

**REDUÇÃO NO VÍCIO DA DISTRIBUIÇÃO DA *DEVIANCE* PARA
DADOS DE CONTAGEM**

DENISE NUNES VIOLA

Dissertação apresentada à Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, para a obtenção do título de Mestre em Agronomia, Área de Concentração: Estatística e Experimentação Agronômica.

PIRACICABA
São Paulo - Brasil
Setembro - 2001

**REDUÇÃO NO VÍCIO DA DISTRIBUIÇÃO DA *DEVIANCE* PARA
DADOS DE CONTAGEM**

DENISE NUNES VIOLA

Bacharel em Estatística

Orientador: Prof. Dr. **José Eduardo Corrente**

Dissertação apresentada à Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, para a obtenção do título de Mestre em Agronomia, Área de Concentração: Estatística e Experimentação Agronômica.

PIRACICABA
São Paulo - Brasil
Setembro - 2001

**Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Viola, Denise Nunes
Redução no vício da distribuição da *Deviance* para dados de contagem /
Denise Nunes Viola. - - Piracicaba, 2001.
59 p.

Dissertação (mestrado) - - Escola Superior de Agricultura Luiz de Queiroz,
2001.
Bibliografia.

1. Distribuição (teoria da probabilidade) 2. Estatística aplicada 3. Inferência
estatística 4. Modelo matemático 5. Teoria de Poisson I. Título

CDD 519.54

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

DEDICATÓRIA

Dedico este trabalho à minha mãe e ao meu pai que eu amo muito e foram os principais responsáveis para eu ser o que sou hoje.

AGRADECIMENTOS

Em especial ao Prof. Dr. José Eduardo Corrente, por ter sido um excelente orientador e um ótimo amigo.

A todo o pessoal do HCFMRP - USP, principalmente à Cláudia, Ilza, Luciane, Maria Eulália e Zara que muito contribuíram para a conclusão deste trabalho.

Aos professores e amigos da UFSCar pelo apoio, principalmente à Angela, que sempre me animava nas horas difíceis.

Ao Laurinho que sempre me ouvia e dava conselhos nas horas de desespero.

Aos meus pais que sempre acreditaram em mim e me apoiaram.

À minha avó, minha irmã, meus tios e meus primos pelo apoio e pelo carinho.

Aos professores e funcionários da ESALQ - USP do Departamento de Ciências Exatas.

A todo o pessoal da Pós Graduação em Estatística e Experimentação Agronômica da ESALQ - USP.

A todos os meus amigos da ESALQ - USP que me ajudaram muito.

A todos os amigos que, de um jeito ou de outro, contribuíram para a conclusão deste trabalho.

SUMÁRIO

	Página
LISTA DE FIGURAS	v
RESUMO	vii
SUMMARY	ix
1 INTRODUÇÃO	1
2 REVISÃO DE LITERATURA	3
2.1 A distribuição Poisson	3
2.2 Modelos Lineares Generalizados	9
2.2.1 Seleção do modelo	10
2.2.2 Estimação	11
2.2.3 Predição	12
2.3 Os componentes de um Modelo Linear Generalizado	12
2.4 Medindo a qualidade do ajuste	16
2.5 Fatores de Correção	19
2.5.1 Correção de Bartlett	19
2.5.2 Propostas de Cordeiro (1986)	21
2.5.3 Outras Propostas para Fatores de Correção	23
2.6 Desenvolvimentos Assintóticos	24
2.6.1 Série de Taylor para funções de uma variável	25
2.6.2 Série de Taylor para funções de duas ou mais variáveis	26

	iv
3 METODOLOGIA	31
3.1 Expansão da <i>deviance</i> para uma variável sem correção	31
3.2 Expansão da <i>deviance</i> para uma variável com correção	34
3.3 Expansão da <i>deviance</i> para n variáveis sem correção	39
3.4 Expansão da <i>deviance</i> para n variáveis com correção	43
4 RESULTADO E DISCUSSÃO	48
5 CONCLUSÃO	56
REFERÊNCIAS BIBLIOGRÁFICAS	57

LISTA DE FIGURAS

Página

1	Gráfico do valor esperado da <i>deviance</i> para $Y \sim P(\lambda)$, $\lambda = 0(0,1)15$. . .	49
2	Gráfico do valor esperado da <i>deviance</i> para uma variável aleatória $Y \sim P(\lambda)$ para $\lambda = 0(0,1)15$ com e sem correção (— = sem correção, = com correção).	49
3	<i>QQ-plot</i> para a distribuição da <i>deviance</i> com e sem correção para uma variável aleatória Poisson(2), para uma amostra de tamanho 10 e 1000 replicações.	50
4	<i>QQ-plot</i> para a distribuição da <i>deviance</i> com e sem correção para uma variável aleatória Poisson(2), para uma amostra de tamanho 20 e 1000 replicações.	51
5	<i>QQ-plot</i> para a distribuição da <i>deviance</i> com e sem correção para uma variável aleatória Poisson(5), para uma amostra de tamanho 10 e 1000 replicações.	52
6	<i>QQ-plot</i> para a distribuição da <i>deviance</i> com e sem correção para uma variável aleatória Poisson(5), para uma amostra de tamanho 20 e 1000 replicações.	52
7	<i>QQ-plot</i> para a distribuição da <i>deviance</i> com e sem correção para uma variável aleatória Poisson(10), para uma amostra de tamanho 10 e 1000 replicações.	53
8	<i>QQ-plot</i> para a distribuição da <i>deviance</i> com e sem correção para uma variável aleatória Poisson(10), para uma amostra de tamanho 20 e 1000 replicações.	53

9	<i>QQ-plot</i> para a distribuição da <i>deviance</i> com e sem correção para uma variável aleatória Poisson(15), para uma amostra de tamanho 10 e 1000 replicações.	54
10	<i>QQ-plot</i> para a distribuição da <i>deviance</i> com e sem correção para uma variável aleatória Poisson(15), para uma amostra de tamanho 20 e 1000 replicações.	54

REDUÇÃO NO VÍCIO DA DISTRIBUIÇÃO DA *DEVIANCE* PARA DADOS DE CONTAGEM

Autora: DENISE NUNES VIOLA

Orientador: Prof. Dr. JOSÉ EDUARDO CORRENTE

RESUMO

Dados de contagem podem ser considerados, em geral, como provenientes de uma distribuição de Poisson. Neste contexto, a análise de tais dados apresenta certas dificuldades, pois não segue as pressuposições básicas para o ajuste de um modelo matemático. Desse modo, algumas transformações são sugeridas, mas nem sempre bons resultados são obtidos. No enfoque de Modelos Lineares Generalizados, a estatística que mede a qualidade do ajuste do modelo para os dados é chamada *deviance*. Porém, a distribuição da *deviance* é, em geral, desconhecida. No entanto, para dados com distribuição de Poisson, pode-se mostrar que a distribuição da *deviance* se aproxima de uma distribuição χ^2 , mas tal aproximação não é boa para tamanhos pequenos de amostra. Para melhorar essa aproximação, alguns fatores de correção para os dados são sugeridos, mas os resultados obtidos ainda não são satisfatórios. Assim, o objetivo deste trabalho é propor um novo fator de correção para os dados seguindo uma distribuição de Poisson, de modo a se

obter uma melhora na distribuição da *deviance* para qualquer tamanho de amostra. Para isto, será adicionada uma constante à variável resposta e, através do valor esperado da *deviance*, calcula-se tal constante de modo a reduzir o erro cometido na aproximação. Para verificar a melhora na aproximação da distribuição da *deviance* a uma distribuição χ^2 , dados de uma distribuição de Poisson foram simulados e o valor da *deviance* foi calculado com e sem o fator de correção. Construiu-se, assim, os *QQ-plots* dos valores obtidos da *deviance* para a comparação com a distribuição χ^2 .

BIAS REDUCTION IN THE DISTRIBUTION OF THE DEVIANCE FOR COUNT DATA

Author: DENISE NUNES VIOLA

Adviser: Prof. Dr. JOSÉ EDUARDO CORRENTE

SUMMARY

Analysis of count data presents, in general, can be supposed coming from a Poisson distribution. The analysis of such data have some problems once the underlying distribution of them does not follow the basic assumptions to fit a model. Some transformations can be suggested, but good results are not always obtained. In the approach of the Generalized Linear Models, the deviance is the statistics that measures the goodness of fit, but its distribution is unknown. Furthermore, considering Poisson distribution data, it is possible to approximate the distribution of the deviance for a chi-square distribution, but such approximation is not good for small sample size. In order of improve this approximation, corrections for the data are suggested, but the results are not good yet. Then, the aim of this work is to propose a new correction factor for data following a Poisson distribution in order to obtain an improvement in the distribution of the deviance for any sample size. For this, just adding a constant at the response variable and, through the

expected value of the deviance, such constant is obtained in order to reduce the error in the approximation. Simulated data from the Poisson distribution were made to calculate the deviance with and without the correction and QQ-plots were used to compare the values of the deviance with the chi-square distribution.

1 INTRODUÇÃO

Dados de contagem, em geral, podem ser supostos como provenientes de uma distribuição Poisson e nesses casos, é comum fazer uma transformação estabilizadora da variância a fim de validar as pressuposições básicas para o ajuste do modelo a ser analisado.

Quando esses dados são coletados seguindo algum tipo de delineamento experimental, pode-se usar o enfoque de Modelos Lineares Generalizados para proceder a análise utilizando-se a *deviance*, que mede a qualidade do ajuste para o modelo considerado. A *deviance* é dada pela diferença do logaritmo da função de verossimilhança do modelo superparametrizado em relação ao logaritmo da função de verossimilhança do modelo sob pesquisa.

A distribuição da *deviance* é desconhecida, mas para alguns casos particulares como no caso da distribuição de Poisson com parâmetro λ , pode-se mostrar que sua distribuição é aproximadamente χ^2 com $n - p$ graus de liberdade, em que n é o número de observações e p é o número de parâmetros envolvidos no modelo sob pesquisa. Tal aproximação apresenta um resto da ordem de $O(\lambda^{-1})$, em que λ representa o valor médio da distribuição de Poisson.

Fatores de correção têm sido propostos na literatura de modo a melhorar a aproximação da distribuição da *deviance* a uma distribuição χ^2 mas sem muito sucesso. O problema fica ainda mais complicado quando se trata de pequenas amostras. Nesses casos, as correções propostas ainda não são eficazes para melhorar a aproximação.

Assim, a fim de reduzir o vício na distribuição da *deviance* e torná-la

mais próxima de uma distribuição χ^2 , o objetivo deste trabalho será propor um novo fator de correção a ser adicionado à variável resposta. Esse fator será obtido através da expansão em série de Taylor da função *deviance* e o cálculo de seu valor esperado e da sua variância de modo que, se a distribuição da *deviance* converge para uma distribuição χ^2 , o valor esperado e a variância devem se aproximar do valor esperado e da variância da distribuição χ^2 . A melhora na aproximação da distribuição da *deviance* para uma distribuição de χ^2 também será verificada através da simulação de dados de uma distribuição de Poisson com parâmetro λ e a construção de *QQ-plots* para vários valores de λ e vários tamanhos de amostra. Tais verificações serão feitas com e sem o fator de correção obtido.

2 REVISÃO DE LITERATURA

2.1 A distribuição Poisson

Segundo Johnson e Kotz (1969), uma variável aleatória Y tem distribuição Poisson com parâmetro λ se sua distribuição de probabilidade é dada por:

$$P(Y = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{para } k = 0, 1, \dots \quad (1)$$

Essa distribuição pode ser obtida da seguinte forma: seja Y uma variável aleatória com distribuição Binomial com parâmetros m e p , dada por

$$P[Y = k] = \binom{m}{k} p^k (1 - p)^{m-k} \quad (k = 0, 1, \dots, m).$$

Fazendo $m \rightarrow +\infty$ e $p \rightarrow 0$, de modo que $mp = \lambda$ permaneça constante, a distribuição limite é Poisson com parâmetro λ , como dada em (1).

Meyer (1984), James (1981), dentre outros, mostram que a distribuição de Poisson pode ser obtida através do Processo de Poisson, onde são feitas algumas hipóteses, chamadas *hipóteses de Poisson*, e uma equação diferencial é calculada, cuja solução, obtida por recorrência, gera (1).

Os diferentes valores de λ correspondem a diferentes formações de distribuições da família; assim, λ é o parâmetro da distribuição de Poisson.

As probabilidades de se observar Y igual a $0, 1, 2, \dots$, respectivamente, são:

$$P(Y = 0) = e^{-\lambda},$$

$$P(Y = 1) = \lambda e^{-\lambda},$$

$$P(Y = 2) = \frac{\lambda^2 e^{-\lambda}}{2},$$

...

$$P(Y = m) = \frac{\lambda^m e^{-\lambda}}{m!},$$

...

e mostra-se que

$$\sum_{k=0}^{+\infty} P(Y = k) = \sum_{k=0}^{+\infty} \frac{\lambda^k e^{-\lambda}}{k!} = 1$$

definindo uma verdadeira distribuição de probabilidade.

A função de distribuição acumulada correspondente à função de probabilidade dada em (1) é :

$$F(y) = P(Y \leq y) = \sum_{k; k \leq y} \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{para todo } y \in \mathcal{R}, \quad k = 0, 1, \dots \quad .$$

As funções característica e geradora de momentos da distribuição Poisson são definidas por

$$\phi_Y(t) = e^{\lambda(e^{it}-1)} \quad \text{e} \quad M_Y(t) = e^{\lambda(e^t-1)}$$

e através delas podem-se calcular os momentos em torno de zero. Assim, por exemplo, os momentos de ordem 1, 2, 3 e 4 são:

$$E(Y) = \lambda,$$

$$E(Y^2) = \lambda + \lambda^2,$$

$$E(Y^3) = \lambda + 3\lambda^2 + \lambda^3,$$

$$E(Y^4) = \lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4,$$

Podem-se ainda, calcular os momentos centrados na média, sendo que os de ordem 1, 2, ..., 8 são dados por:

$$E(Y - E(Y)) = 0,$$

$$E(Y - E(Y))^2 = \text{Var}(Y) = \lambda,$$

$$E(Y - E(Y))^3 = \lambda,$$

$$E(Y - E(Y))^4 = \lambda + 3\lambda^2,$$

$$E(Y - E(Y))^5 = \lambda + 10\lambda^2,$$

$$E(Y - E(Y))^6 = \lambda + 25\lambda^2 + 15\lambda^3,$$

$$E(Y - E(Y))^7 = \lambda + 56\lambda^2 + 105\lambda^3,$$

$$E(Y - E(Y))^8 = \lambda + 119\lambda^2 + 490\lambda^3 + 105\lambda^4,$$

e assim sucessivamente.

Tem-se também que a distribuição de Poisson define uma família exponencial a um parâmetro. Isto porque, segundo Bickel e Doksum (1977), uma

variável aleatória Y com uma distribuição qualquer P_θ pertencente a uma família de distribuições de um modelo $\{P_\theta : \theta \in \Theta\}$, em que θ é um parâmetro unidimensional, é chamada *família exponencial com um parâmetro*, se existem funções de valores reais $c(\theta)$, $d(\theta)$ em Θ , funções de valores reais T e S em \mathcal{R}^n e um conjunto $A \subset \mathcal{R}^n$ tal que a função de densidade (ou a distribuição de probabilidade) $p(y, \theta)$ de P_θ pode ser escrita como:

$$p(y, \theta) = \{\exp[c(\theta)T(y) + d(\theta) + S(y)]\}I_A(y),$$

em que I_A é a função indicadora dos valores possíveis da variável aleatória Y pertencentes a um conjunto A . Note que as funções c , d , S e T não são únicas e o conjunto A não depende do parâmetro θ .

Numa família exponencial com um parâmetro, a estatística $T(Y)$ é suficiente para θ . Isto é claro desde que é necessário somente identificar $e^{c(\theta)T(y)+d(\theta)}$ com $g(T(y), \theta)$ e $[\exp S(y)]I_A(x)$ com $h(y)$ no Teorema da Fatorização de Fisher-Neyman. Neste caso, T é a *estatística suficiente natural da família*.

Assim, supondo que Y tenha uma distribuição Poisson(λ), $\lambda > 0$. Então

$$\begin{aligned} P(Y = y, \lambda) = p(y, \lambda) &= \frac{\lambda^y e^{-\lambda}}{y!} I_A(y) \\ &= \exp\{y \log(\lambda) - \lambda - \log(y!)\} I_A(y) \end{aligned} \quad (2)$$

em que $A = \{0, 1, \dots\}$ que corresponde aos valores possíveis de Y . Reescrevendo (2), obtém-se

$$c(\lambda) = \log(\lambda) \quad d(\lambda) = -\lambda,$$

$$T(y) = y \quad S(y) = -\log(y!).$$

Bickel e Doksum (1977) ainda mostram que existe uma extensão natural da família exponencial com um parâmetro para n variáveis.

