United States Department of Agriculture

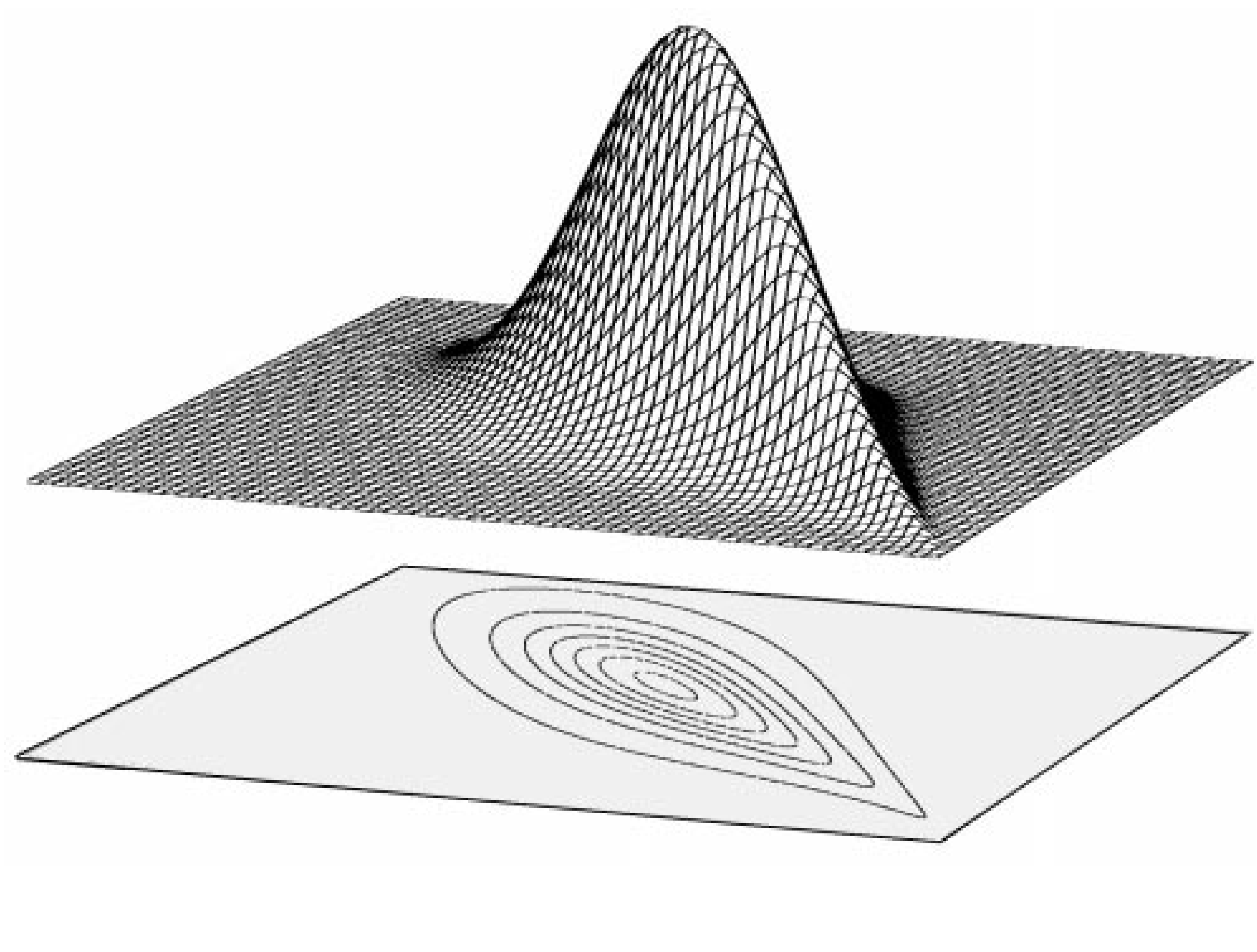
Forest Service

Forest Products Laboratory

Research Paper FPL–RP–575

Some Bivariate Distributions for Modeling the Strength Properties of Lumber

Richard A. Johnson
James W. Evans
David W. Green



Abstract

Accurate modeling of the joint stochastic nature of the strength properties of dimension lumber is essential to the determination of reliability-based design safety factors. This report reviews the major techniques for obtaining bivariate distributions and then discusses bivariate distributions whose marginal distributions suggest they might be useful for modeling the joint distribution of two strength properties. Finally, we pick a bivariate Weibull distribution and show that we can write its likelihood function under a proof loading scheme, offering the possibility that it can be used to model the joint distribution of two properties that must each be measured using a destructive test.

Keywords: bivariate distribution, Weibull, estimation, proof loading

August 1999

Johnson, Richard A.; Evans, James W.; Green, David W. 1999. Some bivariate distributions for modeling the strength properties of lumber. Res. Pap. FPL–RP–575. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 11 p.

A limited number of free copies of this publication are available to the public from the Forest Products Laboratory, One Gifford Pinchot Drive, Madison, WI 53705–2398. Laboratory publications are sent to hundreds of libraries in the United States and elsewhere.

The Forest Products Laboratory is maintained in cooperation with the University of Wisconsin.

Contents

Some Bivariate Distributions for Modeling the Strength Properties of Lumber

Richard A. Johnson, Professor
University of Wisconsin, Madison, Wisconsin

James W. Evans, Supervisory Mathematical Statistician
David W. Green, Supervisory Research General Engineer
Forest Products Laboratory, Madison, Wisconsin

Introduction

Essential to any realistic determination of reliability-based safety factors for wood structures is the accurate modeling of the joint stochastic nature of two or more strength properties of single members. In this report, we will review some existing bivariate distributions whose marginal distributions[1] suggest that they might be useful for modeling the joint distribution of two strength properties. We must address two situations in assessing the possible usefulness of a bivariate distribution: (1) It is sometimes possible to observe two properties on the same specimen, such as modulus of rupture (MOR) and modulus of elasticity (MOE). In this case, information is often lacking on the behavior of estimators even when both components can be observed simultaneously. (2) But often it is impossible to measure both properties on the same experimental unit, such as ultimate tensile stress (UTS) and ultimate compressive stress (UCS). For a bivariate distribution to be useful in this case, the parameters in it must be amenable to estimation under a proof load scheme where a specimen is loaded in strength mode 1 up to a specified load L. If it fails, we record its mode 1 strength. Otherwise, we load the specimen in strength mode 2 until failure. With regard to estimation when strength data must be obtained from proof loading schemes, only the bivariate normal distribution has been previously considered (de Amorim and Johnson 1986, Evans and others 1984, Green and others 1984, Green and Evans 1983).

[1]If two random variables X and Y have a joint density $f(x, y)$, the marginal distribution of X is the distribution defined by the density where the y variable is integrated out, that is,

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

In practical terms, it is the distribution we would get for X if we never measured Y.

