EST-46115: Modelación Bayesiana

Profesor: Alfredo Garbuno Iñigo — Primavera, 2022.
Objetivo. Estudiar integración en el contexto probabilistico. Estudiar el método Monte Carlo y entender sus bondades y limitaciones en el contexto de inferencia Bayesiana.
Lectura recomendada: Sección 6.1 de [1].

1. INTRODUCCIÓN

En inferencia bayesiana lo que queremos es poder resolver

$$\mathbb{E}[f] = \int_\Theta f(\theta)\, \pi(\theta|y)\, \mathrm{d}\theta \,. \tag{1}$$

Lo que necesitamos es resolver integrales con respecto a la distribución de interés.

- La pregunta clave (I) es: ¿qué distribución?
- La pregunta clave (II) es: ¿con qué método numérico resuelvo la integral?
- La pregunta clave (III) es: ¿y si no hay método numérico?

2. INTEGRACIÓN NUMÉRICA

Recordemos la definición de integrales Riemann:

$$\int f(x)\mathrm{d}x \,.$$

La aproximación utilizando una malla de N puntos sería:

$$\sum_{n=1}^{N} f(u_n)\Delta u_n.$$

El método útil cuando las integrales se realizan cuando tenemos pocos parámetros. Es decir, $\theta \in \mathbb{R}^p$ con p pequeña.

2.1. Ejemplo: Proporción

Supongamos que $p(S_n = k|\theta) \propto \theta^k(1-\theta)^{n-k}$ cuando observamos k éxitos en n pruebas independientes. Supongamos que nuestra inicial es $p(\theta) = 2\theta$ (`checa que es una densidad`).

```
crear_log_post <- function(n, k){
  function(theta){
    verosim <- k * log(theta) + (n - k) * log(1 - theta)
```

