Math-Net.Ru

Общероссийский математический портал

А. В. Угланов, Нахождение расстояния между двумя случайными множествами, *Матем. заметки*, 2002, том 72, выпуск 6, 918–923

DOI: <https://doi.org/10.4213/mzm477>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 52.90.164.192

26 апреля 2023 г., 15:44:35

УДК 519.2

НАХОЖДЕНИЕ РАССТОЯНИЯ МЕЖДУ ДВУМЯ СЛУЧАЙНЫМИ МНОЖЕСТВАМИ

А. В. Угланов

Два независимых случайных множества образованы конечными системами независимых равномерно распределенных на единичной окружности случайных величин. Выводится формула для распределения расстояния между множествами.
Библиография: 6 названий.

На окружности единичной длины равномерно и независимо распределены $n + m$ точек. Пусть M_1, M_2 суть множества, образованные соответственно n и m точками. Положим $\rho = \rho(M_1, M_2) = \min\{\rho(x, y) : x \in M_1, y \in M_2\}$, где $\rho(x, y)$ есть дуговое расстояние между точками x и y . Наш основной результат – нахождение функции распределения F_ρ случайной величины ρ .

Для $t \in \mathbb{R}^1$ и целого неотрицательного k положим

$$t_+^k = \begin{cases} 0, & \text{если } t < 0, k > 0, \\ 1, & \text{если } k = 0, \\ t^k, & \text{если } t \geq 0, k > 0. \end{cases}$$

ТЕОРЕМА. Для $s \geq 0$ справедливо равенство

$$F_\rho(s) = P(\rho \leq s) = 1 - \frac{(n-1)!(m-1)!}{(n+m-1)!} \sum_{j=1}^{\min(n,m)} C_n^j C_m^j j (1-2js)_+^{n+m-1}.$$

ДОКАЗАТЕЛЬСТВО. Точки множества M_1 разбивают окружность на n дуг; занумеруем эти дуги в порядке их последовательного расположения (первая дуга и направление обхода окружности произвольные) и обозначим через $\xi_1, \dots, \xi_n$ длины дуг. Известно [1, гл. 9, § 1], что плотность совместного распределения величин $\xi_1, \dots, \xi_{n-1}$ дается формулой

$$f(x_1, \dots, x_{n-1}) = \begin{cases} (n-1)!, & \text{если } x_1, \dots, x_{n-1} \geq 0, \sum_{i=1}^{n-1} x_i \leq 1, \\ 0 & \text{в противном случае.} \end{cases}$$

Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант № 01-01-00701.

Далее используется обозначение

$$X^{n-1} = \left\{ x \in \mathbb{R}^{n-1} : x_1, \dots, x_{n-1} \geq 0, \sum_{i=1}^{n-1} x_i \leq 1 \right\}.$$

Введем гипотезы $H_x = \{(\xi_1, \dots, \xi_{n-1}) = x\}$, где $x \in \mathbb{R}^{n-1}$. Для $c \geq 0$ по формуле полной вероятности имеем

$$\begin{aligned} P\left(\rho > \frac{c}{2}\right) &= \int_{\mathbb{R}^{n-1}} P\left(\rho > \frac{c}{2} \mid H_x\right) f(x_1, \dots, x_{n-1}) dx_1 \dots dx_{n-1} \\ &= (n-1)! \int_{X^{n-1}} P\left(\rho > \frac{c}{2} \mid H_x\right) dx_1 \dots dx_{n-1}. \end{aligned} \quad (1)$$

Обозначим через η_i , $i = 1, \dots, n$, число точек множества M_2 , попавших в i -ю дугу, и пусть $\eta = (\eta_1, \dots, \eta_n)$. Для целых неотрицательных чисел $m_1, \dots, m_n$, удовлетворяющих равенству $m_1 + \dots + m_n = m$, положим $\vec{m} = (m_1, \dots, m_n)$, $G_{\vec{m}} = \{\eta = \vec{m}\}$. Пусть p_i , $i = 1, \dots, n$, есть вероятность попадания любой точки множества M_2 в i -ю дугу при условии H_x ; по формулам обобщенной схемы Бернулли получаем

$$P(G_{\vec{m}} \mid H_x) = \frac{m!}{m_1! \dots m_n!} p_1^{m_1} \dots p_n^{m_n}. \quad (2)$$

