REVISTA DE LA FACULTAD DE INGENIERIA

www.ingenieria.uda.cl
20(2006) 27-31

Calibración bayesiana lineal bajo la distribución Skew-generalizada-normal

H. Gómez[1], J. Olivares[1]

1 Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó, Chile

E-mail: hgomez@matematica.uda.cl (H. Gómez)

Resumen

En este artículo estudiamos el modelo de calibración bayesiana lineal con errores Skew-Generalizados-Normales. Consideramos prioris Normales para el parámetro de asimetría y mostramos que para algunas subfamilias es equivalente al modelo de calibración bayesiana lineal bajo la distribución Normal, la cual fue estudiada por Hoadley (1970). Finalmente realizamos una aplicación utilizando prioris Skew-Normales para el parámetro de calibración y comparamos estos resultados con los clásicos y de Hoadley.

Palabras claves: Inferencia Bayesiana, Modelo de Calibración, Distribución Skew-Generalizada-Normal y Metropolis-Hastings.

Abstract
In this article we studied the model of linear bayesiana calibration with Skew-Generalize-Normal errors. We considered prioris Normal for the asymmetry parameter and showed that for some subfamilies he is equivalent to the model of linear bayesiana calibration under the Normal distribution, which was studied by Hoadley (1970). Finally we made an application using prioris Skew-Normal for the calibration parameter and compared these results with the classic ones and of Hoadley.

Keywords: Bayesiana Inference, Model of Calibration, Skew-Generalize-Normal Distribution and Metropolis-Hastings

1. Introducción

Un problema de gran interés práctico con el modelo de regresión es el problema de calibración. Hoadley (1970) muestra que el estimador inverso es un estimador de Bayes cuando se considera un modelo de regresión Normal con distribución a priori t-Student para el parámetro de calibración. Posteriormente Branco (1997) y Branco, Bolfarine e Iglesias (1998) trabajan el modelo de calibración lineal bajo la distribución t-Student. Una revisión de los principales resultados sobre este tema son encontrados en Brown (1993).

El artículo es desarrollado de la siguiente manera. En la sección 2 entregamos la distribución Skew-Generalizada-Normal y algunas subfamilias importantes, además mostramos un resultado crucial cuando se aborda un modelo de calibración lineal con errores distribuidos con esas familias. En la sección 3 planteamos el modelo de calibración lineal con errores Skew-Generalizados-Normal y entregamos una solución solamente con las subfamilias. Realizamos una aplicación de este modelo con distribución a priori Skew-Normal en el parámetro de calibración en la sección 4. En la sección 5 comparamos estos resultados con los resultados clásicos y de Hoadley.

2. Distribución skew-generalizada normal

Definición: Si una variable aleatoria X tiene función de densidad a

$$\phi(x;\alpha,\beta) = 2\phi(x)\Phi\left(\frac{\alpha x}{\sqrt{1+\beta x^2}}\right), \qquad [1]$$

con $x \in R$, $\alpha \in R$ y $\beta \geq 0$, donde ϕ y Φ son al densidad $N(0,1)$ y su función de distribución, respectivamente, entonces decimos que X es una variable aleatoria Skew-Generalizada-Normal con parámetros α y β, la denotaremos como $X \sim SGN(\alpha,\beta)$.

Esta familia $\{SGN(\alpha,\beta) : \alpha \in R, \beta \geq 0\}$ es introducida por Arellano-Valle, Gómez y Quintana (2004), el cual contiene dos familias importantes.

2.2 Subfamilias Importantes

Sea $X \sim SGN(\alpha,0)$, entonces X es una variable aleatoria Skew-Normal con parámetro de asimetría α, se denota por $X \sim SN(\alpha)$ y su densidad es:

$$\phi(x;\alpha) = 2\phi(x)\Phi(\alpha x), \qquad [2]$$

donde ϕ y Φ son la densidad $N(0,1)$ y su función de distribución, respectivamente.

Esta familia $\{SN(\alpha) : \alpha \in R\}$ fue introducida por Azzalini (1985), la cual contiene la distribución Normal estándar ($SN(0)$). Algunas propiedades de esta distribución son:

$$E(X) = k\delta \qquad [3]$$

$$V(X) = 1-(k\delta)^2 \qquad [4]$$

$$\gamma_1 = \frac{1}{2}(4-\pi)\left(\frac{E^2(X)}{V(X)}\right)^{3/2} \qquad [5]$$

$$\gamma_2 = 2(\pi-3)\left(\frac{E^2(X)}{V(X)}\right)^2 \qquad [6]$$

$$M_X(t) = 2\exp\left(\frac{t^2}{2}\right)\Phi(t\delta) \qquad [7]$$

con $\delta = \alpha/\left(1+\alpha^2\right)^{1/2}$ y $k = \sqrt{2/\pi}$.

