Delta Method and Generalized Linear Models

Patrick Lam

Suppose we have a random variable X, and we know $E(X)$ and $\text{Var}(X)$.

Let Y be another random variable such that $Y = g(X)$.

How do we find $E(Y)$ and $\text{Var}(Y)$? (useful for reparameterizations and quantities of interest)

- Find $E(Y)$ via $\int_{-\infty}^{\infty} g(x)p(x)dx$
- Simulation
- Delta Method

If g(X) is linear...

this is pretty straightforward:

$$\begin{aligned} Y &= a + bX \\ E(Y) &= a + bE(X) \\ \mathrm{Var}(Y) &= b^2 \mathrm{Var}(X) \end{aligned}$$

What if g(X) isn't linear?

We will use the **delta method**, which relies on a linear approximation to $g(X)$ near the mean of X.

Delta Method

Denote μ_X as the mean of X. We use a first-order Taylor series approximation around μ_X:

$$\begin{aligned} Y &= g(X) \\ &\approx g(\mu_X) + (X - \mu_X)g'(\mu_X) \end{aligned}$$

$$\begin{aligned} E(Y) &\approx E[g(\mu_X)] + E[(X - \mu_X)g'(\mu_X)] \\ &= g(\mu_X) \text{ since } E(X - \mu_X) = 0 \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(Y) &\approx \mathrm{Var}[g(\mu_X)] + \mathrm{Var}[(X - \mu_X)g'(\mu_X)] \\ &= 0 + \mathrm{Var}[Xg'(\mu_X)] - \mathrm{Var}[\mu_X g'(\mu_X)] \\ &= \mathrm{Var}(X)[g'(\mu_X)]^2 \end{aligned}$$

One More Step...

We have

$$E(Y) \approx g(\mu_X)$$

but we know generally (from Jensen's inequality)

$$E(g(X)) \neq g(E(X))$$

so we use the second order Taylor expansion for $E(Y)$

$$\begin{aligned} Y &\approx g(\mu_X) + (X-\mu_X)g'(\mu_X) + \frac{1}{2}(X-\mu_X)^2 g''(\mu_X) \\ E(Y) &\approx E[g(\mu_X)] + E[(X-\mu_X)g'(\mu_X)] + \frac{1}{2}E[(X-\mu_X)^2 g''(\mu_X)] \\ &= g(\mu_X) + \frac{1}{2}\mathrm{Var}(X)g''(\mu_X) \end{aligned}$$

So we have

$$
\begin{aligned}
E(Y) &\approx g(\mu_X) + \frac{1}{2}\mathrm{Var}(X)g''(\mu_X) \\
\mathrm{Var}(Y) &\approx \mathrm{Var}(X)[g'(\mu_X)]^2
\end{aligned}
$$

How good the approximations are depend on how nonlinear $g(X)$ is in the neighborhood of μ_X and on the size of Var(X).

Generalized Linear Models

All of the models we've talked about so far (and for the rest of the class) belong to a class called **generalized linear models (GLM)**.

Three elements of a GLM:

- A distribution for Y
- A linear predictor $X\beta$
- A link function that relates the linear predictor to the mean of the distribution.

Steps to running a GLM:

1. Specify a distribution for Y

Assume our data was generated from some distribution.

Examples:

- Continuous and Unbounded: Normal
- Binary: Bernoulli
- Event Count: Poisson
- Duration: Exponential
- Ordered Categories: Normal with observation mechanism
- Unordered Categories: Multinomial

2. Specify a linear predictor

$$X\beta = \beta_0 + x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k$$

3. Specify a link function

The link function relates the linear predictor to the mean of the distribution for Y.

Let $g(\cdot)$ be the link function and let $E(Y) = \theta$ be the mean of distribution for Y.

$$\begin{aligned} g(\theta) &= X\beta \\ \theta &= g^{-1}(X\beta) \end{aligned}$$

Note that we usually use the **inverse link function** $g^{-1}(X\beta)$ rather than the link function.

This is the systematic component that we've been talking about all along.

Example Link Functions

Identity:

- Link: $\mu = X\beta$
- Inverse Link: $\mu = X\beta$

Inverse:

- Link: $\lambda^{-1} = X\beta$
- Inverse Link: $\lambda = (X\beta)^{-1}$

Logit:

- Link: $\ln\left(\frac{\pi}{1-\pi}\right) = X\beta$
- Inverse Link: $\pi = \frac{1}{1+e^{-X\beta}}$

Probit:

- Link: $\Phi^{-1}(\pi) = X\beta$
- Inverse Link: $\pi = \Phi(X\beta)$

Log:

- Link: $\ln(\lambda) = X\beta$
- Inverse Link: $\lambda = \exp(X\beta)$

4. Estimate Parameters via ML

Do it.

5. Quantities of Interest

1. Simulate parameters from multivariate normal.
2. Run $X\beta$ through inverse link function to get expected values.
3. Draw from distribution of Y for predicted values.

Binary Dependent Variable

Let our dependent variable be a binary random variable that can take on values of either 0 or 1.

1. Specify a distribution for Y

$$\begin{aligned} Y_i &\sim \text{Bernoulli}(\pi_i) \\ p(\mathrm{y}|\boldsymbol{\pi}) &= \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i} \end{aligned}$$

2. Specify a linear predictor: $X\beta$

3. Specify a link (or inverse link) function.
    - Logit: $\pi_i = \frac{1}{1+e^{-x_i\beta}}$
    - Probit: $\pi_i = \Phi(x_i\beta)$
    - Complementary Log-log (cloglog): $\pi_i = 1 - \exp(-\exp(X\beta))$
    - Scobit: $\pi_i = (1 + e^{-x_i\beta})^{-\alpha}$

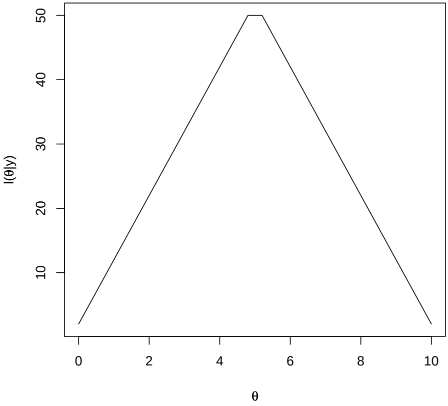
4. Estimate parameters via ML.

$$l(\beta|\mathbf{y}) = \sum_{i=1}^{n} y_i \ln\left(\frac{1}{1+e^{-x_i\beta}}\right) + (1-y_i)\ln\left(1 - \frac{1}{1+e^{-x_i\beta}}\right)$$

5. Simulate Quantities of Interest

Identification

Suppose we have the following log-likelihood function:



What's wrong with this?

The Identification Problem

There are more than one set of parameters that give the same maximum likelihood value, so our model is **unidentified**.

Ordered Probit/Logit:

- If we estimate all the βs and τs, we can get many sets of parameters that have the same likelihood.
- We can make our model identified in two ways:
  - Fix $\beta_0 = 0$ and estimate all the τs (basically don't estimate an intercept)
  - Fix $\tau_1 = 0$ and estimate an intercept.