

REGRESSÃO POLINOMIAL E  
ALOCÇÃO ÓTIMA

FERNANDO FERRARI

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ORIENTADOR: PROF.DR.ADOLPHO WALTER PIMAZONI CANTON

= SÃO PAULO, AGOSTO DE 1981 -

*Aos meus pais  
e irmãos*

## AGRADECIMENTOS

*O presente trabalho resultou da colaboração que recebemos ao longo de sua elaboração, de professores e colegas aos quais desejamos apresentar a nossa gratidão, em especial:*

- ao Professor Doutor Adolpho Walter Pimazoni Canton, nosso orientador, não apenas pela orientação segura, como também pela amizade e incentivo que nos proporcionou.*
- aos docentes e colegas do Instituto de Matemática e Estatística da Universidade de São Paulo, que sempre nos deram apoio durante a realização do nosso programa de Mestrado.*
- ao Professor Doutor Antonio Espada Filho, Chefe do Departamento de Análise Numérica e Estatística do Instituto de Biociências, Letras e Ciências Exatas de São José do Rio Preto, pelo inestimável apoio moral com que sempre se propôs a nos ajudar.*
- aos docentes e aos funcionários dos Departamentos de Análise Numérica e Estatística e de Álgebra, Cálculo e Geometria, pelo incentivo e colaboração na realização deste trabalho.*

## Í N D I C E

CAP. 1 - INTRODUÇÃO .....	1
CAP. 2 - REGRESSÃO POLINOMIAL .....	4
2.1 - Introdução .....	4
2.2 - O problema de mal-condicionamento .....	5
2.3 - Escolha do grau de um polinômio .....	9
2.4 - Polinômios Ortogonais .....	12
CAP. 3 - ALOCAÇÃO ÓTIMA DE PONTOS .....	24
3.1.1 - Introdução .....	24
3.1.2 - Notações .....	25
3.2 - Critérios de Otimidade .....	27
CAP. 4 - DETERMINAÇÃO DO GRAU DE UMA FUNÇÃO POLINOMIAL. ....	40
4.1.1 - Introdução .....	40
4.1.2 - Notações .....	41
4.2 - Teste Sequencial F .....	43
4.3 - Teste com duas Amostras .....	48
4.4 - Uma modificação na decisão múltipla ..	50
4.5 - Perda de precisão por alocar as obser- vações para um grau maior .....	54
CAP. 5 - EXEMPLOS NUMÉRICOS .....	61
5.1 - Introdução .....	61
BIBLIOGRAFIA .....	77

## CAPÍTULO 1

### INTRODUÇÃO

Este trabalho tem por objetivo, fazer um estudo sobre Regressão Polinomial, destacando o uso de polinômios ortogonais e a alocação ótima de pontos no intervalo de variação da variável independente.

A justificativa para o uso de polinômios ortogonais no ajuste de uma Regressão Polinomial é motivada pelas dificuldades de cálculo que normalmente surgem na determinação da inversa da matriz  $X'X$  e conseqüentemente de  $\hat{\beta} = (X'X)^{-1}X'Y$ , o estimador de mínimos quadrados dos coeficientes da equação de regressão. Além disso, quando, uma equação da forma polinomial está sendo ajustada aos dados, é usualmente conveniente incluir termos adicionais de graus sucessivamente maiores na equação, até que um ajuste satisfatório tenha sido obtido. Este processo de cálculo pode ser simplificado se escolhemos as variáveis da regressão como sendo não as potências sucessivas da variável independente  $x$  mas, polinômios de graus crescentes em  $x$  os quais são não correlacionados uns aos outros. Quando as variáveis independentes são definidas desta maneira, a

análise é simplificada, pois:

Cada coeficiente de regressão sobre cada polinômio de grau maior do que o anterior pode ser calculado independentemente dos outros.

A soma de quadrados da regressão atribuída a cada polinômio é do mesmo modo independentemente calculada e, representa a quantidade em que a soma de quadrados da regressão aumenta, pela introdução de uma equação de grau maior do que o polinômio considerado anteriormente.

No capítulo 2, apresentamos um estudo sobre os polinômios ortogonais, especialmente para os polinômios de Chebyshev, como uma maneira de reduzir o problema de mal-condicionamento.

O problema relacionado com a alocação ótima de pontos é estudado no capítulo 3, onde procuramos obter os pontos sobre o eixo  $x$  (valores da variável independente) baseados nos critérios: o que minimiza a variância generalizada de  $\hat{\beta}$ , o que minimiza a variância da estimativa do coeficiente do termo de maior grau, o que minimiza a variância do valor previsto pelo modelo tanto para os casos de interpolação como, extrapolação.

No capítulo 4 é apresentado um estudo sobre a determinação do grau de uma função polinomial, especialmente de duas técnicas sequenciais devidas a Hoel (1968) e, também estudamos a perda de precisão que incorremos quando ajustamos um polinômio de grau maior do que o adequado.

Finalmente, algumas aplicações dos capítulos anteriores são apresentadas no capítulo 5.

## CAPÍTULO 2

### REGRESSÃO POLINOMIAL

#### 2.1. Introdução

Neste capítulo estudamos o problema do ajuste de uma função de regressão polinomial a um conjunto de dados.

Inicialmente, um comentário é feito sobre o problema de mal-condicionamento que é associado com o ajuste de um polinômio. No caso, das variáveis  $x_i$  estarem distribuídas aproximadamente uniforme sobre o eixo  $x$ , resulta que a matriz planejamento associada ao modelo de regressão é a conhecida matriz de Hilbert, que dificulta sobremaneira a determinação dos estimadores de mínimos quadrados dos coeficientes  $\beta_i$ ,  $i=1,2,\dots,k$ , do modelo polinomial, isto devido ao mal-condicionamento da referida matriz. Este problema poderá ser tratado, isto é, facilitado em termos de cálculo se usarmos os polinômios de Chebyshev, que são facilmente obtidos, por meio de relações de recorrência.

Fórmulas para os estimadores de mínimos quadrados dos coeficientes de regressão, assim como, para as somas de quadrados são obtidas tanto para os casos dos valores de  $x$  se

rem igualmente e desigualmente espaçados para planejamentos ba lanceados e não balanceados.

## 2.2. O problema de mal-condicionamento

Ao ajustarmos um modelo polinomial da forma

$$Y_i = \beta_0 + \beta_1 x_i + \dots + \beta_k x_i^k + \varepsilon_i \quad (i=1, 2, \dots, n) \quad (2.2.1)$$

inúmeras dificuldades práticas surgem quando  $k$  é "grande", em bora, teoricamente seja possível ajustar um polinômio de grau  $n-1$ .

Para valores de  $k$  maior do que 6, a matriz de plane-  
jamento  $X$  associada ao modelo (2.2.1)

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^k \\ 1 & x_2 & x_2^2 & \dots & x_2^k \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & x_n & x_n^2 & \dots & x_n^k \end{bmatrix}$$

torna-se mal-condicionada; isto é, "pequenas" mudanças nos ele-  
mentos de  $X$  podem causar "grandes" mudanças em  $(X'X)^{-1}$  e con-  
seqüentemente em  $\hat{\beta} = (X'X)^{-1}X'Y$ . Dessa forma quaisquer erros  
na formação de  $X'X$  afetará a estabilidade e precisão da solu-  
ção  $\hat{\beta}$ . Por exemplo, supondo que  $x_i$  é distribuída aproximadamen-  
te uniforme sobre  $[0,1]$  e  $n$  é "grande", então o elemento da  
 $i$ -ésima linha e  $s$ -ésima coluna da matriz  $X'X$  tem o seguinte  
valor aproximado (Forsythe, 1957):

$$(X'X)_{rs} = n \sum_{i=1}^n x_i^r x_i^s \frac{1}{n} \approx n \int_0^1 x^r x^s dx =$$

$$= n \int_0^1 x^{r+s} dx = \frac{n}{r+s+1} \quad (2.2.2)$$

Assim  $X'X$  é aproximadamente  $n$  vezes a matriz  $\left[ \left( \frac{1}{r+s+1} \right) \right]$ , para  $r, s = 0, 1, \dots, k$  a qual é o menor principal de ordem  $k+1$  da matriz de Hilbert

$$H = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \dots \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \dots \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

Frequentemente tem sido observado que sistemas de equações lineares envolvendo menores principais de  $H$  são muito difíceis de serem resolvidos. Para  $k=9$ , por exemplo, a inversa de  $H_{10}$ , o menor principal  $10 \times 10$  de  $H$ , tem elementos de magnitude  $3 \cdot 10^{12}$ , (Savage e Lucacs, 1954). Assim, um pequeno erro de  $10^{-10}$  em um elemento de  $X'Y$  conduzirá a um erro de aproximadamente 300 em um elemento de  $\hat{\beta} = (X'X)^{-1}X'Y$ , mostrando assim que a matriz  $H$  é muito mal-condicionada.

Como foi destacado anteriormente, uma maneira de reduzir o efeito de mal-condicionamento, é trabalhar com polinômios de Chebyshev e ajustar um modelo da forma

$$E[Y] = \frac{1}{2}\gamma_0 T_0(x) + \gamma_1 T_1(x) + \dots + \gamma_k T_k(x) \quad (2.2.3)$$

onde  $T_r(x)$  é um polinômio de Chebyshev de 1ª espécie de grau  $r$ . Esses polinômios podem ser gerados pela relação de recorrência

$$T_{r+1}(x) = 2xT_r(x) - T_{r-1}(x), \text{ para } r=1,2,\dots \quad (2.2.4)$$

começando com  $T_0(x) = 1$  e  $T_1(x) = x$ .

Assim:

$$T_2(x) = 2x^2 - 1,$$

$$T_3(x) = 4x^3 - 3x,$$

$$T_4(x) = 8x^4 - 8x^2 + 1,$$

$$T_5(x) = 16x^5 - 20x^3 + 5x, \text{ etc.}$$

As expansões de Chebyshev para monômios são:

$$1 = T_0(x)$$

$$x = T_1(x)$$

$$x^2 = \frac{1}{2}\{T_2(x) + T_0(x)\}$$

$$x^3 = \frac{1}{4}\{T_3(x) + 3T_1(x)\}$$

$$x^4 = \frac{1}{8}\{T_4(x) + 4T_2(x) + 3T_0(x)\}$$

$$x^5 = \frac{1}{16}\{T_5(x) + 5T_3(x) + 10T_1(x)\}, \text{ etc.}$$

Para fins práticos, tendo como propósito a estabilidade numérica, os  $x_i$  são "normalizados" de modo que sua variação é restrita ao intervalo  $[-1,1]$ . O valor de  $x$  normalizado é dado por:

$$x' = \frac{2x - \max\{x_i\} - \min\{x_i\}}{\{\max\{x_i\} - \min\{x_i\}\}}$$

Podemos, também escrever  $x' = \cos \theta$ , para algum  $\theta$  em  $[0, \pi]$ , e temos então que

$$T_r(x') = \cos(r\theta), \text{ para } r=0,1,2,\dots$$

Considerando os polinômios de Chebyshev, definidos pela relação de recorrência (2.2.4) as seguintes propriedades

são destacadas, quando  $r \neq s$

$$\int_{-1}^1 T_r(x) T_s(x) (1-x^2)^{-\frac{1}{2}} dx = 0$$

e

$$\sum_{i=1}^n T_r(z_i) T_s(z_i) = 0$$

onde os  $z_i$  são os zeros de  $T_n(x)$ , isto é,  $z_i = \cos \left[ (i - \frac{1}{2}) \left( \frac{\pi}{n} \right) \right]$ .

Essas duas propriedades sugerem que para valores de  $x$  razoavelmente espaçados, a matriz  $X$  associada ao modelo (2.2.3) (expresso na forma  $E(Y) = X\gamma$ ) terá colunas que são aproximadamente ortogonais, de modo que  $X'X$  terá elementos relativamente pequenos fora da diagonal, tais matrizes são geralmente bem-condicionadas. Portanto, um procedimento recomendado, para ajustar polinômios é usar polinômios de Chebyshev juntamente com um dos métodos de decomposição ortogonal. Neste caso o algoritmo modificado de Gram-Schmidt (ver p. ex. Seber cap.11) pode ser escolhido.

Juntamente com o problema de mal-condicionamento, existe também a questão de interpretar a forma de um polinômio quando polinômios de Chebyshev são usados e  $k$  é "grande"; onde interpretação física é importante pode ser apropriado reexpressar o polinômio ajustado em termos de monômios (variáveis originais). Entretanto para alguns programas com dupla precisão o ganho pode ser pouco trabalhando com polinômios de Chebyshev e transformando posteriormente para monômios, ao invés de trabalhar com monômios diretamente (v. p. ex. Beaton e

Tukey, 1974). Entretanto, se monômios são usados os  $x_i$ 's devem obviamente serem normalizados; Hayes (1970) ilustra algumas das dificuldades que podem surgir com dados que não são normalizados.

De uma maneira equivalente, os polinômios de Chebyshev de 1.<sup>a</sup> espécie poderiam ter sido definidos por:

$$T_r(x) = \cos(r \arccos x), \text{ para } r=0,1,2,\dots, x \in [-1,1].$$

Usando a identidade trigonométrica elementar  $\cos(r+1)\theta + \cos(r-1)\theta = 2 \cos \theta \cos r \theta$ , e colocando  $\theta = \arccos x$ , a relação dada em (2.2.4) assim como as propriedades citadas para esses polinômios são verificadas facilmente.

### 2.3. Escolha do Grau de um Polinômio

Um guia para escolher o grau  $k$  de um polinômio é considerar a  $SQR_{r+1}$ , a soma de quadrados residual para um polinômio de grau  $r$ , quando  $r$  aumenta ( $r+1$  parâmetros são estimados). Em geral,  $SQR_{r+1}$  decresce consistentemente no princípio e, então nivela-se para um valor razoavelmente constante em cujo estágio é usualmente claro quando parar (ex. Hayes, 1970). Em casos de dúvida podemos testar a significância do coeficiente do último monômio acrescentado ao modelo; este é o chamado procedimento de seleção "forward". Contudo, este procedimento de teste deve ser usado cautelosamente pois ele pode conduzir a paradas prematuras. Por exemplo, no ajuste de um polinômio para uma função quase simétrica, os coeficientes das potências

Ímpares serão pequenos de modo que, podemos ter uma situação em que termos de ordem ímpar não são significantes e, termos significantes de ordem par necessários para o ajuste do modelo poderiam ser desprezados.

Temos também a possibilidade que a  $SQR_{r+1}$  pode manter-se nivelado para vários valores de  $k$  antes de decrescer novamente. Por esse motivo é preferível irmos vários passos além do primeiro termo não significante.

Um outro procedimento de teste que podemos usar é o chamado procedimento de eliminação "backward". Neste caso o grau máximo que será ajustado é determinado "a priori" e então termos (monômios) de graus mais altos são eliminados um de cada vez. Este procedimento é mais eficiente do que a seleção "forward", e é sugerido que o melhor nível de significância a ser usado em cada passo é  $\alpha = 0,10$  (Kenedy e Bancroft, 1971). Contudo, permanece ainda o problema de decidir o grau máximo a ser ajustado. Infelizmente os procedimentos "forward" e "backward" não conduzem necessariamente ao mesmo modelo. Um outro procedimento que começa com grau máximo foi dado por Hoel (1968) e é descrito no capítulo (IV) deste trabalho.

A adequação de um dado modelo pode ser examinada fazendo vários gráficos residuais tais como  $Y_i$  versus  $\hat{Y}_i$  e, em particular,  $e_i$  versus  $x_i$ , onde  $e_i = Y_i - \hat{Y}_i$ . Suponhamos por exemplo, que o polinômio ajustado é de grau  $k_1$  e o verdadeiro modelo é de grau  $k_2$ . Se  $k_1 < k_2$ , teremos um viés ("underfitting"). Se  $E(\epsilon) = 0$ , então  $E(Y) = X\beta$  e a estimativa de mínimos quadrados

$\hat{\beta} = (X'X)^{-1}X'Y$  é uma estimativa não-viesada de  $\beta$ . Entretanto, se o modelo é "underfitted" de modo que o modelo verdadeiro é

$$E(Y) = X\beta + Z\gamma \quad (2.3.1)$$

onde as colunas de  $Z$  são linearmente independentes das colunas de  $X$ , então  $\hat{\beta}$  é viesado e

$$\begin{aligned} E(\hat{\beta}) &= (X'X)^{-1}X'(X\beta + Z\gamma) \\ &= \beta + (X'X)^{-1}X'Z\gamma \\ &= \beta + C\gamma \end{aligned} \quad (2.3.2)$$

Assim  $\hat{\beta}$  é agora uma estimativa viesada de  $\beta$  com viés  $C\gamma$ , onde  $C = (X'X)^{-1}X'Z$ .

