Chapter 8

Principle in Bayesian Statistics

8.1 Basic Principles

Bayesian statistics is a subset of the field of statistics in which the evidence about the true state of the world is expressed in terms of degrees of belief, or more specifically, "Bayesian probabilities". The essential characteristic of Bayesian methods is their explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis.
Story: the testimony of three men creates a tiger in the market.
In ancient China (475-221BC), kingdoms fight against each other. In order to make the alliance to keep their promise, it is common to send the princes to the alliances as hostage. The Wei Kingdom send his prince to Zhao Kingdom as a hostage. The accompanied minister, Cong Pang, approached the king.

```
Cong: "If there is one person said that there is a wild tiger in the street outside,
will you, the Majesty,  believe it?",
King: "no.
Cong: "If there is another one said the same thing, will you believe that?"
King: "maybe".
Cong: "If there is the third one said the wild tiger is there. Will you believe that?"
King: "Of course".
Cong: "There is no tiger on the street. But it sounds like real if three men repeat it.
Now, the capital of Zhao is very far away from here, and there will be more than three
people slandering me. Hope the Majesty see through this".
```

Let A be the event that there is really a tiger on the street. In frequentist, $P(A)$ is a fixed, unknown parameters, and it never varies. They believe that

Figure 8.1: The testimony of three men creates a tiger.

the probability that there is a tiger or not does not depend on the later observation.

In Bayesian approach, $P(A)$ is viewed as the King's belief of the event that there is a wild tiger on the street. Initially, the King's believe hat $P(A)$ is very small based on the historical data (*prior*). However, his believe keeps on updating when more and more data are collected. Eventually, he believes that there indeed a tiger on the street.

Heated philosophical debates occur between frequentist statisticians and Bayesian statisticians. However, we prefer to concentrate on the pragmatic advantages of the Bayesian framework.

The process of bayesian data analysis can be idealized by dividing it into the following three steps:

1. Setting up a full probability model– a joint probability distribution for all observable and unobservable quantities in a problem. The model should be consistent with knowledge about the underlying scientific problem and the data collection process;

2. Giving a prior distribution on the parameters;

3. Conditioning on observed data: calculating and interpreting the appropriate *posterior distribution*–the conditional probability distribu-

tion of the unobserved quantities of ultimate interest,given the observed data;

4. Evaluating the fit of the model and the implications of the resulting posterior distribution. If necessary, one can alter or expand the model and repeat the three steps.

8.2 Posterior distribution

We use $\boldsymbol{y}$ to denote the observations, $\boldsymbol{\theta}$ to denote the parameter. In Bayesian inference, a probability model include two parts:

1. Likelihood: $\boldsymbol{Y}|\boldsymbol{\theta} \sim f(\boldsymbol{y}|\boldsymbol{\theta})$;

2. Prior distribution: $\boldsymbol{\theta} \sim f(\boldsymbol{\theta})$.

Then the joint density of $(\boldsymbol{Y}, \boldsymbol{\theta})$ is given as

$$f(\boldsymbol{y}, \boldsymbol{\theta}) = f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}).$$

Theorem 8.2.1. Bayes Theorem
The posterior density of $\boldsymbol{\theta}$ given the observation $\boldsymbol{y}$ is

$$f(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta})}{f(\boldsymbol{y})} = \frac{f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta})}{\int f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}}.$$

Consider a simple case when we only have one parameter θ and one observation Y.

Example 8.2.1. Binomial model
Assume that

$$\begin{cases} X|p \sim Bin(n, p) \\ p \sim Beta(\alpha, \beta) \end{cases}$$

Here $\theta \sim Beta(\alpha, \beta)$ if $f(\theta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$. See Figure 8.2 for the shape of the density for various parameters. For the Beta distribution, it is known that

$$E\theta = \frac{\alpha}{\alpha + \beta}, V(\theta) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Assume the above model, then $p|X \sim Beta(\alpha + X, \beta + n - X)$.

Conjugate prior distributions

Conjugacy is formally defined as follows. If $\mathcal{F}$ is a class of sampling distributions $p(y|\theta)$, and $\mathcal{P}$ is a class of prior distributions for θ, then the class $\mathcal{P}$ is conjugate for $\mathcal{F}$ if

$$f(\theta|y) \in \mathcal{P}, \forall f(\cdot|\theta) \in \mathcal{F}, f(\cdot) \in \mathcal{P}.$$

Example 8.2.2. Normal-normal model
Assume that $Y|\theta \sim N(\theta, \sigma^2)$ and $\theta \sim N(\mu, \tau^2)$. Then

$$\theta|Y \sim N(\frac{\tau^2}{\sigma^2 + \tau^2}Y + \frac{\sigma^2}{\sigma^2 + \tau^2}\mu, \frac{\tau^2}{\sigma^2 + \tau^2}).$$

Let $M = \frac{\tau^2}{\sigma^2+\tau^2}$, then

$$\theta|Y \sim N(MY + (1 - M)\mu, M\sigma^2).$$

8.3 Bayesian inference

8.3.1 Estimation

How to estimate θ?

