Estadística Bayesiana

Clase 13: Comparación de modelos

Isabel Cristina Ramírez Guevara

Escuela de Estadística
Universidad Nacional de Colombia, Sede Medellín

Medellín, 30 de septiembre de 2020

Comparación de modelos

Se tiene una muestra aleatoria $y_1, \cdots, y_n$ y se ajusta el modelo:

$$p(\mathbf{y}|\theta) = \prod_{i=1}^{n} p(y_i|\theta)$$
$$\log(p(\mathbf{y}|\theta)) = \log\left(\prod_{i=1}^{n} p(y_i|\theta)\right)$$

Suponga que se ajustan varios modelos, se selecciona el modelo con mayor, $\log(p(\mathbf{y}|\theta))$ pues tendrá una mayor probabilidad posterior. En la práctica el parámetro θ es desconocido, por lo tanto no se puede calcular $\log(p(\mathbf{y}|\theta))$.

Comparación de modelos

Por esta razón se trabaja con la distribución posterior $p(\theta|\mathbf{y})$ y se calcula la exactitud de la predicción del modelo ajustado así:

$$\log \prod_{i=1}^{n} p_{\text{post}}(y_i) = \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{j=1}^{S} p(y_i|\theta^j) \right)$$

donde θ^j son las simulaciones posteriores, $j = 1, \cdots, S$.

A continuación vamos a ver algunos medidas que nos sirven para comparar modelos.

Deviance Information Criterion (DIC)

En la mayoría de la literatura estadística sobre exactitud de la predicción se utiliza como estimador de θ el estimador de máxima verosimilitud. Sea K el número de parámetros estimados en el modelo, el AIC (criterio de información de Akaike) está dado por:

$$\mathrm{AIC} = -2\log\left(p(\mathbf{y}|\hat{\theta}_{\mathrm{MLE}})\right) + 2K$$

El DIC es una versión Bayesiana del AIC. Reemplaza el estimador de máxima verosimilitud por la media posterior de θ, esto es: $\hat{\theta}_{\mathrm{BAYES}} = E(\theta|\mathbf{y})$ y reemplaza K con un factor de corrección de sesgo.

Deviance Information Criterion (DIC)

El DIC está dado por:

$$\text{DIC} = -2\log\left(p(\mathbf{y}|\hat{\theta}_{\text{BAYES}})\right) + 2\text{p}_{\text{DIC}}$$

donde p_{DIC} es el número efectivo de parámetros y está definido como:

$$\text{p}_{\text{DIC}} = 2\log\left(p(\mathbf{y}|\hat{\theta}_{\text{BAYES}}) - E_{\text{post}}\left(\log p(\mathbf{y}|\theta)\right)\right)$$

$$\hat{\text{p}}_{\text{DIC}} = 2\log\left(p(\mathbf{y}|\hat{\theta}_{\text{BAYES}}) - \frac{1}{S}\sum_{j=1}^{S}\log p(\mathbf{y}|\theta^{j})\right)$$

Deviance Information Criterion (DIC)

Se prefiere el modelo con menor DIC, se recomienda:

- Diferencias de más de 10 en el DIC permiten descartar el modelo con el DIC más alto.
- Diferencias entre 5 y 10 pueden considerarse sustanciales.
- Cuando la diferencia es menores que cinco y los modelos hacen inferencias muy diferentes, reportar sólo el modelo con el menor DIC puede llevar a conclusiones erróneas.

Deviance Information Criterion (DIC) - Ejemplo

Suponga que se tiene una m.a. de tamaño n una variable Normal(θ,1) y $p(\theta) \propto 1$. En este caso los cálculos para el AIC son:

$$\begin{aligned}
p(\mathbf{y}|\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(y_i-\theta)^2\right] \\
&= \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^2\right] \\
\log p(\mathbf{y}|\theta) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^2
\end{aligned}$$

Deviance Information Criterion (DIC) - Ejemplo

Se tiene que $\hat{\theta}_{\mathrm{MLE}} = \bar{y}$ por lo tanto:

$$\begin{aligned}\log p(\mathbf{y}|\hat{\theta}_{\mathrm{MLE}}) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i - \bar{y})^2 \\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}(n-1)s_y^2\end{aligned}$$

donde $s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$. Como solamente se está estimando un parámetro, $K = 1$, por lo tanto:

$$\widehat{\mathrm{AIC}} = -2\left(-\frac{n}{2}\log(2\pi) - \frac{1}{2}(n-1)s_y^2\right) + 2.$$

