LECTURE 19

November 17, 2008

- **Readings:**
  Revisit Section 4.3 (pp. 222-226)
  Sections 8.1, 8.2

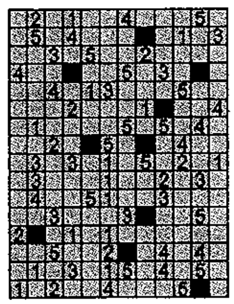
Lecture outline

- Statistics and Inference
  - Representative applications
  - Classical versus Bayesian statistics
- Bayes rule
  - continuous/discrete X
  - continuous/discrete Y
- Reporting a distribution, versus a single number

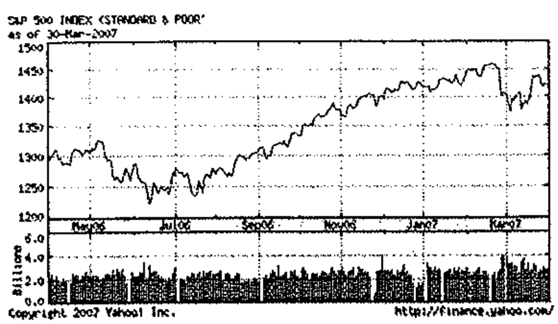
"It is the mark of truly educated people to be deeply moved by statistics."
(Oscar Wilde)

Sample Applications

- Polling
  - Design of experiments/sampling methodologies
  - Lancet study on Iraq death toll
- Medical/pharmaceutical trials
- Data mining

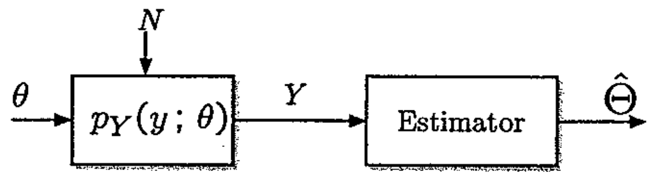


  - Netflix competition
- Finance



- Signal processing
  - Tracking
  - Echo cancellation

Classical statistics:



θ: unknown parameter (not a r.v.)

- $\theta = \pi$, in Buffon needle problem
- $\theta =$ mass of electron

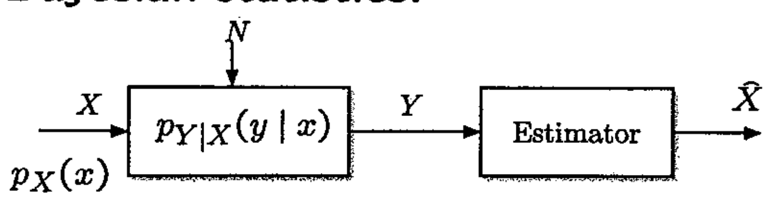
Example: $Y_i = \theta + W_i, \qquad W_i$: i.i.d.

$$\mathbf{E}[W_i] = 0, \qquad \text{var}(W_i) = \sigma^2$$

Natural estimator: $\hat{\Theta} = \frac{1}{n}(Y_1 + \cdots + Y_n)$

Properties: $\mathbf{E}[\hat{\Theta}; \theta] = \theta, \quad \text{var}(\hat{\Theta}; \theta) = \frac{\sigma^2}{n}$

- **Bayesian statistics:**



X discrete (hypothesis testing) Y discrete

$$p_{X|Y}(x \mid y) = \frac{p_X(x) p_{Y|X}(y \mid x)}{p_Y(y)}$$

$$p_Y(y) = \sum_x p_X(x) p_{Y|X}(y \mid x)$$

Example:

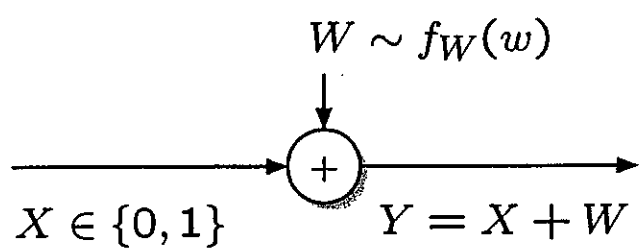
- $X = 1, 0$: airplane present/not present
- $Y = (Y_1, \ldots, Y_n)$
  $Y_i = 0, 1$: ith radar detected something

