
Modelos multivariados para dados de contagem com excessos de
zeros

Rogério Alves Santana

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

ROGÉRIO ALVES SANTANA

**Modelos multivariados para dados de contagem
com excesso de zeros**

Tese apresentada ao Departamento de Estatística – Des/UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.

Orientador: Prof. Dr. Marinho Gomes de Andrade Filho

**São Carlos
Janeiro de 2020**

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

ROGÉRIO ALVES SANTANA

Multivariate models for count data with zeros excess

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

Advisor: Prof. Dr. Marinho Gomes de Andrade Filho

**São Carlos
January 2020**



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Rogério Alves Santana, realizada em 10/12/2019:



Prof. Dr. Marinho Gomes de Andrade Filho
USP



Prof. Dr. Carlos Alberto Ribeiro Diniz
UFSCar



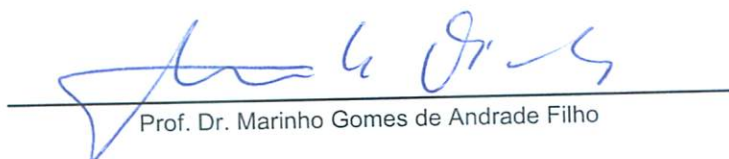
Prof. Dr. Anderson Luiz Ara Souza
UFBA

Prof. Dr. Filidor Edilfonso Vilca Labra
UNICAMP



Prof. Dr. Paulo Henrique Ferreira da Silva
UFBA

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Filidor Edilfonso Vilca Labra e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ao) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.



Prof. Dr. Marinho Gomes de Andrade Filho

Simplesmente a Deus.

AGRADECIMENTOS

Agradeço aos meus familiares; a minha esposa Margarete, por seu companheirismo e compreensão nos momentos mais difíceis do desenvolvimento desta tese. Agradeço também aos amigos que fiz durante o doutorado: Juan Pablo Bustamante; Marcos Antônio Pereira; Rafael Paixão, Marcelo Andrade; Fabiano Coelho, Demerson Poli. A todos os professores do programa em especial aos professores que fiz as disciplinas.

Aos professores: Anderson Luiz Ara Souza; Carlos Alberto Ribeiro Diniz; Filidor Edilson Vilca Labra; Paulo Henrique Ferreira da Silva, integrantes da comissão avaliadora, pelas suas valiosas sugestões.

Em especial agradeço ao meu orientador Marinho Gomes de Andrade Filho, pela orientação e paciência.

RESUMO

SANTANA, R. A. **Modelos multivariados para dados de contagem com excesso de zeros**. 2020. 105 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Nesta tese, apresentamos duas novas distribuições para modelar dados de contagem multivariados com sobredispersão ou subdispersão, excesso de zeros e correlacionados. Nomeadas de distribuições de COM-Poisson multivariada zero-inflacionado (ZICOMP Tipo I) e (ZICOMP Tipo III), suas construções foram baseadas na extensão das distribuições de Poisson multivariado zero-inflacionado ZIP Tipo I e ZIP Tipo III em ((LIU; TIAN, 2015); (TIAN *et al.*, 2014)). Desenvolvemos importantes propriedades teóricas das duas distribuições, seus modelos de regressão, testes de hipóteses da razão de verossimilhanças para ajudar na escolha do melhor modelo, além de uma análise de diagnóstico para o modelo de regressão da distribuição ZICOMP Tipo I. As distribuições propostas apresentaram bons resultados, tanto no estudo de simulação quanto na análise de dados reais. E, em uma análise de dados reais, as distribuições propostas apresentaram melhores ajustes quando comparadas com as distribuições ZIP Tipo I e ZIP Tipo III, segundo os critérios de informação de Akaike (AIC) e Bayesiano (BIC).

Palavras-chave: COM-Poisson multivariado; Zero-inflacionado; Regressão; ZICOMP Tipo I; ZICOMP Tipo III.

ABSTRACT

SANTANA, R. A. **Multivariate models for count data with zeros excess**. 2020. 105 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

In this thesis, we present two new distributions for modeling multivariate counting data with overdispersion or underdispersion, zeros excess and correlation. Named the zero-inflated multivariate COM-Poisson (ZICOMP Type I) and (ZICOMP Type III) distributions, their constructs were based on the extent in of the zero-inflated multivariate Poisson distributions ZIP Type I and ZIP Type III in ((LIU; TIAN, 2015); (TIAN *et al.*, 2014)). We developed important theoretical properties of the two distributions, their regression models, likelihood ratio hypothesis testing to help choose the best model, and a diagnostic analysis for the ZICOMP Type I distribution regression model. The proposed distributions showed in both simulation study good results and real data analysis. In a real data analysis the proposed distributions showed better adjustments when compared to the ZIP Type I and ZIP Type III distributions, according to the Akaike (AIC) and Bayesian (BIC) information criteria.

Keywords: Multivariate COM-Poisson; Zero-inflated; Regression; ZICOMP Type I; ZICOMP Type III.

LISTA DE ILUSTRAÇÕES

Figura 1 – Mapa Político do Brasil com o Estado de Alagoas.	44
Figura 2 – Mapa Político do Brasil com o Estado do Ceará.	61
Figura 3 – Análise de influência	75

LISTA DE TABELAS

Tabela 1 – Dados bivariados com subdispersão e excesso de zeros.	29
Tabela 2 – Resultados da simulação para o modelo da distribuição ZICOMP Tipo I sem covariáveis.	41
Tabela 3 – Resultados da simulação para o modelo de regressão linear da distribuição multivariada ZICOMP Tipo I.	42
Tabela 4 – Resultado do teste da razão de verossimilhanças.	43
Tabela 5 – Frequências dos casos confirmados de AIDS, hepatite e sífilis congênita no estado de Alagoas-Brasil disponível em (ALAGOAS, 2014).	44
Tabela 6 – EMV para os modelos $\text{ZIP}_3^{(1)}(\phi; \lambda_1, \lambda_2, \lambda_3)$ e $\text{ZICOMP}_3^{(1)}(\phi; \lambda_1, \lambda_2, \lambda_3; \nu_1, \nu_2, \nu_3)$	45
Tabela 7 – Critério de seleção dos modelos multivariados ZIP Tipo I e ZICOMP Tipo I.	45
Tabela 8 – Estimativas de máxima verossimilhança e os intervalos de confiança $\text{ICa}(95\%)$ e $\text{IBCa}(95\%)$ para o modelo ZICOMP Tipo I.	46
Tabela 9 – Estimativas das correlações entre as variáveis Y_1, Y_2 e Y_3	46
Tabela 10 – Estimativas do modelo de regressão para o vetor (Y_1, Y_3) associado à covariável X	47
Tabela 11 – Resultado da simulação para o modelo ZICOMP Tipo III com sobredispersão.	58
Tabela 12 – Resultados da simulação para o modelo de regressão multivariada ZICOMP Tipo III.	60
Tabela 13 – Resultados do teste da razão de verossimilhanças.	61
Tabela 14 – Frequências cruzadas das variáveis Y_1 e Y_2 em 175 cidades do interior do estado do Ceará-Brasil disponível em ((CEARÁ, 2013); (SAÚDE, 2014)).	62
Tabela 15 – Critério de seleção dos modelos multivariados ZIP Tipo III e ZICOMP Tipo III.	62
Tabela 16 – Estimativas do modelo ZICOMP Tipo III sem covariáveis para as variáveis Y_1 e Y_2	63
Tabela 17 – Estimativas do modelo de regressão ZICOMP Tipo III para as variáveis Y_1 e Y_2 com a covariável IDH.	64
Tabela 18 – Estimativas do modelo de regressão ZICOMP Tipo III sem o parâmetro de inflação ϕ_1	65
Tabela 19 – Estimativas de máxima verossimilhança para os modelos de regressão.	74
Tabela 20 – Estimativas do modelo de regressão, para as variáveis Y_1 e Y_2 com a covariável IDH.	75
Tabela 21 – Estimativas do modelo de regressão final.	76

Tabela 22 – Frequência das variáveis Y_1 , Y_2 e X em 97 cidades do interior do estado de Alagoas-Brasil 105

LISTA DE ABREVIATURAS E SIGLAS

AIC	Críério da Informação de Akaike
BIC	Crítério de Informação Bayesiano
BZIGP	bivariada Poisson Generalizada zero-inflacionado
COMP	COM-Poisson
DST	Doenças Sexualmente Transmissíveis
iid	variáveis são independentes e identicamente distribuídas
MBZI Tipo I	Binomial multivariada zero-inflacionado Tipo I
MPGZI Tipo I	Poisson Generalizada multivariada zero-inflacionado Tipo I
MZINB	Binomial Negativa multivariada zero-inflacionado
MZIP	Poisson multivariado zero-inflacionado
ZIBernoulli	Bernoulli zero-inflacionado
ZICOMP	Conway-Maxwell-Poisson zero-inflacionado
ZICOMP Tipo I	COM-Poisson multivariada zero-inflacionado Tipo I
ZICOMP Tipo III	COM-Poisson multivariada zero-inflacionado Tipo III
ZIG	Geométrica zero-inflacionado
ZIGP	Poisson Generalizada zero-inflacionado
ZINB	Binomial Negativa zero-inflacionado
ZIP	Poisson zero-inflacionado
ZIP Tipo I	Poisson multivariado zero-inflacionado Tipo I
ZIP Tipo II	Poisson multivariado zero-inflacionado Tipo II
ZIP Tipo III	Poisson multivariado zero-inflacionado Tipo III

SUMÁRIO

1	INTRODUÇÃO	23
2	MODELO DE REGRESSÃO COM A DISTRIBUIÇÃO MULTIVARIADA DE CONWAY-MAXWELL-POISSON ZERO-INFLACIONADO TIPO I	31
2.1	Introdução	31
2.2	Distribuição de COM-Poisson multivariada zero-inflacionado Tipo I	32
2.2.1	<i>Função de massa de probabilidade conjunta</i>	33
2.2.2	<i>Momentos mistos e função geradora de momentos</i>	34
2.2.3	<i>Distribuição marginal</i>	35
2.2.4	<i>Distribuição condicional</i>	35
2.2.4.1	<i>Distribuição condicional $\mathbf{Y}^{(1)} \mathbf{Y}^{(2)}$</i>	35
2.2.4.2	<i>Distribuição condicional $Z \mathbf{Y}$</i>	36
2.2.4.3	<i>Distribuição condicional $\mathbf{X} \mathbf{Y}$</i>	36
2.2.4.4	<i>Distribuição condicional $X_i (Y_i = y_i = 0), i=1, \dots, m$</i>	37
2.2.4.5	<i>Distribuição condicional $X_i (Y_i = y_i > 0), i=1, \dots, m$</i>	37
2.3	Modelo de regressão para a distribuição ZICOMP Tipo I	37
2.3.1	<i>Estimação dos parâmetros</i>	38
2.3.2	<i>Intervalo de confiança</i>	38
2.3.3	<i>Teste da razão de verossimilhanças para o vetor de dispersão do modelo ZICOMP Tipo I</i>	40
2.4	Estudo de simulação	40
2.4.1	<i>Ajuste do modelo sem covariáveis</i>	40
2.4.2	<i>Ajuste do modelo com covariáveis</i>	41
2.4.3	<i>Teste de hipóteses</i>	43
2.5	Aplicação com dados reais	43
2.5.1	<i>Ajuste do modelo sem covariável</i>	45
2.5.2	<i>Análise das distribuições marginais</i>	46
2.5.3	<i>Ajuste do modelo com covariáveis</i>	47
2.6	Conclusão	48

3	MODELO DE REGRESSÃO COM A DISTRIBUIÇÃO MULTIVARIADA DE CONWAY-MAXWELL-POISSON ZERO-INFLACIONADO TIPO III	49
3.1	Introdução	49
3.2	Distribuição de COM-Poisson multivariada zero-inflacionado Tipo III	50
3.2.1	<i>Função de massa de probabilidade conjunta</i>	51
3.2.2	<i>Momentos mistos e função geradora de momentos</i>	52
3.2.3	<i>Distribuição marginal</i>	52
3.2.4	<i>Distribuição condicional</i>	53
3.2.4.1	<i>Distribuição condicional $\mathbf{Y}^{(1)} \mathbf{Y}^{(2)}$</i>	53
3.2.4.2	<i>Distribuição condicional $Z_0 \mathbf{Y}$</i>	54
3.2.4.3	<i>Distribuição condicional $\mathbf{X} \mathbf{Y}$</i>	54
3.2.4.4	<i>Distribuição condicional $X_i (Y_i = y_i = 0), \quad i=1, \dots, m$</i>	54
3.2.4.5	<i>Distribuição condicional $X_i (Y_i = y_i > 0), \quad i=1, \dots, m$</i>	55
3.3	Modelo de regressão para a distribuição ZICOMP Tipo III	55
3.3.1	<i>Estimação dos parâmetros</i>	55
3.3.2	<i>Intervalo de confiança</i>	56
3.3.3	<i>Teste da razão de verossimilhanças para o vetor de dispersão do modelo multivariado ZICOMP Tipo III</i>	57
3.4	Estudo de simulação	57
3.4.1	<i>Ajuste do modelo sem covariáveis</i>	58
3.4.2	<i>Ajuste do modelo com covariáveis</i>	59
3.4.3	<i>Teste de hipóteses</i>	60
3.5	Aplicação com dados reais	61
3.5.1	<i>Ajuste do modelo sem covariável</i>	62
3.5.2	<i>Análise das distribuições marginais</i>	63
3.5.3	<i>Ajuste do modelo com covariáveis</i>	64
3.6	Conclusão	65
4	UM SIMPLES ESTUDO DE PONTOS INFLUENTES NO MODELO DE REGRESSÃO DA DISTRIBUIÇÃO MULTIVARIADA ZICOMP TIPO I	67
4.1	Introdução	67
4.2	Modelo de regressão da distribuição multivariada ZICOMP Tipo I	68
4.3	Teste da bondade do ajuste	69
4.3.1	<i>Teste da razão de verossimilhanças para o parâmetro de inflação total de zeros</i>	70
4.3.2	<i>Teste da razão de verossimilhanças para o vetor de inflação do modelo multivariado ZICOMP Tipo III</i>	70

4.4	Análise de influência	71
4.4.1	<i>Influência global</i>	71
4.4.2	<i>Influência local</i>	72
4.4.3	<i>Ponderação de casos</i>	73
4.5	Aplicação com dados reais	73
4.5.1	<i>Ajuste do modelo com covariável</i>	73
4.6	Conclusão	76
5	CONCLUSÃO	77
REFERÊNCIAS		79
APÊNDICE A	DETALHES TÉCNICOS DA DISTRIBUIÇÃO MULTI-VARIADA ZICOMP TIPO I	89
A.1	Casos especiais da distribuição multivariada ZICOMP Tipo I	89
A.2	Momentos	91
A.3	Distribuições condicionais	92
A.3.1	<i>Distribuição condicional $Y^{(1)} Y^{(2)}$</i>	92
A.3.2	<i>Distribuição condicional $X Y$</i>	93
APÊNDICE B	DETALHES TÉCNICOS DA DISTRIBUIÇÃO MULTI-VARIADA ZICOMP TIPO III	95
B.1	Casos especiais da distribuição multivariada ZICOMP Tipo III	95
B.2	Momentos	98
B.3	Distribuições condicionais	100
B.3.1	<i>Distribuição condicional $Y^{(1)} Y^{(2)}$</i>	100
B.3.2	<i>Distribuição condicional $X Y$</i>	101
APÊNDICE C	INFLUÊNCIA LOCAL	103
C.1	Derivação da matriz de pesos	103
C.1.1	<i>Dados do capítulo 4</i>	105

INTRODUÇÃO

Dados de contagem são comuns nas áreas da agricultura, biologia, educação, engenharia, epidemiologia, indústria, psicologia, saúde pública, entre outras. A literatura envolvendo modelos com dados de contagem é bastante ampla e envolve os modelos de Poisson, Binomial, Binomial Negativa, Poisson Generalizada, COM-Poisson, entre outros. O modelo de regressão linear da distribuição de Poisson é um dos modelos mais utilizados para modelar dados de contagem, pois a maioria das áreas das ciências costumam cumprir os postulados de Poisson (SHMUELI *et al.*, 2005). Entretanto, esse modelo acomoda apenas a equidispersão (variância igual a média). Nas situações em que os dados exibem sobredispersão (variância maior que a média), os modelos Binomial Negativa, Poisson Generalizada e COM-Poisson são os mais indicados por acomodarem sobredispersão. Podem ainda ocorrer situações em que os dados exibem subdispersão (variância menor que a média); neste caso, os modelos COM-Poisson e Poisson Generalizada são mais indicados por possuírem um parâmetro de dispersão mais flexível. Entretanto, para determinados valores do parâmetro de dispersão da distribuição de Poisson Generalizada com subdispersão é necessário realizar reparatrização nos parâmetros, detalhes veja (Consul e Jain (1973); Nelson (1975); Consul e Shoukri (1984)).

A distribuição COM-Poisson (COMP) foi proposta por Conway e Maxwell (1962) na década de 1960, na modelagem de taxas de serviços, e, então, esquecida. Quatro décadas depois, a COM-Poisson é revivida com os trabalhos de (Minka *et al.* (2002); Shmueli *et al.* (2005); Kadane *et al.* (2006)), e, desde então, inúmeros trabalhos têm sido desenvolvidos: em modelos lineares generalizados (Guikema e Goffelt (2008); Lord, Guikema e Geedipally (2008); Francis *et al.* (2012)); análise de sobrevivência Rodrigues *et al.* (2009a); inferência Bayesiana (Chaniavidis (2015); Wu *et al.* (2015)); excesso de zeros (Barriga e Louzada (2014); Sellers e Raim (2016); Choo-Wosoba *et al.* (2017)); dados agrupados ((CHOO-WOSOBA; DATTA, 2017); (CHOO-WOSOBA; LEVY; DATTA, 2016)). Uma variável aleatória Y possui distribuição de COM-Poisson com parâmetros (λ, ν) , e escreve-se $Y \sim \text{COM-Poisson}(\lambda, \nu)$ ou

$Y \sim \text{COMP}(\lambda, \nu)$, se sua função de massa de probabilidade é dada por

$$\Pr(Y = y) = \frac{\lambda^y}{S(\lambda, \nu)(y!)^\nu}, \quad y = 0, 1, 2, \dots, \quad (1.1)$$

em que $\nu \geq 0$ é um parâmetro de dispersão, $\lambda > 0$ é um parâmetro de centro, o qual é aproximadamente a média quando $\nu \approx 1$ (Sellers e Shmueli (2008); Francis *et al.* (2012)); $S(\lambda, \nu)$ é uma constante de normalização, que é dada por

$$S(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}. \quad (1.2)$$

A distribuição de COM-Poisson é mais flexível quando comparada com as outras distribuições de contagem, pois o seu parâmetro de dispersão ν pode acomodar tanto sobredispersão ($\nu < 1$); quanto a subdispersão ($\nu > 1$); equidispersão ($\nu = 1$) e quando $\nu = 0$ e $\lambda > 1$, $S(\lambda, \nu)$ não converge e a distribuição não é definida ((GIUFFRÈ *et al.*, 2013); Gupta, Sim e Ong (2014)).

A sobredispersão nos dados também pode ser provocada pelo excesso de zeros, que ocorre quando a quantidade de zeros é maior que a quantidade esperada pelo modelo proposto. Pode ainda ocorrer uma situação menos comum, quando a quantidade de zeros é menor do que a esperada pelo modelo (deflação de zeros); trabalhos relevantes envolvendo deflação de zeros são encontrados em ((ANGERS; BISWAS, 2003); (CONCEIÇÃO *et al.*, 2017); (DIETZ; BÖHNING, 2000)). O excesso de zeros ocorre com bastante frequência nas áreas de agricultura Ridout, Demétrio e Hinde (1998); bioestatística Broek (1995); criminologia Famoye e Singh (2006); epidemiologia Böhning *et al.* (1999a); indústria Lambert (1992); horticultura Hall (2000); medicina Böhning *et al.* (1999b); saúde pública Zhou e Tu (2000); zoologia Zipkin *et al.* (2014) entre outras. A literatura que visa à modelagem de dados com excesso de zeros é bastante ampla e envolve os modelos: Binomial Negativa zero-inflacionado (ZINB) em Ridout, Hinde e Demétrio (2001); Poisson zero-inflacionado (ZIP) em (Jr (1960); Lambert (1992); Shonkwiler e Shaw (1996); Ridout, Demétrio e Hinde (1998)); Poisson Generalizada zero-inflacionado (ZIGP) em Gupta, Gupta e Tripathi (2005); Conway-Maxwell-Poisson zero-inflacionado (ZICOMP) ou COM-Poisson zero-inflacionado (Barriga e Louzada (2014); Sellers e Raim (2016)), entre outros. Quando a variância dos dados é maior que a média, e independente da sobredispersão provocada pelo excesso de zeros, o modelo ZIP não é indicado e por possuírem um parâmetro de dispersão, os modelos ZINB, ZIGP e ZICOMP são mais apropriados para representar os dados, com destaque para os modelos ZIGP e ZICOMP que conseguem captar tanto sobredispersão quanto subdispersão nos dados.

O modelo da distribuição ZICOMP foi proposto por Barriga e Louzada (2014); em uma abordagem Bayesiana, eles apresentaram o modelo de regressão para distribuição, além da condução de uma análise de diagnóstico de influência. Sellers e Raim (2016) apresentaram o modelo numa abordagem frequentista. Choo-Wosoba, Levy e Datta (2016) incorporaram o efeito aleatório no modelo para dados agrupados. Choo-Wosoba e Datta (2017) apresentaram o modelo

hurdle COM-Poisson com efeito aleatório para estudar o uso de flúor em crianças no estado de Iowa-EUA, [Sim, Gupta e Ong \(2018\)](#) desenvolveram o teste *score* para o parâmetro de inflação do modelo. Segundo [Barriga e Louzada \(2014\)](#) a distribuição ZICOMP é uma distribuição resultante da mistura da distribuição COM-Poisson, $X \sim \text{COMP}(\lambda, \nu)$ com uma distribuição degenerada no ponto zero $\xi \sim \text{Degenerado}(0)$, em que ξ e X são independentes. Dessa forma, a variável aleatória Y segue distribuição ZICOMP(ϕ, λ, ν) se sua função de massa de probabilidade é dada por

$$\begin{aligned} \Pr(Y = y) &= \phi \Pr(\xi = y) + (1 - \phi) \Pr(X = y) \\ &= \begin{cases} \phi + (1 - \phi) \frac{1}{S(\lambda, \nu)}, & y = 0, \\ \frac{(1 - \phi)}{S(\lambda, \nu)} \frac{\lambda^y}{(y!)^\nu}, & y = 1, 2, \dots, \end{cases} \\ &= \left[\phi + \frac{(1 - \phi)}{S(\lambda, \nu)} \right] I(y = 0) + \left[\frac{(1 - \phi)}{S(\lambda, \nu)} \frac{\lambda^y}{(y!)^\nu} \right] I(y > 0). \end{aligned} \quad (1.3)$$

Em que $\phi \in [0, 1)$ é um parâmetro desconhecido que incorpora a proporção de zeros extras além do permitido pelo modelo COM-Poisson, $I(y)$ é uma função indicadora do tipo $I(y) = 1$ se $y = 0$ e $I(y) = 0$ caso contrário, $\lambda > 0$ e $\nu \geq 0$ são parâmetros da distribuição COM-Poisson, especificamente a distribuição ZICOMP equivale a ZIP quando $\nu=1$ ([Barriga e Louzada \(2014\)](#); [Sellers e Raim \(2016\)](#)), $S(\lambda, \nu)$ é a constante de normalização da COM-Poisson.

A literatura envolvendo dados de contagem univariados com e sem o excesso de zeros é bastante consagrada com aplicações nas mais variadas áreas. Entretanto, para a literatura envolvendo dados multivariados a quantidades de trabalhos publicados não ocorrem na mesma proporção. Em geral, existe um interesse do pesquisador em modelar dados que exibem uma certa dependência como por exemplo, os casos de Dengue, Zica e Microcefalia onde existe uma suspeita clínica sobre a dependência entre essas doenças [Beltrán-Silva et al. \(2018\)](#), bem como as doenças crônicas nas articulações tais como, Artrite, Miosite e Tenossivite que estão associadas as infecções causadas por Chikungunya (([MORRISON et al., 2011](#)); ([HAWMAN et al., 2013](#))). Tudo isso motiva o uso de dados multivariados. Trabalhos envolvendo dados multivariados de contagem são encontrados nas áreas da medicina [Shenson e Joshi \(2012\)](#); marketing [Drèze e Hoch \(1998\)](#); esporte ([Lee \(1997\)](#); [Karlis e Ntzoufras \(2003\)](#)); indústria [Ho e Singer \(2001\)](#); transporte [Mehta \(2014\)](#) entre outras. Dentre os modelos multivariados para dados de contagem o modelo de Poisson é o mais popular veja ([Kawamura \(1979\)](#); [Mahamunulu \(1967\)](#); [JOHNSON e KOTZ \(1969\)](#); [Karlis e Meligkotsidou \(2005\)](#)), com destaque para as distribuições bivariadas de Poisson. Existem várias definições da distribuição bivariada de Poisson como as encontradas nos trabalhos de ([M'Kendrick \(1925\)](#); [Campbell \(1934\)](#); [Maritz \(1952\)](#); [Holgate \(1964\)](#); [Kawamura \(1973\)](#); [Marshall e Olkin \(1985\)](#)). Entretanto, as estruturas mais populares são as estruturas

propostas por [Holgate \(1964\)](#), que utiliza três variáveis de Poisson independentes para formar duas variáveis de Poisson dependentes e a de [Kawamura \(1973\)](#), que considera a distribuição bivariada de Poisson como o limite da distribuição binomial bivariada quando n o número de provas de Bernoulli tende ao infinito.

Quanto ao excesso de zeros em conjuntos de dados multivariados, o mesmo ocorre com bastante frequência, pois pode ser do interesse do pesquisador em modelar a ocorrência das frequências de zeros, como por exemplo, na área da saúde. Imagine uma situação onde a imunidade a uma doença A acarreta a imunidade a outras doenças, tais como B e C, por exemplo, casos de AIDS que podem provocar o surgimento de outras infecções tais como tuberculose e diarreia. Isso introduz uma dependência entre as notificações dessas doenças. O mesmo pode ocorrer com as Doenças Sexualmente Transmissíveis (DST), pois ao se proteger da AIDS o indivíduo se protege de outras DST, isso causa uma dependência entre as notificações dessas doenças. Dessa forma, ao considerar os casos de sífilis e hepatite C estudados em [Chen et al. \(2015\)](#), pode se pensar nos indivíduos que são profissionais do sexo e que utilizam prevenção, para esse conjunto de dados, existirá pessoas que contraíram sífilis mas não contraíram hepatite C; não contraíram sífilis e contraíram hepatite C e não contraíram nenhuma dessas doenças, esse tipo de situação motiva a análise estatística em conjuntos multivariados com excesso de zeros. Trabalhos envolvendo dados multivariados de contagem com excesso de zeros são encontrados nas áreas da saúde ([Cameron e Trivedi \(2013\)](#); [Mohammadi, Kheiri e Sedehi \(2016\)](#)); horticultura [Sun \(2016\)](#); indústria [Li et al. \(1999\)](#); transporte [Dong et al. \(2014\)](#) entre outras. A maioria dos modelos multivariados com excesso de zeros são bivariados: Poisson zero-inflacionado ([Karlis, Ntzoufras et al. \(2005\)](#); [Walhin \(2001\)](#)); Poisson Generalizada zero-inflacionado ([Zhang, Tian e Huang \(2015\)](#); [Faroughi e Ismail \(2017\)](#)); Série de Potência zero-inflacionado [Krishna e Tukaram \(2011\)](#); binomial negativa zero-inflacionado [Kim \(2013\)](#), entre outras.

O modelo de Poisson multivariado zero-inflacionado (MZIP) é proposto em [Li et al. \(1999\)](#), cuja estrutura é formada pela mistura de $m+2$ distribuições, sendo uma distribuição degenerada no ponto zero, m distribuições de Poisson univariado com $m - 1$ zeros e uma distribuição multivariada de Poisson sem considerar o excesso de zeros. Dessa forma, a estrutura do vetor aleatório $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)^\top$ pode ser representada da seguinte forma

$$\begin{aligned}
 (Y_1, Y_2, \dots, Y_m)^\top & \\
 & \sim (0, 0, \dots, 0) && \text{com probabilidade } p_0 \\
 & \sim (\text{Poisson}(\lambda_1), 0, \dots, 0) && \text{com probabilidade } p_1 \\
 & \sim (0, \text{Poisson}(\lambda_2), 0, \dots, 0) && \text{com probabilidade } p_2 \\
 & \sim (0, 0, \text{Poisson}(\lambda_3), 0, \dots, 0) && \text{com probabilidade } p_3 \\
 & \sim \quad \vdots && \quad \vdots \quad \vdots \quad \vdots \\
 & \sim (0, \dots, 0, \text{Poisson}(\lambda_m)) && \text{com probabilidade } p_m \\
 & \sim \text{Poisson Multivariada } (\lambda_{10}, \lambda_{20}, \dots, \lambda_{m0}, \lambda_{00}) && \text{com probabilidade } p_{11}, \quad (1.4)
 \end{aligned}$$

em que $p_0 + p_1 + \dots + p_m + p_{11} = 1$ e $\lambda_{10}, \lambda_{20}, \dots, \lambda_{m0}$ e λ_{00} são as médias das variáveis de Poisson independentes X_1, X_2, \dots, X_m e X_0 , com $Y_i = X_i + X_0$, $i = 1, \dots, m$ na estrutura da distribuição de multivariada de Poisson em (JOHNSON; KOTZ, 1969). Considerando a estrutura da distribuição MZIP proposta por Li *et al.* (1999), Dong *et al.* (2014) construiu o modelo da distribuição Binomial Negativa multivariada zero-inflacionado (MZINB). Considerando a estrutura de Olkin (1994) para a distribuição de Poisson Multivariada $(\lambda_{10}, \lambda_{20}, \dots, \lambda_{m0}, \lambda_{00})$ sem considerar o excesso de zeros na Equação (1.4), a distribuição MZIP torna-se bastante complicada, pois o número de parâmetros cresce rapidamente com o aumento no número dimensões m , e, para manter sua estrutura de dependência são necessárias $2^m - 1$ variáveis independentes de Poisson. O que torna difícil a implementação e obtenção dos estimadores de máxima verossimilhança devido à complexidade da função de log-verossimilhança ((LIU; TIAN, 2015); (TIAN *et al.*, 2014)).

Para contornar esses problemas, Liu e Tian (2015) apresentam uma nova estrutura para a distribuição MZIP, para modelar dados multivariados de contagem correlacionados e com excesso de zeros. Nomeada de Poisson multivariado zero-inflacionado Tipo I (ZIP Tipo I), sua estrutura possui um vetor de variáveis Poisson independentes e uma variável aleatória Z para capturar a correlação entre os componentes. Dessa forma, a função de massa de probabilidade conjunta da distribuição multivariada ZIP Tipo I para o vetor $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)^\top$ é dada por

$$\begin{aligned} f(\mathbf{y}|\phi, \boldsymbol{\lambda}) &= \phi \Pr(\boldsymbol{\xi} = \mathbf{y}) + (1 - \phi) \Pr(\mathbf{X} = \mathbf{y}) \\ &= \left[\phi + (1 - \phi)e^{-\lambda_+} \right] I(\mathbf{y} = \mathbf{0}) + \left[(1 - \phi)e^{-\lambda_+} \prod_{i=1}^m \frac{\lambda_i^{y_i}}{y_i!} \right] I(\mathbf{y} \neq \mathbf{0}), \end{aligned}$$

em que $\mathbf{y} = \mathbf{0}$ representa um vetor m -dimensional quanto todos os componentes são iguais a zero e $\mathbf{y} \neq \mathbf{0}$, quando pelo menos um dos elementos do vetor é diferente de zero, dessa forma, $I(\mathbf{y})$ é uma função indicadora do tipo $I(\mathbf{y}) = \mathbf{1}$ se $\mathbf{y} = \mathbf{0}$ e $I(\mathbf{y}) = \mathbf{0}$ caso contrário; $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m)^\top$ com $\{\xi_i\}_{i=1}^m \stackrel{iid}{\sim}$ Degenerado(0); cujas variáveis são independentes e identicamente distribuídas (iid); $\lambda_+ = \sum_{i=1}^m \lambda_i$; ϕ é o parâmetro de inflação de zeros e \mathbf{X} é um vetor de variáveis Poisson independentes (detalhes veja, (LIU; TIAN, 2015)). Devido à simples estrutura do modelo, diversos pesquisadores têm estendido distribuições para classe de distribuições Tipo I, como as distribuições multivariadas: Binomial multivariada zero-inflacionado Tipo I (MBZI Tipo I) em Sun (2016); Poisson Generalizada multivariada zero-inflacionado Tipo I (MPGZI Tipo I) em Huang *et al.* (2017); ZIP Tipo I com restrição nos parâmetros em Ju (2016); e a Poisson inflacionada no zero-um que tem como caso particular a ZIP Tipo I em Zhang *et al.* (2019). A estrutura de inflação das distribuições da família Tipo I considera apenas a inflação conjunta dos componentes do vetor aleatório de contagem, e os demais componentes do vetor aleatório são consideradas como variáveis de Poisson, como por exemplo $(0, \dots, \text{Po}(\lambda_i), \text{Po}(\lambda_j), 0, \dots, 0)$; casos particulares como $(0, \dots, \text{Po}(\lambda_i), \text{ZIP}(\phi_j, \lambda_j), 0, \dots, 0)$ não são considerados pelo modelo (TIAN *et al.*, 2014).

Nesse contexto, Tian *et al.* (2014) propuseram a distribuição de Poisson multivariado

zero-inflacionado Tipo III (ZIP Tipo III), que é mais flexível que a ZIP Tipo I, ao considerar casos particulares do tipo $(0, \dots, \text{Po}(\lambda_i), \text{ZIP}(\phi_j, \lambda_j), 0, \dots, 0)$. A função de massa de probabilidade conjunta da distribuição multivariada ZIP Tipo III para o vetor $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)^\top$ é dada por

$$\begin{aligned} f(\mathbf{y}|\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}) &= \phi_0 \Pr(\boldsymbol{\xi} = \mathbf{y}) + (1 - \phi_0) \Pr(\mathbf{X} = \mathbf{y}) \\ &= \prod_{i=1}^m \left\{ \phi_0 + (1 - \phi_0) \left[\phi_i + (1 - \phi_i) e^{-\lambda_i} \right] \right\} I(\mathbf{y} = \mathbf{0}) + (1 - \phi_0) \\ &\quad \times \prod_{i=1}^m \left\{ \left[\phi_i + (1 - \phi_i) e^{-\lambda_i} \right] I_{(y_i=0)} + \left[(1 - \phi_i) \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \right] I_{(y_i>0)} \right\} I(\mathbf{y} \neq \mathbf{0}), \quad (1.5) \end{aligned}$$

em que $I(\mathbf{y})$ é uma função indicadora do tipo $I(\mathbf{y}) = \mathbf{1}$ se $\mathbf{y} = \mathbf{0}$ e $I(\mathbf{y}) = \mathbf{0}$ caso contrário; $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m)^\top$ com $\{\xi_i\}_{i=1}^m \stackrel{iid}{\sim} \text{Degenerado}(0)$; ϕ_0 é um parâmetro de inflação conjunta de zeros; $\boldsymbol{\phi}$ é um vetor de parâmetros de inflação adicionais, e \mathbf{X} é um vetor de variáveis ZIP independentes.

Quanto a distribuição de Poisson multivariado zero-inflacionado Tipo II (ZIP Tipo II), ela ainda não existe, uma possível estrutura para a ZIP Tipo II poderia ser obtida pela extensão distribuição bivariada Poisson Generalizada zero-inflacionado (BZIGP) em [Zhang, Tian e Huang \(2015\)](#) se considerarmos os componentes do vetor aleatório \mathbf{Y} como variáveis ZIP independentes. Dessa forma, a função de massa de probabilidade conjunta da distribuição multivariada ZIP Tipo II para o vetor $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)^\top$ é dada por

$$f(\mathbf{y}|\boldsymbol{\phi}, \boldsymbol{\lambda}) = \prod_{i=1}^m \left\{ \left[\phi_i + (1 - \phi_i) e^{-\lambda_i} \right] I_{(y_i=0)} + \left[(1 - \phi_i) \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \right] I_{(y_i>0)} \right\}, \quad (1.6)$$

em que $I(y_i)$ é uma função indicadora do tipo $I(y_i) = 1$ se $y_i = 0$ e $I(y_i) = 0$ caso contrário; ϕ_i e λ_i corresponde aos parâmetro de inflação de zeros e a média da Poisson na i -ésima dimensão.

Na prática, conjuntos de dados multivariados com excesso de zeros que exibem sobre-dispersão maior que a provocada pelo excesso de zeros ocorrem com bastante frequência. No entanto, exemplos envolvendo dados multivariados ou univariados com subdispersão e excesso de zeros são difíceis de serem encontrados. Um exemplo envolvendo conjuntos de dados univariados é encontrado em [Choo-Wosoba, Levy e Datta \(2016, p. 14\)](#), que utilizaram o modelo ZICOMP para modelar dados de híbridos de milho, com excesso de zeros e subdispersão. Como ilustração para o caso multivariado, pode-se considerar o conjunto de dados bivariados hipotéticos que aparentemente são subdispersos e com excesso de zeros, conforme pode ser visto na Tabela 1.

Tabela 1 – Dados bivariados com subdispersão e excesso de zeros.

$Y_1 Y_2$	0	1	2	3	Total
0	13	9	11	2	35
1	12	9	7	2	30
2	8	12	5	2	27
3	3	3	0	0	6
4	2	0	0	0	2
Total	38	33	23	6	100

Pode-se observar, na Tabela 1, que as frequências das células (0,0), (0, Y_2) e (Y_1 ,0) são relativamente maiores que as frequências das outras células e, em uma análise descritiva, as variáveis Y_1 e Y_2 apresentaram média maior que a variância amostral: $\bar{y}_1 = 1,10$; $\bar{y}_2 = 0,97$; $s_1^2 = 1,04$ e $s_2^2 = 0,86$. Provavelmente, o modelo da distribuição multivariada MPGZI Tipo I em Huang *et al.* (2017) se ajustaria melhor a este conjunto de dados do que os modelos MZIP em Li *et al.* (1999); MZIP Tipo I em Liu e Tian (2015); MZIP Tipo III em Tian *et al.* (2014) e MZINB em Dong *et al.* (2014), devido à subdispersão existente nos dados.