Comentario 2.1: Henze (1986) entrega una expresión general para los momentos impares de X; los momentos pares coinciden con el caso Normal, ya que X^2 tiene distribución Chi-cuadrado con un grado de libertad.

Por otro lado, sea $X \sim SGN(\alpha, \alpha^2)$, entonces X es una variable aleatoria Skew-Normal-Curvada con parámetro de asimetría α, la denotaremos por $X \sim SNC(\alpha)$ y su densidad es:

$$\phi(x;\alpha,\alpha^2) = 2\phi(x)\Phi\left(\frac{\alpha x}{\sqrt{1+\alpha^2 x^2}}\right) \quad [8]$$

con $x \in R$ y $\alpha \in R$, donde ϕ y Φ son la densidad $N(0,1)$ y su función de distribución, respectivamente.

Comentario 2.2: Las familias $\{SN(\alpha) : \alpha \in R\}$ y $\{SNC(\alpha) : \alpha \in R\}$ son identificables y ambas contienen a la distribución Normal estándar ($SN(0)$ y $SNC(0)$).

Proposición 2.1: Sean

$$X|\theta,\mu,\sigma^2,\delta^2 \sim SN(\mu,\sigma^2,\theta) \quad [9]$$

$$X|\theta,\mu,\sigma^2,\delta^2 \sim SNC(\mu,\sigma^2,\theta) \quad [10]$$

$$\theta|\delta^2 \sim N(0,\delta^2) \quad [11]$$

$$X \amalg \delta^2 | \theta,\mu,\sigma^2 \quad [12]$$

Y

$$\theta \amalg (\mu,\sigma^2) | \delta^2 \quad [13]$$

Entonces

$$X|\mu,\sigma^2 \sim N(\mu,\sigma^2) \quad [14]$$

3. Modelo de calibración lineal

Sea $y_i = \delta + \gamma x_i + \varepsilon_i$ para todo $i = 0,1,\ldots,n$ donde $\varepsilon_i | x_0,\delta,\gamma,\sigma^2,\alpha_i,\beta_i \sim SGN(0,\sigma^2,\alpha_i,\beta_i)$ independientes con $x = (x_1,\ldots,x_n)$, $y = (y_1,\ldots,y_n)$, $\alpha = (\alpha_0,\ldots,\alpha_n)$, $\beta = (\beta_0,\ldots,\beta_n)$, $x_0 \amalg \alpha,\beta,\gamma,\sigma^2$ y $x_0 \amalg y_1,\ldots,y_n | \alpha,\beta,\sigma^2,\delta,\gamma$.

El objetivo principal, es hacer inferencias sobre x_0, a partir de x, y e y_0. Esto es, determinar la distribución a posteriori de x_0, la cual es dada por

$$p(x_0|y_0,y,x) \propto p(y_0,y|x_0,x)\,p(x_0|x). \quad [15]$$

Donde

$$p(y_0,y|x_0,x) = \int_{\alpha,\beta,\sigma^2,\delta,\gamma} p(y_0,y|x_0,x,\alpha,\beta,\sigma^2,\delta,\gamma)\,\pi(\alpha,\beta,\sigma^2,\delta,\gamma)\,d\alpha\, d\beta\, d\sigma^2 d\delta\, d\gamma$$

con $(\alpha,\beta,\sigma^2) \amalg \delta,\gamma$ y $\pi(\delta,\gamma,\sigma^2) \propto 1/\sigma^2$. Utilizando las subfamilias dadas en (2), (8) y prioris para $\alpha_i \sim N(0,\theta_i^2)$, tenemos que la integral con respecto a α se hace trivial gracias a la proposición. Luego se tiene

$$p(y_0,y|x_0,x) \propto \int \frac{1}{\sigma^{n+3}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=0}^{n}(y_i-\delta-\gamma x_i)^2\right\} d\delta\, d\gamma\, d\sigma^2$$

la cual fue calculada por Hoadley (1970).

