

<p>UFSC CTC PPGEP</p>	<p>Universidade Federal de Santa Catarina Programa de Pós-Graduação em Engenharia de Produção</p>
<p>Tese</p>	<p>Rosangela Getirana Santana</p> <p>Matriz de Covariância Corrigida para os Modelos Não-Lineares da Família Exponencial</p>
<p>R O S A N G E L A</p> <p>G E T I R A N A</p> <p>S A N T A N A</p>	
<p>2005</p>	<p>Florianópolis Santa Catarina 19/12/2005</p>

Matriz de Covariância Corrigida para os Modelos Não-Lineares da Família Exponencial

por

Rosangela Getirana Santana

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Produção da
Universidade Federal de Santa Catarina como parte dos requisitos necessários para a obtenção
do grau de Doutor em Engenharia de Produção.

Área de Concentração: Gestão de Qualidade e Produtividade

Orientador: Paulo José Ogliari, *Dr.*

Banca Examinadora:

Paulo Ogliari, Dr.
(Orientador)

Gauss Cordeiro, Ph.D
(Co-orientador)

Robert Samohyl, Ph.D

Adriano F. Borgatto , Dr

Pedro Barbetta, Dr

Lucia Barroso, Dra.

Denise Botter, Dra.

As Minhas irmãs

Agradecimentos

Ao Prof. Dr. Gauss Cordeiro, pelo exemplo de profissionalismo, pela sua orientação e pelas palavras de incentivo foi uma honra ter sido sua aluna.

Ao Prof. Paulo Ogliari, um dos principais motivadores deste trabalho, pois foi através do trabalho final de sua disciplina que tomei a decisão de estudar modelos não-lineares da família exponencial.

A Margareth, minha amiga há mais de 20 anos, com quem posso contar e que esteve presente me apoiando durante todas as fases deste trabalho.

A Isolde, uma amiga com quem aprendi a compartilhar e a desenvolver trabalhos em equipe, obrigada, por tudo, mas em especial, pela sua sensibilidade de criticar no momento certo.

Amigas, Isolde e Margareth, estou certa que sem vocês não teria concluído este trabalho.

A minha filha Sara que é tudo para mim e que deu um sentido novo a minha vida, por você e para você todos os dias tento ser uma pessoa melhor. Eu te amo filha.

Aos meus alunos que, sem dúvida alguma, foram os maiores motivadores para começar este doutorado;

Às amigas e ex-alunas Clédina, Valentina, que sempre tiveram palavras de incentivo, obrigada meninas.

Aos colegas de departamento, Josmar, pelas dicas no SAS, Darlei, pela força e compreensão, Doherty, pelas dicas no Maple.

Aos professores, Roberty Samohyl, Lucia Barroso, Dalton F.de Andrada pelas correções no exame de qualificação.

Resumo

Os modelos não-lineares da família exponencial é uma extensão dos modelos lineares generalizados, permitindo que o preditor da média seja não-linear. Esses modelos, por serem menos restritivos, têm sido utilizados para modelar sistemas produtivos como mais uma ferramenta na tomada de decisão. Usualmente, os parâmetros desses modelos são estimados pelo método de máxima verossimilhança, que têm propriedades assintóticas de $O(n^{-1})$, onde n é o tamanho da amostra. Portanto, para tamanhos de amostras pequenos, pode haver erros consideráveis, nas inferências. Essa Tese tem como objetivo obter uma expressão analítica para a matriz de covariância de segunda ordem do estimador de máxima verossimilhança para os parâmetros dos modelos não-lineares da família exponencial que contribuirá no procedimento de inferência da verossimilhança, quando o tamanho da amostra é pequeno. Esse estimador, que nada mais é do que uma correção do que vem sendo utilizado, tem propriedades assintóticas de $O(n^{-2})$. A metodologia adotada consistiu em obter os cumulantes desses modelos e substituí-los na função geratriz dos cumulantes, que, pela propriedade de invariância sob permutação de índices nos modelos não-lineares da família exponencial, pode ser simplificada e expressa em termos de matrizes. A expressão obtida é de fácil implementação computacional, uma vez que consiste de operações com matrizes. O estimador de segunda ordem da matriz de covariância foi avaliado por um estudo de simulação que mostrou que esse é indispensável para amostras de tamanho pequeno a moderado. Para ilustrar o uso da técnica proposta, uma aplicação na avaliação da qualidade do papel cujo modelo que descreve a variável resposta grau de refino das fibras é log-linear e componente aleatória gama. Nessa aplicação evidenciou-se a necessidade dos estimadores de $O(n^{-2})$.

Abstract

The exponential family non-linear models are an extension of the generalized linear models, allowing the average predictor to be non-linear. These models, for being less restrictive, have been used to model productive system, being one more tool on the decision-making. Usually, the parameters of these models are estimated through the maximum likelihood method, which have asymptotic properties of n^{-1} order, where n is the sample size; thus, for small sample sizes, there might be considerable errors in the inferences. This work has the objective to obtain an analytic expression for the covariance matrix of second-order of the maximum likelihood estimator for the exponential family non-linear models, which will contribute to the likelihood inference proceeding when the sample size is small. This estimator, which is anything but a correction of what has been used, has asymptotic properties of $O(n^{-2})$. The adopted methodology consisted in developing the cumulative of these models and substituting them in the cumulative generative function that, through the invariance property under index permutation in the exponential family models, may be simplified and expressed in matrix terms. The obtained expression is easily computational-implemented, seen that it consists in matrix operations. The covariance matrix second-order estimator was evaluated through a simulation study, which showed that this is indispensable for samples from small to moderate size. In order to illustrate the use of the proposed technique, one application in paper quality, whose model that describes the answer variable fiber refining degree is log-linear, and gamma random component. Into these application, the necessity of $O(n^{-2})$ estimators was evidenced.

Sumário

1	Introdução	10
1.1	Formulação do problema	10
1.2	Objetivos	13
1.2.1	Objetivo geral	13
1.2.2	Objetivos específicos	13
1.3	Caráter inédito, contribuição científica e relevância	13
1.4	Estrutura da tese	14
2	Modelo linear generalizado	15
2.1	Descrição e definição do modelo	16
2.2	Estimação dos parâmetros do MLG	17
2.2.1	Estimador de máxima verossimilhança para β	18
2.2.2	<i>Deviance</i> ou função desvio e critério de informação de Akaike (AIC)	19
2.3	Distribuição assintótica de $\hat{\beta}$	21
2.4	Estimadores corrigidos	22
2.5	Matriz de covariância corrigida de $\hat{\beta}$	24
2.5.1	Expressão da correção para matriz de covariância em termos dos cumulantes	24
2.5.2	Obtendo $\sigma_{ij}^{(1)}$	24
2.5.3	Obtendo $\sigma_{ij}^{(2)}$	26
2.5.4	Obtenção de $\sigma_{ij}^{(3)}$	28
2.5.5	Matriz de covariância Σ em notação matricial	28
2.6	Aplicação	29
2.6.1	Modelo	30
2.6.2	Algoritmo	30

2.6.3	1° Passo: Obter as estimativas de $O(n^{-1})$	30
2.6.4	2° Passo: Resultados da <i>procedure</i> GENMOD	30
2.6.5	3° Passo: <i>procedure</i> IML	30
2.6.6	Resultados	31
3	Modelos Não-Lineares da Família Exponencial	33
3.1	Modelo	34
3.2	Matriz de informação de Fisher	35
3.2.1	Matriz de informação de Fisher de β	35
3.2.2	Matriz de informação de Fisher de ϕ	36
3.3	EMV dos parâmetros do modelo	36
3.4	Viés de $O(n^{-1})$ dos EMV para parâmetros β, η e μ	39
4	Matriz de covariância de $O(n^{-2})$	43
4.1	Função geradora de cumulantes	43
4.2	Obtenção de $\sigma_{ij}^{(1)}$	46
4.2.1	Obtendo a parcela de $\sigma_{ij}^{(1)}$ devida a $L1$	46
4.2.2	Obtendo a parcela de $\sigma_{ij}^{(1)}$ devida a $NL1$	48
4.2.3	Obtendo a parcela de $\sigma_{ij}^{(1)}$ devida a $NL2$	49
4.2.4	Matriz $\Sigma^{(1)}$	50
4.3	Obtenção de $\sigma_{ij}^{(2)}$	50
4.3.1	Obtendo $\kappa_{rst} (\frac{3}{2}\kappa_{abc} + 5\kappa_{a,bc})$	50
4.3.2	Obtendo $3\kappa_{ab,c}\kappa_{r,st}$	51
4.3.3	Matriz $\Sigma^{(21)}$	52
4.3.4	Matriz $\Sigma^{(22)}$ e $\Sigma^{(23)}$	53
4.4	Obtenção de $\sigma_{ij}^{(3)}$	53
4.4.1	Obtendo $(\kappa_{a,bc} + \kappa_{abc})$ e $(2\kappa_{r,st} + \kappa_{rst})$	53
4.4.2	Obtendo $(\kappa_{a,bc} + \kappa_{abc})(2\kappa_{r,st} + \kappa_{rst})$	54
4.4.3	Obtendo $\sigma_{ij}^{(31)}, \sigma_{ij}^{(32)}, \sigma_{ij}^{(33)}$ e $\sigma_{ij}^{(34)}$	54
4.5	Matriz de covariância de $O(n^{-2})$ dos EMV nos MNLFE	55
4.6	Caso especial	55

5	Estudo de simulação	57
5.1	Modelo simulado	57
5.2	Algoritmo	58
5.2.1	Algoritmo para obtenção das estimativas de $O(n^{-1})$ e $O(n^{-2})$	58
5.2.2	1º Passo: Obter as estimativas de $O(n^{-1})$	58
5.2.3	2º Passo: Resultados da <i>procedure</i> NLMIXED	58
5.2.4	3º Passo: <i>procedure</i> IML	59
5.3	Resultados	61
6	Modelando grau de refino das fibras de celulose e avaliando os estimadores de $O(n^{-2})$ para os MNLFE	63
6.1	Componentes do MNLFE	64
6.2	Algoritmo para obtenção das estimativas de $O(n^{-1})$ e $O(n^{-2})$	65
6.3	Resultados	67
7	Considerações finais e recomendações	70
A	Família exponencial	73
A.1	Função de densidade	73
A.2	Função de verossimilhança	73
A.3	Função score e matriz de informação de Fisher	74
A.4	<i>Deviance</i> ou função desvio	77
B	Propriedades do EMV	79
C	Identidades de Bartlett	81
C.1	Viés de 1ª ordem do EMV	83
D	Função geratriz dos cumulantes	86
E	Derivadas da log-verossimilhança	88
E.1	Log-verossimilhança dos MNLFE	88
E.2	Derivada primeira da log-verossimilhança U_r	89
E.3	Derivada segunda da log-verossimilhança U_{rs}	89

E.4	Derivada terceira da log-verossimilhança U_{rst}	90
F	Cumulantes	93
F.1	Cumulante κ_{rs}	93
F.2	Cumulante κ_{rst}	93
F.3	Cumulante $\kappa_{bc}^{(ad)}$	94
F.4	Cumulante $\kappa_{bcd}^{(a)}$	94
F.5	Cumulante $\kappa_{ac,bd}$	96
F.6	Cumulante $\kappa_{a,bc}$	97

Lista de siglas

MLG	Modelo Linear Generalizado
MNLFE	Modelo Não-Linear da Família Exponencial
EMV	Estimador de Máxima Verossimilhança
$O(n^{-1})$	Primeira Ordem
$O(n^{-2})$	Segunda Ordem

Lista de siglas

β	Parâmetros do modelo
$\hat{\beta}$	Estimador de máxima verossimilhança
$\tilde{\beta}$	Estimador corrigido
μ	Parâmetro média
ϕ	Parâmetro dispersão fi
η	Preditor

Capítulo 1

Introdução

1.1 Formulação do problema

Os métodos estatísticos, cada vez mais, vêm sendo reconhecidos para modelar sistemas de produção sendo de grande valia na tomada de decisões, uma vez que estes permitem especificar, prever e avaliar resultados obtidos do sistema produtivo.

Tendo a consciência de que vivemos em um mercado altamente competitivo e globalizado, o uso da modelagem estatística, que consiste em relacionar um vetor resposta, \mathbf{Y} , em função de covariáveis X_1, X_2, \dots, X_p , possibilita significativo diferencial para o crescimento das organizações no cenário atual. Considera-se que a utilização da modelagem estatística às áreas produtivas certamente trarão luz às dificuldades, tornando-as fontes de oportunidades.

Muitos são os trabalhos publicados, nas diversas áreas, que utilizam os modelos lineares e não-lineares normal, mas estes são restritos a variáveis resposta com distribuição normal, não correlacionadas e variância constante. Em muitos destes problemas, a hipótese de homocedasticidade não se verifica. Recorre-se, então, às transformações, que em alguns casos pode não estabilizar a variância do processo, e além disso, dificulta a interpretação do modelo. Portanto a necessidade de modelos menos restritos.

Os MLG (Modelos Lineares Generalizados) e os MNLFE (Modelos Não-Lineares da Família Exponencial) generalizam os modelos tradicionais de regressão normal linear e não-linear, abrindo o leque de opções para a distribuição da variável resposta e permitindo maior flexibilidade para ligação entre a média e a componente sistemática, que para os modelos lineares generalizados é linear, enquanto que nos modelos não-lineares da família exponencial é não-linear, isto é, os MLG e MNLE são caracterizados por duas componentes descritas a seguir:

- i) componente aleatória, do vetor resposta \mathbf{Y} com elementos Y_1, \dots, Y_n pertencente à família exponencial de distribuições;
- ii) a componente sistemática é descrita pelo preditor $\boldsymbol{\eta}$, que pode ser linear, caso dos MLG, ou

não-linear, caso dos MNLFE

Para a obtenção do estimador dos parâmetros destes modelos, várias metodologias de inferência têm sido propostas; uma das mais importantes é decorrente da teoria de verossimilhança que, para Cordeiro (1999, p.3), é fundamentada nos princípios:

- de suficiência, isto é, vetores de dados distintos com os mesmos valores das estatísticas suficientes para um vetor θ de parâmetros fornecem conclusões idênticas para θ ;
- fraco de verossimilhança que estabelece que vetores de dados com verossimilhanças proporcionais produzem as mesmas conclusões sobre θ ;
- forte de verossimilhança que garante, para variáveis aleatórias, X e Y , distintas, que dependem de um mesmo parâmetro e de um mesmo espaço paramétrico; supondo que dois modelos são adequados aos vetores de dados \mathbf{x} e \mathbf{y} em questão e que fornecem verossimilhanças proporcionais, então as conclusões sobre θ tiradas destes dois vetores de dados são idênticas.

A metodologia baseada na verossimilhança será empregada para ajuste e análise dos MLG e dos MNLFE.

Uma das etapas do processo para ajuste e análise destes modelos é a de inferência, onde o analista deve verificar a precisão e a interdependência das estimativas, construir regiões de confiança e testes sobre os parâmetros de interesse, analisar estatisticamente os resíduos e realizar previsões, podendo, assim, avaliar o modelo ajustado. Nesta etapa, é necessário conhecer a distribuição de probabilidade do estimador, que, em geral, é obtida por meio de resultados assintóticos, que são necessários em dois casos distintos:

- 1º quando não se tem solução exata para o problema estatístico ou ela é complicada. Então, poderá ser de interesse obter uma aproximação simples em grandes amostras;
- 2º o mais freqüente revela o papel central da teoria assintótica da verossimilhança na inferência estatística e ocorre quando não se tem solução exata, sendo, assim, inevitável a obtenção da solução aproximada.

Os resultados da teoria assintótica de primeira ordem são válidos quando $n \rightarrow \infty$, onde n é o tamanho da amostra, e são decorrentes de técnicas de linearização local baseadas nas expansões da série de Taylor e nos teoremas centrais do limite. Estes, por sua vez, são à base da construção de testes e intervalo de confiança em grandes amostras e amplamente utilizados em trabalhos aplicados nas áreas de econometria, agronomia, farmacocinética, etc. O erro da aproximação assintótica para a distribuição de $\hat{\beta}$, estimador de máxima verossimilhança de β , é da ordem de n^{-1} , assim sendo os cálculos de probabilidades baseados na função de distribuição

de probabilidade assintótica apresentam erros da ordem n^{-1} que, para amostras pequenas, podem ser consideráveis. A importância dos resultados assintóticos de primeira ordem é que produzem simplificações consideráveis para problemas em grandes amostras, implicando em resultados simples e elegantes, e, em alguns casos, mostram-se bastante razoáveis, mesmo quando n , tamanho da amostra, não é grande. Devem ser avaliadas as seguintes questões:

- a aproximação de 1ª ordem é uma boa aproximação para a distribuição dos estimadores ou de alguma estatística de teste sob a hipótese nula?
- E se, é possível melhorá-la?

É sabido que para amostras finitas as aproximações de primeira ordem podem não funcionar bem, levando assim a distorções de tamanho. Quanto à questão se é possível melhorá-la, será abordada nesta Tese, uma vez que nela será obtida a matriz de covariância de $O(n^{-2})$ dos EMV dos MNLFE.

Condições para assegurar fraca consistência e normalidade assintótica dos EMV (Estimador de Máxima Verossimilhança) nos MLG são apresentadas por Fahrmeir e Kaufmam (1985). McCullagh e Nelder (1989) estabeleceram resultados assintóticos sem fornecer provas ou suposições exatas. Nos modelos lineares generalizados, os estimadores e as inferências sobre os β 's, como a razão de verossimilhança e a estatística escores têm como função densidade de probabilidade assintótica χ^2 , que usualmente não são satisfatórias em pequenas amostras.

Assim, a teoria assintótica de segunda ordem é um refinamento dos resultados da teoria assintótica de primeira ordem, onde se busca melhorar os estimadores e aproximações de estatística-teste. Há alguma literatura no assunto, com artigos publicados, teses e livros, dentre os quais destacam-se os artigos de Cordeiro (1983), que utiliza a correção de Bartellett para melhorar a estatística-teste da razão de verossimilhança nosMLG, Cordeiro e McCullagh (1991), que apresentam os estimadores de segunda ordem para os parâmetros de dispersão e média dosMLG, Cordeiro Ferrari e Cribari que melhoram a estatística escore nos MLG, Paula, (1993), que apresenta uma melhoria dos estimadores dos parâmetros dos modelos não lineares da família exponencial. Cordeiro e Botter (2000), que tratam dos estimadores de segunda ordem para os parâmetros dos modelos lineares generalizados superdispersados. Os livros de McCullagh e Nelder (1991) que é uma referência completa em MLG e Wei (1999), que, apresenta, alguns resultados de segunda ordem para os MNLFE, através geometria diferencial.

1.2 Objetivos

1.2.1 Objetivo geral

Avaliar os estimadores de $O(n^{-2})$ dos parâmetros dos modelos não-lineares da família exponencial.

1.2.2 Objetivos específicos

- Apresentar os estimadores de $O(n^{-1})$ dos parâmetros dos modelos não lineares da família exponencial;
- Apresentar os estimadores de $O(n^{-2})$ dos parâmetros dos modelos não lineares da família exponencial;
- Obter uma fórmula geral para a matriz de covariância de segunda ordem dos modelos não lineares da família exponencial;
- Apresentar a matriz de covariância de segunda ordem, de alguns modelos, tais como o não-linear de regressão normal e os lineares generalizados;
- Avaliar os estimadores obtidos por meio de um estudo de simulação;
- Avaliar a qualidade do papel, modelando a variável resposta grau de refino, proveniente de um experimento realizado na Klabin-Papeis;

1.3 Caráter inédito, contribuição científica e relevância

Esta Tese estende os resultados apresentados em Cordeiro (2004) e deverá contribuir na melhoria dos estimadores por intervalo e testes de hipóteses em MNLFE, uma vez que, para n , tamanho da amostra não muito grande a estrutura de covariância destes modelos é de $O(n^{-1})$.

O avanço tecnológico, no planejamento e execução de pesquisas, requer que técnicas cada vez mais refinadas de análises estatísticas sejam utilizadas. As posições críticas dos estatísticos são, em geral, desconsideradas pelos pesquisadores, das diferentes áreas, devido às dificuldades matemáticas e estatísticas para a implementação de técnicas refinadas de análise de dados experimentais. Por isto, é importante obter modelos menos restritivos e inferências mais precisas. Tradicionalmente, têm-se adotado os modelos de regressão normal, que são bastante restritivos, perdendo informações preciosas que, por vezes, não são detectadas, prejudicando assim as previsões. Além disso, a evolução acentuada da Ciência da Computação e, conseqüentemente, o desenvolvimento de sistemas sofisticados de processamento de dados, com algoritmos rápidos e precisos, tem contribuído muito no desenvolvimento de *softwares*, principalmente de natureza livre, o que possibilita ampla utilização desses recursos computacionais a custos praticamente

nulos. Baseando-se nesses fatos, uma Tese com a finalidade de apresentar uma metodologia de modelagem, na Engenharia de Produção, e indicar melhorias das inferências em modelos com menos restrições, é de relevância, permitindo assim, promover a divulgação da teoria assintótica de $O(n^{-2})$.

1.4 Estrutura da tese

Um resumo do que é apresentado, em cada um dos capítulos é relacionado a seguir:

Capítulo 2: apresenta os estimadores de $O(n^{-1})$ e de $O(n^{-2})$ para os MLG. Trata-se de uma revisão de literatura que vai dar suporte aos Capítulos 3 e 4;

Capítulo 3: apresenta os modelos não-lineares da família exponencial e os estimadores de $O(n^{-1})$ e o de $O(n^{-2})$, para os parâmetros destes modelos. Também trata-se de uma revisão de literatura;

Capítulo 4: é a contribuição teórica desta tese que apresenta a matriz de covariância de $O(n^{-2})$ dos EMV para os parâmetros do modelo não-linear da família exponencial;

Capítulo 5: é apresentado um estudo de simulação, para avaliar o desempenho dos estimadores de $O(n^{-2})$;

Capítulo 6: é apresentada uma aplicação, onde espera-se contribuir na divulgação destes modelos na área de Engenharia de Produção, mais especificamente na indústria de papel.

Capítulo 2

Modelo linear generalizado

Os modelos lineares de regressão normal e de análise de variância são amplamente utilizados em modelagem. Em alguns casos, mesmo quando não é razoável, a suposição de normalidade ou a homocedasticidade para a variável resposta, tentava-se algum tipo de transformação para estabilizar a variância e, conseqüentemente, a normalidade.

Para fazer inferência, as suposições de normalidade e variância constante são necessárias, e isto nem sempre é garantido ou razoável, em muitas aplicações, restringindo, assim, a adoção destes modelos.

Pelo fato dos MLG serem mais flexíveis, permitindo outras distribuições de probabilidade para a componente aleatória e assumindo que existe uma relação entre a variância e a média, estes se tornaram uma ferramenta poderosa na modelagem estatística.

Esta classe de modelos é devida à Nelder e Wedderburn (1972), os quais propuseram uma extensão do método dos escores, para estimar os parâmetros dos modelos da família exponencial. Estes modelos generalizam os modelos lineares tradicionais e possuem as seguintes características:

- a média da população μ depende de um preditor linear η , através de uma função g que é denominada de ligação, possivelmente não-linear;
- a distribuição de probabilidade da variável resposta (\mathbf{Y}) pertence à família exponencial.

Segundo McCullagh e Nelder (1989), algumas propriedades, como a linearidade e a existência de um método comum para obtenção das estimativas dos parâmetros, permitem a unificação dos modelos lineares de regressão dos de análise de variância e dos log-lineares em uma única classe denominada de MLG.

2.1 Descrição e definição do modelo

Considere y_l como a realização de uma variável aleatória Y_l . Assim, y_1, \dots, y_n denotará n independentes observações da vetor resposta \mathbf{Y} , com as seguintes suposições:

- i) A distribuição de cada Y_l pertence à família exponencial na sua forma canônica, algumas característica desta família é apresentada no Apêndice A, e sua função densidade de probabilidade é dada por:

$$\pi(y_l; \theta_l, \phi) = \exp [\phi \{y_l \theta_l - b(\theta_l)\} + c(y_l, \phi)], \quad (2.1)$$

- ii) Todos os Y_l têm a mesma distribuição, porém com médias diferentes $E(Y_l) = \mu_l$ para $l = 1, 2, \dots, n$,

Seja X uma matriz particionada em vetor coluna de variáveis independentes ou covariáveis, isto é, $X = [x_1 \dots x_p]$, cabe ressaltar que $x_1 \dots x_p$ são variáveis conhecidas e consideradas fixas. A componente sistemática é definida como o modelo linear tradicional, isto é:

$$\boldsymbol{\eta} = X\boldsymbol{\beta}$$

Seja g uma função monótona diferenciável que será denominada de função de ligação, a função g descreve como o valor esperado \mathbf{Y} está relacionado ao preditor linear $\boldsymbol{\eta}$, isto é:

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} = X\boldsymbol{\beta}$$

Na formulação dos MLG deve-se escolher a distribuição da probabilidade da variável resposta, as covariáveis (matriz modelo) e a função de ligação.

Observações:

- i) a matriz modelo X é suposta de posto completo;
- ii) como as observações são supostas independentes, ou pelo menos são não correlacionadas, conclui-se que, dados exibindo autocorrelações, como séries de tempo, devem ser cuidadosamente analisados;
- iii) como por hipótese a estrutura do erro é suposta única, outras técnicas estatísticas devem ser consideradas para analisar os dados onde os experimentos foram planejados para apresentarem mais de uma fonte de erro, como, por exemplo, as parcelas subdivididas;
- iv) se os dados têm distribuição não pertencente à família exponencial, como a distribuição de Cauchy e estruturas não-lineares, são excluídos dos MLG.

Função de Ligação Canônica

A função de ligação é dita ser canônica quando $\theta_l = \eta_l$. Uma das vantagens em adotá-las, no modelo é que há garantia da concavidade do logarítimo da função de verossimilhança e, conseqüentemente, muitos resultados assintóticos são obtidos mais facilmente, cuja demonstração é apresentado por Fahmeir e Kaufmann (1985). A ligação canônica g é tal que $\left(\frac{\partial b(\theta_l)}{\partial \theta_l}\right)^{-1} \circ g^{-1}(\theta_l)$ é igual a função identidade, sendo que o operador \circ , significa composição de funções, portanto tem-se que $g^{-1}(\cdot) = \frac{\partial b(\cdot)}{\partial \cdot}$.

2.2 Estimação dos parâmetros do MLG

Nos MLG, as observações são independentes, sendo assim, a função escore e a matriz de informação são obtidas como soma das contribuições individuais sobre os parâmetros.

Os parâmetros de interesse são β , ϕ , θ e μ , e como a componente aleatória do modelo pertence à família exponencial, demonstra-se que os parâmetros β e ϕ são ortogonais, e, portanto, podem ser estimados separadamente.

Log-verossimilhança O logarítimo da função de verossimilhança, simplificado por log-verossimilhança e denotado por $\ell(\beta)$, para os parâmetros dos MLG é dada por:

$$\ell(\beta) = \ell(\mu(\beta)) = \sum_{l=1}^n \phi \{y_l \theta_l - b(\theta_l)\} + c(y_l, \phi).$$

Função de escore de β A função escore é a derivada primeira da log-verossimilhança, e será obtida pela regra da cadeia:

$$U_\beta = U(\beta) = \sum_{l=1}^n \frac{\partial \ell}{\partial \beta_r} = \frac{\partial \ell}{\partial \theta_l} \frac{\partial \theta_l}{\partial \mu_l} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial \eta_l}{\partial \beta_r}, \quad (2.2)$$

mas, por outro lado:

$$\frac{\partial \ell}{\partial \theta_l} = \phi \left(y_l - \frac{db(\theta_l)}{d\theta_l} \right).$$

O valor esperado de Y_l é $\mu_l = \frac{db(\theta_l)}{d\theta_l}$, que derivando com relação a θ_l , obtém-se:

$$\frac{d\mu_l}{d\theta_l} = \frac{d}{d\theta_l} \left(\frac{db(\theta_l)}{d\theta_l} \right) = \frac{\partial^2 b(\theta_l)}{\partial \theta_l^2} = V,$$

sendo V denominada de função de variância que depende de μ .

