Data 102 Spring 2022
Lecture 7

Graphical Models

Weekly Outline

- Last week: Frequentist vs Bayesian, intro to Bayesian modeling

- **Today: Graphical Models**
  - Setting up and understanding more interesting/complicated Bayesian models

- Next time: How do we figure unknowns in more complicated models?

Announcements

- Lab 0 grades are out!
  - Please carefully read the Ed post
  - Regrades are due by 2/16: **no exceptions!**
- Updates on Lab deadline policy
  - From Lab 2 (this week) onwards, all students can submit any lab assignment 24 hours after the deadline without penalty
  - Please don’t use this as an excuse to procrastinate on the lab!
- HW2 released: please start early!

Recap: Statistical modeling

- Goal: find unknown parameter θ using observed data x
- How:
  - Define a probability model for the data/parameters, then use it to estimate θ from x
- Likelihood function p(x|θ): captures how likely our data are for each parameter
  - Used in both frequentist and Bayesian models
- Frequentist modeling
  - MLE (Maximum Likelihood Estimate): value of θ that makes p(x|θ) as large as possible
- Bayesian modeling
  - Define a prior p(θ): what we believe about the parameter before we see any data
  - Compute posterior p(θ|x): what we believe about the parameter after observing data
  - To get a single estimate for θ from the posterior, we can use the MAP or LMSE estimates
  - MAP: value of θ that makes p(θ|x) as large as possible
  - *LMSE:* $E_{\theta|x}[\theta]$ (*expectation of θ according to posterior p(θ|x)*)

Estimating the Mean with a Gaussian Likelihood

Heights: $x_1, \ldots, x_n$

Want: μ (pop. mean)

Bayesians want

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

$$\propto p(x|\theta)p(\theta)$$

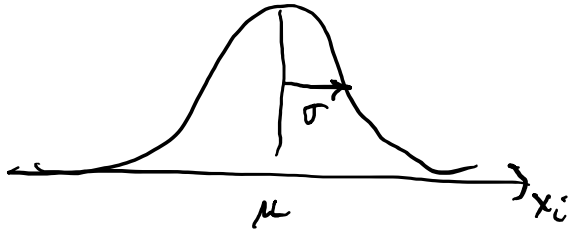
If likelihood and prior for mean are both normal, posterior is also normal.

$\theta|x \sim N(\_\_\_, \_\_\_)$

Normal likelihood: $p(x_i|\mu) = \mathcal{N}(x_i; \mu, \sigma)$

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma}(x_i-\mu)^2}$$

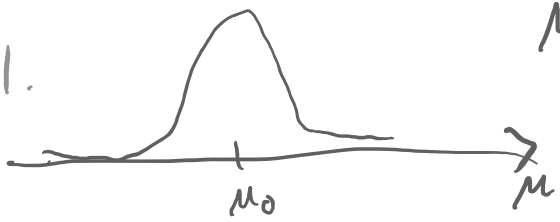
$\sigma = 1$



Normal prior: $p(\mu) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2\sigma_0}(\mu-\mu_0)^2}$

$\mu_0 = 5'5''$

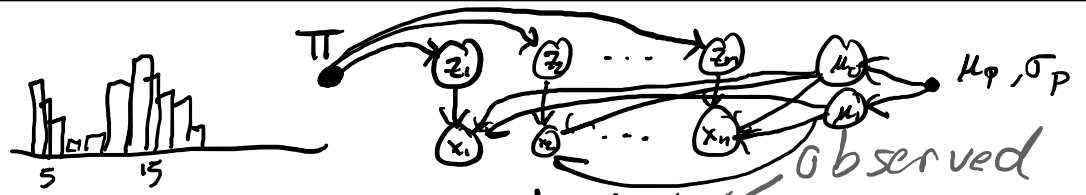
$\sigma_0 = 1''$



Conjugate Priors

Prior	Likelihood	Posterior
Beta	Bernoulli	Beta
Normal	Normal	Normal

Exoplanet Model



$x_1, \ldots, x_n$: x_i is radius of exoplanet i

$z_1, \ldots, z_n$: z_i is $\begin{cases} 0 & \text{if exoplanet } i \text{ is small} \\ 1 & \text{if big} \end{cases}$ ← unknown

μ_0, σ_0: Mean & SD for "small" planets ← unknown

μ_1, σ_1: Mean & SD for "big" planets

fix

posterior $p(\theta \mid x)$ (z, μ)

$z_i \sim \text{Ber}(\pi)$: Prior for unknown z

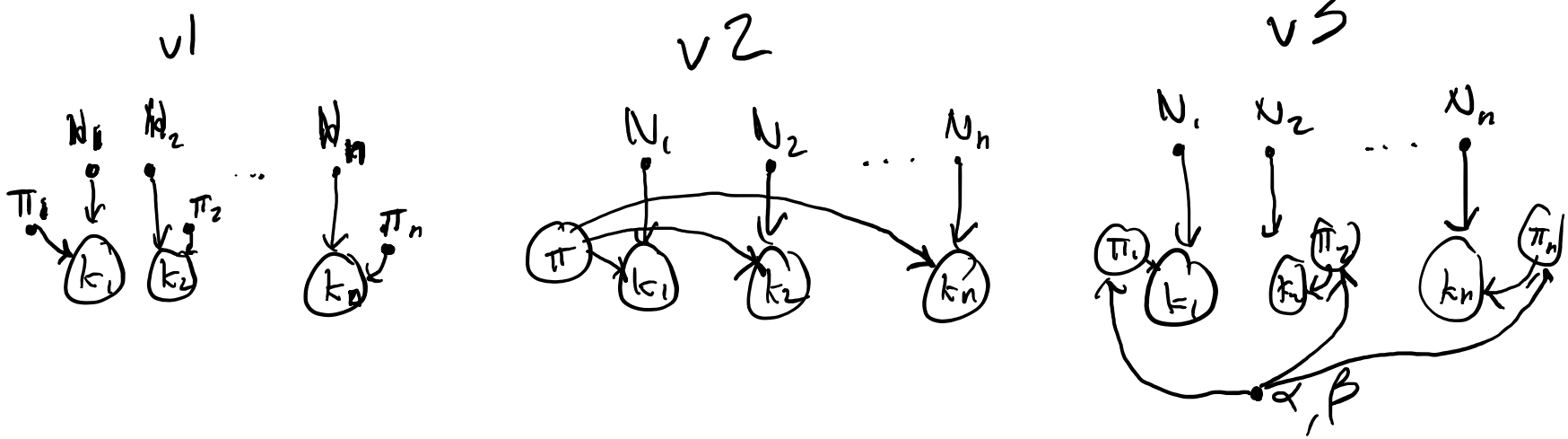
$\mu_k \sim \mathcal{N}(\mu_p, \sigma_p)$, $k=0,1$: Prior for unknown μ_0, μ_1

$p(z_{1:n}, \mu_0, \mu_1 \mid x_{1:n})$

$x_i \mid z_i, \mu_0, \mu_1 \sim \mathcal{N}(\mu_{z_i}, \sigma)$: likelihood

A Graphical Model for COVID Meta-Analysis

- Question: if someone has COVID, how likely are they to transmit it to someone else in their household?
- Goal: Use data from multiple studies to get a better estimate
- Definitions:
  - Secondary Attack Rate (SAR, **π**): Probability of transmitting COVID to someone in household
  - N_i: Number of participants in study i (all sharing households w/COVID-positive people)
  - k_i: How many of them test positive



Graphical Models Recap

Conditional Independence