Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: http://www.tandfonline.com/loi/lssp20

Bayesian Estimation of Incomplete Data Using Conditionally Specified Priors

José María Sarabia & Golnaz Shahtahmassebi

To cite this article: José María Sarabia & Golnaz Shahtahmassebi (2015): Bayesian Estimation of Incomplete Data Using Conditionally Specified Priors, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2015.1091076

To link to this article: http://dx.doi.org/10.1080/03610918.2015.1091076

Accepted author version posted online: 23 Oct 2015.

Submit your article to this journal

Article views: 7

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=lssp20

Download by: [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] **Date:** 24 November 2015, At: 00:37

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

Bayesian Estimation of Incomplete Data Using Conditionally Specified Priors

July 22, 2015

José María Sarabia[a1], **Golnaz Shahtahmassebi**[b]

[a]*Department of Economics, University of Cantabria*
Avda. de los Castros s/n. 39005-Santander, Spain

[b]*School of Science and Technology, Nottingham Trent University*
Nottingham, UK

Abstract

In this paper, a class of conjugate prior for estimating incomplete count data based on a broad class of conjugate prior distributions is presented. The new class of prior distributions arises from a conditional perspective, making use of the conditional specification methodology and can be considered as the generalisation of the form of prior distributions that have been used previously in the estimation of incomplete count data well. Finally, some examples of simulated and real data are given.

Key Words: Conditional specification, Bayesian analysis, truncated gamma distribution, confluent hypergeometric distribution.

[1]Corresponding author. Tel.: +34-942-201635; Fax: +34-942-201603. E-mail address: sarabiaj@unican.es (J.M. Sarabia), golnaz.shahtahmassebi@ntu.ac.uk (Golnaz Shahtahmassebi).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

1 Introduction

Incomplete count data arises in problems where the true sample size is unknown and the true count may be under-reported. The estimation of such data is an important task in many fields of science, including biology (Barker and Sauer,1995), insurance and criminology (Smit *et al.*, 1997; Smyth and Carleton, 2011). Bayesian methodologies for the study of this problem have been proposed by Armero and Bayarri (1997) and Moreno and Girón (1998).

Let x_i be an unobserved count variable, for example the number of homicide crimes or accidents by the customers of a given insurance company in a given time period i, which follows a Poisson distribution with unknown parameter $\lambda > 0$. Assume that we only observe an unknown proportion of x_i, a count random variable denoted by y_i. For example, y_i can be the actual number of reported crimes or the claims to the insurance company. Suppose that the number of events x_i are reported independently of each other with probability θ and the same for all the events, then a possible model is (Moreno and Girón, 1998):

$$\Pr(X = x_i; \lambda) = \frac{(\lambda)^{x_i} e^{-\lambda}}{x_i!}, \quad x_i = 0, 1, 2, \ldots \tag{1}$$

with $\lambda > 0$ and

$$\Pr(Y = y_i | X = x_i, \theta) = \binom{x_i}{y_i} \theta^{y_i} (1-\theta)^{x_i - y_i}, \quad y_i = 0, 1, \ldots, x_i, \tag{2}$$

where $0 < \theta < 1$. If we assume that X is independent of θ given λ and Y is independent of λ given X and θ we have,

$$\begin{aligned} f(y_i|\lambda, \theta) &= \sum_{x_i = y_i}^{\infty} \Pr(Y = y_i | X = x_i, \lambda, \theta) \Pr(X = x_i | \lambda, \theta) \\ &= \sum_{x_i = y_i}^{\infty} \Pr(Y = y_i | X = x_i, \theta) \Pr(X = x_i; \lambda), \end{aligned}$$

and basic computations using (1) and (2) leads to,

$$f(y_i|\lambda, \theta) = \frac{(\lambda\theta)^{y_i} \exp(-\lambda\theta)}{y_i!}, \quad y_i = 0, 1, 2, \ldots, \tag{3}$$

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

where $\lambda > 0$ and $\theta \in [0,1]$, which will be denoted by $y_i \sim \mathcal{P}o(\lambda\theta)$. Thus, the distribution of the reported counts y_i given λ and θ is a Poisson with parameter $\lambda\theta$.

Then, if $y = (y_1, \ldots, y_n)$ is a sample of size n from (3), the likelihood function is given by,

$$f(y|\lambda, \theta) = \frac{(\lambda\theta)^{n\bar{y}} \exp(-n\lambda\theta)}{\prod_{i=1}^{n} y_i!}. \tag{4}$$

Note that in this likelihood it is obviously impossible to distinguish among all the pairs (λ, θ) with the same product $\lambda\theta$, and then the maximum likelihood method cannot estimate λ and θ, separately.

Alternatively, by defining a model in a Bayesian framework we are able to distinguish between λ and θ. A possible model can be set as follows,

$$\underline{y}|\lambda, \theta \sim \mathcal{P}o(\lambda\theta), \tag{5}$$
$$(\lambda, \theta) \sim \pi(\lambda, \theta), \tag{6}$$

where $\pi(\lambda, \theta)$ is a prior distribution to be specified.

If a conjugate prior distributions are adopted, a classical solution was provided by Armero and Bayarri (1994) and Moreno and Girón (1998), which considered the class of prior distributions,

$$\pi(\lambda, \theta) \propto \theta^{a_0 - 1}(1-\theta)^{b_0 - 1}\lambda^{c_0 - 1}\exp(-d_0\lambda - e_0\lambda\theta). \tag{7}$$

The prior distribution (7) has three important properties: (i) it is conjugate for the likelihood (4), (ii) it includes the independent case and (iii) only five parameters must be elicited.

The first property established a congruent model, which present important computational advantages and has a large tradition in classical Bayesian analysis. The fact of the model includes the independent case looks natural. With respect to the third property, five parameters can be sometimes insufficient in practice. In many contexts, the expert has a great deal of information for eliciting the prior distribution $\pi(\lambda, \theta)$. On the other hand (7) only admits negative correlations and this fact makes the models some restrictive.

As an alternative, the incomplete data can also be considered as missing data. The missing data are also parameters to be estimated. Thus, under the Bayesian framework, the prior distribution for missing data have its prior

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

distribution. The Gibbs sampler for the approach only has the conditional posterior distribution for θ and λ.

In this paper, a general methodology for estimating incomplete count data based on a broad class of conjugate prior distributions is presented. The new class of prior distributions arise in a natural way from a conditional perspective, making use of the conditional specification methodology proposed by Arnold, Castillo and Sarabia (1999, 2001). The new family of prior distributions is very flexible and contains as special cases many of the usual priors used previously in the estimation of incomplete count data, including the independence case and the proposals of Armero and Bayarri (1997) and Moreno and Girón (1998). As well as its flexibility, one of the main advantages of this distribution is that, because of its dependence on a large number of parameters, it is possible to incorporate a wide amount of prior information.

