Bayesian inference for a mean - derivation notes

1 Only the mean μ is unknown: Normal conjugate prior

- The likelihood:

$$
\begin{aligned}
y_1, \cdots, y_n \mid \mu, \sigma &\overset{i.i.d.}{\sim} \text{Normal}(\mu, \sigma) && (1) \\
L(\mu) = p(y_1, \cdots, y_n \mid \mu, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \phi^{\frac{1}{2}} \exp\left(-\frac{\phi}{2}(y_i - \mu)^2\right) \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \phi^{\frac{n}{2}} \exp\left(-\frac{\phi}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right)
\end{aligned}
\tag{2}
$$

Note that σ is assumed known, therefore the likelihood function is only in terms of μ, i.e. $L(\mu)$.

- The prior distribution:

$$
\begin{aligned}
\mu \mid \sigma &\sim \text{Normal}(\mu_0, \sigma_0) && (3) \\
\pi(\mu) &= \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right) \\
&= \frac{1}{\sqrt{2\pi}} \phi_0^{\frac{1}{2}} \exp\left(-\frac{\phi_0}{2}(\mu - \mu_0)^2\right) && (4)
\end{aligned}
$$

Similarly the prior is expressed as $\mu \mid \sigma$ to represent that σ is known. When Bayes' rule is applied to derive the posterior of μ, these manipulations require careful consideration regarding what is known. Therefore any "constants" or known quantities can be dropped/added with the proportionality sign "$\propto$".

- The posterior:

$$
\begin{aligned}
\pi(\mu \mid y_1, \cdots, y_n, \sigma) &\propto \pi(\mu) L(\mu) \\
&\propto \exp\left(-\frac{\phi_0}{2}(\mu - \mu_0)^2\right) \times \exp\left(-\frac{\phi}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right) \\
&\propto \exp\left(-\frac{1}{2}(\phi_0 + n\phi)\mu^2 + \frac{1}{2}(2\phi_0\mu_0 + 2n\phi\bar{y})\mu\right) \\
\texttt{[complete the square]} &\propto \exp\left(-\frac{1}{2}(\phi_0 + n\phi)\left(\mu - \frac{\phi_0\mu_0 + n\phi\bar{y}}{\phi_0 + n\phi}\right)^2\right)
\end{aligned}
\tag{5}
$$

Looking at the final expression closely, one recognizes this as a Normal density with a precision instead of variance parameter. Specifically we recognize $(\phi_0 + n\phi)$ as the normal precision and $(\frac{\phi_0\mu_0 + n\bar{y}\phi}{\phi_0 + n\phi})$ as the normal mean, and we see the posterior distribution of μ,

$$\mu \mid y_1, \cdots, y_n, \sigma \sim \text{Normal}\left(\frac{\phi_0\mu_0 + n\phi\bar{y}}{\phi_0 + n\phi}, \sqrt{\frac{1}{\phi_0 + n\phi}}\right). \tag{6}$$

2 Only the standard deviation σ is unknown: Gamma conjugate prior

- The likelihood:

$$\begin{aligned}
y_1, \cdots, y_n \mid \mu, \sigma &\overset{i.i.d.}{\sim} \text{Normal}(\mu, \sigma) &(7)\\
L(1/\sigma^2) = p(y_1, \cdots, y_n \mid \mu, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right)\\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n (1/\sigma^2)^{\frac{n}{2}} \exp\left(-\frac{(1/\sigma^2)}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right) &(8)
\end{aligned}$$

Note that μ is assumed known, therefore the likelihood function is only in terms of $1/\sigma^2$, i.e. $L(1/\sigma^2)$.

- The prior distribution:

$$\begin{aligned}
1/\sigma^2 \mid \mu &\sim \text{Gamma}(\alpha, \beta) &(9)\\
\pi(1/\sigma^2) &= \frac{\beta^\alpha}{\Gamma(\alpha)}(1/\sigma^2)^{\alpha-1}\exp(-\beta(1/\sigma^2)) &(10)
\end{aligned}$$

Similarly the prior is expressed as $1/\sigma^2 \mid \mu$ to represent that μ is known. When Bayes' rule is applied to derive the posterior of $1/\sigma^2$, these manipulations require careful consideration regarding what is known. Therefore any "constants" or known quantities can be dropped/added with the proportionality sign "$\propto$".

- The posterior:

$$\begin{aligned}
\pi(1/\sigma^2 \mid y_1, \cdots, y_n, \mu) &\propto \pi(1/\sigma^2)L(1/\sigma^2)\\
&\propto (1/\sigma^2)^{\alpha-1}\exp(-\beta(1/\sigma^2))(1/\sigma^2)^{\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i - \mu)^2(1/\sigma^2)\right)\\
\texttt{[combine the powers]} &= (1/\sigma^2)^{\alpha+\frac{n}{2}-1}\exp\left(-\left(\beta + \frac{1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right)(1/\sigma^2)\right) &(11)
\end{aligned}$$

Looking at the final expression closely, one recognizes this as a Gamma density:

$$1/\sigma^2 \mid y_1, \cdots, y_n, \mu \sim \text{Gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right). \tag{12}$$