5-2017

A Distribution of the First Order Statistic When the Sample Size is Random

Vincent Z. Forgo Mr
East Tennessee State University

A Distribution of the First Order Statistic when the Sample Size is Random

---

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

Vincent Forgo

May 2017

Bob Price, Ph.D., Chair

Nicole Lewis, Ph.D.

JeanMarie Hendrickson, Ph.D.

Keywords: First order statistic, random sample size, Cumulative density function,

Probability density function, Expectation, Variance, Percentile

ABSTRACT

A Distribution of the First Order Statistic when the Sample Size is Random

by

Vincent Forgo

Statistical distributions also known as probability distributions are used to model a random experiment. Probability distributions consist of probability density functions (pdf) and cumulative density functions (cdf). Probability distributions are widely used in the area of engineering, actuarial science, computer science, biological science, physics, and other applicable areas of study. Statistics are used to draw conclusions about the population through probability models. Sample statistics such as the minimum, first quartile, median, third quartile, and maximum, referred to as the five-number summary, are examples of order statistics. The minimum and maximum observations are important in extreme value theory. This paper will focus on the probability distribution of the minimum observation, also known as the first order statistic, when the sample size is random.

ACKNOWLEDGMENTS

I am grateful for the great support from the faculty of Department of Mathematics and Statistics at East Tennessee State University. I appreciate the tremendous work by committee members: Dr. Robert Price (committee and department Chair), Dr. Nicole Lewis, and Dr. JeanMarie Hendrickson for making this paper possible. I am more grateful to Dr. Robert Price for how instrumental He has being throughout this paper.

TABLE OF CONTENTS

1 INTRODUCTION

The concept of order statistics is familiar in areas of finance and insurance (Risk assessment). The order statistics of a random sample $X_1, X_2, ...X_n$ are defined as $X_{(1)} \leq X_{(2)} \leq .... \leq X_{(n)}$. A situation can occur in actuarial science with a joint life insurance. The policy pays out when one of the spouse's dies. In this problem, we want to know the distribution of the minimum payment, which is the random variable of the two life spans. Another form of application of order statistics is about a machine, which may run on 15 batteries and shuts off when the seventh battery dies. We may want to know the distribution of $X_{(7)}$. Thus, the distribution of the random variable of the seventh longest lasting battery.

In order statistics the variables are considered as independent and identically distributed, *iid*. The cumulative distribution function of the n^{th} order statistic is given as

$$F_n(x) = P\left\{all X_i \leq x\right\} = [P(X \leq x)]^n = [F_n(x)]^n.$$

This implies the cumulative distribution is $F_n(x)$ for this random variable. The cumulative distribution function (cdf) of the first order statistic or the minimum is

$$(1 - P[(X_i > x)])^n.$$

The general formula of the cumulative distribution for the $k^{(th)}$ order statistic is given as

$$\sum_{j=k}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j}.$$

Order statistics is among the most essential functions of a set of random variables that are studied in probability and statistics. There is natural interest in studying the highs and lows of a sequence, and the order statistics help in understanding concentration of probability in a distribution[4]. It is important to note that the variables in the sample are independent and identically distributed but because of the the sequential order associated with order statistics, the order statistics is not distributed identically and independently. Since the variables in the sample appear in order, there is a minimum and maximum order statistics. Therefore, the n^{th}(maximum) order statistic has a pdf of

$$f_{x(n)} = n[F(x)]^{n-1}f(x)$$

and the first order statistic(minimum) will have a pdf of

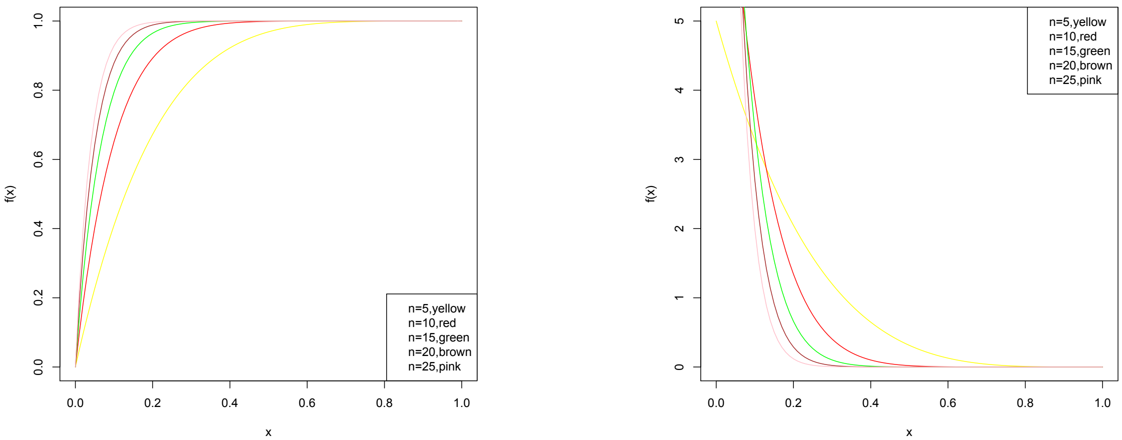
$$f_{x(1)} = n[1 - F(x)]^{n-1}.$$

2 FIXED SAMPLE SIZE

Now let us consider X_i to be *iid* continuous random variables i.e. all the random variables have equal probability distribution and mutually independent of each other such that the random variables follow a uniform distribution, $X_i \sim unif(0,1)$ and $Z = min(X_1, ..., X_n)$. If n is fixed then

$$P(Z \leq z) = 1 - P(Z > z) = 1 - [1 - F_X(z)]^n = 1 - [1 - z]^n, \quad 0 \leq z \leq 1.$$

where $F_x(z)$ is the cdf of the uniform distribution. The idea, of starting with $1 - P(Z > z)$ is because we can say that if the first order statistics is the smallest, then automatically z is less than Z. Hence we can find $P(Z \leq z)$ by starting with $1 - P(Z > z)$. The pdf is generated by taking the derivative of the cdf which gives f(z) $= n[1 - z]^{n-1}$. The cdf and pdf is graphically shown in figure 1 and becomes steeper as n gets large.



(a) Cummulative distribution function plots for different sample sizes.

(b) Probability density function plots for different sample sizes.

Figure 1: The cdf and pdf of the smallest order statistic when the underlying distribution is uniform.

