베이지안 2주차

Bayesian Approach

- A parameter θ is viewed as a random variable whose distribution is unknown.
- Data are observed from the realized sample
- Goal : Estimate the distribution of θ conditional on the observed data, the posterior distribution of θ
- Inference is based on summaries of the posterior distirubtion of θ
  - Induction from $P(\theta|data)$, starting with $P(\theta)$
  - $P(\theta)$ is the prior distirbution of the parameter (before the data are observed) and $P(\theta|data)$ is the posterior distribution of the parameter (after the data are observed).
  - Broad descriptions of the posteriror distribution such as means and quantiles.
- The idea is to assume a prior probability distribution of θ; that is, a distribution representing the plausibility of each possible value of θ before the data are observed.
- To make inferences about θ, one simply considers the conditional distribution of θ given the observed data, referred to as the posterior distirubiotn representing the plausibility of each possible value of θ after seeing the data.
- This provides a coherent framework for makign inferences about unknown parameter θ as well as any future data or missing data, and for makign rational decisions based on such inferences.

Notation

- θ : parameter
- y : observed data
- $p(y|\theta)$: likelihood function of y
- $p(\theta)$: prior distribution (before data observed)
- $p(\theta|y)$: posteriror distirubiton of θ given y (after data observed)

Bayes' Theorem

- Bayes' Theorem $p(\theta|y) = \frac{p(\theta,y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}$ where $p(y)$ is marginal distribution of y and either $p(y) = \sum_\theta p(\theta)p(y|\theta)$ or $p(y) = \int p(\theta)p(y|\theta)d\theta$
- In calculating, $p(\theta|y) \propto p(y|\theta)p(\theta)$

Bayesian Modeling

1. Model specification $p(y|\theta)$: likelihood function of y $p(\theta)$: prior distribution of θ
2. Performing inference $p(\theta|y)$: posterior distirbution of θ given y $p(\theta|y) \propto p(y|\theta)p(\theta)$

- How to model?
  - analytically-only possible for certain models
  - using simulation when we are not able to write down the exact form of the posterior density

3. Inference results

Motivating Examples : Spelling Correction

- Suppose that someone types `radom`. How to should that be read? It could be a mispelling or mistyping of `random` or `radon` or some other alternative, or it could be the intentional typing of `radom`. What is the probability that `radom` actually means `random`?

- If we label y as the data and θ as the word that the person was intending to type, then: $P(\theta|y =' radom') \propto p(\theta)p(y =' random'|\theta)$
- For simplicity, we consider only three possibilities for the intened word. $\theta($random, radon or radom$)$
- We compute the posterior probability of interest by first computing the unnormalized density for all three valeus of θ and then normalizing: $P(\theta_1|'radom') = \frac{P(\theta_1)P('radom'|\theta_1)}{\sum_{j=1}^{3} P(\theta_j)P('radom'|\theta_j)}$ where $\theta_1 =$ random, $\theta_2 =$ radon, $\theta_3 =$ radom
- Prior information strongly affect posterior rhater than liklihood function

Binomial Model

- Goal : estimate an unknown proportion from the results of a sequence of 'bernoulli tirals' (data $y_1, ...y_n$ that are either 1s or 0s)
- Assume that the data arise from a sequence of n independent trials or draws from a large population where each trial is classified as a "success" ($y_i = 1$) or a "failure" ($y_i = 0$)
- We can characterize the data by the toal number of success, denoted by y, in n trials.
- Binomial sampling model $p(y|\theta) = Bin(y|n, \theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$ where the parameter θ represents the proportion of successes in the population (equivalently, the probability of success in each trial).

Frequentist

- $L(\theta|y) = \frac{n!}{y!(n-y)!}\theta^y(1-\theta)^{n-y}$
- $\frac{\partial logL}{\partial\theta} = \frac{y}{\theta} - \frac{n-y}{1-\theta}$
- $\hat{\theta}_{ML} = \frac{y}{n} = \bar{y}$ (sample proportion)
- $\hat{\theta}_{ML} \sim N(\theta, \frac{\theta(1-\theta)}{n})$

Bayesian

- First, we need to specify the prior distribution for θ
  - For now, we assume $p(\theta) = Unif(0, 1)$
- Second: apply Bayes' Rule $p(\theta|y) \propto p(y|\theta)p(\theta)$
  $= \binom{n}{y}\theta^y(1-\theta)^{n-y}$
  $\propto c\theta^y(1-\theta)^{n-y}$
  - How to calculate c
    1. use definition $\int_0^1 c\theta^y(1-\theta)^{n-y}d\theta = 1$ $c = \frac{1}{p(y)}$
    2. $\theta^y(1-\theta)^{n-y} = \theta^{y+1-1}(1-\theta)^{n-y+1-1}$
       $\sim Beta(y+1, n-y+1)$
  - Posterior distribution : $Beta(y+1, n-y+1)$
- Posetrior mean $E[\theta|y] = \int_0^1 \theta p(\theta|y)d\theta$
  $= \frac{y+1}{n+2} = \hat{\theta}_{Bayes}$
  $= \frac{n}{n+2}\frac{y}{n} + \frac{2}{n+2}\frac{1}{2}$
  $= \frac{n}{n+2}\hat{\theta}_{MLE} + \frac{2}{n+2}($ prior mean$)$
  - Posterior mean is the weighted average of MLE & prior mean
  - It means
    * n (sample size) is large, $E[\theta|y] \to MLE$
    * n (sample size) is small, $E[\theta|y] \to$ prior mean

