Bayesian Analysis of Change Point Problems Using Conditionally Specified Priors

G. Shahtahmassebi[1] · **José María Sarabia[2]**

Abstract

In data analysis, change point problems correspond to abrupt changes in stochastic mechanisms generating data. The detection of change points is a relevant problem in the analysis and prediction of time series. In this paper, we consider a class of conjugate prior distributions obtained from conditional specification methodology for solving this problem. We illustrate the application of such distributions in Bayesian change point detection analysis with Poisson processes. We obtain the posterior distribution of model parameters using general bivariate distribution with gamma conditionals. Simulation from the posterior are readily implemented using a Gibbs sampling algorithm. The Gibbs sampling is implemented even when using conditional densities that are incompatible or only compatible with an improper joint density. The application of such methods will be demonstrated using examples of simulated and real data.

Keywords Conditional specification · Change point · Bayesian analysis · Gamma conditional distributions

1 Introduction

Data science, as an interdisciplinary field, plays a pivotal role in extracting knowledge and insights from complex and large-scale datasets. It encompasses a wide range of techniques, methodologies, and tools to analyse and interpret data, enabling evidence-based decision making and driving innovation across various domains [1–4].
Data science involves the integration of statistical analysis, machine learning, data mining, and visualization techniques to extract actionable insights from data. It encompasses the entire data lifecycle, including data collection, cleaning, integra-

✉ G. Shahtahmassebi
golnaz.shahtahmassebi@ntu.ac.uk

1   School of Science and Technology, Nottingham Trent University, Nottingham, UK

2   Department de Métodos Cuantitativos, CUNEF Universidad, Madrid, Spain

tion, transformation, modeling, and interpretation. By applying rigorous statistical methods and computational algorithms, Data Scientists can uncover patterns, trends, and correlations within data, enabling organizations to make data-driven decisions and gain a competitive advantage [5, 6].

The rise of big data has presented both opportunities and challenges in the field of Data Science. On one hand, Big Data offers vast amounts of information that can uncover valuable insights and patterns. On the other hand, the sheer volume, velocity, and variety of data pose significant computational and analytical challenges, which means traditional statistical methods may not be scalable or efficient enough to handle such data [7–10].

The application of statistical methods in data science provides the foundation for understanding data patterns, making inferences, and predicting outcomes. By using statistical techniques, meaningful insights, robust models, and informed decisions based on data-driven evidence can be extracted [6, 11–14].

Within the scope of data science and statistics, change point detection serves as a pivotal tool for identifying shifts or transitions in data patterns. By effectively detecting change points, Data Scientists can gain profound insights into anomalies, events, or structural transformations embedded within the data [15].

The application of change point detection methods, including Bayesian approaches, holds significant importance in data science. Bayesian change point detection employs Bayesian inference principles to estimate the posterior probability of change points by incorporating prior knowledge and updating it based on observed data. Such a framework offers a flexible and robust approach to detecting change points, and its applicability extends across diverse domains such as genetics, environmental monitoring, finance, and quality control.

Change points are defined as abrupt variations in the generative parameters of a data sequence. The detection of change points is a relevant problem in the analysis and prediction of time series. The application can be found in copy number variation detection [16], air quality control [17] and signal processing [18].

Among the methods available, the Bayesian approach is particularly appealing because it automatically captures the trade-off between model complexity (number of change points) and model fit. It also allows one to express uncertainty about the number and location of change points.

In this paper, we propose a novel approach for detecting change points in a Poisson process. We introduce conditionally specified priors and derive the posterior distribution of model parameters using the general bivariate distribution with gamma conditionals. Such prior distribution have also been used in the comparison of gamma scale parameters [19] and the estimation of incomplete data [20]. Although these prior distributions have been extensively used in Bayesian analysis for classical distributions such as the normal distribution, Pareto distribution, and linear regression [21], their application in the context of Poisson processes is novel.

The remainder of this is organised as follows. Section 2 includes a short review on change point techniques and the motivation. Section 3 provides design of our proposed model. Section 4 we introduce the general bivariate distribution with gamma conditionals as the prior distribution for the parameters of a Poisson change point model and discuss the full Bayesian analysis of the change point problem. The appli-

cation of our methodology is illustrated using simulated and real data sets in Sect. 5. Finally, Sect. 6 concludes the paper and suggests future extensions to our proposed approach.

2 Time Series Change Point Detection Techniques

Analysing time series data is crucial in various domains such as finance, economics, environmental science, and healthcare as it allows for the detection of underlying patterns, trends, and anomalies. In this context, change point detection techniques play a vital role in identifying points in a time series where the statistical properties significantly change.

Change point detection methods aim to pinpoint abrupt or gradual shifts in the underlying characteristics of a time series, which may indicate changes in mean, variance, distribution, or other structural properties. Detecting change points is essential for understanding system behavior, forecasting future trends, and detecting unusual events or anomalies.

Numerous change point detection methods have been developed, each with its strengths and assumptions. Classic techniques include cumulative sum methods, likelihood ratio tests, and sequential analysis methods. More advanced approaches leverage Bayesian inference, nonparametric methods, and machine learning algorithms to detect change points in complex and high-dimensional time series data.

Chen and Gupta [22, 23] conducted a detailed study of parametric models for change point analysis, exploring the univariate and multivariate normal cases, regression models, exponential-type models, and models with discrete data, from both classical and Bayesian perspectives. They also presented applications in genetics, medicine, and finance.

The use of Bayesian techniques in these types of models is extensively studied in [24]. One popular Bayesian change point detection method is the Bayesian Online Change point Detection algorithm proposed by [25], which provides a probabilistic framework for online change point detection and has been widely applied in signal processing, finance, and environmental monitoring. Bootstrap techniques have also been used by various authors to estimate these types of problems, as highlighted by [26].

Modern machine learning approaches, such as deep learning models, have shown promise in change point detection. For instance, [27] present a Long Short-Term Memory based model for detecting change points in multivariate time series data, demonstrating its effectiveness in identifying shifts in complex temporal patterns. In another study [28] reviewed models for the detection of change points in time series, including supervised and unsupervised algorithms proposed in the literature, and propose suitable criteria for comparing these algorithms.

In time series trend analysis, [29] propose a solution for change point detection, assessing four methods: wild binary segmentation, E-agglomerative algorithms, iterative robust detection methods, and Bayesian analysis.

In this study, we propose a Bayesian approach for detecting change points in a Poisson process. Unlike previous studies, we assume that the parameters of the Poisson process

before and after the change are jointly distributed. Therefore, we propose a joint conjugate prior distribution based on the conditional specification methodology and introduce the application of conditionally specified prior distributions in change point detection problems. The general aspects of conditional specification can be found in [30]. The following section provides the details of obtaining the proposed prior distribution.

