

MOMENTS OF RANDOM ALLOCATION PROCESSES REACHING A BOUNDARY

G. S. TSITSIASHVILI,* *Russian Academy of Sciences*

Abstract

In this paper we develop some results presented by Gani (2004), deriving moments for random allocation processes. These moments correspond to the allocation processes reaching some domain boundary. Exact formulae for means, variances, and probability generating functions as well as some asymptotic formulae for moments of random allocation processes are obtained. A special choice of the asymptotics and of the domain allows us to reduce a complicated numerical procedure to a simple asymptotic one.

Keywords: Infective; susceptible; allocation process; urn scheme; law of large numbers; domain boundary

2000 Mathematics Subject Classification: Primary 60G20

Secondary 92C60

1. Introduction

In this paper we develop some aspects of [1]; this article is devoted mainly to an analysis of infective numbers by means of probability generating functions. Here, the infective numbers are represented as a random allocation process in an urn scheme. Such an analysis allows us to consider a detailed one-dimensional model with a single group of susceptibles or a single kind of infection. However, a multidimensional model with a few groups of susceptibles or a few kinds of infections becomes too complicated both numerically and analytically. So the problem is to find a convenient approach to the multidimensional model analysis.

This article is based on the concept of moments for random allocation processes reaching some domain boundary. Analogously to [2], these moments are represented as finite sums of independent random variables with geometric distributions. In the one-dimensional case this representation gives exact formulae for means, variances, and probability generating functions as well as some asymptotic formulae for the moments of random allocation processes when they reach the domain boundary. In the multidimensional model a special numerical algorithm is constructed which allows us to calculate these means and variances, but in a somewhat complicated way. So the special asymptotics and the domain are chosen to reduce the complicated multidimensional numerical problem to a simple asymptotic one-dimensional problem based on the law of large numbers.

2. One-dimensional model: mean, variance, and characteristic function

Suppose that n denotes the number of susceptibles, s denotes the number of infectives, and that $h_s(n)$ denotes the probability that a new infection reaches one of the infectives, and so does not increase the number of infectives (in the simplest case $h_s(n) = s/n$), and

Received 11 August 2005; revision received 24 September 2007.

* Postal address: Institute of Applied Mathematics, Far Eastern Branch, Russian Academy of Sciences, Radio str. 7, 690041 Vladivostok, Russia. Email address: guram@iam.dvo.ru

let $g_s(n) = 1 - h_s(n)$. Let $Y_i(n)$ denote the number of infectives in an initial group of susceptibles at the moment i (when the i th infection arrives), $Y_0(n) = 0$, and let $T_{l,n} = \min(i : Y_i(n) = l)$. We write

$$M_{l,n} = E T_{l,n}, \quad V_{l,n} = \text{var } T_{l,n}, \quad \Phi_{l,n}(z) = E z^{T_{l,n}},$$

and note that

$$T_{l,n} = \sum_{k=0}^{l-1} \tau_{k,n},$$

where $\tau_{k,n}$, $k \geq 0$, are independent random variables with the geometric distribution parameter $g_k(n)$: $\tau_{k,n} \rightarrow G(g_k(n))$. Then it is easy to prove that

$$M_{l,n} = \sum_{k=0}^{l-1} E \tau_{k,n} = \sum_{k=0}^{l-1} \frac{1}{g_k(n)}, \quad V_{l,n} = \sum_{k=0}^{l-1} \text{var } \tau_{k,n} = \sum_{k=0}^{l-1} \frac{h_k(n)}{g_k^2(n)}, \tag{1}$$

$$\Phi_{l,n}(z) = E \prod_{k=0}^{l-1} z^{\tau_{k,n}} = \prod_{k=0}^{l-1} E z^{\tau_{k,n}} = \prod_{k=0}^{l-1} \frac{g_k(n)(z)}{1 - h_k(n)(z)}. \tag{2}$$

If $h_s(n) = s/n$ then from (1)–(2) we obtain

$$M_{l,n} = \sum_{k=0}^{l-1} \frac{1}{1 - k/n}, \quad V_{l,n} = \sum_{k=0}^{l-1} \frac{k/n}{(1 - k/n)^2}, \quad \Phi_{l,n}(z) = \prod_{k=0}^{l-1} \frac{kz/n}{1 - z(1 - k/n)}. \tag{3}$$

Now consider (3) for $n \rightarrow \infty$.

