5-2014

Some New Probability Distributions Based on Random Extrema and Permutation Patterns

Jie Hao
East Tennessee State University

Some New Probability Distributions Based on Random Extrema and Permutation Patterns

A thesis

presented to

the faculty of the Department of Mathematics

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Mathematical Sciences

by

Jie Hao

May 2014

Anant Godbole, Ph.D., Chair

Robert Price, Ph.D.

Edith Seier, Ph.D.

Keywords: probability distribution, permutation patterns, random extrema

ABSTRACT

Some New Probability Distributions Based on Random Extrema and Permutation Patterns

by

Jie Hao

In this paper, we study a new family of random variables, that arise as the distribution of extrema of a random number N of independent and identically distributed random variables $X_1, X_2, \ldots, X_N$, where each X_i has a common continuous distribution with support on $[0, 1]$. The general scheme is first outlined, and SUG and CSUG models are introduced in detail where X_i is distributed as $U[0, 1]$. Some features of the proposed distributions can be studied via its mean, variance, moments and moment-generating function. Moreover, we make some other choices for the continuous random variables such as Arcsine, Topp-Leone, and N is chosen to be Geometric or Zipf. Wherever appropriate, we estimate of the parameter θ in the one-parameter family in question and test the hypotheses about θ. In the last section, two permutation distributions are introduced and studied.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincerest appreciation to my thesis advisor, Dr. Anant Godbole, for the support of my research. This thesis would not have been done without his patience, encouragement and guidance. I could not have expected to have a better advisor during my Master's studies. Also, I would like to thank the other committee members, Dr. Seier Edith and Dr. Robert Price, for the assistance and constructive comments they offered. Furthermore, I must acknowledge the Department of Mathematics and Statistics at East Tennessee State University for providing me with financial support. In the end, a very special thanks goes to my parents, Yongli Hao and Xiaomei Li, for always standing behind me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Extreme Value Problems

Extreme value theory has been well developed. Gumbel [1] indicates this area of research might go back to as early as 1709. N. Bernoulli was one of the pioneers who discussed extreme value problems. Classical extreme limit theory about only three types of extreme value distributions was firstly proposed by Fisher and Tippet [2]. If $X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables with cumulative distribution function (cdf) F, then in elementary texts, the distribution of the maxima $Y = \max(X_1, X_2, \ldots, X_n)$ can be derived:

$$\begin{aligned}
\mathbb{P}(Y \leq y) &= \mathbb{P}(X_1 \leq y, X_2 \leq y, \ldots, X_n \leq y) \\
&= \mathbb{P}(X_1 \leq y)\mathbb{P}(X_2 \leq y) \cdots \mathbb{P}(X_n \leq y) \\
&= [F_X(y)]^n, \qquad (1)
\end{aligned}$$

which converges to 1 if $F(x) = 1$ and to 0 if $F(x) < 1$. If there exists a sequence of constants $a_n > 0$, and $b_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x), \qquad (2)$$

where $G(x)$ is a nondegenerate distribution function, then these distributions with G occurring as a limit in (2) are called extreme value distributions [3], which are considered as the following three types of distributions:

Gumbel-type distribution:

$$\mathbb{P}(X \leq x) = \exp(-e^{-(x-\mu)/\sigma}), \qquad -\infty < x < \infty$$

Fréchet-type distribution:

$$\mathbb{P}(X \leq x) = \begin{cases} 0, & x < \mu \\ \exp\{-(\frac{x-\mu}{\sigma})^{-\xi}\}, & x \geq \mu \end{cases}$$

Weibull-type distribution:

$$\mathbb{P}(X \leq x) = \begin{cases} \exp\{-(\frac{\mu-x}{\sigma})^{\xi}\}, & x \leq \mu \\ 1, & x < \mu \end{cases}$$

where $\mu, \sigma(> 0)$ and $\xi(> 0)$ are parameters [4].

Obviously, X with the above three types of distributions takes values in an infinite interval, but this thesis studies the case where $X \in [0, 1]$ and we do not need to use the classical theory. Let us consider a sequence $X_1, X_2, \ldots$ of i.i.d. random variables with support on $[0, 1]$ and having distribution function F. Let Y denote the maximum, and the minimum is denoted by Z. For any fixed n, the distribution of the maximum

$$Y = \max_{1 \leq i \leq n} \{X_i\}$$

has the cdf $F(y) = [F_X(y)]^n$; and the distribution of the minimum

$$Z = \min_{1 \le i \le n} \{X_i\}$$

has the cdf $F(z) = 1 - [1 - F_X(z)]^n$.

But what if we have a situation where N of X_i's is a random number, and we are instead considering the extrema

$$Y = \max_{1 \le i \le N} \{X_i\} \tag{3}$$

and

$$Z = \min_{1 \le i \le N} \{X_i\} \tag{4}$$

of a random number of i.i.d. random variables? Now consider the *sum* S of a random number of i.i.d. variables, defined as

$$S = \sum_{i=1}^{N} X_i$$

satisfies, according to Wald's Lemma [5], the equation

$$\mathbb{E}(S) = \mathbb{E}(N)\mathbb{E}(X),$$

provided that N is independent of the sequence $\{X_i\}$ and assuming that the means of X and N exist.

The main objective of this thesis is to show that the distributions in (3) and (4) can be studied in many canonical cases, even if N and $\{X_i\}_{i=1}^N$ are correlated. The main deviation from the papers [6], [7] and [8], where similar questions are studied, is that the variable X is concentrated on the interval $[0, 1]$ – unlike the above references, where X has lifetime-like distributions on $[0, \infty)$. Even then, we find that many new and interesting distributions arise, none of them to be found, e.g., in [9] or [10]. In another deviation from the theory of extremes of random sequences (see, i.e., [11]), we find that the tail behavior of the extreme distributions is not relevant due to the fact that the distributions have compact support.

We next cite three examples where our methods might be useful. First, we might be interested in the strongest earthquake in a given region in a given year. The number of earthquakes in a year, N, is usually modeled using a Poisson distribution, and, ignoring aftershocks and similarly correlated events, the intensities of the earthquakes can be considered to be i.i.d. random variables in $[a, b]$ whose distribution can be modeled using, e.g., the data set maintained by Caltech [12]. Second, many "small world" phenomena have recently been modeled by power law distributions, also sometimes termed discrete Pareto or Zipf distributions. See, for example, the body of work by [13], [14], and the references therein, where vertex degrees $d(v)$ in "internet-like graphs" G (e.g., the vertices of G are individual webpages, and there is an edge between v_1 and v_2 if one of the webpages has a link to the other) are shown to be modeled by

$$\mathbb{P}(d(v) = n) = \frac{[\zeta(k)]^{-1}}{n^k}$$

for some constant $k > 1$, where $\zeta(\cdot)$ is the Riemann Zeta function

$$\zeta(k) = \sum_{n=1}^{\infty} \frac{1}{n^k}.$$

Thus, the vertices v in a large internet graph have some bounded i.i.d. property X_i, then the maximum and minimum values of X_i for the neighbors of a randomly chosen vertex can be modeled using the methods of this paper. Third, we note that N and the X_i may be correlated, as in the CSUG example (studied systematically in Section 3) where $X_i \sim U[0,1]$ and $N = \inf\{n \geq 2 : X_n > (1-\theta)\}$ follows the geometric distribution with success probability θ. This is an example of a situation where we might be modeling the maximum load that a device might have carried before it breaks down due to an excessive weight or current. It is also feasible in this case that the parameter θ might be unknown.

1.2 General Scheme

Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed (i.i.d.) random variables following a continuous distribution on $[0,1]$ with common probability density function (pdf)

$$p_{X_i}(x) = f(x), \quad i = 1, 2, \ldots, N.$$

Let N is a random number following a discrete distribution on $\{1, 2, \ldots\}$ with probability mass function (pmf) given by

$$\mathbb{P}(N = n) = p(n), \quad n = 1, 2, \ldots .$$

Then the cumulative distribution function (cdf) of $Y = \max(X_1, X_2, \ldots, X_N)$ given $N = n$ is

$$\mathbb{P}(Y \leq y | N = n) = [F(y)]^n, \tag{5}$$

where

$$F(y) = \int_0^y f(x)dx.$$

Taking the first derivative of (5), we have

$$g(y|N = n) = n[F(y)]^{n-1}f(y).$$

Hence, the marginal density function of Y is

$$\begin{aligned} g(y) &= \sum_{n=1}^{\infty} g(y|N = n)\mathbb{P}(N = n) \\ &= f(y)\sum_{n=1}^{\infty} n[F(y)]^{n-1}p(n). \end{aligned} \tag{6}$$

In a similar fashion, the marginal density function of $Z = \min(X_1, X_2, \ldots, X_N)$ can be shown as follows:

$$g(z) = f(z)\sum_{n=1}^{\infty} n[1 - F(z)]^{n-1}p(n), \tag{7}$$

since

$$\mathbb{P}(Z \leq z | N = n) = 1 - \mathbb{P}(Z > z | N = n) = 1 - [1 - F(z)]^n.$$

1.3 Permutation Problems

Before we continue with our study of the distribution of extrema of a random number of random variables, we will also study a few distributions that arise from the theory of permutations and pattern avoidance in permutations. For instance, if $X_1, X_2, \ldots$ are i.i.d. random variables with some common continuous distribution(i.e. $U[0,1]$), and if $X_1 > X_2 > X_3 > \cdots > X_t < X_{t+1}$, then we can say the position of the first ascent is at t. For $n \geq 0$, the Catalan numbers C_n are given by

$$C_n = \frac{1}{n+1}\binom{2n}{n};$$

generalizing this fact, Catalan [15] proved the k-fold Catalan convolution formula

$$C_{n,k} := \sum_{i_1+\ldots+i_k=n} \prod_{r=1}^{k} C_{i_r-1} = \frac{k}{2n-k}\binom{2n-k}{n}.$$

The theory of pattern avoidance in permutations is now well-established and thriving, and a survey of the many results in that area may be found in the text by Kitaev [16]. One of the earliest and most fundamental results in the field is that the number of permutations in which the longest increasing sequence is of length 2, the so-called 123-avoiding permutations, is given by the Catalan numbers. The following is a very natural question: in how many permutations in which the longest increasing subsequence is of length 2, does the first ascent occur in positions $j, j+1$? In [18] it is proved that there are $C_{n,k}$ permutations on $[n] := \{1, 2, \ldots, n\}$ with longest increasing subsequence of size 2 and for which the first ascent occurs at positions $k, k+1$. It is this fact that leads to the distribution in Section 6. In Section 5, however, we will study a simpler distribution that deals with first ascent in ordinary non-restricted permutations.

