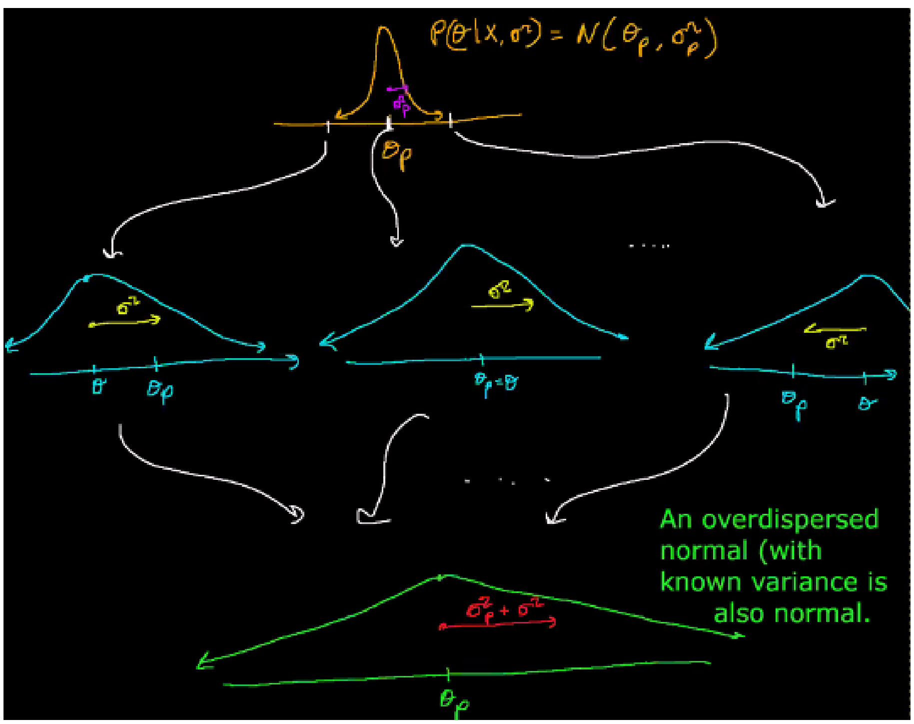
Lecture 17

Amir ElTabakh

Point Estimation in the normal-inverse gamma model

Recall the posterior of the normal distribution is defined as follows,

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



So we have found everything, from Bayesian point estimates, to kernels, to posterior predictive distributions for the normal-normal model. Now we consider the iid normal model with the θ known with σ^2 unknown and θ known i.e.,

$$\mathcal{F}: \ iid \ N(\theta, \sigma^2)$$

$$P(X|\theta,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(x_i-\theta)^2}$$

$$= (2\pi)^{\frac{-n}{2}} (\sigma^2)^{\frac{-n}{2}} e^{\frac{-1}{2\sigma^2} \sum (x_i - \theta)^2}$$

Consider the Laplace prior of indifference with respect to σ^2. A distribution on σ^2 which has support $(0, \infty)$. This prior would be $P(\sigma^2|\theta) \propto 1$.

$$P(\sigma^2|X,\theta) \propto P(X|\theta,\sigma^2)P(\sigma^2|\theta)$$

$$\propto P(X|\theta,\sigma^2$$

$$\propto (\sigma^2)^{\frac{-n}{2}} e^{\frac{\frac{\sum(x_i-\theta)^2}{2}}{\sigma^2}}$$

Let's take a break and find the MLE for σ^2.

$$l(\sigma^2; X, \theta) = \frac{-n}{2} ln(2\pi) - \frac{n}{2} ln(\sigma^2) - \frac{1}{2\sigma^2} \sum (x_i - \theta)^2$$

$$l'(\sigma^2; X, \theta) = \frac{-n}{2\sigma^2} + \frac{\sum(x_i - \theta)^2}{2(\sigma^2)^2} = 0 \; Setequalto0$$

$$l'(\sigma^2; X, \theta) = \frac{-n}{2\sigma^2} + \frac{\sum(x_i - \theta)^2}{2(\sigma^2)^2}$$

$$= 0 \; (Set\ equal\ to\ 0)$$

$$\rightarrow \frac{\sum(x_i - \theta)^2}{\sigma^2}$$

$$= n$$

Therefore, the average squared deviation is,

$$\hat{\sigma^2}_{MLE} = \frac{\sum(x_i - \theta)^2}{n}$$

Let's explore the kernel of the posterior using probability theory.

