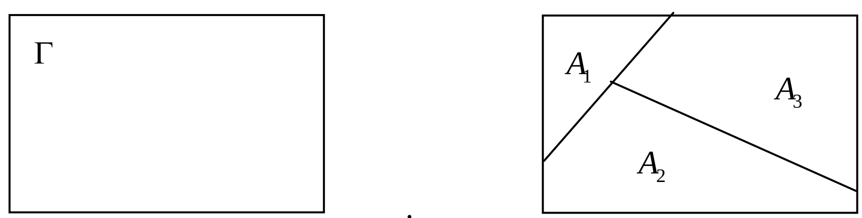
Entropy

Following Gibbs, this section defines entropy as a concept within the theory of probability, rather than a physical quantity following a long line of historical accidents. As such, entropy is a mathematical object independent of the concept of energy and temperature. It is just as sensible to talk about the entropy of rolling a die as talking about the entropy of diamond.

A dissection of a sample space. An experiment has many possible outcomes, each called a sample point. The set of all possible outcomes of the experiment is called the sample space. A subset of possible outcomes is called an event.

Denote the sample space of an experiment by Γ. By a *dissection* of the sample space is meant a class $\mathcal{A} = \{A_1, A_2, ..., A_i, ...\}$ of disjoint events whose union is the sample space. A special case is to use individual sample points to dissect the sample space.



A sample space A dissection

Denote the probability of an event A_i by $P(A_i)$. Clearly the probabilities of all the events in the dissection $\mathcal{A}$ sum to unity:

$$P(A_1) + P(A_2) + ... + P(A_i) + ... = 1.$$

Entropy of a dissection of a sample space. Define the *entropy S* of the dissection $\mathcal{A}$ of a sample space Γ by

$$S(A_1, A_2, ...) = -\sum P(A_i) \log P(A_i).$$

The sum is taken over all the events in the dissection $\mathcal{A}$. As a convention we will use the logarithm of the natural base. Because $x\log x \to 0$ as $x \to 0$, we set $0\log 0 = 0$. For any event A_i, $0 \le P(A_i) \le 1$, so that the number $-\log P(A_i)$ is nonnegative.

To calculate the entropy, we need a sample space, a probability measure defined on the sample space, and a dissection of the sample space. The physical nature of the experiment is irrelevant, be it tossing a coin, rolling a dice, or heating a pile of molecules. Much confusion surrounding entropy disappears if we explicitly state the sample space and the dissection.

When we dissect the sample space into individual sample points, the entropy is

$$S(\gamma_1, \gamma_2, ...) = -\sum_i P(\gamma_i) \log P(\gamma_i).$$

Entropy interpreted as the mean surprise. Historically, the definition of entropy was motivated in several contexts. We will describe one such a context. If you find the motivation contrived, you are not alone. You may as well accept the definition, and move on to learn the properties of entropy. After all, it is these properties that make entropy useful.

Let's say we wish to find a measure of the surprise evoked when an event occurs. It is very surprising when an event of small probability occurs. Naturally the surprise may be quantified by a decreasing function of the probability. Let *p* be the probability of the event, and $f(p)$ be the surprise evoked when the event occurs. When *p* = 1, the occurrence of the event evokes no surprise, so that we may specify $f(1) = 0$. When *p* = 0, the occurrence of the event evokes a great surprise, so that $f(0) = +\infty$. These are all we can say about the function $f(p)$ if we only consider a single event.

