Specifying Interdependence in Networked Systems

Nozer D. Singpurwalla and Chung-Wai Kong

***Abstract*—Realistic assessments of the reliability of networked systems, series and parallel systems being special cases, require that we account for interdependence between the component life-lengths. The key to doing this is the specification and use of a suitable probability model in two or more dimensions. Consequently, several multivariate probabilistic models have been proposed in the literature. Many of these models have marginal distributions that are exponential; the ones by Gumbel, and by Marshall and Olkin being some of the earliest and the best known. The purpose of this paper is two fold:**

The first purpose is to articulate the nature of dependence encapsulated by such models, using a perspective which is best appreciated by a user. Specifically, we anchor on the bivariate case, and focus attention on the *conditional mean* as a measure of dependence. The second purpose, motivated by the first, is to introduce a new family of multivariate distributions with exponential marginals, whose conditional mean fills a void in the general forms of the conditional means of the available models. The method of "copulas" is used to generate this new family of distributions. Attention is focused on the case of exponential marginals, because the notion of "hazard potentials" enables us to use multivariate distributions with exponential marginals as a seed for generating multivariate distributions with marginals other than the exponential.

***Index Terms*—Bivariate exponential distributions, conditional expectaion, copulas, hazard potentials, interaction, regression function, reliability.**

I. Interdependent Life-Times

CONSIDER a two component system with component i having life-time T_i, $i = 1, 2$. When these components operate in a common environment as a system, or when they share similarities due to commonalities in design and/or manufacturing, T_1 & T_2 cannot be judged independent. Rather T_1 & T_2 are said to experience *interaction* or *inter-dependence*. Such inter-dependence is (probabilistically) encapsulated by a bivariate probability distribution having the property that for any $t_1, t_2 \geq 0$, $\mathcal{P}(T_1 \geq t_1, T_2 \geq t_2) \neq \mathcal{P}(T_1 \geq t_1) \cdot \mathcal{P}(T_2 \geq t_2)$; $\mathcal{P}(T_i \geq t_i)$ is the marginal distribution of T_i, $i = 1, 2$. When $\mathcal{P}(T_1 \geq t_1, T_2 \geq t_2) \geq (\leq) \mathcal{P}(T_1 \geq t_1) \cdot \mathcal{P}(T_2 \geq t_2)$, the life-times are said to be *positively* (*negatively*) *dependent*. With positive dependence, an assessor of these probabilities holds the view that a knowledge of the failure of one component exacerbates the assessment of failure of the surviving component. It is often the case that life-times of physical units are judged positively dependent. Negative dependence is rare, save for some certain scenarios in the biological sciences involving a competition for resources.

Manuscript received October 23, 2002; revised June 30, 2003. This work was supported by Grants DAAD19-02-1-0195 and DAAD19-01-1-0502 from the U.S. Army Research Office. Associate Editor: W. Kuo.

The authors are with the Institute for Reliability and Risk Analysis, Department of Statistics, The George Washington University, Washington, D.C. 20052, USA.

Digital Object Identifier 10.1109/TR.2004.833423

The mere judgment that life-times are positively or negatively dependent is purely a qualitative one. At best, its only use is an ability to determine if a system reliability calculation, assuming independent life-times, provides an upper or a lower bound. To obtain sharp assessments of system reliability, we need to specify the degree, or the extent, to which the life-times are dependent. For this, several measures such as (Pearson's) linear correlation, Kendall's Tau, and Spearman's Rho have been proposed. However, from the point of view of a user, say an engineer, these measures are not as intuitively appealing as another measure of interdependence, namely, the *conditional mean*, also known as the *regression function*. For example, Pearson's linear correlation ρ only encapsulates the degree to which T_1 and T_2 bear a *linear* relationship to each other. In principle, T_1 and T_2 could be highly dependent (though not linearly), and yet their ρ could be close to zero; the bivariate exponential distribution of [6] is a case in point. Furthermore, a knowledge of ρ provides little guidance as to what family of bivaraite distributions is a meaningful one to choose. By contrast, as will be shown here, the conditional mean is able to provide a nice discriminatory capability for model selection. This, plus the fact that the conditional mean value function can be elicited in lay terms, makes it a viable candidate for specifying the nature of interdependence in multi-component systems.