This report proceeds in four parts. First we review some of the major techniques for obtaining bivariate distributions. Next we look at some candidate bivariate distributions. Then we pick a promising bivariate distribution and evaluate our ability to estimate its parameters when both strength properties can be measured on the same specimen. Finally, we show for this distribution that we can write its likelihood function under a proof loading scheme, offering the possibility that its parameters can be estimated through a search procedure that maximizes the likelihood.

Procedures for Obtaining Bivariate Distributions

When searching for bivariate distributions to model strength properties of lumber, it seems reasonable to restrict our search to those distributions whose marginal distributions have proven satisfactory for describing a single strength property. A number of families of bivariate distributions are known that have marginal distributions of a specified form. Johnson and Kotz (1972) discuss several procedures of construction, and we reviewed those that seem most promising for our purposes.

Transformation Methods

In one method of construction, distributions with specified marginals are generated from families of bivariate distributions that have uniform marginal distributions. For instance, if (U_1, U_2) has a bivariate distribution and U_1 and U_2 are each distributed uniformly on (0, 1), then

$$X = F_1^{-1}(U_1) \quad \text{and} \quad Y = F_2^{-1}(U_2)$$

have a joint distribution where X has marginal cumulative distribution function (cdf) F_1 and Y has marginal cdf F_2. Genest and MacKay (1986) studied the construction of bivariate distributions with uniform marginals.

In a closely related transformation approach, families of bivariate distributions are obtained by transforming to standard bivariate normal variables (Z_1, Z_2) having correlation ρ. Johnson and Kotz (1972) review the original work of Johnson (1949a). In particular, they consider the transformations

$$Z_1 = \gamma_1 + \delta_1 h_1\left(\frac{X-\zeta_1}{\lambda_1}\right)$$

$$Z_2 = \gamma_2 + \delta_2 h_2\left(\frac{Y-\zeta_2}{\lambda_2}\right)$$

where, in their approach, $h_1(v)$ and $h_2(v)$ are each restricted to be one of the following four transformations: log(v) for a lognormal, v for a normal, $\log(v/(1-v))$ for the S_B, and $\sinh^{-1}(v)$ for the so-called S_U distribution. Thus, the choice is somewhat limited.

In general, there are nine parameters, the correlation coefficient $\rho = \mathrm{Corr}(Z_1, Z_2)$, and the four parameters associated with each marginal distribution (γ_1, δ_1, ζ_1, λ_1) and (γ_2, δ_2, ζ_2, λ_2). Most data sets of less than a few hundred observations do not yield enough information to fit nine parameters. Fitting such distributions to small or moderate-sized data sets will result in "overfit." That is, the fit to the particular data will look good (too good), but it may not produce accurate predictions for future values from the same population of strength values. The case $\zeta_1 = \zeta_2 = 0$ and $\lambda_1 = \lambda_2 = 1$ yields the standard forms of these distributions. If the standard form is fit, the number of parameters estimated from the data is reduced to five.

Johnson and Kotz (1972) presented several examples including the case where (X, Y) is normal–lognormal so

$$Z_1 = \gamma_1 + \delta_1\left(\frac{X-\zeta_1}{\lambda_1}\right)$$

$$Z_2 = \gamma_2 + \delta_2 \log\left(\frac{Y-\zeta_2}{\lambda_2}\right)$$

For these transformations, γ_1, δ_1, and γ_2 are redundant and there are six parameters in the reduced form.

For this whole class of distributions, it is the median rather than the mean regression functions of Y on X that have easily described forms (Johnson and Kotz 1972, table 2, p. 17). Median regressions are equations to predict the median of Y when $X = x$. They are functions of the parameters and x.

Farlie–Gumbel–Morgenstern Families

Another approach is to judiciously choose a formula for the joint cdf that includes the marginal cdf's and some parameters. A series of papers, by several different authors, produced the so-called Farlie–Gumbel–Morgenstern families of bivariate cdf's,

$$F(x,y) = F_1(x)F_2(y)\left[1+\delta\,\{1-F_1(x)\}\{1-F_2(y)\}\right]$$

where $|\delta| \le 1$ for a bivariate density to exist. The marginal distributions are F_1 and F_2. Although the case of independence, $\delta = 0$, is included, the range of dependence is quite restricted. Schucany and others (l978) show that the correlation of X and Y can never exceed 1/3 in absolute value, and in many cases, such as the bivariate exponential distribution, the bound reduces to 0.25.

This bivariate exponential distribution has cdf

$$F(x,y) = \left(1-e^{-x/\theta_1}\right)\left(1-e^{-y/\theta_2}\right) \times\left[1+\delta\, e^{(-x/\theta_1)+(-y/\theta_2)}\right]$$

The term $\delta\{1-F_1(x)\}\{1-F_2(y)\}$ can be replaced by a slightly more general product term (Conway 1983).

Huang and Kotz (1984) studied an iterated Farlie–Gumbel–Morgenstern family with a single iteration,

$$\begin{aligned} F(x,y) &= F_1(x)F_2(y) + \delta F_1(x)F_2(y)\{1-F_1(x)\}\{1-F_2(y)\} \\ &\quad + \beta\left[F_1(x)F_2(y)\right]^2\{1-F_1(x)\}\{1-F_2(y)\} \\ &= F_1(x)F_2(y)\left[1+\delta\{1-F_1(x)\}\{1-F_2(y)\}\right. \\ &\quad \left. + \beta F_1(x)F_2(y)\{1-F_1(x)\}\{1-F_2(y)\}\right] \end{aligned}$$

and showed that the maximum correlation can be substantially increased. However, the complicated form of the density is likely to preclude the development of statistical procedures.

Mixture Models

Marshall and Olkin (1988) provided a unified approach to obtain many of the previously derived families of distributions as well as new ones. This approach begins with a bivariate distribution of a special form,

$$G(x,y|\theta_1,\theta_2) = F_1^{\theta_1}(x)F_2^{\theta_2}(y)$$

which depends on the parameters (θ_1, θ_2). Let $K(\theta_1, \theta_2)$ be a mixing distribution that assigns probability 1 to positive values of θ_1 and θ_2. Then the mixture

$$H(x,y) = \int_0^\infty\int_0^\infty F_1^{\theta_1}(x)F_2^{\theta_2}(y)\,dK(\theta_1,\theta_2)$$

defines a bivariate cdf.

Let

$$\phi_1(s) = \int_0^\infty e^{-s\theta_1} dK(\theta_1)\ , \qquad \phi_2(s) = \int_0^\infty e^{-s\theta_2} dK(\theta_2)$$

be the Laplace transforms of the marginal distributions of $K(\theta_1, \theta_2)$. If it is desired that $H(x,y)$ have specified marginal distributions $H_1(x)$ and $H_2(y)$, this is accomplished by taking

$$F_1 = e^{-\phi_1^{-1}(H_1(x))} \qquad \text{and} \qquad F_2 = e^{-\phi_2^{-1}(H_2(y))}$$

as verified by Marshall and Olkin (1988). They also allow for a more general integrand than

$$F_1^{\theta_1}(x)\, F_2^{\theta_2}(x)$$

Numerous examples of general families are given, but inference procedures are not discussed. The mixture approach does not necessarily lead to nice forms for the densities that are amenable to inference. It does, however, often lead to simple procedures for simulating observations (X,Y) from $H(x,y)$.

