$$(\sigma^2)^{-\frac{n+n_0}{2}-1} e^{-\frac{1}{2\sigma^2}\underbrace{((n-1)s^2+n_0\sigma_0^2+n\bar{x}^2)}_{A}} = N\left(\frac{a}{2b},\frac{1}{2b}\right) =$$

$$=\sqrt{\frac{b}{\pi}}\, e^{-\frac{a^2}{4b}} e^{a\theta-b\theta^2}, \Rightarrow P(\theta|X,\sigma^2) = N\left(\frac{\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right)$$

$$\propto P(\theta|X,\sigma^2)\underbrace{(\sigma^2)^{-\frac{n+n_0}{2}-1} e^{-\frac{1}{2\sigma^2}((n-1)s^2+n_0\sigma_0^2+n\bar{x}^2)} \left(\frac{n}{2\sigma^2}+\frac{1}{2\tau^2}\right)^{-\frac{1}{2}}\cdot \square}$$

$$\square = \underbrace{e^{\left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\tau^2}\right)^2 \Big/ \left(\frac{2n}{\sigma^2}+\frac{2}{\tau^2}\right)}} \leftarrow k(\sigma^2|x) = (\sigma^2)^{-\alpha-1} e^{-\beta/\sigma^2}\left(\frac{\gamma}{\sigma^2}+\delta\right)^{-\frac{1}{2}} e^{(\epsilon/\sigma^2} \ldots\ldots$$

Is this kernel prop to anything we know?

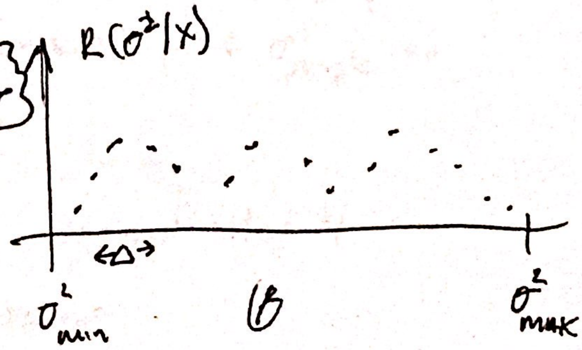
8 parameters. No. This means we can't sample from it using the R-code table we have seen. Not a known distribution. Trouble, cannot sample from $P(\sigma^2|x)$ this cannot sample from the posterior. Cannot use priors from Lec 20. Need a general way to sample from kernels.

Lec 20 ends here.

Grid Sampling Algorithm:

1) Create a grid by first picking σ^2_{min}, σ^2_{max}, Δ

$G = \{\sigma^2_{min}, \sigma^2_{min}+\Delta, \sigma^2_{min}+2\Delta, \ldots, \sigma^2_{max}\}$



For example let $\sigma^2_{min} = 0$, $\sigma^2_{max} = 1000$, $\Delta = 0.1$

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

Step 2) approximate the value of c, the normalization constant

$$c = \frac{1}{\int_{Support[\sigma^2]} k(\sigma^2 | x) d\sigma^2} \approx \frac{1}{\Delta \sum_{\sigma^2 \in G} k(\sigma^2 | x)}$$

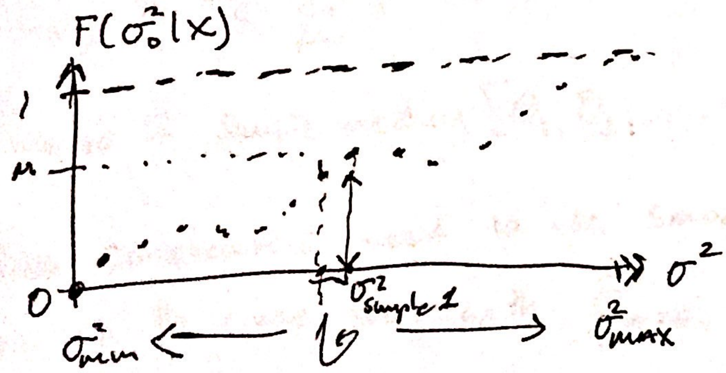
← Riemann Sum (if cts)

If discrete Δ becomes values of support.