De (2.3.2) temos então que

$$\begin{aligned} E(e_i) &= E(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \dots - \hat{\beta}_{k_1} x_i^{k_1}) = \\ &= \sum_{r=0}^{k_2} \beta_r x_i^r - \sum_{r=0}^{k_1} (\beta_r + \delta_r) x_i^r \end{aligned}$$

onde os  $\delta_r$  tenderão a ser "pequenos". Neste caso um gráfico de  $e_i$  versus  $x_i$  exibirá um comportamento sistemático ao invés de aleatório e terá a característica de um gráfico polinomial (alto grau). No caso  $k_1 > k_2$ , teremos um viés ("overfitting"). Supondo que o modelo é  $E(Y) = X_1\beta_1$ , onde  $X_1$  consiste das primeiras  $k$  colunas de  $X$ ; assim  $X = (X_1, X_2)$ , digamos. Então:

$$\begin{aligned} E(\hat{\beta}) &= (X'X)^{-1}X'X_1\beta_1 \\ &= (X'X)^{-1}X'X \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix} \end{aligned} \quad (2.3.3)$$

e  $\hat{\beta}_1$ , consistindo dos primeiros  $k$  elementos de  $\hat{\beta}$ , é uma estimativa não-viesada de  $\beta_1$ . Também

$$E(\hat{Y}) = E(X\hat{\beta}) = X \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix} = X_1 \beta_1 \quad (2.3.4)$$

de modo que o modelo ajustado é uma estimativa não viesada do verdadeiro modelo. Entretanto, o uso da fórmula  $\sigma^2(X'X)^{-1}$  conduz a variâncias maiores para os elementos de  $\hat{\beta}_1$  do que se o modelo adequado (menor grau) fosse adaptado. De (2.3.4) implica que  $E(e_i) = 0$  e não existirá padrão ou tendência no gráfico residual.

## 2.4. Polinômios Ortogonais

### 2.4.1. Propriedades Estatísticas Gerais

Algumas das dificuldades de cálculo relacionadas com mal-condicionamento (secção 2.2.) podem ser evitadas pelo uso de polinômios ortogonais. Por exemplo, consideremos o modelo:

$$Y_i = \gamma_0 \phi_0(x_i) + \gamma_1 \phi_1(x_i) + \dots + \gamma_k \phi_k(x_i) + \epsilon_i,$$

onde  $\phi_r(x_i)$  é um polinômio de grau  $r$  em  $x_i$  ( $r=0,1,\dots,k$ ) e os polinômios são ortogonais sobre o conjunto dos  $x$ , isto é

$$\sum_{i=1}^n \phi_r(x_i) \phi_s(x_i) = 0, \text{ para todo } r,s,r \neq s \quad (2.4.1)$$

Esta condição é um caso especial da propriedade mais geral de ortogonalidade, onde  $\phi_r(x)$  e  $\phi_s(x)$  são chamados ortogonais com relação a função peso  $w(x)$  se

$$\sum_{i=1}^n w(x_i) \phi_r(x_i) \phi_s(x_i) = 0, \text{ para todo } r,s,r \neq s$$

Se além desta condição tivermos,

$\sum_{i=1}^n w(x_i) \phi_r^2(x_i) = 1$  para todo  $r$ , os polinômios são chamados ortonormais.

Então na equação  $Y = X\gamma + \epsilon$ , a matriz

$$X = \begin{bmatrix} \phi_0(x_1) & \phi_1(x_1) & \dots & \phi_k(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \dots & \phi_k(x_2) \\ \dots & \dots & \dots & \dots \\ \phi_0(x_n) & \phi_1(x_n) & \dots & \phi_k(x_n) \end{bmatrix}$$

tem colunas mutuamente ortogonais, e

$$X'X = \begin{bmatrix} \sum_{i=1}^n \phi_0^2(x_i) & 0 & \dots & 0 \\ 0 & \sum_{i=1}^n \phi_1^2(x_i) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sum_{i=1}^n \phi_k^2(x_i) \end{bmatrix}$$

Assim, de  $\hat{\gamma} = (X'X)^{-1}X'Y$ , temos que

$$\hat{\gamma}_r = \frac{\sum_{i=1}^n \phi_r(x_i) Y_i}{\sum_{i=1}^n \phi_r^2(x_i)}, \text{ para } r = 0, 1, \dots, k \quad (2.4.2)$$

A estrutura ortogonal de  $X$  implica que a estimativa de mínimos quadrados de  $\gamma_r$  ( $r \leq k$ ) é independente do grau  $k$  do polinômio.

Sendo  $\phi_0(x_i)$  um polinômio de grau zero podemos colocar  $\phi_0(x) \equiv 1$  e obtermos

$$\hat{\gamma}_0 = \frac{\sum_{i=1}^n 1 \cdot Y_i}{\sum_{i=1}^n 1} = \frac{\sum_i Y_i}{n} = \bar{Y}$$

A soma de quadrados do resíduo, é então

$$\begin{aligned} \text{SQR}_{k+1} &= (Y - X\hat{\gamma})' (Y - X\hat{\gamma}) = Y'Y - \hat{\gamma}' X' X \hat{\gamma} \\ &= \sum_{i=1}^n Y_i^2 - \sum_{r=0}^k \left[ \sum_i \phi_r^2(x_i) \right] \hat{\gamma}_r^2 \\ &= \sum_i (Y_i - \bar{Y})^2 - \sum_{r=1}^k \left[ \sum_i \phi_r^2(x_i) \right] \hat{\gamma}_r^2 \end{aligned} \quad (2.4.3)$$

Se desejarmos testar a hipótese  $H: \gamma_k = 0$ , então a soma de quadrados para o modelo H é:

$$\begin{aligned} \text{SQR}_k &= \sum_{i=1}^n (Y_i - \bar{Y})^2 - \sum_{r=1}^{k-1} \left[ \sum_i \phi_r^2(x_i) \right] \hat{\gamma}_r^2 \\ &= \text{SQR}_{k+1} + \left[ \sum_i \phi_k^2(x_i) \right] \hat{\gamma}_k^2 \end{aligned}$$

e a estatística F apropriada, é

$$F = \frac{\text{SQR}_k - \text{SQR}_{k+1}}{\text{SQR}_{k+1} / (n-k+1)} = \frac{\sum_i \phi_k^2(x_i) \hat{\gamma}_k^2}{\text{SQR}_{k+1} / (n-k+1)}$$

Conforme mencionamos em (2.3), podemos usar seleção "forward" ou eliminação "backward" como procedimentos para de terminar o grau apropriado. Estabelecido o procedimento, pode mos prontamente determinar o grau máximo a ser ajustado. O procedimento "backward" é mais eficiente e também evita a pos sibilidade associada com o procedimento "forward" de paradas prematuras.

### 2.4.2. Obtenção de Polinômios Ortogonais

Polinômios ortogonais podem ser obtidos de várias maneiras. Seguindo Forsythe (1957), um pioneiro neste campo, Hayes (1974) sugere usar a relação de recorrência

$$\phi_{r+1}(x) = 2(x-a_{r+1})\phi_r(x) - b_r\phi_{r-1}(x) \quad (2.4.4)$$

começando com os polinômios

$$\phi_0(x) = 1 \quad \text{e} \quad \phi_1(x) = 2(x-a_1)$$

onde  $x$  é normalizado de modo que  $-1 \leq x \leq +1$ , e os  $a_{r+1}$  e  $b_r$  são escolhidos de modo a tornar as relações ortogonais (2.4.1) válidas.

Provaremos a relação (2.4.4) por indução.

Se colocarmos  $\phi_{-1}(x) = 0$ , então  $\phi_1(x) = 2(x-a_1)$  é o caso especial  $r=0$  de (2.4.4). Suponhamos por uma hipótese de indução, que escolhamos  $a_1, a_2, \dots, a_r; b_1, b_2, \dots, b_{r-1}$  de modo que  $\phi_0(x), \phi_1(x), \dots, \phi_r(x)$  sejam dois a dois ortogonais. Vejamos então como escolheremos  $a_{r+1}$  e  $b_r$  em (2.4.4) de modo que

$$\sum_{i=1}^n \phi_{r+1}(x_i) \phi_j(x_i) = 0 \quad \text{para todo } j=0,1,\dots,r \quad (2.4.4.1)$$

Colocando-se  $x = x_i$  na relação (2.4.4); multiplicando ambos os membros por  $\phi_j(x_i)$  e somando para  $i = 1, 2, \dots, n$ , temos

$$\begin{aligned} \sum_{i=1}^n \phi_{r+1}(x_i) \phi_j(x_i) &= 2 \sum_{i=1}^n (x_i - a_{r+1}) \phi_r(x_i) \phi_j(x_i) - b_r \sum_{i=1}^n \phi_{r-1}(x_i) \phi_j(x_i) = \\ &= 2 \sum_{i=1}^n x_i \phi_r(x_i) \phi_j(x_i) - 2a_{r+1} \sum_{i=1}^n \phi_r(x_i) \phi_j(x_i) - b_r \sum_{i=1}^n \phi_{r-1}(x_i) \phi_j(x_i) \end{aligned} \quad (2.4.4.2)$$

Para  $j < r-1$ , pela hipótese de indução os dois últimos termos da equação (2.4.4.2) são zeros. Entretanto, como  $x\phi_j(x)$  é um polinômio em  $x$  de grau  $j+1 < r$ , podemos expressá-lo como uma combinação linear dos polinômios  $\phi_0(x), \dots, \phi_{r-1}(x)$ . Então a soma  $\sum_{i=1}^n x_i \phi_r(x_i) \phi_j(x_i)$  será igual a zero, pois  $\phi_r$  é ortogonal a cada um dos polinômios  $\phi_0(x), \phi_1(x), \dots, \phi_{r-1}(x)$ , pela hipótese de indução. Provamos portanto, que  $\phi_{r+1}(x)$  definido por (2.4.4) é ortogonal a  $\phi_0(x), \phi_1(x), \dots, \phi_{r-1}(x)$ , para quaisquer  $a_{r+1}$  e  $b_r$  escolhidos.

Tomando  $j = r$  em (2.4.4.2), resulta que, se

$$a_{r+1} = \frac{\sum_{i=1}^n x_i \phi_r^2(x_i)}{\sum_{i=1}^n \phi_r^2(x_i)} \quad (2.4.5)$$

$\phi_{r+1}(x)$  é ortogonal a  $\phi_r(x)$ . Entretanto, se colocarmos  $j = r-1$  em (2.4.4.2), resulta que, se

$$b_r = \frac{2 \sum_{i=1}^n x_i \phi_r(x_i) \phi_{r-1}(x_i)}{\sum_{i=1}^n \phi_{r-1}^2(x_i)} \quad (2.4.6)$$

$\phi_{r+1}(x)$  é ortogonal a  $\phi_{r-1}(x)$

Assim, se  $\phi_{r+1}, b_r$  são escolhidos de acordo com (2.4.5) e (2.4.6),  $\phi_{r+1}$  é ortogonal a  $\phi_0(x), \phi_1(x), \dots, \phi_r(x)$ , o que completa a indução.

Tomando (2.4.4) para  $x = x_i$ , multiplicando membro a membro por  $\phi_{r+1}(x_i)$ , e somando para  $i=1, 2, \dots, n$  resulta a identidade

$$\sum_{i=1}^n \phi_{r+1}^2(x_i) = 2 \sum_{i=1}^n x_i \phi_r(x_i) \phi_{r+1}(x_i)$$

Poderemos então calcular  $b_r$  pela fórmula alternativa,

$$b_r = \frac{\sum_{i=1}^n \phi_r^2(x_i)}{\sum_{i=1}^n \phi_{r-1}^2(x_i)} \quad (2.4.7)$$

onde  $r=0,1,\dots,k-1$ ;  $b_0=0$  e  $a_1=\bar{x}$

(Forsythe usou a variação  $-2a+2$  e o fator unidade ao invés do fator 2 dado na equação (2.4.4)).

"Essas duas diferenças são essencialmente compensatórias, pois existe um fator constante arbitrário associado a cada polinômio ortogonal, conforme Hayes (1969)". Um programa de computador baseado no método de Forsythe é dado por Cooper (1968, 1971).

Usando as expressões (2.4.5) e (2.4.6) ou (2.4.7) na fórmula (2.4.4) poderemos gerar os polinômios ortogonais  $\phi_r(x_i)$ , recursivamente.

A mesma técnica poderá ser usada para gerar polinômios ortogonais usando-se a relação  $\sum_{i=1}^n w(x_i) \phi_r(x_i) \phi_s(x_i) = 0$ , para todo  $r \neq s$ .

### 2.4.3. Mínimos Quadrados Ponderados

Algumas vezes é necessário fazer um ajuste de mínimos quadrados ponderados, particularmente se um gráfico residual do ajuste não-ponderado sugere que a suposição básica,

$\text{Var}(\varepsilon_i) = \sigma^2$  e  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$  para  $i \neq j$  não é verificada. Neste caso podemos transformar  $Y$  (p. ex. tomar  $\log Y$ ) para obter um bom ajuste polinomial, isto é, tentar homogeneizar a variabilidade dos  $\varepsilon_i$ .

Mínimos quadrados ponderados consiste em minimizar  $\sum_{i=1}^n w(x_i) (Y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2 - \dots - \beta_k x_i^k)^2$ , onde  $w(x_i) > 0, i=1, 2, \dots, n$ . Este problema pode ser resolvido uma vez mais pelo uso de polinômios  $\tilde{\phi}_r(x_i)$ , digamos, os quais satisfazem a condição  $\sum_{i=1}^n w(x_i) \tilde{\phi}_r(x_i) \tilde{\phi}_s(x_i) = 0$  para  $r \neq s$

Neste caso, o polinômio ajustado é:

$$\tilde{f}_k(x) = \tilde{\gamma}_0 + \tilde{\gamma}_1 \tilde{\phi}_1(x) + \tilde{\gamma}_2 \tilde{\phi}_2(x) + \dots + \tilde{\gamma}_k \tilde{\phi}_k(x),$$

onde

$$\tilde{\gamma}_r = \frac{\sum_{i=1}^n w(x_i) Y_i \tilde{\phi}_r(x_i)}{\sum_{i=1}^n w(x_i) \tilde{\phi}_r^2(x_i)}, \quad r = 0, 1, \dots, k \quad (2.4.8)$$

e a matriz de variância-covariância de  $\tilde{\gamma}_r$  será uma matriz diagonal com elementos  $\sigma^2 \left\{ \sum_{i=1}^n w(x_i) \tilde{\phi}_r^2(x_i) \right\}^{-1}$ . Os polinômios podem ser gerados por uma relação análoga a (2.4.4), que é

$$\tilde{\phi}_{r+1}(x) = 2(x - \tilde{a}_{r+1}) \tilde{\phi}_r(x) - \tilde{b}_r \tilde{\phi}_{r-1}(x) \quad (2.4.9)$$

começando com  $\tilde{\phi}_0(x) = 1$  e  $\tilde{\phi}_1(x) = 2(x - \tilde{a}_1)$ , onde

$$\tilde{a}_{r+1} = \frac{\sum_{i=1}^n w(x_i) x_i \tilde{\phi}_r^2(x_i)}{\sum_{i=1}^n w(x_i) \tilde{\phi}_r^2(x_i)} \quad \text{e} \quad \tilde{b}_r = \frac{\sum_{i=1}^n w(x_i) \tilde{\phi}_r^2(x_i)}{\sum_{i=1}^n w(x_i) \tilde{\phi}_{r-1}^2(x_i)}$$

O resto da teoria, com relação ao uso de polinômios de Chebyshev, é feito como para o caso não-ponderado; as úni-

cas mudanças são nos coeficientes  $\tilde{a}_r$ ,  $\tilde{b}_r$  e  $\tilde{\gamma}_r$ .

#### 2.4.4. Polinômios Ortogonais com Valores $x$ igualmente Espaçados

Se os valores de  $x$  são igualmente espaçados, poderemos transformá-los para

$$x_i = i - \bar{i} = i - \frac{1}{2}(n+1), \text{ para } i = 1, 2, \dots, n \quad (2.4.10)$$

Então, temos o seguinte sistema de polinômios ortogonais (geralmente atribuído a Chebyshev):

$$\phi_0(x) = 1$$

$$\phi_1(x) = \lambda_1 x$$

$$\phi_2(x) = \lambda_2 \left( x^2 - \frac{1}{12}(n^2 - 1) \right)$$

$$\phi_3(x) = \lambda_3 \left( x^3 - \frac{1}{20}(3n^2 - 7)x \right)$$

$$\phi_4(x) = \lambda_4 \left( x^4 - \frac{1}{14}(3n^2 - 13)x + \frac{3}{560}(n^2 - 1)(n^2 - 9) \right), \text{ etc.}$$

onde os  $\lambda_r$  são escolhidos de modo que os valores  $\phi_r(x_i)$  sejam todos inteiros.

Esses polinômios podem ser obtidos da resolução de sistemas de equações, quando igualamos os coeficientes de  $i^p$  ( $p=0, 1, \dots, k$ ) nos modelos,

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_k x_i^k + \epsilon_i$$

$$Y_i = \gamma_0 + \gamma_1 \phi_1(i - \bar{i}) + \gamma_2 \phi_2(i - \bar{i}) + \dots + \gamma_k \phi_k(i - \bar{i}) + \epsilon_i,$$

para  $i = 1, 2, \dots, n$ .

Sendo que, na primeira equação, pelo fato dos  $x_i$  serem igualmente espaçados, colocamos  $x_i = a + ih$ , para  $a$  e  $h$  constantes e, na segunda equação  $\phi_r(i - \bar{i})$  é um polinômio de grau  $r$  em  $i - \bar{i}$  e,

será representado por

$$\phi_r(i - \bar{i}) = \phi_r(i - \frac{n+1}{2}) = \alpha_0 + \alpha_1(i - \frac{n+1}{2}) + \alpha_2(i - \frac{n+1}{2})^2 + \dots + \alpha_r(i - \frac{n+1}{2})^r$$

para  $r = 1, 2, \dots, k$ .

