The resampling approach application to complex systems reliability analysis and forecasting

Alexander Andronov
Dept. Mathematical Support of Transport System Management
Riga Technical University
Kalku 1, Riga, LV-1658
Latvia
Aleksandrs.Andronovs@rtu.lv

Abstract

Estimation of function expectation $\theta = Ef(X_1, X_2, ..., X_m)$ is considered. For independent variables $X_1, X_2, ..., X_m$, sample populations $H_1, H_2, ..., H_m$ are available as the primary data. The resampling approach uses the usual simulation procedure. For that, the random variables $\{X_i\}$ are not generated by random number generators in accordance with the estimated probabilistic distributions, but are extracted from the present samples at random. In the presented paper some theoretical results of the author for the case of small samples will be reviewed. Ones concern the hierarchical resampling, the controlled resampling, the interval estimation, the case of a dependent observations etc.

1 Introduction

Let us suppose that a considered task is to estimate an expectation θ of a function f of independent continuous random variables $X_1, X_2, ..., X_m$: $\theta= Ef(X_1, X_2, ..., X_m)$. This function describes reliability characteristic of a complex system. For independent variables $X_1, X_2, ..., X_m$ sample populations $H_1, H_2, ..., H_m$ are available as the primary data.
For $\theta = Ef(X_1, X_2, ..., X_m)$ estimation, the *resampling* approach uses the usual simulation procedure. For that, the random variables $\{X_i\}$ are not generated by random number generators in accordance with the estimated probabilistic distributions, but are extracted from the present samples at random.
It is possible to distinguish resampling methods varieties via two directions: the technique of a resampling application and the aim of a statistical procedure. In the first case, there exists a resampling with and without replacement. In the second case, the problems of point (as above) or interval estimation, hypothesis testing, classification and so on can be considered.
A practical usage of the resampling methods is very simple and intuitive. The book (*Good 2001*) can be used as an excellent manual here. On the other hand, a theoretical investigation of the resampling methods properties is a complex problem (*Efron and Tibshirani 1993, Davison and Hinkley 1997, Belyaev 2000*). In the presented paper some theoretical results of the author for the case of small samples will be reviewed. Ones concern the hierarchical resampling, the interval estimation, the controlled resampling, the case of a dependence in initial data etc.

2 Point estimation

In this case the resampling procedure includes k trials. During the ν-th trial, $\nu = 1, 2, ..., k$, some element $X_{j(i,\nu)}$ is extracted (with or without replacement) from sample H_i at random, $i = 1, 2, ..., m$. After k trials we calculate empirical mean as

$$\widetilde{\theta} = \frac{1}{k} \sum_{\gamma=1}^{k} f(X_{j(1,\gamma)}, X_{j(2,\gamma)}, ..., X_{j(m,\gamma)}). \tag{1}$$

One of the earliest works on the application of the resampling to system reliability point estimation belongs to *V.A. Ivnitsky*. The main problem is to calculate a variance of the corresponding estimate and to compare it with those of traditional estimates. *Andronov* (*1995*) introduced a *notion of ω-pair* that turns out to be very useful. Let $M = 1, 2, ..., m$ be the set of integers, $\omega \subset M$ is a subset of M, $\mathbf{X} = (X_1, X_2, ..., X_m)$ and $\mathbf{X}' = (X'_1, X'_2, ..., X'_m)$, where $X_i, X'_i \in H_i$, $i = 1, 2, ..., m$, be two sub-samples from sample populations $\{H_i\}$. The sub-samples $\mathbf{X}$ and $\mathbf{X}'$ are said to be a *ω-pair* if $X_i = X'_i$ for

$i \in \omega$ and $X_i \neq X_i'$ otherwise. Let $cov(f(\mathbf{X}), f(\mathbf{X}')|\omega)$ be the conditional covariance for $f(\mathbf{X})$ and $f(\mathbf{X}')$ given the condition that $\mathbf{X}$ and $\mathbf{X}'$ union the ω-pair, $p(\omega)$ be a probability to get the ω-pair. Then,

$$Cov(f(X), f(X')) = \sum_{\omega} Cov(f(X), f(X')|\omega)p(\omega), \tag{2}$$

where a sum is taken over the set of all ω-pairs.
Now, to calculate this covariance (and therefore, the variance of the estimate (1)) it is necessary to calculate the probability $p(\omega)$ for all possible ω. The corresponding technique depends on the used resampling procedure. *The hierarchical resampling* has been introduced by *Andronov and Merkuryev (2000)* for the case when the function f is described via a tree . ...
Above, each random variable X_i had its unique sample population H_i. Later (*Andronov 2006*), a more general case has been considered, when some different random variables $\{X_i\}$ have the same sample population H_j. Another generalisation of considered setting concerns dependence in the sample data (*Andronov 2007*) and control of the extraction procedure (*Andronov and Merkuryev 1997*). A case, when distributions of some random variables are known, was considered by *Andronov* and *Fioshin* (*1999*).

