Lecture 03: Entropy from Information

In these notes, we will use a simple number guessing game to develop intuition behind the definition of entropy in physics. The number guessing game here is similar to the question guessing game in Chapter 2 of [1].

1 The definition of entropy

In the early 19th century, scientists and engineers[1] sought to understand the properties of steam and heat engines using the recently developed theories of how gas volumes, temperatures, and pressures are interrelated. Ultimately, these engineers found that heat always flowed from higher temperatures to lower temperatures, and, without the input of external work, the reverse was not possible (Fig. 1).

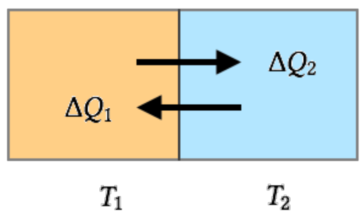


Figure 1: An isolated system experiencing heat exchange. T_1 and T_2 are the temperatures of the left and right system, and ΔQ_1 and ΔQ_2 are the heats added to the left and right systems. Scientists of the 19th century observed that when $T_1 > T_2$, then $\Delta Q_1 < 0$ and $\Delta Q_2 > 0$ and vice versa for $T_1 < T_2$. Clausius explained this result by defining a new concept called entropy and introducing the second law of thermodynamics.

In 1865 Rudolf Clausius, in an attempt to model what was happening in these systems, introduced a new quantity and a new law to explain what fundamentally was being transferred from one temperature to the next. He first stated that any transfer of heat ΔQ to or away from a system (where the sign of ΔQ defines the direction of transfer) is associated with an increase in entropy of

$$\Delta S = \frac{\Delta Q}{T} \qquad \text{[Clausius definition of entropy]}, \tag{1}$$

where T is the temperature at which this transfer occurs[2] . He then asserted what has become known as the **second law of thermodynamics**:

> The entropy of an isolated system (that is, a system which exchanges neither matter nor energy with its surroundings) either increases or stays the same. The entropy of such an isolated system never decreases.

The explanation for why heat flowed from higher temperatures to lower temperatures was as follows. Given the heat changes shown in Fig. 1 and Eq.(1), the total entropy change of the system in Fig. 1 is

$$\Delta S_{\text{tot}} = \Delta S_1 + \Delta S_2 = \frac{\Delta Q_1}{T_1} + \frac{\Delta Q_2}{T_2}. \tag{2}$$

[1] In particular the engineer Sadi Carnot.

[2] More accurately, Eq.(1) applies primarily to *reversible* processes, that is processes in which the time-reversed version is still physically feasible. The spontaneous expansion of a small amount of gas into an open room is not reversible, but the expansion of a gas in a closed container resulting in an increase in the height of a piston is.

For an isolated system, no heat enters or leaves the system. That is, the total heat change of the system is zero: $\Delta Q_1 + \Delta Q_2 = 0$. We can thus define $\Delta Q_1 = \Delta Q$ and $\Delta Q_2 = -\Delta Q$ for some ΔQ. From this fact and Eq.(2), we then have

$$\Delta S_{\text{tot}} = \left(\frac{1}{T_1} - \frac{1}{T_2}\right)\Delta Q. \tag{3}$$

By the second law of thermodynamics, ΔS_{tot} must be greater than or equal to zero. From relative values of T_1 and T_2 we can then determine the sign of ΔQ

- If $T_1 > T_2$, then ΔQ must be negative. Thus, $\Delta Q_1 < 0$ and $\Delta Q_2 > 0$ meaning heat flows from 1 to 2.
- If $T_1 < T_2$, then ΔQ must be positive. Thus, $\Delta Q_1 > 0$ and $\Delta Q_2 < 0$ meaning heat flows from 2 to 1.

In either case, heat flows from the higher temperature to the lower temperature. Eq.(1) together with the second law of thermodynamics well modeled why heat flows from high temperatures to low temperatures, but there still remained the question of what entropy was at the level of particle properties. In particular, Eq.(1) defines *changes* in entropy, but it does not define entropy absolutely.

A definition of entropy grounded in particle properties came a few years later with the work of Ludwig Boltzmann and Josiah Gibbs[3].

