
Una revisión de los metodólogos de integración

Modelos Bayesianos con aplicaciones ecológicas

Dr. Cole Monnahan

University of Concepción, Chile

Enero, 2018

Por que integración?

- Recuerda que para calcular probabilidades tenemos que integrar una distribución continua

$$P(a < X < b) = \int_a^b f(x)dx$$

- Que es la altura de la curva en este caso?
- Que es $P(X=a)=?$
- **La área bajo de la curva es la probabilidad**

Método 1: Integración analítica

- Suponga que hay una v.a. uniforme:
 $X \sim U(0,5)$
- Entonces la pdf $f(x)=1/5$ por x en $(0,5)$ y 0 de lo contrario
- $P(a < X < b) = \int_a^b 1/5 dx = (b - a)/5$
- Si $a=0$, $b=5$; $P=1$
- Si $a=2.5$, $b=5$; $P=0.5$
- Este método es exacto y rápido
- ...pero es arduo para modelos reales (complejos)

Método 1: Integración analítica

- La mediana m , y otros cuartiles:

$$P(X < m) = P(X > m) = \int_{-\infty}^m f(x)dx = 1/2$$

- La media:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

- P.ej., si $X \sim N(\mu, \sigma)$ entonces es (complejos)

```
> qnorm(.5, mean=mu, sd=sigma)
```

```
> qnorm(p=c(0.025, .975), mean=mu, sd=sigma)
```

Método 2: Integración por Monte Carlo

Idea = generar muestras aleatorias y calcular porcentajes para aproximar probabilidades

■ Strong Law of Large Numbers

Given a function $h(x)$ and a distribution $f(x)$, want to find expected value

$$\mathbb{E}_f[h(X)] = \int_{\mathcal{X}} h(x) f(x) \, dx,$$

which converges surely to

$$\bar{h}_n = \frac{1}{n} \sum_{j=1}^n h(x_j),$$

for large n .

Una muestra

Stanislaw Ulam



Las implicaciones de Monte Carlo

- Eso significa que se puede usar Monte Carlo muestras para aproximar varias integrales:
 - Varianza y desviación típica
 - Cuartiles, incluyendo la mediana
 - Funciones de X , como la media
- En lugar de usar integrales exactas, se usa estas aproximaciones para hacer inferencia.
- Una forma de integrar: integración numérica

Método 2: Integración por Monte Carlo

- Idea = generar muestras aleatorias y calcular porcentajes para aproximar probabilidades

```
> x <- rnorm(10, mean=mu1, sd=tau1)
```

```
> mean(x<0)      [1] 1
```

Porcentaje de $x < 0$ (por que?)

```
> quantile(x, probs=c(0.025, 0.975))
```

```
    -1.14731923  -0.08842335
```

La cantidad de las muestras controla la precisión

```
> x <- rnorm(1e6, mean=mu1, sd=tau1)
```

```
> mean(x<0)      [1] 0.940683
```

```
> quantile(x, probs=c(0.025, 0.975))
```

```
    -1.5768345   0.1790194
```

Método 2: Monte Carlo integración

- Para poder implementarlo necesitamos conocer la forma exacta de la distribución a posteriori como en el ejemplo anterior con `rnorm`
- En este método cuanto más muestras generamos mejor será la aproximación
- **Sin embargo, esta situación es muy raro**
- Que haríamos cuando no podamos usar los métodos anteriores?

Homework 1.1

1. Let $X \sim U(-\pi, 1)$. Integrate this analytically for $P(a < X < b)$.

2. Now let X be a triangle:

$$f(x) = \begin{cases} cx & \text{if } 0 < x < 1/2 \\ c(1-x) & \text{if } 1/2 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

3. Draw this or use R to plot it.

4. Integrate it from 0 to 1. What does c need to be so this is a probability distribution? Calculate $P(.35 < X < .98)$

Homework 1.2

1. Assume $X \sim N(0,1)$. Use `rnorm` (numerical) to estimate $P(X < -5)$. Repeat but use `pnorm` (analytical). Why different?
2. Generate 500 samples and plot the cumulative mean of samples vs replicate. How many samples are “good enough” to approximate the mean?
3. Using samples from #2, approximate mean of the function $h(x)=x^2$. Is this equal to $\text{mean}(x)^2$?
4. [Extra: find the pdf for $Y=h(x)$ and add to histogram]

Homework 1.3

- Install JAGS 4.3.0:
<https://sourceforge.net/projects/mcmc-jags/files/JAGS/4.x/>
- Download user manual:
- https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/jags_user_manual.pdf/download
- Install R packages R2jags, coda
- Start R and ensure library(R2jags) works