Scientific
Research

Distribution of Geometrically Weighted Sum of Bernoulli Random Variables

Deepesh Bhati[1], Phazamile Kgosi[2], Ranganath Narayanacharya Rattihalli[1]
[1]*Department of Statistics, Central University of Rajasthan, Kishangarh, India*
[2]*Department of Statistics, University of Botswana, Gaborone, Botswana*
E-mail: *dipesh089@gmail.com, KGOSIPM@mopipi.ub.bw, rnr5@rediffmail.com*

Abstract

A new class of distributions over (0,1) is obtained by considering geometrically weighted sum of independent identically distributed (i.i.d.) Bernoulli random variables. An expression for the distribution function (d.f.) is derived and some properties are established. This class of distributions includes $U(0,1)$ distribution.

1. Introduction

Uniform distribution plays an important role in Statistics. The existence of uniform random variable (r.v.) over the interval (0,1), using $B(1,1/2)$ r.v.s is indicated in [1]. As a generalization, in this paper we consider the following geometrically weighted sum of i.i.d. Bernoulli r.v.s

$$X = \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^j Z_j, \qquad (1)$$

where $Z_j's$ are i.i.d. $B(1,p)$ r.v's. The remainder of the paper is organized as follows. In Section 2 we obtain the characteristic function of X and give an interpretation for the variable X. In Section 3 we derive the distribution function of X and prove some of its properties. In Section 4 we discuss the existence of the density function. In Section 5 distribution of sum of a finite number of variables is considered and the graphs of its probability mass function (p.m.f.) and distribution function (d.f.) are given in the Appendix.

2. The Characteristic Function and an Application of the Model

2.1. The Characteristic Function

Let $F(t) = P\{X \le t\}$ be the d.f. of *X*. Then by the definition of *X* we have

$$F(t) = P(X \le t | Z_1 = 0) P(Z_1 = 0) + P(X \le t | Z_1 = 1) P(Z_1 = 1)$$
$$= P\left(\frac{X}{2} \le t\right)(1-p) + P\left(\frac{1}{2} + \frac{X}{2} \le t\right)p. \qquad (2)$$

Hence the characteristic function (c.f.) $\Psi(t)$ of *X* satisfies the equation

$$\Psi(t) = (1-p)\Psi(t/2) + pe^{it/2}\Psi(t/2).$$

That is,

$$\Psi(t) = \Psi(t/2)\left[1 - p + pe^{it/2}\right].$$

Repeating this and replacing *t* by $t/2$ each time, we get, for $n = 1, 2, \cdots$.

$$\Psi(t) = \Psi(t/2^n)\prod_{k=1}^{n}\left[1 - p + pe^{it/2^k}\right]. \qquad (3)$$

The reproductive property of the characteristic function exhibited by (3) is comparable to the characteristic function of an infinitely divisible distribution. For details one may refer to Section 7 of [2].

Since $\Psi(t/2^n) \to 1$ as $n \to \infty$ we have

$$\Psi(t) = \prod_{k=1}^{\infty}\left[1 - p + pe^{it/2^k}\right]. \qquad (4)$$

Note that if *p* = 0, this infinite product is 1, and, if *p* = 1, the infinite product is e^{it}. Thus *p* = 0 results in *X* being degenerate at 0 while *p* = 1 implies that *X* is degenerate at 1. If *p* = 1/2, then the product term in (4) is

$$\frac{1}{2^n}\frac{e^{it}-1}{e^{it/2^n}-1} \to \frac{e^{it}-1}{it} \quad \text{as} \quad n \to \infty$$

Thus if *p* = 1/2 then *X* has $U(0,1)$ distribution.

2.2. An Application

This resulting distribution can be used as a model in a

situation similar to the following. Suppose a particle has linear movement on the interval [0,1]. To capture the particle suppose the following binary capturing technique of dividing the existing interval into two equal halves is used. Suppose initially there are two barriers put at 0 and 1. After one unit of time a barrier is put at the midpoint of 0 and 1. Further the interval in which the particle is found is divided into two equal halves by placing a barrier at their midpoint and the process is continued. The intervals containing the particle keep on shrinking and finally shrink to *X*, the point at which the particle is captured in the long run. The behavior of the particle is known only to the extent that at the moment of placing a barrier after exactly one unit of time the particle is on the right side of the inserted barrier with probability *p*.