In 1877, Boltzmann stated that the entropy of a system was a function of the number of configurations, and the probability to be in any particular configuration was a function of the energy of the configuration. When configurations have a common energy E, then they are equally likely, and the total number of configurations $\Omega(E)$ can be counted as a function of that energy. Boltzmann then asserted that the entropy of a system with energy E and possible configurations $\Omega(E)$. is

$$S_{\text{Boltz}} = k_B \ln \Omega(E) \qquad \text{[Boltzmann entropy]}, \tag{4}$$

where $k_B = 1.3807 \times 10^{-23}$ J/K is Boltzmann's constant. The definition Eq.(4) only applied when all the configurations in a a system had the same total energy, and, consequently, the same probability of occurrence.

In 1902, in his treatise *Elementary Principles of Statistical Mechanics*, Gibbs provided a more general definition under the circumstances that the probabilities to be in various configurations were not equal because the energy of each configuration was not equal. Gibbs asserted that the entropy could more generally be defined as

$$S_{\text{Gibbs}} = -k_B \sum_{\{i\}} p_i \ln p_i \qquad \text{[Gibbs entropy]}, \tag{5}$$

where k_B is Boltzmann's constant, i denotes a particular configuration of a system, p_i is the probability to be in that configuration, and $\sum_{\{i\}}$ is a summation over all configurations. If there are Ω configurations in a system, and all of them are equally likely then $p_i = 1/\Omega$ and Eq.(5) reduces to Eq.(4).

These days Gibbs's definition of entropy Eq.(5) is taken as the most general form for entropy in physics. It is consistent with Eq.(4) and can be used to derive the Eq.(1) (for reversible processes). However, it is often difficult to see how the formula in Eq.(5) represents what we qualitatively understand as entropy. Entropy is often seen as a proxy for disorder or uncertainty, and in this sense the second law of thermodynamics is asserting that isolated systems tend to become more disordered as time goes on. However, how is this sense of disorder encompassed in the definition given in Eq.(5)? That is what we aim to answer in these notes.

Framing Question

Qualitatively, how can we understand Gibbs's definition of entropy given in Eq.(5)?

Since Eq.(5) is a more fundamental entity than Eq.(4) and Eq.(1), if we can find a way to qualitatively understand the Gibbs definition of entropy, we should also find better qualitative interpretation of all the

[3]See "History and outlook of statistical physics" for a short the history of the subject.

others. Our approach to finding a qualitative interpretation of Eq.(5) will seem like a roundabout one. We will first use a game to motivate a conceptual explanation of the mathematical form of the equation, and then we will reinterpret the results of the game from a physical perspective.

2 Guess That Number! - Round 1

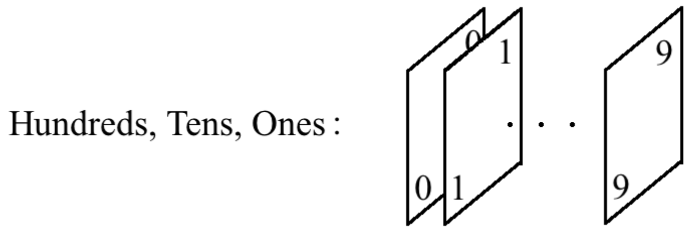


Figure 2: We have three decks of cards each of which represents a digits place for a three digit number.

Let's play a game. I have three decks of cards, and each deck contains ten cards with a single number from 0 to 9 on each card (Fig. 2). We will use the decks collectively to represent a number from 0 to 999. The first deck represents the hundreds digit; the second deck the tens digit; the third deck the ones digit. I place each deck in a box and shake the box so that the cards are no longer ordered. From each box I pull out a single card, and with the cards together I have some hidden number from 0 to 999.

> Your task is to determine a questioning strategy which will lead you, on average, to discern the hidden number in the fewest number of "yes" or "no" questions.

