Applied Mathematical Sciences, Vol. 15, 2021, no. 12, 605 - 614
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2021.914598

A Method of Fitting Distributions to Bivariate Non-negative Data

A. H. Pooi

Centre for Mathematical Sciences
Universiti Tunku Abdul Rahman, Malaysia

Abstract

Given a set of observations from the pair of variables y_1 and y_2 which are continuous and non- negative, an orthogonal transformation is applied to the deviations of y_1 and y_2 from their respective estimated means to form the transformed variables s_1 and s_2 . Each s_i is then a linear function of y_1 and y_2. Let s_1 be the transformed variable of which the coefficients are of different signs. The marginal distribution of the variable s_1 is estimated by a power-normal distribution of which the first four moments are the same as those based on the observed s_1. Given a value of s_1, the conditional distribution of the variable s_2 is estimated by a distribution derived from a quadratic function of two power-normal random variables. The derived conditional distribution is chosen to have about the same first four moments as those computed from some selected observed values of s_2. From the product of the above marginal and conditional distributions, we can obtain the joint probability density function of y_1 and y_2. The proposed fitting procedure is applied to a set of 100 values of (y_1 , y_2) generated from a bivariate Weibull distribution. A comparison of the fitted and actual marginal and conditional distributions indicates that the proposed procedure is satisfactory.

Mathematics Subject Classification: 62F10, 62H12

Keywords: bivariate non-negative data, method of moments, power-normal distribution, quadratic function of power-normal random variables

1. Introduction

Bivariate data have a wide application in the situation when we are interested in the relationship between two variables, denoted as y_1 and y_2 , or the probability of certain events which are expressed in terms of y_1 and y_2 .The following are some examples.

(1) When y_1 is an explanatory variable and y_2 the response variable, we might be interested in the distribution of y_2 when the value of y_1 is given.

(2) When y_1 is the strength of a system which is subjected to stress y_2, the probability $P(y_1 - y_2 > 0)$ measures the system performance.

(3) In the situation in which y_i denotes the lifetime of the i-th component in a series (or parallel) system of two correlated components, the probability that the time to failure of the system is bigger than a chosen value T is given by $P(y_1 > T \text{ and } y_2 > T)$ $(\text{or } P(y_1 > T \text{ or } y_2 > T))$.

(4) When y_1 and y_2 are respectively the future lifetimes of the husband and wife, the insurance companies need to know the distribution of the joint life status given by the minimum of y_1 and y_2 in determining the premium of a fixed term joint life assurance policy.

A direct way of investigating the relationship between two variables y_1 and y_2 ,and the probabilities of the events of interest, is one which makes use of the bivariate probability density function (pdf) of the two variables. There are various methods of constructing bivariate pdf . In what follows some of the important methods are outlined.

One of the methods is based on the copula (see [1]) which is a bivariate distribution function $C(u_1, u_2)$ with uniform marginal over the unit square in the $u_1 - u_2$ plane. Let $F_i(y_i)$ be the marginal distribution function of y_i , $i = 1,2$. Then $F(y_1, y_2)$=$C[F_1(y_1), F_2(y_2)]$ is a bivariate distribution of y_1 and y_2.

Another method is one which is based on the standard bivariate normal variables (u_1, u_2) with correlation ρ (see [3,4]). A bivariate distribution for (y_1, y_2) can be obtained by transforming to u_i to y_i via

$$u_i=\gamma_i+\delta_i h_i((y_i-\zeta_i)/\lambda_i),\ i = 1,2,$$

where $\gamma_i, \delta_{i,}\zeta_i$ *and* λ_i are parameters, and $h_i(v)$ is a selected one-to-one function of v.

In [5], Marshall and Olkin obtained bivariate distribution H for y_1 and y_2 by forming integral using the mixture distribution $K(\theta_1, \theta_2)$ for non-negative values of the parameters θ_1 and θ_2 , together with the distribution functions $F_1(y_1)$ and $F_2(y_2)$:

$$\text{H}(y_1,y_2) = \int_0^\infty \int_0^\infty F_1^{\theta_1}(y_1) F_2^{\theta_2}(y_2) dK(\theta_1, \theta_2).$$

When a set of bivariate non-negative data is given, we may need to go through the process of finding the distribution functions $F_1(y_1)$ and $F_2(y_2)$ and a suitable copula, (or mixture distribution) in order to find a joint pdf to fit the data. Such a process in practice could be quite challenging.