3 TRUNCATED POISSON MIXTURE

The Poisson distribution is a discrete probability distribution. The Poisson distribution is some times truncated, i.e. the random variables are assigned numbers that are greater than zero. The Poisson distribution is a discrete distribution used for the interval counts of events that randomly occur in given interval (or space)[3]. The probability mass function (pmf) is

$$P(N = n) = \frac{\lambda^n e^{-\lambda}}{n!}, n = 0, 1, 2, 3...; \lambda > 0.$$

with expectation $E(N) = \lambda$ and variance $V(N) = \lambda$. The probability generating function of the Poisson distribution is $G(t) = e^{\lambda(t-1)}$ and the moomemt generating function (mgf) is $M(t) = e^{\lambda(e^t-1)}$, where the events occur on a given time t.

The truncated Poisson is a discrete probability distribution which is used to describe events that occur per unit time and can not be a zero event. In this case, the starting point will not be zero but 1. This process is termed as the truncated Poisson distribution or the zero truncated Poisson distribution. The pmf of the zero truncated Poisson is given below as

$$P(N = n) = \frac{\lambda^n e^{-\lambda}}{(1 - e^{-\lambda})n!}, \quad n = 1, 2 \ldots.$$

with an expectation of $E(N) = \frac{\lambda}{1-e^{-\lambda}}$ and a variance of $V(N) = \frac{\lambda}{(1-e^{-\lambda})^2}$.

If the random variable X_i follows a continuous probability distribution and $Z|N = min(X_1, ..., X_n)$, then we can find a distribution for the first other statistic $X_{(1)}$ when the sample size is fixed or random. In the next section, the paper will focus more on a general formula for finding the cdf and pdf of a random variable with any continuous probability distribution like uniform, exponential, etc. and a random sample size(N).

If N is a random sample size and follows a truncated Poisson distribution then for any continuous distribution of X_i, we can find the cdf of the distribution by using the generalised formula for $P(Z > z)$, i.e,

$$P(Z > z) = \sum_n [1 - F_x(z)]^n \frac{e^{-\lambda}\lambda^n}{(1-e^{-\lambda})n!} = \frac{e^{-\lambda}}{(1-e^{-\lambda})}[e^{(1-F_x(z))\lambda} - 1].$$

The general cdf will be

$$F(z) = 1 - \frac{e^{-\lambda}}{(1-e^{-\lambda})}[e^{(1-F_x(z))\lambda} - 1]$$

with a pdf

$$f(z) = \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}} f_x(z)[e^{(1-F_x(z)\lambda}], \quad \lambda > 0,$$

where $F_x(z)$ and $f_x(z)$ are the cdf and pdf of the continuous random variable X_i, respectively. If X_i follows a continuous distribution which is not closed like the normal distribution, then we can use R functions like the pnorm and dnorm to find the cdf and pdf respectively, i.e., $F_x(z) = pnorm(x)$ and $f_x(z) = dnorm(x)$.

3.1 Uniform - Truncated Poisson Mixture Distribution

If the random variable X_i follows a continuous probability distribution and $Z = min(X_1, ..., X_N)$, then we can find a distribution for the first other statistic $X_{(1)}$. In this section we are focused on a distribution of the first order statistics with an underlying uniform distribution and a random sample which follows a truncated Poisson distribution. If N is random then

$$P(Z > z) = \sum_n P(Z > z|N = n)P(N = n), 0 < z < 1$$

where $P(Z > z|N)$ is the conditional distribution and $P(N = n)$ is the marginal distribution. The idea of N being random will be widely explored in this paper. Our new distribution is mainly based on the idea of first order statistics and N following truncated Poisson. Let us consider $X_i \sim Unif(0,1)$ and $Z|N = min\{X_1, X_2, ..., X_N\}$ where N is sample size and is random with distribution

$$P(N = n) = \frac{e^{-\lambda}\lambda^n}{(1 - e^{-\lambda})n!}, n = 1, 2, 3, \dots.$$

Then

$$P(Z > z) = \sum_{n=1}^{\infty}(1 - z)^n \frac{e^{-\lambda}\lambda^n}{(1 - e^{-\lambda})n!}$$

$$= \sum_{n=0}^{\infty}(1 - z)^n \frac{e^{-\lambda}\lambda^n}{(1 - e^{-\lambda})n!} - (1 - z)^0 \frac{e^{-\lambda}\lambda^0}{(1 - e^{-\lambda})0!}$$

$$= \frac{e^{-\lambda}}{(1 - e^{-\lambda})} \frac{\sum_{n=0}^{\infty}((1 - z)\lambda)^n}{n!} - \frac{e^{-\lambda}}{(1 - e^{-\lambda})}.$$

Using the definition of $\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{\lambda}$ we can simplify the above equation as

$$\frac{e^{-\lambda}}{(1 - e^{-\lambda})}[e^{(1-z)\lambda} - 1].$$

Hence the cumulative distribution function (cdf) is given as $F(z) = 1 - \frac{e^{-\lambda}}{(1-e^{-\lambda})}[e^{(1-z)\lambda} - 1]$. where $0 < z < 1, \lambda > 0$.

The probability density function (pdf) can be derived by taking the derivative of the cdf with respect to z. The pdf of a first order statistic when the underlying distribution is uniform with a random sample that is a truncated Poisson is

$$f(z) = \frac{\lambda e^{-\lambda}}{1 - e^{-\lambda}}[e^{(1-z)\lambda}], \quad 0 < z < 1, \lambda > 0.$$

It can be proven that $f(z)$ satisfies the conditions of a pdf i.e. $\int_0^1 f(z)dz = 1$

$$\int_0^1 f(z)dz = \int_0^1 \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}[e^{(1-z)\lambda}]dz = \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\int_0^1 [e^{(1-z)\lambda}]$$

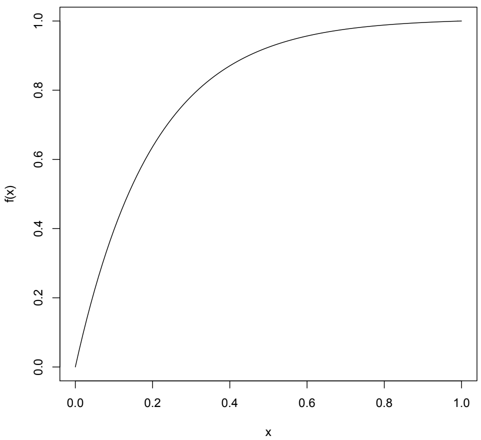
$$\frac{1}{e^{\lambda}-1}[e^{\lambda}-1] = 1.$$

