Lecture 23 MATH 341/641 (FINAL)

Posterior Predictive Intervals

Given likelihood, $\underbrace{X_1, \ldots, X_n}_{\text{data}}$ $\quad \underbrace{X_*}_{\text{future observation}}$

Use ~~data~~ likelihood and prior $\Rightarrow$ posterior distn $p(\vec{\theta} \mid \vec{x})$

Use likelihood + posterior $\Rightarrow$ posterior distn $p(X_* \mid \vec{x})$

What if I want a 2-sided range of possible values for X_*?

$$PI_{X_*, 1-\alpha} := \left[Q\left[X_* \mid \vec{x}, \frac{\alpha}{2}\right], Q\left[X_* \mid \vec{x}, 1 - \frac{\alpha}{2}\right] \right]$$

PI: predictive interval; $1-\alpha$: prob.

The three you are expected to know...

For $X \sim \text{Binom}(n, \theta)$, $\theta \sim \text{Beta}(\alpha, \beta) \Rightarrow \theta \mid x \sim \text{Beta}(\alpha + x, \beta + n - x)$

$\Rightarrow X_* \mid x \sim \text{BetaBinom}(n, \alpha + x, \beta + n - x)$

$$\Rightarrow PI_{X_*, 1-\alpha} = \left[\text{qbetabinom}\left(\frac{\alpha}{2}, n, \alpha + x, \beta + n - x\right), \text{qbetabinom}\left(1 - \frac{\alpha}{2}, n, \alpha + x, \beta + n - x\right) \right]$$

For $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\theta)$, $\theta \sim \text{Gamma}(\alpha, \beta) \Rightarrow \theta \mid \vec{x} \sim \text{Gamma}(\alpha + \sum x_i, \beta + n)$

$\Rightarrow X_* \mid \vec{x} \sim \text{ExtNegBin}\left(\alpha + \sum x_i, \frac{\beta + n}{\beta + n + 1}\right)$

$$\Rightarrow PI_{X_*, 1-\alpha} = \left[\text{qnbinom}\left(\frac{\alpha}{2}, \alpha + \sum x_i, \frac{\beta + n}{\beta + n + 1}\right), \text{qnbinom}\left(1 - \frac{\alpha}{2}, \alpha + \sum x_i, \frac{\beta + n}{\beta + n + 1}\right) \right]$$

For $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$, $f(\theta, \sigma^2) \propto (\sigma^2)^{-1}$ (Jeffreys) $\Rightarrow \theta, \sigma^2 \mid \vec{x} \sim \text{NormInvGamma}(\cdot, \cdot, \cdot, \cdot)$

$\Rightarrow X_* \mid \vec{x} \sim T_{n-1}\left(\bar{x}, \frac{n+1}{n} s^2\right)$

$$\Rightarrow PI_{X_*, 1-\alpha} = \left[\text{qt.scaled}\left(\frac{\alpha}{2}, n-1, \bar{x}, \frac{n+1}{n} s^2\right), \text{qt.scaled}\left(1 - \frac{\alpha}{2}, n-1, \bar{x}, \frac{n+1}{n} s^2\right) \right]$$

$\boxed{\text{END}}$

Pbl: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$, θ, σ^2 both unknown

$f(\theta, \sigma^2) = f(\theta) f(\sigma^2)$

Let, where $f(\theta) = N(\mu_0, \tau^2)$ indep of $f(\sigma^2) = \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0 \sigma_0^2}{2}\right)$

This is differs from $f(\theta \mid \sigma^2) = N\left(\mu_0, \frac{\sigma^2}{n_0}\right)$ which resulted in the conjugate NormInvGamma distr. What happens now?

$$f(\theta, \sigma^2 \mid x) \propto f(\vec{x} \mid \theta, \sigma^2) f(\theta) f(\sigma^2)$$

$$n\bar{x}^2 - 2n\bar{x}\theta + n\theta^2 \qquad -\frac{1}{2\tau^2}\theta^2 + \frac{\mu_0}{\tau^2}\theta - \frac{\mu_0^2}{2\tau^2}$$