Next we consider two independent events with probabilities *p* and *q*, respectively. The probability for both events to occur is *pq*, so that the joint occurrence evokes a surprise of $f(pq)$. We stipulate that the surprises evoked by two independent events be additive, namely,

$$f(pq) = f(p) + f(q).$$

To satisfy this equation for all numbers *p* and *q*, the function must be $f(p) = -k \log p$, where *k* is a positive constant independent of the probability. A choice of the number *k* amounts to prescribing a unit for the measure of surprise; as a convention we choose *k* = 1. Thus, we assign the surprise evoked by the occurrence of an event of probability *p* by

$$f(p) = -\log p.$$

Given a dissection of the sample space, $\mathcal{A} = \{A_1, A_2, ..., A_i, ...\}$, the occurrence of event A_i evokes surprise $f(P(A_i)) = -\log P(A_i)$. The mean surprise defines the entropy of the dissection:

$$S(A_1, A_2, ...) = -\sum P(A_i) \log P(A_i) = \langle -\log P(A_i) \rangle.$$

The entropy quantifies uncertainty about the outcomes of an experiment.

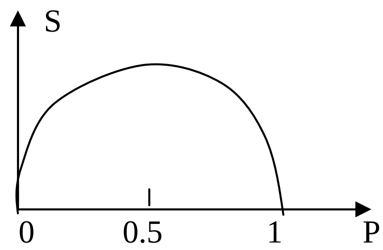
Examples. *(a) A fair die.* For a given sample space and a given probability measure, the entropy depends on how the sample space is dissected. Consider several dissensions of the sample space of rolling a fair die once. First, for a dissection that regards every face as an event, the entropy is $\log 6 = 1.792$. Second, for a dissection that regards all the odd numbers as one event, and all the even numbers as another event, the entropy of this even-odd dissection is $\log 2 = 0.693$. Third, if the sample space is dissected into two events $A_1 = \{1,2,3,4,5\}$ and $A_2 = \{6\}$, the probabilities are $P(A_1) = 5/6$ and $P(A_2) = 1/6$, and the entropy of the dissection is $-(5/6)\log(5/6) - (1/6)\log(1/6) = 0.430$.

(b) Dissecting a sample space into Ω *equally probable events.* The probability of each event is $P(A_i) = 1/\Omega$, and the entropy of the dissection is

$$S = \log \Omega.$$

The more (equally probable) events a sample space is dissected into, the larger the entropy.

(c) Dissecting a sample space into two events. When a sample space is dissected into two events with probabilities P and $1-P$, respectively, the entropy of the dissection is



$$S = -P\log P - (1-P)\log(1-P).$$

The entropy depends on how the sample space is dissected. When the sample space is dissected into one event with probability 1 and the other event with probability 0, the entropy of the dissection vanishes. Of all possible ways to dissect the sample space into two events, the dissection resulting in two equally probable events maximizes entropy. More generally, of all dissections consisting of a fixed number of events, the dissection consisting of equally probable events maximizes the entropy.

(d) Entropy of mixing. A mixture of N_B black marbles and N_W white marbles are placed in $N_B + N_W$ slots. We regard all possible arrangements as a dissection, and assume that every arrangement is equally likely. The number of arrangements is

$$\Omega = \frac{N_B! N_W!}{(N_B + N_W)!}.$$

The entropy of dissection is $S = \log\Omega$. For large numbers, recalling Stirling's formula $\log x! = x\log x - x$, we obtain that

$$S = -N_B \log\frac{N_B}{N_B + N_W} - N_W \log\frac{N_W}{N_B + N_W}.$$

Let the fraction of black marbles be $c = N_B/(N_B + N_W)$. The entropy per marble is

$$\frac{S}{N_B + N_W} = -c\log c - (1-c)\log(1-c).$$

(e) Entropy of a substance. A substance switches between many quantum states. If we regard the set of all quantum states as a dissection, the entropy of this dissection is known as the entropy of the substance. As we will see later, this entropy can be measured experimentally.