From the distribution generated, the expectation is given as

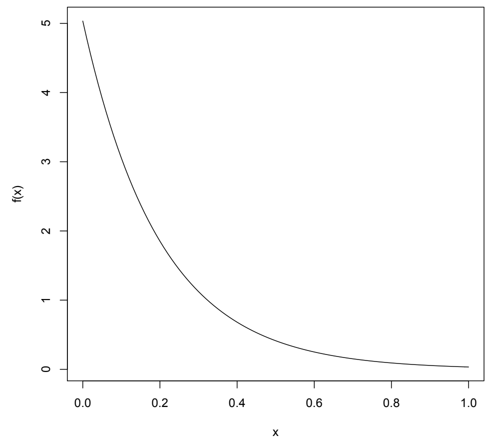
$$E(Z) = \int_0^1 \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}} z[e^{(1-z)\lambda}]dz = \frac{e^{\lambda}-\lambda-1}{\lambda(e^{\lambda}-1)}.$$

The moment generating function (mgf) can be used to estimate both the expectation and variance. The mgf of the distribution is given as

$$M(t) = E(e^{tz}) = \int_0^1 e^{tz} f(z)dz = \int_0^1 e^{tz}\frac{e^{-\lambda}}{1-e^{-\lambda}}\left(e^{(1-z)\lambda}\right)dz$$
$$= \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\left(\frac{e^{\lambda}-e^{t}}{\lambda-t}\right).$$



(a) Cumulative distribution function plot



(b) Probability density function plot

Figure 2: The cdf and pdf of the smallest order statistic when the underlying distribution is uniform and random sample size which follows a truncated Poisson where rate $(\lambda) = 5$.

From Figure 3, it is clear that the cdf in both the random and fixed cases tend to be the same as λ and n increases. The behaviour of the cdf as λ and n increases shows a steep and sharp turn closer to 1. This implies that the larger n and λ gets, the more steeper the curve becomes and the fixed sample size (n) and the random sample size (N) all tend to have the same cdf.

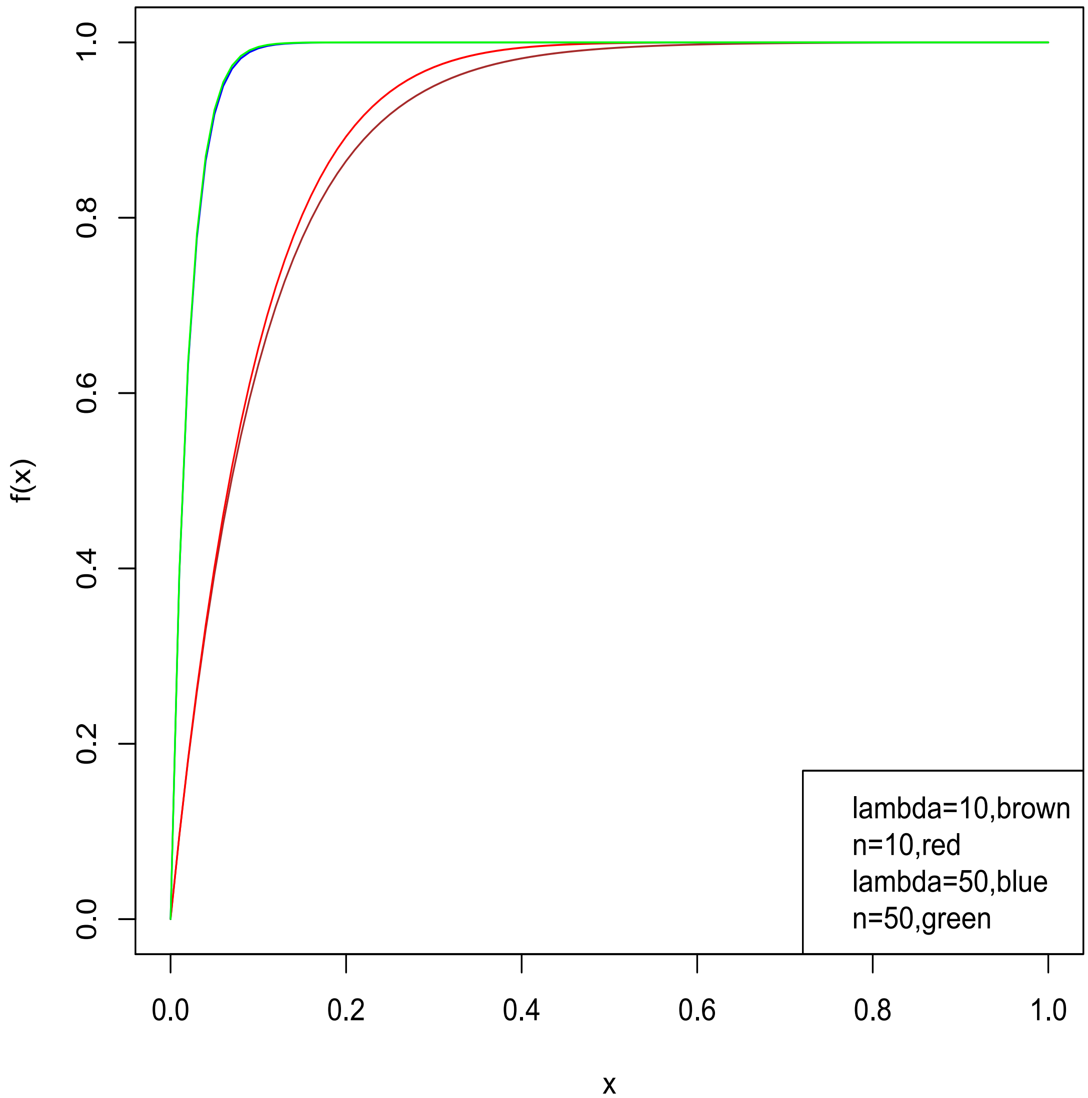


Figure 3: A cumulative distribution function plots of different samples (n) and different rates (λ) when the underlying distribution is uniform for fixed sample and random sample which follows a truncated Poisson.

