

# Resumo

Dados envolvendo medidas angulares estão presentes nas mais diversas áreas do conhecimento. Para analisá-los é necessário utilizar uma teoria estatística específica e apropriada, diferente da que utilizamos para dados lineares. Particularmente, quando o interesse for formular, ajustar e fazer diagnósticos em modelos de regressão, uma vez que, neste contexto, a natureza da variável deve ser considerada. Neste trabalho, utilizamos os modelos de regressão von Mises para investigar a associação tipo circular-linear e apresentamos dois resíduos padronizados que foram obtidos a partir da componente da função desvio e cujas distribuições de probabilidades podem ser aproximadas pela distribuição normal padrão, definida para dados lineares.

# Abstract

Data involving angular are present in the most diverse areas of science. To analyze them is necessary to introduce an appropriate theory and to study specific and appropriate statistics as well, different from that we use for linear data. When the interest is to formulate, to adjust and to make diagnostics on regression models, the nature of the variables must be considered. In this work, we use the von Mises regression models to investigate the circular-linear association and discuss two standardized residuals defined from the component of the deviance function whose probability distributions can be approximated by the normal standard distribution defined for linear data.

Universidade Federal de Campina Grande  
Centro de Ciências e Teconologia  
Programa de Pós-Graduação em Matemática  
Curso de Mestrado em Matemática

# Análise de Resíduos em Modelos de Regressão von Mises

por

Grayci-Mary Gonçalves Leal

sob orientação do

Prof. Dr. Francisco Antônio Morais de Souza

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

Campina Grande - PB

Abril/2006

# Análise de Resíduos em Modelos de Regressão von Mises

por

**Grayci-Mary Gonçalves Leal**

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de Concentração: Matemática

Linha de Pesquisa: Probabilidade e Estatística

Aprovada por:

**Prof. Dr. André Gustavo Campos Pereira**

**Prof. Dr. Antonio José da Silva**

**Prof. Dr. Francisco Antônio Morais de Souza**

**Orientador**

**Universidade Federal de Campina Grande  
Centro de Ciências e Tecnologia  
Programa de Pós-Graduação em Matemática  
Curso de Mestrado em Matemática**

**Abril/2006**

# Agradecimentos

Primeiramente, ao meu Senhor e Salvador Jesus Cristo que me concede mais esta vitória, pois sem Ele nada posso fazer;

Ao meu esposo Janilson por seu amor, incentivo e compreensão sempre dispensado e ao meu primeiro filho Carlos Alberto, por está presente neste momento tornando-o mais especial;

Aos meus pais, Afonso e Fátima, por todo amor, confiança e apoio em todas as áreas, não só agora, mas sempre. Aos meus irmãos Kelly e Affonso por sempre acreditarem e confiarem em mim, enfim agradeço a toda minha família;

Ao professor Francisco Morais, pela dedicação, humildade e paciência ao me orientar em mais este trabalho;

Aos professores André Gustavo Campos Pereira e Antonio José da Silva, por gentilmente aceitarem ao convite de avaliar este trabalho;

A todos os meus professores de graduação e pós-graduação da UAME/UFCG, entre eles Aparecido Jesuíno, Antonio José, Daniel Pelegrino e Marco aurélio. E aos funcionários Salete, D. Argentina, Valdir e Marcelino;

Aos meu colegas Lya, Jaqueline, Areli, Tatiana, Ana Cristina, Rosângela, Marta, Jesualdo, Lino, Aluízio, Cícero, Luciano, Moisés, Orlando e Lauriclécio, pela amizade e companherismo de todos;

À CAPES, pelo apoio financeiro.

# Dedicatória

Ao meu esposo Janilson e aos  
meus pais Afonso e Fátima.

# Conteúdo

<b>Introdução</b> . . . . .	4
<b>1 Dados Direcionais</b>	<b>6</b>
1.1 Introdução . . . . .	6
1.2 Representação Gráfica de Dados Direcionais . . . . .	7
1.3 Medidas Descritivas no Círculo . . . . .	9
1.3.1 Medidas de Locação . . . . .	10
1.3.2 Medidas de Concentração . . . . .	11
1.3.3 Medidas de Dispersão . . . . .	12
1.3.4 Outras Medidas de Dispersão . . . . .	12
1.4 Momentos Trigonométricos . . . . .	14
1.5 Conceitos Teóricos Básicos para Dados Direcionais . . . . .	15
<b>2 Modelos Probabilísticos para Dados Circulares</b>	<b>21</b>
2.1 Introdução . . . . .	21
2.2 Modelo Exponencial . . . . .	22
2.3 Modelos de Transformação . . . . .	23
2.4 A Distribuição Circular Uniforme - $\mathcal{U}$ . . . . .	23
2.5 As Distribuições Arqueadas . . . . .	24
2.5.1 A Distribuição Normal Arqueada - $\mathcal{WN}(\mu, \sigma^2)$ . . . . .	25
2.5.2 A Distribuição Poisson Arqueada - $\mathcal{WP}(\lambda)$ . . . . .	27
2.5.3 A Distribuição Cauchy Arqueada - $\mathcal{WC}(\mu, \rho)$ . . . . .	28
2.6 A Distribuição von Mises - $\mathcal{VM}(\mu, \lambda)$ . . . . .	29
2.6.1 Estimação dos parâmetros da Distribuição von Mises . . . . .	31

	2
2.7 Outras Distribuições Circulares . . . . .	33
<b>3 Modelos de Regressão para uma Resposta Angular</b>	<b>34</b>
3.1 Introdução . . . . .	34
3.2 Alguns Modelos para Regressão Angular . . . . .	35
3.3 Os Modelos de Regressão von Mises . . . . .	36
3.3.1 Modelo de Médias von Mises . . . . .	37
3.3.2 Modelo de Dispersão von Mises . . . . .	41
3.3.3 Modelo Misto von Mises . . . . .	44
<b>4 Resíduos para uma Resposta Angular</b>	<b>46</b>
4.1 Introdução . . . . .	46
4.2 A componente da função desvio . . . . .	47
4.3 Resíduos . . . . .	48
4.3.1 Resíduo $d_i^*$ . . . . .	49
4.3.2 Resíduo $r_i$ . . . . .	50
<b>5 Aplicação</b>	<b>55</b>
5.1 Introdução . . . . .	55
5.2 Descrição e Análise dos Dados . . . . .	56
<b>Conclusões</b> . . . . .	<b>58</b>
<b>A Demonstração das Equações (1.5) e (1.6)</b>	<b>61</b>
A.1 Demonstração . . . . .	61
<b>B Alguns Resultados sobre Dados Direcionais</b>	<b>63</b>
B.1 Demonstração das Propriedades da Função Característica $\phi_Y(p)$ . . . . .	63
B.2 Demonstração para a Desigualdade tipo Tchebychev . . . . .	65
<b>C Demonstração da Equação (4.10)</b>	<b>66</b>
C.1 Enunciado do Teorema da Transformação Integral . . . . .	66
C.2 Demonstração da Equação (4.10) . . . . .	66
<b>D Alguns resultados sobre as funções de Bessel</b>	<b>68</b>
D.1 Derivadas . . . . .	68
D.2 Fórmulas de recorrência . . . . .	69



	1
<b>E Programas Computacionais</b>	<b>70</b>
E.1 Comandos . . . . .	70
<b>Bibliografia</b>	<b>78</b>

# Introdução

Os modelos de regressão fornecem meios capazes de realizar previsões dos valores da variável resposta, através de informações sobre as variáveis explicativas, ou seja, eles nos permite investigar possíveis associações entre estas variáveis. No entanto, na prática, nem sempre encontramos fenômenos que nos forneçam apenas dados lineares, pois existem muitos exemplos, nas mais diversas áreas de conhecimento, em que os dados a serem modelados são também direcionais, como no Meio Ambiente, na Meteorologia, na Geologia, entre outras. Por isso, faz-se necessário o desenvolvimento de uma teoria apropriada para dados como esses. Portanto, quando se investiga associação envolvendo variáveis direcionais e variáveis lineares, é necessário levar em consideração os seguintes tipos de associações:

- **circular-circular** que mede o grau de associação entre duas variáveis direcionais;
- **circular-linear** que pode ser usada para prever valores de uma variável direcional a partir de valores de variáveis lineares;
- **linear-circular** que pode ser usada para prever valores de uma variável linear dado o valor de uma variável direcional.

Após a formulação do modelo, uma etapa de grande importância na análise de um ajuste de regressão é a aplicação das técnicas de diagnósticos, dentre elas a análise de resíduos que tem como objetivo detectar a presença de pontos extremos e avaliar a adequação da distribuição proposta para a variável resposta. Em se tratando de dados direcionais, faz-se necessário a utilização de resíduos convenientes, que destaquem as características desses dados.

Neste trabalho, consideramos a associação do tipo circular-linear, por isso foi conveniente considerar os modelos de regressão von Mises, por eles gozarem de boas propriedades. Apresentamos dois resíduos padronizados que foram obtidos a partir da componente da função desvio e cujas distribuições de probabilidades podem ser comparadas com a distribuição normal padrão.

A dissertação encontra-se estruturada da seguinte forma.

No Capítulo 1, apresentamos os principais conceitos relacionados com os dados direcionais e suas estatísticas resumo.

No Capítulo 2, mostramos as principais famílias dos modelos probabilísticos para os dados direcionais com as principais distribuições que ocorrem sobre o círculo.

No Capítulo 3, discutimos os modelos de regressão von Mises.

No Capítulo 4, apresentamos duas caracterizações para os resíduos de dados direcionais.

No Capítulo 5, fazemos uma aplicação a respeito da associação entre direções de ventos e concentração de ozônio na atmosfera.

# Capítulo 1

## Dados Direcionais

Neste capítulo introduzimos os conceitos relacionados com os dados direcionais, bem como apresentar as principais estatísticas resumo a esse tipo de observação.

### 1.1 Introdução

Em diversas áreas de conhecimento há o interesse por medidas angulares. Podemos citar como exemplo a Geologia, Biologia, Meteorologia, Medicina, Física, Engenharia, Meio Ambiente, entre outras. Alguns exemplos típicos de dados direcionais nestas áreas são:

- (i) Direção de ventos e correntes marinhas;
- (ii) Direção de migrações de animais;
- (iii) Orientação da fratura de uma rocha;
- (iv) Estudos de mapas mentais;
- (v) Transporte de materiais geológicos.

Observações considerando direções podem ser interpretadas como dados direcionais. Cada observação direcional pode ser representada, geometricamente, como um ponto no círculo unitário, no caso bidimensional, ou como um ponto sobre a superfície de uma esfera, no caso tridimensional, desde que a orientação e a direção inicial sejam

definidas. Na verdade, os dados direcionais podem ser tratados em um espaço de qualquer dimensão. No entanto, neste trabalho, nosso interesse será trabalhar no espaço de dimensão 2.

Os dados direcionais podem ser expressos em qualquer medida de ângulo, como por exemplo: graus, grado, radiano, etc. Mas, por convenção, trabalha-se com os dados em radianos, com intervalos de variação  $[0, 2\pi)$  ou  $(-\pi, \pi]$ , dependendo do interesse da análise.

Uma observação de uma variável circular é representada por um ponto  $P$  sobre a circunferência de um círculo unitário centrado na origem do sistema de coordenadas cartesianas e, portanto, tem-se que  $\overrightarrow{OP}$  é um vetor unitário em  $\mathbb{R}^2$ . Podemos ver esta observação ainda como um ângulo  $y$ , formado pelo vetor  $\overrightarrow{OP}$  e o semi-eixo positivo  $O_x$ , medido no sentido anti-horário. Sendo assim, uma observação de uma variável circular pode ser representada pelo par ordenado  $(\cos(y), \sin(y))$ , em coordenadas cartesianas. Existem casos de observações em eixos, em que não é possível distinguir qual é o início ou o fim da observação, como por exemplo: o estudo de fraturas em rochas. Esse tipo de dado é denominado *axial*, diferentemente do anterior, denominado *vetorial* (Fisher, 1995). Para se trabalhar com dados axiais deve-se, inicialmente, convertê-los a dados vetoriais, transformando a observação  $y$  em  $2y$ , ou seja, duplicando o ângulo (Mardia e Jupp, 1999).

## 1.2 Representação Gráfica de Dados Direcionais

Os dados direcionais, assim como os dados lineares, podem ser representados por diagramas e gráficos, com a finalidade de facilitar sua interpretação e análise. Quando os dados não são agrupados, sua representação consiste em corresponder a cada observação bruta um ponto no círculo unitário. No entanto, quando os dados são agrupados eles podem ser representados por uma das três maneiras seguintes: histograma circular, histograma linear ou diagrama de rosas.

Os histogramas circulares são formados por um conjunto de barras que representam as freqüências de cada classe, cujas bases estão sobre a circunferência do círculo unitário dividido de acordo com o tamanho de cada classe, com áreas proporcionais às freqüências. Se fizermos um corte no eixo deste histograma e o transportá-lo ao

eixo de coordenadas  $X$  iremos transformá-lo num histograma linear, semelhante ao de dados lineares.

O diagrama de rosas é um tipo de histograma circular formado por setores circulares cujas áreas são proporcionais às frequências de cada classe e os raios são proporcionais à raiz quadrada de cada frequência.

**Exemplo:** Consideremos o conjunto de dados representando 19 observações de direções de ventos, expresso em graus, da tabela abaixo. Representaremos os dados não-agrupados, e em seguida os agrupados em histograma circular e diagrama de rosas.

Tabela 1.1: Medidas de Direção de Ventos

Direção	327	8	91	204	88	86	305	333	344	18
	270	57	67	6	21	11	281	27	84	

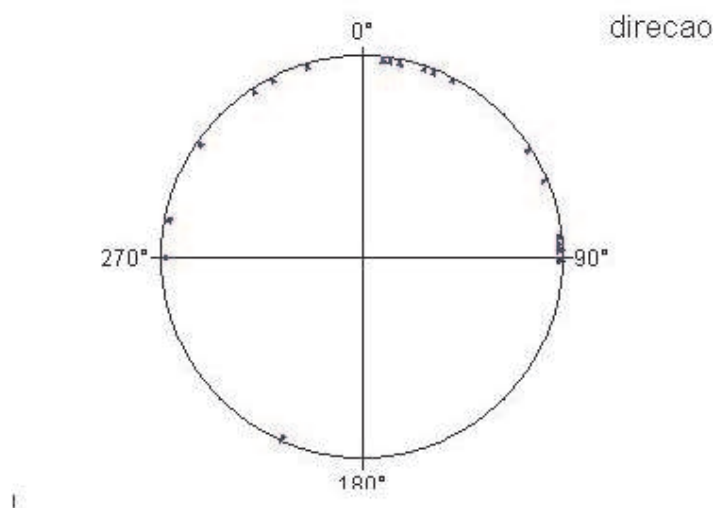


Figura 1.1: Representação Gráfica dos Dados Não-Agrupados

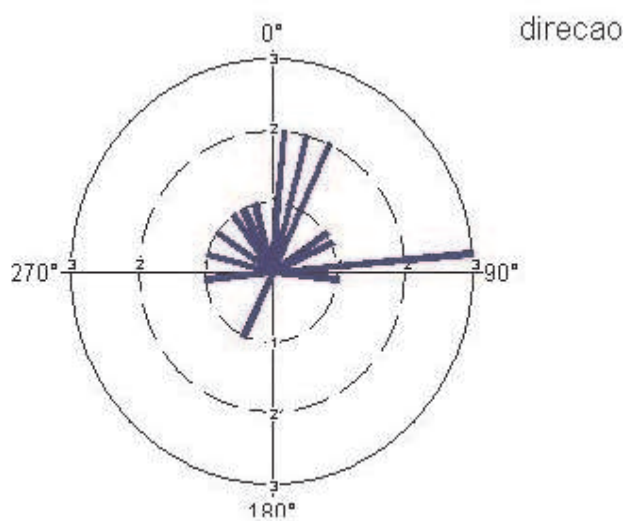


Figura 1.2: Representação Gráfica em Histograma Circular

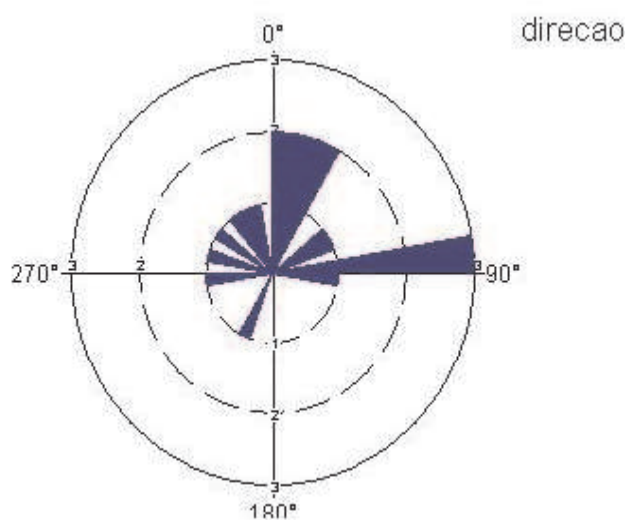


Figura 1.3: Representação Gráfica em Diagrama de Rosas

### 1.3 Medidas Descritivas no Círculo

Depois que os dados direcionais são coletados e representados graficamente, precisaremos estabelecer algumas estatísticas descritivas adequadas. A princípio, é tentador querermos utilizar as mesmas medidas empregadas para os dados lineares. Entretanto, pelo exemplo fictício, a seguir, podemos verificar a inadequação destas medidas no círculo.

**Exemplo Fictício:** Imaginemos uma amostra de tamanho 2 em que os ângulos

observados tenham sido  $y_1 = 1^0$  e  $y_2 = 359^0$  que são “próximos entre si” e equidistantes da direção  $0^0$ . Neste caso, a média aritmética amostral é  $\bar{y} = (1^0 + 359^0)/2 = 180^0$  que corresponde ao ponto no círculo diametralmente oposto ao resultado intuitivo. Em contra partida, o desvio padrão amostral fica dado por:  $S_{n-1} = 253^0$ , ou seja, um valor que não descreve o fato de  $y_1$  e  $y_2$  estarem “próximos entre si”.

Como as medidas resumo usadas para variáveis lineares não fazem sentido para as variáveis circulares, pelo que vimos no exemplo anterior, precisamos estabelecer algumas medidas descritivas relacionadas com o conceito de posição, dispersão e concentração que façam sentido para as variáveis circulares.

### 1.3.1 Medidas de Locação

#### 1.3.1.1 Média Direcional

Considere uma amostra aleatória de  $n$  observações circulares  $y_1, y_2, \dots, y_n$ , às quais podemos associar  $n$  vetores unitários  $\overrightarrow{OP}_1, \overrightarrow{OP}_2, \dots, \overrightarrow{OP}_n$ , conforme mencionamos anteriormente. Em termos algébricos, calculemos as quantidades

$$S = \sum_{i=1}^n \text{sen}(y_i), \quad C = \sum_{i=1}^n \text{cos}(y_i) \quad \text{e} \quad R^2 = C^2 + S^2.$$

A *média direcional* de  $y_1, y_2, \dots, y_n$  é definida como sendo o ângulo  $\bar{\mu}$ , correspondente ao vetor resultante da soma  $\overrightarrow{OP}_1 + \overrightarrow{OP}_2 + \dots + \overrightarrow{OP}_n$ . Esse ângulo deve satisfazer as condições  $\text{cos}(\bar{\mu}) = C/R$  e  $\text{sen}(\bar{\mu}) = S/R$ , ou seja,  $\bar{\mu} = \arctan(S/C)$  (Fisher, 1995). Assim,

$$\bar{\mu} = \begin{cases} \arctan(S/C), & \text{se } S \geq 0 \text{ e } C > 0 \\ \arctan(S/C) + \pi, & \text{se } C < 0 \\ \arctan(S/C) + 2\pi, & \text{se } S < 0 \text{ e } C > 0. \end{cases} \quad (1.1)$$

A quantidade  $R = \sqrt{C^2 + S^2}$  representa o comprimento do vetor resultante e fornece informação sobre o grau de concentração dos ângulos observados. Em vez de  $R$ , é mais comum usar o comprimento médio do vetor resultante, definido por  $\bar{R} = R/n$ , o qual tem a vantagem de variar no intervalo  $[0,1]$ . Portanto,  $\bar{\mu}$  pode ser determinado também como soluções das equações

$$\bar{C} = \bar{R} \cos \bar{\mu} \quad (1.2)$$

$$\bar{S} = \bar{R} \text{sen} \bar{\mu} \quad (1.3)$$



onde

$$\bar{S} = \frac{S}{n} \quad \text{e} \quad \bar{C} = \frac{C}{n}. \quad (1.4)$$

Das expressões (1.2), (1.3) e (1.4), podemos mostrar (vide Apêndice A) que

$$\frac{1}{n} \sum_{i=1}^n \cos(y_i - \bar{\mu}) = \bar{R} \quad \text{e} \quad (1.5)$$

$$\frac{1}{n} \sum_{i=1}^n \text{sen}(y_i - \bar{\mu}) = 0. \quad (1.6)$$

Para dados axiais, a média axial é definida da seguinte forma: Sejam  $y_1, y_2, \dots, y_n$  ângulos que representam observações de dados axiais e seja  $\bar{\mu}$  a média direcional dos ângulos vetoriais correspondentes. Então a média axial é dada por  $\frac{\bar{\mu}}{2}$  e  $\frac{\bar{\mu}}{2} + \pi$  (Mardia e Jupp, 1999).

