ON APPLICATION OF THE SUFFICIENT EMPIRICAL AVERAGING METHOD TO SYSTEM SIMULATION

Alexander Andronov and Catherine Zhukovskaya
Faculty of Transport and Mechanical Engineering
Riga Technical University
Kalku 1, LV-1658, Riga, Latvia
E-mail: lora@mailbox.riga.lv, kat_zuk@hotmail.com

Evgeniy V. Chepurin
Faculty of Mechanics & Mathematics
Moscow State University
Vorobyevie Gori, 119899 Moscow, Russia
E-mail: chepurin@nw.math.msu.su

KEYWORDS
Estimation, complete sufficient statistic, gamma-distribution

ABSTRACT
A new approach to preparing the input data in simulation is being considered – so called Sufficient Empirical Averaging method. It assumes the existence of the complete sufficient statistics for unknown parameters of input variable distributions. Its application allows getting unbiased estimates with minimum variance.

INTRODUCTION

Many problems of reliability, queueing theory, inventory control, insurance and so on may be described by this informal outline. The considered random process $\boldsymbol{Y}(t)$ has k components: $\boldsymbol{Y}(t) = (Y_1(t), Y_2(t), ..., Y_k(t))$. Its realization has been determined completely by mutually independent sequences $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_m$, which are called generating sequences. Each generating sequence is an infinite sequence of independent identically distributed random variables (i.i.d.r.v.): $\boldsymbol{X}_i = (\boldsymbol{X}_{i,1}, \boldsymbol{X}_{i,2}, ...)$:

$$\boldsymbol{Y}(t) = \boldsymbol{Y}(t, \boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_m).$$

There is known m-dimensional function $\boldsymbol{N}(t, \boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_m)$, which i-th component $\boldsymbol{N}_i(t, \boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_m)$ gives a number of sequence $\boldsymbol{X}_i$ random variables from which depends $\boldsymbol{Y}(\tau)$ till time moment t. In particular $\boldsymbol{Y}(t)$ value for time moment t is determined by the finite part of generating sequence:

$$\boldsymbol{Y}(t) = \boldsymbol{Y}(t, \boldsymbol{X}_{i,1}, \boldsymbol{X}_{i,2}, ..., \boldsymbol{X}_{i,N_i(t)} : i = 1, ..., m).$$

Furthermore there is known some real function of the considered process $\boldsymbol{G}(t, \boldsymbol{Y}(\tau) : 0 < \tau < t)$. Our aim is to estimate its expectation $\boldsymbol{g}(t) = \mathrm{E}\, \boldsymbol{G}(t, \boldsymbol{Y}(\tau) : 0 < \tau < t)$. With that the distributions of the generating sequences are unknown, but we have samples of random variables from these sequences.

There are three approaches of these samples used for $\boldsymbol{g}(t)$ estimation. All these suppose a simulation of the random process $\boldsymbol{Y}(t)$. In the traditional approach the "plug-in" method is usually used (Gentle 2002). This approach is parametric and supposes that the type of each distribution is known (but not parameters). The unknown parameters are estimated from given samples $\{\boldsymbol{X}_i\}$. The received estimators are used instead of the unknown parameters. On the base of such estimated distributions, these random variables are generated during the simulation. This approach has the following defects: it gives a biase of output indices if a size of sample is small or a type of some sample has been determined mistakenly.

If the resampling approach is applied, then observations of the input data are used themselves, in various combinations, as realizations of input variables in simulation runs (Andronov and Merkuryev 2002). This method is nonparametric and doesn't use information about the type of distributions. It decreases an efficiency of simulation results if such information is available. In this case the *Sufficient Empirical Averaging* (SEA) method should be used, that has been proposed by Chepurin (1994, 1995, 1999). Under some conditions this parametric method gives unbiased estimators of $\boldsymbol{g}(t)$ with minimum variance. Let us describe one formally.

We will use the same notations for the sample elements as for the elements of the generating sequences: $\boldsymbol{X}_i = (\boldsymbol{X}_{i,1}, \boldsymbol{X}_{i,2}, ..., \boldsymbol{X}_{i,n_i})$ for the i-th sample, where n_i is equal to sample size of the i-th sequence (further n_i-sample). We assume that the family $\{f_i(\boldsymbol{x}; \boldsymbol{\theta}_i)\}$ of probability density function for each generating sequence is known and admits the sufficient statistic for unknown parameters $\boldsymbol{\theta}_i$. We denote one by $\boldsymbol{S}_i$ for $\boldsymbol{\theta}_i$, that has been calculated on the base of $\boldsymbol{X}_i = (\boldsymbol{X}_{i,1}, \boldsymbol{X}_{i,2}, ..., \boldsymbol{X}_{i,n_i})$, $i = 1, 2, ..., m$. Let us denote $h_i(\boldsymbol{s}; n_i, \boldsymbol{\theta}_i)$ its probability density function (p.d.f.).

Below we show how it can get an unbiased estimator of $\boldsymbol{g}(t)$ with uniformly minimum variance.

In the next two Sections the SEA approach will be presented. The Section 4 considers important gamma-distribution. The last section contains numerical example.

