Chapter 3
Input Parameter Distributions

In this chapter we will explore how we can use the principles of statistics and probability to model input parameters to simulation models. This discussion will require that we understand how random variables depend on each other, how we can model this dependence when we have limited information, and how we can approximate a collection of random variables, or even a random process, based on some underlying structure.

In a computer simulation there will be typically be several random variables as inputs. For a collection of random variables it is common to not have an expression for the joint distribution functions (CDF or PDF) for the collection. Rather, the best one can do is hope to have some measure of the dependence between the pairs of variables. As we will see, the dependence measures we use are not enough to uniquely determine the relationship between random variables.

Additionally, later when we try to model the distribution of output quantities of interest based on input uncertainties, we will see that the number of random variables we have as input determine the accuracy we can achieve with our uncertainty quantification given a fixed computational budget. Therefore, we would like to determine if we can eliminate input random variables if there is an underlying correlation or approximation. Methods for this type of reduction will be discussed in this chapter as well.

3.1 Dependence Between Variables

So far we have discussed probability distributions and multivariate distributions in some detail. For collections of random variables, we are often interested in how they vary together. We already have a measure for this: the covariance. One issue with the covariance between two random variables, X and Y,

$$\Sigma(X,Y) = E[XY] - E[X]E[Y], \tag{3.1}$$

is that it has units that are the product of the units of X and Y. This can make it difficult to compare covariances. For instance, $\Sigma(X,Y) > \Sigma(X,Z)$ does not imply that there is a stronger relationship between X and Z than X and Y because of the units.

3.1.1 Pearson Correlation

A normalized measure of the relation between two random variables, is the Pearson correlation coefficient, ρ. Oftentimes, this is simply called the correlation coefficient or correlation. Considering two random variables, X, and Y, the correlation coefficient is

$$\rho(X,Y) = \frac{E[XY] - E[X]E[Y]}{\sigma_X \sigma_Y}. \tag{3.2}$$

That is, the Pearson correlation is the covariance normalized by the standard deviation of each variable. On this normalized scale, we can say things about how two variables change together. If the variables are independent, then $\rho(X,Y) = 0$. As with covariance, a correlation of zero between variables does not imply that the variables are independent.

One property of the correlation coefficient is that if X and Y are linearly related, i.e., there exist an $a > 0$ and b such that $Y = aX + b$, then $\rho(X,Y) = 1$. As a corollary, we have the relation

$$\rho(X,Y) = \text{sign}(a)\rho(aX+b,Y).$$

When we have a collection of random variables, $\mathbf{X} = (X_1, X_2, \ldots, X_p)^{\mathrm{T}}$, we can define a correlation matrix $\mathbf{R}$ in terms of the covariance matrix as

$$R_{ij} = \frac{\Sigma_{ij}}{\sigma_{X_i}\sigma_{X_i}}, \tag{3.3}$$

where $\sigma^2_{X_i} = \Sigma(X_i, X_i)$ is the variance in X_i.

The benefit of the Pearson correlation coefficient is that it is easy to calculate, as simple as the covariance matrix. However, there are some downsides. One is that it is not defined if the expected value of XY is not defined. This is the case with Cauchy random variables, given by the PDF with parameters x_0, γ,

$$f(X) = \frac{1}{\pi\gamma}\left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]^{-1}. \tag{3.4}$$

The mean and variance of the distribution are undefined because the distribution goes to zero too slowly, but the median and mode are x_0. The PDF for a Cauchy distribution, and it's comparison to the standard normal, are given in Figure 3.1.

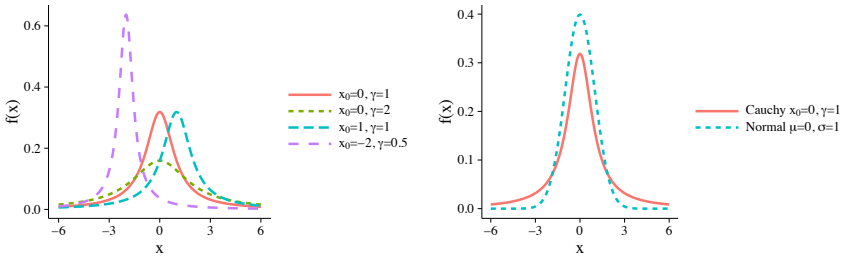

Fig. 3.1 The Cauchy distribution with various parameters and compared with the standard normal.

Another, potentially more important, downside of the Pearson correlation coefficient is that if X is transformed by a nonlinear, strictly increasing function, $g(X)$, the correlation $\rho(X,Y)$ will be different than $\rho(g(X),Y)$. This means that if there is a nonlinear relation between X and Y the Pearson correlation coefficient may under- or over-estimate the relation between the two variables.

3.1.2 Spearman Rank Correlation

An alternative to the Pearson correlation is the Spearman rank correlation, or Spearman correlation. In this measure we look for general, monotonic relationships between two variables. This is defined by looking at the correlation between the marginal CDF of each variable:

$$\rho_{\mathrm{S}}(X,Y) = \rho(F_X(x), F_Y(y)). \tag{3.5}$$

If we do not know the marginal CDF, but we have samples of the random variables we can still estimate the Spearman correlation. Given N samples of X and Y we create a function that takes sample x_i or y_i and gives the rank of the that sample amongst the N samples:

$$\mathrm{rank}(x_i) = \text{Rank of } x_I \text{ in sample population.}$$

Using this function we then define the Spearman correlation coefficient for the samples:

$$\rho_{\mathrm{S}}(X,Y) = \frac{\sum_{i=1}^{N}(\mathrm{rank}(x_i) - \bar{r}_X)(\mathrm{rank}(y_i) - \bar{r}_Y)}{\sqrt{\sum_{i=1}^{N}(\mathrm{rank}(x_i) - \bar{r}_X)^2}\sqrt{\sum_{i=1}^{N}(\mathrm{rank}(y_i) - \bar{r}_Y)^2}}, \tag{3.6}$$

where

$$\bar{r}_X = \frac{1}{N}\sum_{i=1}^{N}\mathrm{rank}(x_i).$$

When computing ρ_{S} any ties in the data are assigned the average rank of the tied scores.

One of the important properties of the Spearman correlation is that if there exists a strictly increasing function $g(X)$ that relates X to Y as $Y = g(X)$, then $\rho_{\mathrm{S}}(X,Y) = 1$. Furthermore, a strictly monotonic transformation of X or Y will not affect the Spearman correlation

As with the Pearson correlation, we can compute a Spearman correlation matrix for a collection of random variables $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$. We will call this matrix $\mathbf{R}_{\mathrm{S}}$ and it is given by

$$R_{\mathrm{S},ij} = \rho_{\mathrm{S}}(X_i, X_j).$$

3.1.3 Kendall's Tau

The final measure of correlation that we will use is Kendall's tau or the Kendall rank correlation coefficient. Similar to the Spearman correlation it tries to measure the relation between two variables in terms of the ranks. It is best for looking at a sample population of random variables because it requires looking at pairs of samples of random variables. To define Kendall's tau be consider N samples of random

variables X and Y. We examine all the pairs of samples (x_i, y_i) for $i \neq j$. There are $\frac{1}{2}n(n-1)$ such pairs. We look at each pair and say that a pair ij is concordant if $x_i > x_j$ and $y_i > y_j$ or if $x_i < x_j$ and $y_i < y_j$. A pair is discordant if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. If either $x_i = x_j$ or $y_i = y_j$ then the pair is a tie.

Using this comparison of pairs we define Kendall's tau as

$$\tau = \frac{(\#\text{ of concordant pairs}) - (\#\text{ of discordant pairs})}{\frac{1}{2}N(N-1)}. \tag{3.7}$$

The range of τ is $[-1,1]$. Kendall's tau has the property that it is not affected by performing a nonlinear, increasing transformation on either random variable: this is the same property Spearman correlation has. We can relate τ to the Pearson correlation coefficient if the variables X and Y are jointly normally distributed through the equation

$$\tau(X,Y) = \frac{2}{\pi}\arcsin\rho(X,Y).$$

We will use Kendall's tau when we want to relate two random variables through copulas.

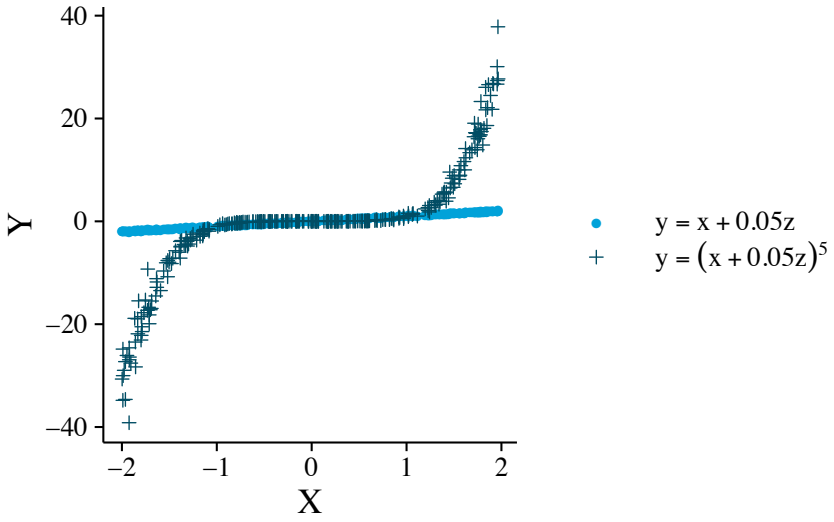

Fig. 3.2 The comparison of Pearson, Spearman, and Kendall's tau correlation measures on 300 samples of two pairs of random variables: $(X, X+0.05Z)$ and $(X, (X+0.05Z)^5)$, where Z is a standard normal random variable. The three measures give a correlation of $\rho = 0.999$, $\rho_S = 0.999$, and $\tau = 0.973$ for the correlation of $(X, X+0.05Z)$. The correlation of $(X, (X+0.05Z)^5)$, the Spearman correlation and Kendall's tau values do not change, but $\rho = 0.843$ for this data.

As comparison of the correlation measures, Figure 3.2 shows how a strictly increasing transformation of a variable changes the Pearson correlation, but not the

Spearman correlation or Kendall's tau. In the figure the correlation between random variables $(X, X+0.05Z)$ and the correlation between $(X, (X+0.05Z)^5)$, where Z is a standard normal random variable is computed. The Spearman and Kendall measures do not change, whereas the Pearson correlation drops by 15%.

3.1.4 Tail Dependence

Another important characterization of how two variables vary together is tail dependence. This is a measure of the correlation between variables as their lower and upper bounds are approached. The lower tail dependence, λ_l is

$$\lambda_l(X,Y) = \lim_{q\to 0} \mathrm{P}(Y \le F_Y^{-1}(q) \mid X \le F_X^{-1}(q)). \tag{3.8}$$

This is the probability that Y goes to its lower bound as X goes to its lower bound. The upper tail dependence is

$$\lambda_u(X,Y) = \lim_{q\to 1} \mathrm{P}(Y > F_Y^{-1}(q) \mid X > F_X^{-1}(q)), \tag{3.9}$$

and measures the probability the X and Y go to their upper bound together.

Tail dependence is different than typical correlation measures in that it is only interested in extreme values. For example two variables could have a Pearson correlation of 0.5, but a tail dependence much larger, say 0.9. This has been observed, for example, in the returns of stocks. Many stocks that had low correlation in typical times had very high lower tail dependence during the financial crises (they all went down a lot).

The lower tail dependence can be written in terms of the joint CDF for two variables. Using the definition of the CDF and law of total probability we get that

$$\mathrm{P}(Y \le F_Y^{-1}(q) \mid X \le F_X^{-1}(q)) = \frac{F_{XY}(F_X^{-1}(q), F_Y^{-1}(q)}{F_X(F_X^{-1}(q)))} = \frac{F_{XY}(F_X^{-1}(q), F_Y^{-1}(q))}{q}, \tag{3.10}$$

where F_{XY} is the joint CDF for X and Y. Similarly, the upper tail dependence can be written in terms of the joint CDF as

$$\begin{aligned}
\mathrm{P}(Y > F_Y^{-1}(q) \mid X > F_X^{-1}(q)) &= \frac{\mathrm{P}(Y > F_Y^{-1}(q), X > F_X^{-1}(q))}{P(Y > F_Y^{-1}(q)} \\
&= \frac{1 - P(X \le F_X^{-1}(q)) - P(Y \le F_Y^{-1}(q)) + F_{XY}(F_X^{-1}(q), F_Y^{-1}(q))}{1 - F_X(F_X^{-1}(q))} \\
&= \frac{1 - 2q + F_{XY}(F_X^{-1}(q), F_Y^{-1}(q))}{1-q}.
\end{aligned} \tag{3.11}$$

These equations give us formulas for the tail dependences in terms of the joint and marginal CDFs for each of these variables.

