

ON THE DISTRIBUTION LAW OF THE NUMBER OF RUNS IN A HOMOGENEOUS MARKOV CHAIN

V. K. Zakharov, O. V. Sarmanov

1968

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196801.69236>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

UDC 519.21

MATHEMATICS

V. K. ZAKHAROV, O. V. SARMANOV

ON THE DISTRIBUTION LAW OF THE NUMBER OF RUNS IN A HOMOGENEOUS MARKOV CHAIN

(Presented by Academician S. N. Bernstein on 23 V 1967)

Consider a homogeneous chain with number of states $s + 1$; p_i are the initial probabilities; p_{ij} are the transition probabilities.

In a chain of n trial outcomes, let m_i denote the number of occurrences of the i -th state, and let m_{ij} denote the number of transitions from the i -th to the j -th state, $i, j = 1, 2, \dots, s + 1$. Then

$$\xi_i = m_i - m_{ii} \quad (1)$$

is the number of runs of the i -th state in the chain.

Since, according to (1), the joint distribution law of m_i and m_{ij} is asymptotically normal, the joint distribution law of ξ_i is also asymptotically normal.

Exact distribution laws for the numbers of runs can be found by means of combinatorial formulas given in ^(1, 2); however, in the general case these expressions, as well as the expressions for the parameters of the normal law, are very cumbersome.

In this paper we consider a number of cases in which the exact laws and their parameters have a simple form.

1. Let $s = 1$ (a two-state chain), and let

$$\begin{pmatrix} \alpha & 1 - \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

be the matrix of transition probabilities; then the exact distribution law of the number of runs ξ_1 has the form

$$P(\xi_1 = l) = P_n^{(2)}(l) = \sum_m C_{m-1}^{l-1} C_{n-m-1}^{l-1} \alpha^{m-l} (1 - \alpha)^{l-1} \beta^{l-1} (1 - \beta)^{n-m-l-1} \times$$

$$\times \left[p_1(1-\alpha)(1-\beta) + (1-p_1)\beta(1-\beta) + p_1 \frac{(l-1)(1-\beta)^2}{n-m-l+1} + (1-p_1) \frac{(n-m-l)\beta(1-\alpha)}{l} \right], \quad (2)$$

$$l = 0, 1, \dots, \left[\frac{n+1}{2} \right],$$

and the mean and variance are equal to

$$M\xi_1 = (n-1)(1-\alpha)P + P + (p_1 - P) [Q + P(\alpha - \beta)^{n-1}], \quad (3)$$

$$\begin{aligned} D\xi_1 = & PQ\{(n-1)[\alpha\beta - (1-\alpha)(P-Q)] + Q - P(P-Q) \\ & + 2P^2(\alpha - \beta)^{n-1}\} + (p_1 - P)\{-2(n-1)PQ\beta(1 + \alpha - \beta)(\alpha - \beta)^{n-2} \\ & + P[3\alpha Q - (1 + 2\alpha - \beta)P](\alpha - \beta)^{n-2} + Q(P-Q)^2\} \\ & - (p_1 - P)^2 [Q + P(\alpha - \beta)^{n-1}]^2, \end{aligned} \quad (4)$$

where

$$P = \frac{\beta}{1-\alpha+\beta}, \quad Q = 1 - P = \frac{1-\alpha}{1-\alpha+\beta}. \quad (5)$$

If $\beta = 1 - \alpha$ (we shall call such a chain factorizable), then the sum (2) contracts and

$$\begin{aligned} P_n^{(2)}(l) = & C_{n-1}^{2l-1} \alpha^{n-2l} (1-\alpha)^{2l-1} + p_1 C_{n-1}^{2l-2} \alpha^{n-2l+1} (1-\alpha)^{2l-2} \\ & + (1-p_1) C_{n-1}^{2l} \alpha^{n-2l-1} (1-\alpha)^{2l}. \end{aligned} \quad (2')$$

2. Let us dwell in more detail on the general case of a factorizable chain, in which the transition probabilities have the form

$$p_{ii} = \alpha, \quad p_{ij} = (1-\alpha)/s, \quad i \neq j, \quad 0 \leq \alpha \leq 1. \quad (6)$$

We shall derive the distribution law of the total number of runs

$$\eta = \sum_{i=1}^{s+1} \xi_i.$$

The number of ways in which γ runs can be arranged among $s + 1$ states, so that a chain of length n begins with a run of any fixed state and two runs of the same kind are not adjacent, is $s^{\gamma-1}$; since γ runs from n trial results can be formed in $C_{n-1}^{\gamma-1}$ ways, the number of all chains of length n , beginning with one fixed state and containing exactly γ runs, is $C_{n-1}^{\gamma-1} s^{\gamma-1}$.

