Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v11n2p655

Bivariate Basu-Dhar geometric model for survival data with a cure fraction
By Martinez, Achcar, Icuma

Published: 14 October 2018

Electronic Journal of Applied Statistical Analysis
Vol. 11, Issue 02, October 2018, 655-673
DOI: 10.1285/i20705948v11n2p655

Bivariate Basu-Dhar geometric model for survival data with a cure fraction

Edson Zangiacomi Martinez*, Jorge Alberto Achcar, and Tatiana Reis Icuma

Department of Social Medicine, University of São Paulo (USP), Ribeirão Preto School of Medicine, Av. Bandeirantes 3900, Monte Alegre, 14049-900, Brazil.

Published: 14 October 2018

Under a context of survival lifetime analysis, we introduce in this paper Bayesian and maximum likelihood approaches for the bivariate Basu-Dhar geometric model in the presence of covariates and a cure fraction. This distribution is useful to model bivariate discrete lifetime data. In the Bayesian estimation, posterior summaries of interest were obtained using standard Markov Chain Monte Carlo methods in the OpenBUGS software. Maximum likelihood estimates for the parameters of interest were computed using the "maxLik" package of the R software. Illustrations of the proposed approaches are given for two real data sets.

keywords: Basu-Dhar distribution, cure fraction, discrete distributions, MCMC methods, lifetime data.

1. Introduction

Discrete-time survival analysis has been applied in many contexts including educational research (Singer and Willett, 1993), clinical psychology (Herzog et al., 1997) and medical studies (Scheike and Jensen, 1997). These analyses treat time to event data not as a continuous variable. Instead, it is considered discrete values in which the event of interest could occur. In this way, let T be a discrete random variable that takes positive integer values k, $k = 1, 2, 3, ...$, with probabilities $P(T = k)$. The survival function at time k is thus defined as

$$S(k) = P(T \geq k) = \sum_{j=k}^{\infty} P(T = j).$$

*Corresponding author: edson@fmrp.usp.br

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

The discrete Weibull distribution (Nakagawa and Osaki, 1975) is widely employed in modeling and analyzing lifetime data, but a large number of new discrete distributions has been proposed by several researchers (Almalki and Nadarajah, 2014; Dewan et al., 2016; Kulasekera and Tonkyn, 1992; Roy, 2004). In a special case, we may suppose that the data consists of two lifetimes T_1 and T_2 associated with each unit. The assumption of independence between these random variables can be unrealistic since the value of one variable can influence value of the other variable. In this context, a considerable number of parametric distributions assuming continuous bivariate lifetimes is found in the literature (Block and Basu, 1974; Freund, 1961; Hawkes, 1972; Sarkar, 1987), including the presence of censoring. Considering discrete lifetimes in place of continuous lifetime observations, Basu and Dhar (1995) derived the bivariate geometric distribution as an analogue to the bivariate distribution introduced by Marshall and Olkin (1967). Achcar et al. (2016) considered the parameter estimation of the Basu-Dhar distribution using Bayesian and maximum-likelihood methods in the presence of covariates and censored data. A discussion about computational issues related to the use of the Basu-Dhar distribution in applications with complete and censored datasets is provided by Oliveira and Achcar (2018).

In the present article we consider the bivariate Basu-Dhar geometric model for survival data in presence of a cure fraction. The cure fraction models are useful to model lifetime data with long-term survivors (Boag, 1949), where it is believed that a proportion of individuals will not experience the event of interest (Lambert et al., 2007). The presence of these individuals in a data set is suggested by a Kaplan-Meier plot of the survival function showing a long and stable plateau with heavy censoring at the extreme right of the plot (Martinez et al., 2013). We implemented the statistical model under a Bayesian framework with parameter estimation based on Markov Chain Monte Carlo (MCMC) methods. In a similar context, Davarzani et al. (2015) introduced a Bayesian analysis for the bivariate geometric distribution proposed by Arnold (1975) in the presence of covariates, censored data and cure fraction using MCMC methods. Although the distribution introduced by Arnold (1975) is also derived from the bivariate distribution by Marshall and Olkin, its joint mass probability has a different formulation from the Basu-Dhar geometric distribution.

2. Methods

2.1. The Basu-Dhar bivariate geometric distribution

Let T_1 and T_2 be two random variables that represent the time to events of interest. The joint survival function of the Basu-Dhar bivariate geometric (BDBG) distribution is given by

$$S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2) = p_1^{t_1} p_2^{t_2} p_3^{\max(t_1, t_2)},$$

where $t_1 \geq 1$ and $t_2 \geq 1$ are respectively discrete observations of T_1 and T_2, $0 < p_1 < 1$, $0 < p_2 < 1$ and $0 < p_3 \leq 1$. Achcar et al. (2016) described that this function satisfies

the loss of memory property without any additional parameter restrictions, or say, for any integer $\delta > 0$,

$$P\left(T_1 > s_1 + \delta, T_2 > s_2 + \delta | T_1 > s_1, T_2 > s_2\right) = P(T_1 > \delta, T_2 > \delta) = \left(p_1 p_2 p_3\right)^{\delta}.$$

The respective bivariate mass function is given by

$$P(T_1 = t_1, T_2 = t_2) = \begin{cases} p_1^{t_1 - 1}\left(1 - p_1\right)\left(p_2 p_3\right)^{t_2 - 1}\left(1 - p_2 p_3\right) & \text{for} \quad T_1 < T_2 \\ p_2^{t_2 - 1}\left(1 - p_2\right)\left(p_1 p_3\right)^{t_1 - 1}\left(1 - p_1 p_3\right) & \text{for} \quad T_1 > T_2 \\ \left(p_1 p_2 p_3\right)^{t_1 - 1}\left(1 - p_1 p_3 - p_2 p_3 + p_1 p_2 p_3\right) & \text{for} \quad T_1 = T_2 \end{cases}.$$

The marginal distributions for T_k are given by

$$P(T_k = t_k) = \left(1 - p_k p_3\right)\left(p_k p_3\right)^{t_k - 1}, \quad k = 1, 2, \tag{1}$$

and the respective means are given by

$$E\left(T_k\right) = \sum_{j=1}^{\infty}\left[j \times P(T_k = j)\right] = \frac{1}{1 - p_k p_3}, \quad k = 1, 2. \tag{2}$$