Quanto ao modelo de COM-Poisson multivariado zero-inflacionado, o mesmo ainda não existe em nenhuma das estruturas dos modelos de ((LI *et al.*, 1999); (LIU; TIAN, 2015); (TIAN *et al.*, 2014)). O campo multivariado para a distribuição COM-Poisson é bastante escasso, com a existência de poucos trabalhos sem o excesso de zeros. Uma proposta para o modelo multivariado é encontrada em Mehta (2014, p. 42) por meio de uma abordagem Bayesiana. Raz (2016) apresenta o modelo bivariado da distribuição para modelagem de dados agrupados via cópula. Pelo conhecimento do autor a única proposta formal é encontrada em Sellers, Morris e Balakrishnan (2016); para a distribuição bivariada de COM-Poisson, em que apresentam a distribuição, suas propriedades teóricas e a obtenção de seus estimadores. Baseado na estrutura da distribuição bivariada proposta por Sellers, Morris e Balakrishnan (2016), Pereira *et al.* (2019, p. 81) propuseram a distribuição bivariada correlacionada de COM-Poisson.

Mediante ao exposto, este trabalho teve por objetivo geral construir as distribuições COM-Poisson multivariada zero-inflacionado Tipo I (ZICOMP Tipo I) e a COM-Poisson multivariada zero-inflacionado Tipo III (ZICOMP Tipo III), considerando as estruturas das distribuições ZIP Tipo I, proposta por Liu e Tian (2015), e a ZIP Tipo III, proposta por Tian *et al.* (2014). Especificamente, desenvolvemos o modelo de regressão para as duas distribuições, testes da razão de verossimilhanças para ajudar na escolha do melhor modelo, além de uma análise de diagnóstico para o modelo de regressão da distribuição ZICOMP Tipo I. Em uma análise de um conjunto com dados reais, envolvendo AIDS e hepatite viral em 97 cidades do interior do estado de Alagoas-Brasil, os modelos destas duas novas distribuições apresentaram melhores ajustes quando comparadas com as distribuições ZIP Tipo I e Tipo III, segundo os critérios de informação Critério da Informação de Akaike (AIC) e Critério de Informação Bayesiano (BIC).

Este trabalho está organizado da seguinte maneira: No Capítulo 2, apresentamos detalha-

damente a construção da distribuição multivariada ZICOMP Tipo I, com o modelo de regressão, teste da razão de verossimilhanças para o parâmetro de dispersão, um estudo de simulação e uma aplicação com dados reais. No Capítulo 3, construímos a distribuição multivariada ZICOMP Tipo III, com o seu modelo de regressão e o teste da razão de verossimilhanças para o parâmetro de dispersão, um estudo de simulação e uma aplicação com dados reais. Já no Capítulo 4, apresentamos mais dois testes da razão de verossimilhanças para a escolha do melhor modelo, além da condução de uma análise de influência global e local para o modelo de regressão da distribuição ZICOMP Tipo I. A conclusão é apresentada no Capítulo 5, e alguns detalhes técnicos das distribuições nos Apêndices A, B e C.

MODELO DE REGRESSÃO COM A DISTRIBUIÇÃO MULTIVARIADA DE CONWAY-MAXWELL-POISSON ZERO-INFLACIONADO TIPO I

2.1 Introdução

Dados de contagem multivariados são encontrados nas áreas da agricultura [Dias \(2014\)](#); esporte ([Lee \(1997\)](#); [Karlis e Ntzoufras \(2003\)](#)); indústria [Ho e Singer \(2001\)](#); marketing [Drèze e Hoch \(1998\)](#); medicina ([Qin et al. \(2016\)](#); [Cha et al. \(2017\)](#); [Russell et al. \(2019\)](#); [Li et al. \(2010\)](#); [Lucar et al. \(2018\)](#); [Shenson e Joshi \(2012\)](#)), entre outras. Assim como no caso univariado o excesso de zeros também ocorre com bastante frequência, como por exemplo, em dados da: saúde [Cameron e Trivedi \(2013\)](#), indústria [Li et al. \(1999\)](#), horticultura [Sun \(2016\)](#); transporte [Dong et al. \(2014\)](#), lesões ocupacionais [Wang et al. \(2003\)](#), entre outras. O modelo da distribuição multivariada de Poisson zero-inflacionado Tipo I (ZIP Tipo I) em [Liu e Tian \(2015\)](#) surge como uma ótima opção para modelagem desses dados, pois o mesmo possui uma estrutura bastante flexível e de fácil implementação quando comparado com o modelo de Poisson multivariado zero-inflacionado em [Li et al. \(1999\)](#). O excesso de zeros pode provocar sobredispersão nos dados, quando a sobredispersão nos dados ocorre de maneira independente da provocada pelo excesso de zeros, o modelo da distribuição ZIP Tipo I não é indicado por acomodar apenas a equidispersão nos dados. Uma possível solução seria estender distribuições univariadas mais flexíveis que acomodassem sobredispersão ou sobredispersão para a família de distribuições multivariadas Tipo I, tais como as distribuições zero-inflacionado ZICOMP e a ZIGP em ([Barriga e Louzada \(2014\)](#); [Gupta, Sim e Ong \(2014\)](#)).

Recentemente o modelo da distribuição multivariada de Poisson Generalizada Tipo I

(MPGZI Tipo I) foi proposto por [Huang *et al.* \(2017\)](#), em sua construção eles consideraram a estrutura da Poisson Generalizada em (([CONSUL; JAIN, 1973](#)); ([CONSUL; SHOUKRI, 1984](#))). Entretanto, para determinados valores do parâmetro de dispersão da distribuição de Poisson Generalizada com subdispersão é necessário realizar reparatrização nos parâmetros, detalhes veja (([CONSUL; JAIN, 1973](#)); ([NELSON, 1975](#)); ([CONSUL; SHOUKRI, 1984](#))). Por esse motivo as distribuições bivariadas de Poisson Generalizada em ([Zhang, Tian e Huang \(2015\)](#); [Faroughi e Ismail \(2017\)](#)) consideram apenas a estrutura de sobredispersão em ([CONSUL, 1989](#)).

Quanto à distribuição de COM-Poisson multivariada zero-inflacionado, esta ainda não existe em nenhuma das estruturas das distribuições de (([LI *et al.*, 1999](#)); ([LIU; TIAN, 2015](#))). O campo multivariado para a distribuição COM-Poisson é bastante escasso com a existência de poucos trabalhos sem o excesso de zeros, pois uma estrutura multivariada para a distribuição de COM-Poisson seria bastante complicada ou muito sofisticada para acomodar as distribuições de Poisson, geométrica e Bernoulli, que são seus casos particulares. Alguns trabalhos envolvendo a COM-Poisson no campo multivariado são encontrados em [Raz \(2016\)](#) e [Khan *et al.* \(2017\)](#) que apresentam estruturas para o modelo bivariado via cópula. Pelo conhecimento do autor a única proposta formal para a distribuição bivariada de COM-Poisson é encontrada em [Sellers, Morris e Balakrishnan \(2016\)](#), que apresentam a distribuição, suas propriedades teóricas e a obtenção de seus estimadores, os quais são difíceis de serem obtidos devido à complexidade da estrutura do modelo.

Mediante ao exposto, este Capítulo teve como objetivo construir a distribuição de COM-Poisson multivariada zero-inflacionado Tipo I (ZICOMP Tipo I), considerando a estrutura da distribuição ZIP Tipo I, proposta em ([LIU; TIAN, 2015](#)). Este Capítulo está organizado da seguinte maneira: na seção [2.2](#), é apresentada a distribuição multivariada ZICOMP Tipo I e algumas propriedades. Na seção [2.3](#), apresentamos, respectivamente, o modelo de regressão para a distribuição ZICOMP Tipo I e o teste da razão de verossimilhanças para ajudar na escolha do melhor modelo. Nas seções [2.4](#) e [2.5](#), apresentamos um estudo de simulação da distribuição proposta, além da análise de um conjunto de dados reais. A conclusão é apresentada na seção [2.6](#).

2.2 Distribuição de COM-Poisson multivariada zero-inflacionado Tipo I

Considere o vetor $\mathbf{X} = (X_1, \dots, X_m)^\top$, com $X_i \stackrel{iid}{\sim} \text{COM-Poisson}(\lambda_i, \nu_i)$ ou $\text{COMP}(\lambda_i, \nu_i)$, para $i = 1, \dots, m$. Dizemos que o vetor aleatório $\mathbf{Y} = (Y_1, \dots, Y_m)^\top$ possui distribuição COM-Poisson multivariada zero-inflacionado Tipo I ou ZICOMP Tipo I, com parâmetros $\phi \in [0, 1)$,

$\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top \in \mathfrak{R}_+^{*m}$ e $\mathbf{v} = (v_1, \dots, v_m)^\top \in \mathfrak{R}_+^m$ se:

$$\mathbf{Y} \stackrel{d}{=} Z\mathbf{X} = \begin{cases} \mathbf{0}, & \text{com probabilidade } \phi, \\ \mathbf{X}, & \text{com probabilidade } (1 - \phi), \end{cases} \quad (2.1)$$

em que $Z \sim \text{Bernoulli}(1 - \phi)$, Z é independente de \mathbf{X} , isto é, $(Z \perp \mathbf{X})$. E escrevemos $\mathbf{Y} \sim \text{ZICOMP}_m^{(1)}(\phi, \boldsymbol{\lambda}, \mathbf{v})$ ou $\mathbf{Y} \sim \text{ZICOMP}^{(1)}(\phi; \lambda_1, \dots, \lambda_m; v_1, \dots, v_m)$, \mathbf{X} é um vetor base do vetor \mathbf{Y} , e “ $\stackrel{d}{=}$ ” significa que em ambos os lados da igualdade o vetor \mathbf{Y} possui a mesma distribuição.

Assim como a distribuição COM-Poisson em [Conway e Maxwell \(1962\)](#) possui seus casos especiais, a distribuição ZICOMP Tipo I também possui seus casos especiais listados a seguir:

- (i) Se $\phi \neq 0$ e $\forall v_i = 1$ elemento de \mathbf{v} , então $\mathbf{Y} \sim \text{ZIP}_m^{(1)}(\phi, \boldsymbol{\lambda})$;
- (ii) Se $\phi \neq 0$ e $(v_i = 0, \lambda_i < 1)$, $\forall \lambda_i$ elemento de $\boldsymbol{\lambda}$ e v_i elemento de \mathbf{v} , então $\mathbf{Y} \sim \text{ZIG}_m^{(1)}(\phi, \mathbf{1} - \boldsymbol{\lambda})$ ou $\mathbf{Y} \sim \text{zero-inflacionado Geométrica}_m^{(1)}(\phi, \mathbf{1} - \boldsymbol{\lambda})$;
- (iii) Se $\phi \neq 0$ e $v_i \rightarrow \infty, \forall v_i$ elemento de \mathbf{v} , então $\mathbf{Y} \sim \text{ZIBernoulli}_m^{(1)}(\phi, \boldsymbol{\lambda} \odot (\mathbf{1} + \boldsymbol{\lambda}))$, em que $\frac{\lambda_i}{1 + \lambda_i}$ é um elemento do vetor $\boldsymbol{\lambda} \odot (\mathbf{1} + \boldsymbol{\lambda})$, \odot é um operador elemento a elemento do produto de Hadamard, para detalhes veja [\(CYGANEK, 2013\)](#),
- (iv) Se $\phi = 0$, então $\{Y_i\}_{i=1}^m$ são variáveis aleatórias COM-Poisson independentes com diferentes parâmetros, isto é, $\{Y_i\}_{i=1}^m \stackrel{iid}{\sim} \text{COM-Poisson}(\lambda_i, v_i)$.

As demonstrações dos itens (i)-(iii), são apresentadas no Apêndice (A.1). E o item (iv), é obtido de forma direta ao substituir $\phi = 0$ na Equação (2.2).

2.2.1 Função de massa de probabilidade conjunta

Seja $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m)^\top \sim \text{Degenerado}(\mathbf{0}_m)$, isto é, $\{\xi_i\}_{i=1}^m \stackrel{iid}{\sim} \text{Degenerado}(0)$. Para a Equação (2.1), a função de massa de probabilidade de $\mathbf{Y} \sim \text{ZICOMP}_m^{(1)}(\phi, \boldsymbol{\lambda}, \mathbf{v})$, é dada por

$$\begin{aligned} f(\mathbf{y}|\phi, \boldsymbol{\lambda}, \mathbf{v}) &= \phi \Pr(\boldsymbol{\xi} = \mathbf{y}) + (1 - \phi) \Pr(\mathbf{X} = \mathbf{y}) \\ &= \left[\phi + (1 - \phi) \prod_{i=1}^m \frac{1}{S(\lambda_i, v_i)} \right] I(\mathbf{y} = \mathbf{0}) + \left[(1 - \phi) \prod_{i=1}^m \frac{\lambda_i^{y_i}}{S(\lambda_i, v_i)(y_i!)^{v_i}} \right] I(\mathbf{y} \neq \mathbf{0}), \end{aligned} \quad (2.2)$$

em que $I(\mathbf{y})$ é uma função indicadora do tipo $I(\mathbf{y}) = \mathbf{1}$ se $\mathbf{y} = \mathbf{0}$ e $I(\mathbf{y}) = \mathbf{0}$ caso contrário, $S(\lambda_i, v_i)$ é o valor da constante de normalização da distribuição de COM-Poisson em [Conway e Maxwell \(1962\)](#), na i -ésima dimensão, $\lambda_i > 0$ é um parâmetro de centro, $v_i \geq 0$ é um parâmetro de dispersão, com sobredispersão quando $(v_i < 1)$, subdispersão com $(v_i > 1)$ e equidispersão quando $(v_i = 1)$ ([\(SHMUELI et al., 2005\)](#); [\(BARRIGA; LOUZADA, 2014\)](#)).

Por sua vez, a função distribuição acumulada de $\mathbf{Y} \sim \text{ZICOMP}_m^{(1)}(\phi, \boldsymbol{\lambda}, \mathbf{v})$, é dada por

$$\begin{aligned} F(\mathbf{Y}|\phi, \boldsymbol{\lambda}, \mathbf{v}) &= \Pr(\mathbf{Y} \leq \mathbf{y}) \\ &= \phi + (1 - \phi) \prod_{i=1}^m \left\{ \sum_{k_i=0}^{y_i} \frac{\lambda_i^{k_i}}{S(\lambda_i, v_i)(k_i!)^{v_i}} \right\}. \end{aligned} \quad (2.3)$$

Note que $\sum_{k_i=0}^{y_i} \frac{\lambda_i^{k_i}}{S(\lambda_i, v_i)(k_i!)^{v_i}}$, depende dos parâmetros λ_i e v_i , então a forma fechada de $F(\mathbf{Y}|\phi, \boldsymbol{\lambda}, \mathbf{v})$, depende dos valores particulares dos parâmetros, pois quando $\lambda_i < 1$ e $v_i = 0$, a distribuição COM-Poisson equivale à distribuição geométrica, então para $\sum_{k_i=0}^{y_i} \frac{\lambda_i^{k_i}}{S(\lambda_i, v_i)(k_i!)^{v_i}}$ resulta na distribuição Binomial Negativa.

2.2.2 Momentos mistos e função geradora de momentos

Seja $\mathbf{Y} \sim \text{ZICOMP}_m^{(1)}(\phi, \boldsymbol{\lambda}, \mathbf{v})$, então $Y_i \stackrel{d}{=} ZX_i$ para $i = 1, \dots, m$. Logo, a função geradora de momentos de \mathbf{Y} é dada por

$$\mathbf{M}_{\mathbf{Y}}(\mathbf{t}) = \phi + (1 - \phi) \prod_{i=1}^m \frac{S(\lambda_i e^{t_i}, v_i)}{S(\lambda_i, v_i)}. \quad (2.4)$$

Dessa forma, para Y_i o i -ésimo elemento do vetor \mathbf{Y} os momentos são dados por

$$\left\{ \begin{array}{l} E(Y_i) = (1 - \phi) \lambda_i \frac{\partial \log S(\lambda_i, v_i)}{\partial \lambda_i}; \\ E(Y_i^2) = (1 - \phi) \left\{ \lambda_i^2 \frac{\partial^2 S(\lambda_i, v_i) | \partial \lambda_i^2}{S(\lambda_i, v_i)} + \lambda_i \frac{\partial \log S(\lambda_i, v_i)}{\partial \lambda_i} \right\}, \quad e \\ \text{Var}(Y_i) = (1 - \phi) \left\{ \lambda_i^2 \left[\frac{\partial^2 S(\lambda_i, v_i) | \partial \lambda_i^2}{S(\lambda_i, v_i)} - (1 - \phi) \left(\frac{\partial \log S(\lambda_i, v_i)}{\partial \lambda_i} \right)^2 \right] + \frac{\partial \log S(\lambda_i, v_i)}{\partial \lambda_i} \right\} \end{array} \right. \quad (2.5)$$

Assim, $\text{Cov}(Y_i, Y_j) = \phi(1 - \phi)\mu_i\mu_j$, onde $\mu_i = \lambda_i \frac{\partial \log S(\lambda_i, v_i)}{\partial \lambda_i}$ e $\mu_j = \lambda_j \frac{\partial \log S(\lambda_j, v_j)}{\partial \lambda_j}$ com $i \neq j$ e $i, j = 1, \dots, m$. Dessa forma, a correlação entre Y_i e Y_j , é dada por

$$\rho = \text{Corr}(Y_i, Y_j) = \frac{\phi(1 - \phi)\mu_i\mu_j}{\sqrt{\text{Var}(Y_i)\text{Var}(Y_j)}}, \quad \text{com } i \neq j \quad (2.6)$$

As demonstrações da função geradora de momentos em (2.4), momentos em (2.5) e da correlação em (2.6) são apresentadas no Apêndice (A.2).

2.2.3 Distribuição marginal

As distribuições marginais tem sua importância porque permitem a análise individual de cada variável isoladamente, mesmo que seja mandatário o ajuste dos modelos com a distribuição multivariada devido a presença das correlações das variáveis envolvidas.

Seja $\mathbf{Y} \sim \text{ZICOMP}_m^{(1)}(\phi; \lambda_1, \dots, \lambda_m; \nu_1, \dots, \nu_m)$, então o i -ésimo componente Y_i de \mathbf{Y} , é dado por

$$Y_i \stackrel{d}{=} ZX_i \sim \text{ZICOMP}(\phi, \lambda_i, \nu_i). \quad (2.7)$$

Se particionarmos $\mathbf{Y} = (\mathbf{Y}^{(1)\top}, \mathbf{Y}^{(2)\top})^\top$, com $\mathbf{Y}^{(1)} = (Y_1, \dots, Y_r)^\top$ e $\mathbf{Y}^{(2)} = (Y_{r+1}, \dots, Y_m)^\top$, temos:

$$\begin{cases} \mathbf{Y}^{(1)} \sim \text{ZICOMP}^{(1)}(\phi; \lambda_1, \dots, \lambda_r; \nu_1, \dots, \nu_r), & \text{e} \\ \mathbf{Y}^{(2)} \sim \text{ZICOMP}^{(1)}(\phi; \lambda_{r+1}, \dots, \lambda_m; \nu_{r+1}, \dots, \nu_m). \end{cases} \quad (2.8)$$

2.2.4 Distribuição condicional

As distribuições condicionais são importantes em análises preditivas de uma das variáveis quando se tem o conhecimento prévio de valores das outras variáveis.

2.2.4.1 Distribuição condicional $\mathbf{Y}^{(1)} | \mathbf{Y}^{(2)}$

Para as Equações (2.2) e (2.8), a distribuição condicional de $\mathbf{Y}^{(1)} | \mathbf{Y}^{(2)}$ é simplesmente dada por

$$\begin{aligned} \Pr(\mathbf{Y}^{(1)} = \mathbf{y}^{(1)} | \mathbf{Y}^{(2)} = \mathbf{y}^{(2)}) &= \frac{f(\mathbf{y} | \phi, \boldsymbol{\lambda}, \boldsymbol{\nu})}{\Pr(\mathbf{Y}^{(2)} = \mathbf{y}^{(2)})} \\ &= \frac{\prod_{i=1}^m \left[\phi + \frac{(1-\phi)}{S(\lambda_i, \nu_i)} \right] I_{(\mathbf{y}=\mathbf{0})} + \prod_{i=1}^m \left[\frac{(1-\phi)}{S(\lambda_i, \nu_i)} \frac{\lambda_i^{y_i}}{(y_i!)^{\nu_i}} \right] I_{(\mathbf{y} \neq \mathbf{0})}}{\prod_{i=r+1}^m \left[\phi + \frac{(1-\phi)}{S(\lambda_i, \nu_i)} \right] I_{(\mathbf{y}^{(2)}=\mathbf{0})} + \prod_{i=r+1}^m \left[\frac{(1-\phi)}{S(\lambda_i, \nu_i)} \frac{\lambda_i^{y_i}}{(y_i!)^{\nu_i}} \right] I_{(\mathbf{y}^{(2)} \neq \mathbf{0})}}. \end{aligned} \quad (2.9)$$

Para encontrar a distribuição condicional de $\mathbf{Y}^{(1)} | \mathbf{Y}^{(2)}$, consideraremos dois casos:

Caso I: $\mathbf{y}^{(2)} \neq \mathbf{0}$ e $\mathbf{y}^{(1)} \neq \mathbf{0}$, temos

$$\begin{aligned} \Pr(\mathbf{Y}^{(1)} = \mathbf{y}^{(1)} | \mathbf{Y}^{(2)} \neq \mathbf{0}) &= \frac{\Pr(\mathbf{Y}^{(1)} = \mathbf{0}, \mathbf{Y}^{(2)} \neq \mathbf{0})}{\Pr(\mathbf{Y}^{(2)} \neq \mathbf{0})} + \frac{\Pr(\mathbf{Y}^{(1)} \neq \mathbf{0}, \mathbf{Y}^{(2)} \neq \mathbf{0})}{\Pr(\mathbf{Y}^{(2)} \neq \mathbf{0})} \\ &= \frac{(1-\phi) \prod_{i=1}^m \left[\frac{\lambda_i^{y_i}}{S(\lambda_i, \nu_i) (y_i!)^{\nu_i}} \right]}{(1-\phi) \prod_{i=r+1}^m \left[\frac{\lambda_i^{y_i}}{S(\lambda_i, \nu_i) (y_i!)^{\nu_i}} \right]} = \prod_{i=1}^r \frac{\lambda_i^{y_i}}{S(\lambda_i, \nu_i) (y_i!)^{\nu_i}}. \end{aligned} \quad (2.10)$$

Isso implica que dado $\mathbf{Y}^{(1)} | \mathbf{Y}^{(2)} \stackrel{d}{=} \mathbf{X}^{(1)}$, não depende de Z , ou seja, os componentes de $\mathbf{Y}^{(1)} = (Y_1, \dots, Y_r)^\top$ são variáveis COMP(λ_i, v_i) mutuamente independentes, quando $\mathbf{y}^{(2)} \neq \mathbf{0}$.

Caso II: $\mathbf{y}^{(2)} = \mathbf{0}$, temos

$$\mathbf{Y}^{(1)} | (\mathbf{Y}^{(2)} = \mathbf{0}) \sim \text{ZICOMP}^{(I)}(\phi^*; \lambda_1, \dots, \lambda_r; v_1, \dots, v_r), \quad (2.11)$$

em que $\phi^* = \frac{\phi S^{(2)}(\boldsymbol{\lambda}, \mathbf{v})}{\phi S^{(2)}(\boldsymbol{\lambda}, \mathbf{v}) + (1-\phi)}$; $S(\boldsymbol{\lambda}, \mathbf{v}) = \prod_{i=1}^m S(\lambda_i, v_i)$; $S^{(1)}(\boldsymbol{\lambda}, \mathbf{v}) = \prod_{i=1}^r S(\lambda_i, v_i)$; $S^{(2)}(\boldsymbol{\lambda}, \mathbf{v}) = \prod_{i=r+1}^m S(\lambda_i, v_i)$. A prova de (2.11) é apresentada no Apêndice (A.3.1).

2.2.4.2 Distribuição condicional $Z | \mathbf{Y}$

Se $Z \sim \text{Bernoulli}(1 - \phi)$, e Z assume os valores 0 e 1, então

$$\begin{aligned} \Pr(Z = 1 | \mathbf{Y} = \mathbf{y}) &= \frac{\Pr(Z = 1, \mathbf{X} = \mathbf{y})}{f(\mathbf{y} | \phi, \boldsymbol{\lambda}, \mathbf{v})} = \frac{(1-\phi) \prod_{i=1}^m \frac{\lambda_i^{y_i}}{(y_i!)^{v_i} S(\lambda_i, v_i)}}{f(\mathbf{y} | \phi, \boldsymbol{\lambda}, \mathbf{v})} \\ &= \begin{cases} \frac{(1-\phi)}{\phi S(\boldsymbol{\lambda}, \mathbf{v}) + (1-\phi)}, & \mathbf{y} = \mathbf{0}, \\ 1, & \mathbf{y} \neq \mathbf{0}, \end{cases} \end{aligned} \quad (2.12)$$

então

$$Z | (\mathbf{Y} = \mathbf{y}) \sim \begin{cases} \text{Bernoulli}(\psi), & \text{se } \mathbf{y} = \mathbf{0}, \\ \text{Degenerado}(1), & \text{se } \mathbf{y} \neq \mathbf{0}, \end{cases} \quad (2.13)$$

sendo $\psi = \frac{(1-\phi)}{\phi S(\boldsymbol{\lambda}, \mathbf{v}) + (1-\phi)}$.

2.2.4.3 Distribuição condicional $\mathbf{X} | \mathbf{Y}$

Se $\mathbf{y} = \mathbf{0}$, temos

$$\mathbf{X} | (\mathbf{Y} = \mathbf{0}) \sim \text{ZICOMP}^{(I)}(\psi; \lambda_1, \dots, \lambda_m; v_1, \dots, v_m). \quad (2.14)$$

A prova de (2.14) é apresentada no Apêndice (A.3.2).

Se $\mathbf{y} \neq \mathbf{0}$, temos

$$\Pr(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}) = \frac{\Pr(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})}{\Pr(\mathbf{Y} = \mathbf{y})} = \frac{\Pr(\mathbf{X} = \mathbf{y}, Z = 1)}{f(\mathbf{y} | \phi, \boldsymbol{\lambda}, \mathbf{v})} = 1. \quad (2.15)$$

Assim, (X_1, \dots, X_m) são mutuamente independentes e se $\mathbf{Y} = \mathbf{y} \neq \mathbf{0}$ é dado então,

$$X_i | (\mathbf{Y} = \mathbf{y} \neq \mathbf{0}) \sim \text{Degenerado}(y_i), \quad i = 1, \dots, m. \quad (2.16)$$

2.2.4.4 Distribuição condicional $X_i | (Y_i = y_i = 0)$, $i=1, \dots, m$

Para a Equação (2.7), temos $Y_i \sim \text{ZICOMP}(\phi, \lambda_i, \nu_i)$, para $i = 1, \dots, m$. Assim,

$$\Pr(X_i = x_i | Y_i = 0) = \frac{\Pr(X_i = x_i, Y_i = 0)}{\Pr(Y_i = 0)},$$

então,

$$X_i | (Y_i = 0) \sim \text{ZICOMP}^{(1)}(\psi_i^*, \lambda_i, \nu_i), \quad \text{com } \psi_i^* = \frac{(1 - \phi)}{\phi S(\lambda_i, \nu_i) + (1 - \phi)}. \quad (2.17)$$

A prova de (2.17) é análoga ao item (2.14).

2.2.4.5 Distribuição condicional $X_i | (Y_i = y_i > 0)$, $i=1, \dots, m$

Para a Equação (2.7), temos $Y_i \sim \text{ZICOMP}(\phi, \lambda_i, \nu_i)$, para $i = 1, \dots, m$. Assim,

$$\Pr(X_i = x_i | Y_i = y_i) = \frac{\Pr(X_i = x_i, Y_i = y_i)}{\Pr(Y_i = y_i)} = \frac{\Pr(X_i = y_i, Z = 1)}{f(y_i | \phi, \lambda_i, \nu_i)} = 1,$$

e obtemos

$$X_i | (Y_i = y_i > 0) \sim \text{Degenerado}(y_i). \quad (2.18)$$

2.3 Modelo de regressão para a distribuição ZICOMP Tipo I

O modelo de regressão linear da distribuição ZICOMP Tipo I, que investiga a relação existente entre o vetor de parâmetros $\theta = (\phi, \boldsymbol{\lambda}^\top, \mathbf{v}^\top)^\top$ e as covariáveis ou variáveis explanatórias é dado por

$$\left\{ \begin{array}{l} \mathbf{Y}_j \stackrel{iid}{\sim} \text{ZICOMP}_m^{(1)}(\phi_j, \boldsymbol{\lambda}_j, \mathbf{v}_j), \quad j = 1, \dots, n; \\ \log\left(\frac{\phi_j}{1 - \phi_j}\right) = \mathbf{z}_j^\top \boldsymbol{\gamma}; \\ \log(\lambda_{ij}) = \mathbf{w}_j^\top \boldsymbol{\beta}_i, \quad i = 1, \dots, m; \\ \log(\nu_{ij}) = \mathbf{u}_j^\top \boldsymbol{\zeta}_i, \end{array} \right. \quad (2.19)$$

em que $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{mj})^\top$, é um vetor de observações, $\boldsymbol{\lambda}_j = (\lambda_{1j}, \dots, \lambda_{mj})^\top$, $\mathbf{v}_j = (\nu_{1j}, \dots, \nu_{mj})^\top$ são vetores de parâmetros do modelo, ϕ_j é parâmetro de inflação de zeros; $\mathbf{z}_j = (1, z_{1j}, \dots, z_{pj})^\top$; $\mathbf{w}_j = (1, w_{1j}, \dots, w_{qj})^\top$ e $\mathbf{u}_j = (1, u_{1j}, \dots, u_{rj})^\top$ são vetores de covariáveis não necessariamente idênticos com respeito à j ; $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_p)^\top$, $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1}, \dots, \beta_{iq})^\top$ e $\boldsymbol{\zeta}_i = (\zeta_{i0}, \zeta_{i1}, \dots, \zeta_{ir})^\top$ são vetores de coeficientes da regressão.

2.3.1 Estimação dos parâmetros

Seja $\{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$ uma amostra aleatória de tamanho n da distribuição m -dimensional $\text{ZICOMP}_m^{(1)}(\phi, \boldsymbol{\lambda}, \mathbf{v})$, onde $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{mj})^\top$ para $j = 1, \dots, n$. E seja $\mathbf{y}_j = (y_{1j}, \dots, y_{mj})^\top$, que consiste na realização do vetor aleatório \mathbf{Y}_j , e seja $\mathbf{Y}_{obs} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^\top$ um vetor de dados observados. E, dessa forma, considere $\mathbb{J} = \{j \mid \mathbf{y}_j = \mathbf{0}, j = 1, \dots, n\}$, com $k = \sum_{j=1}^n I(\mathbf{y}_j = \mathbf{0})$. Assim a função de verossimilhança dos dados observados do vetor de parâmetros $\boldsymbol{\theta} = (\phi, \boldsymbol{\lambda}^\top, \mathbf{v}^\top)^\top$, é dada por

$$\begin{aligned} L(\boldsymbol{\theta} | \mathbf{Y}_{obs}) &= \prod_{j \in \mathbb{J}} \left\{ \phi + (1 - \phi) \prod_{i=1}^m \left[\frac{1}{S(\lambda_i, v_i)} \right] \right\} \times \\ &\quad (1 - \phi) \prod_{j \notin \mathbb{J}} \prod_{i=1}^m \left\{ \frac{\lambda_i^{y_{ij}}}{S(\lambda_i, v_i) (y_{ij}!)^{v_i}} \right\} \\ &= \left(\left\{ \phi + (1 - \phi) \prod_{i=1}^m \left[\frac{1}{S(\lambda_i, v_i)} \right] \right\} \right)^k \times \\ &\quad (1 - \phi)^{n-k} \prod_{j=k+1}^n \prod_{i=1}^m \left\{ \left[\frac{\lambda_i^{y_{ij}}}{S(\lambda_i, v_i) (y_{ij}!)^{v_i}} \right]^{1 - I(y_{ij}=0)} \right\}, \end{aligned} \quad (2.20)$$

cujo logaritmo da função de verossimilhança dos dados observados é dado por

$$\begin{aligned} \ell(\boldsymbol{\theta} | \mathbf{Y}_{obs}) &= k \log \left(\left\{ \phi + (1 - \phi) \prod_{i=1}^m \left[\frac{1}{S(\lambda_i, v_i)} \right] \right\} \right) + (n - k) \log(1 - \phi) + \\ &\quad \sum_{j=k+1}^n \sum_{i=1}^m \left\{ [I(y_{ij} > 0)] [-\log S(\lambda_i, v_i) + y_{ij} \log(\lambda_i) - v_i \log(y_{ij}!)] \right\}. \end{aligned} \quad (2.21)$$

Para obter as estimativas do vetor de parâmetros $\boldsymbol{\theta}$, maximizamos o logaritmo da função de verossimilhança apresentada na Equação (2.21), com respeito aos parâmetros $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_i$ e \mathbf{v}_i , para $i = 1, \dots, m$. Devido à complexidade da função log-verossimilhança, que só pode ser maximizada numericamente, neste trabalho optamos pelo método Quasi-Newton com o algoritmo BFGS (veja Nocedal e Wright (2006)), implementado no *software* R (R Development Core Team (2009)) por meio da função *optim*. Na implementação do código para o modelo proposto, utilizamos, como estimativas iniciais, as estimativas do modelo Poisson para o vetor $\boldsymbol{\beta}_i$ e $\mathbf{v} = \exp(\boldsymbol{\zeta}_i = \mathbf{0}) = \mathbf{1}$ para os parâmetros da COM-Poisson em Conway e Maxwell (1962), além da estimativa do modelo Binomial para o parâmetro de inflação $\boldsymbol{\gamma}$.

2.3.2 Intervalo de confiança

Seja $\hat{\boldsymbol{\theta}}$ o estimador de máxima verossimilhança do vetor $\boldsymbol{\theta}$. O intervalo de confiança assintótico $(1 - \alpha)100\%$ tipo Wald para os parâmetros do vetor $\boldsymbol{\theta}$, é dado por

$$\left[\hat{\boldsymbol{\theta}} - z_{\alpha/2} \sqrt{[\mathbf{J}(\boldsymbol{\theta})]^{-1}}, \hat{\boldsymbol{\theta}} + z_{\alpha/2} \sqrt{[\mathbf{J}(\boldsymbol{\theta})]^{-1}} \right], \quad (2.22)$$

em que $[\mathbf{J}(\boldsymbol{\theta})]^{-1}$ é a matriz inversa da informação observada, $z_{\alpha/2}$ é o $\alpha/2$ -ésimo quantil da distribuição normal padrão (LIU; TIAN, 2015). Entretanto, o intervalo de confiança tipo Wald para os parâmetros $\phi \in [0, 1)$, $\boldsymbol{\lambda} \in \mathfrak{R}_+^{*m}$ e $\mathbf{v} \in \mathfrak{R}_+^m$ pode fornecer estimativas fora de seu espaço paramétrico. Nas situações em que os dados possuem tendência e uma forte assimetria, o intervalo de confiança *bootstrap* corrigido pelo viés acelerado (IBCa) surge como uma ótima opção de intervalo de confiança (para mais detalhes, veja (DICICCIO; EFRON, 1996); (EFRON; TIBSHIRANI, 1994)).

Considere $\mathbf{Y}_{obs} = \{\mathbf{y}_1 \dots, \mathbf{y}_n\}$ uma amostra aleatória de observações do vetor \mathbf{Y} , oriunda da distribuição acumulada F e seja \hat{F} a distribuição empírica dos dados. Considere também $\boldsymbol{\theta} = h(F)$ como o vetor de parâmetros de F , e $\hat{\boldsymbol{\theta}} = s(\mathbf{y}_1 \dots, \mathbf{y}_n)$ seu estimador. Então, ao reamostrarmos com reposição $\mathbf{Y}_{obs}^* = \{\mathbf{y}_1^* \dots, \mathbf{y}_n^*\}$ amostras de \mathbf{Y}_{obs} , de tal forma que podemos obter $\hat{\boldsymbol{\theta}}^*$ estimativa de $\hat{\boldsymbol{\theta}}$. Repetindo esse processo B vezes independentemente, obtemos B replicações de $\{\hat{\boldsymbol{\theta}}_b^*\}_{b=1}^B$. E, assim, o intervalo de confiança (IBCa) é dado por

$$\left[\hat{\boldsymbol{\theta}}_{b(\alpha_1)}^*, \hat{\boldsymbol{\theta}}_{b(\alpha_2)}^* \right],$$

em que

$$\alpha_1 = \Phi \left\{ \hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(\alpha)}{1 - \hat{a}[\hat{z}_0 + \Phi^{-1}(\alpha)]} \right\} \quad \text{e} \quad \alpha_2 = \Phi \left\{ \hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(1 - \alpha)}{1 - \hat{a}[\hat{z}_0 + \Phi^{-1}(1 - \alpha)]} \right\}, \quad (2.23)$$

em que α_1 e α_2 são os percentis corrigidos pelo *bootstrap*; $\Phi(\cdot)$ e $\Phi^{-1}(\cdot)$ correspondem à função de distribuição acumulada da normal padrão e sua inversa; \hat{z}_0 é uma constante de correção de viés, que é obtida pela proporção direta das estimativas *bootstrap* $\hat{\boldsymbol{\theta}}_b^*$ menores que $\hat{\boldsymbol{\theta}}$, isto é,

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\#\{\hat{\boldsymbol{\theta}}_b^* < \hat{\boldsymbol{\theta}}\}}{B} \right), \quad b = 1, \dots, B. \quad (2.24)$$

Já \hat{a} é uma constante de aceleração, que por sua vez ajusta o intervalo de confiança em relação à assimetria existente nos dados, e é obtida por

$$\hat{a} = \frac{\sum_{j=1}^n \left[\hat{\boldsymbol{\theta}}_{(\cdot)} - \hat{\boldsymbol{\theta}}_{(j)} \right]^3}{6 \left\{ \left[\hat{\boldsymbol{\theta}}_{(\cdot)} - \hat{\boldsymbol{\theta}}_{(j)} \right]^2 \right\}^{3/2}}, \quad (2.25)$$

em que $\hat{\boldsymbol{\theta}}_{(j)}$ é a estimativa de $\boldsymbol{\theta}$ com a j -ésima observação eliminada de \mathbf{Y}_{obs} , e $\hat{\boldsymbol{\theta}}_{(\cdot)} = \sum_{j=1}^n \hat{\boldsymbol{\theta}}_j / n$; para mais detalhes, veja ((DICICCIO; EFRON, 1996); (EFRON; TIBSHIRANI, 1994)).

