Point Estimation with Quantal Response Data:

Parametric Bootstrap Estimator Beats the Corresponding MLE

Amy Schrader M.S., Ishwori Dhakal M.S., Reid D. Landes Ph.D.
Department of Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AR

Abstract

Maximum likelihood estimation (MLE) is often used for estimating lethal dose 50% (LD50) and dose reduction factor (DRF) in toxicity studies. We investigated two point estimators of LD50 and DRF: the MLE and the median of a parametric bootstrap distribution. In a Monte Carlo experiment, we simulated quantal response data from different experimental settings. We then compared mean squared error (MSE) between MLE and the bootstrap estimator of both LD50 and DRF. The bootstrap estimator of both LD50 and DRF generally has a lower MSE than the MLE, especially in smaller sample sizes. After investigating the variances and biases of these estimators, the differences between the MSE of the bootstrap estimator and the MSE of the MLE are attributable to the variances. We recommend using the median of the parametric bootstrap for estimating LD50 and DRF over the MLE.

Key Words: relative potency; LD50; dose reduction factor; mean squared error.

1. Introduction

Quantal response analyses relate the probability of an event (e.g., death) to a quantitative factor (e.g., toxin dose). In toxicology studies, a parameter of interest is the dose that is lethal to half of the population; i.e. lethal dose 50% or LD50. When interpreting LD50 values, a smaller LD50 indicates greater toxicity. A parameter useful for comparing two LD50s is the dose reduction factor (DRF, a.k.a. relative potency). A motivating example is comparing the LD50 of total body irradiation between animals receiving a potential radioprotectant and animals receiving no treatment. DRF is $LD50_1$ / $LD50_0$, where 1 and 0 subscript treatment and control, respectively. If $LD50_1 > LD50_0$ then DRF>1 which means the treatment reduces the toxicity from that experienced by the controls. The focus of this paper is to compare two point estimators of LD50 and DRF: (*i*) the MLE and (*ii*) the median of a parametric bootstrap.

2. Methods

2.1 Model

Let Y be the number of events out of n. Model $Y_{jk} \sim \mathrm{Bin}(n_{jk}, p(x_{jk}; \alpha_j, \beta))$ for $k = 1, 2, \ldots, m_j$, where m is the number of distinct x values in group j = 0, 1, and n_{jk} is the number of subjects receiving dose x_{jk}. We use probit regression to estimate $(\alpha_0, \alpha_1, \beta)$ in $p(x_{jk}; \alpha_j, \beta) = \Phi(\alpha_0 + \alpha_1 \mathrm{TRT}_j + \beta x_{jk})$ where $\Phi(*)$ is the standard normal cumulative distribution function and TRT indicates receipt of treatment (1) or not (0). Probit regression was chosen over logistic regression since LD50 estimation was primarily developed using probit regression and continues to be often used for LD50 estimation.

Using maximum likelihood, we fit a given data set with a probit regression, obtaining $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta})$. These MLEs are then used in two ways: (*i*) to obtain MLEs of LD50 and DRF, and (*ii*) to generate B bootstrap data sets under the model

$$Y^*_{jk} \sim \mathrm{Bin}(n_{jk}, p(x_{jk}; \hat{\alpha}_j, \hat{\beta})),$$

where n and x are as in the original data set. This latter usage is a type of parametric bootstrap. From each of the bootstrapped data sets, we obtain the MLEs of $(\alpha_0, \alpha_1, \beta)$, say $(\hat{\alpha}_0^*, \hat{\alpha}_1^*, \hat{\beta}^*)$, compute LD50* and DRF*, and finally take the median from each of the bootstrapped distributions of the parameters of interest (Figure 1).