2 STANDARD UNIFORM GEOMETRIC (SUG) MODEL

Let $X_1, X_2, \ldots, X_N$ be i.i.d. random variables following the Standard Uniform distribution with pdf is given by

$$f(x) = 1,\ 0 \leq x \leq 1.$$

And N is a random number following the Geometric distribution with probability mass function given by

$$\mathbb{P}(N = n) = \theta(1-\theta)^{n-1},\ 0 < \theta < 1,\ n = 1, 2, \ldots .$$

According to general scheme (Section 1.2), we have from (6) that the density function of Y in the SUG Model is given by

$$\begin{aligned} g(y) &= \sum_{n=1}^{\infty} \theta(1-\theta)^{n-1} \times n y^{n-1} \\ &= \theta \times \sum_{n=1}^{\infty} n[(1-\theta)y]^{n-1} \\ &= \frac{\theta}{[1-(1-\theta)y]^2}, \quad 0 < y < 1. \end{aligned} \tag{8}$$

The following figure shows that a graphical probability density function of the SUG maximum model on various values of parameter θ.

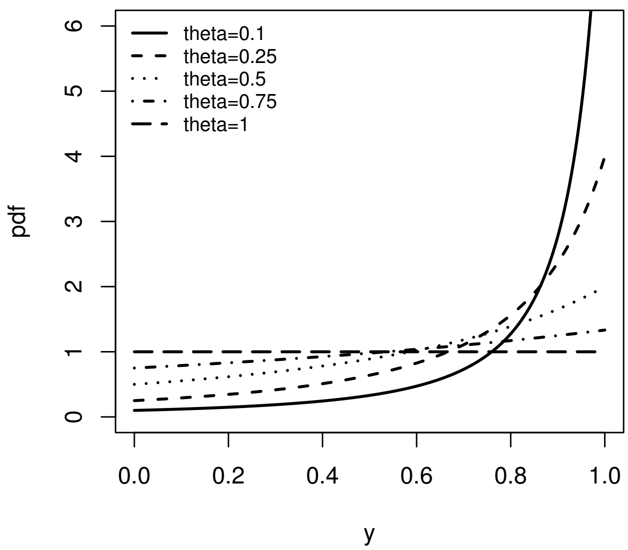

Figure 1: Plot of the SUG maximum density for some values of θ

Similarly, from (7), the density function of Z in the SUG Model is derived as follows:

$$\begin{aligned} g(z) &= \sum_{n=1}^{\infty} \theta(1-\theta)^{n-1} \times n(1-z)^{n-1} \\ &= \theta \times \sum_{n=1}^{\infty} n[(1-\theta)(1-z)]^{n-1} \\ &= \frac{\theta}{[1-(1-\theta)(1-z)]^2} \\ &= \frac{\theta}{[\theta+(1-\theta)z]^2}, \quad 0 < z < 1. \end{aligned} \tag{9}$$

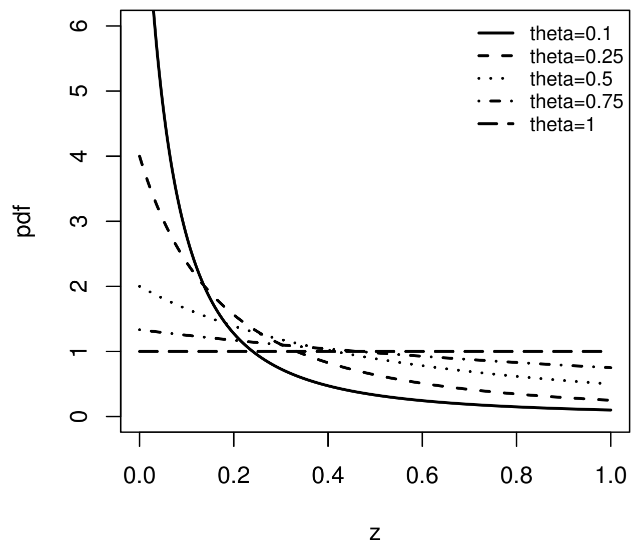

Figure 2: Plot of the SUG minimum density for some values of θ

See Figure 2 for a graphical probability density function of the SUG minimum model on various values of parameter θ.

2.1 Some Properties of the Distributions of Y and Z Under the SUG Model

Proposition 2.1. *For a random variable Y with "SUG maximum distribution" (8) and $k \in \mathbb{N}$, we have that*

$$\mathbb{E}(Y^k) = \frac{\theta}{(1-\theta)^{k+1}} \sum_{j=0}^{k} \binom{k}{j} \int_{\theta}^{1} (-u)^{j-2} du,$$

where $u = 1 - (1-\theta)y$.

Proof.

$$\begin{aligned}
\mathbb{E}(Y^k) &= \int_0^1 y^k \times \frac{\theta}{[1-(1-\theta)y]^2} dy \\
&= \int_1^\theta \left(\frac{1-u}{1-\theta}\right)^k \times \frac{\theta}{u^2} \times \left(-\frac{1}{1-\theta}\right) du \\
&= \frac{\theta}{(1-\theta)^{k+1}} \int_\theta^1 \frac{(1-u)^k}{u^2} du \\
&= \frac{\theta}{(1-\theta)^{k+1}} \int_\theta^1 \frac{\sum_{j=0}^k \binom{k}{j}(-u)^j}{u^2} du \\
&= \frac{\theta}{(1-\theta)^{k+1}} \sum_{j=0}^k \binom{k}{j} \int_\theta^1 (-u)^{j-2} du,
\end{aligned}$$

as claimed. □

Proposition 2.2. *The random variable Y with density function given by (8) has mean and variance obtained, respectively, by*

$$\mathbb{E}(Y) = \frac{\theta(\ln\theta + \frac{1}{\theta} - 1)}{(1-\theta)^2} \quad \textit{and} \quad \mathbb{V}(Y) = \frac{\theta^3 - 2\theta^2 - \theta^2 \ln^2\theta + \theta}{(1-\theta)^4}.$$

Proof. Using Proposition 2.1, we can directly obtain the mean and variance by setting $k = 1$ and $k = 2$, and using the fact that $\mathbb{V}(W) = \mathbb{E}(W^2) - [\mathbb{E}(W)]^2$ for any random variable W. □

Proposition 2.3. *For a random variable Z with "SUG minimum distribution" (9) and $k \in \mathbb{N}$, we have that*

$$\mathbb{E}(Z^k) = \frac{\theta}{(1-\theta)^{k+1}} \sum_{j=0}^k \binom{k}{j} (-\theta)^j \int_\theta^1 u^{k-j-2} du,$$

where $u = \theta + (1-\theta)z$.

Proof.

$$\begin{aligned}
\mathbb{E}(Z^k) &= \int_0^1 z^k \times \frac{\theta}{[\theta + (1-\theta)z]^2} dz \\
&= \int_\theta^1 \left(\frac{u-\theta}{1-\theta}\right)^k \times \frac{\theta}{u^2} \times \frac{1}{1-\theta} du \\
&= \frac{\theta}{(1-\theta)^{k+1}} \int_\theta^1 \frac{(u-\theta)^k}{u^2} du \\
&= \frac{\theta}{(1-\theta)^{k+1}} \int_\theta^1 \frac{\sum_{j=0}^k \binom{k}{j} u^{k-j}(-\theta)^j}{u^2} du \\
&= \frac{\theta}{(1-\theta)^{k+1}} \sum_{j=0}^k \binom{k}{j} (-\theta)^j \int_\theta^1 u^{k-j-2} du.
\end{aligned}$$

as claimed. □

Proposition 2.4. *The random variable Z with density function given by (9) has mean and variance obtained, respectively, by*

$$\mathbb{E}(Z) = \frac{\theta(\theta - 1 - \ln\theta)}{(1-\theta)^2} \quad \textit{and} \quad \mathbb{V}(Z) = \frac{\theta^3 - 2\theta^2 - \theta^2 \ln^2\theta + \theta}{(1-\theta)^4}.$$

Proof. Using Proposition 2.3 and the fact that $\mathbb{V}(W) = \mathbb{E}(W^2) - [\mathbb{E}(W)]^2$ for any random variable W, we directly obtain this proof. □

Moment-generating functions of the Y, Z variables are easily to calculate too. Notice that the logarithmic terms above arise due to the contributions of the $j = 1$ and $j = k - 1$ terms, and it is precisely these logarithmic terms that make, e.g., method of moments estimates for θ to be intractable in a closed (i.e., non-numerical) form. Similar difficulties arise when analyzing the likelihood function and likelihood ratios.

