Matemática Discreta

Pedro Hokama

Fontes

- Gomide, Anamaria; Stolfi, Jorge. Elementos de Matematica Discreta para Computação.
- Rosen, Kenneth H. Discrete mathematics and its applications. McGraw-Hill Education, 8th Edition, 2019.

Valor Esperado - continuação

$$\mathcal{E}(X) = \sum_{v \in D} v \Pr(X = v) \tag{1}$$

Exercício: Furar um poço de petróleo em determinada região custa R$500.000, e tem 30% de chance de encontrar óleo. Se isso acontecer, o poço pode ser vendido por R$800.000. Caso contrário o investimento é totalmente perdido. Qual o ganho esperado por poço?

- Antes de continuar vamos mostrar 2 propriedades:
- É fácil ver que [1]

$$\sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6}$$

- e também

$$\sum_{x=1}^{n} \frac{1}{x} \geq \int_1^n \frac{1}{x}\, dx = \ln x \Big|_1^n = \ln n - \ln 1 = \ln n$$

[1] mentira, é bem difícil ver: `https://en.m.wikipedia.org/wiki/Basel_problem`

- Quando o domínio da variável é um conjunto infinito, o valor esperado pode ser infinito, mesmo que todos os seus valores possíveis sejam finitos.
- Por exemplo, considere a variável X cujo valor é um inteiro positivo, tal que $\Pr(X = k) = (6/\pi^2)/k^2$ para todo $k \in \mathbb{N} \setminus \{0\}$.
- Esta distribuição de probabilidades é válida, pois verifica-se que a soma de todas as probabilidades é 1.

$$\sum_{k=1}^{\infty} \frac{6/\pi^2}{k^2} = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{6}{\pi^2} \frac{\pi^2}{6} = 1$$

Entretanto, o valor esperado de X deveria ser a somatória

$$\mathcal{E}(X) = \sum_{k=1}^{\infty} k \cdot \frac{A}{k^2} = A \sum_{k=1}^{\infty} \frac{1}{k}$$

que, como sabemos, não tem valor finito.

- O valor esperado pode ser definido para qualquer variável cujos valores podem ser somados e multiplicados por um número real.
- Por exemplo, suponha que o valor de uma variável aleatória X é um par (u, v), onde u é o resultado de lançar uma moeda ($0 =$ cara, $1 =$ coroa), e v é o resultado de lançar um dado (um inteiro entre 1 e 6); sendo que cada par possível tem a mesma probabilidade $1/12$.
- Note que esses pares podem ser considerados vetores do espaço $\mathbb{R}^2$. Portanto podemos calcular o valor esperado de X

$$\mathcal{E}(X) = \frac{1}{12}((0,1) + (0,2) + \cdots + (1,5) + (1,6)) = (\frac{1}{6}, \frac{7}{2})$$

Propriedades do valor esperado

- Seja X uma variável aleatória com domínio numérico, sejam α e β dois números reais quaisquer, e seja Z a variável aleatória $\alpha X + \beta$.
- Nesse caso, pode-se provar que

$$\mathcal{E}(Z) = \mathcal{E}(\alpha X + \beta) = \alpha\,\mathcal{E}(X) + \beta \tag{2}$$

- Porém, se uma variável aleatória Z depende de X de maneira não linear (por exemplo, se Z é o quadrado de X), não existe uma fórmula geral que relacione $\mathcal{E}(Z)$ a $\mathcal{E}(X)$

- Sejam X e Y duas variáveis aleatórias com valores numéricos, e seja Z a variável aleatória, denotada por $X + Y$, cujo valor é a soma dos valores de X e de Y.
- Verifica-se que

$$\mathcal{E}(Z) = \mathcal{E}(X) + \mathcal{E}(Y) \tag{3}$$

Prova:

$$\mathcal{E}(Z) = \sum_z z Pr(Z = z) \tag{4}$$

$$= \sum_x \sum_y (x + y) Pr(X = x \wedge Y = y) \tag{5}$$

$$= \sum_x \sum_y x Pr(X = x \wedge Y = y) + y Pr(X = x \wedge Y = y) \tag{6}$$

$$= \sum_x x \left(\sum_y Pr(X = x \wedge Y = y) \right) + \sum_y y \left(\sum_x Pr(X = x \wedge Y = y) \right) \tag{7}$$

$$= \sum_x x Pr(X = x) + \sum_y y Pr(Y = y) = \mathcal{E}(X) + \mathcal{E}(Y) \tag{8}$$

Fim.