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

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

$$= \underbrace{(\sigma^2)^{-\frac{n}{2}-\frac{n_0}{2}-1} e^{-\frac{A}{\sigma^2}} \sqrt{\frac{\pi}{b}}\, e^{\frac{a^2}{4b}}}_{k(\sigma^2 \mid \vec{x})} \underbrace{N\left(\frac{a}{2b}, \frac{1}{2b}\right)}$$

$$N\left(\frac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right) = f(\theta \mid \vec{x}, \sigma^2)$$

same as before...

$$k(\sigma^2 \mid \vec{x}) \propto (\sigma^2)^{-\frac{n+n_0}{2}-1} e^{-\frac{A}{\sigma^2}} \left(\frac{n}{2\sigma^2} + \frac{1}{\tau^2}\right)^{-\frac{1}{2}} e^{\frac{\left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)^2}{4\left(\frac{n}{2\sigma^2} + \frac{1}{2\tau^2}\right)}}$$

$$= (\sigma^2)^{-\theta_1 - 1} e^{-\frac{\theta_2}{\sigma^2}} \left(\frac{\theta_3}{\sigma^2} + \theta_4\right)^{-\frac{1}{2}} e^{\frac{\left(\frac{\theta_5}{\sigma^2} + \theta_6\right)^2}{4\left(\frac{\theta_3}{\sigma^2} + \theta_4\right)}}$$

which is not the kernel of any known distr

In fact, it has 6 params!

What can we do? We can still find $\begin{bmatrix} \hat{\theta}_{map} \\ \hat{\sigma}^2_{map} \end{bmatrix}$ Can we find $\hat{\theta}_{mmse}$?

If we can sample from $f(\theta, \sigma^2 \mid \vec{x})$, we can get samples of $\langle \theta_1, \sigma_1^2 \rangle, \langle \theta_2, \sigma_2^2 \rangle, \ldots, \langle \theta_S, \sigma_S^2 \rangle$. Then $f(\theta \mid \vec{x})$ is approximated via $\theta_1, \theta_2, \ldots, \theta_S$ and $f(\sigma^2 \mid \vec{x})$ is approximated via $\sigma_1^2, \sigma_2^2, \ldots, \sigma_S^2$.

$\sum_{i=1}^{S}\theta_i$ sample; median$[\theta_1,\ldots,\theta_S]$

You can use these samples to then approximate $\hat{\theta}^{\text{MMSE}}$, $\hat{\theta}^{\text{MMAE}}$, $\hat{\sigma}^{2\,\text{MMSE}}$, $\hat{\sigma}^{2\,\text{MMAE}}$

$CR_{\theta,1-\alpha_0}$, $CR_{\sigma^2,1-\alpha_0}$ and pvals for tests. How to sample?

$H_0: \theta \le \theta_0$

$P(H_0 \mid \vec{x}) = \text{pval} \approx \frac{1}{S}\sum_{i=1}^{S} \mathbb{1}_{\theta_i \le \theta_0}$

$$f(\theta, \sigma^2 \mid \vec{x}) = f(\theta \mid \vec{x}, \sigma^2)\, f(\sigma^2 \mid \vec{x})$$

1. sample σ^2 from $f(\sigma^2 \mid \vec{x})$
2. sample θ from $f(\theta \mid \vec{x}, \sigma^2)$
3. Repeat 1-2 S times, to obtain $\langle \theta_1, \ldots, \theta_S \rangle$, $\langle \sigma^2_1, \ldots, \sigma^2_S \rangle$

How to sample from $f(\sigma^2 \mid \vec{x})$? Grid sample $k(\sigma^2 \mid \vec{x})$ to find normalization constant. Then build $\hat{F}(\sigma^2 \mid \vec{x})$. Then sample u from $U(0,1)$ and find min σ^2 s.t. $\hat{F}(\sigma^2 \mid \vec{x}) \ge u$.