3. Main Results

3.1. Notation

It is known that every number t, $0 \leq t \leq 1$ has a binary representation through $\{a_n\}_1^\infty, a_n = 0$ or 1, as

$t = \sum_{i=1}^\infty a_i \left(\frac{1}{2}\right)^i$. If a number t has the representation

$t = \sum_{i=1}^k a_i \left(\frac{1}{2}\right)^i$, $a_k = 1$, we refer t as a finite binary terminating number and such a number can be represented by $r2^{-k}$, for some $r = 1, 2, \cdots, 2^k - 1$. However such a number also can be represented as

$$\sum_{i=1}^\infty b_i \left(\frac{1}{2}\right)^i, b_i = a_i, i = 1, 2, \cdots, k-1,$$
$$b_k = 0, b_i = 1, i = k+1, k+2, \cdots$$

It is to be noted that the right tail of the sequence $\{a_i\}$ is of the form $(0,0,0\cdots)$ while that of the sequence $\{b_i\}$ is of the form $(1,1,1,\cdots)$. In the following as a matter of convention we do not consider representation with the right tail of the form $(1,1,1,\cdots)$. Under such a convention, $t \in [0,1]$ corresponds to a unique binary sequence $\{a_i\}_1^\infty$ and conversely. If $t = \sum_{i=1}^\infty a_i \left(\frac{1}{2}\right)^i$, then we shall

denote this relation as $t \overset{BR}{\Leftrightarrow} \{a_n\}$, (*BR* to mean the binary representation).

3.2. Properties

Theorem 1:

Let $t \overset{BR}{\Leftrightarrow} \{a_n\}$ and $s_k = \sum_{j=1}^k a_j$, $k = 1, 2, \cdots, s_0 = 0$. Then

the d.f. of *X* defined in (1) is given by

$$F(t) = P\{X \leq t\} = \sum_{j=1}^\infty a_j \left\{q^{j-s_{j-1}} p^{s_{j-1}}\right\}. \quad (5)$$

Proof:

Let, $t = \sum_{i=1}^\infty a_i \left(\frac{1}{2}\right)^i$ and a_{k_r} be the rth non-zero element

in the sequence $\{a_i\}, r = 1, 2, \cdots$. Let t_r be the number having the binary representation $(a_1, a_2, \cdots a_{k_r}, 0, 0, \cdots)$ and $t_0 = 0$. It is to be noted that t_r is a finite binary terminating number. If t is not a finite binary terminating number then the sequence $\{t_r\}$ increases to t. If t is a finite binary terminating number then we note that $t = t_r$ for some finite *r*. For example, let t have a binary representation $00010011000\cdots$ and

$k_1 = 4, k_2 = 7, k_3 = 8, k_4 = 11$ so that $t_1 \overset{BR}{\Leftrightarrow} (0010000\cdots)$,

$t_2 \overset{BR}{\Leftrightarrow} (0001001000\cdots)$, $t_3 \overset{BR}{\Leftrightarrow} (00010011000\cdots)$ and

so on. We note that $a_i = 1$ for $i = k_1, k_2, k_3, \cdots$ and 0 otherwise. Note that

$P\{X = t_r\} =$

$P\{Z_i = a_i, i = 1, 2, \cdots, k_r - 1, Z_{k_r} = 1, Z_{k_r+j} = 0, j = 1, 2, \ldots\}$

$\leq P\{Z_{k_r} = 1, Z_{k_r+j} = 0, j = 1, 2, \cdots\} = 0.$

Thus *F* does not have a jump at t_r and $F(t_r) \uparrow F(t)$.

Let t be not a finite binary representing number. We note that

$$P(0 \leq X < t_1) = P(Z_1 = 0, \cdots, Z_{k_1} = 0) = q^{k_1}$$

$$P(t_1 \leq X < t_2) = P(Z_1 = 0, \cdots, Z_{k_1-1} = 0, Z_{k_1} = 1, Z_{k_1+1} = 0,$$
$$\cdots, Z_{k_2} = 0) = q^{k_2-1}p$$

$$P(t_2 \leq X < t_3) = P(Z_1 = 0, \cdots, Z_{k_1-1} = 0, Z_{k_1} = 1, Z_{k_1+1} = 0, \cdots$$
$$\cdots, Z_{k_2-1} = 0, Z_{k_2} = 1, Z_{k_2+1} = 0, \cdots, Z_{k_3} = 0) = q^{k_3-2}p^2.$$

