Lecture 13 3/28/17 Math 341

$$\Rightarrow X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$$

$$\theta \mid \sigma^2 \sim N(\mu_0, \tau^2)$$

$$\Rightarrow \theta \mid X_1, \ldots, X_n, \sigma^2 \sim N\left(\frac{\frac{\bar{x}n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \right)$$

$$\hat{\theta}_{MMSE} = \frac{\frac{\bar{x}n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

$\hat{\theta}_{MAE} =$ "

$\hat{\theta}_{MAP} =$ "

Why? Normal is symmetric & unimodal $\Rightarrow$ mean = median = mode

Shrinkage...

$$\hat{\theta}_{MMSE} = \frac{\frac{\bar{x}n}{\sigma^2}}{\left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)} \cdot \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n}} + \frac{\frac{\mu_0}{\tau^2}}{\left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)} \frac{\tau^2}{\tau^2}$$

$$= \frac{1}{1 + \frac{\sigma^2}{n\tau^2}} \bar{x} + \frac{1}{\frac{n\tau^2}{\sigma^2} + 1} \mu_0$$

$$= \underbrace{\frac{\sigma^2}{n\tau^2 + \sigma^2}}_{\rho} \underbrace{\mu_0}_{E(\theta)} + \underbrace{\frac{n\tau^2}{n\tau^2 + \sigma^2}}_{1-\rho} \underbrace{\bar{x}}_{\hat{\theta}_{MLE}}$$

$$= \rho \, E(\theta) + (1 - \rho) \, \hat{\theta}_{MLE}$$

weighted arithmetic avg shrinkage

Hyperparameter interpretation

μ_0 : prior mean

τ^2 : prior variance

do they represent pseudodata?

Imagine pseudodata $Y_1, \ldots, Y_{n_0}$ you "saw" before

let $\mu_0 = \bar{Y} = \frac{1}{n_0}\sum_{k=1}^{n_0} Y_k$

Remember σ^2 is known ...

let $\tau^2 = \frac{\sigma^2}{n_0} \Rightarrow n_0 = \frac{\sigma^2}{\tau^2}$. So μ_0, τ^2 can be thought of as $\bar{Y}$ and n_0 pseudodata

$$\hat{\theta}_{MMSE} = \frac{\frac{\bar{X}n}{\sigma^2} + \frac{\bar{Y}n_0}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{n_0}{\sigma^2}} \cdot \frac{\sigma^2}{\sigma^2} = \frac{\bar{X}n + \bar{Y}n_0}{n + n_0} = \frac{\sum x_i + \sum y_i}{n + n_0}$$

$\Rightarrow$ the avg of all obs's ... prior and at hand

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

$\Rightarrow \quad \theta \mid \sigma^2 \sim N(\mu_0, \frac{\sigma^2}{n_0})$

$\theta \mid X_1, \ldots, X_n, \sigma^2 \sim N\left(\frac{\bar{X}n + \mu_0}{n + n_0}, \left(\frac{\sigma}{\sqrt{n+n_0}}\right)^2\right)$

What prior should be chosen? Laplace?

lets play the trick again ...

$P(\theta \mid \sigma^2) \propto 1$

Improper!!!

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

$$\propto \underbrace{e^{\frac{\bar{X}n}{\sigma^2}\theta}}_{\frac{c}{v^2}}\ \underbrace{e^{-\frac{n}{2\sigma^2}\theta^2}}_{\frac{1}{2v^2}} \propto N\left(\bar{X}, \frac{\sigma^2}{n}\right)$$

always proper!!!

$$\frac{1}{2v^2} = \frac{n}{2\sigma^2} \Rightarrow v^2 = \frac{\sigma^2}{n}$$

$$\frac{c}{v^2} = \frac{\bar{X}n}{\sigma^2} \Rightarrow c = \frac{\bar{X}n}{\sigma^2}v^2 = \frac{\bar{X}\cancel{n}}{\cancel{\sigma^2}}\cdot\frac{\cancel{\sigma^2}}{\cancel{n}}$$

Under Laplace $\hat{\theta}_{MMSE} = \hat{\theta}_{MAE} = \hat{\theta}_{MAP} = \hat{\theta}_{MLE} = \bar{X}$

the ultimate confluence!

Jeffreys prior...

$$\ell'(\theta; X, \sigma^2) = \frac{\bar{X}n}{\sigma^2} - \frac{n\theta}{\sigma^2}$$

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

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

Jeffreys...

$$P(\theta|\sigma^2) \propto \sqrt{I(\theta)} = \sqrt{\frac{n}{\sigma^2}} \propto 1$$ = Laplace prior!

