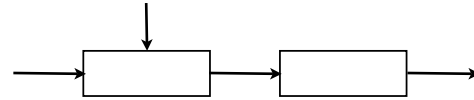


## LECTURE 23

- **Readings:** Section 9.1  
(not responsible for  $t$ -based confidence intervals, in pp. 471-473)
- **Outline**
  - Classical statistics
  - Maximum likelihood (ML) estimation
  - Estimating a sample mean
  - Confidence intervals (CIs)
  - CIs using an estimated variance

## Classical statistics



- also for vectors  $x$  and  $v$ :  
 $p_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$
- **These are NOT conditional probabilities;  $\theta$  is NOT random**
  - mathematically: many models, one for each possible value of  $\theta$
- **Problem types:**
  - Hypothesis testing:  
 $H_0 : \theta = 1/2$  versus  $H_1 : \theta = 3/4$
  - Composite hypotheses:  
 $H_0 : \theta = 1/2$  versus  $H_1 : \theta \neq 1/2$
  - Estimation: design an **estimator**  $\hat{\Theta}$ , to keep estimation **error**  $\hat{\Theta} - \theta$  small

### Maximum Likelihood Estimation

- Model, with unknown parameter(s):  
 $X \sim p_X(x; \theta)$
- Pick  $\theta$  that “makes data most likely”  
 $\hat{\theta}_{ML} = \arg \max_{\theta} p_X(x; \theta)$
- Compare to Bayesian MAP estimation:  
 $\hat{\theta}_{MAP} = \arg \max_{\theta} p_{\Theta|X}(\theta | x)$   
 $\hat{\theta}_{MAP} = \arg \max_{\theta} \frac{p_{X|\Theta}(x|\theta)p_{\Theta}(\theta)}{p_X(x)}$
- **Example:**  $X_1, \dots, X_n$ : i.i.d., exponential( $\theta$ )

$$\max_{\theta} \prod_{i=1}^n \theta e^{-\theta x_i}$$

$$\max_{\theta} \left( n \log \theta - \theta \sum_{i=1}^n x_i \right)$$

$$\hat{\theta}_{ML} = \frac{n}{x_1 + \dots + x_n} \quad \hat{\Theta}_n = \frac{n}{X_1 + \dots + X_n}$$

### Desirable properties of estimators (should hold FOR ALL $\theta$ !!!)

- **Unbiased:**  $E[\hat{\Theta}_n] = \theta$ 
  - exponential example, with  $n = 1$ :  
 $E[1/X_1] = \infty \neq \theta$   
(biased)
- **Consistent:**  $\hat{\Theta}_n \rightarrow \theta$  (in probability)
  - exponential example:  
 $(X_1 + \dots + X_n)/n \rightarrow E[X] = 1/\theta$
  - can use this to show that:  
 $\hat{\Theta}_n = n/(X_1 + \dots + X_n) \rightarrow 1/E[X] = \theta$
- **“Small” mean squared error (MSE)**  
 $E[(\hat{\Theta} - \theta)^2] = \text{var}(\hat{\Theta} - \theta) + (E[\hat{\Theta} - \theta])^2$   
 $= \text{var}(\hat{\Theta}) + (\text{bias})^2$

### Estimate a mean

- $X_1, \dots, X_n$ : i.i.d., mean  $\theta$ , variance  $\sigma^2$   
 $X_i = \theta + W_i$   
 $W_i$ : i.i.d., mean, 0, variance  $\sigma^2$   
 $\hat{\Theta}_n = \text{sample mean} = M_n = \frac{X_1 + \dots + X_n}{n}$

#### Properties:

- $E[\hat{\Theta}_n] = \theta$  (unbiased)
- WLLN:  $\hat{\Theta}_n \rightarrow \theta$  (consistency)
- MSE:  $\sigma^2/n$
  
- Sample mean often turns out to also be the ML estimate.  
E.g., if  $X_i \sim N(\theta, \sigma^2)$ , i.i.d.

### Confidence intervals (CIs)

- An estimate  $\hat{\Theta}_n$  may not be informative enough
- An  $1 - \alpha$  **confidence interval** is a (random) interval  $[\hat{\Theta}_n^-, \hat{\Theta}_n^+]$ ,  
s.t.  $P(\hat{\Theta}_n^- \leq \theta \leq \hat{\Theta}_n^+) \geq 1 - \alpha, \quad \forall \theta$ 
  - often  $\alpha = 0.05$ , or 0.25, or 0.01
  - interpretation is subtle
- CI in estimation of the mean  
 $\hat{\Theta}_n = (X_1 + \dots + X_n)/n$ 
  - normal tables:  $\Phi(1.96) = 1 - 0.05/2$

$$P\left(\frac{|\hat{\Theta}_n - \theta|}{\sigma/\sqrt{n}} \leq 1.96\right) \approx 0.95 \quad (\text{CLT})$$

$$P\left(\hat{\Theta}_n - \frac{1.96\sigma}{\sqrt{n}} \leq \theta \leq \hat{\Theta}_n + \frac{1.96\sigma}{\sqrt{n}}\right) \approx 0.95$$

More generally: let  $z$  be s.t.  $\Phi(z) = 1 - \alpha/2$

$$P\left(\hat{\Theta}_n - \frac{z\sigma}{\sqrt{n}} \leq \theta \leq \hat{\Theta}_n + \frac{z\sigma}{\sqrt{n}}\right) \approx 1 - \alpha$$

### The case of unknown $\sigma$

- Option 1: use upper bound on  $\sigma$ 
  - if  $X_i$  Bernoulli:  $\sigma \leq 1/2$
- Option 2: use ad hoc estimate of  $\sigma$ 
  - if  $X_i$  Bernoulli( $\theta$ ):  $\hat{\sigma} = \sqrt{\hat{\Theta}(1 - \hat{\Theta})}$
- Option 3: Use generic estimate of the variance
  - Start from  $\sigma^2 = E[(X_i - \theta)^2]$

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \theta)^2 \rightarrow \sigma^2$$

(but do not know  $\theta$ )

$$\hat{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\Theta}_n)^2 \rightarrow \sigma^2$$

(unbiased:  $E[\hat{S}_n^2] = \sigma^2$ )

MIT OpenCourseWare  
<http://ocw.mit.edu>

6.041 / 6.431 Probabilistic Systems Analysis and Applied Probability  
Fall 2010

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.