

EST-46115: Modelación Bayesiana

Profesor: Alfredo Garbuno Iñigo — Primavera, 2022 — Muestreo predictivo.

Objetivo: Que veremos.

Lectura recomendada: Referencia.

1. INTRODUCCIÓN

Como hemos visto podemos utilizar la distribución predictiva para traducir nuestros supuestos distribucionales (estado de conocimiento) en cantidades observables. En esta sección del curso usaremos la **distribución predictiva previa** para evaluar consistencia de cómo matematizamos nuestro conocimiento previo con respecto al de un experto. Por otro lado, utilizaremos la **distribución predictiva posterior** con el objetivo de evaluar y criticar el ajuste del modelo propuesto.

2. MUESTREO PREDICTIVO – PREVIA

Usualmente definimos la distribución predictiva previa de manera que refleje lo que sabemos del problema que estamos modelando. Esperaríamos que fuera fácil en situaciones sencillas. Pero incluso en modelos medianamente complejos es difícil 1) poder definir dicha distribución y 2) entender las consecuencias de esa elección.

2.1. Muestreo predictivo

Recordemos que la distribución predictiva previa es

$$\pi(y) = \int \pi(y|\theta) \pi(\theta) d\theta. \quad (1)$$

El mecanismo que utilizamos para generar muestras de dicha distribución es el siguiente:

- Para $i = 1, \dots, N$:
 1. Generar $\tilde{\theta}_i \sim \pi(\theta)$.
 2. Generar $\tilde{y}_i \sim \pi(y|\tilde{\theta}_i)$.

Al final de este proceso podemos descartar las simulaciones de $\tilde{\theta}$ y obtener una colección de realizaciones aleatorias de $\tilde{y} \sim \pi(y)$.

2.1.1. Para pensar: ¿Por qué este proceso iterativo nos deja muestras de la marginal?
Hint: considera el caso para dos variables y considera los gráficos de dispersión y los histogramas individuales.

2.2. Interpretación

El objetivo de utilizar la distribución predictiva previa es evaluar la interacción de la previa con la verosimilitud.

2.3. En la práctica

Usualmente consideramos algún resumen informativo de los datos. Es decir, un resumen a través de un **estadístico** de nuestra distribución, y en consecuencia lo que nos interesaría es evaluar la **distribución de muestreo** de dicho estadístico. La idea es poder definir regiones donde esperaríamos ver nuestra simulación de posibles estadísticos.

Por ejemplo, consideremos el caso de los cantantes de ópera. El modelo asume una distribución previa de la forma

$$\mu|\sigma \sim \text{Normal}\left(\mu_0, \frac{\sigma}{n_0}\right), \quad (2)$$

$$\sigma^{-1} \sim \text{Gamma}(a_0, b_0). \quad (3)$$

Donde los parámetros tienen la siguiente especificación:

```
1 previa.params <- within(list(),
2 {
3   mu0 <- 175
4   n0 <- 5
5   a0 <- 3
6   b0 <- 140
7 })
```

Una manera práctica es hacer remuestreo de la distribución predictiva previa para poder evaluar ciertos estadísticos de resumen. La idea es que podamos definir una región donde la previa genere datos razonables. Por ejemplo, podemos definir una zona donde esperamos observar el promedio de la altura de 200 cantantes. Ver Fig. 1.

```
1 data {
2   real mu0;
3   real<lower=0> n0;
4   real<lower=0> a0;
5   real<lower=0> b0;
6 }
7 generated quantities {
8   real tau = gamma_rng(a0, b0);
9   real sigma = 1/tau;
10  real mu = normal_rng(mu0, sigma/n0);
11  real y_tilde = normal_rng(mu, sigma);
12 }
```

```
1 replica <- function(id){
2   previa <- modelo$sample(previa.params,
3                           fixed_param = TRUE,
4                           refresh = 0, chains = 1,
5                           show_messages = FALSE)
6   mean(previa$draws(format = "df")$y_tilde)
7 }
8
9 resultados <- tibble(id = 1:200) >
10   mutate(ymean = map_dbl(id, replica))
```

Por supuesto escoger el resumen correcto es complicado pero con la respuesta clara que queremos responder nos indicará qué resumen escoger.

Para una discusión mas profunda sobre el estado del arte en elicitación y prácticas para definir las distribuciones previas consultar [1].

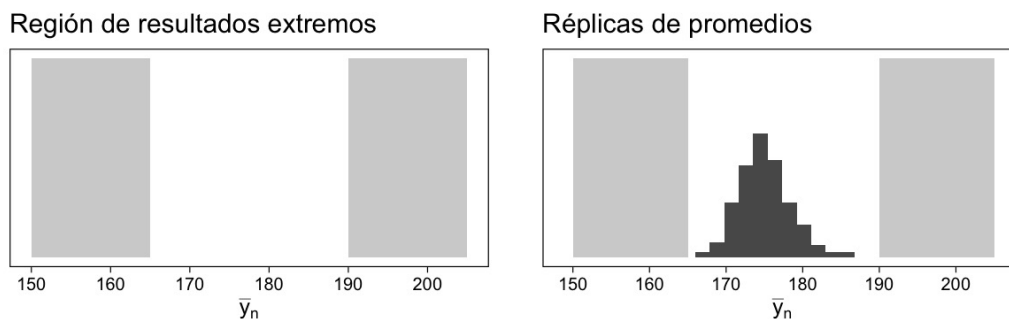


FIGURA 1. Definimos una región donde consideremos resultados extremos de una estadística resumen para poder evaluar la distribución muestral de nuestro estadístico producto de la distribución predictiva previa.

3. MUESTREO PREDICTIVO – POSTERIOR

REFERENCIAS

- [1] P. Mikkola, O. A. Martin, S. Chandramouli, M. Hartmann, O. A. Pla, O. Thomas, H. Pesonen, J. Corander, A. Vehtari, S. Kaski, P.-C. Bürkner, and A. Klami. Prior knowledge elicitation: The past, present, and future. *arXiv:2112.01380 [stat]*, dec 2021. [2](#)