Improper priors can be thought of as the limit of proper priors

$X|\theta \sim Bin(n, p)$

$\theta \sim Beta(\alpha, \beta)$

$\theta | X \sim Beta(x+\alpha, n-x+\beta)$

$$\lim_{\substack{\alpha \to 0 \\ \beta \to 0}} p(\theta | x) = Beta(x, n-x)$$

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

$\theta | \sigma^2 \sim N(\mu_0, \tau^2)$

$$\theta | X_1, \ldots, X_n, \sigma^2 \sim N\left(\underbrace{\frac{\frac{\bar{x} n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}}_{\hat{\theta}_{MMSE}}, \underbrace{\frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}}_{\sigma^2_p} \right) = N(\hat{\theta}_{MMSE}, \sigma^2_p)$$

What's this limit?

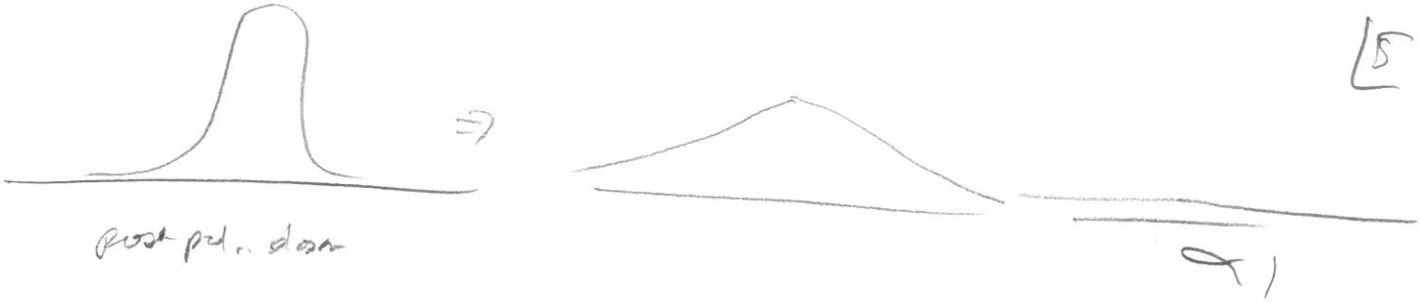
$$\lim_{\tau^2 \to \infty} p(\theta | X_1, \ldots, X_n, \sigma^2) = N\left(\bar{x}, \frac{\sigma^2}{n}\right)$$

let's see

$$\lim_{\tau^2 \to \infty} \frac{\frac{\bar{x} n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \cdot \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n}} = \lim \frac{\bar{x} + \frac{\mu_0 \sigma^2}{\tau^2 n}}{1 + \frac{\sigma^2}{n \tau^2}} = \bar{x} \checkmark$$

$$\lim_{\tau^2 \to \infty} \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} = \frac{1}{\frac{n}{\sigma^2}} = \frac{\sigma^2}{n} \checkmark$$

$$\lim_{\tau^2 \to \infty} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}(\theta - \mu_0)^2} = \underbrace{0 \propto 1}_{\text{weird!!!}} \Rightarrow p(\theta | \sigma^2) \propto 1$$



$$P(X^*|X,\sigma^2) = \int_\Theta \underbrace{P(X^*|\theta,\sigma^2)}_{N(\theta,\sigma^2)} \underbrace{P(\theta|X,\sigma^2)}_{N(\theta_p,\sigma_p^2)} d\theta$$

for $n^*=1$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x^*-\theta)^2} \frac{1}{\sqrt{2\pi\sigma_p^2}} e^{-\frac{1}{2\sigma_p^2}(\theta-\theta_p)^2} d\theta$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\sqrt{2\pi\sigma_p^2}} \int_{\mathbb{R}} e^{-\frac{1}{2\sigma^2}(x^*-\theta)^2 - \frac{1}{2\sigma_p^2}(\theta-\theta_p)^2} d\theta$$

difficult!
Easier way?

$X_1, X_2 \overset{iid}{\sim} U(1,2,3,4,5,6)$

$S = X_1 + X_2 \sim ?$

Not correct..

