The entropy is maximised for uniform distributions

Assume that the considered random variable can take M outcomes

$$|\mathcal{A}_X| = M$$

Then the maximum entropy over all distributions with M possible outcomes is

$$\max_{P(x)} S[P(x)] = S[P_U(x)] = \ln M$$

where $P_U(x) = \frac{1}{M} \quad \forall x \in \mathcal{A}_X$.

The entropy is given by

$$S = -\sum_{x \in \mathcal{A}_X} P(x) \ln P(x)$$

where $\sum_{x \in \mathcal{A}_x} P(x) = 1.$

We use the lagrangian multiplier ν and we consider the functional

$$\mathcal{F} = S - \nu \left(\sum_{x \in \mathcal{A}_x} P(x) - 1 \right)$$

$$= \underbrace{- \sum_{x' \in \mathcal{A}_x} P(x') \ln P(x')} - \nu \underbrace{\left(\sum_{x' \in \mathcal{A}_x} P(x') - 1 \right)}$$

Differentiating with respect to $P(x)$, ν

$$\boxed{\frac{\partial F}{\partial P(x)} = - \ln P(x) - 1 - \nu = 0} \quad *$$

Two outcomes

$$S = - \underbrace{P(x_1) \ln P(x_1)} - \underbrace{P(x_2) \ln P(x_2)}_{=0}$$

$$\frac{\partial S}{\partial P(x_1)} = - \ln P(x_1) - 1$$

$$\boxed{\frac{\partial F}{\partial \nu} = \sum_{x \in \mathcal{A}_x} P(x) - 1 = 0} \quad (**)$$

From $* \implies P(x) = e^{-1-\nu}$

From $(**)$ $$\sum_{x \in \mathcal{A}_x} P(x) = 1$$

$$\sum_{x \in \mathcal{A}_x} e^{-1-\nu} = 1$$

$$e^{-1-\nu} M = 1$$

$$P(x) = e^{-1-\nu} = \frac{1}{M}$$

$$P(x) = P_U(x) = \frac{1}{M}$$

Exponential families

Consider the maximum entropy distribution in which we impose the constraints

$$\boxed{\sum_{x \in \mathcal{A}_x} P(x)\, f_\mu(x) = \mathbb{E}\left(f_\mu(x)\right) = C_\mu} \quad \mu \in 1,2 \ldots \hat{P}$$

6 possible outcomes $\mathcal{A}_x = \{1,2,3,4,5,6\}$

$$E(x) = 4.5 = \sum_{x \in \{1,2,3,4,5,6\}} P(x) \overbrace{x}^{f_\mu(x) = x} = \underbrace{4.5}_{C_\mu}$$

The maximum entropy distribution is

$$P(x) = \frac{e^{-\sum_{\mu=1}^{\hat{P}} \lambda_\mu f_\mu(x)}}{Z}$$

<u>Example</u> $\hat{P} = 1 \quad f_\mu(x) = x$

$$P(x) = \frac{e^{-\lambda_1 x}}{Z}$$

Z is the normalization constant

$$Z = \sum_{x \in \mathcal{A}_x} e^{-\sum_{\mu=1}^{\hat{p}} \lambda_\mu f_\mu(x)}$$

λ_μ are Lagrangian multipliers fixed by the constraint

$$\mathbb{E}(f_\mu(x)) = C_\mu = \sum_{x \in \mathcal{A}_x} P(x)\, f_\mu(x)$$

$$= \sum_{x \in \mathcal{A}_x} \frac{e^{-\sum_\mu \lambda_\mu f_\mu(x)}}{Z} f_\mu(x)$$

$$= \boxed{-\frac{\partial \ln Z}{\partial \lambda_\mu} = C_\mu}$$

<u>Proof</u> We want to maximize the entropy $S = -\sum_{x \in \mathcal{A}_x} P(x) \ln P(x)$

given the constraints

$$C_\mu = \sum_{x \in \mathcal{A}_\mu} P(x)\, f_\mu(x)$$

or equivalently

$$\left[\sum_{x \in \mathcal{A}_x} P(x)\, f_\mu(x)\right] - C_\mu = 0 \qquad (*) \qquad \mu \in \{1, 2 \ldots \hat{P}\}$$

and the normalization constraint

$$\sum_{x \in \mathcal{A}_x} P(x) = 1$$

$$\left[\sum_{x \in \mathcal{A}_x} P(x)\right] - 1 = 0$$

I maximize the functional

$$\mathcal{F} = S - \sum_{\mu'=1}^{\hat{P}} \lambda_{\mu'} \left(\sum_{x' \in \mathcal{A}_x} P(x')\, f_{\mu'}(x') - C_\mu \right) - \nu \left(\sum_{x' \in \mathcal{A}_x} P(x') - 1 \right)$$

$$S = -\sum_{x' \in \mathcal{A}_x} P(x') \ln P(x')$$

Differentiat with respect to $P(x)$, λ_μ, ν

$$\boxed{\frac{\partial F}{\partial P(x)} = -\ln P(x) - 1 - \sum_{\mu=1}^{p} \lambda_\mu f_\mu(x) - \nu = 0} \quad *$$

For 2 outcomes

$$\frac{\partial}{\partial P(x_1)} \left(P(x_1) f_\mu(x_1) + \underbrace{P(x_2) f_\mu(x_2)} \right) = f_\mu(x_1)$$

$$\frac{\partial F}{\partial \lambda_\mu} = \left(\sum_{x \in \mathcal{A}_x} P(x) f_\mu(x) - C_\mu \right) = 0 \leftarrow \mu \in \{1, 2, \dots p\}$$

$$\frac{\partial F}{\partial \nu} = \left(\sum_{x \in \mathcal{A}_x} P(x) - 1 \right) = 0$$

From $*$ $\quad P(x) = e^{-\nu - 1} \, e^{-\sum_{\mu=1}^{p} \lambda_\mu f_\mu(x)}$

By putting $\quad e^{\nu + 1} = Z \Rightarrow \boxed{P(x) = \frac{1}{Z} e^{-\sum_{\mu=1}^{p} \lambda_\mu f_\mu(x)}}$

Gibbs distribution