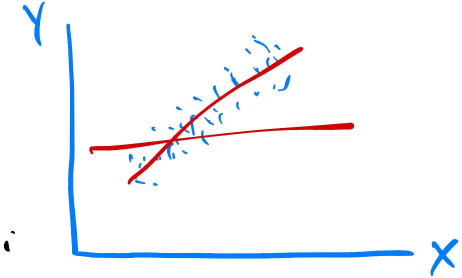
Lecture Note 6: Maximum Likelihood

Least squares vs. MLE



Random var. X with pdf $f(x;\theta)$

Sample X_i iid: $\{X_i\}_{i=1}^{N}$ → each i has x_i

Likelihood:

$$\max_{\tilde{\theta}} L = \max_{\tilde{\theta}} \prod_{i=1}^{N} f(x_i;\tilde{\theta})$$

$$\max_{\tilde{\theta}} \ln L = \max_{\tilde{\theta}} \sum_{i=1}^{N} \ln\left[f(x_i;\tilde{\theta})\right]$$

Solution: $\hat{\theta}$ satisfies $\frac{\partial \ln L}{\partial \hat{\theta}} = 0$

Properties of MLE:

1. Consistency: $\hat{\theta} \xrightarrow{p} \theta$
2. Asymptotic normality (CLT): $\hat{\theta} \xrightarrow{d} N(\theta, \Sigma)$
3. Asymptotic efficiency

Bernoulli $X_i = \begin{cases} 1 & \text{with prob } p \\ 0 & \text{" \quad " \quad } 1-p \end{cases}$

Sample $N=3$, i.i.d., $(1,1,0)$

$$L = \Pr[X_1 = 1] \cdot \Pr[X_2 = 1] \cdot \Pr[X_3 = 0]$$

$$= p \cdot p \cdot (1-p)$$

$$= p^2(1-p)$$

MLE: $\max\limits_{\tilde{p}} L = \max\limits_{\tilde{p}} \tilde{p}^2(1-\tilde{p})$

$\max\limits_{\tilde{p}} \ln L = \max\limits_{\tilde{p}} 2\ln(\tilde{p}) + \ln(1-\tilde{p})$

FOC: $\dfrac{d \ln L}{d \hat{p}} = \dfrac{2}{\hat{p}} - \dfrac{1}{1-\hat{p}} = 0 \Rightarrow \hat{p} = \dfrac{2}{3}$

N Bernoulli variables, S successes, $N-S$ failures

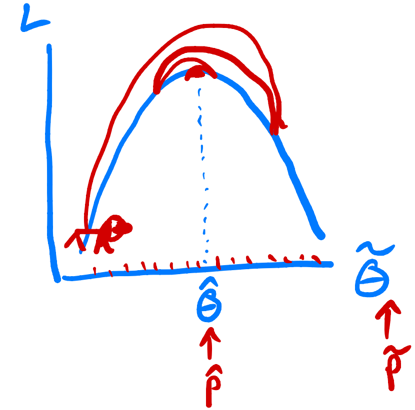
$$L = p^S (1-p)^{N-S}$$

$$\ln L = S \ln(p) + (N-S)\ln(1-p)$$

$$\hat{p} = \frac{S}{N}$$

Methods for finding MLE

1. Analytic optimization
2. (Undirected) grid search
3. (Directed) numerical optimization



Newton - Raphson