On the equivalence of Pareto and Exponential Distribution

Fang Wang

January 26, 2020

Contents

1 Introduction **2**
- 1.1 A short summary what I have done and what I am plan to do . . . 2
  - 1.1.1 What I have done . . . 2
  - 1.1.2 What I am doing now . . . 2
  - 1.1.3 What I am aiming to finish in this semester . . . 2
- 1.2 Purpose and organization . . . 2
- 1.3 Contamination model and Slippage Alternative . . . 2

2 Discordancy Test for the Exponential Case **4**
- 2.1 Problem Setup . . . 4
- 2.2 Mathematical Preliminary . . . 4
- 2.3 Test Statistics and Its Distribution . . . 7
  - 2.3.1 Test and critical values . . . 12
- 2.4 Simulation study . . . 12
  - 2.4.1 Sample generation algorithm . . . 12
  - 2.4.2 Numerical verification of Theorem 1 . . . 13
  - 2.4.3 Critical values . . . 13
- 2.5 Power Estimation . . . 15

3 Discordancy Test for Pareto Case **16**
- 3.1 Introduction . . . 16
- 3.2 Test Statistics . . . 16
- 3.3 Distribution of JZ_r . . . 16
- 3.4 Distribution of JD_r . . . 17

4 On the Equivalence of two statistics **20**

Chapter 1

Introduction

1.1 A short summary what I have done and what I am plan to do

1.1.1 What I have done

I have been working on paper by A.Zerbet and M.Nikulin [1] and [2] by Mehdi Jabari Nooghbi for the last semester. I corrected the proof of the major theorem in [1] and filled in with necessary details, which is given as theorem 1. In particular, the proof given in [1] critically relies on a result given in [3], but our school does not have access to it, therefore I proved it as Lemma 4. I have also verified one of two major theorems proved in [2], which I gave as theorem 2 in the chapter 3.

With the correct result, I gave a correct estimation of power the test statistic proposed in [1] and corrected the critical values and gave it in Table 2.1. I found that the test they proposed in [1] does not have a better performance comparing with the classic benchmark test; in fact, they have very similar performance as shown in Figure 2.2.

1.1.2 What I am doing now

I noticed that the techniques used and results given in [2] and [1] are very similar, therefore, I want to prove that statistics tests proposed in theses two papers are equivalent in the sense they would have have the same power. Furthermore, I wish to simplifying the proof of a theorem given in [2] by showing it's a corollary of another theorem given in [1].

1.1.3 What I am aiming to finish in this semester

I want to give a leaner proof for theorems in [2] and estimate power for tests proposed with simulation study. If time permits, I will try to improve the test proposed and apply it to a real-life data set.

1.2 Purpose and organization

In this report I summarized the two statistics for testing slippage alternative hypothsis proposed by Chikkagoudar et.al. [1] and Zerbet et.al. [3]. The report is organized as follows: a short introduction to the problem and some mathemtical result is given in the first chapter; the test statistics used for the exponential samoles is given in the chapter 2; the test statistics for the Pareto samples are given in the chapter 3; the proof of both statistics are equivalent is given in the chapter 4.

1.3 Contamination model and Slippage Alternative

The outliers of a samples of observations, is a subset of observations appears to be inconsistent with the rest of the observations. A useful framework for examing the outliers is the contamination model, where we assume all samples arose independently from some disribution F and use statical procedure to examing the presence of observations from some other disribution $\overline{F}$, where $\overline{F} \neq F$. Here we give the definition of null hypothesis we will use for the contamination model.

Definition 1 (null hypothesis of contamination model)**.** Let $x_1, \ldots, x_n$ be a sample of n observations. Then under the null hypothsis H_0, $x_1, \ldots, x_n$ are observation of $X_1, \ldots, X_n$, where $X_1, \ldots, X_n$ are independent random variable with disribution F.

There are multiple ways to specify the alternative hypothsis for the null hypothesis H_0, we restrict ourself to the slippage alternative in this report. Slippage alterative hypothsis assuming that, except for a fixed r observation arose independently from disribution F, most observations arose independently from modifed version of F, say G. Since we only considered the location-shift disribution, we futher assume that the observations arose from G are the extreme observation in the samples. We formalize the slippage alternative hypothesis in the follwing definition:

Definition 2 (slippage alterative of the contamination model)**.** Let $x_1, \ldots, x_n$ be a sample of n observations with null hypothesis $x_1, \ldots, x_n$ arose independently from a disribution F. Let $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ be the order statistics of $x_1, \ldots, x_n$. Then under the slippage alterative H_r, the sample $x_{(1)}, \ldots x_{(n-r)}$ are independent observation disribution F and $x_{(n-r+1)}, \ldots, x_{(n)}$ are independent observation of disribution G with $F \neq G$.

Chapter 2

Discordancy Test for the Exponential Case

In this chapter we derive and summary the test statistics and corresponding discordancy test proposed by A.Zerbet and M.Nikulin [1].

2.1 Problem Setup

Recall the contamination model and slippage hypothesis introduced in the Chapter 1, here we consider the case where the underline disribution $F = \mathrm{Exp}(\theta)$ with density function $F(x;\theta) = \mathrm{P}(X \leqslant x) = 1 - \exp\{-\theta x\}$. and the contaminated model $G = \mathrm{Exp}(\theta/b)$ for some unknown constant b, where $0 < b \leqslant 1$.

Hence, with a sample of independent observation $X_1, \dots, X_n$, we wish to test the follwing hypothsis:

H_0: all $X_1, \dots, X_n$ are independent identically distributed $\mathrm{Exp}(\theta)$ random variable

against the slippage alterative hypothesis H_r:

$X_{(1)}, X_{(2)}, \dots, X_{(n-r)}$ derived from $\mathrm{Exp}(\theta)$ disribution

$X_{(1)}, X_{(2)}, \dots, X_{(n-r)}$ derived from $\mathrm{Exp}(\theta/b)$ disribution

2.2 Mathematical Preliminary

In this section we give some lemmas that will be used for derivation letter.

Lemma 1. *Let $\{a_k\}_{k=1}^m$ be a finite number of distinct real number; $x \in \mathbb{C}$, then*

$$\prod_{k=1}^{m} \frac{1}{x - a_k} = \sum_{n=1}^{m} \frac{1}{x - a_n} \prod_{i=1, i \neq n}^{m} \frac{1}{a_n - a_i}.$$

proof of the Lemma1. We will prove the result by induction on k; the case for $k = 1$ clearly holds; case for $k = 2$ also holds, since

$$\frac{1}{(x - a_1)(x - a_2)} = \frac{1}{(x - a_1)(a_1 - a_2)} + \frac{1}{(x - a_2)(a_2 - a_1)}.$$

Now suppose the case holds for $k = m - 1$ and consider the case for $k = m$. Since

$$
\begin{aligned}
\prod_{k=1}^{m} \frac{1}{x-a_k} &= \frac{1}{x-a_m} \prod_{k=1}^{m-1} \frac{1}{x-a_k} \\
&= \frac{1}{x-a_m} \sum_{n=1}^{m-1} \frac{1}{x-a_n} \prod_{i=1, i\neq n}^{m-1} \frac{1}{a_n-a_i} && \text{by induction hypothesis;} \\
&= \sum_{n=1}^{m-1} \left(\frac{1}{a_n-a_m} \frac{1}{x-a_n} + \frac{1}{x-a_m} \frac{1}{a_m-a_n} \right) \prod_{i=1, i\neq n}^{m-1} \frac{1}{a_n-a_i} \\
&= \left[\sum_{n=1}^{m-1} \frac{1}{x-a_n} \prod_{i=1, i\neq n}^{m} \frac{1}{a_n-a_i} \right] + \sum_{n=1}^{m-1} \frac{1}{x-a_m} \frac{1}{a_m-a_n} \prod_{i=1, i\neq n}^{m-1} \frac{1}{a_n-a_i} \\
&= \left[\sum_{n=1}^{m-1} \frac{1}{x-a_n} \prod_{i=1, i\neq n}^{m} \frac{1}{a_n-a_i} \right] + \frac{1}{x-a_m} \sum_{n=1}^{m-1} \frac{1}{a_m-a_n} \prod_{i=1, i\neq n}^{m-1} \frac{1}{a_n-a_i},
\end{aligned}
$$

