Semi-conjugacy
No ridge regression

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

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

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

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

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

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

$$= \int_0^\infty p(\theta, \sigma^2 | X)\, d\sigma^2$$

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

4/27/17

Grid sampling

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

$p(\theta, \sigma^2) \propto \frac{1}{\sigma^2}$

$\quad \| $

$p(\theta)p(\sigma^2)$

$\propto 1 \quad \propto \frac{1}{\sigma^2}$ } Uninformative/Jeffreys

$\Rightarrow p(\theta, \sigma^2|X) = N\text{InvG}$

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

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

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

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

CR? Hyp. Test?

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

not given θ $\rightarrow$ $P(X^*|X) = \int_0^\infty\int_{\mathbb{R}} P(X^*|\theta,\sigma^2)\,P(\theta,\sigma^2|X)\,d\theta\,d\sigma^2$

$(X^*|\theta,\sigma^2 \sim N(\theta,\sigma^2))$

$$\propto \int\int \left((\sigma^2)^{-1/2} e^{-\frac{1}{2\sigma^2}(X^*-\theta)^2}\right)(\sigma^2)^{-\frac{n}{2}-1} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\theta)^2}\,d\theta\,d\sigma^2$$

$$= \int_0^\infty (\sigma^2)^{-(\frac{n}{2}+1)-1}\int_{\mathbb{R}} \underbrace{e^{-\frac{1}{2\sigma^2}\left((X^*-\theta)^2+\sum(X_i-\theta)^2\right)}}_{\text{Normal}}\,d\theta\,d\sigma^2$$

$$= \ldots\ldots \propto t_{n-1}\left(\overline{X},\ S\sqrt{\frac{n+1}{n}}\right) = N(\theta = \overline{X}, \sigma^2 = S^2)$$

$t_n \approx N \quad (n \approx 30)$

$\frac{n+1}{n} \rightarrow 1, \quad \overline{X} \rightarrow \theta, \quad S \rightarrow \sigma$

$$= \int_0^\infty\int_{\mathbb{R}} \underbrace{P(X^*|\theta,\sigma^2)}_{N(\theta,\sigma^2)}\ \underbrace{P(\theta|X,\sigma^2)}_{N(\overline{X},(\frac{\sigma}{\sqrt{n}})^2)}\ \underbrace{P(\sigma^2|X)}_{\text{InvGamma}(\frac{n-1}{2},\frac{(n-1)S^2}{2})}\,d\theta\,d\sigma^2$$

How to sample from $P(X^*|X)$?

Step 1: Sample σ_0^2 from $\text{InvGamma}\left(\frac{n-1}{2}, \frac{(n-1)S^2}{2}\right)$

Step 2: Sample θ_0 from $N\left(\overline{X}, \left(\frac{\sigma_0}{\sqrt{n}}\right)^2\right)$

Step 3: Sample X^* from $N(\theta_0, \sigma_0^2)$

Step 4: Repeat steps 1-3 many times (S) and return only $X_1^*, \ldots, X_S^*$.

$P(\theta,\sigma^2) = P(\theta)P(\sigma^2)$ a priori independence

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

$\sigma^2 \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right)$

$\frac{n_0\sigma_0^2 + n\hat{\sigma}^2}{2}$

$$P(\theta, \sigma^2) = P(\theta \mid \sigma^2)\, P(\sigma^2)$$

If $\tau^2 = \dfrac{\sigma^2}{m}$

Consider the case $\tau^2 \neq \dfrac{\sigma^2}{m}$

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

$$\propto \left((\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\left((n-1)s^2 + n(\bar{X}-\theta)^2\right)}\right)\left(e^{-\frac{1}{2\tau^2}(\theta-\mu_0)^2}\right)$$

$$\left((\sigma^2)^{-\left(\frac{n_0}{2}+1\right)} e^{-\frac{n_0\sigma_0^2/2}{\sigma^2}}\right) = (\sigma^2)^{-\frac{n}{2}-\left(\frac{n_0}{2}+1\right)}$$

$$e^{-\frac{1}{2\sigma^2}\left((n-1)s^2 + n_0\sigma_0^2\right)}\, e^{\overbrace{-\frac{n}{2\sigma^2}(\bar{X}-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu_0)^2}^{-\frac{n\bar{X}^2}{2\sigma^2} + \frac{n\bar{x}\theta}{\sigma^2} - \frac{n\theta^2}{2\sigma^2} - \frac{\theta^2}{2\tau^2} + \frac{\theta\mu_0}{\tau^2} - \frac{\mu_0^2}{2\tau^2}}}$$