Some Candidate Bivariate Models for Strength Properties

In previous investigations, we studied statistical estimation procedures for the case where the two strength properties follow a bivariate normal distribution (Evans and others 1984, Green and others 1984, Green and Evans 1983). Our ultimate goal is to develop similar procedures for non-normal distributions. Based on our literature search thus far, we found that at least four major categories of distributions merit serious consideration as candidates for further study. The properties of the models still require further study, and procedures for statistical inference need to be investigated.

Bivariate Weibull Distributions

One potential category of bivariate distributions is the family of bivariate Weibull distributions. One potentially useful bivariate Weibull distribution has marginal distributions that are two-parameter Weibull distributions. Its survival function is given by

$$\begin{aligned}\bar{F}(x,y) &= P\left[X > x, Y > y\right] \\ &= \exp\left\{-\left[\left(\frac{x}{\theta_1}\right)^{\beta_1/\delta} + \left(\frac{y}{\theta_2}\right)^{\beta_2/\delta}\right]^{\delta}\right\}, \qquad 0 < \delta \le 1\end{aligned}$$

Hougaard (1986) derives this model as a mixture.

While the marginal distributions are well understood, the reasonableness of the bivariate model should be confirmed by fitting some actual data sets. Its density is complicated, and we will have to determine if any estimation procedure is numerically tractable. Lu and Bhattacharyya (1990) give the moments

$$\begin{aligned}E(X) &= \theta_1\Gamma(1/\beta_1 + 1) \\ \mathrm{Var}(X) &= \theta_1^2\left\{\Gamma(2/\beta_1 + 1) - \Gamma^2(1/\beta_1 + 1)\right\} \\ E(Y) &= \theta_2\Gamma(1/\beta_2 + 1) \\ \mathrm{Var}(Y) &= \theta_2^2\left\{\Gamma(2/\beta_2 + 1) - \Gamma^2(1/\beta_2 + 1)\right\}\end{aligned}$$

$$\begin{aligned}&\mathrm{Cov}(X,Y) \\ &= \theta_1\theta_2\left[\Gamma\left(\frac{\delta}{\beta_1}+1\right)\Gamma\left(\frac{\delta}{\beta_2}+1\right)\Gamma\left(\frac{1}{\beta_1}+\frac{1}{\beta_2}+1\right)\right. \\ &\qquad \left. - \Gamma\left(\frac{1}{\beta_1}+1\right)\Gamma\left(\frac{1}{\beta_2}+1\right)\Gamma\left(\frac{\delta}{\beta_1}+\frac{\delta}{\beta_2}+1\right)\right] \\ &\quad \div \Gamma\left(\frac{\delta}{\beta_1}+\frac{\delta}{\beta_2}+1\right)\end{aligned}$$

where Γ is a gamma function. The correlation

$$\mathrm{Corr}(X,Y) = \mathrm{Cov}(X,Y) \Big/ \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$$

is a complicated function of δ, β_1, and β_2. Lu and Bhattacharyya (1990) looked at the correlation when $\beta_1 = \beta_2 = \beta$ for β = 0.1, 0.25, 0.5, 1, 2, 4, and 10 and showed that the correlation is almost linear in δ for equal shape parameters $\beta_1 = \beta_2 = 1$, which could enhance estimation and/or simulation procedures. A bivariate distribution having the three-parameter Weibull distribution as marginals is obtained by replacing x by $x - \alpha_1$ and y by $y - \alpha_2$.

The bivariate Weibull distribution is also related to the bivariate extreme value distribution (Gumbel l960). If (X,Y) has the bivariate Weibull distribution discussed above, then $(\log(X), \log(Y))$ has the bivariate extreme value distribution. In particular, setting $X_2 = \log X$ and $Y_2 = \log Y$, we have

$$\begin{aligned}\bar{F}_{X_2Y_2}(x_2, y_2) &= P[X > e^{x_2}, Y > e^{y_2}] \\ &= \exp\left\{-\left[\exp[(x_2 - \log\theta_1)/(\delta/\beta_1)]\right.\right. \\ &\qquad \left.\left. + \exp[(y_2 - \log\theta_2)/(\delta/\beta_2)]\right]^{\delta}\right\}\end{aligned}$$

Notice that the survivor functions for the marginal distributions

$$\begin{aligned}\bar{F}_{X_2}(x_2) &= \exp\left\{-\exp\left[(x_2 - \log\theta_1)/(1/\beta_1)\right]\right\} \\ \bar{F}_{Y_2}(y_2) &= \exp\left\{-\exp\left[(y_2 - \log\theta_2)/(1/\beta_2)\right]\right\}\end{aligned}$$

are location-scale families. This fact allows us to find alternative estimators for the marginal parameters (Lawless 1982).

Another bivariate Weibull family, given by Lu and Bhattacharyya (1990), has survival function

$$\overline{F}(x,y) = P[X > x, Y > y]$$
$$= \exp\left\{-\left((x/\theta_1)^{\beta_1} + (y/\theta_2)^{\beta_2} + \delta\left[1 - e^{-(x/\theta_1)^{\beta_1}}\right]\right.\right.$$
$$\left.\left.\times\left[1 - e^{-(y/\theta_2)^{\beta_2}}\right]\right)\right\}, \quad 0 < |\delta| \le 1$$

This distribution has correlation restricted to the range –0.20 to 0.32 and is unlikely to be amenable to inferences.

The bivariate exponential distributions of Gumbel (1960) (the Farlie–Gumbel–Morgenstern family) and Marshall and Olkin (1967) can be converted to Weibull distributions by replacing x by x^{β_1} and y by y^{β_2} (Lu and Bhattacharyya 1990). The Marshall and Olkin (1967) model suffers from the drawback that $P[X = Y]$ is positive. In the context of system life length, the probabilistic interpretation is that a Poisson shock will kill component one (X), a second Poisson shock will kill component two (Y), and a third Poisson shock will kill both components. As a consequence of this third shock, there is positive probability that $X = Y$, a property unlikely to hold for two strength properties of lumber. The Freund (1961) bivariate extension does not have exponential marginal distributions.