3 THE CORRELATED STANDARD UNIFORM GEOMETRIC (CSUG) MODEL

The Correlated Standard Uniform Geometric (CSUG) model is related to the SUG model, as the name suggests, but X and N are correlated as indicated in Section 2. The CSUG problems arise in two cases. One case is that we conduct standard uniform trials until a variable X_i exceeds $1-\theta$, where θ is the parameter of the correlated geometric variable, and the maximum of $X_1, X_2, \ldots, X_{i-1}$ is what we seek. The maximum is between 0 and $1-\theta$. The other case is where standard uniform trials are conducted until X_i is less than θ, and we are looking for the minimum of $X_1, X_2, \ldots, X_{i-1}$. The minimum is between θ and 1.

Specifically, let $X_1, X_2, \ldots$ be a sequence of standard uniform variables and define

$$N = \inf\{n \geq 2 : X_i > 1-\theta\},$$

or

$$N = \inf\{n \geq 2 : X_i < \theta\}.$$

In either case N has probability mass function given by

$$\mathbb{P}(N=n) = \theta(1-\theta)^{n-2}, 0 < \theta < 1, n = 2, 3, \ldots; \tag{10}$$

note that this is simply a geometric random variable conditional on the success having occurred at trial 2 or later. Considering (6), X has probability density function given by

$$\text{for maximum}: \ f(x) = \frac{1}{1-\theta}, \quad 0 \le x \le 1-\theta;$$

$$\text{for minimum}: \ f(x) = \frac{1}{1-\theta}, \quad \theta \le x \le 1.$$

Clearly N is dependent on the X sequence.

Proposition 3.1. *Under the CSUG model, the density function of Y, defined by (3), is given by*

$$g(y) = \frac{\theta}{(1-\theta)(1-y)^2}, \quad 0 \le y \le 1-\theta. \tag{11}$$

Proof. The conditional cumulative distribution function of Y given that $N = n$ is given by

$$\mathbb{P}(Y \le y | N = n) = \left(\frac{y}{1-\theta}\right)^{n-1}, \quad n = 2, 3, \ldots . \tag{12}$$

Taking the derivative of (12), we see that the conditional density function is given by

$$g(y|N = n) = \frac{n-1}{1-\theta}\left(\frac{y}{1-\theta}\right)^{n-2}, \quad n = 2, 3, \ldots$$

Consequently, the marginal density function of Y in the CSUG model is given by

$$\begin{aligned} g(y) &= \sum_{n=2}^{\infty} \theta(1-\theta)^{n-2} \times \frac{n-1}{1-\theta}\left(\frac{y}{1-\theta}\right)^{n-2} \\ &= \frac{\theta}{1-\theta} \times \sum_{n=2}^{\infty}(n-1)y^{n-2} \\ &= \frac{\theta}{(1-\theta)(1-y)^2}, \end{aligned}$$

which completes the proof. $\square$

Figure 3 shows the probability density function of the CSUG maximum model on various values of parameter θ.

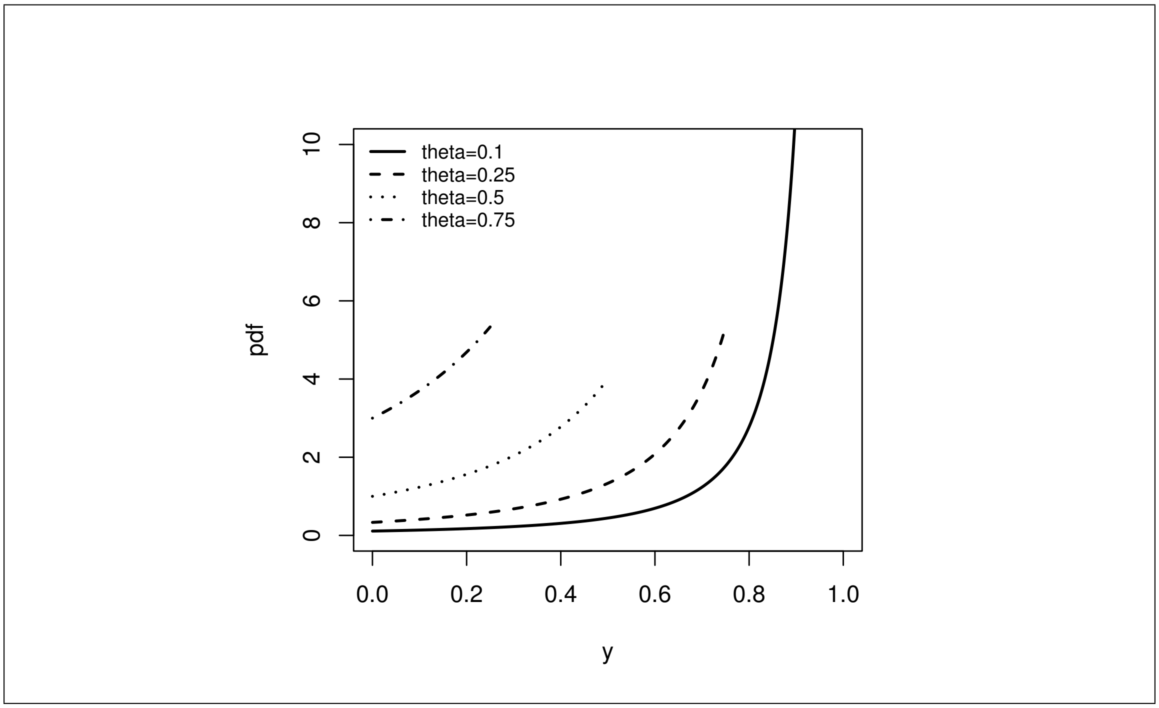

Figure 3: Plot of the CSUG maximum density for some values of θ

Proposition 3.2. *The density function of Z under the CSUG model is given by*

$$g(z) = \frac{\theta}{(1-\theta)z^2}, \ \theta \leq z \leq 1, \ n = 2, 3, \ldots. \tag{13}$$

Proof. The conditional cumulative distribution function of Z given that $N = n$ is given by

$$\mathbb{P}(Z \leq z | N = n) = 1 - \mathbb{P}(Z > z | N = n) = 1 - \left(\frac{1-z}{1-\theta}\right)^{n-1}, \ n = 2, 3, \ldots$$

Thus, the conditional density function is given by

$$g(z|N=n)=\frac{n-1}{1-\theta}\left(\frac{1-z}{1-\theta}\right)^{n-2},\ n=2,3,\ldots,$$

which yields the density function of Z under the CSUG model as

$$\begin{aligned} g(z) &= \sum_{n=2}^{\infty}\theta(1-\theta)^{n-2}\times\frac{n-1}{1-\theta}\left(\frac{1-z}{1-\theta}\right)^{n-2} \\ &= \frac{\theta}{1-\theta}\times\sum_{n=2}^{\infty}(n-1)(1-z)^{n-2} \\ &= \frac{\theta}{(1-\theta)z^2}, \end{aligned}$$

which finishes the proof. □

Figure 4 shows the probability density function of the CSUG minimum model on various values of parameter θ.

3.1 Some Properties of the Distributions of Y and Z Under the CSUG Model

Proposition 3.3. *For a random variable Y with "CSUG maximum distribution" (11) and $k\in\mathbb{N}$, we have that*

$$\mathbb{E}(Y^k)=\frac{\theta}{1-\theta}\sum_{j=0}^{k}\binom{k}{j}\int_{\theta}^{1}(-u)^{j-2}du,$$

where$u=1-y$.

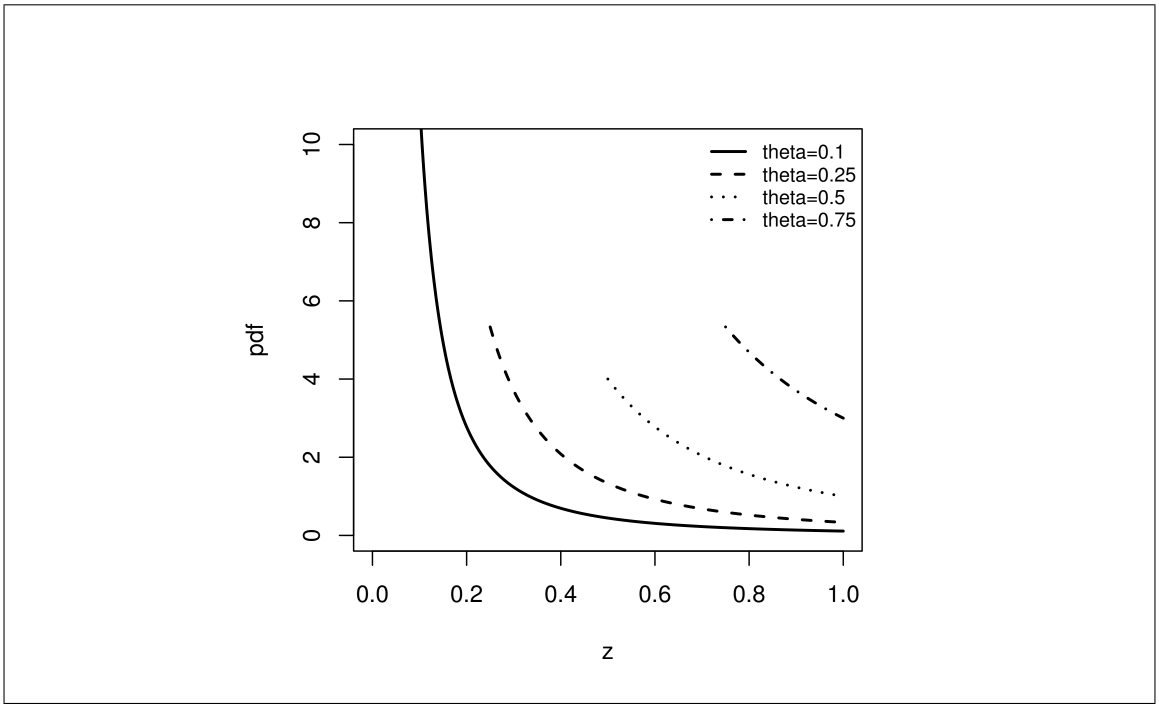

Figure 4: Plot of the CSUG minimum density for some values of θ

Proof.

