

Random ベクトルの和の長さの分布について

(3次元と2次元の場合)

脇本 和昌

(1968年12月 受付)

The Distribution of Length of the Sum of n Random Vectors

(three-dimension and two-dimension)

Kazumasa Wakimoto

A man starts from a point O and walks a distance l in a straight line; he then turns through any angle whatever and walks a distance l in a second straight line, he repeats this process n times.

We require the probability distribution for the distance between initial and terminal points in two or three dimensions.

This problem has been examined by various authors for the long time.

In this paper, we report the various other's results and some results of our calculations.

Particularly, it is the purpose of this paper to point out use of the projection of a vector with random direction in three or two dimensions on the fixed line, say the x -axis.

The Institute of Statistical Mathematics

[1] 3次元の場合

(1,1) Ordinary calculation

図1に示すように3次元において原点から出発する一定の長さ l で方向が random である互いに独立な n ステップのベクトル R_1, R_2, \dots, R_n の和の長さ $T_n = \left| \sum_{i=1}^n R_i \right| = \overline{Op_n}$ を考えると(これは始点と終点とを結ぶ線分の長さである), この T_n の分布は S. Chandrasakhar [1] によって次のように与えられている。

$$P_r\{T_n < t\} = G_n(t); \quad 0 \leq t \leq n$$

$$\frac{dG_n(t)}{dt} = g_n(t) = \frac{2t}{\pi} \int_0^{\infty} \sin(xt) \left(\frac{\sin(xl)}{xl} \right)^n x dx; \quad 0 < t \leq n$$

なお G_n は分布関数 (以後 c. d. f. とかく), $g_n(t)$ は密度関数 (以後 p. d. f. とかく), 特に $n \geq 1$ のときは

$$g_n(t) \sim \frac{\sqrt{2} t^2}{\left(\frac{n l^2}{3}\right)^{2/3} \sqrt{\pi}} \exp\left(-\frac{3 t^2}{2 n l^2}\right); \quad t > 0$$

Chandrasekhar [1] は $g_n(t)$ の計算を $n=6$ まで exact に与えており, さらに S. A. Vincenz and J. McG. Bruckshaw [2] は $n=7, 8$ について exact に計算して示している。

(1,2) 1次元 (x 軸) への Projection を利用した Calculation

$g_n(t)$ については Chandrasekhar, Vincenz and Brackshaw は $n=8$ まで非常に複雑な計

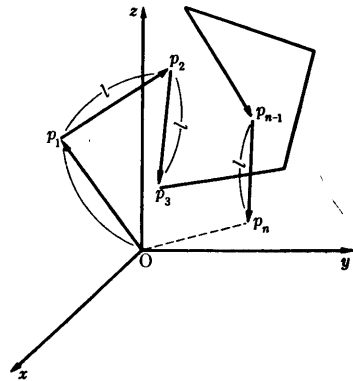


図 1

算の結果, exact に求めているが, W. Feller [3] が次のような考えによって要易に一般の n について exact な計算ができることを示した. ここではその考えを紹介すると共にさらに拡張していろいろの計算を試みその結果を記すことにする.

いま, $\overline{op}_n (=T_n)$ の分布を求めるために, 3次元での長さの分布とそれを1次元(ここでは x 軸を考える)に project したときの長さの分布の間の関係を使う. 特に3次元での長さの分布と1次元での長さの分布との間には大変面白い関係がある.

(イ) 半径 l の球面上に一樣に分布するような原点からの random ベクトルを考え, それを x 軸上に project するとその projection の長さは $(0, l)$ 上の一様分布となる.

この理由を簡単に説明しておこう.

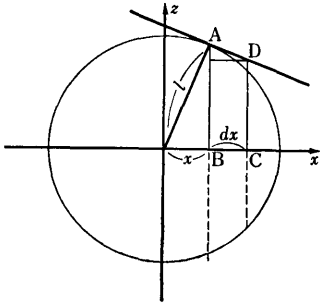


図 2

図2において AD は点 A における接線とする. 台形 ABCD を x 軸のまわりに回転してできる円錐台の側面積は $2\pi l dx$ となり, x 軸上の p. d. f. は

$$f(x)dx = \frac{1}{4\pi l^2} 2\pi l dx = \frac{1}{2l} dx; \quad -l \leq x \leq l$$

となる. 長さの分布は絶対値をとればよいから $(0, l)$ 上の一様分布となる.

(ロ) 次に3次元の原点からのベクトルの長さの c. d. f, p. d. f. を V, v と書き, そのベクトルを x 軸に project したときの長さの c. d. f., p. d. f. を F, f と書くことにしよう.

