Irzam Hardiansyah
irzam.hardiansyah@kth.se

FD3447: Assignment 2

Problem 1 Conjugate Priors

Q1.1 Prove that the posterior distribution of the mean of normal-distribution data with a known variance is as follows:

$$N\left(\mu\Big|\frac{N\sigma_0^2}{\sigma^2+N\sigma_0^2}\bar{x}+\frac{\sigma^2}{\sigma^2+N\sigma_0^2}\mu_0,\left(\frac{1}{\sigma_0^2}+\frac{N}{\sigma^2}\right)^{-1}\right)$$

Answer: The posterior of the mean can be written as:

$$f(\mu|\mathcal{D};\sigma^2,\sigma_0^2)\propto f(\mathcal{D}|\mu;\sigma^2)f(\mu;\sigma_0^2)$$

The likelihood pdf for a sample of N obs from normally-distributed data is:

$$\begin{aligned}f(\mathcal{D}|\mu;\sigma^2)&\propto\prod_{i=1}^{N}\exp^{-\frac{1}{2}\left(\frac{\mu-x_i}{\sigma}\right)^2}\\&\propto\exp^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{\mu-x_i}{\sigma^2}\right)^2}=\exp^{-\frac{1}{2}\left(\frac{\mu-\bar{x}}{\sigma/\sqrt{N}}\right)^2}\end{aligned}$$

where $\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i$ as obtained from the Maximum Likelihood Estimation (MLE) of the mean, and the variance of the mean of N i.i.d random variables is σ^2/N (proof not given as it can easily be found elsewhere)

And the conjugate prior pdf is:

$$f(\mu;\sigma_0^2)\propto\exp^{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2}$$

Thus, the posterior pdf can be written as:

$$f(\mu|\mathcal{D};\sigma^2,\sigma_0^2)\propto\exp^{-\frac{1}{2}\left(\left(\frac{\mu-\bar{x}}{\sigma/\sqrt{N}}\right)^2+\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right)}$$

and by the same ML estimation of the mean and variance (see Appendix for detailed derivation), we obtain
$\bar{\bar{x}}=\frac{\sigma_0^2}{\sigma^2/N+\sigma_0^2}\bar{x}+\left(1-\frac{\sigma_0^2}{\sigma^2/N+\sigma_0^2}\right)\mu_0$ and $\bar{\bar{\sigma}}^2=\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/N}\right)^{-1}$

Thus, posterior pdf is the normal distribution with mean and variance as given above:

$$f(\mu|\mathcal{D})\sim N\left(\bar{\bar{x}},\bar{\bar{\sigma}}^2\right)$$

Q1.2 Prove that the posterior distribution of the variance of normal-distribution data with a known mean is as follows:

$$IG\left(\sigma^2\middle|\alpha+\frac{N}{2},\beta+\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\right)$$

Answer: The posterior of the variance can be written as:

$$f(\sigma^2|\mathcal{D};\mu)\propto f(\mathcal{D}|\sigma^2;\mu)f(\sigma^2)$$

The likelihood pdf for a sample of N obs from normally-distributed data is:

$$\begin{aligned}f(\mathcal{D}|\sigma^2;\mu)&\propto\prod_{i=1}^{N}\frac{1}{\sqrt{\sigma^2}}\exp^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}\\&\propto\left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}}\exp^{-\frac{1}{\sigma^2}\left(\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\right)}\end{aligned}$$

And the conjugate prior pdf is:

$$\begin{aligned}f(\sigma^2)&\sim IG\left(\alpha,\beta\right)\\&\propto\left(\frac{1}{\sigma^2}\right)^{\alpha+1}\exp^{-\frac{\beta}{\sigma^2}}\end{aligned}$$

Thus, the posterior pdf can be written as:

$$\begin{aligned}f(\sigma^2|\mathcal{D};\mu)&\propto\left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}}\exp^{-\frac{1}{\sigma^2}\left(\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\right)}*\left(\frac{1}{\sigma^2}\right)^{\alpha+1}\exp^{-\frac{\beta}{\sigma^2}}\\&\propto\left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{N}{2}+1}\exp^{-\frac{1}{\sigma^2}\left(\beta+\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\right)}\end{aligned}$$

or,

$$f(\sigma^2|\mathcal{D})\sim IG\left(\bar{\alpha},\bar{\beta}\right)$$

where $\bar{\alpha}=\alpha+\frac{N}{2}$ and $\bar{\beta}=\beta+\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2$.

