ON LARGE INTERVALS OF A RANDOM BOOLEAN FUNCTION

MATHEMATICS

1968

SovietRxiv

Abstract

Full Text

UDC 519.95

MATHEMATICS

G. S. PLESNEVICH

ON LARGE INTERVALS OF A RANDOM BOOLEAN FUNCTION

(Presented by Academician S. L. Sobolev, January 4, 1968)

As is well known, a Boolean function f of n variables may be regarded as the characteristic function of the corresponding subset N_f of the set E_n of vertices of the n-dimensional unit cube ([1]). A k-dimensional face $a \subset E_n$ of the n-dimensional cube (otherwise, an interval of length k, or a k-interval) is called an **interval of the Boolean function** f if $N_f \supset a$.

With each vertex $x \in E_n$ we associate a random variable ω_x, taking the values 0 and 1 with equal probabilities, i.e.

$$P(\omega_x = 0) = P(\omega_x = 1) = \frac{1}{2}.$$

Suppose that the random variables ω_x, $x \in E_n$, are mutually independent. We shall call a realization of the collection $\{\omega_x\}$, $x \in E_n$, of these random variables a **random Boolean function** (of n variables). Thus, the probability that a random Boolean function coincides with an arbitrary fixed Boolean function is equal to 2^{-2^n}.

Introduce the random variables $\xi_{n,k}$ and ζ_n, expressing respectively the number of k-intervals and the maximum of the lengths of intervals of a random Boolean function depending on n variables.

In the present note it is shown that, under certain conditions, the distribution of the random variable $\xi_{n,k}$ is approximated by a Poisson distribution. As a consequence of this fact we obtain an asymptotic formula for the mean value of the random variable ζ_n (in other words, for the mean dimension of a Boolean function depending on n variables).

Theorem 1. Let $n \to \infty$ and $k \to \infty$ in such a way that the quantity

$$\lambda = \lambda(n,k) = \binom{n}{k} \cdot 2^{\,n-k-2^k}$$

remains bounded. Then, for any fixed r,

$$P(\xi_{n,k} = r) = \frac{\lambda^r}{r!} e^{-\lambda} + o(1).$$

Theorem 2. Let $\chi(n) = \min\{k : \lambda(n,k) \leq 1\}$. Then

$$M\zeta_n = \chi(n) - e^{-\lambda(n,\chi(n)-1)} - e^{-\lambda(n,\chi(n))} + o(1), \qquad n \to \infty.$$

Remark 1. It is easy to show that $\chi(n)$ is equal to $[\log_2 n]$ or $[\log_2 n] + 1$ (depending on n).

Remark 2. If $1 \leq \xi_{n,k} \leq 2k+1$, then $\zeta_n = k$. Consequently, in this case $\xi_{n,k}$ is in fact equal to the number of largest intervals of the random Boolean function. (Indeed, if $\zeta_n > k$, then the random Boolean function has at least one interval of length $k+1$, and consequently at least $2(k+1)$ intervals of length k.)

It is known that the n-dimensional cube E_n contains exactly

$$m = \binom{n}{k} \cdot 2^{n-k}$$

distinct k-intervals. With each interval a associate the random variable X_a, equal to one when a is an interval of the random Boolean function, and equal to zero otherwise. Then it is obvious that

$$\xi_{n,k} = \sum X_a, \tag{1}$$

where the sum is taken over all k-intervals a in E_n.

Since $MX_\alpha = 2^{-2k}$, by (1) we have $M\xi_{n,k} = m \cdot 2^{-2k} = \lambda(n,k)$. In proving Theorem 1 we shall use the following formula, valid for any random variable Y taking nonnegative integer values:

$$P(Y = r) = \frac{1}{r!}\sum_{s=0}^{\infty}(-1)^s \frac{M^{(s+r)}(Y)}{s!}. \tag{2}$$

In this formula $M^{(t)}(Y)$ denotes the factorial moment of order t of the random variable Y, i.e.

