Notes 9/21

Liz Austin

9/23/2017

Comments on Assignment 1:

1) Produce a single, self-contained file
    - R markdown or python equivalent
    - create images using that same software
2) Solution set will be posted
3) Use slack/office hours!!
    - Work on equations/latex/knitting, share interesting realizations
    - slack board for assignment 2
4) We will begin using HW to cover things not covered in class

Casey's question: Confidence interval = coverage probability? (i.e. 95% when n -> inf?)

Topics for Today:

1) Diagnostics (*read book, too much to cover in class)
2) Newton-Raphson
3) Inference

Diagnostics

1) Deviance
    - measure of goodness of fit based on likelihood
    - most useful as a comparison between models
    a) Equation:

$$D = -2[L(\hat{\mu}|y) - L(y|y)]$$

$L(y|y)$ is log likelihood, saturated model, "best case scenario"
LRTS- comparing your model over a saturated model

$$D_{guassian} = \sum(y - \hat{\mu})^2$$

same Kernel of MSE

$$D_{poisson} = 2\sum(y\log(\frac{y}{\hat{\mu}}) - (y - \hat{\mu}))$$

b) Deviance test for "nested" models

$M_0 \text{−>} \hat{\mu}$ and $M_1 \text{−>} \hat{\mu}_1$
M_0 is a special case of M_1
ex: $\eta_{i1} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$
$\eta_{i0} = \beta_0 + \beta_1 X_{i1} (\text{ i.e. } \beta_2 = 0)$

- simpler models have larger deviance, lower log-likelihoods

- $H_0 : \mu_1$ explains the data as well as μ_0
  $D(y|\hat{\mu}_0) = D(y|\hat{\mu}_1) \sim \chi^2_{df}$
  df = # of parameter difference = # of parts in M_1 - # parts in M_0

2) Residuals
   - Pearson, deviance, standardized
   a) Equation:
      - Pearson: $e_i = \frac{y-\hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$
      - Standardized: $r_i = \frac{e_i}{\sqrt{(1-\hat{h}_i)}}$
        - Advantage of standardized residuals: r_i (approx) $\sim N(0, 1)$
   b) Binned residual plots
      - "Is my model fitting reasonably?"
      - In R –> arm::binnedplot()
        1. Group observations into groups ordered by whatever x value $(x_k, x_{ik}, \hat{y})$ with the same number of observations per group
        2. Compute the group-wide average residual
        3. Plot
   c) Ex. 6.2.2 in book, Heart Disease and BP
      - Sample of male residents of Framingham, MA age 40-59
        - Y=[1, developed heart disease during 6 yr. follow-up or 0, no HD]
        - X = Systolic blood pressure (continuous)
      - Independence model: $\text{logit}(\pi_i) = \alpha$
      - Systolic BP model: $\text{logit}(\pi_i) = \alpha + \beta X_i$

X level	n	Ind. Model	SBP Model
<117	156	-2.62	-1.11
117-126	252	-0.12	2.37
127-136	.	-2.02	-0.95
137-146	.	-0.74	-0.57
147-156	.	-0.74	-0.57
157-166	.	0.93	-0.33
167-186	99	3.67	0.65
>187	43	3.07	-0.18

Pretty strong linear of residuals in independence model so we need to use SBP model

Newton-Raphson

1) What is it?
   - A general iteraive method for solving non-linear equations
   a) Steps:
      1. Initial guess at solution
      2. Approximate the function locally –> find maximum
      3. The maximum becomes your next "guess"
      4. Complete steps 2-3 until convergence
   b) Mathematically:
      - We are maximizing $L(\beta)$

$$u = \left(\frac{\partial L(\beta)}{\partial \beta_1}, \frac{\partial L(\beta)}{\partial \beta_2}, ...\right), H = \text{ Hessian Matrix where element i,j is: } \frac{\partial^2 L(\beta)}{\partial \beta_i \partial \beta_j}$$

      - Iterative process, t=1...

$$u^{(t)} = u(\beta^{(t)})$$
$$H^{(t)} = H(\beta^{(t)})$$
$$\text{where } \beta^{(t)} \text{ is our } t^{th} \text{ "guess" of } \beta$$

- Taylor series approximation in local region of $\beta^{(t)}$:

$$L(\beta) \approx L(\beta^{(t)}) + u^{(t)}(\beta - \beta^{(t)}) + \frac{1}{2}(\beta - \beta^{(t)})^T H^{(t)}(\beta - \beta^{(t)}) \text{ (*multidimensional quadratic!)}$$
$$\text{solve: } \frac{\partial L(\beta)}{\partial \beta} = u^{(t)} + H^{(t)}(\beta - \beta^{(t)}) = 0 \text{ –> } \beta^{(t+1)}$$
$$\beta^{(t+1)} = \beta^{(t)} - (H^{(t)})^{-1}u^{(t)}$$
$$H^{(t)}\text{–>}(I^{(t)})^{-1} \text{ *Fisher's uses this instead of Hessian matrix}$$