図3において $|X|=x$ を fix すると event $\{OH \leq t\}$ と event $\{OP \leq \frac{t}{x}\}$ は等しい event であり

$$F(t) = \int_0^1 V\left(\frac{t}{x}\right) dx; \quad t > 0.$$

両辺を微分すると

$$f(t) = \int_0^1 v\left(\frac{t}{x}\right) \frac{dx}{x}; \quad t > 0$$

$\frac{t}{x} = y$ とおくと

$$f(t) = \int_t^\infty \frac{v(y)}{y} dy$$

$$f'(t) = -\frac{v(t)}{t} \quad \therefore v(t) = -t f'(t); \quad t > 0.$$

(1, 2, 1) この (イ), (ロ) の関係を使うと 3次元の原点を出発点とする長さ l の独立な n ステップのベクトル R_1, R_2, \dots, R_n の和の長さ $\left| \sum_{i=1}^n R_i \right|$ の p. d. f. $v_n(x)$ を求めるには $(-l, l)$

上で一様分布をする独立な n 個の random variable を X_1, X_2, \dots, X_n とするとき $\sum_{i=1}^n X_i$ の p. d. f. を求めればただちに求めることができることがわかる.

$\sum_{i=1}^n X_i$ の p. d. f. を $f_n(t)$ とすると

$$f_n(t) = \frac{1}{(2l)^n (n-1)!} \sum_{k=0}^n (-1)^k \binom{n}{k} (t + (n-2k)l)_+^{n-1}; \quad -nl \leq t \leq nl$$

$$\left(\text{但し } t_+ = \frac{|t|+t}{2} \right)$$

となることはよく知られている。

だから (ii) の関係を使うと

$$v_n(t) = -2t f'_n(t) = \frac{-2t}{(2l)^n (n-2)!} \sum_{k=0}^n (-1)^k \binom{n}{k} (t + (n-2k)l)_+^{n-2}; \quad 0 < t \leq nl$$

この式から Chandrasekher, S.A. Vincenz and J. McG. Bruckshaw の exact に計算した結果は要易に求められ、 $n \geq 9$ についても exact に計算することは要易であることがわかる。

(1, 2, 2) 原点から出発する方向が random でその長さが p. d. f.

$$\sqrt{\frac{2}{\pi}} x^2 e^{-x^2/2}; \quad x > 0$$

なる分布に従うとき、独立な n ステップのベクトル R_1, R_2, \dots, R_n を考え原点 (O) と終点

(p_n) を結ぶ線分の長さ $0 \leq \overline{p_n} = T_n = \left| \sum_{i=1}^n R_i \right|$ の分布を求めてみよう。

前記 (ii) に示した関係から 3次元での random ベクトルの長さの分布が p. d. f.

$\sqrt{\frac{2}{\pi}} x^2 e^{-x^2/2}$ をもつとき、その x 軸への projection の長さの分布は p. d. f.

$$f(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}; \quad x > 0$$

になることがわかり、したがって project した点の x 座標は p. d. f.

$$g(x) = \frac{1}{\sqrt{2}x} e^{-x^2/2}; \quad -\infty < x < \infty$$

をもつ。

p. d. f. $g(x)$ をもつような n 個の独立な random variable を X_1, X_2, \dots, X_n とする。

そのとき $\sum_{i=1}^n X_i$ の p. d. f. は

$$f_n(t) = \frac{1}{\sqrt{2\pi} \sqrt{n}} e^{-\frac{t^2}{2n}}; \quad -\infty < t < \infty$$

となることがわかり、

T_n の p. d. f. は

$$v_n(t) = -2t f'_n(t) = \frac{1}{n^{3/2}} \sqrt{\frac{2}{\pi}} t^2 e^{-\frac{t^2}{2n}}; \quad t > 0$$

となる。

平均値と分散は

$$\text{平均値} = E_n(T) = \frac{4\sqrt{n}}{\sqrt{2\pi}}$$

$$\text{分散} = V_n(T) = n \left(3 - \frac{8}{\pi} \right)$$

となり、変動係数は

$$\frac{\sqrt{V_n(T)}}{E_n(T)} = \sqrt{\frac{3}{8}\pi - 1} = 0.4220$$

となる。

(1, 2, 3) 次に同様に原点から出発し方向が random で、その長さが $(0, 1)$ の上の一様分布に従うようなベクトルを考えると n ステップののち原点と終点を結ぶ線分の長さ $\overline{0p_n} = T_n$ の分布をステップ数がある程度大きいときに近似的に求めてみよう。

