Lec 21

5/5/21

From last time

$$= (\sigma^2)^{-\frac{n+n_0}{2}-1} e^{-\frac{1}{2\sigma^2}\left((n-1)s^2+n_0\sigma_0^2+n\bar{x}^2\right)} \cdot \sqrt{\frac{\pi}{b}}\, e^{-\frac{a^2}{4b}}$$

$$\boxed{N\left(\frac{a}{2b}, \frac{1}{2b}\right) = \sqrt{\frac{b}{\pi}}\, e^{\frac{-a^2}{4b}} e^{-a\theta - b\theta^2}}$$

$$\underbrace{N\left(\frac{\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right)}_{P(\theta \mid X, \sigma^2)}$$

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

$$\propto P(\theta \mid X, \sigma^2) \underbrace{(\sigma^2)^{-\frac{n+n_0}{2}-1} e^{-\frac{A}{2\sigma^2}} \left(\frac{n}{2\sigma^2}+\frac{1}{\tau^2}\right)^{-\frac{1}{2}} e^{\left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\tau^2}\right)^2 / \left(\frac{2n}{\sigma^2}+\frac{2}{\tau^2}\right)}}_{K(\sigma^2 \mid X) = (\sigma^2)^{-\alpha-1} e^{-\frac{\beta}{\sigma^2}} \left(\frac{\gamma}{\sigma^2}+\delta\right)^{-\frac{1}{2}} e^{\left(\frac{\xi}{\sigma^2}+\eta\right)^2 / \left(\frac{m}{\sigma^2}+\kappa\right)}}$$

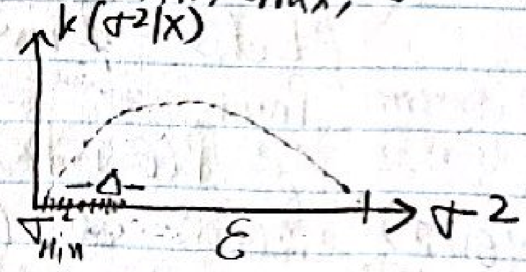
Is this kernel $k(\sigma^2 \mid X)$ proportional to any distribution we know? NO!! This means we can't sample from it using the table you've seen. Get a bigger table? No! This isn't a known distribution. So, we're in trouble... because we can't sample from $P(\sigma^2 \mid X)$ thus we can't sample from the posterior.

We need a general way to sample from kernels of unknown variables.

Grid Sampling Algorithm:

1. Create a grid by first picking $\sigma^2_{min}, \sigma^2_{max}, \Delta$

$$\mathcal{E} := \{\sigma^2_{min}, \sigma^2_{min}+\Delta, \sigma^2_{min}+2\Delta, \ldots \sigma^2_{max}-\Delta, \sigma^2_{max}\}$$



For ex, in our case let

$\sigma^2_{min} = 0, \sigma^2_{max} = 100, \Delta = 0.1$

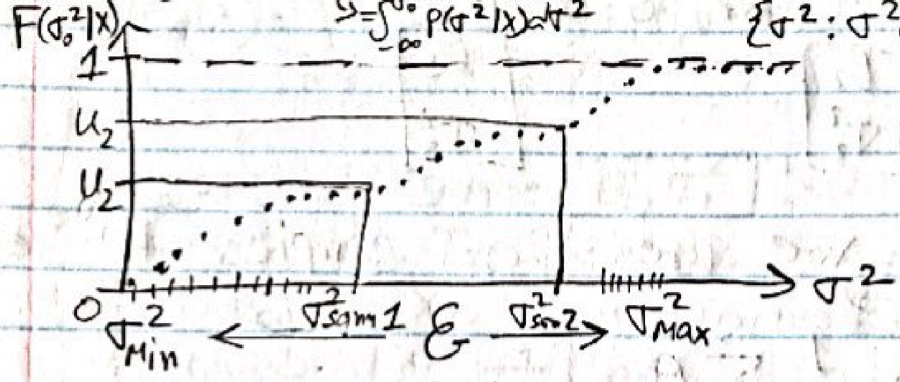
$$\mathcal{E} = \{0, 0.1, 0.2, \ldots 999.9, 1000\}$$

② Approximate the value of c, the normalization constant via Riemann Sum (if continuous)

$$c = \frac{1}{\int k(\sigma^2|X)d\sigma^2} \approx \frac{i}{\Delta \sum_{\sigma^2 \in \mathcal{E}} k(\sigma^2|X)}$$

③ Compute the "sampling CDF" for all grid points:

$$F(\sigma_0^2|X) := \underbrace{P(\sigma^2 \leq \sigma_0^2|X)}_{=\int_{-\infty}^{\sigma_0^2} P(\sigma^2|X)d\sigma^2} \approx \sum_{\{\sigma^2 : \sigma^2 \in \mathcal{E}, \sigma^2 \leq \sigma_0^2\}} c\,k(\sigma^2|X)$$



④ Draw u from $U(0,1)$ and locate

$$\sigma^2_{sample} = \min_{\sigma^2 \in \mathcal{E}} \{F(\sigma^2|X) \geq u\}$$

If you want to draw many samples, repeat step 4. You only need to do step 1-3 once.

Returning to our Problem, how do we sample from the Posterior $P(\theta, \sigma^2|X)$?

