YOKOHAMA MATHEMATICAL
JOURNAL VOL. 36, 1938

SIZE DISTRIBTTIONS OF FRACTIONS GENERATED BY RANDOM SUBDIVISION OF THE UNIT INTERVAL

By

YASUSHI TAGA

(Received September, 30, 1988; Revised October 20, 1988)

Summary Size distributions of fractions generated by random subdivision of the unit interval and their approximation method using gamma distributions are studied.

1. Introduction

Pyke [1] showed that the normalized empirical distribution function of the $n+1$ spacings, random subdivision of the unit interval (0, 1) generated by the first n variables of a sequence of random variables $\{x_m; m\geqq 1\}$ under the Kakutani's model, converges uniformly to the uniform distribution over the interval (0, 2)—this result is partially dependent on the idea of van Zwet [3] on the same problem. On the other hand, Blum showed in 1955 that the normalized empirical distribution of the $n+1$ spacings, the random subdivision of the unit interval by n independent uniform random variables x_i's over (0, 1), converges in probability to the exponential distribution $H(x)=1-e^{-x}(x>0)$.

In this paper size distributions of fractions, generated by the random sequential bisection, and their tail probabilities are studied, where "random sequential bisection" means random sequential subdivision of the unit interval (0, 1) into $2, 2^2, \cdots, 2^m$ $(m=1, 2, \cdots)$ fractions based on some p. d. f. $1/l\ f(x/l)$ over each subinterval of length l.

In the so-called "fractal theory", the upper tail probabilities of size distributions of fractions having length larger than x have been observed in various real situations as $O(x^{-\alpha})$ $(\alpha>0)$. We are also interested in evaluating the upper tail probabilities and their approximation methods, if the above observation or conjecture on the upper tail probabilities could be found through our method.

Let $X_1, X_2, \cdots, X_n$ be mutually independent and identically distributed according to a probability density function $f(x)$ symmetric over the interval (0, 1), i. e. satisfying $f(1-x)=f(x)$, and define Z by

$$Z=X_1X_2\cdots X_n. \tag{1.1}$$

Then Z may be regarded as a random variable representing the length of one of fractional subintervals generated by the n-times bisection procedures of the initial interval (0, 1). Since X_k and $1-X_k$ are identically distributed, every subinterval among 2^n subintervals has the same distribution as Z.

Taking logarithm of Z

$$\log Z=\sum_{i=1}^{n} \log X_i\,, \tag{1.2}$$

and we can represent the distribution function $G(y)$ of Z such that

$$G(y)=P\{Z\leqq y\}=P\left\{-\sum_{i=1}^{n} \log X_i \geqq -\log y\right\}. \tag{1.3}$$

Now the moment generating function $M(t)$ of $(-\log Z)$ is expressed as

$$\begin{aligned} M(t)&=E\{\exp(-t\log Z)\} \\ &=E\{(X_1X_2\cdots X_n)^{-t}\}=\prod_{i=1}^{n} E\{X_i^{-t}\}=(E\{X_1^{-t}\})^n. \end{aligned} \tag{1.4}$$

Each term $E\{X_i^{-t}\}$ in (1.4) is the moment generating function of $(-\log X_i)$ for $i=1, 2, \cdots, n$.

For a while let us cosider the case where each X_i follows the beta distribution (not necessarily symmetric), with parameters p and q. Then we can get $E\{X_i^{-t}\}$ in the following way.

$$\begin{aligned} E\{X_i^{-t}\}&=\int_0^1 \frac{1}{B(p,\, q)} x^{p-t-1}(1-x)^{q-1}dx \\ &=B(p-t,\, q)/B(p,\, q)\,. \end{aligned} \tag{1.5}$$

If $p\geqq 1$ and q is a positive integer, it is easily seen that $E\{X_i^{-t}\}$ is expressed as

$$E\{X_i^{-t}\}=\prod_{k=1}^{q}\left(1-\frac{1}{p+q-k}t\right)^{-1}, \tag{1.6}$$

which is the product of q moment generating functions of the exponential distribution with parameters $(p+q-k)^{-1}$, i.e. the distribution with p.d.f.

$$\frac{1}{p+q-k}e^{-x/(p+q-k)},\quad x>0,\ (k=1, 2, \cdots, q)\,. \tag{1.7}$$

Substituting (1.6) into (1.4) we can get $M(t)$ as

$$M(t)=\prod_{k=1}^{q}\left(1-\frac{1}{p+q-k}t\right)^{-n}. \tag{1.8}$$

From (1.8) $M(t)$ is regarded as the product of q moment generating functions of gamma distributions with parameters n and $(p+q-k)^{-1}$, i.e. the distribution with p.d.f.

(1.9) $$\frac{1}{(p+q-k)^n \Gamma(n)} u^{n-1} e^{-u/(p+q-k)}, \qquad (u>0),$$

for $k=1, 2, \cdots, q$.

Accordingly the distribution of $(-\log Z)$ is regrded as the q-th convolution of gamma distributions given in (1.9).