Estas fórmulas valem mesmo que as variáveis X e Y tenham alguma dependência entre si. Note que não há fórmulas análogas para outras operações (como produto, divisão, etc.).

Exercício: Um dado vai ser lançado, e a seguinte aposta é oferecida: o cliente paga $\mathrm{R\$}7,00$ ao banqueiro, e recebe em reais o dobro do valor que sair no dado. Por exemplo, se sair um 4, o cliente recebe $\mathrm{R\$}8,00$, obtendo um ganho líquido de $\mathrm{R\$}1,00$. Qual é o ganho esperado do cliente?

Exercício: Na mesma situação do exercício anterior, uma outra aposta é oferecida: cliente paga $\mathrm{R\$}49,00$ ao banqueiro, e recebe em reais o dobro do quadrado do valor que sair no dado. Por exemplo, se sair um 6, o cliente recebe $2 \times 6^2 = \mathrm{R\$}72,00$, obtendo um ganho líquido de $\mathrm{R\$}23,00$. Qual é o ganho esperado do cliente?

Teoria da informação

- Hoje em dia todos conhecem o conceito de **bit** e outras unidades derivadas, como **byte** (8 bits), **megabyte** (10^6 ou 2^{20} bytes, conforme o contexto), **gigabyte** (10^9 ou 2^{30} bytes) etc.
- Em geral esses conceitos são usados para descrever tamanhos de arquivos, capacidade de memória, taxas de transmissão, etc.
- Porém é necessário distinguir entre a **capacidade de armazenamento de informação** de tais sistemas, e a **quantidade de informação** contida neles em determinado momento. Este segundo conceito é o centro da **teoria da informação**, desenvolvida principalmente pelo matemático e engenheiro americano Claude Shannon (1916–2001), em meados do século 20.

Capacidade de informação

- Considere um sistema físico (real ou imaginário) que em qualquer momento pode assumir um único estado dentre uma coleção finita de estados possíveis; sendo que esse estado pode ser identificado com precisão por algum tipo de teste ou medida.
- Por exemplo, uma moeda sobre uma mesa, que pode estar na posição ‘cara’ ou ‘coroa’; um dado de jogar, que pode estar virado com qualquer face, entre 1 e 6, para cima; uma chave elétrica, que pode estar ‘desligada’ ou ‘ligada’; um fio elétrico, que pode estar a zero volts ou a $+5$ volts; uma barra de ferro, que pode estar magnetizada em dois sentidos diferentes; e assim por diante. Tal objeto é dito um **sistema discreto**.

- Suponha que o sistema tem apenas dois estados possíveis (ou seja, é um **sistema binário**). Por definição, a capacidade de informação de tal sistema é 1 bit. Se o sistema tem 2^b estados possíveis, sua capacidade é b bits.
- Observe que podemos numerar os estados de tal sistema em base 2 usando b algarismos, cada qual 0 ou 1: — $0\cdots00 = 0$, $0\cdots01 = 1$, $0\cdots10 = 2$, $0\cdots11 = 3$, …, $1\cdots11 = 2^b - 1$. Daí o nome “bit”, que é abreviação do inglês **binary digit**.

- Mais geralmente, se o número de estados possíveis n, a capacidade de informação é definida como $\log_2 n = (\ln n)/(\ln 2)$, o logaritmo de n na base 2. Assim, por exemplo, a capacidade de informação de um dado de jogar, em repouso sobre a mesa, é $\log_2 6 = 2,5849625007\ldots$ bits.
- Note que, se n não é uma potência de 2, a capacidade em bits não é um número inteiro (e, na verdade, é um número irracional). Note também que se o sistema tem apenas um estado possível, sua capacidade de armazenar informação é (como se pode esperar) zero bits.