2.3.3 Teste da razão de verossimilhanças para o vetor de dispersão do modelo ZICOMP Tipo I

O excesso de zeros pode provocar sobredispersão nos dados, e quando essa dispersão é provocada apenas pelo excesso de zeros o modelo ZIP pode ser indicado no ajuste desses dados. Entretanto, existem situações em que a dispersão nos dados é maior que a provocada pelo excesso de zeros. Suponha que estamos interessados em testar qual dos modelos ZIP Tipo I ou ZICOMP Tipo I se ajusta melhor a um conjunto de dados multivariados. Uma alternativa para esse problema é estender o teste da razão de verossimilhanças proposto em [Sellers e Raim \(2016\)](#), para o caso multivariado. Sob a hipótese nula H_0 , testa-se a equidispersão com $\mathbf{v} = \mathbf{1}$. E sob a hipótese alternativa H_1 , testa-se $\mathbf{v} \neq \mathbf{1}$, isto é, testa-se a subdispersão ou sobredispersão dos dados. Se a hipótese nula for rejeitada, neste caso, o modelo ZICOMP Tipo I é o mais apropriado para explicar os dados. Assim, a estatística do teste da razão de verossimilhanças é escrita na seguinte forma

$$T_{\mathbf{v}} = -2\{\ell(\hat{\phi}, \hat{\lambda}, \hat{\mathbf{v}} = \mathbf{1}; \mathbf{Y}) - \ell(\hat{\phi}, \hat{\lambda}, \hat{\mathbf{v}}; \mathbf{Y})\}, \quad (2.26)$$

em que $\ell(\hat{\phi}, \hat{\lambda}, \hat{\mathbf{v}} = \mathbf{1}; \mathbf{Y})$ e $\ell(\hat{\phi}, \hat{\lambda}, \hat{\mathbf{v}}; \mathbf{Y})$ correspondem às funções de log-verossimilhanças sob as hipóteses H_0 e H_1 respectivamente, ambas avaliadas nas estimativas dos parâmetros $\hat{\phi}$, $\hat{\lambda}$ e $\hat{\mathbf{v}}$. A estatística do teste $T_{\mathbf{v}}$ possui distribuição aproximadamente qui-quadrado com r graus de liberdade χ_r^2 [Johnson, Wichern et al. \(2002, p. 188\)](#), que corresponde ao número de dimensões do vetor \mathbf{v} . Dessa forma, o p-valor $_{\mathbf{v}}$ é dado por

$$\Pr(T_{\mathbf{v}} > t_{\mathbf{v}} | H_0) = \Pr(\chi_r^2 > t_{\mathbf{v}}). \quad (2.27)$$

Se p-valor $_{\mathbf{v}} < \alpha$, rejeita-se a hipótese nula H_0 ao nível α de significância.

2.4 Estudo de simulação

Nesta Seção, apresentamos um estudo de simulação com o objetivo de avaliar o comportamento dos estimadores de máxima verossimilhança (EMV) em relação ao viés e a raiz quadrada do erro quadrático médio (REQM), para o modelo da distribuição multivariada ZICOMP Tipo I com e sem covariáveis. Também conduzimos um estudo de simulação Monte Carlo com o objetivo de verificar a funcionalidade do teste de hipótese da razão de verossimilhanças proposto na Seção (2.3.3). Toda a análise foi realizada pelo *software* R.

2.4.1 Ajuste do modelo sem covariáveis

Para o ajuste do modelo sem covariáveis consideramos apenas o caso trivariado com sobredispersão, ou seja, geramos Y_1 , Y_2 e Y_3 amostras de tamanhos ($n=100$ e 500) da distribuição

proposta, por meio da marginal $ZICOMP_i^{(1)}(\phi, \lambda_i, v_i)$ em (2.7), em que a mesma pode ser obtida por meio dos *quantis* da distribuição normal padrão multivariada, veja (YAHAV; SHMUELI, 2012). Para o vetor (Y_1, Y_2, Y_3) , consideramos o vetor de parâmetros $\theta = (\phi; \lambda_1, \lambda_2, \lambda_3; v_1, v_2, v_3)$ com os valores (0.20;3.00,3.60,4.00;0.74,0.76,0.78). Para avaliar a qualidade pontual dos EMV, com relação ao viés e a REQM, realizamos 5000 réplicas *bootstrap* não paramétrico para obter as estimativas do viés, REQM e o intervalo de confiança *bootstrap* corrigido acelerado IBCa(95%). A Tabela 2 apresenta os valores do vetor de parâmetros θ , juntamente com os EMV, viés, REQM e os intervalos com 95% de confiança assintóticos ICa(95%) e *bootstrap* IBCa(95%) para o modelo da distribuição multivariada ZICOMP Tipo I.

Tabela 2 – Resultados da simulação para o modelo da distribuição ZICOMP Tipo I sem covariáveis.

n	θ	Valor	EMV	Viés	REQM	ICa(95%)	IBCa(95%)
100	ϕ	0,20	0,21	0,01	0,04	(0,20; 0,22)	(0,12; 0,28)
	λ_1	3,00	2,97	0,24	0,90	(-1,03; 6,98)	(1,93; 4,47)
	λ_2	3,60	3,49	0,19	1,07	(-2,34; 9,32)	(2,34; 5,60)
	λ_3	4,00	3,90	-0,21	1,07	(-3,77; 11,56)	(2,56; 6,55)
	v_1	0,74	0,70	-0,01	0,14	(0,48; 0,92)	(0,46; 0,96)
	v_2	0,76	0,71	-0,02	0,13	(0,53; 0,89)	(0,50; 0,97)
	v_3	0,78	0,74	-0,01	0,13	(0,55; 0,93)	(0,52; 0,99)
500	ϕ	0,20	0,20	0,00	0,02	(0,20; 0,21)	(0,17; 0,24)
	λ_1	3,00	2,94	-0,01	0,33	(1,18; 4,71)	(2,39; 3,63)
	λ_2	3,60	3,53	0,00	0,42	(0,82; 6,63)	(2,75; 4,31)
	λ_3	4,00	3,94	0,00	0,49	(0,41; 7,46)	(3,19; 5,10)
	v_1	0,74	0,72	-0,01	0,07	(0,63; 0,80)	(0,61; 0,86)
	v_2	0,76	0,74	-0,01	0,06	(0,65; 0,83)	(0,62; 0,86)
	v_3	0,78	0,76	-0,01	0,06	(0,67; 0,85)	(0,66; 0,91)

Observamos, na Tabela 2, que a sobredispersão dos dados é significativa ao nível $\alpha = 5\%$ de probabilidade, isto é, os intervalos com 95% de confiança ICa e IBCa para os parâmetros de dispersão v_1 , v_2 e v_3 não incluíram o termo um. O mesmo não ocorreu com os parâmetros de centro λ_1 , λ_2 e λ_3 no intervalo ICa que incluiu valores para λ_i fora do espaço paramétrico, pois para $\lambda_i > 0$. Quanto aos EMV, os mesmos são ligeiramente viesados para amostras de tamanho $n=100$. Entretanto, à medida que aumentamos o tamanho das amostras como, por exemplo, para $n=500$, o viés e a REQM diminuem, e os parâmetros λ_i 's tornam-se significativos ao nível $\alpha = 5\%$ de probabilidade.

2.4.2 Ajuste do modelo com covariáveis

Para o ajuste do modelo com covariáveis consideramos o caso bivariado com $m=2$ dimensões associado ao efeito da covariável x_i . Dessa forma, o modelo de regressão em (2.19) para o vetor aleatório $\mathbf{Y}_j = (Y_{1j}, Y_{2j})^\top$, associado ao vetor $\mathbf{X}_j = (1, x_{1j})^\top$, com $j = 1, \dots, n$, pode

ser escrito por

$$\left\{ \begin{array}{l} (Y_{1j}, Y_{2j})^\top \stackrel{iid}{\sim} \text{ZICOMP}_2^{(I)}(\phi_j, \lambda_{1j}, \lambda_{2j}, \nu_{1j}, \nu_{2j}), \quad j = 1, \dots, n; \\ \phi_j = \frac{\exp(\gamma_0 + \gamma_{10} x_{1j})}{(1 + \exp(\gamma_0 + \gamma_{10} x_{1j}))}; \\ \lambda_{1j} = \exp(\beta_{10} + \beta_{11} x_{1j}); \quad \lambda_{2j} = \exp(\beta_{20} + \beta_{21} x_{1j}); \\ \nu_{1j} = \exp(\zeta_{10j}), \quad \nu_{2j} = \exp(\zeta_{20j}), \end{array} \right. \quad (2.28)$$

em que x_{1j} é uma realização da variável X com distribuição normal $N(0, 0.25)$, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)^\top$, $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1})^\top$ e $\boldsymbol{\zeta} = (\zeta_{10}, \zeta_{20})^\top$ são vetores de coeficientes da regressão. Então, baseando-se nos vetores de parâmetros: $\boldsymbol{\gamma} = (-1.30, 0.80)^\top$, $\boldsymbol{\beta}_1 = (1.09, 0.30)^\top$, $\boldsymbol{\beta}_2 = (1.40, 0.35)^\top$, $\zeta_{10} = -0.23$ e $\zeta_{20} = -0.17$, geramos duas amostras Y_1 e Y_2 com os tamanhos ($n=100$ e 500) da distribuição proposta com sobredispersão, ou seja, $\nu_{1j} = \exp(\zeta_{10j}) = \exp(-0.23) \approx 0.79$ e $\nu_{2j} = \exp(\zeta_{20j}) = \exp(-0.17) \approx 0.84$. O parâmetro γ_0 do vetor de inflação está associado à proporção de zeros sem covariável, que corresponde $\exp(\gamma_0)/(1 + \exp(\gamma_0)) = \exp(-1.30)/(1 + \exp(-1.30)) \approx 0.21$. Foram realizadas 5000 réplicas *bootstrap* não paramétrico para obter as estimativas do viés, REQM e o intervalo de confiança *bootstrap* corrigido acelerado IBCa(95%). A Tabela 3 apresenta os valores do vetor de parâmetros $(\boldsymbol{\gamma}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\nu})^\top$, juntamente com os EMV, viés, REQM e os intervalos com 95% de confiança assintóticos ICa(95%) e *bootstrap* IBCa(95%), para o modelo de regressão da distribuição multivariada ZICOMP Tipo I.

Tabela 3 – Resultados da simulação para o modelo de regressão linear da distribuição multivariada ZICOMP Tipo I.

n	$\boldsymbol{\theta}$	Valor	EMV	Viés	REQM	ICa(95%)	IBCa(95%)
100	γ_0	-1,30	-1,24	0,03	0,26	(-1,66; -0,72)	(-1,83; -0,81)
	γ_1	0,80	0,76	-0,02	0,57	(-0,31; 1,93)	(-0,32; 1,89)
	β_{10}	1,09	1,10	0,07	0,26	(0,58; 1,52)	(0,63; 1,55)
	β_{11}	0,30	0,34	0,06	0,15	(0,10; 0,58)	(0,07; 0,59)
	β_{20}	1,40	1,46	0,14	0,31	(0,83; 1,80)	(0,96; 1,96)
	β_{21}	0,35	0,44	0,11	0,17	(0,19; 0,63)	(0,21; 0,69)
	ν_1	0,79	0,77	0,02	0,15	(0,55; 1,00)	(0,58; 0,96)
	ν_2	0,84	0,81	0,02	0,14	(0,58; 1,04)	(0,61; 1,01)
	500	γ_0	-1,30	-1,30	-0,01	0,11	(-1,42; -0,98)
γ_1		0,80	0,80	0,00	0,22	(0,53; 1,37)	(0,36; 1,24)
β_{10}		1,09	1,10	0,02	0,11	(0,88; 1,32)	(0,89; 1,30)
β_{11}		0,30	0,33	0,01	0,05	(0,27; 0,47)	(0,21; 0,42)
β_{20}		1,40	1,43	0,04	0,12	(1,16; 1,64)	(1,21; 1,66)
β_{21}		0,35	0,38	0,03	0,06	(0,31; 0,49)	(0,28; 0,47)
ν_1		0,79	0,82	0,04	0,07	(0,70; 0,93)	(0,73; 0,91)
ν_2		0,84	0,81	-0,01	0,06	(0,71; 0,92)	(0,73; 0,90)

Podemos observar na Tabela 3 que, para as amostras de tamanho $n=100$, os EMV são ligeiramente viesados e os coeficientes da regressão γ_{10} para a inflação, v_1 e v_2 os parâmetros de dispersão, são não significativos ao nível de 5% de probabilidade, ou seja, para γ_{10} seus intervalos de confiança ICA e IBCa incluíram o termo zero. Já para os parâmetros v_1 e v_2 , seus intervalos de confiança como, por exemplo, o ICA incluíram valores maiores que um, isto é, a sobredispersão é não significativa. Entretanto, à medida que aumentamos o tamanho das amostras, como por exemplo, para $n=500$, o viés e a REQM diminuem, e os parâmetros γ_1 , v_1 e v_2 tornam-se significativos ao nível de 5% de probabilidade.

2.4.3 Teste de hipóteses

Para ilustrar o teste da razão de verossimilhanças proposto no item (2.3.3), geramos 100 amostras da distribuição proposta para o caso trivariado, considerando apenas os cenários de equidispersão e sobredispersão com os valores (0.20;3.00,3.50,4.00;1.00,1.00,1.00) e (0.20;3.00,3.20,4.00;0.75,0.80,0.85). As hipóteses a serem testadas pelo teste da razão de verossimilhanças são $H_0 : (v_1, v_2, v_3) = (1, 1, 1)$ versus $H_1 : (v_1, v_2, v_3) \neq (1, 1, 1)$. Note que testar a hipótese H_0 , equivale a testar a equidispersão do vetor $(v_1, v_2, v_3) = (1, 1, 1)$ quando o modelo ZICOMP Tipo I corresponde ao modelo ZIP Tipo I. Na Tabela 4, apresentamos as estatísticas do teste da razão de verossimilhanças com $r=3$ graus de liberdade, juntamente com seus respectivos p-valores.

Tabela 4 – Resultado do teste da razão de verossimilhanças.

Dispersão	T_v	p-valor _v
Equidispersão	0,0600	0,9960
Sobredispersão	13,0200	0,0046

Podemos observar, na Tabela 4 que a hipótese nula H_0 foi rejeitada para a sobredispersão. Entretanto, a hipótese H_0 não foi rejeitada para a equidispersão, quando o modelo ZICOMP Tipo I equivale ao modelo ZIP Tipo I. Isso demonstra a flexibilidade da distribuição ZICOMP Tipo I, ao considerar as distribuições que são seus casos particulares, como a ZIP Tipo I.

2.5 Aplicação com dados reais

Nesta seção apresentamos uma aplicação com dados reais, que corresponde ao número de casos confirmados de AIDS, hepatite viral e sífilis congênita em uma amostra com 98 cidades do interior do estado de Alagoas-Brasil. Os dados foram coletados pelo autor no anuário estatístico do ano de 2014, disponível em (ALAGOAS, 2014). A Figura 1 apresenta o mapa político do Brasil com a localização do estado de Alagoas.



Figura 1 – Mapa Político do Brasil com o Estado de Alagoas.

Trabalhos envolvendo dados multivariados com DST são encontrados em ((QIN *et al.*, 2016); (CHA *et al.*, 2017); (RUSSELL *et al.*, 2019); (LI *et al.*, 2010); (LUCAR *et al.*, 2018)). Geralmente esses tipos de dados estão associados às características sócio-demográficas de regiões em crescente desenvolvimento, como por exemplo, em Tan *et al.* (2011) que analisaram a distribuição espacial da sífilis e da gonorréia na província de Guangdong na China. Na Tabela 5, apresentamos as variáveis Y_1 , Y_2 e Y_3 , como o número de casos confirmados de AIDS, hepatite viral e sífilis congênita, f_i representa a frequência das observações do vetor (Y_1, Y_2, Y_3) .

Tabela 5 – Frequências dos casos confirmados de AIDS, hepatite e sífilis congênita no estado de Alagoas-Brasil disponível em (ALAGOAS, 2014).

(Y_1, Y_2, Y_3)	f_i	(Y_1, Y_2, Y_3)	f_i	(Y_1, Y_2, Y_3)	f_i	(Y_1, Y_2, Y_3)	f_i
(0,0,0)	22	(0,0,3)	1	(0,4,0)	1	(8,5,5)	1
(0,1,0)	5	(0,3,0)	1	(0,11,0)	1	(1,1,2)	1
(0,0,1)	5	(0,1,4)	2	(0,11,1)	1	(1,4,12)	1
(1,0,0)	3	(3,6,1)	1	(0,8,1)	1	(1,2,5)	1
(1,1,0)	2	(3,6,2)	1	(5,0,8)	1	(1,0,2)	1
(0,1,1)	2	(0,2,0)	3	(2,5,1)	1	(6,1,6)	1
(0,2,1)	2	(2,2,3)	1	(0,7,0)	1	(0,2,0)	1
(0,5,1)	2	(5,0,1)	1	(5,2,2)	1	(3,3,5)	2
(1,0,1)	1	(2,1,1)	1	(1,1,3)	1	(1,0,3)	2
(0,6,1)	1	(0,13,0)	1	(0,8,0)	1	(2,1,0)	1
(4,2,5)	1	(8,4,10)	1	(3,2,0)	1	(3,1,0)	1
(2,0,2)	2	(0,0,2)	2	(3,8,3)	1	(0,3,1)	1
(0,1,5)	1	(2,1,5)	1	(4,1,0)	1	(1,1,1)	1
(3,2,6)	1	(2,4,3)	1	(1,3,2)	1	(1,3,4)	1

De acordo com a Tabela 5, podemos perceber que as frequências das células $(0,0,0)$, $(Y_1, 0, 0)$, $(0, Y_2, 0)$ e $(0, 0, Y_3)$ são relativamente maiores quando comparadas com as outras

células. Nesses termos, é razoável considerar o excesso de zeros neste conjunto de dados. Em uma análise descritiva, as variáveis Y_1 , Y_2 e Y_3 apresentaram média amostral: $\bar{y}_1 = 1,03$; $\bar{y}_2 = 1,86$ e $\bar{y}_3 = 1,52$; variância amostral: $s_1^2 = 2,96$; $s_2^2 = 7,28$ e $s_3^2 = 5,12$, e aparentemente são sobredispersas, pois suas variâncias são maiores que suas médias.

2.5.1 Ajuste do modelo sem covariável

Para explicar a relação existente entre as variáveis Y_1 , Y_2 e Y_3 , consideramos os ajustes dos modelos multivariados ZIP Tipo I e ZICOMP Tipo I, ou seja, para o vetor de observações $\mathbf{Y}_j = (Y_{1j}, Y_{2j}, Y_{3j})^\top$, com $j = 1, \dots, 98$, ajustamos os modelos $\text{ZIP}_3^{(I)}(\phi; \lambda_1, \lambda_2, \lambda_3)$ e $\text{ZICOMP}_3^{(I)}(\phi; \lambda_1, \lambda_2, \lambda_3; \nu_1, \nu_2, \nu_3)$. Na Tabela 6, são apresentados os EMV para os parâmetros dos modelos $\text{ZIP}_3^{(I)}(\phi; \lambda_1, \lambda_2, \lambda_3)$ e $\text{ZICOMP}_3^{(I)}(\phi; \lambda_1, \lambda_2, \lambda_3; \nu_1, \nu_2, \nu_3)$.

Tabela 6 – EMV para os modelos $\text{ZIP}_3^{(I)}(\phi; \lambda_1, \lambda_2, \lambda_3)$ e $\text{ZICOMP}_3^{(I)}(\phi; \lambda_1, \lambda_2, \lambda_3; \nu_1, \nu_2, \nu_3)$.

	$\text{ZIP}_3^{(I)}(\phi; \lambda_1, \lambda_2, \lambda_3)$	$\text{ZICOMP}_3^{(I)}(\phi; \lambda_1, \lambda_2, \lambda_3; \nu_1, \nu_2, \nu_3)$
θ	EMV	EMV
ϕ	0,22	0,19
λ_1	1,32	0,56
λ_2	2,39	0,70
λ_3	1,95	0,65
ν_1	-	$4,94 \times 10^{-7}$
ν_2	-	$6,80 \times 10^{-7}$
ν_3	-	$1,90 \times 10^{-6}$

Podemos observar, na Tabela 6, que os dois modelos apresentaram valores próximos para o parâmetro de inflação ϕ , com destaque para o modelo da distribuição ZIP Tipo I com $\hat{\phi} = 0,22$, pois $22/98 \approx 0,22$. Entretanto, o modelo ZICOMP Tipo I conseguiu captar toda a sobredispersão provocada além do excesso de zeros, pois seus parâmetros de dispersão apresentaram as estimativas $\hat{\nu}_1 \approx 0,00$, $\hat{\nu}_2 \approx 0,00$ e $\hat{\nu}_3 \approx 0,00$. Quanto aos critérios de escolha do modelo que melhor se ajusta aos dados, utilizamos, como critérios de escolha, os critérios de informação Akaike (AIC) e Bayesiano (BIC); para mais detalhes, veja (Akaike (1974); Schwarz (1978)), além do teste da razão de verossimilhanças proposto na seção (2.3.3). Na Tabela 7, apresentamos os valores dos critérios AIC, BIC e a estatística $T_{\mathbf{v}}$ com seu respectivo p-valor $_{\mathbf{v}}$.

Tabela 7 – Critério de seleção dos modelos multivariados ZIP Tipo I e ZICOMP Tipo I.

Modelo	$T_{\mathbf{v}}$	p-valor $_{\mathbf{v}}$	AIC	BIC
ZIP Tipo I	165,95	$p < 0,0001$	1113,78	1142,48
ZICOMP Tipo I	-	-	953,85	1004,04

Na Tabela 7, observamos que o modelo ZICOMP Tipo I apresentou menor valor para os critérios de informação AIC e BIC. Quanto ao teste da razão de verossimilhanças, rejeitou a hipótese para a equidispersão nos dados. Nesse caso, o modelo ZICOMP Tipo I é o mais

indicado para explicar a relação existente entre as variáveis AIDS, hepatite e sífilis congênita em cidades do interior do estado do Alagoas, segundo os critérios AIC, BIC e o teste da razão de verossimilhanças. Na Tabela 8, apresentamos os EMV e os intervalos com 95% de confiança a: ICa(95%) e IBCa(95%) com 5000 réplicas *bootstrap* para o modelo escolhido.

Tabela 8 – Estimativas de máxima verossimilhança e os intervalos de confiança ICa(95%) e IBCa(95%) para o modelo ZICOMP Tipo I.

θ	EMV	ICa(95%)	IBCa(95%)
ϕ	0,19	(0,17; 0,20)	(0,11; 0,29)
λ_1	0,56	(0,52; 0,60)	(0,48; 0,70)
λ_2	0,70	(0,66; 0,74)	(0,63; 0,85)
λ_3	0,65	(0,61; 0,69)	(0,53; 0,74)
v_1	$4,94 \times 10^{-7}$	$(4,94 \times 10^{-7}; 4,95 \times 10^{-7})$	$(1,76 \times 10^{-11}; 0,30)$
v_2	$6,80 \times 10^{-7}$	$(6,80 \times 10^{-7}; 6,81 \times 10^{-7})$	$(9,72 \times 10^{-15}; 0,14)$
v_3	$1,90 \times 10^{-6}$	$(1,90 \times 10^{-6}; 1,91 \times 10^{-6})$	$(3,16 \times 10^{-18}; 0,18)$

Observamos na Tabela 8, que todos os parâmetros do modelo são significativos ao nível $\alpha = 5\%$, e os dados são sobredispersos, pois os intervalos com 95% de confiança para os parâmetros de dispersão: v_1 , v_2 e v_3 , não incluíram valores maiores ou iguais a um. Em hipótese, outro modelo que poderia se ajustar a esse conjunto de dados seria o modelo ainda não existente da distribuição multivariada geométrica zero-inflacionado Tipo I (ZIG Tipo I), uma vez que a distribuição COM-Poisson equivale à distribuição geométrica com probabilidade $(1 - \lambda_i)$ quando $\lambda_i < 1$ e $v_i = 0$.

2.5.2 Análise das distribuições marginais

Com o objetivo de estimar a correlação existente entre as variáveis Y_1 , Y_2 e Y_3 , estimamos a correlação de Pearson e a correlação proposta pelo modelo ZICOMP Tipo I dada na Equação (2.6). A Tabela 9 apresenta as estimativas das correlações de Pearson e a do modelo para as variáveis Y_1 , Y_2 e Y_3 .

Tabela 9 – Estimativas das correlações entre as variáveis Y_1 , Y_2 e Y_3 .

Correlação	Y_1 e Y_2	Y_1 e Y_3	Y_2 e Y_3
Pearson	0,11	0,57	0,10
Modelo	0,20	0,20	0,19

Na Tabela 9, observamos que as variáveis Y_1 , Y_2 e Y_3 são correlacionadas positivamente, o que justifica a utilização do modelo. Entretanto, as estimativas do coeficiente de correlação propostas pelo modelo ZICOMP Tipo I apresentaram valores maiores em que a correlação de Pearson apresentou menor valor e vice-versa. Duas situações semelhantes a essa são encontradas no modelo ZIP Tipo I em Liu e Tian (2015) e no modelo bivariado ZIGP Tipo I em Zhang, Tian e Huang (2015), em que a estimativa da correlação de Pearson é maior que a do modelo. Isso

é uma característica comum da família das distribuições Tipo I, pois são correlacionadas na variável Z , quando condicionada em $Z = 1$. E ao estimarmos a correlação do modelo ZIP Tipo I proposta em Liu e Tian (2015), as variáveis (Y_1, Y_2) , (Y_1, Y_3) e (Y_2, Y_3) apresentaram as seguintes estimativas 0.28, 0.26 e 0.32.

De acordo com a distribuição marginal proposta na Equação (2.7), teoricamente, cada $Y_i (i = 1, 2, 3) \sim \text{ZICOMP}(\phi, \lambda_i, \nu_i)$, com parâmetro de inflação comum ϕ . No entanto, baseado nas estimativas do modelo ZICOMP univariado em Barriga e Louzada (2014), as variáveis $Y_i (i = 1, 2, 3) \sim \text{ZICOMP}(\hat{\phi}_i, \hat{\lambda}_i, \hat{\nu}_i)$, isto é, $Y_1 \sim \text{ZICOMP}(0.36, 0.67, 0.07)$, $Y_2 \sim \text{ZICOMP}(0.15, 0.69, 0.00)$ e $Y_3 \sim \text{ZICOMP}(0.18, 0.65, 0.00)$. As estimativas dos parâmetros $\hat{\phi}_2$ e $\hat{\phi}_3$ apresentaram valores próximos de $\hat{\phi}$. No entanto, o parâmetro de inflação ϕ , conseguiu captar apenas parte da inflação captada por $\hat{\phi}_1$ para o componente Y_1 do vetor $\mathbf{Y}_i (i = 1, 2, 3)$. Este conjunto de dados nos motiva a estender o modelo ZICOMP Tipo I para o caso Tipo III, considerando a estrutura do modelo da distribuição de Poisson multivariada zero-inflacionado Tipo III ou ZIP Tipo III em Tian *et al.* (2014), por ser mais flexível ao considerar parâmetros de inflação extras para cada componente Y_i do vetor \mathbf{Y} .

2.5.3 Ajuste do modelo com covariáveis

Finalmente, consideramos o modelo de regressão em (2.19), associado ao efeito de uma covariável. Consideramos apenas a relação entre as variáveis Y_1 : AIDS e Y_3 : sífilis congênita com a covariável X (IDH: Índice de Desenvolvimento Humano), pois em uma análise descritiva da correlação de Pearson das variáveis Y_1, Y_2 e Y_3 com a covariável X , obtiveram-se as seguintes correlações $\text{corr}(Y_1, X) = 0,41$, $\text{corr}(Y_2, X) = -0,02$ e $\text{corr}(Y_3, X) = 0,33$. Na Tabela 10, apresentamos: os EMV e os intervalos com 95% de confiança: ICa(95%) e IBCa(95%), com 5000 réplicas *bootstrap* não-paramétrico para o modelo de regressão da distribuição ZICOMP Tipo I.

Tabela 10 – Estimativas do modelo de regressão para o vetor (Y_1, Y_3) associado à covariável X .

θ	EMV	ICa(95%)	IBCa(95%)
ϕ	0,27	(0,27; 0,30)	(0,13; 0,39)
β_{10}	-4,55	(-7,08; -2,02)	(-7,70; -2,08)
β_{11}	7,31	(2,63; 11,99)	(3,10; 13,19)
β_{30}	-2,71	(-4,46; -0,96)	(-4,54; -1,10)
β_{31}	4,43	(1,19; 7,67)	(1,46; 7,87)
ν_1	0,19	(0,13; 0,25)	$(1,26 \times 10^{-5}; 0,46)$
ν_3	0,17	(0,12; 0,21)	$(4,19 \times 10^{-6}; 0,46)$

Observamos, na Tabela 10, que, para os coeficientes da regressão β_{11} e β_{31} , nenhum de seus intervalos de confiança incluíram o valor zero. O mesmo ocorreu com os parâmetros de dispersão ν_1 e ν_3 , pois seus intervalos não incluíram o termo um, ou seja, a sobredispersão existente nos dados é significativa. Nesse caso, todos os parâmetros do modelo são significativos

ao nível $\alpha = 5\%$, ou seja, existe uma relação entre o número de casos confirmados de AIDS e sífilis congênita com IDH, e a mesma é sobredispersa.

2.6 Conclusão

Neste Capítulo, apresentamos a distribuição multivariada de Conway-Maxwell-Poisson zero-inflacionado Tipo I (ZICOMP Tipo I); sua construção foi baseada na extensão da distribuição de Poisson multivariada zero-inflacionado Tipo I. Desenvolvemos importantes propriedades da distribuição, um modelo de regressão e um teste de razão de verossimilhanças para ajudar na escolha do melhor modelo. E concluímos que esta nova distribuição pode produzir melhores ajustes em conjuntos de dados multivariados de contagem com excesso de zeros e correlacionados.

MODELO DE REGRESSÃO COM A DISTRIBUIÇÃO MULTIVARIADA DE CONWAY-MAXWELL-POISSON ZERO-INFLACIONADO TIPO III

3.1 Introdução

Dados de contagem multivariados com excesso de zeros são comuns em diversas áreas do conhecimento, tais como, saúde (Cameron e Trivedi (2013)), indústria (Li *et al.* (1999)), colisões de carros em rodovia (Dong *et al.* (2014)), horticultura (Sun (2016)), lesões ocupacionais (Wang *et al.* (2003)), entre outras. As distribuições multivariadas de Poisson zero-inflacionado (MZIP) em Li *et al.* (1999) e a de Poisson zero-inflacionado Tipo I (ZIP Tipo I) em Liu e Tian (2015) aparecem como fortes candidatas para a análise e modelagem de dados multivariados com excesso de zeros, com destaque para a distribuição ZIP Tipo I, que possui uma estrutura mais flexível e fácil de ser implementada. Entretanto, sua estrutura de inflação considera apenas a inflação conjunta de zeros dos componentes do vetor aleatório de contagem, e as demais componentes do vetor aleatório são consideradas como variáveis de Poisson.

Nesse contexto, Tian *et al.* (2014) propuseram uma distribuição mais flexível, nomeada de distribuição multivariada de Poisson zero-inflacionado Tipo III (ZIP Tipo III). Ela generaliza a distribuição ZIP Tipo I em Liu e Tian (2015) com a adição de outras componentes de inflação de zeros no vetor aleatório de contagem. Nas situações em que a sobredispersão dos dados é maior que a provocada pelo o excesso de zeros, os modelos ZIP Tipo I e ZIP Tipo III não são indicados por acomodarem somente a equidispersão. Podem ainda existir situações em que os dados exibem subdispersão, quando a variância amostral é menor que a média. Nesse contexto, a extensão do modelo da distribuição univariada de COM-Poisson zero-inflacionado (ZICOMP)

em [Barriga e Louzada \(2014\)](#), para o caso multivariado, ou a extensão da distribuição COM-Poisson bivariada proposta por [Sellers, Morris e Balakrishnan \(2016\)](#) para o zero-inflacionado, poderia contornar problemas de dispersão nos dados, uma vez que a distribuição de COM-Poisson bivariada consegue acomodar tanto equidispersão quanto sobredispersão e subdispersão. O campo multivariado da COM-Poisson é bastante escasso, e a mesma ainda não existe em nenhuma das estruturas das distribuições MZIP, ZIP Tipo I e Tipo III. Uma estrutura multivariada para a distribuição de COM-Poisson seria bastante complicada ou muito sofisticada para acomodar as distribuições de Poisson, geométrica e Bernoulli que são seus casos particulares.

Mediante ao exposto, este Capítulo teve por objetivo geral construir a distribuição multivariada COM-Poisson zero-inflacionado Tipo III (ZICOMP tipo III), considerando a estrutura da distribuição ZIP Tipo III, proposta por [Tian et al. \(2014\)](#). Especificamente desenvolvemos o modelo de regressão e o teste da razão de verossimilhanças para ajudar na escolha do melhor modelo.

Este Capítulo está organizado da seguinte maneira: na Seção 3.2, introduzimos a distribuição multivariada ZICOMP Tipo III e algumas propriedades. O modelo de regressão da distribuição ZICOMP Tipo III e o teste da razão de verossimilhanças são apresentados na Seção 3.3. Nas Seções 3.4 e 3.5, conduzimos um estudo de simulação da distribuição proposta e a análise de um conjunto de dados reais. A conclusão é apresentada na Seção 3.6.

3.2 Distribuição de COM-Poisson multivariada zero-inflacionado Tipo III

Considere o vetor $\mathbf{X} = (X_1, \dots, X_m)^\top$ com $\{X_i\}_{i=1}^m \stackrel{iid}{\sim} \text{ZICOMP}(\phi_i, \lambda_i, \nu_i)$. Dizemos que o vetor aleatório $\mathbf{Y} = (Y_1, \dots, Y_m)^\top$ possui distribuição multivariada ZICOMP Tipo III com parâmetros $\phi_0 \in [0, 1)$, $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)^\top \in [0, 1)^m$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top \in \mathfrak{R}_+^{*m}$ e $\mathbf{v} = (\nu_1, \dots, \nu_m)^\top \in \mathfrak{R}_+^m$. E escrevemos $\mathbf{Y} \sim \text{ZICOMP}^{(III)}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v})$, se

$$\mathbf{Y} \stackrel{d}{=} Z_0 \mathbf{X} = Z_0 (X_1, \dots, X_m)^\top = \begin{cases} \mathbf{0}, & \text{com probabilidade } \phi_0, \\ \mathbf{X}, & \text{com probabilidade } (1 - \phi_0), \end{cases} \quad (3.1)$$

em que $Z_0 \sim \text{Bernoulli}(1 - \phi_0)$, $Z_0 \perp \mathbf{X}$ (Z_0 independente de \mathbf{X}), \mathbf{X} é um vetor base do vetor \mathbf{Y} , e “ $\stackrel{d}{=}$ ” significa que, em ambos os lados da igualdade, o vetor \mathbf{Y} possui a mesma distribuição. Assim como a distribuição COM-Poisson em [Conway e Maxwell \(1962\)](#) possui seus casos especiais, a distribuição ZICOMP Tipo III também possui seus casos especiais:

- (i) Se $\phi_0 \neq 0$, $\boldsymbol{\phi} = \mathbf{0}$, e $\mathbf{v} = \mathbf{1}$, então $\mathbf{Y} \sim \text{ZIP}_m^{(I)}(\phi_0, \boldsymbol{\lambda})$;
- (ii) Se $\phi_0 \neq 0$ e $\boldsymbol{\phi} = \mathbf{0}$, então $\mathbf{Y} \sim \text{ZICOMP}_m^{(I)}(\phi_0, \boldsymbol{\lambda}, \mathbf{v})$;
- (iii) $\phi_0 \neq 0$, $\boldsymbol{\phi} = \mathbf{0}$, e $\forall \nu_i = 0$ elemento de \mathbf{v} e $\lambda_i < 1$ elemento de $\boldsymbol{\lambda}$, então $\mathbf{Y} \sim \text{ZIG}_m^{(I)}(\phi_0, \mathbf{1} - \boldsymbol{\lambda})$;

(iv) Se $\phi_0 \neq 0$, $\boldsymbol{\phi} = \mathbf{0}$, e $\forall v_i \rightarrow \infty$ elemento de \mathbf{v} , então $\mathbf{Y} \sim \text{ZIBernoulli}_m^{(I)}(\phi_0, \boldsymbol{\lambda} \otimes (\mathbf{1} + \boldsymbol{\lambda}))$, em que $\frac{\lambda_i}{1+\lambda_i}$ é um elemento do vetor $\boldsymbol{\lambda} \otimes (\mathbf{1} + \boldsymbol{\lambda})$;

(v) Se $\phi_0 \neq 0$, $\boldsymbol{\phi} \neq \mathbf{0}$, e $\mathbf{v} = \mathbf{1}$, então $\mathbf{Y} \sim \text{ZIP}_m^{(III)}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda})$;

(vi) Se $\phi_0 \neq 0$, $\boldsymbol{\phi} \neq \mathbf{0}$, e $\forall v_i = 0$ elemento de \mathbf{v} e $\lambda_i < 1$ elemento de $\boldsymbol{\lambda}$, então $\mathbf{Y} \sim \text{ZIG}_m^{(III)}(\phi_0, \boldsymbol{\phi}, \mathbf{1} - \boldsymbol{\lambda})$;

(vii) Se $\phi_0 \neq 0$, $\boldsymbol{\phi} \neq \mathbf{0}$, e $\forall v_i \rightarrow \infty$ elemento de \mathbf{v} , então $\mathbf{Y} \sim \text{ZIBernoulli}_m^{(III)}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda} \otimes (\mathbf{1} + \boldsymbol{\lambda}))$, em que $\frac{\lambda_i}{1+\lambda_i}$ é um elemento de $\boldsymbol{\lambda} \otimes (\mathbf{1} + \boldsymbol{\lambda})$, \otimes é um operador elemento a elemento do produto de Hadamard, para detalhes veja (CYGANNEK, 2013),

(viii) Se $\phi_0 = 0$, então $\{Y_i\}_{i=1}^m$ são variáveis aleatórias ZICOMP independentes com diferentes parâmetros, isto é, $\{Y_i\}_{i=1}^m \stackrel{iid}{\sim} \text{ZICOMP}(\phi_i, \lambda_i, v_i)$.

As demonstrações dos itens (i) - (vii) são apresentadas no Apêndice (B.1). E o item (viii), é obtido de forma direta ao substituir $\phi_0 = 0$ na Equação (3.2).