We can observe that

$$P(T_1 = t_1, T_2 > t_2) = \begin{cases} p_1^{t_1 - 1}\left(1 - p_1\right)\left(p_2 p_3\right)^{t_2} & \text{for} \quad T_1 \leq T_2 \\ p_2^{t_2}\left(p_1 p_3\right)^{t_1 - 1}\left(1 - p_1 p_3\right) & \text{for} \quad T_1 > T_2 \end{cases},$$

and

$$P(T_1 > t_1, T_2 = t_2) = \begin{cases} p_1^{t_1}\left(p_2 p_3\right)^{t_2 - 1}\left(1 - p_2 p_3\right) & \text{for} \quad T_1 < T_2 \\ p_2^{t_2 - 1}\left(1 - p_2\right)\left(p_1 p_3\right)^{t_1} & \text{for} \quad T_1 \geq T_2 \end{cases}.$$

Since

$$P(T_k < t_k) = \left(1 - p_k p_3\right)\sum_{j=1}^{t_k - 1}\left(p_k p_3\right)^{j - 1} = 1 - \left(p_k p_3\right)^{t_k - 1}, \quad k = 1, 2,$$

we obtain the marginal survival function for the lifetime k given by

$$S_k(t_k) = 1 - P(T_k < t_k) = \left(p_k p_3\right)^{t_k - 1}, \quad k = 1, 2, \tag{3}$$

and, from (1), the respective marginal hazard function is given by

$$h_k(t_k) = \frac{P(T_k = t_k)}{S_k(t)} = \frac{\left(1 - p_k p_3\right)\left(p_k p_3\right)^{t_k - 1}}{\left(p_k p_3\right)^{t_k - 1}} = \left(1 - p_k p_3\right), \quad k = 1, 2.$$

Thus, the BDBG distribution is characterized by a constant hazard function.

2.2. Presence of right-censored data

Let us suppose that either T_1 and T_2 can be censored and that censoring is independent of the time to the events considered by the study researchers. Let us define two indicator variables, denoted by d_{1i} and d_{2i}, where $d_{ki} = 1$ if t_{ki} is an observed lifetime and $d_{ki} = 0$ if t_{ki} is a censored observation, $k = 1, 2$ and $i = 1, ..., n$. In this way, if a sample of size n is sampled from a population of interest, the likelihood function is given by

$$\begin{aligned} L(\mathbf{p}) &= \prod_{i=1}^{n} \left[P(T_{1i} = t_{1i}, T_{2i} = t_{2i})\right]^{d_{1i}d_{2i}} \left[P(T_{1i} = t_{1i}, T_{2i} > t_{2i})\right]^{d_{1i}(1-d_{2i})} \\ &\quad \times \left[P(T_{1i} > t_{1i}, T_{2i} = t_{2i})\right]^{(1-d_{1i})d_{2i}} \left[S(t_{1i}, t_{2i})\right]^{(1-d_{1i})(1-d_{2i})}, \end{aligned}$$

where $\mathbf{p} = (p_1, p_2, p_3)'$ is the vector of parameters. Let us consider that the sample is partioned into three subsets, where $I_0 : \{i|t_{1i} = t_{2i} = t_i\}$, $I_1 : \{i|t_{1i} < t_{2i}\}$ and $I_2 : \{i|t_{1i} > t_{2i}\}$. The likelihood function can be thus written by

$$L(\mathbf{p}) = \prod_{i \in I_0} L_0(\mathbf{p}, \mathbf{w}_i) \prod_{i \in I_1} L_1(\mathbf{p}, \mathbf{w}_i) \prod_{i \in I_2} L_2(\mathbf{p}, \mathbf{w}_i), \tag{4}$$

where $\mathbf{w}_i = (t_{1i}, d_{1i}, t_{2i}, d_{2i})'$ and the components $L_0(\mathbf{p}, \mathbf{w}_i)$, $L_1(\mathbf{p}, \mathbf{w}_i)$ and $L_2(\mathbf{p}, \mathbf{w}_i)$ are respectively given by

$$\begin{aligned} L_0(\mathbf{p}, \mathbf{w}_i) &= p_1^{t_i - d_{1i}} p_2^{t_i - d_{2i}} p_3^{t_i - d_{1i}d_{2i}} (1 - p_1)^{d_{1i}(1-d_{2i})} (1 - p_2)^{(1-d_{1i})d_{2i}} \\ &\quad \times (1 - p_1p_3 - p_2p_3 + p_1p_2p_3)^{d_{1i}d_{2i}}, \end{aligned}$$

$$L_1(\mathbf{p}, \mathbf{w}_i) = p_1^{t_{1i} - d_{1i}} (1 - p_1)^{d_{1i}} (p_2p_3)^{t_{2i} - d_{2i}} (1 - p_2p_3)^{d_{2i}}$$

and

$$L_2(\mathbf{p}, \mathbf{w}_i) = p_2^{t_{2i} - d_{2i}} (1 - p_2)^{d_{2i}} (p_1p_3)^{t_{1i} - d_{1i}} (1 - p_1p_3)^{d_{1i}}.$$

From (4), the log-likelihood function is given by

$$\ln L(\mathbf{p}) = \sum_{i \in I_0} \ln L_0(\mathbf{p}, \mathbf{w}_i) + \sum_{i \in I_1} \ln L_1(\mathbf{p}, \mathbf{w}_i) + \sum_{i \in I_2} \ln L_2(\mathbf{p}, \mathbf{w}_i).$$

Based on this expression, maximum likelihood and Bayesian estimators of the parameters p_1, p_2 and p_3 were developed and compared by Achcar et al. (2016).

2.3. The cure fraction model

The survival function of the standard mixture cure fraction model has the form

$$S_C(t) = \theta + (1 - \theta)S(t) = 1 - (1 - \theta)F(t), \tag{5}$$

where θ is a parameter which represents the proportion of "long-term survivors" or "cured" individuals, regarding the event of interest ($0 < \theta < 1$), $S(t)$ is the baseline survival function for the susceptible or "non-cured" individuals and $F(t) = 1 - S(t)$ is

the respective cumulative distribution function (Boag, 1949). A generalization of (5) considering bivariate lifetimes T_1 and T_2 is given by

$$\begin{aligned} S_C(t_1,t_2) &= P_C(T_1 > t_1, T_2 > t_2) \\ &= \phi_{00} + \phi_{11}S_{12}(t_1,t_2) + \phi_{10}S_1(t_1) + \phi_{01}S_2(t_2) \end{aligned} \quad (6)$$

where $S_{12}(t_1,t_2)$ is the joint survival function for the susceptible individuals to both events, $S_k(t_k)$ is the baseline marginal survival function for T_k, $k = 1, 2$, and ϕ_{00}, ϕ_{11}, ϕ_{10} and ϕ_{01} are parameters to be estimated. Let $V_1$and V_2 be two indicator variables such that $V_k = 1$ for a susceptible individual in the lifetime T_k and $V_k = 0$ for a immune individual in T_k, $k = 1, 2$. Thus, $P(V_k = 0) = \theta_k$ and $P(V_k = 1) = 1 - \theta_k$. Let ω be the covariance between V_1 and V_2. Thus,

$$\phi_{00} = P(V_1 = 0, V_2 = 0) = P(V_1 = 0)P(V_2 = 0) + \omega = \theta_1\theta_2 + \omega,$$

$$\phi_{10} = P(V_1 = 1, V_2 = 0) = P(V_1 = 1)P(V_2 = 0) - \omega = (1 - \theta_1)\theta_2 - \omega,$$

$$\phi_{01} = P(V_1 = 0, V_2 = 1) = P(V_1 = 0)P(V_2 = 1) - \omega = \theta_1(1 - \theta_2) - \omega$$

and

$$\phi_{11} = P(V_1 = 1, V_2 = 1) = P(V_1 = 1)P(V_2 = 1) + \omega = (1 - \theta_1)(1 - \theta_2) + \omega.$$