### 1.3.1.1 Mediana Direcional

**Definição:** A mediana direcional  $\tilde{Y}$  dos ângulos  $y_1, y_2, \dots, y_n$  é um ângulo  $\phi$  que satisfaça as seguintes condições:

- (i) A metade dos pontos situam-se no arco  $[\phi, \phi + \pi)$ ;
- (ii) A maioria dos pontos são mais próximos de  $\phi$  do que  $\phi + \pi$ .

Para  $n$  ímpar, a mediana é um único ângulo e, quando  $n$  for par a mediana é a média aritmética entre os dois ângulos centrais.

## 1.3.2 Medidas de Concentração

### 1.3.2.1 O Comprimento Médio do Vetor Resultante

Como definido anteriormente,  $\bar{R} = R/n$  onde  $R$  é o comprimento do vetor resultante  $\vec{OP}_1 + \vec{OP}_2 + \dots + \vec{OP}_n$ . Como cada  $\vec{OP}_i$  é unitário, temos

$$0 \leq \| \vec{OP}_1 + \vec{OP}_2 + \dots + \vec{OP}_n \| \leq \| \vec{OP}_1 \| + \dots + \| \vec{OP}_n \| = 1 + \dots + 1 = n.$$

Portanto,

$$0 \leq R \leq n \Rightarrow 0 \leq R/n \leq 1 \Rightarrow 0 \leq \bar{R} \leq 1.$$

A quantidade  $\bar{R}$  fornece informações sobre o grau de concentração dos ângulos observados. No caso em que todas as observações forem iguais a um mesmo ângulo,

digamos  $y$ , tem-se que  $\bar{R} = 1$ , representando uma concentração máxima dos dados, e corresponde a uma variabilidade mínima entre eles. Outro caso extremo é quando os dados encontram-se uniformemente distribuídos no círculo, nesta situação tem-se uma concentração mínima ( $\bar{R} = 0$ ) e uma variabilidade máxima. É importante ressaltarmos que  $\bar{R} = 0$  não significa uma distribuição uniforme no círculo, pode acontecer algum tipo de estrutura de agrupamento, como o exemplo citado em Fisher (1995, p.32).

### 1.3.3 Medidas de Dispersão

#### 1.3.3.1 A Variância Circular

A variância circular amostral é definida por

$$V = 1 - \bar{R}. \quad (1.7)$$

O conceito é análogo à variância de dados lineares, no sentido de que quanto menor o valor da variância circular, maior a concentração nos dados. Observe que

$$0 \leq V \leq 1, \quad \text{uma vez que } 0 \leq \bar{R} \leq 1.$$

#### 1.3.3.2 O Desvio Padrão Circular

Para dados direcionais, a definição de desvio padrão é diferente da definição para dados lineares, que corresponde à raiz quadrada da variância. O desvio padrão circular amostral é definido por

$$\sigma = \{-2 \log(1 - V)\}^{1/2} = \{-2 \log \bar{R}\}^{1/2}.$$

Uma boa aproximação para o desvio padrão circular amostral é

$$\sigma = \begin{cases} (2V)^{1/2}, & \text{para } V \text{ pequeno } (V < 0,18) \\ \{2(1 - \bar{R})\}^{1/2}, & \text{para } \bar{R} \text{ grande } (\bar{R} > 0,82) \end{cases}$$

### 1.3.4 Outras Medidas de Dispersão

#### 1.3.4.1 A Dispersão Circular Amostral

A dispersão circular amostral é definida por

$$\hat{\delta} = \frac{1 - \bar{R}_2}{2\bar{R}^2}, \quad (1.8)$$

onde  $\bar{R}_2$  é o comprimento médio resultante dos ângulos duplicados  $2y_1, \dots, 2y_n$ . Com  $\bar{R}_2 = \sqrt{\bar{C}_2^2 + \bar{S}_2^2}$  onde  $\bar{C}_2 = \frac{1}{n} \sum_{i=1}^n \cos(2y_i)$  e  $\bar{S}_2 = \frac{1}{n} \sum_{i=1}^n \sin(2y_i)$ .

### 1.3.4.2 O Desvio Médio Circular

Uma alternativa para medir a distância entre dois ângulos  $\alpha$  e  $\beta$  é

$$\min((\alpha - \beta), 2\pi - (\alpha - \beta)) = \begin{cases} |\alpha - \beta|, & \text{se } |\alpha - \beta| \leq \pi \\ 2\pi - |\alpha - \beta|, & \text{se } |\alpha - \beta| > \pi \end{cases}$$

Desta forma, podemos escrever

$$\min((\alpha - \beta), 2\pi - (\alpha - \beta)) = \begin{cases} \pi - \pi + |\alpha - \beta|, & \text{se } \pi - |\alpha - \beta| \geq 0 \\ \pi + \pi - |\alpha - \beta|, & \text{se } \pi - |\alpha - \beta| < 0 \end{cases}$$

Então,

$$\min((\alpha - \beta), 2\pi - (\alpha - \beta)) = \begin{cases} \pi - (\pi - |\alpha - \beta|), & \text{se } \pi - |\alpha - \beta| \geq 0 \\ \pi - [-(\pi - |\alpha - \beta|)], & \text{se } \pi - |\alpha - \beta| < 0 \end{cases}$$

Portanto,

$$\min((\alpha - \beta), 2\pi - (\alpha - \beta)) = \pi - |\pi - |\alpha - \beta||.$$

Com isto, uma medida de dispersão dos ângulos  $y_1, \dots, y_n$  em torno de um dado ângulo  $\alpha$  é

$$d_0(\alpha) = \frac{1}{n} \sum_{i=1}^n \{\pi - |\pi - |y_i - \alpha||\} = \pi - \frac{1}{n} \sum_{i=1}^n |\pi - |y_i - \alpha||$$

A mediana amostral  $\tilde{y}$  é o ponto que minimiza  $d_0$  (Mardia e Jupp, 1999), logo, o desvio médio circular é definido por

$$d_0(\tilde{y}) = \frac{1}{n} \sum_{i=1}^n \{\pi - |\pi - |y_i - \tilde{y}||\}. \quad (1.9)$$

### 1.3.4.3 A Diferença Circular Média

A diferença circular média é a distância média entre observações circulares é dada por

$$\bar{D}_0 = \frac{1}{n} [d_0(y_1) + d_0(y_2) + \dots + d_0(y_n)] = \frac{1}{n} \sum_{j=1}^n d_0(y_j).$$

Portanto,

$$\bar{D}_0 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \{\pi - |\pi - |y_i - y_j||\}, \quad (1.10)$$

que representa a distância média entre pares de observações.

### 1.3.4.4 A Amplitude Circular

A amplitude circular é o comprimento do menor arco que contém todas as observações. Para determinarmos a distância circular inicialmente consideraremos os ângulos  $y_1, \dots, y_n$  de forma ordenada  $y_{(1)} \leq \dots \leq y_{(n)}$ . Os comprimentos de arco entre observações adjacentes são

$$T_i = \begin{cases} y_{(i+1)} - y_{(i)}, & i = 1, \dots, n-1 \\ 2\pi - y_{(n)} + y_{(1)}, & i = n \end{cases}$$

Assim, definimos a amplitude circular  $w$  por:

$$w = 2\pi - \max(T_1, \dots, T_n) \quad (1.11)$$

## 1.4 Momentos Trigonométricos

As expressões  $\bar{S}$  e  $\bar{C}$  definidos em (1.4), são importantes para determinação da média direcional e variância amostral, como visto anteriormente. Definimos o primeiro momento trigonométrico em torno de  $0^0$  por

$$m'_1 = \bar{C} + i\bar{S}. \quad (1.12)$$

Observe que, de (1.2) e (1.3)

$$m'_1 = \bar{R} \cos \bar{\mu} + i\bar{R} \sin \bar{\mu} = \bar{R}(\cos \bar{\mu} + i \sin \bar{\mu}) = \bar{R}e^{i\bar{\mu}}.$$

Por extensão, definimos o  $p$ -ésimo momento trigonométrico em torno de  $0^0$ , como sendo

$$m'_p = a_p + ib_p, \quad \text{para } p = 1, 2, \dots \quad (1.13)$$

em que

$$a_p = \frac{1}{n} \sum_{i=1}^n \cos(py_i) \quad \text{e} \quad b_p = \frac{1}{n} \sum_{i=1}^n \sin(py_i).$$

Portanto,

$$m'_p = \bar{R}_p e^{i\bar{\mu}_p}, \quad (1.14)$$

onde  $\bar{\mu}_p$  e  $\bar{R}_p$  denotam a média direcional e o comprimento médio do vetor resultante amostral de  $py_1, \dots, py_n$ , respectivamente. o  $p$ -ésimo momento trigonométrico em torno da média direcional  $\bar{\mu}$  é

$$m_p = \bar{a}_p + i\bar{b}_p, \quad (1.15)$$

onde

$$\bar{a}_p = \frac{1}{n} \sum_{j=1}^n \cos[p(y_j - \bar{\mu})] \quad \text{e} \quad \bar{b}_p = \frac{1}{n} \sum_{j=1}^n \text{sen}[p(y_j - \bar{\mu})].$$

Em particular, temos

$$m_1 = \bar{a}_1 + i\bar{b}_1 = \frac{1}{n} \sum_{j=1}^n \cos[p(y_j - \bar{\mu})] + i \frac{1}{n} \sum_{j=1}^n \text{sen}[p(y_j - \bar{\mu})].$$

Dos resultados, (1.5) e (1.6) temos,

$$m_1 = \bar{R} + i.0 = \bar{R}.$$

Posteriormente, apresentaremos as versões populacionais dos momentos trigonométricos, que desempenham importante papel na teoria de distribuições de probabilidade sobre o círculo.

## 1.5 Conceitos Teóricos Básicos para Dados Direcionais

### 1.5.1 A Função de Distribuição

Entendemos por uma variável aleatória circular como sendo uma variável aleatória já convertida em ângulos e assumindo valores no intervalo  $[0, 2\pi)$ .

**Definição:** Definimos a função de distribuição de uma variável aleatória circular como sendo uma função  $F$  tal que

$$\begin{aligned} F(y) &= P(0 < Y \leq y), \quad \forall y \in [0, 2\pi) \\ F(y + 2\pi) - F(y) &= 1, \quad \forall y \in \mathbb{R}. \end{aligned} \tag{1.16}$$

Sejam  $y_1$  e  $y_2$  satisfazendo  $-\infty < y_1 < y_2 < \infty$ , com  $y_2 - y_1 < 2\pi$ . A probabilidade de  $Y$  pertencer ao arco  $(y_1, y_2)$  é dada por

$$P(y_1 < Y \leq y_2) = F(y_2) - F(y_1) = \int_{y_1}^{y_2} dF(y), \tag{1.17}$$

onde a integral é a integral de Lebesgue-Stieltjes.

De fato, (1.17) é verdadeira, pois para  $y_1 < y_2$  temos que,

$$\begin{aligned} P(Y \leq y_2) &= P(Y \leq y_1) + P(y_1 < Y \leq y_2) \\ \Rightarrow P(y_1 < Y \leq y_2) &= P(Y \leq y_2) - P(Y \leq y_1). \end{aligned}$$

Logo,

$$P(y_1 < Y \leq y_2) = F(y_2) - F(y_1).$$

**Observação:** Embora  $F(0) = 0$  e  $F(2\pi) = 1$ , para as variáveis aleatórias circulares, a função de distribuição no círculo não satisfaz as condições  $\lim_{x \rightarrow -\infty} F(x) = 0$  e  $\lim_{x \rightarrow \infty} F(x) = 1$ , que são satisfeitas por funções de distribuições de variáveis aleatórias lineares.

**Definição:** Seja  $F$  a função de distribuição de uma variável aleatória circular  $Y$ . Se  $F$  for absolutamente contínua, definimos a função densidade de probabilidade de  $Y$  como sendo a função  $f$  tal que

$$\int_{y_1}^{y_2} f(y) dy = F(y_2) - F(y_1), \quad -\infty < y_1 \leq y_2 < \infty. \quad (1.18)$$

**Observação:** Uma função  $f$  é função de densidade de probabilidade de uma distribuição absolutamente contínua se, e somente se:

- (i)  $f(y) \geq 0$ , quase sempre em  $(-\infty, \infty)$ ;
- (ii)  $f(y + 2\pi) = f(y)$ , quase sempre em  $(-\infty, \infty)$ ;
- (iii)  $\int_0^{2\pi} f(y) dy = 1$ .

**Definição:** Sejam  $Y$  uma variável aleatória circular. Os momentos circulares de  $Y$  são dados por

$$\alpha_p = E[\cos(pY)] = \int_0^{2\pi} \cos(py) dF(y) \quad (1.19)$$

$$\beta_p = E[\text{sen}(pY)] = \int_0^{2\pi} \text{sen}(py) dF(y) \quad (1.20)$$

onde  $p \in \mathbb{Z}$ . Como conseqüência das propriedades das funções trigonométricas tem-se que

$$\alpha_p = \alpha_{-p}, \quad \beta_p = -\beta_{-p}, \quad |\alpha_p| \leq 1 \text{ e } |\beta_p| \leq 1.$$

### 1.5.2 A Função Característica

**Definição:** Definimos a função característica de uma variável aleatória circular  $Y$  como sendo o conjunto de números complexos  $\{\phi_Y(p) : p \in \mathbb{R}\}$  dada por

$$\phi_Y(p) = E(e^{ipY}) = \int_0^{2\pi} e^{ipY} dF(y), \quad \text{com } p \in \mathbb{R}. \quad (1.21)$$

Podemos escrever (1.21) da forma,

$$\phi_Y(p) = \alpha_p + i\beta_p,$$

onde  $\alpha_p$  e  $\beta_p$  são os momentos circulares de  $Y$  definidos em (1.19) e (1.20). Como está mostrado no Apêndice B, a função característica goza das seguintes propriedades:

**Propriedades:**

- (i)  $\phi_Y(0) = 1$
- (ii)  $\overline{\phi_Y(p)} = \phi_{-Y}(p)$  onde  $\overline{\phi_Y(p)}$  é o conjugado complexo de  $\phi_Y(p)$
- (iii)  $|\phi_Y(p)| \leq 1$ ,
- (iv) É suficiente considerarmos a função  $\phi_Y(p)$  apenas para valores inteiros e positivos de  $p$ .

### 1.5.3 Independência e Convolução

**Definição:** Sejam  $Y_1$  e  $Y_2$  duas variáveis aleatórias circulares. A função característica conjunta  $(Y_1, Y_2)$  é definida pela seqüência dupla  $\{\phi_{Y_1 Y_2}(p, q)$  com  $p, q \in \mathbb{Z}\}$  dada por

$$\phi_{Y_1 Y_2}(p, q) = E(e^{ipY_1 + iqY_2}). \quad (1.22)$$

**Definição:** As variáveis  $Y_1$  e  $Y_2$  são independentes se, e somente se,  $\phi_{Y_1 Y_2}(p, q) = \phi_{Y_1}(p)\phi_{Y_2}(q)$  para todo  $p, q \in \mathbb{Z}$ .

Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias circulares independentes. A função característica da soma  $S_n = \sum_{i=1}^n Y_i$  é análoga ao caso linear. Se as variáveis forem identicamente distribuídas temos que

$$\phi_{S_n}(p) = \prod_{i=1}^n \phi_{Y_i}(p) = \phi_{Y_1}(p) \cdot \dots \cdot \phi_{Y_n}(p) = [\phi_Y(p)]^n.$$

**Definição:** Sejam  $Y_1$  e  $Y_2$  variáveis aleatórias circulares independentes com respectivas funções de distribuição  $F_1$  e  $F_2$  e considere  $S = Y_1 + Y_2$ . A função de distribuição  $F$  de  $S$  é dada por

$$dF_S(s) = \int_0^{2\pi} dF_2(s - y_1) dF_1(y_1).$$

### 1.5.4 Momentos e Medidas de Locação e Dispersão

Vamos considerar as versões populacionais das medidas resumo e dos momentos trigonométricos que foram apresentadas nas seções (1.3) e (1.4).

#### 1.5.4.1 Momentos Trigonométricos

Sejam  $\alpha_p = E[\cos(pY)]$  e  $\beta_p = E[\sin(pY)]$ , com  $p \in \mathbb{Z}$ , os momentos trigonométricos de uma variável aleatória circular  $Y$ . A sequência  $\{(\alpha_p, \beta_p), p \in \mathbb{Z}\}$  caracteriza a função característica de  $Y$ . Portanto, no círculo, qualquer distribuição é determinada pelos seus momentos trigonométricos.

Para  $p \geq 0$  temos que a versão populacional do  $p$ -ésimo momento em torno de  $0^0$  é dada por

$$\phi_Y(p) = \rho_p e^{i\mu_p}, \quad \rho_p \geq 0 \quad (1.23)$$

onde  $\rho_p$  é o comprimento médio do vetor resultante. O  $p$ -ésimo momento trigonométrico populacional em torno da média direcional é definido por

$$\phi_Y(p) = \bar{\alpha}_p + i\bar{\beta}_p, \quad (1.24)$$

onde,  $\bar{\alpha}_p = E\{\cos[p(Y - \bar{\mu})]\}$  e  $\bar{\beta}_p = E\{\sin[p(Y - \bar{\mu})]\}$ . Logo, a equação (1.24) pode ser escrito como

$$\phi_Y(p) = E[e^{ip(Y - \bar{\mu})}].$$

#### 1.5.4.2 Medidas de Locação e Dispersão

##### 1.5.4.2.1 A Média Direcional e o Comprimento Médio Resultante

Para  $p = 1$ , em (1.23), temos

$$\phi_Y(1) = \rho_1 e^{i\bar{\mu}_1} = \rho e^{i\bar{\mu}},$$

onde  $\mu$  é a versão populacional da média direcional  $\bar{\mu}$  e  $\rho$  é a versão populacional do comprimento médio resultante  $\bar{R}$  e pode ser calculado pela expressão  $\rho = (\alpha_1^2 + \beta_1^2)^{1/2}$ .

Observe que, ao considerarmos o efeito de rotação por  $-\psi$  na variável circular  $Y$ , temos  $Y^* = Y - \psi$ . A média direcional de  $Y^*$  é  $\mu^* = \mu - \psi$ , que é equivariante sob rotação.

Então

$$\phi_Y^*(p) = e^{-ip\psi} \phi_Y(p). \quad (1.25)$$



De fato,

$$\phi_Y^*(p) = E(e^{ipY^*}) = E(e^{ip(Y-\psi)}) = e^{-ip\psi} E(e^{ipY}) = e^{-ip\psi} \phi_Y(p).$$

Considere  $p = 1$  em (1.25), então

$$\phi_Y^*(1) = e^{-i\psi} \phi_Y(1) = e^{-i\mu} \rho e^{i\mu} = \rho + i0.$$

Mas,

$$\phi_Y^*(1) = E[\cos(Y - \mu)] + iE[\text{sen}(Y - \mu)].$$

Logo,

$$E[\cos(Y - \mu)] + iE[\text{sen}(Y - \mu)] = \rho + i0.$$

Portanto,

$$E[\cos(Y - \mu)] = \rho, \text{ e} \tag{1.26}$$

$$E[\text{sen}(Y - \mu)] = 0. \tag{1.27}$$

#### 1.5.4.2.2 A Variância Circular

**Definição:** A variância circular  $\nu$ , de uma variável aleatória circular  $Y$  é definida como

$$\nu = 1 - \rho. \tag{1.28}$$

De (1.26) temos,

$$\nu = 1 - E[\cos(Y - \mu)].$$

E, analogamente à situação amostral,  $0 \leq \nu \leq 1$ .

#### 1.5.4.2.3 O Desvio Padrão Circular

A versão populacional do desvio padrão circular é dada por

$$\sigma = \{-2 \log(1 - \nu)\}^{1/2} = \{-2 \log \rho\}^{1/2}. \tag{1.29}$$

Esta definição é motivada pelo fato de que a distribuição normal ao redor do círculo gera a distribuição normal arqueada, como veremos posteriormente.

#### 1.5.4.2.4 A Dispersão Circular

A versão populacional da dispersão circular é

$$\delta = \frac{1 - \rho_2}{2\rho_2^2}, \quad (1.30)$$

onde  $\rho_2 = \sqrt{\alpha_2^2 + \beta_2^2}$ , com  $\alpha_2 = E[\cos(2Y)]$  e  $\beta_2 = E[\sin(2Y)]$ .

#### 1.5.4.2.5 A Mediana Circular

A versão populacional da mediana circular  $\tilde{Y}$  é a mediana direcional  $\tilde{\mu}$  definida como a direção  $\phi$  que minimiza  $E[\pi - |\pi - |Y - \phi||]$  e satisfaz

$$P(Y \in (\tilde{\mu}, \tilde{\mu} + \pi]) \geq \frac{1}{2} \text{ e } P(Y \in (\tilde{\mu} - \pi, \tilde{\mu})) \geq \frac{1}{2}.$$

#### 1.5.5 A Desigualdade Tipo Tchebyshev

Várias desigualdades para distribuições são observadas no círculo, muitas delas foram estudadas por Marshal & Olkin (1961). Usando a desigualdade de Tchebyshev na variável aleatória  $\text{sen}(\frac{Y-\mu}{2})$ , temos:

$$P(|\text{sen}(\frac{Y-\mu}{2})| \geq \varepsilon) \leq \frac{\nu}{2\varepsilon^2}, \quad 0 < \varepsilon \leq 1. \quad (1.31)$$

A demonstração encontra-se no Apêndice B.