O preditor linear do modelo linear generalizado $\eta = X\beta$, que derivando com respeito a β_r ,

obtém-se:

$$\frac{\partial \eta_l}{\partial \beta_r} = x_{lr}.$$

Substituindo estes valores em (2.2) obtém-se a r -ésima componente do vetor escore de β , que é dada por:

$$U(\beta_r) = \sum_{l=1}^n \phi(\mathbf{y}_l - \mu_l) \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} x_{lr}.$$

Considerando W uma matriz diagonal, cujos elementos são com $w_l = \frac{1}{V(\mu_l)} \left(\frac{d\mu_l}{d\eta_l} \right)^2$, a função escore para β na forma matricial é dada por:

$$U(\beta) = \phi X^T (\mathbf{y} - \boldsymbol{\mu}) V^{-1/2} W^{1/2}.$$

Matriz de informação de Fisher de β A matriz informação de Fisher de β que será denotada por $K(\beta)$, cujo elemento da linha r coluna s é:

$$K_{rs} = \sum_l -E \left(\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} \right) = E(-U_{rs}),$$

demonstra-se que $-U_{rs}$ para esses modelos é dada por:

$$-U_{rs} = \frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} = \sum_{l=1}^n \phi \left[-(y_l - \mu_l) \frac{\partial^2 \theta_l}{\partial \eta_l^2} + w_l \right] x_{lr} x_{ls}, \quad (2.3)$$

aplicando a esperança na equação (2.3) obtém-se o elemento K_{rs} da matriz de informação de Fisher para β_{rs} , que é dado por:

$$K_{rs} = \sum_{l=1}^n \phi w_l x_{lr} x_{ls},$$

que em notação matricial é:

$$K(\beta) = \phi (X^T W X).$$

2.2.1 Estimador de máxima verossimilhança para β

As equações de verossimilhança podem ser expressas por $U(\hat{\beta}) = 0$ sendo $\hat{\beta}$ o estimador de máxima verossimilhança.

Estas equações são geralmente não-lineares e neste caso as soluções de $U(\hat{\beta}) = 0$ devem ser obtidas por técnicas iterativas. No SAS estão disponibilizados dois métodos:

- O método iterativo de Newton-Raphson na versão multivariada, nas *procedures* NLP e NLMIXED;

- O método escore de Fisher, na *procedure* GENMOD que, em geral, é mais simples e coincide com o de Newton-Raphson quando a função de ligação é a canônica, pois substitui na expansão em série de Taylor a derivada segunda da log-verossimilhança pela sua esperança negativa, como será mostrado a seguir.

Algoritmo para obtenção das estimativas de β Para as distribuições pertencentes à família exponencial, as duas primeiras derivadas da log-verossimilhança existem. Dessa forma, será adotado um procedimento baseado na expansão de Taylor em torno de alguma estimativa inicial, $\beta^{(0)}$, expandindo $U(\hat{\beta})$ em série de Taylor até primeira ordem tem-se:

$$U(\hat{\beta}) \approx \frac{\partial \ell}{\partial \beta}(\beta^{(0)}) + \frac{\partial^2 \ell}{\partial \beta \partial \beta^T}(\beta^{(0)}) (\hat{\beta} - \beta^{(0)}) \approx 0. \quad (2.4)$$

A solução deste processo iterativo é o EMV para β . Salienta-se aqui que as *procedures* NLP e NLMIXED, fornecem $\hat{\beta}$ diretamente da equação (2.4).

O método escore de Fisher consiste na substituição da matriz de derivadas parciais de segunda ordem, pela matriz de valores esperados das derivadas parciais, isto é, a quantidade $\frac{\partial^2 \ell}{\partial \beta \partial \beta^T}$ é substituída por $K(\beta)$. Considere $\hat{\beta}^{(i)}$ o valor de $\hat{\beta}$, na iteração i e substituindo $\beta^{(0)}$ por $\hat{\beta}^{(i-1)}$ na equação (2.4) obtem-se:

$$\hat{\beta}^{(i+1)} = \hat{\beta}^{(i)} + K^{-1} \left(\hat{\beta}^{(i)} \right) U \left(\hat{\beta}^{(i)} \right),$$

substituindo os valores da matriz de informação e do vetor escore tem-se:

$$\hat{\beta}^{(i+1)} = \{ (X^T W X)^{-1} X^T W Z \}^i, \quad (2.5)$$

sendo:

$$z_l = \eta_l + (y_l - \mu_l) \frac{\partial \eta_l}{\partial \mu_l}.$$

A expressão (2.5) mostra que o procedimento iterativo pode ser considerado como estimador dos mínimos quadrados ponderados, conforme demonstraram Nelder e Wedderburn (1972). Em geral, a convergência é rápida, mas depende da escolha do valor inicial $\beta^{(0)}$.

2.2.2 *Deviance* ou função desvio e critério de informação de Akaike (AIC)

A estatística *deviance* é utilizada para avaliar a qualidade do ajuste de um MLG e também no EMV de ϕ . Para obtê-la, considere a seguinte notação para a log-verossimilhança:

$$\ell(\mu; y) = \sum_{l=1}^n \ell(\mu_l; y_l), \text{ onde } \mu_l = g^{-1}(\eta_l) \text{ e } \eta_l = X\beta.$$

Considere como modelo saturado aquele em que o número de parâmetros p é igual ao número de observações n , isto é, $p = n$. Para este modelo a $\ell(\mu; y)$ é estimada por:

$$\ell(y; y) = \sum_{l=1}^n \ell(y_l; y_l),$$

ou seja, a estimativa de máxima verossimilhança de μ_l fica no modelo saturado dada por $\hat{\mu}_l = y_l$. Quando $p < n$, a estimativa de $\ell(\mu; y)$ será denotada por $\ell(\hat{\mu}; y)$. Neste caso, as estimativas de máxima verossimilhança de $\hat{\mu}_l = g^{-1}(\hat{\eta}_l)$, sendo $\hat{\boldsymbol{\eta}} = X\hat{\boldsymbol{\beta}}$.

A *deviance*, que é uma distância entre a log-verossimilhança do modelo saturado e o modelo sob investigação avaliado na estimativa de máxima verossimilhança de β , é dada por:

$$\begin{aligned} D^*(y; \hat{\mu}) &= \phi D(y; \hat{\mu}) = 2 \{ \ell(y; y) - \ell(\hat{\mu}; y) \} \\ &= 2\phi \sum_{l=1}^n \{ v(y_l) - v(\hat{\mu}_l) + (\hat{\mu}_l - y_l) q(\hat{\mu}_l) \}, \end{aligned} \quad (2.6)$$

sendo $v(\boldsymbol{\mu}) = \boldsymbol{\mu}q(\boldsymbol{\mu}) - b[q(\boldsymbol{\mu})]$ e $q(\boldsymbol{\mu}) = \boldsymbol{\theta} = \int V^{-1}d\boldsymbol{\mu}$, definidos no Apêndice A.

Observação:

- Valor pequeno para a *deviance* indica que, para um número menor de parâmetros, obtém-se um ajuste tão bom quanto o ajuste para o modelo saturado.

Outra estatística para avaliar modelos com número de parâmetros diferentes é o AIC, que penaliza a *deviance* pelo número de parâmetros e é dada por:

$$AIC = -2 \log l - 2p$$

- Quanto menor o AIC melhor é o ajuste do modelo esse critério pode ser encontrado em Pinheiro e Bates (2000, p.84).

EMV para ϕ Para os modelos Poisson, binomial e exponencial, tem-se que $\phi = 1$. Para outros modelos pertencentes à família exponencial, tais como gama, normal e inversa gaussiana, o estimador de máxima verossimilhança para ϕ será obtido utilizando o procedimento apresentado por Cordeiro e McCullagh (1991). Para isto, considere $\hat{\phi}$ o EMV de ϕ . A função escore para ϕ , $U(\phi, \beta) = \frac{\partial \ell(\phi, \beta)}{\partial \phi}$ pode ser obtida em função da *deviance*, isto é,

$$U(\phi, \mu) \cdot D(\mathbf{y}, \hat{\mu}) = 2U(\hat{\phi}, \mathbf{y}).$$

Dessa forma a *deviance* é 2 vezes a função escore de ϕ no ponto $(\hat{\phi}, \mathbf{y})$. Por outro lado, a equação de máxima verossimilhança para $\hat{\phi}$, $U(\hat{\phi}) = 0$ é dada por:

$$\begin{aligned} d(\hat{\phi}) &= n^{-1} \sum_{l=1}^n [y_l q(\hat{\mu}_l) - b(q(\hat{\mu}_l))] + c(y_l) \text{ ou} \\ D(y, \hat{\mu}) &= 2 \sum_{l=1}^n \left[v(y_l) + c(y_l) + 2nd(\hat{\phi}) \right]. \end{aligned} \quad (2.7)$$

Assim, $\hat{\phi}$ é uma função da *deviance* do modelo.

Uma solução aproximada da estimativa do parâmetro de dispersão $\sigma^2 = \phi^{-1}$ é apresentada por Cordeiro e McCullagh (1991), que é dada por:

$$\hat{\phi} = \frac{n[1 + \{1 + 2D(y, \hat{\mu})/3n\}^{1/2}]}{2D(y, \hat{\mu})}$$

2.3 Distribuição assintótica de $\hat{\beta}$

Para esta classe de modelos, resultados assintóticos podem ser encontrados em McCullagh e Nelder (1983), Cordeiro (1983, 1987) e Fahrmeir e Kaufmann (1985). Em particular, Cordeiro (1986, pp.39, 40), demonstra que $\hat{\beta}$ é consistente para β , expandindo $U(\hat{\beta})$ em série de Taylor ao redor do parâmetro verdadeiro β , obtendo a seguinte expressão aproximada:

$$\hat{\beta} - \beta = (X^T W X)^{-1} X^T W Z + O(n^{-1});$$

cuja demonstração é apresentada no Apêndice C, portanto:

$$\begin{aligned} E(\hat{\beta}) &= \beta + O(n^{-1}) \quad \text{e} \\ Cov(\hat{\beta}) &= E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T\right] = (X^T \phi W X)^{-1} + O(n^{-1}), \end{aligned}$$

estas expressões refletem, que para pequenas amostras $\hat{\beta}$, é tendencioso e têm poucas propriedades que são satisfeitas para todos os tamanhos de amostras.

Com argumentos do teorema do limite central adaptados à função escore $U(\beta)$, e a lei fraca dos grandes números, aplicados à matriz de informação de Fisher $K_{\beta\beta}$ e, sob as condições de regularidades apresentadas no Apêndice B, demonstra-se que $\hat{\beta}$ é assintoticamente normal, com média β e matriz de covariância $(X^T \phi W X)^{-1}$, quando ϕ for desconhecido pode ser substituído por alguma estimativa consistente.

Regiões aproximadas de confiança e diversos testes são construídos através de $\hat{\beta}$ ou de estatísticas de $\hat{\beta}$, tais como a estatística da razão de máxima verossimilhança, a estatística

escore ou a estatística de Wald. Geralmente, a estatística da razão de máxima verossimilhança é mais utilizada, uma vez que a mesma é invariante com reparametrizações, o que não ocorre com as outras duas. Estas estatísticas são apresentadas por Demétrio (1993).

2.4 Estimadores corrigidos

Como mencionado, os estimadores de máxima verossimilhança são, geralmente, tendenciosos para amostras de tamanho pequeno ou moderado. É de interesse dimensionar o viés, possibilitando, assim, a melhoria nas inferências dos parâmetros. Cordeiro e McCullagh (1991) apresentam o viés e os estimadores de segunda ordem dos parâmetros β , η , μ e ϕ , dos MLG. Nesta seção, são apresentadas as expressões para o cálculo do viés e o leitor pode ter mais detalhes no Apêndice C.

Viés de $\hat{\beta}$ O viés do EMV do parâmetro β que será denotado por $B(\hat{\beta})$ é dado por :

$$\mathbf{B}(\hat{\beta}) = (X^T W X)^{-1} X^T W \boldsymbol{\xi}, \quad (2.8)$$

sendo:

- $F = \{f_{il}\}$ uma matriz diagonal, com elemento $f_l = V^{-1} \frac{d\mu_l}{d\eta_l} \frac{d^2\mu_l}{d\eta_l^2}$;
- $Z = \{z_{ij}\} = X (X^T W X)^{-1} X^T$, $Z_d = \text{diag}\{z_{11}, z_{22}, \dots, z_{nn}\}$;
- $\mathbf{1} = (1, 1, \dots, 1)^T$;
- $\boldsymbol{\xi} = -\frac{1}{2\phi} W Z_d F \mathbf{1}$.

Viés de $\hat{\eta}$ Como $\hat{\eta} = X\hat{\beta}$, o valor esperado de $\hat{\eta} - \eta$ é:

$$\begin{aligned} \mathbf{B}(\hat{\eta}) &= E(\hat{\eta} - \eta) = E(\hat{\eta}) - \eta = E(X\hat{\beta}) - X\beta = X\mathbf{B}(\hat{\beta}) \\ &= X (X^T W X)^{-1} X^T W \left(-\frac{1}{2\phi} W^{-1} Z_d F \mathbf{1} \right), \end{aligned}$$

resultando em

$$\mathbf{B}(\hat{\eta}) = -(2\phi)^{-1} Z Z_d F \mathbf{1}.$$

Observe que para o cálculo do viés de $\hat{\eta}$, é necessário somente a variância e a função de ligação com as suas primeiras e segundas derivadas.

Viés de $\hat{\boldsymbol{\mu}}$ Para o cálculo do viés para $\hat{\boldsymbol{\mu}}$, faz-se a expansão em série de Taylor da $g^{-1}(\hat{\boldsymbol{\eta}}) = \hat{\boldsymbol{\mu}}$, isto é,

$$g^{-1}(\hat{\boldsymbol{\eta}}) = g^{-1}(\boldsymbol{\eta}) + \frac{\partial g^{-1}(\boldsymbol{\eta})}{\partial \eta_r} (\hat{\eta}_r - \eta_r) + \frac{1}{2} \frac{\partial^2 g^{-1}(\boldsymbol{\eta})}{\partial \eta_r \partial \eta_s} (\hat{\eta}_r - \eta_r) (\hat{\eta}_s - \eta_s),$$

fazendo $g^{-1}(\hat{\boldsymbol{\eta}}) - g^{-1}(\boldsymbol{\eta})$ e, em seguida, aplicando esperança, tem-se:

$$B(\hat{\boldsymbol{\mu}}) = B(\hat{\boldsymbol{\eta}}) \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} + \frac{1}{2} V(\hat{\boldsymbol{\eta}}) \frac{\partial^2 \mu_l}{\partial \eta_i^2},$$

uma vez que $\boldsymbol{\mu}$ é função bijetora de $\boldsymbol{\eta}$ e considerando:

1. $V(\hat{\boldsymbol{\eta}})$ o termo de ordem n^{-1} na variância de $\hat{\boldsymbol{\eta}}$;
2. $D_1 = \text{diag} \left\{ \frac{\partial \mu_l}{\partial \eta_i} \right\}$;
3. $D_2 = \text{diag} \left\{ \frac{\partial^2 \mu_{ll}}{\partial \eta_{ii}^2} \right\}$.

O viés de $\hat{\boldsymbol{\mu}}$ em notação matricial fica:

$$B(\hat{\boldsymbol{\mu}}) = (2\phi)^{-1} (D_2 - D_1 Z F) Z_d \mathbf{1}. \quad (2.9)$$

Os estimadores corrigidos são obtidos a partir dos vieses, e serão denotados com um (\sim) acima do parâmetro.

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \hat{B}(\hat{\boldsymbol{\beta}}), \quad \tilde{\boldsymbol{\eta}} = \hat{\boldsymbol{\eta}} - \hat{B}(\hat{\boldsymbol{\eta}}) \quad \text{e} \quad \tilde{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}} - B(\hat{\boldsymbol{\mu}})$$

Viés de $\hat{\phi}$ O viés para o EMV do parâmetro de dispersão ϕ é obtido analogamente ao viés de $\hat{\boldsymbol{\beta}}$, utilizando a expressão creditada à Cordeiro e McCullagh (1991), que é válida para membros da família exponencial em que $c(y_l, \phi)$ da função densidade de probabilidade dada em (2.1), que pode ser decomposta em duas funções d_1 e d_2 , isto é,

$$c(\mathbf{y}, \phi) = d_1(\phi) + d_2(\mathbf{y}),$$

que neste caso é dado por:

$$B(\hat{\phi}) = \frac{\phi \frac{\partial^3 d_1(\phi)}{\partial \phi^3} - p \frac{\partial^2 d_1(\phi)}{\partial \phi^2}}{2n\phi \left(\frac{\partial^2 d_1(\phi)}{\partial \phi^2} \right)^2}. \quad (2.10)$$

Portanto, o estimador corrigido de ϕ é dado por:

$$\tilde{\phi} = \hat{\phi} - B(\hat{\phi}).$$

2.5 Matriz de covariância corrigida de $\hat{\beta}$

Para os EMV dos parâmetros do MLG, será apresentado o resultado de Cordeiro (2004) que obtém a matriz de covariância de segunda ordem.

Considerando $\hat{\beta}$, a estimativa de máxima verossimilhança dos parâmetros do MLG, será assumido que $\hat{\beta}$ converge para β , verdadeiro valor, quando $n \rightarrow \infty$ e que sua distribuição assintótica é $N\left(\beta, \phi^{-1} (X^T W X)^{-1}\right)$. Bowman e Shenton (1997, p.71) enumera o número de termos para o cálculo da matriz de covariância de segunda ordem do EMV, por exemplo, quando o número de parâmetros é igual a 2, têm-se 169 termos para cada variância e 193 termos para as covariância.

2.5.1 Expressão da correção para matriz de covariância em termos dos cumulantes

A expressão para obter a matriz de covariância de $O(n^{-2})$ está apresentada no Apêndice D, é devida a Peers e Iqbal (1985). O elemento da linha i coluna j , que é correção na covariância entre $\hat{\beta}_i$ e $\hat{\beta}_j$ é dado por:

$$\sigma_{ij} = \sigma_{ij}^{(1)} + \sigma_{ij}^{(2)} + \sigma_{ij}^{(3)}, \quad (2.11)$$

sendo:

$$\sigma_{ij}^{(1)} = -\kappa^{ia} \kappa^{jb} \kappa^{cd} (\kappa_{abcd} + \kappa_{a,bcd} + 2\kappa_{abc,d} + 2\kappa_{a,bc,d} + 3\kappa_{ac,bd}), \quad (2.12)$$

$$\sigma_{ij}^{(2)} = \kappa^{ia} \kappa^{jr} \kappa^{bs} \kappa^{ct} \left(\frac{3}{2} \kappa_{abc} \kappa_{rst} + 4\kappa_{ab,c} \kappa_{rst} + \kappa_{a,bc} \kappa_{rst} + 2\kappa_{ab,c} \kappa_{r,st} + \kappa_{ab,c} \kappa_{rt,s} \right), \quad (2.13)$$

e

$$\sigma_{ij}^{(3)} = \kappa^{ia} \kappa^{jb} \kappa^{rs} \kappa^{ct} (2\kappa_{a,bc} \kappa_{r,st} + \kappa_{a,bc} \kappa_{rst} + \kappa_{abc} \kappa_{rst} + 2\kappa_{abc} \kappa_{r,st}). \quad (2.14)$$

2.5.2 Obtendo $\sigma_{ij}^{(1)}$

Inicialmente, os cumulantes dentro do parêntese de (2.12) serão simplificados usando as identidades de Bartlett e a invariância dos cumulantes sob a permutação de índices. Para isto, considere as seguintes identidades:

$$\begin{aligned} \kappa_{a,bcd} &= -\kappa_{abcd} + \kappa_{bcd}^{(a)}, \\ \kappa_{a,bc,d} &= \kappa_{a,d,bc} = \kappa_{abcd} - \kappa_{abc}^{(d)} - \kappa_{bcd}^{(a)} + \kappa_{bc}^{(ad)} - \kappa_{ad,bc}, \\ \kappa_{abc,d} &= -\kappa_{abcd} + \kappa_{abc}^{(d)}. \end{aligned} \quad (2.15)$$

Usando estas identidades, os cumulantes dentro do parêntese de $\sigma_{ij}^{(1)}$ podem ser expressos

da forma seguinte:

$$\begin{aligned}
\kappa_{abcd} &= \kappa_{abcd}, \\
\kappa_{a,bcd} &= -\kappa_{abcd} + \kappa_{abc}^{(d)}, \\
2\kappa_{a,bc,d} &= 2\kappa_{abcd} - 2\kappa_{abc}^{(d)} - 2\kappa_{bcd}^{(a)} + 2\kappa_{bc}^{(ad)} - 2\kappa_{ad,bc}, \\
3\kappa_{ac,bd} &= 3\kappa_{ad,bc}, \\
2\kappa_{abc,d} &= 2\kappa_{abcd} + 2\kappa_{bcd}^{(a)}.
\end{aligned}$$

Somando as 5 equações obtêm-se o termo entre parênteses de $\sigma_{ij}^{(1)}$, que já simplificado fica:

$$\sigma_{ij}^{(1)} = \kappa^{ia} \kappa^{jb} \kappa^{cd} \left(2\kappa_{bc}^{(ad)} - \kappa_{bcd}^{(a)} + \kappa_{ac,bd} \right). \quad (2.16)$$

Os cumulantes dentro do parênteses de (2.16) são apresentados em Cordeiro (1984,1987) e Cordeiro e Barroso (2003). Destes artigos têm-se então que os cumulantes κ_{rs} , κ_{rst} , $\kappa_{bc}^{(ad)}$, $\kappa_{ab,cd}$ e $\kappa_{bcd}^{(a)}$ são:

- $\kappa_{rs} = -\phi \sum_{l=1}^n (w_l x_{lr} x_{ls});$
- $\kappa_{rst} = -\phi \sum_{l=1}^n \left\{ \frac{1}{V} \left(\frac{d^3 \mu_l}{d\eta_l^3} \right) + 2 \left[\frac{1}{V} \left(\frac{d^3 \mu_l}{d\eta_l^3} \right) \frac{V^{(1)}}{V^2} \left(\frac{d\mu_l}{d\eta_l} \right)^3 \right] \right\} x_{lr} x_{ls} x_{lt}.$

Sendo, como já apresentado anteriormente,

$$w_l = \left(\frac{d\mu_l}{d\eta_l} \right)^2 V^{-1}, \quad V^{(r)} = \frac{\partial^{(r)} \mu_l}{\partial \theta_l^{(r)}}, \quad x_{lr} = \frac{\partial \eta_l}{\partial \beta_r},$$

$V^{(r)}$ denota a derivada de ordem r da função de variância V com respeito à média.

- $\kappa_{bc}^{(ad)}$

$$\kappa_{bc}^{(ad)} = -\phi \sum_{l=1}^n \left\{ \frac{2}{V} \left[\left(\frac{d^2 \mu_l}{d\eta_l^2} \right) + \left(\frac{d\mu_l}{d\eta_l} \right) \left(\frac{d^3 \mu_l}{d\eta_l^3} \right) \right] - \frac{1}{V^2} \left[5 \left(\frac{d\mu_l}{d\eta_l} \right)^2 \left(\frac{d^2 \mu_l}{d\eta_l^2} \right) V^{(1)} + \left(\frac{d\mu_l}{d\eta_l} \right)^4 V^{(2)} \right] + \frac{2}{V^3} \left(\frac{d\mu_l}{d\eta_l} \right)^4 (V^{(1)})^2 \right\} x_{lr} x_{ls} x_{lt} x_{lv} \quad (2.17)$$

- $\kappa_{bcd}^{(a)}$

$$\kappa_{bcd}^{(a)} = \phi \sum_{l=1}^n \left\{ \frac{2}{V^3} (V^{(1)})^2 \left(\frac{d\mu_l}{d\eta_l} \right)^4 - \frac{V^{(2)}}{V^2} \left(\frac{d\mu_l}{d\eta_l} \right)^4 - \frac{3V^{(1)}}{V^2} \left(\frac{d\mu_l}{d\eta_l} \right)^2 \left(\frac{d^2 \mu_l}{d\eta_l^2} \right) + \frac{1}{V} \left(\frac{d\mu_l}{d\eta_l} \right) \left(\frac{d^3 \mu_l}{d\eta_l^3} \right) \right\} x_{la} x_{lb} x_{lc} x_{ld}, \quad (2.18)$$

- $\kappa_{ac,bd}$

$$\kappa_{ac,bd} = \phi \sum_{l=1}^n \left\{ \begin{aligned} & \frac{1}{V^3} (V^{(1)})^2 \left(\frac{d\mu_l}{d\eta_l} \right)^4 - \frac{2V^{(1)}}{V^2} \left(\frac{d\mu_l}{d\eta_l} \right)^2 \left(\frac{d^2\mu_l}{d\eta_l^2} \right) \\ & + \frac{1}{V} \left(\frac{d\mu_l}{d\eta_l} \right)^2 \end{aligned} \right\} x_{la}x_{lb}x_{lc}x_{ld}. \quad (2.19)$$

Substituindo os cumulantes na expressão $2\kappa_{bc}^{(ad)} - \kappa_{bcd}^{(a)} + \kappa_{ac,bd}$ e considerando o elemento h_l dada pela equação (2.20), obtém-se:

$$2\kappa_{bc}^{(ad)} - \kappa_{bcd}^{(a)} + \kappa_{ac,bd} = h_l,$$

sendo:

$$h_l = -\frac{1}{V} \left(\frac{d\mu_l}{d\eta_l} \right) \left(\frac{d^3\mu_l}{d\eta_l^3} \right) - \frac{V^{(1)}}{V^2} \left(\frac{d\mu_l}{d\eta_l} \right)^2 \left(\frac{d^2\mu_l}{d\eta_l^2} \right) + \frac{(V^{(1)})^2}{V^3} \left(\frac{d\mu_l}{d\eta_l} \right)^4. \quad (2.20)$$

Substituindo $2\kappa_{bc}^{(ad)} - \kappa_{bcd}^{(a)} + \kappa_{ac,bd}$ em $\sigma_{ij}^{(1)}$ tem-se:

$$\begin{aligned} \sigma_{ij}^{(1)} &= -\kappa^{ia} \kappa^{jb} \kappa^{cd} \phi \sum_{l=1}^n h_l x_{la} x_{lb} x_{lc} x_{ld}, \\ &= -\phi \sum_{l=1}^n h_l (-\kappa^{ia} \kappa^{jb} x_{la} x_{lb}) (\kappa^{cd} x_{lc} x_{ld}). \end{aligned}$$

Denotando por $\Sigma^{(1)}$ a matriz cujos elementos são $\sigma_{ij}^{(1)}$, e fazendo os somatórios tem-se:

$$\Sigma^{(1)} = \frac{1}{\phi^2} P H Z_d P^T, \quad (2.21)$$

sendo:

$$\begin{aligned} P &= (X^T W X)^{-1} X^T, \\ Z &= X P = X (X^T W X)^{-1} X^T, \\ H &= \text{diag} \{h_1, h_2, \dots, h_n\} \text{ e} \\ Z_d &= \text{diag} \{z_{11}, z_{22}, \dots, z_{nn}\}. \end{aligned}$$

2.5.3 Obtendo $\sigma_{ij}^{(2)}$

O procedimento para obter $\sigma_{ij}^{(2)}$ é análogo ao adotado para obtenção de $\sigma_{ij}^{(1)}$, isto é, serão utilizadas as identidades de Bartlett e as propriedades de invariância dos cumulantes sob a permutação de índices. Inicialmente será simplificada a expressão dentro do parêntese de $\sigma_{ij}^{(2)}$

que é dada por:

$$\begin{aligned} & \frac{3}{2}\kappa_{abc}\kappa_{rst} + 4\kappa_{ab,c}\kappa_{rst} + \kappa_{a,bc}\kappa_{rst} + 2\kappa_{ab,c}\kappa_{r,st} + \kappa_{ab,c}\kappa_{rt,s} \\ &= \frac{3}{2}\kappa_{abc}\kappa_{rst} + 5\kappa_{a,bc} + 3\kappa_{ab,c}\kappa_{r,st} = \kappa_{rst} \left(\frac{3}{2}\kappa_{abc} + 5\kappa_{a,bc} \right) + 3\kappa_{ab,c}\kappa_{r,st}. \end{aligned} \quad (2.22)$$

Os cumulantes necessários para obter $\sigma_{ij}^{(2)}$ são apresentados em Cordeiro (1983, p.204). O cumulante $\kappa_{a,bc}$ é dado por:

$$\kappa_{a,bc} = \phi \sum_{l=1}^n g_l x_{la} x_{lb} x_{lc}.$$

Substituindo os cumulantes em (2.22) tem-se:

$$\begin{aligned} \kappa_{rst} \left(\frac{3}{2}\kappa_{abc} + 5\kappa_{a,bc} \right) + 3\kappa_{ab,c}\kappa_{r,st} &= -\phi \sum_{m=1}^n (f_m + g_m) x_{mr} x_{ms} x_{mt} \\ &+ \left\{ \begin{array}{l} \left[\frac{-3}{2}\phi \sum_{l=1}^n (f_l + g_l) x_{la} x_{lb} x_{lc} \right] + \\ \left[5\phi \sum_{l=1}^n g_l x_{la} x_{lb} x_{lc} \right] \end{array} \right\} \\ &+ 3 \left\{ \begin{array}{l} \left[\phi \sum_{l=1}^n g_l x_{la} x_{lb} x_{lc} \right] \\ \left[\phi \sum_{m=1}^n g_m x_{ma} x_{mb} x_{mc} \right] \end{array} \right\}, \end{aligned}$$

Simplificando tem-se:

$$\kappa_{rst} \left(\frac{3}{2}\kappa_{abc} + 5\kappa_{a,bc} \right) + 3\kappa_{ab,c}\kappa_{r,st} = \phi^2 \sum_{l,m} \left(\frac{3}{2}f_l f_m - 2g_l f_m - 4g_l g_m \right) x_{la} x_{lb} x_{lc} x_{mr} x_{ms} x_{mt}.$$

Substituindo na equação de $\sigma_{ij}^{(2)}$ obtém-se:

$$\begin{aligned} \sigma_{ij}^{(2)} &= \kappa^{ia} \kappa^{jr} \kappa^{bs} \kappa^{ct} \phi^2 \sum_{l,m} \left(\frac{3}{2}f_l f_m - 2g_l f_m - 4g_l g_m \right) x_{la} x_{lb} x_{lc} x_{mr} x_{ms} x_{mt} \\ &= \phi^2 \sum_{l,m} (\kappa^{ia} x_{la}) (\kappa^{jr} x_{mr}) (\kappa^{bs} x_{lb} x_{ms}) (\kappa^{ct} x_{lc} x_{mt}) \left(\frac{3}{2}f_l f_m - 2g_l f_m - 4g_l g_m \right), \end{aligned}$$

que pode ser escrita em notação matricial como:

$$\Sigma^{(2)} = \frac{1}{\phi^2} P \left(\frac{3}{2} F Z^{(2)} F + 2 P G Z^{(2)} - 4 P Z^{(2)} G \right) P^T. \quad (2.23)$$

sendo

$$F = \text{diag}(f_1 \cdots f_n); \quad G = \text{diag}(g_1 \cdots g_n)$$

$$Z = X P, \quad \text{e} \quad Z^{(2)} = Z \odot Z$$

o símbolo \odot denota o produto de Hadamard.