$$M^{(t)}(Y) = M[Y(Y-1)\dots(Y-t+1)],$$

if $t \geq 1$, and $M^{(0)}(Y) = 1$; here we put $M^{(t)}(Y) = 0$ if all possible values of Y are less than t. The validity of formula (2) is easily established. Expanding in a Taylor series at $z = 1$ the r-th derivative $f^{(r)}(z)$ of the generating function $f(z)$ of the random variable Y, we obtain

$$f^{(r)}(z) = \sum_{s=0}^{\infty}\frac{f^{(r+s)}(1)}{s!}(z-1)^s = \sum_{s=0}^{\infty}\frac{M^{(s+r)}(Y)}{s!}(z-1)^s. \tag{3}$$

Moreover,

$$\left|\sum_{s=t}^{\infty}\frac{M^{(s+r)}(Y)}{s!}(z-1)^s\right| \leq \frac{f^{(t+r)}(\theta|z-1|)}{t!}|z-1|^t \quad (0 \leq \theta \leq 1). \tag{4}$$

Substituting $z = 0$ in (3) and taking into account that $f^{(r)}(0) = r!P(Y = r)$, we obtain (2). From (4), also by substituting $z = 0$, we obtain the estimate

$$\left|\sum_{s=t}^{\infty} \frac{M^{(s+r)}(Y)}{s!}(-1)^s\right| \leq \frac{M^{(t+r)}(Y)}{t!}. \tag{5}$$

In proving Theorem 1, we may assume that $k < \log_2 n + 1$. Indeed,

$$P(\xi_{n,k} = 0) = P(\pi\{X_\alpha = 0\}) = 1 - P(U\{X_\alpha = 1\}) >$$

$$> 1 - \sum P(X_\alpha = 1) = 1 - m \cdot 2^{-2k} = 1 - \lambda(n, k). \tag{6}$$

Therefore, if $k \geq \log_2 n + 1$, then $P(\xi_{n,k} = 0) \to 1$, $e^{-\lambda} \to 1$, since obviously $\lambda \to 0$. Consequently, for any $r \geq 1$

$$P(\xi_{n,k} = r) \to 0, \qquad \frac{\lambda^r}{r!} e^{-\lambda} \to 0.$$

It is clear that, in view of (2) and (5), to prove Theorem 1 it is sufficient to show that for every $s = 0, 1, 2, \ldots$

$$M^{(s)}(\xi_{n,k}) = \lambda^s + o(1), \tag{7}$$

if n and k increase as indicated in the hypothesis of the theorem.

It is easy to verify that if $0 \leq s \leq m$, then

$$\sum X_\alpha \left(\sum X_\alpha - 1\right) \ldots \left(\sum X_\alpha - s + 1\right) = \sum X_{\alpha_1} X_{\alpha_2} \ldots X_{\alpha_s},$$

where the sum on the right is taken over all ordered sets of distinct k-intervals. Hence

$$M^{(s)}(\xi_{n,k}) = \sum M\left(X_{\alpha_1} X_{\alpha_2} \ldots X_{\alpha_s}\right) \quad (0 \leq s \leq m). \tag{8}$$

Before proving (7), let us introduce a definition. We shall call a set of intervals in E_n connected if, for any two intervals α and β from this set, there exists a sequence $\alpha_1, \alpha_2, \ldots, \alpha_p$ of intervals, also belonging to this set, such that $\alpha_1 = \alpha$, $\alpha_p = \beta$, and $\alpha_{i-1} \cap \alpha_i \neq \emptyset$ $(1 < i \leq p)$. It is not difficult to show that for a given connected t-element set of k-intervals $\{\alpha_1, \alpha_2, \ldots, \alpha_t\}$ there -

there exists a suitable interval β of length kl such that $\beta \supset \alpha_i$ $(1 \le i \le t)$. Hence it follows that there exist no more than