3 Motivation and Design of the Model

The general ideal of the conditional specification methodology can be described as follows. Suppose x_i, for $i = 1, \ldots, n$, be n independently distributed random variables with probability distribution functions of $f(\theta_i)$, where $\theta_i \in \Theta \subset \mathbb{R}^n$ and let likelihood function be denoted as

$$f(x_1, \ldots, x_n | \theta_1, \ldots, \theta_n).$$

Our interest lies in specifying a class of conjugate prior distributions such that the conditional distribution of $\theta_i | \theta_{-i}$ belongs to the family class of $f(\theta_i | \theta_{-i})$, where $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$.

We consider a candidate family of prior distributions for $\theta_1, \ldots, \theta_n$ such that, for each i, the conditional density of θ_i given θ_{-i} belongs to the family f_i. We will show that the posterior distribution will be in the same family and will also be conditionally specified, resulting in a broad and flexible conjugate prior family for $\theta_1, \ldots, \theta_n$.

Simulation from the posterior will be readily implemented using a Gibbs sampling algorithm. The Gibbs sampling can be implemented even when using conditional densities that are incompatible or only compatible with an improper joint density [31].

Now, let X_i be a Poisson random variable where $i = 1, \ldots, T$. Assume that the change point occurs at time k such that

$$x_i \sim \mathcal{P}o(\lambda), \quad i = 1, 2, \ldots, k,$$

and

$$x_i \sim \mathcal{P}o(\alpha), \quad i = k+1, \ldots, n,$$

where $\lambda > 0, \alpha > 0$ and $k \in \{1, 2, \ldots, n\}$ are unknown parameters and $X \sim \mathcal{P}o(\lambda)$ denotes a Poisson distribution with parameter λ. In can be seen if $k = n$, then $\lambda = \alpha$ and we do not have change. Thus, the likelihood function takes the form,

$$L(\lambda, \alpha, k) \propto \lambda^{\sum_{i=1}^{k} x_i} \alpha^{\sum_{i=k+1}^{n} x_i} \exp\left(-k\lambda - (n - k + 1)\alpha\right). \tag{1}$$

Our goal is to obtain a conjugate prior for λ and α in a way that if α is known, a conjugate prior for λ is a gamma distribution, and if λ is known, the conjugate

prior for α is again a gamma distribution. Consequently, we have conjugate prior distributions for each parameter conditioned on the other parameter.

In order to obtain a bivariate conjugate distribution for λ and α, we will have to obtain the most general bivariate distribution whose conditional distributions are gamma. The class of distributions whose conditionals are gamma has been obtained in [30]. Let us assume the likelihood in Eq. (1) only depends on parameters α and λ. We assume parameter k is known. According to the Theorem 1 in [32], the most general bivariate distribution with all conditionals of λ given α and, of α given λ, being gamma distribution has the following form

$$\pi(\lambda, \alpha; \boldsymbol{m}) \propto (\lambda\alpha)^{-1} \exp[-m_{10}\lambda - m_{01}\alpha + m_{20}\log\lambda + m_{02}\log\alpha$$
$$+ m_{11}\lambda\alpha - m_{12}\lambda\log\alpha - m_{21}\alpha\log\lambda + m_{22}\log\lambda\log\alpha]. \quad (2)$$

which is a conjugate prior for the family of likelihood given in (1). A bivariate distribution with joint pdf (2) will known as (general bivariate distribution with gamma conditionals) which is denoted by $(\lambda, \alpha) \sim \mathcal{GBGC}(\boldsymbol{m})$.

As can be seen in (2), the new class of distributions depends on eight parameters with the following conditional distributions are

$$\lambda|\alpha \sim \mathcal{Ga}(a_1(\alpha), p_1(\alpha)), \quad (3)$$
$$\alpha|\lambda \sim \mathcal{Ga}(a_2(\lambda), p_1(\lambda)). \quad (4)$$

It can be seen that the conditional distributions of the proposed conjugate prior also conjugate distributions for the parameters λ and α. Moreover, it is important to note that the model incorporates informative priors, allowing the incorporation of prior knowledge about the parameters. Lastly, due to the specific construction of the model, both conditional distributions follow a gamma distribution, which facilitates their estimation using Gibbs sampling algorithm. The conditional parameters in (3)–(4) are given by,

$$\begin{aligned}
a_1(\alpha) &= m_{20} - m_{21}\alpha + m_{22}\log\alpha, \\
p_1(\alpha) &= m_{10} - m_{11}\alpha + m_{12}\log\alpha, \\
a_2(\lambda) &= m_{02} - m_{12}\lambda + m_{22}\log\lambda, \\
p_2(\lambda) &= m_{01} - m_{11}\lambda + m_{21}\log\lambda.
\end{aligned} \quad (5)$$

In order to ensure that this prior distribution is proper, the hyperparameters must satisfy the following list of constraints,

$$\begin{aligned}
m_{11} &\leq 0, \\
m_{12} &\leq 0, \\
m_{21} &\leq 0, \\
m_{22} &\leq 0, \\
m_{10} &> m_{12}[1 - \log(m_{12}/m_{11})], \\
m_{20} &> m_{22}[1 - \log(m_{22}/m_{21})], \\
m_{01} &> m_{21}[1 - \log(m_{21}/m_{11})], \\
m_{02} &> m_{22}[1 - \log(m_{22}/m_{12})].
\end{aligned} \quad (6)$$

Note that for a gamma random variable Y with the probability function,

$$f(x; a, p) = \frac{p^a x^{a-1} e^{-px}}{\Gamma(a)}, \quad x \geq 0. \tag{7}$$

the expected value and variance are $E(Y) = \frac{a}{p}$ and $var(Y) = \frac{a}{p^2}$, respectively.

3.1 A Relevant Submodel

Consider the bivariate continuous distribution with joint pdf,

$$f(\lambda, \alpha) = \frac{k_{r,s}(\phi) m_1^r m_2^s}{\Gamma(r)\Gamma(s)} \lambda^{r-1} \alpha^{s-1} e^{-m_1\lambda - m_2\alpha - \phi m_1 m_2 \lambda \alpha}, \quad \lambda, \alpha \geq 0 \tag{8}$$

where $r, s, m_1, m_2 > 0$ and $\phi \geq 0$, which is a submodel of (2) with a new parameterization (see [30]). A random variable with joint probability density function (8) will be denoted by,

$$(\lambda, \alpha) \sim \mathcal{BGC}(r, s, m_1, m_2, \phi) \tag{9}$$

The conditional distributions are:

$$\lambda | \alpha \sim \mathcal{G}a(r, m_1(1 + \phi m_2 \alpha)), \tag{10}$$
$$\alpha | \lambda \sim \mathcal{G}a(s, m_2(1 + \phi m_1 \lambda)). \tag{11}$$

The marginal distributions of (8) are given by,

$$f_\lambda(\lambda) = \frac{k_{r,s}(\phi) m_1^r}{\Gamma(r)} \frac{\lambda^{r-1} e^{-m_1\lambda}}{(1 + \phi m_1 \lambda)^s}, \quad \lambda > 0 \tag{12}$$

and

$$f_\alpha(\alpha) = \frac{k_{r,s}(\phi) m_2^s}{\Gamma(s)} \frac{\alpha^{s-1} e^{-m_2\alpha}}{(1 + \phi m_2 \alpha)^r}, \quad \alpha > 0 \tag{13}$$

The normalizing constant is given by,

$$k_{r,s}(\phi) = \frac{\phi^r}{U(r, r - s + 1, 1/\phi)}, \tag{14}$$

where $U(a, b, z)$ denotes the hypergeometric function, which is defined by $(a, z > 0)$,

$$U(a, b, z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1 + t)^{b-a-1} dt \tag{15}$$

[33].