2.5.4 Obtenção de $\sigma_{ij}^{(3)}$

Como nas duas últimas sub seções, a expressão dentro do parêntese de $\sigma_{ij}^{(3)}$ apresentado em (2.14) será simplificada usando as identidades de Bartlett e a invariância dos cumulantes sob a permutação de índices, isto é,

$$\begin{aligned} 2\kappa_{a,bc}\kappa_{r,st} + \kappa_{a,bc}\kappa_{rst} + \kappa_{abc}\kappa_{rst} + 2\kappa_{abc}\kappa_{r,st} &= \kappa_{a,bc}(2\kappa_{r,st} + \kappa_{rst}) + \kappa_{abc}(\kappa_{rst} + 2\kappa_{r,st}) \\ &= (\kappa_{a,bc} + \kappa_{abc})(2\kappa_{r,st} + \kappa_{rst}), \end{aligned}$$

substituindo os cumulantes apresentados em Cordeiro (1983), na expressão acima, tem-se:

$$(\kappa_{a,bc} + \kappa_{abc})(2\kappa_{r,st} + \kappa_{rst}) = \phi^2 \sum_{l,m} (g_l - f_l - 2g_l) \quad (2.24)$$

$$(2g_m - f_m - g_m) x_{la}x_{lb}x_{lc}x_{ms}x_{mr}x_{mt} \quad (2.25)$$

$$= \phi^2 \sum_{l,m} (g_l + f_l) f_m x_{la}x_{lb}x_{lc}x_{ms}x_{mr}x_{mt}. \quad (2.26)$$

Substituindo (2.26) na equação (2.14) tem-se:

$$\begin{aligned} \sigma_{ij}^{(3)} &= \phi^2 \kappa^{ia} \kappa^{jb} \kappa^{rs} \kappa^{ct} \sum_{l,m} (g_l + f_l) f_m x_{la}x_{lb}x_{lc}x_{ms}x_{mr}x_{mt} \\ &= \phi^2 \kappa^{ia} \kappa^{jb} \sum_{l,m} x_{la}x_{lb} (f_m \kappa^{rs} x_{ms}x_{mr}) (g_l + f_l) (\kappa^{ct} x_{mt}x_{lc}) \end{aligned} \quad (2.27)$$

que em notação matricial é dada por:

$$\Sigma^{(3)} = \frac{1}{\phi^2} P (F + G) Z Z_d F P^T. \quad (2.28)$$

2.5.5 Matriz de covariância Σ em notação matricial

A matriz de covariância corrigida do estimador de máxima verossimilhança para os parâmetros do modelo linear generalizado é obtida como a soma das equações (4.5), (2.23) e (2.28). É importante salientar que a matriz apresentada em (2.28) envolve operações com matrizes e pode ser facilmente implementada na *procedure IML* do SAS.

$$\Sigma = \Sigma^{(1)} + \Sigma^{(2)} + \Sigma^{(3)} = \phi^{-2} P \Lambda P^T,$$

sendo:

$$\Lambda = H Z_d + \frac{3}{2} F Z^{(2)} F + G Z^{(2)} F - G Z^{(2)} G + (F + G) Z Z_d F;$$

$$F = \text{diag} \left\{ V^{-1} \left(\frac{\partial \mu}{\partial \eta} \right) \left(\frac{\partial^2 \mu}{\partial \eta^2} \right) \right\};$$

$$G = \text{diag} \left\{ V^{-1} \left(\frac{\partial \mu}{\partial \eta} \right) \left(\frac{\partial^2 \mu}{\partial \eta^2} \right) - \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu}{\partial \eta} \right)^3 \right\};$$

$$P = (X^T W X)^{-1} = X^T;$$

$$Z = X P;$$

$$h_l = \frac{1}{V} \left(\frac{\partial \mu}{\partial \eta} \right) \left(\frac{\partial^3 \mu}{\partial \eta^3} \right) - \left(\frac{\partial \mu}{\partial \eta} \right)^2 \left(\frac{\partial^2 \mu}{\partial \eta^2} \right)^2 V^{(1)} + \left(\frac{\partial \mu}{\partial \eta} \right)^4 \frac{(V^{(1)})^2}{V^3};$$

$$H = \text{diag} \{h_1, h_2, \dots, h_n\};$$

$$Z_d = \text{diag} \{z_{11}, z_{22}, \dots, z_{nn}\};$$

$Z^{(2)} = Z \odot Z$, o operador \odot denota o produto de Hadamard.

Matriz de covariância de $O(n^{-2})$ A matriz de covariância de $O(n^{-1})$ dos EMV nos MLG é dada por

$$\text{Cov}_1(\hat{\beta}) = \phi^{-1} (X^T W X)^{-1},$$

assim, a matriz de covariância dessas estimativas de $O(n^{-2})$ é dada por:

$$\text{Cov}_2(\hat{\beta}) = \phi^{-1} (X^T W X)^{-1} + \phi^{-2} P \Lambda P^T.$$

2.6 Aplicação

Para ilustrar os resultados obtidos neste capítulo, será apresentado uma aplicação com variável resposta, simulada de uma gama, e matriz de covariáveis simuladas da uniforme [0 1]. É importante salientar que muitos processos em finanças e econometria exibem assimetria e heterocedasticidade. Exemplos incluem modelos de volatilidade, modelos de fronteira de produção estocástica, etc. Neste contexto, não é razoável supor:

- modelo com erro Normal ;
- com variância constante.

Portanto, é razoável nesses casos usar a metodologia de MLG nessas aplicações. A decisão importante é a escolha da distribuição da variável resposta, matriz modelo e função de ligação.

Para selecionar a função de densidade, uma análise exploratória dos dados é necessária e, além disso, conhecer o comportamento da distribuição, como adotou-se o modelo gama que é usado para descrever dados:

1. contínuos não-negativos;
2. que apresentam variância crescente com a média;

3. e que o coeficiente de variação seja aproximadamente constante.

A função de ligação foi a log.

Objetivos desta ilustração são:

avaliar os modelos ajustados pelas distribuições, gama e normal através do AIC;

avaliar as estimativas de $O(n^{-1})$ e $O(n^{-2})$, comparando duas amostras sendo uma delas de tamanho pequeno.

2.6.1 Modelo

componente aleatória: $Y \sim \text{Gama}(\mu, \phi)$, sendo μ a média e $\phi = 5$, parâmetro de precisão.

componente sistemática: $\eta_l = \log(\mu_l) = \beta_1 X_{1l} + \beta_2 X_{2l} + \beta_3 X_{3l}$, sendo $l = 1, \dots, 30$ para amostra pequena.

Os parâmetros β 's foram fixados em $\beta_1 = 1$, $\beta_2 = 2$ e $\beta_3 = 3$.

2.6.2 Algoritmo

2.6.3 1º Passo: Obter as estimativas de $O(n^{-1})$

A *procedure* GENMOD fornece estas estimativas, pelo método score de Fisher.

2.6.4 2º Passo: Resultados da *procedure* GENMOD

As estimativas de ordem $O(n^{-1})$ dos parâmetros do modelo e da matriz de covariância são disponibilizados em arquivos. Estes arquivos são utilizados na *procedure* IML para obter as estimativas de $O(n^{-2})$.

2.6.5 3º Passo: *procedure* IML

Esta *procedure* permite ao usuário operar com matrizes neste passo, as matrizes necessárias para as estimativas de $O(n^{-2})$ são calculadas. A seguir serão enumeradas as matrizes necessárias:

$V(\mu_l) = \mu_l^2$, para a distribuição Gama.

Matriz diagonal peso, W , cujo elemento da diagonal é:

$$w_{ii} = \frac{1}{V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = 1.$$

Matriz diagonal, F , cujo elemento diagonal destas matrizes é:

$$f_{ii} = \frac{1}{V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial^2 \mu_i}{\partial \eta_i^2} = 1;$$

Matriz diagonal, G , cujo elemento diagonal destas matrizes é:

$$g_{ii} = \frac{1}{V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial^2 \mu_i}{\partial \eta_i^2} - \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^3 = 0;$$

Matriz diagonal, H , cujo elemento diagonal destas matrizes é:

$$h_{ii} = 3 + 2\mu_i^2$$

2.6.6 Resultados

Para avaliar a qualidade do ajuste o GENMOD apresenta as estatísticas *deviance*, *scaled deviance*, *Pearson chi-square*, *Scaled Pearson* e log-verossimilhança. Todas essas estatísticas são descritas em Demétrio (1993), que utiliza como critério de ajuste o quociente entre o *scaled deviance* e os graus de liberdade. Quando este quociente for próximo de um, considera-se que o modelo ajusta-se aos dados.

A tabela a seguir apresenta estas estatísticas para a amostra simulada. Observe que o modelo ajusta-se bem à amostra simulada.

Tabela 2.1: Critérios para avaliar o ajuste

Critério	Graus de liberdade	Valor	Valor/graus de liberdade.
<i>deviance</i>	27	3,0582	0,1133
<i>scaled deviance</i>	27	30,5008	1,1297
<i>Pearson chi-square</i>	27	2,5784	0,0955
<i>Scaled Pearson</i>	27	25,7153	0,9524
log -verossimilhança		-36,7708	

A Tabela 2 fornece as estimativas de $O(n^{-1})$ e $O(n^{-2})$. Observe que as magnitudes dos vieses são pequenos quando comparados aos erros-padrão. No entanto, as estimativas corrigidas estão mais próximas dos parâmetros, justificando assim o uso dos vieses.

As matrizes de covariância de $O(n^{-1})$ e $O(n^{-2})$. São apresentadas a seguir. Observe que os elementos da expansão são bem maiores que os da primeira ordem, portanto deve ser considerados.

Tabela 2.2: Estimativas do vetor de parâmetros β

Parâmetros	Estimativas de $O(n^{-1})$	Erros-padrão da Estimativa de $O(n^{-1})$	Vieses	Estimativas de $O(n^{-2})$	Quocientes (Viés/ erro-padrão)*100
β_0	0,743974	0,496154	-0,004900	0,748873	0,98%
β_1	1,677566	0,546083	-0,004647	1,682213	0,85%
β_2	2,994055	0,602070	-0,007938	3,001993	1,30%

Matriz de covariância de $O(n^{-1})$ de $\hat{\beta}$ para amostra simulada

$$Cov_1(\hat{\beta}) = \begin{bmatrix} & \hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 \\ \hat{\beta}_1 & 0,0658 & -0,0164 & -0,0344 \\ \hat{\beta}_2 & & 0,04653 & -0,0223 \\ \hat{\beta}_3 & & & 0,0566 \end{bmatrix}.$$

Matriz de covariância de $O(n^{-2})$ de $\hat{\beta}$ para amostra simulada

$$Cov_2(\hat{\beta}) = \begin{bmatrix} & \hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 \\ \hat{\beta}_1 & 1,0556 & -0,2203 & -0,3405 \\ \hat{\beta}_2 & & 0,5083 & -0,241712 \\ \hat{\beta}_3 & & & 0,8394 \end{bmatrix}$$

Diante do exposto, pode-se dizer que para a amostra simulada as correções são necessárias.

Capítulo 3

Modelos Não-Lineares da Família Exponencial

Esta classe de modelos é mais abrangente que as dos MLG, uma vez que permite para o preditor η qualquer função da matriz modelo. Quando η é linear, tem-se a classe dos MLG. Assim, a teoria unificada para os MNLFE, que será apresentada neste capítulo, é válida para modelos de regressão normal linear, incluindo MLG e modelos de regressão não-linear normal.

Os MNLFE são uma ferramenta útil em aplicações onde as suposições e/ou comportamento da variável resposta não se adequam aos MLG ou não-linear de regressão. Como aplicação destes modelos têm-se:

- i)* os modelos de curva de crescimento apresentados por Ratkowsky (1983), que são amplamente utilizados nas áreas de finanças, agricultura, dentre outras;
- ii)* os modelos *one compartment*, que descrevem a concentração de uma determinada droga no organismo. Estes modelos são utilizados pelas indústrias farmacêuticas para avaliar as características de um novo medicamento;
- iii)* os modelos de demanda de energia;
- iv)* os modelos biexponenciais que são usualmente utilizados para avaliar o prazo de validade de um medicamento.

Observação Outras aplicações são encontrados em Paula (1991).

Os procedimentos de estimação e obtenção das estatísticas necessárias para o ajuste e adequabilidade do modelo são análogos aos apresentados nos MLG. O sistema de equações é não-linear e necessitam de um processo iterativo que são:

- pelo método de Newton Raphson, usando um comando ou *procedure* de otimização, que no SAS estão disponibilizadas nas *procedure* NLP e NLMIXED;

- pelo método escore de Fisher, como nos MLG, é também um processo iterativo, onde o sistema de equações pode ser visto como as equações de mínimo quadrados reponderados, substituindo a matriz modelo X pela matriz das primeiras derivadas de $\boldsymbol{\eta}$ com relação aos parâmetros $\boldsymbol{\beta}$, que será denotada por \tilde{X} e o vetor ajustado \mathbf{y}^* passa a ter mais uma componente devido a não-linearidade do modelo. Este método no SAS está disponível na *procedure* GENMOD, sendo necessário o recurso da *offset*, que é apresentado em Cordeiro e Paula (1989).

3.1 Modelo

Os MNLF ficam caracterizados pelo trinômio:

- componente aleatória, onde supõe-se que, as componentes de $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ sejam variáveis aleatórias independentes e cada \mathbf{Y}_l tem uma distribuição pertencente à família exponencial;
- componente sistemática, onde as covariáveis \mathbf{x}_i fornecem um conjunto de preditores

$$\eta_l = f(\mathbf{x}_i; \boldsymbol{\beta}), \quad l = 1, \dots, n, \quad (3.1)$$

sendo: $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ definido num subconjunto B de \mathfrak{R}^p onde $p < n$, vetor de parâmetros e $f(\mathbf{x}_i; \boldsymbol{\beta})$ uma função qualquer da matriz modelo dado $\boldsymbol{\beta}$;

- função de ligação entre média e o preditor não-linear η_l é dado por:

$$g(\mu_l) = \eta_l,$$

em que: μ_l é a esperança de Y_l , $l = 1, 2, \dots, n$ e $g(\cdot)$ é chamada de função de ligação do modelo.

Observa-se, assim, que este modelo pode ser aplicado a qualquer membro da família exponencial, e a presença da função de ligação, quando conectada ao preditor $\boldsymbol{\eta}$ para esperança de \mathbf{Y} .

Como nos MLG, quando $\theta_l = \eta_l$, a função de ligação é dita ser canônica. Em qualquer problema poderia utilizar várias funções de ligação, contudo, é conveniente utilizar a ligação canônica, pois uma das vantagens de usar esta é que as mesmas garantem a concavidade da log-verossimilhança e, conseqüentemente, muitos resultados assintóticos são obtidos mais facilmente. Além disso, a interpretação dos parâmetros do modelo é em termos do parâmetro canônico.

Assume-se que os MNLFE satisfazem as condições de regularidade, apresentadas em Wei (1998).

3.2 Matriz de informação de Fisher

As matrizes de informação de Fisher de β e ϕ para os MNLFE são obtidas de modo análogo ao MLG. São necessárias no método score de Fisher e ainda esta é a matriz covariância assintótica de $\hat{\beta}$.

3.2.1 Matriz de informação de Fisher de β

Esta matriz será denotada por $K(\beta)$ que por definição é dada por:

$$K(\beta) = -E \left(\frac{\partial^2 \ell}{\partial \beta \partial \beta^T} \right),$$

a derivada segunda da log-verossimilhança é obtida pela regra da cadeia de segunda ordem, cuja demonstração é apresentada no Apêndice E, isto é,

$$\begin{aligned} U_{rs} &= \frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} = \phi \sum_l^n -\frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 (r, s) \\ &\quad - (y_l - \mu_l) \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 (r, s) + (y_l - \mu_l) \frac{1}{V^2} \frac{\partial^2 \mu_l}{\partial \eta_l^2} (r, s) \\ &\quad + (y_l - \mu_l) \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} (rs), \end{aligned} \quad (3.2)$$

sendo:

- que r e s variam de $1, \dots, p$;
- $(r, s) = \frac{\partial \eta_l}{\partial \beta_r} \frac{\partial \eta_l}{\partial \beta_s}$;
- $(rs) = \frac{\partial^2 \eta_l}{\partial \beta_r \partial \beta_s}$;
- $V^{(r)} = \frac{\partial^r V}{\partial \mu^r}$.

Fazendo a esperança de (3.2), tem-se:

$$E(-U_{rs}) = \phi \sum_l^n \frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 (r, s),$$

portanto, a matriz de informação de Fisher de β que será denotada por $K(\beta)$, em notação matricial, fica:

$$K(\beta) = E \left(-\frac{\partial^2 \ell}{\partial \beta \partial \beta^T} \right) = \phi \tilde{X}^T W \tilde{X};$$

- a matriz W é diagonal com elemento da diagonal igual $\frac{1}{V(\mu_l)} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2$;

- a matriz \tilde{X} é a matriz jacobiana da forma

$$\tilde{X} = \begin{bmatrix} \frac{\partial f(\mathbf{x}_{11};\boldsymbol{\beta})}{\partial \beta_1} & \frac{\partial f(\mathbf{x}_{21};\boldsymbol{\beta})}{\partial \beta_2} & \dots & \frac{\partial f(\mathbf{x}_{p1};\boldsymbol{\beta})}{\partial \beta_p} \\ \frac{\partial f(\mathbf{x}_{12};\boldsymbol{\beta})}{\partial \beta_1} & \frac{\partial f(\mathbf{x}_{22};\boldsymbol{\beta})}{\partial \beta_2} & \dots & \frac{\partial f(\mathbf{x}_{p2};\boldsymbol{\beta})}{\partial \beta_p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial f(\mathbf{x}_{1m};\boldsymbol{\beta})}{\partial \beta_1} & \frac{\partial f(\mathbf{x}_{2m};\boldsymbol{\beta})}{\partial \beta_2} & \dots & \frac{\partial f(\mathbf{x}_{pm};\boldsymbol{\beta})}{\partial \beta_p} \end{bmatrix} \text{ ou } \tilde{X} = \left[\frac{\partial f(X_1;\boldsymbol{\beta})}{\partial \beta_1}, \frac{\partial f(X_2;\boldsymbol{\beta})}{\partial \beta_2}, \dots, \frac{\partial f(X_p;\boldsymbol{\beta})}{\partial \beta_p} \right]$$

3.2.2 Matriz de informação de Fisher de ϕ

Por definição tem-se que:

$$K(\phi) = E \left(-\frac{\partial^2 \ell}{\partial \phi^2} \right),$$

como:

$$\frac{\partial \ell}{\partial \phi} = -\frac{1}{2} D'(Y, \mu(\beta)) - \frac{1}{2} \sum_{i=1}^n \frac{\partial c(y_i, \phi)}{\partial \phi},$$

logo, $\hat{\phi}$ satisfaz a:

$$\sum_{i=1}^n \frac{\partial c(y_i, \hat{\phi})}{\partial \phi} = -D'(Y, \hat{\mu}).$$

e pode-se encontrar $\hat{\phi}$ após obter $\hat{\beta}$. Aplicando a derivada segunda na equação do log-verossimilhança obtém-se:

$$\frac{\partial^2 \ell}{\partial \phi^2} = -\frac{1}{2} \sum_{i=1}^n \frac{\partial^2 c(y_i, \phi)}{\partial \phi^2},$$

portanto,

$$K(\phi) = -\frac{1}{2} \sum_{i=1}^n \frac{\partial^2 c(y_i, \phi)}{\partial \phi^2}.$$

3.3 EMV dos parâmetros do modelo

Para o MNLFE, os parâmetros de interesse são:

$\boldsymbol{\mu}$: vetor de parâmetros média do vetor resposta \mathbf{Y} ;

$\boldsymbol{\theta}$: vetor de parâmetros canônico, da função densidade de probabilidade do vetor resposta \mathbf{Y} ;

ϕ : parâmetro de dispersão, da função densidade de probabilidade do vetor resposta \mathbf{Y} ;

$\boldsymbol{\beta}$: vetor de parâmetros da função linear que relaciona o preditor não-linear η ao vetor de parâmetros média $\boldsymbol{\mu}$.

Observação: Cabe ressaltar que os parâmetros β e ϕ , são ortogonais e, portanto, podem ser estimados separadamente, como nos MLG.

EMV de β

O procedimento para obter o EMV de β , para esta classe de modelos é análogo ao adotado para a classe dos MLG, sendo portanto necessário obter a derivada de primeira ordem da log-verossimilhança, e igualar a zero.

A expressão da primeira derivada da log-verossimilhança para um parâmetro β_r , que é apresentada com mais detalhes no Apêndice D, é dada por:

$$U_r = \frac{\partial \ell}{\partial \beta_r} = \sum_l^n \phi(y_l - \mu_l) \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l}(r), \text{ com}$$

$$r = 1, \dots, p. \text{ e } (r) = \frac{\partial f(X; \beta)}{\partial \beta_r}$$

O sistema das p equações pode ser escrito em notação matricial. Para isto, considere $\tilde{X}_{n \times p}$ a matriz de derivadas parciais, como descrito na seção anterior. Assim, este sistema pode ser escrito em notação matricial como:

$$U(\beta) = \frac{\partial \ell}{\partial \beta^T} = (V^{-1} \tilde{X})^T \phi(\mathbf{y} - \boldsymbol{\mu}(\beta)).$$

O sistema de equações de máxima verossimilhança é obtido igualando a zero $U(\beta)$, como V é uma matriz diagonal e ϕ uma constante, obtem-se:

$$\tilde{X}^T(\hat{\beta}) V^{-1}(\hat{\beta}) (Y - \mu(\hat{\beta})) = 0,$$

a solução deste sistema de equações é o estimador de máxima verossimilhança de β .

Método iterativo para solução do sistema $U(\hat{\beta}) = 0$

Este sistema de equações é não linear, ou seja, para obter $\hat{\beta}$ é necessário um método numérico e usualmente adota-se o método de Newton Raphson, que consiste em expandir em série de Taylor, até primeira ordem, numa vizinhança de uma tentativa inicial β_0 , a equação de verossimilhança $U(\hat{\beta}) = 0$.

A expansão de $U\left(\hat{\beta}\right)$ em série de Taylor até ordem 1, na vizinhança de β_0 , é:

$$U\left(\hat{\beta}\right) \approx U\left(\beta_0\right) + \frac{\partial^2 \ell}{\partial \beta \partial \beta^T}\left(\beta_0\right)\left(\hat{\beta} - \beta_0\right),$$

que igualando a zero resulta:

$$\hat{\beta} \approx \beta_0 + \left[-\frac{\partial^2 \ell}{\partial \beta \partial \beta^T}\left(\beta_0\right) \right]^{-1} \frac{\partial \ell}{\partial \beta}\left(\beta_0\right). \quad (3.3)$$

O sistema de equações acima fornece um processo iterativo para obtenção de $\hat{\beta}$. Para isto, basta substituir na iteração i o valor de $\hat{\beta}$ obtido na iteração $(i - 1)$.

O processo converge, quando $abs\left[\left(\hat{\beta}\right)^i - \left(\hat{\beta}\right)^{i-1} \right] < \epsilon$.

Pelo método score de Fisher, pode-se substituir na equação (3.3) a parcela da derivada de ordem 2 da log-verossimilhança pela matriz de informação de Fisher. Portanto, na iteração i , o estimador de $\hat{\beta}$ é dado por:

$$\hat{\beta} = \left(\tilde{X}^T W \tilde{X} \right)^{-1} \tilde{X} W \mathbf{y}^*,$$

sendo, as matrizes \tilde{X}, W e o vetor $\mathbf{y}^* = \tilde{X} \hat{\beta} + \text{diag}\left(\frac{\partial \eta_l}{\partial \mu_l}\right) (\mathbf{y} - \boldsymbol{\mu})$ calculadas em $\hat{\beta}$ da iteração anterior.

Cordeiro e Paula (1989) apresentam um procedimento onde é possível obter as estimativas dos parâmetros dos MNLFE usando uma *procedure* dos MLG iterativamente. Para isto, eles consideram como *offset*:

$$\boldsymbol{\tau} = f(X; \boldsymbol{\beta}) - \tilde{X} \hat{\boldsymbol{\beta}},$$

assim o vetor ajustado \mathbf{y}^* pode ser escrito como:

$$\mathbf{y}^* = \boldsymbol{\eta} - \boldsymbol{\tau} + \text{diag}\left(\frac{\partial \eta_l}{\partial \mu_l}\right) (\mathbf{y} - \boldsymbol{\mu}).$$

No SAS, pode-se, por meio de uma macro, obter em cada iteração os vetores $\boldsymbol{\tau}$, $\boldsymbol{\mu}$ e $\boldsymbol{\eta}$, a matriz \tilde{X} , e pela *procedure* GENMOD, obter o valor de $\hat{\boldsymbol{\beta}}$ nessa iteração.

EMV de ϕ

Para obter o EMV de ϕ é análogo aos modelos lineares generalizados anterior. Assim, a derivada de primeira ordem do log-verossimilhança é dada por:

$$U(\phi) = \frac{\partial \ell}{\partial \phi} = -\frac{1}{2}D(\mathbf{y} - \mu(\boldsymbol{\beta})) - \frac{1}{2} \sum_{i=1}^n \frac{\partial c(\mathbf{y}, \phi)}{\partial \phi},$$

igualando a zero, tem-se que $\hat{\phi}$ deve satisfazer:

$$\sum_{i=1}^n \frac{\partial c(\mathbf{y}, \phi)}{\partial \phi} = -\frac{1}{2}D'(y - \mu(\hat{\beta})).$$

Cordeiro e McCullagh (1991) apresentam métodos aproximados para se obter as estimativas do parâmetro de dispersão $\sigma^2 = \phi^{-1}$, que é dado por:

$$\hat{\phi} = \frac{n[1 + \{1 + 2D(Y, \hat{\mu})/3n\}^{1/2}]}{2D(Y, \hat{\mu})}.$$

Observa-se aqui que, nas procedures NLP e NLMIXED, não há necessidade desta aproximação uma vez que, a log-verossimilhança é maximizada diretamente.

Propriedades dos EMV de $\boldsymbol{\beta}$ e ϕ

As propriedades destes estimadores são assintóticas e a demonstração destas são apresentadas em Fahrmeir e Kaufmann (1985), como segue:

- (a) $\hat{\boldsymbol{\beta}}$ é assintoticamente normal, com média $\boldsymbol{\beta}$ e matriz de covariância $(\tilde{X}^T \phi W \tilde{X})^{-1}$, quando ϕ for desconhecido pode ser substituído por alguma estimativa consistente, como já apresentada.
- (b) $\hat{\sigma}_1^2 = (n - p)^{-1}D(Y, \hat{\mu})$ sendo um estimador não-viesado assintótico de σ^2 .
- (c) $\hat{\sigma}_2^2 = (n - p)^{-1}X^2$, onde $X^2 = (Y - \hat{\mu})^T V(\hat{\mu})^{-1}(Y - \hat{\mu})$ é a estatística generalizada de Pearson, que é um estimador de momento, e pode ser utilizada para avaliar a qualidade do ajuste.

