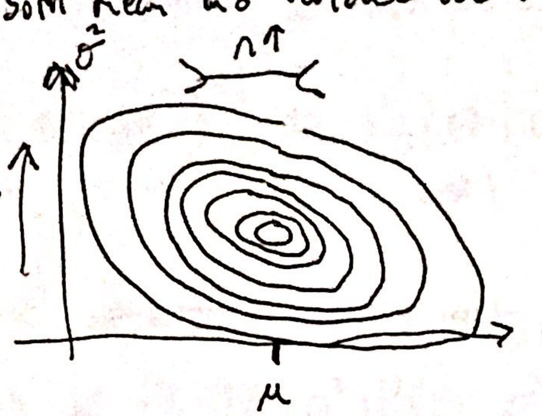
LEC 19 4/21 

Normal but both mean and variance are unknown

(Picture)




high n squeezes dist/picture together

high α↑ squeezes this way

$P(\theta, \sigma^2 | x) = \text{Norm Inv Gamma}(\underbrace{\mu, \lambda}_{\theta \text{ controlling dis}}, \overbrace{\alpha, \beta}^{\text{controlling } \sigma^2})$

$\frac{\sigma^2}{n}$ ↙ as $n \uparrow$ $\sigma^2$ gets smaller everything gets

Point Estimation: 2-d point estimate

$\begin{bmatrix} \theta \\ \sigma^2 \end{bmatrix}_{MAP}$ = is the highest point on the mountain. $= \hat{\begin{bmatrix} \mu \\ \square \end{bmatrix}}$ ??

CREDIBLE REGION: Some 2d area ... hard to define (skip it)

High Density Region: you can visualize this on the picture (oval shaped region)

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

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

this is rarely done so skip it.

Conjugate Model: $\underbrace{P(\theta, \sigma^2 | X)}_{\text{Norm Inv Gamma}} \propto P(X | \theta, \sigma^2) \underbrace{P(\theta, \sigma^2)}_{\text{Norm Inv Gamma}}$ ⟸

The normal inverse gamma model is conjugate for the normal ~~model~~ model with both mean and variance unknown. Now we usually specify hyper parameters for prior and derive general posterior with parameters that combine hyper and data

We will skip this also. Instead we will only consider the Jeffrey's prior

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

Lets derive posterior using only Jeffrey's Prior.

$$P(\theta, \sigma^2 \mid x) \propto P(x \mid \theta, \sigma^2) P_J(\theta, \sigma^2)$$

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

From Lec 18

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

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

this is the only posterior we will need for this.

From Lec 18 $(\sigma^2)^{-\frac{n}{2}} \longrightarrow$ go back to video to correct. (20-25 min)

This concludes unit on 2D inference for both $\theta$ and $\sigma^2$.

Now we transition to 1D inference for either $\theta$ or $\sigma^2$.

Say we want inference for $\theta$, how do we do this given a 2D posterior? This is the most common situation. You care about the mean and don't care about variance. Just avg over $\sigma^2$.

$$\int_0^\infty P(\theta, \sigma^2 \mid x)\, d\sigma^2 = P(\theta \mid x) \leftarrow \text{called a marginal posterior of theta}$$

this solves all our problems.

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

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

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

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

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

adding this in, its ok because s is a function of x which is given.

$$= \left(1 + \underset{\nu}{\frac{1}{(n-1)}} \cdot \underset{\text{scale}}{\frac{(\bar{x}-\theta)^2}{\frac{s^2}{n}}}\right)^{-n/2}$$

write as $-\frac{\overset{\nu}{(n-1)}+1}{2}$

$$\underbrace{\propto T_{n-1}\left(\bar{x}, \frac{s^2}{n}\right) \quad \text{Shifted and scaled students T distribution.}}_{\text{This is our posterior!}}$$

$n-1 > 20$

$$\sim N\left(\bar{x}, \frac{s^2}{n}\right)$$

$\hat{\theta}_{MMSE} = \hat{\theta}_{MMAE} = \hat{\theta}_{MAP} = \bar{x}$ ← same as if $\sigma^2$ was known with Jeff prior.

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

$$H_0: \theta \le \theta_0 \Rightarrow P_{val} = P(\theta \le \theta_0 | X) = pt.scaled\left(\theta_0, \bar{x}, \sqrt{\frac{s^2}{n}}\right)$$

Done with this. What if we wanted inference on $\sigma^2$ and we did not care about the mean? We derive the other marginal distribution.

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

$$= (\sigma^2)^{-\frac{n}{2}-1} e^{\frac{-(n-1)s^2/2}{\sigma^2}} \int_{\mathbb{R}} \underbrace{e^{-\frac{1}{2\sigma^2/n}(\bar{x}-\theta)^2}}_{\text{kernel for normal distribution}} d\theta$$

same trick

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

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

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

Formula comparisons under Jeffrey's Priors.

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

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

$$P(\sigma^2 \mid X, \theta) = \text{InvGamma}\left(\frac{n}{2}, \frac{n\hat{\sigma}^2_{MLE}}{2}\right)$$

$$P(\sigma^2 \mid X) = \text{InvGamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$

---

Posterior predictive Distribution

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