Bayes Example: Conjugate Normal

We are given the following setup:

Likelihood: $p(x \mid \mu) \sim N(\mu, \sigma^2)$, σ known;

Prior: $p(\mu \mid \theta, \tau) \sim N(\theta, \tau^2)$, hyperparameters θ, τ known.

In this problem we are interested in making inferences about μ. In Bayesian statistics, inference is based on the *posterior distribution* of the parameter of interest, so we need to find the posterior distribution for μ. The posterior is the result of applying Bayes' Rule.

It is convenient to rewrite the variances in terms of *precisions* (precision = 1/variance). This gives an equivalent parameterization, while at the same time avoids a lot of tedious calculations with fractions. So let $S = 1/\sigma^2$, $T = 1/\tau^2$.

Using Bayes' Rule, the posterior $p(\mu \mid x)$ is given by:

$$p(\mu \mid x) = \frac{p(x \mid \mu)\, p(\mu)}{\int_\mu p(x \mid \mu)\, p(\mu)\; d\mu}\,.$$

The denominator is a constant with respect to the unknown parameter μ, so doesn't tell us anything about the form of the distribution for μ and we can just treat it as a constant of proportionality. (You don't have to believe me, you could instead keep track of the constant throughout the calculations. Assuming that you don't make any mistakes you will end up with the same answer, it will just take you very much longer to get to it!)

So, we now have that

$$\begin{aligned}
p(\mu \mid x) &\propto p(x \mid \mu)\, p(\mu) && \text{proportionality} \\
&= \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \times \frac{1}{\sqrt{2\pi}\tau}\, e^{-\frac{1}{2}\left(\frac{\mu-\theta}{\tau}\right)^2} && \text{substitute normal densities} \\
&= C\, e^{-\frac{S}{2}(x-\mu)^2}\, e^{-\frac{T}{2}(\mu-\theta)^2} && \text{factor out constants} \\
&= C\, e^{-\frac{1}{2}\left[S(x-\mu)^2 + T(\mu-\theta)^2\right]} && \text{regroup terms.} \qquad (1)
\end{aligned}$$

Now just focus on the exponent term inside [ ] in (1) above:

$$
\begin{aligned}
\left[S\,(x-\mu)^2 + T\,(\mu-\theta)^2\right] &= S(x^2 - 2x\mu + \mu^2) + T(\mu^2 - 2\theta\mu + \theta^2) && \text{expand squared terms} \\
&= (S+T)\mu^2 - 2(Sx + T\theta)\mu + (Sx^2 + T\theta^2) && \text{combine terms} \\
&= (S+T)\left[\mu^2 - 2\frac{Sx+T\theta}{S+T}\mu + \frac{Sx^2+T\theta^2}{S+T}\right] && \text{factor out } S+T \\
&= (S+T)\left[\mu - \left(\frac{Sx+T\theta}{S+T}\right)\right]^2 + \text{other stuff} && \text{rewrite for convenience.}
\end{aligned}
$$

Remember this last calculation was on part of the exponent from (1); plugging in we get:

$$(1) = K\, e^{-\frac{1}{2}(S+T)\left[\mu-\left(\frac{Sx+T\theta}{S+T}\right)\right]^2},$$

which we recognize as having the form of a $N\left(\frac{Sx+T\theta}{S+T}, \frac{1}{S+T}\right)$.

Thus, the posterior mean is $\frac{Sx+T\theta}{S+T} = \frac{T}{S+T}\,\theta + \frac{S}{S+T}\,x$. A nice interpretation of the posterior mean is that it is a weighted average of the prior mean and observed data, with weights proportional to the precisions.

A very similar calculation can be carried out to get the posterior for μ when there is a sample of $n > 1$ data points.

This 'trick' of identifying the form of the posterior is heavily used in Bayesian inference when conjugate families are used. This is used by Lönnstedt and Speed in their derivation of the B-stat.

Bayesian inference is not restricted to conjugate families though, any distribution may be used as a prior. It is even possible to use priors that are not true probability densities (*i.e.* don't integrate to 1). However, when nonconjugate distributions are used there is usually not a closed-form representation of the posterior. In this case, numerical/computational methods are used (*e.g.* numerical integration, Markov Chain Monte Carlo).