Bivariate Inverse Gaussian Distribution

The univariate inverse Gaussian distribution has been proposed as a model for strength properties of various materials but has not yet been emphasized in the wood literature. Recently, Kocherlakota (1986) developed a bivariate distribution whose marginal distributions are univariate inverse Gaussian distributions. Kocherlakota (1986) gives expressions for conditional distributions and the joint density

$$f(x,y) = \frac{1}{4\pi}\sqrt{\frac{\lambda_1\lambda_2}{x^3y^3(1-\rho^2)}}$$
$$\times\left[\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{\lambda_1(x-\mu_1)^2}{\mu_1^2 x} - \frac{2\rho}{\mu_1\mu_2}\left(\frac{\lambda_1\lambda_2}{xy}\right)^{1/2}\right.\right.\right.$$
$$\left.\left.\times(x-\mu_1)(y-\mu_2) + \frac{\lambda_2(y-\mu_2)^2}{\mu_2^2 y}\right]\right\}$$
$$+\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{\lambda_1(x-\mu_1)^2}{\mu_1^2 x} + \frac{2\rho}{\mu_1\mu_2}\left(\frac{\lambda_1\lambda_2}{xy}\right)^{1/2}\right.\right.$$
$$\left.\left.\left.\times(x-\mu_1)(y-\mu_2) + \frac{\lambda_2(y-\mu_2)^2}{\mu_2^2 y}\right]\right\}\right]$$

Kocherlakota (1986) also mentions a disturbing fact that, with high correlation, the distribution may have several peaks. This behavior should be verified and its extent better understood. If true for realistic ranges of correlation, it is unlikely to be a reasonable model for two strength properties of lumber.

Bivariate S_{BB} Distribution

It has been suggested (Pearson 1980, Warren 1979) that the S_B family of distributions provides a reasonable model for a single strength property and that a bivariate version, the S_{BB} family (Johnson 1949a), provides a reasonable model for two strength properties. This bivariate distribution is related via a transformation to the bivariate normal distribution. Specifically, X has the S_B distribution on the interval a_1 to b_1, if for some γ_1 and δ_1

$$Z_1 = \gamma_1 + \delta_1 \log\left[\frac{X - a_1}{b_1 - X}\right]$$

is distributed as a standard normal distribution. Draper (1952) and Johnson (1949b) discuss estimation and give numerical examples.

Sometimes practitioners can use previous studies to fix the lower bound a_1 and thus reduce the number of parameters that must be estimated. However, this may not be easy in lumber applications, and when four parameters are fit to small or moderate-sized data sets, there is likely to be "overfit." If also

$$Z_2 = \gamma_2 + \delta_2 \log\left[\frac{Y - a_2}{b_2 - Y}\right]$$

is distributed as a standard normal and (Z_1, Z_2) has a bivariate normal distribution with correlation ρ, then (X, Y) is said to have the bivariate S_{BB} distribution on the intervals (a_1, b_1) and (a_2, b_2). The bivariate S_{BB} distribution has the single additional parameter ρ, which is the correlation between the normal variables Z_1 and Z_2. On this transformed scale, the parameters (γ_1, δ_1, a_1, b_1) are estimated from the observations on X and the parameters (γ_2, δ_2, a_2, b_2) are estimated from the observations on Y.

Because there are nine parameters to estimate, there may be "overfit" unless the data set is large. The excessive number of parameters may also cause additional difficulties with estimation from proof load data.

Bivariate Normal–Lognormal Distribution

Johnson (1949a) also describes a transformed distribution called a normal–lognormal distribution. If for some (γ_1, δ_1) and (γ_2, δ_2)

$$Z_1 = \gamma_1 + \delta_1 X$$

$$Z_2 = \gamma_2 + \delta_2 \log Y$$

are standard normal variables that have a joint normal distribution with correlation ρ, then (X, Y) is said to have a bivariate normal–lognormal distribution. Slightly more generally, Y could be replaced by $Y - a_2$ if a_2 is a lower bound for the values of Y.

Picking a Promising Bivariate Distribution

Because of the dearth of bivariate distributions with appropriate marginal distributions, no other good candidates are available at this time. The correlation structure of the Farlie–Gumbel–Morgenstern distributions is too restrictive. The inverse Gaussian has a disturbing property that the density has several peaks when the correlation is high. Consequently, we will not pursue estimation of the bivariate inverse Gaussian at this time. The only viable candidate distributions appear to be the bivariate Weibull, the bivariate S_{BB} distributions, and the bivariate normal–lognormal distribution on the basis of complete samples $(x_1, y_1), \ldots, (x_n, y_n)$. The bivariate Weibull has the advantage over the S_{BB} distribution of having fewer parameters that we would need to estimate, meaning we might not need as large of sample sizes in a data set. In addition, the univariate Weibull distribution has proved very useful for modeling individual strength properties. We need to evaluate our capability of estimating the five parameters of this distribution to know if it can be done effectively.

Estimation of the Bivariate Weibull Parameters

Each marginal distribution has a shape parameter β_i and a scale parameter θ_i. The fifth parameter δ introduces a non-negative correlation between the two components. This correlation is the complicated function of δ, β_1, and β_2 discussed in the Bivariate Weibull Distributions section. The probability density function (pdf) for the bivariate Weibull is

$$\frac{\beta_1}{\theta_1}\left(\frac{x}{\theta_1}\right)^{(\beta_1/\delta)-1}\frac{\beta_2}{\theta_2}\left(\frac{y}{\theta_2}\right)^{(\beta_2/\delta)-1}\left\{\left(\frac{x}{\theta_1}\right)^{\beta_1/\delta}+\left(\frac{y}{\theta_2}\right)^{\beta_2/\delta}\right\}^{\delta-2}$$
$$\times\left\{\left[\left(\frac{x}{\theta_1}\right)^{\beta_1/\delta}+\left(\frac{y}{\theta_2}\right)^{\beta_2/\delta}\right]^{\delta}+\frac{1}{\delta}-1\right\}$$
$$\times\exp\left\{-\left[\left(\frac{x}{\theta_1}\right)^{\beta_1/\delta}+\left(\frac{y}{\theta_2}\right)^{\beta_2/\delta}\right]^{\delta}\right\}$$

so the log of the likelihood for a random, uncensored sample is

$$\log L = n\log\left(\frac{\beta_1}{\theta_1}\right)+n\log\left(\frac{\beta_2}{\theta_2}\right)$$
$$+\left(\frac{\beta_1}{\delta}-1\right)\sum_{i=1}^{n}\log\frac{x_i}{\theta_1}+\left(\frac{\beta_2}{\delta}-1\right)\sum_{i=1}^{n}\log\frac{y_i}{\theta_2}$$
$$+(\delta-2)\sum_{i=1}^{n}\log\left[\left(\frac{x_i}{\theta_1}\right)^{\beta_1/\delta}+\left(\frac{y_i}{\theta_2}\right)^{\beta_2/\delta}\right]$$
$$+\sum_{i=1}^{n}\log\left\{\left[\left(\frac{x_i}{\theta_1}\right)^{\beta_1/\delta}+\left(\frac{y_i}{\theta_2}\right)^{\beta_2/\delta}\right]^{\delta}+\frac{1}{\delta}-1\right\}$$
$$-\sum_{i=1}^{n}\left[\left(\frac{x_i}{\theta_1}\right)^{\beta_1/\delta}+\left(\frac{y_i}{\theta_2}\right)^{\beta_2/\delta}\right]^{\delta}$$

Maximum likelihood estimates of the parameters can be obtained by minimizing –2 log L. To study estimation of the parameters for this Weibull distribution, we wrote a computer program to calculate these maximum likelihood estimates. The program first estimated the shape and scale parameters from the marginal distributions. Given these parameter estimates, a univariate search was performed using the Davies, Swann, and Campey algorithm (Box and others 1969) to find δ, which minimized –2 log L. These five parameter values were then used as a starting point for a five-dimensional modified Nelder–Mead search (Evans 1979) to find the five parameter values that minimized –2 log L.

