総合都市研究　第43号　1991

線分をランダムな点で区切った区間の長さの確率分布に関する分析

吉川　徹*
岡部篤行**

要　約

この研究の目的は，一定の長さの線分をある分布に従ってランダムに分布する点で分割してつくった区間の長さの統計学的な性質について解析することにある。この目的のため，まず，ひとつの区間の長さの確率密度関数の持っている性質について検討した。この結果，この確率密度関数はランダムな点の確率密度関数を含んだ積分式のかたちに書きあらわされることがわかった。つづいて，長さの異なる2本の線分をおのおの一様ランダムな点で分割してつくった区間から，それぞれ1個づつを取り出したときに，その長さの比に関する指標の確率密度関数を計算した。

1．はじめに

この研究の目的は，一定の長さの線分をある分布に従ったランダムな点で分割してつくった区間の長さの統計学的な性質を解析することにある。

都市や地域の数理的な分析を行なう場合，ある分布に従った点によって分割された区間の長さが問題になる場面がある。この例としては，鉄道駅やインターチェンジ，バス停や交差点の間隔について議論する場合があげられる。このような場面で，もしランダムな点で分割された区間の長さの統計学的性質がわかっていれば，これとの比較対照によって実際の区間の長さの分布の特徴を定量的に論じることができる。たとえば交差点の間隔については，交差点がランダムに配置されていたと仮定した場合の交差点間距離の性質と実際の交差点間距離を比較することによって，その都市の道路網の特徴を定量化することが可能になる。

そこでこの研究では，このランダムな点で分割された区間の長さがどのような統計学的性質を帯びているかを調べた。この論文では，この研究の結果について次の順序で報告する。まずはじめにランダムな点による分割のモデルを定式化する。つづいてこのモデルのもとでの区間の長さの確率密度関数を計算し，さらに点が一様ランダムである場合について実際に確率密度関数を計算する。最後に長さの異なる2本の線分をおのおの一様ラ

*東京都立大学都市研究センター・教養部
**東京大学工学部都市工学科

ンダムな点で分割してつくった区間から，それぞれ1個づつの区間を取り出したときに，その長さの比に関する指標の確率密度関数を計算する。

2．ランダムな点による分割のモデル

一定の長さの線分がランダムな点によって分割されるモデルとしては，大きく分けて2種類のものを想定することができる。

第1の種類は，きわめて短い単位長さ当りの点の生成確率が決まっている一方で，点の総数は前もっては決まっていないものである。これに対して第2は，点の総数があらかじめ与えられていて，これらの点が特定の確率密度関数に従って互いにランダムに線分上に分布する場合である。

実際の都市や地域の状況を分析する際には，この2種類の状況はどちらも発生しうる。たとえばバス停の配置を分析することを考えよう。もしバス会社が配置するバス停の総数をあらかじめ決定しているのではなく，沿道状況にしたがってバス停を置くか置かないかを判断しているのであれば，これと比較対照すべきランダムな状況は第1の種類の方である。これに対して，あらかじめバス会社が設置するバス停の総数を決めているのであれば比較すべきランダムな状況は第2の種類である。

第1の種類のモデルで，特に線分のあらゆる部分にわたって生成確率が一定である場合には，一定の長さ当りの点の数はポアソン分布に従い，その結果として分割された区間の長さは指数分布に従うことが知られている。これに対して第2の種類のランダムな分割はこれらの確率密度分布には従わない。そこで以下では，この第2の種類のランダムな状況について解析を行なうことにしよう。

3．第2の種類のランダムな状況での区間の長さの確率密度関数

第2の種類のランダムな状況では区間の長さはどのような確率密度関数に従うのであろうか。

互いにランダムに特定の確率密度分布に従う確率変数の最小値と最大値，および最大値と最小値の差（範囲）についてはその確率密度関数をもとの確率密度関数から計算する方法が知られている（ホーエル，1978：pp.269-273）。この結果自体はこの問題には適用できないがその計算の進め方に沿ってこの問題を計算することは可能である。

まず始めに問題を定式化しよう。なお，計算を簡単にするために線分全体の長さを長さの基準にとって1とする。こうしても計算は一般性を失わない。

［問題］長さ1の線分上に，互いにランダムに特定の確率密度分布に従う点が$n-1$個落とされた。この結果として生じるn個の区間の長さの確率密度関数を求めよ。

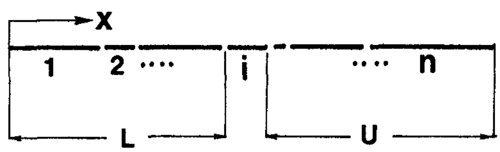

図1　長さ1の線分の分割

この問題を解くために，線分の左端を0として座標を設定し，左端から距離xの位置をxであらわすことにしよう（図1）。また区間の区切りになる点（以下では区切り点と呼ぶ）が従う確率密度関数を$b(x)$であらわす。さらに，左から数えてi番目の区間の左の区切り点から線分左端までの距離をL，右の区切り点から線分右端までの距離をUとしよう。ただし，もっとも左ともっとも右の区間については特別な取り扱いが必要になるので，当面はこれらを除いて（つまり$2\leq i\leq n-1$として）考えることにする。

