Lecture 8A
MTH6102: Bayesian Statistical Methods

Eftychia Solea

Queen Mary University of London

2023

Today's agenda

Today's lecture

- Learn how simulation can be used to approximate integrals.
- Learn how to compute numerically integrals in Bayesian inference e.g., expectations, probabilities.
- Learn two integration methods
  - Basic Monte Carlo integration
  - Monte Carlo integration.

Bayesian inference

$$p(\theta \mid y) \propto p(\theta)\, p(y \mid \theta)$$

Posterior distribution $\propto$ prior distribution $\times$ likelihood

- In the Bayesian framework, all our inferences about θ are based on the posterior distribution $p(\theta \mid y)$.
- The posterior mean is

$$\hat{\theta}_{\mathrm{B}} = \int_\theta \theta p(\theta \mid y)\, d\theta.$$

Bayesian inference

- If $\theta = (\theta_1, \ldots, \theta_K)$ is a K-dimensional vector, then we might be interested in the posterior for one of the components, θ_1, say.

pdf θ_1

- The marginal posterior density is

$$p(\theta_1 \mid y) = \int \int \cdots \int f(\theta_2, \ldots, \theta_K \mid y) d\theta_2 \ldots d\theta_K.$$

- Sometimes it might be not feasible to calculate these integrals analytically.

- Simulation methods will often be helpful.

Example: Comparing two binomials

- Suppose we have data from a clinical trial of two treatments for a serious illness.
- The data are the number of deaths after each treatment.
- Let the data be k_i deaths out of n_i patients, $i = 1, 2$ for the two treatments.
- The two unknown parameters are q_1 and q_2, the probability of death with each treatment.

Example: Comparing two binomials

- We can assume that for each $i = 1, 2$

$$k_i \sim \text{Bin}(n_i, q_i)$$

- Take as independent prior distributions

$$q_i \sim \text{Beta}(\alpha_i, \beta_i), \;\; i = 1, 2$$

- Then the posterior distributions are

$$q_i \mid k_i \sim \text{Beta}(k_i + \alpha_i, n_i - k_i + \beta_i), \;\; i = 1, 2$$

Example: Comparing two binomials

- For each q_i, we have the exact posterior, so we can make exact inferences (point estimates and credible intervals) as in examples we have seen.

- Suppose we want to know the posterior probability

$$P(q_2 < q_1 \mid k_1, k_2) = \iint I(q_2 < q_1)\, p(q_1, q_2 \mid x_1, x_2)\, dq_1\, dq_2$$

- Or suppose we want to estimate the difference in proportions $\delta = q_2 - q_1$.
- There is no simple formula or beta distribution function we can use now.
- But we can use simulation (i.e. a Monte Carlo method).

Monte Carlo methods

- Monte Carlo method refers to the theory and practice of using random samples to approximate a quantity:
  - Expectations.
  - Integrals.
  - Probabilities.
  - Other summaries of distributions.
- Named due to casinos in Monte Carlo.

Basic Monte Carlo integration

- Suppose we want to evaluate the integral

$$I = \int_a^b h(x)\,dx.$$

$$= \int_a^b \underbrace{h(x)\,b-a}_{w(x)} \; \underbrace{\frac{1}{b-a}}_{f(x)} \, dx$$

- Suppose we are unable to compute I in closed form.
- We can rewrite I as

$$I = \int_a^b w(x) f(x)\,dx,$$

where $w(x) = h(x)(b-a)$, $f(x) = \frac{1}{b-a}$, $x \in [a, b]$.

Basic Monte Carlo integration

- Noticing that f is the pdf for a uniform random variable $X \sim \mathsf{U}(a,b)$
- Hence,

$$I = E[w(X)] = \int_a^b w(x) f(x)\,dx, \quad f \sim U(a,b)$$

- If we generate $X_1, \ldots, X_N$ iid from $\mathsf{U}(a,b)$, by the WLLN

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} w(X_i) \xrightarrow{P} E[w(X)] = I, \quad \text{as } N \to \infty.$$

- This is the basic Monte Carlo integration.

