Article

Time Series of Counts under Censoring: A Bayesian Approach

Isabel Silva [1] , **Maria Eduarda Silva** [2] , **Isabel Pereira** [3,*] **and Brendan McCabe** [4]

1   Faculdade de Engenharia, Universidade do Porto, CIDMA, 4200-465 Porto, Portugal
2   Faculdade de Economia, Universidade do Porto, LIADD-INESC TEC, 4200-464 Porto, Portugal
3   Departamento de Matemática, Universidade de Aveiro, CIDMA, 3810-193 Aveiro, Portugal
4   School of Management, University of Liverpool, Liverpool L69 3BX, UK
*   Correspondence: isabel.pereira@ua.pt

Abstract: Censored data are frequently found in diverse fields including environmental monitoring, medicine, economics and social sciences. Censoring occurs when observations are available only for a restricted range, e.g., due to a detection limit. Ignoring censoring produces biased estimates and unreliable statistical inference. The aim of this work is to contribute to the modelling of time series of counts under censoring using convolution closed infinitely divisible (CCID) models. The emphasis is on estimation and inference problems, using Bayesian approaches with Approximate Bayesian Computation (ABC) and Gibbs sampler with Data Augmentation (GDA) algorithms.

Keywords: Bayesian estimation; censored time series; convolution closed infinitely divisible; Poisson INAR(1) model

1. Introduction

Observations collected over time or space are usually correlated rather than independent. Time series are often observed with data irregularities such as missing values or detection limits. For instance, a monitoring device may have a technical detection limit and it records the limit value when the true value exceeds/precedes the detection limit. Such data is called censored (type 1) data and are common in environmental monitoring, physical sciences, business and economics. In particular, in the context of time series of counts, censored data arise in call centers. In fact, the demand measured by the number of calls is limited by the number of operators. When the number of calls is higher than the number of operators the data is right censored and the call center incurs under-staffing and poor service to the costumers.

The main consequence of neglecting censoring in the time series analysis is the loss of information that is reflected in biased and inconsistent estimators and altered serial correlation. These consequences can be summarized as problems in inference that lead to model misspecification, biased parameter estimation, and poor forecasts.

These problems have been solved in regression settings (i.i.d.) and partially solved for Gaussian time series (see for instance [1–7]). However, the problem of modelling time series under censoring in the context of time series of counts has, as yet, received little attention in the literature even though its relevance for inference. Count time series occur in many areas such as telecommunications, actuarial science, epidemiology, hydrology and environmental studies where the modelling of censored data may be invaluable in risk assessment.

In the context of time series of counts, Ref. [8] deal with correlated under-reported data through INAR(1)-hidden Markov chain models. A naïve method of parameter estimation was proposed, jointly with the maximum likelihood method based on a revised version of the forward algorithm. Additionally, Ref. [9] propose a random-censoring Poisson model for under-reported data, which accounts for the uncertainty about both the count and the data reporting processes.

Here, the problem of modelling count data under censoring is considered under a Bayesian perspective. In this paper, we consider a general class of convolution closed infinitely divisible (CCID) models as proposed by [10].

We investigate two natural approaches to analyse censored convolution closed infinitely divisible models of first order, CCID(1), using the Bayesian framework: the Approximate Bayesian Computation (ABC) methodology and the Gibbs sampler with Data Augmentation (GDA).

Since the CCID(1) under censoring presents an intractable likelihood, we resort to the Approximate Bayesian Computation methodology for estimating the model parameters. The presupposed model is simulated by using sample parameters taken from the prior distribution, then a distance between the simulated dataset and the observations is computed and when the simulated dataset is *very close* to the observed, the corresponding parameter samples are accepted as part of the posterior.

In addition, a widely used strategy to deal with censored data is to fill in censored data in order to create a data-augmented (complete) dataset. When the data-augmented posterior and the conditional pdf of the latent process are both available in a tractable form, the Gibbs sampler allows us to sample from the posterior distribution of the parameters of the complete dataset. This methodology is called Gibbs sampler with Data Augmentation (GDA). Here, a modified GDA, in which the data augmentation is achieved by multiple sampling of the latent variables from the truncated conditional distributions (GDA-MMS), is adopted.

The Poisson integer-valued autoregressive models of first-order, PoINAR(1), is one of the most popular classes of CCID models. It was proposed by [11,12] and extensively studied in the literature and applied to many real-world problems because of its ease of interpretation. To motivate the proposed approaches, we present in Figure 1 a synthetic dataset with $n = 350$ observations generated from a PoINAR(1) process with parameters $\alpha = 0.5$ and $\lambda = 5$ (X_t, blue line) and the respective right-censored dataset (Y_t, red line), at $L = 11$, corresponding to 30% of censoring. If we disregard the censoring, the estimates for the parameters (assuming an PoINAR(1) model without censoring) present a strong bias. For instance, in the frequentist framework, the conditional maximum likelihood estimates are $\hat{\alpha}_{CML} = 0.6174$ and $\hat{\lambda}_{CML} = 3.4078$, while in the Bayesian framework, the Gibbs sampler gives $\hat{\alpha}_{Bayes} = 0.6242$ and $\hat{\lambda}_{Bayes} = 3.3297$. On the other hand, if we assume a PoINAR(1) model under censoring, the parameter estimates given by the proposed approaches described in this work are, respectively, $\hat{\alpha}_{ABC} = 0.4623$ and $\hat{\lambda}_{ABC} = 5.2259$, and $\hat{\alpha}_{GDA} = 0.4834$ and $\hat{\lambda}_{GDA} = 4.9073$. Therefore, it is important to consider the censoring in data in order to avoid some inference issues that lead to a poor time series analysis.
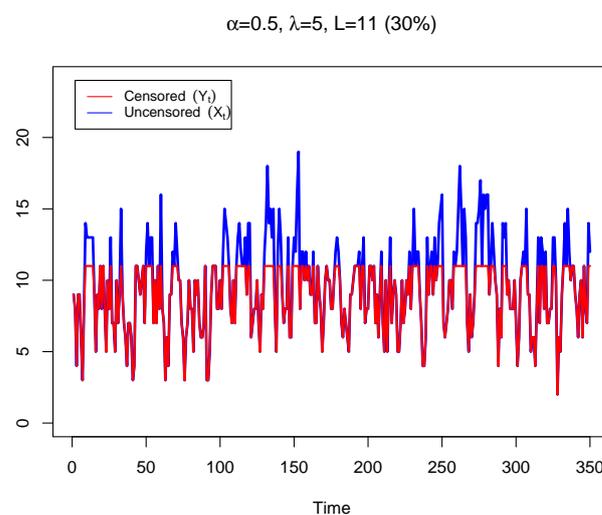
Figure 1. Synthetic dataset with $n = 350$ observations generated from a PoINAR(1) process with parameters $\alpha = 0.5$ and $\lambda = 5$ (X_t, blue line) and the respective right censored dataset (Y_t, red line), at $L = 11$.

The remainder of this work is organized as follows. Section 2 presents a general class of convolution closed infinitely divisible (CCID) models under censoring. Two Bayesian approaches proposed to estimate the parameters of the censored CCID(1) model are described in Section 3. The proposed methodologies are illustrated and compared with synthetic data in Section 4. Finally, Section 5 concludes the paper.

2. A Model for Time Series of Counts under Censoring

This section introduces a class of models adequate for censored time series of counts based on the convolution closed infinitely divisible (CCID) models as proposed by [10].

