

UNA INTRODUCCIÓN AL MODELO DE REGRESIÓN LINEAL NORMAL-ASIMÉTRICO. PARTE 1: TEORÍA, PARAMETRIZACIÓN DIRECTA

AN INTRODUCTION TO THE NORMAL-SKEW LINEAR REGRESSION MODEL. PART 1: THEORY, DIRECT PARAMETERIZATION

Omar Chocotea Poca¹

Carrera de Estadística, Universidad Mayor de San Andrés, Bolivia

✉ ochocoteap@fcpn.edu.bo

Artículo recibido: 13-08-2022

Artículo aceptado: 22-09-2022

RESUMEN

En el espectro de las distribuciones elípticas-asimétricas se encuentra la distribución normal-asimétrica que tiene como caso particular a la distribución normal. El objetivo principal de este trabajo es presentar al modelo de regresión lineal con variable aleatoria de error con distribución normal-asimétrica. Este trabajo se divide en tres partes. En esta parte, se revisa los aspectos teóricos base asociados con la parametrización directa. En la Parte 2, se revisa los aspectos teóricos base asociados con la parametrización centrada. Y en la Parte 3, se efectúa la ilustración con un conjunto de datos reales.

Palabras clave: Distribuciones elípticas-asimétricas; Distribución normal-asimétrica; Modelo de regresión lineal.

ABSTRACT

In the spectrum of elliptic-asymmetric distributions is the skew-normal distribution, which has the normal distribution as a particular case. The main objective of this work is to present the linear regression model with random error variable with skew-normal distribution. This work is divided into three parts. In this part, the basic theoretical aspects associated with direct parameterization are reviewed. In Part 2, the basic theoretical aspects associated with centered parameterization are reviewed. And in Part 3, the illustration is done with a real data set.

Key words: Elliptic-asymmetric distributions; Skew-normal distribution; Linear regression model

1. INTRODUCCIÓN

El modelo de regresión lineal (MRL) permite modelar la función de regresión como una combinación lineal de covariables. La teoría del MRL está bien establecida. Además, el MRL es la base de muchas herramientas de modelado modernas. Por ejemplo, Naseem et al. (2010) presentan un enfoque novedoso de identificación de rostros mediante la formulación del problema de reconocimiento

de patrones en términos del MRL.

Azzalini & Capitanio (1999), Sahu et al. (2003), Rubio & Genton (2016) y Ferreira et al. (2018) entre otros reportan que el supuesto clásico de normalidad de los errores resulta restrictivo en situaciones donde se evidencia asimetría para los datos

Para superar este posible inconveniente, en este de trabajo dividido en tres partes, se

¹ ORCID: 0000-0002-6490-7491.

presenta una introducción al MRL normal-asimétrico bajo la distribución normal-asimétrica de Azzalini (1985), siguiendo los trabajos de Azzalini (1985), Azzalini & Capitanio (1999), y Azzalini & Capitanio (2014). Este modelo anida al MRL habitual.

Este artículo está organizado de la siguiente manera. La Sección 2, revisa los momentos de la distribución normal-asimétrica. La sección 3, define la formulación del modelo con la parametrización directa. Finalmente, la Sección 4, revisa la inferencia por el método de máxima verosimilitud hasta obtener la matriz de información.

2. MOMENTOS

El Lema 1 de Azzalini (1985) permite producir la función de densidad de la normal-asimétrica.

Una v.a. z tiene una distribución normal-asimétrica estándar con parámetro de asimetría $\alpha \in \mathbb{R}$ y representaremos por $z \sim \text{NA}(\alpha)$, si su función de densidad está dada por

$$\varphi(z; \alpha) = 2\varphi(z)\Phi(\alpha z),$$

donde $\varphi(\cdot)$ y $\Phi(\cdot)$ denotan la función de densidad y función de distribución de la distribución normal estándar, respectivamente. La apariencia gráfica de la función de densidad de la $\text{NA}(\alpha)$ se muestra en la Figura 1 para algunos valores de α .