where by induction hypothesis with $x = a_m$ that

$$
\begin{aligned}
\sum_{n=1}^{m-1} \frac{1}{a_m-a_n} \prod_{i=1, i\neq n}^{m-1} \frac{1}{a_n-a_i} &= \prod_{k=1}^{m-1} \frac{1}{a_m-a_k} \\
&= \prod_{i=1, i\neq m}^{m} \frac{1}{a_m-a_i}
\end{aligned}
$$

and therefore

$$
\begin{aligned}
\prod_{k=1}^{m} \frac{1}{x-a_k} &= \left[\sum_{n=1}^{m-1} \frac{1}{x-a_n} \prod_{i=1, i\neq n}^{m} \frac{1}{a_n-a_i} \right] + \frac{1}{x-a_m} \prod_{i=1, i\neq m}^{m} \frac{1}{a_m-a_i} \\
&= \sum_{n=1}^{m} \frac{1}{x-a_n} \prod_{i=1, i\neq n}^{m} \frac{1}{a_n-a_i}
\end{aligned}
$$

which proves the lemma. □

Lemma 2 (Inversion Formula [4]). *Let X be a continous random variable with charatersitic function $f(x)$ such that*

$$\int_{-\infty}^{\infty} |f(x)|\,\mathrm{d}x < \infty.$$

Then the pdf of X, $p(x)$, is given by

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t)\,\mathrm{d}t.$$

Lemma 3. *Let b and θ be real number and $\theta \neq 0$, then*

$$\int_0^{\infty} \frac{e^{-iwz}}{(b/\theta - iz)^r}\,\mathrm{d}z = \frac{2\pi w^{r-1}}{(r-1)!} e^{-wb/\theta}.$$

proof of the Lemma 3. Let X be a random variable follows Gamma$(r, \theta/b)$ distribution, the the pdf $p(x)$ and charatersitic function $f(x)$ of X is given by

$$p(x) = \frac{w^{r-1} e^{-w/b}}{\theta^{\alpha}}, \quad x > 0 \qquad \text{and} \qquad f(x) = (1 - ix\theta/b)^{-r}.$$

Then it follows from Lemma 2 that

$$
\begin{aligned}
2\pi p(w) &= \int_{-\infty}^{\infty} e^{-izw} f(t)\,\mathrm{d}z \\
2\pi \frac{w^{r-1}e^{-w\theta/b}}{(\theta/b)^r(r-1)!} &= \int_0^{\infty} \frac{e^{-iwz}}{(1-iz\theta/b)^r}\,\mathrm{d}z \\
2\pi \frac{w^{r-1}e^{-w\theta/b}}{(r-1)!} &= \int_0^{\infty} \frac{e^{-iwz}}{(b/\theta)^r(1-iz\theta/b)^r}\,\mathrm{d}z \\
2\pi \frac{w^{r-1}e^{-w\theta/b}}{(r-1)!} &= \int_0^{\infty} \frac{e^{-iwz}}{(b/\theta-iz)^r}\,\mathrm{d}z
\end{aligned}
$$

which proves the Lemma 3 □

Lemma 4. *Let $X_1,\dots,X_n$ be idependent* $\mathrm{Exp}(\theta)$ *random variable,* $X_{(1)},\dots X_{(n)}$ *be the order statistics of* $X_1,\dots,X_n$ *and* $Y_j = X_{(j)} - X_{(j-1)}$. *Assuming the slippage alternative H_r holds :*

$$
\begin{aligned}
&X_{(1)}, X_{(2)}, \dots, X_{(n-r)} \text{ derived from } \mathrm{Exp}(\theta); \\
&X_{(n-r+1)}, X_{(n-r+2)}, \dots, X_{(n)} \text{ derived from } \mathrm{Exp}(\theta/b), \\
&\ 0 < b \leqslant 1,\ b \text{ is unknown.}
\end{aligned}
$$

Then $Y_j \sim \mathrm{Exp}(\theta(rb+n-r-j+1))$ *for* $1 \leqslant j \leqslant n-r$; $Y_j \sim \mathrm{Exp}((\theta/b)(r-j+1)^{-1})$ *for* $1 \leqslant j \leqslant r$.

Proof of the Lemma 4. Let $\{N_t^k\}_{t\geq 0}$ be independent Poisson counting processes with stage space $\{0,1\}$, rate $1/\theta$ for $k \in \{1,\dots,n-r\}$ and rate $b/\theta/b$ for $k \in \{n-r+1,\dots n\}$. Then each X_k is the sojourn time of N_t^k in the stage 0. Let

$$
N_t = \sum_{k=1}^{n} N_t^k,
$$

then N_t is a Poisson processes with stage space $\{0,1,\dots,n\}$. Assuming $X_{(0)} = 0$, then $Y_i = X_{(i)} - X_{(i-1)}$ is the sojourn time of N_t in the stage $i-1$; that is,

$$
\begin{aligned}
\mathrm{P}(Y_i > t) &= \mathrm{P}(N_t < i \mid N_t = i-1) \\
&= \mathrm{P}\left(\sum_{k=i}^{n} N_t^k - \sum_{k=i}^{n} -N_0^k = 0\right) \\
&= \mathrm{P}\left(\sum_{k=i}^{n} N_t^k - N_0^k = 0\right) \\
&= \prod_{k=i}^{n} \mathrm{P}\left(N_t^k - N_0^k = 0\right).
\end{aligned}
$$

If we assume $i \in \{1,\dots,n-r\}$, then

$$
\begin{aligned}
\mathrm{P}(Y_i > t) &= \left[\prod_{k=n-r}^{n} \mathrm{P}\left(N_t^k - N_0^k = 0\right)\right]\left[\prod_{k=i}^{n-r} \mathrm{P}\left(N_t^k - N_0^k = 0\right)\right] \\
&= \left[\prod_{k=n-r}^{n} \exp\left\{-\frac{b}{\theta}t\right\}\right]\left[\prod_{k=i}^{n-r} \exp\left\{-\frac{t}{\theta}\right\}\right] \\
&= \exp\left\{-\frac{rb}{\theta}t - \frac{n-r+1-i}{\theta}t\right\}
\end{aligned}
$$

and $Y_i \sim \mathrm{Exp}(\theta(rb+n-r-i+1))$ as desired. If $i = n-r+j$ for $j \in \{1,\dots,r\}$ then

$$
\begin{aligned}
\mathrm{P}\left(Y_{n-r+j} > t\right) &= \prod_{k=n-r+j}^{n} \mathrm{P}\left(N_t^k - N_0^k = 0\right) \\
&= \exp\left\{-(r+j+1)\frac{b}{\theta}t\right\}
\end{aligned}
$$

and $Y_j \sim \mathrm{Exp}((\theta/b)(r-j+1)^{-1})$ for $1 \leqslant j \leqslant r$ as desired. □

2.3 Test Statistics and Its Distribution

Chikkagoudar et.al. proposed the following statistic to test H_r against H_0 and its distribution is derived in the Theorem 1.

$$Z_r = \frac{X_{(n-r)} - X_{(1)}}{\sum_{j=n-r+1}^{n} (X_{(j)}) - X_{(1)})}. \tag{2.1}$$

Theorem 1. *The distribution of the Statistic Zr under H_r is given by the formula:*

$$\begin{aligned} \mathrm{P}\left(Z_r < Z \mid H_r\right) = {} & \frac{b^r \Gamma(rb+n+r)}{\Gamma(rb+1)} \\ & \times \sum_{j=2}^{n-r} \frac{(-1)^{n-j-r}\{b^{-r} - [(rb+n-r-j+1)(z/(1-rz)) + b]^{-r}\}}{(j-2)!(n-j-r)!(rb+n-r-j+1)}, \qquad 0 < x < \frac{1}{r} \end{aligned}$$