2.1. Convolution Closed Models for Count Time Series

First we introduce some notation. Consider a random variable X with a distribution F_μ, $\mu > 0$, belonging to the convolution closed infinitely divisible (CCID) parametric family [10]. This means, in particular, that the distribution F_μ is closed under convolution, $F_{\mu_1} * F_{\mu_2} = F_{\mu_1 + \mu_2}$, where $*$ is the convolution operator. Let $R(\cdot)$ denote a random operator on X such that $R(X) \sim F_{\alpha\mu}$, $0 < \alpha < 1$ and the conditional distribution of $R(X)$ given $X = x$ is $G_{\alpha\mu,(1-\alpha)\mu,x}$, $R(X)|X = x \sim G_{\alpha\mu,(1-\alpha)\mu,x}$. As an example, consider a Poisson random variable, $X \sim \text{Po}(\mu)$ and a binomial thinning operation, $R(X) = \alpha \circ X = \sum_{i=1}^{X} \xi_i$, $\xi_i \sim^{iid} Ber(\alpha)$. Then F_μ is the Poisson distribution with parameter μ, $R(X) \sim \text{Po}(\alpha\mu)$ and $R(X)|X = x \sim \text{Bi}(x, \alpha)$, $G_{\alpha\mu,(1-\alpha)\mu,x}$ is the Binomial distribution with parameters x and α.

A stationary time series, $\{X_t; t = 0, \pm1, \pm2, \ldots\}$ with margin F_μ, $X_t \sim F_\mu$, is called a convolution closed infinitely divisible process of order 1, CCID(1), if it satisfies the following equation

$$X_t = R_t(X_{t-1}) + e_t, \tag{1}$$

where the innovations e_t are independently and identically distributed (i.i.d.) with distribution $F_{(1-\alpha)\mu}$ and $\{R_t(\cdot) : t = 0, \pm1, \pm2, \ldots\}$ are independent replications of the random operator $R(\cdot)$ [10]. Note that the above construction leads to time series with the same marginal distribution as that of the innovations.

Model (1) encompasses many AR(1) models proposed in the literature for integer valued time series. In particular, the Poisson INAR(1), PoINAR(1), the negative binomial INAR(1), NBINAR(1), and the generalised Poisson INAR(1), GPINAR(1) [13], summarized in Table 1 (marginal distribution, random operation and its pmf $g(\cdot|\cdot)$, set of parameters $\boldsymbol{\theta}$), have been widely used in the literature to model time series of counts, see *inter alia* [14,15], among others.

Table 1. Methods for constructing integer valued AR(1) models with specified marginals F_μ and innovations $e_t \overset{iid}{=} F_\lambda$, $\lambda = \mu(1 - \alpha)$. $B(\cdot, \cdot)$ denotes the beta function.

Marginal Distribution	Random Operator	$g(s|X_{t-1}; \alpha)$	Innovations	$\boldsymbol{\theta}$
Poison Po(μ)	binomial thinning	$\binom{X_{t-1}}{s}\alpha^s(1-\alpha)^{X_{t-1}-s}$	Po(λ)	(μ, α)
Negative binomial NB(μ, ξ)	beta binomial thinning	$\binom{X_{t-1}}{s}\alpha^s(1-\alpha)^{X_{t-1}-s}$	NB(λ, ξ)	(μ, α, ξ)
Generalised Poisson GP(μ, ξ)	quasi binomial thinning	$\binom{X_{t-1}}{s}\alpha(\alpha + s(\xi/\mu))^{s-1}$ $(1 - \alpha - s(\xi/\mu))^{X_{t-1}-s}$	GP(λ, ξ)	(μ, α, ξ)

If one chooses $F_{(1-\alpha)\mu}$ as Poisson$((1-\alpha)\mu)$, and the random operation as the usual binomial thinning operation (based on underlying Bernoulli random variables) $R_t(X_{t-1}) = \alpha \circ X_{t-1} = \sum_{i=1}^{X_{t-1}} \xi_{ti}$, $\xi_{ti} \sim^{iid} Ber(\alpha)$, then F_μ is Poisson(μ) and the Poisson integer-valued autoregressive model, PoINAR(1), as proposed by [11,12], is recovered with the familiar representation

$$X_t = \alpha \circ X_{t-1} + \epsilon_t. \tag{2}$$

Since model (1) is Markovian [10], given a time series $x = (X_1, \ldots, X_n)$, the conditional likelihood is as follows

$$L(\boldsymbol{\theta}) = \prod_{t=2}^{n} f_{X_t|X_{t-1}}(x_t|x_{t-1}), \tag{3}$$

with

$$f_{X_t|X_{t-1}}(k|l) = P(X_t = k|X_{t-1} = l) = \sum_{j=0}^{\min\{k,l\}} g(j|l)P(e_t = k - j). \tag{4}$$

2.2. Modelling Censoring in CCID(1) Time Series

Given a model as in (1), a basic question is whether it properly describes all the observations of a given time series, or whether some observations have been affected by censoring. Here, we describe a model for dealing with censored observations in CCID(1) processes and study some of its properties.

Exogenous censoring can be modelled assuming (1) as a latent process and $Y_t = \min\{L, X_t\}$ as the observed process, where L is a constant that is assumed to be known. For simplicity of exposition we assume exogenous right censoring but all the results are easily extended to left-censoring or interval censoring. Hence, for right exogenous censoring

$$
\begin{aligned}
Y_t &= \min\{X_t, L\} = \begin{cases} X_t, & \text{if } X_t < L, \\ L, & \text{if } X_t \geq L, \end{cases} \\
X_t &= R_t(X_{t-1}) + e_t.
\end{aligned}
\tag{5}
$$

Although X_t, a CCID(1) process is Markovian, the exogenous censoring implies that Y_t is not Markovian because Y_t depends on X_t and L. Furthermore, Y_t is not CLAR (Conditionally Linear AutoRegressive). In fact,

$$
\begin{aligned}
E[Y_t|Y_{t-1} = y_{t-1}] &= E[Y_t|Y_{t-1} = y_{t-1}]\mathrm{I}_{\{y_{t-1}<L\}} + E[Y_t|Y_{t-1} = L]\mathrm{I}_{\{y_{t-1}=L\}} \\
&= \left(E[X_t|X_{t-1} = y_{t-1}] - \sum_{j=0}^{+\infty} jP[X_t = L + j|X_{t-1} = y_{t-1}] \right)\mathrm{I}_{\{y_{t-1}<L\}} \\
&\quad + \left(E[X_t|X_{t-1} \geq L] - \sum_{j=0}^{+\infty} jP[X_t = L + j|X_{t-1} \geq L] \right)\mathrm{I}_{\{y_{t-1}=L\}}
\end{aligned}
$$

The authors Zeger and Brookmeyer [1] established a procedure to obtain the likelihood of an observed time series under censoring, $\boldsymbol{y} = (Y_1, \ldots, Y_n)$, which becomes infeasible when the proportion of censoring is large. To overcome this issue, this work considers a Bayesian approach.

3. Bayesian Modelling

The Bayesian approach to the inference of an unknown parameter vector $\boldsymbol{\theta}$ is based on the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, defined as

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto L(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

where $L(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function of the observed data $\boldsymbol{y}$ and $\pi(\boldsymbol{\theta})$ is the prior distribution of the model parameters.

When the likelihood is computationally prohibitive or even impossible to handle, but it is feasible to simulate samples from the model (bypass the likelihood evaluation), as is the case of censored CCID(1) processes, Approximate Bayesian Computation (ABC) algorithms are an alternative. This methodology accepts the parameter draws that produce a match between the observed and the simulated sample, depending on a set of summary statistics, a chosen distance and a selected tolerance. The accepted parameters are then used

to estimate (an approximation of) the posterior distribution (conditioned on the summary statistics that afforded the match).