3 Interval estimation

Now consider a problem of *the interval* estimation for expectation $\theta = Ef(X_1, ..., X_m)$, that corresponds confidence probability γ. This problem is a main subject of the mathematical statistics and reliability theory. Often the bootstrap approach is applied for confidence interval construction (*Efron and Tibshirani 1993, DiCiccio and Efron 1996, Davison and Hinkley 1997*). In the papers (*Andronov 2002, Andronov and Fioshin 2004*) the resampling method is used. One makes use of the following procedure. Series of experiments are produced. Each experiment includes k trials, which have been described above. After k trials we calculate empirical mean θ_l for the current, for instance the l-th, experiment. Then, we return all extracted elements into the corresponding sample populations and repeat the described experiment r times, where the number r is chosen such that αr is whole number. As result we have a sequence of the estimates $\theta_1, \theta_2, ..., \theta_r$. It gives us the order statistics $\theta_{(1)}, \theta_{(2)}, ..., \theta_{(r)}$ and corresponding α-quantile $\theta_{(\alpha r)}$, of estimate $\widetilde{\theta}$ distribution. We set $\alpha = 1 - \gamma$ and accept $(\theta_{(\alpha r)}, \infty)$ as γ-confidence upper interval for the original value θ.
In the paper (*Andronov 2002*) a two-dimensional simple case is considered when $f(X_1, X_2)$ is the indicator function of the random event $\{X_1 < X_2\}$. For instance, X_1 means a value of shock, X_2 means a strength of a construction or X_1 means a term of the exploitation, X_2 means a lifetime of a construction. Our aim is to construct upper confidence interval $\theta_{(\alpha r)}$ for $\theta = P\{X_1 < X_2\}$ and to calculate a true probability of covering $R = P\{\theta_{(\alpha r)} \leq \theta\}$. For that purpose a notion of *protocol* is used. Let $H_1 = \{x_1^{(1)}, x_1^{(2)}, ..., x_1^{(n)}\}$ and $H_2 = \{x_2^{(1)}, x_2^{(2)}, ..., x_2^{(n')}\}$ are an order presentation of initial samples H_1 and H_2. The value $x_1^{(i)}$ is named as *the i-th point of a protocol*, $i = 0, 1, ..., n+1$, $x_1^{(0)} = -\infty$, $x_1^{(n+1)} = \infty$. Let c_i be a number of values $x_2^{(j)}$ which satisfy the inequalities $x_1^{(i)} < x_2^{(j)} \leq x_1^{(i+1)}$. Then, the $(n+1)$-dimensional vector $C = (c_0, c_1, ..., c_n)$ is said to be *a protocol*. Using the protocols simplifies a calculation of the covering probabilities. The following sequence of elements is used in the corresponding numerical procedure:

- the algorithm of enumeration of all protocols,
- the probability $P_{\mathbf{C}}$ to have fixed protocol $\mathbf{C}$,
- the conditional probability $q_{\mathbf{C}}$ of event $\{f(X_1, X_2) = 1\}$ by condition that protocol $\mathbf{C}$ is fixed,
- the conditional probability $\rho_{\mathbf{C}}$ of the event $\{\theta_l < \theta\}$ for the l-th experiment given $\mathbf{C}$,
- the conditional covering probability $R_{\mathbf{C}}$ by condition that protocol $\mathbf{C}$ is fixed.

