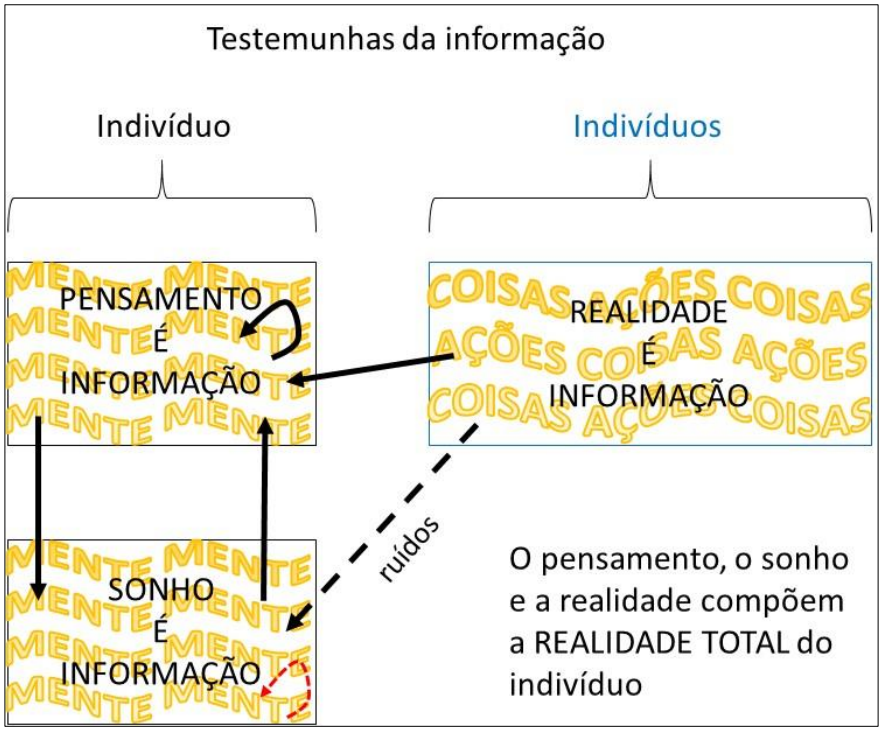
O QUE É INFORMAÇÃO

Informação não é algo físico. Tudo está impregnado de informações. O universo físico está impregnado de informações As próprias informações, como o pensamento e o sonho, estão impregnadas de informação, obviamente.
Pensamentos e sonhos são informações, mas, nem toda informação é pensamento e sonhos. Porém, desconsiderando-se a fonte das informações, não há como discernir se a informação que chega é um pensamento ou não.



Apenas o indivíduo isolado de um grupo recebe as informações de seus pensamentos e de seus sonhos. Os pensamentos são elaborados e/ou formado a partir de fragmentos, tendo como portador (o meio que entrega as informações) a **mente**. Os sonhos, por sua vez, são pensamentos criados por um EU secundário de cada indivíduo, tendo como portador a **mente** também.
A realidade é aquilo que o indivíduo vê e compartilha com um grupo de indivíduos. É aqui que eu e você podemos ter uma conversa coerente. Os portadores da realidade são as **coisas** (objetos) e as **ações** de cada indivíduo.
Conforme mostrado pelas setas na figura, a realidade influencia o pensamento e os sonhos (com menor intensidade – você pode estar sonhando com um lugar calmo e, de repente, passa um avião sobre tua casa e, então, um avião ou qualquer outro objeto causado pelo barulho real do avião vai aparecer no teu sonho). É o que estou chamando de ruídos (seta tracejada).

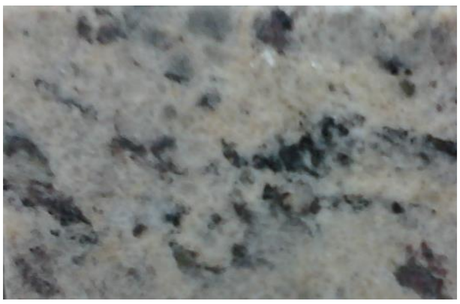
Pensamentos influenciam os próprios pensamentos e influenciam os sonhos, enquanto sonhos influenciam pensamentos e podem, até, influenciar os próprios sonhos. É o que mostra a seta tracejada vermelha. Não é incomum sonhar que estamos sonhando.
A realidade de um indivíduo é sempre maior do que a realidade do grupo de indivíduos, pois ela se compõe de seus pensamentos, sonhos e aquilo que ele compartilha simultaneamente com os demais indivíduos.
Veja estes quadros:

Um texto escrito em uma língua qualquer passa informações para aqueles que conseguem interpretar o arranjo dos caracteres desse texto.	`¤¨¬°´¸ØðĐĦħļļĿŀŁłŊőţŧųƆƍ ƏƕƚƢƣƪƭƸǀǁǏǘǜȸɁɅɉɟɭʉʲʵʴ˘˟ ƏƕƚƢƣƪƭƸǀĦħļļĿŀŁł	

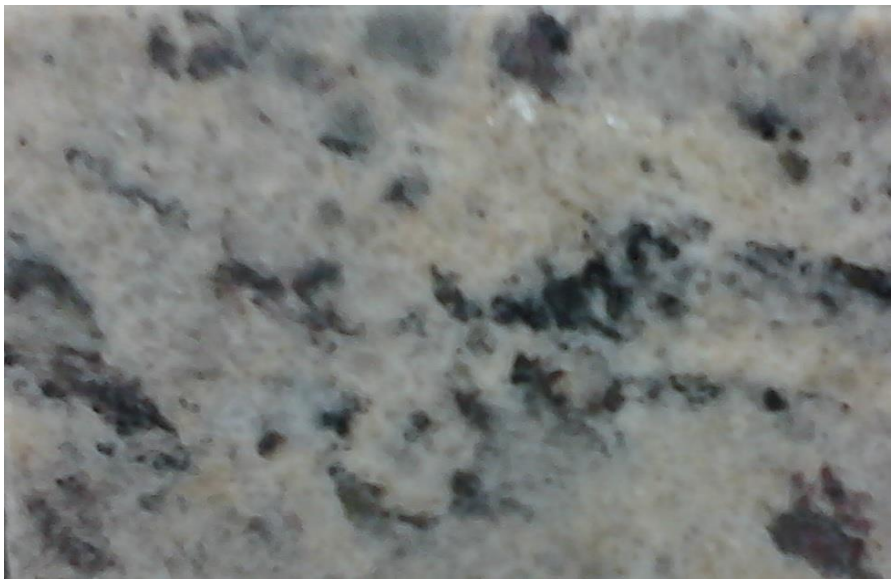
Você tem certeza que o primeiro contém informações impregnadas em um conjunto físico de caracteres. E sabe que o segundo tem informações também, só não sabe de que tipo. A parte física é a mesma nos dois quadros (caracteres, estranhos para uns, mas, não para outros, e vice-versa).

Você sabe que até o terceiro quadro pode conter alguma informação. E quanto ao quarto quadro? E o quinto?
Não tem quarto quadro e nem quinto quadro, então não tem informação. Veja que nem tem parte física. Assim, se tem físico, tem informação, mas, se tem informação, não precisa ter físico.
A informação não é o conjunto de caracteres físicos. É o que se *desprende* deles, conforme o arranjo. É impalpável, não é possível de ser medida. Se a informação fosse o conjunto de caracteres, qualquer arranjo resultaria nas mesmas informações. Os caracteres e seu arranjo são os portadores da informação, não são a própria informação.

O portador da informação nem sempre está ciente do conteúdo das informações que ele carrega. Veja a figura a seguir:



Isto é uma seção de uma pedra de mármore de uma pia. É um pedaço de pedra que está entregando várias informações. Eu não sei você, mas, eu consigo ver um animal bem no centro inferior. Eu poderia classifica-lo como um cachorro ou como um boi. Se você olhar bem, poderá discernir outras figuras. Eis a pedra um pouco aumentada:



A pedra entrega informação sem estar ciente disso, sem ter sido a criadora da informação.

Informação é a interpretação dada pelo receptor.

A interpretação pode "bater" com o que a fonte originadora quis entregar ou não.