In the interest of simplicity, we have restricted attention to the bivariate case by focusing on a two-component system. Whereas this by itself is of limited practical value it helps us lay the groundwork for addressing the case of specifying interdependencies in multi-component systems.

II. The Regression Function

The regression function of T_2 on T_1, denoted $E(T_2|T = t_1)$, is a function of t_1, giving us the expected value of T_2 *were* $T_1 = t_1$. This function is easy to elicit because all one need do is provide ones best assessment of T_2, the time to failure of the component labeled 2, were T_1, the time to failure of the component labeled 1, is $t_1, t_1 \geq 0$. The elicited $E(T_2|T = t_1)$ is a way to encapsulate an expert's view of the relationship between T_2 and T_1 in a form easy to articulate. The expert is free to specify the general shape of $E(T_2|T = t_1)$, be it linear, polynomial, or exponential, each shape corresponding to a suitable bivariate distribution. Possible shapes of the regression function for some well known bivariate distributions with exponential marginals are described below.

A. Gumbel's (Type I) Bivariate Exponential Distribution

Gumbel [3] proposed a bivariate distribution with exponential marginals whose general form is

$$\mathcal{P}(T_1 \geq t_1, T_2 \geq t_2|\delta) = \exp[-(t_1 + t_2 + \delta t_1 t_2)] \quad (1)$$

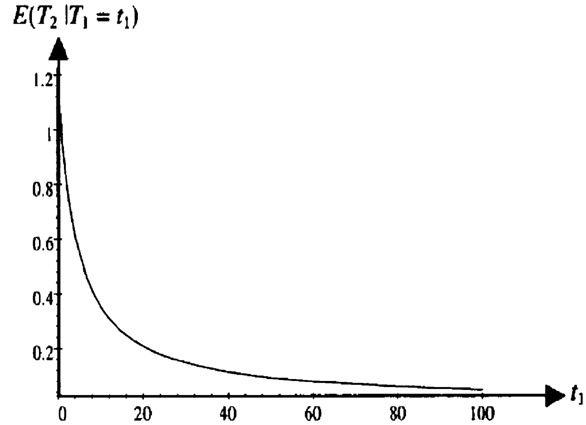

Fig. 1. Regression function of Gumbel's type I bivariate exponential with $\delta = 0.2$.

for some $\delta \in [0,1]$; δ is a parameter whose value describes the nature of dependency between T_1 and T_2. For this model, the regression of T_2 on T_1 is of the form

$$E_\delta(T_2|T = t_1) = \frac{1 + \delta + \delta t_1}{(1 + \delta t_1)^2}, \quad t_1 \geq 0. \tag{2}$$

This function of t_1 is polynomially decreasing, taking the value 2 at $t_1 = 0$, and converging to 0 as $t_1 \to \infty$; see Fig. 1. The model encapsulates negative dependence, and consequently is of a limited interest in reliability. The parameter δ controls the rate at which the regression function decays to zero.

B. Gumbel's (Type II) Bivariate Exponential Distribution

To incorporate both positive and negative dependence, and as a boundary case independence, Gumbel [3] has also proposed another bivariate distribution with exponential marginals. Here, for some dependency parameter $\alpha \in [-1, +1]$, and $t_1, t_2 \geq 0$,

$$\mathcal{P}(T_1 \geq t_1, T_2 \geq t_2|\alpha) = e^{-(t_1+t_2)} \times [1 + \alpha(1 - e^{-t_1} - e^{-t_2} + e^{-(t_1+t_2)})]. \tag{3}$$

For this model, the regression function is

$$E_\alpha(T_2|T = t_1) = 1 + \frac{\alpha}{2} - \alpha e^{-t_1}, \tag{4}$$

which for $\alpha > (<)$ is an increasing (decreasing) function of t_1, and is a constant for $\alpha = 0$; see Fig. 2. Consequently, $\alpha > 0$ encapsulates positive dependence, and $\alpha < 0$ encapsulates negative dependence; we have independence when $\alpha = 0$. Furthermore, because $-1 \leq \alpha \leq +1$, the regression function is bounded between 1/2 & 3/2; when it increases or decreases, it does so exponentially.