前記 (ii) に示した関係から、原点から方向が random でその長さが $(0, 1)$ の上の一様分布に従うような vector を考えるとき、そのベクトルの x 軸への projection の長さの p. d. f. は要易に

$$f(x) = \log \frac{1}{x}; \quad 0 < x < 1$$

なることがわかる。

ところで、project した点の x 座標は $(-1, 1)$ の間の値をとりその p. d. f. は

$$g(x) = \frac{1}{2} \log \frac{1}{|x|}; \quad -1 < x < 1$$

よって、この $g(x)$ なる p. d. f. をもつような n 個の独立な random variable を X_1, X_2, \dots, X_n とすると、そのとき $\sum_{i=1}^n X_i$ の p. d. f. は n がある程度大きいときは、中心極限定理を用いると漸的に

$$f_n(t) \sim \frac{1}{\sqrt{2\pi n \sigma_n}} e^{-\frac{t^2}{2n^2 \sigma_n^2}}; \quad -\infty < t < \infty$$

$$\left(\text{但し } \sigma_n^2 = \frac{1}{9n} \right)$$

となることがわかり T_n の p. d. f. は

$$v_n(t) \sim -2t f_n'(t) = \frac{2t^2}{\sqrt{2\pi n^3 \sigma_n^3}} e^{-\frac{t^2}{2n^2 \sigma_n^2}}; \quad t > 0$$

となる。

平均値と分散を求めてみると

$$\text{平均値} = E_n(T) \sim \frac{4n\sigma_n}{\sqrt{2\pi}} = \frac{4\sqrt{n}}{3\sqrt{2\pi}}$$

$$\text{分散} = V_n(T) \sim n^2 \sigma_n^2 \left(3 - \frac{8}{\pi} \right) = n \left(\frac{1}{3} - \frac{8}{9\pi} \right)$$

となり変動係数は

$$\frac{\sqrt{V_n(T)}}{E_n(T)} \sim \sqrt{\frac{3}{8}\pi - 1} = 0.4220$$

が導かれる。

なお各ステップの長さが1である場合も n がある程度大きいときは $\overline{0p_n}$ の p. d. f. は同様な方法で

$$v_n(t) \sim \frac{\sqrt{2} t^2}{\left(\frac{n}{3}\right)^{3/2} \sqrt{\pi}} e^{-\frac{3}{2n} t^2}; \quad t > 0$$

となる。これは (1, 1) で求めた Chandrasekhar の結果と一致する。計算は Chandrasekhar に比べると大変簡単である。その平均値、分散は次のようになる。

$$\text{平均値} = E_n(T) \sim \frac{4\sqrt{n}}{\sqrt{6\pi}}$$

$$\text{分散} = V_n(T) \sim n \left(1 - \frac{8}{3\pi} \right)$$

よって、変動係数は

$$\frac{\sqrt{V_n(T)}}{E_n(T)} \sim \sqrt{\frac{3}{8}\pi - 1} = 0.4220$$

となり、これは前に述べた各ステップのベクトルの長さが他の分布に従う場合と一致する。

2次元の場合にも変動係数は各ステップのベクトルの長さの分布に無関係に求められることが、(2, 3)の結果よりわかる、

[2] 2次元の場合

(2, 1) Ordinary calculation

図4のように原点から出発する方向がrandomでステップの長さが l_1, l_2, \dots, l_n の random ベクトル R_1, R_2, \dots, R_n において原点と終点とを結ぶ長さ $\overline{Op_n} = T_n = \left| \sum_{i=1}^n R_i \right|$ の分布は J.C. Kluyver

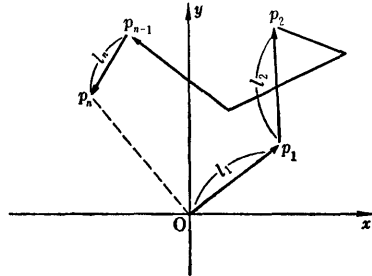


図 4

[4] によって次のように求められている。

$$P_r\{T_n < t\} = t \int_0^\infty J_1(tx) \left[\prod_{i=1}^n J_0(l_i x) \right] dx$$

(但し $J_n(z) = \frac{1}{2\pi} \int_0^{2\pi} \cos(n\theta - z \sin \theta) d\theta$ なる Bessel's function)

$l_1=l_2=\dots=l_n=l$ のときは

$$P_r\{T_n < t\} = t \int_0^\infty J_1(tx) [J_0(lx)]^n dx$$

となり、面白い結果として

$$P_r\{T_n < l\} = \frac{1}{n+1}$$

が示されている。

さらに $l_1=l_2=\dots=l_n=1$ の場合は J. A. Greenwood and D. Durand [5] が

$$P_r\{T_n < t\} = t \int_0^\infty [J_0(x)]^n J_1(tx) dx$$

の値を $n=6, 7, 8, \dots, 24$ まで計算し数値表にしている。また彼等は次に示す L. Rayleigh [6] の求めた近似式で計算した結果と比較している。

L. Rayleigh は $P_r\{T_n < t\}$ の近似式として

$$P_r\{T_n < t\} \sim 1 - e^{-t^2/n}; \quad t > 0$$

を導びいている。

この近似度は例えば $n=7$ の場合は次のようである。

t	$t \int_0^\infty J_1(tx) [J_0(lx)]^n dx$	$1 - e^{-t^2/n}$
1	0.125	0.133
2	0.418	0.435
3	0.714	0.724
4	0.900	0.898
5	0.979	0.972
6	0.998	0.994
7	1.000	0.999

