

# Physics/Neuro 141

Aravi Samuel

Week 1

## Abstract

Sensory perception begins when stimulus energy from the environment is absorbed by a molecular or cellular receptor. The trouble is that these receptors are not at zero energy prior to receiving the stimulus. Because most biology happens at 300K, the receptors have, on average,  $k_B T/2$  amount of energy at any given time. Perhaps the stimulus energy is much larger than mean thermal energy, perhaps it is smaller. But even if the average amount of thermal energy is small, the amount of energy also fluctuates to lower or higher values. No matter where you set an “energy threshold” for detecting a stimulus, thermal energy on its own will be able to exceed that threshold. The reason is that thermal energy is distributed by the Boltzmann distribution.

## 1 Equipartition theorem

In classical statistical mechanics, the equipartition theorem relates the temperature of a system to its average energies. In thermal equilibrium, energy is shared equally among all independent quadratic degrees of freedom; for example, the average kinetic energy per degree of freedom in translational motion of a molecule with mass  $m$  should be equal. For gas molecules in motion, the mean kinetic energy along each axis is equal to  $k_B T/2$ :

$$\left\langle \frac{mv_x^2}{2} \right\rangle = \left\langle \frac{mv_y^2}{2} \right\rangle = \left\langle \frac{mv_z^2}{2} \right\rangle = \frac{k_B T}{2} \quad (1)$$

Thus, the mean total kinetic energy of a mono-atomic gas particle (one without rotational energy or vibrational energy),  $\langle mv^2/2 \rangle$  is  $3k_B T/2$ .

**Question.** Consider the force on the walls of a cubic container of volume  $V = L^3$  containing  $N$  gas molecules at temperature  $T$ . Calculating the change in momentum of the particles bouncing against the walls of the container and the rate at which they bounce against the walls of the container, derive the ideal gas law  $PV = Nk_B T$ .

## 2 The Boltzmann distribution

The energy of a system in thermal equilibrium is not fixed, but will fluctuate around its mean value. Boltzmann discovered that the probability distribution of energies is an exponential:

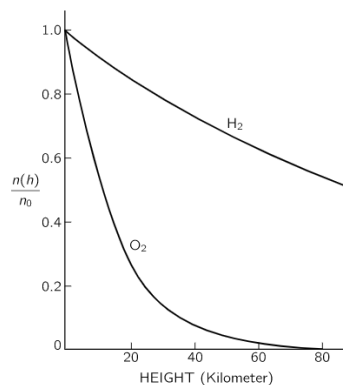


Fig. 40-2. The normalized density as a function of height in the earth's gravitational field for oxygen and for hydrogen, at constant temperature.

Figure 1: Exponential atmosphere

$$P(E) \propto e^{-E/k_B T} \quad (2)$$

Consider the energy of a gas molecule with mass  $m$  in a gravitational field. It will have a potential energy  $mgh$ , and so:

$$P(h) \propto e^{-mgh/k_B T} \quad (3)$$

This result gives us the famous “exponential atmosphere” (see Figure 1), which you can also read about it in the Feynman Lectures on Physics.

## 2.1 The energy states of a sensory receptor

The Boltzmann distribution also governs the likelihood of a sensory receptor like rhodopsin having different thermal energies. We describe the different states of a molecule using a “reaction coordinate”, and assign an energy to each point along the reaction coordinate (Figure 2).

From the Boltzmann distribution, we can infer the relative likelihood of being in different states. Often, going from one stable state (A) to another state (B) requires going over an activation barrier. The height of this activation barrier will slow this reaction rate. When the system has reached equilibrium, the relative likelihood of the two states is given by the Boltzmann distribution:

$$\frac{p_A}{p_B} = e^{-\frac{\Delta E}{k_B T}} \quad (4)$$

To estimate the transition rate from state A to B, we need the relative likelihood of being at the top of the activation barrier relative to the likelihood of being in state A. This is an estimate of the fraction of particles in state A that are able to move over the barrier:

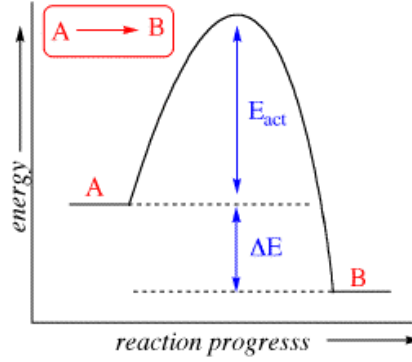


Figure 2: Reaction coordinate

$$k_{A \Rightarrow B} \propto e^{-\frac{E_{act}}{k_B T}} \quad (5)$$

Thus, the Boltzmann distribution gives us insight into the distribution of energies of sensory receptors before the arrival of stimulus energy, as well as the speed of the reactions that characterize every step in sensory transduction.

