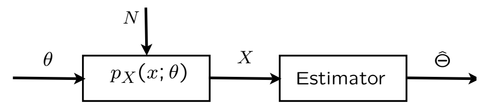
LECTURE 23

- **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 θ:
  $p_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta_1,\ldots,\theta_m)$
- These are NOT conditional probabilities; θ is NOT random
  - mathematically: many models, one for each possible value of θ
- **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 θ that "makes data most likely"

$$\hat{\theta}_{\text{ML}} = \arg\max_{\theta} p_X(x;\theta)$$

- Compare to Bayesian MAP estimation:

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} p_{\Theta|X}(\theta \mid x)$$

$$\hat{\theta}_{\text{MAP}} = \arg\max_{\theta} \frac{p_{X|\Theta}(x|\theta)p_{\Theta}(\theta)}{p_X(x)}$$

- **Example:** $X_1,\ldots,X_n$: i.i.d., exponential(θ)

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

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

$$\hat{\theta}_{\text{ML}} = \frac{n}{x_1 + \cdots + x_n} \qquad \hat{\Theta}_n = \frac{n}{X_1 + \cdots + X_n}$$

Desirable properties of estimators (should hold FOR ALL θ !!!)

- **Unbiased:** $\mathbf{E}[\hat{\Theta}_n] = \theta$
  - exponential example, with $n = 1$:
    $\mathbf{E}[1/X_1] = \infty \neq \theta$
    (biased)
- **Consistent:** $\hat{\Theta}_n \to \theta$ (in probability)
  - exponential example:
    $(X_1 + \cdots + X_n)/n \to \mathbf{E}[X] = 1/\theta$
  - can use this to show that:
    $\hat{\Theta}_n = n/(X_1 + \cdots + X_n) \to 1/\mathbf{E}[X] = \theta$
- **"Small" mean squared error (MSE)**

$$\begin{aligned}\mathbf{E}[(\hat{\Theta} - \theta)^2] &= \text{var}(\hat{\Theta} - \theta) + (\mathbf{E}[\hat{\Theta} - \theta])^2 \\ &= \text{var}(\hat{\Theta}) + (\text{bias})^2\end{aligned}$$

Estimate a mean

- $X_1, \ldots, X_n$: i.i.d., mean θ, variance σ^2

$X_i = \theta + W_i$

W_i: i.i.d., mean, 0, variance σ^2

$$\hat{\Theta}_n = \text{sample mean} = M_n = \frac{X_1 + \cdots + X_n}{n}$$

Properties:

- $\mathbf{E}[\hat{\Theta}_n] = \theta$ (unbiased)
- WLLN: $\hat{\Theta}_n \to \theta$ (consistency)
- MSE: σ^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. $\mathbf{P}(\hat{\Theta}_n^- \le \theta \le \hat{\Theta}_n^+) \ge 1-\alpha, \quad \forall\ \theta$

  - often $\alpha = 0.05$, or 0.25, or 0.01
  - interpretation is subtle

- CI in estimation of the mean $\hat{\Theta}_n = (X_1 + \cdots + X_n)/n$
  - normal tables: $\Phi(1.96) = 1 - 0.05/2$

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

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

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

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

The case of unknown σ

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

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

(but do not know θ)

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

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