How to get $f(x_* \mid \vec{x})$?

Do sampling above. Step

3. Sample x_* from $N(\theta, \sigma^2)$ (θ from step 1, σ^2 from step 2)
4. Repeat steps 1-3 to obtain $\langle x_{*_1}, \ldots x_{*_S} \rangle$.

You can ~~then~~ estimate $E[x_* \mid \vec{x}]$, $PI_{x_*, 1-\alpha}$, etc...

How can this fail? Right now you used grid sampling to approx. $F(\sigma^2 \mid \vec{x})$ from $k(\sigma^2 \mid \vec{x})$. If there were many dimensions, grid search fails. Why? $\dim[\vec{\theta}] = 10$. You have 1,000,000,000 samples, that's only $\sqrt[10]{10^9} \approx 8$ points for each dimension $\Rightarrow$ ~~doesn't~~ have good resolution.

Is there another way?

We know that

$$f(\theta|\vec{x},\sigma^2) = N(\theta_p, \sigma_p^2)$$

$$f(\sigma^2|\vec{x},\theta) = \text{InvGamma}\left(\frac{\nu_0+n}{2}, \frac{\nu_0\sigma_0^2 + n\hat{\sigma}^2}{2}\right)$$

The systematic sweep Gibbs Sampler approximates $f(\theta,\sigma^2|\vec{x})$ via

Step 1: Begin θ_0 at an initial value

Step 2: Draw σ_1^2 from $f(\sigma^2|\vec{x},\theta)$

Step 3: Draw θ_1 from $f(\theta|\vec{x},\sigma^2)$

Step 4: Draw σ_2^2 from $f(\sigma^2|\vec{x},\theta)$

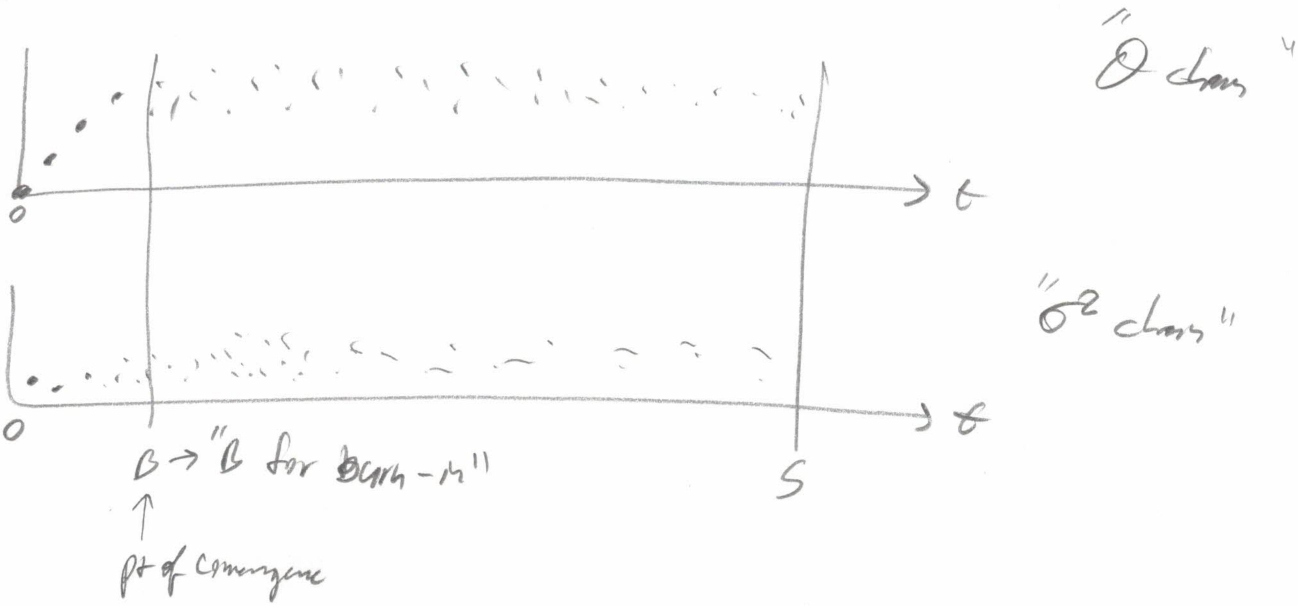
Step 5: Draw θ_2 from $f(\theta|\vec{x},\sigma^2)$