Proceedings 12th International Conference ASMTA 2005
Khalid Al-Begain, Gunter Bolch, Miklos Telek © ECMS, 2005
ISBN 1-84233-112-4 (Set) / ISBN 1-84233-113-2 (CD)

SUFFICIENT EMPIRICAL AVERAGING METHOD

It is well known that conditional distribution of sample elements, given a fixed value of sufficient statistic, doesn't depend on the unknown parameter (Lehman 1959, Cox and Hinkley 1974). Let us denote $\boldsymbol{X_i}(\boldsymbol{s_i}) = \boldsymbol{X_i} \mid \{\boldsymbol{S_i} = \boldsymbol{s_i}\}$ the *i*-th conditional generating sequence, that has been calculated on condition $\{\boldsymbol{S_i} = \boldsymbol{s_i}\}$. Note that $\boldsymbol{X_i}(\boldsymbol{s_i})$ "... is statistically equivalent to the data ($\boldsymbol{X_i}$), i.e. containing the same amount of information ..." (Chepurin 1999, p.182). Let $q_i(\cdot;\, n_i,\, \boldsymbol{s_i})$ be the notation for p.d.f. of $\boldsymbol{X_i}(\boldsymbol{s_i})$. $\boldsymbol{X_i}^* = \boldsymbol{X_i}(\boldsymbol{s_i})$ is called *the data variant for the i-th generating sequence*, but the sequence of these variants $\boldsymbol{X_i}^*(1), \boldsymbol{X_i}^*(2), ..., \boldsymbol{X_i}^*(B)$, where $\boldsymbol{X_i}^*(j) = (\boldsymbol{X}_{i,1}^*(j), \boldsymbol{X}_{i,2}^*(j), ..., \boldsymbol{X}_{i,n}^*(j))$, with common p.d.f. $q_i(\cdot;\, n_i,\, \boldsymbol{s_i})$ – *the sample of data variants for the i-th generating sequence*. The number of repetitions (runs) B may be arbitrary here.

From the sufficient statistics properties the following theorem may be drawn.

Theorem 1. The unconditional distribution (in respect to $h_i(\boldsymbol{s};\, n_i,\, \boldsymbol{\theta_i})$-measure) of each data variant for the *i*-th generating sequence $\boldsymbol{X_i}^*(j) = (\boldsymbol{X}_{i,1}^*(j), \boldsymbol{X}_{i,2}^*(j), ..., \boldsymbol{X}_{i,n}^*(j))$ coincides with the distribution of the initial sample $\boldsymbol{X_i} = (\boldsymbol{X}_{i,1}, \boldsymbol{X}_{i,2}, ..., \boldsymbol{X}_{i,n})$, i.e. it is the sequence of i.i.d.r.v. with common p.d.f. $f_i(\boldsymbol{x};\, \boldsymbol{\theta_i})$.

If we consider all generating sequences then let $\boldsymbol{S} = (\boldsymbol{S}_1, \boldsymbol{S}_2, ..., \boldsymbol{S_m})$ be total sufficient statistic, $q(\cdot;\, n,\, \boldsymbol{s}) = q_1(\cdot;\, n_1,\, \boldsymbol{s}_1) \times q_2(\cdot;\, n_2,\, \boldsymbol{s}_2) \times ... \times q_m(\cdot;\, n_m,\, \boldsymbol{s_m})$. We call $\boldsymbol{X}^* = (\boldsymbol{X_1}^*, \boldsymbol{X_2}^*, ..., \boldsymbol{X_m}^*)$ with p.d.f. $q(\cdot;\, n,\, \boldsymbol{s})$ *the data variant* and the sequence of $\boldsymbol{X}^*(1), \boldsymbol{X}^*(2), ..., \boldsymbol{X}^*(B)$ with common p.d.f. $q(\cdot;\, n,\, \boldsymbol{s})$ – *the sample of data variants*. Obviously the distribution of each data variant $\boldsymbol{X}^* = (\boldsymbol{X_1}^*, \boldsymbol{X_2}^*, ..., \boldsymbol{X_m}^*)$ coincides with the distribution of the initial samples $\boldsymbol{X_1}, \boldsymbol{X_2}, ..., \boldsymbol{X_m}$. Simultaneously all variants of the sample $\boldsymbol{X}^*(1), \boldsymbol{X}^*(2), ..., \boldsymbol{X}^*(B)$ are mutually independent on condition $\boldsymbol{S} = \boldsymbol{s}$.

Note that the last property allows us to use the data variants $\boldsymbol{X}^*(1), \boldsymbol{X}^*(2), ..., \boldsymbol{X}^*(B)$ as copies of the generating sequences $(\boldsymbol{X_1}, \boldsymbol{X_2}, ..., \boldsymbol{X_m})$ and get the unbiased estimators that are any amount of close to the unbiased estimator of $\boldsymbol{g}(t) = \mathrm{E}\, \boldsymbol{G}(t,\, \boldsymbol{Y}(\tau){:}\ 0 < \tau < t)$ with uniformly minimum variance. For this we assume this condition: the time moment t is chosen so that with the probability 1 simultaneously for all generating sequences (with current number i) number $N_i(t)$ of the random variables $\{X_{i,\cdot}\}$, that are necessary for $\boldsymbol{Y}(t)$ calculation does not greater than n_i. Now as the unbiased expectation of $\boldsymbol{g}(t) = \mathrm{E}\, \boldsymbol{G}(t,\, \boldsymbol{Y}(\tau){:}\ 0 < \tau < t)$ we use $\boldsymbol{Z}(t, \boldsymbol{X}) = \boldsymbol{G}(t,\ \boldsymbol{Y}(t, \boldsymbol{X}_{i,1}, \boldsymbol{X}_{i,2}, ..., \boldsymbol{X}_{i,N_i(t)}{:}\ 0 < \tau < t$, $i = 1, \ldots, m)$, that has been calculated on given samples.