3) Compute the "sampling CDF" which is

$$F(\sigma_0^2 | x) := P(\sigma^2 \le \sigma_0^2 | x) = \int_{-\infty}^{\sigma_0^2} p(\sigma^2|x)\, d\sigma^2 \approx \sum_{\{\sigma^2 | \sigma^2 \in G,\ \sigma^2 \le \sigma_0^2\}} c\, k(\sigma^2 | x)$$

For all grid points.



← Approximate CDF

4) Draw u from $U(0,1)$ and locate $\sigma^2_{sample} = \min_{\sigma^2 \in G} \{F(\sigma^2|x) \ge u\}$

pick random number (u) between $(0,1)$ (on y-axis) take σ^2_{sample} as the next $\sigma^2 \in G$ greater than u.

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

Returning to our problem. How do we sample from the posterior $P(\theta, \sigma^2 | X)$?

1) sample from $P(\sigma^2 | X)$ using the grid sampler
2) sample from $P(\theta | \sigma^2, X)$ via r-norm

Return $[\theta_{samp}, \sigma^2_{sample}]$. At this point we are not limited to distributions we know about or limited to using conjugate priors. We agree we can sample now from any arbitrary posterior. How do we get point estimates?

I now have as many as I wish, know posterior to arbitrary precision.

$\begin{bmatrix}\theta_1 \\ \sigma_1^2\end{bmatrix}, \begin{bmatrix}\theta_2 \\ \sigma_2^2\end{bmatrix}, \ldots, \begin{bmatrix}\theta_S \\ \sigma_S^2\end{bmatrix}$ ← you have S samples (S large)

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

$\hat{\theta}_{MMAE} \approx$ sample median $[\theta_1, \theta_2, \ldots, \theta_S]$ ← order them and look for middle value.

$\hat{\theta}_{MAP}$ complicated, need to use smoothing.

can do the same thing with $\hat{\sigma}^2_{MMSE}$ / $\hat{\sigma}^2_{MMAE}$

How do we get CR's.

$CR_{\theta, 1-\alpha_0} := \left[Q[\theta|X, \frac{\alpha_0}{2}], Q[\theta|X, 1-\frac{\alpha_0}{2}]\right] \approx$
$\approx \left[\text{Sample Quantile}[\theta_1, \theta_2, \ldots \theta_S, \frac{\alpha_0}{2}], \text{Sample Quantile}[\theta_1 \ldots \theta_S, 1-\frac{\alpha_0}{2}]\right]$

How do we get pvals for hypothesis testing?

$H_0 : \theta \in \Theta_0$, $\text{pval} := P(\theta \in \Theta_0 | X) = \int_{\Theta} P(\theta|X)\, d\theta \approx \rightarrow$

$\rightarrow \frac{1}{S} \sum_{k=1}^{S} \mathbb{1}_{\theta_k \in \Theta_0}$ for example:

$H_0 : \theta \leq 5.89$ then $\text{pval} := \frac{\# \theta_k \leq 5.89}{\text{All samples}}$ = proportion ≤ 5.89.

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

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

difficult before as $P(\theta|X) = \int_{\text{Supp of } \sigma^2} P(\theta, \sigma^2|X)\, d\sigma^2 \approx \{\theta_1, \ldots, \theta_S\}$

re samples

$= \mathcal{U}(\{\theta_1, \ldots \theta_S\})$

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

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

You use $[\theta_1, \sigma_1^2]$ a draw from, and then draw X_1^* from

$\text{rnorm}(\theta_1, \sigma_1^2)$ then use sample $[\theta_2, \sigma_2^2]$ to repeat

and get X_2^* all the way up to X_S^* hence

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

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

Disadvantages of Grid Sampling Algorithm.

① In many dimensions how do you pick a min, max? For parameters in multi-para models not simple.

② Computers have numerical underflow and overflow. Limit to min, max in a way a computer represents numbers.

③ "Curse of dimensionality" Say ~~sample~~ you want to sample 10,000 pts per dimension, with 10 different thetas

$(10,000)^{10} = 10^{50}$ → impossible for computer.

Another ex. Lets say I want 1 billion pts in 10 dimensions.

$\sqrt[10]{10^9} \approx 8$ per dimension. Not good resolution need another solution. Grid sampling looks good but does not scale well to problems with a lot of dimensions.