Lecture 16

The Normal-Normal Conjugate Model Part 2

$$\hat{\theta}_{MMSE} = \frac{\frac{n\bar{x}}{\sigma^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}} + \frac{\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}} = \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n}} \frac{\frac{n\bar{x}}{\sigma^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}} \hat{\theta}_{MLE} + \frac{\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}} \frac{\tau^2}{\tau^2} E[\theta|\sigma^2]$$

Given τ^2 is our variance in the normal distribution, if τ is high, rho goes down, as in there is less shrinkage, and therefore less weight given to the data. Furthermore, as n approaches infinity, rho approached 0.

$$\rho = \frac{\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}$$

We said before the Laplace Prior is uniform (0,1), in terms of sigma, we get the following,

$$P(\theta|\sigma^2)\alpha 1$$

$$P(\theta|X,\sigma^2)\alpha P(X|\theta,\sigma^2)P(\theta|\sigma^2)\alpha P(X|\theta,\sigma^2)\alpha N(\bar{x},\frac{\sigma^2}{n})$$

$$N(\bar{x},\frac{\sigma^2}{n}) = N(\frac{\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\sigma^2}}{\frac{n\bar{x}}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{\frac{n\bar{x}}{\sigma^2}+\frac{1}{\tau^2}}) \rightarrow \frac{\sigma^2}{n} = \frac{1}{\frac{n}{\sigma^2}} \rightarrow \tau \text{ approaches } \infty$$

This doesn't exist, we cannot have a distribution with infinite variance. The mean of this normal distribution simplifies to $\bar{x}$, therefore, μ_0 could be any value, by convention we let it equal zero. Therefore, Laplace's prior is improper, however the posterior is always proper.

$$P(\theta|\sigma^2) = N(0,\infty)$$

Now we move onto Jeffrey's prior,

$$P_J(\theta|\sigma^2)\alpha\sqrt{I(\theta|\sigma^2)} = \sqrt{\frac{n}{\sigma^2}}\alpha 1 \alpha N(0,\infty)$$

This makes Jeffrey's prior the same as Laplace's prior.

In order to do Haldane's prior, we must do the pseudo-count interpretation. We want a pseudo-count interpretation of the hyperparameters μ_0 and τ_0. The best way to do this is to do a small reparameterization of the prior's τ^2. Recall that we know σ^2.

$$\tau^2 = \frac{\sigma^2}{n_0}$$

$$P(\theta|X,\sigma^2) = N(\frac{\frac{n\bar{x}}{\sigma^2} + \frac{n_0\mu_0}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{n_0}{\sigma^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{n_0}{\tau^2}})$$

$$= N(\frac{n\bar{x} + n_0\mu_0}{n+n_0}, \frac{\sigma^2}{n+n_0})$$

So n_0 represents the number of pseudoobservations. What does μ_0 represent? We let $Y_1, Y_2, ..., Y_{n0}$ be the "pseudodata". Let $\mu_0 = \bar{Y}$ be the sample average of the pseydodata and n_0 and μ_0 is the sum of the pseudodata.

$$\hat{\theta}_{MMSE} = \frac{n}{n+n_0}\bar{x} + \frac{n_0}{n+n_0}\mu_0$$

This is a very edible formula for $\hat{\theta}_{MMSE}$. What's the Haldane prior of total ignorance? $n_0 = 0$. Therefore,

$$N(\mu_0, \frac{\sigma^2}{n_0}) = N(0,\infty)\alpha 1$$

This means that all three objective priors we have studied are the same, Laplace's, Jeffrey's, and Haldane's priors are all the same, and proportional to 1. We are only missing the posterior predictive distribution for $n^* = 1$ observations. We want to see what tomorrow will bring as a distribution, so we are interested in the following quantity.

$$P(X_*|X,\sigma^2) = \int_\Theta P(X_*|\theta,\sigma^2)P(\theta|X,\sigma^2)d\theta$$

