

# EE 278: Probability and Statistical Inference:

- 1) A second course in probability
- 2) Introduction to Statistical Inference

## Probability

A) Concentration Inequalities / High-dimensional probability

EE178: Law of Large Numbers

Ex:  $X_1, X_2, \dots, X_n$  iid Unif  $[0, 1]$

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} \mathbb{E}[X_i] = \frac{1}{2}$$

Ex: Every student in this class attends <sup>exactly!</sup> 50% of the lectures in person.

Ex: four lectures  $\{1, 2, 3, 4\}$ :  $\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}$

What is the probability that a given student attends the  $k$ 'th lecture? Assume there are  $m$  lectures

$$p = \frac{\binom{m-1}{m-1 - \frac{m}{2} - 1}}{\binom{m}{\frac{m}{2}}} = \frac{(m-1)!}{\left(\frac{m}{2}\right)! \left(\frac{m}{2}-1\right)!} = \frac{1}{m} \frac{m/2}{1} = \frac{1}{2}$$

Alternatively, define

$X_{ik} = \begin{cases} 1 & \text{if the } i\text{'th student attends the } k\text{'th lecture} \\ 0 & \text{otherwise} \end{cases}$

↑ student      ↓ lecture

$\mathbb{E}[X_{ik}] = p$ : the probability that the  $i$ 'th student attends the  $k$ 'th lecture

$$X_{i1} + X_{i2} + X_{i3} + \dots + X_{im} = \frac{m}{2}$$

$$\mathbb{E}\{X_{i1} + X_{i2} + X_{i3} + \dots + X_{im}\} = \frac{m}{2}$$

$$\underbrace{\mathbb{E}\{X_{i1}\}}_p + \underbrace{\mathbb{E}\{X_{i2}\}}_p + \mathbb{E}\{X_{i3}\} + \dots + \mathbb{E}\{X_{im}\} = \frac{m}{2}$$

$$\Rightarrow mp = \frac{m}{2} \Rightarrow p = \frac{1}{2}$$

$\sum_{i=1}^{60} X_{i9}$  : # of number of students that show up for the 9'th lecture

LLN  $\frac{1}{60} \sum_{i=1}^{60} X_{i9} \rightarrow \mathbb{E}\{X_{i9}\} = p = \frac{1}{2}$

$$\sum_{i=1}^{60} X_{i9} \approx 30$$

For a class size of  $n=60$ , how reasonable it is to expect the # of students that attend a particular lecture to be close to 30, e.g.  $[25, 35]$ .

B) Random vectors :  $\bar{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$  → a random variable

- Gaussian random vectors
- Covariance matrix (principal component analysis & dimensionality reduction)

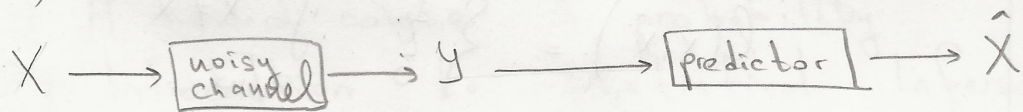
c) Random process :  $X[k] \quad k=1, \dots, \infty$

$$X(t) \quad t \in \mathbb{R}$$

- Basic concepts : Stationarity, Covariance function

- Gaussian processes, White noise

## 2) Inference Problems



$X$ : State of a physical system  
Communicated message

Label of an image, e.g. cat or dog  
State of the patient

- A)  $X$  is discrete
- $X \in \{0, 1\}$  a bit
  - $X \in \{\text{cat}, \text{dog}\}$
  - $X \in \{\text{healthy}, \text{infected}\}$
  - $X \in \{\text{fire}, \text{no fire}\}$

Hypothesis Testing / Detection / Decoding / Classification  
Decision-making

Performance Criteria: Probability of error  $P(\hat{X} \neq X)$

Find the best predictor that minimizes the probability of error.

- B)  $X$  is a continuous random variable

$X$ : location of a robot arm

$X$ : the value of a stock

Estimation problems

$$P(\hat{X} \neq X) = 1$$

$l(x, \hat{x})$ : cost / loss / error function

$$l(x, \hat{x}) : \mathbb{R}^2 \rightarrow \mathbb{R}^+$$

Find a predictor s.t.  $\mathbb{E}[l(x, \hat{x})]$  is minimized.

Probability of error is a special case

$$l(x, \hat{x}) = \begin{cases} 1 & \text{if } \hat{x} \neq x \\ 0 & \text{o/w} \end{cases}$$

Popular loss function:  $l(\hat{x}, x) = (x - \hat{x})^2$   
squared loss

Minimum mean-squared estimation

- I)  $X$  - random variable       $Y$  - random variable
- II)  $X$  - random vector       $Y$  - random vector
- III)  $X$  - random process       $Y$  - random process

Model-based Approach

Data-based Approach (Machine Learning Approach)

1) data  $\rightarrow$  model  $\rightarrow$  predictor  
physics  $\nearrow$

data  $\rightarrow$  predictor

Eg.  $Y = \alpha X + W$   
noise  $\nwarrow$

Eg.  $\{(X_i, Y_i)\}_{i=1}^n \rightarrow \{f_i, i=1, \dots, n\}$   
labels  $\nwarrow$  images  $\nearrow$

2) linear models  $\rightarrow$  optimal linear predictors

non-linear predictors  
e.g. neural networks

3) quantify performance analytically

cross-validation