このLとUの挙動が知られれば，i番目の区間の長さの挙動が判明する。そこで，$L\leq l_0$かつ$U\leq u_0$である確率を$F(l_0, u_0)$とおいて，この確率関数を調べることにしよう。この関数の定義より，

$$F(l_0,u_0)=\binom{n-1}{i-1}\left(\int_0^{l_0}b(x)dx\right)^{i-1}\left(\int_{1-u_0}^{1}b(x)dx\right)^{n-i} \quad (1)$$

となる。というのは，F (l_0, u_0) は，$n-1$ 個の区切り点のうち $i-1$ 個が l_0 より左側に生じて，残りの $n-i$ 個が u_0 より右側に生じる確率に等しいからである。

一方，L, U の確率密度関数を f (l, u) とおけば，その定義より，

$$F(l_0,u_0)=\int_0^{l_0}\int_0^{u_0}f(l,u)du\,dl \tag{2}$$

である。したがって(1)(2)をおのおの l_0 で微分してもその右辺どうしは等しい。

$$\binom{n-1}{i-1}(i-1)\left(\int_0^{l_0}b(x)dx\right)^{i-2} b(l_0)\left(\int_{1-u_0}^1 b(x)dx\right)^{n-i}=\int_0^{u_0}f(l_0,u)du \tag{3}$$

さらにこれを u_0 で微分すれば，$f(l_0,\ u_0)$ が $b(x)$ であらわされる。

$$f(l_0,u_0)=\binom{n-1}{i-1}(i-1)\left(\int_0^{l_0}b(x)dx\right)^{i-2}b(l_0)(n-i)\left(\int_{1-u_0}^1 b(x)dx\right)^{n-i-1}b(1-u_0) \tag{4}$$

この $f(l_0,\ u_0)$ を利用すると，i 番目の区間の長さの確率密度関数が計算できる。このために，まず i 番目の区間の長さを D とおくと，次式が成立する。

$$D=1-L-U \tag{5}$$

一方，D と L の同時確率密度関数を $h(l,\ d)$ とおく。すると，一般に，

$$h=\frac{f(l,u)}{\left|\frac{\partial d}{\partial u}\right|} \tag{6}$$

の関係が成り立つ（ホーエル，1978：pp.244-247）ことから，hとfの関係は，

$$h=f(l,\ 1-l-d) \tag{7}$$

となることがわかる。したがって，

$$h(l,d)=\binom{n-1}{i-1}(i-1)\left(\int_0^1 b(x)dx\right)^{i-2}b(l)(n-i)\left(\int_{d+l}^1 b(x)dx\right)^{n-i-1}b(d+l) \tag{8}$$

である。この式を制約条件として $0\leq l\leq 1-d$（こうしないと区間 i の長さが負になってしまう）を考慮しつつ l に関して積分すれば，d の確率密度関数 $g(d)$ が得られる。

なお，第1番目（もっとも左），あるいは第 n 番目（もっとも右）の区間の長さについては次のように計算すればよい。なお，両方の計算結果は同じになるので，ここでは第1番目のみについて計算過程を示そう。

ここで，第1番目の区間の右端点から区間の右端まで長さ U が u_0 以下である確率を $F(u_0)$ とおく。すると，(1)式で i に1を代入して，

$$F(u_0)=Pr(U\leq u_0)=\left(\int_{1-u_0}^1 b(x)dx\right)^{n-1} \tag{9}$$

であることがわかる。これは，$n-1$ 個のランダム点が u_0 より右にある確率である。これより，$F(u_0)$ の確率密度関数を $f(u_0)$ とおけば，(9)を u_0 で微分して，

$$f(u_0)=(n-1)\left(\int_{1-u_0}^1 b(x)dx\right)^{n-2}b(1-u_0) \tag{10}$$

が得られる。ここで，第1番目の区間の長さを d とすると，$d=1-u_0$ となる。このとき，d の確率密度関数を $g_1(d)$ とおくと，一般に

$$g_1(d)=f(u_0)\left|\frac{\partial d}{\partial u_0}\right| \tag{11}$$

の関係が成り立つ（ホーエル，1978：pp.241-243）ことから，

$$g_1(d)=(n-1)\left(\int_d^1 b(x)dx\right)^{n-2}b(d) \tag{12}$$

である。

以上より，ランダムに確率密度関数$b(x)$に従って分布する点によって区切られた区間の長さの確率密度関数は，$b(x)$の積分式であらわされることがわかった。ところで，$b(x)$として最も基本的なのは一様分布であろう。そこで，ここでは点が一様ランダムに分布する場合について区間の長さの確率密度関数を求めよう。なお，後述のように，この場合はいわゆるbroken-stick問題（たとえば寺本，1990：50）と呼ばれるものに相当する。