3.2 Copulas

A common occurrence when evaluating collections of random variables, it is often much easier to determine the marginal CDF or PDF of each variable, rather than determine the joint distribution functions. Moreover, given a sample of data it is possible to compute correlations between the random variables (in either of the three flavors we mentioned in the previous section). The question is, given the scenario where one has

- An estimate of the marginal CDF of each random variable, and
- An estimate of the correlation between the random variables,

can one generate a joint distribution between the variables and generate samples from the joint distribution? Clearly, there is not a unique way of creating this joint distribution because many functions could replicate the marginal distributions and have a defined correlation.

To answer this question we turn to copulas (or copulæ if one is a fan of Latinisms). We will begin with discussing bivariate copulas before generalizing the idea to general collections of random variables. The word copula comes from a Latin for linking together; in our context it will link marginal distributions to a joint distribution.

A copula, $C(u,v)$, joins random variables X and Y if the joint CDF can be written as

$$F_{XY}(x,y) = C(F_X(x), F_Y(y)). \tag{3.12}$$

This definition takes the marginal CDF for each variable and creates a joint CDF. A result known as Sklar's theorem tells us that such a copula will exist for any joint CDF and it is unique if the marginal CDFs are continuous. A copula has the domain $u,v \in [0,1]$, and a range of $[0,1]$. For a given copula we can define the joint PDF as

$$f(x,y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y), \tag{3.13}$$

where the copula density, $c(u,v)$, is given by

$$c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v). \tag{3.14}$$

This definition is a special case of Eq. (2.25). Additionally, the conditional CDF $C(v|u)$ is

$$C(v|u) = \frac{\partial}{\partial u} C(u,v). \tag{3.15}$$

The tail dependence for a copula can be obtained by plugging Eq. (3.12) into the definition for tail dependence to get

$$\lambda_l = \lim_{q \to 0} \frac{C(q,q)}{q}, \tag{3.16}$$

and

$$\lambda_u = \lim_{q \to 1} \frac{1 - 2q + C(q,q)}{1-q}, \tag{3.17}$$

The simplest copula is the independent copula:

$$C_{\mathrm{I}}(u,v) = uv.$$

Copulas are widely used in the finance and insurance industries to model the joint distributions of risks. Because the fact that mapping marginal distributions to joint distributions is not unique, the way we use a copulas requires choices by the user. The considerations of ease of use, matching observed correlation, and tail dependence have to be weighed when choosing a copula.

3.2.1 Normal Copula

A simple, but useful copula, is the normal (or Gaussian) copula

$$C_N(u,v) = \Phi_{\mathbf{R}}(\Phi^{-1}(u), \Phi^{-1}(v)), \tag{3.18}$$

where $\mathbf{R}$ is a correlation matrix for the intended joint distribution. The normal copula is simple to sample. Given two random variables X and Y with marginal CDFs $F_X(x)$ and $F_Y(y)$, we can generate a sample from $C_N(F_X(x), F_Y(y))$ using the following procedure

1. Sample from the collection of two random variables $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ using the Cholesky factorization approach in the previous chapter.
2. Compute $u = \Phi(z_1)$ and $v = \Phi(z_2)$.
3. The samples are $x = F_X(u)$ and $y = F_Y(v)$.

Therefore, via the normal copula we can create a joint distribution that has a prescribed Pearson correlation where the underlying marginal distributions do not have to be normal. This is different than saying that the two variables are a multivariate normal with a known correlation. Note that the matrix $\mathbf{R}$ has only one degree of freedom because the diagonal is 1 and it is symmetric, we can call this degree of freedom ρ. It can be shown that for a normal copula, the value of Kendall's tau is

$$\tau(X,Y) = \frac{2}{\pi} \arcsin \rho, \qquad F(X,Y) = C_N(F_X(x), F_Y(y)). \tag{3.19}$$

Therefore, given a desired value of Kendall's tau for the joint distribution, one can produce it using the normal copula.

The normal copula has zero tail dependence: as one variable approaches $\pm\infty$ the probability that the other variable does the same goes to zero. Therefore, if we are modeling a system where tail dependence could matter greatly, e.g., analyzing how the system behaves under input variables near their extremes, the normal copula may not be appropriate. In fact the normal has been blamed for the financial crisis of 2008 [Jones(2009)] because it does not account for the fact that mortgage defaults, while not being correlated under normal circumstances, have strong lower tail dependence because if everyone in a neighborhood is foreclosed, the housing prices fall, and more mortgages then default: a fact that risk assessors never understood. This needs to be carefully analyzed when quantifying uncertainty in a physical system. In many cases tail dependence could be present, and we need to understand how this may affect our predictions.

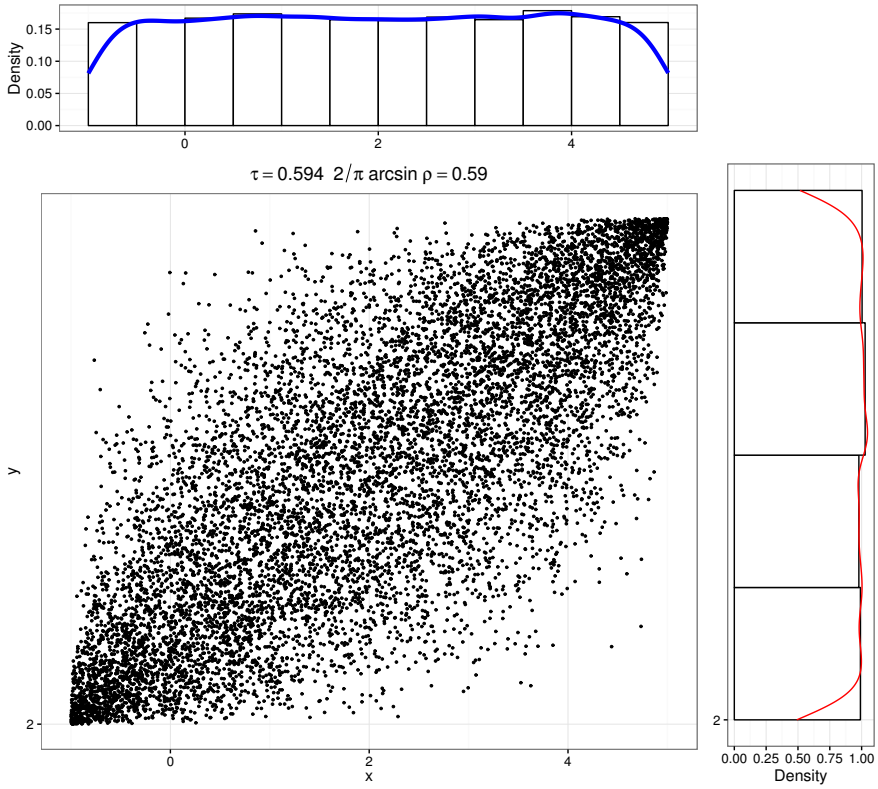

Fig. 3.3 Samples from uniform random variables $X \sim \mathscr{U}(-1,5)$ and $Y \sim \mathscr{U}(2,3)$ joined by a normal copula with $\rho = 0.8$. From these 10^4 samples the empirical value of τ and the predicted value from Eq. (3.19) are shown also.

In Figure 3.3 two uniform distributions are joined by a normal copula with $\rho = 0.8$ are shown. Notice how there is a clear correlation between the two random variables and, as a result, a clustering in the corners of the distributions. An important property of these samples is that they are not normal, we have just used a normal copula to join them.

3.2.2 t-Copula

A distribution similar to the normal is the t-distribution: it is unimodal but has more kurtosis than a normal random variable. This distribution can be used to define a t-copula with a scale parameter $\nu > 0$, and a positive definite, symmetric scale matrix $\mathbf{S}$ with a diagonal of ones as

$$C_t(u,v) = F_{\mathbf{t}}(F_t^{-1}(u), F_t^{-1}(v)), \tag{3.20}$$

where $F_{\mathbf{t}}$ is the joint CDF for a t-distribution with parameters $\boldsymbol{\mu} = \mathbf{0}$, $\mathbf{S}$ and ν. The CDF $F_t(x)$ is the CDF of the t-distribution with parameter ν. The degree of freedom in the $\mathbf{S}$ matrix will be written as r.

To sample from random variables joined by the t-copula we use a similar procedure to that for the normal copula:

1. Sample from the collection of two random variables $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$ using the Cholesky factorization approach in the previous chapter.
2. Compute $\hat{Z} = \sqrt{w}Z$, where w is a sample from the inverse gamma distribution, $W \sim \text{IG}(\nu/2, \nu/2)$.
3. Compute $u = F_t(z_1)$ and $v = F_t(\hat{z}_2)$.
4. The samples are $x = F_X(u)$ and $y = F_Y(v)$.

The t-copula has the same form for Kendall's tau as the normal copula. In particular if we replace $r \to \rho$ in Eq. (3.19) we can relate Kendall's tau to the matrix $\mathbf{S}$.

In Figure 3.4 two uniform distributions are joined by a normal copula with $r = 0.8$ are shown. Notice how there is a clear correlation between the two random variables and, as a result, a clustering in the corners of the distributions. Notice there are more samples farther off the diagonal than in the normal case. This is due to the fact that the t-distribution with a small value of ν has more kurtosis than a normal distribution. Therefore, it is more likely to get anti-correlated values as samples. The fact that the t-copula has tail dependence can also be observed in this figure in the concentration of points near the lower-left and upper-right corners.

The tail dependence can be seen even more clearly if we use a t-copula to couple two normal random variables. In Figures 3.5 the t-copula and normal copulas are compared. Here, we see that the tail dependence appears as the area that the samples occupy narrowing as the upper right and lower left corners are approached in the t-copula, but this not present in the normal copula. This discrepancy in the tails exists even though both distributions have the same value for τ and the same marginal distributions for X and Y.

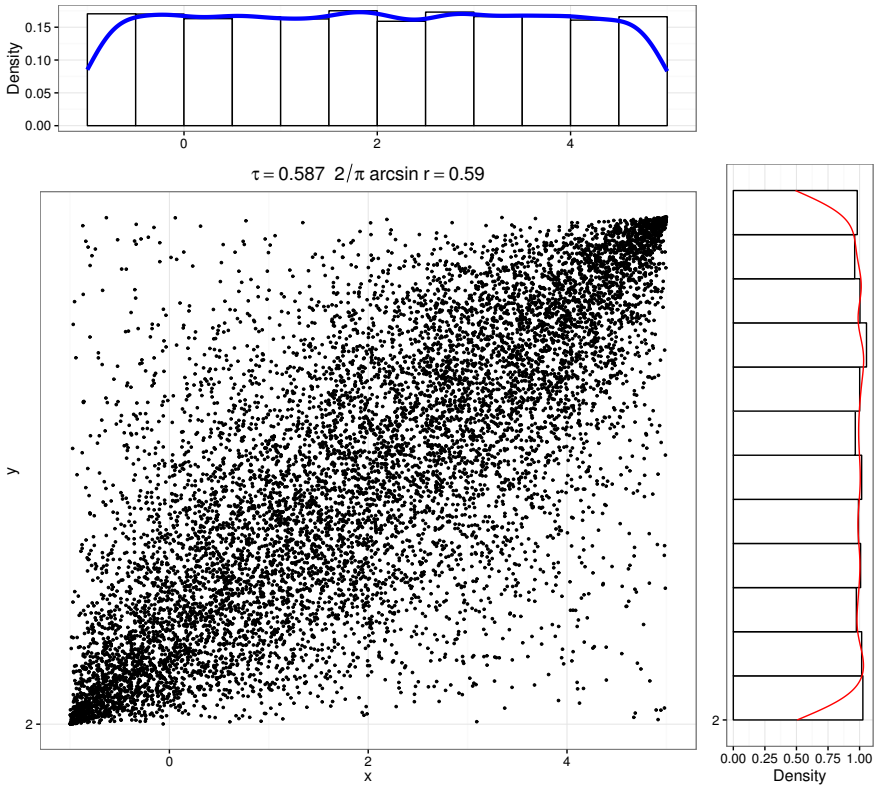

Fig. 3.4 Samples from uniform random variables $X \sim \mathscr{U}(-1,5)$ and $Y \sim \mathscr{U}(2,3)$ joined by a t-copula with $r = 0.8$ and $\nu = 4$. From these 10^4 samples the empirical value of τ and the predicted value from Eq. (3.19) are shown also.

