Monday April 5 – Lect. 32 Maximum Likelihood Est.

$Y \in \Gamma \qquad f_\theta(y) \qquad \theta \in \Lambda$

$\Longrightarrow f_\theta(\cdot)$ as a prob. density on Γ

$\Longrightarrow f_\cdot(y)$ as a funct. of θ on Λ where y is now fixed.

↳ here called the likelihood function

Might slightly change notation

$$f_\theta(y) = f(y, \theta) \text{ or } f(y; \theta)$$

Then the max likelihood estimate (ML)

$$\hat{\theta}_{ML}(y) = \underset{\theta \in \Lambda}{\operatorname{argmax}} \, f_\theta(y)$$

possible: * there is no maximizing θ

* may not be unique.

If $G(\cdot)$ is an increasing funct. on $\mathbb{R}$

$$\underset{\theta}{\operatorname{argmax}}\, f_\theta(y) = \underset{\theta}{\operatorname{argmax}}\, G\left(f_\theta(y)\right)$$

Pick $G(\cdot)$ to simplify... Most common to use natural log....

$$\hat{\theta}_{ML}(y) = \underset{\theta}{\operatorname{argmax}}\, \underbrace{\log f_\theta(y)}_{\text{log-likelihood function}}.$$

Subject to "smoothness conditions" a necc. cond. that $\hat{\theta}_{ML}(y)$ be an ML estimator is that it is a soln to the <u>likelihood equation</u>

$$\left.\frac{\partial}{\partial\theta} \log f_\theta(y)\right|_{\theta = \hat{\theta}_{ML}(y)} = 0.$$

- If θ is vector ... replace $\frac{\partial}{\partial \theta}$ with gradient...
- $\frac{\partial}{\partial \theta} \log f_\theta(y)$ is called the <u>score function</u>. The roots have interesting properties.

<u>Example</u> $Y = [Y_1 \ldots Y_n]^T$ iid sample $N(\mu, \sigma^2)$ where $\theta = [\mu, \sigma^2]^T$ is unknown ... $(\mu \in \mathbb{R}, \sigma^2 > 0)$

$$f(y; \mu, \sigma^2) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \right\}$$

$$\Rightarrow \log f(y; \mu, \sigma^2) = \text{const} - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2$$

Take $\frac{\partial}{\partial \mu}$ and $\frac{\partial}{\partial \sigma}$ and set equal to 0

$$\Rightarrow \left.\begin{aligned} &\frac{1}{\sigma^2} \sum (y_i - \mu) = 0 \\ &-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum (y_i - \mu)^2 = 0 \end{aligned}\right\} \Rightarrow \begin{aligned} \hat{\mu} &= \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\mu})^2 \end{aligned}$$

Computation of ML

Define $L(\theta, y) = \log f_\theta(y)$. ∇_θ = gradient.

$$\nabla_\theta L(\theta, y) = \left[\frac{\partial}{\partial \theta_1} L(\theta, y), \frac{\partial}{\partial \theta_2} L(\theta, y), \ldots, \frac{\partial}{\partial \theta_m} L(\theta, y)\right]^T$$

and similarly ....

$$\nabla_\theta^2 L(\theta, y) = \left[\frac{\partial^2 L(\theta, y)}{\partial \theta_i \partial \theta_j}\right]_{i,j=1}^m$$

Want to solve ... $\nabla_\theta L(\theta, y) = 0$.

Consider Newton Raphson method for finding these zeros.

Taylor Series ... $\hat{\theta}^{(j)}$ denote an estimate computed @ jth iteration

$$\nabla_\theta L(\theta, y) = \nabla_\theta(\hat{\theta}^{(j)}, y) + \nabla_\theta^2 L(\hat{\theta}^{(j)}, y)\left(\theta - \hat{\theta}^{(j)}\right) + \text{H.O.T.s.}$$

Ignore HOTS set left hand side to zero and solve for $\theta = \hat{\theta}^{(j+1)}$

$$0 = \nabla_\theta L(\hat{\theta}^{(j)}, y) + \nabla_\theta^2 L(\hat{\theta}^{(j)}, y) \cdot \left(\hat{\theta}^{(j+1)} - \hat{\theta}^{(j)}\right)$$

$$\Rightarrow \hat{\theta}^{(j+1)}(y) = \hat{\theta}^{(j)}(y) - \left[\nabla_\theta^2 L(\hat{\theta}^{(j)}, y)\right]^{-1} \nabla_\theta L(\hat{\theta}^{(j)}, y)$$

This gives a seq. $\{\hat{\theta}^{(j)}\}$ which:

- may or may not converge.
- even if it does converge ... the limit could corresp. to

  local min
  local max
  inflection pt.

<u>Fact</u> $L(\theta, y)$ is concave $\Longrightarrow$ N.R. converges to a global max.

[If $L(\theta, y)$ is twice differentiable then
concave $\Longleftrightarrow$ $\nabla^2_\theta L(\theta, y) \leq 0 \quad \forall \theta \in \Delta$]

Possible Simplification

In N.R. must invert $m \times m$ matrix @ every iteration:

- If $\hat{\theta}^{(0)}$ is a good estimate then might have

$$\nabla_\theta^2 L(\hat{\theta}^{(j)}, y) \approx \nabla_\theta^2 L(\hat{\theta}^{(0)}, y).$$

so might just use one ...

- $$\nabla_\theta^2 L(\hat{\theta}^{(0)}, y) \cong E_{\hat{\theta}^{(0)}}\left\{\nabla_\theta^2 L(\hat{\theta}^{(0)}, Y)\right\}$$

$$= -I_{\hat{\theta}^{(0)}}$$ info. matrix

$$\Rightarrow \hat{\theta}^{(j+1)} = \hat{\theta}^{(j)} + I_{\hat{\theta}^{(0)}}^{-1}\left(\nabla_\theta L(\hat{\theta}^{(j)}, y)\right).$$

ML Properties

* ML estimates may be biased.

$$\hat{\mu} = \frac{1}{n} \sum y_i \qquad \hat{\sigma}^2 = \frac{1}{n} \sum (y_i - \hat{\mu})^2$$

$$E_\theta\{\hat{\mu}\} = \mu \quad \forall \theta \qquad E_\theta\{\hat{\sigma}^2\} = \frac{n-1}{n}\sigma^2$$

ie biased.

* In "regular" sit. where an efficient est. exists it will be an ML estimator