Entropy and its connection to probability.

The celebrated Boltzmann's expression for entropy reads as:

$$S = k \ln W$$

where k is a suitable constant usually realized as the Boltzmann constant. W is the multiplicity of states.

Ques. Why is the dependence of S on W logarithmic?

Ans. Because we want the entropy to be an extensive quantity.

We explain this below.

If A and B are two systems with multiplicities W_A and W_B respectively. Then we can say that the joint system will have the multiplicity $W = W_A W_B$

$$\therefore \underbrace{k \ln W_A W_B}_{\substack{\text{Total entropy} \\ \text{of} \\ \text{'A + B'}}} = \underbrace{k \ln W_A}_{\substack{\text{Entropy of} \\ \text{'A'}}} + \underbrace{k \ln W_B}_{\substack{\text{Entropy of} \\ \text{'B'}}}$$

"Logarithmic dependence of S on W guarantees additivity of entropies of subsystems."

For discrete probabilities we have additionally

$$S = k \ln W \simeq -N k \sum_i \ln p_i \, p_i$$

We derive this below.

Roll a die (t-faced) N times.

Then multiplicity $W = \dfrac{N!}{\prod_{i=1}^{t} n_i!}$

where n_i is the number of times the i^{th} face comes up.

In the limit of large N and n_i we can use the Stirling's approximation $\left(N! = (N/e)^N\right)$ to get:

$$W \simeq \frac{(N/e)^N}{\prod_{i=1}^{t} (n_i/e)^{n_i}} = \frac{N^N}{\prod_{i=1}^{t} n_i^{n_i}} = \prod_{i=1}^{t} \left(\frac{N}{n_i}\right)^{n_i}$$

$$= \prod_{i=1}^{t} \left(\frac{1}{p_i}\right)^{n_i}$$

Taking natural logarithm on both sides, we get:

$$\ln W \simeq -\sum_{i=1}^{t} n_i \ln(p_i) = -N \sum_{i=1}^{t} p_i \ln p_i$$

$$\because p_i = \frac{n_i}{N}$$

Before we discuss the consequences of entropy maximization with/without constraints, lets take a brief digression to learn the concept of Lagrange multipliers.

Concept of function maximization using Lagrange multipliers

Suppose there exists a function $f \equiv f(x,y)$

We want to maximize f. Then we must find the values $(x,y) = (x^*, y^*)$ which gives

$$\left.\frac{\partial f}{\partial x}\right|_{x=x^*} = \left.\frac{\partial f}{\partial y}\right|_{y=y^*} = 0$$

Now suppose you want to maximize $f(x,y)$ under the constraint:

$$g(x,y) = \text{constant.}$$

Clearly: $df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = 0$

If dx and dy are independent then it is easy to see that $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$ { As this is the only way to make $df = 0$ }

If dx & dy are not independent but are constrained by the condition: $dg = \frac{\partial g}{\partial x} dx + \frac{\partial g}{\partial y} dy$

then, $$\frac{\left(\frac{\partial f}{\partial x}\right)}{\left(\frac{\partial g}{\partial x}\right)} = \frac{\left(\frac{\partial f}{\partial y}\right)}{\left(\frac{\partial g}{\partial y}\right)} = \text{common ratio '}\lambda\text{'}$$

then: $$\frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} = \frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} = 0$$

here λ is the Lagrange multiplier corresponding to each constraint condition.

The above technique is easily generalized to multiple constraints each of them effected by a separate Lagrange multiplier.

Maximization of entropy :

(i) When there are no constraints : Roll an unbiased t-faced die N times. The only thing we can say here is that the probabilities of all possible outcomes must add up to 1.

$$\sum_{i=1}^{t} p_i = 1.$$

{ Actually this is a constraint in the sense of Lagrange multipliers! }

To maximise $S = -\sum_{i=1}^{t} \ln p_i \, p_i$, we use the concept of Lagrange multiplier as below:

$$\left. \frac{\partial S}{\partial p_i} - \lambda \frac{\partial}{\partial p_i}\left(\sum_{i=1}^{t} p_i\right) \right|_{p_i = p_i^*} = 0$$

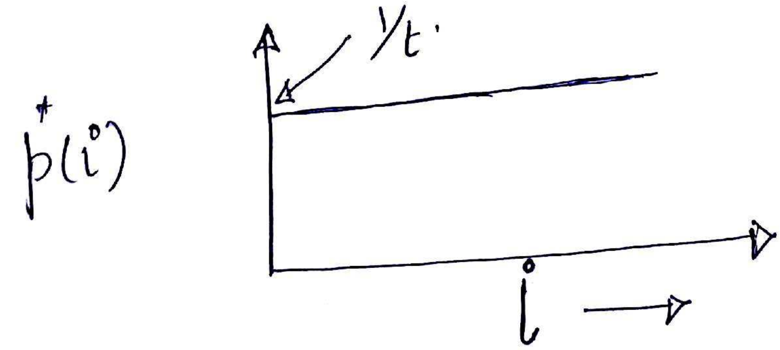
thus we get: $-(1 + \ln p_i^*) - \lambda = 0$

solving for p_i^*: $p_i^* = e^{-1-\lambda}$

dividing by $1 = \sum_{i=1}^{t} p_i^*$, we get:

$$p_i^* = \frac{e^{-1-\lambda}}{1} = \frac{e^{-1-\lambda}}{\sum_{i=1}^{t} p_i^*} = \frac{e^{-1-\lambda}}{e^{-1-\lambda}\, t} = \frac{1}{t}$$

$\therefore p_i^* = 1/t$ {Uniform or "flat" distribution}



(ii) When there are constraints:

This means we now have an extra constraint apart from the conservation of total probabilities.

We list both of them below:

$$\sum_{i=1}^{t} p_i = 1.$$

$$\frac{1}{N}\sum_{i=1}^{t} \epsilon_i n_i = \sum_{i=1}^{t} p_i \epsilon_i = \text{Average of } N \text{ die rolls is a known constant.}$$

Then maximizing S using the concept of Lagrange multipliers

$$\left.\frac{\partial S}{\partial p_i} - \lambda \frac{\partial}{\partial p_i}\left(\sum_{i=1}^{t} p_i\right) - \beta \frac{\partial}{\partial p_i}\left(\sum_{i=1}^{t} p_i \epsilon_i\right)\right|_{p_i = p_i^*} = 0$$

$\because S = -\sum_{i=1}^{t} p_i \ln p_i$, we have:

$$-(1+\ln p_i) - \lambda - \beta \epsilon_i^* = 0$$

Solving for p_i^*: $\quad p_i^* = e^{-1-\lambda-\beta\epsilon_i^*}$.

Again dividing by $1 = \sum_{i=1}^{t} p_i^*$, we get:

$$p_i^* = \frac{e^{-\beta\epsilon_i^*}}{\sum_{i=1}^{t} e^{-\beta\epsilon_i^*}} = \frac{e^{-\beta\epsilon_i^*}}{Z}$$

{Boltzmann distribution}

with

$Z \equiv$ Partition function

"Hence entropy maximization with constraints gives exponential distribution"