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} w(X_i), \quad X_1, \ldots, X_n \overset{iid}{\sim} U(a,b)$$

Basic Monte Carlo integration

$X \sim U(0,1)$

- Let $h(x) = x^3$ and

$$I = \int_0^1 x^3 \, dx = \int_0^1 x^3 f(x) dx = \mathbb{E}(X^3)$$

$f(x) \sim U[0,1]$

$f(x) = 1$

$x \in [0,1]$

- Obviously, $I = 1/4$.
- Simulate $x_1, \ldots, x_N$ from U$(0, 1)$, $N = 10,000$.
- Compute $\hat{I} = \frac{1}{10^4} \sum_{i=1}^{10^4} x_i^3 = 0.248$

Show R code

WLLN

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} h(x_i) = \frac{1}{N} \sum_{i=1}^{N} x_i^3$$

Monte Carlo integration

- A generalisation of the basic Monte Carlo integration is to estimate a quantity based on a probability distribution f.

- We want to compute

$$I = \int h(x) f(x)\, dx = E[h(X)], \quad X \sim f(x),$$

where f is the pdf of a random variable X.

Monte Carlo integration

- Now, we generate an iid random sample $X_1, \ldots, X_N$ from f and use this sample to estimate I by

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} h(X_i).$$

- By the WLLN

$$\hat{I} \xrightarrow{P} E[h(X)] = I, \quad \text{as } N \to \infty.$$

Example: Monte Carlo integration

$P(X \in A) = \mathbb{E}(I(X \in A))$

- Let $f(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}x^2\}$ be the standard normal density.
- We want

$$I = \Phi(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)\, ds, \quad X \sim N(0,1), x \in \mathbb{R}.$$

- We can rewrite

$h(s) = I(s \leq x)$

$$I = \int_{-\infty}^{\infty} h(s) f(s)\, ds,$$

where $h(s) = 1$ if $s < x$ and $h(s) = 0$ otherwise.

$$I = P(X \leq x) = \int_{-\infty}^{+\infty} \underbrace{I(s \leq x)}_{h(s)} f(x)\, dx = \mathbb{E}_f(I(X \leq x))$$

$X \sim N(0,1)$

Example: Monte Carlo integration

$$\frac{1}{N}\sum_{i=1}^{N} h(X_i) = \frac{1}{N}\sum_{i=1}^{N} I(X_i \le x)$$

- We generate $X_1, \ldots, X_N$ iid from $N(0,1)$ and compute

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} h(X_i) = \frac{\text{number of observations less than x}}{N}.$$

- If $x = 2$, then $\Phi(2) = 0.9772$ and $\hat{I} = 0.9781$ with $N = 10,000$.

Example: Comparing two binomials

- Recall that

$$k_1 \sim \text{Bin}(n_1, q_1), \quad k_2 \sim \text{Bin}(n_2, q_2)$$

- Assume the flat prior on (q_1, q_2)

$$p(q_1, q_2) = 1$$

- Then the posterior distribution $p(q_1, q_2 \mid k_1, k_2)$ is

$$p(q_1, q_2 \mid k_1, k_2) = c_1 q_1^{k_1} (1 - q_1)^{n_1 - k_1} q_2^{k_2} (1 - q_2)^{n_2 - k_2}.$$

Example: Comparing two binomials

- Note that $p(q_1, q_2 \mid k_1, k_2) = p(q_1 \mid k_1)p(q_2 \mid k_2)$.
- Thus, q_1 and q_2 are independent under the posterior.
- Also

$$p(q_1 \mid k_1) \sim \text{beta}(1 + k_1, 1 + n_1 - k_1),$$
$$p(q_2 \mid k_2) \sim \text{beta}(1 + k_2, 1 + n_2 - k_2).$$

- We want to compute $\delta = q_2 - q_1$. $= g(q_1, q_2)$
- Note that δ is random parameter with posterior density $p(\delta \mid k_1, k_2)$
- We can estimate δ using its posterior mean which is

$$\begin{aligned} I = E(\delta) &= E(q_2 - q_1) \\ &= E(g(q_1, q_2)) = \int_0^1 \int_0^1 g(q_1, q_2) p(q_1, q_2 \mid k_1, k_2)\, dq_1 dq_2, \end{aligned}$$

where $g(q_1, q_2) = q_2 - q_1$.

