

Machine Learning

Lecture 2 (Spring-2021)

EDA – Part 2

Dr. Tariq Mahmood

Assistant Professor IBA and Head of BDA-LAB

tmahmood@iba.edu.pk

Anova

$$SS_{between} = \sum n_j (\bar{X}_j - \bar{X})^2$$

where

- $\bar{X}_j$ denotes a group mean;
- $\bar{X}$ is the overall mean;
- n_j is the sample size per group.

For our example, this results in

$$SS_{between} = 10 (99.2 - 101.7)^2 + 10 (112.6 - 101.7)^2 + 10 (93.3 - 101.7)^2 = 1956.2$$

Next, for m groups,

$$df_{between} = m - 1$$

so $df_{between} = 3 - 1 = 2$ for our example data.

$$MS_{between} = \frac{SS_{between}}{df_{between}}$$

For our example, that'll be

$$\frac{1956.2}{2} = 978.1$$

Anova

We now turn to the **within-groups** variance. First off,

$$SS_{within} = \sum (X_i - \bar{X}_j)^2$$

where

- $\bar{X}_j$ denotes a group mean;
- X_i denotes an individual observation (“data point”).

For our example, this'll be

$$SS_{within} = (90 - 99.2)^2 + (87 - 99.2)^2 + \dots + (96 - 93.3)^2 = 4294.1$$

for n independent observations and m groups,

$$df_{within} = n - m$$

So for our example that'll be $= 30 - 3 = 27$.

$$MS_{within} = \frac{SS_{within}}{df_{within}}$$

Anova

For our example, this results in

$$\frac{4294.1}{27} = 159$$

We're now ready to calculate the **F-statistic**:

$$F = \frac{MS_{between}}{MS_{within}}$$

which results in

$$\frac{978.1}{159} = 6.15$$

Finally,

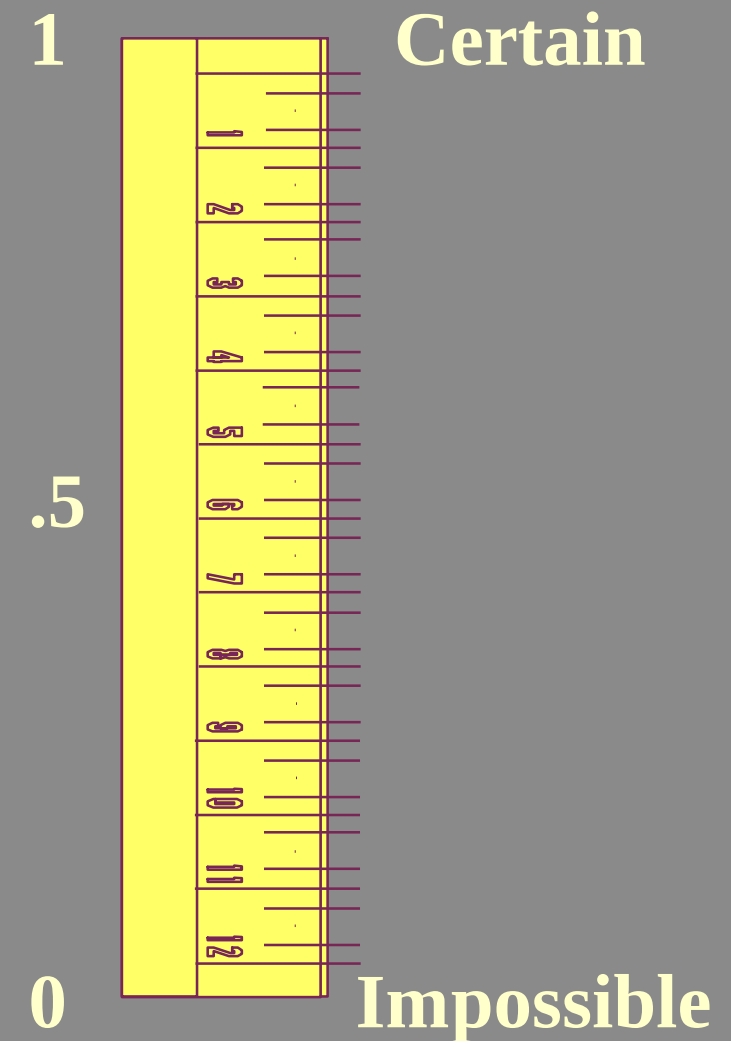
$$P = P(F(2, 27) > 6.15) = 0.0063$$

Optionally, the **effect size** η^2 is calculated as

$$\text{Effect size } \eta^2 = \frac{SS_{between}}{SS_{between} + SS_{within}}$$

Probability

- Probability is the likelihood that the event will occur.
- Two Conditions:
 - Value is between 0 and 1.
 - Sum of the probabilities of all events must be 1.