Esta definição implica na seguinte propriedade:

Teorema: Se um sistema S consiste de dois sub-sistemas discretos A e B independentes (no sentido de que cada estado possível de A pode co-existir com qualquer estado possível de B, e vice-versa), então a capacidade de S é a soma das capacidades de A e de B.

Exercício: Determine a capacidade de informação dos seguintes sistemas:

1. Um odômetro (mostrador de quilometragem) de automóvel com 6 algarismos decimais.
2. Um dado em forma de octaedro, com faces numeradas de 1 a 8, em respouso sobre a mesa.
3. Uma cadeia de DNA com 100 elementos (**nucleotídeos**), cada qual podendo ter quatro estruturas químicas possíveis — adenosina (A), timina (T), guanina (G), ou citosina (C).

Exercício: Determine a capacidade de informação dos seguintes sistemas, constituídos de 4 moedas, cada qual podendo ser de 5, 10, 25, ou 50 centavos, que somente podem ser distinguidas pelo seu valor:

1. Uma pilha, em qualquer ordem.
2. Uma pilha, em ordem crescente de valor.
3. Uma coleção em um saco.
4. Uma pilha onde todas as moedas tem o mesmo valor.

Exercício: Refaça o exercício 16, supondo que todas as moedas de mesmo valor estão marcadas com letras distintas entre ‘A’ e ‘D’. Assim, por exemplo, na alternativa 1, as moedas poderiam ser, na ordem, $(10, D), (25, C), (10, B), (10, C)$ mas não poderiam ser $(10, D), (25, C), (10, B), (10, D)$.

Exercício: Qual é a capacidade de informação de uma carta retirada de um baralho com 13 cartas? E de um baralho com 52 cartas? Se acrescentarmos um coringa ao baralho, de quanto aumenta a capacidade, em cada caso?

Quantidade de informação

- A capacidade de informação de um sistema discreto diz apenas o limite máximo de informação que pode ser armazenada nele. Porém, dependendo de como o sistema é usado, nem toda a capacidade pode ser utilizada.
- Por exemplo, considere uma lâmpada que, ao meio-dia, pode estar acesa ou apagada conforme o sol tenha nascido ou não naquele dia. Embora a capacidade de informação desse sistema seja 1 bit, intuitivamente a notícia de que essa lâmpada está acesa não traz muita informação. Por outro lado, uma lâmpada que indica se está chovendo ou não fora do prédio parece fornecer mais informação — muito embora sua **capacidade** de informação seja exatamente a mesma.

- A diferença entre estes dois exemplos está na probabilidade que atribuímos aos dois estados do sistema. No primeiro caso, é natural atribuir probabilidade bem próxima a 1 à afirmação “a lâmpada está acesa” (menos que sejamos extremamente pessimistas!).
- Por isso, a notícia de que essa informação é verdadeira não muda muito nosso estado de conhecimento. Já, no segundo exemplo, faz sentido atribuir probabilidade bem menor que 1 a essa afirmação (menos que estejamos na Bolívia, onde nunca chove!).

- Para tornar esta intuição mais precisa, suponha que X é uma variável aleatória que pode assumir um certo valor v. A **quantidade de informação** trazida pela notícia "o valor de X é v" é, por definição,

$$Q(X = v) = \log_2 \frac{1}{\Pr(X = v)} = -\log_2 \Pr(X = v)$$

Este valor, como a capacidade de informação, é medido em bits, e nunca é negativo. Em particular, se X pode assumir n valores distintos com igual probabilidade $\Pr(X = v) = 1/n$, a quantidade de informação que recebemos quando ficamos sabendo o valor de X (qualquer valor de X) é exatamente $Q(X = v) = \log_2 n$ bits — ou seja, a capacidade da variável X.