## Capítulo 2

# Modelos Probabilísticos para Dados Circulares

Neste capítulo, nosso objetivo é fazer um estudo das duas principais famílias de modelos probabilísticos para dados direcionais. Os chamados *modelos exponenciais* e os *modelos de transformação*. Além disso, apresentar algumas das mais importantes distribuições sobre o círculo, dando um destaque à distribuição von Mises.

### 2.1 Introdução

Em inferência estatística, a análise de modelos probabilísticos adequados é de grande relevância, tanto para dados lineares quanto para dados direcionais. A maioria dos modelos para dados direcionais pertence a uma das duas classes de modelos paramétricos, ou os modelos exponenciais ou os modelos de transformação (Mardia e Jupp, 1999). Assim sendo, faremos um estudo das principais propriedades dessas duas classes de modelos. Dentre as famílias de distribuições no círculo, temos a distribuição uniforme circular como a mais simples e a distribuição von Mises como a mais importante, pois desempenha um papel em inferência no círculo semelhante à distribuição normal na reta. Apresentaremos também, outras distribuições circulares importantes que são as distribuições arqueada.

## 2.2 Modelo Exponencial

**Definição:** Um modelo é do tipo exponencial se sua função densidade de probabilidade for da forma

$$f(x; \boldsymbol{\omega}) = b(x) \exp\{\boldsymbol{\phi}(\boldsymbol{\omega})' \mathbf{t}(x) - \psi(\boldsymbol{\omega})\}, \quad x \in \chi \text{ e } \boldsymbol{\omega} \in \Omega, \quad (2.1)$$

com relação a algum domínio de medida  $\lambda$ , onde  $\chi$  é o espaço amostral e  $\Omega$  é o espaço paramétrico. A estatística  $\mathbf{t} : \chi \rightarrow \mathbb{R}^m$  é a estatística canônica e  $\boldsymbol{\phi}(\boldsymbol{\omega})$  é o parâmetro canônico ( $\boldsymbol{\phi} : \Omega \rightarrow \mathbb{R}^m$ ).

**Observação:** Segundo Mardia e Jupp (1999), considere  $d = \dim[\boldsymbol{\phi}(\Omega)]$ .

- (i) Se (2.1) for minimal, então o modelo é chamado um modelo exponencial  $(m, d)$ ;
- (ii) Se  $d < m$ , o modelo é chamado modelo exponencial curvo;
- (iii) Se  $d = m$  podemos fazer uma reparametrização para o modelo exponencial (2.1).

Considere  $\theta_i = \phi_i(\boldsymbol{\omega})$  para  $i = 1, \dots, m$ , então (2.1) pode ser escrita na forma

$$f(x; \boldsymbol{\theta}) = b(x) \exp\{\boldsymbol{\theta}' \mathbf{t}(x) - \psi(\boldsymbol{\theta})\}, \quad (2.2)$$

onde  $\boldsymbol{\theta}$  varia em um subconjunto  $\Theta \subset \mathbb{R}^m$ . Quando  $\Theta$  for um aberto em  $\mathbb{R}^m$ , e  $\psi$  pertencer ao conjunto das funções integráveis, então o modelo é chamado de *Modelo Exponencial Regular*.

### Principais Propriedades dos Modelos Exponenciais Regulares

(1) Os primeiros momentos da estatística canônica são dados por:

$$E_{\boldsymbol{\theta}}(\mathbf{t}) = \frac{\partial \psi}{\partial \boldsymbol{\theta}} \quad \text{e} \quad \text{Var}_{\boldsymbol{\theta}}(\mathbf{t}) = \frac{\partial^2 \psi}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$$

(2) A matriz inversa de Informação de Fisher é  $\text{Var}_{\boldsymbol{\theta}}(\mathbf{t})$ .

(3) Se  $X_1, \dots, X_n$  forem independentes e identicamente distribuídas com função de densidade de probabilidade dada por (2.2), então a distribuição amostral de  $\bar{\mathbf{t}}$  (média amostral da estatística  $\mathbf{t}$ ) tem função densidade de probabilidade proporcional a

$$\exp\{n[\boldsymbol{\theta}' \bar{\mathbf{t}}(x) - \psi(\boldsymbol{\theta})]\},$$

e  $\bar{t}$  é suficiente para  $\theta$ .

(4) O estimador de máxima verossimilhança  $\hat{\theta}$  de  $\theta$  é único e é dado pela solução da equação

$$E_{\hat{\theta}}(\mathbf{t}) = \bar{t}.$$

## 2.3 Modelos de Transformação

**Definição:** O modelo de transformação é um modelo em que um grupo  $G$  atua tanto sobre o espaço amostral  $\chi$  quanto sobre o espaço paramétrico  $\Omega$ , de modo que, se a variável aleatória  $X$  tem função densidade de probabilidade  $f(x, \omega)$ , então a variável aleatória  $gX \in G$  tem função densidade de probabilidade  $f(gx, g\omega)$  em que

$$f(gx, g\omega) = f(x, \omega)\chi(g, x), \quad (2.3)$$

para algumas funções  $g$  sobre  $G \times \chi$ .

Em outras palavras, um modelo de transformação é um modelo composto de transformação em que o grupo  $G$  age transitivamente sobre o espaço paramétrico, isto é, para quaisquer  $\omega$  e  $\omega' \in \Omega$ , existe um  $g \in G$  tal que  $g\omega = \omega'$ . Em particular, quando o grupo  $G$  é o grupo aditivo  $\mathbb{R}$  e a medida é a medida de Lebesgue, então (2.3) fica da forma

$$f(x + g, \omega + g) = f(x, \omega). \quad (2.4)$$

Para nossos propósitos, a principal propriedade dos modelos compostos de transformação é que o estimador de máxima verossimilhança  $\hat{\omega}$  de  $\omega$  é equivariante, isto é,

$$\hat{\omega}(gx_1, \dots, gx_n) = g\hat{\omega}(x_1, \dots, x_n).$$

## 2.4 A Distribuição Circular Uniforme - $\mathcal{U}$

Como mencionado anteriormente, a distribuição uniforme é a distribuição mais simples sobre o círculo. Ela é a única que tem a propriedade de ser equivariante sob rotação e reflexão (Mardia e Jupp, 1999).

**Definição:** Uma variável aleatória circular  $Y$  tem distribuição uniforme sobre o círculo se sua função densidade de probabilidade for da forma

$$f(y) = \frac{1}{2\pi}, \quad \alpha \leq y < \alpha + 2\pi, \quad \text{com } \alpha \in [0, 2\pi). \quad (2.5)$$

Observe que, se  $\alpha \leq \beta \leq \alpha + 2\pi$ , então

$$P_r(\alpha < Y \leq \beta) = F(\beta) - F(\alpha) = \int_{\alpha}^{\beta} f(y)dy.$$

Logo,

$$P_r(\alpha < Y \leq \beta) = \frac{1}{2\pi} y|_{\alpha}^{\beta} = \frac{\beta - \alpha}{2\pi},$$

ou seja, essa probabilidade é proporcional ao comprimento de arco.

A função característica da distribuição uniforme é dada por

$$\phi_Y(p) = \begin{cases} 0, & \text{se } p = \pm 1, \pm 2, \dots \\ 1, & \text{se } p = 0. \end{cases}$$

De fato, pela definição de função característica temos que

$$\phi_Y(p) = \int_0^{2\pi} e^{ipY} dF(y) = \int_0^{2\pi} e^{ipY} \frac{1}{2\pi} dy = \frac{1}{2\pi} \int_0^{2\pi} e^{ipY} dy.$$

Então, para  $p = 0$ , temos

$$\phi_Y(p) = \frac{1}{2\pi} \int_0^{2\pi} d(y) = 1.$$

Para  $p = \pm 1, \pm 2, \dots$ , temos

$$\phi_Y(p) = \frac{1}{2\pi} \frac{1}{ip} e^{ipy} \Big|_0^{2\pi} = \frac{e^{2ip\pi} - 1}{2ip\pi} = 0.$$

Portanto,

$$\phi_Y(p) = \begin{cases} 0, & \text{se } p = \pm 1, \pm 2, \dots \\ 1, & \text{se } p = 0 \end{cases}$$

## 2.5 As Distribuições Arqueadas

Muitos modelos probabilísticos para dados direcionais são derivados a partir de transformações de modelos probabilísticos lineares usuais. Em outras palavras, dada uma distribuição sobre a reta podemos arqueá-la em torno da circunferência do círculo de raio unitário e assim obter uma distribuição de probabilidade angular.

**Definição:** Seja  $X$  uma variável aleatória sobre a reta, definimos a variável aleatória  $X_w$  da distribuição arqueada por

$$X_w = X \{\text{mod}(2\pi)\}. \quad (2.6)$$

Se consideramos o círculo como o conjunto de números complexos de módulo unitário, a função de arqueamento  $X \mapsto X_\omega$  pode ser escrita da seguinte forma

$$X \mapsto e^{2\pi i X}.$$

Se  $X$  tem função de distribuição  $F_X$ , então a função de distribuição  $F_Y$  de  $Y$  é dada por

$$F_Y(y) = \sum_{k=-\infty}^{\infty} \{F(y + 2k\pi) - F(2k\pi)\}, \quad 0 \leq y < 2\pi. \quad (2.7)$$

Se  $X$  tem uma função densidade de probabilidade  $f$  então a função densidade de probabilidade correspondente  $f_Y$  de  $Y$  é

$$f_Y(y) = \sum_{k=-\infty}^{\infty} f(y + 2k\pi), \quad 0 \leq y \leq 2\pi. \quad (2.8)$$

**Propriedades:**

(i)  $(X + Z)_\omega = X_\omega + Z_\omega$

(ii) Se  $X$  tem função característica  $\phi$  então a função característica  $\phi_Y(p)$  de  $X_\omega$  é:  
 $\phi_Y(p) = \phi(p)$

(iii) Se  $\phi$  é integrável então  $Y$  tem densidade com  $\sum_{p=1}^{\infty} (\alpha_p^2 + \beta_p^2) < \infty$ , e

$$f_\omega(y) = \sum_{k=-\infty}^{\infty} f(y + 2k\pi) = \frac{1}{2\pi} \left[ 1 + 2 \sum_{p=1}^{\infty} (\alpha_p \cos py + \beta_p \operatorname{sen} py) \right]$$

### 2.5.1 A Distribuição Normal Arqueada - $\mathcal{WN}(\mu, \sigma^2)$

Um caso especial dentre as distribuições arqueadas é aquela que se obtém a partir de uma distribuição Gaussiana, por meio do seu arqueamento sobre o círculo.

Se uma variável aleatória  $X$  tem distribuição normal  $\mathcal{N}(\mu, \sigma^2)$ , então a variável aleatória  $Y = X \{ \bmod (2\pi) \}$  tem distribuição normal arqueada com parâmetros  $\mu$  e  $\sigma^2$ . Sua função densidade de probabilidade é dada por

$$f_Y(y; \mu, \sigma^2) = \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y + 2k\pi - \mu)^2}{2\sigma^2}\right\}. \quad (2.9)$$

De fato, se  $X \sim \mathcal{N}(\mu, \sigma^2)$  sua função densidade de probabilidade é da forma

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}.$$

Então, de (2.8) a função densidade de probabilidade de  $Y$  é dada por

$$f_Y(y; \mu, \sigma^2) = \sum_{k=-\infty}^{\infty} f(y + 2k\pi) = \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y + 2k\pi - \mu)^2}{2\sigma^2}\right\}$$

A função característica de uma variável aleatória  $X \sim \mathcal{N}(\mu, \sigma^2)$  é  $\phi_X(t) = \exp\{i\mu t - t^2\sigma^2/2\}$ . Da propriedade (ii) da seção 2.5, segue-se que a função característica de  $Y$  é

$$\begin{aligned} \phi(p) &= \exp\{i\mu p - p^2\sigma^2/2\} \\ &= e^{i\mu p} e^{-p^2\sigma^2/2} \\ &= e^{i\mu p} e^{(-\sigma^2/2)p^2} \\ &= e^{i\mu p} \rho^{p^2} \\ &= [\cos(\mu p) + i\text{sen}(\mu p)]\rho^{p^2} \\ &= \rho^{p^2} \cos(\mu p) + i\rho^{p^2} \text{sen}(\mu p). \end{aligned}$$

Portanto,

$$\phi(p) = \alpha_p + i\beta_p, \quad (2.10)$$

onde

$$\begin{cases} \alpha_p &= \rho^{p^2} \cos(\mu p) \\ \beta_p &= \rho^{p^2} \text{sen}(\mu p). \end{cases} \quad (2.11)$$

Com este resultado podemos representar  $f_\omega$  de outra maneira, a saber,

$$f_\omega(y; \mu, \rho) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos p(y - \mu) \right\}. \quad (2.12)$$

Para verificarmos isso, basta ver que, da propriedade (iii) da seção 2.5, temos que

$$f_\omega(y) = \sum_{k=-\infty}^{\infty} f(y + 2k\pi) = \frac{1}{2\pi} \left[ 1 + 2 \sum_{p=1}^{\infty} (\alpha_p \cos(py) + \beta_p \text{sen}(py)) \right].$$

Logo, de (2.11), temos que

$$\begin{aligned} \alpha_p \cos(py) + \beta_p \text{sen}(py) &= \rho^{p^2} \cos(\mu p) \cos(py) + \rho^{p^2} \text{sen}(\mu p) \text{sen}(py) \\ &= e^{-p^2\sigma^2/2} \cos(\mu p) \cos(py) + e^{-p^2\sigma^2/2} \text{sen}(\mu p) \text{sen}(py) \\ &= e^{-p^2\sigma^2/2} \cos p(y - \mu). \end{aligned}$$

Logo,

$$f_\omega(y; \mu, \rho) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos[p(y - \mu)] \right\}.$$

**Conclusões:**



- Podemos determinar a função de distribuição  $F_\omega$  integrando, termo a termo, a expressão (2.12).
- Quando  $\sigma^2 \geq 2\pi$ ,  $f_\omega$  pode ser aproximada pelos três primeiros termos de (2.12).
- Quando  $\sigma^2 \leq 2\pi$ ,  $f_\omega$  pode ser aproximada pelo termo  $k = 0$  de (2.9).

**Observações:**

- (i) A distribuição normal arqueada é unimodal e simétrica em torno da moda  $\mu$ .
- (ii) Quando  $\rho \rightarrow 0$  a  $\mathcal{WN}(\mu, \sigma^2)$  tende à distribuição uniforme.
- (iii) Quando  $\rho \rightarrow 1$  a  $\mathcal{WN}(\mu, \sigma^2)$  tende à distribuição degenerada, concentrada em  $\mu$ .

### 2.5.2 A Distribuição Poisson Arqueada - $\mathcal{WP}(\lambda)$

A redução módulo  $2\pi$  “transforma” a reta num círculo. Da mesma forma, temos que se  $m \in \mathbb{Z}_+^*$  a redução módulo  $2\pi m$  “transforma” os inteiros no grupo das  $m$ -ésimas raízes de 1, visto como um subgrupo do círculo (Mardia e Jupp, 1999). Em particular, se  $X$  é uma variável aleatória sobre os inteiros, então  $X_\omega$ , definida por

$$X_\omega = 2\pi X \{\text{mod}(2\pi)\}, \quad (2.13)$$

é uma variável aleatória distribuída sobre o círculo. A função de probabilidade de  $X_\omega$  é:

$$P_r(X_\omega = \frac{2\pi r}{m}) = \sum_{k=-\infty}^{\infty} P_X(r + km), \quad r = 0, 1, \dots, m-1, \quad (2.14)$$

Desta forma, se  $X \sim \text{Poisson}(\lambda)$ ,  $X_\omega$  tem distribuição Poisson Arqueada com função de probabilidade dada por

$$P_r(X_\omega = \frac{2\pi r}{m}) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{r+km}}{(r+km)!}, \quad r = 0, 1, \dots, m-1, \quad (2.15)$$

e denotamos por  $X_\omega \sim \mathcal{WP}(\lambda)$ .

A função característica de  $X_\omega \sim \mathcal{WP}(\lambda)$  é dada por

$$\phi(p) = \exp\{-\lambda(1 - e^{2\pi ip/m})\}, \quad \text{onde } p = 0, 1, 2, \dots \quad (2.16)$$

Uma propriedade importante da distribuição Poisson Arqueada é que o conjunto das suas funções densidades de probabilidades é fechado com relação à soma.

De fato, isto ocorre, pois se considerarmos  $Y_1 \sim \mathcal{WP}(\lambda_1)$  e  $Y_2 \sim \mathcal{WP}(\lambda_2)$  independentes, então, a função característica de  $Y_1 + Y_2$  é dada por

$$\begin{aligned}\phi_{Y_1+Y_2}(p) &= \phi_{Y_1}(p)\phi_{Y_2}(p) \\ &= \exp\{-\lambda_1(1 - e^{2\pi ip/m})\} \exp\{-\lambda_2(1 - e^{2\pi ip/m})\} \\ &= \exp\{-(\lambda_1 + \lambda_2)(1 - e^{2\pi ip/m})\}.\end{aligned}$$

Portanto,  $Y_1 + Y_2 \sim \mathcal{WP}(\lambda_1 + \lambda_2)$ .

### 2.5.3 A Distribuição Cauchy Arqueada - $\mathcal{WC}(\mu, \rho)$

Na reta, a distribuição de Cauchy tem função densidade de probabilidade dada por

$$f(x; \mu, a) = \frac{1}{\pi} \frac{a}{a^2 + (x - \mu)^2}, \quad a > 0, -\infty < \mu < \infty, x \in \mathbb{R}.$$

onde  $\mu$  é a mediana (parâmetro de locação) e  $a$  representa a distância entre a mediana e o primeiro quartil (parâmetro de escala).

A função característica da distribuição de Cauchy é

$$\phi(t) = e^{-a|t|+it\mu}, \quad \text{onde } t \in \mathbb{R}$$

Daí, a função característica da distribuição de Cauchy Arqueada para  $p \in \mathbb{Z}$  é da forma

$$\begin{aligned}\phi(p) &= e^{-a|p|+ip\mu} \\ &= e^{-a|p|} e^{ip\mu} \\ &= e^{-a|p|} (\cos p\mu + i \operatorname{sen} p\mu).\end{aligned}$$

Logo,

$$\phi(p) = e^{-a|p|} (\cos p\mu) + e^{-a|p|} (i \operatorname{sen} p\mu) = \alpha_p + i\beta_p.$$

onde,

$$\begin{aligned}\alpha_p &= e^{-a|p|} \cos(p\mu) \quad \text{e} \\ \beta_p &= e^{-a|p|} \operatorname{sen}(p\mu).\end{aligned} \tag{2.17}$$

Assim, a função densidade de probabilidade da distribuição Cauchy Arqueada fica dada por

$$c(y; \mu, \rho) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{k=-\infty}^{\infty} \rho^k \cos p(y - \mu) \right\}. \quad (2.18)$$

Da propriedade (iii) da seção (2.5) temos que,

$$c(y; \mu, \rho) = \sum_{k=-\infty}^{\infty} f(y + 2k\pi; \mu, a) = \frac{1}{2\pi} \left[ 1 + 2 \sum_{p=1}^{\infty} (\alpha_p \cos py + \beta_p \sin py) \right].$$

Substituindo (2.17) na expressão acima, e considerando  $\rho = e^{-a}$ , segue-se que

$$\begin{aligned} c(y; \mu, \rho) &= \frac{1}{2\pi} \left\{ 1 + 2 \sum_{k=-\infty}^{\infty} e^{-a|k|} \cos p(y - \mu) \right\} \\ &= \frac{1}{2\pi} \left\{ 1 + 2 \sum_{k=-\infty}^{\infty} \rho^{|k|} \cos p(y - \mu) \right\}. \end{aligned}$$

### Propriedades:

- (i) A média direcional de  $\mathcal{WC}(\mu, \rho)$  é  $\mu \{ \text{mod } (2\pi) \}$  e o comprimento médio do vetor resultante é  $\rho$ ;
- (ii) A distribuição  $\mathcal{WC}(\mu, \rho)$  é unimodal e simétrica em torno de  $\mu$ ;
- (iii) Quando  $\rho \rightarrow 0$  a  $\mathcal{WC}(\mu, \rho)$  converge para a distribuição circular uniforme;
- (iv) Quando  $\rho \rightarrow 1$  a  $\mathcal{WC}(\mu, \rho)$  fica concentrada no ponto  $\mu$ ;
- (v) A função de distribuição da  $\mathcal{WC}(\mu, \rho)$  é dada por

$$F(y) - F(\mu) = \frac{1}{2\pi} \cos^{-1} \left\{ \frac{(1 + \rho^2) \cos(y - \mu) - 2\rho}{1 + \rho^2 - 2\rho \cos(y - \mu)} \right\};$$

- (vi) Tal como a Distribuição Poisson Arqueada, o conjunto das funções densidade de probabilidades das distribuições de Cauchy Arqueadas é fechado com relação à soma.