3.2.1 Função de massa de probabilidade conjunta

Seja $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m)^\top \sim \text{Degenerado}(\mathbf{0}_m)$, isto é, $\{\xi_i\}_{i=1}^m \stackrel{iid}{\sim} \text{Degenerado}(0)$. Então para (3.1), a função de massa de probabilidade conjunta de $\mathbf{Y} \sim \text{ZICOMP}_m^{(III)}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v})$, é dada por

$$\begin{aligned} f(\mathbf{y}|\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v}) &= \phi_0 \Pr(\boldsymbol{\xi} = \mathbf{y}) + (1 - \phi_0) \Pr(\mathbf{X} = \mathbf{y}) \\ &= \prod_{i=1}^m \left\{ \phi_0 + (1 - \phi_0) \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, v_i)} \right] \right\} I(\mathbf{y} = \mathbf{0}) + (1 - \phi_0) \\ &\quad \times \prod_{i=1}^m \left\{ \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, v_i)} \right] I_{(y_i=0)} + \left[\frac{(1 - \phi_i)}{S(\lambda_i, v_i)} \frac{\lambda_i^{y_i}}{(y_i!)^{v_i}} \right] I_{(y_i>0)} \right\} I(\mathbf{y} \neq \mathbf{0}), \end{aligned} \quad (3.2)$$

em que $I(\mathbf{y})$ é uma função indicadora do tipo $I(\mathbf{y}) = \mathbf{1}$ se $\mathbf{y} = \mathbf{0}$ e $I(\mathbf{y}) = \mathbf{0}$ caso contrário, $S(\lambda_i, v_i)$ é o valor da constante de normalização da distribuição de COM-Poisson em Conway e Maxwell (1962) na i -ésima dimensão, $\lambda_i > 0$ é um parâmetro de centro, $v_i \geq 0$ é um parâmetro de dispersão, com sobredispersão quando ($v_i < 1$), subdispersão com ($v_i > 1$) e equidispersão quando ($v_i = 1$) ((SHMUELI *et al.*, 2005); (BARRIGA; LOUZADA, 2014)).

Por sua vez, a função de distribuição acumulada $\mathbf{Y} \sim \text{ZICOMP}_m^{(III)}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v})$ é dada por

$$\begin{aligned} F(\mathbf{Y}|\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v}) &= \Pr(\mathbf{Y} \leq \mathbf{y}) \\ &= \phi_0 + (1 - \phi_0) \prod_{i=1}^m \left\{ \phi_i + (1 - \phi_i) \sum_{k_i=0}^{y_i} \frac{\lambda_i^{k_i}}{S(\lambda_i, v_i)(k_i!)^{v_i}} \right\}. \end{aligned} \quad (3.3)$$

A forma fechada de $F(\mathbf{Y}|\phi, \boldsymbol{\lambda}, \mathbf{v})$ depende de valores particulares dos parâmetros λ_i e v_i , pois quando $v_i = 1$, $\sum_{k_i=0}^{y_i} \frac{\lambda_i^{k_i}}{S(\lambda_i, v_i)(k_i!)^{v_i}}$ equivale à soma de distribuições de Poisson, e quando $\lambda_i < 1$ e $v_i = 0$ essa soma equivale a distribuição Binomial Negativa.

3.2.2 Momentos mistos e função geradora de momentos

Seja $\mathbf{Y} \sim \text{ZICOMP}_m^{(\text{III})}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v})$, com $Y_i \stackrel{d}{=} Z_0 Z_i X_i$ para $i = 1, \dots, m$. Então a função geradora de momentos de \mathbf{Y} é dada por

$$M_{\mathbf{Y}}(\mathbf{t}) = \phi_0 + (1 - \phi_0) \prod_{i=1}^m \left\{ \phi_i + (1 - \phi_i) \frac{S(\lambda_i e^{t_i}, \mathbf{v}_i)}{S(\lambda_i, \mathbf{v}_i)} \right\}. \quad (3.4)$$

Dessa forma, para o i -ésimo elemento do vetor \mathbf{Y} os momentos são dados por

$$\left\{ \begin{array}{l} E(Y_i) = (1 - \phi_0)(1 - \phi_i) \lambda_i \frac{\partial \log S(\lambda_i, \mathbf{v}_i)}{\partial \lambda_i}; \\ E(Y_i^2) = (1 - \phi_0) \left\{ (1 - \phi_i) \left[\lambda_i^2 \frac{\partial^2 S(\lambda_i, \mathbf{v}_i)}{S(\lambda_i, \mathbf{v}_i)} \frac{\partial \lambda_i^2}{\partial \lambda_i} + \lambda_i \frac{\partial \log S(\lambda_i, \mathbf{v}_i)}{\partial \lambda_i} \right] \right\}, \text{ e} \\ \text{Var}(Y_i) = (1 - \phi_0) \left\{ (1 - \phi_i) \left\{ \lambda_i^2 \left[\frac{\partial^2 S(\lambda_i, \mathbf{v}_i)}{S(\lambda_i, \mathbf{v}_i)} \frac{\partial \lambda_i^2}{\partial \lambda_i} - (1 - \phi_0)(1 - \phi_i) \times \right. \right. \right. \\ \left. \left. \left. \left(\frac{\partial \log S(\lambda_i, \mathbf{v}_i)}{\partial \lambda_i} \right)^2 \right] + \frac{\partial \log S(\lambda_i, \mathbf{v}_i)}{\partial \lambda_i} \right\} \right\}. \end{array} \right. \quad (3.5)$$

Assim, $\text{Cov}(Y_i, Y_j) = \phi_0(1 - \phi_0) \mu_i \mu_j$, em que $\mu_i = (1 - \phi_0)(1 - \phi_i) \lambda_i \frac{\partial \log S(\lambda_i, \mathbf{v}_i)}{\partial \lambda_i}$ e $\mu_j = (1 - \phi_0)(1 - \phi_j) \lambda_j \frac{\partial \log S(\lambda_j, \mathbf{v}_j)}{\partial \lambda_j}$, com $i \neq j$ e $i, j = 1, \dots, m$. Dessa forma, a correlação entre Y_i e Y_j , é dada por

$$\rho = \text{Corr}(Y_i, Y_j) = \frac{\phi_0(1 - \phi_0) \mu_i \mu_j}{\sqrt{\text{Var}(Y_i) \text{Var}(Y_j)}}, \quad \text{com } i \neq j. \quad (3.6)$$

As demonstrações da função geradora de momentos em (3.4), momentos em (3.5) e da correlação em (3.6) são apresentadas no Apêndice (B.2).

3.2.3 Distribuição marginal

Se $\mathbf{Y} \sim \text{ZICOMP}_m^{(\text{III})}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v})$, então para (3.2), o i -ésimo componente Y_i de \mathbf{Y} , é dado por

$$Y_i \stackrel{d}{=} Z_0 Z_i X_i \sim \text{ZICOMP}_i^{(\text{III})}(\phi_0 + \phi_i - \phi_0 \phi_i, \lambda_i, \mathbf{v}_i). \quad (3.7)$$

Particionando $\mathbf{Y} = (\mathbf{Y}^{(1)\top}, \mathbf{Y}^{(2)\top})^\top$, com $\mathbf{Y}^{(1)} = (Y_1, \dots, Y_r)^\top$ e $\mathbf{Y}^{(2)} = (Y_{r+1}, \dots, Y_m)^\top$, temos:

$$\begin{cases} \mathbf{Y}^{(1)} \sim \text{ZICOMP}^{(\text{III})}(\phi_0; \phi_1, \dots, \phi_r; \lambda_1, \dots, \lambda_r; \nu_1, \dots, \nu_r), & \text{e} \\ \mathbf{Y}^{(2)} \sim \text{ZICOMP}^{(\text{III})}(\phi_0; \phi_{r+1}, \dots, \phi_m; \lambda_{r+1}, \dots, \lambda_m; \nu_{r+1}, \dots, \nu_m). \end{cases} \quad (3.8)$$

3.2.4 Distribuição condicional

Considere \mathbf{Y} um vetor aleatório de eventos do tipo $\mathbf{Y} = (\mathbf{Y}^{(1)\top}, \mathbf{Y}^{(2)\top})^\top$ e seja $\mathbf{Y}^{(1)}$ e $\mathbf{Y}^{(2)}$ subvetores de \mathbf{Y} de eventos como, por exemplo, pessoas que residem em uma determinada região e são portadoras das doenças C e D. Seleciona-se uma amostra aleatória e deseja-se obter a probabilidade de os indivíduos possuírem a doença C dado que eles não possuem a doença D, para obter essa probabilidade basta calcular a probabilidade condicional de $\mathbf{Y}^{(1)}$ dado $\mathbf{Y}^{(2)}$.

3.2.4.1 Distribuição condicional $\mathbf{Y}^{(1)} | \mathbf{Y}^{(2)}$

Para as Equações (3.2) e (3.8), a distribuição condicional de $\mathbf{Y}^{(1)} | \mathbf{Y}^{(2)}$ é simplesmente dada por

$$\Pr(\mathbf{Y}^{(1)} = \mathbf{y}^{(1)} | \mathbf{Y}^{(2)} = \mathbf{y}^{(2)}) = \frac{f(\mathbf{y} | \phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\nu})}{\Pr(\mathbf{Y}^{(2)} = \mathbf{y}^{(2)})}. \quad (3.9)$$

Para encontrar a distribuição condicional $\mathbf{Y}^{(1)} | \mathbf{Y}^{(2)}$, consideraremos dois casos:

Caso I: Quando $\mathbf{y}^{(1)} \neq \mathbf{0}$ e $\mathbf{y}^{(2)} \neq \mathbf{0}$, temos

$$\begin{aligned} \Pr(\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \neq \mathbf{0}) &= \frac{\Pr(\mathbf{Y}^{(1)} = \mathbf{y}^{(1)}, \mathbf{Y}^{(2)} \neq \mathbf{0})}{\Pr(\mathbf{Y}^{(2)} \neq \mathbf{0})} \\ &= \frac{(1 - \phi_0) \prod_{i=1}^m \left\{ \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \right] I_{(y_i=0)} + \left[\frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \frac{\lambda_i^{y_i}}{(y_i!)^{\nu_i}} \right] I_{(y_i>0)} \right\}}{(1 - \phi_0) \prod_{i=r+1}^m \left\{ \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \right] I_{(y_i=0)} + \left[\frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \frac{\lambda_i^{y_i}}{(y_i!)^{\nu_i}} \right] I_{(y_i>0)} \right\}} \\ &= \prod_{i=1}^r \left\{ \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \right] I_{(y_i=0)} + \left[\frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \frac{\lambda_i^{y_i}}{(y_i!)^{\nu_i}} \right] I_{(y_i>0)} \right\}. \end{aligned} \quad (3.10)$$

Isso implica que dado $\mathbf{Y}^{(1)} | \mathbf{Y}^{(2)} \stackrel{d}{=} \mathbf{X}^{(1)}$ não depende de $Z_0 = 1$, ou seja, os componentes de $\mathbf{Y}^{(1)} = (Y_1, \dots, Y_r)^\top$ são variáveis ZICOMP(ϕ_i, λ_i, ν_i) mutuamente independentes, quando $\mathbf{y}^{(2)} \neq \mathbf{0}$.

Caso II: $\mathbf{y}^{(2)} = \mathbf{0}$, temos

$$\mathbf{Y}^{(1)} | (\mathbf{Y}^{(2)} = \mathbf{0}) \sim \text{ZICOMP}^{(\text{III})}\left(\frac{\phi_0}{\omega}; \phi_1, \dots, \phi_r; \lambda_1, \dots, \lambda_r; \nu_1, \dots, \nu_r\right), \quad (3.11)$$

em que $\omega = \phi_0 + (1 - \phi_0) \prod_{i=r+1}^m \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \right]$. A prova do item (3.11) é apresentada no Apêndice (B.3.1).

3.2.4.2 Distribuição condicional $Z_0 | \mathbf{Y}$

Se $Z_0 \sim \text{Bernoulli}(1 - \phi_0)$ e Z_0 assume os valores 0 e 1, então

$$\begin{aligned} \Pr(Z_0 = 1 | \mathbf{Y} = \mathbf{y}) &= \frac{\Pr(Z_0=1, \mathbf{X}=\mathbf{y})}{f(\mathbf{y} | \phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v})} = \frac{(1-\phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, v_i)} \right]}{f(\mathbf{y} | \phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v})} \\ &= \begin{cases} \frac{(1-\phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, v_i)} \right]}{\phi_0 + (1-\phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, v_i)} \right]}, & \text{se } \mathbf{y} = \mathbf{0}, \\ 1, & \text{se } \mathbf{y} \neq \mathbf{0}, \end{cases} \end{aligned} \quad (3.12)$$

logo,

$$Z | (\mathbf{Y} = \mathbf{y}) \sim \begin{cases} \text{Bernoulli}(\tau), & \text{se } \mathbf{y} = \mathbf{0}, \\ \text{Degenerado}(1), & \text{se } \mathbf{y} \neq \mathbf{0}, \end{cases} \quad (3.13)$$

$$\text{em que } \tau = \frac{(1-\phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, v_i)} \right]}{\phi_0 + (1-\phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, v_i)} \right]}.$$

 3.2.4.3 Distribuição condicional $\mathbf{X} | \mathbf{Y}$

Se $\mathbf{y} = \mathbf{0}$, temos

$$\mathbf{X} | (\mathbf{Y} = \mathbf{0}) \sim \text{ZICOMP}_m^{(\text{III})}(\tau, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v}). \quad (3.14)$$

A prova do item (3.14) é apresentada no Apêndice (B.3.2).

Se $\mathbf{y} \neq \mathbf{0}$, temos

$$\Pr(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}) = \frac{\Pr(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})}{\Pr(\mathbf{Y} = \mathbf{y})} = \frac{\Pr(\mathbf{X} = \mathbf{y}, Z_0 = 1)}{f(\mathbf{y} | \phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v})} = 1. \quad (3.15)$$

Assim como (X_1, \dots, X_m) são mutuamente independentes e se $\mathbf{Y} = \mathbf{y} \neq \mathbf{0}$ é dado, então

$$X_i | (\mathbf{Y} = \mathbf{y} \neq \mathbf{0}) \sim \text{Degenerado}(y_i), \quad i = 1, \dots, m. \quad (3.16)$$

 3.2.4.4 Distribuição condicional $X_i | (Y_i = y_i = 0)$, $i=1, \dots, m$

Para a Equação (3.7), temos $Y_i \sim \text{ZICOMP}_i^{(\text{III})}(\phi_0, \phi_i, \lambda_i, v_i)$, $i = 1, \dots, m$. Assim,

$$\Pr(X_i = x_i | Y_i = 0) = \frac{\Pr(X_i = x_i, Y_i = 0)}{\Pr(Y_i = 0)},$$

o que implica

$$X_i | (Y_i = 0) \sim \text{ZICOMP}_i^{(\text{III})}(\tau_i, \phi_i, \lambda_i, v_i), \quad (3.17)$$

com $\tau_i = \frac{(1-\phi_0) \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, v_i)} \right]}{\phi_0 + (1-\phi_0) \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, v_i)} \right]}$. A prova do item (3.17) é análoga à prova do item (3.14).

3.2.4.5 Distribuição condicional $X_i | (Y_i = y_i > 0)$, $i=1, \dots, m$

Para a Equação (3.7), temos $Y_i \sim \text{ZICOMP}_i^{(\text{III})}(\phi_0, \phi_i, \lambda_i, \nu_i)$, $i = 1, \dots, m$. Então

$$\Pr(X_i = x_i | Y_i = y_i) = \frac{\Pr(X_i = x_i, Y_i = y_i)}{\Pr(Y_i = y_i)} = \frac{\Pr(X_i = y_i, Z_0 = 1)}{f(y_i | \phi_0, \phi_i, \lambda_i, \nu_i)} = 1,$$

logo

$$X_i | (Y_i = y_i > 0) \sim \text{Degenerado}(y_i). \quad (3.18)$$

3.3 Modelo de regressão para a distribuição ZICOMP Tipo III

O modelo de regressão linear da distribuição ZICOMP Tipo III que investiga a relação existente entre o vetor de parâmetros $(\phi_0, \boldsymbol{\phi}^\top, \boldsymbol{\lambda}^\top, \mathbf{v}^\top)^\top$ e as covariáveis é dado por

$$\left\{ \begin{array}{l} \mathbf{Y}_j \stackrel{iid}{\sim} \text{ZICOMP}_m^{(\text{III})}(\phi_{0j}, \boldsymbol{\phi}_j, \boldsymbol{\lambda}_j, \mathbf{v}_j), \quad j = 1 \dots, n; \\ \log\left(\frac{\phi_{0j}}{1-\phi_{0j}}\right) = \mathbf{z}_j^\top \boldsymbol{\gamma}_0; \quad \log\left(\frac{\phi_{ij}}{1-\phi_{ij}}\right) = \mathbf{z}_j^\top \boldsymbol{\gamma}_i, \quad i = 1 \dots, m; \\ \log(\lambda_{ij}) = \mathbf{w}_j^\top \boldsymbol{\beta}_i; \quad \text{e} \quad \log(\nu_{ij}) = \mathbf{u}_j^\top \boldsymbol{\zeta}_i, \end{array} \right. \quad (3.19)$$

em que $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{mj})^\top$ é um vetor de observações, ϕ_{0j} é o parâmetro de inflação conjunta do vetor $\mathbf{0}$, $\boldsymbol{\phi}_j = (\phi_{1j}, \dots, \phi_{mj})^\top$ é um vetor de parâmetros de inflação adicional dos componentes do vetor \mathbf{Y}_j , $\boldsymbol{\lambda}_j = (\lambda_{1j}, \dots, \lambda_{mj})^\top$ e $\mathbf{v}_j = (\nu_{1j}, \dots, \nu_{mj})^\top$ são vetores de parâmetros do modelo, $\mathbf{z}_j = (1, z_{1j}, \dots, z_{pj})^\top$, $\mathbf{w}_j = (1, w_{1j}, \dots, w_{qj})^\top$ e $\mathbf{u}_j = (1, u_{1j}, \dots, u_{rj})^\top$ são vetores de covariáveis não necessariamente idênticos com respeito à j , $\boldsymbol{\gamma}_0 = (\gamma_{00}, \gamma_{01}, \dots, \gamma_{0p})^\top$, $\boldsymbol{\gamma}_i = (\gamma_{i0}, \gamma_{i1}, \dots, \gamma_{ip})^\top$, $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1}, \dots, \beta_{iq})^\top$ e $\boldsymbol{\zeta}_i = (\zeta_{i0}, \zeta_{i1}, \dots, \zeta_{ir})^\top$, são vetores de coeficientes da regressão.

3.3.1 Estimação dos parâmetros

Seja $\{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$ uma amostra aleatória de tamanho n da distribuição m -dimensional $\text{ZICOMP}_m^{(\text{III})}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v})$, onde $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{mj})^\top$ para $j = 1, \dots, n$. E seja $\mathbf{y}_j = (y_{1j}, \dots, y_{mj})^\top$, um vetor de realizações do vetor aleatório \mathbf{Y}_j e $\mathbf{Y}_{obs} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ um vetor de dados observados. E considere $\mathbb{J} = \{j | \mathbf{y}_j = \mathbf{0}, j = 1, \dots, n\}$ com $k_0 = \sum_{j=1}^n I(\mathbf{y}_j = \mathbf{0})$. Desta forma, a função de verossimilhança dos dados observados \mathbf{Y}_{obs} do vetor $\boldsymbol{\theta} = (\phi_0, \boldsymbol{\phi}^\top, \boldsymbol{\lambda}^\top, \mathbf{v}^\top)^\top$, é dada por

$$\begin{aligned}
 L(\boldsymbol{\theta}|\mathbf{Y}_{obs}) &= \prod_{j \in \mathbb{J}} \left\{ \prod_{j=1}^n \left\{ \phi_0 + (1 - \phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right] \right\} \right\} \times \\
 &\quad \prod_{j \neq \mathbb{J}} \left\{ (1 - \phi_0) \prod_{i=1}^m \left\{ \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right]^{I_{(y_{ij}=0)}} \left[\frac{(1 - \phi_i)}{S(\lambda_i, \mathbf{v}_i)} \frac{\lambda_i^{y_{ij}}}{(y_{ij}!)^{v_i}} \right]^{I_{(y_{ij}>0)}} \right\} \right\} \\
 &= \left\{ \phi_0 + (1 - \phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right] \right\}^{k_0} \times \\
 &\quad (1 - \phi_0)^{n-k_0} \prod_{j=k_0+1}^n \prod_{i=1}^m \left\{ \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right]^{I_{(y_{ij}=0)}} \left[\frac{(1 - \phi_i)}{S(\lambda_i, \mathbf{v}_i)} \frac{\lambda_i^{y_{ij}}}{(y_{ij}!)^{v_i}} \right]^{I_{(y_{ij}>0)}} \right\}, \quad (3.20)
 \end{aligned}$$

cujo logaritmo da função de verossimilhança dos dados observados é dado por

$$\begin{aligned}
 \ell(\boldsymbol{\theta}|\mathbf{Y}_{obs}) &= k_0 \log \left\{ \phi_0 + (1 - \phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right] \right\} + (n - k_0) \log(1 - \phi_0) + \\
 &\quad \sum_{j=k_0+1}^n \sum_{i=1}^m \left\{ I_{(y_{ij}=0)} \log \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right] \right\} + \sum_{j=k_0+1}^n \sum_{i=1}^m \left\{ [I_{(y_{ij}>0)}] \times \right. \\
 &\quad \left. [\log(1 - \phi_i) - \log S(\lambda_i, \mathbf{v}_i) + y_{ij} \log(\lambda_i) - v_i \log(y_{ij}!)] \right\}. \quad (3.21)
 \end{aligned}$$

Para a obtenção das estimativas do vetor de parâmetros $\boldsymbol{\theta}$, maximizamos o logaritmo da função de máxima verossimilhança em (3.21), com respeito aos parâmetros $\boldsymbol{\gamma}_0$; $\boldsymbol{\gamma}_i$, $\boldsymbol{\beta}_i$ e $\boldsymbol{\zeta}_i$. Devido à complexidade da função de log-verossimilhança que só pode ser feita numericamente neste trabalho, optamos pelo método Quasi-Newton com o algoritmo BFGS (veja Nocedal e Wright (2006)) implementado no *software* R (R Development Core Team (2009)) por meio da função *optim*. Na implementação do código para o modelo proposto, utilizamos como estimativas iniciais, as estimativas do modelo Poisson para o vetor $\boldsymbol{\beta}_i$ e $\mathbf{v} = \exp(\boldsymbol{\zeta}_i = \mathbf{0}) = \mathbf{1}$ para os parâmetros da COM-Poisson em Conway e Maxwell (1962), além das estimativas do modelo binomial para os parâmetros de inflação $\boldsymbol{\gamma}_0$ e $\boldsymbol{\gamma}_i$.

3.3.2 Intervalo de confiança

Seja $\hat{\boldsymbol{\theta}}$ o EMV do vetor $\boldsymbol{\theta}$. O intervalo de confiança assintótico $(1 - \alpha)100\%$ para os parâmetros do vetor $\boldsymbol{\theta}$, é dado por

$$\left[\hat{\boldsymbol{\theta}} - z_{\alpha/2} \sqrt{[\mathbf{J}(\boldsymbol{\theta})]^{-1}}, \hat{\boldsymbol{\theta}} + z_{\alpha/2} \sqrt{[\mathbf{J}(\boldsymbol{\theta})]^{-1}} \right], \quad (3.22)$$

em que $[\mathbf{J}(\boldsymbol{\theta})]^{-1}$ é a matriz inversa da informação observada, $z_{\alpha/2}$ é α -ésimo quantil da distribuição normal padrão. Entretanto, o intervalo de confiança assintótico para os parâmetros do modelo ZICOMP Tipo III $\phi_0 \in (0, 1)$; $\boldsymbol{\phi} \in [0, 1]^m$; $\boldsymbol{\lambda} \in \mathfrak{R}_+^{*m}$ e $\mathbf{v} \in \mathfrak{R}_+^m$ pode fornecer estimativas fora do espaço paramétrico. Dessa forma, o intervalo de confiança bootstrap IBCa, proposto na Seção 2.3.2 do Capítulo 2, pode ser utilizado para contornar esses tipos de problemas nos intervalos de confiança.

3.3.3 Teste da razão de verossimilhanças para o vetor de dispersão do modelo multivariado ZICOMP Tipo III

Suponha que estamos interessados em testar qual dos modelos ZIP Tipo III ou ZICOMP Tipo III se ajusta melhor a um conjunto de dados. Uma alternativa para esse problema é estender o teste da razão de verossimilhanças proposto por [Sellers e Raim \(2016\)](#), para o caso multivariado. Sob a hipótese nula H_0 , testa-se a equidispersão com $\mathbf{v} = \mathbf{1}$, ou seja, o modelo ZIP Tipo III se ajusta melhor aos dados. E sob a hipótese alternativa H_1 , testa-se $\mathbf{v} \neq \mathbf{1}$, a subdispersão ou sobredispersão dos dados. Se a hipótese nula for rejeitada, neste caso, o modelo ZICOMP Tipo III é o mais apropriado para explicar os dados. Assim, a estatística do teste da razão de verossimilhanças é dada por

$$T_{\mathbf{v}} = -2\{\ell(\hat{\phi}_0, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\lambda}}, \hat{\mathbf{v}} = \mathbf{1}; \mathbf{Y}) - \ell(\hat{\phi}_0, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\lambda}}, \hat{\mathbf{v}}; \mathbf{Y})\}, \quad (3.23)$$

em que $\ell(\hat{\phi}_0, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\lambda}}, \hat{\mathbf{v}} = \mathbf{1}; \mathbf{Y})$ e $\ell(\hat{\phi}_0, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\lambda}}, \hat{\mathbf{v}}; \mathbf{Y})$ correspondem, às funções de log-verossimilhança sob as hipóteses H_0 e H_1 , ambas avaliados nas estimativas dos parâmetros $\hat{\phi}_0$, $\hat{\boldsymbol{\phi}}$, $\hat{\boldsymbol{\lambda}}$ e $\hat{\mathbf{v}}$.

A estatística do teste $T_{\mathbf{v}}$ possui distribuição aproximadamente de qui-quadrado $\chi_{(r)}^2$ com (r) graus de liberdade [Johnson, Wichern et al. \(2002, p. 188\)](#) que correspondem ao número de dimensões do vetor \mathbf{v} . Desta forma, o p-valor $_{\mathbf{v}}$ é dado por

$$\Pr(T_{\mathbf{v}} > t_{\mathbf{v}} | H_0) = \Pr(\chi_r^2 > t_{\mathbf{v}}). \quad (3.24)$$

Se p-valor $_{\mathbf{v}} < \alpha$, rejeita-se a hipótese nula H_0 ao nível α de significância.

3.4 Estudo de simulação

Neste Capítulo apresentamos um estudo de simulação com o objetivo de avaliar o comportamento dos EMV em relação ao viés e a REQM, para o modelo da distribuição multivariada ZICOMP Tipo III com e sem covariáveis. Também conduzimos um estudo de simulação Monte Carlo com objetivo de verificar a funcionalidade do teste de hipóteses na Seção 3.3.3. Todas as análises foram realizadas pelo *software* R ([R Development Core Team, 2009](#)).

3.4.1 Ajuste do modelo sem covariáveis

Para o ajuste do modelo sem covariáveis, consideramos apenas o caso trivariado com sobredispersão, ou seja, geramos Y_1, Y_2 e Y_3 amostras de tamanhos ($n=100$ e 500) da distribuição proposta, por meio da marginal $ZICOMP_i^{(III)}(\phi_0 + \phi_i - \phi_0\phi_i, \lambda_i, \nu_i)$ em (3.7). Para o vetor (Y_1, Y_2, Y_3) consideramos o vetor de parâmetros $\theta = (\phi_0, \phi_1, \phi_2, \phi_3, \lambda_1, \lambda_2, \lambda_3, \nu_1, \nu_2, \nu_3)^\top$, com os valores (0.10,0.20,0.20,0.20;3.20,3.60,4.00;0.75,0.85,0.87) e número de parâmetros igual a $3m+1$. O parâmetro ϕ_0 possui o valor 0.10 e representa 10% do excesso de zeros conjuntos na amostra. Já os parâmetros ϕ_1, ϕ_2 e ϕ_3 possuem os valores 0.20 e representam 20% do excesso de zeros extra para cada marginal.

Para cada amostra gerada da distribuição marginal $ZICOMP_i^{(III)}(\phi_0 + \phi_i - \phi_0\phi_i, \lambda_i, \nu_i)$, utilizou-se o teste da distribuição de qui-quadrado $\chi_g^2(1 - \alpha)$ com $g = 4$ graus de liberdade para verificar a aderência das amostras simuladas, todas as amostras não tiveram a hipótese nula rejeitada, pois seus p-valores apresentaram valor maior que o nível $\alpha = 5\%$ de probabilidade. Então, realizamos 5000 réplicas *bootstrap* não paramétrico para obter as estimativas do viés, REQM e o intervalo *bootstrap* com 95% de confiança corrigido e acelerado IBCa(95%). Na Tabela 11, apresentamos os valores dos componentes do vetor de parâmetros θ , e os resultados dos EMV, viés, REQM e os intervalos de confiança assintótico ICa(95%) e *bootstrap* IBCa(95%) para o modelo da distribuição multivariada ZICOMP Tipo III.

Tabela 11 – Resultado da simulação para o modelo ZICOMP Tipo III com sobredispersão.

n	θ	Valor	EMV	Viés	REQM	ICa (95%)	IBCa (95%)
100	ϕ_0	0,10	0,12	0,01	0,03	(0,11; 0,13)	(0,07; 0,21)
	ϕ_1	0,20	0,18	-0,01	0,05	(0,17; 0,19)	(0,08; 0,27)
	ϕ_2	0,20	0,20	0,01	0,05	(0,19; 0,21)	(0,10; 0,29)
	ϕ_3	0,20	0,20	0,01	0,05	(0,19; 0,21)	(0,10; 0,29)
	λ_1	2,80	2,85	0,35	1,09	(-2,01; 7,71)	(1,65; 5,09)
	λ_2	3,10	3,00	0,17	1,01	(-2,18; 8,18)	(1,79; 5,16)
	λ_3	3,30	3,24	0,24	1,15	(-3,03; 9,51)	(1,89; 5,71)
	ν_1	0,68	0,73	0,08	0,19	(0,48; 0,98)	(0,43; 1,07)
	ν_2	0,74	0,70	-0,02	0,15	(0,48; 0,92)	(0,43; 1,00)
	ν_3	0,79	0,75	-0,02	0,16	(0,50; 1,00)	(0,47; 1,07)
500	ϕ_0	0,10	0,10	0,00	0,01	(0,09; 0,10)	(0,08; 0,13)
	ϕ_1	0,20	0,20	0,00	0,02	(0,19; 0,21)	(0,17; 0,24)
	ϕ_2	0,20	0,20	0,00	0,02	(0,19; 0,21)	(0,17; 0,24)
	ϕ_3	0,20	0,21	0,01	0,02	(0,20; 0,22)	(0,17; 0,25)
	λ_1	2,80	2,79	0,04	0,36	(0,80; 4,78)	(2,20; 3,54)
	λ_2	3,10	3,04	0,00	0,40	(0,58; 5,51)	(2,39; 3,90)
	λ_3	3,30	3,15	-0,07	0,46	(0,42; 5,87)	(2,40; 4,08)
	ν_1	0,68	0,69	0,02	0,07	(0,59; 0,79)	(0,56; 0,82)
	ν_2	0,74	0,67	0,01	0,07	(0,58; 0,77)	(0,61; 0,88)
	ν_3	0,79	0,79	0,00	0,08	(0,65; 0,90)	(0,63; 0,94)

Podemos observar, na Tabela 11, que para as amostras de tamanho $n = 100$, os EMV são ligeiramente viesados, o parâmetro v_3 é não-significativo para a sobredispersão ao nível $\alpha = 5\%$ de probabilidade, isto é, seus intervalos com 95% de confiança ICa e IBCa incluíram valores maiores ou iguais a um. O mesmo ocorreu com os parâmetros λ_1 , λ_2 e λ_3 no intervalo ICa que incluiu valores para λ_i fora do espaço paramétrico, pois $\lambda_i > 0$. Entretanto, à medida que aumentamos o tamanho das amostras como, por exemplo, para $n = 500$, o viés e a REQM diminuem, e os parâmetros λ_i e v_i tornam-se significativos ao nível $\alpha = 5\%$ de probabilidade.

3.4.2 Ajuste do modelo com covariáveis

Para o ajuste do modelo com covariáveis, consideramos o caso bivariado, associado ao efeito da covariável x_i . Dessa forma, o modelo de regressão em (3.19) para o vetor aleatório $\mathbf{Y}_j = (Y_{1j}, Y_{2j})^\top$, associado ao vetor $\mathbf{X}_j = (1, x_{1j})^\top$, com $j = 1, \dots, n$, pode ser escrito por

$$\left\{ \begin{array}{l} (Y_{1j}, Y_{2j})^\top \stackrel{iid}{\sim} \text{ZICOMP}_2^{(\text{III})}(\phi_{0j}, \phi_{1j}, \phi_{2j}, \lambda_{1j}, \lambda_{2j}, v_{1j}, v_{2j}), \quad j = 1, \dots, n; \\ \phi_{0j} = \frac{\exp(\gamma_0)}{(1 + \exp(\gamma_0))}; \quad \phi_{1j} = \frac{\exp(\gamma_0 + \gamma_1 x_{1j})}{(1 + \exp(\gamma_0 + \gamma_1 x_{1j}))}; \quad \phi_{2j} = \frac{\exp(\gamma_2 + \gamma_2 x_{1j})}{(1 + \exp(\gamma_2 + \gamma_2 x_{1j}))}; \\ \lambda_{1j} = \exp(\beta_{10} + \beta_{11} x_{1j}); \quad \lambda_{2j} = \exp(\beta_{20} + \beta_{21} x_{1j}); \\ v_{1j} = \exp(\zeta_{10j}); \quad v_{2j} = \exp(\zeta_{20j}). \end{array} \right. \quad (3.25)$$

Em que x_{1j} é uma realização da variável X com distribuição normal $N(0, 0.25)$, $\boldsymbol{\gamma}_0 = \gamma_0$; $\boldsymbol{\gamma}_i = (\gamma_{i0}, \gamma_{i1})^\top$, $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1})^\top$ e $\boldsymbol{\zeta} = (\zeta_{10}, \zeta_{20})^\top$ são vetores de coeficientes da regressão. Então, baseado nos valores dos parâmetros: $\boldsymbol{\gamma}_0 = -2.50$, $\boldsymbol{\gamma}_1 = (-1.60, -0.81)^\top$; $\boldsymbol{\gamma}_2 = (-1.41, 0.92)^\top$; $\boldsymbol{\beta}_1 = (1.16, -0.26)^\top$; $\boldsymbol{\beta}_2 = (1.10, 0.30)^\top$, $\zeta_{10} = -0.25$ e $\zeta_{20} = -0.30$, geramos duas amostras Y_1 e Y_2 com os tamanhos $n = 100$ e 500 da distribuição proposta. Dizemos que o vetor aleatório $\mathbf{Y} = (Y_1, Y_2)^\top$ é sobredisperso e zero inflacionado Tipo III, pois $v_{1j} = \exp(\zeta_{10j}) = \exp(-0.25) \approx 0,78$ e $v_{2j} = \exp(\zeta_{20j}) = \exp(-0.30) \approx 0,74$. E os parâmetros de inflação conjunta ϕ_{00} , e de inflação adicionais ϕ_{10} e ϕ_{20} , em Y_1 e Y_2 , correspondem, respectivamente a $\exp(\gamma_0)/(1 + \exp(\gamma_0)) = \exp(-2.50)/(1 + \exp(-2.50)) \approx 0.08$, $\exp(\gamma_{10})/(1 + \exp(\gamma_{10})) = \exp(-1.60)/(1 + \exp(-1.60)) \approx 0.17$ e $\exp(\gamma_{20})/(1 + \exp(\gamma_{20})) = \exp(-1.41)/(1 + \exp(-1.41)) \approx 0.20$. Foram realizadas 5000 réplicas bootstrap não paramétrico para obter as estimativas do viés, REQM e o intervalo de confiança bootstrap corrigido acelerado IBCa(95%).

A Tabela 12 apresenta os valores do vetor de parâmetros $(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{v})^\top$, juntamente com os EMV, viés, REQM, ICa(95%) e IBCa(95%) para o modelo de regressão da distribuição multivariada ZICOMP Tipo III.

Tabela 12 – Resultados da simulação para o modelo de regressão multivariada ZICOMP Tipo III.

n	θ	Valor	EMV	Viés	REQM	ICa(95%)	IBCa(95%)
100	γ_{00}	-2,50	-2,59	-0,09	0,72	(-3,35; -1,82)	(-3,73; -1,91)
	γ_{10}	-1,60	-1,59	0,03	0,37	(-2,21; -0,96)	(-2,60; -1,06)
	γ_{11}	-0,81	-0,87	0,22	0,91	(-2,27; 0,52)	(-3,58; 0,42)
	γ_{20}	-1,41	-1,41	0,01	0,35	(-2,00; -0,81)	(-2,36; -0,89)
	γ_{21}	0,92	1,00	0,26	0,80	(-0,35; 2,35)	(-0,43; 2,39)
	β_{10}	1,16	1,21	0,30	0,46	(0,60; 1,82)	(0,31; 1,61)
	β_{11}	-0,26	-0,20	0,05	0,19	(-0,45; 0,05)	(-0,58; 0,12)
	β_{20}	1,10	1,07	0,22	0,41	(0,47; 1,67)	(0,33; 1,48)
	β_{21}	0,30	0,30	0,07	0,19	(0,05; 0,55)	(-0,09; 0,56)
	v_1	0,78	0,79	0,09	0,21	(0,52; 1,08)	(0,39; 1,08)
	v_2	0,74	0,76	0,13	0,23	(0,49; 1,03)	(0,25; 0,98)
500	γ_{00}	-2,50	-2,53	-0,04	0,18	(-2,86; -2,19)	(-2,86; -2,18)
	γ_{10}	-1,60	-1,55	0,04	0,15	(-1,83; -1,27)	(-1,85; -1,29)
	γ_{11}	-0,81	-0,85	-0,01	0,29	(-1,37; -0,26)	(-1,41; -0,20)
	γ_{20}	-1,41	-1,42	-0,02	0,14	(-1,69; -1,15)	(-1,71; -1,16)
	γ_{21}	0,92	0,96	0,05	0,26	(0,46; 1,96)	(0,47; 1,46)
	β_{10}	1,16	1,17	0,03	0,15	(0,90; 1,44)	(0,87; 1,42)
	β_{11}	-0,26	-0,26	-0,01	0,06	(-0,36; -0,17)	(-0,38; -0,15)
	β_{20}	1,10	1,09	0,01	0,14	(0,82; 1,35)	(0,82; 1,35)
	β_{21}	0,30	0,30	0,00	0,05	(0,20; 0,40)	(0,19; 0,29)
	v_1	0,78	0,79	0,03	0,09	(0,67; 0,92)	(0,61; 0,94)
	v_2	0,74	0,75	0,02	0,08	(0,63; 0,86)	(0,59; 0,90)

Podemos observar na Tabela 12, que para as amostras de tamanho $n = 100$, os EMV são ligeiramente viesados e os coeficientes da regressão γ_{11} , γ_{21} , β_{11} e β_{21} , com os parâmetros de dispersão v_1 e v_2 , são não significativos ao nível de 5% de probabilidade, ou seja, os intervalos de confiança para os coeficientes da regressão incluíram o termo zero. Já para os parâmetros v_1 e v_2 seus intervalos de confiança incluíram valores maiores ou iguais a um, isto é, a sobredispersão é não significativa. Entretanto, à medida que aumentamos o tamanho das amostras, como por exemplo, para $n = 500$, o viés e a REQM diminuem e os parâmetros γ_{11} , γ_{21} , β_{11} , β_{21} , v_1 e v_2 tornam-se significativos ao nível de 5% de probabilidade.