Informação cria, porém, ela cria algo mental. Informar significa dar uma forma mental à coisa passada pela informação, aquilo que se desprende da fonte da informação.

Informação é qualquer evento que afeta o estado de um sistema.

Ela não pode vir do nada. Mesmo quando você olhou para o quadro vazio lá em cima, uma informação que possa ter saído dele, na verdade, foi forjada na tua cabeça. Assim, informação é sempre criada e pode nascer, sim, em uma dimensão desconhecida tanto quanto em dimensões conhecidas.

Apenas seres vivos percebem informação?
Em outras palavras, coisas e objetos percebem informação?

> Se perceber é *interpretar*, *analisar*, então, sim, só seres vivos percebem.
> Se perceber é *reagir*, então coisas e objetos também percebem informação.

Ora, interpretar, analisar são reações, então não são apenas os seres vivos que percebem informação.

> O magneto A se aproxima do magneto B. O magneto B vai reagir, afastando-se ou se aproximando de A.

O DNA tem informação? Ele entrega informação?

> O DNA é uma sequência de moléculas eletromagnéticas que causará uma série de associações/dissociações, resultando num novo composto eletromagnético que conterá uma informação que pode resultar em outras informações, conforme as associações que forem feitas.
> Ele é como um algoritmo, um programa, uma receita de bolo. Portanto, ele contém informação e entrega informação.

Informação é mais importante do que tempo?

> Sem tempo pode haver informação, mas, ela será inútil, pois, não evoluirá, não será entregue, não agirá, por não persistir. Por outro lado, tempo é informação (mudança de estado, mesmo mental). Como algo que é entregue e causa uma mudança sim, informação é mais importante do que tempo, mas, não pode ser dissociada dele.

Informação é uma diferença (percebida) que faz (causa) uma diferença[1].

> Uma placa de contramão numa rua informa a diferença entre trafegar num sentido e no outro, obrigando o fluxo em um sentido e não no outro (uma diferença).
> Neste aspecto, a informação pode ser quantificada como sim/não, 1/0, etc.

Informação não precisa, necessariamente, ser entregue/interpretada.

> Isso implica que a informação pode existir, porém, não está disponível para acesso.
> O que ainda não sabemos pode estar em portadores com esse tipo de informação.

[1] Gregory Bateson (1904-1980)

Existe portador sem informação?

Não, pois, estar sem um tipo de informação é um tipo de informação. Um portador sempre passa uma informação. No mínimo fica a próprio portador (ou o próprio dado), vazio, mas informativo.
Se o portador não existe, então não existe informação ali. Depois que você morre, você não entre mais informação e nem recebe: não vê a realidade, não pensa e não testemunha seus sonhos.

Existe informação sem portador?

Sim. Veja o corpo. Ele entrega muitas informações, mas, ele, em si, é uma informação que não é entregue por algo.
Existe o dado e a informação que ele entrega. O que entrega o dado? Ele mesmo? Retire a informação que um dado passa e ainda fica ele como informação.

Retiremos do corpo todas as informações que ele entrega. O que sobra? O corpo! Mas, mesmo assim, ele está entregando a informação "corpo". Só temos duas alternativas:

1) Ele é informação.
2) Ele entrega a si mesmo.

Se ele é informação, concluímos que existe informação sem portador. Ponto final.
Se ele entrega a informação "corpo", vamos removê-la então. Aí vão sobrar duas alternativas:

a) O vazio. Então ele era informação, que implica que existe informação sem portador.
b) Fica o corpo, mas, ainda passando a informação "corpo". O que implica que não é possível apagar essa informação, o que implica que ele é a própria informação, o que implica que existe informação sem portador.

Sonho é informação.

Quando você pensa, seus pensamentos são gravados na sua memória. O pensamento é presente e é gravado para um possível replay depois. Isso é necessário para manter o controle mental, para você não ficar preso em um loop de pensamentos.
Quando você sonha, seus sonhos são gravados na sua memória. O sonho é presente e é gravado para um possível replay depois.

A questão é: você sabe que sonhou só se acionar o replay (essa ação normalmente é automática na primeira vez – o replay é executado logo que você acorda) ou você é testemunha presente de seus sonhos?
Testes médicos indicam que os fatos dos sonhos são presentes e indicam também que você é testemunha presente. Essa última indicação é dada pelos chamados *sonhos agitados*, que se refletem no corpo, provando que você está participando.
Depois que você acorda e aciona o replay, o mesmo sonho agitado não te afeta do tanto que afetou enquanto dormindo. Então, você é sempre testemunha presente de seus sonhos, da mesma maneira que você é testemunha presente de seus pensamentos. Se, no momento em que acordar, você morrer, você terá lembrado aquele sonho, pois, você o viveu.

Informação é o tudo e o nada: TUDO + NADA +NULO + VAZIO.

Teoria da Informação

A *Teoria da Informação* é uma formalização matemática da informação quanto ao modo de armazená-la, transmiti-la e recuperá-la no outro lado. Esta formalização foi feita pelo americano Claude Elwood Shannon, em 1948 (A Mathematical Theory of Communication), quando ele era funcionário do Laboratórios Bell. Shannon era matemático e engenheiro eletrônico. Nasceu em 1916 e morreu em 2001.
A Teoria da Informação de Shannon tem pouco a ver com o que discutimos anteriormente. Aqui se fala em como transmitir informação com eficiência e segurança, através de um meio que é chamado de canal:

Canal = Banda + Ruído

Banda é a capacidade que o canal tem, a quantidade de informações simultâneas que ele pode carregar sem perdas.
Ruído é qualquer interferência capaz de alterar o sinal, e daí a informação, seja esse sinal elétrico, sonoro, visual ou mesmo táctil.

A Teoria da Informação não se preocupa com o significado da informação. Isso fica por conta do receptor/interpretador. Ela trata mais do quantitativo da informação que será realmente útil para o receptor. Porém, como o receptor, antes de receber uma mensagem é algo passivo, uma espécie de espectador esperando por um ato, por evento, a Teoria da Informação faz uso de probabilidades e logaritmos para formalizar a definição.
Em termos de quantidade de informação útil, é óbvio que uma mensagem que tem grande probabilidade de chegar já tem seu conteúdo, de certo modo, conhecido por você, e assim não será tão útil, pois nada vai acrescentar de novo ao teu conhecimento. Por outro lado, se a probabilidade é baixa, se a mensagem traz uma informação rara, difícil de acontecer, ela, provavelmente, preencherá um grande vazio em seu conhecimento, tendo, assim, uma grande carga de informação, uma grande *quantidade de informação*.

Veja: quantidade de dados contendo informações é uma coisa; quantidade de informação é outra - o sentido aqui é o de ser novo, de ser útil. É disso que Shannon trata.

Exemplos

- Notícias sobre coisas que estão acontecendo agora são novidades, porque eram inesperadas. Uma notícia de ontem, que você já ouviu, não é novidade, então não tem informação (nova) tanto quanto as primeiras.

- Considere uma lâmpada que, ao meio-dia, pode estar acesa ou apagada, conforme tenha sol lá fora ou não. Depois, considere uma lâmpada que indica se está chovendo lá fora ou não, conforme esteja acesa ou apagada. No caso da primeira lâmpada, você não obterá muita informação, pois, a probabilidade de ter sol ao meio-dia é bem mais alta do que a probabilidade de estar chovendo lá fora. Você sabe que o sol é sempre (quase) certo, mas, chuva é mais raro, isto é, a probabilidade de o Sol brilhar todos os dias é muito maior do que a probabilidade de chover todos os dias. É por isso que a segunda lâmpada vai trazer uma maior quantidade de informação útil para você do que a primeira lâmpada.

Isso te leva a concluir, acertadamente, que quanto mais rara for uma mensagem, maior quantidade de informação ela contém (que ela vai acrescentar ao receptor).

