EST-46115: Modelación Bayesiana

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?

1.1. 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. 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)
    inicial <- log(theta)
    verosim + inicial
  }
}
```

```
# observamos 3  xitos  en 4 pruebas:
log_post <- crear_log_post(4, 3)
```

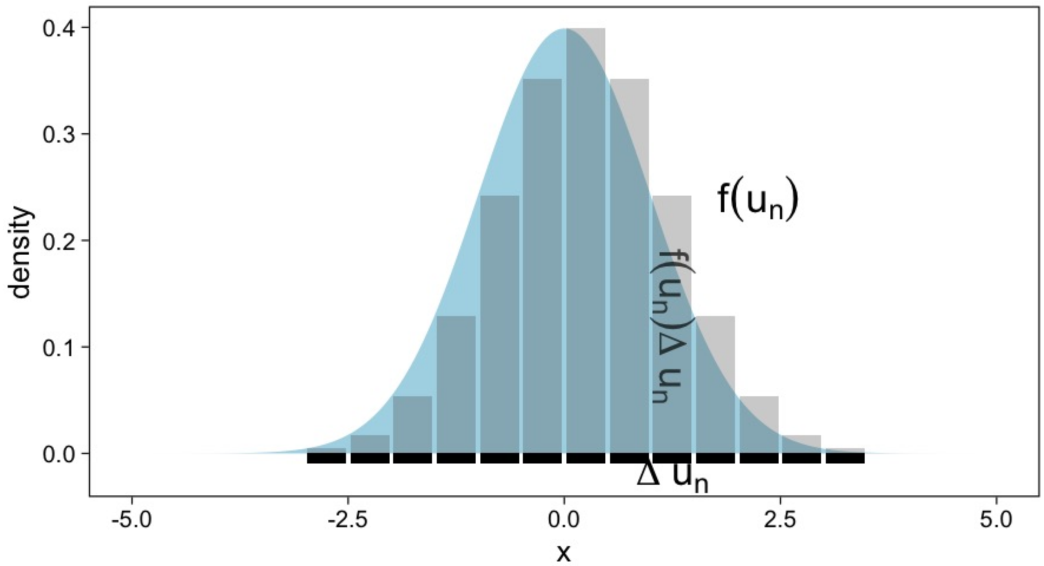

Figura 1. *Integral por medio de discretización.*

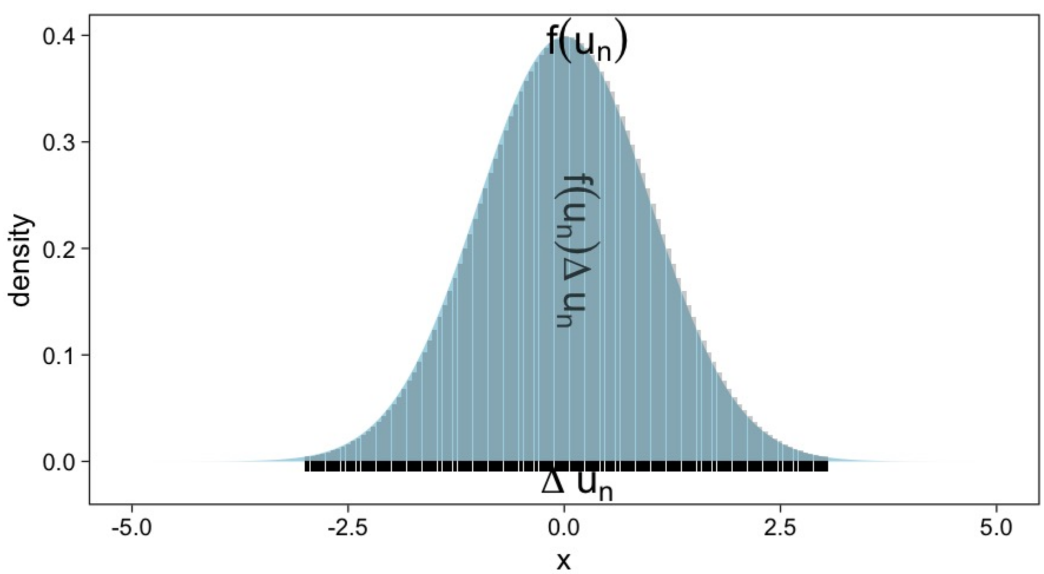

Figura 2. *Integral por medio de una malla fina.*

```r
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:

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

```r
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
```

3. 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.

