# TDA231 Homework 1: Theory

May 11, 2018

## Question 1

The likelihood of a parameter value $\theta = t$ given a dataset $D$ is the probability of obtaining $D$ given that $t$ is the true value (and the underlying model is correct, obviously). In this case, the probability of obtaining any specific datapoint $\mathbf{x}_i$ given parameters $\mu$, $\sigma^2$ is given by the distribution

$$P(\mathbf{X}_i = \mathbf{x}_i | \mu = \mathbf{m}, \sigma^2 = s) = ((2\pi)^p s)^{-\frac{1}{2}} \exp\Big( - \frac{(\mathbf{x}_i - \mathbf{m})^2}{2s} \Big)$$

where the "square" $(\mathbf{x}_i - \mathbf{m})^2$ is actually the vector product $(\mathbf{x}_i - \mathbf{m})^\top (\mathbf{x}_i - \mathbf{m})$. The observations in the dataset are independent, which means that the likelihood of a certain $\sigma^2 = s$ is given by the product of the probabilities of the observations given that $s$:

$$P(\sigma^2 = s | D, \mu = \mathbf{m}) = \prod_i P(\mathbf{X}_i = \mathbf{x}_i | \mu = \mathbf{m}, \sigma^2 = s) = ((2\pi)^p s)^{-\frac{n}{2}} \exp\Big( - \frac{\sum_i (\mathbf{x}_i - \mathbf{m})^2}{2s} \Big)$$

The Maximum Likelihood Estimator for $\sigma$, $\hat{\sigma}$, is then the value of $\sqrt{s}$ which maximises the above quantity. We will find $\hat{\sigma}^2$ instead, since obtaining $\hat{\sigma}$ is then easy. We seek:

$$\hat{\sigma}^2 = \max_s \big( ((2\pi)^p s)^{-\frac{n}{2}} \exp\Big( - \frac{\sum_i (\mathbf{x}_i - \mathbf{m})^2}{2s} \Big) \big)$$

We can take the logarithm of the quantity to be maximised without changing the result, and in doing so we find a constant term which can be dropped.

$$\hat{\sigma}^2 = \max_s \big( - \frac{n}{2} \log s - \frac{\sum_i (\mathbf{x}_i - \mathbf{m})^2}{2s} \big)$$

To find the maximising value of $s$, we now differentiate w.r.t. $s$ and set the result to equal 0. We arrive at:

$$\frac{n}{2s} = \frac{\sum_i (\mathbf{x}_i - \mathbf{m})^2}{2s^2}$$

which is solved by

$$s = \frac{\sum_i (\mathbf{x}_i - \mathbf{m})^2}{n} = \hat{\sigma}^2$$

We can then obtain $\hat{\sigma}$ by taking the square root.

# Question 2

## 2 (a)

We want the posterior distribution $P(\sigma^2 = s|D, \alpha, \beta)$ when using an inverse-Gamma$(\alpha, \beta)$ prior and obtaining data $D$. The posterior is defined as

$$P(\sigma^2 = s|D, \alpha, \beta) \propto P(D|\sigma^2 = s)P(\sigma^2 = s|\alpha, \beta)$$

where the first factor in the right-hand side is the likelihood, computed as part of Question 1; and the second factor is the prior, which is the distribution given by (1) in the assignment. Since the inverse-Gamma is a conjugate prior to the Gaussian distribution, we expect another inverse-Gamma distribution for the result:

$$\begin{aligned}
P(D|\sigma^2 = s)P(\sigma^2 = s|\alpha, \beta) = ((2\pi)^p s)^{-\frac{n}{2}} \exp\Big(-\frac{\sum_i(\mathbf{x}_i - \mathbf{m})^2}{2s}\Big)\frac{\beta^\alpha s^{-\alpha-1}}{\Gamma(\alpha)} \exp\Big(-\frac{\beta}{s}\Big) = \\
= \frac{(2\pi)^{\frac{pn}{2}}\beta^\alpha}{\Gamma(\alpha)} s^{-\alpha-\frac{n}{2}-1} \exp\Big(-\frac{\beta + \frac{1}{2}\sum_i(\mathbf{x}_i - \mathbf{m})^2}{s}\Big) \propto \\
\propto s^{-\alpha-\frac{n}{2}-1} \exp\Big(-\frac{\beta + \frac{1}{2}\sum_i(\mathbf{x}_i - \mathbf{m})^2}{s}\Big) \propto \\
\propto \text{inverse-Gamma}\Big(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_i(\mathbf{x}_i - \mathbf{m})^2\Big)
\end{aligned}$$

In other words, the posterior is a new inverse-Gamma distribution, with $\alpha$ updated by half the number of data points, and $\beta$ updated using the sample variance.

## 2 (b) **The posterior parameters for the two models will be different, otherwise the calculations are identical.**