$$\begin{aligned}
\mathbb{E}(Y^k) &= \int_0^{1-\theta} y^k \times \frac{\theta}{(1-\theta)(1-y)^2} dy \\
&= \frac{\theta}{1-\theta} \int_1^{\theta} \frac{(1-u)^k}{u^2}(-du) \\
&= \frac{\theta}{1-\theta} \int_{\theta}^{1} \frac{\sum_{j=0}^{k} \binom{k}{j} (-u)^j}{u^2} du \\
&= \frac{\theta}{1-\theta} \sum_{j=0}^{k} \binom{k}{j} \int_{\theta}^{1} (-u)^{j-2} du,
\end{aligned}$$

as asserted. $\square$

Proposition 3.4. *The random variable Y with density function given by (11) has* *mean and variance obtained, respectively, by*

$$\mathbb{E}(Y) = \frac{\theta \ln \theta - \theta + 1}{1 - \theta}$$

and

$$\mathbb{V}(Y) = \frac{\theta^3 - 2\theta^2 - \theta^2 \ln^2 \theta + \theta}{(1-\theta)^2}.$$

Proof. Using Proposition 3.3, we can directly compute the mean and variance by setting $k = 1, 2$. Notice that the variance of Y is smaller than that of Y under the SUG model, with an identical numerator term. Also, the expected value is smaller under the CSUG model than in the SUG case. □

Proposition 3.5. *For a random variable Z with "CSUG minimum distribution" (13) and $k \in \mathbb{N}$, we have that*

$$\mathbb{E}(Z^k) = \frac{\theta}{1-\theta} \int_\theta^1 z^{k-2} dz.$$

Proof.

$$\begin{aligned} \mathbb{E}(Z^k) &= \int_\theta^1 z^k \times \frac{\theta}{(1-\theta)z^2} dz \\ &= \frac{\theta}{1-\theta} \int_\theta^1 z^{k-2} du \end{aligned}$$

□

Proposition 3.6. *The random variable Z with density function given by (13) has* *mean and variance given, respectively, by*

$$\mathbb{E}(Z) = \frac{-\theta \ln \theta}{1-\theta} \quad \textit{and} \quad \mathbb{V}(Z) = \frac{\theta^3 - 2\theta^2 - \theta^2 \ln^2 \theta + \theta}{(1-\theta)^2}.$$

Proof. A special case of Proposition 3.5; note that as in the SUG model, $\mathbb{V}(Y) = \mathbb{V}(Z)$. $\square$

3.2 Parameter Estimation

The intermingling of polynomial and logarithmic terms makes method of moments estimation difficult in closed form, as in the SUG case. However, if θ is unknown, the maximum likelihood estimate of θ can be found in a satisfying form, both in the CGUG maximum and CSUG minimum cases. Suppose that $Y_1, Y_2, \ldots, Y_n$ form a random sample from the CSUG Maximum distribution with unknown θ. Since the pdf of each observation has the following form:

$$f(y|\theta) = \begin{cases} \frac{\theta}{(1-\theta)(1-y)^2}, & \text{for } 0 \leq y \leq 1-\theta \\ 0, & \text{otherwise} \end{cases}$$

the likelihood function is given by

$$\ell(\theta) = \begin{cases} (\frac{\theta}{1-\theta})^n \frac{1}{\prod_{i=1}^n (1-y_i)^2}, & \text{for } 0 \leq y_i \leq 1-\theta \ (i = 1, 2, \ldots, n) \\ 0, & \text{otherwise} \end{cases}$$

The MLE of θ is a value of θ, where $\theta \leq 1 - y_i$ for $i = 1, 2, \ldots, n$, which maximizes $\frac{\theta}{1-\theta}$. Let $\varphi(\theta) = \frac{\theta}{1-\theta}$. Since $\varphi'(\theta) \geq 0$, it follows that $\varphi(\theta)$ is a increasing function, which means the MLE is the largest possible value of θ such that $\theta \leq 1 - y_i$ for $i = 1, 2, \ldots, n$.

Thus, this value should be $1 - \max(Y_1, \dots, Y_n)$, i.e., $\hat{\theta} = 1 - Y_{(n)}$.

Suppose next that $Z_1, Z_2, \dots, Z_n$ form a random sample from the CSUG minimum distribution. Since the pdf of each observation has the following form:

$$f(z|\theta) = \begin{cases} \frac{\theta}{(1-\theta)z^2}, & \text{for } \theta \le z \le 1 \\ 0 & \text{otherwise,} \end{cases}$$

it follows that the likelihood function is given by

$$\ell(\theta) = \begin{cases} (\frac{\theta}{1-\theta})^n \frac{1}{\prod_{i=1}^n z_i^2}, & \text{for } \theta \le z_i \le 1 \ (i = 1, 2, \dots, n) \\ 0 & \text{otherwise.} \end{cases}$$

As above, it now follows that $\hat{\theta} = Y_{(1)}$.

4 A SUMMARY OF SOME OTHER MODELS

The general scheme given by (6) and (7) is quite powerful. For example, suppose (using the example from Section 1.1) that the pmf of random numbers is

$$p(n) = \frac{6}{\pi^2}\frac{1}{n^2}$$

and the random variables $X \sim U[0,1]$. Then it is easy to show that

$$g(y) = \frac{6}{\pi^2}\frac{1}{y}\ln\left(\frac{1}{1-y}\right), 0 \le y \le 1,$$

and that $\mathbb{E}(Y) = \frac{6}{\pi^2}$. (The expected value of Y can also be calculated by using the identity $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|N))$.

In this section, we collect some more results of this type, without proof:

UNIFORM-POISSON MODEL Here we let $X \sim U[0,1]$ and $p(n) = \frac{e^{-\lambda}\lambda^n}{(1-e^{-\lambda})n!}, n = 1, 2, \ldots$, so that N follows a left-truncated Poisson distribution.

Proposition 4.1. *Under the Uniform-Poisson model,*

$$g(y) = \frac{\lambda e^{-\lambda}e^{\lambda y}}{1-e^{-\lambda}}; g(z) = \frac{\lambda e^{-\lambda z}}{1-e^{-\lambda}};$$

$$\mathbb{E}(Y) = \frac{1}{1-e^{-\lambda}} - \frac{1}{\lambda}; \mathbb{E}(Z) = \frac{1}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}};$$

$$\mathbb{V}(Y) = \frac{1}{\lambda^2} + \frac{1}{1-e^{-\lambda}} - \frac{1}{(1-e^{-\lambda})^2}; \mathbb{V}(Z) = \frac{1}{\lambda^2} - \frac{e^{-\lambda}}{\lambda(1-e^{-\lambda})} - \frac{e^{-2\lambda}}{(1-e^{-\lambda})^2};$$

$$M_Y(t) = \mathbb{E}(e^{ty}) = \frac{\lambda e^{-\lambda}(e^{t+\lambda} - 1)}{(t+\lambda)(1-e^{-\lambda})}; M_Z(t) = \mathbb{E}(e^{tz}) = \frac{\lambda(e^{t-\lambda} - 1)}{(t-\lambda)(1-e^{-\lambda})}.$$

In some sense, the primary motivation of this paper was to produce extreme value distributions that did not fall into the Beta family (such as $f(y) = nt^{n-1}$ for the maximum of n i.i.d. $U[0,1]$ variables). A wide variety of non-Beta-based distributions may be found in [10]. Can we add extreme value distributions to that collection? In what follows, we use both the Beta families $B(2,2)$ and $B(1/2,1/2)$, the arcsine distribution, and a "Beyond Beta" distribution, the Topp-Leone distribution, as "input variables" to make further progress in this direction.

GEOMETRIC-BETA(2,2) MODEL. Here $X \sim B(2,2)$ and $N \sim \text{Geo}(\theta)$. In this case we get

$$g(y) = \frac{6y(1-y)\theta}{[1-(1-\theta)y^2(3-2y)]^2}$$

and

$$g(z) = \frac{6z(1-z)\theta}{[1-(1-\theta)(2z^3-3z^2+1)]^2}.$$

POISSON-BETA(2,2) MODEL. Here $X \sim B(2,2)$ and $N \sim \text{Po}_0(\theta)$, the Poisson($\theta$) distribution left-truncated at 0. In this case we get

$$g(y) = \frac{6\theta y(1-y)e^{-\theta(2y^3-3y^2+1)}}{1-e^{-\theta}}$$

and

$$g(z) = \frac{6\theta z(1-z)e^{-\theta(3z^2-2z^3)}}{1-e^{-\theta}}.$$

GEOMETRIC-ARCSINE MODEL. Here $X \sim B(1/2,1/2)$ and $N \sim \text{Geo}(\theta)$. In this case we get

$$g(y) = \frac{\theta\pi^{-1}[y(1-y)]^{-1/2}}{[1-(1-\theta)\frac{2}{\pi}\arcsin\sqrt{y}]^2}$$

and

$$g(z) = \frac{\theta\pi^{-1}[z(1-z)]^{-1/2}}{[1-(1-\theta)(1-\frac{2}{\pi}\arcsin\sqrt{z})]^2}.$$

POISSON-ARCSINE MODEL. Here $X \sim B(1/2, 1/2)$ and $N \sim \mathrm{Po}_0(\theta)$. Here we have

$$g(y) = \frac{\theta\pi^{-1}[y(1-y)]^{-1/2}e^{-\theta(1-\frac{2}{\pi}\arcsin\sqrt{y})}}{1-e^{-\theta}}$$

and

$$g(z) = \frac{\theta\pi^{-1}[z(1-z)]^{-1/2}e^{-\frac{2\theta\arcsin\sqrt{z}}{\pi}}}{1-e^{-\theta}}.$$

Before we move on to the next model, we study Topp-Leone distribution first.

