

Using Copulas to Model Dependence in Simulation Risk Assessment

2007 ASME International Mechanical Engineering Congress and Exposition

Dana L. Kelly

November 2007

The INL is a
U.S. Department of Energy
National Laboratory
operated by
Battelle Energy Alliance



This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint should not be cited or reproduced without permission of the author. This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights. The views expressed in this paper are not necessarily those of the United States Government or the sponsoring agency.

IMECE2007-41284

USING COPULAS TO MODEL DEPENDENCE IN SIMULATION RISK ASSESSMENT

Dana L. Kelly

Idaho National Laboratory
P. O. Box 1625
Idaho Falls, ID 83415-3850
Dana.Kelly@inl.gov

ABSTRACT

Typical engineering systems in applications with high failure consequences such as nuclear reactor plants often employ redundancy and diversity of equipment in an effort to lower the probability of failure and therefore risk. However, it has long been recognized that dependencies exist in these redundant and diverse systems. Some dependencies, such as common sources of electrical power, are typically captured in the logic structure of the risk model. Others, usually referred to as intercomponent dependencies, are treated implicitly by introducing one or more statistical parameters into the model. Such common-cause failure models have limitations in a simulation environment. In addition, substantial subjectivity is associated with parameter estimation for these models. This paper describes an approach in which system performance is simulated by drawing samples from the joint distributions of dependent variables. The approach relies on the notion of a copula distribution, a notion which has been employed by the actuarial community for ten years or more, but which has seen only limited application in technological risk assessment. The paper also illustrates how equipment failure data can be used in a Bayesian framework to estimate the parameter values in the copula model. This approach avoids much of the subjectivity required to estimate parameters in traditional common-cause failure models. Simulation examples are presented for failures in time. The open-source software package R is used to perform the simulations. The open-source software package WinBUGS is used to perform the Bayesian inference via Markov chain Monte Carlo sampling.

INTRODUCTION

Tools for risk assessment are becoming more sophisticated as computer hardware and software continue their rapid advancement. Fault trees and event trees, while still prevalent, are quickly becoming the tools of the past. They are being replaced with tools such as discrete event simulation, which can capture finer details of the problem at hand. Consideration of uncertainty in the outputs of risk assessments is also important when these outputs are the inputs to decision-making. In even simple problems, there are often dependencies among the variables in the risk assessment that must be accounted for if the analysis is to yield an accurate assessment of uncertainty.

For example, typical engineering systems in applications with high failure consequences such as nuclear reactor plants often employ redundancy and diversity of equipment in an effort to lower the probability of failure and therefore risk. However, it has long been recognized that dependencies exist in these redundant and diverse systems. Some dependencies, such as common sources of electrical power, are typically captured in the logic structure of the risk model. Others, usually referred to as intercomponent dependencies, are treated implicitly by introducing one or more statistical parameters into the model. Such common-cause failure models have limitations, such as inability to model common-cause failure among diverse (as opposed to redundant and identical) components. In addition, substantial subjectivity is associated with parameter estimation for these models.

This paper describes an approach in which system performance is simulated by drawing samples from the joint distributions of dependent variables. The approach relies on

the notion of a copula distribution, a notion which has been employed by the actuarial community for ten years or more, but which has seen only limited application in technological risk assessment. The paper also illustrates how equipment failure data can be used in a Bayesian framework to estimate the parameter values in the copula model. This approach avoids much of the subjectivity required to estimate parameters in traditional common-cause failure models. Simulation examples are presented for failures in time. The open-source software package R (www.r-project.org) is used to perform the simulations. The open-source software package WinBUGS (<http://mathstat.helsinki.fi/openbugs/Home.html>) is used to perform the Bayesian inference via Markov chain Monte Carlo sampling.

NOMENCLATURE

cdf : cumulative distribution function
 $C(u, v)$: bivariate copula distribution function
 $F(\bullet)$: cumulative distribution function
 $F^{-1}(\bullet)$: inverse cumulative distribution function
 $\log(\bullet)$: log to base e
 $P(\bullet)$: probability of an event
 u, v : uniform(0, 1) random variables

DEFINITION OF COPULA FUNCTION

The basic problem to be examined in this paper is simulation of component failure times from a joint distribution in the general case in which the failure times are not independent. The exposition begins with the simplest case in which there are two components arranged in parallel, and generalizes this later to the case of three or more components.

With two components in a parallel arrangement, with a 1-of-2 success criterion, failure is defined as the event in which neither component operates for the required period of time, denoted t_m for mission time. In terms of the joint cumulative distribution function, one can write

$$P(\text{failure}) = F(t_m, t_m) \quad (1)$$

Because the failure times are dependent, the joint distribution function cannot be factored as the product of the marginal distributions. In a simulation environment, one randomly samples pairs of failure times from the joint distribution, and then estimates the probability of system failure as the fraction of samples in which both failure times are less than or equal to t_m . In general, the marginal distributions will not be Gaussian; typically they will be exponential, Weibull, gamma, lognormal, etc. And they need not be identical. This complicates the process of drawing Monte Carlo samples from the joint distribution.