It is well known that given $\boldsymbol{S} = \{\boldsymbol{S}_1, \boldsymbol{S}_2, ..., \boldsymbol{S_m}\}$ condition expectation $\boldsymbol{H^0}(\boldsymbol{S}) = E(\boldsymbol{Z}(t,\, \boldsymbol{X}) \mid \{\boldsymbol{S_1}, \boldsymbol{S_2}, ..., \boldsymbol{S_m}\})$ of unbiased estimator $\boldsymbol{Z}(t,\, \boldsymbol{X})$ is unbiased estimator with uniformly minimum variance. In it's turn this condition expectation may be estimated by arithmetical mean for B data variants $\boldsymbol{X}^*(1), \boldsymbol{X}^*(2), ..., \boldsymbol{X}^*(B)$:

$$\boldsymbol{H}(\boldsymbol{S}) = \mathrm{B}^{-1} \sum_{j=1}^{B} \boldsymbol{Z}(t, \boldsymbol{X}^*(j)). \qquad (1)$$

Theoretical reasons of such approach are given by the following theorem (Chepurin 1995).

Theorem 2. If $\boldsymbol{S} = \{\boldsymbol{S}_1, \boldsymbol{S}_2, ..., \boldsymbol{S_m}\}$ is the complete sufficient statistics and the variance of $\boldsymbol{Z}(t,\, \boldsymbol{X})$ is finite then:
(*i*) the estimate (1) is unbiased;
(*ii*) the variance of estimate (1) has the following presentation by means of variance $\mathrm{Var}(\boldsymbol{H^0}(\boldsymbol{S}))$ of $\boldsymbol{H^0}(\boldsymbol{S})$:

$$\mathrm{Var}(\boldsymbol{H}(\boldsymbol{S})) = (1 - B^{-1})\, \mathrm{Var}(\boldsymbol{H^0}(\boldsymbol{S})) + B^{-1}\mathrm{Var}(\boldsymbol{Z}(t,\, \boldsymbol{X})); \qquad (2)$$

(*iii*) $\boldsymbol{H}(\boldsymbol{S})$ is consistent estimate of $\boldsymbol{H^0}(\boldsymbol{S})$ in respect to $q(\cdot;\, n,\, \boldsymbol{s})$-measure:

$$\boldsymbol{H}(\boldsymbol{S}) = \boldsymbol{H^0}(\boldsymbol{S})) + \boldsymbol{0}_q(1) \text{ as } B \to \infty. \qquad (3)$$

SIMULATION OF DATA VARIANTS

Also to realize the above-described approach we must simulate data variants. Since the various generated sequences and corresponding variants are mutually independent, we may confine ourselves to one sequence. We will omit an index of this sequence and describe the initial sample as $\boldsymbol{X}_{\cdot} = (\boldsymbol{X}_{\cdot,1}, \boldsymbol{X}_{\cdot,2}, ..., \boldsymbol{X}_{\cdot,n})$. There are n copies of i.i.d.r.v. with common p.d.f. $f(\boldsymbol{x};\, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. According to our assumptions, the sufficient statistic $\boldsymbol{S}(n)$ for $\boldsymbol{\theta}$ exists with p.d.f. $h(\boldsymbol{s};\, n,\, \boldsymbol{\theta})$. Here n shows the size of the sample.

Conditional p.d.f. of the random sequence $\boldsymbol{X}_{\cdot}$ on condition $\{\boldsymbol{S} = \boldsymbol{s}\}$ has been denoted $q(\cdot;\, n,\, \boldsymbol{s})$. Obviously if $h(\boldsymbol{s};\, n,\, \boldsymbol{\theta}) \neq 0$ in the point $\boldsymbol{s}$, then

$$q(\boldsymbol{x};\, n,\, \boldsymbol{s}) = f(\boldsymbol{x}_1;\, \boldsymbol{\theta})\, f(\boldsymbol{x}_2;\, \boldsymbol{\theta}) \ldots \ldots f(\boldsymbol{x}_n;\, \boldsymbol{\theta}) \,/\, h(\boldsymbol{s}; n, \boldsymbol{\theta}), \qquad (4)$$

where $\boldsymbol{s} = \boldsymbol{s}(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n)$.