Suppose that $l = o(n)$; from (3) we obtain

$$M_{l,n} \sim l, \quad V_{l,n} \sim \frac{l^2}{2n}, \quad \frac{V_{l,n}}{M_{l,n}^2} \sim \frac{1}{2n} \rightarrow 0. \tag{4}$$

Now suppose that $q = n - l \geq 1$ (where $q = o(n)$ as $n \rightarrow \infty$). Then

$$M_{l,n} \sim n \ln n, \tag{5}$$

and, for some positive and finite numbers $C_1 \leq C_2$, we have

$$\frac{C_1 n^2}{q} \leq V_{l,n} \leq \frac{C_2 n^2}{q} \sim n \ln n \implies \frac{C_1}{q \ln^2 n} \leq \frac{V_{l,n}}{M_{l,n}^2} \leq \frac{C_2}{q \ln^2 n} \rightarrow 0. \tag{6}$$

Suppose that $l = cn$, $0 < c < 1$, then

$$M_{l,n} \sim n \ln \frac{1}{1 - c}, \quad V_{l,n} \sim n \left(\frac{c}{1 - c} + \ln(1 - c) \right), \quad \frac{V_{l,n}}{M_{l,n}^2} \sim \frac{c/(1 - c) + \ln(1 - c)}{n \ln^2(1 - c)} \rightarrow 0. \tag{7}$$

Consequently, from the asymptotics (4), (6), and (7), we have

$$\frac{T_{l,n}}{M_{l,n}} \xrightarrow{P} 1 \quad \text{as } n \rightarrow \infty, \tag{8}$$

where $M_{l,n}$ satisfies (4), (5), and (7). Here ‘ \xrightarrow{P} ’ denotes convergence in probability.

3. Multidimensional model with one-sided boundary

Model 1. Suppose that the initial group of n susceptibles consists of subgroups with sizes n_1, \dots, n_m , where $n_1 + \dots + n_m = n$. Each new infection arises in these subgroups with the probabilities p_1, \dots, p_m , where $p_1 + \dots + p_m = 1$. Set $\mathbf{n} = (n_1, \dots, n_m)$, and let $Y_{i,n} = (Y_{i,n}^j, j = 1, \dots, m)$ be the vector of infective numbers in these subgroups at the moment i . We define

$$T_{l,j,n} = \min(i : Y_{i,n}^j = l),$$

and writing

$$M_{l,j,n} = E T_{l,j,n}, \quad V_{l,j,n} = \text{var } T_{l,j,n}, \quad \Phi_{l,j,n}(z) = E z^{T_{l,j,n}},$$

then we obtain

$$T_{l,j,n} = \sum_{k=0}^{l-1} \tau_{k,j,n},$$

where $\tau_{k,j,n}, k \geq 0$, are independent random variables with the geometric distribution parameter $p_j g_k(n_j)$: $\tau_{k,j,n} \rightarrow G(p_j g_k(n_j))$. By analogy with (1)–(2), it is easy to prove that

$$\begin{aligned} M_{l,j,n} &= \sum_{k=0}^{l-1} \frac{1}{p_j g_k(n_j)}, & V_{l,j,n} &= \sum_{k=0}^{l-1} \frac{1 - p_j g_k(n_j)}{[p_j g_k(n_j)]^2}, \\ \Phi_{l,j,n}(z) &= \prod_{k=0}^{l-1} \frac{p_j g_k(n_j)(z)}{1 - z(1 - p_j g_k(n_j))}. \end{aligned} \tag{9}$$

So, if, for some $j, 1 \leq j \leq m$ (where $l_j = o(n_j)$ as $n_j \rightarrow \infty$) and n_i is a constant, where $1 \leq i \neq j \leq m$, then by analogy with (8), we obtain

$$\frac{T_{l,j,n}}{M_{l,j,n}} \xrightarrow{P} 1. \tag{10}$$

Model 2. Now suppose that the initial group of n susceptibles is uniform but that there are m types of infections arising with the probabilities p_1, \dots, p_m , where $p_1 + \dots + p_m = 1$. Denote by $Z_{i,n} = (Z_{i,n}^j, j = 1, \dots, m)$ the vector of the infective numbers, and set

$$R_{l,n}^j = \min(i : Z_{i,n}^j = l),$$

$$M_{l,n}^j = E R_{l,n}^j, \quad V_{l,n}^j = \text{var } R_{l,n}^j, \quad \Phi_{l,n}^j(z) = E z^{R_{l,n}^j},$$

then we obtain

$$R_{l,n}^j = \sum_{k=0}^{l-1} \alpha_{k,n}^j,$$

where $\alpha_{k,n}^j, k \geq 0$, are independent random variables with the geometric distribution parameter $p_j g_k(n)$: $\alpha_{k,n}^j \rightarrow G(p_j g_k(n))$. Then, by analogy with (9), it is easy to prove that

$$\begin{aligned} M_{k,n}^j &= \sum_{k=0}^{l-1} \frac{1}{p_j g_k(n)}, & V_{k,n}^j &= \sum_{k=0}^{l-1} \frac{1 - p_j g_k(n)}{[p_j g_k(n)]^2}, \\ \Phi_{k,n}^j(z) &= \prod_{k=0}^{l-1} \frac{p_j g_k(n)(z)}{1 - z(1 - p_j g_k(n))}. \end{aligned}$$

