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

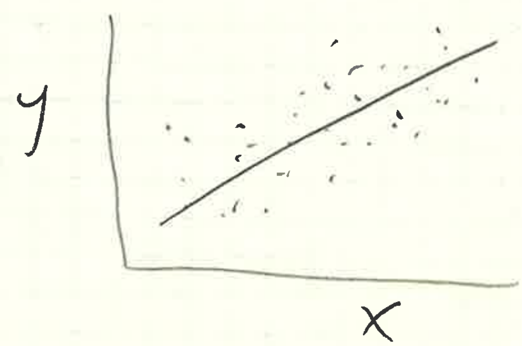
④ Postulate model

⑤ Fit model

⑥ Check model

⑦ Iterate (if needed)

⑧ Use model



Components of Bayesian Models

heights $n=15$ men

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

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

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

prior $p(\theta)$

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

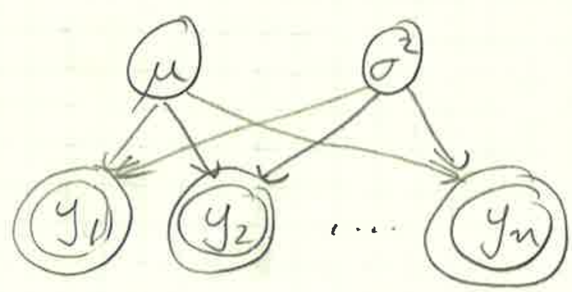
Model Specification

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

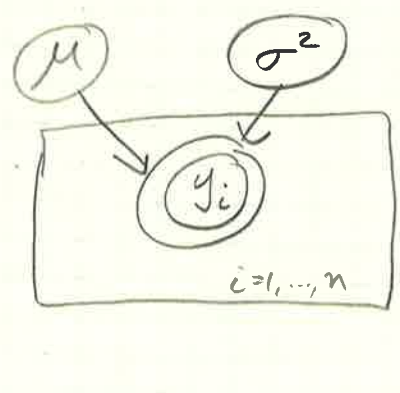
$p(\mu,\sigma^2) = p(\mu)\,p(\sigma^2)$ (independence)

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

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



$\Updownarrow$



y's must be exchangeable

STAEDTLER No. 937 811E Engineer's Computation Pad

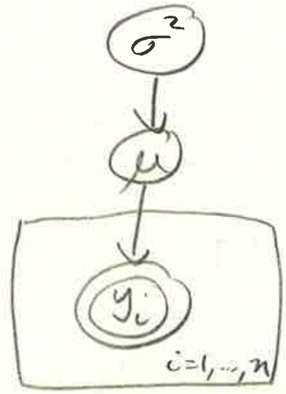
STAEDTLER® No. 937 811E Engineer's Computation Pad

Posterior Derivation

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

$$\mu \mid \sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{w_0}\right)$$

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



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

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

$$= \prod_{i=1}^{n} \left[N(y_i \mid \mu, \sigma^2) \right] N\left(\mu \mid \mu_0, \frac{\sigma^2}{w_0}\right) IG(\sigma^2 \mid \nu_0, \beta_0)$$

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

Non-conjugate Models

$$n = 10$$

$$y_i \mid \mu \overset{iid}{\sim} N(\mu, 1)$$

$$\mu \sim t(\underbrace{0}_{\text{location}}, \underbrace{1}_{\text{scale}}, \underbrace{1}_{\text{df}})$$

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

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

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

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

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

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

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

Monte Carlo Integration

$\theta \sim Ga(a, b)$

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

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

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

$$\bar{\theta}^* = \frac{1}{m}\sum_{i=1}^{m} \theta_i^*$$

$$Var(\theta) = \int_0^\infty (\theta - E(\theta))^2 p(\theta)\, d\theta$$

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 $h(\theta) = I_{\theta<5}(\theta)$

$$E(h(\theta)) = \int_0^\infty I_{\theta<5}(\theta)\, p(\theta)\, d\theta$$

$$= \int_0^5 1 \cdot p(\theta) d\theta + \int_5^\infty 0 \cdot p(\theta) d\theta$$

$$= Pr[0 < \theta < 5]$$

$$\approx \frac{1}{m}\sum_{i=1}^{m} I_{\theta^*<5}(\theta_i^*)$$

Monte Carlo Error and Marginalization

$$\bar{\theta}^* \stackrel{\cdot}{\sim} N\left(E(\theta), \frac{Var(\theta)}{m}\right)$$

$$\widehat{Var(\theta)} = \frac{1}{m}\sum_{i=1}^{m} (\theta_i^* - \bar{\theta}^*)^2 \qquad \sqrt{\frac{\widehat{Var(\theta)}}{m}} = \text{standard error}$$

