Bayesian Inference for Normal - unknown mean and variance

Jeff Mills

October, 2018

Abstract

Posterior analysis for Normally distributed variable with unknown mean and variance using the Normal Inverted-Gamma (NIG) prior.

1 TL;DR What to do

If x is assumed to be normally distributed with unknown mean, μ and unknown variance, τ^{-1}, adopting a Normal-Gamma prior,

$$N(\mu_0, \lambda_0^{-1})\text{Gamma}(\alpha_0, \beta_0)$$

The posterior for μ is the t-distribution,

$$p(\mu|x) = \text{t}(\mu_1, \beta_1/(\lambda_1\alpha_1), 2\alpha_1),$$

where

$$\mu_1 = \frac{\lambda_0\mu_0 + n\bar{x}}{\lambda_0 + n}$$

$$\lambda_1 = \lambda_0 + n$$

$$\alpha_1 = \alpha_0 + n/2$$

$$\beta_1 = \beta_0 + 1/2\left(ns^2 + \frac{\lambda_0 n(\bar{x} - \mu_0)^2}{\lambda_0 + n}\right),$$

with $\bar{x}$ = the sample mean, and s^2 = the sample variance.

Computing the above six quantities: $\mu_1, \lambda_1, \alpha_1, \beta_1, \bar{x}$ and s^2 allows exact inference for μ. One can either proceed

(1) analytically using knowledge of the noncentral Student-t distribution, or

(2) numerically, drawing a large set of values, M, from the posterior t-distribution and then computing pseudo-sample values.

Approach (2) is generally easier in practice. Example R code is given in **NIGprior_t_posterior.R**.

2 Inference with Gaussian distribution - NIG prior

- see Jaynes (2003), ch. 7, for justification of using the Gaussian in most cases (unless we have very good reason not to).

Suppose we have data $y = [y_1, y_2, ..., y_n]$ consisting of n observations on the variable y, and wish to draw inferences about the mean value of y. Since we have no compelling reason to think that a Gaussian sampling distribution is inappropriate, we are compelled by the maximum entropy principle to adopt a Gaussian, so that the model for observation i is $y_i \sim N(\mu, \sigma^2)$.

2.1 Uninformative prior

With no prior information about the possible mean value of y (which is really an unusual case - we usually have at least some vague idea - but for the sake of "scientific objectivity" we will ignore relevant information!), we adopt an uninformative prior distribution, $p(\mu, \sigma^2) \propto \sigma^{-2}$.

This leads to the joint posterior [see Gelman *et al.* (2014), p.64, equation (3.2)],

$$p(\mu, \sigma^2|y) \propto \sigma^{n-2} \exp(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2]),$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y})^2$, $\bar{y} = \sum_{i=1}^n y_i/n$ are sufficient statistics in this case.

N.B. check exponent of σ in next formula:

2.2 Normal Inverted-Gamma or Normal Gamma prior

Instead, we could adopt an informative conjugate prior,

$$p(\mu, \sigma^2) \propto \sigma^{-v_0-1} \exp(\frac{1}{2\sigma^2}[v_0\sigma_0^2 + n_0(\mu_0 - \mu)^2]).$$

[Gelman *et al.* (2014), equation (3.6)]

This would lead to a joint posterior of exactly the same ('Normal-Inverted Gamma') functional form as the above posterior. For both mathematical convenience and computational stability, we rewrite this as the Normal-Gamma prior for the mean and precision, where the precision is the inverse of the variance ($\tau = 1/\sigma^2$).

For

$$x \sim N(\mu, \tau^{-1}),$$

and Normal-Gamma prior

$$p(\mu, \tau|\mu_0, \lambda_0, \alpha_0, \beta_0) = N(\mu_0, \lambda_0^{-1})\text{Gamma}(\alpha_0, \beta_0)$$

for with the density is given by,

$$p(\mu, \tau) \propto \tau^{\alpha_0 - 1/2} \exp[-\beta_0 \tau] \exp\left(-\frac{\lambda_0 \tau (\mu - \mu_0)^2}{2}\right).$$

The posterior can be written in the same form

$$p(\mu, \tau | \mu_1, \lambda_1, \alpha_1, \beta_1) = N(\mu_1, \lambda_1^{-1})\text{Gamma}(\alpha_1, \beta_1)$$

where

$$\mu_1 = \frac{\lambda_0 \mu_0 + n\bar{x}}{\lambda_0 + n}$$
$$\lambda_1 = \lambda_0 + n$$
$$\alpha_1 = \alpha_0 + n/2$$
$$\beta_1 = \beta_0 + 1/2\left(ns^2 + \frac{\lambda_0 n(\bar{x} - \mu_0)^2}{\lambda_0 + n}\right),$$

with $\bar{x}$ = the sample mean, and s^2 = the sample variance.

2.3 Marginal Posterior Densities

The marginal posterior for μ is obtained by integrating the joint posterior with respect to τ. This results in a Student-t posterior,

$$p(\mu|x) = \text{t}(\mu_1, \beta_1/(\lambda_1\alpha_1), 2\alpha_1).$$

The marginal posterior for τ is,

$$p(\tau|x) = \text{Gamma}(\alpha_1, \beta_1)$$

.

2.4 Bayesian Updating

Further, suppose we now obtain a second sample. Let's relabel the first sample y_1 (not to be confused with the first observation of the sample) with n_1 observations, and the second sample y_2 with n_2 observations. Using past data, i.e. the posterior from the first data set, (3.2) above as the new prior, comparing Gelman et al. (2014), equations (3.2) and (3.6) with parameters

$$v_0 = n_1 - 1,\ n_0 = n_1,\ \mu_0 = \bar{y_1},\ \sigma_0^2 = s_1^2.$$

This informative prior can also be written as [comparing notation more closely with Greenberg (2013)],

The joint posterior with the above informative prior is then of the same form as the prior,

$$p(\mu, \sigma^2) \propto \sigma^{-v_n - 1} \exp(\frac{1}{2\sigma^2}[v_n\sigma_n^2 + N(\mu_n - \mu)^2]),$$

where

$$N = n_1 + n_2,$$
$$\mu_n = \frac{n_1}{N}\bar{y}_1 + \frac{n_2}{N}\bar{y}_2,$$
$$v_n = N - 1,$$
$$v_n\sigma_n^2 = v_0\sigma_0^2 + (n-1)s^2 + (\frac{n_0 n}{n_0 + n})(\bar{y} - \mu_0)^2$$
$$= (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (\frac{n_1 n_2}{N})(\bar{y}_2 - \bar{y}_1)^2.$$

[see Gelman *et al.* (2014), p.68]

Julia code to implement the above in:
OneDrive, Julia Code, Compare-means-Strawn, pcbo-mean-compare.jl

3 Comparison of means

See Gregory (2005) chapter 9. Also data on river sediment toxin measurments, p. 182, Table 7.5, and World Cup goal statistics, p. 177, Table 7.4

Comparing two independent samples, the trial and control samples. The Bayesian approach developed by Bretthorst (1993) extends earlier work by Dayal (1972) and Dayal and Dickey (1976). [see Gregory (2005). p.228+]