Regresión Clásica Versus Regresión Bayesiana

July 12, 2020

Suponemos una cantidad n de pares de observaciones (Y, X) en base a las que queremos analizar a Y como función de X.

El **Enfoque Clásico** postula **parámetros fijos** y target Y aleatoria para el modelo:

$Y \sim N(\alpha + \beta X, \sigma)$

donde

α, β, σ son parámetros fijos y desconocidos

El enfoque consigue estimaciones (realizaciones) de los estimadores

$\hat{\alpha}(Y, X) \sim N(\alpha, \propto \frac{\sigma}{\sqrt{n}})$
$\hat{\beta}(Y, X) \sim N(\beta, \propto \frac{\sigma}{S_X \sqrt{n}})$
$\frac{\hat{\sigma}(Y,X)}{\sigma} \sim \chi^2_{n-2}$

que son variables aleatorias con distribución inducida por la distribución de la target Y(si n es chico) o por el Teorema Central del Límite (si n es grande).
De este enfoque se desprenden:

- Estimaciones puntuales (realización del estimador)
- Intervalos de confianza (basado en la distribución del estadístico)
- Test de hipótesis y p-valor (basados en la distribución del estadístico)

Ventaja del enfoque: Pocos supuestos. Incluso la normalidad se puede eliminar y todo sigue valiendo, si n es razonablemente grande.

Contra del enfoque: La información que se obtiene del paramétro es escasa e indirecta. Por ejemplo, para calcular el p-valor del parámetro β debemos suponer un valor poblacional igual a 0 y computar la probabilidad (usando la distribución del estimador, bajo la hipótesis de $\beta = 0$) de observar una realización tanto o más extrema que la observada.

El **Enfoque Bayesiano** propone **parámetros aleatorios** para el mismo modelo:

$Y \sim N(\alpha + \beta x, \sigma)$

donde

$\alpha \sim N(0, \tau_\alpha)$ con prior (arbitrariamente) normal
$\beta \sim N(0, \tau_\beta)$ con prior (arbitrariamente) normal
$\sigma \sim Exp(\lambda)$ con prior (arbitrariamente) exponencial
τ_α es fijo y propuesto por el investigador
τ_β es fijo y propuesto por el investigador
λ es fijo y propuesto por el investigador

Este enfoque consigue distribuciones a posteriori de los parámetros cuyas distribuciones son inducidas por la distribución de Y y por las distribuciones a priori de los parámetros. Las distribuciones a posteriori son:

$\alpha \mid Y, X, \tau_\alpha \sim \mathrm{P}_\alpha$ la posterior de α
$\beta \mid Y, X, \tau_\beta \sim \mathrm{P}_\beta$ la posterior de β
$\sigma \mid Y, X, \lambda \sim \mathrm{P}_\sigma$ la posterior de σ

De este enfoque se desprenden:

- Estimaciones puntuales (basado en la distribución a posteriori)
- Intervalos de credibilidad (basado en la distribución a posteriori)

Contra del enfoque: Mochos supuestos y muy fuertes.

Ventaja del enfoque: La información que se obtiene del paramétro es muy rica. Por ejemṕlo, la posterior del parámetro β brinda una medida directa de la certeza de los posibles valores de β, condicional a las observaciones.

Ejemplo

Veamos, mediante un ejemplo práctico, una comparación de ambos enfoques. Tratamos de entender la relación entre el perímetro cefálico (target Y)de bebes nacidos con bajo peso (*headcirc*) y la edad gestacional (X=*gestage*). Datos publicados en Leviton, Fenton, Kuban, y Pagano [1991], tratados en el libro de Pagano et al. [2000].

En nuestro modelo:

$Y \sim N(\alpha + \beta x, \sigma)$

donde

$\alpha \sim N(0, \tau_\alpha = 100)$ con prior (arbitrariamente) normal
$\beta \sim N(0, \tau_\beta = 100)$ con prior (arbitrariamente) normal
$\sigma \sim Exp(\lambda = 0.1)$ con prior (arbitrariamente) exponencial

