$= e^{-(m+\beta)\lambda_1} \lambda_1^{a+\alpha-1} e^{-(n-m+\beta)\lambda_2} \lambda_2^{b+\alpha-1}$ unknown distribution

$P(\lambda_1 | X_1, \ldots, X_n, \lambda_2, m) \propto e^{-(m+\beta)\lambda_1} \lambda_1^{a+\alpha-1} \propto$ Gamma$(a+\alpha, m+\beta)$

$P(\lambda_2 | X_1, \ldots, X_n, \lambda_1, m) \propto e^{-(n-m+\beta)\lambda_2} \lambda_2^{b+\alpha-1}$

$\propto$ Gamma$(b+\alpha, n-m+\beta)$

dependent on m

$P(m | X_1, \ldots, X_n, \lambda_1, \lambda_2) \propto \underbrace{e^{-m(\lambda_1-\lambda_2)} \lambda_1^{a} \lambda_2^{b}}_{h(m)}$

$\propto \dfrac{h(m)}{\sum_{k=0}^{n} h(k)}$

$\left\langle \begin{bmatrix} \lambda_{0,1} \\ \lambda_{0,2} \\ m_0 \end{bmatrix}, \begin{bmatrix} \lambda_{1,1} \\ \lambda_{1,2} \\ m_1 \end{bmatrix}, \ldots \right]$

Burn-in line: points follow gray line when it burned-in

Delete other data points; now all correlated b/c one vector comes from other

Take posterior and graph credible region

Middle bar is mean ($\hat{\theta}_{MMSE}$)

$CR_{\theta, 95\%}$: capture θ value 95% of the time

5/16/17

Recall the "Bayesian protocol"

① Pick F, the lik. model, $P(X|\theta)$

② Pick $P(\theta)$ prior.

③ Collect data.

④ Compute posterior $P(\theta|X)$ and you use it to get inference

→ a) do it directly in closed form; e.g. conjugacy

b) If only $k(\theta|X)$, then use grid sampling if you think it will be accurate

c) Gibbs sampling

Inference
- Point estimation
- Credible regions
- Hypothesis testing

What if #1 & #2 are wrong?

i.e. the model is wrong. How do you assess the degree of departure from reality?

Model checking

<u>First check (easy to pass)</u>

Recall $P(X) = \int_{\Theta} P(X|\theta)\,P(\theta)\,d\theta$

prior predictive distr.

This shows you what data (x) looks like assuming $\mathcal{F}$ (i.e. $P(X|\theta)$) and your prior $P(\theta)$.

$P(X|\theta) = \text{Binom}(100, \theta)$

$P(\theta) = U(0,1) = \text{Beta}(1,1)$

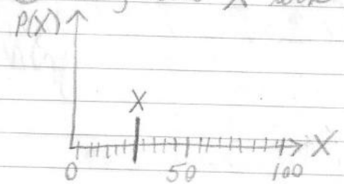
$\Rightarrow P(X) = \text{BetaBinom}(100, 1, 1)$

<u>How to check</u>

① Sample many replications from $P(X)$

② Plot your data (x)

③ ask, "does X look plausible in the context" of $P(X)$?



Uninformative, diffuse prior, everything looks plausible

<u>Second check (after passing first check)</u>

Recall $P(X^*|X) = \int_{\Theta} P(X^*|\theta)\,P(\theta|X)\,d\theta$

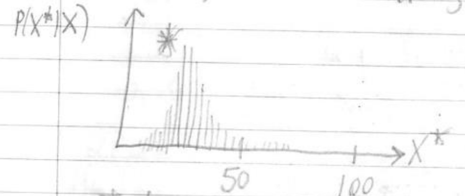
the "posterior predictive distr." AKA "posterior replicative distr." only if $\dim(x^*) = \dim(x)$

① Sample many replications from $P(X^*|X)$

② Plot your data (X).

③ ask, "does your data appear plausible in the context of $P(X^*|X)$?"

$P(X^*|X) = \text{BetaBinom}(100, 30, 72)$



⇒ "Yes"

Test question: would "x" at a point mark? (eg., to the right, no, unreasonable)

Gibbs Sampler

Want to sample a posterior $P(\theta_1, \dots, \theta_p|X)$ which is not easily sampled from directly and you have

$\forall j \;\; P(\theta_j | \theta_{-j}, X)$

i.e. all the conditional dist. that are easy to sample from.