一様分布の場合，線分の長さを1とおいたので，

$$b(x)=1(0\leq x\leq 1) \tag{13}$$

である。まず，$2\leq i\leq n-1$，つまり1番目，n番目以外の区間について計算するために，(13)式を(8)式に代入すると次の式が得られる。

$$h(l,\mathrm{d})=\binom{n-1}{i-1}(i-1)(n-i)l^{i-2}(1-d-l)^{n-i-1} \tag{14}$$

これをlについて区間［0, $1-d$］で積分すれば，$g(d)$が得られる。つまり，

$$g(d)=\binom{n-1}{i-1}(i-1)(n-i)\int_0^{1-d}l^{i-2}(1-d-l)^{n-i-1}dl \tag{15}$$

である。以下ではこの式の右辺の積分を求めよう。

もしdが1でなければ，$l=(1-d)m$とおけば，右辺の積分式は

$$\begin{aligned}&\int_0^{1-d}l^{i-2}(1-d-l)^{n-i-1}dl\\&=\int_0^1(1-d)^{n-3}m^{i-2}(1-m)^{n-i-1}(1-d)dm\\&=(1-d)^{n-2}\int_0^1 m^{i-2}(1-m)^{n-i-1}dm\end{aligned}$$

であり，この積分がベータ関数であることを考慮すれば，

$$=(1-d)^{n-2}\frac{(i-2)!(n-i-1)!}{(n-2)!} \tag{16}$$

となる。これを(15)式に代入して整理すれば，

$$g(d)=(n-1)(1-d)^{n-2} \tag{17}$$

となる。ところで(17)式は$d<1$という条件のもとで計算していた。一方で$d=1$のときには(15)式より$g(d)=0$であるが，これは(17)式と一致している。したがって，$0\leq d\leq 1$で(17)式は成立する。

さらに$i=1$あるいは$i=n$，つまり左端，右端の区間においては，(12)式に$b(x)=1$を代入すれば，やはり(17)式が得られる。

以上より，長さ1の線分が$n-1$個の一様ランダムな点によって分けられたときには，任意の区間の長さの確率密度関数は，

$$g(d)=(n-1)(1-d)^{n-2}(0\leq d\leq 1) \tag{18}$$

であることがわかった。

なお，上記のように，$b(x)=1$，つまり一様ランダムな点によって分割された区間については，broken-stick問題として研究が行われてきた。この式(18)は，これらの結果として得られた式（たとえばWebb，1974）と一致している。

4．ふたつの線分からひとつづつ区間を取り出した場合

前節ではひとつの線分をランダムな点で分割して得られた区間の長さの確率密度関数を求めたが，都市，地域の研究では，複数の線分について区間の長さを比べる必要が生じる場合もある。この例としては，異なるバス会社についてバス停の間隔を比較する場合があげられる。そこでこの節では，ふたつの線分からひとつづつ区間を取り出して比較する場合について検討しよう。（なお，この問題は，紅白の千歳飴を割って赤い破片と白い破片をひとつづつもらったときに，両方の長さを比べる問題である。そこで以下ではこの問題を千歳飴問題と呼ぶことにする。）

ここで，長さL_r，L_w（rとwはそれぞれ赤と白の略である）のふたつの線分を，おのおのn_r-1，

n_w-1 個の一様ランダムな点で分割して得られた区間をひとつづつ取り出したところ，その長さがおのおの S_r, S_w であったとしよう。

この場合，ふたつの区間を比較するための指標として，次式で定義される R を利用することができる。

$$R=\frac{S_r}{S_r+S_w} \tag{19}$$

この R の確率密度関数はどのようにあらわされるのだろうか。

［千歳飴問題］(19)式の R の確率密度関数 $r(R)$ を求めよ。

まず，S_r, S_w の確率密度関数は，(18)式より次の形になることがわかる。

$$g(S_r)=(n_r-1)\left(1-\frac{S_r}{L_r}\right)^{n_r-1}$$
$$g(S_w)=(n_w-1)\left(1-\frac{S_w}{L_w}\right)^{n_w-1} \tag{20}$$

