ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIII, 1 (1972)

B. KOPOCIŃSKI (Wrocław)

ON THE NUMBER OF LOCAL EXTREMA IN A SEQUENCE OF INDEPENDENT RANDOM VARIABLES

0. The problem. Let $X_0, X_1, \ldots$ be independent, continuous, equally distributed random variables. Let us define the sequence of random variables $Z_1, Z_2, \ldots$

$$(1)\qquad Z_k = \begin{cases} 1 & \text{if } \dfrac{X_{k+1}-X_k}{X_k-X_{k-1}} < 0, \quad k = 1, 2, \ldots, \\ 0 & \text{otherwise,} \end{cases}$$

and the sequence of random variables $S_1, S_2, \ldots$

$$(2)\qquad S_n = \sum_{k=1}^{n} Z_k, \quad n = 1, 2, \ldots$$

We may regard the random variables $X_0, X_1, \ldots$ as successive values of an investigated characteristic, whereas sequence (1) is defined so that the random variable Z_k takes the value 1 if in the sequence $X_0, X_1, \ldots$ there is a local extremum (maximum or minimum) at place k, and Z_k is zero otherwise. The random variable S_n is defined as the number of local extrema in the sequence $X_0, X_1, \ldots, X_{n+1}$.

Thus S_n may be used as a statistics for testing stability of the characteristic in question. A deficiency in the number of local extrema may suggest the existence of a systematic trend while their excess could be explained as an effect of a servomechanism striving to maintain a stabilization. To construct a statistical test it is necesary first to calculate the probability distribution of S_n. This was suggested by M. Krzanowski from the Institute of Zoology in Cieszyn who wanted to test in this way the stability of some biochemical characteristic.

The first two moments of S_n were given in the unpublished diploma work by Stępniak [1], written under the direction of J. Łukaszewicz. There was also given an argument for the limiting distribution of S_n while $n \to \infty$. In the present paper, along with former results, there is also presented a method of calculating the exact probability distribution of S_n. For $n = 1, 2, \ldots, 20$, the distributions of S_n are tabulated at the end of the paper.

1. The distribution of S_n. To find the distribution of the random variable S_n ($n = 1, 2, \ldots$), let us consider the sequence of triplets of random variables

$$(S_n, Z_{n+1}, Z_{n+2}), \quad n = 1, 2, \ldots \tag{3}$$

Using (1) and (2), it is easy to see that sequence (3) is a Markov chain. This property will enable us to find the probability distribution of the states of the chain, and then the probability distribution of its first component.

Denote by

$$P_{ijk}^{(n)} = \Pr(S_n = i, Z_{n+1} = j, Z_{n+2} = k), \quad i = 0, 1, \ldots, n, \quad j, k = 0, 1, \tag{4}$$

the probability distribution of (3). Introduce also the conditional probabilities

$$P_{ij.k} = \Pr(Z_{n+3} = k \mid Z_{n+1} = i, Z_{n+2} = j), \quad i, j, k = 0, 1, \tag{5}$$

which of course do not depend on n.

Using this notation, we may write the obvious recurrent relations

$$P_{ijk}^{(n+1)} = P_{i0j}^{(n)} P_{0j.k} + P_{i-1,1j}^{(n)} P_{1j.k}, \quad i = 0, 1, \ldots, n+1, \quad j, k = 0, 1, \quad n = 1, 2, \ldots, \tag{6}$$

where $P_{-1,1j}^{(n)} = 0$ and $P_{n+1,0j}^{(n)} = 0$ for $j = 0, 1$.

To complete these formulas we have still to find

$$P_{ijk}^{(1)} = P_{ijk} = \Pr(Z_1 = i, \ Z_2 = j, \ Z_3 = k), \quad i, j, k = 0, 1, \tag{7}$$

and also the conditional probabilities (5).

Consider the realizations x_0, x_1, x_2, x_3, x_4 of the random variables X_0, X_1, X_2, X_3, X_4, and let these numbers be pairwise different (the continuity assumption of the random variables enables, without loss of generality, the restriction to such realizations only). We consider now all $5! = 120$ permutations of these numbers and also the values z_1, z_2, z_3 which are realizations of the random variables Z_1, Z_2, Z_3 as calculated for permutations.