On the other hand, the idea of imputation arises naturally in the context of censored data. The Gibbs sampler with Data Augmentation (GDA) allows us to obtain an augmented dataset from the censored data by using a modified version of the Gibbs sampler, which samples not only the parameters of the model from its complete conditional but also the censored observations. The usual inference procedures may then be applied to the augmented data set.

3.1. Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) is based on an acceptance–rejection algorithm. ABC is used to compute a draw from an approximation of the posterior distributions, based on simulated data obtained from the assumed model in situations where its likelihood function is intractable or numerically difficult to handle. Summary statistics from the synthetic data are compared with the corresponding statistics from the observed data and a parameter draw is retained when there is a *match* (in some sense) between the simulated sample and the observed time series observation.

Recently, Ref. [16] provided the asymptotic results pertaining to the ABC posterior, such as Bayesian consistency and asymptotic distribution of the posterior mean.

Let $\mathbf{y^0} = (Y_1^0, \ldots, Y_n^0)$ be the fixed (observed) data and $\eta(\cdot)$ the model from which the data is generated. The most basic *approximate acceptance/rejection algorithm*, based on the works of [17,18], is as follows:

1. draw a value $\boldsymbol{\theta}$ from the prior distribution, $\pi(\boldsymbol{\theta})$,
2. simulate a sample $\mathbf{y} = (Y_1, \ldots, Y_n)$ from the model $\eta(.|\boldsymbol{\theta})$,
3. accept $\boldsymbol{\theta}$ if $d(S(\mathbf{y}), S_0)) \leq \delta$ for some distance measure $d(.,.)$ and some non-negative tolerance value δ, where $S(\cdot)$ is a summary statistic and $S_0 = S(\mathbf{y^0})$ is a fixed value.

It is well known that, if we use a proper distance measure, then as δ tends to zero, the distribution of the accepted values tends to the posterior distribution of the parameter given the data. When the summary statistics are sufficient for the parameter, then the distribution of the accepted values tends to the true posterior as δ tends to zero, assuming a proper distance measure on the space of sufficient statistics. The latent structure of the thinning operator means that the reduction to a sufficient set of statistics of dimension smaller than the sample size is not feasible and, therefore, informative summary statistics are often used [19].

In this work, given the characteristics of the data under study to compare the observed data ($\mathbf{y^0}$) and the synthetic (simulated) data ($\mathbf{y}$), we consider two distinctive characteristics of CCID(1) time series which are affected by the censoring: (i) the empirical marginal distribution and (ii) lag 1 auto-correlation.

To measure the similarity between the empirical marginal distributions the Kullback-Leibler (Note that Kullback-Leibler distance measures the dissimilarity between two probability distributions.) distance is calculated as

$$S_1(\mathbf{y}) = d_{KL}(\hat{p}^0, \hat{p}) = \sum_j \ln\left(\frac{\hat{p}_j^0}{\hat{p}_j}\right)\hat{p}_j^0, \tag{6}$$

where $\hat{p}_j^0$ and $\hat{p}_j$ denote the empirical marginal distribution of the observed time series and that of the simulated time series, respectively, estimated by the corresponding sample proportions, $\left(\hat{p}_j^0 = \frac{1}{n}\sum_{j=1}^n I_{\{Y_j^0=j\}} \text{ and } \hat{p}_j = \frac{1}{n}\sum_{j=1}^n I_{\{Y_j=j\}}\right)$. Whenever $\hat{p}_j^0$ is zero, the contribution of the jth term is interpreted as zero because $\lim_{p\to 0} p\ln(p) = 0$.

On the other hand, lag 1 sample autocorrelations, $S_2(\mathbf{y^0}) = \hat{\rho}_{Y^0}(1)$ and $S_2(\mathbf{y}) = \hat{\rho}_Y(1)$, are compared by their squared difference.

Additionally, we estimated the censoring rates, $S_3(\boldsymbol{y^0}) = \frac{1}{n}\sum_{t=1}^{n} I_{\{y_t^0 = L\}}$ and $S_3(\boldsymbol{y}) = \frac{1}{n}\sum_{t=1}^{n} I_{\{Y_t = L\}}$, which are also compared by their squared difference.

Thus, for each set of parameters, $\boldsymbol{\theta}^{(k)}$, a time series $\boldsymbol{x}^{(k)}$ is generated from the model CCID(1) and right censored at L, yielding $\boldsymbol{y}^{(k)} = (Y_1^{(k)}, \ldots, Y_n^{(k)})$ and the above statistics, $S_1(\boldsymbol{y}^{(k)})$, $S_2(\boldsymbol{y}^{(k)})$ and $S_3(\boldsymbol{y}^{(k)})$ are computed. Combining these statistics in a metric leading to the choice of the parameters $\boldsymbol{\theta}$ requires scaling. Thus, we propose the following metric

$$d_S^{(k)} = \frac{S_1(\boldsymbol{y}^{(k)})^2}{V(S_1(\boldsymbol{y}))} + \sum_{i=2}^{3} \frac{[S_i(\boldsymbol{y^0}) - S_i(\boldsymbol{y}^{(k)})]^2}{V(S_i(\boldsymbol{y^0}) - S_i(\boldsymbol{y}))} \qquad (7)$$

where $S_i(\boldsymbol{y^0})$ and $S_i(\boldsymbol{y}^{(k)})$ are the ith statistics obtained respectively from the observed and kth simulated data and $V(S_1(\boldsymbol{y}))$ and $V(S_i(\boldsymbol{y^0}) - S_i(\boldsymbol{y}))$ are the corresponding sample variances across the replications.

In summary, we propose Algorithm 1 for ABC approach based on [20]:

Algorithm 1 ABC for censored CCID(1)

For $k = 1, \ldots, N$
    Sample $\boldsymbol{\theta}^{(k)}$ from the prior distribution $\pi(\boldsymbol{\theta})$
    Generate a time series with n observations, $\boldsymbol{x}^{(k)}$ from the CCID(1) model
    Right truncate at L $\boldsymbol{x}^{(k)}$ to obtain the simulated data $\boldsymbol{y}^{(k)}$
    Compute $S_1(\boldsymbol{y}^{(k)})$, $S_2(\boldsymbol{y}^{(k)})$ and $S_3(\boldsymbol{y}^{(k)})$
Compute $d_S^{(k)} = \frac{S_1(\boldsymbol{y}^{(k)})^2}{V(S_1(\boldsymbol{y}))} + \sum_{i=2}^{3} \frac{[S_i(\boldsymbol{y^0}) - S_i(\boldsymbol{y}^{(k)})]^2}{V(S_i(\boldsymbol{y^0}) - S_i(\boldsymbol{y}))}$, $k = 1, \ldots, N$
Select the values $\boldsymbol{\theta}^{(k)}$ corresponding to the 0.1% quantile of $d_S^{(k)}$, $k = 1, \ldots, N$

Implementation issues regarding the prior distributions and the number of draws N for the CensPoINAR(1) model are addressed in Sections 3.3 and 4.1.

