STAT 431 — Applied Bayesian Analysis — Course Notes

Bayesian Computation: Monte Carlo

Fall 2022

Numerical integration works well for low-dimensional problems (few parameters).

For high-dimensional problems, simulation is often better ...

Any method using randomized simulation (sampling) for approximation is called a **Monte Carlo** method.

Independent Sampling

Idea: Randomly generate samples from the posterior.

Then use the **empirical distribution** of the samples to estimate aspects of the actual posterior.

Let data be $\boldsymbol{y}$, and let the sample from the posterior for parameter θ be

$$\theta^{(1)}, \ \cdots \ , \ \theta^{(S)}$$

A posterior sample may be used to approximate many things —

- a mean:

$$\mathrm{E}(\theta \mid \boldsymbol{y}) \;\approx\; \frac{1}{S}\sum_{s=1}^{S}\theta^{(s)} \;=\; \text{sample mean of } \theta^{(s)}\text{s}$$

- a variance:

$$\mathrm{Var}(\theta \mid \boldsymbol{y}) \;\approx\; \text{sample variance of } \theta^{(s)}\text{s}$$

- a (lower) quantile $Q_\theta(\tau)$:

  [Draw quantile ...]

  Round τS to the nearest integer: $[\tau S]$

  Then use the $[\tau S]$th order statistic from the sample.

  (The order statistics are the $\theta^{(s)}$s re-ordered from least to greatest.)

- a 95% equal-tailed credible interval: form an estimated version of

$$\big(Q_\theta(0.025),\ Q_\theta(0.975)\big)$$

- probability of $H_0 : \theta \in \Theta_0$

$$\text{fraction of } \theta^{(s)}\text{s in } \Theta_0 \quad = \quad \frac{1}{S}\sum_{s=1}^{S} I\big(\theta^{(s)} \in \Theta_0\big)$$

where I represents the indicator function:

$$I(\theta \in \Theta_0) \quad = \quad \begin{cases} 1 & \theta \in \Theta_0 \\ 0 & \theta \notin \Theta_0 \end{cases}$$

- a mean of a function of θ:

$$\mathrm{E}\big(g(\theta) \mid \boldsymbol{y}\big) \quad \approx \quad \frac{1}{S}\sum_{s=1}^{S} g\big(\theta^{(s)}\big)$$

This is valid because

$$g\big(\theta^{(1)}\big), \quad \ldots, \quad g\big(\theta^{(S)}\big)$$

is a random sample from the posterior of $g(\theta)$.

Since these approximations are subject to random sampling variability, we need an assessment of their accuracy.

For example, the approximation

$$\frac{1}{S}\sum_{s=1}^{S} g\big(\theta^{(s)}\big) \qquad \text{of} \qquad \mathrm{E}(g(\theta) \mid \boldsymbol{y})$$

has an approximate *standard error* of

$$\frac{sd_g}{\sqrt{S}} \qquad \text{where} \qquad sd_g \;=\; \sqrt{\text{sample var. of } g\big(\theta^{(s)}\big)\text{s}}$$

(why?)

This is the **Monte Carlo error** for the mean approximation.

Example: Jevons's Coins Comparison

For $n_1 = 24$ coins minted before 1830:

$$\bar{y}_1 \; = \; 7.8730 \qquad s_1 \; = \; \text{sample s.d.} \; = \; 0.05353$$

For $n_2 = 123$ newer coins (1860's):

$$\bar{y}_2 \; = \; 7.9725 \qquad s_2 \; = \; \text{sample s.d.} \; = \; 0.01409$$

Assume independent samples from $\mathrm{Normal}(\mu_1, \sigma_1^2)$ and $\mathrm{Normal}(\mu_2, \sigma_2^2)$.

We want inference about $\mu_1 - \mu_2$.

Take “independent” “standard” (product-Jeffreys) priors:

$$\mu_1,\, \sigma_1^2,\, \mu_2,\, \sigma_2^2 \quad \sim \quad \frac{1}{\sigma_1^2} \cdot \frac{1}{\sigma_2^2}\, d\mu_1\, d\sigma_1^2\, d\mu_2\, d\sigma_2^2$$

Then it turns out that (later?)

$$(\mu_1, \sigma_1^2) \quad \text{and} \quad (\mu_2, \sigma_2^2) \quad \text{are posterior-independent}$$

with posterior distributions ($i = 1, 2$)

$$\mu_i \mid \sigma_i^2, \boldsymbol{y} \;\; \sim \;\; \mathrm{Normal}(\bar{y}_i, \sigma_i^2/n_i)$$

$$\sigma_i^2 \mid \boldsymbol{y} \;\; \sim \;\; \mathrm{InvGamma}\left(\frac{n_i - 1}{2},\, \frac{n_i - 1}{2} s_i^2\right)$$

where $\boldsymbol{y}$ represents both samples together.

As it turns out, under the posterior, μ_1 and μ_2 have independent (location-scale) t distributions. (Later?)

Nonetheless, $\mu_1 - \mu_2$ has no simple-form posterior density.

We can easily randomly sample from the posterior distribution of $\mu_1 - \mu_2$ using R ...

R Example 3.3:

Comparing Normal Means: Independent Sampling

The 95% *Welch interval* is an approximate frequentist confidence interval for $\mu_1 - \mu_2$, used here for comparison:

$$\bar{y}_1 - \bar{y}_2 \quad \pm \quad t_{0.025,\,\mathrm{df}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where

$$\mathrm{df} \;=\; \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)}$$