A informação entregue é dependente do contexto, isto é, a mensagem tem que carregar o contexto também, ou, de alguma maneira, o receptor tem que conhecer o contexto. Por exemplo:

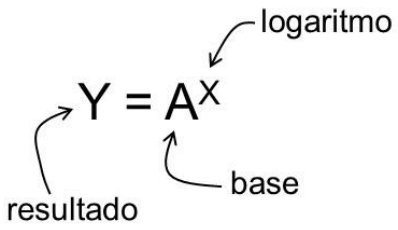
Em uma Escola, se você pede para chamar o *Silva* fica mais complicado do que você pedir para chamar o *zelador Silva*. Então, a palavra *Silva* carrega muito menos informação do que a expressão *zelador Silva*, pois *Silva* é um nome muito comum, pouco raro. No contexto em que foi proferida, praticamente não vai causar uma ação em ninguém, enquanto *zelador Silva* acrescenta muito mais informações ao ouvinte.

Quando queremos apontar uma localização onde tem várias pessoas próximas e vemos um baixinho (pessoa de pequena estatura) perto da localização, costumamos dizer: *Lá, perto daquele baixinho*. Porém, se tiver vários baixinhos ali por perto, a informação estará enfraquecida. O contexto, no caso, é a localização que a outra pessoa espera descobrir, ou seja, é a probabilidade da mensagem correta ser obtida. No caso da Escola, a probabilidade de chamar o *zelador Silva* é bem menor do que chamar o *Silva*. Assim, quanto mais rara for a mensagem, mais informação ela estará carregando.

Antes de continuarmos, vamos recordar logaritmos e probabilidades.

Logaritmos

Um número X é o *logaritmo* de um número Y quando uma *base* A elevada ao número X dá como *resultado* o número Y:



Mas, que operação é essa aí? Sim, *exponenciação* ou *potenciação*. Logaritmo é outro nome que se dá a um expoente e a operação que usamos para encontrar esse expoente é chamada de **logaritmo** (também! – como o resultado de um produto entre dois números é chamado de *produto* numa operação de multiplicação [outro nome para produto, mas, logaritmos não tem outro nome!]. Talvez a operação devesse ser chamada de *logaritmação*, mas, não é).
A operação de logaritmo é montada assim:

$$\log_A Y = X$$

Que se lê: *o logaritmo de Y na base A é igual a X*.

A **operação de logaritmos** retorna o expoente desconhecido (x) ao qual devo elevar a base conhecida (a) para obter o montante conhecido (y).
A **operação de potenciação** retorna o montante desconhecido (y) quando elevo a base conhecida (a) ao expoente conhecido (x).

Seja $2^x = 128$. Que operação podemos usar para encontrar o valor de x? Como não sabemos o valor de x, não dá para usar potenciação, claro. Talvez possamos usar radiciação: calculando a x-ésima raiz dos dois lados da equação, temos que

$$\sqrt[x]{2^x} = \sqrt[x]{128} \Rightarrow 2 = \sqrt[x]{128}$$

Não dá para resolver.

Sabemos que $y = a^x \Rightarrow y = \underbrace{a \times a \times a \times \ldots \times a}_{x\ vezes}$. O número x é o expoente, o valor procurado, o logaritmo. Para resolver, perguntamos: *Quantas vezes devo multiplicar a por ele mesmo para obter y*? Da mesma maneira, podemos perguntar: *Quantas vezes devo multiplicar 2 por ele mesmo para obter 128?* A resposta será x, e será obtida pela operação

$$\log_2 128 = x$$

Onde *log* é a operação; 2 é a base; 128 é o número a ser operado; x é o resultado procurado. Normalmente, se usa tábuas de logaritmos já calculados, mas, no nosso exemplo, sabemos que 2 produz 128 quando multiplicamos 2 por ele mesmo por 7 vezes. Assim, 7 é o logaritmo de 128 na base 2:

$$\log_2 128 = 7 \Rightarrow 2^7 = 128$$

As propriedades a seguir se aplicam a qualquer base.

- Quando a base não é informada, é assumida a base **10**.
- Quando se usa a base ***e*** (número neperiano) usa-se o operador ln, em vez de log.

$$1)\ \log(a \times b) = \log a + \log b$$
$$2)\ \log a \div b = \log a - \log b$$
$$3)\ \log x^n = n \times \log x$$

- Como $a^0 = 1$, concluímos, facilmente, que o logaritmo de **1** em qualquer base ***a*** é igual a **0**:

$$\log_a 1 = 0$$

- Sabemos que $a^1 = a$, então $\log_a a = 1$.

- Sabemos que $a^{-1} = \frac{1}{a}$, pois, pela propriedade da divisão de potências de mesma base (conserva-se a base e subtrái-se os expoentes), temos que:

$$\frac{a}{a^2} = a^{1-2} = a^{-1} = \frac{a}{a^2} = \frac{1}{a}$$

Usando-se a propriedade 2 ou 3, temos que $\boldsymbol{\log \frac{1}{a} = \log a^{-1} = -\log a}$.
As bases mais usadas, para as quais existem tábuas, são a base 10 e a base e, mas, é possível calcular logaritmos em qualquer base usando-se a expressão a seguir, que calcula o logaritmo em uma base ***a*** conhecendo-se uma base ***b***:

$$\log_a x = \frac{\log_b x}{\log_b a}$$

Se x for igual à base b: $\log_a x = \frac{1}{\log_x a}$, pois $\log_b b = 1$

Probabilidades

- **Evento** – Um fato, um acontecimento. Um valor ou uma mensagem que surge de um sistema ou de uma variável que pode assumir diversos valores é um evento.
- **Eventos Independentes** – Quando um evento que ocorre não altera o próximo evento que pode ocorrer. Por exemplo, lançar uma moeda duas vezes ou lançar duas moedas.
- **Eventos Dependentes ou Condicionais** – Quando um evento que ocorre sofre uma alteração causada por um evento anterior. Por exemplo, em uma urna com 2 bolas azuis e 3 bolas vermelhas, qual é a chance de tirar uma bola azul depois que tiramos uma bola vermelha?
- **Eventos Mutuamente Exclusivos** – Quando a ocorrência de um evento impede a ocorrência do outro. Por exemplo, ao tirar uma carta de uma pilha de baralho, essa carta não pode ser um rei e um ás. Ou é um ou é o outro, não tem como ser ambos. Por outro lado, a carta retirada pode ser um rei e uma carta de copas ao mesmo tempo, ou seja, estes dois últimos eventos não são mutuamente exclusivos.
- **Espaço de Amostragem** – É o conjunto de eventos possíveis, ou valores possíveis, que uma variável aleatória pode assumir.
- **Variável Aleatória** – Sistema ou variável que pode apresentar eventos sem que haja um controle que determine que evento será apresentado. O mesmo que sistema randômico ou variável randômica. Uma variável aleatória pode assumir um de vários valores do seu espaço de amostragem.

Se você joga uma moeda de 50 centavos para cima, qual é a chance dela mostrar *cara* para cima quando ela cair no chão?
Ora, como ela tem apenas 2 faces, uma com *cara* e a outra com *coroa*, a chance dela mostrar *cara* é de 1 possibilidade entre duas, ou seja, 50%, ou 50/100, ou 1/2. Igualmente, a chance dela mostrar coroa é de 1/2. A chance dela mostrar *cara* ou *coroa* é igual a 1/2 + 1/2 = 1, igual a 100%.

O espaço de amostragem é **{cara, coroa}**. *Cara* e *coroa* são os eventos possíveis. A variável aleatória é uma variável qualquer que pode assumir o valor *cara* ou o valor *coroa*.