It is necessary to generate the data variant $\boldsymbol{X}^*_{\cdot} = (\boldsymbol{X}^*_{\cdot,1}, \boldsymbol{X}^*_{\cdot,2}, ..., \boldsymbol{X}^*_{\cdot,n})$ in accordance with this distribution. We do it recurrently. At first we generate one conditional random variable $\boldsymbol{X}_{\cdot,1}^* = \boldsymbol{X}_{\cdot,1} \mid \{\boldsymbol{S} = \boldsymbol{s}\}$ in accordance with p.d.f.

$$q^X(\boldsymbol{x}_1 | \boldsymbol{s};\, n) = f(\boldsymbol{x}_1;\, \boldsymbol{\theta})\, h(\boldsymbol{s}\hat{\ }(\boldsymbol{s}, \boldsymbol{x}_1);\, n - 1,\, \boldsymbol{\theta}) \,/\, h(\boldsymbol{s};\, n,\, \boldsymbol{\theta}), \quad (5)$$

where $\boldsymbol{s}\hat{\ }(\boldsymbol{s}, \boldsymbol{x}_1)$ means the sufficient statistics that has been calculated on the base of a $(n - 1)$-sample without $\boldsymbol{x}_1$, instead of complete n-sample.

Let us remind that $q^X(\boldsymbol{x}_1|\boldsymbol{s}; n)$ does not depend from the unknown parameter $\boldsymbol{\theta}$. Further we generate the second conditional random variable $\boldsymbol{X}_{\cdot,2}^* = \boldsymbol{X}_{\cdot,2} \mid \{\boldsymbol{S} = \boldsymbol{s}, \boldsymbol{X}_{\cdot,1}^* = \boldsymbol{x}_1\}$. By this we replace magnitude of the sufficient statistics $\boldsymbol{s}$ by $\boldsymbol{s}\hat{}(\boldsymbol{s}, \boldsymbol{x}_1)$, n by $n - 1$ and realize simulation as earlier. The continuation of this process gives $n - r$ conditional random variables $\boldsymbol{X}_{i,1}^*$, $\boldsymbol{X}_{i,2}^*$, ..., $\boldsymbol{X}_{i,n-r}^*$, where r is a minimal size of the sample that is necessary to calculate the sufficient statistics for $\boldsymbol{\theta}$. The last r conditional random variable $\boldsymbol{X}^*_{\cdot,n-r+1}$, $\boldsymbol{X}^*_{\cdot,n-r+2}$, ..., $\boldsymbol{X}^*_{\cdot,n}$ are determined as r magnitudes that give getting value of sufficient statistics $\boldsymbol{s}\hat{}(\boldsymbol{s}, \boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_{n-r})$. If we have several solutions, then randomization is used: each solution is chosen with equal probability (corresponding example will be considered below).

However it is usually very difficult to find the explicit form for p.d.f. $h(\boldsymbol{s}; n, \boldsymbol{\theta})$, since it is a distribution on hupersurfaces in space of high dimensions. On the other hand, the generation of the corresponding random variables is a complicated problem too. In this connection the two ways are possible. Firstly, it is quite possible to generate the random variables of interest directly, without knowledge of the according distribution. The corresponding examples were given in (Chepurin 1995, 1999, Engen and Lillegard 1997).

Secondly, we may decrease the dimension n using the following resampling procedure. Let $n = vd$, where v and d are natural numbers, $d \geq r$. Then we divide the sample $(\boldsymbol{X}_{\cdot,1}, \boldsymbol{X}_{\cdot,2}, ..., \boldsymbol{X}_{\cdot,n})$ from n elements into v subsamples, each from d elements. Applying the usual stochastic procedure (Kennedy and Gentle 1980), such partition may be performed. Now for each subsample we calculate sufficient statistic $\boldsymbol{s}'$ and generate d random variables with p.d.f. $q(\boldsymbol{x}; d, \boldsymbol{s}')$. Then we unify r getting group of the random variables to form a data variant $\boldsymbol{X}^*_{\cdot} = (\boldsymbol{X}^*_{\cdot,1}, \boldsymbol{X}^*_{\cdot,2}, ..., \boldsymbol{X}^*_{\cdot,n})$. We are able to reiterate this procedure with either previous or new partitions.

SPECIAL CASE: GAMMA-DISTRIBUTION

The *gamma-distribution* is considered because it includes exponential, Erlang and chi-square distributions, those have a main role in the reliability, queueing theory, inventory, insurance and so on. Gamma-distribution $G(\lambda, l)$ (Wilks 1944, Cox and Hinkley 1974, Lehman 1959) is determined by the probability density function

$$f(x) = \frac{\lambda}{\Gamma(l)}(\lambda x)^{l-1}e^{-\lambda x},\ x > 0\,, \tag{6}$$

where $\Gamma(l)$ is a gamma-function, $l, \lambda > 0$.
The three cases are possible: 1) l parameter is known, λ parameter is unknown ($\boldsymbol{\theta} = \lambda$); 2) λ parameter is known, l parameter is unknown ($\boldsymbol{\theta} = l$); 3) both parameters l and λ are unknown ($\boldsymbol{\theta} = (l, \lambda)$).
The first case is well known (Wilks 1944, Cox and Hinkley 1974). The sufficient statistics for λ on bases of sample $X_1, X_2, ..., X_n$ is the sum $T=X_1 + X_2 + ... + X_n$. It has gamma-distribution $G(\lambda, ln)$, therefore $h(\cdot; n, \lambda) = G(\lambda, ln)$. Joint conditional distribution of $k < n$ components by condition $T = t$ is Dirichle distribution $D(v_1, v_2, ..., v_k ; v_{k+1})$ for $v_1 = v_2 = ... = v_k = l$, $v_{k+1} = l(n - k)$:

$$q(x_1, x_2, ..., x_k, n, t) = t^{nl-k}\frac{\Gamma(nl)}{(\Gamma(l))^k\,\Gamma(l(n-k))}\cdot$$

$$\cdot\, x_1^{\nu_1-1}x_2^{\nu_2-1}...x_k^{\nu_k-1}(t - x_1 - x_2 - ... - x_k)^{\nu_{k+1}-1}\,,$$

$$0 < x_1, x_2, ..., x_k < t,\ x_1 + x_2 + ... + x_k < t.$$

If $k = 1$ then we have beta-distribution $Be(v_1, v_2)$ with p.d.f.