Finally, the unconditional probability of the covering is calculated as

$$R = \sum_{C} P_C R_C.$$

The numerical results show that true probability of covering is close to appointee value.
A dissemination of this approach on multivariate case is described by *Andronov and Fioshin (2004).* Here the function $f(x_1, ..., x_m)$ describes an efficiency of a logical system, when f is a predicate (with value 0 or 1), that contains real numbers as arguments $x_1, ..., x_m$, operations over them (such as minimum, maximum, order statistics), the sub-predicates $<, >, =$ and so on.
In the paper (*Andronov and Afanasyeva 2004*) the resampling approach is used to statistical inferences of order statistics, those have an important role in various applications: reliability theory, insurance, storage control and so on. It is sufficient, for example, to refer to the well-known *k-out-of-n-system* (*Gertsbakh 2000*). In fact, an elaborated method allows calculating the distributions of order statistic estimates and on this basis calculating the confidence intervals for quintiles of order statistics.

4 Statistical inference for stochastic processes

In this case the total approach is the same as earlier mentioned. To simulate one trajectory of a considered process, we extract (without replacement) necessary random variables from the given samples at random and calculate the efficiency characteristics of interest θ_l (it corresponds to one trial). Then, we return all extracted values into the initial samples and repeat this procedure many times. As a result, the estimate sequence $\theta_1, \theta_2, ..., \theta_r$ gives a base for various statistical inferences. In a case of the multiple linear regressions (*Afanasyeva and Andronov 2006*), an estimate of unknown coefficients during one trial is performed using a part of the given observations. Now, to get a robustness estimate, we can take the median of getting coefficient estimates. Numerical results testify an advantage of such an approach. A case of a nonparametric interval estimation of the regression function is considered by *Andronov (2007).* Successful results have place for an analogous application to the estimates of the renewal function (*Andronov 2005*), system reliability (*Andronov, Afanasyeva and Fioshin 2009*), efficiency characteristics of queuing systems and so on. In particular it was showed that resampling estimates of the renewal function have less bias in comparison with usual estimates that use the empirical distribution function of time between renewals and its successive convolution.
Andronov, Fioshina and Fioshin (2009) considered the following known reliability problem (*Gertsbakh 2000*). The model supposes two types of failures: *initial* and *terminal failures.* Initial failures (or damages) appear according to homogeneous Poisson process with the rate λ. Each initial failure degenerates into a terminal failure after a random time B. So if an initial failure appears at time τ_i then a terminal failure appears at the instant $B_i + \tau_i$. The terminal failure and the corresponding initial failure are eliminated instantly. We assume that $\{B_i\}$ is mutually independent identically distributed random variables, independent on $\{\tau_i\}$. We take an interest in the number of initial failures $Y(t)$ at time t, which have not degenerated to the terminal failures, and the number of terminal failures $Z(t)$ that have occurred till time t. Let $EY(t)$ and $EZ(t)$ be the corresponding expectations, $P_i(t) = P\{Y(t) = i\}$, $R_i(t) = P\{Z(t) = i\}$ be the corresponding probability distributions, $i = 0, 1, ...$. Our aim is to estimate these characteristics on the basis of the sample $H_1 = \left\{X_1^{(1)}, X_1^{(2)}, ..., X_1^{(n)}\right\}$ of the intervals between initial failures appearances and of the sample $H_2 = \{B_1, B_2, ..., B_{n'}\}$. The authors testify that the proposed resampling-approach is a good alternative to the traditional plug-in estimation. It is especially remarkable increasing the size of the given samples.

Acknowledgements

These researches were supported in part by grants from the Latvian Council of Science (0798, 0842, and 1637) during 1998 - 2008 years.

References

[1] Afanasyeva, H., Andronov, A. (2006). On Robustness of Resampling Estimators for Linear Regression Models. *Communication on Dependability and Quality Management, 9*(1), 5–11.

[2] Andronov, A. (1995). Non-parametric statistical estimation of system's numerical efficiency under the test of its components, *Automatic Control and Computer Sciences, 5, 28-38* .

[3] Andronov, A., Merkuryev, Yu. (1997). Controlled Bootstrap Method and Its Application in Simulation. In *Proceedings of the 11th European Simulation Multiconference*, SCS, Istanbul, Turkey,

160–164.

[4] Andronov, A., Fioshin, M. (1999). Simulation technology under small samples for unknown distributions. In *Proceedings of 10 GI/ITG Special Interest Conference "Measurement, Modelling and Evaluation of Computer and Communication Systems"*, Trier, Germany, 153–162.

[5] Andronov, A., Merkuryev, Yu. (2000). Optimisation of Statistical Sample Sizes in Simulation. *Journal of Statistical Planning and Inference*, *85*, 93–102.

