Introduction to theory of probability

Jayanta Mukhopadhyay
Dept. of Computer Science and Engg.

Courtesy: Prof. Pabitra Mitra, CSE, IIT Kharagpur

Outline

- Important concepts in probability theory
- Bayes' rule
- Random variables and distributions

Definition of Probability

- ***Experiment***: toss a coin twice
- ***Sample space***: possible outcomes of an experiment
  - S = {HH, HT, TH, TT}
- ***Event***: a subset of possible outcomes
  - A={HH}, B={HT, TH}
- ***Probability of an event*** :
  - a number assigned to an event Pr(A)
    - Axiom 1: Pr(A) $\geq$ 0
    - Axiom 2: Pr(S) = 1
    - Axiom 3: For every sequence of disjoint events

$$\Pr\left(\bigcup_i A_i\right) = \sum_i \Pr(A_i)$$

  - Example: Pr(A) = n(A)/N: frequentist statistics

Conditional Probability

- For events A and B, **conditional probability** Pr(A|B) stands for the probability that A happens if B occurs.
  - Prob. of A given B.
- Example: A={HH}, B={HT, TH}, what is the conditional probability Pr(A|B)?
  - 0
- Example: A={HH}, B={HH, TH}, what is the conditional probability Pr(B|A)?
  - 1

Joint Probability

- For events A and B, **joint probability** Pr(AB) stands for the probability that both events happen.
- Example: A={HH}, B={HT, TH}, what is the joint probability Pr(A,B)?
  - 0
- Pr(AB)=Pr(A)Pr(B|A)
  - =Pr(B)Pr(A|B)
- A={HH}, B={HH, TH}, the joint probability Pr(A,B)?
  - =Pr(A)Pr(B|A)
  - =1/4 x 1=1/4

Independence

- Two events ***A and B are independent*** in case

$$\Pr(A|B)=\Pr(A)$$

OR

$$\Pr(B|A)=\Pr(B)$$

OR

$$\Pr(AB) = \Pr(A)\Pr(B)$$

- A set of events $\{A_i\}$ is independent in case

$$\Pr\left(\bigcap_i A_i\right) = \prod_i \Pr(A_i)$$

Independence

- Two events ***A and B are independent*** in case

$$\Pr(AB) = \Pr(A)\Pr(B)$$

- A set of events $\{A_i\}$ is independent in case

$$\Pr\left(\bigcap_i A_i\right) = \prod_i \Pr(A_i)$$

- Example: Drug test

	Woman	Man
Success	200	1800
Failure	1800	200

A = {A patient is a Woman}

B = {Drug fails}

Will event A be independent from event B ?

Independence

- Consider the experiment of tossing a coin twice
- Example I:

  HT or HH

  - A = {HT, HH}, B = {HT}
  - Will event A independent from event B?

  $P(B|A)=1$
  $P(A|B)=0$
  $P(A)=0.5$
  $P(B)=0.25$

- Example II:
  - A = {HT}, B = {TH}
  - Will event A independent from event B?

  $P(B|A)=0$
  $P(A|B)=0$
  $P(A)=0.25$
  $P(B)=0.25$

- Disjoint $\neq$ Independence
- If A is independent from B, B is independent from C, will A be independent from C?

Independence mutual, but not transitive!

$(P(AB)=P(A)P(B))$ & $(P(BC)=P(B)P(C)) \neq> (P(AC)=P(A)P(C))$

Conditioning

- If A and B are events with Pr(A) > 0, the ***conditional probability of B given A*** is

$$\Pr(B \mid A) = \frac{\Pr(AB)}{\Pr(A)}$$

Conditioning

- If A and B are events with Pr(A) > 0, the ***conditional probability of B given A*** is

$$\Pr(B \mid A) = \frac{\Pr(AB)}{\Pr(A)}$$

- Example: Drug test

	Women	Men
Success	200	1800
Failure	1800	200

A = {Patient is a Women}

B = {Drug fails}

Pr(B|A) = ?

Pr(A|B) = ?