Note that $\phi_{00} + \phi_{11} + \phi_{10} + \phi_{01} = 1$. In addition, $\theta_1 = \phi_{00} + \phi_{01}$ and $\theta_2 = \phi_{00} + \phi_{10}$ are interpreted as the cure fractions related to the lifetimes T_1 and T_2, respectively. Let us consider that each ith set of observations $(t_{1i}, d_{1i}, t_{2i}, d_{2i})$, $i = 1, ..., n$, can be classified into one of four groups:

(a) C_1: both t_{1i} and t_{2i} are complete observations, or say, they are uncensored lifetimes;

(b) C_2: t_{1i} is a complete observation and t_{2i} is a censored lifetime;

(c) C_3: t_{1i} is a censored lifetime and t_{2i} is a complete observation;

(d) C_4: both t_{1i} and t_{2i} are censored lifetimes.

Therefore, the likelihood function in presence of cure function and right censoring is given by

$$L(\mathbf{p},\theta_1,\theta_2,\omega) = \prod_{i\in C_1} P_1(t_{1i},t_{2i}) \prod_{i\in C_2} P_2(t_{1i},t_{2i}) \prod_{i\in C_3} P_3(t_{1i},t_{2i}) \prod_{i\in C_4} P_4(t_{1i},t_{2i}), \quad (7)$$

where $\mathbf{p} = (p_1, p_2, p_3)'$,

$$P_1(t_{1i},t_{2i}) = \phi_{11}P(T_{1i} = t_{1i}, T_{2i} = t_{2i}), \quad i \in C_1,$$

$$P_2(t_{1i}, t_{2i}) = \phi_{10} P(T_{1i} = t_{1i}) + \phi_{11} P(T_{1i} = t_{1i}, T_{2i} > t_{2i}), \quad i \in C_2,$$

$$P_3(t_{1i}, t_{2i}) = \phi_{01} P(T_{2i} = t_{2i}) + \phi_{11} P(T_1 > t_1, T_2 = t_2), \quad i \in C_3,$$

and, from (6),

$$P_4(t_{1i}, t_{2i}) = S_C(t_{1i}, t_{2i}), \quad i \in C_4.$$

Considering the BDBG distribution, we have

$$P_1(t_{1i}, t_{2i}) = \begin{cases} \phi_{11} p_1^{t_{1i}-1} (1-p_1) (p_2 p_3)^{t_{2i}-1} (1-p_2 p_3) & \text{for} \quad T_{1i} < T_{2i} \\ \phi_{11} p_2^{t_{2i}-1} (1-p_2) (p_1 p_3)^{t_{1i}-1} (1-p_1 p_3) & \text{for} \quad T_{1i} > T_{2i} \\ \phi_{11} (p_1 p_2 p_3)^{t_{1i}-1} (1-p_1 p_3 - p_2 p_3 + p_1 p_2 p_3) & \text{for} \quad T_{1i} = T_{2i} \end{cases},$$

$$P_2(t_{1i}, t_{2i}) = \begin{cases} \phi_{10} P(T_{1i} = t_{1i}) + \phi_{11} p_1^{t_{1i}-1} (1-p_1) (p_2 p_3)^{t_{2i}} & \text{for} \quad T_{1i} \leq T_{2i} \\ \phi_{10} P(T_{1i} = t_{1i}) + \phi_{11} p_2^{t_{2i}} (p_1 p_3)^{t_{1i}-1} (1-p_1 p_3) & \text{for} \quad T_{1i} > T_{2i} \end{cases},$$

$$P_3(t_{1i}, t_{2i}) = \begin{cases} \phi_{01} P(T_{2i} = t_{2i}) + \phi_{11} p_1^{t_{1i}} (p_2 p_3)^{t_{2i}-1} (1-p_2 p_3) & \text{for} \quad T_{1i} < T_{2i} \\ \phi_{01} P(T_{2i} = t_{2i}) + \phi_{11} p_2^{t_{2i}-1} (1-p_2) (p_1 p_3)^{t_{1i}} & \text{for} \quad T_{1i} \geq T_{2i} \end{cases},$$

where, from (1), $P(T_{ki} = t_{ki}) = (1-p_k p_3) (p_k p_3)^{t_{ki}-1}$, $k = 1, 2$, and

$$\begin{aligned} P_4(t_{1i}, t_{2i}) &= \phi_{00} + \phi_{11} P(T_{1i} > t_{1i}, T_{2i} > t_{2i}) + \phi_{10} S_1(t_{1i}) + \phi_{01} S_2(t_{2i}) \\ &= \phi_{00} + \phi_{11} p_1^{t_{1i}} p_2^{t_{2i}} p_3^{\max(t_{1i}, t_{2i})} + \phi_{10} S_1(t_{1i}) + \phi_{01} S_2(t_{2i}). \end{aligned}$$

From (3), we have that

$$P_4(t_{1i}, t_{2i}) = \phi_{00} + \phi_{11} p_1^{t_{1i}} p_2^{t_{2i}} p_3^{\max(t_{1i}, t_{2i})} + \phi_{10} (p_1 p_3)^{t_{1i}-1} + \phi_{01} (p_2 p_3)^{t_{2i}-1}.$$

Maximum-likelihood estimates for the parameters of interest are given by deriving the logarithm of the likelihood function $L(\mathbf{p}, \theta_1, \theta_2)$ with respect to each parameter and by equalizing the obtained expressions to zero. Considering that the resulting equations does not have an explicit solution, the estimates of p_1, p_2, p_3, θ_1 and θ_2 could be found using numerical methods.