In this paper, a straightforward method of fitting distributions to the observed bivariate non-negative data is proposed. In the method, deviations of the variables y_1 and y_2 from their respective observed means are transformed by using an orthogonal matrix to s_1 and s_2. Suppose the range of s_1 is $(-\infty, \infty)$ and that of s_2 is $[s_2^*(s_1), \infty)$ where $s_2^*(s_1)$ is dependent on s_1. The first four moments of s_1 are calculated using the observed values of s_1. The method of moments is then used to find a power-normal distribution (see [6]) to estimate the marginal distribution of s_1. When s_1 takes on a particular value, say s_1^*, the observed values of s_2 of which (s_1, s_2) lies in a selected region in the $s_1 - s_2$ plane are used to calculate the first four moments of s_2 given that $s_1 = s_1^*$ (denoted as $s_2|s_1 = s_1^*$). From the distribution derived from a certain quadratic function of two power-normal random variables, we use the method of moments to find a conditional distribution for $s_2|s_1 = s_1^*$. From the marginal distribution of s_1 and the conditional distribution for $s_2|s_1 = s_1^*$, we find the joint pdf for the point (s_1^*, s_2).

This paper has five sections. Section 2 describes the computation of the pdf of the variable formed by a certain quadratic function of two power-normal random variables. Section 3 deals with the estimation of the marginal distribution of s_1 and the conditional distribution of $s_2|s_1 = s_1^*$. In Section 4, the method in Section 3 is used to fit distributions to the data generated by using the bivariate Weibull distribution. Section 5 concludes the paper.

2. Pdf of Quadratic Function of Two Power-normal Variables

In [6], the authors introduced the following power transformation of the standard normal random variable z:

$$\tilde{\varepsilon} = \psi(\lambda^+, \lambda^-, z) = \begin{cases} \left[(z+1)^{\lambda^+} - 1\right]/\lambda^+, & (z \geq 0, \lambda^+ \neq 0) \\ \log(z+1), & (z \geq 0, \lambda^+ = 0) \\ -\left[(-z+1)^{\lambda^-} - 1\right]/\lambda^-, & (z < 0, \lambda^- \neq 0) \\ -\log(-z+1), & (z < 0, \lambda^- = 0) \end{cases}$$

where λ^+ and λ^- are constants. The distribution of the resulting random variable $\tilde{\varepsilon}$ is referred to as the power-normal distribution with parameters λ^+ and λ^-.

Let $w = dv + e$

where d and e are constants, $v = v_1 + v_2$, $v_i = \varepsilon_i^{+2}$, $\varepsilon_i^+ = c_i + \varepsilon_i$,

c_1 and c_2 are constants,

$\varepsilon_i = [\tilde{\varepsilon}_i - E(\tilde{\varepsilon}_i)]/[var(\tilde{\varepsilon}_i)]^{1/2}$,

$\tilde{\varepsilon}_i$ has a power-normal distribution with parameters λ_i^+ and λ_i^- , and $\tilde{\varepsilon}_1$, $\tilde{\varepsilon}_2$ are independent. The above definition of w implies that the pdf of w is zero for $w < e$ and takes on a non-negative value if $w \geq e$.

To find the pdf of v, we first find the pdf of v_i from the pdf f_{ε_i} of ε_i using

$$f_{v_i}(v_i) = \frac{f_{\varepsilon_i}(-c_i - \sqrt{v_i}) + f_{\varepsilon_i}(-c_i + \sqrt{v_i})}{2\sqrt{v_i}}.$$

The pdf of v evaluated at a given value, say v^*, may be expressed as

$$f_v(v^*) = \int_0^{v^*} f_{v_1}(t)\, f_{v_2}(v^* - t)dt.$$

When t is near zero, $f_{v_i}(t)$ may be approximated by the following power function:

$$f_{v_i}(t) \cong \alpha_i t^{-\beta_i}, i = 1,2$$

where α_i and β_i are constants. A way to estimate the values of α_i and β_i is by solving the equations

$$f_{v_i}(a) = \alpha_i a^{-\beta_i}$$

and

$$f_{v_i}(2a) = \alpha_i (2a)^{-\beta_i}$$

for a small positive value of a. Let m be a large positive integer. Suppose $h > 0$ is such that $mh = v^*$. Then the interval $(0, v^*)$ may be partitioned into m intervals each of length h, and $f_v(v^*)$ may be approximated by

$$f_v(v^*) = \alpha_1 \frac{h^{1-\beta_1}}{1-\beta_1} f_{v_2}\left(mh - \frac{h}{2}\right) + \sum_{j=1}^{m-2} f_{v_1}(jh) f_{v_2}(v^* - jh)h + \alpha_2 \frac{h^{1-\beta_2}}{1-\beta_2} f_{v_1}\left(mh - \frac{h}{2}\right).$$