Thus in general for $r = 1, 2, 3 \cdots$, we have

$$P(t_{r-1} \leq X < t_r) = q^{k_r-(r-1)}p^{r-1}.$$

Hence

$$P(X < t) = \sum_{r=1}^\infty P(t_{r-1} \leq X < t_r) = \sum_{r=1}^\infty q^{k_r-(r-1)}p^{r-1}.$$

If we let $k_r = j$, then $r - 1 = \sum_{i=1}^{j-1} a_i = s_{j-1}$, then we

will have $P(X < t) = \sum_{j=1}^\infty a_j \left(q^{j-s_{j-1}} p^{s_{j-1}}\right).$

In fact *F* does not have jump at ant *t*, (0 < *t* < 1).

Hence $P(X \leq t) = P(X < t) = \sum_{j=1}^\infty a_j \left(q^{j-s_{j-1}} p^{s_{j-1}}\right).$

However, if t is a finite binary termination number then $t = t_r$ for some finite number r and

$$P(X \leq t) = P(X < t) + P(X = t_r)$$

$$= P(X < t_r) = \sum_{j=1}^{r} a_j \left(q^{j - s_{j-1}} p^{s_{j-1}} \right) = \sum_{j=1}^{\infty} a_j \left(q^{j - s_{j-1}} p^{s_{j-1}} \right)$$

since $a_j = 0, j = r+1, r+2, \cdots$.

3.3. Particular Case

If $p = \dfrac{1}{2}$, then $P(X < t) = \sum_{j=1}^{\infty} a_j \left(\dfrac{1}{2} \right)^j = t, t \in (0,1)$.

Hence $X \sim U(0,1)$.

Remarks:

In Fact it can be verified that (5) satisfies (2). It follows from the fact that if t has binary representation $a = (a_1, a_2, \cdots)$ then for

1) $0 < t < 1$, $t/2$ has binary representation $u = (u_1, u_2, \cdots)$ where $u_1 = 0$ and $u_{i+1} = a_i, i = 1, 2, \cdots$ and

2) $1/2 < t < 1$, $2t - 1$ has binary representation $v = (v_1, v_2, \cdots)$ where $v_i = a_{i+1}, i = 1, 2, \cdots$.

Theorem 2:

If u and v have binary representations $(a_1, a_2, \cdots, a_n, 0, 0, \cdots, 0)$ and $(a_1, a_2, \cdots, a_n, 1, 1, \cdots, 1)$ respectively then the conditional distribution of X given $u \leq X \leq v$ is that of $u + 2^{-n} X$.

Proof: Follows by the definition of X.

Theorem 3:

Let $t \overset{BR}{\Leftrightarrow} \{a_n\}$ with $(a_1, a_2, \cdots, a_k, 0, 0, \cdots)$ with $a_k = 1$.

Then for $0 \leq s \leq \dfrac{1}{2^k}$

$$P\left\{ 0 < X - t < s \mid 0 < X - t < \frac{1}{2^k} \right\} = P\left\{ 0 < X < 2^k s \right\}.$$

Proof:

$$P\left\{ 0 < X - t < s \mid 0 < X - t < 2^{-k} \right\}$$

$$= P\left\{ 0 < X < t + s \mid 0 < t < t + 2^{-k} \right\} = \frac{P\left\{ 0 < X < t + s \right\}}{P\left\{ 0 < X < t + 2^{-k} \right\}}$$

$$= \frac{P\left\{ Z_1 = a_1, Z_2 = a_2, \cdots, Z_k = a_k, \sum_{j=k+1}^{\infty} \left(\frac{1}{2} \right)^j Z_j \leq s \right\}}{P\left\{ Z_1 = a_1, Z_2 = a_2, \cdots, Z_k = a_k \right\}}$$

$$= \frac{P\left\{ Z_1 = a_1, Z_2 = a_2, \cdots, Z_k = a_k \right\} P\left\{ \left(\frac{1}{2} \right)^k \sum_{j=0}^{\infty} \left(\frac{1}{2} \right)^j Z_j \leq s \right\}}{P\left\{ Z_1 = a_1, Z_2 = a_2, \cdots, Z_k = a_k \right\}}$$

$$= P\left\{ \left(\frac{1}{2^k} \right) X \leq s \right\} = P\left\{ X \leq 2^k s \right\}.$$