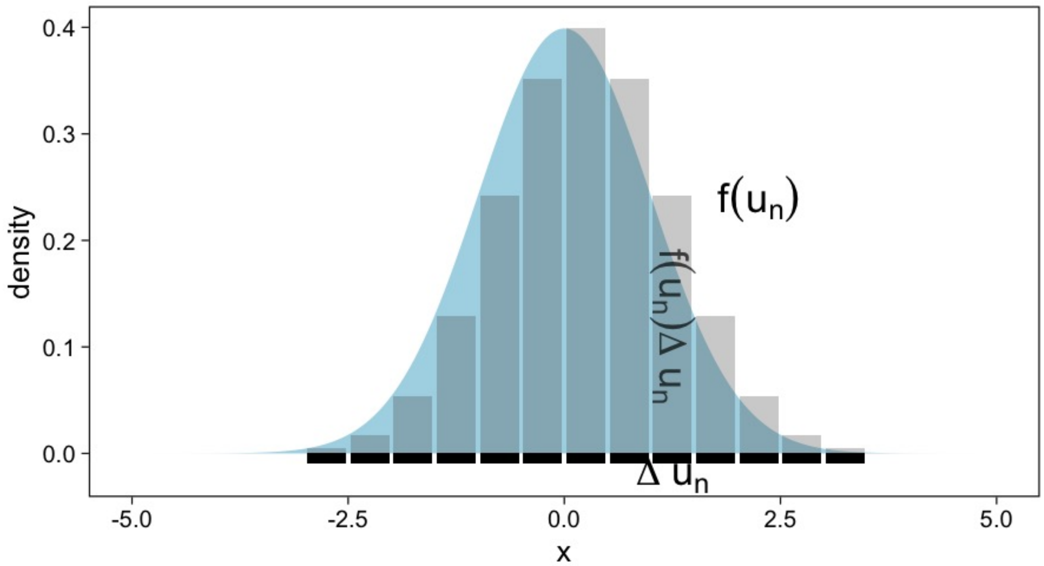

FIGURA 1. *Integral por medio de discretización.*

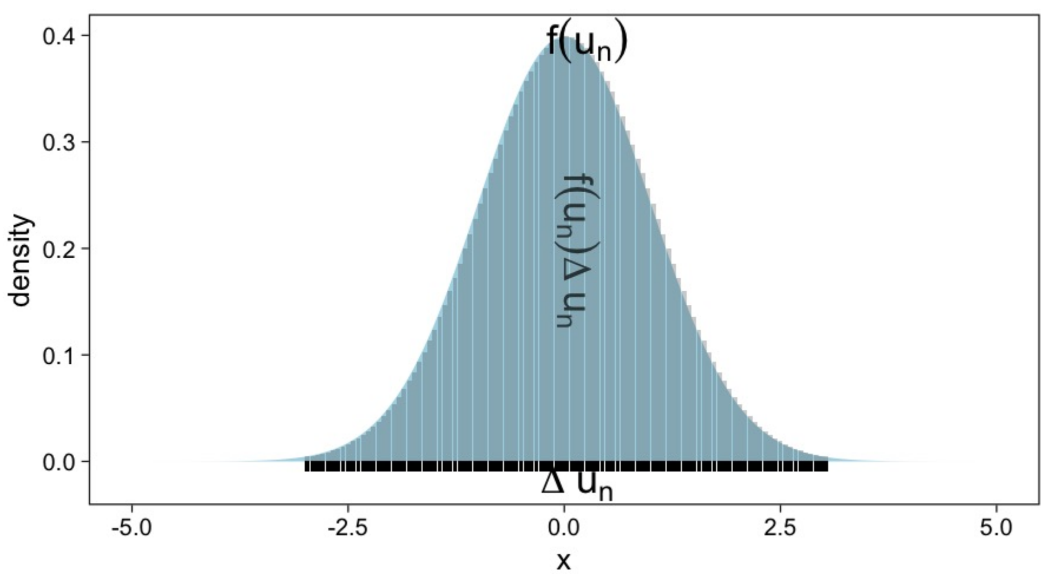

FIGURA 2. *Integral por medio de una malla fina.*

```
    inicial <- log(theta)
    verosim + inicial
  }
}
```

```
# observamos 3  xitos  en 4 pruebas:
log_post <- crear_log_post(4, 3)
prob_post <- function(x) { exp(log_post(x))}
# integramos num ricamente
p_x <- integrate(prob_post, lower = 0, upper = 1, subdivisions = 100L)
p_x
```

```
0.033 with absolute error < 3.7e-16
```

Y ahora podemos calcular la media posterior:

$$\mathbb{E}[\theta] = \int \theta\, \pi(\theta|S_n)\, \mathrm{d}\theta\,. \tag{2}$$

```
media_funcion <- function(theta){
  theta * prob_post(theta) / p_x$value
}
integral_media <- integrate(media_funcion,
                            lower = 0, upper = 1,
                            subdivisions = 100L)
media_post <- integral_media$value
c(Numerico = media_post, Analitico = 5/(2+5))
```

```
Numerico Analitico
    0.71      0.71
```

2.2. Más de un parámetro

Consideramos ahora un espacio con $\theta \in \mathbb{R}^p$. Si conservamos N puntos por cada dimensión, ¿cuántos puntos en la malla necesitaríamos? Lo que tenemos son recursos computacionales limitados y hay que buscar hacer el mejor uso de ellos. En el ejemplo, hay zonas donde no habrá contribución en la integral.

3. INTEGRACIÓN MONTE CARLO

$$\pi(f) = \mathbb{E}_\pi[f] = \int f(x)\pi(x)\mathrm{d}x\,,$$

$$\pi_N^{\mathsf{MC}}(f) = \frac{1}{N}\sum_{n=1}^{N} f(x^{(n)}), \qquad \text{donde } x^{(n)} \overset{\text{iid}}{\sim} \pi, \qquad \text{con } n = 1, \ldots, N\,,$$