Conditioning

- If A and B are events with Pr(A) > 0, the ***conditional probability of B given A*** is

$$\Pr(B \mid A) = \frac{\Pr(AB)}{\Pr(A)}$$

- Example: Drug test

	Women	Men
Success	200	1800
Failure	1800	200

A = {Patient is a Women}

B = {Drug fails}

Pr(B|A) = ?

Pr(A|B) = ?

- Given A is independent from B, what is the relationship between Pr(A|B) and Pr(A)?

Bayes' Rule

- Given two events A and B and suppose that Pr(A) > 0. Then

$$\Pr(B \mid A) = \frac{\Pr(AB)}{\Pr(A)} = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)}$$

- Example:

Pr(R) = 0.8

Pr(W|R)	R	¬R
W	0.7	0.4
¬W	0.3	0.6

R: It is a rainy day

W: The grass is wet

Pr(R|W) = ?

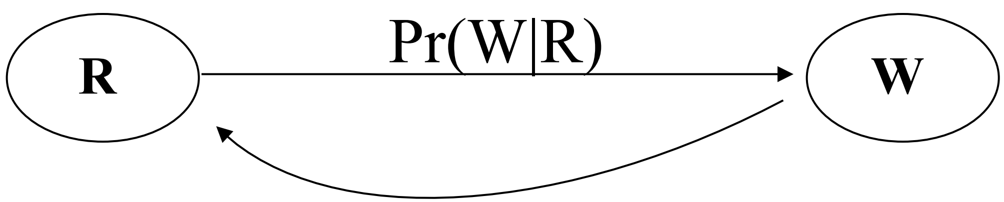
Bayes' Rule

	R	¬R
W	0.7	0.4
¬W	0.3	0.6

R: It rains

W: The grass is wet

Information



Inference

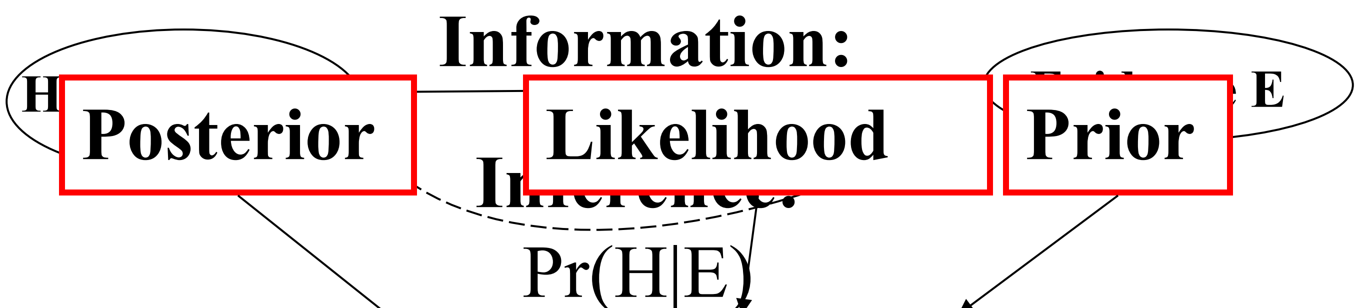
Pr(R|W)

Bayes' Rule

	R	¬R
W	0.7	0.4
¬W	0.3	0.6

R: It rains

W: The grass is wet



$$\boxed{\Pr(H \mid E)} = \frac{\boxed{\Pr(E \mid H)}\boxed{\Pr(H)}}{\Pr(E)}$$

Bayes' Rule: Expanded

- Suppose that B_1, B_2, … B_k form a partition of S:

$$B_i \bigcap B_j = \varnothing; \quad \bigcup_i B_i = S$$

Suppose that $\Pr(B_i) > 0$ and $\Pr(A) > 0$. Then

$$\Pr(B_i \mid A) = \frac{\Pr(A \mid B_i)\Pr(B_i)}{\Pr(A)}$$

Bayes' Rule: Expanded

- Suppose that B_1, B_2, … B_k form a partition of S:

