1 Conjugate Posterior for the Normal Mean

Here we derive the update formula from Lesson 10.1 where the likelihood is normal and we use a conjugate normal prior on the mean. Specifically, the model is $x_1, x_2, \ldots, x_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma_0^2)$, $\mu \sim \mathrm{N}(m_0, s_0^2)$ with σ_0^2, m_0, and s_0^2 known. First consider the case in which we have only one data point x. The posterior is then

$$
\begin{aligned}
f(\mu|x) &= \frac{f(x|\mu)f(\mu)}{\int_{-\infty}^{\infty} f(x|\mu)f(\mu)d\mu} \\
&\propto f(x|\mu)f(\mu) \\
&= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2\sigma_0^2}(x-\mu)^2\right\} \frac{1}{\sqrt{2\pi s_0^2}} \exp\left\{-\frac{1}{2s_0^2}(\mu-m_0)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_0^2}(x-\mu)^2\right\} \exp\left\{-\frac{1}{2s_0^2}(\mu-m_0)^2\right\} \\
&= \exp\left\{-\frac{1}{2\sigma_0^2}(x-\mu)^2 - \frac{1}{2s_0^2}(\mu-m_0)^2\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_0^2}(x^2-2x\mu+\mu^2) + \frac{1}{s_0^2}(\mu^2-2m_0\mu+m_0^2)\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\frac{\mu^2}{\sigma_0^2} + \frac{\mu^2}{s_0^2} + \frac{-2x}{\sigma_0^2}\mu + \frac{-2m_0}{s_0^2}\mu + \frac{x^2}{\sigma_0^2} + \frac{m_0^2}{s_0^2}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{s_0^2}+\frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{x}{\sigma_0^2}+\frac{m_0}{s_0^2}\right)\mu + \frac{x^2}{\sigma_0^2} + \frac{m_0^2}{s_0^2}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{s_0^2}+\frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{x}{\sigma_0^2}+\frac{m_0}{s_0^2}\right)\mu\right]\right\} \cdot \exp\left\{-\frac{1}{2}\left[\frac{x^2}{\sigma_0^2} + \frac{m_0^2}{s_0^2}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{s_0^2}+\frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{x}{\sigma_0^2}+\frac{m_0}{s_0^2}\right)\mu\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\frac{1}{s_1^2}\mu^2 - 2\left(\frac{x}{\sigma_0^2}+\frac{m_0}{s_0^2}\right)\mu\right]\right\} \qquad \text{where } s_1^2 = \left(\frac{1}{s_0^2}+\frac{1}{\sigma_0^2}\right)^{-1} \\
&= \exp\left\{-\frac{1}{2}\left[\frac{1}{s_1^2}\mu^2 - 2\frac{s_1^2}{s_1^2}\left(\frac{x}{\sigma_0^2}+\frac{m_0}{s_0^2}\right)\mu\right]\right\} \\
&= \exp\left\{-\frac{1}{2s_1^2}\left[\mu^2 - 2m_1\mu\right]\right\} \qquad \text{where } m_1 = s_1^2\left(\frac{m_0}{s_0^2}+\frac{x}{\sigma_0^2}\right).
\end{aligned}
$$

The next step is to "complete the square" in the exponent:

$$
\begin{aligned}
f(\mu|x) &\propto \exp\left\{-\frac{1}{2s_1^2}\left[\mu^2 - 2m_1\mu\right]\right\} \\
&= \exp\left\{-\frac{1}{2s_1^2}\left[\mu^2 - 2m_1\mu + m_1^2 - m_1^2\right]\right\} \\
&= \exp\left\{-\frac{1}{2s_1^2}\left[\mu^2 - 2m_1\mu + m_1^2\right]\right\} \cdot \exp\left\{\frac{m_1^2}{2s_1^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2s_1^2}\left[\mu^2 - 2m_1\mu + m_1^2\right]\right\} \\
&= \exp\left\{-\frac{1}{2s_1^2}(\mu - m_1)^2\right\}
\end{aligned}
$$

which, except for a normalizing constant not involving μ, is the PDF of a normal distribution with mean m_1 and variance s_1^2.

The final step is to extend this result to accommodate n independent data points. The likelihood in this case is

$$
\begin{aligned}
f(\boldsymbol{x}|\mu) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2\sigma_0^2}(x_i - \mu)^2\right\} \\
&= (2\pi\sigma_0^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i + n\mu^2\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_0^2}\left[-2n\bar{x}\mu + n\mu^2\right]\right\} .
\end{aligned}
$$

We can repeat the steps above *or* notice that the data contribute only through the sample mean $\bar{x}$ (and n which we assume is known). This means that $\bar{x}$ is a "sufficient statistic" for μ, allowing us to use the distribution of $\bar{x}$ as the likelihood (analogous to using a binomial likelihood in place of a sequence of Bernoullis). The model then becomes

$$
\bar{x} \mid \mu \sim \mathrm{N}\left(\mu, \frac{\sigma_0^2}{n}\right) , \quad \mu \sim \mathrm{N}(m_0, s_0^2) .
$$

We now apply our result derived above, replacing x with $\bar{x}$ and σ_0^2 with σ_0^2/n. This yields the update equation presented in Lesson 10.1.