$$\pi(f) \approx \pi_N^{\mathsf{MC}}(f)\,.$$

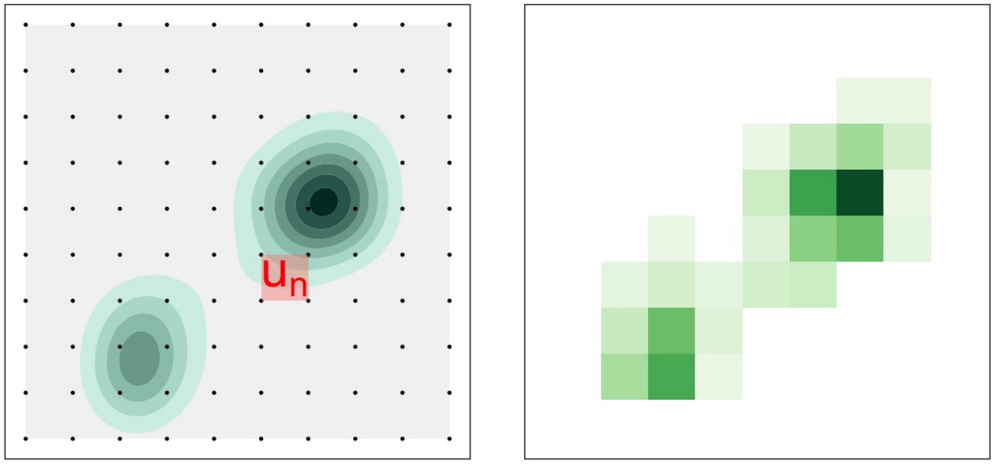

Figura 3. *Integral por método de malla.*

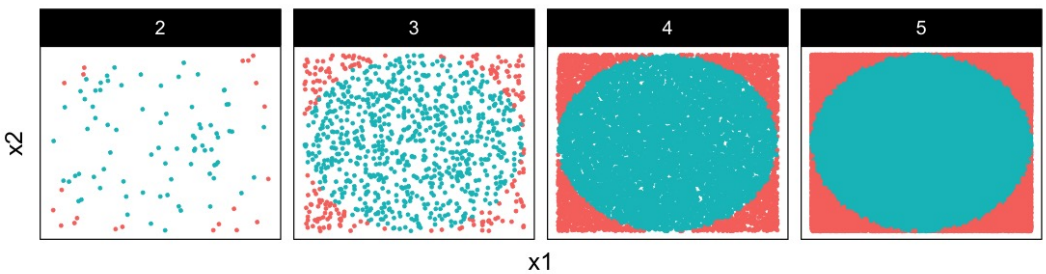

Figura 4. *Integración Monte Carlo para aproximar π.*

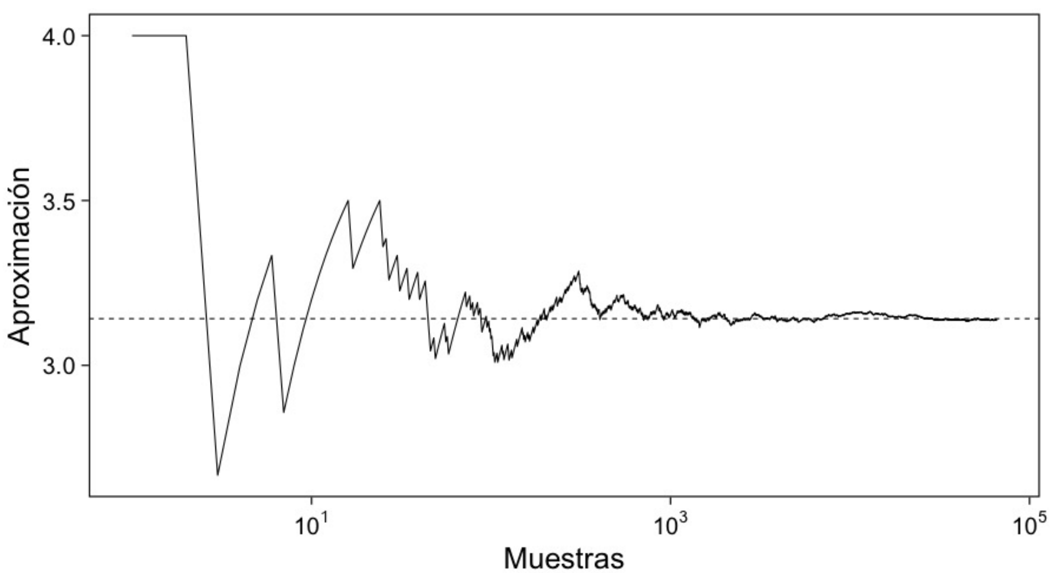

Figura 5. *Estimación $\pi_N^{MC}(f)$ con $N \to \infty$.*

3.1. Ejemplo: Dardos

Consideremos el experimento de lanzar dardos uniformemente en un cuadrado de tamaño 2, el cual contiene un circulo de radio 1.