The assumptions of independence and equidistribution of X_0, X_1, X_2, X_3, X_4 leads us to the conclusion that every permutation has a probability of occurrence in the conditional distribution defined on the considered permutation set. From this it follows directly that the frequences of the possible permutations of the numbers z_1, z_2, z_3 are equal to the sought probabilities in the distribution of Z_1, Z_2, Z_3.

The random variables Z_1, Z_2, Z_3 take on the values 000, 001, 010, 011, 100, 101, 110, 111, and the probabilities of those values (found as frequencies in the above-mentioned manner) are

$$(8)\qquad \begin{aligned} P_{000} &= 2/120, & P_{100} &= 8/120,\\ P_{001} &= 8/120, & P_{101} &= 22/120,\\ P_{010} &= 12/120, & P_{110} &= 18/120,\\ P_{011} &= 18/120, & P_{111} &= 32/120. \end{aligned}$$

Now probabilities (5) may be calculated. Thus, from

$$(9)\qquad P_{ij.k} = \Pr(Z_3 = k | Z_1 = i,\ Z_2 = j) = \frac{P_{ijk}}{P_{ij0} + q_{ij1}}, \qquad i, j, k = 0, 1,$$

we have

$$(10)\qquad \begin{aligned} P_{00.0} &= 0,2, & P_{00.1} &= 0,8,\\ P_{01.0} &= 0,4, & P_{01.1} &= 0,6,\\ P_{10.0} &= 0,2(6), & P_{10.1} &= 0,7(3),\\ P_{11.0} &= 0,36, & P_{11.1} &= 0,64. \end{aligned}$$

The recurrent formulas (6) and probabilities (8) and (10) enable us to find the probability distributions (4) of the triples (S_n, Z_{n+1}, Z_{n+2}). Now we find the distribution of random variable S_n as the marginal distribution

$$(11)\qquad \Pr(S_n = i) = \sum_{j=0}^{1} \sum_{k=0}^{1} P_{ijk}^{(n)}, \qquad i = 0, 1, \ldots, n.$$

Tables 1 and 2 contain these distributions and their cummulative distribution functions for $n = 1, 2, \ldots, 20$.

2. Mean and variance of S_n. From the distributions $P_{ijk} = \Pr(Z_1 = i, Z_2 = j, Z_3 = k)$, $i, j, k = 0, 1$, we may easily find the distributions: of the random variable Z_1 (or Z_2 or Z_3);

$$P_0 = \Pr(Z_1 = 0) = 1/3, \quad P_1 = \Pr(Z_1 = 1) = 2/3;$$

of the pair of consecutive random variables Z_1 and Z_2 (or Z_2 and Z_3);

$$P_{00} = \Pr(Z_1 = 0, Z_2 = 0) = 1/12, \quad P_{01} = \Pr(Z_1 = 0, Z_2 = 1) = 3/12,$$
$$P_{10} = \Pr(Z_1 = 1, Z_2 = 0) = 3/12, \quad P_{11} = \Pr(Z_1 = 1, Z_2 = 1) = 5/12;$$

and of the pair of non-consecutive random variables Z_1 and Z_3;

$$P'_{00} = \Pr(Z_1 = 0, Z_3 = 0) = 4/120, \quad P'_{01} = \Pr(Z_1 = 0, Z_3 = 1) = 26/120,$$
$$P'_{10} = \Pr(Z_1 = 1, Z_3 = 0) = 26/120, \quad P'_{11} = \Pr(Z_1 = 1, Z_3 = 1) = 54/120.$$

We have thus

$$\mathrm{E}(Z_1) = 2/3, \quad \mathrm{D}^2(Z_1) = 2/9,$$
$$\mathrm{cov}(Z_1, Z_2) = -1/36, \quad \mathrm{cov}(Z_1, Z_3) = 1/180.$$