Onde  $\alpha_1, \alpha_2, \dots, \alpha_r$  são constantes, as quais são determinadas de modo que a matriz  $X'X$  seja diagonal e todos os elementos em  $\phi_r$  sejam números inteiros. Esses polinômios estão tabelados em Pearson e Hartley (1970) para  $n = 1$  a 52 e  $r = 1$  a 6, ( $r \leq n-1$ ); uma parte desta tabela é dada na Tabela 2.4.1. abaixo. Para ilustrar seu uso suponhamos que  $n = 3$ . Então,

$$x_i = -1, 0, 1$$

$\phi_0(x) = 1$ ,  $\phi_1(x) = \lambda_1 x = x$ ,  $\phi_2(x) = \lambda_2(x^2 - \frac{2}{3}) = 3x^2 - 2$ , e o polinômio ajustado é:

$$\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2(3x^2 - 2), \text{ onde}$$

$$\hat{\beta}_0 = \bar{Y}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^3 \phi_1(x_i) Y_i}{\sum_{i=1}^3 \phi_1^2(x_i)} = \frac{1}{2} \{ (-1)Y_1 + (0)Y_2 + (1)Y_3 \} = \frac{1}{2}(Y_3 - Y_1) \quad \text{e}$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^3 \phi_2(x_i) Y_i}{\sum_{i=1}^3 \phi_2^2(x_i)} = \frac{1}{6} \{ (1)Y_1 + (-2)Y_2 + (1)Y_3 \} = \frac{1}{6} \{ Y_1 - 2Y_2 + Y_3 \}$$

Tabela 2.4.1 - Valores dos Polinômios Ortogonais,  $\phi_r(x)$ , para valores de x igualmente espaçados da eq. (2.4.10)

	n=3		n=4			n=5			
	$\phi_1$	$\phi_2$	$\phi_1$	$\phi_2$	$\phi_3$	$\phi_1$	$\phi_2$	$\phi_3$	$\phi_4$
	-1	1	-3	1	-1	-2	2	-1	1
	0	-2	-1	-1	3	-1	-1	2	-4
	1	1	1	-1	-3	0	-2	0	6
			3	1	1	1	-1	-2	-4
						2	2	1	1
$\sum_{i=1}^3 \phi_r^2(x_i)$	2	6	20	4	20	10	14	10	70
$\lambda_r$	1	3	2	1	$\frac{10}{3}$	1	1	$\frac{5}{6}$	$\frac{35}{12}$

A soma de quadrados do resíduo dada pela equação (2.4.3), é:

$$\begin{aligned} \text{SQR}_3 &= \sum_{i=1}^3 (Y_i - \bar{Y})^2 - \hat{\beta}_1^2 \sum_{i=1}^3 \phi_1^2(x_i) - \hat{\beta}_2^2 \sum_{i=1}^3 \phi_2^2(x_i) = \\ &= \sum_{i=1}^3 (Y_i - \bar{Y})^2 - 2\hat{\beta}_1^2 - 6\hat{\beta}_2^2 \end{aligned}$$

#### 2.4.5. Polinômios Ortogonais com os Valores x igualmente espaçados e não balanceados

Neste caso, a condição de ortogonalidade é  $\sum_{i=1}^n w(x_i) \phi_r(x_i) \phi_s(x_i) = 0$  para todo r, s,  $r \neq s$ , onde  $\phi_r(x_i)$  e  $\phi_s(x_i)$  são polinômios de grau r e s respectivamente,

ortogonais em relação a função peso  $w(x)$ .

Se  $w(x_i) = n_i$ , o número de observações em  $x_i$ , então o critério de mínimos quadrados será expresso como a condição que  $\sum_{i=1}^n w(x_i) [\bar{Y}_i - \hat{Y}_i]^2$  seja mínimo, onde  $\bar{Y}_i$  é a média das observações para o nível  $x_i$ , e  $\hat{Y}_i$  é o valor previsto pelo modelo nesse ponto. Então  $\hat{\gamma}_j$  a estimativa de mínimos quadrados de  $\gamma_j$  na equação de regressão:

$$Y_i = \gamma_0 \phi_0(x_i) + \gamma_1 \phi_1(x_i) + \dots + \gamma_k \phi_k(x_i) + \epsilon_i$$

é dada por

$$\hat{\gamma}_j = \frac{\sum_{i=1}^n w(x_i) \bar{Y}_i \phi_j(x_i)}{\sum_{i=1}^n w(x_i) \phi_j^2(x_i)} \quad (2.4.11)$$

e a  $SQ_j$ , a porção da soma de quadrados entre  $x$  representada pelo termo  $\gamma_j \phi_j(x)$  é dada por

$$SQ_j = \frac{\left[ \sum_{i=1}^n w(x_i) \bar{Y}_i \phi_j(x_i) \right]^2}{\sum_{i=1}^n w(x_i) \phi_j^2(x_i)} \quad (2.4.12)$$

com 1 grau de liberdade.

#### 2.4.6. Polinômios Ortogonais com os Valores $x$ desigualmente espaçados e não balanceados

Nesse caso,  $\hat{\gamma}_j$  a estimativa de mínimos quadrados de  $\gamma_j$  na equação de regressão

$$Y_i = \gamma_0 \phi_0(x_i) + \gamma_1 \phi_1(x_i) + \dots + \gamma_k \phi_k(x_i) + \epsilon_i \quad \text{é dada}$$

por (2.4.11) sendo que a  $SQ_j$ , é

$$SQ_j = \frac{r \left\{ \sum_{i=1}^n w(x_i) \bar{Y}_i \phi_j(x_i) \right\}^2}{\sum_{i=1}^n w(x_i) \phi_j^2(x_i)} \quad (2.4.13)$$

com 1 grau de liberdade.

onde existem  $rw(x_i)$  observações de  $Y$  para o ponto  $x_i$ , sendo  $\bar{Y}_i$ , a média das mesmas.

Dois métodos para a obtenção de polinômios ortogonais para um conjunto de pontos com espaços e pesos arbitrários, foram apresentados por Emerson (1965).

O método baseado na relação de recorrência de Christoffel-Darboux (Szegő, 1959), para quaisquer três polinômios ortogonais consecutivos,

$$\phi_j(x) = (A_j x + B_j) \phi_{j-1}(x) - C_j \phi_{j-2}(x), \quad j=2,3,\dots \quad (2.4.14)$$

onde  $A_j$ ,  $B_j$  e  $C_j$  são constantes para um dado  $j$ , dadas por:

$$A_j = \left\{ \sum_{i=1}^n w(x_i) x_i^2 \phi_{j-1}^2(x_i) - \left[ \sum_{i=1}^n w(x_i) x_i \phi_{j-1}^2(x_i) \right]^2 - \left[ \sum_{i=1}^n w(x_i) x_i \phi_{j-1}(x_i) \phi_{j-2}(x_i) \right]^2 \right\}^{-\frac{1}{2}} \quad (2.4.15)$$

$$B_j = -A_j \sum_{i=1}^n w(x_i) x_i \phi_{j-1}^2(x_i)$$

$$C_j = A_j \sum_{i=1}^n w(x_i) x_i \phi_{j-1}(x_i) \phi_{j-2}(x_i)$$

parece ser preferido, tanto em relação ao tempo, como em precisão, comparado ao método de Gram-Schmidt.

## CAPÍTULO 3

### ALOCAÇÃO ÓTIMA DE PONTOS

#### 3.1.1. Introdução

Ao planejarmos um experimento para ajustar uma curva de regressão polinomial, nos encontramos com um problema de alocação de pontos. Se os pontos  $x$ , devem pertencer ao intervalo  $[-1,1]$  por exemplo, então, a questão é, para que valores de  $x$  deveremos observar os  $n$  valores de  $Y$  e que proporção dessas observações devem ser tomadas em cada ponto.

Obviamente, a resposta para esta questão depende do nosso propósito no ajuste do polinômio: estimação (com relação a todos os coeficientes do polinômio, ou apenas um subconjunto de coeficientes), interpolação (por exemplo, calibração de curvas), ou extrapolação (previsão).

Para o estudo dos objetivos destacados, uma variedade de critérios de otimidade foram sugeridos, dentre os quais os mais citados na literatura são: D-otimidade, G-otimidade e C-otimidade.

### 3.1.2. Notações

Para o estudo dos problemas relativos a alocação ótima de pontos, não haverá perda de generalidades se tomarmos a "região de interesse" como sendo o intervalo de -1 a +1, desde que os critérios que consideraremos são invariantes sob transformações lineares sobre a variável independente.

Inicialmente vamos supor que

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_k x_i^k + \epsilon_i, \text{ para } i=1, 2, \dots, n \quad (3.1.1)$$

é um polinômio de grau  $k$ , conhecido, e os  $\beta_i$ 's são desconhecidos. Denotando por  $Y$ ,  $\beta$  e  $\epsilon$  os vetores colunas,  $Y' = (Y_1, Y_2, \dots, Y_n)$ ,  $\beta' = (\beta_0, \beta_1, \dots, \beta_k)$ ,  $\epsilon' = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$  e  $X$  a matriz  $n \times (k+1)$  dada por,

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^k \\ 1 & x_2 & x_2^2 & \dots & x_2^k \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & x_n & x_n^2 & \dots & x_n^k \end{bmatrix}$$

então o modelo (3.1.1) pode ser escrito na forma matricial  $Y = X\beta + \epsilon$ , onde o vetor  $\epsilon$  tem, por hipótese, média 0 e matriz de covariância  $\sigma^2 S$  ( $S$  é uma matriz de constantes conhecidas), e os  $x_i$ 's tomam pelo menos  $k+1$  valores distintos.

Nestas condições, a estimativa de mínimos quadrados de  $\beta$ , é dada por  $\hat{\beta} = (X'S^{-1}X)^{-1}X'S^{-1}Y = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)'$  a qual tem média  $\beta$  e matriz de covariância  $(X'S^{-1}X)^{-1}$ .

Desde que o modelo de regressão clássico supõe que  $Y_1, Y_2, \dots, Y_n$  são variáveis aleatórias não correlacionadas com variância comum  $\sigma^2$ , a matriz  $S$  é a matriz identidade.

Um primeiro passo em direção ao encontro dos planejamentos ótimos que veremos neste capítulo foi feito por De La Garza (1954). Ele mostrou que correspondendo a qualquer alocação de  $n$  valores  $x$ , existe uma outra alocação, de apenas  $k+1$  pontos  $x_i$  ( $i=0, 1, \dots, k$ ) com a mesma matriz  $X'S^{-1}X$  e pertencentes ao intervalo de variação dos valores de  $x$  originais. Denotando por  $n_i$  o número de observações a ser tomadas em  $x_i$ , onde  $\sum_{i=0}^k n_i = n$ , teremos que as matrizes  $X$  e  $S$  serão dadas por

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^k \\ 1 & x_1 & x_1^2 & \dots & x_1^k \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & x_k & x_k^2 & \dots & x_k^k \end{bmatrix}, \quad S = \begin{bmatrix} \frac{1}{n_0} & 0 & \dots & 0 \\ 0 & \frac{1}{n_1} & \dots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & \frac{1}{n_k} \end{bmatrix}$$

Um planejamento discreto (ou exato) é definido como uma medida de probabilidade  $\xi$  sobre  $[-1, 1]$ , concentrando pesos  $p_0, p_1, \dots, p_k$  nos pontos distintos  $x_0, x_1, \dots, x_k$  respectivamente, tais que  $np_i = n_i$ ,  $i=0, 1, \dots, k$  são inteiros e  $\sum_{i=0}^k p_i = 1$ , onde  $n$  é o número (finito) de observações não correlacionadas  $Y_i$ ,  $i=1, 2, \dots, n$ , as quais tem variância comum  $\sigma^2$ . Nota-se que observações diferentes tomadas no mesmo ponto  $x$  são não-correlacionadas. Portanto, a medida (ou planejamento) discreto  $\xi$  especifica os diferentes pontos onde as observações serão tomadas, e número das mesmas em cada ponto.

Analogamente, correspondendo a uma medida de probabilidade discreta  $\xi$ , temos o planejamento que considera  $n\xi(x)$  medidas independentes para o valor  $x$ , para cada  $x$  em  $[-1,1]$ . Por razões de conveniência matemática, frequentemente descrevemos um planejamento por uma medida  $\xi$ , sem a restrição que  $n\xi(x)$  seja inteiro. Por exemplo, se por algum critério de otimidade, o melhor planejamento para regressão linear é:  $\xi(-1)=\xi(1)=\frac{1}{2}$ , teremos  $n\xi(1)$  não inteiro, a menos que  $n$  seja par. Para situações como esta é recomendado que uma aproximação para o planejamento estabelecido seja usado (no exemplo acima com  $n$  ímpar, tomaríamos  $(n-1)/2$  medidas para  $x=-1$  e  $(n+1)/2$  para  $x=1$ ). Para a maioria dos critérios de otimidade tais aproximações produzirão planejamentos "aproximadamente melhores" se  $n$  é "grande".

### 3.2. Critérios de Otimidade

#### 3.2.1. Planejamentos D-ótimo e G-ótimo

Um dos critérios de otimidade mais estudado é o critério de D-otimidade. Diremos que um planejamento  $\xi^*$  é D-ótimo para o modelo (3.1.1) se este minimiza o determinante da matriz de covariância de  $\hat{\beta}$  (o qual é também chamado a variância generalizada de  $\hat{\beta}$ ); isto é, se  $\xi^*$  maximiza  $|X'S^{-1}X|$ .

Um planejamento  $\xi'$  é chamado G-ótimo (ou minimax) para o modelo (3.1.1) se  $\xi'$  minimiza

$$\max_{-1 \leq x \leq 1} \text{Var} \left[ \sum_{r=0}^k \hat{\beta}_r x^r \right]$$

A equivalência dos critérios D-ótimo e G-ótimo (quando  $\xi$  é expresso como uma medida sobre  $[-1,1]$ ) foi provada por Kiefer e Wolfowitz (1960) (o chamado teorema da equivalência), encontrando que um planejamento é D-ótimo se, e somente se, ele é G-ótimo.

Para um planejamento D-ótimo, Hoel e Guest (1958) mostraram que  $np_i$  observações ( $i=0,1,\dots,k$ ) devem ser feitas para  $x_i$ , onde  $p_i = \frac{1}{k+1}$  (isto é, alocação igual) e os  $x_i$  são os  $k+1$  zeros de  $(1-x^2) \cdot P'_k(x)$ ,  $P_k(x)$  sendo o polinômio de Legendre de grau  $k$ .

A Tabela 3.2.1. abaixo nos fornece os valores dos  $x_i$  (para  $k=1,2,\dots,5$ ) onde deveremos fazer as observações para a obtenção do planejamento D-ótimo ou do planejamento minimax (G-ótimo).

Tabela 3.2.1. - Zeros de  $(1-x^2)P'_k(x)$ , para os casos  $k=1,2,3,4$  e  $5$ .

k	Valores de $x_i$
1	$\pm 1$
2	$\pm 1, 0$
3	$\pm 1, \pm \sqrt{\frac{1}{5}}$
4	$\pm 1, \pm \sqrt{\frac{3}{7}}, 0$
5	$\pm 1, \pm \sqrt{\frac{1}{3}(1+\sqrt{\frac{4}{7}})}, \pm \sqrt{\frac{1}{3}(1-\sqrt{\frac{4}{7}})}$

Observamos que os polinômios de Legendre são ortogonais sobre o intervalo  $[-1,1]$  com respeito a função peso constante  $w(x)=1$ , e que a fórmula de Rodrigues para esses polinô-

mios é

$$P_k(x) = \frac{1}{2^k k!} \frac{d^k}{dx^k} (x^2 - 1)^k,$$

onde a expansão de  $P_k(x)$  em potências de  $x$  é:

$$P_k(x) = \frac{(2k)!}{2^k (k!)^2} x^k - \frac{(2k-2)!}{2^k (k-1)! (k-2)!} x^{k-2} + \frac{(2k-4)!}{2^k (k-2)! (k-3)!} x^{k-4} + \dots$$

Em particular, temos que

$$P_0(x) = 1; P_1(x) = x; P_2(x) = \frac{3}{2} x^2 - \frac{1}{2}; P_3(x) = \frac{5}{2} x^3 - \frac{3}{2} x;$$

$$P_4(x) = \frac{35}{8} x^4 - \frac{30}{8} x^2 + \frac{3}{8}; P_5(x) = \frac{63}{8} x^5 - \frac{70}{8} x^3 + \frac{15}{8} x$$

e estes resultados acima foram usados para a elaboração da Tabela 3.2.1.

### 3.2.2. Planejamentos para fazer inferências sobre $\beta_k$

Se estivermos interessados especificamente apenas em  $\beta_k$ , o coeficiente do termo de maior grau da função de regressão polinomial, então, um critério de planejamento mais razoável reduz-se a encontrar o planejamento que minimiza  $\text{Var}(\hat{\beta}_k)$ . Neste caso a solução ótima é (Kiefer e Wolfowitz, 1959)

$$p_i = \frac{1}{2k}, \quad i=0, k \\ = \frac{1}{k}, \quad i=1, 2, \dots, k-1$$

nos pontos  $x_i = -\cos(\frac{i\pi}{k})$ ,  $i=0, 1, \dots, k$ , os quais são chamados pontos de Chebyshev.

A principal desvantagem deste critério é que ele é apropriado apenas para um problema muito limitado: estimar  $\beta_k$  ou testar hipóteses sobre  $\beta_k$ .

