Lecture Notes 7/5/2012

July 5, 2012

Derivations of OLS and MLE optimization problems

Model:

$$y_i = x_i'\beta + \varepsilon_i$$

in vector form:

$$y = X\beta + \varepsilon$$

We want to estimate β because it has some sort of economic significance: e.g. elasticity of demand for a good; monetary returns to additional years of schooling; etc.

Estimation procedure for recovering β depends on what assumptions we make on the error term, ε.

OLS

Assume ε is mean-zero and uncorrelated with X. We want to find the β that minimizes the distance between $X\beta$ and y. This problem can be written as

$$\begin{aligned}
\hat{\beta} &= \arg\min_{\beta} \sum_i \varepsilon_i^2 \\
&= \arg\min_{\beta} \varepsilon'\varepsilon \\
&= \arg\min_{\beta} (y - X\beta)'(y - X\beta)
\end{aligned}$$

We can solve this optimization problem with calculus:

$$\begin{aligned}
& \min_{\beta} (y - X\beta)'(y - X\beta) \\
= & \min_{\beta} y'y - \beta'X'y - y'X\beta - \beta'X'X\beta
\end{aligned}$$

taking first-order conditions for β:

$$\begin{aligned}
[\beta] \quad &: \quad -\frac{\partial}{\partial\beta}\beta' X' y - \frac{\partial}{\partial\beta} y' X\beta - \frac{\partial}{\partial\beta}\beta' X' X\beta \\
&= \quad -X'y - X'y - 2\left(X'X\right)\hat{\beta} \\
0 \quad &= \quad -2X'y - 2\left(X'X\right)\hat{\beta} \\
X'y \quad &= \quad \left(X'X\right)\hat{\beta} \\
\hat{\beta} \quad &= \quad \left(X'X\right)^{-1} X'y
\end{aligned}$$

which is the OLS estimator. Second-order conditions for a minimum:

$$\begin{aligned}
\frac{\partial^2}{\partial\beta\partial\beta'} \quad &> \quad 0 \\
\left(X'X\right) \quad &> \quad 0 \ \text{(trivially satisfied)}
\end{aligned}$$

Maximum Likelihood

Now let's look at a linear regression model when $\varepsilon_i \overset{iid}{\sim} N\left(0, \sigma^2\right)$. The likelihood function is

$$\begin{aligned}
\mathscr{L}\left(\varepsilon_i;\beta,\sigma\right) \quad &= \quad \prod_i f\left(\varepsilon_i\right) \\
&= \quad \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right) \\
&= \quad \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{\sum_i \varepsilon_i^2}{2\sigma^2}\right) \\
&= \quad \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{\varepsilon'\varepsilon}{2\sigma^2}\right), \ \text{and taking logs,} \\
\ell\left(y,X;\beta,\sigma\right) \quad &= \quad -\frac{n}{2}\ln\left(2\pi\right) - n\ln\left(\sigma\right) - \frac{1}{2\sigma^2}\left(y - X\beta\right)'\left(y - X\beta\right)
\end{aligned}$$

So the objective function is to maximize the likelihood (or the log likelihood) with respect to β and σ.

$$\max_{\beta,\sigma} -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)$$

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta} &= \frac{1}{2\sigma^2}\frac{\partial}{\partial \beta}(y-X\beta)'(y-X\beta) \\
0 &= \frac{1}{2\sigma^2}\left[-2X'y - 2\left(X'X\right)\hat{\beta}\right] \\
\hat{\beta} &= \left(X'X\right)^{-1}X'y \\
\frac{\partial \ell}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{(y-X\beta)'(y-X\beta)}{\sigma^3} \\
0 &= -\frac{n}{\sigma} + \frac{(y-X\beta)'(y-X\beta)}{\sigma^3} \\
\frac{n}{\sigma} &= \frac{(y-X\beta)'(y-X\beta)}{\sigma^3} \\
\hat{\sigma}^2 &= \frac{\left(y-X\hat{\beta}\right)'\left(y-X\hat{\beta}\right)}{n}
\end{aligned}$$

Second-order conditions require that the matrix

$$\begin{bmatrix} \frac{\partial^2 \ell}{\partial\beta\partial\beta'} & \frac{\partial^2 \ell}{\partial\beta\partial\sigma} \\ \frac{\partial^2 \ell}{\partial\sigma\partial\beta'} & \frac{\partial^2 \ell}{(\partial\sigma)^2} \end{bmatrix} = \begin{bmatrix} -\frac{X'X}{\sigma^2} & \frac{(X'X\beta - X'Y)}{\sigma^3} \\ \frac{\left(\hat{\beta}'X'X - Y'X\right)}{\sigma^3} & \frac{n}{\sigma^2} - \frac{3\left(Y-X\hat{\beta}\right)'\left(Y-X\hat{\beta}\right)}{\sigma^4} \end{bmatrix}$$

be negative definite. This is satisfied, but we won't go through it today.