C. Marshall and Olkin's Multivariate Exponential Distribution

Marshall and Olkin [6] introduced a multivariate distribution with exponential marginals for describing positive dependence between the life-times of a multi-component system. In the bivariate case, their model takes the form

$$\mathcal{P}(T_1 \geq t_1, T_2 \geq t_2|\lambda_1, \lambda_2, \lambda_{12}) = e^{-(\lambda_1 t_1 + \lambda_2 t_2 + \lambda_{12} \max(t_1, t_2))}, \tag{5}$$

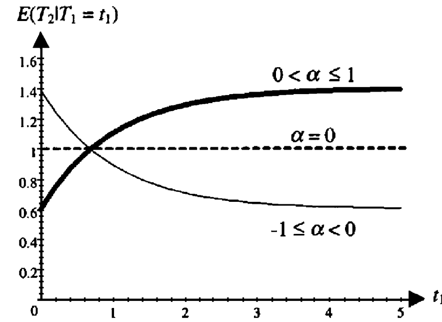

Fig. 2. Regression function for Gumbel's type II bivariate exponential for different values of α.

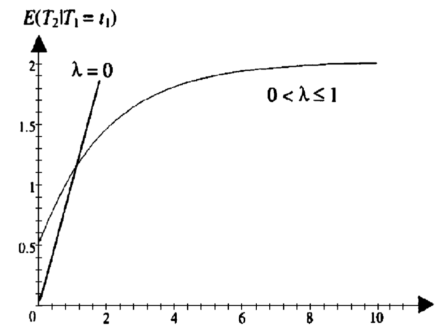

Fig. 3. Regression function of Marshall and Olkin's bivariate exponential for $\lambda = 0$ and $\lambda = 0.5$.

where λ_1, λ_2, and $\lambda_{12} \geq 0$ are the parameters of the model, λ_{12} being the dependency parameter. With $\lambda_1 = \lambda_2 = \lambda$ and $\lambda_{12} = 1 - \lambda$, the regression function takes the form

$$E_\lambda(T_2|T = t_1) = \frac{1 - (1 - \lambda^2)\exp(-\lambda t_1)}{\lambda}, \text{ for } \lambda > 0$$
$$= t_1, \text{ for } \lambda = 0. \tag{6}$$

The regression function increases linearly in t_1 when $\lambda = 0$, and exponentially in t_1 when $\lambda > 0$; see Fig. 3. For $\lambda > 0$, $E_\lambda(T_2|T = t_1)$ is bounded between λ and $1/\lambda$, allowing it to take a wider range of values than those taken by the regression function of Gumbel's Type II bivariate exponential distribution with $\alpha \in (0, 1]$.

D. Singpurwalla and Youngren's Bivariate Distribution

Whereas Gumbel's bivariate exponential distributions have no physical motivation, Marshall and Olkin's distribution does, in the sense that it is derived from Poisson shock generating processes. This is also true of a distribution introduced by Singpurwalla and Youngren [9]. This distribution is derived via a shock generating mechanism governed by a shot-noise process; it has the form of (7) (at the bottom of the next page) for some dependency parameter $m > 0$.

For this model, the regression function $E_m(T_2|T = t_1)$ cannot be expressed in closed form. Fig. 4 illustrates the general form of the regression function, numerically evaluated for $m = 1$. It is initially convex, and asymptotes as a linear function of t_1. The initial convexity signals the fact that small values of t_1 provide little insight about the behavior of T_2; for large values of t_1, the behavior of T_2 parallels that of T_1.