The remainder of the paper is organised as follows. In Section 2, we present the gamma, truncated gamma and confluent hypergeometric distributions, which will be used in the rest of the paper. Section 3 introduces and investigates the so-called gamma confluent hypergeometric conditionals distribution (BGCHC). We describe the Bayesian approach used to this problem in Section 4. The application of BGCHC distribution is illustrated though a sets of simulated and real data in Section 5. Finally, some conclusions are included in Section 6.

2 Basic distributions

2.1 Classical and truncated gamma distributions

A classical gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ will be denoted as $X \sim \mathcal{G}a(\alpha, \beta)$, with probability density function (PDF),

$$f(x; \alpha, \beta) = \frac{\beta^{\alpha} x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)}, \quad x > 0. \tag{8}$$

If $X \sim \mathcal{G}a(\alpha, \beta)$ then $E(X) = \frac{\alpha}{\beta}$.

A truncated gamma distribution in the interval $[0, 1]$ has the following PDF,

$$f(x; \alpha, \beta) = \frac{\beta^{\alpha} x^{\alpha-1} \exp(-\beta x)}{\gamma(\alpha, \beta)}, \quad 0 \leq x \leq 1, \tag{9}$$

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

and $\gamma(a,x)$ denotes the lower incomplete gamma function defined as,

$$\gamma(a,x) = \int_0^x t^{a-1} \exp(-t) dt,$$

where $a > 0$. A truncated gamma distribution with PDF (9) will be denoted as $X \sim \mathcal{TG}a(\alpha, \beta)$.

If $X \sim \mathcal{TG}(\alpha, \beta)$, the raw moments are given by,

$$E(X^r) = \frac{\gamma(\alpha + r, b\beta)}{\beta^r \gamma(\alpha, \beta)}, \quad r > 0.$$

2.2 Confluent Hypergeometric distribution

A random variable X is said to have a confluent hypergeometric distribution if its PDF is given by,

$$f(x; a, b, c) = Kx^{a-1}(1-x)^{b-1} \exp(-cx), \quad 0 \leq x \leq 1, \tag{10}$$

where $a, b > 0$, $c \in \mathbb{R}$, where

$$K^{-1} = B(a,b) {}_1F_1[a; a+b; -c], \tag{11}$$

and $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ denotes the beta function and ${}_1F_1[a,b,c]$ the confluent hypergeometric function (Abramowitz and Stegun, 1964), defined as,

$${}_1F_1[a; a+b; -c] = \frac{1}{B(a,b)} \int_0^1 t^{a-1}(1-t)^{b-1} \exp(-ct) dt = \sum_{k=0}^{\infty} \frac{(a)_k}{(a+b)_k} \cdot \frac{(-c)^k}{k!},$$

and $(a)_k = a(a+1)\cdots(a+k-1)$ is the ascending factorial. A distribution with PDF (10) will be denoted as $X \sim \mathcal{CH}(a,b,c)$ and was considered by Armero and Bayarri (1997), Gordy (1998) and Ng and Kotz (1995).

The confluent hypergeometric distribution can have the following particular forms if in (11):

- $c = 0$, then we obtain the classical beta distribution.
- $b = 1$, then we obtain the truncated gamma distribution in $[0,1]$ defined previously.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

Finally, the raw moments are given by,

$$E(X^r) = \frac{\Gamma(a+r)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+r)} \cdot \frac{{}_1F_1[a+r; a+b+r, -c]}{{}_1F_1[a; a+b, -c]},$$

where $r > 0$.

3 Prior Distributions based on Conditional Specification

A flexible and conjugate prior distribution for the specification (5)-(6) is based on the following reasoning. If θ is a known parameter, the gamma distribution is a conjugate distribution for λ and if λ is known, the confluent hypergeometric distribution is a conjugate prior distribution for θ. Consequently, it has sense to ask for the most general prior distribution $\pi(\lambda, \theta)$ such that the distribution of $\lambda|\theta$ is a gamma and distribution of $\theta|\lambda$ is a confluent hypergeometric distribution such that,

$$\lambda|\theta \sim \mathcal{G}a(\alpha_1(\theta), \beta_1(\theta)), \quad (12)$$

$$\theta|\lambda \sim \mathcal{CH}(\alpha_2(\lambda), \beta_2(\lambda), \gamma(\lambda)), \quad (13)$$

where $\alpha_1(\theta), \beta_1(\theta) : [0,1] \to \mathbb{R}_+$, $\alpha_2(\lambda), \beta_2(\lambda) : \mathbb{R}_+ \to \mathbb{R}_+$ and $\gamma(\lambda) : \mathbb{R} \to \mathbb{R}$.

Theorem 1 *The most general distribution with conditional distributions (12) and (13) is given by,*

$$\pi(\lambda, \theta; \boldsymbol{a}) = \lambda^{-1}\theta^{-1}(1-\theta)^{-1} \exp\{\underline{u}_\lambda^\top A \underline{v}_\theta\}, \quad \lambda > 0, \quad 0 \le \theta \le 1, \quad (14)$$

where the vectors $\underline{u}_\lambda$ *and* $\underline{v}_\theta$ *are given by,*

$$\underline{u}_\lambda^\top = (1, \log\lambda, -\lambda), \quad (15)$$

$$\underline{v}_\theta^\top = (1, \log\theta, \log(1-\theta), -\theta), \quad (16)$$

and $A = \{a_{ij}\}$, $i = 0, 1, 2$ *and* $j = 0, 1, 2, 3$. *The parameter* a_{00} *is the normalizing constant and must be chosen to satisfy* $\int\int \pi(\lambda, \theta; \boldsymbol{a}) d\lambda d\theta = 1$. *The parameters* $\{a_{ij}\}$ *must be selected to satisfy* $\int \pi(\lambda, \theta; \boldsymbol{a}) d\lambda < \infty$ *or* $\int \pi(\lambda, \theta; \boldsymbol{a}) d\theta < \infty$.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

Proof: The gamma and the confluent hypergeometric distributions can be written as,

$$f(\lambda|\theta) \propto \lambda^{\alpha_1-1}e^{-\beta_1\lambda} = \lambda^{-1}\exp\{\alpha_1\log\lambda - \beta_1\lambda\},$$

and

$$f(\theta|\lambda) \propto \theta^{\alpha_2-1}(1-\theta)^{\beta_2-2}e^{-\gamma\theta} = \theta^{-1}(1-\theta)^{-1}\exp\{\alpha_2\log\theta+\beta_2\log(1-\theta)-\gamma\theta\},$$

and then both families belong to the exponential family of distributions with canonical functions $(\log\lambda, -\lambda)$ and $\log\theta, \log(1-\theta), -\theta)$, respectively.

