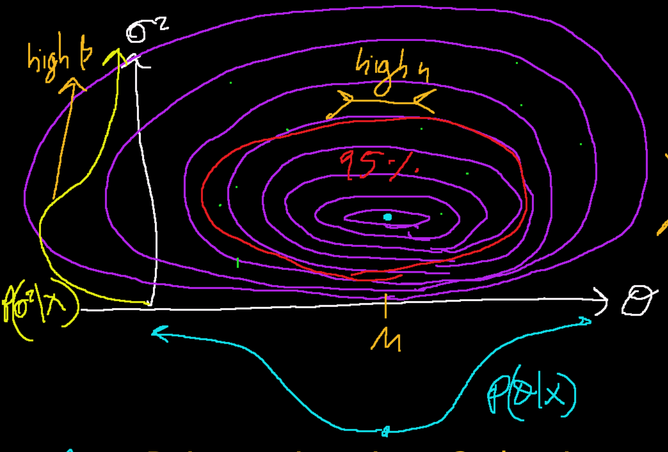


$$p(\theta,\sigma^2|x) = \text{NormInvG}(\mu,\lambda,\alpha,\beta)$$

Point estimation: 2-d point estimate.

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

Credible Region: some 2d area... but hard to define (we skip it)
High density region: you can visualize this

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

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

this is rarely done... so we skip it

$$\underbrace{p(\theta,\sigma^2|x)}_{\text{NormInvGamma}} \propto p(X|\theta,\sigma^2)\,\underbrace{p(\theta,\sigma^2)}_{\text{NormInvGamma}}$$

The NormalInverseGamma is conjugate for the normal model with both mean and variance 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 the Jeffrey's prior and we won't even derive it.

$$p_J(\theta,\sigma^2) \propto (\sigma^2)^{-1} = \overset{\propto 1}{p_J(\theta|\sigma^2)}\;\overset{\propto(\sigma^2)^{-1}}{p_J(\sigma^2)}$$

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

$$p(\theta,\sigma^2|x) \propto p(X|\theta,\sigma^2)\,p_J(\theta,\sigma^2)$$

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

from lec 18

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

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

This concludes the unit on 2-dim inference (for both theta and sigsq). Yes, we didn't do that much.

Now we transition to 1-dim inference for either theta or sigsq. Let's say we want inference for theta. How do we do this given a 2-dim posterior? This is the most common situation. You care about inference for the mean and you don't care about the variance (it's a nuisance). Why don't we... average over sigsq i.e.

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

marginal posterior of theta

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

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

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

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

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

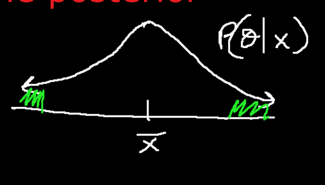
$$\propto \left(\frac{n(\bar{x}-\theta)^2}{2}+\frac{(n-1)s^2}{2}\right)^{-n/2}\left(\frac{2}{(n-1)s^2}\right)^{-n/2}$$

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

$$\propto T_{n-1}\left(\bar{x},\tfrac{s^2}{n}\right)$$ shifted and scaled Student's T distr.

$$\overset{n-1>20}{\approx} N\left(\bar{x},\tfrac{s^2}{n}\right)$$

the posterior



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

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

$$H_0: \theta\le\theta_0 \Rightarrow p_{val} = P(\theta\le\theta_0|x) = pt.scaled\left(\theta_0,\bar{x},\tfrac{s}{\sqrt{n}}\right)$$

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

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

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

Kernel for normal $\propto N\left(\bar{x},\frac{\sigma^2}{n}\right)$

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

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

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

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

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

Formula comparisons under Jeffrey's prior:

$$p(\theta|X,\sigma^2) = N\left(\bar{x},\tfrac{\sigma^2}{n}\right)$$

$$p(\theta|X) = T_{n-1}\left(\bar{x},\tfrac{s^2}{n}\right) \qquad /\!/ \ \tfrac{1}{n}\textstyle\sum(x_i-\theta)^2 \approx \sigma^2$$

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

worse

$$p(\sigma^2|X) = \text{InvGamma}\left(\frac{n-1}{2},\frac{(n-1)s^2}{2}\right) \qquad = \tfrac{1}{n-1}\textstyle\sum(x_i-\bar{x})^2 \approx \sigma^2$$

Posterior predictive distribution

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