4. INTEGRACIÓN MONTE CARLO

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

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

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

4.1. Ejemplo: Dardos

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

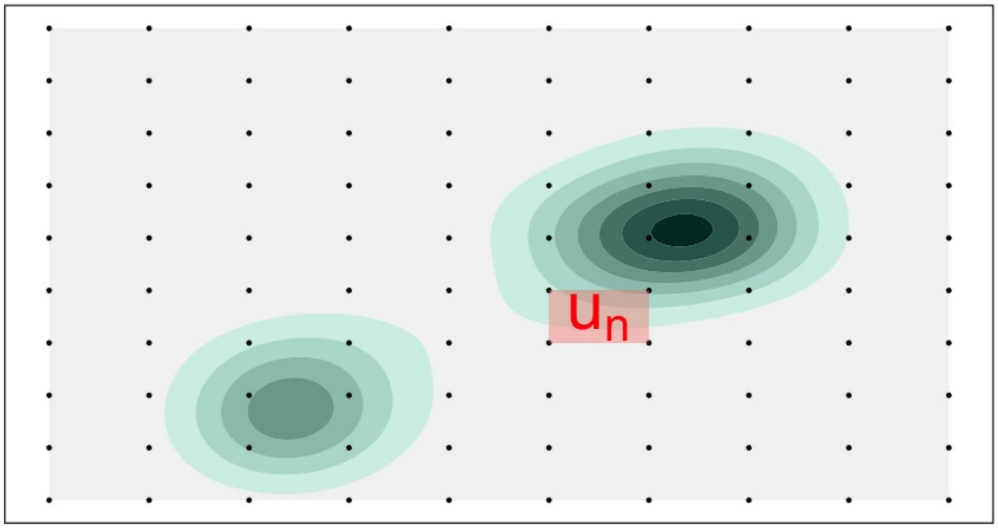

Figura 3. *Integral por método de malla.*

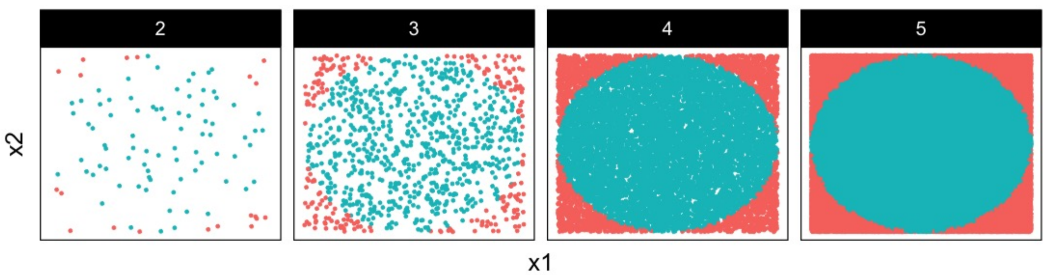

Figura 4. *Integración Monte Carlo para aproximar π.*

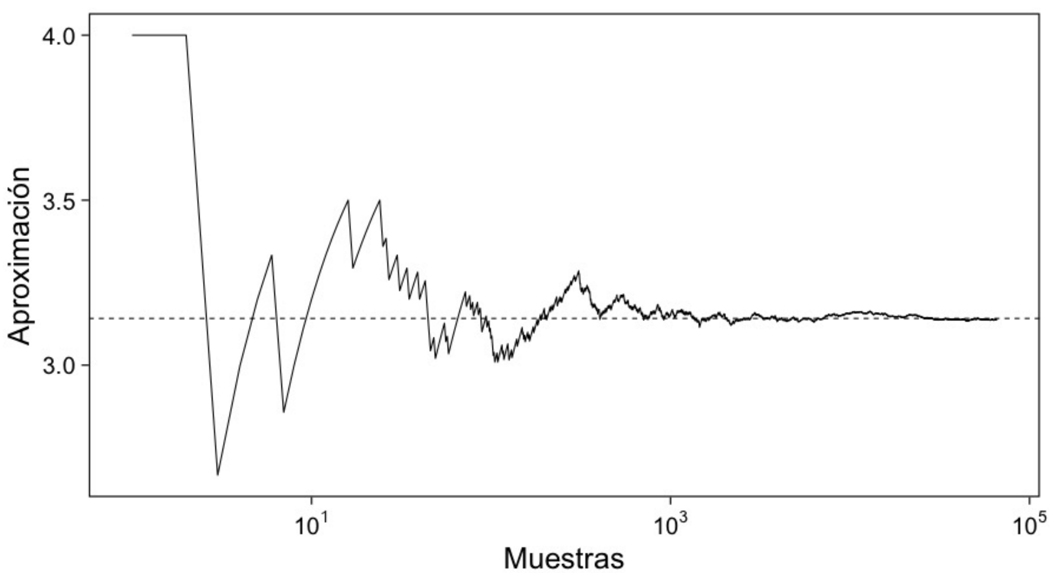

Figura 5. *Estimación $\pi_N^{MC}(f)$ con $N \to \infty$.*

4.2. 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

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