Where,

$$P(X_*|\theta,\sigma^2) = N(\theta,\sigma^2)$$

$$P(\theta|X,\sigma^2) = N(\theta_P,\sigma_P^2)$$

$$\int_\Theta P(X_*|\theta,\sigma^2)P(\theta|X,\sigma^2)d\theta = \int_{\mathbb{R}} \frac{1}{\sqrt{s\pi\sigma^2}} e^{1\frac{1}{2\sigma^2}(x_*-\theta)^2} \frac{1}{\sqrt{s\pi\sigma_P^2}} e^{1\frac{1}{2\sigma_P^2}(\theta-\theta_P)^2} d\theta$$

$$\alpha \int_{\mathbb{R}} e^{\frac{-x_*^2}{2\sigma^2}} e^{\frac{x_*\theta}{\sigma^2}} e^{\frac{-\theta^2}{2\sigma^2}} e^{\frac{-\theta^2}{2\sigma_P^2}} e^{\frac{\theta\theta_P}{\sigma_P^2}} e^{\frac{-\theta_P^2}{2\sigma_P^2}} d\theta$$

$$\alpha e^{\frac{X_*^2}{2\sigma^2}} \int_{\mathbb{R}} e^{a\theta - b\theta^2}$$

Where,

$$a = \frac{X_*}{\sigma^2} + \frac{\theta_p}{\sigma_p^2}$$

$$b = \frac{1}{2}(\frac{1}{\sigma^2} + \frac{1}{\sigma_p^2})$$

Everything in the integral above is proportional to the normal distribution. Say we have the following

$$P(\theta) = N(\frac{a}{2b}, \frac{1}{2b}) = \frac{1}{\sqrt{2\pi(\frac{1}{2b})}} e^{-\frac{1}{2(\frac{1}{2b})}(\theta - \frac{a}{2b})^2} = \sqrt{\frac{2}{b}} e^{-b(\theta^2 - \frac{\theta a}{b} + \frac{a^2}{4b^2})} = \sqrt{\frac{b}{\pi}} e^{-b\theta^2 + \theta a - \frac{a^2}{4b}}$$

$$= \sqrt{\frac{b}{\pi}} e^{\frac{-a^2}{4b}} e^{a\theta - b\theta^2}$$

All we do now is decompose the pdf into the normalization constant and the kernel.

$$e^{\frac{-X_*^2}{2\sigma^2}} \frac{1}{C} \int_{\mathbb{R}} C e^{a\theta - b\theta^2} = e^{\frac{-X_*^2}{2\sigma^2}} \sqrt{\frac{\pi}{b}} e^{\frac{a^2}{4b}}$$

We let $A = 2(\frac{1}{\sigma} + \frac{1}{\sigma_p^2})$.

$$\alpha e^{\frac{-X_*^2}{2\sigma^2}} e^{\frac{X_*^2}{2A\sigma^4}} e^{\frac{X_* \theta p}{A\sigma^2\sigma_p^2}} e^{\frac{\theta_p^2}{2A\sigma_p^2}} \alpha e^{\frac{\theta p}{A\sigma^2\sigma_p^2} x_* - (\frac{1}{2\sigma^2} - \frac{1}{2A\sigma^4}) X_*^2}$$

$$A\sigma^2\sigma_p^2 = \sigma^2\sigma_p^2(\frac{1}{\sigma^2} + \frac{1}{\sigma_p^2}) = \sigma_p^2 + \sigma^2$$

$$A\sigma^4 = \sigma^4(\frac{1}{\sigma_p^2} + \frac{1}{\sigma^2}) = \sigma^2 + \frac{\sigma^4}{\sigma 2_p}$$

This is all proportional to the following,

$$\alpha N(\frac{n}{2v}, \frac{n}{2v}) = N(\frac{\frac{\theta_p}{\sigma_p^2 + \sigma^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma^2 + \frac{\sigma^4}{\sigma_p^2}}})$$

$$2v = \frac{1}{\sigma^2} - \frac{1}{\sigma^2 + \frac{\sigma^4}{\sigma_p^2}} = \frac{\sigma_p^4 + \sigma^2\sigma_p^2 + \sigma^4}{\sigma^2\sigma_p^4 + \sigma^4\sigma_p^2}$$

$$\frac{\sigma^2}{\sigma^2(\sigma_p^2 + \sigma^2)} \to \frac{1}{2v} = \sigma_p^2 + \sigma^2$$