(2,2) 1次元 (x軸) への Projection を利用した calculation

(1,2) で考えたと同じようなことをここでも考えてみるが、3次元のときと比べると相当複雑な関係を持つことがわかる。

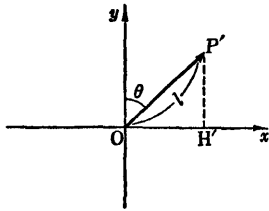


図 5

(1) 簡単のために図5のような random unit ベクトルを考えると、その x 軸への projection の長さの分布は p. d. f. で示すと

$$f(x) = \frac{2}{\pi \sqrt{1-x^2}}; \quad 0 < x < 1$$

となることはよく知られている。

理由は図5において $\overline{OP'}$ と y 軸との間の角度を θ とすると $\overline{OH'} = |\sin \theta|$ となり、よって

$$P_r \{ |\sin \theta| \leq x \} = \frac{2}{\pi} \sin^{-1} x = F(x); \quad 0 < x < 1$$

両辺を微分すると $f(x) = \frac{2}{\pi \sqrt{1-x^2}}; \quad 0 < x < 1$ を求めることができる。

(ii) 次に2次元の原点からのベクトルの長さの c. d. f., p. d. f. を V, v と書き、そのベクトルを x 軸に project したときの長さの p. d. f., c. d. f. を F, f と書くことにすると

図6よる、 $\overline{OH'} = \sin \theta$ を fix して考えると event $\{\overline{OH} \leq t\}$ と event $\{\overline{OP} \leq \frac{t}{\sin \theta}\}$ は等しい event であり、 \overline{OP} が random vector であることから θ の p. d. f. は

$$h(\theta) = \frac{1}{2\pi}; \quad 0 \leq \theta \leq 2\pi$$

であることがわかり、

$$\begin{aligned} F(t) &= 4 \int_0^{\pi/2} V\left(\frac{t}{\sin \theta}\right) \frac{1}{2\pi} d\theta \\ &= \frac{2}{\pi} \int_0^{\pi/2} V\left(\frac{t}{\sin \theta}\right) d\theta; \quad t > 0 \end{aligned}$$

なる関係が導ける。

さらにこの式は

$$V(t) = 1 - t \int_0^{\pi/2} f\left(\frac{t}{\sin \theta}\right) \frac{d\theta}{\sin^2 \theta}; \quad t > 0$$

と書きなおすことができる。

(2,2,1) ここでこの (i), (ii) の関係を使って random unit ベクトルの n 個の和の長さの分布 (各ステップの長さが1である n ステップの random ベクトルにおける始点と終点を結ぶ線分の長さの分布) を n がある程度大きいときに求めてみよう。

Unit vector の x 軸への projection の長さの p. d. f. は

$$f(x) = \frac{2}{\pi \sqrt{1-x^2}}; \quad 0 < x < 1$$

である。だから project した点の x 座標は $(-1, 1)$ の間の値をとりその p. d. f. は

$$g(x) = \frac{1}{\pi \sqrt{1-x^2}}; \quad -1 < x < 1$$

と書ける。

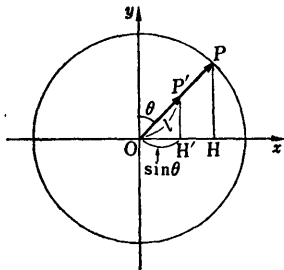


図 6

よってこの $g(x)$ なる p. d. f. をもつような n 個の独立な random variable を X_1, X_2, \dots, X_n とすると $\sum_{i=1}^n X_i$ の p. d. f. は n がある程度大きいときに漸近式は

$$f_n(t) \sim \frac{1}{\sqrt{2\pi n \sigma_n}} e^{-\frac{t^2}{2n^2 \sigma_n^2}}; \quad -\infty < t < \infty$$

$$\left(\text{但し } \sigma_n^2 = \frac{1}{2n} \right)$$

となり、長さの p. d. f. は $\sum_{i=1}^n X_i$ の絶対値をとればよいから

$$f_n(t) \sim \frac{2}{\sqrt{2\pi n \sigma_n}} e^{-\frac{t^2}{2n^2 \sigma_n^2}}; \quad t > 0, \quad \text{但し } \sigma_n^2 = \frac{1}{2n}$$

となる。

よって (4) の関係を使うと

$$\begin{aligned} V_n(t) &\sim 1 - t \int_0^{\pi/2} f_n\left(\frac{t}{\sin \theta}\right) \frac{1}{\sin^2 \theta} d\theta \\ &= 1 - t \int_0^{\pi/2} \frac{2}{\sqrt{\pi} \sqrt{n}} e^{-\frac{t^2}{n \sin^2 \theta}} \frac{1}{\sin^2 \theta} d\theta \\ &= 1 - \frac{1}{\sqrt{\pi}} e^{-\frac{t^2}{n}} \int_0^{\infty} y^{-1/2} e^{-y} dy \\ &= 1 - \frac{1}{\sqrt{\pi}} e^{-\frac{t^2}{n}} \Gamma\left(\frac{1}{2}\right) \\ &= 1 - e^{-\frac{t^2}{n}}; \quad t > 0 \end{aligned}$$