3.4 Viés de $O(n^{-1})$ dos EMV para parâmetros $\boldsymbol{\beta}$, η e μ

Os EMV para os MNLFE são viesados de $O(n^{-1})$. Assim, é útil, para amostras pequenas, obter uma estimativa do viés, pois em alguns casos o viés pode tomar valores de mesma magnitude que o erro-padrão da estimativa.

Paula (1991) apresenta o viés de $O(n^{-1})$. Neste artigo, o viés é obtido pela expressão apresentada por Cox Snell (1968), conforme mostrdo no Apêndice C para maiores detalhes, mas pode-se também obtê-lo pela expressão creditada a Peers e Iqbal (1885), que é apresentada no Apêndice D.

Nos dois procedimentos são necessários os cumulantes κ_{rtu} e $\kappa_{rt,u}$ para o cálculo do viés do estimador de máxima verossimilhança de β , denotado por $B(\hat{\beta}_a)$ de ordem n^{-1} , cuja expressão são:

$$\begin{aligned}\kappa_{rtu} &= -\phi \sum_l (f + 2g)_l(r, t, u) + w_l \{(rt, u) + (tu, r) + (ru, t)\}_l, \\ \kappa_{rt,u} &= \phi \sum_l g_l(r, t, u) + w_l(rt, u),\end{aligned}$$

sendo:

$$f_l : \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2};$$

$$g_l : \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2} - \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3;$$

$$(r) \frac{\partial \eta_l}{\partial \beta_r};$$

$$(r, t, u) : \frac{\partial \eta_l}{\partial \beta_r} \frac{\partial \eta_l}{\partial \beta_s} \frac{\partial \eta_l}{\partial \beta_u};$$

$$(ru, t) : \frac{\partial^2 \eta_l}{\partial \beta_r \partial \beta_u} \frac{\partial \eta_l}{\partial \beta_t}.$$

Substituindo estes cumulantes na expressão creditada a Peers e Iqbal (1985), e, após algumas manipulações algébricas, tem-se que o viés de $O(n^{-1})$ de $\hat{\beta}_a$ é dado por:

$$\begin{aligned}B(\hat{\beta}_a) &= -(2\phi)^{-1} \sum_l f_l \sum_r \kappa^{-ar} \frac{\partial \eta}{\partial \beta_r} z_{lu} \\ &\quad - (2\phi)^{-1} \sum_l w_l \sum_r \kappa^{-ar} \frac{\partial \eta}{\partial \beta_r} \sum_{t,u} \kappa^{-tu} \frac{\partial \eta}{\partial \beta_l},\end{aligned}$$

sendo r, t e u variam de $1, 2, \dots, p$ e l varia de $1, 2, \dots, n$.

Pode-se reescrever a expressão acima em notação matricial. Para isto, considere:

Matriz Diagonal F \rightarrow é uma matriz cujo elemento da diagonal é $f_{ll} = V^{-1} \frac{\partial \mu}{\partial \eta} \frac{\partial^2 \mu}{\partial \eta^2}$ esta matriz tem dimensão $n \times n$;

Matriz Z $\rightarrow Z_{n \times n} = \{z_{ij}\} = \tilde{X}^T \left(\tilde{X}^T W \tilde{X} \right)^{-1} \tilde{X}$;

Matriz Z_d $\rightarrow Z_d = \text{diag} \{z_{11}, z_{22}, \dots, z_{nn}\}$;

Matriz $C \rightarrow C = \text{diag} \{c_1, c_2, \dots, c_n\}$, $c_l = \text{tr} \left\{ \tilde{X}_l \left(\tilde{X}^T W \tilde{X} \right)^{-1} \right\}$, \tilde{X}_l é uma matriz $p \times p$ cujos elementos são $\frac{\partial^2 \eta}{\partial \beta_r \partial \beta_s}$;

Vetor $\mathbf{1} \rightarrow \mathbf{1} = (1, 1, \dots, 1)^T$..

Fazendo o somatório, obtém-se a seguinte expressão para o viés de $\hat{\beta}$

$$B(\hat{\beta}) = -(2\phi)^{-1} \left(\tilde{X}^T W \tilde{X} \right)^{-1} \tilde{X}^T W Z_d W^{-1} F \mathbf{1} \\ - (2\phi)^{-1} \left(\tilde{X}^T W \tilde{X} \right)^{-1} \tilde{X}^T W \text{tr} \left\{ \tilde{X}_l \left(\tilde{X}^T W \tilde{X} \right)^{-1} \right\} \mathbf{1},$$

fazendo $\xi_1 = -(2\phi)^{-1} Z_d W^{-1} F \mathbf{1}$ e $\xi_2 = -(2\phi)^{-1} C \mathbf{1}$ obtém-se o seguinte vetor do viés

$$B(\hat{\beta}) = -(2\phi)^{-1} \left(\tilde{X}^T W \tilde{X} \right)^{-1} \tilde{X}^T W (\xi_1 + \xi_2). \quad (3.4)$$

Para o cálculo do viés de $\hat{\eta}$, expande-se em série de Taylor até primeira ordem em uma vizinhança de $\hat{\beta}$ o preditor não-linear $\hat{\eta}$,

$$f(x_l, \hat{\beta}_r) = f(x_l, \beta) + \frac{\partial f(x_l, \beta)}{\partial \beta_r} (\hat{\beta}_r - \beta_r) \\ + \frac{1}{2} \frac{\partial^2 f(x_l, \beta)}{\partial \beta_r \partial \beta_s} (\hat{\beta}_r - \beta_r) (\hat{\beta}_s - \beta_s),$$

aplicando a esperança, tem-se:

$$B(\hat{\eta}_a) = \frac{\partial \eta_l}{\partial \beta_r} B(\hat{\beta}_a) + \frac{1}{2} \frac{\partial^2 \eta_l}{\partial \beta_r \partial \beta_s} V(\hat{\beta}_a),$$

que substituindo as matrizes e fazendo algumas operações algébricas obtém-se a expressão para o viés de $\hat{\eta}$ que é dado por:

$$B(\hat{\eta}) = -(2\phi)^{-1} Z Z_d F - (2\phi)^{-1} Z W C + (2\phi)^{-1} C \\ = -(2\phi)^{-1} \{ Z Z_d F + (Z W - I) C \} \mathbf{1}. \quad (3.5)$$

Para o cálculo do viés de ordem n^{-1} para $\hat{\mu}$ expande-se $g^{-1}(\hat{\eta}) = \hat{\mu}$ em série de Taylor,

$$g^{-1}(\hat{\eta}) - g^{-1}(\eta) = \frac{dg^{-1}(\eta)}{d\eta_r} (\hat{\eta}_r - \eta_r) + \frac{1}{2} \frac{d^2 g^{-1}(\eta)}{d\eta_r d\eta_s} (\hat{\eta}_r - \eta_r) (\hat{\eta}_s - \eta_s),$$

aplicando esperança, tem-se:

$$B(\hat{\mu}) = \mathbf{B}_1(\hat{\eta}) \frac{d\mu_l}{d\eta_l} + \frac{1}{2} V_l(\hat{\eta}) \frac{d^2 \mu_l}{d\eta_l^2}$$

considerando I a matriz identidade de ordem n , Z a matriz de ordem $n \times n$, semi-definida positiva de posto p , $G_1 = \text{diag} \left\{ \frac{d\mu}{d\eta} \right\}$ e $G_2 = \text{diag} \left\{ \frac{d^2\mu}{d\eta^2} \right\}$, a notação matricial fica

$$\begin{aligned} B(\hat{\boldsymbol{\mu}}) &= G_1 \left[-(2\phi)^{-1} \{ZZ_d F + (ZW - I)C\} + (2\phi)^{-1} G_2 Z_d \right] \\ &= (2\phi)^{-1} \{(G_2 - G_1 Z F) Z_d - G_1 (ZW - I) C\} \mathbf{1}. \end{aligned} \quad (3.6)$$

Substituindo ϕ por um estimador consistente nas equações (3.4), (3.5) e (3.6) obtém-se os estimadores de máxima verossimilhança de segunda ordem para $\boldsymbol{\beta}$, $\boldsymbol{\eta}$ e μ dados, respectivamente por:

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - B(\hat{\boldsymbol{\beta}}), \quad \tilde{\boldsymbol{\eta}} = \hat{\boldsymbol{\eta}} - B(\hat{\boldsymbol{\eta}}) \quad \text{e} \quad \tilde{\mu} = \hat{\mu} - B(\hat{\mu}).$$

O viés de ordem n^{-1} para o estimador de máxima verossimilhança para o parâmetro de dispersão ϕ a partir da equação apresentada no Capítulo 2 que é devida a Cordeiro e McCullagh (1991).

Capítulo 4

Matriz de covariância de $O(n^{-2})$

Neste capítulo, a matriz de covariância de $O(n^{-2})$ é apresentada em termos dos cumulantes e de matrizes.

A metodologia deve-se a Cordeiro (2004) e consiste em substituir, os cumulantes apresentados no Apêndice F, na expressão obtida por Peers e Iqbal (1985).

4.1 Função geradora de cumulantes

A matriz de covariância de $O(n^{-2})$ será obtida por meio da expressão proposta por Peers e Iqbal (1985) e cujo esboço de sua demonstração é apresentada no Apêndice D.

A expressão para obtenção do elemento σ_{ij} é dada por:

$$\sigma_{ij} = \sigma_{ij}^{(1)} + \sigma_{ij}^{(2)} + \sigma_{ij}^{(3)}$$

sendo:

$$\sigma_{ij}^{(1)} = -\kappa^{ia} \kappa^{jb} \kappa^{cd} (\kappa_{abcd} + \kappa_{a,bcd} + 2\kappa_{abc,d} + 2\kappa_{a,bc,d} + 3\kappa_{ac,bd}), \quad (4.1)$$

$$\sigma_{ij}^{(2)} = \kappa^{ia} \kappa^{jr} \kappa^{bs} \kappa^{ct} \left(\frac{3}{2} \kappa_{abc} \kappa_{rst} + 4\kappa_{ab,c} \kappa_{rst} + \kappa_{a,bc} \kappa_{rst} + 2\kappa_{ab,c} \kappa_{r,st} + \kappa_{ab,c} \kappa_{rt,s} \right) \text{ e} \quad (4.2)$$

$$\sigma_{ij}^{(3)} = \kappa^{ia} \kappa^{jb} \kappa^{rs} \kappa^{ct} (2\kappa_{a,bc} \kappa_{r,st} + \kappa_{a,bc} \kappa_{rst} + \kappa_{abc} \kappa_{rst} + 2\kappa_{abc} \kappa_{r,st}). \quad (4.3)$$

As expressões de $\sigma_{ij}^{(1)}$, $\sigma_{ij}^{(2)}$ e $\sigma_{ij}^{(3)}$ podem ser simplificadas através das identidades de Bartlett e pela invariância dos cumulantes sob a permutação de índices. Para simplificar $\sigma_{ij}^{(1)}$, considere as seguintes identidades:

$$\begin{aligned} \kappa_{a,bcd} &= -\kappa_{abcd} + \kappa_{bcd}^{(a)}, \\ \kappa_{a,bc,d} &= \kappa_{a,d,bc} = \kappa_{abcd} - \kappa_{abc}^{(d)} - \kappa_{bcd}^{(a)} + \kappa_{bc}^{(ad)} - \kappa_{ad,bc}, \\ \kappa_{abc,d} &= -\kappa_{abcd} + \kappa_{abc}^{(d)}, \end{aligned}$$

usando estas identidades, os cumulantes dentro do parêntese de $\sigma_{ij}^{(1)}$ podem ser expressos da seguinte forma:

$$\begin{aligned}
\kappa_{abcd} &= \kappa_{abcd}, \\
\kappa_{a,bcd} &= -\kappa_{abcd} + \kappa_{abc}^{(d)}, \\
2\kappa_{a,bc,d} &= 2\kappa_{abcd} - 2\kappa_{abc}^{(d)} - 2\kappa_{bcd}^{(a)} + 2\kappa_{bc}^{(ad)} - 2\kappa_{ad,bc}, \\
3\kappa_{ac,bd} &= 3\kappa_{ad,bc}, \\
2\kappa_{abc,d} &= 2\kappa_{abcd} + 2\kappa_{bcd}^{(a)},
\end{aligned}$$

somando estas 5 equações, obtém-se uma simplificação para a expressão (4.1).

Para simplificar $\sigma_{ij}^{(2)}$, as propriedades de invariância dos cumulantes sob a permutação de índices, será utilizada no termo entre parênteses da expressão (4.2), isto é:

$$\begin{aligned}
P1 &= \frac{3}{2}\kappa_{abc}\kappa_{rst} + 4\kappa_{ab,c}\kappa_{rst} + \kappa_{a,bc}\kappa_{rst} + 2\kappa_{ab,c}\kappa_{r,st} + \kappa_{ab,c}\kappa_{rt,s} \\
&= \frac{3}{2}\kappa_{abc}\kappa_{rst} + 5\kappa_{a,bc} + 3\kappa_{ab,c}\kappa_{r,st} \\
&= \kappa_{rst} \left(\frac{3}{2}\kappa_{abc} + 5\kappa_{a,bc} \right) + 3\kappa_{ab,c}\kappa_{r,st}.
\end{aligned} \tag{4.4}$$

que é substituído na equação (4.2). Este procedimento é também adotado para simplificar a expressão de $\sigma_{ij}^{(3)}$. As expressões simplificadas são:

$$\sigma_{ij}^{(1)} = -\kappa^{ia}\kappa^{jb}\kappa^{cd} \left(2\kappa_{bc}^{(ad)} - \kappa_{bcd}^{(a)} + \kappa_{ac,bd} \right), \tag{4.5}$$

$$\sigma_{ij}^{(2)} = \kappa^{ia}\kappa^{jr}\kappa^{bs}\kappa^{ct} \left(\kappa_{rst} \left(\frac{3}{2}\kappa_{abc} + 5\kappa_{a,bc} \right) + 3\kappa_{ab,c}\kappa_{r,st} \right) e \tag{4.6}$$

$$\sigma_{ij}^{(3)} = \kappa^{ia}\kappa^{jb}\kappa^{rs}\kappa^{ct} (\kappa_{a,bc} + \kappa_{abc}) (2\kappa_{r,st} + \kappa_{rst}). \tag{4.7}$$

Os índices a, b, c, d, r, s, t variam de $1, \dots, p$, e os cumulantes da expressão acima são apresentados com detalhes no Apêndice F e são dados por:

Cumulante κ_{rst}

$$\kappa_{rst} = \phi \sum_l^n - (f_l + 2g_l)(r, s, t) - w_l [(rt, s) + (st, r) + (rs, t)];$$

Cumulante $\kappa_{a,bc}$

$$\kappa_{a,bc} = \phi \sum_l^n (\kappa_{a,bc})_{L1} + (\kappa_{a,bc})_{NL1};$$

sendo: $(\kappa_{a,bc})_{L1} = g_l(a, b, c)$ e $(\kappa_{a,bc})_{NL1} = w_l(a, bc)$.

Cumulante $\kappa_{ac,bd}$

$$\kappa_{ac,bd} = \phi \sum_l^n (\kappa_{ac,bd})_{L1} + (\kappa_{ac,bd})_{NL1} + (\kappa_{ac,bd})_{NL2},$$

sendo:

$$(\kappa_{ac,bd})_{L1} = \left[\begin{array}{c} -\frac{2V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} + \frac{1}{V} \left(\frac{\partial^2 \mu_l}{\partial \eta_l^2} \right)^2 \\ + \frac{(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 \end{array} \right] (a, b, c, d),$$

$$(\kappa_{ac,bd})_{NL1} = g_l[(a, c, bd) + (ac, b, d)] \quad \mathbf{e} \quad (\kappa_{ac,bd})_{NL2} = w_l(ac, bd).$$

Cumulante $\kappa_{ac,bd}$

$$\kappa_{bcd}^{(a)} = \frac{\partial \kappa_{bcd}}{\partial \beta_a} = \phi \sum_l^n \left(\kappa_{bcd}^{(a)} \right)_{L1} + \left(\kappa_{bcd}^{(a)} \right)_{NL1} + \left(\kappa_{bcd}^{(a)} \right)_{NL2} + \left(\kappa_{bcd}^{(a)} \right)_{NL3},$$

sendo:

$$\left(\kappa_{bcd}^{(a)} \right)_{L1} = \left[\begin{array}{c} \frac{9V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} + \frac{3}{V} \left(\frac{\partial^2 \mu_l}{\partial \eta_l^2} \right)^2 - \frac{3}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} \\ - \frac{4(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 + \frac{2V^{(2)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 \end{array} \right] (a, b, c, d),$$

$$\left(\kappa_{bcd}^{(a)} \right)_{NL1} = -f_l \left[\begin{array}{c} (ab, c, d) + (ac, b, d) + (ad, b, d) \\ + (a, bc, d) + (a, bd, c) + (a, cd, b) \end{array} \right],$$

$$\left(\kappa_{bcd}^{(a)} \right)_{NL2} = -g_l \left[\begin{array}{c} 2(ab, c, d) + 2(ac, b, d) + 2(ad, b, d) \\ + (a, bc, d) + (a, bd, c) + (a, cd, b) \end{array} \right] \mathbf{e}$$

$$\left(\kappa_{bcd}^{(a)} \right)_{NL3} = -w_l[(abc, d) + (bd, ad) + (abd, c) + (bd, ac) + (cd, ab)].$$

Cumulante $\kappa_{bc}^{(ad)}$

$$\kappa_{bc}^{(ad)} = \frac{\partial^2 \kappa_{bc}}{\partial \beta_a \partial \beta_d} = \phi \sum_l^n \left(\kappa_{bc}^{(ad)} \right)_{L1} + \left(\kappa_{bc}^{(ad)} \right)_{NL1} + \left(\kappa_{bc}^{(ad)} \right)_{NL2}$$

sendo:

$$\left(\kappa_{bc}^{(ad)} \right)_{L1} = \left[\begin{array}{c} \frac{5V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} - \frac{2}{V} \left(\frac{\partial^2 \mu_l}{\partial \eta_l^2} \right)^2 - \frac{2}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} - \\ \frac{2(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 + \frac{V^{(2)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 \end{array} \right] (a, b, c, d);$$

$$\begin{aligned} \left(\kappa_{bc}^{(ad)} \right)_{NL1} &= -(g_l + f_l) \left[\begin{array}{c} (ab, c, d) + (ac, b, d) + \\ (ad, b, d) + (dc, a, b) + (dc, a, b) + (db, c, a) \end{array} \right] \mathbf{e} \\ \left(\kappa_{bc}^{(ad)} \right)_{NL2} &= -w_l [(adc, b) + (adb, c) + (dc, ab) + (db, ac)]. \end{aligned}$$

4.2 Obtenção de $\sigma_{ij}^{(1)}$

Inicialmente, será obtida a soma dentro do parêntese de $\sigma_{ij}^{(1)}$, que é apresentada na equação (4.5), isto é, $2\kappa_{bc}^{(ad)} - \kappa_{bcd}^{(a)} + \kappa_{ac,bd}$, como esta soma tem uma expressão grande, será adotada a seguinte notação:

$$2\kappa_{bc}^{(ad)} - \kappa_{bcd}^{(a)} + \kappa_{ac,bd} = L1 + NL1 + NL2,$$

sendo:

$$\begin{aligned} L1 &: 2 \left(\kappa_{bc}^{(ad)} \right)_{L1} - \left(\kappa_{bcd}^{(a)} \right)_{L1} + (\kappa_{ac,bd})_{L1}; \\ NL1 &: 2 \left(\kappa_{bc}^{(ad)} \right)_{NL1} - \left(\kappa_{bcd}^{(a)} \right)_{NL1} - \left(\kappa_{bcd}^{(a)} \right)_{NL2} + (\kappa_{ac,bd})_{NL1}; \\ NL2 &: 2 \left(\kappa_{bc}^{(ad)} \right)_{NL2} - \left(\kappa_{bcd}^{(a)} \right)_{NL3} + (\kappa_{ac,bd})_{NL2}. \end{aligned}$$

4.2.1 Obtendo a parcela de $\sigma_{ij}^{(1)}$ devida a $L1$

Esta parcela será denotada por $\left(\sigma_{ij}^{(1)} \right)_{L1}$. Para obtê-la, basta substituir parcelas dos cumulantes em $L1$ e fazer as operações, isto é:

$$\begin{aligned} 2 \left(\kappa_{bc}^{(ad)} \right)_{L1} &= \left[\begin{array}{c} \frac{10V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} - \frac{4}{V} \left(\frac{\partial^2 \mu_l}{\partial \eta_l^2} \right)^2 - \frac{4}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} \\ - \frac{4(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 + \frac{2V^{(2)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 \end{array} \right] (a, b, c, d) \\ - \left(\kappa_{bcd}^{(a)} \right)_{L1} &= \left[\begin{array}{c} -\frac{9V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} + \frac{3}{V} \left(\frac{\partial^2 \mu_l}{\partial \eta_l^2} \right)^2 + \frac{3}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} \\ + \frac{4(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 - \frac{2V^{(2)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 \end{array} \right] (a, b, c, d) \\ (\kappa_{ac,bd})_{L1} &= \left[\begin{array}{c} \frac{9V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} + \frac{3}{V} \left(\frac{\partial^2 \mu_l}{\partial \eta_l^2} \right)^2 - \frac{3}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} - \\ \frac{4(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 + \frac{2V^{(2)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 \end{array} \right] (a, b, c, d), \end{aligned}$$

fazendo as operações, tem-se o valor de $L1$, que é dado por:

$$L1 = \phi \sum_l^n \left[-\frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} - \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} + \frac{(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 \right] (a, b, c, d).$$

O termo dentro do colchetes será denotado por h_l , portanto $L1$ pode se escrito como:

$$L1 = \phi \sum_l^n h_l(a, b, c, d),$$

sendo:

$$h_l = -\frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} - \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} + \frac{(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4.$$

A parcela $\left(\sigma_{ij}^{(1)} \right)_{L1}$ é, então,

$$\left(\sigma_{ij}^{(1)} \right)_{L1} = -\kappa^{ia} \kappa^{jb} \kappa^{cd} \phi \sum_l^n h_l(a, b, c, d),$$

para obtê-la em notação matricial é necessário reescrevê-la como:

$$\left(\sigma_{ij}^{(1)} \right)_{L1} = \phi \sum_l^n h_l \left(\sum_a -\kappa^{ia}(a) \right) \left(\sum_b \kappa^{jb}(b) \right) \left(\sum_{c,d} \kappa^{cd}(c, d) \right).$$

Por outro lado, utilizando a mesma notação da matriz de covariância de $O(n^{-2})$ dos MLG e fazendo os somatórios a matriz $\Sigma_{L1}^{(1)}$, cujos elementos são $\left(\sigma_{ij}^{(1)} \right)_{L1}$, é dada por:

$$\Sigma_{L1}^{(1)} = \frac{1}{\phi^2} P H Z_d P^T,$$

sendo:

$$\begin{aligned} P &= \left(\tilde{X}^T W \tilde{X} \right)^{-1} \tilde{X}^T, \\ Z &= \tilde{X} P = \tilde{X} \left(\tilde{X}^T W \tilde{X} \right)^{-1} \tilde{X}^T, \\ H &= \text{diag} \{ h_1, h_2, \dots, h_n \} \text{ e} \\ Z_d &= \text{diag} \{ z_{11}, z_{22}, \dots, z_{nn} \}. \end{aligned}$$

4.2.2 Obtendo a parcela de $\sigma_{ij}^{(1)}$ devida a $NL1$

Esta parcela será denotada por $\left(\sigma_{ij}^{(1)}\right)_{NL1}$ e sua obtenção é análoga a de $\left(\sigma_{ij}^{(1)}\right)_{L1}$, isto é, basta substituir parcelas dos cumulantes em $NL1$ e fazer as operações da seguinte maneira:

$$\begin{aligned} 2 \left(\kappa_{bc}^{(ad)}\right)_{NL1} &= -2(g_l + f_l) \left[\begin{array}{l} (ab, c, d) + (ac, b, d) + (ad, b, d) \\ + (dc, a, b) + (dc, a, b) + (db, c, a) \end{array} \right] \\ - \left(\kappa_{bcd}^{(a)}\right)_{NL1} &= f_l \left[\begin{array}{l} (ab, c, d) + (ac, b, d) \\ + (ad, b, d) + (dc, a, b) + (dc, a, b) + (db, c, a) \end{array} \right] \\ - \left(\kappa_{bcd}^{(a)}\right)_{NL2} &= g_l \left[\begin{array}{l} 2(ab, c, d) + 2(ac, b, d) \\ + 2(ad, b, d) + (dc, a, b) + (db, c, a) \end{array} \right] \\ \left(\kappa_{ac,bd}\right)_{NL1} &= g_l [(a, c, bd) + (ac, b, d)]. \end{aligned}$$

Pelo fato dos termos dentro do colchetes serem iguais quando se faz o somatório tem-se que $NL1$ é dado por:

$$NL1 = \phi \sum_l^n (-4f_l + g_l) (a, b, cd).$$

A parcela $\left(\sigma_{ij}^{(1)}\right)_{NL1}$ é, então,

$$\left(\sigma_{ij}^{(1)}\right)_{NL1} = -\kappa^{ia} \kappa^{jb} \kappa^{cd} \phi \sum_l^n (-4f_l + g_l) (ab, c, d),$$

para obter uma expressão em notação matricial, é feito a soma em a, b, c e d , obtendo-se:

$$\left(\sigma_{ij}^{(1)}\right)_{NL1} = \phi \sum_l^n \sum_l^n \left[\begin{array}{l} (-4f_l + g_l) (ab, c, d) (\sum_a -\kappa^{ia} (a)) \\ (\sum_b \kappa^{jb} (b)) (\sum_{c,d} \kappa^{cd} (cd)) \end{array} \right],$$

considerando a matriz $\Sigma_{NL1}^{(1)}$, cujos elementos são $\left(\sigma_{ij}^{(1)}\right)_{NL1}$, é dada por:

$$\Sigma_{NL1}^{(1)} = \frac{1}{\phi^2} P (-4F + G) CP^T,$$

sendo:

$$\begin{aligned}
F &= \text{diag} \left(\frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2} \right), \\
G &= \text{diag} \left(\frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2} - \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3 \right), \\
C &= \text{diag} (c_l), \text{ sendo } c_l = \text{tr} \left\{ \tilde{X}_l \left(\tilde{X}^T W \tilde{X} \right)^{-1} \right\}, \\
\left(\tilde{X}_l \right)_{p \times p} &= \text{cujo elemento da linha } r \text{ e coluna } s \text{ é } \left(\frac{\partial^2 \eta_l}{\partial \beta_r \partial \beta_s} \right).
\end{aligned}$$

4.2.3 Obtendo a parcela de $\sigma_{ij}^{(1)}$ devida a $NL2$

Esta parcela será denotada por $\left(\sigma_{ij}^{(1)} \right)_{NL2}$, como nas subseções anteriores, obtém-se $NL2$, isto é,

$$\begin{aligned}
2 \left(\kappa_{bc}^{(ad)} \right)_{NL2} &= -2w_l [(abd, c) + (acd, b)] - 2w_l [(ab, cd) + (bd, ac)], \\
- \left(\kappa_{bcd}^{(a)} \right)_{NL2} &= w_l [(abc, d) + (abd, c) + (acd, b)] + w_l [(bc, ad) + (bd, ac) + (cd, ab)] \text{ e} \\
\left(\kappa_{ac, bd} \right)_{NL1} &= w_l (ac, bd).
\end{aligned}$$

cujo resultado é:

$$NL2 = \phi \sum_l^n -w_l (adc, b).$$

Para obter uma expressão em notação matricial, é feito a soma em a, b, c e d , obtendo-se:

$$\left(\sigma_{ij}^{(1)} \right)_{NL2} = \phi \sum_l^n w_l \left(\sum_{a,c,d} -\kappa^{ia} \kappa^{cd} (acd) \right) \left(\sum_b \kappa^{jb} (b) \right),$$

considerando a matriz $\Sigma_{NL2}^{(1)}$, cujos elementos são $\left(\sigma_{ij}^{(1)} \right)_{NL2}$, é dada por:

$$\Sigma_{NL2}^{(1)} = -\frac{1}{\phi^2} PW (DER3) K^{-1},$$

sendo: $(DER3)_{n \times p} = \left[\left(\text{tr} [(D3)_1 K^{-1}] \right) \cdots \left(\text{tr} [(D3)_p K^{-1}] \right) \right];$

$[(D3)_a]_{p \times p}$ é a matriz cujo elemento da linha r e coluna s é: $\left(\frac{\partial^3 \eta_l}{\partial \beta_a \partial \beta_r \partial \beta_s} \right);$

$\text{tr} [(D3)_a K^{-1}]$ é o traço da matriz $[(D3)_a K^{-1}]$ para cada $a = 1, \dots, p$, avalia-se esta matriz nas n observações.

