\right] \end{aligned}$$

así se tiene que

$$F_W(w)=\begin{cases} \frac{1}{2} P\left[\sqrt{X} \le w\right] + \frac{1}{2} & si \quad w \ge 0 \\ \frac{1}{2} P\left[\sqrt{X} \le w\right] & si \quad w < 0 \end{cases}$$

$$F_W(w)=\begin{cases} \frac{1}{2} P\left[X \le w^2\right] + \frac{1}{2} & si \quad w \ge 0 \\ \frac{1}{2}\left(1 - P\left[X \le w^2\right]\right) & si \quad w < 0 \end{cases}$$

$$F_W(w)=\begin{cases} \frac{1}{2} F_X\left(w^2\right) + \frac{1}{2} & si \quad w \ge 0 \\ \frac{1}{2} - \frac{1}{2} F_X\left(w^2\right) & si \quad w < 0 \end{cases}$$

derivando la función $F_W(w)$ con respecto a w, para encontrar la densidad $f_W(w)$, se tiene que:

$$f_W(w)=\begin{cases} w\, f_X\left(w^2\right) & si \quad w \ge 0 \\ -w\, f_X\left(w^2\right) & si \quad w < 0 \end{cases}$$

Obteniendo que:

$f_W(w) = |w|\, f_X\left(w^2\right)$ para todo $w \in \mathbb{R}$

donde $f_X(x)$ corresponde a la función de densidad de una chi-cuadrado con 3 grados de libertad, dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}} x^{\frac{1}{2}} e^{\frac{1}{2}}$$

de tal forma que

$$\begin{aligned} f_W(w) &= |w|\, f_X\left(w^2\right) \\ &= |w| \frac{1}{\sqrt{2\pi}} (w^2)^{\frac{1}{2}} \exp\left(\tfrac{w^2}{2}\right) \\ &= |w||w|\phi(w) \\ &= w^2\phi(w) \end{aligned}$$

A continuación presentamos uno de los resultados mas importante y tiene relación con la representación estocástica del modelo NP, que se ha logrado, inspirado en la metodología empleada por Henze[2] con el modelo $SN(\lambda)$ presentado por Azzalini[1].

Proposición 3.3 Considere las variables aleatorias independientes Z e Y donde $Z \sim N(0,1)$ e $Y \sim NB$.

Si se cumple que

$$W = \frac{\sqrt{3}}{2} Z + \frac{1}{2} Y$$

entonces

$$W \sim NP$$

Demostración. Sea $\Gamma_W(w)$ la función generatriz de momentos de la variable aleatoria W, se tiene entonces que:

$$\begin{aligned} \Gamma_W(w) &= E\left[\exp(t\,W)\right] \\ &= E\left[\exp\left(t\left[\tfrac{\sqrt{3}}{2} Z + \tfrac{1}{2} Y\right]\right)\right] \\ &= E\left[\exp\left(\tfrac{t\sqrt{3}}{2} Z\right)\right] E\left[\exp\left(\tfrac{t}{2} Y\right)\right] \\ &= \exp\left(\tfrac{3t^2}{8}\right)\left[\exp\left(\tfrac{t^2}{8}\right)\left(1+\tfrac{t^2}{4}\right)\right] \\ &= \exp\left(\tfrac{t^2}{2}\right)\left(1+\tfrac{t^2}{4}\right) \end{aligned}$$

Observe que la función $\Gamma_W(w)$ es igual a la función generatriz de momentos del modelo NP, por lo tanto $W \sim NP$.

A través de la representación estocástica del modelo NP se generaron 2000 datos los que fueron representados en un histograma de 16 intervalos, enseguida se ajustó la función de densidad del modelo NP a dichos datos dando origen al siguiente gráfico.

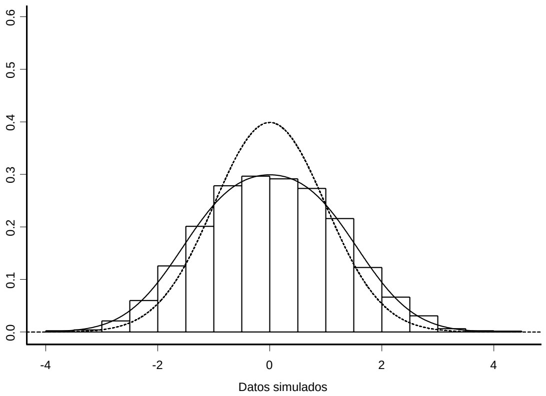

Figura 2. densidad normal (línea segmentada) y densidad del modelo NP(línea sólida)

Es importante destacar que la densidad del modelo NP ajusta muy bien los 2000 datos debido a que estos fueron generados por su propio modelo (representación estocástica), por otra parte, la densidad normal descrita por la línea segmentada se ha superpuesto sobre el histograma solamente para advertir la diferencia que presenta al ser comparada, su ajuste, con el modelo NP. A la vez se entrega una idea de cómo se deben presentar, en situaciones reales, los histogramas para aplicar uno u otro modelo.

Observación 3.1 Un tema interesante de abordar en este modelo NP, para fines inferenciales, es incorporar un parámetro de localización μ y un parámetro de escala σ mediante la reparametrización

$$Y = \mu + \sigma X \quad donde \quad X \sim NP$$

con $\mu \in \mathbb{R}$ y $\sigma > 0$

Así el nuevo modelo se presentaría como normal platicúrtico de parámetros μ y σ denotado como $NP(\mu, \sigma)$, donde: $NP(0,1) = NP$

4. REFERENCIAS

[1] Azzalini, A. (1985). A Class of Distributions which includes the Normal Ones. Scan. J. Statist. **12**, 171-8.

[2] Henze, N. (1986) A Probability Representation of the Skew-normal Distribution. Scand. J. Statist. **13**, 271-275.