Proof [1]. Suppose H_r holds and consider the random variable U_r given by

$$U_r = \frac{X_{(n-r)} - X_{(1)}}{\sum_{j=n-r+1}^{n} (X_{(j)}) - X_{(n-r)})},$$

then

$$\frac{1}{Z_r} - \frac{1}{U_r} = \frac{1}{X_{(n-r)} - X_{(1)}} \sum_{j=1}^{r} X_{(n-r)} - X_{(1)} = r \tag{2.2}$$

so $U_r = \frac{Z_r}{1-r}$ and we can obtain the disribution of Z_r by considering the disribution of U_r. Let

$$V = X_{(n-r)} - X_{(1)} \qquad \text{and} \qquad W = \sum_{j=n-r+1}^{n} X_{(j)} - X_{(n-r)}$$

then the test statistics $Z_r = V/W$. Let $Y_j = X_{(j)} - X_{(1)}$. Since

$$\begin{aligned} \sum_{j=2}^{n-r} Y_j &= \sum_{j=2}^{n-r} X_{(j)} - X_{(j-1)} \\ &= \sum_{j=2}^{n-r} X_{(j)} - \sum_{j=2}^{n-r} X_{(j-1)} \\ &= X_{(n-r)} - X_{(1)} + \sum_{j=1}^{n-r-1} X_{(j)} - \sum_{j=1}^{n-r-1} X_{(j)} \\ &= X_{(n-r)} - X_{(1)}. \end{aligned}$$

and

$$\begin{aligned}
\sum_{j=n-r+1}^{n}(n-j+1)Y_j &= \sum_{j=n-r+1}^{n} X_{(j)} - X_{(j-1)} \\
&= \sum_{j=n-r+1}^{n} X_{(j)} - \sum_{j=n-r}^{n-1} X_{(j-1)} \\
&= \sum_{j=n-r+1}^{n} X_{(j)} - \sum_{j=n-r}^{n-1} X_{(j-1)} \\
&= (n-n+1)X_{(n)} - (n-n+r-1+1)X_{(n-r+1)} + \sum_{j=n-r}^{n-1}(n-j+1)X_{(j)} - (n-j)X_j \\
&= X_{(n)} - rX_{(n-r)} + \sum_{j=n-r}^{n-1} X_{(j)} \\
&= \sum_{j=n-r}^{n} X_{(j)} - X_{(n-r)},
\end{aligned}$$

hence

$$Z_r = \frac{V}{W} = \frac{\sum_{j=2}^{n-r} Y_j}{\sum_{j=n-r+1}^{n}(n-j+1)Y_j}. \tag{2.3}$$

The joint charatersitic function (V, W) is

$$\begin{aligned}
\varphi_{V,W}(t,z) &= \mathrm{E}\left[\exp\{i(Vt+Wz)\}\right] \\
&= \mathrm{E}\left[\exp\left\{i(\sum_{j=2}^{n-r} Y_j t + \sum_{j=n-r+1}^{n}(n-j+1)Y_j z)\right\}\right] \\
&= \int_{\mathbb{R}^{n-1}} \exp\left\{i(\sum_{j=2}^{n-r} Y_j t + \sum_{j=n-r+1}^{n}(n-j+1)Y_j z)\right\} \\
&\quad \times f_{(Y_2,\ldots,Y_n)}(y_2,\ldots,y_n)\,\mathrm{d}y_2\cdots\mathrm{d}y_n
\end{aligned}$$

It follows from the Lemma 4 that Y_j, $j = 1, \cdots, n-r$ follows $\mathrm{Exp}(\theta(rb+n-r-j+1))$ and Y_{n-r+j} follows $\mathrm{Exp}((\theta/b)(r-$

$j+1)^{-1})$ disribution. Let $a_j = \theta(rb+n-r-j+1)^{-1}$ and $b_j = (\theta/b)(r-j+1)^{-1}$, we can then write $\varphi_{(V,W)}(t,z)$ as

$$
\begin{aligned}
\varphi_{V,W}(t,z) = & \int_{\mathbb{R}^{n-1}} \exp\left\{ it(\sum_{j=2}^{n-r} Y_j) \right\} \left[\prod_{k=2}^{n-r} \frac{1}{a_k} e^{-yk/ak} \right] \\
& \times \exp\left\{ iz \sum_{j=n-r+1}^{n} (n-j+1) y_j \right\} \left[\prod_{k=1}^{r} \frac{1}{b_k} e^{-y_{n-r+1}/b_k} \right] \mathrm{d}y_2 \cdots \mathrm{d}y_n \\
= & \prod_{j=2}^{n-r} \int_0^\infty \frac{1}{a_j} \exp\left\{ -\frac{y_j}{a_j} + ity_j \right\} \mathrm{d}y_j \\
& \times \prod_{j=1}^{r} \int_0^\infty \frac{1}{b_j} \exp\{ -iz(r-j+1) y_{n-r+j} - y_{n-r+j}/b_j \} \mathrm{d}y_{n-r+j} \\
= & \prod_{j=2}^{n-r} \int_0^\infty \frac{1}{a_j} \exp\left\{ -\frac{y_j}{a_j} + ity_j \right\} \mathrm{d}y_j \\
& \times \prod_{j=1}^{r} \int_0^\infty \frac{1}{b_j} \exp\left\{ -y_{n-r+j} \left(\frac{1}{b_j} - iz(r-j+1) \right) \right\} \mathrm{d}y_{n-r+j} \\
= & \left[\prod_{j=2}^{n-r} \frac{1}{a_j} (1/a_j - it)^{-1} \right] \times \left[\prod_{j=1}^{r} \frac{1}{b_j} (1/b_j - iz(r-j+1))^{-1} \right]
\end{aligned}
$$

It then follows from the Lemma 2 that the joint pdf of (V,W) is given by

$$
\begin{aligned}
f_{(V,W)}(v,w) &= \frac{1}{(2\pi)^2} \int_0^\infty \int_0^\infty \varphi_{(V,W)}(t,z) \exp\{it(tv+wz)\} \mathrm{d}t\,\mathrm{d}z \\
&= \frac{1}{(2\pi)^2} \int_0^\infty \left[\prod_{j=2}^{n-r} \frac{1}{a_j} (1/a_j - it)^{-1} \exp\{itv\} \mathrm{d}t \right] \times \int_0^\infty \left[\prod_{j=1}^{r} \frac{1}{b_j} (1/b_j - iz(r-j+1))^{-1} e^{-iwz} \mathrm{d}z \right]
\end{aligned}
\tag{2.4}
$$

To simplify (2.4), we first use Lemma 1 to calculate

$$
\begin{aligned}
\prod_{j=2}^{n-r} \frac{1}{1/a_j - it} &= (-1)^{n-r-1} \prod_{j=2}^{n-r} \frac{1}{it - 1/a_j} \\
&= (-1)^{n-r-1} \sum_{j=2}^{n-r} \frac{1}{it-1/a_j} \prod_{i\neq 1, i\neq j}^{n-r} \frac{1}{1/a_j - 1/a_i} \\
&= (-1)^{n-r-1} \sum_{j=2}^{n-r} \frac{1}{it-1/a_j} \prod_{i\neq 1, i\neq j}^{n-r} \frac{1}{\theta^{-1}(i-j)} \\
&= (-1)^{n-r-1} \sum_{j=2}^{n-r} \frac{\theta^{n-r-2}}{it-1/a_j} \left[\prod_{i=2}^{j-1} \frac{1}{1-j} \right] \left[\prod_{i=j+1}^{n-r} \frac{1}{1-j} \right] \\
&= (-1)^{n-r-1} \sum_{j=2}^{n-r} \frac{\theta^{n-r-2}}{it-1/a_j} \left[\frac{(-1)^{j+2}}{((j-2)!)} \right] \left[\frac{1}{(n-r-j)!} \right] \\
&= \sum_{j=2}^{n-r} \frac{(-1)^{n+j-r-1} \theta^{n-r-2}}{(it-1/a_j)(j-2)!(n-j-r)!};
\end{aligned}
$$

$$\begin{aligned}\prod_{j=1}^{r}(1/b_j - iz(r-j+1))^{-1} &= \prod_{j=1}^{r}\left(\frac{b}{\theta}(r-j-1) - iz(r-j+1)\right)^{-1}\\ &= \prod_{j=1}^{r}((b/\theta - iz)(r-j+1))^{-1}\\ &= (b/\theta - iz)^{r}\prod_{j=1}^{r}(r-j+1)^{-1}\\ &= \frac{(b/\theta - iz)^{r}}{r!};\end{aligned}$$

$$\begin{aligned}\prod_{j=1}^{r}\frac{1}{a_j} &= \prod_{j=2}^{n-r}\frac{rb+n-r-j+1}{\theta}\\ &= \frac{1}{\theta^{n-r-1}}\frac{\Gamma(rb+n-r)}{\Gamma(rb+1)};\end{aligned}$$

and

$$\prod_{j=1}^{r}\frac{1}{b_j} = \prod_{j=1}^{r}\frac{b}{\theta}(r-j+1) = \left(\frac{b}{\theta}\right)^{r} r! \tag{2.5}$$