[6] Andronov, A. (2002). On Resampling Approach to a Construction of Approximate Confidence Intervals for System Reliability. In *Third International Conference on Mathematical Methods in Reliability. Methodology and Practice.* Trondheim, Norway, Norwegian University of Science and Technology, 39–42.

[7] Andronov, A., Afanasyeva, H. (2004). Resampling-based nonparametric statistical inferences about the distribution of ordered statistics. In *Transactions of the XXXIV International Seminar on Stability Problems for Stochastic Models.* Jurmala, Latvia, 300–307.

[8] Andronov, A., Fioshin, M. (2004). Applications of resamping approach to statistical problems of logical systems. *ACTA et Commentationes Universitatis Tartuensis de Mathematica*, *8*, 63–71.

[9] Andronov, A. (2005). Resampling-Estimators of the Renewal Function. In *Transactions of the XXV International Seminar on Stability Problems for Stochastic Models.* Maiori (Salerno), Italy, 17–24.

[10] Andronov, A. (2006). About nonparametrical estimations for an expectation of a function of random variables having the same distributions. In *Statistical methods of estimation and hypothesis testing*, Perm (Russia): Perm's State University, 34–48. (In Russian)

[11] Andronov, A. (2007). On nonparametric estimation of expectation of random variables function under dependent observations. *Journal of Statistical Planning and Inference*, *137*, 3828–3837.

[12] Andronov, A. (2007). On nonparametric interval estimation of a regression function based on the resampling. *Computer Modeling and New Technologies*, *11*(1), 47–54.

[13] Andronov, A., Fioshina, H., Fioshin, M. (2009). Statistical estimation for a failure model with damage accumulation in a case of small samples. *Journal of Statistical Planning and Inference*, *139*, 1685–1692.

[14] Belyaev, Yu.K. (2000). Computer Intensive Method Based on Resampling in Analysis of Reliability and Survival Data. In N. Limnios and M. Nikulin (Eds.), *ResentAdvances in Reliability Theory: Methodology, Practice and Inference*, Boston: Birkhauser, 183–198.

[15] Davison, A.C., Hinkley, D.V. (1997). *Bootstrap methods and Their Applications.* Cambridge: Cambridge Press University.

[16] DiCiccio, T.J., Efron, B. (1996). Bootstrap Confidence Intervals. *Statistical Science*, *11*, 189–228.

[17] Efron, B., Tibshirani, R.J. (1993). *Introduction to the Bootstrap.* London: Chapman&Hall.

[18] Gertsbakh, I. (2000). *Reliability Theory: With Applications to Preventive Maintenance.* Berlin: Springer.

[19] Good, P. (2001). *Resampling method: a practical guide to data analysis.* Second edition, Boston: Birkhauser.

[20] Wu, C.F.J. Jackknife. (1986). Bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, *14*, (3), 1261–1295.

On a copula for failure times of system elements

Alexander Andronov
Dept. Mathematical Support of Transport System Management
Riga Technical University
Kalku 1, Riga, LV-1658
Latvia
Aleksandrs.Andronovs@rtu.lv

Abstract

A considered reliability system consists from n identical elements. Each element can be available or failed. A failure rate of element equals λ/i, where i is a number of available elements. A time till a failure of an element has exponential distribution. Independence property has place. Therefore total failure rate is the constant λ, until at last one element is available. Initially all elements are available. The main aim of the paper is to determine a joint distribution of the available time for all elements, as so as a corresponding copula. It allows us to generalize gotten results on unexponential case.

1 Introduction

Usually described in the literature reliability systems are considered under supposition that system's elements fail independently (*Gertsbakh 2000, Limnios and Nikulin 2000*). However, often a case takes place when failure of some elements increases a load on worked elements, so element failure times are dependent random variable. In connection with that a chousing problem of corresponding multidimensional distribution for fitting given statistical data is actual.
In econometric often such problem's solution uses so-called *copulas* (*Nelsen 1999, Embrechts 2000*). Joint distribution function $C(u_1, u_2, ..., u_n) = P\{U_1 \le u_1, U_2 \le u_2, ..., U_n \le u_n\}$ is called *copula* if the marginal distributions of all components $U_1, U_2, ..., U_n$ are uniform on [0, 1]. The following fact is basic (*Sklar 1959*): any multivariate continue distribution function $G(x_1, x_2, ..., x_n) = P\{X_1 \le x_1, X_2 \le x_2, ..., X_n \le x_n\}$ can be presented uniquely via cumulative distribution function of its component $F_i(x_i) = P\{X_i \le x_i\}$ by corresponding copula C:

