Lab 4

PSTAT 115, Fall 2019

October 24, 2019

Objectives

- Predictive Posterior Distribution
- Monte Carlo Method

Posterior Predictive Distribution

Notation

1. y_{obs}: observed data $y_1, y_2, \cdots, y_n$.
2. y^{rep}: replicated data that could have been observed. y^{rep} is from the same model that produced y_{obs} and related to a posterior distribution.
3. predicative value of y: any future observable value.

Difference: y^{rep} is a special case of the predicative value. Example, if the model has explanatory variables x, then y^{rep} is generated by the same x as y^{obs}, whereas a predictive value of y might have its own new x.

PPD

Posterior predictive distribution:

$$p(y^{rep}|y_{obs}) = \int p(y^{rep}|\theta)p(\theta|y_{obs})d\theta.$$

It has both uncertainty from the sampling of θ, and the uncertainty from y^{rep} if θ is known.

Monte Carlo

Monte Carlo is a simulation method that uses random numbers to solve many computational problems.

Monte Carlo Method for Computing Integrals

- $\bar{\theta} = \sum_{s=1}^{S} \theta^{(s)}/S \rightarrow \mathrm{E}[\theta|y_1, \ldots, y_n]$
- $\sum_{s=1}^{S} \left(\theta^{(s)} - \bar{\theta}\right)^2/(S-1) \rightarrow \mathrm{Var}[\theta|y_1, \ldots, y_n]$
- $\#\left(\theta^{(s)} \leq c\right)/S \rightarrow \Pr(\theta \leq c|y_1, \ldots, y_n)$
- the α-percentile of $\left\{\theta^{(1)}, \ldots, \theta^{(S)}\right\} \rightarrow \theta_\alpha$

Monte Carlo integration

We can approximate integrals through simulation!

- Want to estimate the integral: $I = \int_a^b f(x)dx$
- Make an estimate: $\tilde{I} = \frac{f(x)}{p(x)}$, where $p(x)$ is a density function on $[a, b]$.
- So, $I = \int_a^b \tilde{I}p(x)dx = E_{p(x)}(\tilde{I}) \approx \frac{1}{N}\sum_1^N \tilde{I}(x_j)$