

Statistical Model

Imitates + approximates data generating process

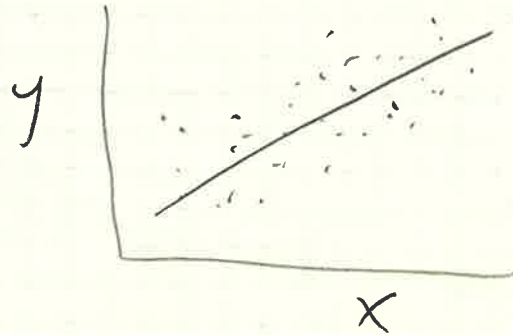
- ① Quantify uncertainty
- ② Inference
- ③ Measure support for hypotheses

55% women favor candidate
59% men favor candidate

- ④ Prediction

Statistical Modeling Process

- ① Understand the problem
- ② Plan and collect data
- ③ Explore data
- ④ Postulate model
- ⑤ Fit model
- ⑥ Check model
- ⑦ Iterate (if needed)
- ⑧ Use model



Components of Bayesian Modelsheights $n=15$ men

$$y_i = \mu + \varepsilon_i, \quad \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2), \quad i=1, \dots, n$$

$$y_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$$

likelihood $p(y|\theta)$

$$p(y, \theta) = p(\theta) p(y|\theta)$$

prior $p(\theta)$

$$\text{posterior } p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta, y)}{\int p(\theta, y) d\theta} = \frac{p(y|\theta) p(\theta)}{\int p(y|\theta) p(\theta) d\theta}$$

Model Specification

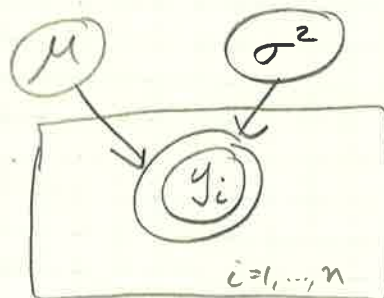
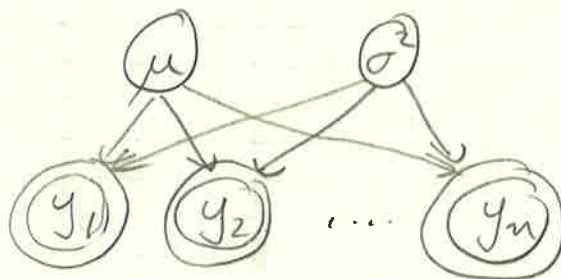
$$y_i | \mu, \sigma^2 \stackrel{iid}{\sim} N(\mu, \sigma^2), \quad i=1, \dots, n$$

$$p(\mu, \sigma^2) = p(\mu) p(\sigma^2) \quad (\text{independence})$$

$$\mu \sim N(\mu_0, \sigma_0^2)$$

$$\sigma^2 \sim \text{IG}(\nu_0, \beta_0)$$

Inverse gamma



y 's must
be exchangeable

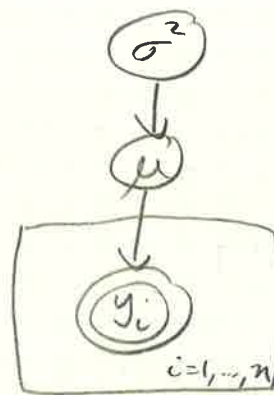
Posterior Derivation

$$y_i | \mu, \sigma^2 \stackrel{iid}{\sim} N(\mu, \sigma^2)$$

$$\mu | \sigma^2 \sim N(\mu_0, \frac{\sigma^2}{\omega_0})$$

$$\sigma^2 \sim IG(\nu_0, \beta_0)$$

$$p(\theta | y) = \frac{p(y | \theta) p(\theta)}{\int p(y | \theta) p(\theta) d\theta}$$



$$p(y_1, \dots, y_n, \mu, \sigma^2) = p(y_1, \dots, y_n | \mu, \sigma^2) p(\mu, \sigma^2) p(\sigma^2)$$

$$= \prod_{i=1}^n [N(y_i | \mu, \sigma^2) N(\mu | \mu_0, \frac{\sigma^2}{\omega_0})] IG(\sigma^2 | \nu_0, \beta_0)$$