Porém, se as probabilidades dos valores de X não são iguais, a quantidade de informação pode ser menor ou maior, dependendo do valor. Por exemplo:

Exemplo: Suponha que um dado está para ser lançado, e X é uma variável que vale 100 se o resultado do dado é 1, e 200 caso contrário. Então as notícias "$X = 100$" e "$X = 200$" carregam as seguintes quantidades de informação:

$$\begin{aligned} Q(X = 100) &= -\log_2 \Pr(X = 100) = -log_2 \frac{1}{6} \approx 2,5849625\ldots \\ Q(X = 200) &= -\log_2 \Pr(X = 200) = -log_2 \frac{5}{6} \approx 0,2630344\ldots \end{aligned}$$

Neste exemplo, observe que a notícia "$X = 200$" traz muito menos informação do que a notícia "$X = 100$", porque tem probabilidade maior — $5/6$ em vez de $1/6$.

Quantidade esperada de informação

- No exemplo 22, observe também que a notícia "$X = 100$" traz mais que 1 bit de informação — muito embora a variável X tenha apenas dois valores possíveis, e portanto tenha apenas 1 bit de capacidade.

- Este paradoxo é resolvido se considerarmos a **quantidade esperada de informação**, ou **entropia**, da variável X. Ou seja, a quantia

$$\mathcal{H}(X) = \sum_v \Pr(X = v)\, Q(X = v) = \sum_v -\Pr(X = v) \log_2 \Pr(X = v) \qquad (9)$$

- Nesta fórmula, o índice v do somatório assume todos os valores possíveis da variável X. Observe que, como na fórmula (1), cada termo desta soma é a quantidade de informação trazida pela notícia "$X = v$", vezes a probabilidade de recebermos essa notícia. Pode-se verificar que $\mathcal{H}(X)$, assim como cada termo $Q(X = v)$, é um valor real não negativo.

- No exemplo 22, a quantidade esperada de informação que recebemos ao conhecer o valor de X é

$$\begin{aligned}\mathcal{H}(X) &= \Pr(X=100)\,Q(X=100) + \Pr(X=200)\,Q(X=200)\\ &= \tfrac{1}{6}\log_2\tfrac{6}{1} + \tfrac{5}{6}\log_2\tfrac{6}{5}\\ &\approx \tfrac{1}{6}2,5849625\ldots + \tfrac{5}{6}0,2630344\ldots\\ &\approx 0,65002241\ldots\end{aligned}$$

- Observe que, embora a notícia "$X = 100$" forneça mais de 2,5 bits de informação, ela é muito menos provável que a notícia "$X = 200$", que fornece menos que 0,27 bits de informação. Assim, a quantidade esperada de informação que ganhamos ao saber o valor de X é cerca de 0,65 bits, ou seja abaixo da capacidade de X (1 bit). Esta última observação é um resultado importante:

  Teorema: Se uma variável aleatória X pode assumir n valores distintos, então a quantidade esperada de informação que ganhamos conhecendo o valor de X é no máximo a capacidade de X, $\log_2 n$; e é exatamente $\log_2 n$ apenas quando todos esses valores podem ocorrer com igual probabilidade $1/n$.

- Devido a este teorema, a fórmula (9) é muito usada para medir a "uniformidade" da distribuição de probabilidades de uma variável aleatória X. O valor de $\mathcal{H}(X)$ varia entre 0 e $\log_2 n$, onde n é o número de valores possíveis de X. Quanto maior $\mathcal{H}(X)$, mais uniforme a distribuição. Na verdade, a fórmula (9) pode ser usada com qualquer lista de n valores reais $p_0, p_1, \ldots p_{n-1}$ não negativos cuja soma é 1.

- Observe que se X tem uma distribuição degenerada — com $\Pr(X = v) = 1$ para um único valor v, e zero para os demais valores — então $\mathcal{H}(X)$ é zero. Ou seja, se temos certeza de qual vai ser o valor de X, nossa expectativa é que a revelação desse valor não vai nos trazer nenhuma informação.