3. Estimation of Marginal and Conditional Distributions

Suppose the non-negative random variables y_1 and y_2 have respectively the means μ_1 and μ_2, and the variance-covariance matrix

$$\boldsymbol{M} = \left\{E\{[y_i - E(y_i)][y_j - E(y_j)]\}\right\}.$$

Let $\boldsymbol{y}_k = (y_{1k}, y_{2k})^T, k = 1, \ldots, n,$ be a set of bivariate observations from the variables y_1 and y_2. The mean μ_i of y_i may be estimated by

$$\hat{\mu}_i = \bar{y}_i = \frac{1}{n}\sum_{k=1}^{n} y_{ik}, i = 1,2$$

while the variance-covariance matrix M may be estimated by using

$$\widehat{M}_{ij} = \frac{1}{n}\sum_{k=1}^{n} [y_{ik} - \bar{y}_i]\,[y_{jk} - \bar{y}_j]$$

Let $\widehat{\boldsymbol{H}}$ be the 2× 2 orthogonal matrix formed by the two 2×1 eigenvectors of $\widehat{\boldsymbol{M}}$. The vector formed by the deviation of $\boldsymbol{y}_k$ from $\widehat{\boldsymbol{\mu}}$ may be rotated by $\widehat{\boldsymbol{H}}^T$ to the rotated variable $\boldsymbol{s_k} = [s_{1k} \quad , \quad s_{2k}]^T$:

$$\boldsymbol{s_k} = \widehat{\boldsymbol{H}}^T[\boldsymbol{y}_k - \widehat{\boldsymbol{\mu}}].$$

We may consider $\boldsymbol{s_k}$ as the k-th observed value of the rotated variable $\boldsymbol{s}$. The rows in the matrix $\widehat{\boldsymbol{H}}^T$ are typically such that the entries in one of the rows have opposite signs while those of the other row have the same sign. Thus the range of one of the component, say s_1, of $\boldsymbol{s}$ will typically be (-∞, ∞) while that of the other component s_2 is $[s_2^*(s_1), \infty)$ where $s_2^*(s_1)$ is dependent on s_1.

To find an approximate distribution for s_1 ,we may first use the n observed values of s_1 to compute the observed mean $\hat{\mu}_{s1} = m_{s1}^{(1)}$ and the second to fourth central moments $m_{s1}^{(l)}, l = 2,3,4$ of s_1:

$$m_{s1}^{(1)} = \frac{1}{n}\sum_{k=1}^{n} s_{1k}$$

$$m_{s1}^{(l)} = \frac{1}{n}\sum_{k=1}^{n}\left(s_{1k} - m_{s1}^{(1)}\right)^{l}, \; l = 2,3,4.$$

From the computed central moments, we find the following estimated standard deviation, coefficient of skewness and coefficient of kurtosis of s_1:

$$\hat{\sigma}_{s1} = \left[m_{s1}^{(2)}\right]^{1/2}, \widehat{SK}_{s1} = \frac{m_{s1}^{(3)}}{\hat{\sigma}_{s1}{}^{3}}, and\ \hat{K}_{s1} = \frac{m_{s1}^{(4)}}{\hat{\sigma}_{s1}{}^{4}}.$$

We next find a power-normal distribution of which the corresponding coefficients of skewness and kurtosis are the same as those of s_1. Let the resulting parameters of this power-normal distribution be λ_1^+ and λ_1^- . The random variable s_1 may then be equated to

$$s_1 = \hat{\mu}_{s1} + \hat{\sigma}_{s1}\,[\tilde{\varepsilon}_1 - E(\tilde{\varepsilon}_1)]/[var(\tilde{\varepsilon}_1)]^{1/2} \quad (1)$$

where $\tilde{\varepsilon}_1$has a power-normal distribution with parameters λ_1^+ and λ_1^-. Equation (1) then provides the estimated marginal distribution of s_1.