## 2.6 A Distribuição von Mises - $\mathcal{VM}(\mu, \lambda)$

**Definição:** Dizemos que uma variável aleatória circular  $Y$  tem distribuição von Mises com parâmetros  $\mu$  e  $\lambda$  se sua função densidade de probabilidade for dada por

$$f_Y(y; \mu, \lambda) = \frac{1}{2\pi I_0(\lambda)} e^{\lambda \cos(y - \mu)}, \quad 0 \leq y < 2\pi, \quad 0 \leq \mu < 2\pi, \quad \text{e } \lambda > 0, \quad (2.19)$$

onde,  $I_0(\lambda)$  é a função de Bessel modificada de primeiro tipo e ordem zero, que é dada pela série

$$I_0(\lambda) = \sum_{r=0}^{\infty} \frac{1}{r!^2} \left(\frac{1}{2}\lambda\right)^{2r}.$$

A função de distribuição de  $Y$ , é dada por

$$F_Y(y) = \frac{1}{2\pi I_0(\lambda)} \int_0^y e^{\lambda \cos(t-\mu)} dt, \quad (2.20)$$

que não admite uma forma fechada. Algumas características da distribuição von Mises envolvem a função  $A_p(\lambda) = \frac{I_p(\lambda)}{I_0(\lambda)}$ , onde  $I_p(\lambda)$  é a função de Bessel modificada do primeiro tipo e ordem  $p$ , que é definida pela série

$$I_p(\lambda) = \sum_{r=0}^{\infty} \frac{1}{(p+r)!r!} \left(\frac{1}{2}\lambda\right)^{2r+p}, \quad p = 0, 1, \dots$$

As propriedades desta função podem ser encontradas em Abramowitz e Stegun (1970). Na distribuição von Mises  $\mathcal{VM}(\mu, \lambda)$ , o parâmetro  $\mu$  representa a média circular de  $Y$  enquanto que  $\lambda$  é denominado parâmetro de concentração, e é tal que o comprimento resultante médio satisfaz  $\rho = A_1(\lambda)$ . A dispersão circular de  $Y$  é dada por  $\delta = [\lambda A_1(\lambda)]^{-1}$  e as componentes de seus momentos trigonométricos são  $\alpha_p = A_p(\lambda)$  e  $\beta_p = 0$ , para  $p \geq 1$ . A distribuição von Mises goza de propriedades bastante interessantes, dentre as quais, destacamos

- (i) Ela é unimodal e simétrica em torno da média direcional  $\mu$ ;
- (ii) Quando  $\lambda \rightarrow 0$ , a distribuição de von Mises converge para a distribuição circular uniforme;
- (iii) Quando  $\lambda \rightarrow \infty$ , a distribuição converge para a distribuição degenerada no ponto  $y = \mu$ ;
- (iv) Segundo Stephens (1963), para valores não muito pequenos de  $\lambda$ , a distribuição von Mises  $\mathcal{VM}(\mu, \lambda)$  é próxima de uma distribuição normal arqueada  $\mathcal{WN}(\mu, 1/\lambda)$ .
- (v) A razão entre a densidade na moda e a densidade na antimoda é  $e^{2\lambda}$ . Isto implica que, quanto maior for o valor de  $\lambda$  maior será o agrupamento em torno da moda.

### 2.6.1 Estimação dos parâmetros da Distribuição von Mises

Seja  $Y_1, \dots, Y_n$  uma amostra aleatória independente de distribuição  $\mathcal{VM}(\mu, \lambda)$ , com valores observados  $y_1, \dots, y_n$ . A função de log-verossimilhança fica dada por

$$l(\mu, \lambda; y_1, \dots, y_n) = -n \log 2\pi - n \log I_0(\lambda) + \lambda \sum_{i=1}^n \cos(y_i - \mu).$$

Utilizando a equação (1.5) na expressão acima temos

$$l(\mu, \lambda; y_1, \dots, y_n) = -n \{ \log 2\pi - \log I_0(\lambda) + \lambda \bar{R} \cos(\bar{\mu} - \mu) \}. \quad (2.21)$$

O estimador de máxima verossimilhança para  $\mu$  é obtido quando a função de log-verossimilhança atinge seu valor máximo. Como  $\cos x$  tem valor máximo para  $x = 0$ , então

$$\cos(\bar{\mu} - \mu) = 0 \Rightarrow \bar{\mu} - \mu = 0 \Rightarrow \hat{\mu} = \bar{\mu}$$

ou seja, o estimador de máxima verossimilhança da média direcional  $\mu$  é exatamente a média circular amostral, definida pela equação (1.1). A propósito, Bingham e Mardia (1975) mostram que a distribuição von Mises é a única distribuição circular onde a direção média amostral é o estimador de máxima verossimilhança para a direção média populacional.

Por sua vez, obtemos o estimador de máxima verossimilhança para  $\lambda$  da seguinte forma

$$\frac{\partial l}{\partial \lambda} = n \{ \bar{R} \cos(\bar{\mu} - \mu) - \frac{1}{I_0(\lambda)} I_0'(\lambda) \}.$$

Mas,  $I_0'(\lambda) = I_1(\lambda)$ , então

$$\frac{\partial l}{\partial \lambda} = n \{ \bar{R} \cos(\bar{\mu} - \hat{\mu}) - \frac{I_1(\lambda)}{I_0(\lambda)} \}.$$

Como  $A_1(\lambda) = \frac{I_1(\lambda)}{I_0(\lambda)}$  temos,

$$n \{ \bar{R} \cos(\bar{\mu} - \hat{\mu}) - A_1(\lambda) \} = 0 \Rightarrow \bar{R} \cos(\bar{\mu} - \hat{\mu}) - A_1(\lambda) = 0 \Rightarrow A_1(\lambda) = \bar{R}.$$

Portanto, o estimador de máxima verossimilhança  $\hat{\lambda}$  para  $\lambda$  é tal que

$$A_1(\hat{\lambda}) = \bar{R} \text{ ou seja, } \hat{\lambda} = A_1^{-1}(\bar{R})$$

O estimador  $\hat{\lambda}$  é viciado e uma aproximação para o vício foi obtida por Best e Fisher (1981). Eles concluem que, para amostras pequenas e valores pequenos de  $\lambda$ , a distribuição de  $\hat{\lambda}$  tem uma grande assimetria, com uma longa cauda à direita, conseqüentemente,

as inferências baseadas em valores pequenos de  $\lambda$  podem ser suspeitas, a menos que  $n$  seja suficientemente grande. Best e Fisher (1981) sugerem a seguinte aproximação para  $A_1^{-1}$ .

$$A_1^{-1}(r) \simeq \begin{cases} 2r + r^3 + (5r^5)/6, & r < 0,53 \\ -0,4 + 1,39r + 0,43/(1-r), & 0,53 \leq r \leq 0,85 \\ (r^3 - 4r^2 + 3r)^{-1}, & r > 0,85. \end{cases}$$

Para pequenas amostras,  $n \leq 15$ , Fisher(1995, p.88) sugere o seguinte estimador

$$\hat{\lambda}^* = \begin{cases} \max\{\hat{\lambda} - 2(n\hat{\lambda})^{-1}, 0\}, & \text{se } \hat{\lambda} < 2 \\ \frac{(n-1)^3 \hat{\lambda}}{n^3+n}, & \text{se } \hat{\lambda} \geq 2, \end{cases}$$

onde  $\hat{\lambda}$  é o estimador de máxima verossimilhança,  $\hat{\lambda} = A_1^{-1}(\bar{R})$ .

### Distribuições Assintóticas

A partir da teoria padrão dos *estimadores de máxima verossimilhança*, segue-se que a distribuição assintótica para grandes amostras de  $(\hat{\mu}, \hat{\lambda})$  é

$$\sqrt{n} (\hat{\mu} - \mu, \hat{\lambda} - \lambda) \sim N_2(\mathbf{0}, I^{-1}), \quad (2.22)$$

onde  $I$  é a matriz de informação de Fisher, dada por

$$I = E \left[ - \begin{pmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial \lambda} \\ \frac{\partial^2 l}{\partial \lambda \partial \mu} & \frac{\partial^2 l}{\partial \lambda^2} \end{pmatrix} \right].$$

Calculando as derivadas parciais e considerando as propriedades do valor esperado, verificamos que a matriz de informação de Fisher é representada da seguinte forma:

$$I = \begin{pmatrix} \lambda A(\lambda) & 0 \\ 0 & 1 - A^2(\lambda) - \frac{A(\lambda)}{\lambda} \end{pmatrix}. \quad (2.23)$$

Logo, para  $n$  grande, temos

$$\begin{cases} nVar(\hat{\mu}) & \approx \frac{1}{\lambda A(\lambda)} \\ nVar(\hat{\lambda}) & \approx \frac{1}{1 - A^2(\lambda) - \frac{A(\lambda)}{\lambda}} \\ nCov(\hat{\mu}, \hat{\lambda}) & \approx 0 \end{cases} \quad (2.24)$$

Portanto,  $\hat{\mu}$  e  $\hat{\lambda}$  são aproximadamente distribuídos de acordo com duas normais independentes, ou seja

$$\sqrt{n}(\hat{\mu} - \mu) \sim N\left(0, \frac{1}{\lambda A(\lambda)}\right) \text{ e}$$

$$\sqrt{n}(\hat{\lambda} - \lambda) \sim N\left(0, \frac{1}{1 - A^2(\lambda) - \frac{A(\lambda)}{\lambda}}\right).$$

## 2.7 Outras Distribuições Circulares

Apresentamos aqui as distribuições mais importantes para variáveis aleatórias que assumem valores no círculo. Além dessas, existem outras distribuições, tais como Cardióide, Lattice, entre outras. Mais detalhes podem ser encontrados em Batschelet (1981), Fisher (1995) e Mardia (1972).

# Capítulo 3

## Modelos de Regressão para uma Resposta Angular

Neste capítulo, consideraremos modelos de regressão onde a variável resposta é uma variável angular sobre um conjunto de variáveis explicativas lineares. Apresentaremos modelos em que tanto a média direcional quanto a dispersão, de uma variável com distribuição von Mises, estão relacionadas com as variáveis explicativas por meio de funções de ligação apropriadas. Finalizaremos fazendo inferências para os modelos de média, de dispersão e misto.

### 3.1 Introdução

Os modelos de regressão têm como objetivo fornecer um mecanismo capaz de fazer previsões dos valores de uma variável, denominada resposta, utilizando informações sobre outras variáveis, que chamamos explicativas. Quando a variável resposta é angular, a teoria dos modelos de regressão apropriada tem sido pouco considerada. Em geral, a variável é tratada como sendo linear, embora problemas desse tipo não são raros e estão presentes nas mais diversas áreas. Por exemplo, a partir de informações sobre velocidades de ventos, fazer previsões com respeito a direções, ou ainda, a partir de dados de distâncias percorridas por determinados animais, prever direções a serem seguidas, entre outras aplicações. Seguindo esta abordagem, podemos destacar alguns estudos como pioneiros.



- Gould (1969), Laycock (1975) e Mardia (1972) consideraram várias propostas de modelos em que a média da variável resposta  $Y$  como função de  $x$  é uma curva em volta de infinitos espirais na superfície de um cilindro infinito.
- Jonhson e Wehrly (1978) propuseram uma classe de modelos onde a variável resposta preenche apenas uma única espiral ao longo de  $x$ .

Devido a relevância destes modelos discutiremos os resultados dos mesmos e apresentaremos algumas extensões dos modelos de Jonhson e Wehrly.

## 3.2 Alguns Modelos para Regressão Angular

A modelagem de uma variável aleatória circular  $Y$  em termos de variáveis explicativas pode ser feita sob as três abordagens seguintes:

- Modelo de Médias: Modelam-se a média direcional de  $Y$  em termos do vetor das covariáveis  $\mathbf{X} = (X_1, \dots, X_p)^T$ .
- Modelo de Dispersão: Modelam-se a dispersão de  $Y$  em termos do vetor das covariáveis  $\mathbf{X} = (X_1, \dots, X_p)^T$ .
- Modelo Misto: Modelam-se ambas, a média direcional e a dispersão de  $Y$ , em termos do vetor das covariáveis  $\mathbf{X} = (X_1, \dots, X_p)^T$ .

Segundo Fisher e Lee (1992), quando a resposta é uma variável circular, os modelos de regressão apresentam algumas características peculiares, dentre elas podemos destacar a possibilidade de se modelar a dispersão circular tão bem quanto se modela a média, uma vez que conjuntos de dados encontrados na prática exibem características de aumento de variabilidade de  $Y$  para pequenos valores de  $X$ , e outra razão é o fato de existir distribuições, tais como a uniforme, em que a média direcional não está definida. Vale salientar que modelar a dispersão não é uma tarefa simples, apresenta certas dificuldades, porque para as distribuições circulares não existe uma medida natural de escala. Por isso, é conveniente trabalhar com a família de distribuição von Mises, que apresentam uma medida de dispersão intrínseca ( $\lambda$ ) e, como foi dito no capítulo anterior, esta distribuição compartilha de propriedades inferenciais semelhantes à distribuição normal nos dados lineares.

Fazendo uma analogia à teoria de regressão normal linear, Gould (1969) propõe um modelo para média direcional que apresenta a estrutura

$$\mu = \mu_0 + \sum_{j=1}^p \beta_j x_j, \quad (3.1)$$

onde  $\mu_0, \beta_1, \dots, \beta_p$  são os parâmetros a serem estimados. Gould (1969) obteve um sistema de equações de estimação referente aos estimadores de máxima verossimilhança para os parâmetros desse modelo e desenvolveu um método iterativo para determinar  $\hat{\beta}$ , que depende do ponto de iniciação do processo. Laycok (1975), estudando este modelo, verificou que o método da máxima verossimilhança era equivalente ao método dos mínimos quadrados.

Posteriormente, Jonhson e Wehrly (1978) verificaram que a função de verossimilhança para este modelo não atendia ao critério de unicidade, podendo gerar muitos picos igualmente importantes. Sendo assim, sugeriram uma abordagem para uma variável explicativa, por meio de um modelo específico para a distribuição conjunta de  $Y$  e  $X$ , com uma função de distribuição marginal  $F(x)$  completamente especificada. A distribuição condicional de  $Y$  dado  $X = x$  é  $\mathcal{VM}(\mu + 2\pi F(x), \lambda)$ , um modelo que permite estimação direta de  $\mu$  e  $\lambda$ , o qual chamamos de modelo de médias de Jonhson e Wehrly. Para o modelo de dispersão, com apenas uma variável explicativa, eles sugeriram modelar a distribuição condicional de  $Y$  dado  $X = x$  como  $\mathcal{VM}(\mu, \lambda x)$ , que também permite a estimação direta dos parâmetros.

### 3.3 Os Modelos de Regressão von Mises

Sejam  $Y_1, \dots, Y_n$  observações angulares, que seguem distribuição von Mises com médias direcionais  $\mu_1, \dots, \mu_n$  e parâmetros de concentração  $\lambda_1, \dots, \lambda_n$ , respectivamente. Segundo Fisher e Lee (1992), esses parâmetros direcionais podem ser modelados pelos modelos de regressão von Mises. Basta considerarmos que as  $n$  observações independentes estão relacionadas com  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , em que  $\mathbf{x}_i$  é um vetor de  $p$  variáveis explicativas, com  $p \leq n$ , através de funções de ligação adequadas.

### 3.3.1 Modelo de Médias von Mises

No modelo de médias von Mises assumimos que todos os parâmetros de concentração sejam iguais  $\lambda_1 = \dots = \lambda_n = \lambda$  e que as médias direcionais  $\mu_1, \dots, \mu_n$  são relacionados com  $p$ -covariáveis explicativas pela forma

$$\mu_i = \mu + g(\boldsymbol{\beta}^T \mathbf{x}_i), \quad (3.2)$$

onde  $g$  é uma função de ligação monótona, duas vezes continuamente diferenciável, cujo objetivo é mapear a linha real para o círculo e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  é um vetor de parâmetros a serem estimados. Outra importante consideração que precisa ser feita sobre a função de ligação  $g$  é que  $g(0) = 0$ , fazendo com que o parâmetro  $\mu$  possa ser visto como uma origem, uma vez que este conceito não está bem definido quando a variável resposta assume valores no círculo. Fisher e Lee (1992) discutem possíveis candidatas para a função  $g$ , no entanto uma função de ligação bastante utilizada é  $g(t) = 2\arctan(t)$ .

#### Inferências para o Modelo de Média von Mises

Vamos considerar o modelo em que as variáveis aleatórias  $Y_1, \dots, Y_n$  são independentes e  $Y_i \sim \mathcal{VM}(\mu_i, \lambda)$ , para  $i = 1, \dots, n$ , onde  $\mu_i = \mu + g(\boldsymbol{\beta}^T \mathbf{x}_i)$  e  $g$  é conhecida. Neste caso, a log-verossimilhança é

$$l \equiv l(\mu, \boldsymbol{\beta}, \lambda; \mathbf{y}) = -n \log I_0(\lambda) + \lambda \sum_{i=1}^n \cos[y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)] \quad (3.3)$$

e o vetor paramétrico é dado por  $\boldsymbol{\theta} = (\mu, \boldsymbol{\beta}^T, \lambda)^T$ .

## Estimadores de Máxima Verossimilhança para o Modelo de Médias von Mises

Inicialmente definimos:

$$\begin{aligned} u_i &= \text{sen}(y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)) \\ \mathbf{u} &= (u_1, \dots, u_n)^T \\ \mathbf{X} &= (\mathbf{x}_1, \dots, \mathbf{x}_n)^T \\ \mathbf{G} &= \text{diag}(g'(\boldsymbol{\beta}^T \mathbf{x}_1), \dots, g'(\boldsymbol{\beta}^T \mathbf{x}_n)) \\ S &= \sum_{i=1}^n \text{sen}(y_i - g(\boldsymbol{\beta}^T \mathbf{x}_i))/n \end{aligned} \quad (3.4)$$

$$C = \sum_{i=1}^n \text{cos}(y_i - g(\boldsymbol{\beta}^T \mathbf{x}_i))/n \quad (3.5)$$

$$R = (S^2 + C^2)^{1/2}. \quad (3.6)$$

Calculando as derivadas parciais de (3.3), com relação aos parâmetros  $\mu$ ,  $\boldsymbol{\beta}$  e  $\lambda$ , fazendo as devidas substituições temos que

$$\begin{aligned} \frac{\partial l}{\partial \mu} &= \lambda \sum_{i=1}^n \text{sen}(y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)) \\ &= \lambda \sum_{i=1}^n \text{sen}[(y_i - g(\boldsymbol{\beta}^T \mathbf{x}_i)) - \mu] \\ &= \lambda \sum_{i=1}^n \{[\text{sen}(y_i - g(\boldsymbol{\beta}^T \mathbf{x}_i)) \text{cos} \mu] - [\text{sen} \mu \text{cos}(y_i - g(\boldsymbol{\beta}^T \mathbf{x}_i))]\} \\ &= \lambda(nS \text{cos} \mu - nC \text{sen} \mu) \\ &= \lambda n(S \text{cos} \mu - C \text{sen} \mu). \end{aligned}$$

Então,

$$\frac{\partial l}{\partial \mu} = 0 \Rightarrow S \text{cos} \hat{\mu} - C \text{sen} \hat{\mu} = 0 \Rightarrow S \text{cos} \hat{\mu} = C \text{sen} \hat{\mu}.$$

Desta forma, o estimador de máxima verossimilhança  $\hat{\mu}$  deve satisfazer

$$S^2 \text{cos}^2 \hat{\mu} = C^2 \text{sen}^2 \hat{\mu}.$$

Sabemos que  $\text{cos}^2 \theta + \text{sen}^2 \theta = 1$ , substituindo este resultado na expressão acima temos,

$$S^2 \text{cos}^2 \hat{\mu} = C^2 (1 - \text{cos}^2 \hat{\mu}) \Rightarrow S^2 \text{cos}^2 \hat{\mu} + C^2 \text{cos}^2 \hat{\mu} = C^2.$$

Assim,

$$R^2 \text{cos}^2 \hat{\mu} = C^2 \Rightarrow R |\text{cos} \hat{\mu}| = |C| \Rightarrow R \text{cos} \hat{\mu} = C.$$

De fato, temos que se

$$0 \leq \hat{\mu} < \frac{\pi}{2} \text{ ou } \frac{3\pi}{2} \leq \hat{\mu} < 2\pi \Rightarrow \cos \hat{\mu} > 0 \text{ e } C > 0 \Rightarrow |\cos \hat{\mu}| = \cos \hat{\mu} \text{ e } |C| = C,$$

e se

$$\frac{\pi}{2} \leq \hat{\mu} < \pi \text{ ou } \pi \leq \hat{\mu} < \frac{3\pi}{2} \Rightarrow \cos \hat{\mu} < 0 \text{ e } C < 0 \Rightarrow |\cos \hat{\mu}| = -\cos \hat{\mu} \text{ e } |C| = -C.$$