2. Main Results

First let us consider the case where $f(x)$ is uniform over the interval (0, 1). In this case, from (1.8), the moment generating function $M(t)$ of $(-\log Z)$ is given by

(2.1) $$M(t)=(1-t)^{-n},$$

and then $(-\log Z)$ has the gamma density

(2.2) $$\frac{1}{\Gamma(n)} u^{n-1} e^{-u},$$

Therefore the upper tail probability of Z is obtained in the following way.

(2.3) $$\begin{aligned} P\{Z>y\} &= P\{-\log Z<-\log y\} \\ &= \int_0^{-\log y} \frac{1}{\Gamma(n)} t^{n-1} e^{-t} dt. \end{aligned}$$

Therefore, if y is sufficiently near to 1, the above tail probability is evaluated such that

(2.4) $$P\{Z>y\}=\int_0^{-\log y} \frac{1}{(n-1)!} t^{n-1} e^{-t} dt = \sum_{k=n}^{\infty} \frac{(-\log y)^k}{k!} y \doteqdot \frac{(-\log y)^n}{n!}.$$

In the similar way we can get tail probability for $y=1$ in the case where the initial interval is of length $x(>1)$:

(2.5) $$\begin{aligned} P\{xZ>1\} &= P\{-\log Z<\log x\} \\ &= \int_0^{\log x} \frac{1}{\Gamma(n)} t^{n-1} e^{-t} dt \\ &= \sum_{k=n}^{\infty} \frac{(\log X)^k}{k!} \frac{1}{x}. \end{aligned}$$

This result is identical to the proportion of the expected number of "internal nodes" obtained by Sibuya and Itoh [1]. Besides the orders of upper tail probabilities given in (2.4) and (2.5) is somewhat different from those as conjectured in "fractal theory" as stated in Introduction.

Let $f(x)$ be

$$f(x)=\frac{1}{B(p,q)}x^{p-1}(1-x)^{q-1}, \quad (p, q\geqq 1) \tag{2.6}$$

then it can be shown after some calculations that

$$E\{-\log X_i\}=\sum_{k=1}^{q}\frac{1}{p+q-k} \tag{2.7}$$

and

$$V[(-\log X_i)]=\sum_{k=1}^{q}\left(\frac{1}{p+q-k}\right)^2. \tag{2.8}$$

From (2.7) and (2.8) the expectation and variance of $(-\log Z)$ is expressed as

$$E\{-\log Z\}=nE\{-\log X_i\} \tag{2.9}$$

and

$$V(-\log Z)=n\,V(-\log X_i). \tag{2.10}$$

Using (2.8) and (2.9) we could get the approximate distribution of $(-\log Z)$ by a suitable gamma distribution through the moment method. We shall show such an approximation in the case where $p=q=2$.

From (2.9) and (2.10) we can get easily

$$E\{-\log Z\}=\frac{5}{6}n \tag{2.11}$$

and

$$V(-\log Z)=\frac{13}{36}n. \tag{2.12}$$

By the moment method, parameters θ and ν of a suitable gamma distribution for approximating the distribution of $(-\log Z)$ can be obtained such that

$$\theta=V(-\log Z)/E\{-\log Z\}=\frac{13}{30}=0.4333$$

and

$$\nu=[E\{-\log Z\}]^2/V(-\log Z)=\frac{25}{13}n\doteqdot 1.9231\,n.$$

Hence we could get the approximate value of the upper tail probability for $p=q=2$, $n=2$ and $y=0.8$ such that

(2.13) $$P\{Z>y\}=P\left\{-\frac{2}{\theta}\log Z<-\frac{2}{\theta}\log y\right\}\doteqdot 0.0022\,.$$

On the other hand, the exact value of the upper tail probability above can be obtained by calculating the following double integral directly, i. e.

(2.14) $$\begin{aligned}P\{Z>y\}&=p\{X_1X_2>y\}\\&=\iint_{y<x_1x_2<1}36x_1(1-x_1)x_2(1-x_2)dx_1dx_2\\&=1+27y^2+18y^2\log y-28y^3+12y^3\log y\,.\end{aligned}$$

Putting $y=0.8$ we can get the exact tail probability $P\{Z>y\}=0.0024$, which is quite near to the approximate value of it given in (2.13).

3. Concluding Remarks

We shall study to extend the results stated in the section 2 as follows.

1) To remove the restriction of symmetry of the basic distribution $f(x)$, i. e. the beta distribution with parameters $p\neq q$.

2) To evaluate the upper tail probability as given in (2.14) by the Monte-Carlo method as correctly as desired.

3) To exammine the goodness of approximation method as shown in (2.11)$\sim$(2.14) using the gamma distribution in the general case.

References

[1] Pyke, R.: *The asymptotic behavior of spacings under Kakutani's model for interval subdivision,* Ann. Probab., **8** (1980), 157-163.

[2] Sibuya, M. and Itoh, Y.: *Random sequential bisection and its associated binary tree,* Ann. Inst. Statist. Math., **39**, Part **A**, 69-84.

[3] van Zwet, W.R.: *A proof of Kakutani's conjecture on random subdivision of longest intervals,* Ann. Probab., **6** (1978), 133-137.

Department of Mathematics
Yokohama City University
22-2, Seto, Kanazawa-ku
Yokohama 236, JAPAN