ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIV, 2 (1974)

G. TRYBUŚ (Wrocław)

# ON THE DISTANCE RANDOM VARIABLE

**1. Introduction.** Let $f(\boldsymbol{x})$ denote the probability density function for $\boldsymbol{x} = (x_1, x_2, \ldots, x_k) \in R^k$ and let $P_j = (x_{j1}, x_{j2}, \ldots, x_{jk})$, $j = 0, 1, \ldots, n$, be independent random points having this density. We define the *distance random variable* by

$$C_{n,k} = \min_{1 \leqslant i \leqslant n} \Big(\sum_{j=1}^{k} (x_{0j} - x_{ij})^2\Big)^{1/2}.$$

The distance random variable denotes thus the distance of the random point $P_0$ to the nearest of the points $P_1, P_2, \ldots, P_n$. Because of its simplicity and interesting properties, this random variable has numerous applications in the experimental sciences (see [3], p. 63-132).

The notion of distance random variable has been introduced by Hellwig in [3]. The paper [3] contains a theorem on the limit distribution of $\sqrt{n}\, C_{n,2}$ for $f(\boldsymbol{x})$ uniform on a unit square with $n$ increasing to infinity. The distribution of the distance random variable has also be analyzed by Kopociński in [4], where, among others, the limit distribution of $\sqrt[k]{n}\, C_{n,k}$ for an arbitrary $f(\boldsymbol{x})$ with $n$ increasing to infinity has been given. In the present paper in sections 2 and 3 we remind the fundamental facts about the distance random variable and deal afterwards with the moments of this random variable, the moments of the limit distribution and their approximation using Monte Carlo methods. A comparison of the results obtained by the Monte Carlo methods with the exact ones indicates the usefulness of the Monte Carlo method in this problem.

**2. Distribution of $C_{n,k}$.** Denote by $\Gamma_{n,k}(c) = \mathrm{P}(C_{n,k} \leqslant c)$ the distribution function of $C_{n,k}$ and by $\mathrm{P}(C_{n,k} > c \mid P_0 = \boldsymbol{x})$ the conditional probability that $C_{n,k}$ exceeds $c$ under the condition that the random point $P_0$ equals $\boldsymbol{x}$.

THEOREM 1. *If the random points $P_0, P_1, \ldots, P_n$ are independent and if $f(\boldsymbol{x})$ is the probability density of the random point $P_i$, $i = 0, 1, \ldots, n$,* *then*

$$\mathrm{P}(C_{n,k} > c \mid P_0 = x) = \big(1 - R(x, c)\big)^n$$

*and* [1]

$$\Gamma_{n,k}(c) = 1 - \int \big(1 - R(x, c)\big)^n f(x)\,dx, \tag{1}$$

*where*

$$R(x, c) = \int_{|u|<c} f(x+u)\,du.$$

The proof of this theorem is omitted. It is similar to that of theorem 1 in [4].

COROLLARY 1. *If $f(x)$ is the uniform distribution on the unit interval,*

$$f(x) = \begin{cases} 1, & 0 \leqslant x \leqslant 1, \\ 0, & \textit{otherwise}, \end{cases} \tag{2}$$

*then*

$$\Gamma_{n,1}(c) = \begin{cases} 0, & c < 0, \\ 1 - \dfrac{1}{n+1}\big(2(1-c)^{n+1} + (n-1)(1-2c)^{n+1}\big), & 0 \leqslant c \leqslant \dfrac{1}{2}, \\ 1 - \dfrac{2}{n+1}(1-c)^{n+1}, & \dfrac{1}{2} < c \leqslant 1, \\ 1, & c > 1. \end{cases}$$

Proof. From (1) we have

$$\Gamma_{n,1}(c) = 1 - \int_0^1 \Big(\int_{|u|\leqslant c} f(x+u)\,du\Big)^n dx$$
$$= 1 - \int_0^1 \big(1 - \min(x+c, 1) + \max(0, x-c)\big)^n dx.$$