Now, we define $r_1(\lambda) = \lambda^{-1}$, $r_2(\theta) = \theta^{-1}(1-\theta)^{-1}$ and $u_\lambda^\top$ and $\underline{v}_\theta^\top$ as (15) and (16), respectively. Since both conditional distributions belong to the exponential family, and using Theorem 4.1 in Arnold, Castillo and Sarabia (1999), the most general bivariate distribution with conditionals (12) and (13) is given by,

$$\pi(\lambda,\theta;\boldsymbol{a}) = r_1(\lambda)r_2(\theta)\exp\{\underline{u}_\lambda^\top A\underline{v}_\theta\},$$

which is (14), being $A = \{a_{ij}\}$, $i = 0,1,2$ and $j = 0,1,2,3$. ∎

We call this type of prior distribution *bivariate gamma confluent hypergeometric conditionals distribution* (BGCHC), denote it by $(\lambda,\theta) \sim \mathcal{BGCHC}(\boldsymbol{a})$.

3.1 Properties of the BGCHC distribution

3.1.1 Joint PDF

If we expand (14) we obtain (changing the signs of some coefficients),

$$\begin{aligned}\pi(\lambda,\theta;\boldsymbol{a}) \quad = \quad & \lambda^{-1}\theta^{-1}(1-\theta)^{-1}\exp\{a_{00}+a_{10}\log\lambda - a_{20}\lambda + a_{01}\log\theta + a_{02}\log(1-\theta) \\ & -a_{03}\theta + a_{11}\log\lambda\log\theta + a_{12}\log\lambda\log(1-\theta) - a_{13}\theta\log\lambda \\ & -a_{21}\lambda\log\theta - a_{22}\lambda\log(1-\theta) - a_{23}\lambda\theta\},\end{aligned} \tag{17}$$

where $\lambda > 0$, $\theta \in [0,1]$ and a_{00} is the normalizing constant, which is a function of the rest of the parameters.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

3.1.2 Conditional distributions

The conditional distributions of (17) are (12) and (13), where the conditional parameters are given by,

$$\alpha_1(\theta) = a_{10} + a_{11}\log\theta + a_{12}\log(1-\theta) - a_{13}\theta, \tag{18}$$

$$\beta_1(\theta) = a_{20} + a_{21}\log\theta + a_{22}\log(1-\theta) + a_{23}\theta, \tag{19}$$

$$\alpha_2(\lambda) = a_{01} + a_{11}\log\lambda - a_{21}\lambda, \tag{20}$$

$$\beta_2(\lambda) = a_{02} + a_{12}\log\lambda - a_{22}\lambda, \tag{21}$$

$$\gamma(\lambda) = a_{03} + a_{13}\log\lambda + a_{23}\lambda, \tag{22}$$

where these parameters must satisfy one of the two following sets of constraints:

Case 1 (independent case):

$$a_{11} = a_{12} = a_{13} = a_{21} = a_{22} = a_{23} = 0,$$
$$a_{10}, a_{20}, a_{01}, a_{02} > 0$$

Case 2 (dependent case):

$$a_{11} < 0, \quad a_{21} < 0, \quad a_{01} + a_{11}\log(\frac{a_{11}}{a_{11}}) > a_{11},$$
$$a_{12} < 0, \quad a_{22} < 0, \quad a_{02} + a_{12}\log(\frac{a_{12}}{a_{22}}) > a_{12},$$
$$a_{13} < 0, \quad a_{23} > 0, \quad a_{03} + a_{13}\log(-\frac{a_{13}}{a_{23}}) > a_{13},$$

and

$$g(x_0^{(1)}; a_{10}, a_{11}, a_{12}, -a_{13}) > 0, \quad g(x_0^{(2)}; a_{20}, a_{21}, a_{22}, a_{23}) > 0,$$

where: $x_0^{(1)} = x_0(a_{10}, a_{11}, a_{12}, -a_{13})$, $x_0^{(1)} = x_0(a_{20}, a_{21}, a_{22}, a_{23})$ being

$$x_0(a,b,c,d) = \frac{-b-c+d-\sqrt{(b+c-d)^2+4bd}}{2d}$$

and $g(x; a, b, c, d) = a + b\log(x) + c\log(1-x) + dx$, with $0 < x < 1$.

3.1.3 Marginal distributions

The marginal distributions of λ is given by,

$$\pi_1(\lambda; \boldsymbol{a}) = \frac{e^{a_{00}}\lambda^{a_{10}-1}e^{-a_{20}\lambda}}{K(\alpha_2(\lambda), \beta_2(\lambda), \gamma(\lambda))}, \quad \lambda > 0,$$

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

where $K(a,b,c)$ is defined in (11) and $\alpha_2(\lambda)$, $\beta_2(\lambda)$ and $\gamma(\lambda)$ are defined in (20), (21) and (22) respectively.

The marginal distribution of θ is,

$$\pi_2(\theta;\boldsymbol{a}) = \frac{e^{a_{00}}\theta^{a_{01}-1}(1-\theta)^{a_{02}-1}e^{-a_{03}\theta}\Gamma(\alpha_1(\theta))}{\beta_1(\theta)^{\alpha_1(\theta)}}, \quad 0 \leq \theta \leq 1,$$

where $\alpha_1(t)$ and $\beta_1(t)$ are defined in (18) and (19) respectively.

3.1.4 Conditional expectations

The conditional expectations are,

$$E(\lambda|\theta) = \frac{\alpha_1(\theta)}{\beta_1(\theta)},$$

and

$$E(\theta|\lambda) = \frac{\alpha_2(\lambda)}{\alpha_2(\lambda)+\beta_2(\lambda)} \cdot \frac{{}_1F_1[\alpha_2(\lambda)+1;\alpha_2(\lambda)+\beta_2(\lambda)+1,-\gamma(\lambda)]}{{}_1F_1[\alpha_2(\lambda);\alpha_2(\lambda)+\beta_2(\lambda),-\gamma(\lambda)]},$$

where $\alpha_1(\theta)$, $\beta_1(\theta)$, $\alpha_2(\lambda)$, $\beta_2(\lambda)$ and $\gamma(\lambda)$ are defined in (18) to (22).

3.1.5 Mode

The mode is the solution of the system of equations,

$$1 - a_{10} + a_{20}\lambda + (a_{13}+a_{23}\lambda)\theta - (a_{12}-a_{22}\lambda)\log(1-\theta) - (a_{11}-a_{21}\lambda)\log\theta = 0,$$
$$-1 + a_{01} + (2 - a_{0+} + a_{21} - a_{2+}\lambda)\theta + (a_{03}+a_{23}\lambda)\theta^2 + (a_{11} - a_{1+}\theta + a_{13}\theta^2)\log\lambda = 0.$$

Numerical computations show that the solutions of previous equations do not need to be unique, and bimodality is possible. The property of multimodality appears in other models with conditional specification (see Arnold et al., 2000, Arnold, Castillo and Sarabia (2001) and Sarabia et al. (2005)).