The notion of a copula function is extremely useful for simulation, as it avoids the need for assumptions of normality and, most importantly, allows the analyst to proceed in two separate stages: 1) model the one-dimensional marginal

distributions, and 2) model the dependence structure (1). This is in contrast to the frailty approach, another popular approach to dependence, in which the failure times are modeled as being conditionally independent, given a frailty variable. A copula function can be viewed as generalizing to two or more dimensions the well known result that the cumulative distribution function (cdf) of a random variable X , denoted $F(x)$, is uniformly distributed on the interval $[0, 1]$. This is the basis for simple Monte Carlo sampling routines that use the inverse cdf to obtain samples from a specified distribution.

Two random variables T_1 and T_2 are joined by a copula function, C , if their joint cdf can be written as

$$F(t_1, t_2) = C[F_1(t_1), F_2(t_2)] \quad (2)$$

This result is known as Sklar's Theorem. As shown in (2), every bivariate distribution has a copula representation, which is unique if the variables are continuous, and given a copula function, along with marginal distributions F_1 and F_2 , the function obtained by eq. 2 is the joint cdf. Also, if the joint and marginal distributions are known, the copula function can be written as

$$C(u, v) = F[F_1^{-1}(u), F_2^{-1}(v)] \quad (3)$$

In eq. 3, u and v are uniformly distributed on the interval $[0, 1]$. Thus, the copula function is a distribution function defined on the unit square, possessing uniform marginals.

Many different copulas are available to express dependence between marginal distributions. A partial list is shown in (2). This paper will illustrate two choices, beginning with the Marshall-Olkin copula, which is derived from the Marshall-Olkin shock model. The second is Frank's copula, which is not based on a shock model, and has great flexibility along with a convenient closed functional form.

MARSHALL-OLKIN COPULA

As the first specific example, consider the bivariate Marshall-Olkin model, described in (3). In this model there are three shocks. Shock 1 occurs at time S_1 and fails component 1. Shock 2 occurs at time S_2 and fails component 2. Shock 3 occurs at time S_{12} and fails both components. These three shocks are assumed to be independently distributed as follows:

$$\begin{aligned} S_1 &\sim \exp(\lambda_{11}) \\ S_2 &\sim \exp(\lambda_{21}) \\ S_{12} &\sim \exp(\lambda_{12}) \end{aligned}$$

The following result is easily obtained for the failure times of components 1 and 2:

$$P(T_1 > t_1, T_2 > t_2) = \exp[-\lambda_{11}t_1 - \lambda_{21}t_2 - \lambda_{12}\max(t_1, t_2)] \quad (4)$$

Thus, one has the following result for the failure time of each component.

$$T_1 = \min(S_1, S_{12}) \sim \exp(\lambda_{11} + \lambda_{12}) \quad (5)$$

$$T_2 = \min(S_2, S_{12}) \sim \exp(\lambda_{21} + \lambda_{12}) \quad (6)$$

One can use eq. 4 to find the joint cdf:

$$F(t_1, t_2) = \bar{F}(t_1, t_2) + F_1(t_1) + F_2(t_2) - 1 \quad (7)$$

Because of Sklar's Theorem, eq. 2, the bivariate exponential distribution defined in this way implies a corresponding (unique) bivariate copula function. To obtain the copula function, let $u = 1 - F_1(t_1)$ and $v = 1 - F_2(t_2)$. Furthermore, let

$$\alpha = \frac{\lambda_{12}}{\lambda_{11} + \lambda_{12}} \quad \text{and}$$

$$\beta = \frac{\lambda_{12}}{\lambda_{21} + \lambda_{12}}$$

Then, as is shown in (2), the copula function is given by

$$C(u, v) = \begin{cases} u^{1-\alpha} v, & u^\alpha \geq v^\beta \\ uv^{1-\beta}, & u^\alpha \leq v^\beta \end{cases} \quad (8)$$

To generate pairs of failure times using eq. 8, one can use the following algorithm. First, generate three independent, uniform(0, 1) random variables: r , s , and z . Then, generate t_1 and t_2 according to the following:

$$\begin{aligned} t_1 &= \min[-\log(r)/\lambda_{11}, -\log(z)/\lambda_{12}] \\ t_2 &= \min[-\log(s)/\lambda_{21}, -\log(z)/\lambda_{12}] \end{aligned} \quad (9)$$

One can then show that u and v , given by the following equations, are uniform(0, 1) random variables, whose joint distribution is the Marshall-Olkin copula defined by eq. 8.

$$\begin{aligned} u &= \exp[-(\lambda_{11} + \lambda_{12})t_1] \\ v &= \exp[-(\lambda_{21} + \lambda_{12})t_2] \end{aligned} \quad (10)$$

Being a shock model, the bivariate Marshall-Olkin distribution has a singular component. In terms of simulation, this means that one can expect to find $t_1 = t_2$ in a proportion of the samples generated from this distribution. One can show that this proportion is given by

$$P(u^\alpha = v^\beta) = \frac{\alpha\beta}{\alpha + \beta - \alpha\beta} = \frac{\lambda_{12}}{\lambda_{11} + \lambda_{21} + \lambda_{12}} \quad (11)$$

The figure below shows a plot of 500 values from a Marshall-Olkin copula with $\alpha = 0.5$ and $\beta = 0.75$. The singular component (where $t_1 = t_2$) is clearly visible.