Table 1 Hyperparameter updating of the prior (2) with likelihood (1)

Parameter	Prior value	Posterior value
m_{10}	$m_{10}^{(0)}$	$m_{10}^{(0)} + k$
m_{01}	$m_{01}^{(0)}$	$m_{01}^{(0)} + n - k + 1$
m_{20}	$m_{20}^{(0)}$	$m_{20}^{(0)} + \sum_{i=1}^{n} x_i$
m_{02}	$m_{02}^{(0)}$	$m_{02}^{(0)} + \sum_{i=k+1}^{n} x_i$
m_{11}	$m_{11}^{(0)}$	$m_{11}^{(0)}$
m_{12}	$m_{12}^{(0)}$	$m_{12}^{(0)}$
m_{21}	$m_{21}^{(0)}$	$m_{21}^{(0)}$
m_{22}	$m_{22}^{(0)}$	$m_{22}^{(0)}$

The raw moments of the marginal distributions (12) and (13) are given by,

$$E(\lambda^k) = \frac{\Gamma(r + k)k_{r,s}(\phi)}{\Gamma(r)k_{r+k,s}(\phi)m_1^k} \tag{16}$$

and

$$E(\alpha^k) = \frac{\Gamma(s + k)k_{r,s}(\phi)}{\Gamma(s)k_{r,s+k}(\phi)m_2^k} \tag{17}$$

respectively.

4 Bayesian Analysis with GBGC Prior

Let us assign the joint distribution of λ and α

$$(\lambda, \alpha) \sim \mathcal{GBGC}(\boldsymbol{m}^{(0)}),$$

as a prior distribution and incorporate Eq. (2) with (1). Thus, we obtain the following posterior distribution

$$(\lambda, \alpha)|x \sim \mathcal{GBGC}(\boldsymbol{m}^*),$$

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

Now, if we consider the submodel

$$(\lambda, \alpha) \sim \mathcal{BGC}(r^{(0)}, s^{(0)}, m_1^{(0)}, m_2^{(0)}, \phi^{(0)}) \tag{18}$$

Table 2 Hyperparameter updating of the prior (8) with likelihood (1)

Parameter	Prior value	Posterior value
r	$r^{(0)}$	$r^{(0)} + \sum_{i=1}^{n} x_i$
s	$s^{(0)}$	$s^{(0)} + \sum_{i=k+1}^{n} x_i$
m_1	$m_1^{(0)}$	$m_1^{(0)} + \sum_{i=1}^{k} t_i$
m_2	$m_2^{(0)}$	$m_2^{(0)}$
ϕ	$\phi^{(0)}$	$\dfrac{\phi^{(0)} m_1^{(0)} m_2^{(0)} \sum_{i=k+1}^{n} t_i}{(m_1^{(0)} + \sum_{i=1}^{k} t_i) m_2^{(0)}}$

and we combine with (2) we obtain

$$(\lambda, \alpha)|x \sim \mathcal{BGC}(r^{(*)}, s^{(*)}, m_1^{(*)}, m_2^{(*)}, \phi^{(*)}) \tag{19}$$

where the parameters are updated according to Table 2

4.1 Parameter Estimation

In the conditional context, Gibbs sampling is a natural estimation methodology. The idea in Gibbs sampling is to generate posterior samples by sweeping through each variable (or block of variables) to sample from its conditional distribution with the remaining variables fixed to their current values. Assume that we are interested in approximating the posterior moments of a given function of λ and α, say $\delta(\lambda, \alpha)$.

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

$$\lambda_1, \alpha_1, \lambda_2, \alpha_2, \ldots, \lambda_{m_0+m}, \alpha_{m_0+m}$$

using the conditional gamma distributions,

$$\lambda|(\alpha, x) \sim \mathcal{G}a(a_1(\alpha), p_1(\alpha))$$
$$\alpha|(\lambda, x) \sim \mathcal{G}a(a_2(\lambda), p_2(\lambda)),$$

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

$$E(\delta(\lambda, \alpha)|x) \approx \frac{1}{m} \sum_{i=m_0+1}^{m_0+m} \delta(\lambda_i, \alpha_i).$$

At each iteration l, for $l = 1, \ldots, M$, the following steps are repeated:

(1) Set $l = 0$ and initial values as $\lambda^{(0)}, \alpha^{(0)}$ and $k^{(0)}$.
(2) Repeat for $l = 1, \ldots, M$

  (a) Draw $\lambda^{(l)}|(\alpha^{(l-1)}, k^{(l-1)})$ from $\mathcal{G}a(a_1(\alpha)^{(l-1)}, b_1(\alpha)^{(l-1)}; k^{(l-1)})$.

(b) sample from $\alpha^{(l)}|(\lambda^{(l)}, k^{(l-1)})$ from $\mathcal{G}a(a_2(\lambda^{(l)}), b_2(\lambda^{(l)}); k^{(l-1)})$.

(c) sample from $k^{(l)}|(\lambda^{(l)}, \alpha^{(l)})$ from $\pi(k|\lambda^{(l)}, \alpha^{(l)})$.

where $\pi(k|\lambda^{(l)}, \alpha^{(l)})$ is the conditional posterior distribution of k given in (20). Let us assume $k \sim unif\{1, n\}$ then k can be updated using its posterior conditional distribution

$$\pi(k^{(l)}|x, \lambda^{(l)}, \alpha^{(l)}) = \frac{\exp\left[k^{(l)}(\alpha^{(l)} - \lambda^{(l)})\right]\left(\frac{\lambda^{(l)}}{\alpha^{(l)}}\right)^{S_{k^{(l)}}}}{\sum_{j=1}^{n} \exp\left[j(\alpha^{(l)} - \lambda^{(l)})\right]\left(\frac{\lambda^{(l)}}{\alpha^{(l)}}\right)^{S_j}}, \tag{20}$$

where m^{th} iteration, for $m = 1, \ldots, M$, and S_k is $\sum_{i=1}^{k} x_i$ [34]. As a previous step, we must elicit the hyperparameters m_{ij} using the methods proposed in Sect. 4.2.

