Robust Estimators for Location and Scale Parameters

Why Robust Statistics?

- To understand about a population, we cannot assess it all, but just through a representative of population, which is a sample/dataset collected randomly from the population.

- We may make some assumptions about the underlying distribution (population distribution).

- However, outlier(s) may appear in the sample, hence the sample distribution may depart from the underlying distribution assumptions.

- Therefore, the conclusion derived from this sample using some statistical methods might not be reliable if these statistical methods are not robust.

What Is a Robust Statistics?

- A statistical method is robust if the statistic is insensitive to slight departures from the assumptions that justify the use of the statistic.

- The robustness of a robust statistic can be measured by measures such as breakdown point, influence curve and gross error sensitivity.

- Consider the following example about formulating a 95% confidence interval for population mean to make it clear about the robustness.

An Example: 95% CI for the Population Mean

Definition (Confidence Interval for Mean)

- Suppose that we have a sample of size n
  - This sample must be obtained by randomization, either by a random sample or a randomized experiment.
  - The distribution of the data should be approximately Normal.
- Then a 95% confidence interval for a population mean μ is

$$\bar{X} \pm t_{n-1,0.975} \times \frac{s}{\sqrt{n}}$$

where $t_{n-1,0.975}$ corresponds to the 0.975-quantile of a t-distribution with $n-1$ degrees of freedom.

- Some may use the z-score $z_{0.975}$ instead of $t_{n-1,0.975}$. We'll not discuss about this matter here.
- The formula of construction CI for the population mean given above is well-known.
- However, do note that the constructed interval is good if the assumptions stated above are satisfied. If they are not satisfied strictly, then how?

Robustness of Assumptions

Definition (Robustness of Assumptions)

A statistical method is said to be **robust** with respect to a particular assumption if it performs adequately even when that assumption is modestly violated.

The procedure of using a t-distribution to compute confidence intervals for the mean of a population involves 2 assumptions (see previous slide).

- The assumption that data are obtained via a random sample or randomization is crucial. The procedure is **not robust** to this assumption.
- The assumption that the data are from a Normal population is not crucial. The procedure is **robust** to this assumption. We only need to ensure that there are no extreme outliers in the dataset and sample size n is relatively large enough, then we can proceed.

From the example about the robustness of assumptions above, similar understanding is applied to the robustness of statistical methods in general.

Location Parameter

- A measure of location of a sample is used as a simple summary of that sample–for example, "the average score of the midterm test was 17.5/30".

- There are some robust estimators of the location parameter.

- Among the robust estimators of the location parameter, the most commonly used is the arithmetic mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Sample Mean for a Normal Distribution Location

- Let $X_1, ..., X_n$ be IID $N(\mu, \sigma^2)$. The joint pdf is

$$f(x_1, ..., x_n|\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/(2\sigma^2)}$$

- The loglikelihood function is

$$l(\mu, \sigma) = -n\log(\sigma) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

- Taking partial derivatives w.r.t μ and then let it equal to 0

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

- Solving the equation above, we have the estimators

$$\hat{\mu} = \bar{x}$$

- Sample mean $\bar{x}$ is an estimator of the location parameter μ and it is the MLE estimator.

Robust Estimators of a Location Parameter

- Trimmed mean
- Winsorized mean
- M-Estimates (or Huber's M-estimates)
- Tukey's bisquare estimator
- Humpe's M-estimator

Example: Heats of Sublimation of Platinum

Consider a set of 26 measuments of the heat of sublimation of platinum from an experiment done by Hampson and Walker (1961). The measurements are in the unit of kcal/mol. The data are stored in the file `heats.csv`.

```
136.3 136.6 135.8 135.4 134.7 135.0 134.1 143.3 147.8 148.8 134.8 135.2
134.9 146.5 141.2 135.4 134.8 135.8 135.0 133.7 134.4 134.9 134.8 134.5
134.3 135.2
```

- The 26 measurements are all attempts to measure the “true” heat of sublimation, and we see that there is variability among them.
- Intuitively, it may seem that a measure of location or center for this batch of numbers would give a more accurate estimate of the heat of sublimation than any one of the numbers alone.
- The arithmetic mean for this sample is 137.05

Trimmed Mean

- Trimmed mean is a simple robust estimator of location.
- The $100\alpha\%$ trimmed mean calculated by: discarding the lowest $100\alpha\%$ and the highest $100\alpha\%$ and take the arithmetic mean of the remaining data.
- It is recommeded that we choose α from 0.1 to 0.2.
- For the heats data, the 20% trimmed mean is 135.29.