$$B_i \bigcap B_j = \varnothing; \quad \bigcup_i B_i = S$$

Suppose that $\Pr(B_i) > 0$ and $\Pr(A) > 0$. Then

$$\Pr(B_i \mid A) = \frac{\Pr(A \mid B_i)\Pr(B_i)}{\Pr(A)}$$

$$= \frac{\Pr(A \mid B_i)\Pr(B_i)}{\sum_{j=1}^{k} \Pr(AB_j)}$$

Bayes' Rule: Expanded

- Suppose that B_1, B_2, … B_k form a partition of S:

$$B_i \bigcap B_j = \varnothing; \quad \bigcup_i B_i = S$$

Suppose that $\Pr(B_i) > 0$ and $\Pr(A) > 0$. Then

$$\Pr(B_i \mid A) = \frac{\Pr(A \mid B_i)\Pr(B_i)}{\Pr(A)}$$

$$= \frac{\Pr(A \mid B_i)\Pr(B_i)}{\sum_{j=1}^{k} \Pr(AB_j)}$$

$$= \frac{\Pr(A \mid B_i)\Pr(B_i)}{\sum_{j=1}^{k} \Pr(B_j)\Pr(A \mid B_j)}$$

Random Variable and Distribution

- A ***random variable X*** is a numerical outcome of a random experiment
- The ***distribution*** of a random variable is the collection of possible outcomes along with their probabilities:
  - Discrete case: $\Pr(X = x) = p_\theta(x)$
  - Continuous case: $\Pr(a \le X \le b) = \int_a^b p_\theta(x)dx$
  - Probability density function
  - Probability mass function

Random Variable: Example

- Let S be the set of all sequences of three rolls of a die. Let X be the sum of the number of dots on the three rolls.
- What are the possible values for X?
- $\Pr(X = 5) = ?$, $\Pr(X = 10) = ?$

Expectation

- A random variable X~Pr(X=x). Then, its expectation is

$$E[X] = \sum_x x \Pr(X = x)$$

  - In an empirical sample, $x_1, x_2, \ldots, x_N$,

$$E[X] = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- Continuous case: $E[X] = \int_{-\infty}^{\infty} x p_\theta(x) dx$

- Expectation of sum of random variables

$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

Expectation: Example

- Let S be the set of all sequence of three rolls of a die. Let X be the sum of the number of dots on the three rolls.
- What is E(X)?

- Let S be the set of all sequence of three rolls of a die. Let X be the product of the number of dots on the three rolls.
- What is E(X)?

Variance

- The variance of a random variable X is the expectation of $(X\text{-}E[X])^2$:

$$
\begin{aligned}
Var(X) &= E((X - E[X])^2) \\
&= E(X^2 + E[X]^2 - 2XE[X]) \\
&= E(X^2 - E[X]^2) \\
&= E[X^2] - E[X]^2
\end{aligned}
$$

Bernoulli Distribution

- The outcome of an experiment can either be success (i.e., 1) and failure (i.e., 0).
- $Pr(X=1) = p$, $Pr(X=0) = 1-p$
- $E[X] = p$, $Var(X) = p(1-p)$