$$q(x_1, n, t) = \frac{1}{t^{\nu_1+\nu_2-1}}\frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)}\cdot$$

$$\cdot\, x_1^{\nu_1-1}(t - x_1)^{\nu_2-1},\ 0 < x_1 < t, \tag{7}$$

where $v_1 = l$, $v_2 = l(n\text{-}1)$.

Now to generate random variables with this distribution, we are able to apply usual methods, for example *Acceptance/Rejection* or *Inverse Cumulative Distribution Function* methods (Kennedy and Gentle 1980, Gentle 2002). But for the considered case it is more effectively to use the direct generation of the corresponding random variables (Chepurin 1999). Let $X^0_1, X^0_2, ..., X^0_n$ be a sequence of independent random variables with gamma-distribution $G(1, l)$. The generation technique of these variables is well known. After the generation, the sum $T^0 = X^0_1 + X^0_2 + ... + X^0_n$ should be calculated. Then the random variables of interest $X^*_1, X^*_2, ..., X^*_n$ may be calculated by formula

$$X^*_i = X^0_i\, T / T^0,\ i = 1, 2, ..., n. \tag{8}$$

Now we consider *the second case*. Here (Lehman 1959) the sufficient statistic for parameter l is a product $Z = \prod_{i=1}^{n} X_i$. Corresponding probability distribution function is given by the following theorem, the proof of which is similar to the Theorem 4.

Theorem 3. Let $X_1, X_2, ..., X_n$ be i.i.d.r.v. which have gamma-distribution $G(\lambda, l)$. Then their product $Z = \prod_{i=1}^{n} X_i$ has the following p.d.f. for $x \geq 0$:

$$h(z; n, l) = \left(\frac{\lambda^l}{\Gamma(l)}\right)^n z^{l-1}\int_0^\infty\int_0^\infty ... \int_0^\infty \frac{1}{\prod_{i=1}^{n-1} x_i}\cdot \tag{9}$$

$$\cdot \exp\left(-\lambda\left(x_1 + \sum_{i=1}^{n-2} \frac{x_{i+1}}{x_i} + \frac{z}{x_{n-1}}\right)\right) dx_1 dx_2 \ldots dx_{n-1}.$$

Conditional p.d.f. of r.v. X_i, $i = 1, 2, ..., n$, on condition $Z = z$ is determined by the formula

$$q^X(x|z,n) = \frac{1}{xh(z;n,l)} \cdot$$

$$\cdot f(x) h(\frac{z}{x}; n-1, l), \; x, z > 0. \quad (10)$$

In particular for $n = 2$ we have the formulas

$$h(z;2,l) = \left(\frac{\lambda^l}{\Gamma(l)}\right)^2 z^{l-1} \int_0^\infty \frac{1}{x} \cdot$$

$$\cdot \exp\left(-\lambda\left(x + \frac{z}{x}\right)\right) dx, \; z \ge 0, \quad (11)$$

$$q^X(x|z,n) =$$

$$= \frac{\exp\left(-\lambda\left(x + \frac{z}{x}\right)\right)}{x \int_0^\infty \frac{1}{u} \exp\left(-\lambda\left(u + \frac{z}{u}\right)\right) du}, \; x, z \ge 0. \quad (12)$$

In *the third case*, when both parameters are unknown i.e. $\boldsymbol{\theta} = (\lambda, l)$, a sufficient statistic for $\boldsymbol{\theta}$ is pair $\boldsymbol{S} = (T, Z)$. To formalize a necessary result we need to use the following Lemma, the proof of which is presented in the Appendix.

Lemma 1. If $z < t^3/27$ then the cubic equation

$$x^3 - 2tx^2 + t^2x - 4z = 0 \quad (13)$$

has three real roots

$$x_1(t,z) = \frac{2}{3} t \left[1 + \cos\left(\frac{1}{3}(\alpha + 2\pi)\right)\right],$$

$$x_2(t,z) = \frac{2}{3} t \left[1 + \cos\left(\frac{1}{3}(\alpha - 2\pi)\right)\right], \quad (14)$$

$$x_3(t,z) = \frac{2}{3} t \left[1 + \cos\left(\frac{1}{3}\alpha\right)\right],$$

where

$$\cos(\alpha) = 54 \frac{z}{t^3} - 1.$$

By this $0 \le x_1(t, z) \le x_2(t, z) \le t$, $x_1(t, z) + x_2(t, z) \le t$, $x_3(t, z) \ge t$, $x_1(t,z) + x_2(t,z) + x_3(t, z) = 2t$.

Using this Lemma we are able to present the joint distribution which is connected with the sufficient statistic $\boldsymbol{\theta} = (\lambda, l)$.