3.2. Propiedades

Teorema (`Error Monte Carlo`). Sea $f : \mathbb{R}^p \to \mathbb{R}$ cualquier función bien comportada†. Entonces, el estimador Monte Carlo es **insesgado**. Es decir, se satisface

$$\mathbb{E}\left[\pi_N^{\mathsf{MC}}(f) - \pi(f)\right] = 0,$$

para cualquier N. Además tiene **error cuadrático medio acotado**

$$\sup_{f \in \mathcal{F}} \; \mathbb{E}\left[\left(\pi_N^{\mathsf{MC}}(f) - \pi(f)\right)^2\right] \leq \frac{1}{N}.$$

En particular, la varianza del estimador (**error estándar**) satisface la igualdad

$$\mathsf{ee}^2\left(\pi_N^{\mathsf{MC}}(f)\right) = \frac{\mathbb{V}_\pi(f)}{N}.$$

Teorema (`TLC para estimadores Monte Carlo`). Sea f una función **bien comportada** ††, entonces bajo una N suficientemente grande tenemos

$$\sqrt{N}\left(\pi_N^{\mathsf{MC}}(f) - \pi(f)\right) \sim \mathsf{N}\left(0, \mathbb{V}_\pi(f)\right)\,. \tag{3}$$

3.3. Ejemplo: Proporciones

Consideramos la estimación de una proporción θ, tenemos como inicial $p(\theta) \propto \theta$, que es una $\mathsf{Beta}(2,1)$. Si observamos 3 éxitos en 4 pruebas, entonces sabemos que la posterior es $p(\theta|x) \propto \theta^4(1-\theta)$, que es una $\mathsf{Beta}(5,2)$. Si queremos calcular la media y el segundo momento posterior para θ, en teoría necesitamos calcular

$$\mu_1 = \int_0^1 \theta\ p(\theta|X=3)\,\mathrm{d}\theta, \qquad \mu_2 = \int_0^1 \theta^2\ p(\theta|X=3)\,\mathrm{d}\theta. \tag{4}$$

Utilizando el `método Monte Carlo`:

```
theta <- rbeta(10000, 5, 2)
media_post <- mean(theta)
momento_2_post <- mean(theta^2)
c(mu_1 = media_post, mu_2 = momento_2_post)
```

```
mu_1 mu_2
0.72 0.54
```

Incluso, podemos calcular cosas mas *exóticas* como

$$P(e^\theta > 2|x)\,. \tag{5}$$

```
mean(exp(theta) > 2)
```

```
[1] 0.61
```

3.4. Ejemplo: Sabores de helados

Supongamos que probamos el nivel de gusto para 4 sabores distintos de una paleta. Usamos 4 muestras de aproximadamente 50 personas diferentes para cada sabor, y cada uno evalúa si le gustó mucho o no. Obtenemos los siguientes resultados:

```
      sabor  n gusto prop_gust
1     fresa 50    36      0.72
2     limon 45    35      0.78
3     mango 51    42      0.82
4 guanabana 50    29      0.58
```

Listing 1. *Resultados de las encuestas.*

Usaremos como inicial $\mathsf{Beta}(2,1)$ (pues hemos obervado cierto sesgo de cortesía en la calificación de sabores, y no es tan probable tener valores muy bajos) para todos los sabores, es decir $p(\theta_i)$ es la funcion de densidad de una $\mathsf{Beta}(2,1)$. La inicial conjunta la definimos entonces, usando `independencia inicial`, como

$$p(\theta_1, \theta_2, \theta_3, \theta_4) = p(\theta_1)p(\theta_2)p(\theta_3)p(\theta_4)\,.$$

Pues inicialmente establecemos que ningún parámetro da información sobre otro: saber que mango es muy gustado no nos dice nada acerca del gusto por fresa. Bajo este supuesto, y el supuesto adicional de que las muestras de cada sabor son independientes, podemos mostrar que las `posteriores son independientes`:

$$p(\theta_1, \theta_2, \theta_3, \theta_4|k_1, k_2, k_3, k_4) = p(\theta_4|k_1)p(\theta_4|k_2)p(\theta_4|k_3)p(\theta_4|k_4)$$

```
      sabor  n gusto prop_gust a_post b_post media_post
1     fresa 50    36      0.72     38     15       0.72
2     limon 45    35      0.78     37     11       0.77
3     mango 51    42      0.82     44     10       0.81
4 guanabana 50    29      0.58     31     22       0.58
```