$$\binom{kt}{t}^t \binom{n}{kt} 2^{n-kt} \quad (< n^{2kt} \cdot 2^n)$$

connected sets consisting of t intervals of length k. An arbitrary set a of k-intervals in E_n decomposes into connected components (i.e., into maximal connected subsets). Denote by $[a]_i$ the number of all connected components of the set a that consist of i elements. If $\langle t_1, t_2, \dots, t_s \rangle$ is an arbitrary set of nonnegative integers such that

$$\sum_{i=1}^{s} i t_i = s, \tag{9}$$

then, obviously, there exist no more than

$$m^{t_1} n^{2k(t_2+\cdots+t_s)} \cdot 2^{n(t_2+\cdots+t_s)} = m^{t_1} \cdot 2^{(n+2k\log_2 n)(t_2+\cdots+t_s)} \tag{10}$$

distinct s-element sets a of k-intervals in which $[a]_i = t_i$ $(1 \le i \le s)$. We split the sum in (8) into two partial sums, assigning to the first partial sum (Σ_1) all terms of (8) that correspond to s-tuples of pairwise nonintersecting k-intervals, and to the second partial sum (Σ_2) all the remaining terms of (8). For a given k-interval there exist exactly

$$\sum_{i=0}^{k-1} \binom{n-k}{i} \binom{n}{i} 2^i \quad (< (2nk)^k)$$

distinct k-intervals intersecting it. Therefore the number of terms in the sum Σ_1 (coinciding with the number of all ordered s-tuples of pairwise nonintersecting k-intervals) is less than m^s, but greater than

$$m\{m-(2nk)^k\} \cdots \{m-(s-1)(2nk)^k\} \quad (> m^s - s^2(2nk)^k m^{s-1}).$$

Hence, and from the fact that

$$M(X_{\alpha_1} X_{\alpha_2} \cdots X_{\alpha_s}) = MX_{\alpha_1} MX_{\alpha_2} \cdots MX_{\alpha_s} = 2^{-s \cdot 2^k}$$

for every term of Σ_1, the following estimate for the sum Σ_1 follows:

$$\lambda^s - s^2(2nk)^k \cdot 2^{-2^k} \lambda^{s-1} < \Sigma_1 < \lambda^s. \tag{11}$$

We shall now show that the sum Σ_2 is infinitely small when n and k increase as indicated in the condition of Theorem 1. Let $a = \{\alpha_1, \alpha_2, \dots, \alpha_s\}$ be an arbitrary set of k-intervals. Then

$$M(X_{\alpha_1} X_{\alpha_2} \cdots X_{\alpha_s}) \le 2^{-t_1 \cdot 2^k - 3 \cdot 2^{k-1}(t_2+\cdots+t_s)}, \tag{12}$$

where $t_i = [a]_i$. Indeed, if the intervals α and β belong to different connected components of the set a, then they have no common vertices and, consequently,

$M(X_\alpha X_\beta) = MX_\alpha \cdot MX_\beta$. On the other hand, for any k-intervals α and β one always has

$$M(X_\alpha X_\beta) \le 2^{-3\cdot 2^{k-1}},$$

since the union of these intervals contains at least

$$2^k + 2^k - 2^{k-1} = 3 \cdot 2^{k-1}$$

vertices. Therefore the mathematical expectation of the product of random variables X_α corresponding to distinct k-intervals α from any connected set does not exceed $2^{-3\cdot 2^{k-1}}$, provided only that this product contains at least two factors. Hence (12) follows.

From (10) and (12) it follows that

$$\sum_{[a]_i = t_i} M(X_{\alpha_1} X_{\alpha_2} \cdots X_{\alpha_s}) \le \lambda^{t_1} \cdot 2^{(n + k\log_2 n - 3\cdot 2^{k-1})(t_2 + \cdots + t_s)}. \tag{13}$$