Deviance Information Criterion (DIC) - Ejemplo

Para el calculo del DIC tenemos que la distribución a priori es $p(\theta) \propto 1$ por lo tanto:

$$
\begin{aligned}
p(\theta|\mathbf{y}) &\propto p(\mathbf{y}|\theta)p(\theta) \\
&\propto \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^2\right] * 1 \\
&= \exp\left[-\frac{n}{2}(\bar{y}-\theta)^2\right] \\
\theta|\mathbf{y} &\sim \mathrm{N}\left(\bar{y}, \frac{1}{n}\right)
\end{aligned}
$$

Se tiene $\mathrm{E}(\theta|\mathbf{y}) = \bar{y} = \hat{\theta}_{\mathrm{MLE}} = \hat{\theta}_{\mathrm{BAYES}}$ entonces $\log p(\mathbf{y}|\hat{\theta}_{\mathrm{BAYES}}) = \log p(\mathbf{y}|\hat{\theta}_{\mathrm{MLE}})$.

Deviance Information Criterion (DIC) - Ejemplo

Para calcular el segundo término de $\mathrm{p}_{\mathrm{DIC}}$ iniciamos con:

$$
\begin{aligned}
\log p(\mathbf{y}|\theta) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta)^2 \\
&= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\left[n(\bar{y} - \theta)^2 + (n-1)s_y^2\right].
\end{aligned}
$$

Ahora queremos calcular:

$$
E_{\text{post}}(\log p(\mathbf{y}|\theta)) = E_{\text{post}}\left(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\left[n(\bar{y} - \theta)^2 + (n-1)s_y^2\right]\right)
$$

Deviance Information Criterion (DIC) - Ejemplo

Para calcular el segundo término de $\mathrm{p}_{\mathrm{DIC}}$ iniciamos con:

$$
\begin{aligned}
\log p(\mathbf{y}|\theta) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^2 \\
&= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\left[n(\bar{y}-\theta)^2 + (n-1)s_y^2\right].
\end{aligned}
$$

Ahora queremos calcular:

$$
E_{\text{post}}(\log p(\mathbf{y}|\theta)) = E_{\text{post}}\left(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\left[n(\bar{y}-\theta)^2 + (n-1)s_y^2\right]\right)
$$

$$
= -\frac{n}{2}\log(2\pi) - \frac{1}{2}nE_{\text{post}}\left((\bar{y}-\theta)^2\right) - \frac{1}{2}(n-1)s_y^2
$$

Deviance Information Criterion (DIC) - Ejemplo

Para calcular el segundo término de $\mathrm{p}_{\mathrm{DIC}}$ iniciamos con:

$$
\begin{aligned}
\log p(\mathbf{y}|\theta) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^2 \\
&= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\left[n(\bar{y}-\theta)^2 + (n-1)s_y^2\right].
\end{aligned}
$$

Ahora queremos calcular:

$$
E_{\text{post}}(\log p(\mathbf{y}|\theta)) = E_{\text{post}}\left(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\left[n(\bar{y}-\theta)^2 + (n-1)s_y^2\right]\right)
$$

$$
= -\frac{n}{2}\log(2\pi) - \frac{1}{2}nE_{\text{post}}\left((\bar{y}-\theta)^2\right) - \frac{1}{2}(n-1)s_y^2
$$

$$
= -\frac{n}{2}\log(2\pi) - \frac{1}{2}n\left(\frac{1}{n}\right) - \frac{1}{2}(n-1)s_y^2
$$

Deviance Information Criterion (DIC) - Ejemplo

Para calcular el segundo término de p_{DIC} iniciamos con:

$$\begin{aligned}\log p(\mathbf{y}|\theta) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^2\\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\left[n(\bar{y}-\theta)^2 + (n-1)s_y^2\right].\end{aligned}$$

Ahora queremos calcular:

$$E_{\text{post}}(\log p(\mathbf{y}|\theta)) = E_{\text{post}}\left(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\left[n(\bar{y}-\theta)^2 + (n-1)s_y^2\right]\right)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}nE_{\text{post}}\left((\bar{y}-\theta)^2\right) - \frac{1}{2}(n-1)s_y^2$$

$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}n\left(\frac{1}{n}\right) - \frac{1}{2}(n-1)s_y^2$$

$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\left(1 + (n-1)s_y^2\right)$$

