STAT 6324 Fall 2020: Project 1

Robust Estimation for Location Parameter

1 Introduction

An estimator which is insensitive to departures from the assumed distribution is said to be a robust estimator.

For example, let $X_1, X_2, \ldots, X_n$ be Normal (μ, σ^2) variables. Then $\bar{X}$ is an unbiased estimator of μ and $var(\bar{X}) = \frac{\sigma^2}{n}$. Also, $\bar{X}$ is also the MVUE (minimum variance unbiased estimator) of μ. We will introduce some robust estimators of μ in the following section.

2 Alternate estimators of μ

Let $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ be the order statistics obtained from $X_1, X_2, \ldots, X_n$.

2.1 r-trimmed mean

$$T_r = \frac{X_{r+1:n} + X_{r+2:n} + \ldots + X_{n-r:n}}{n-2r}$$

2.2 r-Winsorized mean

$$W_r = \frac{(r+1)X_{r+1:n} + X_{r+2:n} + \ldots + X_{n-r-1:n} + (r+1)X_{n-r:n}}{n}$$

2.3 r-linearly weighted mean

$$L_r = \frac{1}{2\left(\frac{n}{2} - r\right)^2} \sum_{i=1}^{\left[\frac{n}{2}\right]-r} (2i-1)(X_{r+i:n} + X_{n-r-i+1:n})$$

2.4 Median

$$M = \begin{cases} X_{\frac{n+1}{2}:n} & \text{if } n \text{ is odd} \\ \frac{1}{2}\left[X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}\right] & \text{if } n \text{ is even} \end{cases}$$

3 Outlier Models

3.1 A Fixed Outlier Model

Suppose k is a fixed integer from $1, 2, \ldots, n$:

$$\begin{cases} (n-k)X's & \sim F \\ kX's & \sim G \end{cases}$$

Location-outlier model if $G(x) = F(x - \lambda)$
Scale-outlier model if $G(x) = F(x/\tau)$
For example: $X_1, X_2, \ldots, X_{n-1} \sim N(\mu, \sigma^2)$ and $X_n \sim N(\mu + \lambda\sigma, \sigma^2)$, where λ is the location-shift in the outlier.

3.2 A Random Outlier Model

$X_1, X_2, \ldots, X_n \sim$ a contaminated distribution with distribution function:

$$H(x) = pF(x) + (1-p)G(x)$$

where $F(x)$ is the true distribution and $G(x)$ is the contaminating distribution.

The number of contaminants in the sample is a random variable with $Binomial(n, 1-p)$.

4 Simulation Study

We are interested to the performance (bias, variance and mean square errors) of some robust estimators for the location under difference situations.

4.1 Robust Estimators proposed

$$\bar{X}, T_{0.1n}, T_{0.2n}, W_{0.1n}, W_{0.2n}, L_{0.1n}, L_{0.2n}, M$$

4.2 Models

1. Normal (0, 1)

2. Random outlier model: $0.9N(0,1) + 0.1N(0,4)$

3. Random outlier model: $0.8N(0,1) + 0.2N(0,4)$

4. Fixed outlier model: $(n-1)N(0,1)$ and $1N(0,4)$

5. Fixed outlier model: $(n-2)N(0,1)$ and $2N(0,4)$

6. Logistic distribution with p.d.f:

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$E(X) = 0, \quad Var(X) = \frac{\pi^2}{3}$$

7. Laplace distribution with p.d.f:

$$f(x) = \frac{1}{2}e^{-|x|}$$

$$E(X) = 0, \quad Var(X) = 2$$

- Remarks: For models 6 and 7, standardize your generated random variates by the corresponding mean and variance of the distribution. That is, $Y_i = \frac{X_i}{\pi/\sqrt{3}}$ for model 6 and $Y_i = \frac{X_i}{\sqrt{2}}$ for model 7.

4.3 Sample sizes

$n = 10, 20, 30$ and 50