The Akaike information criterion (AIC), given by $AIC = 2k - 2 \ln L(\widehat{\beta})$, where k is the number of parameters in $\widehat{\beta}$ included in the model and $\ln L(\widehat{\beta})$ is the maximum value of the log-likelihood function, is used to compare fits of the model under different formulations. The model with the lowest values of AIC is selected as fitting the sample data better.

In this article, MLE for the parameters p_1, p_2, p_3, θ_1, θ_2 and ω were computed using the "maxLik" package of the R software (Henningsen and Toomet, 2011).

2.4. Estimating the mean lifetime

In applications of the model to real data, estimates of the mean lifetime with the respective precision can be very useful to the researcher. In the frequentist case, let us consider the application of the "delta method" in estimating the sampling variances of the mean of the lifetime T_k, $k = 1, 2$, considering only the susceptible individuals. From (2),

$$g(p_k, p_3) = \frac{1}{1 - p_k p_3}, \quad k = 1, 2,$$

and an approximation of the variance of $g(p_k, p_3)$ is obtained by

$$Var\left[g(p_k, p_3)\right] \approx \left[\begin{array}{cc} \frac{\partial}{\partial p_k} g(p_k, p_3) & \frac{\partial}{\partial p_3} g(p_k, p_3) \end{array} \right] \mathbf{\Sigma}_{(p_k, p_3)} \left[\begin{array}{c} \frac{\partial}{\partial p_k} g(p_k, p_3) \\ \frac{\partial}{\partial p_3} g(p_k, p_3) \end{array} \right],$$

where

$$\begin{aligned} \frac{\partial}{\partial p_k} g(p_k, p_3) &= \frac{p_3}{(p_k p_3 - 1)^2}, \\ \frac{\partial}{\partial p_3} g(p_k, p_3) &= \frac{p_k}{(p_k p_3 - 1)^2} \end{aligned}$$

and $\mathbf{\Sigma}_{(p_k, p_3)}$ is the respective maximum-likelihood estimated variance-covariance matrix. Therefore, the Wald-type asymptotic 95% confidence limits for $E\left(T_k\right)$ are given by

$$g(\widehat{p}_k, \widehat{p}_3) - 1.96\sqrt{Var\left[g(\widehat{p}_k, \widehat{p}_3)\right]} \;\; , \;\; g(\widehat{p}_k, \widehat{p}_3) + 1.96\sqrt{Var\left[g(\widehat{p}_k, \widehat{p}_3)\right]},$$

where $g(\widehat{p}_k, \widehat{p}_3)$ is the point maximum likelihood estimate of the mean lifetime and $\widehat{p}_k$ and $\widehat{p}_3$ are the MLE for the parameters p_k and p_3, respectively, for $k = 1, 2$. An estimate of the variance-covariance matrix $\mathbf{\Sigma}_{(p_k, p_3)}$ is obtained as the inverse of the Fisher information matrix, that is, the matrix of the expected second derivatives of the log-likelihood function.

2.5. Bayesian analysis

According to Bayes' theorem, the posterior knowledge about the parameters of interest is proportional to the product of the respective prior knowledge and the likelihood function. Thus, let us suppose that the parameters p_1, p_2, p_3, θ_1 and θ_2 follow independent beta distributions. In this case,

$$p_j \sim Beta\left(a_j, b_j\right), \quad j = 1, 2, 3,$$

and

$$\theta_k \sim Beta\left(c_k, d_k\right), \quad k = 1, 2,$$

where $a_1 > 0$, $a_2 > 0$, $a_3 > 0$, $b_1 > 0$, $b_2 > 0$, $b_3 > 0$, $c_1 > 0$, $c_2 > 0$, $d_1 > 0$ and $d_2 > 0$ are known hyperparameters. In addition, considering that $0 \leq \omega \leq min(p1, p2) - p1p2$,

we can assume a generalized beta (GB) prior distribution for the covariance parameter ω. The GB distribution is a continuous probability distribution ranged in an interval not necessary equals to $[0, 1]$. The joint posterior density function for p_1, p_2, p_3, θ_1, θ_2 and ω is written as

$$f\left(\mathbf{p},\theta_1,\theta_2,\omega|\mathbf{w}\right) \propto L(\mathbf{p},\theta_1,\theta_2,\omega) \times f(\mathbf{p},\theta_1,\theta_2,\omega),$$

where the likelihood function $L(\mathbf{p},\theta_1,\theta_2,\omega)$ is given by (7), $f(\mathbf{p},\theta_1,\theta_2,\omega)$ is given by the product of the prior probability functions and $\mathbf{w} = (t_1, d_1, t_2, d_2)'$ is the vector of observed variables. Expressions for the full conditional posterior distributions are obtained by deriving the posterior distribution of each parameter given the data and all other parameters included in the model. In this article, we use Markov Chain Monte Carlo (MCMC) algorithms to get the posterior summaries of interest. In this way, we generated 510,000 Gibbs samples for each parameter, where the first 10,000 simulated samples were discarded as a burn-in period, which is used to minimize the effect of the initial values. The Bayesian estimates for the parameters of interest are obtained using the means of the respective posterior distributions. In addition, the 95% credible interval for a given parameter is estimated from the $2.5th$ and $97.5th$ centiles of its posterior distribution.

The deviance information criterion (DIC) is a generalization of the AIC for Bayesian analysis, easily obtained from the samples generated by MCMC simulation (Spiegelhalter et al., 2014). The DIC value is computed as $DIC = D(\widehat{\beta}) + 2n_p = 2\overline{D} - D(\widehat{\beta})$, where $D(\widehat{\beta})$ is the deviance evaluated using the mean of the parameters obtained from the MCMC samples, $\overline{D}$ is the posterior mean of the deviance and $n_p = \overline{D} - D(\widehat{\beta})$ is a measure of model complexity given by the effective number of parameters of the model. The model with the lowest values of DIC provides the best fit to data. Alternatively, comparison between different models can be also assessed using the log pseudo marginal likelihood ($LPML$) measure, derived from the conditional predictive ordinate (CPO) statistics. The larger is the value of $LPML$, the better is the fit of the model (Geisser and Eddy, 1979).

All Bayesian parameter estimates and their respective 95% credible intervals ($95\%CrI$) were obtained using the OpenBUGS software version 3.2.2. Convergence of the simulated samples was verified by traceplots. Details of our OpenBUGS program code are provided in an Appendix at the end of the paper.

3. Results

In this section, we present a numerical illustration of the proposed model to real medical data from two previously published studies.