$$\propto p(\mu, \sigma^2 | y_1, \dots, y_n)$$

Non-conjugate Models

$$n=10$$

$$y_i | \mu \stackrel{iid}{\sim} N(\mu, 1)$$

$$\mu \sim t(0, 1, \frac{1}{2})$$

\uparrow \uparrow \uparrow
 location scale df

$$p(\mu | y_1, \dots, y_n) \propto \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i - \mu)^2\right) \right] \frac{1}{\pi(1+\mu^2)}$$

$$\propto \exp\left[-\frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2\right] \frac{1}{1+\mu^2}$$

$$\propto \exp\left[-\frac{1}{2} \left(\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2 \right)\right] \frac{1}{1+\mu^2}$$

$$\propto \exp\left[n(\bar{y}\mu - \mu^2/2)\right] \frac{1}{1+\mu^2}$$

$$y_i | \mu, \sigma^2 \stackrel{iid}{\sim} N(\mu, \sigma^2), i=1, \dots, n$$

$$\mu \sim N(\mu_0, \sigma_0^2)$$

$$\sigma^2 \sim IG(\nu_0, \beta_0)$$

Monte Carlo Integration

$$\Theta \sim \text{Ga}(a, b)$$

$$a=2, \quad b=\frac{1}{3}$$

$$E(\Theta) = \int_0^{\infty} \theta p(\theta) d\theta = \int_0^{\infty} \theta \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} d\theta = \frac{a}{b}$$

$$\Theta_i^* \quad i=1, \dots, m$$

$$\bar{\Theta}^* = \frac{1}{m} \sum_{i=1}^m \Theta_i^*$$

$$\text{Var}(\Theta) = \int_0^{\infty} (\theta - E(\Theta))^2 p(\theta) d\theta$$

Consider $h(\theta)$

$$\int h(\theta) p(\theta) d\theta = E[h(\theta)] \approx \frac{1}{m} \sum_{i=1}^m h(\Theta_i^*)$$

Example $h(\theta) = \mathbb{I}_{\theta < 5}(\theta)$

$$\begin{aligned} E(h(\theta)) &= \int_0^{\infty} \mathbb{I}_{\theta < 5}(\theta) p(\theta) d\theta \\ &= \int_0^5 1 \cdot p(\theta) d\theta + \int_5^{\infty} 0 \cdot p(\theta) d\theta \\ &= \Pr[0 < \theta < 5] \\ &\approx \frac{1}{m} \sum_{i=1}^m \mathbb{I}_{\Theta_i^* < 5}(\Theta_i^*) \end{aligned}$$

Monte Carlo Error and Marginalization

$$\bar{\Theta}^* \sim N(E(\Theta), \frac{\text{Var}(\Theta)}{m})$$

$$\widehat{\text{Var}}(\Theta) = \frac{1}{m} \sum_{i=1}^m (\Theta_i^* - \bar{\Theta}^*)^2 \quad \sqrt{\frac{\widehat{\text{Var}}(\Theta)}{m}} = \text{standard error}$$

$$y|\phi \sim \text{Bin}(10, \phi) \quad p(y, \phi) = p(\phi) p(y|\phi)$$

$$\phi \sim \text{Beta}(2, 2)$$

Simulate: ① ϕ_i^* from Beta

② given ϕ_i^* , draw $y_i^* \sim \text{Bin}(10, \phi_i^*)$

(y_i^*, ϕ_i^*) drawn from $p(y, \phi)$

If we discard the ϕ_i^* , then we draw y_i^* from its marginal dist

Algorithm

$p(\theta) \propto q(\theta)$
Metropolis-Hastings

① Select initial value θ_0

② For $i=1, \dots, m$, repeat:

a) Draw candidate $\theta^* \sim q(\theta^* | \theta_{i-1})$

$$b) \alpha = \frac{q(\theta^*) / q(\theta_{i-1} | \theta^*)}{q(\theta_{i-1}) / q(\theta_{i-1} | \theta^*)} = \frac{q(\theta^*) q(\theta_{i-1} | \theta^*)}{q(\theta_{i-1}) q(\theta^* | \theta_{i-1})}$$

c) $\alpha \geq 1$: accept θ^* and set $\theta_i \leftarrow \theta^*$

$0 < \alpha < 1$: $\begin{cases} \text{accept } \theta^* \text{ and set } \theta_i \leftarrow \theta^* \text{ with prob } \alpha \\ \text{reject } \theta^* \text{ and set } \theta_i \leftarrow \theta_{i-1} \text{ with prob } 1-\alpha \end{cases}$