The percentile function is relevant in statistics because it can be used to indicate the value below a certain percentage. The percentile function can also be used to calculate the lower quartile(Q_1), median(Q_2) and the upper quartile(Q_3). The percentile function of the first order statistic when the underlying distribution is uniform and a random sample size that follows a truncated Poisson is

$$P = F(\mu) = \int_{-\infty}^{\mu} f(y)dy.$$

where $f(y)$ is the probability density function(pdf). From the uniform truncated Poisson distribution, the probability distribution function pdf is

$\frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}[e^{(1-z)\lambda} - 1], \lambda > 0$. The percentile function is generated by

$$\int_0^{\mu} \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}[e^{(1-z)\lambda}]dz = \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}} \int_0^{\mu} [e^{(1-z)\lambda}]dz$$

$$= \frac{e^{-\lambda}}{1-e^{-\lambda}}(e^{\lambda} - e^{\lambda(1-\mu)})$$

$$P = \frac{e^{-\lambda}}{1-e^{-\lambda}}(e^{\lambda} - e^{\lambda(1-\mu)}), \lambda > 0, \mu > 0$$

From the above percentile equation, the 50th percentile(median) is calculated in terms of λ as

$$0.5 = \frac{e^{\lambda} - e^{\lambda(1-\mu)}}{e^{\lambda} - 1} \Rightarrow \mu = \frac{\lambda - log(1-e^{\lambda})\left(0.5 + \frac{e^{\lambda}}{1-e^{\lambda}}\right)}{\lambda}.$$

3.2 Exponential - Truncated Poisson Mixture Distribution

In this section, we are focused on a distribution of the first order statistics with an underlying exponential distribution and a random sample which follows a truncated Poisson. If $Z|N = min\{X_1, ..., X_N\}$, where N is the sample size which is random with a distribution

$$P(N = n) = \frac{e^{-\lambda}\lambda^n}{(1 - e^{-\lambda})n!}.$$

We have

$$P(Z > z) = \sum_{n=1}^{\infty}(e^{-z\mu})^n \frac{e^{-\lambda}\lambda^n}{(1 - e^{-\lambda})n!}$$

$$P(Z > z) = \sum_{n=1}^{\infty} \frac{e^{-\lambda - zn\mu}\lambda^n}{(1 - e^{-\lambda})n!} = \sum_{n=0}^{\infty} \frac{e^{-\lambda - zn\mu}\lambda^n}{(1 - e^{-\lambda})n!} - \frac{e^{-\lambda}}{1 - e^{-\lambda}}$$

$$= \frac{e^{-\lambda}}{1 - e^{-\lambda}}\left[\sum_{n=0}^{\infty} \frac{e^{-zn\mu}\lambda^n}{n!} - 1\right]$$

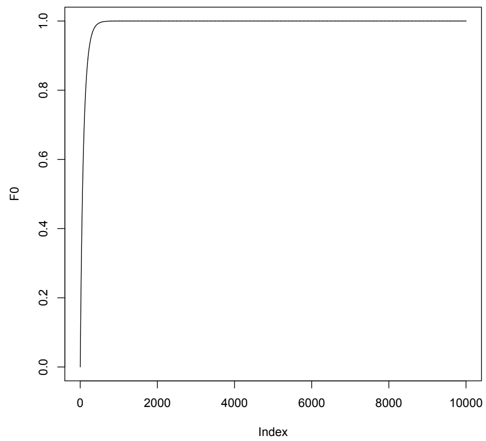
$$= \frac{e^{-\lambda}}{(1 - e^{-\lambda})}[e^{\lambda e^{-z\mu}} - 1].$$

The cdf of an underlying exponential distribution with a random sample size which follows a truncated Poisson is

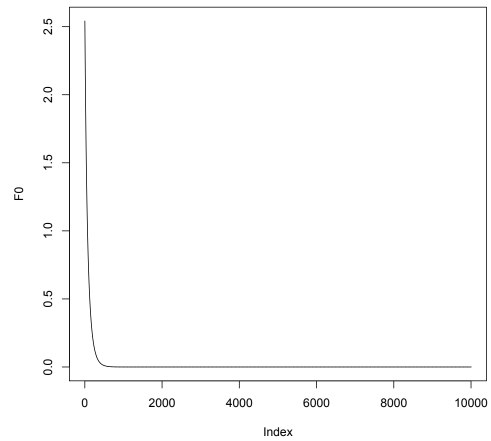
$$F(z) = 1 - \frac{e^{-\lambda}}{1 - e^{-\lambda}}[e^{\lambda e^{-z\mu}} - 1], \quad \lambda > 0, \quad z > 0, \quad \mu > 0.$$

and a pdf of

$$f(z) = \frac{d}{dz}\left(1 - \frac{e^{-\lambda}}{1 - e^{-\lambda}}[e^{\lambda e^{-z\mu}} - 1]\right) = \frac{\lambda\mu}{e^{\lambda} - 1}[e^{\lambda e^{-z\mu} - z\mu}], \quad \lambda > 0, \quad z > 0, \quad \mu > 0.$$



(a) Cumulative density function plot.



(b) Probability density function plot.

Figure 4: A figure representing the cdf and pdf with an underlying exponential distribution and a random sample size which follows a truncated Poisson when rate $(\lambda) = 0.5$ and $\mu = 1$.

4 TRUNCATED BINOMIAL MIXTURE

The truncated binomial distribution is a discrete probability distribution with a probability mass function

$$P(N = n) = \frac{\binom{k}{n} p^n (1-p)^{k-n}}{1-(1-p)^k}$$

for $n = 1, 2, ...k$. Where k is the number of success and p is the probability of success with an expectation

$$E(N) = \frac{kp}{1-(1-p)^k}$$

and a variance

$$V(N) = \frac{kp(1-p-(1-p+kp))(1-p)^k}{(1-(1-p)^k)^2}.$$

The binomial distribution is often used to model the number of success(k) among a sample of size(n).

In this section of the paper, we will focus on finding a general cdf and pdf when the random sample size follows a truncated binomial distribution and the random variable X_i follows a continuous distribution that can be exponential, uniform, etc. The general formula for $P(Z > z)$ will be

$$P(Z > z) = \frac{\sum_n [1 - F_x(z)]^n \binom{k}{n} p^n (1-p)^{k-n}}{1-(1-p)^k} = \frac{1}{1-(1-p)^k}\Big[(1-F_x(z)p)^k - (1-p)^k\Big].$$

The general cdf will be

$$F(z) = 1 - \frac{1}{1-(1-p)^k}\Big[(1-F_x(z)p)^k - (1-p)^k\Big]$$

and a pdf of

$$f(z) = \frac{kp(1 - F_x(z)p)^{k-1} f_x(z)}{1 - (1-p)^k}, \quad 0 \le p \le 1,$$

where $F_x(z)$ and $f_x(z)$ are the cdf and pdf of the continuous random variable X_i respectively.

