

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Modelos de mistura para dados com distribuições Poisson truncadas no zero

Andressa do Carmo Gigante

Dissertação de Mestrado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Andressa do Carmo Gigante

Modelos de mistura para dados com distribuições Poisson truncadas no zero

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Luis Aparecido Milan

USP – São Carlos
Novembro de 2017

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

G333m Gigante, Andressa do Carmo
Modelos de mistura para dados com distribuições
Poisson truncadas no zero / Andressa do Carmo
Gigante; orientador Luis Aparecido Milan. -- São
Carlos, 2017.
103 p.

Dissertação (Mestrado - Programa
Interinstitucional de Pós-graduação em Estatística) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2017.

1. variável truncada no zero. 2. modelo de
mistura. 3. métodos de agrupamento ou segmentação.
4. mistura de Poisson truncadas no zero. 5.
algoritmo EM adaptado. I. Milan, Luis Aparecido ,
orient. II. Título.

Andressa do Carmo Gigante

Mixture models for data with zero truncated Poisson
distributions

Master dissertation submitted to the Institute of
Mathematics and Computer Sciences – ICMC-USP
and to the Department of Statistics – DEs-UFSCar, in
partial fulfillment of the requirements for the degree of
the Master Interagency Program Graduate in Statistics.
FINAL VERSION

Concentration Area: Statistics

Advisor: Prof. Dr. Luis Aparecido Milan

USP – São Carlos
November 2017

RESUMO

GIGANTE, A. C. **Modelos de mistura para dados com distribuições Poisson truncadas no zero**. 2017. 103 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2017.

Modelo de mistura de distribuições tem sido utilizado desde longa data, mas ganhou maior atenção recentemente devido ao desenvolvimento de métodos de estimação mais eficientes. Nesta dissertação, o modelo de mistura foi utilizado como uma forma de agrupar ou segmentar dados para as distribuições Poisson e Poisson truncada no zero. Para solucionar o problema do truncamento foram estudadas duas abordagens. Na primeira, foi considerado o truncamento em cada componente da mistura, ou seja, a distribuição Poisson truncada no zero. E, alternativamente, o truncamento na resultante do modelo de mistura utilizando a distribuição Poisson usual. As estimativas dos parâmetros de interesse do modelo de mistura foram calculadas via metodologia de máxima verossimilhança, sendo necessária a utilização de um método iterativo. Dado isso, implementamos o algoritmo EM para estimar os parâmetros do modelo de mistura para as duas abordagens em estudo. Para analisar a performance dos algoritmos construídos elaboramos um estudo de simulação em que apresentaram estimativas próximas dos verdadeiros valores dos parâmetros de interesse. Aplicamos os algoritmos à uma base de dados real de uma determinada loja eletrônica e para determinar a escolha do melhor modelo utilizamos os critérios de seleção de modelos *AIC* e *BIC*. O truncamento no zero indica afetar mais a metodologia na qual aplicamos o truncamento em cada componente da mistura, tornando algumas estimativas para a distribuição Poisson truncada no zero com viés forte. Ao passo que, na abordagem em que empregamos o truncamento no zero diretamente no modelo as estimativas apontaram menor viés.

Palavras-chave: variável truncada no zero, modelo de mistura, métodos de agrupamento ou segmentação, mistura de Poisson truncadas no zero, algoritmo EM adaptado.

ABSTRACT

GIGANTE, A. C. **Mixture models for data with zero truncated Poisson distributions**. 2017. 103 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2017.

Mixture models has been used since long but just recently attracted more attention for the estimations methods development more efficient. In this dissertation, we consider the mixture model like a method for clustering or segmentation data with the Poisson and Poisson zero truncated distributions. About the zero truncation problem we have two emplacements. The first, consider the zero truncation in the mixture component, that is, we used the Poisson zero truncated distribution. And, alternatively, we do the zero truncation in the mixture model applying the usual Poisson. We estimated parameters of interest for the mixture model through maximum likelihood estimation method in which we need an iterative method. In this way, we implemented the EM algorithm for the estimation of interested parameters. We apply the algorithm in one real data base about one determined electronic store and towards determine the better model we use the criterion selection AIC and BIC. The zero truncation appear affect more the method which we truncated in the component mixture, return some estimates with strong bias. In the other hand, when we truncated the zero directly in the model the estimates pointed less bias.

Keywords: mixture model, clustering methods, zero truncated Poisson mixture, EM algorithm.

LISTA DE ILUSTRAÇÕES

Figura 1 – Gráfico da distribuição de probabilidade Poisson para dados simulados com diferentes valores do parâmetro θ , ilustrando uma mistura de Poisson. . . .	39
Figura 2 – Gráficos <i>Box-Plot</i> e Histograma da variável quantidade de compras por cliente.	59
Figura 3 – Gráficos do modelo de mistura de Poisson truncadas no zero para K componentes ajustado através do algoritmo EMMPTZ.	61
Figura 4 – Gráficos do modelo de mistura de Poisson truncadas no zero para K componentes ajustado através do algoritmo EMPMTZ.	62

LISTA DE ALGORITMOS

Algoritmo 1 – Algoritmo EM	29
Algoritmo 2 – Algoritmo EM para Modelos de Mistura	30
Algoritmo 3 – Algoritmo EM para Poisson truncada no zero	35
Algoritmo 4 – Algoritmo EM para Mistura de Poisson	38
Algoritmo 5 – Algoritmo EM para Mistura de Poisson truncadas no zero	41
Algoritmo 6 – Algoritmo EM para Mistura de Poisson com truncamento na resultante do modelo e no ponto zero	46

LISTA DE CÓDIGOS-FONTE

Código-fonte 1 – Códigos para gerar Gráfico representativo de Mistura de Poisson . . .	95
Código-fonte 2 – Função para gerar dados da distribuição Poisson truncada no zero . .	96
Código-fonte 3 – Função para estimar o parâmetro para distribuição Poisson truncada no zero	97
Código-fonte 4 – Função para gerar dados com distribuição Poisson truncada no zero .	97
Código-fonte 5 – Algoritmo EM para estimar parâmetros do modelo de mistura de Poisson truncadas no zero	98
Código-fonte 6 – Algoritmo EM para estimar parâmetros do modelo de mistura de Poisson para variável truncada no zero	100
Código-fonte 7 – Funções utilizadas no Código-fonte 6	102

LISTA DE TABELAS

Tabela 1 – Valores da Esperança e da Variância de X , sendo $X \sim PoisTrun(\theta)$, para diversos valores de θ	33
Tabela 2 – Valores fixados na simulação dos dados para os parâmetros θ_k da Poisson truncada no zero e p_k para a mistura.	48
Tabela 3 – Valores iniciais na implementação do algoritmo EM para as estimativas dos parâmetros θ_k	49
Tabela 4 – Amostras em que o algoritmo EM resultou em $K - 1$ componentes na mistura.	53
Tabela 5 – Resultados do algoritmo EM para $\hat{\theta}_k, \hat{p}_k$ e \hat{n}_k para a amostra M^* , $N = 10$ e $\theta_k = (2, 3, 4, 5, 6)$	54
Tabela 6 – Valores da quantidade de compras por cliente, proporção do total de clientes e a proporção acumulada.	58
Tabela 7 – Resultados do algoritmo EMMPTZ aplicado à quantidade de compras por cliente para K componentes na mistura e medidas descritivas dentro de cada componente.	60
Tabela 8 – Resultados do algoritmo EMPMTZ aplicado à quantidade de compras por cliente para K componentes na mistura e medidas descritivas dentro de cada componente.	61
Tabela 9 – Resultados da quantidade de iterações e do tempo de processamento, em segundos, na aplicação do algoritmo EMMPTZ e do algoritmo EMPMTZ.	63
Tabela 10 – Valores AIC e BIC para os modelos ajustados.	64
Tabela 11 – Valores das estimativas dos parâmetros $\hat{\theta}_k$ e \hat{p}_k para o modelo ajustado através do algoritmo EMMPTZ.	64
Tabela 12 – Valores das estimativas dos parâmetros $\hat{\lambda}_k$ e \hat{q}_k para o modelo ajustado através do algoritmo EMPMTZ.	65
Tabela 13 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson truncadas no zero e o tempo de processamento em segundos para $K = 2$ componentes.	71
Tabela 14 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson truncadas no zero e o tempo de processamento em segundos para $K = 3$ e 4 componentes.	72
Tabela 15 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson truncadas no zero e o tempo de processamento em segundos para $K = 5$ componentes.	73

Tabela 16 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 2 componentes.	74
Tabela 17 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 2 componentes.	75
Tabela 18 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 3 componentes.	76
Tabela 19 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 3 componentes.	77
Tabela 20 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 3 componentes.	78
Tabela 21 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 3 componentes.	78
Tabela 22 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 4 componentes.	79
Tabela 23 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 4 componentes.	80
Tabela 24 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 4 componentes.	81
Tabela 25 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 4 componentes.	82
Tabela 26 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 5 componentes.	83
Tabela 27 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 5 componentes.	84
Tabela 28 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 5 componentes.	85
Tabela 29 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 5 componentes.	86
Tabela 30 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson para modelo truncado no zero e o tempo de processamento em segundos para $K = 2$	87
Tabela 31 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson para modelo truncado no zero e o tempo de processamento em segundos para $K = 3$ e 4.	88
Tabela 32 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson para modelo truncado no zero e o tempo de processamento em segundos para $K = 5$	89
Tabela 33 – Resultados do algoritmo EM para as estimativas $\hat{\lambda}_k$, \hat{n}_k e \hat{q}_k para mistura de Poisson para modelo truncado no zero com 2 componentes.	90

Tabela 34 – Resultados do algoritmo EM para as estimativas $\hat{\lambda}_k$, \hat{n}_k e \hat{q}_k para mistura de Poisson para modelo truncado no zero com 3 componentes.	91
Tabela 35 – Resultados do algoritmo EM para as estimativas $\hat{\lambda}_k$, \hat{n}_k e \hat{q}_k para mistura de Poisson para modelo truncado no zero com 4 componentes.	92
Tabela 36 – Resultados do algoritmo EM para as estimativas $\hat{\lambda}_k$, \hat{n}_k e \hat{q}_k para mistura de Poisson para modelo truncado no zero com 5 componentes.	93
Tabela 37 – Resultados do algoritmo EM para as estimativas $\hat{\lambda}_k$, \hat{n}_k e \hat{q}_k para mistura de Poisson para modelo truncado no zero com 5 componentes.	94

SUMÁRIO

1	INTRODUÇÃO	21
2	MODELO DE MISTURA	25
2.1	Estimativas via máxima verossimilhança	27
2.2	Algoritmo EM	28
3	MODELO POISSON	31
3.1	Distribuição Poisson	31
3.2	Distribuição Poisson truncada no zero	31
4	MISTURA DE POISSON	37
4.1	Mistura de Poisson com truncamento no zero	40
5	MISTURA DE POISSON PARA MODELO TRUNCADO NO ZERO	43
5.1	Mistura de Poisson com truncamento na resultante do modelo e no ponto zero	43
6	ESTUDO DE SIMULAÇÃO	47
6.1	Simulação dos dados	47
6.2	Implementação do Algoritmo EM	48
6.3	Discussão dos Resultados	50
6.3.1	<i>Algoritmo EMMPTZ</i>	50
6.3.2	<i>Algoritmo EMPMTZ</i>	54
7	APLICAÇÃO	57
7.1	Descrição dos dados	57
7.2	Algoritmo EMMPTZ	58
7.3	Algoritmo EMPMTZ	60
7.4	Comparação dos modelos ajustados	62
8	DISCUSSÃO	67
	REFERÊNCIAS	69
	APÊNDICE A RESULTADOS DO ALGORITMO EMMPTZ	71

APÊNDICE B	RESULTADOS DO ALGORITMO EMPMTZ	87
APÊNDICE C	CÓDIGOS EM R	95
C.1	Códigos para Gráfico de Mistura de Poisson	95
C.2	Códigos para Poisson truncada no zero	96
C.3	Códigos para Mistura de Poisson truncadas no zero	97
C.4	Códigos para mistura de Poisson com variável truncada no zero . .	100

INTRODUÇÃO

No cenário de agrupamentos e segmentações pode se discorrer em análises estatísticas mais simplificadas como distância e similaridade, análise discriminante e outras técnicas multivariadas mais robustas. Todavia, a maioria dessas técnicas são válidas somente para os casos de independência entre as variáveis e para a distribuição normal, como estudado por [Anderson *et al.* \(1958\)](#).

O problema de discriminação e classificação de informações em geral tem ganho grande importância devido a facilidade que encontramos atualmente em acumular grandes massas de dados. Resumir essas informações, de forma que as discrimine e então classifique, define segmentos com comportamentos distintos.

[McLachlan e Basford \(1988\)](#) abordam o modelo de mistura também como uma metodologia de agrupamentos, uma vez que cada distribuição do modelo representa cada grupo ou segmento. Além disso, o modelo de mistura pode ser ajustado para variáveis que seguem distintas distribuições e para variáveis dependentes, de acordo com [Frühwirth-Schnatter \(2006\)](#).

Em um modelo de mistura o conjunto de dados é considerado proveniente de uma população composta por K subpopulações, em que cada subpopulação é modelada por uma distribuição de probabilidade. Cada grupo de dados se aglutina em função de sua distribuição, ou seja, dados com características semelhantes entre si formam conjuntos de indivíduos – ou alguma outra unidade de interesse – sendo que cada grupo apresenta uma distribuição.

O conjunto de dados que motivou este estudo é originário da área comercial, em que a classificação de clientes tem grande importância para abordagem publicitária. A distribuição Poisson é forte candidata para modelar dados de contagem de produtos vendidos, no entanto, são informações nas quais não se observa o valor zero na contagem das vendas. Desse modo, fez-se uso da distribuição Poisson truncada no zero, como introduzida por [David e Johnson \(1952\)](#).

O objetivo inicial é estimar não apenas os parâmetros das distribuições do modelo de

mistura, como também estimar a proporção de cada componente do modelo.

De acordo com [McLachlan e Peel \(2004\)](#) os primeiros estudos sobre modelo de mistura foram realizados com a distribuição Normal e, posteriormente, com as distribuições de mistura para a família exponencial. [McLachlan e Peel \(2004\)](#) discutem também que o método dos momentos apresentou estimativas inferiores ao método de máxima verossimilhança.

Tendo isso em vista, fez-se o uso do método de máxima verossimilhança para estimar os parâmetros de interesse. No entanto, é necessário também empregar um método iterativo, em que se considera o algoritmo EM – *expectation maximization algorithm* – discutido por [McLachlan e Krishnan \(2007\)](#), devido às dificuldades encontradas em isolar os parâmetros de interesse.

O algoritmo EM foi implementado na estimação da variável não observável S , que determina em que componente pertence cada elemento, bem como na estimação dos parâmetros de interesse. Para modelos de mistura, denominamos cada grupo como um componente da mistura. Assim como desenvolvido por [Meira \(2014\)](#), definiu-se cada componente a partir da variável S . Para isso, fez-se uso da probabilidade de S condicionada à variável de interesse.

Como solução para a restrição em que a variável não assume valor zero empregou-se a metodologia discutida por [Dietz e Böhning \(2000\)](#). E, alternativamente, estudou-se o caso do truncamento como discutido por [III e Pollock \(1996\)](#), [Pledger \(2000\)](#) e [Link \(2003\)](#).

O objetivo desta dissertação é calcular e explorar as estimativas dos parâmetros do modelo de mistura de Poisson para os casos com truncamento no zero, dispondo de dois tipos de modelo: o modelo de mistura de distribuições Poisson truncadas no zero e o modelo de mistura de Poisson com truncamento na resultante.

O modelo de mistura é definido formalmente no [Capítulo 2](#) e são formuladas as estimativas dos parâmetros do modelo para o método de máxima verossimilhança, associado ao algoritmo EM.

No [Capítulo 3](#) introduz-se a distribuição Poisson, assim como a distribuição Poisson truncada no zero, considerando os casos para a variável assumindo valor zero como valor faltante.

O modelo de mistura de Poisson é apresentado no [Capítulo 4](#), bem como o modelo de mistura de Poisson truncadas no zero e as estimativas dos parâmetros de interesse.

Como outro caminho para trabalhar com o caso do truncamento, a metodologia discutida por [III e Pollock \(1996\)](#), [Pledger \(2000\)](#) e [Link \(2003\)](#) foi abordada no [Capítulo 5](#), em que o truncamento é realizado na resultante do modelo de mistura e o modelo é definido com distribuições não truncadas.

No [Capítulo 6](#) as duas metodologias temas, [Capítulo 4](#) e [Capítulo 5](#), foram exploradas através de um estudo de simulação. Os dados que motivaram este estudo foram apresentados de forma breve através de análise exploratória no [Capítulo 7](#). No qual, as teorias discutidas foram aplicadas e examinados os resultados obtidos. Para finalizar, no [Capítulo 8](#), as resoluções das

metodologias desta pesquisa são comentadas apontando os principais resultados.

MODELO DE MISTURA

Um modelo estatístico permite traçar um padrão geral para os dados; no entanto, tais dados podem pertencer a uma população composta de subpopulações, dificultando a modelagem. Dessa maneira, pode se utilizar múltiplas distribuições de probabilidade em que cada uma descreve o comportamento de cada subpopulação. Estas distribuições, geralmente pertencentes à mesma família, formam o modelo de mistura de distribuições. Tal técnica permite modelar os dados com maior precisão, pois a mistura de K distribuições tem maior flexibilidade do que apenas uma distribuição de probabilidade, como discutido por [McLachlan e Peel \(2004\)](#).

[McLachlan e Peel \(2004\)](#) definem mistura de distribuições para a variável aleatória X por

$$f(x; \mathbf{p}, \boldsymbol{\theta}) = \sum_{k=1}^K p_k f_k(x_k; \boldsymbol{\theta}_k) \quad (2.1)$$

em que $f_k(x_k; \boldsymbol{\theta}_k)$ é a função de densidade do k -ésimo componente da mistura, $\boldsymbol{\theta}_k$ são os parâmetros da k -ésima componente, $p_k > 0$ a proporção de cada componente com restrição $\sum_{k=1}^K p_k = 1$, $0 \leq p_k \leq 1$, para $k = 1, \dots, K$ e K a quantidade de componentes do modelo.

No caso contínuo, pode se escrever a função de densidade da variável X em termos de probabilidades condicionais de uma determinada variável S não observável, como discutido no Exemplo 1.

Exemplo 1: Considere o experimento no qual foi escolhido ao acaso uma amostra com tamanho N de cisnes brancos – *Cygnus olor* – e medimos o peso para cada ave. Definiu se o vetor de observações de pesos $x = (x_1, x_2, \dots, x_N)$ para $x_i \in (4, 30)$ Kg.

Essa espécie de cisne é de penagem inteira branca e com o bico laranja, sendo dificilmente diferenciados entre fêmeas e machos sem um especialista. Assim, definiu se S_i como a variável aleatória não observável

$$S_i = \begin{cases} 1, & \text{se o } i\text{-ésimo cisne é fêmea} \\ 2, & \text{se o } i\text{-ésimo cisne é macho.} \end{cases}$$

Considerando que as fêmeas pesam em média 8 Kg e os machos 12 Kg pode se inferir sobre o gênero do animal com as informações de seus pesos. E formular em termos de probabilidades condicionais

$$\begin{aligned} f_{X_i}(x) &= Pr[S_i = 1] f_{X_i}(x | S_i = 1) + Pr[S_i = 2] f_{X_i}(x | S_i = 2) \\ &= \sum_{k=1}^2 Pr[S_i = k] f_{X_i}(x | S_i = k). \end{aligned} \quad (2.2)$$

Pode se estender a [Equação 2.2](#) para o caso em que a variável S possui K grupos, assumindo valores de $1, \dots, K$. Deste modo, reescreve se a [Equação 2.2](#) pela fórmula

$$f_{X_i}(x) = \sum_{k=1}^K Pr[S_i = k] f_{X_i}(x | S_i = k). \quad (2.3)$$

Com relação a soma de distribuições, [Equação 2.3](#) e [Equação 2.2](#), diz se que a função de probabilidade da variável X é uma mistura de K distribuições condicionais de X dado S , sendo que $f_{X_i}(x | S_i = k)$ é o k -ésimo componente da mistura para a i -ésima observação. Note que, se $k = 2$ a [Equação 2.3](#) é equivalente a [Equação 2.1](#) escrita em termos de probabilidades condicionais.

No Exemplo 1, a variável S assume os valores 1 ou 2 com $K = 2$, pois existem dois resultados possíveis para o gênero do cisne branco. Com tal característica, obtém se dois componentes de mistura para modelar a variável X .

Como definido por [Meira \(2014\)](#), considera se que X assume a sequência de valores $x = (x_1, \dots, x_N)$ e para a variável não observável S em que s_i assume valores s_1, \dots, s_N sendo N a quantidade total de observações. Sendo cada S_i independente de S_j para i diferente de j , em que $i, j = 1, \dots, N$. O vetor de parâmetros para a distribuição de probabilidade ou função densidade de X é dado por $\theta = (\theta_1, \dots, \theta_K)$ e para a distribuição de probabilidade de S por $(1, \mathbf{p})$ em que $\mathbf{p} = (p_1, \dots, p_K)$.

No caso em que a variável X é estritamente contínua define se a função de densidade marginal de X para cada elemento i por

$$f_{X_i}(x | \theta) = \sum_{k=1}^K Pr[S_i = k | \mathbf{p}] f_{X_i}(x | S_i = k, \theta, \mathbf{p}). \quad (2.4)$$

Similarmente, para o caso em que a variável aleatória X apresenta-se como uma variável discreta, redefinimos a [Equação 2.3](#) para cada observação i pela equação

$$Pr[X_i = x_i | \theta] = \sum_{k=1}^K Pr[S_i = k | \mathbf{p}] Pr[X_i = x_i | S_i = k, \theta, \mathbf{p}]. \quad (2.5)$$

[Meira \(2014\)](#) também define a variável S com distribuição Multinomial e vetor de parâmetros $(1, \mathbf{p})$, em que $\mathbf{p} = (p_1, \dots, p_K)$ e $\sum_{k=1}^K p_k = 1$, ou seja, $Pr[S_i = k] = p_k$. Assim, é

escrita a distribuição conjunta de X e S para o modelo de mistura, definido na [Equação 2.5](#) como

$$Pr[X_i = x_i, S_i = s_i | \theta, \mathbf{p}] = Pr[S_i = s_i | \mathbf{p}] Pr[X_i = x_i | S_i = s_i, \theta, \mathbf{p}].$$

Utilizando suas respectivas funções indicadoras obtém se

$$Pr[X_i = x_i, S_i = s_i | \theta, \mathbf{p}] = \prod_{k=1}^K p_k^{\mathbb{1}_{s_i}(k)} Pr[X_i = x_i | S_i = s_i, \theta_k]^{\mathbb{1}_{s_i}(k)} \quad (2.6)$$

em que a função indicadora $\mathbb{1}_{s_i}(k)$ assume valor 1 se $s_i = k$ for verdadeiro e 0 caso contrário.

Lembrando que para cada observação i temos um valor para s_i de S e um valor para x_i de X , assim define se os vetores $\mathbf{X} = (X_1, \dots, X_N)$ e $\mathbf{S} = (S_1, \dots, S_N)$ e calculamos a distribuição conjunta de \mathbf{X} e \mathbf{S} partindo da equação

$$Pr[\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s} | \theta, \mathbf{p}] = \prod_{i=1}^N \prod_{k=1}^K p_k^{\mathbb{1}_{s_i}(k)} Pr[X_i = x_i | S_i = s_i, \theta_k]^{\mathbb{1}_{s_i}(k)}. \quad (2.7)$$

2.1 Estimativas via máxima verossimilhança

Nesta seção, foram calculadas as estimativas dos parâmetros de interesse para os modelos de mistura, θ e \mathbf{p} , aplicando a metodologia de máxima verossimilhança.

Assim como discutido por [Meira \(2014\)](#), na [Equação 2.7](#) os produtórios foram manipulados de forma que uma das funções indicadoras, $\mathbb{1}_{s_i}(k)$, expresse a quantidade de observações, n_k , presente em cada componente do modelo de mistura. Observando que $n_k = \sum_{i=1}^N \mathbb{1}_{s_i}(k)$ é o número de elementos do k -ésimo componente da mistura reescrevemos a [Equação 2.7](#) da seguinte forma

$$\begin{aligned} Pr[\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s} | \theta, \mathbf{p}] &= \prod_{k=1}^K \prod_{i=1}^N p_k^{\mathbb{1}_{s_i}(k)} Pr[X_i = x_i | S_i = s_i, \theta_k]^{\mathbb{1}_{s_i}(k)} \\ &= \prod_{k=1}^K p_k^{\sum_{i=1}^N \mathbb{1}_{s_i}(k)} \prod_{i=1}^N Pr[X_i = x_i | S_i = s_i, \theta_k]^{\mathbb{1}_{s_i}(k)} \\ &= \prod_{k=1}^K p_k^{n_k} \prod_{i=1}^N Pr[X_i = x_i | S_i = s_i, \theta_k]^{\mathbb{1}_{s_i}(k)}. \end{aligned} \quad (2.8)$$

Através da função de verossimilhança, [Equação 2.8](#), calcula se o ponto de máximo para cada um dos parâmetros de interesse, resultando em estimativas para $\theta_1, \dots, \theta_K$ e p_1, \dots, p_K . Portanto, a [Equação 2.8](#) pode ser reescrita como

$$L(\theta, \mathbf{p} | \mathbf{x}, \mathbf{s}) = \prod_{k=1}^K p_k^{n_k} \prod_{i=1}^N Pr[X_i = x_i | S_i = s_i, \theta_k]^{\mathbb{1}_{s_i}(k)}. \quad (2.9)$$

O ponto que maximiza a função de verossimilhança também maximiza o logaritmo da função de verossimilhança, portanto calcula se

$$l(\theta, \mathbf{p} | \mathbf{x}, \mathbf{s}) = \ln \left(\prod_{k=1}^K p_k^{n_k} \prod_{i=1}^N Pr[X_i = x_i | S_i = s_i, \theta_k]^{\mathbb{1}_{s_i}(k)} \right). \quad (2.10)$$

Devido a restrição de que $\sum_{k=1}^K p_k = 1$ Meira (2014) utiliza multiplicadores de Lagrange para a solução da Equação 2.10 e obtém

$$l(\theta, \mathbf{p} | \mathbf{x}, \mathbf{s}) = \sum_{k=1}^K n_k \ln(p_k) + \sum_{k=1}^K \sum_{i=1}^N \mathbb{1}_{s_i}(k) \ln(\Pr[X_i = x_i | S_i = s_i, \theta_k]) + \zeta \left(1 - \sum_{k=1}^K p_k \right) \quad (2.11)$$

em que $\zeta = \sum_{k=1}^K n_k = N$.

A derivada da Equação 2.11 com relação a θ_k é dada por

$$\frac{\partial l(\theta, \mathbf{p} | \mathbf{x}, \mathbf{s})}{\partial \theta_k} = \sum_{i=1}^N \frac{\mathbb{1}_{s_i}(k)}{\Pr[X_i = x_i | S_i = s_i, \theta_k]} \frac{\partial \Pr[X_i = x_i | S_i = s_i, \theta_k]}{\partial \theta_k}.$$

Igualando a derivada obtida a zero e isolando θ_k encontra se

$$\sum_{i=1}^N \frac{\mathbb{1}_{s_i}(k)}{\Pr[X_i = x_i | S_i = s_i, \theta_k]} \frac{\partial \Pr[X_i = x_i | S_i = s_i, \theta_k]}{\partial \theta_k} \Bigg|_{\theta_k = \hat{\theta}_k} = 0. \quad (2.12)$$

Com relação ao parâmetro p_k , formula se

$$\frac{\partial l(\theta, \mathbf{p} | \mathbf{x}, \mathbf{s})}{\partial p_k} = n_k \frac{1}{p_k} - \zeta$$

e

$$n_k \frac{1}{\hat{p}_k} - \zeta = 0$$

e desta forma, a estimativa de p_k é dada por

$$\hat{p}_k = \frac{n_k}{N}. \quad (2.13)$$

Repare que as estimativas de \mathbf{p} e θ dependem da variável não observável S e para a solução deste problema recorre se ao algoritmo EM. A seguir discute se o algoritmo EM no caso geral e para o modelo de mistura.

2.2 Algoritmo EM

O algoritmo EM – *expectation maximization* – é uma forma de cálculo iterativo para encontrar a solução de estimativas de parâmetros via máxima verossimilhança, utilizado para dados não observáveis, dados incompletos, observações com censura, distribuições truncadas entre outros. Existe uma grande diversidade de remodelamentos do algoritmo que podem ser consultados em McLachlan e Krishnan (2007).

Com a finalidade de iniciar o processo iterativo foram estabelecidos valores iniciais para as informações faltantes. Assim, a técnica administra os dados incompletos como dados

completos tornando-o de fácil aplicação computacional. Este recurso é feito estabelecendo relação entre as funções de verossimilhança obtidas em cada iteração.

A metodologia se organiza na repetição dos passos E e M até que a convergência ocorra. Considerando a $(j+1)$ -ésima iteração devemos encontrar, no passo E, o valor esperado do logaritmo da função de verossimilhança, $E[l(\theta)]$, usando o cálculo mais recente de $\theta^{(j)}$ para θ . Isto é realizado substituindo θ_i em $l(\theta)$ por $\theta_i^{(j)}$. No passo M escolhemos $\theta^{(j+1)}$ que maximiza a quantidade $Q(\theta, \theta^{(j)}) = E[l(\theta)]$, com respeito ao parâmetro θ .

A convergência é alcançada dentro de duas restrições. Quando a diferença entre os parâmetros obtidos na j -ésima e na $(j+1)$ -ésima iterações é próxima de zero e quando a diferença entre as quantidades Q obtidas na j -ésima e na $(j+1)$ -ésima iterações é próxima de zero.

Tomando os passos descritos acima, podemos formular o algoritmo em etapas como apresentado no [Algoritmo 1](#).