### 3.2.3. Planejamentos Ótimos para Previsão

#### Previsão Ótima

O problema básico que consideraremos agora é, como de terminar onde os  $x_i$  devem ser escolhidos no intervalo  $[-1,1]$  de modo a minimizar a variância do valor previsto de  $Y$ , isto é, minimizar  $\text{Var} \left[ \sum_{r=0}^k \hat{\beta}_r x^r \right]$  em algum ponto  $x$ , tal que  $|x| < 1$ . Esta variância é dada por,  $\text{Var}[\hat{Y}(x)] = \sigma^2 x' (X'S^{-1}X)^{-1} x$ , onde  $x' = (1, x, x^2, \dots, x^k)$ .

Se o número total de observações a ser tomado,  $n$ , é fixado, as proporções  $p_i$  a serem tomadas nos  $k+1$  pontos distintos  $x_i$  do intervalo  $[-1,1]$ , podem não produzir valores inteiros para o número de observações a ser tomado nos vários pontos, assim, se um planejamento ótimo requer que  $n_i = np_i$  observações sejam feitas em  $x_i$  e  $n_i$  não é inteiro, deveremos escolher o inteiro próximo a  $n_i$  para o número real de observações e, portanto, os resultados que obteremos serão considerados como aproximados apenas.

Quando somente  $k+1$  pontos de observações são escolhidos para estimar um polinômio de grau  $k$ , o estimador de mínimos quadrados  $\hat{Y}(x)$  passa através dos  $k+1$  pontos médios  $(x_i, \bar{Y}_i)$ ,  $i=0,1,\dots,k$ ; portanto, sua equação pode ser escrita na forma

$$\hat{Y}(x) = \sum_{i=0}^k L_i(x) \bar{Y}_i, \quad (3.2.1)$$

onde  $L_i(x)$  é o polinômio de Lagrange dado por,

$$L_i(x) = \frac{(x-x_0) \dots (x-x_{i-1})(x-x_{i+1}) \dots (x-x_k)}{(x_i-x_0) \dots (x_i-x_{i-1})(x_i-x_{i+1}) \dots (x_i-x_k)} \quad (3.2.2)$$

e portanto, segue de (3.2.1) que

$$\text{Var}[\hat{Y}(x)] = \sum_{i=0}^k L_i^2(x) \text{Var}[\bar{Y}_i] = \frac{\sigma^2}{n} \sum_{i=0}^k \frac{L_i^2(x)}{p_i} \quad (3.2.3)$$

O problema então reduz-se na escolha dos  $k+1$  valores de  $x_i$ ,  $i=0,1,\dots,k$  e os correspondentes valores  $p_i$  que minimizam (3.2.3). Os dois resultados que seguem devidos a Hoel e Levine (1964) nos dão esses valores.

Se os  $p_i$ ,  $i=0,1,\dots,k$  variam continuamente em  $(0,1)$  sob a restrição que  $\sum_{i=0}^k p_i = 1$ , então para um fixado  $x \neq x_i$ ,  $i=0,1,\dots,k$ , a escolha

$$p_i = \frac{|L_i(x)|}{\sum_{i=0}^k |L_i(x)|}, \text{ para } i=0,1,\dots,k \quad (3.2.4)$$

minimizará  $\text{Var}[\hat{Y}(x)]$ .

A prova deste resultado é feita, usando técnicas usuais de cálculo e, verificando por meio das derivadas segundas que o ponto crítico produz um mínimo relativo que é também mínimo absoluto.

Embora estamos assumindo que  $x$  é um ponto dado fora do intervalo  $[-1,1]$ , e portanto seria desnecessário inserir a restrição  $x \neq x_i$ , isto é feito para assinalar o fato que esses  $p$ 's produzem um mínimo se  $x$  é interno ou externo a  $[-1,1]$ , de tal modo que  $x$  não seja escolhido como um ponto de observação.

Se os  $p$ 's dados pela fórmula (3.2.4) são usados, os  $k+1$  pontos de observação que minimizarão  $\text{Var}[\hat{Y}(x)]$  para  $x > 1$  são os pontos de Chebyshev

$$x_i = -\cos \frac{i\pi}{k}, \quad \text{para } i=0,1,\dots,k \quad (3.2.5)$$

Para ilustrar a vantagem de usar o espaçamento e ponderação ótimos dados pelas fórmulas (3.2.4) e (3.2.5), consideremos o problema de prever o valor de um polinômio do 3º grau no ponto  $x = 2$  se 52 observações são tomadas no intervalo  $[-1,1]$  e se  $\sigma=1$ .

Sob a tradicional aproximação de usar espaços e pesos iguais, escolheríamos

$$\begin{array}{cccc} x_0 = -1, & x_1 = -\frac{1}{3}, & x_2 = \frac{1}{3}, & x_3 = 1 \\ n_0 = 13, & n_1 = 13, & n_2 = 13, & n_3 = 13 \end{array}$$

Os valores correspondentes para a solução de Chebyshev dados pelas fórmulas (3.2.4) e (3.2.5) serão:

$$\begin{array}{cccc} x_0 = -1, & x_1 = -\frac{1}{2}, & x_2 = \frac{1}{2}, & x_3 = 1 \\ n_0 = 5, & n_1 = 12, & n_2 = 20, & n_3 = 15 \end{array}$$

Os valores correspondentes da  $\text{Var}[\hat{Y}(2)]$  são aproximadamente, 20 e 13. Assim, para este exemplo o método tradicional produz um valor de variância que é cerca de 50% maior do que a solução ótima apresentada pelo método de Hoel e Levine (1964).

### 3.2.4. Planejamentos C-ótimos

Para obter os planejamentos C-ótimos, vamos supor que para cada  $x \in [-1,1]$  um experimento possa ser realizado, o resultado do qual é uma variável aleatória  $Y(x)$  com

$$E[Y(x)] = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k; \text{ Var}[Y(x)] = \sigma^2 \quad (3.2.6)$$

Com o objetivo de estimar os coeficientes de regressão  $\beta_i$ ,  $i=0,1,\dots,k$ , pelo método dos mínimos quadrados, utilizaremos um planejamento  $\xi$ . Para isso, faremos  $n$  observações não-correlacionadas nos pontos  $x_i$ ,  $i=0,1,\dots,k$  com  $x_i \in [-1,1]$  tais que  $np_i = n_i$  observações serão feitas em  $x_i$ , onde  $p_i > 0$  e  $\sum_{i=0}^k p_i = 1$ .

Para um dado planejamento  $\xi$ , definiremos a matriz  $M(\xi)$  de ordem  $(k+1) \times (k+1)$  por

$$M_{rs}(\xi) = \sum_{i=0}^k p_i x_i^{r+s} \quad \text{para } r=0,1,\dots,k \quad (3.2.7) \\ s=0,1,\dots,k$$

$M(\xi)$  é chamada matriz planejamento.

A matriz de variância-covariância das estimativas de mínimos quadrados dos  $\beta_i$  é dada por  $\frac{\sigma^2}{n} M^{-1}(\xi)$ . A variância do estimador de mínimos quadrados, de uma função linear arbitrária dos coeficientes de regressão,  $(c, \beta) = \sum_{i=0}^k c_i \beta_i$  é proporcional a

$$c' M^{-1}(\xi) c = \text{tr } M^{-1}(\xi) c c' \quad (3.2.8)$$

onde  $c'$  representa o vetor linha  $(c_0, c_1, \dots, c_k)$  e  $c$  é sua transposta e,  $\text{tr}$  representa o traço da matriz.

Um planejamento  $\xi^*$  que minimiza (3.2.8) é chamado um planejamento C-ótimo, para estimar a forma linear  $(c, \beta)$ .

Para obter o planejamento  $\xi^*$  que minimiza (3.2.8) usaremos a caracterização de Kiefer e Wolfowitz (1959, 1965) de tais planejamentos, procedendo do seguinte modo:

Seja  $\xi$  um planejamento para  $k+1$  pontos distintos  $x_i$ ,  $i=0,1,\dots,k$  com pesos correspondentes  $p_i$ . Consideremos a matriz de ordem  $(k+1) \times (k+2)$ , cuja  $(i+1)^{a}$  coluna é  $(1, x_i, x_i^2, \dots, x_i^k)$ ;  $i=0,1,\dots,k$  e a última coluna é  $(c_0, c_1, \dots, c_k)$ . Sem perda de generalidades supomos que  $x_0 < x_1 < \dots < x_k$ , onde cada  $x_i$  é um ponto da região experimental  $[-1, 1]$ .

Retirando a  $(i+1)^{a}$  coluna desta matriz, obtemos uma matriz quadrada de ordem  $(k+1)$ , cujo determinante será denotado por  $D_i(c)$ . Se os  $(k+1)$  determinantes  $D_i(c)$ ;  $i=0,1,\dots,k$  têm o mesmo sinal, isto é, todos positivos ou todos negativos, ou se esses determinantes alternam em sinal, então o planejamento  $\xi$  nos  $(k+1)$  pontos  $x_i$ ,  $i=0,1,\dots,k$  com pesos  $p_i = \frac{|D_i(c)|}{\sum_{i=0}^k |D_i(c)|}$ , é ótimo.

Ao verificarmos a permanência ou alternância dos sinais dos determinantes  $D_i(c)$ ;  $i=0,1,\dots,k$ , um determinante igual a zero poderá ser considerado positivo ou negativo arbitrariamente. Para um dado  $c' = (c_0, c_1, \dots, c_k)$  para o qual os determinantes ou mantêm o mesmo sinal ou alternam em sinal, os pontos  $x_i$  são dados por  $x_i = -\cos \frac{i\pi}{k}$ , para  $i=0,1,\dots,k$ .

Temos portanto, que um experimento realizado para os  $(k+1)$  pontos  $x_i$  é C-ótimo se o vetor  $c' = (c_0, c_1, \dots, c_k)$  é tal

que os determinantes  $D_i(c)$ , ou têm o mesmo sinal ou alternam em sinal. Esta é uma condição apenas suficiente para C-otimidade.

Como uma aplicação do planejamento C-ótimo vamos estudar os planejamentos ótimos para estimar a inclinação de uma regressão polinomial. A inclinação da regressão polinomial  $E[Y(x)] = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k$ , em um ponto fixado  $x \in [-1, 1]$  é

$$\beta_1 + 2\beta_2 x + \dots + k\beta_k x^{k-1} \quad (3.2.9)$$

Tomando  $c' = (0, 1, 2x, \dots, kx^{k-1})$ , (3.2.9) pode ser representada como  $(c, \beta)$  e assim a variância da estimativa de mínimos quadrados de (3.2.9) usando um planejamento  $\xi$  pode ser calculada utilizando (3.2.8). Estamos portanto interessado em obter o planejamento  $\xi^*$  que minimiza  $\text{tr } M^{-1}(\xi) c c'$  para esta escolha especial de  $c'$ .

Segue-se um estudo detalhado para a inclinação da regressão polinomial de 2º e 3º graus.

Para regressão quadrática tomamos  $k=2$  em (3.2.9) de modo que,  $c' = (0, 1, 2x)$  e para a regressão cúbica,  $k=3$ , e  $c' = (0, 1, 2x, 3x^2)$ .

#### Regressão Quadrática

O planejamento para os pontos  $x_i = -\cos \frac{i\pi}{2}$ ;  $i=0, 1, 2$ , isto é,  $x_0 = -1$ ,  $x_1 = 0$  e  $x_2 = 1$  será C-ótimo para estimar a inclinação  $(c, \beta)$ , onde  $c' = (0, 1, 2x)$  se os determinantes

$$D_0(c) = \begin{vmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2x \end{vmatrix} = 2x-1$$

$$D_1(c) = \begin{vmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \\ 1 & 1 & 2x \end{vmatrix} = 4x$$

e

$$D_2(c) = \begin{vmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 2x \end{vmatrix} = 2x+1$$

ou mantêm o mesmo sinal ou alternam em sinal. Se  $2x > 1$  ou  $2x < -1$  esses três determinantes são de mesmo sinal. Se  $|2x| < 1$ , então para  $x \in (0, \frac{1}{2})$ ,  $D_0(c) < 0$ ,  $D_1(c) > 0$  e  $D_2(c) > 0$  e para  $x \in (-\frac{1}{2}, 0)$ ,  $D_0(c) < 0$ ,  $D_1(c) < 0$  e  $D_2(c) > 0$  e para  $x=0$ ,  $D_0(c) < 0$ ,  $D_1(c) = 0$  e  $D_2(c) > 0$ , de modo que qualquer que seja o sinal que atribuirmos a  $D_1(c)$  não existe alternância de sinais. Assim o planejamento  $\xi$  concentrando-se sobre os pontos

$$x_0 = -1, \quad x_1 = 0 \quad \text{e} \quad x_2 = 1, \quad \text{com pesos}$$

$$p_i = \frac{|D_i(c)|}{\sum_{i=0}^2 |D_i(c)|}; \quad \text{para } i=0,1,2$$

é ótimo para estimar a inclinação em um ponto  $x$ , onde  $|2x| \geq 1$ .

Os pesos reais são

$$p_0 = \frac{1}{4} - \frac{1}{8x}; \quad p_1 = \frac{1}{2}; \quad p_2 = \frac{1}{4} + \frac{1}{8x}$$

### Regressão Cúbica

Tomando  $k=3$ , temos que  $c' = (0, 1, 2x, 3x^2)$ .

Calculando os determinantes  $D_0(c)$ ,  $D_1(c)$ ,  $D_2(c)$  e  $D_3(c)$  nos pontos  $x_i = -\cos \frac{i\pi}{3}$ ,  $i=0,1,2,3$ , isto é,  $x_0=-1$ ,  $x_1=-\frac{1}{2}$ ,  $x_2=\frac{1}{2}$  e  $x_3=1$ , temos

$$D_0(c) = \begin{vmatrix} 1 & 1 & 1 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 1 & 1 \\ \frac{1}{4} & \frac{1}{4} & 1 & 2x \\ \frac{1}{8} & \frac{1}{8} & 1 & 3x^2 \end{vmatrix} = \frac{1}{16} [36x^2 - 24x - 3]$$

$$D_1(c) = \begin{vmatrix} 1 & 1 & 1 & 0 \\ -1 & \frac{1}{2} & 1 & 1 \\ 1 & \frac{1}{4} & 1 & 2x \\ -1 & \frac{1}{8} & 1 & 3x^2 \end{vmatrix} = \frac{1}{2} [9x^2 - 3x - 3]$$

$$D_2(c) = \begin{vmatrix} 1 & 1 & 1 & 0 \\ -1 & -\frac{1}{2} & 1 & 1 \\ 1 & \frac{1}{4} & 1 & 2x \\ -1 & -\frac{1}{8} & 1 & 3x^2 \end{vmatrix} = \frac{1}{2} [9x^2 + 3x - 3]$$

$$D_3(c) = \begin{vmatrix} 1 & 1 & 1 & 0 \\ -1 & -\frac{1}{2} & \frac{1}{2} & 1 \\ 1 & \frac{1}{4} & \frac{1}{4} & 2x \\ -1 & -\frac{1}{8} & \frac{1}{8} & 3x^2 \end{vmatrix} = \frac{1}{36} [36x^2 + 24x - 3]$$

Esses determinantes mantêm o mesmo sinal se a)  $x \in [-1, -2 - \frac{\sqrt{7}}{6}]$  ou b)  $x \in [2 - \frac{\sqrt{7}}{6}, -2 + \frac{\sqrt{7}}{6}]$  ou c)  $x \in [2 + \frac{\sqrt{7}}{6}, 1]$ . Assim nesses três casos o planejamento será ótimo para os pontos  $x_i = -\cos \frac{i\pi}{3}$ ,  $i=0,1,2,3$ , isto é,  $x_0 = -1$ ,  $x_1 = -\frac{1}{2}$ ,  $x_2 = \frac{1}{2}$  e  $x_3 = 1$  com pesos

$$p_i = \frac{|D_i(c)|}{\sum_{i=0}^3 |D_i(c)|}; \quad i=0,1,2,3.$$

Como um caso particular do planejamento C-ótimo, temos que quando  $c' = (1, 0, \dots, 0)$  este planejamento torna-se o planejamento ótimo para fazer inferências sobre  $\beta_k$ , isto é,

$$\begin{aligned} p_i &= \frac{1}{2k}, & i=0, k \\ &= \frac{1}{k}, & i=1, 2, \dots, k-1 \end{aligned}$$

nos pontos  $x_i = -\cos(\frac{i\pi}{k})$ ,  $i = 0, 1, \dots, k$  o qual foi dado em (3.2.2).

As soluções acima para o problema de planejamento tem uma séria desvantagem. Nós supomos que  $k$  é conhecido e  $Y$  é observado em apenas  $k+1$  valores diferentes de  $x$ , de modo que o ajuste de mínimos quadrados (ponderado) usa que a média dos  $Y$  para cada  $x_i$  é exata, com nenhum resíduo para examinar as su posições básicas. Desde que na prática,  $k$  não é geralmente co nhecido, uma solução ótima pode não ser satisfatória quando esta não toma precaução para examinar a adequação do ajuste. Para contornar esse problema Box e Draper (1959; 1963) definiram otimidade em termos de minimizar o erro médio quadrático integrado da previsão; a atenção foi dada ao termo viés qua-

drático porque este termo pareceu ser o fator de maior contribuição. Karson e outros (1969) e Cote e outros (1973) tomaram um ponto de vista análogo e consideraram planejamentos que, na primeira instância, minimizavam o viés quadrático integrado. Stigler (1971), entretanto, criticou o critério de viés e sugeriu modificações dos planejamentos usuais D-ótimos e minimax. Sua modificação é denominada um planejamento C-restrito quando é imposta a condição que, se o termo adicional  $\beta_{k+1}x^{k+1}$  é ajustado,

$$\text{Var}[\hat{\beta}_{k+1}] \leq \frac{\sigma^2 C}{n},$$

onde C é pré-escolhido. A escolha de C reflete um compromisso entre duas metas conflitantes: inferências precisas sobre  $\beta_{k+1}$ , e inferências precisas sobre o polinômio de grau k. Por outro lado, C deveria ser escolhido suficientemente pequeno de modo que afastamentos praticamente significantes do modelo pudessem ser detectados com uma precisão especificada (por ex., o teste  $H:\beta_{k+1}=0$  com um poder especificado); por outro lado, valores grandes de C produzirão planejamentos mais eficientes para o ajuste polinomial de grau k. Infelizmente, encontrar a solução ótima C-restrita não é fácil, e apenas a solução para  $k=1$  é dada por Stigler.

