this morning time: | optional lecture next Tue 1:30-3:05 pm (usual room) (webcast) (?) | AMS 206 15 Mar 18

why Metropolis sampling works; Bayesian data science (1)

$(Y_i \mid \lambda_i\, \mathcal{B}) \stackrel{I}{\sim} \text{Poisson}(\lambda_i)$
$(i = 1, \dots, n)$

$\log \lambda_i = \underline{\gamma_0} + \underline{\gamma_1}\, x_i$ $\begin{pmatrix} 1 & \text{if } T/E \\ 0 & c \end{pmatrix}$

$(\gamma_0, \gamma_1) \sim$ diffuse

fixed-effects Poisson regression model (FEPR)

$(Y_i \mid \lambda_i\, \mathcal{B}) \stackrel{I}{\sim} \text{Poisson}(\lambda_i)$

age

$\log \lambda_i = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \gamma_3 x_{i3} + \dots + \gamma_k x_{ik}$

random-effects Poisson reg. model (REPR)

sickness at beginning of experiment

e_i

$\text{IID } N(0, \underline{\underline{\sigma_e^2}})$

model complexity is somewhere between 3 & (572 + 3).

$\boxed{\text{DIC}}$ p_D

use data to estimate

(unobserved) → latent variables

random effects

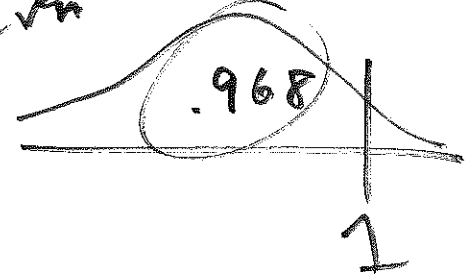
This REPR model is also a lognormal <u>mixture</u> of Poissons; an alternative would be Gamma mixture of Poissons (generalization of Negative Binomial)

Purpose of random effects is to <u>describe</u> within-(E/C)-group heterogeneity in λ in relation to Poisson

	DIC
FEPR	1500
REPR	1419.6

massive difference

posterior mult. effect



IHGA is statsig beneficial

REPR wins

Regression models

outcome

predictors (statistics)

= features (machine learning (ML))

(i) n

$\leftarrow j \rightarrow$

linear: $\textcircled{1}$ $y \in \mathbb{R}$

logistic: $y \in \{0, 1\}$

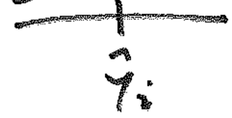
$$y \qquad \underbrace{x_1 \quad \cdots \quad x_k}$$

some of these x_j are quantitative, some dichotomous, ...

$\underline{\text{goal}}$: predict y from $(x_1, \dots, x_k)$ as accurately as possible (ML) $\rightarrow \hat{y}$

predicted value (point prediction)

& attach well-calibrated measures of predictive uncertainty (error bands) to the point predictions (Bayesian statistics)



$p(y_i \mid D\, \mathcal{B} \cdots)$

"Bayes on a small budget"

(1) create m bootstrap copies of D

(2) run ML algorithm on copy ℓ, obtaining $\hat{y}^*_\ell$;

(3) collect $\hat{y}^*_\ell$ together to form an approx. posterior predictive dist.

---

linear regression

$$Y_i = \underline{\underline{\beta_0}} + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + e_i$$

"error"

e_i IID $N(0, \sigma_e^2)$

$$Y_i = \sum_{j=0}^{k} x_{ij}\beta_j + e_i, \quad x_0 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\underset{n \times 1}{\underset{\sim}{Y}} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n \times (k+1)}{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ \vdots & \vdots & & \ddots & \\ 1 & x_{1n} & & & \end{bmatrix}$$

(fixed constants)

$$\underset{n \times 1}{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} \qquad \underset{\sim}{Y} = X \underset{\sim}{\beta} + \underset{\sim}{e}, \quad \underset{(k+1) \times 1}{\underset{\sim}{\beta}} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}$$

maximum likelihood

$$\hat{\underset{\sim}{\beta}}_{MLE} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underset{\sim}{y}$$

$$\hat{\sigma}^2_{MLE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\hat{\underset{\sim}{y}} = \underline{X} \hat{\underset{\sim}{\beta}}$$

$$= \underbrace{\underline{X} (\underline{X}^T \underline{X})^{-1} \underline{X}^T}_{\text{hat matrix}} \underset{\sim}{y}$$

Bayesian story: like MLE but with prior on $\underset{\sim}{\beta}$ & σ^2_e (see Banerjee's notes)

binary outcomes

$$(Y_i \mid p_i, B) \overset{I}{\sim} \text{Bernoulli}(p_i) \quad (i = 1, \ldots, n)$$

any inverse CDF $F^{-1}(p_i)$

on $\mathbb{R}$

$$\log \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$$

logistic regression

closed-form MLE don't exist; iterative methods needed model uncertainty

(1) variable selection : (6)

optimal subset of $(X_1, \dots, X_k)$