$P(S=1)=0$

$P(S=2) = P(X_1=1) \cdot P(X_2=1) = \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36}$

$P(S=3) = P(X_1=1)P(X_2=2) + P(X_1=2)P(X_1=1)$

$$= \sum_{x\in \text{supp}(X_1)} P(X_1=x)\, P(X_2=3-x)$$

$$\Rightarrow P(S=s) = \sum_{x \in \mathrm{Supp}(X_1)} P(X_1 = x)\, P(X_2 = s - x)$$

$$= \sum_{x \in \mathrm{Supp}(X_2)} P(X_2 = x)\, P(X_1 = s - x)$$

for iid, doesn't matter the order

For cont. r.v.'s,

$$S = X_1 + X_2 \sim \int_{\mathrm{Supp}[X_1]} f_{X_1}(x)\, f_{X_2}(s-x)\, dx =: f_{X_1} * f_{X_2}$$

convolution operation

Recall fact from Stat 201

$X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

thus

$$f_{X_1} * f_{X_2} = \int_{\mathbb{R}} f_{X_1}(x)\, f_{X_2}(s-x)\, dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(s-x-\mu_2)^2} dx$$

Note: HARD PROOF

$$\overset{\text{must}}{=} \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} e^{-\frac{1}{2(\sigma_1^2+\sigma_2^2)}(x-\mu_1-\mu_2)^2}$$

Let's return to our problem...

$$p(X^* | X, \sigma^2) = \int_{\mathbb{R}} \underbrace{\frac{1}{\sqrt{2\pi\sigma_p^2}} e^{-\frac{1}{2\sigma_p^2}(\theta - \theta_p)^2}} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x^* - \theta - 0)}} d\theta$$

$S = X^*$

$\sigma_1^2 = \sigma_p^2$ $\quad$ $\sigma_2^2 = \sigma^2$

$\mu_1 = \theta_p$ $\quad$ $\mu_2 = 0$

$$= N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$= N(\theta_p, \sigma_p^2 + \sigma^2)$$

$\Rightarrow$ A mixture of normal whose mean is draw normal is normal

Sum up....

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

$\theta | \sigma^2 \sim N(\mu_0, \tau^2)$

$\theta | X_1, \dots, X_n, \sigma^2 \sim N(\theta_p, \sigma_p^2)$ $\quad$ $\theta_p = \dfrac{\frac{\bar{X}n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$, $\sigma_p^2 = \dfrac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$

$X^* | X_1, \dots, X_n, \sigma^2 \sim N(\theta_p, \sigma_p^2 + \sigma^2)$

if $\tau^2 \to 0$

$\Rightarrow \theta | X_1, \dots X_n, \sigma^2 \sim N\left(\bar{X}, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$

$X^* | X_1, \dots X_n, \sigma^2 \sim N\left(\bar{X}, \left(\sqrt{\frac{\sigma^2}{n} + \sigma^2}\right)^2\right)$

$$\| $$

$$\left(\sigma\sqrt{\tfrac{1}{n} + 1}\right)^2$$

$$\left(\sigma\sqrt{\tfrac{n+1}{n}}\right)^2 \qquad \sqrt{\frac{n+1}{n}} \to 1$$

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

if θ known, σ^2 unknown, what is MLE of σ^2?

$$\mathcal{L}(\sigma^2; X,\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X_i-\theta)^2}$$

$(2\pi\sigma^2)^{-\frac{n}{2}}$

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

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

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2)$$

$$\ell'(\sigma^2; X,\theta) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum(X_i-\theta)^2 = 0$$

$$\Rightarrow -n + \frac{1}{\sigma^2}\sum(x_i-\theta)^2 = 0 \Rightarrow \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\theta)^2$$

$$\Rightarrow n\hat{\sigma}^2_{MLE} = \sum(X_i-\theta)^2$$ AKA sum of sqd error (SSE)

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

$Y = \frac{1}{\theta} \sim ?$ Use c.o.v. $y = t(\theta) = \frac{1}{\theta} \Rightarrow \theta = t^{-1}(y) = \frac{1}{y} = y^{-1}$

InvGamma(α,β) !!

$|-y^{-2}| = y^{-2}$

$$f_Y(y) = f_\theta(t^{-1}(y))\left|\frac{d}{dy}(t^{-1}(y))\right| = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\frac{\beta}{y}} \left(\frac{1}{y}\right)^{\alpha-1} \left|\frac{d}{dy}[y^{-1}]\right| = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\frac{\beta}{y}} y^{-\alpha-1}$$

$y^{1-\alpha}$

$E[Y] = \frac{\beta}{\alpha-1}$ if $\alpha>1$, $\text{Mode}(Y) = \frac{\beta}{\alpha+1}$, $\text{Median}(Y) = \text{qinvgamma}(0.5, \alpha, \beta)$

$\text{Supp}(Y) = (0,\infty)$

Param Space $\alpha, \beta > 0$

σ^2 unknown. Use Bayesian inference

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

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

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

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

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

matching kernels ...

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

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

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

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

However... we usually don't use α, β. We use a parameterization that mirrors the pseudocount interpretation

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

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