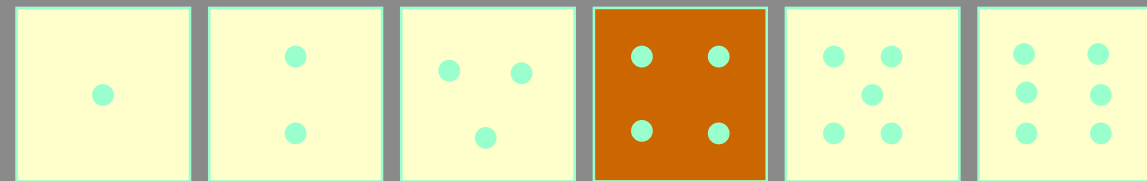


Random Variable

- A numerical description of the outcome of an experiment

Example: Discrete RV: countable # of outcomes

Throw a die twice: Count the number of times 4 comes up (0, 1, or 2 times)



Discrete Random Variable

- **Discrete Random Variable:**
 - **Obtained by Counting (0, 1, 2, 3, etc.)**
 - **Usually finite by number of different values**

e.g.

Toss a coin 5 times. Count the number of tails. (0, 1, 2, 3, 4, or 5 times)

Random Variable

- A numerical description of the outcome of an experiment

Example: Continuous RV:

- The Value of the DJIA
- Time to repair a failed machine
- RV Given by Capital Letters X & Y
- Specific Values Given by lower case

Probability Distribution

Characterization of the possible values that a RV may assume along with the probability of assuming these values.

Discrete Probability Distribution

- List of all possible $[x_i, p(x_i)]$ pairs
 - x_i = value of random variable
 - $P(x_i)$ = probability associated with value
- Mutually exclusive (nothing in common)
- Collectively exhaustive (nothing left out)
 - $0 \leq p(x_i) \leq 1$
 - $\Sigma P(x_i) = 1$

Weekly Demand of a Slow-Moving Product

Probability Mass Function

Demand, x	Probability, $p(x)$
0	0.1
1	0.2
2	0.4
3	0.3
4 or more	0

Weekly Demand of a Slow-Moving Product

A Cumulative Distribution Function:

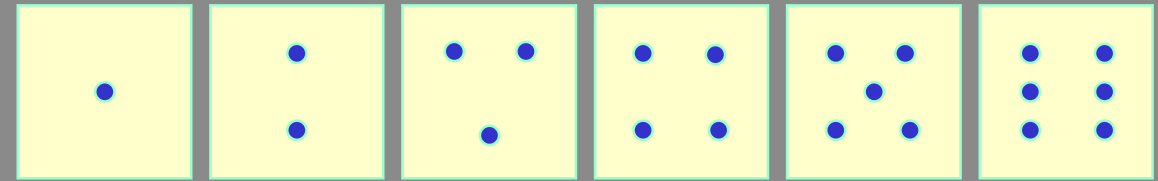
Probability that RV assume a value $\leq$ a given value, x

