

Método dos Mínimos Quadrados e Distribuição Gaussiana

Esmerindo Bernardes^{1,*}

¹L.I.A. – LABORATÓRIO DE INSTRUMENTAÇÃO ALGÉBRICA
 Instituto de Física de São Carlos
 Universidade de São Paulo
 13560-970 São Carlos, SP, Brazil
 (Dated: 13 de Abril de 2015)

O método dos mínimos quadrados nos permite encontrar a melhor função que descreve um certo conjunto de informações experimentais. Este texto apresenta o método dos mínimos quadrados aplicado ao caso linear de uma forma que possa ser facilmente generalizado. Funções transcendentais são calculadas por séries de potência. Faremos aqui uma brevíssima introdução ao mecanismo desenvolvido por Taylor para determinarmos a série de potência de uma função transcendental. O método dos mínimos quadrados e as séries de Taylor são duas outras aplicações fundamentais de derivadas.

CONTENTS

I. Método dos Mínimos Quadrados	1
II. Valor médio e desvio médio quadrático	3
Referências	4

I. MÉTODO DOS MÍNIMOS QUADRADOS

No caso linear, as previsões que nossos modelos nos fornecem podem ser resumidas na forma

$$f(u) = c_0 + c_1 u, \quad (1)$$

onde c_0 e c_1 são constantes arbitrárias e u é a variável independente. A função $f(u)$ representa uma determinada resposta de um sistema, como a posição de um corpo deslizando sobre um plano inclinado ou em queda livre, ou ainda o quadrado do período de um pêndulo ideal (pequenas amplitudes). A Tabela I mostra quem é a função resposta $f(u)$, bem como a variável independente u , em dois casos típicos. As duas últimas colunas desta tabela mostra os valores esperados para as constantes arbitrárias c_0 e c_1 .

Tabela I. Função resposta $f(u)$ em dois modelos lineares: queda livre com aceleração g e o quadrado do período de um pêndulo ideal de comprimento l (pequenas amplitudes).

$x(t) = \frac{1}{2}gt^2$	$u = t^2$	$f = x$	$c_0 = 0$	$c_1 = \frac{1}{2}g$
$\tau^2(l) = \frac{4\pi^2}{g}l$	$u = l$	$f = \tau^2$	$c_0 = 0$	$c_1 = \frac{4\pi^2}{g}$

Suponha agora que já temos um conjunto com N medidas da variável independente u_i e da função resposta $f(u_i) = f_i$, $i \in \{1, 2, \dots, N\}$. Como podemos extrair os melhores valores para as constantes c_0 e c_1 presentes na função linear (1)? Gauss mostrou (em 1809) que a melhor forma de calcular es-

tas constantes é minimizando o erro quadrático dado por

$$E_2(c_0, c_1) = \sum_{i=1}^N \left[f_i - (c_0 + c_1 u_i) \right]^2. \quad (2)$$

Este procedimento é bastante razoável, pois para cada par de valores (f_i, u_i) esperamos que a diferença $f_i - (c_0 + c_1 u_i)$ seja a menor possível para valores apropriados das constantes c_0 e c_1 . Em princípio, numa busca manual, teríamos N valores (distintos) para estas constantes. Depois poderíamos escolher um valor único, digamos, usando uma média aritmética. O que Gauss mostrou (e não discutiremos os detalhes aqui) é que podemos determinar os melhores valores destas constantes em uma tacada única.