## CAPÍTULO 4

### DETERMINAÇÃO DO GRAU DE UMA FUNÇÃO POLINOMIAL

#### 4.1.1. Introdução

Em análise de regressão, frequentemente supomos que a função de regressão é um polinômio de grau conhecido. Frequentemente, ocorre entretanto, que estamos incertos sobre que grau polinomial será adequado para nossos propósitos. Experiências têm mostrado que apenas polinômios com graus menores ou iguais a cinco têm valor prático. O problema então reduz-se a decidir que grau polinomial, não excedendo cinco, deve ser escolhido para a função de regressão.

Várias técnicas tem sido sugeridas para determinar o grau apropriado. Uma das últimas é devida a Anderson (1962), onde ele trata o problema como um problema de decisão múltipla e, encontra que a solução ótima reduz-se a uma sequência de testes para o grau, começando com o maior grau possível. Uma desvantagem desses vários procedimentos é que eles garantem proteção apenas para o erro de tipo I. Assim, no teste de Anderson, somos protegidos contra a afirmação que o grau é menor do que  $k$ , quando este é na verdade, pelo menos  $k$ .

Se o tamanho da amostra que estamos empregando é pequeno, a probabilidade que um polinômio de grau baixo seja compatível com os dados será grande, mesmo que a função de regressão seja de grau maior, assim existirá uma tendência para testes baseados sobre uma amostra de tamanho fixado (pequeno), para aceitar modelos polinomiais de grau menor do que o adequado.

Para amostras bastante grandes, existirá uma tendência para testes baseados sobre uma amostra de tamanho fixado (grande), para decidir qual será o grau máximo, mesmo que um grau polinomial menor produzisse um ajuste satisfatório de um ponto de vista prático. Esta tendência para testes com tamanho de amostras fixado, ser influenciado pelo tamanho da amostra ao tomarmos uma decisão, poderá ser eliminada se empregarmos um teste do tipo sequencial.

Nosso objetivo neste capítulo é mostrar como duas técnicas sequenciais poderão ser usadas para produzir uma solução prática satisfatória, para o problema de determinação do grau de uma função polinomial, como também, verificar a perda de precisão que estaremos sujeitos, quando alocamos as observações para um grau polinomial maior do que o grau suposto verdadeiro.

#### 4.1.2. Notações

Para os dois métodos que apresentaremos, para decidirmos sobre o grau apropriado de uma função de regressão polino

mial, vamos supor que as variáveis  $Y(x_i)$ ,  $i=1,2,\dots,n$  são distribuídas de acordo com o modelo clássico de regressão, com função de regressão

$$E[Y(x)] = \gamma_0\phi_0(x) + \gamma_1\phi_1(x) + \dots + \gamma_k\phi_k(x),$$

onde  $\phi_i(x)$  denota um polinômio de grau  $i$ .

Por conveniência, escolhemos os  $\phi$ 's formando um sistema ortonormal com respeito aos pontos sobre o eixo  $x$ , onde as observações serão tomadas. Vamos supor também que o grau máximo permitido é  $k$  e, que tomaremos as observações em pontos do intervalo  $[-1,1]$ .

Em geral, os dados usados para determinar o grau da regressão polinomial, são usados para estimar o polinômio e, este fato deve ser levado em conta, na escolha dos pontos de observação; se réplicas para uns poucos pontos de observações no intervalo, são mais viáveis de se obter, do que o mesmo número total de observações em pontos distintos do intervalo, então devemos escolher os pontos onde réplicas ocorrem, por algum critério de otimismo. Por exemplo, se o polinômio estimado é para ser usado para propósitos de interpolação, uma escolha ótima é dada pelo conjunto de pontos de Legendre e, se é para ser usado para extrapolação, os pontos de Chebyshev possuem propriedades ótimas. Em ambos os casos,  $k+1$  pontos são usados, sob a suposição que o grau máximo a ser considerado é  $k$ . Embora os pontos ótimos, são ótimos para grau  $k$  apenas, eles ainda possuem vantagens sobre pontos igualmente espaçados, quando o grau é menor do que  $k$ . Apesar da solução ótima para

extrapolação produzir réplicas desiguais nos  $k+1$  pontos, o uso de réplicas iguais não reduzirá a eficiência da extrapolação apreciavelmente.

Para o estudo dos pontos destacados acima, suporemos que as observações serão tomadas nos pontos  $x_0, x_1, \dots, x_k$ , onde  $x_0 = -1$  e  $x_k = 1$  e, os  $x$ 's restantes serão determinados de acordo com nosso objetivo no ajuste da função de regressão estimada,

$$\hat{Y}(x) = \hat{\gamma}_0 \phi_0(x) + \hat{\gamma}_1 \phi_1(x) + \dots + \hat{\gamma}_k \phi_k(x)$$

Em virtude da ortogonalidade dos  $\phi$ 's com relação aos pontos  $x_0, x_1, \dots, x_k$ , os  $\hat{\gamma}$ 's serão independentes e normalmente distribuídos, e, se  $m$  réplicas forem feitas, teremos:

$$\hat{\gamma}_j = \frac{1}{m} \sum_{\alpha=1}^m \sum_{i=0}^k Y_{i\alpha} \phi_j(x_i)$$

$$\text{Var}\{\hat{\gamma}_j\} = \frac{\sigma^2}{m} \tag{4.1.1}$$

$$\text{Var}\{\hat{Y}(x)\} = \frac{\sigma^2}{m} \sum_{j=0}^k \phi_j^2(x)$$

#### 4.2. Teste Sequencial F

Suponhamos que estamos bastante seguros que o grau da função de regressão polinomial é  $k$ , ou possivelmente  $k-1$ , mas certamente não maior do que  $k$ . Então nosso problema é testar a hipótese  $H_0: \gamma_k = 0$ . Com o propósito de encontrar uma hipótese alternativa apropriada, consideremos as seguintes rela

ções:

Se  $\sigma$  fosse conhecido, das fórmulas (4.1.1) implicaria que, com uma probabilidade de aproximadamente 0,95, o erro de estimar  $Y(x)$  por meio de  $\hat{Y}(x)$  não excederá

$$2 \frac{\sigma}{\sqrt{m}} \left[ \phi_0^2(x) + \phi_1^2(x) + \dots + \phi_k^2(x) \right]^{\frac{1}{2}}$$

Se supormos que  $\gamma_k = 0$ , esse erro máximo torna-se

$$2 \frac{\sigma}{\sqrt{m}} \left[ \phi_0^2(x) + \phi_1^2(x) + \dots + \phi_{k-1}^2(x) \right]^{\frac{1}{2}}$$

Denotando a diferença desses dois erros por  $d(x)$ , então se  $d(x)$  é maior do que a diferença dos valores ordenados da verdadeira função de regressão e, a função com  $\gamma_k = 0$  no ponto  $x$ , existe pouca evidência para mantermos o termo com coeficiente  $\gamma_k$ . Assim, parece apropriado considerar, se

$$2 \frac{\sigma}{\sqrt{m}} \left\{ \left[ \phi_0^2(x) + \dots + \phi_k^2(x) \right]^{\frac{1}{2}} - \left[ \phi_0^2(x) + \dots + \phi_{k-1}^2(x) \right]^{\frac{1}{2}} \right\} > \left| \gamma_k \phi_k(x) \right| \quad (4.2.1)$$

que é equivalente a

$$\left| \frac{\gamma_k}{\sigma} \right| < \frac{2 \left\{ \left[ \phi_0^2(x) + \dots + \phi_k^2(x) \right]^{\frac{1}{2}} - \left[ \phi_0^2(x) + \dots + \phi_{k-1}^2(x) \right]^{\frac{1}{2}} \right\}}{\sqrt{m} \left| \phi_k(x) \right|} \quad (4.2.2)$$

Se soubermos aproximadamente, quantas réplicas deveremos tomar, o lado direito de (4.2.2) pode ser avaliado para qualquer ponto  $x$  no intervalo  $[-1,1]$ . O valor para  $x=1$  seria suficiente para interpolação, desde que o erro neste ponto, provavelmente está próximo de seu valor máximo. Para extrapolação será necessário usar o valor de  $x$ , onde extrapolação é

observada. Se o valor para  $x=1$ , ou o valor no ponto de extrapolação, sobre o lado direito de (4.2.2) for denotado por  $\delta$ , então temos, a menos que  $|\gamma_k| > \sigma\delta$ , não existirá vantagem apreciável em usar um polinômio de grau  $k$  sobre um de grau  $k-1$ . Poderemos então, formular o problema como um teste de

$$H_0: \gamma_k = 0 \quad \text{contra} \quad H_1: \left| \frac{\gamma_k}{\sigma} \right| > \delta$$

No caso de não termos certeza sobre o grau  $k$  ou  $k-1$  a hipótese mais adequada é  $H_0: \gamma_s = \dots = \gamma_k = 0$ , onde  $s < k$ . De modo análogo, a relação (4.2.1) terá  $k-1$  substituído por  $s-1$ , isto é,

$$2 \frac{\sigma}{\sqrt{m}} \left\{ \left[ \phi_0^2(x) + \dots + \phi_k^2(x) \right]^{\frac{1}{2}} - \left[ \phi_0^2(x) + \dots + \phi_{s-1}^2(x) \right]^{\frac{1}{2}} \right\} > \left| \gamma_s \phi_s(x) + \dots + \gamma_k \phi_k(x) \right|$$

Temos que:  $\sum_j \gamma_j \phi_j = \cos \theta \left[ \sum_j \gamma_j^2 \sum_j \phi_j^2 \right]^{\frac{1}{2}}$ , onde  $\theta$  é o ângulo entre os vetores  $\gamma$  e  $\phi$ . Assim, a desigualdade análoga a (4.2.2) será:

$$\frac{\sum_{j=s}^k \gamma_j^2}{\sigma^2} < \frac{4 \left\{ \left[ \phi_0^2(x) + \dots + \phi_k^2(x) \right]^{\frac{1}{2}} - \left[ \phi_0^2(x) + \dots + \phi_{s-1}^2(x) \right]^{\frac{1}{2}} \right\}^2}{m \left[ \phi_s^2(x) + \dots + \phi_k^2(x) \right] \cos^2 \theta} \quad (4.2.3)$$

O valor mais comumente usado para  $\cos \theta$  é 1, embora seja improvável que os vetores  $\gamma$  e  $\phi$  sejam paralelos para qualquer valor de  $x$ , que torne a diferença dos erros grande. Se o valor em  $x=1$ , ou o valor no ponto de extrapolação, sobre o lado direito de (4.2.3) com  $\cos \theta = 1$  é denotado por  $\lambda_0$ , temos que

não existirá vantagem apreciável em usar um polinômio de grau  $s$  ou maior sobre um de grau  $s-1$ , a menos que,  $\sum_{j=s}^k \gamma_j^2 > \sigma^2 \lambda_0$ . Consequentemente poderemos formular o problema como um teste de

$$H_0: \gamma_s = \dots = \gamma_k = 0 \quad \text{contra} \quad H_1: \frac{\sum_{j=s}^k \gamma_j^2}{\sigma^2} > \lambda_0 \quad (4.2.4)$$

O teste sequencial  $F$  que apresentaremos a seguir, é capaz de tratar esses dois problemas com valores fixados dos tamanhos dos erros de tipo I e tipo II escolhidos "a priori". Se  $\delta$  e  $\lambda_0$  são muito pequenos, o teste sequencial pode requerer um valor grande de  $E(n)$  antes que uma decisão seja tomada. É aconselhável considerar valores maiores de  $\delta$  e  $\lambda_0$  do que aqueles dados pelas relações acima e, verificar se tais valores são aceitáveis.

O teste sequencial  $F$  (Hoel, 1954) foi designado para testar a hipótese linear geral na forma canônica. Este teste supõe que as variáveis básicas  $Y_1, Y_2, \dots, Y_\ell$  tem médias  $\mu_1, \mu_2, \dots, \mu_\ell$  e variâncias comuns  $\sigma^2$  e possui a função de densidade

$$f(Y_1, Y_2, \dots, Y_\ell) = (\sqrt{2\pi} \sigma)^{-\ell} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^k (Y_i - \mu_i)^2 + \sum_{i=k+1}^{\ell} Y_i^2 \right] \right\} \quad (4.2.5)$$

e testa

$$H_0: \mu_1 = \dots = \mu_p = 0 \quad \text{contra} \quad H_1: \frac{\sum_{i=1}^p \mu_i^2}{\sigma^2} > \lambda_0, \quad p \leq k$$

Nosso problema de regressão, poderá ser reduzido para a forma canônica por meio da transformação  $Z = \Phi'Y$  onde  $\Phi$  é a matriz  $(\phi_j(x_i))$ . Desde que os  $\phi$ 's formam um sistema ortonormal com respeito aos  $x$ 's,  $\Phi'$  é uma matriz ortogonal, e portan

to os Z's são independentes e normalmente distribuídos com variâncias  $\sigma^2$  e médias dadas por:

$$v = E[Z] = \Phi' E[Y] = \Phi' \Phi \gamma = \gamma$$

A distribuição dos Z's é portanto, dada por:

$$f(Z_0, \dots, Z_k) = (\sqrt{2\pi} \sigma)^{-(k+1)} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=0}^k (Z_i - \gamma_i)^2\right\} \quad (4.2.6)$$

Uma comparação de (4.2.5) e (4.2.6) nos mostra que (4.2.4) é um caso especial de um teste da hipótese linear geral na forma canônica.

O teste sequencial descrito por Hoel (1954) por (4.2.5) quando adaptado para o problema que estamos tratando e suas notações, torna-se o teste sequencial da razão de probabilidade, baseado sobre

$$\frac{P_{1n}}{P_{0n}} = e^{-\frac{n\lambda_0}{2}} F\left(\frac{1}{2}[nk+n-s-2], \frac{k-s+1}{2}, \frac{\delta^2}{2\lambda_0}\right) \quad (4.2.7)$$

onde

$$\delta^2 = n \lambda_0^2 \frac{n \sum_{i=s}^k \bar{z}_i^2}{\sum_{\alpha=1}^n \sum_{i=0}^k z_{i\alpha}^2 - n \sum_{i=0}^{s-1} \bar{z}_i^2} \quad (4.2.8)$$

sendo F, a função hipergeométrica confluyente, que é dada por:

$$F(a, b, x) = 1 + \frac{a}{b} x + \frac{a(a+1)}{b(b+1)} \frac{x^2}{2!} + \frac{a(a+1)(a+2)}{b(b+1)(b+2)} \frac{x^3}{3!} + \dots$$

Uma vez que os tamanhos dos erros de tipo I e tipo II

foram escolhidos "a priori", existe proteção contra decidir que o grau é s ou maior, quando ele é menor, e também existe proteção contra decidir que o grau é menor do que s quando ele é de fato maior, no sentido que  $\sum_{i=s}^k \gamma_i^2 > \sigma^2 \lambda_0$ .

### 4.3. Teste com duas Amostras

Se não ficamos satisfeitos com a maneira de escolher  $\delta$  ou  $\lambda_0$  no teste anterior e, preferíssemos ter uma desigualdade do tipo  $|\gamma_i| > n$  satisfeita, antes de considerar  $\gamma_i$  diferente de zero, então um teste que nos dá esse tipo de proteção, pode ser construído usando o procedimento de duas amostras de Stein (1945), quando este é aplicado para o problema da hipótese linear geral. Quando aplicarmos este teste para o problema dado por (4.2.4) e (4.2.6), devemos proceder do seguinte modo:

Escolhemos uma constante  $c > 0$  e um tamanho amostral  $n_0$ . Tomamos uma amostra deste tamanho e calculamos

$$S^2 = \frac{1}{(n_0 - 1)(k + 1)} \left\{ \sum_{j=1}^{n_0} \sum_{i=0}^k z_{ij}^2 - \frac{1}{n_0} \sum_{i=0}^k \left( \sum_{j=1}^{n_0} z_{ij} \right)^2 \right\}$$

Seja  $n = \max \left\{ \left[ \frac{S^2}{c} \right] + 1, n_0 + 1 \right\}$ . Escolhemos números  $a_1, \dots, a_n$  satisfazendo  $\sum_{j=1}^n a_j = 1$ ,  $a_1 = \dots = a_{n_0}$ ,  $\sum_{j=1}^n a_j^2 = \frac{c}{S^2}$ .  
Então,

$$F' = \frac{\left( \sum_{i=s}^k \sum_{j=1}^n a_j z_{ij} \right)^2}{c(n_0 - 1)(k + 1)} \quad (4.3.1)$$

possuirá uma distribuição F não-central, dada por

$P\{F' < \ell\} = \phi_{p,q}(\ell, m)$ , onde

$$\phi_{p,q}(\ell, m) = \frac{\Gamma(\frac{p+q}{2})}{\sqrt{\pi} \Gamma(\frac{q}{2}) \Gamma(\frac{p-1}{2})}$$

$$\int_0^{\ell} \int_{-\sqrt{F'}}^{\sqrt{F'}} [F' - t^2]^{(p-3)/2} [1 + F' + 2t\sqrt{m} + m]^{-\frac{(p+q)}{2}} dt dF'$$

onde  $p = k - s + 1$ ,  $q = (n_0 - 1)(k + 1)$ , e  $m = \sum_{i=s}^k \frac{\gamma_i^2}{c(n_0 - 1)(k + 1)}$ .