4.1 Uniform-Truncated Binomial Mixture Distribution

In this section, we are focused on finding a distribution of the first order statistics with an underlying uniform distribution and a random sample which follows a truncated binomial distribution. If $Z|N = min\{X_1, ..., X_N\}$ where N is the sample size which is random with a distribution

$$P(N = n) = \frac{\binom{k}{n} p^n (1-p)^{k-n}}{1 - (1-p)^k}$$

for n=1, 2, ...k. This implies that

$$P(Z>z) = \frac{\sum_{n=1}^{k} (1-z)^n \binom{k}{n} p^n (1-p)^{k-n}}{1-(1-p)^k} = \frac{1}{1-(1-p)^k}\left[\sum_{n=1}^{k} (1-z)^n \binom{k}{n} p^n (1-p)^{k-n}\right]$$

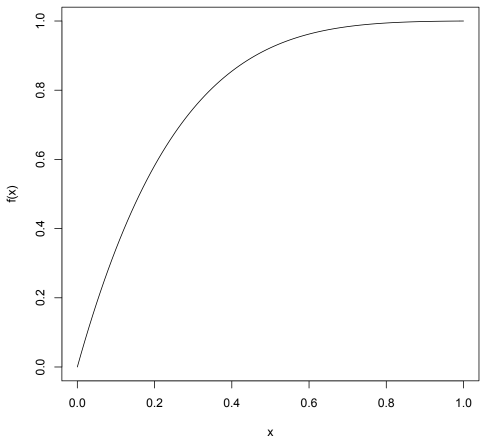
$$= \frac{1}{1-(1-p)^k}\left[\sum_{n=0}^{k} (1-z)^n \binom{k}{n} p^n (1-p)^{k-n} - (1-p)^k\right]$$

$$= \frac{1}{1-(1-p)^k}\left[\sum_{n=0}^{k} \binom{k}{n} ((1-z)p)^n (1-p)^{k-n} - (1-p)^k\right]$$

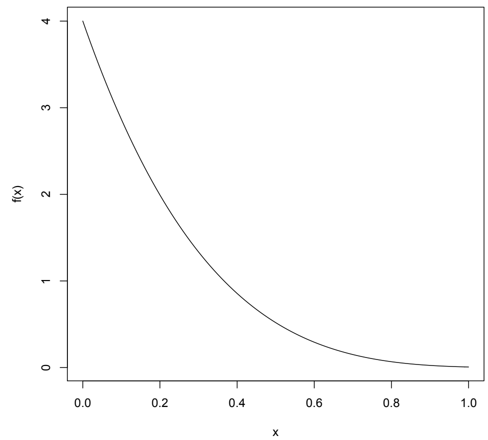
$$= \frac{1}{1-(1-p)^k}\left[((1-z)p + (1-p))^k - (1-p)^k\right] = \frac{1}{1-(1-p)^k}\left[(1-zp)^k - (1-p)^k\right].$$

Thus the cdf of an underlying uniform distribution with a random sample which follows a truncated binomial distribution is

$$F(z) = 1 - \frac{1}{1-(1-p)^k}\left[(1-zp)^k - (1-p)^k\right]$$



(a) Cumulative distribution function plot with k= 5 and p=0.8.



(b) Probability density function plot with k=5 and p=0.8.

Figure 5: A figure representing the cdf and pdf with an underlying uniform distribution and random sample size which follows truncated binomial.

with a probability density function

$$f(z)=\frac{kp(1-zp)^{k-1}}{1-(1-p)^k}$$

and

$$E(Z)=\int_0^1 z\frac{kp(1-zp)^{k-1}}{1-(1-p)^k}dz=\frac{1-(1-p)^k(kp+1)}{p(k+1)}.$$

Figure 5a shows the cdf of an underlying uniform distribution and a random sample size that follows truncated binomial distribution. Figure 5b is the pdf of an underlying uniform distribution and a random sample size which follows a truncated binomial distribution.