Definition 4.2. *Topp-Leone distribution is defined by Topp and Leone [20], which is a family of distributions with cdf*

$$F(x;a,b) = \begin{cases} 0, & x \leq 0 \\ \left(\frac{x}{b}\right)^a \left(2-\frac{x}{b}\right)^a, & 0 < x < b \\ 1, & x \geq b \end{cases} \tag{14}$$

where $0 < x < b < \infty$ and $a > 0$. *The pdf of the distribution obtained by differentiation of (14) given by*

$$f(x;a,b) = \frac{2a}{b}(1-\frac{x}{b})\left(\frac{x}{b}\right)^{a-1}\left(2-\frac{x}{b}\right)^{a-1} \tag{15}$$

Genç [21] stated that the support of the Topp-Leone distribution is bounded. For

$b = 1$, the pdf of the standard Topp-Leone (STL) distribution is given by

$$f(x; a) = 2a(1 - x)x^{a-1}(2 - x)^{a-1}, \quad 0 < x < 1, \ a > 0 \tag{16}$$

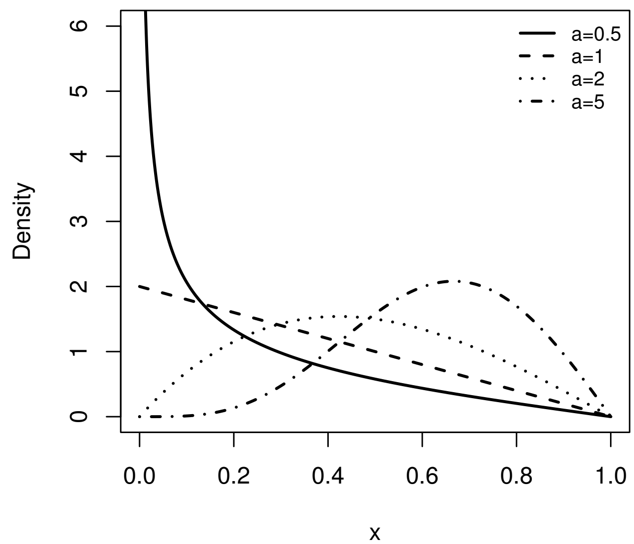

Figure 5: Plot of the standard Topp-Leone density for some values of a

See Figure 5 for a graphical probability density function of the standard Topp-Leone distribution on various values of parameter a.

GEOMETRIC-TOPP-LEONE MODEL. Here $X \sim STL(a)$ and $N \sim \text{Geo}(\theta)$:

$$g(y) = \frac{2a(1 - y)y^{a-1}(2 - y)^{a-1}\theta}{[1 - (1 - \theta)y^a(2 - y)^a]^2}$$

and

$$g(z) = \frac{2a(1-z)z^{a-1}(2-z)^{a-1}\theta}{\{1-(1-\theta)[1-z^a(2-z)^a]\}^2}.$$

POISSON-TOPP-LEONE MODEL. $X \sim TL(a)$ and $N \sim \mathrm{Po}_0(\theta)$:

$$g(y) = \frac{2\theta a(1-y)y^{a-1}(2-y)^{a-1}e^{-\theta[1-y^a(2-y)^a]}}{1-e^{-\theta}}$$

and

$$g(z) = \frac{2\theta a(1-z)z^{a-1}(2-z)^{a-1}e^{-\theta[z^a(2-z)^a]}}{1-e^{-\theta}}.$$

5 PERMUTATION BASED DISTRIBUTION 1

5.1 First Ascent Distribution

Consider there are $n!$ permutations on $\{1, 2, 3, \cdots, n\}$. For example, the probability that a randomly chosen permutation has the first ascent at position 1 is

$$\mathbb{P}(n_1 < n_2) = \frac{1}{2!} = \frac{1}{2};$$

the probability that a randomly chosen permutation has the first ascent at position 2 is

$$\mathbb{P}(n_3 > n_1 > n_2, n_1 > n_3 > n_2) = \frac{2}{3!}.$$

In general, the probability that a random permutation on has its first ascent at position k is given, for $1 \leq k \leq n-1$, by $\frac{k}{(k+1)!}$. To see this, choose any one of the $k+1$ elements in positions 1 through $k+1$, except for the smallest, to occupy the $k+1$st position, and then arrange the other elements in a monotone decreasing fashion. The chance that the first ascent is at position n is, of course, $\frac{1}{n!}$.

We will find it more convenient in this section to consider an infinite analogs of this distribution. An infinite permutation may be realized, e.g., by considering the order statistics $X_{(1)} < X_{(2)} < \ldots$ of a sequence $X_1, X_2 \ldots$ of i.i.d. random variables with say a uniform distribution on [0,1]. Under this scheme we get the first ascent distribution as being

$$f(x) = \frac{x}{(x+1)!}, x = 1, 2, \ldots. \tag{17}$$

It is easy to verify that this is a genuine probability distribution on $1, 2, \cdots$, and that it is similar to the unit Poisson distribution on 0,1,... with mass function $f(x) = e^{-1}/x!$.

5.2 Some Properties of the First Ascent Distribution

Proposition 5.1. *For i.i.d. random variable X with the first ascent distribution on $\{1, 2, \cdots\}$, the expected value is given by*

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x=1}^{\infty} x \times \frac{x}{(x+1)!} \\
&= \sum_{x=1}^{\infty} \frac{1}{(x-1)!} - \sum_{x=1}^{\infty} \frac{1}{x!} + \sum_{x=1}^{\infty} \frac{1}{(x+1)!} \\
&= e - (e-1) + (e-1-1) \\
&= e - 1.
\end{aligned}$$

Note that the exponential function formula, which is

$$e^x = \sum_{x=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots, \text{ valid for all real } x.$$

Proposition 5.2. *For i.i.d. random variable X with the first ascent distribution on $\{1, 2, \cdots\}$, the variance is given by*

$$\mathbb{V}(X) = e(3 - e).$$

Proof. Since

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{x=1}^{\infty} x^2 \times \frac{x}{(x+1)!} \\
&= \sum_{x=2}^{\infty} \frac{1}{(x-2)!} + \sum_{x=1}^{\infty} \frac{x}{(x+1)!} \\
&= e+1,
\end{aligned}$$

then by the alternative definition of the variance,

$$\begin{aligned}
\mathbb{V}(X) &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \\
&= (e+1) - (e-1)^2 \\
&= e(3-e).
\end{aligned}$$

□

Proposition 5.3. *For i.i.d. random variable X with the first ascent distribution on $\{1, 2, \cdots\}$, the moment-generating function (m.g.f) is given by*

$$\begin{aligned}
\mathbb{E}(e^{tX}) &= \sum_{x=1}^{\infty} e^{tx} \times \frac{x}{(x+1)!} \\
&= \sum_{x=1}^{\infty} \frac{e^{tx}}{x!} - \sum_{x=1}^{\infty} \frac{e^{tx}}{(x+1)!} \\
&= \sum_{x=1}^{\infty} \frac{e^{tx}}{x!} - \frac{1}{e^t} \sum_{x=1}^{\infty} \frac{(e^t)^{x+1}}{(x+1)!} \\
&= (e^{e^t} - 1) - \frac{1}{e^t}(e^{e^t} - 1 - e^t) \\
&= e^{-t}(1 - e^{e^t} + e^{t+e^t}),
\end{aligned}$$

which is quite a strange m.g.f.!

5.3 The θ Model for the First Ascent Distribution

Meanwhile, it is similar as in the Poisson distribution, to define a θ-analog of the above distribution defined by

$$f(x) = \frac{\theta^x}{x!} - \frac{\theta^{x+1}}{(x+1)!},\ 0 < \theta < 1,\ x = 0, 1, 2, \ldots \tag{18}$$

Let us discuss the relationship among our new density function, sub-Poisson and super-Poisson. If we assume our new density function $f(x)$ is a sub-Poisson function,

$$\frac{\theta^x}{x!} - \frac{\theta^{x+1}}{(x+1)!} < \frac{e^{-\theta}\theta^x}{x!}$$

by standard manipulations,

$$1 - \frac{\theta}{x+1} - e^{-\theta} < 0$$

If we set $x = 0$, then

$$1 - \theta - e^{-\theta} < 0$$

Let $\alpha(\theta) = 1 - \theta - e^{-\theta}$, we have $\alpha(0) = 0$.

Taking the first derivative of $\alpha(\theta)$,

$$\alpha'(\theta) = -1 + e^{-\theta} < 0 \text{ by } e^{-1} < e^{-\theta} < 1,\ \text{and } e^{-1} \approx 0.37$$

Hence, $\alpha(\theta)$ is a decreasing function. Due to $0 < \theta < 1, \alpha(\theta) < 0$.

Therefore, $f(x)$ is always a sub-Poisson function when $x = 0$.