Sabemos que uma função de uma única variável apresenta uma derivada nula (coeficiente angular nulo) em seus pontos extremos (máximos e mínimo). Esta situação se mantém para funções com mais de uma variável com relação às derivadas em cada uma de suas variáveis independentes. Nossa função erro dada em (2) tem duas variáveis independentes: c_0 e c_1 . O mínimo desta função ocorre para os valores das constantes c_0 e c_1 são tais que o sistema (linear) de equações

$$\frac{dE_2}{dc_0} = -2 \sum_{i=1}^N \left[f_i - (c_0 + c_1 u_i) \right] = 0, \quad (3)$$

$$\frac{dE_2}{dc_1} = -2 \sum_{i=1}^N \left[f_i - (c_0 + c_1 u_i) \right] u_i = 0, \quad (4)$$

é satisfeito simultaneamente. Este sistema pode ser resolvido facilmente. Primeiro isolamos c_0 da Eq. (3),

$$\sum_{i=1}^N \left[f_i - (c_0 + c_1 u_i) \right] = \sum_{i=1}^N f_i - Nc_0 - c_1 \sum_{i=1}^N u_i = 0, \quad (5)$$

e, conseqüentemente,

$$c_0 = \frac{1}{N} \sum_{i=1}^N f_i - c_1 \frac{1}{N} \sum_{i=1}^N u_i = \bar{f} - c_1 \bar{u}, \quad (6)$$

(verifique) onde introduzimos as médias (aritméticas),

$$\bar{u} = \frac{1}{N} \sum_{i=1}^N u_i, \quad \bar{f} = \frac{1}{N} \sum_{i=1}^N f_i. \quad (7)$$

Vale lembrar que as quantidades (u_i, f_i) representam os valores de medidas realizadas no laboratório, portanto as médias (\bar{u}, \bar{f}) são números já conhecidos.

Devemos substituir o valor de c_0 encontrado em (6) na Eq. (4) para determinar c_1 . Antes, porém, vamos preparar melhor a Eq. (4) contendo as médias dadas em (7),

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N [f_i - (c_0 + c_1 u_i)] u_i = \\ \frac{1}{N} \sum_{i=1}^N f_i u_i - c_0 \frac{1}{N} \sum_{i=1}^N u_i - c_1 \frac{1}{N} \sum_{i=1}^N u_i^2 = \\ \overline{f u} - c_0 \bar{u} - c_1 \bar{u}^2 = 0, \end{aligned} \quad (8)$$

onde introduzimos as médias

$$\overline{f u} = \frac{1}{N} \sum_{i=1}^N f_i u_i, \quad \bar{u}^2 = \frac{1}{N} \sum_{i=1}^N u_i^2. \quad (9)$$

Agora sim podemos substituir na Eq. (8) a expressão de c_0 dada em (6) e isolar c_1 ,

$$c_1 = \frac{\overline{f u} - \bar{f} \bar{u}}{\bar{u}^2 - \bar{u}^2}, \quad c_0 = \frac{\bar{u}^2 \bar{f} - \bar{u} \overline{f u}}{\bar{u}^2 - \bar{u}^2}, \quad (10)$$

onde aproveitamos para substituir c_1 de volta na Eq. (6) para c_0 (verifique). Assim, após calcularmos as quatro médias dadas em (7) e (9), podemos obter diretamente de (10) os melhores valores para as constantes c_0 e c_1 que minimizam o erro quadrático (2). Um bom teste para as expressões dadas em (10) é a obtenção do conhecido coeficiente angular de uma reta determinado por dois pontos. Faça o Exercício 1.

Para completar este procedimento, precisamos determinar o erro Δc_k associado a cada um dos parâmetros c_k dados em (10). Estes erros são provenientes dos erros presentes nas medidas de u_i e f_i . Uma possível estimativa para os erros propagados para c_0 e c_1 é dada pelo erro quadrático

$$(\Delta c_k)^2 = \sum_{i=1}^N \left(\frac{dc_k}{df_i} \right)^2 (\Delta f_i)^2 + \sum_{i=1}^N \left(\frac{dc_k}{du_i} \right)^2 (\Delta u_i)^2. \quad (11)$$

Isto é uma definição que generaliza o uso de derivada para a propagação de erros em uma função com apenas uma variável. Em alguns casos pode ser introduzida uma outra forma, usando módulos ao invés de quadrados,

$$\Delta c_k = \sum_{i=1}^N \left| \frac{dc_k}{df_i} \right| \Delta f_i + \sum_{i=1}^N \left| \frac{dc_k}{du_i} \right| \Delta u_i. \quad (12)$$

Este é o erro absoluto. Trabalharemos com o erro quadrático.

