Análisis de datos con Métodos Bayesianos

Instituto Bayesiano para la Investigación en Desarrollo (BayesGroup.org)

Rolando Gonzales
rgonzales@bayesgroup.org

21 de octubre de 2015

Contenido del curso

Este es un curso corto e introductorio sobre Análisis de datos con métodos Bayesianos.

- Un poco de historia como introducción al curso.
- Homogeneización.
- Introducción a la inferencia Bayesiana.
- Introducción a la econometría Bayesiana.
- Aplicaciones.

Ejercicios prácticos con datos del anuario estadístico del municipio de La Paz 2013 y otros datos relevantes.

Introducción

- La Econometría puede definirse como la unificación de la teoría económica, la estadística y las matemáticas.[1]
- El objetivo de la econometría es la **modelización empírica** de los fenómenos económicos.[2]

[1] Frisch, R. , Note on the term "Econometrics", Econometrica, Volume 4, Issue 1 (jan., 1936), 95. Frisch se atribuye el uso temprano del término en la forma de la palabra francesa *économétrie* en su artículo "Sur un problème d'économie pure", pero reconoce que existe un uso anterior de la palabra, en la forma alemana " Oekonometrie"por Pawel Ciompa en su libro de 1910.

[2] La **modelización empírica** es la descripción parsimoniosa de un fenómeno estocástico observable utilizando **modelos estadísticos**.

Modelo de Regresión Lineal

Función de regresión (definición). *La función de regresión se define como la media condicional de Y dado $X = x$, interpretada como una función de x:*
$\mathbb{E}\left(Y|X = x\right) = h\left(x\right), x \in \mathbb{R}_X$

Función de regresión lineal (definición). *Si la media condicional $h\left(x\right)$ toma la forma,*
$\mathbb{E}\left(Y|X = x\right) = \beta_0 + \beta_1 x$, $Var\left(Y|X = x\right) = \sigma^2$, $x \in \mathbb{R}$,
$\beta_0 = \left(\mu_1 - \beta_1\mu_2\right) \in \mathbb{R}$, $\beta_1 = \frac{\sigma_{12}}{\sigma 22} \in \mathbb{R}$, $\sigma^2 = \left(\sigma_{11} - \frac{\sigma_{12}^2}{\sigma 22}\right) \in \mathbb{R}_+$
la función de regresión se dice lineal *en x.*

Modelo de Regresión Lineal Gaussiano Bivariante

La anterior definición viene de considerar una distribución gaussiana bivariante de la forma,

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \right)$$

con $\mu_1 = E\left(Y\right)$, $\mu_2 = E\left(X\right)$, $\sigma_{11} = Var(Y)$, $\sigma_{22} = Var(X)$, $\sigma_{12} = Cov(X, Y)$. Las distribuciones condicionales y marginales toman la forma:

$(Y|X = x) \sim \mathcal{N}\left(\beta_0 + \beta_1 x, \sigma^2\right), x \in \mathbb{R}, X \sim \mathcal{N}\left(\mu_2, \sigma_{22}\right)$

$\beta_0 = \mu_1 - \beta_1 \mu_2, \beta_1 = \frac{\sigma_{12}}{\sigma 22}$, $\sigma^2 = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma 22}$.

Nótese que pueden realizarse extensiones al caso gaussiano multivariante.

Modelo de Regresión Lineal Gaussiano Multivariante

El Modelo de Regresión Lineal Gaussiano (MRLG) multivariante será

(i) DGP estadístico: $Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_j$

(ii) Modelo de probabilidad:

$$\Phi = \left\{ f\left(y_i | x_{1i}, ..., x_{ki}; \boldsymbol{\theta}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{\left(y_t - \beta_0 - \beta_1 x_{1i} - \cdots - \beta_k x_{ki}\right)^2}{2\sigma^2} \right\}, \boldsymbol{\theta} \in \boldsymbol{\Theta}, y_i, x_{ki} \in \mathbb{R} \right\}$$

$\beta := (\beta_1, ..., \beta_n)$

$\boldsymbol{\theta} := \left(\beta, \sigma^2\right), \boldsymbol{\Theta} := \mathbb{R}^n \times \mathbb{R}_+$

(iii) Modelo muestral: $(Y_1, ..., Y_n)$ es una muestra de $f\left(y_i | x_{1i}, ..., x_{ki}; \boldsymbol{\theta}\right)$