3.1. The cervical carcinoma study

Let us consider the data from a study that included a total of 148 women diagnosed and treated for invasive cervical carcinoma between 1992 and 2002 (Brenna et al., 2004). This dataset was also used by Martinez and Achcar (2014) to illustrate the performance of a bivariate model based on the generalized Lindley distribution and copula functions. For the purposes of the present study, it was considered a subsample of 118 women who received the standard treatment recommended by the International Federation of Gynecology and Obstetrics (FIGO). Let us define the disease-free survival (DFS) as the time from the date of surgery to the first event of disease recurrence and the overall survival (OS) as the time from the date of surgery to the death. Thus, T_1 and T_2 are random variables denoting the DFS and the OS times, respectively. Note that, in this example, it is not possible the case in which $T_1 > T_2$, given that T_2 is the time to death. However, there are six observations in which $T_1 = T_2$. Both lifetimes are expressed in complete months.

In Table 1, we have the Bayesian (posterior summaries) and maximum likelihood estimates of the parameters for the model based on the BDBG distribution in the presence of censored data and a cure fraction. Note that the Bayesian and maximum likelihood estimates are very similar. Attempting to consider non-informative prior distributions for the parameters, in the Bayesian analysis we assumed uniform $U(0,1)$ prior distribution for the parameters p_1, p_2, p_3, θ_1 and θ_2 (remember that the uniform $U(0,1)$ distribution is an especial case of the beta distribution with both hyperparameters set to 1). In addition, we also assume a generalized beta prior distribution for ω, given by $Beta(1,1) \times min(p1, p2) - p1p2$. In a brief sensitivity analysis, we also assumed beta prior distributions with both hyperparameters set to the value 0.5 for the parameters, as a special case, but the results did not change substantially. In the frequentist case, standard errors are given Table 1 for the parameters ϕ_{00}, ϕ_{01}, ϕ_{10} and ϕ_{11}, and also for the quantities $E(T_1)$ and $E(T_2)$, which were obtained using the "delta method". We can note that the Wald method produced unrealistic confidence intervals (CI) for p_3, ϕ_{01} and ϕ_{10}, given that the upper limits can be higher than 1 and the lower limits can be lower than 0.

We also fitted alternative models assuming independence between the probabilities V_1 and V_2, which leads to $\omega = 0$. In this case, the Bayesian model generated a DIC value equals to 1189.0 and a $LPML$ value equals to -595.41, suggesting that the model that includes the term ω best fits the data, as indicated by the lowest DIC value and highest $LPML$ value (Table 1). In addition, the log-likelihood value under the frequentist model that considers $\omega = 0$ is -592.11 and the respective AIC value is 1194.21, which reinforces this conclusion.

Figure 1 shows plots of the survival function considering the DFS (panel A) and the OS times (panel B). This figure compares the survival functions obtained by the Kaplan-Meier (KM) estimator and the Bayesian parametric model based on the BDBG distribution in the presence of censored data and a cure fraction. The grey region in the

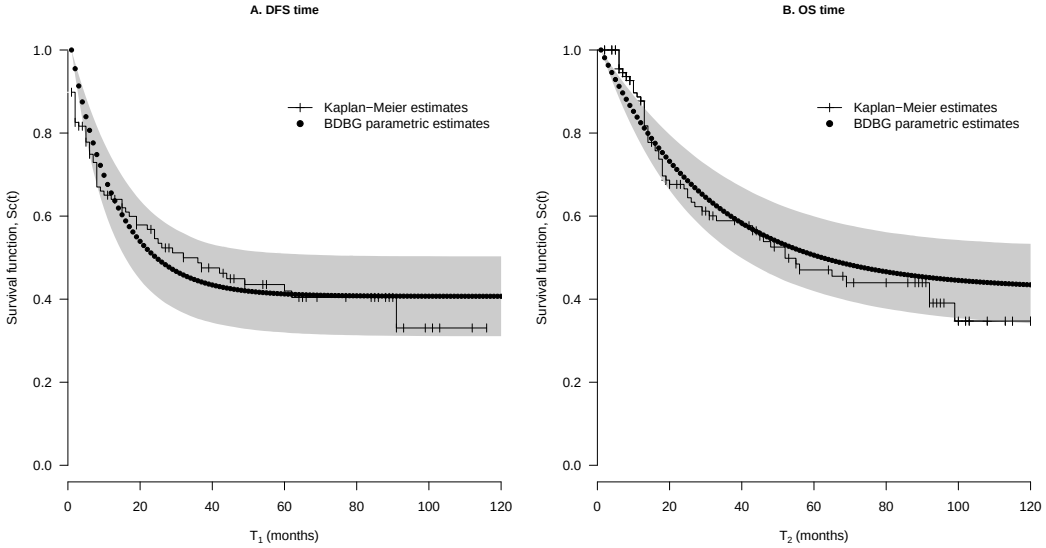

Figure 1: Plots of the survival functions estimated by the Kaplan-Meier method and assuming the Bayesian bivariate model based on the BDBG distribution. Censoring points are marked by vertical lines on the Kaplan-Meier survival curves. The grey region in the plots describes 95% credible regions for the parametric marginal survival functions. Data from the cervical carcinoma study (Brenna et al., 2004)

Table 1: Bayesian and maximum likelihood estimates. Data from the cervical carcinoma study (Brenna et al., 2004).

Parameter	Bayesian estimates			Maximum likelihood estimates		
	mean	standard deviation	95%CrI	estimate	standard error	Wald-type 95%CI
p_1	0.9395	0.0078	(0.9232 , 0.9539)	0.9399	0.0077	(0.9249 , 0.9551)
p_2	0.9838	0.0097	(0.9652 , 0.9992)	0.9735	0.0043	(0.9651 , 0.9819)
p_3	0.9836	0.0098	(0.9652 , 0.9991)	0.9979	0.0033	(0.9914 , 1.0043)
ϕ_{00}	0.5659	0.0509	(0.4662 , 0.6650)	0.5652	0.0724[(a)]	(0.4232 , 0.7071)
ϕ_{01}	0.0273	0.0205	(0.0012 , 0.0778)	0.0281	0.0753[(a)]	(-0.1196 , 0.1758)
ϕ_{10}	0.0112	0.0113	(0.0002 , 0.0420)	0.0128	0.0698[(a)]	(-0.1240 , 0.1495)
ϕ_{11}	0.3957	0.0490	(0.3004 , 0.4925)	0.3940	0.0899[(a)]	(0.2176 , 0.5703)
ω	0.2214	0.0137	(0.1899 , 0.2431)	0.2223	0.0497	(0.1249 , 0.3197)
θ_1	0.4069	0.0491	(0.3107 , 0.5031)	0.4067	0.0698	(0.2698 , 0.5435)
θ_2	0.4230	0.0511	(0.3249 , 0.5231)	0.4221	0.0929	(0.2398 , 0.6043)
$E(T_1)$	13.49	2.204	(10.05 , 18.47)	16.13	1.686[(a)]	(12.82 , 19.43)
$E(T_2)$	31.42	4.712	(23.44 , 41.84)	35.05	4.208[(a)]	(26.80 , 43.29)
	$LPML$: −549.54, DIC : 1099.0			Log-likelihood : −547.90, AIC: 1107.807		

[(a)]Standard errors obtained using the \ delta method".