3.4.3 Teste de hipóteses

Para ilustrar o teste da razão de verossimilhanças em (3.3.3), geramos 200 amostras da distribuição proposta para o caso bivariado sem covariáveis, e para o vetor $(\phi_0; \phi_1, \phi_2; \lambda_1, \lambda_2; v_1, v_2)^\top$ com $3m+1$ o número de dimensões, consideramos os valores $(0.10; 0.15, 0.15; 2.00, 3.00; 1.00, 1.00)^\top$ com equidispersão e $(0.10; 0.20, 0.20; 3.20, 4.00; 0.75, 0.78)^\top$ com sobredispersão. As hipóteses a serem testadas pelo teste da razão de verossimilhanças são $H_0 : (v_1, v_2) = (1, 1)$ versus $H_1 : (v_1, v_2) \neq (1, 1)$. No teste de razão de verossimilhanças, testar a hipótese H_0 equivale a testar a equidispersão do vetor $(v_1, v_2) = (1, 1)$, quando o modelo ZICOMP Tipo III equi-

vale ao modelo ZIP Tipo III. Na Tabela 13, apresentamos as estatísticas do teste da razão de verossimilhanças com $r=2$ graus de liberdade, juntamente com seus respectivos p-valores.

Tabela 13 – Resultados do teste da razão de verossimilhanças.

Dispersão	$T_{\mathbf{v}}$	p-valor $_{\mathbf{v}}$
Equidispersão	0,02	0,99
Sobredispersão	8,08	0,02

Podemos observar na Tabela 13, que a hipótese nula H_0 foi rejeitada para a sobredispersão. Entretanto, H_0 não foi rejeitada para a equidispersão, quando o modelo ZICOMP Tipo III equivale ao modelo ZIP Tipo III. Isso demonstra a flexibilidade da distribuição ZICOMP Tipo III, ao considerar as distribuições que são seus casos particulares, como a ZIP Tipo III.

3.5 Aplicação com dados reais

Nesta Seção, apresentamos uma aplicação com dados reais, que corresponde ao número de casos confirmados de AIDS e sífilis congênita em uma amostra com 175 cidades do interior do estado do Ceará-Brasil. Os dados foram coletados pelo autor no anuário estatístico do ano de 2013 e na página do DATASUS disponível em (([CEARÁ, 2013](#)); ([SAÚDE, 2014](#))). A Figura 2 apresenta o mapa político do Brasil com a localização do estado do Ceará.

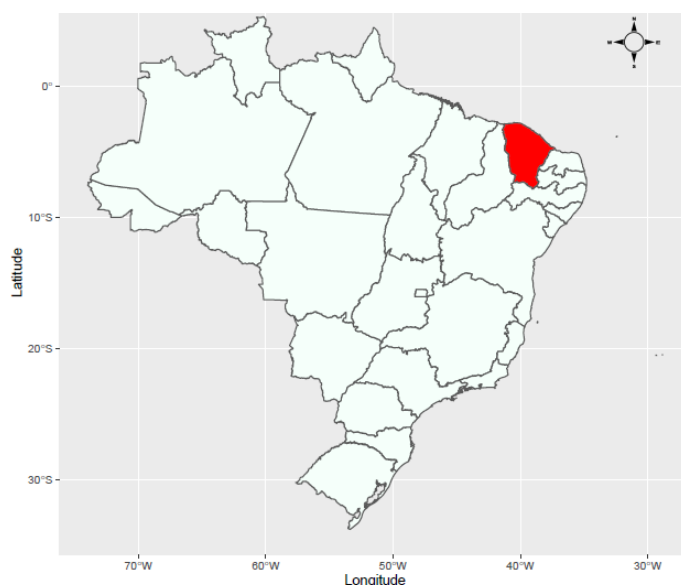


Figura 2 – Mapa Político do Brasil com o Estado do Ceará.

Trabalhos envolvendo modelos multivariados para modelar DST como, por exemplo, o modelo de regressão multivariado de Poisson são encontrados em (([LI et al., 2010](#)); ([LUCAR et al., 2018](#)); ([RAUCH et al., 2005](#)); ([RODRIGUES et al., 2009b](#)); ([SANTOS et al., 2017](#))). Geralmente, dados envolvendo DST podem estar associados às características espaciais e sócio-demográficas de cada região, veja os trabalhos de (([TAN et al., 2011](#)); ([FEDE et al., 2011](#))).

Na Tabela 14, apresentamos as variáveis Y_1 e Y_2 , como o número de casos confirmados de AIDS e sífilis congênita.

Tabela 14 – Frequências cruzadas das variáveis Y_1 e Y_2 em 175 cidades do interior do estado do Ceará-Brasil disponível em ((CEARÁ, 2013); (SAÚDE, 2014)).

$Y_1 Y_2$	0	1	2	3	4	5	6	7	9	10	11	12	13	15	Total
0	52	15	8	4	3	0	1	0	1	0	1	0	0	0	85
1	14	5	3	2	1	2	0	0	1	0	0	1	1	0	30
2	15	7	2	1	2	1	0	0	0	1	0	0	0	0	29
3	4	3	3	0	0	1	0	0	0	1	0	0	0	0	12
4	4	0	0	0	0	0	1	1	0	0	0	0	0	0	6
5	0	2	0	0	0	1	1	0	0	0	0	0	0	0	4
6	1	0	0	2	0	0	0	1	0	0	0	1	0	0	5
9	0	0	1	0	0	1	0	0	0	0	0	0	0	1	3
10	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Total	90	32	17	9	6	7	3	2	2	2	1	2	1	1	175

Observamos, na Tabela 14, que as frequências das células (0,0), $(Y_1, 0)$ e $(0, Y_2)$ são relativamente maiores quando comparadas com as outras células, ou seja, existem indícios de inflação de zeros da célula (0,0) e inflação nos demais componentes $(Y_1, 0)$ e $(0, Y_2)$. Em uma análise descritiva, as variáveis Y_1 e Y_2 aparentemente são sobredispersas, pois suas variâncias são maiores que suas médias: $\bar{y}_1 = 1,34$; $\bar{y}_2 = 1,63$; $s_1^2 = 3,78$ e $s_2^2 = 7,81$.

3.5.1 Ajuste do modelo sem covariável

Para explicar a relação existente entre as variáveis Y_1 e Y_2 , consideramos os modelos ZIP Tipo III e ZICOMP Tipo III, isto é, para o vetor de observações $\mathbf{Y}_j = (Y_{1j}, Y_{2j})^\top$, com $j = 1, \dots, 175$, ajustamos os modelos $\text{ZIP}_3^{(III)}(\phi_0, \phi_1, \phi_2; \lambda_1, \lambda_2)$ e $\text{ZICOMP}_3^{(III)}(\phi_0, \phi_1, \phi_2; \lambda_1, \lambda_2; \nu_1, \nu_2)$. Quanto aos critérios de escolha do modelo que melhor se ajusta aos dados, utilizamos como critérios de escolha os critérios de informação AIC e BIC (para mais detalhes, veja Akaike (1974) e Schwarz (1978)), além do teste da razão de verossimilhanças proposto na Seção (3.3.3). Na Tabela 15, apresentamos os valores dos critérios AIC, BIC e a estatística $T_{\mathbf{v}}$ com seu respectivo p-valor $_{\mathbf{v}}$.

Tabela 15 – Critério de seleção dos modelos multivariados ZIP Tipo III e ZICOMP Tipo III.

Modelo	$T_{\mathbf{v}}$	p-valor $_{\mathbf{v}}$	AIC	BIC
ZIP Tipo III	41,60	p<0,0001	1260,38	1302,03
ZICOMP Tipo III	-	-	1152,61	1210,92

Na Tabela 15, observamos que o modelo ZICOMP Tipo III apresentou menor valor para os critérios de informação AIC e BIC. Quanto ao teste da razão de verossimilhanças, ele rejeitou a hipótese para a equidispersão nos dados. Nesse caso, o modelo ZICOMP Tipo III é o mais indicado para explicar a relação existente entre as variáveis AIDS e sífilis congênita

em cidades do interior do estado do Ceará, segundo os critérios AIC, BIC e o teste razão de verossimilhanças. Na Tabela 16, apresentamos os EMV e os intervalos com 95% de confiança ICa(95%) e IBCa(95%) com 5000 réplicas *bootstrap* para o modelo escolhido.

Tabela 16 – Estimativas do modelo ZICOMP Tipo III sem covariáveis para as variáveis Y_1 e Y_2 .

θ	EMV	ICa(95%)	IBCa(95%)
ϕ_0	0,16	(0,14; 0,17)	(0,05; 0,26)
ϕ_1	0,04	(0,03; 0,06)	$(1,21 \times 10^{-25}; 0,25)$
ϕ_2	0,18	(0,15; 0,20)	(0,03; 0,34)
λ_1	0,66	(0,46; 0,86)	(0,56; 0,92)
λ_2	0,70	(0,66; 0,74)	(0,63; 0,76)
ν_1	0,05	(0,03; 0,06)	$(9,57 \times 10^{-11}; 0,33)$
ν_2	$3,02 \times 10^{-6}$	$(3,02 \times 10^{-6}; 3,02 \times 10^{-6})$	$(7,00 \times 10^{-8}; 0,22)$

Observamos na Tabela 16, que a estimativa do parâmetro de inflação conjunta ϕ_0 , que representa a proporção de zeros da célula (0,0), apresentou valor $\hat{\phi}_0 = 0,16$, diferente da proporção de zeros nos dados, pois $52/175 \approx 0,30$. Isso ocorreu porque parte dos (0,0) correspondem à sobredispersão provocada pelo excesso de zeros segundo o modelo, pois de acordo com as estimativas dos parâmetros de dispersão, $\hat{\nu}_1 = 0,05$ e $\hat{\nu}_2 \approx 0,00$. Já em relação ao parâmetro ϕ_1 , o mesmo foi não-significativo ao nível $\alpha = 5\%$, pois seu intervalo com 95% de confiança IBCa incluiu o termo zero. Entretanto, o mesmo não ocorreu com o parâmetro ϕ_2 , o que justifica a utilização do modelo ZICOMP Tipo III. Quanto aos demais parâmetros, todos são significativos ao nível $\alpha = 5\%$ e os dados são sobredispersos, ou seja, os intervalos de confiança para os parâmetros de dispersão: ν_1 e ν_2 , não incluíram valores maiores ou iguais a um.

Outro modelo que poderia se ajustar a esse conjunto de dados seria o modelo ainda não existente da distribuição multivariada geométrica zero-inflacionado Tipo III (ZIG Tipo III), uma vez que a distribuição COM-Poisson equivale à distribuição geométrica com probabilidade $(1 - \lambda_i)$ quando $\lambda_i < 1$ e $\nu_i = 0$. E a condução de um teste de hipótese tipo razão de verossimilhanças, ou a implementação do teste empírico multivariado proposto por [McAssey \(2013\)](#) para dados discretos, poderiam verificar essa suposição.

3.5.2 Análise das distribuições marginais

As variáveis Y_1 e Y_2 são correlacionadas positivamente, o que justifica a utilização do modelo. De acordo com as estimativas das correlações de Pearson e a proposta pelo modelo ZICOMP Tipo III na Equação (3.6), as variáveis Y_1 e Y_2 apresentaram as seguintes estimativas $r_p = 0,36$ e $r_m = 0,07$. A correlação proposta pelo modelo ZICOMP Tipo III apresentou valor menor do que a correlação de Pearson. Isso é uma característica comum da família da distribuição Tipo III, pois são correlacionadas na variável Z_0 . E, ao estimarmos a correlação do modelo ZIP Tipo III para as variáveis Y_1 e Y_2 , obteve-se valor igual a 0,12.

De acordo com a distribuição marginal proposta na Equação (3.7), teoricamente, cada $Y_i (i = 1, 2) \sim \text{ZICOMP}^{(\text{III})}(\phi_0 + \phi_i - \phi_0\phi_i, \lambda_i, \nu_i)$ possui parâmetro de inflação equivalente ao do modelo ZICOMP univariado em (BARRIGA; LOUZADA, 2014). Então, baseado nas estimativas do modelo ZICOMP em Barriga e Louzada (2014): $Y_1 \sim \text{ZICOMP}(0.20, 0.66, 0.05)$ e $Y_2 \sim \text{ZICOMP}(0.31, 0.70, 0.00)$. E são idênticas às do modelo ZICOMP Tipo III, isto é, $Y_1 \sim \text{ZICOMP}^{(\text{III})}(0.20, 0.66, 0.05)$ e $Y_2 \sim \text{ZICOMP}^{(\text{III})}(0.31, 0.70, 0.00)$. Nesse caso, o modelo ZICOMP Tipo III produz resultados satisfatórios para a análise das marginais neste conjunto de dados.

3.5.3 Ajuste do modelo com covariáveis

Finalmente, consideramos o modelo de regressão na Equação (3.19) associado ao efeito de uma covariável para explicar a relação existente entre as variáveis Y_1 e Y_2 com a covariável X (IDH Índice de Desenvolvimento Humano, disponível em (Ceará (2013); Saúde (2014))) das 175 cidades do interior do estado do Ceará. Em uma análise descritiva da correlação de Pearson das variáveis Y_1 e Y_2 com a covariável X , obtiveram-se as seguintes correlações: $\text{corr}(Y_1, X) = 0,30$ e $\text{corr}(Y_2, X) = 0,46$. Aparentemente, as variáveis Y_1 e Y_2 são correlacionadas positivamente com a covariável X . Na Tabela 17, apresentamos os EMV e os intervalos com 95% de confiança $\text{ICa}(95\%)$ e $\text{IBCa}(95\%)$ com 5000 réplicas *bootstrap* não-paramétrico para o modelo de regressão da distribuição ZICOMP Tipo III.

Tabela 17 – Estimativas do modelo de regressão ZICOMP Tipo III para as variáveis Y_1 e Y_2 com a covariável IDH.

θ	EMV	ICa(95%)	IBCa(95%)
ϕ_0	0,16	(0,14; 0,17)	(0,06; 0,27)
ϕ_1	0,10	(0,08; 0,12)	$(4,77 \times 10^{-9}; 0,30)$
ϕ_2	0,19	(0,16; 0,21)	(0,03; 0,36)
β_{10}	-4,46	(-6,72; -2,20)	(-6,74; -2,25)
β_{11}	6,91	(2,88; 10,94)	(2,93; 10,80)
β_{20}	-5,65	(-8,08; -3,22)	(-5,99; -4,39)
β_{21}	8,93	(4,60; 13,26)	(5,28; 9,00)
ν_1	0,24	(0,16; 0,32)	$(1,99 \times 10^{-8}; 0,60)$
ν_2	0,21	(0,16; 0,27)	(0,08; 0,39)

Assim como na Tabela 16, observamos, na Tabela 17, que o parâmetro de inflação ϕ_1 é não significativo ao nível $\alpha = 5\%$, pois seu intervalo de confiança IBCa incluiu o termo zero. Entretanto, isso não ocorreu com o parâmetro ϕ_2 , o que justifica a utilização do modelo ZICOMP Tipo III. Quanto aos coeficientes da regressão β_{11} e β_{21} , os mesmos são significativos ao nível $\alpha = 5\%$, ou seja, seus intervalos de confiança não incluíram o valor zero. O mesmo ocorreu com os parâmetros de dispersão ν_1 e ν_2 , isto é, seus intervalos não incluíram valores maiores ou iguais a um, ou seja, a sobredispersão existente nos dados é significativa.

Como o parâmetro ϕ_1 foi não significativo, neste caso, optamos pela sua retirada no modelo. A Tabela 18 apresenta os EMV, e os intervalos de confiança assintóticos ICa(95%) e *bootstrap* corrigido e acelerado IBCa(95%), com 5000 réplicas *bootstrap* para o modelo sem o parâmetro de inflação adicional ϕ_1 .

Tabela 18 – Estimativas do modelo de regressão ZICOMP Tipo III sem o parâmetro de inflação ϕ_1 .

θ	EMV	ICa(95%)	IBCa(95%)
ϕ_0	0,17	(0,14; 0,17)	(0,07; 0,26)
ϕ_2	0,18	(0,15; 0,20)	(0,01; 0,34)
β_{10}	-4,16	(-6,10; -2,22)	(-6,16; -2,26)
β_{11}	6,17	(2,87; 9,47)	(2,93; 9,95)
β_{20}	-5,65	(-8,08; -3,22)	(-8,03; -2,87)
β_{21}	8,94	(4,60; 13,28)	(4,14; 13,26)
v_1	0,14	(0,11; 0,17)	$(6,53 \times 10^{-6}; 0,37)$
v_2	0,21	(0,16; 0,27)	$(3,60 \times 10^{-5}; 0,48)$

Observamos na Tabela 18, que todos os parâmetros do modelo são significativos ao nível $\alpha = 0,05$ de probabilidade, segundo seus intervalos ICa(95%) e IBCa(95%) com 95% de confiança, ou seja, existe uma relação entre o número de casos confirmados de AIDS e sífilis congênita com IDH, e elas são sobredispersas.

3.6 Conclusão

Neste Capítulo, apresentamos a distribuição multivariada de Conway-Maxwell-Poisson zero-inflacionado Tipo III (ZICOMP Tipo III), e sua construção foi baseada na extensão da distribuição de Poisson multivariada zero-inflacionado Tipo III. Desenvolvemos importantes propriedades teóricas da distribuição, um modelo de regressão e um teste de razão de verossimilhanças para ajudar na escolha do melhor modelo. E concluímos que esta nova distribuição pode produzir melhores ajustes em conjuntos de dados multivariados de contagem com excesso de zeros e correlacionados.

UM SIMPLES ESTUDO DE PONTOS INFLUENTES NO MODELO DE REGRESSÃO DA DISTRIBUIÇÃO MULTIVARIADA ZICOMP TIPO I

4.1 Introdução

Dados de contagem multivariados com excesso de zeros são comuns nas áreas da saúde (Cameron e Trivedi (2013)), indústria (Li *et al.* (1999)), colisões de carros em rodovia (Dong *et al.* (2014)), horticultura (Sun (2016)), lesões ocupacionais (Wang *et al.* (2003)), entre outras. Os modelos das distribuições multivariadas de Poisson zero-inflacionado: ZIP Tipo I Liu e Tian (2015); ZIP Tipo III Tian *et al.* (2014) e as distribuições ZICOMP Tipo I e a ZICOMP Tipo III apresentadas nos Capítulos 2 e 3, respectivamente, aparecem como fortes candidatas para modelagem desse tipo de dados, com destaque para as distribuições ZICOMP Tipo I e ZICOMP Tipo III, que são mais flexíveis devido ao seu vetor de parâmetros de dispersão \mathbf{v} , que consegue acomodar vários níveis de dispersão. Existem situações em que o modelo ZICOMP Tipo I ou ZICOMP Tipo III se ajusta melhor a um conjunto de dados e os critérios de informação de AIC e BIC podem ser utilizados como critérios de escolha do modelo. Entretanto, podem ocorrer situações em que o AIC apresenta menor valor para um modelo, e o BIC para o outro modelo, ou vice-versa. Nesse caso, dois testes do tipo da razão de verossimilhanças para testar o vetor de parâmetros de inflação ϕ e parâmetro de inflação conjunta ϕ_0 ou ϕ poderiam ser utilizados para ajudar na escolha do melhor modelo, pois quando $\phi = \mathbf{0}$, o modelo ZICOMP Tipo III equivale ao modelo ZICOMP Tipo I, e $\phi_0 = 0$ ou $\phi = 0$ nenhum dos modelos são indicados para representar os dados.

Após a etapa da escolha do modelo, é preciso verificar a existência de pontos discrepantes no conjunto de dados, que exercem influência significativa nas estimativas dos parâmetros do

modelo ((MARTINEZ, 2008); (PEREIRA, 2012)). Essa etapa é conhecida como análise de diagnóstico e tem o objetivo de determinar as observações que são influentes nas estimativas dos parâmetros do modelo. Existem várias técnicas de diagnóstico, dentre as quais se destacam as técnicas: influência global (distância generalizada de Cook em Cook (1977), e o afastamento da função de verossimilhança em Cook e Weisberg (1982)) e influência local (ponderação de casos; perturbação na variável resposta; perturbação na variável explicativa e perturbação na matriz de variância e covariância; detalhes, veja Cook (1986)).

Mediante ao exposto, este Capítulo teve por objetivo geral construir os testes da razão de verossimilhanças para o vetor de parâmetros de inflação ϕ do modelo ZICOMP Tipo III e para o parâmetro de inflação ϕ do modelo ZICOMP Tipo I, além da condução de uma análise de diagnóstico para o modelo de regressão da distribuição ZICOMP Tipo I via influência global (distância generalizada de Cook) e local (ponderação de pesos), devido à complexidade da estrutura do modelo.

Este Capítulo está organizado da seguinte maneira. Na seção 4.2, apresentamos o modelo de regressão linear da distribuição ZICOMP Tipo I. Já na Seção 4.3, apresentamos os testes da razão de verossimilhanças para testar o parâmetro de inflação ϕ do modelo ZICOMP Tipo I e o vetor de parâmetros de inflação ϕ do modelo ZICOMP Tipo III. Na Seção 4.4, apresentamos a análise de influência global e local para o modelo de regressão via distância generalizada de Cook e ponderação de casos. Na Seção 4.5, conduzimos uma análise com dados reais. E a conclusão é apresentada na Seção 4.6.

4.2 Modelo de regressão da distribuição multivariada ZICOMP Tipo I

A distribuição multivariada ZICOMP Tipo I, introduzida no Capítulo 2, é uma distribuição resultante da mistura de um vetor de variáveis aleatórias de COM-Poisson e uma variável aleatória Z com distribuição de Bernoulli($1-\phi$), para captar a estrutura de correlação entre os componentes do vetor de contagem m -dimensional \mathbf{Y} . A função de massa de probabilidade de $\mathbf{Y} \sim \text{ZICOMP}_m^{(1)}(\phi, \boldsymbol{\lambda}, \mathbf{v})$, é dada por

$$f(\mathbf{y}|\phi, \boldsymbol{\lambda}, \mathbf{v}) = \left[\phi + (1 - \phi) \prod_{i=1}^m \frac{1}{S(\lambda_i, v_i)} \right] I(\mathbf{y} = \mathbf{0}) + \left[(1 - \phi) \prod_{i=1}^m \frac{\lambda_i^{y_i}}{S(\lambda_i, v_i)(y_i!)^{v_i}} \right] I(\mathbf{y} \neq \mathbf{0}), \quad (4.1)$$

em que $I(\mathbf{y})$, é uma função indicadora do tipo $I(\mathbf{y}) = \mathbf{1}$ se $\mathbf{y} = \mathbf{0}$ e $I(\mathbf{y}) = \mathbf{0}$ caso contrário, $I(\mathbf{y} \neq \mathbf{0}) = (\mathbf{1} - I(\mathbf{y} = \mathbf{0}))$, $S(\lambda_i, v_i)$ é o valor da constante de normalização da distribuição de COM-Poisson em Conway e Maxwell (1962), na i -ésima dimensão, $\lambda_i > 0$ é um parâmetro de centro, $v_i \geq 0$ é um parâmetro de dispersão, com sobredispersão quando ($v_i < 1$), subdispersão com ($v_i > 1$) e equidispersão quando ($v_i = 1$) ((SHMUELI *et al.*, 2005); (BARRIGA; LOUZADA, 2014); (SELLERS; SWIFT; WEEMS, 2017)).

Para uma amostra aleatória observada $(\mathbf{Y}_1, \mathbf{z}_1, \mathbf{w}_1, \mathbf{u}_1) \dots, (\mathbf{Y}_n, \mathbf{z}_n, \mathbf{w}_n, \mathbf{u}_n)$, se considerarmos a estrutura de regressão da distribuição ZICOMP Tipo I, o parâmetro de inflação de zeros ϕ_j , relacionam-se com a covariável \mathbf{z}_j pela função de ligação logística, já os vetores de parâmetros $\boldsymbol{\lambda}_j = (\lambda_{1j}, \dots, \lambda_{mj})^\top$ e $\mathbf{v}_j = (v_{1j}, \dots, v_{mj})^\top$, se relacionam com as covariáveis \mathbf{w}_j e \mathbf{u}_j , por meio da função de ligação logarítmica, isto é,

$$\log\left(\frac{\phi_j}{1 - \phi_j}\right) = \mathbf{z}_j^\top \boldsymbol{\gamma}; \quad \log(\lambda_{ij}) = \mathbf{w}_j^\top \boldsymbol{\beta}_i; \quad \log(v_{ij}) = \mathbf{u}_j^\top \boldsymbol{\zeta}_i; \quad i = 1 \dots, m, \quad (4.2)$$

em que $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{mj})^\top$ é um vetor de observações; $\mathbf{z}_j = (1, z_{1j}, \dots, z_{pj})^\top$; $\mathbf{w}_j = (1, w_{1j}, \dots, w_{qj})^\top$ e $\mathbf{u}_j = (1, u_{1j}, \dots, u_{rj})^\top$ são vetores de covariáveis não necessariamente idênticos com respeito à j ; $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_p)^\top$, $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1}, \dots, \beta_{iq})^\top$ e $\boldsymbol{\zeta}_i = (\zeta_{i0}, \zeta_{i1}, \dots, \zeta_{ir})^\top$ são vetores de coeficientes da regressão. Dessa forma, o logaritmo da função verossimilhança do modelo é dado por

$$\begin{aligned} \ell(\boldsymbol{\theta} | \mathbf{Y}_{obs}) = k \log \left(\left\{ \frac{\exp(\mathbf{z}_j^\top \boldsymbol{\gamma})}{1 + \exp(\mathbf{z}_j^\top \boldsymbol{\gamma})} + \frac{1}{1 + \exp(\mathbf{z}_j^\top \boldsymbol{\gamma})} \prod_{i=1}^m \left[\frac{1}{S(\exp(\mathbf{w}_j^\top \boldsymbol{\beta}_i), \exp(\mathbf{u}_j^\top \boldsymbol{\zeta}_i))} \right] \right\} \right) + (n - k) \times \\ \log(1 + \exp(\mathbf{z}_j^\top \boldsymbol{\gamma})) + \sum_{j=k+1}^n \sum_{i=1}^m \left\{ [I_{(y_{ij} > 0)}] \left[-\log S(\exp(\mathbf{w}_j^\top \boldsymbol{\beta}_i), \exp(\mathbf{u}_j^\top \boldsymbol{\zeta}_i)) + y_{ij} \mathbf{w}_j^\top \boldsymbol{\beta}_i + \right. \right. \\ \left. \left. - \exp(\mathbf{u}_j^\top \boldsymbol{\zeta}_i) \log(y_{ij}!) \right] \right\}, \end{aligned} \quad (4.3)$$

em que $\mathbf{Y}_{obs} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$ é um vetor de dados observados; $\mathbb{J} = \{j \mid \mathbf{y}_j = \mathbf{0}, j = 1, \dots, n\}$ e $k = \sum_{j=1}^n I(\mathbf{y}_j = \mathbf{0})$.

A obtenção da estimativa do vetor de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta}_i, \mathbf{v}_i)$ desconhecido do modelo, pode ser feita pela maximização da função de log-verossimilhança na Equação (4.3), usando o método Quasi-Newton com o algoritmo BFGS (veja [Nocedal e Wright \(2006\)](#)), implementado no *software R* ([R Development Core Team \(2009\)](#)) por meio da função *optim*. Na implementação do código para o modelo proposto, utilizamos, como estimativas iniciais, as estimativas do modelo Poisson para o vetor $\boldsymbol{\beta}_i$ e $\mathbf{v}_i = \exp(\boldsymbol{\zeta}_i = \mathbf{0}) = \mathbf{1}$ para os parâmetros da COM-Poisson em [Conway e Maxwell \(1962\)](#), além da estimativa do modelo Binomial para o parâmetro de inflação $\boldsymbol{\gamma}$.

4.3 Teste da bondade do ajuste

Testes da bondade do ajuste de modelos são testes utilizados para verificar se um determinado conjunto de dados ou amostra é proveniente de uma determinada distribuição a certo nível de probabilidade desejado, [Olmos \(2017\)](#). O teste da razão de verossimilhanças se revela como uma ótima ferramenta na escolha do modelo de certa distribuição, cuja variável resposta pode ser tanto contínua quanto discreta. Nesta Seção, apresentamos dois testes da razão de verossimilhanças para o modelo de regressão da distribuição ZICOMP Tipo I.

4.3.1 Teste da razão de verossimilhanças para o parâmetro de inflação total de zeros

Suponha que estamos interessados em testar se o parâmetro inflação de zeros ϕ do modelo de regressão da distribuição ZICOMP Tipo I é igual a zero, pois para o parâmetro $\phi = 0$, na Equação (2.2) do Capítulo 2, isso implica dizer que todos os elementos do vetor \mathbf{Y} são variáveis COM-Poisson independentes, e neste caso cada elemento Y_i do vetor \mathbf{Y} pode ser ajustado por um modelo univariado de COM-Poisson. O teste da razão de verossimilhanças pode ser utilizado para verificar essa suposição sobre o parâmetro ϕ . Sob a hipótese nula testa-se $H_0 : \phi = 0$, versus a hipótese alternativa $H_1 : \phi > 0$. Dessa forma, a estatística do teste da razão de verossimilhanças para o parâmetro de inflação ϕ é dada por

$$T_\phi = -2\{\ell(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}}; \mathbf{Y}) - \ell(\hat{\phi}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{v}}; \mathbf{Y})\}, \quad (4.4)$$

em que $\ell(\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}}; \mathbf{Y})$ é o logaritmo da função de verossimilhança sob H_0 , e $\ell(\hat{\phi}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{v}}; \mathbf{Y})$ é o logaritmo da função de verossimilhança sob H_1 , ambos avaliados nas estimativas dos parâmetros $\hat{\phi} = \exp(\mathbf{z}^\top \hat{\boldsymbol{\gamma}}) / (1 + \exp(\mathbf{z}^\top \hat{\boldsymbol{\gamma}}))$, $\hat{\boldsymbol{\beta}}$ e $\hat{\mathbf{v}} = \exp(\mathbf{u}^\top \hat{\boldsymbol{\xi}})$; para mais detalhes, veja a Equação (2.19) no Capítulo (2). Assim como nos testes da razão de verossimilhança para o parâmetro ϕ em (Liu e Tian (2015); Huang *et al.* (2017); Zhang, Tian e Huang (2015)) a estatística T_ϕ possui distribuição $\frac{1}{2}(\chi_0^2 + \chi_1^2)$, para mais detalhes, veja os trabalhos de ((JANSAKUL; HINDE, 2002); (FENG; MCCULLOCH, 1992); (SELF; LIANG, 1987); (JOE; ZHU, 2005); (BUSE, 1982)). Dessa forma, o p-valor $_\phi$ é dado por

$$\text{p-valor}_\phi = \frac{1}{2} \Pr(T_\phi > t_\phi | H_0) = \frac{1}{2} \Pr(\chi_1^2 > t_1). \quad (4.5)$$

Se p-valor $_\phi < \alpha$, rejeita-se a hipótese nula H_0 ao nível α de significância.

4.3.2 Teste da razão de verossimilhanças para o vetor de inflação do modelo multivariado ZICOMP Tipo III

A estrutura de inflação do modelo de regressão da distribuição multivariada ZICOMP Tipo I considera apenas a inflação conjunta dos componentes do vetor aleatório de contagem \mathbf{Y} , e os demais componentes são considerados como variáveis de COM-Poisson, ou seja, ele considera a estrutura do tipo $(0, \dots, \text{COMP}(\lambda_i, v_i), \text{COMP}(\lambda_j, v_j), 0, \dots, 0)$. E por ser mais flexível em relação à inflação de zeros, o modelo da distribuição multivariada ZICOMP Tipo III, considera a estrutura do tipo $(0, \dots, \text{ZICOMP}(\phi_i, \lambda_i, v_i), \text{ZICOMP}(\phi_j, \lambda_j, v_j), 0, \dots, 0)$. Se estivermos interessados em testar qual dos modelos ZICOMP Tipo I ou ZICOMP Tipo III se ajusta melhor a um conjunto de dados, o teste da razão de verossimilhanças surge como uma opção para verificar a bondade do ajuste do modelo ZICOMP Tipo I versus ZICOMP Tipo III.

Sob a hipótese nula H_0 , testa-se $\boldsymbol{\phi} = \mathbf{0}$, ou seja, se o modelo ZICOMP Tipo I se ajusta melhor aos dados. E sob a hipótese alternativa H_1 , testa-se $\boldsymbol{\phi} > \mathbf{0}$. Se a hipótese nula for rejeitada, neste caso o modelo ZICOMP Tipo III é o mais apropriado para explicar os dados. Assim, a estatística do teste da razão de verossimilhanças é dada por

$$T_{\boldsymbol{\phi}} = -2\{\ell(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{v}}; \mathbf{Y}) - \ell(\hat{\boldsymbol{\phi}}_0, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{v}}; \mathbf{Y})\}, \quad (4.6)$$

em que $\ell(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{v}}; \mathbf{Y})$ e $\ell(\hat{\boldsymbol{\phi}}_0, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{v}}; \mathbf{Y})$ correspondem aos logaritmos das funções de verossimilhanças sob H_0 e H_1 , ambos avaliados nas estimativas dos parâmetros: $\hat{\boldsymbol{\beta}} = \exp(\mathbf{z}^\top \hat{\boldsymbol{\gamma}})/(1 + \exp(\mathbf{z}^\top \hat{\boldsymbol{\gamma}}))$; $\hat{\boldsymbol{\phi}}_0 = \exp(\mathbf{z}^\top \hat{\boldsymbol{\gamma}}_0)/(1 + \exp(\mathbf{z}^\top \hat{\boldsymbol{\gamma}}_0))$; $\hat{\boldsymbol{\phi}} = \exp(\mathbf{z}^\top \hat{\boldsymbol{\gamma}})/(1 + \exp(\mathbf{z}^\top \hat{\boldsymbol{\gamma}}))$ e $\hat{\boldsymbol{v}} = \exp(\mathbf{u}^\top \hat{\boldsymbol{\zeta}})$; para detalhes, veja as Equações (2.19) e (3.19) nos Capítulos 2 e 3 respectivamente.

A estatística do teste $T_{\boldsymbol{\phi}}$, possui distribuição aproximadamente qui-quadrado $\chi_{(p)}^2$ com (p) graus de liberdade Johnson, Wichern *et al.* (2002, p. 188), que corresponde ao número de dimensões do vetor de parâmetro de inflação $\boldsymbol{\gamma}_i$ da distribuição ZICOMP Tipo III. Desta forma, o p-valor $_{\boldsymbol{\phi}}$ é dado por

$$\text{p-valor}_{\boldsymbol{\phi}} = \Pr(T_{\boldsymbol{\phi}} > t_{\boldsymbol{\phi}} \mid H_0) = \Pr(\chi_p^2 > t_{\boldsymbol{\phi}}). \quad (4.7)$$

Se $\text{p-valor}_{\boldsymbol{\phi}} < \alpha$, rejeita-se a hipótese nula H_0 ao nível α de significância desejado.

4.4 Análise de influência

Uma observação pode ser considerada como influente quando ela altera de forma significativa as estimativas dos parâmetros do modelo. Nesta Seção, apresentamos duas medidas de influência para o modelo de regressão da distribuição multivariada ZICOMP Tipo I.

4.4.1 Influência global

A influência global tem a finalidade de analisar o impacto causado nas estimativas dos parâmetros do modelo, ao se retirar uma ou mais observações consideradas como influentes. Uma das medidas mais utilizadas para a detecção de pontos influentes na influência global é a distância generalizada de Cook em (COOK, 1977).

Considere o modelo de regressão da distribuição multivariada ZICOMP Tipo I, e seja $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\zeta}}^\top)^\top$ e $\hat{\boldsymbol{\theta}}_{(j)} = (\hat{\boldsymbol{\gamma}}_j, \hat{\boldsymbol{\beta}}_j^\top, \hat{\boldsymbol{\zeta}}_j^\top)^\top$ as estimativas de máxima verossimilhança para $\boldsymbol{\theta}$ com e sem a j -ésima observação. Então a distância generalizada de Cook para a j -ésima observação é dada por

$$GD_j(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_{(j)} - \hat{\boldsymbol{\theta}})^\top [\hat{\mathbf{J}}(\boldsymbol{\theta})] (\hat{\boldsymbol{\theta}}_{(j)} - \hat{\boldsymbol{\theta}}), \quad (4.8)$$

em que $\hat{\mathbf{J}}(\boldsymbol{\theta})$ é a matriz de informação observada, quanto maior a medida GD_j , mais influente será a j -ésima observação no ajuste do modelo (GARAY *et al.*, 2011).

4.4.2 Influência local

A influência local é uma metodologia de diagnóstico proposta por Cook (1986), que permite verificar a existência de observações influentes, ao inserir pequenas perturbações no modelo ou nos dados ((MARTINEZ, 2008); (PEREIRA, 2012)). Um dos motivos do método de influência local ser um dos métodos mais utilizados na análise de diagnóstico é o fato de que ele pode ser aplicado em qualquer modelo desde que se tenha a função de verossimilhança (PEREIRA, 2012). Desde então, inúmeros trabalhos foram desenvolvidos para classe de modelos univariados e multivariados, utilizando a metodologia proposta por (COOK, 1986). Trabalhos com modelos multivariados usando influência local são encontrados em ((ZHU; ZHU; SONG, 2008); (CHEN *et al.*, 2012); (MATOS *et al.*, 2015); (PATRIOTA; LEMONTE; CASTRO, 2010); (SÁNCHEZ; VILCA, 2014); (MARCHANT *et al.*, 2016)).