Algoritmo 1 – Algoritmo EM

- 1: inicialize $\theta^{(0)}$
 - 2: para $j = 1, 2, 3, \dots$ ▷ j -ésima iteração
 - 3: **enquanto** $|\theta^{(j+1)} - \theta^{(j)}| \geq \tau_\theta$
 - 4: **ou** $|Q(\theta^{(j+1)}, \theta^{(j)}) - Q(\theta^{(j)}, \theta^{(j)})| \geq \tau_\theta$ **faça** ▷ critério de parada
 - 5: definir $Q(\theta, \theta^{(j)}) = E[l(\theta)]$ com θ_i substituído por $\theta_i^{(j)}$ ▷ Passo E
 - 6: escolher $\theta^{(j+1)}$ que maximiza $Q(\theta, \theta^{(j)}) = E[l(\theta)]$ com respeito a θ ▷ Passo M
 - 7: **fim enquanto**
-

Dentro da metodologia de modelos de mistura é necessário fazer adaptações no algoritmo EM. De modo que S é uma variável não observável, como já definida, é necessário estimar valores para cada s_i .

Para o cálculo da atualização de s_i aplicou-se o teorema de *Bayes*, o teorema das probabilidades totais e a distribuição conjunta de X e S obtida na [Equação 2.8](#), resultando na probabilidade condicional de S dado X , como discutido por [Meira \(2014\)](#). Primeiramente, a distribuição marginal de X é manipulada, pode-se realizar o seguinte cálculo

$$\begin{aligned}
 Pr[X_i = x_i | \theta] &= \sum_{j=1}^K Pr[X_i = x_i, S_i = j | \theta, \mathbf{p}] \\
 &= \sum_{j=1}^K Pr[S_i = j | \mathbf{p}] Pr[X_i = x_i | S_i = j, \theta, \mathbf{p}] \\
 &= \sum_{j=1}^K p_j Pr[X_i = x_i | S_i = j, \theta, \mathbf{p}].
 \end{aligned} \tag{2.14}$$

Utilizando o teorema de *Bayes* e a [Equação 2.8](#), obtemos a distribuição condicional de X

dado S em

$$Pr[S_i = k | X_i = x_i, \theta, \mathbf{p}] = \frac{Pr[S_i = k, X_i = x_i | \theta, \mathbf{p}]}{\sum_{j=1}^K Pr[S_i = j, X_i = x_i | \theta, \mathbf{p}]}$$

para $i = 1, \dots, N$ e $j = 1, \dots, K$ e podemos escrever

$$Pr[S_i = k | X_i = x_i, \theta, \mathbf{p}] = \frac{p_k Pr[X_i = x_i | S_i = k, \theta, \mathbf{p}]}{\sum_{j=1}^K p_j Pr[X_i = x_i | S_i = j, \theta, \mathbf{p}]} \quad (2.15)$$

O algoritmo EM usual, [Algoritmo 1](#), maximiza a quantidade $Q(\theta, \theta^{(j)})$ ao passo que o algoritmo EM modificado faz uso da função argumento máximo para determinar os valores das probabilidades condicionais de S dado X , [Equação 2.15](#). Ou seja, para se obter \hat{s}_i calcula se a probabilidade do i -ésimo elemento pertencer a cada componente da mistura e escolhe se \hat{s}_i para a maior probabilidade resultante dentre os K componentes. Assim a quantidade $s_i^{(j)}$ é atualizada por $\hat{s}_i^{(j)}$.

Na $(j+1)$ -ésima iteração é utilizado o valor estimado $\hat{s}_i^{(j)}$ para computar as estimativas de máxima verossimilhança, $\hat{\mathbf{p}}^{(j)}$ e $\hat{\theta}^{(j)}$, dos parâmetros do modelo de mistura, obtidos através da [Equação 2.13](#) e da [Equação 2.12](#), respectivamente.

Os passos descritos devem ser repetidos até que a convergência ocorra. Para isso, foi estipulado um valor para τ_p e para τ_θ embasado na variação máxima das estimativas obtidas para os parâmetros \mathbf{p} e θ entre a j -ésima iteração e a $(j+1)$ -ésima iteração. Fixamos as medidas τ_p e τ_θ em 10^{-4} como tolerância máxima para essa variação.

Logo, escreve se o algoritmo EM modificado para o modelo de mistura, dispondo da probabilidade condicional de S dado X na atualização da variável não observável.

Algoritmo 2 – Algoritmo EM para Modelos de Mistura

- 1: atribua um vetor de valores iniciais para $\mathbf{s}^{(0)}$ tal que $s_i^{(0)} \in \{1, \dots, K\}$ para $i = 1, \dots, N$
 - 2: inicialize $\hat{\theta}_k^{(0)}$ e $\hat{p}_k^{(0)}$
 - 3: para $j = 1, 2, \dots$ ▷ j -ésima iteração
 - 4: **enquanto** $\max_k \left\{ |\hat{p}_k^{(j)} - \hat{p}_k^{(j+1)}| \right\} \geq \tau_p$
 - 5: **ou** $\max_k \left\{ |\hat{\theta}_k^{(j)} - \hat{\theta}_k^{(j+1)}| \right\} \geq \tau_\theta$ **faça** ▷ critério de parada
 - 6: estime $\theta_k^{(j)}$, [Equação 2.12](#), e $p_k^{(j)}$, [Equação 2.13](#)
 - 7: calcule $\hat{s}_i^{(j)} = \operatorname{argmax}_k \left\{ Pr[S_i = k | X_i = x_i, \hat{\theta}, \hat{\mathbf{p}}] \right\}$, [Equação 2.15](#)
 - 8: **fim enquanto**
-

Maiores detalhes em [McLachlan e Krishnan \(2007\)](#), [Bohning, Schlattmann e Lindsay \(1992\)](#) entre outros estudos encontrados na literatura sobre o método EM e método EM para modelos de mistura.

MODELO POISSON

A distribuição Poisson é comumente usada para descrever dados em que a variável ajustada seja do tipo discreta e tenha valores inteiros e positivos, como os fenômenos que envolvem contagem.

3.1 Distribuição Poisson

Uma variável aleatória X assume distribuição Poisson com parâmetro θ , $X \sim Pois(\theta)$, se sua função de probabilidade é dada por

$$Pr[X = x | \theta] = \frac{e^{-\theta} \theta^x}{x!} \quad (3.1)$$

em que $x \in \{0, 1, 2, 3, \dots\}$ e $\theta > 0$, ou seja uma constante positiva. A distribuição de probabilidade definida na [Equação 3.1](#) é discreta com média e variância iguais a θ , isto é, $E[X] = Var[X] = \theta$.

A função de verossimilhança, como descrito por [Bickel e Doksum \(2015\)](#), é utilizada para obtenção das estimativas dos parâmetros de interesse, calculando o valor que a maximiza. Dado isso, o estimador via máxima verossimilhança para o parâmetro θ é dado por

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \quad (3.2)$$

assim como definido por [Casella e Berger \(2002\)](#).

3.2 Distribuição Poisson truncada no zero

Definiu-se a distribuição truncada no ponto a para uma variável X , tal como foi definida por [Grogger e Carson \(1991\)](#), pela equação

$$f_a(x_i) = \frac{f(x_i)}{1 - F(a)}, \quad (3.3)$$

em que $f(\cdot)$ é a função densidade de X e $F(a)$ é a função de distribuição acumulada de X no ponto a .

Considerando X variável aleatória com distribuição Poisson, definida na [Equação 3.1](#), em que x assume valores inteiros e positivos. No entanto, existem casos nos quais a variável não assume valor zero, sendo que na Poisson $Pr[X \leq 0] = Pr[X = 0] = e^{-\theta}$. Desse modo, com a [Equação 3.3](#) modifica se a distribuição Poisson, definida na [Equação 3.1](#), de forma que x não assuma o valor 0 e obtém se

$$Pr[X = x | \theta] = \frac{1}{1 - Pr[X \leq 0]} \frac{e^{-\theta} \theta^x}{x!}.$$

Logo, diz se que uma variável aleatória X tem distribuição Poisson truncada no zero de parâmetro θ , $X \sim PoisTrun(\theta)$, se sua função de probabilidade é dada por

$$\begin{aligned} Pr[X = x | \theta] &= \frac{e^{-\theta} \theta^x}{(1 - e^{-\theta}) x!} \\ &= \frac{1}{(e^\theta - 1)} \frac{\theta^x}{x!} \end{aligned} \quad (3.4)$$

para $x \in \{1, 2, 3, \dots\}$ e $\theta > 0$.

Considerando a [Equação 3.4](#) calcula se o valor esperado para a variável aleatória X

$$\begin{aligned} E[X] &= \sum_{i=1}^n x_i Pr(X = x_i | \theta) \\ &= \frac{1}{1 - e^{-\theta}} \sum_{i=1}^n \frac{x_i}{x_i!} e^{-\theta} \theta^{x_i} \\ &= \frac{\theta}{1 - e^{-\theta}}. \end{aligned} \quad (3.5)$$

E a variância de X é dada por

$$\begin{aligned} Var[X] &= E[X^2] - (E[X])^2 \\ &= \frac{\theta + \theta^2}{1 - e^{-\theta}} - \frac{\theta^2}{(1 - e^{-\theta})^2}. \end{aligned} \quad (3.6)$$

As formulações obtidas para o valor esperado e para a variância de uma variável aleatória com distribuição Poisson truncada no zero, [Equação 3.5](#) e [Equação 3.6](#), respectivamente, indicam diferenças quando comparadas às mesmas estatísticas da distribuição Poisson sem o truncamento.

De maneira que a [Equação 3.5](#) e a [Equação 3.6](#) dependem apenas do valor do parâmetro θ , foram fixados alguns valores para θ e calculadas a esperança e a variância de X , expostos na [Tabela 1](#). Nota se que quanto maior o valor do parâmetro θ a esperança e a variância se aproximam de θ , assim como na distribuição Poisson em que $E[X] = Var[X] = \theta$. Em outras palavras, pode se afirmar que o efeito do truncamento no zero é maior para valores pequenos de θ , como para $\theta = 1$ e $\theta = 2$, sendo que para $\theta \geq 8$ o efeito do truncamento é praticamente nulo,

Tabela 1 – Valores da Esperança e da Variância de X , sendo $X \sim PoisTrun(\theta)$, para diversos valores de θ .

θ_{real}	$E[X]$	$Var[X]$
1	1,581977	0,6613031
2	2,313035	1,5889736
3	3,571870	2,6609180
4	4,074629	3,7705421
5	5,033918	4,8631765
6	6,014909	5,9252304
7	7,006389	6,9616252
8	8,002685	7,9812006
9	9,001111	8,9911122
10	10,000454	9,9959136
11	11,000184	10,9981627
12	12,000074	11,9991890

Fonte: Elaborada pelo autor.

pensando em 3 casas decimais. Além do mais, o truncamento no zero influencia a variabilidade de X , pois foi encontrado $E[X] > Var[X]$ para $1 \leq \theta \leq 37$.

O efeito do truncamento no zero deve ser avaliado também no parâmetro de interesse θ , pois as estimativas usuais – por exemplo, via máxima verossimilhança dada na [Equação 3.2](#) – do parâmetro da distribuição Poisson podem estar com algum viés, como observado por [David e Johnson \(1952\)](#). Para isso, usa-se a função de verossimilhança

$$\begin{aligned}
 L(\theta | \mathbf{x}) &= \prod_{i=1}^n \frac{e^{-\theta}}{(1 - e^{-\theta})} \frac{\theta^{x_i}}{x_i!} \\
 &= \left(\frac{e^{-\theta}}{1 - e^{-\theta}} \right)^n \frac{\theta^{\left(\sum_{i=1}^n x_i\right)}}{\prod_{i=1}^n x_i!}.
 \end{aligned} \tag{3.7}$$

O logaritmo da função de verossimilhança é

$$\begin{aligned}
 l(\theta | \mathbf{x}) &= n \ln \left(\frac{e^{-\theta}}{1 - e^{-\theta}} \right) + \ln(\theta) \sum_{i=1}^n x_i - \ln \left(\prod_{i=1}^n x_i! \right) \\
 &= -n\theta - n \ln(1 - e^{-\theta}) + \ln(\theta) \sum_{i=1}^n x_i - \ln \left(\prod_{i=1}^n x_i! \right).
 \end{aligned} \tag{3.8}$$

Para a derivada de $l(\theta | \mathbf{x})$ com relação ao parâmetro θ , calcula-se

$$\begin{aligned}
 \frac{\partial l(\theta | \mathbf{x})}{\partial \theta} &= -n + \frac{1}{\theta} \sum_{i=1}^n x_i - n \frac{e^{-\theta}}{(1 - e^{-\theta})} \\
 &= \frac{1}{\theta} \sum_{i=1}^n x_i - \frac{n}{(1 - e^{-\theta})}.
 \end{aligned} \tag{3.9}$$

Para se obter o valor candidato a maximizar a função de verossimilhança, [Equação 3.7](#), iguala se o resultado da [Equação 3.9](#) a zero

$$\frac{1}{\hat{\theta}} \sum_{i=1}^n x_i - \frac{n}{(1 - e^{-\hat{\theta}})} = 0. \quad (3.10)$$

A [Equação 3.10](#) indica dificuldade em isolar $\hat{\theta}$. Uma solução para esse problema é reescrevê-la de forma que expresse uma função recursiva do parâmetro como

$$\hat{\theta}^{(j)} = \left(\frac{1 - e^{-\hat{\theta}^{(j-1)}}}{n} \right) \sum_{i=1}^n x_i \quad (3.11)$$

para $j = 1, 2, 3, \dots$

Como opção alternativa a [Equação 3.11](#) pode se utilizar a técnica proposta por [Dietz e Böhning \(2000\)](#). Trata se o tamanho da amostra n_0 em que x assume o valor zero como valores faltantes, n o tamanho da amostra para $x = \{1, 2, 3, \dots\}$ e utiliza-se a estimativa \hat{n} , em que $\hat{n} = n + n_0$, como uma amostra aumentada.

Para estimar a quantidade de valores faltantes n_0 parte se do cálculo de seu valor esperado condicionado à θ e n . [Dietz e Böhning \(2000\)](#) utilizam a distribuição Binomial Negativa com parâmetros n e $(1 - e^{-\theta})$ para o número de observações zero dado θ e n . Assim, escreve se o valor esperado do número de observações dado θ e n

$$E[n_0 | \theta, n] = \frac{n e^{-\theta}}{1 - e^{-\theta}}. \quad (3.12)$$

Desenvolvendo o valor esperado para o tamanho da amostra, considerando θ e n conhecidos, calcula se

$$\begin{aligned} \hat{n} &= n + E[n_0 | \theta, n] \\ &= n + \frac{n e^{-\theta}}{1 - e^{-\theta}} \\ &= \frac{n}{1 - e^{-\theta}}. \end{aligned} \quad (3.13)$$

Através da [Equação 3.11](#) e da [Equação 3.13](#) o estimador de θ pode ser reescrito em

$$\hat{\theta} = \frac{1}{\hat{n}} \sum_{i=1}^n x_i. \quad (3.14)$$

Com a [Equação 3.14](#) pode se realizar a aplicação iterativa para estimação de θ e para estimar os valores faltantes, n_0 . Na implementação do algoritmo esta formulação foi preferida

$$E[n_0 | \theta, n] = \frac{n}{e^{\theta} - 1} \quad (3.15)$$

para a [Equação 3.12](#) com a finalidade de diminuir a atividade computacional. Além disso, foi utilizada a função *round* do **R** para que o resultado obtido na [Equação 3.15](#) seja um número inteiro.

Utiliza-se o algoritmo EM, como definido no [Capítulo 2](#), sendo necessárias algumas alterações. Para a $(j+1)$ -ésima iteração, no passo E computar o número esperado de zeros da distribuição Poisson truncada. No passo M, estimar via máxima verossimilhança o parâmetro de interesse θ , aplicado para as observações positivas e diferentes de zero e atualizar com o número esperado de zeros, obtido na [Equação 3.14](#).

Deve-se repetir os passos E e M até que a convergência ocorra. Para isso, considera-se como critério de parada o módulo da diferença dos parâmetros estimados entre a j -ésima e a $(j+1)$ -ésima iteração menor do que 10^{-4} .

[Dietz e Böhning \(2000\)](#) discutem que $\hat{\theta}^{(j+1)}$, [Equação 3.14](#), converge para algum valor $\hat{\theta}$ para a estimativa de máxima verossimilhança satisfazendo $\hat{\theta}$, [Equação 3.11](#), dado o valor esperado da quantidade total de observações, [Equação 3.13](#). A única suposição é que pelo menos um elemento da amostra deve ser positivo e diferente de zero.

Algoritmo 3 – Algoritmo EM para Poisson truncada no zero

- 1: atribua um valor inicial $\theta^{(0)}$ para θ e para $n_0^{(0)}$
 - 2: para $j = 1, 2, \dots$ ▷ j -ésima iteração
 - 3: **enquanto** $|\hat{\theta}^{(j)} - \hat{\theta}^{(j+1)}| \geq \tau_\theta$ **faça** ▷ critério de parada
 - 4: obtenha $\hat{\theta}^{(j)}$, [Equação 3.14](#)
 - 5: calcule $E[n_0 | \theta, n]$, [Equação 3.15](#)
 - 6: **fim enquanto**
-

A variância de $\hat{\theta}$ foi obtida através da Informação de Fisher, representada por $I[\theta]$ e definida por [Bickel e Doksum \(2015\)](#), e foi escrito

$$\begin{aligned}
 \text{Var}[\hat{\theta}] &= (I[\theta])^{-1} \\
 &= - \left(E \left[\frac{\partial^2 l(\theta | x)}{\partial \theta^2} \right] \right)^{-1} \\
 &= \frac{1}{n} \frac{\theta (1 - e^{-\theta})^2}{(1 - e^{-\theta} - \theta e^{-\theta})}
 \end{aligned} \tag{3.16}$$

em que $l(\theta | \mathbf{x})$ é o logaritmo da função determinada pela [Equação 3.7](#), n é o total de observações da amostra e parâmetro $\theta > 0$, como conceituado por [David e Johnson \(1952\)](#). Ao passo que a variância para a variável aleatória X , [Equação 3.6](#), depende apenas do parâmetro de interesse θ , a variância do estimador do parâmetro depende também do tamanho da amostra n , definida na [Equação 3.16](#).

MISTURA DE POISSON

Neste capítulo especificou-se o modelo de mistura para a distribuição Poisson e para a distribuição Poisson truncada no zero. As estimativas dos parâmetros de interesse foram calculadas via máxima verossimilhança, no qual será necessário novamente a implementação do algoritmo EM.

O modelo de mistura de distribuições Poisson, $X_k \sim Pois(\theta_k)$, foi definido pela equação

$$Pr[X_i = x_i | \mathbf{p}, \boldsymbol{\theta}] = \sum_{k=1}^K p_k \frac{e^{-\theta_k} \theta_k^{x_i}}{x_i!} \quad (4.1)$$

em que $x_i = \{0, 1, 2, \dots\}$, $\sum_{k=1}^K p_k = 1$, $\theta_k > 0$ e $k = 1, \dots, K$.

Para o cálculo dos estimadores dos parâmetros do modelo de mistura via máxima verossimilhança o logaritmo da função de verossimilhança foi utilizado para a [Equação 4.1](#) e pode-se escrever

$$\begin{aligned} l(\boldsymbol{\theta}, \mathbf{p} | X, S) &= \sum_{k=1}^K n_k \ln(p_k) + \sum_{i=1}^N \sum_{k=1}^K \ln \left(\frac{e^{-\theta_k} \theta_k^{x_i}}{x_i!} \right) \mathbb{1}_{s_i}(k) + \zeta \left(1 - \sum_{k=1}^K p_k \right) \\ &= \sum_{k=1}^K n_k \ln(p_k) - \sum_{k=1}^K n_k \theta_k + \sum_{k=1}^K \sum_{i=1}^N x_i \ln(\theta_k) \mathbb{1}_{s_i}(k) \\ &\quad - \sum_{i=1}^N \ln(x_i!) \mathbb{1}_{s_i}(k) + \zeta \left(1 - \sum_{k=1}^K p_k \right). \end{aligned} \quad (4.2)$$

Derivando $l(\boldsymbol{\theta}, \mathbf{p} | X, S)$, obtida na [Equação 4.2](#), primeiramente em relação a θ_k tem-se

$$\frac{\partial l(\boldsymbol{\theta}, \mathbf{p} | X, S)}{\partial \theta_k} = \frac{1}{\theta_k} \sum_{i=1}^N x_i \mathbb{1}_{s_i}(k) - n_k \quad (4.3)$$

e derivando a [Equação 4.2](#) com relação a p_k

$$\frac{\partial l(\boldsymbol{\theta}, \mathbf{p} | X, S)}{\partial p_k} = \frac{1}{p_k} - \zeta = \frac{1}{p_k} - \sum_{k=1}^K n_k. \quad (4.4)$$

Através da [Equação 4.3](#) e da [Equação 4.4](#) obtemos as estimativas, $\hat{\theta}$ e \hat{p} , respectivamente, para o k -ésimo componente da mistura via máxima verossimilhança, calculou se

$$\frac{1}{\hat{\theta}_k} \sum_{i=1}^N x_i \mathbb{1}_{s_i}(k) - n_k = 0$$

$$\hat{\theta}_k = \frac{1}{n_k} \sum_{i=1}^N x_i \mathbb{1}_{s_i}(k) \quad (4.5)$$

e

$$\frac{n_k}{\hat{p}_k} - \sum_{l=1}^K n_l = 0$$

$$\hat{p}_k = \frac{n_k}{\sum_{l=1}^K n_l} = \frac{n_k}{N} \quad (4.6)$$

em que $l = 1, \dots, K$.

Note que a estimativa $\hat{\theta}_k$ é a média do k -ésimo componente e \hat{p}_k é a proporção de elementos do k -ésimo componente dentre todos os N elementos de X . Assim como no modelo de mistura para o caso geral, como definido no [Capítulo 2](#), as estimativas $\hat{\theta}_k$ e \hat{p}_k , dependem do valor da variável não observável s_i , para $i = 1, \dots, N$. Para estimar s_i maximiza se a [Equação 2.15](#) para a distribuição Poisson.

A seguir o algoritmo EM foi apresentado dentro das particularidades descritas neste capítulo para o modelo de mistura de Poisson. A diferença para o algoritmo EM descrito no [Capítulo 2](#) está na estimação do vetor de parâmetros θ e na estimação da variável não observável S .

Algoritmo 4 – Algoritmo EM para Mistura de Poisson

- 1: atribua um vetor de valores iniciais para $\mathbf{s}^{(0)}$ restrição $s_i^{(0)} \in \{1, \dots, K\}$ para $i = 1, \dots, N$
 - 2: para $j = 1, 2, \dots$ ▷ j -ésima iteração
 - 3: **enquanto** $\max_k \left\{ \left| \hat{p}_k^{(j)} - \hat{p}_k^{(j+1)} \right| \right\} \geq \tau_p$
 - 4: **ou** $\max_k \left\{ \left| \hat{\theta}_k^{(j)} - \hat{\theta}_k^{(j+1)} \right| \right\} \geq \tau_\theta$ **faça** ▷ critério de parada
 - 5: calcule $\hat{\theta}^{(j)}$, [Equação 4.5](#), e $\hat{p}^{(j)}$, [Equação 4.6](#)
 - 6: estime $s_i^{(j)} = \operatorname{argmax}_k \left\{ Pr[S_i = k | X_i = x_i, \hat{\theta}, \hat{p}] \right\}$
 - 7: **fim enquanto**
-

Para ilustrar um modelo de mistura de Poisson com 3 componentes foi apresentado na [Figura 1](#) com o histograma de dados gerados para a família Poisson e cada componente da mistura foi representado por diferentes símbolos. Os parâmetros foram fixados em $\theta = 3$ com a ilustração dos círculos, $\theta = 9$ para os quadrados e $\theta = 15$ representado pelos triângulos. Desse modo, pode se visualizar as informações disponíveis de uma forma geral bem como cada componente da mistura.

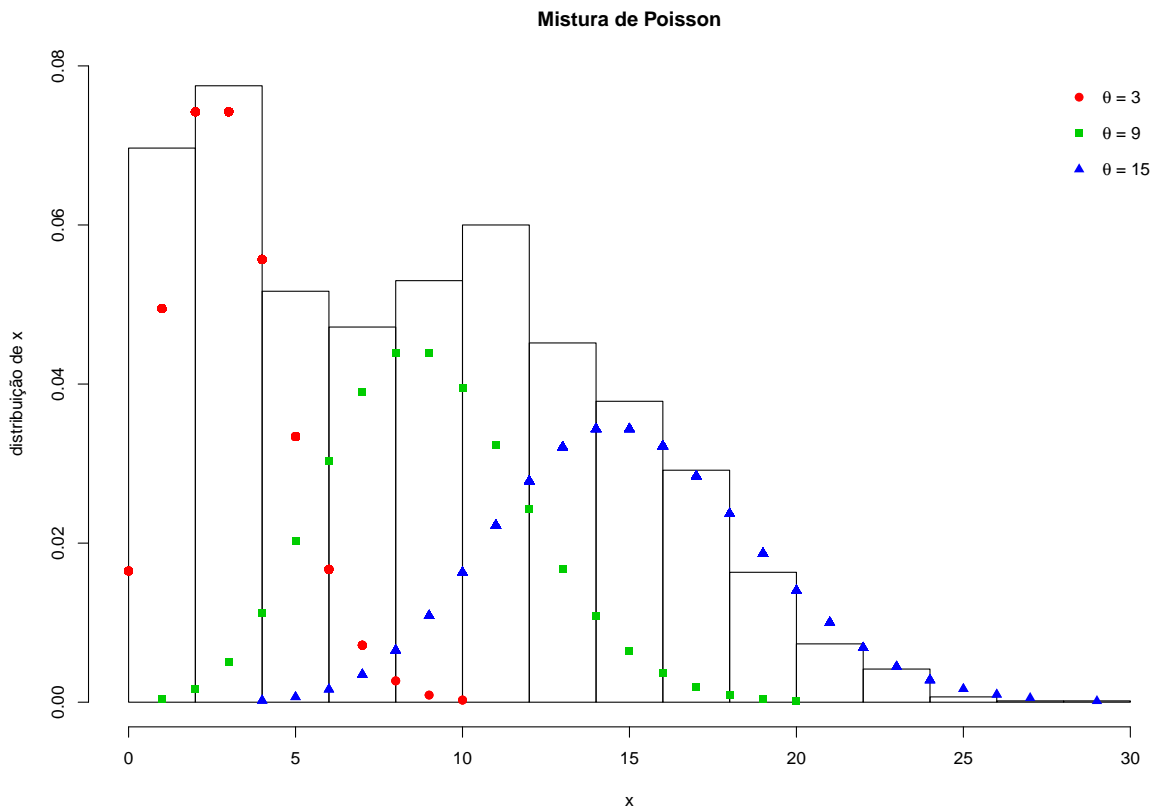


Figura 1 – Gráfico da distribuição de probabilidade Poisson para dados simulados com diferentes valores do parâmetro θ , ilustrando uma mistura de Poisson.

Fonte: Elaborada pelo autor.

O tamanho de amostra $N = 3.000$ foi utilizado de forma que cada grupo teve a mesma quantidade de elementos na amostra, ou seja, cada componente da mistura foi composto de aproximadamente 1.000 observações.

Para isso, definiu-se a variável U reproduzida a partir da distribuição Uniforme com parâmetros 0 e 1, $U \sim Unif(0, 1)$, e tamanho N . Em seguida, determinou-se cada grupo a partir das probabilidades

$$Pr[S = s_i] = \begin{cases} 0 \leq u_i < \frac{1}{3} & \text{para } s_i = 1 \\ \frac{1}{3} \leq u_i < \frac{2}{3} & \text{para } s_i = 2 \\ \frac{2}{3} \leq u_i < 1 & \text{para } s_i = 3 \end{cases}$$

em que $i = 1, 2, \dots, N$. Isto posto, amostrou-se cada componente da mistura com distribuição e parâmetros especificados.

Os comandos em **R** para a simulação dos dados e desenho do gráfico para a [Figura 1](#) encontram-se na [Seção C.1](#).

Para maiores detalhes para mistura de Poisson pode-se consultar [Karlis e Xekalaki \(2005\)](#), [Wedel et al. \(1993\)](#) entre outros.

4.1 Mistura de Poisson com truncamento no zero

O objetivo aqui foi ajustar um modelo de mistura em que os componentes se adequam à distribuição Poisson sem a ocorrência de valores zero para a variável de interesse.

Sendo $X_k \sim \text{PoisTrun}(\theta_k)$, define-se este modelo de mistura de Poisson truncadas no zero pela equação

$$\Pr[X_i = x_i | \mathbf{p}, \boldsymbol{\theta}] = \sum_{k=1}^K p_k \frac{e^{-\theta_k}}{(1 - e^{-\theta_k})} \frac{\theta_k^{x_i}}{x_i!} \quad (4.7)$$

em que $x_i \in \{1, 2, \dots\}$, $\sum_{k=1}^K p_k = 1$ e $\theta_k > 0$.