The probability of obtaining a particular chain containing exactly γ runs, under the condition that the first trial has the i -th outcome, is

$$p_i \alpha^{n-\gamma} \left(\frac{1-\alpha}{s} \right)^{\gamma-1},$$

therefore

$$P(\eta = \gamma) = \sum_{i=1}^{s+1} p_i \alpha^{n-\gamma} \left(\frac{1-\alpha}{s} \right)^{\gamma-1} C_{n-1}^{\gamma-1} s^{\gamma-1} = C_{n-1}^{\gamma-1} \alpha^{n-\gamma} (1-\alpha)^{\gamma-1}, \quad (7)$$

$$\gamma = 1, 2, \dots, n.$$

Thus, $\eta - 1$ has a binomial distribution.

In a chain with transition probabilities (6), the parameters of the exact distribution laws of the numbers of runs ξ_i have the form

$$\mathbf{M}\xi_i = \frac{1}{s+1} \left[(n-1)(1-\alpha) + 1 + \left(p_i - \frac{1}{s+1} \right) (s + a^{n-1}) \right], \quad (8)$$

$$\mathbf{D}\xi_i = \frac{1}{(s+1)^2} \left\{ (n-1)(1-\alpha) \left(\alpha + \frac{s(s-1)}{s+1} + s - \frac{2s}{(s+1)^2} + \frac{2s}{(s+1)^2} a^{n-1} \right) \right.$$

$$\left. - \left(p_i - \frac{1}{s+1} \right) \left[2(n-1)(1-\alpha) \left(1 + \alpha - \frac{1-\alpha}{s} \right) a^{n-2} - \frac{s(s-1)^2}{s+1} - \frac{(s-1)(1+3s)}{s+1} a^{n-1} \right] - \left(p_i - \frac{1}{s+1} \right)^2 \right\} \quad (9)$$

$$\begin{aligned} \text{cov } \xi_i \xi_j &= \frac{1}{(s+1)^2} \left\{ (n-1)(1-\alpha) \left(\alpha - \frac{s-1}{s+1} \right) + \frac{2}{(s+1)^2} - 1 - \frac{2}{(s+1)^2} a^{n-1} \right. \\ &+ \left(p_i + p_j - \frac{2}{s+1} \right) \left[(n-1)(1-\alpha) \frac{2-(s+1)\alpha}{s} a^{n-2} - \frac{s(s-1)}{s+1} - \frac{1+3s}{s+1} a^{n-1} \right] \\ &\left. - \left(p_i - \frac{1}{s+1} \right) \left(p_j - \frac{1}{s+1} \right) (s + a^{n-1})^2 \right\}, \end{aligned} \quad (10)$$

where

$$a = [(s+1)\alpha - 1]/s.$$

From (9) and (10) it follows that the correlation coefficient between ξ_i and ξ_j is equal to

$$R(\xi_i, \xi_j) = \frac{\alpha - (s-1)/(s+1)}{\alpha + s(s-1)/(s+1)} + O\left(\frac{1}{n}\right), \quad i \neq j, \quad s = 1, 2, \dots \quad (11)$$

Remark 1. For $s = 1$, $R(\xi_1, \xi_2) \approx 1$, since the numbers of runs of successes and failures differ by no more than one. For $s > 1$ and $\alpha = (s-1)/(s+1)$, the numbers of runs ξ_i and ξ_j , as (11) shows, are asymptotically independent.

Remark 2. If $p_i = \alpha = 1/(s+1)$, $i = 1, 2, \dots, s+1$, then the factorizable chain turns into a polynomial scheme with equal probabilities of all outcomes.

