Generalized Linear Models

The general linear model approach used so far:

- consists of specifications for the mean and covariance structure;

- mean response is a linear function of parameters and covariates;

- has strong theoretical support when the responses are multivariate normal;

- works well when the response is continuous.
Generalized Linear Models

When the response is *discrete*, or the mean response is naturally modeled as *nonlinear* in the parameters, we need an alternative.

Other members of the *exponential family* of distributions provide various alternatives.

Each distribution has a natural way to relate the mean response to the parameters and covariates, and implies a relationship between the mean response and the variance.

Begin with *independent* responses; longitudinal data handled by allowing for dependence or with random effects.
Distributions

The most important distributions are:

- normal
- Bernoulli
- binomial
- Poisson

where the last three are used for modeling count data.
Variance Functions

For each of these, the variance is in the form $\phi v(\mu)$ for some dispersion parameter $\phi$ and variance function $v(\mu)$:

- **normal**: $\text{Var}(Y) = \sigma^2$, so $v(\mu) = 1$, $\phi = \sigma^2$;

- **Bernoulli**: $E(Y) = p$, $\text{Var}(Y) = p(1 - p)$, so $v(\mu) = \mu(1 - \mu)$, $\phi = 1$;

- **binomial**: $E(Y) = np$, $\text{Var}(Y) = np(1 - p)$, so $v(\mu) = \mu(n - \mu)/n$, $\phi = 1$;

- **Poisson**: $E(Y) = \text{Var}(Y) = \mu$, so $v(\mu) = \mu$, $\phi = 1$. 
Link Functions

The *link function* links the mean response to the parameters and covariates:

\[ g(\mu) = \eta = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p = \mathbf{X}'\beta. \]

Each distribution has a *canonical* link function, derived from its natural parametrization in the exponential family.

Other link functions can also be used, when the subject matter suggests one.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Variance Function, ( v(\mu) )</th>
<th>Canonical Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>( v(\mu) = 1 )</td>
<td>Identity: ( \mu = \eta )</td>
</tr>
<tr>
<td>Bernoulli</td>
<td>( v(\mu) = \mu(1 - \mu) )</td>
<td>Logit: ( \log\left(\frac{\mu}{1-\mu}\right) = \eta )</td>
</tr>
<tr>
<td>Poission</td>
<td>( v(\mu) = \mu )</td>
<td>Log: ( \log(\mu) = \eta )</td>
</tr>
</tbody>
</table>
Logistic Regression

- Response $Y$ is Bernoulli: $P(Y = 1) = 1 - P(Y = 0) = \mu$.

- Variance function: $\text{Var}(Y) = \mu(1 - \mu)$ (whence $\phi = 1$).

- Covariates: $\mu = \mu(X, \beta)$.

- Require $0 \leq \mu(X, \beta) \leq 1$: rules out linear model $\mu = X'\beta$. 
Logistic Link Function

The log odds ratio

\[
\log \left( \frac{\mu}{1 - \mu} \right) = \text{logit}(\mu) = \eta
\]

can be solved for \( \mu \):

\[
\mu = \frac{\exp(\eta)}{1 + \exp(\eta)}.
\]

Any value for the log odds gives \( \mu \) satisfying \( 0 < \mu < 1 \).

So a linear model for \( \text{logit}(\mu) \)

\[
\eta = X'\beta
\]

always gives valid values of \( \mu \).
Interpretation of Parameters

If one covariate changes by $+1$ and the others are not changed, $\eta$ changes by the corresponding $\beta$ coefficient.

If $\mu \approx 0$, logit($\mu$) $\approx$ log($\mu$), so log($\mu$) also changes by approximately $\beta$.

If $\beta$ is also small, a change of $\beta$ in log($\mu$) is approximately a fractional change of $\beta$ in $\mu$.

For example, if $\beta = 0.1$, a change of $+1$ in the covariate implies a 10% increase in $\mu$.

At the other extreme, if $\mu \approx 1$, logit($\mu$) $\approx$ $-\log(1 - \mu)$. So now, if $\beta$ is still 0.1, a change of $+1$ in the covariate implies a 10% decrease in $1 - \mu$. 
Binomial Case

If $Y_1, Y_2, \ldots, Y_m$ are independent Bernoulli with the same covariates (or at least, the same $\mu$), then

$$Y = Y_1 + Y_2 + \cdots + Y_m$$

is Bin$(m, \mu)$.

- The logistic link function is still reasonable.

- The binomial variance function is

  $$\text{Var}(Y) = m\mu(1 - \mu).$$
• In practice, $Y$ is often over-dispersed:

\[ \text{Var}(Y) > m\mu(1 - \mu). \]

• The variance function therefore often includes the dispersion parameter:

\[ v(\mu) = \phi m\mu(1 - \mu). \]

• Over-dispersion can be caused by:

  **dependence:** conditionally on the covariates, the individual $Y_j$s are not independent (e.g., common cause of failure);

  **heterogeneity:** the individual $Y_j$s do not have exactly the same covariates, or covariates are measured with error.