ISSN 0971-0388

Aligarh Journal of Statistics
Vol. 44(2024), 103-112

Bayes Analysis of Hadwiger Fertility Model Using Markov Chain Monte Carlo Simulation

Shambhavi Singh[1], Akanksha Gupta[2] and Satyanshu K. Upadhyay[3]
[Received on February, 2024. Accepted on May, 2024]

ABSTRACT

The paper provides the Bayes analysis of the Hadwiger fertility model using the Gaussian copula prior to quantifying the information available in terms of marginal probability into a comprehensive joint probability framework for the parameters of the considered Hadwiger model. The resulting posterior, being analytically intractable, is analyzed using Markov chain Monte Carlo simulation, in particular, the Metropolis algorithm. The findings are supported by a real data example of reported age-specific fertility rate data. The approach is useful in the sense that it allows for an apt representation and analysis of the relationship between the priors.

1. Introduction

Modelling the fertility curve has always drawn the attention of demographers. It is both interesting and challenging because of the varying nature of age-specific fertility rate (ASFR) patterns in different countries over time. As far as the study of fertility is concerned, numerous models with a number of parameters are proposed in the literature to depict the fertility curve. One of the earliest ASFR models proposed by Gilje (1969) and Hoem *et al.* (1981) is based on the Hadwiger function, which can be written as

$$f_h(x|\boldsymbol{\theta}) = \left(\frac{ab}{c}\right)\left(\frac{c}{x}\right)^{\frac{3}{2}} \exp\left\{-b^2\left(\frac{c}{x}+\frac{x}{c}-2\right)\right\}, \tag{1}$$

where x refers to the age of a woman at the time of birth of her child, and $a, b,$ and c are the three parameters associated with the model. In (1), $\boldsymbol{\theta}$ is used to denote the parameter vector, that is, $\boldsymbol{\theta} = (a, b, c)$.

Chandola *et al.* (1999) studied the demographic interpretation of these parameters. They concluded that the parameter a is 0.56 times the total fertility rate (TFR), b is strongly associated with the maximum ASFR whereas c is related to the mean age of motherhood of the specified country. The following year Osona and Kohler (2000), while commenting on the same distribution, gave the relationship between the parameters of the distribution. The variance of the fertility structure was found to give a relation between the two parameters, b and c, by means of the equation.

$$\sigma^2 = \frac{1}{2}\frac{c^2}{b^2} \tag{2}$$

Furthermore, the authors also derived the formula for the parameter a by relating it to the TFR of the age-specific fertility schedules as

Corresponding author : Shambhavi Singh, Department of Statistics, Banaras Hindu University, Varanasi - 221 005, India. Email: shambhavisingh010@gmail.com
[2] Department of Statistics, Banaras Hindu University, Varanasi - 221 005, India.
[3] Department of Statistics and DST-Centre for Interdisciplinary Mathematical Sciences Banaras Hindu University, Varanasi - 221 005, India.

TFR= $\sqrt{\pi}a$. (3)

Obviously, the parameter a has the same relation with TFR as observed by Chandola *et al*. (1999). Chandola *et al.* (1999), in their paper studied the inferential aspects of the model from a classical perspective. The same model was later taken by Mishra and Upadhyay (2019), but the authors were mostly concerned with the Bayesian developments based on Markov chain Monte Carlo (MCMC) simulation-based approaches, particularly using the Gibbs sampler and the Metropolis algorithm (see Smith and Roberts (1993), Upadhyay *et al.* (2001)) as the resulting posterior was intractable analytically. The key idea of these simulation-based algorithms involves the construction of a Markov chain such that the chain's stationary distribution corresponds to the posterior of interest. The corresponding samples generated converge in distribution to the random sample from the posterior distribution of interest.

