# Maximization of Entropy

Let be the functional $H[P] = -\int_{-\infty}^{\infty} P(x)\log P(x)dx$ we are interest in what distribution P(x) with mean $\mu$ and variance $\sigma^2$ that maximizes $H[P]$, we know that:

$$\int\limits_{-\infty}^{\infty} P(x)dx = 1$$
$$\int\limits_{-\infty}^{\infty} x \cdot P(x)dx = \mu$$
$$\int\limits_{-\infty}^{\infty} x^2 \cdot P(x)dx = \sigma^2 + \mu^2$$

Let be $\lambda_0$, $\lambda_1$ and $\lambda_2$ the Lagrange multipliers associated with the constraints conditions.

$H_L(P) = -\int_{-\infty}^{\infty} P(x)\log P(x)dx - \lambda_0(\int\limits_{-\infty}^{\infty} P(x)dx - 1) - \lambda_1(\int\limits_{-\infty}^{\infty} x \cdot P(x)dx - \mu) - \lambda_2(\int\limits_{-\infty}^{\infty} x^2 \cdot P(x)dx - \sigma^2 - \mu^2)$

Now we are interest in analyse small variations of $H_L[P]$:

$$\delta(H_L[P])) = -\int\limits_{-\infty}^{\infty} (+1 + \log P(x) + \lambda_0 + \lambda_1 x + \lambda_2 x^2)\delta P(x)dx$$
$$\delta(H_L[P])) = 0 \implies (1 + \log P(x) + \lambda_0 + \lambda_1 x + \lambda_2 x^2) = 0$$
$$P(x) = e^{-1-\lambda_0-\lambda_1 x-\lambda_2 x^2}$$

Using the boundary conditions we have:

$$\int\limits_{-\infty}^{\infty} P(x)dx = 1 = \int\limits_{-\infty}^{\infty} e^{-1-\lambda_0-\lambda_1 x-\lambda_2 x^2} = \frac{\sqrt{\pi}\mathrm{e}^{-\frac{(4\lambda_0+4)\lambda_2-\lambda_1^2}{4\lambda_2}}}{\sqrt{\lambda_2}}$$
$$\frac{\sqrt{\lambda_2}}{\sqrt{\pi}} = e^{-\frac{(4\lambda_0+4)\lambda_2-\lambda_1^2}{4\lambda_2}}$$

$$\int_{-\infty}^{\infty} x \cdot P(x)dx = \mu = \int_{-\infty}^{\infty} xe^{-1-\lambda_0-\lambda_1 x-\lambda_2 x^2} = -\frac{\sqrt{\pi}\lambda_1 \mathrm{e}^{-\frac{(4\lambda_0+4)\lambda_2-\lambda_1^2}{4\lambda_2}}}{2\lambda_2^{\frac{3}{2}}}$$

$$\frac{\sqrt{\lambda_2}}{\sqrt{\pi}} = e^{-\frac{(4\lambda_0+4)\lambda_2-\lambda_1^2}{4\lambda_2}} = -\frac{2\lambda_2^{\frac{3}{2}}\mu}{\sqrt{\pi}\lambda_1}$$

$$-2\mu\lambda_2 = \lambda_1$$

$$\int_{-\infty}^{\infty} x^2 \cdot P(x)dx = \sigma^2 + \mu^2 = \int_{-\infty}^{\infty} x^2 e^{-1-\lambda_0-\lambda_1 x-\lambda_2 x^2} = \frac{\sqrt{\pi}\left(2\lambda_2+\lambda_1^2\right)\mathrm{e}^{-\frac{(4\lambda_0+4)\lambda_2-\lambda_1^2}{4\lambda_2}}}{4\lambda_2^{\frac{5}{2}}}$$

$$4\lambda_2^2(\sigma^2+\mu^2) = 2\lambda_2 + \lambda_1^2 = 2\lambda_2 + 4\mu^2\lambda_2^2$$

$$\lambda_2 = \frac{1}{2\sigma^2} \text{ , } \lambda_1 = -\frac{\mu}{\sigma^2}$$

$$\frac{\sqrt{\lambda_2}}{\sqrt{\pi}} = e^{-\frac{(4\lambda_0+4)\lambda_2-\lambda_1^2}{4\lambda_2}} = e^{-(1+\lambda_0)+\frac{\lambda_1^2}{4\lambda_2}}$$

$$e^{-(1+\lambda_0)} = \frac{\sqrt{\lambda_2}}{\sqrt{\pi}} e^{-\frac{\lambda_1^2}{4\lambda_2}} = \frac{\sqrt{\lambda_2}}{\sqrt{\pi}} e^{-\mu^2\lambda_2}$$

So we have:

$$P(x) = e^{-1-\lambda_0-\lambda_1 x-\lambda_2 x^2} = e^{-1-\lambda_0} \cdot e^{-\lambda_1 x-\lambda_2 x^2} = \frac{\sqrt{\lambda_2}}{\sqrt{\pi}} e^{-\mu^2\lambda_2-\lambda_1 x-\lambda_2 x^2}$$

$$P(x) = \frac{\sqrt{\lambda_2}}{\sqrt{\pi}} e^{-\lambda_2(\mu^2-2\mu x+x^2)} = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

So Gaussian distribution maximizes the Shannon entropy, this is the reason why in data science we use a lot of normal distributions, using any other distribution means implicit use of unknown information.