Problem 1

(1)
Solving the maximum entropy distribution with constraints is essentially

$$\begin{aligned} \max & \int p(x)\ln p(x)\mathrm{d}x \\ \text{s.t.} & \int p(x)\mathrm{d}x = 1 \\ & \int xp(x)\mathrm{d}x = 0 \\ & \int x^2 p(x)\mathrm{d}x = 1 \end{aligned} \tag{1}$$

Lemma(functional derivative): We first show how to take the derivative of a function. Suppose we have

$$F(p(x)) = \int p(x)f(x)\mathrm{d}x. \tag{2}$$

By substituting $p(x)$ with $p(x)+\epsilon\eta(x)$, we have

$$F(p(x)+\epsilon\eta(x)) = \int p(x)f(x)\mathrm{d}x + \epsilon\int \eta(x)f(x)\mathrm{d}x, \tag{3}$$

and thus,

$$\frac{\partial F}{\partial p(x)} = f(x). \tag{4}$$

Similarly, we have

$$\begin{aligned} G(p(x)) &= \int p(x)\ln p(x)\mathrm{d}x \\ \frac{\partial G}{\partial p(x)} &= \ln p(x) + 1. \end{aligned} \tag{5}$$

From the fact that the above optimization question is a concave maxmization one, we construct its Lagrangian function and set its derivative w.r.t. λ to zero.

$$\begin{aligned} L(x,\lambda) &= \int p(x)\ln p(x)\mathrm{d}x - \lambda_1(\int p(x)\mathrm{d}x - 1) - \lambda_2\int xp(x)\mathrm{d}x - \lambda_3(\int x^2p(x)\mathrm{d}x - 1) \\ \frac{\partial L}{\partial \lambda} &= \ln p(x) + 1 - \lambda_1 - \lambda_2 x - \lambda_3 x^2 \\ &= 0. \end{aligned} \tag{6}$$

We obtain

$$p(x) = \exp(-1+\lambda_1+\lambda_2 x+\lambda_3 x^2) \tag{7}$$

To solve λ, we substitute the above $p(x)$ into constraints

$$\begin{aligned} \int \exp(-1+\lambda_1+\lambda_2 x+\lambda_3 x^2)\mathrm{d}x = 1 \\ \int x\exp(-1+\lambda_1+\lambda_2 x+\lambda_3 x^2)\mathrm{d}x = 0 \\ \int x^2\exp(-1+\lambda_1+\lambda_2 x+\lambda_3 x^2)\mathrm{d}x = 1 \end{aligned} \tag{8}$$

We notice that $\lambda = (1-\frac{1}{2}\ln(2\pi\sigma^2),\ 0,\ \frac{1}{2\sigma^2})$ from $N(0,1)$ does satisfy above equations(and thus no need to check conditions for Lagrangian method).
This shows that $X\ N(0,1)$ is the maximum entropy distribution under such constraints.

Problem 1 continued on next page...

(2)
Similar to **(1)**, we have

$$\begin{aligned} p(x) &= \exp(-1 + \sum_{0}^{n} \lambda_i x^i) \\ \text{s.t.} \int p(x) \mathrm{d}x &= 1 \\ \int x^i p(x) \mathrm{d}x &= m_i \quad \text{for} 1 \leq i \leq n \end{aligned} \tag{9}$$

Notice that in some cases, such distribution does not exist and the maximum entropy only serves as an upper bound.