An important drawback of the study performed by Mishra and Upadhyay (2019) was that all three parameters were treated a priori independently while formulating the necessary Bayesian modelling. The assumption of a priori independence, however, may not be justified in the light of relationships between the parameters given by Osona and Kohler (2000) (see also (2)-(3)). The present paper attempts to overcome this drawback by considering the parameters to be dependent on each other and reflecting this dependence by defining an appropriate a priori distribution. The possible solution involves considering the joint prior of the parameters, but defining the joint prior requires the knowledge of conditional probability structuring, a situation that is generally difficult to anticipate by the demographers. The task becomes even more difficult if one has more than two parameters. Among various other possibilities to reflect the dependence structure between the parameters, one can consider copula-based priors, the simplest and the foremost being the Gaussian copula.

The plan of the paper is given below. The next section briefly discusses the copula model to be considered as the prior distribution for the concerned model parameters. Section 2 discusses the Bayesian model formulation, outlining the prior specification and then specifying the posterior distribution up to proportionality. A brief comment is also given on implementing the Metropolis algorithm for drawing the relevant sample-based inferences from the posterior specified up to proportionality only. Model compatibility study is discussed briefly in Section 3. A numerical illustration of the proposed study is given in Section 4, based on a real data set of women from Denmark. Finally, a brief conclusion is given in the last section.

1.1 The Copula

Quite often, while working with multidimensional models, it is difficult to arrive at their joint probability distributions given the marginal distributions. Copula comes into play in such cases and according to Nelsen (2007), it is a term that describes the functions that bond the marginal distributions to give the joint distribution. Copula can thus be perceived as a function used for coupling the marginal distributions. Statistically, these are distribution functions having marginals as Uniform (0,1). It is a linking function that takes into account the dependence measure having scale free approach. Sklar (1996), Hoeffding (1941), Kimeldori and Sampson (1975) and Deheuvels (1978), among others, independently studied copula in different forms. However, Sklar’s theorem (1959) (see also Schweizer and Sklar (2011)) could be regarded as a milestone in the development of the statistical significance of copulas. As per the theorem, given a joint distribution function, there exists a copula such that for all the random variables $Z_1, Z_2, \ldots, Z_k$, we have

$$H(z_1, z_2, \dots, z_k) = C\big(F_1(z_1), F_2(z_2), \dots, F_k(z_k)\big), \tag{4}$$

where $F_i(\cdot)$s are the marginal distribution functions and H is the joint distribution function. The converse of this theorem provides the basis for developing the joint priors. It states that given a copula *C*, the known marginal distribution functions $F_i(\cdot)s$ will lead us to the joint distribution H using (4). In this paper, we have used the inversion method (see Nelsen (2007)) to find the copula as

$$C(z_1, z_2, \dots, z_k) = H\left(F_1^{-1}(z_1), F_2^{-1}(z_2), \dots, F_k^{-1}(z_k)\right). \tag{5}$$

Geometric and algebraic methods are also popular, but in our case, the inversion method of copula generation is used since it allows one to get copulas with arbitrary margins. Gaussian copula is considered one of the best inversion copulas among others (see also Lin and Li (2014)).

Copulas are dynamic in that they model how the variables are related to each other, even in the case of a complex structural dependence, giving us the joint distributions without assuming anything about the marginal. Using copula as a prior has garnered increasing interest among researchers in recent years. Sharma and Das (2017) in their study have used the Gaussian as well as the *t*-copula priors with Lasso to conclude that in the case of regularization and variable selection, the two priors provided an improvement over the elastic net and *g*-priors. Elfadaly and Garthwaite (2017) compared the Dirichlet prior and the Gaussian copula prior in the case of multinomial model and observed that the Gaussian copula outperforms the former. Klein and Smith (2021) used the marginally calibrated copula prior in case of non-Gaussian responses.

