

# Lecture 2 : *Probability Theory and Random Variables*

---

Hee Yong Youn

School of Information and Communication Engineering

Sungkyunkwan University

Suwon, Korea

youn@ece.skku.ac.kr

# Probability theory

## □ Probability theory as a model

- Functional aspect (not scale)  
because deals with the process of the object
- Abstract representation (not concrete)  
because averages large number of non-deterministic outcomes
- Analytical techniques (neither physical nor simulation)  
because uses set theory

## □ A series of observations can characterize the relative frequency of the possible outcomes (ex) Program execution time

# Experiment

## ❑ Experiment

### ❑ Discrete/ continuous outcomes

- ✓ Discrete outcomes: rolling a dice ( 6 different outcomes)
- ✓ Continuous outcomes: uncountably infinite no. of outcomes even with the range

### ❑ Element : instance of an object of interest

(ex) Object: color

Element; red, yellow, ...

# Set theory notation

## □ Set theory notation

- Sample space (or universe) ( $\Omega$ ): The universal set containing all possible outcomes considered
- $\{a,b\}$ : a set of distinct elements,  $a$  and  $b$
- $[a,b]$  (or  $(a,b)$ ): a set of infinite, uncountable values between and including (or excluding)  $a$  and  $b$
- Empty set ( $\phi$ ): a set of no element
- Union ( $\cup$ )
- Intersection ( $\cap$ )
- Complement ( $'$ )
- Membership ( $\in$ )
- Subset ( $\subset$ )

# Set relationships

## □ Set relationships

- Mutually exclusive:  $A \cap B = \emptyset$
- Mutually exhaustive:  $A \cup B = \underline{\Omega}$
- Partition: mutually exclusive and exhaustive
- Interpretation using Venn diagram

# Law of set theory

## □ Law of set theory

- Commutative (for same operators):  $A \cap B = B \cap A$
- Associative (for same operators):  $A \cup (B \cup C) = (A \cup B) \cup C$
- Distributive (for different operators):  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- **Identities:**  $A \cap \Omega = A$ ,  $A \cup \underline{\phi} = A$
- **Inverse:**  $(A')' = A$   
 $A \cup A' = \underline{\Omega}$  (inclusion),  $A \cap A' = \phi$  (exclusion)
- **DeMorgan's Law:**  $(A \cap B)' = A' \cup B'$   
 $(A \cup B)' = A' \cap B'$

# Sample space & event

## □ Sample space

- For an experiment
- Set of all possible outcomes
- (ex) tossing two coins: { HH,HT,TH,TT }

## □ Event

- A set of outcomes which is a subset of  $\Omega$
- (ex) The faces are not same in tossing two coins: { HT,TH }

# Power set & probability measure

- Power set of Set-A: a set of all possible subsets of A
- Probability measure ( $P$ ): the fraction of a large number of repetitions ( relative frequency ) that a prescribed event or outcome may occur

# Law of probability

## □ Law of probability

- $P[\Omega] = \underline{1}$
- $0 \leq P[A] \leq \underline{1}$  for  $A \subseteq \Omega$
- $P[A \cup B] = P[A] + P[B] - P[\underline{A \cap B}]$  for  $A, B \subseteq \Omega$
- $P[\bigcup_{m=1}^{\infty} A_m] = \sum_{m=1}^{\infty} P[A_m]$  if  $A_i$ 's are mutually disjoint

# Conditional probability

## □ Conditional Probability

$$\square P[A|B] = \frac{P[A \cap B]}{P[B]}$$

- (Em 2.21) Two coins are flipped. What is the prob. of having 2 heads if at least one is head?

*A: two heads;  $P[A] = 1/4$*

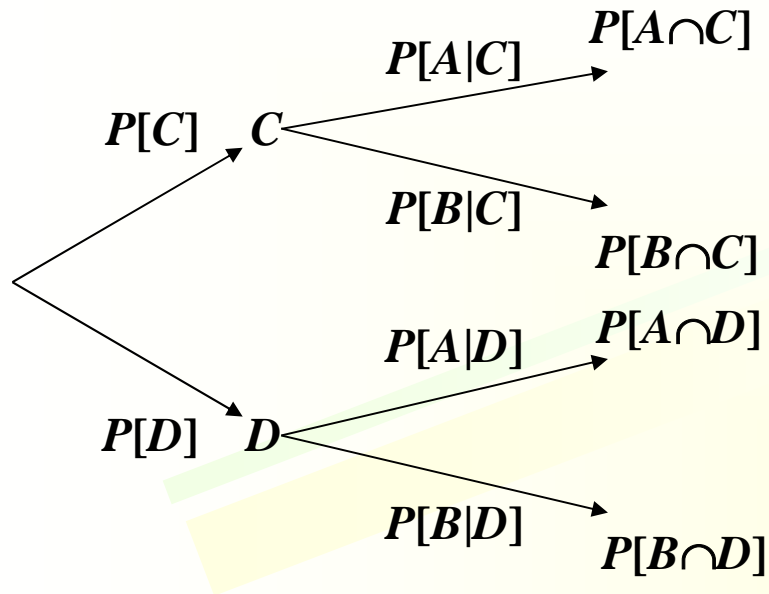
*B: at least one is head;  $P[B] = 3/4$*

$$P[A \cap B] = 1/4$$

$$P[A|B] = P[A \cap B]/P[B] = (1/4)/(3/4) = 1/3$$

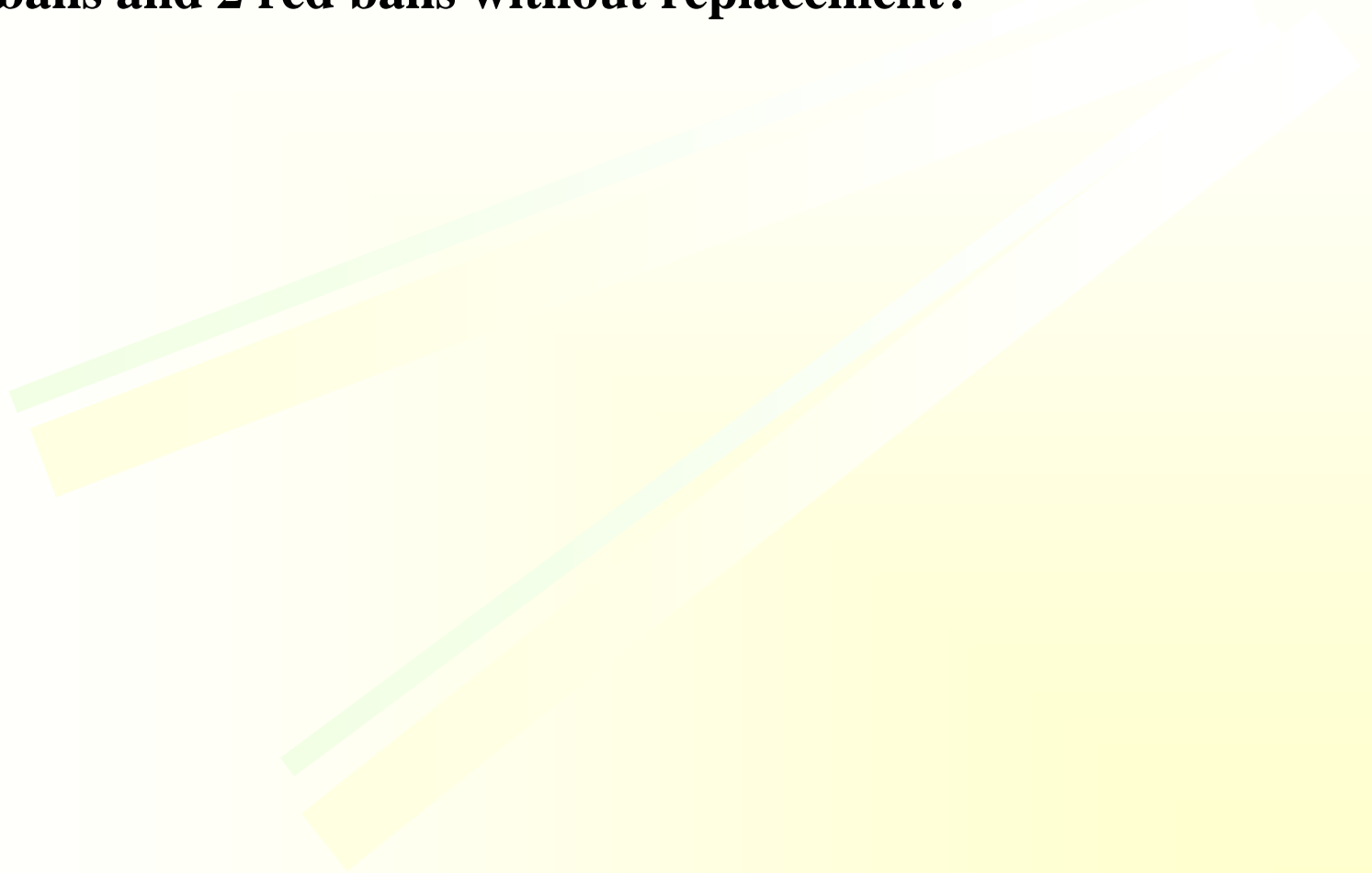
# Probability tree

## □ Probability tree



# Probability tree (exercise)

- (Ex 2.9) Prob. of drawing 2 white balls from a bucket containing 3 white balls and 2 red balls without replacement?



