Exponential Family

지수족 분포는 기본적으로 다음과 같이 표현 가능하다.

$$p(x|\eta) = h(x)g(\eta)\exp\eta^T u(x)$$

위 식은 pdf이기 때문에 확률의 정의를 만족한다.

$$\int p(x|\eta) = 1$$

$$\int h(x)g(\eta)\exp[\eta^T u(x)] = 1$$

$$g(\eta)\int h(x)\exp[\eta^T u(x)dx] = 1$$

지수족은 기본적으로 완비충분통계량 u(x)를 포함하고 있기에 위 형태로 인수분해한다면 완비충분통계량을 쉽게 구할 수 있으며 이에 따라 MVUE도 쉽게 구할 수 있다.

1. Bernoulli Distribution

베르누이 분포는 다음과 같다.

$$p(x|\mu) = Bern(x|\mu) = \mu^x (1-\mu)^{1-x}$$

양변에 ln을 취해주면,

$$\ln p(x|\mu) = x\ln\mu + (1-x)\ln(1-\mu)$$

양변에 exp를 취해주면,

$$p(x|\mu) = \exp[x\ln\mu + (1-x)\ln(1-\mu)] = (1-\mu)\exp[\ln(\frac{\mu}{1-\mu})x]$$

위 식에서 $\eta = \ln(\frac{\mu}{1-\mu})$로 두면, $\mu = \frac{\exp(\eta)}{1+\exp(\eta)} = \frac{1}{1+\exp(-\eta)} = \sigma(\eta)$로 표현 가능하다.

그런데 $1-\mu = 1-\sigma(\eta) = \sigma(-\eta)$이므로 다음과 같이 표현가능하다.

$$p(x|\mu) = \sigma(-\eta)\exp(\eta x)$$

$$\text{where } u(x) = x, h(x) = 1, g(\eta) = \sigma(-\eta)$$

2. Gaussian Distribution

가우시안 분포는 식 자체에 exp항이 있어 바로 표현가능하다.

$$\begin{aligned} p(x|\mu,\sigma^2) &= \frac{1}{(2\pi\sigma^2)^{1/2}}\exp[-\frac{1}{2\sigma^2}(x-\mu)^2] \\ &= \frac{1}{(2\pi\sigma^2)^{1/2}}\exp[-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}\mu^2] \end{aligned}$$

$$\text{where, } \eta = \begin{pmatrix} \mu/\sigma^2 \\ -1/2\sigma^2 \end{pmatrix}$$

$$g(\eta) = (-2\eta_2)^{1/2}\exp(\frac{\eta_1^2}{4\eta_2})$$

$$u(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}$$

$$h(x) = (2\pi)^{-1/2}$$

3. Maximum Likelihood Estimator and Sufficient Statistics

MLE를 찾기 위해 일반적인 지수족의 pdf식을 미분하면,

$$\frac{d}{d\eta}(g(\eta)\int h(x)\exp[\eta^T u(x)]dx = 1)$$

$$\nabla g(\eta)\int h(x)\exp[\eta^T u(x)]dx + g(\eta)\int h(x)\exp[\eta^T u(x)]u(x)dx = 0$$

$\int h(x)\exp[\eta^T u(x)]dx = \frac{1}{g(\eta)}$이므로,

$$-\frac{1}{g(\eta)}\nabla g(\eta) = g(\eta)\int h(x)\exp[\eta^T u(x)]u(x)dx = E(u(x))$$

$\frac{1}{g(\eta)}\nabla g(\eta)$는 $\nabla \ln(g(\eta))$이므로,

$$-\nabla \ln g(\eta) = g(\eta)\int h(x)\exp[\eta^T u(x)]u(x)dx = E(u(x))$$

만약 여러 개의 관측 데이터라면, likelihood는 다음과 같다.

$$p(x|\eta) = (\prod_{n=1}^{N} h(x[n]))g(\eta)^N \exp[\eta^T \sum_{n=1}^{N} u(x[n])]$$

앞선 경우와 같이 미분 후 0으로 두고 정리하면 다음과 같고, 해당 식을 만족하는 η값이 MLE가 된다.

$$-\nabla \ln g(\eta_{ML}) = \frac{1}{N}\sum_{n=1}^{N} u(x[n])$$