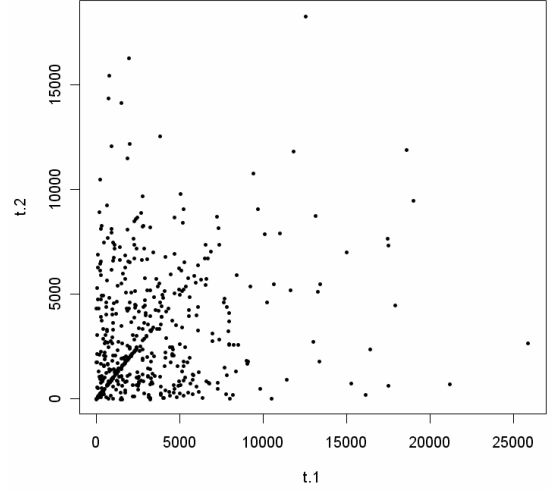


Fig. 1 Plot of 500 samples from Marshall-Olkin copula illustrating singular component

FRANK'S COPULA

In some situations, the shock model may not be appropriate. For example, the failure mechanism may be gradual in nature, so that dependence between the failure times arises more from shared conditions than from a common shock. In these cases, the Marshall-Olkin model is inappropriate. However, other copula functions exist which are absolutely continuous and thus may be appropriate models. One such model that is extremely flexible is Frank's copula. As described in (1, 2), Frank's copula is in the family of so-called Archimedean copulas, and is defined by one parameter, denoted θ . It has a closed functional form, which is convenient for simulation, and the copula parameter, θ , can be related to various popular measures of bivariate dependence, such as Spearman's rank correlation coefficient, ρ_r . Fig. 2 below shows how Spearman's rank correlation coefficient varies with increasing values of θ . This relation is symmetric for negative values of θ , with ρ_r being negative in that case.

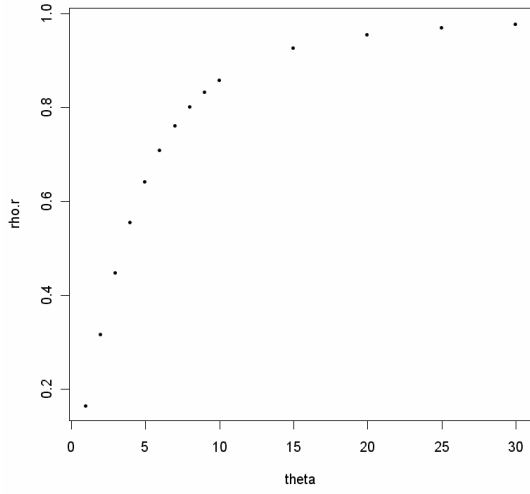


Fig. 2 Plot of Spearman rank correlation coefficient vs. parameter of Frank's copula

The functional form of Frank's copula is given by

$$C(u, v; \theta) = -\frac{1}{\theta} \log \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right] \quad (12)$$

In eq. 12, $u = F_1(t_1)$ and $v = F_2(t_2)$.

Because Frank's copula is absolutely continuous, one can write the bivariate copula density as

$$\begin{aligned} c(u, v) &= \frac{\partial^2 C(u, v)}{\partial u \partial v} \\ &= \frac{\theta(1 - e^{-\theta})e^{-\theta(u+v)}}{[1 - e^{-\theta} - (1 - e^{-\theta u})(1 - e^{-\theta v})]^2} \end{aligned} \quad (13)$$

In the limit as $\theta \rightarrow 0$, the failure times are independent and the copula reduces to the independent copula, $C(u, v) = uv$. For $\theta > 0$, the failure times exhibit positive quadrant dependence (3), with $\rho_r > 0$. For $\theta < 0$, the failure times exhibit negative quadrant dependence (3) and $\rho_r < 0$. The following figures illustrate the copula density function for selected values of θ .