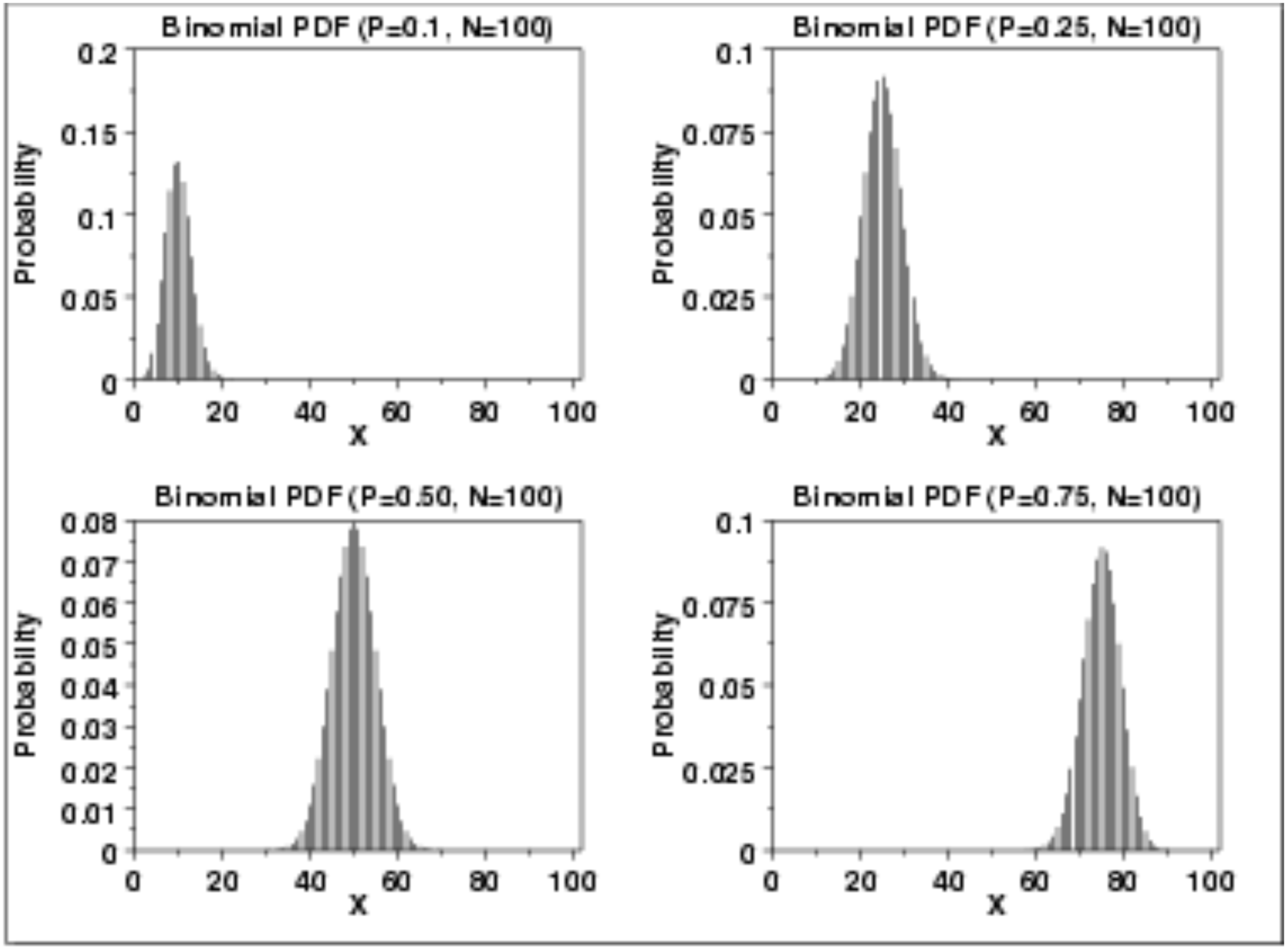
Binomial Distribution

- n draws of a Bernoulli distribution
  - X_i~Bernoulli(p), $X=\sum_{i=1}^{n} X_i$, X~Bin(p, n)
- Random variable X stands for the number of times that experiments are successful.

$$\Pr(X=x)=p_\theta(x)=\begin{cases}\binom{n}{x}p^x(1-p)^{n-x} & x=1,2,\ldots,n\\ 0 & \text{otherwise}\end{cases}$$