Theorem 4. Let $n \ge 2$ and $X_1, X_2, ..., X_n$ be i.i.d.r.v. which have gamma-distribution $G(\lambda, l)$. Then the joint p.d.f. of the sum $T = \sum_{i=1}^n X_i$ and the product $Z = \prod_{i=1}^n X_i$ is given by formulas (15) - (16) for $t, z > 0$:

for $n = 2$, $z \le \frac{1}{4} t^2$:

$$h(t,z,2) = \frac{2}{\sqrt{t^2 - 4z}} \frac{1}{(\Gamma(l))^2} \lambda^{2l} z^{l-1} e^{-\lambda t} \; ; \quad (15)$$

for $n > 2$, $z < \left(\frac{t}{n}\right)^n$:

$$h(t,z,n) = 2\left(\frac{\lambda^l}{\Gamma(l)}\right)^n z^{l-1} e^{-\lambda t} \int_{x_1^{(1)}}^{x_2^{(1)}} \int_{x_1^{(2)}}^{x_2^{(2)}} \ldots \quad (16)$$

$$\int_{x_1^{(n-2)}}^{x_2^{(n-2)}} \frac{dx_{n-2} \ldots dx_2 dx_1}{\sqrt{[x_1 \ldots x_{n-2}(t - x_1 - \ldots - x_{n-2})]^2 - 4zx_1 \ldots x_{n-2}}},$$

in particular for $n = 3$, $z \le \frac{1}{27} t^3$:

$$h(t,z,3) = 2\left(\frac{\lambda^l}{\Gamma(l)}\right)^3 z^{l-1} e^{-\lambda t}.$$

$$\cdot \int_{x_1(t,z)}^{x_2(t,z)} \frac{1}{\sqrt{u\left[u(t-u)^2 - 4z\right]}} du. \quad (17)$$

In these formulas we used the following notations for roots (14) of the cubic equation (13):

$$x_1^{(i)} = x_1\left(t - x_1 - \ldots - x_{i-1}, \frac{z}{x_1 \ldots x_{i-1}}\right),$$

$$x_2^{(i)} = x_2\left(t - x_1 - \ldots - x_{i-1}, \frac{z}{x_1 \ldots x_{i-1}}\right), \quad (18)$$

where for $i = 1$ $x_1 + \ldots + x_{i-1} = 0$, $x_1 \cdot \ldots \cdot x_{i-1} = 1$.

The conditional p.d.f. of r.v. X_i, $i = 1, 2, ..., n$, on condition $T = t$, $Z = z$, is determined by formulas (19) - (21):

for $n = 2$, $z < \frac{1}{4} t^2$: the conditional distribution of X_1 and X_2 is discrete (see formula (19)); two possible values are equal-probable:

$$P\left\{X_1=\frac{1}{2}\left(t-\sqrt{t^2-4z}\right),\ X_2=\frac{1}{2}\left(t+\sqrt{t^2-4z}\right)\right\}=P\left\{X_1=\frac{1}{2}\left(t+\sqrt{t^2-4z}\right),\ X_2=\frac{1}{2}\left(t-\sqrt{t^2-4z}\right)\right\}=\frac{1}{2}; \quad (19)$$

for $n > 2,\ z < \left(\frac{t}{n}\right)^n,\ x \geq 0$:

$$q^X(x|t,z,n)=\frac{\int_{x_1^{(1)}}^{x_2^{(1)}}\int_{x_1^{(2)}}^{x_2^{(2)}}\ldots\int_{x_1^{(n-3)}}^{x_2^{(n-3)}}\frac{dx_{n-3}\ldots dx_2dx}{\sqrt{[xx_{n-3}\ldots x_2x_1(t-x-x_{n-3}-\ldots-x_2-x_1)]^2-4zxx_{n-3}\ldots x_2x_1}}}{\int_{x_1^{(1)}}^{x_2^{(1)}}\int_{x_1^{(2)}}^{x_2^{(2)}}\ldots\int_{x_1^{(n-2)}}^{x_2^{(n-2)}}\frac{dx_{n-2}\ldots dx_2dx}{\sqrt{[x_{n-2}\ldots x_2x_1(t-x_{n-2}-x_{n-3}-\ldots-x_2-x_1)]^2-4zx_{n-2}\ldots x_2x_1}}}; \quad (20)$$

in particular, for $n = 3,\ z \leq \frac{1}{27}t^3,\ x \geq 0$:

$$g^X(x|t,z,3)=\frac{1}{\sqrt{[x(t-x)]^2-4zx}\int_{x_1(t,z)}^{x_2(t,z)}\frac{1}{\sqrt{[u(t-u)]^2-4zu}}du},\ x_1(t,z)<x<x_2(t,z); \quad (21)$$

NUMERICAL EXAMPLE

Let us consider queueing system with two servers. Interarrival times A_1, A_2, ... are i.i.d. random variables, which have shifted exponential distribution with unknown mean rate λ and known parameter of shift $\delta > 0$:

$$f_A(x)=\begin{cases}0, & x\leq\delta,\\ \lambda e^{-\lambda(x-\delta)}, & x\geq\delta.\end{cases}$$