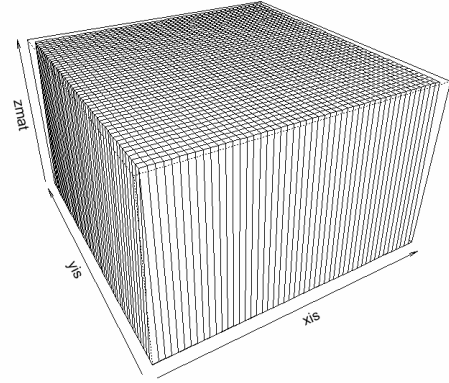


Fig. 3 Density function of Frank's copula with $\theta = 0$ (independence)

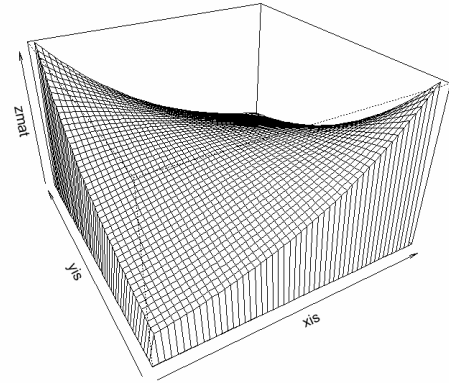


Fig. 4 Density function of Frank's copula with $\theta = -2$ (negative quadrant dependence, $\rho_r < 0$)

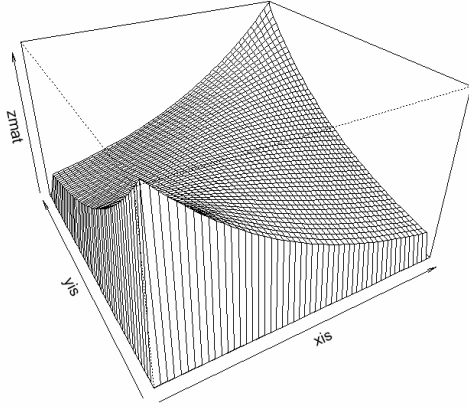


Fig. 5 Density function of Frank's copula with $\theta = 2$ (positive quadrant dependence, $\rho_r > 0$)

To generate failure times from a bivariate Frank's copula, one can use the following algorithm, adapted from (4). First generate u_1 and u_2 from independent uniform(0, 1) distributions. Then, set $t_1 = F_1^{-1}(u_1)$. Calculate t_2 as the solution to the following equation.

$$u_2 = e^{-\theta u_1} \left(\frac{e^{-\theta} - e^{-\theta F_2(t_1)}}{e^{-\theta F_2(t_1)} - 1} + 1 \right)^{-1} \quad (14)$$

That is, $t_2 = F_x^{-1}(u_{*2})$ where u_{*2} is given by

$$u_{*2} = \frac{u_2 e^{-\theta} - e^{-\theta u_1} (1 - u_2)}{u_2 + e^{-\theta u_1} (1 - u_2)} \quad (15)$$

To illustrate the use of copulas for simulation, consider the following example. Assume a portion of a system contains two redundant components, both of which have a marginal failure time distribution that is exponential with rate 2×10^{-4} per hour. Assume that the mission time is 1000 hours. Failure of the system is given by the joint cdf, evaluated at the mission time of 1000 hours. If the components were independent, this would just be the product of the marginal cdfs. However, in the case of dependence, the joint cdf cannot be factored into the product of the marginals. Assume that the dependence can be represented by Frank's copula with exponential marginals and $\theta = 10$. Using the relationship between θ and ρ_r illustrated in fig. 2 and specified in (2) in terms of numerical integrals known as Debye functions, one can show that this corresponds to a rank correlation between the failure times of 0.86.

The R package was used to generate 100 samples from the bivariate exponential distribution (via Frank's copula) in this case, resulting in the figure below.

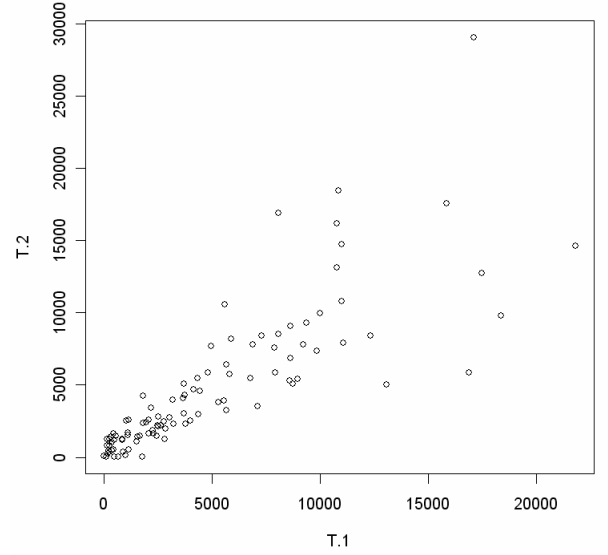


Fig. 6 Scatterplot of 100 failure times from Frank's copula with $\theta = 10$ illustrating dependence between failure times and lack of singular component (no pairs of times are equal)

Expanding the lower region of this figure shows the cases in which both simulated failure times were less than the mission time of 1000 hours, indicating failure.