We indicate the exact joint distribution law of the numbers of runs ξ_1, ξ_2, ξ_3 for a three-valued ($s = 2$) factorizable Markov chain:

$$\begin{aligned} P(\xi_1 = l_1, \xi_2 = l_2, \xi_3 = l_3) &= C_{n-1}^{l_1+l_2+l_3-1} \alpha^{n-l_1-l_2-l_3} \left(\frac{1-\alpha}{2} \right)^{l_1+l_2+l_3-1} \times \\ &\times \sum_{\beta} \left\{ \left[(1-p_3) C_{2\beta-1}^{l_1+l_2+l_3-1} + (1+p_3) C_{2\beta-1}^{l_1+l_2-l_3} + 2p_3 C_{2\beta-1}^{l_1+l_2-l_3+1} \right] C_{l_1-1}^{\beta-1} C_{l_2-1}^{\beta-1} + \right. \\ &+ \left[p_2 C_{2\beta}^{l_1+l_2-l_3-1} + (1-p_1) C_{2\beta}^{l_1+l_2-l_3} + p_3 C_{2\beta}^{l_1+l_2-l_3+1} \right] C_{l_1-1}^{\beta-1} C_{l_2-1}^{\beta} + \\ &\left. + \left[p_1 C_{2\beta}^{l_1+l_2-l_3-1} + (1-p_2) C_{2\beta}^{l_1+l_2-l_3} + p_3 C_{2\beta}^{l_1+l_2-l_3+1} \right] C_{l_1-1}^{\beta} C_{l_2-1}^{\beta-1} \right\} \end{aligned} \quad (12)$$

From (12) one obtains the following one-dimensional distribution law for the number of runs ξ_i :

$$P(\xi_i = k) = \sum_{\gamma=2k-1}^n \alpha^{n-\gamma} \left(\frac{1-\alpha}{2} \right)^{\gamma-1} C_{n-1}^{\gamma-1} \left[(1-p_i)2^k C_{\gamma-1-k}^k + (1+p_i)2^{k-1} C_{\gamma-1-k}^{k-1} + 2p_i 2^{k-2} C_{\gamma-1-k}^{k-2} \right], \quad (13)$$

$$i = 1, 2, 3, \quad k = 0, 1, \dots, \left\lceil \frac{n+1}{2} \right\rceil.$$

3. Let us also consider the polynomial scheme, i.e., put

$$p_{ii} = p_i, \quad p_{ij} = p_j.$$

In this case the parameters of the exact distribution laws for the numbers of runs take the form

$$\mathbf{M}\xi_i = (n-1)p_i(1-p_i) + p_i, \quad (14)$$

$$\mathbf{D}\xi_i = (n-1)p_i(1-p_i)(1-3p_i+3p_i^2) + p_i(1-p_i)(1-2p_i^2), \quad (15)$$

$$\text{cov } \xi_i \xi_j = (n-1)p_i p_j (2p_i + 2p_j - 1 - 3p_i p_j) + 2p_i^2 p_j^2 - p_i p_j, \quad (16)$$

$$i \neq j, \quad i, j = 1, 2, \dots, s+1.$$

The correlation coefficient is equal to

$$R(\xi_i, \xi_j) = \frac{2p_i + 2p_j - 1 - 3p_i p_j}{\sqrt{(1-3p_i+3p_i^2)(1-3p_j+3p_j^2)}} \sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}} + O\left(\frac{1}{n}\right). \quad (17)$$

We also give the expression for the parameters of the distribution law of the total number of runs

$$\eta = \sum_{i=1}^{s+1} \xi_i :$$

$$\mathbf{M}\eta = (n-1)(1-s_2) + 1, \quad (18)$$

$$\mathbf{D}\eta = (n-1)(s_2 + 2s_3 - 3s_2^2) + 2(s_2^2 - s_3), \quad (19)$$

where

$$s_2 = \sum_{i=1}^{s+1} p_i^2, \quad s_3 = \sum_{i=1}^{s+1} p_i^3. \quad (20)$$

Mathematical Institute named after V. A. Steklov
Academy of Sciences of the USSR

Received
14 V 1967

REFERENCES

1. N. V. Smirnov, O. V. Sarmanov, V. K. Zakharov, DAN, 167, No. 6, 1238 (1966).
2. O. V. Sarmanov, V. K. Zakharov, DAN, 176, No. 3, 530 (1967).

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

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.