The servers are different (nonhomogeneous). Service times are i.i.d. random variables having Erlang distributions $G(\mu, 2)$ for the first server and $G(v, 3)$ for the second server. Parameters μ and v are unknown. The customers are served one at a time, and if both servers are busy when a customer arrives, then customer must joint a queue (i.e., wait in line). From the queue customers go to server in accordance with their own arrival time. If both servers are empty when a customer arrives then customer goes to a server that has been cleared earlier. Let's find the nonstationary distribution of the number $Y(t)$ of customers in the system at the time moment $t > 0$:

$$P_j(t) = P\{Y(t) = j\},\quad j = 0, 1, \ldots .$$

We assume that originally the system was empty.
Let us remind that values λ, μ and υ are unknown but we have three samples of sizes n_0, n_1, n_2: interarrival times, service times of the first and the second servers. Let $t<n_0\delta$, $n_0 \leq n_1 \leq n_2$. We know that the sums of sample elements are complete sufficient statistics for the considered case. Let us denote they s_0, s_1, s_2.

Our example corresponds to a general outline of the Section 1. We have three generating sequences of interarrival times, service times of the first and the second servers. Random process $\boldsymbol{Y}(t)$ is univariate and means the number of customers in the system at the time moment t. The probabilities of interest $\{P_j(t)\}$ are expectations $\{q_j(t)\}$ of indicator functions $\{\chi_j(t)\}$ of events $\{Y(t) = j\}$: $\chi_j(t) = 1$ if event $\{Y(t) = j\}$ takes place and $\chi_j(t) = 0$ otherwise.

As usually these expectations are estimated by simulation of corresponding queueing process. The received results are given in the Table below. They correspond to the following input data: $n_0 = 30$, $n_1 = 22$, $n_2 = 22$, $s_0 = 20$, $s_1 = 30$, $s_2 = 40$, $\delta = 0.5$, $t = 8$, $B = 1000$. According to the above, presented estimates are close to unbiased estimates of $\{P_j(t)\}$with minimum variance.

Table 1: Minimum variance unbiased estimators
for the probabilities $\{P_j(t)\}$ of customer number in the system

	$t=0$	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$
$P_0(t)$		0.532	0.263	0.178	0.158	0.148	0.144	0.136	0.133
$P_1(t)$		0.468	0.517	0.472	0.444	0.423	0.411	0.401	0.377
$P_2(t)$			0.208	0.283	0.305	0.310	0.315	0.324	0.327
$P_3(t)$			0.012	0.065	0.084	0.101	0.104	0.107	0.116
$P_{>3}(t)$				0.002	0.009	0.018	0.026	0.032	0.047

REFERENCES

Andronov, A. and Yu. Merkuryev. 2002. "Use of a Resampling Approach to Systems Simulation." In *Proceedings of the 16th European Simulation Multiconference.* (Darmstadt, Germany, June 3-5). SPS, 150 - 155.

Chepurin, E.V. 1994. "The Statistical Methods in Theory of Reliability." *Obozrenije Prikladnoj i Promishlennoj Matematiki, Ser. Verojatnost i Statistika*, Vol.1, N 2, Moscow, 279 - 330. (In Russian.)

Chepurin, E.V. 1995. "The Statistical Analysis of the Gauss Data Based on the Sufficient Empiric Averaging Method." *Proceeding of the Russian University of People's Friendship. Series Applied Mathematics and Informatics*, N 1, Moscow, 112 - 125. (In Russian.)

Chepurin, E.V. 1999. "On Analitic-Computer Methods of Statistical Inferences of Small Size Data Samples." In Proceedings of the International Conference Probabilistic Analysis of Rare Events. (Riga, Latvia, June 28 – July 3), V.V. Kalashnikov and A.M. Andronov (Eds.). Riga Aviation University, Riga, 180 - 194.

Engen, S. and M. Lillegrad. 1997. Stochastic Simulations Conditioned of Sufficient Statistics. Biometrica, Vol 84., N1, 235-240.

Gentle, J.E. 2002. Elements of Computational Statistics. Springer-Verlag, New York – Berlin - Heidelberg.

Good, P.I. 2001. Resampling Methods: a Practical Guide to Data Analysis. Birkhäuser, Boston-Basel-Berlin.

Kennedy, W.J. and J.E. Gentle. 1980. Statistical Computing. Marcel Dekker, Inc., New York.

Lehman, E.L. 1983. Theory of Point Estimation. John Wiley and Sons, New York.

Rao, C.R. 1973. Linear Statistical Inferences and its Applications. Wiley, New York.

Srivastava, M. 2002. Methods of Multivariate Statistics. John Wiley&Sons, New York.

Wilks, S.S. 1944. Mathematical Statistics. Princeton University Press, New York.

