Bayesian computations for the mean of a normal distribution

Matthew Stephens

University of Chicago

January 9, 2026

See here for a PDF version of this vignette.

Introduction

We consider computing the posterior distribution of μ given data $X \sim N(\mu, \sigma^2)$ where σ^2 is known. You should be familiar with the idea of a conjugate prior.

Preliminaries

This problem is really about algebraic manipulation.

There are two tricks to making the algebra a bit simpler. The first is to work with the precision $\tau = 1/\sigma^2$ instead of the variance σ^2. So consider $X \sim N(\mu, 1/\tau)$.

The second trick is to rewrite the normal density slightly. First, let us recall the usual form for the normal density. If $Y \sim N(\mu, 1/\tau)$ then it has density:

$$p(y) = (\tau/2\pi)^{0.5} \exp(-0.5\tau(y - \mu)^2)$$

We can rewrite this:

$$p(y) \propto \exp(-0.5\tau y^2 + \tau\mu y).$$

Or, equivalently:

$$p(y) \propto \exp(-0.5Ay^2 + By)$$

where $A = \tau$ and $B = \tau\mu$.

Thus if $p(y) \propto \exp(-0.5Ay^2 + By)$ then Y is normal with precision $\tau = A$ and mean $\mu = B/A$.

Posterior Calculation

Now, back to the problem. Assume we observe a single data point $X \sim N(\mu, 1/\tau)$, with τ known, and our goal is to do Bayesian inference for the mean μ.

As we will see, the conjugate prior for the mean μ turns out to be a normal distribution. So we will assume a prior:

$$\mu \sim N(\mu_0, 1/\tau_0).$$

(Here the 0 subscript is being used to indicate that μ_0, τ_0 are parameters in the prior.)

Now we can compute the posterior density for μ using Bayes Theorem:

$$p(\mu|X) \propto p(X|\mu)p(\mu)$$

$$\propto \exp[-0.5\tau(X - \mu)^2]\exp[-0.5\tau_0(\mu - \mu_0)^2]$$
$$\propto \exp[-0.5(\tau + \tau_0)\mu^2 + (X\tau + \mu_0\tau_0)\mu]$$

From the result in "Preliminaries" above we see that

$$\mu|X \sim N(\mu_1, 1/\tau_1)$$

where

$$\tau_1 = \tau + \tau_0$$

and

$$\mu_1 = (X\tau + \mu_0\tau_0)/(\tau + \tau_0).$$

Interpretation

Although the algebra may look a little messy the first time you see this, in fact this result has some simple and elegant interpretations.

First, let us deal with the precision. Note that the Posterior precision (τ_1) is the sum of the Data precision (τ) and the Prior precision (τ_0). This makes sense: the more precise your data, and the more precise your prior information, the more precise your posterior information. Also, this means that the data always improves your posterior precision compared with the prior: noisy data (small τ) improves it only a little, whereas precise data improves it a lot.

Second, let us deal with the mean. We can rewrite the posterior mean as:

$$\mu_1 = wX + (1 - w)\mu_0,$$

where $w = \tau/(\tau + \tau_0)$. Thus μ_1 is a *weighted average* of the data X and the prior mean μ_0. And the weights depend on the relative precision of the data and the prior. If the data are precise compared with the prior ($\tau >> \tau_0$) then the weight w will be close to 1 and the posterior mean will be close to the data.

In contrast, if the data are imprecise compared with the prior ($\tau << \tau_0$) then the weight w will be close to 0 and the posterior mean will be close to the prior mean.

You can see a visual illustration of this result in this shiny app.