4.2.4 Matriz $\Sigma^{(1)}$

A matriz $\Sigma^{(1)}$ cujo elemento é $\sigma_{ij}^{(1)}$ é:

$$\Sigma^{(1)} = \Sigma_{L1}^{(1)} + \Sigma_{NL1}^{(1)} + \Sigma_{NL2}^{(1)},$$

substituindo os valores de $\Sigma_{L1}^{(1)}$, $\Sigma_{NL1}^{(1)}$ e $\Sigma_{NL2}^{(1)}$, na equação acima tem-se:

$$\Sigma^{(1)} = \frac{1}{\phi^2} P H Z_d P^T + P (-4F + G) C P^T - P W (DER3) K^{-1}.$$

4.3 Obtenção de $\sigma_{ij}^{(2)}$

Este elemento da matriz de covariância de $O(n^{-2})$ é obtido de modo análogo ao elemento $\sigma_{ij}^{(1)}$ da seção anterior. Os cumulantes necessários foram apresentados no início deste capítulo.

4.3.1 Obtendo $\kappa_{rst} \left(\frac{3}{2} \kappa_{abc} + 5 \kappa_{a,bc} \right)$

A soma dentro do parêntese de $\sigma_{ij}^{(2)}$, que é apresentada na equação (4.6), será obtida a partir de $\kappa_{rst} \left(\frac{3}{2} \kappa_{abc} + 5 \kappa_{a,bc} \right)$, isto é:

$$\begin{aligned} \frac{3}{2} \kappa_{abc} + 5 \kappa_{a,bc} &= \phi \sum_l^n \left(-\frac{3}{2} f_l + 2g_l \right) (a, b, c) \\ &+ w_l \left[-\frac{3}{2} (ab, c) - \frac{3}{2} (ac, b) + \frac{7}{2} (bc, a) \right]. \end{aligned}$$

Por outro lado tem-se que:

$$\kappa_{rst} = \phi \sum_m^n (-2f_m - 2gm) (r, s, t) - w_m [(rs, t) + (rt, s) + (ts, r)],$$

fazendo então o produto:

$$\begin{aligned}
\kappa_{rst} \left(\frac{3}{2} \kappa_{abc} + 5 \kappa_{a,bc} \right) &= \phi^2 \sum_{l,m}^n \left(-\frac{3}{2} f_l + 2g_l \right) (-f_m - 2g_m) (a, b, c, r, s, t) \\
&- \left(-\frac{3}{2} f_l + 2g_l \right) w_m \left[\begin{array}{c} (a, b, c, rs, t) \\ + (a, b, c, rt, s) + (a, b, c, ts, r) \end{array} \right] \\
&+ (w_l f_m - 2w_l g_m) \left[\begin{array}{c} \frac{3}{2} (r, s, t, ab, c) \\ + \frac{3}{2} (r, s, t, ac, b) - \frac{7}{2} (r, s, t, bc, a) \end{array} \right] \\
&+ w_l w_m \left[\begin{array}{c} \frac{3}{2} (ab, c, rs, t) + \frac{3}{2} (ab, c, rt, s) \\ + \frac{3}{2} (ab, c, rs, t) + \frac{3}{2} (ac, b, ts, r) \\ - \frac{7}{2} (bc, a, rs, t) - \frac{7}{2} (bc, a, rt, s) - \frac{7}{2} (bc, a, ts, e) \end{array} \right].
\end{aligned}$$

4.3.2 Obtendo $3\kappa_{ab,c}\kappa_{r,st}$

Agora será obtido $3\kappa_{ab,c}\kappa_{r,st}$, um dos termos da equação (4.6), que é dado por:

$$3\kappa_{ab,c}\kappa_{r,st} = \phi^2 \sum_{l,m}^n 3g_l g_m (a, b, c, r, s, t) + (6w_m g_l + 3w_l w_m) (ab, rs, c, t).$$

Para obter o elemento entre parêntese de $\sigma_{ij}^{(2)}$, basta somar $\kappa_{rst} \left(\frac{3}{2} \kappa_{abc} + 5 \kappa_{a,bc} \right)$ com $3\kappa_{ab,c}\kappa_{r,st}$, que simplificando uma vez que a, b e c variam de $1, \dots, p$, tornando os termos das derivadas de η com relação a β , iguais. Tem-se a seguinte equação:

$$\begin{aligned}
\kappa_{rst} \left(\frac{3}{2} \kappa_{abc} + 5 \kappa_{a,bc} \right) + 3\kappa_{ab,c}\kappa_{r,st} &= \phi^2 \sum_{l,m}^n \left(\frac{3}{2} f_l f_m - f_l g_m - g_l g_m \right) (a, b, c, r, s, t) \\
&+ 5w_l f_m (ab, c, rs, t) + w_l g_m (ab, r, c, st) \\
&+ \frac{9}{2} w_l w_m (ct, rs, a, b).
\end{aligned}$$

Para obter o elemento entre parêntese de $\sigma_{ij}^{(2)}$, basta somar $\kappa_{rst} \left(\frac{3}{2} \kappa_{abc} + 5 \kappa_{a,bc} \right)$ com $3\kappa_{ab,c}\kappa_{r,st}$, que simplificando, uma vez que a, b e c variam de $1, \dots, p$, tornando os termos das derivadas de η com relação a β iguais. Tem-se a seguinte equação:

$$\begin{aligned}
\sigma_{ij}^{(2)} &= \kappa^{ia} \kappa^{jr} \kappa^{bs} \kappa^{ct} \phi^2 \sum_{l,m}^n \left(\frac{3}{2} f_l f_m - f_l g_m - g_l g_m \right) (a, b, c, r, s, t) \\
&- w_l f_m (r, s, t, bc, a) + 2w_l g_m (r, s, t, bc, a) \\
&+ w_m \left(\frac{3}{2} f_l - 3g_l \right) (a, b, c, rs, t).
\end{aligned}$$

Considere:

$$\sigma_{ij}^{(21)} = \kappa^{ia} \kappa^{jr} \kappa^{bs} \kappa^{ct} \phi^2 \sum_{l,m}^n \left(\frac{3}{2} f_l f_m - f_l g_m - g_l g_m \right) (a, b, c, r, s, t);$$

$$\sigma_{ij}^{(22)} = \kappa^{ia} \kappa^{jr} \kappa^{bs} \kappa^{ct} \phi^2 \sum_{l,m}^n 5w_l f_m (ab, c, rs, t) + w_l g_m (ab, r, c, st);$$

$$\sigma_{ij}^{(23)} = \kappa^{ia} \kappa^{jr} \kappa^{bs} \kappa^{ct} \phi^2 \sum_{l,m}^n w_m \left(\frac{3}{2} f_l - 3g_l \right) (a, b, c, rs, t).$$

Assim, $\sigma_{ij}^{(2)}$ pode ser escrito como:

$$\sigma_{ij}^{(2)} = \sigma_{ij}^{(21)} + \sigma_{ij}^{(22)} + \sigma_{ij}^{(23)}.$$

4.3.3 Matriz $\Sigma^{(21)}$

A matriz $\Sigma^{(21)}$ tem como elemento $\sigma_{ij}^{(21)}$. Para obtê-la, basta agrupar convenientemente os termos de $\sigma_{ij}^{(21)}$, isto é,

$$\sigma_{ij}^{(21)} = \phi^2 \sum_{l,m}^n \left[\begin{array}{c} \left(\frac{3}{2} f_l f_m - f_l g_m - g_l g_m \right) \left(\sum_a \kappa^{ia} (a) \right) \left(\sum_r \kappa^{jr} (r) \right) \\ \left(\sum_{b,s} \kappa^{bs} (b, s) \right) \left(\sum_{c,t} \kappa^{ct} (c, t) \right) \end{array} \right].$$

Nesta equação, os termos $\sum_a \kappa^{ia} (a)$, $\sum_r \kappa^{jr} (r)$, $\sum_{b,s} \kappa^{bs} (b, s)$ e $\sum_{b,s} \kappa^{bs} (b, s)$ podem ser escritos em notação matricial como segue, sendo o vetor $\rho_i^T = \left[0 \ \cdots \ 0 \ \cdots \ 1 \ \cdots \ 0 \right]^T$:

- $\sum_a \kappa^{ia} (a) = -\rho_i^T \frac{K^{-1}}{\phi} \tilde{X}_l$;
- $\sum_r \kappa^{jr} (r) = -\rho_j \frac{K^{-1}}{\phi} \tilde{X}_l$;
- $\sum_{b,s} \kappa^{bs} (b, s) = -\rho_j \tilde{X}_l \frac{K^{-1}}{\phi} \tilde{X}_l^T$;
- $\sum_{b,s} \kappa^{bs} (b, s) = -\rho_j \tilde{X}_l \frac{K^{-1}}{\phi} \tilde{X}_l^T$;

Assim, $\Sigma^{(21)}$ é dada por:

$$\Sigma^{(21)} = P \left(\frac{3}{2} FF + GF - GG \right) Z^{(2)} P^T,$$

sendo $Z^{(2)} = Z \odot Z$ o operador \odot significa o produto de Hadamard.

4.3.4 Matriz $\Sigma^{(22)}$ e $\Sigma^{(23)}$

A matriz $\Sigma^{(22)}$ tem como elemento $\sigma_{ij}^{(22)}$. Para obtê-la, basta agrupar convenientemente os termos de $\sigma_{ij}^{(22)}$, isto é,

$$\sigma_{ij}^{(22)} = \phi^2 \sum_{l,m}^n \left[\begin{array}{c} (w_l (-f_m + 2g_m)) (\sum_a \kappa^{ia} (a)) (\sum_r \kappa^{jr} (r)) \\ \left(\sum_{b,s} \kappa^{bs} (bs) \right) \left(\sum_{c,t} \kappa^{ct} (c, t) \right) \end{array} \right].$$

Nesta equação, os termos $\sum_a \kappa^{ia} (a)$, $\sum_r \kappa^{jr} (r)$, $\sum_{b,s} \kappa^{bs} (bs)$ foram apresentados na seção anterior. O termo $\sum_{b,s} \kappa^{bs} (bs)$ pode ser escritos em notação matricial como segue:

- $\sum_a \kappa^{bs} (bs) = -\rho_j \tilde{X}_l \frac{K^{-1}}{\phi} C_l.$

Assim, a expressão da matriz $\Sigma^{(22)}$ é dada por:

$$\Sigma^{(22)} = PWC (-F + 2G) P^T.$$

O elemento $\sigma_{ij}^{(23)}$ difere do $\sigma_{ij}^{(22)}$ pela parcela $w_l (-f_m + 2g_m)$ que em $\sigma_{ij}^{(23)}$ é $w_m (\frac{3}{2}f_l - 3g_l)$. Assim, a expressão de $\Sigma^{(23)}$ é dada por:

$$\Sigma^{(23)} = PWC \left(\frac{3}{2}F - 3G \right) P^T.$$

Somando $\Sigma^{(21)}$, $\Sigma^{(22)}$ e $\Sigma^{(23)}$ obtém-se $\Sigma^{(2)}$, cuja expressão é dada por:

$$\begin{aligned} \Sigma^{(2)} &= \Sigma^{(21)} + \Sigma^{(22)} + \Sigma^{(23)} \\ &= \frac{1}{\phi^2} \left[P \left(\frac{3}{2}FF + GF - GG \right) Z^{(2)} P^T. + PWC \left(\frac{7}{2}F - 2G \right) P^T \right]. \end{aligned}$$

4.4 Obtenção de $\sigma_{ij}^{(3)}$

Este elemento da matriz de covariância de $O(n^{-2})$ será obtido a partir da expressão (4.7), e os cumulantes necessários foram apresentados no início deste capítulo.

4.4.1 Obtendo $(\kappa_{a,bc} + \kappa_{abc})$ e $(2\kappa_{r,st} + \kappa_{rst})$

Para isto, basta somar os cumulantes apresentados no início deste capítulo tem-se então que:

$$\kappa_{a,bc} + \kappa_{abc} = \phi \sum_l^n (-f_l - 2g_l) (a, b, c) - w_l [(ab, c) + (ac, b)],$$

e

$$2\kappa_{r,st} + \kappa_{rst} = \phi \sum_m^n -f_m(r, s, t) - w_m(rs, t)$$

4.4.2 Obtendo $(\kappa_{a,bc} + \kappa_{abc})(2\kappa_{r,st} + \kappa_{rst})$

As parcelas deste produto são dadas por:

$$\mathbf{P1} = \phi^2 \sum_{l,m}^n (f_l f_m + g_l g_m)(a, b, c, r, st);$$

$$\mathbf{P2} = \phi^2 \sum_{l,m}^n (f_l w_m + g_l w_m)(a, b, c, rs, t);$$

$$\mathbf{P3} = \phi^2 \sum_{l,m}^n (f_m w_l + g_l w_m)[(ab, c, r, s, t) + (ac, b, r, s, t)];$$

$$\mathbf{P4} = \phi^2 \sum_{l,m}^n w_m w_l [(ab, c, rs, t) + (ac, b, rs, t)].$$

4.4.3 Obtendo $\sigma_{ij}^{(31)}, \sigma_{ij}^{(32)}, \sigma_{ij}^{(33)}$ e $\sigma_{ij}^{(34)}$

O elemento $\sigma_{ij}^{(3)}$ da matriz $\Sigma^{(3)}$ é dado pela soma de $\sigma_{ij}^{(31)}, \sigma_{ij}^{(32)}, \sigma_{ij}^{(33)}$ e $\sigma_{ij}^{(34)}$ sendo:

$$\sigma_{ij}^{(31)} = \phi^2 \kappa^{ia} \kappa^{jb} \kappa^{rs} \kappa^{ct} \sum_{l,m}^n (f_l f_m + g_l g_m)(a, b, c, r, st);$$

$$\sigma_{ij}^{(32)} = \phi^2 \kappa^{ia} \kappa^{jb} \kappa^{rs} \kappa^{ct} \phi^2 \sum_{l,m}^n (f_l w_m + g_l w_m)(a, b, c, rs, t);$$

$$\sigma_{ij}^{(33)} = \phi^2 \kappa^{ia} \kappa^{jb} \kappa^{rs} \kappa^{ct} \phi^2 \sum_{l,m}^n (f_m w_l + g_l w_m)[(ab, c, r, s, t) + (ac, b, r, s, t)];$$

$$\sigma_{ij}^{(34)} = \phi^2 \kappa^{ia} \kappa^{jb} \kappa^{rs} \kappa^{ct} \phi^2 \sum_{l,m}^n w_m w_l [(ab, c, rs, t) + (ac, b, rs, t)].$$

A próxima etapa é escrever cada uma dessas equações na forma matricial, que, após fazer os somatórios, tem-se:

$$\Sigma^{(31)} = \frac{1}{\phi^2} P (F + G) Z Z_d F P^T;$$

$$\Sigma^{(32)} = \frac{1}{\phi^2} P (F + G) W C P^T;$$

$$\Sigma^{(33)} = \frac{2}{\phi^2} P W Z C F P^T;$$

$$\Sigma^{(34)} = \frac{2}{\phi^2} P W W_c P^T.$$

Somando $\Sigma^{(31)}$, $\Sigma^{(32)}$, $\Sigma^{(33)}$ e $\Sigma^{(34)}$ tem-se a terceira parcela da matriz de covariância de $O(n^{-2})$, isto é,

$$\begin{aligned}\Sigma^{(3)} &= \frac{1}{\phi^2} P(F+G)ZZ_dFP^T + \frac{1}{\phi^2} P(F+G)WCP^T \\ &\quad + \frac{2}{\phi^2} PWZCFP^T + \frac{1}{\phi^2} P(F+G)WCP^T\end{aligned}$$

4.5 Matriz de covariância de $O(n^{-2})$ dos EMV nos MNLFE

Essa matriz será denotada por $\Sigma = \Sigma^{(1)} + \Sigma^{(2)} + \Sigma^{(3)}$ e depende explicitamente da matriz \tilde{X} , do parâmetro de precisão ϕ e da média desconhecida. É possível simplificar essa matriz quando o modelo adotado possui uma expressão para Z . Para alguns MNLFE, o termo de $O(n^{-2})$ pode ser relativamente grande devendo então ser considerado nas inferências.

A matriz de covariância de primeira ordem dos MNLFE será denotada por $Cov_1(\hat{\beta})$ e é dada por:

$$Cov_1(\hat{\beta}) = \phi^{-1} (\tilde{X}^T W \tilde{X})^{-1}.$$

Assim, a matriz de covariância destas estimativas de $O(n^{-2})$, que é denotada por $Cov_2(\hat{\beta})$, é:

$$Cov_2(\hat{\beta}) = \phi^{-1} (\tilde{X}^T W \tilde{X})^{-1} + \Sigma.$$

4.6 Caso especial

Como já discutido, os MLG é um caso particular dos MNLFE, ou seja, que a matriz de covariância apresentada em Cordeiro (2004) passa a ser um caso especial dos aqui apresentados.

Observe que para o MLG as matrizes da parcela $\Sigma^{(1)}$, $\Sigma^{(2)}$ e $\Sigma^{(3)}$ são:

1. $P(-4F+G)CP^T = 0$;
2. $K^{-1}diag(\sum_a^p tr[(D3)_a K^{-1}])P^T = 0$;
3. $\tilde{X} = X$;
4. $PWC(\frac{7}{2}F - 2G)P^T = 0$;
5. $P(F+G)WCP^T = 0$;
6. $PWZCFP^T = 0$;
7. $P(F+G)WCP^T = 0$.

Portanto, as expressões de $\Sigma^{(1)} = \frac{1}{\phi^2} P H Z_d P^T$, $\Sigma^{(2)} = \frac{1}{\phi^2} P \left(\frac{3}{2} F F + G F - G G \right) Z^{(2)} P^T$ e $\Sigma^{(3)} = \frac{1}{\phi^2} P (F + G) Z Z_d F P^T$, que corroboram com a expressão obtida por Cordeiro (2004).

Capítulo 5

Estudo de simulação

A simulação via Monte Carlo (MC) é uma ferramenta indispensável quando o propósito é avaliar um estimador, uma vez que, pela lei dos grandes números, é possível, na análise das amostras simuladas, se forem tomadas amostras suficientes, em média, o estimador está próximo do verdadeiro valor do parâmetro e, além disso, pelo teorema do limite central, a distribuição da média do estimador nas amostras simuladas é normal.

Neste capítulo, é apresentado um estudo de simulação via MC para avaliar a matriz de covariância de $O(n^{-2})$.

5.1 Modelo simulado

A componente aleatória do MNLFE \mathbf{Y} , com distribuição de Poisson com média $\boldsymbol{\mu}$, isto é,

$$Y_l \sim P(\mu_l), \text{ para } l = 1, \dots, n,$$

sendo n o tamanho da amostra simulada.

A distribuição de Poisson é geralmente utilizada na análise de dados em forma de contagem, e é satisfatória para descrever dados experimentais em que a média é proporcional à variância.

Este modelo tem um papel importante na análise de dados categorizados dispostos em uma tabela de contingência. As variáveis aleatórias neste caso tem distribuição multinomial, mas pode-se demonstrar, segundo Cordeiro (1986, p.8), que isto equivale a um conjunto de distribuições de Poisson com a condição do total de frequência observada ser fixado.

Neste estudo, foram simuladas 10.000 amostras de tamanhos $n = 20$, $n = 40$ e $n = 60$.

A matriz modelo X foi simulada de uma distribuição uniforme $[0, 1]$ e fixada para as 10.000 amostras de igual tamanho.

O preditor não-linear η_l é dado por:

$$\eta_l = \mu_l = \exp(\beta_1 X_{1l} + \beta_2 X_{2l} + \beta_3 X_3),$$

sendo: $\beta_1 = 3$, $\beta_2 = 4$, $\beta_3 = 5$.

5.2 Algoritmo

5.2.1 Algoritmo para obtenção das estimativas de $O(n^{-1})$ e $O(n^{-2})$

Os passos a seguir descrevem como foi elaborado o programa do SAS para modelar a variável resposta \mathbf{Y} :

5.2.2 1º Passo: Obter as estimativas de $O(n^{-1})$

A *procedure* NLMIXED fornece estas estimativas, maximizando a log-verossimilhança, diretamente, isto é, obtendo as raízes do sistema de equações da função escore pelo método de Newton Raphson.

Equações necessárias na *procedure* NLMIXED

i) log-verossimilhança, de uma Poisson, com parâmetros $\boldsymbol{\mu}$ e $\phi = 1$, que é dada por:

$$l(\boldsymbol{\mu}, \phi; \mathbf{y}) = \mathbf{y} \log(\boldsymbol{\mu}) - \boldsymbol{\mu};$$

ii) preditor não-linear:

$$\eta_l = \mu_l = \exp(\beta_1 x_{1l} + \beta_2 x_{2l} + \beta_3 x_{3l}),$$

Tentativa inicial,

iii) foi utilizada a *procedure* NLIM, para obter a tentativa inicial, pois a convergência do método de Newton Raphson depende dos valores iniciais.

5.2.3 2º Passo: Resultados da *procedure* NLMIXED

- As estimativas de ordem $O(n^{-1})$ dos parâmetros do modelo e da matriz de covariância, são disponibilizados em arquivos. Estes arquivos são utilizados na *procedure* IML, para obter as estimativas de $O(n^{-2})$.

5.2.4 3º Passo: *procedure* IML

Esta *procedure* permite ao usuário operar com matrizes. Neste passo, as matrizes necessárias para as estimativas de $O(n^{-2})$ são calculadas, a seguir serão enumeradas as matrizes necessárias:

Matriz das derivadas primeiras (\tilde{X}), cada elemento desta matriz é um vetor de n elementos

$$\tilde{X} = \left[\frac{\partial \boldsymbol{\eta}}{\partial \beta_1}, \frac{\partial \boldsymbol{\eta}}{\partial \beta_2}, \frac{\partial \boldsymbol{\eta}}{\partial \beta_3} \right], \text{sendo} \left(\frac{\partial \boldsymbol{\eta}}{\partial \beta_r} \right)^T = \left[\frac{\partial \eta_1}{\partial \beta_r} \quad \frac{\partial \eta_2}{\partial \beta_r} \quad \cdots \quad \frac{\partial \eta_{s00}}{\partial \beta_r} \right].$$

Neste problema $\tilde{X} = [\mathbf{x}_1 \boldsymbol{\eta}, \mathbf{x}_2 \boldsymbol{\eta}, \mathbf{x}_3 \boldsymbol{\eta}]$, com dimensão de $(n \times 3)$

Matrizes das derivadas segundas, \tilde{X}_l , cuja dimensão para cada l é (3×3)

$$\tilde{X}_l = \begin{bmatrix} \frac{\partial^2 \boldsymbol{\eta}_l}{\partial \beta_1^2} & \frac{\partial^2 \boldsymbol{\eta}_l}{\partial \beta_1 \partial \beta_2} & \frac{\partial^2 \boldsymbol{\eta}_l}{\partial \beta_1 \partial \beta_3} \\ & \frac{\partial^2 \boldsymbol{\eta}_l}{\partial \beta_2^2} & \frac{\partial^2 \boldsymbol{\eta}_l}{\partial \beta_2 \partial \beta_3} \\ & & \frac{\partial^2 \boldsymbol{\eta}_l}{\partial \beta_3^2} \end{bmatrix} = \begin{bmatrix} \eta_l & \eta_l X_{1l} & \eta_l X_{2l} \\ & \eta_l X_{1l}^2 & \eta_l X_{1l} X_{2l} \\ & & \eta_l X_{2l}^2 \end{bmatrix}, \text{ para } l = 1, \dots, n$$

Traço da matriz de covariância, que será denotada por C , que nada mais é que o traçado produto entre a matriz inversa de Fisher (KDEBETA) e a matriz das segundas derivadas, isto é, $C = tr(KDEBETA * \tilde{X}_l)$

$V(\mu_l) = \mu_l$, para a distribuição Poisson;

$$V^{(1)}(\mu_l) = \frac{\partial V(\mu_l)}{\partial \mu_l} = 1;$$

Matriz diagonal peso, W , cujos elementos da diagonal são:

$$w_{ll} = \frac{1}{V(\mu_l)} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 = \frac{1}{\mu_l};$$

Matriz diagonal, F , cujo elementos diagonais desta matriz são:

$$f_{ll} = \frac{1}{V(\mu_l)} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2} = 0;$$

Matriz diagonal, G , cujos elementos diagonais destas matrizes são:

$$g_{ll} = \frac{1}{V(\mu_l)} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2} - \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3 = 0 + \frac{1}{\mu_l^2};$$

Matriz diagonal, H , cujos elementos diagonais destas matrizes são:

$$h_{ll} = -\frac{1}{V(\mu_l)} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} - \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} + \frac{(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 = \frac{1}{\mu_l};$$

Uma das matrizes mais trabalhosa na obtenção da matriz de covariância de $O(n^{-2})$ é a matriz, $(DER3)$. Para obtê-la, deve-se calcular suas colunas, que neste caso são:

$$(D3)_{a=1} = \begin{bmatrix} \frac{\partial^3 \eta}{\partial \beta_1^3} & \frac{\partial^3 \eta}{\partial \beta_1^2 \partial \beta_2} & \frac{\partial^3 \eta}{\partial \beta_1^2 \partial \beta_3} \\ \frac{\partial^3 \eta}{\partial \beta_1 \partial \beta_2^2} & \frac{\partial^3 \eta}{\partial \beta_1 \partial \beta_3^2} & \frac{\partial^3 \eta}{\partial \beta_1 \partial \beta_3^2} \end{bmatrix} = \begin{bmatrix} \eta x_1^3 & x_1^2 x_2 \eta & x_1^2 x_3 \eta \\ & x_1 x_2^2 \eta & x_1 x_3^2 \eta \\ & & x_1 x_3^2 \eta \end{bmatrix},$$

$$(D3)_{a=2} = \begin{bmatrix} \frac{\partial^3 \eta}{\partial \beta_1^2 \partial \beta_2} & \frac{\partial^3 \eta}{\partial \beta_2^2 \partial \beta_1} & \frac{\partial^3 \eta}{\partial \beta_1 \partial \beta_2 \partial \beta_3} \\ \frac{\partial^3 \eta}{\partial \beta_2^3} & \frac{\partial^3 \eta}{\partial \beta_2 \partial \beta_1 \partial \beta_3} & \frac{\partial^3 \eta}{\partial \beta_2 \partial \beta_3^2} \end{bmatrix} = \begin{bmatrix} \eta x_1^2 x_2 & x_1 x_2^2 \eta & x_1 x_2 x_3 \eta \\ & x_2^3 \eta & x_1 x_2 x_3 \eta \\ & & x_2 x_3^2 \eta \end{bmatrix}$$

e, finalmente,

$$(D3)_{a=3} = \begin{bmatrix} \frac{\partial^3 \eta}{\partial \beta_1^2 \partial \beta_3} & \frac{\partial^3 \eta}{\partial \beta_3 \partial \beta_1 \partial \beta_2} & \frac{\partial^3 \eta}{\partial \beta_3^2 \partial \beta_1} \\ \frac{\partial^3 \eta}{\partial \beta_2^2 \partial \beta_3} & \frac{\partial^3 \eta}{\partial \beta_1 \partial \beta_3^2} & \frac{\partial^3 \eta}{\partial \beta_3^3} \end{bmatrix} = \begin{bmatrix} \eta x_1^2 x_3 & x_1 x_2 x_3 \eta & x_3^2 x_1 \eta \\ & x_2^2 x_3 \eta & x_1 x_3^2 \eta \\ & & x_2^3 \eta \end{bmatrix}$$

Portanto, coluna a da $(DER3)_a = (tr [(D3)_a K^{-1}])$ é dada por:

$$\begin{bmatrix} (DER3)_{1,l=1} & (DER3)_{2,l=1} & (DER3)_{3,l=1} \\ (DER3)_{1,l=2} & (DER3)_{2,l=2} & (DER3)_{1,l=2} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ (DER3)_{1,l=n} & (DER3)_{2,l=n} & (DER3)_{3,l=n} \end{bmatrix}.$$

Observação: a notação $(DER3)_{a,l=i}$, significa que a coluna a de $(DER3)$ é avaliada para a observação i .

5.3 Resultados

Na Tabela 5.1 são apresentados os resultados obtidos para a amostra de tamanho $n = 20$. A primeira entrada desta tabela são as covariâncias de $O(n^{-1})$ e $O(n^{-2})$, avaliadas nos parâmetros. Na segunda encontram-se as médias das covariâncias de $O(n^{-1})$ e $O(n^{-2})$ para as 10.000 amostras e na linha 3 é apresentada a covariância amostral. Observe que as variâncias da matriz de $O(n^{-2})$ são sempre positivas e a quase todas as covariâncias estão mais próximas da covariância amostral do que as de $O(n^{-1})$.