As estimativas dos parâmetros da mistura foram obtidas através da função de verossimilhança, partindo da [Equação 4.7](#), calcula-se

$$\begin{aligned} l(\boldsymbol{\theta}, \mathbf{p} | \mathbf{X}, \mathbf{S}) &= \sum_{k=1}^K n_k \ln(p_k) + \sum_{i=1}^N \sum_{k=1}^K \ln \left[\frac{e^{-\theta_k}}{(1 - e^{-\theta_k})} \frac{\theta_k^{x_i}}{x_i!} \right] \mathbb{1}_{s_i}(k) + \zeta \left(1 - \sum_{k=1}^K p_k \right) \\ &= \sum_{k=1}^K n_k \ln(p_k) - \sum_{k=1}^K n_k \theta_k - \sum_{k=1}^K n_k \ln(1 - e^{-\theta_k}) \\ &\quad - \sum_{i=1}^N \sum_{k=1}^K x_i \ln(\theta_k) \mathbb{1}_{s_i}(k) - \sum_{i=1}^N \ln(x_i!) \mathbb{1}_{s_i}(k) + \zeta \left(1 - \sum_{k=1}^K p_k \right). \end{aligned} \quad (4.8)$$

Derivando a [Equação 4.8](#), primeiramente com relação ao k -ésimo parâmetro θ_k , obteve-se

$$\frac{\partial l(\boldsymbol{\theta}, \mathbf{p} | \mathbf{X}, \mathbf{S})}{\partial \theta_k} = -n_k + \frac{1}{\theta_k} \sum_{i=1}^n x_i \mathbb{1}_{s_i}(k) - n_k \left(\frac{e^{-\theta_k}}{1 - e^{-\theta_k}} \right). \quad (4.9)$$

Deparou-se com a mesma particularidade no cálculo da estimativa do parâmetro para a distribuição Poisson truncada no zero, pois a fórmula obtida não permite isolar o parâmetro θ . Desse modo, tratou-se a estimação do k -ésimo parâmetro para o k -ésimo componente da mistura com a mesma metodologia da distribuição Poisson truncada, discutida no [Capítulo 4](#) e formulou-se

$$-n_k + \frac{1}{\hat{\theta}_k} \sum_{i=1}^n x_i \mathbb{1}_{s_i}(k) - n_k \left(\frac{e^{-\hat{\theta}_k}}{1 - e^{-\hat{\theta}_k}} \right) = 0. \quad (4.10)$$

Assim, a estimativa de θ_k pelo método recursivo é definida por

$$\hat{\theta}_k^{(j)} = \left(\frac{1 - e^{-\hat{\theta}_k^{(j-1)}}}{n_k} \right) \sum_{i=1}^n x_i \mathbb{1}_{s_i}(k) \quad (4.11)$$

para $j = 1, 2, 3, \dots$

A estimativa \hat{p}_k para o k -ésimo componente da mistura é dada por

$$\frac{n_k}{\hat{p}_k} - \sum_{k=1}^K n_k = 0$$

e

$$\hat{p}_k = \frac{n_k}{\sum_{k=1}^K n_k} = \frac{n_k}{N}. \quad (4.12)$$

Para o caso das misturas de Poisson truncadas foi necessário utilizar o algoritmo EM para a estimação dos parâmetros de interesse. Sendo que, no passo para estimação do k -ésimo parâmetro θ_k utilizou se novamente o procedimento do algoritmo EM, devido ao truncamento da distribuição Poisson.

A seguir descreve se a proposta do algoritmo EM dentro das características apresentadas para misturas de Poisson truncadas no zero.

Algoritmo 5 – Algoritmo EM para Mistura de Poisson truncadas no zero

- 1: atribua um vetor de valores iniciais para $s^{(0)}$ e para o vetor de parâmetros $\theta^{(0)}$ restrição $s_i^{(0)} \in \{1, \dots, K\}$ para $i = 1, \dots, N$
 - 2: $j = 1, 2, \dots$ ▷ j -ésima iteração
 - 3: **enquanto** $\max_k \left\{ |\hat{p}_k^{(j)} - \hat{p}_k^{(j+1)}| \right\} \geq \tau_p$
 - 4: **ou** $\max_k \left\{ |\hat{\theta}_k^{(j)} - \hat{\theta}_k^{(j+1)}| \right\} \geq \tau_\theta$ **faça** ▷ critério de parada
 - 5: calcule $\hat{s}_i^{(j)} = \operatorname{argmax}_k \{Pr[S_i = k | X_i = x_i, \hat{\theta}, \hat{\mathbf{p}}]\}$, Equação 2.15
 - 6: para o k -ésimo componente da mistura
 - 7: **enquanto** $\max_k \left\{ |\hat{\theta}_k^{(j)} - \hat{\theta}_k^{(j+1)}| \right\} \geq \tau_\theta$ **faça** ▷ critério de parada
 - 8: calcule $\hat{\theta}_k^{(j)}$, Algoritmo 3
 - 9: **fim enquanto**
 - 10: estime $\hat{p}^{(j)}$, Equação 4.12
 - 11: **fim enquanto**
-

MISTURA DE POISSON PARA MODELO TRUNCADO NO ZERO

Outra perspectiva para o problema do truncamento em modelo de mistura é declarar o valor omitido no modelo usual ao invés de utilizar a distribuição truncada como componente da mistura, de acordo com estudos de [III e Pollock \(1996\)](#), [Pledger \(2000\)](#) e [Link \(2003\)](#).

De tal maneira, foram utilizadas as definições de modelo de mistura e de distribuição truncada no zero, [Equação 2.6](#) e [Equação 3.3](#), respectivamente, e formulou se mistura de distribuições não truncadas para modelo truncado no ponto a por

$$Pr[X_i = x_i | \mathbf{q}, \lambda] = \frac{\sum_{k=1}^K q_k \mathbb{1}_{s_i(k)} Pr[X_i = x_i | S_i = s_i, \lambda_k] \mathbb{1}_{s_i(k)}}{1 - \sum_{k=1}^K q_k Pr[X_i \leq a | S_i = k, \lambda_k]} \quad (5.1)$$

em que $x_i > a$, q_k é a proporção do k -ésimo componente, com restrições $\sum_{k=1}^K q_k = 1$, $0 \leq q_k \leq 1$, $\lambda_k > 0$ e $k = 1, \dots, K$.

5.1 Mistura de Poisson com truncamento na resultante do modelo e no ponto zero

Considerando na [Equação 5.1](#) a k -ésima variável aleatória X_k com distribuição Poisson, $X_k \sim Pois(\lambda_k)$, e ponto de truncamento $a = 0$, definiu se mistura de distribuições Poisson não truncadas para modelo com truncamento na resultante do modelo e no ponto zero por

$$Pr[X_i = x_i | \mathbf{q}, \lambda] = \frac{\sum_{k=1}^K q_k \mathbb{1}_{s_i(k)} \left(\frac{e^{-\lambda_k} \lambda_k^{x_i}}{x_i!} \right) \mathbb{1}_{s_i(k)}}{1 - \sum_{k=1}^K q_k e^{-\lambda_k}} \quad (5.2)$$

em que $x_i \in \{1, 2, \dots\}$, $\sum_{k=1}^K q_k = 1$, $\lambda_k > 0$ e $k = 1, \dots, K$.

A função de verossimilhança para o modelo apresentado na [Equação 5.2](#) é dada por

$$\begin{aligned} L(\lambda, \mathbf{q} | \mathbf{X}, \mathbf{S}) &= \prod_{i=1}^N \frac{\sum_{k=1}^K q_k \mathbb{1}_{s_i}(k) \left(\frac{e^{-\lambda_k} \lambda_k^{x_i}}{x_i!} \right)^{\mathbb{1}_{s_i}(k)}}{1 - \sum_{k=1}^K q_k e^{-\lambda_k}} \\ &= \frac{\prod_{k=1}^K q_k^{n_k} \sum_{i=1}^N \left(\frac{e^{-\lambda_k} \lambda_k^{x_i}}{x_i!} \right)^{\mathbb{1}_{s_i}(k)}}{\left(1 - \sum_{k=1}^K q_k e^{-\lambda_k} \right)^N}. \end{aligned} \quad (5.3)$$

Calculou se o logaritmo da função de verossimilhança definida na [Equação 5.3](#) e escreve se

$$\begin{aligned} l(\lambda, \mathbf{q} | \mathbf{X}, \mathbf{S}) &= \ln \left[\prod_{k=1}^K q_k^{n_k} \sum_{i=1}^N \left(\frac{e^{-\lambda_k} \lambda_k^{x_i}}{x_i!} \right)^{\mathbb{1}_{s_i}(k)} \right] - N \ln \left(1 - \sum_{k=1}^K q_k e^{-\lambda_k} \right) \\ &= \sum_{k=1}^K n_k \ln(q_k) + \sum_{k=1}^K \sum_{i=1}^N \mathbb{1}_{s_i}(k) \ln \left(\frac{e^{-\lambda_k} \lambda_k^{x_i}}{x_i!} \right) + \zeta \left(1 - \sum_{k=1}^K q_k \right) \\ &\quad - N \ln \left(1 - \sum_{k=1}^K q_k e^{-\lambda_k} \right) \\ &= \sum_{k=1}^K n_k \ln(q_k) - \sum_{k=1}^K n_k \lambda_k + \sum_{i=1}^N \sum_{k=1}^K \mathbb{1}_{s_i}(k) x_i \ln(\lambda_k) - \sum_{i=1}^N \mathbb{1}_{s_i}(k) \ln(x_i!) \\ &\quad + N \left(1 - \sum_{k=1}^K q_k \right) - N \ln \left(1 - \sum_{k=1}^K q_k e^{-\lambda_k} \right). \end{aligned} \quad (5.4)$$

Para primeira derivada da [Equação 5.4](#) em relação a λ_k calcula se

$$\frac{\partial l(\lambda, \mathbf{q} | \mathbf{X}, \mathbf{S})}{\partial \lambda_k} = \frac{1}{\lambda_k} \sum_{i=1}^N x_i \mathbb{1}_{s_i}(k) - n_k + N \left(\frac{q_k e^{-\lambda_k}}{1 - q_k e^{-\lambda_k}} \right). \quad (5.5)$$

Estimou se λ_k via máxima verossimilhança encontrando o valor que maximiza a [Equação 5.4](#), para isso igualou se o resultado obtido na [Equação 5.5](#) a zero. A estimativa $\hat{\lambda}_k$, expressa na forma iterativa para implementação do algoritmo EM, é definida por

$$\hat{\lambda}_k = \frac{1}{n_k - N \left(\frac{\hat{q}_k e^{-\hat{\lambda}_k}}{1 - \hat{q}_k e^{-\hat{\lambda}_k}} \right)} \sum_{i=1}^N x_i \mathbb{1}_{s_i}(k). \quad (5.6)$$

A primeira derivada em relação a q_k é dada por

$$\frac{\partial l(\lambda, \mathbf{q} | \mathbf{X}, \mathbf{S})}{\partial q_k} = \frac{n_k}{q_k} - N + N \left(\frac{e^{-\lambda_k}}{1 - q_k e^{-\lambda_k}} \right). \quad (5.7)$$

Similarmente, para a estimativa \hat{q}_k na forma iterativa, formula se

$$\hat{q}_k = \frac{n_k}{N \left(1 - \frac{e^{-\hat{\lambda}_k}}{1 - \hat{q}_k e^{-\hat{\lambda}_k}} \right)}. \quad (5.8)$$

Visto que a segunda derivada da Equação 5.4 em relação ao k -ésimo parâmetro λ_k é menor do que zero e a segunda derivada da Equação 5.7 com relação à q_k é menor do que zero indica que $\hat{\lambda}_k$ e \hat{q}_k – definidos na Equação 5.6 e na Equação 5.8, respectivamente – são pontos de máximo.

Na implementação do algoritmo EM os estimadores via máxima verossimilhança de λ_k e q_k foram reescritos para a j -ésima iteração por

$$\hat{\lambda}_k^{(j)} = \frac{1}{n_k - N \left(\frac{\hat{q}_k^{(j-1)} e^{-\hat{\lambda}_k^{(j-1)}}}{1 - \hat{q}_k^{(j-1)} e^{-\hat{\lambda}_k^{(j-1)}}} \right)} \sum_{i=1}^N x_i \mathbb{1}_{S_i}(k) \quad (5.9)$$

e

$$\hat{q}_k^{(j)} = \frac{n_k}{N \left(1 - \frac{e^{-\hat{\lambda}_k^{(j)}}}{1 - \hat{q}_k^{(j-1)} e^{-\hat{\lambda}_k^{(j)}}} \right)}. \quad (5.10)$$

Para o cálculo de \hat{s}_i utilizou se novamente o teorema de Bayes, assim como o raciocínio desenvolvido no Capítulo 2, iniciando pelo cálculo da distribuição marginal de X e considerando o truncamento no ponto zero, formulou se

$$Pr[X_i = x_i | \lambda] = \sum_{j=1}^K q_j Pr[X_i = x_i | S_i = j, \lambda, \mathbf{q}]. \quad (5.11)$$

Com a Equação 5.3 pode se escrever a distribuição condicional de S_i dado X_i por

$$\begin{aligned} Pr[S_i = k | X_i = x_i, \lambda, \mathbf{q}] &= \frac{q_k \frac{e^{-\lambda_k} \lambda_k^{x_i}}{(1 - q_k e^{-\lambda_k}) x_i!}}{\sum_{j=1}^K q_j Pr[X_i = x_i | S_i = j, \lambda, \mathbf{q}]} \\ &= \frac{q_k \frac{e^{-\lambda_k} \lambda_k^{x_i}}{(1 - q_k e^{-\lambda_k}) x_i!}}{\sum_{j=1}^K q_j \frac{e^{-\lambda_j} \lambda_j^{x_i}}{(1 - e^{-\lambda_j}) x_i!}}. \end{aligned} \quad (5.12)$$

De tal modo, foi empregado o algoritmo EM para mistura de Poisson não truncadas para modelo com truncamento no ponto zero.

Algoritmo 6 – Algoritmo EM para Mistura de Poisson com truncamento na resultante do modelo e no ponto zero

- 1: atribua um vetor de valores iniciais para $\mathbf{s}^{(0)}$ e para os vetores de parâmetros $\lambda^{(0)}$ e $\mathbf{q}^{(0)}$ com restrições $s_i^{(0)} \in \{1, \dots, K\}$ para $i = 1, \dots, N$, $\lambda_k > 0$ e $\sum_{k=1}^K q_k = 1$
 - 2: para $j = 1, 2, \dots$ ▷ j -ésima iteração
 - 3: **enquanto** $\max_k \left\{ |\hat{q}_k^{(j)} - \hat{q}_k^{(j+1)}| \right\} \geq \tau_q$
 - 4: **ou** $\max_k \left\{ |\hat{\lambda}_k^{(j)} - \hat{\lambda}_k^{(j+1)}| \right\} \geq \tau_\lambda$ **faça** ▷ critério de parada
 - 5: calcule $\hat{s}_i^{(j)} = \operatorname{argmax}_k \left\{ Pr[S_i = k | X_i = x_i, \hat{\lambda}, \hat{\mathbf{q}}] \right\}$, Equação 5.12
 - 6: estime $\lambda_k^{(j)}$, Equação 5.9, e $q_k^{(j)}$, Equação 5.10
 - 7: **fim enquanto**
-

ESTUDO DE SIMULAÇÃO

Todo o estudo estatístico computacional foi desenvolvido no *software* estatístico **R** com o auxílio do *IDE – integrated development environment* – **RStudio** e do **Tinn-R**. Os comandos desenvolvidos constam no [Apêndice C](#).

6.1 Simulação dos dados

Para representar um modelo de mistura de Poisson truncadas no zero foram simuladas diversas amostras, sendo cada uma gerada com grupos de tamanhos iguais – sub amostras – para a distribuição Poisson truncada no zero com distintos parâmetros.

Inicialmente, foram replicadas amostras com distribuição Uniforme entre 0 e 1, com o mesmo tamanho N da amostra requerida e a utilizamos para estabelecer se o i -ésimo elemento pertence ao k -ésimo componente. Para isso, comparou se o i -ésimo valor obtido da Uniforme com probabilidades acumuladas, determinada inicialmente por $\frac{1}{K}$. Assim como exemplificado no [Capítulo 4](#).

Com K estipulado, cada elemento foi gerado como uma sub amostra seguindo distribuição Poisson de parâmetro θ_k . O truncamento foi efetuado descartando todos os zeros obtidos e esse procedimento foi repetido até que a amostra esteja completa, ou seja, de tamanho n_k , sem valores *missings* e sem valor zero.

Foram gerados variados tamanhos de amostra N assumindo os valores 10, 30, 50, 100, 1.000 e 10.000. Em cada amostra gerada foram fixadas as quantidades de sub amostras em 2, 3, 4 e 5, representando os K componentes na mistura.

A [Tabela 2](#) apresenta os verdadeiros valores dos parâmetros da Poisson na geração dos dados, sendo que os parâmetros foram construídos dentro de 3 diferentes cenários.

No cenário I foram estabelecidas sub amostras com parâmetros próximos entre si, isto

Tabela 2 – Valores fixados na simulação dos dados para os parâmetros θ_k da Poisson truncada no zero e p_k para a mistura.

K	p_k	$\theta = (\theta_1, \dots, \theta_K)$		
		I	II	III
2	0,5000	(2, 3)	(3, 9)	(4, 20)
3	0,3333	(2, 3, 4)	(3, 9, 15)	(4, 20, 36)
4	0,2500	(2, 3, 4, 5)	(3, 9, 15, 21)	(4, 20, 36, 52)
5	0,2000	(2, 3, 4, 5, 6)	(3, 9, 15, 21, 27)	(4, 20, 36, 52, 68)

Fonte: Elaborada pelo autor.

é, fixado o k -ésimo parâmetro em $\theta_k = k + 1$ para $k = 1, \dots, K$. Para o Cenário II, assumiu se $\theta_1 = 3$ e $\theta_k = \theta_{k-1} + 6$ para $k = 2, \dots, K$. E, distanciando mais os parâmetros das Poisson entre as sub amostras, no Cenário III foram atribuídos $\theta_1 = 4$ e $\theta_k = \theta_{k-1} + 16$ para $k = 2, \dots, K$.

Foram geradas 100 amostras da distribuição Poisson truncada no zero dentro de cada especificidade apresentada na Tabela 2 para os parâmetros θ_k e p_k fixo em $\frac{1}{K}$, com diversificados tamanhos de amostras N e diferentes quantidades de sub amostras K , de tamanho aproximado $\frac{N}{K}$. Totalizando 90 condições amostrais distintas a serem analisadas na perspectiva de mistura de Poisson truncadas no zero.

6.2 Implementação do Algoritmo EM

Todo código para o algoritmo EM, apresentado no Apêndice C, foi desenvolvido primeiramente nas especificações da metodologia para estimação dos parâmetros do modelo de mistura de Poisson truncadas no zero, discutida na Seção 4.1, e o algoritmo foi chamado de EMMPTZ.

Partindo dos resultados obtidos, implementou se para o caso de misturas de Poisson para modelo com truncamento no zero, discutido na Seção 5.1, nomeado por algoritmo EMPMTZ.

Nesta Seção é apresentado alguns detalhes da construção lógica nos passos apresentados e os principais obstáculos obtidos na implementação e execução dos algoritmos.

A quantidade total de componentes na mistura K foi atribuída como valor fixo, conhecido e de entrada. E a quantidade de elementos do k -ésimo componente foi restrita como um valor inteiro e diferente de zero.

O primeiro passo para o algoritmo EM, é estipular valores iniciais para os parâmetros em que tem se interesse em estimar. Para a estimativa de S , que discriminou em qual componente da mistura o elemento pertenceu, foram gerados valores aleatórios entre 1 e K , determinando $\hat{s}_i^{(0)}$.

Para as estimativas dos parâmetros $\hat{\theta}_k^{(0)}$, Equação 4.11, foram calculadas algumas medi-

das descritivas da variável de interesse. A [Tabela 3](#) apresenta em detalhes as medidas descritivas e em qual componente ela foi aplicada como valor inicial $\hat{\theta}_k^{(0)}$, relacionado a quantidade total de componentes. Em que, Q_2 representa a mediana da variável de interesse e $Q_{(\cdot)}$ são os quantis. Para melhor comparação entre os dois algoritmos desenvolvidos os valores iniciais de $\hat{\lambda}_k^{(0)}$, [Equação 5.9](#), foram fixados iguais a $\hat{\theta}_k^{(0)}$.

Tabela 3 – Valores iniciais na implementação do algoritmo EM para as estimativas dos parâmetros θ_k .

K	$\hat{\theta}_1^{(0)}$	$\hat{\theta}_2^{(0)}$	$\hat{\theta}_3^{(0)}$	$\hat{\theta}_4^{(0)}$	$\hat{\theta}_5^{(0)}$
2	Min	Max	-	-	-
3	Min	Q_2	Max	-	-
4	Min	$Q_{0,4}$	$Q_{0,6}$	Max	-
5	$Q_{0,15}$	$Q_{0,3}$	$Q_{0,45}$	$Q_{0,6}$	$Q_{0,75}$

Fonte: Elaborada pelo autor.

Ao definir valores aleatórios também para os valores iniciais $\hat{\theta}_k^{(0)}$ o algoritmo indicou maior tempo de processamento e maior quantidade de iterações até a convergência dos parâmetros. Aparentemente, as estimativas se desenvolveram por muito tempo dentro da aleatoriedade dos valores iniciais antes de alcançar a convergência. A mudança na inicialização de $\hat{\theta}_k^{(0)}$ pelas medidas descritivas da variável em estudo apresentou diferenças significativas para esse obstáculo.

Com tal característica, o valor inicial de $\hat{\theta}_k^{(0)}$ foi a primeira informação para o algoritmo sobre os dados em análise. A partir disso, no Passo 2, foi estimado se a variável não observável \hat{S} . Estimar $\hat{S}^{(1)}$ e posteriormente estimar $\hat{\theta}_k^{(1)}$ e $\hat{p}_k^{(1)}$ para prosseguir com os próximos passos indicou direcionar o algoritmo com mais eficácia para a convergência dos parâmetros. Similarmente, foram implementados os cálculos das estimativas $\hat{S}^{(j)}$ anterior a operação de $\hat{\lambda}_k^{(j)}$, [Equação 5.9](#), e $\hat{q}_k^{(j)}$, [Equação 5.10](#).

A implementação do algoritmo EMMPTZ indicou menor complexidade, comparada ao algoritmo EMPMTZ. Pois, estimar cada $\hat{\theta}_k$ requer a execução de um algoritmo EM adaptado uma vez que o truncamento foi efetuado no componente da mistura e, além disso, o truncamento foi tratado como valor faltante na distribuição Poisson. Assim, para calcular $\hat{\theta}_k$ formulou-se uma função no R, [Código-fonte 3](#), que descreve o algoritmo para a estimação do parâmetro θ para a distribuição Poisson truncada no zero, detalhado na [Seção 3.2](#).

Ao passo que, na metodologia discutida na [Seção 5.1](#) considerou-se o truncamento no modelo de mistura e apenas realizou-se o cálculo de $\hat{\lambda}_k$ de forma recursiva, assim como o cálculo de \hat{q}_k .

O critério de parada foi estipulado em 10^{-4} para o máximo da diferença das estimativas dos parâmetros $\hat{\theta}_k$ e \hat{p}_k entre a j -ésima e a $(j+1)$ -ésima iterações. E, equivalentemente para $\hat{\lambda}_k$ e \hat{q}_k .

Em alguns casos os algoritmos elaborados apresentaram problemas na convergência dos parâmetros, interrompidos manualmente entre 2 e 5 dias, aproximadamente. Observamos que apenas algumas das 100 amostras estavam apresentando problemas na convergência das estimativas dos parâmetros de interesse.

Tal fato, apresentou-se mais evidente na convergência de $\hat{\lambda}_k$ e \hat{q}_k . Sendo assim, a quantidade de iterações foi limitada para cada amostra em 200.000 iterações.

Desse modo, pode se pensar que há indícios de que a metodologia apresentada na [Seção 4.1](#) parece solucionar melhor o problema do truncamento no zero no modelo de mistura. Possivelmente, por tratar o truncamento como valor faltante.

6.3 Discussão dos Resultados

Por intermédio dos dados simulados, descritos na [Seção 6.1](#), analisou se a performance dos dois algoritmos desenvolvidos para estimação dos parâmetros da mistura de Poisson para dados truncados no zero. Nesta Seção foram discutidos os resultados obtidos para o algoritmo EMMPTZ e para o algoritmo EMPMTZ, nomeados na [Seção 6.2](#). As tabelas com os resultados constam no [Apêndice A](#) e no [Apêndice B](#).

6.3.1 Algoritmo EMMPTZ

A [Tabela 13](#), a [Tabela 14](#) e a [Tabela 15](#) apresentam medidas descritivas da quantidade total de iterações para que o algoritmo EM obtenha convergência na estimação dos parâmetros θ_k e p_k de cada amostra. Informando também o tempo total de processamento, em segundos, para cada grupo de 100 amostras. E cada grupo de amostras é especificado dentre os diferentes tamanhos N , quantidade de componentes K e verdadeiro valor do parâmetro θ_k da variável simulada.

No geral, o tamanho da amostra N e a quantidade de componentes K tem forte influência no tempo de processamento, variando entre milésimos de segundos e 17 horas 10 minutos e 38,77 segundos; tempo máximo obtido na estimação dos parâmetros para dados simulados dentro do Cenário II, com $K = 5$ e $N = 10.000$. Em contrapartida, a quantidade máxima de iterações obtidas foi de 9.500 iterações em uma das 100 amostras.

Tal fato, pode ser explicado pelo processo na estimativa de S , pois é necessário obter a probabilidade de cada elemento i pertencer a cada grupo k . Em que, a [Equação 2.15](#) é computada $N \times K$ vezes a cada iteração. O máximo obtido para a quantidade de iterações foi de 170.600 iterações com tempo total de processamento para o grupo de 100 amostras em 3 horas 6 minutos e 49,95 segundos e tamanho de amostra $N = 100$, $K = 5$ e Cenário I para θ_k .

Observou se também para os casos das simulações amostrais realizadas dentro do Cenário I, com parâmetros próximos, maior variabilidade na quantidade de iterações. Nas amostras

simuladas para θ_k no Cenário III, o algoritmo resultou em menor variabilidade na quantidade de iterações, obtendo para $N = 1.000$ e $K = 5$ máximo de 5.545 iterações em uma das 100 amostras.

Considerando $K = 2$ componentes na mistura e para θ_k no Cenário II e no Cenário III a quantidade de iterações foi baixa. Sendo que atingiu a convergência entre 1 e 4 iterações, mesmo para as pequenas amostras, com $N = 10$.

Um fato provável do algoritmo EM alcançar a convergência em apenas uma iteração pode ser explicado pelos valores iniciais $\hat{\theta}_k^{(0)}$, pois são medidas de entrada estabelecidas pelos dados.

Para θ_k no Cenário I e dentro das condições em que $K = 4$ e $N = 10.000$ e para $K = 5$, $N = 1.000$ e $N = 10.000$ não ocorreu a convergência na estimação dos parâmetros. O algoritmo permaneceu por volta de três dias para cada uma das 3 condições apresentadas, sendo que o algoritmo permaneceu na primeira amostra para cada grupo, não finalizando o cálculo das estimativas de interesse. Tais situações estão representadas pelo símbolo (o), [Tabela 13](#), [Tabela 14](#) e [Tabela 15](#), e podem ser explicadas devido ao fato da proximidade dos verdadeiros valores dos parâmetros θ_k e pelo valor de N ser muito grande.

Com relação a p_k , devido aos valores fixos definidos na [Tabela 2](#), considera-se que o algoritmo EMMPTZ aponte uma boa performance caso $\hat{p}_k \approx \frac{1}{K}$. Na [Tabela 16](#) são apresentadas medidas descritivas dos resultados das estimativas de \hat{p}_k e \hat{n}_k para 2 componentes na mistura e dentro de cada Cenário de θ_k . A [Tabela 18](#) e a [Tabela 20](#) exibem as estimativas para 3 componentes, a [Tabela 22](#) e a [Tabela 24](#) para 4 componentes e a [Tabela 26](#) e a [Tabela 28](#) para $K = 5$.

Em linhas gerais, o algoritmo obteve bons resultados para \hat{p}_k . Sendo que, para θ_k dentro do Cenário I e $K = 2$, [Tabela 16](#), os componentes indicaram maior distância de p_k .

Para θ_k no Cenário II e $K = 3$ a diferença entre as estimativas \hat{p}_k apresentou-se mais evidente para o segundo componente da mistura. Sendo que a mediana da quantidade de elementos para o componente 2, com $\theta_2 = 9$, resultou em maior número de elementos para todos os valores de N em estudo.

Sendo mais perceptível para amostra de tamanho $N = 10.000$ em que o componente 2 apresentou mediana de 6.039 observações para \hat{n}_2 e o componente 3, com $\theta_3 = 15$, revelou poucos elementos com mediana de 843 observações para \hat{n}_3 .

Para a mistura com 4 componentes, [Tabela 22](#) e [Tabela 24](#), os resultados de \hat{p}_k para as amostras N entre 10 e 100 foram dentro do esperado. No entanto, com a amostra de tamanho 1.000 os componentes de valores extremos para θ_k obtiveram, em média, maiores quantidades de elementos, ou seja, $\hat{n}_1 \approx \hat{n}_4$ e ainda $\hat{n}_2 \approx \hat{n}_3$, em média.

E com $N = 10.000$ o componente 3 de parâmetro $\theta_3 = 15$ obteve, em média, as maiores quantidades de elementos. Sendo 3.598 observações a mediana para \hat{n}_3 .

Considerando 5 componentes na mistura, para N entre 10 e 100, o algoritmo EMMPTZ apresentou bom desempenho analisando as estimativas \hat{p}_k e \hat{n}_k . Ao passo que, para as amostras de tamanho 1.000 e 10.000 a proximidade dessas estimativas com o verdadeiro valor do parâmetro não foi tão grande.

Em que, na média, as maiores quantidades de observações foram estipuladas para o componente 5 no Cenário II. E para o componente 3, com $\theta_3 = 15$, resultou em menores quantidades de observações. Assim como no Cenário III, para $N = 1.000$ o componente 3 com $\theta_3 = 36$, igual a média de θ_k , obteve menores quantidades de observações.

A Tabela 17 exibe medidas descritivas das estimativas $\hat{\theta}_k$ para a mistura com dois componentes. A Tabela 19 e a Tabela 21 apresentam os resultados para três componentes na mistura. A Tabela 23 e a Tabela 25 exibem medidas descritivas de $\hat{\theta}_k$ para a mistura com quatro componentes; e a Tabela 27 e a Tabela 29 para 5 componentes.

No geral, quanto maior o tamanho da amostra a estimativa $\hat{\theta}_k$ se aproxima de θ_k . Além disso, quando θ_k assume valor baixo, isto é, valores próximos de 1, $\hat{\theta}_k$ apresenta maior viés. Fato que pode ser justificado pelo efeito do truncamento no zero da distribuição Poisson.

As melhores performances para as estimativas $\hat{\theta}_k$ foram obtidas dentro do Cenário III e para $K = 2$ e 3, em que $\theta_k \approx \hat{\theta}_k$, mesmo para as amostras pequenas $N = 10$.