$$G(x_1, x_2, ..., x_n) = C(F_1(x_1), F_2(x_2), ..., F_n(x_n)).$$

An aim of current paper is using copula-based approach to description of reliability function of system that elements have dependent available times. We consider concrete system of such a sort and derive corresponding copula $C(u_1, u_2, ..., u_n)$. Then we can use this copula with various marginal distributions of components getting a distribution family for fitting aim.
Let us describe a considered system. One consists from n elements. Each element can be in two states: available and failed. A failure rate for each element is λ/i, where i is a number of available elements, with that corresponding time till failure has exponential distribution. Independence property has place as well. Therefore total failure rate is constant λ, until at least one element is available. We would like to answer on the following questions: 1) What is distribution of time moment T when the last element fails? 2) What is marginal distribution of failure time X for fixed element? 3) What is joint distribution of failure times $\{X_i\}$ for fixed elements? 4) What is corresponding copula?
To get answers on two first questions, we take in mind the following. 1) Failure flow is a Poisson process with intensity λ, till a time moment when the last element fails. 2) The time moment of the i-th failure has Erlang distribution with parameters λ and i (*Ross 1996*). 3) Each such moment is a failure time moment of some element. The fixed element fails as i-th in failure succession with probability $1/n$. It allows us to give answer for two first questions.
Distribution of time moment T when the last element fails:

$$R\{T \le t\} = 1 - \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!} e^{-\lambda t}, \;\; t > 0. \tag{1}$$

The marginal probability density function for failure time X of fixed element:

$$p_\lambda(x) = \frac{1}{n}\sum_{i=1}^{n} \lambda \frac{(\lambda x)^{i-1}}{(i-1)!} e^{-\lambda x}, \quad x \geq 0. \tag{2}$$

A corresponding distribution function is

$$P_\lambda(x) = P\{X \leq x\} = \int_0^x p_\lambda(z)dz = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \sum_{j=0}^{i-1}\frac{(\lambda x)^j}{j!}e^{-\lambda x}\right) =$$

$$= 1 - \frac{1}{n}\sum_{j=0}^{n-1}\sum_{i=j+1}^{n}\frac{(\lambda x)^j}{j!}e^{-\lambda x} = 1 - \frac{1}{n}\sum_{j=0}^{n-1}(n-j)\frac{(\lambda x)^j}{j!}e^{-\lambda x} = \tag{3}$$

$$= 1 - (\lambda x)^{n-1}\frac{1}{(n-1)!}e^{-\lambda x} - \left(1 - \frac{\lambda x}{n}\right)\sum_{i=0}^{n-2}\frac{(\lambda x)^i}{i!}e^{-\lambda x}, \quad x \geq 0.$$

2 Joint distribution of failure times

Joint density function for $X_1, X_2, ..., X_n$ obviously is

$$g(x_1, x_2, ..., x_n) = \frac{1}{n}x_{i^*}^{-(n-1)}\frac{1}{(n-1)!}\lambda(\lambda x_{i^*})^{n-1}e^{-\lambda x_{i^*}} = \frac{1}{n!}\lambda^n e^{-\lambda x_{i^*}}, \ \ \forall \ \ 0 \leq x_i \leq x_{i^*}, \tag{4}$$

where $x_{i^*} = \max\{x_1, x_2, ..., x_n\}$.
Let us verify the normalization condition. For $i^* = 1$ we have

$$\int_0^\infty \int_0^z ... \int_0^z g(z, x_2, ..., x_n)dx_2...dx_n dz = \int_0^\infty z^{n-1}\frac{1}{n!}\lambda^n e^{-\lambda z}dz = \frac{1}{n}. \tag{5}$$