If we set $x = 1$, then

$$1 - \frac{\theta}{2} - e^{-\theta} < 0$$

Let $\beta(\theta) = 1 - \frac{\theta}{2} - e^{-\theta}$, we have $\beta(0) = 0$.

Taking the first derivative of $\beta(\theta)$,

$$\beta'(\theta) = -0.5 + e^{-\theta}$$

Now, we have to discuss $\beta'(\theta)$ into two cases. If $\theta > ln2$, and $ln2 \approx 0.69$, $\beta'(\theta) < 0$. Thus, $\beta(\theta)$ is a decreasing function, which implies $\beta(\theta) < 0$. We conclude that $f(x)$ is a sub-Poisson function. If $\theta < ln2, \beta'(\theta) > 0$, which implies $\beta(\theta)$ is an increasing function. Thus, $\beta(\theta) > 0$, which means our assumption is invalid. So, $f(x)$ is a super-Poisson function.

Therefore, $f(x)$ is either sub-Poisson or super-Poisson when $x = 1$. It depends on the value of θ. If $\theta > ln2$, $f(x)$ is a sub-Poisson function; if $\theta < ln2$, $f(x)$ is a super-Poisson function.

If we set $x = 2$, then

$$1 - \frac{\theta}{3} - e^{-\theta} < 0$$

Let $\gamma(\theta) = 1 - \frac{\theta}{3} - e^{-\theta}$, we have $\gamma(0) = 0$.

Taking the first derivative of $\gamma(\theta)$,

$$\gamma'(\theta) = -\frac{1}{3} + e^{-\theta}$$

we know $\gamma'(\theta)$ would be greater than zero if $\theta > ln3$. Since $0 < \theta < 1,$ and $ln3 \approx 1.09,$ $\gamma'(\theta)$ is positive. It implies $\gamma(\theta)$ is an increasing function, and then it is greater than zero. Thus, the assumption is false.

So, $f(x)$ is a super-Poisson function when $x = 2$. Similarity, $f(x)$ is still a super-Poisson function when $x = 3, 4, 5, \ldots$

Therefore, $f(x)$ is always a super-Poisson function when $x \geq 2$.

5.4 Some Properties of the θ Model

Proposition 5.4. *Under the θ-analog of the first ascent model, the m.g.f of X is*

$$\begin{aligned} M_X(t) = \mathbb{E}(e^{tX}) &= \sum_{x=0}^{\infty} e^{tx} \times (\frac{\theta^x}{x!} - \frac{\theta^{x+1}}{(x+1)!}) \\ &= \sum_{x=0}^{\infty} \frac{(\theta e^t)^x}{x!} - e^{-t} \sum_{x=0}^{\infty} \frac{(\theta e^t)^{x+1}}{(x+1)!} \\ &= e^{\theta e^t} - e^{-t}(e^{\theta e^t} - 1) \\ &= e^{\theta e^t}(1 - e^{-t}) + e^{-t} \end{aligned}$$

Proposition 5.5. *Under the θ-analog of the first ascent model, the expected value and variance are given, respectively, by*

$$\mathbb{E}(X) = e^{\theta} - 1 \quad \textit{and} \quad \mathbb{V}(X) = e^{\theta}\left(2\theta + 1 - e^{\theta}\right).$$

Proof. If we differentiate $M_X(t)$ once and then set $t = 0$, we get the expected value is

$$\begin{aligned}
\mathbb{E}(X) &= M'(0) \\
&= \left(e^{\theta e^t}\theta e^t\right) \times \left(1 - t^{-t}\right) + e^{\theta e^t} \times e^{-t} - e^{-t}\bigg|_{t=0} \\
&= e^{\theta} - 1;
\end{aligned}$$

if we differentiate $M_X(t)$ twice and then set $t = 0$, we get $\mathbb{E}(X^2)$ is

$$\begin{aligned}
\mathbb{E}(X^2) &= M''(0) \\
&= \left(\theta^2 e^{2t} - \theta^2 e^t + \theta e^t + \theta - e^{-t}\right) e^{\theta e^t} + e^{-t}\bigg|_{t=0} \\
&= (2\theta - 1)\, e^{\theta} + 1.
\end{aligned}$$

Therefore, using $\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$, the variance is

$$\mathbb{V}(X) = (2\theta - 1)\, e^{\theta} + 1 - \left(e^{\theta} - 1\right)^2 = e^{\theta}\left(2\theta + 1 - e^{\theta}\right)$$

Let's verify the variance above is greater than zero. Let $\varphi(\theta) = 2\theta + 1 - e^{\theta}$, then we have $\varphi(0) = 0$, and $\varphi'(\theta) = 2 - e^{\theta}$ which is greater than zero if $\theta < \ln 2$. Thus,$\varphi(\theta)$ is an increasing function and greater than zero when $\theta < \ln 2$. Also, we have $\varphi(1) = 3 - e$ which is greater than zero since $e \approx 2.718$. Since $0 < \theta < 1$, $\varphi(\theta)$ is always greater than zero. And we know $e^{\theta} > 0$, therefore we claim that the variance is positive value. $\square$

5.5 Inference for the θ Model

5.5.1 Method of Moments

Suppose θ is unknown, we can estimate parameter θ based on the method of moments (MM) since moment expressions for this model is not cumbersome.

Let $\bar{x} = \mathbb{E}(X) = e^{\theta} - 1$, then $\hat{\theta} = \ln(\bar{x} + 1)$, where $\bar{x} < e - 1 \approx 1.718$ based on the domain of θ. But, what is the probability when $\bar{x} < e - 1$? The probability of $\bar{x} < e - 1$ can be represented as $\mathbb{P}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \frac{e-1-(e^{\theta}-1)}{\sqrt{e^{\theta}\left(2\theta+1-e^{\theta}\right)}/\sqrt{n}}\right)$, which implies $\mathbb{P}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \frac{(e-e^{\theta})\sqrt{n}}{e^{\theta/2}\sqrt{2\theta+1-e^{\theta}}}\right)$. Define $k_{\theta} = \frac{e-e^{\theta}}{e^{\theta/2}\sqrt{2\theta+1-e^{\theta}}}$, then for any fixed θ where $0 < \theta < 1$, k_{θ} should be greater than zero. Furthermore, $\mathbb{P}\left(z < \sqrt{n}k_{\theta}\right)$ would converge to 1 as n tends to ∞ by the law of large number. So, the MM estimator is good as the sample size is very large. In summary, the MM estimator is expressed as

$$\hat{\theta}_{MM} = \begin{cases} \ln(\bar{x}+1) & \text{if } \ln(\bar{x}+1) < 1 \\ 1 & \text{if } \ln(\bar{x}+1) \geq 1 \end{cases}$$

5.5.2 Maximum Likelihood Estimation

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from the θ model. By definition, the likelihood function is represented as

$$\mathcal{L}(x;\theta) = \frac{\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n}(x_i)!} \prod_{i=1}^{n} \left(1 - \frac{\theta}{1+x_i}\right) \tag{19}$$

Taking the logarithm of (19),

$$\log \mathcal{L}(\theta) = \left(\sum_{i=1}^{n} x_i\right) \log\theta - \log\left(\prod_{i=1}^{n}(x_i)!\right) + \sum_{i=1}^{n} \log\left(1 - \frac{\theta}{1+x_i}\right) \tag{20}$$

Taking the derivative with respect to θ we obtain

$$\frac{\partial \log \ell(\theta)}{\partial\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{1}{1-\theta+x_i}. \tag{21}$$

Setting $\dfrac{\partial \log \ell(\theta)}{\partial\theta} = 0$, we obtain

$$\frac{n\bar{x}}{\theta} = \sum_{i=1}^{n} \frac{1}{1-\theta+x_i}. \tag{22}$$

Unfortunately, it is difficult to solve for θ by standard manipulation. Next, we use simulation technique to evaluate the maximum likelihood estimate (MLE).

Suppose we assume the value of θ is 0.25. Substituting $\theta = 0.25$ into (18), the density function yields that

$$f(x) = \frac{0.25^x}{x!}\left(1 - \frac{0.25}{x+1}\right).$$

The simulation technique is as follows: Using R, we generate random number from 0 to 9999. And choose 100 numbers randomly without replace by three times. The following table shows how we process the simulation.

Based on the simulated data, we obtain the MLE $\hat{\theta} \approx 0.24659$ by solving for the parameter in (22), which is approximately to the origin $\theta = 0.25$.

Table 1: Simulated data for $\theta = 0.25$

x	f(x)	intervals	Sample 1	Sample 2	Sample 3	Total
0	0.7500	0-7499	76	74	77	227
1	0.2188	7500-9687	21	22	19	62
2	0.0286	9688-9973	3	4	4	11
3	0.0024	9974-9997	0	0	0	0
4	0.0002	9998-9999	0	0	0	0
5	0.0000	n/a	0	0	0	0

Similarity, if $\theta = 0.5$, then the data is simulated by

Table 2: Simulated data for $\theta = 0.5$

x	f(x)	intervals	Sample 1	Sample 2	Sample 3	Total
0	0.5000	0-4999	56	48	53	157
1	0.3750	5000-8749	37	22	38	97
2	0.1042	8750-9791	9	13	6	28
3	0.0182	9792-9973	1	0	3	4
4	0.0023	9974-9996	0	2	0	2
5	0.0002	9997-9998	0	0	0	0
6	0.0001	9999	0	0	0	0

Based on the simulated data, we obtain the MLE $\hat{\theta} \approx 0.466992$ by solving for the parameter in (22), which is approximately to the origin $\theta = 0.5$.