- $E[X] = np$, $Var(X) = np(1-p)$

Plots of Binomial Distribution



Poisson Distribution

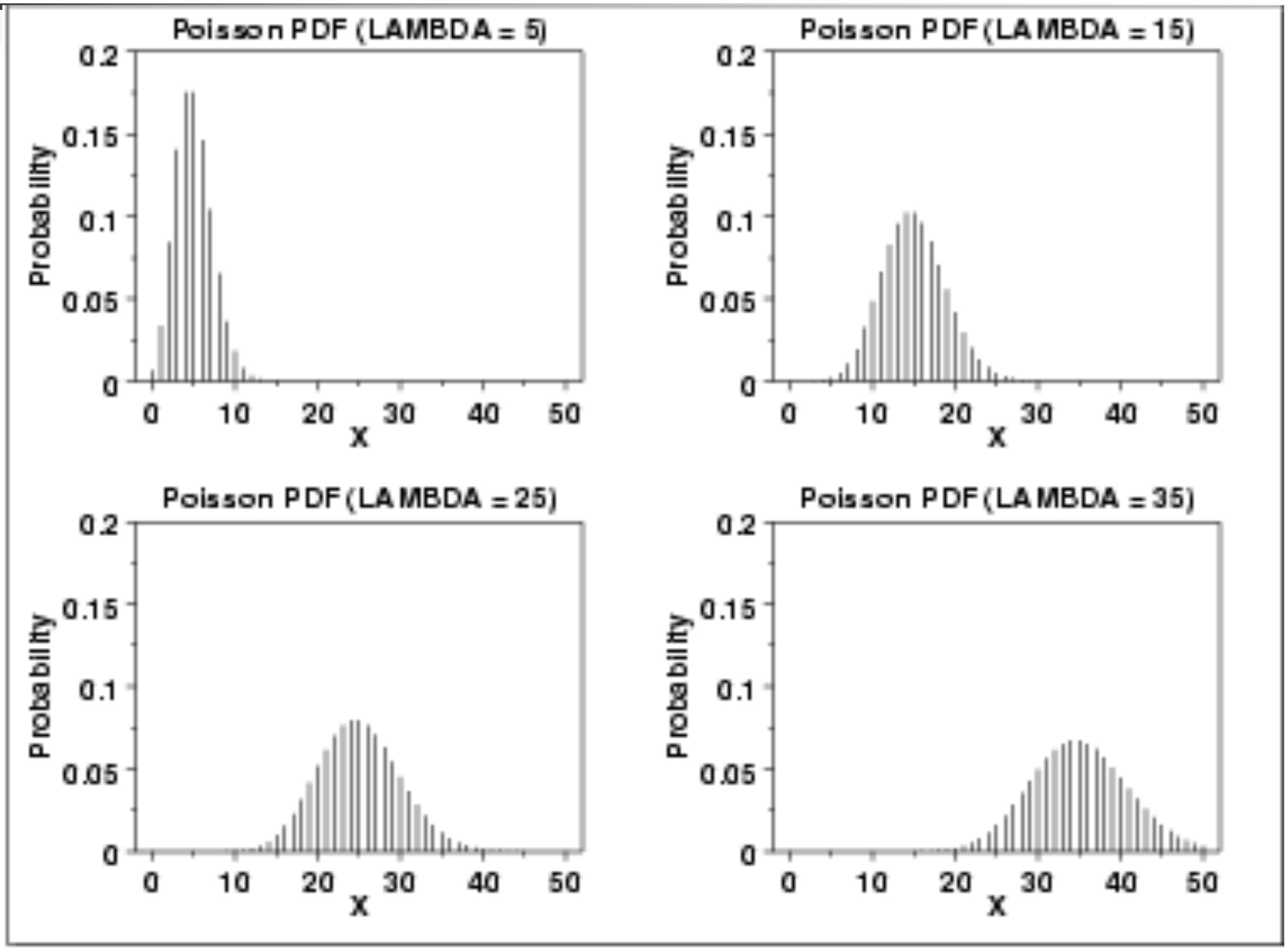
- Coming from Binomial distribution
  - Fix the expectation $\lambda = np$
  - Let the number of trials $n \to \infty$

  A Binomial distribution will become a Poisson distribution

$$\Pr(X = x) = p_\theta(x) = \begin{cases} \dfrac{\lambda^x}{x!} e^{-\lambda} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- E[X] = λ, Var(X) = λ

Plots of Poisson Distribution



Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma^2)$

$$p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

$$\Pr(a \le X \le b) = \int_a^b p_\theta(x)dx = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} dx$$

- E[X]= μ, Var(X)= σ^2
- If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, $X = X_1 + X_2$?

Not sum of Gaussian distributions.

Sum of normally distributed random variables.
If independent, also Gaussian. $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Summary

- Basic concepts in probability theory
- Bayes' rule
- Random variable and probability distributions

Thank you!