No Cenário II e para $K = 2$ as estimativas $\hat{\theta}_1$ e $\hat{\theta}_2$ resultaram em valores próximos do verdadeiro valor do parâmetro. Sendo que, o máximo da diferença entre θ_1 e $\hat{\theta}_1$ é dada por aproximadamente 0,5 em média e módulo. E, o máximo da diferença entre θ_2 e $\hat{\theta}_2$ é dada por aproximadamente 1 unidade, em média e módulo, para todos os valores de N em estudo.

Com 4 componentes na mistura, Tabela 23, observamos na amostra de tamanho 10 que as médias das estimativas $\hat{\theta}_k$ estão bem distantes do valor de θ_k , principalmente para os Cenários II e III. Sendo que, os componentes 1 e 4 apresentaram as estimativas mais discrepantes de θ_k .

Para as amostras de tamanhos 30, 50 e 100 dentro do Cenário II, Tabela 23 e Tabela 25, as estimativas $\hat{\theta}_1$ indicam maiores valores em relação a θ_k . Consequentemente, as estimativas $\hat{\theta}_k$ de alguns dos outros componentes resultaram em valores menores que o verdadeiro valor do parâmetro. Tais efeitos também foram apresentados para os modelos ajustados com cinco componentes na mistura, Tabela 27 e a Tabela 29. Em que, $\hat{\theta}_1$ indicou maior viés para a amostra de tamanho 10.000 dentro do Cenário II.

A variância para as estimativas $\hat{\theta}_k$, não obteve valores muito elevados. No entanto, para os casos em que o tamanho da amostra é pequeno, $N = 10$, tanto a média quanto a mediana das variâncias de $\hat{\theta}_k$, para cada grupo de 100 amostras, apresentaram valores um pouco mais elevados.

Um resultado interessante foi que em algumas amostras, dentro de 4 diferentes condições, o algoritmo estimou dois parâmetros θ_k de mesmo valor. Ou seja, para uma mesma amostra a

estimativa de θ_k resultou exatamente no mesmo valor de θ_l , para $k \neq l$. Tal fato indica que para essas amostras o algoritmo aponta o modelo com $K - 1$ componentes na mistura.

A [Tabela 4](#) apresenta em detalhes quais as condições amostrais resultaram em $K - 1$ componentes na mistura e a quantidade de amostras, M_\star , dentre cada grupo de $M = 100$ amostras analisadas.

Tabela 4 – Amostras em que o algoritmo EM resultou em $K - 1$ componentes na mistura.

K	N	M_\star	Cenário
4	1.000	7	I
		40	I
5	10	9	II
		16	I
		12	I
		6	I

Fonte: Elaborada pelo autor.

Foram discriminados na [Tabela 14](#) e na [Tabela 15](#) a quantidade de iterações que tais amostras necessitaram para alcançar a convergência e indicadas pelo símbolo (\star). As amostras M_\star revelam necessidade de mais iterações para a convergência quando comparadas com as amostras do mesmo grupo. E, para esses casos, foram retirados tais resultados da análise das estimativas de θ_k e p_k .

O caso em que o algoritmo apontou $M_\star = 40$ é significativo, pois representa 40% das amostras modeladas. O parâmetro verdadeiro na geração dos dados foi de $\theta_1 = 2$, $\theta_2 = 3$, $\theta_3 = 4$, $\theta_4 = 5$ e $\theta_5 = 6$ para a Poisson truncada no zero e para a mistura $p_k = 0,2$ em que $k = 1, \dots, K$ e $K = 5$. O algoritmo EMMPTZ foi executado para as amostras M_\star e $K = 4$. E, novamente, o algoritmo indicou $K - 1$ componentes na mistura para 15% das amostras M_\star , isto é, 6 das 40 amostras resultaram em uma mistura com 3 componentes.

Diminuir a quantidade de componentes na mistura parece ser um bom indício, pois neste caso o estudo foi realizado com amostra de tamanho $N = 10$ e dividí-la em muitas distribuições, com 4 ou 5 componentes, pode ser inviável e desnecessário. Além da amostra ser pequena, tem se a peculiaridade da proximidade dos parâmetros, Cenário I.

Algumas medidas descritivas das estimativas \hat{p}_k e $\hat{\theta}_k$ para as amostras que indicaram redução no número de componentes M_\star são apresentados na [Tabela 5](#). As estimativas dos parâmetros que definem a quantidade de elementos em cada componente, indicam que dois componentes, para $k = 2$ e $k = 3$, obtiveram maiores valores para \hat{p}_k e \hat{n}_k .

Mesmo com a diminuição na quantidade de componentes, em média, os valores estimados

Tabela 5 – Resultados do algoritmo EM para $\hat{\theta}_k$, \hat{p}_k e \hat{n}_k para a amostra M^* , $N = 10$ e $\theta_k = (2, 3, 4, 5, 6)$.

k	$\hat{\theta}_k$						Mediana		
	Min	Q_1	Q_2	Média	Q_3	Max	$Var[\hat{\theta}_k]$	\hat{p}_k	\hat{n}_k
1	0,0303	0,8500	1,3330	1,5400	2,5000	4,6670	0,9392	0,2000	2
2	1,0000	2,5000	3,0000	3,0100	3,5000	5,2000	1,2220	0,3000	3
3	2,5000	3,5000	4,0000	4,2730	4,6670	9,0000	1,4460	0,3000	3
4	3,5000	4,3750	5,0000	6,4020	8,5000	13,0000	3,0380	0,2000	2

Fonte: Elaborada pelo autor.

$\hat{\theta}_k$ foram próximos ao verdadeiro valor do parâmetro, desconsiderando o componente em que $\theta_k = 5$. A variância de $\hat{\theta}_k$ obteve maior valor para o componente 4, em que foi estimado o maior parâmetro $\hat{\theta}_k$ dentre as 38 amostras.

6.3.2 Algoritmo EMPMTZ

A Tabela 30, a Tabela 31 e a Tabela 32 apresentam medidas descritivas da quantidade de iterações necessárias para que o algoritmo EM alcance a convergência nas estimativas dos parâmetros para a mistura de Poisson com modelo truncado no zero. Assim como, o tempo de processamento para as 100 amostras geradas, em segundos, e o percentual de amostras que obtiveram a convergência em até 200 mil iterações. Tais informações estão detalhadas para variados tamanhos de amostra N , diferentes Cenários para o parâmetro de interesse λ_k e distintas quantidades de componentes na mistura K .

De forma global, o algoritmo EMPMTZ apresentou necessidade de mais iterações para convergência na estimativa dos parâmetros e, conseqüentemente, maior tempo de processamento comparado ao algoritmo EMMPTZ.

Tal situação foi enfatizada para os casos dos parâmetros λ_k próximos, ou seja, dentro do Cenário I. Em que, na maioria das situações em estudo obteve a convergência em menos da metade das 100 amostras geradas, variando entre 2 e 44 amostras; exceto pelos casos em que nenhuma das amostras obteve convergência dos parâmetros de interesse, sinalizados por (o).

Em decorrência, o tempo de processamento foi de 1 hora 13 minutos e 30,57 segundos para amostras pequenas, $N = 10$, com apenas 2 componentes na mistura e parâmetros λ_k no Cenário I.

O algoritmo também apresentou dificuldades na convergência dos parâmetros, entre 1 e 14 amostras, para os casos dos parâmetros λ_k fixados dentro do Cenário II, para $N = 10$ com $K = 4$ componentes na mistura e para $N = 10$ e $N = 30$ com $K = 5$ componentes. O pode ser explicado pelo fato das amostras serem pequenas e a quantidade de componentes da mistura alta.

O algoritmo EMPMTZ apresentou bom desempenho para todos os casos estudados

com parâmetro λ_k com valores distantes, Cenário III. Atingindo o tempo máximo para $K = 5$ componentes e o maior tamanho de amostra em estudo $N = 10.000$ em pouco menos de 1 hora.

A [Tabela 33](#) apresenta medidas descritivas das estimativas dos parâmetros $\hat{\lambda}_k$ e \hat{q}_k que atingiram a convergência através do algoritmo EMPMTZ para 2 componentes na mistura. Assim como o verdadeiro valor do parâmetro λ_k e a quantidade de observações em cada componente da mistura n_k .

Em média, as estimativas $\hat{\lambda}_k$ resultaram em valores próximos do verdadeiro valor λ_k , apresentando maior viés em $\hat{\lambda}_2$ para as amostras pequenas $N = 10$ e $N = 30$ com λ_k no Cenário I.

As variâncias das estimativas dos parâmetros $\hat{\lambda}_k$ apresentaram valores baixos, exceto para alguns casos em que λ_k foi gerado dentro do Cenário I, obtendo o máximo em $Var[\hat{\lambda}_2] = 30,6933$.

A [Tabela 34](#) exhibe algumas medidas descritivas das estimativas dos parâmetros $\hat{\lambda}_k$ e \hat{q}_k para $K = 3$ componentes na mistura. De forma geral, as estimativas apresentaram-se próximas do verdadeiro valor do parâmetro.

Dentro do Cenário II ocorreu um fato interessante em que o componente da mistura com $\lambda_k = 9$ obteve maior número de observações comparado aos outros dois componentes. Sendo que, com amostras de tamanho grande $N = 1.000$ e $N = 10.000$ tal diferença fica mais evidente resultando, em média, em mais da metade das observações da amostra.

O terceiro componente resultou em valores pequenos para a estimativa \hat{q}_3 com média 0,1280 e mediana 0,0130. Tal situação também influenciou a estimativa dos parâmetros $\hat{\lambda}_2$ e $\hat{\lambda}_3$ em que apresentaram maior distância do verdadeiro valor do parâmetro.

Além do fato do valor da variância ser inversamente proporcional ao tamanho da amostra, as variâncias das estimativas dos parâmetros $\hat{\lambda}_k$ apresentaram valores baixos. No entanto, para os casos em que o verdadeiro valor do parâmetro é alto, para $\lambda_k \geq 20$, as variâncias apresentaram valores maiores.

Já para $K = 4$ componentes na mistura, resultados na [Tabela 35](#), nenhuma das amostras geradas com parâmetros λ_k próximos, Cenário I, alcançou a convergência na estimativa dos parâmetros, simbolizado por (\circ).

E para parâmetros λ_k dentro dos Cenários II e III, algumas das estimativas $\hat{\lambda}_k$ resultaram em valores mais distantes do verdadeiro valor do parâmetro, por exemplo para $\lambda_2 = 9$ e $N = 10$ com $\hat{\lambda}_2 = 7,8550$.

Para os casos de amostras de tamanho grande, N a partir de 30 observações, o valor das variâncias de $\hat{\lambda}_k$ apresentaram valores baixos para λ_k pequeno e valores altos para λ_k grande. Com mínimo de 0,0028 para $\lambda_1 = 4$ e $N = 10.000$ e máximo de 68,7170 para $\lambda_4 = 52$ e $N = 10$.

Considerando as estimativas de \hat{q}_k , os resultados apresentaram-se dentro do esperado com $\hat{q}_k \approx 0,2500$.

Para finalizar com 5 componentes na mistura, resultados na [Tabela 36](#) para $N = 10, 30$ e 50 e na [Tabela 37](#) para $N = 100, 1.000$ e 10.000 , a convergência nas estimativas dos parâmetros do modelo foram obtidas para os casos onde λ_k estão dentro do Cenário II e do Cenário III e, apenas para alguns casos do Cenário I.

No Cenário III, para $\lambda_4 = 52$ as estimativas $\hat{\lambda}_4$ apresentaram por volta de 45 para os diferentes tamanhos de amostra, indicando a maior distância do verdadeiro valor do parâmetro dentre todos os componentes do modelo.

Em relação a \hat{q}_k , o qual foi esperado resultado próximo a 0,2000, foram obtidos maiores valores para os componentes com λ_k de valores extremos, por exemplo para $N = 10$, $\lambda_5 = 68$ resultou na estimativa $\hat{q}_5 = 0,3000$. Consequentemente, as estimativas de \hat{q}_2 , \hat{q}_3 e \hat{q}_4 apresentaram se abaixo de 0,2000.

APLICAÇÃO

7.1 Descrição dos dados

A evolução dos meios de comunicação trouxe algumas facilidades como o acesso à informação e à comercialização virtual. Neste contexto, é crescente o uso da internet para comercializações em geral – a informatização das vendas – definidas como comércio eletrônico e mais comumente chamado pelos especialistas no termo em inglês *e-commerce*. Em vista disso, o conjunto de dados que motivou este estudo é originário de um *site* de venda de diversos produtos por intermédio da internet.

Em linhas gerais, ao realizar uma compra em uma determinada página da internet o usuário adquire um número de identificação e então são armazenadas as informações de navegação no site e de ocorrência de venda.

Neste ambiente, é importante identificar padrões de compra de clientes para abordagem de publicidades em geral, como por exemplo ações para fidelização de compradores. O objetivo desta análise é determinar grupos homogêneos de clientes de acordo com a quantidade de compra realizadas em um determinado período.

Temos disponíveis as informações de vendas de 685 dias, trata-se do período de maio de 2.012 a abril de 2.014, em que 20.302 clientes realizaram 60.158 compras.

Apresentamos na [Tabela 6](#) as informações da quantidade de compras e o número de clientes, indicando que os clientes com apenas uma compra são os mais frequentes, somando 44% do total de clientes.

A [Figura 2](#) mostra o *Box-Plot* e o Histograma da variável em estudo. Os valores de compra analisados estão entre 1 e 26, com maior concentração entre 1 e 3, mediana de 2 compras e média de 3 compras. O *Box-Plot*, [Figura 2](#), indica alguns pontos discrepantes entre 6 e 26. A variância da variável quantidade de compras resultou em 11,2007.

Tabela 6 – Valores da quantidade de compras por cliente, proporção do total de clientes e a proporção acumulada.

compras	clientes	proporção	acumulada
1	8.958	0,4412	0,4412
2	4.152	0,2045	0,6457
3	2.315	0,1140	0,7598
4	1.410	0,0695	0,8292
5	885	0,0436	0,8728
6	638	0,0314	0,9042
7	401	0,0198	0,9240
8	282	0,0139	0,9379
9	237	0,0117	0,9496
10	172	0,0085	0,9580
11	138	0,0068	0,9648
12	131	0,0065	0,9713
13	95	0,0047	0,9760
14	92	0,0045	0,9805
15	64	0,0032	0,9836
16	60	0,0030	0,9866
17	56	0,0028	0,9894
18	30	0,0015	0,9908
19	33	0,0016	0,9925
20	26	0,0013	0,9937
21	31	0,0015	0,9953
22	14	0,0007	0,9960
23	20	0,0010	0,9969
24	20	0,0010	0,9979
25	21	0,0010	0,9990
26	21	0,0010	1,0000
total	20.302	1,0000	

Fonte: Dados da pesquisa.

7.2 Algoritmo EMMPTZ

Aplicou-se o algoritmo EMMPTZ, discutido na [Seção 6.2](#), à variável quantidade de compras para o número de componentes na mistura K variando entre 2 e 5 componentes.

A [Tabela 7](#) apresenta as estimativas dos parâmetros para o modelo de mistura de Poissons truncadas no zero, assim como a quantidade de componentes na mistura e medidas descritivas dentro de cada componente. O algoritmo EMMPTZ alcançou convergência para todas as K quantidades de componentes estudadas.

Para o modelo de mistura ajustado com 2 componentes, o componente 1 somou cerca de 90% dos clientes da base de dados, com valores entre 1 e 6 compras e estimativa $\hat{\theta}_1$ próximo a 2 compras. O componente 2 apresentou os clientes com maiores valores de compras, entre 7 e 26, com $\hat{\theta}_2$ próximo a 11 compras.

Com 3 componentes na mistura, o componente 1 somou cerca de 65% dos clientes, com valores de 1 e 2 compras e $\hat{\theta}_1$ próximo a 1 compra. O componente 2 totalizou aproximadamente 30% dos clientes, em que realizaram entre 3 e 10 compras e $\hat{\theta}_2$ próximo a 5 compras. E o componente 3 com menos de 5% do total de clientes, quantidade de compras entre 11 e 26 e $\hat{\theta}_3$ próximo a 15 compras.

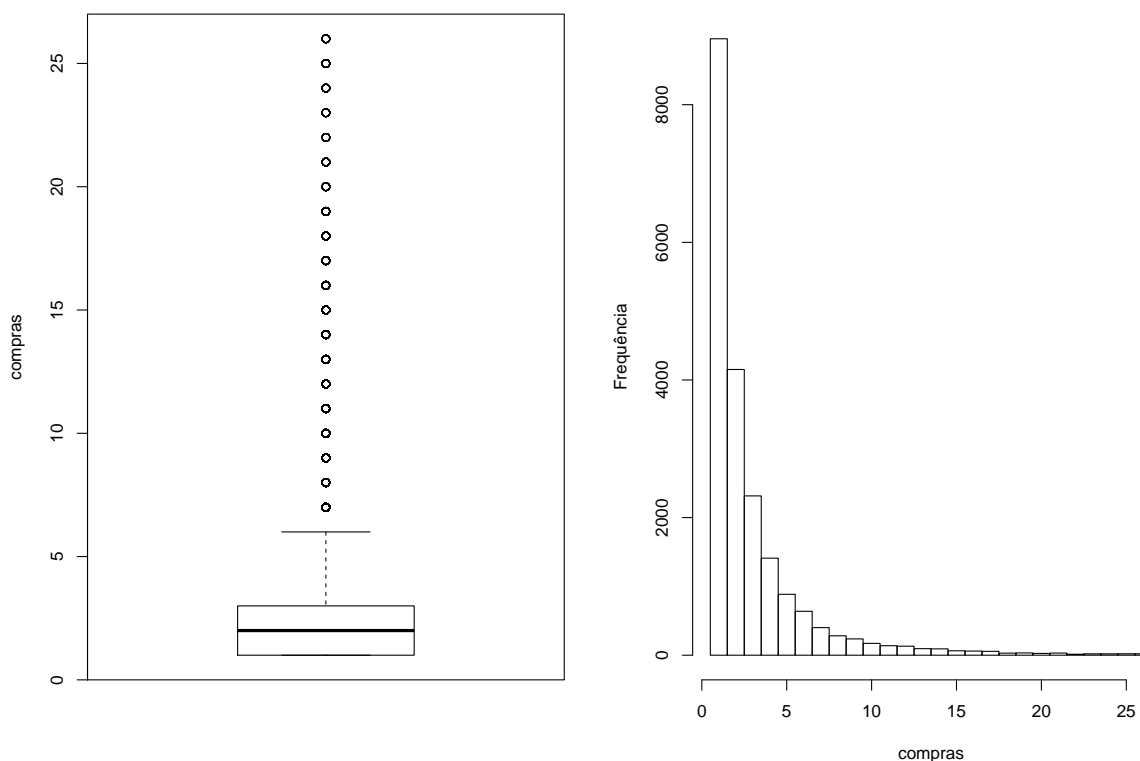


Figura 2 – Gráficos *Box-Plot* e Histograma da variável quantidade de compras por cliente.

Fonte: Dados da pesquisa.

O modelo ajustado com 4 componentes classificou todos os clientes com 1 compra para o componente 1, com 44% do total de clientes, e os clientes com 2 compras para o componente 2, cerca de 20% do total. Para o componente 3 resultou em clientes que fizeram entre 3 e 10 compras, $\hat{\theta}_3$ próximo a 5 compras e 30% do total de clientes. No componente 4, tem-se os clientes com 11 a 26 compras, $\hat{\theta}_4$ próximo a 15 e menos de 5%.

Note que para o modelo ajustado para 3 componentes na mistura e para o modelo com 4 componentes, os modelos selecionaram exatamente os mesmos clientes para os componentes com maiores valores de quantidade de compras. Ao passo que, os clientes do componente 1 para o modelo com $K = 3$ foram divididos em 2 componentes no modelo com $K = 4$.

Para 5 componentes na mistura o modelo agrupou novamente os mesmos clientes dos modelos ajustados para $K = 3$ e $K = 4$ nos componentes 1 e 2. O componente 3 resultou em cerca de 30% do total de indivíduos, com compras entre 3 e 9 e $\hat{\theta}_3$ próximo a 5. O componente 4 com menos de 5% do total dos clientes, compras entre 10 e 22 e $\hat{\theta}_4$ próximo a 14. E, para o componente 5, com apenas 0,4% do total, compras entre 23 e 26 e $\hat{\theta}_5$ próximo a 25 compras por cliente.

A [Figura 3](#) mostra o histograma da variável em estudo com sinalização do modelo

Tabela 7 – Resultados do algoritmo EMMPTZ aplicado à quantidade de compras por cliente para K componentes na mistura e medidas descritivas dentro de cada componente.

K	estimativas					quantidade de compras						
	\hat{s}	\hat{p}_k	\hat{n}_k	$\hat{\theta}_k$	$Var[\hat{\theta}_k]$	Min	Q_1	Q_2	Média	Q_3	Max	Var
2	1	0,9042	18.358	1,6943	0,0001	1	1	2	2	3	6	1,9056
	2	0,0958	1.944	11,3467	0,0058	7	8	10	11	14	26	21,2569
3	1	0,6457	13.110	0,5804	0,0001	1	1	1	1	2	2	0,2164
	2	0,3123	6.340	4,6483	0,0008	3	3	4	5	6	10	3,6196
	3	0,0420	852	15,4249	0,0181	11	12	14	15	18	26	17,3680
4	1	0,4412	8.958	0,0434	0	1	1	1	1	1	1	0
	2	0,2045	4.152	1,5942	0,0005	2	2	2	2	2	2	0
	3	0,3123	6.340	4,6483	0,0008	3	3	4	5	6	10	3,6196
	4	0,0420	852	15,4249	0,0181	11	12	14	15	18	26	17,3680
5	1	0,4412	8.958	0,0434	0	1	1	1	1	1	1	0
	2	0,2045	4.152	1,5942	0,0005	2	2	2	2	2	2	0
	3	0,3038	6.168	4,4941	0,0008	3	3	4	5	6	9	2,9131
	4	0,0464	942	13,6423	0,0145	10	11	13	14	16	22	10,5807
	5	0,0040	82	24,5244	0,2991	23	24	25	25	26	26	1,2648

Fonte: Elaborada pelo autor.

ajustado para os distintos tamanhos de componentes na mistura, em que cada componente da mistura é diferenciado por símbolos.

No geral, as estimativas dos modelos ajustados apresentaram proximidade com os dados, sendo $\hat{p}_k = \frac{n_k}{N}$ e $\hat{\theta}_k$ próxima a média de compras para o k -ésimo componente, exceto em alguns casos.

Os gráficos apresentados, [Figura 3](#), assim como as estimativas apresentadas na [Tabela 7](#) indicam que quanto maior o número de componentes na mistura as estimativas de θ para o primeiro componente tendem a apresentar maior distância da média da quantidade de compras, tornando o modelo menos preciso.

Tal fato pode ser explicado pelo efeito do truncamento no zero, pois as menores estimativas foram obtidas para θ_1 . Sendo $\hat{\theta}_1 = 0,0434$ em que os modelos ajustados selecionaram no componente 1 apenas os clientes com 1 compra.

7.3 Algoritmo EMPMTZ

A aplicação do algoritmo EMPMTZ alcançou a convergência nas estimativas dos parâmetros da mistura de Poisson para dados truncados no zero para 2 e 3 componentes na mistura.

Os resultados das estimativas do modelo ajustado obtidos na implementação do algoritmo

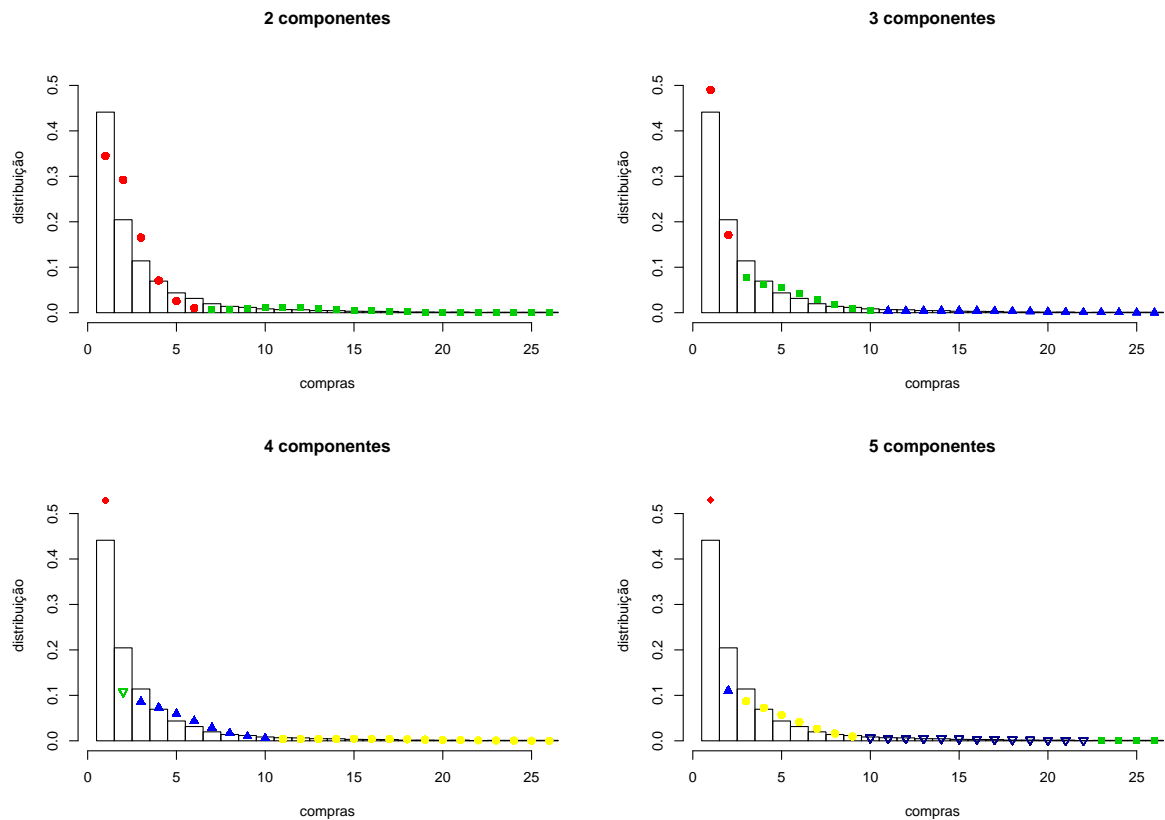


Figura 3 – Gráficos do modelo de mistura de Poisson truncadas no zero para K componentes ajustado através do algoritmo EMMPTZ.

Fonte: Elaborada pelo autor.

EMPMTZ constam na [Tabela 8](#), assim como medidas descritivas para cada componente da mistura.

Tabela 8 – Resultados do algoritmo EMPMTZ aplicado à quantidade de compras por cliente para K componentes na mistura e medidas descritivas dentro de cada componente.

K	estimativas				quantidade de compras						
	\hat{s}	\hat{q}_k	\hat{n}_k	$\hat{\lambda}_k$	Min	Q_1	Q_2	Média	Q_3	Max	Var
2	1	0,9379	19.041	2,5109	1	1	2	2	3	8	2,8313
	2	0,0621	1.261	13,4774	9	10	12	13	16	26	19,7164
3	1	0,9240	18.759	2,4391	1	1	2	2	3	7	2,3723
	2	0,0668	1.357	11,1527	8	9	10	11	13	18	8,0940
	3	0,0092	186	22,1344	19	20	22	22	24	26	5,6954

Fonte: Elaborada pelo autor.

As estimativas $\hat{\lambda}_k$ resultaram em valores próximos à média do k -ésimo componente e para as estimativas \hat{q}_k obtivemos valores iguais a $\frac{n_k}{N}$, apontando indícios de bom ajuste.

A Figura 4 mostra os histogramas da variável quantidade de compras com indicação dos modelos ajustados para $K = 2$ componentes na mistura à esquerda e para $K = 3$ à direita.

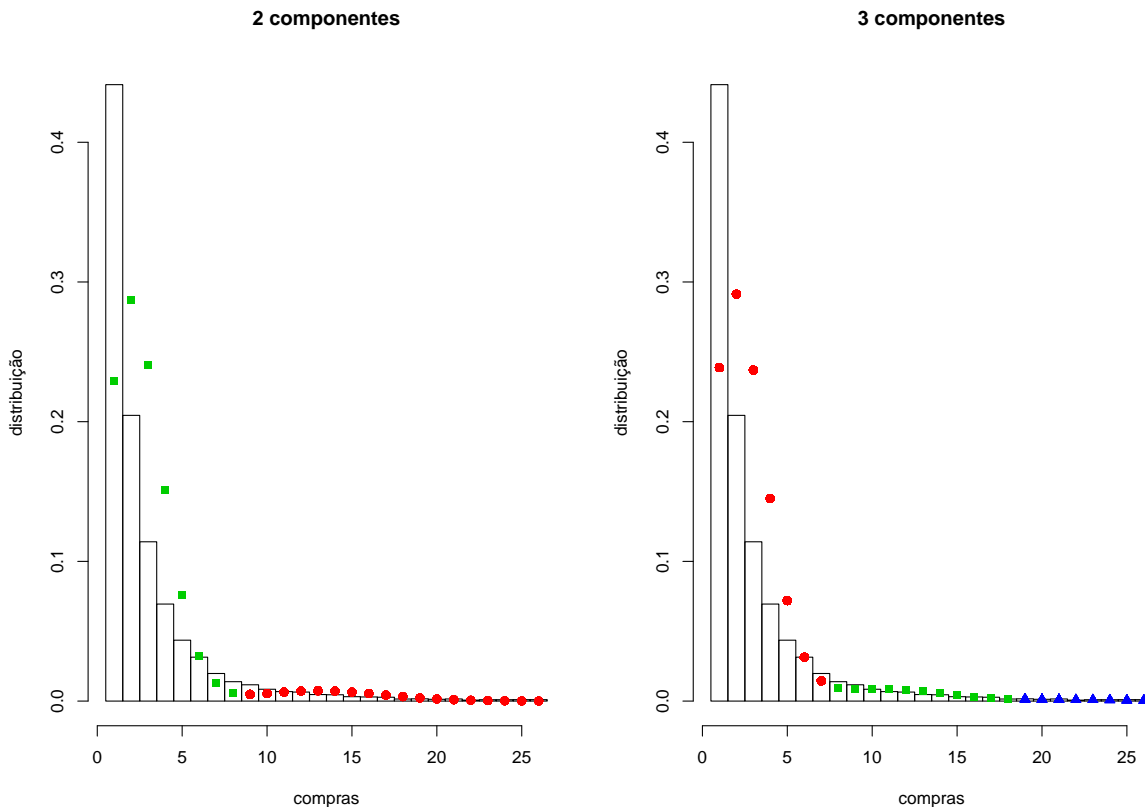


Figura 4 – Gráficos do modelo de mistura de Poisson truncadas no zero para K componentes ajustado através do algoritmo EMPMTZ.