How would you go about this? Assuming you had no knowledge of the *most* efficient way to search for the correct number, you might first try to determine the number's general properties. (With 1000 numbers to choose from you'd probably immediately realize that asking questions like "Is it 251?" is a bad way to proceed). For example, you might ask the following sequence (with my corresponding answers):

1. Is the number a prime number? (Yes)
2. Is the last digit of the number 7? (Yes)
3. Is the hundreds digit an odd number? (No)
4. Is the tens digit an odd number? (No)

With these questions and their answers, you'd be able to discover that the hidden number is 457. All in four questions, which, given the spectrum of numbers, is quite good. The only problem is this low number of questions mostly depended on the fact that the number was a prime number, and "Is the number a prime number?" was the first question you asked. With that question you reduced the number of possible guesses from 1000 to 168. In other words, you got lucky, and if, bolstered by your success, you doubled down on this strategy the next time you played this game, you'd probably find the correct number in many more than four questions.

Because we're trying to find the hidden number in the fewest number of questions on average, there are good questions which consistently represent the best strategy towards eliminating numbers, and there are bad questions which only work if we're lucky. The question "Is the number a prime number" is a bad question because given that each number between 0 and 999 is equally likely, the question is more likely to leave us with 832 remaining numbers than with168 which is a poor start if we're trying minimize the number of questions we ask.

To see this play out we'll continue our example, but let's now say the number drawn from the boxes was not 457 and not prime. How would you proceed after the first question, now that you have 832 guesses to choose from? Another sequence of questions might be:

1. **[First Question]** Is the number prime (No; 832 #s left; Bad Question)
2. Is the number even? (Yes; 499 #s left; Bad question)
3. Is it divisible by 3? (No; 332 #s left; Bad question)
4. Is it divisible by 5? (No; 266 #s left; Bad question)
5. Is it greater than 505? (Yes; 133 #s left; Good question)
6. Does it have a 9 anywhere? (No; 95 #s left; Bad question)
7. Does the number have an 8 anywhere? (No; 47 #s left; Typically Bad, but this time Good)

You have just asked seven questions, and you still have 47 numbers to choose from (Also, as a practical matter, by the middle of your questioning, unless you had a running list of numbers, you most likely became confused as to which numbers were eliminated and which you could still guess from.) Having 47 numbers left might seem OK unless you knew that if you had employed a more efficient strategy, you would now have only eight numbers to choose from.

What led you astray is that you asked quite a few "bad questions". Specifically, any question which divides the spectrum of numbers into unequal (in number of elements) sets of numbers is a bad question because the hidden number will more likely be in the larger set of numbers than in the smaller set, and therefore, on average, you would have to ask more questions than you optimally would have in order to find the hidden number. Phrased differently, **the good questions are those which consistently eliminate *half* the spectrum of numbers after each answer**[4], and if you had consistently employed these good questions from the beginning you would be down to eight numbers rather than 47.

We still don't know the new hidden number, so let's start over. But this time, we'll go all the way through with good questions. Here's a possible sequence of questions:

1. Is the number greater than 500? (Yes; 500 #s left)
2. Is the number greater than 750? (No; 250 #s left)
3. Is the number greater than 625? (No; 125 #s left)
4. Is the number greater than 563? (No; 63 #s left)
5. Is the number greater than 531? (Yes; 32 #s left)
6. Is the number greater than 547? (No; 16 #s left)
7. Is the number greater than 539? (No; 8 #s left)
8. Is the number greater than 535? (No; 4 #s left)
9. Is the number greater than 533? (No; 2 #s left)
10. Is the number 533? (No) $\longrightarrow$ The number is 532.

[4]To be precise, it is those questions which eliminate half the *probability-weighted* spectrum of numbers, but we'll get to that later. Here, for a uniform distribution, the probability weigthed spectrum is just the spectrum itself.