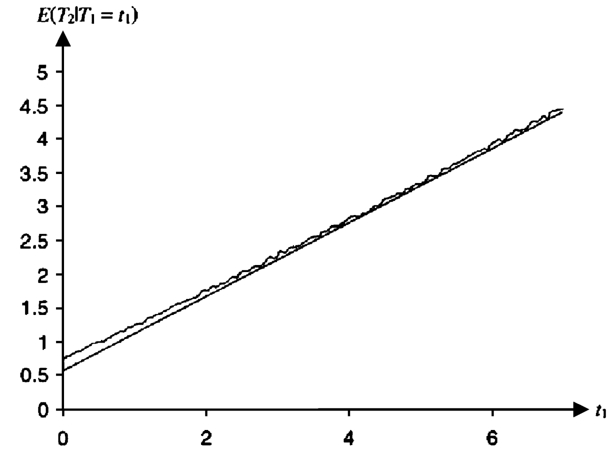

Fig. 4. Regression function of Singpurwalla and Youngren's bivariate exponential, with $m = 1$.

A linear regression function is also characteristic of a bivariate exponential distribution proposed by Downton [4]. Here, $E_\delta(T_2|T = t_1) = (1-\rho) + \rho t$, where $\rho \in [0,1]$ is a dependency parameter.

There are several other bivariate distributions with exponential marginals which have been proposed in the literature; a comprehensive review exists in [4]. For the purposes of this paper, we have confined attention to the four illustrations given above.

E. Discussion: The Elicitation of Regression Functions

Independence and dependence are judgments made by a single assessor or a group of assessors acting as one. Thus the nature of dependence has to be elicited, and the regression function is a vehicle for doing so. The regression function is attractive because it profiles an assessor's view (via the mean) about the time to failure of a surviving component, given the several possible times to failure of the failed component. A constant regression function signals the lack of dependence, because it states that no matter what the time to failure of a failed component is, the time to failure of the surviving component is assessed at a constant. Similarly, an increasing (decreasing) regression function signals positive (negative) dependence in the sense that large values of the times to failure of the failed component signal large (small) times to failure of the surviving component. Corresponding to every bivariate distribution, there exists a regression function, and vice versa, though not necessarily unique. Consequently, we would like to portray as many functional forms of the regression function as is possible. Gumbel's Type I & Type II (with $-1 \leq \alpha < 0$) encapsulate different forms of negative dependence, whereas Gumbel's Type II (with $0 < \alpha \leq 1$), Marshall & Olkin's bivariate exponential, and Singpurwalla & Youngren's bivariate distributions encapsulate different forms of positive dependence. However, not all forms of positive dependence are covered by the above candidates. Specifically, the (nondecreasing) regression functions corresponding to these distributions are either constant, or *concave* increasing to a constant asymptote. The asymptotic nature of these regression functions suggests that, in the opinion of an assessor, the time to failure of the surviving component is independent of the time to failure t_1 of the failed component, when t_1 is large. What therefore seems to be missing is a bivariate distribution with exponential marginals whose regression function is *convex* and increasing. Once such a distribution is identified, we would have a catalogue of bivariate distributions whose regression functions cover a spectrum of possible shapes which a user is able to employ. The purpose of the Section II-F is to propose one such family of bivariate distributions with exponential(1) marginals; i.e., exponential distributions with scale parameter one.

F. The Hazard Potential of Items

Before proceeding to the material of Section III, it is appropriate to comment on the fact that the bivariate distributions catalogued here have their parameters chosen so that the marginal distributions are exponential(1). This is because dependent life-times are the manifestation of what are known as dependent "hazard potentials", and it can be argued that the hazard potential of any component has an exponential(1) distribution; see [8]. The *hazard potential* of an item is interpreted as an unknown amount of resource possessed by the item upon its inception; the item fails when this resource gets exhausted. Using this argument, we can generate multivariate failure models having marginal distributions which are not exponential, using a multivariate distribution with exponential(1) marginals as a seed. The details are in [8].

III. Multivariate Exponentials With Convex and Increasing Regressions

In this section, we introduce a new family of multivariate distributions whose marginal distributions are exponential(1), and where regression functions (in the bivariate case) are convex & increasing. The former requirement is per the dictates of hazard potentials, and the latter is to fill a gap in our catalogue of shapes for regression functions.

