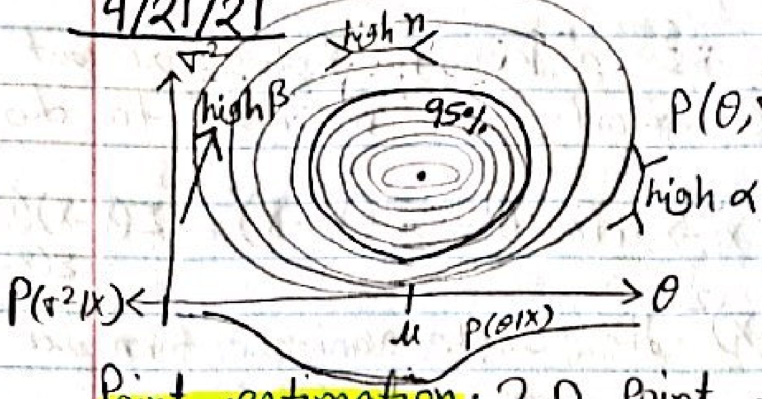
4/21/21 Lec 19



$P(\theta, \sigma^2|X) = NormInvG(\mu, \lambda, \alpha, \beta)$

Point estimation: 2-D point estimator.

$\begin{bmatrix} \hat{\theta} \\ \sigma^2 \end{bmatrix}_{MAP}$ is the higest point on the mountain.

Credible region: Some 2D area... but hard to define so we skip it!

High Density Region: You can visualize this on the graph.

Hypothesis testing: $H_0: \theta \in \Theta_A$ and $\sigma^2 \in \Theta_B$

$$P\text{-val} = P(\quad |X) = \iint_{\Theta_A \, \Theta_B} P(\theta, \sigma^2|X)\, d\theta\, d\sigma^2$$

this is rarely done... So we skip it!

$$\overbrace{P(\theta, \sigma^2|X)}^{\text{NormInv Gamma}} \propto P(X|\theta, \sigma^2)\overbrace{P(\theta, \sigma^2)}^{\text{NormInvGamma}}$$

The NormInvGamma is conjugate for the normal model with both mean and varience unknown. Now we usually specified hyperparameters for the prior and derived the general posterior which will have parameters that combine the prior hyperparameters with the data. We will skip this too! Instead, we will only consider Jeffrey's Prior and we won't even derive it.

$$P_J(\theta, \sigma^2) \propto (\sigma^2)^{-1} = \underbrace{P_J(\theta|\sigma^2)}_{\propto 1} \underbrace{P_J(\sigma^2)}_{\propto (\sigma^2)^{-1}}$$

Let's derive the posterior for only the Jeffrey's Prior:

$$P(\theta,\sigma^2|X) \propto P(X|\theta,\sigma^2)P_{J_1}(\theta,\sigma^2)$$

$$\propto \left((\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\Sigma(X_i-\theta)^2}\right)(\sigma^2)^{-1}$$

from Lec 18 $\underset{\downarrow}{=}$ $e^{-\frac{1}{2\sigma^2/n}(\bar{X}-\theta)^2}(\sigma^2)^{-\frac{n}{2}-1} e^{\frac{-(n-1)s^2/2}{\sigma^2}}$

$$\propto \text{NormInvGamma}\left(\mu=\bar{X}, \lambda=n, \alpha=\frac{n}{2}, \beta=\frac{(n-1)s^2}{2}\right)$$

This concludes the unit of 2-D Inference (for both θ and σ^2). Yes, we didn't do much.

Now we transition to 1-D inference for either θ or σ^2. Let's say we want inference for θ. How do we do this given a 2-D posterior? This is most common situation.

- You care about inference for the mean and you don't care about the variance (it's a nuissance).

Let's average over σ^2:

$$\rightarrow P(\theta|X) = \int_0^\infty P(\theta,\sigma^2|X)\,d\sigma^2 \qquad g(x)=\int_{\mathbb{R}} f(x,y)\,dy$$

marginal posterior of θ

$$\propto \int_0^\infty (\sigma^2)^{-\frac{n}{2}-1} e^{-\frac{1}{2\sigma^2/n}(\bar{X}-\theta)^2} e^{-\frac{(n-1)s^2/2}{\sigma^2}}\,d\sigma^2$$

$$= \int_0^\infty (\sigma^2)^{-\overset{=\alpha}{\frac{n}{2}}-1} e^{-\frac{\overbrace{n(\bar{X}-\theta)^2/2+(n-1)s^2/2}^{=\beta}}{\sigma^2}}\,d\sigma^2$$

$$= \int_0^\infty \underbrace{(\sigma^2)^{-\alpha-1} e^{-\frac{\beta}{\sigma^2}}}_{\text{kernel inverse gamma}}\,d\sigma^2$$

$$= \frac{\Gamma(\alpha)}{\beta^\alpha} \underbrace{\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} e^{-\frac{\beta}{\sigma^2}} \, d\sigma^2}_{=1}$$