$X_1, \dots, X_n | \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho \overset{iid}{\sim} \rho N(\theta_1, \sigma_1^2) + (1-\rho) N(\theta_2, \sigma_2^2)$

Priors $P(\theta_1) \propto 1$ — know nothing prior

$P(\theta_2) \propto 1$

$P(\sigma_1^2) \propto \frac{1}{\sigma_1^2}$

$P(\sigma_2^2) \propto \frac{1}{\sigma_2^2}$

$P(\rho) = U(0, 1) \propto 1$

Want $P(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho | X)$

EM algorithm gives you $\hat{\theta}_{MAP}$

Use data expression...

$P(I_1, \dots, I_n, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho | X)$

$\propto P(X | I_1, \dots, I_n, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, \rho)$

Each I is a Bernoulli

$$\cdot\, p(I_1,\dots,I_n|\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho)\cdot \underbrace{p(\theta_1,\theta_2,\sigma_1^2,\sigma_2^2,\rho)}_{\frac{1}{\sigma_1^2}\cdot\frac{1}{\sigma_2^2}}$$

$$= \prod_{i=1}^{n} \rho^{I_i}(1-\rho)^{1-I_i}$$

$$\underbrace{\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\, e^{-\frac{1}{2\sigma_1^2}(X_i-\theta_1)^2}\right)^{I_i} \cdot \left(\frac{1}{\sqrt{2\pi\sigma_2^2}}\, e^{-\frac{1}{2\sigma_2^2}(X_i-\theta_2)^2}\right)^{1-I_i}}_{\text{likelihood}}$$

$$\cdot\left(\prod_{i=1}^{n} \rho^{I_i}(1-\rho)^{1-I_i}\right)\left(\frac{1}{\sigma_1^2}\cdot\frac{1}{\sigma_2^2}\right)$$

$I_1,\dots,I_n \overset{iid}{\sim} \mathrm{Bern}(\rho)$

$$\propto \frac{1}{\sigma_1^2}\frac{1}{\sigma_2^2}\prod_{i=1}^{n}\left(\rho\,\frac{1}{\sqrt{\sigma_1^2}}\, e^{-\frac{1}{2\sigma_1^2}(X_i-\theta_1)^2}\right)^{I_i} \cdot\left((1-\rho)\frac{1}{\sqrt{\sigma_2^2}}\, e^{-\frac{1}{2\sigma_2^2}(X_i-\theta_2)^2}\right)^{1-I_i}$$

$$= \frac{1}{\sigma_1^2}\frac{1}{\sigma_2^2}\prod_{i=1}^{n}\left(\frac{\rho}{\sqrt{\sigma_1^2}}\right)^{I_i} e^{-\frac{1}{2\sigma_1^2}I_i(X_i-\theta_1)^2}\left(\frac{1-\rho}{\sqrt{\sigma_2^2}}\right)^{1-I_i} e^{-\frac{1}{2\sigma_2^2}(1-I_i)(X_i-\theta_2)^2}$$

$(\sigma_1^2)^{-1}(\sigma_1^2)^{-\frac{1}{2}\sum I_i}$

$$= \frac{1}{\sigma_1^2}\frac{1}{\sigma_2^2}\left(\frac{\rho}{\sqrt{\sigma_1^2}}\right)^{\sum I_i} e^{-\frac{1}{2\sigma_1^2}\sum I_i(X_i-\theta_1)^2}\left(\frac{1-\rho}{\sqrt{\sigma_2^2}}\right)^{\sum 1-I_i} e^{-\frac{1}{2\sigma_2^2}\sum(1-I_i)(X_i-\theta_2)^2}$$

$$p(\theta_1|\theta_2,\sigma_1^2,\sigma_2^2,\rho,I_1,\dots,I_n,X)\propto$$

$$e^{-\frac{1}{2\sigma_1^2}\sum I_i(X_i-\theta_1)^2}\propto e^{\frac{\sum I_iX_i}{\sigma_1^2}\theta_1-\frac{\sum I_i}{2\sigma_1^2}\theta_1^2}$$

Take kernels

$$\propto N\left(\frac{\sum I_i X_i}{\sum I_i}, \frac{\sigma_1^2}{\sum I_i}\right)$$

$$P(\theta_2 \mid ——) \propto N\left(\frac{\sum(1-I_i)X_i}{\sum 1-I_i}, \frac{\sigma_2^2}{\sum 1-I_i}\right)$$

$$P(\sigma_1^2 \mid ——) \propto (\sigma_1^2)^{-\frac{\sum I_i}{2}-1} e^{-\frac{\sum I_i (X_i-\theta_1)^2/2}{\sigma_1^2}}$$

$$\propto \text{InvGamma}\left(\frac{\sum I_i}{2}, \frac{\sum I_i (X_i-\theta_1)^2}{2}\right)$$