2. Bayesian Model Formulation

It is assumed that each conception results in only one live birth and that the occurrence of birth is treated as success. Thus, births can be modelled as binomially distributed (see, Mishra and Upadhyay (2019)). Here, the probability of success will be given by the considered fertility graduating function. Next, it is assumed that the level of fecundity among each woman is uniform and it is not affected by any external forces. Thus, we can write the probability of birth at age x, that is, b_x, given the parameter vector $\boldsymbol{\theta}$ as

$$p(b_x|\boldsymbol{\theta}) = \binom{n_x}{b_x}\big(f_h(x|\boldsymbol{\theta})\big)^{b_x}\big(1 - f_h(x|\boldsymbol{\theta})\big)^{(n_x - b_x)}, \tag{6}$$

where n_x is the total number of women at age x and b_x is the number of births to n_x women. Furthermore, the age of child-bearing lies in the range of 15 to 49 years. Now, one can write the likelihood function as

$$L_h(\boldsymbol{b}, \boldsymbol{\theta}) = \prod_{x=15}^{49}\binom{n_x}{b_x}\big(f_h(x|\boldsymbol{\theta})\big)^{b_x}\big(1 - f_h(x|\boldsymbol{\theta})\big)^{(n_x - b_x)}, \tag{7}$$

where $\boldsymbol{b} = (b_{15}, b_{16}, \dots, b_{49})$ is the vector showing births to women at different child-bearing ages. Now, substituting $f(x|\boldsymbol{\theta})$ from (1) in (7), one can get the likelihood function corresponding to Hadwiger model and the same can be written as

$$\begin{aligned} L_h(\boldsymbol{b}, \boldsymbol{\theta}) = &\prod_{n=15}^{49}\binom{n_x}{b_x}\left[\left(\frac{ab}{c}\right)\left(\frac{c}{x}\right)^{(3/2)}\exp\left\{-b^2\left(\frac{c}{x} + \frac{x}{c} - 2\right)\right\}\right]^{b_x} \\ \times &\left[1 - \left(\frac{ab}{c}\right)\left(\frac{c}{x}\right)^{(3/2)}\exp\left\{-b^2\left(\frac{c}{x} + \frac{x}{c} - 2\right)\right\}\right]^{(n_x - b_x)}. \end{aligned} \tag{8}$$

2.1 The Prior and the Posterior Distribution

As mentioned, this paper considers the use of Gaussian copula as the prior density. Let us have a random vector consisting of k random variables defined as $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_k)$, then the Gaussian copula at $(z_1, z_2, \dots, z_k)$, with $G_i(z_i)$ indicating a cumulative distribution can be given as

$$C\big(G_1(z_1), G_2(z_2), \dots, G_k(z_k)\big) = \Phi\left(\phi^{-1}\big(G_1(z_1)\big), \phi^{-1}\big(G_2(z_2)\big), \dots, \phi^{-1}\big(G_k(z_k)\big)\right), \qquad (9)$$

where Φ denotes the cumulative distribution function (cdf) of a k-variate normal density having zero mean and unit variance and ϕ is the univariate marginal cdf of the standard normal variate. To use the Gaussian copula as the prior density, one has to procure the density function from the cdf. Thus, differentiating the copula given in (9) with respect to z_i, one gets

$$f(z_1, z_2, \dots, z_k|R) = \frac{\prod_{j=1}^{k} g_j(z_j)}{|R|^{1/2}} \exp\left\{\frac{-1}{2} y_k^T (R^{-1} - I_k) y_k\right\}, \qquad (10)$$

where $y_k^T = \left(\phi^{-1}\big(G_1(z_1)\big), \phi^{-1}\big(G_2(z_2)\big), \dots, \phi^{-1}\big(G_k(z_k)\big)\right)$, $g_j(z_j)$ is the density function corresponding to z_j and I_k is the identity matrix of order k. The correlation matrix R should be positive definite given the distribution as multivariate normal.

For the present study, since the parameter vector $\boldsymbol{\theta}$ has three components (a, b, c), one can consider $k = 3$ and, correspondingly, $z_1 = a, z_2 = b$ and $z_3 = c$. Therefore, the joint prior for $\boldsymbol{\theta}$ becomes

$$f(\boldsymbol{\theta}|R) = \frac{\prod_{j=1}^{3} g_j(z_j)}{|R|^{1/2}} \exp\left\{\frac{-1}{2} y_3^T (R^{-1} - I_3) y_3\right\}. \qquad (11)$$