When the value of $s_1$is fixed at a certain value, say s_1^*, the value of s_2 will lie in $[s_2^*(s_1^*), \infty)$ where $s_2^*(s_1^*)$ is the value of s_2 for the point of intersection of the line $y_1 = 0$ (or $y_2 = 0$, depending on the value of s_1^*) and the line $s_1 = s_1^*$. In what follows, we attempt to derive an approximate distribution for $s_2|s_1 = s_1^*$.

Let $\bar{s}_1^*$ be the value of $s_1$for the point of intersection of the line $y_2 = 0$ (or $y_1 = 0$) and the line $s_2 = s_2^*(s_1^*)$. We next let

$P(s_1^*) = \{(y_1, y_2): s_2 \geq s_2^*(s_1^*), s_1 \text{ is in between } \bar{s}_1^* \text{ and } s_1^*\}$,

and denote the number of observed values of (y_1, y_2) in $P(s_1^*)$ as $N(s_1^*)$.

It is possible that the distribution of s_2 in $[s_2^*(s_1^*), \infty)$ is the same for any given value of s_1 in between $\bar{s}_1^*$ and s_1^* . To check the equality of distributions of s_2 in $[s_2^*(s_1^*), \infty)$, we may first

1. regress the values of s_2 on s_1 for all the values of (s_1, s_2) derived from the values of (y_1, y_2) in $P(s_1^*)$ to obtain the line of fit
$$s_2 = \hat{\beta}_0 + \hat{\beta}_1^{(1)} s_1,$$
2. regress the values of $\left[s_2 - \hat{\beta}_0 - \hat{\beta}_1^{(1)} s_1\right]^{l}$ on s_1 for all the values of (s_1, s_2) derived from the values of (y_1, y_2) in $P(s_1^*)$ to obtain the slope $\hat{\beta}_1^{(l)}$ of the regression line, $l = 2,3,4$

In the case when we can establish that $\beta_1^{(l)}$ for $l = 1,..,4$ are all insignificant when the relevant t-tests are applied, then we get an indication that the distribution of s_2 in $[s_2^*(s_1^*), \infty)$ is the same for any given value of s_1 in between $\bar{s}_1^*$ and s_1^* . We may then use the observed values of s_2 in $P(s_1^*)$ to compute the first four moments of s_2. A quadratic function w of two power-normal random variables introduced in Section 2 may next be found such that the computed first two moments of s_2 are equal to those of the quadratic function w, while the computed third and fourth moments of s_2 are as close as possible to those of w. The distribution of w may now be used to approximate that of $s_2|s_1 = s_1^*$.

After transforming $(y_1, y_2)^T$ to $(s_1, s_2)^T$, the joint pdf of $(y_1, y_2)^T$ may be computed as the product of the marginal pdf of s_1 and the conditional pdf of s_2 given the value of s_1.

4. A Numerical Example

Consider the two-parameter Weibull distribution (see [2]) which has the following survival function:

$$\bar{F}(y_1, y_2) = exp\left\{-\left[\left(\frac{y_1}{\lambda_1}\right)^{\frac{\gamma_1}{\alpha}} + \left(\frac{y_2}{\lambda_2}\right)^{\frac{\gamma_2}{\alpha}}\right]^{\alpha}\right\} \quad (2)$$

where $0 \le \alpha \le 1;\ 0 < \lambda_1, \lambda_2 < \infty;\ 0 < \gamma_1, \gamma_2 < \infty.$

A total of $n = 100$ observations of (y_1, y_2) are generated from the above bivariate Weibull distribution with parameters $\lambda_1 = \lambda_2 = 1, \gamma_1 = \gamma_2 = 1.5\ and\ \alpha = 0.8$. The corresponding values of $[\bar{y}_1, \bar{y}_2]$ and $\widehat{\boldsymbol{H}}$ are

$$[\bar{y}_1, \bar{y}_2] = [0.9321, 0.9258],$$

$$\text{and } \widehat{\boldsymbol{H}} = \begin{bmatrix} 0.5593 & 0.8289 \\ -0.8289 & 0.5593 \end{bmatrix}$$

Figure 1 exhibits the marginal pdf of s_1 estimated from the 100 observed values of s_1, and that based on the theoretical Weibull pdf obtained by differentiating partially the survival function given in (2). The figure shows that the estimated pdf bears a fairly striking resemblance to the theoretical pdf.

