Bayesian Analysis

Dr Niamh Cahill

Bayesian Inference

Bayesian Inference

- Bayesian point estimates are often given by:
  - the posterior mean $E(\theta|y)$
  - or the posterior median θ^* with $P(\theta < \theta^*|y) = 0.5$
- Uncertainty is quantified with credible intervals (CIs)
  - An interval is a 95% credible interval if the posterior probability the θ is in the interval is 0.95.
  - Often quantile based, given by posterior quantiles with $P(\theta < \theta_{\alpha/2}|y) = P(\theta > \theta_{1-\alpha/2}|y) = \alpha/2$

Bayesian Inference for Happiness example

- $\theta | y \sim Beta(y + a, n - y + b)$
- Posterior mean $E(\theta|y) = \frac{y+a}{n+a+b}$
- For quantile estimates we can use `qbeta()` in `R`

```
## data
n = 20
y = 14

## prior parameters
a = 1
b = 1

## posterior parameters
a_post = y + a
b_post = n - y + b

## posterior mean
(y+a)/(n+a+b)

## quantiles
qbeta(c(0.025,0.5,0.975),a_post,b_post)
```

Simulation-based inference

- The general idea in simulation-based inference: We can make inference about a parameter θ , using a sample $\{\theta^{(1)} \dots \theta^{(S)}\}$ from its probability distribution.
- Assessing the properties of a target (e.g., posterior) distribution by generating representative samples is called Monte Carlo simulation.
- Based on the law of large numbers we know that: $\frac{1}{S}\sum_{s=1}^{S}\theta^{(s)} = E(\theta)$ as sample size $S \to \infty$
  - The error in the MC approximation goes to zero as $S \to \infty$ because $\frac{var(\theta)}{S} \to 0$
- Just about any aspect of the distribution of θ can be approximated arbitrarily exactly with a large enough Monte Carlo sample, e.g.
  - the α-percentile of the distribution of θ
  - $Pr(\theta \geq x)$ for any constant x

Simulation-based inference for the Happiness example

For the Happiness example, we can approximate the mean and quantiles of θ using samples from a $Be(y + a, n - y + b)$ distribution (i.e., the posterior)

```
## data
n = 20
y = 16

## prior parameters
a = 1
b = 1

## posterior parameters
a_post = y + a
b_post = n - y + b

## sample
samp_theta <- rbeta(1000,a_post,b_post)

## sample mean and quantiles
mean(samp_theta)
quantile(samp_theta, probs = c(0.025,0.5,0.975))
```

Monte Carlo approximation: some more details

- With a simulation, it also becomes very easy to analyze the distributions of any function of your parameter,
  - e.g. the distribution of the odds $\frac{\theta}{1-\theta}$ by using samples from $\frac{\theta^{(s)}}{1-\theta^{(s)}}$

```
## sample
samp_theta <- rbeta(1000,a_post,b_post)

## get odds based on samples
samp_odds <- samp_theta/(1-samp_theta)

## sample mean and quantiles
mean(samp_odds)
quantile(samp_odds, probs = c(0.025,0.5,0.975))
```

Class Exercise

We are interested in the proportion of people that approve of the Irish government's pandemic response. Suppose you surveyed a sample of $n = 50$ people (students and staff) at Maynooth University and ($y = 20$) responded saying they approve. You wish to assume a Binomial likelihood for these data such that $y \sim Binomial(n, \theta)$

$$p(y|\theta) = c\theta^y(1-\theta)^{n-y} \text{ with } c = \binom{n}{y}$$

Now suppose a previous study carried out at another university found that the approval proportion was 0.5 ± 0.1 and you wish to define a Beta prior that incorporates this prior information such that $\theta|a, b \sim Beta(a, b)$

$$p(\theta|a, b) = k\theta^{a-1}(1-\theta)^{b-1} \text{ with } k = \frac{1}{B(a, b)}$$

Recall if $\theta \sim Be(a, b)$ then $E(\theta) = \frac{a}{a+b}$ and $Var(\theta) = \frac{ab}{(a+b)^2(a+b+1)}$

1. What values would you assign to a and b?
2. Write down the posterior distribution for θ
3. What is the expected value of θ based on the posterior distribution?