Considere $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\gamma}, \boldsymbol{\beta}_i, \boldsymbol{\zeta}_i)$, o logaritmo da função de verossimilhança do modelo de regressão da distribuição ZICOMP Tipo I, e $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_i$ e $\boldsymbol{\zeta}_i$ são vetores de parâmetros desconhecidos do modelo com dimensões $1 \times (p+1)$; $1 \times (q+1)$ e $1 \times (1+r)$, $i = 1, \dots, m$. E seja $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^\top$ um vetor de perturbações restrito a um conjunto aberto Ω , de forma que $\boldsymbol{\delta} \in \Omega \subseteq \mathbb{R}^n$. E considere também $\ell(\boldsymbol{\gamma}, \boldsymbol{\beta}_i, \boldsymbol{\zeta}_i | \boldsymbol{\delta})$, o logaritmo da função de verossimilhança do modelo perturbado, e $\boldsymbol{\delta}_0 = (1, \dots, 1)^\top$, o vetor de não perturbações de dimensão $1 \times n$, tal que $\ell(\boldsymbol{\gamma}, \boldsymbol{\beta}_i, \boldsymbol{\zeta}_i | \boldsymbol{\delta}_0) = \ell(\boldsymbol{\gamma}, \boldsymbol{\beta}_i, \boldsymbol{\zeta}_i)$. Para avaliar o efeito de $\boldsymbol{\delta}$ sobre $(\boldsymbol{\gamma}, \boldsymbol{\beta}_i, \boldsymbol{\zeta}_i)$, pode-se utilizar o afastamento da função de verossimilhança, que é dado por

$$LD(\boldsymbol{\delta}) = 2 \left\{ \ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}} | \boldsymbol{\delta}) \right\}, \quad (4.9)$$

em que $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\theta}} | \boldsymbol{\delta}$ são as estimativas de máxima verossimilhanças de $\boldsymbol{\theta}$ dos modelos postulado e perturbado, respectivamente. A ideia básica da influência local consiste em inserir pequenas perturbações no modelo ou nos dados e então estudar o comportamento da função $LD(\boldsymbol{\delta})$ em torno de $\boldsymbol{\delta}_0$. O gráfico de $LD(\boldsymbol{\delta})$ corresponde à superfície $\{\boldsymbol{\delta}, LD(\boldsymbol{\delta})\}$, que desvia de seu plano tangente em $\boldsymbol{\delta}_0$ à medida que $\boldsymbol{\delta}$ se afasta de $\boldsymbol{\delta}_0$ ((CARRASCO, 2012); (FAGUNDES *et al.*, 2017); (PAULA, 2015, p. 56)).

Cook (1986) sugere analisar a curvatura normal associada à maior direção \mathbf{d}_{\max} , e apresenta a curvatura normal C_d na direção \mathbf{d} , na forma

$$C_d = 2 | \mathbf{d}^\top \mathbf{F} \mathbf{d} |, \quad (4.10)$$

em que $\mathbf{F} = \Delta^\top \hat{\mathbf{J}}(\boldsymbol{\theta})^{-1} \Delta$, $\hat{\mathbf{J}}(\boldsymbol{\theta})$ é a matriz de informação observada do modelo postulado, Δ é uma matriz de dimensão $\{(p+1) + i((q+1) + (r+1))\} \times n$, com elementos $\Delta_{lj} = \frac{\partial^2 \ell(\boldsymbol{\theta} | \boldsymbol{\delta})}{\partial \theta_l \partial \delta_j}$,

avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\delta} = \boldsymbol{\delta}_0$, $l = 1, \dots, (p+1) + i((q+1) + (r+1))$ e $j = 1, \dots, n$, que depende do esquema de perturbação. A análise gráfica da direção do autovetor normalizado \mathbf{d}_{\max} associado à maior curvatura $C_{\mathbf{d}_{\max}}$, que corresponde ao maior autovalor da matriz \mathbf{F} , permite encontrar os elementos que, sob um esquema de perturbação, exercem influência desproporcional na função de afastamento da verossimilhança $LD(\boldsymbol{\delta})$ (PAULA, 2015, p. 56).

4.4.3 Ponderação de casos

Considere $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\gamma}, \boldsymbol{\beta}_i, \boldsymbol{\zeta}_i)$ o logaritmo da função de verossimilhança do modelo de regressão linear da distribuição ZICOMP Tipo I. E seja $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^\top$ um vetor de perturbações, com $0 \leq \delta_j \leq 1$, e $j = 1, \dots, n$. Neste caso, o logaritmo da função de verossimilhança perturbada é dado por

$$\ell(\boldsymbol{\theta} | \boldsymbol{\delta}) = \sum_{j=1}^n \delta_j \ell_j(\boldsymbol{\theta}). \quad (4.11)$$

A matriz de perturbação Δ , para o modelo de regressão da distribuição ZICOMP Tipo I, pode ser particionada na forma $\Delta = (\Delta\boldsymbol{\gamma}, \Delta\boldsymbol{\beta}_i, \Delta\boldsymbol{\zeta}_i)^\top$, em que cada uma das matrizes que compõem Δ possui n linhas e número de colunas igual ao número de dimensões do correspondente vetor de parâmetros. Os elementos da matriz Δ são apresentados no Apêndice C.

4.5 Aplicação com dados reais

Nesta Seção, apresentamos uma aplicação com um conjunto de dados reais, que corresponde ao número de casos confirmados de AIDS e hepatite viral em 97 cidades do interior do estado de Alagoas. As informações de AIDS e hepatite viral foram coletadas pelo autor no anuário estatístico do estado de Alagoas-Brasil referente ao ano de 2015, disponível em (ALAGOAS, 2015). Na Tabela 22, do item C.1.1, no Apêndice C, apresentamos as variáveis Y_1 , Y_2 e X , como o número de casos confirmados de AIDS, hepatite viral e o IDH das 97 cidades em estudo respectivamente. Em uma análise descritiva, a célula $(Y_1, Y_2) = (0, 0)$ apresentou frequência igual a 28, sendo maior que as outras células. Aparentemente, as variáveis Y_1 e Y_2 são sobredispersas e correlacionadas positivamente com a covariável X , isto é, as estimativas das correlações de Pearson, das médias e das variâncias foram: $\text{corr}(Y_1, X) = 0,40$; $\text{corr}(Y_2, X) = 0,31$; $\bar{y}_1 = 2,49$; $\bar{y}_2 = 1,36$; $s_1^2 = 15,44$ e $s_2^2 = 15,23$.

4.5.1 Ajuste do modelo com covariável

Para explicar a relação existente entre as variáveis Y_1 e Y_2 , com a covariável X , consideramos os modelos de regressão das distribuições ZIP Tipo I; ZIP Tipo III; ZICOMP Tipo I e ZICOMP Tipo III. Na Tabela 19, apresentamos os EMV para os parâmetros dos modelos propostos, juntamente com as estimativas dos critérios de informação AIC e BIC para os respectivos modelos.

Tabela 19 – Estimativas de máxima verossimilhança para os modelos de regressão.

	ZIP Tipo I	ZICOMP Tipo I	ZIP Tipo III	ZICOMP Tipo III
θ	EMV	EMV	EMV	EMV
ϕ_0	0,25	0,14	0,18	0,13
ϕ_1	-	-	0,18	0,00
ϕ_2	-	-	0,35	0,13
β_{10}	-5,79	-2,02	-4,58	-2,07
β_{11}	12,17	2,98	10,22	3,04
β_{21}	-9,91	-3,93	-9,59	-3,58
β_{22}	18,13	5,86	18,20	5,34
v_1	-	$4,50 \times 10^{-5}$	-	$2,26 \times 10^{-10}$
v_2	-	$9,56 \times 10^{-7}$	-	$1,71 \times 10^{-7}$
AIC	886,26	670,42	843,36	673,10
BIC	899,70	688,44	893,41	696,27

Observamos na Tabela 19, que os modelos ZIP Tipo I e Tipo III conseguiram captar melhor a proporção de zeros da célula (0,0), segundo a estimativa do parâmetro ϕ_0 , pois $28/97 \approx 0,29$. Entretanto, os modelos ZICOMP Tipo I e Tipo III, consideraram que parte dos (0,0) corresponde à sobredispersão provocada pelo excesso de zeros, pois, de acordo com as estimativas dos seus parâmetros de dispersão, $\hat{v}_1 \approx 0,00$ e $\hat{v}_2 \approx 0,00$. Para sanar a dúvida em qual dos modelos devemos utilizar para explicar os dados, uma vez que os modelos ZIP Tipo I e Tipo III conseguiram captar melhor a proporção de zeros, enquanto que os modelos ZICOMP Tipo I e Tipo III conseguiram captar melhor a sobredispersão existente nos dados, utilizamos os critérios de informação AIC e BIC. De acordo com as estimativas dos critérios AIC e BIC na Tabela 19, o modelo ZICOMP Tipo I foi o que apresentou menor valor para os critérios, sendo portanto, o mais indicado. Quanto ao teste da razão de verossimilhanças, na Seção 4.3.2, ele apresentou a estatística do teste $T_\phi = 1,32$ e p-valor $_\phi = 0,52$, ou seja, os dados se ajustam melhor ao modelo ZICOMP Tipo I, quando comparado ao modelo ZICOMP Tipo III.

Após a escolha do modelo ZICOMP Tipo I, pelos critérios AIC e BIC, e o teste da razão de verossimilhanças na Seção 4.3.2, utilizamos o teste da razão de verossimilhanças na Seção 4.3.1 para verificar se o parâmetro (ϕ ou ϕ_0) é significativo ao nível $\alpha = 0,05$ de probabilidade; caso o parâmetro ϕ não seja significativo, a relação das variáveis Y_1 e Y_2 , com a covariável X , deveria ser avaliada independentemente. A estatística do teste da razão de verossimilhanças na Seção 4.3.1, apresentou o valor $T_\phi = 6,15$ e p-valor $_\phi = 0,007$, sendo, portanto, significativa ao nível $\alpha = 0,05$, e a relação das variáveis Y_1 e Y_2 , com a covariável X , não deve ser avaliada independentemente.

A Tabela 20 apresenta os EMV, e os intervalos de confiança assintótico ICa(95%) e *bootstrap* corrigido e acelerado IBCa(95%) (para mais detalhes, veja Efron e Tibshirani (1994); DiCiccio e Efron (1996)), com 5000 réplicas *bootstrap* para o modelo de regressão da distribuição ZICOMP Tipo I.

Tabela 20 – Estimativas do modelo de regressão, para as variáveis Y_1 e Y_2 com a covariável IDH.

θ	EMV	ICa(95%)	IBCa(95%)
ϕ	0,14	(0,13; 0,15)	$(7,03 \times 10^{-10}; 0,31)$
β_{10}	-2,02	(-2,85; -1,19)	(-7,45; -0,58)
β_{11}	2,98	(1,62; 4,34)	(0,40; 12,37)
β_{20}	-3,93	(-5,19; -2,67)	(-23,53; -1,59)
β_{21}	5,86	(3,86; 7,86)	(1,89; 7,00)
v_1	$4,50 \times 10^{-5}$	$(4,43 \times 10^{-5}; 4,56 \times 10^{-5})$	$(6,80 \times 10^{-15}; 2,72)$
v_2	$9,56 \times 10^{-7}$	$(9,56 \times 10^{-7}; 9,57 \times 10^{-7})$	$(7,77 \times 10^{-12}; 2,72)$

Podemos observar na Tabela 20, que o intervalo de confiança IBCa(95%) apresentou valores desproporcionais em relação ao intervalo IBCa(95%) para os parâmetros β_{20} ; v_1 ; v_2 e ϕ , uma vez que o teste da razão de verossimilhanças na Seção 4.3.1 considerou o parâmetro ϕ como significativo e, ao incluir no termo zero no intervalo IBCa(95%), ϕ foi não significativo. De maneira semelhante, o mesmo ocorreu com os parâmetros de dispersão v_1 e v_2 , pois seus intervalos de confiança incluíram valores maiores que um, o que torna a sobredispersão não significativa. Isso ocorreu porque, provavelmente, podem existir pontos discrepantes que estão influenciando no ajuste do modelo. Para avaliar a presença de pontos influentes no ajuste do modelo, utilizamos a distância generalizada de Cook na Seção 4.4.1 e influência local via ponderação de casos na Seção 4.4.2. Na Figura 3, apresentamos os gráficos da distância de Cook (item (a)) e influência local (item (b)) *versus* a ordem das observações do modelo ajustado.

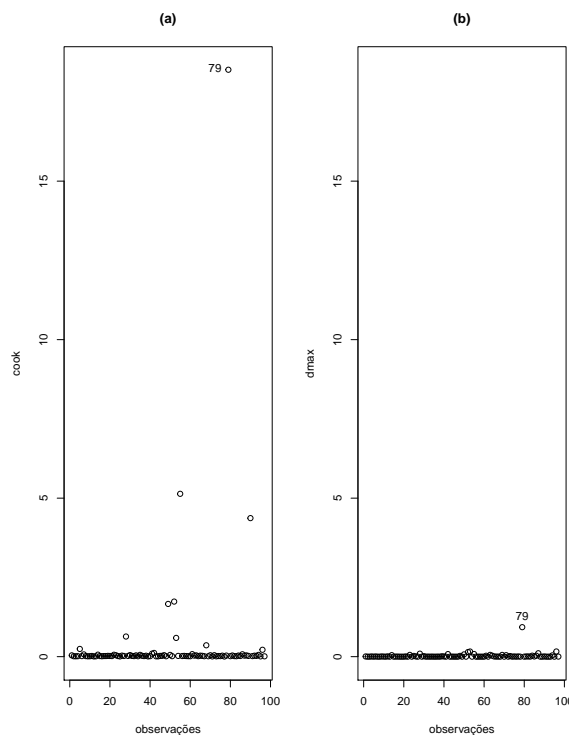


Figura 3 – Análise de influência

A Figura 3 revelou o ponto #79 como um possível ponto influente na estimativa dos parâmetros. Esse ponto corresponde à notificação $(Y_1, Y_2) = (16, 35)$ da cidade de Rio Largo, com IDH de 0,643 e população de 68.481 habitantes. Na Tabela 21, apresentamos o resultado final da análise inferencial com a eliminação do ponto #79 e com os menores valores das estimativas do AIC e BIC, com os valores 645 e 661.

Tabela 21 – Estimativas do modelo de regressão final.

θ	EMV	ICa(95%)	IBCa(95%)
ϕ	0,17	(0,15; 0,19)	(0,05; 0,28)
β_{10}	-1,84	(-2,74; -0,94)	(-3,03; -0,60)
β_{11}	2,66	(1,16; 4,16)	(0,40; 4,63)
β_{20}	-0,60	(-0,74; -0,46)	(-0,80; -0,42)
v_1	$1,60 \times 10^{-5}$	(0,00; 0,00)	$(2,56 \times 10^{-16}; 0,12)$
v_2	$2,79 \times 10^{-10}$	$(2,79 \times 10^{-10}; 2,79 \times 10^{-10})$	$(1,85 \times 10^{-48}; 8,67 \times 10^{-6})$

Observamos, na Tabela 21, que todos os parâmetros do modelo são significativos ao nível $\alpha = 0,05$ de probabilidade segundo os dois intervalos ICa(95%) e IBCa(95%), ou seja, existe uma relação entre o número de casos confirmados de AIDS e hepatite viral, e elas são sobredispersas. Entretanto, quando associadas com a covariável IDH, a relação entre hepatite viral com IDH não é existente nas 97 cidades do interior do estado de Alagoas.

4.6 Conclusão

Neste Capítulo, apresentamos uma análise de diagnóstico para o modelo de regressão da distribuição multivariada ZICOMP Tipo I, proposta no Capítulo 2. Consideramos uma análise de influência global (distância generalizada de Cook) e local (ponderação de casos). Também desenvolvemos dois testes da razão de verossimilhanças para ajudar na escolha do melhor modelo: um teste da razão de verossimilhanças para testar o parâmetro de inflação ϕ da distribuição ZICOMP Tipo I e o outro para testar o vetor de inflação ϕ da distribuição ZICOMP Tipo III.

As metodologias propostas neste Capítulo apresentaram bons resultados na análise de um conjunto de dados reais. E, por fim, concluímos que estas metodologias podem ajudar na escolha e no ajuste de modelos para dados multivariados com excesso de zeros.

CONCLUSÃO

Nesta tese, foram apresentadas duas novas distribuições multivariadas com o excesso de zeros e com os seus modelos de regressão. A construção das duas novas distribuições ZICOMP Tipo I e ZICOMP Tipo III foi baseada na extensão das distribuições ZIP Tipo I e ZIP Tipo III, sendo a distribuição ZICOMP Tipo I obtida com a substituição dos componentes do vetor base \mathbf{X} com variáveis Poisson por variáveis COM-Poisson, e ZICOMP Tipo III com a substituição dos componentes do vetor base \mathbf{X} com variáveis ZIP por variáveis ZICOMP.

Uma ampla descrição das propriedades e particularidades destas duas novas distribuições multivariadas, ZICOMP Tipo I e ZICOMP Tipo III, foi apresentada nos Capítulos 2, 3 e 4, para o melhor entendimento dessa distribuições, tais como, as distribuições marginais; distribuições condicionais; modelo de regressão; estimação pelo método da máxima verossimilhança; intervalo de confiança *bootstrap*, testes de hipóteses da razão de verossimilhanças para ajudar na escolha do melhor modelo, e uma análise de influência global e local para o modelo de regressão da distribuição ZICOMP Tipo I.

As duas novas distribuições apresentaram resultados satisfatórios, tanto no estudo de simulação quanto na análise com dados reais, como pôde ser visto nos Capítulos 2 e 3. Na Seção 2.5 do Capítulo 2, a distribuição ZICOMP Tipo I apresentou melhor ajuste que a distribuição ZIP Tipo I para explicar a relação entre AIDS, hepatite viral e sífilis congênita em 98 cidades do interior do estado de Alagoas-Brasil, segundo os critérios de informações AIC e BIC, e o teste da razão de verossimilhanças proposto na Seção 2.3.3. O mesmo ocorreu na Seção 3.5 do Capítulo 3 em que a distribuição ZICOMP Tipo III apresentou melhor ajuste que a distribuição ZIP Tipo III para explicar a relação entre AIDS e sífilis congênita em 175 cidades do interior do estado do Ceará-Brasil, segundo os critérios de informação AIC e BIC, e o teste da razão de verossimilhanças proposto na Seção 3.3.3. O mesmo pôde ser comprovado em um terceiro conjunto de dados reais no Capítulo 4, em que os modelos de regressão das distribuições ZICOMP Tipo I e Tipo III, apresentaram melhores ajustes que os modelos das distribuições ZIP

Tipo I e Tipo III, segundo os critérios de informações AIC e BIC.

E, por fim, concluímos que estas duas novas distribuições podem produzir melhores ajustes em dados multivariados de contagem com o excesso de zeros. Nesse sentido, além da utilização destas distribuições para modelagem de dados nas áreas de epidemiologia, biometria, entre outras, várias pesquisas podem ser realizadas como continuação da presente tese, dentre as quais propomos:

- Desenvolver uma análise de diagnóstico com influência global e local para o modelo de regressão da distribuição ZICOMP Tipo III;
- Apresentar uma abordagem Bayesiana para os modelos das distribuições ZICOMP Tipo I e ZICOMP Tipo III, uma vez que, na inferência clássica, as estimativas dos parâmetros são obtidas por métodos iterativos os quais podem não convergir. Além de terem suas inferências fundamentadas na teoria assintótica, o que a torna questionável para determinado tamanho de amostra;
- Construir as distribuições multivariadas Geométricas zero-inflacionado Tipo I e Tipo III baseado nas estruturas das distribuições ZICOMP Tipo I e ZICOMP Tipo III;
- Construir os testes: razão de verossimilhanças, score e gradiente para o parâmetro de inflação ϕ_0 da distribuição multivariada ZICOMP Tipo III;
- Construir os pacotes em linguagem R, para uma maior divulgação e aplicação das distribuições propostas.

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 nas páginas 45 e 62.
- ALAGOAS, S. do Estado de. **Anuário Estatístico do Estado de Alagoas 2014**. Alagoas: [s.n.], 2014. Disponível em: <<http://dados.al.gov.br/dataset/anuario-estatistico-do-estado-de-alagoas/resource/0af79f2d-d42c-41e5-9e55-76e74adfc1a7>>. Citado nas páginas 15, 43 e 44.
- _____. **Anuário Estatístico do Estado de Alagoas 2015**. Alagoas: [s.n.], 2015. Disponível em: <<http://dados.al.gov.br/dataset/anuario-estatistico-do-estado-de-alagoas/resource/e7015c87-944f-46f6-8ee5-d0727578ee7f>>. Citado nas páginas 73 e 105.
- ANGERS, J.-F.; BISWAS, A. A bayesian analysis of zero-inflated generalized poisson model. **Computational statistics & data analysis**, Elsevier, v. 42, n. 1, p. 37–46, 2003. Citado na página 24.
- BARRIGA, G. D.; LOUZADA, F. The zero-inflated conway–maxwell–poisson distribution: Bayesian inference, regression modeling and influence diagnostic. **Statistical Methodology**, Elsevier, v. 21, p. 23–34, 2014. Citado nas páginas 23, 24, 25, 31, 33, 47, 50, 51, 64 e 68.
- BELTRÁN-SILVA, S.; CHACÓN-HERNÁNDEZ, S.; MORENO-PALACIOS, E.; PEREYRA-MOLINA, J. Clinical and differential diagnosis: Dengue, chikungunya and zika. **Revista Médica del Hospital General de México**, Elsevier, v. 81, n. 3, p. 146–153, 2018. Citado na página 25.
- BÖHNING, D.; DIETZ, E.; SCHLATTMANN, P.; MENDONCA, L.; KIRCHNER, U. The zero-inflated poisson model and the decayed, missing and filled teeth index in dental epidemiology. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, Wiley Online Library, v. 162, n. 2, p. 195–209, 1999. Citado na página 24.
- _____. The zero-inflated poisson model and the decayed, missing and filled teeth index in dental epidemiology. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, Wiley Online Library, v. 162, n. 2, p. 195–209, 1999. Citado na página 24.
- BROEK, J. Van den. A score test for zero inflation in a poisson distribution. **Biometrics**, JSTOR, p. 738–743, 1995. Citado na página 24.
- BUSE, A. The likelihood ratio, wald, and lagrange multiplier tests: An expository note. **The American Statistician**, Taylor & Francis, v. 36, n. 3a, p. 153–157, 1982. Citado na página 70.
- CAMERON, A. C.; TRIVEDI, P. K. **Regression analysis of count data**. [S.l.]: Cambridge university press, 2013. v. 53. Citado nas páginas 26, 31, 49 e 67.
- CAMPBELL, J. The poisson correlation function. **Proceedings of the Edinburgh Mathematical Society**, Cambridge University Press, v. 4, n. 1, p. 18–26, 1934. Citado na página 25.
- CARRASCO, J. M. F. **Modelos de regressão beta com erro nas variáveis**. Tese (Doutorado) — Universidade de São Paulo, 2012. Citado na página 72.

- CEARÁ, S. do Estado de. **Anuário Estatístico do Ceará 2013**. Ceará: [s.n.], 2013. Disponível em: <http://www2.ipece.ce.gov.br/publicacoes/anuario/anuario2013/qualidade/saude/vigilancia_epidemiologica.htm>. Citado nas páginas 15, 61, 62 e 64.
- CHA, S.; MALIK, T.; ABARA, W. E.; DESIMONE, M. S.; SCHUMANN, B.; MALLADA, E.; KLEMME, M.; AGUON, V.; SANTOS, A. M.; PETERMAN, T. A. *et al.* Screening for syphilis and other sexually transmitted infections in pregnant women—guam, 2014. **MMWR. Morbidity and mortality weekly report**, Centers for Disease Control and Prevention, v. 66, n. 24, p. 644, 2017. Citado nas páginas 31 e 44.
- CHANIALIDIS, C. **Bayesian mixture models for count data**. Tese (Doutorado) — University of Glasgow, 2015. Citado na página 23.
- CHEN, Y.; SHEN, Z.; MORANO, J. P.; KHOSHNOOD, K.; WU, Z.; LAN, G.; ZHU, Q.; ZHOU, Y.; TANG, S.; LIU, W. *et al.* Bridging the epidemic: a comprehensive analysis of prevalence and correlates of hiv, hepatitis c, and syphilis, and infection among female sex workers in guangxi province, china. **PloS one**, Public Library of Science, v. 10, n. 2, p. e0115311, 2015. Citado na página 26.
- CHEN, Z.; LIU, J.; LIANG, K.; LIANG, W.; MA, S.; ZENG, G.; XIAO, S.; HE, J. The diagnostic value of a multivariate logistic regression analysis model with transvaginal power doppler ultrasonography for the prediction of ectopic pregnancy. **Journal of International Medical Research**, SAGE Publications Sage UK: London, England, v. 40, n. 1, p. 184–193, 2012. Citado na página 72.
- CHOO-WOSOBA, H.; DATTA, S. Analyzing clustered count data with a cluster-specific random effect zero-inflated conway–maxwell–poisson distribution. **Journal of Applied Statistics**, Taylor & Francis, p. 1–16, 2017. Citado nas páginas 23 e 24.
- CHOO-WOSOBA, H.; GASKINS, J.; LEVY, S.; DATTA, S. A bayesian approach for analyzing zero-inflated clustered count data with dispersion. **Statistics in Medicine**, Wiley Online Library, 2017. Citado na página 23.
- CHOO-WOSOBA, H.; LEVY, S. M.; DATTA, S. Marginal regression models for clustered count data based on zero-inflated conway–maxwell–poisson distribution with applications. **Biometrics**, Wiley Online Library, v. 72, n. 2, p. 606–618, 2016. Citado nas páginas 23, 24 e 28.
- CONCEIÇÃO, K. S.; LOUZADA, F.; ANDRADE, M.; HELOU, E. Zero-modified power series distribution and its hurdle distribution version. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 87, n. 9, p. 1842–1862, 2017. Citado na página 24.
- CONSUL, P.; SHOUKRI, M. Maximum likelihood estimation for the generalized poisson distribution. **Communications in statistics. Theory and methods**, Taylor & Francis, v. 13, n. 12, p. 1533–1547, 1984. Citado nas páginas 23 e 32.
- CONSUL, P. C. **Generalized Poisson Distributions**. [S.l.]: Dekker New York, 1989. Citado na página 32.
- CONSUL, P. C.; JAIN, G. C. A generalization of the poisson distribution. **Technometrics**, Taylor & Francis, v. 15, n. 4, p. 791–799, 1973. Citado nas páginas 23 e 32.
- CONWAY, R. W.; MAXWELL, W. L. A queuing model with state dependent service rates. **Journal of Industrial Engineering**, v. 12, n. 2, p. 132–136, 1962. Citado nas páginas 23, 33, 38, 50, 51, 56, 68 e 69.

COOK, R. D. Detection of influential observation in linear regression. **Technometrics**, Taylor & Francis, v. 19, n. 1, p. 15–18, 1977. Citado nas páginas 68 e 71.

_____. Assessment of local influence. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 133–169, 1986. Citado nas páginas 68 e 72.

COOK, R. D.; WEISBERG, S. **Residuals and influence in regression**. [S.l.]: New York: Chapman and Hall, 1982. Citado na página 68.

CYGANEK, B. **Object detection and recognition in digital images: theory and practice**. [S.l.]: John Wiley & Sons, 2013. Citado nas páginas 33 e 51.

DIAS, R. A. P. **Simulação estocástica de variáveis aleatórias Poisson correlacionadas: aplicação ao controle populacional do percevejo (*Euschistus heros* Fabricius) da soja (*Glycine max* L.)**. Tese (Doutorado) — Universidade de São Paulo, 2014. Citado na página 31.

DICICCIO, T. J.; EFRON, B. Bootstrap confidence intervals. **Statistical science**, JSTOR, p. 189–212, 1996. Citado nas páginas 39 e 74.

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 na página 24.

DONG, C.; CLARKE, D. B.; YAN, X.; KHATTAK, A.; HUANG, B. Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections. **Accident Analysis & Prevention**, Elsevier, v. 70, p. 320–329, 2014. Citado nas páginas 26, 27, 29, 31, 49 e 67.

DRÈZE, X.; HOCH, S. J. Exploiting the installed base using cross-merchandising and category destination programs. **International Journal of Research in Marketing**, Elsevier, v. 15, n. 5, p. 459–471, 1998. Citado nas páginas 25 e 31.

EFRON, B.; TIBSHIRANI, R. J. **An introduction to the bootstrap**. [S.l.]: CRC press, 1994. Citado nas páginas 39 e 74.

FAGUNDES, R. S. *et al.* Distribuição slash multivariada aplicada a dados agrícolas. Universidade Estadual do Oeste do Paraná, 2017. Citado na página 72.

FAMOYE, F.; SINGH, K. P. Zero-inflated generalized poisson regression model with an application to domestic violence data. **Journal of Data Science**, v. 4, n. 1, p. 117–130, 2006. Citado na página 24.

FAROUGHI, P.; ISMAIL, N. Bivariate zero-inflated generalized poisson regression model with flexible covariance. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 46, n. 15, p. 7769–7785, 2017. Citado nas páginas 26 e 32.

FEDE, A. L.-D.; STEWART, J. E.; HARDIN, J. W.; MAYFIELD-SMITH, K.; SUDDUTH, D. Spatial visualization of multivariate datasets: an analysis of std and hiv/aids diagnosis rates and socioeconomic context using ring maps. **Public health reports**, SAGE Publications Sage CA: Los Angeles, CA, v. 126, n. 3_suppl, p. 115–126, 2011. Citado na página 61.

FENG, Z.; MCCULLOCH, C. E. Statistical inference using maximum likelihood estimation and the generalized likelihood ratio when the true parameter is on the boundary of the parameter space. **Statistics & Probability Letters**, Elsevier, v. 13, n. 4, p. 325–332, 1992. Citado na página 70.

- FRANCIS, R. A.; GEEDIPALLY, S. R.; GUIKEMA, S. D.; DHAVALA, S. S.; LORD, D.; LAROCCA, S. Characterizing the performance of the conway-maxwell poisson generalized linear model. **Risk Analysis**, Wiley Online Library, v. 32, n. 1, p. 167–183, 2012. Citado nas páginas 23 e 24.
- GARAY, A. M.; HASHIMOTO, E. M.; ORTEGA, E. M.; LACHOS, V. H. On estimation and influence diagnostics for zero-inflated negative binomial regression models. **Computational Statistics & Data Analysis**, Elsevier, v. 55, n. 3, p. 1304–1318, 2011. Citado na página 72.
- GIUFFRÈ, O.; GRANÀ, A.; GIUFFRÈ, T.; MARINO, R. Accounting for dispersion and correlation in estimating safety performance functions. an overview starting from a case study. **Modern Applied Science**, Canadian Center of Science and Education, v. 7, n. 2, p. 11, 2013. Citado na página 24.
- GUIKEMA, S. D.; GOFFELT, J. P. A flexible count data regression model for risk analysis. **Risk analysis**, Wiley Online Library, v. 28, n. 1, p. 213–223, 2008. Citado na página 23.
- GUPTA, P. L.; GUPTA, R. C.; TRIPATHI, R. C. Score test for zero inflated generalized poisson regression model. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 33, n. 1, p. 47–64, 2005. Citado na página 24.
- GUPTA, R. C.; SIM, S.; ONG, S. Analysis of discrete data by conway–maxwell poisson distribution. **ASta Advances in Statistical Analysis**, Springer, v. 98, n. 4, p. 327–343, 2014. Citado nas páginas 24 e 31.
- HALL, D. B. Zero-inflated poisson and binomial regression with random effects: a case study. **Biometrics**, Wiley Online Library, v. 56, n. 4, p. 1030–1039, 2000. Citado na página 24.
- HAWMAN, D. W.; STOERMER, K. A.; MONTGOMERY, S. A.; PAL, P.; OKO, L.; DIAMOND, M. S.; MORRISON, T. E. Chronic joint disease caused by persistent chikungunya virus infection is controlled by the adaptive immune response. **Journal of virology**, Am Soc Microbiol, v. 87, n. 24, p. 13878–13888, 2013. Citado na página 25.
- HO, L. L.; SINGER, J. d. M. Generalized least squares methods for bivariate poisson regression. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 30, n. 2, p. 263–277, 2001. Citado nas páginas 25 e 31.
- HOLGATE, P. Estimation for the bivariate poisson distribution. **Biometrika**, Oxford University Press, v. 51, n. 1-2, p. 241–287, 1964. Citado nas páginas 25 e 26.
- HUANG, X.-F.; TIAN, G.-L.; ZHANG, C.; JIANG, X. Type i multivariate zero-inflated generalized poisson distribution with applications. **Statistics and Its Interface**, International Press of Boston, v. 10, n. 2, p. 291–311, 2017. Citado nas páginas 27, 29, 32 e 70.
- JANSAKUL, N.; HINDE, J. Score tests for zero-inflated poisson models. **Computational statistics & data analysis**, Elsevier, v. 40, n. 1, p. 75–96, 2002. Citado na página 70.
- JOE, H.; ZHU, R. Generalized poisson distribution: the property of mixture of poisson and comparison with negative binomial distribution. **Biometrical Journal: Journal of Mathematical Methods in Biosciences**, Wiley Online Library, v. 47, n. 2, p. 219–229, 2005. Citado na página 70.

- JOHNSON, N.; KOTZ, S. Distributions in statistics: Discrete distributions (boston, ma, houghtonmifflin). **Johnson Distributions in Statistics: Discrete Distributions** 1969, 1969. Citado nas páginas 25 e 27.
- JOHNSON, R. A.; WICHERN, D. W. *et al.* **Applied multivariate statistical analysis**. [S.l.]: Prentice hall Upper Saddle River, NJ, 2002. v. 5. Citado nas páginas 40, 57 e 71.
- JR, A. C. C. Estimating the parameters of a modified poisson distribution. **Journal of the American Statistical Association**, Taylor & Francis, v. 55, n. 289, p. 139–143, 1960. Citado na página 24.
- JU, D. Likelihood-based methods for constrained parameter problems. **HKU Theses Online (HKUTO)**, The University of Hong Kong (Pokfulam, Hong Kong), 2016. Citado na página 27.
- KADANE, J. B.; SHMUELI, G.; MINKA, T. P.; BORLE, S.; BOATWRIGHT, P. *et al.* Conjugate analysis of the conway-maxwell-poisson distribution. **Bayesian analysis**, International Society for Bayesian Analysis, v. 1, n. 2, p. 363–374, 2006. Citado na página 23.
- KARLIS, D.; MELIGKOTSIDOU, L. Multivariate poisson regression with covariance structure. **Statistics and Computing**, Springer, v. 15, n. 4, p. 255–265, 2005. Citado na página 25.
- KARLIS, D.; NTZOUFRAS, I. Analysis of sports data by using bivariate poisson models. **Journal of the Royal Statistical Society: Series D (The Statistician)**, Wiley Online Library, v. 52, n. 3, p. 381–393, 2003. Citado nas páginas 25 e 31.
- KARLIS, D.; NTZOUFRAS, I. *et al.* Bivariate poisson and diagonal inflated bivariate poisson regression models in r. **Journal of Statistical Software**, Foundation for Open Access Statistics, v. 14, n. 10, p. 1–36, 2005. Citado na página 26.
- KAWAMURA, K. The structure of bivariate poisson distribution. In: DEPARTMENT OF MATHEMATICS, TOKYO INSTITUTE OF TECHNOLOGY. **Kodai Mathematical Seminar Reports**. [S.l.], 1973. v. 25, n. 2, p. 246–256. Citado nas páginas 25 e 26.
- _____. The structure of multivariate poisson distribution. **Kodai Mathematical Journal**, Department of Mathematics, Tokyo Institute of Technology, v. 2, n. 3, p. 337–345, 1979. Citado na página 25.
- KHAN, N. M.; RUMJAUN, W.; SUNECHER, Y.; JOWAHEER, V. Computing with bivariate com-poisson model under different copulas. **Monte Carlo Methods and Applications**, De Gruyter, v. 23, n. 2, p. 131–146, 2017. Citado na página 32.
- KIM, D. A simple zero inflated bivariate negative binomial regression model with different dispersion parameters. **Journal of the Korean Data and Information Science Society**, Korean Data and Information Science Society, v. 24, n. 4, p. 895–900, 2013. Citado na página 26.
- KRISHNA, P. M.; TUKARAM, S. D. Bivariate zero-inflated power series distribution. **Applied Mathematics**, Scientific Research Publishing, v. 2, n. 07, p. 824, 2011. Citado na página 26.
- LAMBERT, D. Zero-inflated poisson regression, with an application to defects in manufacturing. **Technometrics**, Taylor & Francis, v. 34, n. 1, p. 1–14, 1992. Citado na página 24.
- LEE, A. J. Modeling scores in the premier league: is manchester united really the best? **Chance**, Taylor & Francis Group, v. 10, n. 1, p. 15–19, 1997. Citado nas páginas 25 e 31.