MRLG multivariante: Estimación clásica

Sea un modelo de regresión,

$$\underset{(n\times 1)}{\mathbf{y}} = \underbrace{\underset{(n\times k)}{\mathbf{X}}\underset{(k\times 1)}{\beta}}_{(n\times 1)} + \underset{(n\times 1)}{\varepsilon}$$

en el que,

$$\underset{(n\times 1)}{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \underset{(n\times 1)}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \underset{(n\times k)}{\mathbf{X}} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ x_{21} & \cdots & x_{2k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix},$$

MRLG multivariante: Estimación clásica

se busca obtener estimadores OLS ($\widehat{\beta}$) de,

$$\underset{(k\times 1)}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}.$$

Defínase la suma de residuos al cuadrado (SSR),

$$SSR(\widehat{\beta}) = (\mathbf{y} - \mathbf{X}\widehat{\beta})'(\mathbf{y} - \mathbf{X}\widehat{\beta})$$

MRLG multivariante: Estimación clásica

El estimador OLS $\widehat{\beta}$ es aquel que minimiza esta función,

$$\widehat{\beta} \equiv \arg\min_{\widehat{\beta}} SSR(\widehat{\beta})$$

matricialmente este resultado se obtiene con,

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

MRLG: Estimación clásica - Máxima Verosimilitud

Si $\varepsilon|\mathbf{X} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$, y dado que $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, entonces,

$$\mathbf{y}|\mathbf{X} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$$

La densidad condicional de $\mathbf{y}$ dado $\mathbf{X}$ será,

$$f(\mathbf{y}|\mathbf{X}) = \left(2\pi\sigma^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)\right\}$$

y se tendrá la función de verosimilitud,

$$\mathcal{L}\left(\beta, \sigma^2 \middle| \mathbf{X}, \mathbf{y}\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)\right]$$

y los estimadores de máxima verosimilitud de $\left(\beta, \sigma^2\right)$ serán los que maximizen esta función. Para el caso de β se tiene que,

$$\widehat{\beta}_{MV} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

Estimación Bayesiana

La estimación Bayesiana del modelo de regresión lineal gaussiano complementa el análisis máximo verosímil con información adicional—contenida en una distribución prior— a través del Teorema de Bayes,

$$\text{posterior} \propto \text{prior} \times \text{verosimilitud}$$

"Econometrics Should Always and Everywhere Be Bayesian"

Christopher Albert ("Chris") Sims

2011 Nobel Prize in Economic Sciences

Estimación Bayesiana: priors no informativos

Priors no informativos: Si se emplean priors no informativos $\mathbb{P}(\beta) \propto c$ y $\mathbb{P}(\sigma^2) \propto \sigma^{-1}$ en el soporte $[-\infty : \infty]$ y $[0 : \infty]$, respectivamente, la densidad posterior conjunta será,

$$\mathbb{P}(\beta, \sigma^2) \propto \mathcal{L}\left(\beta, \sigma^2 \middle| \mathbf{X}, \mathbf{y}\right)\mathbb{P}(\beta)\mathbb{P}(\sigma^2)$$

$$\propto \sigma^{-n-1} \exp\left[-\frac{1}{2\sigma^2}(\hat{\sigma}^2(n-k) + (\beta - \hat{\beta})'\mathbf{X}'\mathbf{X}(\beta - \hat{\beta}))\right],$$

y los estimadores Bayesianos serán iguales a los estimadores clásicos máximo verosímiles.

Estimación Bayesiana MCMC con priors informativos

Retomando $X = [\mathbf{1}_n \;\; (x_1, ..., x_n)']$, $\beta = [\beta_0 \;\; \beta_1 \cdots \beta_k]$, $y = (y_1, ..., y_n)'$,

$$y \sim \mathcal{N}_n(X\beta, \sigma^2 I_n),$$

para realizar una estimación Bayesiana MCMC con priors informativos puede utilizarse las distribuciones prior,

$$\beta \sim \mathcal{N}_k(\beta_0, B_0), \;\; \sigma^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2),$$

se tendrá por conjugados naturales,

$$\beta|\sigma^2, y \sim \mathcal{N}(\overline{\beta}, B_1), \qquad \sigma^2|\beta, y \sim \mathcal{IG}(\alpha_1/2, \delta_1/2)$$

Estimación Bayesiana MCMC con priors informativos

En $\beta|\sigma^2, y \sim \mathcal{N}(\overline{\beta}, B_1)$ y $\sigma^2|\beta, y \sim \mathcal{IG}(\alpha_1/2, \delta_1/2)$,

$$\begin{aligned} B_1 &= [s^{-2}X'X + B_0^{-1}]^{-1}, \\ \overline{\beta} &= B_1[\sigma^{-2}X'y + B_0^{-1}\beta_0], \\ \alpha_1 &= \alpha_0 + n, \\ \delta_1 &= \delta_0 + (y - X\beta)'(y - X\beta), \end{aligned}$$

y Markov Chain Monte Carlo (MCMC) puede emplearse para encontrar la distribución posterior de (θ, σ^2) utilizando un algoritmo de Gibbs en bloques.

