BIOST 572 Introductory Talk

David Benkeser

University of Washington Department of Biostatistics

April 4, 2012

Enjoy every sandwich

Warren Zevon (1947-2003)
American singer-songwriter

MODEL-ROBUST REGRESSION AND A BAYESIAN "SANDWICH" ESTIMATOR

BY ADAM A. SZPIRO, KENNETH M. RICE AND THOMAS LUMLEY

University of Washington, University of Washington and University of Washington

The Annals of Applied Statistics
2010, Vol. 4, No. 4, 2099–2113

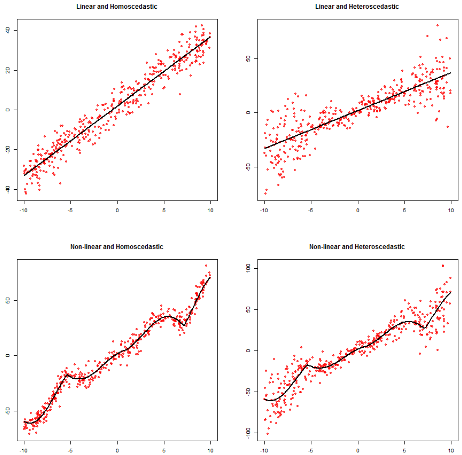
Classical Assumptions

- Outcome variable linearly related to covariates on average
- Random variations from the linear trend are homoscedastic
- Random variations are Normally distributed

Under these assumptions...

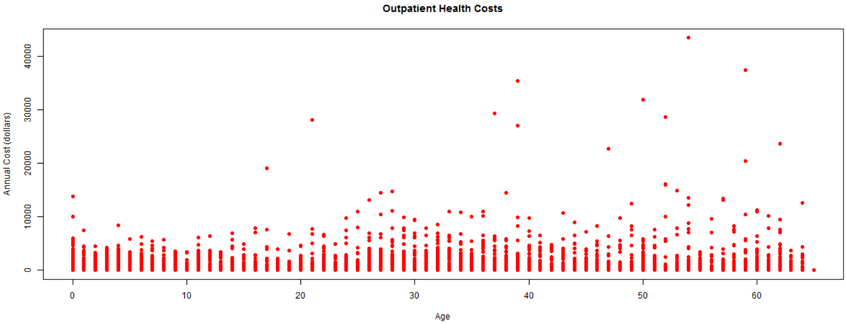
- Classical frequentist methods allow us to make exact probability statements
- Bayesian methods allow us to make uncertainty statements

Motivation
Data



Motivation
Health Care Cost Data

Data from evaluation Washington State Basic Health Plan, a program offering subsidized health insurance for low income residents (Diehr et al 1993)



Do we really need these assumptions?

- Central limit theorem guarantees asymptotically Normal sampling distribution
- Without linearity it is not even clear what quantity we are estimating
- Linearity and homoscedasticity assumptions ensure valid standard errors

Suppose we have Y and $\mathbf{X}$ sampled from a population F and

$$\mathbb{E}_F(Y|\mathbf{X}=\mathbf{x}) = \phi(\mathbf{X})$$

Define our quantity of interest as

$$\boldsymbol{\beta} = \underset{\boldsymbol{\beta}}{argmin}\ \ \mathbb{E}_F\left[\left(\phi(\mathbf{X}) - \mathbf{X}^T\beta\right)^2\right]$$

i.e. the set of coefficients minimizing average squared error in approximating the mean value of Y by linear function of $\mathbf{X}$

Definition of $\boldsymbol{\beta}$ is statement about the scientific question of interest, not details of random sampling

β is the best linear approximation to the curve $\phi(\cdot)$.

- This idea extends back to the early work of Gauss (1809), Legendre (1805), and Jacobi (1841)

Jacobi (1841) showed that

$$\beta_X = \frac{\mathrm{Cov}_F[X, Y]}{\mathrm{Var}_F[X]}$$

- Berman (1988) points out that Jacobi's result has been largely overlooked

Can also view β as solution to estimating equations (Liang and Zeger (1986))

$$\mathbb{E}_F\left[G(\beta, Y, \mathbf{X})\right] = \mathbb{E}_F\left[\mathbf{X}^T(Y - \mathbf{X}\beta)\right] = 0$$

Note that, in 1D, substituting $\mathbb{E}_F$ with summation over the sample, this gives

$$\hat{\beta}_X = \frac{Cov(X, Y)}{Var(X)}$$

Estimating equations give motivation for use of "sandwich" estimator of Huber (1967) and White (1980)

Allow for robust, frequentist inference on β in a model agnostic setting

Freedman (2006) points out the danger of fitting incorrect linear model

A Bayesian analogue of estimating equations has been an open problem for some time

Recall our definition of β

$$\boldsymbol{\beta} = \underset{\boldsymbol{\beta}}{argmin} \;\; \mathbb{E}_F \left[\left(\phi(\mathbf{X}) - \mathbf{X}^T \boldsymbol{\beta} \right)^2 \right]$$

If we knew $\phi(\cdot)$ and F, we could calculate β directly
Instead, model data using a flexible Bayesian model such that we can derive a posterior for $\phi(\cdot)$ and F
Likelihood:

$$Y|\mathbf{X} = \mathbf{x}, \phi(\cdot), \sigma^2(\cdot) \sim N(\phi(\mathbf{x}), \sigma^2(\mathbf{x}))$$

Priors: Saturated model for $\phi(\cdot)$ and $\sigma^2(\cdot)$ in discrete case, spline based priors in continuous case

The posteriors for $\phi(\cdot), \sigma^2(\cdot)$ and F *induce* a posterior for $\boldsymbol{\beta}$ from which we can use standard posterior summaries

It turns out that for random $\mathbf{X}$...

- The posterior mean for $\boldsymbol{\beta}$ is asymptotically equivalent to the ordinary least squares slope estimate $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y$
- The posterior standard deviation of $\boldsymbol{\beta}$ is asymptotically equivalent to the usual sandwich estimate $\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\Sigma\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\right]^{1/2}$, where Σ is a diagonal matrix of squared residuals

Conclusions

Classical assumptions of linear models not always reasonable and are not always necessary

This method provides Bayesian analogue of frequentist estimating equations and sandwich-based errors

In the random $\mathbf{X}$ case, posterior standard deviation of β is asymptotically equivalent to the "sandwich"-estimator

Payoffs for fitting very flexible Bayesian models (difficult) will be minimal when studying trends in large samples using standard sandwich-based inference (easy)

Questions?