となつて、これは (2, 1) で述べた L. Rayleigh の結果と一致する。
両辺を微分すると

$$v_n(t) \sim \frac{2}{n} t e^{-\frac{t^2}{n}}; \quad t > 0$$

この $v_n(t)$ の平均値、分散を求めてみると

$$\text{平均値} = E_n(T) \sim \frac{\sqrt{n\pi}}{2}$$

$$\text{分散} = V_n(T) \sim n \left(1 - \frac{\pi}{4} \right)$$

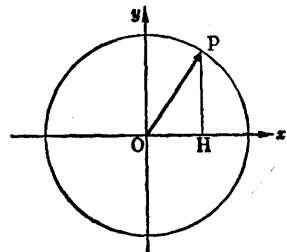
となり、変動体数は

$$\frac{\sqrt{V_n(T)}}{E_n(T)} \sim \sqrt{\frac{4}{\pi} - 1} = 0.5227$$

となる。

(2, 2, 2) 次に原点から出発するその長さが $(0, 1)$ の上の一様分布に従うような独立な n 個の random ベクトル R_1, R_2, \dots, R_n の和の長さ $T_n = \left| \sum_{i=1}^n R_i \right| = \overline{OP}_n$ の分布を求めてみよう。

図 7 において \vec{OP} を長さが $(0, 1)$ の上の一様分布に従うような random ベクトルとする。



を p. d. f. で示すと

$$f(x) = \frac{2}{\pi} \log \frac{1 + \sqrt{1-x^2}}{x}; \quad 0 < x < 1$$

となり、したがって project した点の x 座標は $(-1, 1)$ の間の値をとり、その p. d. f. は

$$g(x) = \frac{1}{\pi} \log \frac{1 + \sqrt{1-x^2}}{|x|}; \quad -1 < x < 1$$

と考えることができる。分散は要易に $V(X) = E(X^2) = \frac{1}{6}$ と計算できる。この $g(x)$ なる p. d. f. をもつような独立な n 個の random variable を X_1, X_2, \dots, X_n とするとき、 $\sum_{i=1}^n X_i$ の p. d. f. は、 n がある程度大きいとき漸近的に

$$g_n(t) \sim \frac{1}{\sqrt{2\pi n} \sigma_n} e^{-\frac{t^2}{2n^2 \sigma_n^2}}; \quad -\infty < t < \infty$$

$$\left(\text{但し } \sigma_n = \frac{1}{6n} \right)$$

となる。

長さの p. d. f. は $\sum_{i=1}^n X_i$ の絶対値をとればよいから

$$f_n(t) \sim \frac{2}{\sqrt{2\pi n} \sigma_n} e^{-\frac{t^2}{2n^2 \sigma_n^2}}; \quad t > 0$$

と書ける。

このとき T_n の c. d. f. は

$$V_n(t) \sim 1 - t \int_0^{\pi/2} f_n\left(\frac{t}{\sin \theta}\right) \frac{d\theta}{\sin^2 \theta}$$

$$= 1 - \frac{1}{\sqrt{\pi}} e^{-\frac{3t^2}{n}} \Gamma\left(\frac{1}{2}\right)$$

$$= 1 - e^{-\frac{3}{n} t^2}; \quad t > 0$$

となり

したがって p. d. f. は

$$v_n(t) \sim \frac{6t}{n} e^{-\frac{3}{n} t^2}; \quad t > 0$$

となる。

この分布の平均値と分散を求めてみると

$$\text{平均値} = E_n(T) \sim \sqrt{\frac{n\pi}{12}}$$

$$\text{分散} = V_n(T) \sim \frac{n}{3} \left(1 - \frac{\pi}{4}\right)$$

となり、変動係数は

$$\frac{\sqrt{V_n(T)}}{E_n(T)} \sim \sqrt{\frac{4}{\pi} - 1} = 0.5227$$

となる。これは各ベクトルの長さが1である場合と一致する。

(2, 3) またこれらの計算は次のようにも考えられる. すなわちいま図 8 において各ベクトル R_1, R_2, \dots, R_n は独立で $\theta_i (i=1, 2, \dots, n)$ は $(0, 2\pi)$ の上で一様分布に従って分布しているとする.