Fonte: Elaborada pelo autor.

7.4 Comparação dos modelos ajustados

O algoritmo EMMPTZ e do algoritmo EMPMTZ performaram com resultados bastante distintos. A Tabela 9 mostra a quantidade de iterações para a convergência dos parâmetros de interesse, assim como o tempo de processamento, em segundos.

Notou-se que para o algoritmo EMMPTZ foi necessário um número bem menor de iterações, comparado ao algoritmo EMPMTZ. Conseqüentemente, o algoritmo EMMPTZ demandou menos tempo de processamento.

Apesar dos dois algoritmos implementados apresentarem indícios de bons ajustes, o algoritmo EMMPTZ indicou menores limitações para a solução do problema do truncamento na mistura. Sendo que, o algoritmo EMPMTZ não alcançou a convergência nas estimativas dos parâmetros do modelo para os casos com $K = 4$ e $K = 5$ componentes na mistura, representados na Tabela 9 pelo símbolo (\circ), denotando maior limitação. Além disso, a comparação dos

Tabela 9 – Resultados da quantidade de iterações e do tempo de processamento, em segundos, na aplicação do algoritmo EMMPTZ e do algoritmo EMPMTZ.

K	Algoritmo EMMPTZ		Algoritmo EMPMTZ	
	iterações	tempo	iterações	tempo
2	3	1,52	10	6,73
3	2	1,45	12	11,48
4	131	112,54	(o)	(o)
5	3.131	3.347,46	(o)	(o)

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros

Fonte: Elaborada pelo autor.

resultados dos métodos EMMPTZ e EMPMTZ ficou prejudicada pela não convergência para valores de $K \geq 4$ no algoritmo EMPMTZ.

Para selecionar o melhor modelo, foram considerados os critérios de seleção de modelos *AIC* e *BIC*. A medida *AIC* – *Akaike information criterion* – desenvolvida por Akaike (1974) é definida por

$$AIC(K)_{EMMPTZ} = -2l(\hat{\theta}, \hat{\mathbf{p}} | \mathbf{X}, \hat{\mathbf{S}}) + \gamma_K \quad (7.1)$$

para o modelo ajustado através do algoritmo EMMPTZ em que γ_K é a quantidade de parâmetros do modelo ajustado e $l(\hat{\theta}, \hat{\mathbf{p}} | \mathbf{X}, \hat{\mathbf{S}})$ é o logaritmo natural da função de verossimilhança dado pela Equação 4.2.

Para o modelo estipulado pelo algoritmo EMPMTZ, definiu se

$$AIC(K)_{EMPMTZ} = -2l(\hat{\lambda}, \hat{\mathbf{q}} | \mathbf{X}, \hat{\mathbf{S}}) + \gamma_K, \quad (7.2)$$

em que $l(\hat{\lambda}, \hat{\mathbf{q}} | \mathbf{X}, \hat{\mathbf{S}})$ é definida pela Equação 5.4.

E a medida *BIC* – *Bayesian information criterion* – introduzido por Schwarz *et al.* (1978) é formulada por

$$BIC(K)_{EMMPTZ} = -2l(\hat{\theta}, \hat{\mathbf{p}} | \mathbf{X}, \hat{\mathbf{S}}) + \gamma_K \ln(N) \quad (7.3)$$

para o modelo ajustado por intermédio do algoritmo EMMPTZ e por

$$BIC(K)_{EMPMTZ} = -2l(\hat{\lambda}, \hat{\mathbf{q}} | \mathbf{X}, \hat{\mathbf{S}}) + \gamma_K \ln(N) \quad (7.4)$$

para o modelo determinado pelo algoritmo EMPMTZ.

A Tabela 10 apresenta os valores obtidos nas medidas *AIC* e *BIC* para os dois modelos em discussão. Sendo que, o menor valor para *AIC* e para *BIC* são indicativos de melhor ajuste do modelo proposto aos dados. Considerou se, segundo as medidas *AIC* e *BIC*, que o melhor modelo é o modelo ajustado através do algoritmo EMPMTZ com 3 componentes na mistura.

Tabela 10 – Valores AIC e BIC para os modelos ajustados.

K	Algoritmo EMMPTZ		Algoritmo EMPMTZ	
	AIC	BIC	AIC	BIC
2	373.397,3	373.433,0	119.233,0	119.268,6
3	369.383,4	369.436,9	118.206,3	118.259,8
4	317.910,2	317.981,6	(o)	(o)
5	320.000,3	320.089,5	(o)	(o)

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros

Fonte: Elaborada pelo autor.

O modelo selecionado pelos critérios *AIC* e *BIC* para o método EMMPTZ é o modelo de mistura de Poisson truncada no zero com 4 componentes, dado por

$$\begin{aligned}
 Pr[X_i = x_i, S_i = s_i | \hat{\mathbf{p}}, \hat{\boldsymbol{\theta}}] &= \hat{p}_1 \frac{\hat{\theta}_1^{x_i}}{(e^{\hat{\theta}_1} - 1)x_i!} \mathbb{1}_{[S_i=1]} + \hat{p}_2 \frac{\hat{\theta}_2^{x_i}}{(e^{\hat{\theta}_2} - 1)x_i!} \mathbb{1}_{[S_i=2]} \\
 &+ \hat{p}_3 \frac{\hat{\theta}_3^{x_i}}{(e^{\hat{\theta}_3} - 1)x_i!} \mathbb{1}_{[S_i=3]} + \hat{p}_4 \frac{\hat{\theta}_4^{x_i}}{(e^{\hat{\theta}_4} - 1)x_i!} \mathbb{1}_{[S_i=4]} \\
 &= 0,4412 \frac{0,0434^{x_i}}{(e^{0,0434} - 1)x_i!} \mathbb{1}_{[S_i=1]} + 0,2045 \frac{1,5942^{x_i}}{(e^{1,5942} - 1)x_i!} \mathbb{1}_{[S_i=2]} \\
 &+ 0,3123 \frac{4,6483^{x_i}}{(e^{4,6483} - 1)x_i!} \mathbb{1}_{[S_i=3]} + 0,0420 \frac{15,4249^{x_i}}{(e^{15,4249} - 1)x_i!} \mathbb{1}_{[S_i=4]}
 \end{aligned} \tag{7.5}$$

em que $x_i \in \{1, 2, \dots, 26\}$, $s_i \in \{1, 2, 3, 4\}$ e $\mathbb{1}_{[S_i=s_i]}$ é a função indicadora que determina em qual componente da mistura pertence o i -ésimo elemento da variável X e X é a quantidade de compras por cliente em uma determinada loja.

A Tabela 11 apresenta os valores das estimativas dos parâmetros de interesse para o modelo formulado na Equação 7.5 para cada componente da mistura.

Tabela 11 – Valores das estimativas dos parâmetros $\hat{\theta}_k$ e \hat{p}_k para o modelo ajustado através do algoritmo EMMPTZ.

k	estimativas	
	$\hat{\theta}_k$	\hat{p}_k
1	0,0434	0,4412
2	1,5942	0,2045
3	4,6483	0,3123
4	15,4249	0,0420

Fonte: Elaborada pelo autor.

Considerando o ajuste feito através do algoritmo EMPMTZ o modelo selecionado foi o modelo de mistura de Poisson para variável truncada no zero com 3 componentes, formulado por

$$\begin{aligned}
 Pr[X_i = x_i, S_i = s_i | \mathbf{q}, \lambda] &= \frac{\hat{q}_1}{\delta} \left(\frac{e^{-\hat{\lambda}_1} \hat{\lambda}_1^{x_i}}{x_i!} \right)^{\mathbb{1}_{[S_i=1]}} + \frac{\hat{q}_2}{\delta} \left(\frac{e^{-\hat{\lambda}_2} \hat{\lambda}_2^{x_i}}{x_i!} \right)^{\mathbb{1}_{[S_i=2]}} \\
 &+ \frac{\hat{q}_3}{\delta} \left(\frac{e^{-\hat{\lambda}_3} \hat{\lambda}_3^{x_i}}{x_i!} \right)^{\mathbb{1}_{[S_i=3]}} \\
 &= \frac{0,9240}{\delta} \left(\frac{e^{-2,4391} 2,4391^{x_i}}{x_i!} \right)^{\mathbb{1}_{[S_i=1]}} + \frac{0,0668}{\delta} \left(\frac{e^{-11,1527} 11,1527^{x_i}}{x_i!} \right)^{\mathbb{1}_{[S_i=2]}} \\
 &+ \frac{0,0092}{\delta} \left(\frac{e^{-22,1344} 22,1344^{x_i}}{x_i!} \right)^{\mathbb{1}_{[S_i=3]}} \quad (7.6)
 \end{aligned}$$

em que $\delta = 1 - \sum_{k=1}^K \hat{q}_k e^{-\hat{\lambda}_k}$.

Os valores para as estimativas dos parâmetros de cada componente constam na [Tabela 12](#), para o modelo representado na [Equação 7.6](#), para cada componente da mistura.

Tabela 12 – Valores das estimativas dos parâmetros $\hat{\lambda}_k$ e \hat{q}_k para o modelo ajustado através do algoritmo EMPMTZ.

k	estimativas	
	$\hat{\lambda}_k$	\hat{q}_k
1	2,4391	0,9240
2	11,1527	0,0668
3	22,1344	0,0092

Fonte: Elaborada pelo autor.

DISCUSSÃO

Na pesquisa desenvolvida para essa dissertação foram combinadas variadas metodologias para estimar os parâmetros do modelo de mistura de Poisson com truncamento no zero. Para a mistura, empregou-se a variável não observável S para definir em qual componente cada elemento pertence. E, especificando o estudo para a distribuição Poisson, foram estudadas duas abordagens para resolver o problema do truncamento no zero.

Na primeira abordagem efetuou-se o truncamento no zero em cada componente da mistura e, além disso, foi estipulado tal truncamento como valor faltante. Posteriormente, foi discutida a visão em que o truncamento é feito diretamente no modelo. Sendo assim, foram desenvolvidos o algoritmo EMMPTZ e o algoritmo EMPMTZ, com a finalidade de analisar suas performances na estimação dos parâmetros de interesse.

Em comparação, o algoritmo EMMPTZ apresentou menor tempo de processamento e estimativas próximas do verdadeiro valor do parâmetro. Ao passo que, o algoritmo EMPMTZ apresentou maior tempo de processamento e estimativas próximas do verdadeiro valor do parâmetro. Ambos os algoritmos apresentaram maior viés nas estimativas para parâmetros, θ_k ou λ_k , próximos de 1. Sendo que, neste caso as estimativas $\hat{\lambda}_k$ indicaram menor viés comparadas a $\hat{\theta}_k$.

Além do mais, o algoritmo EMMPTZ apontou o diferencial de indicação de menores quantidades de componentes na mistura, principalmente para os casos de amostras pequenas. O algoritmo EMMPTZ apresentou boa performance, apesar da maior complexidade em sua implementação. Sendo o algoritmo EMPMTZ menos complexo, porém um método mais limitado devido à não convergência dos parâmetros de interesse em alguns casos.

Apesar da boa performance no estudo de simulação para as duas abordagens os modelos levaram à conclusões diferentes no que confere a quantidade de componentes do modelo de mistura.

A aplicação dos métodos no conjunto de dados em estudo apontou como melhor modelo o modelo ajustado com o truncamento na resultante do modelo, como discutido no [Capítulo 5](#). O melhor modelo foi selecionado usando os critérios *AIC* e *BIC*. As duas abordagens estudadas levaram à conclusões diferentes no que confere ao número de componentes na mistura e também na classificação das observações nas suas respectivas composições (componentes).

Como propostas para continuidade da pesquisa recomendamos a aplicação da metodologia Bayesiana, com a implementação do método *Reversible Jump* que permite a seleção do modelo que melhor se ajusta aos dados ao mesmo tempo que estima os seus parâmetros, e também fornece uma medida desse ajuste na forma de probabilidade à *posteriori* (condicionada aos dados observados).

O método *Reversible Jump* na sua variante *DDRJ – Data Driven Reversible Jump* – ver [Saraiva e Milan \(2012\)](#) e [Zuanetti e Milan \(2017\)](#), é ideal para trabalhar com modelos de mistura quando o número de componentes, K , é desconhecido.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, Ieee, v. 19, n. 6, p. 716–723, 1974. Citado na página 63.
- ANDERSON, T. W.; ANDERSON, T. W.; ANDERSON, T. W.; ANDERSON, T. W.; MATHÉ-MATICIEN, E.-U. **An introduction to multivariate statistical analysis**. [S.l.]: Wiley New York, 1958. v. 2. Citado na página 21.
- BICKEL, P. J.; DOKSUM, K. A. **Mathematical statistics: basic ideas and selected topics**. [S.l.]: CRC Press, 2015. v. 2. Citado nas páginas 31 e 35.
- BOHNING, D.; SCHLATTMANN, P.; LINDSAY, B. Computer-assisted analysis of mixtures (ca man): Statistical algorithms. **Biometrics**, JSTOR, p. 283–303, 1992. Citado na página 30.
- CASELLA, G.; BERGER, R. L. **Statistical inference**. [S.l.]: Duxbury Pacific Grove, CA, 2002. v. 2. Citado na página 31.
- DAVID, F.; JOHNSON, N. The truncated poisson. **Biometrics**, JSTOR, v. 8, n. 4, p. 275–285, 1952. Citado nas páginas 21, 33 e 35.
- DIETZ, E.; BÖHNING, D. On estimation of the poisson parameter in zero-modified poisson models. **Computational Statistics & Data Analysis**, Elsevier, v. 34, n. 4, p. 441–459, 2000. Citado nas páginas 22, 34 e 35.
- FRÜHWIRTH-SCHNATTER, S. **Finite mixture and Markov switching models**. [S.l.]: Springer Science & Business Media, 2006. Citado na página 21.
- GROGGER, J. T.; CARSON, R. T. Models for truncated counts. **Journal of applied econometrics**, Wiley Online Library, v. 6, n. 3, p. 225–238, 1991. Citado na página 31.
- III, J. L. N.; POLLOCK, K. H. Nonparametric mle under two closed capture-recapture models with heterogeneity. **Biometrics**, JSTOR, p. 639–649, 1996. Citado nas páginas 22 e 43.
- KARLIS, D.; XEKALAKI, E. Mixed poisson distributions. **International Statistical Review**, Wiley Online Library, v. 73, n. 1, p. 35–58, 2005. Citado na página 39.
- LINK, W. A. Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. **Biometrics**, Wiley Online Library, v. 59, n. 4, p. 1123–1130, 2003. Citado nas páginas 22 e 43.
- MCLACHLAN, G.; KRISHNAN, T. **The EM algorithm and extensions**. [S.l.]: John Wiley & Sons, 2007. v. 382. Citado nas páginas 22, 28 e 30.
- MCLACHLAN, G.; PEEL, D. **Finite mixture models**. [S.l.]: John Wiley & Sons, 2004. Citado nas páginas 22 e 25.
- MCLACHLAN, G. J.; BASFORD, K. E. **Mixture models. Inference and applications to clustering**. [S.l.: s.n.], 1988. v. 1. Citado na página 21.

- MEIRA, S. A. **Modelo de mistura com dependência Markoviana de primeira ordem**. Tese (Doutorado) — Universidade Federal de São Carlos, 2014. Disponível em: <<https://repositorio.ufscar.br/handle/ufscar/4587>>. Citado nas páginas 22, 26, 27, 28 e 29.
- PLEDGER, S. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. **Biometrics**, Wiley Online Library, v. 56, n. 2, p. 434–442, 2000. Citado nas páginas 22 e 43.
- SARAIVA, E. F.; MILAN, L. A. Clustering gene expression data using a posterior split-merge-birth procedure. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 39, n. 3, p. 399–415, 2012. Citado na página 68.
- SCHWARZ, G. *et al.* Estimating the dimension of a model. **The annals of statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978. Citado na página 63.
- WEDEL, M.; DESARBO, W. S.; BULT, J. R.; RAMASWAMY, V. A latent class poisson regression model for heterogeneous count data. **Journal of Applied Econometrics**, Wiley Online Library, v. 8, n. 4, p. 397–411, 1993. Citado na página 39.
- ZUANETTI, D. A.; MILAN, L. A. A generalized mixture model applied to diabetes incidence data. **Biometrical Journal**, Wiley Online Library, 2017. Citado na página 68.

RESULTADOS DO ALGORITMO EMMPTZ

Tabela 13 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson truncadas no zero e o tempo de processamento em segundos para $K = 2$ componentes.

K	N	θ_{real}	quantidade de iterações					tempo	
			Min	Q_1	Q_2	Média	Q_3		Max
2	10	(2, 3)	1	2	2	6	3	69	0,39
		(3, 9)	1	1	2	2	2	4	0,26
		(4, 20)	1	2	2	2	2	3	0,21
	30	(2, 3)	1	2	2	3	4	18	0,51
		(3, 9)	1	2	2	2	3	4	0,37
		(4, 20)	1	2	2	2	2	4	0,36
	50	(2, 3)	1	2	3	5	6	55	1,55
		(3, 9)	1	2	2	2	2	4	0,43
		(4, 20)	1	2	2	2	2	4	0,44
	100	(2, 3)	2	2	5	6	8	28	1,97
		(3, 9)	1	2	2	2	2	3	0,78
		(4, 20)	1	2	2	2	3	4	0,77
	1.000	(2, 3)	2	2	7	8	9	45	22,14
		(3, 9)	2	2	2	2	2	2	5,62
		(4, 20)	1	2	2	2	2	3	6,30
	10.000	(2, 3)	6	8	9	11	11	32	317,22
		(3, 9)	2	2	2	2	2	2	56,98
		(4, 20)	1	2	2	2	3	4	64,84

Fonte: Elaborada pelo autor.

Tabela 14 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson truncadas no zero e o tempo de processamento em segundos para $K = 3$ e 4 componentes.

K	N	θ_{real}	quantidade de iterações						tempo
			Min	Q_1	Q_2	Média	Q_3	Max	
3	10	(2, 3, 4)	1	2	15	29	38	171	1,72
		(3, 9, 15)	2	2	2	5	3	114	0,32
		(4, 20, 36)	2	2	2	3	2	30	0,30
	30	(2, 3, 4)	1	5	24	59	66	508	7,98
		(3, 9, 15)	2	2	3	5	4	143	0,88
		(4, 20, 36)	1	2	3	3	4	8	0,54
	50	(2, 3, 4)	2	10	31	78	81	900	17,01
		(3, 9, 15)	2	2	3	3	4	21	0,75
		(4, 20, 36)	2	3	4	4	4	10	0,87
	100	(2, 3, 4)	3	14	37	68	68	1.048	28,59
		(3, 9, 15)	2	2	3	7	4	256	2,88
		(4, 20, 36)	3	4	4	5	5	9	2,18
	1.000	(2, 3, 4)	4	14	29	49	60	374	194,36
		(3, 9, 15)	2	2	3	4	3	35	16,83
		(4, 20, 36)	4	5	6	6	8	9	23,80
	10.000	(2, 3, 4)	4	17	35	59	70	388	2.348,06
		(3, 9, 15)	3	3	3	8	11	45	300,06
		(4, 20, 36)	5	6	7	7	8	9	273,97
4	10	(2, 3, 4, 5)	2	48	138	315	377	2.586	22,95
		(3, 9, 15, 21)	2	3	12	44	33	523	3,32
		(4, 20, 36, 52)	2	3	4	13	9	212	1,10
	30	(2, 3, 4, 5)	2	141	498	1.468	1.368	21.050	470,44
		(3, 9, 15, 21)	2	3	12	73	49	1.635	13,08
		(4, 20, 36, 52)	2	2	3	6	5	92	1,26
	50	(2, 3, 4, 5)	2	128	756	2.014	2.356	18.040	531,68
		(3, 9, 15, 21)	2	3	22	159	69	3.892	42,16
		(4, 20, 36, 52)	2	3	4	5	5	45	2,00
	100	(2, 3, 4, 5)	2	259	1.327	6.475	5.590	59.930	3.545,62
		(3, 9, 15, 21)	2	3	8	52	66	469	27,89
		(4, 20, 36, 52)	2	3	4	4	5	11	2,51
	1.000	(2, 3, 4, 5)	2	434	78	122	160	781	8,79
		(3, 9, 15, 21)	3	27	74	99	133	519	(*)
		(4, 20, 36, 52)	1	26	59	79	111	383	3.532,86
	10.000	(2, 3, 4, 5)	8	18	24	32	38	137	114,36
		(3, 9, 15, 21)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
		(4, 20, 36, 52)	3	16	52	61	90	255	2.651,10
		(4, 20, 36, 52)	8	16	31	44	57	203	1.973,67

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

(*) : amostras que indicaram $K - 1$ componentes na mistura.

Fonte: Elaborada pelo autor.

Tabela 15 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson truncadas no zero e o tempo de processamento em segundos para $K = 5$ componentes.

K	N	θ_{real}	quantidade de iterações						tempo
			Min	Q_1	Q_2	Média	Q_3	Max	
5	10	(2, 3, 4, 5, 6)	3	26	74	98	133	519	8,79
			2	44	78	122	160	781	(*)
		(3, 9, 15, 21, 27)	2	8	19	39	48	392	4,06
			2	64	88	134	225	348	(*)
		(4, 20, 36, 52, 68)	2	3	8	28	18	422	2,14
	30	(2, 3, 4, 5, 6)	33	937	1.624	2.656	3.008	15.660	504,16
			742	1.456	3.295	3.088	4.426	6.402	(*)
		(3, 9, 15, 21, 27)	2	63	160	639	547	10.700	621,85
			2	3	4	33	14	455	6,52
		(4, 20, 36, 52, 68)							
	50	(2, 3, 4, 5, 6)	13	1.981	4.590	6.714	9.292	33.880	1.855,40
			990	3.399	5.444	5.069	6.234	9.383	(*)
		(3, 9, 15, 21, 27)	2	87	309	1.220	892	19.410	348,64
			2	3	5	53	30	1.057	15,98
		(4, 20, 36, 52, 68)							
	100	(2, 3, 4, 5, 6)	1.009	6.648	15.060	17.060	15.800	51.360	11.209,95
			441	5.886	15.230	24.970	36.100	170.600	(*)
		(3, 9, 15, 21, 27)	2	215	1.052	2.348	2.403	37.730	1.085,14
			2	3	5	77	20	2.411	42,40
		(4, 20, 36, 52, 68)							
1.000	(2, 3, 4, 5, 6)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	
		2	2	1.450	3.980	4.773	93.760	19.656,26	
	(3, 9, 15, 21, 27)	2	6	7	168	11	5.545	873,92	
	(4, 20, 36, 52, 68)								
10.000	(2, 3, 4, 5, 6)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	
		2	71	781	1.355	1.837	9.500	61.838,77	
	(3, 9, 15, 21, 27)	2	7	8	46	8	2.339	2.009,27	
	(4, 20, 36, 52, 68)								

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

(*) : amostras que indicaram $K - 1$ componentes na mistura.

Fonte: Elaborada pelo autor.

Tabela 16 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 2 componentes.

N	θ_{real}	\hat{n}_k		\hat{p}_k					
		Q_2	Média	Min	Q_1	Q_2	Média	Q_3	Max
10	2	6	6	0,1000	0,4750	0,6000	0,5620	0,7000	0,9000
	3	4	4	0,1000	0,3000	0,4000	0,4380	0,5250	0,9000
	3	5	5	0,1000	0,4000	0,5000	0,5050	0,6000	0,9000
	9	5	5	0,1000	0,4000	0,5000	0,4950	0,6000	0,9000
	4	5	5	0,1000	0,4000	0,5000	0,5080	0,6250	0,8000
	20	5	5	0,2000	0,3750	0,5000	0,4920	0,6000	0,9000
30	2	19	19	0,1330	0,5250	0,6333	0,6400	0,7420	0,9670
	3	11	11	0,0333	0,2580	0,3670	0,3600	0,4750	0,8670
	3	15	15	0,1670	0,4000	0,5000	0,4960	0,5750	0,8670
	9	15	15	0,1333	0,4250	0,5000	0,5040	0,6000	0,8333
	4	15	15	0,2000	0,4333	0,5000	0,5111	0,5670	0,8000
	20	15	15	0,2000	0,4333	0,5000	0,4890	0,5670	0,8000
50	2	32	31	0,1600	0,4950	0,6400	0,6250	0,7600	0,9800
	3	18	19	0,0200	0,2400	0,3600	0,3750	0,5050	0,8400
	3	26	26	0,3400	0,4600	0,5200	0,5150	0,5600	0,7400
	9	24	24	0,2600	0,4400	0,4800	0,4850	0,5400	0,6600
	4	25	24	0,3200	0,4400	0,5000	0,4860	0,5400	0,6400
	20	25	25	0,3600	0,4600	0,5000	0,5140	0,5600	0,6800
100	2	67	61	0,1700	0,4880	0,6700	0,6090	0,7200	0,9900
	3	33	39	0,0100	0,2800	0,3300	0,3910	0,5130	0,8300
	3	53	53	0,3400	0,4980	0,5300	0,5310	0,5700	0,7300
	9	47	47	0,2700	0,4300	0,4700	0,4690	0,5030	0,6600
	4	50	50	0,4000	0,4600	0,5000	0,5030	0,5400	0,6300
	20	50	50	0,3700	0,4600	0,5000	0,4970	0,5400	0,6000
1.000	2	715	643	0,2090	0,4960	0,7150	0,6430	0,7370	0,9999
	3	285	357	0,0001	0,2630	0,2860	0,3570	0,5040	0,7910
	3	582	583	0,4870	0,5720	0,5820	0,5830	0,5950	0,6790
	9	418	417	0,3210	0,4050	0,4180	0,4170	0,4280	0,5130
	4	502	501	0,4420	0,4880	0,5020	0,5010	0,5130	0,5340
	20	498	499	0,4660	0,4870	0,4990	0,4990	0,5120	0,5580
10.000	2	5.134	6.096	0,2278	0,5039	0,5134	0,6096	0,9998	0,9999
	3	4.866	3.904	0,0001	0,0002	0,4866	0,3904	0,4961	0,7722
	3	5.880	6.018	0,5754	0,5834	0,5880	0,6018	0,5950	0,7195
	9	4.120	3.982	0,2805	0,4050	0,4120	0,3982	0,4166	0,4246
	4	5.010	5.009	0,4850	0,4979	0,5010	0,5009	0,5053	0,5132
	20	4.990	4.991	0,4868	0,4947	0,4990	0,4991	0,5021	0,5150

Fonte: Elaborada pelo autor.

Tabela 17 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 2 componentes.

N	θ_{real}	$\hat{\theta}_k$						$Var[\hat{\theta}_k]$	
		Min	Q_1	Q_2	Média	Q_3	Max	Q_2	Média
10	2	0,0300	0,7780	1,1430	1,1840	1,5000	3,7140	0,2520	0,2940
	3	1,6000	3,4210	3,6667	4,1010	4,8500	9,0000	0,9520	1,4980
	3	0,0440	1,8000	2,2500	2,5120	3,2130	5,8750	0,5840	0,6400
	9	5,5000	7,8540	9,0000	9,1380	10,1800	14,0000	1,8810	2,5230
	4	1,2500	3,4210	3,7750	3,9820	4,5790	9,5000	0,7810	1,0550
	20	12,6000	18,0000	20,0000	19,7200	21,0000	27,0000	4,0200	4,6010
30	2	0,0400	1,0430	1,4290	1,4080	1,7520	2,6470	0,0950	0,0999
	3	2,4000	3,7140	4,6790	4,8110	5,2060	10,0000	0,4640	1,3530
	3	1,0667	2,3078	2,7500	2,7420	3,1010	4,5910	0,2130	0,2222
	9	7,1111	8,8655	9,4175	9,4065	10,1500	12,8800	0,6190	0,6887
	4	2,7500	3,7444	4,0620	4,0490	4,3720	6,0000	0,2830	0,2930
	20	16,3300	19,1300	20,0000	19,8900	20,6800	23,3300	1,3070	1,4400
50	2	0,042	0,9755	1,3400	1,3040	1,7575	2,8655	0,0555	0,0540
	3	2,3230	3,4950	4,6550	4,8490	5,2980	10,0000	0,2630	1,2270
	3	1,7240	2,5210	2,7777	2,7920	3,0290	3,9674	0,1222	0,1270
	9	7,8400	8,8887	9,2890	9,3690	9,7450	12,0000	0,3830	0,4030
	4	1,9520	3,6040	3,8540	3,9920	4,3830	5,1111	0,1720	0,1780
	20	17,1600	19,6100	19,9700	19,9600	20,5000	21,8800	0,7860	0,7910
100	2	0,0420	0,9864	1,4374	1,2690	1,5974	2,6790	0,0290	0,0260
	3	2,3460	3,7060	4,5742	4,6950	4,8811	11,0000	0,1474	0,8560
	3	2,2370	2,6990	2,9080	2,9320	3,1060	3,7090	0,0630	0,0630
	9	8,3750	9,1510	9,4300	9,4920	9,8140	11,4100	0,1990	0,2070
	4	3,2750	3,7690	3,9380	3,9740	4,1810	4,7200	0,0850	0,0850
	20	18,2200	19,5100	19,9800	19,9500	20,3800	21,6500	0,3980	0,4060
1.000	2	0,0440	0,9200	1,5500	1,3520	1,6300	2,6120	0,0033	0,0022
	3	2,9900	3,8170	4,6850	5,6820	4,7984	14,0000	0,0170	1,7150
	3	2,6960	3,1880	3,2400	3,2300	3,2950	3,7460	0,0066	0,0066
	9	9,0920	9,8400	9,9360	9,9300	10,0300	10,7500	0,0244	0,0244
	4	3,7430	3,9111	3,9888	3,9811	4,0380	4,2322	0,0088	0,0088
	20	19,5700	19,8600	20,0100	20,0100	20,1300	20,4200	0,0400	0,0400
10.000	2	0,0434	0,9234	0,9411	1,2470	2,4910	2,5560	0,0003	0,0002
	3	3,0890	3,8700	3,9080	6,1390	12,0000	13,0000	0,0009	3,0510
	3	3,1580	3,2140	3,2390	3,3230	3,2730	4,0960	0,0006	0,0006
	9	9,8410	9,9050	9,9310	10,0600	9,9860	11,2300	0,0024	0,0026
	4	3,9080	3,9520	4,0210	4,0070	4,0520	4,1080	0,0009	0,0009
	20	19,8200	19,9700	20,0700	20,0400	20,1000	20,2300	0,0044	0,0044

Fonte: Elaborada pelo autor.