What we're doing here is exemplative of a **binary search algorithm** where one searches for a particular element in a set by successively eliminating half the elements. For a set with 1000 elements, the total number of binary-valued questions needed to specify a single element is the number of times you have to divide 1000 by 2 in order to get 1. This is simply the logarithm base 2 of 1000, so

$$\text{\# of Qs to specify 532} = \log_2 1000 \approx 9.97 \tag{6}$$

rounded up, which is 10, as we listed above. We can think of Eq.(6) as the amount of *information* we needed to determine the number 532. But the number 532 is not special; we would have needed the same number of questions to specify any number we drew:

$$\text{\# of Qs to specify any number} = \log_2 1000. \tag{7}$$

We can now ask a more general question: What is the *average* number of questions we need to specify a number? Given Eq.(7) the answer may be clear, but for future reference let's go through this in some detail. We recall that, in general, to compute the average of a quantity we have

$$\langle \text{Quantity} \rangle = \sum_{\text{state}} p(\text{state}) \times \text{Quantity}(\text{state}), \tag{8}$$

which is a sum over the values of the quantity for each state weighted by the state's probability. Using Eq.(7) and the fact that each number from 0 to 999 could be drawn with probability 1/1000, we then find

$$\langle \text{\# of Qs} \rangle = \sum_{i=1}^{1000} \frac{1}{1000} \log_2 1000 = \log_2 1000. \tag{9}$$

Although the values of Eq.(9) and Eq.(7) are the same, Eq.(9) is representative of our entire set of numbers while Eq.(7) only represents any particular number. Now, let's make our game a little more interesting.

3 Guess That Number! Round 2

Although the discussion in the previous section might have seemed involved, it was greatly simplified by the fact that all of the numbers occurred with equal probability. How should our questioning strategy change if they didn't?

We currently still have 10 cards (with numbers from 0 to 9) in each box. Let's now add to the hundreds and the tens box, each, eight cards with the number 9. So, now the hundreds and the tens box each have 18 cards, half of which are the number 9 (Fig. 3). We'll play the same game: I draw a card from each digit box to create a three digit number. What questions should you ask to guess the card in the fewest number of questions?

Having learned something from the last round of this game, and recognizing that the spectrum of numbers is unchanged (i.e., it is still from 0 to 999), you might proceed with an exact replica of the binary search strategy we previously employed. Namely, you'd ask if the number is greater than 500, then if it's greater than 750 and so on. And while this questioning strategy will definitely allow you to determine the hidden number in 10 questions, it is a sub-optimal strategy for this new distribution of numbers. This is because the *minimum* number of questions we need in order to determine a number is dependent on the probability that number occurs, and our probability distribution is no longer the same.

We can get a sense of this with an extreme example. Consider the case where for the same spectrum of numbers, we had a 0.99 probability of finding the number 792 and a $0.01/999 \approx 0.0001$ probability of finding any other number. Then, although the spectrum of numbers is the same (from 0 to 999), since one number occurs so much more frequently than the others, we now require fewer questions (on average) to determine the correct number. Noting the relative probabilities, the first question we would ask is "Is the number 792?". The answer would most likely be "Yes", and we would have determined the correct number

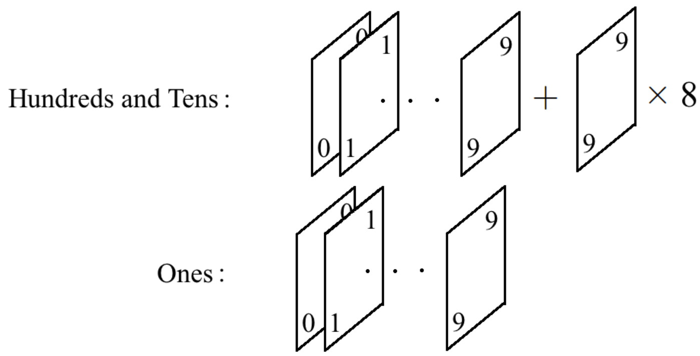
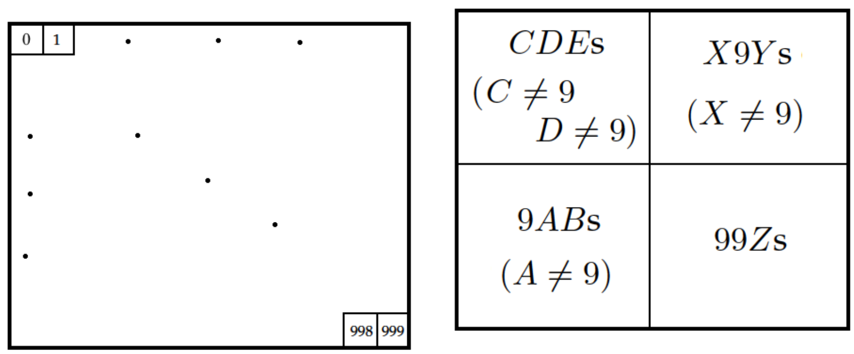


Figure 3: To the hundreds and tens decks we add eight "9" cards, thus giving us a 50% chance of drawing a card with a 9 for each of these digits.



