

LAPLACE e GAUSS introduzem no início do século XIX duas importantes ferramentas na Estatística, a utilização da distribuição normal para caracterizar os erros (e não apenas como uma aproximação da binomial) e a utilização da distribuição normal como uma distribuição aproximada da média em amostras de grande dimensão (Teorema Limite Central de âmbito geral).

Os erros de observação

Quando se pretende medir o valor de uma grandeza há, regra geral, erros associados a essa medição que podem ser **sistemáticos** (atuam sempre no mesmo sentido e, habitualmente, estão associados ao método de medição utilizado) ou **fortuitos** (que têm origem em causas aleatórias e, como tal, atuam em ambos os sentidos de forma não previsível).

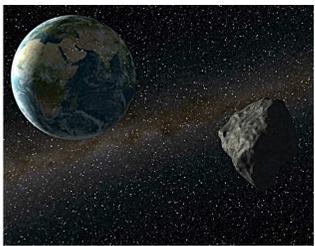
Seja y a quantidade que pretendemos medir (desconhecida) que é uma função de outras quantidades (conhecidas) x_j ($j = 1, \dots, m$), na qual há k incógnitas (parâmetros) β_i ($i = 1, \dots, k$). Contudo, não observamos o valor exato de y , mas cada uma das n observações y_i , $i = 1, \dots, n$ desta medida tem um erro fortuito associado que representaremos por $\varepsilon_i = y_i - y$. Deste modo, considerando que não há erros de ordem sistemática, podemos modelar as observações através de

$$y_i = f(x_{i1}, x_{i2}, \dots, x_{im}, \beta_1, \beta_2, \dots, \beta_k) + \varepsilon_i, \quad i = 1, \dots, n,$$

onde y_i representa a i -ésima observação de y que lhe tem associado o erro (fortuito) ε_i . Pretende-se analisar qual é a quantidade, definida em função das observações, que deve ser utilizada de forma a melhor *estimar* a quantidade y . A forma mais usual da função f é considerar um modelo linear,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im} + \varepsilon_i, \quad i = 1, \dots, n.$$

Este tipo de problema surgia frequentemente no século XVII em astronomia. Nestas aplicações pretendia-se determinar os valores dos parâmetros (coeficientes) com base num conjunto de observações. Deste modo, se tivermos o mesmo número de observações que coeficientes podemos determinar os valores desses coeficientes (teremos o mesmo número de equações e coeficientes e, como tal, supondo independência entre as diferentes equações, teremos um sistema possível e determinado).



Todavia, se o número de observações for superior ao número de coeficientes (mais equações que incógnitas) o sistema será impossível (inconsistente) devido aos erros que as observações contêm. Muitas vezes escolhia-se apenas algumas observações (em igual número que os coeficientes) e determinava-se os coeficientes resolvendo esse sistema.

A distribuição dos erros fortuitos

THOMAS SIMPSON (1710–1761) foi o primeiro, em 1755, a aplicar a teoria da probabilidade na análise dos erros (fortuitos) de observação, considerando que os erros deste tipo são igualmente prováveis de serem positivos e negativos, limitados e contínuos, tendo utilizado uma distribuição triangular para os caracterizar (concluído a forma da distribuição da média de erros com esta distribuição). Deduziu igualmente que a média de um conjunto de observações de determinada quantidade muito provavelmente terá um erro associado menor do que o erro de cada observação individual, uma vez que os erros compensam-se, contrariando uma ideia usual na época de que a média, por resultar da soma de muitas observações (e consequentemente de muitos erros) teria necessariamente associado um erro maior do que cada observação.

