A Probabilidade Inversa
Teorema de Bayes

Rui Santos *rui.santos@ipleiria.pt*, Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, CEAUL — Centro de Estatística e Aplicações da Universidade de Lisboa

Trabalho financiado por Fundos Nacionais através da Fundação para a Ciência e a Tecnologia, no âmbito do projeto PEst-OE/MAT/UI0006/2011.

Bayes e Laplace são pioneiros num raciocínio central em Estatística, ao deduzirem como, com base na informação obtida pela observação do fenómeno aleatório em análise, se pode tirar conclusões acerca das probabilidades associadas a cada acontecimento, um resultado basilar na fundamentação da inferência estatística.

Problema das probabilidades das causas

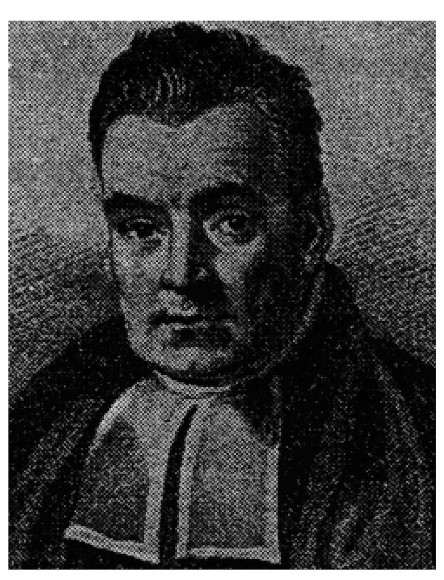
Recorrendo à notação atual, vamos representar por $\mathbb{P}(\mathbf{A}|\mathbf{B})$ a probabilidade de o acontecimento $\mathbf{A}$ ocorrer sabendo que o acontecimento $\mathbf{B}$ ocorreu, um conceito que só foi rigorosamente definido por Andrey Kolmogoroff (1903−1987) em 1933, mas que já era utilizado desde o século XVII em sequências, organização cronológica dos acontecimentos, onde tipicamente é definida a probabilidade de um acontecimento posterior se realizar sabendo que um acontecimento prévio se realizou. Com esta noção, podemos enunciar o problema das probabilidades das causas.

Problema das probabilidades das causas: Considere-se um conjunto de urnas que contêm bolas brancas e pretas. Sejam $\mathbb{P}(\mathbf{U}_1), \mathbb{P}(\mathbf{U}_2), \cdots, \mathbb{P}(\mathbf{U}_n)$ as probabilidades *a priori* das n urnas (causas) que dão lugar à saída de uma bola branca com probabilidade, respetivamente, $\mathbb{P}(\mathbf{B}|\mathbf{U}_1), \mathbb{P}(\mathbf{B}|\mathbf{U}_2), \cdots, \mathbb{P}(\mathbf{B}|\mathbf{U}_n)$. Tire-se, à sorte, uma urna e, da urna que sair, tire-se, à sorte, uma bola que, por hipótese, sai branca. Qual é a probabilidade $\mathbb{P}(\mathbf{U}_i|\mathbf{B})$ de que a bola tirada pertença a uma urna cuja probabilidade de sair bola branca seja $\mathbb{P}(\mathbf{U}_i)$?

A solução é dada pelo Teorema de Bayes, que neste caso (discreto) pode ser apresentado por:

$$\mathbb{P}(\mathbf{U}_i|\mathbf{B}) = \frac{\mathbb{P}(\mathbf{B}|\mathbf{U}_i) \times \mathbb{P}(\mathbf{U}_i)}{\sum_{j=1}^{n} \mathbb{P}(\mathbf{B}|\mathbf{U}_j) \times \mathbb{P}(\mathbf{U}_j)}, \quad i = 1, \ldots, n.$$

O artigo póstumo de Bayes



Bayes

Em 1764 é publicado o artigo póstumo do Reverendo Thomas Bayes (1701−1761) intitulado *An Essay Toward Solving a Problem in the Doctrine of Chances*. Neste artigo surge, pela primeira vez, o princípio da probabilidade inversa (atualmente denominado por Teorema ou regra de Bayes) para o caso contínuo (o caso discreto, previamente apresentado, foi unicamente deduzido por Laplace). Este resultado permite atualizar as probabilidades associadas a cada acontecimento (probabilidades *a posteriori*) depois de ter sido realizada uma experiência (observação do fenómeno em estudo), com base nos resultados observados nessa experiência e nas probabilidades associadas a cada resultado antes da realização da experiência (probabilidades *a priori*).

Neste artigo Bayes deduz a probabilidade de p (probabilidade de a próxima bola retirada da urna ser branca) se situar num dado intervalo $[p_1, p_2]$, sabendo que das n bolas previamente retiradas $n_\mathbf{B}$ eram brancas. Nesta dedução considera que, uma vez que não tem qualquer informação acerca da proporção de bolas brancas na urna antes da observação das bolas retiradas, todos os possíveis valores para proporção de bolas brancas são igualmente prováveis e, como tal, utilizou a distribuição uniforme no intervalo $[0, 1]$ para caracterizar a informação *a priori* sobre p (distribuição *a priori*). Desta forma ficou resolvido o problema de como analisar o parâmetro da binomial p em função de um conjunto de dados, i.e. Bayes apresentou a forma de deduzir a probabilidade *a posteriori* para o parâmetro p da binomial utilizando a sua regra para a inferência indutiva.