Tabela 18 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 3 componentes.

N	θ_{real}	\hat{n}_k		\hat{p}_k					
		Q_2	Média	Min	Q_1	Q_2	Média	Q_3	Max
10	2	3	3	0,1000	0,2000	0,3000	0,3130	0,4000	0,7000
	3	4	4	0,1000	0,3000	0,4000	0,3800	0,5000	0,7000
	4	3	3	0,1000	0,2000	0,3000	0,3070	0,4000	0,6000
	3	3	3	0,1000	0,2000	0,3000	0,2760	0,3000	0,8000
	9	4	4	0,1000	0,3750	0,4500	0,4390	0,5000	0,8000
	15	3	3	0,1000	0,2000	0,3000	0,2850	0,4000	0,7000
	4	3	3	0,1000	0,2000	0,3000	0,3150	0,4000	0,6000
	20	4	4	0,1000	0,3000	0,4000	0,4030	0,5000	0,8000
	36	3	3	0,1000	0,2000	0,3000	0,2820	0,4000	0,6000
30	2	10	10	0,1333	0,2333	0,3167	0,3177	0,4000	0,6000
	3	11	12	0,1667	0,3000	0,3667	0,3843	0,4333	0,8333
	4	9	9	0,0333	0,2333	0,3000	0,2980	0,3667	0,6333
	3	7	7	0,0333	0,1667	0,2333	0,2320	0,2667	0,6333
	9	17	16	0,1000	0,4583	0,5667	0,5440	0,6333	0,7667
	15	6	7	0,0333	0,1667	0,2167	0,2240	0,3000	0,5667
	4	10	10	0,1333	0,2667	0,3333	0,3203	0,3667	0,5000
	20	12	12	0,2000	0,3000	0,4000	0,3950	0,4667	0,7333
	36	8	9	0,0333	0,2333	0,2667	0,2847	0,3333	0,5667
50	2	16	16	0,0400	0,2150	0,3200	0,3118	0,4000	0,6400
	3	16	17	0,1400	0,2600	0,3200	0,3468	0,4200	0,7200
	4	17	17	0,0200	0,2600	0,3400	0,3414	0,4200	0,7800
	3	11	12	0,0400	0,1800	0,2200	0,2442	0,3000	0,5000
	9	29	28	0,2800	0,4800	0,5800	0,5570	0,6400	0,8400
	15	10	10	0,0200	0,1600	0,2000	0,1988	0,2400	0,5800
	4	17	17	0,1600	0,2800	0,3400	0,3300	0,3650	0,5200
	20	20	19	0,1600	0,3200	0,4000	0,3874	0,4400	0,6600
	36	14	14	0,0600	0,2200	0,2800	0,2826	0,3400	0,4800
100	2	34	32	0,0300	0,2100	0,3400	0,3236	0,4200	0,6100
	3	33	34	0,1300	0,2500	0,3300	0,3385	0,4000	0,8000
	4	34	34	0,0100	0,2800	0,3400	0,3379	0,3925	0,8400
	3	25	26	0,1200	0,2100	0,2500	0,2631	0,3100	0,5200
	9	56	54	0,1500	0,5000	0,5600	0,5495	0,6225	0,8100
	15	19	19	0,0100	0,1400	0,1900	0,1874	0,2225	0,4700
	4	32	32	0,2300	0,2800	0,3200	0,3241	0,3600	0,4600
	20	36	35	0,1800	0,3075	0,3600	0,3503	0,3900	0,5000
	36	33	33	0,1500	0,2900	0,3300	0,3256	0,3700	0,5300

Fonte: Elaborada pelo autor.

Tabela 19 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 3 componentes.

N	θ_{real}	$\hat{\theta}_k$						$Var[\hat{\theta}_k]$	
		Min	Q_1	Q_2	Média	Q_3	Max	Q_2	Média
10	2	0,0303	0,0441	1,6750	1,9530	3,0000	6,0000	0,6825	0,8146
	3	0,0375	2,0000	3,0000	2,8060	3,5000	6,0000	0,8988	0,9849
	4	0,0435	3,0000	4,0000	4,3220	5,0830	11,0000	1,4140	1,8880
	3	0,0303	1,3330	1,8750	2,4740	3,0000	10,5000	0,8647	1,2100
	9	1,0000	6,5000	8,5500	8,2780	10,0400	13,6700	1,8620	2,2530
	15	3,0000	13,5800	16,3300	16,0700	18,6200	26,0000	5,5560	7,5760
	4	0,0303	2,7290	3,7080	4,1440	4,7500	29,0000	1,2220	1,5720
	20	4,2500	17,5000	20,0000	19,9600	22,8100	32,7500	5,1490	6,2870
	36	5,5000	34,0000	37,0000	36,4500	40,0800	50,5000	13,1700	16,1400
30	2	0,0397	0,0441	1,5810	1,6620	2,6670	5,6360	0,1697	0,2012
	3	0,0397	2,4170	2,8970	3,0020	3,6250	5,9000	0,3009	0,3136
	4	0,0397	2,7620	4,5000	4,3510	5,5430	12,0000	0,4397	1,2090
	3	0,0303	1,3750	1,9300	1,9700	2,3640	11,0000	0,3386	0,3826
	9	0,0375	7,4580	8,5810	8,5530	9,9110	12,5500	0,5077	0,5458
	15	2,1110	16,0000	17,1900	17,4300	19,0000	28,0000	2,5510	4,3300
	4	2,0000	3,3940	4,0000	3,9890	4,4710	6,1250	0,4494	0,4636
	20	13,1000	19,2500	20,6100	20,8700	22,3700	28,4500	1,7630	1,8600
	36	31,5000	35,7700	37,2200	37,4900	39,0000	46,0000	4,3110	5,3760
50	2	0,0407	0,0440	1,0730	1,6270	2,7050	9,0000	0,0751	0,1568
	3	0,0407	2,1240	2,7600	2,7400	3,2000	5,5000	0,1788	0,1919
	4	0,0407	3,1280	4,5180	4,3300	5,3740	12,0000	0,2503	0,6565
	3	0,0435	1,6360	1,9510	1,9750	2,2420	3,5830	0,1996	0,2170
	9	4,0530	7,8440	8,8460	8,9440	9,9560	13,8800	0,3082	0,3351
	15	13,7600	16,5600	17,8100	18,2300	19,1100	28,0000	1,7740	3,6170
	4	2,3330	3,5140	4,0000	3,9230	4,3120	5,0000	0,2567	0,2629
	20	16,8600	19,5500	20,5100	20,8600	21,6700	27,3000	1,0620	1,1420
	36	32,1100	36,1900	37,5000	37,9600	39,2900	47,3300	2,6150	3,1570
100	2	0,0422	0,0439	1,0490	1,5420	2,3310	9,0000	0,0365	0,0768
	3	0,0422	2,1720	2,8000	2,7110	3,2410	5,8710	0,0905	0,0961
	4	0,0423	2,8720	4,8100	4,0270	5,0990	11,0000	0,1281	0,3846
	3	1,1760	1,6880	2,0000	2,3970	2,3900	14,8600	0,1006	0,1100
	9	2,0000	8,3170	9,1690	9,1030	9,9360	12,6800	0,1621	0,1736
	15	6,3080	17,0000	17,8900	18,2200	19,1900	29,0000	0,9300	2,6640
	4	3,2690	3,7190	3,9280	3,9610	4,2240	4,8500	0,1271	0,1333
	20	17,7000	19,2700	20,0000	20,0300	20,6200	23,6200	0,5690	0,5905
	36	33,7400	35,8900	36,6900	36,7800	37,7400	39,6500	1,1050	1,1800

Fonte: Elaborada pelo autor.

Tabela 20 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 3 componentes.

N	θ_{real}	\hat{n}_k		\hat{p}_k					
		Q_2	Média	Min	Q_1	Q_2	Média	Q_3	Max
1.000	2	247	303	0,0010	0,1860	0,2470	0,3026	0,4212	0,8230
	3	373	349	0,0020	0,2235	0,3725	0,3485	0,4420	0,8330
	4	363	349	0,0010	0,3435	0,3625	0,3490	0,3828	0,8300
	3	278	272	0,1920	0,2215	0,2775	0,2720	0,2962	0,4230
	9	614	613	0,2640	0,5868	0,6135	0,6131	0,6452	0,7210
	15	124	115	0,0020	0,0895	0,1235	0,1149	0,1470	0,3400
	4	334	334	0,2960	0,3250	0,3340	0,3337	0,3430	0,3640
	20	345	348	0,3050	0,3350	0,3450	0,3486	0,3632	0,3940
	36	321	318	0,2740	0,3028	0,3210	0,3177	0,3320	0,3580
10.000	2	1873	2799	0,1743	0,1811	0,1873	0,2799	0,3753	0,8177
	3	4382	3663	0,0001	0,2231	0,4382	0,3663	0,4467	0,8253
	4	3710	3538	0,0001	0,3638	0,3710	0,3538	0,3766	0,5882
	3	2910	3150	0,0495	0,2854	0,2910	0,3150	0,3441	0,5034
	9	6270	6039	0,1287	0,6012	0,6270	0,6039	0,6350	0,6606
	15	843	810	0,0001	0,0045	0,0843	0,0811	0,0975	0,5074
	4	3348	3347	0,3205	0,3306	0,3348	0,3347	0,3382	0,3487
	20	3453	3463	0,3252	0,3389	0,3453	0,3463	0,3544	0,3662
	36	3228	3189	0,2935	0,3096	0,3228	0,3189	0,3282	0,3367

Fonte: Elaborada pelo autor.

Tabela 21 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 3 componentes.

N	θ_{real}	$\hat{\theta}_k$						$Var[\hat{\theta}_k]$	
		Min	Q_1	Q_2	Média	Q_3	Max	Q_2	Média
1.000	2	0,0434	0,0434	0,9479	1,4630	2,2000	13,0000	0,0034	0,2641
	3	0,0434	2,1870	2,5460	2,8060	3,0050	12,0000	0,0085	0,0692
	4	0,0434	2,2050	4,9990	4,0430	5,0920	14,0000	0,0139	0,4698
	3	1,6270	1,8050	2,2670	2,2270	2,4000	3,3870	0,0101	0,0101
	9	7,9430	9,2320	10,0400	10,0700	10,3300	12,9900	0,0166	0,0166
	15	15,6600	18,4100	19,0700	19,9100	19,7700	28,5000	0,1573	0,9523
	4	3,6650	3,8900	3,9990	4,0000	4,1070	4,4050	0,0127	0,0127
	20	19,2000	19,7700	20,0300	20,1000	20,4000	21,2800	0,0580	0,0578
	36	35,6100	36,3300	36,6500	36,6900	37,0300	37,9500	0,1141	0,1160
10.000	2	0,0434	0,0434	0,0434	1,3990	2,2040	5,0880	0,0001	0,0004
	3	0,0434	1,5940	2,2070	2,3040	2,8210	15,0000	0,0006	0,1507
	4	0,0434	2,8210	5,0470	4,3850	5,0680	17,0000	0,0014	0,5411
	3	0,0434	2,3380	2,3770	2,7740	2,8240	13,8600	0,0010	0,0010
	9	1,0470	10,4100	10,5000	10,9600	12,1100	15,6800	0,0017	0,0018
	15	1,7820	19,7600	20,0600	21,8400	25,9200	33,0000	0,0237	1,5500
	4	3,8100	3,9670	4,0190	4,0070	4,0520	4,1240	0,0013	0,0013
	20	19,6800	19,8400	19,9800	20,0700	20,3200	20,4500	0,0058	0,0058
	36	36,2600	36,4600	36,5800	36,6900	36,9500	37,2200	0,0113	0,0115

Fonte: Elaborada pelo autor.

Tabela 22 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 4 componentes.

N	θ_{real}	\hat{n}_k		\hat{p}_k					
		Q_2	Média	Min	Q_1	Q_2	Média	Q_3	Max
10	2	2	2	0,0000	0,2000	0,2000	0,2420	0,3000	0,5000
	3	2	2	0,0000	0,2000	0,2000	0,2430	0,3000	0,8000
	4	3	3	0,0000	0,2000	0,3000	0,2600	0,3000	0,6000
	5	3	3	0,0000	0,2000	0,3000	0,2550	0,3000	0,6000
	3	3	2	0,0000	0,2000	0,3000	0,2490	0,3000	0,6000
	9	2	2	0,1000	0,2000	0,2000	0,2410	0,3000	0,5000
	15	2	2	0,1000	0,2000	0,2000	0,2470	0,3000	0,6000
	36	3	3	0,1000	0,2000	0,3000	0,2630	0,3000	0,5000
	4	3	3	0,1000	0,2000	0,3000	0,2590	0,3000	0,7000
	20	2	2	0,1000	0,1000	0,2000	0,2350	0,3000	0,5000
	36	2	2	0,1000	0,2000	0,2000	0,2480	0,3000	0,6000
	52	3	3	0,1000	0,2000	0,3000	0,2580	0,3000	0,5000
30	2	8	8	0,1333	0,2000	0,2500	0,2597	0,3000	0,8333
	3	7	7	0,0333	0,2000	0,2333	0,2480	0,3000	0,6333
	4	7	7	0,1000	0,2000	0,2333	0,2450	0,2750	0,4333
	5	7	7	0,0333	0,2000	0,2333	0,2473	0,3000	0,4667
	3	8	8	0,0333	0,2333	0,2667	0,2757	0,3333	0,6000
	9	7	7	0,0667	0,1667	0,2333	0,2337	0,2667	0,6333
	15	7	8	0,0333	0,1667	0,2333	0,2523	0,3000	0,7667
	21	8	7	0,0333	0,1667	0,2667	0,2383	0,3000	0,5000
	4	8	7	0,0333	0,2000	0,2500	0,2473	0,3000	0,5000
	20	7	7	0,0333	0,1667	0,2333	0,2283	0,2667	0,4667
	36	7	7	0,0333	0,1667	0,2167	0,2300	0,3000	0,5667
	52	9	9	0,0667	0,2667	0,3000	0,2943	0,3333	0,5667
50	2	12	12	0,0200	0,2200	0,2400	0,2498	0,3000	0,4600
	3	12	12	0,1200	0,2200	0,2400	0,2472	0,2800	0,4800
	4	13	13	0,1600	0,2200	0,2500	0,2562	0,2800	0,6400
	5	13	12	0,0200	0,2200	0,2600	0,2468	0,2800	0,3800
	3	14	13	0,0200	0,1800	0,2800	0,2526	0,3200	0,5400
	9	11	12	0,0200	0,1750	0,2200	0,2382	0,2850	0,5400
	15	11	13	0,0200	0,1800	0,2200	0,2602	0,3200	0,6800
	21	13	12	0,0200	0,1800	0,2600	0,2490	0,3200	0,6000
	4	12	13	0,0600	0,2200	0,2400	0,2584	0,3000	0,4800
	20	11	11	0,0200	0,1800	0,2200	0,2196	0,2650	0,4200
	36	11	12	0,0600	0,1800	0,2200	0,2360	0,2800	0,4000
	52	15	14	0,1400	0,2400	0,3000	0,2860	0,3200	0,4200

Fonte: Elaborada pelo autor.

Tabela 23 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 4 componentes.

N	θ_{real}	$\hat{\theta}_k$						$Var[\hat{\theta}_k]$	
		Min	Q_1	Q_2	Média	Q_3	Max	Q_2	Média
10	2	0,0000	1,3120	3,0000	2,7930	4,0000	9,0000	1,1720	1,4860
	3	0,0000	2,5000	3,0000	3,2440	4,0000	10,0000	1,4780	1,8240
	4	0,0000	2,5000	3,2250	3,1970	4,0000	7,0000	1,4140	1,5550
	5	0,0000	2,9170	4,0000	4,0210	4,6670	12,0000	1,4950	2,2260
	3	0,0000	1,5750	4,8330	6,6260	9,0830	26,0000	1,7060	3,5130
	9	0,0303	6,4580	9,0000	9,9040	12,3800	26,5000	4,0000	5,1120
	15	0,0303	10,0000	13,5000	13,3400	17,0000	29,0000	5,3890	6,5450
	21	1,3330	15,4200	18,8800	17,9900	22,0500	32,0000	6,7780	8,2770
	4	0,0303	3,1880	6,2920	14,9700	21,0800	64,0000	2,5120	8,5510
	20	0,0303	16,1900	21,0000	22,0100	25,0000	61,0000	9,5420	11,9700
	36	1,3330	22,0000	34,0000	32,2100	39,1700	71,0000	13,3600	17,5000
	52	3,0000	37,5000	48,6700	44,3000	57,0000	66,0000	18,7500	21,4900
30	2	0,0397	2,1430	2,7390	2,8450	3,7570	6,6360	0,4408	0,4327
	3	0,0440	2,8800	3,4290	3,6940	3,8570	12,0000	0,4896	0,7804
	4	0,0375	2,7750	3,7320	3,4230	4,0280	8,7500	0,5180	0,5266
	5	0,0375	3,2860	4,1670	4,3640	4,8620	13,0000	0,5555	0,9058
	3	0,0303	2,3650	3,4000	6,8630	10,6700	27,0000	0,4607	1,0040
	9	1,0000	6,6560	8,7500	9,7260	10,8100	29,2500	1,3560	1,8000
	15	0,0303	10,9500	13,5500	12,7100	15,3500	29,0000	1,7840	2,1140
	21	0,0303	12,3100	20,6700	18,1800	22,5300	32,0000	2,4630	3,5670
	4	0,0303	3,5860	4,1270	6,6290	4,7570	61,2500	0,6070	1,1390
	20	4,0000	17,2800	19,1200	19,5800	21,5800	46,5000	2,7330	3,2690
	36	4,1670	30,4400	34,1100	33,1800	37,0000	50,4600	4,7530	6,4290
	52	4,0000	48,7400	53,2200	50,7200	55,9000	67,5000	5,7740	6,6950
50	2	0,0407	1,1200	2,9280	2,6800	3,7380	13,0000	0,2452	0,3581
	3	0,0407	2,7270	3,1020	3,1850	3,8140	7,6670	0,2881	0,2925
	4	0,0441	3,2620	3,8090	3,7670	4,0730	7,8750	0,3245	0,3309
	5	0,0407	3,5340	4,1540	4,4220	5,9500	12,0000	0,3536	0,5074
	3	0,0303	2,3560	2,9430	5,0880	4,3870	28,5000	0,2529	0,5598
	9	2,2630	6,9770	8,4720	9,8250	10,7300	31,0000	0,8039	1,2240
	15	0,0435	11,5900	13,5900	13,1900	15,0600	30,0000	1,1530	1,3980
	21	0,0303	18,6200	21,3000	19,8600	23,3200	35,0000	1,4890	3,2280
	4	0,0375	3,5650	4,0000	4,1360	4,3810	21,2900	0,3279	0,3485
	20	3,5000	18,0000	19,4100	18,9500	20,5500	26,0600	1,8040	2,0270
	36	19,0800	32,3200	34,7000	34,5300	37,2000	55,2000	2,8910	3,2930
	52	37,5000	50,4400	53,2700	52,8500	55,4400	60,5000	3,5440	3,9000

Fonte: Elaborada pelo autor.

Tabela 24 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 4 componentes.

N	θ_{real}	\hat{n}_k		\hat{p}_k						
		Q_2	Média	Min	Q_1	Q_2	Média	Q_3	Max	
100	2	25	25	0,1100	0,2100	0,2500	0,2479	0,2800	0,4100	
	3	24	25	0,0100	0,2100	0,2400	0,2502	0,2725	0,6300	
	4	24	24	0,0100	0,2075	0,2400	0,2405	0,2700	0,5300	
	5	25	26	0,1400	0,2300	0,2500	0,2614	0,2800	0,6300	
	3	27	25	0,0100	0,2100	0,2700	0,2513	0,3000	0,5700	
	9	23	24	0,0100	0,1900	0,2300	0,2373	0,2900	0,4300	
	15	23	26	0,0300	0,1875	0,2250	0,2605	0,3125	0,5500	
	21	28	25	0,0100	0,1900	0,2800	0,2509	0,3200	0,5800	
	4	24	25	0,1500	0,2200	0,2400	0,2506	0,2900	0,3800	
	20	25	25	0,1500	0,2175	0,2500	0,2488	0,2800	0,3800	
	36	23	22	0,0700	0,2000	0,2300	0,2248	0,2600	0,4300	
	52	28	28	0,1400	0,2400	0,2800	0,2758	0,3100	0,4300	
	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	1.000	3	290	284	0,2190	0,2660	0,2900	0,2839	0,3012	0,3290
		9	209	196	0,0260	0,1880	0,2090	0,1960	0,2310	0,2900
		15	203	221	0,0440	0,1970	0,2030	0,2207	0,2140	0,4860
21		301	299	0,0040	0,2910	0,3010	0,2993	0,3113	0,4360	
4		249	249	0,2190	0,2390	0,2490	0,2487	0,2550	0,2900	
20		247	247	0,1130	0,2290	0,2465	0,2469	0,2670	0,3170	
36		238	235	0,0500	0,2105	0,2380	0,2349	0,2738	0,4390	
52		268	269	0,0050	0,2300	0,2675	0,2695	0,3105	0,4560	
I		(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
10.000		3	2.580	2.372	0,0356	0,2112	0,2580	0,2372	0,2965	0,3728
		9	2.510	2.520	0,0873	0,2025	0,2510	0,2520	0,3008	0,4206
		15	3.598	3.471	0,1229	0,2004	0,3598	0,3471	0,4997	0,6084
		21	2.182	1.637	0,0001	0,0006	0,2182	0,1637	0,2935	0,4588
		4	2.502	2.506	0,2413	0,2476	0,2502	0,2506	0,2538	0,2596
		20	2.456	2.492	0,0930	0,2340	0,2456	0,2492	0,2583	0,3267
		36	2.474	2.578	0,0444	0,2415	0,2474	0,2578	0,2598	0,4466
	52	2.599	2.424	0,0006	0,2438	0,2599	0,2424	0,2752	0,4258	

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

Tabela 25 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 4 componentes.

N	θ_{real}	$\hat{\theta}_k$						$Var[\hat{\theta}_k]$	
		Min	Q_1	Q_2	Média	Q_3	Max	Q_2	Média
100	2	0,0417	0,9237	2,9550	2,3530	3,4790	6,9520	0,1328	0,1056
	3	1,0550	2,8550	3,2770	3,6450	3,7820	13,0000	0,1487	0,3401
	4	0,0417	3,2190	3,5280	3,4610	4,0000	13,0000	0,1523	0,2843
	5	0,0425	3,2690	3,8910	4,1620	5,6950	6,6560	0,1680	0,1676
	3	0,0397	2,5900	3,1040	5,5670	4,1180	34,0000	0,1306	0,9038
	9	1,1760	7,6480	8,7660	10,0800	9,7850	35,0000	0,3743	0,8461
	15	2,0870	12,5700	13,8900	13,8800	16,1900	30,3300	0,6082	0,6771
	21	1,0830	19,2500	21,2700	19,5500	22,9600	36,0000	0,6857	1,9670
	4	3,0910	3,7230	4,0290	4,0300	4,3200	5,1200	0,1700	0,1764
	20	16,4100	18,6500	19,3800	19,5000	20,3500	25,8600	0,7922	0,8058
	36	28,9300	33,0400	34,5400	34,5000	35,7400	40,0500	1,5020	1,6510
	52	46,3700	52,0700	53,3900	53,5000	55,3600	59,4300	1,9300	2,0110
1.000	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	2,5600	2,8210	3,2440	3,1170	3,3260	3,5410	0,0123	0,0123
	9	6,0000	8,5340	8,8540	8,6660	9,0970	9,7230	0,0433	0,0525
	15	11,4900	13,1200	13,9300	13,6700	14,0400	18,1500	0,0673	0,0668
	21	19,7400	21,0800	21,2500	21,3500	21,4300	33,2500	0,0710	0,1530
	4	3,6190	3,8780	3,9600	3,9610	4,0790	4,3210	0,0170	0,0170
	20	16,4800	19,2200	19,6600	19,6100	20,0600	22,0300	0,0798	0,0806
	36	27,4800	33,1800	34,6900	34,8200	35,8900	46,9000	0,1457	0,1759
	52	46,5500	50,6000	52,0300	52,3900	53,4100	71,8000	0,1943	0,4961
	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	0,0434	2,3360	3,1960	3,3960	3,2670	12,0300	0,0012	0,0015
	9	3,3240	7,4290	9,0430	8,5540	9,6240	12,1000	0,0037	0,0039
15	9,5170	12,8300	15,9700	15,6200	18,3000	21,4500	0,0047	0,0054	
21	12,1500	20,7800	22,7900	27,6300	36,7800	43,0000	0,0104	6,2880	
10.000	4	3,8420	3,9550	3,9860	4,6970	4,0240	27,6600	0,0017	0,0020
	20	15,4700	19,2100	19,5100	19,9700	20,0000	27,7600	0,0080	0,0081
	36	25,1600	34,0800	34,9400	35,5700	35,8200	46,5200	0,0140	0,0164
	52	27,9500	51,7000	52,2300	53,0000	52,8100	75,0000	0,0200	0,2909

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

Tabela 26 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 5 componentes.

N	θ_{real}	\hat{n}_k		\hat{p}_k					
		Q_2	Média	Min	Q_1	Q_2	Média	Q_3	Max
10	2	1	2	0,1000	0,1000	0,1000	0,1661	0,2000	0,4000
	3	2	2	0,1000	0,2000	0,2000	0,2115	0,3000	0,4000
	4	2	2	0,1000	0,2000	0,2000	0,2230	0,3000	0,4000
	5	2	2	0,1000	0,1000	0,2000	0,2186	0,3000	0,5000
	6	2	2	0,1000	0,1000	0,2000	0,1800	0,2000	0,4000
	3	2	2	0,1000	0,1000	0,2000	0,2088	0,3000	0,5000
	9	2	2	0,1000	0,1000	0,2000	0,1945	0,2000	0,5000
	15	2	2	0,1000	0,1000	0,2000	0,1989	0,3000	0,4000
	21	2	2	0,1000	0,1000	0,2000	0,2055	0,3000	0,5000
	27	2	2	0,1000	0,1000	0,2000	0,1923	0,2500	0,4000
	4	2	2	0,1000	0,1000	0,2000	0,2040	0,3000	0,5000
	20	2	2	0,1000	0,1000	0,2000	0,2030	0,3000	0,5000
	36	2	2	0,1000	0,1000	0,2000	0,1880	0,2000	0,5000
	52	2	2	0,1000	0,1000	0,2000	0,1840	0,2000	0,5000
	68	2	2	0,1000	0,1000	0,2000	0,2210	0,3000	0,4000
30	2	6	6	0,0667	0,1667	0,2000	0,1992	0,2333	0,3667
	3	6	6	0,0667	0,1667	0,2000	0,2024	0,2333	0,3333
	4	6	6	0,1000	0,1667	0,2000	0,2048	0,2333	0,3333
	5	6	6	0,1000	0,1667	0,2000	0,1984	0,2333	0,3667
	6	6	6	0,0333	0,1667	0,2000	0,1952	0,2333	0,3667
	3	5	6	0,0333	0,1333	0,1667	0,1937	0,2333	0,4000
	9	6	6	0,0333	0,1333	0,2000	0,2010	0,2333	0,4667
	15	6	6	0,0333	0,1333	0,1833	0,1987	0,2333	0,4667
	21	7	6	0,0667	0,1667	0,2000	0,2260	0,2667	0,5667
	27	6	6	0,0333	0,1000	0,1667	0,1807	0,2417	0,4333
	4	6	6	0,0667	0,1667	0,2000	0,2140	0,2667	0,4667
	20	5	5	0,0333	0,1000	0,1667	0,1673	0,2333	0,3667
	36	4	5	0,0333	0,1000	0,1333	0,1553	0,2000	0,4000
	52	5	6	0,0333	0,1333	0,1667	0,1937	0,2333	0,5000
	68	8	8	0,0333	0,2333	0,2667	0,2697	0,3333	0,4000
50	2	10	10	0,0800	0,1600	0,2000	0,1900	0,2200	0,3000
	3	10	10	0,1400	0,1750	0,2000	0,1941	0,2200	0,3000
	4	11	11	0,1200	0,1800	0,2200	0,2141	0,2400	0,3000
	5	10	10	0,1000	0,1800	0,2000	0,2039	0,2200	0,3600
	6	10	10	0,0400	0,1800	0,2000	0,1980	0,2200	0,2800
	3	10	10	0,0200	0,1400	0,2000	0,1980	0,2600	0,4800
	9	9	9	0,0600	0,1200	0,1700	0,1838	0,2250	0,4600
	15	10	10	0,0600	0,1600	0,2000	0,2050	0,2250	0,5000
	21	9	11	0,0600	0,1600	0,1800	0,2102	0,2400	0,4400
	27	10	10	0,0200	0,1400	0,2000	0,2030	0,2600	0,4400
	4	10	10	0,0800	0,1800	0,2000	0,2088	0,2400	0,3600
	20	8	9	0,0400	0,1400	0,1600	0,1752	0,2200	0,3600
	36	8	8	0,0200	0,1200	0,1600	0,1606	0,2000	0,3200
	52	8	8	0,0600	0,1200	0,1600	0,1856	0,2200	0,5200
	68	15	13	0,0200	0,2350	0,2900	0,2698	0,3400	0,4200

Fonte: Elaborada pelo autor.