$$P(\sigma_2^2 \mid ——) \propto \text{InvGamma}\left(\frac{\sum 1-I_i}{2}, \frac{\sum(1-I_i)(X_i-\theta_2)^2}{2}\right)$$

$$P(\rho \mid ——) \propto \rho^{\sum I_i}(1-\rho)^{\sum 1-I_i}$$

$$\propto \text{Beta}(1+\sum I_i, 1-\sum 1-I_i)$$

$$P(I_1 \mid ——) \propto \rho^{I_1}\left(e^{-\frac{1}{2\sigma_1^2}(X_1-\theta_1)^2}\right)^{I_1}(1-\rho)^{1-I_i} \cdot \left(e^{-\frac{1}{2\sigma_2^2}(X_1-\theta_2)^2}\right)^{1-I_i}$$

$$= \left(\rho e^{-\frac{1}{2\sigma_1^2}(X_1-\theta_1)^2}\right)^{I_i}\left((1-\rho)e^{-\frac{1}{2\sigma_2^2}(X_1-\theta_2)^2}\right)^{1-I_i}$$

$$= a^{I_1} b^{1-I_i} \propto \text{Bern}\left(\frac{a}{a+b}\right)$$

$$P(I_2 \mid - \ldots) = \ldots = P(I_n \mid - \ldots)$$

Gibbs:

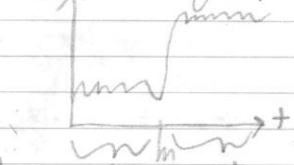
1. Burned in
2. Remove dependence (should like iid) w/ autocorrelation (thin by "#", keep only 1 of #)

Thick line is posterior expectation
Gray is 95% CR

COURSE END

Metropolitan Hastings

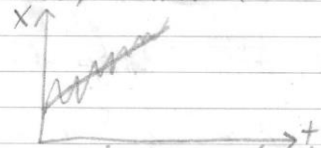
Charge pt model



$\overset{iid}{\sim}$ Poisson (λ_1) $\overset{iid}{\sim}$ Poisson (λ_2)

What if λ was a function of time?

$\lambda(t) = a+bt$ (linear model)



$\vec{\theta} = \{a, b\}$

$\overset{iid}{\sim}$ Poisson $(a+bt)$

Poisson regression

$P(a,b|X,t) \propto P(X|a,b,t)\, P(a,b|t)$

$P(a) \propto$

$P(b) \propto$

$= P(X|a,b,t)$

$$= \prod_{i=1}^{n} \frac{e^{-a+bt_i} \cdot (a+bt_i)^{x_i}}{x_i!} = \frac{e^{-\Sigma a+bt_i} (a+bt_i)^{\Sigma x_i}}{\prod x_i!}$$

$$\propto e^{-\Sigma a+bt_i} (a+bt_i)^{\Sigma x_i}$$

$$P(a|X,t,b) \propto e^{-na} \prod_{i=1}^{n} (a+bt_i)^{x_i} = k(a|\cdots)$$

$$P(b|X,t,a) \propto e^{-b\Sigma t_i} \prod_{i=1}^{n} (a+bt_i)^{x_i} = k(b|\cdots)$$

both conditions are easy to sample from

Metropolis - Hastings Algorithm

Metropolis et. al 1953

Hastings 1970

○ The "Metropolis Step" w/in a Gibbs sampler samples from $k(a|-)$.

Init. a_0

Step 1: Draw a_1 from $q(a=a_0, \emptyset)$ → tuning parameters

↳ conditional density $\sim k(a|-)$

e.g. $q = N(a_0, \sigma^2 = 1)$

Step 2: Calculate

$$r := \frac{\frac{k(a=a_1, b=b_0|X)}{q(a_1; a_0, \emptyset)}}{\underbrace{\frac{k(a=a_0, b=b_0|X)}{q(a_0; a_0, \emptyset)}}_{\text{ratio of kernel prob. of previous } (a_0) \text{ to candidate prob. of previous } (a_0)}}$$

} ratio of kernel prob. of proposal (a_1) to candidate prob. of proposal (a_1)

Step 3

Accept a_1 w.p. r

(if $r \geq 1$, accept w.p. 1)

Gibbs is special type of M-H

Note:

If iterations are same for a spell, there was a rejection

Break dependence

If always reject, go back to candidate

5/18/17

Review Day

HW#7, 8, 9 (focus of final)

HW#9, 5f - Similar to histograms

Look at histogram of marg. dist. of x

Inference for e

CR of 50% (example) approximate is

q ~ .25 to .75, best from picture

Draw 2 gray lines

HW#7, 5n