4.2 Hyperparameter Elicitation

This section provides a brief description of eliciting hyperparameters. Similar to Arnold01, we consider inconsistency in the prior information provided by the expert or experts. This will allow us to assume a sample space for the conditional conjugate space rather than exactly one value that agrees with the prior information elicited from our expert or experts.

Let μ, σ^2, μ' and σ'^2 denote the conditional mean and variance of $\lambda|\alpha$ and $\alpha|\lambda$, respectively. We assume a series of $\alpha_1, \ldots, \alpha_{m_1}$ and $\lambda_1, \ldots, \lambda_{m_2}$ based on the experts' information are available. This results in a system of equations will allow us to obtain a realisation of hyperparametrs.

$$\mu_j = E(\lambda|\alpha = \alpha_j)$$
$$\sigma_j^2 = Var(\lambda|\alpha = \alpha_j) \quad j = 1, \ldots, m_1, \tag{21}$$
$$\mu_i' = E(\alpha|\lambda = \lambda_i)$$
$$\sigma_i'^2 = Var(\alpha|\lambda = \lambda_i) \quad i = 1, \ldots, m_2 \tag{22}$$

In order to satisfy conditions in Eq. (6) sequential quadratic programming using `nloptr` [35] in R was used to solve the above system of equations and elicit hyperparameters. Sequential Quadratic Programming (SQP) is one of the most successful methods for the numerical solution of constrained nonlinear optimization problems. It relies on a profound theoretical foundation and provides powerful algorithmic tools for the solution of large-scale technologically relevant problems.

5 Numerical Experiments

The numerical experiments is designed as follows. The method will be applied to sets of simulated data and real data. The simulation study investigated the performance of the model under three cases: 1. all parameters of the model are fixed, 2. varying k, 3. varying n and k. In all examples, 100 sets of data are simulated and number

Table 3 Posterior mean (sd) for single Gibbs sampler of the full model and sub-model with $n = 200$, $M = 15000$ after 5000 burnin where change happens at $k = 25\%$, 50% and 75%

Parameters	Full-model			Sub-model		
	25%	50%	75%	25%	50%	75%
λ	0.91 (0.008)	0.95 (0.008)	1.02 (0.007)	1.13 (0.007)	1.04 (0.007)	1.04 (0.007)
α	3.05 (0.014)	2.8 (0.018)	2.96 (0.029)	2.87 (0.015)	3.01 (0.016)	2.66 (0.023)
k	49.9 (0.83)	98.17 (3.83)	149.7 (1.94)	49.9 (1.76)	100.13 (1.8)	150.42 (1.54)

of iterations are 20,000. Based on Gelman-Rubin convergence diagnostics [36, 37], chains were considered to be converging after 5000 iterations, in all simulations. The accuracy of parameter estimates using root mean square error (RMSE).

The real data application will use the well known mine data available in R.

5.1 Simulation Study

Trace plots and marginal posterior density of model parameters for the full model and the submodel are presented in.

From Table 2, it can be seen that model is not generally affected by the number of observations and where the change happens.

Finally, we extended the simulation study to sets of MC simulation studies for $N = 10000$ and obtained the RMSE for the parameters of the model. In order to present the convergence diagnostics, we selected Figs. 3 and 4 as examples of the Gelman-Rubin convergence diagnostics plots which confirm the convergence after 5000 burn-in [38].

Case 1: $n = 200$, $k = 100$, $\lambda = 1$ and $\alpha = 3$

Case 1 simulates data from a model with $\lambda = 1, \alpha = 3, n = 200$ and $k = 100$. A realisation of the conditional posterior distributions of model parameters was obtained using Gibbs sampler, and graphical presentation of the results including trace plots, marginal posterior densities and Gelman–Rubin diagnostics plot are given in Figs. 6–4. It can be seen for both models the posterior mean of parameters are close to true values. However, the posterior variance for parameters λ and α were small.

Case 2: $n = 200$, $\lambda = 1$, $\alpha = 3$ and $k = 50$, 100, 150

The simulation study was extended where λ, α and n are fixed with values of 1, 3 and 200, respectively and k varies at 25%, 50% and 75% of the total number of observations. This will allow us to assess the performance of the model based on where the change happens.

Results for this case are summarised in Table 3. It can be seen k has hardly been affected by the position of change. However, the estimation of parameters λ and α was less accurate when the location of change was at the 25% of the total sample size for both full- and sub-model.
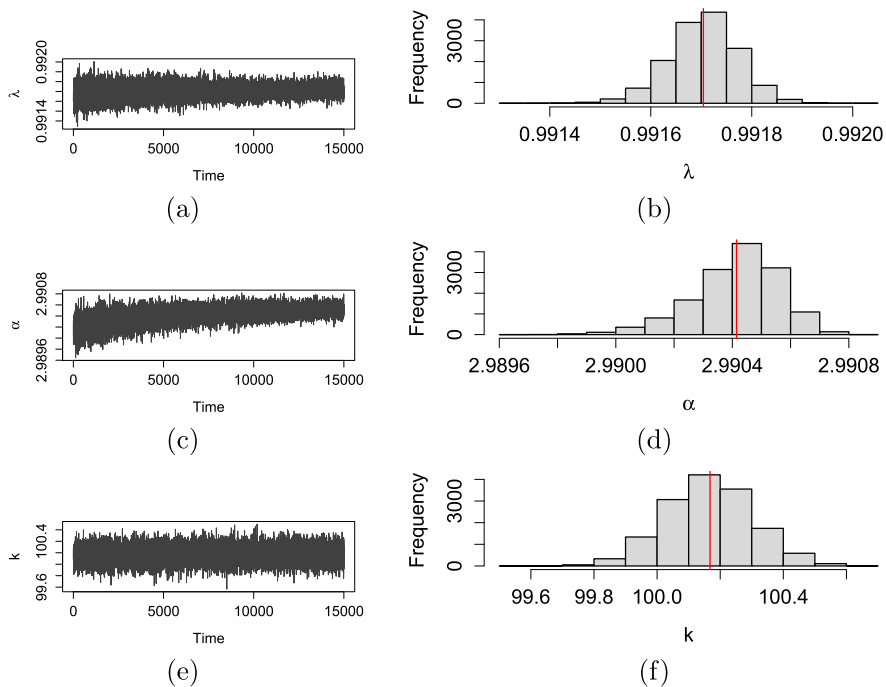
Fig. 1 Monte Carlo simulation results of the full model

Table 4 Residual mean square error (RMSE) of the full model, where the number of observations $N = 50, 100, 150$ and change points happening at the following percentages 25%, 50%, 75%.