$$K(y) = y^{-a} e^{\frac{-b}{y}}$$

where $y \in (0, \infty)$. Lets try to find the actual density by finding the norm. constant c.

$$\frac{1}{c} = \int_0^\infty K(y)dy$$

$$= \int_0^\infty y^{-a} e^{\frac{-b}{y}} dy$$

Let z equal the following,

$$z = \frac{1}{y}$$

$$y = \frac{1}{z}$$

$$\frac{dy}{dz} = -z^{-2}$$
$$dy = -z^{-2}$$
$$y = 0 \to z = \infty$$
$$y = \infty \to z = 0$$

Now that we have done the u-substitution, lets return to the integral.

$$\int_0^\infty y^{-a} e^{\frac{-b}{y}} dy = \int_\infty^0 z^a e^{-bz}(-z^{-2})$$
$$= \int_0^\infty z^{(a-1)-1} e^{-bz} dz$$

Apply the u-substitution,

$$\frac{\gamma(a-1)}{b^{a-1}} \to P(y)$$

$$P(y) = \frac{b^{a-1}}{\gamma(a-1)} y^{-a} e^{\frac{-b}{y}}$$
$$\to \frac{\beta^\alpha}{\gamma(\alpha)} y^{-\alpha-1} e^{\frac{-\beta}{y}}$$
$$= InvGamma(\alpha, \beta)$$

This is called the "inverse gamma" distribution. Note,

$$W \sim Gamma(\alpha, \beta) \longleftrightarrow \frac{1}{W} \sim InvGamma(\alpha, \beta)$$

$$Y \sim InvGamma(\alpha, \beta)$$

$$E[Y] = \frac{\beta}{\alpha - 1} for \alpha > 1$$
$$Med[Y] = qinvgamma(0.5, \alpha, \beta)$$
$$Mode[Y] = \frac{\beta}{\alpha + 1} \; \forall \; \alpha, \beta > 0$$

Back to the regularly scheduled program...

$$P(\sigma^2 | X, \theta) \; \alpha \; (\sigma^2)^{\frac{-n}{2}} e^{\frac{\frac{-n\hat{\sigma^2}}{2}MLE}{\sigma^2}}$$

$$= (\sigma^2)^{\frac{-n-2}{2}-1} e^{\frac{\frac{-n\hat{\sigma^2}}{2}MLE}{\sigma^2}}$$

$$\alpha \; InvGamma(\frac{n-2}{2}, \frac{-n\hat{\sigma^2}_{MLE}}{2})$$

That's the posterior under Laplace's prior. Let's get the conjugate model now:

$$P(\sigma^2|X,\theta) \propto P(X|\theta,\sigma^2)P(\sigma^2|\theta)$$

$$\propto (\sigma^2)^{\frac{-n-2}{2}-1} e^{\frac{\frac{-n\hat{\sigma^2}_{MLE}}{2}}{\sigma^2}} P(\sigma^2|\theta)$$