Q1.3 Prove that the posterior distributions of the mean and variance of normal-distribution data are, respectively, as follows:

$$N\left(\mu\middle|\frac{N}{N+n_0}\bar{x}+\frac{n_0}{N+n_0}\mu_0,(N\tau+n_0\tau)^{-1}\right)$$

$$Ga\left(\tau\middle|\alpha+\frac{N}{2},\beta+\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2+\frac{Nn_0}{2(N+n_0)}(\bar{x}-\mu_0)^2\right)$$

Answer: The posteriors can be found by treating them in a hierarchical way - first, we obtain the posterior of the mean and then use the posterior mean to obtain the posterior of the variance:

$$f(\mu|\mathcal{D};\tau) \propto f(\mathcal{D}|\mu;\tau)f(\mu;\tau)$$

$$f(\tau|\mathcal{D};\bar{\bar{x}}) \propto f(\mathcal{D}|\tau;\bar{\bar{x}})f(\tau;\bar{\bar{x}})$$

where $\bar{\bar{x}}$ is the posterior mean and we use precision, $\tau = \frac{1}{\sigma^2}$, as the dispersion parameter.

** Then, for the posterior mean:*

The likelihood pdf for a sample of N obs from normally-distributed data is (from Q1.1 by substituting $\frac{1}{\sigma^2/N} = N\tau$):

$$\begin{aligned} f(\mathcal{D}|\mu;\tau) &\propto \prod_{i=1}^{N} \exp^{-\frac{\tau}{2}(\mu-x_i)^2} \\ &\propto \exp^{-\frac{\tau}{2}\sum_{i=1}^{N}(\mu-x_i)^2} = \exp^{-\frac{N\tau}{2}(\mu-\bar{x})^2} \end{aligned}$$

And the conjugate prior pdf for the mean is:

$$f(\mu;\tau) \propto \exp^{-\frac{\tau_0}{2}(\mu-\mu_0)^2} = \exp^{-\frac{n_0\tau}{2}(\mu-\mu_0)^2}$$

Therefore, the posterior pdf for the mean is:

$$f(\mu|\mathcal{D};\tau) \propto \exp^{-\frac{1}{2}\left(N\tau(\mu-\bar{x})^2+n_0\tau(\mu-\mu_0)^2\right)}$$

and analogously with Q1.1, by substituting $\sigma^2/N = (N\tau)^{-1}$ and $\sigma_0^2 = (n_0\tau)^{-1}$, we obtain

$\bar{\bar{x}} = \frac{(n_0\tau)^{-1}}{(N\tau)^{-1}+(n_0\tau)^{-1}}\bar{x} + \left(1 - \frac{(n_0\tau)^{-1}}{(N\tau)^{-1}+(n_0\tau)^{-1}}\right)\mu_0 = \frac{N}{N+n_0}\bar{x} + \left(1 - \frac{N}{N+n_0}\right)\mu_0$

and $\bar{\bar{\tau}} = (n_0\tau + N\tau)$

hence, posterior pdf of the mean is the normal distribution with mean and variance as given above:

$$f(\mu|\mathcal{D}) \sim N\left(\bar{\bar{x}}, \bar{\bar{\tau}}^{-1}\right)$$

** Now, for the posterior precision:*

The likelihood pdf for a sample of N obs from normally-distributed data, by plugging $\bar{\bar{x}}$ in place of μ, is:

$$
\begin{aligned}
f(\mathcal{D}|\tau;\bar{\bar{x}}) &\propto \prod_{i=1}^{N} \sqrt{\tau}\exp^{-\frac{\tau}{2}(x_i-\bar{\bar{x}})^2} \\
&\propto (\tau)^{\frac{N}{2}} \exp^{-\tau\left(\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{\bar{x}})^2\right)}
\end{aligned}
$$

And the conjugate prior pdf for the precision is:

$$
\begin{aligned}
f(\tau;\bar{\bar{x}}) &\sim N - Ga\left(\alpha,\beta,\bar{\bar{x}}\right) \\
&\propto \left[(n_0\tau)^{\frac{1}{2}}\exp^{-\frac{n_0\tau}{2}(\bar{\bar{x}}-\mu_0)^2} \times (\tau)^{\alpha-1}\exp^{-\beta\tau}\right]
\end{aligned}
$$

where N-Ga(.) is a Normal-Gamma distribution function.