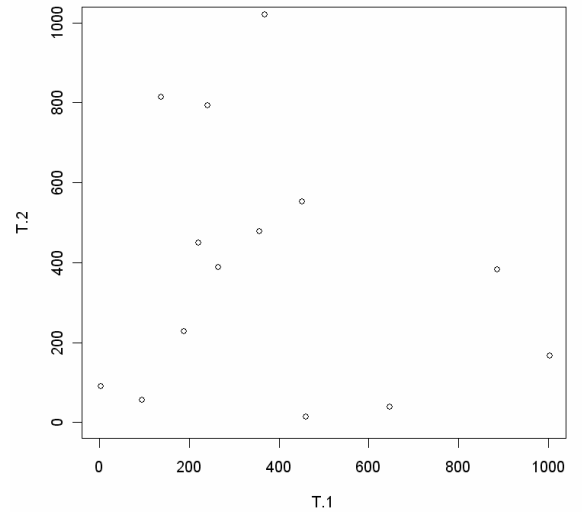


Fig. 7 Magnified region showing 14 failure times less than 1000 hours

Thus, the estimated joint failure probability would be 0.14. Of course, 100 samples are not sufficient to estimate the failure

probability with high accuracy. A run of 100,000 samples resulted in an estimated failure probability of 0.12. This can be compared to a probability of about 0.03 when the failure times are independent.

BAYESIAN ESTIMATION OF COPULA PARAMETERS

The inverse of this problem is also important. In this case, one is presented with pairs of failure times and the problem is to estimate the bivariate distribution, capturing any dependence exhibited by the failure times. The Bayesian approach to this problem is illustrated here, for both the Marshall-Olkin shock model and the absolutely continuous Frank's copula. The WinBUGS package is used to carry out the analysis via Markov chain Monte Carlo sampling.

A key strength of the copula approach is that it allows the estimation to proceed in two stages. First, the marginal distributions are estimated, assuming the failure times are independent. Second, the dependence structure is estimated, which corresponds to estimating the parameter(s) of the copula function.

Assuming that the failure times have marginal exponential distributions, one first specifies a prior distribution for each of the failure rates, and then uses the observed failure times to update these rates. This is a straightforward process, as described for example in (5). Jeffreys noninformative priors were used for each of the failure rates.

With the posterior distributions for the failure rates in hand, the next step is to find the posterior distribution for the copula parameter. For Frank's copula, a reasonable prior belief for a system with redundant components is that $\theta > 0$, representing positive quadrant dependence between the failure times. A value of 50 for θ corresponds to ρ_r of 0.99, very near the maximum value of 1. Therefore, a uniform distribution between 0 and 50 was used as a minimally informative prior for θ . The joint density for the failure times, which is the likelihood function for Bayesian inference, can be written as

$$f(t_1, t_2) = f_1(t_1)f_2(t_2)c[F_1(t_1), F_2(t_2)] \quad (16)$$

In eq. 16, one has $f_i(t_i) = \lambda_i \exp(-\lambda_i t_i)$ and c is the copula density, given by eq. 13. The "zeros trick" described in the WinBUGS user manual is used to encode this likelihood function into WinBUGS. This trick entails assigning a Poisson distribution to a vector of zeros, with parameter (Poisson mean) equal to the negative of the logarithm of the likelihood function. The WinBUGS script is shown in the table below.

```
model {
#Estimate parameters of marginal distributions
for(i in 1:N) {
  t.1[i] ~ dexp(lambda.11)
  t.2[i] ~ dexp(lambda.21)
}
#Set up "zeros trick"
```

```
for(i in 1:N) {
  zeros[i] <- 0
  zeros[i] ~ dpois(phi[i])
#Following lines calculate likelihood with exponential
marginals
  u[i] <- 1 - exp(-lambda.11*t.1[i])
  v[i] <- 1 - exp(-lambda.21*t.2[i])
  phi[i] <- -log(lambda.11) - log(lambda.21) + lambda.11*t.1[i]
+ lambda.21*t.2[i] - log(theta) - log(1 - exp(-theta)) +
theta*(u[i] + v[i]) + 2*log(1 - exp(-theta)) - (1 - exp(-
theta*u[i]))*(1 - exp(-theta*v[i]))
}
theta ~ dunif(0,50) #Uniform prior on copula parameter
lambda.11 ~ dgamma(0.0001, 0.0001) #Jeffreys prior on
marginal parameters
lambda.21 ~ dgamma(0.0001, 0.0001)
```

Tab. 1 WinBUGS script for estimating parameters of joint distribution of failure times using Frank's copula to model dependence

The R package was used to generate 50 pairs of failure times from Frank's copula with $\theta = 10$, corresponding to $\rho_r = 0.86$. The marginal distributions were both exponential with rate 2×10^{-4} per hour. These simulated failure times were used to test the WinBUGS script. Two Markov chains were used and were run until convergence was achieved, after 2,000 iterations. Another 20,000 iterations were used to estimate the marginal failure rates and the copula parameter, θ . The two marginal failure rates both had a posterior mean of 1.7×10^{-4} , with a 90% interval of $(1.4, 2.0) \times 10^{-4}$. The posterior mean of θ was found to be 9.3, with a 90% interval of (6.97, 11.79). The posterior distribution of θ was approximately Gaussian, with a standard deviation of 1.5.