Para valores negativos de α se tiene una distribución asimétrica a la izquierda, y para valores positivos de α se tiene una distribución asimétrica a la derecha.

Son varias las propiedades que se derivan inmediatamente de la definición anterior. Si $z \sim \text{NA}(\alpha)$, se cumplen las siguientes propiedades:

- a) $\varphi(z; 0) = \varphi(z)$ para todo z ;
- b) $-z \sim \text{NA}(-\alpha)$; y
- c) $z^2 \sim \chi_1^2$, independientemente de α .

La función generatriz de cumulantes de $z \sim \text{NA}(\alpha)$ es

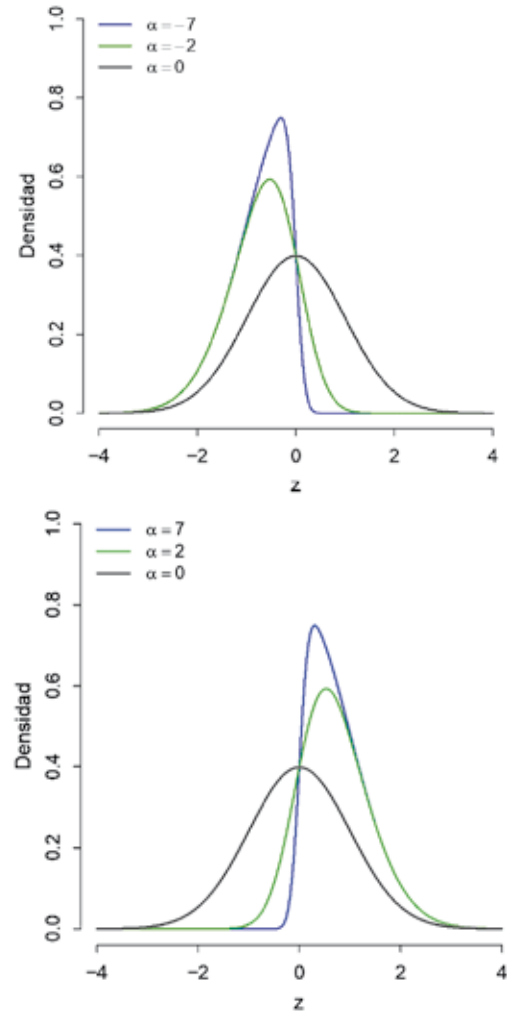
$$\text{fgc}_z(t) = \frac{1}{2}t^2 + \zeta_0(\delta t),$$

donde $\delta = \delta(\alpha) = \alpha/\sqrt{1 + \alpha^2} \in [-1, 1]$ y

$$\zeta_0(x) = \log\{2\Phi(x)\}.$$

Figura 1

Función de densidad de la $\text{NA}(\alpha)$ para algunos valores de α .



Fuente: Elaboración propia

El siguiente resultado se utilizará en la última sección.

Si

$$\zeta_r(x) = \frac{d^r}{dx^r} \zeta_0(x)$$

con $r = 1, 2, \dots$. Entonces

$$\begin{aligned}\zeta_1(x) &= \frac{\varphi(x)}{\Phi(x)}, \\ \zeta_2(x) &= -\zeta_1(x)\{x + \zeta_1(x)\}, \\ \zeta_3(x) &= -\zeta_2(x)\{x + \zeta_1(x)\} - \zeta_1(x)\{1 + \zeta_2(x)\}, \\ \zeta_4(x) &= -\zeta_3(x)\{x + 2\zeta_1(x)\} - 2\zeta_2(x)\{1 + \zeta_2(x)\}.\end{aligned}$$

y

$$\begin{aligned}\zeta_1(0) &= \sqrt{\frac{2}{\pi}}, \\ \zeta_2(0) &= -\zeta_1^2(0), \\ \zeta_3(0) &= 2\zeta_1^3(0) - \zeta_1(0), \\ \zeta_4(0) &= -6\zeta_1^4(0) + 4\zeta_1^2(0).\end{aligned}$$