Tabela 5.1: Matriz de covariância de $\hat{\beta}$ resultados da simulação para $n = 20$.

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
	0,000599 (0,000601)	-0,000373 (-0,000378)	-0,000207 (-0,000206)
$\hat{\beta}_1$	0,000599 (0,000601)	-0,000373 (-0,000378)	-0,000207 (-0,000206)
	0,000591	-0,000382	-0,000193
		0,000853 (0,000884)	-0,000386 (-0,000406)
$\hat{\beta}_2$		0,000853(0,000888)	-0,000386 (-0,000406)
		0,000877	-0,000397
			0,000517 (0,000523)
$\hat{\beta}_3$			0,000517 (0,000523)
			0,000517

Na Tabela 5.2 são apresentados os resultados obtidos para a amostra de tamanho $n = 40$ as entradas são as mesmas da tabela 5.1 e para este tamanho de amostra não há muita diferença entre as matrizes de covariâncias de $O(n^{-1})$ e $O(n^{-2})$.

Tabela 5.2: Matriz de covariância de $\hat{\beta}$ resultados da simulação para $n = 40$.

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
	0,000180 (0,000184)	-0,000029 (-0,000027)	-0,000121 (-0,000124)
$\hat{\beta}_1$	0,000180 (0,000183)	-0,000029 (-0,000027)	-0,000121 (-0,000124)
	0,000179	-0,000028	-0,000120
		0,000167 (0,000169)	-0,000110 (-0,000112)
$\hat{\beta}_2$		0,000167(0,000169)	-0,000110 (-0,000112)
		0,000167	-0,000101
			0,000198 (0,000201)
$\hat{\beta}_3$			0,000198 (0,000201)
			0,000199

Na Tabela 5.3, são apresentados os resultados para $n = 60$. Como para a amostra de tamanho 40 não houve grande diferença entres os elementos das três matrizes.

Tabela 5.3: Matriz de covariância de $\hat{\beta}$ resultados da simulação para $n = 60$.

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
$\hat{\beta}_1$	0,000904 (0,000091) 0,000090 (0,000091) 0,000090	-0,000040 (-0,000040) -0,000040 (-0,000040) -0,000040	-0,000037 (-0,000037) -0,000037 (-0,000037) -0,000038
$\hat{\beta}_2$		0,000110 (0,000110) 0,000110(0,000110) 0,000110	-0,000062 (-0,000063) -0,000062 (-0,000063) -0,000063
$\hat{\beta}_3$			0,0000923 (0,000928) 0,000092 (0,000093) 0,000093

Observando as tabelas acima, este estudo de simulação mostra que a covariância de $O(n^{-2})$ é indispensável para amostras de tamanho pequeno.

Capítulo 6

Modelando grau de refino das fibras de celulose e avaliando os estimadores de $O(n^{-2})$ para os MNLFE

Como ilustração e para mostrar que os resultados obtidos são de fácil implementação em problemas na área de produção, a metodologia apresentada será aplicada a um conjunto de dados provenientes do processo de fabricação de papel. O interesse é mostrar que para amostras pequenas os estimadores de $O(n^{-2})$ são indispensáveis.

O Grau de Refino (°SR) das fibras de celulose (grau de drenabilidade da polpa celulósica) é considerado o mais importante componente do processo, pois dá ao papel parte preponderante de suas características finais. O refino é um tratamento mecânico dado às fibras em suspensão com o intuito de modificar sua estrutura e melhorar as características das fibras para a produção do papel. Estas modificações são irreversíveis e as propriedades mais afetadas são a resistência à tração, resistência à tensão e à opacidade do papel.

O procedimento foi aplicado a um conjunto de dados de operação das variáveis que teoricamente teriam influência sobre o valor do Grau de Refino (°SR) das fibras de celulose, variável resposta. Aqui, serão consideradas as covariáveis: carga do refinador e condutividade elétrica da solução.

Neste problema, o objetivo é encontrar um modelo que possa prever em amostras pequenas o grau de refino e, com isto, permitir a tomada de decisão em tempo real.

Delimitação amostral

O delineamento inteiramente casualizado foi adotado, para obter uma amostra de 800 observações, que foram coletados na Klabin Papéis Monte Alegre durante o ano de 2003, e foram registradas durante o processo de produção por meio de um aparelho que mede as covariáveis e a variável resposta.

Variáveis de interesse

Variável resposta Grau de Refino ($^{\circ}\text{SR}$) das fibras de celulose (grau de drenabilidade da polpa celulósica);

Covariáveis carga do refino (x_1) e condutividade elétrica da solução (x_2).

Análise de dados

Inicialmente, avaliou-se a distribuição da variável resposta, por meio de uma análise exploratória dos dados e foram consideradas as distribuições gama e normal.

A não-linearidade foi detectada, através de diagrama de dispersão, sendo esta não muito acentuada.

O objetivo do experimento é o de usar o modelo obtido para previsões em tempo real. Portanto em observar uma amostra pequena e decidir se a qualidade do papel é satisfatória. Retirou-se da amostra observada uma amostra aleatória de tamanho 30 para avaliar o modelo e o desempenho dos estimadores de $O(n^{-2})$ em amostras pequenas.

6.1 Componentes do MNLFE

A componente aleatória o grau do refino das fibras de celulose ($^{\circ}\text{SR}$), denotada por \mathbf{Y} , com função densidade de probabilidade distribuição Gama(μ_l, ϕ), sendo μ_l a média de Y_l e ϕ o parâmetro de dispersão desconhecido. Para $l = 1, 2, \dots, 800$.

A componente sistemática é:

$$\eta_l = \mu_l = \exp(\beta_1 + \beta_2 x_{1l} + \beta_3 x_{2l}),$$

sendo:

x_{1l} \rightarrow a covariável carga do refino;

x_{2l} \rightarrow a covariável condutividade elétrica;

$\beta_1, \beta_2, \beta_3$ e ϕ , \rightarrow os parâmetros do modelo.

6.2 Algoritmo para obtenção das estimativas de $O(n^{-1})$ e $O(n^{-2})$

Os passos a seguir descrevem como foi elaborado o programa do SAS para modelar a variável resposta \mathbf{Y} : Grau de refino.

1º Passo: Obter as estimativas de $O(n^{-1})$ A *procedure* NLMIXED calcula estas estimativas, maximizando a log-verossimilhança diretamente, isto é, obtendo as raízes do sistema de equações da função escore pelo método de Newton Raphson.

2º Passo: Equações necessárias na *procedure* NLMIXED

- log-verossimilhança, que neste problema foram:

Normal:

$$l(\boldsymbol{\mu}, \phi; \mathbf{y}) = -\frac{1}{2} \log \left(\frac{1}{\phi} \right) - \frac{\phi}{2} (\mathbf{y} - \boldsymbol{\mu});$$

Gama:

$$l(\boldsymbol{\mu}, \phi; \mathbf{y}) = -\log(\Gamma(\phi)) + \phi \log \left(\frac{\mathbf{y}\phi}{\boldsymbol{\mu}} \right) - \frac{\mathbf{y}\phi}{\boldsymbol{\mu}}.$$

- equação que descreve a média

$$\eta_i = \mu_i = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})$$

3º Passo: Tentativa inicial

- foi utilizada a *procedure* NLIM para obter a tentativa inicial, pois a convergência do método de Newton Raphson depende dos valores iniciais.

4º Passo: Resultados da *procedure* NLMIXED

- As estimativas de ordem $O(n^{-1})$ dos parâmetros do modelo e da matriz de covariância são disponibilizados em arquivos. Estes arquivos são utilizados na *procedure* IML para obter as estimativas de $O(n^{-2})$.

5º Passo: proc IML Esta *procedure* permite ao usuário operar com matrizes, neste passo as matrizes necessárias para as estimativas de $O(n^{-2})$ são calculadas. A seguir serão enumeradas as matrizes necessárias:

- Para as estimativas de $O(n^{-2})$ dos parâmetros β' s

Matriz das derivadas primeiras (\tilde{X}) cada elemento desta matriz é um vetor de n elementos

$$\tilde{X} = \left[\frac{\partial \boldsymbol{\eta}}{\partial \beta_1}, \frac{\partial \boldsymbol{\eta}}{\partial \beta_2}, \frac{\partial \boldsymbol{\eta}}{\partial \beta_3} \right], \text{sendo } \left(\frac{\partial \boldsymbol{\eta}}{\partial \beta_1} \right)^T = \left[\frac{\partial \eta_1}{\partial \beta_1} \quad \frac{\partial \eta_2}{\partial \beta_1} \quad \dots \quad \frac{\partial \eta_{800}}{\partial \beta_1} \right]. \text{ Neste problema } \tilde{X} = [\boldsymbol{\eta}, \mathbf{X}_1 \boldsymbol{\eta}, \mathbf{X}_2 \boldsymbol{\eta}],$$

com dimensão de $(n \times 3)$, sendo $n = 30$ e $n = 800$.

Matrizes das derivadas segundas, \tilde{X}_l , que tem dimensão (3×3)

$$\tilde{X}_l = \begin{bmatrix} \frac{\partial^2 \eta_l}{\partial \beta_1^2} & \frac{\partial^2 \eta_l}{\partial \beta_1 \partial \beta_2} & \frac{\partial^2 \eta_l}{\partial \beta_1 \partial \beta_3} \\ & \frac{\partial^2 \eta_l}{\partial \beta_2^2} & \frac{\partial^2 \eta_l}{\partial \beta_2 \partial \beta_3} \\ & & \frac{\partial^2 \eta_l}{\partial \beta_3^2} \end{bmatrix} = \begin{bmatrix} \eta_l & \eta_l X_{1l} & \eta_l X_{2l} \\ & \eta_l X_{1l}^2 & \eta_l X_{1l} X_{2l} \\ & & \eta_l X_{2l}^2 \end{bmatrix}, \text{ para } l = 1, 2, \dots, n$$

Matriz de inversa de Fisher que será denotada por KDEBETA e a matriz das segundas derivadas que será denotada por \tilde{X}_l , portanto $C = tr(KDEBETA * \tilde{X}_l)$. Para o modelo com componente aleatória normal, a matriz diagonal obtida será denotada por $trN(KDEBETA_N * \tilde{X}_l)$ e, para o modelo com componente aleatória gama, $trG(KDEBETA_G * \tilde{X}_l)$.

Matriz diagonal peso, para as componentes aleatória normal e gama, denotados por W_N e W_G , respectivamente, cujos elementos da diagonal são:

$$W_N = \frac{1}{V(\boldsymbol{\mu})} \frac{d\boldsymbol{\mu}}{d\boldsymbol{\eta}} = 1 = \mathbf{I};$$

$$W_G = \frac{1}{V(\boldsymbol{\mu})} \frac{d\boldsymbol{\mu}}{d\boldsymbol{\eta}} = \left\{ \frac{1}{\mu_{il}^2} \right\}.$$

Matriz diagonal, F_N e F_G , cujos índices N e G denotam normal e gama. O elemento diagonal destas matrizes são:

$$F_N = \frac{1}{V(\boldsymbol{\mu})} \frac{d\boldsymbol{\mu}}{d\boldsymbol{\eta}} \frac{d^2 \boldsymbol{\mu}}{d\boldsymbol{\eta}^2} = 0$$

$$F_G = \frac{1}{V(\boldsymbol{\mu})} \frac{d\boldsymbol{\mu}}{d\boldsymbol{\eta}} \frac{d^2 \boldsymbol{\mu}}{d\boldsymbol{\eta}^2} = 0$$

Matriz diagonal Z_d , cujos elementos são a diagonal da matriz $Z = \{z_{ij}\} = (\tilde{X}^T W \tilde{X})^{-1} \tilde{X}_l$

$$\text{Vetor } \boldsymbol{\xi}_1 = -(2\phi)^{-1} Z_d W^{-1} F \mathbf{1} = \mathbf{0}$$

$$\text{Vetor } \boldsymbol{\xi}_2 = -(2\phi)^{-1} tr(KDEBETA * \tilde{X}_l) \mathbf{1}$$

6º Passo: Procedure REG O viés é a solução do sistema de equações normais de mínimos quadrados ponderados, isto é, a solução do sistema a seguir, que são as equações normais com matriz modelo \tilde{X} , peso W e vetor resposta $\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2$. Portanto o *procedure* REG calcula o viés de $\hat{\boldsymbol{\beta}}$, usando a expressão,

$$B(\hat{\boldsymbol{\beta}}) = -(2\phi)^{-1} \left(\tilde{X}^T W \tilde{X} \right)^{-1} \tilde{X} W (\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2),$$

7º Passo: procedure IML para obter a matriz covariância corrigida Para essa matriz, são necessárias as matrizes $P, H, Z_d, Z^{(2)}, C = tr(KDEBETA * \tilde{X}_l)$, que praticamente já foram obtidas no passo anterior. Neste passo, a dificuldade é obter a matriz $diag(tr[(D3)_a K^{-1}])$, que neste modelo para

- β_0 fixado, resulta:

$$(D3)_1 = \begin{bmatrix} \frac{\partial^3 \eta}{\partial \beta_0^3} & \frac{\partial^3 \eta}{\partial \beta_0^2 \partial \beta_1} & \frac{\partial^3 \eta}{\partial \beta_0^2 \partial \beta_2} \\ \frac{\partial^3 \eta}{\partial \beta_0 \partial \beta_1^2} & \frac{\partial^3 \eta}{\partial \beta_0 \partial \beta_1 \partial \beta_2} & \frac{\partial^3 \eta}{\partial \beta_0 \partial \beta_2^2} \end{bmatrix} = \begin{bmatrix} \eta & \eta x_1 & \eta x_2 \\ & \eta x_1^2 & \eta x_1 x_2 \\ & & \eta x_2^2 \end{bmatrix};$$

- β_1 fixado, resulta:

$$(D3)_2 = \begin{bmatrix} \frac{\partial^3 \eta}{\partial \beta_0^2 \partial \beta_1} & \frac{\partial^3 \eta}{\partial \beta_1^2 \partial \beta_0} & \frac{\partial^3 \eta}{\partial \beta_0 \partial \beta_1 \partial \beta_2} \\ \frac{\partial^3 \eta}{\partial \beta_1^3} & \frac{\partial^3 \eta}{\partial \beta_1^2 \partial \beta_2} & \frac{\partial^3 \eta}{\partial \beta_1 \partial \beta_2^2} \end{bmatrix} = \begin{bmatrix} \eta x_1 & x_1^2 \eta & x_1 x_2 \eta \\ & x_1^3 \eta & x_1^2 x_2 \eta \\ & & x_1 x_2^2 \eta \end{bmatrix}$$

- e β_2 fixado, resulta:

$$(D3)_3 = \begin{bmatrix} \frac{\partial^3 \eta}{\partial \beta_2 \partial \beta_0^2} & \frac{\partial^3 \eta}{\partial \beta_2 \partial \beta_0 \partial \beta_1} & \frac{\partial^3 \eta}{\partial \beta_2^2 \partial \beta_0} \\ \frac{\partial^3 \eta}{\partial \beta_1^2 \partial \beta_2} & \frac{\partial^3 \eta}{\partial \beta_2^2 \partial \beta_1} & \frac{\partial^3 \eta}{\partial \beta_2^3} \end{bmatrix} = \begin{bmatrix} x_2 \eta & \eta x_1 x_2 & \eta x_2^2 \\ & x_1^2 x_2 \eta & x_1 x_2^2 \eta \\ & & x_2^3 \eta \end{bmatrix}.$$

6.3 Resultados

Avaliando as estimativas de $O(n^{-1})$ e $O(n^{-2})$ do parâmetro $\boldsymbol{\beta}$

Adotou-se o modelo gama, uma vez que este apresenta menor AIC. Para este modelo, os resultados das estimativas do parâmetro $\boldsymbol{\beta}$, quando, amostra é de tamanho $n = 30$ encontra-se na tabela abaixo. A coluna quociente avalia se as magnitudes dos vieses são apreciáveis quando comparada ao erros-padrão, a qual apresenta um intervalo de variação relativamente grande, mostrando que, neste caso, a correção se faz necessária.

Tabela 6.1: Estimativas do vetor de parâmetros β para amostra de tamanho 30.

Parâmetros	Estimativas de $O(n^{-1})$	Erros-padrão da Estimativa de $O(n^{-1})$	Vieses	Estimativas de $O(n^{-2})$	Quocientes (Viés/ erro-padrão)*100
β_0	2,2239	0,2211	-1,7340	3,9579	784%
β_1	0,0060	0,0023	0,0211	-0,0150	917%
β_2	0,0343	0,0156	0,0226	0,0118	14%

A Tabela 6.2 apresenta os mesmos resultados da Tabela 6.1. Para amostras de tamanho $n = 800$, observa-se que as magnitudes dos vieses são bem menores quando comparada aos erros-padrão, isto é, apresentam um intervalo de variação menor como se esperava, mostrando que para amostras grandes o viés é dispensável.

Tabela 6.2: Estimativas do vetor de parâmetros β para amostra de tamanho 800.

Parâmetros	Estimativas de $O(n^{-1})$	Erros-padrão da Estimativa de $O(n^{-1})$	Vieses	Estimativas de $O(n^{-2})$	Quocientes (Viés/ erro-padrão)*100
β_0	2,4591	0,0431	-0,0029	2,4619	0,11%
β_1	0,0050	0,0003	0,0001	0,0049	33,33%
β_2	0,0174	0,0041	-0,0008	0,0182	19,51%

Avaliando as estimativas de $O(n^{-1})$ e $O(n^{-2})$ da matriz de covariância de $\hat{\beta}$

Para as amostras de tamanho $n = 30$ e $n = 800$ são obtidas as matrizes de covariância de $O(n^{-1})$ e $O(n^{-2})$. Como é de interesse, conhecer a magnitude entre Cov_2 com Cov_1 , em percentual, utilizou-se como medida a seguinte expressão:

$$medida = abs [(Cov_2/Cov_1) * 100\% - 100],$$

Observa-se que para:

- $n = 30$ o intervalo de variação foi de [12% 49%];
- $n = 800$ o intervalo de variação foi de [0% 0,7%]

A seguir, são apresentadas as matrizes de covariância de $\hat{\beta}$ de $O(n^{-1})$ e de $O(n^{-2})$ para os dois tamanhos de amostras.

Matriz de covariância de $O(n^{-1})$ de $\hat{\beta}$ para amostra de tamanho 30

$$Cov_1(\hat{\beta}) = \begin{bmatrix} \hat{\beta}_0 & \hat{\beta}_1 & \hat{\beta}_2 \\ \hat{\beta}_0 & 0,04890 & 0,00040 & 0,00232 \\ \hat{\beta}_1 & & 0,53E-5 & 0,16E-5 \\ \hat{\beta}_2 & & & 0,00024 \end{bmatrix}.$$

Matriz de covariância de $O(n^{-2})$ de $\hat{\beta}$ para amostra de tamanho 30

$$Cov_2(\hat{\beta}) = \begin{bmatrix} & \hat{\beta}_0 & \hat{\beta}_1 & \hat{\beta}_2 \\ \hat{\beta}_0 & 0,03383 & 0,00024 & 0,00187 \\ \hat{\beta}_1 & & 0,35E-5 & 0,08E-5 \\ \hat{\beta}_2 & & & 0,00021 \end{bmatrix}.$$

Pode-se observar que a matriz de $O(n^{-2})$, tem variâncias menores que a de $O(n^{-1})$.

Matriz de covariância de $O(n^{-1})$ de $\hat{\beta}$ para amostra de tamanho 800

$$Cov_1(\hat{\beta}) = \begin{bmatrix} & \hat{\beta}_0 & \hat{\beta}_1 & \hat{\beta}_2 \\ \hat{\beta}_0 & 0,001861 & -7,86E-6 & -0,000147 \\ \hat{\beta}_1 & & 0,12E-6 & -6,654E-8 \\ \hat{\beta}_2 & & & 0,00017 \end{bmatrix}.$$

Matriz de covariância de $O(n^{-2})$ de $\hat{\beta}$ para amostra de tamanho 800

$$Cov_2(\hat{\beta}) = \begin{bmatrix} & \hat{\beta}_0 & \hat{\beta}_1 & \hat{\beta}_2 \\ \hat{\beta}_0 & 0,001852 & -7,64E-6 & -0,000146 \\ \hat{\beta}_1 & & 0,12E-6 & -6,606E-8 \\ \hat{\beta}_2 & & & 0,000017 \end{bmatrix}.$$

Para $n = 800$ os elementos das matrizes $Cov_1(\hat{\beta})$ e $Cov_2(\hat{\beta})$ são praticamente iguais.

O modelo ajustado deverá ser utilizado na indústria para avaliação do processo de produção do papel, isto é, durante o processo retira-se amostras pequenas, pois a produção não pode parar e, através do modelo, estima-se o valor médio do grau de refino.

É importante salientar que a tomada de decisão tem que ser em tempo real, pois a demora pode implicar em produção de baixa qualidade. Diante disto e dos resultados obtidos, os estimadores de $O(n^2)$ são indispensáveis.

Capítulo 7

Considerações finais e recomendações

Em relação ao objetivo geral desta Tese, considera-se que o mesmo foi alcançado na medida em que técnica proposta é viável e de fácil implementação, conforme mostrado no estudo de simulação e na aplicação. Em ambos, a necessidade de se obter a matriz de covariância de $O(n^{-2})$ ficou evidenciada.

Para implementação, há necessidade de um aplicativo matemático como o MAPPLE, para calcular as derivadas, uma vez que a expressão obtida envolve as matrizes das derivadas segundas e terceiras, que dependendo do preditor não-linear pode ser complicada. É também imprescindível um aplicativo estatístico que permita ao usuário operar com matrizes e tenha algum comando ou *procedure* de otimização, como é o caso do SAS.

Os programas que foram utilizados nesta Tese serão disponibilizados no site da autora para que a técnica possa ser aplicada por um maior número de pesquisadores.

A expressão obtida nesta Tese é válida para observações independentes e que o parâmetro de dispersão seja constante. Portanto para experimentos que não possuam estas características não de ser adotado. Portanto recomenda-se estender esta técnica para modelos em que tais suposições não sejam necessárias, como os modelos não-lineares generalizados superdispersados e ou os modelos não-lineares da família exponencial com covariáveis de dispersão.

Referências Bibliográficas

- [1] Bowman, K.O. e Shentom, L.R. (1977). **Maximu, Likelihood Estimation in Small Sample**, Charles Grillin and Company Ltd, Londom.
- [2] CORDEIRO, G.M.(1983). **Improved Likelihood Ratio Statistics for Generalized Linear Models**. Journal of the Royal Statistical Society B, 45, 404-413.
- [3] CORDEIRO, G.M. (1986). **Modelos Lineares Generalizados**. VII Simpósio Nacional de Probabilidade e Estatística, UNICAMP, Campinas, São Paulo.
- [4] CORDEIRO, G.M. (1987). **On the Correction to the likelihood ratio statistics**. Biometrika, 74, 265-274.
- [5] CORDEIRO, G.M. (1999). **Introdução à Teoria Assintótica**. 22^o Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro.
- [6] CORDEIRO, G.M.(2004). **Second-Order Covariance Matrix of Maximum Likelihood Estimate in Generalized Linear Models**. Statistics and Probability Letters, não publicado.
- [7] CORDEIRO, G.M. e BOTTER, D.A. (2001). **Second-Order Biases of Maximun Likelihood Estimates in Overdispersed Generalized Linear Models**. Statistics and Probability Letters, 55, 269-280.
- [8] CORDEIRO, G.M. e MCCULLAGH, P. (1991). **Bias Correction in Generalized Linear Models**. Journal of the Royal Statistical Society B, 53, 629-643.
- [9] CORDEIRO, G.M. e BARROSO, L.P. (2003). **A Third-Order Bias Corrected Estimate in Generalized Linear Models**. Relatório Técnico.RT-MAE -USP- SP
- [10] CORDEIRO, G.M. e PAULA, G.A. (1989). **Fitting Non-Exponential Family Nnlinear Models in GLIM by Using the Offset-Facility**. Lectures Notes in Statistics, 57, 105-114.
- [11] CORDEIRO, G.M. e VASCONCELLOS, K.L.P. (1997). **Bias Correction for a Class of Multivariate Nonlinear Regression Models**. Statistics an Probability Letters, 35, 155-164.

- [12] DEMÉTRIO, C. G. B., (1993)- Modelos lineares generalizados na experimentação agronômica- 5º Simpósio de estatística Aplic. à Exp. Agr. e 38º Reunião anual da sociedade internacional de biometria.
- [13] FAHMEIR, L. e KAUFMANN, H. (1985). **Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models.** Ann. Statist., 13, 342-368.
- [14] GELFAND, A. E., DALAL, S.R., (1991). **A note on overdispersed exponential families.** Biometrika 77, 55-64.
- [15] JØRGENSEN, B. (1997). **The Theory of Dispersion Models.** 1ª Edição. Chapman & Hall. Great Britain.
- MCCULLAGH, P. e NELDER, J. A. (1983). **Generalized Linear Models.** 1ª Edição. Chapman & Hall. London.
- [16] MCCULLAGH, P. e NELDER, J. A. (1989). **Generalized Linear Models.** 2ª Edição. Chapman & Hall. London.
- [17] NELDER, J.A. e WEDDERBURN, R.W.M. (1972). **Generalized Linear Models.** Journal of the Royal Statistical Society A, 135, 370-384. 1972.
- [18] PACE, L. e SALVAN, A. (1989). **Principles of Statistical Inference from a Neo-Fisherian Perspective** 2ª Edição. Chapman & Hall. London.
- [19] PAULA, G.A.(1992). **Bias Correction for Exponential Family Nonlinear Models.** Journal of Statistical Computation and Simulation, 40, 43-54.
- [20] PEERS, H.A. e IQBAL, M. (1985). **Asymptotic Expansions for Confidence Limits in the Presence of Nuisance of Parameters, with Applications.** Journal of the Royal Statistics Society, B47, 547-554.
- [21] PINHEIRO, J. C. e BATES, D. M. (2000). **Mixed-Effects Models in S and S-PLUS.** Springer-Verlag. New York.
- [22] PREGIBON, D. (1984). **Review of Generalized Linear Models.** Annals of Statistics, 12, 1589-1596.
- [23] RATKOWSKY, D. A. (1983). **Nonlinear Regression Analysis.** Marcel Decker Inc. New York.
- [24] WEI, B.C. (1999). **Exponential Family Nonlinear Models.** Springer-Verlag Singapore Pte. Ltd. China.

Apêndice A

Família exponencial

A família exponencial de distribuições de probabilidade desempenha um papel importante na inferência estatística. Como a componente aleatória dos MNLFE tem distribuição de probabilidade pertencente à esta família, é de interesse apresentar alguns resultados.

Essa classe de distribuições foi proposta independentemente por Koopman, Pitman e Dar-mois através do estudo de propriedades de suficiência estatística.

O conceito de família exponencial foi introduzido por Fisher, que posteriormente muitos aspectos desta família foram estudados por diversos autores. Jørgensen (1992), resume e discute alguns aspectos desta classe de família.

A.1 Função de densidade

A família de distribuições $\pi_{\theta, \phi}$ de um vetor aleatório \mathbf{Y} tem função de densidade ou função de probabilidade dada por:

$$\pi(\mathbf{y}; \boldsymbol{\theta}, \phi) = \exp [\phi \{ \mathbf{y}\boldsymbol{\theta} - b(\boldsymbol{\theta}) \} + c(\mathbf{y}, \phi)], \quad (\text{A.1})$$

onde $b(\cdot)$ e $c(\cdot)$ são funções específicas, $\boldsymbol{\theta}$ é o parâmetro natural e ϕ é o parâmetro de precisão, ϕ pertence a um subconjunto dos reais positivos, o qual pode ser denotado por σ^{-2} , onde σ^2 é o parâmetro de dispersão. Neste modelo, $\mu = \frac{db(\boldsymbol{\theta})}{d\boldsymbol{\theta}}$ é a média e pertence a um subconjunto dos reais.