Так как точки множества M_2 распределены на окружности равномерно и независимо, то

$$p_i = x_i, \quad i = 1, \dots, n-1, \quad p_n = 1 - \sum_{i=1}^{n-1} x_i.$$

Теперь равенство (2) дает

$$P(G_{\vec{m}} \mid H_x) = \frac{m!}{m_1! \dots m_n!} \left[\prod_{i=1}^{n-1} x_i^{m_i} \right] \left[1 - \sum_{i=1}^{n-1} x_i \right]^{m_n}.$$

Отсюда, из (1) и формулы полной вероятности находим

$$\begin{aligned} P\left(\rho > \frac{c}{2}\right) &= (n-1)! \int_{X^{n-1}} \sum_{\vec{m}} P\left(\rho > \frac{c}{2} \mid H_x G_{\vec{m}}\right) P(G_{\vec{m}} \mid H_x) dx_1 \dots dx_{n-1} \\ &= (n-1)! \int_{X^{n-1}} \sum_{\vec{m}} \frac{m!}{m_1! \dots m_n!} \left[\prod_{i=1}^{n-1} x_i^{m_i} \right] \left[1 - \sum_{i=1}^{n-1} x_i \right]^{m_n} \\ &\quad \times P\left(\rho > \frac{c}{2} \mid H_x G_{\vec{m}}\right) dx_1 \dots dx_{n-1}. \end{aligned} \quad (3)$$

Ясно, что для выполнения неравенства $\rho > c/2$ необходимо и достаточно, чтобы для любого $i = 1, \dots, n$ все точки множества M_2 , попавшие в i -ю дугу, лежали в $(c/2)$ -внутренности этой дуги. Тогда, используя независимость и равномерную распределенность точек множества M_2 , находим

$$P\left(\rho > \frac{c}{2} \mid H_x G_{\vec{m}}\right) = \left[\prod_{i=1}^{n-1} \left(\frac{x_i - c}{x_i} \right)_+^{m_i} \right] \left[\frac{1 - \sum_{i=1}^{n-1} x_i - c}{1 - \sum_{i=1}^{n-1} x_i} \right]_+^{m_n}.$$

Подставляя это равенство в (3), получаем

$$\begin{aligned}
 P\left(\rho > \frac{c}{2}\right) &= (n-1)! \int_{X^{n-1}} \sum_{\vec{m}} \frac{m!}{m_1! \cdots m_n!} \left[\prod_{i=1}^{n-1} x_i^{m_i} \right] \left[1 - \sum_{i=1}^{n-1} x_i \right]^{m_n} \\
 &\quad \times \left[\prod_{i=1}^{n-1} \left(\frac{x_i - c}{x_i} \right)_+^{m_i} \right] \left[\frac{1 - \sum_{i=1}^{n-1} x_i - c}{1 - \sum_{i=1}^{n-1} x_i} \right]_+^{m_n} dx_1 \dots dx_{n-1} \\
 &= (n-1)! \sum_{\vec{m}} \frac{m!}{m_1! \cdots m_n!} \int_{X^{n-1}} \left[\prod_{i=1}^{n-1} (x_i - c)_+^{m_i} \right] \\
 &\quad \times \left[1 - \sum_{i=1}^{n-1} x_i - c \right]_+^{m_n} dx_1 \dots dx_{n-1}. \tag{4}
 \end{aligned}$$

Назовем индекс i *хорошим*, если $m_i > 0$. Пусть хороших индексов j штук, $j = 1, \dots, \min(n, m)$. Обозначим последний интеграл в равенствах (4) через $I(\vec{m})$ и сделаем в этом интеграле замену, полагая

$$y_i = \begin{cases} x_i - c, & \text{если индекс } i \text{ хороший,} \\ x_i & \text{в противном случае.} \end{cases}$$