Na prática, quando se pode supor que os dados provêm de uma distribuição de Poisson, as pressuposições básicas para uma análise de variância clássica não são satisfeitas, principalmente devido à heterogeneidade de variâncias.

Algumas vezes, uma transformação nesses dados pode ser utilizada a fim de estabilizar a variância. Cox e Snell (1981) comentam sobre a existência de três métodos para se fazer transformações, como segue:

- i) uma transformação não linear de uma variável resposta efetivamente contínua para uma nova forma;
- ii) uma transformação não linear do parâmetro, que é usualmente um valor esperado, numa representação da variação sistemática;
- iii) uma transformação não linear de uma variável explanatória.

Em (i), pode-se supor que algumas formas transformadas da variável resposta satisfazem um modelo linear teoricamente normal, mas na prática isso nem sempre é possível. Quando se determina uma transformação, deseja-se encontrar uma estrutura linear simples, variância constante e distribuições normais. No caso de ocorrer um conflito entre essas estruturas, deve-se concentrar a atenção na análise dos dados não transformados.

Em (ii), a transformação não linear não é aplicada à variável resposta, mas em um parâmetro, que normalmente é a média.

A distinção entre (i) e (ii) é ilustrada no contexto da teoria normal supondo-se que:

- a) alguma função $g(Y_i)$ da resposta Y_i , por exemplo, $g(Y_i) = \log Y_i$, é normalmente distribuída com variância constante, e com a média contendo uma estrutura simples em termos de variáveis explanatórias;
- b) Y_i é normalmente distribuída com variância constante e que alguma função $g\{E(Y_i)\}$ de sua média, por exemplo, $\log E(Y_i)$, tem estrutura simples em termos de variáveis explanatórias.

Nesse contexto particular, (a) é o procedimento mais usado e, além disso, McCullagh e Nelder (1989) comentam que, para uma variável aleatória Y com distribuição de Poisson com parâmetro λ , quando $\lambda \rightarrow +\infty$

$$\frac{(Y - \lambda)}{\lambda^{1/2}} \sim N(0, 1) + O_p(\lambda^{-1/2})$$

em que $O_p(\lambda^{-1/2})$ é $P(\lim_{\lambda \rightarrow \infty} \frac{\lambda}{a_n} = k) \rightarrow 1$ e pode-se utilizar uma transformação estabilizadora da variância dada por $Y^{1/2}$ no sentido de que, para λ grande

$$E(Y^{1/2}) \simeq \lambda^{1/2} \quad \text{e} \quad \text{var}(Y^{1/2}) \simeq \frac{1}{4}.$$

Para mostrar este fato, considere $g(y) = \sqrt{y}$. A expansão em série de Taylor de $g(y)$ em torno de λ é dada por:

$$\begin{aligned} \sqrt{y} &= \sqrt{\lambda} + \frac{1}{2\sqrt{\lambda}}(y - \lambda) - \frac{1}{8\sqrt{\lambda}^3}(y - \lambda)^2 + \frac{1}{16\sqrt{\lambda}^5}(y - \lambda)^3 - \frac{5}{128\sqrt{\lambda}^7}(y - \lambda)^4 + \\ &+ \frac{7}{256\sqrt{\lambda}^9}(y - \lambda)^5 + \dots \end{aligned}$$

Considerando Y como uma variável aleatória e calculando o valor esperado da expansão de $g(Y) = \sqrt{Y}$, tem-se:

$$\begin{aligned} E(\sqrt{Y}) &= E \left[\sqrt{\lambda} + \frac{1}{2\sqrt{\lambda}}(Y - \lambda) - \frac{1}{8\sqrt{\lambda}^3}(Y - \lambda)^2 + \frac{1}{16\sqrt{\lambda}^5}(Y - \lambda)^3 \right. \\ &\quad \left. - \frac{5}{128\sqrt{\lambda}^7}(Y - \lambda)^4 + \dots \right] \\ &= \sqrt{\lambda} + \frac{1}{2\sqrt{\lambda}}E(Y - \lambda) - \frac{1}{8\sqrt{\lambda}^3}E(Y - \lambda)^2 + \frac{1}{16\sqrt{\lambda}^5}E(Y - \lambda)^3 \\ &\quad - \frac{5}{128\sqrt{\lambda}^7}E(Y - \lambda)^4 + \dots \\ &= \sqrt{\lambda} + 0 - \frac{1}{8\sqrt{\lambda}^3}\lambda + \frac{1}{16\sqrt{\lambda}^5}\lambda - \frac{5}{128\sqrt{\lambda}^7}\lambda(3\lambda + 1) + \dots \\ &= \sqrt{\lambda} - \frac{1}{8\sqrt{\lambda}} + \frac{1}{16\sqrt{\lambda}^3} - \frac{15}{128\sqrt{\lambda}^3} - \frac{5}{128\sqrt{\lambda}^5} + \dots \end{aligned}$$

$$\begin{aligned}
&= \sqrt{\lambda} - \frac{1}{8\sqrt{\lambda}} + \frac{128 - 240}{2048\sqrt{\lambda}^3} - \frac{5}{128\sqrt{\lambda}^5} + \dots \\
&= \sqrt{\lambda} - \frac{1}{8\sqrt{\lambda}} - \frac{7}{128\sqrt{\lambda}^3} - \frac{5}{128\sqrt{\lambda}^5} + \dots \\
&= \sqrt{\lambda} + O(\lambda^{-1/2}).
\end{aligned}$$

Do mesmo modo, calculando-se a variância de \sqrt{Y} , têm-se

$$\begin{aligned}
Var(\sqrt{Y}) &= E\left(\sqrt{Y^2}\right) - E^2(\sqrt{Y}) \\
&= E(Y) - \left[\sqrt{\lambda} - \frac{1}{8\sqrt{\lambda}} - \frac{7}{128\sqrt{\lambda}^3} + \dots\right]^2 \\
&= \lambda - \left[\lambda - \frac{1}{8} - \frac{7}{128\lambda} - \frac{1}{8} + \frac{1}{64\lambda} + \frac{7}{1024\lambda^2} - \frac{7}{128\lambda} + \frac{7}{124\lambda^2} + \dots\right] \\
&= \lambda - \lambda + \frac{1}{4} - \frac{7}{54\lambda} - \frac{1}{64\lambda} - \frac{7}{512\lambda^2} + \dots \\
&= \frac{1}{4} + O(\lambda^{-1}).
\end{aligned}$$

Desse modo, ao se fazer uma transformação raiz quadrada numa variável aleatória Y com distribuição de Poisson com parâmetro λ , a variância de \sqrt{Y} torna-se constante ou seja, a transformação considerada satisfaz a pressuposição de homogeneidade de variância.

McCullagh & Nelder (1989) comentam ainda que pode-se usar a transformação $Y^{2/3}$ afim de melhorar a simetria de distribuição.

2.2 Modelos Lineares Generalizados

A classe dos Modelos Lineares Generalizados é uma extensão de Modelos Lineares Clássicos e foi desenvolvida por Nelder & Wedderburn (1972). Na realidade eles mostram que uma série de técnicas comumente estudadas separadamente podem ser reunidas sob o nome de Modelos Lineares Generalizados, utilizando as distribuições que pertencem à Família Exponencial (Demétrio, 1993).

Segundo McCullagh e Nelder (1989), existem três processos de ajuste de um modelo:

- i) Seleção do modelo
- ii) Estimação dos parâmetros
- iii) Predição de valores futuros.

2.2.1 Seleção do modelo

Os modelos selecionados para se ajustar um conjunto de dados são usualmente escolhidos de uma classe particular e, se o processo de ajuste do modelo é utilizado, essa classe deve ser relevante aos tipos de dados em estudo. Uma característica importante de Modelos Lineares Generalizados é que eles assumem independência entre as observações (ou pelo menos são não correlacionadas). Como consequência disso, dados exibindo autocorrelações de séries temporais e processos espaciais são excluídos. A suposição de independência é característica de um modelo linear da análise de regressão clássica, sendo conduzida sem modificações para uma classe mais ampla de Modelos Lineares Generalizados.

A escolha de escala para análise é um aspecto importante para a seleção do modelo. A escolha mais comum é entre a análise de Y , isto é, a escala original, ou $\log Y$. Para a pergunta "O que caracteriza uma boa escala?" deve-se responder que tudo depende do propósito para o qual a escala está sendo usada. Numa análise de regressão linear clássica, uma boa escala deve combinar homogeneidade de variâncias, normalidade aproximada dos erros e aditividade dos efeitos sistemáticos. No caso de análise de dados discretos, em que as observações são aproximadas pela distribuição de Poisson, os efeitos sistemáticos são, na maioria das vezes, multiplicativos. Aqui, $Y^{1/2}$ dá uma melhora na homogeneidade das variâncias, $Y^{2/3}$ melhora a simetria ou normalidade, e $\log Y$ produz a aditividade dos efeitos sistemáticos.

Com a introdução dos Modelos Lineares Generalizados, problemas com escala são grandemente reduzidos.

Note que se podemos definir exatamente o significado de um modelo

ótimo num dado contexto, é menos provável que não se tenha um que melhor se ajuste aos dados dentre todos os modelos que estão entre os escolhidos.

2.2.2 Estimação

Após escolhido o modelo, é necessário estimar os parâmetros e conhecer a precisão dos estimadores. Em Modelos Lineares Generalizados, a estimação se processa definindo uma medida de qualidade do ajuste entre dados observados e os valores gerados pelo modelo. Os estimadores dos parâmetros são os valores que minimizam o critério de ajuste.

Se $f(y; \theta)$ é a função de densidade ou distribuição de probabilidade para a observação y dado o parâmetro θ , então o logaritmo da função de verossimilhança, expressa como uma função do valor médio de Y , $\mu = E(Y)$, é:

$$l(\mu; y) = \log f(y; \mu).$$

O logaritmo da função de verossimilhança baseada num conjunto de observações independentes y_1, \dots, y_n é então, a soma das contribuições individuais, ou seja:

$$l(\mu; y) = \sum_i \log f(y; \mu),$$

em que $\mu = (\mu_1, \dots, \mu_n)$. Note que a função de densidade $f(y; \mu)$ é considerada como uma função de y para μ fixo, enquanto que o logaritmo da função de verossimilhança é primeiramente considerada como uma função de μ para os y 's observados.

Existem vantagens em usar como critério de qualidade de ajuste, não o logaritmo da função de verossimilhança $l(y; \mu)$, mas uma função linear particular, i.é,

$$D^*(y, \mu) = 2l(y; y) - 2l(\mu; y),$$

que é chamada *deviance* escalonada. Note que, para os modelos da família exponencial, $l(y, y)$ é a máxima verossimilhança calculada para um ajuste exato no qual os

valores ajustados são iguais aos dados observados. Devido a $l(y, y)$ não depender dos parâmetros, maximizar $l(y, \mu)$ é equivalente a minimizar a $D^*(y, \mu)$ em relação a μ , sujeitos à restrições impostas pelo modelo.

2.2.3 Predição

A predição está concentrada em informações sobre valores parecidos de eventos não observados, não necessariamente aqueles no futuro. Para utilizar isso, tem-se que quantidades previsivas precisam ser acompanhadas por medidas de precisão. Há a suposição de que o conjunto de dados permanece constante, e o modelo usado na análise deve ser substancialmente correto.

2.3 Os componentes de um Modelo Linear Generalizado

Um vetor de observações y com n componentes é suposto ser uma realização de uma variável aleatória Y cujos componentes são independentemente distribuídos com média μ . Em modelos lineares ordinários, esta média toma a forma

$$E(Y) = \mu = \sum_{j=1}^p x_j \beta_j,$$

em que os x_j são variáveis explanatórias e os β_j 's são parâmetros cujos valores são usualmente desconhecidos e serão estimados através dos dados. Se tomarmos i como os índices das observações, então a parte sistemática do modelo deve ser escrita como:

$$E(Y_i) = \mu_i = \sum_{j=1}^p x_{ij} \beta_j, \quad \text{para } i = 1, \dots, n,$$

em que x_{ij} é o valor da j -ésima variável para a observação i . Em notação matricial (em que μ é $n \times 1$, X é $n \times p$ e β é $p \times 1$) pode-se escrever

$$\mu = X\beta$$

em que X é a matriz do modelo e β é o vetor de parâmetros. Isto completa a parte sistemática do modelo.

Para a parte aleatória, supõe-se que os resíduos são independentes com variâncias constantes. Estas suposições são fortes e precisam ser bem justificadas utilizando-se os dados. Do mesmo modo, a estrutura da parte sistemática supõe que as covariáveis que influenciam na média são conhecidas e podem ser medidas efetivamente sem erros.

Uma adaptação posterior do modelo envolve a mais forte suposição: a de que os resíduos seguem uma distribuição Normal ou Gausseana com variância constante σ^2 .

Pode-se assim, resumir o Modelo Linear Clássico da seguinte forma: os componentes de Y são variáveis normais independentes com variância constante σ^2 e

$$E(Y) = \mu \quad \text{para} \quad \mu = X\beta.$$

Para generalizar essa idéia para Modelos Lineares Generalizados, deve-se fazer uma nova especificação desse modelo composta de três partes:

1. *Componente Aleatório*: os componentes de Y têm distribuições normais independentes com $E(Y)=\lambda$ e variância constante σ^2 ;
2. *Componente Sistemático*: as covariáveis x_1, x_2, \dots, x_n produzem um preditor linear η dado por

$$\eta = \sum_{j=1}^p x_j \beta_j;$$

3. *Ligação*: a ligação entre os componentes aleatório e sistemático é dado por:

$$\mu = \eta.$$

Essa generalização introduz um novo símbolo η que é o preditor linear, e o terceiro componente especifica que μ e η são de fato idênticos. Logo:

$$\eta_i = g(\mu_i),$$

e, então, $g(\cdot)$ será chamada de *função de ligação*. Dentro desta formulação, Modelos Lineares Clássicos têm uma distribuição normal (ou Gausseana) no componente aleatório e a função identidade para a ligação entre os componentes aleatório e sistemático. Os Modelos Lineares Generalizados permitem duas extensões: primeiro, a distribuição no componente aleatório pode vir de uma outra distribuição que não somente a normal, e segundo, a função de ligação entre os componentes aleatório e sistemático pode ser qualquer função monotônica diferenciável.

Suponha, agora, que cada componente de Y tem uma distribuição na família exponencial canônica, tomando a forma:

$$f_Y(y, \theta, \phi) = \exp\{(y\theta - b(\theta))/a(\phi) + c(y, \phi)\}, \quad (3)$$

para algumas funções específicas $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$. Se ϕ é conhecido, este é um modelo de família exponencial com parâmetro canônico θ . Se ϕ é desconhecido, ela pode ser ou não uma família exponencial biparametrizada. Assim, para a distribuição Normal com parâmetros μ e σ^2 , a função de densidade é dada por:

$$\begin{aligned} f_Y(y, \theta, \phi) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y - \mu)^2}{2\sigma^2}\right\} \\ &= \exp\left\{\frac{(y\mu - \mu^2/2)}{\sigma^2} - \frac{y^2/\sigma^2 + \log(2\pi\sigma^2)}{2}\right\}, \end{aligned}$$

e tem-se que:

$$\begin{aligned} \theta &= \mu & \phi &= \sigma^2, \\ a(\phi) &= \phi, & b(\theta) &= \frac{\theta^2}{2}, & c(y, \phi) &= \frac{1}{2}\left(\frac{y^2}{\phi^2} + \log 2\pi\sigma^2\right). \end{aligned}$$

Fazendo $l(\theta, \phi; y) = \log f_Y(y; \theta, \phi)$ para a função de verossimilhança considerada como uma função de θ e ϕ para y fixado, a média e variância de Y podem ser obtidas facilmente das relações:

$$E\left(\frac{\partial l}{\partial \theta}\right) = 0, \quad (4)$$

$$E\left(\frac{\partial^2 l}{\partial \theta^2}\right) - E\left(\frac{\partial l}{\partial \theta}\right)^2 = 0. \quad (5)$$

De (3) tem-se que

$$l(\theta, y) = \left\{ \frac{(y\theta - b(\theta))}{a(\phi)} + c(y, \phi) \right\},$$

que resulta em

$$\frac{\partial l}{\partial \theta} = \frac{\{y - b'(\theta)\}}{a(\phi)}, \quad (6)$$

e

$$\frac{\partial^2 l}{\partial \theta^2} = \frac{-b''(\theta)}{a(\phi)}, \quad (7)$$

em que $b'(\cdot)$ e $b''(\cdot)$ denotam a diferenciação de $l(\theta, y)$ em relação a θ .

De (4) e (6) tem-se que

$$E\left(\frac{\partial l}{\partial \theta}\right) = \frac{\{\mu - b'(\theta)\}}{a(\phi)} = 0,$$

obtendo:

$$E(Y) = \mu = b'(\theta).$$

Do mesmo modo, de (5) e (6) e (7), tem-se que:

$$\frac{-b''(\theta)}{a(\phi)} + \frac{Var(Y)}{a^2(\phi)} = 0,$$

que resulta em:

$$Var(Y) = b''(\theta)a(\phi).$$

Assim, a variância de Y é o produto de duas funções: uma, $b''(\theta)$, que depende do parâmetro canônico somente (e também da média) e será chamada de

função de variância, enquanto que a outra é independente de θ e depende somente de ϕ . A função de variância considerada como uma função de μ será escrita por $V(\mu)$.