As a second test, a set of 50 independent failure times ($\rho_r = -0.003$) was generated, each from an exponential distribution with rate 2×10^{-4} per hour. The posterior mean of θ was found to be 0.63, with a 90% interval of (0.05, 1.57). Recall that variables joined by Frank's copula are independent in the limit $\theta \rightarrow 0$, so estimated values of θ near zero are suggestive of independence..

Turning now to Bayesian estimation for the bivariate Marshall-Olkin distribution, there are again three parameters to estimate: the failure rates of the marginal exponential distributions, denoted here as λ_{11} and λ_{21} , and the rate of the common shock, denoted λ_{12} . Jeffreys noninformative priors were used for each of these rates. The likelihood function has three pieces, corresponding to cases where $t_1 > t_2$, $t_1 < t_2$, and $t_1 = t_2$. The WinBUGS script is shown in the table below. The "zeros trick" is again used to encode the likelihood function.

```
model {
for(i in 1:N) {
  zeros[i] <- 0
  zeros[i] ~ dpois(phi[i])
```

```

logL.1[i] <- log(lambda.21*(lambda.11 +
lambda.12))*step(t.1[i] - t.2[i]) - log(lambda.21*(lambda.11 +
lambda.12))*equals(t.1[i], t.2[i]) + log(lambda.11*(lambda.21
+ lambda.12))*step(t.2[i] - t.1[i]) - log(lambda.11*(lambda.21 +
lambda.12))*equals(t.2[i], t.1[i])
logL.2[i] <- - lambda.11*t.1[i] - lambda.21*t.2[i] -
lambda.12*t.1[i]*step(t.1[i] - t.2[i]) +
lambda.12*t.1[i]*equals(t.1[i], t.2[i]) -
lambda.12*t.2[i]*step(t.2[i] - t.1[i]) +
lambda.12*t.2[i]*equals(t.2[i], t.1[i])
logL.3[i] <- log(lambda.12)*equals(t.1[i], t.2[i]) -
lambda.12 *t.1[i]*equals(t.1[i], t.2[i])
logL[i] <- logL.1[i] + logL.2[i] + logL.3[i]
phi[i] <- -logL[i]
}
#Priors
lambda.11 ~ dgamma(0.0001, 0.0001) #Diffuse priors
lambda.21 ~ dgamma(0.0001, 0.0001)
lambda.12 ~ dgamma(0.0001, 0.0001)
#Parameters of Marshall-Olkin copula
alpha <- lambda.12/(lambda.11 + lambda.12)
beta <- lambda.12/(lambda.21 + lambda.12)
}

```

Tab. 2 WinBUGS script for estimating parameters of bivariate Marshall-Olkin joint distribution of failure times

The R package was used to generate 50 pairs of failure times from a bivariate Marshall-Olkin distribution with $\lambda_{11} = \lambda_{21} = 2 \times 10^{-4}$ per hour and $\lambda_{12} = 5 \times 10^{-5}$ per hour. These simulated failure times were used to test the WinBUGS script. One thousand WinBUGS iterations were run for convergence, followed by 20,000 iterations to estimate the parameters. The posterior mean of λ_{11} and λ_{21} was 2.0×10^{-4} with a 90% interval of $(1.5, 2.6) \times 10^{-4}$. The posterior mean of λ_{12} was 4.8×10^{-5} with a 90% interval of $(2.1, 8.4) \times 10^{-5}$. The posterior means and 90% intervals for the parameters of the Marshall-Olkin copula (α and β) were as follows.

α : 0.19 (0.08, 0.33)
 β : 0.18 (0.08, 0.31)

MODELING STATE-OF-KNOWLEDGE DEPENDENCE

As first discussed in (6), Monte Carlo sampling for propagation of uncertainties in a risk model should account for state-of-knowledge dependencies among parameters in the model. For example, in a system with redundant, nominally identical components, a single probability distribution typically is developed to represent epistemic uncertainty in a desired parameter value (e.g., failure rate), and this distribution is applied to each of the redundant components. This represents complete state-of-knowledge dependence between the redundant components, and is reflected in the Monte Carlo uncertainty propagation by sampling a single value from this

distribution at each iteration of the Monte Carlo sampling and applying this single value to each of the redundant components. This is the opposite extreme from independence, in which separate values would be sampled for each component at each iteration.

There are cases where an intermediate level of state-of-knowledge dependence is desired. For example, a system may have redundant, but nonidentical components, for which separate probability distributions are developed to represent parameter uncertainty. However, state-of-knowledge independence may not be justifiable. Fragilities of components in seismic analysis are one example. In these cases, copulas provide an easy means to carry out the Monte Carlo sampling. Frank's copula is attractive for this application because the copula parameter can be related to Spearman's rank correlation coefficient, which is a useful means of expressing the state-of-knowledge dependence, as shown below.