A. The Notion of Copulas

There are many approaches for constructing multivariate distributions with specified marginals, one of which is the method of "copulas" [7]. Briefly, copulas are functions which couple univariate distribution functions to form multivariate distribution functions. In the two dimensional case, a *copula* is a function C which maps $[0,1] \times [0,1]$ to $[0,1]$ with the property that C is continuous and nondecreasing in each variable, and that $C(0,t) = C(t,0) = 0$ & $C(1,t) = C(t,1) = t$ for all t in $[0,1]$. The concept of copulas was initially used as a means for characterizing and measuring the dependence between two random variables. In particular, it can be seen that Spearman's Rho and Kendall's Tau are functions of their copulas alone, so

$$\mathcal{P}(T_1 \geq t_1, T_2 \geq t_2|m) = \begin{cases} \sqrt{\frac{1-mt_1+mt_2}{1+mt_1+mt_2}}\exp(-mt_2), & 0 \leq t_1 \leq t_2 \\ \sqrt{\frac{1-mt_2+mt_1}{1+mt_1+mt_2}}\exp(-mt_1), & 0 \leq t_2 \leq t_1, \end{cases} \tag{7}$$

that knowing the latter is equivalent to knowing the former [7]. By contrast, Pearson's correlation coefficient depends on both the copula, and the marginal distributions. The importance of copulas to this section can be seen from Sklar's [10] famous theorem.

Theorem 1 (Sklar): Let H be a two-dimensional distribution function with marginal distribution functions F & G. Then, there exists a copula C such that $H(x,y) = C(F(x), G(y))$. Conversely, for any univariate distribution functions F & G, and any copula C, the function H defined above is a two-dimensional distribution function with marginals F & G. Furthermore, if F & G are continuous, then C is unique.

Thus, in the bivariate case, Sklar's theorem plus the fact that $C(1,t) = C(t,1) = t$ implies that a copula is itself a bivariate distribution function with uniform marginals on $[0,1]$.

Another consequence of Sklar's theorem is that two random variables X & Y are independent, if and only if $C(u,v) = u \cdot v$, and X & Y are *positively quadrant dependent* (i.e., $\mathcal{P}(X \leq x, Y \leq y) \geq \mathcal{P}(X \leq x) \cdot \mathcal{P}(Y \leq y)$ for all x & y) if and only if $C(u,v) \geq u \cdot v$, for all u, $v \in [0,1]$; similarly with *negative quadrant dependence*.

B. Copula of a Bivariate Pareto and Its Induced Bivariate Exponential

The multivariate Pareto distribution, as a model for dependent life-times T_i, $i = 1, \ldots, n$, was proposed by Lindley and Singpurwalla [5]. This model is based on the premise that dependent life-times are a consequence of the action of a common but unknown environment on all components. When the unknown environment is characterized by a gamma distribution with a shape parameter $a > -1$ & a scale parameter $b > 0$, the resulting distribution of the T_i is a multivariate Pareto. In the bivariate case, this results in the form

$$\mathcal{P}(T_1 \geq t_1, T_2 \geq t_2 | a, b) = \left(\frac{b}{b + t_1 + t_2} \right)^{a+1}, \quad (8)$$

with marginal distribution functions

$$\mathcal{P}(T_i \geq t_i | a, b) = \left(\frac{b}{b + t_i} \right)^{a+1}. \quad (9)$$

Applying Sklar's theorem to the above yields the copula

$$C_a(u,v) = u + v - 1 + \left((1-u)^{-(a+1)} + (1-v)^{-(a+1)} - 1 \right)^{-(a+1)}. \quad (10)$$

Observe that it is only the shape parameter a which goes into determining this copula; the role of b is immaterial.