3.2. Gibbs Sampler with Data Augmentation

Gibbs sampling is a Markov chain Monte Carlo (MCMC) algorithm that can generate samples of the posterior distribution from their full conditional distributions [21]. When the data are under censoring or there are missing values, both cases leading to an incomplete data set, Ref. [22] proposed to combine the Gibbs sampler with data augmentation. This methodology implies imputing the censored (or missing) data, thus obtaining a complete dataset, and then dealing with the posterior of the complete data through the iterative Gibbs sampler. Therefore, the Gibbs sampler is modified in order to sample not only the parameters of the model from their complete conditionals but also the censored observations, obtaining an augmented (complete) dataset $\mathbf{z} = (z_1, \ldots, z_n)$ where

$$z_t = \begin{cases} Y_t, & \text{if } Y_t < L \\ z_t \sim F_\mu(x|x \geq L), & \text{if } Y_t = L \end{cases} \qquad (8)$$

where $F_\mu(x|x \geq L)$ is the truncated marginal distribution of the CCID(1) model with support in $[L, +\infty[$. Furthermore, we consider a modified sampling procedure for the imputation, designated as Mean of Multiple Simulation (MMS), proposed by [23] consisting in sampling from $F_\mu(x|x \geq L)$ multiple times, say m, and then imputing with the (nearest integer value) median of the m samples. This procedure is designated by **GDA-MMS** .

The augmented dataset can be considered as a CCID(1) time series and with a conditional likelihood function given by Equation (3). The posterior distribution of $\boldsymbol{\theta}$ is given by

$$p(\boldsymbol{\theta}|\boldsymbol{z}) \propto L(\boldsymbol{z}|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta}) \qquad (9)$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution of the parameters. In CCID(1) models the complexity of $p(\boldsymbol{\theta}|\boldsymbol{z})$ requires resorting to Markov Chain Monte Carlo (MCMC) techniques for sampling

from the full conditional distributions. The procedure is summarized in Algorithm 2 and detailed for the CensPoINAR(1) case in Sections 3.3 and 4.1.

Algorithm 2 GDA-MMS for censored CCID(1)

Initialize with $\boldsymbol{y} = (Y_1, \ldots, Y_n)$, $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \ldots, \theta_p^{(0)})$, $L \in \mathbb{R}$, and $n, \ m, \ N \in \mathbb{N}$
Set $\boldsymbol{z}^{(0)} = \boldsymbol{y}$
For $k = 1, \ldots, N$
    Sample $\theta_i^{(k)} \sim \pi(\theta_i | \boldsymbol{\theta}_{(-i)}^{(k-1)}, z^{(k-1)})$ ($\boldsymbol{x}_{(-i)}$ represents the vector $\boldsymbol{x}$ with the ith element removed.), $i = 1, \ldots, p$
    For $t = 1, \ldots, n$
        If $Y_t = L$
            For $j = 1, \ldots, m$
                Sample $z_t^{(j)} \sim F(x | \boldsymbol{\theta}^{(k)}, x \geq L)$
            $z_t^{(k)} := \frac{1}{m} \sum_{j=1}^m z_t^{(j)}$
        Else
            $z_t^{(k)} := Y_t$
Return $\boldsymbol{\theta} = [\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(N)}]'$ and $\boldsymbol{z}^{(N)}$.

3.3. The Particular Case of CensPoINAR(1)

This section details the ABC and GDA-MMS procedures to estimate a censored CCID(1) with the binomial thinning operation and Poisson marginal distribution, the censored Poisson INAR(1), CensPoINAR(1), model.

Consider the censored observations $\boldsymbol{y} = (Y_1, \ldots, Y_n)$ from a PoINAR(1) time series $\boldsymbol{x} = (X_1, \ldots, X_n)$ defined as

$$
\begin{aligned}
Y_t &= \min\{X_t, L\} = \begin{cases} X_t, & \text{if } X_t < L \\ L, & \text{if } X_t \geq L \end{cases} \\
X_t &= \alpha \circ X_{t-1} + e_t,
\end{aligned}
\tag{10}
$$

with $\alpha \circ X_{t-1} = \sum_{i=1}^{X_{t-1}} \xi_{ti}$, $\xi_{ti} \sim^{iid} Ber(\alpha)$, $e_t \sim Po(\lambda)$ and $X_t \sim Po(\frac{\lambda}{1-\alpha})$. Then $\boldsymbol{\theta} = (\alpha, \lambda)$ and given $\boldsymbol{x}$, the conditional likelihood function is given by

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \prod_{t=2}^n f_{X_t | X_{t-1}}(x_t | x_{t-1}) \\
&= \prod_{t=2}^n \sum_{j=0}^{\min\{x_t, x_{t-1}\}} \binom{x_{t-1}}{j} \alpha^j (1-\alpha)^{(x_{t-1}-j)} \frac{e^{-\lambda} \lambda^{x_t - j}}{(x_t - j)!}.
\end{aligned}
\tag{11}
$$

Under a Bayesian approach, we need a prior distribution for $\boldsymbol{\theta}$. In the absence of prior information, we use weakly informative prior distributions for $\boldsymbol{\theta}$ detailed below.

3.3.1. ABC for Censored PoINAR(1)

The ABC procedure described in Algorithm 1 is now implemented for the censored PoINAR(1). For the parameter $0 < \alpha < 1$, we choose a non-informative prior $U(0, 1)$, while for the positive parameter λ, we choose a non-informative $U(0, 10)$. The former allows us to explore all the support space for α. The choice of $U(0, 10)$ as a prior for $\lambda > 0$ allows us to explore a restricted support for the parameter that is in accordance with small counts.

3.3.2. GDA-MMS for Censored PoINAR(1)

Under the GDA-MMS approach, we first need to obtain a complete data set $\boldsymbol{z} = (z_1, \ldots, z_n)$ by imputing the censored observations, see (8). In this work, we draw $m = 10$ replicates of the right truncated at L Poisson distribution with parameter $\frac{\lambda}{1-\alpha}$, $w_i \sim \text{Po}\left(\frac{\lambda}{1-\alpha}\right) \times I_{(w_i \geq L)}$ and set $z_t = \lceil median(\boldsymbol{w}) \rceil$ ($\lceil c \rceil$ represents the integer ceiling of c),

$w = (w_1, \ldots, w_m)$, if $Y_t = L$. Figure 2 shows an augmented dataset (Z_t, black line) from the synthetic data presented in Figure 1.
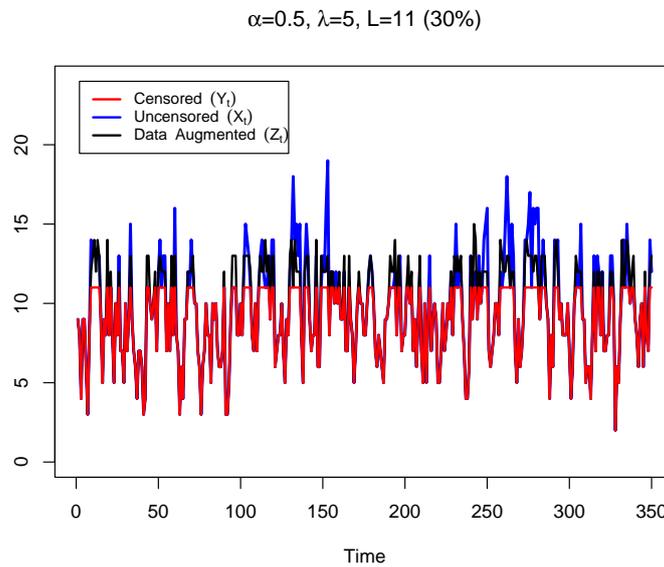
Figure 2. Synthetic dataset with $n = 350$ observations generated from a PoINAR(1) process with parameters $\alpha = 0.5$ and $\lambda = 5$ (X_t, blue line), the respective right-censored dataset (Y_t, red line), at $L = 11$, and an example of data augmentation (Z_t, black line).