A.2 Função de verossimilhança

Sejam y_1, y_2, \dots, y_n n variáveis aleatórias independentes com função de probabilidade dada por (A.1) com parâmetro de precisão constante para as n observações e parâmetro canônico θ_l para $l = 1, 2, \dots, n$. Em algumas situações y_l pode depender de uma variável x_{li} fixa, chamada de variável independente. A função de verossimilhança dos parâmetros do mod-

elo definido em (A.1) é apresentada na (A.2 e A.3) , onde vetor $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ e $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)^T$ obtida como o produto de (A.1) para $l = 1, 2, \dots, n$, que em notação matricial é dada por:

$$L(\boldsymbol{\theta}, \phi) = \exp [\phi \{ \mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta}) \} + c(\mathbf{y}, \phi)]$$

ou equivalentemente:

$$L(\boldsymbol{\theta}, \phi) = \exp \left[\sum_{l=1}^n [\phi \{ y_l \theta_l - b(\theta_l) \} + c(y_l, \phi)] \right].$$

Uma vez que existe uma função biúnivoca entre $\boldsymbol{\mu}$ e $\boldsymbol{\theta}$, pode-se denotar o logaritmo da verossimilhança de \mathbf{y} em termos de $\boldsymbol{\mu}$ com relação o parâmetro $\boldsymbol{\theta}$, ϕ que será denotado por $\ell(\boldsymbol{\mu}, \phi) = \log L(\boldsymbol{\theta}, \phi)$ para um fixado parâmetro de precisão ϕ é dado por:

$$\ell(\boldsymbol{\mu}, \phi) = \phi \{ \mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta}) \} + c(\mathbf{y}, \phi), \quad (\text{A.2})$$

ou equivalentemente,

$$\ell(\boldsymbol{\mu}, \phi) = \sum_{i=1}^n \phi \{ y_i \theta_i - b(\theta_i) \} + c(y_i, \phi). \quad (\text{A.3})$$

A.3 Função escore e matriz de informação de Fisher

A função escore e a matriz de informação de Fisher são úteis na obtenção dos estimadores de máxima verossimilhança para os parâmetros da distribuição de probabilidade de um vetor aleatório \mathbf{Y} .

A função escore de um parâmetro é definida com sendo a derivada primeira da log-verossimilhança com relação ao parâmetro.

A matriz de informação de Fisher é o valor esperado negativo da derivada segunda da log-verossimilhança.

Para obtenção da função escore e matriz de informação de Fisher, foram utilizados os resultados apresentados a seguir, os quais consistem das derivadas com relação a $\boldsymbol{\mu}$ e $\boldsymbol{\theta}$, válidas para os modelos exponenciais:

$$\begin{aligned} V(\boldsymbol{\theta}) &= V = \frac{\partial^2 b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}}, & \frac{\partial^2 \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T \partial \boldsymbol{\theta}} &= S \\ \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}^T} &= V(\boldsymbol{\theta}), & \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\mu}^T} &= V^{-1}(\boldsymbol{\mu}) = V^{-1}, \\ \text{e } \frac{\partial^2 \boldsymbol{\theta}}{\partial \boldsymbol{\mu}^T \partial \boldsymbol{\mu}} &= - [V^{-1}] [V^{-1} S V^{-1}] \end{aligned}$$

Função escore de θ

Derivando (A.2), com relação a θ tem-se:

$$U_{\theta} = \frac{\partial \ell}{\partial \theta} = \phi(\mathbf{y}^T - \boldsymbol{\mu}) = \phi(\mathbf{y} - \boldsymbol{\mu}).$$

Para a componente θ_r é dada por:

$$U_{\theta_r} = \frac{\partial \ell}{\partial \theta_r} = \phi(y_r - \mu_r). \quad (\text{A.4})$$

Função escore para μ

Para obter a derivada de primeira da log-verossimilhança com relação a μ , é necessário utilizar a regra da cadeia, assim tem-se:

$$U_{\mu} = \frac{\partial \ell}{\partial \theta} \frac{\partial \theta}{\partial \mu},$$

substituindo, nesta equação, $\frac{\partial \ell}{\partial \theta}$ por $\phi(y_r - \mu_r)$, conforme (A.4) e $\frac{\partial \theta}{\partial \mu}$ por V^{-1} , a função escore de μ fica:

$$U_{\mu} = \phi(\mathbf{y} - \boldsymbol{\mu})V^{-1}(\boldsymbol{\mu}) = \phi V^{-1} \mathbf{e} \quad (\text{A.5})$$

onde e denota o erro aleatório $(\mathbf{y} - \boldsymbol{\mu})$.

Função escore para ϕ

Derivando a equação (A.2) com relação a ϕ tem-se:

$$U_{\phi} = \frac{\partial \ell}{\partial \phi} = \{\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta})\} + c(\mathbf{y}, \phi). \quad (\text{A.6})$$

Estimador de máxima verossimilhança para μ , θ e ϕ

Os estimadores de máxima verossimilhança de μ , θ e ϕ , serão denotados por $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\theta}}$ e $\hat{\phi}$, respectivamente. Para obter $\hat{\boldsymbol{\mu}}$, basta igualar (A.5) a zero. Obtendo-se como estimador de máxima verossimilhança para μ :

$$\hat{\boldsymbol{\mu}} = \mathbf{y}.$$

Pela propriedade de invariância dos estimadores de máxima verossimilhança e pelo fato de que $\theta = \frac{\partial b^{-1}(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}$, tem-se que:

$$\hat{\boldsymbol{\theta}} = \frac{\partial b^{-1}(\hat{\boldsymbol{\mu}})}{\partial \hat{\boldsymbol{\mu}}}.$$

Finalmente, igualando (A.6) a zero obtem-se:

$$c(\mathbf{y}, \tilde{\phi}) = -\mathbf{y}^T \tilde{\boldsymbol{\theta}} + b(\tilde{\boldsymbol{\theta}}),$$

e ainda se a função for da família exponencial linear, a estimativa $\tilde{\phi}$ é dada por:

$$\tilde{\phi} = \dot{s}^{-1} \left[2 \left\{ \mathbf{y}^T \tilde{\boldsymbol{\theta}} - b(\tilde{\boldsymbol{\theta}}) - c(\mathbf{y}) \right\} \right].$$

Matriz de informação de Fisher de θ e ϕ

A matriz de informação de Fisher, que será denotada por $K(\boldsymbol{\theta}, \phi)$, pode ser escrita como uma matriz em blocos, da seguinte forma: $\begin{bmatrix} K_{\theta\theta} & K_{\theta\phi} \\ K_{\theta\phi} & k_{\phi\phi} \end{bmatrix}$ onde, $K_{\theta\theta} = -E \left[\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]$, $K_{\theta\phi} = -E \left[\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}^T \partial \phi} \right]$ e $k_{\phi\phi} = -E \left(\frac{\partial^2 \ell}{\partial \phi^2} \right)$. Observa-se que:

$$K_{\theta\theta} = E \left(-\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right) = E \left(\phi \frac{\partial^2 b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right) = \phi V(\boldsymbol{\mu}).$$

Para obter $k_{\phi\phi}$ deriva-se (A.6) com relação a ϕ obtendo-se:

$$k_{\phi\phi} = E \left(-\frac{\partial^2 \ell}{\partial \phi^2} \right) = E(-\ddot{c}(\mathbf{y}, \phi)),$$

onde $\ddot{c}(\mathbf{y}, \phi)$ é a derivada segunda de $c(\mathbf{y}, \phi)$.

Por fim, para obter $K_{\phi\theta}$, deriva-se a log-verossimilhança com relação a ϕ e $\boldsymbol{\theta}$ assim

$$K_{\phi\theta} = E \left(-\frac{\partial^2 \ell}{\partial \phi \partial \boldsymbol{\theta}^T} \right) = E(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}.$$

Matriz de informação de Fisher de $\boldsymbol{\mu}$ e ϕ

Para obter a matriz de informação de Fisher de $\boldsymbol{\mu}$ e ϕ , que será denotada por $K(\boldsymbol{\mu}, \phi)$, o procedimento é análogo ao da matriz $K(\boldsymbol{\theta}, \phi)$.

Considere então a forma: $\begin{bmatrix} K_{\boldsymbol{\mu}\boldsymbol{\mu}} & K_{\boldsymbol{\mu}\phi} \\ K_{\boldsymbol{\mu}\phi} & k_{\phi\phi} \end{bmatrix}$ sendo, $K_{\boldsymbol{\mu}\boldsymbol{\mu}} = -E \left[\frac{\partial^2 \ell}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T} \right]$, $K_{\boldsymbol{\mu}\phi} = -E \left[\frac{\partial^2 \ell}{\partial \boldsymbol{\mu}^T \partial \phi} \right]$ e $k_{\phi\phi} = -E \left(\frac{\partial^2 \ell}{\partial \phi^2} \right)$.

A matriz $K_{\boldsymbol{\mu}\boldsymbol{\mu}}$, é obtida pela regra da cadeia para derivadas de 2ª ordem de um vetor, isto é,

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T} = \left(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\mu}^T} \right)^T \left(\frac{-\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right) \left(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\mu}^T} \right) - \left[\left(\frac{\partial \ell}{\partial \boldsymbol{\theta}} \right)^T \right] \left[\frac{\partial^2 \boldsymbol{\theta}}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T} \right].$$

Substituindo as derivadas do lado direito da equação pelos resultados apresentados nesta

seção tem-se:

$$\begin{aligned}\frac{\partial^2 l}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T} &= -\{V^{-1}(\phi V)V^{-1} + [\phi(y - \boldsymbol{\mu})^T][[V^{-1}][V^{-1}SV^{-1}]]\} \\ &= -\{\phi V^{-1} + \phi[e^T V^{-1}][V^{-1}SV^{-1}]\},\end{aligned}$$

aplicando a esperança obtem-se $K_{\boldsymbol{\mu}\boldsymbol{\mu}}$ que é dada por:

$$K_{\boldsymbol{\mu}\boldsymbol{\mu}} = E\left(-\frac{\partial^2 l}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T}\right) = -E\{\phi V^{-1} + \phi[e^T V^{-1}][V^{-1}SV^{-1}]\} = \phi V^{-1}(\boldsymbol{\mu}).$$

A matriz $K_{\boldsymbol{\mu}\phi} = E(V^{-1}(\mathbf{y} - \boldsymbol{\mu})) = \mathbf{0}$.

As matrizes $K(\boldsymbol{\theta}, \phi)$ e $K(\boldsymbol{\mu}, \phi)$ são bloco diagonal. Portanto, mostra que, tanto $\boldsymbol{\theta}$ e ϕ , quanto $\boldsymbol{\mu}$ e ϕ , são ortogonais e, portanto, podem ser estimados separadamente.

A.4 *Deviance* ou função desvio

A *deviance* para vetores aleatórios pertencentes à família exponencial é uma estatística importante em modelos lineares generalizados e nos modelos não-lineares da família exponencial, pois é através desta estatística que será avaliada a qualidade do ajuste destes modelos.

A *deviance* pode ser obtida através da estatística da razão de verossimilhanças. Considere o modelo definido em (A.1) com parâmetro de precisão ϕ , fixado, e o problema de testar as hipóteses:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0.$$

A estatística da razão de verossimilhanças de Neyman e Pearson (LR), definida em McCullag e Nelder (1989, Apêndice C), para testar a hipótese nula, é obtida pela diferença do log-verossimilhança calculada em $\boldsymbol{\mu} = \tilde{\boldsymbol{\mu}}$ e $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, isto é,

$$\begin{aligned}LR(\boldsymbol{\mu}_0) &= 2\{\ell(\tilde{\boldsymbol{\mu}}; \mathbf{y}) - \ell(\boldsymbol{\mu}_0; \mathbf{y})\} \\ &= 2\phi\{\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta})\}_{\boldsymbol{\mu}=\mathbf{y}} - 2\phi\{\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta})\}_{\boldsymbol{\mu}=\boldsymbol{\mu}_0} \\ &= 2\phi\{\mathbf{y}^T q(\mathbf{y}) - b(q(\mathbf{y}))\} - 2\phi\{\mathbf{y}^T q(\boldsymbol{\mu}_0) - b(q(\boldsymbol{\mu}_0))\},\end{aligned}\tag{A.7}$$

sendo $\boldsymbol{\theta} = \int V^{-1} d\boldsymbol{\mu} = q(\boldsymbol{\mu})$.

A *deviance* que é uma distância entre o log da função de verossimilhança do modelo saturado (com n parâmetros) e do modelo sob investigação (com p parâmetros) avaliado na estimativa de máxima verossimilhança $\hat{\boldsymbol{\mu}}$ é dada por:

$$\begin{aligned}
D(\mathbf{y}, \boldsymbol{\mu}) &= 2\phi\{y^T q(\mathbf{y}) - b(q(\mathbf{y}))\} - 2\{\mathbf{y}^T q(\boldsymbol{\mu}) - b(q(\boldsymbol{\mu}))\} \\
&= 2\{\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta}) - c(\mathbf{y})\}_{\boldsymbol{\mu}=\mathbf{y}} - 2\{\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta}) - c(\mathbf{y})\} \\
&= 2\{\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta}) - c(\mathbf{y})\}_{\boldsymbol{\mu}=\mathbf{y}} - 2\{\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta}) - c(\mathbf{y})\},
\end{aligned} \tag{A.8}$$

$D(y, \mu)$ é chamado de *deviance* ou desvio.

A distribuição assintótica desta estatística é uma χ^2 com n graus de liberdade, a *deviance* para qualquer $\boldsymbol{\mu}$ pode ser escrita como:

$$D^*(\mathbf{y}, \boldsymbol{\mu}) = \phi D(\mathbf{y}, \boldsymbol{\mu}) \sim \chi^2(n) \text{ (assintoticamente).}$$

Wei (1998, p.9 e10) mostra que a log-verossimilhança é proporcional a *deviance*. Assim pode-se obter o EMV de $\boldsymbol{\theta}$ e μ minimizando a *deviance* ao invés de maximizar a log-verossimilhança.

Apêndice B

Propriedades do EMV

A metodologia baseada na verossimilhança é adotada para a obtenção dos estimadores dos parâmetros dos modelos sob investigação. Portanto, é importante descrever alguns aspectos desta metodologia.

Considere \mathbf{y} o vetor observado do vetor aleatório \mathbf{Y} pertencente à família exponencial com vetor de parâmetro canônico desconhecido θ . Seja $\Theta \subset R^p$ o espaço paramétrico representando o conjunto de valores possíveis do vetor θ .

As funções densidade de probabilidade que serão consideradas para a classe dos MLG e dos MNLFE satisfazem, as seguintes condições de regularidade:

- i)* Θ é fechado, compacto e tem dimensão finita sendo o parâmetro verdadeiro θ_0 um ponto interior de Θ ;
- ii)* a função de densidade de probabilidade do vetor \mathbf{Y} é uma função bijetora de θ ;
- iii)* a matriz de informação de Fisher para θ é finita e positiva definida numa vizinhança de θ_0 ;
- iv)* as três primeiras derivadas da log-verossimilhança existem numa vizinhança de θ_0 ;
- v)* para modelos contínuos a igualdade $\frac{\partial}{\partial \theta} E \{t(\mathbf{Y})\} = \int t(\mathbf{y}) \frac{\partial}{\partial \theta} f(\mathbf{y}; \theta) d\mathbf{y}$ é válida para quaisquer estatísticas $t(\mathbf{y})$;
- vi)* para modelos discretos a igualdade $\frac{\partial}{\partial \theta} E \{t(\mathbf{Y})\} = \sum t(\mathbf{y}) \frac{\partial}{\partial \theta} f(\mathbf{y}; \theta)$ é válida para quaisquer estatísticas $t(\mathbf{y})$;

as duas últimas condições garantem que as operações de diferenciação com respeito a θ e a integração ou soma em \mathbf{y} são permutáveis.

As condições são usadas para justificar expansões em séries de Taylor, procedimentos similares e para a demonstração das identidades de Bartlett.

Propriedades dos EMV

Cordeiro (1992, seções 3.2 e 3.4) demonstra as propriedades de consistência, unicidade, invariância e suficiência dos estimadores de máxima verossimilhança.

A normalidade e eficiência assintótica, são apresentadas por Cramér (1946, seção 33.3) e Lehmann (1953, seção 6.4). Estes autores demonstram que, usualmente o estimador de máxima verossimilhança são viesados.

Com objetivo de melhorar os resultados da teoria assintótica de primeira ordem Cordeiro (1999) apresenta alguns resultados referentes à teoria assintótica de segunda ordem, que neste caso os erros associados as propriedades estatísticas são em geral de ordem $O(n^{-2})$. As identidades de Bartlett facilitam a obtenção dos cumulantes, pois determinada parametrização pode conduzir a um cálculo direto e simples de alguns cumulantes, sendo os demais calculados indiretamente através destas identidades. Esses cumulantes têm como aplicabilidade principal a obtenção de estimadores de segunda ordem.

Apêndice C

Identidades de Bartlett

Seja $L : L(\theta)$ a verossimilhança total de um problema regular. Será adotado a seguinte notação para as derivadas da log-verossimilhança ℓ , onde todos índices variam de $1 \cdots p$:

$$U_r = \frac{\partial \ell}{\partial \theta_r}, \quad U_{rs} = \frac{\partial^2 \ell}{\partial \theta_r \partial \theta_s}, \quad \text{etc.}$$

Os momentos conjuntos de derivadas de $\ell(\theta)$ são

$$\mu_r = E(U_r), \quad \mu_{rs} = E(U_{rs}), \quad \mu_{r,s} = E(U_r U_s), \quad \mu_{r,st} = E(U_r U_{st})$$

e assim, por diante.

Como

$$\mu_r = 0$$

os correspondentes cumulantes conjuntos (κ' 's) expresso em termos dos momentos são:

$$\begin{aligned} \kappa_{r,s} &= \mu_{r,s}, \\ \kappa_{rs} &= \mu_{rs}, \\ \kappa_{rs,t} &= \mu_{rs,t}, \\ \kappa_{rs,tu} &= \mu_{rs,tu} - \mu_{rs} \mu_{tu}, \\ \kappa_{r,s,t} &= \mu_{r,s,t} \text{ e} \\ \kappa_{r,s,t,u} &= \mu_{r,s,t,u} - \sum_{(3)} \mu_{r,s} \mu_{t,u}, \end{aligned}$$

onde a $\sum_{(k)}$ representa o somatório sobre todas as k combinações de índices.

Os momentos e cumulantes definidos não são independentes, mas satisfazem à certas equações, as quais facilitam seus cálculos. Estas equações, que representam condições de regularidade,

são denominadas de identidades de Bartlett.

Na dedução das identidades de Bartlett será usada a condição de regularidade (v) definida no Apêndice B, ou seja,

$$\frac{\partial}{\partial \theta} E \{t(Y)\} = \int t(y) \frac{\partial f(y; \theta)}{\partial \theta} dy$$

que é válida em problemas regulares

- $\kappa_r = 0$.

Para verificar esta idntidade, considere $U_r = \frac{L_r}{L}$ onde $L_r = \frac{\partial L}{\partial \theta_r}$.

Como a função de verosimilhança é uma função de densidade de probabilidade, tem-se que $\int L dy = 1$.

Diferenciando com relação a θ_r esta igualdade tem-se $\frac{\partial}{\partial \theta_r} \int L dy = 0$. Aplicando a condição de regularidade acima, isto é, permutando a derivada com a integral tem-se que, $\int \frac{\partial L}{\partial \theta_r} dy = 0$.

Logo pode-se concluir que $\kappa_r = E(U_r) = 0$.

- $\kappa_{rs} + \kappa_{r,s} = 0$.

A sua dedução é análoga à anterior, isto é, diferenciando a $\int L_r dy = 0$ com relação a θ_s .

Obtem-se $\int \frac{\partial}{\partial \theta_s} L_r dy = 0$, substituindo L_r por $U_r L$ tem-se que:

$$\int \frac{\partial}{\partial \theta_s} (U_r L) dy = \int \left\{ \left(\frac{\partial U_r}{\partial \theta_s} \right) L + \left(\frac{\partial L}{\partial \theta_s} \right) U_r \right\} dy = 0,$$

assim, $(U_{rs} L + U_r U_s) dy = 0$, ou seja, $\kappa_{rs} + \kappa_{r,s} = 0$.

Denotando-se as derivadas dos cumulantes com sobrescritos, isto é,

$$\kappa_{rs}^{(t)} = \frac{\partial \kappa_{rs}}{\partial \theta_t}, \quad \kappa_{rs}^{(tu)} = \frac{\partial^2 \kappa_{rs}}{\partial \theta_t \partial \theta_u}, \quad \kappa_{rst}^{(u)} = \frac{\partial \kappa_{rst}}{\partial \theta_u},$$

etc, pode-se deduzir outras identidades de Bartlett de forma análoga, destacando-se:

- $\kappa_{r,s,t} + \kappa_{rst} + \sum_{(3)} \kappa_{r,st} = 0$,
- $\kappa_{r,st} + \kappa_{rst} - \kappa_{st}^{(r)} = 0$,
- $\kappa_{r,s,t} - 2\kappa_{rst} + \sum_{(3)} \kappa_{rs}^{(t)} = 0$,
- $\kappa_{rst}^{(u)} = \kappa_{rst,u}$,
- $\kappa_{r,stu} + \kappa_{rstu} - \kappa_{stu}^{(r)} = 0$,
- $\kappa_{rstu} + \sum_{(4)} \kappa_{r,stu} + \sum_{(3)} \kappa_{rs,tu} + \sum_{(6)} \kappa_{r,s,tu} + \kappa_{r,s,t,u} = 0$,

- $\kappa_{r,s,t,u} = -3\kappa_{rstu} + 2\sum_{(4)} \kappa_{rst}^{(u)} - \sum_{(6)} \kappa_{rs}^{(tu)} + \sum_{(3)} \kappa_{rs,tu}$,
- $\kappa_{r,s,tu} = \kappa_{rstu} - \kappa_{rtu}^{(s)} - \kappa_{stu}^{(r)} + \kappa_{tu}^{(rs)} - \kappa_{rs,tu}$.

C.1 Viés de 1ª ordem do EMV

Como foi visto no comportamento assintótico dos estimadores de máxima verossimilhança, em geral, estes estimadores são viesados o quando o tamanho da amostra, n , é pequeno ou a informação de Fisher é reduzida. Porém, segundo Cordeiro (1998) o viés é algumas vezes ignorado na prática, uma vez que comparando a ordem de seu viés, que é n^{-1} , com a do erro padrão da estimativa de máxima verossimilhança, que é $n^{-1/2}$, observa-se que o viés é desprezível se n for grande. Mas em pequenas amostras, este viés pode apresentar uma magnitude comparável ao do erro-padrão de sua estimativa. Assim, é interessante estimar este viés neste caso.

A fórmula do viés de ordem n^{-1} da estimativa de máxima verossimilhança considerando observações identicamente distribuídas para modelos uniparamétricos foi deduzida por Bartlett (1953), e, para modelos multiparamétricos com observações não necessariamente identicamente distribuídas a dedução é apresentada por Cox e Snell (1968).

Para determinar o viés de $O(n^{-1})$ do estimador de máxima verossimilhança, Cox e Snell (1968) expandem o sistema de equações

$$Ur = 0$$

até o termo de segunda ordem da série de Taylor, isto è,

$$\begin{aligned} Ur &= \frac{\partial \ell(\theta_r)}{\partial \theta} + \frac{\partial^2 \ell(\theta_r)}{\partial \theta^2} (\hat{\theta}_r - \theta_r) \\ &= U_r + \sum_s U_{rs} (\hat{\theta}_s - \theta_s) + O_p(1), \end{aligned}$$

que na forma matricial fica,

$$U = K (\hat{\theta}_s - \theta_s) + O_p(1).$$

Resultando,

$$(\hat{\theta}_s - \theta_s) = K^{-1}U + O_p(1). \quad (\text{C.1})$$

Assim, expandindo Ur até segunda ordem, tem-se,

$$Ur = U_r + \sum_s U_{rs} (\hat{\theta}_s - \theta_s) + \frac{1}{2} \sum_{s,t} U_{rst} (\hat{\theta}_s - \theta_s) (\hat{\theta}_t - \theta_t) + o_p(1). \quad (\text{C.2})$$

Aplicando a esperança em (C.2) e como $Ur = 0$, tem-se

$$E(U_r) + \sum_s E \left\{ U_{rs} (\hat{\theta}_s - \theta_s) \right\} + \frac{1}{2} \sum_s E \left\{ U_{rst} (\hat{\theta}_s - \theta_s) (\hat{\theta}_t - \theta_t) \right\} + o_p(1) = 0. \quad (C.3)$$

Por outro lado a segunda expressão de (C.3) fica,

$$\sum_s E \left\{ U_{rs} (\hat{\theta}_s - \theta_s) \right\} = \sum_s E(U_{rs}) \sum_s E(\hat{\theta}_s - \theta_s) + Cov \left\{ U_{rs}, (\hat{\theta}_s - \theta_s) \right\}, \quad (C.4)$$

e por (C.1) tem-se que $\hat{\theta}_s - \theta_s = - \sum_t \kappa^{st} U_t$, então (C.4) resulta,

$$\begin{aligned} Cov \left\{ U_{rs}, (\hat{\theta}_s - \theta_s) \right\} &= \sum_s E \left\{ U_{rs} \left(- \sum_t \kappa^{st} U_t \right) \right\} \\ &= - \sum_t (\kappa^{st}) E(U_{rs} U_t) \\ &= - \sum_t \kappa^{st} \kappa_{rs,t}. \end{aligned} \quad (C.5)$$

E o terceiro termo da expressão (C.2) é calculado analogamente, resultando

$$\frac{1}{2} \sum_s E \left\{ U_{rst} (\hat{\theta}_s - \theta_s) (\hat{\theta}_t - \theta_t) \right\} = \frac{1}{2} \sum_{s,t} \kappa_{rst} (-\kappa^{st}). \quad (C.6)$$

Assim, a equação (C.3), lembrando que $E(U_r) = 0$, é a soma das parciais (C.4), (C.5) e (C.6) resultando em:

$$\sum_s \kappa_{rs} \sum_s E(\hat{\theta}_s - \theta_s) - \sum_t \kappa^{st} \kappa_{rs,t} + \frac{1}{2} \sum_{s,t} \kappa_{rst} (-\kappa^{st}) + o_p(1) = 0,$$

e, como $\sum_s E(\hat{\theta}_s - \theta_s) = B(\hat{\theta})$ e como $o_p(1)$ é um valor desprezível, obtém-se

$$\sum_s \kappa_{rs} B(\hat{\theta}) = \left(\sum_{s,t} \kappa^{st} \kappa_{rs,t} + \frac{1}{2} \sum_{s,t} \kappa_{rst} \right),$$

cuja inversão resulta em:

$$B(\hat{\theta}) = \left(\sum_{s,t,u} \kappa^{ru} \kappa^{st} \left(\kappa_{rs,t} + \frac{1}{2} \kappa_{rst} \right) \right), \quad (C.7)$$

para $r = 1, 2, \dots, p$ até ordem n^{-1} . Esta fórmula (C.7) é devida a Cox e Snell (1968) e através desta, obtém-se o viés, da estimativa de máxima verossimilhança em modelos multiparamétricos. Para obter o viés de ordem n^{-1} basta conhecer a matriz inversa da informação de Fisher, os cumulantes com relação a todos os parâmetros. Da equação (C.7) pode-se obter a estimativa

de segunda ordem de θ ou estimador de máxima verossimilhança corrigido, $O(n^{-2})$, que nada mais é que a diferença entre a estimativa de máxima verossimilhança de ordem n^{-1} e o viés, isto é, estimativa corrigida de $\hat{\theta}$ é dada por:

$$\tilde{\theta}_r = \hat{\theta} - \hat{B}(\hat{\theta}_r),$$

onde $\tilde{\theta}_r$ denota a estimativa corrigida de $\hat{\theta}_r$ e $\hat{B}(\cdot)$ é o viés $B(\cdot)$ avaliado em $\hat{\theta}$. O estimador de máxima verossimilhança corrigida $\tilde{\theta}_r$ tem viés de ordem dois, $O(n^{-2})$.

Apêndice D

Função geratriz dos cumulantes

Esta função é apresentada em Peers e Iqbal (1985), e será utilizada nesta tese para obter a matriz de covariância de segunda ordem. Assumindo as condições de regularidade, será apresentado um esboço de sua dedução.