Then, the posterior pdf of the precision can be written as,

$$
\begin{aligned}
f(\tau|\mathcal{D};\bar{\bar{x}}) &\propto (\tau)^{\frac{N}{2}} \exp^{-\tau\left(\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{\bar{x}})^2\right)} \times (n_0\tau)^{\frac{1}{2}}\exp^{-\frac{n_0\tau}{2}(\bar{\bar{x}}-\mu_0)^2} \times (\tau)^{\alpha-1}\exp^{-\beta\tau} \\
&\propto (\tau)^{\alpha+\frac{N+1}{2}-1}\exp^{-\tau\left(\beta+\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{\bar{x}})^2+\frac{n_0}{2}(\bar{\bar{x}}-\mu_0)^2\right)} \\
&\propto (\tau)^{\alpha+\frac{N}{2}-1}\exp^{-\tau\left(\beta+\frac{1}{2}\sum_{i=1}^{N}(x_i-\frac{N}{N+n_0}\bar{x}-\frac{n_0}{N+n_0}\mu_0)^2+\frac{n_0}{2}(\frac{N}{N+n_0}\bar{x}+\frac{n_0}{N+n_0}\mu_0-\mu_0)^2\right)} \\
&\propto (\tau)^{\alpha+\frac{N}{2}-1}\exp^{-\tau\left(\beta+\frac{1}{2}\sum_{i=1}^{N}\left(\frac{N(x_i-\bar{x})+n_0(x_i-\mu_0)}{N+n_0}\right)^2+\frac{n_0}{2}\left(\frac{N}{N+n_0}(\bar{x}-\mu_0)\right)^2\right)} \\
&\propto (\tau)^{\alpha+\frac{N}{2}-1}\exp^{-\tau\left(\beta+\frac{1}{2}\sum_{i=1}^{N}\left(\frac{N(x_i-\bar{x})+n_0(x_i-\bar{x}+\bar{x}-\mu_0)}{N+n_0}\right)^2+\frac{n_0}{2}\left(\frac{N}{N+n_0}(\bar{x}-\mu_0)\right)^2\right)} \\
&\propto (\tau)^{\alpha+\frac{N}{2}-1}\exp^{-\tau\left(\beta+\frac{1}{2}\sum_{i=1}^{N}\left(\frac{N(x_i-\bar{x})+n_0(x_i-\bar{x})+n_0(\bar{x}-\mu_0)}{N+n_0}\right)^2+\frac{n_0}{2}\left(\frac{N}{N+n_0}(\bar{x}-\mu_0)\right)^2\right)} \\
&\propto (\tau)^{\alpha+\frac{N}{2}-1}\exp^{-\tau\left(\beta+\frac{1}{2}\sum_{i=1}^{N}\left((x_i-\bar{x})+\frac{n_0}{N+n_0}(\bar{x}-\mu_0)\right)^2+\frac{n_0}{2}\left(\frac{N}{N+n_0}(\bar{x}-\mu_0)\right)^2\right)} \\
&\propto (\tau)^{\alpha+\frac{N}{2}-1}\exp^{-\tau\left(\beta+\frac{1}{2}\sum_{i=1}^{N}\left((x_i-\bar{x})^2+2(x_i-\bar{x})\frac{n_0}{N+n_0}(\bar{x}-\mu_0)+\frac{n_0^2}{(N+n_0)^2}(\bar{x}-\mu_0)^2\right)+\frac{n_0}{2}\left(\frac{N^2}{(N+n_0)^2}(\bar{x}-\mu_0)^2\right)\right)} \\
&\propto (\tau)^{\alpha+\frac{N}{2}-1}\exp^{-\tau\left(\beta+\frac{1}{2}\left(\sum_{i=1}^{N}(x_i-\bar{x})^2+2\sum_{i=1}^{N}(x_i-\bar{x})\frac{n_0}{N+n_0}(\bar{x}-\mu_0)+\frac{Nn_0^2}{(N+n_0)^2}(\bar{x}-\mu_0)^2\right)+\frac{1}{2}\frac{n_0N^2}{(N+n_0)^2}(\bar{x}-\mu_0)^2\right)} \\
&\propto (\tau)^{\alpha+\frac{N}{2}-1}\exp^{-\tau\left(\beta+\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2+\frac{1}{2}\frac{Nn_0^2}{(N+n_0)^2}(\bar{x}-\mu_0)^2+\frac{1}{2}\frac{n_0N^2}{(N+n_0)^2}(\bar{x}-\mu_0)^2\right)} \\
&\propto (\tau)^{\alpha+\frac{N}{2}-1}\exp^{-\tau\left(\beta+\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2+\frac{Nn_0}{2(N+n_0)}(\bar{x}-\mu_0)^2\right)} \\
&\propto (\tau)^{\bar{\bar{\alpha}}-1}\exp^{-\bar{\bar{\beta}}\tau}
\end{aligned}
$$

where $\bar{\bar{\alpha}} = \alpha + \frac{N}{2}$ and $\bar{\bar{\beta}} = \beta + \frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2 + \frac{Nn_0}{2(N+n_0)}(\bar{x}-\mu_0)^2$