(a) Spectrum of numbers

(b) Probability-weighted spectrum of numbers

Figure 4: Space and Probability-weighted space of numbers for the case where the number 9 comprises half of both the hundreds and tens deck of cards. We note that although the spectrum of numbers goes from 0 to 999, the $99Z$s for example take up 1/4 of the probability-weighted space of numbers (rather than 1/100 of that space as in the previous section). We also note that Fig. 4a would also represent the probability-weighted space of numbers if each number were equally likely.

after a single question.

This example gives us a clue of how to proceed in this new round of questioning. Recall that we have the same spectrum of numbers as before except our hundreds digit and our tens digit boxes each are half composed of cards with a 9. Thus the probability to get a number with a hundreds digit of 9 is 1/2, and the probability to get a number with a tens digit of 9 is 1/2. Thus the probability breakdown for all the numbers from 0 to 999 is as follows:

- Probability to get any $99Z$ number: $\frac{1}{2} \times \frac{1}{2} \times 1 = \frac{1}{4}$
- Probability to get any $9AB$ number (where $A \neq 9$): $\frac{1}{2} \times (1 - \frac{1}{2}) \times 1 = \frac{1}{4}$
- Probability to get any $X9Y$ number (where $X \neq 9$): $(1 - \frac{1}{2}) \times \frac{1}{2} \times 1 = \frac{1}{4}$
- Probability to get any CDE number (where $C,\ D \neq 9$): $(1 - \frac{1}{2}) \times (1 - \frac{1}{2}) \times 1 = \frac{1}{4}$

where, Z, B, Y, and E can be any digit. We now see that we have a *probability-weighted* spectrum of states which is distinct from the basic spectrum of states. We schematically depict the two representations of our state space in Fig. 4.

In determining a hidden number, we want, as before, to eliminate half of *something* after each step, but because the numbers 0-999 are not uniformly distributed it would not be an optimal questioning strategy to sequester the numbers as if they were. Instead, after each turn, what we actually want to eliminate is half of the *probability-weighted* state space. Again, this is because (as in our extreme example with a 0.99 probability of obtaining 792), the amount of information (i.e., the average number of questions) needed to determine the hidden number depends on the probabilities of all possible numbers.

Thus, for this new set of numbers the first and second questions we *should* ask (in either order) are

1. Does the number have a 9 in the hundreds digit?
2. Does the number have a 9 in the tens digit?

These questions are optimal given our distribution of numbers because they have 1/2 a chance of being answered "Yes" and hence 1/2 a chance of being answered "No", and thus will always eliminate half of the probability-weighted state space. After these questions, we would be left with numbers of the form $99Z, X9Y\ (X \neq 9), 9AB\ (A \neq 9)$, or $CDE\ (C,\ D \neq 9)$. The number of elements and associated relative probability for each element in these sets of numbers are

- $99Z$s: There are $1 \times 1 \times 10$ possible numbers each of which occur with equal probability.
- $9AB$s $(A \neq 9)$: There are $1 \times 9 \times 10 = 90$ possible numbers each of which occur with equal probability.
- $X9Y$s $(X \neq 9)$: There are $9 \times 1 \times 10 = 90$ possible numbers each of which occur with equal probability.
- CDEs $(C,\ D \neq 9)$: There are $9 \times 9 \times 10 = 810$ numbers each of which occur with equal probability.

Given that the numbers in each of these sets occur with equal probability, we can simply use the standard binary search strategy within these sets to determine our hidden number. Through this questioning strategy (the first two questions and then a binary search), we would then have optimally determined the hidden number in the fewest number of questions on average.