En consecuencia: la esperanza y la varianza están dadas por

$$\mu_z = \sqrt{\frac{2}{\pi}} \delta$$

y

$$\sigma_z^2 = 1 - \frac{2}{\pi} \delta^2,$$

respectivamente; y los coeficientes de asimetría y curtosis están dados por

$$\gamma_1 = \frac{4 - \pi}{2} \left(\frac{\mu_z}{\sigma_z} \right)^3$$

y

$$\gamma_2 = 2(\pi - 3) \left(\frac{\mu_z}{\sigma_z} \right)^4,$$

respectivamente.

El coeficiente de asimetría es una función impar, y el coeficiente de curtosis es una función par. Efectuando una evaluación de los valores mínimo y/o máximo de los coeficientes de asimetría y curtosis, se tiene que los coeficientes de asimetría y curtosis varían en los intervalos $[-0.995; 0.995]$ y $[0; 0.869]$, respectivamente.

De acuerdo con la teoría de la familia de localización-escala (ver, por ejemplo, Lehmann & Casella, 1998, p. 17), la distribución normal-asimétrica con parámetro de localización $\xi \in \mathbb{R}$ parámetro de escala $\omega > 0$ y parámetro de asimetría α se genera con la transformación

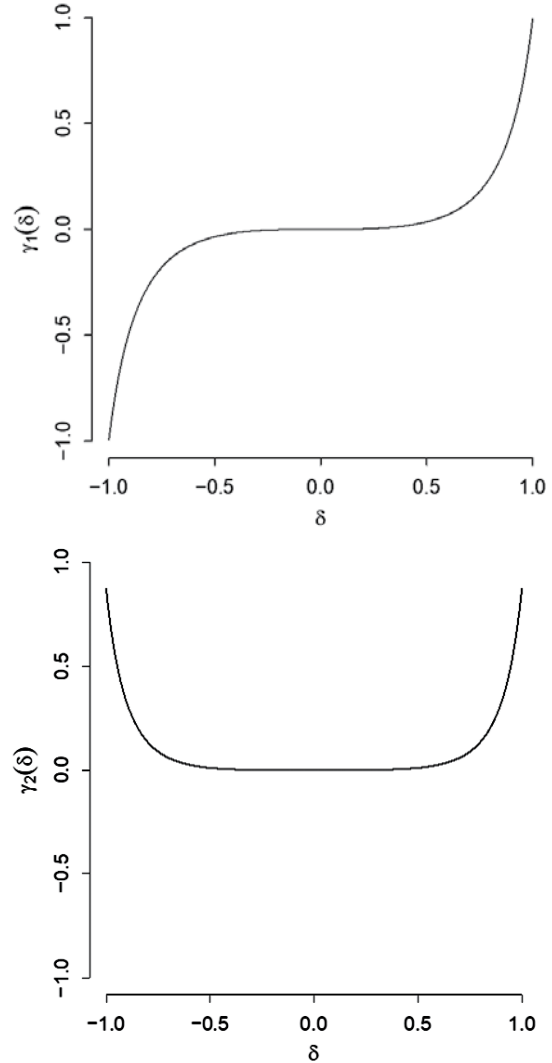
$$y = \xi + \omega z,$$

y representaremos por $y \sim \text{NA}(\xi, \omega^2, \alpha)$, si su función de densidad es

$$\frac{2}{\omega} \varphi\left(\frac{y - \xi}{\omega}\right) \Phi\left(\alpha \frac{y - \xi}{\omega}\right) = \frac{1}{\omega} \varphi\left(\frac{y - \xi}{\omega}; \alpha\right).$$

Figura 2

Coeficientes de asimetría y curtosis de la $\text{NA}(\alpha)$.



Fuente: Elaboración propia

A continuación, con base en la anterior transformación se extiende las características de la de $NA(\alpha)$.

