Maximizing the Entropy predicts flat distributions when there are no constraints (beyond normalized probability distributions).

We want to maximize $S(p_1, p_2, \ldots p_t) = -k \sum_{i=1}^{t} p_i \ln p_i$

But we have a normalization
Constraint : $g(p_1, p_2, \ldots p_t) = \left(\sum_{i=1}^{t} p_i\right) - 1 = 0$

We create a new objective function:

Lagrange Multiplier

$$F = S - \alpha g$$

And minimize F:

minimizes F

$$\frac{\partial F}{\partial p_i} = \frac{\partial S}{\partial p_i} - \alpha \frac{\partial g}{\partial p_i} = 0$$

Let's figure out these partials:

$$\frac{\partial g}{\partial p_i} = \frac{\partial}{\partial p_i}\left(\sum_{i=1}^{t} p_i - 1\right)$$

$$= \frac{\partial}{\partial p_i}\left(p_1 + p_2 + p_3 + \ldots p_t - 1\right)$$

$$= 1$$ ← no matter which p_i

$$\frac{\partial S}{\partial p_i} = \frac{\partial}{\partial p_i}\left(-k \sum_{i=1}^{t} p_i \ln p_i\right)$$

$$= -k\left(p_i \frac{\partial}{\partial p_i}(\ln p_i) + \frac{\partial p_i}{\partial p_i} \ln p_i\right)$$

$$= -k\left(\frac{p_i}{p_i} + \ln p_i\right)$$

$$= -k(1 + \ln p_i)$$

Now, we can put them together

$$\frac{\partial F}{\partial p_i} = -k - k\ln p_i - \alpha \qquad = 0$$

$$\ln p_i = \frac{-\alpha - k}{k} = -\frac{\alpha}{k} - 1$$

$$p_i = e^{-(\alpha/k + 1)}$$

What does this mean for the relative probabilities?

$$\frac{p_i}{\sum_{i=1}^{t} p_i} = \frac{e^{-(\alpha/k + 1)}}{\underbrace{\sum_{i=1}^{t}\left(e^{-(\alpha/k)+1}\right)}_{\text{all of these are the same!}}} = \frac{e^{-(\alpha/k + 1)}}{t\left(e^{-(\alpha/k+1)}\right)} = \frac{1}{t}$$

All of the p_i are the same $\longrightarrow$ flat distribution

---

Maximizing the entropy predicts exponential distributions when there are constraints.

Suppose we roll a t-sided die N times

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

Now, lets assign a "score" to each side, ε_i

Total score:

$$E = \sum_{i=1}^{t} n_i \varepsilon_i$$

Average score / roll:

$$\langle E \rangle = \frac{E}{N} = \sum_{i=1}^{t} \frac{n_i}{N}\varepsilon_i = \sum_{i=1}^{t} p_i \varepsilon_i$$

What is the distribution that maximizes the entropy **and** is consistent with an observed average score?

Now we have two constraints:

$$g(p_1, p_2, \ldots p_t) = \sum_{i=1}^{t} p_i - 1 = 0$$

$$h(p_1, p_2, \ldots p_t) = \sum_{i=1}^{t} p_i \varepsilon_i - \langle \varepsilon \rangle = 0$$

So our new objective function has 2 undetermined multipliers:

this k is just here for convenience

$$F = S - \alpha g - k\beta h$$

Like before, we have

$$\frac{\partial g}{\partial p_i} = 1$$

$$\frac{\partial S}{\partial p_i} = -k - k\ln p_i$$

$$\frac{\partial h}{\partial p_i} = \varepsilon_i$$

So:

$$\frac{\partial F}{\partial p_i} = -k - k\ln p_i - \alpha - k\beta\varepsilon_i = 0$$

$$-k\ln p_i = k + \alpha + k\beta\varepsilon_i$$

$$\ln p_i = -1 - \frac{\alpha}{k} - \beta\varepsilon_i$$

$$p_i = e^{-1-\alpha/k-\beta\varepsilon_i} = e^{-(1+\frac{\alpha}{k})} e^{-\beta\varepsilon_i}$$

The relative probabilities

$$\frac{p_i}{\sum_i p_i} = \frac{e^{-(1+\frac{\alpha}{k})} e^{-\beta\varepsilon_i}}{\sum_i e^{-(1+\alpha/k)} e^{-\beta\varepsilon_i}} = \frac{e^{-\beta\varepsilon_i}}{\sum_{i=1}^{t} e^{-\beta\varepsilon_i}}$$

This is the Boltzmann distribution law:

$$\boxed{p_i = \frac{e^{-\beta \varepsilon_i}}{\sum_{i=1}^{t} e^{-\beta \varepsilon_i}}} \leftarrow$$

the "score" ε_i is just the energy associated with a state

The denominator is so important, we call it the partition function:

$$q = \sum_{i=1}^{t} e^{-\beta \varepsilon_i}$$

$$\langle \varepsilon \rangle = \sum_{i=1}^{t} \varepsilon_i p_i = \frac{1}{q} \sum_{i=1}^{t} \varepsilon_i e^{-\beta \varepsilon_i}$$

Let's do an example with a biased coin:

$t = 2$

$\varepsilon_H = 2 \qquad \varepsilon_T = 1$ $\leftarrow$ heads are more valuable

(For an unbiased coin, $p_H = p_T = \frac{1}{2}$ and $\langle \varepsilon \rangle = \frac{3}{2}$)

Using the Boltzmann distribution, however:

$$p_T = \frac{e^{-\beta(1)}}{e^{-\beta(1)} + e^{-\beta(2)}} = \frac{x}{x + x^2} \qquad \text{where } x = e^{-\beta}$$

$$p_H = \frac{e^{-\beta(2)}}{e^{-\beta(1)} + e^{-\beta(2)}} = \frac{x^2}{x + x^2}$$

$$\langle \varepsilon \rangle = (1)\, p_T + (2)\, p_H = \frac{x + 2x^2}{x + x^2} = \frac{1 + 2x}{1 + x}$$

Solving for x gives:

$$x = \frac{\langle E \rangle - 1}{2 - \langle E \rangle}$$

If you observe that $\langle E \rangle = \frac{3}{2}$, then $x = 1$ and $P_H = P_T = \frac{1}{2}$ as expected (unbiased).

If instead, you observe $\langle E \rangle = 1.2$, then $x = \frac{1}{4}$ and $P_H = \frac{1}{5}$ with $P_T = 4/5$ (biased).