A regra da sucessão de Laplace

Pierre Laplace (1749−1827) redescobriu o Teorema de Bayes, aparentemente de forma independente de Bayes, numa publicação de 1774, na qual apresenta a sua célebre **Regra da Sucessão** associada ao seguinte problema.

Problema da Regra da Sucessão: De uma urna contendo uma infinidade de bolas brancas e pretas com percentagem desconhecida retiramos $n = n_\mathbf{B} + n_\mathbf{P}$ bolas, das quais $n_\mathbf{B}$ são brancas e $n_\mathbf{P}$ são pretas. Qual a probabilidade de a próxima bola a ser retirada seja branca?

Recorrendo a notação atual, Laplace utilizou uma sucessão de variáveis aleatórias $X_1, X_2, \cdots, X_n, \cdots$, onde X_i representa a i-ésima bola retirada, que assume o valor 1 no caso de a bola ser branca e assume o valor 0 caso seja preta (isto é, as variáveis aleatórias X_i representam provas de Bernoulli com probabilidade de sucesso p desconhecida) e considerou que:

1. as variáveis aleatórias X_i são dependentes se p é desconhecido (só são consideradas independentes quando condicionalmente a um valor de p fixo);
2. a proporção de bolas brancas é caracterizada por uma variável aleatória p com distribuição uniforme no intervalo $[0, 1]$ (pelo facto de desconhecermos o verdadeiro valor da proporção p).

Neste contexto, Laplace conclui que

$$\mathbb{P}(X_{n+1} = 1 \mid X_1 + \cdots + X_n = n_\mathbf{B}) = \frac{n_\mathbf{B} + 1}{n + 2}.$$

Este resultado é denominado por regra da sucessão (*rule of succession*) de Laplace. Uma das razões do seu aparecimento na maioria das obras publicada no século XIX deve-se ao facto de ser controverso. Não é a autenticidade do resultado, sob as hipótese referidas, que é questionada. O que é problemático são duas hipóteses utilizadas na sua resolução.

A primeira hipótese assume enorme importância na Estatística bayesiana, que não aceita independência entre as variáveis aleatórias X_i se p for desconhecido, uma vez que, se fossem independentes, nada se aprenderia com a experiência uma vez que, nesse caso, obteríamos facilmente

$$\mathbb{P}(X_{n+1} = 1 \mid X_1 + \cdots + X_n = n_\mathbf{B}) = \mathbb{P}(X_{n+1} = 1),$$

não havendo qualquer processo de aprendizagem (isto é, mesmo que se observe tiragens de bolas da urna não se pode concluir nada sobre o valor da proporção de bolas brancas). A inexistência de um processo de aprendizagem é fortemente criticada pelos defensores da interpretação bayesiana de probabilidade. Se não sabemos qual a probabilidade de cada cor e se retirarmos 1000 bolas e todas elas forem brancas, será que continuamos sem saber nada sobre a probabilidade de uma bola ser branca? Para os bayesianos, como a probabilidade de sucesso é desconhecida, o facto de saírem bolas brancas ou pretas altera o nosso grau de credibilidade sobre a cor que vai sair a seguir. Se, em 1000 bolas retiradas todas forem brancas, teremos um forte grau de credibilidade de que a próxima bola a sair também será branca.

O artigo de 1774 de Laplace é considerado, frequentemente, como o primeiro artigo bayesiano, no qual é apresentada uma metodologia geral de inferência estatística baseada na probabilidade inversa. Posteriormente, Laplace generalizou este resultado, aplicando a diversos parâmetros (não apenas à proporção de sucessos) e recorrendo a distintas distribuições para caracterizar a probabilidade *a priori* (quer discretas quer contínuas). Obtemos, deste modo, uma metodologia geral para a estimação de um parâmetro.

As discussões em torno da aplicabilidade da regra da sucessão

Bertrand

Todavia, mesmo aceitando as hipóteses previamente consideradas, são ainda discutíveis os limites de aplicação da fórmula deduzida. Deste modo, durante muitos anos diversos Matemáticos discutiram os limites da sua aplicação, sendo extremamente polémica a utilização em algumas situações famosas, como ilustram as célebres discussões filosóficas sobre a probabilidade de o sol nascer amanhã. Joseph Bertrand (1822−1900) em 1888, considerando que já observamos o sol há seis mil anos nos quais o sol nasceu todos os dias (todas as 2191500 bolas observadas são brancas) estimou a probabilidade de o sol nascer amanhã através de

$$\mathbb{P}(\text{“Sol nascer amanhã”}) = \frac{2\,191\,501}{2\,191\,502} \approx 0.9999995437.$$

Esta análise é efetuada em diversas obras ao longo dos séculos XVIII e XIX, uns defendendo a sua aplicação, outros criticando fortemente a sua utilização. Por exemplo, Antoine Cournot (1801−1877), em 1843, refere que não faz qualquer sentido apostar 2 para 1 (probabilidade igual a $\frac{2}{3}$) que vai sair cara num segundo lançamento de uma moeda apenas pelo facto de no primeiro lançamento ter saído cara!