# Independence

□ Independence:  $A, B \subseteq \Omega$  are independent iff  
 $P[A \cap B] = P[A] P[B]$

□ (Proof)

If  $A$  and  $B$  are independent,  $P[A|B] = P[A]$  -----(a)

By definition,  $P[A|B] = \frac{P(A \cap B)}{P(B)}$  -----(b)

(a) = (b) results in

$$P[A] = \frac{P(A \cap B)}{P(B)}$$

Finally,  $P[A \cap B] = P[A] P[B]$

# Independence (example)

- (Em 2.23) When we toss two coins, what is the probability getting Head on the second coin given that Tail on the first coin?

$A$ : getting Head on the second coin;  $P[A] = \{TH, HH\} = 1/2$

$B$ : Tail on the first coin;  $P[B] = \{TT, TH\} = 1/2$

$P[A|B] = P[A \cap B] / P[B] = (1/4) / (1/2) = 1/2 = P[A]$

# Independence (exercise)

- (Ex 2.10) If  $A$  and  $B$  are independent, what is  $P[A \cup B]$ ? Here  $P[A] = 0.2$  and  $P[B] = 0.3$ .

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] = 0.5 - 0.2 \times 0.3 = 0.44$$

# Independence of a set of events

## □ Independence of a set of events

□ Mutually independent

□ Pairwise independent

□ (ex) Experiment: tossing 2 dices

Event  $A$ : 1<sup>st</sup> dice = 1, 2, or 3

Event  $B$ : 1<sup>st</sup> dice = 3, 4, or 5

Event  $C$  :  $\Sigma = 9$

□  $A = \{(1,*),(2,*),(3,*)\}, P[A] = \underline{1/2}$

$B = \{(3,*),(4,*),(5,*)\}, P[B] = \underline{1/2}$

$C = \{(3,6),(4,5),(5,4),(6,3)\}, P[C] = \underline{1/4}$

$A \cap B = \{ \underline{(3,*)} \}, P[A \cap B] = \underline{1/6} \neq P[A]P[B]$

$A \cap C = \{ \underline{(3,6)} \}, P[A \cap C] = \underline{1/36} \neq P[A]P[C]$

$B \cap C = \{ \underline{(3,6),(4,5),(5,4)} \}, P[B \cap C] = \underline{1/12} \neq P[B]P[C]$

$A \cap B \cap C = \{ \underline{(3,6)} \}, P[A \cap B \cap C] = \underline{1/36} = P[A]P[B]P[C]$

The events are not mutually independent since they are not pairwise independent

□ Does pairwise independency guarantee mutual independency?

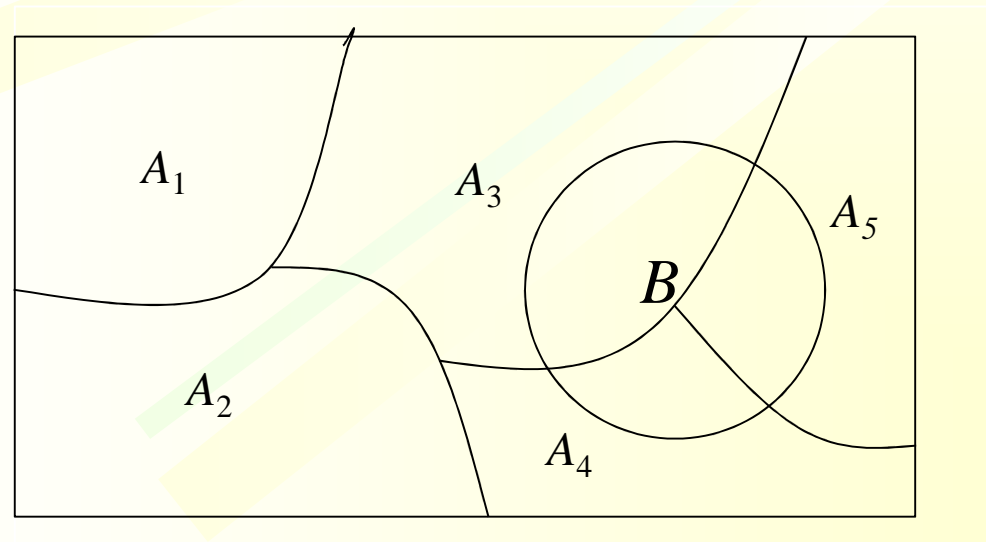
# Bayes' theorem

## □ Bayes' Theorem (Posteriori probability)

$$P[A_i | B] = \frac{P[A_i \cap B]}{P[B]} = \frac{P[A_i] P[B | A_i]}{\sum_j P[A_j \cap B]} = \frac{P[A_i] P[B | A_i]}{\sum_j P[A_j] P[B | A_j]}$$

### □ Conditions for applying the theorem

- i) Partition by  $A_i$ 's
- ii)  $P[ \textcolor{red}{B} ] \neq 0$



# Bayes' theorem (example)

- (Em 2.24) Three programmers submit jobs to a system, and sometimes their jobs fail to be executed. Assume that a job failed to be executed. What is the prob. that Programmer-1 sent the job?

Event- $A_i$  : program was submitted by Programmer- $i$

Event- $B$ : program failed

$$P[A_1] = 0.2, P[A_2] = 0.3, P[A_3] = 0.5$$

$$P[B/A_1] = 0.1, P[B/A_2] = 0.7, P[B/A_3] = 0.1$$

$$\begin{aligned} P[A_1|B] &= (P[A_1] P[B|A_1]) / (P[A_1] P[B|A_1] + P[A_2] P[B|A_2] + P[A_3] P[B|A_3]) \\ &= (0.2 \times 0.1) / (0.2 \times 0.1 + 0.3 \times 0.7 + 0.5 \times 0.1) = 0.02 / (0.02 + 0.21 + 0.05) \\ &= 0.02 / 0.28 = 0.071 \end{aligned}$$

# Combinatorics

## □ Combinatorics:

### □ Sum (Product) rule:

The total number of outcomes is the sum (product) of the number of outcomes of each \_\_\_\_\_ if they are \_\_\_\_\_ (combined).

# Combinatorics (exercise)

- (Ex 2.11) What is the probability to pick up an ace card after two decks of cards are shuffled together?
- (Ex 2.12) How many different combinations of cards do we have by picking one card from each of two decks of cards?

# Sampling with replacement

## □ Sampling with replacement: $N^R$

□  $N$ : number of elements,  $R$ : length of sequence (no. of samplings)

(ex)  $N = 4$  for  $\{1, 2, 3, 4\}$ ,  $R = 2$

11	12	13	14
21	22	23	24
31	32	33	34
41	42	43	44

}  $4 \times 4 = 4^2$

# Sampling without replacement

□ Sampling without replacement:  $\frac{N!}{(N - R)!}$

□ (ex)

$$\begin{array}{ccc} 12 & 13 & 14 \\ 21 & & 23 \\ 31 & 32 & \\ 41 & 42 & 43 \end{array} \left. \begin{array}{c} 24 \\ 34 \end{array} \right\} 4 \times 3 = \frac{4!}{(4 - 2)!}$$

□  ${}_N P_R = N(N-1)\dots(N-(R-1))$

□ (Ex 2.13) How many different combinations of cards do we have when we draw five cards from a deck of cards?

# Permutations & Combinations

❑ **Permutations:**  $N!$  (Sampling without replacement for length  $N$ )

❑ (ex)  $4 \times 3 \times 2 \times 1$

❑ **Combinations:**  $\binom{N}{R} = \frac{N!}{R!(N-R)!}$

❑  ${}_NC_R$  : Binomial coefficient of  $R$ th term of  $(x+y)^N$

(ex)  $(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$

$$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$$

❑ **Size of power set:**  $\sum_{R=0}^N \binom{N}{R} = 2^N$

$$(x+y)^N = \binom{N}{0}x^N y^0 + \binom{N}{1}x^{N-1}y^1 + \binom{N}{2}x^{N-2}y^2 + \dots + \binom{N}{N}x^0 y^N$$

$$\sum_{R=0}^N \binom{N}{R} = \binom{N}{0} + \binom{N}{1} + \dots + \binom{N}{N} = (x+y)^N \Big|_{x=1 \text{ and } y=1} = 2^N$$

# Combinations

- (Ex 2.14) How many different poker hands do we have if we draw five cards from a deck of cards?

# Random variables

## □ Random Variables ( $X$ )

□ A \_\_\_\_\_ that assigns a real number to each possible \_\_\_\_\_ in the sample space

## □ (ex) $X$ : number of heads in tossing two coins

	<u>Outcome</u>	<u>Probability</u>	<u>Value of <math>X</math></u>	<u>Prob[<math>X</math>]</u>
H	H	$1/4$	2	Prob[ $X=2$ ]=__
	T	$1/4$	1	Prob[ $X=1$ ]=__
T	H	$1/4$	1	
	T	$1/4$	0	Prob[ $X=0$ ]=__

Notation  $[X = x] = \{s \in \Omega \mid X(s) = x\}$

□ (ex)  $[X = 1] = \{HT \mid X(HT) = 1\}$

□ Random variable carries info about events using \_\_\_\_\_ in order to simplify the manipulation of them

## Random variables (cont'd)

- (Em 2.30) In dart throwing random variable  $x$  is the distance from the left side,  $l$ , normalized by the width,  $w$ . What is the value of  $x$ ?
- (Ex 2.15) What is the value of random variable,  $x$ , which is the sum of the dots of two dices rolled?