We can now add some theoretical structure to this setup. Using results similar to Eq.(7) and fact that the numbers in each of our four sets occur with equal probability, we can determine how many questions (after the first two listed above) we would need to determine the value of the hidden number. Including the two questions, we have

To determine a	You need
$99Z$	$2 + \log_2 10$ questions
$9AB\ (A \neq 9)$	$2 + \log_2 90$ questions
$X9Y\ (X \neq 9)$	$2 + \log_2 90$ questions
$CDE\ (C,\ D \neq 9)$	$2 + \log_2 810$ questions

Table 1

where the result of each logarithm is rounded up to the nearest whole number. Now, given that each set of these numbers occur with probability 1/4, the average number of questions needed to specify a number is

$$\begin{aligned}\langle \#\text{ of Qs } \rangle &= \sum_{\text{value}} p(\text{value}) Q(\text{value}) \\ &= \tfrac{1}{4}\left(2 + \log_2 10\right) + \tfrac{1}{4}\left(2 + \log_2 90\right) \\ &\quad + \tfrac{1}{4}\left(2 + \log_2 90\right) + \tfrac{1}{4}\left(2 + \log_2 810\right) \\ &= 2 + \tfrac{1}{4}\log_2(6.561 \times 10^7) \approx 8.491, \end{aligned} \tag{10}$$

which gives 9 when rounded up. So, on average, we'll need 9 (as opposed to the 10 of the previous section) questions to determine the hidden number in this setup.

4 Generalizing Arguments - Round 3

We've been through two rounds of "Guess That Number" and in each round we determined what questioning strategy would, on average, have us asking the fewest number of questions to find the hidden number. But we also determined something else: we specifically calculated what that fewest number of questions should be. From here we can pursue generality and ask something which can encompass any probability distribution of numbers.

> **Question:** If the hidden number has a probability p_i of being i, what average number of questions (under an optimal questioning strategy) would we need to ask in order to determine the hidden number?

We can answer this question by induction, namely by inspecting the results of the previous section and presenting them with a more general notation. We start by determining the probabilities for the hidden number to be each kind of number listed in Table 1. From the bulleted lists in the previous section, we see that the probability p_{99Z} for the hidden number to be a *particular* number of the form $99Z$ is $\frac{1}{4} \times \frac{1}{10} = \frac{1}{40}$. In general the probabilities p_i to find a *particular* number i are

$$p_{99Z} = \frac{1}{4} \times \frac{1}{10} = \frac{1}{40} \tag{11}$$
$$p_{9AB} = \frac{1}{4} \times \frac{1}{90} = \frac{1}{360} \tag{12}$$
$$p_{X9Y} = \frac{1}{4} \times \frac{1}{90} = \frac{1}{360} \tag{13}$$
$$p_{CDE} = \frac{1}{4} \times \frac{1}{810} = \frac{1}{3240}. \tag{14}$$

The word *particular* is important here, for while there is a $\frac{1}{4}$ chance of getting *any* of the 810 numbers of the form CDE with $(C,\, D \neq 9)$, there is only a $\frac{1}{4} \times \frac{1}{810} = \frac{1}{3240}$ chance of getting, for example, the number 523. With these more general representations of the elements of our analys we are now ready present our previous results in a more general notation. For example, from an inspection of Table 2 we see that the "number of questions" needed to determine a number in each set can be written in terms of the probabilities to find that number: We can then rewrite Eq.(10) with an eye toward this notation. Specifically we have

To determine a	Number of questions you need
$99Z$	$2 + \log_2 10 = \log_2 40 = -\log_2 p_{99Z}$
$9AB\,(A \neq 9)$	$2 + \log_2 90 = \log_2 360 = -\log_2 p_{9AB}$
$X9Y\,(X \neq 9)$	$2 + \log_2 90 = \log_2 360 = -\log_2 p_{X9Y}$
$CDE\,(C,\, D \neq 9)$	$2 + \log_2 810 = \log_2 3240 = -\log_2 p_{CDE}$

