Bayesian Applications Of Monte Carlo Integration

Dr. Uma Ravat
PSTAT 194CS

Simple example

Simple example

Consider a box with blue and yellow marbles. We want to

1. estimate the proportion of marbles that are blue and
2. Find the probability that if we draw three marbles we get exactly two blue marbles.

Data/Sample: We draw 10 marbles with replacement and 6 are blue.

$\theta =$ proportion of marbles that are blue.

We need an estimate for θ.

Frequentist approach: parameter is unknown but fixed

By MLE, we get $\hat{\theta} = 0.6$ (verify this on your own).

Next, before drawing three more marbles with replacement, what is the probability of two out of three being blue?

$X =$ Number of marbles out of 3 that are blue

$X \sim bino(n = 3, p = \hat{\theta} = 0.6)$

Required probability $= \binom{3}{2} 0.6^2 (1 - 0.6)^{(3-1)} = 0.4323$

And you can create a confidence interval: for eg: "I am 95% confident that the proportion of blue marbles that I will get is between 40% and 46%"

Drawback of frequentist approach: assuming the proportion is exactly 0.6 after just 10 observations and use this proportion for all calculations/inferences.

Bayesian approach : Parameters have probability distribution

In Bayesian approach, the proportion of blue marbles in the box, θ, will be a random variable and have a probability distribution.

After observing the data, we know this probability distribution for θ will be tallest at 0.6, but there is high probability that this proportion is .55 or 0.65.

We take into account the data we observed, but realize that the true proportion could very well be some other value.

Bayesian approach: Step 1. Prior: Distribtuion of θ before observing data

Assume equally likely distribution for proportion of blue marbles.

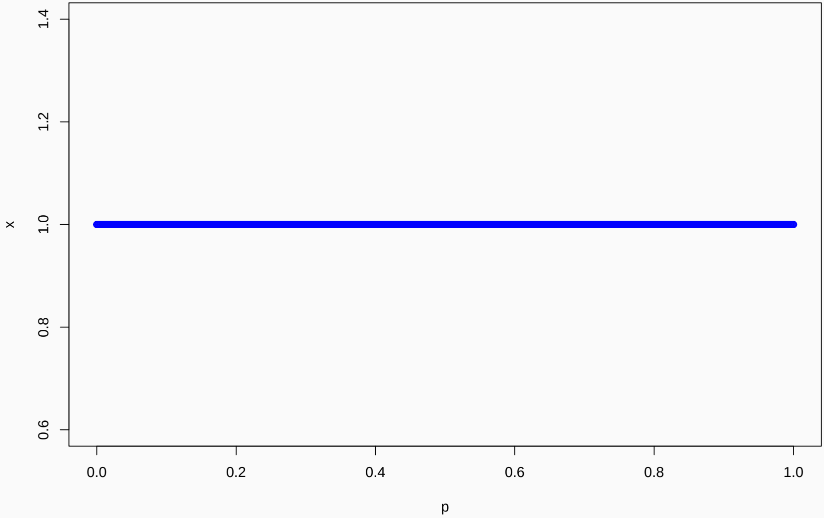
Assume the prior distribution for θ , $P(\theta)$ is $\mathit{Uniform}(0, 1) = \beta(1, 1)$.

Recall p.d.f of $\beta(\alpha, \beta)$ distribution is

$$\frac{1}{B(\alpha, \beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

This prior distribution looks like

Uniform or $\beta(1, 1)$ prior



Bayesian approach: Step 2. Decide a model for the data

The probability(likelihood) of getting $X = 2$ blue marbles when we pick three marbles from the urn with θ proportion of blue marbles will be

$$P(X = 2 \mid \theta) = \binom{3}{2}\theta^2(1-\theta)^1$$

Basic Bayesian Terminology

The starting point for Bayesian models is **Bayes Theorem**

$$P(\theta \mid x) = \frac{P(\theta)P(x \mid \theta)}{P(x)}$$

where θ is the parameter(proportion of blue marbles in the box) and $X = x$ is the data we observed.

- **Posterior:** $P(\theta \mid x)$ is the probability distribution of the parameter θ that we want to find out.
- **Likelihood:** $P(x \mid \theta)$ is the probability of observing the observed data for a given value of θ.
- **Prior:** $P(\theta)$ is the prior probability distribution of θ.
- **Marginal probability:** $P(x)$ is the probability of observing the data we observed regardless of the value of θ. By Law of Total Probability, this is the integral of the numerator across all possible values of θ. This is a constant.

Bayesian approach: Step 3: Posterior distribution of θ.