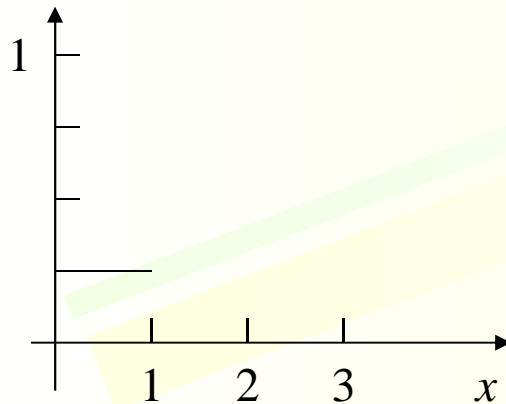
# Cumulative distribution function

## □ Cumulative distribution function (CDF), $F$

□  $F(x) = P[X \leq x]$

□ (ex) Coin tossing

$F(x)$



$$F(x) = \begin{cases} 0, & x < 0 \\ 1/4, & 0 \leq x < 1 \\ 3/4, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}$$
$$\begin{cases} F(-\infty) = 0 \\ F(\infty) = 1 \\ F(x_1) \leq F(x_2), \quad x_1 < x_2 \end{cases}$$

# Cumulative distribution function (exercise)

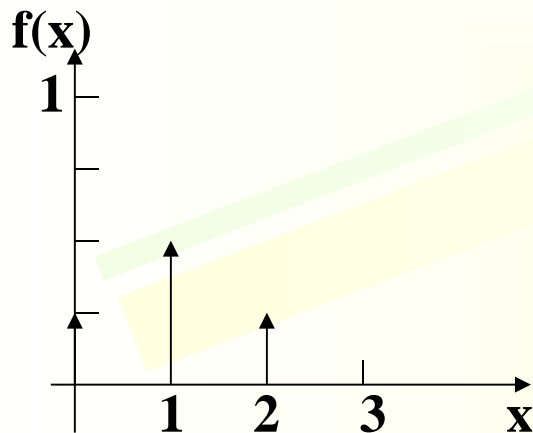
□ (Ex 2.16) Define the CDF of  $x$  of the dart problem.

# Probability density function

## □ Probability density function (PDF), $f$

□  $f(x) = \frac{d}{dx} F(x)$  or  $F(x) = \int_{-\infty}^x f(y) dy$

□ (ex) Coin tossing



- when a r.v. is discrete

$$f(x) = P[X = x] \\ = \int_x f(y) dy$$

$$f(x) = \begin{cases} \frac{1}{2}, & x = 0, 2 \\ \frac{1}{2}, & x = 1 \\ 0, & \text{elsewhere} \end{cases}$$

□ Since  $F(\infty) = 1$ ,  $\int_{-\infty}^{\infty} f(x) dx = 1$

□ Since  $F(x)$  is nondecreasing,  $f(x) \geq 0$

# Distribution of random variable

- ❑ Specified by the condition under which the r.v. is defined
- ❑ Geometric/ Binomial/ Exponential/ Poisson distribution
- ❑ Discrete/ continuous, finite/ infinite distribution

# Geometric distribution

- ❑ Experiment: a trial succeeds (1) with probability  $p$  or fails(0) with probability  $(1-p)$ . The trial continues until it succeeds.
- ❑  $\Omega: \{ 0^{i-1}1 \mid i = 1,2,3,\dots \}$
- ❑ r.v.  $K$ : no. of trials \_\_\_\_\_ the first success
- ❑  $P[K = k] = (1-p)^{k-1} p$  for  $k = 1,2,\dots$

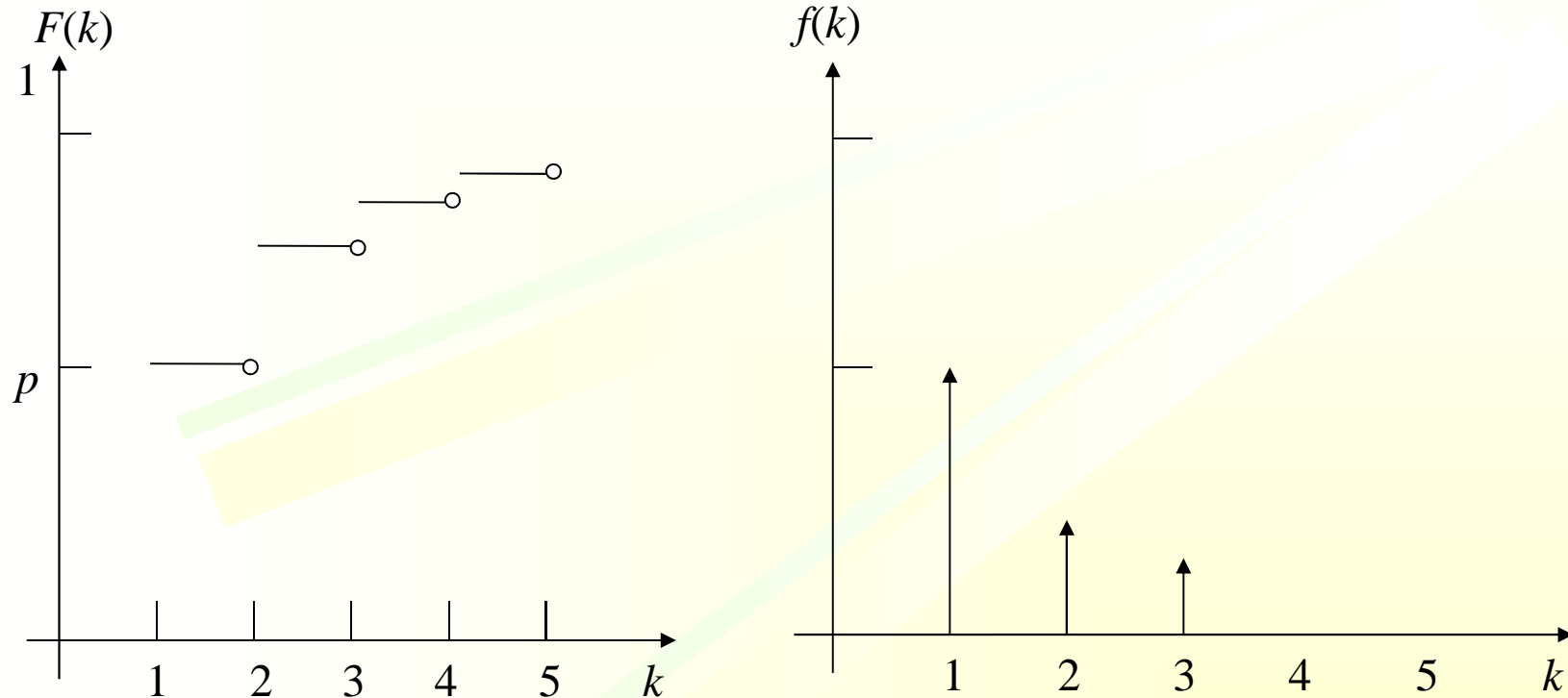
- ❑  $F(k) = P[K \leq k] = \sum_{i=1}^k (1-p)^{i-1} p = 1 - (1-p)^k$  for  $k \geq 1$

(Proof)

*Let  $q = 1 - p$*

$$\begin{aligned} \sum_{i=1}^k (1-p)^{i-1} p &= \sum_{i=1}^k q^{i-1} (1-q) = \sum_{i=1}^k (q^{i-1} - q^i) = \\ &= (q^0 - q^1) + (q^1 - q^2) + \dots + (q^{k-1} - q^k) = 1 - q^k \end{aligned}$$

# Geometric distribution



# Modified geometric distribution

□ r.v.: no. of trials \_\_\_\_\_ the first success

□  $P[K = k] = (1 - p)^k p$  for  $k = \text{____}, 2, 3, \dots$

□  $F(k) = P[K \leq k] = \sum_{i=0}^k (1-p)^i p = 1 - (1-p)^{k+1}$  for  $k \geq 0$

(Proof)

*Let  $q = 1 - p$*

$$\begin{aligned} \sum_{i=0}^k (1-p)^i p &= \sum_{i=0}^k q^i (1-q) = \sum_{i=0}^k (q^i - q^{i+1}) = \\ &= (q^0 - q^1) + (q^1 - q^2) + \dots + (q^k - q^{k+1}) = 1 - q^{k+1} \end{aligned}$$

## Binomial distribution ( $b(k;N,p)$ )

❑ **Experiment:** a trial succeeds (1) with prob.  $p$  or fails(0) with prob.  $(1-p)$ . The trial continues for  $N$  times.