Se estivermos testando  $\gamma_s = \dots = \gamma_k = 0$  e o tamanho do erro de tipo I,  $\alpha$ , foi escolhido, é necessário determinar o valor de  $\ell$  que satisfaz

$$P\{F_{k-s+1, (n_0-1)(k+1)} > \ell\} = \alpha \tag{4.3.2}$$

onde F é a variável F central. O tamanho do erro de tipo II deste teste é dado por

$$\phi_{k-s+1, (n_0-1)(k+1)} \left[ \ell, \frac{\sum_{i=s}^k \gamma_i^2}{c(n_0-1)(k+1)} \right] \tag{4.3.3}$$

Se pudermos prescrever um limite superior, digamos  $\eta^2$ , para o valor de  $\sum_{i=s}^k \gamma_i^2$  além do qual, trataremos esses coeficientes como não sendo todos iguais a zero, então poderemos calcular o tamanho do erro de tipo II com base na hipótese alternativa  $H_1: \sum_{i=s}^k \gamma_i^2 > \eta^2$ . Assim, é possível escolher  $\ell$ ,  $n_0$  e  $c$  antes da amostragem, a fim de determinar os tamanhos dos dois tipos de

erros para quaisquer níveis desejados. Alguns cuidados devem ser tomados nestas escolhas, pois  $E(n)$  poderá tornar-se grande se os tamanhos desses erros forem escolhidos muito pequenos.

#### 4.4. Uma Modificação na Decisão Múltipla

Os dois procedimentos tipo sequencial que apresentamos produzem duplo erro de proteção contra uma decisão incorreta, contudo eles estão restritos para testar a hipótese  $\gamma_k=0$  ou a hipótese  $\gamma_s=\dots=\gamma_k=0$ , onde  $s < k$ . Eles não permitem testar graus menores crescentemente, se um grau maior é rejeitado. Isto não é um problema sério do ponto de vista prático, pois se o grau máximo permitido é escolhido como sendo  $k=5$ , usualmente estaríamos satisfeitos com grau 4, ou certamente, com grau 3, e portanto a técnica precedente com  $s \geq 3$  seria satisfatória.

É possível, contudo, modificar os dois testes anteriores de modo a torná-los um teste de decisão múltipla e, ao mesmo tempo permitir-nos calcular os tamanhos dos dois tipos de erros que podem surgir neste teste. Suponhamos que estamos testando  $H_0: \gamma_s = \dots = \gamma_k = 0$ . Se  $H_0$  é aceita, o problema está terminado. Se  $H_0$  é rejeitada, permanece ainda o problema de determinar o grau apropriado. A região crítica do teste de  $H_0$  dado por (4.3.2) é o exterior da esfera  $c(n_0-1)(k+1)F' = \ell$  nas variáveis  $u_i = \sum_{j=1}^n a_j Z_{ij}$ ,  $i = s, \dots, k$ , onde  $F'$  é dado por (4.3.1). A rejeição de  $H_0$ , portanto ocorrerá quando os valores

amostrais dos  $u$ 's produzirem um ponto fora da esfera  $u_s^2 + \dots + u_k^2 = r^2 = \ell c (n_0 - 1)(k + 1)$ . Quando  $H_0$  é rejeitada devemos primeiro verificar se  $u_k^2 < r^2$ ; se for, a seguir verificamos se  $u_{k-1}^2 + u_k^2 < r^2$ , e continuamos desta maneira até que a desigualdade não seja satisfeita. Se  $u_{m+1}^2 + \dots + u_k^2 < r^2$  denota a última desigualdade a ser satisfeita, concluímos então, que o grau é  $m$ .

Podemos calcular os tamanhos dos dois tipos de erros que podem surgir no procedimento anterior, em vista da natureza da distribuição dos  $u$ 's. Dado  $s^2$ , eles possuem distribuição normal independente com  $E(u_i) = \gamma_i$  e com uma variância comum  $\sigma^2 c / s^2$ . Resulta então que, sua distribuição incondicional é dada por

$$f(u_s, \dots, u_k) = \frac{(vc)^{\frac{v}{2}} \Gamma\left(\frac{v+k-s+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right) \pi^{\frac{k-s+1}{2}}} (vc + \sum_{i=s}^k (u_i - \gamma_i)^2)^{-\frac{(v+k-s+1)}{2}}$$

onde  $v = (n_0 - 1)(k + 1)$ . Assim, se  $D$  denotar grau e,  $D_m$  denota decidir se o grau é  $m$ , resultará

$$P\{D_m | D < m\} = P\{u_{m+1}^2 + \dots + u_k^2 < r^2, u_m^2 + \dots + u_k^2 > r^2 |$$

$$\gamma_m = \dots = \gamma_k = 0\} = \int_{R_m} \dots \int f(u_s, \dots, u_k) du_s \dots du_k$$

onde  $R_m$  representa a região especificada pelas desigualdades e na qual  $\gamma_m, \dots, \gamma_k$  foram dados valores zero. Em vista da geometria da região  $R_m$ , esta integral é independente dos  $\gamma$ 's

desconhecidos, cujos valores podem portanto serem escolhidos como sendo zeros. Uma vez que esta integral, então depende somente de constantes conhecidas, a probabilidade de decidir sobre grau  $m$  quando o grau  $\hat{e}$  menor do que  $m$  pode ser avaliada por métodos de integração numérica. A seguir, consideremos

$$P\{D_m | D > m\} = P\{u_{m+1}^2 + \dots + u_k^2 < r^2, u_m^2 + \dots + u_k^2 > r^2 \mid \sum_{j=m+1}^k \gamma_j^2 > \eta^2\}$$

Como antes, a geometria de  $R_m$  mostra que a integral necessária para avaliar esta probabilidade  $\hat{e}$  independente dos valores de  $\gamma_s, \dots, \gamma_{m-1}$ , de modo que podemos atribuir-lhes valores zero. Por simetria, a integral tem o mesmo valor para todos os pontos da esfera  $\sum_{j=m+1}^k \gamma_j^2 = \eta^2$ ; assim esta probabilidade será maximizada, quando um dos  $\gamma$ 's nesta soma tiver o valor  $\eta$  e os restantes forem colocados iguais a zero. Esta integral  $\hat{e}$  então maximizada, permitindo  $\gamma_m$  tornar-se infinito, que  $\hat{e}$  numericamente equivalente a colocar  $\gamma_m = 0$  e integrar sobre a variação inteira dos valores de  $u_m$ , sem a restrição imposta sobre  $u_m$  por  $R_m$ . Assim, a probabilidade máxima, de decidir sobre grau  $m$ , quando o grau  $\hat{e}$  maior do que  $m$ , pode também ser avaliada por métodos numéricos.

$\hat{E}$  improvável que este procedimento modificado tenha quaisquer propriedades ótimas úteis, contudo este  $\hat{e}$  o teste natural a se aplicar se insistimos sobre um teste que possua proteção contra os dois tipos de erros e, se não ficamos satisfeitos com os testes simples de  $H_0$ , dados pelo teste sequencial  $F$  ou pelo teste com duas amostras.

No capítulo seguinte, um exemplo é apresentado para ilustrar os métodos dados acima, para um polinômio do quarto grau.

Uma vez que esses procedimentos para testar o grau de uma curva de regressão polinomial não são fáceis de serem aplicados, além de produzirem erros de proteção contra uma decisão incorreta, o que acontece é que, quando escolhemos os pontos de observação no intervalo  $[-1,1]$  com o propósito de estimar curvas de regressão polinomial, frequentemente estamos incertos sobre o grau  $k$  do polinômio na variável independente  $x$ . Veremos agora a influência desta incerteza sobre o problema de alocar observações, a fim de minimizar o máximo da variância do valor previsto pelo modelo, para um dado  $x$  pertencente ao intervalo. Por exemplo, se  $k=1$  a alocação minimax consiste em tomar metade das observações possíveis em cada um dos extremos do intervalo (conforme Tab. 3.2.1.), mas esta alocação não nos proporciona oportunidade de detectarmos evidência que  $k=2$ , por exemplo.

A seguir, estudaremos a perda de precisão que incorremos por usar alocação minimax para um grau polinomial maior do que o grau verdadeiro, com atenção particular para os casos onde o grau verdadeiro  $k$  é igual a 1 ou 2. Discutiremos também, que se ocorresse um possível grau alternativo  $k_1$  um ou dois graus maiores do que o presumido grau  $k_0$ , nós alocaríamos as observações como se  $k=k_1$ , desde que se de fato  $k=k_0$  a perda de precisão que incorremos por usar alocações para  $k_1$  se

rã relativamente pequena.

Seja novamente o modelo usual de regressão na forma

$$Y = \beta_0 + \beta_1 x + \dots + \beta_k x^k + \varepsilon$$

onde os  $\beta$ 's são estimados pelo método dos mínimos quadrados de uma amostra de  $n$  observações. Seleccionados  $n$  valores  $x$  para as observações, temos que a variância de  $\hat{Y}$  em um ponto  $x = x_0$  é dada por

$$\text{Var}(\hat{Y}) = x' C x \sigma^2$$

onde  $x' = (1 \quad x_0 \quad x_0^2 \quad \dots \quad x_0^k)$  e  $C$  denota a inversa da matriz informação ou equivalentemente, a matriz de covariância das estimativas dos  $\beta$ 's. Definiremos

$$V_k = \frac{n \text{Var}(\hat{Y})}{\sigma^2}$$

como sendo uma padronização conveniente, com respeito a  $n$  e  $\sigma^2$  da variância do valor previsto pelo modelo, o subíndice  $k$  representará sempre o verdadeiro grau da regressão polinomial.

#### 4.5. Perda de Precisão por Alocar as Observações para um grau maior

A alocação que minimaximiza  $V_k$ , consiste de  $\frac{n}{k+1}$  observações em cada um dos  $k+1$  pontos de observações (esses pontos são dados na Tab. 3.2.1. para os casos  $k=1,2,3,4$  e  $5$ ). Para evitar aproximações na alocação de pontos vamos supor que

$\frac{n}{k+1}$  é inteiro.

Examinaremos agora a possibilidade, de usar uma alocação minimax correspondente a um grau maior do que o grau su posto verdadeiro. Vamos denotar por  $k_0$ , o inteiro positivo que corresponde ao verdadeiro valor de  $k$ , o grau do polinômio; e seja  $k_1$  denotando um inteiro maior do que  $k_0$ , que é um possível valor alternativo para  $k$ . Se alocação minimax para  $k = k_0$  é feita, não teremos condições de examinar a possibilidade de que  $k=k_1$ , enquanto que, se alocação minimax para  $k=k_1$  é usada estaremos arriscando uma perda de precisão se de fato  $k=k_0$ . Definiremos uma medida desta perda de precisão por  $E(k_1, k_0)$ , como sendo a razão entre: o valor máximo de  $V_{k_0}$  sobre  $[-1, 1]$  usando alocação minimax para  $k_0$  e o correspondente máximo de  $V_{k_0}$  usando alocação minimax para  $k_1$ .

#### 4.5.1. O caso $k_0=1$

Neste caso, a variância em um ponto  $X_0$  é igual a  $V_1 = 1 + \left[ \frac{n}{\sum_{i=1}^n x_i^2} \right] X_0^2$ . Para uma alocação simétrica arbitrária em torno de  $x = 0$ ;  $x_i$  representa a alocação da observação  $i$  em  $[-1, 1]$ . Sob alocação minimax para  $k_0=1$  temos conforme a Tabela 3.2.1. que os valores de  $x_i$  são 1 e -1 e portanto,  $V_1=1+X_0^2$ . Estamos interessados em examinar a perda de precisão ( quando  $k = k_0 = 1$ ) incorrida por alocarmos para vários valores alternativos de  $k = k_1$ . Sob alocação minimax para  $k_1 = 2$ , os valores

de  $x_i$  são  $\pm 1$  e 0 e assim,  $V_1 = 1 + (\frac{3}{2})X_0^2$ , resultando que

$$E(2,1) = \frac{1 + 1^2}{1 + \frac{3}{2} \cdot 1^2} = \frac{2}{\frac{5}{2}} = \frac{4}{5}$$

Sob alocação minimax para  $k_1 = 3$  os valores de  $x_i$  são  $\pm 1$  e  $\pm \sqrt{\frac{1}{5}}$  e então,  $V_1 = 1 + (\frac{4}{\frac{12}{5}})X_0^2$ , resultando que

$$E(3,1) = \frac{1 + 1^2}{1 + \frac{5}{3} \cdot 1^2} = \frac{2}{\frac{8}{3}} = \frac{6}{8}$$

e para  $k_1$  em geral pode ser mostrado (do mesmo modo que o esboçado na secção 4.5.2. abaixo) que

$$E(k_1,1) = \frac{2k_1}{3k_1 - 1}$$

#### 4.5.2. O caso $k_0 = 2$

Neste caso a variância em um ponto  $X_0$ , para uma alocação simétrica em torno de  $x = 0$ , é

$$V_2 = \frac{\Sigma x^2 \Sigma x^4 + [n \Sigma x^4 - 3(\Sigma x^2)^2] X_0^2 + [n \Sigma x^2] X_0^4}{\Sigma x^2 \Sigma x^4 - \frac{1}{n} (\Sigma x^2)^3}$$

Sob alocação minimax para  $k_0 = 2$  verificamos que

$V_2 = (6 - 9X_0^2 + 9X_0^4)/2$ . Sob alocação minimax para  $k_1 = 3$  encontramos que  $V_2 = (39 - 70X_0^2 + 75X_0^4)/12$  resultando que

$$E(3;2) = \frac{9}{11}$$

Do mesmo modo encontramos que

$$E(4;2) = \frac{12}{16}$$

e para  $k_1$  em geral, mostraremos que

$$E(k_1; 2) = \frac{3k_1}{5k_1 - 4}$$

Desde que, alocação minimax para  $k_1$  sempre distribui as  $n$  observações igualmente sobre os  $k_1 + 1$  valores  $x$ , consideraremos então as somas na expressão para  $V_2$  estendidas sobre precisamente esses  $k_1 + 1$  valores (assim, colocamos  $n = k_1 + 1$ ). Usando a expressão para  $V_2$  desta forma em conjunção com a definição anterior de  $E(k_1, 2)$  resulta

$$E(k_1; 2) = \frac{3[(k_1 + 1)\Sigma x^2 \Sigma x^4 - (\Sigma x^2)^3]}{(k_1 + 1)[\Sigma x^2 \Sigma x^4 - 3(\Sigma x^2)^2 + (k_1 + 1)(\Sigma x^4 + \Sigma x^2)]}$$

A equivalência desta expressão com a anterior é demonstrada do seguinte modo:

Dois dos pontos de observação minimax são sempre  $\pm 1$ ; Guest (1958) indicou que os  $k_1 - 1$  pontos restantes são dados pelas raízes da primeira derivada do polinômio de Legendre de grau  $k$ ; isto é, pelas raízes da equação (para simplicidade usamos  $k$  ao invés de  $k_1$ ).

$$\frac{d^{k+1}}{dx^{k+1}}(x^2 - 1)^k = 0$$

Multiplicando ambos os membros da expressão acima por  $\frac{(k-1)!}{(2k)!}$  e denotando por  $C_i$  o coeficiente do termo de grau  $i$ , então,  $C_{k-1} = 1$  e  $C_{k-2} = C_{k-4} = 0$ . Um resultado da teoria das funções simétricas, sobre as raízes de uma equação, estabelece que a so

ma das quartas potências das raízes de  $\frac{d^{k+1}}{dx^{k+1}}(x^2-1)^k$  é igual a  $2C_{k-3}^2 - 4C_{k-5}$ , e portanto, na expressão de  $E(k_1; 2)$  podemos escrever

$$\begin{aligned} \Sigma x^4 &= 2 + 2C_{k-3}^2 - 4C_{k-5} \\ &= 2 + \frac{(k-1)^2(k-2)^2}{2(2k-1)^2} - \frac{(k-1)(k-2)(k-3)(k-4)}{2(2k-1)(2k-3)} \\ &= \frac{k(k+1)(3k^2-5k+1)}{(2k-1)^2(2k-3)} \end{aligned}$$

É analogamente verdadeiro que  $\Sigma x^2 = k(k+1)/(2k-1)$ , e usando estes resultados  $E(k_1, 2)$  pode ser simplificado para o resultado desejado, ou seja,  $E(k_1, 2) = \frac{3k_1}{5k_1 - 4}$

#### 4.5.3. Comentários

Modelos de regressão polinomial são tipicamente usados como aproximações simples para modelos desconhecidos e para manter este espírito de simplicidade geralmente não nos preocupamos com valores alternativos de  $k$  excedendo  $k_0$ . O valor alternativo  $k_0+1$  frequentemente é de interesse; assim como também o é o valor  $k_0+2$ , particularmente quando o sinal da primeira derivada de  $Y$  com respeito a  $x$  é bem estabelecido em ambos os extremos do eixo  $x$ . Valores alternativos maiores são de interesse apenas em circunstâncias não usuais.