In general, we estimate θ as the mean or median of $\theta|Y$. In the normal-normal model, the estimator of θ is

$$\hat{\theta} = E\theta|Y = MY + (1-M)\mu.$$

What about the Binomial example?

8.3.2 Credible interval

In Bayesian statistics, the interval is called "credible interval". There are two widely used credible intervals (i) equal-tail interval; (ii) highest posterior density (HPD) interval. Let $f(\theta|y)$ be the posterior distribution of the parameter given the data. Let $F(\theta|y)$ be the cumulative distribution function. Then $1-\alpha$ equal-tail interval is defined as

$$[F^{-1}(1-\alpha/2), F^{-1}(\alpha/2).$$

The $1-\alpha$ HPD interval is defined as $[a, b]$ where a and b satisfy the following conditions:

i. $f(a) = f(b)$;

ii. $F(b) - F(a) = 1 - \alpha$.

For the normal-normal model, the $1 - \alpha$ interval is

$$\frac{\tau^2}{\sigma^2 + \tau^2} Y + \frac{\sigma^2}{\sigma^2 + \tau^2} \mu \pm z_{1-\alpha/2} \sqrt{\frac{\tau^2}{\sigma^2 + \tau^2}}.$$

Interpretation of credible intervals and confidence intervals.

8.3.3 Hypothesis testing

Consider the hypothesis

$$H_0 : \theta \in \Theta_0, H_1 : \theta \in \Theta_0^c.$$

When assuming θ be a continuous random variable, it is very rare to test $H_0 : \theta = c$ because θ is almost surely never equal to c. It is more reasonable to test whether θ falls in to a range or not. One way to do the Bayesian hypothesis testing:

- If $P(\theta \in \Theta_0|y) \geq P(\theta \in \Theta_0^c|y)$ or, $P(\theta \in \Theta_0|y) > 0.5$, fail to reject H_0;
- If $P(\theta \in \Theta_0|y) < P(\theta \in \Theta_0^c|y)$, reject H_0.

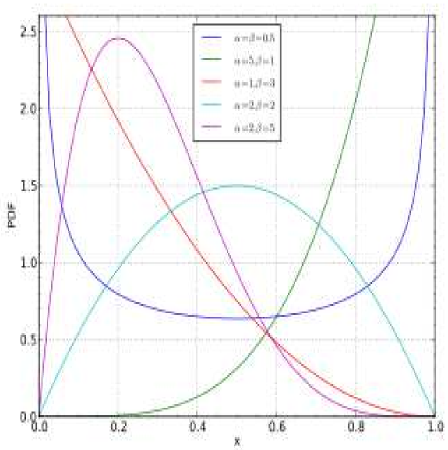

Figure 8.2: Beta distribution.

8.3.4 Example

Example 8.3.1. Female birth rate
We consider the maternal condition placenta previa, an unusual condition of pregnancy in which the placenta is implanted very low in the uterus, obstructing the fetus from a normal vaginal delivery. This factor may influence the sex ratio. An early study concerning the sex of placenta previa births in Germany found that of at a total of 980 births, 437 were female. How much evidence does this provide for the claim that the proportion of female births in the population of placenta previa births is less than 0.485, the proportion of female births in the general population?

Let Y be the number of female, θ be the proportion of female births in the population of placenta previa births.
Model:

$$\begin{cases} Y|\theta \sim Bin(980, \theta) \\ \theta \sim Beta(\alpha, \beta). \end{cases}$$

Beta distribution is conjugate to the Binomial distribution. It can be shown that

$$\theta|y \sim Beta(y + \alpha + 1, n - y + \beta + 1).$$

When $n = 980$, $y = 437$, then

$$\theta|y \sim Beta(437 + \alpha, 543 + \beta),$$

Parameters of the prior		Summaries of the Posterior	
$\frac{\alpha}{\alpha+\beta}$	$\alpha+\beta$	Posterior median	95% posterior interval
0.500	2	0.446	[0.415, 0.477]
0.485	2	0.446	[0.415, 0.477]
0.485	5	0.446	[0.415, 0.477]
0.485	200	0.453	[0.424, 0.481]

Table 8.1: Estimator of female birth rate.

and $E\theta|y = \frac{437+\alpha}{980+\alpha+\beta}$. See Table

Two problems:

1. How to choose the prior distribution?
2. Naturally, we are interested in the odds $(\frac{\theta}{1-\theta})$ or logit $(\log\frac{\theta}{1-\theta})$. How to estimate them and the corresponding intervals.

