# DD2447 Statistical Methods in Applied Computer Science Assignment 2

Leonardo De Clara, Riccardo Sena

December 4, 2023

## Exercise 1

### Q 1.1

Let's denote our data points as $x_i s$ which are i.i.d. and normal distributed with parameter $\mu$ and $\sigma^2$, where $\sigma^2$ is known. Hence the likelihood of $X = (x_1, ..., x_N)$ is:

$$p(X|\mu) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2)$$

The prior over the parameter $\mu$ is still a normal distribution with parameter $\mu_0$ and $\sigma_0^2$:

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}$$

Since remembering that the Normal distribution is a conjugate prior to itself, the posterior distribution will also be a Normal. By the definition of posterior distribution (from Bayes' theorem) and by the fact that the marginal $p(X)$ does not depend on $\mu$ we know that:

$$p(\mu|X) = \frac{p(X|\mu)p(\mu)}{p(X)} \propto p(X|\mu)p(\mu)$$

$$p(\mu|X) \propto \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2) \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp(-\frac{(\mu - \mu_0)^2}{2\sigma_0^2})$$

By the property of the exponential we can sum the exponent and if we denote the sample mean with $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$, we obtain:

$$-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2 - \frac{(\mu-\mu_0)^2}{2\sigma_0^2} = -\frac{\sum_{i=1}^{N} x_i^2}{2\sigma^2} + \frac{2N\bar{x}\mu}{2\sigma^2} - \frac{N\mu^2}{2\sigma^2} - \frac{\mu^2 - 2\mu\mu_0 + \mu_0^2}{2\sigma_0^2} =$$

$= C - \frac{\mu^2(N\sigma_0^2+\sigma^2)}{2\sigma^2\sigma_0^2} + \frac{\mu(2\sigma_0^2 N\bar{x}+2\sigma^2\mu_0)}{2\sigma^2\sigma_0^2}$,

where we denoted with C the terms that don't depend on $\mu$.
Since we are looking for a Normal distribution our exponent should be a square of the type $-\frac{1}{2\psi^2}(\mu-\epsilon)^2$, where $\epsilon$ and $\psi$ are respectively the mean and the standard deviation of the distribution. If we gather together all the terms that are constant with respect to $\mu$ into a constant K we can rewrite the posterior as this:

$$p(\mu|X) \propto K \exp(-\frac{1}{\frac{2\sigma^2\sigma_0^2}{N\sigma_0^2+\sigma^2}}(\mu^2 - \frac{\mu(2\sigma_0^2 N\bar{x}+2\sigma^2\mu_0)}{N\sigma_0^2+\sigma^2}))$$

The term $\frac{\mu(2\sigma_0^2 N\bar{x}+2\sigma^2\mu_0)}{N\sigma_0^2+\sigma^2}$ represents the double product between $\mu$ and its mean. So if we multiply and divide the expression by $\exp(-\frac{1}{\frac{2\sigma^2\sigma_0^2}{N\sigma_0^2+\sigma^2}}(\frac{(\sigma_0^2 N\bar{x}+\sigma^2\mu_0)}{N\sigma_0^2+\sigma^2})^2)$ (which is constant w.r.t. $\mu$) we get:

$$p(\mu|X) \propto K \exp(-\frac{1}{\frac{2\sigma^2\sigma_0^2}{N\sigma_0^2+\sigma^2}}(\mu - \frac{(\sigma_0^2 N\bar{x}+\sigma^2\mu_0)}{N\sigma_0^2+\sigma^2})^2)$$

Setting $\epsilon = \frac{(\sigma_0^2 N\bar{x}+\sigma^2\mu_0)}{N\sigma_0^2+\sigma^2}$ and $\psi^2 = \frac{\sigma^2\sigma_0^2}{N\sigma_0^2+\sigma^2} = (\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2})^{-1}$, the posterior is exactly a Normal distribution: $\mu|X \sim N(\epsilon, \psi^2)$.

## Q 1.2

As in exercise Q 1.1 we have that our data points are normally distributed but the difference now is that $\mu$ is known and $\sigma$ is not:

$$p(X|\sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2)$$

The prior on $\sigma^2$ is:

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}(1/\sigma^2)^{\alpha+1}\exp(-\beta/\sigma^2)$$

The posterior will be:

$$p(\sigma^2|X) \propto p(X|\sigma^2)p(\sigma^2) = \frac{1}{(2\pi)^{N/2}}\frac{1}{(\sigma^2)^{N/2+\alpha+1}}\frac{\beta^\alpha}{\Gamma(\alpha)}\exp(-\frac{1}{\sigma^2}(\beta + \sum_{i=1}^{N}\frac{(x_i-\mu)^2)}{2})$$

Since the Inverse Gamma distribution for the variance is a conjugate prior to the Normal distribution, also the posterior distribution will be an Inverse Gamma distribution:

$$p(\sigma^2|X) \propto \frac{1}{(\sigma^2)^{N/2+\alpha+1}}\exp(-\frac{1}{\sigma^2}(\beta + \sum_{i=1}^{N}\frac{(x_i-\mu)^2)}{2})$$

Which is (with the correct normalizing constant) an Inverse Gamma distribution with $\alpha_0 = \alpha + \frac{N}{2}$ and $\beta_0 = \beta + \frac{1}{2}\sum_{i=1}^{N}(x_i - \mu)^2$, leading to $\sigma^2|X \sim IG(\alpha_0, \beta_0)$.