Si $y \sim NA(\xi, \omega^2, \alpha)$, se cumplen las siguientes propiedades:

- a) $\varphi(y; \xi, \omega^2, 0) = \varphi(y; \xi, \omega^2)$ para todo y ;
- b) $-y \sim NA(-\xi, \omega^2, -\alpha)$;
- c) $(y - \xi) / \omega \sim NA(\alpha)$;
- d) $\{(y - \xi)/\omega\}^2 \sim \chi_1^2$, independientemente de α .

La esperanza y la varianza de la $NA(\xi, \omega^2, \alpha)$ son

$$\mu_y = \xi + \omega \mu_z$$

y

$$\sigma_y^2 = \omega^2 \sigma_z^2,$$

respectivamente. Los coeficientes de asimetría y curtosis de la $NA(\xi, \omega^2, \alpha)$ son iguales a los coeficientes de asimetría y curtosis de la $NA(\alpha)$, respectivamente.

3. FORMULACIÓN

La sección modifica el supuesto del MRL habitual, en formulación.

Sea $\mathbf{y} = (y_1, y_2, \dots, y_n)^T \in \mathbb{R}^n$ un vector de variables aleatorias independientes, $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik})^T \in \mathbb{R}^k$ un vector de covariables donde x_{i1} es igual a uno (corresponde a una intersección), $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^T \in \mathbb{R}^k$ un vector de coeficientes de regresión, y $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^T \in \mathbb{R}^n$ un vector de errores aleatorios independientes tales que $\epsilon_i \sim NA(\alpha)$,

$$\varphi(\epsilon_i; \alpha) = 2\varphi(\epsilon_i) \Phi(\alpha \epsilon_i),$$

con $i = 1, \dots, n$. Suponiendo x_i completamente observado, el i -ésimo predictor lineal es $\mathbf{x}_i^T \boldsymbol{\beta}$. El modelo de regresión lineal normal-asimétrico vía parametrización directa está

dado por

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \omega \epsilon_i.$$

La función de densidad de y_i es

$$\frac{2}{\omega} \varphi(z_i) \Phi(\alpha z_i),$$

donde $z_i = (y_i - \mathbf{x}_i^T \boldsymbol{\beta})/\omega$.

4. INFERENCIA

La sección revisa la inferencia del modelo de regresión lineal normal-asimétrico vía parametrización directa con el método de máxima verosimilitud.

Sea $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \omega, \alpha)^T \in \mathbb{R}^{k+2}$. El logaritmo natural de la verosimilitud del modelo de regresión lineal normal-asimétrico vía parametrización directa es

$$\ell(\boldsymbol{\theta}) = \text{constante} - n \log \omega - \frac{1}{2} \mathbf{z}^T \mathbf{z} + \boldsymbol{\lambda}_0^T \mathbf{1}_n,$$

donde $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$,

$$\boldsymbol{\lambda}_0 = (\zeta_0(\alpha z_1), \zeta_0(\alpha z_2), \dots, \zeta_0(\alpha z_n))^T, \text{ y}$$

$$\mathbf{1}_n = (1, 1, \dots, 1)^T.$$

En consecuencia: la primera derivada de $\ell(\boldsymbol{\theta})$ es

$$\mathcal{S} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \mathcal{S}_1 \\ \mathcal{S}_2 \\ \mathcal{S}_3 \end{bmatrix}$$

con

$$\mathcal{S}_1 = \frac{1}{\omega} \mathbf{X}^T (\mathbf{z} - \alpha \boldsymbol{\lambda}_1)$$

$$\mathcal{S}_2 = \frac{1}{\omega} (-n + \mathbf{z}^T \mathbf{z} - \alpha \boldsymbol{\lambda}_1^T \mathbf{z})$$

$$\mathcal{S}_3 = \boldsymbol{\lambda}_1^T \mathbf{z},$$

donde $\boldsymbol{\lambda}_1 = (\zeta_1(\alpha z_1), \zeta_1(\alpha z_2), \dots, \zeta_1(\alpha z_n))^T$; y la matriz de información es

