Lec 16 4/18/18 Bolt 391 $\quad X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$

θ known, σ^2 unknown. $\Rightarrow \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2 = \frac{SSE}{n}$. Bayesian inference:

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

$$= \left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X_i - \theta)^2}\right) P(\sigma^2 \mid \theta)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\overbrace{\sum (X_i - \theta)^2}^{n\hat{\sigma}^2_{MLE}}} P(\sigma^2 \mid \theta)$$

$$\propto \underbrace{(\sigma^2)^{-\frac{n}{2}} e^{-\frac{n\hat{\sigma}^2}{2\sigma^2}}} P(\sigma^2 \mid \theta)$$

Kernel of InvGamma $\left(\frac{n}{2} - 1, \frac{n\hat{\sigma}^2}{2}\right)$.

$Y \sim \text{Gamma}(\alpha, \beta) := \frac{\beta^\alpha}{\Gamma\alpha} e^{-\beta y} y^{\alpha - 1}$

Recall: $Y \sim \text{InvGamma}(\alpha, \beta) := \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\frac{\beta}{y}} y^{-\alpha - 1}$

match kernels...

$\Rightarrow \sigma^2 \mid \theta \sim \text{InvGamma}(\alpha, \beta)$

$$P(\sigma^2 \mid X, \theta) \propto \left((\sigma^2)^{-\frac{n}{2}} e^{-\frac{n\hat{\sigma}^2}{2\sigma^2}}\right)\left(\frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\frac{\beta}{\sigma^2}} (\sigma^2)^{-\alpha - 1}\right)$$

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

$$\propto \text{InvGamma}\left(\frac{n}{2} + \alpha, \frac{n\hat{\sigma}^2}{2} + \beta\right)$$

However... we ~~usually don't use~~ the canonical form of this gamma doesn't use α, β. We use a parametrization that mirrors the pseudocount interpretation

if $\alpha = \frac{n_0}{2}$, $\beta = \frac{n_0 \sigma_0^2}{2} \Rightarrow \sigma^2 \mid \theta \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right)$

$\Rightarrow P(\sigma^2 \mid X, \theta) = \text{InvGamma}\left(\frac{n + n_0}{2}, \frac{n\hat{\sigma}^2 + n_0\sigma_0^2}{2}\right)$

Interpretation

n_0: # prior trials of ... $Y_1, \ldots, Y_{n_0} \stackrel{iid}{\sim} N(\theta, \sigma^2)$

σ_0^2: previous $\hat{\sigma}^2_{MLE}$ for σ^2, $\sigma_0^2 := \frac{1}{n_0} \sum_{i=1}^{n_0} (Y_i - \theta)^2$

$$\Rightarrow n_0 \sigma_0^2 = \sum_{i=1}^{n_0} (Y_i - \theta)^2 = SSE_0$$

$$\Rightarrow \sigma^2 | X, \theta \sim \text{InvGamma}\Big(\underbrace{\frac{n + n_0}{2}}_{\alpha'}, \underbrace{\frac{SSE + SSE_0}{2}}_{\beta'}\Big)$$

$$\hat{\sigma}^2_{MMSE} = \frac{\frac{n\hat{\sigma}^2_{MLE} + n_0\sigma_0^2}{2}}{\frac{n+n_0}{2} - 1} = \frac{n\hat{\sigma}^2_{MLE} + n_0\sigma_0^2}{n + n_0 - 2}$$

Shrinkage ... easy ... HW ...

$$\hat{\sigma}^2_{MAP} = \frac{n\hat{\sigma}^2_{MLE} + n_0\sigma_0^2}{n + n_0 + 2}$$

$$\hat{\sigma}^2_{MAE} = \text{qinvgamma}\left(0.5, \frac{n+n_0}{2}, \frac{n\hat{\sigma}^2 + n_0\sigma_0^2}{2}\right)$$

CR's, hypothesis tests follow ...

Uninformative prior? If $n_0 = 0$ .. as if seeing "nothing"

$\sigma^2 | \theta \sim \text{InvGamma}(0, 0)$ Improper ...

But ...

$$P(\sigma^2 | X, \theta) = \text{InvGamma}\left(\frac{n}{2}, \frac{n\hat{\sigma}^2}{2}\right)$$ always proper ...

$$\hat{\sigma}^2_{MMSE} = \frac{\frac{n\hat{\sigma}^2}{2}}{\frac{n}{2}-1} = \frac{n\hat{\sigma}^2}{n-2} = \frac{1}{n-2}\sum(X_i-\theta)^2 \approx \hat{\sigma}^2_{MLE}$$

What is another uniform prior?

$$\sigma^2|\theta \sim \text{InvGamma}(-2, 0) \qquad n_0 = 2,\ \sigma_0^2 = 0, \text{ weird...}$$

$$\Rightarrow \sigma^2|X_1,\ldots,X_n,\theta \sim \text{InvGamma}\left(\frac{n+2}{2}, \frac{n\hat{\sigma}^2}{2}\right)$$

$$\hat{\sigma}^2_{MMSE} = \frac{\frac{n\hat{\sigma}^2}{2}}{\frac{n+2}{2}-1} = \hat{\sigma}^2_{MLE}$$

How about Jeffrey's prior?