This result can be multiplying by n because full integral contains possibilities that maximal component can take any place from n ones ($i^* = 1, 2, ..., n$). So the normalization condition is fulfilled.
One can get that joint density function for $X_1, X_2, ..., X_k$, $1 < k \leq n$, has the following form:

$$g(x_1, x_2, ..., x_k) = \frac{1}{n}x_{i^*}^{-(k-1)}\sum_{i=k}^{n}\frac{(i-1)(i-2)...(i-k+1)}{(n-1)(n-2)...(n-k+1)!}\frac{1}{(i-1)!}\lambda(\lambda x_{i^*})^{i-1}e^{-\lambda x_{i^*}},$$

where $x_{i^*} = \max\{x_1, x_2, ..., x_k\}$.
The last expression is valid for $k = 1$ as well, if we set $(i-1)(i-2)...i = 1$, and $(n-1)(n-2)...n = 1$.
Now we intend to get an expression for joint distribution function $G(x_1, x_2, ..., x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n\}$. At first we consider a case $x_1 < x_2 < ... < x_n$. The event $\{X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n\}$ consists from n disjoint events, when the random variable X_i, $i = 1, 2, ..., n$, takes the maximal value. If we denote this maximal value by z then

$$G(x_1, x_2, ..., x_n) = \int_0^{x_1}\int_0^z ... \int_0^z g(z, \nu_2, ..., \nu_n)d\nu_n...d\nu_2 dz +$$

$$+\sum_{i=2}^{n}\left\{\int_0^{x_1}\int_0^z ... \int_0^z g(\nu_1, \nu_2, ..., \nu_{i-1}, z, \nu_{i+1}, ..., \nu_n)d\nu_n d\nu_{n-1}...d\nu_{i+1}d\nu_{i-1}...d\nu_1 dz +\right.$$

$$+\int_{x_1}^{x_2}\int_0^{x_1}\int_0^z ... \int_0^z g(\nu_1, \nu_2, ..., \nu_{i-1}, z, \nu_{i+1}, ..., \nu_n)d\nu_n d\nu_{n-1}...d\nu_{i+1}d\nu_{i-1}...d\nu_1 dz + ...$$

$$\left. ... + \int\limits_{x_{i-1}}^{x_i} \int\limits_0^{x_1} ... \int\limits_0^{x_{i-1}} \int\limits_0^{z} ... \int\limits_0^{z} g(\nu_1, \nu_2, ..., \nu_{i-1}, z, \nu_{i+1}, ..., \nu_n) d\nu_n d\nu_{n-1} ... d\nu_{i+1} d\nu_{i-1} ... d\nu_1 dz \right\}.$$

A substitution (4) in the integral gives the following expression:

$$G(x_1, x_2, ..., x_n) = \frac{1}{n} \left(1 - \sum_{j=0}^{n-1} \frac{1}{j!} (\lambda x_1)^j e^{-\lambda x_1} \right) + \sum_{i=2}^{n} \left\{ \left(1 - \sum_{j=0}^{n-1} \frac{1}{j!} (\lambda x_1)^j e^{-\lambda x_1} \right) + \right.$$

$$+ \int\limits_{x_1}^{x_2} \int\limits_0^{x_1} z^{n-2} \frac{\lambda^n}{n!} e^{-\lambda z} d\nu_1 dz + \int\limits_{x_2}^{x_3} \int\limits_0^{x_1} \int\limits_0^{x_2} z^{n-3} \frac{\lambda^n}{n!} e^{-\lambda z} d\nu_2 d\nu_1 dz + ...$$

$$\left. ... + \int\limits_{x_{i-1}}^{x_i} \int\limits_0^{x_1} ... \int\limits_0^{x_{i-1}} z^{n-i} \frac{\lambda^n}{n!} e^{-\lambda z} d\nu_{i-1} ... d\nu_1 dz \right\} =$$

$$= 1 - \sum_{j=0}^{n-1} \frac{1}{j!} (\lambda x_1)^j e^{-\lambda x_1} + \sum_{i=2}^{n} \left\{ x_1 \frac{\lambda}{n(n-1)} \left[\sum_{j=0}^{n-2} (x_1^j e^{-\lambda x_1} - x_2^j e^{-\lambda x_2}) \frac{\lambda^j}{j!} \right] + \right.$$

$$+ x_1 x_2 \frac{\lambda^2}{n(n-1)(n-2)} \left[\sum_{j=0}^{n-3} (x_2^j e^{-\lambda x_2} - x_3^j e^{-\lambda x_3}) \frac{\lambda^j}{j!} \right] + ...$$

$$\left. ... + x_1 x_2 ... x_{i-1} \frac{\lambda^{i-1}}{n(n-1)...(n-i+1)} \left[\sum_{j=0}^{n-i} (x_{i-1}^j e^{-\lambda x_{i-1}} - x_i^j e^{-\lambda x_i}) \frac{\lambda^j}{j!} \right] \right\}.$$