If we set $u = 1 - \exp(-t_1)$ & $v = 1 - \exp(-t_2)$, and use $C_a(u,v)$ for invoking Sklar's theorem in reverse, we produce the following bivariate distribution with exponential(1) marginals:

$$\mathcal{P}(T_1 \geq t_1, T_2 \geq t_2 | a) = \left(\exp\left(\frac{t_1}{a+1} \right) + \exp\left(\frac{t_2}{a+1} \right) - 1 \right)^{-(a+1)}. \quad (11)$$

The regression function $E_a(T_2 | T_1 = t_1)$ of the bivariate exponential of (11) does not exist in closed form. A plot of $E_a(T_2 | T_1 = t_1)$ with $a = 1$, evaluated numerically, is shown in Fig. 5; it is a convex, increasing function of t_1. Furthermore, a plot of $\ln E_a(T_2 | T_1 = t_1)$ is linear in t_1, suggesting that $E_a(T_2 | T_1 = t_1)$, for $a = 1$, grows exponentially in t_1.

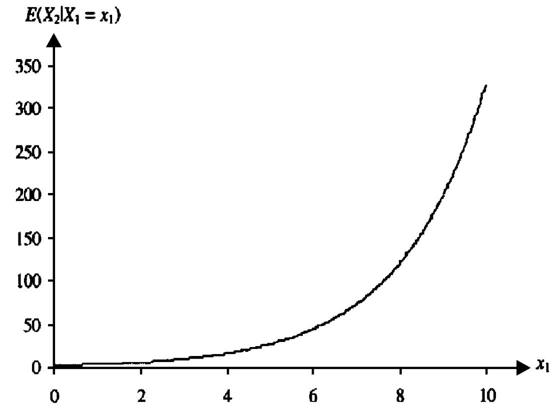

Fig. 5. A convex, increasing regression function generated by the new family of bivariate exponential distributions.

C. Invariance Properties of Copula Induced Regressions

The approach of Section III-B is general in the sense that we may start with any bivariate distribution with known marginals, and determine its copula. We then use this copula to generate a bivariate distribution with any desired marginals. In our case, we started with a bivariate Pareto as a seed, and using Sklar's theorem obtained its copula $C_a(u,v)$. We then used this $C_a(u,v)$ to invoke the converse of Sklar's theorem to induce a multivariate distribution with exponential(1) marginals, which is what we wanted. We shall refer to this process of producing the desired multivariate distributions as "copula induced transformation of a parent distribution". The question now arises as to whether the conditional mean of the induced distribution retains the general form of the conditional mean of the parent distribution. In our particular case we see that the regression function of the bivariate Pareto—$(b + t_1)/a$ [5]—is a linearly increasing function of t_1, whereas that of the bivariate exponential of (11) is, for $a = 1$, a convex increasing function of t_1. This observation leads us to claim that the conditional mean is *not* invariant under the copula induced transformations of the parent distribution.

Thus, unlike Spearman's Rho and Kendall's Tau, but like Pearson's correlation, the conditional mean, as a measure of dependence, cannot be characterized by a copula alone.

IV. Summary and Conclusions

It is well known [1] that for series systems, the assumption of independence (of life-times) underestimates system reliability when in actuality the assumption of positive dependence is germane; the reverse is true of parallel systems. But to capture the import of interdependence, one needs to specify a multivariate probability distribution for the component life-times. The best way to do this is to develop a probabilistic model which incorporates all aspects of the basic failure mechanisms. This could be a difficult task, because it calls for a deep appreciation of the physics of failure of each component. An alternative is to elicit the informed judgments of experienced individuals who may have views about the relationships between the failure times of the various components. It is with the above in mind that the material of this paper has been developed.

Specifically, we have advocated a use of the conditional mean as a measure of dependence. Our reason for advocating this choice is a naturalness of interpretation, and an ease of elicitation. Consequently, we have portrayed via Figs. 1–4, the conditional means of several bivariate exponential distributions whose marginal distributions are exponential(1). To complement this catalogue of conditional means, we have introduced, in Section III, a new family of multivariate exponential distributions, and have shown via Fig. 5, the general form of its conditional mean. We focus on multivariate distributions with exponential(1) marginals because such distributions can serve as a seed for generating multivariate life distributions having marginals that are other than the exponential.