As remarked above, given the complexity of the posterior distribution, Markov Chain Monte Carlo techniques are required for sampling from the full conditional distributions. Thus, the Adaptive Rejection Metropolis Sampling (ARMS) is used inside the Gibbs sampler [24]. Also in this approach, in the absence of prior information, we use weakly informative prior distributions for (α, λ). Thus, for the parameter $0 < \alpha < 1$, we choose a non-informative beta prior, conjugate of the binomial distribution, with parameters (a, b), while for the positive parameter μ, we choose a non-informative Gamma (shape, rate) prior, conjugate of the Poisson distribution, with parameters (c, d). The full conditional of λ is given by

$$p(\lambda|\alpha, z) = \frac{p(\lambda, \alpha|z)}{p(\alpha|z)} \propto \exp[-(d + (n-1))\lambda]\lambda^{c-1} \prod_{t=2}^{n} \sum_{i=0}^{\min\{z_t, z_{t-1}\}} C(t, i)\lambda^{(z_t)-i}, \quad (12)$$

where

$$C(t, i) = \frac{1}{((z_t) - i)!}\binom{z_{t-1}}{i}\alpha^i(1 - \alpha)^{(z_{t-1})-i} \quad \text{and} \quad \lambda > 0.$$

The full conditional distribution of α is given by

$$p(\alpha|\lambda, z) = \frac{p(\lambda, \alpha|z)}{p(\lambda|z)} \propto \alpha^{a-1}(1 - \alpha)^{b-1} \prod_{t=2}^{n} \sum_{i=0}^{\min\{z_t, z_{t-1}\}} K(t, i)\alpha^i(1 - \alpha)^{(z_{t-1})-i}, \quad (13)$$

where

$$K(t, i) = \frac{\lambda^{(z_t)-i}}{((z_t) - i)!}\binom{X_{t-1}}{i} \quad 0 < \alpha < 1.$$

The parameters α and λ are computed as the posterior mean.

The GDA-MMS procedure to estimate a censored PoINAR(1) process is detailed in Algorithm 3.

Algorithm 3 GDA-MMS for CensPoINAR(1)

Initialize with $\boldsymbol{y}$, $\boldsymbol{\theta}^{(0)} = (\alpha^{(0)}, \lambda^{(0)})$, $L \in \mathbb{R}$, and N, $m \in \mathbb{N}$

Set $\boldsymbol{z}^{(0)} = \boldsymbol{y}$

For $k = 1, ..., N$

    Using ARMS

        Sample $\lambda^{(k)} \sim p(\lambda | \alpha^{(k-1)}, \boldsymbol{z}^{(k-1)})$

        Sample $\alpha^{(k)} \sim p(\alpha | \lambda^{(k)}, \boldsymbol{z}^{(k-1)})$

    For $t = 1, ..., n$

        If $Y_t = L$

            For $j = 1, ..., m$

                Sample $w^{(j)} \sim \text{Po}\left(\frac{\lambda^{(k)}}{1-\alpha^{(k)}}\right) \times I_{(w^{(j)} \geq L)}$

            $z_t^{(k)} := \lceil median(\boldsymbol{w}) \rceil$, $\boldsymbol{w} = (w^{(1)}, ..., w^{(m)})$,

        Else

            $z_t^{(k)} := Y_t$

Return $\boldsymbol{\theta} = [\boldsymbol{\theta}^{(1)}, ..., \boldsymbol{\theta}^{(N)}]'$ and $\boldsymbol{z}^{(N)}$.

---

4. Illustration

This section illustrates the procedures proposed above to model CCID(1) right-censored time series in the particular case of Poisson distribution and binomial thinning operation.

4.1. Illustration with CensPoINAR(1)

In this section, the performance of the Bayesian approaches previously proposed is illustrated via synthetic data. Thus, realizations with $n = 100, 350, 1000$ observations of CensPoINAR(1) models were simulated, with parameters $\theta = (0.2, 3)$ and $\theta = (0.5, 5)$, considering for each case two levels of censorship, namely 30% and 5%.

For the ABC estimates, we run $N = 10^6$ replications and choose the pairs (α, λ) corresponding to the 0.1% lower quantile of $d_S^{(k)}$, Equation (7), in total of 1000 values from which the estimates are computed as the mean value. Software R [25] was used to implement the ABC algorithm.

To implement GDA-MMS algorithm we consider the initial values $\boldsymbol{\theta}^{(0)} = (\alpha^{(0)}, \lambda^{(0)})$ given by the Conditional Least squares estimates of α and λ [24]. The hyper-parameters for the prior distributions of α and λ are the following $\alpha \sim Beta(a = 2, b = 2)$ and $\lambda \sim Gamma(c = 0.1, d = 0.1)$. In this work, the function *armspp* was used from the package *armspp* [26] in R to sample from the full conditional distributions. Several experiments were carried out to analyse the size that the chain should have in order to be stable and, thus, the number of Gibbs sampler iterations used in this work is $N = 15,000$. Among these, we ignored the first 5000 simulations as burn-in time and, to reduce autocorrelation between MCMC observations, we considered only simulations from every 30 iterations. Therefore, we use a simulated sample with size 323 to obtain the Bayesian estimates. A convergence analysis with the usual diagnostic tests was performed with the package *coda* [27] in R [25].

Tables 2 and 3 summarize ABC and GDA-MMS results for the several scenarios described above: point estimates, $\hat{\alpha}$ and $\hat{\lambda}$, obtained as sample means, the corresponding bias, standard deviation and the coefficient of variation. The results indicate that the bias tends to decrease for large sample sizes and small censoring rates. The results also indicate that overall ABC presents estimates with smaller bias but larger variability when compared with GDA-MMS.

Table 2. ABC and GDA-MMS results for parameter α (sample mean, and the corresponding bias, standard deviation and coefficient of variation) for synthetic data generated from CensPoINAR(1) models.

|				ABC				GDA-MMS			
α	λ	L	n	$\bar{\bar{\alpha}}$	Bias($\hat{\alpha}$)	s.d.($\hat{\alpha}$)	CV($\hat{\alpha}$)	$\bar{\bar{\alpha}}$	Bias($\hat{\alpha}$)	s.d.($\hat{\alpha}$)	CV($\hat{\alpha}$)
0.2	3	4 (30%)	100	0.2571	0.0571	0.0911	0.3544	0.3155	0.1155	0.0323	0.1024
			350	0.2067	0.0067	0.0579	0.2803	0.2274	0.0274	0.0178	0.0783
			1000	0.1793	−0.0207	0.0398	0.2217	0.2025	0.0025	0.0157	0.0775
0.2	3	6 (5%)	100	0.2268	0.0268	0.0760	0.3350	0.2738	0.0738	0.0270	0.0986
			350	0.2302	0.0302	0.0511	0.2221	0.2309	0.0309	0.0140	0.0606
			1000	0.1931	−0.0069	0.0327	0.1692	0.1915	−0.0085	0.0112	0.0585
0.5	5	11 (30%)	100	0.5304	0.0304	0.0800	0.1508	0.5596	0.0596	0.0170	0.0304
			350	0.4637	−0.0363	0.0535	0.1153	0.4834	−0.0166	0.0124	0.0257
			1000	0.5115	0.0115	0.0320	0.0626	0.5050	0.0050	0.0072	0.0143
0.5	5	14 (5%)	100	0.5230	0.0230	0.0815	0.1559	0.5363	0.0363	0.0175	0.0326
			350	0.4671	−0.0329	0.0461	0.0987	0.4796	−0.0204	0.0107	0.0223
			1000	0.4992	−0.0008	0.0291	0.0584	0.5008	0.0008	0.0070	0.0140

Table 3. ABC and GDA-MMS results for the parameter λ (sample mean, and the corresponding bias, standard deviation and coefficient of variation) for synthetic data generated from CensPoINAR(1) models.

