

Modelos Mixtos Bayesianos

Federico Alvarez

17 de septiembre de 2019

Reintroducción a Bayes

Frecuentismo vs. bayesianismo

La diferencia principal entre el enfoque bayesiano y el frecuentista es la interpretación del concepto de probabilidad. La perspectiva frecuentista, como lo indica su nombre, concibe a la probabilidad (de un evento) como un número que denota su “frecuencia relativa” en un alto número de repeticiones del mismo experimento. La perspectiva bayesiana, en cambio, ve a la probabilidad como un número que denota nuestro estado de conocimiento sobre el evento. Entre los defensores de la estadística bayesiana hay dos grandes grupos. Están los que consideran a la probabilidad bayesiana como una medida objetiva y racional del estado de conocimiento (objetivistas) y los que la ven como una indicación de la creencia personal (subjetivistas). Esto implica que los objetivistas abogan por una inferencia uniforme a partir de los datos por parte de investigadores que comparten un mismo nivel de conocimiento sobre un problema, mientras que los subjetivistas sostienen que esta inferencia puede diferir y depende fuertemente de las creencias individuales. Tanto subjetivistas como objetivistas realizan inferencias a través de la aplicación de la regla de Bayes. Los objetivistas, sin embargo, procuran minimizar la influencia que una única persona puede tener en el proceso de inferencia, de manera tal que las conclusiones finales sean determinadas por los datos tanto como sea posible. Los intentos de minimizar esta influencia a veces se llevan a cabo introduciendo *priors* de referencia o no informativos.

¿Por qué usar modelos bayesianos?

De acuerdo con Nicenboim y Vasisht hay dos ventajas principales en el uso de métodos bayesianos. En primer lugar, nos permiten reponder directamente la pregunta que nos importa: ¿Cuán plausible es nuestra hipótesis dados nuestros datos? En segundo lugar, y quizá más importante, es más fácil definir modelos mixtos en el marco bayesiano que en el frecuentista.

En un marco bayesiano nuestro objetivo es obtener una estimación de un parámetro de la población (μ) dados los datos disponibles y con una estimación de la incertidumbre sobre el valor del parámetro (como el intervalo creíble) que nos da un rango dentro del cual podemos estar seguros de que está el valor verdadero. Estas estimaciones se obtienen a partir de los datos y dado nuestro conocimiento previo sobre valores plausibles del parámetro. La idea fundamental es que la distribución de μ (el *posterior*) puede ser expresada en términos del *prior* y la verosimilitud:

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

Entonces, dada información a priori sobre el parámetro de interés y la verosimilitud, podemos calcular la distribución posterior del parámetro. El foco no es rechazar una hipótesis nula sino encontrar lo que la distribución posterior nos dice sobre el rango de valores posibles del parámetro de interés. Mientras que la estadística frecuentista se enfoca en calcular el valor p , o sea, la probabilidad de los datos dada una hipótesis nula, la estadística bayesiana nos permite hablar sobre los posibles valores del parámetro μ de acuerdo con los datos a través de su distribución a posteriori. Algo importante es que el *posterior* es básicamente una media pesada entre el *prior* y la verosimilitud. Cuando el *prior* está disperso en un rango amplio de valores a los que asigna una misma probabilidad, el *posterior* termina asemejándose más a la verosimilitud. Cuando tenemos *priors* débiles con datos escasos, el *posterior* se acerca más al *prior*, pero si la cantidad de datos aumenta, nuevamente se acerca a la verosimilitud. Esto implica que, cuando tenemos pocos datos, vale la pena invertir tiempo en seleccionar *priors* informados por conocimiento previo, pero cuando tenemos muchos, la verosimilitud va a dominar el *posterior*. De hecho, en casos de una gran cantidad de muestras,

usualmente la distribución a posteriori bayesiana y la media frecuentista, con sus respectivas estimaciones de incertidumbre, son casi idénticas, o muy similares (aunque sus interpretaciones son muy diferentes).

¿Por qué no usar NHST?

Los valores de p obtenidos utilizando tests frecuentistas suponen la asignación de probabilidad a la totalidad de eventos en el espacio muestral de acuerdo con la hipótesis nula. Eso implica que operar sobre el espacio muestral, necesariamente modificamos la distribución de valores de probabilidad para un mismo conjunto de datos. Las maneras de operar sobre el espacio muestral son muchas, como por ejemplo el criterio para detener la recolección de muestras o la cantidad y tipo de tests a llevar a cabo (CITA KRUSHKE). Un problema es que (en una investigación apropiadamente llevada a cabo) las observaciones no dependen de las decisiones de la persona que investiga y, por lo tanto, las conclusiones que obtengamos de ellas tampoco deberían estar sujetas a esa dependencia.

En un marco bayesiano, la probabilidad de un conjunto de observaciones es simplemente el producto de las probabilidades de las observaciones individuales. La función de verosimilitud captura todos los aspectos que, consideramos, influyen sobre los datos. Nuestras hipótesis, por otro lado, se operacionalizan a través de la distribución a priori que supongamos que tienen las distintas fuentes de variabilidad.