Note: 95%CrI, 95% credible interval; 95%CI, 95% confidence interval.

plots describes 95% credible regions for the marginal survival functions, obtained from the $2.5th$ and $97.5th$ centiles of the posterior distribution for $S_C(t)$ (see equation (5)) at a lifetime $t = t^*$, where $t^* \geq 1$ is an integer value. The plots in Figure 1 suggest a reasonable approximation between the KM estimates and the values predicted by the parametric model.

3.2. The diabetic retinopathy study

In this subsection we consider the model based on the BDBG distribution in the presence of censoring and a cure fraction for the data from the Diabetic Retinopathy Study (Huster et al., 1989), where the objective was to assess the efficacy of photocoagulation treatment for proliferativeretinopathy. Each patient had one eye randomized to laser treatment and the other eye received no treatment. The main endpoint of this study was the severe visual loss in each eye. Let T_1 and T_2 be two random variables denoting the time up to visual loss for the control eye and the time up to visual loss for the treatment eye, respectively. Censoring was caused by death, dropout or end of the study. Table 2 compares the results obtained from the Bayesian analysis and the maximum likelihood estimation. In the Bayesian analysis, the choice of the prior distributions was similar to that of the previous application. As can be seen, the Bayesian and maximum likelihood estimates are very close to each other. We note that the upper limit of the 95%CI for p_3 is higher than 1, and the lower limit of the 95%CI for ϕ_{01} is lower than 0, reinforcing that

Table 2: Bayesian and maximum likelihood estimates. Data from the diabetic retinopathy study (Huster et al., 1989).

Parameter	Bayesian estimates			Maximum likelihood estimates		
	mean	standard deviation	95%CrI	estimate	standard error	Wald-type 95%CI
p_1	0.9720	0.0083	(0.9548 , 0.9872)	0.9711	0.0084	(0.9546 , 0.9875)
p_2	0.9713	0.0057	(0.9594 , 0.9815)	0.9703	0.0059	(0.9586 , 0.9820)
p_3	0.9937	0.0029	(0.9871 , 0.9982)	0.9952	0.0026	(0.9902 , 1.0003)
ϕ_{00}	0.3119	0.0620	(0.2162 , 0.4618)	0.2981	0.0516[(a)]	(0.1968 , 0.3992)
ϕ_{01}	0.0594	0.0319	(0.0069 , 0.1289)	0.0559	0.0360[(a)]	(-0.0147 , 0.1265)
ϕ_{10}	0.3472	0.0672	(0.2070 , 0.4755)	0.3493	0.0551[(a)]	(0.2413 , 0.4573)
ϕ_{11}	0.2815	0.0519	(0.1727 , 0.3765)	0.2967	0.0493[(a)]	(0.2001 , 0.3933)
ω	0.0667	0.0239	(0.0212 , 0.1156)	0.0689	0.0236	(0.0226 , 0.1152)
θ_1	0.6286	0.0669	(0.4626, 0.7288)	0.6461	0.0540	(0.5401 , 0.7520)
θ_2	0.3409	0.0616	(0.2115 , 0.4503)	0.3529	0.0613	(0.2323 , 0.4728)
$E(T_1)$	31.76	10.170	(19.19 , 57.49)	29.82	7.952[(a)]	(14.23 , 45.41)
$E(T_2)$	29.68	5.488	(21.08 , 42.28)	29.14	5.596[(a)]	(18.17 , 40.11)
	$LPML$: −824.74, DIC : 1649.0			Log-likelihood : −819.64, AIC: 1651.28		

[(a)]Standard errors obtained from the Delta method.

Note: 95%CrI, 95% credible interval; 95%CI, 95% confidence interval.

the Wald-type confidence intervals can produce unrealistic results when the respective parameter is estimated to be close to 0 or 1.

If we assume independence between the probabilities V_1 and V_2, a model fitted using $\omega = 0$ has a DIC value of 1655.0, and a $LPML$ of −828.14, as opposed to respective values of 1649.0 and −824.74 from the results in Table 2.

In the plots in Figure 2 we compared the survival functions for the T_1 and T_2 lifetimes estimated by the KM method and the marginal survival functions obtained from the Bayesian parametric model based on the BDBG distribution. We can note that the estimated survival functions are very close to each other. As in the previous application example, the grey region in the plots describes 95% credible regions for the parametric marginal survival functions.

4. Discussion and remarks

In this paper, we extend the problem of parameter estimation for the model based on the BDBG distribution in presence of censoring, as discussed by Achcar et al. (2016), to the case where a cure fraction is present. We showed that the frequentist and Bayesian approaches can be used for the purpose of estimating the parameters of the model, and

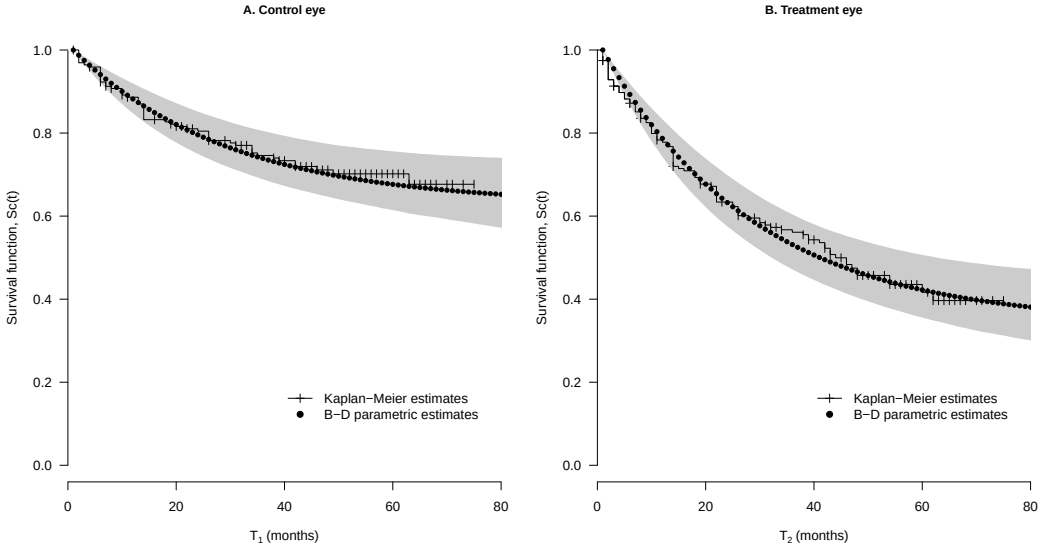

Figure 2: Plots of the survival functions estimated by the Kaplan-Meier method and assuming the Bayesian bivariate model based on the BDBG distribution. Censoring points are marked by vertical lines on the Kaplan-Meier survival curves. The grey region in the plots describes 95% credible regions for the parametric marginal survival functions. Data from the diabetic retinopathy study (Huster et al., 1989).

the use of free available software like the popular R and OpenBUGS software provide a great facility for this task.