3.1.6 Special Cases

The gamma confluent hypergeometric conditionals distribution includes as particular cases the following models:

1. The Armero-Bayarri (1997) and Moreno-Girón (1998) prior distribution: $a_{03} = a_{11} = a_{12} = a_{13} = a_{21} = a_{22} = 0$,

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

2. The independence case. This situation corresponds to the choice: $a_{11} = a_{12} = a_{12} = a_{21} = a_{22} = a_{23} = 0$.

3. The gamma truncated gamma conditionals distribution, which is obtained when $a_{02} = a_{12} = a_2 = 0$

4. The submodel considered by Gómez-Déniz et al. (2014), which corresponds to $a_{20} = a_{03} = a_{11} = a_{12} = a_{13} = a_{21} = a_{22} = 0$.

3.1.7 A simple submodel

Let consider the following bivariate distribution,

$$\pi(\lambda, \theta; \boldsymbol{a}, b) = \frac{a_{23}^{a_{10}}}{B(a_{01} - a_{10}, a_{02})\Gamma(a_{10})} \lambda^{a_{10}-1} \theta^{a_{01}-1} (1-\theta)^{a_{02}-1} \exp(-a_{23}\lambda\theta), \tag{23}$$

where $\lambda > 0$, $0 \leq \theta \leq 1$ and the parameters must satisfy the constraints,

$$a_{10}, a_{01}, a_{02}, a_{23} > 0, \quad a_{01} > a_{10}.$$

This model was considered by Gómez-Déniz et al. (2014) in a contest of risk theory.

In relation with (23) we have the following properties:

- The marginal distributions are:

$$\pi(\lambda; \boldsymbol{a}) = \frac{a_{23}^{a_{10}} \Gamma(a_{01} + a_{02} - a_{10})}{\Gamma(a_{01} + a_{02}) B(a_{01} - a_{10}, a_{02})} \lambda^{a_{10}-1} e^{-a_{23}\lambda} \, {}_1F_1[a_{02}, a_{01}+a_{02}, a_{23}\lambda],$$

  and $\theta \sim \mathcal{B}e(a_{01} - a_{10}, a_{02})$ with PDF,

$$\pi(\theta; \boldsymbol{a}) = \frac{\theta^{a_{01}-a_{10}-1}(1-\theta)^{a_{02}-1}}{B(a_{01} - a_{10}, a_{02})}, \quad 0 \leq \theta \leq 1.$$

- The conditional distribution of $\lambda|\theta$ is a classical gamma with parameters,

$$\lambda|\theta \sim \mathcal{G}a(a_{10}, a_{23}\theta),$$

  and the PDF of $\theta|\lambda$ is,

$$\pi(\theta|\lambda; \boldsymbol{a}) \propto \theta^{a_{01}-1}(1-\theta)^{a_{02}-1} e^{-a_{23}\lambda\theta}, \quad 0 \leq \theta \leq 1, \tag{24}$$

  that is $\theta|\lambda \sim \mathcal{CH}(a_{01}, a_{02}, a_{23}\lambda)$. It can be seen that (24) reduces to the truncated gamma distribution for values of $a_{02} = 1$.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

- The mathematical expectation of θ and λ are,

$$E(\lambda) = \frac{a_{10}(a_{01} - a_{10} + a_{02} - 1)}{a_{23}(a_{01} - a_{10} - 1)}, \tag{25}$$

$$E(\theta) = \frac{a_{01} - a_{10}}{a_{01} - a_{10} + a_{02}}. \tag{26}$$

- Since the first cross moment is $E(\lambda\theta) = \frac{a_{10}}{a_{23}}$, the covariance of (λ, θ) is,

$$cov(\lambda, \theta) = \frac{a_{10}b}{a_{23}(a_{01} - a_{10} + a_{02})(a_{10} - a_{01} + 1)}.$$

- The distribution (23) is unimodal with modal value (λ_0, θ_0) given by,

$$mode(\lambda) = \frac{(a_{10} - 1)(a_{01} - a_{10} + a_{02} - 1)}{a_{23}(a_{01} - a_{10})}, \tag{27}$$

$$mode(\theta) = \frac{a_{01} - a_{10}}{a_{01} - a_{10} + a_{02} - 1}. \tag{28}$$

4 Bayesian analysis with BGCHC prior

In order to obtain the posterior distribution, the likelihood function can be written as,

$$f(y|\lambda, \theta) \propto \exp(n\bar{y}\log\lambda + n\bar{y}\log\theta - n\lambda\theta), \tag{29}$$

where $\bar{y}$ denotes the sample mean.

If we assign

$$(\lambda, \theta) \sim \mathcal{BGCHC}(\boldsymbol{a}^{(0)}),$$

as a prior distribution and incorporate equation (17) with (29) we obtain the following posterior distribution

$$(\lambda, \theta)|y \sim \mathcal{BGCHC}(\boldsymbol{a}^{*}),$$

where the hyperparameters vector $\boldsymbol{a}^{(0)}$ is updated to $\boldsymbol{a}^{*}$ using the expression in Table 1. Note that only three of the eleven parameters are updated by data: a_{10}, a_{01} and a_{23}. The rest of the parameters do no change. However, the existence of these parameters permits more flexibility when we select the prior distribution.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

Table 1: Hyperparameter updating of the prior (17) with likelihood (29)

Parameter	Prior value	Posterior value
a_{10}	$a_{10}^{(0)}$	$a_{10}^{(0)} + \sum_{i=1}^{n} y_i$
a_{20}	$a_{20}^{(0)}$	$a_{20}^{(0)}$
a_{01}	$a_{01}^{(0)}$	$a_{01}^{(0)} + \sum_{i=1}^{n} y_i$
a_{02}	$a_{02}^{(0)}$	$a_{02}^{(0)}$
a_{03}	$a_{03}^{(0)}$	$a_{03}^{(0)}$
a_{11}	$a_{11}^{(0)}$	$a_{11}^{(0)}$
a_{12}	$a_{12}^{(0)}$	$a_{12}^{(0)}$
a_{13}	$a_{13}^{(0)}$	$a_{13}^{(0)}$
a_{21}	$a_{21}^{(0)}$	$a_{21}^{(0)}$
a_{22}	$a_{22}^{(0)}$	$a_{22}^{(0)}$
a_{23}	$a_{23}^{(0)}$	$a_{23}^{(0)} + n$

4.1 Parameter estimation

In the conditional context, Gibbs sampling is a natural estimation methodology. Assume that we are interested in approximating the posterior moments of a given function of λ and θ, say $\delta(\lambda, \theta)$.