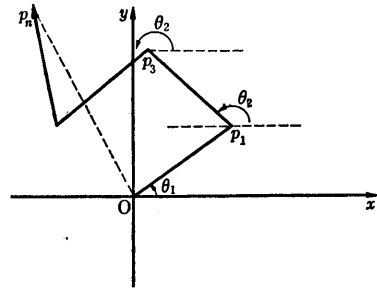


図 8

(a) 各ベクトルの長さが 1 のとき.

$$\begin{aligned} X_i &= \cos \theta_i \\ Y_i &= \sin \theta_i \quad (i = 1, 2, \dots, n) \end{aligned}$$

$$\overline{Op_n} = \left| \sum_{i=1}^n R_i \right| = \sqrt{\left(\sum_{i=1}^n X_i \right)^2 + \left(\sum_{i=1}^n Y_i \right)^2}$$

の分布を求めてみよう.

$$\left. \begin{aligned} E(X_i) &= E(Y_i) = 0 \\ V(X_i) &= V(Y_i) = \frac{1}{2} \\ \text{Cov}(X_i, Y_i) &= 0 \end{aligned} \right\} (i = 1, 2, \dots, n)$$

中心極限定理により

$$\sum_{i=1}^n X_i, \quad \sum_{i=1}^n Y_i \quad \text{の同時分布は漸近的に} \quad N \begin{pmatrix} \frac{n}{2} & 0 \\ 0 & \frac{n}{2} \end{pmatrix} \quad \text{に従う.}$$

よって

$$\left(\sqrt{\frac{2}{n}} \sum_{i=1}^n X_i \right)^2 + \left(\sqrt{\frac{2}{n}} \sum_{i=1}^n Y_i \right)^2$$

は漸近的に自由度 2 の χ^2 分布に従う.

$\overline{Op_n} = T_n$ におくと, T_n の p. d. f. $f_n(t)$ は漸近的に次のような形に書ける.

$$f_n(t) \sim \frac{2}{n} t e^{-\frac{t^2}{n}}; \quad t > 0.$$

よって T_n の平均値と分散を計算してみると,

$$E_n(T) \sim \frac{\sqrt{n\pi}}{2}$$

$$V_n(T) \sim n \left(1 - \frac{\pi}{4} \right)$$

となって前記の方法で求めたものと一致する.

(b) 次に各ベクトルの長さ (L) がある確率分布をもっている場合を考えよう.

$$\begin{aligned} X_i &= L \cos \theta_i \\ Y_i &= L \sin \theta_i \quad (i = 1, 2, \dots, n), \quad L \text{ と } \theta_i \text{ は独立} \end{aligned}$$

とおくと

$$E(X_i) = E(Y_i) = 0$$

$$V(X_i) = V(Y_i) = \frac{1}{2} E(L^2) \quad (i = 1, 2, \dots, n)$$

$$\text{Cov}(X_i, Y_i) = 0$$

となり, 中心極限定理により

$$\sum_{i=1}^n X_i, \quad \sum_{i=1}^n Y_i \quad \text{の同時分布は漸近的に} \quad N \begin{pmatrix} \frac{n}{2} E(L^2) & 0 \\ 0 & \frac{n}{2} E(L^2) \end{pmatrix} \quad \text{に従う,}$$

よって

$$\left(\sqrt{\frac{2}{nE(L^2)}} \sum_{i=1}^n X_i \right)^2 + \left(\sqrt{\frac{2}{nE(L^2)}} \sum_{i=1}^n Y_i \right)^2$$

は漸近的に自由度2の χ^2 分布に従う。

$\overline{Op}_n = T_n$ とおくと T_n の p.d.f. $f_n(t)$ は漸近的に次のように求まる。

$$f_n(t) \sim \frac{2}{nE(L^2)} t e^{-\frac{t^2}{nE(L^2)}}; \quad t > 0 \quad .$$

よって、平均値、分散は次のように計算できる。

$$E_n(T) \sim \frac{\sqrt{nE(L^2)}\pi}{2}$$

$$V_n(T) \sim nE(L^2) \left(1 - \frac{\pi}{4} \right)$$

L が $(0, 1)$ の上の一様分布に従う場合は $E[L^2] = \frac{1}{3}$ となって $(2, 2, 2)$ の方法で求めたものと一致する。

ここで \overline{Op}_n の平均値 $E_n(T)$ について3次元と2次元の間の関係を調べてみよう。

3次元の場合 $(1, 2, 3)$ に示した各ステップの長さが $(0, 1)$ の上の一様分布に従う場合と $(2, 2, 2)$ で示した2次元の場合とを比較してみよう。

$$\frac{\text{3次元の平均値 } E_n(T)}{\text{2次元の平均値 } E_n(T)} \sim \frac{\frac{4\sqrt{n}}{3\sqrt{2\pi}}}{\frac{n\pi}{12}} = \frac{4\sqrt{6}}{3\pi} = 1.0396$$

これが $(1, 2, 3)$, $(2, 2, 1)$ で示した各ステップの長さが1のときも同様に

$$\frac{\text{3次元の平均値 } E_n(T)}{\text{2次元の平均値 } E_n(T)} \sim \frac{\frac{4\sqrt{n}}{\sqrt{6\pi}}}{\frac{\sqrt{n\pi}}{2}} = \frac{4\sqrt{6}}{3\pi} = 1.0396$$