Portanto,

$$R|\cos \hat{\mu}| = |C| \Rightarrow R \cos \hat{\mu} = C.$$

De forma análoga,

$$\begin{aligned} S^2 \cos^2 \hat{\mu} = C^2 \sin^2 \hat{\mu} &\Rightarrow S^2(1 - \sin^2 \hat{\mu}) = C^2 \sin^2 \hat{\mu} \\ &\Rightarrow R^2 \sin^2 \hat{\mu} = S^2 \\ &\Rightarrow R \sin \hat{\mu} = S. \end{aligned}$$

Podemos verificar que se

$$0 \leq \hat{\mu} < \frac{\pi}{2} \text{ ou } \frac{\pi}{2} \leq \hat{\mu} < \pi \Rightarrow \sin \hat{\mu} > 0 \text{ e } S > 0 \Rightarrow |\sin \hat{\mu}| = \sin \hat{\mu} \text{ e } |S| = S,$$

e se

$$\pi \leq \hat{\mu} < \frac{3\pi}{2} \text{ ou } \frac{3\pi}{2} \leq \hat{\mu} < 2\pi \Rightarrow \sin \hat{\mu} < 0 \text{ e } S < 0 \Rightarrow |\sin \hat{\mu}| = -\sin \hat{\mu} \text{ e } |S| = -S.$$

Portanto,

$$R|\sin \hat{\mu}| = |S| \Rightarrow R \sin \hat{\mu} = S.$$

Logo, o estimador  $\hat{\mu}$  deve satisfazer o sistema de equações

$$\begin{cases} R \cos \hat{\mu} = C \\ R \sin \hat{\mu} = S. \end{cases} \quad (3.7)$$

Por outro lado,

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \sum_{i=1}^n \sin(y_i - \mu - g(\beta^T \mathbf{x}_i)) g'(\beta^T \mathbf{x}_i) \mathbf{x}_i \\ &= \sum_{i=1}^n \mathbf{x}_i g'(\beta^T \mathbf{x}_i) \sin(y_i - \mu - g(\beta^T \mathbf{x}_i)). \end{aligned}$$

Então,

$$\frac{\partial l}{\partial \beta} = \mathbf{0} \Rightarrow \sum_{i=1}^n \mathbf{x}_i g'(\beta^T \mathbf{x}_i) \sin(y_i - \mu - g(\beta^T \mathbf{x}_i)) = \mathbf{0}.$$

Logo,

$$\sum_{i=1}^n \mathbf{x}_i g'(\boldsymbol{\beta}^T \mathbf{x}_i) \mathbf{u}_i = \mathbf{0} \Rightarrow \mathbf{x}^T \mathbf{G} \mathbf{u} = \mathbf{0}.$$

Agora, com relação ao parâmetro de concentração  $\lambda$ , temos

$$\begin{aligned} \frac{\partial l}{\partial \lambda} &= -nA_1(\lambda) + \sum_{i=1}^n \cos[(y_i - g(\boldsymbol{\beta}^T \mathbf{x}_i)) - \mu] \\ &= -nA_1(\lambda) + \cos \mu \sum_{i=1}^n \cos(y_i - g(\boldsymbol{\beta}^T \mathbf{x}_i)) + \text{sen} \mu \sum_{i=1}^n \text{sen}(y_i - g(\boldsymbol{\beta}^T \mathbf{x}_i)) \\ &= -nA_1(\lambda) + nC \cos \mu + nS \text{sen} \mu \\ &= n[-A_1(\lambda) + C \cos \mu + S \text{sen} \mu]. \end{aligned}$$

Portanto,

$$\frac{\partial l}{\partial \lambda} = 0 \Rightarrow C \cos \hat{\mu} + S \text{sen} \hat{\mu} = A_1(\hat{\lambda}).$$

Utilizando a expressão (3.7) temos,

$$C^2/R + S^2/R = A_1(\hat{\lambda}) \Rightarrow A_1(\hat{\lambda}) = R.$$

Portanto, os estimadores de máxima verossimilhança de  $\mu, \boldsymbol{\beta}$  e  $\lambda$ , devem satisfazer o seguinte sistema de equações

$$\mathbf{X}^T \mathbf{G} \mathbf{u} = \mathbf{0} \tag{3.8}$$

$$R \text{sen} \hat{\mu} = S \tag{3.9}$$

$$R \cos \hat{\mu} = C \tag{3.10}$$

$$A_1(\hat{\lambda}) = R. \tag{3.11}$$

Este sistema não possui uma solução explícita e deve ser resolvido por um processo iterativo. Podemos obter uma solução através do algoritmo dos mínimos quadrados iterativamente reponderados (IRLS), Green (1984). O vetor  $\hat{\boldsymbol{\beta}}$  é atualizado pela equação

$$\mathbf{X}^T \mathbf{G}^2 \mathbf{X} (\hat{\boldsymbol{\beta}}^{(m+1)} - \hat{\boldsymbol{\beta}}^{(m)}) = \mathbf{X}^T \mathbf{G}^2 \mathbf{y}^*, \tag{3.12}$$

onde  $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^T$  e  $y_i^* = \frac{\hat{\mu}_i}{[A_1(\hat{\lambda})g'(\hat{\boldsymbol{\beta}}^T \mathbf{x}_i)]}$ .

### Processo iterativo

A obtenção das estimativas de máxima verossimilhança  $\hat{\mu}, \hat{\boldsymbol{\beta}}$  e  $\hat{\lambda}$ , é feita através do seguinte processo iterativo:

- (i) Obter uma estimativa inicial  $\widehat{\beta}_0$ ;
- (ii) Calcular  $S$ ,  $C$  e  $R$  usando as expressões (3.4), (3.5) e (3.6);
- (iii) Determinar  $\widehat{\mu}$  e  $\widehat{\lambda}$  por (3.9)-(3.11);
- (iv) Atualizar  $\widehat{\beta}$  pela equação (3.12);
- (v) Repetir os passos (ii) a (iv) até obter a convergência.

### 3.3.2 Modelo de Dispersão von Mises

Como dito anteriormente, existem situações em que o interesse consiste em modelar o parâmetro de concentração  $\lambda$  da distribuição von Mises. Para o modelo de dispersão, assumimos que as médias direcionais são todas iguais ( $\mu_1 = \dots = \mu_n = \mu$ ) e os parâmetros de concentração  $\lambda_1, \dots, \lambda_n$  são modelados por

$$\lambda_i = h(\boldsymbol{\gamma}^T \mathbf{z}_i), \quad (3.13)$$

onde  $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_q)^T$  é um vetor de parâmetros a serem estimados,  $\mathbf{z}_i = (1, z_{i1}, \dots, z_{iq})^T$  é o vetor das covariáveis e  $h : \mathbb{R} \rightarrow [0, \infty)$  duas vezes continuamente diferenciável e atua como uma função de ligação.

#### Inferências para o Modelo de Dispersão von Mises

Vamos considerar o modelo em que as variáveis aleatórias  $Y_1, \dots, Y_n$  são independentes e  $Y_i \sim \mathcal{VM}(\mu, \lambda_i)$ , para  $i = 1, \dots, n$ , onde  $\lambda_i = h(\boldsymbol{\gamma}^T \mathbf{z}_i)$  e  $h$  é conhecida. Neste caso, a log-verossimilhança é expressa como

$$l \equiv l(\mu, \boldsymbol{\gamma}; \mathbf{y}) = - \sum_{i=1}^n \log I_0(h(\boldsymbol{\gamma}^T \mathbf{z}_i)) + \sum_{i=1}^n h(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos(y_i - \mu). \quad (3.14)$$

e o vetor paramétrico é dado por  $\boldsymbol{\theta} = (\mu, \boldsymbol{\gamma}^T)^T$ .

## Estimadores de Máxima Verossimilhança para o Modelo de Dispersão von Mises

Inicialmente definimos:

$$S = \sum_{i=1}^n \lambda_i \text{sen}(y_i), \quad (3.15)$$

$$C = \sum_{i=1}^n \lambda_i \cos(y_i) \quad \text{e} \quad (3.16)$$

$$R = (S^2 + C^2)^{1/2}. \quad (3.17)$$

Calculando as derivadas parciais de (3.14), com relação aos parâmetros  $\mu, \gamma_0, \gamma_1, \dots, \gamma_q$  e fazendo as devidas substituições temos que

$$\begin{aligned} \frac{\partial l}{\partial \mu} &= \sum_{i=1}^n h(\boldsymbol{\gamma}^T \mathbf{z}_i) \text{sen}(y_i - \mu) \\ &= \sum_{i=1}^n \lambda_i \text{sen}(y_i - \mu) \\ &= \sum_{i=1}^n \lambda_i (\text{sen} y_i \cos \mu - \cos y_i \text{sen} \mu) \\ &= \cos \mu \sum_{i=1}^n \lambda_i \text{sen} y_i - \text{sen} \mu \sum_{i=1}^n \lambda_i \cos y_i \\ &= S \cos \mu - C \text{sen} \mu. \end{aligned}$$

Então,

$$\frac{\partial l}{\partial \mu} = 0 \Rightarrow S \cos \hat{\mu} = C \text{sen} \hat{\mu} \Rightarrow S^2 \cos^2 \hat{\mu} = C^2 \text{sen}^2 \hat{\mu}.$$

Sabemos que  $\cos^2 \theta + \text{sen}^2 \theta = 1$ . Substituindo este resultado na expressão acima temos,

$$S^2 \cos^2 \hat{\mu} = C^2 (1 - \cos^2 \hat{\mu}) \Rightarrow S^2 \cos^2 \hat{\mu} + C^2 \cos^2 \hat{\mu} = C^2.$$

Assim,

$$R^2 \cos^2 \hat{\mu} = C^2 \Rightarrow R |\cos \hat{\mu}| = |C| \Rightarrow R \cos \hat{\mu} = C.$$

De forma análoga,

$$S^2 \cos^2 \hat{\mu} = C^2 \text{sen}^2 \hat{\mu} \Rightarrow S^2 (1 - \text{sen}^2 \hat{\mu}) = C^2 \text{sen}^2 \hat{\mu} \Rightarrow R^2 \text{sen}^2 \hat{\mu} = S^2 \Rightarrow R \text{sen} \hat{\mu} = S.$$

Portanto, o estimador  $\hat{\mu}$  deve satisfazer

$$\begin{cases} R \cos \hat{\mu} = C \\ R \text{sen} \hat{\mu} = S. \end{cases}$$



Por outro lado, considerando que  $\frac{d}{dt}I_0(t) = I_1(t)$  e  $A_p(t) = \frac{I_p(t)}{I_0(t)}$ , conforme Apêndice D, tem-se que

$$\begin{aligned} \frac{\partial l}{\partial \gamma_0} &= -\sum_{i=1}^n \frac{I_1(\lambda_i)}{I_0(\lambda_i)} \frac{\partial \lambda_i}{\partial \gamma_0} + \sum_{i=1}^n \cos(y_i - \mu) \frac{\partial \lambda_i}{\partial \gamma_0} \\ &= -\sum_{i=1}^n A_1(\lambda_i) h'(\boldsymbol{\gamma}^T \mathbf{z}_i) \frac{\partial}{\partial \gamma_0} (\gamma_0 + \gamma_1 \mathbf{z}_{i1} + \cdots + \gamma_q \mathbf{z}_{iq}) + \\ &\quad \sum_{i=1}^n \cos(y_i - \mu) h'(\boldsymbol{\gamma}^T \mathbf{z}_i) \frac{\partial}{\partial \gamma_0} (\gamma_0 + \gamma_1 \mathbf{z}_{i1} + \cdots + \gamma_q \mathbf{z}_{iq}) \\ &= \sum_{i=1}^n h'(\boldsymbol{\gamma}^T \mathbf{z}_i) \{-A_1(\lambda_i) + \cos(y_i - \mu)\}. \end{aligned}$$

Então o estimador de  $\gamma_0$  deve satisfazer

$$\sum_{i=1}^n h'(\hat{\boldsymbol{\gamma}}^T \mathbf{z}_i) \{-A_1(\hat{\lambda}_i) + \cos(y_i - \hat{\mu})\} = 0$$

Agora, para  $j = 1, \dots, q$ ,

$$\begin{aligned} \frac{\partial l}{\partial \gamma_j} &= -\sum_{i=1}^n \frac{I_1(\lambda_i)}{I_0(\lambda_i)} h'(\boldsymbol{\gamma}^T \mathbf{z}_i) z_{ij} + \sum_{i=1}^n \cos(y_i - \mu) h'(\boldsymbol{\gamma}^T \mathbf{z}_i) z_{ij} \\ &= -\sum_{i=1}^n A_1(\lambda_i) h'(\boldsymbol{\gamma}^T \mathbf{z}_i) z_{ij} + \sum_{i=1}^n \cos(y_i - \mu) h'(\boldsymbol{\gamma}^T \mathbf{z}_i) z_{ij} \\ &= \sum_{i=1}^n h'(\boldsymbol{\gamma}^T \mathbf{z}_i) \{-A_1(\lambda_i) + \cos(y_i - \mu)\} z_{ij}. \end{aligned}$$

Desta forma, o estimador de máxima verossimilhança de  $\gamma_j$  deve satisfazer

$$\sum_{i=1}^n h'(\hat{\boldsymbol{\gamma}}^T \mathbf{z}_i) \{-A_1(\hat{\lambda}_i) + \cos(y_i - \hat{\mu})\} z_{ij} = 0 \quad \text{para } j = 1, \dots, q.$$

Portanto, o estimador de máxima verossimilhança  $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\boldsymbol{\gamma}}^T)^T$  deve satisfazer o sistema

$$\sum_{i=1}^n h'(\hat{\boldsymbol{\gamma}}^T \mathbf{z}_i) \{-A_1(\hat{\lambda}_i) + \cos(y_i - \hat{\mu})\} = 0, \quad (3.18)$$

$$\sum_{i=1}^n h'(\hat{\boldsymbol{\gamma}}^T \mathbf{z}_i) \{-A_1(\hat{\lambda}_i) + \cos(y_i - \hat{\mu})\} z_{ij} = 0 \quad \text{para } j = 1, \dots, q, \quad (3.19)$$

$$R \cos \hat{\mu} = C, \quad (3.20)$$

$$R \text{sen} \hat{\mu} = S. \quad (3.21)$$

Como no modelo de médias, este sistema não admite solução explícita, logo a mesma deve ser obtida através de algum método iterativo. A solução através do algoritmo do

mínimos quadrados reponderados (IRLS), Green (1984), usa a equação de atualização

$$\mathbf{Z}^T \mathbf{W} \mathbf{Z} (\hat{\boldsymbol{\gamma}}^{(m+1)} - \hat{\boldsymbol{\gamma}}^{(m)}) = \mathbf{Z}^T \mathbf{W} \mathbf{y}^*, \quad (3.22)$$

onde  $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^T$ ,  $\mathbf{W}$  é uma matriz diagonal com elementos  $w_i = \{h'(\hat{\boldsymbol{\gamma}}^T \mathbf{z}_i)\}^2 A_1(\hat{\lambda}_i)$  e  $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^T$ , com  $y_i^* = \frac{\cos(y_i - \hat{\mu}) - A_1(\hat{\lambda}_i)}{h'(\hat{\boldsymbol{\gamma}}^T \mathbf{z}_i) A_1(\hat{\lambda}_i)}$ .

### Processo iterativo

A obtenção das estimativas de máxima verossimilhança  $\hat{\mu}$  e  $\hat{\boldsymbol{\gamma}}$ , é feita através do seguinte processo iterativo:

- (i) Obter uma estimativa inicial  $\hat{\boldsymbol{\gamma}}_0$ ;
- (ii) Calcular  $S$ ,  $C$  e  $R$  usando (3.15) - (3.17) ;
- (iii) Determinar  $\hat{\mu}$  por (3.20)-(3.21);
- (iv) Atualizar  $\hat{\boldsymbol{\gamma}}$  pela equação (3.22);
- (v) Repetir os passos (ii) a (iv) até obter a convergência.

### 3.3.3 Modelo Misto von Mises

Segundo Fisher (1995), podemos encontrar situações, envolvendo variáveis direcionais, onde tanto a média direcional quanto o parâmetro de concentração dependem das covariáveis. Nesses casos, nos referimos a mais uma classe de modelos de regressão, denominada *modelo misto* von Mises, que combina o modelo de médias com o modelo de dispersão. Para este modelo, a função de log-verossimilhança é dada por

$$l \equiv l(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y}) = - \sum_{i=1}^n \log I_0(h(\boldsymbol{\gamma}^T \mathbf{z}_i)) + \sum_{i=1}^n h(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos(y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)), \quad (3.23)$$

onde o vetor paramétrico é  $\boldsymbol{\theta} = (\mu, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ . Essa função é maximizada pela utilização simultânea dos dois processos iterativos citados nas Secções 3.3.1 e 3.3.2, embora sejam necessárias fazer as seguintes modificações:

- (i) A matriz  $\mathbf{G}^2$  da equação de atualização para  $\boldsymbol{\beta}$ , (3.12), deve ser substituída por  $\mathbf{G}\boldsymbol{\Lambda}\mathbf{G}$ , onde  $\boldsymbol{\Lambda}$  é uma matriz diagonal com elementos  $\hat{\lambda}_i A_1(\hat{\lambda}_i)$ ;
- (ii) Na definição do vetor  $\mathbf{y}^*$  em (3.12), deve-se substituir  $A_1(\hat{\lambda})$  por  $A_1(\hat{\lambda}_i)$ ;

- (iii) Na equação (3.22), equação de atualização para  $\hat{\gamma}$ , a  $i$ -ésima coordenada de  $\mathbf{y}^*$  fica dada por

$$y_i^* = \frac{\cos(y_i - \hat{\mu} - g(\hat{\beta}^T \mathbf{z}_i)) - A_1(\hat{\lambda}_i)}{h'(\hat{\gamma}^T \mathbf{z}_i) A_1(\hat{\lambda}_i)}.$$

### Processo iterativo

Para obtermos as estimativas de máxima verossimilhança  $\hat{\mu}$ ,  $\hat{\beta}$  e  $\hat{\gamma}$ , devemos seguir o seguinte processo iterativo:

- (i) Obter estimativas iniciais  $\hat{\beta}_0$  e  $\hat{\gamma}_0$ ;
- (ii) Obter uma estimativa atualizada  $\hat{\beta}$ , por (3.12), considerando a devida modificação;
- (iii) Obter uma estimativa atualizada  $\hat{\gamma}$ , por (3.22), considerando a devida modificação;
- (iv) Determinar  $R$ ,  $S$  e  $C$ , por (3.15)-(3.17);
- (v) Determinar  $\hat{\mu}$  por (3.20)-(3.21);
- (vi) Repetir os passos (ii) a (v) até obter a convergência.

# Capítulo 4

## Resíduos para uma Resposta Angular

Neste capítulo estudaremos os resíduos para respostas angulares. Inicialmente, faremos um estudo da componente da função desvio para a distribuição von Mises e em seguida, apresentaremos duas formas padronizadas de resíduos para o modelo de regressão von Mises e investigaremos suas respectivas distribuições de probabilidade.

### 4.1 Introdução

Podemos destacar a análise de diagnóstico como uma etapa de grande relevância no ajuste de um modelo de regressão, uma vez que nela é possível verificar possíveis afastamentos das suposições feitas para o modelo, bem como detectar a presença de observações extremas, nos resultados do ajuste. A análise de diagnóstico foi iniciada com a análise de resíduos, com a finalidade de detectar a presença de pontos extremos e avaliar a adequação da distribuição proposta para a variável resposta, vide Paula (2004, p.29). Quando a variável de interesse é direcional, como enfatizado anteriormente, já podemos esperar que a metodologia utilizada para variáveis lineares não faça sentido. Logo, faz-se necessário o desenvolvimento de uma teoria apropriada para dados direcionais.