It may be noted that some of the symbols on the right-hand side of (11) are retained from (10) for notational convenience. Further, $g_j(z_j)$ is proportional to a constant as the marginal distribution of the parameters is taken to be uniform with limits depending on their respective demographic interpretations. As such, since a is 0.56 times the TFR, the marginal density for a is taken to be $U(0,5)$. Similarly, given that b is proportional to the height of the ASFR curve, b is assumed to follow $U(0,4)$. Finally, since c is the mean age of motherhood, the density of c is assumed to follow $U(15,49)$. The correlation matrix R will be a 3×3 matrix and $y_3^T = \left(\phi^{-1}\big(G_1(z_1)\big), \phi^{-1}\big(G_2(z_2)\big), \phi^{-1}\big(G_3(z_3)\big)\right)$ where $z_i's$ are as described in the previous paragraph. Also, $U(*,**)$ in the previous description denotes the uniform density in the range $(*,**)$.

Once the prior distribution is specified, the next issue is the specification of prior hyperparameters. In this paper, we have instead considered a two-stage hierarchical Bayes approach. The prior for the first stage consists of the joint prior for the parameter vector $\boldsymbol{\theta}$ and the same for the second stage is introduced because of the correlation matrix R. Moreover, it is known that the correlation coefficient lies between -1 and 1 and no other information is available a priori. It is, therefore, appropriate to consider a second stage prior for r_{ij} as the uniform density $p(R)$ between the range $r_{ijl} = -1$ and $r_{iju} = 1$. Thus, considering the above a priori structure, one can obtain the posterior distribution up to proportionality as

$$p(\boldsymbol{\theta}, R|\boldsymbol{b}) \propto L_h(\boldsymbol{b}, \boldsymbol{\theta}) \times f(\boldsymbol{\theta}|R) \times p(R). \tag{12}$$

Combining (8), (11) and (12), the posterior distribution up to proportionality can be written as,

$$\begin{aligned} p(\boldsymbol{\theta}, R|\boldsymbol{b}) \propto \quad & \prod_{n=15}^{49} \binom{n_x}{b_x} \left[\left(\frac{ab}{c}\right) \left(\frac{c}{x}\right)^{(3/2)} \exp\left\{-b^2 \left(\frac{c}{x} + \frac{x}{c} - 2\right)\right\}\right]^{b_x} \\ & \times \left[1 - \left(\frac{ab}{c}\right) \left(\frac{c}{x}\right)^{(3/2)} \exp\left\{-b^2 \left(\frac{c}{x} + \frac{x}{c} - 2\right)\right\}\right]^{(n_x - b_x)} \\ & \times \frac{1}{|R|^{1/2}} \exp\left\{\frac{-1}{2} y_3^T (R^{-1} - I_3) y_3\right\} \prod_{i \neq j=1}^{3} I_{(r_{ijl}, r_{iju})}(r_{ij}), \end{aligned} \tag{13}$$

where $I_{(r_{ijl}, r_{iju})}(r_{ij})$ denotes an indicator function that takes value unity if r_{ij} belongs to the interval (r_{ijl}, r_{iju}). The above posterior is analytically intractable; thus, working with the same appears difficult. As an option, a sample-based approach may be used to get the posterior samples, hence the sample-based posterior inferences. One of the most widely used techniques for the generation of samples is the Metropolis algorithm (see, for example, Chib and Greenberg (1995) and Upadhyay et al. (2001) for more details and related diagnostic issues). Given that $\varphi(\boldsymbol{\theta}|d)$ is the posterior distribution specified up to proportionality. One needs to generate samples from the same. We take a symmetric Markov kernel $q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ given that $\boldsymbol{\theta}$ is the initial value and $\boldsymbol{\theta}'$ is the proposed realisation to be generated. The next step consists of accepting $\boldsymbol{\theta}'$ with probability α equals to

$$\alpha = \min\left[\frac{\varphi(\boldsymbol{\theta}'|d)}{\varphi(\boldsymbol{\theta}|d)}, 1\right]. \tag{14}$$

Now, the chain moves from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}'$ with probability α, otherwise it stays at $\boldsymbol{\theta}$. In the present study, we have used $MvN(\boldsymbol{\theta}, c_s\Sigma)$ as the Markovian kernel for simulating the chain where c_s is a scaling constant often taken between 0.5 and 1.0 and $MvN(\boldsymbol{\theta}, c_s\Sigma)$ denotes the multivariate normal distribution with mean vector $\boldsymbol{\theta}$ and variance-covariance matrix $c_s\Sigma$. For the initial values of $\boldsymbol{\theta}$ and Σ, one can consider, for example, maximum likelihood (ML) estimates and the corresponding Hessian-based approximation.