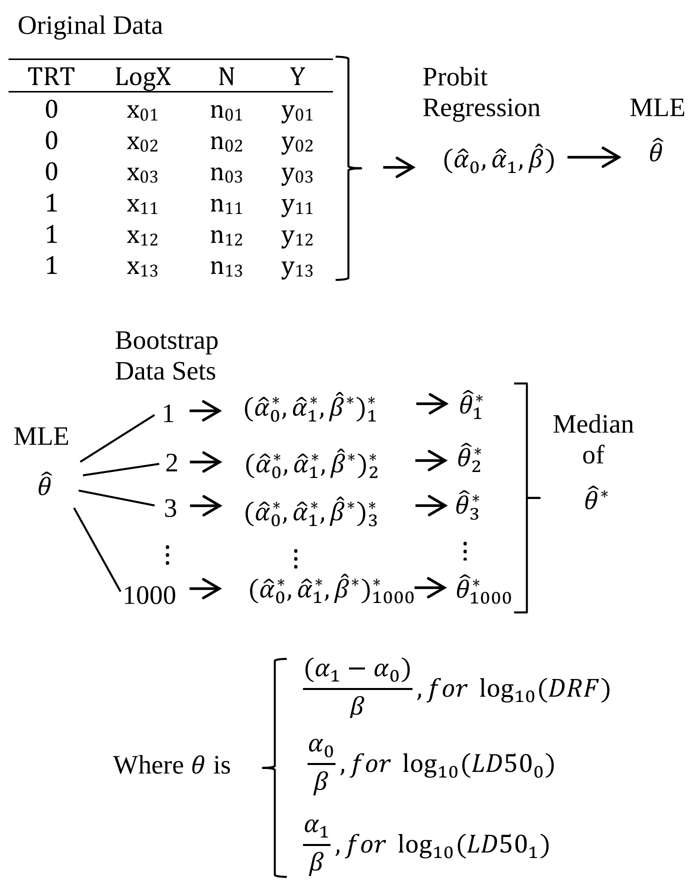

Figure 1. Schematic of how bootstrap data were generated.

2.2 Simulation Study

We considered the number of doses (*M*), the number of animals per treatment×dose (*N*), steepness of the response (*SLOPE*), and dose reduction factor (*DRF*).

We set:

- *M* at 3, 5 and 7,
- *N* at 4, 9 and 16,
- *SLOPE* at (α, β) = (1.5714, 2.91) for shallow control and (-23.25, 23.25) for steep control, and
- *DRF* contingent upon *SLOPE*,
  - at 1, 2.2, and 3.26 for shallow, and
  - at 1, 1.1, and 1.16 for steep.

The settings for *M* and *N* were selected based upon the number of radiation doses and samples sizes that are often used in the radiation countermeasure studies (see Landes et al., 2013 for a review). And the SLOPE and DRF settings were also chosen based on observed values reported in the radiation countermeasure literature. There were 54 different settings (3 *M/N* × 3 *N/SLOPE* × 2 *SLOPE/DRF* × 3 *DRF*). We chose doses to be centered on the true LD50 for each treatment group and equally spaced, and generated random responses under a probit regression model defined by the given setting. The data were analyzed with probit regression and estimates computed and recorded. We repeated this process 1,000 times for each of the 54 settings.

3. Results

We compared the mean square error (MSE) between the MLE and the bootstrap estimator of both LD50 and DRF. The bootstrap estimators of both LD50 and DRF generally had lower MSE than the MLE (Figures 2 and 3). For smaller sample sizes, the bootstrap estimators of both LD50 and DRF had considerably lower MSE than the MLE. However, as the sample size increases, the difference of the MSEs between the two estimators decreases.

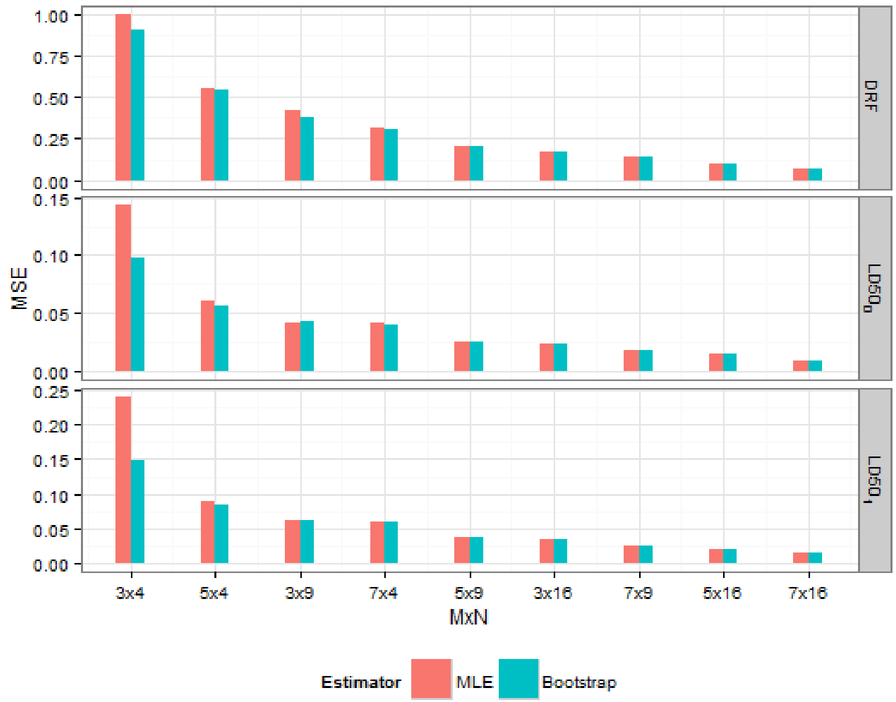

Figure 2: For each $M \times N$ combination, MSE compared between MLE and bootstrap median estimators of DRF and LD50.