We then performed a simulation experiment to check the accuracy of the maximum likelihood estimation procedure. Data were generated for the bivariate Weibull distribution. Lu and Bhattacharyya (1990), following Lee (1979), showed that (X, Y) can be represented as

$$X = U^{\delta/\beta_1}V^{1/\beta_1}\theta_1 \quad , \qquad Y = (1-U)^{\delta/\beta_2}V^{1/\beta_2}\theta_2$$

where U is distributed as a uniform (0,1) and independently, V is distributed as the mixture of a standard exponential and standard gamma (2). The pdf of V is

$$f_v(v) = \delta v e^{-v} + (1-\delta)e^{-v} \quad , \qquad v > 0$$

The procedure begins by generating five uniform random variables $(U_1, U_2, U_3, U_4, U_5)$. We then take

$$U = U_1$$

and

$$V = \begin{cases} -\ln(U_2)-\ln(U_3), & \text{if } U_5 \le \delta \\ -\ln(U_4), & \text{if } U_5 > \delta \end{cases}$$

We showed that this procedure could be implemented on a computer and samples generated from the bivariate Weibull distribution. We thus have an approach to use in future studies of both estimation procedures and goodness of fit procedures. Moreover, simulation studies are apt to provide the main approach for studying statistical methods in the presence of proof loading.

A simulation experiment to study the maximum likelihood estimators employed

sample sizes of 20, 100

pairs of shape (2.0, 2.0) (2.0, 3.6) (2.0, 6.5)
(3.6, 3.6) (3.6, 6.5) (6.5, 6.5)

$\delta = 0.1, 0.5, 0.9$

and, without loss of generality, scale parameters 7. That is true because neither the sample correlation nor the population correlation depends on the values of the underlying scale parameters of the marginal distributions. Further, the distribution of the estimates for the shape parameters does not depend on the true scale parameters θ_1, θ_2 and neither does the distribution of

$$\left(\hat{\theta}_1/\theta_1,\, \hat{\theta}_2/\theta_2\right)$$

The values selected for the shape parameter of each marginal distribution were meant to span the range of those typically encountered for the strength properties of lumber. See for example Green and Evans (1988a,b,c,d) and Evans and Green (1988a,b,c,d) for fits of two- and three-parameter Weibull distributions to lumber strength properties for several wood species.

A series of 100 trials were conducted for each combination of sample size, pair of shape parameters, and value of δ. The results of this simulation experiment are summarized in the Appendix. The entry Corrf corresponds to the calculation of correlation using the formulas in the Bivariate Weibull Distributions section evaluated at the maximum likelihood estimates, and the entry Corr corresponds to Pearson's correlation coefficient. Generally, all the parameters can be estimated with reasonable accuracy when the sample size $n = 100$. For sample size $n = 20$, the results were far less satisfactory. Our algorithm was based on a simplex method (Evans 1979) and did not contain the restraint that $0 < \delta \leq 1$. For instance, in the case where $n = 20$ and both shape parameters are 6.5, 21 of the 100 trials resulted in estimates of $\delta > 1$ so the estimate Corrf was negative when the true $\delta = 0.9$. This was reduced to 1 in a 100 when $\delta = 0.5$, and there were no negative estimates when $\delta = 0.1$. The situation improved for the larger sample size $n = 100$ where there were only 6 of 100 cases with estimates of $\delta > 1$ when the true $\delta = 0.9$.

Because θ_1 and θ_2 are scale parameters,

$$\left(\hat{\theta}_1/\theta_1,\, \hat{\theta}_2/\theta_2\right)$$

has a distribution that is free of (θ_1, θ_2). For any value of (θ_1, θ_2), $\left(\hat{\theta}_1, \hat{\theta}_2\right)$ is distributed as $\left(7^{-1}\theta_1\hat{\theta}_{1[7]}, 7^{-1}\theta_2\hat{\theta}_{2[7]}\right)$ where $\left(\hat{\theta}_{1[7]}, \hat{\theta}_{2[7]}\right)$ are the maximum likelihood estimates for the case when the true scale parameters are 7.0 as in our simulation study. Therefore, the properties of the maximum likelihood estimators for any other case can be deduced from the case considered in the simulation by multiplying each mean and standard deviation entry by $7^{-1}\theta_i$.

Example Using Real Data

As a final test, we employed our computer program to fit a data set consisting of MOR and MOE for 412 specimens of visually graded Southern Pine 2 by 10 No. 2 boards tested in bending. We first fit a bivariate normal distribution and found that

for MOR, mean = 5.916×10^3 lb/in^2 (40.79 MPa)
standard deviation = 1.857

for MOE, mean = 1.491×10^6 lb/in^2 (10.28 GPa)
standard deviation = 0.379

correlation = 0.71625

Marginal normality was investigated by calculating the Kolmogorov–Smirnov (KS), the Anderson–Darling (AD), and the Shapiro–Wilks (SW) test for each component. The MOE goodness-of-fit values were nonsignificant for all these tests. The P values for MOR are less than 0.01 on the KS test and 0.05 on the AD test, respectively, indicating a poor fit to MOR.

The maximum likelihood estimates of the parameters in the bivariate Weibull distribution are

for MOR, shape = 3.5771, scale = 6.5672

for MOE, shape = 4.3916, scale = 1.6284

$\delta = 0.5137$, Corrf = 0.68818

where Corrf is the estimate of the correlation using the formula for the bivariate Weibull distribution. Marginal Weibull fits were investigated by calculating the KS, AD, and SW tests for each component. For MOE, all these goodness-of-fit values were nonsignificant. For MOR, only the KS test showed significance with $P < 0.01$.