Then, to approximate $E(\delta(\lambda, \theta)|y)$, we generate random values

$$\lambda_1, \theta_1, \lambda_2, \theta_2, \dots, \lambda_{m_0+m}, \theta_{m_0+m}$$

using the conditional distributions gamma and confluent hypergeometric,

$$\begin{array}{rcl} \lambda|(\theta, y) & \sim & \mathcal{G}a(\alpha_1^*(\theta), \beta_1^*(\theta)) \\ \theta|(\lambda, y) & \sim & \mathcal{CH}a(\alpha_2^*(\lambda), \beta_2^*(\lambda), \gamma^*(\lambda)), \end{array}$$

where the expressions for obtaining a_{ij} are given in Table 1 and m_0 is the number of iterations before burn-in. Thus, we have the estimator,

$$E(\delta(\lambda, \theta)|y) \approx \frac{1}{m} \sum_{i=m_0+1}^{m_0+m} \delta(\lambda_i, \theta_i).$$

As a previous step, me must elicit the hyperparameters a_{ij} using the methods proposed in Section 4.3.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

4.2 Posterior distribution of x given y

The posterior distribution of x given y can be computed using the formulation,

$$\pi(x|y) \propto \pi(x)\pi(y|x).$$

Then,

$$\begin{aligned}
\pi(x) &= \int \pi(x|\lambda)\pi(\lambda)d\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!}\frac{e^{a_{00}}\lambda^{a_{10}-1}e^{-a_{20}\lambda}}{K(\alpha_2(\lambda),\beta_2(\lambda),\gamma(\lambda))}d\lambda \\
&= \frac{1}{x!}\exp(a_{00}-\tilde{a}_{00}),
\end{aligned}$$

where $\tilde{a}_{00}$ is the normalizing constant changing a_{10} by $a_{10}+x$ and a_{20} by $a_{20}+1$ keeping the rest of the hyperparameters constant.

On the other hand,

$$\begin{aligned}
\pi(y|x) &= \int \pi(y|x,\theta)\pi(\theta|x)d\theta \\
&= \int_0^1 \binom{x}{y}\theta^y(1-\theta)^{x-y}\frac{e^{a_{00}}\theta^{a_{01}-1}(1-\theta)^{a_{02}-1}e^{-a_{03}\theta}\Gamma(\alpha_1(\theta))}{\beta_1(\theta)^{\alpha_1(\theta)}}d\theta \\
&= \binom{x}{y}\exp(a_{00}-\tilde{\tilde{a}}_{00}),
\end{aligned}$$

where we have assumed that $\pi(\theta|x)=\pi(\theta)$ and $\tilde{\tilde{a}}_{00}$ is the normalizing constant changing a_{01} by $a_{01}+y$ and a_{02} by $a_{20}+x-y$ keeping the rest of the hyperparameters constant. Finally, multiplying the last two formulas we obtain,

$$\pi(x|y) \propto \frac{\exp(2a_{00}-\tilde{a}_{00}-\tilde{\tilde{a}}_{00})}{y!(x-y)!}, \quad x=y,y+1,\ldots,\infty$$

In relation with previous formulation we point out the following difficulty. For the computation of $\pi(\theta|x)$ the conjugate prior for (θ,λ) does not simplify the computation. In order to avoid this difficulty, we have assumed that θ and X are independent so that $\pi(\theta|x)=\pi(\theta)$. However, note that this equation does not hold when λ and θ are not independent. In fact, θ and X are independent if and only if θ and λ are independent (this follows from Theorem 1 in Moreno and Girón, 1998).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

4.2.1 Expected value of x given y

An important quantity to be computed is the expected value of $x|y$. We have,

$$E(x|y) = y + \frac{\int_0^\infty \int_0^1 \lambda(1-\theta) f(y|\lambda,\theta)\pi(\lambda,\theta)d\theta d\lambda}{\int_0^\infty \int_0^1 f(y|\lambda,\theta)\pi(\lambda,\theta)d\theta d\lambda}. \tag{30}$$

The proof of this formula can be found in Moreno and Girón (1998) and some aspects of the proof have been included in the Appendix.

In the case of the general BGTG distribution, previous formula (30) is given by,

$$E(x|y) = y + \exp(a_{00}^{(1)} - a_{00}^{(2)}), \tag{31}$$

where $a_{00}^{(1)}$ corresponds to the normalizing constant changing a_{10} by $a_{10}+y+1$, a_{02} by $a_{02}+1$, a_{01} by $a_{01}+y$ and a_{23} by $a_{23}+1$ and $a_{00}^{(2)}$ corresponds to the normalizing constant changing a_{10} by $a_{10}+y$, a_{01} by $a_{01}+y$ and a_{23} by $a_{23}+1$.

In the case of the simple submodel (23), the quantity (31) can be written as,

$$E(x|y) = y + \frac{a_{02}(a_{10}+y)}{(a_{23}+1)(a_{01}-a_{10}-1)}.$$

4.3 Hyperparameter elicitation

In this section we consider the elicitation of the hyperparameters a_{ij}.

We begin with the submodel considered in Section 3.1.7, which depends on four parameters: a_{10}, a_{01}, a_{02} and a_{23}. For the elicitation we consider the means and modes of λ and θ given by the equations (25), (26), (27) and (28) respectively. We consider the nonlinear system,

$$\begin{aligned} E(\lambda) &= \bar{\lambda}, && (32)\\ E(\theta) &= \bar{\theta}, && (33)\\ mode(\lambda) &= \lambda_0, && (34)\\ mode(\theta) &= \theta_0. && (35) \end{aligned}$$

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

If we solve (32) to (35) for a_{10}, a_{01}, a_{02} and a_{23} we obtain,

$$
\begin{aligned}
a_{10} &= \frac{\bar{\lambda}(\bar{\theta} - \theta_0 + \theta_0\bar{\theta})}{\theta_0(\bar{\lambda}\bar{\theta} - \bar{\lambda} - \lambda_0\bar{\theta}) + \bar{\lambda}\bar{\theta}}, \\
a_{01} &= \frac{\theta_0^2(\lambda_0\bar{\theta}^2 + \bar{\lambda} - \bar{\lambda}\bar{\theta}^2) - 2\theta_0\bar{\lambda}\bar{\theta} + \bar{\theta}^2\bar{\lambda}}{(\theta_0 - \bar{\theta})(\lambda_0\theta_0\bar{\theta} + \theta_0\bar{\lambda} - \bar{\lambda}\bar{\theta} - \theta_0\bar{\lambda}\bar{\theta})}, \\
a_{02} &= \frac{\theta_0(1 - \bar{\theta})}{\theta_0 - \bar{\theta}}, \\
a_{23} &= \frac{\bar{\theta}}{\theta_0(\bar{\lambda}\bar{\theta} - \lambda_0\bar{\theta} - \bar{\lambda}) + \bar{\lambda}\bar{\theta}}
\end{aligned}
$$

It can be seen that in order to have $a_{02} > 0$ we must have $\theta_0 > \bar{\theta}$.