Estimación Bayesiana: Gibbs Sampling

(a) Sea $\sigma^{2(0)}$ un valor inicial de σ^2.
(b) En la g-iteración,

$$\beta^{(g)} \sim \mathcal{N}_k(\bar{\beta}^{(g)}, B_1^{(g)}),$$
$$\sigma^{2(g)} \sim \mathcal{IG}(\alpha_1/2, \delta^{(g)}/2),$$

con,

$$B_1^{(g)} = [\sigma^{-2(g-1)} X'X + B_0^{-1}]^{-1},$$
$$\overline{\beta}^{(g)} = B_1^{(g)}[\sigma^{-2(g-1)} X'y + B_0^{-1}\beta_0],$$
$$\delta^{(g)} = \delta_0 + (y - X\beta^{(g)})'(y - X\beta^{(g)}).$$

Con una repetición del algoritmo hasta $g = \mathcal{B} + \mathcal{G}$ —para $\mathcal{B}$ el burn-in— es posible simular la distribución posterior de β y σ^2.

Ejemplo: Estimación de parámetros en una función CES

Una función de producción con una $\sigma \in [0, \infty)$-elasticidad de sustitución constante (CES) entre capital (k) y trabajo (l) tendrá una función de producción,

$$q = A\left[\rho l^{\frac{\sigma-1}{\sigma}} + (1-\rho)k^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}},$$

para q el producto real, A la productividad, $\rho \in [0, 1]$ un parámetro de distribución. La primera condición del problema de optimización de la firma es,

$$pA\frac{\sigma}{\sigma-1}\left(\rho\ell^{\frac{\sigma-1}{\sigma}} + (1-\rho)\,k^{\frac{1}{\sigma-1}}\right)\rho\frac{\sigma-1}{\sigma}\ell^{-\frac{1}{\sigma}} = w,$$

por lo que,

$$\ell = \left(\frac{p}{w}\right)^{\sigma}\rho^{\sigma}A^{\sigma}\left(\rho\ell^{\frac{\sigma-1}{\sigma}} + (1-\rho)\,k^{\frac{1}{\sigma-1}}\right)^{\frac{\sigma}{\sigma-1}},$$

Ejemplo: Estimación de parámetros en una función CES

reemplazando la función CES en la ecuación de demanda de trabajo se tiene,

$$\ell = \left(\frac{p}{w}\right)^{\sigma} \rho^{\sigma} A^{\sigma-1} q,$$

por lo que,

$$\frac{q}{\ell} = \left(\frac{w}{p}\right)^{\sigma} (\rho)^{-\sigma} A^{1-\sigma}.$$

Esta ecuación puede ser log-linealizada,

$$y = \theta_0 + \theta_1 x,$$

para $y = \ln(q/\ell)$, $x = \ln(w/p)$, $\theta_0 = -\sigma \ln \rho + (1-\sigma) \ln A$ y $\theta_1 := \sigma$. Añadiendo un término de error ϵ, la anterior ecuación se vuelve un modelo de regresión lineal.

Ejemplo: Estimación de parámetros en una función CES

Sea $X = [\mathbf{1}_n \ \ (x_1, ..., x_t)']$, $\theta = [\theta_0 \ \ \theta_1]$, $y = (y_1, ..., y_t)'$, $y \sim \mathcal{N}(X\theta, s^2 I_n)$, con los priors,

$$\theta \sim \mathcal{N}(\theta_0, B_0), \quad s^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2).$$

Un estimador Bayesiano de la elasticidad de sustitución capital-trabajo será,

$$\hat{\theta}|s^2, y \sim \mathcal{N}(\bar{\theta}, B_1),$$

con,

$$B_1 = [s^{-2}X'X + B_0^{-1}]^{-1},$$
$$\bar{\theta} = B_1[s^{-2}X'y + B_0^{-1}\theta_0],$$

que puede calcularse empleando MCMC.