⋮

Note: θ_{t+1} is dependent of θ_t

In 343, I hope we will prove this converges using Markov Chain theory.

Repeat until "convergence" and then collect many samples.



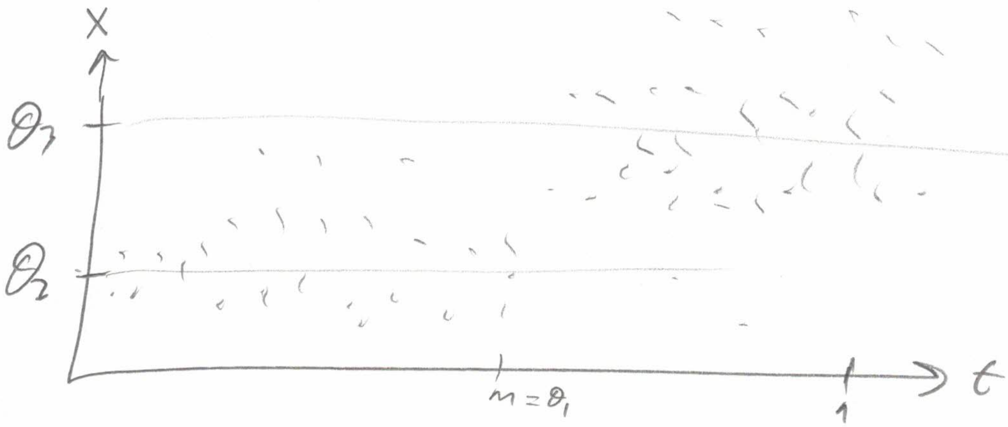
Record but "thinned" i.e. only 1 of every T are kept since the samples are dependent on those previously. Thinning by T breaks this dependence and it results in iid samples from $p(\vec{\theta}|\vec{x})$.

Real world example. DGP $X_1, \ldots, X_{\theta_1} \overset{iid}{\sim} \text{Poisson}(\theta_2)$ indep of

$X_{\theta_1+1}, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\theta_3)$

Parameter space

$\theta_1 \in \{1, 2, \ldots, n\}$, $\theta_2 \in (0, \infty)$, $\theta_3 \in (0, \infty)$



Let's use Laplace for all three params:

[illegible] & [illegible] $\rightarrow$ $p(\theta_1, \theta_2, \theta_3) \propto 1$

$$p(\theta_1, \theta_2, \theta_3 \mid \vec{X}) \propto p(\vec{X} \mid \theta_1, \theta_2, \theta_3)$$

$$= \prod_{t=1}^{\theta_1} \frac{e^{-\theta_2} \theta_2^{x_t}}{x_i} \prod_{t=\theta_1+1}^{n} \frac{e^{-\theta_3} \theta_3^{x_t}}{x_i}$$

$$= \frac{e^{-\theta_1 \theta_2} \theta_2^{\sum_{t=1}^{\theta_1} x_t}}{\prod_{t=1}^{\theta_1} x_i} \frac{e^{-(n-\theta_1)\theta_3} \theta_3^{\sum_{t=\theta_1+1}^{n} x_t}}{\prod_{t=\theta_1+1}^{n} x_i}$$