Тогда

$$\sum_{i=1}^{n-1} x_i = \sum_{i=1}^{n-1} y_i + jc, \quad \sum_{i=1}^{n-1} y_i \leq 1 - jc.$$

Пусть $m_n > 0$. Произведя необходимую перенумерацию, будем считать, что хорошие индексы – это $n - j + 1, \dots, n$. Временно положим $h = 1 - jc$. Теперь получаем

$$\begin{aligned}
 I(\vec{m}) &= \int_{\sum_{i=1}^{n-1} y_i \leq h} \left[\prod_{i=n-j+1}^{n-1} (y_i)_+^{m_i} \right] \left[h - \sum_{i=1}^{n-1} y_i \right]^{m_n} dy_1 \dots dy_{n-1} \\
 &= \int_0^h \left(\int_0^{h-y_1} \left(\dots \int_0^{h-\sum_{i=1}^{n-j} y_i} y_{n-j+1}^{m_{n-j+1}} \right. \right. \\
 &\quad \times \left. \left(\dots \int_0^{h-\sum_{i=1}^{n-2} y_i} y_{n-1}^{m_{n-1}} \left(h - \sum_{i=1}^{n-1} y_i \right)^{m_n} dy_{n-1} \right) \dots dy_{n-j+1} \right) \dots dy_2 \Big) dy_1. \tag{5}
 \end{aligned}$$

Временно считая, что $1 < j < n$, полагая для $k = 1, \dots, j-1$ $l_k = m_n + \dots + m_{n-k+1}$ ($l_1 = m_n$), $t_k = h - \sum_{i=1}^{n-k-1} y_i$, выполняя замену $y_{n-k} = t_k s$ и используя известное значение эйлерова интеграла (бета-функции), находим

$$\begin{aligned}
 &\int_0^{h-\sum_{i=1}^{n-k-1} y_i} y_{n-k}^{m_{n-k}} \left(h - \sum_{i=1}^{n-k} y_i \right)^{l_k+k-1} dy_{n-k} \\
 &= \int_0^{t_k} y_{n-k}^{m_{n-k}} (t - y_{n-k})^{l_k+k-1} dy_{n-k} = t_k^{l_k+1+k} \int_0^1 s^{m_{n-k}} (1-s)^{l_k+k-1} ds \\
 &= \left(h - \sum_{i=1}^{n-k-1} y_i \right)^{l_k+1+k} \frac{m_{n-k}! (l_k + k - 1)!}{(l_k + 1 + k)!}.
 \end{aligned}$$

Подставляя полученное выражение в (5) последовательно для $k = 1, \dots, j-1$, получим

$$\begin{aligned}
 I(\vec{m}) &= \frac{m_n! m_{n-1}! \cdots m_{n-j+1}!}{(m_n + \cdots + m_{n-j+1} + j - 1)!} \int_0^h \left(\int_0^{h-y_1} \left(\cdots \right. \right. \\
 &\quad \times \left. \left. \int_0^{h-\sum_{i=1}^{n-j-1} y_i} \left(h - \sum_{i=1}^{n-j} y_i \right)^{m_n + \cdots + m_{n-j+1} + j - 1} dy_{n-j} \right) \cdots \right) dy_2 \cdots dy_1 \\
 &= \frac{m_n! m_{n-1}! \cdots m_{n-j+1}!}{(m_n + \cdots + m_{n-j+1} + j - 1)!} \\
 &\quad \times \int_{y_i \geq 0, \sum_{i=1}^{n-j} y_i \leq h} \left(h - \sum_{i=1}^{n-j} y_i \right)^{m_n + \cdots + m_{n-j+1} + j - 1} dy_1 \cdots dy_{n-j} \\
 &= \frac{m_n! m_{n-1}! \cdots m_{n-j+1}!}{(n + m_n + \cdots + m_{n-j+1} - 1)!} (1 - jc)_+^{n + m_n + \cdots + m_{n-j+1} - 1}.
 \end{aligned}$$