3.2.3 Fréchet Copulas

The Fréchet copula C_L and C_U are simple copulas that join random variables with Spearman correlation ± 1. Furthermore, any other copula is bounded by the relation $C_L \leq C \leq C_U$. The Fréchet copulas are

$$C_L(u,v) = \max(u+v-1,0), \qquad C_U(u,v) = \min(u,v). \tag{3.21}$$

C_L will give perfect negative dependence between variables and C_U will give perfect positive correlation between variables. We can then combine Fréchet copulas to describe something with a Spearman correlation between $[-1,1]$:

$$C_A(u,v) = (1-A)C_L(u,v) + AC_U(u,v), \qquad A \in [0,1]. \tag{3.22}$$

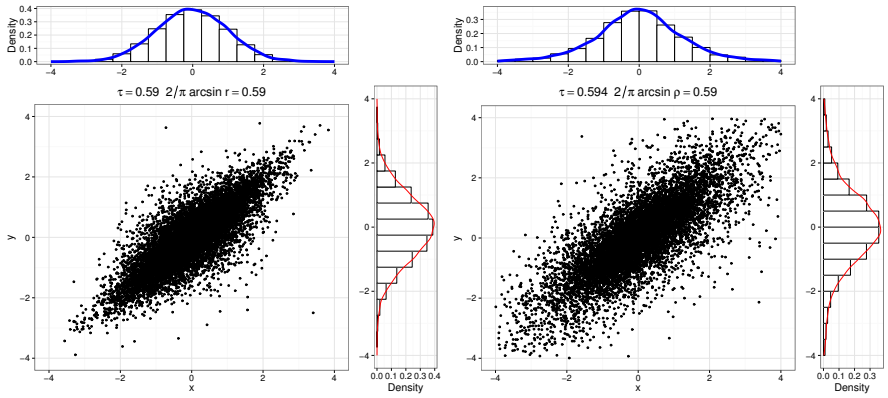

Fig. 3.5 Samples from standard normal random variables $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ joined by a t-copula with $r = 0.8$ and $\nu = 4$ (right) and the normal copula with $\rho = 0.8$ (left) . From these 10^4 samples the empirical value of τ and the predicted value from Eq. (3.19) are shown also. Note the tail dependence in the t-copula that is lacking in the normal copula: when one variable is close to ± 4 the other variable is also likely to be close to ± 4.

These are a simple combination and can give a Spearman correlation given by $2A - 1$.

3.2.4 Archimedean Copulas

There is another class of copulas that easily generalize to an arbitrary number of dimensions and have an explicit formula. These copulas, called Archimedean copulas, and they are defined by a generator function, $\varphi(t)$ for $t \in [0,\infty)$. Given a generator, we define the quasi-inverse

$$\hat{\varphi}^{-1}(t) \equiv \begin{cases} \varphi^{-1}(t) & 0 \le t \le \varphi(0) \\ 0 & \varphi(0) < t < \infty \end{cases}. \tag{3.23}$$

With the generator and quasi-inverse the Archimedean copula for $\varphi(t)$ is

$$C_\varphi(u,v) = \hat{\varphi}^{-1}\left(\varphi(u) + \varphi(v)\right). \tag{3.24}$$

The term Archimedean arises from the development of the triangle inequality for probability spaces, in that context Archimedes of Syracuse's name is attached a particular norm that has the form of Eq. (3.24).

Archimedean copulas are commutative:

$$C_\varphi(u,v) = C_\varphi(v,u),$$

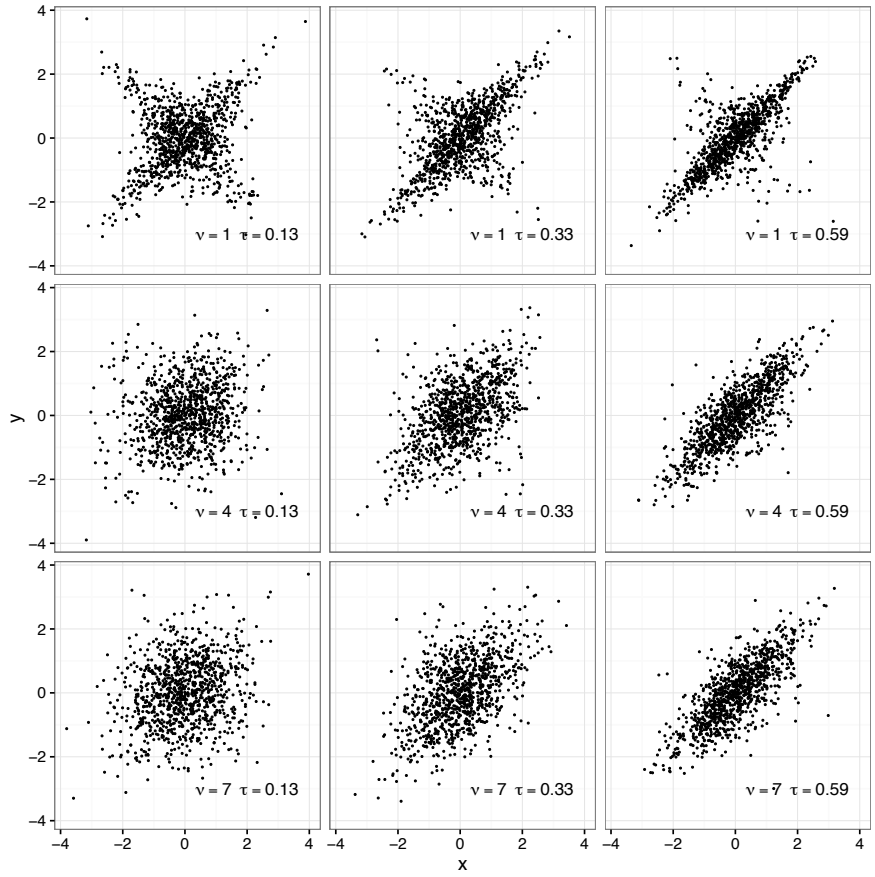

Fig. 3.6 Samples from standard normal random variables $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ joined by a t copula with several values of r and ν. The value of ν is constant in a column and the value of r (and the corresponding τ) are constant in each row.

associative

$$C_\varphi(C_\varphi(u,v),w) = C_\varphi(u,C_\varphi(v,w)),$$

and are order preserving

$$C(u_1,v_1) > C(u_2,v_2), \qquad u_1 > u_2, \quad v_1 > v_2.$$

The associative property will be used later to easily create Archimedean copulas for arbitrary numbers of variables.

Furthermore, an Archimedean copula can be related to Kendall's tau via the formula

$$\tau(U,V) = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}\,dt. \tag{3.25}$$

There are many Archimedean copulas one could define, we will discuss two below that are commonly used.

3.2.4.1 The Frank Copula

One common Archimedean copula is the Frank copula. This copula has a single parameter, $\theta \neq 0$, and a generator function given by

$$\varphi_{\mathrm{F}}(t) = -\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right). \tag{3.26}$$

The inverse is

$$\hat{\varphi}^{-1}(t) = -\frac{1}{\theta}\log\left(1 + e^{-t}(e^{-\theta}-1)\right). \tag{3.27}$$

This makes the copula

$$C_{\mathrm{F}}(u,v) = -\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right). \tag{3.28}$$

One property of the Frank copula is that as $\theta \to \infty$, the copula becomes the upper Fréchet copula: $C_{\mathrm{F}} \to C_U$. As $\theta \to -\infty$, then the Frank copula approaches the lower Fréchet copula: $C_{\mathrm{F}} \to C_L$.

The value of Kendall's tau for a Frank copula is can be calculated from the Eq. (3.25) as

$$\tau_F(U,V) = 1 - \frac{2\left(3\theta^2 - 6i\pi\theta + 6\theta - 6\theta\log\left(e^{\theta}-1\right) - 6\mathrm{Li}_2\left(e^{\theta}\right) + \pi^2\right)}{3\theta^2}, \tag{3.29}$$

where $\mathrm{Li}_s(z)$ is the polylogarithm function. A table for matching a desired value of τ_{F} to θ is given in Table 3.1. Additionally, the value of τ_{F} as a function of θ is shown in Figure 3.7. The Frank copula has a tail dependence of zero.

In Figure 3.9 samples from Frank copula are shown with the values of θ given in Table 3.1. In this figure we can see that as θ gets larger, the distribution is pinched in the middle, but the tails of the distribution remain spread out.

3.2.4.2 The Clayton Copula

The Clayton copula generator function has a single parameter, $\theta > 0$, with generator function

$$\varphi_{\mathrm{C}}(t) = t^{-\theta} - 1 \tag{3.30}$$

τ_F	θ
0.1	0.907368
0.2	1.860880
0.3	2.917430
0.4	4.161060
0.5	5.736280
0.6	7.929640
0.7	11.411500
0.8	18.191500
0.9	26.508600

Table 3.1 The corresponding value of θ for different values of Kendall's tau using the Frank copula. Note that negative values of τ_F will have a corresponding negative value of θ.

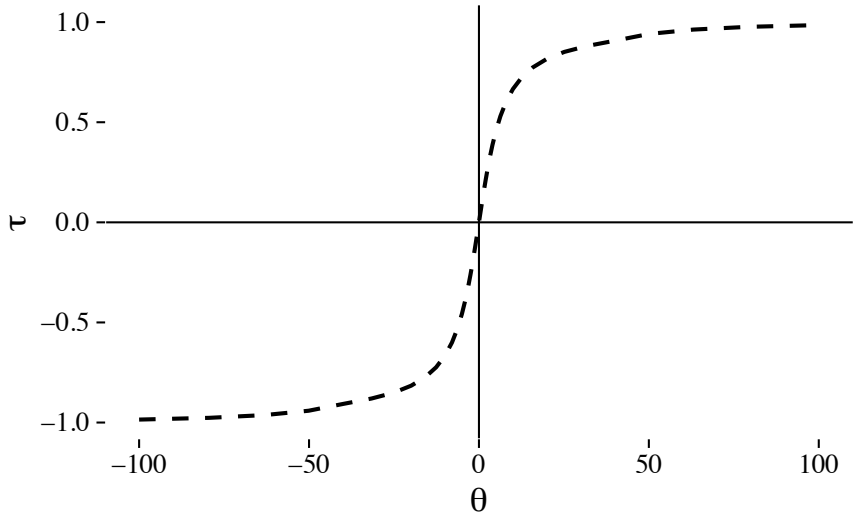

Fig. 3.7 Kendall's tau as a function of θ for the Frank copula.

and inverse

$$\hat{\varphi}_C^{-1}(t) = (1+t)^{-1/\theta}. \tag{3.31}$$

The resulting copula is

$$C_C(u,v) = \max\left(0, u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}. \tag{3.32}$$

The Clayton copula has Kendall's tau for the resulting joint distribution

$$\tau_C(U,V) = \frac{\theta}{\theta+2}. \tag{3.33}$$

Additionally, the Clayton copula has zero upper tail dependence and non-zero lower tail dependence:

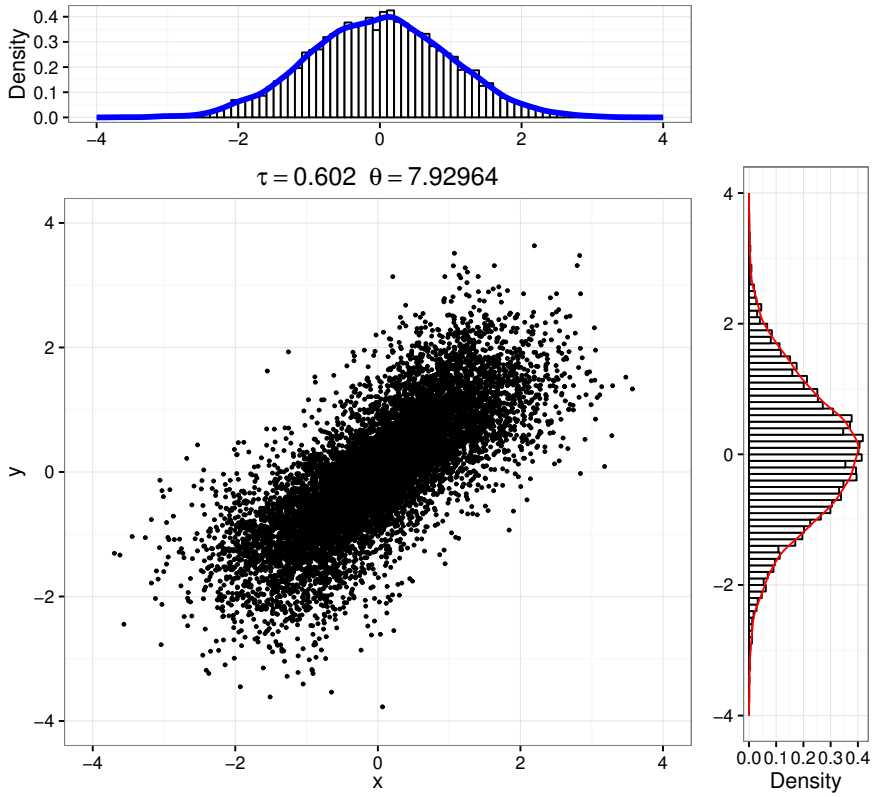

Fig. 3.8 Samples from standard normal random variables $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ joined by a Frank copula with θ chosen to get $\tau = 0.6$. Note the lack of tail dependence in the lack of concentration near the upper right and lower left corners. Note that relative to the normal copula and the t-copula, these points form a rectangular-shaped band. The lack of tail dependence is also apparent in the lack of points along the diagonal.

$$\lambda_l = 2^{-1/\theta}. \tag{3.34}$$

We can use the Clayton copula to produce joint distributions with upper tail dependence and no lower tail dependence by using the copula $C_C(1-u, 1-v)$.

The Clayton copula with different values of θ that correspond with values of Kendall's tau from 0.1 to 0.9 is shown in Figure 3.11. As θ increases, the shape of the distribution becomes more tapered in the middle and makes the lower tail dependence more prominent, as predicted, making the samples form something akin to the celebrate emoji 🎉.

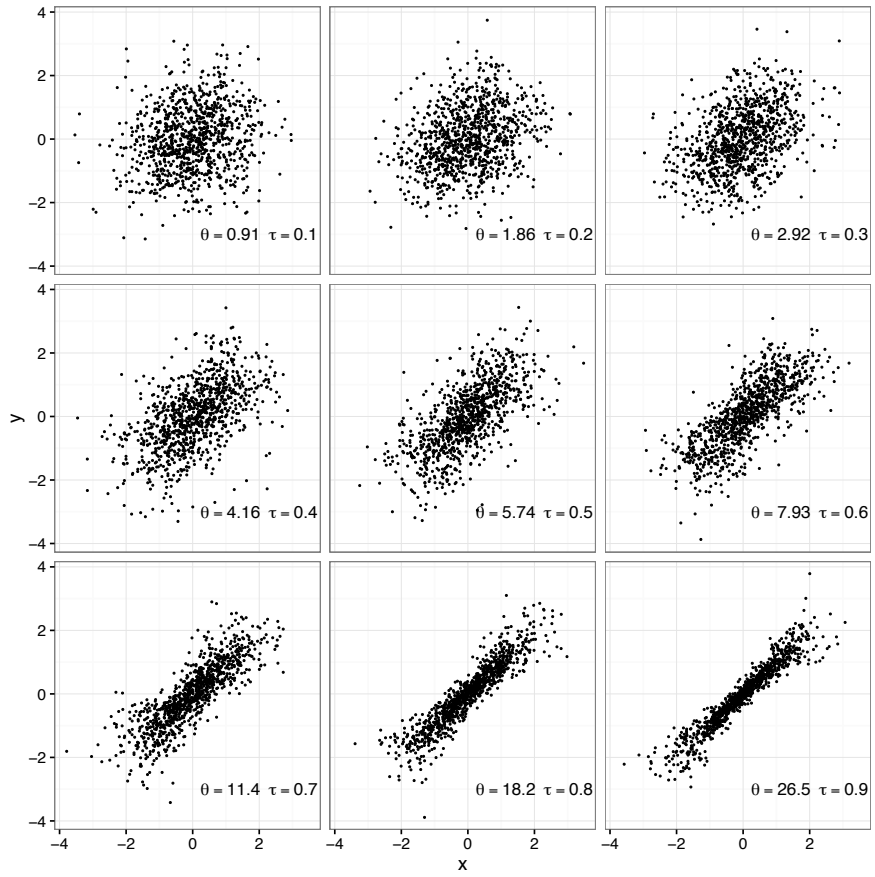

Fig. 3.9 Samples from standard normal random variables $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ joined by a Frank copula with several values of θ taken from Table 3.1 .

3.2.5 Sampling from Bivariate Copulas

We have discussed how to sample from the t- and normal copulas, but these procedures do not extend to general copulas. There is a straightforward way to produce samples from a joint distribution produced by copulas. Consider the marginal CDFs for random variables X and Y, $F_X(x)$ and $F_Y(y)$, and a copula $C(u,v)$. The procedure to produce samples from the joint distribution given by $C(F_X(x), F_Y(y))$ is

1. Produce two uniform random variables ξ_1 and ξ_2 where $\xi_i \sim \mathscr{U}(0,1)$.
2. Set
$$w \equiv C^{-1}(\xi_2|\xi_2).$$
3. Then the samples x and y are $x = F_X^{-1}(\xi_1)$ and $y = F_Y^{-1}(w)$.

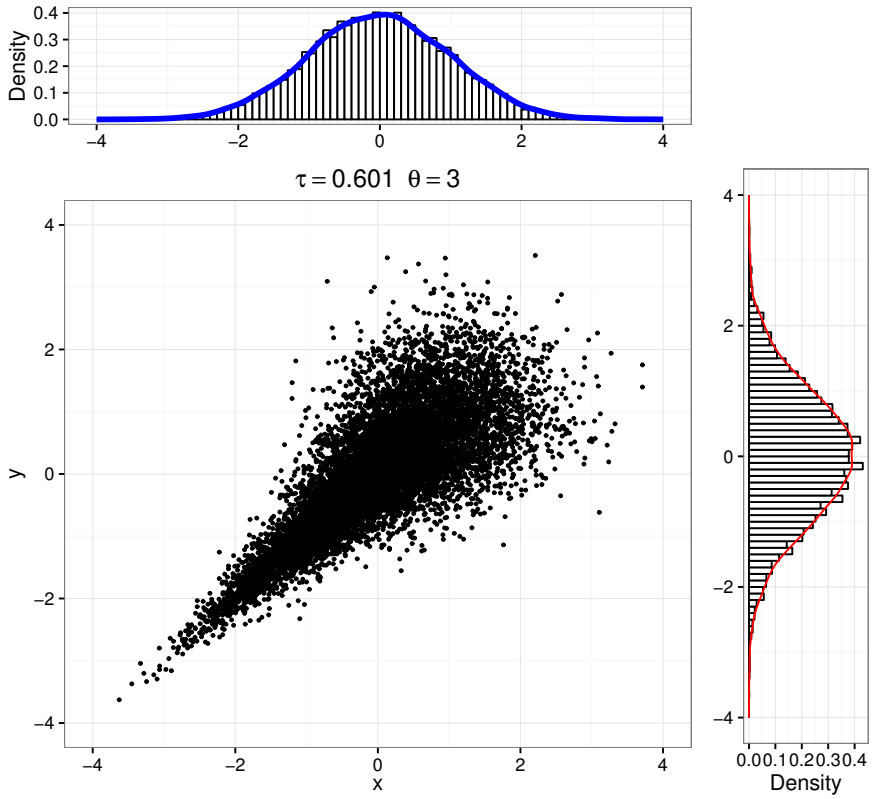

Fig. 3.10 Samples from standard normal random variables $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ joined by a Clayton copula with θ chosen to get $\tau = 0.6$. There is strong lower tail dependence in the samples and the zero upper tail dependence.

This sampling procedure is simple to perform with the possible exception of not knowing $C^{-1}(v|u)$. In this case we can use a nonlinear solver to perform the inversion.

As a demonstration we will show how this works for the Frank copula. This is a case where the inverse of the conditional CDF, $C^{-1}(v|u)$ can be explicitly calculated. For the Frank copula, we have

$$C_{\mathrm{F}}(v|u) = \frac{e^{\theta}\left(e^{\theta v}-1\right)}{-e^{\theta}+e^{\theta+\theta u}-e^{\theta(u+v)}+e^{\theta+\theta v}}. \tag{3.35}$$

The inverse of this function is

$$C_{\mathrm{F}}^{-1}(\xi|u) = \frac{1}{\theta}\log\left(\frac{e^{\theta}\left(\xi\left(e^{\theta u}-1\right)+1\right)}{\xi e^{\theta u}-e^{\theta}(\xi-1)}\right). \tag{3.36}$$

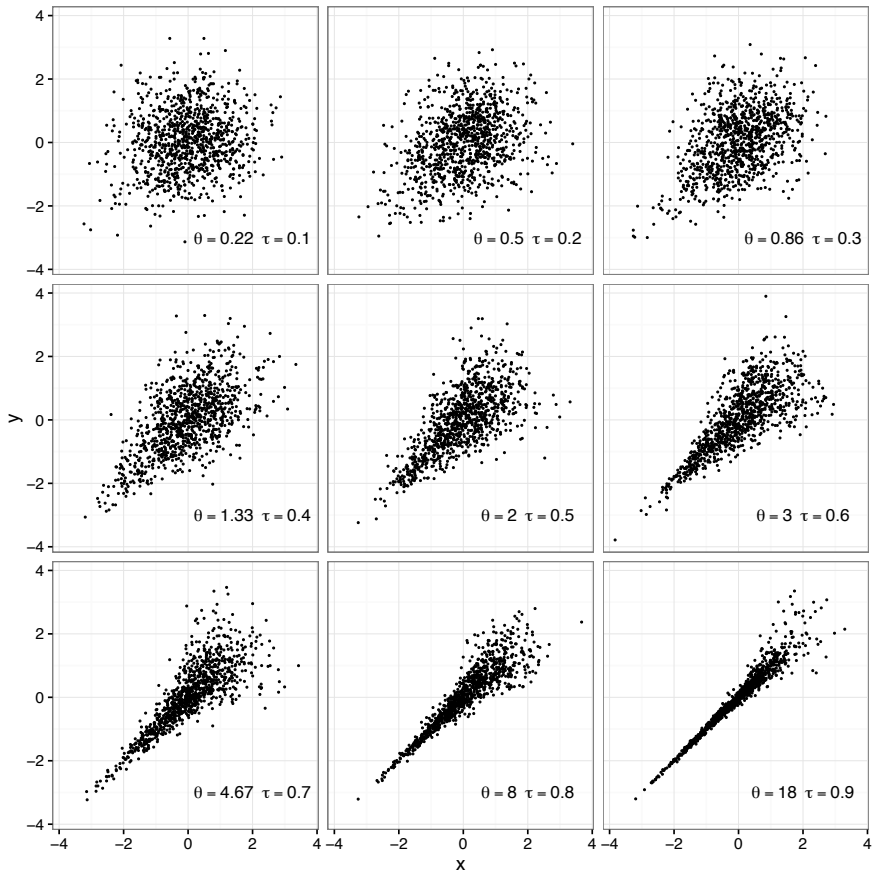

Fig. 3.11 Samples from standard normal random variables $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ joined by a Clayton copula with several values of θ.

This algorithm was used to generate the samples in Figure 3.8.