Table 2

$$\begin{aligned} \langle \#\text{ of Qs } \rangle &= \sum_{\text{value}} p(\text{value}) Q(\text{value}) \\ &= -\sum_{\{99Z\}} p_{99Z} \log_2 p_{99Z} - \sum_{\{9AB\}} p_{9AB} \log_2 p_{9AB} \\ &\quad - \sum_{\{X9Y\}} p_{X9Y} \log_2 p_{X9Y} - \sum_{\{CDE\}} p_{CDE} \log_2 p_{CDE} \end{aligned} \tag{15}$$

where each sum runs over all elements in the respective set. Collectively the sums run over all the numbers from $i = 0$ to $i = 999$. Therefore we can simply write

$$\langle \#\text{ of Qs} \rangle = -\sum_i p_i \log_2 p_i. \tag{16}$$

Eq.(16) provides a rigorous definition of information. With our optimal questioning strategy we have heuristically derived a formula for the average amount of "information" needed to determine a hidden number.

Noting that our hidden number can more generally be considered to be the hidden *state*, and that the collection of numbers we can draw from the three boxes can be considered a *system*, we can extrapolate a precise definition of information to be as follows:

Definition 1 *The quantity*

$$I = -\sum_i p_i \log_2 p_i \tag{17}$$

is the average number of binary-valued questions (under an optimal questioning strategy) needed to specify the state of a system. An 'optimal questioning strategy' refers to a series of questions which seek to divide the probability-weighted space of states in half after each question.

Information: associates the mathematical idea (and somewhat unintuitive label) "information" with the more precise definition "the number of binary questions (under an optimal questioning strategy) needed to specify the state."

We can even return to the case studied in Sec. 2, and see if the general formula Eq.(16) is consistent with our results. For this case, each number from 0 to 999 is equally likely and hence each number has a probability $\frac{1}{1000}$. Thus the information is

$$I = -\sum_i p_i \log_2 p_i = -\sum_{i=0}^{999} \frac{1}{1000} \log_2 \frac{1}{1000} = \log_2 1000 \tag{18}$$

which, as we found in Eq.(9), is the average number of questions needed to determine the hidden number.

Exercise: Calculating Information
Calculate the amount of information (defined by Eq.(17)) needed to specify the outcome of the following random trials:

1. A single coin flip for a fair coin
2. A single dice roll for a fair die
3. A single dice roll where there is a 90% chance of getting a 6 and 2% chance of getting any other value

5 Returning to Physics

In moving to statistical physics, we need to introduce some vocabulary. When we are studying a physical system, we can characterize the system on at least two size scales. On a large scale we can describe the macroscopic properties of the system such as the total energy the system contains, its total volume, or its total magnetization. Such properties constitute descriptions of the system's **macrostate**. However, for each macrostate there are often many different smaller length-scale states associated with it. For example, a box of particles which have total energy E (a description of the macrostate) can contain many particles with various energies. For N particles, a particular set of energies $E_1, E_2, \ldots, E_N$ defines a **microstate** of the macrostate E so long as $E = E_1 + E_2 + \cdots + E_N$.

Definition 2 *A **macrostate** of a system is a description of the total system (i.e., at a "macroscopic" scale),*

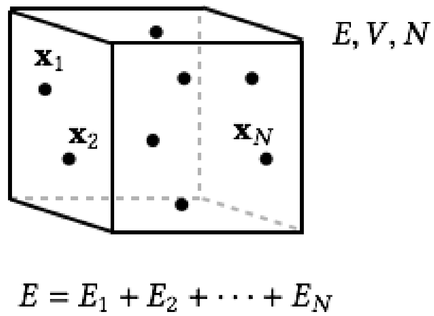


Figure 5: The macrostate of a system of gas particles can be defined by total energy E. A particular microstate is a set of energies $E_1, E_2, \ldots E_N$, such that $E = E_1 + E_2 + \cdots E_N$.

and often is associated with many different configurations of the system. A ***microstate*** *of a system is a particular configuration of a macrostate.*

Below are some mathematical examples of the relationship between microstates and macrostates. In each case, the macrostate refers to a description of the system as a whole.