To try to visually show this improved fit, we calculated the difference between the empirical cumulative distribution of the data and the predicted values based on the bivariate normal and Weibull distributions. So for any percentile of the data, we had the actual data point and estimated values based on the two different fits. We then plotted the absolute value of the difference in the actual and predicted values, so smaller values indicate a better fit. Since we are most interested in predicting the lower tail for MOR and mean values

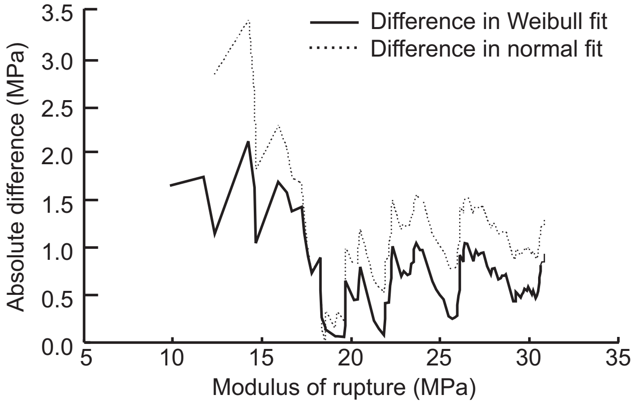


Figure 1—Comparison of bivariate normal and Weibull fits (first 25% only).

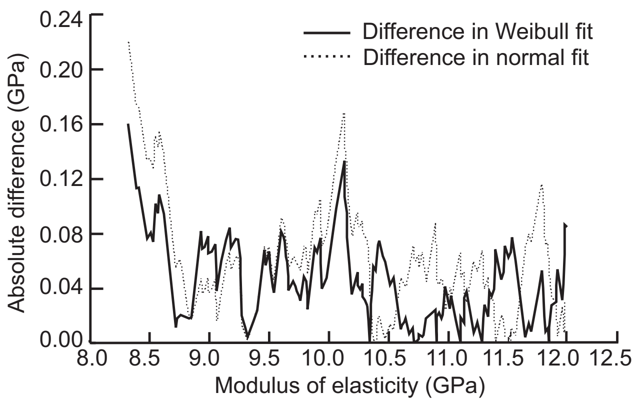


Figure 2—Comparison of bivariate normal and Weibull fits (25th to 75th percentiles).

for MOE, we will confine our plot to those regions. Figure 1 shows the absolute difference in MOR predictions for the lower 25% of the data. Figure 2 shows the absolute difference in MOE predictions for the data between the 25th and 75th percentile. As Figures 1 and 2 show, it is clear that the bivariate Weibull distribution fits these parts of the data better.

Using the Bivariate Weibull Distribution in a Proof Loading Procedure

As discussed in the introduction, it is often not possible to measure two strength properties on the same specimen and we must rely on a proof loading procedure to give us parameter estimates. We shall now look at whether the bivariate Weibull distribution appears amenable to estimation under such a scheme. We begin with the bivariate Weibull distribution. The likelihood is the product of contributions from each specimen. If the specimen fails in mode 1 at $X = x$, the contribution is

$$f_X\left(x \middle| \theta_1, \beta_1\right) = \frac{\beta_1}{\theta_1}\left(\frac{x}{\theta_1}\right)^{\beta_1 - 1} e^{-(x/\theta_1)^{\beta_1}}$$

When the specimen survives the proof load to load L in mode 1 and fails in mode 2 at $Y = y$, the contribution to the likelihood is

$$-\frac{\partial \overline{F}(L,y)}{\partial y} = \overline{F}(L,y)\left\{\left(\frac{L}{\theta_1}\right)^{\beta_1/\delta} + \left(\frac{y}{\theta_2}\right)^{\beta_2/\delta}\right\}^{\delta - 1} \times \left(\frac{\beta_2}{\theta_2}\right)\left(\frac{y}{\theta_2}\right)^{(\beta_2/\delta)-1}$$

That is, if $R = \{i: X_i < L\}$ denotes the index set of the specimens that failed the proof load, then the likelihood is

$$\prod_{i \in R} \frac{\beta_1}{\theta_1}\left(\frac{x_i}{\theta_1}\right)^{\beta_1 - 1} e^{-(x_i/\theta_1)^{\beta_1}}$$
$$\times \prod_{i \notin R} \overline{F}(L, y_i)\left\{\left(\frac{L}{\theta_1}\right)^{\beta_1/\delta} + \left(\frac{y_i}{\theta_2}\right)^{\beta_2/\delta}\right\}^{\delta - 1} \left(\frac{\beta_2}{\theta_2}\right)\left(\frac{y_i}{\theta_2}\right)^{(\beta_2/\delta)-1}$$

Thus, the likelihood for the proof load case can be expressed simply. What remains is to assess our ability to estimate the parameters. It may be that more than one proof load is necessary or that symmetric loading schemes are required to adequately address the estimation problem. The situation where the marginal distributions are known may be realistic in cases where a large amount of data is available on the individual strength properties.

Future Work

It is clear from the skewness exhibited by the distribution of many lumber strength properties that statistical estimation procedures need to be developed for bivariate and even higher dimensional distributions. The current investigation established the feasibility of fitting a five-parameter bivariate Weibull distribution. Many directions remain to be explored if we are to make maximum use of the bivariate Weibull distribution in modeling lumber strength properties:

1. Perform a more extensive study of the maximum likelihood estimator and incorporate a better search procedure that would not have as much difficulty with boundaries for the range of δ.

2. Expand the study to include proof loading as indicated in the section on using a bivariate Weibull distribution in a proof loading procedure. The first step would be to treat the marginal distributions as known. The proof load scheme could vary in complexity. It could be (i) a scheme that proof loads only on one variable, (ii) a symmetric scheme that proof loads part of a sample on one variable and the rest of the sample on the other

variable, or (iii) a scheme that combines proof load data with data on the marginal distributions.

3. Expand the study to the bivariate distribution that has as marginals the three-parameter Weibull distribution.

4. Consider and evaluate other bivariate distributions, such as the inverse Gaussian, the S_{BB}, and the normal–log-normal distributions.

5. Develop confidence statements concerning percentiles of one variable, say bending strength, given that the specimen was proof loaded (or tested) in another variable, say tension. This should be done after acceptable models are obtained and goodness of fit is investigated.

References

Box, M.J.; Davies, D.; Swann, W.H. 1969. Non-linear optimization techniques. Edinburgh, Scotland: Oliver & Boyd.

Conway, D. 1983. Farlie–Gumbel–Morgenstern distributions. Encyclopedia of Statistical Sciences. New York, NY: John Wiley & Sons. 3: 28–31.

de Amorim, S.; Johnson, R.A. 1986. Experimental designs for estimating the correlation between two destructively tested variables. Journal of the American Statistical Association. 81 (395): 807–812.

Draper, J. 1952. Properties of distributions resulting from simple transformations of the normal distributions. Biometrika. 39: 290–301.

Evans, J.W. 1979. Computer augmentation of experimental designs to maximize |X'X|. Technometrics. 21: 321–330.

Evans, J.W.; Green, D.W. 1988a. Mechanical properties of visually graded dimension lumber: Vol. 2. Douglas Fir–Larch. Pub. No. PB–88–159–397. Springfield, VA: National Technical Information Service. 476 p.