Summary

- Binomial model: $p(y|\theta) = Bin(y|n, \theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$ where the parameter θ represents the proportion of successes in the population.
- Prior distribution of θ: $\theta\ Unif(0, 1)$
- Posterior distribution of θ given y: $\theta|y \sim Beta(y+1, n-y+1)$
- Posterior mean $E[\theta|y] = \frac{y+1}{n+2}$ -> Weighted average of sample mean and prior mean

Binomial Model with Beta Piror

- Use the different prior distribution $p(\theta) = Beta(\alpha, \beta)$
- Posterior distribution $p(\theta|y) \propto p(y|\theta)p(\theta)$
  $= (\binom{n}{y}\theta^y(1-\theta)^{n-y})(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1})$
  $\propto \theta^y(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1}$
  $= \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$
  $\sim Beta(y+\alpha, n-y+\beta)$
  - $p(\theta|y) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$
- Compare $unif(0,1) = Beta(1,1) \to \theta|y \sim Beta(y+1, n-y+1)$
- $E[\theta|y] = \frac{y+\alpha}{n+\alpha+\beta} = \frac{n}{n+\alpha+\beta}\frac{y}{\alpha} + \frac{\alpha+\beta}{n+\alpha+\beta}\frac{\alpha}{\alpha+\beta}$
  - It is a weighted average of sample mean & prior mean
  - It means
    * n (sample size) is large, $E[\theta|y] \to MLE$
    * n (sample size) is small, $E[\theta|y] \to$ prior mean
    * $(\alpha+\beta)$: amount of prior information, $(\alpha+\beta)$ is large, $E[\theta|y] \to$ prior mean $E[\theta]$

Example: Placenta Previa

- Placenta previa is an unusual pregnancy condition in which the placenta is implanted very low in the uteras, obstructing the fetus from a normal vaginal delivery.
- A study of the sex of placentas previa births in Genrmany found that there were 437 females among 980 births.
- How much evidence does this data provide for the claim that the proportion of female births in the population is less than the proportion of female births in the general population, which is approximately 0.485?

Frequentist

- $H_0 : \theta = 0.485$
- $H_1 : \theta < 0.485$

```
theta_ml = 437 / 980
theta_0 = .485
z = (theta_ml - theta_0) / sqrt(theta_0 * (1 - theta_0) / 980)
p_value = pnorm(z)
p_value
```

```
## [1] 0.007182601
```

- Reject H_0
- Let θ be the probability of a female births among placenta previa pregnancies.
  - Assuming a uniform prior what is $p(\theta|y)$ $\theta|y \sim Beta(y+1, n-y+1)$
    $= Beta(438, 544)$
  - What is the posterior mean of $theta$? $hat\theta_{Bayes} = E[\theta|y] = \frac{438}{438+544} = 0.446$
  - What is the 95% posterior interval?
    1. Integration $\int_a^b p(\theta|y)d\theta = 0.95$ Find a, b.
    2. Normal approximation
    3. Numerical Method (quantile based C.I)
    4. HPD(Highest Posterior Density) Interval
- Use Different prior distributions: $\theta \sim Beta(\alpha, \beta) \to \theta|y \sim Beta(437+\alpha, 543+\beta)$

Specification of a Prior Distribution

```r
par(mfrow=c(2,2))
x <- seq(0, 1, length.out=51)
alpha = 1
beta = 1
n = 5
y = 1
plot(x, dbeta(x, y + alpha, n - y + beta), col='red', type='l') + lines(x, dbeta(x, alpha, beta), col='l
```

```
## integer(0)
```

```r
alpha = 3
beta = 2
n = 5
y = 1
plot(x, dbeta(x, y + alpha, n - y + beta), col='red', type='l') + lines(x, dbeta(x, alpha, beta), col='l
```

```
## integer(0)
```

```r
alpha = 1
beta = 1
n = 100
y = 20
plot(x, dbeta(x, y + alpha, n - y + beta), col='red', type='l') + lines(x, dbeta(x, alpha, beta), col='l
```

```
## integer(0)
```

```r
alpha = 3
beta = 2
n = 100
y = 20
plot(x, dbeta(x, y + alpha, n - y + beta), col='red', type='l') + lines(x, dbeta(x, alpha, beta), col='l
```