For the Bayesian analysis of bivariate models with cure fractions, a number of authors assume a Dirichlet prior for the parameters ϕ_{00}, ϕ_{11}, ϕ_{10} and ϕ_{01}, since $\phi_{00} + \phi_{11} + \phi_{10} + \phi_{01} = 1$ (for examples, see Achcar et al. (2013) or Davarzani et al. (2015)). In the present article, we alternatively assumed prior beta distributions for the parameters θ_1 and θ_2, such that $\theta_1 = \phi_{00} + \phi_{01}$ and $\theta_2 = \phi_{00} + \phi_{10}$ are the cure fractions related to the lifetimes T_1 and T_2, respectively. As an advantage of this choice, we observed that assuming dependence between the random variables V_1 and V_2 by the introduction of the parameter ω denoting the covariance between these indicator variables resulted in much better fit of the proposed model for the numerical examples.

Achcar et al. (2016) state that models based on bivariate discrete distributions could be good alternatives to the analysis of bivariate real data in presence of censoring, since we usually find computational difficulties to get inferences for the parameters of interest using standard continuous lifetime distributions as the popular Block and Basu or the Marshall and Olkin bivariate exponential distributions. In this way, the parametric models based on the BDBG distribution are a good alternative to analyse bivariate discrete lifetime data, considering the ease of implementation using computational software as R or OpenBUGS. However, an important limitation of the BDBG distribution is that their marginal hazard functions are constant, and this can be unrealistic in many applications. Thus, future studies will be required to extend the BDBG distribution to provide a model with a more flexible hazard function while maintaining the ease of computational implementation.

Acknowledgement

The authors are very grateful to the Editor and a referee for their helpful and useful comments that improved the manuscript. The first and second authors were partially supported by CNPq-Brazil.

References

Achcar, J. A., Coelho-Barros, E. A. and Mazucheli, J. (2013). Block and Basu bivariate lifetime distribution in the presence of cure fraction. *Journal of Applied Statistics*, 40(9):1864-1874.

Achcar, J. A., Davarzani, N. and Souza, R. M. (2016). Basu-Dhar bivariate geometric distribution in the presence of covariates and censored data: a Bayesian approach. *Journal of Applied Statistics*, 43(9):1636-1648.

Almalki, S. J. and Nadarajah, S. (2014). A new discrete modified Weibull distribution. *IEEE Transactions on Reliability*, 63(1):68-80.

Arnold, B. (1975). A characterization of the exponential distribution by multivariate geometric compounding. *Sankhyā: The Indian Journal of Statistics, Series A*, 37(1):164-173.

Basu, A. P. and Dhar, S. (1995). Bivariate geometric distribution. *Journal of Applied Statistical Science*, 2(1):33-44.

Block, H. W. and Basu, A. P. (1974). A continuous bivariate exponential extension. *Journal of the American Statistical Association*, 69(348):1031-1037.

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):15-53.

Brenna, S. M., Silva, I. D., Zeferino, L. C., Pereira, J. S., Martinez, E. Z. and Syrjänen, K. J. (2004). Prognostic value of P53 codon 72 polymorphism in invasive cervical cancer in Brazil. *Gynecologic Oncology*, 93(2):374-380.

Davarzani, N., Achcar, J. A., Simirnov, E. N. and Peeters, R. (2015). Bivariate lifetime geometric distribution in presence of cure fraction. *Journal of Data Science*, 13(4):755-770.

Dewan, I., Sudheesh, K. K. and Anisha, P. (2016). Proportional hazards model for discrete data: some new developments. *Communications in Statistics - Theory and Methods*, 45(21):6481-6493.

Freund, J. E. (1961). A bivariate extension of the exponential distribution. *Journal of the American Statistical Association*, 56(296):971-977.

Geisser, S. and Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365):153-160.

Gupta, R. D. and Kundu, D. (1999). Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, 41(2):173-188.

Hawkes, A. G. (1972). A bivariate exponential distribution with applications to reliability. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(1):129-131.

Henningsen, A. and Toomet, O. (2011). maxLik: A package for maximum likelihood estimation in R. *Computational Statistics*, 26(3):443-458.

Herzog, W., Schellberg, D. and Deter, H. C. (1997). First recovery in anorexia nervosa patients in the long-term course: a discrete-time survival analysis. *Journal of*

Consulting and Clinical Psychology, 65(1):169-177.

Huster, W. J., Brookmeyer, R. and Self, S. G. (1989). Modelling paired survival data with covariates. *Biometrics*, 45(1):145-156.

Kulasekera, K. B. and Tonkyn, D. W. (1992). A new discrete distribution, with applications to survival, dispersal and dispersion. *Communications in Statistics-Simulation and Computation*, 21(2):499-518.

Lambert, P. C., Thompson, J. R., Weston, C. L. and Dickman, P. W. (2007). Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3):576-594.

Marshall, A. W. and Olkin, I. (1967). A generalized bivariate exponential distribution. *Journal of Applied Probability*, 4(2):291-302.

Martinez, E. Z. and Achcar J. A. (2014). Bayesian bivariate generalized Lindley model for survival data with a cure fraction. *Computer Methods and Programs in Biomedicine*, 117(2):145-157.

Martinez, E. Z., Achcar, J. A., Jácome, A. A. and Santos, J. S. (2013). Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. *Computer Methods and Programs in Biomedicine*, 112(3):343-355.

Nakagawa, T. and Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability*, 24(5):300-301.

Oliveira, R. P. and Achcar, J. A. (2018). Basu-Dhar's bivariate geometric distribution in presence of censored data and covariates: some computational aspects. *Electronic Journal of Applied Statistical Analysis*, 11(1):108-136.

Roy, D. (2004). Discrete Rayleigh distribution. *IEEE Transactions on Reliability*, 53(2):255-260.

Sarkar, S. K. (1987). A continuous bivariate exponential distribution. *Journal of the American Statistical Association*, 82(398):667-675.