(Here the sum is taken over all ordered s-tuples $\langle \alpha_1, \alpha_2, \dots, \alpha_s \rangle$ of distinct intervals of length k, and such that $[a]_i = t_i$, where $a = \{\alpha_1, \alpha_2, \dots, \alpha_s\}$.) We show that the sum (13) tends to zero if the set a contains at least one pair of intersecting intervals. Indeed, in this case $t_1 \le s - 2$ and $t_2 + \cdots + t_s > 0$, and it is sufficient

but show that $n + 2k\log_2 n - 3 \cdot 2^{k-1} \to -\infty$. Thanks to the assumption that $k < \log_2 n + 1$, and to the fact that $n - k - 2^k < \log_2 \lambda$, we have

$$\begin{aligned} n + 2k\log_2 n - 3\cdot 2^{k-1} &= n + k(\log_2 n + 3/2) + {}^3/2(-k - 2^k) < \\ &< n + (\log_2 n + 1)(2\log_2 n + {}^3/2) - {}^3/2 n + {}^3/2 \log_2 \lambda \to -\infty. \end{aligned} \tag{14}$$

Obviously,

$$\Sigma_2 = \sum_{[a]_j = t_j} \sum M(X_{a_1} X_{a_2} \dots X_{a_s}),$$

where the outer summation on the right-hand side is taken over all solutions $\langle t_1, t_2, \dots, t_s \rangle$ of equation (9) in nonnegative integers, moreover such that $t_1 \le s - 2$. For fixed s there are, obviously, only finitely many such solutions. (More precisely, their number is not greater than $s^s/s!$.) Hence it follows that Σ_2 tends to zero. (More precisely, by (13) and (14) we have, starting from some n:

$$\Sigma_2 < \frac{s^s}{s!} \lambda^{s-2} \cdot 2^{-(1/2 - \varepsilon)n},$$

where ε is any positive number.) Together with (11) this gives (7). Theorem 1 is proved.

Remark 3. From the proof of Theorem 1 there essentially follows the following estimate of the order of approximation of the distribution of the random variable $\xi_{n,k}$ by the Poisson distribution:

$$P(\xi_{n,k} = r) - \frac{\lambda^r}{r!}e^{-\lambda} = O\left(2^{-(1/2-\varepsilon)n}\right),$$

where ε is any fixed positive number.

Passing to the proof of Theorem 2, note that

$$M\xi_n = \sum_{k=1}^{n} P(\xi_n \geq k) =$$

$$= \sum_{k=1}^{\chi(n)-2} P(\xi_n \geq k) + P(\xi_n \geq \chi(n)-1) + P(\xi_n \geq \chi(n)) + \sum_{k=\chi(n)+1}^{n} P(\xi_n \geq k).$$

Since, obviously, $\xi_n < k$ if and only if $\xi_{n,k} = 0$, by Theorem 1 we have

$$M\xi_n = \chi(n) - \sum_{k=1}^{\chi(n)-2} P(\xi_{n,k} = 0) - e^{-\lambda(n,\chi(n)-1)} - e^{-\lambda(n,\chi(n))} +$$
$$+ \sum_{k=\chi(n)+1}^{n} P(\xi_n \geq k) + o(1). \tag{15}$$

But, by (6),

$$\sum_{k=\chi(n)+1}^{n} P(\xi_n \geq k) = \sum_{k=\chi(n)+1}^{n} \{1 - P(\xi_{n,k} = 0)\} <$$
$$< \sum_{k=\chi(n)+1}^{n} \lambda(n,k) < n\lambda(n,\chi(n)+1) = o(1). \tag{16}$$

On the other hand, since in E_n there exists a system of 2^{n-k} pairwise nonintersecting intervals of length k, we have

$$P(\xi_{n,k} = 0) < (1 - 2^{-2^k})^{2^{n-k}}.$$

Using the last estimate, it is easy to show that

$$\sum_{k=1}^{\chi(n)-2} P(\xi_{n,k} = 0) = o(1). \tag{17}$$

Theorem 2 follows from (15), (16), and (17).

Institute of Mathematics
Siberian Branch of the Academy of Sciences of the USSR

Received
27 XII 1967

REFERENCES

1. S. V. Yablonskii, *Tr. Mat. Inst. im. V. A. Steklova AN SSSR*, **51**, 5 (1958).

Note: Figure translations are in progress. See original paper for figures.