Tabela 27 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 5 componentes.

N	θ_{real}	$\hat{\theta}_k$						$Var[\hat{\theta}_k]$	
		Min	Q_1	Q_2	Média	Q_3	Max	Q_2	Média
10	2	0,0303	1,0000	1,0000	1,5450	2,5000	3,7500	1,4780	1,2150
	3	0,0435	2,5000	3,0000	3,0170	3,5000	5,0000	1,6910	1,8870
	4	2,5000	3,0000	4,0000	3,8700	4,5000	5,5000	1,9050	2,1980
	5	3,0000	4,0000	5,0000	4,9310	5,8330	9,0000	2,3430	2,9680
	6	3,6670	5,5000	6,5830	6,7290	7,5000	11,0000	3,6410	4,7720
	3	0,0303	2,0000	3,5000	4,5650	6,7500	15,0000	1,9050	2,6640
	9	1,0000	6,1670	10,0000	10,0000	13,0000	20,0000	5,1410	6,0680
	15	3,0000	10,5000	14,0000	14,4600	17,6700	27,0000	7,5000	8,6240
	21	7,6670	16,0000	19,3300	20,0200	23,8800	34,0000	10,2500	12,9900
	27	14,0000	23,7500	27,0000	27,2100	31,2500	46,0000	15,5000	18,1500
	4	0,0303	3,0000	4,5000	7,6120	7,2500	39,3300	2,4570	4,0470
	20	3,0000	15,8800	20,0000	21,5300	26,7900	47,5000	10,7100	12,3600
	36	13,0000	27,0000	35,0000	34,9300	41,6200	64,0000	20,6200	23,2200
	52	22,5000	41,0000	47,5000	47,9400	54,2500	75,0000	28,1200	32,0400
	68	29,5000	59,2500	66,5800	64,8100	72,0000	89,0000	31,2500	37,2700
30	2	0,0435	2,1250	2,7780	2,7120	3,3440	4,5710	0,5363	0,5541
	3	2,0000	3,2000	3,6000	3,5290	4,0000	4,6670	0,6505	0,6869
	4	2,5000	3,6500	4,0000	4,0300	4,5000	5,2000	0,7072	0,7435
	5	3,2500	4,1670	4,5000	4,5470	5,0000	6,2860	0,7677	0,8636
	6	3,6670	4,8510	5,2500	5,2890	5,6670	8,6670	0,9081	1,0450
	3	0,0303	1,7880	2,8250	4,2560	4,3080	17,5700	0,5242	0,8277
	9	3,0000	6,5710	8,5360	9,0600	10,8500	19,1700	1,5430	1,9460
	15	8,8000	11,7700	13,9200	13,9200	15,8800	21,0000	2,5820	2,7050
	21	10,6700	17,0400	19,2500	19,5800	22,0000	30,4000	3,0680	3,2990
	27	17,0000	23,9500	27,7100	28,0700	32,5000	42,0000	4,7600	9,2280
	4	0,6667	3,4860	4,0000	4,7000	5,0360	16,3300	0,6988	0,8267
	20	5,0000	16,4800	19,0000	18,9300	21,1800	34,0000	3,7400	4,8230
	36	17,8000	26,5100	32,0000	31,4400	36,0000	44,7500	6,9060	8,3150
	52	30,5000	40,3300	44,4200	44,9300	49,3100	58,7500	7,9200	9,1770
	68	54,0900	62,3200	65,8000	66,0400	69,7100	80,0000	7,9810	9,6080
50	2	0,0397	2,6670	3,0000	2,9390	3,5670	4,5000	0,3569	0,3537
	3	1,6360	3,4290	3,7210	3,5760	4,0000	4,7000	0,4156	0,4120
	4	2,6670	3,9000	4,0830	4,0480	4,2980	4,8890	0,3930	0,4168
	5	3,6250	4,2220	4,4490	4,5380	4,7800	6,0620	0,4598	0,4862
	6	3,8000	4,6120	5,0000	5,2980	5,4460	14,0000	0,5091	0,6351
	3	0,0303	2,0000	2,9670	4,5510	4,7190	15,1200	0,3215	0,5084
	9	3,2500	7,2810	8,5920	8,9560	11,0500	15,7100	1,0470	1,1410
	15	8,3750	12,2700	14,2300	13,9100	15,6800	18,6200	1,4380	1,5520
	21	13,3000	16,8200	18,6500	19,3300	21,2800	29,2700	1,8950	2,0440
	27	15,1000	23,8000	27,4100	27,1900	30,5800	44,0000	2,2360	5,0400
	4	2,1250	3,5420	4,0000	4,3260	4,4040	29,2200	0,4128	0,4657
	20	5,6670	17,4200	19,4200	19,0400	20,9400	31,7100	2,2630	2,4020
	36	18,6200	28,6200	32,5000	31,8800	36,0000	43,1900	3,9550	4,7720
	52	32,8600	41,6200	45,6700	45,6800	49,0800	58,8800	5,3440	5,9450
	68	50,7500	62,4500	66,6500	66,0400	68,9500	82,6700	4,5360	6,5520

Fonte: Elaborada pelo autor.

Tabela 28 – Resultados do algoritmo EM para as estimativas \hat{p}_k e \hat{n}_k para o modelo de mistura de Poisson truncadas no zero com 5 componentes.

N	θ_{real}	\hat{n}_k		\hat{p}_k						
		Q_2	Média	Min	Q_1	Q_2	Média	Q_3	Max	
100	2	19	19	0,0900	0,1625	0,1900	0,1907	0,2100	0,3100	
	3	19	20	0,1400	0,1800	0,1900	0,1970	0,2100	0,2900	
	4	20	20	0,1300	0,1800	0,2000	0,2040	0,2300	0,3100	
	5	20	21	0,1400	0,1900	0,2000	0,2088	0,2300	0,4600	
	6	20	20	0,0100	0,1800	0,2000	0,1994	0,2200	0,3200	
	3	18	18	0,0200	0,1300	0,1800	0,1794	0,2300	0,3700	
	9	18	19	0,0600	0,1500	0,1800	0,1872	0,2225	0,3500	
	15	18	19	0,0700	0,1400	0,1800	0,1897	0,2200	0,4000	
	21	19	23	0,0900	0,1500	0,1900	0,2344	0,2800	0,5300	
	27	21	21	0,0100	0,1450	0,2100	0,2093	0,3100	0,4100	
	4	20	20	0,1200	0,1800	0,2000	0,2019	0,2300	0,3100	
	20	18	18	0,0800	0,1500	0,1800	0,1839	0,2100	0,3500	
	36	15	15	0,0400	0,1175	0,1500	0,1519	0,1900	0,3400	
	52	18	20	0,0500	0,1300	0,1800	0,1975	0,2525	0,4000	
	68	29	26	0,0200	0,2175	0,2900	0,2648	0,3325	0,4200	
	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	1.000	3	196	155	0,0220	0,0790	0,1960	0,1551	0,2250	0,2920
		9	175	183	0,0550	0,1415	0,1750	0,1834	0,2238	0,3530
15		166	182	0,0630	0,1417	0,1655	0,1821	0,2095	0,3810	
21		172	257	0,1290	0,1570	0,1715	0,2572	0,3908	0,5500	
27		287	222	0,0010	0,0270	0,2870	0,2222	0,3273	0,4380	
4		201	202	0,1700	0,1927	0,2010	0,2017	0,2120	0,2370	
20		186	185	0,1080	0,1740	0,1855	0,1852	0,1970	0,2530	
36		143	149	0,0640	0,1300	0,1430	0,1488	0,1602	0,2980	
52		218	217	0,1180	0,1898	0,2180	0,2169	0,2490	0,3440	
68		266	247	0,0010	0,2270	0,2655	0,2474	0,2860	0,3480	
I		(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
3		1.991	1.991	0,1909	0,1962	0,1991	0,1991	0,2016	0,2123	
9		1.845	1.848	0,1626	0,1821	0,1845	0,1848	0,1876	0,2168	
15		1.420	1.425	0,0872	0,1376	0,1420	0,1425	0,1461	0,2579	
21		2.128	2.147	0,1019	0,2066	0,2128	0,2147	0,2271	0,3228	
27		2.617	2.589	0,0001	0,2492	0,2617	0,2589	0,2709	0,2963	
4		2.008	1.860	0,0294	0,1965	0,2008	0,1860	0,2109	0,3296	
20		1.976	1.840	0,0679	0,1624	0,1976	0,1840	0,2023	0,2638	
36	1.976	1.934	0,0704	0,1618	0,1976	0,1934	0,2019	0,3413		
52	1.990	2.401	0,1308	0,1627	0,1990	0,2401	0,2057	0,5479		
68	2.017	1.965	0,0001	0,1962	0,2017	0,1965	0,3268	0,4360		

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

Tabela 29 – Resultados do algoritmo EM para as estimativas $\hat{\theta}_k$ para o modelo de mistura de Poisson truncadas no zero com 5 componentes.

N	θ_{real}	$\hat{\theta}_k$						Var[$\hat{\theta}_k$]		
		Min	Q_1	Q_2	Média	Q_3	Max	Q_2	Média	
100	2	0,0419	2,9250	3,2190	2,8390	3,4520	4,1880	0,1786	0,1634	
	3	1,5830	3,3470	3,6220	3,4820	3,9460	4,6840	0,2021	0,1973	
	4	2,8330	3,8700	4,1280	4,0250	4,3300	4,8100	0,2121	0,2144	
	5	3,5220	4,2150	4,4750	4,5230	4,7020	6,8000	0,2259	0,2310	
	6	3,7600	4,5220	4,8220	5,4210	5,1340	17,0000	0,2402	0,7629	
	3	0,0375	1,8920	2,8890	3,8780	3,8160	14,5900	0,1710	0,2292	
	9	3,0620	6,6360	8,2980	8,4860	9,9350	14,8400	0,4828	0,5025	
	15	7,1110	11,9500	13,4000	13,2500	14,510	18,9300	0,7678	0,7690	
	21	13,3100	16,5700	18,9100	19,1700	21,6100	26,2100	0,9227	0,9464	
	27	14,0500	25,1300	27,5900	27,7700	29,7700	43,0000	1,1000	3,9090	
	4	2,8120	3,5760	4,0000	4,0140	4,3180	10,7300	0,2067	0,2173	
	20	14,1100	18,0900	18,9100	19,0800	19,9200	30,4200	1,0320	1,1020	
	36	21,9200	30,1000	32,5400	32,4700	35,1500	39,9400	2,1580	2,4680	
	52	36,8700	42,5100	45,0900	46,1800	48,0800	60,2100	2,4580	2,7650	
	68	59,1200	63,2600	65,7700	66,8400	69,7500	83,5000	2,2380	3,5900	
	1.000	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
		3	0,0426	0,9858	2,6270	2,0330	2,9990	4,3210	0,0154	0,0136
		9	1,5940	4,9760	7,6070	6,8940	8,5820	10,4900	0,0452	0,0428
15		6,3170	11,530	12,5100	12,3600	13,2600	16,7900	0,0768	0,0759	
21		13,3800	16,9900	18,0900	19,2300	21,6500	27,5100	0,1007	0,0901	
27		23,5900	25,8400	26,8300	30,6900	36,1500	48,0000	0,0935	6,0320	
4		3,5360	3,8730	3,9850	3,9760	4,0920	4,4770	0,0210	0,0210	
20		17,0200	18,8500	19,2000	19,2600	19,7300	20,7500	0,1040	0,1051	
36		24,4700	31,0300	32,1800	32,3400	33,2200	40,3100	0,2236	0,2286	
52		41,0600	44,3300	46,1700	46,9000	47,5800	65,1200	0,2177	0,2255	
68		62,5200	64,9600	66,3400	67,6600	67,7700	94,0000	0,2489	2,3870	
10.000		I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
		3	0,0434	2,8000	3,2560	7,4480	14,7100	15,0200	0,0018	0,0039
		9	2,2510	8,0500	8,8090	10,4900	14,8900	15,1400	0,0066	0,0059
		15	7,8600	12,5100	13,5200	13,7300	15,0500	16,5400	0,0077	0,0076
		21	14,5100	15,1600	16,9900	18,0500	20,4200	26,3100	0,0077	0,0085
		27	14,8700	15,3100	25,7300	26,0800	28,4100	48,0000	0,0078	2,8500
		4	3,8660	3,9380	3,9710	3,9680	3,9960	4,1240	0,0021	0,0021
	20	18,5300	19,0900	19,1600	19,1600	19,2300	20,3800	0,0104	0,0104	
	36	29,1000	31,9400	32,0200	32,0800	32,1600	39,2300	0,0226	0,0228	
	52	42,8300	45,5100	45,9900	46,2000	46,5400	63,3800	0,0216	0,0218	
	68	64,6600	65,8400	66,3000	66,6000	66,9000	93,0000	0,9552	0,0269	

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

RESULTADOS DO ALGORITMO EMPMTZ

Tabela 30 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson para modelo truncado no zero e o tempo de processamento em segundos para $K = 2$.

K	N	λ_{real}	quantidade de iterações					tempo	percentual de amostras	
			Min	Q_1	Q_2	Média	Q_3			Max
2	10	(2, 3)	6	9	28	19.980	26.800	188.800	4.410,57	41
		(3, 9)	3	5	6	7	7	38	0,50	100
		(4, 20)	4	4	5	5	6	16	0,42	100
	30	(2, 3)	7	11	14	27.660	18.770	172.800	13.643,67	44
		(3, 9)	3	4	5	5	6	9	0,54	100
		(4, 20)	2	3	4	4	4	6	0,47	100
	50	(2, 3)	5	9	11	18.890	38	161.300	17.232,75	39
		(3, 9)	3	4	5	5	5	7	0,77	100
		(4, 20)	2	3	4	4	4	6	0,66	100
	100	(2, 3)	5	9	11	20.450	15	165.800	39.297,14	37
		(3, 9)	3	4	4	4	5	7	1,69	100
		(4, 20)	2	3	4	4	4	5	1,40	100
	1.000	(2, 3)	7	7	7	22.640	33.960	67.910	532.917,78	3
		(3, 9)	2	3	4	4	4	5	11,87	100
		(4, 20)	2	3	4	4	4	4	11,54	100
	10.000	(2, 3)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9)	3	4	4	4	4	4	128,61	100
		(4, 20)	3	4	4	4	4	4	134,88	100

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

Tabela 31 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson para modelo truncado no zero e o tempo de processamento em segundos para $K = 3$ e 4.

K	N	λ_{real}	quantidade de iterações						tempo	percentual de amostras
			Min	Q_1	Q_2	Média	Q_3	Max		
3	10	(2, 3, 4)	7	38	1.009	1.329	1.340	10.260	3.736,99	27
		(3, 9, 15)	3	6	9	16.200	15	1.439.000	1.039,50	100
		(4, 20, 36)	3	4	5	20	6	1.386	1,52	100
	30	(2, 3, 4)	17	72	177	10.590	3.450	105.100	19.237,41	23
		(3, 9, 15)	4	7	7	22	10	1.097	3,61	100
		(4, 20, 36)	4	5	5	5	6	8	0,80	100
	50	(2, 3, 4)	15	46	2.945	39.770	68.280	180.600		16
		(3, 9, 15)	5	7	7	11	12	68	2,48	100
		(4, 20, 36)	3	5	5	5	5	10	1,30	100
	100	(2, 3, 4)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15)	6	7	8	15	10	367	6,03	100
		(4, 20, 36)	4	5	5	6	6	9	2,34	100
	1.000	(2, 3, 4)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15)	6	8	11	15	19	44	54,91	100
		(4, 20, 36)	5	5	6	6	6	9	22,71	100
	10.000	(2, 3, 4)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15)	10	12	14	18	19	61	688,56	100
		(4, 20, 36)	5	6	7	7	7	8	261,10	100
4	10	(2, 3, 4, 5)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15, 21)	5	10	19	2.436	87	166.200	391,72	99
		(4, 20, 36, 52)	5	7	9	35	13	1.309	2,77	100
	30	(2, 3, 4, 5)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15, 21)	5	13	34	2.144	79	199.000	404,20	100
		(4, 20, 36, 52)	5	7	8	10	12	66	2,10	100
	50	(2, 3, 4, 5)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15, 21)	6	16	41	83	82	1.495	25,53	100
		(4, 20, 36, 52)	5	7	11	13	16	52	4,37	100
	100	(2, 3, 4, 5)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15, 21)	7	22	63	117	126	986	71,15	100
		(4, 20, 36, 52)	6	7	10	13	17	56	8,49	100
	1.000	(2, 3, 4, 5)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15, 21)	7	28	65	103	131	625	591,72	100
		(4, 20, 36, 52)	6	12	21	27	33	126	155,39	100
	10.000	(2, 3, 4, 5)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15, 21)	7	28	94	131	200	834	8.528,53	100
		(4, 20, 36, 52)	7	14	22	31	44	123	1.671,80	100

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

Tabela 32 – Medidas descritivas da quantidade de iterações do algoritmo EM aplicado à estimação dos parâmetros para mistura de Poisson para modelo truncado no zero e o tempo de processamento em segundos para $K = 5$.

K	N	λ_{real}	quantidade de iterações						tempo	percentual de amostras
			Min	Q_1	Q_2	Média	Q_3	Max		
5	10	(2, 3, 4, 5, 6)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15, 21, 27)	5	25	102	1.111	372	42.470	2.734,34	86
		(4, 20, 36, 52, 68)	2	5	8	63	20	2.214	191,01	99
	30	(2, 3, 4, 5, 6)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15, 21, 27)	3	89	351	6.173	1.050	166.800	1.998,14	99
		(4, 20, 36, 52, 68)	3	4	7	27	24	267	6,64	100
	50	(2, 3, 4, 5, 6)	29.650	67.510	105.400	105.400	143.200	181.100	61.306,13	2
		(3, 9, 15, 21, 27)	4	218	834	3.089	1.876	47.710	1.342,10	100
		(4, 20, 36, 52, 68)	3	4	6	113	23	7.449	44,44	100
	100	(2, 3, 4, 5, 6)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15, 21, 27)	3	356	2.138	6.537	5.307	116.700	4.634,85	100
		(4, 20, 36, 52, 68)	3	4	5	96	9	2.985	72,72	100
	1.000	(2, 3, 4, 5, 6)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15, 21, 27)	4	5	3.830	7.482	9.879	88.460	46.855,75	100
		(4, 20, 36, 52, 68)	3	6	7	168	11	5.542	1.233,00	100
	10.000	(2, 3, 4, 5, 6)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	0
		(3, 9, 15, 21, 27)	4	156	2.076	6.318	6.760	88.940	386.164,98	100
		(4, 20, 36, 52, 68)	5	7	8	46	8	2.339	2.726,72	100

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

Tabela 33 – Resultados do algoritmo EM para as estimativas $\hat{\lambda}_k$, \hat{n}_k e \hat{q}_k para mistura de Poisson para modelo truncado no zero com 2 componentes.

N	λ_{real}	$\hat{\lambda}_k$							\hat{n}_k		\hat{q}_k	
		Min	Q_1	Q_2	Média	Q_3	Max	Var	Q_2	Média	Q_2	Média
10	2	1,4740	1,5060	1,9490	2,1250	2,5660	3,7580	0,3753	5	6	0,5412	0,6001
	3	2,7040	3,3260	3,8350	4,3940	5,0330	9,0010	2,1728	5	4	0,3020	0,5190
	3	1,6830	2,8480	3,1460	3,3190	3,7010	6,1250	0,4287	6	6	0,6061	0,5871
	9	6,3450	9,0010	10,0000	9,9970	11,0000	16,0000	1,3399	4	4	0,4004	0,4411
	4	1,8590	3,3720	3,8980	4,1040	4,4530	19,2000	0,2210	5	5	0,5072	0,5135
	20	11,0000	18,7100	19,7100	19,8800	21,2100	26,5000	1,5032	5	5	0,5000	0,5000
30	2	1,4740	1,8180	1,9040	2,0790	2,5660	2,9940	0,2198	13	14	0,5167	0,5301
	3	2,6480	3,4580	3,8360	4,0990	4,0940	9,0010	2,2360	17	16	0,5765	0,5581
	3	2,1370	2,7020	3,1190	3,1310	3,5430	4,4800	0,3073	15	15	0,5329	0,5398
	9	6,7450	8,7620	9,3020	9,4760	10,3500	13,3300	1,3911	15	15	0,5000	0,4871
	4	3,2740	3,8150	4,1580	4,1320	4,4210	5,2770	0,1985	15	15	0,5080	0,5014
	20	17,4100	19,0000	19,9700	20,0200	20,9500	22,4600	1,4794	15	15	0,5000	0,5077
50	2	1,7120	1,9110	1,9670	2,1790	2,5960	3,1660	0,1752	22	25	0,5246	0,5792
	3	2,6220	3,7030	3,9740	4,5150	4,1560	12,0000	4,8001	28	25	0,5692	0,5072
	3	2,3950	2,9220	3,1010	3,1470	3,3750	4,5240	0,1441	25	26	0,5302	0,5434
	9	7,5240	8,7480	9,2040	9,3860	10,0000	12,5700	0,8863	25	24	0,5000	0,4817
	4	3,2650	3,7940	4,0880	4,1270	4,4120	5,2350	0,1564	25	25	0,5089	0,5109
	20	18,3900	19,5600	20,1700	20,0800	20,6000	21,6900	0,5607	25	25	0,5000	0,4980
100	2	1,7570	1,8840	1,9610	2,2500	2,8080	3,0310	0,2443	46	52	0,5345	0,5951
	3	2,6290	3,7170	3,9590	4,5400	4,1220	10,0000	4,5966	54	48	0,5583	0,4888
	3	2,7410	3,0400	3,1900	3,2550	3,4250	4,0350	0,0861	53	54	0,5456	0,5604
	9	8,0920	9,1880	9,5060	9,5720	9,9170	11,5700	0,3826	48	46	0,4750	0,4623
	4	3,4890	3,9920	4,1680	4,1740	4,3590	4,7560	0,0698	50	50	0,5050	0,5054
	20	18,2600	19,5900	20,1900	20,1100	20,6100	21,7100	0,4743	50	50	0,4950	0,4946
1.000	2	2,8840	2,8920	2,9010	2,9080	2,9190	2,9380	0,0007	511	511	0,5111	0,5100
	3	2,9430	7,4720	12,0000	9,3140	12,5000	13,0000	30,6933	487	486	0,4873	0,4861
	3	2,9910	3,0920	3,1240	3,1330	3,1540	3,4900	0,0056	512	513	0,5372	0,5373
	9	9,0800	9,2650	9,3310	9,3620	9,4320	9,9820	0,0244	489	487	0,4885	0,4868
	4	3,9000	4,0710	4,1680	4,1560	4,2250	4,4520	0,0125	502	503	0,5105	0,5109
	20	19,4100	19,9100	20,1000	20,0600	20,2200	20,4600	0,0580	498	497	0,4980	0,4973
10.000	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	3,0960	3,1220	3,1310	3,1320	3,1420	3,1710	0,0003	5.138	5.138	0,5381	0,5379
	9	9,2660	9,3210	9,3480	9,3460	9,3710	9,4340	0,0015	4.862	4.862	0,4862	0,4862
	4	4,0660	4,1570	4,1750	4,1740	4,1930	4,2500	0,0009	5.020	5.030	0,5100	0,5110
	20	19,8500	20,0500	20,0900	20,0900	20,1200	20,2200	0,0045	4.980	4.970	0,4980	0,4970

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

Tabela 34 – Resultados do algoritmo EM para as estimativas $\hat{\lambda}_k$, \hat{n}_k e \hat{q}_k para mistura de Poisson para modelo truncado no zero com 3 componentes.

N	λ_{real}	$\hat{\lambda}_k$							\hat{n}_k		\hat{q}_k	
		Min	Q_1	Q_2	Média	Q_3	Max	Var	Q_2	Média	Q_2	Média
10	2	1,4740	1,4750	1,4900	1,7530	1,9780	3,0350	0,1482	3	3	0,3359	0,3837
	3	2,2720	2,7010	3,1440	3,4960	3,6840	8,0030	1,6738	4	4	0,4070	0,3978
	4	3,6950	4,5490	6,0150	6,2150	7,2550	10,5000	4,0808	4	3	0,4006	0,3102
	3	1,4590	2,2680	2,9660	3,2350	3,6020	10,0000	2,9448	3	3	0,3163	0,3231
	9	3,1440	7,0060	8,8010	8,7380	10,5200	17,0000	7,4769	4	4	0,4002	0,4080
	15	9,2510	13,6300	15,8800	16,1100	18,4200	25,0000	10,7864	3	3	0,3000	0,2920
	4	1,4740	3,3720	4,0710	4,1120	5,0330	6,5100	1,1999	3	3	0,3100	0,3291
	20	4,0710	17,9500	19,6700	19,6200	22,3800	29,2500	22,8905	4	4	0,4000	0,3871
	36	18,4000	33,6300	37,4500	36,6000	40,1200	52,0000	41,4268	3	3	0,3000	0,2930
30	2	1,4690	1,4740	1,7660	1,7270	1,9240	2,0880	0,0509	8	8	0,3257	0,3271
	3	2,2690	2,4860	3,1440	3,1150	3,5350	4,3510	0,4174	10	11	0,3580	0,3734
	4	4,0080	4,9520	5,4520	5,9880	5,7570	12,0000	4,3641	12	12	0,4026	0,3809
	3	1,4510	2,6030	2,9660	3,1110	3,4040	7,1230	0,9461	11	11	0,3762	0,3681
	9	4,3090	7,7890	9,3330	9,1970	10,2600	15,2500	4,2394	11	12	0,3670	0,3929
	15	12,6700	15,0900	16,8300	17,3900	19,1700	29,0000	8,8426	8	8	0,2500	0,2580
	4	2,8090	3,7520	4,1620	4,1700	4,5210	6,1560	0,3786	10	10	0,3392	0,3373
	20	12,4000	18,9500	20,0000	20,2400	21,7200	27,6200	6,3582	11	11	0,3667	0,3613
	36	31,3100	35,6200	37,2500	37,4600	39,1200	46,2000	7,8876	9	9	0,3000	0,3077
50	2	1,4710	1,4720	1,8920	1,8140	1,9850	2,7800	0,1196	14	16	0,3312	0,3735
	3	2,2690	2,7690	3,1440	3,6120	4,0480	9,0010	2,5937	17	18	0,3494	0,3796
	4	4,4000	5,0500	5,6560	6,9190	9,2510	12,0000	7,1817	19	16	0,3812	0,3247
	3	1,4510	2,6920	3,0180	3,0610	3,5220	4,4340	0,3778	17	17	0,3475	0,3524
	9	4,5050	8,1270	9,4080	9,2140	10,4700	13,5900	3,7248	20	20	0,4001	0,4026
	15	11,6800	15,4300	16,9200	17,4200	18,8800	27,0000	9,1130	13	13	0,2600	0,2624
	4	3,0740	3,8320	4,2440	4,1940	4,5250	5,3730	0,2837	17	17	0,3453	0,3484
	20	17,0800	19,5200	20,6400	21,0300	21,9700	33,1100	6,4980	18	19	0,3600	0,3710
	36	31,6400	36,1600	37,2900	37,5400	38,4800	51,0000	7,8540	14	14	0,2800	0,2866
100	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	1,7720	2,7850	3,0280	3,0750	3,2950	5,1520	0,3005	33	33	0,3462	0,3444
	9	4,0710	8,5890	9,4570	9,4870	10,3900	14,9600	3,3695	46	44	0,4551	0,4369
	15	12,6900	15,5100	17,1000	18,0200	19,2800	29,0000	12,2899	22	24	0,2200	0,2354
	4	3,1990	3,8830	4,1300	4,1220	4,3480	5,3960	0,1352	33	33	0,3371	0,3391
	20	16,7500	19,3500	20,0000	20,1900	20,8900	25,8600	1,5951	36	36	0,3600	0,3572
	36	33,2400	36,0700	37,3000	37,0200	38,0100	40,1700	1,9297	31	31	0,3100	0,3096
	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	1,9010	3,1140	3,4050	3,2360	3,4850	3,9510	0,1810	381	355	0,3932	0,3696
9	5,1760	9,2310	11,0700	10,6400	12,6200	12,9900	5,2422	574	492	0,5740	0,4919	
15	13,320	16,0400	20,8800	21,8700	27,6300	32,0000	34,6046	67	154	0,0650	0,1534	
1.000	4	3,9110	4,0970	4,1750	4,1790	4,2560	4,4540	0,0121	336	335	0,3411	0,3400
	20	19,1500	19,8900	20,2800	20,2000	20,4600	21,3800	0,1819	348	347	0,3475	0,3470
	36	35,6600	36,4700	36,8200	36,8100	37,1000	38,0300	0,2181	317	318	0,3170	0,3182
	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	1,9480	3,4660	3,4810	3,3950	3,4990	3,9040	0,1046	3,918	3,781	0,4043	0,3915
	9	5,7390	9,0660	12,6700	11,2600	12,7200	12,8600	5,1041	6,036	4,971	0,6036	0,4972
	15	13,7500	15,6200	28,3500	24,7600	29,4600	32,5000	45,0936	13	1,248	0,0013	0,1248
	4	4,0840	4,1520	4,1820	4,1790	4,1990	4,2620	0,0015	3,368	3,363	0,3420	0,3416
	20	19,8000	19,9400	20,0100	20,1200	20,3200	20,4600	0,0437	3,426	3,445	0,3426	0,3445
36	36,2400	36,4500	36,5500	36,6700	36,9300	37,1200	0,0653	3,222	3,192	0,3222	0,3192	

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

Tabela 35 – Resultados do algoritmo EM para as estimativas $\hat{\lambda}_k$, \hat{n}_k e \hat{q}_k para mistura de Poisson para modelo truncado no zero com 4 componentes.