Theorem 4:

For $0 \leq t < s \leq \dfrac{1}{2}$ we have

$$P\{t < X < s\} = \frac{q}{p} P\left\{ t + \frac{1}{2} < X < s + \frac{1}{2} \right\}.$$

Proof:

Let $Z = \left(Z_1, Z^{(2)} \right) = \left(Z_1, Z_2, Z_3, \cdots \right)$

$$P\{t < X < s\} = P\left\{ Z_1 = 0, Z^{(2)} \in A \right\}$$

$$= P\left\{ Z_1 = 0 \right\} P\left\{ Z^{(2)} \in A \right\} = q P\left\{ Z^{(2)} \in A \right\}$$

and

$$P\left\{ t + \frac{1}{2} < X < s + \frac{1}{2} \right\} = P\left\{ Z_1 = 1, Z^{(2)} \in A \right\}$$

$$P\left\{ Z_1 = 1 \right\} P\left\{ Z^{(2)} \in A \right\} = p P\left\{ Z^{(2)} \in A \right\}.$$

In the above in fact $P\left\{ Z^{(2)} \in A \right\} = P\left\{ \dfrac{X}{2} < s - t \right\}$.

Hence the result.

Theorem 5:

If $F(., p)$ and $G(., p)$ are the distribution functions of X and $1-X$ respectively then $G(x, p) = F(x, 1-p)$.

Proof:

If $X = \sum_{j=1}^{\infty} \left(\dfrac{1}{2} \right)^j Z_j$ and Z_j's are i.i.d. $B(1, p)$ then

$1 - X = \sum_{j=1}^{\infty} \left(\dfrac{1}{2} \right)^j \left(1 - Z_j \right) = \sum_{j=1}^{\infty} \left(\dfrac{1}{2} \right)^j Y_j$ where Y_j's are i.i.d. $B(1, 1-p)$.

Hence the result.

Mean and Variance of X:

$$E(X) = \sum_{j=1}^{\infty} \left(\frac{1}{2} \right)^j E(Z_j) = \sum_{j=1}^{\infty} \left(\frac{1}{2} \right)^j p = p$$

$$E(X^2) = \sum_{j=1}^{\infty} \left(\frac{1}{2} \right)^{2j} E(Z_j) + 2 \sum_{i=1}^{\infty} \sum_{j=1}^{i-1} \left(\frac{1}{2} \right)^{i+j} E(Z_i Z_j)$$

$$= \sum_{j=1}^{\infty} \left(\frac{1}{2} \right)^{2j} p + 2 p^2 \sum_{i=1}^{\infty} \left(\frac{1}{2} \right)^i \sum_{j=1}^{i-1} \left(\frac{1}{2} \right)^j$$

$$= \frac{p}{3} + 2 p^2 \sum_{i=1}^{\infty} \left(\frac{1}{2} \right)^i \left\{ 1 - \left(\frac{1}{2} \right)^{i-1} \right\}$$

$$= \frac{p}{3} + 2 p^2 \left(\frac{1}{3} \right) = \frac{p(1 + 2p)}{3}$$

Hence $Var(X) = \dfrac{p(1 - p)}{3}$.

By using the c.f. $\Psi(t) = \prod_{k=1}^{\infty} \left[1 - p + p e^{it/2^k} \right]$ the cumulants can also be obtained.

4. Nonexistence of Density Function

We have proved that the distribution function of X is given by $F(t) = \sum_{j=1}^{\infty} a_j \left\{ q^{j-s_{j-1}} p^{s_{j-1}} \right\}$, $t \overset{BR}{\Leftrightarrow} \{a_n\}$. Let the left derivative and the right derivative of F at t exist. These be denoted by $f(t+)$ and $f(t-)$ respectively. Consider $f\left(\frac{1}{2}+\right)$ and $f\left(\frac{1}{2}-\right)$, the right and left derivatives of F at 1/2.

$$f\left(\frac{1}{2}-\right) = \lim_{k\to\infty} \frac{F\left(\frac{1}{2}\right) - F\left(\frac{1}{2} - \frac{1}{2^k}\right)}{2^{-k}}$$

$$= \lim_{k\to\infty} \frac{qp^{k-1}}{2^{-k}} = \lim_{k\to\infty}(2q)(2p)^{k-1}$$

and

$$f\left(\frac{1}{2}+\right) = \lim_{k\to\infty} \frac{F\left(\frac{1}{2} + \frac{1}{2^k}\right) - F\left(\frac{1}{2}\right)}{2^{-k}}$$

$$= \lim_{k\to\infty} \frac{pq^{k-1}}{2^{-k}} = \lim_{k\to\infty}(2p)(2q)^{k-1}.$$