For the elicitation of the hyperparameters in the general case we can use the methods proposed in Arnold, Castillo and Sarabia (1999) and Sarabia et al. (2005). Assume that we have information about the mean and variance of the two conditional distributions $\lambda|\theta$ and $\theta|\lambda$, i.e.,

$$
\begin{aligned}
E(\lambda|\theta_i) &= \xi_i, \quad i = 1, 2, \ldots, n_1, \\
var(\lambda|\theta_i) &= \eta_i, \quad i = 1, 2, \ldots, n_1, \\
E(\theta|\lambda_j) &= \psi_j, \quad j = 1, 2, \ldots, n_2, \\
var(\theta|\lambda_j) &= \chi_j, \quad j = 1, 2, \ldots, n_2,
\end{aligned}
$$

where the values ξ_i, η_i, ψ_i and χ_i are known and $2n_1 + 2n_2 \geq 11$. Then, the a_{ij} value can be obtained by nonlinear least squares, minimizing the sum of the squares of the differences between the components of the pairs $(E(\lambda|\theta_i), \xi_i)$, $(var(\lambda|\theta_i), \eta_i)$, $(E(\theta|\lambda_j), \psi_j)$ and $(var(\theta|\lambda_j), \chi_j)$, for $i = 1, 2, \ldots, n_1$ and $j = 1, 2, \ldots, n_2$.

5 Numerical experiments

In this section we describe the implementation of the model incorporating simulated and reals sets of data. These examples illustrated the proposed methodology is easily automated, widely applicable and flexible with respect to the choice of function structure. We obtain the posterior distribution of λ and θ for special cases:

(a) The truncated gamma conditional prior distribution ($a_{02} = a_{12} = a_{22} = 0$).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

Table 2: Posterior mean (SD) and 95% HPD regions of θ and λ for the simulated data with $\lambda = 3$, $\theta = 0.3$ for special case (a).

	θ_{post}		λ_{post}	
	mean (SD)	95% CI	mean (SD)	95% CI
$n = 50$	0.361 (0.114)	(0.205,0.584)	2.878 (0.996)	(1.427,5.023)
$n = 100$	0.362 (0.114)	(0.205,0.583)	2.717 (0.900)	(1.428,4.600)
$n = 150$	0.371 (0.115)	(0.206, 0.585)	2.815 (0.923)	(1.539,4.742)

(b) The simple submodel where $\theta|\lambda \sim \mathcal{CH}(a_{01}, a_{02}, a_{23}\lambda)$.

All computations and simulations were done using `R`2.15.2 on a 3.20 GHz i3 processor desktop. Since confluent hypergeometric distribution is not a predefined distribution in `R`, functions for evaluating the $\mathcal{CH}$ probability distribution function and generating random samples from such distribution are written by the authors. A rejection sampling algorithm (Gamerman and Lopez, 2006) was used to generate random samples from a $\mathcal{CH}$ distribution. In our analysis the Gelman and Rubin's convergence diagnostic with the statistic values of 1.01 suggested that we can consider the convergence of the Gibbs sampler after $m_0 = 10000$ iterations (Gelman and Rubin, 1992). A further 5000 samples were collected after burn-in.

5.1 Simulation Study

We simulated random sets of incomplete data with the probability distribution given by (3) of size $n = 50, 100$ and 150 with parameters $\lambda = 3$ and $\theta = 0.3, 0.8$. Tables 2-5 present the posterior means, standard deviations (SD) and 95% highest probability density (HPD) region for the parameters of interest for special cases (a) and (b). It can be seen for all cases both models provide a close posterior mean to the true values of the parameters.

Finally, the posterior summary of $\delta(\lambda, \theta) = \lambda\delta$ are provided in Tables 6 and 7. Visualisations of the posterior density of $\delta(\lambda, \theta)$ at different values of parameters and $n = 50, 100$ and 150 are given in Figures 1 and 2. It can be seen that our belief about $\delta(\lambda, \theta)$ becomes stronger by the number of samples n. However, $\delta(\lambda, \theta)$ was overestimated (the posterior distribution was not centred around the true value) for larger values of θ when n is small.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

Table 3: Posterior mean (SD) and 95% HPD regions of θ and λ for the simulated data with $\lambda = 3$, $\theta = 0.3$ for special case (b).

	θ_{post}		λ_{post}	
	mean (SD)	95% CI	mean (SD)	95% CI
$n = 50$	0.399 (0.114)	(0.211,0.589)	2.626 (0.933)	(1.405, 4.860)
$n = 100$	0.404 (0.113)	(0.211,0.590)	2.430 (0.832)	(1.392,4.440)
$n = 150$	0.403 (0.118)	(0.208, 0.591)	2.40 (0.901)	(01.512,4.693)

Table 4: Posterior mean (SD) and 95% HPD regions of θ and λ for the simulated data with $\lambda = 3$, $\theta = 0.8$ for special case (a).

	$\delta(\lambda, \theta) = 0.3 \cdot 3 = 0.9$		$\delta(\lambda, \theta) = 0.8 \cdot 3 = 2.4$	
	mean (SD)	95% CI	mean (SD)	95% CI
$n = 50$	0.792 (0.145)	(0.560, 0.919)	3.369(0.766)	(2.387,5.339)
$n = 100$	0.781 (0.147)	(0.467,0.991)	3.234 (0.738)	(2.338,5.159)
$n = 150$	0.781 (0.150)	(0.461, 0.991)	3.240 (0.755)	(2.359, 5.252)

Table 5: Posterior mean (SD) and 95% HPD regions of θ and λ for the simulated data with $\lambda = 3$, $\theta = 0.8$ for special case (b).

	θ_{post}		λ_{post}	
	mean (SD)	95% CI	mean (SD)	95% CI
$n = 50$	0.85 (0.57)	(0.587, 0.987)	3.18 (0.58)	(2.39,4.69)
$n = 100$	0.84 (0.10)	(0.611, 0.987)	2.93 (0.45)	(2.33, 4.07)
$n = 150$	0.85 (0.10)	(0.803, 1.110)	2.92 (0.48)	(2.35, 4.21)

Table 6: Posterior mean (SD) and 95% HPD regions of $\delta(\lambda, \theta) = \lambda\theta$ for the simulated data for special case (a).