A função $a(\phi)$ é comumente da forma:

$$a(\phi) = \frac{\phi}{\omega},$$

em que ϕ , também denotado por σ^2 e chamado de *parâmetro de dispersão*, é constante sobre as observações, e ω é um peso *a priori* conhecido, que varia de observação para observação. Assim, para um modelo normal, no qual cada observação é a média de n leituras independentes, têm-se

$$a(\phi) = \frac{\sigma^2}{n},$$

com $\omega = n$.

A função de ligação relaciona o preditor linear η ao valor esperado μ de uma variável Y . Em Modelos Lineares Clássicos, a média e o preditor linear são idênticos, fazendo com que a ligação identidade seja plausível. Em ambos η e μ podem tomar qualquer valor real. No entanto, quando se trata de contagens considerando uma distribuição de Poisson com parâmetro λ , deve-se ter $\lambda > 0$, o que torna a ligação identidade é menos atrativa, pois η pode ser negativo enquanto que λ não pode ser. Modelos para dados de contagens baseados na independência em classificação cruzada levam naturalmente a efeitos multiplicativos, e isso é expresso pela ligação logarítmica, $\eta = \log \lambda$, com sua inversa $\lambda = e^\eta$. Porém, efeitos aditivos contribuindo para η tornam-se efeitos multiplicativos, contribuindo para λ , e λ é necessariamente positivo.

2.4 Medindo a qualidade do ajuste

O processo para ajustar um modelo aos dados pode ser entendido como um meio de trocar valores de um conjunto de dados y por um conjunto de

valores ajustados λ oriundos de um modelo envolvendo normalmente um número de parâmetros. Em geral os λ 's não serão exatamente iguais aos y 's. A questão então surge de quão discrepantes eles são, porque uma pequena discrepância pode ser tolerável enquanto que uma grande discrepância não o é. As medidas de discrepância ou de qualidade de ajuste podem ser fornecidas de vários modos. Aqui será usada a medida formada pelo logaritmo da razão de verossimilhanças, chamada *deviance*.

Assim, considerando n observações, pode-se ajustar modelos com n parâmetros. O modelo mais simples chamado *modelo nulo*, tem um parâmetro comum λ para todos os y 's; o modelo nulo concentra toda a variação dos y 's no componente aleatório. Em outro extremo, o modelo completo tem n parâmetros, um por observação, e os λ 's obtidos deles representam os dados exatamente. O modelo completo concentra então toda a variação dos y 's no componente sistemático não sobrando nenhuma para o componente aleatório. Na prática o modelo nulo é simples demais e o modelo completo é não informativo, porque ele não representa os dados, apenas os repete. No entanto, o modelo completo fornece uma boa base para se medir a discrepância para um modelo intermediário com p parâmetros.

É mais conveniente então, expressar o logaritmo da função de verossimilhança em termos do valor médio do parâmetro λ do que do parâmetro canônico θ . Seja $l(\lambda, \phi, y)$ o logaritmo da função de verossimilhança maximizada sobre β para um valor fixo do parâmetro de dispersão ϕ . A função de máxima verossimilhança num modelo completo com n parâmetros é $l(\lambda, \phi, y)$, que é comumente finita. A discrepância de um ajuste é proporcional a duas vezes a diferença entre o máximo do logaritmo da função de verossimilhança do modelo completo e a do modelo sob pesquisa. Isto é dado por:

$$D(y, \hat{\lambda}) = 2 [l(y, y) - l(\hat{\lambda}, y)],$$

em que $D(y; \hat{\lambda})$ é conhecida como a *deviance* para os valores observados y e que depende somente dos dados.

Geralmente pode-se supor que dados de contagem provêm de uma

distribuição de Poisson com parâmetro desconhecido λ . Assim, se $Y \sim P(\lambda)$, a distribuição de probabilidade é dada por:

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!} \quad \text{para } y = 0, 1, \dots,$$

e o logaritmo da função de verossimilhança para o modelo sob investigação é dada por:

$$l(\hat{\lambda}, y) = -\hat{\lambda} + \log \hat{\lambda}^y - \log y!,$$

enquanto para o modelo superparametrizado, é dado por:

$$l(y, y) = -y + \log y^y - \log y!.$$

Assim, a *deviance* para uma variável aleatória $Y \sim P(\lambda)$ é obtida por:

$$\begin{aligned} D(y, \hat{\lambda}) &= 2 \left[l(y, y) - l(y, \hat{\lambda}) \right] = \\ &= 2 \left[-y + \log y^y - \log y! + \hat{\lambda} - \log \hat{\lambda}^y + \log y! \right] = \\ &= 2 \left[-y + \log y^y + \hat{\lambda} - \log \hat{\lambda}^y \right] = \\ &= 2 \left[y \log \left(\frac{y}{\hat{\lambda}} \right) - (y - \hat{\lambda}) \right], \end{aligned}$$

em que $\hat{\lambda}$ é o valor médio estimado do modelo sob pesquisa.

A *deviance* tem uma distribuição desconhecida, mas no caso de grandes amostras, ela apresenta uma distribuição assintoticamente χ_{n-p}^2 , onde $n - p$ é o número de graus de liberdade do modelo, sendo $n =$ número de variáveis e $p =$ número de parâmetros envolvidos no modelo, mas essa aproximação apresenta um resto da ordem de $O(m^{3/2})$, portanto ela não é boa, principalmente no caso de pequenas amostras.

2.5 Fatores de Correção

2.5.1 Correção de Bartlett

McCullagh e Nelder (1989) apresentam a seguinte proposta: uma hipótese nula simples $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ pode ser testada usando a estatística da razão de verossimilhanças, que é dada por duas vezes a diferença entre o logaritmo da função de verossimilhança em $\boldsymbol{\theta}$ e o valor obtido em $\boldsymbol{\theta}_0$. Para modelos lineares generalizados nos quais o parâmetro de dispersão é conhecido, esta diferença pode ser escrita em termos das *deviances* como segue:

$$\Lambda = 2l(\hat{\boldsymbol{\theta}}; Y) - 2l(\boldsymbol{\theta}_0; Y) = D(Y; \boldsymbol{\theta}_0) - D(Y; \hat{\boldsymbol{\theta}}).$$

Supõe-se aqui, para simplificar a notação, que o parâmetro de dispersão é igual à unidade.

Sob as condições usuais de regularidade assintótica para grandes amostras a média assintótica desta estatística é

$$\begin{aligned} E\{D(Y, \boldsymbol{\theta}_0) - D(Y; \hat{\boldsymbol{\theta}})\} &= p + \epsilon_p + O(n^{-2}) \\ &= p\{1 + b_p(\boldsymbol{\theta}_0)\} + O(n^{-2}), \end{aligned}$$

em que $p = \dim(\boldsymbol{\theta})$ e $b(\boldsymbol{\theta})$ é conhecido como o fator de Correção de Bartlett. O cumulante de ordem r é dado por:

$$\kappa_r\{D(Y; \boldsymbol{\theta}_0) - D(Y; \hat{\boldsymbol{\theta}})\} = (r - 1)!2^{r-1}p\{1 + b_p(\boldsymbol{\theta}_0)\}^r + O(n^{-2}). \quad (8)$$

O principal termo nesta expressão é o r -ésimo cumulante da dis-

tribuição de χ_p^2 . Da propriedade multiplicativa de cumulantes, pode ser visto imediatamente que os cumulantes da estatística ajustada

$$\Lambda' = \frac{\Lambda}{1 + b_p},$$

são os mesmos da distribuição de χ_p^2 , quando termos de ordem $O(n^{-2})$ são desprezados. Tem-se que b_p e ϵ_p são ambos $O(n^{-1})$, por suposição.

Apesar da convergência de os cumulantes implicar na convergência à distribuição requerida, a ordem de magnitude da discrepância neles não é necessariamente do mesmo tamanho do erro na função de distribuição acumulada. Contudo, parece razoável concluir que a distribuição da estatística ajustada é dada por

$$\Lambda' \sim \chi_p^2 + O(n^{-2}),$$

Para hipóteses nulas compostas, que são mais comuns em aplicações, uma correção similar pode ser feita. A estatística pode ser escrita da forma

$$\Lambda(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_0) = 2l(\hat{\boldsymbol{\theta}}; Y) - 2l(\hat{\boldsymbol{\theta}}_0; Y) = D(Y; \hat{\boldsymbol{\theta}}_0) - D(Y; \hat{\boldsymbol{\theta}}), \quad (9)$$

em que $\hat{\boldsymbol{\theta}}_0$ é a estimativa dos parâmetros *nuisance* sob H_0 , e $\hat{\boldsymbol{\theta}}$ é a estimativa irrestrita. Supondo que as hipóteses são aninhadas, e que $q < p$ é a dimensão do espaço paramétrico sob H_0 , a média desta estatística é

$$\begin{aligned} E\{\Lambda(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}_0)\} &= p + \epsilon_p + O(n^{-2}) - (q + \epsilon_q + O(n^{-2})) \\ &= p - q + (\epsilon_p - \epsilon_q) + O(n^{-2}) \\ &= (p - q)\{1 + b_{pq}(\boldsymbol{\theta})\} + O(n^{-2}). \end{aligned}$$

Assim o fator de correção requerido é agora

$$b_{pq} = (pb_p - qb_q)/(p - q) = (\epsilon_p - \epsilon_q)/(p - q). \quad (10)$$

Os cumulantes da estatística da razão de verossimilhanças (9) obedecem a (8), com p substituído por $p - q$. Assim os cumulantes da estatística corrigida continuam os mesmos da distribuição de χ_{p-q}^2 , separados dos termos de ordem $O(n^{-2})$.

2.5.2 Propostas de Cordeiro (1986)

Cordeiro (1986) relata que para examinar propriedades da *deviance* em pequenas amostras, torna-se necessária uma aproximação de ordem superior para o seu valor esperado, supondo o modelo sob pesquisa verdadeiro. Tem-se que a definição da *deviance* é $D_p = 2(\hat{L}_n - L) - 2(L - \hat{L}_p)$, em que \hat{L}_n , \hat{L}_p e L são os valores de $L(\beta)$ dados a seguir, nos pontos y , $\hat{\mu}$ e no parâmetro verdadeiro, respectivamente:

$$L(\beta) = \sum_{l=1}^n \phi_l [y_l \theta_l - b(\theta_l) + c(y_l, \phi_l)],$$

em que $\theta_l = q(\mu_l)$, $\mu_l = f^{-1}(\eta)$, $\eta = \sum_{i=1}^p x_{li} \beta_i$ e ϕ_l é suposto conhecido para cada observação.

Então, o valor esperado da *deviance* é expresso, até termos de ordem n^{-1} , por

$$E(D_p) = 2E(\hat{L}_n - L) - (p + \epsilon_p) + O(n^{-2}).$$

O termo $2(\hat{L}_n - L)$ representa a *deviance* no parâmetro verdadeiro e o seu valor esperado pode ser calculado a partir da distribuição proposta para os dados. Dado que:

$$\hat{L}_n = \sum_{l=1}^n \phi_l [y_l q(y_l) - b(q(y_l)) + c(y_l, \phi_l)],$$

vem, até ordem n^{-1} ,

$$E(D_p) = 2 \sum_{l=1}^n \phi_l \{E[v(Y_l)] - v(\mu_l)\} - (p + \epsilon_p), \quad (11)$$

em que

$$v(x) = xq(x) - b(q(x)),$$

com $b(\cdot)$ sendo uma função conhecida e $q(\cdot) = \theta$.

Todos os termos em (11) devem ser calculados de acordo com o modelo sob pesquisa.

O termo de ordem n em (11) é uma função das médias verdadeiras e desconhecidas, para o Modelo Binomial. Este termo pode ser calculado exatamente e a aproximação assintótica considerada é somente quando n tende a infinito.

A vantagem de (11) é que se pode deduzir um fator de correção:

$$c = (n - p)/E(D_p).$$

Então, pode-se definir uma *deviance* modificada $D_p^* = cD_p$ para o modelo sob pesquisa, tal que $E(D_p^*)$ é melhor aproximada por $(n - p)$ do que $E(D_p)$. Isto não implica, necessariamente, que D_p^* é melhor aproximada por uma distribuição χ_{n-p}^2 .

Frequentemente, $E(D_p)$ é uma função das médias desconhecidas e estas são estimadas a partir de $\tilde{\beta} = (\tilde{\beta}_1 \dots \tilde{\beta}_p)^T$ para calcular um valor numérico para c . Isto claramente não afetará a ordem da aproximação obtida. Pelo menos, com grandes amostras, os momentos de D_p^* devem ser mais próximos dos correspondentes momentos da distribuição χ_{n-p}^2 do que aqueles de D_p . Mas, para n pequeno, isto, em geral, não é verdadeiro.

2.5.3 Outras Propostas para Fatores de Correção

Botter & Cordeiro (1998) propõem a redução no vício de estimadores de máxima verossimilhança para os parâmetros do modelo linear generalizado com dispersão nas covariáveis. Para esta classe de modelos, foram obtidas fórmulas gerais para os vieses de tais estimadores, expandindo-os até termos de segunda ordem dos parâmetros do modelo linear e da dispersão bem como para os preditores lineares, parâmetros de precisão e valores da média. Alguns casos especiais são tratados utilizando a distribuição normal para a variável resposta. Estudos de simulação também são feitos para variáveis com distribuição gama e normal inversa. Já em Cordeiro & Cribari-Neto (1998), a mesma proposta foi feita considerando estimadores de máxima verossimilhança para os parâmetros dos modelos de regressão não-lineares em famílias não-exponenciais, em que foi estudado a redução no vício para os preditores lineares e as médias. Os comportamentos dos estimadores com vício reduzido foi estudado através de simulação.

Gimenes (2000) propõe um novo fator de correção para a distribuição da *deviance* para dados de proporções, que podem ser considerados como provenientes de uma distribuição binomial. Tal fator é obtido adicionando uma constante à variável resposta, supondo ainda que tal variável siga uma distribuição binomial e calculando a *deviance* para a variável corrigida. Como essa função não possui valor esperado de forma explícita, foi obtida a expansão em série de Taylor da *deviance* para a nova variável e calculado seu valor esperado, zerando-se os termos de $O(m^{-1})$. Através de dados simulados de uma distribuição binomial, *QQ-plots* foram construídos para verificar a melhora na aproximação da distribuição da *deviance* para uma distribuição χ^2 , para diversos tamanhos de amostra e diversos valores dos parâmetros.

Colosimo et al (2000) consideraram estudos de redução de vício em estimativas de parâmetros para o modelo de riscos proporcionais de Cox, considerando o estimador de máxima verossimilhança parcial como uma primeira abordagem

e métodos de reamostragem para uma segunda abordagem, sugerindo que os estimadores de máxima verossimilhança parcial têm uma performance melhor do que aqueles obtidos por reamostragem.

2.6 Desenvolvimentos Assintóticos

Seja $f(y)$ a soma de séries ascendentes com intervalo de convergência $a - r^* < y < a + r^*$ ($r^* > 0$):

$$f(y) = \sum_{n=0}^{\infty} c_n (y - a)^n, \quad a - r^* < y < a + r^*. \quad (12)$$

Esta série, segundo Kaplan (1973), é chamada de série de Taylor de $f(y)$ para $y = a$ se os coeficientes c_n são dados pela regra:

$$c_0 = f(a), \quad c_1 = \frac{f'(a)}{1!}, \quad c_2 = \frac{f''(a)}{2!}, \dots, \quad c_n = \frac{f^{(n)}(a)}{n!}, \dots,$$

tais que

$$f(x) = f(a) + \frac{f'(a)}{1!}(y - a) + \dots + \frac{f^{(n)}(a)}{n!}(y - a)^n + \dots \quad (13)$$

Teorema. Todas as séries ascendentes com convergência de área não nula são a série de Taylor desta soma.

Prova. Seja $f(y)$ dada por (12). Então, por repetidas diferenciações, encontra-se que:

$$f(y) = c_0 + c_1(y - a) + \dots + c_n(y - a)^n + \dots,$$

$$f'(y) = c_1 + 2c_2(y - a) + \dots + n.c_n(y - a)^{n-1} + \dots,$$

$$f''(y) = 2c_2 + 6c_3(y - a) + \dots + n(n - 1).c_n(y - a)^{n-2} + \dots,$$

.....,

$$\begin{aligned} f^{(n)}(y) &= n(n-1)(n-2)\dots 2.1.c_n \\ &+ (n+1)n(n-1)\dots 2.c_{n+1}(y-a) + \dots \end{aligned}$$

Aqui todas as séries convergem para $a - r^* < y < a + r^*$. Fazendo agora $y = a$, segue que

$$f(a) = c_0, \quad f'(a) = c_1, \quad f''(a) = 2c_2, \dots, \quad f^{(n)}(a) = n!c_n, \dots$$

Assim, $c_0 = f(a)$ e

$$c_n = \frac{f^{(n)}(a)}{n!}, \quad n = 1, 2, \dots,$$

encerrando a prova \diamond .