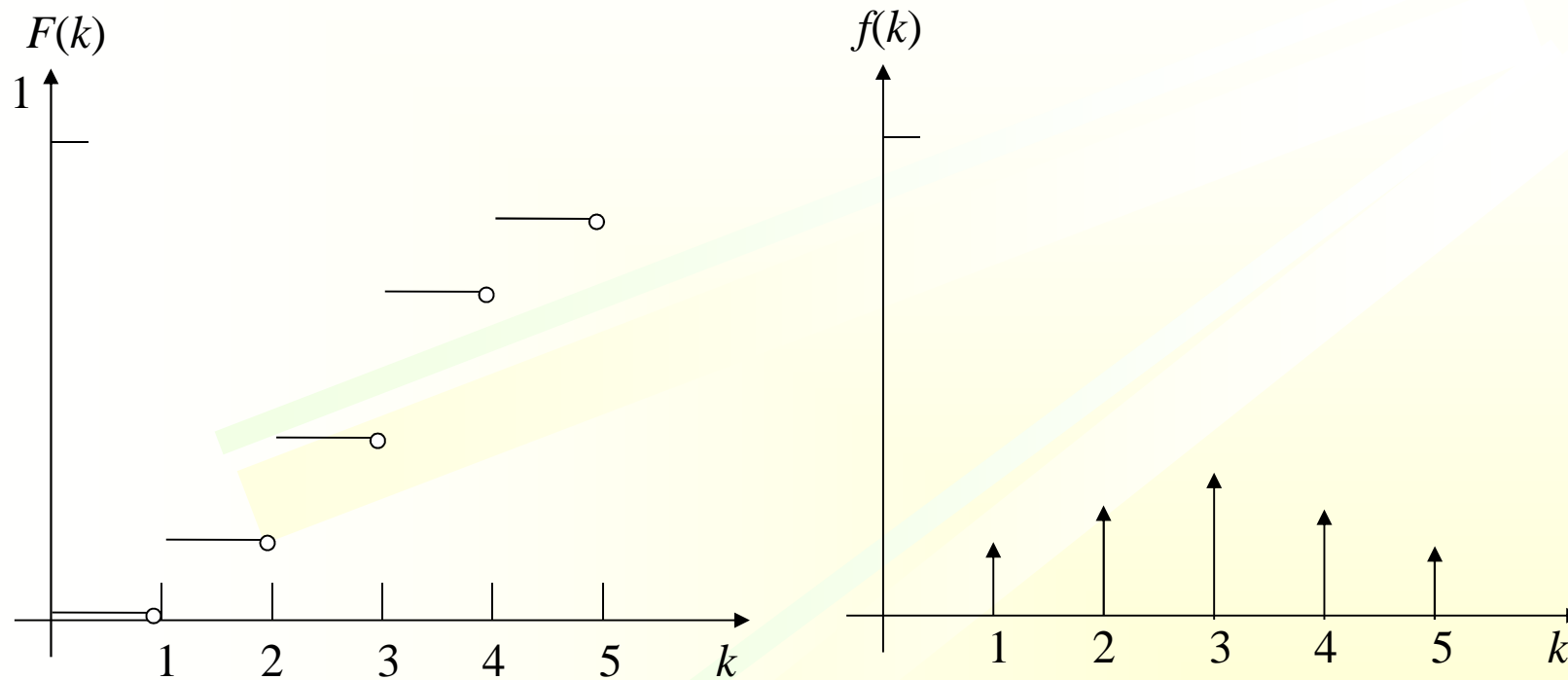
❑  $\Omega: \{0^i 1^{N-i} \mid i = 0, 1, \dots, N\}$

❑ **r.v.  $K$ :** no. of successes out of  $N$  trials

❑  $P[K = k] = \binom{N}{k} p^k (1-p)^{N-k}$  for  $0 \leq k \leq N$

❑  $F(k) = P[K \leq k] = \sum_{i=0}^k \binom{N}{i} p^i (1-p)^{N-i}$  (no closed form solution)

# Binomial distribution ( $b(k;N,p)$ )



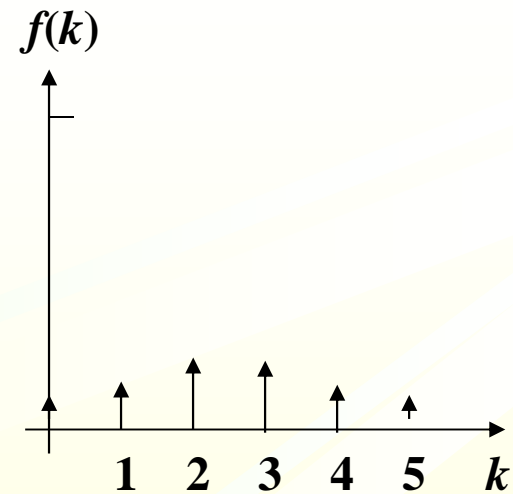
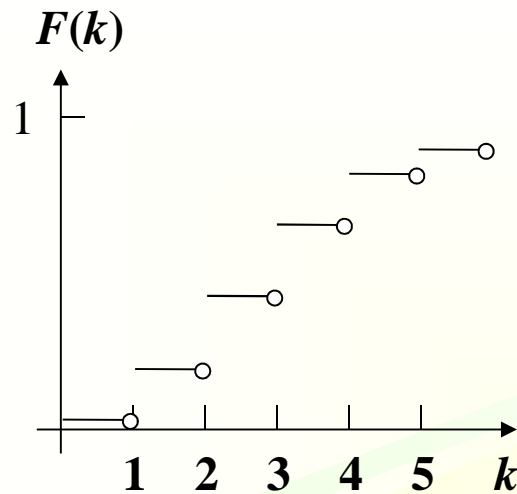
# Poisson distribution

- ❑ Experiment: success occurs at the rate of  $\lambda$
- ❑  $\Omega$ : {0,1,2,...successes }
- ❑ r.v.  $K$ : no. of successes in time  $T$

- ❑  $P[K = k \text{ in } T] = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$

- ❑  $F(k) = \sum_{i=0}^k \frac{(\lambda T)^i}{i!} e^{-\lambda T}$

# Poisson distribution (cont'd)



$$P = \lambda \Delta t = \frac{\lambda T}{n} : \text{prob. of a success in } \Delta t \ (n \gg \lambda T)$$

# Poisson distribution (cont'd)

$$\begin{aligned}
 P[K \text{ in } n] &= \binom{n}{k} \left( \frac{\lambda T}{n} \right)^k \left( 1 - \frac{\lambda T}{n} \right)^{n-k} \\
 &= \frac{n!}{k!(n-k)!} \frac{(\lambda T)^k}{n^k} \left( 1 - \frac{\lambda T}{n} \right)^{n-k} \\
 &= \frac{n(n-1)\dots(n-k+1)(n-k)!}{k!(n-k)!} \frac{(\lambda T)^k}{n^k} \left( 1 - \frac{\lambda T}{n} \right)^n \left( 1 - \frac{\lambda T}{n} \right)^{-k} \\
 &= \frac{n^k \left( 1 - \frac{1}{n} \right) \left( 1 - \frac{2}{n} \right) \dots \left( 1 - \frac{k-1}{n} \right)}{k! n^k} (\lambda T)^k \left( 1 - \frac{\lambda T}{n} \right)^{-k} \left( \left( 1 - \frac{\lambda T}{n} \right)^{\frac{-n}{\lambda T}} \right)^{-\lambda T}
 \end{aligned}$$

$$* \lim_{n \rightarrow \infty} \left( 1 - \frac{a}{n} \right)^{\frac{n}{a}} = e$$

$$P[k \text{ in } T] = \lim_{n \rightarrow \infty} P[k \text{ in } n] = \frac{1}{k!} (\lambda T)^k 1 \cdot e^{-\lambda T}$$

# Poisson distribution (cont'd)

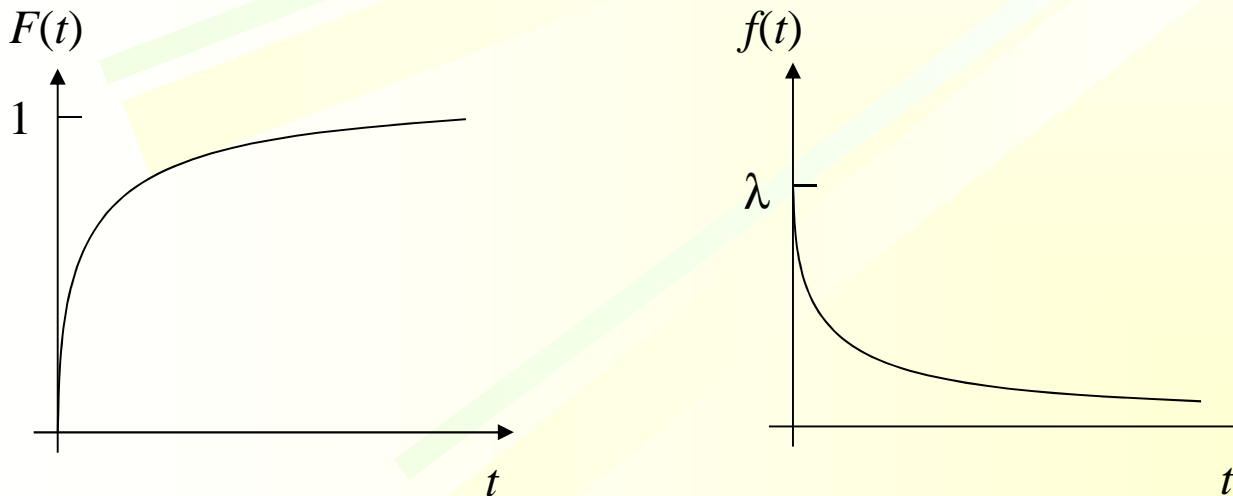
❑ Rule of thumb: Use Poisson for binomial if  $n \geq 20$  and  $p \leq 0.05$

❑ (ex)

$k$	$b(k; 5, 0.2)$	$b(k; 20, 0.05)$	Poisson( $k; \lambda T=1$ )
0	0.328	0.359	0.368
1	0.410	0.377	0.368
2	0.205	0.189	0.184
3	0.051	0.060	0.061

# Exponential distribution

- ❑ Continuous case of \_\_\_\_\_ distribution
- ❑ Experiment: success occurs at the rate of  $\lambda$
- ❑  $\Omega: \{t \mid t \geq 0\}$
- ❑ r.v.  $t$ : time to the first success
- ❑  $F(t) = P[T \leq t] = 1 - e^{-\lambda t}$  for  $0 \leq t$
- ❑  $f(t) = \lambda e^{-\lambda t}$



