Estadística Bayesiana

Clase 12: Evaluación y diagnóstico de modelos

Isabel Cristina Ramírez Guevara

Escuela de Estadística
Universidad Nacional de Colombia, Sede Medellín

Medellín, 28 de septiembre de 2020

Evaluación y diagnóstico de modelos

En general, cuando se realiza un ajuste de un modelo estadístico, se debe realizar un diagnóstico de que tan adecuado es ese ajuste.

Evaluación y diagnóstico de modelos

En general, cuando se realiza un ajuste de un modelo estadístico, se debe realizar un diagnóstico de que tan adecuado es ese ajuste. Cuando hablamos de modelo nos referimos a todos sus componentes: distribución de muestreo (verosimilitud), la distribución a priori y cualquier estructura jerárquica que se esté empleando.

Evaluación y diagnóstico de modelos

En general, cuando se realiza un ajuste de un modelo estadístico, se debe realizar un diagnóstico de que tan adecuado es ese ajuste. Cuando hablamos de modelo nos referimos a todos sus componentes: distribución de muestreo (verosimilitud), la distribución a priori y cualquier estructura jerárquica que se esté empleando. Cualquiera de estos elementos puede causar que el modelo no cumpla las expectativas para las cuales fue propuesto, y puede estar sujeto a cambio.

Evaluación y diagnóstico de modelos

En general, cuando se realiza un ajuste de un modelo estadístico, se debe realizar un diagnóstico de que tan adecuado es ese ajuste. Cuando hablamos de modelo nos referimos a todos sus componentes: distribución de muestreo (verosimilitud), la distribución a priori y cualquier estructura jerárquica que se esté empleando. Cualquiera de estos elementos puede causar que el modelo no cumpla las expectativas para las cuales fue propuesto, y puede estar sujeto a cambio. Por lo tanto, es necesario poder comparar y seleccionar entre diferentes modelos aquel (o aquellos) que mejor se ajustan al problema estudiado.

Evaluación y diagnóstico de modelos

En general, cuando se realiza un ajuste de un modelo estadístico, se debe realizar un diagnóstico de que tan adecuado es ese ajuste. Cuando hablamos de modelo nos referimos a todos sus componentes: distribución de muestreo (verosimilitud), la distribución a priori y cualquier estructura jerárquica que se esté empleando. Cualquiera de estos elementos puede causar que el modelo no cumpla las expectativas para las cuales fue propuesto, y puede estar sujeto a cambio. Por lo tanto, es necesario poder comparar y seleccionar entre diferentes modelos aquel (o aquellos) que mejor se ajustan al problema estudiado. Es importante destacar que no siempre existe un "mejor" modelo, se puede tener un conjunto de modelos que poseen un desempeño similar.

Diagnóstico de modelos

Al evaluar un modelo, la pregunta pertinente no es si el modelo es verdadero o falso, ya que en la mayoría de los casos ningún modelo es totalmente correcto, aún cuando sea útil en la práctica. La pregunta relevante será, entonces, si las deficiencias del modelo tienen un impacto importante en la inferencia.

Diagnóstico de modelos

Al evaluar un modelo, la pregunta pertinente no es si el modelo es verdadero o falso, ya que en la mayoría de los casos ningún modelo es totalmente correcto, aún cuando sea útil en la práctica. La pregunta relevante será, entonces, si las deficiencias del modelo tienen un impacto importante en la inferencia.

La primera herramienta para el diagnóstico de un modelo es la distribución posterior que produce. Se pueden hacer los siguientes diagnósticos con esta:

- Comparar la distribución posterior de los parámetros con un conocimiento importante del problema o con otros datos que no hayan sido incluidos en la distribución a priori o en la verosimilitud. Por ejemplo, comparar la probabilidad posterior de $\theta_j=$ probabilidad del jugador j de hacer un gol en un partido de fútbol, con el rendimiento de los jugadores en partidos anteriores.

Diagnóstico de modelos

- Comparar la distribución predictiva posterior de observaciones futuras con el conocimiento del problema. Por ejemplo, comparar las predicciones por departamento de una elección presidencial con un conocimiento político importante sobre la preferencia de los candidatos en cada estado.

Diagnóstico de modelos

- Comparar la distribución predictiva posterior de observaciones futuras con el conocimiento del problema. Por ejemplo, comparar las predicciones por departamento de una elección presidencial con un conocimiento político importante sobre la preferencia de los candidatos en cada estado.
- Comparar la distribución predictiva posterior de observaciones futuras con los datos que ya han ocurrido. Esto implica que si el modelo es adecuado, los datos observados deben ser factibles bajo la distribución predictiva posterior. En este caso no se usa información adicional.

Comparación de los datos con la distribución predictiva posterior

El chequeo predictivo posterior trata de responder a la pregunta de si el modelo es consistente con los datos. La técnica básica consiste en simular muestras de la distribución predictiva posterior y comparar estas muestras de datos replicados con los datos observados. Se va a utilizar la siguiente notación:

- $\mathbf{y}$: datos observados.
- $\boldsymbol{\theta}$: vector de parámetros incluyendo todos los hiperparámetros si el modelo es jerárquico.
- $\mathbf{y}^{rep}$: una réplica de $\mathbf{y}$.
- $\tilde{y}$: una observación futura.