$y|\phi \sim Bin(10, \phi) \qquad p(y,\phi) = p(\phi)\, p(y|\phi)$

$\phi \sim Beta(2,2)$

Simulate: ① ϕ_i^* from Beta

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

(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

$$p(\theta) \propto g(\theta)$$

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

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

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

Demonstration

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

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

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

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

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

$$= \frac{0.0125\, I_{\{\theta=\text{fair}\}} + 0.00794\, I_{\{\theta=\text{loaded}\}}}{0.0125 + 0.00794}$$

$$= 0.612\, I_{\{\theta=\text{fair}\}} + 0.388\, I_{\{\theta=\text{loaded}\}}$$

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

STAEDTLER No. 937 811E Engineer's Computation Pad

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

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

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

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

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

q is a deterministic proposal in this case

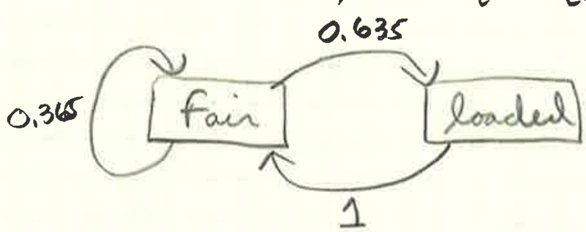
If $\theta^* = $ loaded $\quad \alpha = \frac{0.00794}{0.0125} = 0.635$

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

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

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

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



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

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

$$P = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} \text{fair} & \text{loaded} \\ \left[\begin{array}{cc} 0.365 & 0.635 \\ 1 & 0 \end{array}\right] \end{array} \begin{array}{l} \text{— fair} \\ \text{— loaded} \end{array}$$

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

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

Multiple parameter sampling and full conditional distributions

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

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

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

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

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

$$y_i|\mu,\sigma^2 \overset{iid}{\sim} N(\mu,\sigma^2), \quad i=1,\dots,n$$
$$\mu \sim N(\mu_0,\sigma_0^2)$$
$$\sigma^2 \sim IG(\nu_0,\beta_0)$$

$$p(\mu,\sigma^2|y_1,\dots,y_n) \propto p(y_1,\dots,y_n|\mu,\sigma^2)\,p(\mu)\,p(\sigma^2)$$
$$= \prod_{i=1}^{n}\left[N(y_i|\mu,\sigma^2)\right] N(\mu|\mu_0,\sigma_0^2)\, IG(\nu_0,\beta_0)$$
$$= \prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-1}{2\sigma^2}(y_i-\mu)^2\right)\right]\frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left(\frac{-1}{2\sigma_0^2}(\mu-\mu_0)^2\right)\frac{\beta_0^{\nu_0}}{\Gamma(\nu_0)}(\sigma^2)^{-(\nu_0+1)}\exp\left(\frac{-\beta_0}{\sigma^2}\right)$$
$$\propto (\sigma^2)^{-n/2}\exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]\exp\left[\frac{-1}{2\sigma_0^2}(\mu-\mu_0)^2\right](\sigma^2)^{-(\nu_0+1)}\exp\left[\frac{-\beta_0}{\sigma^2}\right]$$

$$p(\mu|\sigma^2,y_1,\dots,y_n) \propto p(\mu,\sigma^2|y_1,\dots,y_n)$$
$$\propto \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]\exp\left[\frac{-1}{2\sigma_0^2}(\mu-\mu_0)^2\right]$$
$$= \exp\left[-\frac{1}{2}\left(\frac{\sum(y_i-\mu)^2}{\sigma^2}+\frac{(\mu-\mu_0)^2}{\sigma_0^2}\right)\right]$$
$$\propto N\left(\mu \,\middle|\, \frac{n\bar{y}/\sigma^2+\mu_0/\sigma_0^2}{n/\sigma^2+1/\sigma_0^2},\ \frac{1}{n/\sigma^2+1/\sigma_0^2}\right)$$