Another way to view the differences of MSE between the MLE and the bootstrap estimator is shown in Figure 3. The MSE of the bootstrap estimator is expressed as a percentage of the MSE of the MLE. When the percentage is below 100 percent, the bootstrap estimator has lower MSE compared to the MLE. Percentages over 100 percent indicate the MLE has lower MSE compared to the bootstrap estimator.

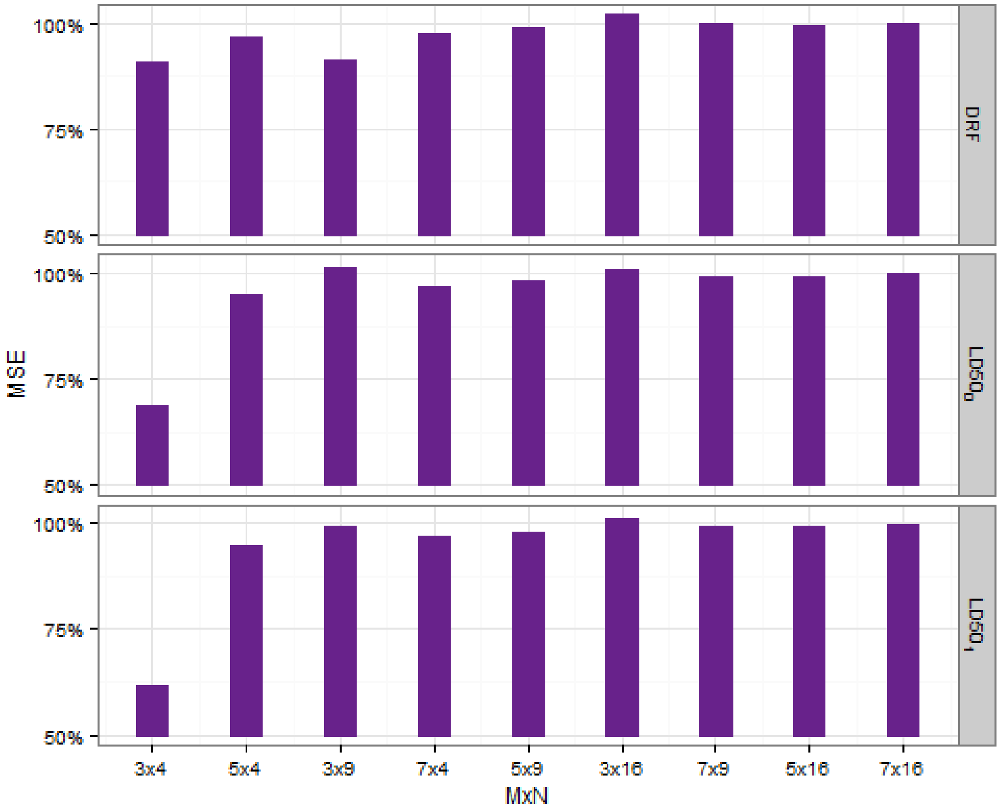

Figure 3: For each $M \times N$ combination, MSE of bootstrap median estimators expressed as a percentage of the MLE's MSE. Note: Vertical axis starts at 50 percent to highlight differences.

In continuation of expressing the MSE of the bootstrap estimator as a percentage of the MLE's MSE, we compared the variance and bias^2 of the two estimators since the MSE is the sum of the variance and bias^2. In all cases, the bias^2 accounted for less than 5 percent of the MSE. On average, both components (variance and bias^2) of the bootstrap median estimators were less than that for MLEs. The bootstrap median estimator was substantially better for the smallest $M \times N$ combinations.

