Lecture 16:

4/27/17

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

$P(\theta, \sigma^2) \propto \frac{1}{\sigma^2}$ => uniformative / jeffrey's

$\underbrace{P(\theta) P(\sigma^2)}_{\propto\, 1 \cdot \frac{1}{\sigma^2}}$

=> $P(\theta, \sigma^2 | X)$ = NormInvGamma → factors into InvGam & Normal

$P(\theta | X, \sigma^2) = N(\bar{x}, (\frac{\sigma}{\sqrt{n}})^2)$

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

$P(\theta | X) = T_{n-1}(\bar{x}, \frac{s}{\sqrt{n}})$

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

} take out the affect of uncertainty of θ <much more realistic>

Credible Region? Hypothesis testing?

- can be calculated directly since $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$

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

$$\propto \iint \left((\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(x^* - \theta)^2}\right) \underbrace{\left((\sigma^2)^{-\frac{n}{2}-1} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \theta)^2}\right)}_{\text{joint likelihoods}} d\theta \, d\sigma^2$$

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

$$= \ldots\ldots \propto T_{n-1}\left(\bar{x}, s\sqrt{\frac{n+1}{n}}\right)$$

$T_n \rightarrow Z \sim N(0,1)$ ↳ can approximate this to $N(\theta, \sigma^2)$

$\frac{n+1}{n} \rightarrow 1$

$\bar{x} \rightarrow \theta$

$s \rightarrow \sigma$

For computer:

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

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

Step 1: sample σ_0^2 from $\text{InvGam}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$

Step 2: sample θ_0 from $N\left(\bar{x}, \left(\frac{\sigma_0}{\sqrt{n}}\right)^2\right)$

Step 3: sample x^* from $N(\theta_0, \sigma_0^2)$

Step 4: repeat step 1-3 many $\overset{(S)}{\text{times}}$ and return only $x_1^*, \dots, x_S^*$

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

$$\underbrace{P(\theta, \sigma^2) = P(\theta)\, P(\sigma^2)}_{\text{a priori independence}}$$

$\theta \sim N(\mu_0, \tau^2) \Rightarrow$ used as prior of θ when variance is known

$\sigma^2 \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0 \sigma_0^2}{2}\right) \Rightarrow$ used as prior when θ is known

Consider the case $\tau^2 \neq \frac{\sigma^2}{m}$ not allowing θ & σ^2 distributions to affect one another

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

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

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

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

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

$$\propto A \overbrace{\sqrt{2\pi\sigma_p^2}\, e^{-\frac{\theta_p^2}{2\sigma_p^2}}}^{\text{putting back constants that were taken out from the kernel}} N(\theta_p, \sigma_p^2)$$

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

$\not\propto$ to known distribution

$$= N(\theta_p, \sigma_p^2)\, K(\sigma^2 | x)$$

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

Step 1: Sample σ_0^2 from $K(\sigma^2 | X)$

Step 2: sample θ_0 from $N\left(\dfrac{\frac{n\bar{x}}{\sigma_0^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma_0^2} + \frac{1}{\tau^2}}, \dfrac{1}{\frac{n}{\sigma_0^2} + \frac{1}{\tau^2}} \right)$

Step 3: record $\langle \theta_0, \sigma_0^2 \rangle$

Step 4: repeat steps 1-3 S times.

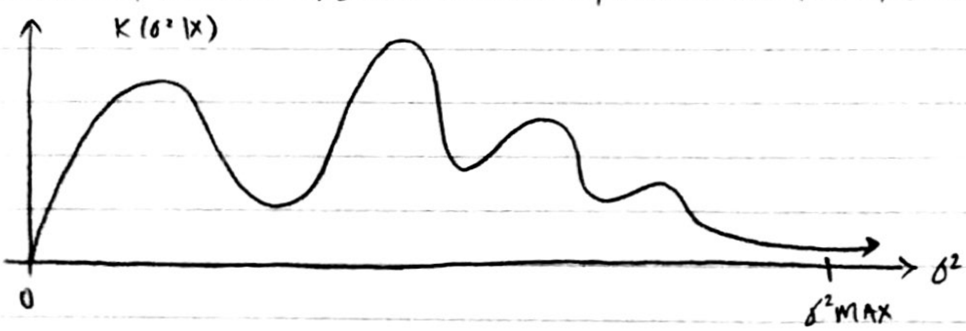
$P(\sigma^2 | X) = c\, K(\sigma^2 | X)$ => if hyperparameter is proper, posterior must be proper

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

step 0

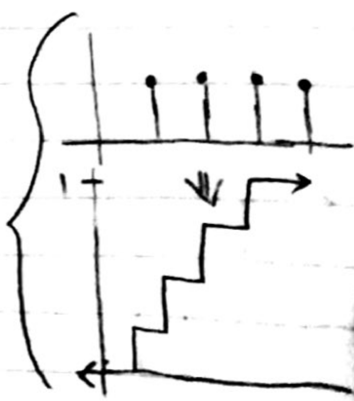
Step 1: Pick σ^2_{min}, σ^2_{MAX}, $\Delta\sigma^2$

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



step 3: compute $c \approx \dfrac{1}{\sum_{\sigma^2 \in \mathcal{G}} K(\sigma^2 | X)}$

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



step 1

step 5: Draw y from $U(0,1)$

step 6: compute $\sigma^2_* \approx F^{-1}(y)$

Grid Sampling Disadvantages:

1. Numerically unstable (computers have trouble with very big numbers and #s & very small #s)
2. Arbitrary decisions for $\theta_{min}, \theta_{max}, \Delta\theta$ $\begin{cases} \text{bad decision for } \theta_{min,max} \Rightarrow \text{miss part of supp of parameter} \\ \text{bad decision for } \Delta\theta \Rightarrow \text{bad resolution.} \end{cases}$
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.

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