Winsorized Mean

- Trimmed mean is the mean of trimmed data. Winsorized mean is the mean of **trimmed and replaced** data.
- Winsorization replaces extreme data values with less extreme values.
- Let $[a]$ denote the nearest integer of number a; Let dataset of observations $x_1, ..., x_n$ are sorted by $x_{(1)}, ..., x_{(n)}$.
- The winsorized mean is computed after all the $[n\alpha]$ smallest observations are replaced by $x_{([n\alpha]+1)}$, and the $[n\alpha]$ largest observations are replaces by $x_{(n-[n\alpha]))}$.
- It is recommeded that we choose α from 0.1 to 0.2.
- For the heats data, the 20% Winsorized mean is 135.43.

M-Estimators for Location Parameter (1)

- We know that for a population with underlying distribution approximately normal, location parameter μ is estimated well by sample mean, $\bar{x}$.
- $\bar{x}$ is the MLE of μ, where $\bar{x}$ is found by minimizing

$$\sum_{i=1}^{n} \left(x_i - \mu\right)^2$$

  which is called by ''sum of squared errors".
- Huber (1964) (can find here: `https://projecteuclid.org/euclid.aoms/1177703732`) proposed that, when estimating the location parameter μ, one can obtain more robustness by another function of error than the sum of their squares.
- Hence, instead of minimizing the sum of squared error, Huber proposed we can find the estimator denoted by T - which is a function of $x_1, ..., x_n$ and this T is the minimizer of

$$\sum_{i=1}^{n} \rho(x_i - T)$$

  where ρ is a non-constant function that is meaningful.

M-Estimators for Location Parameter (2)

- The class of M-estimator proposed by Huber contains the sample mean, sample median and all the maximumlikelihood estimators:

- If we set function $\rho(x) = x^2$, the minimizer of $\sum_{i=1}^{n} \rho(x_i - T) = \sum_{i=1}^{n} (x_i - T)^2$ is $\bar{x}$;

- If we set function $\rho(x) = |x|$, the minimizer of $\sum_{i=1}^{n} \rho(x_i - T) = \sum_{i=1}^{n} |x_i - T|$ is the sample median;

- If we set function $\rho(x) = -\log f(x)$ where f is the assumed density function, the minimizer of $\sum_{i=1}^{n} \rho(x_i - T) = \sum_{i=1}^{n} -\log f(x_i - T)$ is the MLE;

M-Estimators for Location Parameter (3)

- If we set function $\rho(x)$ as

$$\rho(x) = \begin{cases} 1/2x^2 & \text{for} |x| \leq k \\ k|x| - 1/2k^2 & \text{for} |x| > k \end{cases}$$

  then the estimator corresponds to a Winsorized mean;

- If we set function $\rho(x)$ as

$$\rho(x) = \begin{cases} 1/2x^2 & \text{for} |x| \leq k \\ 1/2k^2 & \text{for} |x| > k \end{cases}$$

  then the estimator corresponds to a trimmed mean;

  The detailed proof can be found in the original paper by Huber (Annals of Mathematical Statistics, Volume 35, Number 1 (1964), 73-101).

Sample Standard Deviation

- The sample standard deviation which usually is denoted as s is a commonly used estimator of the population scale parameter, σ.

- However, the usual sample standard deviation is not robust, it is sesitive to outliers and may not remain bounded when a single data point is replaced by an arbitrary number.

- With robust scale estimators, the estimates remain bounded even when a portion of the data points are replaced by arbitrary numbers.