Demand, x	Cumulative Probability, $P(x)$
0	0.1
1	0.3
2	0.7
3	1

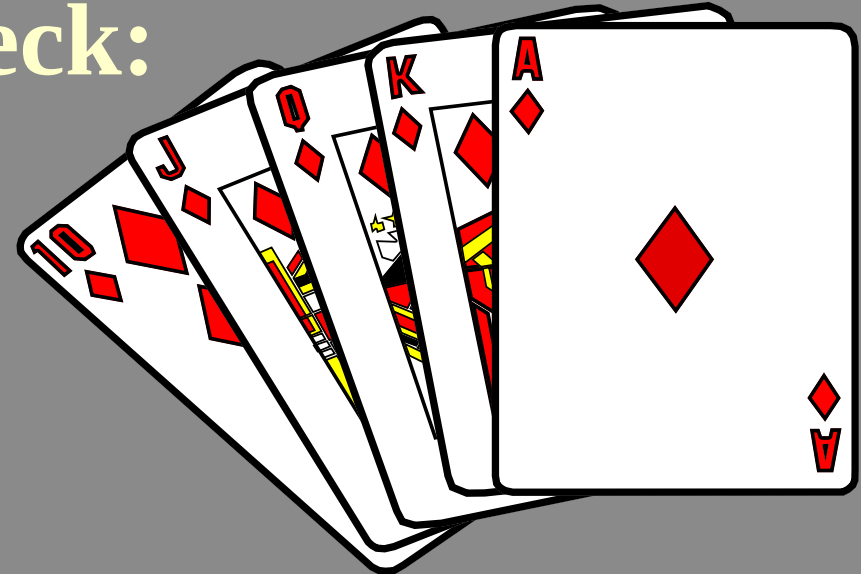
Sample Spaces

Collection of all Possible Outcomes

e.g. All 6 faces of a die:



e.g. All 52 cards of a bridge deck:



Events

- Simple Event: Outcome from a Sample Space with 1 Characteristic

e.g. A *Red Card* from a deck of cards.

- Joint Event: Involves 2 Outcomes Simultaneously

e.g. An *Ace* which is also a *Red Card* from a deck of cards.