となる。

この結果、原点から出発する n ステップのランダムベクトルの和の長さは、平均値の意味で3次元の場合が2次元の場合よりわずかに大きいことがいえる。

さらに4次元、5次元と次元をあげても同様なことが言えると思っているが、この点についてはまだ確かめていない。

[3] Random ベクトルの和の長さの最大値の分布

いま2次元における原点から出発する方向が random で長さが一定またはある分布に従うような独立な n ステップのベクトル R_1, R_2, \dots, R_n を考えよう。

そのとき

$$T_n = \max_{1 \leq k \leq n} \left| \sum_{i=1}^k R_i \right|$$

の分布を求めたいというのが問題である。

この分布を exact に計算することは大変やっかいである。

1次元の場合にはかなり $n \rightarrow \infty$ とした場合に計算されている。ちなみに主なものを紹介しておく

[a] X_1, X_2, \dots, X_n が p.d.f. $(2\pi)^{-1/2} e^{-x^2/2}$ をもつ独立な確率変数とし

$$T_n = X_1 + X_2 + \cdots + X_n$$

とする。そのとき

$$\begin{aligned} \lim_{n \rightarrow \infty} P_r \left\{ \max_{1 \leq k \leq n} |T_k| < t\sqrt{n} \right\} &= \frac{4}{\pi} \sum_{j=0}^{\infty} \frac{(-1)^j}{2j+1} \exp \left\{ -\frac{(2j+1)^2 \pi^2}{8t^2} \right\} \\ &= L(t); \quad t > 0 \end{aligned}$$

となることが L. Bachelier [7] によって示されている。さらに P. Erdős and M. Kac [8] は X_1, X_2, \dots, X_n が平均値 0, 分散 1 である一様分布に従うときも

$$\frac{\max_{1 \leq k \leq n} |T_k|}{\sqrt{n}}$$

の分布関数が $n \rightarrow \infty$ で $L(t)$ に収束することを示している。

[b] また D.A. Darling [9] によって次のことが示されている。

X_1, X_2, \dots, X_n を Cauchy 分布, $\frac{1}{\pi(1+x^2)}$ をもつ独立な確率変数とし

$$T_n = X_1 + X_2 + \cdots + X_n$$

とおくと

$$\max_{1 \leq k \leq n} \frac{|T_k|}{n}$$

の p. d. f. は $n \rightarrow \infty$ のとき

$$f(t) = \frac{1}{\pi t^{1/2} (1+t^2)^{3/4}} \exp \left(-\frac{1}{\pi} \int_0^1 \frac{\log x}{1+x^2} dx \right); \quad t > 0$$

$$f(t) \sim \frac{1}{\pi t^2} \quad (t \rightarrow \infty)$$

$$f(t) \sim \frac{1}{\pi \sqrt{t}} \quad (t \rightarrow 0)$$

を導いている。

さてこれが 2 次元になって

$$T_n = \max_{1 \leq k \leq n} \left| \sum_{i=1}^k R_i \right|$$

の分布を exact に求めるとなると大変である。非常に簡単な場合, すなわち

(イ) $|R_1|, |R_2|, \dots, |R_n|$ がすべて constant な長さ (簡単のために 1 とする) であるとき

(ロ) $|R_1|, |R_2|, \dots, |R_n|$ がすべて独立に $N(0, 1)$ の分布に従うとき

などのときにも大変やっかいである。

例えば $|R_1|, |R_1+R_2|, \dots, |R_1+R_2+\cdots+R_n|$ の同時 p. d. f. を求めても Bessel 関数を含んだ複雑な式となり T_n の分布を計算するのは大変である。また幾何学的に考えて原点から出発した n ステップの random unit ベクトル (イの場合) を考えると

$$T = \max_{1 \leq k \leq n} \left| \sum_{i=1}^k R_i \right|$$

の分布は c. d. f. を使って書くと

$$P_r \{ T \leq t \} = F(t) = 1 - \left[\begin{array}{l} n \text{ ステップまでにベクトルの} \\ \text{和の終点が少なくとも 1 回,} \\ \text{半径 } t \text{ より大きくなる確率} \end{array} \right]$$

と書ける。

だから一般的には点 $P(x_0, y_0)$ (図9参照) から出発する各ベクトルの長さが1の n ステップのベクトルを考える.

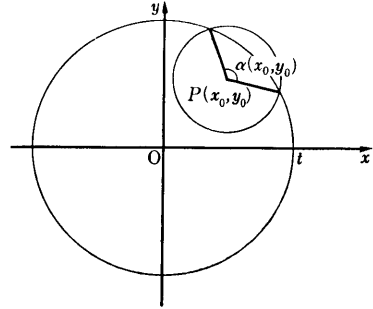


図 9

いま $H_{n,t}(x_0, y_0)$ を点 (x_0, y_0) から出発する n ステップのベクトルについて、 n ステップまでにそのベクトルの和の終点が少なくとも1回原点を中心とする半径 t の円の外に出る確率とすると

$$H_{n,t}(x_0, y_0) = H_{1,t}(x_0, y_0) + \int H_{n-1,t}(x_0 + \cos \theta, y_0 + \sin \theta) \frac{2\pi - a_t(x_0, y_0)}{2\pi} d\theta$$

ある方程式を解けばよいことになる. しかしこれは大変やっかいでまだ未解決である.