Podría suponerse que las creencias a priori no son menos misteriosas que los criterios de finalización y testeo, pero no es el caso: las creencias a priori no son caprichosas ni idiosincráticas. Los *priors* deben ser explícitos, debatibles, y fundamentados sobre la base de investigación previa. En la medida en que un investigador seleccione un *prior* determinado para sus pruebas, deberá argumentar a favor de su selección, convencer a los revisores y editores de su trabajo, y arriesgarse a no ser publicado si no consideran adecuada su elección. Entonces surge la pregunta: ¿si en un análisis bayesiano se permiten *priors* informativos, por qué no aceptar las elecciones subjetivas en NHST? La respuesta a esto es que las elecciones subjetivas de NHST afectan al espacio muestral, esto es, a los datos no observados. En cambio, las creencias a priori no son sobre cosas que no ocurren sino sobre la manera en que los datos influyen sobre las creencias a posteriori.

Parametrizar un modelo

```
data <- read_csv('gibsonwu2012data.txt') %>%
  filter(region == 'headnoun') %>%
  mutate(cond = ifelse(type2 == "object relative", 1, -1))
```

```
## Parsed with column specification:
## cols(
##   subj = col_double(),
##   item = col_double(),
##   type = col_double(),
##   pos = col_character(),
##   word = col_double(),
##   correct = col_character(),
##   rt = col_double(),
##   region = col_character(),
##   type2 = col_character()
## )
```

```
mod <- stan_lmer(
  formula = log(rt) ~ cond + (cond|subj) + (cond|item),
  prior_intercept = normal(0,10),
  prior = normal(0,1),
  prior_covariance = decov(regularization = 2),
  data = data,
  chains = 4,
  iter = 2000,
```

```
cores = 4,  
adapt_delta = 0.99,  
control = list(max_treedepth = 20)  
)
```

```
## Warning: There were 4 chains where the estimated Bayesian Fraction of Missing Information was low. See  
## http://mc-stan.org/misc/warnings.html#bfmi-low  
  
## Warning: Examine the pairs() plot to diagnose sampling problems  
  
## Warning: Markov chains did not converge! Do not analyze results!
```

Especificar *priors*

Para parametrizar un modelo bayesiano necesitamos especificar una distribución a priori sobre cada parámetro. Estas distribuciones expresan nuestro estado de conocimiento inicial sobre los posibles valores del parámetro. Es posible especificar *priors* completamente no informativos, como *priors* planos, pero esto está lejos de ser lo ideal porque concentran mucha masa de probabilidad por fuera de cualquier valor razonable de la distribución a posteriori. Esto puede causar que, si no tenemos datos suficientes, el *prior* va a dominar el cálculo de la media del *posterior* y la incertidumbre asociada. Los *priors* que aportan una cantidad mínima de información mejoran la inferencia y se llaman *priors débilmente informativos* o *regularizadores*.

Fundamentalmente, es importante especificar los *priors* explícitamente. Los *priors* por defecto asumidos por la función pueden no ser apropiados para nuestros datos en particular.

Posterior e inferencia estadística

El resultado de un análisis bayesiano es una distribución a posteriori que muestra la plausibilidad relativa de cada posible valor del parámetro de interés, condicionada por los datos, las distribuciones a priori, y el modelo. Cada parámetro del modelo (los efectos fijos, los aleatorios, la forma de la distribución) va a tener su propia distribución a posteriori. Típicamente las funciones de análisis devuelven muestras del *posterior* que podemos usar para realizar inferencias.

Para comunicar los resultados, necesitamos resumir e interpretar la distribución a posteriori. Podemos reportar estimaciones puntuales de la probabilidad a posteriori como la media o la mediana (en algunos casos también se reporta la moda). Cuando el *posterior* es simétrico y aproximadamente normal, no importa cuál reportemos.

Intervalos del posterior

Es posible (y deseable) reportar un intervalo de probabilidad posterior, o sea, dos parámetros que contienen entre ellos una cantidad especificada de probabilidad posterior. Este tipo de intervalo puede ser un intervalo creíble, que es un intervalo de percentiles que marca el rango dentro del cual podemos tener la certeza, con cierta probabilidad, de que se encuentra el “valor real” de un parámetro. Es el valor real del mundo lógico del modelo, no del mundo real. Otra opción es reportar el intervalo más angosto que contenga una masa de probabilidad predeterminada. Esto se llama intervalo posterior de densidad más elevada (HDPI). El uso de uno u otro tipo de intervalo depende fuertemente de la audiencia a la cual se le comuniquen los resultados. El intervalo creíble es más similar al intervalo de confianza frecuentista, es menos demandante computacionalmente y es más sensible a variaciones en la toma de muestras del *posterior*, mientras que el HDPI es mucho más angosto. En cualquier caso, ambos deberían ser relativamente similares. Si la elección de intervalo hace una gran diferencia, entonces no debería usarse ningún intervalo; la estimación bayesiana consiste en la totalidad de la distribución a posteriori de todos los modos.

Reportar resultados

Primero hay que reportar toda la información relevante para el modelo, o sea, los *priors* y la cantidad de cadenas e iteraciones. De haberlo hecho, se puede reportar el uso de un análisis de sensibilidad para asegurar

que el *posterior* no se ve excesivamente afectado por el *prior*. Aparte de esto, se reportan una estimación del efecto, un intervalo creíble o de máxima densidad, y la probabilidad de un efecto negativo.

Testeo de hipótesis con el factor de Bayes

El factor de Bayes es una forma de cuantificar la evidencia según la cual los datos observados son más probables respecto de otro modelo. Esto se consigue calculando la proporción de la verosimilitud marginal de dos modelos que corresponden a hipótesis diferentes.