- ❑ Application: interarrival time, service time, time to failure, repair time

# Conditional PDF

$$\square f(x|A) = \frac{f(x)}{P[A]}, \quad x \in A$$

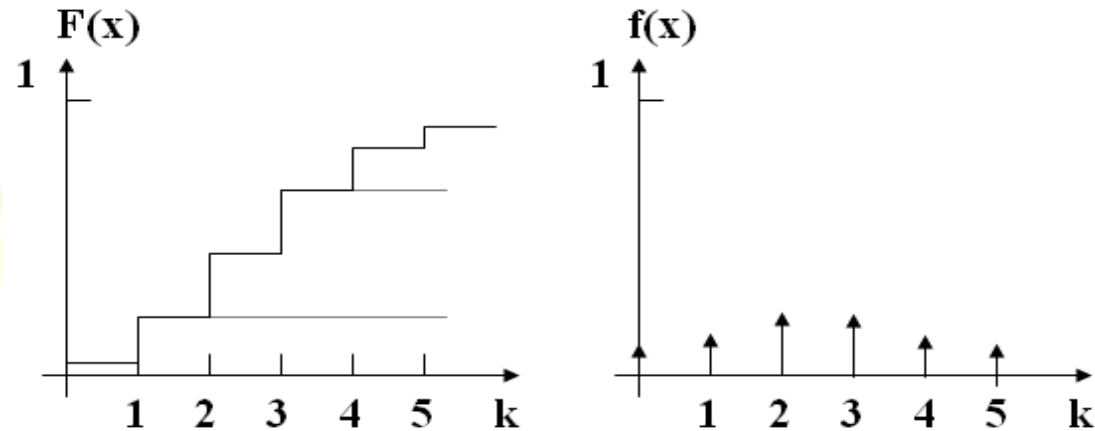
# Using CDF and PDF

## □ Using CDF and PDF

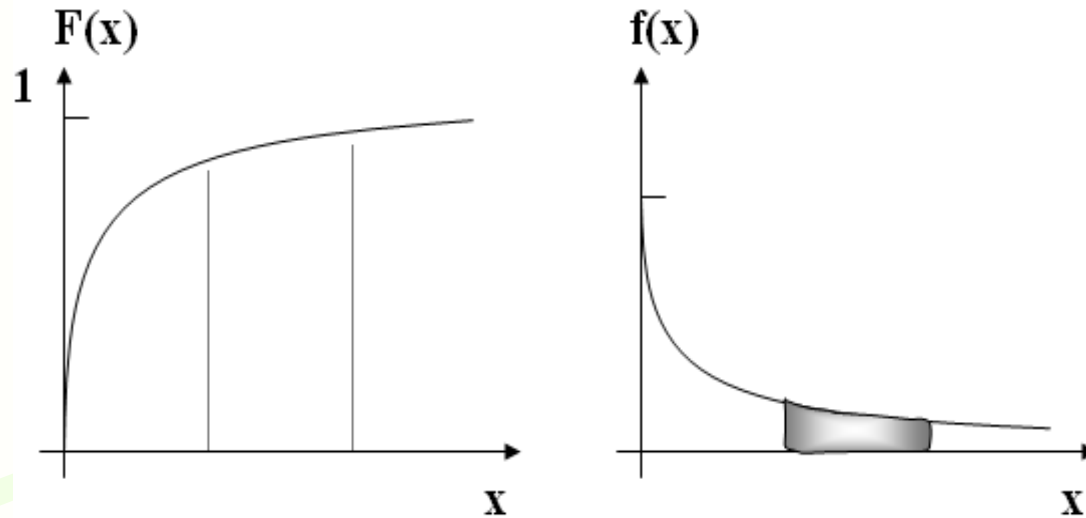
□ Calculate prob. of events and expectations

□ Use \_\_\_\_\_ for prob. and \_\_\_\_\_ for expectation

□  $P[a < X \leq b] = P[X \leq b] - P[X \leq a] = F(b) - F(a) = \int_a^b f(x) dx$  or  $\sum_{i=a+1}^b f(i)$



# Using CDF and PDF (cont'd)

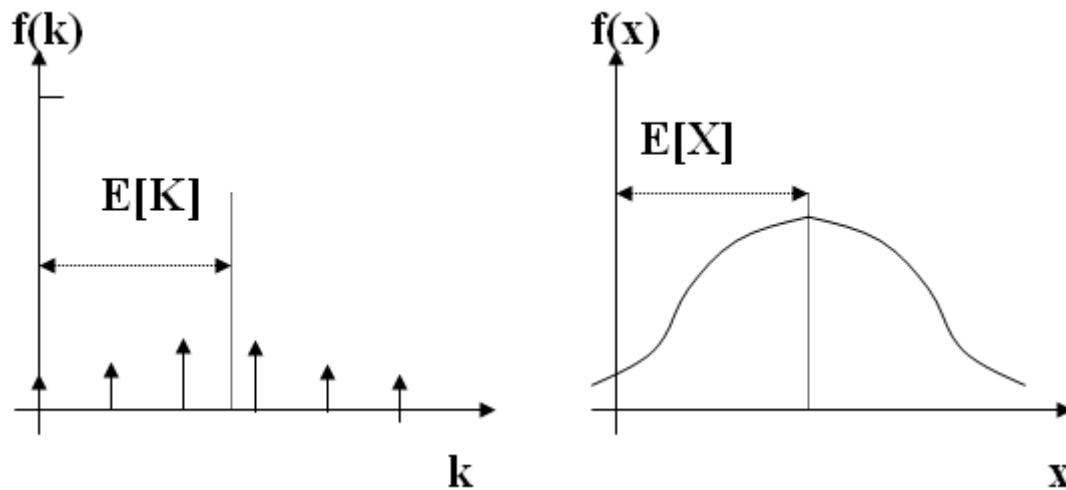


## Using CDF (exercises)

- ❑ (Ex 2.17) What is the prob that a dart lands in the middle third of the dart board?
- ❑ (Ex 2.18) What is the prob that a dart lands precisely in the middle of the dart board?
- ❑ (Ex 2.19) What is the prob that for a geometrically distributed random variable, the value is 4, 5, or 6?

# Expectations

$$\square E[K] = \sum_{-\infty}^{\infty} kf(k), \quad E[X] = \int_{-\infty}^{\infty} xf(x) dx$$



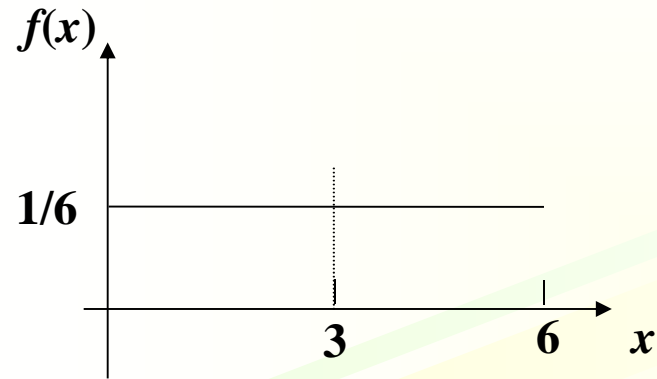
## $\square E[K]$

- $\square$  Expected value (average) of r.v.  $K$
- $\square$  Center of mass of the PDF
- $\square$  First moment of r.v.  $K$

## $\square E[K^2]$ : Second moment, $E[K^3]$ : Third moment

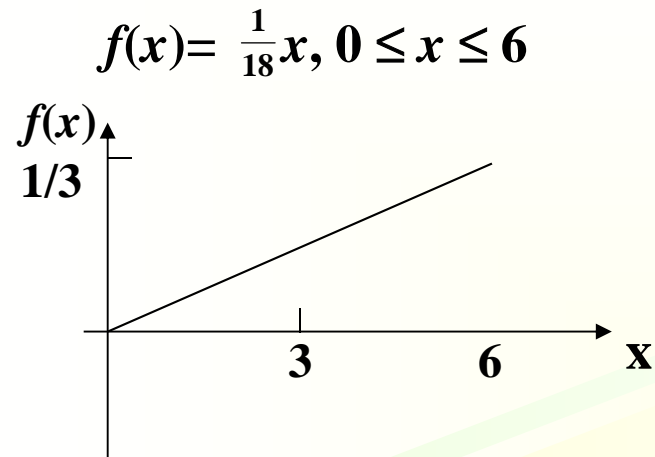
# Expectations (example)

□ (Ex)  $f(x) = 1/6, 0 \leq x \leq 6$



$$\begin{aligned} E[X] &= \int_0^6 x \cdot \frac{1}{6} dx \\ &= \frac{1}{6} \left. \frac{x^2}{2} \right|_0^6 \\ &= \frac{1}{6} \left( \frac{36}{2} - 0 \right) \\ &= 3 \end{aligned}$$

# Expectations (example)



$$\begin{aligned} E[X] &= \int_0^6 x \cdot \frac{1}{18} x \, dx \\ &= \frac{1}{18} \left. \frac{x^3}{3} \right|_0^6 \\ &= \frac{1}{18} (2 \times 6 \times 6) = 4 \end{aligned}$$

