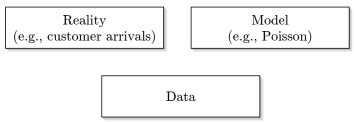
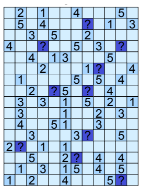
LECTURE 21

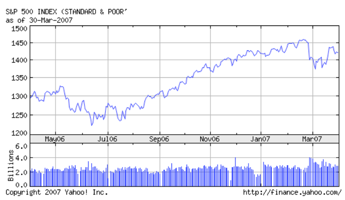
- **Readings:** Sections 8.1-8.2

"It is the mark of truly educated people to be deeply moved by **statistics**."
(Oscar Wilde)



- Design & interpretation of experiments
  – polling, medical/pharmaceutical trials...
- Netflix competition
- Finance



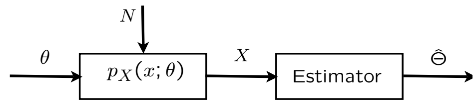


- Signal processing
  – Tracking, detection, speaker identification,...

Types of Inference models/approaches

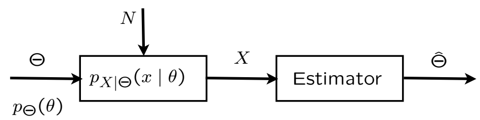
- Model building versus inferring unknown variables. E.g., assume $X = aS + W$
  – Model building: know "signal" S, observe X, infer a
  – Estimation in the presence of noise: know a, observe X, estimate S.
- **Hypothesis testing:** unknown takes one of few possible values; aim at small probability of incorrect decision
- **Estimation:** aim at a small estimation error

- **Classical statistics:**



θ: unknown parameter (not a r.v.)
  - E.g., θ = mass of electron

- **Bayesian:** Use priors & Bayes rule



Bayesian inference: Use Bayes rule

- **Hypothesis testing**
  – discrete data

$$p_{\Theta|X}(\theta \mid x) = \frac{p_\Theta(\theta)\, p_{X|\Theta}(x \mid \theta)}{p_X(x)}$$

  – continuous data

$$p_{\Theta|X}(\theta \mid x) = \frac{p_\Theta(\theta)\, f_{X|\Theta}(x \mid \theta)}{f_X(x)}$$

- **Estimation;** continuous data

$$f_{\Theta|X}(\theta \mid x) = \frac{f_\Theta(\theta)\, f_{X|\Theta}(x \mid \theta)}{f_X(x)}$$

$$\begin{aligned} Z_t &= \Theta_0 + t\Theta_1 + t^2\Theta_2 \\ X_t &= Z_t + W_t, \qquad t = 1, 2, \ldots, n \end{aligned}$$

Bayes rule gives:

$$f_{\Theta_0,\Theta_1,\Theta_2|X_1,\ldots,X_n}(\theta_0, \theta_1, \theta_2 \mid x_1, \ldots, x_n)$$

Estimation with discrete data

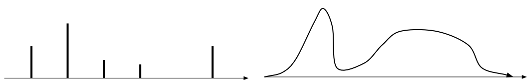
$$f_{\Theta|X}(\theta \mid x) = \frac{f_\Theta(\theta)\, p_{X|\Theta}(x \mid \theta)}{p_X(x)}$$

$$p_X(x) = \int f_\Theta(\theta) p_{X|\Theta}(x \mid \theta)\, d\theta$$

- **Example:**
  – Coin with unknown parameter θ
  – Observe X heads in n tosses
- What is the Bayesian approach?
  – Want to find $f_{\Theta|X}(\theta \mid x)$
  – Assume a prior on Θ (e.g., uniform)

Output of Bayesian Inference

- Posterior distribution:
  - pmf $p_{\Theta|X}(\cdot \mid x)$ or pdf $f_{\Theta|X}(\cdot \mid x)$



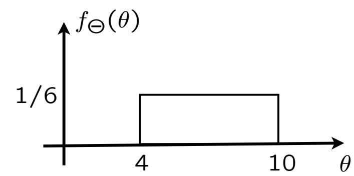
- If interested in a single answer:
  - Maximum a posteriori probability (MAP):
    - $p_{\Theta|X}(\theta^* \mid x) = \max_\theta p_{\Theta|X}(\theta \mid x)$
      minimizes probability of error;
      often used in hypothesis testing
    - $f_{\Theta|X}(\theta^* \mid x) = \max_\theta f_{\Theta|X}(\theta \mid x)$
  - Conditional expectation:

$$\mathbf{E}[\Theta \mid X = y] = \int \theta f_{\Theta|X}(\theta \mid x)\, d\theta$$

  - Single answers can be misleading!

Least Mean Squares Estimation

- Estimation in the absence of information



- find estimate c, to:

$$\text{minimize } \mathbf{E}\left[(\Theta - c)^2\right]$$

- Optimal estimate: $c = \mathbf{E}[\Theta]$
- Optimal mean squared error:

$$\mathbf{E}\left[(\Theta - \mathbf{E}[\Theta])^2\right] = \mathrm{Var}(\Theta)$$

LMS Estimation of Θ based on X

- Two r.v.'s Θ, X
- we observe that $X = x$
  - new universe: condition on $X = x$
- $\mathbf{E}\left[(\Theta - c)^2 \mid X = x\right]$ is minimized by
  $c =$
- $\mathbf{E}\left[(\Theta - \mathbf{E}[\Theta \mid X = x])^2 \mid X = x\right]$
  $\leq \mathbf{E}[(\Theta - g(x))^2 \mid X = x]$

◦ $\mathbf{E}\left[(\Theta - \mathbf{E}[\Theta \mid X])^2 \mid X\right] \leq \mathbf{E}\left[(\Theta - g(X))^2 \mid X\right]$

◦ $\mathbf{E}\left[(\Theta - \mathbf{E}[\Theta \mid X])^2\right] \leq \mathbf{E}\left[(\Theta - g(X))^2\right]$

$\mathbf{E}[\Theta \mid X]$ minimizes $\mathbf{E}\left[(\Theta - g(X))^2\right]$
over all estimators $g(\cdot)$

LMS Estimation w. several measurements

- Unknown r.v. Θ
- Observe values of r.v.'s $X_1, \ldots, X_n$
- Best estimator: $\mathbf{E}[\Theta \mid X_1, \ldots, X_n]$
- Can be hard to compute/implement
  - involves multi-dimensional integrals, etc.