S_r と S_w は互いに独立であるから，両者の同時密度関数 $p(S_r,\ S_w)$ は

$$p(S_r,\ S_w)=g(S_r)g(S_w) \tag{21}$$

である。一方，R は(19)式から S_r, S_w の単調な関数であることから，S_r と R の同時密度関数を $q(S_r,\ R)$ とおくと，(6)式と同様に

$$q(S_r,\ R)=\frac{p(S_r,S_w)}{\left|\dfrac{\partial R}{\partial S_w}\right|}$$
$$=\left[p(S_r,\ S_w)\frac{(S_r,S_w)^2}{-S_r}\right]_{S_w=S_r\left(\frac{1}{R}-1\right)} \tag{22}$$

の関数が成り立つ（ホーエル，1978：pp. 244-247)。したがって，

$$q(S_r,R)=$$
$$(n_r-1)(n_w-1)\left(1-\frac{S_r}{L_r}\right)^{n_r-2}$$
$$\left\{1-\frac{S_r}{L_w}\left(\frac{1}{R}-1\right)\right\}^{n_w-2}\frac{S_r}{R^2} \tag{23}$$

である。これを S_r について 0 から L_r まで積分すれば R の確率密度関数 $r(R)$ が求められる。

$$r(R)=$$
$$\frac{(n_r-1)(n_w-1)}{R^2}\int_0^{L_r}\left(1-\frac{S_r}{L_r}\right)^{n_r-2}$$
$$\left\{1-\frac{S_r}{L_w}\left(\frac{1}{R}-1\right)\right\}^{n_w-2}S_r\,dS_r \tag{24}$$

ここで，$S_r=L_r m$（m が変数）とおき，さらに

$$\alpha=\frac{L_r}{L_w}\left(\frac{1}{R}-1\right) \tag{25}$$

とおくと，右辺の積分式は，

$$L_r^2\int_0^1(1-m)^{n_r-2}(1-\alpha m)^{n_w-2}m\,dm=$$
$$=L_r^2\sum_{j=0}^{n_w-2}(-\alpha)^j\binom{n_w-2}{j}\int_0^1(1-m)^{n_r-2}m^{j+1}dm$$

となり，この積分がやはりベータ関数を与えることを考慮すれば，

$$=L_r^2\sum_{j=0}^{n_w-2}(-\alpha)^j\binom{n_w-2}{j}\frac{(n_r-2)!\,(j+1)!}{(n_r+j)!} \tag{26}$$

であることがわかる。これを式(24)に代入すれば，次の式が得られる。

$$r(R)=\frac{L_r^2}{R^2}\sum_{j=0}^{n_w-2}\left\{-\frac{L_r}{L_w}\left(\frac{1}{R}-1\right)\right\}^j$$
$$(j+1)\frac{(n_w-1)!\,(n_r-1)!}{(n_w-2-j)!\,(n_r+j)!} \tag{27}$$

5．おわりに

以上に述べたように，一定の長さの線分をランダムな点で分割した区間の長さの統計学的性質についての基礎的な知見が得られた。さらなる課題としては，この確率密度関数をもとにした，統計的検定の手法の開発があげられる。

なお，この研究の成果の一部概要は，（吉川他，1991）に発表されるものである。

文 献 一 覧

ホーエル，P.G.
1978　『入門数理統計学』（浅井晃・村上正康訳）培風館。
寺本英
1990　『ランダムな現象の数学』　吉岡書店。
吉川徹・岡部篤行
1991　「都市・地域解析のための線分のランダム分割モデルでの区間の長さの確率分布」『日本建築学会大会学術講演梗概集』　F分冊，pp. 419-420。
Webb, D. J.
1974　"The Statistics of Relative Abundance and Diversity" Journal of Theoretical Biology 43：277-291.

Key Words（キー・ワード）

Random Point（ランダムな点），**Interval**（区間），**Probability Distribution**（確率分布），**Broken-Stick Problem**（折れ棒問題），**Index about the Ratio of Two Intervals**（ふたつの区間の比に関する指標）

Analysis of the probability Distribution of Interval Length of Random Points on a Line Segment

Tohru Yoshikawa* and Atsuyuki Okabe**

*Center for Urban Studies, Tokyo Metropolitan University
**Department of Urban Engineering, The University of Tokyo
Comprehensive Urban Studies, No. 43, 1991, pp.

The objective of this study is to obtain the probability distribution of the length of intervals of random points on line segments. First, the characteristics of the probability density function of the length is studied, and it is shown that the function is the integrated form of a function including the probability density function of the random points. Second, the probability density function of an index for the ratio of two intervals from different line segments is calculated.