X discrete (hypothesis testing)
Y continuous

$$p_{X|Y}(x \mid y) = \frac{p_X(x) f_{Y|X}(y \mid x)}{f_Y(y)}$$

$$f_Y(y) = \sum_x p_X(x) f_{Y|X}(y \mid x)$$

Example:



$$f_{Y|X}(y \mid x) = f_W(y - x)$$

X continuous (estimation); Y continuous

$$f_{X|Y}(x \mid y) = \frac{f_X(x) f_{Y|X}(y \mid x)}{f_Y(y)}$$

$$f_Y(y) = \int_x f_X(x) f_{Y|X}(y \mid x)\, dx$$

Example:

$$\begin{aligned} Z_t &= X_0 + tX_1 + t^2 X_2 \\ Y_t &= Z_t + W_t, \qquad t = 1, 2, \ldots, n \end{aligned}$$

$$f_{Y_t|X_0,X_1,X_2}(y \mid x_0, x_1, x_2) = f_{W_t}(y - x_0 - tx_1 - t^2 x_2)$$

Bayes rule gives:

$$f_{X_0,X_1,X_2|Y_1,\ldots,Y_n}(x_0, x_1, x_2 \mid y_1, \ldots, y_n)$$

Use this to get:

$$f_{Z_n|Y_1,\ldots,Y_n}(z \mid y_1, \ldots, y_n)$$

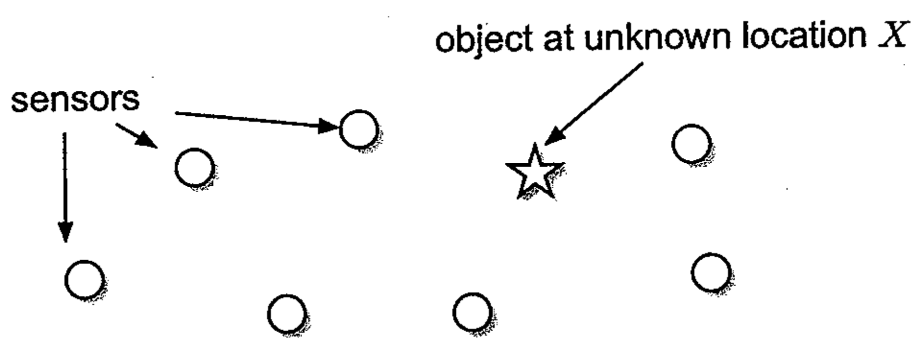
(derived distribution problem)

X continuous (estimation); Y discrete

$$f_{X|Y}(x \mid y) = \frac{f_X(x)p_{Y|X}(y \mid x)}{p_Y(y)}$$

$$p_Y(y) = \int_x f_X(x)p_{Y|X}(y \mid x)\,dx$$

Example:

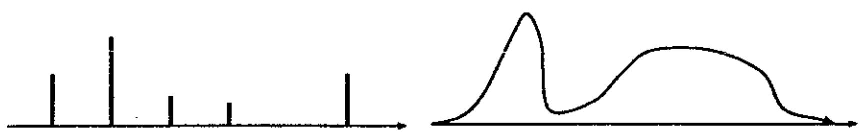


$\mathbf{P}$(sensor i "senses" the object | X)

$= f$(distance of X from sensor)

Output of Bayesian Inference

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

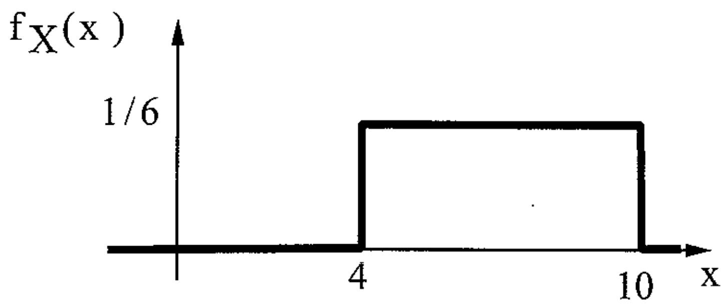


- If interested in a single answer:
  - Maximum a posteriori probability:
    - $p_{X|Y}(x^* \mid y) = \max_x p_{X|Y}(x \mid y)$ (minimizes probability of error)
    - $f_{X|Y}(x^* \mid y) = \max_x f_{X|Y}(x \mid y)$
  - Conditional expectation:

$$\mathbf{E}[X \mid Y = y] = \int_x x f_{X|Y}(x \mid y)\, dx$$

  - Single answers can be misleading!

Prediction in the absence of information



- prediction c

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

- $c = \mathbf{E}[X]$

- Optimal mean squared error:

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

Predicting X based on Y

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

  $c =$

- $\mathbf{E}\left[(X - \mathbf{E}[X \mid Y=y])^2 \mid Y=y\right]$

$$\leq \mathbf{E}[(X-g(y))^2 \mid Y=y]$$

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

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

$\mathbf{E}[X \mid Y]$ minimizes

$$\mathbf{E}\left[(X-g(Y))^2\right]$$

over all predictors $g(\cdot)$