Dado um evento A qualquer, a probabilidade de A ocorrer é definida assim:

$$P(A) = \frac{quantidade\ que\ A\ pode\ ocorrer}{total\ de\ eventos\ possíveis\ no\ conjunto\ em\ que\ A\ está}$$

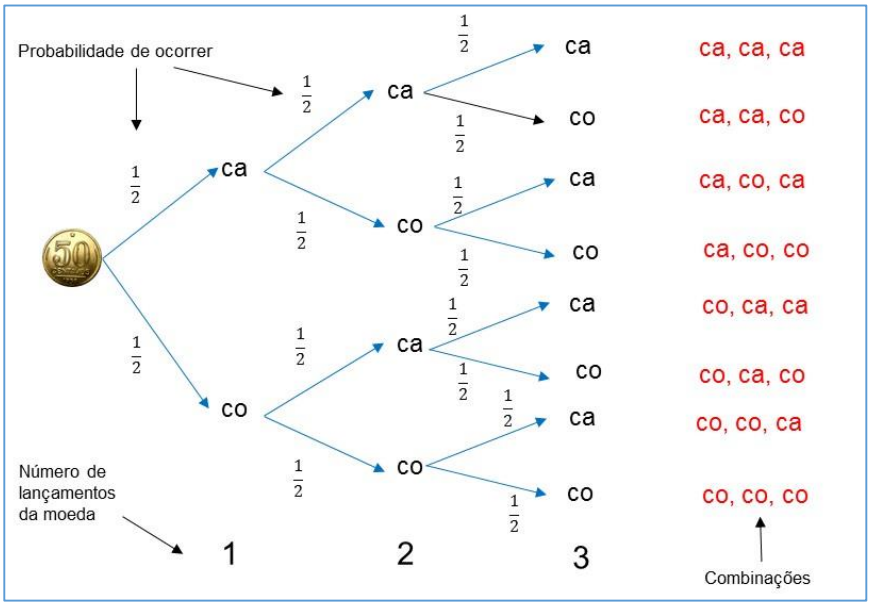
A probabilidade de um evento ocorrer mais a probabilidade de ocorrer todos os demais eventos do conjunto é sempre igual a 1. Isso implica que o somatório de todas as probabilidades de um espaço de amostragem é igual a 1.

O que implica também que um valor de probabilidade varia entre 0 e 1, o que significa que probabilidades podem ser escritas em termos de porcentagem: de 0% a 100%. Disso você pode concluir que não existe probabilidade negativa.

Eventos Independentes

Quando há **m** maneiras de fazer uma coisa A e **n** maneiras de fazer uma coisa B, então há **m x n** maneiras de fazer A junto com B, desde que A e B sejam eventos independentes (A não afeta B, e vice-versa).

Quais são as probabilidades de *cara* e *coroa* se você jogar uma moeda mais que uma vez? O jeito mais fácil de fazer o cálculo é visualizar o que acontece através de um diagrama de árvore:



Como os eventos são independentes (sair cara no primeiro lançamento não afeta sair cara de novo no segundo lançamento, ou coroa), a probabilidade de sair a sequência **ca**, **co**, **ca** é igual à multiplicação das chances de cada evento:

$$P = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

Disso, você vê que a probabilidade de ocorrer qualquer outra combinação é a mesma e igual a 1/8. Vê também que:

$$P(1\ cara) = \frac{1}{2}$$

$$P(2\ caras) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

E assim por diante.

- Quando dois eventos A e B são independentes, a probabilidade de ambos ocorrerem é

$$\boldsymbol{P(A \cap B) = P(A).P(B)}$$

P(*A* ∩ *B*) tem o mesmo significado de *P*(*A* e *B*).

No diagrama acima, qual é a probabilidade de obter pelo menos 1 cara em 2 lançamentos?

Ora, as combinações possíveis para dois lançamentos são: caca, caco, coca e coco. Nestas 4 combinações, temos cara em 3 delas, com a probabilidade de 1/2 cada uma.
A probabilidade de dar cara nestas 3 combinações é igual à soma das probabilidades de cada combinação sair. Como cada combinação tem 1/4 de chance:

$$P(1\ cara\ ou\ mais) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} = 0.75$$

E se você jogar duas moedas, quais são as chances? E se jogar 3? É a mesma coisa de jogar uma moeda 2 vezes ou 3 vezes. A independência continua. Porém, vamos ver de outra maneira e chegaremos ao mesmo resultado.

Duas moedas equivalem a uma moeda com 4 faces (não dá para visualizar, mas, dá para teorizar). Quer ver?
Suponha que A seja cara e B seja coroa. Então, a jogar as duas moedas, podem sair:

AA, **AB**, **BA**, **BB** (cara+cara, cara+coroa, coroa+cara, coroa+coroa)

É como se fosse um objeto de 4 faces em que cada face fosse: AA, AB, BA e BB. Cada uma das faces tem uma chance em quarto de ser mostrada: 1/4. A soma total representa a chance de qualquer face sair, ou seja, 1, ou 100%.

E se você jogar 3 moedas. O raciocínio será o mesmo:

AAA, **AAB**, **ABA**, **BAA**, **BBB**, **BBA**, **BAB**, **ABB**

Equivale a um objeto de 8 faces, e a chance de cada face será, portanto, de 1/8.

Se você continuar aumentando o número de faces, chegará à conclusão que jogar **N** moedas corresponderá a um objeto de $\mathbf{2^N}$ faces, e que a chance de cada face aparecer será igual a $\mathbf{1/2^N}$ (é de se esperar, pois, num conjunto com N objetos, a chance de você retirar um objeto qualquer é de 1 em N).

Podemos generalizar e dizer que, se um objeto tem **Qf** faces e você lança **Qo** objetos iguais a esse, a quantidade de faces que podem ser mostradas (as combinações possíveis) serão $\mathbf{Qf^{Qo}}$, e a chance de cada face aparecer será $\mathbf{1/Qf^{Qo}}$.

Você pode até calcular a chance de você acertar a sena da megasena considerando o volante como um objeto de 60 faces em que você lança 6 desses objetos. Você terá que subtrair as repetições (23, 27, 50, 23, 10, 33 não é uma sena válida) para chegar no valor correto da chance.

Veja que **Qf** é sempre maior que zero (não existe um objeto sem face – ele teria que ser invisível – o valor será 1 ou maior) e **Qo** pode ser igual a zero (só matematicamente) ou maior. Assim, não existe chance menor que zero, isto é, **não existe probabilidade negativa**. Não tem sentido existir.

<u>Eventos Mutuamente Exclusivos</u>

- Quando dois eventos são mutuamente exclusivos (se um ocorrer, o outro não ocorre), a probabilidade de A ou B (não ambos) ocorrer é igual à soma das probabilidades dos dois eventos:

  P(A ∪ B) = P(A) + P(B)

  O símbolo U significa “ou”.

Ao jogar um dado, qual a probabilidade de sair um 3 ou um 5?
Como um dado tem 6 faces, a chance de qualquer face é 1/6. Daí,

$$P(face3\ ou\ face5) = P(face3) + P(face5) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

Em um baralho de 52 cartas temos Rei e Ás. Ao tirarmos uma carta, temos o seguinte:

$$P(Rei) = \frac{4}{52} = \frac{1}{13}$$
$$P(\text{Ás}) = \frac{4}{52} = \frac{1}{13}$$

Qual é a probabilidade da carta retirada ser um Rei Ás?

Seja K o conjunto de todos os Reis (4 cartas) de um baralho e C o conjunto de todas as Copas (13 cartas, incluindo um Rei de Copas) do baralho. Seja A o conjunto de todos os Ases.

A intersecção do conjunto K com o conjunto A é vazia. Combinando A e K, qual é a probabilidade de, tirando uma carta, ela ser um Rei Ás? É impossível. Ou sai um Rei ou sai um Ás. São **eventos mutuamente exclusivos**. A probabilidade de K e A é nula:

P(K ∩ A) = 0

Qual é a probabilidade de sair um Rei ou um Ás? Como são eventos mutuamente exclusivos:

$$P(K \cup A) = P(K) + P(A) = \frac{1}{13} + \frac{1}{13} = \frac{2}{13}$$

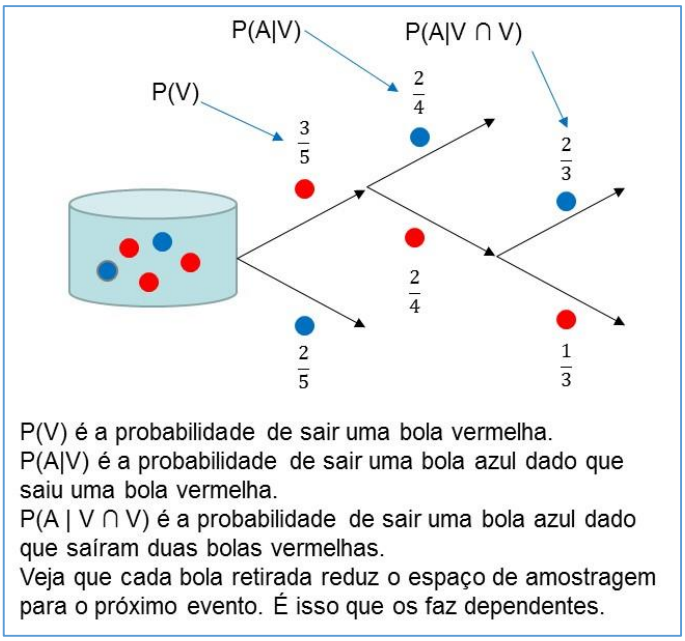
A intersecção do conjunto K com o conjunto C tem um único elemento, que é o Rei de Copas. Uma carta tirada de K e uma carta tirada de C são eventos **não mutuamente exclusivos**. A probabilidade de K e C não é nula: P(K ∩ R) > 0. Qual é a probabilidade de se retirar um Rei ou uma carta de copas do conjunto K + C?