- Not easy to do analytically but we can use Monte Carlo integration.

$$\frac{1}{N} \sum_{i=1}^{N} g(q_{1i}, q_{2i})$$

$$(q_{i1}, q_{2i}) \sim p(q_1, q_2 \mid x_1, x_2)$$

Example: Comparing two binomials

- Hence, using Monte Carlo, we can simulate an iid sample $(Q_1^{(1)}, Q_2^{(1)}), \dots, (Q_1^{(N)}, Q_2^{(N)})$ from $p(q_1, q_2 \mid k_1, k_2)$ by drawing

$$Q_1^{(1)}, \dots, Q_1^{(n)} \text{ iid} \sim \text{beta}(1 + k_1, 1 + n_1 - k_1)$$
$$Q_2^{(1)}, \dots, Q_2^{(n)} \text{ iid} \sim \text{beta}(1 + k_2, 1 + n_2 - k_2)$$

- We can estimate I by

$$\hat{I} = \underbrace{\frac{1}{N}\sum_{i=1}^{N} g(Q_1^{(i)}, Q_2^{(i)})}_{} = \frac{1}{N}\sum_{i=1}^{N}(Q_2^{(i)} - Q_1^{(i)}). = \frac{1}{N}\sum_{i=1}^{N} \delta^{(i)}$$

Monte Carlo integration estimator of $\delta = q_2 - q_1$

Example: Comparing two binomials

- Also, note that $\delta^{(i)} = Q_2^{(i)} - Q_1^{(i)}$, $i = 1, \ldots, N$ can be viewed as an iid sample from δ.

- Then the posterior density of δ, $p(\delta \mid k_1, k_2)$, can be approximated by plotting the histogram of $\delta^{(1)}, \ldots, \delta^{(N)}$.

- A 95% quantile credible intervals of δ can be obtained by sorting the simulated values and finding the 0.025 and 0.975 sample quantiles of $\delta^{(1)}, \ldots, \delta^{(N)}$.

Board question: Comparing two binomials

- $k_1 = 8 \sim \text{Bin}(n_1, q_1)$, $k_2 = 6 \sim \text{Bin}(n_2, q_2)$.
- Assume $n_1 = n_2 = 10$.
- Describe how you would estimate $\delta = q_2 - q_1$ and $I = P(q_2 < q_1 \mid k_1, k_2)$ using simple Monte Carlo integration.
- Compute a 95% quantile credible interval for δ.

$= \mathbb{E}(I(q_2 < q_1))$

$$I = P(q_2 < q_2 | x_1, x_2) = \iint I(q_2 < q_1)\, p(q_1, q_2 | x_1, x_2)\, dq_1 dq_2$$

By MC, $\hat{I} = \frac{1}{N} \sum_{i=1}^{N} I(q_2^{(i)} < q_1^{(i)})$

Board question: binomial data, flat prior

- Let $k \sim \text{binom}(n, q)$.
- Assume flat prior on q.
- Let $n = 860$ and $k = 441$
- R code below

```
a=1
b=1
n=860
k=441
N=10000
beta.post.sample=rbeta(N, shape1=a+k,shape2=b+n-k)
gamma.sample=log((beta.post.sample/(1-beta.post.sample)))
mean(gamma.sample)
c(quantile(gamma.sample,0.025),quantile(gamma.sample,0.975))
```

Board question: binomial data, flat prior

- When this code has run, what will `beta.post.sample` contain? What will `gamma.sample` contain?

- Describe the estimator $\hat{\theta}$ for a quantity θ (which you should also determine) that would be obtained by the following R commands

```
gamma.sample=log((beta.post.sample/(1-beta.post.sample)))
mean(gamma.sample)
```

- In statistical terms, what quantity will the last line of code output?

- See also, **Question 3, final exam Jan 2023**

- A sample, $\theta_1, \ldots, \theta_N$, from the posterior density
  $p(\theta|x) \sim \text{beta}(1+x, 1+n-x)$

It contains, $\log\left(\frac{\theta_i}{1-\theta_i}\right)$ sample $i=1,\ldots,N$
which can be viewed as an iid sample from
$\log\left(\frac{\theta}{1-\theta}\right)$ (log of odds)