3. Model Compatibility

Once a model is proposed for the data in hand, it is essential to check if it is compatible with the data. A compatible model, of course, provides reasonable estimates and appears justified for the entertained data analysis. There are a number of techniques available within the Bayesian paradigm to assess the appropriateness of a model. One can use, for instance, the visual representations based on model- and data-based characteristics and check if the two characteristics appear to be similar when plotted on the same scale. Although informal, this similarity provides an impression that the entertained model can be appropriately considered for the data in hand. This paper considers the visual characterization of the proposed model to examine its suitability.

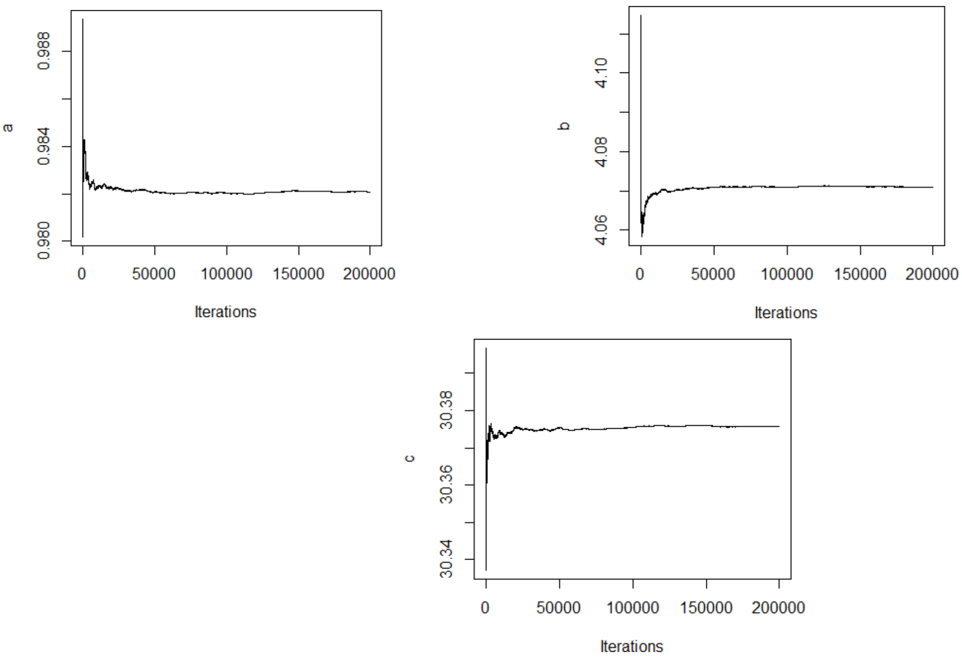

Figure 1: Trace plots for posterior samples of a, b and c exhibiting very good convergence in all the cases.

4. Numerical Illustration

For the numerical illustration of the proposed methodology, the paper considers a data set on ASFR for the year 2012 of women from Denmark in the child bearing age group 15 to 49 years. The data can be found in the repository of the human fertility collection database by the Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). The complete data set is not shown in the paper, though one can download it from the web page www.fertilitydata.org. The given data on ASFR were finally multiplied by 10^3 in order to provide the total number of births per 10^3 women. Also, the TFR was calculated from the data and the value was found to be 1.73.

Table 1: Estimated posterior characteristics of the parameters of Hadwiger density based on simulated posterior sample.