Como o mesmo Rei de Copas aparece duas vezes, uma em cada conjunto, é necessário subtrair a probabilidade da interseção:

P(K U C) = P(K) + P(C) – P(K ∩ C)

Eventos Dependentes e Probabilidade Condicional (ou inversa)

Seja uma urna com 3 bolas vermelhas e 2 bolas azuis. Ao retirarmos uma bola e não a repormos na urna (dizemos que a experiência é **sem reposição**) isso vai alterar o espaço de amostragem e faz com que o próximo **evento** fique **dependente** do evento anterior. Se a experiência é feita **com reposição**, os **eventos** se tornam **independentes**.



P(V) é a probabilidade de sair uma bola vermelha.
P(A|V) é a probabilidade de sair uma bola azul dado que saiu uma bola vermelha.
P(A | V ∩ V) é a probabilidade de sair uma bola azul dado que saíram duas bolas vermelhas.
Veja que cada bola retirada reduz o espaço de amostragem para o próximo evento. É isso que os faz dependentes.

Probabilidade condicional é a probabilidade de um evento B ocorrer dado que um evento A já ocorreu.

Exemplos:

Foram dados dois testes de matemática para uma classe. 25% dos alunos passaram em ambos os testes e 46% passaram só no primeiro teste. Qual é a porcentagem dos que passaram no segundo teste tendo também passado o primeiro?

Uma urna X contém 10 bolas vermelhas e 1 bola branca, enquanto uma urna Y contém 10 bolas brancas e 1 bola vermelha. Qual é a chance de sair uma bola vermelha, dado que saiu a urna Y?

A probabilidade de um evento ***B*** ocorrer, dado que um evento ***A*** ocorreu (o evento B é dependente do evento A), denotada por ***P(B|A)*** ou PA(B), é dada por:

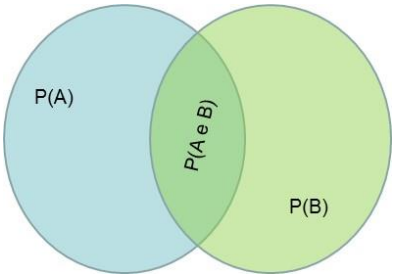
$$P(B|A) = \frac{P(A\ e\ B)}{P(A)}$$

O evento A tem que ocorrer, obrigatoriamente, pois P(A) não pode ser zero.

- Quando dois eventos <u>A e B são dependentes</u>, a probabilidade de ambos ocorrerem é

$$P(A \cap B) = P(A) \times P(B|A)$$

P(*A* ∩ *B*) significa P(A e B), ou seja, a probabilidade de A e B ocorrerem simultaneamente. Em teoria dos conjuntos, é o conjunto formado pela intersecção dos conjuntos A e B.



Dividindo os dois lados da equação acima por P(A), obtemos a equação da probabilidade condicional mais acima.
Para 3 eventos dependentes: P(A e B e C) = P(A)•P(B|A)•P(C|A∩B) e assim para N eventos dependentes.
O ponto (●) significa multiplicação (x).

No caso da probabilidade condicional em que os eventos A e B são independentes (tipo: *qual é a chance de tirar uma bola vermelha da urna X, dado que choveu ontem*?), a probabilidade de B ocorrer dado que A ocorreu é, exatamente, a probabilidade de B ocorrer:

P(A∩B) = P(B).

Vamos agora resolver os dois problemas iniciais de probabilidade condicional.

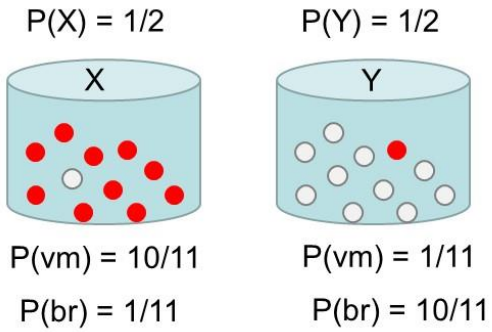
Foram dados dois testes de matemática para uma classe. 25% dos alunos passaram em ambos os testes e 46% passaram só no primeiro teste. Qual é a porcentagem dos que passaram no segundo teste tendo também passado o primeiro?

Evento A: Passaram no primeiro teste. P(A) = 0.46
Evento B: Passaram no Segundo teste.
Passaram no primeiro e no segundo: P(A ∩ B) = 0.25
Questão: P(B|A)?

Usando a formula da condicional, teremos que

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.25}{0.46} = 0.54, ou\ 54\%$$

Uma urna X contém 10 bolas vermelhas e 1 bola branca, enquanto uma urna Y contém 10 bolas brancas e 1 bola vermelha. Qual é a chance de sair uma bola vermelha, dado que saiu a urna Y?



Probabilidade de sair urna X e bola vermelha: P(X ∩ vm) = P(X).P(vm|X)

$$P(X \cap vm) = \frac{1}{2} \times \frac{10}{11} = \frac{10}{22}$$

Probabilidade de sair urna X e bola branca = 1/2 x 1/11 = 1/22
Probabilidade de sair urna Y e bola vermelha = 1/2 x 1/11 = 1/22
Probabilidade de sair urna Y e bola branca = 1/2 x 10/11 = 10/22
Probabilidade de sair bola vermelha = P(vm|X) + P(vm|Y) = 11/22 = 1/2

<u>Esperança</u>

Esperança é o valor médio esperado quando um experimento probabilístico é realizado muitas vezes. O cálculo da esperança é igual ao cálculo da média ponderada, sendo definido por:

$$E(X) = \sum_{i}^{n} x_i \times P(x_i)$$

Se todos os $\mathbf{x_i}$ têm a mesma probabilidade, *E(X)* se resume ao cálculo da média aritmética simples.

Por exemplo, no lançamento de um dado, o valor esperado será dado por:

$$E(X) = 1.\frac{1}{6} + 2.\frac{1}{6} + 3.\frac{1}{6} + 4.\frac{1}{6} + 5.\frac{1}{6} + 6.\frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$

As probabilidades de cada face são iguais, 1/6. Calculando a média aritmética simples:

$$M = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6}$$

Variância

A variância de uma variável aleatória nos diz o quanto os valores ficaram distantes (se desviam) do valor esperado (da esperança).
Seja v o valor esperado de uma variável aleatória X. A variância de X é dada por:

$$V(X) = E[(X - v)^2]$$

Ou seja, a variância é o valor esperado do desvio elevado ao quadrado.

Desvio Padrão

O desvio padrão mede a tendência dos valores em relação ao valor esperado. Ele igual à raiz quadrada da variância:

$$DP(X) = \sqrt{V(X)}$$

Um baixo DP indica que os dados tendem a estar próximos da média.

Moda

É o valor que aparece mais vezes num experimento aleatório. O valor mais comum entre todos. A moda não é, necessariamente, um valor único.

Redundância

Trechos repetindo o mesmo significado da mensagem, caracteres que forçam a mensagem a obedecer regras sintáticas, por exemplo, carregam redundância. A redundância faz parte de um esforço, mesmo inconsciente, para que a informação fique imune ao ruído do canal. Quanto maior a redundância, menor a chance do ruído corromper a mensagem. Como o ruído é analógico, ele se mistura à mensagem, corrompendo-a. Mesmo que se amplifique a mensagem, não há como amplificar apenas ela. O ruído é amplificado também, não resolvendo o problema.