Para $a = 0$, a expressão (13) para a série de Taylor de $f(y)$ torna-se

$$f(y) = f(0) + \frac{f'(0)}{1!}y + \frac{f''(0)}{2!}y^2 + \dots + \frac{f^{(n)}(0)y^n}{n!} + \dots$$

Esta expressão é chamada série de MacLaurin de $f(y)$ que, para muitos propósitos, é mais fácil de se usar. A substituição $t = y - a$ reduz a série de Taylor geral à forma de MacLaurin.

2.6.1 Série de Taylor para funções de uma variável

De acordo com Lang (1976), pela teoria das funções de uma variável, obtém-se a seguinte expressão para os valores de uma função f na vizinhança de um ponto a , por meio das derivadas:

$$f(a+h) = f(a) + f'(a)h + \frac{f^{(2)}(a)}{2!}h^2 + \dots + \frac{f^{(r-1)}(a)}{(r-1)!}h^{r-1} + R_r,$$

em que

$$R_r = \frac{f^{(r)}(c)}{r!} h^r,$$

para algum ponto c entre a e $a + h$.

2.6.2 Série de Taylor para funções de duas ou mais variáveis

Lang (1976) diz que é possível obter uma função de duas ou mais variáveis expandida em série de Taylor. Seja uma função f de n variáveis definida num conjunto aberto U , e tendo derivadas parciais contínuas até a ordem r . Sejam P um ponto de U , e H um vetor. Suponha que o segmento de reta

$$P + tH, \quad 0 \leq t \leq 1,$$

esteja contido em U . Então existe um número τ entre 0 e 1 tal que

$$\begin{aligned} f(P + H) &= f(P) + \frac{(H \cdot \nabla) f(P)}{1!} + \dots + \frac{(H \cdot \nabla)^{r-1} f(P)}{(r-1)!} + \\ &+ \frac{(H \cdot \nabla)^r f(P + \tau H)}{r!}. \end{aligned}$$

Esta forma é obtida substituindo na Fórmula de Taylor para uma variável as derivadas da função $g(t) = f(P + tH)$ pelas suas expressões. Tem-se, assim, que

$$g^{(s)}(0) = (H \cdot \nabla)^s f(P),$$

e

$$g^{(r)}(\tau) = (H \cdot \nabla)^r f(P + \tau H).$$

Reescrevendo em termos das notações $\partial/\partial x$ e $\partial/\partial y$, segue que

$$\begin{aligned}
f(a+h, b+k) &= f(a, b) + \left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right) f(a, b) + \dots + \\
&+ \left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)^{r-1} f(a, b) + \\
&+ \left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)^r f(a + \tau h, b + \tau k).
\end{aligned}$$

As potências dos operadores diferenciais

$$\left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right),$$

são calculadas pelo desenvolvimento do binômio. Por exemplo:

$$\left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)^2 = h^2 \frac{\partial^2}{\partial x^2} + 2hk \frac{\partial^2}{\partial x \partial y} + k^2 \frac{\partial^2}{\partial y^2},$$

$$\begin{aligned}
\left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)^3 &= h^3 \left(\frac{\partial}{\partial x} \right)^3 + 3h^2k \left(\frac{\partial}{\partial x} \right)^2 \left(\frac{\partial}{\partial y} \right) + \\
&+ 3hk^2 \left(\frac{\partial}{\partial x} \right) \left(\frac{\partial}{\partial y} \right)^2 + k^3 \left(\frac{\partial}{\partial y} \right)^3,
\end{aligned}$$

e assim por diante.

Segundo Kaplan (1973), há uma fórmula de Taylor com resíduo para funções de n variáveis:

$$\begin{aligned}
F(x, y) &= F(x_1, y_1) + dF(x_1, y_1; x - x_1, y - y_1) \\
&+ \dots + \frac{1}{n!} d^n F(x_1, y_1; x - x_1, y - y_1) \\
&+ \frac{1}{(n+1)!} d^{n+1} F(x^*, y^*; x - x_1, y - y_1),
\end{aligned} \tag{14}$$

com $x^* = x_1 + t^*(x - x_1)$, $y^* = y_1 + t^*(y - y_1)$, $0 < t^* < 1$.

O ponto (x^*, y^*) se encontra entre (x_1, y_1) e (x, y) na linha de junção desses pontos. Para $n = 1$, a fórmula se torna:

$$F(x, y) = F(x_1, y_1) + (x - x_1)F_x(x^*, y^*) + (y - y_1)F_y(x^*, y^*). \quad (15)$$

que é conhecida como a Lei da Média para funções de duas variáveis.

Para provar (15), pode-se escrever:

$$\phi(t) = F[x_1 + t(x - x_1), y_1 + t(y - y_1)], \quad 0 \leq t \leq 1.$$

Tem-se que x e y são considerados fixos e ϕ depende somente de t .

Pela lei da média para ϕ ,

$$\phi(1) = \phi(0) + \phi'(t^*), \quad 0 < t^* < 1$$

Mas $\phi(1) = F(x, y)$, $\phi(0) = F(x_1, y_1)$ e

$$\begin{aligned} \phi'(t) &= (x - x_1)F_x[x_1 + t(x - x_1), y_1 + t(y - y_1)] \\ &\quad + (y - y_1)F_y[x_1 + t(x - x_1), y_1 + t(y - y_1)]. \end{aligned}$$

Se t for substituído por t^* , obtém-se (15). A fórmula geral (14) é provada da mesma maneira, baseada na fórmula de Taylor para ϕ :

$$\phi(1) = \phi(0) + \phi'(0) + \dots + \frac{\phi^{(n)}(0)}{n!} + \frac{\phi^{(n+1)}(t^*)}{(n+1)!},$$

em que $0 < t^* < 1$. Pode-se encontrar, por indução, que

$$\phi^{(n)}(t) = d^n F[x_1 + t(x - x_1), y_1 + t(y - y_1); x - x_1, y - y_1], \quad (16)$$

e a validade desta equação é assegurada se $F(x, y)$ tem derivadas contínuas através da $(n+1)$ -ésima ordem num domínio D contendo o segmento de linha que liga (x, y) a (x_1, y_1) .

A Série de Taylor ou Fórmula de Taylor pode ser usada para estudar a natureza de uma função perto de um ponto particular. Como foi dito anteriormente, os termos lineares resultam em dF , a melhor "aproximação linear" para

$F(x, y) - F(x_1, y_1)$. Se $dF = 0$, os termos quadráticos $d^2F/2!$ tornam-se de grande importância. Em particular, se a expressão quadrática

$$d^2F = A(x - x_1)^2 + 2B(x - x_1)(y - y_1) + C(y - y_1)^2,$$

for positiva, exceto para $x = x_1, y = y_1$, então $F(x, y)$ tem um mínimo no ponto (x_1, y_1) .

Em estatística, segundo Leite e Singer (1990), uma das ferramentas mais utilizadas no estudo de métodos assintóticos é a aproximação de uma função real f de variável real por um polinômio. Embora a expansão de f em série de potências na vizinhança de um ponto x_0 seja importante sob o ponto de vista teórico, o que realmente interessa nas aplicações é o estudo do resto proveniente da aproximação de f por um polinômio.

Para uma função expandida em série de Taylor, para todo $x \in U$, $R_n(x) = f(x) - P(x)$ é o resto ou o erro da aproximação de $f(x)$ por $P(x)$. Logo, se f for derivável até a ordem n em um ponto x_0 , tem-se:

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + R_n(x),$$

para todo $x \in U$. A expressão acima é denominada fórmula de Taylor, com resto R_n , de f em torno de x_0 . Tem-se que, para $n = 1$ e 2 , $R_n(x)$ tende a zero *mais rapidamente* do que $(x - x_0)^n$ quando x tende a x_0 .

Teorema. Seja f uma função real de variável real derivável até a ordem $n + 1$ em um intervalo I e sejam $x_0, x \in I$. Então existe pelo menos um ponto c entre x_0 e x tal que

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}.$$

Em outras palavras, o resto na fórmula de Taylor é dado por

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} (x - x_0)^{n+1}.$$

Observa-se que, quando $n = 0$, este teorema é precisamente o teorema do valor médio. (A prova deste teorema pode ser encontrada em Courant, 1958, cap. VI).

Leite e Singer (1990) definem, a seguir, a notação $O(\cdot)$ e $o(\cdot)$, que é muito usada no estudo de métodos assintóticos, quando se comparam sequências de variáveis aleatórias, e que será utilizada na próxima seção: sejam $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ sequências de números reais. Pode-se dizer que:

- i) $a_n = O(b_n)$ se existirem um número real $K > 0$ e um número inteiro positivo $n_0 = n_0(K)$ tal que $|a_n/b_n| \leq K, \quad \forall n \geq n_0$;
- ii) $a_n = o(b_n)$ se para todo $\epsilon > 0$ existir um número inteiro positivo $n_0 = n_0(\epsilon)$ tal que $|a_n/b_n| < \epsilon, \quad \forall n \geq n_0$.

Ou seja, pode-se dizer que $a_n = O(b_n)$ se a razão $|a_n/b_n|$ for limitada para todo n suficientemente grande e que $a_n = o(b_n)$ se $a_n/b_n \rightarrow 0$ quando $n \rightarrow \infty$. Em particular, $a_n = O(1)$ se existir um número real $K > 0$ tal que $|a_n| \leq K$ para todo n suficientemente grande e $a_n = o(1)$ se $a_n \rightarrow 0$ quando $n \rightarrow \infty$. Essencialmente, a definição anterior está relacionada com a comparação das *ordens de magnitude* (ou de grandeza) de $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$. A grosso modo, afirmar que $a_n = O(b_n)$ corresponde a dizer que a ordem de magnitude de $\{a_n\}_{n \geq 1}$ é, *no máximo, igual* à de $\{b_n\}_{n \geq 1}$ para todo n suficientemente grande; do mesmo modo, afirmar que $a_n = o(b_n)$ corresponde a dizer que a ordem de magnitude de $\{a_n\}_{n \geq 1}$ é *menor do que* a de $\{b_n\}_{n \geq 1}$, para todo n suficientemente grande.

Em Leite e Singer (1990) ainda, pode-se encontrar uma série de aplicações sobre o uso das expansões em Taylor e MacLaurin em estatística, como, por exemplo, a Lei dos Grandes Números e o Teorema Central do Limite.

3 METODOLOGIA

3.1 Expansão da *deviance* para uma variável sem correção

Como já visto anteriormente, McCullagh & Nelder (1989) citam que sua a distribuição da *deviance* é desconhecida, mas no caso de uma variável aleatória com distribuição de Poisson com parâmetro λ , ela poder ser aproximada por uma distribuição com erro da ordem $O(\lambda^{-1})$. Também, se a verdadeira distribuição da *deviance* fosse χ^2 , seus momentos calculados deveriam coincidir com os momentos da distribuição χ^2 .

Para verificar essa afirmativa, considere a variável aleatória $Y \sim P(\lambda)$. A *deviance* calculada para essa variável é dada por:

$$D(y, \lambda) = 2 \left[y \ln \frac{y}{\lambda} - (y - \lambda) \right]$$

e o valor esperado da *deviance* é dado por:

$$E(D(Y, \lambda)) = 2 \sum_{y=1}^{+\infty} \left(y \ln \frac{y}{\lambda} - (y - \lambda) \right) \frac{e^{-\lambda} \lambda^y}{y!}, \quad (17)$$

que não tem valor explícito. Note que o somatório inicia em um, uma vez que não existe logaritmo de zero.

Um valor aproximado para a esperança da *deviance* dada em (17), pode ser obtido pela expansão em série de Taylor. Seja

$$g(y) = D(y, \lambda) = 2 \left[y \ln \left(\frac{y}{\lambda} \right) - (y - \lambda) \right].$$

A expansão de $g(y)$ em série de Taylor em torno de λ , é dada por:

$$g(y) = g(\lambda) + g'(\lambda)(y - \lambda) + g''(\lambda)\frac{(y - \lambda)^2}{2!} + g'''(\lambda)\frac{(y - \lambda)^3}{3!} + g^{iv}(\lambda)\frac{(y - \lambda)^4}{4!} + \dots$$

Fazendo-se os cálculos, obtêm-se:

$$g(y) = 2 \left[y \ln\left(\frac{y}{\lambda}\right) - (y - \lambda) \right] \implies g(\lambda) = 2 [\lambda \ln(1) - 0] = 0,$$

$$g'(y) = 2 \left[\ln\left(\frac{y}{\lambda}\right) + y \frac{1}{\lambda} \frac{1}{y/\lambda} - 1 \right] \implies g'(\lambda) = 2[0 + 1 - 1] = 0,$$

$$g''(y) = 2 \left[\frac{1}{\lambda} \frac{1}{y/\lambda} \right] \implies g''(\lambda) = 2 \left[\frac{1}{\lambda} \right],$$

$$g'''(y) = 2 \left[-\frac{1}{y^2} \right] \implies g'''(\lambda) = 2 \left[-\frac{1}{\lambda^2} \right],$$

$$g^{iv}(y) = 2 \left[\frac{2}{y^3} \right] \implies g^{iv}(\lambda) = 2 \left[\frac{2}{\lambda^3} \right].$$

Logo:

$$\begin{aligned} g(y) &= 2 \left[0 + 0(y - \lambda) + \frac{1}{\lambda} \frac{(y - \lambda)^2}{2!} - \frac{1}{\lambda^2} \frac{(y - \lambda)^3}{3!} + \frac{2}{\lambda^3} \frac{(y - \lambda)^4}{4!} + \dots \right] \\ &= 2 \left[0 + 0 + \frac{1}{2\lambda} (y - \lambda)^2 - \frac{1}{6\lambda^2} (y - \lambda)^3 + \frac{1}{12\lambda^3} (y - \lambda)^4 + \dots \right]. \end{aligned}$$

Logo, calculando o valor esperado, tem-se:

$$\begin{aligned} E(D(Y, \lambda)) &= 2E \left[\frac{1}{2\lambda} (Y - \lambda)^2 - \frac{1}{6\lambda^2} (Y - \lambda)^3 + \frac{1}{12\lambda^3} (Y - \lambda)^4 + \dots \right] \\ &= 2 \left[\frac{1}{2\lambda} E(Y - \lambda)^2 - \frac{1}{6\lambda^2} E(Y - \lambda)^3 + \frac{1}{12\lambda^3} E(Y - \lambda)^4 + \dots \right] \\ &= 2 \left[\frac{1}{2\lambda} \lambda - \frac{1}{6\lambda^2} \lambda + \frac{1}{12\lambda^3} (3\lambda^2 + \lambda) + \dots \right] \\ &= 2 \left[\frac{1}{2} - \frac{1}{6\lambda} + \frac{1}{4\lambda} + \frac{1}{12\lambda^2} \dots \right] \\ &= 2 \left[\frac{1}{2} + \frac{1}{12\lambda} + \frac{1}{12\lambda^2} + \dots \right] \\ &= \left[1 + \frac{1}{6\lambda} + \frac{1}{6\lambda^2} + \dots \right] \\ &= 1 + \frac{1}{6\lambda} + O(\lambda^{-2}). \end{aligned}$$

Da mesma forma pode-se calcular a variância da *deviance*. Isto é:

$$\begin{aligned}
 \text{Var}(D(Y, \lambda)) &= E [D(Y, \lambda) - E(D(Y, \lambda))]^2 \\
 &= E(D^2(Y, \lambda)) - 2E(D(Y, \lambda))D(Y, \lambda) + E^2(D(Y, \lambda)) \\
 &= E(D^2(Y, \lambda)) - E^2(D(Y, \lambda)).
 \end{aligned}$$