$$\mathcal{J} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \begin{bmatrix} \mathcal{J}_{11} & \mathcal{J}_{12} & \mathcal{J}_{13} \\ \mathcal{J}_{21} & \mathcal{J}_{22} & \mathcal{J}_{23} \\ \mathcal{J}_{31} & \mathcal{J}_{32} & \mathcal{J}_{33} \end{bmatrix},$$

con

$$\begin{aligned}
 \mathcal{J}_{11} &= -\frac{1}{\omega^2} \mathbf{X}^\top \{\mathbf{I}_n + \alpha^2 \mathbf{D}(\boldsymbol{\lambda}_2)\} \mathbf{X} & \mathcal{J}_{32} &= -\frac{1}{\omega} \{\boldsymbol{\lambda}_1^\top \mathbf{z} + \alpha \mathbf{z}^\top \mathbf{D}(\boldsymbol{\lambda}_2) \mathbf{z}\} \\
 \mathcal{J}_{21} &= -\frac{1}{\omega^2} \mathbf{X}^\top \{2\mathbf{z} - \alpha \boldsymbol{\lambda}_1 - \alpha^2 \mathbf{D}(\boldsymbol{\lambda}_2) \mathbf{z}\} & \mathcal{J}_{33} &= \mathbf{z}^\top \mathbf{D}(\boldsymbol{\lambda}_2) \mathbf{z}, \\
 \mathcal{J}_{22} &= -\frac{1}{\omega^2} \{-n + 3\mathbf{z}^\top \mathbf{z} - 2\alpha \boldsymbol{\lambda}_1^\top \mathbf{z} - \alpha^2 \mathbf{z}^\top \mathbf{D}(\boldsymbol{\lambda}_2) \mathbf{z}\} & \mathcal{J}_{12} &= \mathcal{J}_{21}^\top, \quad \mathcal{J}_{13} = \mathcal{J}_{31}^\top, \quad \mathcal{J}_{23} = \mathcal{J}_{32}^\top, \text{ donde} \\
 \mathcal{J}_{31} &= -\frac{1}{\omega} \mathbf{X}^\top \{\boldsymbol{\lambda}_1 + \alpha \mathbf{D}(\boldsymbol{\lambda}_2) \mathbf{z}\} & \boldsymbol{\lambda}_2 &= (\zeta_2(\alpha z_1), \zeta_2(\alpha z_2), \dots, \zeta_2(\alpha z_n))^\top \quad \text{y} \\
 & & \mathbf{D}(\boldsymbol{\lambda}_2) &= \text{diagonal}(\boldsymbol{\lambda}_2).
 \end{aligned}$$

REFERENCIA BIBLIOGRÁFICA

- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2), 171–178. <http://www.jstor.org/stable/4615982>.
- Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3), 579–602. <https://doi.org/10.1111/1467-9868.00194>
- Azzalini, A. and Capitanio, A. (2014). *The Skew-Normal and Related Families*. New York: Cambridge University Press.
- Ferreira, C., Vilca, F., and Bolfarine, H. (2018). Diagnostics analysis for skew-normal linear regression models: Applications to a quality of life dataset. *Brazilian Journal of Probability and Statistics*, 32(3). <https://doi.org/10.1214/17-BJPS352>
- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. New York: Springer, 2nd edition.
- Naseem, I., Togneri, R., and Bennamoun, M. (2010). Linear regression for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11), 2106–2112. <https://ieeexplore.ieee.org/document/5506092>
- Rubio, F. J. and Genton, M. G. (2016). Bayesian linear regression with skew-symmetric error distributions with applications to survival analysis. *Statistics in Medicine*, 35(14), 2441–2454. <https://doi.org/10.1002/sim.6897>
- Sahu, S. K., Dey, D. K., and Branco, M. D. (2003). A new class of multivariate skew distributions with applications to bayesian regression models. *Canadian Journal of Statistics*, 31(2), 129–150. <https://doi.org/10.2307/3316064>