One way to attach a meaning to Spearman's rank correlation, ρ_r , is to use the following nonparametric regression expression.

$$E[F(X|y)] = \rho_r[G(y) - 0.5] + 0.5 \quad (17)$$

In this equation, $G(y)$ is the cdf of Y , and $F(X|y)$ is the cdf of X , conditional upon a value of y . So the equation relates the expected percentile of X , given y , to the unconditional percentile of y . Assume that a value of y is sampled, and it is at the 80th percentile, so $G(y) = 0.8$. If $\rho_r = 0.5$, this means we would expect to sample a value of X corresponding to the 65th percentile. If $\rho_r = 1$, we get $E[F(X|y)] = G(y)$, as expected (the two values are equal). This explains why the distributions representing epistemic uncertainty have to be identical if there is complete correlation. If $\rho_r = 0$, corresponding to independence, we get $E[F(X|y)] = 0.5$, which is what we expect, because in this case $F(X|y) = F(X)$, which has a uniform(0, 1) distribution and an expected value of 0.5.

As an example, consider an AND gate in a fault tree with two inputs, each of which represents failure of a redundant component to start on demand. Assume that the epistemic uncertainty distribution for the failure-to-start probability, p , is beta(0.957, 190). If no state-of-knowledge correlation exists between the two components, then the mean probability from the AND gate is 2.3×10^{-5} . With complete dependence, the probability increases to 4.7×10^{-5} . Using a two-dimensional Frank's copula with beta(0.957, 190) marginal distributions to represent intermediate dependence, we ran 100,000 Monte Carlo samples to obtain the plot below, which shows how the AND gate probability varies with Spearman's rank correlation, ρ_r .

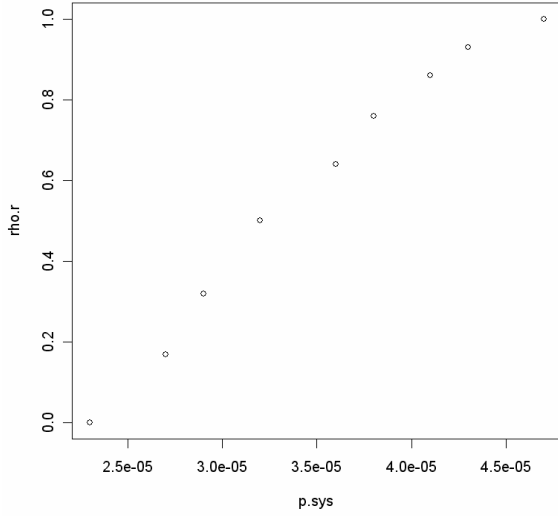


Fig. 8 Plot showing how mean probability of failure of two redundant components increases with increasing state-of-knowledge dependence

MULTIVARIATE EXTENSION

As shown in (2), Frank's copula can be extended to n dimensions in a straightforward manner for $\theta > 0$. For n dimensions, the copula function becomes

$$C(u_1, u_2, \dots, u_n) = -\frac{1}{\theta} \log \left[1 + \frac{\prod_{i=1}^n (e^{-\theta u_i} - 1)}{(e^{-\theta} - 1)^{n-1}} \right] \quad (18)$$

The n -dimensional copula is still specified by the single parameter, θ , which determines the correlation between each pair of variables. As there is only one parameter, each pair of variables has the same correlation coefficient, so that the correlation matrix has a 1 for each diagonal entry, and ρ_i for each off-diagonal entry.

As an example, consider a portion of a system containing three components arranged in parallel. Assume that two of these components are redundant (i.e., identical) and the other is diverse. Assume that the redundant components are exponentially distributed with a failure rate of 5×10^{-5} per hour, while the diverse component has a failure rate that increases with operating time. Assume that the mean time to failure for this diverse component is 5,000 hours and that the failure time distribution is gamma with a shape parameter of 2, giving the desired increase in failure rate with time. Assume that the dependence between each pair of components is moderate, expressed by a rank correlation of about 0.3, which corresponds to a parameter of about 2 for Frank's copula. With a mission time of 1,000 hours, the joint failure probability for this group of three parallel components is estimated to be 6.4×10^{-4} , based

on 10^6 samples from a Frank's copula with the specified marginals and $\theta = 2$. This is compared with a joint failure probability assuming independence of 1.5×10^{-4} .

One can also use dependence trees, as described in (1, 7), to specify a more complex bivariate dependence structure, in which the pair-wise correlations are not identical. In this approach, the user can specify rank correlations between each pair of nodes in the dependence tree, and then bivariate copulas are used to sample each pair of nodes in a conditional fashion.

As an example, consider the dependence tree shown in the figure below. In this example, there are again three components, each of which has the marginal distributions for time to failure listed above. However, the rank correlations are not identical. Rank correlation coefficients are specified between t_1 and t_2 (the redundant components) and between t_1 and t_3 , where component 3 is the diverse component. The sampling proceeds as follows, assuming that a bivariate Frank's copula joins each of these pairs.