Statistically independent experiments. Consider two statistically independent experiments, such as rolling a die and throwing a coin (or quantum states in a glass of wine and quantum states in a piece of cheese). Each experiment generates a distinct sample space. Let $\mathcal{A} = \{A_1, A_2, ..., A_i, ...\}$ be a dissection of one sample space, and $\mathcal{B} = \{B_1, B_2, ..., B_j, ...\}$ be a dissection of another sample space. Each event in dissection $\mathcal{A}$ is *statistically independent* of each event in dissection $\mathcal{B}$, namely,

$$P(A_i B_j) = P(A_i)P(B_j).$$

We next regard the combination of the two experiments as an experiment, and denote $\mathcal{AB}$ as the class of events $\{A_1B_1, A_2B_1, ..., A_iB_j ...\}$, which consist of a dissection of the sample space of the composite experiment. The entropy of this dissection is

$$S(\mathcal{AB}) = -\sum_{i,j} P(A_i B_j)\log P(A_i B_j).$$

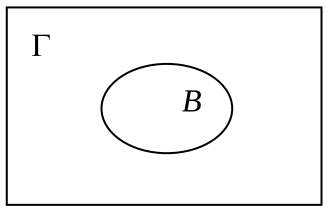
We can confirm that

$$S(\mathcal{AB}) = S(\mathcal{B}) + S(\mathcal{A}).$$

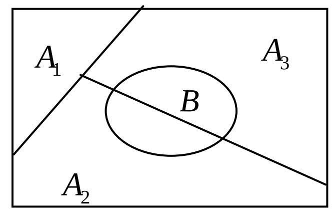
The joint entropy of two independent dissections is the sum of the entropies of the individual dissections.

Conditional entropy. Let $\mathcal{A} = \{A_1, A_2, ..., A_i, ...\}$ be a dissection of a sample space Γ, and *B* be an event in Γ. If we view *B* as a sample space, the class of events $\{A_1B, A_2B, ..., A_jB ...\}$ is a dissection of *B*. The probability of event A_i conditional on event *B* is defined by

$$P(A_i|B)=\frac{P(A_iB)}{P(B)}.$$



A sample space and an event



A dissection

Note that $BA_1, BA_2, ..., BA_i ...$ are mutually exclusive with one another, so that

$$P(B)=\sum_i P(A_iB),$$

and

$$\sum_i P(A_i|B)=1.$$

When the event B coincides with the sample space Γ, we may call $P(A_i|\Gamma)$ the absolute probability, and drop the reference to Γ.

The conditional entropy of the dissection $\mathcal{A}$ given that event B occurs is

$$S(\mathcal{A}|B)=-\sum_i P(A_i|B)\log P(A_i|B).$$

Example. *A fair die.* We dissect the sample space of rolling a die once as $A_1=\{1,2,3\}$, $A_2=\{4,5\}$, and $A_3=\{6\}$. Event B is "the face is an even number". Thus, $A_1B=\{2\}$, $A_2B=\{4\}$, and $A_3B=\{6\}$. The probabilities are

$$P(B)=1/2$$

$$P(A_1B)=P(A_2B)=P(A_3B)=1/6,$$

$$P(A_1|B)=P(A_2|B)=P(A_3|B)=1/3.$$

The conditional entropy of the dissection $\mathcal{A}$ given that event B occurs is $S(\mathcal{A}|B) = \log 3$.

Entropy of an event. A special case of the conditional entropy $S(\mathcal{A}|B)$ is when we use the sample points to dissect the sample space, namely, we use the sample space itself $\Gamma = \{\gamma_1, \gamma_2, ...\}$ as a dissection. The conditional probability is

$$P(\gamma_i|B) = \begin{cases} P(\gamma_i)/P(B), & \text{when } \gamma_i \in B \\ 0, & \text{when } \gamma_i \notin B \end{cases}$$

We define the entropy of the event B by

$$S_B = -\sum_i P(\gamma_i|B) \log P(\gamma_i|B).$$

Thus, we view the event B as a sample space, and dissect the sample space using the sample points that below to B.