# Expectations (cont'd)

□  $E[K]$  for geometric distribution

$$E[K] = \sum_{k=0}^{\infty} k (1-p)^{k-1} p$$

$$= p \sum_{k=1}^{\infty} k (1-p)^{k-1} = -p \sum_{k=1}^{\infty} \frac{d}{dp} (1-p)^k$$

$$= -p \frac{d}{dp} \sum_{k=1}^{\infty} (1-p)^k \quad (* \sum_{i=0}^{\infty} x^i = \frac{1}{1-x}, x < 1 *)$$

$$= -p \frac{d}{dp} \left[ \frac{1}{1-p} - 1 \right] \quad (* \frac{d}{dx} \frac{f(x)}{g(x)} = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2} *)$$

$$= -p \frac{-1}{(1-p)^2} = \frac{p}{1-p}$$

# Expectations (cont'd)

## □ $E[T]$ for exponential distribution

$$\begin{aligned} E[T] &= \int_0^{\infty} t \lambda e^{-\lambda t} dt = -\lambda \int_0^{\infty} \left( \frac{d}{d\lambda} e^{-\lambda t} \right) dt = -\lambda \frac{d}{d\lambda} \int_0^{\infty} e^{-\lambda t} dt \\ &= -\lambda \frac{d}{d\lambda} \left( \frac{e^{-\lambda t}}{-\lambda} \Big|_0^{\infty} \right) = -\lambda \frac{d}{d\lambda} \left( \frac{1}{\lambda} \right) = -\lambda \frac{-1}{\lambda^2} = \frac{1}{\lambda} \end{aligned}$$

## Second moment ( $E[X^2]$ )

- $E[X]$ : first moment about the origin
- $X - E[X]$ : first moment about the mean

$$\int_{-\infty}^{\infty} (x - E[X]) f(x) dx = \int_{-\infty}^{\infty} x f(x) dx - E[X] = 0$$

- Second moment about the mean (variance)

$$\begin{aligned} & \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx - 2E[X] \int_{-\infty}^{\infty} x f(x) dx + (E[X])^2 \int_{-\infty}^{\infty} f(x) dx \\ &= E[X^2] - 2E[X]E[X] + (E[X])^2 \\ &= E[X^2] - (E[X])^2 \\ &= \sigma^2 \text{ (measure of the spread of the distribution)} \end{aligned}$$

- $\sigma = \sqrt{E[X^2] - (E[X])^2}$  (standard deviation)

## Second moment ( $E[K^2]$ ) (cont'd)

□  $E[K^2]$  for geometric distribution

$$\begin{aligned} E[K^2] &= \sum_{k=0}^{\infty} k^2 (1-p)^{k-1} p = p \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} \\ &= p \sum_{k=2}^{\infty} k^2 (1-p)^{k-1} + p \\ &= p \sum_{k=2}^{\infty} (k^2 - k) (1-p)^{k-1} + p + p \sum_{k=2}^{\infty} k (1-p)^{k-1} \\ &= p (1-p) \sum_{k=2}^{\infty} (k^2 - k) (1-p)^{k-2} + p + p \left( \frac{1}{p^2} - 1 \right) \end{aligned}$$

$$= p(1-p) \sum_{k=2}^{\infty} \frac{d^2}{d^2 p} (1-p)^k + \frac{1}{p}$$

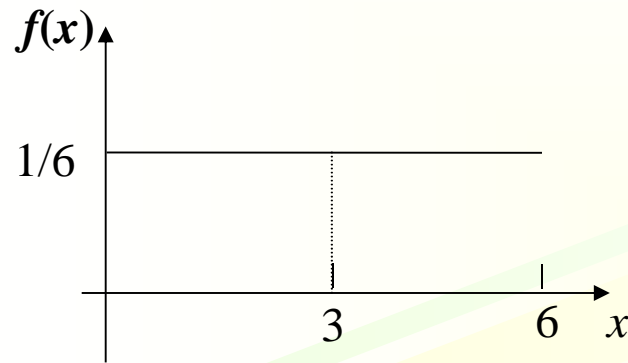
$$= p(1-p) \frac{d^2}{d^2 p} \left[ \frac{1}{p} - (1-p) - 1 \right] + \frac{1}{p}$$

$$= p(1-p) \frac{d}{dp} \left[ -\frac{1}{p^2} + 1 \right] + \frac{1}{p} = p(1-p) \left( -\frac{-2p}{p^4} \right) + \frac{1}{p} = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}$$

$$\square \sigma^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2-2p+p-1}{p^2} = \frac{1-p}{p^2}$$

## Second moment ( $E[X^2]$ ) (example)

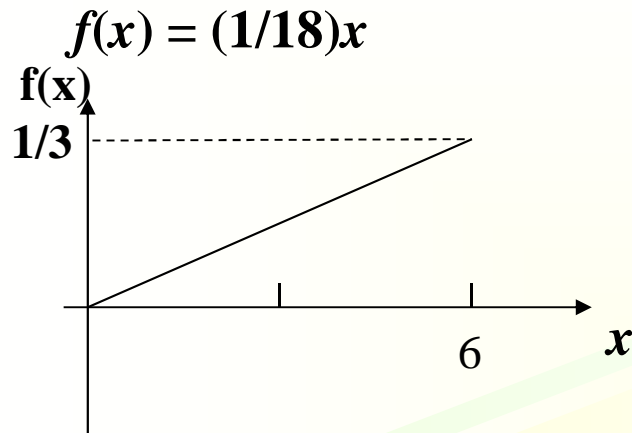
□  $(E\mathbf{x}) f(x) = 1/6$



$$\begin{aligned} E[X] &= 3 \\ E[X^2] &= \int_0^6 x^2 f(x) dx \\ &= \int_0^6 x^2 \frac{1}{6} dx \\ &= \frac{1}{6} \frac{x^3}{3} \Big|_0^6 = 12 \end{aligned}$$

$$\sigma^2 = 12 - 9 = 3$$

## Second moment ( $E[X^2]$ ) (example)



$$E[X] = 4$$

$$E[X^2] = \int_0^6 x^2 \frac{1}{18} x \, dx$$
$$= \frac{1}{18} \frac{x^4}{4} \Big|_0^6 = 18$$

$$\sigma^2 = 18 - 16 = 2$$

## Second moment ( $E[X^2]$ ) (exercise)

□ (Ex 2.20) Find the average of a random variable  $K$  whose discrete prob function is the Poisson density function.

(Sol)

$a = \lambda\tau$ , Differentiate both sides on  $a : e^a = \sum_{k=0}^{\infty} \frac{a^k}{k!}$

$$e^a = \sum_{k=0}^{\infty} k \frac{a^{k-1}}{k!} = \frac{1}{a} \sum_{k=0}^{\infty} k \frac{a^k}{k!} \quad \therefore \sum_{k=1}^{\infty} k \frac{a^k}{k!} = a e^a$$

$$E[K] = \sum_{k=1}^{\infty} k f(k) = \sum_{k=1}^{\infty} k \frac{a^k}{k!} e^{-a} = a e^a e^{-a} = a$$

Differentiate both sides on  $a : e^a = \sum_{k=0}^{\infty} k \frac{a^{k-1}}{k!}$

$$e^a = \sum_{k=1}^{\infty} k(k-1) \frac{a^{k-2}}{k!} = \frac{1}{a^2} \sum_{k=1}^{\infty} k^2 \frac{a^k}{k!} - \frac{1}{a^2} \sum_{k=1}^{\infty} k \frac{a^k}{k!} = \frac{1}{a^2} \sum_{k=1}^{\infty} k^2 \frac{a^k}{k!} - \frac{1}{a^2} a e^a = \frac{1}{a^2} \sum_{k=1}^{\infty} k^2 \frac{a^k}{k!} - a^{-1} e^a$$

$$E[K^2] = \sum_{k=1}^{\infty} k^2 \frac{a^k}{k!} e^{-a} = e^{-a} (e^a + a^{-1} e^a) a^2 = a^2 + a$$

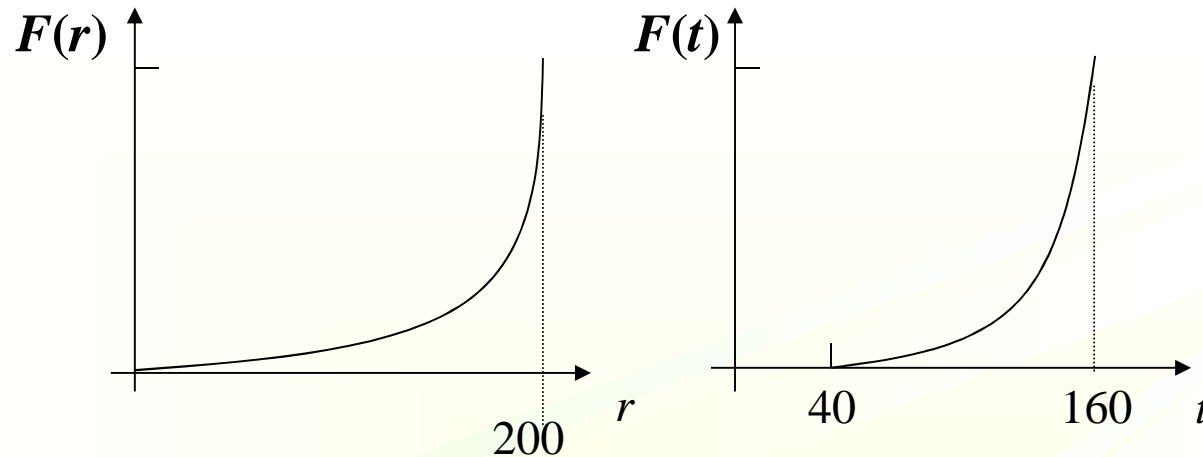
# Joint CDF and PDF

□  $F(x,y) \triangleq P[X \leq x \text{ and } Y \leq y]$

□  $f(x,y) \triangleq \frac{\partial^2 F(x,y)}{\partial x \partial y}$

□ (ex)  $R$ : rainfall  
           $T$ : temperature } independent

## Joint CDF and PDF (cont'd)



$$F(r) = \frac{10^{-6}}{8} r^3$$

$$f(r) = \frac{3}{8} 10^{-6} r^2$$

$$F(t) = \frac{(t - 40)^2}{14,400}$$

$$f(t) = \frac{2(t - 40)}{14,400}$$

$$F(r, t) = sr^3(t - 40)^2, s = 8.68 \times 10^{-12}$$

$$f(r, t) = \frac{\partial^2 (sr^3(t - 40)^2)}{\partial r \partial t} = s \frac{\partial}{\partial t} (3r^2(t - 40)^2) = 6sr^2(t - 40)$$

