ՀԱՅԿԱԿԱՆ ՍՍՀ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԿԱԴԵՄԻԱՅԻ ԶԵԿՈՒՅՑՆԵՐ
ДОКЛАДЫ АКАДЕМИИ НАУК АРМЯНСКОЙ ССР
LXXII 1981 3

УДК 513.82+519.2+681.142.37

ПРИКЛАДНАЯ МАТЕМАТИКА

Д. О. Аветисян

О четырех теоремах, связанных со статистической теорией информатики

(Представлено чл.-корр. АН Армянской ССР Р. А. Александряном 25/XI 1980)

Рассматриваются задачи о случайных точках в n-мерном евклидовом пространстве E_n. Случайные точки либо распределены на n-мерной единичной сфере, либо же выбираются случайно из совокупности вершин n-мерного куба. Задачи второй группы являются дискретными аналогами задач первой группы и представляют определенный интерес для теории и практики информационного поиска.

Пусть $x=(x_1x_2\ldots x_n)$ и $y=(y_1y_2\ldots y_n)$ радиус-векторы независимых случайных точек x и y, имеющих равномерные распределения на n-мерной единичной сфере с центром в начале координат. Легко показать, что при этом имеют место

$$M[\cos(xy)]=0, \quad D[\cos(xy)]=\frac{1}{n}. \tag{1}$$

Пусть $x=(x_1x_2\ldots x_n)$ и $y=(y_1y_2\ldots y_n)$ радиус-векторы независимых случайных точек x и y, имеющих равномерные распределения на соответствующих окружностях сферы, определенных заданными значениями $\cos(xz_0)$ и $\cos(yz_0)$, где $z_0=(z_{01}z_{02}\ldots z_{0n})$ радиус-вектор произвольно выбранной точки z_0. Покажем, что при этом имеет место следующая теорема:

Теорема 1.

а) $M[\cos(xy)]=\cos(xz_0)\cos(yz_0)$;

б) $D[\cos(xy)]=\dfrac{R_1^2R_2^2}{n-1}$;

в) при $\cos^2(xz_0)+\cos^2(yz_0)<1$ *величина* $\left\{P[\cos(xy)>0]-\frac{1}{2}\right\}$ *является монотонно возрастающей нечетной функцией аргумента* $\cos(xz_0)\cos(yz_0)$;

г) при $\cos^2(xz_0)+\cos^2(yz_0)\geq 1$ *имеет место:*

$$\left\{P[\cos(xy)>0]-\frac{1}{2}\right\}=\frac{1}{2}\operatorname{sgn}[\cos(xz_0)\cos(yz_0)].$$

Приняв в качестве базы n-мерного пространства систему векторов, один из которых (например, с индексом 1) проходит через точку z_0, получим

$$\cos(xy)=\cos(xz_0)\cos(yz_0)+\sum_{i=2}^{n}x_iy_i. \tag{2}$$

Закон распределения случайной величины $\sum_{i=2}^{n}x_iy_i$ совпадает с законом распределения случайной величины $R_1R_2\cos(uv)$, где $R_1=\sqrt{1-\cos^2(xz_0)}$, $R_2=\sqrt{1-\cos^2(yz_0)}$, а $u=(u_1u_2\ldots u_{n-1})$ и $v=(v_1v_2\ldots v_{n-1})$ радиус-векторы независимых случайных точек u и v, имеющих равномерные распределения на $(n-1)$-мерной единичной сфере с центром в начале координат. Отсюда с учетом (1), (2) легко убедиться в справедливости всех четырех пунктов теоремы.

В частном случае, когда $n=2$, формулировка пункта а) теоремы 1 приводит к известному соотношению:

$$\cos(\alpha+\beta)+\cos(\alpha-\beta)=2\cos\alpha\cos\beta.$$

Последний пункт теоремы 1 остается справедливым при произвольных законах распределений случайных точек x и y на упомянутых выше окружностях. При некоторых обобщениях законов распределений x и y остаются справедливыми также другие пункты теоремы 1, что придает теореме в целом определенную прикладную значимость.

