Lec 23 5/7/19 Math 381

1

Another problem (this's solvable)

Heyic not independent since

θ_t depends on θ_{t-1}. The θ_t also depends ~~on~~ θ_{t-2} and θ_{t-3}...

But at some point the dependence is ~~so~~ low it's negligible.

How can we reason it?

Back to basic stat:

Two r.v.'s X_1, X_2

$\sigma_{12} := \mathrm{Cov}[X_1, X_2] := \mathbb{E}(X_1 - \mu_1)(X_2 - \mu_2)$ degree of linear dependence

$\rho := \mathrm{Corr}[X_1, X_2] = \dfrac{\mathrm{Cov}[X_1, X_2]}{SE(X_1)\, SE(X_2)}$ Reasons see this be $\in [-1, 1]$ and unitless

to ~~Reason~~ estimate these parameters,

$$S_{12} = \frac{1}{S-1} \sum (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)$$

$$S_1 = \sqrt{\frac{1}{S-1} \sum (X_{1i} - \bar{X}_1)^2}, \quad S_2 = \sqrt{\frac{1}{S-1} \sum (X_{2i} - \bar{X}_2)^2}$$

$$r = \frac{S_{12}}{S_1 S_2} = \frac{\sum (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\sqrt{\sum (X_{1i} - \bar{X}_1)^2 \sum (X_{2i} - \bar{X}_2)^2}}$$

Back to our problem... how does θ_t depend on θ_{t-1}?
Estimate its correlation...

$$r_{\theta_1} := \frac{\sum_{t=B+2}^{S} (\theta_t - \bar{\theta})(\theta_{t-1} - \bar{\theta})}{\sum_{t=B+1}^{S} (\theta_t - \bar{\theta})^2}, \quad \bar{\theta} := \frac{1}{S-B} \sum_{t=B+1}^{S} \theta_t$$

autocorrelation: correlation with itself, lag = 1 : one step lookback

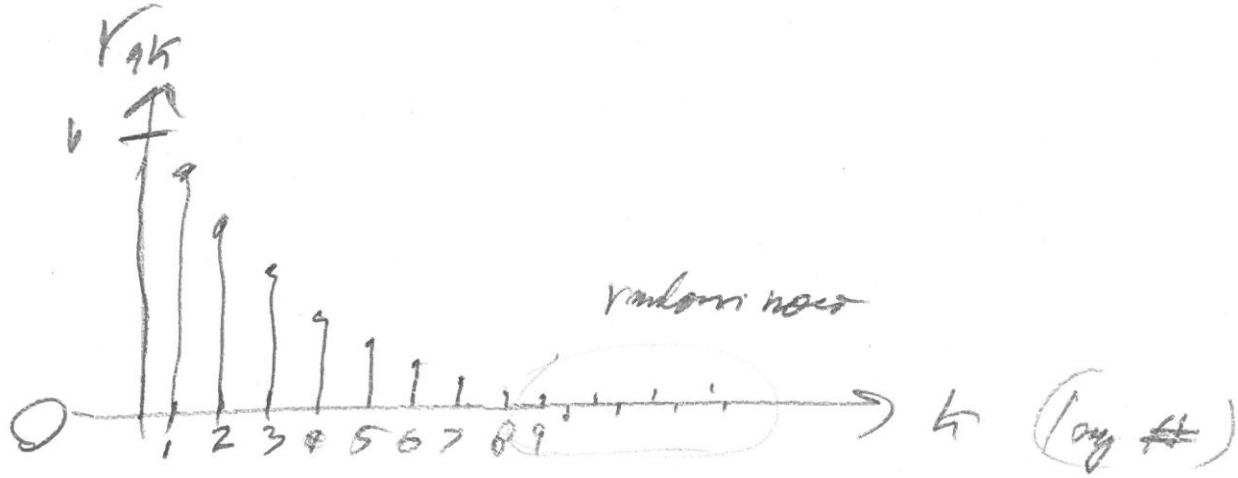
$r_{\theta 1}$? LARGE & POSITIVE

$$r_{\theta 2} = \frac{\sum_{t=B+3}^{S} (\theta_t - \bar{\theta})(\theta_{t-2} - \bar{\theta})}{\sum_{t=B+1}^{S} (\theta_t - \bar{\theta})^2}$$

$\vdots$

$$r_{\theta k} \quad \frac{\sum_{t=B+k+1}^{S} (\theta_t - \bar{\theta})(\theta_{t-k} - \bar{\theta})}{\sum_{t=B+1}^{S} (\theta_t - \bar{\theta})^2}$$

for some lag k, $r_{\theta k} \approx 0$. Why? Eventually samples are $\approx$ independent. How to check? Create an autocorrelation plot.



