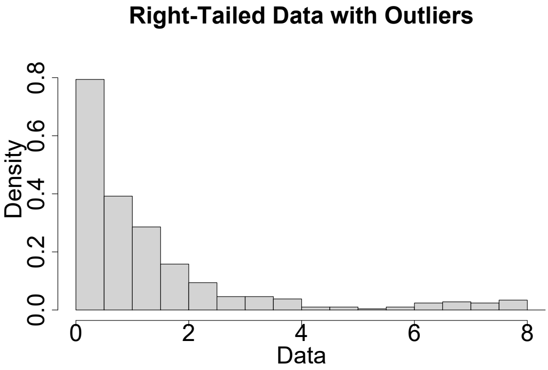
Trimming the Mean of Data with Long Right Tails: The Bias-Variance Tradeoff

Xiaoya Wang, Bertrand Sodjahin, and Gradon Nicholls

June 19, 2023

Motivation



- Goal: Compute sample average.
- Challenge: Are outliers “true” outliers or not?
- Even if true outliers, sample average may have large variance.

Trimming the Mean

Let $Y_i \stackrel{\text{iid}}{\sim} \mathrm{F}(\cdot)$. The sample mean is

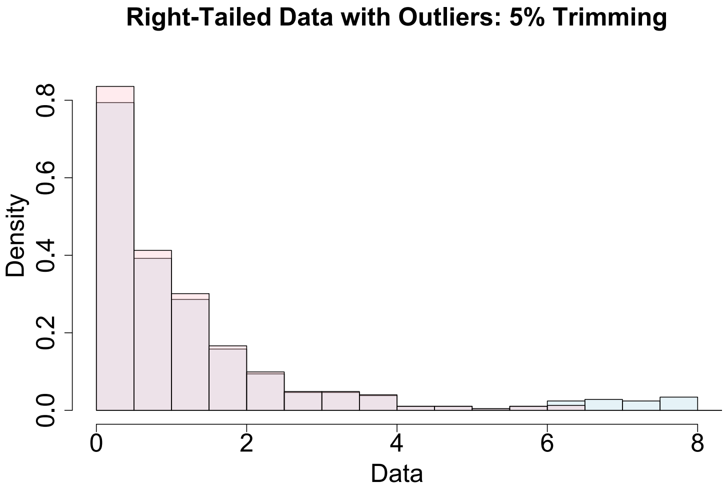
$$\hat{\mu}_0 = \frac{\sum_{i=1}^{n} Y_i}{n}.$$

To reduce the influence of outliers in the right tail, could use trimmed mean: remove the top $100p\%$ of observations:

$$\hat{\mu}_p = \frac{\sum_{i=1}^{n} I(Y_i \leq Y_{(1-p)})Y_i}{\sum_{i=1}^{n} I(Y_i \leq Y_{(1-p)})}$$

where $Y_{(1-p)}$ is the $1-p^{\text{th}}$ quantile from the sample data.

Example: 5% Trimming



Bias-Variance Tradeoff

- Is there any value to trimming if there are no “true” outliers?
- Trimming reduces large values $\rightarrow$ decrease in variance, increase in bias.
- Overall, may observe reduction in MSE:

$$\text{MSE} = \text{Bias}^2 + \text{Variance}$$

Research Question: How does trimming affect bias, variance, MSE of estimator if there are no outliers?
Assumption: We assume underlying data generated from a Weibull random variable.

The Weibull Distribution e.g. Bain and Engelhardt (1992)

$X \sim \text{Wei}(\theta, \beta)$:

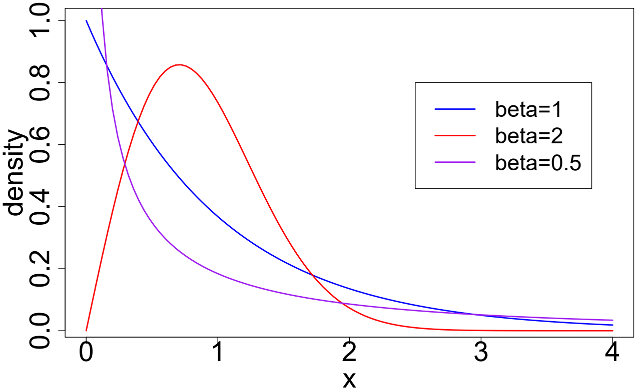
$$f(x;\, \theta, \beta) = \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1} e^{-(x/\theta)^{\beta}}$$

- θ: scale parameter–i.e. $X \sim \text{Wei}(\theta, \beta) \implies \frac{X}{\theta} \sim \text{Wei}(1, \beta)$
    - Interested in relative efficiency–WLOG assume $\theta = 1$
- β: shape parameter
    - $\beta = 1$: exponential distribution
    - $\beta > 1$: polynomial times exponential
    - $\beta < 1$: inverse polynomial times exponential – asymptote at 0

$$\mathbb{E}[X] = \theta\Gamma(1 + 1/\beta)$$

$$\text{Var}(X) = \theta^2\left(\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)\right)$$

Shape of the Weibull



We will consider only values $\beta \in (0, 1)$.

Maximum Likelihood Estimation of the Mean

- Assume $Y_i \overset{\text{iid}}{\sim} \text{Wei}(\theta, \beta)$. Let $\mu = \mathbb{E}[Y_i] = \theta\Gamma(1 + 1/\beta)$.
- $\hat{\mu}_{\text{MLE}} = \hat{\theta}_{\text{MLE}}\Gamma(1 + 1/\hat{\beta}_{\text{MLE}})$ by invariance property.
- $\hat{\theta}_{\text{MLE}}, \hat{\beta}_{\text{MLE}}$ have no closed form, rely on numerical methods
- For large n, $\text{Var}(\hat{\mu}_{\text{MLE}})$ achieves Cramer-Rao lower bound, and so can be used as "gold standard" baseline.
- For small n, less certain.

Simulation

Parameters:

- $\beta \in (0, 1)$
- $n \in \{20, 200, 2000\}$
- $p \in \{0.01, 0.025, 0.05, 0.1\}$

Simulation:

- for $s = 1, \ldots, S$
  - gen $Y_i^s \overset{\text{iid}}{\sim} \text{Wei}\,(1, \beta),\ i = 1, \cdots, n$
  - estimate $\hat{\mu}_0^s,\ \hat{\mu}_p^s,\ \hat{\mu}_{\text{MLE}}^s$

Estimate:

- Bias: $\frac{\sum_s \hat{\mu}^s}{S} - \Gamma(1 + 1/\beta)$
- Variance: $\frac{\sum_s \left(\hat{\mu}^s - (1/n) \sum_s \hat{\mu}^s\right)^2}{S-1}$
- MSE: $\text{Bias}^2 + \text{Var}$

(Preliminary?) Results

- . . . . . .

Conclusions

- ......

Bain, Lee J and Max Engelhardt. 1992. *Introduction to probability and mathematical statistics*. Vol. 4 Duxbury Press Belmont, CA.