## 4.2 A componente da função desvio

Após o ajuste de um modelo de regressão a um conjunto de dados, uma avaliação da qualidade desse ajuste sempre é realizada por meio de uma estatística especial. Nos modelos lineares generalizados, uma medida que avalia a qualidade do ajuste bastante utilizada é a função desvio, definida por  $D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n d_i^2$ , onde

$$d_i \equiv d_i(y_i; \hat{\mu}_i) = \pm\sqrt{2}[l_i(y_i; \tilde{\mu}_i) - l_i(y_i; \hat{\mu}_i)]^{1/2}, \quad (4.1)$$

em que  $l_i(y_i; \dots)$  é a contribuição de  $y_i$  para a log-verossimilhança,  $\tilde{\mu}_i$  é o estimador de máxima verossimilhança baseado apenas em  $y_i$  e  $\hat{\mu}_i$  é o estimador de máxima verossimilhança baseado na amostra completa, segundo (Souza, 1999) e (McCullagh e Nelder, 1989). O sinal de  $d_i$  representa a direção de discrepância. Em particular, se  $Y_1, \dots, Y_n$  representam variáveis aleatórias independentes com  $Y_i \sim \mathcal{VM}(\mu_i, \lambda)$ , então a função de log-verossimilhança é dada por

$$l(\mu_i; \lambda, y_i) = -n \log I_0(\lambda) + \lambda \sum_{i=1}^n \cos(y_i - \mu_i),$$

de onde podemos escrever

$$l_i(y_i; \mu_i, \lambda) = -\log I_0(\lambda) + \lambda \cos(y_i - \mu_i).$$

Logo,

$$\begin{aligned} l_i(y_i; \tilde{\mu}_i, \lambda) &= -\log I_0(\lambda) + \lambda \cos(y_i - \tilde{\mu}_i) \quad \text{e} \\ l_i(y_i; \hat{\mu}_i, \lambda) &= -\log I_0(\lambda) + \lambda \cos(y_i - \hat{\mu}_i). \end{aligned}$$

Portanto, substituindo as expressões acima em (4.1), temos

$$d_i = \pm\sqrt{2}[-\log I_0(\lambda) + \lambda \cos(y_i - \tilde{\mu}_i) + \log I_0(\lambda) - \lambda \cos(y_i - \hat{\mu}_i)]^{1/2}$$

Como  $\tilde{\mu}_i$  é o estimador de máxima verossimilhança baseado apenas em  $y_i$ , temos que  $\cos(y_i - \tilde{\mu}_i) = 1$ . Logo,

$$\begin{aligned} d_i &= \pm\sqrt{2}[\lambda - \lambda \cos(y_i - \hat{\mu}_i)]^{1/2} \\ &= \pm\sqrt{2}[\lambda(1 - \cos(y_i - \hat{\mu}_i))]^{1/2}. \end{aligned}$$

Então, a componente da função desvio é

$$d_i \equiv d_i(y_i; \hat{\mu}_i, \lambda) = \pm\sqrt{2\lambda}[1 - \cos(y_i - \hat{\mu}_i)]^{1/2}. \quad (4.2)$$

Utilizando algumas propriedades das funções trigonométricas podemos escrever (4.2) de uma outra maneira mais apropriada,

$$d_i = d_i(y_i; \hat{\mu}_i, \lambda) = \pm 2\sqrt{\lambda} \text{sen} \left[ \frac{1}{2}(y_i - \hat{\mu}_i) \right]. \quad (4.3)$$

De fato, observe inicialmente que

$$\cos(y_i - \hat{\mu}_i) = \cos\left[\frac{1}{2}(y_i - \hat{\mu}_i) + \frac{1}{2}(y_i - \hat{\mu}_i)\right],$$

de onde se obtém

$$\cos(y_i - \hat{\mu}_i) = \cos^2\left[\frac{1}{2}(y_i - \hat{\mu}_i)\right] - \text{sen}^2\left[\frac{1}{2}(y_i - \hat{\mu}_i)\right].$$

Assim,

$$1 - \cos(y_i - \hat{\mu}_i) = 1 - \left\{ \cos^2\left[\frac{1}{2}(y_i - \hat{\mu}_i)\right] - \text{sen}^2\left[\frac{1}{2}(y_i - \hat{\mu}_i)\right] \right\} = 2\text{sen}^2\left[\frac{1}{2}(y_i - \hat{\mu}_i)\right].$$

Portanto, substituindo esta expressão em (4.2) temos,

$$\begin{aligned} d_i = d_i(y_i; \hat{\mu}_i, \lambda) &= \pm\sqrt{2\lambda} \left[ 2\text{sen}^2\left[\frac{1}{2}(y_i - \hat{\mu}_i)\right] \right]^{1/2} \\ &= \pm 2\sqrt{\lambda} \text{sen} \left[ \frac{1}{2}(y_i - \hat{\mu}_i) \right]. \end{aligned}$$

### 4.3 Resíduos

De uma maneira geral, o resíduo para a  $i$ -ésima observação pode ser definido como uma função  $r_i = r(y_i, \hat{\mu}_i)$ , que tem como objetivo medir a discrepância entre o valor observado e o valor ajustado para a  $i$ -ésima observação (Cox e Snell, 1968). No modelo de regressão normal linear, em que  $Y_i \sim \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2)$ , o resíduo ordinário é definido por  $r_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$ . Todavia há outras formas de definir resíduos. Nesse modelo, a estimativa do vetor de médias é  $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ , onde  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ . Logo podemos escrever  $\hat{\boldsymbol{\mu}} = \mathbf{H}\mathbf{y}$ , com  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ . A matriz  $\mathbf{H}$  é a matriz de projeção ortogonal dos vetores de  $\mathbb{R}^n$  no subespaço gerado pelas colunas da matriz modelo  $X$ , conhecida como matriz *hat*. Portanto, considerando o vetor de resíduos ordinários definido por  $\mathbf{r} = (r_1, \dots, r_n)^T$ , segue que

$$\mathbf{r} = \mathbf{y} - \hat{\boldsymbol{\mu}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}.$$

Para este modelo, segue que  $E(\mathbf{r}) = \mathbf{0}$  e  $Var(\mathbf{r}) = \sigma^2(\mathbf{I} - \mathbf{H})$ , ou seja,  $r_i$  tem distribuição normal com média zero e variância  $Var(r_i) = \sigma^2(1 - h_{ii})$ . Como os  $r_i$ 's possuem variâncias distintas, é conveniente que eles sejam expressos de uma forma padronizada para que seja possível fazer comparações entre os mesmos. Uma forma natural é obter o resíduo *studentizado*, dividindo  $r_i$  pelo respectivo desvio padrão, logo

$$t_i = \frac{r_i}{s(1 - h_{ii})^{1/2}}, \quad i = 1, \dots, n, \quad (4.4)$$

onde,  $s^2 = \sum_{i=1}^n \frac{r_i^2}{(n-p)}$  e  $p$  é o número das covariáveis. No entanto,  $t_i$  não segue uma distribuição *t-student*, já que  $r_i$  não é independente de  $s^2$ . Para contornar este problema, substituímos  $s^2$  por  $s_{(i)}^2$ , que é o erro quadrático médio correspondente ao modelo sem a  $i$ -ésima observação. Assim o novo resíduo *studentizado* é dado por

$$t_i^* = \frac{r_i}{s_{(i)}(1 - h_{ii})^{1/2}}, \quad (4.5)$$

que segue uma distribuição  $t_{n-p-1}$  central.

Para os modelos lineares generalizados, a definição de um resíduo *studentizado* pode ser feita de forma análoga à regressão normal linear. Todavia algumas propriedades não continuam valendo, por isso faz-se necessária a definição de outros resíduos cujas propriedades sejam conhecidas ou se aproximam das propriedades de  $t_i^*$ . Nos MLGs os resíduos mais utilizados são definidos a partir das componentes da função desvio, cuja versão padronizada (vide McCullagh, 1987; Davison e Gigli, 1989) é dada por

$$t_{Di} = \frac{d(y_i; \mu_i)}{\sqrt{1 - \hat{h}_{ii}}}. \quad (4.6)$$

Williams(1984) verificou, por meio de simulações, que a distribuição de  $t_{Di}$  tende a estar mais próxima da normalidade do que as distribuições de outros resíduos citados na literatura.

Podemos encontrar muitas informações sobre resíduos em Cox e Snell (1968), McCullagh e Nelder (1989), entre outros.

### 4.3.1 Resíduo $d_i^*$

McCullagh (1987) mostra que, para os modelos lineares generalizados, a distribuição de probabilidade de

$$\frac{d(y_i; \mu_i) + \rho_{3i}/6}{\sqrt{1 + (14\rho_{3i}^2 - 9\rho_{4i})/36}}$$

é aproximadamente  $\mathcal{N}(0, 1)$ , em que  $\rho_{3i}$  e  $\rho_{4i}$  são os coeficientes de assimetria e curtose de  $\frac{\partial L(\eta_i)}{\partial \eta_i}$ , respectivamente, e  $d^*(y_i; \mu_i)$  é a  $i$ -ésima componente da função desvio  $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ . Usando os resultados de Cox e Snell (1986) é possível mostrar que  $E(d^*(y_i; \mu_i)) = 0$  e  $Var(d^*(y_i; \mu_i)) = 1 - h_{ii}$ , em que os termos negligenciados são  $o(n^{-1})$ . Para este caso, a matriz *hat* é dada por  $\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$  onde  $\mathbf{W}$  é uma matriz diagonal de pesos  $w_i = E(-\partial^2 l_i / \partial \eta_i^2)$ , calculado na convergência do processo iterativo do MLG. Baseado nesta característica da matriz *hat*, Souza e Paula (2002) consideraram uma correção para a componente da função desvio, equação (4.3), obtendo o resíduo padronizado  $d_i^*$  dado por

$$d_i^* = \pm 2\sqrt{\lambda} \frac{\text{sen} \frac{1}{2}(y_i - \hat{\mu}_i)}{(1 - \hat{h}_{ii}^*)^{1/2}}, \quad (4.7)$$

onde a matriz *hat* é  $\mathbf{H}^* = \dot{\mathbf{G}} \mathbf{X} (\mathbf{X}' \dot{\mathbf{G}}^2 \mathbf{X})^{-1} \mathbf{X}' \dot{\mathbf{G}}$  com  $\dot{\mathbf{G}} = \text{diag}(g'(\boldsymbol{\beta} \mathbf{x}'_i))$  e  $\hat{h}_{ii}^*$  é o elemento da diagonal principal de  $\mathbf{H}^*$ , avaliada na estimativa de máxima verossimilhança. Utilizando procedimentos análogos a Williams (1984), para os modelos lineares generalizados, Souza e Paula (2002) investigaram a distribuição de probabilidade do resíduo (4.7), por meio de estudo de simulação, e verificaram que a distribuição deste resíduo apresenta concordância com a distribuição normal padrão.

### 4.3.2 Resíduo $r_i$

Souza e Paula (2002) fizeram outra transformação na componente da função desvio, baseando-se na abordagem empregada por Davison e Gigli (1989). Tal abordagem tem como objetivo expressar a componente do desvio como função de uma variável aleatória com distribuição conhecida, e em seguida fazer uma expansão em série de Taylor até segunda ordem. Considerando a transformação  $y - \mu = \frac{z}{\sqrt{\lambda}}$ , na função densidade de probabilidade da distribuição von Mises, dada por (2.19), e usando o fato de que  $\cos t = 1 - \frac{1}{2}t^2 + o(t^2)$ , quando  $t \rightarrow 0$ , verificamos que, como a função densidade de probabilidade da von Mises é da forma

$$f_Y(y; \mu, \lambda) = \frac{1}{2\pi I_0(\lambda)} e^{\lambda \cos(y - \mu)},$$

tem-se

$$\begin{aligned} \exp\{\lambda \cos(y - \mu)\} &= \exp\left\{\lambda \cos\left(\frac{z}{\sqrt{\lambda}}\right)\right\} \\ &= \exp\left\{\lambda - \frac{z^2}{2} + o\left(\frac{z^2}{\lambda}\right)\right\}. \end{aligned}$$



Mas, pela continuidade da função exponencial  $\exp\{o(z^2)\} = 1 + o(z^2)$ . Então a função densidade de probabilidade  $\mathcal{VM}(\mu, \lambda)$  fica da forma

$$\begin{aligned} f\left(\frac{z}{\sqrt{\lambda}}; \mu, \lambda\right) &= \frac{1}{2\pi I_0(\lambda)} \exp\left\{\lambda - \frac{z^2}{2} + 1 + o(z^2)\right\} \\ &= \frac{e^{-z^2/2}}{\sqrt{2\pi}} \frac{e^\lambda}{\sqrt{2\pi I_0(\lambda)}} \{1 + o(z^2)\} \end{aligned} \quad (4.8)$$

Assim verificamos que a função densidade de probabilidade da  $\mathcal{VM}(\mu, \lambda)$  é aproximadamente proporcional à função densidade de probabilidade da normal padrão  $\mathcal{N}(0, 1)$  para valores próximos de  $\mu$ . É válido lembrar que, na propriedade (iv) da seção (2.6), Stephens (1963) mostrou que existe relação entre a distribuição von Mises e a distribuição normal.

A abordagem de Davison e Gigli (1989) requer que seja conhecida a mediana da distribuição von Mises  $\mathcal{VM}(\mu, \lambda)$ , mas, devido à simetria desta distribuição, vamos considerar a seguinte transformação para podermos definir esta medida. Seja  $Y \sim \mathcal{VM}(\mu, \lambda)$  no intervalo  $[-\pi, \pi)$ , considere  $W_0 = (Y - \mu) \pmod{2\pi}$  e a transformação

$$W = \begin{cases} W_0 + 2\pi, & \text{se } -2\pi < W_0 < -\pi \\ W_0, & \text{se } -\pi \leq W_0 < \pi \\ W_0 - 2\pi, & \text{se } \pi \leq W_0 < 2\pi. \end{cases} \quad (4.9)$$

Então,  $W \sim \mathcal{VM}(0, \lambda)$  no intervalo  $[-\pi, \pi)$ , logo a mediana de  $W$  é  $w_m = 0$ , similar à média direcional. De acordo com a abordagem usada por Davison & Gigli (1989), considere que  $W$  tem função de distribuição acumulada  $F_W(w; \nu, \lambda)$  onde  $\nu$  representa a média direcional.

Pelo teorema da Transformação Integral, podemos escrever o desvio residual da seguinte forma:

$$d(W; \nu, \lambda) = d(F_W^{-1}(\Phi(Z)); \nu, \lambda) = T(Z; \nu, \lambda). \quad (4.10)$$

onde,  $Z \sim \mathcal{N}(0, 1)$  e  $T$  é uma função desconhecida. A equação (4.10) está demonstrada no Apêndice C. De (4.10) podemos concluir que o desvio residual  $d$  pode ser obtido pelo estudo da transformação  $T$ . Mas, como na  $\mathcal{VM}(\mu, \lambda)$  a função de distribuição acumulada não possui uma forma explícita, então não podemos determinar uma forma analítica para a função  $T$ . No entanto, podemos recorrer aos dois primeiros termos da expansão em série de Taylor de  $T(z; \nu, \lambda)$  em torno de  $z = 0$ . Como  $w_n$  é a mediana de

$W \sim \mathcal{VM}(\nu, \lambda)$  e  $f$  como sendo sua respectiva função de densidade de probabilidade, pela demonstração de Davison & Gigli temos,

$$T(z; \nu, \lambda) = T(0; \nu, \lambda) + T'(z; \nu, \lambda)z + \frac{1}{2}T''(z; \nu, \lambda)z^2 + R_2(z, 0), \quad (4.11)$$

onde,

$$T(0; \nu, \lambda) = d(w_n; \nu, \lambda) \quad (4.12)$$

$$T'(0; \nu, \lambda) = \frac{d(w_n; \nu, \lambda)}{(2\pi)^{1/2}f(w_n)} \quad (4.13)$$

$$T''(0; \nu, \lambda) = \frac{1}{2\pi f^2(w_n)} \left( d''(w_n; \nu, \lambda) - \frac{f'(w_n)d'(w_n; \nu, \lambda)}{f(w_n)} \right) \quad (4.14)$$

**Demonstração das expressões (4.12) - (4.14):**

(4.12) De (4.10) temos que,

$$T(0; \nu, \lambda) = d(F_W^{-1}(\Phi(0)); \nu, \lambda) = d(w_n; \nu, \lambda)$$

(4.13) Calculando a derivada da função  $T$  em relação à variável  $z$ , temos,

$$\frac{\partial T}{\partial z} = \frac{\partial[d(F_W^{-1}(\Phi(z)))]}{\partial z} = d'[F_W^{-1}(\Phi(z))] \frac{1}{f(w_n)} \frac{1}{(2\pi)^{1/2}} e^{-z^2/2}.$$

Quando  $z \rightarrow 0$ ,

$$T'(0; \nu, \lambda) = \frac{d'(w_n; \nu, \lambda)}{(2\pi)^{1/2}f(w_n)}.$$

(4.14) Encontramos a expressão (4.14), derivando a expressão (4.13) e analisando-a quando  $z \rightarrow 0$ .

$$\begin{aligned} T''(z; \nu, \lambda) &= \left[ \frac{d'(w_n; \nu, \lambda)e^{-z^2/2}}{(2\pi)^{1/2}f(w_n)} \right]' \\ &= \frac{[d'(w_n; \nu, \lambda)e^{-z^2/2}]'(2\pi)^{1/2}f(w_n) - d'(w_n; \nu, \lambda)e^{-z^2/2}[(2\pi)^{1/2}f(w_n)]'}{2\pi f^2(w_n)} \end{aligned}$$

Agora

$$\begin{aligned} [d'(w_n; \nu, \lambda)e^{-z^2/2}]' &= \frac{d}{d[F_W^{-1}(\Phi(z))]} d'([F_W^{-1}(\Phi(z))]) \frac{d[F_W^{-1}(\Phi(z))]}{d[\Phi(z)]} \frac{d[\Phi(z)]}{dz} e^{-z^2/2} \\ &= +d'(d[F_W^{-1}(\Phi(z))])[-ze^{-z^2/2}] \\ &= \frac{d''(w_n; \nu, \lambda)e^{-z^2/2}}{(2\pi)^{1/2}f(w_n)} - zd'(w_n; \nu, \lambda)e^{-z^2/2} \end{aligned}$$

$$\begin{aligned} [(2\pi)^{1/2}f(w_n)]' &= f'(F_W^{-1}(\Phi(z))) \frac{d[F_W^{-1}(\Phi(z))]}{d[\Phi(z)]} \frac{d[\Phi(z)]}{dz} \\ &= \frac{f'(w_n)e^{-z^2/2}}{f(w_n)}. \end{aligned}$$

Substituindo estas expressões verificamos que,

$$T''(z; \nu, \lambda) = \frac{e^{-z^2/2} \left[ d''(w_n; \nu, \lambda) - z d'(w_n; \nu, \lambda) f(w_n) (2\pi)^{1/2} - d'(w_n; \nu, \lambda) \frac{f'(w_n) e^{-z^2/2}}{f(w_n)} \right]}{2\pi f^2(w_n)}.$$

Portanto, quando  $z \rightarrow 0$  temos,

$$T''(0; \nu, \lambda) = \frac{1}{2\pi f^2(w_n)} \left( d''(w_n; \nu, \lambda) - \frac{f'(w_n) d'(w_n; \nu, \lambda)}{f(w_n)} \right).$$

Para  $W \sim \mathcal{VM}(\nu, \lambda)$ , vimos que a componente da função desvio pode ser determinado pela equação (4.3) ficando, neste caso, da forma

$$d(w; \nu, \lambda) = 2\sqrt{\lambda} \text{sen} \left[ \frac{1}{2}(w - \nu) \right]. \quad (4.15)$$

Derivando a expressão (4.15) duas vezes temos que,

$$\begin{aligned} d'(w; \nu, \lambda) &= \sqrt{\lambda} \cos \left[ \frac{1}{2}(w - \nu) \right] \\ d''(w; \nu, \lambda) &= \frac{-1}{2} \sqrt{\lambda} \text{sen} \left[ \frac{1}{2}(w - \nu) \right]. \end{aligned}$$

Como  $W \sim \mathcal{VM}(\nu, \lambda)$ , sua função de densidade de probabilidade é dada por

$$f(w) = \frac{1}{2\pi I_0(\lambda)} e^{\lambda \cos(w)}$$

e a primeira derivada por

$$f'(w) = \frac{-\lambda \text{sen}(w)}{2\pi I_0(\lambda)} e^{\lambda \cos(w)}.$$

Substituindo estas expressões em (4.11) e usando (4.8) segue-se que

$$d(w; 0, \lambda) = T(z; 0, \lambda) = (2\lambda\pi)^{1/2} I_0(\lambda) e^{-\lambda} z + o(w^2), \quad \text{quando } w \rightarrow 0 \quad (4.16)$$

Uma vez que, de (4.11), temos,

$$T(z; 0, \lambda) = T(0; 0, \lambda) + T'(0; 0, \lambda)z + \frac{1}{2}T''(0; 0, \lambda)z^2 + R_2(z, 0),$$

onde,

$$\begin{aligned} T(0; 0, \lambda) &= d(w_n; 0, \lambda) = 2\sqrt{\lambda} \text{sen} \left( \frac{1}{2}w \right) \\ T'(0; 0, \lambda) &= \frac{d(w_n; 0, \lambda)}{(2\pi)^{1/2} f(w_n)} = \frac{2\sqrt{\lambda} \text{sen} \left( \frac{1}{2}w \right)}{(2\pi)^{1/2} f(w)} \\ T''(0; 0, \lambda) &= \frac{1}{2\pi f^2(w_n)} \left( d''(w_n; 0, \lambda) - \frac{f'(w_n) d'(w_n; 0, \lambda)}{f(w_n)} \right) \\ &= \frac{1}{\frac{e^{2(\lambda \cos(w))}}{2\pi I_0^2(\lambda)}} \left( \frac{-1}{2} \text{sen}(w) - \frac{\frac{-\lambda \text{sen}(w) e^{\lambda \cos(w)}}{2\pi I_0(\lambda)} \sqrt{\lambda} \cos \frac{1}{2}(w)}{\frac{e^{\lambda \cos(w)}}{2\pi I_0(\lambda)}} \right). \end{aligned}$$

Portanto, quando  $w \rightarrow 0$  tem-se

$$d(w; 0, \lambda) = T(z; 0, \lambda) = (2\lambda\pi)^{1/2} I_0(\lambda) e^{-\lambda z} + o(w^2).$$

Com este resultado podemos aproximar a componente da função desvio de  $\mathcal{VM}(0, \lambda)$  por uma função linear da distribuição normal padrão  $\mathcal{N}(0, 1)$ . Mas, temos também que, para  $\mu \neq 0$ , segue de (4.9), que

$$d(w_i; \nu_i, \lambda) = \sqrt{2\lambda}(1 - \cos(y_i - \mu_i))^{1/2} = d(y_i; \nu_i, \lambda), \quad (4.17)$$

ou seja, a componente da função desvio da  $\mathcal{VM}(\mu, \lambda)$  segue a mesma distribuição do componente da função desvio para  $\mathcal{VM}(0, \lambda)$ . Pelos termos acima, Souza e Paula (2002) apresentaram o seguinte resíduo padronizado

$$r_i = \pm \left(\frac{2}{\pi}\right)^{1/2} \frac{\text{sen}\left(\frac{1}{2}(y_i - \hat{\mu}_i)\right)}{I_0(\hat{\lambda})e^{-\hat{\lambda}}}, \quad (4.18)$$

que foi tratado como o resíduo  $d_i^*$  e comprovado que sua distribuição de probabilidade é equivalente à distribuição normal padrão  $\mathcal{N}(0, 1)$ .

