

Lecture 7: Likelihood

- Data X ; model density $f_{\theta}(x)$ (could be discrete).
- *Likelihood function* is

$$L(\theta) = f_{\theta}(X).$$

- It is a statistic.
- The *log-likelihood function* is

$$\ell(\theta) = \log L(\theta).$$

- It is a statistic.
- The *Score function* is the gradient of ℓ :

$$U(\theta) = \nabla \ell(\theta) = \frac{\partial \ell}{\partial \theta}.$$

- It is a statistic.
- The statistics in question are all *functions* not just numbers.



What we do with likelihood

- Estimate θ by finding the value $\hat{\theta} = \hat{\theta}(X)$ w which maximizes L or equivalently ℓ . The MLE is $\hat{\theta}$.
- Get confidence intervals or sets for θ by taking

$$\{\theta : L(\theta)/L(\hat{\theta}) > c\}$$

for some c calibrated to give the desired coverage probability.

- Test hypotheses like $H_o : \theta = \theta_o$ by rejecting if

$$L(\theta)/L(\hat{\theta}) < c$$

for some c calibrated to give the desired level of the test.

- We study these methods by analyzing the behaviour of L , ℓ and U near the true value of θ ; do large sample theory.
- For large n , L is normal, ℓ is quadratic, U is linear *near* true parameter value.



Course coverage

- Chapter 9.3-9.7.