Therefore, the simulation method for MLE is appropriate. Furthermore, it shows that the value of MLE is slightly greater than MM estimator for the same simulated data.

6 PERMUTATION BASED DISTRIBUTION 2

6.1 First Ascent in 123-avoiding Permutation Distribution

What can be said about the location distribution of the first ascent in a random 123-avoiding permutation? We see from the discussion in Section 1.3 that there are C_n 123-avoiding permutation, and $C_{n,k}$ 123-avoiding permutation in which the first ascent is at position $(k, k+1)$. Here the probability that the first ascent is a randomly chosen 123-avoiding permutation is given by

$$f(k) = \frac{C_{n,k}}{C_n} = k\frac{(2n-k-1)!(n+1)!}{(2n)!(n-k)!}, \;\; k = 1, 2, \ldots, n,$$

which, for small k and large n, may be approximated by $f(x) = \frac{k}{2^{k+1}}$. Accordingly, let us define the geometric-like distribution on $\mathbf{Z}^+ = 1, 2, \ldots$ by

$$f(x) = \frac{x}{2^{x+1}}, x = 1, 2, \ldots.$$

We see that

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x=1}^{\infty} x \times \frac{x}{2^{x+1}} \\
&= \frac{1}{8}\sum_{x=1}^{\infty} x(x-1)(\frac{1}{2})^{x-2} + \frac{1}{4}\sum_{x=1}^{\infty} x(\frac{1}{2})^{x-1} \\
&= \frac{1}{8} \times \frac{2}{(1-\frac{1}{2})^3} + \frac{1}{4} \times \frac{1}{(1-\frac{1}{2})^2} \\
&= 3.
\end{aligned}$$

Together with the result from the previous section, we have, roughly speaking, that for a random permutation on a large $[n]$, we expect the first ascent to be at position $e-1 \approx 1.718$, whereas this value increases to 3 for a random 123-avoiding permutation. Also,

$$\begin{aligned}
\mathbb{E}(e^{tX}) &= \sum_{x=1}^{\infty} e^{tx} \times \frac{x}{2^{x+1}} \\
&= \frac{e^t}{4}\sum_{x=1}^{\infty} x(\frac{e^t}{2})^{x-1} \\
&= \frac{e^t}{4} \times \frac{1}{(1-\frac{e^t}{2})^2} \\
&= \frac{e^t}{(2-e^t)^2}, \ \text{if } \frac{e^t}{2} < 1, \ \text{i.e., if } t < \ln 2.
\end{aligned}$$

6.3 The α Model and Its M.G.F

Moreover, it makes sense, as in the geometric distribution, to define a α-analog of the above distribution defined by

$$f(x) = \frac{(\alpha-1)^2 x}{\alpha^{x+1}}, x = 1, 2, \ldots, \alpha > 1, \tag{23}$$

Note that in the case of first ascent in a 123-avoiding permutation we have $\alpha = 2$. In the general case, the m.g.f. given by

$$\begin{aligned}
\mathbb{E}(e^{tX}) &= \sum_{x=1}^{\infty} e^{tx} \times \frac{(\theta-1)^2 x}{\theta^{x+1}} \\
&= \frac{(\theta-1)^2 e^t}{\theta^2} \sum_{x=1}^{\infty} x(\frac{e^t}{\theta})^{x-1} \\
&= \frac{(\theta-1)^2 e^t}{\theta^2} \times \frac{1}{(1-\frac{e^t}{\theta})^2} \\
&= (\frac{\theta-1}{\theta-e^t})^2 \cdot e^t, \text{ if } \frac{e^t}{\theta} < 1, \text{ i.e., if i.e., if } t < \ln\theta.
\end{aligned}$$

6.4 Inference for the α Model

6.4.1 Method of Moments

Consider the population mean is

$$\begin{aligned}
\mu_X &= \sum_{x=1}^{\infty} x \times \frac{(\alpha-1)^2 x}{\alpha^{x+1}} \\
&= \frac{(\alpha-1)^2}{\alpha^3} \sum_{x=1}^{\infty} x(x-1)(\frac{1}{\alpha})^{x-2} + \frac{(\alpha-1)^2}{\alpha^2} \sum_{x=1}^{\infty} x(\frac{1}{\alpha})^{x-1} \\
&= \frac{(\alpha-1)^2}{\alpha^3} \times \frac{2}{(1-\frac{1}{\alpha})^3} + \frac{(\alpha-1)^2}{\alpha^2} \times \frac{1}{(1-\frac{1}{\alpha})^2} \\
&= \frac{2}{\alpha-1} + 1
\end{aligned}$$

If the number of observations n is large, the sample mean should be well approximated by the population mean based on the law of large numbers. Thus,

$$\bar{x} = \mu_X = \frac{2}{\alpha-1} + 1,$$

which is solved for $\hat{\alpha}$. Consequently, we obtain the MM estimate of α

$$\hat{\alpha} = 1 + \frac{2}{\bar{x}-1}$$

6.4.2 Maximum Likelihood Estimation

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from the α model. By definition, the likelihood function is represented as

$$\mathcal{L}(\alpha) = \frac{(\alpha-1)^{2n} \prod_{i=1}^{n} x_i}{\alpha^{n+\sum_{i=1}^{n} x_i}}, \tag{24}$$

Taking the logarithm of (24),

$$\log \mathcal{L}(\alpha) = \sum_{i=1}^{n} \log x_i + 2n \log(\alpha - 1) - \left(n + \sum_{i=1}^{n} x_i\right) \log \alpha, \tag{25}$$

Taking the derivative with respect to α we obtain

$$\frac{\partial \log \ell(\alpha)}{\partial \alpha} = \frac{2n}{\alpha - 1} - \frac{n + \sum_{i=1}^{n} x_i}{\alpha}. \tag{26}$$

Setting $\dfrac{\partial \log \ell(\alpha)}{\partial \alpha} = 0$, we have

$$\frac{2n}{\alpha - 1} = \frac{n + \sum_{i=1}^{n} x_i}{\theta}, \tag{27}$$

Solving for α in (27), we obtain an very interesting MLE

$$\hat{\alpha} = 1 + \frac{2}{\bar{x} - 1}.$$

Note that the estimates of MM and ML are same in α model.

6.4.3 Testing Hypotheses

By Neyman-Pearson Test, we state the null hypothesis $H_o: \quad \alpha = \alpha_0$ versus the alternative hypothesis $H_a: \quad \alpha = \alpha_1$. Assume that $\alpha_1 > \alpha_0$.

The most powerful test rejects H_o if the likelihood ratio $\frac{\mathcal{L}\alpha_1)}{\mathcal{L}(\alpha_0)} > k,$ where k is a constant. Then we have,

$$\begin{aligned}
\frac{\mathcal{L}(\alpha_1)}{\mathcal{L}(\alpha_0)} &= \frac{(\alpha_1 - 1)^{2n} \left(\prod_{i=1}^{n} x_i\right)}{\alpha_1^{n+\sum_{i=1}^{n} x_i}} \times \frac{\alpha_0^{n+\sum_{i=1}^{n} x_i}}{(\alpha_0 - 1)^{2n}(\prod_{i=1}^{n} x_i)} \\
&= \left(\frac{\alpha_1 - 1}{\alpha_0 - 1}\right)^{2n} \left(\frac{\alpha_0}{\alpha_1}\right)^{n+\sum_{i=1}^{n} x_i}
\end{aligned}$$

We would reject H_o if $\left(\frac{\alpha_1-1}{\alpha_0-1}\right)^{2n} \left(\frac{\alpha_0}{\alpha_1}\right)^{n+\sum_{i=1}^{n} x_i} > k$

which yields,

$$\left(\frac{\alpha_0}{\alpha_1}\right)^{n+\sum_{i=1}^{n} x_i} > k'$$

taking the logarithm on both sides,

$$\left(n + \sum_{i=1}^{n} x_i\right) \log \frac{\alpha_0}{\alpha_1} > \log k'$$

since $\alpha_1 > \alpha_0$, which implies $\log \left(\frac{\alpha_0}{\alpha_1}\right) < 0$, then we turn into

$$\left(n + \sum_{i=1}^{n} x_i\right) < k''$$

Therefore, $\bar{x} < k,$ where k is a constant.

7 OPEN PROBLEMS

We are trying to fit all models for appropriate real data sets. The data set we mentioned in Section 1.1 is the magnitude of earthquakes measured at the Southern California Earthquake Data Center [12]. Unfortunately, we failed to fit the model for earthquake data sets from 1962 to 2013. By now, we are still looking for data sets.

BIBLIOGRAPHY

[1] Gumbel, E. J. (1958). *Statistics of Extremes*, Columbia University Press, New York.

[2] Fisher, R. A. and Tippett, L. H. C. (1928). “Limiting forms of the frequency distribution of the largest or smallest member of a sample”, *Proc. Cambridge Philosophical Society*, **24**, 180–190.

[3] Haan, L. D. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer, New York.

[4] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London.

[5] Durrett, R. (1991). *Probability: Theory and Examples.*, Wadsworth and Brooks/Cole, Pacific Grove.

[6] Louzada, F., Roman M. and Cancho, V. G. (2011). “The complementary exponential geometric distribution: Model, properties, and a comparison with its counterpart”, *Computational Statistics and Data Analysis*, **55**, 2516–2524.

[7] Louzada, F., Bereta E. M. P. and Franco, M. A. P. (2012). “On the distribution of the minimum or maximum of a random number of i.i.d. lifetime random variables”, *Applied Mathematics*, **3**, 350–353.

