$X_1, \dots, X_m \mid \lambda_1 \overset{iid}{\sim} \text{Poisson}(\lambda_1)$

indep.

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

Prior

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

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

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

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

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

Let's deal with the prior. Is there any reason to not make our priors independent of each other? Unless you have special knowledge... no... so... $P(\lambda_1, \lambda_2, m) = P(\lambda_1) P(\lambda_2) P(m) \propto 1$

What kind of priors do we use? Without prior knowledge, we should use a principled uninformative prior. How about Laplace?

$$P(\lambda_1) = \text{Gamma}(1, 0) \propto 1$$
$$P(\lambda_2) = \text{Gamma}(1, 0) \propto 1$$
$$P(m) = U(\{0, 1, \dots, n+1\}) \propto 1$$

The support of m is... {0, 1, ..., n} for a total of n+1 values. What is the uniform prior on this r.v.? Uniform discrete where P(m) = 1 / (n+1)

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

this is the kernel for the posterior

Let's now find the conditional distributions for each parameter

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

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

$$P(m \mid X_1, \dots, X_n, \lambda_1, \lambda_2) \propto \underbrace{e^{m(\lambda_2 - \lambda_1)} \lambda_1^{\sum_{t=1}^{m} X_t} \lambda_2^{\sum_{t=m+1}^{n} X_t}}_{k(m \mid X_1, \dots, X_n, \lambda_1, \lambda_2)} \propto \text{Nothing you know!}$$

To sample from P(m | .....) you can use grid sampling. Turns out the disadvantages of grid sampling don't apply. We know where m_min and m_max should be i.e. m_min = 0, m_max = n. There's no choice of the resolution, Δ as Δ must be = 1 because the support of m is discrete {0, 1, ..., n}. There's also no curse of dimensionality since there's only one dimension: m! However, we still could numerically overflow/underflow. So we use a trick for that which I'll show you in the demo.

FINAL ↑↑

EXTRA ↓↓

Consider a Gibbs sampler:

$$P(\theta_1 \mid \theta_2, \dots, \theta_p, X) \checkmark$$
$$P(\theta_2 \mid \theta_1, \theta_3, \dots, \theta_p, X) \propto k(\theta_2 \mid \theta_1, \theta_3, \dots, \theta_p, X)$$
$$\vdots$$
$$P(\theta_p \mid \theta_1, \dots, \theta_{p-1}, X) \checkmark$$

What if you don't know the conditional distribution of the second parameter (you only know its kernel). You could grid sample but it may be slow because you need good resolution, you may not know the min/max of the grid which may change iteration-iteration. For these reasons, you want an alternative to grid-sampling.

You can use the Metropolis (1953) - Hastings (1974) Algorithm:

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

where q is a "proposal distribution" which is not the real conditonal probability distribution and φ are tuning parameter(s) for that proposal distribution e.g. q = N($\theta_{2,t-1}$, 1)

(2) Calculate:

$$r := \frac{\dfrac{P(\theta_{1,t}, \theta_{2,t}^{\text{proposal}}, \theta_{3,t}, \dots, \theta_{p,t} \mid X)}{q(\theta_{2,t}^{\text{proposal}}; \theta_{2,t-1}, \phi)}}{\dfrac{P(\theta_{1,t}, \theta_{2,t-1}, \theta_{3,t}, \dots, \theta_{p,t} \mid X)}{q(\theta_{2,t-1}; \theta_{2,t}^{\text{proposal}}, \phi)}}$$

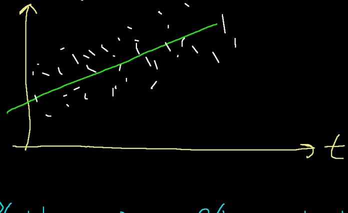
$\theta_{2,t-1} \to \theta_{2,t}^{\text{prop.}}$

$\theta_{2,t}^{\text{prop}} \to \theta_{2,t-1}$

Metropolis Ratio

The Metropolis-Hastings ratio is the same except it demands that the the two q values are the same and hence they cancel in the r expression.

(3) Sample u ~ U(0,1) and if u < r then you accept proposal and if not, you keep the t-1st value of θ_2.



$X_t \overset{ind}{\sim} \text{Poisson}(a + bt)$

$P(a) P(b) \propto 1$

$$P(a, b \mid X_1, \dots, X_n) \propto P(X_1, \dots, X_n \mid a, b)\, P(a, b)$$

$$\propto P(X_1, \dots X_n \mid a, b)$$

$$= \prod_{t=1}^{n} \frac{e^{-(a+bt)} (a+bt)^{x_i}}{x_i!}$$

$$\propto e^{-\sum_{i=1}^{n} a + bt} \prod_{i=1}^{n} (a+bt)^{x_i}$$

$$P(a \mid X_1 \dots X_n, b) \propto e^{-na} \prod_{i=1}^{n} (a+bt)^{x_i} \propto \; ?$$

$$P(b \mid X_1 \dots X_n, a) \propto e^{-bn\bar{t}} \prod_{i=1}^{n} (a+bt)^{x_i} \propto \; ?$$

Let the proposal distributions be normal centered at previous value and variance = 1. If you choose a bad variance then most of the proposals are rejected and you never move the chain.

Metropolis-within-Gibbs sampler