Then can express the $f_{V,W}(v,w)$ given in (2.4) as

$$f_{(V,W)}(v,w) = \left[\frac{1}{(2\pi)^2}\sum_{j=2}^{n-r}\frac{\Gamma(rb+n-r)(-1)^{n+j-r-1}}{\Gamma(rb+1)(j-2)!(n-j-r)!}\int_0^{\infty}\frac{e^{-iv}}{it-1/a_j}\,\mathrm{d}t\right]\times\left[\left(\frac{b}{\theta}\right)^{r+1}\int_0^{\infty}\frac{e^{-iwz}}{(b/\theta - iz)^{r}}\,\mathrm{d}z\right].$$

We claim that

$$\int_0^{\infty}\frac{e^{-iwz}}{(b/\theta - iz)^{r}}\,\mathrm{d}z = \frac{2\pi w^{r-1}}{(r-1)!}e^{-wb/\theta}..$$

To show the the above inteegration identity holds, let X be a random variable follows Gamma$(r,\theta/b)$ distribution, the the pdf $p(x)$ and charatersitic function $f(x)$ of X is given by

$$p(x) = \frac{w^{r-1}e^{-w/b}}{\theta^{\alpha}}, \quad x>0 \qquad \text{and} \qquad f(x) = (1-ix\theta/b)^{-r}.$$

Then it follows from Lemma 2 that

$$\begin{aligned}2\pi p(w) &= \int_{-\infty}^{\infty} e^{-izw} f(t)\,\mathrm{d}z\\ 2\pi\frac{w^{r-1}e^{-w\theta/b}}{(\theta/b)^{r}(r-1)!} &= \int_0^{\infty}\frac{e^{-iwz}}{(1-iz\theta/b)^{r}}\,\mathrm{d}z\\ 2\pi\frac{w^{r-1}e^{-w\theta/b}}{(r-1)!} &= \int_0^{\infty}\frac{e^{-iwz}}{(b/\theta)^{r}(1-iz\theta/b)^{r}}\,\mathrm{d}z\\ 2\pi\frac{w^{r-1}e^{-w\theta/b}}{(r-1)!} &= \int_0^{\infty}\frac{e^{-iwz}}{(b/\theta - iz)^{r}}\,\mathrm{d}z\end{aligned}$$

which proves the claim.

Since $Z_r = V/W$, therefore we can obtain the disribution of Z_r by direct integrate $f_{V,W}(v,w)$ as follows.

$$\begin{aligned}\mathrm{P}(U_r < u) &= \mathrm{P}\left(\frac{V}{W} < u\right)\\ &= \mathrm{P}(V < uW)\\ &= \int_0^{\infty}\int_0^{uw} f_{(V,W)}(v,w)\,\mathrm{d}v\,\mathrm{d}w.\end{aligned}$$

It follows from the Lemma 2 and the charatersitic function of $\operatorname{Exp}(1/a_j)$ that

$$\int_0^\infty \frac{e^{-itv}}{it-1/a_j}\,\mathrm{d}t = -a_j \int_0^\infty \frac{e^{-itv}}{1-ita_j}\,\mathrm{d}t = -2\pi \exp\left\{-\frac{1}{a_j}v\right\},$$

therefore with Proposition 2 we can simplify $f_{V,W}(v,w)$ to

$$\begin{aligned} f_{V,W}(v,w) = &\frac{\Gamma(rb+n-r)}{\Gamma(rb+1)}\left(\frac{b}{\theta}\right)^{r+1}\frac{1}{(r-1)!} \\ &\times \sum_{j=2}^{n-r}\frac{(-1)^{n+j-r}}{(j-2)!(n-j-r)!}g(v,w) \end{aligned}$$

where

$$g(v,w) = w^{r-1}\exp\left\{-\frac{v}{a_j}-\frac{b}{\theta}w\right\}.$$

Since

$$\begin{aligned} \int_0^\infty \int_0^{uw} g(v,w)\,\mathrm{d}v\,\mathrm{d}w &= \int_0^\infty w^{r-1}\exp\left\{-\frac{b}{\theta}w\right\}(-a_j)\left[\exp\left\{-\frac{u}{a_j}w\right\}-1\right]\mathrm{d}w \\ &= -a_j\left[\int_0^\infty w^{r-1}\exp\left\{-\left(\frac{b}{\theta}+\frac{1}{a_j}u\right)w\right\}\mathrm{d}w\right. \\ &\quad \left. -\int_0^\infty w^{r-1}\exp\left\{-\frac{b}{\theta}w\right\}\mathrm{d}w\right] \\ &= -a_j\left[\Gamma(r)\left(\frac{b}{\theta}+\frac{u}{a_j}\right)^{-r} - \Gamma(r)\left(\frac{b}{\theta}\right)^{-r}\right] \\ &= -a_j(r-1)!\left[\left(\frac{b+u\theta/a_j}{\theta}\right)^{-r} - \left(\frac{b}{\theta}\right)^{-r}\right] \\ &= a_j(r-1)!\theta^r\left[b^{-r}-[u\theta/a_j+b]^{-r}\right] \end{aligned}$$

and therefore

$$\begin{aligned} \int_0^\infty \int_0^{uw} f_{(V,W)}(v,w)\,\mathrm{d}v\,\mathrm{d}w &= \frac{\Gamma(rb+n-r)}{\Gamma(rb+1)}\left(\frac{b}{\theta}\right)^{r+1}\frac{1}{(r-1)!}\times\sum_{j=2}^{n-r}\frac{(-1)^{n+j-r}}{(j-2)!(n-j-r)!}\int_0^\infty\int_0^{uw} g(v,w)\,\mathrm{d}v\,\mathrm{d}w \\ &= \frac{\Gamma(rb+n-r)}{\Gamma(rb+1)}\left(\frac{b}{\theta}\right)^{r+1}\frac{1}{(r-1)!}\sum_{j=2}^{n-r}\frac{(-1)^{n-j-r}}{(j-2)!(n-j-r)!} \\ &\quad \times a_j(r-1)!\theta^r\left[b^{-r}-[u\theta/a_j+b]^{-r}\right] \\ &= \frac{\Gamma(rb+n-r)}{\Gamma(rb+1)}\frac{b^{r+1}}{\theta}\sum_{j=2}^{n-r}\frac{(-1)^{n-j-r}}{(j-2)!(n-j-r)!}a_j\left\{b^{-r}-[u\theta/a_j+b]^{-r}\right\} \\ &= \frac{b^{r+1}\Gamma(rb+n-r)}{\Gamma(rb+1)}\sum_{j=2}^{n-r}\frac{(-1)^{n-j-r}}{(j-2)!(n-j-r)!}\frac{\left[b^{-r}-(u\theta/a_j+b)^{-r}\right]}{\theta/a_j} \end{aligned}$$

Then using the fact $\theta/a_j = (rb+n-r-j+1)$ the equality above becomes

$$\mathrm{P}(U_r<u) = \int_0^\infty\int_0^{uw} f_{(V,W)}(v,w)\,\mathrm{d}v\,\mathrm{d}w = \frac{b\Gamma(rb+n-r)}{\Gamma(rb+1)}\sum_{j=2}^{n-r}\frac{(-1)^{n-j-r}\{b^{-r}-[(rb+n-r-j+1)u+b]^{-1}\}}{(j-2)!(n-j-r)!(rb+n-r-j+1)}.$$

and the result follows from $U_r = \frac{Z_r}{1-r}$. $\square$

Remark. We remark that there is a mistake in the derivation of the previous theorem given in [1], their disribution is differ from the correct disribution by a factor of b. This mistake is rather severe, since this eventually lead to a wrong conclusion drawn by the author. We verified this numerically in the next section. In fact, this mistake has been corrected by Mehdi Jabbari Nooghabi, he proved a analogous result for Pareto disribution as theorem 3 in [2], where b is corresponding to β and r is corresponding to k.