Therefore, the posterior pdf of the precision is the following gamma distribution:

$$
f(\tau|\mathcal{D}) \sim Ga\left(\bar{\bar{\alpha}},\bar{\bar{\beta}}\right)
$$

Appendix

Maximum Likelihood Estimation (MLE) of the mean and variance of the posterior pdf of the unknown normal mean parameter (with known variance)

To derive the ML estimate of the mean of this posterior pdf (Q.1.1):

$$f(\mu|\mathcal{D};\sigma^2,\sigma_0^2) \propto (\sigma^2)^{-N/2}(\sigma_0^2)^{-1/2}\exp^{-\frac{1}{2}\left(\left(\frac{\mu-\bar{x}}{\sigma/\sqrt{N}}\right)^2+\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right)}$$

we take the logarithm of the above kernel density function:

$$\begin{aligned} log\Big(f(\mu|\mathcal{D};\sigma^2,\sigma_0^2)\Big) &\propto log((\sigma^2)^{-N/2}) + log((\sigma_0^2)^{-1/2}) + log\left(\exp^{-\frac{1}{2}\left(\left(\frac{\mu-\bar{x}}{\sigma/\sqrt{N}}\right)^2+\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right)}\right) \\ &\propto -\frac{N}{2}log(\sigma^2) - \frac{1}{2}log(\sigma_0^2) - \frac{1}{2}\left(\frac{\mu-\bar{x}}{\sigma/\sqrt{N}}\right)^2 - \frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2 \end{aligned}$$

then we maximize this log-posterior function w.r.t. the mean parameter, μ, by taking the following:

$$\nabla_\mu log\Big(f(\mu|\mathcal{D};\sigma^2,\sigma_0^2)\Big) = 0$$

Thus:

$$\begin{aligned} -\frac{\partial}{\partial\mu}\left(\frac{N}{2}log(\sigma^2) + \frac{1}{2}log(\sigma_0^2) + \frac{1}{2}\left(\frac{\mu-\bar{x}}{\sigma/\sqrt{N}}\right)^2 + \frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right) &= 0 \\ \left(\frac{\mu-\bar{x}}{\sigma/\sqrt{N}}\right)\left(\frac{1}{\sigma/\sqrt{N}}\right) + \left(\frac{\mu-\mu_0}{\sigma_0}\right)\left(\frac{1}{\sigma_0}\right) &= 0 \\ \frac{\sigma_0^2(\mu-\bar{x})}{\sigma_0^2\sigma^2/N} + \frac{\sigma^2/N(\mu-\mu_0)}{\sigma_0^2\sigma^2/N} &= 0 \\ \frac{(\sigma_0^2+\sigma^2/N)\mu}{\sigma_0^2\sigma^2/N} &= \frac{\sigma_0^2\bar{x}+\sigma^2/N\mu_0}{\sigma_0^2\sigma^2/N} \\ \hat{\mu} &= \frac{\sigma_0^2\bar{x}+\sigma^2/N\mu_0}{\sigma_0^2+\sigma^2/N} = \bar{\bar{x}} \end{aligned}$$

Now, to determine the variance, it is easier to work with the precision parameter $\tau = \frac{1}{\sigma^2}$. Each additional observation increases the precision in an additive manner, and because the posterior incorporates *"an additional observation"* from the prior, the precision of the posterior is just the sum of the precisions of the likelihood ($\bar{\tau}$) and of the prior (τ_0). Thus:

$$\begin{aligned} \bar{\bar{\tau}} &= \bar{\tau} + \tau_0 \\ &= \frac{1}{\sigma^2/N} + \frac{1}{\sigma_0^2} \end{aligned}$$

and, hence:

$$\bar{\bar{\sigma}}^2 = \frac{1}{\bar{\bar{\tau}}} = \Big(\frac{1}{\sigma^2/N} + \frac{1}{\sigma_0^2}\Big)^{-1}$$