First, generate a uniform random variable u_1 and set

$t_1 = F_1^{-1}(u_1)$. Next sample u_2 , conditional on u_1 , from Frank's copula with uniform marginals and rank correlation 0.6 ($\theta \approx 5$) and set $t_2 = F_2^{-1}(u_2)$. Generate t_3 similarly using the appropriate rank correlation between t_1 and t_3 . The UNICORN package, described in (1) is very handy for this type of analysis, although such analysis can also be programmed with the R package.

If failure is defined as all three components running for less than 1,000 hours, we find an estimated failure probability, based on 10^6 samples, of about 1.2×10^{-3} . This can be compared with the joint failure probability of 1.5×10^{-4} in the case of mutual independence. Note that in both of the cases just illustrated, as the rank correlation goes to unity, the joint failure probability approaches the minimum of the marginal failure probabilities, as a consequence of the Frechét-Hoeffding upper bound (2). In this case this limit is about 0.05, the marginal failure probability for either of the redundant, exponentially distributed components.

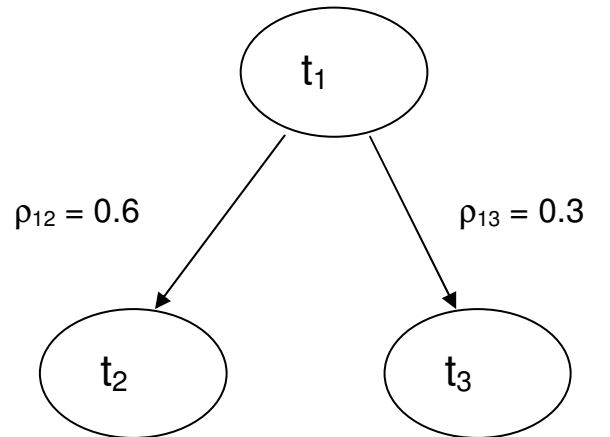


Fig. 9 Dependence tree for three exponential times to failure illustrating pair-wise rank correlations

CONCLUSIONS

This paper has illustrated the use of copulas for generating values from joint distributions of dependent random variables. Both shock and non-shock failure time models have been illustrated. The bivariate Marshall-Olkin distribution was used to illustrate a shock model. The corresponding copula has a singular component, producing samples in which the two failure times are equal. For non-shock models, the absolutely continuous Frank's copula was employed. Bayesian inference for the copula parameters was illustrated for both cases. In the case of Frank's copula, a connection was illustrated between the copula parameter and the rank correlation, which is a commonly used measure of dependence. Guidance for eliciting this correlation from domain experts can be found in (8). It is also possible to use other measures of dependence, as discussed in (9).

Examples were provided illustrating the use of these copulas in a simulation setting, where the desired parameter is the joint failure probability of two or more redundant or diverse components. The examples provided were in terms of a continuous random variable (time to failure); however, the same approach can be applied to failures on demand. The use of copulas in this fashion extends the common-cause failure models of risk assessment to diverse (i.e., non-identical) components arranged in parallel, and has the potential to replace models of common-cause failure currently in use.

ACKNOWLEDGMENTS

I have had the benefit of helpful discussions with Curtis Smith and Bill Galyean at Idaho National Laboratory. Martin Sattison at Idaho National Laboratory provided funding for presenting this work.

REFERENCES

- [1] Kurowicka, Dorota and Cooke, Roger, Uncertainty Analysis with High Dimensional Dependence Modelling, John Wiley & Sons, Ltd., 2006.
- [2] Nelsen, Roger B., An Introduction to Copulas, Second Edition, Springer, 2006.
- [3] Barlow, Richard E. and Proschan, Frank, Statistical Theory of Reliability and Life Testing: Probability Models, McArthur Press, 1981 (reprint of 1975 edition, with corrections).
- [4] Frees, Edward W. and Valdez, Emiliano A., "Understanding Relationships Using Copulas," North American Actuarial Journal, Vol. 2, No. 1, 1998.
- [5] Siu, Nathan O. and Kelly, Dana L., "Bayesian Parameter Estimation in Probabilistic Risk Assessment," Reliability Engineering and System Safety, **62** (1998), pp. 89-116.
- [6] Apostolakis, G. and Kaplan, S., "Pitfalls in Risk Calculations," Reliability Engineering, **2** (1981), pp. 135-145.
- [7] Bedford, Tim and Cooke, Roger, Probabilistic Risk Analysis: Foundations and Methods, Cambridge University Press, 2001.

[8] Clemen, Robert T. and Reilly, Terence, "Correlations and Copulas for Decision and Risk Analysis," Management Science, Vol. 45, No.2, February 1999.

9. Ferson, Scott and Hajagos, Janos G., "Varying Correlation Coefficients Can Underestimate Uncertainty in Probabilistic Models," Reliability Engineering and System Safety, **91** (2006), pp. 1461-146