At $k=9$ samples look independent enough.

Now, if we take every $T=9^{th}$ sample, we have iid draws from $P(\theta_1, \ldots \theta_p | x)$! Deleting the 8 samples in between is called "thinning".

After "burning" and "thinning" we harvest N samples:

$$\Omega := \left\{ \begin{bmatrix} \theta_{1,B+1} \\ \theta_{2,B+1} \\ \vdots \end{bmatrix}, \begin{bmatrix} \theta_{1,B+1+T} \\ \theta_{2,B+1+T} \\ \vdots \end{bmatrix}, \begin{bmatrix} \theta_{1,B+1+2T} \\ \theta_{2,B+1+2T} \\ \vdots \end{bmatrix}, \ldots \begin{bmatrix} \theta_{1,B+1+NT} \\ \theta_{2,B+1+NT} \\ \vdots \end{bmatrix} \right\}$$

braces means no need to order these samples anymore, they're all iid!!

How to do Bayesian inference? EASY!

$$\hat{\theta}_{j,\text{mmse}} = E[\theta_j | X] \approx \frac{1}{N} \sum_{\theta_j \in M} \theta_j$$

$$\hat{\theta}_{j,\text{mmae}} = \text{Med}[\theta_j | X] \approx \text{SampleMedian}[\{\theta_j \in M\}]$$

$$CR_{\theta_j, 1-\alpha} \approx [\text{Samplequantile}[\tfrac{\alpha}{2}, \{\theta_j \in M\}], \text{Samplequantile}[1-\tfrac{\alpha}{2}, \{\theta_j \in M\}]]$$

$H_0: \theta_j \in \Theta_j$

$$p_{val} = P(H_{0,j} | X) \approx \frac{1}{N} \sum_{\theta_j \in M} \mathbb{1}_{\theta_j \in \Theta_j}$$ approximate integral

$$P(X^* | X) = \int_{\Theta} P(X^* | \vec{\theta}) P(\vec{\theta} | X) \, d\vec{\theta} \approx$$ the following sampling algorithm:

(I) select $\vec{\theta}$ from M at random

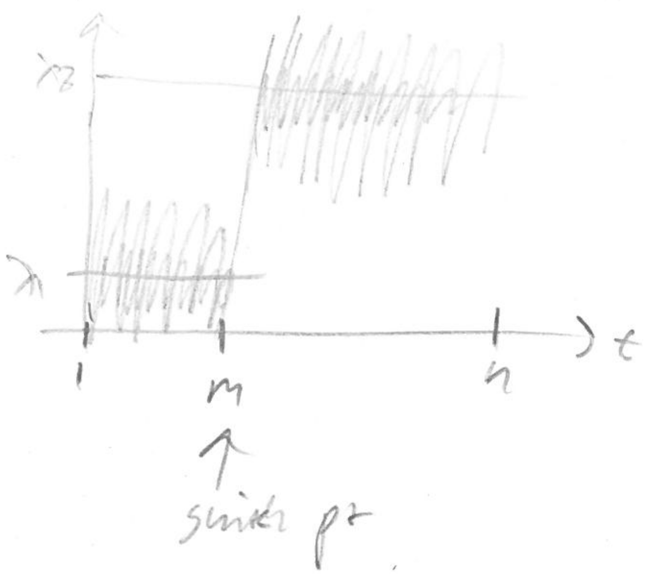
(II) Draw X^*_{samp} from $P(X^* | \vec{\theta})$ and skip X^*_{samp}

(III) Repeat steps I-II S times

Change pt. model

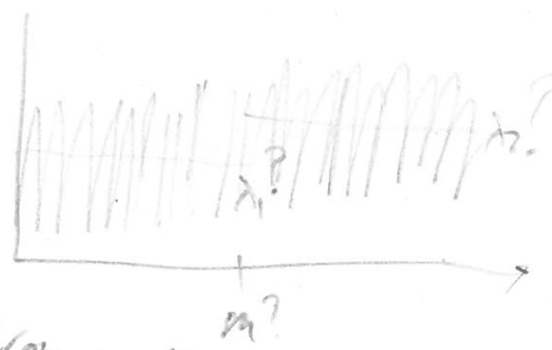
\# of phone cells modeled as a

for a period of time, you have a Poisson iid model with λ_1 and then it switches to a Poisson iid model with λ_2