	$\delta(\lambda, \theta) = 0.3 \cdot 3 = 0.9$		$\delta(\lambda, \theta) = 0.8 \cdot 3 = 2.4$	
	mean (SD)	95% CI	mean (SD)	95% CI
$n = 50$	0.938 (0.138)	(0.684,1.222)	2.641 (0.229)	(2.213, 3.099)
$n = 100$	0.890 (0.093)	(0.720,1.082)	2.458 (0.154)	(2.167, 2.765)
$n = 150$	0.947(0.081)	(0.797, 1.115)	2.437 (0.13)	(2.203, 2.688)

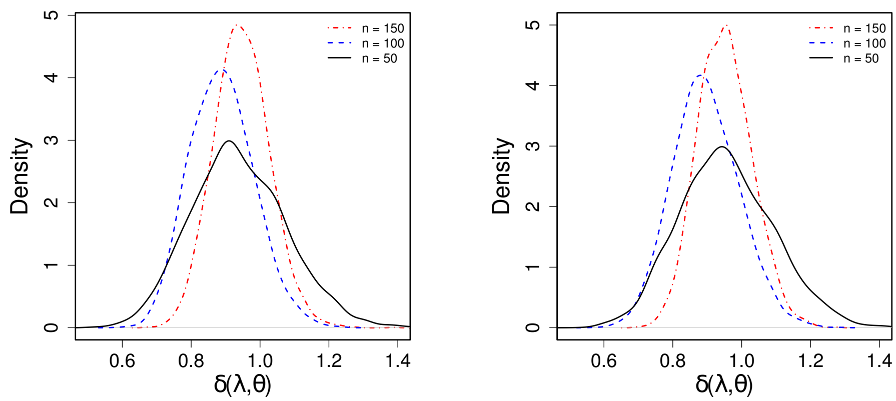
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

Table 7: Posterior mean (SD) and 95% HPD regions of $\delta(\lambda, \theta) = \lambda\theta$ for the simulated data for special case (b).

	$\delta(\lambda,\theta) = 0.3 \cdot 3 = 0.9$		$\delta(\lambda,\theta) = 0.8 \cdot 3 = 2.4$	
	mean (SD)	95% CI	mean (SD)	95% CI
$n = 50$	0.955 (0.137)	(0.710, 1.236)	2.651 (0.232)	(2.216, 3.121)
$n = 100$	0.895 (0.097)	(0.717, 1.092)	2.459 (0.157)	(2.163, 2.776)
$n = 150$	0.951 (0.078)	(0.807, 1.113)	2.442 (0.127)	(2.193, 2.698)



(a) (b)

Figure 1: The posterior density of $\lambda\theta$ with $n = 50$, 100, 150 for special cases (a) and (b) for the simulated data with $\lambda = 3$, $\theta = 0.3$.

5.2 Number of assaults

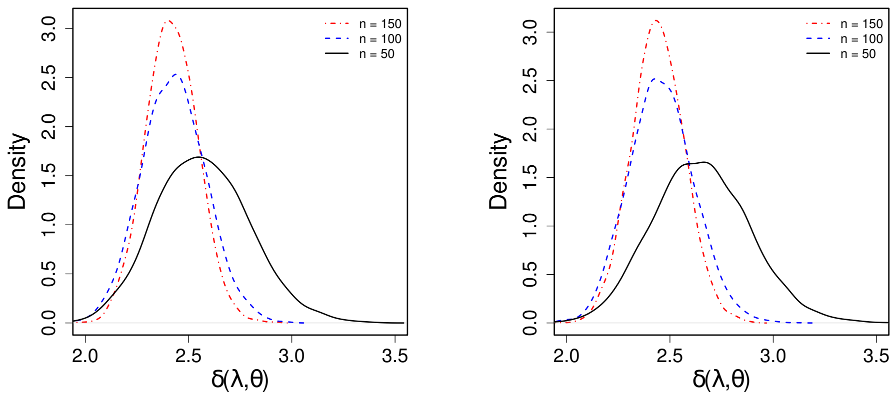
The data in the first column in Tables 8 and 9 are taken from Yannaros (1993) and Moreno and Girón (1998). They mean the reported assaults in Stockholm during seven years period 1980-1986. Let y_i be the number of of reported crimes of year i, for $i = 1980, \ldots, 1986$. For the purpose of parameter identifiability, after some experimentations, we found that for our data a good choice of the lower boundary for θ is $\theta_{\min} = 0.28$ and $\theta_{\max} = 1$ and the median of θ ($\tilde{\theta}$) is set to be 0.5 which was in agreement with Moreno and Girón (1998). Note that values of $\bar{\theta}, \bar{\lambda}$ and $\tilde{\lambda}$ are chosen to satisfy $a_{ij} > 0$ for $i, j = 0, 1, 2$. Our interest in this section specifically lies in $E(x|y)$.

Tables 8 and 9 illustrate the posterior predictive mean, standard deviation and 95% posterior predictive interval (PPI) of x. For this data, the $E(x|y)$

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015



(a) (b)

Figure 2: The posterior density of $\lambda\theta$ with $n = 50$, 100, 150 for special cases (a) and (b) for the simulated data with $\lambda = 3$, $\theta = 0.8$.

for the $\mathcal{CH}$ distribution are larger than the $E(x|y)$ of the $\mathcal{TG}$ distribution. This was compensated by smaller $SD(x|y)$ for the $\mathcal{CH}$ distribution. This may suggest that over 1980-1986, on average from every 2 crimes, only one crime was reported, using the the $\mathcal{CH}$ prior distribution. On the other, the posterior average of unreported crimes is smaller for the $\mathcal{TG}$ prior distribution. It can be seen that the 95% PPIs for the $\mathcal{CH}$ distribution are narrower than the 95% PPIs for the $\mathcal{TG}$ distribution.

y	$E(x\|y)$	$SD(x\|y)$	95% PPI
3303	6615	1751	(4146, 10816)
3334	6677	1761	(4185, 10918)
3931	7873	2083	(4934, 12873)
3857	7725	2044	(4841, 12631)
4154	8320	2303	(5214, 13603)
4345	8702	2303	(5454, 14229)
4224	8460	2239	(5302, 13833)

Table 8: Reported assault y in 1980-1986, posterior predictive mean, standard deviation and 95% posterior predictive intervals, using the $\mathcal{CH}$ prior distribution.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

y	$E(x\|y)$	$SD(x\|y)$	95% PPI
3303	4926	2051	(3307, 11359)
3334	4972	2071	(3338, 11466)
3931	5862	2441	(3936, 13519)
3857	5752	2395	(3862, 13264)
4154	6195	2580	(4160, 14286)
4345	6480	2699	(4351, 14943)
4224	6299	2623	(4230, 14527)

Table 9: Reported assault y in 1980-1986, posterior predictive mean, standard deviation and 95% posterior predictive intervals, using the $\mathcal{TG}$ prior distribution.

6 Conclusions

In the context of Bayesian estimation with incomplete count data, we have introduce a broad class of conjugate prior distributions for the corresponding likelihood. The new class of prior distributions arise in a natural way from a conditional perspective according to the conditional specification methodology proposed by Arnold, Castillo and Sarabia (1999, 2001). The new family of prior distributions depends on eleven parameters and is very flexible. The new family contains as special cases many of the usual priors used previously in the estimation of incomplete count data, including the independence case, the proposals of Armero and Bayarri (1997), Moreno and Girón (1998), the gamma truncated conditionals distribution and the model considered by Gómez-Déniz et al (2014). One of the main advantages of this distribution is that, because of its dependence on a large number of parameters, it is possible to incorporate a wide amount of prior information.

Finally, the applicability of the new distribution with $\mathcal{CH}$ and $\mathcal{TG}$ prior distributions, was illustrated for sets of simulated and real data. It was shown that under both priors, we are able to recover the true population. Possible extension to the application of this distribution would be a time varying θ for time dependent count data. In addition, its application in other areas such biology and criminology can be further explored.

