Case 1 : Thermodynamics

Ill-defined problem

$$\sum_i^n P_i = 1$$
$$\sum_i^n E_i P_i = \langle E \rangle$$

n variables
2 equations

$\longrightarrow$

Well-defined problem

$$0 = 1 - \sum_i^n P_i$$
$$0 = \langle E \rangle - \sum_i^n E_i P_i$$

subject to $\max_{P_i}(S) = \max_{P_i}\left(-k_B \sum_i P_i \ln P_i\right)$

Solution to well-defined problem

$$\mathcal{L} = -k_B \sum_i P_i \ln P_i - \lambda_1 \left(1 - \sum_i P_i\right) - \lambda_2 \left(\langle E \rangle - \sum_i E_i P_i\right)$$

$$\delta \mathcal{L} = -k_B \sum_i (\ln P_i + 1)\,\delta P_i - \lambda_1 \left(-\sum_i \delta P_i\right) - \lambda_2 \left(-\sum_i E_i \delta P_i\right) = 0$$

$$\Rightarrow \delta P_i \cdot \left(\sum_i^n \left(-k_B \ln P_i - k_B + \lambda_1 + \lambda_2 E\right)\right) = 0$$

$$\Rightarrow \ln P_i = \frac{\lambda_1 + \lambda_2 E_i - k_B}{k_B}$$

$$\Rightarrow P_i \propto e^{-1 + \frac{\lambda_1}{k_B} + \frac{\lambda_2}{k_B} E_i}$$

$$\Rightarrow \boxed{P_i \propto e^{c \cdot E_i}}$$

$\Rightarrow$ we can learn alot about the system's structure by imposing $\max_{P_i}$, even with only 2 constraints.

Case 2: Highest-entropy language

Consider language with M symbols.

Maximize $H = -\sum_i p_i \log_M p_i$ w/ constraint ~~constraints~~ $\sum p_i = 1 \Rightarrow \sum p_i - 1 = 0$.

$$\therefore \mathcal{L} = -\sum_i p_i \log_M p_i - \lambda\left(\sum_i p_i - 1\right)$$

$$\Rightarrow \quad \delta\mathcal{L} = -\sum_i \left(\log_M p_i + \frac{1}{\ln M}\right)\delta p_i - \lambda\left(\sum \delta p_i\right) = 0$$

$$\Rightarrow \quad -\delta p_i \sum_i \left(\frac{1}{\ln M} + \log_M p_i + \lambda\right) = 0$$

$$\Rightarrow \quad \log_M p_i = -\lambda - \frac{1}{\ln M}$$

$$\Rightarrow \quad p_i = M^{\left(-\lambda - \frac{1}{\ln M}\right)}$$ $\Rightarrow$ <u>all probabilities are the same.</u>

Thus, highest-entropy language is the "noisiest", i.e. all symbols appear with equal likelihood.

Case 3: MaxEnt can give us the Normal distribution

Consider continuous, 1-D entropy:

$$H = -\sum_i P_i \log P_i \xrightarrow[\text{continuous}]{\text{make}} \boxed{H = -\int p(x)\ln p(x)\,dx.}$$

Subject to constraints:

① $\int p(x)\,dx = 1$

② $\int p(x)(x-\mu)^2\,dx = \sigma^2$] → known mean and std. deviation

$$\therefore \mathcal{L} = -\int p(x)\ln p(x)\,dx - \lambda_1\left(\int p(x)\,dx - 1\right) - \lambda_2\left(\int p(x)(x-\mu)^2\,dx - \sigma^2\right)$$

$$\Rightarrow \delta\mathcal{L} = -\int\left[(\ln p(x)+1)\,\delta p(x)\right]dx - \lambda_1\left(\int[\delta p(x)]\,dx\right) - \lambda_2\left(\int (x-\mu)^2\,\delta p(x)\,dx\right) = 0$$

Integrate across all x and divide by $\delta p(x)$:

$$-\ln p(x) + 1 - \lambda_1 - \lambda_2(x-\mu)^2 = 0$$

$$\Rightarrow p(x) = e^{1-\lambda_1-\lambda_2(x-\mu)^2}$$

From constraints 1 and 2: $\lambda_2 = \frac{1}{2\sigma^2}$, $\lambda_1 = \ln\sqrt{2\sigma^2\pi} + 1$

$$\therefore \boxed{p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

Thus, MaxEnt gives us the normal distribution (validates that it behaves like Bayesian and traditional uncertainty estimates.)

(Method can be extended to 3rd, 4th, etc. moments.)