[8] Morais, A. L. and Barreto-Souza W. (2011). "A compound class of Weibull and power series distributions", *Computational Statistics and Data Analysis*, **55**, 1410–1425.

[9] Johnson, N., Kotz S. and Balakrishnan N. (1995). *Continuous Univariate Distributions, Vol. 2, 2nd Edition*, John Wiley, New York.

[10] Kotz S. and Dorp J. V. (2004). *Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications*, World Scientific Publishing Co., Singapore.

[11] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

[12] Earthquake data set, available at

`http://www.data.scec.org/eq-catalogs/date_mag_loc.php (March 2014).`

[13] Chung, F., Lu, L. and Vu, V. (2003). "Eigenvalues of random power law graphs", *Annals of Combinatorics*, **7**, 21–33.

[14] Aiello, B., Chung, F. and Lu, L. (2001). "A random graph model for power law graphs", *Experimental Mathematics*, **10**, 53–66.

[15] Catalan, E. (1887). "Sur les nombres de Segner", *Rend. Circ. Mat. Palermo*, **1**, 190–201.

[16] Kitaev, S. (2011). *Patterns in Permutations and Words*, Springer, New York.

[17] Hao, J. and Godbole, A. (2014+). "Distribution of the Maximum and Minimum of a Random Number of Bounded Random Variables", submitted, available at `http://arxiv.org/pdf/1403.1302v1.pdf` (March 2014).

[18] Connolly, S., Gabor, Z. and Godbole, A. (2014+). "The location of the first ascent in a 123-avoiding permutation", submitted, available at `http://arxiv.org/pdf/1401.2691.pdf` (March 2014).

[19] Godbole, A. and Hao, J. (2014+). "Telescoping Sums, Permutations, and First Occurrence Distributions", submitted, available at `http://arxiv.org/pdf/1403.7665v1.pdf` (April 2014).

[20] Topp, C. W. and Leone, F. C. (1955). "A Family of J-Shaped Frequency Functions", *Journal of the American Statistical Association*, **50**, 209–219.

[21] Genç, A. İ. (2012). "Moments of order statistics of Topp-Leone distribution", *Statistical Papers*, **53**, 117–131.

APPENDICES

In these Appendices, the code used in R version 3.0.3. The R project for statistical computing and graphics is a free programming software. All comments given in R are preceded by the # symbol.

A Plotting Probability Density Functions

```
#SUG-Y#

y=seq(0,1,0.001)

theta=c(0.1,0.25,0.5,0.75,1)

fy1=theta[1]/(1-(1-theta[1])*y)^2

fy2=theta[2]/(1-(1-theta[2])*y)^2

fy3=theta[3]/(1-(1-theta[3])*y)^2

fy4=theta[4]/(1-(1-theta[4])*y)^2

fy5=theta[5]/(1-(1-theta[5])*y)^2

plot(y,fy1,type="l",ylim=c(0,6),lty=1,lwd=2,ylab="pdf")

lines(y,fy2,type="l",lty=2,lwd=2)

lines(y,fy3,type="l",lty=3,lwd=2)

lines(y,fy4,type="l",lty=4,lwd=2)

lines(y,fy5,type="l",lty=5,lwd=2)

legend("topleft",c("theta=0.1","theta=0.25","theta=0.5","theta=0.75","theta=1"),
        cex=0.8,lty=1:5, lwd=2, bty="n")
```

```
#SUG-Z#

z=seq(0,1,0.001)

theta=c(0.1,0.25,0.5,0.75,1)

fz1=theta[1]/(theta[1]+(1-theta[1])*z)^2

fz2=theta[2]/(theta[2]+(1-theta[2])*z)^2

fz3=theta[3]/(theta[3]+(1-theta[3])*z)^2

fz4=theta[4]/(theta[4]+(1-theta[4])*z)^2

fz5=theta[5]/(theta[5]+(1-theta[5])*z)^2

plot(z,fz1,type="l",ylim=c(0,6),lty=1,lwd=2,ylab="pdf")

lines(z,fz2,type="l",lty=2,lwd=2)

lines(z,fz3,type="l",lty=3,lwd=2)

lines(z,fz4,type="l",lty=4,lwd=2)

lines(z,fz5,type="l",lty=5,lwd=2)

legend("topright",c("theta=0.1","theta=0.25","theta=0.5","theta=0.75","theta=1"),

        cex=0.8,lty=1:5, lwd=2, bty="n")


#CSUG-Y#

theta=c(0.1,0.25,0.5,0.75)

y1=seq(0,1-theta[1],0.001)

y2=seq(0,1-theta[2],0.001)

y3=seq(0,1-theta[3],0.001)

y4=seq(0,1-theta[4],0.001)
```

```
fy1=theta[1]/((1-theta[1])*(1-y1)^2)

fy2=theta[2]/((1-theta[2])*(1-y2)^2)

fy3=theta[3]/((1-theta[3])*(1-y3)^2)

fy4=theta[4]/((1-theta[4])*(1-y4)^2)

plot(y1,fy1,type="l",xlim=c(0,1),ylim=c(0,10),lty=1,lwd=2,xlab="y",ylab="pdf")

lines(y2,fy2,type="l",lty=2,lwd=2)

lines(y3,fy3,type="l",lty=3,lwd=2)

lines(y4,fy4,type="l",lty=4,lwd=2)

legend("topleft",c("theta=0.1","theta=0.25","theta=0.5","theta=0.75"),

        cex=0.8,lty=1:4, lwd=2, bty="n")


#CSUG-Z#

theta=c(0.1,0.25,0.5,0.75)

z1=seq(theta[1],1,0.001)

z2=seq(theta[2],1,0.001)

z3=seq(theta[3],1,0.001)

z4=seq(theta[4],1,0.001)

fz1=theta[1]/((1-theta[1])*(z1^2))

fz2=theta[2]/((1-theta[2])*(z2^2))

fz3=theta[3]/((1-theta[3])*(z3^2))

fz4=theta[4]/((1-theta[4])*(z4^2))

plot(z1,fz1,type="l",xlim=c(0,1),ylim=c(0,10),lty=1,lwd=2,xlab="z",ylab="pdf")
```

```
lines(z2,fz2,type="l",lty=2,lwd=2)

lines(z3,fz3,type="l",lty=3,lwd=2)

lines(z4,fz4,type="l",lty=4,lwd=2)

legend("topright",c("theta=0.1","theta=0.25","theta=0.5","theta=0.75"),

        cex=0.8,lty=1:4, lwd=2, bty="n")


#STL#

x=seq(0,1,0.001)

a=c(0.5,1,2,5)

f1=2*a[1]*(1-x)*x^(a[1]-1)*(2-x)^(a[1]-1)

f2=2*a[2]*(1-x)*x^(a[2]-1)*(2-x)^(a[2]-1)

f3=2*a[3]*(1-x)*x^(a[3]-1)*(2-x)^(a[3]-1)

f4=2*a[4]*(1-x)*x^(a[4]-1)*(2-x)^(a[4]-1)

plot(x,f1,type="l",ylim=c(0,6),lty=1,lwd=2,ylab="pdf")

lines(x,f2,type="l",lty=2,lwd=2)

lines(x,f3,type="l",lty=3,lwd=2)

lines(x,f4,type="l",lty=4,lwd=2)

legend("topright",c("a=0.5","a=1","a=2","a=5"),cex=0.8,lty=1:4, lwd=2, bty="n")
```

B Simulation for Parameter Estimation of Permutation Based Model 1

```
#theta=1/2#

x=c(0,1,2,3,4,5)

f=(0.5^x)/factorial(x)*(1-1/(2*(x+1)))

x<-sample(0:9999,100,replace=F) # Generating random number

x0<-x[x<=4999]

x1<-x[x>=5000 & x<=8749]

x2<-x[x>=8750 & x<=9791]

x3<-x[x>=9792 & x<=9973]

x4<-x[x>=9974 & x<=9996]

x5<-x[x>=9997 & x<=9998]

x6<-x[x>=9999]

l0<-length(x0)

l1<-length(x1)

l2<-length(x2)

l3<-length(x3)

l4<-length(x4)

l5<-length(x5)

l6<-length(x6)

#theta=1/4#

x=c(0,1,2,3,4,5)

f=((1/4)^x)/factorial(x)*(1-(1/4)/(x+1))
```

```
x<-sample(0:9999,100,replace=F)

x0<-x[x<=7499]

x1<-x[x>=7500 & x<=9687]

x2<-x[x>=9688 & x<=9973]

x3<-x[x>=9974 & x<=9997]

x4<-x[x>=9998 & x<=9999]

x5<-x[x>=10000]

l0<-length(x0)

l1<-length(x1)

l2<-length(x2)

l3<-length(x3)

l4<-length(x4)

l5<-length(x5)

#theta=3/4#

x=c(0,1,2,3,4,5)

f=((3/4)^x)/factorial(x)*(1-(3/4)/(x+1))

x<-sample(0:9999,100,replace=F)

x0<-x[x<=2499]

x1<-x[x>=2500 & x<=7187]

x2<-x[x>=7188 & x<=9296]

x3<-x[x>=9297 & x<=9867]

x4<-x[x>=9868 & x<=9979]
```

```
x5<-x[x>=9980 & x<=9996]

l0<-length(x0)

l1<-length(x1)

l2<-length(x2)

l3<-length(x3)

l4<-length(x4)

l5<-length(x5)
```

VITA

JIE HAO

Education:	B.S. Statistics, North China University of Technology China, Beijing 2012 M.S. Mathematics, East Tennessee State University Johnson City, Tennessee 2014
Professional Experience:	Graduate Assistant, East Tennessee State University, Johnson City, Tennessee, 2012–2014