1 (a) $X \sim N(0,1)$ is max. entropy distribution s.t. $\mathbb{E}X = 0$ $\mathbb{E}X^2 = 1$

This is an optimization problem

$$\min \int_{\mathbb{R}} \log p(x)\, p(x)\, dx \quad \text{s.t.} \int_{\mathbb{R}} x^2 p(x) dx = 1 \text{ and } \int_{\mathbb{R}} p(x) dx = 1$$

Lagrangian

$$\mathcal{L}(p(x), \lambda_1, \lambda_2) = \int_{\mathbb{R}} \log p(x)\, p(x) dx + \lambda_1 \left(\int_{\mathbb{R}} p(x) dx - 1 \right) + \lambda_2 \left(\int_{\mathbb{R}} x^2 p(x) dx - 1 \right)$$

KKT condition

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial p} = 0 \\ \int p(x) dx = 1 \\ \int x^2 p(x) dx = 1 \end{cases}$$

$$\frac{\partial \mathcal{L}}{\partial p} = 0 \Rightarrow 1 + \log p(x) + \lambda_1 + \lambda_2 x^2 = 0$$

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

$$\int p(x) dx = 1 \Rightarrow e^{\lambda_1 + 1} = \sqrt{\frac{\pi}{\lambda_2}}$$

$$\int x^2 p(x) dx = 1 \Rightarrow \sqrt{\pi}\, e^{-(\lambda_1 + 1)} = 2 \lambda_2^{\frac{3}{2}}$$

$$\Rightarrow \lambda_2 = \frac{1}{2} \quad e^{\lambda_1 + 1} = \sqrt{2\pi}$$

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \sim N(0,1)$$

$\frac{\partial \mathcal{L}^2}{\partial p^2} = \frac{1}{p(x)} > 0$ So this is maximum

(b) Find max entropy distribution s.t. $\mathbb{E}X^i = m_i$, $i = 1, \cdots k$

$$\min \int_{\mathbb{R}} \log p(x)\, p(x) dx \quad \text{s.t.} \int_{\mathbb{R}} x^i p(x) dx = m_i, \ i = 1, \cdots k. \ \int p(x) dx = 1$$

$$\mathcal{L}(p, \lambda) = \int_{\mathbb{R}} \log p(x)\, p(x) dx + \sum_{i=1}^{k} \left(\int_{\mathbb{R}} x^i p(x) dx - m_i \right) \lambda_i + \mu \left(\int p(x) dx - 1 \right)$$

KKT $\begin{cases} \frac{\partial L}{\partial p} = 0 \\ \int x^i p(x) dx = m_i \\ \int p(x) dx = 1 \end{cases}$

$$\frac{\partial L}{\partial p} = \log p + 1 + \mu + \sum_{i=1}^{k} x^i \lambda_i = 0 \qquad \frac{\partial L^2}{\partial p^2} = \frac{1}{p} > 0$$

$$p(x) = e^{-(\mu + 1 + \sum_{i=1}^{k} \lambda_i x^i)}$$

$p(x)$ should satisfy $\int p(x) dx = 1$

and $\int x^i p(x) dx = m_i$, $i = 1 \cdots k$

but the integral doesn't have closed form. #

2. $p(y_i|\theta_i) = \exp(y_i b(\theta_i) + c(\theta_i) + d(y_i)) \quad i=1,\cdots n$

$\mathbb{E}(Y_i) = \mu_i(\theta_i) \quad g(\mu_i) = x_i^T \beta \quad \beta \in \mathbb{R}^d$

(1) $g(\mu_i) = \eta_i$, S is score func of β

Show $S_j = \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{Var(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \quad j=1,\cdots d$

As definition.

$S_j = \frac{\partial}{\partial \beta_j} \log L(y_1 \cdots y_n | \theta_1 \cdots \theta_n) = \frac{\partial}{\partial \beta_j} \sum_{i=1}^n \log L(y_i|\theta_i)$

$= \sum_{i=1}^n \frac{\partial \log L(y_i|\theta_i)}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}$

$= \sum_{i=1}^n \frac{\partial \log L(y_i|\theta_i)}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} x_{ij} \frac{\partial \mu_i}{\partial \eta_j} \quad (*)$

We have $1 = \int \exp(y_i b(\theta_i) + c(\theta_i) + d(y_i)) \, dy_i$

$0 = \frac{\partial}{\partial \theta_i} \int \exp(y_i b(\theta_i) + c(\theta_i) + d(y_i)) dy_i = \int \exp(y_i b(\theta_i) + c(\theta_i) + d(y_i))(y_i b'(\theta_i) + c'(\theta_i)) dy_i$

$= b'(\theta_i)\mu_i + c'(\theta_i) \quad \mu_i = -\frac{c'(\theta_i)}{b'(\theta_i)}$

$0 = \frac{\partial^2}{\partial \theta_i^2} \int \exp(y_i b(\theta_i) + c(\theta_i) + d(y_i)) = \int \exp(y_i b(\theta_i) + c(\theta_i) + d(y_i))(y_i b''(\theta_i) + c''(\theta_i))$

$\exp(y_i b(\theta_i) + c(\theta_i) + d(y_i)) (y_i b'(\theta_i) - b'(\theta_i)\mu_i)^2 dy_i$

$+ \exp(y_i b(\theta_i) + c(\theta_i) + d(y_i)) + (b'(\theta_i))^2 Var Y_i = \frac{-c'(\theta_i) b''(\theta_i) + c''(\theta_i) b'(\theta_i)}{b'(\theta_i)} + b'(\theta_i)^2 Var Y_i$

$= b''(\theta_i)\mu_i + c''(\theta_i) + (b'(\theta_i))^2 Var Y_i$

$\frac{\partial \mu_i}{\partial \theta_i} = \frac{-c''(\theta_i) b'(\theta_i) + c'(\theta_i) b''(\theta_i)}{(b'(\theta_i))^2} = Var Y_i \, b'(\theta_i) \quad (1)$

$\frac{\partial \log L(y_i|\theta_i)}{\partial \theta_i} = y_i b'(\theta_i) + c'(\theta_i) = (y_i - \mu_i) b'(\theta_i) \quad (2)$

Noticing (1) (2)

We have $(*) \Rightarrow S_j = \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{Var Y_i} \frac{\partial \mu_i}{\partial \eta_j}$ #

(2) I : Fisher information

$$I_{jk} = \mathbb{E}(S_j S_k) = \sum_{i=1}^{n} \frac{x_{ij} x_{ik}}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \quad \forall 1 \le j, k \le d.$$

$$\mathbb{E}(S_j S_k) = \mathbb{E} \sum_{i=1}^{n} \sum_{t=1}^{n} \frac{(y_i - \mu_i) x_{ij}}{Var(Y_i)} \frac{(y_t - \mu_t) x_{tk}}{Var(Y_t)} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \mu_t}{\partial \eta_t}$$

$$= \mathbb{E} \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2 x_{ij} x_{ik}}{(Var\, Y_i)^2} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \quad (Y_i \text{ independent})$$

$$= \sum^{n} \frac{x_{ij} x_{ik}}{Var\, Y_i} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \quad \#$$