PIERRE LAPLACE (1749–1827) utilizou diversas distribuições para caracterizar os erros, tais como a uniforme, a quadrática, a cosseno, a semi-circular ou a exponencial dupla (atualmente denominada por distribuição de Laplace), na procura de obter a distribuição da média dos erros. Para a obtenção de uma estimativa \hat{y} de y , função das observações y_i ($i = 1, \dots, n$), eram utilizados o método das médias (utilizar a média ponderada como estimador — $\hat{y} = \sum \omega_i y_i$ com $\sum \omega_i = 1$, sem haver qualquer justificação para tal procedimento), o método dos mínimos desvios absolutos (minimização de $\sum \omega_i |\varepsilon_i|$) e o método de minimizar o maior desvio absoluto ($\min_{\beta_1, \dots, \beta_m} \max_i |\varepsilon_i|$), que corresponde a uma solução do tipo *minimax* (minimizar o pior cenário possível).

O método dos mínimos quadrados

O método dos mínimos quadrados foi publicado independentemente por ADRIEN-MARIE LEGENDRE (1752–1833), em 1805, e por CARL FRIEDRICH GAUSS (1777–1855), em 1809, na obra que inclui a sua célebre previsão da localização do planeta anão (atualmente asteroide) Ceres. A primazia deste resultado foi disputada por estes dois matemáticos, pois apesar de LEGENDRE o publicar primeiro, GAUSS terá apresentado o resultado antes (em 1795), sendo atualmente este resultado atribuído usualmente a GAUSS.



GAUSS



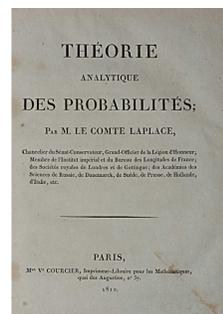
LEGENDRE

O método dos mínimos quadrados tornou-se um sucesso imediato, não só pela sua simplicidade, quer conceptual quer computacional (bem mais acessível que minimizar a soma dos desvios absolutos), bem como pela sua generalidade e relação com outros métodos que já seriam aplicados.

A distribuição normal e o Teorema Limite Central

GAUSS, em 1809, justifica a utilização do método dos mínimos quadrados em termos probabilísticos, demonstrando que a estimativa obtida por este método corresponde ao valor com maior probabilidade *a posteriori* (moda) se os erros forem caracterizados pela distribuição normal (Lei de Gauss). Contudo, GAUSS não conseguiu justificar a utilização da distribuição normal para os erros, referindo que é a única distribuição para o erro que faz com que a média aritmética se torne no valor mais provável quando temos observações de uma única quantidade desconhecida.

LAPLACE desenvolveu metodologias, utilizando funções geradoras (transformadas de Laplace) e análise assintótica de integrais, para deduzir as probabilidades aproximadas para médias de muitas observações e, deste modo, concluindo que independentemente da distribuição que caracteriza os erros, as probabilidades para a sua média podem ser determinadas utilizando a Lei de Gauss, resultado que corresponde à primeira versão geral do Teorema Limite Central (resultado que só foi rigorosamente demonstrado em 1901 por ALEKSANDR LYAPOUNOV (1857–1918)).



PÓLYA

A denominação Teorema Limite Central só surgiu em 1920 através de GEORGE PÓLYA (1887–1985), por considerar que este teorema assume um papel central entre os resultados sobre convergência, sendo por isso fundamental na Teoria da Probabilidade e na Estatística. Todavia, esta denominação é, por vezes, interpretada de forma errónea, ao ser considerado que o nome deriva de o resultado apresentar a convergência do centro (média) dos dados.

Este resultado também permitiu que LAPLACE justificasse a utilização da Lei de Gauss para caracterizar os erros, pois cada erro pode ser visto como a média de muitas influências independentes e, como tal, ser caracterizado por esta Lei, razão pela qual esta distribuição foi, durante muito tempo, denominada por lei dos erros ou lei dos desvios.

Por outro lado, uma vez que as estimativas habituais correspondiam a médias ponderadas das observações, LAPLACE conclui que o estimador será igualmente caracterizado aproximadamente pela lei de Gauss se houver um grande número de observações (independentemente das distribuições que caracterizam os erros individuais). Por fim, demonstra igualmente que estas estimativas têm o menor erro esperado comparativamente com quaisquer estimativas que correspondam a médias ponderadas das observações (entre as estimativas lineares).