Considere as quantidades U_r, U_{rs}, U_{rst} e U_{rstu} . A função conjunta de cumulantes é definida em Graybill et al (1974, p.160) é dada por:

$$M_{(U_r, U_{rs}, U_{rst}, U_{rstu})}(\tau_r, \tau_{rs}, \tau_{rst}, \tau_{rstu}) = E[\exp(\tau_r U_r + \tau_{rs} U_{rs} + \tau_{rst} U_{rst} + \tau_{rstu} U_{rstu})].$$

Expandindo $M_{(U_r, U_{rs}, U_{rst}, U_{rstu})}(\tau_r, \tau_{rs}, \tau_{rst}, \tau_{rstu})$ em série de Taylor e aplicando a esperança obtém-se

$$M_{(U_r, U_{rs}, U_{rst}, U_{rstu})}(\tau_r, \tau_{rs}, \tau_{rst}, \tau_{rstu}) = \exp\left[\kappa_{rs}\tau_{rs} + \frac{1}{2!}\kappa_{r,s}\tau_r\tau_s + \frac{1}{n^2}\Upsilon_1 + \frac{1}{n}\Upsilon_2\right],$$

onde

$$\begin{aligned}\Upsilon_1 &= \kappa_{rst}\tau_{rst} + \kappa_{r,st}\tau_r\tau_{st} + \frac{1}{3!}\kappa_{rs}\tau_{rs} + \kappa_{r,s,t}\tau_r\tau_s\kappa_{rs}\tau_{rs} + \frac{1}{2!}\kappa_{r,s}\tau_r\tau_s\tau_{tr} \\ \Upsilon_2 &= \kappa_{rstu}\tau_{rstu} + \kappa_{r,stu}\tau_r\tau_{stu} + \frac{1}{2!}\kappa_{r,s,tu}\tau_r\tau_s\tau_{tu} \\ &\quad + \frac{1}{2!}\kappa_{rs,tu}\tau_{rs}\tau_{tu} + \frac{1}{4!}\kappa_{r,s,t,u}\tau_r\tau_s\tau_t\tau_u.\end{aligned}$$

Expandindo o sistema de equações de verossimilhança, $U_r = 0$ em série de Taylor até terceira ordem tem-se:

$$\begin{aligned}U_r &= U_r + U_{rs}(\hat{\theta}_s - \theta_s) + \frac{1}{2!}U_{rst}(\hat{\theta}_s - \theta_s)(\hat{\theta}_t - \theta_t) + \\ &\quad \frac{1}{3!}U_{rstu}(\hat{\theta}_s - \theta_s)(\hat{\theta}_t - \theta_t)(\hat{\theta}_u - \theta_u).\end{aligned}\tag{D.1}$$

Pace e Salvam (1999, cap 9), apresentam alguns resultados para as expansões assintóticas, dentre elas a inversão de uma expansão em série de Taylor, utilizando este resultado para inverter (D.1) e substituindo os cumulantes. Obtém-se a função conjunta dos cumulantes $\psi_{(\hat{\theta}_s - \theta_s)}(\tau_r)$, a

$$\begin{aligned}\psi_{(\hat{\theta}_s - \theta_s)}(\tau_r) &= \ln \left[M_{(U_r, U_{rs}, U_{rst}, U_{rstu})}(\tau_r, \tau_{rs}, \tau_{rst}, \tau_{rstu}) \right] \\ &= -\frac{\tau_r \tau_s}{2!} \kappa^{rs} + \frac{1}{n^2} \frac{\tau_r}{1!} Vies + \\ &\quad \frac{1}{n^2} \frac{\tau_r \tau_s \tau_t}{3!} simetria + \\ &\quad \frac{1}{n} \frac{\tau_r \tau_s}{2!} \Sigma + \frac{1}{n} \frac{\tau_r \tau_s \tau_t \tau_u}{4!} curtose + O\left(n^{-\frac{3}{2}}\right),\end{aligned}$$

sendo:

$$\begin{aligned}Vies &= \frac{1}{2} \kappa^{ra} \kappa^{bc} \kappa_{abc} + \kappa^{ra} \kappa^{bc} \kappa_{ab,c}, \\ simetria &= -2\kappa^{ra} \kappa^{sb} \kappa^{tc} \kappa_{abc} - 3\kappa^{ra} \kappa^{sb} \kappa^{tc} \kappa_{ab,c} \\ curtose &= \kappa^{ra} \kappa^{sb} \kappa^{tc} \kappa^{ud} (8\kappa_{a,bcd} + 3\kappa_{abcd} + 9\kappa_{ab,cd} + 6\kappa_{a,b,cd}) \\ &\quad - 12\kappa^{ra} \kappa^{sb} \kappa^{ti} \kappa^{uj} \kappa^{cl} (\kappa_{a,bc} + \kappa_{ab,c} + \kappa_{abc}) (\kappa_{i,j,l} + \kappa_{ijl})\end{aligned}$$

e a matriz de covariância, que foi denotada por Σ é dada por : $\sigma_{ij} = \sigma_{ij}^{(1)} + \sigma_{ij}^{(2)} + \sigma_{ij}^{(3)}$, sendo

$$\sigma_{ij}^{(1)} = -\kappa^{ia} \kappa^{jb} \kappa^{cd} \begin{pmatrix} \kappa_{abcd} + \kappa_{a,bcd} + \\ 2\kappa_{abc,d} + 2\kappa_{a,bc,d} + 3\kappa_{ac,bd} \end{pmatrix},$$

$$\sigma_{ij}^{(2)} = \kappa^{ia} \kappa^{jr} \kappa^{bs} \kappa^{ct} \begin{pmatrix} \frac{3}{2} \kappa_{abc} \kappa_{rst} + 4\kappa_{ab,c} \kappa_{rst} + \\ \kappa_{a,bc} \kappa_{rst} + 2\kappa_{ab,c} \kappa_{r,st} + \kappa_{ab,c} \kappa_{rt,s} \end{pmatrix},$$

e

$$\sigma_{ij}^{(3)} = \kappa^{ia} \kappa^{jb} \kappa^{rs} \kappa^{ct} \begin{pmatrix} 2\kappa_{a,bc} \kappa_{r,st} + \\ \kappa_{a,bc} \kappa_{rst} + \kappa_{abc} \kappa_{rst} + 2\kappa_{abc} \kappa_{r,st} \end{pmatrix}.$$

Apêndice E

Derivadas da log-verossimilhança

Para obtenção das três derivadas, utiliza-se a regra da cadeia, uma vez que a log-verossimilhança destes modelos estão em função dos parâmetros, canônico e de dispersão, é de interesse a derivada com relação aos parâmetros do modelo, β 's.

E.1 Log-verossimilhança dos MNLFE

A função de densidade de probabilidade de um vetor aleatório \mathbf{Y} que pertence à família exponencial é dada por:

$$\pi(\mathbf{y}; \boldsymbol{\theta}, \phi) = \exp [\phi (\mathbf{y}\boldsymbol{\theta} - b(\boldsymbol{\theta}) + c(\mathbf{y}))] + d(\mathbf{y}, \phi),$$

cujas funções de densidade de probabilidade da l -ésima componente Y_l é

$$\pi(y_l; \theta_l, \phi) = \exp [\phi (y_l \theta_l - b(\theta_l) + c(y_l))] + d(y_l, \phi),$$

sendo $b(\cdot)$ e $c(\cdot)$ são funções específicas, $\boldsymbol{\theta}$ é o parâmetro natural e ϕ é o parâmetro de precisão, ϕ pertence a um suconjunto dos reais positivos, que pode ser denotado por σ^{-2} , onde σ^2 é o parâmetro de dispersão.

Nesta família, $E(\mathbf{Y}) = \boldsymbol{\mu} = \frac{db(\boldsymbol{\theta})}{d\boldsymbol{\theta}}$ é o vetor de médias e pertence a um suconjunto dos reais.

A log-verossimilhança é,

$$\ell(\boldsymbol{\mu}, \phi) = \sum_{l=1}^n \phi \{y_l \theta_l - b(\theta_l) + c(y_l)\} + d(y_l, \phi), \quad (\text{E.1})$$

sendo que $\boldsymbol{\mu}$ é descrito pelo preditor não-linear $\boldsymbol{\eta}$ através da relação,

$$\eta_l = g(\mu_l) = f(\mathbf{X}; \boldsymbol{\beta}),$$

e $\boldsymbol{\beta}$ o vetor de parâmetros com p componentes.

E.2 Derivada primeira da log-verossimilhança U_r

Considere β_r a r-ésima componente do vetor $\boldsymbol{\beta}$ a derivada primeira de $\ell(\boldsymbol{\mu}, \phi)$ com relação a β_r , é obtida pela regra da cadeia, isto é,

$$U_r = \frac{\partial \ell}{\partial \beta_r} = \frac{\partial \ell}{\partial \theta_l} \frac{\partial \theta_l}{\partial \mu_l} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial \eta_l}{\partial \beta_r}. \quad (\text{E.2})$$

Obtenção de $\frac{\partial \ell}{\partial \theta_l}$, $\frac{\partial \theta_l}{\partial \mu_l}$, $\frac{\partial \mu_l}{\partial \eta_l}$ e $\frac{\partial \eta_l}{\partial \beta_r}$

Derivando (E.1), com relação a θ_l , obtém-se:

$$\frac{\partial \ell}{\partial \theta_l} = \sum_l^n \phi(y_l - \mu_l), \quad (\text{E.3})$$

conforme já visto no Apêndice A,

$$\frac{\partial \theta_l}{\partial \mu_l} = \frac{1}{V}. \quad (\text{E.4})$$

E a derivada $\frac{\partial \eta_l}{\partial \beta_r}$ será denotada por (r)

$$\frac{\partial \eta_l}{\partial \beta_r} = (r). \quad (\text{E.5a})$$

Substituindo (E.3), (E.4) e (E.5a) em (E.2), tem-se a expressão da primeira derivada da log-verossimilhança que é:

$$U_r = \frac{\partial \ell}{\partial \beta_r} = \sum_l^n \phi(y_l - \mu_l) \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} (r) \quad (\text{E.6})$$

E.3 Derivada segunda da log-verossimilhança U_{rs}

Considere β_r e β_s a, r-ésima e s-ésima componente do vetor $\boldsymbol{\beta}$, respectivamente, a derivada segunda de $\ell(\boldsymbol{\mu}, \phi)$ com relação a β_r e β_s , é obtida derivando U_r , (E.6), com relação a β_s , isto é,

$$U_{rs} = \frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} = \frac{\partial}{\partial \beta_s} \left(\phi \sum_l^n (y_l - \mu_l) \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} (r) \right),$$

a derivada da expressão acima, nada mais é que a derivada do produto dos seguintes termos:

$$\begin{aligned}
A_1 &= (y_l - \mu_l) \frac{1}{V}, \\
A_2 &= \frac{\partial \mu_l}{\partial \eta_l} e \\
A_3 &= \frac{\partial \eta_l}{\partial \beta_r} = (r),
\end{aligned}$$

derivando A_1 , A_2 e A_3 com relação a β_s e fazendo algumas manipulações algébricas obtém-se então U_{rs} ,

$$\begin{aligned}
U_{rs} &= \frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} = \phi \sum_l^n -\frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 (r, s) \\
&\quad - (y_l - \mu_l) \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 (r, s) + (y_l - \mu_l) \frac{1}{V^2} \frac{\partial^2 \mu_l}{\partial \eta_l^2} (r, s) \\
&\quad + (y_l - \mu_l) \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} (rs),
\end{aligned} \tag{E.7}$$

sendo $(rs) = \frac{\partial^2 \eta_l}{\partial \beta_r \partial \beta_s}$, $(r, s) = \frac{\partial \eta_l}{\partial \beta_r} \frac{\partial \eta_l}{\partial \beta_s}$ e $V^{(1)} = \frac{\partial V}{\partial \mu}$.

Para facilitar a obtenção dos cumulantes, U_{rs} será denotado pelas contribuições de cada parcela na equação (E.7), que serão denominadas por $(U_{rs})_{L1}$, $(U_{rs})_{L2}$, $(U_{rs})_{L3}$ e $(U_{rs})_{NL1}$, sendo:

$$(U_{rs})_{L1} = -\frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 (r, s);$$

$$(U_{rs})_{L2} = - (y_l - \mu_l) \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 (r, s);$$

$$(U_{rs})_{L3} = (y_l - \mu_l) \frac{1}{V^2} \frac{\partial^2 \mu_l}{\partial \eta_l^2} (r, s);$$

$$(U_{rs})_{NL1} = (y_l - \mu_l) \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} (rs).$$

Com esta notação, a derivada segunda resulta:

$$U_{rs} = \phi \sum_l^n (U_{rs})_{L1} + (U_{rs})_{L2} + (U_{rs})_{L3} + (U_{rs})_{NL1}.$$

E.4 Derivada terceira da log-verossimilhança U_{rst}

Considere β_r , β_s e β_t a, r -ésima, s -ésima e t -ésima componente do vetor $\boldsymbol{\beta}$, respectivamente, a derivada terceira de $\ell(\mu, \phi)$ com relação a β_r , β_s e β_t , é obtida derivando U_{rs} , (E.7), com relação a β_t , isto é,

$$\begin{aligned}
U_{rst} &= \frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s \partial \beta_t} = \frac{\partial}{\partial \beta_t} \left(\phi \sum_l^n -\frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 (r, s) \right) \\
&\quad - \frac{\partial}{\partial \beta_t} \left(\phi \sum_l^n (y_l - \mu_l) \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 (r, s) \right) \\
&\quad - \frac{\partial}{\partial \beta_t} \left(\phi \sum_l^n (y_l - \mu_l) \frac{1}{V^2} \frac{\partial^2 \mu_l}{\partial \eta_l^2} (r, s) \right) \\
&\quad + \frac{\partial}{\partial \beta_t} \left(\phi \sum_l^n (y_l - \mu_l) \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} (rs) \right),
\end{aligned}$$

cada parcela da expressão acima citada trata-se da derivada de um produto, derivando e escrevendo U_{rst} com a notação de suas contribuições com em U_{rs} tem-se:

$$\begin{aligned}
U_{rst} &= \phi \sum_l^n (U_{rst})_{L1} + (U_{rst})_{L2} + (U_{rst})_{L3} + (U_{rst})_{L4} + (U_{rst})_{L5} + (U_{rst})_{L6} \\
&\quad + (U_{rst})_{L7} + (U_{rst})_{NL2} + (U_{rst})_{NL3} + (U_{rst})_{NL4}.
\end{aligned} \tag{E.8}$$

As contribuições lineares são :

$$\begin{aligned}
(U_{rst})_{L1} &= \frac{2V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3 (r, s, t); \\
(U_{rst})_{L2} &= -\frac{3}{V} \frac{\partial^2 \mu_l}{\partial \eta_l^2} \frac{\partial \mu_l}{\partial \eta_l} (r, s, t); \\
(U_{rst})_{L3} &= (y_l - \mu_l) \frac{V^{(2)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3 (r, s, t); \\
(U_{rst})_{L4} &= -(y_l - \mu_l) \frac{2(V^{(1)})^2}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3 (r, s, t); \\
(U_{rst})_{L5} &= (y_l - \mu_l) \frac{2(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3 (r, s, t); \\
(U_{rst})_{L6} &= -(y_l - \mu_l) \frac{3V^{(1)}}{V^2} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2} (r, s, t); \\
(U_{rst})_{L7} &= -(y_l - \mu_l) \frac{1}{V} \frac{\partial^3 \mu_l}{\partial \eta_l^3} (r, s, t).
\end{aligned}$$

As contribuições devido à não linearidade do modelo são:

$$\begin{aligned}
(U_{rst})_{NL1} &: -(y_l - \mu_l) \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 [(rs, t) + (rt, s) + (st, r)]; \\
(U_{rst})_{NL2} &: -\frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 [(rs, t) + (rt, s) + (st, r)];
\end{aligned}$$

$$(U_{rst})_{NL3} : (y_l - \mu_l) \frac{1}{V} \frac{\partial^2 \mu_l}{\partial \eta_l^2} [(rs, t) + (rt, s) + (st, r)];$$

$$(U_{rst})_{NL4} : (y_l - \mu_l) \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} (rst).$$

Sendo: $(rst) = \frac{\partial^3 \eta_l}{\partial \beta_r \partial \beta_s \partial \beta_t}$, $(rs, t) = \frac{\partial^2 \eta_l}{\partial \beta_r \partial \beta_s} \frac{\partial \eta_l}{\partial \beta_t}$, $V^{(2)} = \frac{\partial^2 V}{\partial \mu^2}$.

Apêndice F

Cumulantes

Os cumulantes dos MNLFE são apresentados neste apêndice. Para obtê-los são utilizadas as derivadas da log-verossimilhança do apêndice E.

Será adotada a seguinte notação:

$$\kappa_{rs} = E(U_{rs}); \kappa_{rst} = E(U_{rst}); \kappa_{r,s} = E(U_r U_s); \kappa_{r,st} = E(U_r U_{st}) \text{ e } \kappa_{rs}^{(t)} = \frac{\partial \kappa_{rs}}{\partial \beta_t};$$

F.1 Cumulante κ_{rs}

Este cumulante é obtido facilmente, uma vez que basta obter a esperança de U_{rs} cuja expressão é apresentada na (E.7).

$$\kappa_{rs} = E(U_{rs}) = \phi \sum_l^n -\frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 (r, s),$$

para simplificar denotar-se-a $\frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2$ por w_l assim o cumulante κ_{rs} pode ser reescrito como

$$\kappa_{rs} = \phi \sum_l^n -\frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 (r, s).$$

F.2 Cumulante κ_{rst}

O cumulante κ_{rst} , é a esperança de U_{rst} cuja expressão é apresentada na (E.8), assim este cumulante é:

$$\begin{aligned}\kappa_{rst} &= E(U_{rst}) = \phi \sum_l^n \frac{2V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3 (r, s, t) - \frac{3}{V} \frac{\partial^2 \mu_l}{\partial \eta_l^2} \frac{\partial \mu_l}{\partial \eta_l} (r, s, t) \\ &\quad - \frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 [(rs, t) + (rt, s) + (st, r)],\end{aligned}$$

considere,

$$\begin{aligned}f_l &= \frac{1}{V} \frac{\partial \mu}{\partial \eta} \frac{\partial^2 \mu}{\partial \eta^2} e \\ g_l &= \frac{1}{V} \frac{\partial \mu}{\partial \eta} \frac{\partial^2 \mu}{\partial \eta^2} - \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu}{\partial \eta} \right)^3,\end{aligned}$$

assim κ_{rst} pode ser reescrito em termos de f_l , g_l e w_l como:

$$\kappa_{rst} = \phi \sum_l^n - (f_l + 2g_l) (r, s, t) - w_l [(rt, s) + (st, r) + (rs, t)]$$

F.3 Cumulante $\kappa_{bc}^{(ad)}$

A obtenção deste cumulante é trabalhoso, uma vez que envolve um número considerável de derivadas. Apesar disso, ele é a derivada segunda do κ_{bc} , com relação a β_a e β_c . Sua expressão é grande e, desta forma, será apresentada em termos das três parcelas que o compõe, isto é:

$$\kappa_{bc}^{(ad)} = \frac{\partial^2 \kappa_{bc}}{\partial \beta_a \partial \beta_d} = \phi \sum_l^n \left(\kappa_{bc}^{(ad)} \right)_{L1} + \left(\kappa_{bc}^{(ad)} \right)_{NL2} + \left(\kappa_{bc}^{(ad)} \right)_{NL3},$$

sendo:

$$\left(\kappa_{bc}^{(ad)} \right)_{L1} = \left[\frac{5V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} - \frac{2}{V} \left(\frac{\partial^2 \mu_l}{\partial \eta_l^2} \right)^2 - \frac{2}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} - \frac{2(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 + \frac{V^{(2)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 \right] (a, b, c, d);$$

$$\left(\kappa_{bc}^{(ad)} \right)_{NL2} = - (g_l + f_l) [(ab, c, d) + (ac, b, d) + (ad, b, d) + (dc, a, b) + (dc, a, b) + (db, c, a)];$$

$$\left(\kappa_{bc}^{(ad)} \right)_{NL3} = -w_l [(adc, b) + (adb, c) + (dc, ab) + (db, ac)].$$

F.4 Cumulante $\kappa_{bcd}^{(a)}$

Para este cumulante é interessante obter as derivadas f_l , g_l e w_l , uma vez que ele é igual a derivada primeira de κ_{bcd} , com relação a β_a , por isto, estas derivadas são apresentadas.

Derivada de $w_l = \frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2$ com relação a β_a

$$\begin{aligned} \frac{\partial w_l}{\partial \beta_a} &= \frac{\partial}{\partial \beta_a} \left[\frac{1}{V} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \right] = \left[\frac{\partial}{\partial \beta_a} \left(\frac{1}{V} \right) \right] \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 + \frac{1}{V} \left[\frac{\partial}{\partial \beta_a} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \right] \\ &= \frac{-V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3 (a) + \frac{2}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2} (a) = (f_l + g_l) (a). \end{aligned} \quad (F.1)$$

Derivada de $f_l = \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2}$ com relação a β_a

$$\begin{aligned} \frac{\partial f_l}{\partial \beta_a} &= \frac{\partial}{\partial \beta_a} \left[\frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2} \right] = \left[\frac{\partial}{\partial \beta_a} \left(\frac{1}{V} \right) \right] \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2} + \frac{1}{V} \left[\frac{\partial}{\partial \beta_a} \left(\frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2} \right) \right] \\ &= \frac{-V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} (a) + \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} (a) + \frac{1}{V} \left(\frac{\partial^2 \mu_l}{\partial \eta_l^2} \right)^2 (a). \end{aligned} \quad (F.2)$$

Derivada de $g_l = \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^2 \mu_l}{\partial \eta_l^2} - \frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3$ com relação a β_a

$$\begin{aligned} \frac{\partial g_l}{\partial \beta_a} &= \frac{\partial}{\partial \beta_a} [f_l] - \frac{\partial}{\partial \beta_a} \left(\frac{V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3 \right) = \\ &= \frac{\partial}{\partial \beta_a} [f_l] - \frac{V^{(2)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3 (a) + \frac{2(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^3 (a) - \frac{3(V^{(1)})^2}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} (a), \end{aligned}$$

simplicando tem-se:

$$\begin{aligned} \frac{\partial g_l}{\partial \beta_a} &= \frac{-4V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} (a) + \frac{1}{V} \left(\frac{\partial^2 \mu_l}{\partial \eta_l^2} \right)^2 (a) + \frac{1}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} (a) \\ &\quad - \frac{V^{(2)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 (a) + \frac{2(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 (a), \end{aligned} \quad (F.3)$$

substituindo estas derivadas (F.1), (F.2) e (F.3) na expressão de $\frac{\partial \kappa_{bcd}}{\partial \beta_a}$, obtem-se :

$$\kappa_{bcd}^{(a)} = \frac{\partial \kappa_{bcd}}{\partial \beta_a} = \phi \sum_l^n \left(\kappa_{bcd}^{(a)} \right)_{L1} + \left(\kappa_{bcd}^{(a)} \right)_{NL2} + \left(\kappa_{bcd}^{(a)} \right)_{NL3} + \left(\kappa_{bcd}^{(a)} \right)_{NLA},$$

sendo:

$$\left(\kappa_{bcd}^{(a)} \right)_{L1} = \left[\frac{9V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} + \frac{3}{V} \left(\frac{\partial^2 \mu_l}{\partial \eta_l^2} \right)^2 - \frac{3}{V} \frac{\partial \mu_l}{\partial \eta_l} \frac{\partial^3 \mu_l}{\partial \eta_l^3} - \frac{4(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 + \frac{2V^{(2)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 \right] (a, b, c, d);$$

$$\begin{aligned} \left(\kappa_{bcd}^{(a)} \right)_{NL2} &= -f_l [(ab, c, d) + (ac, b, d) + (ad, b, d) + (a, bc, d) + (a, bd, c) + (a, cd, b)]; \\ \left(\kappa_{bcd}^{(a)} \right)_{NL3} &= -g_l [2(ab, c, d) + 2(ac, b, d) + 2(ad, b, d) + (a, bc, d) + (a, bd, c) + (a, cd, b)]; \\ \left(\kappa_{bcd}^{(a)} \right)_{NL4} &= -w_l [(abc, d) + (bd, ad) + (abd, c) + (bd, ac) + (cd, ab)]. \end{aligned}$$

F.5 Cumulante $\kappa_{ac,bd}$

Este cumulante é obtido pela esperança do produto U_{ac} por U_{bd} , cuja a expressão é apresentada no Apêndice B, equação (E.7), para U_{ac} será adotado na variação do somatório o índice l e para U_{bd} o índice m .

As expressões para U_{ac} e U_{bd} podem ser escritas como:

$$\begin{aligned} U_{ac} &= \phi \sum_l^n -w_l(a, c) + A_l(y_l - \mu_l)(a, c) + B_l(y_l - \mu_l)(a, c), \\ U_{bd} &= \phi \sum_m^n -w_m(b, d) + A_m(y_m - \mu_m)(b, d) + B_m(y_m - \mu_m)(b, d), \end{aligned}$$

sendo:

$$\begin{aligned} A_l &= \left(\frac{V^{(1)}}{V^2} \right)_l \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2; \\ A_m &= \left(\frac{V^{(1)}}{V^2} \right)_m \left(\frac{\partial \mu_m}{\partial \eta_m} \right)^2; \\ B_l &= \left(\frac{1}{V} \right)_l \frac{\partial \mu_l}{\partial \eta_l}; \\ B_m &= \left(\frac{1}{V} \right)_m \frac{\partial \mu_m}{\partial \eta_m}. \end{aligned}$$

O cumulante $\kappa_{ac,bd}$ é dado por:

$$\kappa_{ac,bd} = E(U_{ac}U_{bd}) - E(U_{ac})E(U_{bd}),$$

portanto é necessário obter o produto $U_{ac}U_{bd}$, que é dado por:

$$\begin{aligned} U_{ac}U_{bd} &= \phi^2 \sum_l^n \sum_m^n w_l w_m(a, c, b, d) - w_l A_m(y_m - \mu_m)(a, c, b, d) - w_l B_m(y_m - \mu_m)(a, c, bd) \\ &+ \phi^2 \sum_l^n \sum_m^n -A_l w_m(y_l - \mu_l)(a, c, b, d) + A_l A_m(y_m - \mu_m)(y_l - \mu_l)(a, c, b, d) + A_l B_m(y_l - \mu_l)(a, c, bd) \\ &+ \phi^2 \sum_l^n \sum_m^n -B_l w_m(y_m - \mu_m)(ac, b, d) + B_l A_m(y_m - \mu_m)(y_l - \mu_l)(ac, b, d) + A_l B_m(y_l - \mu_l)(ac, b, d) \end{aligned}$$

Aplicando a esperança nesta equação e observando que:

1. para $l = m$,

$$E [(y_m - \mu_m) (y_l - \mu_l)] = \frac{V}{\phi};$$

2. para $l \neq m$

$$E [(y_m - \mu_m) (y_l - \mu_l)] = 0;$$

3. a esperança da primeira parcela de $U_{ac}U_{bd}$ é igual a $E(U_{ac})E(U_{bd})$, isto é,

$$E \left[\phi^2 \sum_l^n \sum_m^n w_l w_m (a, c, b, d) \right] - E(U_{ac})E(U_{bd}) = 0.$$

Tem-se a seguinte expressão para o cumulante $\kappa_{ac,bd}$:

$$\kappa_{ac,bd} = \phi \sum_l^n (\kappa_{ac,bd})_{L1} + (\kappa_{ac,bd})_{NL1} + (\kappa_{ac,bd})_{NL2},$$

sendo:

$$(\kappa_{ac,bd})_{L1} = \left[-\frac{2V^{(1)}}{V^2} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^2 \frac{\partial^2 \mu_l}{\partial \eta_l^2} + \frac{1}{V} \left(\frac{\partial^2 \mu_l}{\partial \eta_l^2} \right)^2 + \frac{(V^{(1)})^2}{V^3} \left(\frac{\partial \mu_l}{\partial \eta_l} \right)^4 \right] (a, b, c, d);$$

$$(\kappa_{ac,bd})_{NL1} = g_l [(a, c, bd) + (ac, b, d)];$$

$$(\kappa_{ac,bd})_{NL2} = w_l (ac, bd).$$

F.6 Cumulante $\kappa_{a,bc}$

O procedimento para obtenção deste cumulante é análogo ao cumulante $\kappa_{ac,bd}$. Por isto serão omitidas algumas etapas.

As expressões para U_a e U_{bc} podem ser escritas como:

$$U_a = \phi \sum_l^n \frac{1}{V_l} (y_l - \mu_l) \frac{\partial \mu_l}{\partial \eta_l} (a),$$

$$U_{bd} = \phi \sum_m^n -w_m (b, c) + A_m (y_m - \mu_m) (b, c) + B_m (y_m - \mu_m) (b, c),$$

sendo:

$$A_m : \left(\frac{V^{(1)}}{V^2} \right)_m \left(\frac{\partial \mu_m}{\partial \eta_m} \right)^2;$$

$$B_m : \left(\frac{1}{V} \right)_m \frac{\partial \mu_m}{\partial \eta_m}.$$

O cumulante $\kappa_{a,bc}$ é dado por:

$$\kappa_{a,bc} = E(U_a U_{bc}) - E(U_a) E(U_{bc}),$$

após a obtenção da esperança do produto $U_a U_{bc}$, tem-se a seguinte expressão para o cumulante $\kappa_{a,bc}$:

$$\kappa_{a,bc} = \phi \sum_l^n (\kappa_{a,bc})_{L1} + (\kappa_{a,bc})_{NL1},$$

sendo:

$$(\kappa_{a,bc})_{L1} = g_l(a, b, c);$$

$$(\kappa_{a,bc})_{NL1} = w_l(a, bc).$$