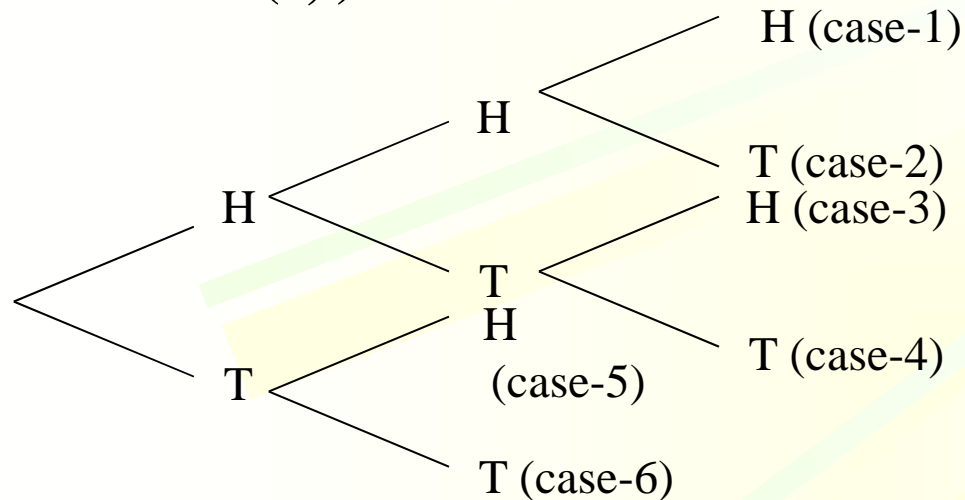
$$\begin{aligned} P[100 \leq r \leq 105 \text{ \& } 50 \leq t \leq 55] &= F(105, 55) - F(100, 50) \\ &= 0.00226 \end{aligned}$$

## Joint CDF and PDF (example)

- ❑ (Em 2.38) The first coin determines the number of subsequent flips such that ‘Head’ two more flips and ‘Tail’ one more flip.

**$T$ : number of Tails;  $C$ : number of flips**

**Obtain  $F(c,t)$ .**



$F(c,t)$	$t = 0$	1	2
$c = 2$			
3			

## Joint CDF and PDF (example)

□ (Em 2.40) Obtain  $f(c,t)$ .

$F(c,t)$	$t = 0$	1	2
$c = 2$			
3			

$f(c,t)$	$t = 0$	1	2
$c = 2$			
3			

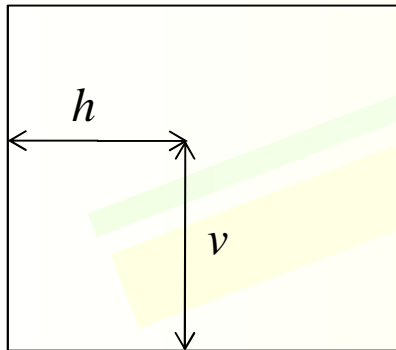
## Joint CDF and PDF (example)

□ (Em 2.39) The dart problem.

$H$ : the ratio of horizontal distance to the bottom side length

$V$ : the ratio of vertical distance to the left side length

Obtain  $F(h,v)$ .



□ (Em 2.41) Obtain  $f(h,v)$ .

# Marginal density function

- To extract the PDF of a single r.v. from a joint PDF, integrate the PDF over its range with respect to the r.v.

- $f(x) = \int_{-\infty}^{\infty} f(x,y) dy$ ,  $f(k) = \sum_{i=0}^{\infty} f(k,i)$

- (Em 2.42) Find  $f(2)$  and  $f(3)$  of the coin flipping problem.

$$f(2) = \sum_{i=0}^2 f(2,i) = f(2,0) + f(2,1) + f(2,2) = 0 + 1/4 + 1/4 = 1/2$$

$$f(3) =$$

- (Em 2.43) Find  $f(h)$  of the dart problem.

$$f(h) = \int_0^1 f(h,v) dv = \underline{\hspace{2cm}}$$

# Conditional PDF

□  $f(x | y) = \frac{f(x,y)}{f(y)}$

□ (Em 2.44) Find  $f(t|c)$  of the coin flipping problem.

$f(c,t)$	$t = 0$	1	2
$c = 2$	0	1/4	1/4
3	1/8	1/4	1/8

□  $f(c=2) = 1/2, f(c=3) = 1/2$

$f(t c)$	$t = 0$	1	2
$c = 2$			
3			

# Independence and unconditioning

□ If  $f(x,y) = f(x)f(y)$ , then  $f(x)$  and  $f(y)$  are \_\_\_\_\_

□ Unconditioning

$$f(x) = \int_{-\infty}^{\infty} f(x,y) dy = \int_{-\infty}^{\infty} f(x|y)f(y) dy$$

$f(x,y)$  is usually \_\_\_\_\_ to get, while  $f(x|y)$  and  $f(y)$  are not.

□ (Em 2.45) For the coin flipping problem, obtain  $f(t=i)$ , ( $i = 0,1,2$ ).

$$f(t=0)=f_{t=0|c=2}f(2) + f_{t=0|c=3}f(3) = 0 \times 1/2 + 1/4 \times 1/2 = 1/8$$

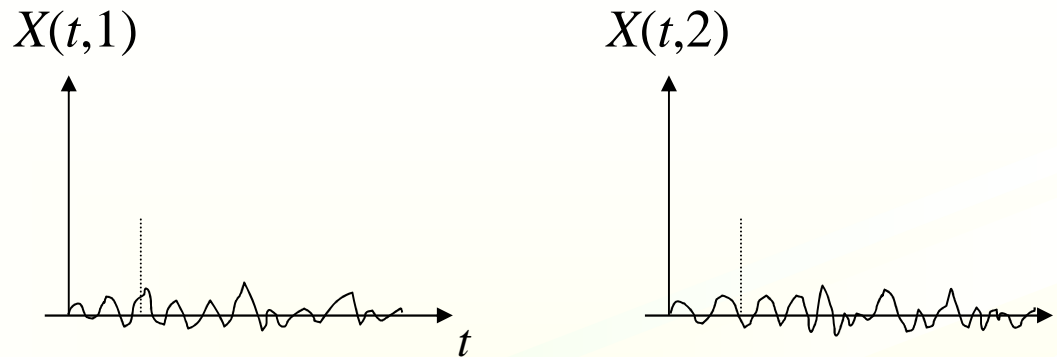
$$f(t=1)=f_{t=1|c=2}f(2) + f_{t=1|c=3}f(3) = \underline{\hspace{2cm}}$$

$$f(t=2)=f_{t=2|c=2}f(2) + f_{t=2|c=3}f(3) = \underline{\hspace{2cm}}$$

# Stochastic processes

- ❑ Process: a series of \_\_\_\_\_
- ❑ A family of r.v.'s  $\{X(t)|t \in T\}$ , defined on a given probability space, indexed by the parameter  $t$ , where  $t$  varies over an index set  $T$  (discrete time process or continuous time process)
- ❑ The values assumed by  $X(t)$  are called \_\_\_\_\_
- ❑ Set of states is \_\_\_\_\_ (discrete (chain) or continuous)
- ❑  $\{X(t,s) \mid s \in \Omega, t \in T\}$ 
  - $t = t_1 ; X_{t_1}(s) = X(t_1,s)$  (random variable)
  - $t = t_2 ; X_{t_2}(s) = X(t_2,s)$
  - $s = s_1 ; X(t) = X(t, s_1)$  (sample function)
- ❑ Sample function is a realization of the process
- ❑ When both  $s$  and  $t$  are varied, we have the family of random variables constituting a stochastic process

# Stochastic processes (example)



- ❑ Select a register  $s$  from a set  $S$ , and measure the noise  $X(t,s)$  at  $t$ .
- ❑ At  $t = t_1$ , count the resistor whose noise is smaller than  $x_1$ , divide it by  $|S|$ .  $F_{x(t_1)}(x_1) = P[X(t_1) \leq x_1]$
- ❑ Repeat for  $t_2, t_3, \dots$ , and get CDF of  $X(t_2), X(t_3), \dots$
- ❑  $F(x_1, x_2) = P[X(t_1) \leq x_1 \text{ and } X(t_2) \leq x_2]$
- ❑ When  $X(t_1, s) = X(t_2, s)$ , then the r.v.'s are \_\_\_\_\_ distributed

## Stochastic processes (cont'd)

❑ Sample path: the set of \_\_\_\_\_ of the r.v.'s for particular outcomes in a stochastic process and the \_\_\_\_\_ associated with those outcomes

❑ (Em 2.46) Rolling two dices

❑ To describe a stochastic process, two r.v.'s which represent an \_\_\_\_\_ and a \_\_\_\_\_ of the events are required

Process type	Event counting	Time between events
Poisson(continuous)	Poisson distribution	_____ distribution
Bernoulli(discrete)	Binomial distribution	_____ distribution