For $0 \leqslant c \leqslant 1/2$, we have

$$\Gamma_{n,1}(c) = 1 - \int_0^c (1-c-x)^n dx - \int_c^{1-c} (1-2c)^n dx - \int_{1-c}^1 (x-c)^n dx$$
$$= 1 - \frac{2}{n+1}(1-c)^{n+1} - \frac{n-1}{n+1}(1-2c)^{n+1},$$

and for $1/2 < c \leqslant 1$, in an analogous way, we obtain

$$\Gamma_{n,1}(c) = 1 - \int_0^{1-c} (1-c-x)^n dx - \int_c^1 (x-c)^n dx = 1 - \frac{2}{n+1}(1-c)^{n+1},$$

which was to be proved.

[1] In the whole paper, if an integral sign is without integral limits, the integration should be performed on the whole space $R^k$.

COROLLARY 2. *If $f(x)$ is the exponential distribution with parameter $\lambda$,*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0, \\ 0, & x < 0, \end{cases} \tag{3}$$

*then*

$$\Gamma_{n,1}(c) = \begin{cases} 0, & c < 0, \\ 1 - \dfrac{1-\exp\big(-\lambda(n+1)c\big)}{n+1}\Bigg(\exp(-\lambda nc) + \\ \qquad\qquad + \dfrac{1+\exp\big(-\lambda(n+1)c\big)}{\exp(\lambda c)-\exp(-\lambda c)}\Bigg), & c \geqslant 0. \end{cases} \tag{4}$$

Proof. From (1) we have

$$\Gamma_{n,1}(c) = 1 - \lambda\int_0^\infty \Big(1-\lambda\int_{\max(0,x-c)}^{x+c} \exp\big(-\lambda(x+u)\big)du\Big)^n e^{-\lambda x}dx$$

$$= 1 - \lambda\int_0^\infty \Big(1+\exp\big(-\lambda(x+c)\big) - \exp\big(-\lambda\max(0, x-c)\big)\Big)^n e^{-\lambda x}dx$$

$$= 1 - \lambda\exp(-\lambda nc)\int_0^c \exp\big(-\lambda(n+1)x\big)dx -$$

$$-\lambda\int_c^\infty \Big(1+\exp\big(-\lambda(x+c)\big) - \exp\big(-\lambda(x-c)\big)\Big)^n e^{-\lambda x}dx.$$

Now

$$\lambda\int_0^c \exp\big(-\lambda(n+1)x\big)dx = \frac{1}{n+1}\Big(1-\exp\big(-\lambda(n+1)c\big)\Big)$$

and

$$\lambda\int_c^\infty \big(1+e^{-\lambda x}(e^{-\lambda c}-e^{\lambda c})\big)^n e^{-\lambda x}dx$$

$$= \frac{1}{(n+1)\big(\exp(\lambda c)-\exp(-\lambda c)\big)}\Big(1-\exp\big(-2\lambda(n+1)c\big)\Big).$$

This completes the proof of formula (4).

**3. Limit distribution of $C_{n,k}$.** Since the random variable $C_{n,k}$ for $n\to\infty$ is stochastically convergent to zero, one takes the random variable $\sqrt[k]{n}\,C_{n,k}$ to investigate the limit properties of $C_{n,k}$. The limit distribution is defined as follows:

$$\Gamma_k(c) = \lim_{n\to\infty} \mathrm{P}\big(\sqrt[k]{n}\,C_{n,k} \leqslant c\big) = \lim_{n\to\infty} \Gamma_{n,k}\big(c/\sqrt[k]{n}\big).$$

THEOREM 2. *If the probability density $f(\boldsymbol{x})$ is bounded, $f(\boldsymbol{x}) \leqslant M < \infty$, then*

$$\Gamma_k(c) = 1 - \int \exp\big(-V_k(c) f(\boldsymbol{x})\big) f(\boldsymbol{x})\, d\boldsymbol{x}, \tag{5}$$