### 3 Using counting statistics to derive the Boltzmann distribution

Find the distribution of  $N$  particles over the energy states available to them. That is, how many,  $n_s$ , of the  $N$  particles in the assembly do we expect to find in the state of  $s$  having energy  $\epsilon_s$ ? Because of conservation of mass:  $\sum_s n_s = N$ . Because of conservation of energy:  $\sum_s n_s \epsilon_s = E$ . The probability of any system being in a given state  $s$  is  $n_s/N$ .

Thus, the average energy of each system is:

$$\bar{\epsilon} = \frac{\sum_s n_s \epsilon_s}{\sum_s n_s} \quad (6)$$

To find the distribution of the particles over their possible energy states, we enumerate the states ( $s = 1, 2, \dots$ ), associate each state with a discrete energy ( $\epsilon_1, \epsilon_2, \dots$ ), and assign a number of particles to each state ( $n_1, n_2, \dots$ ).

What is the probability of observing a specific distribution with say  $n_1$  systems state 1,  $n_2$  in state 2, and so on? The central assumption is that every possible distinct arrangement of the systems that satisfies the conservation laws is equally possible. The probability of observing a particular distribution of  $n_s$  is thus proportional to the number of distinct ways that the distribution can be achieved with the  $N$  particles. The number of permutations that corresponds to the same distribution is a measure of its probability. The number of ways the distribution can be formed is:

$$W = \frac{N!}{n_1!n_2!n_3!\dots} \quad (7)$$

To maximize  $W$  with respect to  $n_s$  and with respect to the conservation laws we need to use a convenient analytical expression for the factorial and we need to use Lagrange Multipliers.

We choose to maximize  $\log W$ <sup>1</sup> subject to the constraints of conservation of energy and mass. Thus,

$$d \left( \log W - \alpha \sum_s n_s - \beta \sum_s \epsilon_s n_s \right) = 0 \quad (8)$$

The  $\alpha$  and  $\beta$  are the Lagrange multipliers. Varying with respect to  $n_s$  and incorporating Stirling's Formula gives:

$$- \sum_s dn_s (\log n_s + \alpha + \beta \epsilon_s) = 0 \quad (9)$$

Because this must hold true for every  $\delta n_s$ , every term in the sum must vanish. The values of  $n_s$  which do this are

$$\log n_s + \alpha + \beta \epsilon_s = 0 \quad (10)$$

We can thus conclude that the occupancy of state  $s$  depends exponentially on its energy:

$$n_s \propto e^{-\beta \epsilon_s} \quad (11)$$

Actually, we have shown that the exponential dependence of state occupancy on energy is true only at the most probable  $W$ . We have not shown that every other arrangement can be ignored, but, in fact, it can be when  $N$  gets to be very large. This is hard to show, and we will not attempt to do so here. We also have not shown that  $\beta$ , introduced here as a Lagrange multiplier, is related to temperature. For our purposes, we define  $\beta = 1/k_B T$ . When you take thermodynamics, you will learn that  $\beta$  behaves like  $1/k_B T$  for various thermodynamic relationships and can have no other meaning. We are simply going to declare victory with the result that energy levels are exponentially distributed at thermal equilibrium.

#### Questions.

- Derive Stirling's Law  $N! \approx \left(\frac{N}{e}\right)^N$  by integrating  $\log N!$  from 1 to  $N$ .
- A more accurate version of Stirling's Law is  $N! \approx \sqrt{2\pi N} (N/e)^N$ . Show why using this more accurate version has no effect on deriving the Boltzmann distribution at large  $N$ .

<sup>1</sup>You will learn when you take a Statistical Mechanics course like Physics 181 that  $\log W = S/k_B$ , the Boltzmann equation for entropy that is etched on his gravestone.

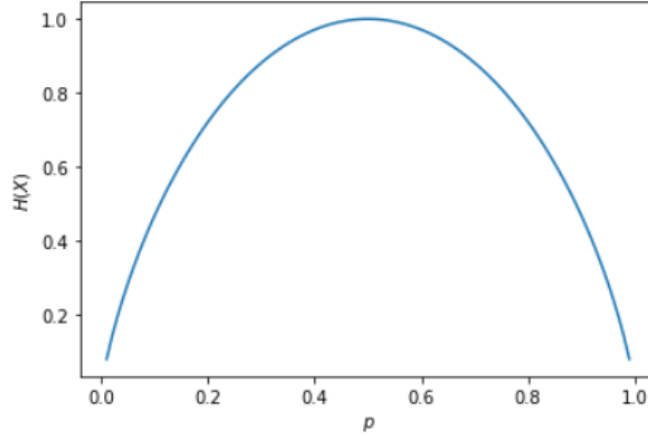


Figure 3: Entropy of a coin toss

## 4 Using information theory to derive the Boltzmann distribution

### 4.1 The maximum entropy distribution with constraints

Claude Shannon who invented *information theory* defined his own notion of *entropy* for any probability distribution.<sup>2</sup> Let  $X$  be a discrete random variable that can take on different possible values,  $x$ . The probability density function,  $p(x)$ , gives us the relative likelihood of taking on these different values.