$$P(\sigma^2|\theta) \propto \sqrt{I(\sigma^2)}$$

$$l'(\sigma^2; X, \theta) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}SSE = -\frac{n}{2}(\sigma^2)^{-1} + \frac{SSE}{2}(\sigma^2)^{-2}$$

$$l''(\sigma^2; X, \theta) = \frac{n}{2}(\sigma^2)^{-2} - SSE(\sigma^2)^{-3}$$

$$I(\sigma^2) = E[-l''(\sigma^2; X, \theta)] = E\left[-\frac{n}{2}(\sigma^2)^{-2} + SSE(\sigma^2)^{-3}\right] = -\frac{n}{2}(\sigma^2)^{-2} + (\sigma^2)^{-3}E[SSE]$$

$$E[SSE] = E\left[\sum_{i=1}^n (X_i-\theta)^2\right] = \sum_{i=1}^n E[(X_i-\theta)^2] \overset{\text{due to iid}}{=} n\,E[(X-\theta)^2] \overset{\text{def of variance!}}{=} n\,Var[X] = n\sigma^2$$

Recall: $X \sim N(\theta, \sigma^2)$

$$\Rightarrow I(\sigma^2) = -\frac{n}{2}(\sigma^2)^{-2} + (\sigma^2)^{-3}(n\sigma^2) = -\frac{n}{2}(\sigma^2)^{-2} + n(\sigma^2)^{-2} = (\underbrace{n-\tfrac{n}{2}}_{n/2})(\sigma^2)^{-2}$$

$\Rightarrow$ Jeffreys $P(\sigma^2|\theta) \propto \sqrt{\frac{n}{2}(\sigma^2)^{-2}} \propto (\sigma^2)^{-1} = \text{InvGamma}(0,0)$

$\Rightarrow n_0 = 0$

Proper? $\int_0^\infty \frac{1}{\sigma^2}\, d\sigma^2 = \infty$ ... NO!

post pred. distr. ... not covered

Finally ... the "more" realistic situation:

$X_1,\dots,X_n|\theta,\sigma^2 \overset{iid}{\sim} N(\theta,\sigma^2)$

but now both θ, σ^2 unknown ...

$$\underbrace{P(\theta,\sigma^2|X)}_{\text{2-dim posterior}} \propto \underbrace{P(X|\theta,\sigma^2)}_{\text{sac likelihood}} \underbrace{P(\theta,\sigma^2)}_{\text{2-dim prior}}$$

$$= \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X_i-\theta)^2}\right) P(\theta,\sigma^2)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n (\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^n (X_i-\theta)^2}{2\sigma^2}} P(\theta,\sigma^2)$$

$$\propto (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(X_i-\theta)^2} P(\theta,\sigma^2)$$

Inverse gamma? No... θ is a free variable!

This is a 2-dim distr... you haven't seen these before...

Consider

$$\sum_{i=1}^{n}(X_i-\theta)^2 = \sum_{i=1}^{n}\left((X_i-\bar{X})+(\bar{X}-\theta)\right)^2$$

$$= \sum_{i=1}^{n}(X_i-\bar{X})^2 + 2(X_i-\bar{X})(\bar{X}-\theta) + (\bar{X}-\theta)^2 = \sum(X_i-\bar{X})^2 + 2\sum(X_i\bar{X} - X_i\theta - \bar{X}^2 + \bar{X}\theta) + n(\bar{X}-\theta)^2$$

Note: $s^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ Variance estimator if both θ, σ^2 unknown

$$= (n-1)s^2 + 2\left(\bar{X}\overset{\bar{X}n}{\overset{\shortparallel}{\sum X_i}} - \theta\overset{\bar{X}n}{\overset{\shortparallel}{\sum X_i}} - \sum\bar{X}^2 + \theta\sum X_i\right) + n(\bar{X}-\theta)^2$$

$$= (n-1)s^2 + 2\left(\cancel{n\bar{X}^2} - \cancel{\theta\bar{X}n} - \cancel{n\bar{X}^2} + \cancel{\theta\bar{X}n}\right) + n(\bar{X}-\theta)^2$$

$$= (n-1)s^2 + n(\bar{X}-\theta)^2$$

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

$$P(\sigma^2,\theta|X) \propto (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\left((n-1)s^2 + n(\bar{X}-\theta)^2\right)} P(\theta,\sigma^2)$$

$$= \underbrace{(\sigma^2)^{-\frac{n}{2}} e^{-\frac{(n-1)s^2/2}{\sigma^2}} e^{-\frac{1}{2\frac{\sigma^2}{n}}(\bar{X}-\theta)^2}}_{\propto \text{ NormalInvGamma}\left(\mu=\bar{X},\ \lambda=n,\ \alpha=\frac{n}{2}+1,\ \beta=\frac{(n-1)s^2}{2}\right)} P(\theta,\sigma^2)$$

4 param distr!

$P(\theta,\sigma^2)$ should be Normal Inv Gamma as well

However... we will only consider the uniformative, Jeffreys prior

$$P(\theta,\sigma^2) = P(\theta|\sigma^2)\,P(\sigma^2) \propto (1)\left(\frac{1}{\sigma^2}\right) = \frac{1}{\sigma^2}$$