A solução que se adota é recuperar (reler) a mensagem digital pelo caminho ante de amplificá-la. Com isso, ela acabava ficando mais forte do que o ruído e podia ser lida na ponta final. Os dispositivos usados para isso são os repetidores regenerativos.

Redundância aumenta o consumo da banda, reduzindo a quantidade de mensagens a serem enviadas na mesma janela de tempo. O código ideal seria aquele que tem redundância mínima com correção máxima.

Qualquer linguagem falada ou escrita contém redundâncias. Estas redundâncias nasceram com o uso, não foram colocadas de proposito na língua, se bem que, às vezes, costumamos reforçar uma expressão tornando-a redundante.
Estamos tão acostumados com a redundância que, quando ela é removida da frase, ainda assim conseguimos preencher os vazios mentalmente. Veja esta frase:

vc cnsg lr st frs?

Quantificando a Informação

Para quantificar uma informação, Shannon usa um conceito de bit (binary digit) um tanto diferente do conceito conhecido na Informática. A diferença fica por conta mais de um método matemático que é usado para a quantificação: o logaritmo de base 2, como veremos a seguir.

Capacidade de (Armazenar) Informação

Suponha que um sistema tem apenas dois estados possíveis (é um *sistema binário*). Por definição, a capacidade de informação de tal sistema é 1 bit (que é capaz de armazenar um valor ou seu oposto definido, mas, não ambos ao mesmo tempo).

Generalizando: **Se o sistema tem 2^b estados possíveis, sua capacidade é *b bits***.
Se fizermos $N = 2^b$, então, pelo que aprendemos de logaritmos:

$$\boldsymbol{b = \log_2 N} \qquad \boldsymbol{(1)}$$

Usando a propriedade para transformação entre bases, podemos, também, escrever que

$$b = \frac{\log_e N}{\log_e 2} = \frac{\ln N}{\ln 2} \qquad (2)$$

Essa fórmula é útil para calcularmos logaritmos na base 2 em calculadoras eletrônicas. Caso a calculadora só tenha a tecla ***log***, você pode usar a transformação para a base 10:

$$b = \log_2 N = \frac{\log N}{\log 2} \qquad (3)$$

Como você pode ver, basta mudar ***ln*** para ***log*** na fórmula anterior.

Podemos definir *capacidade de informação*: **Capacidade de informação é a quantidade de bits necessários para armazenar todos os estados possíveis de um sistema**.

Exemplo: A quantidade de estados possíveis de um dado é 6. Assim, sua capacidade de informação é:

$$\log_2 6 = 2.5849 \ldots bits$$

Não deu um número inteiro de bits porque 6 não é uma potência de 2, porém, matematicamente, esta é a quantidade de bits necessários para armazenar todos os estados possíveis de um dado. Usar 3 bits seria um desperdício de espaço (ainda, matematicamente falando).

Se o sistema tem apenas **1** estado possível (uma moeda com duas caras, ou com duas coroas, tem apenas 1 estado possível – **você concorda**), então sua capacidade de informação é **0** (zero bit!), pois, você sabe que o logaritmo de 1 (**em qualquer base**!) é igual a zero.

Um sistema cuja capacidade de informação é nula, não traz nenhuma informação (nova). Mas, por que *nova*? Ora, imagine uma moeda que tem duas caras (você sabe que ela tem duas caras). Se ela for lançada, você já sabe qual será o resultado, ou seja, você não estará ganhando uma informação que você não tinha. Você estará ganhando zero de informação.
Veja, pegue o sentido de *informação* da *teoria da informação*: não é o que você já sabe, mas, o que você ganha (ou ainda não sabe).

A moeda gira, cai no chão e te mostra uma cara. Apesar de você chamar a apresentação dessa cara de informação (afinal, você a está recebendo), essa *informação* que foi entregue não preencheu um vazio no teu "banco" de conhecimento (o lugar estava ocupado – a *informação* já havia sido recebida num momento passado qualquer). Com isso você concorda. Com isso você consegue entender o que é, realmente, informação Shannon.

O conceito é muito escorregadio, está sempre querendo fugir da nossa mente. Talvez isso se deva ao fato de usarmos a mesma palavra (informação) para duas coisas diferentes que nem são opostas e nem são complementares. Na verdade, uma se adiciona à outra.
O que podemos dizer é que, quando uma informação (Shannon) chega, ela não mais será informação Shannon. Será conhecimento. Se a mesma informação chegar de novo, é como se ela tivesse vindo do seu próprio banco de conhecimento, não sendo, portanto, informação.

Veja de novo: aquilo que você já sabe, aquilo que é certo, **aquilo que tem 100% de probabilidade de acontecer, não é informação**.
Com isso, podemos dizer que, **quanto mais alta for a probabilidade de ocorrer, menor será a quantidade de informação sendo entregue**.
Da mesma maneira, podemos dizer: **Quanto mais baixa for a probabilidade de ocorrer**, quanto mais rara for uma mensagem qualquer que você possa receber, **mais informação** (a mensagem) **conterá.**
Podemos representar isso matematicamente: Seja **Q** a quantidade de informação (útil) e **P** a probabilidade da mensagem ocorrer. Então:

$$Q = \frac{1}{P} \qquad (4)$$

Veja que se P é pequeno, Q é grande (se a probabilidade for baixa, mais informação será obtida) e vice-versa.

Algumas Clarificações

- **Sistema** – Conjunto de elementos que se relacionam com um objetivo final, uma função. Uma moeda é um sistema. Um dado é um sistema. Uma máquina é um sistema. Um sistema pode ser representado por uma letra (X, por exemplo) que pode assumir mais que um valor.
- **Dado** – Qualquer coisa capaz de carregar informação ou de que se pode derivar informações. Todo dado carrega uma mensagem.
- **Mensagem** – Qualquer coisa que se entrega, que comunica, que "fala algo".

Qualquer sistema contém dados, e assim pode passar uma mensagem, uma informação.

SISTEMA → DADO → MENSAGEM → INFORMAÇÃO

Poder passar informação implica em probabilidade de armazenar e passar a informação. Probabilidade é uma medida, um número, podendo, portanto, sofrer operações matemáticas.

Informação é, normalmente, definida como uma mensagem formada de símbolos entregues por uma fonte.

- ✓ Uma moeda lançada é uma fonte que entrega cara (A) ou coroa (B), e uma mensagem com estes símbolos pode ser: **BBABABBB**.
- ✓ Da mesma maneira, os símbolos 1,2,3,4,5,6 podem ser emitidos por uma fonte que é um dado de 6 faces.
- ✓ As letras A, C, G e T são emitidas por uma fonte DNA.

Se um sistema S consiste de dois subsistemas discretos A e B independentes (podem se manifestar juntos ou separados), então a capacidade de S é a soma das capacidades de A e B.

Vimos em probabilidades que, se **A** e **B** são dois eventos independentes, então, a probabilidade de **A** e **B** ocorrerem é dada por: **P(A ∩ B) = P(A) x P(B)**. E, vimos em logaritmos que *log*AB = *log*A + *log*B. Assim, a capacidade de informação de S é dada por:

$$P(S) = P(A \cap B) = \log_2 A \times B = \log_2 A + \log_2 B$$

Quantidade de Informação

Reconsidere a lâmpada que, ao meio-dia, pode estar acesa ou apagada, conforme esteja ensolarado lá fora ou não.
Embora a *capacidade de informação* desse sistema seja 1 bit (2 estados: acesa ou apagada), a notícia de que a lâmpada está acesa não traz muita informação. Por outro lado, uma lâmpada que indica se está chovendo lá fora ou não parece fornecer uma quantidade maior de informação, muito embora sua capacidade de informação seja de 1 bit também.

Esclarecendo a diferença entre os dois sistemas com uma lâmpada

Com base no que vimos sobre informação já conhecida, a lâmpada indicando sol lá fora acrescenta pouca informação porque ocorre muito mais períodos de sol do que de chuva. Então, você espera mais sol do que chuva. A informação de que está ensolarado não te surpreende tanto quanto a informação de que está chovendo. Chover é uma coisa mais rara do que fazer sol. A probabilidade de chover é bem menor do que a probabilidade de fazer sol. Imagine um lugar onde nunca choveu na história e alguém lá vê a lâmpada indicadora acender. Seria surpreendente. Um evento raro, de baixíssima probabilidade. Daqui você vê que quanto mais raro for um evento (quanto menor for sua probabilidade de ocorrer), mais informação ele acrescenta.