A esperança de $D^2(Y, \lambda)$ é dada por:

$$\begin{aligned}
 E(D^2(Y, \lambda)) &= 4E \left[Y \ln \frac{Y}{\lambda} - (Y - \lambda) \right]^2 \\
 &= E \left[\frac{1}{\lambda^2} (Y - \lambda)^4 - \frac{2}{3\lambda^3} (Y - \lambda)^5 + \frac{4}{9\lambda^4} (Y - \lambda)^6 - \frac{14}{45\lambda^5} (Y - \lambda)^7 \right. \\
 &\quad \left. + \frac{41}{180\lambda^6} (Y - \lambda)^8 + O(Y - \lambda)^9 \right] \\
 &= \frac{1}{\lambda^2} E(Y - \lambda)^4 - \frac{2}{3\lambda^3} E(Y - \lambda)^5 + \frac{4}{9\lambda^4} E(Y - \lambda)^6 - \frac{14}{45\lambda^5} E(Y - \lambda)^7 \\
 &\quad + \frac{41}{180\lambda^6} E(Y - \lambda)^8 + \dots \\
 &= \frac{1}{\lambda^2} \lambda(3\lambda + 1) - \frac{2}{3\lambda^3} \lambda(10\lambda + 1) + \frac{4}{9\lambda^4} \lambda(15\lambda^2 + 25\lambda + 1) \\
 &\quad - \frac{14}{45\lambda^5} \lambda(105\lambda^2 + 56\lambda + 1) + \frac{41}{180\lambda^6} \lambda(105\lambda^3 + 490\lambda^2 + 119\lambda + 1) + \dots \\
 &= \frac{1}{\lambda} (3\lambda + 1) - \frac{2}{3\lambda^2} (10\lambda + 1) + \frac{4}{9\lambda^3} (15\lambda^2 + 25\lambda + 1) \\
 &\quad - \frac{14}{45\lambda^4} (105\lambda^2 + 56\lambda + 1) + \frac{41}{180\lambda^5} (105\lambda^3 + 490\lambda^2 + 119\lambda + 1) + \dots \\
 &= 3 + \frac{1}{\lambda} - \frac{20}{3\lambda} + \frac{20}{3\lambda} + O(\lambda^{-2}) \\
 &= 3 + \frac{1}{\lambda} + O(\lambda^{-2}).
 \end{aligned}$$

Como visto anteriormente,

$$\text{Var}(D(Y, \lambda)) = E(D^2(Y, \lambda)) - E^2(D(Y, \lambda)).$$

Logo:

$$\text{Var}(D(Y, \lambda)) = E(D^2(Y, \lambda)) - E^2(D(Y, \lambda))$$

$$\begin{aligned}
&= 3 + \frac{1}{\lambda} - \left[1 + \frac{1}{6\lambda} + \dots\right]^2 \\
&= 3 + \frac{1}{\lambda} - 1 - \frac{2}{6\lambda} - \frac{1}{36\lambda^2} + \dots \\
&= 2 + \frac{3-1}{3\lambda} + O(\lambda^{-2}) \\
&= 2 + \frac{2}{3\lambda} + O(\lambda^{-2}).
\end{aligned}$$

3.2 Expansão da *deviance* para uma variável com correção

Para melhorar a distribuição da *deviance* e, conseqüentemente, o seu valor esperado, vários fatores de correção têm sido propostos, como já visto anteriormente. Aqui, será introduzido um novo fator de correção $a \in \mathcal{R}$, substituindo y por $y+a$ e, conseqüentemente, λ por $\lambda+a$. Assim, supondo $Y+a \sim P(\lambda)$, a *deviance* será dada por:

$$\begin{aligned}
D(y+a, \lambda+a) &= 2 \left[(y+a) \ln \frac{y+a}{\lambda+a} - (y+a-\lambda-a) \right] \\
&= 2 \left[(y+a) \ln \frac{y+a}{\lambda+a} - (y-\lambda) \right]. \tag{18}
\end{aligned}$$

Considerando novamente $D(y+a, \lambda+a) = g(y)$ e calculando a expansão em série de Taylor em torno de λ , segue que

$$g(y) = 2 \left[(y+a) \ln \left(\frac{y+a}{\lambda+a} \right) - (y-\lambda) \right] \quad \implies \quad g(\lambda) = 2 [(\lambda+a) \ln(1) - 0] = 0,$$

$$g'(y) = 2 \left[\ln \left(\frac{y+a}{\lambda+a} \right) + 1 - 1 \right] \quad \implies \quad g'(\lambda) = 2 [\ln 1 + 1 - 1] = 0,$$

$$g''(y) = 2 \left[\frac{1}{(y+a)} \right] \quad \implies \quad g''(\lambda) = 2 \left[\frac{1}{\lambda+a} \right] = \frac{2}{\lambda+a},$$

$$g'''(y) = -\frac{2}{(y+a)^2} \quad \implies \quad g'''(\lambda) = -\frac{2}{(\lambda+a)^2},$$

$$g^{iv}(y) = 2 \left[\frac{2(y+a)}{(y+a)^4} \right] = \frac{4}{(y+a)^3} \quad \implies \quad g^{iv}(\lambda) = \frac{4}{(\lambda+a)^3}.$$

Desta forma:

$$\begin{aligned} g(y) &= 0 + 0(y-\lambda) + \frac{2}{\lambda+a} \frac{(y-\lambda)^2}{2!} - \frac{2}{(\lambda+a)^2} \frac{(y-\lambda)^3}{3!} + \frac{4}{(\lambda+a)^3} \frac{(y-\lambda)^4}{4!} + \dots \\ &= 0 + 0 + \frac{2}{2(\lambda+a)} (y-\lambda)^2 - \frac{1}{3(\lambda+a)^2} (y-\lambda)^3 + \frac{1}{6(\lambda+a)^3} (y-\lambda)^4 + \dots \end{aligned}$$

Calculando o valor esperado, tem-se:

$$\begin{aligned} E(D(Y+a, \lambda+a)) &= E \left[\frac{2}{2(\lambda+a)} (Y-\lambda)^2 - \frac{1}{3(\lambda+a)^2} (Y-\lambda)^3 \right. \\ &\quad \left. + \frac{1}{6(\lambda+a)^3} (Y-\lambda)^4 + \dots \right] \\ &= \frac{2}{2(\lambda+a)} E(Y-\lambda)^2 - \frac{1}{3(\lambda+a)^2} E(Y-\lambda)^3 \\ &\quad + \frac{1}{6(\lambda+a)^3} E(Y-\lambda)^4 + \dots \\ &= \frac{1}{\lambda+a} \lambda - \frac{1}{3(\lambda+a)^2} \lambda + \frac{1}{6(\lambda+a)^3} (3\lambda^2 + \lambda) + \dots \quad (19) \end{aligned}$$

Expandindo em série de Taylor em torno de $a = 0$ os termos da expressão (19), têm-se:

$$\frac{\lambda}{\lambda+a} = 1 - \frac{a}{\lambda} + \frac{a^2}{\lambda^2} + \dots,$$

$$\frac{\lambda}{(\lambda+a)^2} = \frac{1}{\lambda} - \frac{a}{2\lambda^2} + \dots,$$

$$\frac{\lambda^2}{(\lambda+a)^3} = \frac{1}{\lambda} - \frac{3a}{\lambda^2} + \dots,$$

$$\frac{3\lambda^2 + \lambda}{(\lambda+a)^3} = \frac{3}{\lambda} - \frac{9a}{\lambda^2} + \dots$$

Utilizando os termos até $O(\lambda^{-1})$ das expansões consideradas, segue que:

$$\begin{aligned}
E(D(Y + a, \lambda + a)) &= \frac{\lambda}{\lambda + a} - \frac{\lambda}{3(\lambda + a)^2} + \frac{3\lambda^2 + \lambda}{6(\lambda + a)^3} + \dots \\
&= 1 - \frac{a}{\lambda} + \frac{a^2}{\lambda^2} - \frac{a^3}{\lambda^3} + \frac{a^4}{\lambda^4} - \frac{1}{3\lambda} + \frac{2a}{3\lambda^2} - \frac{a^2}{\lambda^3} + \frac{4a^3}{3\lambda^4} + \frac{3\lambda^2 + \lambda}{6\lambda^3} \\
&\quad - \frac{3\lambda^2 a + a\lambda}{2\lambda^4} + \frac{3\lambda^2 a^2 + a^2\lambda}{\lambda^5} + \dots \\
&= 1 - \frac{a}{\lambda} - \frac{1}{3\lambda} + \frac{3\lambda^2}{6\lambda^3} + O(\lambda^{-2}) \\
&= 1 - \frac{6a - 2 + 3}{6\lambda} + O(\lambda^{-2}) \\
&= 1 - \frac{6a - 1}{6\lambda} + O(\lambda^{-2}) \\
&= 1 - \frac{1}{6\lambda}(6a - 1) + O(\lambda^{-2}).
\end{aligned}$$

Para reduzir a ordem de convergência na expansão da *deviance*, o termo da ordem de $O(\lambda^{-1})$ será igualado a zero. Logo:

$$\begin{aligned}
\frac{1}{6\lambda}(6a - 1) &= 0 \\
\implies 6a - 1 &= 0 \\
\implies 6a &= 1 \\
\implies a &= \frac{1}{6},
\end{aligned}$$

obtendo assim o valor de a que reduz a ordem de convergência na expansão do valor esperado da *deviance*, para as variáveis com distribuição de Poisson com parâmetro λ .

Calculando a variância para a variável corrigida, tem-se:

$$Var(D(Y + a, \lambda + a)) = E(D^2(Y + a, \lambda + a)) - E^2(D(Y + a, \lambda + a)).$$

A esperança de $D^2(Y + a, \lambda + a)$ é dada por:

$$E(D^2(Y + a, \lambda + a)) = 4 \left[E[(Y + a) \ln \frac{Y + a}{\lambda + a} - (Y - \lambda)]^2 \right]$$

$$\begin{aligned}
&= E \left[\frac{1}{(\lambda + a)^2} (Y - \lambda)^4 - \frac{2}{3(\lambda + a)^3} (Y - \lambda)^5 \right. \\
&+ \frac{4}{9(\lambda + a)^4} (Y - \lambda)^6 - \frac{14}{45(\lambda + a)^5} (Y - \lambda)^7 \\
&+ \left. \frac{41}{180(\lambda + a)^6} (Y - \lambda)^8 + \dots \right] \\
&= \frac{1}{(\lambda + a)^2} E(Y - \lambda)^4 - \frac{2}{3(\lambda + a)^3} E(Y - \lambda)^5 \\
&+ \frac{4}{9(\lambda + a)^4} E(Y - \lambda)^6 - \frac{14}{45(\lambda + a)^5} E(Y - \lambda)^7 \\
&+ \frac{41}{180(\lambda + a)^6} E(Y - \lambda)^8 + \dots \\
&= \frac{1}{(\lambda + a)^2} \lambda(3\lambda + 1) - \frac{2}{3(\lambda + a)^3} \lambda(10\lambda + 1) \\
&+ \frac{4}{9(\lambda + a)^4} \lambda(15\lambda^2 + 25\lambda + 1) \\
&- \frac{14}{45(\lambda + a)^5} \lambda(105\lambda^2 + 56\lambda + 1) \\
&+ \frac{41}{180(\lambda + a)^6} \lambda(105\lambda^3 + 490\lambda^2 + 119\lambda + 1) + \dots \quad .
\end{aligned}$$

Como $Var(D(Y+a, \lambda+a)) = E(D^2(Y+a, \lambda+a)) - E^2(D(Y+a, \lambda+a))$,

segue que:

$$\begin{aligned}
Var(D(Y + a, \lambda + a)) &= \frac{1}{\lambda + a} (3\lambda + 1) - \frac{2}{3(\lambda + a)^2} (10\lambda + 1) \\
&+ \frac{4}{9(\lambda + a)^3} (15\lambda^2 + 25\lambda + 1) - \frac{14}{45(\lambda + a)^4} (105\lambda^2 + 56\lambda + 1) \\
&+ \frac{41}{180(\lambda + a)^5} (105\lambda^3 + 490\lambda^2 + 119\lambda + 1) + \dots \\
&- \left[1 - \frac{6a - 1}{6\lambda} + \dots \right]^2 \\
&= \frac{1}{180\lambda^2(\lambda + a)^6} \left[-360\lambda^8 - 120\lambda^7 - 1440a\lambda^7 - 300\lambda^6 \right. \\
&+ \left. 780a\lambda^6 \right] - \frac{1}{180\lambda^2(\lambda + a)^6} \left[2520a^2\lambda^6 - 17034\lambda^5 \right] + \dots \\
&= \frac{1}{180\lambda^2(\lambda + a)^6} \left[2270a\lambda^5 + 1860a^2\lambda^5 - 2880a^3\lambda^5 \right. \\
&- \left. 4823\lambda^4 \right] + \frac{1}{180\lambda^2(\lambda + a)^6} \left[2976a\lambda^4 - 1565a^2\lambda^4 + 780a^3\lambda^4 - \dots \right]
\end{aligned}$$

$$\begin{aligned}
& - 234a^4\lambda^4 - 41\lambda^3] + \frac{1}{180\lambda^2(\lambda+a)^6} [56a\lambda^3 - 80a^2\lambda^3 + \\
& + 220a^3\lambda^3 - 480a^4\lambda^3 - 720a^5\lambda^3] + \frac{1}{180\lambda^2(\lambda+a)^6} [75a^4\lambda^2 + \\
& + 720a^6\lambda^2 - 540a^5\lambda^2 + 30a^5\lambda - 300a^6\lambda] + \\
& + \frac{1}{180\lambda^2(\lambda+a)^6} [720a^7\lambda + 5a^6 - 60a^7 + 180a^8] + \dots \quad .
\end{aligned}$$

Expandindo a série (20 em torno de $a=0$, tem-se:

$$\begin{aligned}
Var(D(Y+a, \lambda+a)) &= \frac{1}{540\lambda^5} [1080\lambda^5 + 360\lambda^4 + 9510\lambda^3 + 91282\lambda^2 + 24227\lambda + \\
& + 205] - \frac{a}{18\lambda^6} [72\lambda^5 + 150\lambda^4 + 2129\lambda^3 + 18554\lambda^2 + \\
& + 4851\lambda + 41] + \frac{a^2}{36\lambda^7} [288\lambda^5 + 1068\lambda^4 + 16351\lambda^3 + \\
& + 131382\lambda^2 + 33985\lambda + 287] + \frac{2a^3}{27\lambda^8} [162\lambda^5 + 954\lambda^4 + \\
& + 17610\lambda^3 + 132518\lambda^2 + 34006\lambda + 287] + \frac{a^4}{18\lambda^9} [270\lambda^5 + \\
& + 2490\lambda^4 + 56065\lambda^3 + 400218\lambda^2 + 102067\lambda + 861] + \\
& + \frac{a^5}{45\lambda^{10}} [810\lambda^5 + 10770\lambda^4 + 294175\lambda^3 + 2011786\lambda^2 + \\
& + 510531\lambda + 4305] + O(a^6) = \\
&= \frac{1080\lambda^5}{540\lambda^5} + \frac{360\lambda^4}{540\lambda^5} - \frac{72a\lambda^5}{18\lambda^6} + O(\lambda^{-2}) = \\
&= 2 + \frac{2}{3\lambda} - \frac{4a}{\lambda} + O(\lambda^{-2}) = \\
&= 2 + \frac{1}{\lambda} \left(\frac{2}{3} - 4a \right) + O(\lambda^{-2}).
\end{aligned}$$

Igualando o termo de ordem $O(\lambda^{-1})$ a zero, tem-se:

$$\begin{aligned}
\frac{2}{3} - 4a &= 0 \\
\implies \frac{2}{3} &= 4a \\
\implies \frac{2}{12} &= a \\
\implies a &= \frac{1}{6}
\end{aligned}$$

como visto anteriormente.

3.3 Expansão da *deviance* para n variáveis sem correção

Sejam agora Y_1, Y_2, \dots, Y_n , n variáveis aleatórias independentes e identicamente distribuídas, com distribuição de Poisson com parâmetro λ . A *deviance* para n variáveis aleatórias é dada por:

$$D(\mathbf{y}, \hat{\lambda}) = 2 \sum_{i=1}^n D(y_i, \hat{\lambda}).$$

No caso da distribuição de Poisson, tem-se que:

$$D(\mathbf{y}, \hat{\lambda}) = 2 \sum_{i=1}^n \left[y_i \ln \frac{y_i}{\hat{\lambda}} - (y_i - \hat{\lambda}) \right],$$

em que $\hat{\lambda} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

O valor esperado da *deviance* para n variáveis é dado por:

$$\begin{aligned} E(D(\mathbf{Y}, \hat{\lambda})) &= 2E \sum_{i=1}^n \left[(Y_i \ln \frac{Y_i}{\hat{\lambda}} - (Y_i - \hat{\lambda})) \right] = \\ &= E \left[2 \sum_{i=1}^n Y_i \ln \frac{Y_i}{\hat{\lambda}} - 2 \sum_{i=1}^n (Y_i - \hat{\lambda}) \right] = \\ &= E \left[2 \sum_{i=1}^n Y_i \ln \frac{Y_i}{\hat{\lambda}} - 2 \left(\sum_{i=1}^n Y_i - n\hat{\lambda} \right) \right]. \end{aligned}$$

Mas $E(Y_i) = \lambda$ e, portanto:

$$\begin{aligned} E(D(\mathbf{Y}, \lambda)) &= E \left[2 \sum_{i=1}^n Y_i \ln \frac{Y_i}{\hat{\lambda}} - 2 \left(\sum_{i=1}^n Y_i - n\hat{\lambda} \right) \right] = \\ &= 2E \left[\sum_{i=1}^n Y_i \ln \frac{Y_i}{\hat{\lambda}} \right]. \end{aligned}$$

Considere a função de n variáveis:

$$g(\mathbf{y}) = g(y_1, y_2, \dots, y_n) = \sum_{i=1}^n y_i \ln \frac{y_i}{\hat{\lambda}} =$$

$$\begin{aligned}
&= \sum_{i=1}^n y_i \ln \frac{ny_i}{\sum_{i=1}^n y_i} = \\
&= y_1 \ln \frac{ny_1}{\sum_{i=1}^n y_i} + y_2 \ln \frac{ny_2}{\sum_{i=1}^n y_i} + \dots + y_n \ln \frac{ny_n}{\sum_{i=1}^n y_i}.
\end{aligned}$$

A expansão de $g(\mathbf{y})$ em série de Taylor em torno de $\boldsymbol{\lambda}$ é dada por:

$$g(\mathbf{y}) = g(\boldsymbol{\lambda}) + dg(\boldsymbol{\lambda})(\mathbf{y} - \boldsymbol{\lambda}) + \frac{1}{2!} d^2g(\boldsymbol{\lambda})(\mathbf{y} - \boldsymbol{\lambda})^2 + \frac{1}{3!} d^3g(\boldsymbol{\lambda})(\mathbf{y} - \boldsymbol{\lambda})^3 + \dots$$

em que:

$$dg(\boldsymbol{\lambda}) = \sum_{i=1}^n \left. \frac{\partial dg}{\partial y_i} \right|_{y_i=\lambda},$$

$$d^2g(\boldsymbol{\lambda}) = \sum_{i=1}^n \left. \frac{\partial^2 dg}{\partial y_i^2} \right|_{y_i=\lambda} + 2 \sum_{i<j=2}^n \left. \frac{\partial^2 dg}{\partial y_i \partial y_j} \right|_{y_i=\lambda},$$

$$d^3g(\boldsymbol{\lambda}) = \sum_{i=1}^n \left. \frac{\partial^3 dg}{\partial y_i^3} \right|_{y_i=\lambda} + 3 \sum_{i<j=2}^n \left. \frac{\partial^3 dg}{\partial^2 y_i \partial y_j} \right|_{y_i=\lambda} + 3 \sum_{i<j=2}^n \left. \frac{\partial^3 dg}{\partial y_i \partial^2 y_j} \right|_{y_i=\lambda},$$

e assim sucessivamente.

