Back to Poisson example from last class.....

$X_1, \ldots, X_m \overset{iid}{\sim} \text{Poisson}(\lambda_1)$ — independent of each other as well.

$X_{m+1}, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda_2)$

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

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

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

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

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

Lets deal with prior now. Is there any reason why priors should not be independent of each other? Unless you have special knowledge no. So $P(\lambda_1, \lambda_2, m) = P(\lambda_1) P(\lambda_2) P(m)$

What kind of priors do we want to use on all 3? we should use a principled un-informative prior like Laplace.

$P(\lambda_1) = \text{Gamma}(1,0) \propto 1$

$P(\lambda_2) = \text{Gamma}(1,0) \propto 1$

$P(m) =$ .. the support of m is $\{0,1,2,\ldots,n\}$ for a total of $n+1$ values. Uniform Discrete where

$P(m) = 1/n+1 \rightarrow P(m) = U(\{0,1,2,\ldots,n+1\}) \propto 1$

hence $P(\lambda_1)P(\lambda_2)P(m) \propto 1$ so we don't have to deal with it:

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

this is the kernel for the posterior

Go to Gibbs sampler. need conditional distributions for each parameter.

$P(\lambda_1 \mid X_1 \ldots X_m, \lambda_2, m) \propto$ assume you know λ_2 and m

$$\propto e^{-m\lambda_1} \lambda_1^{\sum_{t=1}^{m} X_t + 1 - 1} \propto \text{Gamma}\left(1 + \sum_{t=1}^{m} X_t,\ m\right)$$

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

$$P(m \mid X_1 \ldots X_n, \lambda_1, \lambda_2) \propto e^{m(\lambda_2 - \lambda_1)} \lambda_1^{\sum_{t=1}^{m} X_t} \lambda_2^{\sum_{t=m+1}^{n} X_t} \propto \underline{\text{nothing}}$$

$\downarrow$

$K(m \mid X_1, \ldots, X_n, \lambda_1, \lambda_2)$

Need to grid this.

To sample from $p(m \mid X_1, \ldots X_n, \lambda_1, \lambda_2)$ use grid sampling
~~[illegible]~~ disadvantages of sampling dont apply.
We know where m_{min} m_{max} $min = 0$ $max = n$
Theres no choice of resolution delta. As delta $= 1$
because the support of m is discrete $\{0, 1, \ldots, n\}$
No curse of dimensionality since $d = 1$. However we still
can numerically overflow - underflow. We use a trick for
that. (use logs)

3:50 Demo Skip mixture model on practice tests. MATERIAL DONE
HW That I don't have to hand in but should do
~~[illegible]~~ to prep for mid term.

Review	Sunday 5/23 6PM

Consider a Gibbs sampleR!

$P(\theta_1 \mid \theta_2, \ldots, \theta_p, X)$, ✓

$P(\theta_2 \mid \theta_1, \theta_3, \ldots, \theta_p, X) \propto k(\theta_2 \mid \theta_1, \theta_3, \ldots, \theta_p, X)$

$\vdots$

$P(\theta_p \mid \theta_1, \ldots, \theta_{p-1}, X)$ ✓

what if you dont know the conditional distribution of the second parameter? You only have kernel. Then Grid sample but you may have missing ~~[illegible]~~ info to process the grid, min, max etc maybe we don't want grid sampling. Alternative?

You can use the Metropolis - Hastings Algorithm

① You draw $\theta_{t,2}^{proposed}$ from $q(\theta_{2,t-1}, \phi)$

where q is a proposal distribution which is not the real conditional probability distribution, ϕ are tuning parameters for that proposed distribution eg. $q = N(\theta_{2,t-1}, 1)$

② Calculate

$r :=$ ———