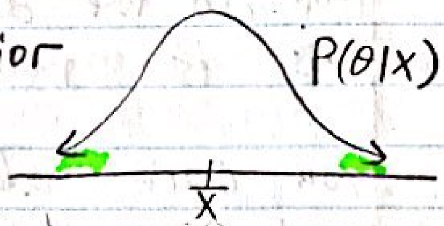
$$= \Gamma(\alpha)\beta^{-\alpha} = \Gamma\left(\frac{n}{2}\right) \left(\frac{n(\bar{x}-\theta)^2+(n-1)S^2}{2}\right)^{-n/2}$$

$$\propto \left(\frac{n(\bar{x}-\theta)^2}{2} + \frac{(n-1)S^2}{2}\right)^{-n/2} \left(\frac{2}{(n-1)S^2}\right)^{-n/2}$$

$$= \left(1+\frac{1}{\underbrace{n-1}_{\nu}} \frac{(\bar{x}-\theta)^2}{\underbrace{\frac{S^2}{n}}_{\text{scale}}}\right)^{-\frac{\overbrace{n-1}^{\nu}+1}{2}}$$

$\propto T_{n-1}\left(\bar{x}, \frac{S^2}{n}\right)$ Shifted and Scaled Student's T distribution.

$\overset{n-1>20}{\approx} N\left(\bar{x}, \frac{S^2}{n}\right)$ the posterior



$$\hat{\theta}_{MMSE} = \hat{\theta}_{MMAE} = \hat{\theta}_{MAP} = \bar{X}$$

$$CR_{\theta|1-\alpha_0} = \left[qt.scaled\left(\frac{\alpha_0}{2}, \bar{x}, \frac{S}{\sqrt{n}}\right), qt.scaled\left(1-\frac{\alpha_0}{2}, \bar{x}, \frac{S}{\sqrt{n}}\right)\right]$$

$$H_0: \theta \leq \theta_0 \Rightarrow P_{val} = P(\theta \leq \theta_0 | X) = Pt.scaled\left(\theta_0, \bar{x}, \frac{S}{\sqrt{n}}\right)$$

What if we wanted inference for σ^2 (varience) and we didn't care about the mean (nuisance)? We derive the other marginal distribution:

$$P(\sigma^2|X) = \int_{\mathbb{R}} P(\theta, \sigma^2 | X) \, d\theta$$

$$\propto \int_{\mathbb{R}} (\sigma^2)^{-\frac{n}{2}-1} e^{-\frac{1}{2\sigma^2/n}(\bar{x}-\theta)^2} e^{-\frac{(n-1)S^2/2}{\sigma^2}} \, d\theta$$

$$= (\sigma^2)^{-\frac{n}{2}-1} e^{-\frac{(n-1)S^2/2}{\sigma^2}} \int_{\mathbb{R}} \overbrace{e^{-\frac{1}{2\sigma^2/n}(\bar{x}-\theta)^2}}^{\text{kernel for normal} \propto N\left(\bar{x}, \frac{\sigma^2}{n}\right)} d\theta$$

$$= (\sigma^2)^{-\frac{n}{2}-1} e^{-\frac{(n-1)S^2/2}{\sigma^2}} \sqrt{2\pi\frac{\sigma^2}{n}} \overbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} e^{-\frac{1}{2\sigma^2/n}(\bar{x}-\theta)^2} \, d\theta}^{=1}$$

$$\propto (\sigma^2)^{-\frac{n}{2}-1} e^{-\frac{(n-1)S^2/2}{\sigma^2}} (\sigma^2)^{\frac{1}{2}}$$

$$= (\sigma^2)^{-\frac{n}{2}+\frac{n}{2}-1} e^{\frac{-(n-1)s^2/2}{\sigma^2}}$$

$$= (\sigma^2)^{-\frac{n-1}{2}-1} e^{-\frac{(n-1)s^2/2}{\sigma^2}}$$

$$\propto \text{InvGamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$

Formula comparisons under Jeffrey's Prior:

$$P(\theta|X,\sigma^2) = N\left(\bar{X}, \frac{\sigma^2}{n}\right)$$

$$P(\theta|X) = T_{n-1}\left(\bar{X}, \frac{s^2}{n}\right)$$

$$P(\sigma^2|X,\theta) = \text{InvGamma}\left(\frac{n}{2}, \frac{n\,\hat{\sigma}^2_{MLE}}{2}\right) \quad \hat{\sigma}^2_{MLE} = \frac{1}{n}\Sigma(X_i-\theta)^2 \approx \sigma^2$$

$$P(\sigma^2|X) = \text{InvGamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right) \quad s^2 = \frac{1}{n-1}\Sigma(X_i-\bar{X})^2 \overset{\text{worst}}{\approx} \sigma^2$$

Posterior Predictive distribution:

$$P(X_*|X) = \int_0^\infty \int_{\mathbb{R}} P(X_*|\theta,\sigma^2)\, P(\theta,\sigma^2|X)\, d\theta\, d\sigma^2$$