				ABC				GDA-MMS			
α	λ	L	n	$\bar{\bar{\lambda}}$	Bias($\hat{\lambda}$)	s.d.($\hat{\lambda}$)	CV($\hat{\lambda}$)	$\bar{\bar{\lambda}}$	Bias($\hat{\lambda}$)	s.d.($\hat{\lambda}$)	CV($\hat{\lambda}$)
0.2	3	4 (30%)	100	2.6623	−0.3377	0.3699	0.1389	2.3265	−0.6735	0.1144	0.0492
			350	2.8530	−0.1470	0.2353	0.0825	2.6639	−0.3361	0.0672	0.0252
			1000	3.1398	0.1398	0.1668	0.0531	2.9757	−0.0243	0.0603	0.0203
0.2	3	6 (5%)	100	2.7918	−0.2082	0.3203	0.1147	2.5719	−0.4281	0.1007	0.0392
			350	2.9507	−0.0493	0.2173	0.0736	2.7846	−0.2154	0.0579	0.0208
			1000	3.1342	0.1342	0.1448	0.0462	3.0417	0.0417	0.0460	0.0151
0.5	5	11 (30%)	100	4.3432	−0.6568	0.7504	0.1728	3.9528	−1.0472	0.1600	0.0405
			350	5.2315	0.2315	0.5265	0.1006	4.9073	−0.0927	0.1177	0.0240
			1000	4.9102	−0.0898	0.3247	0.0661	4.8974	−0.1026	0.0720	0.0147
0.5	5	14 (5%)	100	4.4488	−0.5512	0.7828	0.1760	4.2574	−0.7426	0.1682	0.0395
			350	5.1333	0.1333	0.4286	0.0835	4.9877	-0.0123	0.1088	0.0218
			1000	5.0613	0.0613	0.2826	0.0558	4.9964	−0.0036	0.0708	0.0142

Additionally, Figures 3 and 4 represent the corresponding posterior densities. The plots show unimodal and approximately symmetric distributions, with a dispersion that clearly decreases with increasing sample size and smaller censoring rate. The posterior densities indicate that the ABC approach produces posteriors that are flatter but with modes very close to the true value, while the corresponding GDA-MMS approach, despite producing posteriors which are more concentrated also evidence higher bias. However, the behaviour of GDA-MMS estimates varies with the parameters and even the sample sizes. These results are representative of the properties of GDA-MMS estimates across a large number of experiments, not reported here for conciseness.
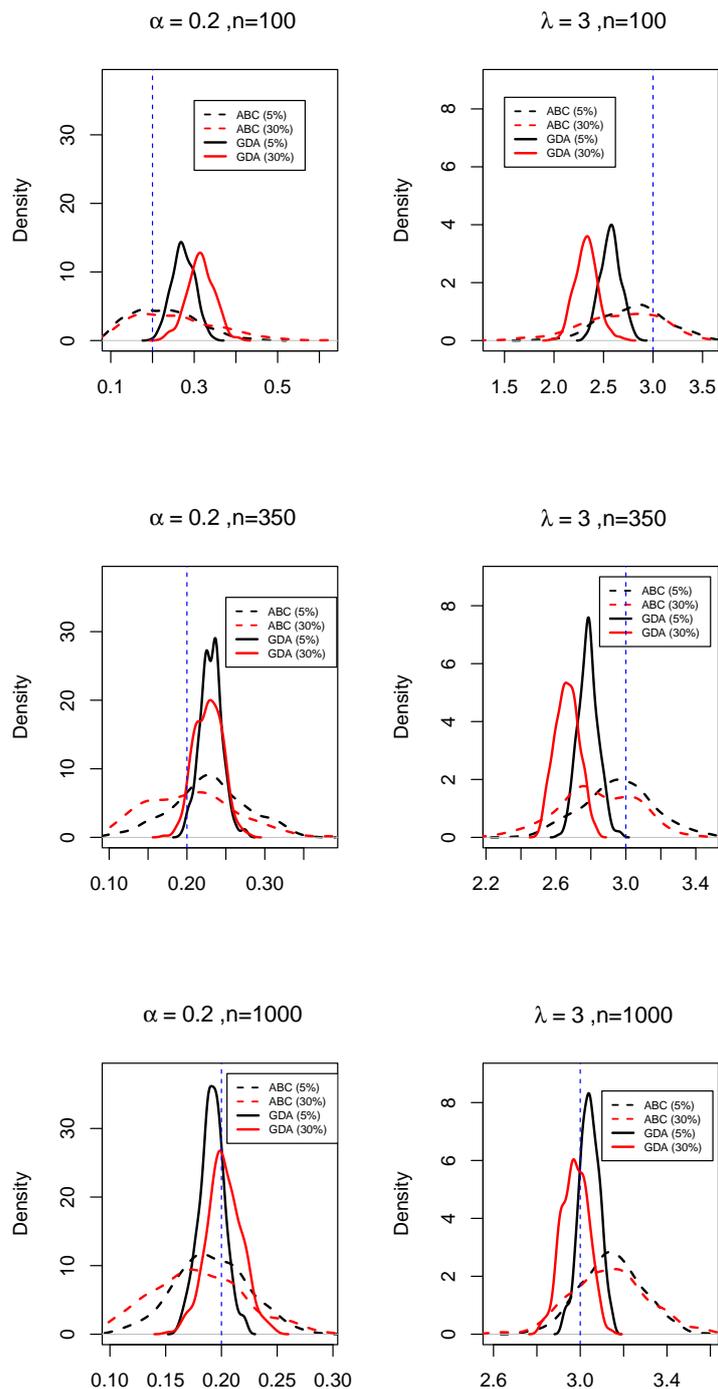
Figure 3. ABC and GDA-MMS posterior densities of the parameters for a realization of 100, 350 and 1000 observations of a CensPoINAR(1) model with $\theta = (0.2, 3)$, considering two levels of censoring. Note that the scale of *x*-axis of the six plots are different.
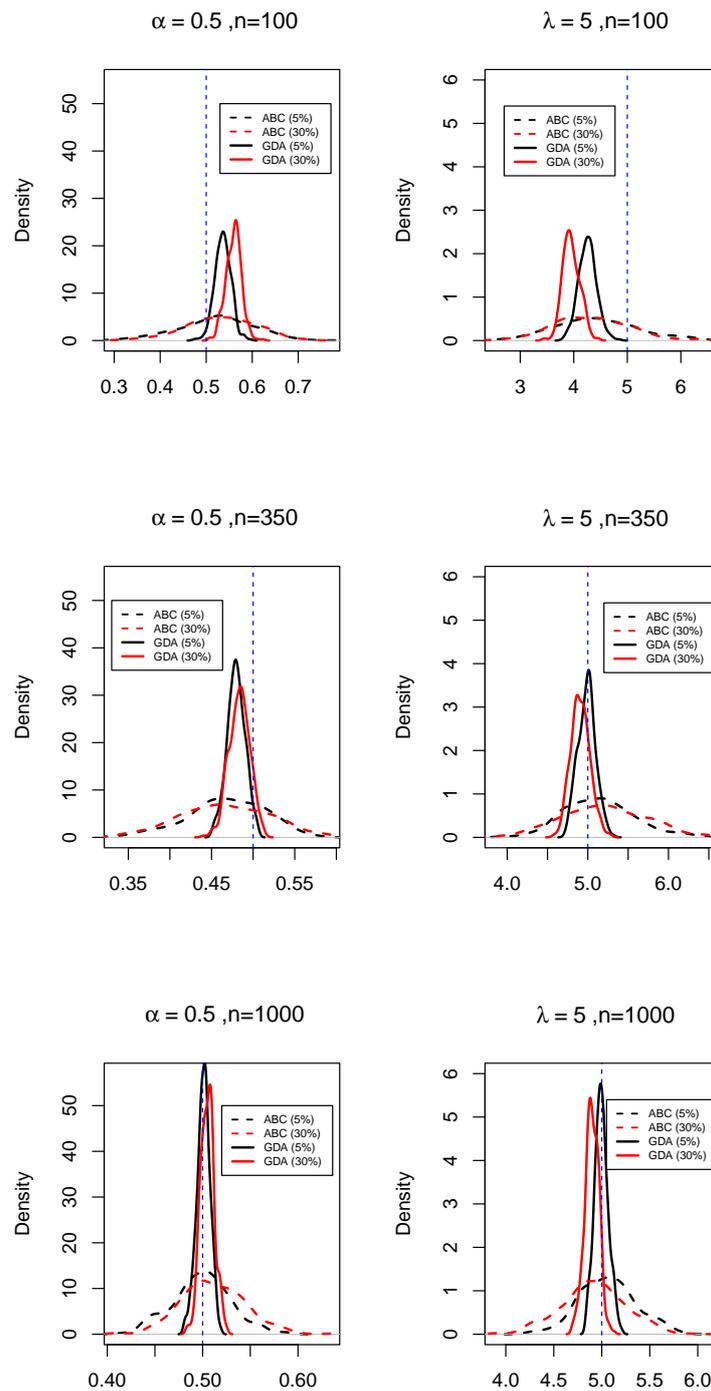
Figure 4. ABC and GDA-MMS posterior densities of the parameters for a realization of 100, 350 and 1000 observations of a CensPoINAR(1) model with $\theta = (0.5, 5)$, considering two levels of censoring. Note that the scale of *x*-axis of the six plots are different.