Evans, J.W.; Green, D.W. 1988b. Mechanical properties of visually graded dimension lumber: Vol. 3 Hem-fir. Pub. No. PB–88–159–405. Springfield, VA: National Technical Information Service. 392 p.

Evans, J.W.; Green, D.W. 1988c. Mechanical properties of visually graded dimension lumber: Vol. 5. Aspen–Cottonwood, Balsam Fir, Douglas Fir (South). Pub. No. PB–88–159–421. Springfield, VA: National Technical Information Service. 510 p.

Evans, J.W.; Green, D.W. 1988d. Mechanical properties of visually graded dimension lumber: Vol. 8. Sitka Spruce, Subalpine Fir, Sugar Pine, Tamarack, Yellow–Poplar. Pub. No. PB–88–159–454. Springfield, VA: National Technical Information Service. 492 p.

Evans, J.W., Johnson, R.A. and Green, D.W. 1984. Estimating the correlation between variables under destructive testing, or how to break the same board twice. Technometrics. 26(3): 285–290.

Freund, J.E. 1961. A bivariate extension of the exponential distribution. Journal of the American Statistical Association. 56: 971–977.

Genest, C.; MacKay, R. 1986. Copules Archimediennes et Families de Lois Bidimensionnelles Dont les Marges Sont Donnees. The Canadian Journal of Statistics. 14: 145–159.

Green, D.W.; Evans, J.W. 1983. Estimating correlation between strength properties, recent advances in engineering mechanics and their impact in civil engineering practice. In: Chen, W.F.; Lewis, A.D.M. eds. Proceedings, fourth engineering mechanics division specialty conference; 1983 May 23–25; Purdue University, West Lafayette, IN. Volume II. New York, NY: American Society of Civil Engineers. 6: 936–939.

Green, D.W.; Evans, J.W. 1988a. Mechanical properties of visually graded dimension lumber: Vol. 1 A summary. Pub. No. PB–88–159–389. Springfield, VA: National Technical Information Service. 131 p.

Green, D.W.; Evans, J.W. 1988b. Mechanical properties of visually graded dimension lumber: Vol. 4. Southern Pine. Pub. No. PB–88–159–413. Springfield, VA. National Technical Information Service. 412 p.

Green, D.W.; Evans, J.W. 1988c. Mechanical properties of visually graded dimension lumber: Vol. 6 Eastern Hemlock, Eastern Spruces, Eastern White Pine, Engelmann Spruce, Idaho White Pine, Jack Pine. Publication No. PB–88–159–439. Springfield, VA: National Technical Information Service. 560 p.

Green, D.W.; Evans, J.W. 1988d. Mechanical properties of visually graded dimension lumber: Vol. 7 (Minor) Southern Pines, Ponderosa Pine, Red Pine. Publication No. PB–88–159–447. Springfield, VA: National Technical Information Service. 542 p.

Green, D.W.; Evans, J.W.; Johnson, R.A. 1984. Investigation of the procedure for estimating concomitance of lumber strength properties. Wood and Fiber Science. 16(3): 427–440.

Gumbel, E.J. 1960. Bivariate exponential distribution. Journal of the American Statistical Association. 55: 698–707.

Hougaard, P. 1986. A class of multivariate failure time distributions. Biometrika. 73: 671–678.

Huang, J.S.; Kotz, S. 1984. Correlation structure in iterated Farlie–Gumbel–Morgenstern distributions. Biometrika. 71: 633–636.

Johnson, N.L. 1949a. Bivariate distributions based on simple translation systems. Biometrika. 36: 297–304.

Johnson, N.L. 1949b. Systems of frequency curves generated by methods of translation. Biometrika. 36: 149–176.

Johnson, N.L.; Kotz, S. 1972. Distributions in statistics: continuous multivariate distributions. New York, NY: John Wiley & Sons.

Kocherlakota, S. 1986. The bivariate inverse Gaussian distribution: An introduction. Communications in Statistics Theory and Methods. 15(4): 1081–1112.

Lawless, J.F. 1982. Statistical models and methods for lifetime data. New York, NY: John Wiley & Sons.

Lee, L. 1979. Multivariate distributions having Weibull properties. Journal of Multivariate Analysis. 9: 267–277.

Lu, J.C.; Bhattacharyya, G.K. 1990. Some new constructions of bivariate Weibull models. Annals of Institute of Statistical Mathematics. 42(3): 543–559.

Marshall, A.; Olkin, I. 1967. A multivariate exponential distribution. Journal of the American Statistical Association. 62: 30–44.

Marshall, A.; Olkin, I. 1988. Families of multivariate distributions. Journal of the American Statistical Association. 83: 834–841.

Pearson, R.G. 1980. Potential of the S_B and S_{BB} distributions for describing mechanical properties of lumber. Wood and Fiber. 12(4): 244–254.

Schucany, W.; Parr W.; Boyer J. 1978. Correlation structure in Farlie–Gumbel–Morgenstern distributions. Biometrika. 65: 650–653.

Warren, W.G. 1979. Some recent developments relating to statistical distributions in forestry and forest products research. In:. Ord, J.K.; Patil, G.P.; Tallie, C. eds. Statistical distributions in ecological work. Fairland, MD: International Co-operative Publishing House: 247–260.

Appendix—Summary of Simulation Results

(100 trials per simulation)