Para simplificar um pouco, vamos considerar que os erros Δu_i cometidos nas medidas da variável independente u sejam desprezíveis, ou seja, muito menores que os erros Δf_i cometidos nas medidas da variável dependente f (e esta é uma situação comum em nossos experimentos). Além disto, vamos considerar também que as incertezas Δf_i sejam praticamente iguais, $\Delta f_i = \Delta f$. Isto é razoável se as medidas f_i estão

numa mesma escala, ou seja, os valores f_i possuem a mesma ordem de grandeza. Uma estimativa muito usada para o erro Δf no presente caso (função linear com dois parâmetros) é o erro quadrático médio

$$\Delta f = \sqrt{\frac{E_2}{N-2}}, \quad (13)$$

onde E_2 é o erro quadrático dado em (2), porém calculado com os valores de c_0 e c_1 dados por (10). Note que temos de usar um número grande de medidas, muito maior que a quantidade de parâmetros sendo determinados, para que o erro (13) seja pequeno. Com estas duas considerações, $\Delta u_i \approx 0$ e $\Delta f_i = \Delta f$, o erro quadrático em (11) pode ser reescrito na forma

$$\Delta c_k = \sqrt{\sum_{i=1}^N \left(\frac{dc_k}{df_i} \right)^2} \Delta f. \quad (14)$$

As derivadas em (14) podem ser calculadas facilmente. Primeiro note que os denominadores (idênticos) em (10) não dependem de f_i , portanto ficam de fora da derivada. Segundo, os valores médios aparecendo nos numeradores em (10) dependem de f_i conforme indicado nas definições (7) e (9). Sendo assim, precisaremos das derivadas

$$\frac{d\bar{f}}{df_i} = \frac{d}{df_i} \left(\frac{1}{N} \sum_{j=1}^N f_j \right) = \frac{1}{N} \sum_{j=1}^N \frac{df_j}{df_i} = \frac{1}{N} \sum_{j=1}^N \delta_{ij} = \frac{1}{N}, \quad (15)$$

e

$$\begin{aligned} \frac{d\overline{f u}}{df_i} &= \frac{d}{df_i} \left(\frac{1}{N} \sum_{j=1}^N f_j u_j \right) = \frac{1}{N} \sum_{j=1}^N u_j \frac{df_j}{df_i} \\ &= \frac{1}{N} \sum_{j=1}^N u_j \delta_{ij} = \frac{u_i}{N}. \end{aligned} \quad (16)$$

Note que introduzimos o símbolo

$$\delta_{ij} = \delta_{ji} = \begin{cases} 1 & \text{se } i = j, \\ 0 & \text{se } i \neq j, \end{cases} \quad (17)$$

denominado de “delta de Kronecker”, para representar o resultado da derivada

$$\frac{df_j}{df_i} = \delta_{ij}. \quad (18)$$

Desta forma, podemos calcular as derivadas das constantes c_0 e c_1 , em relação a f_i , a partir de (10) (verifique),

$$\frac{dc_0}{df_i} = \frac{1}{\bar{u}^2 - \bar{u}^2} \left(\bar{u}^2 \frac{d\bar{f}}{df_i} - \bar{u} \frac{d\overline{f u}}{df_i} \right) = \frac{\bar{u}^2 - \bar{u} u_i}{N(\bar{u}^2 - \bar{u}^2)}, \quad (19)$$

$$\frac{dc_1}{df_i} = \frac{1}{\bar{u}^2 - \bar{u}^2} \left(\frac{d\overline{f u}}{df_i} - \bar{u} \frac{d\bar{f}}{df_i} \right) = \frac{u_i - \bar{u}}{N(\bar{u}^2 - \bar{u}^2)}. \quad (20)$$

