STA 360/602L: MODULE 3.1

MONTE CARLO APPROXIMATION AND SAMPLING

DR. OLANREWAJU MICHAEL AKANDE

MONTE CARLO APPROXIMATION

- Monte Carlo integration is very key for Bayesian computation and using simulations in general.

- While we will focus on using Monte Carlo integration for Bayesian inference, the development is general and applies to any pdf/pmf $p(\theta)$.

- For our purposes, we will want to evaluate expectations of the form

$$H = \int h(\theta) \cdot p(\theta) d\theta,$$

for many different functions $h(.)$ (usually scalar for us).

MONTE CARLO APPROXIMATION

- Procedure:

  1. Generate a random sample $\theta_1, \ldots, \theta_m \overset{ind}{\sim} p(\theta)$.

  2. Estimate H using

  $$\bar{h} = \frac{1}{m} \sum_{i=1}^{m} h(\theta_i).$$

- In this course, $p(\theta)$ would often be the posterior distribution $\pi(\theta|y)$.

MONTE CARLO APPROXIMATION

- We have $\mathbb{E}[h(\theta_i)] = H$.

- Assuming $\mathbb{E}[h^2(\theta_i)] < \infty$, so that the variance of each $h(\theta_i)$ is finite, we have

1. LLN: $\bar{h} \xrightarrow{a.s.} H$.

2. CLT: $\bar{h} - H$ is is asymptotically normal, with asymptotic variance

$$\frac{1}{m} \int (h(\theta) - H)^2 p(\theta) d\theta,$$

which can be approximated by

$$v_m = \frac{1}{m^2} \sum_{i=1}^{m} (h(\theta_i) - \bar{h})^2.$$

- $\sqrt{v_m}$ is often called the Monte Carlo standard error.

MONTE CARLO APPROXIMATION

- That is, generally, taking large Monte Carlo sample sizes m (in the thousands or tens of thousands) can yield very precise, and cheaply computed, numerical approximations to mathematically difficult integrals.

- What this means for us: we can approximate just about any aspect of the posterior distribution with a large enough Monte Carlo sample.

MONTE CARLO APPROXIMATION

- For samples $\theta_1, \ldots, \theta_m$ drawn iid from $\pi(\theta|y)$, as $m \to \infty$, we have

  - $\bar{\theta} = \dfrac{1}{m} \displaystyle\sum_{i=1}^{m} \theta_i \to \mathbb{E}[\theta|y]$

  - $\hat{\sigma}_\theta = \dfrac{1}{m-1} \displaystyle\sum_{i=1}^{m} (\theta_i - \bar{\theta})^2 \to \mathbb{V}[\theta|y]$

  - $\dfrac{1}{m} \displaystyle\sum_{i=1}^{m} 1[\theta_i \leq c] = \dfrac{\#\theta_i \leq c}{m} \to \Pr[\theta \leq c|y]$

  - $[\dfrac{\alpha}{2}$th percentile of $(\theta_1, \ldots, \theta_m), (1 - \dfrac{\alpha}{2})$th percentile of $(\theta_1, \ldots, \theta_m)]$
    $\to 100 \times (1 - \alpha)$ quantile-based credible interval.

BACK TO BIRTH RATES

- Suppose we randomly sample two "new" women, one with degree and one without.

- To what extent do we expect the one without the degree to have more kids than the other, e.g. $\tilde{y}_1 > \tilde{y}_2 | y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}$?

- Using R,

```
set.seed(01222020)
a <- 2; b <- 1; #prior
n1 <- 111; sumy1 <- 217; n2 <- 44; sumy2 <- 66 #data
y1_pred <- rnbinom(100000,size=(a+sumy1),mu=(a+sumy1)/(b+n1))
y2_pred <- rnbinom(10000,size=(a+sumy2),mu=(a+sumy2)/(b+n2))
mean(y1_pred > y2_pred)
```

```
## [1] 0.48218
```

```
mean(y1_pred == y2_pred)
```

```
## [1] 0.21842
```

BACK TO BIRTH RATES

- That is, $\Pr(\tilde{y}_1 > \tilde{y}_2 | y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}) \approx 0.48$ and $\Pr(\tilde{y}_1 = \tilde{y}_2 | y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}) \approx 0.22$.

- Notice that strong evidence of difference between two populations does not really imply the difference in predictions is large.

MONTE CARLO APPROXIMATION

- This general idea of using samples to "approximate" averages (expectations) is also useful when trying to approximate posterior predictive distributions.

- Quite often, we are able to sample from $p(y_i|\theta)$ and $\pi(\theta|\{y_i\})$ but not from $p(y_{n+1}|y_{1:n})$ directly.

- We can do so indirectly using the following Monte Carlo procedure:

$$\text{sample } \theta^{(1)} \sim \pi(\theta|\{y_i\}), \quad \text{then sample } y_{n+1}^{(1)} \sim f(y_{n+1}|\theta^{(1)})$$
$$\text{sample } \theta^{(2)} \sim \pi(\theta|\{y_i\}), \quad \text{then sample } y_{n+1}^{(2)} \sim f(y_{n+1}|\theta^{(2)})$$
$$\vdots \qquad\qquad\qquad\qquad\qquad \vdots$$
$$\text{sample } \theta^{(m)} \sim \pi(\theta|\{y_i\}), \quad \text{then sample } y_{n+1}^{(m)} \sim f(y_{n+1}|\theta^{(m)}).$$

- The sequence $\{(\theta, y_{n+1})^{(1)}, \dots, (\theta, y_{n+1})^{(m)}\}$ constitutes m independent samples from the joint posterior of (θ, Y_{n+1}).

- In fact, $\{y_{n+1}^{(1)}, \dots, y_{n+1}^{(m)}\}$ are independent draws from the posterior predictive distribution we care about.

BACK TO BIRTH RATES

- Let's re-compute $\Pr(\tilde{y}_1 > \tilde{y}_2 | y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2})$ and $\Pr(\tilde{y}_1 = \tilde{y}_2 | y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2})$ using this method.

- Using R,

```
set.seed(01222020)
a <- 2; b <- 1; #prior
n1 <- 111; sumy1 <- 217; n2 <- 44; sumy2 <- 66 #data
theta1_pred <- rgamma(10000,219,112); theta2_pred <- rgamma(10000,68,45)
y1_pred <- rpois(10000,theta1_pred); y2_pred <- rpois(10000,theta2_pred)
mean(y1_pred > y2_pred)
```

```
## [1] 0.4765
```

```
mean(y1_pred == y2_pred)
```

```
## [1] 0.2167
```

- Again, $\Pr(\tilde{y}_1 > \tilde{y}_2 | y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}) \approx 0.48$ and $\Pr(\tilde{y}_1 = \tilde{y}_2 | y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}) \approx 0.22$.

WHAT'S NEXT?

Move on to the readings for the next module!