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

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

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

$$\propto \mathrm{IG}\left(\sigma^2 \,\middle|\, \underbrace{\nu_0 + \frac{n}{2}}_{\text{shape}}, \underbrace{\beta_0 + \frac{\sum_{i=1}^n (y_i - \mu)^2}{2}}_{\text{scale}}\right)$$

no notes for Lesson 6

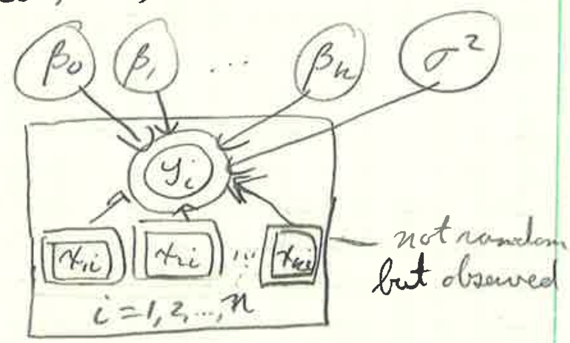
$$y_i = \underbrace{\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}}_{\text{mean}} + \underbrace{\varepsilon_i}_{\text{error}}, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2), \; i = 1, \dots, n$$

$$y_i \mid \underset{\sim}{x}_i, \underset{\sim}{\beta}, \sigma^2 \overset{ind}{\sim} N(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}, \sigma^2)$$

$\beta_0 \sim p(\beta_0)$, $\beta_1 \sim p(\beta_1)$, ...

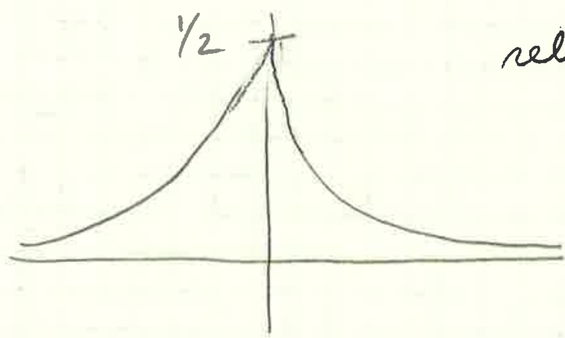
$\sigma^2 \sim p(\sigma^2)$ (maybe IG)

usually normal or multivariate normal



Laplace prior

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



related to the LASSO

Factors:	sound	font size
Levels	music no music	small medium large

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

$$y_i \mid g_i, \underset{\sim}{\mu}, \sigma^2 \overset{ind}{\sim} N(\mu_{g_i}, \sigma^2)$$
$$g_i \in \{1, \dots, G\} \quad , i = 1, \dots, n$$

Alternative

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

x's are indicators or dummy variables

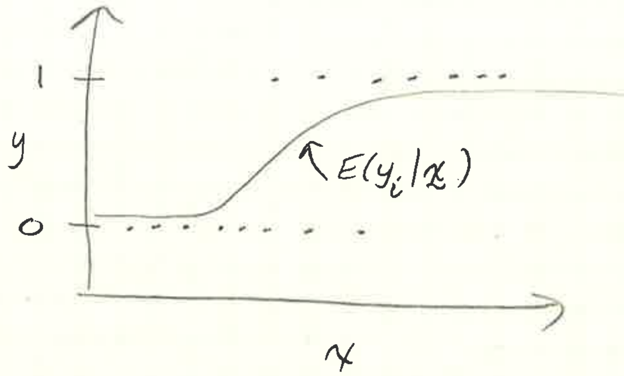
A, B ~ Factors

$$\begin{array}{cc|ccc} & & & B & \\ & & 1 & 2 & 3 \\ \hline A & 1 & \mu_{1,1} & \mu_{1,2} & \mu_{1,3} \\ & 2 & \mu_{2,1} & \mu_{2,2} & \mu_{2,3} \end{array}$$