*where $V_k(c) = \pi^{k/2} c^k / \Gamma(k/2+1)$ is the volume of a $k$-dimensional sphere with radius $c$.*

Proof. From theorem 1, for $f(\boldsymbol{x}) < M < \infty$, we can write the following sequence of equalities:

$$\lim_{n\to\infty} \mathrm{P}\big(\sqrt[k]{n}\, C_{n,k} < c\big) = 1 - \int \Big(1 - \int_{|\boldsymbol{u}| < c/\sqrt[k]{n}} f(\boldsymbol{x}+\boldsymbol{u})\, d\boldsymbol{u}\Big)^n f(\boldsymbol{x}) d\boldsymbol{x}$$

$$= 1 - \lim_{n\to\infty} \int \big(1 - V_k\big(c/\sqrt[k]{n}\big) f(\boldsymbol{x}) + o(n)\big)^n f(\boldsymbol{x})\, d\boldsymbol{x}$$

$$= 1 - \int \exp\big(-V_k(c) f(\boldsymbol{x})\big) f(\boldsymbol{x})\, d\boldsymbol{x}.$$

We have used here the fact that

$$V_k\big(c/\sqrt[k]{n}\big) = \frac{1}{n} V_k(c).$$

This completes the proof of theorem 2.

COROLLARY 3. *If $f(\boldsymbol{x})$ is the uniform distribution on $[0,1]^k$, then*

$$\Gamma_k(c) = 1 - \exp\big(-V_k(c)\big) = 1 - \exp\Big(-\frac{\pi^{k/2} c^k}{\Gamma(k/2+1)}\Big), \quad c \geqslant 0.$$

COROLLARY 4. *If $f(x_1, x_2)$ is the normal distribution,*

$$f(x_1, x_2) = \frac{1}{2\pi} \exp\Big(-\frac{1}{2}(x_1^2 + x_2^2)\Big), \tag{6}$$

*then*

$$\Gamma_2(c) = 1 - \frac{2}{c^2}\Big[1 - \exp\Big(-\frac{1}{2}c^2\Big)\Big], \quad c \geqslant 0. \tag{7}$$

In fact,

$$\Gamma_2(c)$$

$$= 1 - \frac{1}{2\pi} \iint \exp\Big[-\frac{\pi c^2}{2\pi} \exp\Big(-\frac{1}{2}(x_1^2 + x_2^2)\Big)\Big] \exp\Big(-\frac{1}{2}(x_1^2 + x_2^2)\Big) dx_1 dx_2$$

$$= 1 - \int_0^\infty \exp\Big[-\frac{c^2}{2}\exp\Big(-\frac{u^2}{2}\Big)\Big] \exp\Big(-\frac{u^2}{2}\Big) u\,du$$

$$= 1 - \frac{2}{c^2}\Big[1 - \exp\Big(-\frac{1}{2}c^2\Big)\Big].$$

COROLLARY 5. *If $f(x)$ is the exponential density* (3), *then*

$$\Gamma_1(c) = 1 - \frac{1}{2\lambda c}(1 - e^{-2\lambda c}), \qquad c \geqslant 0. \tag{8}$$

Formula (8) is easily obtained by substituting (3) into (5).

**4. Moments in exact and limit distributions.**

COROLLARY 6. *If $f(x)$ is the uniform distribution* (2), *then*

$$\mathrm{E}(C_{n,1}) = \frac{n+3}{2(n+1)(n+2)}, \qquad \mathrm{D}^2(C_{n,1}) = \frac{n^3+11n^2+19n+1}{4(n+1)^2(n+2)^2(n+3)}.$$

Hence

$$\mathrm{E}(C_{n,1}) \sim \frac{1}{2n} \quad \text{and} \quad \mathrm{D}^2(C_{n,1}) \sim \frac{1}{4n^2}.$$