	Shape 1	Scale 1	Shape 2	Scale 2	Delta	Corrf	Corr
			Sample size = 100				
True value	2.0	7.0	2.0	7.0	0.9	0.1406	
Mean	2.022	6.969	2.040	6.990	0.897	0.144	0.141
SD	0.159	0.339	0.170	0.402	0.064	0.089	0.098
True value	2.0	7.0	2.0	7.0	0.5	0.6600	
Mean	2.022	6.978	2.022	6.913	0.496	0.665	0.665
SD	0.147	0.364	0.159	0.341	0.039	0.042	0.060
True value	2.0	7.0	2.0	7.0	0.1	0.9822	
Mean	2.047	7.044	2.045	7.047	0.103	0.981	0.982
SD	0.152	0.394	0.149	0.393	0.011	0.003	0.004
True value	2.0	7.0	3.6	7.0	0.9	0.1493	
Mean	2.042	6.987	3.626	6.974	0.896	0.153	0.157
SD	0.182	0.350	0.304	0.203	0.070	0.102	0.106
True value	2.0	7.0	3.6	7.0	0.5	0.6726	
Mean	2.026	6.959	3.643	6.986	0.507	0.664	0.660
SD	0.148	0.368	0.274	0.201	0.048	0.050	0.055
True value	2.0	7.0	3.6	7.0	0.1	0.9712	
Mean	2.040	6.950	3.665	6.974	0.101	0.971	0.973
SD	0.176	0.352	0.318	0.198	0.013	0.003	0.004
True value	2.0	7.0	6.5	7.0	0.9	0.1531	
Mean	2.017	6.964	6.554	6.992	0.893	0.162	0.158
SD	0.158	0.355	0.453	0.133	0.062	0.093	0.108
True value	2.0	7.0	6.5	7.0	0.5	0.6709	
Mean	2.013	6.940	6.469	6.968	0.497	0.671	0.674
SD	0.156	0.371	0.461	0.111	0.056	0.057	0.066
True value	2.0	7.0	6.5	7.0	0.1	0.9506	
Mean	2.050	7.025	6.657	7.005	0.101	0.951	0.953
SD	0.152	0.402	0.485	0.125	0.011	0.003	0.008
True value	3.6	7.0	3.6	7.0	0.9	0.1606	
Mean	3.669	6.989	3.611	7.014	0.914	0.137	0.143
SD	0.264	0.234	0.265	0.182	0.059	0.093	0.094
True value	3.6	7.0	3.6	7.0	0.5	0.6998	
Mean	3.684	6.997	3.648	7.005	0.505	0.693	0.695
SD	0.270	0.205	0.284	0.198	0.045	0.047	0.055
True value	3.6	7.0	3.6	7.0	0.1	0.9860	
Mean	3.666	6.951	3.659	6.956	0.101	0.986	0.986
SD	0.276	0.202	0.271	0.206	0.012	0.003	0.004
True value	6.5	7.0	3.6	7.0	0.9	0.1661	
Mean	6.589	6.991	3.699	6.979	0.906	0.154	0.165
SD	0.503	0.106	0.314	0.226	0.062	0.099	0.099
True value	6.5	7.0	3.6	7.0	0.5	0.7076	
Mean	6.729	6.998	3.667	7.016	0.502	0.704	0.710
SD	0.560	0.112	0.322	0.202	0.050	0.052	0.054
True value	6.5	7.0	3.6	7.0	0.1	0.9815	
Mean	6.577	7.003	3.646	7.007	0.100	0.982	0.982
SD	0.536	0.125	0.296	0.224	0.012	0.002	0.004
True value	6.5	7.0	6.5	7.0	0.9	0.1729	
Mean	6.562	7.014	6.560	6.978	0.897	0.175	0.181
SD	0.513	0.113	0.528	0.113	0.070	0.117	0.116
True value	6.5	7.0	6.5	7.0	0.5	0.7221	
Mean	6.606	7.004	6.571	6.994	0.498	0.722	0.721
SD	0.479	0.121	0.516	0.115	0.047	0.048	0.057
True value	6.5	7.0	6.5	7.0	0.1	0.9879	
Mean	6.623	7.009	6.621	7.010	0.102	0.987	0.987
SD	0.522	0.111	0.528	0.111	0.012	0.003	0.003

	Shape 1	Scale 1	Shape 2	Scale 2	Delta	Corrf	Corr
			Sample size = 20				
True value	2.0	7.0	2.0	7.0	0.9	0.1406	
Mean	2.084	6.821	2.105	6.833	0.893	0.146	0.136
SD	0.340	0.808	0.406	0.892	0.132	0.182	0.209
True value	2.0	7.0	2.0	7.0	0.5	0.6600	
Mean	2.084	7.070	2.085	7.050	0.501	0.652	0.661
SD	0.372	0.745	0.361	0.840	0.124	0.149	0.160
True value	2.0	7.0	2.0	7.0	0.1	0.9822	
Mean	2.180	7.062	2.178	7.060	0.101	0.982	0.982
SD	0.447	0.816	0.440	0.831	0.026	0.008	0.011
True value	2.0	7.0	3.6	7.0	0.9	0.1493	
Mean	2.088	6.829	3.783	6.894	0.896	0.148	0.149
SD	0.340	0.804	0.724	0.491	0.142	0.209	0.204
True value	2.0	7.0	3.6	7.0	0.5	0.6726	
Mean	2.087	7.071	3.758	7.017	0.502	0.662	0.672
SD	0.372	0.741	0.646	0.463	0.124	0.149	0.153
True value	2.0	7.0	3.6	7.0	0.1	0.9712	
Mean	2.182	7.063	3.924	7.021	0.101	0.971	0.974
SD	0.450	0.826	0.798	0.468	0.026	0.007	0.010
True value	2.0	7.0	6.5	7.0	0.9	0.1531	
Mean	2.107	6.843	6.770	6.937	0.893	0.155	0.153
SD	0.402	0.893	1.086	0.251	0.133	0.194	0.197
True value	2.0	7.0	6.5	7.0	0.5	0.6709	
Mean	2.090	7.054	6.784	7.014	0.501	0.661	0.672
SD	0.362	0.846	1.196	0.231	0.119	0.138	0.149
True value	2.0	7.0	6.5	7.0	0.1	0.9506	
Mean	2.179	7.055	7.088	7.007	0.101	0.952	0.956
SD	0.433	0.834	1.434	0.254	0.026	0.007	0.017
True value	3.6	7.0	3.6	7.0	0.9	0.1606	
Mean	3.761	6.894	3.788	6.893	0.895	0.159	0.162
SD	0.615	0.453	0.722	0.494	0.134	0.208	0.201
True value	3.6	7.0	3.6	7.0	0.5	0.6998	
Mean	3.754	7.031	3.754	7.017	0.501	0.690	0.693
SD	0.676	0.419	0.644	0.466	0.118	0.139	0.151
True value	3.6	7.0	3.6	7.0	0.1	0.9860	
Mean	3.923	7.019	3.921	7.017	0.102	0.985	0.985
SD	0.805	0.461	0.795	0.468	0.026	0.007	0.009
True value	6.5	7.0	3.6	7.0	0.9	0.1661	
Mean	6.779	6.938	3.787	6.893	0.897	0.159	0.168
SD	1.084	0.251	0.727	0.494	0.142	0.229	0.200
True value	6.5	7.0	3.6	7.0	0.5	0.7076	
Mean	6.781	7.015	3.764	7.017	0.502	0.694	0.699
SD	1.203	0.223	0.660	0.465	0.125	0.156	0.150
True value	6.5	7.0	3.6	7.0	0.1	0.9815	
Mean	7.093	7.009	3.924	7.021	0.101	0.981	0.981
SD	1.452	0.255	0.791	0.465	0.026	0.006	0.009
True value	6.5	7.0	6.5	7.0	0.9	0.1729	
Mean	6.775	6.936	6.838	6.937	0.897	0.165	0.176
SD	1.103	0.252	1.304	0.273	0.136	0.224	0.202
True value	6.5	7.0	6.5	7.0	0.5	0.7221	
Mean	6.773	7.012	6.784	7.005	0.503	0.706	0.708
SD	1.215	0.231	1.186	0.256	0.125	0.157	0.151
True value	6.5	7.0	6.5	7.0	0.1	0.9879	
Mean	7.090	7.008	7.083	7.007	0.102	0.987	0.987
SD	1.443	0.257	1.417	0.259	0.026	0.007	0.008