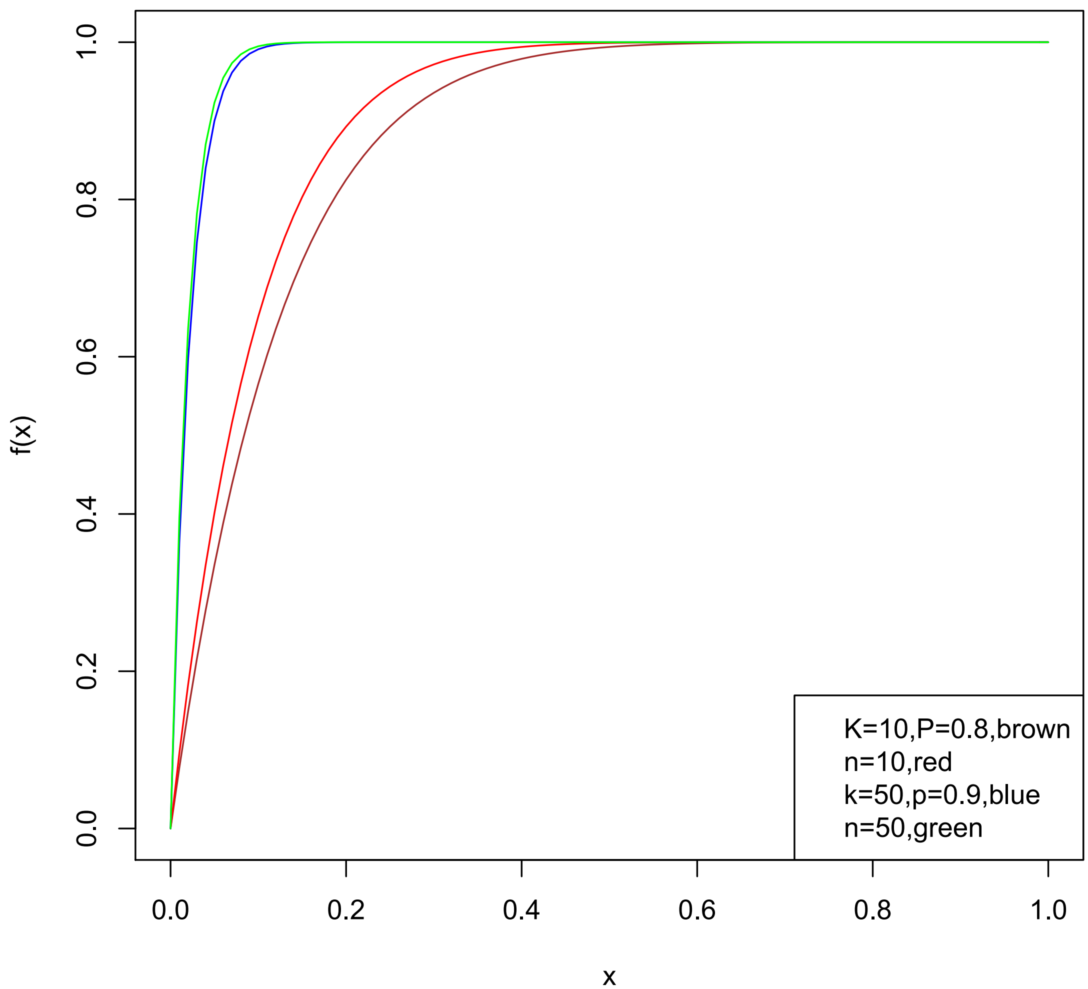


Figure 6: A cumulative distribution function plots of different samples (n) and different p and k when the underlying distribution is uniform for fixed sample and random sample that follows a truncated binomial.

From Figure 6, it is clear that the cdf in both the random and fixed cases tend to be the same as p, k and n increases. The behaviour of the cdf as p, k and n increases shows a steep and sharp turn closer to 1. This implies that the larger p, k and n gets, the steeper the curve becomes.

The percentile function of an underlying uniform distribution with a random sample size which follows a truncated binomial distribution is given as

$$P = \int_0^{\mu} \frac{kp(1-zp)^{k-1}}{1-(1-p)^k} dz = \frac{(1-\mu p)^k - 1}{(1-p)^k - 1}.$$

From the Percentile function, the 50th percentile (μ) or the median is calculated using the relation

$$P = \frac{(1-p\mu)^k - 1}{(1-p)^k - 1}$$

Hence

$$\mu = \frac{1 - (0.5 - \frac{1}{1-(1-p)^k})((1-p)^k - 1))}{p}$$

$0 \leq p \leq 1$, $P=0.5$ and $0<z<1$.

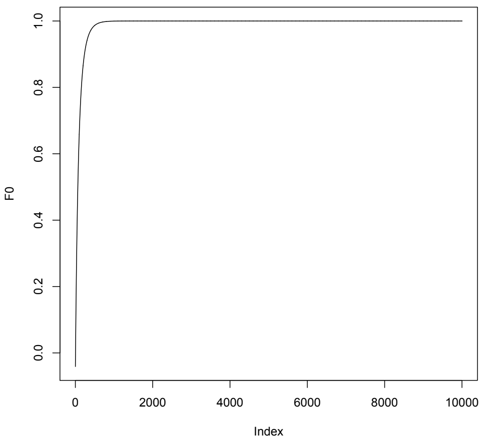
4.2 Exponential-Truncated Binomial Mixture Distribution

In this section, we are focused on finding a distribution of the first order statistics with an underlying exponential distribution and a random sample which follows a truncated binomial distribution. If $Z|N = min\{X_1, ..., X_N\}$ where N is the sample size which is random with a distribution

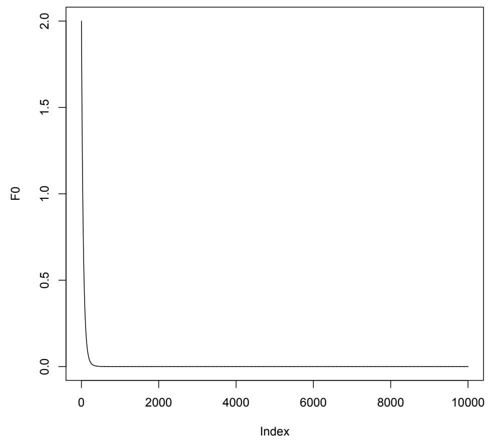
$$P(N=n) = \frac{\binom{k}{n} p^n (1-p)^{k-n}}{1-(1-p)^k}$$

for $n = 1, 2, ...k$. This implies that

$$P(Z>z) = \frac{\sum_{n=1}^{k} (e^{-z\mu})^n \binom{k}{n} p^n (1-p)^{k-n}}{1-(1-p)^k} = \frac{1}{1-(1-p)^k} \left[\sum_{n=1}^{k} (e^{-z\mu})^n \binom{k}{n} p^n (1-p)^{k-n} \right]$$



(a) Cumulative distribution function plot when p= 0.2, k=5 and μ=1.



(b) Probability density function plot when p= 0.8, k=5 and μ=1.

Figure 7: The cdf and pdf of the smallest order statistic when the underlying distribution is exponential and random sample size which follows a truncated binomial.

$$\frac{1}{1-(1-p)^k}\left[\sum_{n=0}^{k}(e^{-z\mu})^n\binom{k}{n}p^n(1-p)^{k-n}-(1-p)^k\right]=\frac{(pe^{-z\mu}+1-p)^k-(1-p)^k}{1-(1-p)^k}.$$

Thus the cdf of an underlying exponential distribution with a random sample which follows a truncated binomial distribution is

$$F(z)=1-\frac{(pe^{-z\mu}+1-p)^k-(1-p)^k}{1-(1-p)^k},\quad \mu>0,\quad 0\leq p\leq 1,\quad z>0$$

with a probability density function

$$f(z)=k\mu p^{1-k}e^{-z\mu}(pe^{-z\mu}-p+1)^{k-1},\quad \mu>0,\quad 0\leq p\leq 1,\quad z>0.$$

5 TRUNCATED GEOMETRIC MIXTURE

In this section of the paper, the focus will be finding a distribution of the first order statistic when X_i is any continuous distribution and the random sample size follows a geometric distribution. The geometric distribution is a discrete probability distribution which is used to represent the first outcome of a specific event with a probability p of the event occurring. The pmf of the geometric distribution is

$$P(N{=}n){=}p(1-p)^n, \quad n{=}0,1,2,\ldots$$

with an expectation

$$E(N){=}\frac{1-p}{p}$$

and variance

$$V(N){=}\frac{1-p}{p^2} \quad 0{\leq}p{\leq}1.$$

The truncated geometric distribution is a modified form of the geometric distribution with a probability mass function (pmf)

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

with an expectation

$$E(N){=}\frac{1}{p}$$

and variance

$$V(N){=}\frac{1-p}{p^2} \quad 0{\leq}p{\leq}1.$$

If N follows a truncated geometric distribution then for any continuous distribution of X_i, the generalised formula for $P(Z>z)$ will be

$$P(Z>z)=\sum_n P(Z>z|N=n)P(N=n)$$

$$=\sum_{n=1}^{\infty}(1-F_x(z))^n p(1-p)^{n-1}=p\sum_{n=0}^{\infty}\left[((1-F_x(z))(1-p))^n-(1-p)^{-1}\right].$$

The general cdf of an underlying continuous distribution with a random sample which follows truncated geometric distribution is

$$F(z)=1-\frac{p}{1-p}\left[\left(\frac{-1}{pF_x(z)-p-F_x(z)}\right)-1\right]$$

and a pdf of

$$f(z)=\frac{p}{f_x(z)(pF_x(z)-p-F_x(z))^2}\quad 0\leq p\leq 1.$$

Where $F_x(z)$ and $f_x(z)$ are the cdf and pdf of the continuous random variable X_i respectively.