N & k	λ	α	K
50 25	$2.28.10^{-4}$	$2.11.10^{-5}$	$2.98.10^{-2}$
50 50	$1.78.10^{-6}$	$4.4.10^{-5}$	$6.1.10^{-3}$
50 75	$1.68.10^{-5}$	$7.95.10^{-5}$	$1.71.10^{-2}$
100 25	$6.71.10^{-6}$	$8.22.10^{-7}$	$2.82.10^{-6}$
100 50	$3.01.10^{-5}$	$2.07.10^{-5}$	$9.21.10^{-2}$
100 75	$1.29.10^{-8}$	$7.04.10^{-4}$	$7.03.10^{-2}$
150 25	$1.25.10^{-6}$	$2.08.10^{-7}$	$2.54.10^{-3}$
150 50	$5.23.10^{-6}$	$1.22.10^{-5}$	$4.23.10^{-2}$
150 75	$2.21.10^{-7}$	$6.04.10^{-7}$	$1.41.10^{-3}$

Case 3: $\lambda = 1$ and $\alpha = 3$, varying n and k

In the final case we assessed the effect of sample size on the parameter estimation, in addition to the location of change. Data was simulated with values of $n = 50, 100$ and 150. Tables 4 and 5 provide summarise performance of full- and sub-models using RMSE.
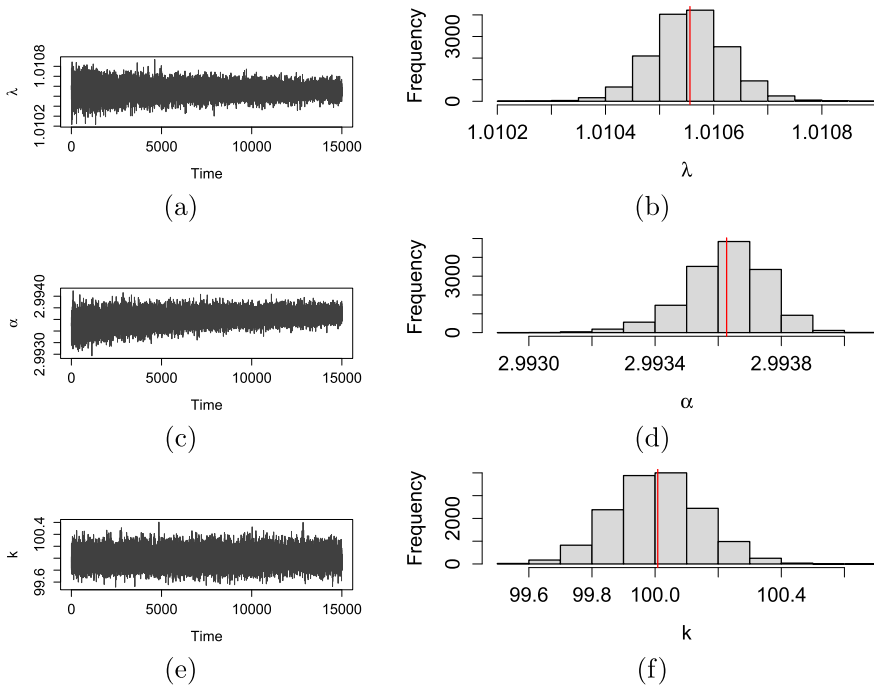
Fig. 2 Monte Carlo simulation results of the sub-model
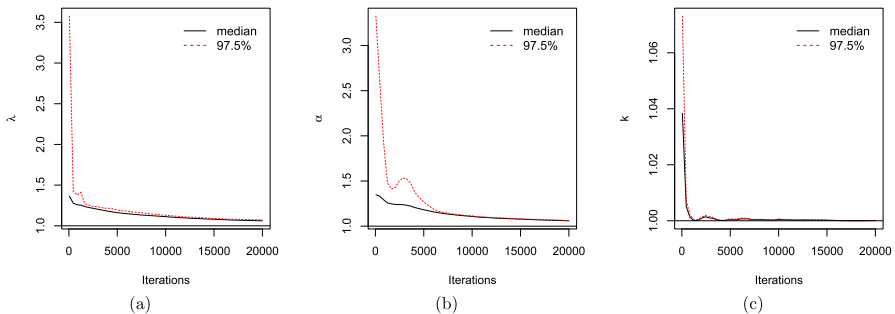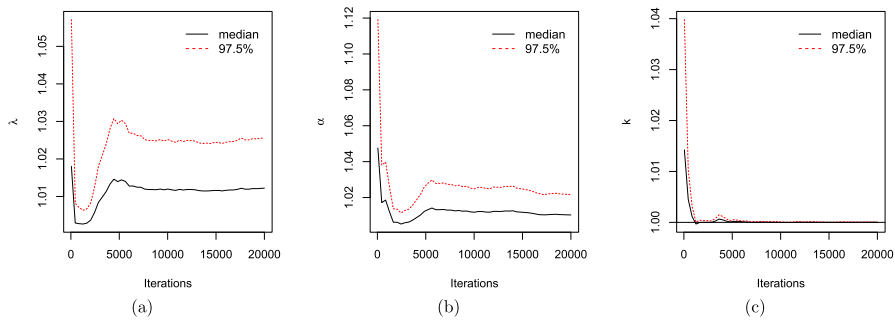

Fig. 3 Gelman–Rubin diagnostics plots and statistics for parameters **a** λ, **b** α and **c** k of the full model with $n = 200$ and $k = 100$. The Gibbs sampler converges after 5000 iterations

5.2 Real Data: Coal Mine Example

The coal mine example adopted from [39]. The data was downloaded from the `boot` library in R [40]. The data set provided the dates of 191 explosions in coal mines which resulted in 10 or more fatalities between March 15, 1851 and March 22, 1962.

Let the number of explosions with more than 10 mortalities, X such that

$$x_i \sim Po(\lambda), \ i = 1, \ldots, k$$

Fig. 4 Gelman–Rubin diagnostics plots and statistics for parameters **a** λ, **b** α and **c** k of the sub-model model with $n = 200$ and $k = 100$. The Gibbs sampler converges after 5000 iterations