Рассмотрим, например, совокупность всех 2^n-2 бинарных функций, определенных на некотором множестве из n элементов и не равных тождественно нулю или единице. Задание каждой из этих функций фиксирует определенное подмножество исходного множества и соответствующую изображающую точку n-мерного евклидового пространства E_n, координаты которой равны значениям данной функции. Геометрическим местом изображающих точек является совокупность всех вершин n-мерного куба за исключением точек (00...0) и (11...1). По определению, коэффициент линейной корреляции $r(xy)$ между произвольно выбранными функциями x и y с изображающими точками $X=(X_1X_2\ldots X_n)$ и $Y=(Y_1Y_2\ldots Y_n)$ равен косинусу угла между векторами $\bar{x}=(\bar{x}_1\bar{x}_2\ldots\bar{x}_n)$ и $\bar{y}=(\bar{y}_1\bar{y}_2\ldots\bar{y}_n)$ с координатами

$$\bar{x}_i=X_i-\frac{1}{n}\sum_{i=1}^{n}X_i;\qquad \bar{y}_i=Y_i-\frac{1}{n}\sum_{i=1}^{n}Y_i$$

или, что то же самое, косинусу угла между радиус-векторами точек $x=\frac{\bar{x}}{|\bar{x}|}$, $y=\frac{\bar{y}}{|\bar{y}|}$. Геометрическим местом этих точек служит единичная сфера с центром в начале координат, скалярные произведения всех радиус-векторов которой с радиус-вектором точки $(11\ldots1)$ равны нулю.

Пусть x и y случайные функции, изображающие точки которых X и Y имеют независимые и равномерные распределения на соответствующих подмножествах вершин куба, определенных заданными значениями $r(xz_0)$ и $r(yz_0)$, где z_0 бинарная функция, соответствующая некоторой произвольно выбранной вершине Z_0. Покажем, что при этом имеет место следующая теорема:

Теорема 2

$$M[r(xy)] = r(xz_0)r(yz_0) \text{ или } M[\cos(xy)] = \cos(xz_0)\cos(yz_0).$$

От теоремы 1 эта теорема отличается тем, что здесь случайные точки x и y имеют *дискретные* распределения на соответствующих окружностях единичной сферы, фактическая размерность которой равна $n-1$.

Для доказательства теоремы 2 выделим из случайной величины $r(xy)$ неслучайную компоненту $r(xz_0)r(yz_0)$, приняв в качестве базы n-мерного пространства систему векторов, один из которых (например, с индексом 1) совпадает с радиус-вектором точки $z_0 = \frac{\bar{z}_0}{|\bar{z}_0|}$:

$$r(xy) = r(xz_0)r(yz_0) + \sum_{i=2}^{n} x_i y_i$$

или

$$r(xy) = r(xz_0)r(yz_0) + uv,$$

откуда с учетом независимости векторов $u = x - \Pi_{p_{z_0}} x$ и $v = y - \Pi_{p_{z_0}} y$ получим

$$M[r(xy)] = r(xz_0)r(yz_0) + M(u)M(v).$$

Отсюда следует, что для доказательства теоремы 2 достаточно показать, что имеют место $M(u) = 0$ или $M(v) = 0$. Для удобства дальнейшего изложения координатные представления n-мерных векторов будем рассматривать как соответствующие n-разрядные коды. Легко показать, что имеет место

$$r(xz_0) = \frac{n a_{xz_0} - \alpha\gamma_0}{\sqrt{\alpha\gamma_0(n-\alpha)(n-\gamma_0)}}, \tag{3}$$

где:

α — число единиц в коде X,

γ_0 — число единиц в коде Z_0,

a_{xz_0} — число разрядов, где коды X и Z_0 одновременно содержат единицы.

Из (3) следует, что при фиксированном Z_0 заданное значение $r(xz_0)$ в общем случае может быть достигнуто при различных парах значений a_{xz_0} и α. Поразрядной проверкой легко убедиться, что для произвольной фиксированной пары значений a_{xz_0} и α имеет место $M(u) = 0$, откуда следует справедливость теоремы 2.

Пусть при фиксированных точках x_0 и y_0 случайная точка z

имеет равномерное распределение на n-мерной единичной сфере. Покажем, что при этом имеет место следующая теорема:

Теорема 3.

$$M[\cos(x_0 z)\cos(y_0 z)] = \frac{\cos(x_0 y_0)}{n}.$$

Примем в качестве базы n-мерного евклидового пространства систему векторов, один из которых (например, с индексом 1) проходит через точку x_0, а другой (например, с индексом 2) принадлежит плоскости, содержащей радиус-векторы точек x_0 и y_0. Тогда имеют место:

$$x_{01} = 1,\ x_{02} = x_{03} = \ldots\ x_{0n} = 0,$$

$$y_{01} = \cos(x_0 y_0),\ y_{02} = \sqrt{1-\cos^2(x_0 y_0)},\ y_{03} = y_{04} = \ldots\ y_{0n} = 0,$$

$$\cos(z x_0), = z_1,\ \cos(z y_0) = z_1 y_{01} + z_2 y_{02},$$

$$\cos(z x_0)\cos(z y_0) = z_1^2 \cos(x_0 y_0) + z_1 z_2 \sqrt{1-\cos^2(x_0 y_0)},$$

$$M[\cos(z x_0)\cos(z y_0)] = \cos(x_0 y) M(z_1^2) + \sqrt{1-\cos^2(x_0 y_0)}\, M(z_1 z_2).$$