3.3 Multivariate Copulas

The idea of a copula can be extended to more than two random variables. For this discussion we will have a collection of p random variables $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$. Each of these random variables has a known marginal CDF $F_{X_i}(x_i)$. A copula, C, on this collection of random variables is function that maps a p-dimensional vector $\mathbf{u}$ with each component in $[0,1]$ to a non-negative real number. With this copula we then define a joint CDF for $\mathbf{X}$ as

$$F(\mathbf{x}) = C(F_{X_1}(x_1), \ldots, F_{X_p}(x_p)). \tag{3.37}$$

In the multivariate setting, the independence copula, C_I, is a product of each input:

$$C_I(\mathbf{u}) = \prod_{i=1}^{p} u_i. \tag{3.38}$$

The normal copula is extended in a simple manner,

$$C_{\mathrm{N}}(\mathbf{u}) = \Phi_{\mathbf{R}}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p)). \tag{3.39}$$

In this case the correlation matrix is of size $p \times p$. In an analogous fashion, the t-copula can be extended as well,

$$C_{\mathrm{t}}(\mathbf{u}) = F_{\mathbf{t}}(F_{\mathrm{t}}^{-1}(u_1), \ldots, F_{\mathrm{t}}^{-1}(u_p)). \tag{3.40}$$

For both of these copulas we have already given a procedure for sampling from the joint distributions. The algorithms that we discussed earlier need to take p samples from a multivariate normal instead of two, and the rest of the algorithm proceeds naturally. The multivariate extensions of these copulas will have the correlation between variables specified by the **R** and **S** matrices.

Archimedean copulas in higher-dimensions also have a natural extension. These copulas can be written as

$$C_{\varphi}(\mathbf{u}) = \hat{\varphi}^{-1}(\varphi(u_1) + \cdots + \varphi(u_p)). \tag{3.41}$$

Note that each generator needs to have the same value of θ in this definition. This means that Kendall's tau, and perhaps the tail dependence, will be the same for all the variables.

3.3.1 Sampling Multivariate Archimedean Copulas

To sample from an Archimedean copula we will use the Marshall-Olkin algorithm. In this algorithm we need to take the inverse Laplace transform of the quasi-inverse of the generator function, $\varphi(t)$. It turns out that the inverse Laplace transform of a generator function is a cumulative distribution function. We denote the Laplace transform of the quasi-inverse of generator as $F(s) \equiv \mathscr{L}^{-1}[\varphi^{-1}(t)]$. Using this cumulative distribution function, this algorithm is given as

1. Sample $s = F^{-1}(\xi)$, where $\xi \sim \mathscr{U}(0,1)$.
2. Create p samples $\mathbf{u}$ where $u_i \sim \mathscr{U}(0,1)$.
3. Create p values$\mathbf{v}$ where $v_i = \varphi^{-1}(-\log(u_i)/s)$.
4. The samples from $F(\mathbf{x})$ are $x_i = F_{X_i}^{-1}(v_i)$.

For the Clayton copula with positive θ, the inverse Laplace transform of the generator yields the CDF for the Gamma distribution with parameter $\alpha = \theta^{-1}$ and

$\beta = 1$. For the Frank copula, the function F, for positive θ, is a discrete random variable with CDF

$$F(k) = \frac{(1-e^{-\theta})^k}{k\theta}, \qquad k = 1, 2, \ldots.$$

With multivariate copulas, Archimedean copulas, as we have defined them, have the same dependence between all variables, as seen in the example in Figure 3.12. Is this figure we show the 2-D projections of a 5-D joint distribution joined by a Clayton copula with $\theta = 3$. The marginal distributions are all standard normal. The

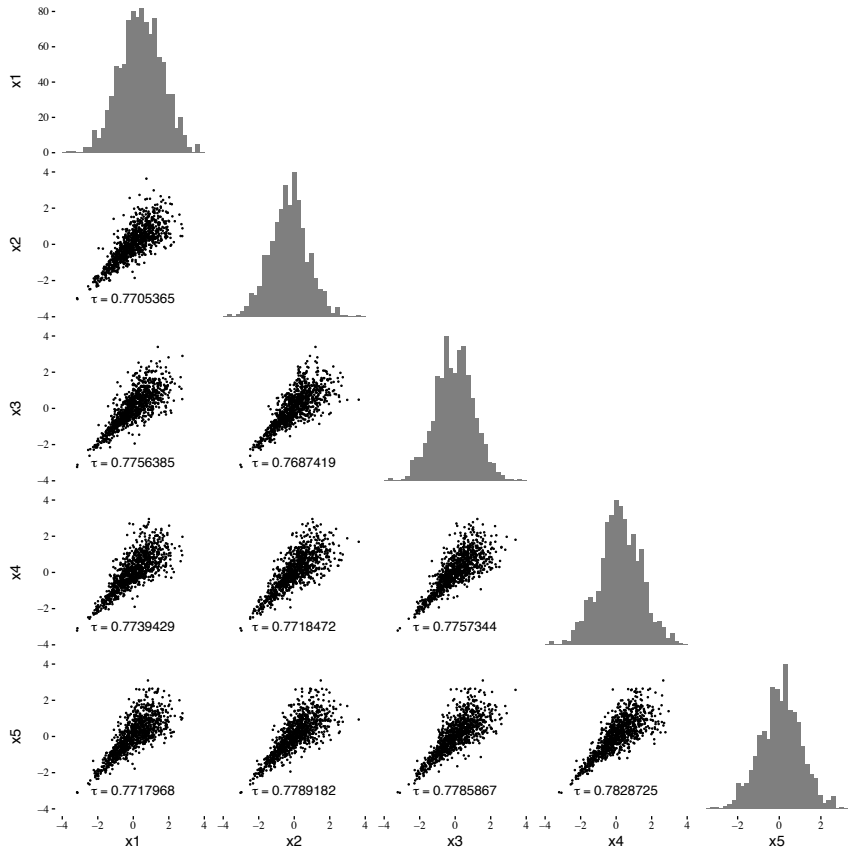

Fig. 3.12 Samples from 5 standard normal random variables joined by a Clayton copula with $\theta = 3$. Note how the Kendall's tau value between each pair of variables is constant.

case is different with a normal copula (or a t-copula) in that the correlation between variables can be different depending on the pair of variables. In Figure 3.13 samples from a 5-D normal copula with correlation matrix given by

$$\mathbf{R} = \begin{pmatrix} 1.00 & 0.75 & 0.50 & 0.25 & 0.12 \\ 0.75 & 1.00 & 0.50 & 0.25 & 0.12 \\ 0.50 & 0.50 & 1.00 & 0.12 & 0.50 \\ 0.25 & 0.25 & 0.12 & 1.00 & 0.12 \\ 0.12 & 0.12 & 0.50 & 0.12 & 1.00 \end{pmatrix}.$$

In this case the dependence between pairs of variables does change.

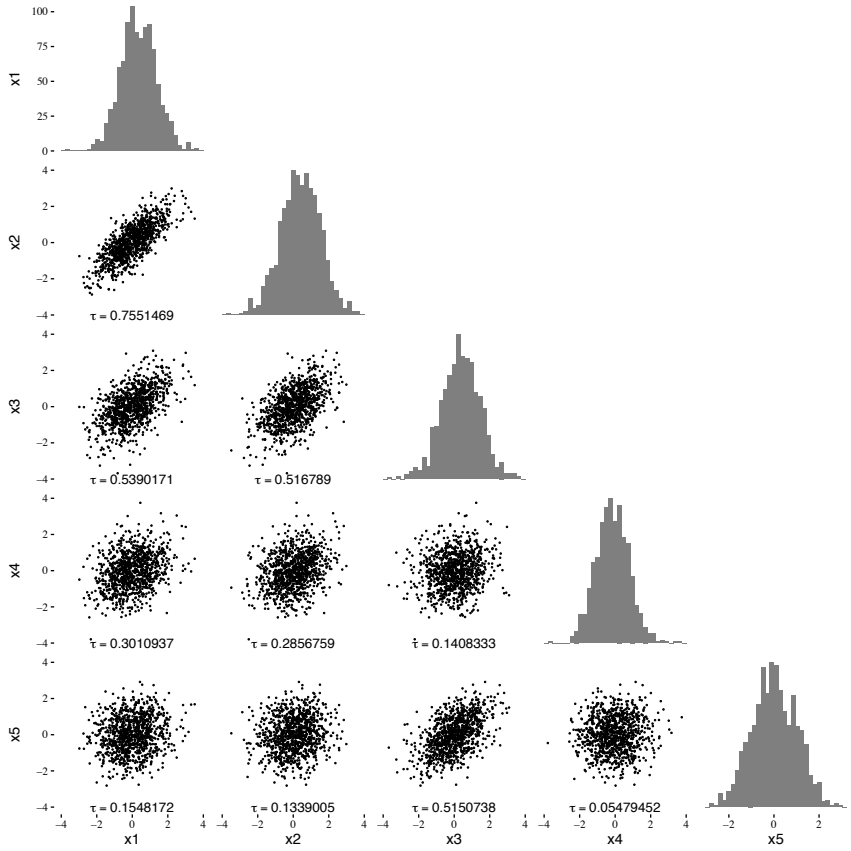

Fig. 3.13 Samples from 5 standard normal random variables joined by a normal copula with a correlation matrix that is not uniform. Note how the Kendall's tau value between each pair of variables is different.

3.4 Random Variable Reduction: The Singular Value Decomposition

In this section we will discuss a way to create uncorrelated random variables from a set of correlated random variables. We will do this use the singular value decomposition of the data. This procedure is known by several names, including principle components analysis, the Hotelling transform, or proper orthogonal decomposition. This procedure can be used to reduce the dimension of the data set by revealing a set of uncorrelated random variables that produce the observed correlated random variables.

Consider that we have a collection of p random variables $\mathbf{X}$ and n samples of these random variables, and $n > p$. We can assemble these samples into a n by p matrix (n rows and p columns) of the form

$$\mathbf{A} = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{p-1}^{(1)} & x_p^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{p-1}^{(2)} & x_p^{(2)} \\ \vdots & & & & \\ x_1^{(n)} & x_2^{(n)} & \dots & x_{p-1}^{(n)} & x_p^{(n)} \end{pmatrix}. \tag{3.42}$$

The matrix $\mathbf{A}$ will be rectangular in general, $n \neq p$. For such a matrix, we can factor it into what is known as the singular value decomposition (SVD):

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}. \tag{3.43}$$

In this decomposition

- $\mathbf{U}$ is a $n \times p$ orthogonal matrix, i.e., $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$ and $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$ where $\mathbf{I}$ is an identity matrix.
- $\mathbf{S}$ is a $p \times p$ diagonal matrix with non-negative entries.
- $\mathbf{V}$ is a $p \times p$ orthogonal matrix.

The singular value decomposition is related to the eigenvalue decomposition of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ or $\mathbf{A}^{\mathrm{T}}\mathbf{A}$. To see this we left multiply Eq. (3.42) by $\mathbf{A}^{\mathrm{T}}$ to get

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \left(\mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}\right)^{\mathrm{T}}\mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} = \mathbf{V}\mathbf{S}\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} = \mathbf{V}\mathbf{S}^2\mathbf{V}^{\mathrm{T}}$$

Similarly, right multiplying by $\mathbf{A}^{\mathrm{T}}$ gives

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{U}\mathbf{S}^2\mathbf{U}^{\mathrm{T}}.$$

Therefore, we can interpret the entries in $\mathbf{S}$ as the square-root of the eigenvalues of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$. Therefore, we will call these $\sqrt{\lambda_i}$ and order them in decreasing magnitude

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r,$$

where r is the number of non-zero eigenvalues of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$.

Note that the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is an approximation of the covariance matrix for the p random variables because the dot product between rows and columns is an approximation to the integral in the definition in covariance.

Given that we know how to interpret the values in the matrix $\mathbf{S}$, we can develop interpretations for the meaning of the $\mathbf{U}$ and $\mathbf{V}$ matrices. We can transform the matrix $\mathbf{A}$ into an orthogonal matrix by multiplying by $\mathbf{V}$ to get the $n \times p$ matrix:

$$\mathbf{T} \equiv \mathbf{AV}.$$

The resulting matrix has columns that are linear combinations of the original columns. The columns in $\mathbf{T}$ will have zero covariance between them. This can be seen by multiplying $\mathbf{T}$ by its transpose. The matrix $\mathbf{T}^{\mathrm{T}}\mathbf{T}$ is the covariance matrix of the data matrix $\mathbf{T}$ and it is given by

$$\mathbf{T}^{\mathrm{T}}\mathbf{T} = (\mathbf{VS})^{\mathrm{T}}\mathbf{VS} = \mathbf{S}^2.$$

Therefore, the matrix $\mathbf{V}$ has columns that give the coefficients for a linear combination of the original variables to create p uncorrelated variables.

