W01 – The maximum entropy principle

Jaynes (1957)

"Information theory and statistical mechanics"

$X = (x_1 \ldots x_N) \to p_1 \ldots p_N$, $\sum_i p_i = 1$

$$H(X) = -\sum_i p_i \log p_i \quad \text{Entropy}$$

Imagine, you don't know $\{p_i\}$,
you only know the average of some function
$f(x_i)$, that is,

$$\bar{f} = \frac{1}{N} \sum_i f(x_i)\, p_i$$

$\bar{f}$ is an actual #

then what can we say about $\{p_i\}$?
and the average of any other function $g(x)$?

The maximum-entropy principle:

take $\{p_i\}$ that maximizes its entropy
given the constrains

$$L = -\sum_i p_i \log p_i - \lambda\left(\sum_i p_i - 1\right)$$
$$- \lambda_f\left(\sum_i f(x_i) p_i - 1\right)$$

λ, λ_f are Lagrange multipliers

$\frac{\delta L}{\delta p_i} = 0 \rightarrow$ solves for p_i^* that
maximizes entropy

$$\frac{\delta L}{\delta p_i} = -\log p_i - p_i \cdot \frac{1}{p_i} - \lambda - \lambda_f f(x_i) = 0$$

$$\log p_i^* = -\left(1 + \lambda + \lambda_f f(x_i)\right)$$

$$p_i^* = e^{-(1+\lambda + \lambda_f f(x_i))}$$

$$\text{or} \quad \sum_i p_i^* = 1 = \sum_i e^{-[1+\lambda + \lambda_f f(x_i)]}$$

$$= e^{-(1+\lambda)} \sum_i e^{-\lambda_f f(x_i)}$$

$$e^{-(1+\lambda)} = \frac{1}{\sum_i e^{-\lambda_f f(x_i)}}$$

$$\boxed{p_i^* = \frac{e^{-\lambda_f f(x_i)}}{\sum_j e^{-\lambda_f f(x_j)}}}$$

$$\sum_i f(x_i) p_i^* = \bar{f} = \frac{\sum_i f(x_i) e^{-\lambda_f f(x_i)}}{\sum_i e^{-\lambda_f f(x_i)}}$$

---

1) $f = 0$ $\quad p_i^* = \frac{1}{n}$ the uniform distribution

If you know nothing assume nothing → all events equally likely

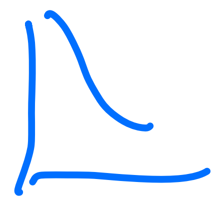
$$H = -\sum_i \frac{1}{n} \log n$$

$$\boxed{H = \log n}$$

i) $f(x) = x$

that you know is mean of the distribution.

$$p^*(x_i) = \frac{e^{-x_i \lambda_f}}{\sum_j e^{-x_j \lambda_f}}$$



an exponential distribution

ii) if you know the var.

$$f(x) = (x-\mu)^2$$

$$\sigma^2 = \langle (x-\mu)^2 \rangle \Rightarrow$$

$$P^* = \frac{e^{-\lambda_f (x-\mu)^2}}{\int e^{-\lambda_f (y-\mu)^2} dy} \rightarrow$$ a Normal distribution