Final formula has the following view:

$$G(x_1, x_2, ..., x_n) = 1 - \sum_{j=0}^{n-1} \frac{1}{j!} (\lambda x_1)^j e^{-\lambda x_1} +$$

$$+ \sum_{i=2}^{n} x_1 x_2 ... x_{i-1} \frac{\lambda^{i-1}}{n(n-1)...(n-i+1)} (n-i+1) \sum_{j=0}^{n-1} \left(x_{i-1}^j e^{-\lambda x_{i-1}} - x_i^j e^{-\lambda x_i} \right) \frac{\lambda^j}{j!}. \tag{6}$$

Now we consider a case when the sequence $x_1, x_2, ... x_n$ is not ordered. Let $\pi = (\pi(1), \pi(2), ..., \pi(n))$ be a permutation of $1, 2, ..., n$. We consider $\pi(i)$ as a number of the element $x_{\pi(i)}$ that take i-th place in the ordered sequence $x^{(1)} < x^{(2)} < ... < x^{(n)} : x_{\pi(i)} = x^{(i)}$. Then the previous formula takes place if we change the index i by $\pi(i)$. In fact to calculate the distribution function $G(x_1, x_2, ..., x_n)$ for arbitrary order of $x_1, x_2, ..., x_n$, we must range this sequence, get ordered one $x^{(1)} < x^{(2)} < ... < x^{(n)}$, and use formula (6) for $x_i = x^{(i)}$.

3 Copula

Our last aim is to find an expression for a copula corresponding to the joint distribution (6). Let $q(p) > 0$ be the root of the equation

$$p = 1 - \frac{1}{n} \sum_{j=0}^{n-1} (n-j) \frac{1}{j!} x^j e^{-x}, \quad 0 < p < 1. \tag{7}$$

Obviously it is the p-qvantile of the distribution (3) for $\lambda = 1$:

$$P_1(q(p)) = 1 - \frac{1}{n}\sum_{j=0}^{n-1}(n-j)\frac{1}{j!}q(p)^j e^{-q(p)} = p, \quad 0 < p < 1.$$

Substitution $p = P_\lambda(x)$ gives

$$P_1(q(P_\lambda(x))) = 1 - \frac{1}{n}\sum_{j=0}^{n-1}(n-j)\frac{1}{j!}q(P_\lambda(x))^j e^{-q(P_\lambda(x))} = P_\lambda(x) \quad x > 0.$$

By a comparison with (3) we get

$$\lambda x = q(P_\lambda(x)).$$

It allows us to represent (6) for $x_1 < x_2 < ... < x_n$ by such way:

$$G(x_1, x_2, ..., x_n) = 1 - \sum_{j=0}^{n-1}\frac{1}{j!}q(P_\lambda(x))^j e^{-q(P_\lambda(x_1))} + \sum_{i=2}^{n} q(P_\lambda(x_1))...q(P_\lambda(x_{i-1}))\times$$

$$\times\frac{1}{n(n-1)...(n-i+2)}\sum_{j=0}^{n-1}\left(q(P_\lambda(x_{i-1}))^j e^{-q(P_\lambda(x_{i-1}))} - q(P_\lambda(x_i))^j e^{-q(P_\lambda(x_i))}\right)\frac{1}{j!}. \tag{8}$$

Therefore a copula of interest for $0 < u_1 < u_2 < ... < u_n < 1$ has the following form:

$$C(u_1, u_2, ..., u_n) = 1 - \sum_{j=0}^{n-1}\frac{1}{j!}q(u_1)^j e^{-q(u_1))} +$$

$$+\sum_{i=2}^{n} q(u_1)...q(u_{i-1})\frac{1}{n(n-1)...(n-i+2)}\sum_{j=0}^{n-1}(q(u_{i-1})^j e^{-q(u_{i-1})} - q(u_i)^j e^{-q(u_i)})\frac{1}{j!}. \tag{9}$$

Using this formula, one can represent the joint distribution (6) for $x_1 < x_2 < ... < x_n$ in the form

$$G(x_1, x_2, ..., x_n) = C(P_\lambda(x_1), P_\lambda(x_2), ..., P_\lambda(x_n)),$$

where $P_\lambda(x_i)$ is calculated by (3).
Now we are able to generalize our results and consider a family of reliability functions of the system as

$$G(x_1, x_2, ..., x_n) = C(F(x_1), F(x_2), ..., F(x_n)),$$

where $F(x)$ is an arbitrary reliability function of elements.
Note a case of unordered sequence $x_1, x_2, ..., x_n$ is considered as earlier.