Os termos do desenvolvimento em série de Taylor envolvendo derivadas mistas não serão utilizados já que tais termos produzem valor esperado nulo. Desse modo, será considerada apenas a derivada mista para a derivada de quarta ordem. Desse modo, o desenvolvimento em série da função $g(\mathbf{y})$ a ser utilizado, terá a forma:

$$\begin{aligned}
g(y_1, y_2, \dots, y_n) &= g(\boldsymbol{\lambda}) + \sum_{i=1}^n \left. \frac{\partial g}{\partial y_i} \right|_{y_i=\lambda} (y_i - \lambda) + \frac{1}{2!} \sum_{i=1}^n \left. \frac{\partial^2 g}{\partial y_i^2} \right|_{y_i=\lambda} (y_i - \lambda)^2 \\
&+ \frac{1}{3!} \sum_{i=1}^n \left. \frac{\partial^3 g}{\partial y_i^3} \right|_{y_i=\lambda} (y_i - \lambda)^3 + \frac{1}{4!} \sum_{i=1}^n \left. \frac{\partial^4 g}{\partial y_i^4} \right|_{y_i=\lambda} (y_i - \lambda)^4 \\
&+ \frac{1}{2!2!} \sum_{i=1}^n \left. \frac{\partial^4 g}{\partial y_i^2 \partial y_j^2} \right|_{y_i=\lambda} (y_i - \lambda)^2 (y_j - \lambda)^2 + \dots \quad . \quad (20)
\end{aligned}$$

Assim, tem-se que:

$$g(\mathbf{y}) = \sum_{i=1}^n y_i \ln \frac{y_i}{\bar{y}} = \sum_{i=1}^n y_i \ln \frac{ny_i}{\sum_{i=1}^n y_i} \quad \Longrightarrow \quad g(\boldsymbol{\lambda}) = 0.$$

Calculando a 1ª derivada de $g(\mathbf{y})$ em relação a y_1 , tem-se:

$$\begin{aligned} \frac{\partial g}{\partial y_1} &= \ln \frac{ny_1}{\sum_{i=1}^n y_i} + y_1 \frac{n \sum_{i=1}^n y_i - ny_1}{(\sum_{i=1}^n y_i)^2} \frac{\sum_{i=1}^n y_i}{ny_1} + y_2 \frac{-ny_2}{(\sum_{i=1}^n y_i)^2} \frac{\sum_{i=1}^n y_i}{ny_2} + \dots \\ &+ y_n \frac{-ny_n}{(\sum_{i=1}^n y_i)^2} \frac{\sum_{i=1}^n y_i}{ny_n} \\ &= \ln \frac{ny_1}{\sum_{i=1}^n y_i}. \end{aligned}$$

Substituindo \mathbf{y} por $\boldsymbol{\lambda}$, tem-se:

$$\left. \frac{\partial g}{\partial y_1} \right|_{\mathbf{y}=\boldsymbol{\lambda}} = \ln \frac{n\lambda}{n\lambda} = \ln 1 = 0$$

e, como as demais derivadas de $g(\mathbf{y})$ no ponto $\mathbf{y} = \boldsymbol{\lambda}$ produzem o mesmo valor, segue que:

$$\sum_{i=1}^n \left. \frac{\partial g}{\partial y_i} \right|_{y_i=\lambda} E(Y_i - \lambda) = 0.$$

Derivando a 2ª vez em relação a y_1 , tem-se:

$$\begin{aligned} \frac{\partial^2 g}{\partial y_1^2} &= \frac{n \sum_{i=1}^n y_i - ny_1}{(\sum_{i=1}^n y_i)^2} \frac{\sum_{i=1}^n y_i}{ny_1} \\ &= \frac{\sum_{i=1}^n y_i - y_1}{y_1 \sum_{i=1}^n y_i} \\ &= \frac{1}{y_1} - \frac{1}{\sum_{i=1}^n y_i} \end{aligned}$$

e fazendo $\mathbf{y} = \boldsymbol{\lambda}$:

$$\left. \frac{\partial^2 g}{\partial y_1^2} \right|_{\mathbf{y}=\boldsymbol{\lambda}} = \frac{1}{\lambda} - \frac{1}{n\lambda} = \frac{n-1}{n\lambda}.$$

Logo, usando os mesmos argumentos anteriores, tem-se que:

$$\frac{1}{2!} \sum_{i=1}^n \frac{\partial^2 g}{\partial y_i^2} \Big|_{\mathbf{y}=\boldsymbol{\lambda}} E(Y_i - \lambda)^2 = \frac{1}{2!} \frac{n(n-1)\lambda}{n\lambda} = \frac{n-1}{2}.$$

Para a 3ª e 4ª derivadas, tem-se:

$$\frac{\partial^3 g}{\partial y_1^3} = \frac{-1}{y_1^2} + \frac{1}{(\sum_{i=1}^n y_i)^2} \quad \Rightarrow \quad \frac{\partial^3 g}{\partial y_1^3} \Big|_{\mathbf{y}=\boldsymbol{\lambda}} = \frac{-1}{\lambda^2} + \frac{1}{(n\lambda)^2} = \frac{1-n^2}{(n\lambda)^2}$$

$$\frac{\partial^4 g}{\partial y_1^4} = \frac{2}{y_1^3} - \frac{2}{(\sum_{i=1}^n y_i)^3} \quad \Rightarrow \quad \frac{\partial^4 g}{\partial y_1^4} \Big|_{\mathbf{y}=\boldsymbol{\lambda}} = \frac{2}{\lambda^3} - \frac{2}{(n\lambda)^3} = 2 \frac{n^3-1}{(n\lambda)^3}.$$

Portanto,

$$\frac{1}{3!} \sum_{i=1}^n \frac{\partial^3 g}{\partial y_i^3} \Big|_{\mathbf{y}=\boldsymbol{\lambda}} E(Y_i - \lambda)^3 = \frac{1}{6} \frac{n(1-n^2)\lambda}{n^2\lambda^2} = \frac{1-n^2}{6n\lambda}$$

e

$$\frac{1}{4!} \sum_{i=1}^n \frac{\partial^4 g}{\partial y_i^4} \Big|_{\mathbf{y}=\boldsymbol{\lambda}} E(Y_i - \lambda)^4 = \frac{1}{12} \frac{(n^3-1)\lambda(3\lambda+1)}{n^2\lambda^3}.$$

Calculando a derivada mista de ordem quatro com relação às derivadas segundas de y_1 e y_2 , tem-se :

$$\frac{\partial^4 g}{\partial y_1^2 \partial y_2^2} = \frac{-2 \sum_{i=1}^n y_i}{(\sum_{i=1}^n y_i)^4} = \frac{-2}{(\sum_{i=1}^n y_i)^3} \quad \Rightarrow \quad \frac{\partial^4 g}{\partial y_1^2 \partial y_2^2} \Big|_{\mathbf{y}=\boldsymbol{\lambda}} = \frac{-2}{(n\lambda)^3}.$$

Portanto,

$$\frac{1}{2!2!} \sum_{i < j} \frac{\partial^4 g}{\partial y_i^2 \partial y_j^2} \Big|_{\mathbf{y}=\boldsymbol{\lambda}} E(y_i - \hat{\lambda})^2 E(y_j - \lambda)^2 = \frac{1}{4} \frac{n(n-1)}{2} \frac{-2}{(n\lambda)^3} \lambda^2 = \frac{-1}{4} \frac{(n-1)}{n^2\lambda}.$$

Logo, o desenvolvimento em série de Taylor em torno do ponto $\boldsymbol{\lambda}$ do valor esperado da *deviance* é:

$$\begin{aligned}
E(D(\mathbf{y}, \lambda)) &= E\left[2 \sum_{i=1}^n D(Y_i, \lambda)\right] = \\
&= 2 \left[0 + 0 + \frac{n-1}{2} + \frac{1-n^2}{6n\lambda} + \frac{1}{12} \frac{(n^3-1)(3\lambda+1)}{n^2\lambda^2} \right. \\
&\quad \left. - \frac{1}{4} \frac{(n-1)}{n^2\lambda} + \dots \right] \\
&= (n-1) + \frac{1-n^2}{3n\lambda} + \frac{1}{6} \frac{(n^3-1)(3\lambda+1)}{n^2\lambda^2} - \frac{1}{2} \frac{(n-1)}{n^2\lambda} + \dots \\
&= (n-1) + \frac{1}{6\lambda} \left[\frac{n^2-1}{n} \right] + O(\lambda^{-2}).
\end{aligned}$$

Assim, nota-se que, no caso do modelo de médias para variáveis aleatórias independentes e identicamente distribuídas segundo uma distribuição de Poisson com parâmetro λ , o valor esperado da *deviance* se aproxima do número de graus de liberdade da distribuição $\chi^2_{(n-1)}$ apenas para n pequeno, uma vez que os termos envolvendo fatores da $O(\lambda^{-1})$ dependem de n . Assim, quando n for grande, pode-se não ter tal convergência. Portanto, um fator de correção pode melhorar essa aproximação.

3.4 Expansão da *deviance* para n variáveis com correção

Novamente, a fim de reduzir o viés na convergência da distribuição da *deviance* a uma distribuição χ^2 , utiliza-se o mesmo modelo de médias para propor um novo fator de correção para as variáveis resposta.

Assim, sejam agora $Y_1 + a, Y_2 + a, \dots, Y_n + a$ variáveis aleatórias independentes e identicamente distribuídas, com distribuição de Poisson com parâmetro λ e considere a *deviance* obtida para as novas variáveis, que é dada por:

$$D(\mathbf{y} + a, \lambda + a) = 2 \sum_{i=1}^n D(y_i + a, \lambda + a)$$

$$\begin{aligned}
&= 2 \sum_{i=1}^n \left[(y_i + a) \ln \frac{y_i + a}{\hat{\lambda} + a} - (y_i - \hat{\lambda}) \right] \\
&= 2 \sum_{i=1}^n \left[(y_i + a) \ln \frac{y_i + a}{\bar{y} + a} - (y_i - \hat{\lambda}) \right] \\
&= 2 \sum_{i=1}^n \left[(y_i + a) \ln \frac{n(y_i + a)}{\sum_{i=1}^n y_i + na} - (y_i - \hat{\lambda}) \right].
\end{aligned}$$

Novamente, seu valor esperado é:

$$\begin{aligned}
E(D(\mathbf{Y} + a, \lambda + a)) &= E \left[2 \sum_{i=1}^n (Y_i + a) \ln \frac{n(Y_i + a)}{\sum_{i=1}^n Y_i + na} - (Y_i - \hat{\lambda}) \right] \\
&= 2E \left[\sum_{i=1}^n (Y_i + a) \ln \frac{n(Y_i + a)}{\sum_{i=1}^n Y_i + na} \right]
\end{aligned}$$

que não possui valor expícito, como já visto.

Seja, então,

$$\begin{aligned}
g(y_1 + a, y_2 + a, \dots, y_n + a) &= \sum_{i=1}^n (y_i + a) \ln \frac{n(y_i + a)}{\sum_{i=1}^n y_i + na} \\
&= \sum_{i=1}^n (y_i + a) \ln \frac{y_i + a}{\bar{y} + a} \\
&= \sum_{i=1}^n (y_i + a) \ln \frac{n(y_i + a)}{\sum_{i=1}^n y_i + na} \\
&= (y_1 + a) \ln \frac{n(y_1 + a)}{\sum_{i=1}^n y_i + na} + (y_2 + a) \ln \frac{n(y_2 + a)}{\sum_{i=1}^n y_i + na} + \dots \\
&+ (y_n + a) \ln \frac{n(y_n + a)}{\sum_{i=1}^n y_i + na}.
\end{aligned}$$

Expandindo novamente a função $g(\mathbf{y})$ em torno de $\boldsymbol{\lambda}$ utilizando a expressão (20), tem-se:

$$g(\mathbf{y}) = \sum_{i=1}^n (y_i + a) \ln \frac{y_i + a}{\bar{y} + a} \implies g(\boldsymbol{\lambda}) = 0.$$

Calculando a 1ª derivada para a função $g(\mathbf{y})$ em relação a y_1 , tem-se:

$$\frac{\partial g}{\partial y_1} = \ln \frac{ny_1 + na}{\sum_{i=1}^n y_i + na} + (y_1 + a) \frac{n(\sum_{i=1}^n y_i + na) - (ny_1 + na) \frac{\sum_{i=1}^n y_i + na}{ny_1 + na}}{(\sum_{i=1}^n y_i + na)^2}$$

$$\begin{aligned}
& - (y_2 + a) \frac{ny_2 + na}{(\sum_{i=1}^n y_i + na)^2} \frac{\sum_{i=1}^n y_i + na}{ny_2 + na} - \dots \\
& - (y_n + a) \frac{ny_n + na}{(\sum_{i=1}^n y_i + na)^2} \frac{\sum_{i=1}^n y_i + na}{ny_n + na} \\
& = \ln \frac{ny_1 + na}{\sum_{i=1}^n y_i + na}.
\end{aligned}$$

Substituindo \mathbf{y} por $\boldsymbol{\lambda}$, tem-se:

$$\left. \frac{\partial g}{\partial y_1} \right|_{\mathbf{y}=\boldsymbol{\lambda}} = \ln \frac{n\lambda + na}{n\lambda + na} = \ln 1 = 0$$

e como todas as outras derivadas de $g(\mathbf{y})$ fornecem o mesmo valor em $\mathbf{y} = \boldsymbol{\lambda}$, segue que:

$$\sum_{i=1}^n \left. \frac{\partial g}{\partial y_i} \right|_{y_i=\lambda} E(Y_i - \lambda) = 0.$$

Calculando a 2ª, 3ª e 4ª derivadas em relação a y_1 , tem-se:

$$\frac{\partial^2 g}{\partial y_1^2} = \frac{n}{ny_1 + na} - \frac{1}{\sum_{i=1}^n y_i + na} \quad \Rightarrow \quad \left. \frac{\partial^2 g}{\partial y_1^2} \right|_{\mathbf{y}=\boldsymbol{\lambda}} = \frac{n}{n\lambda + na} - \frac{1}{n\lambda + na} = \frac{n-1}{n(\lambda+a)}$$

$$\begin{aligned}
\frac{\partial^3 g}{\partial y_1^3} &= \frac{-1}{(y_1 + a)^2} + \frac{1}{(\sum_{i=1}^n y_i + na)^2} \quad \Rightarrow \\
\Rightarrow \left. \frac{\partial^3 g}{\partial y_1^3} \right|_{y_i=\lambda} &= \frac{-1}{(\lambda + a)^2} + \frac{1}{(n\lambda + na)^2} = \frac{1 - n^2}{n^2(\lambda + a)^2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^4 g}{\partial y_1^4} &= \frac{2}{(y_1 + a)^3} - \frac{2}{(\sum_{i=1}^n y_i + na)^3} \quad \Rightarrow \\
\Rightarrow \left. \frac{\partial^4 g}{\partial y_1^4} \right|_{y_i=\lambda} &= \frac{2}{(\lambda + a)^3} - \frac{2}{(n\lambda + na)^3} = \frac{2(n^3 - 1)}{n^3(\lambda + a)^3}.
\end{aligned}$$

