Exponential Family

- Suppose $Y_1, \cdots, Y_n$ are independent random variables.

- Let $f(y_i; \theta_i, \phi)$ be PMF or PDF of $Y_i$, where $\phi$ is a scale parameter.

- If we can write

$$f(y_i; \theta_i, \phi) = \exp \left[ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right],$$

then we call the PMF or the PDF $f(y_i; \theta_i)$ is an exponential family.
Normal Distribution

Assume $X \sim N(\mu, \sigma^2)$. Then, $E(X) = \mu$. $\sigma$ is a scale parameter. Then, the PDF is

$$
\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \exp\left\{\frac{\mu x - \mu^2/2}{\sigma^2} + \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{x^2}{2\sigma^2}\right]\right\}.
$$

We may use $\theta = \mu$, $b(\theta) = \theta^2/2$, $\phi = \sigma^2$, $y = x$, $a(\phi) = \phi$, $c(y, \phi) = -(1/2) \log(2\pi\phi) - y^2/(2\phi)$. 
Binomial Distribution

Assume $X \sim Bin(n, p)$. Then, $E(X) = np$. The PMF is

$$\binom{n}{x} p^x (1 - p)^{n-x}$$

$$= \exp\left\{ x \log \frac{p}{1-p} + n \log(1 - p) + \log \binom{n}{x} \right\}.$$  

So $y = x$, $\theta = \log[p/(1 - p)]$, $b(\theta) = n \log(1 + e^{\theta})$, $\phi = 1$, $a(\phi) = 1$, $c(y, \phi) = \log \binom{n}{x}$. 
Poisson Distribution

Assume $X \sim \text{Poisson}(\lambda)$. Then, $E(X) = \lambda$. The PMF is

$$
\frac{\lambda^x}{x!} e^{-\lambda}
$$

$$
= \exp\{x \log(\lambda) - \lambda - \log(x!}\}.
$$

So $y = x$, $\theta = \log(\lambda)$, $b(\theta) = e^\theta$, $\phi = 1$, $a(\phi) = 1$, $c(y, \phi) = -\log(x!)$. 
Assume $X \sim \Gamma(\alpha, \beta)$, $\beta$ is known. Then, $E(X) = \alpha/\beta$. Then PMF is

$$
\frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}
= \exp\{\alpha \log x + \alpha \log(\beta) - \log(\Gamma(\alpha)) - \log(x) - \beta x\}.
$$

Assume $\alpha$ is known. Then $\theta = -\beta$, $y = x$, $b(\theta) = -\alpha \log(\beta)$, $\phi = 1$ and $a(\phi) = 1$. 
Negative Binomial Distribution

Assume \( X \sim NB(k, p) \). The PDF is

\[
\binom{x - 1}{k - 1} p^k (1 - p)^{x-k}
\]

\[
= \exp\{x \log(1 - p) + k \log \frac{p}{1 - p} + \log \left( \frac{x - 1}{k - 1} \right) \},
\]

for \( x = k, k + 1, \cdots \). So \( y = x, \theta = \log(1 - p), b(\theta) = -k \log[(1 - e^\theta)/e^\theta], \phi = 1, a(\theta) = 1 \) and \( c(y, \phi) = \log \left( \frac{x - 1}{k - 1} \right) \). Then,

\[
E(X) = b'(\theta) = \frac{k}{1 - e^\theta} = \frac{k}{p}
\]

and

\[
V(X) = b''(\theta) = \frac{ke^\theta}{(1 - e^\theta)^2} = \frac{k(1 - p)}{p^2}.
\]
The definition of Generalized Linear Model (GLM) is based on exponential family. There are three components in GLM. They are

- **Random component.** Assume the distributions of the sample. Such as normal, binomial, Poisson and etc.

- **Systematic component.** Describe the form of predictor (independent) variables. Such as
  \[ \eta = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{ip} x_{ip}. \]

- **Link function.** Connect the unknown parameters to model. Such as
  \[ g[\mu(\theta)] = \eta \]

for some \( g(\cdot) \), where \( \mu(\theta) \) is the expected value.
It uses $\theta = \eta$ is the canonical link.

- Normal: identity link $g(\mu) = \mu$.
- Binomial: logistic link $g(\mu) = \log \frac{\mu}{1-\mu}$.
- Poisson: log link $g(\mu) = \log(\mu)$.
- Gamma: negative inverse link $g(\mu) = -1/\mu$.
- Negative binomial: $g(\mu) = \log[\mu/k(1 + \mu/k)]$. 
• The most important cases are binomial and Poisson.

• Canonical link is just one of the link functions.

• Estimation is based on maximum likelihood.
There are three link functions for binomial.

- **Logistic link.**
  \[
  \log \frac{p_i}{1 - p_i} = \beta_0 + \sum_{j=1}^{p} x_{ij} \beta_j
  \]
called logistic linear model or logistic regression.

- **Inverse CDF link.**
  \[
  F^{-1}(p_i) = \beta_0 + \sum_{j=1}^{p} x_{ij} \beta_j.
  \]
  If \( F = \Phi \), it is the probit link, called probit model.

- **Complementary loglog link.**
  \[
  \log[- \log(1 - p_i)] = \beta_0 + \sum_{j=1}^{p} x_{ij} \beta_j.
  \]
Logistic Regression

Consider the simplest case. That is

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_i.$$

Suppose $\hat{\beta}$ and $\hat{\beta}_1$ are the MLEs.

- Odds ratio: as $x$ increases $a$ units, the estimate of odds ratio is $e^{a\hat{\beta}_1}$.
- The significance of the odds ratio can be directly read by the $p$-value of $\beta_1$.
- Confidence interval can also be derived respectively.
### Table 1: Blood Pressure and Heart Disease

<table>
<thead>
<tr>
<th>Blood Pressure</th>
<th>Heart Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
</tr>
<tr>
<td>&lt; 117</td>
<td>3</td>
</tr>
<tr>
<td>117 – 126</td>
<td>17</td>
</tr>
<tr>
<td>127 – 136</td>
<td>12</td>
</tr>
<tr>
<td>137 – 146</td>
<td>16</td>
</tr>
<tr>
<td>147 – 156</td>
<td>12</td>
</tr>
<tr>
<td>157 – 166</td>
<td>8</td>
</tr>
<tr>
<td>167 – 186</td>
<td>16</td>
</tr>
<tr>
<td>&gt; 186</td>
<td>8</td>
</tr>
</tbody>
</table>
Goodness of Fit

Let \( \hat{n}_{ij} \) be the predicted counts of the model.

- Pearson \( \chi^2 \) is

\[
X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}.
\]

- Loglikelihood ratio \( \chi^2 \) is

\[
G^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \log\left(\frac{n_{ij}}{\hat{n}_{ij}}\right).
\]