Exponential Family

$p(x|\theta) = g(T(x), \theta)\, h(x, T(x))$

Let $\vec{x} = (x_1, \dots, x_k)$ and $\vec{\eta} \in H \subseteq \mathbb{R}^D$

$$p(\vec{x}|\vec{\eta}) = h(\vec{x}) \exp\left[\vec{\eta}^T T(\vec{x}) - A(\vec{\eta})\right]$$

Canonical parameters

sufficient statistics

log partitition function

$$= h(\vec{x}) \exp\left[\vec{\eta}^T T(\vec{x}) - \log Z(\vec{\eta})\right]$$

$\frac{1}{Z(\vec{\eta})}$

EF Examples

① Categorical (e.g. Multinomial$(1, \vec{\pi})$)

Let $\vec{x} = (x_1, \dots, x_k)$ be a 1-hot vector, $\sum_i \pi_i = 1$

$\approx [0,0,1,0,0,0]$

$$p(\vec{x}|\vec{\pi}) = \pi_1^{x_1} \pi_2^{x_2} \cdots \pi_k^{x_k}$$

$$= \exp(\ln(\uparrow))$$

$$= \exp\left[\sum_{k=1}^{K} x_k \ln \pi_k\right]$$

Exp Fam: $\vec{\eta} = [\ln \pi_1, \dots, \ln \pi_k]$

$T(\vec{x}) = \vec{x}$

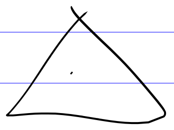
$A(\vec{\eta}) = 0$

$h(\vec{x}) = 1$

Given $\vec{\eta}$: $\vec{\pi} = [\exp[\eta_1], \dots, \exp[\eta_k]]$ $\quad \sum \exp(\eta) = 1$

$\mathbb{E}[X_i] = \pi_i \qquad Var[X_i] = \pi_i(1-\pi_i)$

② Dirichlet



Let $\vec{\pi} \in M-1$ simplex

and $\alpha_1, \dots, \alpha_k > 0$

$$p(\vec{\pi}|\vec{\alpha}) = \frac{1}{B(\vec{\alpha})} \prod_{i=1}^{K} \pi_i^{\alpha_i - 1}$$

Beta function

$$= \exp(\ln(\uparrow))$$

$$= \exp\left[\left(\sum_{i=1}^{K} (\alpha_i - 1) \ln \pi_i\right) - \ln B(\vec{\alpha})\right]$$

Exp Fam: $\vec{\eta} = [\alpha_1 - 1, \dots, \alpha_k - 1]$

$T(\vec{\pi}) = [\ln \pi_1, \dots, \ln \pi_k]$

$A(\vec{\eta}) = \ln B(\vec{\alpha})$

$h(\vec{x}) = 1$

Given $\vec{\eta}$: $\vec{\alpha} = [\eta_1 + 1, \dots, \eta_k + 1]$

Cumulant Generating Function, $A(\vec{\eta})$

Shown for scalar η i.e. $D = 1$

1) First cumulant

$$\frac{dA(\eta)}{d\eta} = \frac{d}{d\eta} \log Z(\eta)$$

$$= \frac{1}{Z(\eta)} \cdot \frac{d}{d\eta} Z(\eta)$$

$$= \frac{1}{Z(\eta)} \frac{d}{d\eta} \left(\int h(\vec{x}) \exp[\eta^T T(\vec{x})] \, dx \right)$$

$$= \int T(\vec{x}) \frac{h(\vec{x}) \exp[\eta^T T(\vec{x})] \, dx}{Z(\eta)}$$

$$= \int T(\vec{x}) p(\vec{x} | \vec{\eta}) \, dx$$

$$= \mathbb{E}_{p(x|\eta)}[T(\vec{x})] \longleftarrow \text{Mean}$$

2) Second Cumulant

$$\frac{d^2 A(\eta)}{d\eta^2} = \frac{d}{d\eta} \int T(x) \underbrace{p(x|\eta)}_{\exp(\eta^T T(x) - A(\eta)) h(x)} \, dx$$

$$= \int T(x) \exp[\eta^T T(x) - A(\eta)] h(\vec{x}) \cdot \left(T(x) - \frac{dA(\eta)}{d\eta} \right) dx$$

$$= \int T(\vec{x})^2 p(\vec{x} | \vec{\eta}) \, dx - \frac{dA(\eta)}{d\eta} \int T(x) p(x|\eta) \, dx$$