2.3.1 Test and critical values

Since we are assuming $b < 1$ and under slippage alternative the sample $X_{(n-r+1)}, \dots, X_{(n)} \sim \mathrm{Exp}(\theta/b)$, and therefore the test statistic Z_r given in (2.1) has the property that

$$\begin{aligned}\lim_{b\to\infty} \mathrm{E}\left[Z_r\right] &= \lim_{b\to\infty} \mathrm{E}\left[\frac{X_{(n-r)} - X_{(1)}}{\sum_{j=n-r+1}^{n}(X_{(j)}) - X_{(1)})}\right] \\ &\leqslant \lim_{b\to\infty} \mathrm{E}\left[\frac{X_{(n-r)}}{X_{(n)} - X_{(1)}}\right] \\ &= 0,\end{aligned}$$

for $X_{(n-r)}$ and $X_{(1)}$ are samples arose from $\mathrm{Exp}(\theta)$ and $X_{(n)}$ arose from $\mathrm{Exp}(\theta/b)$ under the slippage alternative hypothesis H_r. Therefore, one should reject H_0 and if $Z_r > z_c$, where $z_c(\alpha)$ is the significant level corresponding to the significance level α. We have corrected the direction of the test claimed in [1], where they claimed the test direction is a upper tail test.

2.4 Simulation study

As we have pointed out in the previous section, the A.Zerbet and M. Nikulin made a mistake in the derivation and presented a wrong method to calculate the critical values. In this chapter we will first use numerical methods to verify the theorem 1, and then give a table of correct critical values as well as the correct power estimation.

2.4.1 Sample generation algorithm

We used the following algorithm to generate the samples under H_r.

Algorithm 1 Algorithm for generating sample under H_r

Input:
sample size : n
number of contaminated sample: r
the parameter for contaminated sample : b, θ, where $b \in (0,1), \theta > 0$.
Output: A sample $\mathbf{X} = (x_1, \dots, x_n)$ under H_r.

repeat
    generating sample $\mathbf{x} = (x_1, \dots, x_{n-r}) \overset{iid}{\sim} \mathrm{Exp}(\theta)$
    generating sample $\mathbf{x}^b(x_{n-r+1}, \dots, x_n) \overset{iid}{\sim} \mathrm{Exp}(\theta/b)$
until $\max \mathbf{x} < \min \mathbf{x}^b$
$\mathbf{x} := (x_1, \dots, x_{n-r}, x_{n-r+1}, \dots, x_n)$
return x.

The disribution of the samples generated by algorithm 1 is indeed the samples under H_r. Let $\mathbf{X} = (X_1, \dots, X_n)$ with $(X_1, \dots, X_{n-p})$ coming from $\mathrm{Exp}(\theta)$ and $(X_{p+1}, \dots, X_n)$ comes from $\mathrm{Exp}(\theta/b)$. Then

$$\begin{aligned}\mathrm{P}\left(X_{(k)} \leqslant x \mid H_r\right) &= \mathrm{P}\left(X_{(k)} \leqslant x \mid \max\{X_1, \dots, X_{n-p}\} < \min\{X_{n-p+1}, \dots, X_n\}\right) \\ &= \mathrm{P}\left(X_{(k)} \leqslant x \mid \text{accept } \mathbf{X}\right)\end{aligned}$$

We remark that, the efficiency of the algorithm 1 decrease as b tends to 1.

The efficiency of the algorithm

Since the algorithm generate samples by accepting samples that satisfy H_r from samples with contamination but not necessarily satisfy H_r. A natural question would be, how efficient is our algorithm?

Let $Y = \max\{X_1, \ldots, X_{n-p}\}$ and $Z = \min\{X_{n-p}, \ldots, X_n\}$. It then follows from the

Then the sample $\mathbf{X}$ is accepted only if $Y < X$. It then follows from Theorem 5.4.4 of [5] that the pdf for Y and Z is given by

$$f_z(y) = (n-r)[1 - e^{y/\theta}]^{n-r-1} \frac{e^{-y/\theta}}{\theta}$$

and

$$f_Z(z) = r(e^{-bz/\theta})^{r-1} \frac{be^{-bz/\theta}}{\theta}.$$

Hence the accepting probability is given by

$$\begin{aligned}
\mathrm{P}(Y < z) &= \int_0^\infty \mathrm{P}(Z < z \mid Z = z) f_Z(z)\, \mathrm{d}z \\
&= \int_0^\infty (1 - e^{-z/\theta})^{n-r} r \exp\{-zb/\theta\}^{r-1} \frac{be^{-zb/\theta}}{\theta}\, \mathrm{d}z \\
&= \int_0^\infty \frac{rb}{\theta} [1 - \exp\{-z/\theta\}]^{n-r} \exp\{-rbz/\theta\}\, \mathrm{d}z \\
&= \int_0^1 rb[1-w]^{n-r} w^{rb} w^{-1}\, \mathrm{d}w \\
&= rbB(rb, n-r+1),
\end{aligned}$$

where $B(r,s)$ is the beta function. For the special case where $r = 1$, the accepting probability becomes

$$\begin{aligned}
\mathrm{P}(Y < Z) &= rB(r, n-r+1) \\
&= \frac{\Gamma(r)\Gamma(n-r+1)}{\Gamma(n+1)} \\
&= \binom{n}{r}^{-1},
\end{aligned}$$

which rapid decreases to 0 as n increase. Besides giving an estimation for the efficiency of tha algorithm 1, the accepting probability also gives an measure for whether the slippage alternative is appropriate. It could be interpret as the probability we get a sample $(x_1, \ldots, x_n)$ that satisfy H_r if $x_1, \ldots, x_{n-r}$ arose from $\mathrm{Exp}(\theta)$ disribution and $x_{n-r+1}, \ldots, x_n$ arose from $\mathrm{Exp}(\theta b)$ disribution.

2.4.2 Numerical verification of Theorem 1

We numerically verified the theorem 1. This was done by generating 10,000 samples for the case $n = 12, r = 3, b = 1/2$. Then we plot the empirical disribution function (edf) and the derived cdf in the Figure2.1, the empirical disribution agrees with the result we obtained, and the mistake in the [1] is quite evident.

2.4.3 Critical values

As pointed in the previous section, [1] estiamte the critical values in the wrong direction, here we used the bisection method to obtain the correct table of critical values. The code we used is given in the appendix.

To find the critical values and estimate the power for Dixon's statistic, we generated samples with algorithm 1 and then the critical values are found with Monte Carlo methods. The bias and standard error was then estimated using Bootstrap resampling method.

The critical values for Dixon statistics are given in the Table2.2 and the estimate standard error is given in the table 2.3. As can be seen from the Table2.3, the critical values we obtained is accurate to three decimal places, and this agrees with the result given in [1].

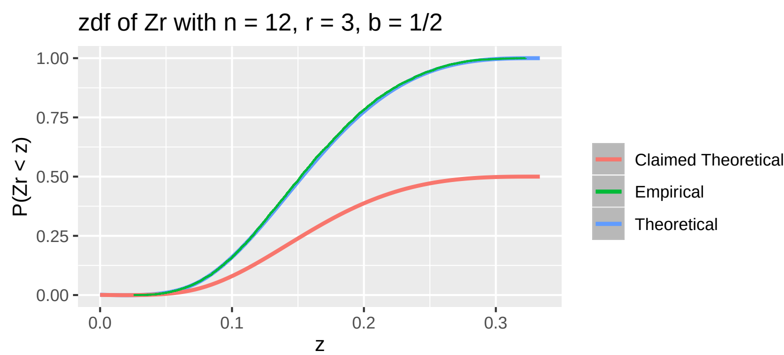

Figure 2.1: The CDF of Z_r for $n = 12, r = 3, b = 1/2$

Table 2.1: The critical values for Z_r for $\alpha = 0.05$

n	r					
	1	2	3	4	5	6
6	0.2179255	0.07271396	0.02257252	0.002554801		
7	0.2541362	0.09761256	0.04158413	0.014258365	0.001703935	
8	0.2827005	0.11738195	0.05767611	0.027187769	0.009843320	0.001217544
9	0.3059432	0.13338088	0.07094024	0.038625121	0.019246229	0.007211060
10	0.3253324	0.14660659	0.08194491	0.048345371	0.027852283	0.014371925
11	0.3418340	0.15775129	0.09119986	0.056587249	0.035351931	0.021105592
12	0.3561090	0.16729823	0.09909488	0.063629961	0.041830932	0.027096803