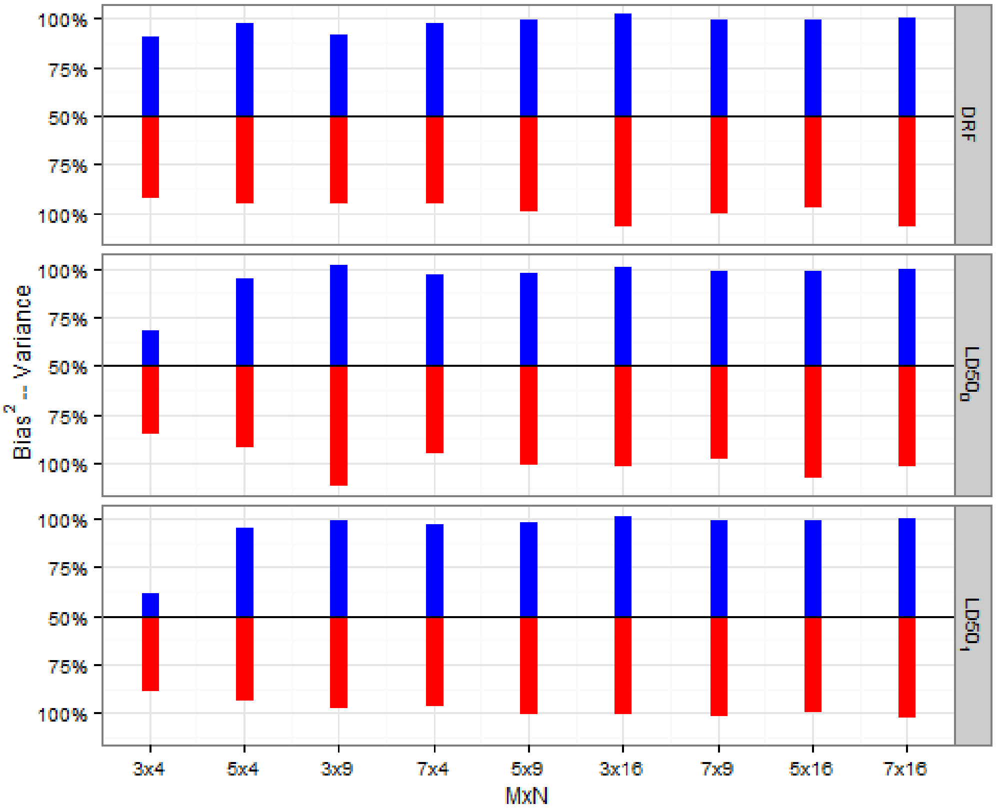

Figure 4: For each $M \times N$ combination, variance (blue) and bias^2 (red) of bootstrap median estimators expressed as a percentage of the MLE's components. Note: Vertical axis starts at 50 percent to highlight differences.

4. Conclusions

4.1 Recommendation

For smaller sample sizes (e.g., < 60), the parametric bootstrap estimator has a notably lower MSE than the MLE for both LD50 and DRF. Therefore, we recommend using the parametric bootstrap estimator over the MLE when sample sizes are low. Otherwise the parametric bootstrap estimator and MLE are comparable when estimating LD50 and DRF.

4.2 Limitations

We used the median of the bootstrap distribution to obtain the point estimator rather than the mean. As sample sizes decrease the probability of astronomically large parameter estimates in the bootstrap distribution increase; the median was resistant to these outliers. Theoretically, the mean of the bootstrap distribution is preferred, but the median tends to be more stable when working with small samples. For theoretically symmetric bootstrap distributions, the mean and median will be the same.

Though the simulation study covers a wide range of scenarios appropriate for radiation countermeasure studies, it does not consider extreme slopes ($|\beta|$ near 0 or ∞), very small sample sizes (e.g., $M \times N = 2 \times 1$), or logistic regression. Further preliminary investigation into these limitations show a similar pattern of the bootstrap estimator having a better MSE than the MLE on which it is based in small sample situations.

Another consideration is that the expected response rate (averaging over all M doses) was 50% - the response rate carrying the most information about the parameters. We did not explore the (real) possibility of the design points missing the truth, and thus having response rates further away from 50%. We are unsure of how the two estimation methods would compare in such cases.

This study considers only point estimation. We are currently investigating whether bootstrap confidence intervals for the desired parameters tend to outperform confidence intervals based on maximum likelihood estimation, especially when working with small samples.

Acknowledgements

This project was supported, in part, by Award Number 1UL1RR029884 from the NCATS.

References

Landes RD, Lensing SY, Kodell RL, Hauer-Jensen, M. (to appear) “Practical advice on calculating confidence intervals for radioprotection effects and reducing animal numbers in radiation countermeasure experiments.” *Radiation Research*.