Portanto, o erro quadrático definido em (14) para $k = 1$ é

$$\begin{aligned} (\Delta c_1)^2 &= \sum_{i=1}^N \left(\frac{dc_1}{df_i} \right)^2 (\Delta f)^2 = \frac{(\Delta f)^2}{N^2(\bar{u}^2 - \bar{u}^2)^2} \sum_{i=1}^N (u_i - \bar{u})^2 = \\ &= \frac{(\Delta f)^2}{N(\bar{u}^2 - \bar{u}^2)^2} \frac{1}{N} \sum_{i=1}^N (u_i^2 - 2\bar{u}u_i + \bar{u}^2) = \\ &= \frac{(\Delta f)^2}{N(\bar{u}^2 - \bar{u}^2)^2} (\bar{u}^2 - \bar{u}^2) = \frac{(\Delta f)^2}{N(\bar{u}^2 - \bar{u}^2)}. \end{aligned} \quad (21)$$

Repetindo este procedimento para c_0 (faça o Exercício 2), as estimativas para os erros associados às constantes c_0 e c_1 são

$$\Delta c_0 = \sqrt{\frac{\bar{u}^2}{N(\bar{u}^2 - \bar{u}^2)}} \Delta f, \quad (22)$$

$$\Delta c_1 = \sqrt{\frac{1}{N(\bar{u}^2 - \bar{u}^2)}} \Delta f. \quad (23)$$

Tendo estabelecido as expressões (10) que precisamos para calcular os melhores valores das constantes c_0 e c_1 [que melhor ajustam os dados experimentais à curva linear (1)] e seus erros (22)–(23), bem como o erro (13) associado à função resposta, precisamos passar estas expressões para uma linguagem computacional (embora tudo isto e muito mais esteja implementado em diversos ambientes computacionais).

Tudo que foi discutido nesta seção faz parte de Estatística e há muitos livros textos que podem ser consultados para se ter uma visão mais ampla destas e outras ferramentas indispensáveis a engenheiros, químicos, físicos e cientistas em geral. Em particular, a Ref. 1 é um excelente texto e apresenta uma abordagem um pouco diferente do que foi apresentado aqui, ao fazer uso de matrizes para determinar os melhores valores das constantes procuradas. O uso de matrizes facilita a generalização do método dos mínimos quadrados a outros tipos de função resposta (além da linear).

Exercício 1

Suponha que os dois pontos $\{(u_1, f_1), (u_2, f_2)\}$ sejam conhecidos ($N = 2$). Use (10) e mostre explicitamente que o coeficiente angular c_1 da reta (1) é $c_1 = (f_2 - f_1)/(u_2 - u_1)$, como esperado da geometria plana.

Exercício 2

Repita o procedimento usado para obter Δc_1 dado na Eq. (23) para obter o erro Δc_0 dado na Eq. (22).

Exercício 3

Implemente o método dos mínimos quadrados (MMQ) em computação algébrica e use seus programas para estudar os casos em que haja necessidade do MMQ em suas atividades de laboratório.

II. VALOR MÉDIO E DESVIO MÉDIO QUADRÁTICO

Dado um conjunto de medidas $A = \{A_1, A_2, \dots, A_N\}$ contendo N informações, o valor médio \bar{A} destas medidas é defi-

nido por

$$\bar{A} = \frac{1}{N} \sum_{i=1}^N A_i. \quad (24)$$

O desvio médio quadrático σ_A destas medidas é definido por

$$\sigma_A = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\bar{A} - A_i)^2}. \quad (25)$$

Este desvio médio fornece uma medida quantitativa do grau de dispersão dos valores A_i em torno do valor médio.