O fato de um sistema fornecer mais ou menos informação implica que nem toda a capacidade de informação do sistema pode ser utilizada. Diferentes quantidades de informações cabem no mesmo "container", ou seja, uma parte do container pode ficar vazia (quando uma parte da informação já é esperada, quando sua probabilidade é alta). Só podemos conseguir determinar o máximo, nunca a quantidade exata.

Podemos concluir o seguinte:

- ✓ Probabilidade alta → Pouca informação e vice-versa.
- ✓ Probabilidade baixa → Muita informação e vice-versa.

Qual é a probabilidade de uma moeda de duas caras te mostrar cara? O que ela vai te dizer além do que você já sabe? Qual é a probabilidade dela te mostrar coroa? E se ela mostrar?

Suponha que **X** seja uma variável aleatória que pode assumir um certo valor v (note que valor não precisa ser um número). **A quantidade de informação, Q, em bits, trazida pela notícia "o valor de X é v" é**, por definição:

$$Q(X = v) = \log_2\left(\frac{1}{P(X = v)}\right) \qquad (5)$$

Veja que o termo entre parêntesis do lado direito da equação 5 é, exatamente, o lado direito da equação 4.
Usando a propriedade de logaritmo da divisão, ou logaritmo de potência negativa, no segundo termo da equação (5), teremos que:

$$\boldsymbol{Q(X = v) = -\log_2 P(X = v)} \qquad \boldsymbol{(6)}$$

Esse valor, como capacidade de informar, é medido em bits e nunca é negativo, pois, o valor de P é, no máximo, igual a 1, o que daria Q = *log*1 = 0 (certeza absoluta implica informação zero!). Por outro lado, P não pode ser zero (como se pode ver na equação 4 – também sabemos que o logaritmo de zero não existe). P estando no intervalo (0, 1], *log*P será um número negativo que será transformado para positivo pelo sinal negativo na equação 6.

Em particular, se X pode assumir n valores distintos com **igual probabilidade** (a frequência de distribuição dos valores é uniforme, isto é, têm a mesma chance de ocorrer) $P(X = v) = \frac{1}{n}$, a quantidade de informação que recebemos quando ficamos conhecendo o valor de X (qualquer valor de X) é, exatamente:

$$\boldsymbol{Q(X = v) = \log_2 n \ bits} \qquad \boldsymbol{(7)}$$

ou seja, a capacidade da variável X.
Para verificar, basta usar a equação 6:

$$Q(X = v) = -\log_2 \frac{1}{n} = -\log_2 n^{-1} = (-1).(-\log_2 n) = \log_2 n$$

Esse resultado é a equação 1 da capacidade de (armazenar) informação.

Porém, se entre os n valores que X pode assumir, alguns tiverem probabilidades diferentes dos demais (e até entre si), a quantidade de informação poderá ser menor ou maior, dependendo do valor que ocorreu (e usamos a equação 6):

> Suponha que um dado está para ser lançado e X uma variável que vale 100 se o dado mostrar 1 e vale 200 se o dado mostrar qualquer outro valor. Então, as notícias *X é 100* e *X é 200*, carregam as seguintes quantidades de informação:
>
> $$Q(X = 100) = -\log_2 P(X = 100) = -\log_2 \frac{1}{6} = \log_2 6 = 2.5849 \ldots$$
> $$Q(X = 200) = -\log_2 P(X = 200) = -\log_2 \frac{5}{6} = -(\log_2 5 - \log_2 6) = 0.26 \ldots$$
>
> Note que a notícia *X é 200* traz bem menos informação do que a notícia *X é 100*, porque tem probabilidade maior (5/6 → qualquer face diferente de 1). *X é 100* traz 2.5 bits de informação, apesar de X ter apenas 2 estados possíveis e, portanto, apenas 1 bit de capacidade.

X é 200, com 0.26 bit (que já seu máximo), traz pouca informação. Se colocarmos a mensagem em 1 bit inteiro, a diferença será redundância.

Reforço

Para reforçar o conceito de informação Shannon, tentando diferenciá-lo do conceito corriqueiro, de vez em quando vou colocar um exemplo:

Se teu cachorro está latindo lá fora, fazendo um *AU*! por segundo, você vai esperar mais *AUs* nos próximos segundos, pois, a probabilidade de um AU sair é grande. Você, praticamente, sabe que ele virá e, se vier, não vai te surpreender, não vai te trazer informação útil.
Porém, se no lugar de um *AU*! vier um *CAIN*!, que você não espera (pois, teu cachorro está no jardim, num lugar seguro), que tem baixa probabilidade de ocorrer, você vai se surpreender, aquilo vai chamar tua atenção, vai te trazer informação nova (mais do que se tivesse saído um *AU*!).

Se 1 bit é suficiente para conter a informação *X é 100*, por que o cálculo mostrou 2.5 bits? Exatamente porque a distribuição não era uniforme. Para corrigir isso é preciso considerar o que Shannon chama de entropia.

Quantidade Esperada da Informação (Entropia)

O que a equação 7 mostra é que a informação de qualquer valor (ou símbolo de uma mensagem) é igual a $\log_2 n$, e esse valor é, também, a média.

$$Se\ você\ tem\ N\ notas, a\ média\ será\ M = \frac{a_1 + a_2 + \cdots + a_n}{N}.$$
$$Se\ cada\ nota\ é\ igual\ a\ \frac{1}{n}, então\ M = \frac{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}{N} = \frac{1}{N}$$

O nome formal para a informação média por símbolo é ENTROPIA, denotada pela letra H, como na entropia da Física.

Quando as **probabilidades** forem **diferentes** (a distribuição não for uniforme, os símbolos não tiverem a mesma probabilidade), então será necessário calcular a média ponderada (esperança), usando a probabilidade de cada símbolo como peso contra a média de cada símbolo dada pela equação 6 (a média final é o somatório das médias):

$$H(X) = \sum_{v}^{n} P(X = v).Q(X = v) \qquad (8)$$

Veja que a equação 8 é, exatamente, a fórmula de cálculo da esperança *E(X)*, ou valor médio esperado, de um experimento aleatório (ver quadro *Probabilidades*)

Usando o valor de Q do lado direito da equação 6:

$$\boldsymbol{H(X) = -\sum_{v}^{n} P(X = v).\log_2 P(X = v)} \qquad (9)$$

Onde ***v*** pode assumir todos os ***n*** valores possíveis para a variável ***X***. Cada termo do somatório é a quantidade de informação trazida pela notícia *X é v* multiplicada pela *probabilidade de recebermos* essa notícia.

Para o exemplo anterior, teremos então:

$$H(X) = -[P(X = 100).\log_2 P(X = 100) + P(X = 200).\log_2 P(X = 200)] \Longrightarrow$$
$$\Rightarrow H(X) = -\left[\frac{1}{6}.\log_2 \frac{1}{6} + \frac{5}{6}.\log_2 \frac{5}{6}\right] = -\left[-\frac{1}{6}.\log_2 6 + \frac{5}{6}.(\log_2 5 - \log_2 6)\right] \Rightarrow$$
$$\Longrightarrow H(X) = -\left[-\frac{2.5849}{6} + \frac{5}{6}.(-0.263)\right] \Longrightarrow H(X) = \frac{2.5849 + 1.315}{6} = 0.65$$

Assim, a quantidade esperada de informação que ganharemos ao saber o valor de X é cerca de 0.65 bits, abaixo da capacidade de X, que vimos ser de 1 bit. Isto nos permite concluir que:

> **Se uma variável aleatória X pode assumir N valores distintos, então, a quantidade esperada de informação que ganhamos conhecendo o valor de X é menor ou igual à capacidade de X (lg_2N). É exatamente igual a lg_2N apenas quando todos os valores de X forem igualmente prováveis (= 1/N).**
>
> *Quando há mais probabilidade, haverá menos informação; quando as probabilidades se igualam, a quantidade de informação se iguala à largura do canal.*

Fica claro também que o valor de H(X) varia entre 0 e lg_2N.
Se P(X=v) = 1 para um dado x, significa que P(X=v) = 0 para os demais x, implicando que H(X) = 0, 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 nova.