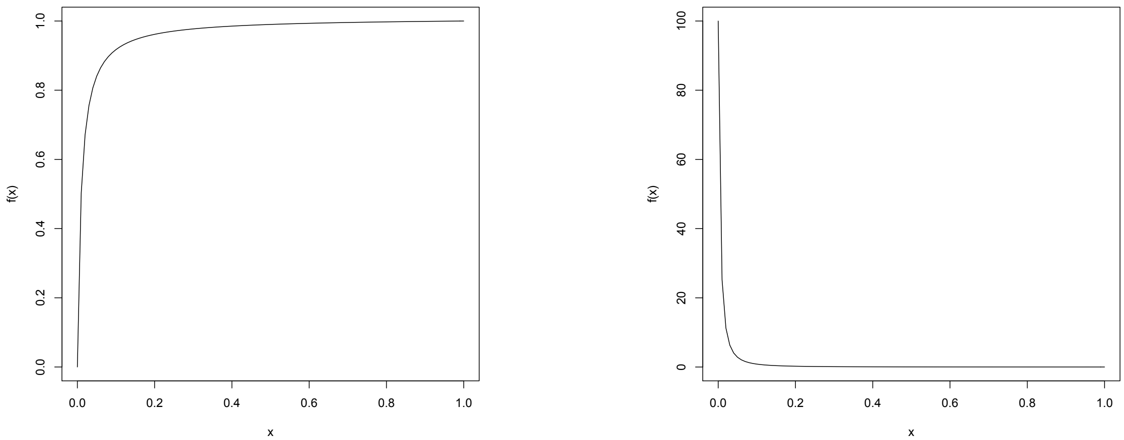
5.1 Uniform-Truncated Geometric Mixture

The distribution of a first order statistic with an underlying uniform distribution if $Z|N=min\{X_1,....,X_N\}$ where N is the random sample size which follows a truncated geometric distribution

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

This implies

$$P(Z>z)=\sum_{n=1}^{\infty}(1-z))^n p(1-p)^{n-1}=p\sum_{n=0}^{\infty}\left[((1-z)(1-p))^n-(1-p)^{-1}\right]$$



(a) Cumulative distribution function plot

(b) Probability density function plot

Figure 8: A Figure representing the cdf and pdf with an underlying uniform distribution and a random sample size which follows a truncated geometric distribution when p=0.01.

$$=\frac{p}{1-p}\left[\left(\frac{-1}{zp-p-z}\right)-1\right].$$

The cdf of a first order statistic with an underlying uniform distribution and a random sample size which follows a truncated geometric distribution is

$$F(z)=1-\frac{p}{1-p}\left[\left(\frac{-1}{zp-p-z}\right)-1\right], \quad 0\leq p\leq 1, \quad 0<z<1,$$

with a pdf of

$$f(z)=\frac{p}{(pz-p-z)^2}, \quad 0\leq p\leq 1, \quad 0<z<1.$$

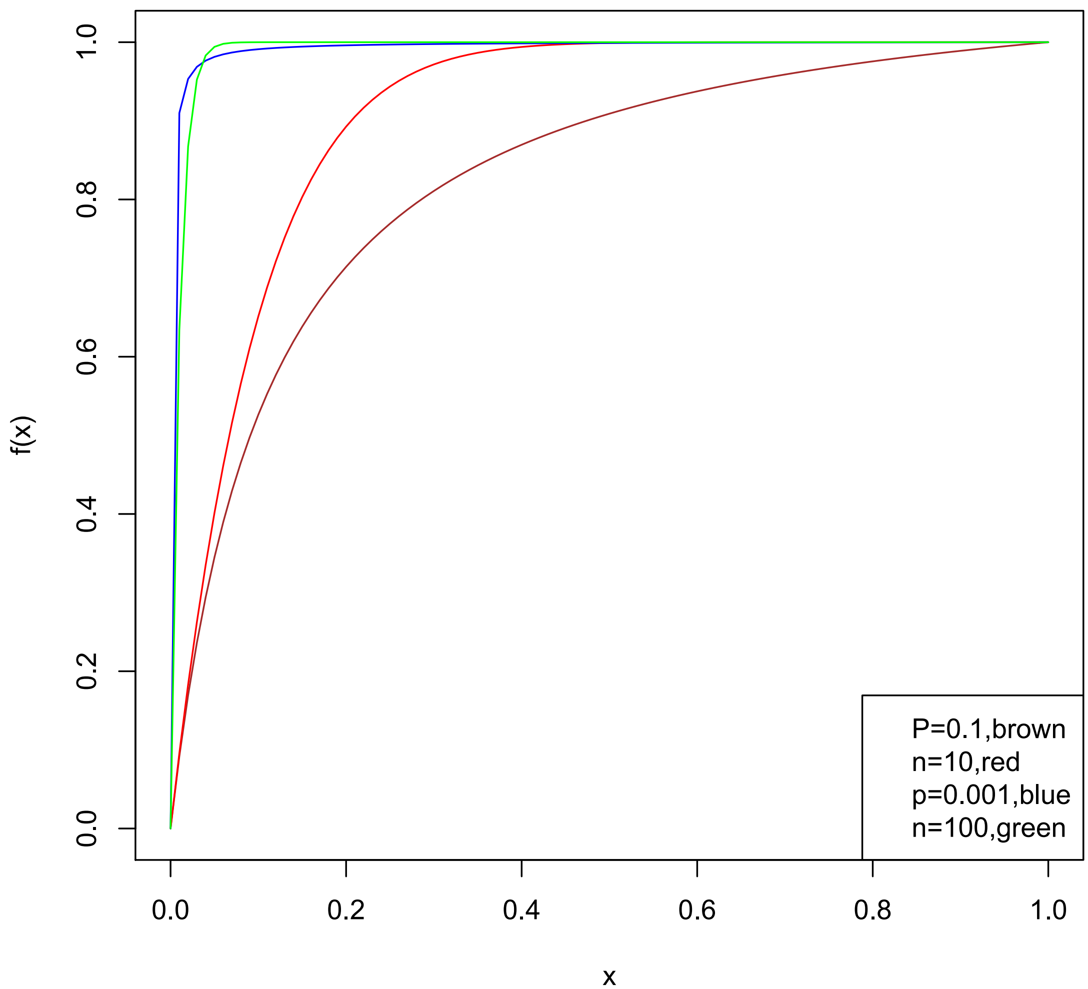


Figure 9: A cumulative distribution function plots of different samples (n) and different probabilities p when the underlying distribution is uniform for fixed and a random sample which follows a truncated geometric .