APPENDIX

Proof of Lemma 1. Using Viète Theorem we verify that (14) gives roots of the cubic equation (13):

$$x_1(t,z) + x_2(t,z) + x_3(t,z) = -(-2t),$$

$$x_1(t,z)x_2(t,z) + x_1(t,z)x_3(t,z) + x_2(t,z)x_3(t,z) = t^2,$$

$$x_1(t,z)x_2(t,z)x_3(t,z) = -(-4z).$$

Further if $\alpha \in (0,\pi)$ then $\cos\left(\frac{1}{3}\alpha\right) \geq \frac{1}{2}$ and

$$x_3(t,z) \geq \frac{2}{3}t\left(1+\frac{1}{2}\right) \geq t.$$

Therefore $x_1(t,z) + x_2(t,z) \leq t$. Finally since $\sin\left(\frac{2}{3}\pi\right) = \frac{\sqrt{3}}{2}$, $\sin\left(\frac{1}{3}\alpha\right) > 0$ for $\alpha \in (0,\pi)$ then

$$x_1(t,z) - x_2(t,z) =$$

$$= \frac{2}{3}t\left[\cos\left(\frac{1}{3}(\alpha+2\pi)\right) - \cos\left(\frac{1}{3}(\alpha-2\pi)\right)\right] =$$

$$= -\frac{2}{3}t2\sin\left(\frac{1}{2}\frac{1}{3}2\alpha\right)\sin\left(\frac{1}{2}\frac{1}{3}4\pi\right) = -\frac{2}{\sqrt{3}}t\sin\left(\frac{1}{3}\alpha\right) \leq 0.$$

.

Proof of the Theorem 4. For the case $n = 2$ the joint distribution of X_1 and X_2 has the following density function for $x_1, x_2 \geq 0$:

$$f(x_1,x_2) = (x_1x_2)^{l-1}\left(\frac{\lambda^l}{\Gamma(l)}\right)^2 e^{-\lambda(x_1+x_2)}.$$

Let us consider the Jacobian of the transformation $t = x_1 + x_2$, $z = x_1x_2$:

$$J((x_1,x_2) \to (t,z)) = 1/\det{}_+\begin{pmatrix} \partial t/\partial x_1 & \partial t/\partial x_2 \\ \partial z/\partial x_1 & \partial z/\partial x_2 \end{pmatrix},$$

where $\det_+$ represents the absolute value of the determinant (Srivastava 2002).

We have

$$J((x_1,x_2)\to(t,z)) = 1/\det_+\begin{pmatrix}1 & 1\\ x_1 & x_2\end{pmatrix} = \frac{1}{|x_2-x_1|}.$$

But $t^2-4z=(x_1-x_2)^2$, therefore

$$|x_2-x_1| = \sqrt{t^2-4z},\quad z\le t^2/4.$$

The considered transformation $(x_1,x_2)\to(t,z)$ has two domains of one-to-one correspondence: for $x_1>x_2$ and for $x_1\le x_2$. Therefore the joint p.d.f. of T and Z is

$$h(t,z;2) = 2J((x_1,x_2)\to(t,z))f(x_1(t,z),x_2(t,z)) =$$

$$= 2\frac{1}{\sqrt{t^2-4z}}z^{l-1}\left(\frac{\lambda}{\Gamma(l)}\right)^2 e^{-\lambda t}, t>0,\ 0\le z\le\frac{1}{4}t^2.$$

Also the formula (15) is proved.

For the case $n > 2$ we use the mathematical induction:

$$h(t,z,n+1) = \int_{x_1^*}^{x_2^*}\lambda\frac{(\lambda x)^{l-1}}{\Gamma(l)}e^{-\lambda x}\cdot \tag{22}$$

$$\cdot h\left(t-x,\frac{z}{x},n\right)\frac{1}{x}dx,\quad n=2,3,...,$$

where $h\left(t-x,\frac{z}{x},n\right)$ is calculated by formula (16).

Here integration limits x_1^* and x_2^* are determined according to the conditions $x\le t$, $\frac{z}{x}<\left(\frac{t-x}{n}\right)^n$.

Let us change the order of integration so that the inner integration is performed according to x. Corresponding limits of integration are choosen so that magnitude under the radical in (16) for $t = t - x$, $z = z/x$ must be positive:

$$[x_1...x_{n-2}((t-x)-x_1-...-x_{n-2})]^2 \ge 4\frac{z}{x}x_1...x_{n-2}.$$

Setting $t' = t-x_1-...-x_{n-2}$, $z' = z/x_1...x_{n-2}$ we get the inequality

$$t'^2 - 2xt' + x^2 \ge 4z'.$$

It corresponds to the cubic equation (13). Therefore in accordance with the Lemma 1, the values of x must lie in the interval $\left(x_1^{(n-1)}, x_2^{(n-1)}\right)$ where (see formula (18)):

$$x_1^{(n-1)} = x_1\left(t-x_1-...-x_{n-2},\frac{z}{x_1...x_{n-2}}\right),$$

$$x_2^{(n-1)} = x_2\left(t-x_1-...-x_{n-2},\frac{z}{x_1...x_{n-2}}\right).$$

After a substitution (16) for $h\left(t-x,\frac{z}{x},n\right)$ in (22) and some simplifications, we get for $h(t, z, n+1)$ the expression (16).

Now we consider a conditional p.d.f. of r.v. X_i, $i = 1, 2, \ldots, n$, on condition $T = t$, $Z = z$. If $n = 2$ then the values of X_1 and X_2 are defined from equations $T = X_1 + X_2$, $Z = X_1X_2$. It gives two pairs of the possible values of X_1 and X_2 that are equal-probable (see formulas (19)). If $n > 2$ then we use the formula (4):

$$q(x;n,(t,z)) = \frac{f(x;(\lambda,l))}{h((t,z);n,(\lambda,l))}.$$

After a substitution (6) for $f(x; (\lambda, l))$ and (16) for $h((t, z); n, (\lambda, l))$ we get the formula (20).