Para muitas quantidades, estes desvios médios quadráticos obedecem uma distribuição gaussiana (normalizada) centrada no valor médio,

$$g(A) = \frac{1}{\sqrt{2\pi}\sigma_A} \exp\left\{-\frac{1}{2} \frac{(A - \bar{A})^2}{\sigma_A^2}\right\}. \quad (26)$$

A variável A , representando uma determinada quantidade, assume valores reais em torno do valor médio \bar{A} . A integral desta distribuição num intervalo $A \in [A_1, A_2]$ fornece a probabilidade que um determinado valor de A seja encontrado dentro deste intervalo. Por isso, a distribuição gaussiana é também conhecida como uma distribuição de probabilidades e é muito importante em Estatística, Engenharia, sistemas de controle, etc.

Como exemplo, considere um pêndulo balístico. Vamos admitir que medimos massas e o comprimento do pêndulo de formas acuradas. Também escolhemos $M' = 10M$, onde M' é a massa do pêndulo e M é a massa de um pequeno objeto que fica alojado no pêndulo após uma colisão. O comprimento do pêndulo é $l = 1.20$ m, o que nos dá uma frequência natural $\omega_0 = \sqrt{g/l} = 2.86$ rad/s. A Tabela II mostra os valores obtidos para os deslocamentos horizontais s do pêndulo e as respectivas velocidades calculadas via $V = (1/M, /M)\omega_0 s$, onde V é a velocidade do pequeno objeto antes da colisão, bem como os valores médios e respectivos desvios médios quadráticos. Novamente podemos ver que o valor $V = (3.77 \pm 0.11)$ m/s concorda com os outros dois valores.

Tabela II. Valores para as cinco distâncias (s) obtidas pelo pêndulo e respectivas velocidades horizontais ($V = (1/M, /M)\omega_0 s$) do objeto antes do impacto. σ representa o desvio médio quadrático.

	1	2	3	4	5	média	σ
s (m)	0.120	0.115	0.118	0.124	0.122	0.120	0.003
V (m/s)	3.77	3.62	3.71	3.90	3.84	3.77	0.11

A Figura 1 mostra que os valores para a velocidade V mostrados na Tabela II satisfaz a distribuição gaussiana dada em (26). Os pontos em vermelho (+) representam os cinco dados da Tabela II. Como exemplo, a probabilidade de encontrarmos um valor de V no intervalo $[3.66, 3.88]$ é de 68%. Esta

probabilidade aumenta rapidamente para intervalos mais largos. Eventualmente, esta probabilidade será 100% se tomarmos todo o intervalo horizontal mostrado na Figura 1. Note que a distribuição é máxima no valor médio $\bar{V} = 3.77$ m/s.

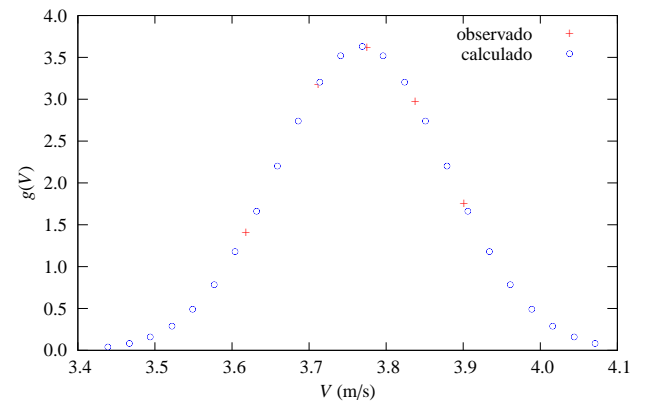


Figura 1. Distribuição de probabilidades gaussianas $g(V)$ [veja a Eq. (26)] para os valores da velocidade V conforme dados exibidos na Tabela II. O valor médio é $\bar{V} = 3.77$ m/s e o desvio médio quadrático é $\sigma_V = 0.11$ m/s.

* sousa@ifsc.usp.br

¹ O. Helene, *Método dos Mínimos Quadrados com Formalismo Matricial* (Ed. Livraria do IFUSP, 2006).