Portanto, com os mesmos argumentos utilizados anteriormente, segue que:

$$\frac{1}{2!} \sum_{i=1}^n \left. \frac{\partial^2 g}{\partial y_i^2} \right|_{y_i=\lambda} E(Y_i - \lambda)^2 = \frac{1}{2!} \frac{n(n-1)\lambda}{n(\lambda+a)} = \frac{\lambda(n-1)}{2(\lambda+a)}$$

$$\frac{1}{3!} \sum_{i=1}^n \frac{\partial^3 g}{\partial y_i^3} \Big|_{y_i=\lambda} E(Y_i - \lambda)^3 = \frac{1}{3!} \frac{n(1-n^2)\lambda}{n^2(\lambda+a)^2} = \frac{1-n^2}{6n(\lambda+a)^2} \lambda$$

$$\frac{1}{4!} \sum_{i=1}^n \frac{\partial^4 g}{\partial y_i^4} \Big|_{y_i=\lambda} E(Y_i - \lambda)^4 = \frac{1}{4!} \frac{2n(n^3-1)\lambda(3\lambda+1)}{n^3(\lambda+a)^3} = \frac{1}{12} \frac{(n^3-1)(3\lambda^2+\lambda)}{n^2(\lambda+a)^3}.$$

Calculando a derivada mista de 4ª ordem com relação a y_1 e y_2 :

$$\frac{\partial^4 g}{\partial y_1^2 \partial y_2^2} = \frac{-2}{(n\lambda + na)^3} \implies \frac{\partial^4 g}{\partial y_1^2 \partial y_2^2} \Big|_{\mathbf{y}=\boldsymbol{\lambda}} = \frac{-2}{n^3(\lambda+a)^3}.$$

Substituindo \mathbf{y} por $\boldsymbol{\lambda}$, tem-se:

$$\frac{1}{2!2!} \sum_{i<j=2}^n \frac{\partial^4 g}{\partial y_i^2 \partial y_j^2} \Big|_{\mathbf{y}=\boldsymbol{\lambda}} E(Y_i - \lambda)^2 E(Y_j - \lambda)^2 = \frac{1}{4!} \frac{n(n-1)}{2} \frac{-2\lambda^2}{n^3(\lambda+1)^3} = \frac{-(n-1)\lambda^2}{4n^2(\lambda+a)^3}.$$

Assim, a expansão em série de Taylor em torno do ponto $\boldsymbol{\lambda}$ do valor esperado da *deviance* para as variáveis corrigidas por a é:

$$\begin{aligned} E(D(\mathbf{Y} + a, \boldsymbol{\lambda} + a)) &= 2 \left[0 + 0 + \frac{\lambda(n-1)}{2(\lambda+a)} + \frac{1-n^2}{6n(\lambda+a)^2} \lambda + \frac{1}{12} \frac{(n^3-1)(3\lambda^2+\lambda)}{n^2(\lambda+a)^3} \right. \\ &\quad \left. - \frac{(n-1)\lambda^2}{4n^2(\lambda+a)^3} + \dots \right] \\ &= 2 \left[\frac{n-1}{2} \frac{\lambda}{\lambda+a} + \frac{1-n^2}{6n} \frac{\lambda}{(\lambda+a)^2} + \right. \\ &\quad \left. + \frac{(n^3-1)}{12n^2} \frac{3\lambda^3+\lambda}{(\lambda+a)^3} - \frac{(n-1)\lambda^2}{4n^2(\lambda+a)^3} + \dots \right]. \end{aligned} \quad (21)$$

Desenvolvendo os termos da expressão (21) em torno de $a = 0$ e fazendo as simplificações necessárias, tem-se que:

$$\begin{aligned} E(D(\mathbf{Y} + a, \boldsymbol{\lambda} + a)) &= E \left[2 \sum_{i=1}^n (Y_i + a) \ln \frac{n(Y_i + a)}{\sum_{i=1}^n Y_i + na} \right] \\ &= (n-1) - \frac{(n-1)a}{\lambda} + \frac{1-n^2}{6n\lambda} + \frac{n^3-1}{4n^2\lambda} - \frac{(n-1)}{4n^2\lambda} + O(\lambda^{-2}) \\ &= (n-1) + \frac{1}{\lambda} \left[-(n-1)a + \frac{1-n^2}{3n} + \frac{n^3-1}{2n^2} - \frac{n-1}{2n^2} \right] + O(\lambda^{-2}). \end{aligned}$$

Igualando a zero o termo envolvendo $O(\lambda^{-1})$, tem-se:

$$-(n-1)a + \frac{1-n^2}{3n} + \frac{n^3-1}{2n^2} - \frac{n-1}{2n^2} = 0$$

resultando em

$$a = \frac{1}{6} + \frac{1}{6n}.$$

Nota-se, novamente que, quando $n \rightarrow +\infty$, o valor de $a \rightarrow \frac{1}{6}$. Assim, tem-se que esse é o valor que corrige a distribuição da *deviance* reduzindo o viés na convergência para a distribuição χ^2 , verificada através o valor esperado. Este valor para o fator de correção foi o obtido para o caso de apenas uma variável aleatória, como foi deduzido anteriormente.

4 RESULTADO E DISCUSSÃO

Conforme visto anteriormente, se $Y \sim \text{Poisson}(\lambda)$, o valor esperado da *deviance* é dado por

$$E(D(Y, \lambda)) = 2 \sum_{y=1}^{+\infty} \left(y \ln \frac{y}{\lambda} - (y - \lambda) \right) \frac{e^{-\lambda} \lambda^y}{y!}.$$

A Figura 1 mostra o comportamento desse valor para λ variando entre 0 e 15, tomados de 0,1 em 0,1. Para a construção do gráfico, foi utilizado o programa S-Plus. Os valores de probabilidade para cada valor da variável aleatória foram obtidos pela função $dpois(y, \lambda)$ e a variação de y foi considerada de 1 a 1000, já que não se pôde fazer $y \rightarrow +\infty$.

Com a adição de uma constante a às variáveis Y , em que novamente foi suposto que $Y + a \sim \text{Poisson}(\lambda)$, encontrou-se $a = 1/6$, que zera os termos envolvendo λ^{-1} na expansão em série de Taylor do valor esperado da *deviance*. A Figura 2 mostra o comportamento desse valor, comparando-o sem a correção, para λ variando novamente entre 0 e 15, tomados de 0,1 em 0,1, utilizando o mesmo programa.

Analisando as Figuras 1 e 2, nota-se que com a adição do valor $\frac{1}{6}$ à variável resposta, o valor esperado da *deviance* quando $Y \sim \text{Poisson}(\lambda)$ tende rapidamente a 1, que é o valor esperado de uma variável aleatória com distribuição χ^2 com 1 grau de liberdade. Isso ocorre mesmo para valores pequenos de λ . Desse modo, pode-se dizer que o fator melhora a aproximação da distribuição da *deviance* a uma distribuição χ^2_1 para qualquer valor de λ .

Para confirmar a aproximação da distribuição da *deviance* à uma dis-

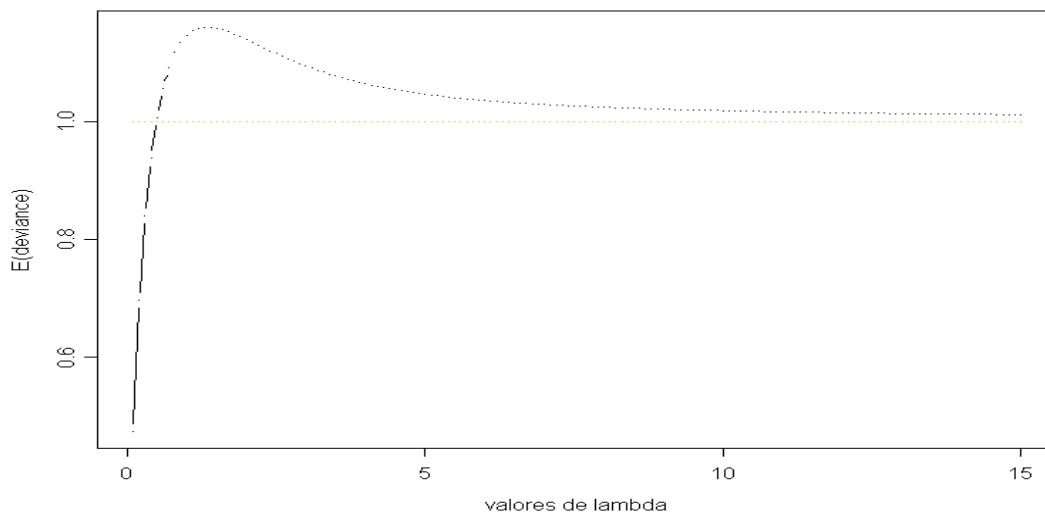


Figura 1 - Gráfico do valor esperado da *deviance* para $Y \sim P(\lambda)$, $\lambda = 0(0,1)15$.

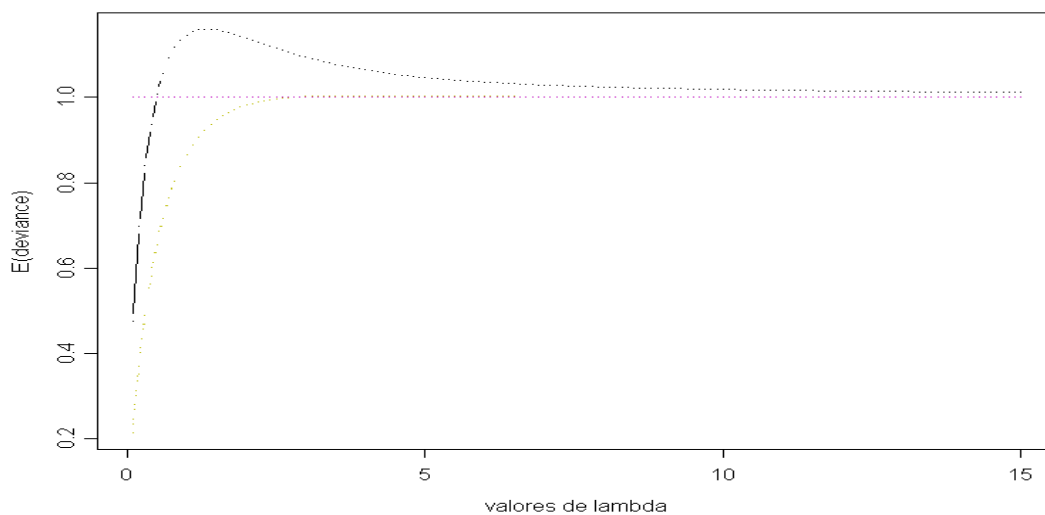


Figura 2 - Gráfico do valor esperado da *deviance* para uma variável aleatória $Y \sim P(\lambda)$ para $\lambda = 0(0,1)15$ com e sem correção (— = sem correção, = com correção).

tribuição χ^2 , as Figuras 3 e 4 mostram os *QQ-plots* para dados simulados de uma distribuição de Poisson com parâmetro $\lambda = 2$ e tamanhos de amostra 10 e 20, respectivamente. Foram gerados 1000 valores da distribuição de Poisson(2) e calculada a *deviance* para cada replicação. Os valores obtidos foram comparados com o quantis de uma distribuição $\chi^2_{(9)}$. Para a simulação, também foi usado o programa S-Plus.

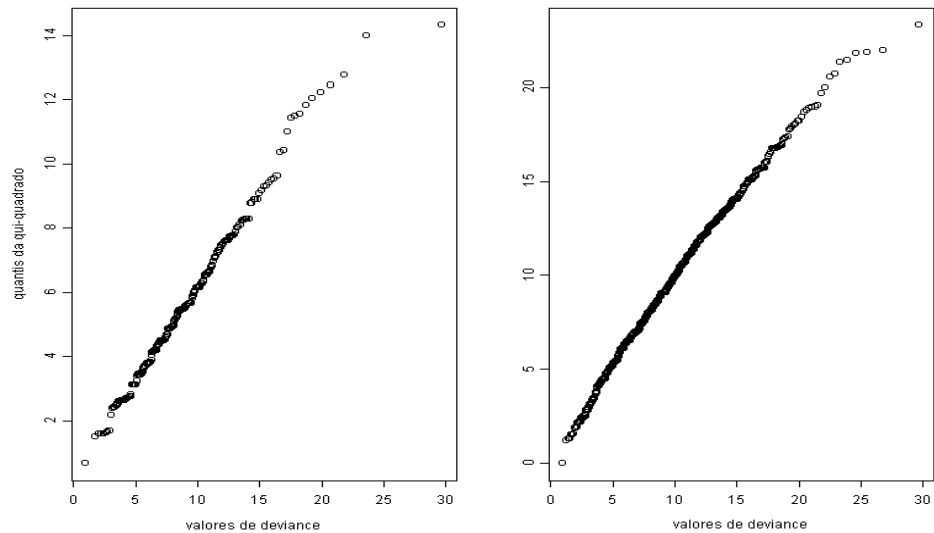


Figura 3 - *QQ-plot* para a distribuição da *deviance* com e sem correção para uma variável aleatória Poisson(2), para uma amostra de tamanho 10 e 1000 replicações.

Como se observa na Figura 3, o *QQ-plot* mostra que a *deviance* com correção apresenta quantis mais próximos ao de uma distribuição χ^2_9 do que a sem correção, para um tamanho de amostra 10, revelando que o fator de correção aproxima melhor a distribuição da *deviance* a uma distribuição χ^2 , considerando um tamanho de amostra relativamente pequeno. Na Figura 4, em que foi considerado

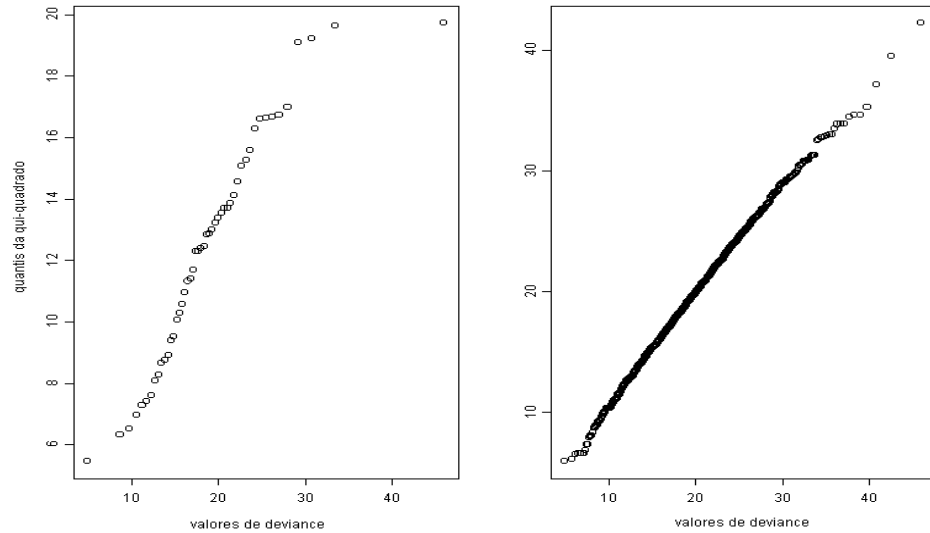


Figura 4 - *QQ-plot* para a distribuição da *deviance* com e sem correção para uma variável aleatória Poisson(2), para uma amostra de tamanho 20 e 1000 replicações.

um tamanho de amostra maior, a aproximação da distribuição da *deviance* a uma distribuição χ^2 sem o fator de correção ainda é ruim. Quando se consideram as variáveis corrigidas por $1/6$, nota-se uma melhora na aproximação da distribuição da *deviance*. Assim, mesmo quando o tamanho de amostra é maior mas o parâmetro da distribuição de Poisson ainda é pequeno, a correção mostra-se necessária para que a aproximação seja boa.

As Figuras de 5 a 10 mostram outras simulações considerando agora $\lambda = 5, 10$ e 15 para tamanhos de amostra 10 e 20 em cada caso, ainda com 1000 replicações.