How to choose the prior?

1. An *informative prior* expresses specific, definite information about a variable. An example is a prior distribution for the temperature at noon tomorrow. A reasonable approach is to make the prior with expected value equal to today's noontime temperature.
2. A *noninformative prior* expresses vague or general information about a variable. Such a prior might be called a *not very informative prior*, or an *objective prior*. In the previous example, one can assume that $\alpha = \beta = 1$, i.e. $\theta \sim U(0, 1)$.
3. Consider hierarchical prior, where $\alpha, \beta \sim Unif(0, 100)$, for instance.

When interested in other quantities as functions of θ, it is difficult to derive the explicit formula for the median and quantiles even if we know the posterior distribution. In such cases, it can be particularly useful to use simulation from the posterior distribution to obtain inferences.

We now consider the non-informative prior, then $\theta|y \sim Beta(438, 544)$. We draw $n = 10,000$ samples from this distribution. In Figure 8.3, we plot the histogram of θ, $\log\frac{\theta}{1-\theta}$ and $\frac{\theta}{1-\theta}$.

Inference

Inference regarding θ,

1. $\hat{\theta} = 0.446$;

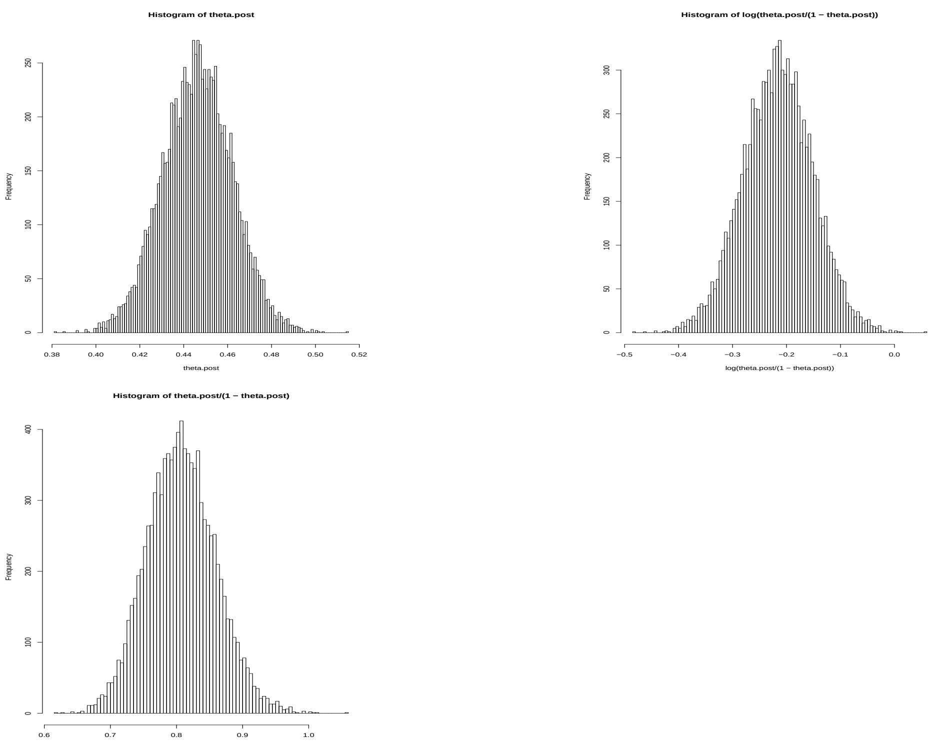

Figure 8.3: Draws from the posterior distribution of θ, $logit(\theta)$, and $\frac{\theta}{1-\theta}$.

2. The 95% credible interval is [0.415, 0.477];

3. $P(\theta < 0.485|y) = 99.2\%$.

Inference regarding $\log \frac{\theta}{1-\theta}$:

1. $\widehat{\log \frac{\theta}{1-\theta}} = -0.216$;

2. The 95% credible interval is [-0.343, -0.092].

Inference regarding $\frac{\theta}{1-\theta}$:

1. $\widehat{\frac{\theta}{1-\theta}} = 0.806$;

2. The 95% credible interval is [0.710, 0.912]. The European-race sex ratio ($\frac{female}{male}$) is 0.943. Consequently, the probability of a female birth given placenta previa is less than in the general population.

Interpretation: does this mean the *placenta previa* causes low rate of female birth?