Table 2.2: The estimated critical values for D_r for $\alpha = 0.05$ with Monte Carlo method

n	r					
	1	2	3	4	5	6
6	0.7451293	0.8613298	0.9295339	0.9721648		
7	0.7174043	0.8333060	0.8997864	0.9454283	0.9782023	
8	0.6937633	0.8084582	0.8758362	0.9217053	0.9569222	0.9819938
9	0.6748915	0.7878169	0.8512355	0.9002023	0.9363351	0.9643261
10	0.6572173	0.7696995	0.8354201	0.8819363	0.9175486	0.9458965
11	0.6438796	0.7539956	0.8176284	0.8643931	0.9012357	0.9296735
12	0.6313994	0.7392545	0.8037565	0.8488094	0.8850763	0.9146531

Table 2.3: The standard error of the estimate of critical values for D_r for $\alpha = 0.05$

n	r					
	1	2	3	4	5	6
6	0.0003654004	0.0006721107	0.0004517287	0.0002250101		
7	0.0003930877	0.0007751294	0.0005556101	0.0003343549	0.0001549828	
8	0.0004040385	0.0008054345	0.0005574572	0.0003964377	0.0002388118	0.0001299860
9	0.0003597688	0.0008568465	0.0006336613	0.0004514928	0.0003180044	0.0002395700
10	0.0003644383	0.0008356480	0.0006315783	0.0004870168	0.0004525706	0.0002701931
11	0.0003900676	0.0008706266	0.0007761038	0.0005363027	0.0004374640	0.0003245950
12	0.0003680189	0.0008683104	0.0007025402	0.0005738111	0.0004305760	0.0003728830

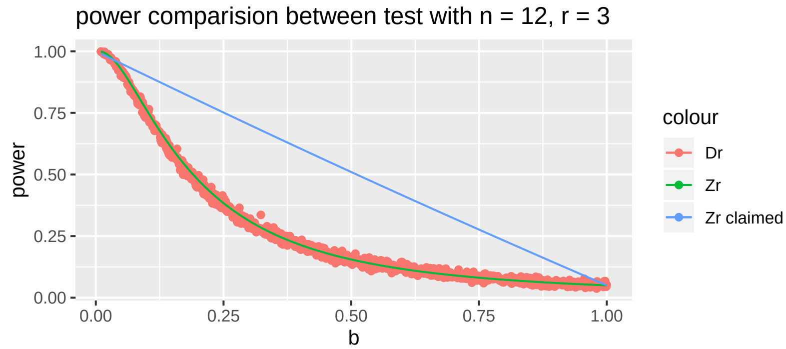

Figure 2.2: The power of Z_r and D_r for $n = 12, r = 3$

2.5 Power Estimation

To properly evaluate the the statistic test with Z_r, we conduct a power comparison with Dixon's statistic given by

$$D_r = \frac{X_{(n)} - X_{(n-r)}}{X_{(n)}}.$$

Since we have derived the exact disribution of Z_r in Theorem 1, therefore we the power function for Z_r can be calculate by

$$\gamma_{Z_r}(b) = \mathrm{P}\left(Z_r < z_\alpha \mid H_r\right) = 1 - F_{Z_r}(z_\alpha; b),$$

where z_α is the corresponding critical values found in the Table 2.1.

Then we used Monte Carlo method to evaluate the power of the Dixon's statistic for different choice of b and we plot it against the power of Z_r for the case of $n = 12$ and $r = 3$. For the sake of the comparison, we also ploted the wrong power estimated given in [1], which is given in the Figure 2.2.

As can be seen from the Figure 2.2, the test proposed in [1] has very similar power comparring with the classical Dixon's statistic, and conclusion given in [1] is invalid.

Chapter 3

Discordancy Test for Pareto Case

In chapter 2 we investigated a statistic test proposed by Nikulin et.al to test the slippage alterative hypothsis for the samples with exponential samples. It turns out that the same kind of statistics can be easily adapat to test the slippage alternative for Pareto disribution. The statistics and test introduced is proposed by Mehdi Jabbari Nooghabi [2].

3.1 Introduction

First we give a definition for Pareto disribution.

Definition 3 (Pareto disribution)**.** A random variable X follows Pareto(α,θ) disribution if its pdf is given by

$$f(x;\alpha,\theta) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}} I_{\{x\geqslant\theta>0\}}$$

where θ and α are both positive.

Let $X_1,\ldots,X_n$ be independent Pareto random variables and suppose we wish to test the slippage alternative hypothesis with H_0 $X_1,\ldots,X_n$ follows Pareto(α,θ) for some unknown α and θ against the slippage alterative H_r that $X_{(1)},\ldots,X_{(n-r)}$ arose from Pareto(α,θ) and $X_{(n-r+1)},\ldots X_{(n)}$ arose from Pareto$(\alpha\beta,\theta)$ disribution for some $\beta>1$.

3.2 Test Statistics

Mehdi Jabbari Nooghabi proposed the statistic JZ_k and JD_r to test H_r against H_0 for Pareto samples, where

$$JZ_r = \frac{\ln\left(X_{(n-r)}\right) - \ln\left(X_{(1)}\right)}{\sum_{j=n-r+1}^{n}(\ln\left(X_{(j)}\right) - \ln\left(X_{(1)}\right))}$$

and

$$JD_r = \frac{\ln\left(X_{(n-r)}\right) - \ln\left(X_{(1)}\right)}{(\ln\left(X_{(n)}\right) - \ln\left(X_{(n-r+1)}\right))}.$$

The main goal of this chapter is to derive the disribution of JZ_k and JD_r. Mehdi Jabbari Nooghabi proved the result in the same as Chikkagoudar et.al did.

3.3 Distribution of JZ_r

In this section we will prove the disribution of JZ_r.

Lemma 5. *Suppose* $X \sim$ Pareto(α,θ) *and* $Y = \log(X/\theta)$*, then* $Y \sim$ Exp(α).

Proof. Let f_X and f_Y be the pdf of X and Y respectively, then we have

$$f(x) = \alpha\theta^{\alpha} x^{-(\alpha+1)}$$

for $x \geqslant 0$. Since the function $g : x \mapsto \log(x/\theta)$ is a bijection on the support of X with inverse $g^{-1} : y \mapsto \theta e^{y}$. Then the pdf of Y is given by

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{\mathrm{d}g^{-1}}{\mathrm{d}y} \right| \\ &= \alpha\theta^{\alpha}[\theta e^{y}]^{-(\alpha+1)} |\theta e^{y}| \\ &= \alpha e^{-\alpha y} \end{aligned}$$

which is a pdf of $\mathrm{Exp}(\alpha)$ random variable and the lemma is proved. □

3.4 Distribution of JD_r

In this section we will prove the disribution of JZ_r.