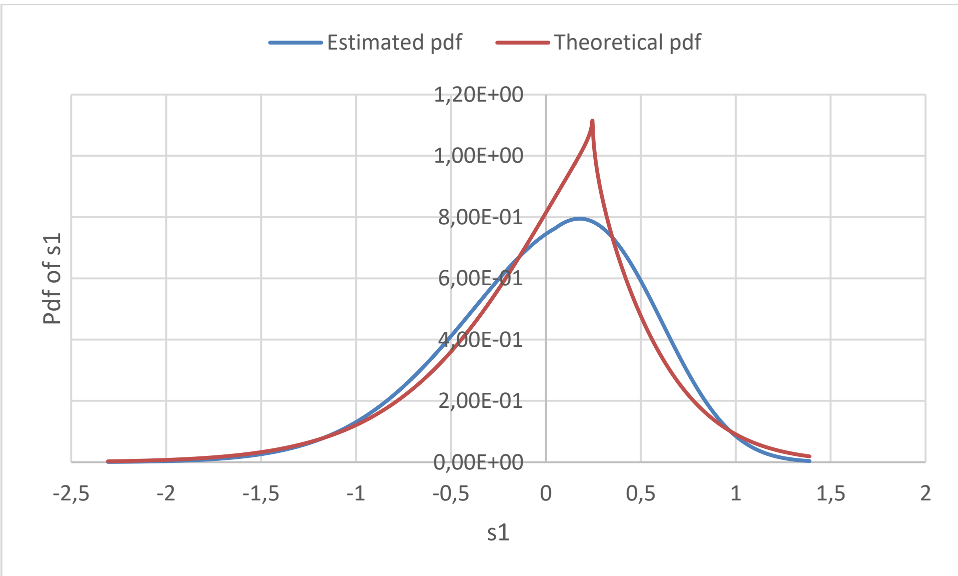

Figure 1. Estimated and theoretical marginal pdf of s_1 in the bivariate Weibull model with parameters $\lambda_1 = \lambda_2 = 1, \gamma_1 = \gamma_2 = 1.5$ and $\alpha = 0.8$. $(n = 100)$

Before estimating the moments of $s_2|s_1 = s_1^*$,we may investigate whether the distribution of s_2 in $[s_2^*(s_1^*), \infty)$ is the same for any given value of s_1 in between $\bar{s}_1^*$ and s_1^* . For example, when $s_1^* = -0.1875,\ N(s_1^*)$ is found to be 40. If we regress the values of s_2 on s_1 for all the 40 values of (s_1, s_2) derived from the values of (y_1, y_2) in (s_1^*) , the line of fit is found to be $s_2 = 0.6089\ s_1 - 0.1680$. The absolute values of the t-statistics are respectively found to be
$|T^{(1)}|$ =0.9507, $|T^{(2)}| = 1.8908$, $|T^{(3)}| = 1.1691$, and $|T^{(4)}| = 1.6909$.
These absolute values show that the slope parameters $\beta_1^{(l)}, l = 1,2,3,4$ are all insignificantly different from zero at the 5% level.

Table 1 shows that when the number $N(s_1^*)$ of points in $P(s_1^*)$ is fairly large, the estimated mean $\hat{\mu}_{s2}$, standard deviation $\hat{\sigma}_{s2}$, coefficient of skewness $\widehat{SK}_{s2}$ and coefficient of kurtosis $\widehat{K}_{s2}$ of $s_2|s_1 = s_1^*$ are quite comparable to the corresponding theoretical values.

Table 1. Estimated and theoretical values of mean, standard deviation, coefficient of skewness, and coefficient of kurtosis of $\mathrm{s}_2|\mathrm{s}_1 = \mathrm{s}_1^*$.