- LI, C.-S.; LU, J.-C.; PARK, J.; KIM, K.; BRINKLEY, P. A.; PETERSON, J. P. Multivariate zero-inflated poisson models and their applications. **Technometrics**, Taylor & Francis Group, v. 41, n. 1, p. 29–38, 1999. Citado nas páginas 26, 27, 29, 31, 32, 49 e 67.
- LI, D.; JIA, Y.; RUAN, Y.; LIU, Y.; LI, Q.; LIANG, H.; ZHOU, Z.; LUO, F.; ZHANG, X.; JIANG, S. *et al.* Correlates of incident infections for hiv, syphilis, and hepatitis b virus in a cohort of men who have sex with men in beijing. **AIDS patient care and STDs**, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 24, n. 9, p. 595–602, 2010. Citado nas páginas 31, 44 e 61.
- LIU, Y.; TIAN, G.-L. Type i multivariate zero-inflated poisson distribution with applications. **Computational Statistics & Data Analysis**, Elsevier, v. 83, p. 200–222, 2015. Citado nas páginas 9, 11, 27, 29, 31, 32, 39, 46, 47, 49, 67 e 70.
- LORD, D.; GUIKEMA, S. D.; GEEDIPALLY, S. R. Application of the conway–maxwell–poisson generalized linear model for analyzing motor vehicle crashes. **Accident Analysis & Prevention**, Elsevier, v. 40, n. 3, p. 1123–1134, 2008. Citado na página 23.
- LUCAR, J.; HART, R.; RAYEED, N.; TERZIAN, A.; WEINTROB, A.; SIEGEL, M.; PARENTI, D. M.; SQUIRES, L. E.; WILLIAMS, R.; CASTEL, A. D. *et al.* Sexually transmitted infections among hiv-infected individuals in the district of columbia and estimated hiv transmission risk: data from the dc cohort. In: OXFORD UNIVERSITY PRESS US. **Open forum infectious diseases**. [S.l.], 2018. v. 5, n. 2, p. ofy017. Citado nas páginas 31, 44 e 61.
- MAHAMUNULU, D. A note on regression in the multivariate poisson distribution. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 62, n. 317, p. 251–258, 1967. Citado na página 25.
- MARCHANT, C.; LEIVA, V.; CYSNEIROS, F. J. A.; VIVANCO, J. F. Diagnostics in multivariate generalized birnbaum-saunders regression models. **Journal of Applied Statistics**, Taylor & Francis, v. 43, n. 15, p. 2829–2849, 2016. Citado na página 72.
- MARITZ, J. Note on a certain family of discrete distributions. **Biometrika**, JSTOR, v. 39, n. 1/2, p. 196–198, 1952. Citado na página 25.
- MARSHALL, A. W.; OLKIN, I. A family of bivariate distributions generated by the bivariate bernoulli distribution. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 80, n. 390, p. 332–338, 1985. Citado na página 25.
- MARTINEZ, R. O. **Modelos de regressão beta inflacionados**. Tese (Doutorado) — Universidade de São Paulo, 2008. Citado nas páginas 68 e 72.
- MATOS, L. A.; BANDYOPADHYAY, D.; CASTRO, L. M.; LACHOS, V. H. Influence assessment in censored mixed-effects models using the multivariate student’s t distribution. **Journal of multivariate analysis**, Elsevier, v. 141, p. 104–117, 2015. Citado na página 72.
- MCASSEY, M. P. An empirical goodness-of-fit test for multivariate distributions. **Journal of Applied Statistics**, Taylor & Francis, v. 40, n. 5, p. 1120–1131, 2013. Citado na página 63.
- MEHTA, G. S. **Analyzing crash frequency and severity data using novel techniques**. Tese (Doutorado) — University of Alabama Libraries, 2014. Citado nas páginas 25 e 29.
- MINKA, T. P.; SHMEULI, G.; KADANE, J. B.; BORLE, S.; BOATWRIGHT, P. Computing with the com-poisson distribution. figshare, 2002. Citado na página 23.

- M'KENDRICK, A. Applications of mathematics to medical problems. **Proceedings of the Edinburgh Mathematical Society**, Cambridge University Press, v. 44, p. 98–130, 1925. Citado na página 25.
- MOHAMMADI, T.; KHEIRI, S.; SEDEHI, M. Analysis of blood transfusion data using bivariate zero-inflated poisson model: a bayesian approach. **Computational and mathematical methods in medicine**, Hindawi, v. 2016, 2016. Citado na página 26.
- MORRISON, T. E.; OKO, L.; MONTGOMERY, S. A.; WHITMORE, A. C.; LOTSTEIN, A. R.; GUNN, B. M.; ELMORE, S. A.; HEISE, M. T. A mouse model of chikungunya virus-induced musculoskeletal inflammatory disease: Evidence of arthritis, tenosynovitis, myositis, and persistence. **The American journal of pathology**, Elsevier, v. 178, n. 1, p. 32–40, 2011. Citado na página 25.
- NELSON, D. L. Some remarks on generalizations of the negative binomial and poisson distributions. **Technometrics**, Taylor & Francis, v. 17, n. 1, p. 135–136, 1975. Citado nas páginas 23 e 32.
- NOCEDAL, J.; WRIGHT, S. **Numerical optimization**. [S.l.]: Springer Science & Business Media, 2006. Citado nas páginas 38, 56 e 69.
- OLKIN, I. Multivariate non-normal distributions and models of dependency. **Lecture Notes-Monograph Series**, JSTOR, p. 37–53, 1994. Citado na página 27.
- OLMOS, Y. M. G. **Extensões de distribuições com aplicação à análise de sobrevivência**. Tese (Doutorado) — Universidade de São Paulo, 2017. Citado na página 69.
- PATRIOTA, A. G.; LEMONTE, A. J.; CASTRO, M. de. Influence diagnostics in a multivariate normal regression model with general parameterization. **Statistical Methodology**, Elsevier, v. 7, n. 6, p. 644–654, 2010. Citado na página 72.
- PAULA, G. A. Modelos de regressão com apoio computacional. 2013. **Citado na pág**, v. 1, n. 9, p. 10, 2015. Citado nas páginas 72 e 73.
- PEREIRA, G. H. d. A. **Modelos de regressão beta inflacionados truncados**. Tese (Doutorado) — Universidade de São Paulo, 2012. Citado nas páginas 68 e 72.
- PEREIRA, G. M. S. *et al.* Modelos com-poisson correlacionados. Universidade Federal de São Carlos, 2019. Citado na página 29.
- QIN, Q.; TANG, W.; GE, L.; LI, D.; MAHAPATRA, T.; WANG, L.; GUO, W.; CUI, Y.; SUN, J. Changing trend of hiv, syphilis and hepatitis c among men who have sex with men in china. **Scientific reports**, Nature Publishing Group, v. 6, p. 31081, 2016. Citado nas páginas 31 e 44.
- R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2009. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>. Citado nas páginas 38, 56, 57 e 69.
- RAUCH, A.; MARTIN, M.; WEBER, R.; HIRSCHL, B.; TARR, P. E.; BUCHER, H. C.; VERNAZZA, P.; BERNASCONI, E.; ZINKERNAGEL, A. S.; EVISON, J. *et al.* Unsafe sex and increased incidence of hepatitis c virus infection among hiv-infected men who have sex with men: the swiss hiv cohort study. **Clinical infectious diseases**, The University of Chicago Press, v. 41, n. 3, p. 395–402, 2005. Citado na página 61.

- RAZ, S. **COM-Poisson Clustering of Correlated Bivariate Over-and Under-Dispersed Counts**. Tese (Doutorado) — University of Calgary, 2016. Citado nas páginas 29 e 32.
- RIDOUT, M.; DEMÉTRIO, C. G.; HINDE, J. Models for count data with many zeros. In: **Proceedings of the XIXth international biometric conference**. [S.l.: s.n.], 1998. v. 19, p. 179–192. Citado na página 24.
- RIDOUT, M.; HINDE, J.; DEMÉTRIO, C. G. A score test for testing a zero-inflated poisson regression model against zero-inflated negative binomial alternatives. **Biometrics**, Wiley Online Library, v. 57, n. 1, p. 219–223, 2001. Citado na página 24.
- RODRIGUES, J.; CASTRO, M. de; CANCHO, V. G.; BALAKRISHNAN, N. Com–poisson cure rate survival models and an application to a cutaneous melanoma data. **Journal of Statistical Planning and Inference**, Elsevier, v. 139, n. 10, p. 3605–3611, 2009. Citado na página 23.
- RODRIGUES, J.; GRINSZTEJN, B.; BASTOS, F. I.; VELASQUE, L.; LUZ, P. M.; SOUZA, C. T. de; GEORG, I.; PILOTTO, J. H.; VELOSO, V. G. Seroprevalence and factors associated with herpes simplex virus type 2 among hiv-negative high-risk men who have sex with men from rio de janeiro, brazil: a cross-sectional study. **BMC infectious diseases**, BioMed Central, v. 9, n. 1, p. 39, 2009. Citado na página 61.
- RUSSELL, N. K.; NAZAR, K.; PINO, S. del; GONZALEZ, M. A.; BERMÚDEZ, X. P. D.; RAVASI, G. Hiv, syphilis, and viral hepatitis among latin american indigenous peoples and afro-descendants: a systematic review. **Revista Panamericana de Salud Pública**, SciELO Public Health, v. 43, p. e17, 2019. Citado nas páginas 31 e 44.
- SÁNCHEZ, L.; VILCA, F. **Modelos Birnbaum-Saunders bivariados**. Tese (Doutorado) — Dissertação de Mestrado. UNICAMP: Programa de Pós-Graduação em Estatística, SP, 2014. Citado na página 72.
- SANTOS, O. P. dos; SOUZA, M. R. de; BORGES, C. J.; NOLL, M.; CAMPOS, F.; LIMA, P. d. S. B. Hepatitis b and c and syphilis: Prevalence and characteristics associated to coinfection among seropositive individuals. **Cogitare Enferm**, v. 22, n. 3, p. e51693, 2017. Citado na página 61.
- SAÚDE, M. da. **Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis**. 2014. [Online; accessed 16-Junho-2018]. Disponível em: <<http://indicadoressifilis.aids.gov.br/>>. Citado nas páginas 15, 61, 62 e 64.
- SCHWARZ, G. Estimating the dimension of a model the annals of statistics 6 (2), 461–464. URL: <http://dx.doi.org/10.1214/aos/1176344136>, 1978. Citado nas páginas 45 e 62.
- SELF, S. G.; LIANG, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. **Journal of the American Statistical Association**, Taylor & Francis, v. 82, n. 398, p. 605–610, 1987. Citado na página 70.
- SELLERS, K. F.; MORRIS, D. S.; BALAKRISHNAN, N. Bivariate conway–maxwell–poisson distribution: formulation, properties, and inference. **Journal of Multivariate Analysis**, Elsevier, v. 150, p. 152–168, 2016. Citado nas páginas 29, 32 e 50.
- SELLERS, K. F.; RAIM, A. A flexible zero-inflated model to address data dispersion. **Computational Statistics & Data Analysis**, Elsevier, v. 99, p. 68–80, 2016. Citado nas páginas 23, 24, 25, 40 e 57.

SELLERS, K. F.; SHMUELI, G. A flexible regression model for count data. 2008. Citado na página 24.

SELLERS, K. F.; SWIFT, A. W.; WEEMS, K. S. A flexible distribution class for count data. **Journal of Statistical Distributions and Applications**, SpringerOpen, v. 4, n. 1, p. 22, 2017. Citado na página 68.

SHENSON, J.; JOSHI, A. Visualizing disease incidence in the context of socioeconomic factors. In: ACM. **Proceedings of the 5th International Symposium on Visual Information Communication and Interaction**. [S.l.], 2012. p. 29–38. Citado nas páginas 25 e 31.

SHMUELI, G.; MINKA, T. P.; KADANE, J. B.; BORLE, S.; BOATWRIGHT, P. A useful distribution for fitting discrete data: revival of the conway–maxwell–poisson distribution. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, Wiley Online Library, v. 54, n. 1, p. 127–142, 2005. Citado nas páginas 23, 33, 51 e 68.

SHONKWILER, J. S.; SHAW, W. D. Hurdle count-data models in recreation demand analysis. **Journal of Agricultural and Resource Economics**, v. 21, n. 1835-2016-148885, p. 210–219, 1996. Citado na página 24.

SIM, S. Z.; GUPTA, R. C.; ONG, S. H. Zero-inflated conway-maxwell poisson distribution to analyze discrete data. **The international journal of biostatistics**, De Gruyter, v. 14, n. 1, 2018. Citado na página 25.

SUN, Y. **New Multivariate Zero-Inflated Binomial Distribution with Application**. Tese (Doutorado) — Faculty of Graduate Studies and Research, University of Regina, 2016. Citado nas páginas 26, 27, 31, 49 e 67.

TAN, N. X.; MESSINA, J. P.; YANG, L.-G.; YANG, B.; EMCH, M.; CHEN, X.-S.; COHEN, M. S.; TUCKER, J. D. A spatial analysis of county-level variation in syphilis and gonorrhea in guangdong province, china. **PLoS One**, Public Library of Science, v. 6, n. 5, p. e19648, 2011. Citado nas páginas 44 e 61.

TIAN, G.; WU, Q.; TANG, M.; NG, K. **Type III multivariate zero-inflated Poisson model with applications**. [S.l.], 2014. (Technical Report). Citado nas páginas 9, 11, 27, 29, 47, 49, 50 e 67.

WALHIN, J. F. Bivariate zip models. **Biometrical Journal**, Berlin, Akademie-Verlag, 1977-, v. 43, n. 2, p. 147–160, 2001. Citado na página 26.

WANG, K.; LEE, A. H.; YAU, K. K.; CARRIVICK, P. J. A bivariate zero-inflated poisson regression model to analyze occupational injuries. **Accident Analysis & Prevention**, Elsevier, v. 35, n. 4, p. 625–629, 2003. Citado nas páginas 31, 49 e 67.

WU, G.; HOLAN, S. H.; NILON, C. H.; WIKLE, C. K. *et al.* Bayesian binomial mixture models for estimating abundance in ecological monitoring studies. **The Annals of Applied Statistics**, Institute of Mathematical Statistics, v. 9, n. 1, p. 1–26, 2015. Citado na página 23.

YAHAV, I.; SHMUELI, G. On generating multivariate poisson data in management science applications. **Applied Stochastic Models in Business and Industry**, Wiley Online Library, v. 28, n. 1, p. 91–102, 2012. Citado na página 41.

ZHANG, C.; TIAN, G.; HUANG, X. Two new bivariate zero-inflated generalized poisson distributions with a flexible correlation structure. **Statistics, Optimization & Information Computing**, v. 3, n. 2, p. 105–137, 2015. Citado nas páginas [26](#), [28](#), [32](#), [46](#) e [70](#).

ZHANG, C.; TIAN, G.-L.; YUEN, K. C.; WU, Q.; LI, T. Multivariate zero-and-one inflated poisson model with applications. **Journal of Computational and Applied Mathematics**, Elsevier, p. 112356, 2019. Citado na página [27](#).

ZHOU, X.-H.; TU, W. Confidence intervals for the mean of diagnostic test charge data containing zeros. **Biometrics**, Wiley Online Library, v. 56, n. 4, p. 1118–1125, 2000. Citado na página [24](#).

ZHU, L.; ZHU, R.; SONG, S. Diagnostic checking for multivariate regression models. **Journal of Multivariate Analysis**, Elsevier, v. 99, n. 9, p. 1841–1859, 2008. Citado na página [72](#).

ZIPKIN, E. F.; LEIRNESS, J. B.; KINLAN, B. P.; O'CONNELL, A. F.; SILVERMAN, E. D. Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation. **Statistical Methodology**, Elsevier, v. 17, p. 67–81, 2014. Citado na página [24](#).

DETALHES TÉCNICOS DA DISTRIBUIÇÃO MULTIVARIADA ZICOMP TIPO I

Neste Apêndice, apresentamos alguns detalhes técnicos da Seção 2.2, do Capítulo 2, tais como, os casos especiais da distribuição; função geradora de momentos; esperança; variância e as distribuições condicionais.

A.1 Casos especiais da distribuição multivariada ZICOMP Tipo I

(i) Se $\phi \neq 0$ e $\forall v_i = 1$ elemento do vetor \mathbf{v} , então $\mathbf{Y} \sim \text{ZIP}_m^{(I)}(\phi, \boldsymbol{\lambda})$;

Considere λ_i e v_i elementos dos vetores $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top$ e $\mathbf{v} = (v_1, \dots, v_m)^\top$ e seja $S(\lambda, \mathbf{v}) = \sum_{j=0}^{\infty} \lambda^j / (j!)^v$. Então para $v_i = 1$, $S(\lambda_i, v_i) = \exp(\lambda_i)$. Então para a Equação (2.2) a função de massa de probabilidade para Y_i o i -ésimo elemento do vetor $\mathbf{Y} = (Y_1, \dots, Y_m)$ é dada por

$$f(y_i | \phi, \lambda_i, v_i) = \left[\phi + (1 - \phi)e^{-\lambda_i} \right] I(y_i = 0) + (1 - \phi) \left[\frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \right] I(y_i > 0),$$

logo $Y_i \sim \text{ZIP}(\phi, \lambda_i)$. Então $\forall, Y_i \in \mathbf{Y}$ e $v_i = 1$ elemento de \mathbf{v} , a função de massa de probabilidade conjunta do vetor \mathbf{Y} para a Equação (2.2) é dada por

$$\begin{aligned} f(\mathbf{y} | \phi, \boldsymbol{\lambda}, \mathbf{v}) &= \phi \Pr(\boldsymbol{\xi} = \mathbf{y}) + (1 - \phi) \Pr(\mathbf{X} = \mathbf{y}) \\ &= \left[\phi + (1 - \phi) \prod_{i=1}^m e^{-\lambda_i} \right] I(\mathbf{y} = \mathbf{0}) + (1 - \phi) \left[\prod_{i=1}^m \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \right] I(\mathbf{y} \neq \mathbf{0}). \end{aligned} \quad (\text{A.1})$$

$$\therefore \mathbf{Y} \sim \text{ZIP}_m^{(I)}(\phi, \boldsymbol{\lambda}).$$

(ii) Se $\phi \neq 0$ e $(v_i = 0, \lambda_i < 1)$, $\forall \lambda_i$ e v_i elementos dos vetores $\boldsymbol{\lambda}$ e \mathbf{v} , então $\mathbf{Y} \sim \text{ZIG}_m^{(I)}(\phi, \mathbf{1} - \boldsymbol{\lambda})$;

Considere $0 < \lambda_i < 1$ e $v_i = 0$, tal que λ_i é um elemento de $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top$ e v_i é um elemento de $\mathbf{v} = (v_1, \dots, v_m)^\top$, com $S(\boldsymbol{\lambda}, \mathbf{v}) = \sum_{j=0}^{\infty} \lambda^j / (j!)^v$. Então para $0 < \lambda_i < 1$ e $v_i = 0$, $S(\lambda_i, v_i) = 1/(1 - \lambda_i)$. Então para a Equação (2.2) a função de massa de probabilidade para Y_i o i -ésimo elemento do vetor $\mathbf{Y} = (Y_1, \dots, Y_m)$ é dada por

$$f(y_i | \phi, \lambda_i, v_i) = [\phi + (1 - \phi)(1 - \lambda_i)] I(y_i = 0) + (1 - \phi) [\lambda_i^{y_i} (1 - \lambda_i)] I(y_i > 0),$$

logo $Y_i \sim \text{ZIG}(\phi, 1 - \lambda_i)$, ou seja Y_i segue a distribuição Geométrica zero-inflacionado (ZIG) com parâmetros ϕ e $1 - \lambda_i$. Então $\forall, 0 < \lambda_i < 1$ e $v_i = 0$, tal que λ_i e v_i são elementos dos vetores $\boldsymbol{\lambda}$ e \mathbf{v} , a função de massa de probabilidade conjunta do vetor \mathbf{Y} para a Equação (2.2) é dada por

$$\begin{aligned} f(\mathbf{y} | \phi, \boldsymbol{\lambda}, \mathbf{v}) &= \phi \Pr(\boldsymbol{\xi} = \mathbf{y}) + (1 - \phi) \Pr(\mathbf{X} = \mathbf{y}) \\ &= \left[\phi + (1 - \phi) \prod_{i=1}^m (1 - \lambda_i) \right] I(\mathbf{y} = \mathbf{0}) + (1 - \phi) \left[\prod_{i=1}^m \lambda_i^{y_i} (1 - \lambda_i) \right] I(\mathbf{y} \neq \mathbf{0}). \end{aligned} \quad (\text{A.2})$$

$$\therefore \mathbf{Y} \sim \text{ZIG}_m^{(I)}(\phi, \mathbf{1} - \boldsymbol{\lambda}).$$

(iii) Se $\phi \neq 0$ e $v_i \rightarrow \infty$, $\forall v_i$ elemento do vetor \mathbf{v} , então $\mathbf{Y} \sim \text{ZIBernoulli}_m^{(I)}(\phi, \boldsymbol{\lambda} \odot (\mathbf{1} + \boldsymbol{\lambda}))$;

Considere $v_i \rightarrow \infty$, tal que v_i elemento do vetor $\mathbf{v} = (v_1, \dots, v_m)^\top$, com $S(\lambda_i, v_i) \rightarrow 1 + \lambda_i$, quando $v_i \rightarrow \infty$. Então para a Equação (2.2) a função de massa de probabilidade para Y_i o i -ésimo elemento do vetor $\mathbf{Y} = (Y_1, \dots, Y_m)$ é dada por

$$f(y_i | \phi, \lambda_i, v_i) = \left[\phi + \frac{(1 - \phi)}{(1 + \lambda_i)} \right] I(y_i = 0) + (1 - \phi) \left[\frac{\lambda_i}{(1 + \lambda_i)} \right] I(y_i > 0),$$

logo $Y_i \sim \text{ZIBernoulli}(\phi, \frac{\lambda_i}{1 + \lambda_i})$, ou seja, Bernoulli zero-inflacionado (ZIBernoulli) com parâmetros ϕ e $\frac{\lambda_i}{1 + \lambda_i}$. Então $\forall v_i$ elemento de $\mathbf{v} = (v_1, \dots, v_m)^\top$, com $S(\lambda_i, v_i) \rightarrow 1 + \lambda_i$, quando $v_i \rightarrow \infty$. A função de massa de probabilidade conjunta do vetor \mathbf{Y} em (2.2) é dada por

$$\begin{aligned} f(\mathbf{y} | \phi, \boldsymbol{\lambda}, \mathbf{v}) &= \phi \Pr(\boldsymbol{\xi} = \mathbf{y}) + (1 - \phi) \Pr(\mathbf{X} = \mathbf{y}) \\ &= \left[\phi + (1 - \phi) \prod_{i=1}^m \frac{1}{(1 + \lambda_i)} \right] I(\mathbf{y} = \mathbf{0}) + (1 - \phi) \left[\prod_{i=1}^m \frac{\lambda_i}{(1 + \lambda_i)} \right] I(\mathbf{y} \neq \mathbf{0}). \end{aligned} \quad (\text{A.3})$$

$$\therefore \mathbf{Y} \sim \text{ZIBernoulli}_m^{(I)}(\phi, \boldsymbol{\lambda} \odot (\mathbf{1} + \boldsymbol{\lambda})).$$

A.2 Momentos

Demonstração da função geradora de momentos para o vetor $\mathbf{Y} \sim \text{ZICOMP}_m^{(I)}(\phi, \boldsymbol{\lambda}, \mathbf{v})$

$$\begin{aligned}
\mathbf{M}_{\mathbf{Y}}(\mathbf{t}) &= \mathbb{E} \left[\exp(\mathbf{t}^\top \mathbf{Y}) \right] = \mathbb{E} \left[\exp(\mathbf{Z} \cdot \mathbf{t}^\top \mathbf{X}) \right] = \mathbb{E} \left[\prod_{i=1}^m M_{X_i}(t_i Z) \right] \\
&= \phi \prod_{i=1}^m M_{X_i}(0) + (1 - \phi) \prod_{i=1}^m M_{X_i}(t_i) \\
&= \phi + (1 - \phi) \prod_{i=1}^m \left\{ \frac{1}{S(\lambda_i, v_i)} \sum_{y_i=1}^{\infty} e^{t_i x_i} \frac{\lambda_i^{x_i}}{(x_i!)^{v_i}} \right\} \\
&= \phi + (1 - \phi) \prod_{i=1}^m \left\{ \frac{1}{S(\lambda_i, v_i)} \sum_{x_i=1}^{\infty} \frac{(\lambda_i e^{t_i})^{x_i}}{(x_i!)^{v_i}} \right\} \\
&= \phi + (1 - \phi) \prod_{i=1}^m \left\{ \frac{S(\lambda_i e^{t_i}, v_i)}{S(\lambda_i, v_i)} \right\}. \tag{A.4}
\end{aligned}$$

$E(Y_i)$

Considere $S(\lambda_i, v_i) = \sum_{j=0}^{\infty} \frac{\lambda_i^j}{(j!)^{v_i}}$, tal que

$$\frac{\partial S(\lambda_i, v_i)}{\partial \lambda_i} = \sum_{j=1}^{\infty} \frac{j \lambda_i^{j-1}}{(j!)^{v_i}} = \sum_{t_i=0}^{\infty} \frac{\lambda_i^{t_i} (t_i + 1)^{1-v_i}}{(t_i!)^{v_i}}, \tag{A.5}$$

e tome $h(t_i) = \lambda_i e^{t_i}$, tal que $\frac{\partial S(\lambda_i e^{t_i}, v_i)}{\partial t_i} = \frac{\partial S(h(t_i), v_i)}{\partial \lambda_i} \frac{\partial h(t_i)}{\partial t_i}$, então substituindo $\frac{\partial S(\lambda_i, v_i)}{\partial t_i}$ e (A.4) em (A.5), temos

$$\begin{aligned}
E(Y_i) &= \frac{\partial \mathbf{M}_{Y_i}(t_i)}{\partial t_i} \Big|_{t_i=0} = (1 - \phi) \left\{ \frac{\partial S(\lambda_i, v_i) | \partial \lambda_i}{S(\lambda_i, v_i)} \lambda_i \right\} \\
&= (1 - \phi) \left\{ \lambda_i \frac{\partial \log S(\lambda_i, v_i)}{\partial \lambda_i} \right\}. \tag{A.6}
\end{aligned}$$

$E(Y_i^2)$

Similarmente considere $S(\lambda_i, v_i)$, tal que

$$\frac{\partial^2 S(\lambda_i, v_i)}{\partial \lambda_i^2} = \sum_{j=2}^{\infty} \frac{j^2 \lambda_i^{j-2}}{(j!)^{v_i}} = \sum_{t_i=0}^{\infty} \frac{\lambda_i^{t_i} [(t_i + 1)(t_i + 2)]^{1-v_i}}{(t_i!)^{v_i}}, \tag{A.7}$$

e seja $\frac{\partial^2 S(\lambda_i, v_i)}{\partial t_i^2} = \frac{\partial^2 S(h(t_i), v_i)}{\partial \lambda_i^2} \frac{\partial h(t_i)}{\partial t_i} + \frac{\partial S(h(t_i), v_i)}{\partial \lambda_i} \frac{\partial^2 h(t_i)}{\partial t_i^2}$, então substituindo $\frac{\partial^2 S(\lambda_i, v_i)}{\partial t_i^2}$ e A.7 em A.4, temos

$$E(Y_i^2) = \frac{\partial^2 \mathbf{M}_{Y_i}(t_i)}{\partial t_i^2} \Big|_{t_i=0} = (1 - \phi) \left\{ \lambda_i^2 \frac{\partial^2 S(\lambda_i, v_i) | \partial \lambda_i^2}{S(\lambda_i, v_i)} + \lambda_i \frac{\partial \log S(\lambda_i, v_i)}{\partial \lambda_i} \right\}, \tag{A.8}$$

Var(Y_i)

Considere $\text{Var}(Y_i) = E(Y_i^2) - E^2(Y_i)$, então

$$\text{Var}(Y_i) = (1 - \phi) \left\{ \lambda_i^2 \left[\frac{\partial^2 S(\lambda_i, \mathbf{v}_i) / \partial \lambda_i^2}{S(\lambda_i, \mathbf{v}_i)} - (1 - \phi) \left(\frac{\partial \log S(\lambda_i, \mathbf{v}_i)}{\partial \lambda_i} \right)^2 \right] + \frac{\partial \log S(\lambda_i, \mathbf{v}_i)}{\partial \lambda_i} \right\}. \quad (\text{A.9})$$

Cov(Y_i, Y_j)

$$\begin{aligned} E(Y_i, Y_j) &= \frac{\partial^2 M_{(Y_i, Y_j)}(0, 0)}{\partial t_i \partial t_j} \\ &= (1 - \phi) \left\{ \lambda_i \frac{\partial S(\lambda_i, \mathbf{v}_i) / \partial \lambda_i}{S(\lambda_i, \mathbf{v}_i)} \right\} \left\{ \lambda_j \frac{\partial S(\lambda_j, \mathbf{v}_j) / \partial \lambda_j}{S(\lambda_j, \mathbf{v}_j)} \right\} \\ &= (1 - \phi) \mu_i \mu_j, \end{aligned} \quad (\text{A.10})$$

e como $\text{Cov}(Y_i, Y_j) = E(Y_i, Y_j) - E(Y_i)E(Y_j)$, então

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= E(Y_i, Y_j) - E(Y_i)E(Y_j) \\ &= \phi(1 - \phi) \mu_i \mu_j \\ &= \phi(1 - \phi) \left\{ \lambda_i \frac{\partial S(\lambda_i, \mathbf{v}_i) / \partial \lambda_i}{S(\lambda_i, \mathbf{v}_i)} \right\} \left\{ \lambda_j \frac{\partial S(\lambda_j, \mathbf{v}_j) / \partial \lambda_j}{S(\lambda_j, \mathbf{v}_j)} \right\}. \end{aligned} \quad (\text{A.11})$$

A.3 Distribuições condicionais

A.3.1 Distribuição condicional $\mathbf{Y}^{(1)} \mid \mathbf{Y}^{(2)}$

Caso II da distribuição condicional $\mathbf{Y}^{(1)} \mid \mathbf{Y}^{(2)}$ em (2.2.4.1). Podemos ter $\mathbf{Y}^{(1)} = \mathbf{0} \mid \mathbf{Y}^{(2)} = \mathbf{0}$, ou $\mathbf{Y}^{(1)} \neq \mathbf{0} \mid \mathbf{Y}^{(2)} = \mathbf{0}$, quando $\mathbf{y}^{(1)} = \mathbf{0}$, temos

$$\begin{aligned} \Pr(\mathbf{Y}^{(1)} = \mathbf{0} \mid \mathbf{Y}^{(2)} = \mathbf{0}) &= \frac{\Pr(\mathbf{Y}^{(1)} = \mathbf{0}, \mathbf{Y}^{(2)} = \mathbf{0})}{\Pr(\mathbf{Y}^{(2)} = \mathbf{0})} \\ &= \frac{\phi + \frac{(1-\phi)}{S(\boldsymbol{\lambda}, \mathbf{v})}}{\phi + \frac{(1-\phi)}{S^{(2)}(\boldsymbol{\lambda}, \mathbf{v})}} = \frac{\phi}{\phi + \frac{(1-\phi)}{S^{(2)}(\boldsymbol{\lambda}, \mathbf{v})}} + \frac{\frac{(1-\phi)}{S(\boldsymbol{\lambda}, \mathbf{v})}}{\phi + \frac{(1-\phi)}{S^{(2)}(\boldsymbol{\lambda}, \mathbf{v})}} \\ &= \phi^* + \frac{(1 - \phi^*)}{S^{(1)}(\boldsymbol{\lambda}, \mathbf{v})}, \end{aligned} \quad (\text{A.12})$$

em que $S(\boldsymbol{\lambda}, \mathbf{v}) = \prod_{i=1}^m S(\lambda_i, v_i)$; $S^{(1)}(\boldsymbol{\lambda}, \mathbf{v}) = \prod_{i=1}^r S(\lambda_i, v_i)$; $S^{(2)}(\boldsymbol{\lambda}, \mathbf{v}) = \prod_{i=r+1}^m S(\lambda_i, v_i)$, $\phi^* = \frac{\phi S^{(2)}(\boldsymbol{\lambda}, \mathbf{v})}{\phi S^{(2)}(\boldsymbol{\lambda}, \mathbf{v}) + (1-\phi)}$ e $(1 - \phi^*) = \frac{(1-\phi)}{\phi S^{(2)}(\boldsymbol{\lambda}, \mathbf{v}) + (1-\phi)}$.

Quando $\mathbf{y}^{(1)} \neq \mathbf{0}$ e $\mathbf{y}^{(2)} = \mathbf{0}$, para a Equação (2.7) temos

$$\begin{aligned} \Pr(\mathbf{Y}^{(1)} \neq \mathbf{0} | \mathbf{Y}^{(2)} = \mathbf{0}) &= \frac{\Pr(\mathbf{Y}^{(1)} \neq \mathbf{0}, \mathbf{Y}^{(2)} = \mathbf{0})}{\Pr(\mathbf{Y}^{(2)} = \mathbf{0})} \\ &= \frac{\frac{(1-\phi)}{S(\boldsymbol{\lambda}, \mathbf{v})} \prod_{i=1}^r \frac{\lambda_i^{y_i}}{(y_i!)^{v_i}}}{\phi + \frac{(1-\phi)}{S^{(2)}(\boldsymbol{\lambda}, \mathbf{v})}} = (1 - \phi^*) \prod_{i=1}^r \frac{\lambda_i^{y_i}}{(y_i!)^{v_i} S(\lambda_i, v_i)}, \end{aligned} \quad (\text{A.13})$$

combinando (A.12) com (A.13), e substituindo em (2.2), temos

$$\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{Y}^{(2)} = \mathbf{0}) = \left[\phi^* + \frac{(1-\phi^*)}{S^{(1)}(\boldsymbol{\lambda}, \mathbf{v})} \right] I_{(\mathbf{y}=\mathbf{0})} + \left[(1 - \phi^*) \prod_{i=1}^r \frac{\lambda_i^{y_i}}{(y_i!)^{v_i} S(\lambda_i, v_i)} \right] I_{(\mathbf{y} \neq \mathbf{0})}, \quad (\text{A.14})$$

então $\mathbf{Y} | (\mathbf{Y}^{(2)} = \mathbf{0}) \sim \text{ZICOMP}^{(1)}(\phi^*; \lambda_1, \dots, \lambda_r; v_1, \dots, v_r)$

A.3.2 Distribuição condicional $\mathbf{X} | \mathbf{Y}$

Prova da distribuição $\mathbf{X} | \mathbf{Y}$ quando $\mathbf{y} = \mathbf{0}$

$$\begin{aligned} \Pr(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{0}) &= \frac{\Pr(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{0})}{\Pr(\mathbf{Y} = \mathbf{0})} \\ &= \frac{\Pr(\mathbf{X} = \mathbf{0}, Z\mathbf{X} = \mathbf{0})}{f(\mathbf{0}, \phi, \boldsymbol{\lambda}, \mathbf{v})} I_{(\mathbf{x}=\mathbf{0})} + \frac{\Pr(\mathbf{X} = \mathbf{x}, Z\mathbf{X} = \mathbf{x})}{f(\mathbf{0}, \phi, \boldsymbol{\lambda}, \mathbf{v})} I_{(\mathbf{x} \neq \mathbf{0})} \\ &= \frac{\Pr(\mathbf{X} = \mathbf{0})}{f(\mathbf{0}, \phi, \boldsymbol{\lambda}, \mathbf{v})} I_{(\mathbf{x}=\mathbf{0})} + \frac{\Pr(\mathbf{X} = \mathbf{x}, Z = 0)}{f(\mathbf{0}, \phi, \boldsymbol{\lambda}, \mathbf{v})} I_{(\mathbf{x} \neq \mathbf{0})} \\ &= \frac{\Pr(\mathbf{X} = \mathbf{0})}{f(\mathbf{0}, \phi, \boldsymbol{\lambda}, \mathbf{v})} I_{(\mathbf{x}=\mathbf{0})} + \frac{\phi \Pr(\mathbf{X} = \mathbf{x})}{f(\mathbf{0}, \phi, \boldsymbol{\lambda}, \mathbf{v})} I_{(\mathbf{x} \neq \mathbf{0})} \\ &= \frac{\frac{1}{S(\boldsymbol{\lambda}, \mathbf{v})}}{\phi + \frac{(1-\phi)}{S(\boldsymbol{\lambda}, \mathbf{v})}} I_{(\mathbf{x}=\mathbf{0})} + \frac{\phi \prod_{i=1}^m \left[\frac{\lambda_i^{y_i}}{(y_i!)^{v_i} S(\lambda_i, v_i)} \right]}{\phi + \frac{(1-\phi)}{S(\boldsymbol{\lambda}, \mathbf{v})}} I_{(\mathbf{x} \neq \mathbf{0})} \\ &= \left[\psi + \frac{(1-\psi)}{S(\boldsymbol{\lambda}, \mathbf{v})} \right] I_{(\mathbf{x}=\mathbf{0})} + (1-\psi) \prod_{i=1}^m \left[\frac{\lambda_i^{y_i}}{(y_i!)^{v_i} S(\lambda_i, v_i)} \right] I_{(\mathbf{x} \neq \mathbf{0})}, \end{aligned} \quad (\text{A.15})$$

ou seja,

$$\mathbf{X} | (\mathbf{Y} = \mathbf{0}) \sim \text{ZICOMP}^{(1)}(\psi; \lambda_1, \dots, \lambda_m; v_1, \dots, v_m) \quad (\text{A.16})$$

em que $\psi = \frac{(1-\phi)}{\phi S(\boldsymbol{\lambda}, \mathbf{v}) + (1-\phi)}$ e $(1 - \psi) = \frac{\phi}{\phi S(\boldsymbol{\lambda}, \mathbf{v}) + (1-\phi)}$.