N	λ_{real}	$\hat{\lambda}_k$							\hat{n}_k		\hat{q}_k	
		Min	Q_1	Q_2	Média	Q_3	Max	Var	Q_2	Média	Q_2	Média
	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
10	3	1,4590	2,2660	3,1420	3,4210	4,0710	12,0000	3,4657	2	2	0,2237	0,2507
	9	2,6980	5,6530	8,0030	7,8550	9,5010	16,2000	9,2828	2	2	0,2002	0,2305
	15	5,0330	12,0000	14,0000	14,3700	16,4200	25,5000	16,3254	2	3	0,2000	0,2839
	21	11,2000	19,0000	22,6700	21,9900	24,7500	33,0000	20,8147	2	3	0,2000	0,2525
	4	1,4600	3,1430	3,9770	4,9080	5,0330	30,0000	12,8675	2	3	0,2237	0,2779
	20	5,2770	17,0000	19,5000	20,4400	22,4200	49,2500	64,1889	2	2	0,2000	0,2170
	36	14,0000	28,0000	33,8800	33,8900	39,4400	61,5000	85,4243	2	2	0,2000	0,2320
	52	26,2500	48,8300	53,7500	52,8300	58,1200	73,0000	68,7170	3	3	0,3000	0,2810
30	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	1,4590	2,5520	3,1430	3,2550	3,9130	5,7680	0,9184	8	8	0,2757	0,2665
	9	4,0710	7,0950	8,7100	8,4780	10,0000	13,7800	4,4549	6	7	0,2001	0,2283
	15	8,8010	12,5900	14,9200	14,8500	16,8800	22,4300	8,3309	8	9	0,2667	0,2977
	21	13,6400	20,4700	22,9100	23,7900	26,8100	36,0000	20,4402	7	7	0,2333	0,2207
	4	2,5530	3,6020	4,0710	4,6620	4,7450	15,5400	5,0605	8	8	0,2686	0,2729
	20	12,0000	18,1500	19,6200	19,5700	21,5000	26,6700	8,8838	6	6	0,2000	0,2083
	36	22,0000	31,0000	33,5000	33,4100	35,6900	53,1000	26,5779	6	6	0,2000	0,2107
52	43,0000	49,7400	52,5000	52,8200	55,1200	73,0000	21,5244	10	9	0,3333	0,3133	
50	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	1,4510	2,5790	3,0780	3,1490	3,5750	6,7580	0,8493	13	13	0,2626	0,2560
	9	4,7100	7,1310	8,5690	8,5860	10,0000	12,5900	3,6559	12	13	0,2409	0,2560
	15	10,1700	13,3000	14,7100	15,0000	16,1900	23,1800	6,8926	12	13	0,2400	0,2670
	21	18,6100	20,8400	22,1900	23,5100	24,1300	36,0000	16,2697	12	11	0,2400	0,2250
	4	2,9180	3,7210	4,0710	4,1060	4,4290	6,0850	0,3137	12	13	0,2468	0,2629
	20	12,0000	17,4400	19,0000	18,8300	20,2100	24,8000	6,1637	10	11	0,2000	0,2086
	36	22,1700	29,7600	32,5900	32,8900	36,0000	50,0000	29,2798	11	11	0,2100	0,2108
52	38,5000	48,9100	51,8200	52,1000	54,3300	72,0000	27,7212	17	16	0,3400	0,3226	
100	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	1,4520	2,6240	3,0620	3,1100	3,5170	6,1030	0,6711	26	26	0,2530	0,2525
	9	4,0210	6,7140	8,5900	8,2230	9,5390	12,7200	3,9999	23	24	0,2400	0,2348
	15	8,8840	12,7200	14,5700	14,7700	17,2300	20,1400	7,4098	25	29	0,2450	0,2267
	21	17,4700	20,9300	22,5000	24,0900	26,1100	42,0000	22,4697	24	22	0,2900	0,2903
	4	3,1800	3,8630	4,1310	4,1620	4,4050	5,2160	0,1743	25	25	0,2500	0,2482
	20	11,5000	18,6900	19,7600	19,5700	20,7900	23,2600	3,8113	24	24	0,2400	0,2348
	36	22,5900	32,0700	34,2300	34,2600	36,3800	51,4700	16,6587	23	23	0,2250	0,2267
52	46,0900	51,1000	53,6600	53,7300	56,1300	75,0000	18,2234	29	29	0,2900	0,2903	
1.000	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	1,4520	2,5920	3,2320	3,0880	3,5110	4,4570	0,4464	274	244	0,2835	0,2561
	9	3,9420	7,2930	8,4590	8,1390	9,5070	10,9600	3,1366	225	231	0,2263	0,2309
	15	9,4550	12,0700	14,4300	14,6700	17,0700	20,3500	8,3129	230	295	0,2295	0,2949
	21	18,3300	20,1200	21,8900	25,1200	32,0000	38,0000	45,4818	269	230	0,2685	0,2301
	4	3,8180	4,0780	4,1760	4,1710	4,2630	4,5160	0,0196	249	249	0,2529	0,2531
	20	11,5500	19,1100	19,7100	19,4700	20,1100	21,9600	2,2487	246	239	0,2455	0,2387
	36	23,2200	33,3000	34,7500	34,8900	36,1100	46,2700	14,7082	249	251	0,2485	0,2509
52	46,6000	50,9100	52,1500	52,7900	53,7300	70,6200	19,3314	267	261	0,2665	0,2612	
10.000	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	1,9100	2,7720	3,1160	3,0440	3,4840	4,3250	0,3298	2.565	2.417	0,2684	0,2539
	9	3,5530	7,0090	8,2920	8,0910	9,1000	11,0900	2,3105	2.382	2.329	0,2383	0,2331
	15	9,0010	12,4700	14,4500	14,7200	17,6900	19,5700	8,5790	2.060	2.952	0,2060	0,2952
	21	17,8300	20,1500	22,0500	26,1000	35,9300	43,0000	65,2322	2.586	2.302	0,2586	0,2302
	4	4,0310	4,1180	4,1620	4,1550	4,1970	4,2650	0,0028	2.514	2.513	0,2553	0,2553
	20	14,8800	19,4600	19,6200	19,8500	20,2900	22,0400	1,1019	2.446	2.510	0,2446	0,2510
	36	25,2500	34,3600	34,9600	36,3200	36,8100	46,3500	16,8638	2.492	2.678	0,2492	0,2678
52	46,4500	51,7100	52,3800	54,4200	53,8900	79,2500	36,0188	2.605	2.300	0,2605	0,2300	

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

Tabela 36 – Resultados do algoritmo EM para as estimativas $\hat{\lambda}_k$, \hat{n}_k e \hat{q}_k para mistura de Poisson para modelo truncado no zero com 5 componentes.

N	λ_{real}	$\hat{\lambda}_k$							\hat{n}_k		\hat{q}_k	
		Min	Q_1	Q_2	Média	Q_3	Max	Var	Q_2	Média	Q_2	Média
	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
10	3	1,4590	2,2660	2,8170	3,3550	3,8160	9,5010	3,3883	2	2	0,2056	0,2253
	9	3,1420	5,5840	8,0030	8,5690	11,0000	17,3300	13,0777	1	2	0,1017	0,1937
	15	4,0710	11,1300	14,7500	14,3600	17,3800	28,0000	22,0633	2	2	0,2000	0,2094
	21	9,0010	17,6300	20,5000	20,3300	23,3800	30,6700	21,9330	2	2	0,2000	0,1907
	27	15,8000	25,1900	29,0000	28,7400	32,0000	40,0000	25,9062	2	2	0,2000	0,1977
	4	1,4600	3,1420	4,0710	5,6150	5,1960	33,0000	26,9816	2	2	0,2056	0,2123
	20	3,1430	14,2500	19,0000	19,2100	24,7500	46,1700	94,6696	1	2	0,1007	0,1750
	36	7,0060	23,0000	31,0000	31,8200	39,2500	60,0000	129,6393	1	2	0,1000	0,1768
	52	20,0000	37,2500	46,0000	45,5700	54,5000	67,3300	124,8786	2	2	0,2000	0,1919
	68	43,5000	59,0000	64,6700	64,8100	71,2500	82,0000	78,3521	3	3	0,3000	0,2515
30	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
	3	1,4500	2,3310	2,9640	3,2960	3,8820	8,2520	1,8565	6	6	0,2351	0,2279
	9	3,7570	6,6420	8,6020	8,6680	10,6300	15,8300	7,0525	5	5	0,1667	0,1829
	15	7,5040	11,9200	14,0000	14,2200	16,0000	22,2500	9,7341	5	5	0,1667	0,1842
	21	13,6000	18,5400	20,0000	20,7000	22,9400	31,0000	10,6110	6	7	0,2000	0,2232
	27	23,0000	26,7200	29,5000	29,5700	32,3500	41,0000	15,2926	6	6	0,2000	0,1943
	4	1,8510	3,4370	4,1860	4,6860	4,8980	16,7500	6,2094	6	6	0,2061	0,2168
	20	5,2770	16,7500	19,1600	19,1000	21,7600	33,7500	21,7601	5	5	0,1667	0,1667
	36	16,7500	27,5900	32,3100	31,8200	36,3000	47,3600	41,1140	4	5	0,1333	0,1660
	52	25,8000	40,1900	46,0000	45,7700	50,6200	64,5000	51,8295	5	6	0,1667	0,1877
68	54,5000	61,4500	65,7800	66,5400	70,2500	93,0000	49,0828	8	8	0,2667	0,2677	
50	2	1,4600	1,4620	1,4640	1,4640	1,4660	1,4680	0,00003	7	7	0,1713	0,1713
	3	2,2680	2,2680	2,2680	2,2680	2,2680	2,2690	0,0000007	9	9	0,2013	0,2013
	4	3,1430	3,2260	3,3090	3,3090	3,3920	3,4750	0,0552	11	11	0,2181	0,2181
	5	4,7740	4,9720	5,1700	5,1700	5,3690	5,5670	0,3142	13	13	0,2617	0,2617
	6	7,7030	7,7570	7,8120	7,8120	7,8660	7,9200	0,0233	11	11	0,2201	0,2201
	3	1,4550	2,5190	3,0760	3,2300	3,7890	7,0060	1,1150	10	11	0,2134	0,2198
	9	3,1420	7,0060	8,6020	8,4310	10,0900	14,3300	5,1827	9	9	0,1800	0,1815
	15	6,2120	12,5500	14,2200	14,0500	16,1000	19,1700	6,8147	9	10	0,1800	0,1988
	21	11,5700	18,3800	20,2200	20,2800	22,0300	29,0000	8,3036	9	10	0,1800	0,2010
	27	21,4400	26,6800	28,0900	28,9100	30,0000	45,0000	14,7730	11	11	0,2200	0,2098
4	2,6490	3,6950	4,0710	4,1110	4,4050	6,7220	0,5142	10	10	0,2033	0,2124	
20	5,1410	17,4300	19,1900	19,0300	20,9400	28,0000	9,6308	9	9	0,1800	0,1764	
36	18,6300	28,7500	32,5000	32,0500	36,2100	43,1900	26,8688	8	8	0,1600	0,1642	
52	32,8600	41,5500	45,6300	46,3300	49,8800	63,6400	41,3226	8	9	0,1600	0,1874	
68	53,5300	62,4100	66,6600	66,9500	69,7900	85,0000	40,2629	15	13	0,2900	0,2640	

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

Tabela 37 – Resultados do algoritmo EM para as estimativas $\hat{\lambda}_k$, \hat{n}_k e \hat{q}_k para mistura de Poisson para modelo truncado no zero com 5 componentes.

N	λ_{real}	$\hat{\lambda}_k$							\hat{n}_k		\hat{q}_k	
		Min	Q_1	Q_2	Média	Q_3	Max	Var	Q_2	Média	Q_2	Média
	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
100	3	1,7160	2,4270	2,7600	2,9060	3,4070	4,6140	0,4797	19	19	0,2030	0,2019
	9	3,1430	6,2180	8,1630	7,7710	9,4330	12,8800	4,3173	18	18	0,1801	0,1848
	15	6,2120	11,5600	13,2500	13,3800	15,4900	18,2000	5,9966	16	18	0,1600	0,1880
	21	13,0700	17,4900	19,6300	19,8800	21,9600	34,0000	11,9830	18	18	0,1800	0,2170
	27	22,9800	25,8000	27,7700	29,2200	30,3600	46,0000	25,2377	24	22	0,2400	0,2199
	4	3,3060	3,8710	4,1460	4,1640	4,4320	5,6360	0,1936	20	20	0,2028	0,2032
	20	10,0000	17,7500	19,0300	18,6600	19,9200	28,5100	6,4074	18	18	0,1750	0,1748
	36	21,6200	28,9500	31,6300	31,4000	34,1100	42,0000	17,5540	15	15	0,1500	0,1506
	52	34,8000	41,2700	45,2900	45,3500	48,4500	60,4700	23,4551	20	20	0,1950	0,1994
	68	58,2100	63,5700	65,6200	66,3000	67,9700	84,0000	21,7004	29	28	0,2900	0,2754
	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
1.000	3	1,8570	2,6730	3,1560	3,0020	3,4420	4,7140	0,4265	214	191	0,2231	0,2010
	9	3,5450	6,9920	8,2090	8,0090	9,1760	11,9900	3,7582	176	186	0,1763	0,1860
	15	8,3640	12,3600	13,0300	13,4600	14,6900	18,5100	3,7540	154	178	0,1540	0,1780
	21	14,9700	17,1100	19,1400	19,7900	22,2900	26,7200	8,8592	160	227	0,1595	0,2272
	27	23,9900	25,9800	27,5900	30,6400	29,7500	49,0000	50,2111	260	218	0,2595	0,2183
	4	3,7340	4,0410	4,1360	4,1300	4,2500	4,5710	0,0223	202	202	0,2054	0,2053
	20	17,0200	18,8700	19,2200	19,2800	19,7800	20,7500	0,4238	185	185	0,1850	0,1848
	36	24,4700	31,0300	32,1800	32,3400	33,2200	40,3100	5,7936	143	149	0,1430	0,1488
	52	41,0600	44,3300	46,1700	46,9000	47,5800	65,1200	21,5941	218	217	0,2180	0,2169
	68	62,5200	64,9600	66,3400	67,6600	67,7700	94,0000	35,1949	266	248	0,2655	0,2474
	I	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)	(o)
10.000	3	1,8990	2,3910	3,1240	2,9670	3,4730	4,7780	0,3948	2.050	1.849	0,2143	0,1949
	9	3,5710	6,9190	8,1020	7,8230	8,7350	11,4600	2,9789	1.682	1.840	0,1684	0,1843
	15	9,0980	12,4700	12,9600	13,2000	14,0200	18,5000	3,1714	1.634	1.760	0,1635	0,1760
	21	13,970	17,0600	18,9700	19,7200	22,8300	27,6200	10,9279	1.648	2.371	0,1648	0,2371
	27	23,4800	25,8400	27,0100	32,2100	43,2200	52,0000	87,5805	2.721	2.180	0,2721	0,2180
	4	4,0220	4,0860	4,1160	4,1150	4,1390	4,2610	0,0019	1.992	1.992	0,2027	0,2025
	20	18,5300	19,0900	19,1600	19,1600	19,2300	20,0300	0,0492	1.845	1.846	0,1845	0,1846
	36	29,1000	31,9400	32,0200	32,0300	32,1600	36,8800	0,7543	1.420	1.416	0,0872	0,1376
	52	42,8300	45,5100	45,9900	46,0900	46,5400	51,9700	1,3672	2.128	2.141	0,2128	0,2141
	68	64,6600	65,8400	66,3000	66,3900	66,9000	71,0800	0,7102	2.617	2.605	0,1594	0,2492

(o) : cenário em que não ocorreu a convergência nas estimativas dos parâmetros.

Fonte: Elaborada pelo autor.

CÓDIGOS EM R

Para maiores informações sobre as ferramentas utilizadas, o *software* estatístico **R** e os *IDEs* **RStudio** e **Tinn-R** consulte:

- <http://www.r-project.org/>,
- <https://www.rstudio.com/> e
- <https://sourceforge.net/projects/tinn-r/>.

C.1 Códigos para Gráfico de Mistura de Poisson

Código-fonte 1 – Códigos para gerar Gráfico representativo de Mistura de Poisson

```
1: # Simulador
2: # Mistura de Poisson
3: N=3000
4: dists = runif(N)
5: probs = c(.333,.778)
6: #
7: xpois = vector(length=N)
8: smult = vector(length=N)
9: #
10: for(i in 1:N){
11:   if(dists[i]<probs[1]){
12:     xpois[i] = rpois(1, lambda=3)
13:     smult[i] = 1
14:   } else if(dists[i]<probs[2]){
15:     xpois[i] = rpois(1, lambda=9)
```

```

16:   smult[i] = 2
17: } else {
18:   xpois[i] = rpois(1, lambda=15)
19:   smult[i] = 3   }}
20: #
21: ## distribuição para cada componente
22: d1.xpois<-(length(xpois[smult==1])/length(xpois))*dpois(x1pois,
   lambda=3)
23: d2.xpois<-(length(xpois[smult==2])/length(xpois))*dpois(x2pois,
   lambda=9)
24: d3.xpois<-(length(xpois[smult==3])/length(xpois))*dpois(x3pois,
   lambda=15)
25: #
26: ## gráfico das distribuições
27: hist(xpois, main="Mistura de Poisson", xlab="x", ylab="
   distribuição de x", freq=F)
28: points(xpois[smult==1], d1.xpois, col=2, pch=19)
29: points(xpois[smult==2], d2.xpois, col=3, pch=15)
30: points(xpois[smult==3], d3.xpois, col=4, pch=17)
31: #
32: legend("topright", legend=c(expression(paste(theta, " = 3")),
   expression(paste(theta, " = 9")), expression(paste(theta, "
   = 15"))), col=c(2,3,4), pch=c(19, 15, 17), bty=FALSE, box.
   lwd = 0,box.col = "white",bg = "white")

```

C.2 Códigos para Poisson truncada no zero

Geramos a distribuição Poisson truncada no zero a partir da função *rTruncPoiss* detalhada no [Código-fonte 2](#).

Código-fonte 2 – Função para gerar dados da distribuição Poisson truncada no zero

```

1: rTruncPoiss=function(n,theta){
2:   nt = 2*n
3:   x = rpois(nt, theta)      # dados gerados da Poisson
4:   x=x[x>0]                  # truncar zeros
5:   while(length(x)<n){ x=c(x, rpois(n, theta)) ; x=x[x>0] }
6:   x[1:n]      }

```

Para uma amostra de tamanho 20 e parâmetro $\theta = 3$, escrevemos $x = rTruncPoiss(n = 20, \theta = 3)$.

Estimamos o parâmetro θ da Poisson truncada no zero através do algoritmo EM com a função descrita no [Código-fonte 3](#).

Código-fonte 3 – Função para estimar o parâmetro para distribuição Poisson truncada no zero

```

1: est.pois.trunc = function(sx, n){
2:   n0=1; tch=0; tch0=1           # valores iniciais
3:   while(abs(tch-tch0)>1e-4){    # critério de parada
4:     tch0 = tch
5:     tch = sx/(n + n0)
6:     etch = exp(-tch)
7:     n0 = round(n*etch/(1-etch)) } # Esperança Binomial Negativa
8: tch    }
```

Para encontrar a estimativa do parâmetro $\hat{\theta}$ de uma amostra de tamanho 20 e valor inicial $n_0 = 1$, codificamos $t.est = est.pois.trunc(sx = sum(x), n = 20)$.

C.3 Códigos para Mistura de Poisson truncadas no zero

O [Código-fonte 4](#) gera dados para 3 componentes na mistura com distribuição Poisson truncada no zero e parâmetros $\theta_1 = 3$, $\theta_2 = 9$ e $\theta_3 = 15$, utiliza o [Código-fonte 3](#).

Código-fonte 4 – Função para gerar dados com distribuição Poisson truncada no zero

```

1: N = 3000           # tamanho da amostra
2: dists = runif(N)   # gera variável Uniforme
3: probs = c(.333, .667)
4: xpoist = vector(length=N)
5: smult = vector(length=N)
6: #
7: for(i in 1:N){
8:   if(dists[i] < probs[1]){
9:     xpoist[i] = rTruncPoiss(n=1, theta=3)
10:    smult[i] = 1
11:   }else if(dists[i] < probs[2]){
12:     xpoist[i] = rTruncPoiss(n=1, theta=9)
13:     smult[i] = 2
14:   }else{
15:     xpoist[i] = rTruncPoiss(n=1, theta=15)
16:     smult[i] = 3    }
```

Apresentamos no [Código-fonte 5](#) os comandos para estimar os parâmetros do modelo de mistura para distribuição Poisson truncada no zero, definido no [Algoritmo 5](#). É necessário gerar os dados no [Código-fonte 2](#).

Código-fonte 5 – Algoritmo EM para estimar parâmetros do modelo de mistura de Poisson truncadas no zero

```

1: ## definição das variáveis de saída
2: ## para cada réplica de amostra
3: tempo0=proc.time() # tempo de processamento
4: qte.it=rep(0,M) # quantidade de iterações do algoritmo para
   obtenção dos parâmetros estimados
5: qte.nk=matrix(0,M,K) # quantidade de observações por componente
6: theta.ch=matrix(0,M,K) # parâmetros estimados da Poisson
   truncada no zero
7: var.theta=matrix(0,M,K) # variância de theta
8: S.ch=matrix(0,M,N) # valores estimados para determinar o grupo
   da variável gerada
9: p.ch=matrix(0,M,K) # estimativa do parâmetro p do modelo de
   mistura
10: #
11: ## para cálculo da prob de cada elemento
12: dpt0<-matrix(0,N,K)
13: deno<-matrix(0,N,K)
14: pr.S<-matrix(0,N,K)
15: #
16: for (w in 1:M){
17:   X=xpoist[w,] # matriz de dados gerados para mistura de
   Poisson truncada no zero
18:   ### valores iniciais
19:   s.est=sample(1:K,N, replace=TRUE) # valores estimados para
   determinar o grupo da variável gerada
20:   n.it=0 # quantidade de iterações do algoritmo para obtenção
   dos parâmetros estimados
21:   t.est=c(min(X), median(X), max(X)) # parâmetros estimados da
   Poisson truncada no zero
22:   Var.t=rep(0,K)
23:   tch0=rep(0, K)
24:   nk=rep(1, K) # quantidade de observações por componente da
   mistura
25:   p.est=rep(1/K, K) # estimativa do parâmetro p do modelo de
   mistura

```

```

26:  pch0=rep(1, K) # utiliza no while
27:  lim.p=1
28:  lim.t=1
29:  #
30:  while((lim.p>1e-4)||lim.t>1e-4){ # critério de parada
31:    pch0=p.est
32:    tch0=t.est
33:    #
34:    ### Estima S
35:    #
36:    exp.t=exp(-t.est)
37:    for(k in 1:K){
38:      for (i in 1:N){
39:        # distribuição Poisson truncada no zero para todos os k
40:        dpt0[i,k]=(exp.t[k]*(t.est[k]^X[i]))/((1-exp.t[k])*
factorial(X[i]))
41:        deno[i,k]=p.est[k]*dpt0[i,k] # denominador para cálculo
da Pr(si)
42:      }}
43:      #dpt0<1 #probabilidades
44:      #
45:      #
46:      deno1=apply(deno, 1, sum) # denominador para cálculo da
Pr(si)
47:      #
48:      for(k in 1:K){
49:        for (i in 1:N){
50:          pr.S[i,k]=(p.est[k]*dpt0[i,k])/deno1[i] #Pr(si) para
cada k
51:        }}
52:        #
53:        max.pr.S=apply(pr.S, 1, max) #máximo Pr(si) para k
54:        ss=(max.pr.S==pr.S)*1
55:        nc=apply(ss,2,sum) # tamanho de cada grupo
56:        #
57:        for(k in 1:K){
58:          ss[,k]=(ss[,k])*k
59:        }
60:        s.est=apply(ss, 1, sum)
61:        #
62:        ### Estima theta_k
63:        #

```

```

64:   for (k in 1:K){
65:     nk[k]=length(s.est[s.est==k])
66:     t.est[k]=est.pois.trunc(sx=sum(X[s.est==k]), n=nk[k])
67:   }
68:   #
69:   ### Estima p_k
70:   #
71:   p.est=nk/N
72: #
73:   n.it=n.it+1
74:   lim.p=max(abs(pch0-p.est))
75:   lim.t=max(abs(tch0-t.est))
76: }
77: #
78:   ### Cálculo da variância do estimador de theta
79:   #
80:   exp.t=exp(-t.est)
81:   for (k in 1:K){
82:     Var.t[k]=(1/nk[k])*(t.est[k]*(1-exp.t[k])^2)/(1-exp.t[k]-(t
      .est[k]*exp.t[k]))
83:   }
84:   #
85:   ##
86:   qte.it[w]=n.it # quantidade de iterações
87:   theta.ch[w,]=t.est # estimativas de theta
88:   var.theta[w,]=Var.t # variância de theta
89:   S.ch[w,]=s.est # estimativas de S
90:   p.ch[w,]=p.est # estimativas de p
91:   qte.nk[w,]=nk # quantidade de elementos em cada componente
92:   tempo.proc=proc.time()-tempo0 # tempo de processamento
93: }
94: #

```

C.4 Códigos para mistura de Poisson com variável truncada no zero

Para a metodologia discutida no [Capítulo 5](#) e para o [Algoritmo 6](#) escrevemos o [Código-fonte 6](#) para estimação dos parâmetros do modelo de mistura de Poisson para variável truncada no zero.

Código-fonte 6 – Algoritmo EM para estimar parâmetros do modelo de mistura de Poisson para variável truncada no zero

```
1: K=max(smult) # variável de entrada
2: #
3: tempo0=proc.time() # tempo de processamento
4: # declaração das variáveis
5: qte.it=rep(0,M) # quantidade de iterações do algoritmo para
   obtenção dos parâmetros estimados
6: qte.nk=matrix(0,M,K) # quantidade de observações por componente
   da mistura
7: lambda.ch=matrix(0,M,K) # parâmetros estimados da Poisson
   truncada no zero
8: q.ch=matrix(0,M,K) # estimativa do parâmetro p do modelo de
   mistura
9: #
10: ##
11: for (w in 1:M){
12:   X=xpoist[w,]
13:   ###
14:   # valores iniciais
15:   n.it=0
16:   # parâmetros
17:   lambda=c(min(X), median(X), max(X)) # valores iniciais
   baseados nos dados
18:   s.est=sample(1:K,N, replace=TRUE) ## inicial aleatório
19:   nk=rep(1, K) # quantidade de observações por componente da
   mistura
20:   #
21:   lim.q=1
22:   lim.l=1
23:   ###
24:   #
25:   while((lim.q>1e-4) || (lim.l>1e-4)){ # critério de parada
26:     #
27:     q.est0=q.est
28:     lambda0=lambda
29:     #####
30:     # Estima S
31:     #
32:     facX=factorial(X)
33:     exp.l=exp(-lambda)
```

```

34:   s.est=estima.S(K, N, X, lambda, q.est, exp.l, facX)
35:   #
36:   #####
37:   # Estima lambda_k
38:   #
39:   nk=qte.comp(s.est, K) # cálculo de nk
40:   lambda=c.lambda(K, X, s.est, nk, N, q.est, exp.l)
41:   #
42:   #####
43:   # Estima q_k
44:   #
45:   exp.l=exp(-lambda)
46:   q.est=c.qest(K, nk, N, exp.l, q.est)
47:   #
48:   n.it=n.it+1
49:   lim.l=max(abs(lambda0-lambda))
50:   lim.q=max(abs(q.est0-q.est))
51: #
52: }# fecha while
53: #
54: ##
55: qte.it[w]=n.it
56: lambda.ch[w,]=lambda
57: q.ch[w,]=q.est
58: qte.nk[w,]=nk
59: tempo.proc=proc.time()-tempo0
60: #

```

Com a finalidade de diminuir o tempo de processamento do algoritmo EM na obtenção das estimativas dos parâmetros de interesse apresentamos no [Código-fonte 7](#) as funções executadas no [Código-fonte 6](#).

Código-fonte 7 – Funções utilizadas no [Código-fonte 6](#)

```

1: ## cálculo de nk
2: qte.comp=function(s.est, K){
3:   for (k in 1:K){
4:     nk[k]=length(s.est[s.est==k])   }
5:   nk   }
6: #
7: ## estima S
8: estima.S=function(K, N, X, lambda, q.est, exp.l, facX){

```

```

9:   lamxi=matrix(0,N,K)
10:  nume=matrix(0,N,K)
11:  deno=matrix(0,N,K)
12:  for (k in 1:K){
13:    for (i in 1:N){
14:      lamxi[i,k]=lambda[k]^X[i]
15:      nume[i,k]=(q.est[k]*exp.l[k]*lamxi[i,k])/((1-(q.est[k]*
      exp.l[k]))*facX[i])
16:      deno[i,k]=q.est[k]*(exp.l[k]/(1-exp.l[k]))*(lamxi[i,k]/
      facX[i])    } }
17:  deno1=apply(deno, 1, sum)
18:  Pr.sik=matrix(0,N,K)
19:  for (k in 1:K){
20:    for (i in 1:N){
21:      Pr.sik[i,k]=nume[i,k]/deno1[i]    }}
22:  max.pr.S=apply(Pr.sik, 1, max)
23:  ss=(max.pr.S==Pr.sik)*1
24:  for(k in 1:K){
25:    ss[,k]=(ss[,k])*k    } # determinação dos grupos
26:    s.est=apply(ss, 1, sum)    }
27:  s.est    }
28: #
29: ## estima lambda_k
30: c.lambda=function(K, X, s.est, nk, N, q.est, exp.l){
31:   somaX=rep(0,K)
32:   est.lamb=rep(0,K)
33:   for (k in 1: K){
34:     somaX[k]=sum(X[s.est==k])
35:     est.lamb[k]=(1/(nk[k] - (N*q.est[k]*exp.l[k])/(1-(q.est[k]*
      exp.l[k]))))*somaX[k]    }
36:   est.lamb    }
37: #
38: ## estima q_k
39: c.qest=function(K, nk, N, exp.l, q.est0){
40:   for (k in 1: K){
41:     q.est[k]=nk[k]/(N*(1-(exp.l[k]/(1-q.est0[k]*exp.l[k]))))
      }
42:   q.est    }
43: #

```
