COS 513, *Scribe Notes*, November 8, 2010

Caleb Bastian

Continuation from previous lecture on $\boxed{\textbf{Exponential families}}$

$$p(x|\xi) = h(x)exp\{\eta^{\intercal}t(x) - a(\eta)\}$$

where $X_n \sim p(x|\eta)$. Take Log, resulting in

$$log\left\{p(X_{1:N}|\xi)\right\} = \sum_{n=1}^{N}\left[log\{h(X_n)\} + \eta^{\intercal}\left(\sum_n t(X_n)\right) - Na(\eta)\right].$$

Take derivative of both sides and set equal to zero,

$$\frac{\partial l}{\partial \eta} = \sum_n t(X_n) - N\nabla_\eta a(\eta) = 0,$$

so

$$\nabla_\eta a(\eta) = \frac{1}{N}\sum_n t(X_n),$$

which is the empirical average of the sufficient statistics, which is equivalent to

$$\mathbb{E}_{\hat{\eta}}\{t(X)\} = \frac{1}{N}\sum_n t(X_n).$$

$\boxed{\textbf{Bayesian setting}}$

- place a prior on the natural parameter
- compute the posterior given data $X_{1:N}$

Conjugate Prior: a prior for which the posterior is in the same family. We previously saw Gaussian/Gaussian *(prior/likelihood)* case.

Beta-Bernoulli Conjugacy:

Bernoulli Distribution: $p(x|\pi) = \pi^x(1-\pi)^{1-x}$, where $x \in \{0, 1\}$

Beta Distribution: $p(\pi|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}$, where $\alpha > 0, \beta > 0, \pi \in (0, 1)$.

Note the Gamma function $\Gamma\{\}$ is the real-valued extension of the factorial function.

Some properties of the Beta distribution:
- $\mathbb{E}\{\pi|\alpha,\beta\} = \frac{\alpha}{\alpha+\beta}$
- can assume uniform, symmetric, and skewed distributions

So, defining our model via

$$\pi \sim Beta(\alpha, \beta)$$
$$X_n \sim Bernoulli(\pi),$$

The *Posterior* distribution is:

$$\begin{aligned} p(\pi|X_{1:N}) &\propto p(\pi)p(X_{1:N}|\pi) \\ &= p(\pi)\prod_{n=1}^{N} p(X_n|\pi) \\ &= \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}\right)\prod_{n=1}^{N}\pi^{X_n}(1-\pi)^{1-X_n} \\ &\propto \pi^{(\alpha+\sum_n X_n-1)}(1-\pi)^{(\beta+N-\sum_n X_n-1)} \\ &= Beta\left(\alpha+\sum_n X_n, \beta+N-\sum_n X_n\right) \end{aligned}$$

Note that

$$\mathbb{E}(\pi|X_{1:N}) = \frac{\alpha+\sum_n X_n}{\alpha+\beta+\sum_n X_n + N - \sum_n X_n} = \frac{\alpha+\sum_n X_n}{\alpha+\beta+N}$$

Conjugate Prior for Exponential Family

$\eta \sim$ Conj.(λ)
$X_n \sim$ Exponential family (η), $n = 1, ..., N$

$$p(X_n|\eta) = h(x)exp\Big\{\eta^\intercal t(x) - a(\eta)\Big\}$$

- $$p(\eta|\lambda) = h(\eta)exp\Big\{\lambda_1^\intercal \eta + \lambda_2(-a(\eta)) - a_c(\lambda)\Big\}$$

($\bullet$) is the conjugate prior for the Exponential family. λ_1 is a vector with the same dimension as η, $\lambda_2 \in \mathbb{R}$, and a_c is log-normalized.

In the conjugate prior, natural parameter: $< \lambda_1, \lambda_2 >$, where λ_1 is dim(η) and $\lambda_2 \in \mathbb{R}$.

sufficient statistics are $< \eta, -a(\eta) >$

To confirm this,

$$p(\eta|X_{1:N}) \propto p(\eta|\lambda) * \prod_n^N p(X_n|\eta)$$
$$= h(\eta) * exp\Big\{\lambda_1^\intercal \eta + \lambda_2(-a(\eta)) - a_c(\lambda)\Big\} * Nh(x)exp\Big\{\eta^\intercal\Big(\sum_n t(X_n)\Big) - Na(\eta)\Big\}$$
$$\propto h(\eta) * exp\Big\{\Big(\lambda_1 + \sum_n t(X_n)\Big)^\intercal \eta + (\lambda_2 + N)(-a(\eta))\Big\}$$

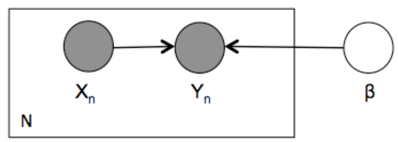
This is the same form as the prior. So, the posterior is in the same family with

$$\hat{\lambda_1} = \lambda_1 + \sum_n t(X_n)$$
$$\hat{\lambda_2} = \lambda_2 + N$$

other conjugates: *(prior/likelihood)*
Normal (on μ)-Inverse Wishart (on Σ) / Normal

Dirichlet/Multinomial
Gamma/Poisson
Beta/Bernoulli

Generalized Linear Models



-Observed input X enters the model through a linear function:

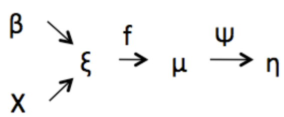
$$\xi = \beta^{\intercal} X$$

-Conditional mean of the response y is a function of ξ

$$\mathbb{E}(Y|X) \triangleq \mu = f(\beta^{\intercal} X)$$

-Response Y is drawn from an exponential family with mean μ.

Diagram:



Ψ maps the mean to the natural parameter.
Usually, we work with an over-dispersed exponential family

$$p(y|\eta) = h(y, \delta) exp \left\{ \frac{\eta^{\intercal} y - a(\eta)}{\delta} \right\}$$

Linear Regression:

$$p(Y|X) = \frac{1}{\sqrt{2\pi\sigma^2}} exp \left\{ -\frac{1}{2\sigma^2} (Y - \beta^{\intercal} X)^2 \right\}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} exp \left\{ \frac{-Y}{2\sigma^2} \right\} * exp \left\{ \frac{Y(\beta^{\intercal} X) - \frac{1}{2}(\beta^{\intercal} X)^2}{\sigma^2} \right\}$$

where

$\delta = \sigma^2$

$h(\delta, Y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\{\frac{-Y}{2\sigma^2}\}$

$\eta = \beta^{\intercal} X$

$a(\eta) = \eta^2/2$

Ψ: Identity

f: Identity

<u>2 decisions to make to define model</u>:
1. Choose the exponential family distribution of Y (this determines Ψ).
2. Choose the response function f ("link function")

<u>Canonical Response Function</u>: $f = \Psi^{-1}$.
Here, the linear function of X <u>is</u> the natural parameter. $\eta = X^{\intercal}\beta$. The only "choice" we now have is 1.