	N & k	λ	α	K
Table 5 Residual mean square error (RMSE) of the sub model, where the number of observations $N = 50, 100, 150$ and change points happening at the following percentages 25%, 50%, 75%.	50 25	$2.67.10^{-4}$	$5.8.10^{-7}$	$1.31.10^{-3}$
	50 50	$8.82.10^{-4}$	$1.91.10^{-4}$	$3.53.10^{-1}$
	50 75	$5.19.10^{-12}$	$1.81.10^{-4}$	$7.3.10^{-3}$
	100 25	$3.59.10^{-3}$	$1.5.10^{-4}$	1.22
	100 50	$2.46.10^{-7}$	$1.23.10^{-5}$	$1.88.10^{-2}$
	100 75	$3.58.10^{-4}$	$7.41.10^{-4}$	2.39
	150 25	$6.01.10^{-7}$	$9.59.10^{-7}$	$4.78.10^{-3}$
	150 50	$2.28.10^{-8}$	$2.14.10^{-6}$	$2.95.10^{-3}$
	150 75	$5.12.10^{-8}$	$2.68.10^{-5}$	$1.43.10^{-2}$
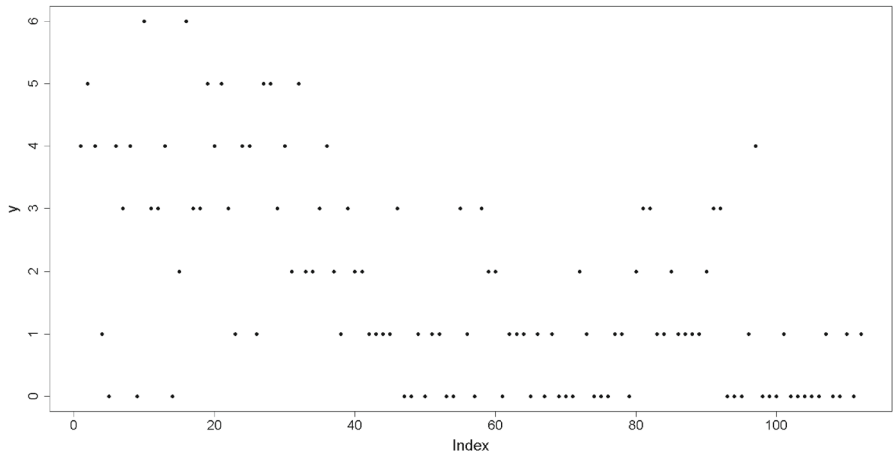
Fig. 5 Mine data adopted from Davison and Hinkley (1997) illustrates number of explosions with more than 10 mortalities from 1851 to 1962

Table 6 Hyperameters for the full modell in the mine data example

Parameter	Prior value	Posterior value
m_{10}	0.872	$0.872 + c^{(j)}$
m_{01}	0.761	$0.761 + (n - c^{(j)})$
m_{20}	1.682	$1.682 + \sum_{i=1}^{c^{(j)}} x_i$
m_{02}	2.390	$2.390 + \sum_{i=c^{(j)}+1}^{n} x_i$
m_{11}	-0.511	-0.511
m_{12}	-1.308	-1.308
m_{21}	-0.429	-0.429
m_{22}	-1.305	-1.305

Table 7 Posterior mean (sd) and 95% credible intervals (C.I.) for parameters of the full model for the min data example

Parameters	Proposed Model		Basic Model	
	Mean (SD)	95% C.I	Mean (SD)	95% C.I
λ	3.12 (0.003)	(3.11, 3.12)	3.06 (0.280)	(2.53, 3.65)
α	0.92 (0.001)	(0.92, 0.93)	0.92 (0.116)	(0.70, 1.16)
k	39.89 (2.25)	(36, 46)	40.14 (2.468)	(36, 46)

$$x_i \sim Po(\alpha), \ i = k + 1, \ldots, 191.$$

The full model was only fitted to the data. In order to solve the system of equations in (6)–(5) elicit hyperparameters, we assumed the following values for the conditional expected mean and variances

$$\mu_1 = E(\lambda|\alpha) = 0.038, \sigma_1^2 = Var(\lambda|\alpha) = 0.00012$$
$$\mu_2 = E(\lambda|\alpha) = 0.038, \sigma_2^2 = Var(\lambda|\alpha) = 0.00011$$
$$\mu_3 = E(\alpha|\lambda) = 0.024, \sigma_3^2 = Var(\alpha|\lambda) = 0.00012$$
$$\mu_4 = E(\alpha|\lambda) = 0.024, \sigma_4^2 = Var(\alpha|\lambda) = 0.000073$$

Therefore, under the general model hyperparameters can be elicited as illustrated in Table 6.

Table 7 provides a summary of the posterior mean, standard deviation and 95% credible intervals (CI) of the model parameters. The results are compared with a basic model where we assume λ and α are independently distributed with non-informative gamma priors. It can be seen that although both models have close posterior means, the proposed model offers smaller standard deviation and a narrower 95% CI.

Trace plots and marginal density plots are presented in Fig. 6. Finally, the posterior mean of the model parameters were added to the coal mine data along with the estimated change point and presented in Fig. 7. It can be seen that the number of explosions resulting in 10 or more mortalities have decreased around year 40 which is 1891.
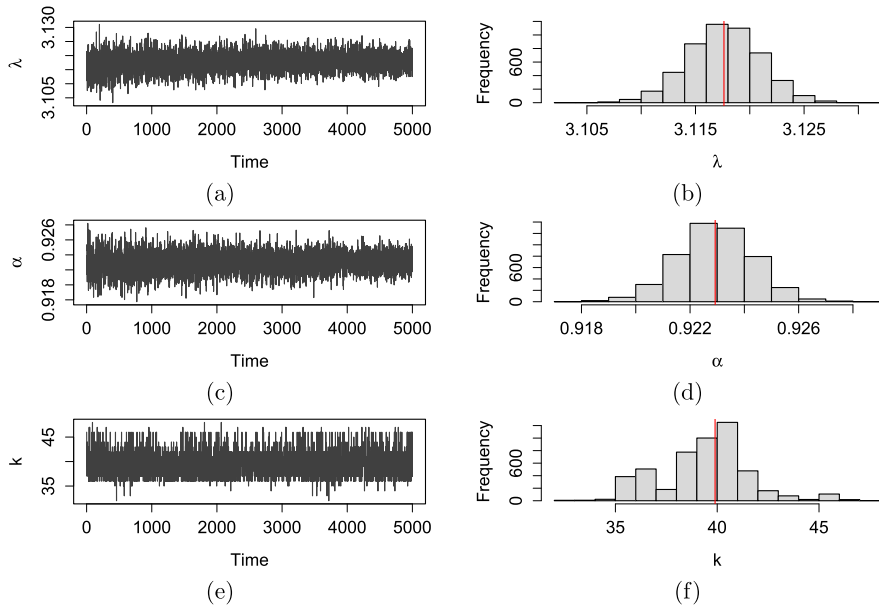
Fig. 6 Mine data Gibbs sampler chain results, where **a**, **c** and **e** are the chain convergence plots for the parameters λ,α amd k
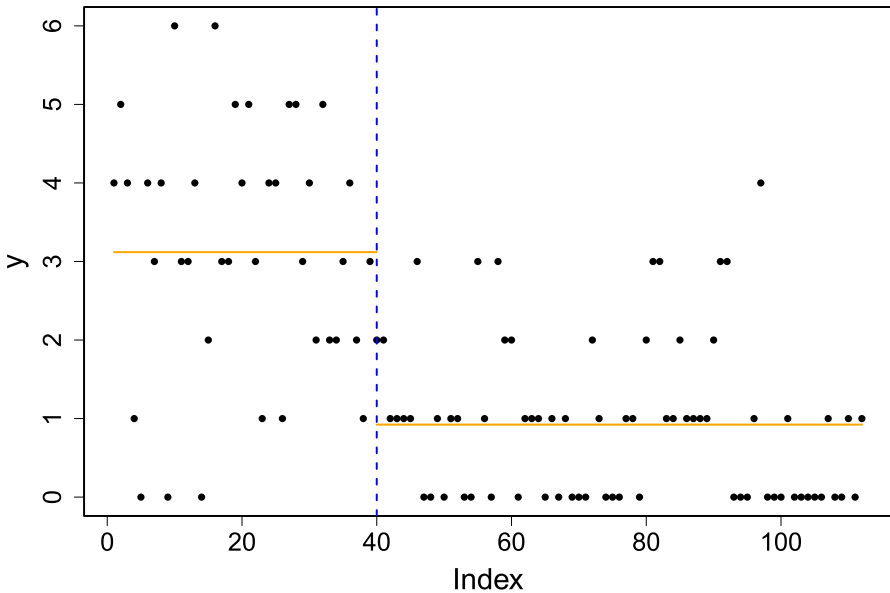