Probabilidades diferentes implicam um desvio nas informações. Dados com desvio carregam menos informações do que dados sem desvio, por isso ocupam menos bits. Em casos em que o desvio é muito forte, como no caso de uma moeda viciada que, em 1000 tacadas mostre 1 cara e 999 coroas, então, basta transmitir apenas as informações das caras, pois, a informação das coroas seria a diferença.

REVISÃO

- ✓ Capacidade de informação de uma mensagem:

$$b = \log_2 N$$

- ✓ Quantidade de informação para mensagens igualmente prováveis:

$$q = \log_2 N \ \ bits$$

Quantidade de Informação equivale a Capacidade de Informação

- ✓ Quantidade de informação para mensagens com probabilidades diferentes:

$$H(X) = -\sum_{i}^{n} P_i . \log_2 P_i$$

- ✓ Quantidade de informação = nível de incerteza = entropia transmitida
- ✓ Informação shannon é aquilo que se acrescenta ao teu conhecimento.

Mas, só ao meu conhecimento ou ao conhecimento de todos? Afinal de contas, sei mais que muita gente e muita gente sabe mais do que eu. O que está errado aqui?

Se o teu cachorro fez cain e você recebeu mais informação do que se ele tivesse feito au-au, supondo que uma pessoa atirou uma pedra nele, essa pessoa, certamente, esperaria o cain em vez do au-au. Assim, você e ela não receberam a mesma informação. O que foi informação para você não foi para ela, e vice-versa. E agora?
O problema, novamente, é de entendimento do que é informação shannon.
Informação shannon não é a informação no sentido comum.
Informação, no sentido de Shannon, é uma quantidade absoluta, que tem o mesmo valor numérico para qualquer observador, seja você ou seja a pessoa que atirou a pedra no teu cachorro. O valor que cada um dá para a informação é, necessariamente, uma quantidade relativa e, assim, terá valores diferentes para diferentes observadores.
Para Shannon, não importa o significado da mensagem ou o que se vai fazer com ela (a sua utilidade). Isso é relativo.
Dissociando o significado e a utilidade de uma mensagem que soa diferente para duas pessoas, a quantidade de informação recebida por cada pessoa é a mesma, seja em um punhado de letras embaralhadas, seja numa página de um jornal onde esse mesmo punhado de letras forma uma notícia.

Arranjar uma sequência de letras em uma frase om significado é por restrições na mensagem, o que diminui a quantidade de informação. As letras embaralhadas provêm mais informação shannon do que quando arranjadas em um significado (você vai prever, o que implica probabilidade alta de ocorrer, o que implica quantidade baixa de informação).

O sentido de entropia dado por Shannon é que quanto mais informação, quanto mais sua quantidade aumenta, mais difícil fica para o receptor receber a mensagem que ele espera. Esse aumento de informação no espaço de amostragem de onde o receptor espera receber a mensagem tem o mesmo efeito de um aumento de desordem ou entropia (como se diz na Termodinâmica). A incerteza de receber a mensagem aumenta. Se há ordem, fica muito mais fácil para você fazer uma previsão acertada. Se há desordem, a probabilidade de acerto diminui muito. A sua escolha se dilui entre tantas outras que se ajuntam ao espaço de amostragem.

Não é o que você absorve, entende, é o que você pode absorver ou não. Nesse sentido, todos ficam expostos à mesma quantidade de informação.

Exemplos Vários

1. Se a fonte pode enviar apenas uma mensagem, a chance que o receptor tem de recebe-la é de 100% (1 em 1, ou 1/1). Em outras palavras, a incerteza do receptor é zero.

$$H(X) = -1 \times \log_2 1 = 0$$

   Zero bits de informação shannon, que corresponde a zero bits de incerteza.

2. Se a fonte pode enviar duas mensagens, então ambas têm a mesma chance de serem enviadas. O receptor terá 1 chance em 2 (1/2) de receber a mensagem que ele espera. A incerteza aumentou, em relação ao caso anterior.

$$H(X) = -\left(\frac{1}{2} \times \log_2 2^{-1} + \frac{1}{2} \times \log_2 2^{-1}\right) = \frac{1}{2} \times 1 + \frac{1}{2} \times 1 = 1$$

   Temos 1 bit de informação shannon, que corresponde a 1 bit de incerteza, ou 1 bit de entropia shannon.

   Como as probabilidades são iguais, poderíamos ter usado a equação 6 (ou 7):

$$Q(X) = -\log_2 \frac{1}{2} = -\log_2 2^{-1} = \log_2 2 = 1$$

3. Se a fonte pode enviar dez (de 1 a 10) mensagens, todas com a mesma probabilidade de serem enviadas, a chance do receptor receber qualquer uma delas é de 1/10. A incerteza aumentou mais ainda.

$$Q(X) = -\log_2 10^{-1} = \log_2 10 = 3.3219 \ldots$$

4. Se de duas respostas, a resposta A tem chance 1/4 de aparecer e a resposta B tem chance 3/4, então:

$$H(A) = Q(A) = -\log_2 \frac{1}{4} = 2\ bits$$

$$H(B) = -\log_2 \frac{3}{4} = -(\log_2 3 - \log_2 4) = \log_2 4 - \log_2 3 = 2 - 1.58 = 0.42\ bits$$

   Podemos agora efetuar:

$$\frac{1}{4} \times 2 + \frac{3}{4} \times 0.42 = 0.500 + 0.315 = 0.815$$

   Fiz passo a passo, em vez de usar a equação 9.

5. Um DNA normal lança a sequência A, C, T, G com as mesmas probabilidades.

   Podemos usar a equação 7 diretamente:

$$H(X) = \log_2 4 = 2 \ \ bits.$$

6. Um DNA alterado lança 90% de A ou T e 10% de C ou G.

   Se sai A com 90%, não sai T com 90% e vice-versa. Se sai C com 10%, não sai G com 10% e vice versa. Então cada dois são eventos mutuamente exclusivos. Daí

$$P(A \cup T) = \frac{9}{10} = P(A) + P(T)$$

$$P(C \cup G) = \frac{1}{10} = P(C) + P(G)$$

   Como cada dois são igualmente prováveis, então:

$$P(A) + P(T) = \frac{4.5}{10} + \frac{4.5}{10} = \frac{9}{10}$$

$$P(C) + P(G) = \frac{0.5}{10} + \frac{0.5}{10} = \frac{1}{10}$$

$$H(X) = -[2.(0.45 \times \log_2 0.45) + 2.(0.05 \times \log_2 0.05)] \Longrightarrow$$
$$\Longrightarrow H(X) = -[2.(0.45 \times -1.152) + 2.(0.05 \times -4.321)] \Longrightarrow$$
$$\Longrightarrow H(X) = -[2 \times -0.5184 + 2 \times -0.21605] \Longrightarrow$$
$$H(X) = -[-1.0368 - 0.4321] = 1.0368 + 0.4321 = \mathbf{1.47\ bits}$$

Compressão de Dados

Foi só após a divulgação da Teoria da Informação de Shannon que foram desenvolvidos os sistemas de compressão de dados usados hoje. Basicamente, existem dois tipos de sistemas:

- Sem Perda (lossless) – Em que a compressão é feita através da remoção de redundâncias no texto. Por exemplo, trocar JOSÉ por ZÉ, QU por Q, etc. O texto original pode ser recuperado facilmente. A mensagem carrega a mesma quantidade de informação que a mensagem não compactada, só que com menos caracteres. Ela tem mais informação (maior entropia) por caracter do que a mensagem não comprimida, por causa da redundância reduzida.
- Com Perda não prejudicial (lossy) – Na compactação de imagens (JPG, por exemplo) e áudio (MP3, por exemplo), alguns pixels aos quais os olhos são insensíveis e sons que o ouvido não capta, são removidos do arquivo original. Não há como recuperar o arquivo original a partir do arquivo compactado.

Brasilio – Agosto/Setembro 2014.

Bibliografia

1. Artigos da Internet
2. Probabilidades e Teoria da Informação – PDF da UNICAMP