So, if, for some j , $1 \leq j \leq m$ (where $l_j = o(n)$ as $n \rightarrow \infty$), then similarly to (10), we obtain

$$\frac{R_{l,n}^j}{M_{l,n}^j} \xrightarrow{P} 1.$$

4. Numerical algorithms in the multidimensional model with multisided boundary

Without loss of generality, consider Model 2 and introduce the set

$$\Gamma = \bigcap_{j=1}^m (0 \leq x_j < l_j).$$

For simplicity suppose that $m = 2$, denote by $R_{\Gamma,n}(x_1, x_2)$ the first moment when the chain $Z_{i,n}$ exits from Γ provided that $Z_{0,n} = (x_1, x_2)$, and set $R_{\Gamma,n} = R_{\Gamma,n}(0, 0)$. Now, if $(x_1, x_2) \in \Gamma$, it is easy to show that, for $M_{\Gamma,n}(x_1, x_2) = E R_{\Gamma,n}(x_1, x_2)$,

$$\begin{aligned} & (p_1 g_{x_1}^1(n) + p_2 g_{x_2}^2(n)) M_{\Gamma,n}(x_1, x_2) \\ &= 1 + p_1 g_{x_1}^1(n) M_{\Gamma,n}(x_1 + 1, x_2) + p_2 g_{x_2}^2(n) M_{\Gamma,n}(x_1, x_2 + 1). \end{aligned} \tag{11}$$

This recurrent formula is to be complemented by the following boundary conditions:

$$M_{\Gamma,n}(x_1, l_2) = 0, \quad 0 \leq x_1 < l_1, \quad M_{\Gamma,n}(l_1, x_2) = 0, \quad 0 \leq x_2 < l_2. \tag{12}$$

So, to find $M_{\Gamma,n}(0, 0) = E R_{\Gamma,n}$, it is necessary to use (11) and the boundary conditions, (12), sequentially at the points

$$\begin{aligned} (x_1, x_2) = & (l_1 - 1, l_2 - 1), (l_1 - 2, l_2 - 1), \dots, (0, l_2 - 1), \\ & (l_1 - 1, l_2 - 2), (l_1 - 2, l_2 - 2), \dots, (0, l_2 - 2), \end{aligned}$$

and so on. As a result, it is possible to obtain $E R_{\Gamma,n}$. This algorithm may also apply to calculations of $E R_{\Gamma,n}^2$, $\text{var } R_{\Gamma,n}$, etc. Therefore, the calculation of $E R_{\Gamma,n}$ and $\text{var } R_{\Gamma,n}$ requires a complicated numerical procedure, and the problem is how to simplify this procedure in some asymptotic cases.

5. Some asymptotics in the multidimensional model with multisided boundary

In this section we consider multidimensional Models 1 and 2 with multisided boundaries. In Model 1 suppose that $n_1/n \rightarrow r_1, \dots, n_m/n \rightarrow r_m$, $n_1 + \dots + n_m = n$, where r_1, \dots, r_m , $r_1 + \dots + r_m = 1$, are fixed. Then it is possible to write $Y_{i,n} = Y_{i,n}$ and $T_{\Gamma,n} = \min(i : Y_{i,n} \notin \Gamma)$, where

$$T_{\Gamma,n} = \min(T_{l_j,j,n}, 1 \leq j \leq m). \tag{13}$$

In Model 2 let $R_{\Gamma,n} = \min(Z_{i,n} \notin \Gamma)$. Then we obtain

$$R_{\Gamma,n} = \min(R_{l_j,n}^j, 1 \leq j \leq m).$$

Theorem 1. *Suppose that, for $n \rightarrow \infty$ and $1 \leq j \leq m$,*

$$l_j \rightarrow \infty, \quad l_j = o(n), \quad 1 \geq g_k(n) \geq 1 - \frac{k}{n}, \quad 0 \leq k \leq n. \tag{14}$$