Deviance Information Criterion (DIC) - Ejemplo

Por lo tanto:

$$\begin{aligned}\hat{\mathrm{p}}_{\mathrm{DIC}} &= 2\left[-\frac{n}{2}\log(2\pi) - \frac{1}{2}(n-1)s_y^2 - \left(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\left(1+(n-1)s_y^2\right)\right)\right] \\ &= 1\end{aligned}$$

Entonces:

$$\widehat{\mathrm{DIC}} = -2\left(-\frac{n}{2}\log(2\pi) - \frac{1}{2}(n-1)s_y^2\right) + 2$$

Para este ejemplo $\widehat{\mathrm{DIC}} = \widehat{\mathrm{AIC}}$.

Deviance Information Criterion (DIC) - Ejemplo

Suponga el mismo ejemplo anterior pero con distribución a priori informativa $\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$. Si la precisión es $m = \frac{1}{\tau_0^2}$ se tiene que $\theta|\mathbf{y} \sim \mathrm{N}\left(\frac{m\mu_0 + n\bar{y}}{m+n}, \frac{1}{m+n}\right)$.

Adicionar información a priori no afecta el estimador de máxima verosimilitud por lo tanto el $\widehat{\mathrm{AIC}}$ es igual al ejemplo anterior.

La media posterior es $\hat{\theta}_{\mathrm{BAYES}} = \frac{m\mu_0 + n\bar{y}}{m+n}$, entonces del primer término del DIC es:

$$
\begin{aligned}
\log p(\mathbf{y}|\hat{\theta}_{\mathrm{BAYES}}) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}(n-1)s_y^2 - \frac{1}{2}n(\bar{y} - \hat{\theta}_{\mathrm{BAYES}})^2 \\
&= -\frac{n}{2}\log(2\pi) - \frac{1}{2}(n-1)s_y^2 - \frac{1}{2}n\left(\frac{m}{m+n}\right)^2(\bar{y} - \mu_0)^2
\end{aligned}
$$

Deviance Information Criterion (DIC) - Ejemplo

Después de realizar algún álgebra se obtiene: $\hat{\mathrm{p}}_{\mathrm{DIC}} = \frac{n}{m+n}$, la cual hace sentido: la distribución a priori plana corresponde a $m = 0$, entonces el $\hat{\mathrm{p}}_{\mathrm{DIC}} = 1$; en el caso que m tenga valores grandes, lo cual corresponde a una a priori informativa, entonces el $\hat{\mathrm{p}}_{\mathrm{DIC}} \to 0$.

Deviance Information Criterion (DIC) - Ejemplo

Nuevamente el ejemplo del tema de regresión. Una compañía tiene máquinas dispensadoras de refrescos. Se quiere ver que variables afectan el tiempo que requiere un empleado para llenar la máquina de productos. Para esto se realiza un estudio donde se registra el tiempo de llenado, el número de artículos a organizar y la distancia caminada por el empleado. Se toma una muestra de 25 observaciones. Se ajustan tres modelos, uno con las dos covariables y los otros dos con una sola covariable. Se obtienen los siguientes resultados.

Deviance Information Criterion (DIC) - Ejemplo

Modelo 1:

$$tiempo[i] \sim dnorm(mu[i], tau)$$

$$mu[i] < -beta0 + beta1 * casos[i] + beta2 * distancia[i]$$

	2.5 %	50 %	97.5 %
β_0	0.1028	2.3430	4.6610
β_1	1.261975	1.616000	1.962025
β_2	0.00692785	0.01447000	0.02202025

$$\text{DIC} = 135.4$$

Deviance Information Criterion (DIC) - Ejemplo

Modelo 2:

$$tiempo[i] \sim dnorm(mu[i], tau)$$

$$mu[i] < -beta0 + beta1 * casos[i]$$

	2.5 %	50 %	97.5 %
β_0	0.4599675	3.3010000	6.1180000
β_1	1.918	2.177	2.430

$$\text{DIC} = 146.7$$

Deviance Information Criterion (DIC) - Ejemplo

Modelo 3:

$$tiempo[i] \sim dnorm(mu[i], tau)$$

$$mu[i] < -beta0 + beta1 * distancia[i]$$

	2.5 %	50 %	97.5 %
β_0	0.07256575	4.92100000	9.72815000
β_1	0.03324950	0.04259000	0.05177025

$$\text{DIC} = 173.7$$