4.2. Simulation Study for GDA-MMS

This section presents the results of a simulation study designed to further analyse the sample properties of GDA-MMS, in particular the bias of the resulting Bayesian estimates.

For that purpose, realizations with sample sizes $n = 100$ and $n = 350$ of CensPoINAR(1) models with parameters $\theta = (0.2, 3)$ and $\theta = (0.5, 5)$, are generated, considering two levels of censorship, namely 30% and 5%. To analyse the performance of the

procedure, the sample posterior mean, standard deviation and mean squared error were calculated over 50 repetitions.

Boxplots of the sample bias for the 50 repetitions of GDA-MMS methodology are presented in Figures 5 and 6. The bias increases with the rate of censoring and the variability decreases with the sample size. Furthermore, in general, the estimates for α presents positive sample mean biases, indicating that α is overestimated, whilst the estimates for λ shows negative sample biases, indicating underestimation for λ. Both bias and dispersion seem larger for λ.
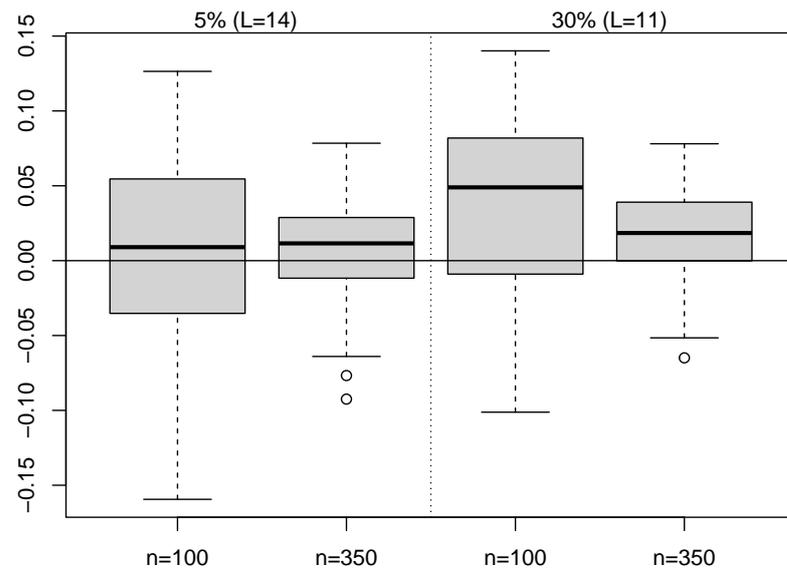
Figure 5. Boxplots of bias for GDA-MMS estimates of α, when $\theta = (0.5, 5)$.
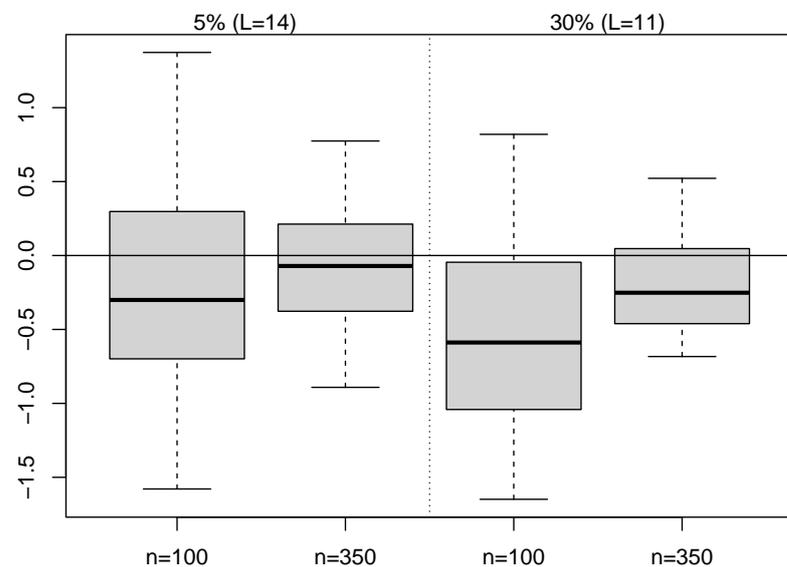

Figure 6. Boxplots of bias for GDA-MMS estimates of λ, when $\theta = (0.5, 5)$.

Tables 4 and 5 present the sample posterior measures for $\hat{\alpha}$ and $\hat{\lambda}$, respectively. We can see improvement of the estimation methods performance when the sample size increases. Additionally, the higher the censoring percentage, the worse the behavior of the proposed methods.

Table 4. Sample posterior mean, standard errors (in brackets) and root mean square error for GDA-MMS estimates of α.

α	λ	L	n	$\bar{\bar{\alpha}}$ (s.e.($\hat{\alpha}$))	**RMSE($\hat{\alpha}$)**
0.2	3	4 (30%)	100	0.2918 (0.0977)	0.1341
			350	0.2385 (0.0698)	0.0797
0.2	3	6 (5%)	100	0.2739 (0.0680)	0.1004
			350	0.2229 (0.0487)	0.0538
0.5	5	11 (30%)	100	0.5404 (0.0632)	0.0750
			350	0.5156 (0.0344)	0.0378
0.5	5	14 (5%)	100	0.5142 (0.0626)	0.0642
			350	0.5066 (0.0386)	0.0392

Table 5. Sample posterior mean, standard errors (in brackets) and root mean square error for GDA-MMS estimates of λ.

