Terceira Lista de Exercícios

Mestrado Professional em Economia — Econometria Avançada
Professor: Hedibert Freitas Lopes — Monitor: Guilherme Piantino
Insper Instituto de Educação e Pesquisa — November/2024

Questão 1: P(dados | hipótese) vs P(hipótese | dados)

Suponha que você acorde um dia com manchas em todo o rosto. O médico lhe diz que 90% das pessoas com varíola apresentam os mesmos sintomas que você. Em outras palavras, a probabilidade de ter esses sintomas, dado que você tem varíola, é de 0,9 (ou seja, 90%). Como a varíola é frequentemente fatal, você naturalmente fica aterrorizado. No entanto, após alguns momentos de reflexão, você decide que não quer saber a probabilidade de ter esses sintomas (afinal, você já sabe que os tem). Em vez disso, o que você realmente quer saber é a probabilidade de ter varíola.

Então você diz ao seu médico: "Sim, mas qual é a probabilidade de eu ter varíola, dado que tenho esses sintomas?". "Ah", diz o médico, "uma pergunta muito boa." Após rabiscar algumas equações, o médico olha para você. "A probabilidade de você ter varíola, dado que tem esses sintomas, é de 1,1%, ou, equivalentemente, 0,011." Claro, isso não é uma boa notícia, mas parece melhor do que 90% e (mais importante) é, pelo menos, uma informação útil. Isso demonstra o contraste marcante entre a probabilidade dos sintomas dado uma doença (que você não quer saber) e a probabilidade da doença dado os sintomas (que você quer saber).

A regra de Bayes transforma probabilidades que parecem úteis (mas muitas vezes não são) em probabilidades que realmente são úteis. No exemplo acima, o médico usou a regra de Bayes para transformar a probabilidade pouco informativa dos seus sintomas, dado que você tem varíola, na probabilidade informativa de que você tem varíola, dado os seus sintomas.

$$Pr(\text{você tem sintomas} \mid \text{você tem varíola}) \neq Pr(\text{você tem varíola} \mid \text{você tem sintomas})$$

Agora, suponha que você é um médico, confrontado com um paciente coberto de manchas. Os sintomas do paciente são consistentes com catapora, mas também são consistentes com outra doença mais perigosa, a varíola. Assim, você enfrenta um dilema. Você sabe que 80% das pessoas com catapora apresentam manchas, mas também que 90% das pessoas com varíola apresentam manchas. Portanto, a probabilidade (0,8) dos sintomas, dado que o paciente tem catapora, é semelhante à probabilidade (0,9) dos sintomas, dado que o paciente tem varíola.

Se você fosse um médico com pouca experiência, talvez achasse que a catapora e a varíola são igualmente prováveis. Mas, como é um médico experiente, sabe que a catapora é comum, enquanto a varíola é rara. Esse conhecimento, ou informação prévia, pode ser usado para decidir qual doença o paciente provavelmente tem. Se tivesse que adivinhar (e você precisa adivinhar, pois é o médico), então combinaria os possíveis diagnósticos implícitos pelos sintomas com seu conhecimento prévio para chegar a uma conclusão (ou seja, que o paciente provavelmente tem catapora).

Podemos calcular as probabilidades associadas a uma doença utilizando estatísticas de saúde pública. Suponha que os médicos sejam solicitados a relatar o número de casos de varíola e catapora, e os sintomas observados. Usando os resultados de tais pesquisas, é uma questão simples encontrar a proporção de pacientes diagnosticados com varíola e catapora, bem como os sintomas de cada paciente (por exemplo, manchas). Com esses dados, podemos descobrir que a probabilidade de um paciente ter manchas, dado que ele tem varíola, é de 90% ou 0,9. Da mesma forma, poderíamos descobrir que manchas são observadas em 80% dos pacientes que têm catapora. Em termos de probabilidade condicional:

$$Pr(\text{manchas} \mid \text{varíola}) = 0,9$$
$$Pr(\text{manchas} \mid \text{catapora}) = 0,8$$

Essas equações não levam em conta nossa experiência anterior sobre a prevalência relativa da varíola e da catapora, e são baseadas apenas nos sintomas observados.

Estatísticas de saúde pública podem nos informar que a prevalência da varíola na população geral é de 0.001, o que significa que há uma chance em mil de que um indivíduo escolhido aleatoriamente tenha varíola. Assim, a probabilidade de que um indivíduo escolhido aleatoriamente tenha varíola é

$$Pr(\text{varíola}) = 0,001.$$

Similarmente, e como catapora é muito mais comum,

$$Pr(\text{catapora}) = 0,1.$$

Problema: Um dos nossos interesses é em saber qual a chance/probabilidade de um paciente aleatório ter varíola, dado que o paciente apresenta manchas, assumindo que 81% da população tem manchas. Similarmente, outro interesse é em saber qual a chance/probabilidade de um paciente aleatório ter catapora, dado que o paciente apresenta manchas, assumindo novamente que 81% da população tem manchas.

Observação: Note que o complemente da proposição/evento "varíola" não é "catapora", mas sim "não varíola".

Questão 2: Retornos dependentes

Em qualquer dia, o retorno de uma ação é positivo ou negativo com probabilidades iguais. Entretanto, a chance do retorno ser positivo dado que no dia anterior também foi positivo é maior, digamos 80%. Similarmente, a chance do retorno ser positivo dado que no dia anterior foi negativo é de 60%. No final do dia de hoje, no fechamento da bolsa, você observou que a ação produziu um retorno negativo, mas se esqueceu de observar se ontem o retorno foi negativo ou positivo. Qual dos dois eventos é mais provável? Ou seja, é mais provável que o retorno de ontem tenha sido positivo ou negativo, dado que o retorno de hoje é negativo?

Questão 3: Eleições municipais

Em uma determinada cidade, 30% das pessoas são conservadoras, 50% são liberais e 20% são independentes. Registros mostram que, em uma eleição específica, 65% dos conservadores votaram, 82% dos liberais votaram e 50% dos independentes votaram. Se uma pessoa na cidade é selecionada aleatoriamente e se descobre que ela não votou na última eleição, qual é a probabilidade de que ela seja uma liberal?

Questão 4: Especificidade/sensibilidade de um novo teste

Um novo teste foi desenvolvido para detectar um tipo específico de câncer. Se o teste for aplicado a uma pessoa que tem esse tipo de câncer, a probabilidade de que a pessoa tenha uma reação positiva é de 0,95 e a probabilidade de que a pessoa tenha uma reação negativa é de 0,05. Se o teste for aplicado a uma pessoa que não tem esse tipo de câncer, a probabilidade de que a pessoa tenha uma reação positiva é de 0,05 e a probabilidade de que a pessoa tenha uma reação negativa é de 0,95. Suponha que, na população geral, uma pessoa a cada 100.000 tenha esse tipo de câncer. Se uma pessoa escolhida aleatoriamente tiver uma reação positiva ao teste, qual é a probabilidade de que ela tenha esse tipo de câncer?