Parameters	ML Estimate	Estimated posterior characteristics			
		Mean	**Median**	**Mode**	**0.95 HPDI**
a	0.981	0.983	0.982	0.981	(0.958,1.007)
b	4.071	4.069	4.070	4.071	(3.982,4.151)
c	30.375	30.371	30.372	30.386	(30.248,30.551)

As discussed in Section 2, the Metropolis algorithm is used to generate samples from the posterior distribution given in (13). As mentioned, ML estimates of the parameters and the corresponding Hessian-based approximation were used as the initial values for the multivariate normal kernel. A value of 0.5 for the scaling constant c_s appeared to provide a good acceptance rate of the Metropolis steps.

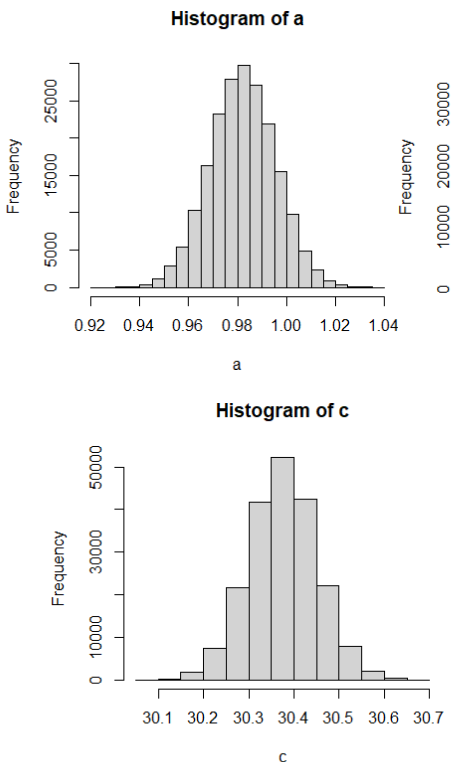

Figure 2: Histogram showing the marginal posterior density estimates of parameters. It is evident from the plot that all the marginals are symmetric.

After getting the convergence with the burn-in period of nearly 50 thousand iterations, a sample of size 10^3 was taken. The sample observations were chosen with a gap of 10 to minimize serial correlation among the generating variates. The trace plot of the Metropolis output for various parameters is shown in Figure 1. Using the finally picked up samples of size 10^3. The estimated posterior mean, median, and mode of the parameters were obtained, and they are shown in Table 1. The table also provides estimated highest posterior density interval with coverage probability 0.95 (0.95 HPrDI). These values can be used to see the concentration of the marginal posteriors towards the estimated central tendency measures.

It is obvious from Table 1 that the estimated posterior mode of the parameter a is 0.981, a value which is approximately equal to 0.56 times TFR. Also, the parameter c comes out to be 30.39, which is approximately equal to the mean age of fertility (see also (3)). Obviously, these estimated values are in accordance with the interpretation of the model parameters given in Section 1.

The estimated marginal posteriors in the form of histograms are shown in Figure 2. These figures are shown to get an overall idea of the marginal posterior densities. Say, for instance, one can immediately conclude from the Figure 2 that the marginal posterior densities are more or less symmetric for all the three parameters.

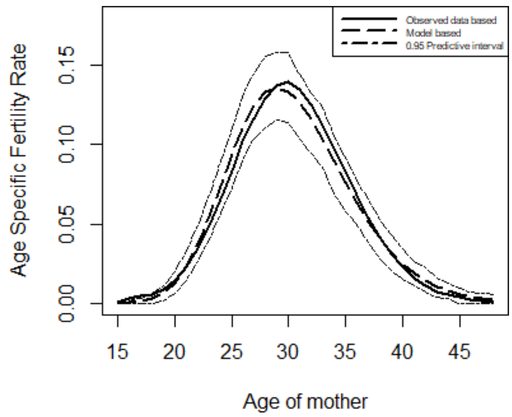

Figure 3: Model based predicted and observed data based ASFR plots. 0.95 HPrDI is also shown.

Finally, Figure 3 provides the visual representation showing the compatibility of the model with the real data. The bold continuous line in the figure represents the real data based ASFR values, whereas the large dashed line depicts the model based ASFR values. The dashed and dotted lines are used to represent the plots corresponding to the highest predictive density interval with coverage probability 0.95 (0.95 HPrDI). It can be seen that the data based and the model based plots show perfect compatibility as discussed in Section 3 and, therefore, it can be concluded that the entertained Hadwiger fertility model appears to be an apt candidate for the data in hand.