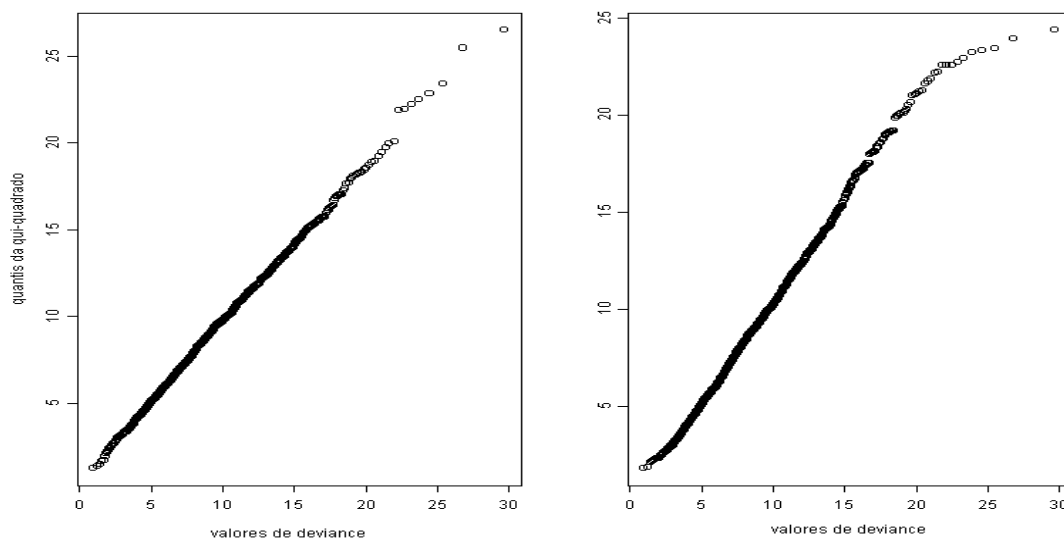


Figura 5 - *QQ-plot* para a distribuição da *deviance* com e sem correção para uma variável aleatória Poisson(5), para uma amostra de tamanho 10 e 1000 replicações.

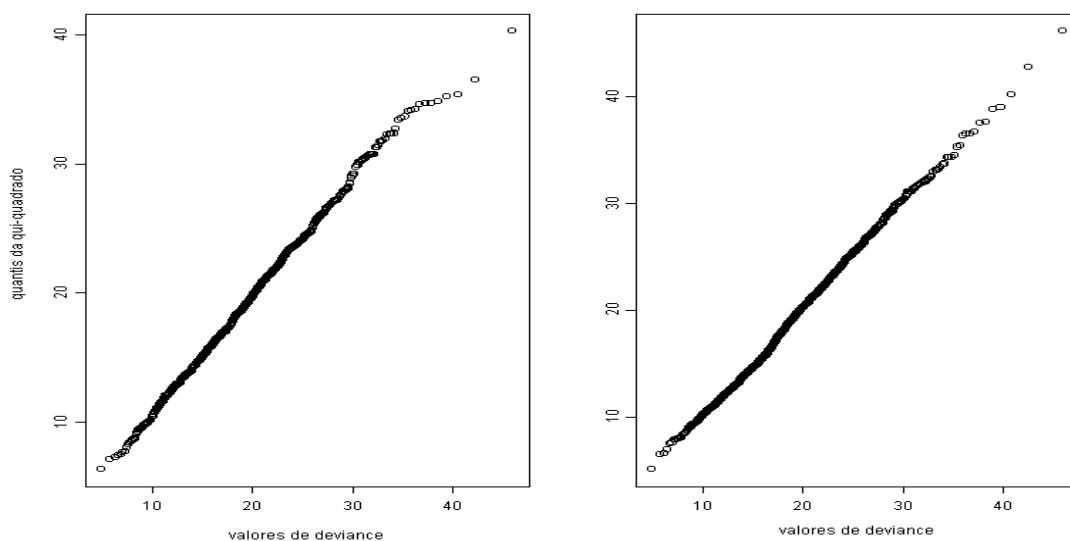


Figura 6 - *QQ-plot* para a distribuição da *deviance* com e sem correção para uma variável aleatória Poisson(5), para uma amostra de tamanho 20 e 1000 replicações.

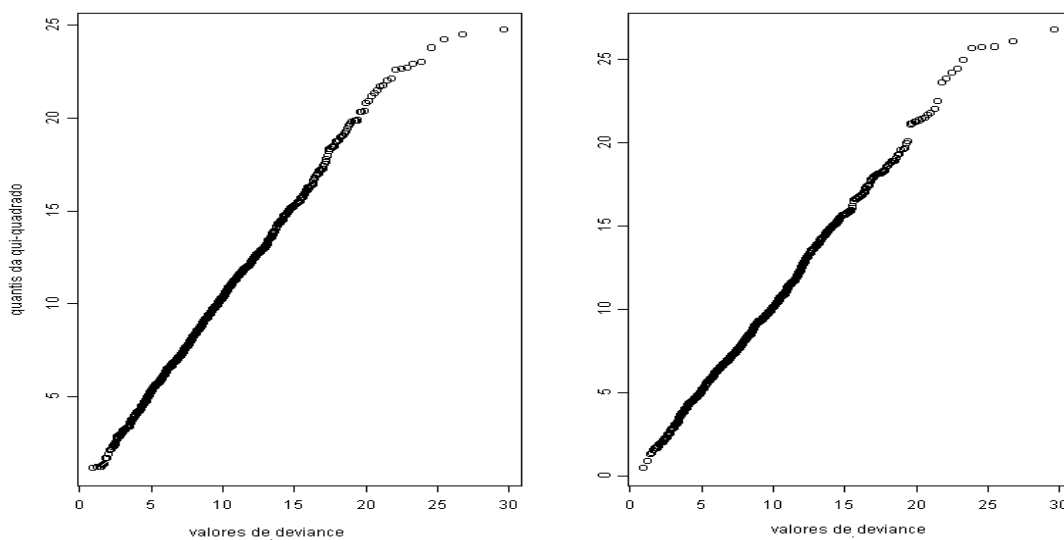


Figura 7 - *QQ-plot* para a distribuição da *deviance* com e sem correção para uma variável aleatória Poisson(10), para uma amostra de tamanho 10 e 1000 replicações.

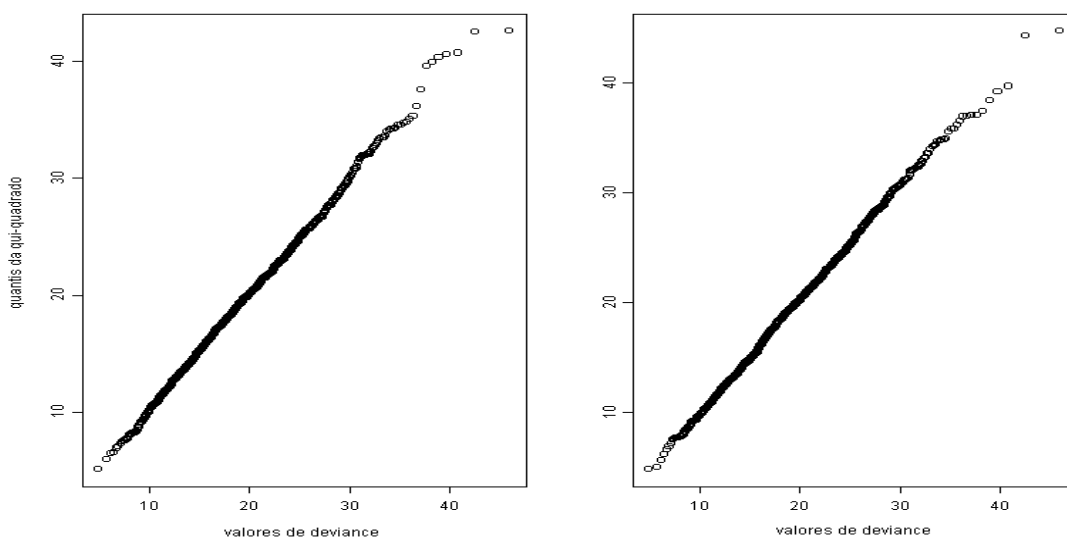


Figura 8 - *QQ-plot* para a distribuição da *deviance* com e sem correção para uma variável aleatória Poisson(10), para uma amostra de tamanho 20 e 1000 replicações.

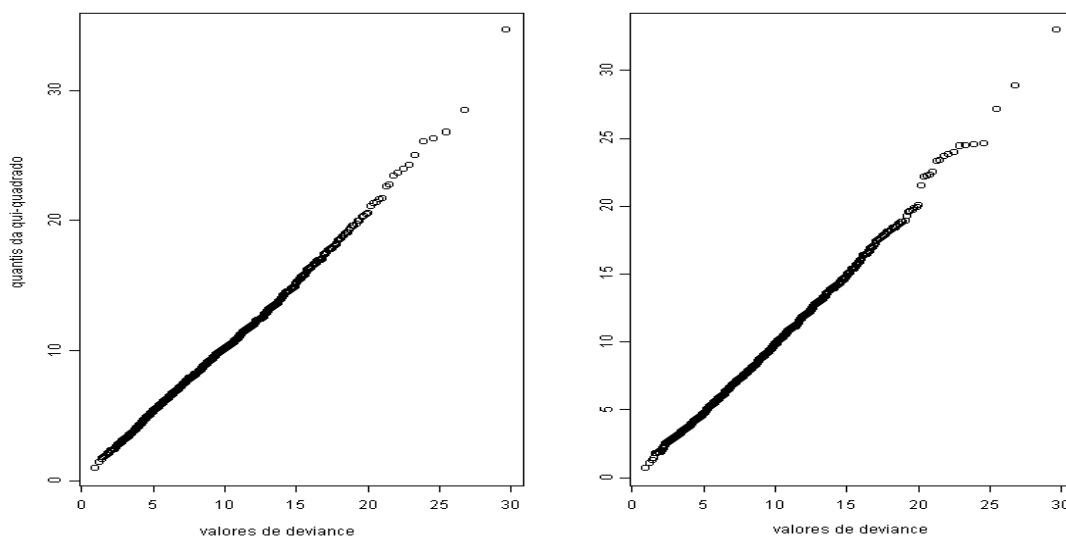


Figura 9 - *QQ-plot* para a distribuição da *deviance* com e sem correção para uma variável aleatória Poisson(15), para uma amostra de tamanho 10 e 1000 replicações.

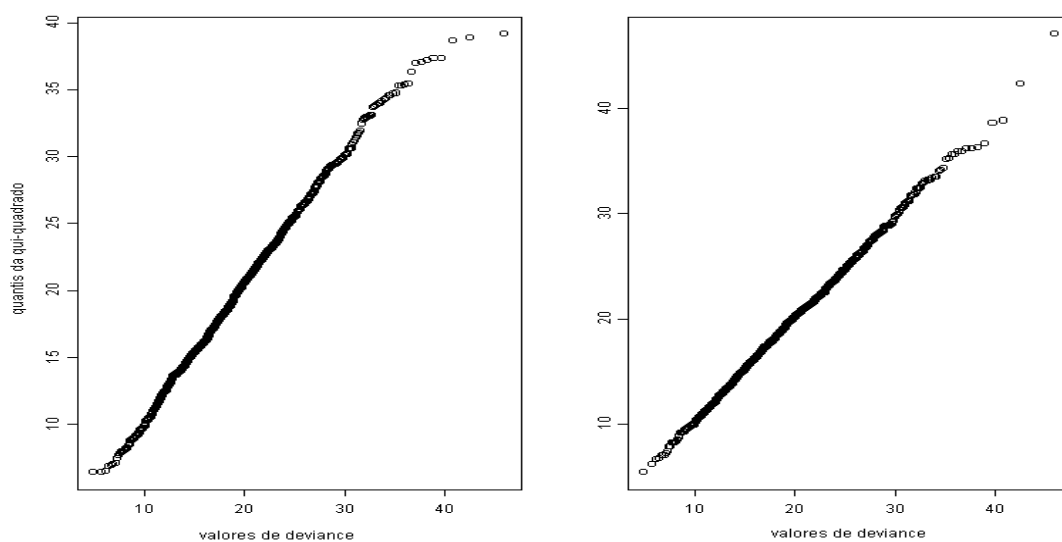


Figura 10 - *QQ-plot* para a distribuição da *deviance* com e sem correção para uma variável aleatória Poisson(15), para uma amostra de tamanho 20 e 1000 replicações.

Analisando as Figuras de 5 a 10, nota-se que quando o valor do parâmetro λ da distribuição de Poisson torna-se maior, o fator de correção já não é tão necessário. Isso porque os valores de *deviance* sem a correção e com a correção fornecem quantis da distribuição $\chi^2_{(9)}$ bastante próximos, indicando que a aproximação é boa considerando os tamanhos de amostra 10 ou 20. Assim, pode-se notar que os tamanhos de amostra não exerceram muita influência na aproximação da distribuição da *deviance* a uma distribuição χ^2 . Na realidade, como visto nas Figuras 1 e 2, para o valor do parâmetro λ aumentando, não existe problema com a aproximação quando se considera uma variável, refletindo também no caso de n variáveis independentes e identicamente distribuídas.

5 CONCLUSÃO

Diante do exposto e dos resultados obtidos neste trabalho, pode-se concluir que:

- A distribuição da *deviance* de uma variável aleatória seguindo uma distribuição de Poisson com parâmetro λ não possui uma boa aproximação à uma distribuição χ^2 para pequenos valores de λ .
- O valor $1/6$ obtido como o fator de correção a ser adicionado às variáveis resposta melhora a aproximação da distribuição da *deviance* à uma distribuição χ^2 , tanto para uma variável como para o modelo de média, considerando n variáveis independentes e identicamente distribuídas.
- O fator de correção se mostrou mais eficaz para o caso em que o valor do parâmetro λ é pequeno na distribuição de Poisson, não havendo necessidade de corrigir as variáveis para grandes valores de λ e grandes tamanhos de amostra.
- Estudos posteriores podem ser conduzidos a fim de generalizar o fator de correção obtido considerando que as variáveis aleatórias sejam agora independentes e cujas médias sigam um modelo linear ou um modelo linear generalizado.

REFERÊNCIAS BIBLIOGRÁFICAS

- BICKEL, P.J.; DOKSUM, K.A. **Mathematical statistics**. São Francisco: Holden Day, 1977. 493p.
- BLOM, G. Transformations of the binomial, negative binomial, Poisson and χ^2 distributions. **Biometrika**, v. 41, n. 3, p. 302-316, 1954.
- BOTTER, D.A.; CORDEIRO, G.M. Improved estimators for generalized linear models with dispersion covariates. **Journal of Statistical Computation and Simulation**, v. 62, p. 91-104, 1998.
- COLOSIMO, E.A. ; SILVA, A.F. ; CRUZ, F.R.B. Bias evaluation in the proportional hazards model. **Journal of Statistical Computation and Simulation**, v. 65, p. 191-201, 2000.
- CORDEIRO, G.M. **Modelos lineares generalizados**. Campinas: ABE, 1986. 286p.
- CORDEIRO, G.M.; CRIBARI-NETO, F. On bias reduction in exponential and non-exponential family regression models. **Communications in Statistics - Simulation**, v. 27, n. 2, p. 485-500, 1998.
- COX, D.R.; SNELL, E.J. A general definition of residuals. **Journal of the Royal Statistical Society**. Serie B, v. 30, n. 2, p. 248-275, 1968.
- COX, D.R.; SNELL, E.J. **Applied statistics**. London: Chapman and Hall, 1981. 189p.

- COURANT, R. **Cálculo diferencial e integral**. Rio de Janeiro: Globo, 1958. 532p.
- DEMÉTRIO, C.G.B. **Modelos lineares generalizados na experimentação agrônômica**. Porto Alegre: UFRGS ; RBRAS, 1993. 125p.
- GIMENES, A.P.G.S. Fator de correção para a distribuição da *deviance* para dados de proporções. Piracicaba, 2000. 63p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- JAMES, B.R **Probabilidade**: um curso em nível intermediário. Rio de Janeiro: IMPA, Projeto Euclides, 1981 304p.
- JOHNSON, N.L.; KOTZ, S. **Discrete distributions**. Boston: Houghton Mifflin Company, 1969. 328p.
- KAPLAN, W. **Advanced calculus**. Reading: Addison Wesley Pub. , 1973. 709p.
- LANG, S. **Cálculo**. Rio de Janeiro: Livros Técnicos e Científicos, 1976. 366p.
- LEITE, J.G.; SINGER, J.M. **Métodos assintóticos em estatística**: fundamentos e aplicações. São Paulo: ABE, 1990. 130p.
- MCCULLAGH, P.; NELDER, J.A. **Generalized linear Models**. London: Chapman and Hall, 1989. 511p.
- MEYER, P.L. **Probabilidade**: aplicações à estatística. Rio de Janeiro: Livros Técnicos e Científicos, 1984. 426p.
- MURTEIRA, B.J.F. **Probabilidades e estatística**. Lisboa: McGraw-Hill, 1990. 423p.
- NELDER. J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society**. Serie A, v.135, n. 3, p. 370-384, 1972.

- ROUSSAS, G.G. **A first course in mathematical statistics**. Reading: Addison-Wesley Pub. , 1973. 506p.
- STATISTICAL SCIENCE **S-Plus for Windows**: user's manual. Seattle: StatiSci, 1993. 2v.
- TAYLOR, J.M.G.; SIQUEIRA, A.L.; WEISS, R.E. The cost of adding parameters to a model. **Journal of the Royal Statistical Society**. Series B, v.58, n. 3, p. 593-609, 1996.
- WEDDERBURN, R.W.M. Quasilikelihood functions, generalized linear models and the Gauss-Newton method. **Biometrika**, v.61, n. 3, p. 439-447, 1974.