So how can a practioner such as a network designer use the results of this paper? Our proposal is that the designer elicit a regression function from a panel of experts (the designer could also act as an expert), and to then match this function against the catalogue of regressions given here. The designer then chooses that probability model whose regression best matches the elicited regression. Clearly, this approach can work well when comparing any two component life-times. The problem becomes trickier when one has three or more components to contend with. For example, in the case of three life-times T_1, T_2, and T_3 one would have, as a process of pairwise elicitation, three bivariate distributions, one for (T_1, T_2), one for (T_2, T_3) and one for (T_1, T_3). How should one combine these bivariate distributions to obtain a trivariate distribution? This question remains to be addressed more carefully, but one possible strategy is to assume conditional independence, so that if T_1 is judged independent of T_3 given T_2, we would have

$$\begin{aligned}\mathcal{P}(T_1 \geq t_1, T_2 \geq t_2, T_3 \geq t_3) &= \mathcal{P}(T_1 \geq t_1 | T_2 \geq t_2, T_3 \geq t_3) \\ &\quad \cdot \mathcal{P}(T_2 \geq t_2 | T_3 \geq t_3) \cdot \mathcal{P}(T_3 \geq t_3) \\ &= \mathcal{P}(T_1 \geq t_1 | T_2 \geq t_2) \mathcal{P}(T_2 \geq t_2 | T_3 \geq t_3) \\ &\quad \cdot \mathcal{P}(T_3 \geq t_3),\end{aligned}$$

and *mutatis mutandis* if T_1 is independent of T_2 given T_3.

Conditional independence is commonly used when analyzing belief networks. However, it should be borne in mind that this too is a simplifying assumption, and as such is only one step away from complete independence per se. The matter of how to go beyond conditional independence remains to be addressed, and so is the matter of how to formally elicit regression functions. Gokhale and Press [2] have discussed methods of formally eliciting correlations, but as we have said before, correlations do not quite convey the intuitive import of the conditional mean.

REFERENCES

[1] R. E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing*. New York: Holt, Rinehart and Winston, 1975.

[2] D. V. Gokhale and S. J. Press, "Assessment of a prior distribution for the correlation coefficient in a bivariate normal distribution," *Journal of the Royal Statistical Association*, vol. 145, pp. 237–249, 1982.

[3] E. J. Gumbel, "Bivariate exponential distributions," *Journal of the American Statistical Association*, vol. 55, pp. 698–707, 1960.

[4] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions*. New York: John Wiley, 2000.

[5] D. V. Lindley and N. D. Singpurwalla, "Multivariate distributions for the life lengths of components of a system sharing a common environment," *J. Appl. Prob.*, vol. 23, no. 2, pp. 418–431, 1986.

[6] A. W. Marshall and I. Olkin, "A multivariate exponential distribution," *Journal of the American Statistical Association*, vol. 62, pp. 30–44, 1967.

[7] R. Nelsen, *An Introduction to Copulas*. New York: Springer, 1999.

[8] N. D. Singpurwalla, *The Hazard Potential of Items and Individuals*.

[9] N. D. Singpurwalla and M. A. Youngren, "Multivariate distributions induced by dynamic environments," *Scandinavian Journal of Statistics*, vol. 20, pp. 251–261, 1993.

[10] A. Sklar, "Random variables, distribution functions, and copulas—A personal look backward and forward," in *Distributions With Fixed Marginals and Related Topics*, Ruschendorff and Taylor, Eds. Hayward, CA: Institute of Mathematical Statcistics, 1996, pp. 1–14.

Nozer D. Singpurwalla is Professor of Statistics and Distinguished Research Professor at The George Washington University. He is Fellow of the American Statistical Association, the Institute of Mathematical Statistics, the American Association for the Advancement of Sciences, and an elected member of the International Statistical Institute. He has authored and coauthored over 170 papers and 2 books.

Chung-Wai Kong was a graduate student at the Institute for Reliability and Risk Analysis at The George Washington University.