

Bayesian Analysis (Optional)

Big Picture

There are two ways to conduct statistical inference

1. Classical method (frequentist), which postulates
 - (a) Probability refers to limiting relative frequencies
 - (b) Parameters are fixed, unknown constants
2. Bayesian method, which postulates
 - (a) Probability describes degree of belief, not limiting frequency
 - (b) Parameters follow distributions

For example, consider flipping a coin. The unknown parameter is the probability of showing head. The classical method refers to the limiting frequency of observing head when the coin is tossed many many times, while the Bayesian method amounts to whether you think the coin is fair, and then tossing coin many times to update that belief.

Bayes' Theorem (BT)

By the definition of conditional probability,

$$P(A|B) \equiv \frac{P(A \cap B)}{P(B)} \equiv \frac{P(A)P(B|A)}{P(B)}$$

So the Bayes' theorem (BT) is

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{1}$$

Basically BT shows how to move from $P(A)$ to $P(A|B)$. Notice that $P(A|B) \gg P(A)$ if $P(B)$ is close to zero. For example, A is death, and B is having an airplane crash.

Bayes' Method

1. Let $A = \theta$, the unknown parameter
2. $P(A) = P(\theta)$ is the prior distribution, or our belief about the parameter before seeing the data
3. Let B be the data (y_1, y_2, \dots, y_T)
4. $P(B|A) = P((y_1, y_2, \dots, y_T)|\theta)$ is the likelihood, or joint density of (y_1, y_2, \dots, y_T) conditional on θ
5. $P(A|B) = P(\theta|(y_1, y_2, \dots, y_T))$ is the posterior distribution, the updated belief about the parameter after seeing the data

Marginal Distribution $P(B)$

Notice that the marginal distribution of the data is free of θ :

$$P(B) = \sum_A P(A \cap B) \quad (2)$$

$$\Rightarrow P(y_1, y_2, \dots, y_T) = \int f(y_1, y_2, \dots, y_T, \theta) d\theta \quad (3)$$

where $f(y_1, y_2, \dots, y_T, \theta)$ is the joint density of data and parameter. Since the parameter θ has been integrated out, we can ignore the marginal distribution $P(y_1, y_2, \dots, y_T)$ for purpose of estimating θ .

Bayes' Method II

Now the Bayes' theorem can be rewritten as

$$P(\boldsymbol{\theta}|(y_1, y_2, \dots, y_T)) \propto P(\boldsymbol{\theta})P(y_1, y_2, \dots, y_T|\boldsymbol{\theta}) \quad (4)$$

or in terms of density function,

$$f(\boldsymbol{\theta}|(y_1, y_2, \dots, y_T)) \propto f(\boldsymbol{\theta})f(y_1, y_2, \dots, y_T|\boldsymbol{\theta}) \quad (5)$$

In words, the posterior distribution $f(\boldsymbol{\theta}|(y_1, y_2, \dots, y_T))$ is proportional to the prior distribution $f(\boldsymbol{\theta})$ times the likelihood $f(y_1, y_2, \dots, y_T|\boldsymbol{\theta})$. To sum up, Bayes method uses the likelihood to update the distribution of $\boldsymbol{\theta}$.

Dominance of Likelihood and Central Limit Theorem

In large sample

1. the likelihood dominates the prior distribution, or the effect of prior distribution is asymptotically negligible.
2. the central limit theorem implies that the likelihood follows normal distribution, so does the posterior distribution.

Bayes' Method III

Bayesian method amounts to extracting point estimates and interval estimates from the posterior distribution. For example, the point estimate can be the mean of the posterior distribution

$$\hat{\theta} = \int \theta f(\theta | (y_1, y_2, \dots, y_T)) d\theta \quad (6)$$

We can do the integration numerically. That is, we can average the many many random draws from the posterior distribution:

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N \theta_i, \quad \theta_i \sim \text{Posterior Distribution} \quad (7)$$

According to the law of large number, this sample average is the point estimate. Alternatively, we can obtain the point estimate by maximizing the posterior distribution (obtaining the mode of the posterior distribution).

