Exponential Family of Distributions

- Parametric density functions

$$P_G(\boldsymbol{x}|\boldsymbol{\theta}) = e^{\boldsymbol{\theta}\cdot\boldsymbol{x} - G(\boldsymbol{\theta})}\; P_0(\boldsymbol{x})$$

- $\boldsymbol{\theta}$ and $\boldsymbol{x}$ vectors in R^d
- Cumulant function $G(\boldsymbol{\theta})$ assures normalization

$$G(\boldsymbol{\theta}) = \ln \int e^{\boldsymbol{\theta}\cdot\boldsymbol{x}}\; P_0(\boldsymbol{x})\, d\boldsymbol{x}$$

- $G(\boldsymbol{\theta})$ is convex function on convex set $\boldsymbol{\Theta} \subseteq R^d$
- G characterizes members of the family
- $\boldsymbol{\theta}$ is *natural* parameter

- Expectation parameter

$$\boldsymbol{\mu} = \int_{\boldsymbol{x}} \boldsymbol{x} P_G(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x} = E_{\boldsymbol{\theta}}(\boldsymbol{x}) = g(\boldsymbol{\theta})$$

  where $g(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} G(\boldsymbol{\theta})$

- Second convex function $F(\boldsymbol{\mu})$ on space $g(\boldsymbol{\Theta})$

$$F(\boldsymbol{\mu}) \;\; = \;\; \boldsymbol{\theta} \cdot \boldsymbol{\mu} - G(\boldsymbol{\theta})$$

- $G(\boldsymbol{\theta})$ and $F(\boldsymbol{\mu})$ are *convex conjugate* functions
- Let $f(\boldsymbol{\mu}) = \nabla_{\boldsymbol{\mu}} F(\boldsymbol{\mu})$
- $f(\boldsymbol{\mu}) = g^{-1}(\boldsymbol{\mu})$

Primal & Dual Parameters

natural		expectation
paramater		parameter

$$\boldsymbol{\theta} \quad \underset{f}{\overset{g}{\rightleftarrows}} \quad \boldsymbol{\mu}$$

$$G(\boldsymbol{\theta}) \qquad\qquad F(\boldsymbol{\mu})$$

- $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$ are dual parameters
- Parameter transformations
  $g(\boldsymbol{\theta}) = \boldsymbol{\mu}$ and $f(\boldsymbol{\mu}) = \boldsymbol{\theta}$ [A,BN]

Gaussian (unit variance)

$$\begin{aligned} P(\boldsymbol{x}|\boldsymbol{\theta}) &\sim e^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{x})^2} \\ &= e^{\boldsymbol{\theta}\cdot\boldsymbol{x}-\frac{1}{2}\boldsymbol{\theta}^2} e^{\frac{1}{2}\boldsymbol{x}^2} \end{aligned}$$

Cumulant function: $G(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^2$

Parameter transformations:

$g(\boldsymbol{\theta}) = \boldsymbol{\theta} = \boldsymbol{\mu}$ and $f(\boldsymbol{\mu}) = \boldsymbol{\mu} = \boldsymbol{\theta}$

Dual convex function: $$\begin{aligned} F(\boldsymbol{\mu}) &= \boldsymbol{\theta}\cdot\boldsymbol{\mu} - G(\boldsymbol{\theta}) \\ &= \frac{1}{2}\boldsymbol{\mu}^2 \end{aligned}$$

Square loss: $L_t(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta}_t - \boldsymbol{x}_t)^2$

Bernoulli

Examples x_t are coin flips in $\{0, 1\}$

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

μ is the probability (expectation) of 1

Natural parameter: $\theta = \ln \frac{\mu}{1-\mu}$

$$P(x|\theta) = \exp\left(\theta x - \ln(1+e^\theta)\right)$$

Cumulant function: $G(\theta) = \ln(1+e^\theta)$

Parameter transformations:

$$\mu = g(\theta) = \frac{e^\theta}{1+e^\theta} \text{ and } \theta = f(\mu) = \ln \frac{\mu}{1-\mu}$$

Dual function: $F(\mu) = \mu \ln \mu + (1-\mu)\ln(1-\mu)$

Log loss:
$$\begin{aligned} L_t(\theta) &= -x_t\theta + ln(1+e^\theta) \\ &= -x_t \ln \mu - (1-x_t)\ln(1-\mu) \end{aligned}$$

Poisson

Examples x_t are natural numbers in $\{0, 1, \dots\}$

$$P(x|\mu) = \frac{e^{-\mu}\mu^x}{x!}$$

μ is expectation of x

Natural parameter: $\theta = \ln \mu$

$$P(x|\theta) = \exp\left(\theta x - e^\theta\right) \frac{1}{x!}$$

Cumulant function: $G(\theta) = e^\theta$

Parameter transformations:

$$\mu = g(\theta) = e^\theta \text{ and } \theta = f(\mu) = \ln \mu$$

Dual function: $F(\mu) = \mu \ln \mu - \mu$

$$\begin{aligned} \text{Loss:} \quad L_t(\theta) &= -x_t\theta + e^\theta + \ln x_t! \\ &= -x_t \ln \mu + \mu + \ln x_t! \end{aligned}$$