$$\propto \overbrace{(\sigma^2)^{-\frac{n}{2}-\left(\frac{n_0}{2}+1\right)} e^{-\frac{1}{2\sigma^2}\left((n-1)s^2 + n_0\sigma_0^2 + n\bar{x}^2\right)}}^{A}\, e^{-\left(\frac{n}{2\sigma^2}+\frac{1}{2\tau^2}\right)\theta^2 + \left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\tau^2}\right)\theta}$$

$$\propto N\Bigg(\underbrace{\frac{\frac{n\bar{X}}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}}_{\theta_p}, \underbrace{\frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}}_{\sigma_p^2}\Bigg)$$

$$\propto A\sqrt{2\pi\sigma_p^2}\, e^{-\frac{\theta_p^2}{2\sigma_p^2}}\, N(\theta_p, \sigma_p^2) = (\sigma^2)^{-\frac{n}{2}-\left(\frac{n_0}{2}+1\right)}$$

$$\underbrace{e^{-\frac{1}{2\sigma^2}\left((n-1)s^2 + n_0\sigma_0^2 + n\bar{x}^2\right)} \sqrt{2\pi \frac{1}{\frac{n}{2\sigma^2}+\frac{1}{2\tau^2}}}\, e^{-\frac{\theta_p^2}{2\sigma_p^2}}}_{K(\sigma^2 \mid X)}\, N(\theta_p, \sigma_p^2)$$

$= N(\theta_p, \sigma_p^2)\, K(\sigma^2 \mid X)$ $\not\propto$ any known distribution

How to sample from the posterior $P(\theta, \sigma^2 \mid X)$.

<u>Step 1:</u> Sample from $K(\sigma^2 \mid X)$

<u>Step 2:</u> Sample θ_0 from $N\left(\dfrac{\frac{n\bar{x}}{\sigma_0^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma_p^2} + \frac{1}{\tau^2}}, \dfrac{1}{\frac{n}{\sigma_0^2} + \frac{1}{\tau^2}}\right)$

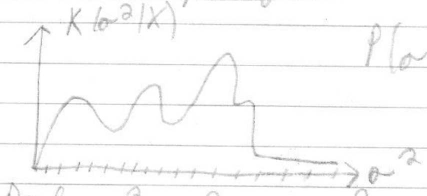
<u>Step 3:</u> Record $\langle \theta_0, \sigma_0^2 \rangle$

<u>Step 4:</u> Repeat steps 1-3 S times.

$$P(\sigma^2 \mid X) = cK(\sigma^2 \mid X)$$

Metropolis - Hastings algorithm
Gibbs Sampling

How to sample from $K(\sigma^2|X)$



$$P(\sigma^2|X) = c\,K(\sigma^2|X)$$

Step 0

① Pick $\sigma^2_{min}, \sigma^2_{max}, \Delta\sigma^2$

② Create grid $G := \langle \sigma^2_{min}, \sigma^2_{min} + \Delta\sigma^2, \sigma^2_{min} + 2\Delta\sigma^2 + \ldots, \sigma^2_{max} \rangle$

③ Compute c. $c \approx \dfrac{1}{\sum_{\sigma^2 \in G} K(\sigma^2|X)}$

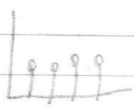
$$1 = \int f(x) = \int c\,g(x) = c\int g(x) = c \cdot \frac{1}{c} = 1 \Rightarrow c = \frac{1}{\int g(x)}$$

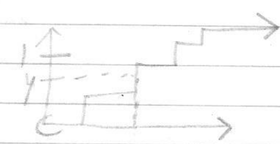
④ Compute $F(\sigma_0^2|X) \approx \sum_{\{\sigma^2 \in G : \sigma^2 \leq \sigma_0^2\}} c\,K(\sigma^2|X) \quad \forall \sigma_0^2 \in G$

Step 1

⑤ Draw y from $U(0,1)$

⑥ Compute $\sigma_0^2 \approx F^{-1}(y)$





Grid Sampling Disadvantages

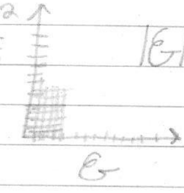
① Numerically unstable (computers have trouble w/ very big #'s & very small #'s)

② Arbitrary decisions for $\theta_{min}, \theta_{max}, \Delta\theta$.

③ Imagine $\theta_{min} = 0, \theta_{max} = 1, \Delta\theta = 10^{-5} \Rightarrow |G| = 10^5$

If posterior had 10 dimensions

$\Rightarrow |G| = 10^{5^{10}} = 10^{50} \Rightarrow$ IMPOSSIBLE



$|G|^2$

Grid sampling only works in low dimensions if you know the support (to pick $\theta_{min}, \theta_{max}$) and you know the slope (to pick $\Delta\theta$).