Then, for $n \rightarrow \infty$, we have

$$\frac{T_{\Gamma,n}}{F} \xrightarrow{p} 1, \quad \frac{R_{\Gamma,n}}{F} \xrightarrow{p} 1, \tag{15}$$

where

$$F = \min\left(\frac{l_j}{p_j}, 1 \leq j \leq m\right).$$

Proof. We introduce the random variables

$$\tau_{k,j,n}^+ \rightarrow G\left(p_j\left(1 - \frac{l_j}{n_j}\right)\right), \quad \tau_{k,j,n}^- \rightarrow G(p_j), \tag{16}$$

so that the random vectors $(\tau_{k,j,n}^-, \tau_{k,j,n}, \tau_{k,j,n}^+)$, $k \geq 0$, are independent and

$$\tau_{k,j,n}^- \leq \tau_{k,j,n} \leq \tau_{k,j,n}^+, \quad 0 \leq k, n, 1 \leq j \leq m, \tag{17}$$

almost surely. Let

$$n_\varepsilon = \max\left(\frac{q_j}{p_j^2(1 - l_j/n_j)^2\varepsilon^3}, \frac{l_j}{r_j\varepsilon}, 1 \leq j \leq m\right), \quad T_{l,j,n}^\pm = \sum_{k=0}^{l-1} \tau_{k,j,n}^\pm,$$

then from Chebyshev's inequality and (16), we obtain

$$P\left(\left|\frac{T_{l,j,n}^-}{l_j} - \frac{1}{p_j}\right| > \varepsilon\right) < \frac{q_j}{p_j^2 l_j \varepsilon^2}.$$

So, from condition (14), for $n > n_\varepsilon$ and $1 \leq j \leq m$,

$$P\left(\frac{l_j(1 - \varepsilon)}{p_j} < T_{l,j,n}^- < \frac{l_j(1 + \varepsilon)}{p_j} < \frac{l_j(1 + \varepsilon)}{p_j(1 - \varepsilon)}\right) > 1 - \varepsilon, \tag{18}$$

and analogously

$$P\left(\frac{l_j(1 - \varepsilon)}{p_j} < T_{l,j,n}^+ < \frac{l_j(1 + \varepsilon)}{p_j(1 - l_j/n_j)} < \frac{l_j(1 + \varepsilon)}{p_j(1 - \varepsilon)}\right) > 1 - \varepsilon. \tag{19}$$

Now, (18) and (19) give the inequality

$$P\left(\frac{l_j(1 - \varepsilon)}{p_j} < T_{l,j,n} < \frac{l_j(1 + \varepsilon)}{p_j(1 - \varepsilon)}, 1 \leq j \leq m\right) > 1 - 2m\varepsilon,$$

and consequently, from (13) and (17), we obtain

$$P\left(F(1 - \varepsilon) < T_{\Gamma,n} < \frac{F(1 + \varepsilon)}{(1 - \varepsilon)}\right) > 1 - 2m\varepsilon.$$

The first part of (15) is proved; the proof of the second part is similar. It is necessary to introduce only the random variables

$$\alpha_{k,j,n}^+ \rightarrow G\left(p_j\left(1 - \frac{l_j}{n}\right)\right), \quad \alpha_{k,j,n}^- \rightarrow G(p_j),$$

so that the random vectors $(\alpha_{k,j,n}^-, \alpha_{k,j,n}, \alpha_{k,j,n}^+)$, $k \geq 0$, are independent and

$$\alpha_{k,j,n}^- \leq \alpha_{k,j,n} \leq \alpha_{k,j,n}^+, \quad 0 \leq k, n, 1 \leq j \leq m,$$

almost surely.

Remark 1. The statement of Theorem 1 may be generalized to the case

$$\Gamma = \bigcap_{k \in K} \bigcup_{j \in J_k} (0 \leq x_j < l_{j,k}),$$

where

$$T_{\Gamma,n} = \min_{k \in K} \max_{j \in J_k} T_{l_{j,k},j,n}, \quad F = \min_{k \in K} \max_{j \in J_k} \frac{l_{j,k}}{p_j}.$$

Acknowledgements

The author would like to thank Professor Joe Gani for useful advice. This work was supported by the Russian Foundation for Basic Research, project 03-01-00512-a.

References

- [1] GANI, J. (2004). Random-allocation and urn models. In *Stochastic Methods and Their Applications* (J. Appl. Prob. Spec. Vol. **41A**), Applied Probability Trust, Sheffield, pp. 313–320.
- [2] TSITSIASHVILI, G. S. (1995). Transformation of an epidemic model to a random walk and its management. *Math. Scientist* **20**, 103–106.