Since Marginal probability $P(x)$ is a constant wrt θ, the posterior distribution of θ is proportional to the likelihood times the prior distribution

$$P(\theta \mid x) \propto P(\theta)P(x \mid \theta)$$

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

Back to the marble example

The likelihood function is:

$$P(X = 2 \mid \theta) \propto \theta^2(1-\theta)^1$$

The prior distribution of $\theta \sim \beta(1,1)$

$$P(\theta) = \frac{1}{B(1,1)}\theta^{1-1}(1-\theta)^{1-1} = 1$$

The posterior

$$P(\theta \mid x) \propto P(\theta)P(x \mid \theta) = \theta^2(1-\theta)^1$$

To make this posterior into a distribution, we must find the normalizing constant

$$\int_0^1 \theta^2(1-\theta)^1 d\theta$$

Normalizing constant for posterior

Recall: $\int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta = B(\alpha,\beta)$

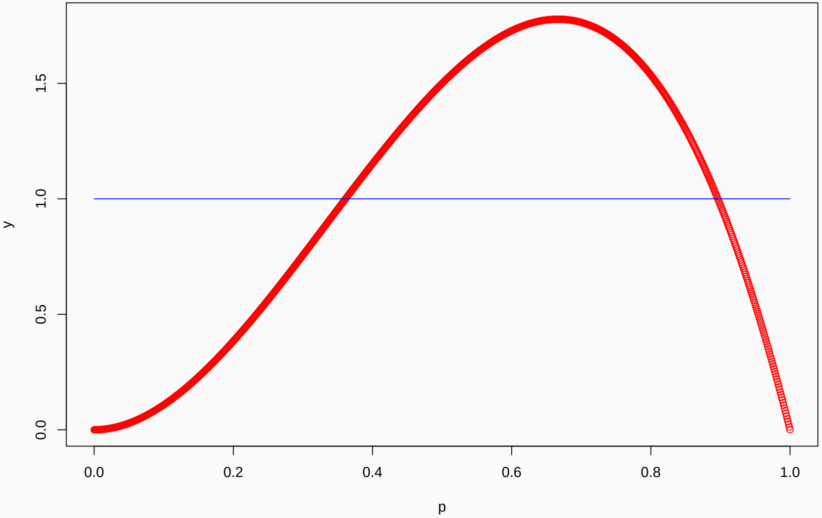
This means the normalizing constant for our posterior distribution is $\int_0^1 \theta^{(3-1)}(1-\theta)^{(2-1)}d\theta = B(\alpha=3,\beta=2)$

Thus, the p.d.f of our posterior distribution can be written as

$$P(\theta \mid x) = \frac{1}{B(3,2)}\theta^{(3-1)}(1-\theta)^{(2-1)}$$

Posterior distribution is a beta distribution with $\alpha = 3$ and $\beta = 2$

Prior and Posterior



Making predictions with the posterior distribution

Find the probability that if we draw three marbles we get exactly two blue.

The answer depends on θ: $g(X = 2|\theta) = \binom{3}{2}\theta^2(1-\theta)^1$ where θ is now a random variable with pdf given by
$P(\theta \mid x) = \frac{1}{B(3,2)}\theta^{(3-1)}(1-\theta)^{(2-1)} \sim \beta(\alpha = 3, \beta = 2)$
We use the posterior distribution of θ to make this probability prediction.

A few strategies for dealing with the unknown θ

1. Use a single value: the mean of the posterior distribution
2. Use a single value: the mode of the posterior distribution
3. Find the expected value $E_\theta(g(X = 2|\theta))$ of the probability analytically
4. Estimate the expected value of the probability $E_\theta(g(X = 2|\theta))$ via Monte Carlo simulation.

1. Use a single value: The mean of the posterior distribution

Posterior distribution is a beta distribution with $\alpha = 3$ and $\beta = 2$

So the mean of the posterior is $\frac{\alpha}{\alpha+\beta} = \frac{3}{3+2} = 0.6$

Find the probability that if we draw three marbles we get exactly two blue.

The answer is: $\binom{3}{2}\theta^2(1-\theta)^1 = \binom{3}{2}0.6^2(1-0.6)^1 = 0.432$

2. Use a single value: the mode of the posterior distribution

Posterior distribution is a beta distribution with $\alpha = 3$ and $\beta = 2$

So the mode of the posterior is $\frac{\alpha-1}{\alpha+\beta-2} = \frac{3-1}{3+2-2} = 0.6667$

Find the probability that if we draw three marbles we get exactly two blue.