It follows from (1) that for $n \geqslant 4$ the random variables Z_1 and Z_n are independent, hence it is easy to calculate that

$$\mathrm{E}(S_n) = \sum_{i=1}^{n} \mathrm{E}(Z_i) = n\mathrm{E}(Z_1) = \frac{2}{3}n, \quad n = 1, 2, \ldots, \tag{12}$$

$$\begin{aligned} \mathrm{D}^2(S_n) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{cov}(Z_i, Z_j) \\ &= n\mathrm{D}^2(Z_1) + 2(n-1)\mathrm{cov}(Z_1, Z_2) + 2(n-2)\mathrm{cov}(Z_1, Z_3) \\ &= \frac{2}{9}n - \frac{2}{36}(n-1) + \frac{2}{180}(n-2), \quad n = 1, 2, \ldots, \end{aligned}$$

thus

$$\mathrm{D}^2(S_n) = \frac{4}{45}n + \frac{1}{30}, \quad n = 1, 2, \ldots \tag{13}$$

3. Limit behaviour. It follows from the central limit theorem that the random variables S_n have, for $n \to \infty$, an asymptotically normal distribution with mean (12) and variance (13). The last columns of Tables 1 and 2 show that already for $n = 20$ this estimation suffices (e.g. for testing hypotheses with significance levels usually accepted).

Reference

[1] C. Stępniak, *Testowanie stabilności cechy biologicznej na przykładzie zawartości kwasu DNA w jądrach komórkowych* (unpublished diploma work), Wrocław University, 1969.

MATHEMATICAL INSTITUTE
UNIVERSITY OF WROCŁAW

Received on 22. 9. 1971

TABLE 1. The probabilities $\Pr(S_n = i)$

1	2	3	4	5	6	7
0.33333	0.08333	0.01667	0.00333	0.00067	0.00013	0.00003
.66667	.50000	.23333	.08000	.02311	.00604	.00149
	.41667	.48333	.33400	.16658	.06684	.02305
		.26667	.41200	.37074	.24069	.12440
			.17067	.32968	.36315	.28787
				.10923	.25324	.32932
					.06991	.18910
						.04474

8	9	10	11	12	13	14
0.00036	0.00008	0.00002	0.00000	0.00000	0.00000	0.00000
.00715	.00206	.00056	.00015	.00004	.00001	.00000
.05423	.02078	.00721	.00232	.00070	.00020	.00006
.18109	.09551	.04383	.01799	.00675	.00235	.00077
.30696	.22652	.13975	.07470	.03549	.01530	.00607
.28324	.30277	.25565	.18013	.10970	.05919	.02884
.13833	.23435	.28218	.26778	.21172	.14444	.08715
.02863	.09960	.18824	.25186	.26502	.23199	.17502
	.01833	.07083	.14769	.21727	.25082	.24057
		.01173	.04987	.11370	.18231	.22896
			.00751	.03482	.08616	.14952
				.00480	.02414	.06443
					.00307	.01664
						.00197

15	16	17	18	19	20	20 Normal approximation
0.00002	0.00000	0.00000	0.00000	0.00000	0.00000	0.0000
.00024	.00007	.00002	.00001	.00000	.00000	.0000
.00225	.00079	.00026	.00008	.00003	.00001	.0000
.01289	.00534	.00208	.00076	.00027	.00009	.0002
.04736	.02352	.01080	.00464	.00187	.00072	.0008
.11661	.06986	.03818	.01925	.00904	.00399	.0044
.19863	.14468	.09465	.05640	.03096	.01580	.0161
.23857	.21379	.16885	.11968	.07718	.04579	.0458
.20284	.22803	.22025	.18736	.14291	.09910	.0992
.12028	.17524	.21131	.21873	.19924	.16264	.1635
.04764	.09519	.14818	.19071	.21051	.20434	.2051
.01141	.03489	.07426	.12300	.16823	.19719	.1860
.00126	.00779	.02533	.05722	.10045	.14547	.1526
	.00081	.00530	.01825	.04362	.08088	.0790
		.00052	.00359	.01306	.03292	.0334
			.00033	.00242	.00930	.0107
				.00021	.00163	.0026
					.00014	.0006