The rows of the matrix $\mathbf{U}$ are the values of the uncorrelated random variables created by the linear combinations defined by the columns of $\mathbf{V}$, divided by the $\sqrt{\lambda_i}$. To see this we can look at the matrix $\mathbf{T}$:

$$\mathbf{T} = \mathbf{AV} = \mathbf{US},$$

or

$$\mathbf{U} = \mathbf{TS}^{-1}.$$

Additionally, mean of each column in the matrix $\mathbf{U}$ is zero. Therefore, each row of $\mathbf{U}$ contains the values of p uncorrelated, mean-zero random variables.

To summarize, the SVD transforms the original data matrix, into a matrix $\mathbf{U}$ of mean-zero, uncorrelated variables that are re-scaled linear combinations of the original variables. The linear combinations are defined in $\mathbf{V}$, and the scaling is given in the diagonal matrix $\mathbf{S}$.

3.4.1 Approximate Data Matrix

To examine the way the SVD works and see how we can use it to approximate the original matrix, we write it as a sum

$$\mathbf{A} = \sum_{i=1}^{r} \sqrt{\lambda_i}\mathbf{u}_i\mathbf{v}_i^{\mathrm{T}}, \tag{3.44}$$

where $\mathbf{u}_i$ is the ith column of $\mathbf{U}$ and $\mathbf{v}_i$ is the ith column of $\mathbf{V}$. Given that that each term of this sum is a $n \times p$ matrix, we can write an approximation to $\mathbf{A}$ using a subset of the terms. Call the matrix using only k terms in the sum as $\mathbf{A}_k$ such that

$$\mathbf{A}_k = \sum_{i=1}^{k} \sqrt{\lambda_i}\mathbf{u}_i\mathbf{v}_i^{\mathrm{T}}.$$

It can be shown that $\mathbf{A}_k$ is the best rank k approximation to $\mathbf{A}$.

We can interpret this truncated expansion that gives $\mathbf{A}_k$ as the SVD

$$\mathbf{A}_k = \mathbf{U}\mathbf{S}_k\mathbf{V}^{\mathrm{T}},$$

where $\mathbf{S}_k$ has the first k entries of $\mathbf{S}$ and zeros afterward.

On a similar note, we can take the first k columns of $\mathbf{V}$ and call this matrix $\mathbf{V}_k$. Therefore, if we multiply $\mathbf{A}$ by this matrix we get a $n \times k$ matrix $\mathbf{T}_k = \mathbf{A}\mathbf{V}_k$. We can interpret the columns of this matrix as k random variables that approximate the full set of p random variables.

3.4.2 Using the SVD to reduce the number of random variables

As we will see later, the number of input random variables is a strong determinant of the computational cost of performing a UQ study. In such an instance it may be possible to use the SVD to reduce the number of input random variables. If we say that there are nominally p input random variables to our simulation. and we have n samples of those random variables, we can form the matrix $\mathbf{A}$ as described above. Then we can perform the SVD on the matrix and determine how many independent variables there are. For instance, if $r < p$, then we know we can exactly represent the matrix $\mathbf{A}$ using fewer than p random variables.

We can also use the SVD to create a small number of solutions based on the numerical solutions to our model equations. It is also the case that if we have the numerical solution to our model equations at a finite number of points, we can use the SVD to represent the variability in the numerical solution with a handful of uncorrelated random variables. Suppose we know the solution to the model equations at p points, these could be points in any number of dimensions, but we write them as a single vector. For each realization of the input random variables, we will get a different vector. Using these vectors, we can create a data matrix as described above.

Regardless of whether one wants to reduce the number of input or output random variables, it is likely that a large fraction of the variance in the data can be adequately represented by k uncorrelated random variables. We measure the fraction of variance explained to determine how many variables we need. We call the total variance in the data the sum of the λ_i from the SVD. The fraction of variance explained by the random variables in $\mathbf{T}_k$ is written as

$$s_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{r} \lambda_i}.$$

Clearly, the fraction of variance explained is 1 if k is equal to or greater than r.

It is often the case that a few uncorrelated random variables can represent the p correlated variables quite well. That is, with $k \ll r$ the value of the fraction of variance explained can be close to 1. The user can select a value of k that explains an appropriate amount of the total variance for the problem of interest. Once we have selected these k variables we can consider these as our uncertain inputs to our model. We will demonstrate how to select these variables below.

A sketch of the procedure for using SVD to reduce the number of input variables is

1. Select a desired fraction of variance explained, s.
2. Perform the SVD on the data matrix $\mathbf{A}$, and determine the value of k that gives a fraction of variance explained greater than or equal to s.
3. The independent random variables are given by $\mathbf{T}_k = \mathbf{A}\mathbf{V}_k$.
4. To transform to the original random variables compute $\mathbf{A}_k = \mathbf{T}_k\mathbf{V}_k^{\mathrm{T}}$.

We note that the uncorrelated variables produced by the SVD will not necessarily be a standard distribution. One exception is if the data where generated from a multivariate normal. In this case, the uncorrelated variables will be standard normal random variables. If the uncorrelated random variables are not normal, it is possible to fit a distribution

As an example of how this works we will consider the data matrix shown in Table 3.2. This data has $p = 9$, and $n \approx 10^4$. Given the range of units in each of these columns we will first normalize our data so that the columns are mean 0 and standard deviation 1. That is, for each column we subtract the column mean and divide the result by the standard deviation of the column. After this normalization, we take the SVD of the data matrix. It turns out for this data set, with $k = 4$ the fraction of variance explained is 0.9. The fraction of variance explained is shown in Figure 3.14.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
13	46	10	0	2	24	0	1	20
45	93	16	0	17	53	0	0	62
20	46	6	2	0	14	8	5	23
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
51	87	20	2	22	60	0	3	17

Table 3.2 Data matrix for the SVD example.

When we create the approximate data matrix $\mathbf{A}_k$, we are then making an approximation to the full data set. To see how this approximation behaves, we can look at scatter plots of X_1 versus X_2 from various approximations, $\mathbf{A}_k$ give. As shown in Figure 3.15, as k increases the reconstructed data set begins to resemble the original data set.

The columns of the matrix $\mathbf{V}$, which we will call v_i for $i = 1 \ldots 9$, tell us what linear combinations of the original variables us the uncorrelated random variables.

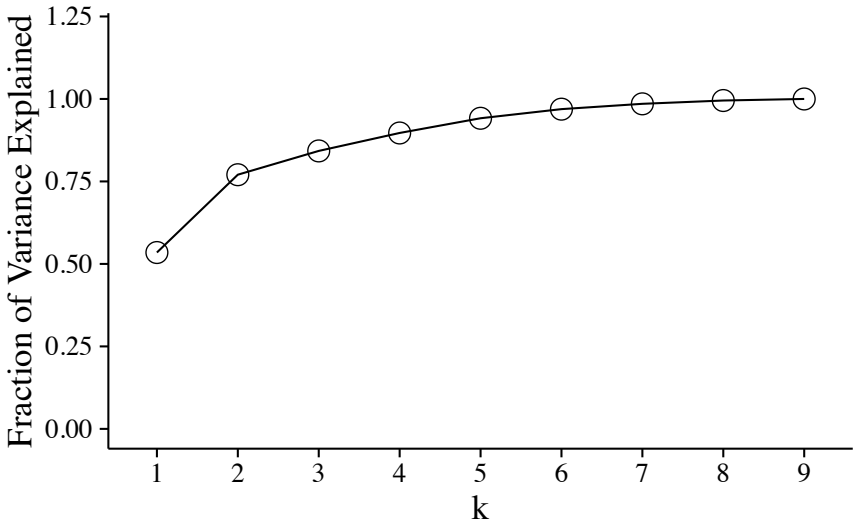

Fig. 3.14 The fraction of variance explained as a function of k for the SVD of the water saturation data.

For this example the matrix **V** is given in Table 3.3. These weights can give us an idea of what are the important features of the data.

	1	2	3	4	5	6	7	8	9
X_1	0.4395	-0.0267	0.0191	-0.0276	0.0497	-0.1530	-0.2828	0.6513	0.5243
X_2	0.4219	-0.0278	-0.2149	0.2823	0.1095	0.0136	-0.6138	-0.1011	-0.5442
X_3	0.3813	0.1060	-0.2970	0.5144	0.2719	0.0045	0.6441	0.0416	-0.0051
X_4	0.1825	-0.4359	-0.6416	-0.5791	-0.0050	0.0939	0.1345	-0.0458	-0.0250
X_5	0.3303	0.3501	0.1525	-0.2686	-0.5914	-0.0170	0.2606	0.2722	-0.4253
X_6	0.3938	0.2765	-0.0438	-0.0159	-0.3143	0.0709	-0.1083	-0.6475	0.4812
X_7	0.1898	-0.5449	0.2943	0.0810	-0.1533	-0.6985	0.1308	-0.2054	-0.0570
X_8	0.1783	-0.5381	0.3456	0.2017	-0.2145	0.6827	0.0648	0.0388	0.0274
X_9	0.3437	0.1089	0.4715	-0.4472	0.6273	0.0907	0.0970	-0.1570	-0.1089

Table 3.3 The matrix **V** for the example data set. Each column gives the weights for a linear combination of the p original random variables X_i.

To aid in the interpretation of the transformed variables, we plot the coefficients for the first three linear combinations (i.e., the first three columns of **V**) in Figure 3.16. From this we can see that most important uncorrelated variable has all positive coefficients, and we can think of this variable as a measure of the overall magnitude of observation i: when all the original variables are large for an observation, then this quantity will be large. Going back to our interpretation of the columns of **U**, an row in the data matrix that has a large value for all the variables will have a large value in the first column of **U** on the appropriate row.

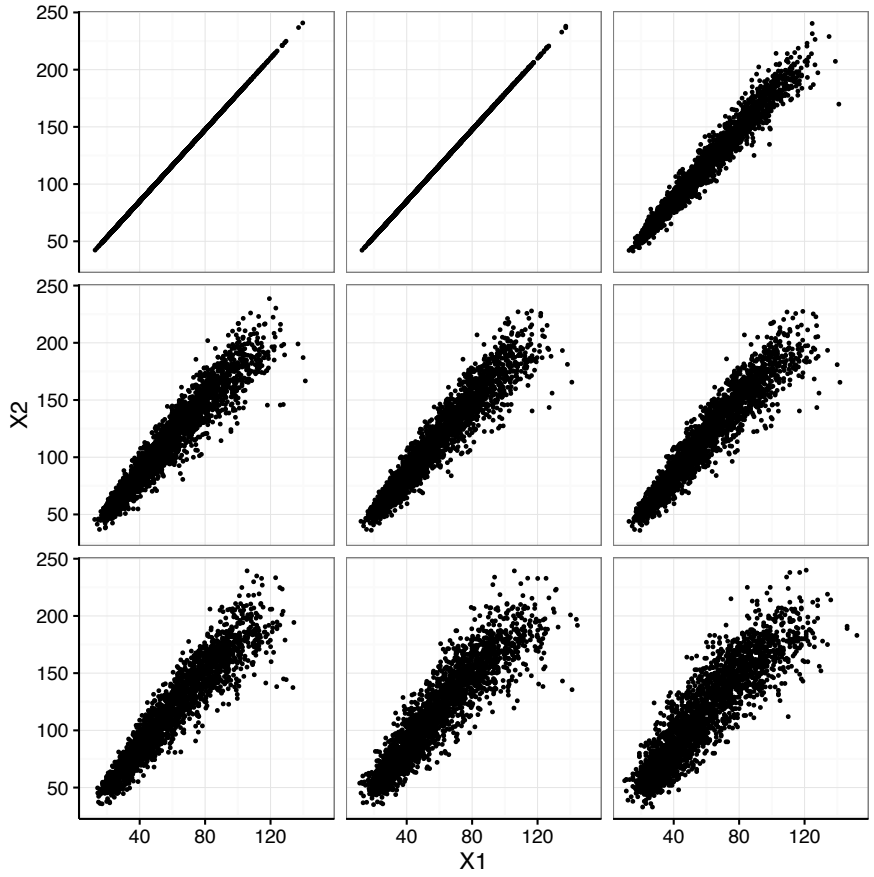

Fig. 3.15 Scatter plots of X_1 versus X_2 for $k = 1 \ldots 9$. The top left plot is $k = 1$ and the bottom right is $k = 9$ (the original data set.