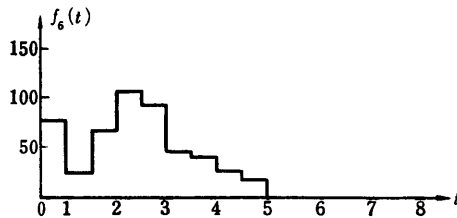
ここではその目安として (4) の場合に関して Sumilation によって求めた結果を紹介しよう.

$$T = \max_{1 \leq k \leq n} \left| \sum_{i=1}^k R_i \right| \text{ の分布についての Sumilation の結果}$$

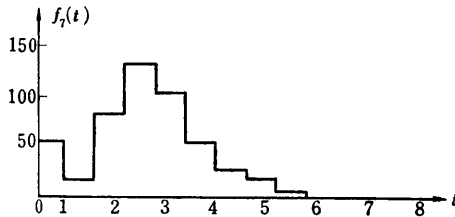
(計算は宮川富美子による)

- n はステップ数
- $f_n(t)$ は n ステップの後に T の実現値の各区間に入った度数
但し $(0, 1)$ の区間の度数は $t=1$ の度数と一致する.
- 使用した乱数の個数は各ステップにつき 500 個, n ステップのとき $500 \times n$ 個, Path の数は 500 (=度数の和)

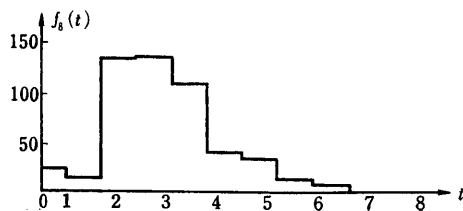
(1) $n = 6$ (区間の中は 0.5)



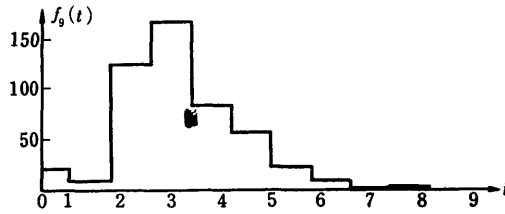
(2) $n = 7$ (区間の中は 0.6)



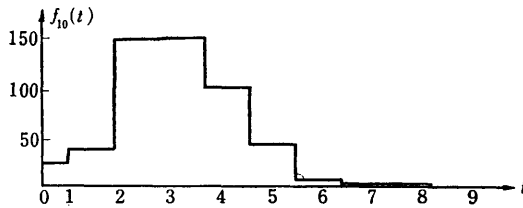
(3) $n = 8$ (区間の中は 0.7)



(4) $n = 9$ (区間の中は 0.8)



(5) $n = 10$ (区間の中は 0.9)



ここでは $T = \max_{1 \leq k \leq n} \left| \sum_{i=1}^k R_i \right|$ の分布についての simulation の結果を非常に簡単な場合について報告したが、方向が random でないとき、また各ステップの長さがいろいろな分布に従うときも、宮川富美子さんによって simulation の結果が出されている。

この結果はうさぎ、鳥などの行動範囲のモデル、また物理学における粒子の運動などのモデルに役立っている。

統計数理研究所

参 考 文 献

- [1] S. Chandrasekher "stochastic problems in physics and astronomy" Rev. Modern Phys., 15, 1-89, (1943)
- [2] S. A. Vincenz and J. McG. Bruckshaw "Note on the probability distribution of a small number of vectors". Pro. Cambridge Philos. Soc., 56, 21-26 (1960)
- [3] W. Feller, "An introduction to probability theory and its application" Vol. 2
- [4] J. C. Kluyver, "A local probability problem" Nederl. Akad. Wetensch. Proc. Vol. 8 (1906) p. 341-350
- [5] J. A. Green wood and D. Durand "The distribution of length and components of the sum of n random unit vectors". Ann. Math. Stas. Vol 26 (1955)
- [6] L. Layleigh "On the problem of random vibrations and random flights in one, two and three dimensions". Philosophical Magazine, Vol. 37, 321-347 (1919)
- [7] L. Bachelier, "Les lois des grands nombres du calcul des probabilités" (Paris, Gauthier-Villars) (1937)
- [8] P. Erdős and M. Kac, "On certain limit theorems of the theory probability" Bull. Amer. Math. Soc. Vol. 52 (1946) 292-302
- [9] D. A. Darling, "The Maximum of sums of stable random variables" Trans. Am. Math. Soc. 83 (1956)