Formulation

$$p(y;\eta) = b(y)e^{\eta^T T(y) - a(\eta)}$$

- Base measure $b(y)$ when exponent is zero
- Sufficient statistic $T(y)$
- Log-partition function $a(\eta)$ normalizes to sum 100%
- Why exponential tilting: lots of convenience
- η natural parameter

Bernoulli Distribution

$$p(y;\phi) = \phi^y(1-\phi)^{1-y} = \exp\log\phi^y(1-\phi)^{1-y} = \exp(y\log\phi + (1-y)\log(1-\phi))$$
$$= \exp(\log\frac{\phi}{1-\phi}y + \log(1-p))$$

- $\eta = \log\frac{\phi}{1-\phi}$, invert to get $\phi = \frac{1}{1+e^{-\eta}}$ logistic function
- $T(y) = y$
- $a(\eta) = -\log(1-\phi) = -\log\left(1-\frac{1}{1+e^{-\eta}}\right) = -\log\left(\frac{e^{-\eta}}{1+e^{-\eta}}\right) = \log(1+e^{-\eta})$
- Hypothesis: $h_\theta(x) = E[y|x,\theta] = \phi = \frac{1}{1+e^{-\eta}} = \frac{1}{1+e^{-\theta^T x}}$

Univariate Gaussian

$$p(y;\mu) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{(y-\mu)^2}{2}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}e^{\mu y-\frac{\mu^2}{2}}$$

- $\eta = \mu$
- $T(y) = y$
- $a(\eta) = \frac{\mu^2}{2} = \frac{\eta^2}{2}$
- $b(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$
- Hypothesis: $h_\theta(x) = E[y|x,\theta] = \mu = \eta = \theta^T x$

Properties

1. $\log p(y;\eta)$ is concave in η, MLE is concave in η (NNL convex in η)
2. $E[y;\eta] = \frac{\partial a(\eta)}{\partial\eta}$, first moment
3. $Var[y;\eta] = \frac{\partial^2 a(\eta)}{\partial^2\eta}$, second moment
4. $a(\eta)$ is moment generating function

Generalized Linear Model

Three assumptions

- $y|x;\theta \sim ExpFamily(\eta)$
- $h_\theta(x) = E[y|x,\theta]$
- Natural parameter is a linear combination: $\eta = \theta^T x$
    - Beyond GLM, can be arbitrarily complex (neural net)