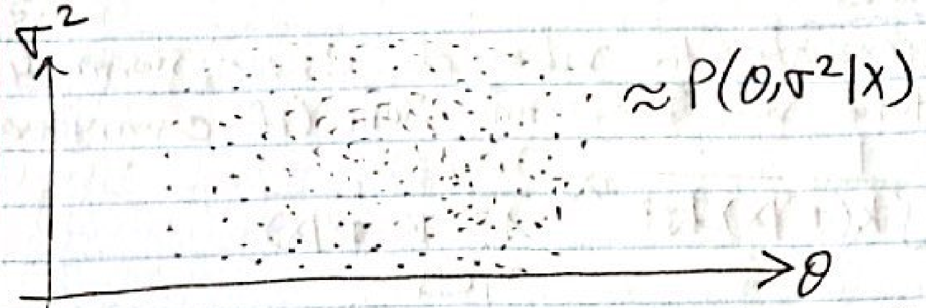
① Sample σ^2_{sample} from $P(\sigma^2|X)$ using the grid sampler

② Sample θ_{sample} from $P(\theta|\sigma^2, X)$ via rnorm.

Return $[\theta_{sample}, \sigma^2_{sample}]$

At this point, we're not limited to distributions we know about nor limited to using conjugate Priors. We're liberated!

We agree we can sample now any arbitury Posterior. But how do we get Point estimates?



I now have as many sample as I wish from my Posterior (you want more? Let the computer churn more). You have S samples:

$$\begin{bmatrix}\theta_1\\ \sigma_1^2\end{bmatrix}, \begin{bmatrix}\theta_2\\ \sigma_2^2\end{bmatrix}, \ldots\ldots \begin{bmatrix}\theta_S\\ \sigma_S^2\end{bmatrix}$$

How do we use these samples to get point estimates?

$$\hat{\theta}_{MMSE} \approx \frac{1}{S}\sum_{k=1}^{S}\theta_k$$

$\hat{\theta}_{MMAE} \approx \text{SampleMedian}[\theta_1, \ldots \theta_S]$ i.e. you order all the samples and return the mid value.

$\hat{\theta}_{MAP}$ is complicated... you need to use smoothing.. forget it!

$$\hat{\sigma}^2_{MMSE} = \frac{1}{S}\sum_{k=1}^{S}\sigma_k^2$$

$$\hat{\sigma}^2_{MMAE} = \text{SampleMedian}[\sigma_1^2, \ldots \sigma_S^2]$$

How do we get CR's?

$$CR_{\theta,1-\alpha_0} = \left[Q\left[\theta|X, \tfrac{\alpha_0}{2}\right], Q\left[\theta|X, 1-\tfrac{\alpha_0}{2}\right]\right]$$

$$\approx \left[\text{SampleQuantile}\left[Q_1, \ldots Q_s, \tfrac{\alpha_0}{2}\right], \text{SampleQuantile}\left[Q_1 \ldots Q_s, 1-\tfrac{\alpha_0}{2}\right]\right]$$

How do we get P-vals for hypothesis tests?

$$H_0: \theta \in \Theta_0, \quad \text{P-val} := P(\theta \in \Theta_0 | X) = \int_{\Theta} p(\theta|X)\, d\theta$$

$$\approx \frac{1}{S}\sum_{k=1}^{S}\mathbb{1}_{\theta_k \in \Theta}$$

e.g. $H_0: \theta \leq 5.89$, $P\text{-val} = \frac{\# \theta_K \leq 5.89}{S}$ = Proportion ≤ 5.89

How do we get the marginal distribution $P(\theta|X)$?

$$P(\theta|X) = \int_{supp[\sigma^2]} P(\theta,\sigma^2|X)\,d\sigma^2 \approx U(\underbrace{\{\theta_1, \dots, \theta_S\}}_{\text{Sample}})$$

uniform discrete distribution

How do we get the posterior Predictive distribution $P(X_*|X)$?

$$P(X_*|X) = \int_0^\infty \int_{\mathbb{R}} \underbrace{P(X_*|\theta,\sigma^2)}_{N(\theta,\sigma^2)} \underbrace{P(\theta,\sigma^2|X)}_{??}\,d\theta\,d\sigma^2 = \iint P(X_*,\theta,\sigma^2|X)\,d\theta\,d\sigma^2$$

You use the sample $[\theta_1, \sigma_1^2]$ and then draw X_{*1} from rnorm(θ_1, σ_1^2). Then use the sample $[\theta_2, \sigma_2^2]$ and then draw X_{*1} from rnorm(θ_2, σ_2^2) .... Then use the sample $[\theta_S, \sigma_S^2]$ and then draw X_{*S} from rnorm(θ_S, σ_S^2):

$$U(\{X_{*_1}, X_{*_2}, \dots, X_{*S}\}) \approx P(X_*|X)$$

$$E[X_*|X] \approx \frac{1}{S}\sum_{K=1}^{S} X_{*K}$$

Disadvantages of the Grid Sampling Algorithm

1) In many dimentions, how do you Pick min, max? For Parameters in multiparameter models, it's not so Simple!

2) Computers have numerical underflow and overflow.

3) "Curse of dimensionlity". Let's Say you want to Sample 10,000 grid Points Per dimention. But You have 10 diffrent θs. $10{,}000^{10} = 10^{50}$ which is Impossible for the computer. As another ex. Let's say I want 1 billion Points in 10 dimentions. That is only : $\sqrt[10]{10^9} \approx 8$ Points Per dimentions.

That's not good resolution at all!

We need another solution to this Problem!!