Advances in Reliability, Safety and Security
ESREL 2024
Collection of Extended Abstracts

Stochastic Three-Dependence. General Analytic Models' Pattern

Jerzy K. Filus[a], Lidia Z. Filus[b]

[a]*Oakton College, Des Plaines, Illinois, USA*
[b]*Northeastern Illinois University, Chicago, USA*

Keywords: stochastic dependence, three-dependence, system reliability modeling, universal joiner representation of multivariate survival (reliability) functions, new stochastic models

---

The well known in statistics "third variable effect" is revisited and developed on the ground of stochastic modelling within statistics free probability theory. In that, a general approach to this modelling (including multicomponent system ***reliability*** modelling) with use of the joiner's methodology is applied.

At the moment, it is hard to find in literature a *general enough analytic* model for phenomena like, between others, those two described in the following examples (Somewhat related models present in literature are, however, those described as the 'conditional dependence'.)

Example 1. A researcher observes that (randomly chosen) cities with more fire hydrants tend also to have more dogs. However, these two variables are only correlated (stochastically dependent) because they both have a high correlation with a third variable: city's population size. Larger cities tend to have more fire hydrants and more dogs. On the other hand, smaller cities tend to have fewer fire hydrants and fewer dogs. □

Example 2. Suppose in an experimental medical trial one investigates the correlation between blood pressure and blood cholesterol level in groups of the patients if the age of patients in each group is the same but is different than in any other group while the ages varies across different groups. It turns out that the correlation coefficient of the original two quantities of interest is a function of the third variable i.e., of the age. □

In this case, the correlation coefficient between the blood pressures and the cholesterol levels depends not only on the fact that all the patients were humans but also on realization a of the random variable A here representing an age.

When considering set of problems associated with ***system reliability*** one may encounter the situation when three components of a system are arranged in such a way that third component may either be considered as a "stabilizer" of work conditions of the two remaining components or may represent an environmental random quantity that induces a positive or negative stress on the two components. See, for example (Lindley and Singpurwalla, 1986). In cases like that, the dependence of the two component life-times also may depend on magnitude of an environmental quantity (temperature, for example). Mathematically the situation described in the last reference is similar like that in Example 2.

This kind of the 3-dependence we will call ***3-dependence of second kind***.

This second kind of 3-dependence is known in statistical investigations. However, mostly it is investigated within pure statistical frameworks which means the problem is reduced to data processing and analyzing. According to our best knowledge, the probabilistic [as opposite to statistic] analytical models, such as joint probability distributions of the underlying three or more random variables involved, are rare and are not general to a degree that would comprehensively describe the essence of the phenomena like those in above examples.

In the coming presentation, we will tend to describe the phenomenon of three or more dependence generally as to comprise all or at least vast majority of cases of 3-dependence of second kind by providing an analytical *pattern* for modelling the 3-dependence. The obtained pattern is very general, comprising system reliability problems on one side and on the other exceeding frameworks of reliability toward full generality. As the analytical tool for that aim, we apply the **'*joiner*** methodology' that was elaborated in our previous works such as, for example (Filus and Filus, 2020).

Recall then the ***universal joiner representation*** of (any) *k*-variate survival function of ***any*** random vector $(X_{\mathbf{1}}, \ldots, X_{\boldsymbol{k}})$ as it is given as follows**:**

$$S(x_1, \ldots, x_k) = P(X_1 \geq x_1, \ldots, X_k \geq x_k) = J_{1,\ldots,k}(x_1, \ldots, x_{k-1}, x_k)\, S_1(x_1)\, S_2(x_2) \ldots S_k(x_k) \tag{1}$$

where $S_1(x_1)\, S_2(x_2) \ldots S_k(x_k)$ are univariate marginal survival functions of the random variables $X_1, \ldots, X_k$ **(**being, for example, life-times of system components in system reliability modeling settings **)** and the continuous "dependence function" $J_{1,\ldots,k}(x_1,\ldots,x_{k-1}, x_k)$, called the ***joiner,*** determines all the stochastic dependences (including the three-dependence if exist) among the random variables $X_1, \ldots, X_k$.

Representation (1) of any survival function is <u>universal</u> and, as such, is competitive to the ***copula*** representation which, in turn, works for cumulative distribution functions.

To see the difference, suppose, the dependences occur pairwise only i.e., among each pair of the variables, say X_i, X_j, where $1 \leq i < j \leq k$, so that the dependence degree [a correlation coefficient, for example] within any pair X_i , X_j does not depend on realization of any remaining random variables within the vector $(X_1, \ldots, X_k)$.