В последнем равенстве использована формула

$$\int_{x_i \geq 0, \sum_{i=1}^r x_i \leq t} \left(t - \sum_{i=1}^r x_i \right)^l dx = \frac{t_+^{r+l} l!}{(r+l)!}.$$

Полученное выражение для $I(\vec{m})$ верно и для $j = 1, n$ (доказывается по той же схеме с очевидными изменениями), а также при $m_n = 0$ (легко проверяется). Подставляя это выражение в равенство (4) и используя равенства $m_1 = \cdots = m_{n-j} = 0$, $m_1 + \cdots + m_n = m$, находим

$$P\left(\rho > \frac{c}{2}\right) = (n-1)! \sum_{\vec{m}} \frac{m!}{(n+m-1)!} (1-jc)_+^{n+m-1}. \quad (6)$$

Мы видим, что при фиксированном j правая часть (6) не зависит от $\vec{m}$. Найдем число слагаемых в сумме (6) при фиксированном j . Выбор j хороших индексов можно осуществить C_n^j различными способами. Решение элементарной комбинаторной задачи показывает, что число m можно представить в виде суммы j ненулевых слагаемых (с учетом порядка) C_{m-1}^{j-1} различными способами. Итак,

$$\begin{aligned}
 P\left(\rho > \frac{c}{2}\right) &= \frac{(n-1)! m!}{(n+m-1)!} \sum_{j=1}^{\min(n, m)} C_n^j C_{m-1}^{j-1} (1-jc)_+^{n+m-1} \\
 &= \frac{(n-1)! (m-1)!}{(n+m-1)!} \sum_{j=1}^{\min(n, m)} C_n^j C_m^j (1-jc)_+^{n+m-1}
 \end{aligned}$$

и теорема доказана.

Величина ρ абсолютно непрерывна: плотность ее распределения дается формулой

$$f_\rho(s) = \frac{dF_\rho(s)}{ds} = 2 \frac{(n-1)! (m-1)!}{(n+m-2)!} \sum_{j=1}^{\min(n, m)} C_n^j C_m^j j^2 (1-2js)_+^{n+m-2}, \quad s \geq 0$$

(здесь $n + m > 2$; в противном случае выражение для f_ρ очевидно).

Найдем еще математическое ожидание величины ρ . Используя формулы

$$\int_0^1 x^k (1-x)^l dx = \frac{k!l!}{(k+l+1)!} \quad \text{и} \quad \sum_{j=0}^{\min(k,l)} C_k^j C_l^j = C_{k+l}^l$$

(в последних двух равенствах соответственно), имеем

$$\begin{aligned} E\rho &= \int_{-\infty}^{\infty} s f_\rho(s) ds = \frac{(n-1)!(m-1)!}{(n+m-2)!} \int_0^1 \sum_{j=1}^{\min(n,m)} C_n^j C_m^j 2j^2 s (1-2js)_+^{n+m-2} ds \\ &= \frac{(n-1)!(m-1)!}{2(n+m-2)!} \sum_{j=1}^{\min(n,m)} C_n^j C_m^j \int_0^1 2js(1-2js)^{n+m-2} d(2js) \\ &= \frac{(n-1)!(m-1)!}{2(n+m)!} \sum_{j=1}^{\min(n,m)} C_n^j C_m^j \\ &= \frac{n!m!}{2nm(n+m)!} [C_{n+m}^n - 1] = \frac{1}{2nm} \left[1 - \frac{1}{C_{n+m}^n} \right]. \end{aligned}$$

ЗАМЕЧАНИЕ 1. Фиксируем $c \geq 0$ и пусть λ есть лебегова мера (дуговой) c -окрестности множества M_1 . Формула полной вероятности, равномерная распределенность и независимость точек множества $M_1 \cup M_2$ и наша теорема дают значения всех моментов случайной величины $1 - \lambda$:

$$\mu_m \stackrel{\text{def}}{=} E(1-\lambda)^m = \frac{(n-1)!(m-1)!}{(n+m-1)!} \sum_{j=1}^{\min(n,m)} C_n^j C_m^j (1-2jc)_+^{n+m-1}$$