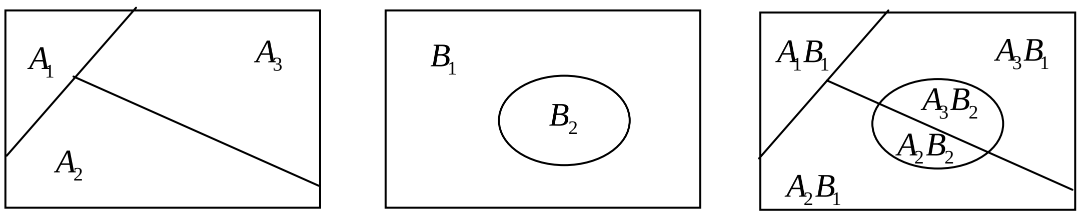
Examples. *(a) Entropy for a die to have an even number.* Event B = "a die have an even number". $P(B) = 1/2$. Three sample points: 2, 4, 6, each with absolute probability 1/6, or conditional probability 1/3. Thus, $S_B = \log 3$, same as the entropy of a dissection with three equally probable events.

(b) Entropy of a phase of a substance. A substance can be in different phases, e.g., solid, liquid, gas. We regard the set of all the quantum states of a substance as a dissection. We will be interested in the entropy conditional on that the substance is in a particular phase, such as the entropy of the liquid phase of the substance.

(c) Entropy of a rare event. A site in a crystal that misses an atom is known as a vacancy. A nearby atom may jump into the vacancy and leave a vacancy behind. The net result is that the vacancy jumps one atomic distance. Such a jump is a rare event: the atom must acquire a large amount of energy to overcome the activation barrier. No matter how rare it is,

however, the event corresponds to a large number of quantum states of the crystal. We will be interested in the entropy of the quantum states conditional on that the vacancy jumps.

Shannon's formula. (I need more time to make this part smooth.) Let $\mathcal{A} = \{A_1, A_2, ..., A_i, ...\}$ and $\mathcal{B} = \{B_1, B_2, ..., B_j, ...\}$ be two dissections of a sample space Γ. Denote $\mathcal{AB}$ as the class of events $\{A_1B_1, A_2B_1, ..., A_iB_j ...\}$, which again consist of a dissection of Γ.



Dissection $\mathcal{A}$ Dissection $\mathcal{B}$ Common refinement, $\mathcal{AB}$

The conditional probability is

$$P(A_i|B_j) = \frac{P(A_iB_j)}{P(B_j)}.$$

Note that

$$\sum_{i,j} P(A_iB_j) = \sum_j P(B_j)\left(\sum_i P(A_i|B_j)\right) = 1.$$

By definition, the entropy of the dissection $\mathcal{AB}$ is

$$S(\mathcal{AB}) = -\sum_{i,j} P(A_iB_j)\log P(A_iB_j).$$

We can readily confirm *Shannon's formula*:

$$S(\mathcal{AB}) = S(\mathcal{B}) + \sum_j P(B_j)S(\mathcal{A}|B_j).$$

Consider a special case when the dissection $\mathcal{A}$ is the sample space $\Gamma = \{\gamma_1, \gamma_2, ...\}$. An event in the dissection $\mathcal{AB}$ is one of the sample points. In this case, Shannon's formula reduces to

$$S(\gamma_1, \gamma_2, ...) = S(B_1, B_2, ...) + \sum P(B_j) S_{B_j} .$$

Example. Consider a crystal with a small fraction of sites being vacant. All the quantum states of this crystal form a sample space. We dissect the sample space is two ways. One dissection consists of individual quantum states, and the other dissection consists of arrangements of the vacancies. Associated with each arrangement of the vacancies are many the quantum states of the crystal. Thus,

A_i is a quantum state of the defect crystal.

B_j is an arrangement of the vacancies in the crystal.

$A_i B_j$ is a quantum state consistent with the a particular arrangement of vacancies B_j.

$S(\mathcal{AB})$ is the entropy of the defect crystal when the dissection consists of individual quantum states.

$S(\mathcal{B})$ is the entropy of the dissection consists of arrangements of the vacancies.

$S(\mathcal{A}|B_j)$ is the entropy that of the crystal given that the vacancies is in a given arrangement B_j.

$P(B_j)$ is the probability of the vacancies in the arrangement B_j.