$$= \mathbb{E}[T(x)^2] - \mathbb{E}[T(x)]^2$$

$$= Var(T(\vec{x}))$$

MLE for EF

Given iid data $D = (\vec{x}^{(1)}, \dots, \vec{x}^{(N)})$

$$\log p(D|\eta) = \log \prod_{i=1}^{N} p(x^{(i)}|\eta)$$

$$= \log \prod_{i=1}^{N} h(\vec{x}^{(i)}) \exp(\vec{\eta}^T T(\vec{x}^{(i)}) - A(\vec{\eta}))$$

$$= \left(\sum_{i=1}^{N} \log h(\vec{x}^{(i)}) \right) + \left(\eta^T \sum_{i=1}^{N} T(\vec{x}^{(i)}) \right) + N A(\vec{\eta})$$

$$\frac{d \log p(D|\eta)}{d\eta} = \sum_{i=1}^{N} T(\vec{x}^{(i)}) - N \frac{dA(\eta)}{d\eta} = 0$$

scalar

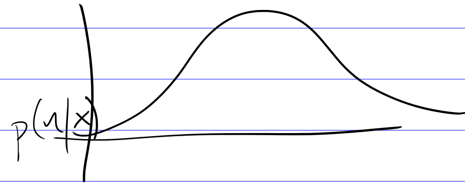
$$\frac{dA(\eta)}{d\eta} = \frac{1}{N} \sum_{i=1}^{N} T(\vec{x}^{(i)}) = \mathbb{E}[T(\vec{x})]$$

$$\hat{\mu}_{MLE} = \frac{1}{N} \sum_{i=1}^{N} T(\vec{x}^{(i)})$$

$$\vec{\eta} = \psi(\vec{\mu})$$

$$p(x, T(x), \Theta) = p(x, \theta) = \psi_1(\ \ )\psi_2(\ \ )$$

Baysian Est.



Part 1: Conjugacy

(x, η, ϕ are all vectors)

$$p(x/\eta) \propto \exp[\eta^T T(x) - A(\eta)] \qquad \text{likelihood}$$
$$p(\eta/\phi) \propto \exp\{\phi^T T(\eta) - A(\phi)] \qquad \text{prior}$$

$$p(\eta|x,\phi) \propto p(x|\eta)\,p(\eta|\phi)$$
$$= \exp[\eta^T T(x) + \phi^T T(\eta) - A(\eta) - A(\phi)]$$

Suppose $\eta \equiv T(\eta)$

$$= \exp[\underbrace{T(\eta)^T}_{\text{s.s.}}\underbrace{(T(x)+\phi)}_{\text{natural parameters}} - \underbrace{A(\eta) - A(\phi)}_{A(\eta,\phi)}]$$

Exp Fam: $\eta =$
$T(x) =$

Part 2: Estimation

(Left as an exercise)

Observe that for $p(D|\eta)$
The s.s. N and $T(D)$ where

$$T(D) = \left[\sum_{i=1}^{N} T_1(x^{(i)}), \ldots, \sum_{i=1}^{N} T_k(x^{(i)})\right]$$

$p(\eta/D)$

GLIMs

$$p(y|\eta,\sigma) = h(y,\sigma)\exp\left[\frac{1}{\sigma^2}\left(\eta\cdot y - A(\eta)\right)\right]$$

Let $\eta \equiv \psi(\mu)$

$\mu = \psi^{-1}(\eta)$

Let $\xi = \theta^T\vec{x}$

$\mu = f(\xi) = f(\theta^T\vec{x})$

☆ $\Rightarrow \eta = \psi(f(\theta^T\vec{x}))$

$$p(y|\vec{x},\vec{\theta},\sigma) = h(y,\phi)\exp\left[\frac{1}{\sigma^2}\left(\psi(\cancel{f}(\theta^T\vec{x}))\cdot y - A(\eta)\right)\right]$$

GLIMs w/ Canonical Response

$$f(\cdot) = \psi^{-1}(\cdot) \Rightarrow \eta = \theta^T\vec{x}$$

Examples

$\vec{x} \longrightarrow Y$

$$\mathbb{E}[Y|\vec{X}] = \mu = f(\theta^T\vec{x})$$

	f	Y
Linear Reg.	identity	$N(\mu,\sigma^2)$
Log. Reg	logistic	Bernoulli(μ)
Probit Reg.	cum. Gau.	〃
Mult. Reg	logistic	Mult([illegible], μ)