$m = 1, 2, \dots$. Поскольку $\sum_{m=1}^{\infty} \mu_{2m}^{-1/(2m)} = \infty$ (условие Карлемана), то распределение величины λ однозначно определяется последовательностью $\{\mu_m\}$ [2, гл. IV, § 3]. Нами получен явный вид этого распределения; соответствующее выражение здесь не приводится ввиду его чрезмерной громоздкости и необходимости некоторых объяснений.

ЗАМЕЧАНИЕ 2. Мы рассмотрели и случай, когда $n + m$ точек равномерно и независимо распределены на единичном отрезке. Решение здесь аналогично приведенному выше; соответствующие статистические характеристики отличаются незначительно.

ЗАМЕЧАНИЕ 3. На практике встречается следующая ситуация. Имеются несколько независимых пуассоновских потоков событий; нужно, чтобы события разных потоков не происходили слишком близко (во времени). Наши рассмотрения позволяют разобраться и в этой ситуации (связь пуассоновских процессов и порядковых статистик независимых равномерно распределенных на отрезке величин хорошо известна: [1, гл. 9, § 1]).

ЗАМЕЧАНИЕ 4. Автор пытался исследовать двумерный аналог рассмотренной задачи (точки распределены на торе или квадрате). Здесь, однако, не удалось получить никакого выражения не только для F_ρ , но даже для $E\rho$ (подчеркнем, что в настоящей статье порядковая структура окружности использовалась самым существенным образом).

ЗАМЕЧАНИЕ 5. В [3] исследовалась задача, похожая на рассмотренную нами, но для одного случайного множества. Точнее, пусть $\xi_1, \dots, \xi_n$ суть независимые одинаково распределенные точки на плоскости, ρ равно минимальному расстоянию между точками ξ_i, ξ_j , $i, j = 1, \dots, n, i \neq j$. Применяя сугубо вероятностный метод, заключающийся в доказательстве некоторой предельной теоремы (типа Пуассона), авторы получили асимптотическое при $n \rightarrow \infty, s \rightarrow 0$ выражение для вероятности $P(\rho > s)$. (На самом деле получен несколько более общий результат.) Аналогичные, но более общие рассмотрения проводились (тем же методом) в [4]–[6]. Не исключено, что этот “асимптотический” метод удастся распространить и для нахождения расстояния между двумя случайными множествами; конечно, получение точных формул (типа даваемой нашей теоремой) на этом пути невозможно, но хотелось бы иметь хотя бы приближенные (см. предыдущее замечание).

Автор искренне благодарит А. Зубкова за проявленное внимание и советы, позволившие улучшить первоначальный вариант статьи.

СПИСОК ЦИТИРОВАННОЙ ЛИТЕРАТУРЫ

- [1] Карлин С. Основы теории случайных процессов. М.: Мир, 1971.
- [2] Прохоров Ю. В., Розанов Ю. А. Теория вероятностей. М.: Наука, 1987.
- [3] Silverman B., Brown T. Short distances, flat triangles and Poisson limits // J. Appl. Probab. 1978. V. 15. № 4. P. 816–826.
- [4] Brown T., Silverman B. Rates of Poisson convergence for U -statistics // J. Appl. Probab. 1979. V. 16. № 2. P. 428–432.
- [5] Barbour A., Eagleson G. Poisson convergence for dissociated statistics // J. Royal Statistics Soc. Ser. B. 1984. V. 46. № 3. P. 397–401.
- [6] Грушо А. А. К вопросу о распределениях U -статистик // Теория вероятн. и ее применения. 1987. Т. 32. № 2. С. 400–404.

Ярославский государственный университет
 E-mail: uglanov@univ.uniyar.ac.ru

Поступило
 21.11.2000