Scheike, T. H. and Jensen, T. K. (1997). A discrete survival model with random effects: an application to time to pregnancy. *Biometrics*, 53(1):318-329.

Singer, J. D. and Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational and Behavioral Statistics*, 18(2):155-195.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2014). The deviance information criterion: 12 years on (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):485-493.

Appendix

A. OpenBUGS

The OpenBUGS code used for estimating the parameters of the BDBG distribution with presence of censoring and a cure fraction is given below.

```
model {
for (i in 1:N) {
 difT[i] <- t1[i] - t2[i]
 k1[i] <- 1 - step(difT[i])
 k3[i] <- equals(t1[i],t2[i])
 k2[i] <- (1-k1[i])*(1-k3[i])
 PT1[i] <- (1-p[1]*p[3])*pow(p[1]*p[3],t1[i]-1)
 PT2[i] <- (1-p[2]*p[3])*pow(p[2]*p[3],t2[i]-1)
 P1A[i] <- phi[1]*pow(p[1],t1[i]-1)*(1-p[1])
              *pow(p[2]*p[3],t2[i]-1)*(1-p[2]*p[3])
 P1B[i] <- phi[1]*pow(p[2],t2[i]-1)*(1-p[2])
              *pow(p[1]*p[3],t1[i]-1)*(1-p[1]*p[3])
 P1C[i] <- phi[1]*pow(p[1]*p[2]*p[3],t1[i]-1)
              *(1-p[1]*p[3]-p[2]*p[3]+p[1]*p[2]*p[3])
 P2A[i] <- phi[2]*PT1[i]+phi[1]*pow(p[1],t1[i]-1)
              *(1-p[1])*pow(p[2]*p[3],t2[i])
 P2B[i] <- phi[2]*PT1[i]+phi[1]*pow(p[2],t2[i])
              *pow(p[1]*p[3],t1[i]-1)*(1-p[1]*p[3])
 P2C[i] <- P2A[i]
 P3A[i] <- phi[3]*PT2[i]+phi[1]*pow(p[1],t1[i])
              *pow(p[2]*p[3],t2[i]-1)*(1-p[2]*p[3])
 P3B[i] <- phi[3]*PT2[i]+phi[1]*pow(p[2],t2[i]-1)
              *(1-p[2])*pow(p[1]*p[3],t1[i])
 P3C[i] <- P3B[i]
 P4[i] <- phi[4]+phi[1]*pow(p[1],t1[i])*pow(p[2],t2[i])
             *pow(p[3],max(t1[i],t2[i]))
             +phi[2]*pow(p[1]*p[3],t1[i]-1)
             +phi[3]*pow(p[2]*p[3],t2[i]-1)
 P1[i] <- pow(P1A[i],k1[i])*pow(P1B[i],k2[i])*pow(P1C[i],k3[i])
 P2[i] <- pow(P2A[i],k1[i])*pow(P2B[i],k2[i])*pow(P2C[i],k3[i])
 P3[i] <- pow(P3A[i],k1[i])*pow(P3B[i],k2[i])*pow(P3C[i],k3[i])
  L[i] <- pow(P1[i],d1[i]*d2[i])*pow(P2[i],d1[i]*(1-d2[i]))
     *pow(P3[i],(1-d1[i])*d2[i])*pow(P4[i],(1-d1[i])*(1-d2[i]))
  logL[i] <- log(L[i])
  zeros[i] <- 0
```

```
  zeros[i] ~ dloglik(logL[i])
}
 phi[1] <- (1-theta[1])*(1-theta[2]) + w
 phi[2] <- (1-theta[1])*theta[2] - w
 phi[3] <- theta[1]*(1-theta[2]) - w
 phi[4] <- theta[1]*theta[2] + w

# Prior distributions
 theta[1] ~  dbeta(1,1)
 theta[2] ~  dbeta(1,1)
 p[1] ~ dbeta(1,1)
 p[2] ~ dbeta(1,1)
 p[3] ~ dbeta(1,1)
 w <- g*(min(theta[1],theta[2]) - theta[1]*theta[2])
 g ~ dunif(0,1)

# Means of the lifetimes
m[1] <- 1/(1-p[1]*p[3])
m[2] <- 1/(1-p[2]*p[3])

# Marginal survival functions
for (t in 1:120)  {
S1[t] <- theta[1]+(1-theta[1])*pow(p[1]*p[3],t-1)
S2[t] <- theta[2]+(1-theta[2])*pow(p[2]*p[3],t-1) }
}
```

The data from the cervical carcinoma study in OpenBUGS format are shown below. This dataset can be freely used by students and researchers on their own works, but the article from (Brenna et al., 2004) should be cited as its source.

```
list(N=118, t1=c(28, 8, 116, 2, 1, 112, 7, 17, 32, 1, 19, 5, 29,
 42, 91, 8, 91, 85, 2, 1, 89, 3, 1, 5, 1, 84, 1, 112, 99, 1, 6,
 2, 62, 6, 103, 24, 1, 4, 19, 24, 26, 1, 8, 2, 2, 1, 77, 27, 65,
 2, 64, 8, 16, 91, 54, 90, 60, 8, 1, 1, 93, 3, 101, 10, 5, 12, 1,
 85, 37, 5, 89, 2, 91, 88, 44, 19, 1, 15, 69, 88, 25, 36, 86, 45,
 7, 22, 90, 2, 49, 9, 8, 66, 6, 15, 11, 1, 46, 23, 5, 28, 4, 4,
 15, 13, 1, 1, 6, 32, 49, 23, 2, 39, 19, 37, 55, 43, 1, 1),

t2=c(30, 14, 120, 18, 8, 115, 16, 20, 40, 7, 29, 13, 94, 46, 99,
 13, 92, 94, 10, 6, 90, 17, 6, 18, 4, 86, 10, 113, 100, 10, 14, 13,
 69, 13, 108, 52, 7, 5, 23, 52, 33, 4, 103, 17, 6, 6, 80, 31, 68,
 102, 64, 19, 25, 96, 56, 95, 65, 55, 6, 18, 94, 9, 102, 13, 31,
 18, 11, 90, 45, 14, 90, 16, 91, 92, 48, 25, 2, 27, 69, 89, 56,
```

```
 43, 88, 45, 13, 26, 93, 8, 92, 14, 12, 71, 9, 24, 15, 8, 49, 29,
 9, 32, 9, 7, 20, 19, 5, 5, 12, 38, 53, 29, 6, 44, 22, 42, 55, 43,
 5, 5),

d1=c(0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0,
 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1,
 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1,
 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0,
 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 1),

d2=c(0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0,
 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0,
 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1,
 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0,
 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0))
```