Such cases describe the joiner having the following form of the product**:**

$$J_{1,\ldots,k}(x_1, \ldots, x_{k-1}, x_k) = \prod_{1\leq i<j\leq k} J_{i,j}(x_i, x_j). \tag{2}$$

This means that all the dependences among $X_1, \ldots, X_k$ reduce to, say **"*bi-dependences*"** and in such a case, there are mostly ***no*** direct dependences within triples X_i, X_j, X_r , treated as wholes, where i, j, r are indexes such that $1 \leq i < j < r \leq k$.

Formulas (1) and (2) are, probably, the most general but for better presentation we restrict [unnecessarily if from generality viewpoint] the considerations to "continuous cases" only.

For that we adopt the <u>convention</u>, that k-variate survival (reliability) function (1) is "continuous" if the continuous hazard [failure] rates $\lambda_1(t_1), \ldots, \lambda_k(t_k)$ of all the underlying random variables $X_1, \ldots, X_k$ $(k \geq 3)$ exist and, moreover, for every **"*bi-joiner*"** $J_{i,j}(x_i, x_j)$, present in formula (2), there exist the following exponential integral representation**:**

$$J_{i,j}(x_i, x_j) = \exp[\, -\int_0^{x1} \int_0^{x2} \Psi_{i,j}(u_i, u_j)\, du_i\, du_j], \tag{3}$$

where nonnegative continuous function $\Psi_{i,j}(u_i, u_j)$ uniquely represents that bi-joiner.

Under that continuity assumption and according to notation (3), general model (1) satisfying assumption (2) takes on the form**:**

$$S(x_1, \ldots, x_k) = \exp[\, -\int_0^{x1} \lambda_1(t)\, dt - \ldots - \int_0^{xk} \lambda_k(t_k)\, dt_k - \Sigma_{1\leq i<j\leq 1}\, c_{ij} \int_0^{xi}\int_0^{xj} \Psi_{i,j}(u_i, u_j)\, du_i\, du_j], \tag{4}$$

where the nonnegative continuous functions $\Psi_{i,j}(u_i, u_j)$ must satisfy**:**

$$\Psi_{i,j}(u_i, u_j) \leq \lambda_i(u_i)\, \lambda_j(u_j) \tag{5}$$

with nonnegative real constants c_{ij} satisfying**:**

$$\Sigma_{1\leq i<j\leq 1}\, c_{ij} \leq 1. \tag{6}$$

For justification of the conditions (5) and (6) see (Filus and Filus, 2020).

The 3-dependence, absent in the model described by formula (4), will take place if we enrich the resulting exponent in (4) by the term $-\gamma \int_0^{x1} \int_0^{x2} \int_0^{x3} \Psi_{123}(u_1, u_2, u_3)\, du_1\, du_2\, du_3$, where γ is a real nonzero parameter satisfying $|\gamma| < 1$.

Now, formula (4) takes on the following form**:**

$$S(x_1, x_2, x_3) = \exp[\, -\int_0^{x1} \lambda_1(t_1)\, dt_1 - \int_0^{x2} \lambda_2(t_2)\, dt_2 - \int_0^{x3} \lambda_3(t_3)\, dt_3$$
$$- \Sigma_{1\leq i<j\leq 3}\, c_{ij} \int_0^{xi}\int_0^{xj} \Psi_{i,j}(u_i, u_j)\, du_i\, du_j - \gamma \int_0^{x1}\int_0^{x2}\int_0^{x3} \Psi_{123}(u_1, u_2, u_3)\, du_1\, du_2\, du_3], \tag{7}$$

where c_{ij} are nonnegative real coefficients and γ is a constant real coefficient such that $|\gamma| < 1$.

Now, we must add the following additional condition to the conditions (5) and (6):

$$0 \leq \Psi_{123}(u_1, u_2, u_3) \leq \lambda_1([u_1]^{2/3})\lambda_2([u_2]^{2/3})\, \lambda_3([u_3]^{2/3}) \tag{8}$$

(or, possibly more generally,

$$0 \leq \Psi_{123}(u_1, u_2, u_3) \leq \lambda_1([u_1]^{a})\lambda_2([u_2]^{b})\, \lambda_3([u_3]^{c}) \tag{8*}$$

with arbitrary non-negative constant exponents a, b, c satisfying $a + b + c \leq 2$**).**

The conditions (5), (6) and (8) or (8*) ensure that formula (7) represents legitimate survival function. This means that the corresponding expression $-\partial^3 / \partial x_1\, \partial x_2\, \partial x_3\, (S(x_1, x_2, x_3))$ is uniformly nonnegative i.e., the joint probability density is well defined.

References

Lindley, D.V., Singpurwalla, N.D. 1986. Multivariate Distributions for Life-Lengths of Components of a System Sharing a Common Environment, J. Appl. Probab. 23, 418-431.

Filus, J.K., Filus, L.Z. 2020. On Construction of k-Variate Survival Functions with k = 3 Special Case Emphasis, Applied Mathematics 11 (07), 678 – 697.