Theorem 2. *Let $X_1, \ldots, X_n$ be a collection of idependent Pareto random variable, and under slippage alternative hypothsis H_r given in the section 3.1. The the disribution of statistics JZ_r is given by*

$$\begin{aligned} \mathrm{P}\left(JZ_r < Z \mid H_r\right) = &\frac{\beta^r \Gamma(r\beta + n - r)}{\Gamma(rb+1)} \\ &\times \sum_{j=2}^{n-r} \frac{(-1)^{n-j-r}\{\beta^{-r} - [(r\beta + n - r - j + 1)(x/(1-rx)) + \beta]^{-r}\}}{(j-2)!(n-j-r)!(r\beta + n - r - j + 1)} \qquad 0 < x < \frac{1}{r} \end{aligned}$$

Proof. We will prove the result in the same way the Theorem 1 is proved. Suppose H_r holds and define U_r to be

$$U_r = \frac{\ln\left(X_{(n-r)}\right) - \ln\left(Z_{(1)}\right)}{\sum_{j=n-r+1}^{n} (\ln\left(X_{(j)}\right)) - \ln\left(X_{(n-r)}\right)} = \frac{V}{W} \tag{3.1}$$

Let $Y_j = \ln\left(X_{(j)}\right) - \ln\left(X_{(j-1)}\right)$, it follows from the lemma 5 and lemma 4 that Y_j, $j = 1, \cdots, n-r$ follows $\mathrm{Exp}(\alpha(r\beta + n - r - j + 1))$ and Y_{n-r+j} follows $\mathrm{Exp}((\alpha\beta)(r-j+1)^{-1})$ disribution. This shows that each Y_j is of the same disribution as Y_j defined in the proof of Theorem 1. The only difference is we used θ and α rather than α and θ as the parameter. It then follows from (2.3) and (3.1) that U_r and Z_r has the exactly same disribution. Then it follows from (2.2) and Theorem 1 that the disribution of JZ_r is given by

$$\begin{aligned} \mathrm{P}\left(JZ_r < Z \mid H_r\right) = &\frac{\beta^r \Gamma(r\beta + n - r)}{\Gamma(rb+1)} \\ &\times \sum_{j=2}^{n-r} \frac{(-1)^{n-j-r}\{\beta^{-r} - [(r\beta + n - r - j + 1)(x/(1-rx)) + \beta]^{-r}\}}{(j-2)!(n-j-r)!(r\beta + n - r - j + 1)} \qquad 0 < x < \frac{1}{r} \end{aligned}$$

□

Theorem 3. *Let $X_1, \ldots, X_n$ be a collection of idependent Pareto random variable, and under slippage alternative hypothsis H_r given in the section 3.1. The the disribution of statistics JZ_k is given by*

$$\begin{aligned} \mathrm{P}\left(JZ_r < Z \mid H_r\right) = &\frac{\beta^r \Gamma(r\beta + n - r)}{\Gamma(rb+1)} \\ &\times \sum_{j=2}^{n-r} \frac{(-1)^{n-j-r}\{\beta^{-r} - [(r\beta + n - r - j + 1)(x/(1-rx)) + \beta]^{-r}\}}{(j-2)!(n-j-r)!(r\beta + n - r - j + 1)} \qquad 0 < x < \frac{1}{r} \end{aligned}$$

Proof. We will prove the result with method of characteristic function. Let $X_{(1)},\ldots,X_{(n)}$ be the order statistic of the samples under $\overline{H}_r$. Let

$$R_r = \frac{\sum_{j=2}^{n-k} Y_j}{\sum_{j=n-k+2}^{n} Y_j} = \frac{P}{Q}$$

where $Y_j = \ln\big(X_{(j)}\big) - \ln\big(X_{(j-1)}\big)$ for $j > 1$. It then follows from lemma 4 and lemma 5 that the joint disribution of P, Q is given by

$$\begin{aligned}
\phi_{P,Q}(t,s) &= \prod_{j=2}^{n-r} \phi_{Y_j} \prod_{j=2}^{r} \phi_{Y_j} \\
&= \prod_{j=2}^{n-r} \left[\frac{1}{a_j} \left(\frac{1}{a_j} - it \right) \right] \prod_{j=2}^{r} \left[\frac{1}{b_j} \left(\frac{1}{b_j} - is \right) \right]
\end{aligned}$$

where $a_j = [\alpha(r\beta + n - r - j + 1)]^{-1}$ and $b_j = [\alpha\beta(r - j + 1)]^{-1}$. Then it follows from lemma 2 that the joint density function of U, V is then given by

$$f_{P,Q}(p,q) = \frac{1}{(2\pi)^2} \int_0^\infty \left[\prod_{j=2}^{n-r} \frac{1}{a_j} (1/a_j - it)^{-1} \exp\{-itp\} \,\mathrm{d}t \right] \times \int_0^\infty \left[\prod_{j=2}^{n-r} \frac{1}{b_j} (1/b_j - it)^{-1} \exp\{-isq\} \,\mathrm{d}t \right].$$

To simplify $f_{P,Q}$, we first use lemma 1 to calculate

$$\begin{aligned}
\prod_{j=2}^{k} \frac{1}{1/b_j - is} &= (-1)^r \prod_{j=2}^{r} \frac{1}{it - 1/b_j} \\
&= (-1)^{r-1} \sum_{j=2}^{r} \frac{1}{is - 1/b_j} \prod_{i=2}^{k} \frac{1}{\alpha\beta(k-i)} \\
&= (-1)^{r-1} \frac{1}{(\alpha\beta)^{r-2}} \sum_{j=2}^{k} \frac{1}{is - 1/b_j} \prod_{k=2}^{r} \frac{1}{k-i} \\
&= (-1)^{r-1} \frac{1}{(\alpha\beta)^{r-2}} \sum_{j=2}^{k} \frac{1}{is - 1/b_j} \frac{(-1)^{j-2}}{(j-2)!(k-j)!} \\
&= \sum_{j=2}^{r} \frac{(-1)^{r+j-1}}{(is - 1/b_j)(j-2)(r-j)!(\alpha\beta)^{r-2}}.
\end{aligned}$$

It follows from (2.5) that

$$\prod_{j=2}^{r} \frac{1}{b_j} = (r-1)!(\alpha\beta)^{r-1}.$$

Therefore with the product identities derived in the proof for Theorem 1 we obtained the simplified pdf for $f(p,q)$ is given by

$$f_{P,Q}(p,q) = \frac{\alpha\Gamma(r\beta + n - r)}{\Gamma(r\beta + 1)} \sum_{j=2}^{n-r} \frac{(-1)^{n-r+j-1}}{(j-2)(n-r-j)!} \exp\{-n_j p\} \times \alpha\beta(r-1)! \sum_{i=1}^{r} \frac{(-1)^{n-r+j-1}}{(r-1)!(i-2)!} \exp\{-m_i q\}$$

where $n_j = -\alpha(r\beta + n - r - j + 1)$ and $m_i = -\alpha\beta(r - i + 1)$. Then the disribution for R is given by

$$\begin{aligned}
\mathrm{P}\left(R_r < x\right) &= \frac{\alpha^2 \beta \Gamma(r\beta + n - r)}{\Gamma(r\beta + 1)} \sum_{j=2}^{n-k} \sum_{i=2}^{r} \frac{-1^{n+j+i-2}}{(n-r-j)!(j-2)!(r-1)!(i-2)!} \int_0^\infty \int_0^{rq} \exp\{-\alpha n_j p + -\alpha m_i q\} \,\mathrm{d}p \,\mathrm{d}q \\
&= \frac{\alpha^2 \beta \Gamma(r\beta + n - r)}{\Gamma(r\beta + 1)} \sum_{j=2}^{n-k} \sum_{i=2}^{r} \frac{-1^{n+j+i-2}}{(n-r-j)!(j-2)!(r-1)!(i-2)!} \alpha^{-2} \frac{x}{m_i(n_j x + m_i)} \\
&= \frac{\beta^r \Gamma(r\beta + n - r)}{\Gamma(rb + 1)} \sum_{j=2}^{n-r} \sum_{i=2}^{r} \frac{(-1)^{n+i+j-2}}{(n-r-j)!(j-2)!(r-i)!(i-2)!} \times \frac{m_i^{-1} - (m_i + x n_j)^{-1}}{n_i} \\
&= \frac{\beta^r \Gamma(r\beta + n - r)}{\Gamma(rb + 1)} \sum_{j=2}^{n-r} \sum_{i=2}^{r} \frac{(-1)^{n+i+j-2}}{(n-r-j)!(j-2)!(r-i)!(i-2)!}. \\
&\times \frac{[\beta(r-i+1)]^{-1} - [\beta(x-i+1) + x(r\beta + n - r - j + 1)]^{-1}}{r\beta + n - r - j + 1}
\end{aligned}$$

Then the result follows by $R = 1/JD_r - 1$ ☐

Remark. I should remark that I am not fully convinced about the last step of change of variable in the theorem above.

Chapter 4

On the Equivalence of two statistics

We noticed that the test statistic proposed by Mehdi Jabbari Nooghabi is indeed very similar to the statistic proposed by Chikkagoudar, and therefore we will derivation their disribution with Theorem 1.

Remark. The Lemma says that the logarithm of a Pareto disribution differs from exponential random variable by at a constant. Such connection suggests that JZ_k and Z_k given in 1 may have very similar disribution.