DETALHES TÉCNICOS DA DISTRIBUIÇÃO MULTIVARIADA ZICOMP TIPO III

B.1 Casos especiais da distribuição multivariada ZICOMP Tipo III

(i) Se $\phi_0 \neq 0$, $\boldsymbol{\phi} = \mathbf{0}$, e $\mathbf{v} = \mathbf{1}$, então $\mathbf{Y} \sim \text{ZIP}_m^{(I)}(\phi_0, \boldsymbol{\lambda})$;

Considere $\phi_0 \neq 0$; $\forall \phi_i = 0$, λ_i e v_i elementos dos vetores $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)^\top$; $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top$ e $\mathbf{v} = (v_1, \dots, v_m)^\top$ e seja $S(\boldsymbol{\lambda}, \mathbf{v}) = \sum_{j=0}^{\infty} \lambda^j / (j!)^v$, para $v_i = 1$, $S(\lambda_i, v_i) = \exp(\lambda_i)$. Então para a Equação (3.2) a função de massa de probabilidade para Y_i o i -ésimo elemento do vetor $\mathbf{Y} = (Y_1, \dots, Y_m)$ é dada por

$$f(y_i | \phi_0, \phi_i, \lambda_i, v_i) = \left\{ \phi_0 + (1 - \phi_0)e^{-\lambda_i} \right\} I(y_i = 0) + (1 - \phi_0) \left\{ \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \right\} I(y_i \neq 0),$$

logo $Y_i \sim \text{ZIP}(\phi_0, \lambda_i)$. Então $\forall, Y_i \in \mathbf{Y}$; $\phi_0 \neq 0$; $\forall \phi_i = 0$ e v_i elementos dos vetores $\boldsymbol{\phi}$ e \mathbf{v} , a função de massa de probabilidade conjunta do vetor \mathbf{Y} para a Equação (3.2) é dada por

$$f(\mathbf{y} | \phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v}) = \prod_{i=1}^m \left\{ \phi_0 + (1 - \phi_0)e^{-\lambda_i} \right\} I(\mathbf{y} = \mathbf{0}) + (1 - \phi_0) \prod_{i=1}^m \left\{ \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \right\} I(\mathbf{y} \neq \mathbf{0}), \quad (\text{B.1})$$

$$\therefore \mathbf{Y} \sim \text{ZIP}_m^{(I)}(\phi_0, \boldsymbol{\lambda}).$$

(ii) Se $\phi_0 \neq 0$ e $\boldsymbol{\phi} = \mathbf{0}$, então $\mathbf{Y} \sim \text{ZICOMP}_m^{(I)}(\phi_0, \boldsymbol{\lambda}, \mathbf{v})$;

Considere $\phi_0 \neq 0$; $\forall \phi_i = 0$, λ_i e v_i elementos dos vetores $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)^\top$; $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top$ e $\mathbf{v} = (v_1, \dots, v_m)^\top$ e seja $S(\boldsymbol{\lambda}, \mathbf{v}) = \sum_{j=0}^{\infty} \lambda^j / (j!)^v$. Então para a Equação (3.2) a

função de massa de probabilidade para Y_i o i -ésimo elemento do vetor $\mathbf{Y} = (Y_1, \dots, Y_m)$ é dada por

$$f(y_i|\phi_0, \phi_i, \lambda_i, \nu_i) = \left\{ \phi_0 + \frac{(1-\phi_0)}{S(\lambda_i, \nu_i)} \right\} I(y_i = 0) + (1-\phi_0) \left\{ \frac{\lambda_i^{y_i}}{S(\lambda_i, \nu_i)(y_i!)^{\nu_i}} \right\} I(y_i \neq 0),$$

logo $Y_i \sim \text{ZICOMP}(\phi_0, \lambda_i)$. Então $\forall, Y_i \in \mathbf{Y}; \phi_0 \neq 0; \forall \phi_i = 0$ elemento de $\boldsymbol{\phi}$, a função de massa de probabilidade conjunta do vetor \mathbf{Y} para a Equação (3.2) é dada por

$$f(\mathbf{y}|\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v}) = \prod_{i=1}^m \left\{ \phi_0 + \frac{(1-\phi_0)}{S(\lambda_i, \nu_i)} \right\} I(\mathbf{y} = \mathbf{0}) + (1-\phi_0) \prod_{i=1}^m \left\{ \frac{\lambda_i^{y_i}}{S(\lambda_i, \nu_i)(y_i!)^{\nu_i}} \right\} I(\mathbf{y} \neq \mathbf{0}), \quad (\text{B.2})$$

$$\therefore \mathbf{Y} \sim \text{ZICOMP}_m^{(1)}(\phi_0, \boldsymbol{\lambda}, \mathbf{v}).$$

(iii) $\phi_0 \neq 0, \boldsymbol{\phi} = \mathbf{0}$, e $(\nu_i = 0, \lambda_i < 1)$ então $\mathbf{Y} \sim \text{Geométrica}_m^{(1)}(\phi_0, \mathbf{1} - \boldsymbol{\lambda})$;

Considere $\phi_0 \neq 0; \phi_i = 0; 0 < \lambda_i < 1$ e $\nu_i = 0$, tal que ϕ_i, λ_i e ν_i são elementos dos vetores $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)^\top$; $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top$ e $\mathbf{v} = (\nu_1, \dots, \nu_m)^\top$ e seja $S(\boldsymbol{\lambda}, \mathbf{v}) = \sum_{j=0}^{\infty} \lambda^j / (j!)^{\nu}$. Então para $0 < \lambda_i < 1$ e $\nu_i = 0$, $S(\lambda_i, \nu_i) = 1/(1 - \lambda_i)$. Então para a Equação (3.2) a função de massa de probabilidade para Y_i o i -ésimo elemento do vetor $\mathbf{Y} = (Y_1, \dots, Y_m)$ é dada por

$$f(y_i|\phi_0, \phi_i, \lambda_i, \nu_i) = \{ \phi_0 + (1-\phi_0)(1-\lambda_i) \} I(y_i = 0) + (1-\phi_0) \{ \lambda_i^{y_i} (1-\lambda_i) \} I(y_i \neq 0),$$

logo $Y_i \sim \text{ZIG}(\phi_0, 1 - \lambda_i)$. Então $\forall, \phi_i = 0; 0 < \lambda_i < 1$ e $\nu_i = 0$, tal que ϕ_i, λ_i e ν_i são elementos dos vetores $\boldsymbol{\phi}; \boldsymbol{\lambda}; \mathbf{v}$, com $\phi_0 \neq 0$ a função de massa de probabilidade conjunta do vetor \mathbf{Y} para a Equação (3.2) é dada por

$$f(\mathbf{y}|\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v}) = \left\{ \phi_0 + (1-\phi_0) \prod_{i=1}^m (1-\lambda_i) \right\} I(\mathbf{y} = \mathbf{0}) + (1-\phi_0) \left\{ \prod_{i=1}^m \lambda_i^{y_i} (1-\lambda_i) \right\} I(\mathbf{y} \neq \mathbf{0}). \quad (\text{B.3})$$

$$\therefore \mathbf{Y} \sim \text{ZIG}_m^{(1)}(\phi_0, \mathbf{1} - \boldsymbol{\lambda}).$$

(iv) Se $\phi_0 \neq 0, \boldsymbol{\phi} = \mathbf{0}$, e $(\nu_i \rightarrow \infty, \lambda_i < 1)$ então $\mathbf{Y} \sim \text{Bernoulli}_m^{(1)}(\phi_0, \boldsymbol{\lambda} \odot (\mathbf{1} + \boldsymbol{\lambda}))$;

Considere $\phi_0 \neq 0$ e $\forall \phi_i = 0$ e $\nu_i \rightarrow \infty$ elementos dos vetores $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)^\top$ e $\mathbf{v} = (\nu_1, \dots, \nu_m)^\top$, tal que $S(\lambda_i, \nu_i) \rightarrow 1 + \lambda_i$, quando $\nu_i \rightarrow \infty$. Então a probabilidade de sucesso para $Y_i = 1$ o i -ésimo elemento do vetor $\mathbf{Y} = (Y_1, \dots, Y_m)$, a Equação (3.2) é dada por

$$f(y_i = 1|\phi_0, \phi_i, \lambda_i, \nu_i) = \left\{ \phi_0 + \frac{(1-\phi_0)}{(1+\lambda_i)} \right\} I(y_i = 0) + (1-\phi_0) \left\{ \frac{\lambda_i}{(1+\lambda_i)} \right\} I(y_i \neq 0),$$

logo $Y_i \sim \text{ZIBernoulli}(\phi, \frac{\lambda_i}{1+\lambda_i})$. Então $\forall \nu_i \rightarrow \infty$ elemento de $\mathbf{v} = (\nu_1, \dots, \nu_m)^\top$ tal que $S(\lambda_i, \nu_i) \rightarrow 1 + \lambda_i$ quando $\nu_i \rightarrow \infty$, e $\forall \phi_i = 0$ elemento de $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)^\top$ e $\phi_0 \neq 0$. A função de massa de probabilidade conjunta do vetor \mathbf{Y} em (3.2) é dada por

$$f(\mathbf{y}|\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v}) = \left\{ \phi_0 + (1 - \phi_0) \prod_{i=1}^m \frac{1}{(1 + \lambda_i)} \right\} I(\mathbf{y} = \mathbf{0}) + \left[(1 - \phi_0) \prod_{i=1}^m \frac{\lambda_i}{(1 + \lambda_i)} \right] I(\mathbf{y} \neq \mathbf{0}). \quad (\text{B.4})$$

$$\therefore \mathbf{Y} \sim \text{ZIBernoulli}_m^{(I)}(\phi_0, \boldsymbol{\lambda} \odot (\mathbf{1} + \boldsymbol{\lambda})).$$

(v) Se $\phi_0 \neq 0$, $\boldsymbol{\phi} \neq \mathbf{0}$, e $\mathbf{v} = \mathbf{1}$, então $\mathbf{Y} \sim \text{ZIP}_m^{(III)}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda})$;

Considere $\phi_0 \neq 0$, e ϕ_i ; λ_i e v_i elementos dos vetores $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)^\top$; $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top$ e $\mathbf{v} = (v_1, \dots, v_m)^\top$, e seja $S(\boldsymbol{\lambda}, \mathbf{v}) = \sum_{j=0}^{\infty} \lambda^j / (j!)^v$ tal que $S(\lambda_i, v_i) = \exp(\lambda_i)$ quando $v_i = 1$. Então para $\phi_i \neq 0$ e $v_i = 1$ elementos dos vetores $\boldsymbol{\phi}$ e \mathbf{v} , com $\phi_0 \neq 0$, a função de massa de probabilidade conjunta para a Equação (3.2) para Y_i o i -ésimo elemento do vetor $\mathbf{Y} = (Y_1, \dots, Y_m)$ é dada por

$$f(y_i|\phi_0, \phi_i, \lambda_i, v_i) = \left\{ \phi_0 + (1 - \phi_0)[\phi_i + (1 - \phi_i)e^{-\lambda_i}] \right\} I(y_i = 0) + (1 - \phi_0) \left\{ (1 - \phi_i) \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \right\} I(y_i \neq 0),$$

logo $Y_i \sim \text{ZIP}(\phi_0, \phi_i, \lambda_i)$. Então $\forall Y_i \in \mathbf{Y}$ com $\phi_0 \neq 0$, e $\forall \phi_i = 0$ e $v_i = 1$ elementos dos vetores $\boldsymbol{\phi}$ e \mathbf{v} a função de massa de probabilidade conjunta do vetor \mathbf{Y} para a Equação (3.2) é dada por

$$\begin{aligned} f(\mathbf{y}|\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v}) &= \prod_{i=1}^m \left\{ \phi_0 + [\phi_i + (1 - \phi_i)e^{-\lambda_i}] \right\} I(\mathbf{y} = \mathbf{0}) + (1 - \phi_0) \\ &\times \prod_{i=1}^m \left\{ [\phi_i + (1 - \phi_i)e^{-\lambda_i}] I_{(y_i=0)} + [(1 - \phi_i) \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!}] I_{(y_i>0)} \right\} I(\mathbf{y} \neq \mathbf{0}), \end{aligned} \quad (\text{B.5})$$

$$\therefore \mathbf{Y} \sim \text{ZIP}_m^{(III)}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}).$$

(vi) Se $\phi_0 \neq 0$, $\boldsymbol{\phi} \neq \mathbf{0}$, e $(v_i = 0, \lambda_i < 1)$ então $\mathbf{Y} \sim \text{ZIG}_m^{(III)}(\phi_0, \boldsymbol{\phi}, \mathbf{1} - \boldsymbol{\lambda})$;

Considere $\phi_0 \neq 0$, e $\phi_i \neq 0$; $0 < \lambda_i < 1$ e $v_i = 0$ elementos dos vetores $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)^\top$; $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top$ e $\mathbf{v} = (v_1, \dots, v_m)^\top$. Para $0 < \lambda_i < 1$ e $v_i = 0$ isso implica $S(\lambda_i, v_i) = 1/(1 - \lambda_i)$, tal que $S(\boldsymbol{\lambda}, \mathbf{v}) = \sum_{j=0}^{\infty} \lambda^j / (j!)^v$. Então para a Equação (3.2) a função de massa de probabilidade para Y_i o i -ésimo elemento do vetor $\mathbf{Y} = (Y_1, \dots, Y_m)$ é dada por

$$\begin{aligned} f(y_i|\phi_0, \phi_i, \lambda_i, v_i) &= \left\{ \phi_0 + (1 - \phi_0)[\phi_i + (1 - \phi_i)(1 - \lambda_i)] \right\} I(y_i = 0) + (1 - \phi_0) \\ &\times \left\{ (1 - \phi_i) \lambda_i^{y_i} (1 - \lambda_i) \right\} I(y_i \neq 0), \end{aligned}$$

logo $Y_i \sim \text{ZIG}(\phi_0, \phi_i, 1 - \lambda_i)$. Então $\forall \phi_i \neq 0$, $0 < \lambda_i < 1$ e $v_i = 0$ elementos dos vetores $\boldsymbol{\phi}$, $\boldsymbol{\lambda}$ e \mathbf{v} , com $\phi_0 \neq 0$ a função de massa de probabilidade conjunta do vetor \mathbf{Y} para a Equação (3.2) é dada por

$$f(\mathbf{y}|\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v}) = \left\{ \phi_0 + (1 - \phi_0) \prod_{i=1}^m [\phi_i + (1 - \phi_i)(1 - \lambda_i)] \right\} I(\mathbf{y} = \mathbf{0}) + (1 - \phi_0) \times \prod_{i=1}^m \{ [\phi_i + (1 - \phi_i)(1 - \lambda_i)] I_{(y_i=0)} + [(1 - \phi_i)\lambda_i^{y_i}(1 - \lambda_i)] I_{(y_i>0)} \} I(\mathbf{y} \neq \mathbf{0}). \quad (\text{B.6})$$

$$\therefore \mathbf{Y} \sim \text{ZIG}_m^{(\text{III})}(\phi_0, \boldsymbol{\phi}, \mathbf{1} - \boldsymbol{\lambda}).$$

(vii) Se $\phi_0 \neq 0$, $\boldsymbol{\phi} \neq \mathbf{0}$, e $(v_i \rightarrow \infty, \lambda_i < 1)$ então $\mathbf{Y} \sim \text{ZIBernoulli}_m^{(\text{III})}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda} \odot (\mathbf{1} + \boldsymbol{\lambda}))$.

Considere $\phi_0 \neq 0$; $\phi_i \neq 0$ e $v_i \rightarrow \infty$ elementos dos vetores $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)^\top$ e $\mathbf{v} = (v_1, \dots, v_m)^\top$. E seja $S(\lambda_i, v_i) \rightarrow 1 + \lambda_i$ quando $v_i \rightarrow \infty$. Então para a Equação (3.2) a função de massa de probabilidade para Y_i o i -ésimo elemento do vetor $\mathbf{Y} = (Y_1, \dots, Y_m)$ é dada por

$$f(y_i|\phi_0, \phi_i, \lambda_i, v_i) = \left\{ \phi_0 + (1 - \phi_0) \left[\phi_i + \frac{(1 - \phi_i)}{(1 + \lambda_i)} \right] \right\} I(y_i = 0) + (1 - \phi_0) \times \left\{ \phi_i + (1 - \phi_i) \frac{\lambda_i}{(1 + \lambda_i)} \right\} I(y_i \neq 0),$$

logo $Y_i \sim \text{ZIBernoulli}(\phi_0, \phi_i, \frac{\lambda_i}{1 + \lambda_i})$. Então para $\phi_0 \neq 0$, e $\forall \phi_i$ e $v_i \rightarrow \infty$, elementos dos vetores $\boldsymbol{\phi}$ e \mathbf{v} . A função de massa de probabilidade conjunta do vetor \mathbf{Y} em (3.2) é dada por

$$f(\mathbf{y}|\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda}, \mathbf{v}) = \left\{ \phi_0 + (1 - \phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1 - \phi_i)}{(1 + \lambda_i)} \right] \right\} I(\mathbf{y} = \mathbf{0}) + (1 - \phi_0) \times \prod_{i=1}^m \left\{ \left[\phi_i + \frac{(1 - \phi_i)}{(1 + \lambda_i)} \right] I_{y_i=0} + \left[\frac{(1 - \phi_i)\lambda_i}{(1 + \lambda_i)} \right] I_{y_i>0} \right\} I(\mathbf{y} \neq \mathbf{0}). \quad (\text{B.7})$$

$$\therefore \mathbf{Y} \sim \text{ZIBernoulli}_m^{(\text{III})}(\phi_0, \boldsymbol{\phi}, \boldsymbol{\lambda} \odot (\mathbf{1} + \boldsymbol{\lambda})).$$

B.2 Momentos

Derivação da função geradora de momentos da distribuição multivariada ZICOMP tipo III

Para o caso univariado sem o excesso de zeros a variável $X_i \sim \text{COMP}(\lambda_i, v_i)$ possui função geradora de momentos $M_{X_i}(t_i) = \frac{S(\lambda_i e^{t_i}, v_i)}{S(\lambda_i, v_i)}$. Então usando a definição $E(\xi) = E[E(\xi | \eta)]$

a função geradora de momentos para o vetor \mathbf{Y} é dada por

$$\begin{aligned}
\mathbf{M}_{\mathbf{y}}(\mathbf{t}) &= E \left[\exp(\mathbf{t}^\top \mathbf{y}) \right] = E \left[\exp \left(\sum_{i=1}^m t_i X_i Z_0 Z_i \right) \right] \\
&= E \left[\prod_{i=1}^m M_{X_i}(t_i X_i Z_0 Z_i) \right] = \phi_0 + (1 - \phi_0) \prod_{i=1}^m [E[M_{X_i}(t_i Z_i)]] \\
&= \phi_0 + (1 - \phi_0) \prod_{i=1}^m \left\{ \phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \sum_{y_i=1}^{\infty} e^{t_i y_i} \frac{\lambda_i^{y_i}}{(y_i!)^{\nu_i}} \right\} \\
&= \phi_0 + (1 - \phi_0) \prod_{i=1}^m \left\{ \phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \sum_{y_i=1}^{\infty} \frac{(\lambda_i e^{t_i})^{y_i}}{(y_i!)^{\nu_i}} \right\} \\
&= \phi_0 + (1 - \phi_0) \prod_{i=1}^m \left\{ \phi_i + (1 - \phi_i) \frac{S(\lambda_i e^{t_i}, \nu_i)}{S(\lambda_i, \nu_i)} \right\}. \tag{B.8}
\end{aligned}$$

$E(Y_i)$

Considere $S(\lambda_i, \nu_i) = \sum_{j=0}^{\infty} \frac{\lambda_i^j}{(j!)^{\nu_i}}$, tal que

$$\frac{\partial S(\lambda_i, \nu_i)}{\partial \lambda_i} = \sum_{j=1}^{\infty} \frac{j \lambda_i^{j-1}}{(j!)^{\nu_i}} = \sum_{t_i=0}^{\infty} \frac{\lambda_i^{t_i} (t_i + 1)^{1-\nu_i}}{(t_i!)^{\nu_i}}, \tag{B.9}$$

e tome $h(t_i) = \lambda_i e^{t_i}$, tal que $\frac{\partial S(\lambda_i e^{t_i}, \nu_i)}{\partial t_i} = \frac{\partial S(h(t_i), \nu_i)}{\partial \lambda_i} \frac{\partial h(t_i)}{\partial t_i}$, então combinando $\frac{\partial S(\lambda_i, \nu_i)}{\partial t_i}$ e (B.9) com (B.8) temos

$$\begin{aligned}
E(Y_i) &= \frac{\partial \mathbf{M}_{Y_i}(t_i)}{\partial t_i} \Big|_{t_i=0} = (1 - \phi_0) \left\{ (1 - \phi_i) \frac{\partial S(\lambda_i, \nu_i)}{\partial \lambda_i} \lambda_i \right\} \\
&= (1 - \phi_0) \left\{ (1 - \phi_i) \lambda_i \frac{\partial \log S(\lambda_i, \nu_i)}{\partial \lambda_i} \right\}. \tag{B.10}
\end{aligned}$$

$E(Y_i^2)$

Similarmente considere $S(\lambda_i, \nu_i)$, tal que

$$\frac{\partial^2 S(\lambda_i, \nu_i)}{\partial \lambda_i^2} = \sum_{j=2}^{\infty} \frac{j^2 \lambda_i^{j-2}}{(j!)^{\nu_i}} = \sum_{t_i=0}^{\infty} \frac{\lambda_i^{t_i} [(t_i + 1)(t_i + 2)]^{1-\nu_i}}{(t_i!)^{\nu_i}}, \tag{B.11}$$

e seja $\frac{\partial^2 S(\lambda_i, \nu_i)}{\partial t_i^2} = \frac{\partial^2 S(h(t_i), \nu_i)}{\partial \lambda_i^2} \frac{\partial h(t_i)}{\partial t_i} + \frac{\partial S(h(t_i), \nu_i)}{\partial \lambda_i} \frac{\partial^2 h(t_i)}{\partial t_i^2}$, então combinando $\frac{\partial^2 S(\lambda_i, \nu_i)}{\partial t_i^2}$ e (B.11) com (B.8) temos

$$E(Y_i^2) = \frac{\partial^2 \mathbf{M}_{Y_i}(t_i)}{\partial t_i^2} \Big|_{t_i=0} = (1 - \phi_0) \left\{ (1 - \phi_i) \left[\lambda_i^2 \frac{\partial^2 S(\lambda_i, \nu_i)}{\partial \lambda_i^2} + \lambda_i \frac{\partial \log S(\lambda_i, \nu_i)}{\partial \lambda_i} \right] \right\}, \tag{B.12}$$

Var(Y_i)

Considere $\text{Var}(Y_i) = E(Y_i^2) - E^2(Y_i)$, então

$$\text{Var}(Y_i) = (1 - \phi_0) \left\{ (1 - \phi_i) \left\{ \lambda_i^2 \left[\frac{\partial^2 S(\lambda_i, \nu_i)}{S(\lambda_i, \nu_i)} \mid \frac{\partial \lambda_i^2}{\partial \lambda_i} - (1 - \phi_0)(1 - \phi_i) \left(\frac{\partial \log S(\lambda_i, \nu_i)}{\partial \lambda_i} \right)^2 \right] + \frac{\partial \log S(\lambda_i, \nu_i)}{\partial \lambda_i} \right\} \right\}. \quad (\text{B.13})$$

Cov(Y_i, Y_j)

$$\begin{aligned} E(Y_i, Y_j) &= \frac{\partial^2 M_{(Y_i, Y_j)}(0, 0)}{\partial t_i \partial t_j} \\ &= (1 - \phi_0) \left\{ (1 - \phi_i) \lambda_i \frac{\partial S(\lambda_i, \nu_i) / \partial \lambda_i}{S(\lambda_i, \nu_i)} \right\} \left\{ (1 - \phi_j) \lambda_j \frac{\partial S(\lambda_j, \nu_j) / \partial \lambda_j}{S(\lambda_j, \nu_j)} \right\} \\ &= (1 - \phi_0) \mu_i \mu_j, \end{aligned} \quad (\text{B.14})$$

e como $\text{Cov}(Y_i, Y_j) = E(Y_i, Y_j) - E(Y_i)E(Y_j)$, então

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= E(Y_i, Y_j) - E(Y_i)E(Y_j) \\ &= \phi_0(1 - \phi_0) \mu_i \mu_j \\ &= \phi_0(1 - \phi_0) \left\{ \lambda_i \frac{\partial S(\lambda_i, \nu_i) / \partial (1 - \phi_i) \lambda_i}{S(\lambda_i, \nu_i)} \right\} \left\{ (1 - \phi_j) \lambda_j \frac{\partial S(\lambda_j, \nu_j) / \partial \lambda_j}{S(\lambda_j, \nu_j)} \right\}. \end{aligned} \quad (\text{B.15})$$

B.3 Distribuições condicionais

B.3.1 Distribuição condicional $\mathbf{Y}^{(1)} \mid \mathbf{Y}^{(2)}$

Caso II da distribuição condicional $\mathbf{Y}^{(1)} \mid \mathbf{Y}^{(2)}$ em (3.2.4.1). Podemos ter $\mathbf{Y}^{(1)} = \mathbf{0} \mid \mathbf{Y}^{(2)} = \mathbf{0}$, ou $\mathbf{Y}^{(1)} \neq \mathbf{0} \mid \mathbf{Y}^{(2)} = \mathbf{0}$, quando $\mathbf{y}^{(1)} = \mathbf{0}$, temos

$$\begin{aligned} \Pr(\mathbf{Y}^{(1)} = \mathbf{0} \mid \mathbf{Y}^{(2)} = \mathbf{0}) &= \frac{\phi_0 + (1 - \phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \right]}{\phi_0 + (1 - \phi_0) \prod_{i=r+1}^m \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \right]} \\ &= \frac{\phi_0 + (1 - \phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \right]}{\omega} \\ &= \frac{\phi_0}{\omega} + \left(1 - \frac{\phi_0}{\omega} \right) \prod_{i=1}^r \left[\phi_i + \frac{(1 - \phi_i)}{S(\lambda_i, \nu_i)} \right], \end{aligned} \quad (\text{B.16})$$

em que $\omega = \phi_0 + (1 - \phi_0) \prod_{i=r+1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right]$.

Se $\mathbf{y}^{(1)} \neq \mathbf{0}$, obtemos

$$\begin{aligned} & \Pr(\mathbf{Y}^{(1)} = \mathbf{Y}^{(1)} | \mathbf{y}^{(2)} = \mathbf{0}) \\ &= \frac{(1 - \phi_0) \prod_{i=1}^m \left\{ \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right] I_{(y_i=0)} + \left[\frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \frac{\lambda_i^{y_i}}{(y_i!)^{v_i}} \right] I_{(y_i>0)} \right\}}{\phi_0 + (1 - \phi_0) \prod_{i=r+1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right]} \\ &= \left(1 - \frac{\phi_0}{\omega}\right) \prod_{i=1}^r \left\{ \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right] I_{(y_i=0)} + \left[\frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \frac{\lambda_i^{y_i}}{(y_i!)^{v_i}} \right] I_{(y_i>0)} \right\}. \end{aligned} \quad (\text{B.17})$$

Combinando (B.16) com (B.17), e substituindo em (3.2), temos

$$\begin{aligned} & \Pr(\mathbf{Y}^{(1)} = \mathbf{Y}^{(1)} | \mathbf{y}^{(2)} = \mathbf{0}) \\ &= \left\{ \frac{\phi_0}{\omega} + \left(1 - \frac{\phi_0}{\omega}\right) \prod_{i=1}^r \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right] \right\} I_{(\mathbf{y}^{(1)}=\mathbf{0})} + \left(1 - \frac{\phi_0}{\omega}\right) \\ & \times \prod_{i=1}^r \left\{ \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right] I_{(y_i=0)} + \left[\frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \frac{\lambda_i^{y_i}}{(y_i!)^{v_i}} \right] I_{(y_i>0)} \right\} I_{(\mathbf{y}^{(1)} \neq \mathbf{0})}, \end{aligned} \quad (\text{B.18})$$

então, isso implica que

$$\mathbf{Y}^{(1)} | (\mathbf{Y}^{(2)} = \mathbf{0}) \sim \text{ZICOMP}^{(\text{III})} \left(\frac{\phi_0}{\omega}; \phi_1, \dots, \phi_r; \lambda_1, \dots, \lambda_r; \mathbf{v}_1, \dots, \mathbf{v}_r \right)$$

B.3.2 Distribuição condicional $\mathbf{X} | \mathbf{Y}$

Prova da distribuição Condicional $\mathbf{X} | \mathbf{Y}$ quando $\mathbf{y} = \mathbf{0}$

$$\begin{aligned} \Pr(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{0}) &= \frac{\Pr(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{0})}{\Pr(\mathbf{y} = \mathbf{0})} \\ &= \frac{\Pr(\mathbf{X} = \mathbf{0}, \mathbf{Y} = \mathbf{0})}{f(\mathbf{0}, \phi, \boldsymbol{\lambda}, \mathbf{v})} I_{(\mathbf{x}=\mathbf{0})} + \frac{\Pr(\mathbf{X} = \mathbf{x}, \mathbf{Z}_0 = \mathbf{0})}{f(\mathbf{0}, \phi, \boldsymbol{\lambda}, \mathbf{v})} I_{(\mathbf{x} \neq \mathbf{0})} \\ &= \frac{\Pr(\mathbf{X} = \mathbf{0})}{f(\mathbf{0}, \phi, \boldsymbol{\lambda}, \mathbf{v})} I_{(\mathbf{x}=\mathbf{0})} + \frac{\phi_0 \Pr(\mathbf{x} = \mathbf{x})}{f(\mathbf{0}, \phi, \boldsymbol{\lambda}, \mathbf{v})} I_{(\mathbf{x} \neq \mathbf{0})} \\ &= \frac{\prod_{i=1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right]}{\phi_0 + (1 - \phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right]} I_{(\mathbf{x}=\mathbf{0})} + \\ &+ \frac{\phi_0 \prod_{i=1}^m \left[\frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \frac{\lambda_i^{y_i}}{(y_i!)^{v_i}} \right]}{\phi_0 + (1 - \phi_0) \prod_{i=1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, \mathbf{v}_i)} \right]} I_{(\mathbf{x} \neq \mathbf{0})} \\ &= \left[\tau + \frac{(1-\tau)}{S^+(\boldsymbol{\lambda}, \mathbf{v})} \right] I_{(\mathbf{x}=\mathbf{0})} + (1-\tau) \prod_{i=1}^m \left[\frac{\lambda_i^{y_i}}{(y_i!)^{v_i} S(\lambda_i, \mathbf{v}_i)} \right] I_{(\mathbf{x} \neq \mathbf{0})}, \end{aligned} \quad (\text{B.19})$$

em que $\tau = \frac{(1-\phi_0)\prod_{i=1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, \nu_i)} \right]}{\phi_0 + (1-\phi_0)\prod_{i=1}^m \left[\phi_i + \frac{(1-\phi_i)}{S(\lambda_i, \nu_i)} \right]}$, então

$$\mathbf{x} | (\mathbf{y} = \mathbf{0}) \sim \text{ZICOMP}_m^{(\text{III})}(\tau, \boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \quad (\text{B.20})$$

INFLUÊNCIA LOCAL

C.1 Derivação da matriz de pesos

Seja $\ell(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança do modelo ZICOMP Tipo I, e considere $\lambda_j = \exp(\mathbf{w}_j^\top \boldsymbol{\beta}_i)$; $v_j = \exp(\mathbf{u}_j^\top \boldsymbol{\zeta}_i)$; $\phi_j = \frac{\exp(\mathbf{z}_j^\top \boldsymbol{\gamma})}{1 + \exp(\mathbf{z}_j^\top \boldsymbol{\gamma})}$, tal que

$$\frac{\partial \lambda_j}{\partial \boldsymbol{\beta}_j} = \exp(\mathbf{w}_j^\top \boldsymbol{\beta}_j) \mathbf{w}_j = \lambda_j \mathbf{w}_j; \quad \frac{\partial v_j}{\partial \boldsymbol{\zeta}_j} = \exp(\mathbf{u}_j^\top \boldsymbol{\zeta}_j) \mathbf{u}_j = v_j \mathbf{u}_j; \quad \frac{\partial \phi_j}{\partial \boldsymbol{\gamma}} = \frac{\exp(\mathbf{z}_j^\top \boldsymbol{\gamma})}{(1 + \exp(\mathbf{z}_j^\top \boldsymbol{\gamma}))^2} \mathbf{z}_j, \quad (\text{C.1})$$

e considere também, $g_j = \frac{\exp(\mathbf{z}_j^\top \boldsymbol{\gamma})}{1 + \exp(\mathbf{z}_j^\top \boldsymbol{\gamma})}$ e $(1 - g_j) = \frac{1}{1 + \exp(\mathbf{z}_j^\top \boldsymbol{\gamma})}$. Dessa forma, as derivadas de $\ell(\boldsymbol{\theta})$ em relação aos parâmetros $\boldsymbol{\gamma}_i$, $\boldsymbol{\beta}_i$ e $\boldsymbol{\zeta}_i$ são dadas por

$$\left\{ \begin{array}{l} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \phi} = k \left\{ \frac{1 - \prod_{i=1}^m \left[\frac{1}{S(\lambda_i, v_i)} \right]}{a_1} \right\} - \frac{n-k}{1-\phi}; \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = \left\{ k \left\{ \frac{1 - \prod_{i=1}^m \left[\frac{1}{S(\lambda_{il}, v_{il})} \right]}{a_1} \right\} - \frac{n-k}{1-\phi} \right\} g_j (1 - g_j) \mathbf{z}_j^\top; \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \lambda_i} = -k \left(\frac{(1-\phi)}{S(\lambda_i, v_i)} \right) \left(\frac{\partial \log S(\lambda_i, v_i)}{\partial \lambda_i} \right) b_i + \sum_{j=k+1}^n \frac{y_{ij}}{\lambda_i} - \frac{\partial \log S(\lambda_i, v_i)}{\partial \lambda_i}; \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i} = \left\{ -k \left(\frac{(1-\phi)}{S(\lambda_{il}, v_{il})} \right) \left(\frac{\partial \log S(\lambda_{il}, v_{il})}{\partial \lambda_{il}} \right) b_i + \sum_{j=k+1}^n \frac{y_{ij}}{\lambda_{il}} - \frac{\partial \log S(\lambda_{il}, v_{il})}{\partial \lambda_{il}} \right\} \lambda_{jl} \mathbf{w}_j^\top; \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial v_i} = -k \left\{ \frac{(1-\phi)}{S(\lambda_i, v_i)} \left[\frac{\partial \log S(\lambda_i, v_i)}{\partial v_i} \right] \right\} b_i - \frac{\partial \log S(\lambda_i, v_i)}{\partial v_i} - \sum_{j=k+1}^n \log(y_{ij}!); \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\zeta}_i} = \left\{ -k \left\{ \frac{(1-\phi)}{S(\lambda_{il}, v_{il})} \left[\frac{\partial \log S(\lambda_{il}, v_{il})}{\partial v_{il}} \right] \right\} b_i - \frac{\partial \log S(\lambda_{il}, v_{il})}{\partial v_{il}} - \sum_{j=k+1}^n \log(y_{ij}!) \right\} v_{jl} \mathbf{u}_j^\top; \end{array} \right. \quad (\text{C.2})$$

em que:

$$k = \sum_{j=1}^n I(\mathbf{y}_j = \mathbf{0}); \quad a_1 = \phi + (1 - \phi) \prod_{i=1}^m \left[\frac{1}{S(\lambda_i, v_i)} \right]; \quad b_i = \prod_{l \neq i} \left[\frac{1}{S(\lambda_l, v_l)} \right], \quad i = 1, \dots, m.$$

Dessa forma, os elementos da matriz de perturbação Δ , para o modelo de regressão da distribuição ZICOMP Tipo I, cujo logaritmo da função perturbada

$$\ell(\boldsymbol{\theta} \mid \boldsymbol{\delta}) = \sum_{j=1}^n \boldsymbol{\delta}_j \ell_j(\boldsymbol{\theta}), \quad (\text{C.3})$$

com $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^\top$ um vetor de perturbações, com $0 \leq \delta_j \leq 1$, e $j = 1, \dots, n$, são dados por

$$\left\{ \begin{array}{l} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_i \partial \delta_j} = \left\{ k \left\{ \frac{1 - \prod_{i=1}^m \left[\frac{1}{S(\lambda_{il}, v_{il})} \right]}{a_1} \right\} - \frac{n-k}{1-\phi} \right\} g_j (1 - g_j) \mathbf{z}_j^\top; \\ \frac{\partial^2 \ell(\boldsymbol{\theta} \mid \boldsymbol{\delta}_j)}{\partial \beta_i} = \left\{ -k \left(\frac{(1-\phi)}{S(\lambda_{il}, v_{il})} \right) \left(\frac{\partial \log S(\lambda_{il}, v_{il})}{\partial \lambda_{il}} \right) b_i + \sum_{j=k+1}^n \frac{y_{ij}}{\lambda_{il}} - \frac{\partial \log S(\lambda_{il}, v_{il})}{\partial \lambda_{il}} \right\} \lambda_{jl} \mathbf{w}_j^\top; \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \zeta_i \partial \delta_j} = \left\{ -k \left\{ \frac{(1-\phi)}{S(\lambda_{il}, v_{il})} \left[\frac{\partial \log S(\lambda_{il}, v_{il})}{\partial v_{il}} \right] \right\} b_i - \frac{\partial \log S(\lambda_{il}, v_{il})}{\partial v_{il}} - \sum_{j=k+1}^n \log(y_{ij}!) \right\} v_{jl} \mathbf{u}_j^\top; \end{array} \right.$$

em que:

$$k = \sum_{j=1}^n I(\mathbf{y}_j = \mathbf{0}); \quad a_1 = \phi + (1 - \phi) \prod_{i=1}^m \left[\frac{1}{S(\lambda_i, v_i)} \right]; \quad b_i = \prod_{t \neq i}^m \left[\frac{1}{S(\lambda_t, v_t)} \right], \quad i = 1, \dots, m. \quad (\text{C.4})$$

C.1.1 Dados do capítulo 4

Na Tabela 22, apresentamos as variáveis Y_1 , Y_2 e X , como o número de casos confirmados de AIDS, hepatite viral e o IDH (índice de desenvolvimento humano) de 97 cidades do interior do estado de Alagoas-Brasil disponível em (ALAGOAS, 2015).

Tabela 22 – Frequência das variáveis Y_1 , Y_2 e X em 97 cidades do interior do estado de Alagoas-Brasil

(Y_1, Y_2, X)	(Y_1, Y_2, X)	(Y_1, Y_2, X)	(Y_1, Y_2, X)
(0,2,0.549)	(0,2,0.532)	(0,0,0.504)	(0,1,0.541)
(1,1,0.568)	(0,0,0.534)	(19,0,0.584)	(3,1,0.551)
(1,1,0.561)	(0,8,0.533)	(3,11,0.568)	(0,0,0.559)
(3,0,0.557)	(0,1,0.527)	(0,0,0.539)	(16,35,0.643)
(1,0,0.615)	(1,3,0.536)	(13,0,0.527)	(0,0,0.505)
(3,0,0.594)	(0,1,0.518)	(1,0,0.521)	(1,2,0.597)
(0,0,0.593)	(2,0,0.564)	(1,0,0.565)	(3,0,0.591)
(0,0,0.517)	(0,0,0.568)	(0,0,0.524)	(2,0,0.519)
(3,1,0.604)	(2,0,0.484)	(0,0,0.503)	(2,4,0.573)
(2,0,0.513)	(0,0,0.583)	(0,0,0.493)	(1,2,0.527)
(0,0,0.531)	(0,0,0.548)	(1,4,0.547)	(6,2,0.536)
(2,2,0.562)	(2,0,0.570)	(0,0,0.558)	(7,7,0.623)
(1,1,0.559)	(0,0,0.552)	(10,1,0.638)	(1,0,0.591)
(9,1,0.570)	(4,1,0.556)	(7,1,0.593)	(1,0,0.549)
(0,0,0.524)	(3,0,0.531)	(0,2,0.548)	(3,1,0.660)
(2,0,0.506)	(0,1,0.562)	(2,0,0.605)	(0,0,0.518)
(2,0,0.573)	(11,3,0.575)	(2,0,0.533)	(0,0,0.555)
(0,0,0.526)	(1,1,0.552)	(0,1,0.589)	(0,0,0.541)
(1,0,0.575)	(2,1,0.580)	(4,5,0.630)	(8,1,0.564)
(0,0,0.533)	(1,0,0.566)	(0,0,0.572)	(1,1,0.532)
(3,0,0.517)	(1,0,0.577)	(5,4,0.610)	(18,4,0.593)
(0,0,0.586)	(0,0,0.574)	(0,0,0.574)	(2,1,0.586)
(10,0,0.626)	(5,2,0.569)	(2,3,0.589)	
(0,0,0.525)	(1,0,0.642)	(0,0,0.526)	
(3,3,0.612)	(11,2,0.594)	(3,0,0.586)	