Listing 2. *Resultado de inferencia Bayesiana.*

Podemos hacer preguntas interesantes como: ¿cuál es la probabilidad de que mango sea el sabor preferido? Para contestar esta pregunta podemos utilizar simulación y responder por medio de un procedimiento Monte Carlo.

```
## Generamos muestras de la posterior
paletas <- datos |>
  mutate(alpha = a_post, beta = b_post) |>
  nest(params.posterior = c(alpha, beta)) |>
  mutate(muestras.posterior = map(params.posterior, modelo_beta)) |>
  select(sabor, muestras.posterior)
```

```
## Utilizamos el metodo Monte Carlo para aproximar la integral.
paletas |>
  unnest(muestras.posterior) |>
  mutate(id = rep(seq(1, 5000), 4)) |> group_by(id) |>
  summarise(favorito = sabor[which.max(muestras.posterior)]) |>
  group_by(favorito) |> tally() |>
  mutate(prop = n/sum(n)) |>
  as.data.frame()
```

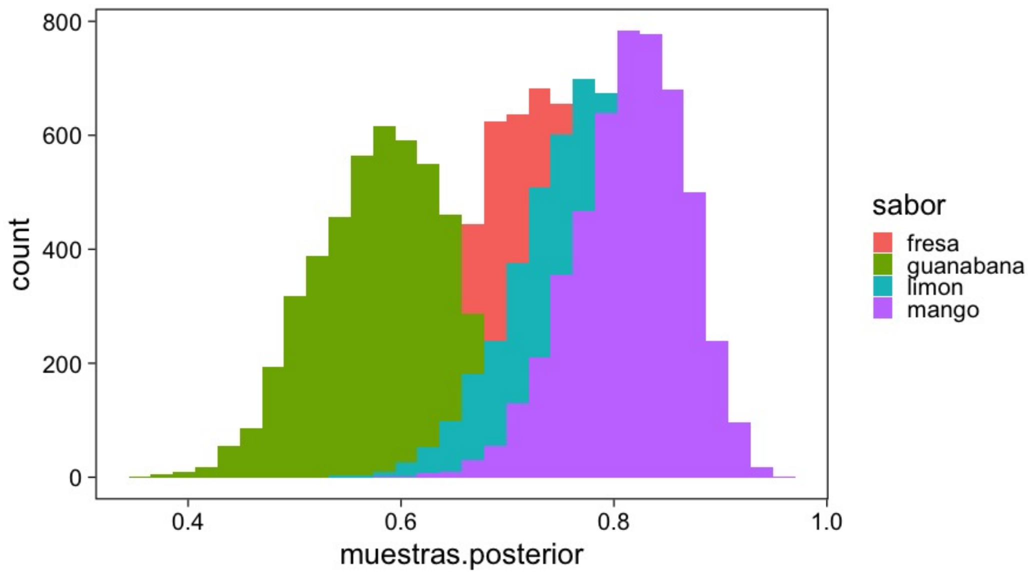

FIGURA 6. *Histogramas de la distribución predictiva marginal para cada θ_j.*

```
   favorito    n   prop
1     fresa  321 0.0642
2 guanabana    2 0.0004
3     limon 1327 0.2654
4     mango 3350 0.6700
```

LISTING 3. *Aproximación Monte Carlo.*

Escencialmente estamos preguntándonos sobre calcular la integral:

$$\mathbb{P}(\text{mango sea preferido}) = \int_{\Theta} f(\theta_1, \ldots, \theta_4)\, p(\theta_1, \ldots, \theta_4 | X_1, \ldots, X_n) d\theta \,, \qquad (6)$$

donde $f(\theta_1, \ldots, \theta_4) = \mathbb{I}_{[\theta_4 \geq \theta_j, j \neq 4]}(\theta_1, \ldots, \theta_4)$.

3.5. Tarea: Sabores de helados

- ¿Cuál es la probabilidad a priori de que cada sabor sea el preferido?
- Con los datos de arriba, calcula la probabilidad de que la gente prefiera el sabor de mango sobre limón.

REFERENCIAS

[1] A. Johnson, M. Ott, and M. Dogucu. *Bayes Rules! An Introduction to Applied Bayesian Modeling.* 2021. 1