В силу симметрии имеют место $M(z_1 z_2) = 0,\ M(z_1^2) = \dfrac{1}{n}$, т. е.

$$M[\cos(z x_0)\cos(z y_0)] = \frac{\cos(x_0 y_0)}{n}.$$

В частном случае, когда $n=2$, формулировка теоремы 3 приводит к известному соотношению:

$$\int\limits_0^{2\pi} \cos x \cos(x-\alpha)\, dx = \pi \cos \alpha.$$

Рассмотрим множество точек—всех вершин n-мерного куба, за исключением вершин $(00\ldots0)$ и $(11\ldots1)$. Пусть при фиксированных точках X_0 и Y_0 случайная точка Z имеет равномерное распределение на элементах этого множества. Покажем, что при этом имеет место следующая теорема:

Теорема 4.

$$M[r(x_0 z) r(y_0 z)] = \frac{r(x_0 y_0)}{n-1} \text{ или } M[\cos(x_0 z)\cos(y_0 z)] = \frac{\cos(x_0 y_0)}{n-1}.$$

От теоремы 3 эта теорема отличается тем, что здесь случайная точка z имеет *дискретное* распределение на единичной сфере, фактическая размерность которой равна $n-1$. Здесь, как и при доказательстве теоремы 3, имеет место

$$M[\cos(x_0 z)\cos(y_0 z)] = \cos(x_0 y_0) M(z_1^2) + \sqrt{1-\cos^2(x_0 y_0)}\, M(z_1 z_2).$$

Однако, в отличие от теоремы 3, где непрерывное и равномерное распределение случайной точки z на $(n-1)$-мерной сфере позволило бы, оперируя условиями симметричности, утверждать справедливость

$M(z_1 z_2)=0$ и $M(z_1^2)=\frac{1}{n-1}$, здесь эти равенства не столь очевидны и требуют своего доказательства. Для простоты изложения и не в ущерб общности доказательства примем, что $X_0=(100\ldots0)$. Сгруппируем все коды точек $Z=(1Z_2Z_3\ldots Z_n)$, содержащих в первых разрядах единицы, по числу единиц $k=0, 1, 2\ldots(n-2)$, содержащихся в остальных $n-1$ разрядах, и покажем, что для каждой из этих групп имеет место $\sum z_1 z_2=0$. Для каждой из этих групп значение z_1 постоянно и равно $z_1=r(x_0 z)=\frac{n-k-1}{\sqrt{(n-1)(k+1)(n-k-1)}}$. Поразрядной проверкой легко убедиться, что для каждой из этих групп сумма векторов типа $z-\Pi_{p_{x_0}}z=\frac{\bar{z}}{|\bar{z}|}-\frac{\bar{x}_0}{|\bar{x}_0|}r(x_0 z)$ равна нулю, т. е. сумма проекций каждой группы векторов z на плоскости, перпендикулярной вектору x_0, равна нулю. Отсюда следует, что для каждой из этих групп имеет место $\sum z_1 z_2=0$. Можно показать, что аналогичная картина имеет место при рассмотрении точек $Z=(0Z_2Z_3\ldots Z_n)$, коды которых в первом разряде содержат нули, что окончательно убеждает в справедливости $M(z_1 z_2)=0$.

Для доказательства условия $M(z_1^2)=\frac{1}{n-1}$ также примем, что $X_0=(100\ldots0)$. Рассматривая, как и в предыдущем случае, группу векторов Z с одинаковыми числами (k) единиц в остальных $n-1$ разрядах, легко показать, что сумма квадратов проекций всех возможных векторов z на вектор x_0 равна величине

$$S=2\sum_{k=0}^{n-2} C_{n-1}^{k}\frac{(n-k-1)^2}{(n-1)(k+1)(n-k-1)}=\frac{2^n-2}{n-1},$$

т. е. имеет место

$$M(z_1^2)=\frac{S}{2^n-2}=\frac{1}{n-1}.$$

Научно-производственное объединение
«Армсельхозмеханизация»

Դ. Հ. ԱՎԵՏԻՍՅԱՆ

Ինֆորմատիկայի վիճակագրական տեսության հետ առնչվող չորս թեորեմների մասին

Էվկլիդյան տարածության մեջ դիտարկվում են պատահական կետեր՝ հավասարաչափ բաշխված n-չափանի գնդային մակերևույթի կամ n-չափանի խորանարդի գագաթների բազմության վրա։ Երկրորդ դեպքում դիտարկվող խնդիրներն ունեն կիրառական մեծ նշանակություն՝ առնչված ինֆորմացիայի փնտրման ավտոմատացված համակարգերի ստեղծման հետ։