"Normal likelihood, normal prior" conjugate pair

Denis Valle

March 2017

Say that we are interested in estimating μ (i.e., we assume that we know σ^2) and that our likelihood is given by:

$$p(x_1, \dots, x_n|\mu) \propto \prod_{i=1}^{n} N(x_i|\mu, \sigma^2)$$

Say we impose a prior for μ which is also normal:

$$p(\mu) = N(\mu|m_0, v_0^2)$$

To obtain the posterior distribution, we need to calculate:

$$p(\mu|x_1, \dots, x_n) \propto \left[\prod_{i=1}^{n} N(x_i|\mu, \sigma^2)\right] N(\mu|m_0, v_0^2)$$

We know that the posterior distribution is also a normal distribution (i.e., $N(\mu|m_1, v_1^2)$). As a result, after discarding everything that does not depend on μ, we expect something of the form:

$$= \left(\frac{1}{\sqrt{2\pi v_1^2}}\right) \exp\left(-\frac{1}{2v_1^2}(\mu - m_1)^2\right)$$

$$\propto \exp\left(-\frac{1}{2v_1^2}(\mu^2 - 2\mu m_1 + m_1^2)\right)$$

$$= \exp\left(-\frac{1}{2v_1^2}(\mu^2 - 2\mu m_1)\right) \exp\left(-\frac{1}{2v_1^2}(m_1^2)\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(\mu^2\left(\frac{1}{v_1^2}\right) - 2\mu\left(\frac{1}{v_1^2}\right)m_1\right)\right) \quad \text{[Eqn. 1]}$$

where the parameters v_1^2 and m_1 characterize the posterior distribution of μ. We will be manipulating our equations to get to something that should resemble Eqn. 1.

Here are the steps to derive this expression:

$$\propto \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)\right] \frac{1}{\sqrt{2\pi v_0^2}} \exp\left(-\frac{1}{2v_0^2}(\mu - m_0)^2\right)$$

After we drop everything that is a constant from the perspective of μ (i.e., $\frac{1}{\sqrt{2\pi\sigma^2}}$ and $\frac{1}{\sqrt{2\pi v_0^2}}$), we obtain:

$$\propto \left[\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)\right] \exp\left(-\frac{1}{2v_0^2}(\mu-m_0)^2\right)$$

Expanding the quadratic terms, we obtain:

$$\propto \left[\prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}\left(x_i^2-2x_i\mu+\mu^2\right)\right)\right] \exp\left(-\frac{1}{2v_0^2}(\mu^2-2\mu m_0+m_0^2)\right)$$

We can get rid of the big product sign by converting it into a big sum sign using the rule $e^{x_1}e^{x_2}e^{x_3} = e^{x_1+x_2+x_3}$:

$$\propto \exp\left(-\sum_{i=1}^{n}\frac{1}{2\sigma^2}\left(x_i^2-2x_i\mu+\mu^2\right)\right) \exp\left(-\frac{1}{2v_0^2}(\mu^2-2\mu m_0+m_0^2)\right)$$

Now sum the exponents and bring $\frac{1}{2\sigma^2}$ in front of the summation sign:

$$\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i^2-2x_i\mu+\mu^2\right)-\frac{1}{2v_0^2}(\mu^2-2\mu m_0+m_0^2)\right)$$

Bring the common factor $-\frac{1}{2}$ upfront:

$$\propto \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i^2-2x_i\mu+\mu^2\right)+\frac{1}{v_0^2}(\mu^2-2\mu m_0+m_0^2)\right]\right)$$

Distribute the summation sign:

$$\propto \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}\left(\sum_i x_i^2-2\mu\sum_i x_i+n\mu^2\right)+\frac{1}{v_0^2}(\mu^2-2\mu m_0+m_0^2)\right]\right)$$

Notice that there are additional pieces that don't depend on μ and that therefore can be dropped:

$$\propto \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}\left(-2\mu\sum_i x_i+n\mu^2\right)+\frac{1}{v_0^2}(\mu^2-2\mu m_0)\right]\right) \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}\left(\sum_i x_i^2\right)+\frac{1}{v_0^2}(m_0^2)\right]\right)$$

$$\propto \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma^2}\left(-2\mu\sum_i x_i+n\mu^2\right)+\frac{1}{v_0^2}(\mu^2-2\mu m_0)\right]\right)$$

Distribute $\frac{1}{\sigma^2}$ and $\frac{1}{v_0^2}$:

$$\propto \exp\left(-\frac{1}{2}\left[-2\mu\frac{1}{\sigma^2}\sum_i x_i+n\frac{1}{\sigma^2}\mu^2+\frac{1}{v_0^2}\mu^2-2\mu\frac{1}{v_0^2}m_0\right]\right)$$

If we now combine like terms (i.e., terms that contain μ^2 and that contain -2μ), we obtain:

$$\propto \exp\left(-\frac{1}{2}\left[\left(\frac{n}{\sigma^2}+\frac{1}{v_0^2}\right)\mu^2 - 2\mu\left[\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i + \frac{1}{v_0^2}m_0\right]\right]\right)$$

Because $\frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$, then $\sum_{i=1}^{n} x_i = \bar{x} \times n$.

$$\propto \exp\left(-\frac{1}{2}\left[\left(\frac{n}{\sigma^2}+\frac{1}{v_0^2}\right)\mu^2 - 2\mu\left[\frac{n}{\sigma^2}\bar{x} + \frac{1}{v_0^2}m_0\right]\right]\right)$$

If you compare this expression to the one we were trying to obtain in Eqn. 1 (in yellow), you should readily see that:

$$\frac{1}{v_1^2} = \frac{n}{\sigma^2} + \frac{1}{v_0^2}$$

$$\left(\frac{1}{v_1^2}\right)m_1 = \left[\frac{n}{\sigma^2}\bar{x} + \frac{1}{v_0^2}m_0\right]$$

Therefore, we find that:

$$v_1^2 = \left(\frac{n}{\sigma^2} + \frac{1}{v_0^2}\right)^{-1}$$

$$m_1 = \left(\frac{n}{\sigma^2} + \frac{1}{v_0^2}\right)^{-1}\left[\frac{n}{\sigma^2}\bar{x} + \frac{1}{v_0^2}m_0\right]$$