No	s_1^*	N(s_1^*)	$s_2^*(s_1^*)$	Estimated Values				Theoretical Values			
				$\hat{\mu}_{s2}$	$\hat{\sigma}_{s2}$	$\widehat{SK}_{s2}$	$\widehat{K}_{s2}$	μ_{s2}	σ_{s2}	SK_{s2}	K_{s2}
1	-1.5	52	-0.11	0.46	0.48	1.15	3.89	0.73	0.53	0.91	3.60
2	-1.3125	59	-0.24	0.36	0.49	1.22	4.19	0.62	0.54	0.93	3.70
3	-1.125	62	-0.37	0.27	0.52	1.22	4.16	0.50	0.55	0.95	3.79
4	-0.9375	65	-0.49	0.21	0.54	1.13	3.92	0.39	0.55	0.97	3.87
5	-0.75	67	-0.62	0.11	0.56	1.18	4.33	0.27	0.56	0.98	3.94
6	-0.5625	68	-0.74	0.01	0.55	1.23	4.82	0.15	0.57	1.00	4.01
7	-0.375	61	-0.87	-0.03	0.61	1.04	4.03	0.03	0.58	1.02	4.08
8	-0.1875	40	-1.00	-0.10	0.66	1.22	4.52	-0.10	0.59	1.03	4.12
9	0	23	-1.12	-0.09	0.73	1.35	4.70	-0.23	0.60	1.05	4.14
10	0.1875	3	-1.25	-0.33	0.41	0.48	1.50	-0.40	0.64	1.04	4.01
11	0.375	30	-1.10	-0.10	0.69	1.27	4.72	-0.18	0.61	1.01	4.01
12	0.5625	65	-0.82	-0.03	0.58	1.10	4.29	0.08	0.58	1.00	4.01
13	0.75	68	-0.54	0.16	0.55	1.16	4.17	0.32	0.56	1.00	3.95
14	0.9375	62	-0.27	0.32	0.50	1.19	4.11	0.56	0.53	0.98	3.84
15	1.125	44	0.01	0.55	0.47	1.12	3.80	0.81	0.51	0.94	3.68
16	1.3125	25	0.29	0.85	0.41	1.09	3.74	1.05	0.49	0.89	3.46
17	1.5	17	0.57	1.05	0.35	1.33	3.90	1.29	0.46	0.82	3.19
18	1.6875	11	0.85	1.21	0.34	1.14	2.91	1.53	0.42	0.72	2.90
19	1.875	4	1.12	1.59	0.30	-0.22	1.89	1.76	0.38	0.60	2.60
20	2.0625	3	1.40	1.73	0.19	0.37	1.50	1.99	0.34	0.45	2.33

When $N(s_1^*)$ is very small, the estimated moments may no longer be reliable. However, it is possible to use the results of the neighboring rows in the table to obtain estimates of the moments by means of interpolation or extrapolation.

Figures 2 and 3 show the estimated and theoretical marginal pdf of $s_2|s_1 = s_1^*$ when s_1^*=-1.5 and -0.1875 respectively. The figures reveal that the estimated pdf bear some resemblance to the theoretical pdf.

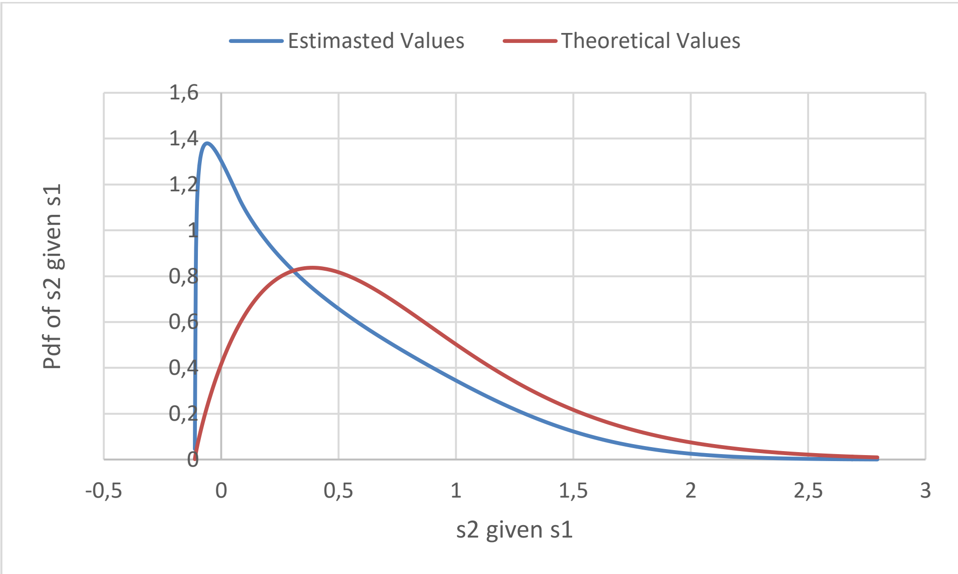

Figure 2. Estimated and theoretical pdf of $s_2|s_1 = s_1^*$ in the bivariate Weibull model with parameters $\lambda_1 = \lambda_2 = 1, \gamma_1 = \gamma_2 = 1.5$ and $\alpha = 0.8$. $(n = 100, s_1^* = -1.5)$