Here we will give a theorem that gives the disribution of continous random variable under strictly monotonic transformation. This theorem is motivated by the Theorem 4.4.1 given by Hogg and Craig [5].

Theorem 4. *Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution of the continuous type having a pdf $f(x)$ that has support (a,b) with $(-\infty \leqslant a < b \leqslant \infty)$. Let g be a differentiable strictly increasing function on (a,b). Define random variable $Y_k = g(X_{(k)})$, then the joint pdf of $Y_1, \ldots Y_n$ is given by*

$$f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) = \begin{cases} n!\left|\frac{\mathrm{d}g^{-1}}{\mathrm{d}y}\right|^n f(y_1)f(y_2)\cdots f(y_n) & g(a) < y_1 < y_2 < \cdots < y_n < g(b) \\ 0 & \textit{elsewhere} \end{cases}$$

Proof. Since g is monotonic increasing and $X_{(1)} < \cdots < X_{(n)}$ and therefore the support of of $Y_1, \ldots, Y_n$ is given by $\mathcal{Y} = \{(y_1, \ldots, y_n) : a < y_1 < \cdots < y_n < b\}$, which is the image of the map $T : (x_1, \ldots, x_n) \mapsto (g(x_1), \ldots, g(x_n))$ from the set $\mathcal{X}_0 = \{a < x_1 < \cdots < x_n < b\}$. In fact, since g is an order preserving bijection, and therefore T is a bijection from set $\mathcal{X}_\sigma = \{(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) : (x_{\sigma(1)} < \cdots < x_{\sigma(n)}\}$ to $\mathcal{Y}$, where σ is any permutation on the set $\{1, \ldots, n\}$.

Consider the transformation of T restrict on $\mathcal{X}_0$, which is given by $y_1 \mapsto g(x_1), \ldots, x_n \mapsto g(x_n)$. Then Jacobian for $T^{-1}_{\mathcal{X}_0}$ is given by

$$|J_0| = \text{abs}\left(\begin{vmatrix} \frac{\mathrm{d}g^{-1}}{\mathrm{d}y_1} & 0 & \cdots & 0 \\ 0 & \frac{\mathrm{d}g^{-1}}{\mathrm{d}y_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\mathrm{d}g^{-1}}{\mathrm{d}y_n} \end{vmatrix}\right),$$

which is $= \left|\frac{\mathrm{d}g^{-1}}{\mathrm{d}y}\right|^n$. Since for for any permutation σ, the Jacobian of $T^{-1}|_{\mathcal{X}_\sigma}$ can be obtained by interchange the row of J_0, and therefore they are differ by J_0 by at most a sign.

Let S_n denoted all permutations on $\{1, \ldots, n\}$, and J_σ denoted the Jacobian of $T^{-1}_{\mathcal{X}_\sigma}$. Since $\{\mathcal{X}_\sigma\}_{\sigma \in S_n}$ are mutually disjoint and partitioned the support of $X_1, \ldots, X_n$, then the pdf of $(Y_1, \ldots, Y_n)$ is given by

$$\begin{aligned} f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) &= \sum_{\sigma \in S_n} |J_\sigma| f(y_1,\ldots,y_n) \\ &= \begin{cases} n!\left|\frac{\mathrm{d}g^{-1}}{\mathrm{d}y}\right|^n f(y_1)f(y_2)\cdots f(y_n) & g(a) < y_1 < y_2 < \cdots < y_n < g(b) \\ 0 & \text{elsewhere,} \end{cases} \end{aligned}$$

where $n!$ comes from S_n has $n!$ elements.

□

By taking g to be the identity map in the previous theorem we obtained the follwing well known fact:

Corollary 1. *Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution of the continuous type having a pdf $f(x)$ that has support (a,b) with $(-\infty \leqslant a < b \leqslant \infty)$. Let $Y_k = X_{(k)}$ for $k = 1, \ldots, n$, then then the joint pdf of $Y_1, \ldots Y_n$ is given by*

$$f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) = \begin{cases} n! f(y_1) f(y_2) \cdots f(y_n) & 0 < y_1 < y_2 < \cdots < y_n < b \\ 0 & \textit{elsewhere} \end{cases}$$

The follwing theorem allows us to use statistics handling exponential samples to the Pareto samples

Theorem 5. *Let $X_1, \ldots X_n$ be independent $\mathrm{Pareto}(\alpha, \theta)$ random variables, $E_1, \ldots, E_n$ be the independent $\mathrm{Exp}(\alpha)$ random variable. Let $Y_k = E_{(k)}$ for $k = 1, \ldots, n$. Define $U_k = \log\big(X_{(k)}/\theta\big)$, then the random vector $\mathbf{U} = (U_1, \ldots, U_n)$ has the same disribution as $\mathbf{Y} = (Y_1, \ldots, Y_n)$*

Proof. Since each X_k is independent and the function $g(y) = \log(y/\theta)$ is differentiable and strictly increasing on $[0, \infty)$, and therefore we can apply the Theorem 4 to obtain the pdf of $\mathbf{U}$. We remark that it follows from the definition that the support of $\mathbf{U}$ and $\mathbf{Y}$ are both $\mathcal{U} = \{(u_1, \ldots, u_n) : 0 < u_1 < \cdots < u_n < \infty\}$. Hence the pdf of $\mathbf{U}$ is given by

$$\begin{aligned} f_{U_1,\ldots,U_n}(u_1,\ldots,u_n) &= n! \left| \frac{\mathrm{d}g^{-1}}{\mathrm{d}u} \right|^n f(u_1) \cdots f(u_n) I_{\mathcal{U}}(u_1,\ldots,u_n) \\ &= n! \prod_{k=1}^{n} \left| \frac{\mathrm{d}g^{-1}}{\mathrm{d}u} \right| f_(u_k) \cdots f(u_k) I_{\mathcal{U}}(u_1,\ldots,u_n) \\ &= n! \prod_{k=1}^{n} \alpha\theta^{\alpha} [\theta e^{u}_{k}]^{-(\alpha+1)} |\theta e^{u_k}| I_{\mathcal{U}}(u_1,\ldots,u_n) \\ &= n! \prod_{k=1}^{n} \alpha e^{-\alpha u_k} I_{\mathcal{U}}(u_1,\ldots,u_n) \\ &= n! \prod_{k=1}^{n} f_Y(u_k) I_{\mathcal{U}}(u_1,\ldots,u_n). \end{aligned}$$

Then it follows from the previous corollary that $\mathbf{U}$ has the same pdf as $\mathbf{Y}$ and the proposition is proved. □

Here we show that analogous result holds under slippage alternatives.

Now we can give can give another proof of disribution of JZ_k.

I wish to prove the following result as a corollary as theorem 1, but there is a gap in the proof.

Corollary 2. *Let $X_1, \ldots, X_n$ be a collection of idependent random variable, and under slippage alternative hypothsis H_r given in the section 3.1. The the disribution of statistics JZ_r is given by*

$$\begin{aligned} \mathrm{P}\left(JZ_r < Z \mid H_r\right) &= \frac{\beta^r \Gamma(r\beta + n - r)}{\Gamma(rb+1)} \\ &\times \sum_{j=2}^{n-r} \frac{(-1)^{n-j-r}\{\beta^{-r} - [(r\beta + n - r - j + 1)(x/(1-rx)) + \beta]^{-r}\}}{(j-2)!(n-j-r)!(r\beta + n - r - j + 1)} \qquad 0 < x < \frac{1}{r} \end{aligned}$$

Bibliography

[1] Aicha Zerbet and Mikhail Nikulin. A new statistic for detecting outliers in exponential case. *Communications in Statistics-Theory and Methods*, 32(3):573–583, 2003.

[2] Mehdi Jabbari Nooghabi. On detecting outliers in the pareto distribution. *Journal of Statistical Computation and Simulation*, 89(8):1466–1481, 2019.

[3] MS Chikkagoudar and SH Kunchur. Distributions of test statistics for multiple outliers in exponential samples. *Communications in Statistics-Theory and Methods*, 12(18):2127–2142, 1983.

[4] Narayanaswamy Balakrishnan and Valery B Nevzorov. *A primer on statistical distributions*. John Wiley & Sons, 2004.

[5] Joseph W McKean Robert V Hogg and Allen T Craig. *Introduction to mathematical statistics*. Pearson, 2019.