The second variable has large positive weights for X_5 and X_6 and large negative weights for X_4, X_7, and X_8. Therefore, we can interpret the variable as differentiating between those observations that have large values of X_5 and X_6 and those that have large values of X_4, X_7, and X_8. We could continue interpreting the uncorrelated random variables, but it is more important to point out that the SVD is set up so that all the variability in the data is mapped onto these uncorrelated variables.

Additional interpretation of the example

The data used in the above example, was not contrived for the example. It is actually the season offensive statistics for Major League baseball since 1980 for all players having over 200 at bats in a season. The variables $(X_1, \ldots, X_9)$ are

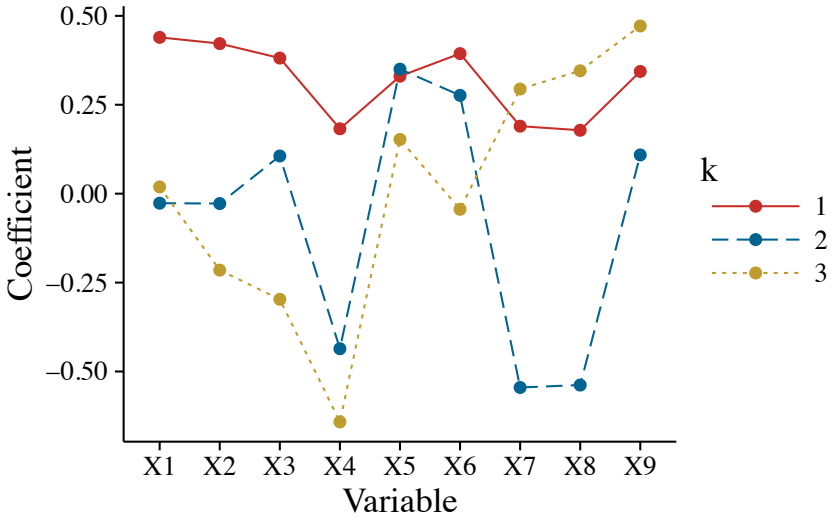

Fig. 3.16 Scatter plots of X_1 versus X_2 for $k = 1 \ldots 9$. The top left plot is $k = 1$ and the bottom right is $k = 9$ (the original data set.

1. Runs
2. Hits
3. Doubles
4. Triples
5. Home runs
6. Runs Batted In (RBI)
7. Stolen Bases
8. Times Caught Stealing
9. Walks

It is useful to know what the data represents so that it can aid in interpreting the variables. Given what the original variables are, we see that the first uncorrelated variable, the value in the first column of **U**, is a measure of the overall magnitude of a player's statistics. Additionally, the second uncorrelated random variable, column 2 of **U**, differentiates between those players with a high number of home runs and runs batted in, the so-called power hitters, and those that have high numbers of triples, stolen bases, and times caught stealing, these are the so-called speedsters. In the data set, the largest value of in the second column of **U** belongs to Mark McGwire in 1998 when he hit 70 home runs in a potentially steroid-tainted campaign. This is the most extreme power hitter in this measure. The lowest value in this column is Rickey Henderson in 1982 when he set the modern-day record for stolen bases with 130 (and was caught 42 times).

When looking at the SVD results in this light, we can see that the coefficients are telling us something about the data. In this case it tells us that one measure of a

baseball player is the amount of power versus speed. These results also indicate that the SVD can be useful even when we are not looking to reduce the data because it can give us a different lens to see how a data is varying.

3.5 The Karhunen-Loève Expansion

The Karhunen-Loève expansion (KL expansion) is the analog of the SVD for a random process. Recall that a random process can be thought of as a collection of random variables where the number of random variables goes to infinity. In this case, we represent the random process as an expansion in basis functions instead of the basis vectors in the **V** matrix in the SVD. To compute the KL expansion we need to know only the mean function, $\mu(x)$, and covariance function, $k(x_1,x_2)$. With this knowledge we can write the KL expansion of a random process $u(x;\boldsymbol{\xi})$ where $x \in [a,b]$ is the deterministic spatial variable and $\boldsymbol{\xi}$ denotes the random component as

$$u(x;\boldsymbol{\xi}) = \mu(x) + \sum_{\ell=0}^{\infty} \sqrt{\lambda_\ell}\xi_\ell g_\ell(x). \tag{3.45}$$

Notice that this form looks nearly identical to the SVD in Eq. (3.44). The ξ_ℓ are random variables with zero mean and unit variance. The ξ_ℓ are also uncorrelated, but they are not necessarily independent.

The λ_ℓ and $g_\ell(x)$ are eigenvalues and eigenfunctions of the covariance function:

$$\int_a^b k(x,y)g_\ell(y)\,dy = \lambda_\ell g_\ell(x). \tag{3.46}$$

The functions $g_\ell(x)$ are orthonormal just like the matrix **V** was orthogonal in the SVD. Also, we order the eigenvalues as we did in the SVD case, $\lambda_1 \geq \lambda_2 \geq \ldots$, and the eigenvalues have a finite sum of squares,

$$\sum_{\ell=0}^{\infty} \lambda_\ell^2 \leq \infty.$$

Determining the eigenvalues and eigenfunctions is not a trivial task as it involves determining the spectrum of an integral operator. There are cases where the solution is known and we will focus on these cases.

For the KL expansion to exist there are some technical details that need to be met by the random process. Firstly, it needs to be square-integrable over the x domain, i.e., the integral of $u^2(x;\boldsymbol{\xi})$ must be finite. Also, the covariance function, must be symmetric, i.e., $k(x,y) = k(y,x)$, and positive definite. If these are satisfied, the KL expansion will exist.

The KL expansion is most useful if the random process is Gaussian. This is because in this case we know that the ξ_ℓ will be independent, standard normal random variables because the sum of normal random variables is normal. If the random

process is not normal, then we know that the ξ_ℓ must not be independent because, by the central limit theorem, that the sum will limit to a normal random variable. Therefore, if the random process is not Gaussian we need more information about the ξ_ℓ.

If we restrict ourselves to a Gaussian random process, we can still model non-Gaussian random processes with the KL expansion. We could do this by writing the random process as a nonlinear transformation of a Gaussian random process. Two possible ways of doing this are with a logarithmic transform, where $\log u(x,\boldsymbol{\xi}) = \hat{u}(x,\boldsymbol{\xi})$, where $\hat{u}(x,\boldsymbol{\xi})$ is a Gaussian random process. The other commonly used approach to transforming a random process is the Nataf transform. This method is beyond the scope of our study, but it does allow a general random process to be represented with a Gaussian random process so that the KL expansion could be used.

3.5.1 Truncated Karhunen-Loève Expansion

The KL expansion turns a random process into a sum over random variables. Therefore, if we truncate the expansion we have effectively discretized it in terms of randomness: rather an infinite collection of random variables we write the process as a finite sum of random variables with known properties. Going back to our definition of the UQ problem in Chapter 1, if we have a calculation that depends on a random process as input, we can consider the ξ_ℓ as our uncertain inputs and get a map to the input random process. The number of terms that we need to keep in the expansion depends on the covariance function, and how fast the λ_ℓ go to zero in magnitude.

3.5.1.1 The exponential covariance

As we mentioned before, the determination of the eigenvalues and eigenvectors of the covariance function is not a trivial task. For a general covariance function this can be quite difficult. There are a handful of cases where the solution is known, and here we will present the results for a simple, but useful, case.

If the covariance function has the form of an exponential of an absolute value:

$$k(x_1,x_2) = ce^{-b|x_1-x_2|}, \tag{3.47}$$

we can find the eigenvalues and eigenvectors exactly. The case we will consider has $x \in [-a,a]$, but we can use these results over any domain provided we define a shifted spatial variable.

The eigenvectors for this covariance function can be expressed in terms of cosines and sines, and we will write the KL expansion in a slightly different way:

$$u(x;\boldsymbol{\xi}) = \mu(x) + \sum_{\ell=0}^{\infty} \left[\sqrt{\lambda_\ell}\xi_\ell g_\ell(x) + \sqrt{\lambda_\ell^*}\xi_\ell^* g_\ell^*(x) \right]. \tag{3.48}$$

The eigenvalues are

$$\lambda_\ell = \frac{2cb}{\omega_\ell^2 + b^2}, \qquad \lambda_\ell^* = \frac{2cb}{\omega_\ell^{*2} + b^2}, \tag{3.49}$$

where the ω_ℓ and ω_ℓ^* are solutions to the transcendental equations

$$b + \omega\tan(\omega a) = 0, \qquad b - \omega^*\tan(\omega^* a) = 0. \tag{3.50}$$

The eigenfunctions are

$$g_\ell = \frac{\cos(\omega_\ell x)}{\sqrt{a + \frac{\sin(2\omega_\ell a)}{2\omega_\ell}}}, \tag{3.51}$$

and

$$g_\ell^* = \frac{\sin(\omega_\ell x)}{\sqrt{a - \frac{\sin(2\omega_\ell a)}{2\omega_\ell}}}. \tag{3.52}$$

The value of b has an important impact on the eigenvalues. A smaller value of b makes the eigenvalues decay to zero faster than a larger value. This is demonstrated in Figure 3.17 where the first ten eigenvalues of the exponential covariance are shown for several values of b. When b is large, the eigenvalues decay slowly. This implies that we can capture the behavior of the random process with a few terms in the KL expansion.

To demonstrate how the KL expansion behaves as more terms are added, we show a single realization of a random process in Figure **??**. In that figure, we see that with two terms in the KL expansion (in this case one λ^* and one λ term) give a smooth, slowly varying function. As the number of terms increases the complexity of the realization increases by having finer scale variations in the solution: at 10 terms there is more variability in the solution, and by 100 terms there are sharp oscillations at a very fine scale.

Another way to look at the behavior of the expansion is to compare several realizations of a random process with different expansion orders to the full process. This comparison is made in Figure 3.18 using the same random process as in Figure 3.17. In this figure the full random process will have different realizations than the approximate process, but the character of the processes should be similar. When comparing the full process with the low order expansions (e.g., two and ten terms), there is much less structure in the low order expansions. However, as more and more terms are added the character of the expansion approaches that of the full process. In many cases, the fine scale structure of the process is not what is important, rather the overall behavior is of interest. If this were the case, we would likely be able to adequately model this process with just a few terms in the expansion.

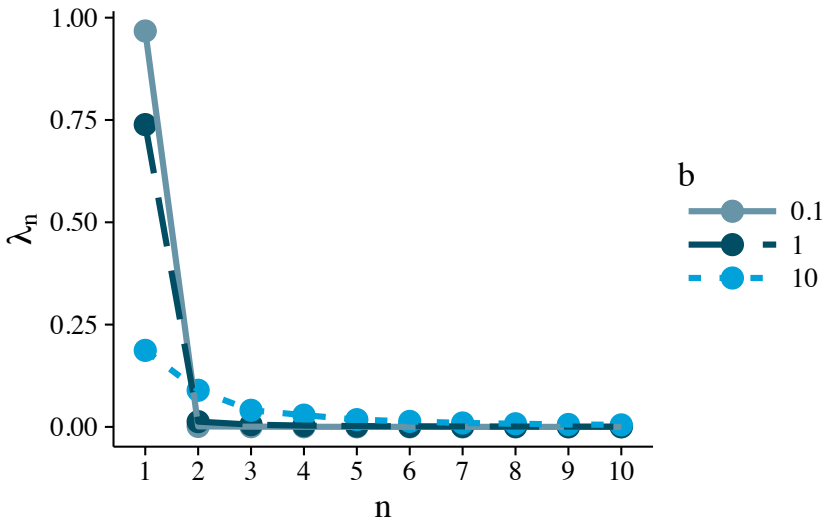

Fig. 3.17 The eigenvalues λ_n and λ_n^* for various values of b and $a = 0.5$ and $c = 1$. The odd n are λ_n^*, and even n are λ_n.