❑ These processes are interested since they have \_\_\_\_\_ property

❑  $k$ -th interval of exponential (geometric) dist is \_\_\_\_\_ ( \_\_\_\_\_ ) dist

# Renewal theory

- A process is  $n$ th order renewal process if the \_\_\_\_\_ is the same after every  $n$  events
- Residual lifetime ( $R$ ) ( $R$  is a r.v. representing residual lifetime of a r.v.  $A$ )

$$r(\tau - t_e | t_e) \triangleq a(\tau | \tau > t_e) = \frac{a(\tau)}{P[A > t_e]} = \frac{a(\tau)}{1 - \int_0^{t_e} a(s) ds}$$

$$t = \tau - t_e$$

$$r(t | t_e) = \frac{a(t + t_e)}{1 - \int_0^{t_e} a(s) ds}$$

$$\bar{r} = \frac{\overline{f^2}}{2\bar{f}} \quad (f \text{ is original density})$$

# Memoryless property

- When the \_\_\_\_\_ of a process has the same distribution as the original process
- (Em 2.49) Show geometric distribution has memoryless property.

$$\begin{aligned} f_k &= (1-p)^{k-1} p \\ r_k &= \frac{f_{k+n}}{1 - \sum_{m=1}^n f_m} = \frac{(1-p)^{k+n-1} p}{1-p \sum_{m=1}^n (1-p)^{m-1}} = \frac{(1-p)^{k+n-1} p}{1-p \sum_{m=0}^{n-1} (1-p)^m} = \frac{(1-p)^{k+n-1} p}{1-p \frac{1-(1-p)^n}{1-(1-p)}} \quad (* \sum_{m=0}^n x^m = \frac{1-x^{n+1}}{1-x} *) \\ &= (1-p)^{k-1} p \end{aligned}$$

( $r_k$  is the number of \_\_\_\_\_ trials until success while  $n$  is the number of failures)

- (Ex 2.21) Show exponential distribution has memoryless property.

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t} \\ r(t | t_e) &= \frac{\lambda e^{-\lambda(t+t_e)}}{1 - \int_0^{t_e} \lambda e^{-\lambda t} dt} = \frac{\lambda e^{-\lambda t} e^{-\lambda t_e}}{1 - (-e^{-\lambda t})_0^{t_e}} = \frac{\lambda e^{-\lambda t} e^{-\lambda t_e}}{e^{-\lambda t_e}} = \lambda e^{-\lambda t} \end{aligned}$$

## Memoryless property (example)

□ (ex) Interarrival time is exponentially distributed with  $\lambda = 2$

(a) What is the prob. that a job arrives in  $t = 1$ ?

$$P(T \leq 1) = \int_0^1 2e^{-2t} dt = -e^{-2t} \Big|_0^1 = 1 - e^{-2}$$

(b) Provided that no job arrives in  $t = 10$ , what is the prob. of a job arrival in  $t = 11$ ?

$$\begin{aligned} P(T \leq 2 | 1 < T) &= \frac{P(1 < T \leq 2)}{P(1 < T)} = \frac{\int_1^2 2e^{-2t} dt}{1 - (1 - e^{-2})} \\ &= \frac{-e^{-2t} \Big|_1^2}{e^{-2}} = \frac{e^{-2} - e^{-4}}{e^{-2}} = 1 - e^{-2} \end{aligned}$$

## Memoryless property (example)

□  $N$ : r.v. for the no. of jobs arriving in  $(0,t)$

$X$ : r.v. for the interarrival time

If  $N$  is poisson distribution with  $\lambda t$ , what is the distribution of  $X$ ?

$$(\text{Ans}) P[X > t] = P[N = 0] = \frac{e^{-\lambda t} (\lambda t)^0}{0!} = e^{-\lambda t}$$

$$F(t) = P[X \leq t] = 1 - P[X > t] = 1 - e^{-\lambda t}$$

Poisson arrival  $\equiv$  \_\_\_\_\_ interarrival

# Poisson arrival takes a random look

□  $k$  arrivals in  $[0,t]$

□ Uniform distribution means an arrival in each  $k$  subintervals;

$$\text{prob} = \frac{h_1}{t} k \frac{h_2}{t} (k-1) \cdots \frac{h_k}{t} 1 = \frac{k!}{t^k} h_1 h_2 \cdots h_k$$

□ An arrival in each subinterval provided  $k$  Poisson arrivals in time  $t$

$$\text{prob} = \frac{\frac{(\lambda h_1)^1 e^{-\lambda h_1}}{1!} \cdots \frac{(\lambda h_k)^1 e^{-\lambda h_k}}{1!}}{\frac{(\lambda t)^k e^{-\lambda t}}{k!}} = \frac{k!}{t^k} h_1 h_2 \cdots h_k$$

## Poisson arrival (example)

- (Em 2.50) At a bus stop, the interarrival time of buses is exponentially distributed with a rate of  $\lambda$ . If I walk up to the bus stop, how long do I have to wait?
- (Ex 2.22) If buses arrive at intervals of  $\frac{1}{2}$  hour and 1 hour alternatively, how long would you wait on the average?

# Erlang distribution

- $r$  sequential phases have independent identical \_\_\_\_\_ distributions
- $F(t) = 1 - \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, t \geq 0, \lambda > 0, r = 1, 2, \dots$   
$$f(t) = \frac{\lambda^r t^{r-1} e^{-\lambda t}}{(r-1)!}$$
- A component has  $N$  peak stresses in  $(0, t]$ , which is Poisson distributed with parameter  $\lambda t$ . For component withstanding  $(r-1)$  peak stresses (so  $r$ th occurrence causes \_\_\_\_\_),
  - $X$ : lifetime
  - $[X > t] = [N < r]$
  - $F(t) = 1 - R(t) = 1 - P[X > t] = 1 - P[N < r] = 1 - \sum_{k=0}^{r-1} P[N = k] = 1 - \sum_{k=0}^{r-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!}$
  - Exponential is a special case of Erlang with  $r = \underline{\hspace{1cm}}$

# Hypoexponential distribution

□ Similar to Erlang distribution, but the time in each sequential phase is independent and \_\_\_\_\_ distributed

□ 2-stage:  $X \sim \text{HYPO}(\lambda_1, \lambda_2)$ ,  $\lambda_1 \neq \lambda_2$

$$f(t) = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} (e^{-\lambda_1 t} - e^{-\lambda_2 t}), \quad t > 0$$

$$F(t) = 1 - \frac{\lambda_2}{\lambda_2 - \lambda_1} e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_2 - \lambda_1} e^{-\lambda_2 t}, \quad t \geq 0$$

# Hyperexponential distribution

- A process consisting of alternate phases, while experiencing one and only one of the independent \_\_\_\_\_ distributed phases

- $f(t) = \sum_{i=1}^k \alpha_i \lambda_i e^{-\lambda_i t}, \quad t > 0, \lambda_i > 0, \alpha_i > 0, \sum_{i=1}^k \alpha_i = 1$

$$F(t) = \sum_{i=1}^k \alpha_i (1 - e^{-\lambda_i t})$$

# Normal (Gaussian) distribution

- ❑ Central limit theorem: Mean of a sample of  $n$  mutually independent r.v.'s is normally distributed in the limit  $n \rightarrow \infty$
- ❑  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ ,  $-\infty < x < \infty$ ,  $\mu$  : mean
- ❑  $X \sim N(\mu, \sigma^2)$
- ❑ (ex) errors of measurement
- ❑ No closed form  $F(x)$ ; use table for  $Z \sim N(0,1)$  (standard normal dist)
- ❑  $F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt$
- ❑  $F_Z(-z) = 1 - F_Z(z)$ ;  $F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right)$
- ❑ (ex)  $N(200,256)$  signal is received. What is the prob that the signal is greater than 240mV?  
(Ans)  $P[X > 240] = 1 - P[X \leq 240] = 1 - F_Z\left(\frac{240-200}{16}\right) = 1 - F_Z(2.5)$   
 $= 0.0062$

# Standard normal distribution table

Table 3 Distribution Function of Standard Normal Random Variable

z	0	1	2	3	4	5	6	7	8	9
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9430	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9648	.9656	.9664	.9671	.9678	.9686	.9693	.9700	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9762	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9874	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.	.9987	.9990	.9993	.9995	.9997	.9998	.9998	.9999	.9999	1.0000

Note 1: If a normal variable  $X$  is not "standard," its value must be standardized:  $Z = (X - \mu)/\sigma$ . Then  $F_X(x) = F_Z\left(\frac{x - \mu}{\sigma}\right)$ .

Note 2: For  $z \geq 4$ , use  $F_Z(z) = 1$  to four decimal places; for  $z \leq -4$ ,  $F_Z(z) = 0$  to four decimal places.

Note 3: The entries opposite  $z = 3$  are for 3.0, 3.1, 3.2, etc.

Note 4: For  $z < 0$  use  $F_Z(z) = 1 - F_Z(-z)$ .

Reprinted by permission from *Introduction to Probability and Statistics*, 2nd ed., by B. W. Lindgren and G. W. McElrath (copyright © 1966 by B. W. Lindgren and G. W. McElrath. Published by Macmillan Publishing Co., Inc.)

# Weibull distribution

- $f(t) = \lambda \alpha t^{(\alpha-1)} e^{-\lambda t^\alpha}$
- $F(t) = 1 - e^{-\lambda t^\alpha}$
- Fault modeling
- Exponential dist is a special case of Weibull dist with  $\alpha = 1$