

Bayesian Normal Stuff

- Set-up of the basic model of a normally distributed random variable with unknown mean and variance (a two-parameter model).
- Discuss philosophies of prior selection
- Implementation of different priors with a discussion of MCMC methods.

Introduction to Applied Bayesian Modeling

ICPSR 2008

Day 8

The normal model with unknown mean and variance

Let's extend the normal model to the case where the variance parameter is assumed to be unknown. Thus, $y_i \sim N(\mu, \sigma^2)$, where μ and σ^2 are both unknown random variables.

The Bayesian set-up should still look familiar:

$$p(\mu, \sigma^2 | y) \propto p(\mu, \sigma^2) p(y | \mu, \sigma^2).$$

Note: we would like to make inferences about the marginal distributions $p(\mu | y)$ and $p(\sigma^2 | y)$ rather than the conditional distribution $p(\mu, \sigma^2 | y)$. Ultimately, we'd like to find:

$$p(\mu | y) = \int p(\mu | \sigma^2, y) p(\sigma^2 | y) d\sigma^2$$

What should we choose for the prior distribution $p(\mu, \sigma^2)$?

Different “types” of Bayesians choose different priors

Classical Bayesians: the prior is a necessary evil.

→ choose priors that interject the least information possible.

Modern Parametric Bayesians: the prior is a useful convenience.

→ choose prior distributions with desirable properties (e.g. conjugacy). Given a distributional choice, prior parameters are chosen to interject the least information possible.

Subjective Bayesians: the prior is a summary of old beliefs

→ choose prior distributions based on previous knowledge—either the results of earlier studies or non-scientific opinion.

The Classical Bayesian and the normal model with unknown mean and variance

$y \sim N(\mu, \sigma^2)$ where μ and σ^2 are both unknown random variables.

What prior distribution would we choose to represent the absence of any knowledge in this instance?

What if we assumed that the two parameters were independent, so $p(\mu, \sigma^2) = p(\mu)p(\sigma^2)$?

Modern Parametric Bayesians and the normal model with unknown mean and variance

$y \sim N(\mu, \sigma^2)$ where μ and σ^2 are both unknown random variables.

What prior distribution would a modern parametric Bayesian choose to satisfy the demands of convenience?

What if we used the definition of conditional probability, so $p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$?

Modern Parametric Bayesians and the normal model with unknown mean and variance

$y \sim N(\mu, \sigma^2)$ where μ and σ^2 are both unknown random variables.

A modern parameteric Bayesian would typically choose a conjugate prior.

For the normal model with unknown mean and variance, the conjugate prior for the joint distribution of μ and σ^2 is the *normal inverse-gamma* (Γ) distribution (i.e. normal-inverse- χ^2)

$$p(\mu, \sigma^2) \sim \text{N-Inv-}\chi^2(\mu_0, \sigma_0^2/k_0; \nu_0, \sigma_0^2)$$



Four Parameters in the prior

Suppose $p(\mu, \sigma^2) \sim \text{N-Inv-}\chi^2(\mu_0, \sigma_0^2/k_0; v_0, \sigma_0^2)$

ICBST the above expression can be factored such that:

$$p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$$

where $\mu|\sigma^2 \sim \text{N}(\mu_0, \sigma^2/k_0)$ and $\sigma^2 \sim \text{Inv-}\chi^2(v_0, \sigma_0^2)$

Because this is a conjugate distribution for the normal distribution with unknown mean and variance, the posterior distribution will also be normal-Inv- χ^2 .

Lazy Modern **Parametric Bayesians** and the **normal model with unknown mean and variance**

Suppose that $y \sim N(\mu, \tau)$ where τ was the prior precision.

From here on when we talk about the normal distribution you should expect that we will speak in terms of the precision τ rather than the variance σ^2 . This is because WinBugs is programmed to use τ rather than σ^2

Suppose also that you don't want to think too hard about the prior joint distribution of μ and τ , and assume that:

$$p(\mu, \tau) = p(\mu)p(\tau)$$

What distributions would you choose for $p(\mu)$ and $p(\tau)$?

Suppose that $y \sim N(\mu, \tau)$

What priors would you choose for μ and τ ?

I would choose:

$\mu \sim N(0, t)$ (where t was a large number)

This is because, if we expect something like the central limit theorem to hold, then the distribution of the sample mean should be approximately normal for large n .

Gamma



$\tau \sim \Gamma(a, b)$ (where a, b are small numbers)

This is because this distribution is bounded below at zero and unlike the χ^2 distribution which shares this property it is not constrained to have an equal mean and variance.

→ Note how we now have to talk about the mean of the distribution of the variance.

```
model {  
  for (i in 1:N)  
    { y[i] ~ dnorm( mu, tau) }  
  mu ~ dnorm(0, .001)  
  tau ~ dgamma(.01 , .001) }  
}
```

The Subjective Bayesian and the normal model with unknown mean and variance

The subjective Bayesian framework provides little guidance about what prior distribution that one should choose.

In a sense, that is the point of the subjective approach—it is subjective.

You are free to pick whatever prior distribution you want: multi-modal, uniform, high or low variance, skewed, constrained to lie between a certain set of values, etc.

One of the key difficulties is that the prior distributions probably are not independent (i.e. $p(\theta_1, \theta_2) \neq p(\theta_1)p(\theta_2)$).

For example, regression coefficients are generally not independent, even if that isn't transparent in your STATA output. If you want to incorporate subjective beliefs, this non-independence should be taken into account.

Some General Guidelines

I recommend the following general guidelines:

- 1) if possible, use standard distributions (i.e. conjugate or semi-conjugate) and choose parameters that fix the mean, variance, kurtosis, etc. to be some desirable level.

- 2) sample from the prior predictive distribution and check to see if your results make sense.
 - Mechanically, perform the following steps:
 - i) take a random draw θ' from the joint prior distribution of θ .
 - ii) take a random draw Y from the pdf $Y|\theta$ with $\theta = \theta'$
 - iii) repeat steps i and ii several thousand times to provide a sample that you can use to summarize the prior predictive distribution.
 - iv) generate various summaries of the prior predictive distribution and check to see if the model's predictions are consistent with your beliefs about the data-generating process.