3 params!
λ_1, λ_2, m

this is obvious! But what if



You would like to capture the uncertainty in all these unknown params

What is data? $X_1, X_2, \dots, X_n$ is # of phone cells at each time $t = 1, 2, \dots, n$

$$P(\lambda_1, \lambda_2, m \mid X_1, \dots, X_n) \propto P(X_1, \dots, X_n \mid \lambda_1, \lambda_2, m)\, P(\lambda_1, \lambda_2, m)$$

What does the likelihood look like? Likelihood assumes params known $P(X|\theta)$

$X_1, \ldots, X_m | \lambda_1 \overset{iid}{\sim} \text{Poisson}(\lambda_1)$

$X_{m+1}, \ldots, X_n | \lambda_2 \overset{iid}{\sim} \text{Poisson}(\lambda_2)$ $\quad X_1, \ldots, X_m$ ind of $X_{m+1}, \ldots X_n$

$$P(X_1, \ldots, X_n | \lambda_1, \lambda_2, m) = P(X_1, \ldots, X_m | \lambda_1) P(X_{m+1}, \ldots, X_n | \lambda_2)$$

$$= \prod_{t=1}^{m} \frac{e^{-\lambda_1} \lambda_1^{x_t}}{x_t!} \prod_{t=m+1}^{n} \frac{e^{-\lambda_2} \lambda_2^{x_t}}{x_t!}$$

Prior

$$P(\lambda_1, \lambda_2, m) \overset{?}{=} P(\lambda_1) P(\lambda_2) P(m)$$ any reason not so?? prob. not

Should we use uninformative priors? Sure...

$P(\lambda_1) = \text{Gamma}(\overset{1}{\alpha_1}, \overset{0}{\beta_1}) \propto 1$ Laplace!

$P(\lambda_2) = \text{Gamma}(\overset{1}{\alpha_2}, \overset{0}{\beta_2}) \propto 1$ "

$m \in \{1, \ldots, n\}$ $\quad \underset{\text{Supp}(m)}{}$ $P(m) = \frac{1}{n}$ discrete uniform from Lec 2 or 3!

Put it all together...

$$P(\lambda_1, \lambda_2, m | X_1 \ldots X_n) \propto \prod_{t=1}^{m} \frac{e^{-\lambda_1} \lambda_1^{x_t}}{x_t!} \prod_{t=m+1}^{n} \frac{e^{-\lambda_2} \lambda_2^{x_t}}{x_t!} (1)(1)\left(\frac{1}{n}\right)$$

$$\propto \frac{\prod_{t=1}^{m} e^{-\lambda_1} \lambda_1^{x_t} \prod_{t=m+1}^{n} e^{-\lambda_2} \lambda_2^{x_t}}{\prod_{t=1}^{n} x_t!} \propto e^{-m\lambda_1} \lambda_1^{\sum_{t=1}^{m} x_t} e^{-(n-m)\lambda_2} \lambda_2^{\sum_{t=m+1}^{n} x_t}$$

$\overset{?}{\propto}$ known distr?? No!

Sol... 3-d grid sampling? No... Gibbs? Let's see...

$$P(\lambda_1 \mid X_1, \dots, X_n, \lambda_2, m) = \lambda_1^{\left(\sum_{t=1}^{m} X_t + 1\right) - 1} e^{-m\lambda_1} \propto \text{Gamma}\left(\sum_{t=1}^{m} X_t + 1, m\right)$$

$$P(\lambda_2 \mid X_1, \dots, X_n, \lambda_1, m) = \lambda_2^{\left(\sum_{t=m+1}^{n} X_t + 1\right) - 1} e^{-(n-m)\lambda_2} \propto \text{Gamma}\left(\sum_{t=m+1}^{n} X_t + 1, n-m\right)$$

$$P(m \mid X_1, \dots, X_n, \lambda_1, \lambda_2) = e^{m(\lambda_2 - \lambda_1)} \lambda_1^{\sum_{t=1}^{m} X_t} \lambda_2^{\sum_{t=m+1}^{n} X_t} \propto \text{? None...}$$

$$= k(m \mid \text{—})$$

$$\parallel \quad \frac{k(m \mid \text{—})}{\sum_{m=1}^{n} k(m \mid \text{—})}$$

Easy to grid sample! Only n pts in support!
No need to select X_{min}, X_{max} or Δ and only 1 dim necessary $\Rightarrow$ No disadvantages...