If q is a symmetric distribution $q(\theta_{i-1} | \theta^*) = q(\theta^* | \theta_{i-1})$

$$\Leftrightarrow \alpha = \frac{q(\theta^*)}{q(\theta_{i-1})}$$

Demonstration

$$\Theta = \{\text{fair, loaded}\}$$

$$\text{Prior } P(\Theta = \text{loaded}) = 0.6$$

$$\text{likelihood } f(x|\theta) = \binom{5}{x} \left(\frac{1}{2}\right)^5 \mathbb{I}_{\{\theta = \text{fair}\}} + \binom{5}{x} (.7)^x (.3)^{5-x} \mathbb{I}_{\{\theta = \text{loaded}\}}$$

$$\text{Posterior } f(\theta | X=2) = \frac{f(X=2|\theta) f(\theta)}{f(X)}$$

$$= \frac{\left(\frac{1}{2}\right)^5 (.4) \mathbb{I}_{\{\theta = \text{fair}\}} + (.7)^2 (.3)^3 (.6) \mathbb{I}_{\{\theta = \text{loaded}\}}}{\left(\frac{1}{2}\right)^5 (.4) + (.7)^2 (.3)^3 (.6)}$$

$$= \frac{0.0125 \mathbb{I}_{\{\theta = \text{fair}\}} + 0.00794 \mathbb{I}_{\{\theta = \text{loaded}\}}}{0.0125 + 0.00794}$$

$$= 0.612 \mathbb{I}_{\{\theta = \text{fair}\}} + 0.388 \mathbb{I}_{\{\theta = \text{loaded}\}}$$

$$\Rightarrow P(\Theta = \text{loaded} | X=2) = 0.388$$

1) Start at either $\theta_0 = \text{fair}$ or $\theta_0 = \text{loaded}$

2) For $i=1, \dots, m$

a) Propose candidate θ^* to be the other state θ_{i-1}

$$b) \alpha = \frac{q(\theta^*) / q(\theta^* | \theta_{i-1})}{q(\theta_{i-1}) / q(\theta_{i-1} | \theta^*)}$$

$$= \frac{f(X=2 | \theta^*) f(\theta^*) / 1}{f(X=2 | \theta_{i-1}) f(\theta_{i-1}) / 1}$$

q is a deterministic proposal in this case

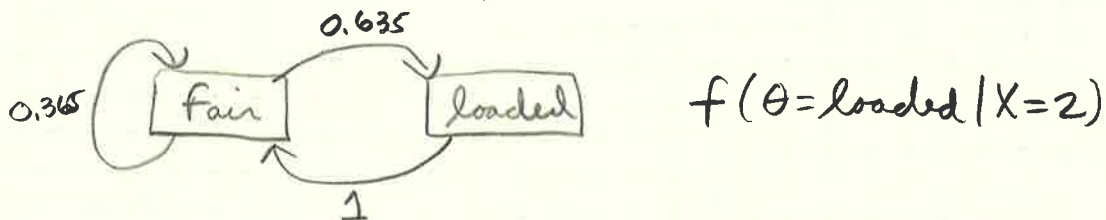
If $\theta^* = \text{loaded}$ $\alpha = \frac{0.00794}{0.0125} = 0.635$

If $\theta^* = \text{fair}$ $\alpha = \frac{0.0125}{0.00794} = 1.574$

c) If $\theta^* = \text{fair}$, $\alpha > 1$ and accept θ^* and set $\theta_i = \text{fair}$

If $\theta^* = \text{loaded}$, $\alpha = 0.635$, and accept θ^* w.p. 0.635 and set $\theta_i = \text{loaded}$

Otherwise, set $\theta_i = \theta_{i-1} = \text{fair}$



Markov chain: π is stationary if $\pi P = \pi$ for transition probability P

$$P = \begin{bmatrix} \text{fair} & \text{loaded} \\ 0.365 & 0.635 \\ 1 & 0 \end{bmatrix} \begin{matrix} \text{fair} \\ \text{loaded} \end{matrix}$$

$$\pi = [0.612 \quad 0.388]$$

$$[0.612 \quad 0.388] \begin{bmatrix} 0.365 & 0.635 \\ 1 & 0 \end{bmatrix} = [0.612 \quad 0.388]$$

