Ignorance priors, functional equations and transformation groups

ignorance prior: impose invariant state of knowledge according to a transformation T.

$$\Rightarrow \quad \boxed{p(T(x))\, dT(x) = p(x)\, dx}$$

- simplest case: $\mathcal{M}_1$, $\mathcal{M}_2$. symmetry under exchange of $\mathcal{M}_1$ and $\mathcal{M}_2$.

$$\Rightarrow p(\mathcal{M}_1) = p(\mathcal{M}_2)$$
$$\oplus \quad p(\mathcal{M}_1) + p(\mathcal{M}_2) = 1 \quad \Rightarrow \left\{ \boxed{\begin{array}{l} p(\mathcal{M}_1) = \frac{1}{2} \\ p(\mathcal{M}_2) = \frac{1}{2}. \end{array}} \right.$$

$\mathbb{Z}_2$ symmetry

- "location parameter": $T(x) = x + s. \ \forall s$

$$\Rightarrow dT = dx.$$
$$p(x) = p(x+s) \quad \forall s. \quad \Rightarrow \boxed{p(x) = C}$$

"flat" prior

- "scale parameter" $T(x) = ax \ \forall a.$

$$dT = a\, dx.$$
$$p(x) = a\, p(ax) \quad \Rightarrow \boxed{p(x) = \frac{C}{x}}$$

Jeffreys' prior. an improper prior.

$U(1)$ symmetry

general case: specify a group of transformations.

exercise: the lighthouse problem (notebook) $\Rightarrow$ maximum ignorance for one variable is not the same thing as maximum ignorance on a non-linear function of that variable.

Maximum entropy principle.

indifference about states of equal knowledge $\oplus$ relevant prior information.

$H[p]$? for a source of info producing N finite "words" with proba p_m.

Desiderata: (i) if all words are equiprobable $\left(p_m = \frac{1}{N} \ \forall m\right)$, $H[p]$ must grow with N.

(ii) if words are generated in two steps:

1) choosing a subset of words.
2) choosing a word in this subset,

the entropy is the sum of the entropy assigned to each step.

$\Rightarrow$ (Shannon) theorem: $\boxed{H[p] = -\sum_m p_m \log_2 p_m}$ (up to normalization)

Loaded dice example. fair dice: $p_n = \frac{1}{6}$ for $n \in [\![1,6]\!]$. The principle of indifference was enough.

The mean value after N trials is ~~not~~ 3.5... (say it is 4).
(fair dice)

what is the probability law giving an average of 4?

$\hookrightarrow$ maximum entropy principle: maximise $H[p]$ given 2 constraints:

$$\langle n \rangle_p = \sum_{n=1}^{6} n\, p_n = 4 \quad \text{and} \quad \sum_{n=1}^{6} p_n = 1.$$

(1) brute force way... • get p_5, p_6 as a fct of p_1, p_2, p_3, p_4.
• express $H[p]$ as a fct of p_1, p_2, p_3, p_4.
• differentiate and solve $\frac{\partial H}{\partial p} = 0$ for $n \in [\![1,4]\!]$.

(2) much better solution: method of Lagrange multipliers.

Notebook: Maximum Entropy Principle.

$$\mathcal{L}(\{p_n\}, \lambda, \mu) = -\sum_{n=1}^{6} p_n \log_2 p_n - \lambda\left(\sum_{n=1}^{6} n\, p_n - 4\right) - \mu\left(\sum_{n=1}^{6} p_n - 1\right)$$

$\frac{\partial \mathcal{L}}{\partial \lambda} = 0$ and $\frac{\partial \mathcal{L}}{\partial \mu}$ are our two constraints.

$\frac{\partial \mathcal{L}}{\partial p_n} = 0$ gives $-1 - \ln p_n - \lambda n - \mu = 0.$
$\Rightarrow p_n = \exp(-\lambda n)/Z$ where $\ln Z = 1 + \mu$. $\rightarrow$ now what are λ and Z?

• normalisation constraint: $\sum_{n=1}^{6} p_n = 1$ fixes: $Z = \sum_{n=1}^{6} \exp(-\lambda n) = \frac{1 - e^{-6\lambda}}{e^{\lambda} - 1}$

• constraint on the mean:
$$-\frac{d \ln Z}{d\lambda} = -\frac{1}{Z}\frac{dZ}{d\lambda} = \sum_{n=1}^{6} n \frac{e^{-\lambda n}}{Z} = \sum_{n=1}^{6} n\, p_n = 4.$$

this gives an equation for λ: $\frac{e^{\lambda}}{e^{\lambda} - 1} - \frac{6}{e^{6\lambda} - 1} = 4.$

solution: $e^{\lambda} \simeq 0.83977$, then $p_1, \ldots p_6$ from $\frac{e^{-\lambda n}}{Z}$ | an "observation" giving a probability assignment without Bayes's theorem

Thermodynamics analogy: fair dice = microcanonical ensemble.
loaded dice = canonical ensemble.

$\beta = \frac{1}{k_B T}$, $p_n = \frac{e^{-\beta E_n}}{Z}$ Z = partition function $\simeq$ evidence in Bayesian stats

Cosmic web example