Bayesian Inference   A statistical model for text-message data.

Data: $C_t \equiv$ # texts on day t

- Let's model this by assuming each value of C_t is "drawn from" a probability distribution
- Natural choice for count data is the Poisson distribution

Imagine the day is chopped up into many time intervals Δt and in each interval there is a small constant probability for a single text to occur. Each time interval is a coin flip and the total count for the day is a binomial distribution. If let $\Delta t \to 0$ (Ie, the number of flips $n \to \infty$) and prob of heads $p \to 0$ such that $np \to \lambda$ (a constant) then we have a poisson distribution

$\Rightarrow$ Poisson is limit of Binomial (see derivation)

- C_t is "drawn from" a poisson distribution:

$$C_t \sim \mathrm{Pois}(\lambda) \Rightarrow \Pr(C_t) = \frac{\lambda^{C_t} e^{-\lambda}}{C_t!} \qquad E[C_t] = \lambda$$

λ is a parameter for the average rate of events

This is a probabilistic or statistical model for the data C_t.

Question: Does the level or volume or rate of text messages change (increase?) over the span of the data?

Let's answer this by extending our model to capture two rates.

Model: $C_t \sim \mathrm{Pois}(\lambda)$ but now λ is not constant.

Instead: $$\lambda = \begin{cases} \lambda_1 & \text{if } t < \tau \\ \lambda_2 & \text{if } t \geq \tau \end{cases}$$ λ changes from λ_1 to λ_2 on day τ.

$\Rightarrow$ 3 parameters: $\lambda_1, \lambda_2, \tau$

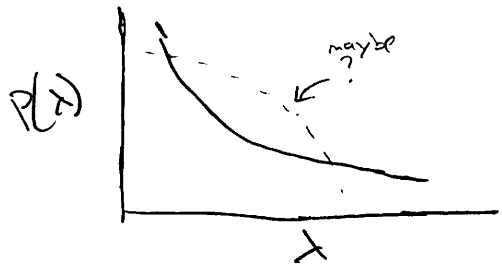
Bayesian Inference: Compare model to data by assuming prior distributions for $\lambda_1, \lambda_2, \tau$ and compute the posterior distribution for the model using Bayes Thm.

What priors?

First, the rate parameters λ_1, λ_2.

- We could assume all λ's are equally likely (uniform prior) but that's not "physical" in the sense that high $\lambda \rightarrow$ many text $\rightarrow$ much effort.

- Instead let's assume we initially think it less likely that the λ's are high and more likely they are low:



A convenient choice is an <u>exponential</u> <u>distribution</u>.

$$P(\lambda_i) = \alpha e^{-\alpha \lambda_i}, \quad \lambda_i \geq 0, \quad i \in \{1, 2\}$$

or $\lambda_1 \sim Exp(\alpha), \lambda_2 \sim Exp(\alpha)$

$\rightarrow$ We've introduced <u>another</u> parameter, α. This is a <u>hyperparameter</u> b/c it parameterizes the parameters of the model

$\rightarrow$ Need to choose either a distribution for α or a constant. Let's do the latter otherwise we'll be here all day. It turns out the particular value of α is not too important, but a good rule of thumb is $\alpha = E[C_t]^{-1}$. This keeps our prior "belief" in the rate centered on the average of the data.

$\rightarrow$ We could also choose $\alpha_1 \neq \alpha_2$ to reflect our current belief that $\lambda_1 \neq \lambda_2$.

Now what about τ?

- Let's assume the "break time" is equally likely at any value of t.

$$Pr(\tau = t) = \begin{cases} \frac{1}{T+1}, & 0 \leq t \leq T \\ 0, & \text{otherwise} \end{cases}$$ Discrete Uniform distribution

or $\tau \sim U[0, T]$ (sometimes written as $\tau \sim \text{DiscreteUniform}[0, T]$)

Now to compute the <u>posterior</u> we need to figure out some potentially <u>nasty</u> distributions:

$$\Pr(\text{data}|\text{model}) \longleftrightarrow \Pr(\{C_t\}|\lambda_1,\lambda_2,\tau)$$

and $$\int \Pr(\text{data}|\text{model})\cdot\Pr(\text{model})\, \text{"}d(\text{model})\text{"}$$

Before we get to true inference (using the posterior(s)) let's make some graphics to understand the priors and posterior $\propto$ likelihood $\cdot$ prior.

[slides]