2 Marginal Distribution of Normal Mean in Conjugate Model

Consider again the model $x|\mu \sim \text{N}(\mu, \sigma_0^2)$, $\mu \sim \text{N}(m_0, s_0^2)$ with σ_0^2 known. Here we derive the marginal distribution for data given by $\int_{-\infty}^{\infty} f(x|\mu)f(\mu)d\mu$. This is the prior predictive distribution for a new data point x^*.

To do so, re-write the model in an equivalent, but more convenient form: $x = \mu + \epsilon$ where $\epsilon \sim \text{N}(0, \sigma_0^2)$ and $\mu = m_0 + \eta$ where $\eta \sim \text{N}(0, s_0^2)$, with ϵ and η independent. Now substitute μ into the first equation to get $x = m_0 + \eta + \epsilon$. Recall that adding two normal random variables results in another normal random variable, so x is normal with $E(x) = E(m_0 + \eta + \epsilon) = m_0 + E(\eta) + E(\epsilon) = m_0 + 0 + 0$ and $Var(x) = Var(m_0 + \eta + \epsilon) = Var(m_0) + Var(\eta) + Var(\epsilon) = 0 + s_0^2 + \sigma_0^2$ (note that we can add variances because of the independence of η and ϵ). Therefore the marginal distribution for x is normal with mean m_0 and variance $s_0^2 + \sigma_0^2$. The posterior predictive distribution is the same, but with m_0 and s_0^2 replaced by the posterior updates given in Lesson 10.1.

3 Inverse-Gamma Distribution

The inverse-gamma distribution is the conjugate prior for σ^2 in the normal likelihood with known mean. It is also the marginal prior/posterior for σ^2 in the model of Lesson 10.2.

As the name implies, the inverse-gamma distribution is related to the gamma distribution. If $X \sim \text{Gamma}(\alpha, \beta)$, then the random variable $Y = 1/X \sim \text{Inverse-Gamma}(\alpha, \beta)$ where

$$\begin{aligned} f(y) &= \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha+1)} \exp\left(-\frac{\beta}{y}\right) I_{\{y>0\}} \\ E(Y) &= \frac{\beta}{\alpha - 1} \quad \text{for } \alpha > 1\,. \end{aligned}$$

The relationship between gamma and inverse-gamma suggest a simple method for simulating draws from the inverse-gamma distribution. First draw X from the Gamma(α, β) distribution and take $Y = 1/X$, which corresponds to a draw from the Inverse-Gamma(α, β).

4 Marginal Posterior Distribution for the Normal Mean when the Variance is Unknown

If we are not interested in inference for an unknown σ^2, we can integrate it out of the joint posterior in Lesson 10.2. This results in a t-distributed marginal posterior as noted at the end of the lesson. This t distribution has $\nu = 2\alpha + n$ degrees of freedom and two additional parameters, a scale γ and a location m^* given by

$$m^* = \frac{n\bar{x} + wm}{n + w} \qquad \text{(the mean of the conditional posterior for } \mu\text{)}$$

$$\gamma = \sqrt{\frac{\beta + \frac{n-1}{2}s^2 + \frac{wn}{2(w+n)}(\bar{x} - m)^2}{(n + w)(\alpha + n/2)}} \qquad \text{(modified scale of the updated inverse-gamma for } \sigma^2\text{)}$$

where $s^2 = \sum(x_i - \bar{x})^2/(n-1)$, the sample variance.

This t distribution can be used to create a credible interval for μ by multiplying the appropriate quantiles of the standard t distribution by the scale γ and adding the location m^*.

Example: Suppose we have normal data with unknown mean μ and variance σ^2. We use the model from Lesson 10.2 with $m = 0$, $w = 0.1$, $\alpha = 3/2$, and $\beta = 1$. The data are $n = 20$ independent observations with $\bar{x} = 1.2$ and $s^2 = 0.7$. Then we have

$$\begin{aligned}
\sigma^2 \mid \boldsymbol{x} &\sim \text{Inverse-Gamma}(11.5, 7.72) \\
\mu \mid \sigma^2, \boldsymbol{x} &\sim \text{N}\left(1.19, \frac{\sigma^2}{20.1}\right) \\
m^* &= 1.19 \\
\gamma &= 0.183
\end{aligned}$$

and $\mu \mid \boldsymbol{x}$ is distributed t with 23 degrees of freedom, location 1.19 and scale 0.183. To produce a 95% equal-tailed credible interval for μ, we first need the 0.025 and 0.975 quantiles of the standard t distribution with 23 degrees of freedom. These are -2.07 and 2.07. The 95% credible interval is then $m^* \pm \gamma(2.07) = 1.19 \pm 0.183(2.07) = 1.19 \pm 0.38$.