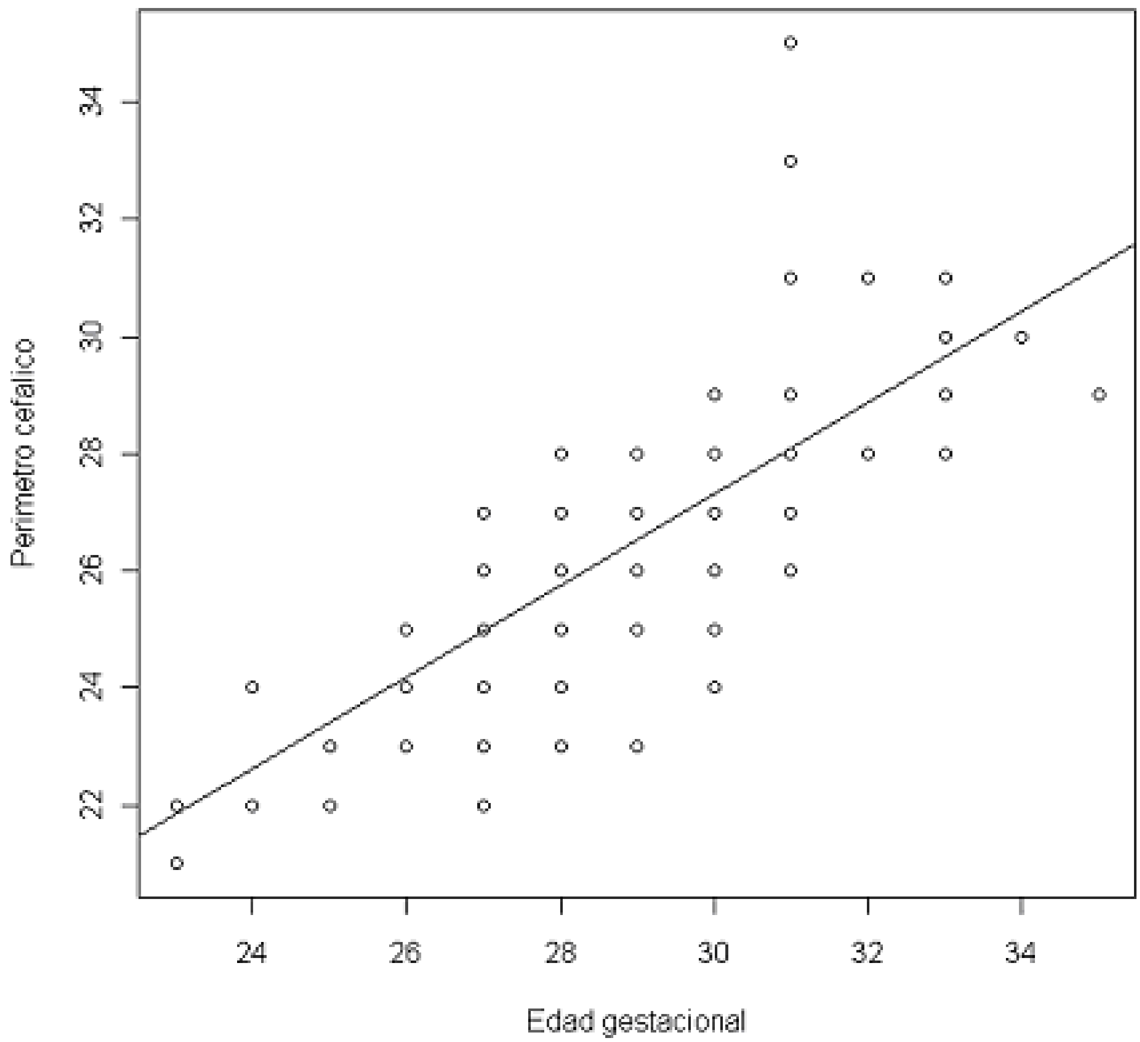

Figure 1: Perímetro Cefálico vs. Edad Gestacional

La Distribución a Posteriori en Regresión Bayesiana

En resumen, en nuestro ejemplo, la distribución a posteriori queda:

$posterior(\alpha, \beta, \sigma \mid Y, X) \propto L(Y, X, \alpha, \beta, \sigma) * p(\alpha) * p(\beta) * p(\sigma)$

La Magia: Monte Carlo Markov Chain

El Algoritmo Metrópolis

Veamos con un ejemplo sencillo el principio fundamental sobre el que se basa la aplicación moderna de la inferencia bayesiana, la **estimación numérica de la distribución a posteriori**.

Modelo de Posición

Dada una muestra $(Y_1 = 2, Y_2 = 3, Y_3 = 4)$, sea

$Y \sim N(\alpha, \sigma = 1)$

donde

$\alpha \sim N(0, 1)$ con prior (arbitrariamente) normal

L a posterior es

$posterior(\alpha) \propto L(Y, \alpha) * p(\alpha) = \prod_i \phi(Y_i - \alpha, 1) * \phi(\alpha, 1)$

En este caso particular, como la normal (likelihood) es conjugada de la normal (prior), se puede calcular analíticamente la distribución:

$posterior(\alpha) = N(\frac{n\overline{Y}}{1+n}, \sqrt{\frac{1}{1+n}})$

El Algoritmo Metrópolis

1. Genero un α_0 aleatorio inicial
2. Sea $\alpha_{viejo} = \alpha_0$
3. Genero un nuevo $\alpha_{nuevo} \sim normal(\alpha_{viejo})$ aleatorio centrado en α_{viejo}
4. Calculo $r = \frac{posterior(\alpha_{nuevo})}{posterior(\alpha_{viejo})} = \frac{\prod_i \phi(Y_i - \alpha_{nuevo}, 1) * \phi(\alpha_{nuevo}, 1)}{\prod_i \phi(Y_i - \alpha_{viejo}, 1) * \phi(\alpha_{viejo}, 1)}$
5. Genero $u \sim U(0, 1)$
6. Si $u < r$ entonces $\alpha_{nuevo} = \alpha_{viejo}$ y guardo α_{nuevo}
7. Vuelovo a 3 hasta alcanzar la cantidad deseada de muestras

Este es un algoritmo (tipo aceptación/rechazo) de Monte Carlo (basado en la generación de números pseudo aleatorios) Markov Chain de grado 1 (cada nyuevo alfa sólo depende del anterior) en el que la distribución de la sucesión de alfas converge a la distribución a posteriori deseada.

El algoritmo sólo requiere contar con una función **proporcional** a la distribución a posteriori que sea **numéricamente evaluale**.