The answer is:
$\binom{3}{2}\theta^2(1-\theta)^1 = \binom{3}{2}0.6667^2(1-0.6667)^1 = 0.4444444$

3. Find the expected value $E_\theta(g(X = 2|\theta))$ of the probability analytically

We want to find

$$E_\theta(g(X = 2|\theta)) = \int_0^1 \binom{3}{2}\theta^2(1-\theta)^1 \frac{1}{B(3,2)}\theta^2(1-\theta)^1 d\theta$$

Exercise:

Evaluate this integral using calculus

4. Estimate the expected value of the probability $E_\theta(g(X = 2|\theta))$ via Monte Carlo simulation.

$$E_\theta(g(X = 2|\theta)) = \int_0^1 \binom{3}{2}\theta^2(1-\theta)^1 \frac{1}{B(3,2)}\theta^2(1-\theta)^1 d\theta$$

$$E_\theta(g(X = 2|\theta)) = \int_A g(X = 2|\theta)P(\theta|X = 2)d\theta \approx \frac{1}{N}\sum_{i=1}^{N} g(\theta_i)$$

Here:

1. $P(\theta|x)$ is the posterior $Beta(3,2)$ distribtion.
2. The function $g(X = 2|\theta$ is the probability of getting 2 blue marbles from 3 marbles chosen. $g(X = 2|\theta) = \binom{3}{2}\theta^2(1-\theta)^1$

5. Activity:

1. Write code for Monte Carlo integration to estimate this expected value $E_\theta(g(X = 2|\theta))$
2. Evaluate this integral using calculus
3. Compare the MC estimate of expected value with the true expected calculated by hand.

Is the MC estimate a good approximation?

The Beta-Binomial model

In the example we had

- The prior was a Beta distribution,
- The observed data came from a binomial distribution
- The posterior distribution also ended up being a Beta distribution(with different parameters from the prior)

This was no coincidence. If the prior distribution of the parameter θ is a beta distribution with parameters α and β

$$P(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

And the observed data come from a binomial distribution. If probability of a success in each trial is θ and there are N trials and x successes, then the Likelihood function of x successess is

$$P(X = x \mid \theta, N) \propto \theta^{x}(1-\theta)^{(N-x)}$$

Then the posterior distribution will be a Beta distribution with parameters $\alpha + x$ and $\beta + N - x$

$$\begin{aligned} P(\theta \mid X = x) &= Beta(\alpha + x, \beta + N - x) \\ &= \frac{1}{B(\alpha + x, \beta + N - x)} \theta^{(\alpha + x - 1)} (1 - \theta)^{(\beta + N - x - 1)} \end{aligned}$$

And the Beta distribution and Binomial Distribution are called **conjugate distributions**

Conjugate priors

If the prior and posterior are the same family of probability distributions then **the prior is called a conjugate prior for the likelihood**.

The beta and binomial distributions are a pair of conjugate distributions.

(See Wikipedia for other important conjugate pairs of conjugate distributions)

Bayesian Models and computation

- The Bayesian model is $\theta \sim g(\theta)$ where g is a *prior distribution*.
- We have the model for the data to be $X_1, X_2, \ldots, X_n \sim f(X = x|\theta)$.
- Observe data $x_1, x_2, \ldots, x_n$. This gives us the *likelihood*

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

- We base inferences on the posterior distribution

$$g(\theta|\text{data}) \propto \prod_{i=1}^{n} f(x_i|\theta) \cdot g(\theta) = \frac{L(\theta)g(\theta)}{C}.$$

- The normalizing constant C is

$$C = \int \prod_{i=1}^{n} f(x_i|\theta) \cdot g(\theta)\, d\theta.$$

We summarize the posterior distribution when it comes time to make inferences.

Bayesian Inference

[Posterior mean:]

$$\mu = E\left(\theta \middle|\, \text{data}\right) = \int \theta g(\theta|\text{data})\, d\theta,$$

$$= \frac{\int \theta\, L(\theta) g(\theta)\, d\theta}{\int L(u) g(u)\, du}$$

[Posterior spread:] variance

$$E\left[(\theta-\mu)^2 \,|\, \text{data}\right] = \int (\theta-\mu)^2\, g(\theta|\text{data})\, d\theta,$$

$$= \frac{\int (\theta-\mu)^2\; L(\theta) g(\theta)\, d\theta}{\int L(u) g(u)\, du},$$

[Posterior probabilities:]

$$\begin{aligned}\mathbb{P}(\theta \in A | \, \mathrm{data}) &= \int_A g(\theta|\mathrm{data})\, d\theta, \\ &= \frac{\int 1\,(\theta \in A)\; L(\theta)g(\theta)\, d\theta}{\int L(u)g(u)\, du},\end{aligned}$$

Concluding remarks.

For many problems the analytic math is intractable or difficult to solve

For these types of problems, we rely on using the Monte Carlo methods to estimate the desired quantity.

Monte Carlo methods are frequently used in the context of Bayesian Statistcs because the parameter is a random variable and solving problems often involve some complex integrals.