Similarly, one can show that if $f(x_1, x_2)$ is the uniform density on $[0,1]^2$, then

$$\mathrm{E}(C_{n,2}) \sim \frac{1}{2\sqrt{n}} \quad \text{and} \quad \mathrm{D}^2(C_{n,2}) \sim \frac{4-\pi}{4\pi n}.$$

COROLLARY 7. *If $f(x)$ is the exponential density* (3), *then*

$$\begin{aligned} \mathrm{E}(C_{n,1}) &= \frac{1}{\lambda n(2n+1)} + \frac{1}{\lambda(n+1)}\sum_{k=1}^{n+1}\frac{1}{2k-1} \\ &\sim \frac{1}{\lambda n(2n+1)} + \frac{1}{2\lambda(n+1)}\left(\log 4(n+1) + C + \frac{1}{24(n+1)^2}\right), \end{aligned} \tag{9}$$

*where $C = 0.5772115\ldots$ is the Euler constant.*

In fact, from (4) we get

$$\mathrm{E}(C_{n,1}) = \frac{1}{n+1}\int_0^\infty \left(\exp(-\lambda nc) - \exp\big(-\lambda(2n+1)c\big) + \frac{1-\exp\big(-2\lambda(n+1)c\big)}{\exp(\lambda c)-\exp(-\lambda c)}\right)dc.$$

Now

$$\frac{1}{n+1}\int_0^\infty \frac{1-\exp\big(-2\lambda(n+1)c\big)}{\exp(\lambda c)-\exp(-\lambda c)}\,dc = \frac{1}{\lambda(n+1)}\sum_{j=1}^{n+1}\frac{1}{2j-1}$$

(see [2], p. 959), and from this formula (9) follows. Using the formula

$$\sum_{j=1}^{n+1}\frac{1}{2j-1} \sim \log\big(2\sqrt{n+1}\big) + \frac{C}{2} + \frac{1}{48(n+1)^2}$$

(see [2], p. 17), we obtain the asymptotic expression of corollary 7.

THEOREM 3. *If in the limit distribution* (5) *the function* $f(x)$ *is the normal density* (6), *then* $\mu_1 = \sqrt{2\pi}$, *and the moments of order* $r > 1$ *do not exist.*

Proof. From (7) we have

$$\mu_r = r\int\limits_0^\infty c^{r-1}\big(1-\Gamma_k(c)\big)\,dc = 2r\int\limits_0^\infty c^{r-3}\left[1-\exp\left(-\frac{1}{2}c^2\right)\right]dc.$$

For $r > 1$, the moments are not finite, and, for $r = 1$, we obtain

$$\mu_1 = 2\int\limits_0^\infty \frac{1}{c^2}\left[1-\exp\left(-\frac{1}{2}c^2\right)\right]dc = 2\int\limits_0^\infty \exp\left(-\frac{1}{2}c^2\right)dc = \sqrt{2\pi},$$

which was to be proved.

THEOREM 4. *If in the limit distribution* (5) *the function* $f(x)$ *is exponential* (3), *then no moment* $\mu_r$, $r > 0$, *is finite.*

Proof. From (8) we have

$$\mu_r = r\int\limits_0^\infty c^{r-1}\frac{1}{2\lambda c}(1-e^{-2\lambda c})\,dc.$$

This integral does not converge for $r > 0$.

**5. Approximation of $\mathrm{E}(C_{n,1})$ by the Monte Carlo method.** The evaluation of the moments of $C_{n,k}$ is usually cumbersome. In complicated cases the Monte Carlo method can be useful. On the base of $(n+1)$-element samples drawn from a given distribution it is possible to estimate the distribution function of $C_{n,k}$ and its moments with sufficient accuracy.