## Q 1.3

The posterior distribution of the mean will be the same as in exercise Q 1.1, since we are considering that $\tau$ (hence the variance) is given. In this case we have that the prior is $\mu|\tau \sim N(\mu|\mu_0, (n_0\tau)^{-1})$. Substituting $\sigma^2$ with $\tau^{-1}$ and $\sigma_0^2$ with $(n_0\tau)^{-1}$ we get for the mean of the posterior:

$$\frac{(\sigma_0^2 N\bar{x}+\sigma^2\mu_0)}{N\sigma_0^2+\sigma^2} = \frac{(\frac{1}{n_0\tau}N\bar{x}+\frac{1}{\tau}\mu_0)}{\frac{1}{n_0\tau}N+\frac{1}{\tau}} = \frac{\tau N\bar{x}+n_0\tau\mu_0}{N\tau+n_0\tau} = \frac{N\bar{x}+n_0\mu_0}{N+n_0}$$

And for the variance of the posterior:

$$(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2})^{-1} = (n_0\tau + N\tau)^{-1}$$

So we get $\mu|\tau, X \sim N(\frac{N\bar{x}+n_0\mu_0}{N+n_0}, (n_0\tau + N\tau)^{-1})$.

Now we want to find the distribution of $\tau|X$. We now have a $Gamma(\alpha, \beta)$ prior for $\tau$ which is a conjugate prior for the precision with respect to a normal distribution.
We can get $p(\tau|X)$ by looking at the joint distribution $p(\tau, \mu|X)$ and then integrating over all possible values of $\mu$. First thanks to Bayes' rule we can write:

$$p(\tau, \mu|X) = \frac{p(x,\mu,\tau)}{p(X)} \propto p(X|\tau, \mu)p(\mu|\tau)p(\tau) =$$

$$= \frac{1}{(2\pi\frac{1}{\tau})^{\frac{N}{2}}} e^{-\frac{\tau}{2}\sum_{i=1}^{N}(x_i-\mu)^2} \sqrt{\frac{n_0\tau}{2\pi}} e^{-\frac{n_0\tau(\mu-\mu_0)^2}{2}} \frac{\tau^{\alpha-1}e^{-\beta\tau}\beta^{\alpha}}{\Gamma(\alpha)} \propto \tau^{\frac{N}{2}} e^{-\frac{\tau}{2}\sum_{i=1}^{N}(x_i-\mu)^2} \sqrt{\tau} e^{-\frac{n_0\tau(\mu-\mu_0)^2}{2}} \tau^{\alpha-1}e^{-\beta\tau}$$

Thanks to the fact that (using the notation $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$):

$$\sum_{i=1}^{N}(x_i - \mu)^2 = \sum_{i=1}^{N}(x_i - \bar{x})^2 + N(\bar{x} - \mu)^2$$

We get that $p(\tau, \mu|X)$ is proportional to (getting rid of all the constants):

$$\tau^{\alpha+\frac{N}{2}-1} e^{-\tau(\beta+\frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2)} \tau^{\frac{1}{2}} e^{-\frac{\tau}{2}(n_0(\mu-\mu_0)^2+N(\bar{x}-\mu)^2)}$$

We notice now that only the last exponential depends on $\mu$. So we can integrate it over all possible values of $\mu$:

$$\int \tau^{\frac{1}{2}} e^{-\frac{\tau}{2}(n_0(\mu-\mu_0)^2+N(\bar{x}-\mu)^2)} d\mu = \int \tau^{\frac{1}{2}} e^{-\frac{\tau(n_0+N)}{2}(\mu^2 - 2\mu\frac{n_0\mu_0+N\bar{x}}{n_0+N} + \frac{n_0\mu_0^2+N\bar{x}^2}{n_0+N})} d\mu =$$

$$= \tau^{\frac{1}{2}} e^{\frac{\tau(n_0+N)}{2}((\frac{n_0\mu_0+N\bar{x}}{n_0+N})^2 - \frac{n_0\mu_0^2+N\bar{x}^2}{n_0+N})} \int e^{-\frac{\tau(n_0+N)}{2}(\mu - \frac{n_0\mu_0+N\bar{x}}{n_0+N})^2} d\mu \propto e^{-\frac{n_0 N\tau}{2(n_0+N)}(\bar{x}-\mu_0)^2}$$

where we used the fact that the integral is equal to a constant times $\tau^{-\frac{1}{2}}$, since $\mu|\tau$ is distributed as a Normal. Therefore, plugging in everything we've got:

$$p(\tau|X) \propto \tau^{\alpha+\frac{N}{2}-1} e^{-\tau(\beta+\frac{1}{2}\sum\limits_{i=1}^{N}(x_i-\bar{x})^2+\frac{n_0 N}{2(n_0+N)}(\bar{x}-\mu_0)^2)}$$

Which is a Gamma with $\alpha_0 = \alpha + \frac{N}{2}$ and $\beta_0 = \beta + \frac{1}{2}\sum\limits_{i=1}^{N}(x_i - \bar{x})^2 + \frac{n_0 N}{2(n_0+N)}(\bar{x} - \mu_0)^2$.