$$p(y_0,y|x_0,x) \propto \frac{(1+n+x_0^2)^{(n-2)/2}}{\left(1+n+R\hat{x}_c^2+(F/(n-2)+1)(x_0-R\hat{x}_c)^2\right)^{(n-1)/2}}$$

con

$$F = \frac{s_X^2 \hat{\gamma}^2}{S^2},$$

$$R = \frac{F}{F + n - 2},$$

$$\hat{x}_c = \frac{y_0 - \hat{\delta}}{\hat{\gamma}}$$

$$s_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Y

$$S^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{\delta} - \hat{\gamma}x_i\right)^2$$

Comentario 3.1: Nosotros hemos resuelto el problema sólo para las subfamilias, falta considerar el caso general descrito en (1). En la siguiente sección realizamos una aplicación considerando prioris Skew-Normal para el parámetro de calibración.

4. Aplicación

Dos mediciones de grasa corporal son consideradas para evaluar los porcentajes de grasa en el cuerpo: DEXA (Dual Energy X-ray Absortiometry) y la ecuación de Durnim. La Tabla siguiente presenta valores de estas mediciones para un grupo de $n = 24$ mujeres con edades entre $16 - 19$ años.

El principal objetivo es predecir el porcentaje de la grasa corporal por DEXA, denotado por x_0, a partir de una medida obtenida usando la ecuación de Durnim, denotado por y_0.

Tabla 1: Mediciones de grasa corporal

Mujer	DEXA (X)	Durnim (y)	Mujer	DEXA (X)	Durnim (y)
1	28.6	28.34	13	24.7	27.51
2	24.8	25.68	14	32.7	31.25
3	30.9	25.26	15	34.2	31.16
4	32.2	29.12	16	28.9	32.86
5	35.6	35.56	17	23.8	25.96
6	18.4	18.86	18	25.6	22.60
7	37.9	35.06	19	34.9	24.53
8	27.0	32.77	20	30.7	26.88
9	37.6	34.84	21	36.7	31.44
10	40.8	33.84	22	42.4	33.84
11	33.7	30.87	23	43.5	35.56
12	20.6	18.63	24	29.7	25.12

Dos nuevas mujeres fueron observadas con medidas de Durnim, dadas por, 26.98 y 33.25. La primera es considerada delgada y la segunda obesa. Las distribuciones a priori para x_0 son $SN(32,70,-2)$ y $SN(32,70,2)$, respectivamente. Se tomo en cuenta la información relacionada con la contextura de la mujer.

Las muestras fueron generadas usando el algoritmo de Metropolis-Hastings que fue implementado en S-Plus.

5. Resultados

Estimaciones puntuales e intercalares (0.95) para DEXA (x_0).

Tabla 2: Comparación en las estimaciones

y_0 **\ E. puntual**	**Clásica**	**Hoadley**	**Skew-Normal**
26.98 Delgada	28.11 (21.22,41.23)	29.25 (19.92,38.58)	25.53 (15.54,35.23)
33.25 Obesa	38.36 (29.04,56.17)	36.04 (24.58,47.50)	36.21 (27.25,47.04)

Con $V(x_0|26.98,x,y) = 26.93$ y

$$V(x_0|33.25,x,y) = 23.94.$$

6. Interpretación

Para propósitos prácticos, la priori Skew-Normal resulta más interesante porque la estimación pertenece a la misma categoría. Es decir, si una mujer es considerada delgada (obesa) usando la medida Durnim ella es también considerada delgada (obesa) usando la medida DEXA. En la escala de DEXA una mujer es considerada delgada si $X < 26$ ($X > 36$), tal estimación 25.53 (36.21) está en la categoría.

7. Referencias

Arellano-Valle, R. B., Gómez, H. W. and Quintana F. A. (2004). A New Class of Skew-Normal Distributions. Communications in Statistics. 33. 1465-1480.

Azzalini, A. (1985). A Class of Distributions which Incluye the Normal Ones. Scand. J. Statist. 12. 171-178.

Branco, M. (1997). Calibraçao: Uma Abordagem Bayesiana. Tese de doutorado IME-USP.

Branco, M., Bolfarine H. and Iglesias P. (1998). Bayesian Calibration Under a Student-t Model. Computacional Statistics. 13. 319-338.

Brown, P. (1993). Measurement, Regresión and Calibration. Oxford University Press.

Hoadley, B. (1970). A Bayesian Look at Inverse Linear Regresión. JASA. 65. 356-369.