Fig. 7 Mine data along with the posterior mean of λ and α presented in yellow

6 Conclusions

In this paper, we presented a novel approach for estimating change point problems by using a broad class of conjugate prior distributions derived from a conditional specification methodology. While previous research extensively demonstrated the application of such prior distributions in problems involving continuous distributions, our contribution lied in exploring their effectiveness in the context of discrete distributions, specifically the Poisson process.

We conducted a comprehensive simulation study and applied the proposed methodology to real mine data. Through Gelman-Rubin diagnostics, we confirmed the convergence of the Gibbs sampler after a burn-in period of 5000 iterations. The simulation results revealed that parameter estimates exhibit smaller error when the change point, denoted as k, is closer to half of the total data points n. We compared the results obtained using our methodology with those of a basic model assuming independent parameters with non-informative priors. The findings demonstrated that our proposed approach yields significantly smaller estimation errors.

Furthermore, our methodology holds potential for extension to estimate multiple change points. Additionally, the introduction of a bivariate Poisson process poses new challenges that warrant further investigation and refinement of our proposed methodology. Overall, our research contributes to the advancement of change point analysis and provides valuable insights for improving the estimation accuracy in various application domains.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval All authors approved this article.

Appendix

Posterior Conditional Distributions

Let X_i be a Poisson random variable where $i = 1, \ldots, T$. Assume that the change point occurs at time k such that

$$x_i \sim \mathcal{P}o(\lambda), \quad i = 1, 2, \ldots, k,$$

and

$$x_i \sim \mathcal{P}o(\alpha), \quad i = k+1, \ldots, n,$$

where $\lambda > 0$, $\alpha > 0$ and $k \in \{1, 2, \ldots, n\}$ are parameters of interest. The likelihood function can be written as:

$$
\begin{aligned}
L(\lambda, \alpha, k) &= \prod_{1}^{k} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \prod_{k+1}^{n} \frac{e^{-\alpha} \alpha^{x_i}}{x_i!} \\
&= \frac{e^{-k\lambda} \lambda^{\sum_1^k x_i}}{\prod_1^k x_i!} \frac{e^{-(n-k+1)\alpha} \alpha^{\sum_{k+1}^n x_i}}{\prod_{k+1}^n x_i!} \\
&\propto \lambda^{\sum_1^k x_i} \alpha^{\sum_{k+1}^n x_i} \exp\left(-k\lambda - (n-k+1)\alpha\right).
\end{aligned}
$$

We assign a bivariate gamma prior distribution to the joint distribution of (λ, α) with the following form:

$$
\begin{aligned}
\pi(\lambda, \alpha; \boldsymbol{m}) \propto (\lambda\alpha)^{-1} \exp[&-m_{10}\lambda - m_{01}\alpha + m_{20}\log\lambda + m_{02}\log\alpha \\
&+ m_{11}\lambda\alpha - m_{12}\lambda\log\alpha - m_{21}\alpha\log\lambda + m_{22}\log\lambda\log\alpha].
\end{aligned}
$$

Thus the joint posterior distribution of $(\lambda, \alpha)|(x, k)$ can be obtain as:

$$
\begin{aligned}
\pi((\lambda, \alpha)|(x, k)) \propto\ & L(\lambda, \alpha, k)\pi(\lambda, \alpha; \boldsymbol{m}) \\
=\ & \lambda^{\sum_1^k x_i} \alpha^{\sum_{k+1}^n x_i} \exp\left(-k\lambda - (n-k+1)\alpha\right) \\
& \times (\lambda\alpha)^{-1} \exp[-m_{10}\lambda - m_{01}\alpha + m_{20}\log\lambda + m_{02}\log\alpha \\
& + m_{11}\lambda\alpha - m_{12}\lambda\log\alpha - m_{21}\alpha\log\lambda + m_{22}\log\lambda\log\alpha] \\
=\ & exp[-(m_{10} + k)\lambda - (m_{01} + n - k + 1)\alpha] \\
& \exp[(m_{20} + \sum_1^k x_i)\log(\lambda) + (m_{02} + \sum_{k+1}^n x_i)\log(\alpha)] \\
& \exp[\lambda\alpha - \lambda\log(\alpha) - \alpha\log(\lambda) + \log(\lambda)\log(\alpha)].
\end{aligned}
$$

Pseudo Code: Gibbs Sampler

```
    ####   Nonlinear Programming ####
## Solve inequality in equations ()−() using nonlinear programming
## to obtain hyperparameters
library(nloptr)
## set the initial solution
x0.hs100 <− answer

## Define posterior function
fn.hs100 <− function(x) {
−0.038^2/0.00012 + x[3]  −0.01 * x[7] + log(0.01) * x[8]−
0.038/0.00012 + x[1] −0.01 * x[5] + log(0.01)*x[6]−
0.038^2/0.00011 + x[3]  −0.08 * x[7] + log(0.08) * x[8]−
0.038/0.00011 + x[1] −0.08 * x[5] + log(0.08)*x[6] −
0.0242^2/0.00012 + x[4] −0.01 * x[6] + log(0.01) * x[8]−
0.0242/0.00012 + x[2] −0.01 * x[5] + log(0.01) * x[7] −
0.0242^2/0.000073 + x[4] −0.08 * x[6] + log(0.08) * x[8]−
0.0242/0.000073 + x[2] −0.08 * x[5] + log(0.08) * x[7]
}

## Assign conditions
hin.hs100 <− function(x) {
y <− numeric(8)
y[5] <− −x[5] ; y[6]<− −x[6]; y[7]<− −x[7]; y[8]<− −x[8]
y[1] <− x[1] − x[6]*(1 − log(x[6]/x[5]))
y[2] <− x[3] − x[8]*(1 − log(x[8]/x[7]))
y[3] <− x[2] − x[7]*(1 − log(x[7]/x[5]))
y[4] <− x[4] − x[8]*(1 − log(x[8]/x[6]))
return(y)
}
## Solve for hyperparameters
hinjac.hs100 <− function(x) nl.jacobian(x, hin.hs100, heps = 1e−2)
S <− slsqp(x0.hs100, fn = fn.hs100,
hin = hin.hs100, hinjac = hinjac.hs100,
control = list(xtol_rel = 1e−12))

    ####    Gibbs sampler    ####
## Storing simulated values
a1<−a2<− b1<− b2<−lambda.g<−alpha.g<− numeric(m)
## Set initial values
a1[1]<− m20; a2[1]<−m02; b1[1]<−m10; b2[1]<−m01
## Gibbs sampler
lambda.g[1] <− 2;alpha.g[1]<− 5
for(i in 2:m){
```

```
## Sample from lambda|alpha
a1[i] <- m20 - m21 *alpha.g[i-1] + m22*log(alpha.g[i-1]) #
b1[i] <- m10 - m11 *alpha.g[i-1] + m12*log(alpha.g[i-1])#
lambda.g[i] <- rgamma(1,a1[i], b1[i])
## Sample from lambda|alpha
a2[i] <- m02 - m12 *lambda.g[i] + m22*log(lambda.g[i]) #
b2[i] <- m01 - m11 *lambda.g[i] + m21*log(lambda.g[i])#
alpha.g[i]<- rgamma(1,a2[i],b2[i])
}
```