The results of the performed experiments are the following. On the computer Odra 1204, using random number generators described in [1] and [5], realizations of independent samples from the uniform distribution on the interval $[0, 1]$ and from the exponential distribution with $\lambda = 1$ have been calculated and the distance $C_{n,1}$ determined. The obtained mean distances can be found in tables 1 and 2. However, drawing the numerous samples is costly. The experimentation performed in [3] and [4] shows that the distribution of the distance random variable can be estimated from the distances of one sample. One may assume that the estimation of expected values on the base of one sample is also sufficiently accurate. The results of our experimentation, shown in tables 1 and 2, confirm this fact. For control purposes, in all considered cases the distribution of $\Gamma_{n,1}(c)$ has also been estimated. In no case, for typical significance levels, the hypothesis of equality of the exact and empirical distributions was rejected.

TABLE 1. Uniform distribution

| $n$ | Mean values | | | No. of samples |
|---|---|---|---|---|
| | exact | independent samples | one sample | |
| 5 | .09524 | .09417 | .09254 | 500 |
| 10 | .04924 | .04842 | .04931 | 500 |
| 20 | .02489 | .02479 | .02853 | 500 |
| 50 | .00999 | .01006 | .01007 | 500 |
| 100 | .00500 | .00469 | .00524 | 500 |
| 200 | .00250 | .00261 | .00227 | 500 |
| 500 | .00100 | .00094 | .00100 | 500 |
| 1000 | .00050 | .00047 | .00051 | 500 |

TABLE 2. Exponential distribution ($\lambda = 1$)

| $n$ | Mean values | | | No. of samples |
|---|---|---|---|---|
| | exact | independent samples | one sample | |
| 5 | .33122 | .32489 | .32720 | 1000 |
| 10 | .20302 | .19806 | .22502 | 1000 |
| 20 | .12046 | .11702 | .13940 | 1000 |
| 50 | .05800 | .05694 | .05188 | 1000 |
| 100 | .03262 | .03311 | .03636 | 1000 |
| 200 | .01809 | .02044 | .01963 | 500 |
| 500 | .00817 | .00764 | .00839 | 500 |
| 1000 | .00443 | .00436 | .00430 | 500 |

**References**

[1] W. I. Berezhnov, L. K. Gorskii and W. K. Korolenok (В. И. Бережнов, Л. К. Горский и В. К. Короленок), *Алгоритмы получения последовательностей псевдослучайных чисел с различными законами распределения*, in *Алгоритмы и программы случайного поиска*, Рига 1969, p. 335-339.

[2] I. S. Gradstein and I. M. Rizhik (И. С. Градштейн и И. М. Рыжик), *Таблицы интегралов, сумм, рядов и произведений*, Москва 1971.

[3] Z. Hellwig, *Aproksymacja stochastyczna*, PWE, Warszawa 1965.

[4] B. Kopociński, *Rozkład zmiennej losowej dystansowej*, Prace Naukowe WSE Wrocław 6 (1967), p. 83-92.

[5] *Systemy programowania maszyny cyfrowej Odra 1204, Odra-Algol*, part 2, *Function unif*, WZE Elwro, Wrocław 1970.

INSTITUTE OF STATISTICS AND ECONOMIC ACCOUNTING
GRADUATE SCHOOL OF ECONOMICS, WROCŁAW

*Received on 1. 3. 1973*

G. TRYBUŚ (Wrocław)

## O ZMIENNEJ LOSOWEJ DYSTANSOWEJ


STRESZCZENIE

W początkowej części pracy podano najważniejsze wyniki dotyczące rozkładu prawdopodobieństwa zmiennej losowej dystansowej, zdefiniowanej jako odległość punktu losowego z próby prostej do najbliższego punktu tej próby. Następnie wyznaczono rozkłady zmiennej losowej dystansowej w przypadku jednostajnego i wykładniczego rozkładu w próbie oraz jej momenty. Pokazano również, że w rozkładach granicznych zmiennej losowej dystansowej mogą nie istnieć momenty. W pracy rozważa się również możliwość aproksymacji momentów zmiennej losowej dystansowej metodą Monte Carlo. Porównanie wyników uzyskanych tą metodą i wyników analitycznych wskazuje na przydatność metody Monte Carlo w omawianym zagadnieniu.