Interquartile Range (IQR)

- The interquartile range (IQR) can be used to estimate or measure the scale parameter, though it is not a robust estimator of σ.
- IQR is defined as $IQR = Q_3 - Q_1$ where Q_1 and Q_3 are the first and third quartiles respectively.
- For a normal distribution, the standard deviation σ can be estimated by dividing the interquartile range by 1.35. **Why?**

  Ans: In general, we can write $X = \mu + \sigma Z$ where μ and σ are the mean and standard deviation of X and Z has mean 0 and variance 1.

  Under the normality assumption, the IQR of a normal population X is $\sigma \times IQR(Z)$.

  Note that $Pr[Z < Q_1(Z)] = 0.25$ hence $Q1(Z) = -0.675$ and $Q_3(Z) = 0.675$. Thus, $IQR(Z) = 1.35$.

  Hence, given a dataset from a normal distribution, σ can be estimated by the ratio of the sample IQR and 1.35.

Median Absolute Deviation (MAD)

- The most popular robust estimator of scale parameter is MAD.

- $MAD = \text{median}_i(|x_i - \text{median}_j(x_j)|)$ where the inner median, $\text{median}_j(x_j)$ is the median of n observations and the outer median, median_i - the median of the n absolute values of the deviations about the median.

- For a normal distribution, $1.4826 \times MAD$ can be used to estimate the standard deviation σ. **Why?**

Ans:
Under the normality assumption, we can show that
$MAD(X) = \sigma \times MAD(Z)$ where $Z \sim N(0, 1)$. Therefore,
$\sigma = MAD(X)/MAD(Z)$.
Theoretically, we can get $MAD(Z) = 1/1.4826$.

Gini's Mean Difference

- The Gini's mean difference equals the mean of all absolute mutual differences of any two observation of the sample. Since we have $n(n-1)/2$ terms while the terms with outliers may be much less than $n(n-1)/2$ so that a robust estimator of σ may be expected using the Gini's mean difference.

- Gini's mean difference is defined as

$$G = \frac{1}{\binom{n}{2}} \sum_{i<j} |x_i - x_j|$$

- If the observations are from a normal distribution, then $\sqrt{\pi}G/2$ is an unbiased estimator of the standard deviation σ. **Why?**

  Ans: Under the normality assumption, one may show that $G(X) = \sigma G(Z)$ where $G(Z)$ is the Gini's mean difference of the standard normal distribution. It can be shown that $G(Z) = 2/\sqrt{\pi}$ asymptotically.

Robust Estimators: R (1)

```
> data<-read.csv("C:/Data/heats.csv")
> x = data$heat
> mean(x)
[1] 137.0462
> mean(x, trim = 0.2) # 20% trimmed mean
[1] 135.2875
> winsor<-function(x, alpha = 0.2)
+ { n = length(x); xq = n * alpha; x = sort(x)
+ m = x[(round(xq)+1)]; M = x[(n - round(xq))]
+ x[which(x<m)] = m; x[which(x>M)] = M
+ return(c(mean(x),var(x))) }
> winsor(x)
[1] 135.4269231   0.5868462
```

Robust Estimators: R (2)

```
> median(abs(x - median(x))) #MAD
[1] 0.65
> mad(x) # estimate of \sigma = 1.4826*MAD
[1] 0.96369
> IQR(x)
[1] 1.375
> library(MASS)
> hubers(x, k= 0.84) # this gives the 20% Winsorized mean
$mu
[1] 135.2469

$s
[1] 0.9935541
```

Robust Estimators: SPSS

"Analyze" → "Descriptive Statistics" → "Explore"... → choose "M-estimator" ...

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
heat	26	100.0%	0	0.0%	26	100.0%

M-Estimators

	Huber's M-Estimator[a]	Tukey's Biweight[b]	Hampel's M-Estimator[c]	Andrews' Wave[d]
heat	135.226	135.013	135.017	135.014

a. The weighting constant is 1.339.

b. The weighting constant is 4.685.

c. The weighting constants are 1.700, 3.400, and 8.500

d. The weighting constant is 1.340*pi.