References

1. Hastie T, Tibshirani R, Friedman JH, Friedman JH (2009) The elements of statistical learning: data mining, inference and prediction, vol 2. Springer, Berlin
2. Shi Y, Tian Y, Kou G, Peng Y, Li J (2011) Optimization based data mining: theory and applications. Springer, Berlin
3. Tien JM (2017) Internet of things, real-time decision making, and artificial intelligence. Ann Data Sci 4:149–178
4. Olhede SC, Wolfe PJ (2018) The future of statistics and data science. Statist Prob Lett 136:46–50
5. Louis D, Shi Y (2007) Introduction to business data mining, vol 10. McGraw-Hill/Irwin, New York
6. Shi Y (2022) Advances in big data analytics: theory, algorithm and practice. Springer, Singapore
7. Irizarry RA (2019) Introduction to data science. Chapman and Hall-CRC Press, Boca Raton
8. Agresti A, Kateri M (2021) Foundations of statistics for data scientists: with R and Python. Chapman and Hall-CRC Press, Boca Raton
9. Bruce P, Bruce A, Gedeck P (2020) Practical statistics for data scientists: 50+ essential concepts using R and Python. O'Reilly Media
10. Mishra S, Datta-Gupta A (2017) Applied statistical modeling and data analytics: a practical guide for the petroleum geosciences. Elsevier, New York
11. Olson DL, Shi Y (2007) Introduction to business data mining. McGraw-Hill/Irwin, New York
12. Shi Y, Tian YJ, Kou G, Peng Y, Li JP (2011) Optimization based data mining: theory and applications. Springer, Berlin
13. Tien JM (2017) Internet of things, real-time decision making, and artificial intelligence. Ann Data Sci 4(2):149–178
14. Chen M, Mao S, Liu Y (2014) Big data: a survey. Mobile Netw Appl 19:171–209
15. Killick R, Fearnhead P, Eckley IA (2012) Optimal detection of changepoints with a linear computational cost. J Am Stat Assoc 107(500):1590–1598
16. Zhang L, Jia S, Yang M, Yao X, Li C, Sun J, Huang Y et al (2014) Detection of copy number variations and their effects in Chinese bulls. BMC Genom 15:1
17. Carslaw DC, Ropkins K, Bell MC (2006) Change-point detection of gaseous and particulate traffic-related pollutants at a roadside location. Environ Sci Technol 40(22):6912–6918
18. Sluss D, Bingham C, Burr M, Bott ED, Riley EA, Reid PJ (2009) Temperature-dependent fluorescence intermittency for single molecules of violamine r in poly (vinyl alcohol). J Mater Chem 19(40):7561–7566
19. Arnold BC, Castillo E, Sarabia JM (1998) The use of conditionally conjugate priors in the study of ratios of gamma scale parameters. Comput Stat Data Anal 27(2):125–139
20. Sarabia JM, Shahtahmassebi G (2017) Bayesian estimation of incomplete data using conditionally specified models. Commun Stat Simul Comput 46:3419–3435
21. Arnold BC, Castillo E, Sarabia JM (1998) Bayesian analysis for classical distributions using conditionally specified priors. Sankhyā Indian J Stat Ser B 228–245
22. Chen J, Gupta AK (2012) Parametric statistical change point analysis: with applications to genetics, medicine, and finance, 2nd edn. Birkhäuser, Boston
23. Chen J, Gupta AK (2001) On change point detection and estimation. Commun Stat Simul Comput 30(3):665–697
24. Barry D, Hartigan JA (1993) A Bayesian analysis for change point problems. J Am Stat Assoc 88(421):309–319

25. Adams RP, MacKay DJC (2007) Bayesian online change point detection. arXiv:0710.3742
26. Antoch J, Husková M, Veraverbeke N (1995) Change-point problem and bootstrap. J Nonparametr Stat 5(2):123–144
27. Malhotra P, Vig L, Shroff G, Agarwal P, et al (2015) Long short term memory networks for anomaly detection in time series. In: ESANN, p 89
28. Aminikhanghahi S, Cook DJ (2017) A survey of methods for time series change point detection. Knowl Inf Syst 51(2):339–367
29. Sharma S, Swayne DA, Obimbo C (2016) Trend analysis and change point techniques: a survey. Energy Ecol Environ 1:123–130
30. Arnold BC, Castillo E, Sarabia JM (2001) Conditionally specified distributions: an introduction (with discussion). Stat Sci 16:249–274
31. Hobert JP, Casella G (1996) The effect of improper prior on gibbs sampling in hierarchical linear mixed models. J Am Stat Assoc 91:1481–1473
32. Arnold BC, Castillo E, Sarabia JM (1999) Conditional specification of statistical models. Springer Series in Statistics. Springer, New York
33. Abramowitz M, Stegun I (1964) Handbook of mathematical functions. Government Publishing Office, Washington, DC
34. Rizzo ML (2007) Statistical computing with R. Chapman and Hall/CRC, New York
35. Johnson SG (2007) The NLopt nonlinear-optimization package. http://github.com/stevengj/nlopt
36. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Stat Sci 7(4):457–472
37. Brooks SP, Gelman A (1998) General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 7(4):434–455
38. Plummer M, Best N, Cowles K, Vines K (2006) Coda: convergence diagnosis and output analysis for mcmc. R News 6:7–11
39. Hand David J, Daly Fergus, McConway K, Lunn D, Ostrowski E (1993) A handbook of small data sets
40. Canty A, Ripley B (2017) boot: Bootstrap r (s-plus) functions. R package version 1(3–20):2017