# Capítulo 5

## Aplicação

Neste capítulo apresentaremos a aplicação que realizamos, considerando o conjunto de dados descrito em Fisher (1995, p.251), representando 19 observações de direção de ventos e concentração de ozônio.

### 5.1 Introdução

Desde os anos 90, o mundo tem se voltado para as questões ambientais, dentre elas, a poluição, os problemas na camada de ozônio, o efeito estufa entre outros. Recentemente, o aumento da temperatura global da Terra, tem preocupado, não apenas os ambientalistas, como também a comunidade científica em todo o mundo. O efeito estufa, bem como o buraco na camada de ozônio, tem aumentado a temperatura no planeta e este aumento tem causado uma série de fenômenos catastróficos marcados pelas mudanças de tempo em vários lugares, intensificação de ventos, chuvas e tempestades. Por isso, nosso objetivo, neste trabalho, é investigar a associação entre a direção dos ventos e a concentração de ozônio, de forma a verificar se as direções dos ventos são influenciadas pela concentração de ozônio da atmosfera. É bem verdade que vale a recíproca, a concentração de ozônio é influenciada pela direção dos ventos, tal abordagem pode ser vista em Fisher (1995).

## 5.2 Descrição e Análise dos Dados

Nesta aplicação, consideramos o conjunto de dados descrito no Apêndice B.18 de Fisher (1995, p. 251), representando 19 observações de direção de ventos, expressa em graus e concentração de ozônio, expressa em  $g/m^3$ . Essas observações foram coletadas sempre às 6:00 h da manhã, num intervalo de 4 dias, no período de 18 de abril a 29 de junho de 1975, por uma estação de tempo em Milwaukee. Os valores desses dados podem ser vistos pela Tabela 5.1 a seguir.

Tabela 5.1: Medidas de Direção de Ventos e Concentração de Ozônio

Direção	Concentração	Direção	Concentração
327	28.0	8	112.0
91	85.2	204	20.0
88	80.5	86	72.5
305	4.7	333	16.0
344	45.9	18	45.9
270	12.7	57	32.6
67	72.5	6	56.6
21	56.6	11	112.0
281	31.5	27	91.8
84	55.2		

Fonte: Fisher (1995)

Como estamos considerando a variável resposta como sendo a direção dos ventos (circular) e a concentração de ozônio como variável explicativa (linear), então a associação entre as variáveis é do tipo circular-linear com o modelo de regressão von Mises.

Para termos melhor visão das componentes sistemáticas nas observações circulares, na Figura 5.1 apresentamos um gráfico sugerido por Fisher (1995), onde o eixo das abscissas corresponde à concentração de ozônio e no eixo das ordenadas consideramos  $y_i$  como sendo a  $i$ -ésima observação de direção de ventos e  $y_i + 2\pi$ .

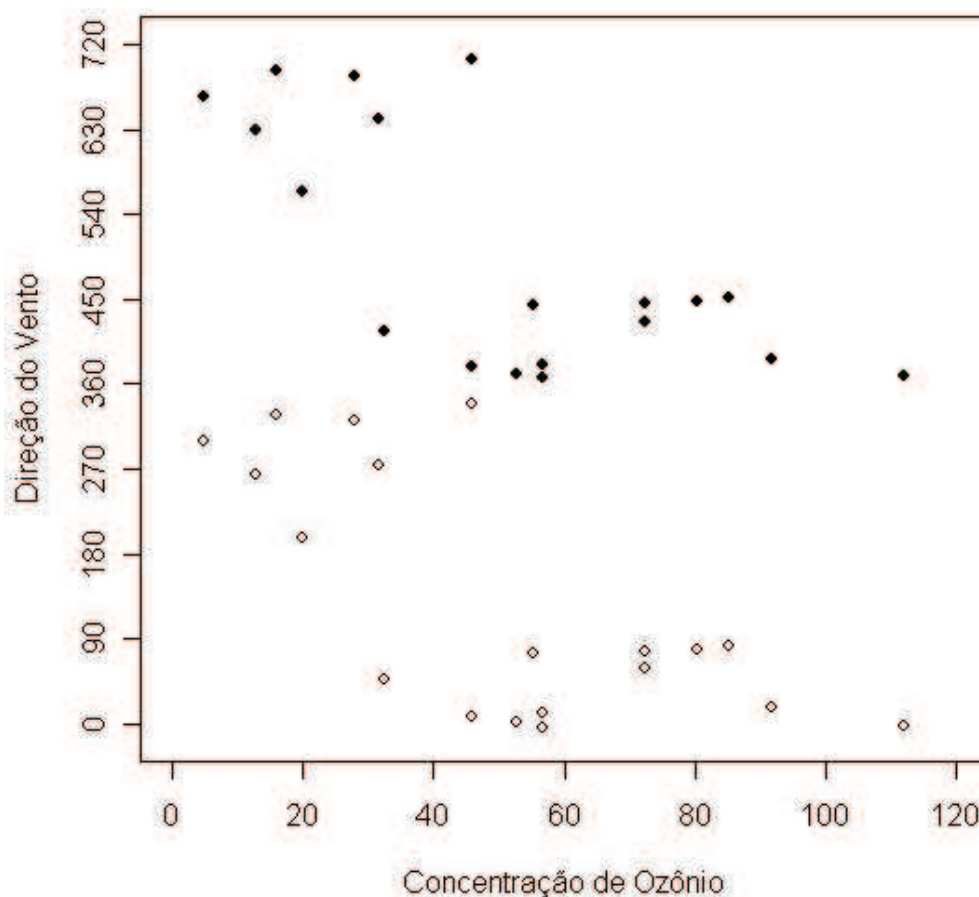


Figura 5.1: Gráfico da Concentração de Ozônio X Direção dos Ventos

Pelo gráfico, percebemos a possível existência de uma tendência na direção média. Desta forma vamos ajustar o seguinte modelo de médias:

- (i)  $y_i$  segue uma distribuição von Mises  $\mathcal{VM}(\mu_i, \lambda)$  para  $i = 1, \dots, 19$ ;
- (ii)  $\mu_i = \mu + 2 \arctan(\beta x_i)$ , onde  $x_i$  representa a concentração de ozônio de  $i$ -ésima observação.

Aplicando o procedimento iterativo descrito em (3.12), através do aplicativo R (disponível em <http://www.r-project.org>), obtivemos as seguintes estimativas de máxima verossimilhança, com respectivos desvios padrão assintóticos, que estão apresentados na Tabela 5.2, indicando que há indícios de que ajuste é aceitável.

Tabela 5.2: Estimativas dos parâmetros do modelo de médias von Mises

Parâmetro	Estimativa	D. Padrão	Estim/ D. Padrão	Nível descritivo
$\mu$	-1.492	0.1903	-7.8402	—
$\beta$	0.025817	0.005707	4.5237	<0.0001
$\lambda$	2.135	0.5988	3.5655	—

### Análise dos Resíduos

Nas Figuras 5.2 e 5.3, apresentamos os gráficos normal de probabilidade para os resíduos padronizados  $r_i$  e  $d_i^*$ , respectivamente. Em cada gráfico apresentamos bandas de confiança construídas através da técnica de envelopes, cujos limites foram obtidos a partir de 19 simulações, segundo Atkison (1985). Desta forma, a probabilidade do resíduo observado  $r_i$  ou  $(d_i^*)$  exceder o limite superior do envelope é aproximadamente  $1/20$ . Ao analisarmos esses gráficos não verificamos afastamentos das suposições feitas inicialmente para o ajuste, indicando assim que o modelo de médias foi ajustado de forma satisfatória.

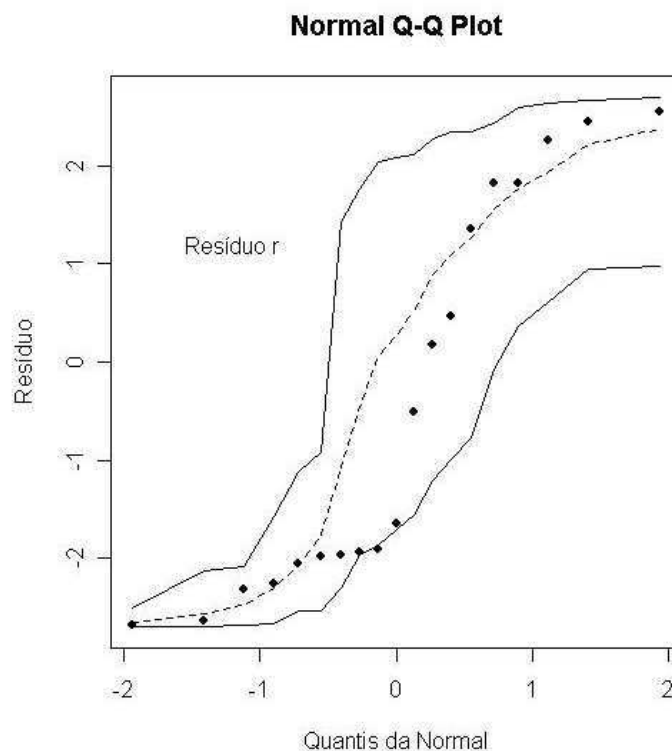


Figura 5.2: Gráfico dos envelopes para o resíduo-r



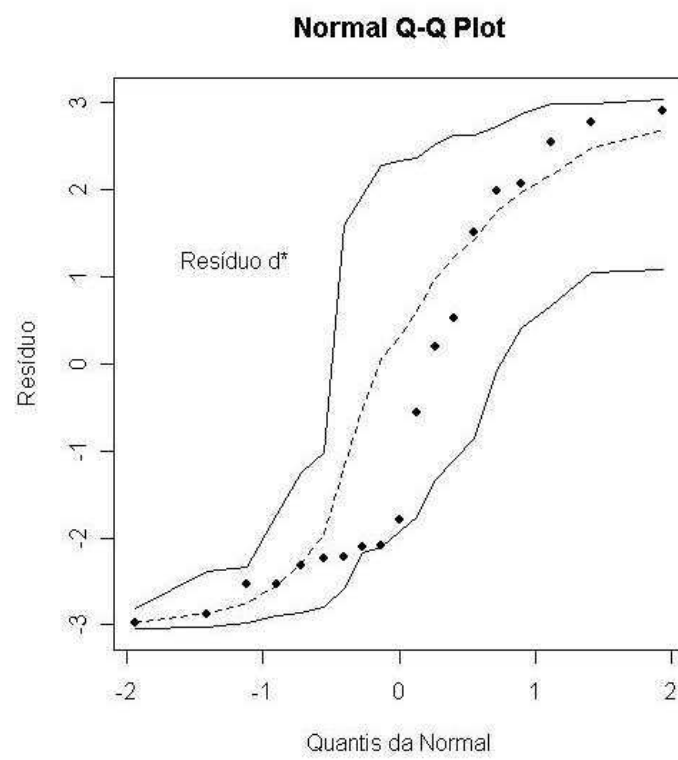


Figura 5.3: Gráfico dos envelopes para o resíduo- $d^*$

# Conclusões

Reforçando o que já foi dito neste trabalho, a análise de dados direcionais não precisa, nem deve, ser feita utilizando a teoria desenvolvida para dados lineares, pois dispomos de ferramentas eficientes e mais apropriadas para análises de dados direcionais, quer sejam análises descritivas exploratórias ou mesmo quando se deseja ajustar modelos de regressão, dentre outras. Os resíduos considerados neste trabalho, são capazes de indicar a presença de informações discrepantes, como também apresentam propriedades que nos permitem fazer inferências baseadas na aproximação pela distribuição normal padrão, tão importante e conhecida da estatística linear.

# Apêndice A

## Demonstração das Equações (1.5) e (1.6)

Neste apêndice apresentaremos a demonstração das equações (1.5) e (1.6) citadas no Capítulo 1.

### A.1 Demonstração

Aplicando a propriedade do cosseno de uma diferença na expressão (1.5), observamos que

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n \cos(y_i - \bar{\mu}) &= \frac{1}{n} \sum_{i=1}^n (\cos y_i \cos \bar{\mu} + \operatorname{sen} y_i \operatorname{sen} \bar{\mu}) \\ &= \frac{1}{n} \left[ \cos \bar{\mu} \sum_{i=1}^n \cos y_i + \operatorname{sen} \bar{\mu} \sum_{i=1}^n \operatorname{sen} y_i \right]\end{aligned}$$

De (1.4) podemos escrever

$$\frac{1}{n} \sum_{i=1}^n \cos(y_i - \bar{\mu}) = \bar{C} \cos \bar{\mu} + \bar{S} \operatorname{sen} \bar{\mu}$$

Substituindo (1.2) e (1.3) na expressão acima temos

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n \cos(y_i - \bar{\mu}) &= \bar{R} \cos^2 \bar{\mu} + \bar{R} \operatorname{sen}^2 \bar{\mu} \\ &= \bar{R} (\cos^2 \bar{\mu} + \operatorname{sen}^2 \bar{\mu}) \\ &= \bar{R}.\end{aligned}$$

Portanto,

$$\frac{1}{n} \sum_{i=1}^n \cos(y_i - \bar{\mu}) = \bar{R}.$$

Por outro lado, de maneira análoga temos

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \operatorname{sen}(y_i - \bar{\mu}) &= \frac{1}{n} \sum_{i=1}^n (\operatorname{sen} y_i \cos \bar{\mu} - \operatorname{sen} \bar{\mu} \cos y_i) \\ &= \frac{1}{n} \left[ \cos \bar{\mu} \sum_{i=1}^n \operatorname{sen} y_i - \operatorname{sen} \bar{\mu} \sum_{i=1}^n \cos y_i \right]. \end{aligned}$$

Fazendo as devidas substituições temos que,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \operatorname{sen}(y_i - \bar{\mu}) &= \bar{S} \cos \bar{\mu} - \bar{C} \operatorname{sen} \bar{\mu} \\ &= \frac{\bar{S}\bar{C}}{\bar{R}} - \frac{\bar{C}\bar{S}}{\bar{R}} \\ &= 0. \end{aligned}$$

Logo,

$$\frac{1}{n} \sum_{i=1}^n \operatorname{sen}(y_i - \bar{\mu}) = 0.$$

# Apêndice B

## Alguns Resultados sobre Dados Direcionais

Neste apêndice demonstraremos as propriedades da função característica  $\phi_Y(p)$  e apresentaremos uma demonstração para desigualdade tipo Tchebychev, que foram citadas no Capítulo 1.

### B.1 Demonstração das Propriedades da Função Característica $\phi_Y(p)$

Como visto anteriormente, a função característica  $\phi_Y(p)$  goza das seguintes propriedades:

- (i)  $\phi_Y(0) = 1$ ;
- (ii)  $\overline{\phi_Y(p)} = \phi_{-Y}(p)$ , onde  $\overline{\phi_Y(p)}$  é o conjugado complexo de  $\phi_Y(p)$ ;
- (iii)  $|\phi_Y(p)| \leq 1$ ;
- (iv) É suficiente considerarmos a função  $\phi_Y(p)$  apenas para valores inteiros e positivos de  $p$ .

**Demonstração:** A propriedade (i) é consequência da definição. Ou seja,

$$\phi_Y(0) = E(e^{i0Y}) = E(1) = 1.$$

(ii) Seja  $\overline{\phi_Y(p)} = \alpha_p - i\beta_p$ . De (1.21) temos que

$$\begin{aligned}\phi_Y(-p) &= E(e^{-ipY}) = E[\cos(-pY)] + iE[\text{sen}(-pY)] \\ &= E[\cos(pY)] - iE[\text{sen}(pY)] \\ &= \alpha_p - i\beta_p \\ &= \overline{\phi_Y(p)}.\end{aligned}$$

(iii) A propriedade (iii) é consequência das propriedades dos momentos trigonométricos e do módulo. Ou seja,

$$|\phi_Y(p)| = |E(e^{ipY})| \leq E|e^{ipY}| = 1.$$

(iv) Usando a teoria de séries de Fourier, verifica-se que é suficiente considerar a função  $\phi_Y(p)$  apenas para valores inteiros positivos de  $p$ . Sejam os números complexos  $\phi_Y(p)$  são os coeficientes de Fourier de  $F$  (Feller, 1996, p.595 ou Zygmund, 1959, p.11). Quando os  $\phi_Y(p)$  são dados por (1.21), temos que a expansão da série de Fourier da  $F$  é

$$dF(y) \sim \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \phi_Y(p) e^{-ipY}.$$

A variável aleatória circular  $Y$  tem função de densidade dada por

$$f(y) = \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \phi_Y(p) e^{-ipY}, \quad (\text{B.1})$$

uma vez que,  $\sum_{p=-\infty}^{\infty} (\alpha_p^2 + \beta_p^2)$  é convergente em  $L^2$ .

Pelo Teorema da Inversão, considerado em variáveis aleatórias contínuas na reta, podemos escrever (1.22) na forma

$$f(y) = \frac{1}{2\pi} \left\{ \sum_{p=-\infty}^{-1} (\alpha_p \cos py + \beta_p \text{sen} py) + 1 + \sum_{p=1}^{\infty} (\alpha_p \cos py + \beta_p \text{sen} py) \right\}.$$

Pelas propriedades das funções trigonométricas e dos momentos circulares, temos que

$$f(y) = \frac{1}{2\pi} \left\{ 1 + \sum_{p=1}^{\infty} (\alpha_p \cos py + \beta_p \text{sen} py) \right\}. \quad (\text{B.2})$$

Portanto, é suficiente considerarmos a função  $\phi_Y(p)$  apenas para valores inteiros e positivos de  $p$ .

## B.2 Demonstração para a Desigualdade tipo Tchebychev

Temos que  $\rho = E[\cos(Y - \mu)] = E\{\cos[(\frac{Y-\mu}{2}) + (\frac{Y-\mu}{2})]\}$ . Aplicando a propriedade do cosseno de uma soma, bem como algumas propriedades do valor esperado temos

$$\begin{aligned}\rho &= E\left[\cos^2\left(\frac{Y-\mu}{2}\right)\right] - E\left[\text{sen}^2\left(\frac{Y-\mu}{2}\right)\right] \\ &= E\left[1 - \text{sen}^2\left(\frac{Y-\mu}{2}\right)\right] - E\left[\text{sen}^2\left(\frac{Y-\mu}{2}\right)\right] \\ &= 1 - 2E\left[\text{sen}^2\left(\frac{Y-\mu}{2}\right)\right].\end{aligned}$$

Logo,

$$E\left[\text{sen}^2\left(\frac{Y-\mu}{2}\right)\right] = \frac{1-\rho}{2} = \frac{\nu}{2}.$$

Assim,

$$\nu = 2E\left[\text{sen}^2\left(\frac{Y-\mu}{2}\right)\right] = 2\int_{-\pi}^{\pi} \text{sen}^2\left(\frac{y-\mu}{2}\right)dF(y).$$

Agora, dado  $0 < \varepsilon \leq 1$ , considere  $A_\varepsilon = \{y : |\text{sen}(\frac{y-\mu}{2})| \geq \varepsilon\}$ . Assim,

$$\nu = 2\left[\int_{A_\varepsilon} \text{sen}^2\left(\frac{y-\mu}{2}\right)dF(y) + \int_{A_\varepsilon^c} \text{sen}^2\left(\frac{y-\mu}{2}\right)dF(y)\right] \geq 2\int_{A_\varepsilon} \text{sen}^2\left(\frac{y-\mu}{2}\right)dF(y).$$

Então,

$$\nu \geq 2\int_{A_\varepsilon} \varepsilon^2 dF(y) = 2\varepsilon^2 \int_{A_\varepsilon} dF(y) = 2\varepsilon^2 P(Y \in A_\varepsilon).$$

Portanto,

$$P\left(|\text{sen}\left(\frac{Y-\mu}{2}\right)| \geq \varepsilon\right) \leq \frac{\nu}{2\varepsilon^2}.$$

# Apêndice C

## Demonstração da Equação (4.10)

Neste apêndice enunciaremos o Teorema da Transformação Integral e uma demonstração para a equação (4.10) citada no Capítulo 4.

### C.1 Enunciado do Teorema da Transformação Integral

**Teorema:** Se  $X$  é uma variável aleatória cuja função de distribuição  $F$  é uma função contínua na reta, então a distribuição de  $Y = F(X)$  é  $U[0, 1]$ .