Multiple parameter sampling and full conditional distributions

$$p(\theta, \phi | y) \propto q(\theta, \phi)$$

$$p(\theta, \phi | y) = \underbrace{p(\phi | y)}_{\text{no } \theta} p(\theta | \phi, y) \quad \uparrow \text{full conditional dist for } \theta$$

$$p(\theta | \phi, y) \propto p(\theta, \phi | y) \propto q(\theta, \phi)$$

$$p(\phi | \theta, y) \propto p(\theta, \phi | y)$$

Gibbs Sampling Algorithm

① Initialize θ_0, ϕ_0

② For $i=1, \dots, m$, repeat:

a) Using ϕ_{i-1} , draw $\theta_i \sim p(\theta | \phi_{i-1}, y)$

b) Using θ_i , draw $\phi_i \sim p(\phi | \theta_i, y)$
get (θ_i, ϕ_i)

Conditionally conjugate prior example with Normal likelihood

$$y_i | \mu, \sigma^2 \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2), \quad i=1, \dots, n$$

$$\mu \sim N(\mu_0, \sigma_0^2)$$

$$\sigma^2 \sim \text{IG}(\nu_0, \beta_0)$$

$$p(\mu, \sigma^2 | y_1, \dots, y_n) \propto p(y_1, \dots, y_n | \mu, \sigma^2) p(\mu) p(\sigma^2)$$

$$= \prod_{i=1}^n [N(y_i | \mu, \sigma^2)] N(\mu | \mu_0, \sigma_0^2) \text{IG}(\nu_0, \beta_0)$$

$$= \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right) \right] \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right) \frac{\beta_0^{\nu_0}}{\Gamma(\nu_0)} (\sigma^2)^{-(\nu_0+1)} \exp\left(-\frac{\beta_0}{\sigma^2}\right)$$

$$\propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right] \exp\left[-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right] (\sigma^2)^{-(\nu_0+1)} \exp\left[-\frac{\beta_0}{\sigma^2}\right]$$

$$p(\mu | \sigma^2, y_1, \dots, y_n) \propto p(\mu, \sigma^2 | y_1, \dots, y_n)$$

$$\propto \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right] \exp\left[-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right]$$

$$= \exp\left[-\frac{1}{2} \left(\frac{\sum (y_i - \mu)^2}{\sigma^2} + \frac{(\mu - \mu_0)^2}{\sigma_0^2} \right)\right]$$

$$\propto N\left(\mu \mid \frac{n\bar{y}/\sigma^2 + \mu_0/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2}, \frac{1}{n/\sigma^2 + 1/\sigma_0^2}\right)$$

$$\begin{aligned}
 p(\sigma^2 | \mu, y_1, \dots, y_n) &\propto p(\mu, \sigma^2 | y_1, \dots, y_n) \\
 &\propto (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum (y_i - \mu)^2\right] (\sigma^2)^{-(\nu_0+1)} \exp\left[-\frac{\beta_0}{\sigma^2}\right] \\
 &= (\sigma^2)^{-(\nu_0 + n/2 + 1)} \exp\left[-\frac{1}{\sigma^2} \left(\beta_0 + \frac{\sum_{i=1}^n (y_i - \mu)^2}{2}\right)\right] \\
 &\propto \text{IG}\left(\sigma^2 \mid \underbrace{\nu_0 + \frac{n}{2}}_{\text{shape}}, \underbrace{\beta_0 + \frac{\sum_{i=1}^n (y_i - \mu)^2}{2}}_{\text{scale}}\right)
 \end{aligned}$$

no notes for Lesson 6

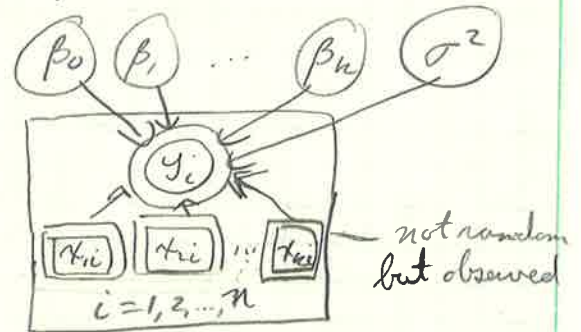
$$y_i = \underbrace{\beta_0 + \beta_1 x_{1i} + \dots + \beta_n x_{ni}}_{\text{mean}} + \underbrace{\varepsilon_i}_{\text{error}}, \quad \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2), i=1, \dots, n$$

$$y_i | x_i, \beta, \sigma^2 \stackrel{\text{i.i.d.}}{\sim} N(\beta_0 + \beta_1 x_{1i} + \dots + \beta_n x_{ni}, \sigma^2)$$