Ao escolhermos uma alocação com um ou ambos os valores alternativos  $k_0+1$  e  $k_0+2$  em mente, interesses tais como

custos de escolhas incorretas e informações "a priori" intuitiva para os vários valores de  $k$ , obviamente relacionam-se com a decisão. Os resultados da secção (4.5.2.) indicam entretanto, que dada qualquer chance apreciável que  $k = k_0 + 1$  ( $k = k_0 + 2$ ) deveremos alocar as observações em conformidade, desde que: a perda na precisão se  $k = k_0$  será pequena, e o custo esperado desta perda provavelmente será pequeno comparado com o custo esperado de examinar os valores alternativos  $k_0 + 1$  ( $k_0 + 2$ ). Tanto para  $k_0 = 1$  como para  $k_0 = 2$  foi mostrado que para a alocação  $k_0 + 2$  a eficiência mínima  $E(k_0 + 2, k_0)$  é igual a 0,75 (equivalentemente, quando de fato  $k = k_0$  o aumento na variância minimax do valor previsto pelo modelo sobre  $[-1, 1]$  é apenas 33% ou 15% em termos de desvio padrão). Este é um pequeno preço a ser pago pela oportunidade de testar e talvez melhorar a adequação do modelo de regressão.

Em particular, quando  $k_0 = 1$  ou 2, temos uma medida da perda na precisão, qualquer que seja o valor alternativo  $k_1$ , isto é,  $E(k_1, 1) > 0,666$  e  $E(k_1, 2) \geq 0,6$ . Equivalentemente, se alocação minimax para  $k_1$  é usada quando de fato  $k = 1$  (2) o aumento na variância minimax do valor previsto pelo modelo sobre  $[-1, 1]$  é quando muito 50% (67%), ou 22% (29%) em termos de desvios padrões.

Com valores particulares de  $k_0$  e  $k_1$  em mente é possível considerar alocações alternativas para as duas alocações. Por exemplo, se  $k_0 = 1$ ,  $k_1 = 2$  e  $n = 60$ , podemos considerar as alocações da Tabela 4.1. e procedermos uma escolha ótima basea

da sobre considerações de custos de erros e informações "a priori" para  $k_0$  e  $k_1$ .

Tabela 4.1. Alocações Alternativas para as duas Alocações Minimax:  $k_0=1$ ,  $k_1=2$ ,  $n=60$

Número de observações em			Variância máxima do valor previsto sobre $[-1,1]$ , quando:	
$x=-1$ ,	0,	+1	$k = 1$	$k = 2$
30	0	30	2	$\infty$
28	4	28	2,07	15
26	8	26	2,15	7,5
24	12	24	2,25	5
22	16	22	2,36	3,75
20	20	20	2,5	3

Na Tabela 4.1., notamos que quando  $k=2$ , a variância do valor previsto aumenta acentuadamente quando nos afastamos da alocação minimax "20 - 20 - 20"; um aumento que será tolerável apenas se não dispomos de informação "a priori" para  $k=2$ , e o custo de examinar  $k=2$  é relativamente pequeno.

## CAPÍTULO 5

### EXEMPLOS NUMÉRICOS

#### 5.1. Introdução

Neste capítulo nós apresentamos exemplos numéricos para ilustrar a aplicação de polinômios ortogonais com pontos de observação igualmente e desigualmente espaçados e, também a aplicação dos vários critérios de alocação ótima. Esses exemplos são encontrados nos livros:

Williams, E.J. - "Regression Analysis" - 1959,

Mendenhall, W. e Scheaffer, R.L. - "Mathematical Statistics with Applications" - 1973, e no artigo de

Hoel, P.G. - "On Testing for the Degree of a Polynomial" - 1968.

#### Exemplo 5.1.

Neste exemplo um estudo da relação entre a resistência à ruptura ( $\text{gr/cm}^2$ ) e a massa específica ( $\frac{\text{gr}}{\text{m}^2}$ ) para certa espécie de papel é apresentado. O estudo foi realizado através de modelos polinomiais usando-se os dados da Tabela 5.1., onde são encontrados os valores médios da resistência à ruptura,

baseados em 3 observações para cada uma das nove diferentes massas específicas do papel, variando de 10 a 90  $\frac{\text{gr}}{\text{m}^2}$ . Embora a resistência à ruptura aumenta com a massa específica, os dados indicam que a mesma não aumenta de forma linear e assim um ajuste com uma curva polinomial de grau maior do que 1 é o ajuste mostrado. Os polinômios ortogonais para  $n=9$  (Fisher e Yates, 1957) foram usados e estão na Tabela 5.1.

Tabela 5.1

Resistência à ruptura,  $Y(\frac{gI}{cm^2})$ , e massa específica,  $x(\frac{gI}{m^2})$  para certa espécie de papel

Y	360	1267	2146	3038	3962	5009	6114	6906	7519	$\Sigma(\phi^2)$	$\Sigma(Y\phi)$	$\beta$	S.Q.	Ajuste da Regressão		
	x	10	20	30	40	50	60	70	80					90	G.L.	S.Q.
$\phi_1$	-4	-3	-2	-1	0	1	2	3	4	60	55.460	924,3	51.263.527	8	51.386.918	
$\phi_2$	28	7	-8	-17	-20	-17	-8	7	28	2772	-4.296	-1,5	6.658	7	123.391	17.627**
$\phi_3$	14	-7	-13	-9	0	9	13	7	-14	990	8.570	8,7	74.187	6	116.733	19.456**
$\phi_4$	14	-21	-11	9	18	9	-11	-21	14	2002	-8.448	-4,2	35.649	5	42.546	8.509(n)
$\phi_5$	4	-11	4	9	0	-9	-4	11	-4	468	-218	-0,5	102	4	6.897	1.724(n)
Ajuste Cúbico											(n): não significativo					
$\hat{Y}$	418	1191	2086	3059	4066	5064	6009	6859	7569	**: significativo ao nível de 1%						

O ajuste da regressão é testado contra a variância estimada das médias dos 3 valores em cada ponto  $x$ . O valor desta variância foi 3804 com 16 graus de liberdade. Os cálculos são dados nas colunas da direita da Tabela 5.1. Vemos que o ajuste para uma curva quadrática é altamente significativa, mas é não significativa para uma cúbica, de modo que uma curva cúbica estabelece um ajuste satisfatório para os dados.

A fim de usar a função de regressão, pode ser desejável convertê-la para um polinômio em  $x$  explicitamente. Fisher e Yates (1957), na introdução de suas tabelas mostram como isto pode ser feito. Entretanto, para muitos propósitos, é igualmente conveniente trabalhar com polinômios ortogonais. Assim, neste exemplo, a equação de regressão cúbica em termos dos  $\phi_i$ , é dada por

$$\hat{Y} = 4035,7 + 924,3\phi_1 - 1,5\phi_2 + 8,7\phi_3$$

Uma vez que os valores dos  $\phi_i$  são dados, os valores estimados correspondentes aos pontos experimentais são facilmente encontrados, por exemplo, para  $x=40$ , o valor estimado é:

$4035,7 + 1 \times 924,3 - 17 \times (-1,5) - 9 \times 8,7 = 3059$  e seu desvio padrão é 34,6.

As estimativas da curva cúbica para os outros pontos são dadas na Tabela 5.1.

Dessa forma podemos adotar o modelo polinomial de grau 3 como um modelo aceitável para explicar a relação entre a resistência à ruptura e a massa específica para esta espécie de

papel.

Exemplo 5.2.

Neste exemplo é mostrado, para um caso particular, como um modelo polinomial é ajustado para valores de  $x$  desigualmente espaçados, a saber:  $x_0=1$ ,  $x_1=2$ ,  $x_2=5$ ,  $x_3=8$ .

Inicialmente, consideramos os polinômios

$$\phi_0(x) = 1$$

$$\phi_1(x) = a_0 + a_1 x$$

$$\phi_2(x) = b_0 + b_1 x + b_2 x^2$$

$$\phi_3(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3$$

A fim de determinar os valores desses coeficientes, devemos resolver os sistemas

$$\sum \phi_0(x_i) \phi_1(x_i) = 0$$

$$\begin{cases} \sum \phi_0(x_i) \phi_2(x_i) = 0 \\ \sum \phi_1(x_i) \phi_2(x_i) = 0 \end{cases}$$

$$\begin{cases} \sum \phi_0(x_i) \phi_3(x_i) = 0 \\ \sum \phi_1(x_i) \phi_3(x_i) = 0 \\ \sum \phi_2(x_i) \phi_3(x_i) = 0 \end{cases}$$

para  $x_i = 1, 2, 5, 8$ .

Obtemos como um possível resultado os polinômios

$$\phi_1(x) = x - 4$$

$$\phi_2(x) = (2x^2 - 18x + 25) / 3$$

$$\phi_3(x) = (55x^3 - 755x^2 + 2846x - 2524) / 42$$

cujos valores e somas de quadrados nos pontos 1, 2, 5 e 8 são:

	S.Q.				
Para $\phi_1(x)$ :	-3	-2	1	4	30
Para $\phi_2(x)$ :	3	-1	-5	3	44
Para $\phi_3(x)$ :	-9	14	-7	2	330

Para os correspondentes valores de  $Y$ :  $Y_1, Y_2, Y_5, Y_8$  as somas de quadrado atribuída a cada aumento do grau no modelo polinomial são apresentadas na Tabela 5.2.

Tabela 5.2.

Modelo Polinomial	S.Q.
Linear	$(-3Y_1 - 2Y_2 + Y_5 + 4Y_8)^2 / 30$
Quadrático	$(3Y_1 - Y_2 - 5Y_5 + 3Y_8)^2 / 44$
Cúbico	$(-9Y_1 + 14Y_2 - 7Y_5 + 2Y_8)^2 / 330$

Um programa de computador para obter polinômios ortogonais para valores de  $x$  desigualmente espaçados e não balanceados foi dado por Emerson (1965).

### Exemplo 5.3.

Neste exemplo é apresentado um caso particular de alocação ótima de pontos pelo critério que consiste em minimizar a variância da estimativa do coeficiente do termo de maior grau do modelo polinomial.

Suponhamos que o problema é ajustar uma reta a um conjunto de  $n$  pontos dados, onde  $n$  é um inteiro par, e que selecionaremos os  $n$  valores de  $x$  no intervalo  $-a \leq x \leq a$ ,  $a$  real posi

tivo, de modo que esta alocação minimize  $\text{Var}(\hat{\beta}_1)$ .

Considerando o modelo  $Y = \beta_0 + \beta_1 x$ , temos pelo critério 3.2.2 que deveremos atribuir os pesos:

$$p_0 = p_1 = \frac{1}{2k} = \frac{1}{2}$$

aos pontos

$$x_0 = -\cos\left(\frac{0\pi}{1}\right) = -\cos 0 = -1 \quad \text{e,}$$

$$x_1 = -\cos\left(\frac{1\pi}{1}\right) = -\cos \pi = 1$$

Portanto, deveremos selecionar

$$n_0 = \frac{n}{2} \text{ observações em } x = -1 \text{ e,}$$

$$n_1 = \frac{n}{2} \text{ observações em } x = 1$$

Usando a transformação linear  $x' = ax$ , teremos  $\frac{n}{2}$  observações em  $x'_0 = -a$  e  $\frac{n}{2}$  em  $x'_1 = a$ .

Exemplo 5.4.

Neste exemplo é apresentado para o modelo linear  $Y = \beta_0 + \beta_1 x$ , uma aplicação do critério D-ótimo, isto é, supondo que selecionaremos  $n$  valores de  $x$  no intervalo  $[-1,1]$ , desejamos alocar os mesmos de modo a minimizar o determinante da matriz de covariância de  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)'$ .

Pelo critério 3.2.1, temos que o planejamento D-ótimo para o modelo  $Y = \beta_0 + \beta_1 x$ , é o que maximiza  $|X'S^{-1}X|$ , onde

$$X = \begin{bmatrix} 1 & x_0 \\ 1 & x_1 \end{bmatrix} \quad \text{e} \quad S = \begin{bmatrix} \frac{1}{n_0} & 0 \\ 0 & \frac{1}{n_1} \end{bmatrix}$$

Em vista dos resultados apresentados no Capítulo 3, a solução do mesmo é dada pelos pesos  $p_0 = p_1 = \frac{1}{k+1} = \frac{1}{2}$  nos pontos  $x_0$  e  $x_1$  respectivamente. Os pontos  $x_0$  e  $x_1$  são os zeros de  $(1-x^2)$ .  $P_1'(x) = (1-x^2)x' = (1-x^2)$ , isto é,  $x_0 = -1$  e  $x_1 = 1$ .

Assim deveremos fazer:

$$n_0 = \frac{n}{2} \text{ observações em } x_0 = -1 \text{ e}$$

$$n_1 = \frac{n}{2} \text{ observações em } x_1 = 1$$

Vemos, portanto, que para o modelo do primeiro grau  $Y = \beta_0 + \beta_1 x$ , o planejamento para fazer inferências sobre  $\beta_1$  (minimizar  $\text{Var}(\hat{\beta}_1)$ ) e o planejamento D-ótimo (maximizar  $|X'S^{-1}X|$ ) coincidem.

Concretamente, suponhamos que desejamos estudar o efeito de "simulantis digitalis" (digitalina) sobre a pressão sanguínea de ratos, sendo a variação das doses de  $x=2$  a  $x=5$  unidades. A resposta é esperada ser linear e seis ratos são utilizados para o experimento, cada rato podendo receber apenas uma dose. Se o critério para alocar as doses de "digitalis" a serem usadas no experimento, e o número de ratos a serem designados em cada dose é o de maximizar a quantidade de informação no experimento (maximizar  $|X'S^{-1}X|$ ) isto implica que o planejamento D-ótimo para o modelo  $Y = \beta_0 + \beta_1 x'$ , onde  $-1 \leq x' \leq 1$  é dado por:

$$p_0 = p_1 = \frac{1}{k+1} = \frac{1}{2},$$

nos pontos  $x'_0 = -1$  e  $x'_1 = 1$ , isto é

$$n'_0 = 6 \cdot \frac{1}{2} = 3 \text{ ratos em } x'_0 = -1 \text{ e}$$

$$n'_1 = 6 \cdot \frac{1}{2} = 3 \text{ ratos em } x'_1 = 1$$

E se a transformação linear  $x' = \frac{2x-7}{3}$  é usada, pois  $2 \leq x \leq 5$ , teremos,  $n_0 = 3$  ratos em  $x_0 = 2$  e  $n_1 = 3$  ratos em  $x_1 = 5$ .

Exemplo 5.5.

Neste exemplo os critérios de minimizar  $\text{Var}(\hat{\beta}_k)$  e D-ótimo são aplicados para modelos polinomiais do 2º grau.

Suponhamos que desejamos ajustar uma curva do 2º grau a um conjunto de  $n$  pontos dados, onde  $n$  é um inteiro par, e que selecionaremos os  $n$  valores de  $x$  no intervalo  $-a \leq x \leq a$ ,  $a$  real positivo.

Considerando o modelo  $Y = \beta_0 + \beta_1 x + \beta_2 x^2$ , temos pelo critério 3.2.2, isto é, minimizar  $\text{Var}(\hat{\beta}_2)$  que:

$$p_0 = p_2 = \frac{1}{2k} = \frac{1}{4} \quad \text{e} \quad p_1 = \frac{1}{k} = \frac{1}{2}$$

nos pontos

$$x_0 = -\cos\left(\frac{0\pi}{2}\right) = -\cos 0 = -1$$

$$x_1 = -\cos\left(\frac{1\pi}{2}\right) = -\cos \frac{\pi}{2} = 0$$

$$x_2 = -\cos\left(\frac{2\pi}{2}\right) = -\cos \pi = -(-1) = 1$$

Portanto deveremos selecionar

$$n_0 = n_2 = \frac{n}{4} \text{ pontos em } x_0 = -1 \text{ e } x_2 = 1 \text{ e}$$

$$n_1 = \frac{n}{2} \text{ pontos em } x_1 = 0$$

Usando a transformação linear  $x' = ax$ , teremos

$$\frac{n}{4} \text{ pontos em } x'_0 = -a \text{ e } x'_2 = a \text{ e}$$

$$\frac{n}{2} \text{ pontos em } x'_1 = 0$$

Devemos observar, no entanto, que a menos que  $n$  seja um múltiplo de 4, não teremos valores inteiros para  $n_0$  e  $n_2$  obtendo então um resultado aproximado para o planejamento ótimo.

No caso de selecionarmos os valores de  $x$  e as proporções de observações nesses pontos, para  $-1 \leq x \leq 1$ , de modo a obter um planejamento D-ótimo, isto é, maximizar  $|X'S^{-1}X|$ , onde

$$X = \begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{bmatrix} \text{ e } S = \begin{bmatrix} \frac{1}{n_0} & 0 & 0 \\ 0 & \frac{1}{n_1} & 0 \\ 0 & 0 & \frac{1}{n_2} \end{bmatrix}$$

teremos que os pesos são:

$$p_0 = p_1 = p_2 = \frac{1}{k+1} = \frac{1}{2+1} = \frac{1}{3}$$

nos pontos  $x_0$ ,  $x_1$  e  $x_2$  que são os zeros de  $(1-x^2)P_2'(x) = (1-x^2)(\frac{3}{2}x^2 - \frac{1}{2})' = (1-x^2)(3x)$ , isto é,  $x_0 = -1$ ,  $x_1 = 0$  e  $x_2 = 1$ .