### C.2 Demonstração da Equação (4.10)

De fato, as igualdades da equação (4.10) são observadas. Seja  $W = F_W^{-1}(F_W(W))$  pelo teorema da Transformação Integral, podemos considerar que,

$$F_W(W) = U_1 \text{ e } \Phi(Z) = U_2$$

onde  $U_i \sim U[0, 1]$  e  $U_1 = U_2$  quase sempre.

Assim,

$$F_W^{-1}(U_1) = W = F_W^{-1}(\Phi(Z))$$

como  $U_1 = U_2 \Rightarrow F_W(W) = \Phi(Z)$  quase sempre. Logo,

$$W = F_W^{-1}(\Phi(Z)).$$



Portanto,

$$\forall W \text{ tal que } F_W(W; \nu, \lambda) > 0 \exists! z \in \mathbb{R} \text{ tal que } W = F_W^{-1}(\Phi(Z))$$

Logo,

$$d(W; \nu, \lambda) = d(F_W^{-1}(\Phi(Z)); \nu, \lambda) = T(Z; \nu, \lambda).$$

# Apêndice D

## Alguns resultados sobre as funções de Bessel

Neste apêndice apresentaremos alguns resultados relacionados com a função de Bessel  $I_p(t)$ , e também com a função  $A_p(t) = \frac{I_p(t)}{I_0(t)}$ . Como visto anteriormente, essas funções são de grande importância para a distribuição von Mises. Destacaremos aqui, alguns resultados que foram extraídos de Fisher (1995) e Abramowits & Stegun (1970).

### D.1 Derivadas

A função de Bessel modificada do primeiro tipo e ordem  $p$ , é definida pela série

$$I_p(\lambda) = \sum_{r=0}^{\infty} \frac{1}{(p+r)!r!} \left(\frac{1}{2}\lambda\right)^{2r+p}, \quad p = 0, 1, \dots$$

As derivadas de maior interesse no nosso caso, de funções envolvendo  $I_0(t)$  e  $I_1(t)$  são:

- (i)  $\frac{d}{dt}I_0(t) = -I_1(t)$ ;
- (ii)  $\frac{d}{dt}I_1(t) = I_0(t) - \frac{I_1(t)}{t} = \frac{1}{2}\{I_0(t) + I_2(t)\}$ ;
- (iii)  $\frac{d}{dt}\{tI_1(t)\} = tI_0(t)$ ;
- (iv)  $\frac{d}{dt}A_1(t) = 1 - \frac{A_1(t)}{t} - A_1^2(t)$ ;
- (v)  $\frac{d^2}{dt^2}A_1(t) = \frac{A_1(t)}{t^2} - \{2A_1(t) + \frac{1}{t}\} \frac{d}{dt}A_1(t)$ .

## D.2 Fórmulas de recorrência

Para  $p = 2, 3, \dots$  e  $t > 0$  dispomos das seguintes fórmulas de recorrência:

$$(i) \quad I_p(t) = I_{p-2}(t) - \frac{2(p-1)}{t} I_{p-1}(t)$$

$$(ii) \quad A_p(t) = A_{p-2}(t) - \frac{2(p-1)}{t} A_{p-1}(t).$$

# Apêndice E

## Programas Computacionais

Neste apêndice apresentaremos os programas computacionais, escritos para o aplicativo R, que foram utilizados para realizar o ajuste do modelo de média von Mises, para construção dos gráficos dos envelopes e dos demais gráficos deste trabalho.

### E.1 Comandos

```
###Este programa contém os comandos para ajustar o modelo de médias,
usando a função "lm.circular" # do pacote CIRCULAR ###

#LEITURA DOS DADOS#
ozonio.dat<-read.table("ozonio.dat",header=TRUE)
ozonio<-ozonio.dat$concentracao direcao<-ozonio.dat$direcao

direcao.real <- circular(ozonio.dat$direcao, type = c("angles"),
units = c("degrees"), modulo = c("2pi"), zero = 90 , rotation =
c("clock"))
ajuste <- lm.circular(y=direcao.real, x=ozonio, init=0,type=c("c-1"))
print(ajuste)
#-----#

### # Este programa faz o gráfico de 'ENVELOPES' # para componente
do desvio padronizado no # modelo de MEDIAS von Mises.

##LEITURA DOS DADOS ##
```

```

ozonio.dat <-read.table("ozonio.dat", header=TRUE)
attach(ozonio.dat)

namos <- length(ozonio.dat$direcao) direcao <-
matrix(c(ozonio.dat$direcao),nrow=namos,ncol=1)

##TRASFORMANDO A DIRECAO DE 'graus' PARA 'radianos'
require(circular)
require(CircStats)
direcao <- rad(direcao)# <--- angulos em [0 , 2*pi)

#TRANSFORMANDO A DIRECAO PARA O INTERVALO [-pi , +pi] ## for(k in
1:namos) {if (direcao[k] > pi) direcao[k] <- direcao[k]-2*pi}

#DEFININDO AS FUNCOES DE LIGACAO E SUAS DERIVADAS
g <-function(a) { 2*atan(a) }
dg <-function(a) { 2/(1+a^2) }
d2g<-function(a) { -(4*a)/((1+a^2)^2) }
h <-function(a) { exp(a) }
dh <-function(a) { exp(a) }
d2h<-function(a) { exp(a) }

#source("bessel1.fun") #source("bestfish.fun") source("modulo.fun")
source("arctg.fun") source("lambda.fun") source("daviggli.fun")

#AJUSTE: estimativas dos parâmetros para os dados de OZÔNIO
#Modelo de MEDIAS
mu <- -1.492 # d.p. =0,1903) # mu ~ ?? graus
beta <- matrix(c(0.025817),nrow=1,ncol=1) # d.p.=0,005707
lambda <- 2.135 # d.p.=0,5988
# #Definindo matrizes e calculando as predicoes 'y_hat' #
x <- matrix(c(ozonio.dat$concentracao),nrow=namos,ncol=1)
X <- x - mean(x)
y <- direcao
Xbeta <- X%*%beta
#A <- sapply(lambda,Bessel.FUN)
#dA <- 1 - A/lambda - A*A

#OBTENDO VALORES PREDITOS EM (0, 2*pi)

```

```

yhat <- array(data=mu + 2*atan(Xbeta) + 2*pi, dim=namos, dimnames=NULL)
for(i in 1:namos)
{
  yhat[i] <- mod.ab(yhat[i], 2*pi)
# if (yhat[i] > pi) yhat[i] <- yhat[i] - 2*pi
}
yhat<-matrix(c(yhat),nrow=namos,ncol=1)

#CALCULO DOS RESIDUOS: 'r' e 'd.estrela' # Formula Original:
d_i<-2*sign(y-yhat)*sqrt(lambda)*abs(sin((y-yhat)/2)) # comp.desvio
<- 2*sqrt(lambda)*sin((y-yhat)/2) # A expressao a seguir esta' com
a correcao # proposta de Davison & Gigli (1989) para # obter uma
aproximacao para a normal N(0,1). #
r <- matrix(NA,nrow=namos,ncol=1)
for(i in 1:namos)
{
  r[i] <- DAVGIGLI.FUN( lambda ) * sin( (y[i]-yhat[i])/2 )
}

gvet <- dg(Xbeta)
G <- diag(c(gvet),nrow=namos)
# K <- lambda * aux1 # K <- diag(c(K), nrow=namos)
G2 <- G %*% G
#H representa a matriz 'hat'
H <- G %*% X %*% solve(t(X) %*% G2 %*% X)%*% t(X) %*% G
Hdiag <- diag(H)
d.estrela <- 2*sqrt(lambda)*sin((y-yhat)/2) / sqrt(1-Hdiag)
#FIM DO CALCULO DOS RESIDUOS DO MODELO AJUSTADO

# SIMULACAO DAS 'BANDAS DE CONFIANCA' #

residuos.r <- matrix(NA,namos,19)
residuos.d <- matrix(NA,namos,19)
y <- array (data=NA,dim=namos,dimnames=NULL)
mu.fit <- yhat
lambda.fit <- lambda

#INICIO DO CALCULO DAS 'BANDAS'

```

```

for(j in 1:19)
{
#USANDO O ALGORITMO DE BEST & FISHER PARA GERAR 'y'
for (i in 1:namos)
{
y[i] <- rvm(1,mu.fit[i],lambda) #BestFisher.FUN(mu.fit[i], lambda.fit)
# if (y[i] < -pi) {y[i] <- y[i] + 2*pi}
# if (y[i] > pi) {y[i] <- y[i] - 2*pi}
}
y<-matrix(c(y),nrow=namos)

### FAZENDO O AJUSTE DA AMOSTRA SIMULADA # # #

* # * Ajuste do Modelo de MEDIAS * #
Inicio do algoritmo GERAL

### Estimativa inicial de 'mu'
Fisher (1993, eq. 2.9)
'mu'= media circular (origem) ###

Sen0 <- sum(sin(y))/namos
Cos0 <- sum(cos(y))/namos
mu <- arctg(Sen0, Cos0)
if (mu > pi) mu <- mu-2*pi #Transf. 'mu' para o intervalo (-pi , pi)

### Inicializacao do vetor beta ###

X <- x
beta.ini <- solve(t(X)%*%X) %*% t(X) %*% tan((y-mu)/2)
ajuste.amostra.simulada <-
lm.circular.cl(y , X, init=beta.ini, verbose=FALSE)#, tol=1e-10)
beta<-ajuste.amostra.simulada$coeff

#CALCULO DAS ESTIMATIVAS 'y_hat' DA 'j-ESIMA' AMOSTRA SIMULADA
#OBTENDO OS VALORES PREDITOS EM (0, 2*pi) E TRANSFORMANDO PARA (-pi,
pi)
Xbeta <- X%*%beta
yhat <- array(data=mu + 2*atan(Xbeta) + 2*pi, dim=namos,dimnames=NULL)
for(i in 1:namos)

```

```

{
  yhat[i] <- mod.ab(yhat[i], 2*pi)
# if (yhat[i] > pi) yhat[i] <- yhat[i] - 2*pi
}
yhat<-matrix(c(yhat),nrow=namos,ncol=1)

#CALCULO DOS RESIDUOS: 'r' e 'd.estrela' DA 'j-ESIMA' AMOSTRA
SIMULADA

r.aux <- matrix(NA,nrow=namos,ncol=1)
for(i in 1:namos)
{
  r.aux[i] <- DAVGIGLI.FUN( lambda ) * sin( (y[i]-yhat[i])/2 )
}

gvet <- dg(Xbeta)
G <- diag(c(gvet),nrow=namos)
G2 <- G %*% G
# K <- lambda * aux1 # K <- diag(c(K), nrow=namos)

#H representa a matriz 'hat'
H <- G %*% X %*% solve(t(X) %*% G2 %*% X)%*% t(X) %*% G
Hdiag <- diag(H)
d.aux <- 2*sqrt(lambda)*sin((y-yhat)/2) / sqrt(1-Hdiag)
#FIM DO CALCULO DOS RESIDUOS DA 'j-ESIMA' AMOSTRA SIMULADA

residuos.r[,j] <- sort( r.aux )
residuos.d[,j] <- sort( d.aux )
} #FIM DO CALCULO DOS RESIDUOS DA 'j-ESIMA' AMOSTRA SIMULADA

#FIM DOS CALCULOS PARA OBTENCAO DAS 'BANDAS' (item 7)

###      INICIO DO 'LOOP-2' (item 8)      . # .

k<-19 alfa<-0.05
linf.r<-numeric(namos) lsup.r<-numeric(namos)
linf.d<-numeric(namos)
lsup.d<-numeric(namos)

```



```

alfa1<-ceiling(k*alfa) alfa2<-ceiling(k*(1-alfa)) for(i in 1:namos)
{
  eo<-sort(residuos.r[i,])
  linf.r[i]<-eo[alfa1]
  lsup.r[i]<-eo[alfa2]
  eo<-sort(residuos.d[i,])
  linf.d[i]<-eo[alfa1]
  lsup.d[i]<-eo[alfa2]
}

xb.r<-apply(residuos.r,1,mean)
xb.d<-apply(residuos.d,1,mean)

#GRAFICO DO RESIDUO 'r' ISOLADO win.graph()
faixa<-range(r,linf.r,lsup.r) par(pty="s")
qqnorm(linf.r,axes=F,xlab="",ylab="",ylim=faixa,type="l",lty=1,lwd=1)
par(new=T)
qqnorm(lsup.r,axes=F,xlab="",ylab="",ylim=faixa,type="l",lty=1,lwd=1)
par(new=T) qqnorm(xb.r,xlab="Quantiles of Standard
Normal",ylab="Deviance Residual",ylim=faixa,type="l",lty=2,lwd=1)
text(-1.2,1.2,"Residual r") par(new=T) #oldpar
<-par(pch=15,cex=0.1,csi=0.1) # set new values, remember old oldpar
<-par(pch=16,mkh=0.1) # set new values, remember old
qqnorm(r,axes=F,xlab="",ylab="",ylim=faixa)

par(oldpar) # set parameters back to remembered values #title("(2a)") #

#GRAFICO DO RESIDUO 'd.estrela' ISOLADO win.graph()
faixa<-range(d.estrela,linf.d,lsup.d) par(pty="s")
qqnorm(linf.d,axes=F,xlab="",ylab="",ylim=faixa,type="l",lty=1,lwd=1)
par(new=T)
qqnorm(lsup.d,axes=F,xlab="",ylab="",ylim=faixa,type="l",lty=1,lwd=1)
par(new=T) qqnorm(xb.d,xlab="Quantiles of Standard
Normal",ylab="Deviance Residual",ylim=faixa,type="l",lty=2,lwd=1)
text(-1.2,1.2,"Residual d*") par(new=T) #oldpar
<-par(pch=15,cex=0.1,csi=0.1) # set new values, remember old oldpar
<-par(pch=16,mkh=0.1) # set new values, remember old
qqnorm(d.estrela,axes=F,xlab="",ylab="",ylim=faixa) par(oldpar) #
set parameters back to remembered values #title("(b)")

```

```

#qqnorm(d.estrela,xlab="Percentis da N(0,1)",ylab="Residuo
d*",ylim=faixa) #title("Grafico dos envelopes para o modelo von
Mises") #

#GRAFICO DOS RESIDUOS 'r' e 'd.estrela' CONJUNTAMENTE #
win.graph() par(mfrow=c(1,2))
#PRIMEIRO: 'r'
faixa<-range(r,linf.r,lsup.r) par(pty="s")
qqnorm(linf.r,axes=F,xlab="",ylab="",ylim=faixa,type="l",lty=1,lwd=1)
par(new=T)
qqnorm(lsup.r,axes=F,xlab="",ylab="",ylim=faixa,type="l",lty=1,lwd=1)
par(new=T) qqnorm(xb.r,xlab="Quantiles of Standard
Normal",ylab="Deviance Residual",ylim=faixa,type="l",lty=2,lwd=1)
text(-1.2,1.2,"Residual r")

par(new=T)#
oldpar<-par(pch=15,cex=0.1,csi=0.1) # set new values, remember old
oldpar<-par(pch=16,mkh=0.1) # set new values, remember old
qqnorm(r,axes=F,xlab="",ylab="",ylim=faixa)
par(oldpar) # set parameters back to remembered values #
title("(2a)")#
qqnorm(r,xlab="Percentis da N(0,1)",ylab="Residuo r",ylim=faixa)
title("Grafico dos envelopes para o modelo von Mises") #
#SEGUNDO: 'd.estrela'
faixa<-range(d.estrela,linf.d,lsup.d) par(pty="s")
qqnorm(linf.d,axes=F,xlab="",ylab="",ylim=faixa,type="l",lty=1,lwd=1)
par(new=T)
qqnorm(lsup.d,axes=F,xlab="",ylab="",ylim=faixa,type="l",lty=1,lwd=1)
par(new=T)
qqnorm(xb.d,xlab="Quantiles of Standard
Normal",ylab="Deviance Residual",ylim=faixa,type="l",lty=2,lwd=1)
text(-1.2,1.2,"Residual d*") par(new=T)
oldpar<-par(pch=15,cex=0.1,csi=0.1) # set new values, remember old
oldpar<-par(pch=16,mkh=0.1) # set new values, remember old
qqnorm(d.estrela,axes=F,xlab="",ylab="",ylim=faixa)
par(oldpar) #set parameters back to remembered values #
title("(b)")
qqnorm(d.estrela,xlab="Percentis da N(0,1)",ylab="Residuo
d*",ylim=faixa)

```

```

title("Grafico dos envelopes para o modelo von Mises")
#-----#

###      Este programa faz o gráfico da Concentração do Ozônio versus
Direção do Vento          ###

#LEITURA DOS DADOS

ozonio.dat <-read.table("ozonio.dat", header=TRUE)
attach(ozonio.dat)

##1° Gráfico##
plot(ozonio.dat$concentracao, ozonio.dat$direcao,
xlim=c(0,120), ylim=c(0,720), xlab="Concentração de Ozônio",
ylab="Direção do Vento",axes=F, pch=1) par(new=TRUE)
plot(ozonio.dat$concentracao, ozonio.dat$direcao+360, xlim=c(0,120),
ylim=c(-10,720), xlab="", ylab="",axes=F, pch=16)
axis(side=1,c(0,20,40,60,80,100,120))
axis(side=2,c(0,90,180,270,360,450,540,630,720)) box()
title(main = "Gráfico . . .")
# Cálculos auxiliares
namos <-length(ozonio.dat$direcao)
direcao <-matrix(c(ozonio.dat$direcao),nrow=namos,ncol=1)

require(circular) teta<-rad(ozonio.dat$direcao) teta<- -teta + pi/2

abcissa<-ozonio.dat$concentracao*cos(teta)
ordenada<-ozonio.dat$concentracao*sin(teta)

win.graph()
##2° Gráfico##
plot(abcissa,ordenada,xlim=c(-50,90),ylim=c(-50,110),xlab="",
ylab="", axes=F) axis(side=1,c(0)) axis(side=2,c(0)) par(new=T)
plot.circular(as.circular(teta), stack=TRUE, bins=150, shrink=0.025,
type="n", axes=F,xlim=c(-50,90),ylim=c(-50,110))

box() points(0, 0, pch="+") #title(main = "Gráfico . . .")

```

# Bibliografia

- [1] Abramowitz, M. e Stegun, I. A. (1970). *Handbook of Mathematical Functions*. New York: Dover Pub.
- [2] Batschelet, E. (1981). *Circular Statistics in Biology*. London: Academic Press.
- [3] Best, D. J. e Fisher, N. I. (1981). The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters. *Communications in Statistics - Simulation and Computations* **10**, 493-502.
- [4] Bingham, M. S. e Mardia, K. V. (1975). Maximum likelihood characterization of the von Mises distribution. *In: Statistical Distributions in Scientific Work* (C. P. Patial et al., eds), Vol. 3, pp.387-398. Reidel, Dordrecht.
- [5] Cox, D. R. e Snell, E. J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society B* **30**, 248-275.
- [6] Davison, A. C. e Gigli, A. (1989). Deviance residuals and normal scores plots. *Biometrika* **76**, 211-221.
- [7] Feller, W. (1966). *An Introduction to Probability Theory and its Applications*, Vol.2 Wiley, New York. (27, 54,57, 64).
- [8] Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge: University Press.
- [9] Fisher, N. I. e Lee, A. J. (1992). Regression models for an angular response. *Biometrics* **48**, 665-667.
- [10] Gould, A. L. (1969). A regression technique for angular variates. *Biometrics* **25**, 683-700.

- [11] Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society B* **46**, 2, 149-192.
- [12] Johnson, R. A. e Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association* **73**, 602-606.
- [13] Laycock, P. J. (1975). Optimal desing: Regression models for directions. *Biometrika* **62**, 305-311.
- [14] Mardia, K. V. (1972). *Statistics of Directional Data*. London: Academic Press.
- [15] Mardia, K. V. e Jupp, P. E. (1999). *Directional Statistics*. Wiley series in probability and statistics.
- [16] Marshal,A. W. e Olkin, I. (1961). Game theoretic proof that Chebyshev inequalities are sharp. *Pacific J. Math.*, **11**, 1421-1429. (31).
- [17] McCullagh, P. (1987). *Tensor Methods in Statistics*. New York: Chapman and Hall.
- [18] McCullagh,P. e Nelder, J. A. (1989). *Generalized Linear Models*. 2nd edition. London: Chapman and Hall.
- [19] Paula, G. A. (2004). *Modelos de Regressão com Apoio Computacional*. Instituto de Matemática e Estatística - USP.
- [20] Souza, F. A. M. (1999). *Influência Local e Análise de Resíduos em Modelos de Regressão von Mises*. Tese de Doutorado. Instituto de Matemática e Estatística - USP.
- [21] Souza, F. A. M. e Paula, G. A. (2002). Deviance Residuals for an Angular Response. *Australian Statistical Publishing Association Inc. 2002* **44**, 345-356.
- [22] Stephens, M. A. (1963). Random walk on a circle. *Biometrika* **50**, 385-390.
- [23] Williams, D. A. (1984). Residuals in generalized linear models. *In: Proceedings of the 12th. International Biometrics Conference, Tokyo*, pp. 59-68.
- [24] Zygmund, A. (1959). *Trigonometrics Series*. Cambridge University Press, Vol.1, Cambridge. (27)