**Definition.** The *entropy*  $H(x)$  of a discrete random variable  $X$  is defined by:

$$H(X) = - \sum_x p(x) \log_2 p(x) \quad (12)$$

For example, say the random variable was the outcome of the toss of a fair coin, so  $X$  can either be heads ( $p = 0.5$ ) or tails ( $p = 0.5$ ). In this case,  $H = 1$ . The entropy indicates the amount of information in *bits* that you need to characterize the outcome of the coin toss.

What if the coin was totally biased such that it always came up heads? In this case,  $H = 0$ . You don't need any information (0 bits) to characterize the outcome of the coin toss, because the outcome is guaranteed. What happens if we vary  $p$ ? The entropy has a minimum at  $p = 0$  or  $p = 1$  and a maximum at  $p = 0.5$  (Figure 3:

$$H(X) = -p \log_2 p - (1 - p) \log_2 (1 - p) \quad (13)$$

Consider the roll of a fair die. In this case, every outcome has  $p = 1/6$  and  $H = \log_2 6 \approx 2.5$  – if we had an 8-sided die from Dungeons and Dragons, the number of bits that you need to characterize one roll would be 3.

<sup>2</sup>Shannon couldn't actually decide what he wanted to call  $H$ . Neumann suggested entropy because "nobody knows what entropy is anyway."

**Question.**

Consider an  $N$ -sided die. The probability distribution of outcomes is normalized,  $\sum_{i=1}^N p_i = 1$ . Use Lagrange multipliers to show that the maximum entropy (i.e., most unbiased) distribution corresponds to  $p_i = 1/N$  for all possible outcomes.

*N.B.: Use natural logarithms for doing calculations. Doing calculus with  $\log_2$  is mathematically unwieldy. Using a different base for your logarithms only changes the amount of information needed to characterize a distribution by a multiplicative factor.*

## 4.2 The entropy of a probability distribution

If you know nothing about a probability distribution except the range of outcomes, the most probable *a priori* description of the probability distribution corresponds to the maximum entropy distribution. In general, this does not mean that the maximum entropy distribution is the right one. But it does mean that if it isn't, you are lacking some pertinent facts about the distribution.

Now reconsider the problem of  $N$  systems exchanging a total amount of energy  $E$  among them, and again that each particle is constrained to discrete energy levels  $\epsilon_1, \epsilon_2, \dots, \epsilon_s, \dots$ . What is the probability distribution that governs the energies of each system? This problem is fully described. We are not missing any facts needed to calculate a distribution. The correct probability distribution must be the one that maximizes its entropy. In other words, we have another Lagrange multiplier problem:

$$\delta \left( H - \alpha \sum_s p_s - \beta \sum_s \epsilon_s p_s \right) = 0 \quad (14)$$

The unavoidable conclusion is that  $p_s \propto e^{-\epsilon_s \beta}$ . If the distribution were anything else, you would be missing salient facts about the physical problem. But because we set the problem without any other facts, only the exponential distribution of energies, the Boltzmann distribution, is possible.

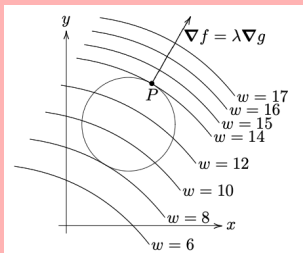
**Question.**

Consider a probability distribution,  $p(x)$ , over all values of  $x$  in one dimension, but with fixed mean,  $\langle x \rangle = \mu$  and fixed variance,  $\langle (x - \mu)^2 \rangle = \sigma^2$ . Show that the maximum entropy distribution corresponds to a Gaussian distribution.

### Lagrange multipliers.

For the function  $w = f(x, y, z)$  constrained by  $g(x, y, z) = c$ , the maxima and minima are those points where  $\nabla f$  is parallel to  $\nabla g$ :

$$\nabla f - \lambda \nabla g = 0 \quad (15)$$



**Geometric proof.** For concreteness, we've drawn the constraint curve,  $g(x, y) = c$ , as a circle and some level curves for  $w = f(x, y) = c$  with explicit (made up) values. Geometrically, we are looking for the point on the circle where  $w$  takes its maximum or minimum values.

Now, start at the level curve with  $w = 17$ , which has no points on the circle. So, clearly, the maximum value of  $w$  on the constraint circle is less than 17. Move down the level curves until they first touch the circle when  $w = 14$ . Call the point where the first touch P. It is clear that P gives a local maximum for  $w$  on  $g = c$ , because if you move away from P in either direction on the circle you'll be on a level curve with a smaller value.

Since the circle is a level curve for  $g$ , we know  $\nabla g$  is perpendicular to it. We also know  $\nabla f$  is perpendicular to the level curve  $w = 14$ , since the curves themselves are tangent, these two gradients must be parallel.

## References

- Cover and Thomas. *Information Theory*.
- Feynman. *Lectures*.
- Kardar. *Statistical Mechanics of Particles*
- Schrodinger. *Statistical Mechanics*.
- Wikipedia.
- MIT OpenCourseWare.