Bayesian AR(1) regression for GDP

The classical method can be cumbersome. We need to run the unit root test, which follows nonstandard distribution, and choose the appropriate model based on the testing result. By contrast, the Bayesian method always implies that the posterior distribution of the autoregressive coefficient is normal distribution if sample is large.

Standardized GDP

For simple exposition, we first standardize the log GDP

$$y = \frac{\log(\text{GDP}) - \mu}{\sigma}$$

where μ and σ are the mean and standard deviation of log GDP. Now the only unknown parameter is the autoregressive coefficient θ in the AR(1) regression

$$y_t = \theta y_{t-1} + e_t.$$

The classical method is interested in the hypothesis $H_0 : \theta = 1$, and the t test follows the Dickey-Fuller distribution.

Likelihood

Without losing generality, we can assume

$$y_t | y_{t-1} \sim N(\theta y_{t-1}, 1) \quad (8)$$

$$f(y_t | y_{t-1}, \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_t - \theta y_{t-1})^2}{2}} \quad (9)$$

$$f((y_1, y_2, \dots, y_t) | \theta) = f(y_1) \prod_{t=2}^T \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_t - \theta y_{t-1})^2}{2}} \quad (10)$$

where $f(y_1)$ can be ignored in large sample. Equation (10) is the likelihood.

Prior

Given the strong persistence of log GDP, we assume the prior distribution of θ is a uniform distribution between 0.8 and 1.2:

$$f(\theta) = \frac{1}{1.2 - 0.8}, \quad (0.8 \leq \theta \leq 1.2) \quad (11)$$

Notice that this prior allows that the series is stationary ($\theta < 1$), having unit root ($\theta = 1$), or exploding ($\theta > 1$).

Posterior

According to BT, the posterior distribution of θ is

$$f(\boldsymbol{\theta} | (y_1, y_2, \dots, y_T)) \propto f(\boldsymbol{\theta}) f(y_1, y_2, \dots, y_T | \boldsymbol{\theta}) \quad (12)$$

$$= \frac{1}{1.2 - 0.8} f(y_1) \prod_{t=2}^T \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_t - \theta y_{t-1})^2}{2}} \quad (13)$$

$$\propto \prod_{t=2}^T \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_t - \theta y_{t-1})^2}{2}} \quad (14)$$

The final result clearly shows the likelihood is the dominant factor for the posterior distribution.

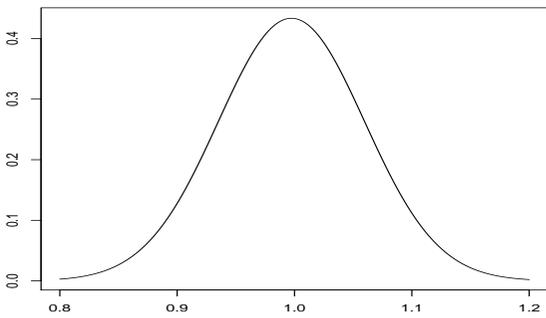
R code of obtaining likelihood

```
lik.ar1 = function(co, y) {  
  a1 = co[1]  
  n = length(y)  
  f = exp(-(y[1])**2/2)  
  for (i in 2:n) {  
    f = f*exp(-(y[i]-a1*y[i-1])**2/2)  
  }  
  return(f)  
}
```

R code of plotting posterior

```
theta = seq(0.8, 1.2, by = 0.002)
k = length(theta)
lik = rep(0, k)
for (i in 1:k) lik[i] = lik.ar1(theta[i], y) * (1 / (1.2 - 0.8))
plot(theta, lik, type = "l")
```

Plotting posterior



We see that the peak of the posterior distribution corresponds to value of θ around 1.0. There is no test to conduct. We do not need to worry about Dickey-Fuller distribution.