Bootstrap Statistics and Simulation

Elise de Doncker
John Kapenga
Joseph McKean
Western Michigan University

Simple Linear Models
(with normality assumptions)

Model: $y = \bar{x}' * \bar{\beta} + \varepsilon$

Sample: $\{\bar{x}_i, y_i\} 0 \leq \mathrm{i} < \mathrm{n}$

Least Squares Solution: $\hat{\beta} = (X'X)^{-1} X' \bar{y}$

(Solving the Normal Equations)

C.I. Use standard deviation and Student's t

Not so simple models

- Non-normal error distributions
- Less than full rank systems
- High leverage designs
- Errors in data (outliners)
- General models $g(y, \overline{x}, \overline{\beta}, \varepsilon) = 0$
- Cis, hypothesis testing, influence functions, outlier detection, power analysis

Common GLMs

Many common exponential models use the pdf

$$f(y_i;\theta_i,\phi) = \exp\left\{\frac{y_i\theta_i - b(\theta)}{a(\phi)} + c(y_i,\phi)\right\}.$$

Which is called the canonical PDF.

These includes the normal, gamma, binomial, Poisson, negative binomial, etc…

If the pdf is binomial this is the logit model, which is the model of interest here

Logit Models

- Responses y_i are 1 (success) or 0 (failure)
- We want to estimate p = Pr[success]
- Can model device failure or treatment success
- The linear term $r_i = \beta_0 + \beta_1 x_i$
  - Provides the odds ratio r in the model
- With link function $r_i = \ln\left(\frac{p_i}{1-p_i}\right)$
- Whose inverse is $p_i = \frac{1}{1+\exp(-r_i)}$

The logit Estimation Problem

- Using the maximum likelihood principle leads to a set of general estimating equations (GEE)
- Generally GEEs can not be solved in closed form
- Iterative methods must be used. Such as reweighted least squares (IRLS) or a quasi-Newton method like L - BFGS
- There can be many problems here.
- Computing a CI is even more daunting.

Logit GEEs

The logit GEEs

$$\sum_{i=1}^{n} w_i(Y_i - \mu_i)\frac{\partial(\boldsymbol{x}_i^T\boldsymbol{\beta})}{\partial\mu_i}x_{ij}, \quad j = 1, \ldots, p,$$

Where

$$w_i = \left[V_i(\partial(\boldsymbol{x}_i^T\boldsymbol{\beta})/\partial\mu_i)^2\right].$$

IRLS

$$\hat{\boldsymbol{\beta}}^{(1)} - \hat{\boldsymbol{\beta}}^{(0)} = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}^{1/2}\mathbf{Z},$$

$$\begin{aligned} V_i &= p_i(1-p_i) \\ \frac{\partial(\boldsymbol{x}_i^T\boldsymbol{\beta})}{\partial\mu_i} &= \frac{1}{p_i(1-p_i)} \\ \sqrt{w_i} &= \sqrt{p_1(1-p_i)}. \end{aligned}$$

IRLS for 2 parameters

Using the matrix inversion formula

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

the IRLS iteration formula can be written in an interesting for using inner products.

Positional problems with this approach would need to be addressed in production code.

Reweighted Least Square Iteration 2 parameter logit model in R

```
###   Loop
I = 1
while(i < nstp){
        I = i + 1
        p0 = 1/(1+exp(-(beta00 + beta01*x)))    ### component wise
        w = p0*(1-p0)                           ### component wise
        v0 = sqrt(w)*1                          ### component wise
        v1 = sqrt(w)*x                          ### component wise
        r = (y-p0)/sqrt(w)                      ### component wise
        a = ip(v0,v0)
        b = ip(v0,v1)
        d = ip(v1,v1)
        delta0 = (1/(a*d - b*b))*(d*ip(v0,r) - b*ip(v1,r))
        delta1 = (1/(a*d - b*b))*(-b*ip(v0,r) + a*ip(v1,r))
        beta00 = beta00 + delta0
        beta01 = beta01 + delta1
}
beta00
beta01
```

The Bootstrap
(90% Confidence Interval for β)
f() could be an IRLS

Input $\{\bar{x}_i, y_i\} 0 \le i < n$

$\hat{\beta} = f(X, \bar{y})$

for(b=0; b<B; b++) {

Resample to get $\{\bar{x}_i^{(b)}, y_i^{(b)}\} 0 \le i < n$

$\hat{\beta}_b = f(X^{(b)}, \bar{y}^{(b)})$

}

C.I. = [quant(0.05, $\{\hat{\beta}_b\}$), quant(0.95, $\{\hat{\beta}_b\}$)]

Bootstrap
(Resampling, Jack Knife)

- Proposed by Effron in 1979
- Requires only an estimator to get CIs (and other information)
- Is compute intensive, B = 3500 typical for Cis
- Is “embarrassingly” parallelizable
- Is very easy to explain and simplifies many situations, often with better results than “usual” methods.

Some Bootstrap Applications

- CIs
- Power analysis
- Hypothesis testing
- Influence functions
- Outlier detection
- Parts of Bayesian procedures
- Algorithm comparisons and validation
- Summary of uncertainty (flu epidemic display)

What is needed for CUDA

- PRNG, cuRAND was a bottlneck for 1 stream per thread
- An iterative algorithm that is SIMD suitable
  - It must survive bad input and
  - Should report non-convergence
  - Uses the GPUs memory effectively
- If 20 parameters could be done there would be more applications
- Make an R plug-in (easy)
- Adapt to Kepler

Timing

- n = 1000 : size of $\bar{x}$ & $\bar{y}$
- B = 3500 : Bootstraps for each CI
- M = 10,000 : number of simulations

# Bootstraps	R	C	C (8 cores)	M2090
1	.14 sec	.01 sec	.01	.3 sec
3500	7.9 min	28 sec	3.4 sec	.3 sec
10000 *3500	55 days	3.2 days	9.4 hr	21 min

References

Bradly Efron and Roberrt J. Tibshirani (1994), *An IntrodUcton to the Bootstrap*, Boca Raton,Cahpman and Hall CRC .

Cox, D. (1970), *Analysis of binary data*, London: Spottiswoode, Ballantyne and Co.

Hettmansperger, T. P., & McKean, J. W. (2011), *Robust Nonparametric Statistical Methods, 2nd Ed.*, New York: Chapman & Hall.

Liang, K-Y. and Zeger, S.L. (1986), Longitudinal data analysis using generalized linear models, *Biometrika*, 73, 13–22.

McCullagh, P. and Nelder, J. (1989), *Generalized linear models*, London: Chapman & Hall.