5. Conclusion

This study aims to test the effectiveness of the copula model as a prior in the Bayesian framework. To accomplish this, Gaussian copula density is employed as a joint prior for the parameters of the Hadwiger fertility model, allowing us to accurately represent the joint prior density considering the marginal densities of the parameters. It has been discovered that the model under consideration successfully explains the fertility pattern observed in Denmark. Also, the compatibility study confirms the appropriateness of the model with the data in hand, where appropriateness is judged based on both the prior and the likelihood. As such, the considered copula prior, which has an inherent dependency among the parameters, is recommended.

Acknowledgement

The authors express their thankfulness to the editor and the anonymous reviewers for their useful suggestions that improved the earlier version of the manuscript.

References

Chandola, T., Coleman, D. A., and Hiorns, R. W. (1999): Recent European fertility patterns: Fitting curves to ‘distorted’ distributions. *Population Studies*, **53(3)**:317–329. PMID: 11624024.

Chib, S. and Greenberg, E. (1995): Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49(4):**327–335.

Deheuvels,P. (1978): Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes. In *Annales de l’ISUP*. **23(3-4)**:1–36.

Elfadaly, F. G. and Garthwaite, P. H. (2017): Eliciting Dirichlet and Gaussian copula prior distributions for multinomial models. *Statistics and Computing*, **27**:449–467.

Gilje, E. (1969): Model for population projections for Norwegian regions. *Finnish Yearbook of Population Research*. 22–32.

Hoeffding, W. (1941): Masstabinvariante korrelationsmasse für diskontinuierliche verteilungen. *Archiv für mathematische Wirtschafts-und Sozialforschung*, **7**:49–70.

Hoem, J. M., Madsen, D., Nielsen, J. L., Ohlsen, E.-M., Hansen, H. O., and Renner- malm, B. (1981): Experiments in modelling recent Danish fertility curves. *Demography*, **18(2)**:231–244.

Kimeldori, G. and Sampson, A. (1975): Uniform representations of bivariate distributions. *Communications in Statistics–Theory and Methods*, **4(7)**:617–627.

Klein, N. and Smith, M. S. (2021): Bayesian variable selection for non-Gaussian responses: a marginally calibrated copula approach. *Biometrics*, **77(3)**:809–823.

Lin, J. and Li, X. (2014): Multivariate generalized Marshall–Olkin distributions and copulas. *Methodology and Computing in Applied Probability*, **16**:53–78.

Mishra, R. K. and Upadhyay, S. K. (2019): Parametric Bayes analyses to study the Age- specific fertility patterns. *American Journal of Mathematical and Management Sciences*, **38(2)**:151–173.

Nelsen, R. B. (2007): An Introduction to Copulas. Springer-Verlag, New York. Inc.; 2nd ed. 2006. Corr. 2nd printing 2007 edition.

Osona, J. A. O. and Kohler, H.-P. (2000): A comment on recent European fertility pat- terns: Fitting curves to distorted distributions by T. Chandola, DA Coleman and RW Hiorns. *Population Studies*, **54(3)**:347–349.

Schweizer, B. and Sklar, A. (2011): Probabilistic metric spaces. *Courier Corporation*.

Sharma, R. and Das, S. (2017): Regularization and variable selection with copula prior.*arXiv preprint arXiv:1709.05514*.

Sklar, A. (1959): Fonctions dé repatition à *n* dimensions et leurs marges. *Publ Inst Statist Univ Paris* **8**:229-231.

Sklar, A. (1996): Random variables, distribution functions, and copulas: a personal look backward and forward. *Lecture notes-monograph series*. 1–14.

Smith, A. F. and Roberts, G. O. (1993): Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, **55(1)**:3–23.

Upadhyay, S., Vasishta, N., and Smith, A. (2001): Bayes inference in life testing and reliability via Markov chain Monte Carlo simulation. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, **63(1)**:15–40.