What form should the prior be so that it's kernel has the same form as the posterior's kernel? It's an inverse gamma. Consider the following,

$$P(\sigma^2|\theta) = InvGamma(\alpha,\beta)$$

$$= (\sigma^2)^{\frac{-n-2}{2}-1} e^{\frac{\frac{-n\hat{\sigma^2}_{MLE}}{2}}{\sigma^2}} \frac{\beta^\alpha}{\gamma(\alpha)} y^{-\alpha-1} e^{\frac{-\beta}{y}}$$

$$\propto (\sigma^2)^{\frac{-n}{2}-\alpha-1} e^{-\frac{\frac{n\hat{\sigma^2}_{MLE}}{2}+\beta}{\sigma^2}}$$

$$\propto InvGamma(\frac{n}{2}+\alpha, \frac{n\hat{\sigma^2}_{MLE}}{2}+\beta)$$

Traditionally, however, we use a different parameterization of the prior. Let the following be true,

$$\alpha = \frac{n_0}{2}$$

$$\beta = \frac{n_0\sigma_0^2}{2}$$

$$P(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2})$$

$$P(\sigma^2|X,\theta) = InvGamma(\frac{n+n_0}{2}, \frac{n\hat{\sigma^2}_{MLE}+n_0\sigma_0^2}{2})$$

Last but not least, let us define the Bayesian point estimates for σ^2.

$$\hat{\theta}_{MMSE} = E[\sigma^2|X,\theta] = \frac{\frac{n\hat{\sigma}^2_{MLE}+n_0\sigma_0^2}{2}}{\frac{n+n_0}{2}-1} = \frac{n\sigma^2_{MLE}+n_0\sigma_0^2}{n+n_0-2} \; if \; n+n_0 > 2$$

$$\hat{\theta}_{MMAE} = Med[\sigma^2|X,\theta] = qinvgamma(0.5, \; \frac{n+n_0}{2}, \frac{n\sigma^2_{MLE}+n_0\sigma_0^2}{2})$$

$$\hat{\theta}_{MAP} = Mode[\sigma^2|X,\theta] = \frac{n\sigma^2_{MLE}+n_0\sigma_0^2}{n+n_0-2}$$

To find credible regions just use the appropriate *qinvgamma*. To hypothesis test use the appropriate *pinvgamma*. With regards to Pseudo-observation interpretation, n_0 is the number of pseudo-observations. Imagine $y_1, y_2, ..., y_{n_0}$, σ_0 is the guess of the value σ^2.

$$\frac{n\hat{\sigma}^2_{MLE}+n_0\sigma_0^2}{2} = \frac{\sum_{i=1}^{n}(x_i-\theta)^2 + \sum_{i=1}^{n_0}(y_i-\theta)^2}{2}$$

$$\rightarrow \sigma_0^2 = \frac{1}{n_0} \sum (y_i - \theta)^2$$

Recall in Haldane's objective prior of absolute ignorance, $n_0 = 0$. In term of the inverse gamma distribution, note the following. Also note that σ_0^2 can be anything, however by convention we let $\sigma_0^2 = 0$.

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

$$= \frac{0^0}{\gamma(0)} (\sigma^2)^{-0-1} e^{-\frac{0}{\sigma^2}}$$

$$= \frac{1}{\sigma^2}$$

$$\rightarrow P(\sigma^2|X,\theta) = InvGamma(\frac{n}{2}, \frac{n\hat{\sigma}^2_{MLE}}{2})$$

Laplace's prior of indifference is proportional to 1 across the distribution, that is, $P(\sigma^2|\theta)\ \alpha\ 1$. Haldane's prior makes much more sense to use in terms of the inverse gamma distribution than Laplace's prior because Laplace's prior implies an infinite variance. For instance, σ^2 has the same weight in $[0,1]$ and in $[1,000,000,000,\ 1,000,000,001]$. No one uses the Laplace prior for practical purposes given the inverse gamma distribution in the real world for this reason. What does this Laplace prior correspond to? Recall it results in a posterior of,

$$P(\sigma^2|X,\theta) = InvGamma(\frac{n-2}{2}, \frac{n\hat{\sigma}^2_{MLE}}{2})$$

$$\rightarrow n_0 = -2,\ \sigma_0^2 = 0$$

$$P(\sigma^2|\theta) = InvGamma(\frac{-2}{2}, \frac{0}{2})$$

$$= InvGamma(-1,0)$$

Its odd that $n_0 = -2$. How is it that the number of pseudo-observations is a negative number? It means that we have to delete two data points, which makes little sense to do.