- - **macrostate:** Choosing four numbers x_1, x_2, x_3, x_4 such that their sum is 10.
  - **microstate:** The numbers $x_1 = 2$, $x_2 = 4$, $x_3 = 3$, $x_3 = 3$.
- - **macrostate:** Getting 5 heads out of 10 coin flips which can be heads (H) or tails (T).
  - **microstate:** Getting HTHHHTTHT in that order for the 10 coin flips.
- - **macrostate:** Any number between 1 and 10
  - **microstate:** The number 3

Returning to the "Guess that number" game, we can use the macrostate and microstate language to describe our results. In the game, our macrostate would be defined as "numbers between 0 and 999" and a microstate of the system would be a particular number in this range (e.g., 532). In the first round of the game, each "microstate" was equally probable. So, given that we had $N = 1000$ microstates to choose from, we found that the information needed to specify a particular microstate of the system was

$$I = \log_2 N. \tag{19}$$

Similarly, in the second and third rounds of the game, we considered microstates which were not equally probable. Instead the microstate i occurred with probability i. In this case, we discerned that the information needed to specify the microstate was

$$I = -\sum_{i=1}^{N} p_i \log_2 p_i. \tag{20}$$

Comparing Eq.(19) and Eq.(20) to Eq.(4) and Eq.(5) (and noting that for Eq.(20), $\sum_{i=1}^{N}$ represents the sum over microstates), we find that our definition of information and the physics definitions of entropy differ merely by a constant. Using the logarithm identity

$$\frac{\ln x}{\ln 2} = \log_2 x \tag{21}$$

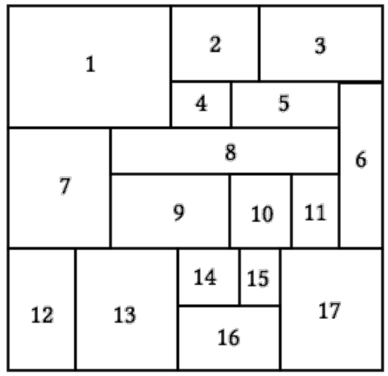


Figure 6: Depiction of a system with 17 microstates whose probabilities are proportional to their areas. In a physical system, we can be given a set of microstates which our system can occupy with various probabilities. We can compute the entropy of this system by applying Eq.(5).

and in particular considering Eq.(20) and Eq.(5), we find

$$S_{\text{Gibbs}} = -k_B \sum_{\{i\}} p_i \ln p_i = -k_B \ln 2 \sum_{\{i\}} p_i \log_2 p_i = (k_B \ln 2)\, I, \tag{22}$$

or

$$\text{Gibbs Entropy} = (\text{Constant}) \times (\text{Information needed to specify microstate}). \tag{23}$$

Since the Gibbs entropy is proportional to what we determined to be the information needed to specify a microstate, we can use this informational interpretation to motivate an understanding of the physics definition of entropy. The fact that entropy and information differ by a dimensional constant can be seen as a historical artifact of how entropy was first experimentally defined and the role the natural logarithm plays in calculus.

Therefore with Eq.(23), we see that our game has already supplied much of the interpretive work in understanding the Gibbs entropy. In a physical system, we will have many different microstates each of which occurs with various probabilities. In a way similar to our depiction of the space of numbers in the second round of the ”Guess that number” game, we can schematically depict these microstates as cells in a grid with the ratio between the area of the cell and the total area of all cells representing the probability to be in a particular microstate. When we compute the entropy of such a system (according to the definition in Eq.(5)) we are essentially computing the amount of information (according to the definition given in Def. 1.) needed to precisely specify a microstate in the system.

So we have an interpretation for

$$S_{\text{Gibbs}} = -k_B \sum_{\{i\}} p_i \ln p_i, \tag{24}$$

and from the second law of thermodynamics we know that this quantity increases for isolated systems. However, this result is not as useful to us as it can be. In particular we know how to compute the entropy of a system given p_i (the probability to be in microstate i) but we don’t know how to calculate p_i itself. In practice, statistical physics goes in the reverse direction: with the definition of entropy in Eq.(24), and knowing the laws of thermodynamics, we are able to compute p_i for a particular microstate. In these notes we introduced the second law of thermodynamics. Now it’s time to discuss the others.

References

[1] A. Ben-Naim, *Entropy and the second law: interpretation and misss-interpretationsss*. World Scientific Publishing Company, 2012.