$$\beta_0 \sim p(\beta_0), \quad \beta_i \sim p(\beta_i), \dots$$

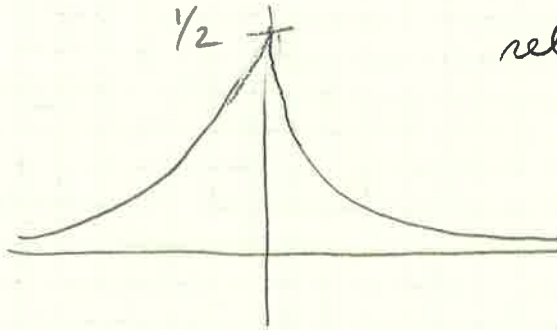
$$\sigma^2 \sim p(\sigma^2) \quad (\text{maybe IG})$$

usually normal or multivariate normal



Laplace prior

$$p(\beta) = \frac{1}{2} e^{-|\beta|}$$



related to the LASSO

Factors: sound

Levels: music
no music

font size

small
medium
large

$g_i \sim$ group of subject i
 $\underline{\mu} \sim$ vector of G group means

$$y_i | g_i, \underline{\mu}, \sigma^2 \overset{\text{ind}}{\sim} N(\mu_{g_i}, \sigma^2)$$

$$g_i \in \{1, \dots, G\}, i = 1, \dots, n$$

Alternative

$$E(y_i) = \beta_0 + \beta_1 \underset{\substack{\uparrow \\ I(g_i=1)}}{x_{1,i}} + \dots + \beta_{G-1} \underset{\substack{\uparrow \\ I(g_i=G-1)}}{x_{G-1,i}}$$

x 's are indicators or dummy variables

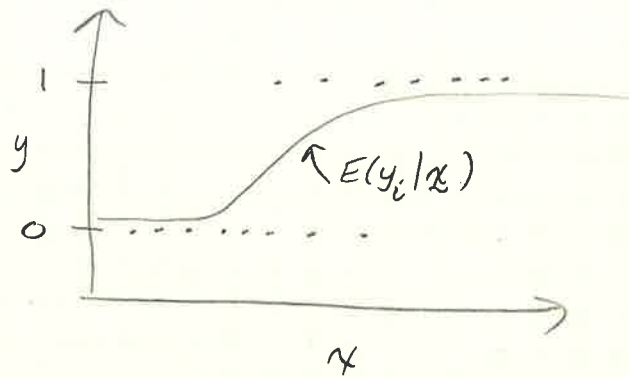
		B		
		1	2	3
A	1	$\mu_{1,1}$	$\mu_{1,2}$	$\mu_{1,3}$
	2	$\mu_{2,1}$	$\mu_{2,2}$	$\mu_{2,3}$

$A, B \sim$ Factors

6 treatment groups

Additive model: (good for no interactions between factors)

$$E(y_i) = \mu + \alpha_2 I(a_i = 2) + \beta_2 I(b_i = 2) + \beta_3 I(b_i = 3)$$



$y_i | \phi_i \overset{\text{ind}}{\sim} \text{Bernoulli}(\phi_i)$, $i=1, \dots, n$
 In linear regression: $E(y_i) = \beta_0 + \beta_1 x_{1i}$

ϕ is probability of success

$$\frac{\phi}{1-\phi} \Leftrightarrow \text{odds}$$

$$\log\left(\frac{\phi}{1-\phi}\right) \Leftrightarrow \text{logit link}$$

$$\text{logit}(\phi_i) = \log\left(\frac{\phi_i}{1-\phi_i}\right) = \beta_0 + \beta_1 x_{1i}$$

$$E(y_i) = \phi_i = \frac{\exp(\beta_0 + \beta_1 x_{1i})}{1 + \exp(\beta_0 + \beta_1 x_{1i})} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1i})}}$$

$$y_i | x_i, \beta \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{1}{1 + \exp[-\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}]}\right)$$

$$y_i | \lambda_i \stackrel{\text{ind}}{\sim} \text{Pois}(\lambda_i), \quad i=1, \dots, n$$

linear regression: $E(y_i) = \beta_0 + \beta_1 x_{1i}$ is a problem

$$\text{log link: } \log(\lambda_i) = \beta_0 + \beta_1 x_{1i}$$

$$\Rightarrow E(y_i) = \lambda_i = e^{\beta_0 + \beta_1 x_{1i}}$$

λ_i is the mean
and variance
of the poisson distribution

$$E(\log(y)) \neq \log(E(y))$$

$$y_i | x_i, \beta \stackrel{\text{ind}}{\sim} \text{Pois}(\exp[\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}])$$