From Figure 9, it is clear that the cdf in both the random and fixed cases tend to take the similar shape as p decreases and n increases. The behaviour of the cdf as p

decreases and n increases shows a steep and sharp turn closer to 1. This implies that the smaller p gets and the larger n gets, both fixed and random sample size tend to have the same cdf.

5.2 Exponential-Truncated Geometric Mixture

The distribution of a first order statistic with an underlying exponential distribution if $(Z|N){=}min\{X_1, ..., X_N\}$ where N is the random sample size which follows a truncated geometric distribution

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

$$P(Z{>}z){=}\sum_{n=1}^{\infty}(e^{-z\mu}))^n p(1-p)^{n-1}{=}p\sum_{n=0}^{\infty}\left[((e^{-z\mu}(1-p))^n-(1-p)^{n-1}\right].$$

This implies

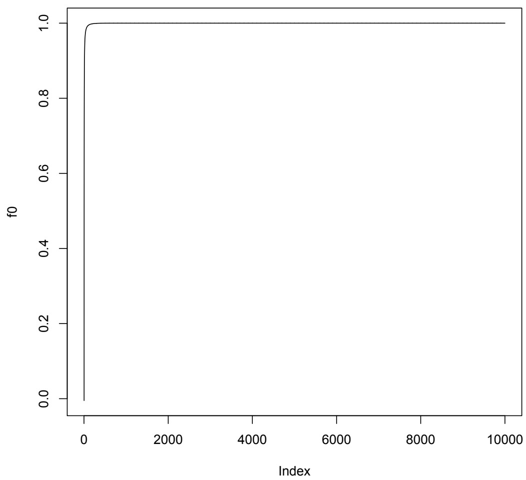
$$P(Z{>}z){=}\frac{pe^{-z\mu}}{(1-(e^{-z\mu}))(1-p)}.$$

Hence, the cdf of a first order statistic with an underlying exponential distribution and a random sample size which follows a truncated geometric distribution is

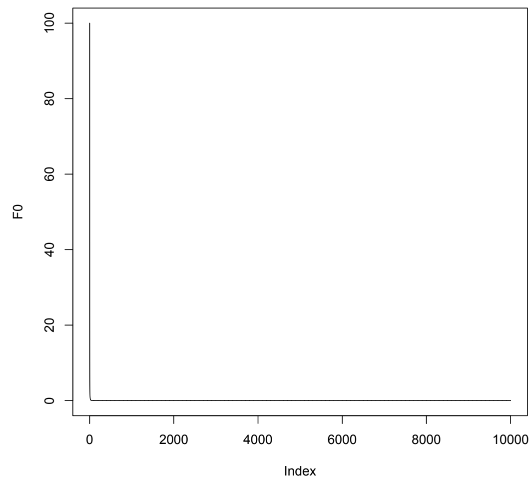
$$F(z){=}1-\frac{pe^{-z\mu}}{(1-(e^{-z\mu}))(1-p)}, \quad rate(\mu){>}0, \quad 0{\le}p{\le}1, \quad z{>}0$$

with a pdf of

$$f(z){=}\frac{p\mu e^{z\mu}}{(e^{z\mu}+p-1)^2}, \quad rate(\mu){>}0, \quad 0{\le}p{\le}1, \quad z{>}0.$$



(a) Cumulative distribution function plot



(b) Probability distribution function plot

Figure 10: The cdf and pdf of the smallest order statistic when the underlying distribution is exponential and random sample size which follows a truncated geometric with p=0.01 and μ=1.

6 CONCLUSION

The paper has focused on finding probability distributions of the first-order statistic when the sample size is random. The pivot of these joint distributions is a merge between the marginal and conditional probability distribution. In some instances, some properties that include the expectation, variance and percentile are calculated. The primary objective of this paper is to consider a random sample size and compare its behaviour to a fixed sample size in terms of their cumulative distribution functions(cdf). A comparison between the cdf when the sample size is fixed and random sample size is shown in figures 3, 6 and 9. It is clear at the end of the comparison in figure 3 and 6 that, as the sample size(n) increases in the fixed case, the cdf approaches one and gets more steep. We see from figures 3 and 6 that as the sample size increases and λ, p and k increases the cdf in both the fixed and random case appear the same. In figure 9, as n increases and p decreases, both cdfs in the fixed and random case take similar shape and becomes more steep and turns sharply close to one.

REFERENCES

[1] *Statistical inference* 2nd Edition by George Cassela,Roger. L. Berger.

[2] *Introduction to Mathematical Statistics*, 7th edition, by Hogg, Robert V. Mckean, and Allen T. Published by Pearson, 2013.

[3] *The Poisson Distribution*, by Jonathan Marchini, Nov 2008.

[4] *Finite sample theory of order statistics and extremes*, by Anirban DasGupta, May 2011.

VITA

VINCENT FORGO

Education:	B.S. Mathematics and Statistics, University of Capecoast, Capecoast, Ghana 2009
	M.S. Applied Mathematics, Indiana University of Pennsylvania, Indiana, Pennsylvania 2014
	M.S. Mathematical Science, East Tennessee State University 2017
Professional Experience:	Mathematics Teacher, S D A Senior High School, Bekwai, Ghana, 2009–2011
	Graduate Assistant, Indiana University of Pennsylvania, Indiana, Pennsylvania, 2012–2013
	Graduate Assistant, East Tennessee State University, Johnson City, Tennessee, 2014–2016
Projects:	Vincent Forgo, "A Distribution of the First Order Statistic when the Sample Size is Random"