TABLE 2. The cummulative distribution function $\sum_{j=0}^{i} \Pr(S_n = j)$

i \ n	1	2	3	4	5	6	7
0	0.33333	0.08333	0.01667	0.00333	0.00067	0.00013	0.00003
1	1.00000	.58333	.25000	.08333	.02311	.00618	.00152
2		1.00000	.73333	.41733	.19036	.07302	.02457
3			1.00000	.82933	.56109	.31371	.14897
4				1.00000	.89077	.67686	.43683
5					1.00000	.93009	.76616
6						1.00000	.95526
7							1.00000

i \ n	8	9	10	11	12	13	14
1	0.00036	0.00008	0.00002	0.00000	0.00000	0.00000	0.00000
2	.00751	.00214	.00058	.00015	.00004	.00001	.00000
3	.06175	.02292	.00779	.00247	.00074	.00021	.00006
4	.24284	.11843	.05162	.02046	.00749	.00257	.00083
5	.54980	.34495	.19138	.09516	.04298	.01786	.00691
6	.83304	.64772	.44702	.27529	.15268	.07705	.03575
7	.97137	.88207	.72920	.54307	.36440	.22150	.12290
8	1.00000	.98167	.91744	.79493	.62941	.45349	.29792
9		1.00000	.98827	.94263	.84668	.70431	.53848
10			1.00000	.99249	.96038	.88662	.76744
11				1.00000	.99520	.97278	.91696
12					1.00000	.99693	.98139
13						1.00000	.99803
14							1.00000

i \ n	15	16	17	18	19	20	20 Normal approximation
3	0.00002	0.00000	0.00000	0.00000	0.00000	0.00000	0.0000
4	.00026	.00008	.00002	.00001	.00000	.00000	.0000
5	.00251	.00086	.00028	.00009	.00003	.00001	.0000
6	.01539	.00621	.00236	.00085	.00030	.00010	.0002
7	.06275	.02973	.01317	.00549	.00217	.00082	.0010
8	.17936	.09959	.05134	.02474	.01121	.00480	.0054
9	.37799	.24427	.14600	.08114	.04217	.02060	.0215
10	.61656	.45805	.31485	.20082	.11935	.06639	.0673
11	.81940	.68609	.53510	.38817	.26225	.16550	.1665
12	.93969	.86133	.74641	.60690	.46150	.32813	.3300
13	.98733	.95652	.89460	.79761	.67200	.53248	.5351
14	.99874	.99140	.96886	.92061	.84023	.72966	.7211
15	1.00000	.99919	.99419	.97783	.94068	.87514	.8737
16		1.00000	.99948	.99608	.98430	.95601	.9527
17			1.00000	.99967	.99736	.98894	.9861
18				1.00000	.99979	.99823	.9968
19					1.00000	.99986	.9994
20						1.00000	1.0000

B. KOPOCIŃSKI (Wrocław)

O LICZBIE LOKALNYCH EKSTREMÓW W CIĄGU NIEZALEŻNYCH ZMIENNYCH LOSOWYCH

STRESZCZENIE

Niech $X_0, X_1, \ldots, X_{n+1}$ będą niezależnymi zmiennymi losowymi ciągłymi o tym samym rozkładzie prawdopodobieństwa. Liczbę ekstremów lokalnych w tym ciągu określamy jako zmienną losową

$$S_n = \sum_{k=1}^{n} Z_k,$$

gdzie

$$Z_k = \begin{cases} 1 & \text{jeżeli } \dfrac{X_{k+1}-X_k}{X_k-X_{k-1}} < 0, \quad k = 1, 2, \ldots, n, \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

W pracy znaleziono wzory rekurencyjne pozwalające obliczyć rozkład zmiennej losowej S_n, wzory na średnią i wariancję w tym rozkładzie oraz oszacowanie asymptotyczne rozkładu. Znajdują się w niej także tablice badanego rozkładu (dla $n = 1, 2, \ldots$ $\ldots, 20$).