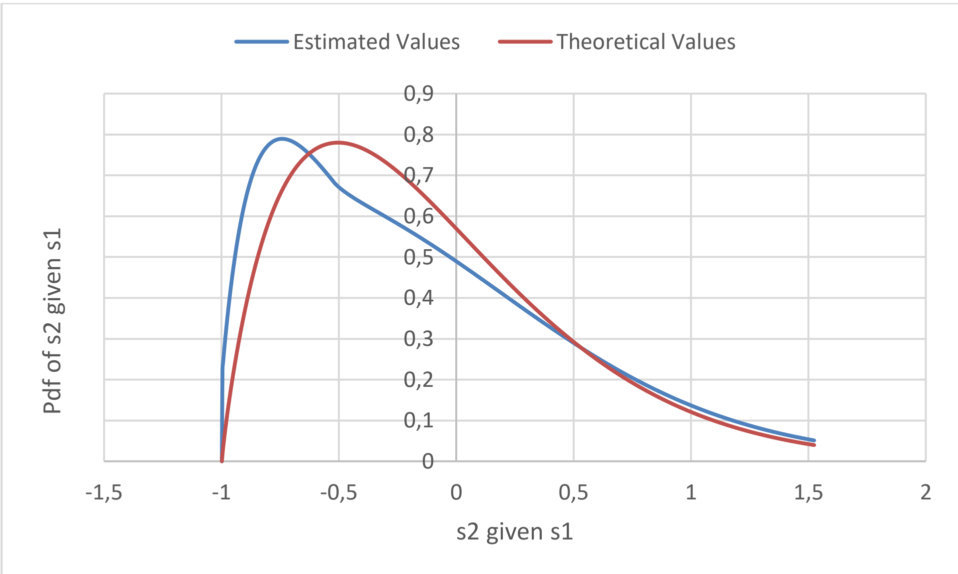

Figure 3. Estimated and theoretical pdf of $s_2|s_1 = s_1^*$ in the bivariate Weibull model with parameters $\lambda_1 = \lambda_2 = 1, \gamma_1 = \gamma_2 = 1.5$ and $\alpha = 0.8$. $(n = 100, s_1^* = -0.1875)$

As the estimated marginal pdf of s_1 and the estimated conditional pdf of $s_2|s_1 = s_1^*$ are both resembling the corresponding theoretical distributions, we would expect the product of the above marginal and conditional pdf to be a reasonable estimate of the theoretical joint pdf of s_1 and s_2 (or y_1 and y_2).

5. Concluding Remarks

The method proposed in this paper works reasonably well for the data generated from the bivariate Weibull distribution. The reason is that the power-normal distribution and the distribution derived from a quadratic function of two power-normal random variables are able to fit the underlying marginal and conditional distribution functions which are unimodal and taking on the zero value at the boundary point in the x-axis (or y-axis). When there is departure from unimodality of distribution and zero pdf at the boundary point, we may consider using a mixture of two power-normal distributions, and non-negative distribution which takes on nonzero value at the boundary point.

In the case of multivariate non-negative distribution with dimension larger than two, we may attempt to extend the method given in this paper. However, there would then be the need to tackle the issue of long computing time.

References

[1] C. Genest and J. McKay, The Joy of Copulas: Bivariate distributions with uniform marginals, *The American Statistician*, **40** (1986), 280-283. https://doi.org/10.2307/2684602

[2] P. Hougaard, A class of multivariate failure time distributions, *Biometrika*, **73** (1986), 671–678. https://doi.org/10.2307/2684602

[3] N.L. Johnson, Bivariate distributions based on simple translation systems, *Biometrika*, **36** (1949), 297–304. https://doi.org/10.2307/2332669

[4] N.L. Johnson and S. Kotz, *Distributions in statistics: continuous multivariate distributions*, John Wiley & Sons, 1972.

[5] A. Marshall and I. Olkin, Families of multivariate distributions, *Journal of the American Statistical Association*, **83** (1988), 834–841. https://doi.org/10.1080/01621459.1988.10478671

[6] I.K. Yeo and R.A. Johnson, A new family of power transformations to improve normality or symmetry, *Biometrika*, **87** (4) (2000), 954-959. https://doi.org/10.1093/biomet/87.4.954

Received: August 28, 2021; Published: September 10, 2021