Logit

Now, instead of being continuous, our dependent variable y only takes on two values: 0 and 1. We can think of y as being distributed Bernoulli with probability p, where $p = \Pr(y=1)$. The pdf of the Bernoulli distribution is $f(y;p) = p^y(1-p)^{1-y}$. Fitting this to our model above, we get

$$\begin{aligned}
\Pr(y=1) &= \Pr(y>0) \\
&= \Pr(X\beta + \varepsilon > 0) \\
&= \Pr(X\beta > -\varepsilon) \\
&= \Pr(-\varepsilon < X\beta) \\
&= \Pr(\varepsilon < X\beta) \text{ if } \varepsilon \text{has a symmetric distribution} \\
&= F(X\beta)
\end{aligned}$$

When we assume $\varepsilon \sim Logistic$, then we get $F(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}$. So $p = \Pr(y=1) = \frac{\exp(X\beta)}{1+\exp(X\beta)}$. Plugging this into our likelihood function, we get

$$\begin{aligned}
\mathscr{L}(y,X;\beta) &= \prod_i p_i^{y_i}(1-p_i)^{1-y_i} \\
&= \prod_i \left(\frac{\exp(X\beta)}{1+\exp(X\beta)}\right)^{y_i}\left(\frac{1}{1+\exp(X\beta)}\right)^{1-y_i} \\
\ell &= \sum_i y_i \ln\left(\frac{\exp(X\beta)}{1+\exp(X\beta)}\right) + (1-y_i)\ln\left(\frac{1}{1+\exp(X\beta)}\right)
\end{aligned}$$

The first order conditions are

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta} &= \frac{\partial \ell}{\partial \beta} y\left[X\beta - \ln(1+\exp(X\beta))\right] - (1-y)\left[\ln(1+\exp(X\beta))\right] \\
0 &= \frac{\partial \ell}{\partial \beta} y[X\beta] - \ln(1+\exp(X\beta)) \\
0 &= X'y - \left[\frac{1}{1+\exp(X\beta)} X\exp(X\beta)\right] \\
0 &= X'y - \left[X\frac{\exp(X\beta)}{1+\exp(X\beta)}\right] \\
0 &= X'y - X'p \\
0 &= X'(y-p) \\
0 &= X'\left(y - \frac{\exp(X\beta)}{1+\exp(X\beta)}\right)
\end{aligned}$$

Now we need to solve for β:

$$X'y = X'\left(\frac{\exp(X\beta)}{1+\exp(X\beta)}\right)$$

But this, unfortunately, does not have a closed-form solution. Hence, we need to use numerical methods.

Probit

For probit, we maintain the assumption that y can only take on two values and is hence distributed Bernoulli. The only difference comes in the fact that now we assume $\varepsilon \sim N(0,1)$, so we get $F(x) = \Phi(x)$. The likelihood is then

$$\begin{aligned}
\mathscr{L}(y,X;\beta) &= \prod_i p_i^{y_i}(1-p_i)^{1-y_i} \\
&= \prod_i \Phi(X\beta)^{y_i}(1-\Phi(X\beta))^{1-y_i} \\
\ell &= \sum_i y_i \ln(\Phi(X\beta)) + (1-y_i)\ln(1-\Phi(X\beta))
\end{aligned}$$

The first order conditions are

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta} &= \frac{\partial \ell}{\partial \beta}\left[y \ln\left(\Phi\left(X\beta\right)\right) + \left(1-y\right)\ln\left(1-\Phi\left(X\beta\right)\right)\right] \\
0 &= y\left(\frac{X\phi\left(X\beta\right)}{\Phi\left(X\beta\right)}\right) - \left(1-y\right)\left(\frac{X\phi\left(X\beta\right)}{1-\Phi\left(X\beta\right)}\right) \\
0 &= X'y\left(\frac{\phi\left(X\beta\right)}{\Phi\left(X\beta\right)}\right) - X'\left(1-y\right)\left(\frac{\phi\left(X\beta\right)}{1-\Phi\left(X\beta\right)}\right)
\end{aligned}$$

So the solution we seek is

$$X'y\left(\frac{\phi\left(X\beta\right)}{\Phi\left(X\beta\right)}\right) = X'\left(1-y\right)\left(\frac{\phi\left(X\beta\right)}{1-\Phi\left(X\beta\right)}\right)$$

where we need to solve for β. Again, no closed-form solution.