$$\propto e^{-\theta_1 \theta_2} \theta_2^{\sum_{t=1}^{\theta_1} x_t} \overbrace{e^{-(n-\theta_1)\theta_3}}^{e^{-n\theta_3} e^{\theta_1 \theta_3}} \theta_3^{\sum_{t=\theta_1+1}^{n} x_t}$$

$$p(\theta_1 \mid \vec{x}, \theta_2, \theta_3) \propto e^{(\theta_3 - \theta_2)\theta_1} \theta_2^{\sum_{t=1}^{\theta_1} x_t} \theta_3^{\sum_{t=\theta_1+1}^{n} x_t} = k(\theta_1 \mid \vec{x}, \theta_2, \theta_3)$$

This is not the kernel of any known distr. Luckily gibbs sampling is trivial since $\theta_1 \in \{1, 2, \ldots, n\}$

$$\Rightarrow \frac{1}{c} = \sum_{\theta_1 \in \{1,2,\ldots,n\}} k(\theta_1 | \vec{x}, \theta_2, \theta_3)$$

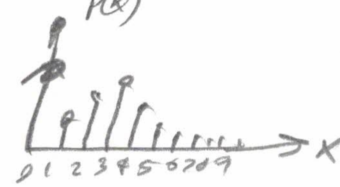
$$\Rightarrow p(\theta_1 | \vec{x}, \theta_2, \theta_3) = c\, k(\theta_1 | \vec{x}, \theta_2, \theta_3)$$ which can then be sampled from

$$f(\theta_2 | \vec{x}, \theta_1, \theta_3) \propto e^{-\theta_1 \theta_2} \theta_2^{\sum_{t=1}^{\theta_1} x_t + 1 - 1} \propto \text{Gamma}\left(1 + \sum_{t=1}^{\theta_1} x_t, \theta_1\right)$$

$$f(\theta_3 | \vec{x}, \theta_1, \theta_2) \propto e^{-(n-\theta_1)\theta_3} \theta_3^{\sum_{t=\theta_1+1}^{n} x_t + 1 - 1} \propto \text{Gamma}\left(1 + \sum_{t=\theta_1+1}^{n} x_t, n - \theta_1\right)$$

Now, we go through the Gibbs sampler... (demo)

Hurdle Model w/ Poisson Counts for data



$$X_1, \dots, X_n \overset{iid}{\sim} \begin{cases} 0 & \text{wp } \theta_1 \\ \text{Poisson}(\theta_2) & \text{wp } (1-\theta_1) \end{cases}$$

i.e. a random data with many comparisons

Using the Laplace prior

$$f(\theta_1, \theta_2 \mid \vec{x}) \propto P(\vec{x} \mid \theta_1, \theta_2)$$

$$= \prod_{i=1}^{n} (\theta_1)^{\mathbb{1}_{x_i=0}} \left((1-\theta_1) \left(\frac{\theta_2^{x_i} e^{-\theta_2}}{x_i} \right) \right)^{\mathbb{1}_{x_i \neq 0}}$$

let $n_0 = \# x_i = 0$

$$\propto \theta_1^{n_0} (1-\theta_1)^{n-n_0} \theta_2^{\sum x_i \mathbb{1}_{x_i \neq 0}} e^{-(n-n_0)\theta_2}$$

$$f(\theta_1 \mid \vec{x}, \theta_2) \propto \theta_1^{n_0+1-1} (1-\theta_1)^{n-n_0+1-1} \propto \text{Beta}(n_0+1, n-n_0+1)$$

$$f(\theta_2 \mid \vec{x}, \theta_1) \propto \theta_2^{\sum x_i \mathbb{1}_{x_i \neq 0} + 1 - 1} e^{-(n-n_0)\theta_2} \propto \text{Gamma}\left(\sum x_i \mathbb{1}_{x_i \neq 0} + 1, n-n_0\right)$$