Note that $f\left(\frac{1}{2}-\right)$ and $f\left(\frac{1}{2}+\right)$ are equal if and only if $p = q = \frac{1}{2}$. Hence F is not differentiable at 1/2 if $p \neq \frac{1}{2}$.

Let $D = \left\{ \frac{r}{2^k} : k = 1,2,3\cdots, r = 0,1,2,\cdots,2^k \right\}$. It can be shown that F is not differentiable at each point of D and the set D is countable dense subset of $[0,1]$. Hence F is nowhere differentiable in $[0,1]$.

5. Distribution of Sum of a Finite Number of Bernoulli Random Variables

Since $F(t)$ is an infinite series for t not in D, the exact evaluation of $F(t)$ is not possible for each $t \in (0,1)$. More-over, the density function of X does not exist on the interval $(0,1)$. Hence in the following we consider the sequence $\{X_k\}$ of r.v. defined by

$$X_k = \sum_{i=1}^{k} \left(\frac{1}{2}\right)^i Z_i. \tag{6}$$

The sequence $\{X_k\}$ increases point wise to X and $X - X_k < 2^{-k}$.

Similar to (5), it can be shown that for $t = \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^j a_j$,

the d.f. of X_k is $F_k(t) = \sum_{j=1}^{k} a_j \left\{ q^{j-s_{j-1}} p^{s_{j-1}} \right\} = F(s)$,

where $s = \sum_{j=1}^{k} a_j \left(\frac{1}{2}\right)^j$. Further, at these values of s,

$F_k(s) = F(s)$ and the difference between two successive values of s's is 2^{-k}, as such the two functions $F_k(t)$ and $F(t)$ are almost alike. Since the sequence $\{X_k\}$ increases point wise to X the sequence $\{F_k(t)\}$ decreases to $F(t)$ for $t \in (0,1)$.

We note that for $p = 1/2$, $F_k(t) = F(s)$. The graphs of p.m.f and d.f. of X_5 and X_{10} for different values of $p \neq \frac{1}{2}$ are given in the Appendix.
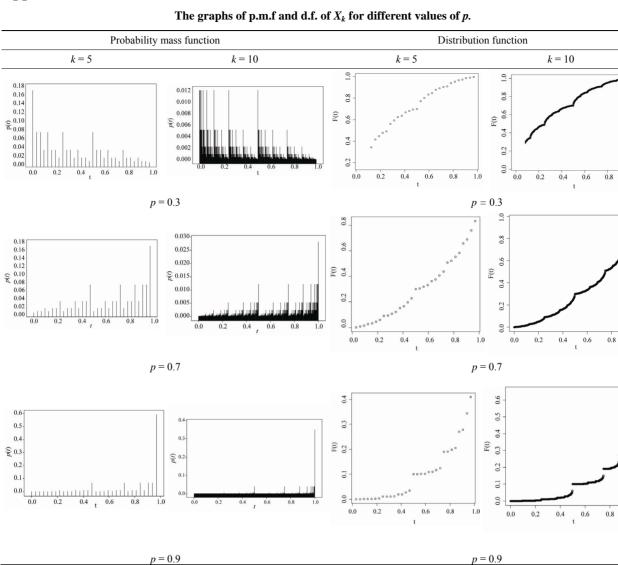
6. Acknowledgements

7. References

[1] S. Kunte and R. N. Rattihalli, "Uniform Random Variable. Do They Exist in Subjective Sense?" *Calcutta Statistical Association Bulletin*, Vol. 42, 1992, pp. 124-128.

[2] K. L. Chung, "A Course in Probability Theory," 3rd Edition, Academic Press, Cambridge, 2001.

Appendix

The graphs of p.m.f and d.f. of X_k for different values of *p*.

Probability mass function		Distribution function	
k = 5	*k* = 10	*k* = 5	*k* = 10

p = 0.3   *p* = 0.3

p = 0.7   *p* = 0.7

p = 0.9   *p* = 0.9

AM