Portanto deveremos tomar  $n_0 = n_1 = n_2 = \frac{n}{3}$  observações em cada um dos pontos  $x_0$ ,  $x_1$  e  $x_2$ , ou seja, deveremos ter  $n$  como sendo múltiplo de 3 para que tenhamos valores inteiros de  $n_0$ ,  $n_1$  e  $n_2$ .

Observação: Os critérios para este caso conduzem a planejamentos distintos.

Exemplo 5.6.

Para ilustrar os métodos apresentados no Capítulo 4, suponhamos que um polinômio do quarto grau tenha sido escolhido e que o mesmo passa pelos cinco pontos  $(-1, 2)$ ,  $(-\frac{1}{2}, 0)$ ,

$(0, \frac{1}{2})$ ,  $(\frac{1}{2}, 0)$  e  $(1, 2)$ .

Os polinômios  $\phi_j(x)$ ,  $j=0,1,\dots,4$ , que são ortonormais com relação a esses pontos igualmente espaçados  $(-1, -\frac{1}{2}, 0, \frac{1}{2}, 1)$  são dados por:

$$\begin{aligned} \phi_0(x) &= \sqrt{\frac{1}{5}}, & \phi_1(x) &= \sqrt{\frac{2}{5}} x, & \phi_2(x) &= \sqrt{\frac{8}{7}} (x^2 - \frac{1}{2}), \\ \phi_3(x) &= \sqrt{\frac{40}{9}} (x^3 - \frac{17}{20} x), & \phi_4(x) &= \sqrt{\frac{280}{9}} (x^4 - \frac{31}{28} x^2 + \frac{9}{70}) \end{aligned}$$

Os coeficientes da regressão polinomial são encontrados resolvendo as cinco equações obtidas impondo que o polinômio passe através dos cinco pontos dados. Essas equações dão a solução

$$E[Y(x)] = \frac{9}{10} + 2(x^2 - \frac{1}{2}) + \frac{14}{3}(x^4 - \frac{31}{28} x^2 + \frac{9}{70})$$

resultando que

$$\gamma_0 = \frac{9}{10} \sqrt{5}, \quad \gamma_1 = 0, \quad \gamma_2 = \frac{\sqrt{14}}{2}, \quad \gamma_3 = 0, \quad \gamma_4 = \frac{\sqrt{70}}{10}$$

A transformação  $Z = \Phi'Y$  necessária para reduzir o problema para a forma canônica é obtida encontrando os valores de  $\phi_j(x)$  para os pontos  $-1, -\frac{1}{2}, 0, \frac{1}{2}, 1$ . Temos então que

$$\Phi' = \begin{bmatrix} \sqrt{\frac{1}{5}} & (1 & 1 & 1 & 1 & 1) \\ \sqrt{\frac{2}{5}} & (-1 & -\frac{1}{2} & 0 & \frac{1}{2} & 1) \\ \sqrt{\frac{8}{7}} & (\frac{1}{2} & -\frac{1}{4} & -\frac{1}{2} & -\frac{1}{4} & \frac{1}{2}) \\ \sqrt{\frac{40}{9}} & (-\frac{3}{20} & \frac{3}{10} & 0 & -\frac{3}{10} & \frac{3}{20}) \\ \sqrt{\frac{280}{9}} & (\frac{3}{140} & -\frac{3}{35} & \frac{9}{70} & -\frac{3}{35} & \frac{3}{140}) \end{bmatrix}$$

Assim, as variáveis canônicas  $Z = \Phi'Y$  são dadas por

$$\begin{aligned} Z_0 &= \sqrt{\frac{1}{5}} (Y_0 + Y_1 + Y_2 + Y_3 + Y_4) \\ Z_1 &= \sqrt{\frac{2}{5}} (-Y_0 - \frac{1}{2}Y_1 + \frac{1}{2}Y_3 + Y_4) \\ Z_2 &= \sqrt{\frac{2}{7}} (Y_0 - \frac{1}{2}Y_1 - Y_2 - \frac{1}{2}Y_3 + Y_4) \\ Z_3 &= \frac{\sqrt{10}}{10} (-Y_0 + 2Y_1 - 2Y_3 + Y_4) \\ Z_4 &= \frac{\sqrt{70}}{70} (Y_0 - 4Y_1 + 6Y_2 - 4Y_3 + Y_4) \end{aligned} \tag{5.5.1}$$

O valor de  $\sigma$  foi tomado como sendo  $\frac{1}{2}$ . Então um conjunto de valores normais aleatórios foi tomado de uma tabela de tais números para  $\mu = 0$  e  $\sigma = 1$ . Esses números foram divididos por 2 para tornar  $\sigma = \frac{1}{2}$ , e então cada conjunto de cinco números foi adicionado às ordenadas de regressão 2, 0,  $\frac{1}{2}$ , 0, 2 para produzir um conjunto de valores aleatórios de  $Y_0, Y_1, Y_2, Y_3, Y_4$ . Esses produziram um conjunto de valores das variáveis canônicas  $Z_0, Z_1, Z_2, Z_3, Z_4$  dadas por (5.5.1). A amostragem continuou até que vinte de tais conjuntos de valores tivessem sido obtidos; entretanto, teoricamente a amostragem ocorreria com um conjunto de cada vez para o teste sequencial. Os resultados desta amostragem produziram os seguintes valores de  $Z$ .

$\lambda_0=0,02$ . O contraste entre o valor real 2,8 e o valor do limite inferior conservativo 0,02 para a quantidade  $(\gamma_3^2 + \gamma_4^2)/\sigma^2$  sugere que  $\lambda_0$  deve ser escolhido consideravelmente maior do que 0,02; por outro lado, uma amostra sequencial muito grande será requerida para atingir uma decisão. Portanto para facilidade de ilustração o valor de  $\lambda_0$  foi tomado como sendo 0,10. Então  $e^{-\lambda_0/2} = 1,05$ . Desde que  $k=4$  e  $s=3$ , então a fórmula (4.2.7) reduz-se para

$$\frac{P_{1n}}{P_{0n}} = (1,05)^n F\left(\frac{5}{2}(n-1), 1, \frac{\delta^2}{2\lambda_0}\right)$$

onde

$$\frac{\delta^2}{2\lambda_0} = 0,05n \frac{n \sum_{i=3}^4 \bar{Z}_i^2}{\sum_{\alpha=1}^n \sum_{i=0}^4 Z_{i\alpha}^2 - n \sum_{i=0}^2 \bar{Z}_i^2}$$

A Tabela 5.3. com os seguintes valores foi obtida.

Tabela 5.3.

n	6	7	8	9	10	11
F	4,4	7,0	6,2	10,9	14,8	27,0
$\frac{P_{1n}}{P_{0n}}$	3,2	4,9	4,1	6,9	8,9	15,0

Desde que os tamanhos dos dois tipos de erros foram escolhidos como sendo 0,10, os limites sequenciais para  $\frac{P_{1n}}{P_{0n}}$  são  $\frac{1}{9}$  e 9. O valor do limite superior foi ultrapassado em  $n=11$

$Z_0$	$Z_1$	$Z_2$	$Z_3$	$Z_4$
2,34	0,40	1,75	-0,62	-0,06
1,83	0,28	2,22	-0,10	1,16
1,97	0,48	2,04	0,23	1,70
2,23	0,28	2,03	-0,59	0,03
1,61	-0,29	2,09	0,20	0,83
3,34	0,14	0,98	0,42	0,70
1,86	0,09	2,39	0,08	0,84
1,84	-1,45	2,01	-0,34	-0,02
2,84	0,52	2,51	0,33	1,87
1,74	1,16	2,14	0,21	0,67
1,96	0,49	1,42	-0,07	1,12
1,67	-0,04	2,20	0,42	1,13
2,80	0,21	1,56	0,34	1,13
2,41	0,31	1,70	1,01	0,87
1,92	-0,20	0,89	-0,03	1,15
1,89	0,23	1,28	-0,35	0,48
2,32	0,68	1,39	-0,56	0,76
1,70	0,56	2,80	0,70	1,65
1,89	0,28	1,07	-0,43	0,53
2,18	0,32	1,66	-0,40	1,34

Inicialmente, consideremos o teste sequencial para testar a hipótese  $H_0: \gamma_3 = \gamma_4 = 0$  contra  $H_1: (\gamma_3^2 + \gamma_4^2)/\sigma^2 > \lambda_0$ . Temos que  $(\gamma_3^2 + \gamma_4^2)/\sigma^2 = 2,8$ . Se  $x = 1$  e  $\cos \theta = 1$  são usados na fórmula (4.2.3) para determinar  $\lambda_0$ , encontramos que  $\lambda_0 = 0,122/m$ . Supondo que seis réplicas são suficientes, isto acarretará

portanto, a amostragem está terminada e  $H_1$  foi aceita neste estágio.

Consideremos agora a solução para este problema por meio do procedimento com duas amostras. O valor de  $n_0$  foi escolhido como sendo 6 e  $c$  como 0,01. Uma vez que  $\sigma = \frac{1}{2}$  é conhecido "a priori" é possível escolher  $c$  de modo que  $n$  muito provavelmente seria menor do que 20. Isto foi o que motivou a escolha de  $c = 0,01$ . Usando esses mesmos valores de  $Z$ , cálculos baseados sobre os seis primeiros conjuntos de valores amostrais resultaram

$$S^2 = \frac{1}{(5)(5)} \{60,1 - 55,8\} = 0,172$$

Assim,

$$n = \text{máx} \left\{ \left[ \frac{0,172}{0,01} \right] + 1; 7 \right\} = 18,$$

Seja agora  $a_1 = \dots = a_6 = a$  e escolhamos  $a_7 = \dots = a_{18} = b$ . Então  $\sum_{j=1}^{18} a_j = 1$  e  $\sum_{j=1}^{18} a_j^2 = \frac{c}{S^2}$  serão verificadas se  $a$  e  $b$  satisfazem a equação  $6a + 12b = 1$  e  $6a^2 + 12b^2 = \frac{0,01}{0,172} = 0,058$ . A solução dessas equações é  $a = 0,041$  e  $b = 0,063$ . Como resultado a fórmula (4.3.1) torna-se

$$F' = \frac{\sum_{i=3}^4 (0,041 \sum_{j=1}^6 Z_{ij} + 0,063 \sum_{j=7}^{18} Z_{ij})^2}{(0,01)(5)(5)}$$

Da fórmula (4.3.2) e da tabela da variável  $F$  central será encontrado que  $P\{F_{2,25} > 2,55\} = 0,10$ , portanto, 2,55 é o valor crítico desejado de  $F'$  para fazer o teste quando o tama

nho do erro de tipo I é 0,10. Cálculos baseados na fórmula de  $F'$  acima produzem o valor  $F' = 3,4$ . Sendo  $F' > 2,55$  a hipótese  $H_0: \beta_3 = \beta_4 = 0$  é rejeitada.

Consideremos agora o teste modificado para determinar o grau apropriado quando  $H_0$  é rejeitada. Temos

$$u_k = u_4 = \sum_{j=1}^{18} a_j Z_{ij} = 0,918; \text{ portanto } u_4^2 = 0,84. \text{ Também}$$

$r^2 = \ell c(n_0 - 1)(k + 1) = 0,64$ . A primeira pergunta que faríamos é se  $u_k^2 < r^2$ . Desde que  $u_4^2 = 0,84 > 0,64$ , a resposta é não, e portanto paramos aqui e concluimos que o grau adequado é 4.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON, T.W. The choice of the degree of a polynomial regression as a multiple decision problem. *The Annals of Mathematical Statistics*, 33(1):255-265, 1962.
- BEATON, A.E. & TUKEY, J.W. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16(2):147-185, 1974.
- BOX, G.E.P. & DRAPER, N.R. A basis for the selection of a response surface design. *Journal of the American Statistical Association*, 54(287):622-654, 1959.
- BOX, G.E.P. & DRAPER, N.R. The choice of a second order rotatable design. *Biometrika*, 50(3/4):335-352, 1963.
- CLENSHAW, C.W. Curve fitting with a digital computer. *The Computer Journal*, 2(4):170-173, 1960.
- COOPER, B.E. A remark on algorithm AS 10: the use of orthogonal polynomials. *Applied Statistics*, 20(2):216, 1971.
- COOPER, B.E. The use of orthogonal polynomials: algorithm AS 10. *Applied Statistics*, 17(3):283-287, 1968.
- COOPER, B.E. The use of orthogonal polynomials with equal x-values: algorithm AS 42. *Applied Statistics*, 20(2):209-213, 1971.
- COTE, R.; MANSON, A.R.; HADER, R.J. Minimum bias approximation of a general regression model. *Journal of the American Statistical Association*, 68(343):633-638, 1973.

- EMERSON, P.L. A Fortran generator of polynomials over unequally spaced and weighted abscissas. *Educational and Psychological Measurement*, 25:867-871, 1965.
- EMERSON, P.L. Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics*, 24(3):695-701, 1968.
- EMERSON, P.L. Orthogonal polynomials for unequally weighted means. *Biometrics*, 21(1):226-230, 1965.
- FORSYTHE, G.E. Generation and use of orthogonal polynomials for data-fitting with a digital computer. *Journal of the Society for Industrial and Applied Mathematics*, 5(2):74-88, 1957.
- GARZA, A. de la. Spacing of information in polynomial regression. *The Annals of Mathematical Statistics*, 25(1):123-130, 1954.
- GUEST, P.G. The spacing of observations in polynomial regression. *The Annals of Mathematical Statistics*, 29(1):294-299, 1958.
- HAYES, D.G. A method of storing the orthogonal polynomials used for curve and surface fitting. *The Computer Journal*, 12(2):148-150, 1969.
- HAYES, J.G. Curve fitting by polynomials in one variable. In: HAYES, J.G., ed. *Numerical approximation to functions and data*. Oxford, University Press, 1970. p.43-64.
- HAYES, J.G., ed. *Numerical approximation to functions and data*. Oxford, University Press, 1970. 177p.

- HAYES, J.G. Numerical methods for curve and surface fitting. *Journal of the Institute of Mathematics and its Applications*, 10(2):144-152, 1972.
- HOEL, P.G. Efficiency problems in polynomial estimation. *The Annals of Mathematical Statistics*, 29(2):1134-1145, 1958.
- HOEL, P.G. On a sequential test for the general linear hypothesis. *The Annals of Mathematical Statistics*, 26:136-139, 1955.
- HOEL, P.G. On testing for the degree of a polynomial. *Technometrics*, 10(4):757-767, 1968.
- HOEL, P.G. & LEVINE, A. Optimal spacing and weighting in polynomial prediction. *The Annals of Mathematical Statistics*, 35(4):1553-1560, 1964.
- JOHN, R.C.St. & DRAPER, N.R. D-optimality for regression designs: a review. *Technometrics*, 17(1):15-23, 1975.
- KARSON, M.J.; MANSON, A.R.; HADER, R.J. Minimum bias estimation and experimental design for response surfaces. *Technometrics*, 11(3):461-475, 1969.
- KENNEDY, W.J. & BANCROFT, T.A. Model building for prediction in regression based upon repeated significance tests. *The Annals of Mathematical Statistics*, 42(4):1273-1284, 1971.
- KIEFER, J. Optimum experimental designs. *Journal of the Royal Statistical Society, Serie B*, 21(1):273-319, 1959.

- KIEFER, J. & WOLFOWITZ, J. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12(3):363-366, 1960.
- KIEFER, J. & WOLFOWITZ, J. On a theorem of Hoel and Levine on extrapolation designs. *The Annals of Mathematical Statistics*, 36(2):1627-1655, 1965.
- KIEFER, J. & WOLFOWITZ, J. Optimum designs in regression problems. *The Annals of Mathematical Statistics*, 30(2): 271-294, 1959.
- KRYLOV, V.I. *Approximate calculation of integrals*. New York, Macmillan, c1962. 357p. (ACM Monograph Series)
- KUSSMAUL, K. Protection against assuming the wrong degree in polynomial regression. *Technometrics*, 11(4):677-682, 1969.
- MURTY, V.N. Minimax designs. *Journal of the American Statistical Association*, 66(334):319-320, 1971.
- MURTY, V.N. & STUDDEN, W.J. Optimal designs for estimating the slope of a polynomial regression. *Journal of the American Statistical Association*, 67(340):869-873, 1972.
- PEARSON, K., ed. *Tables for statisticians and biometricians*. 3.ed. London, University College, 1930. v.1.
- ROBSON, D.S. A simple method for constructing orthogonal polynomials when the independent variable is unequally spaced. *Biometrics*, 15(2):187-191, 1959.

SEBER, G.A.F. *Linear regression analysis*. New York, John Wiley, c1977. 465p. (Wiley Series in Probability and Mathematical Statistics)

STEIN, C. A two-sample test for a linear hypothesis whose power is independent of the variance. *The Annals of Mathematical Statistics*, 16(3):243-258, 1945.

STIGLER, S.M. Optimal experimental design for polynomial regression. *Journal of the American Statistical Association*, 66(334):311-318, 1971.

WILLIAMS, E.J. *Regression analysis*. New York, John Wiley, c1959. 214p. (Wiley Publications in Statistics)