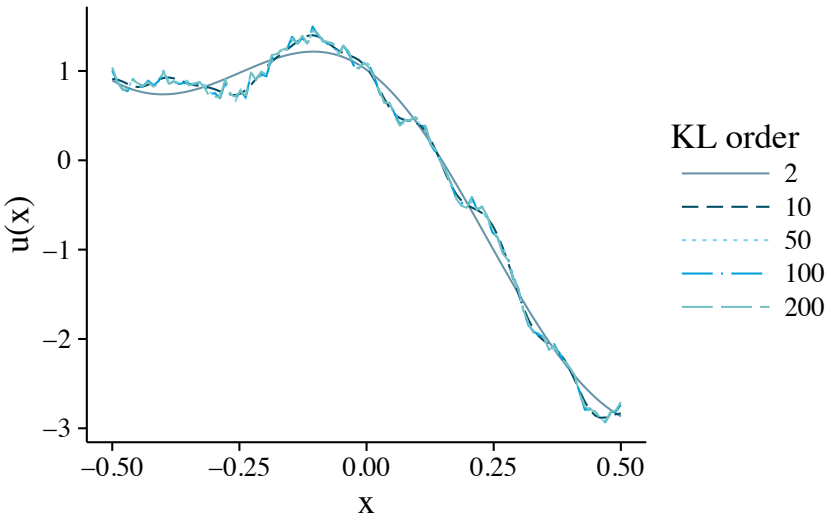

Fig. 3.18 A single realization of a Gaussian random process over $[-0.5, 0.5]$ with $\mu(x) = \cos 2\pi x$, and an exponential covariance with $b = c = 1$ using various number of expansion terms.

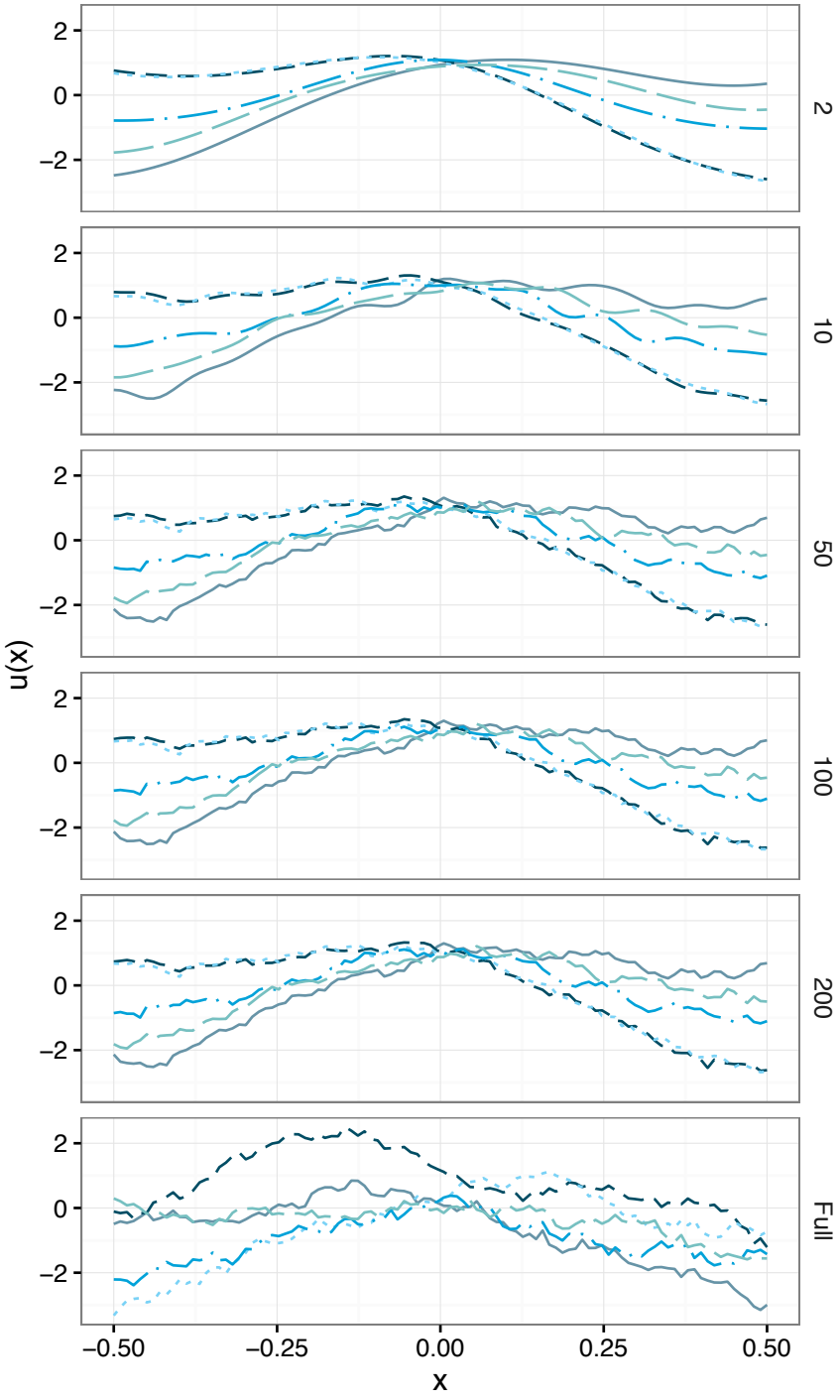

Fig. 3.19 A five realizations of a Gaussian random process over $[-0.5, 0.5]$ with $\mu(x) = \cos 2\pi x$, and an exponential covariance with $b = c = 1$ at different number of expansion terms compared with the full process.

3.6 Choosing Input Parameter Distributions

One basic question regarding an uncertain parameter regards how we would like to represent that uncertainty. From the previous chapter, we know that once we have a CDF or PDF for a random variable that we can then compute quantities like the mean, variance, and any number of other properties of the distribution. Nevertheless, it is generally not possible have a unique mapping the other way: to go from moments of the distribution, e.g., mean, variance, skewness, kurtosis, etc. to produce a PDF or CDF.

Unfortunately, we usually do not know the distribution of our input parameters. It is much more typical to have some number of samples from the distribution. For instance, if the system we are interested in simulating has manufactured parts and the properties of those parts have a distribution we will be able to take a number of parts and measure the properties. This gives us samples from the distribution of the properties, from which we can estimate moments like the mean and variance. However, we cannot robustly quantify the behavior of the tails of the distribution from a small number of samples. This is because, by definition, our samples will, with high probability, not have any values from the tails of the distribution. Therefore, the best we can do is make a guess as to the tail behavior of the system. We need to acknowledge that we have made this assumption about the tail behavior of the distribution, and not make overly specific claims about the probability of a tail event is.

A common approach to modeling a random variable is to select a distribution from the standard set of distributions (such as those provided in the appendix). There are several considerations that are important when selecting a distribution for an input random variable. For a given parameter we want the distribution we assume it follows to be consistent with the parameter in the following regards

1. The range, e.g., real numbers, positive real numbers, or a certain range, and
2. The known moments, or other properties of the distribution, e.g., mean, median, variance, or various quantiles.

The first of these conditions can eliminate many possible distributions. For instance, if we know the parameter can only take on a range of values or is positive, then we know that we cannot use a normal distribution without an ad hoc procedure for ignoring the probability of getting an invalid parameter. The known information about the parameter's behavior will also eliminate some possible distributions. As an example, if the parameter is known to possess some skewness or excess kurtosis, then a normal distribution will not be able to capture those properties. Once a distribution is chosen, then one can fit the remaining known information about the distribution. That is, select the parameters of the distribution so that the input random variable's properties are preserved.

Many times it will not be the case that all of the desired properties of the distribution can be fit with a standard distribution. It may be the case that a standard distribution is not flexible enough to reproduce the desired properties (e.g., there is a fixed relationship between moments of the distribution). In this case one could

compromise and decide to not match all of the desired properties. The other possibility is to blend distributions together to get the desired properties. For instance, if the desired distribution is multi-modal, i.e., it has multiple local maxima in the PDF, one could write this as the sum of normal distributions and fit the mean and standard deviation of each normal to match the desired distribution.

3.6.1 Choosing Joint Distributions

It is potentially even more complicated to choose a joint distribution for a set of inputs. We have already mentioned that in general one will not know much about the joint distribution functions for a collection of random variables. Therefore, it is typical to be less constrained in the selection of the joint distribution, and this freedom can be a double-edged sword. We have already mentioned that choosing a joint distribution, through the selection of copula, that does not have any tail dependence can lead to erroneous conclusions about the probability of the parameters going to extremes together.

One of the measures we want to match for a joint distribution is a measure of correlation between the variables. This could be any of the measures we discussed: Pearson or Spearman correlation or Kendall's tau. We also noted in our discussion of copulas that it may be possible to produce a desired tail dependence in the joint distribution. Nevertheless, it is likely not possible to match both the correlation and tail dependence. Therefore, one often has to make a decision as to which feature is more important for the analysis being performed.

If the uncertainty analysis being performed is looking for understanding the behavior of the system under conditions near the median inputs, then the tail dependence of the distribution is less important than the measure of the correlation. In such a situation it is reasonable to choose a joint distribution without tail dependence. It is not reasonable, however, to then use this distribution to make statements about extreme events using this joint distribution.

In the case where one cares about distribution of system performance near the median inputs and also wants to make assertions of the system behavior near the tails of the distribution, it is possible to use both distributions. For instance, one could perform an analysis using a joint distribution that has zero tail dependence and use this to quantify the system behavior near the nominal inputs. Then, to predict the behavior at the extremes use a different joint distribution that does have tail dependence. This analysis should make clear the caveat that the behavior near the nominal inputs and near the extremes were produced using different assumptions about the distributions.

3.6.2 Distribution Choice as a Source of Epistemic Uncertainty

In the selection of a distribution for input parameters, there are necessarily assumptions that are made. These assumptions are a type of epistemic uncertainty in the uncertainty modeling. For the distribution of a single parameter, i.e., its marginal distribution, the behavior of that distribution in the tails could have an impact on the conclusions of the analysis. For instance, if one is interested in the percentage of time the system's maximum temperature exceeds some threshold, one could get an answer of 0.01% using normal distributions for the input parameters, and 0.05% using a t-distribution for the parameters. Given that we do not actually know which is the is correct distribution to use, the range 0.01 to 0.05% is the epistemic uncertainty in the result.

Furthermore, the assumptions on the joint distribution lead to epistemic uncertainty. For a given measure of relation between two variables there are an infinite number of joint distributions that could match this quantity. In fact, we discussed several possible joint distributions when we discussed copulas. Each of these joint distributions has properties that could affect an uncertainty analysis. For example, both the Frank copula and the normal copula could match any particular value of Kendall's tau, but the behavior of the joint distribution is not the same: when we look at samples from the joint distributions Frank gives an almost rectangular distribution versus the elliptically-shaped normal copula.

Robustness to outliers (underestimating them) Tail dependence

3.7 Problems

1. Assume you have a 100 samples of a pair of random variables (X_1, X_2) that have a positive correlation, call this set of pairs, $\mathbf{A}_1$. You then draw another 100 samples and call this set $\mathbf{A}_2$. The Pearson correlation between (X_1, X_2) in $\mathbf{A}_1$ is positive and the Pearson correlation between (X_1, X_2) in $\mathbf{A}_2$. What can you say about the Pearson correlation for all 200 samples?
2. For the following data, compute by hand the Pearson and Spearman correlations and Kendall's tau.

X_1	X_2
55.01	82.94
54.87	55.02
57.17	85.18
36.01	-84.27
35.88	-106.30
36.33	-119.65
43.49	-112.03
41.44	-71.69
54.43	-3.50
36.47	140.57

3. Demonstrate the tail dependence of a bivariate normal random variable is 0.
4. Another Archimedean copula is the Joe copula with generator

$$\varphi_{\mathrm{J}}(t) = -\log\left(1-(1-t)^{\theta}\right),$$

and

$$\varphi_{\mathrm{J}}^{-1}(t) = 1-(1-\exp(-t))^{1/\theta}.$$

a. Compute the bivariate copula for this generator.
b. Derive the upper and lower tail dependence for this copula.
c. Compute the value of Kendall's tau for this copula
d. Generate 1000 samples from the copula with standard normal marginals and a value of Kendall's tau of 0.6.