6 treatment groups

Additive model: (good for no interactions between factors)

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



$y_i | \phi_i \overset{ind}{\sim} \text{Bernoulli}(\phi_i)$, $i = 1, \dots, n$ $\qquad$ ϕ is probability of success

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

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

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

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

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

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

STAEDTLER® No. 937 811E Engineer's Computation Pad

$y_i \mid \lambda_i \overset{ind}{\sim} Pois(\lambda_i)$, $i = 1, \ldots, n$

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

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

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

$y_i \mid x_i, \underline{\beta} \overset{ind}{\sim} Pois\left(\exp[\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}]\right)$

Correlated Data

150 test cookies (to count # of chocolate chips in each cookie)
30 from each of 5 locations

fully independent model:

$$y_i \mid \lambda \overset{iid}{\sim} \mathrm{Pois}(\lambda) \quad , \; i = 1, \dots, 150$$

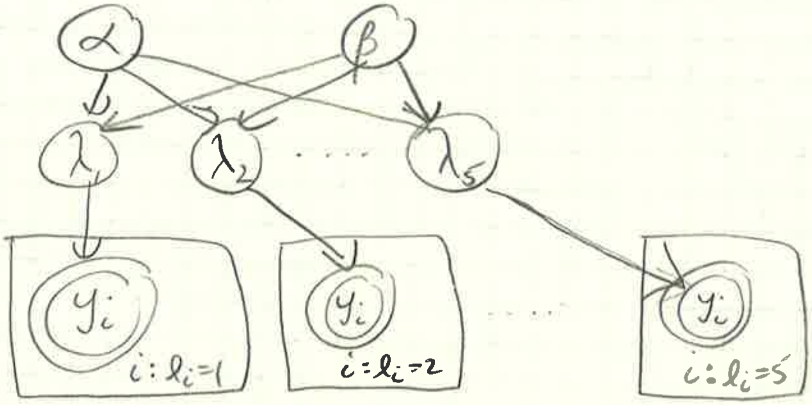
location dependent:

$$y_i \mid l_i, \lambda_{l_i} \overset{ind}{\sim} \mathrm{Pois}(\lambda_{l_i}) \quad , \; l_i \in \{1, \dots, 5\} \quad (\text{locations})$$

$$i = 1, \dots, 150$$

$$\lambda_l \mid \alpha, \beta \overset{iid}{\sim} \mathrm{Ga}(\alpha, \beta) \quad , \; l = 1, \dots, 5$$

$$\alpha \sim p(\alpha) \quad , \; \beta \sim p(\beta) \qquad [\text{priors}]$$



Example

$$y_i \mid g_i, \mu_{g_i}, \sigma^2 \overset{ind}{\sim} N(\mu_{g_i}, \sigma^2) \quad , \; g_i \in \{1, 2, 3\} \quad (3 \text{ groups})$$

$$\mu_g \overset{iid}{\sim} N(\eta, \tau^2) \; , \; g = 1, 2, 3$$

$$\eta \sim p(\eta), \; \tau^2 \sim p(\tau^2)$$

STAEDTLER No. 937 811E Engineer's Computation Pad

Linear Regression Example

linear model: $y_i \mid \underset{\sim}{x}_i, \underset{\sim}{\beta}, \sigma^2 \overset{ind}{\sim} N(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}, \sigma^2)$, $i = 1, \ldots, n$

random intercept:

$$y_i \mid r_i, \underset{\sim}{x}_i, \underset{\sim}{\alpha}, \underset{\sim}{\beta}, \sigma^2 \overset{ind}{\sim} N(\alpha_{r_i} + \beta_1 x_{1i} + \beta_2 x_{2i}, \sigma^2), \quad r_i \in \{1, \ldots, R\}$$

R = # of regions $\qquad i = 1, \ldots, n$

$$\alpha_r \mid \mu, \tau^2 \overset{iid}{\sim} N(\mu, \tau^2), \quad r = 1, \ldots, R$$