4 Example

We consider our basic model for the following data: $n = 4$, $\lambda = 2$. Therefore for initial failure rate $\lambda/n = 0.5$, a fixed element has the following values of the mean μ and the standard deviation σ of a time till a failure: $\mu = \sigma = n/\lambda = 2$. The table contains values of the joint distribution function (6) for different values of argument $x = (x_1\ x_2\ x_3\ x_4)^T$.
Now we show how the same copula (9) can be used for another marginal distributions $F(x)$. As the last uniform distribution $U(x)$ and lognormal distribution $L(x)$ are chosen:

$$U(x) = \begin{cases} 0 & if \quad x < 0, \\ \frac{x}{2\mu} & if \quad 0 < x < 2\mu, \\ 1 & if \quad x > 2\mu, \end{cases} \qquad L(x) = \begin{cases} 0 & if \quad x < 0, \\ \Phi\left(\frac{ln(x)-a}{s}\right) & if \quad x > 0, \end{cases}$$

where $\Phi(z)$ is the cumulative distribution function of the standard normal distribution,
a and $s > 0$ are parameters of lognormal distribution.

The parameters of both distribution $U(x)$ and $L(x)$ were choused in such a way that expectation μ coincide for all three distributions. For the lognormal distribution standard deviation σ coincides as well. Let us remember that μ and σ of the lognormal distribution are calculated with respect to formulas

$$\mu = exp\left(a + \frac{1}{2}s^2\right), \quad \sigma^2 = (exp(s^2) - 1)exp(2a + s^2).$$

For our numerical data $a = 0.3464$ and $s = 0.8326$. Corresponding values of the joint distribution functions $GU(x)$ and $GL(x)$ are represented in the Table. A comparing of all three distributions shows that the considered copula (9) generates different distributions. It allows us to suggest using the considered model for a description of complex reliability systems. An estimation of unknown parameters can be performed, in particular, analogously to the paper (*Andronov 2008*).

Table 1: Values of the joint distribution functions

$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 1.5 \\ 3 \\ 3.5 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 3.5 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 1.5 \\ 2.5 \\ 3.5 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 3 \\ 3.5 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 3 \\ 3 \\ 3 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 3.5 \\ 4 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 6.5 \\ 6.5 \\ 6.5 \end{pmatrix}$
$G(x)$	0.363	0.415	0.567	0.621	0.767	0.849	0.912	0.999
$GU(x)$	0.107	0.136	0.159	0.264	0.400	0.473	0.690	1.000
$GL(x)$	0.198	0.236	0.341	0.410	0.534	0.589	0.677	0.904

5 Conclusion

In the paper new family of multivariate distribution functions for nonnegative random variables was suggested. All distributions have the same copula and differ one from other by marginal distributions. The last allows choosing a multivariate distribution that in best way fits given data of system reliability.

References

[1] Andronov, A. (2008). On a simple method of multivariate distribution estimation defined by copulas. In *Modelling in Business, Industry and Transport*, Riga: Transport and Telecommunication Institute.

[2] Gertsbakh, I. (2000). *Reliability Theory: with Applications to Preventive Maintenance*. Berlin: Springer.

[3] Embrechts, P. (2000). *Some Selected Papers*. http://www.math.ethz.ch/ baltes/ftp/papers.html

[4] Limnios, N., Nikulin, M. (2000). *Recent Advances in Reliability Theory: Methodology, Practice, and Inference*. Boston: Birkhauser.

[5] Nelsen, R.B. (1999). *An Introduction to Copulas. Lecture Notes in Statistics*, *139*, NewYork: Springer.

[6] Ross, S.M. (1996). *Stochastic Processes*. New York: Wiley.

[7] Sklar, A. (1959). *Functions de repartition a n dimensions et leurs marges*. Publications de 1'Institut de Statistique de 1'Universitê de Paris, *8*, 229–231.