Our model currently has a basic form which assumes the unobserved variable x_i follows a Poisson distribution. This can be a restrictive model especially for overdispersed data or data with excess of zeros. Thus, an extension to this model would be a negative binomial or a zero inflated Poisson distri-

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

bution as the distribution of the unobserved variable.

Acknowledgements

First author thanks to the Ministerio de Economía y Competitividad (ECO2013-48326-C2-2-P and Proyecto Puente) for partial support of this work. The authors would like to thank the editor and the reviewers for careful reading and for their comments and suggestions which greatly improved the paper.

References

Abramowitz, M., Stegun, I. (1964). *Handbook of Mathematical Functions.* National Bureau of Standards Applied Mathematical Sciences. U.S. Government Publishing Office, Washington, DC.

Armero, C., Bayarri, M.J. (1997). A Bayesian analysis of a queuing system with unlimited service. *Journal of Statistical Planning and Inference*, 58, 241-264.

Arnold, B.C., Castillo, E., Sarabia, J.M. (1999). *Conditional Specification of Statistical Models.* Springer Series in Statistics, Springer Verlag, New York.

Arnold, B.C., Castillo, E., Sarabia, J.M. (2001). Conditionally specified distributions: An introduction (with discussion). *Statistical Science* 16, 249-274.

Arnold, B.C., Castillo, E., Sarabia, J.M., González-Vega, L. (2000). Multiple Modes in Densities with Normal Conditionals. *Statistics and Probability Letters*, 49, 355-363.

Barker, R.J. and Sauer, J. R. (1995). Statistical aspects of point count sampling. In Ralph, J. C, Sauer, J. R., Droege, S. ed. *Monitoring Bird Populations by Point Counts.* Albany: Pacific Southwest Research Station, 125-130

Gamerman, D. and Lopes, H.F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* Chapman & Hall/CRC.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

Gómez-Déniz, E.; Hernández-Bastida, A.; Fernández-Sánchez, M.P. (2014). Computing credibility bonus-malus premiums using the aggregate claims distribution. *Hacettepe Journal of Mathematics and Statistics*, 43, 6, 1047-1061.

Gordy, M.B. (1998). Computationally convenient distributional assumptions for common-value actions. *Computational Economics*, 12, 61-78.

Moreno, E:, Girón, J. (1998). Estimating with incomplete count data: A Bayesian approach. *Journal of Statistical Planning and Inference*, 66, 147-159.

Ng K.W., Kotz, S. (1995). Kummer-Gamma and Kummer-Beta univariate and multivariate distributions. *Research Report*, No. 84, Department of Statistics, The University of Hong Kong, Hong Kong.

Sarabia, J.M., Castillo, E., Gómez, E., Vázquez, F. (2005). A Class of Conjugate Priors for Log-Normal Claims Based on Conditional Specification. *Journal of Risk and Insurance*, 72, 479-495.

Smit F, Toet J, Van der Heijden P. (1997). Estimating the number of opiate users in Rotterdam using statistical models for incomplete count data. In: Hay G, McKeganey N, Birks E. *Final report EMCDDA project methodological pilot study of local level prevalence estimates.* Glasgow: Centre for Drug Misuse Research, University of Glasgow / EMCDDA.

Smyth, Sara M. and Carleton,Rebecca (2011). *Measuring the Extent of CyberFraud: A Discussion Paper on Potential Methods and Data Sources*, Public Safety Canada.

Yannaros (1993). Analyzing Incomplete Count Data. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 42, 181-187.

Appendix

In this appendix we include the `R` code for computing the Confluent Hypergeometric distribution and a brief indication about the Rejection Sampling as well as the proof of the formula (30).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

Confluent Hypergeometric Distribution

```
library(coda)
library(truncdist)
library(fAsianOptions)

# CH probability density function
f_CHD<- function(x, a,b,g){
(x^(a-1)*(1-x)^(b-1)*exp(-g*x))/(beta(a,b)*Re(kummerM( x =-g,a=a,b=a+b)))
}
# cumulative distribution function of CH
F_CHD<- function(x, a,b,g){
n<- length(x)
F_chd<- numeric(n)
for(i in 1:n){
F_chd[i]<- integrate(CHD, lower = 0, upper = x[i], a,b,g)$value
}
return(F_chd)
}
```

Rejection Sampling

This method uses an auxiliary density for generation of random quantities from distributions not amenable to analytic treatment. Our aim is to generate random samples from the CH distribution, f_{CH}. We chose a uniform distribution ($q(a,b) = U(a,b)$) as the auxiliary distribution from which draws can be made. We use a standard uniform distribution, $U(0,1)$ to make generations from f_{CH}. For the ith sample, $i = 1, \ldots, n$

(1) Draw a random sample $\theta^{\text{cand}} \sim U(0,1)$.

(2) Generate $U \sim (0,1)$.

(3) Calculate

$$r = \frac{f_{\text{CH}}}{C\,U(0,1)}$$

where $C < \infty$ is chosen such that $f_{\text{CH}} \leq Cq(a,b)$ for every possible value of θ.

   (a) If $r \geq U$ then accept θ^{cand} and let $\theta^{(i)} = \theta^{\text{cand}}$.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [Universidad de Cantabria. Biblioteca], [Jose Maria Sarabia] at 00:37 24 November 2015

(b) Otherwise reject θ^{cand} and go back to 1.

Note that we can change the limits of the uniform distribution in order to generate samples from a truncated CH distribution. Also other distributions with same range as the CH distribution can be used as the auxiliary density.

Proof of the formula (30).

This proof can be found in Moreno and Girón (1998). The posterior expectation is,

$$E(x|y) = \frac{\int_0^\infty \int_0^1 \left[\sum_{x=y}^\infty x\pi(y|x,\theta)\pi(x|\lambda)\right] \pi(\lambda,\theta)d\theta d\lambda}{\int_0^\infty \int_0^1 f(y|\lambda,\theta)\pi(\lambda,\theta)d\theta d\lambda}.$$

Now, on the one hand,

$$\frac{\partial f(y|\lambda,\theta)}{\partial\lambda} = \sum_{x=y}^{\infty}(x/\lambda - t)\pi(y|x,\theta)\pi(x|\lambda),$$

and on the other hand directly from the probability mass function,

$$\frac{\partial f(y|\lambda,\theta)}{\partial\lambda} = \frac{y}{\lambda}f(y|\lambda,\theta) - \theta t f(y|\lambda,\theta).$$

Equating previous two equations we get,

$$\sum_{x=y}^{\infty} x\pi(y|x,\theta)\pi(x|\lambda) = [y - \lambda t(1-\theta)]f(y|\lambda,\theta),$$

and then we obtain (30) taking $t = 1$.

ACCEPTED MANUSCRIPT