α	λ	L	n	$\bar{\bar{\lambda}}$ (s.e.($\hat{\lambda}$))	**RMSE($\hat{\lambda}$)**
0.2	3	4 (30%)	100	2.5283 (0.3077)	0.5632
			350	2.7842 (0.2462)	0.3274
0.2	3	6 (5%)	100	2.6814 (0.2984)	0.4365
			350	2.8934 (0.1843)	0.2129
0.5	5	11 (30%)	100	4.4861 (0.6710)	0.8452
			350	4.7976 (0.3357)	0.3920
0.5	5	14 (5%)	100	4.7593 (0.6391)	0.6829
			350	4.9229 (0.4177)	0.4248

5. Final Comments

This work approaches the problem of estimating CCID(1) models for time series of counts under censoring from a Bayesian perspective. Two algorithms are proposed: one is based on ABC methodology and the second a Gibbs Data Augmentation modified with multiple sampling. Experiments with synthetic data allow us to conclude that both approaches lead to estimates that present less bias than those obtained neglecting the censoring. Moreover, the GDA-MMS approach allows us to obtain a complete data set, making it a valuable method in other situations such as missing data.

In this study, we focus on the most popular CCID(1) model, the Poisson INAR(1). However, if the data under study present over- or under-dispersion, other CCID(1) models with appropriate distributions for the innovations, such as Generalised Poisson or Negative Binomial, can easily be entertained. Furthermore, one can consider different models for time series of counts under censoring, based on INGARCH models, ([28,29] using a switching mechanism) if they are more suitable to the data set to be modeled. These issues are beyond the scope of this paper and are a topic for a future research project.

References

1. Zeger, S.; Brookmeyer, R. Regression Analysis with Censored Autocorrelated Data. *J. Am. Stat. Assoc.* **1986**, *81*, 722–729. [CrossRef]
2. Hopke, P.K.; Liu, C.; Rubin, D.R. Multiple imputation for multivariate data with missing and below-threshold measurements: Time-series concentrations of pollutants in the arctic. *Biometrics* **2001**, *57*, 22–33. [CrossRef] [PubMed]
3. Park, J.W.; Genton, M.G.; Ghosh, S.K. Censored time series analysis with autoregressive moving average models. *Can. J. Stat.* **2007**, *35*, 151–168. [CrossRef]
4. Mohammad, N.M. Censored Time Series Analysis. Ph.D. Thesis, The University of Western Ontario, London, ON, Canada, 2014.
5. Schumacher, F.L.; Lachos, V.H.; Dey, D.K. Censored regression models with autoregressive errors: A likelihood-based perspective. *Can. J. Stat.* **2017**, *45*, 375–392. [CrossRef]
6. Wang, C.; Chan, K.S. Carx: An R Package to Estimate Censored Autoregressive Time Series with Exogenous Covariates. *R J.* **2017**, *9*, 213–231. [CrossRef]
7. Wang, C.; Chan, K.S. Quasi-Likelihood Estimation of a Censored Autoregressive Model with Exogenous Variables. *J. Am. Stat. Assoc.* **2018**, *113*, 1135–1145. [CrossRef]
8. Fernández-Fontelo, A.; Cabaña, A.; Puig, P.; Moriña, D. Under-reported data analysis with INAR-hidden Markov chains. *Stat. Med.* **2016**, *35*, 4875–4890. [CrossRef]
9. De Oliveira, G.L.; Loschi, R.H.; Assunção, R.M. A random-censoring Poisson model for underreported data. *Stat. Med.* **2017**, *36*, 4873–4892. [CrossRef]
10. Joe, H. Time series model with univariate margins in the convolution-closed infinitely divisible class. *Appl. Probab.* **1996**, *33*, 664–677. [CrossRef]
11. Al-Osh, M.A.; Alzaid, A.A. First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **1987**, *8*, 261–275. [CrossRef]
12. McKenzie, E. Some ARMA models for dependent sequences of Poisson counts. *Adv. Appl. Probab.* **1988**, *20*, 822–835. [CrossRef]
13. Alzaid, A.A.; Al-Osh, M.A. Some autoregressive moving average processes with generalized Poisson marginal distributions. *Ann. Inst. Stat. Math.* **1993**, *45*, 223–232. [CrossRef]
14. Han, L.; McCabe, B. Testing for parameter constancy in non-Gaussian time series. *J. Time Ser. Anal.* **2013**, *34*, 17–29. [CrossRef]
15. Jung, R.C.; Tremayne, A.R. Useful models for time series of counts or simply wrong ones? *AStA Adv. Stat. Anal.* **2011**, *95*, 59–91. [CrossRef]
16. Frazier, D.T.; Martin, G.M.; Robert, C.P.; Rousseau, J. Asymptotic properties of approximate Bayesian computation. *Biometrika* **2018**, *105*, 593–607. [CrossRef]
17. Plagnol, V.; Tavaré, S. Approximate Bayesian computation and MCMC. In *Monte Carlo and Quasi-Monte Carlo Methods*; Niederreiter, H., Ed.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 99–113.
18. Wilkinson, R.D. Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. *Stat. Appl. Genet. Mol. Biol.* **2013**, *12*, 129–141. [CrossRef]
19. Frazier, D.T.; Maneesoonthorn, W.; Martin, G.M.; McCabe, B. Approximate Bayesian forecasting. *Int. J. Forecast.* **2019**, *35*, 521–539. [CrossRef]
20. Biau, G.; Cérou, F.; Guyader, A. New Insights into Approximate Bayesian Computation. *Ann. Henri Poincaré (B) Probab. Stat.* **2015**, *51*, 376–403. [CrossRef]
21. Gelfand, A.; Smith, A. Sampling-based approaches to calculating marginal densities. *J. R. Stat. Soc. Ser. B* **1990**, *85*, 398–409. [CrossRef]
22. Chib, S. Bayes Inference in the Tobit Censored Regression Model. *J. Econom.* **1992**, *51*, 79–99. [CrossRef]
23. Sousa, R.; Pereira, I.; Silva, M.E.; McCabe, B. Censored Regression with Serially Correlated Errors: A Bayesian approach. *arXiv* **2023**, arXiv:2301.01852.
24. Silva, I.; Silva, M.E.; Pereira, I.; Silva, N. Replicated INAR(1) Processes. *Methodol. Comput. Appl. Probab.* **2005**, *7*, 517–542. [CrossRef]
25. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022. Available online: https://www.R-project.org/ (accessed on 20 March 2023).
26. Bertolacci, M. Armspp: Adaptive Rejection Metropolis Sampling (ARMS) via 'Rcpp'. R Package Version 0.0.2. 2019. Available online: https://CRAN.R-project.org/package=armspp (accessed on 20 March 2023).
27. Plummer, M.; Best, N.; Cowles, K.; Vines, K. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* **2006**, *6*, 7–11.
28. Chen, C.W.S.; Lee, S. Generalized Poisson autoregressive models for time series of counts. *Comput. Stat. Data Anal.* **2016**, *99*, 51–67. [CrossRef]
29. Ferland, R.; Latour, A.; Oraichi, D. Integer-Valued GARCH Process. *J. Time Ser. Anal.* **2006**, *27*, 923–942. [CrossRef]