Comparación de los datos con la distribución predictiva posterior

Los pasos para medir las discrepancias entre los datos y las simulaciones de la distribución predictiva son:

i. Definir una medida de discrepancia $T(\mathbf{y}, \theta)$, la cual es una cantidad escalar que depende de los parámetros y los datos (equivalente a la estadística de prueba desde el punto de vista clásico).

Comparación de los datos con la distribución predictiva posterior

Los pasos para medir las discrepancias entre los datos y las simulaciones de la distribución predictiva son:

i. Definir una medida de discrepancia $T(\mathbf{y}, \theta)$, la cual es una cantidad escalar que depende de los parámetros y los datos (equivalente a la estadística de prueba desde el punto de vista clásico).

ii. Calcular el p-valor Bayesiano predictivo posterior el cual está dado por:

$$\text{p-valor}_B = p(T(\mathbf{y}^{rep}, \boldsymbol{\theta}) \geq T(\mathbf{y}, \boldsymbol{\theta})|\mathbf{y})$$

donde la probabilidad se toma sobre la distribución posterior de $\boldsymbol{\theta}$ y la distribución predictiva posterior de $\mathbf{y}^{rep}$. Esta es la distribución conjunta $p(\boldsymbol{\theta}, \mathbf{y}^{rep}|\mathbf{y})$.

Comparación de los datos con la distribución predictiva posterior

Si se tienen L simulaciones de la densidad predictiva posterior de $\boldsymbol{\theta}$, se puede simular un valor de $\mathbf{y}^{rep}$ para cada valor de $\boldsymbol{\theta}$ simulado, obteniéndose entonces L simulaciones de la distribución $p(\boldsymbol{\theta}, \mathbf{y}^{rep}|\mathbf{y})$. El diagnóstico predictivo posterior consiste en comparar la cantidad $T(\mathbf{y}, \theta)$ con $T(\mathbf{y}^{rep\,l}, \theta^l)$. El p-valor predictivo posterior estimado es la proporción de las L simulaciones para las cuales $T(\mathbf{y}^{rep\,l}, \theta^l) \geq T(\mathbf{y}, \theta)$. Esta es la probabilidad posterior de que las réplicas sean más extremas que los datos.

Comparación de los datos con la distribución predictiva posterior

Si se tienen L simulaciones de la densidad predictiva posterior de $\boldsymbol{\theta}$, se puede simular un valor de $\mathbf{y}^{rep}$ para cada valor de $\boldsymbol{\theta}$ simulado, obteniéndose entonces L simulaciones de la distribución $p(\boldsymbol{\theta}, \mathbf{y}^{rep}|\mathbf{y})$. El diagnóstico predictivo posterior consiste en comparar la cantidad $T(\mathbf{y}, \theta)$ con $T(\mathbf{y}^{rep\ l}, \theta^l)$. El p-valor predictivo posterior estimado es la proporción de las L simulaciones para las cuales $T(\mathbf{y}^{rep\ l}, \theta^l) \geq T(\mathbf{y}, \theta)$. Esta es la probabilidad posterior de que las réplicas sean más extremas que los datos.

Valores de p-valor$_B$ cercanos a 0.5 indican que la distribución de los datos replicados y los datos reales son cercanas, mientras que valores cercanos a cero o uno indican diferencia entre ellos.

Ejemplo Simon Newcomb

Suponga que se analizan nuevamente lo datos del experimento de Simon Newcomb. Se ajusta el modelo:

$$\begin{aligned} y_i|\mu, \sigma^2 &\sim \text{Normal}(\mu, \sigma^2) \\ p(\mu, \sigma^2) &\propto \frac{1}{\sigma^2} \end{aligned}$$

por lo tanto:

$$\begin{aligned} p(\sigma^2|\mathbf{y}) &= \text{Inv-}\chi^2((n-1), s^2) \\ p(\mu|\sigma^2, \mathbf{y}) &= \text{Normal}(\bar{y}, \sigma^2/n) \end{aligned}$$

Ejemplo Simon Newcomb

Para verificar la distribución predictiva posterior hacemos para $i = 1, \cdots, L$ ($L = 100$):

i) Generar $(\mu^{(i)}, \sigma^{2(i)})$ de $p(\sigma^2|\mathbf{y})$ y $p(\mu|\sigma^2, \mathbf{y})$.

ii) Generar $y_1^{rep(i)}, \cdots, y_{66}^{rep(i)}$ de una Normal$(\mu^{(i)}, \sigma^{2(i)})$

Utilizar como medida de discrepancia el mínimo.

Ejemplo Schools

Se lleva a cabo un estudio para analizar el efecto relativo de un programa de entrenamiento en ocho colegios diferentes sobre una prueba de aptitud escolar. Se ajusta el siguiente modelo:

$$y_j \sim \mathrm{N}(\mu + \alpha_j, \sigma_j^2) \quad j = 1, \cdots, J$$

$$\alpha_j \sim \mathrm{N}(0, \sigma_\alpha^2)$$

donde σ_j^2 es conocida.