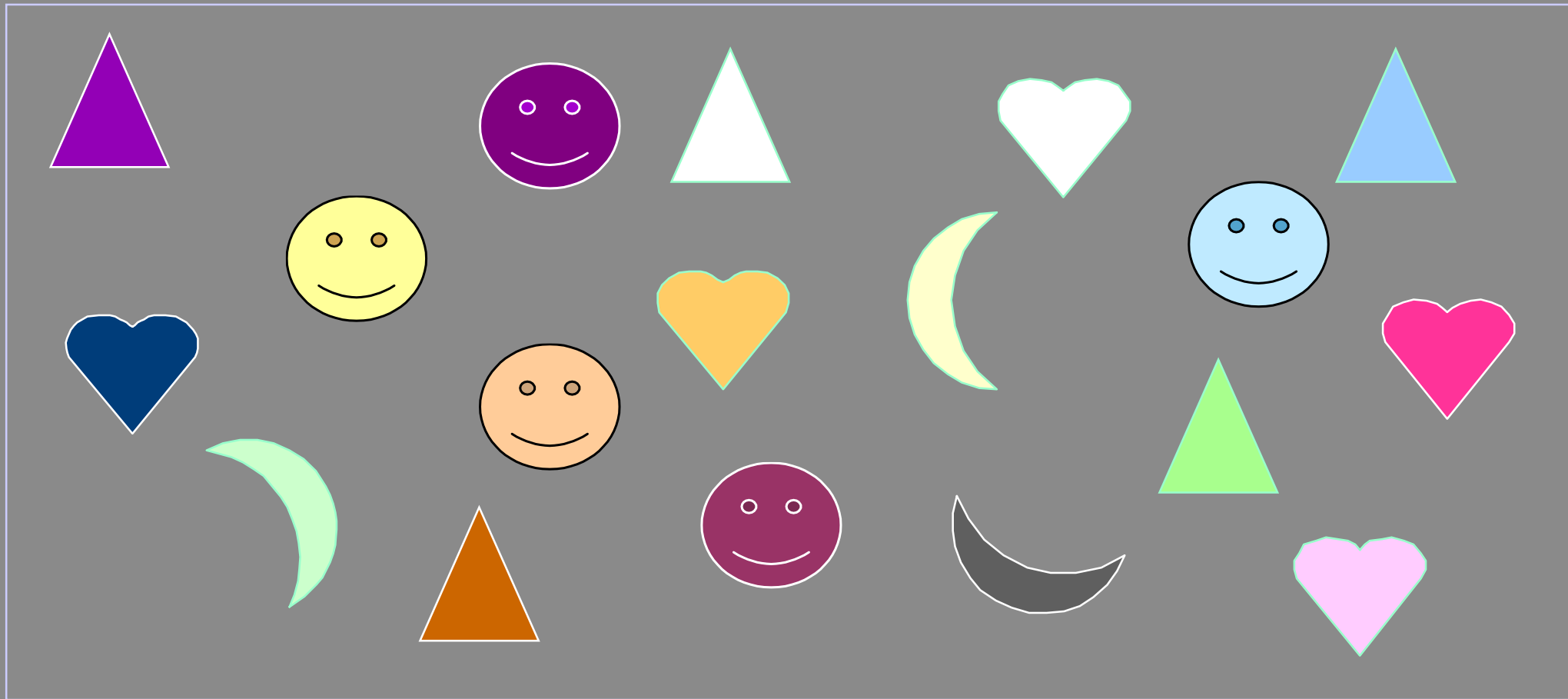
Visualizing Events

- Contingency Tables

	Ace	Not Ace	Total
Black	2	24	26
Red	2	24	26
Total	4	48	52

Simple Events

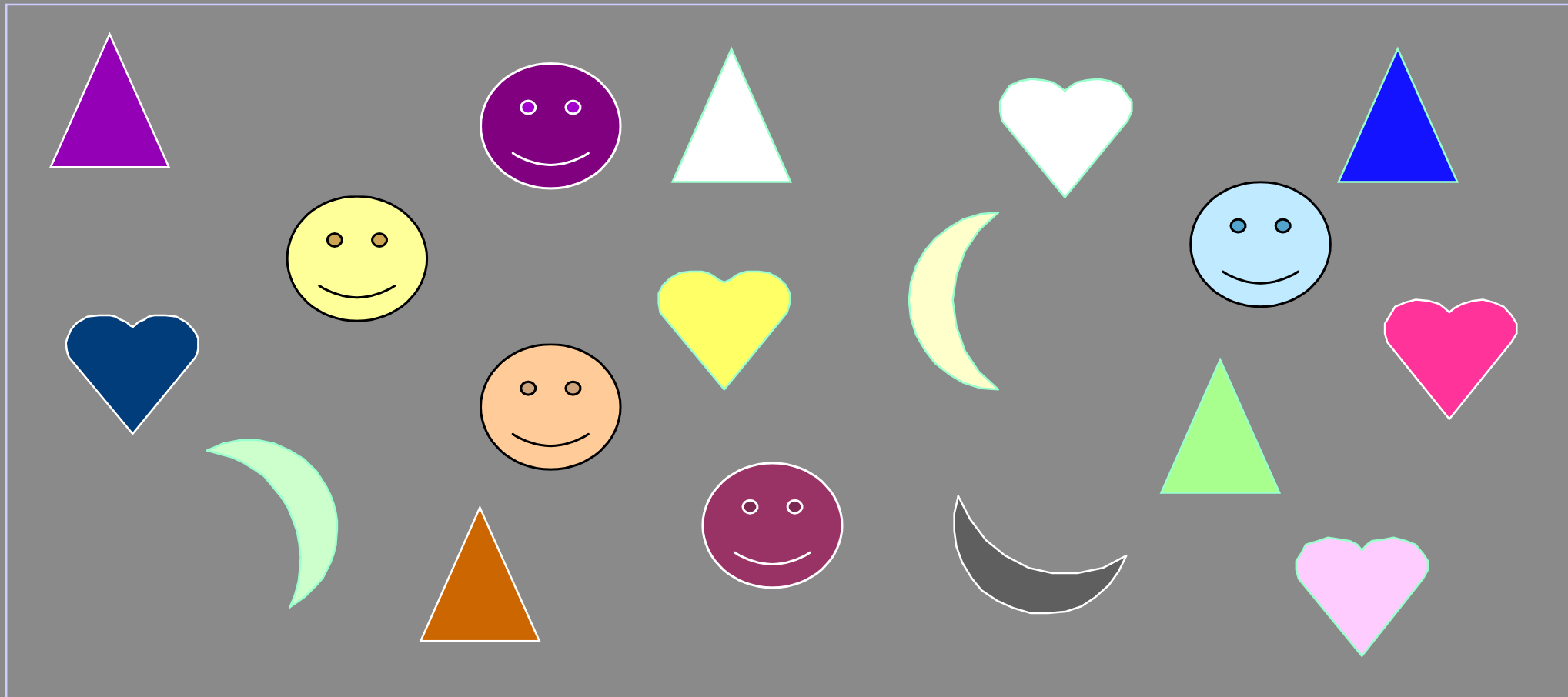
The Event of a Happy Face



There are 5 happy faces in this collection of 18 objects

Joint Events

The Event of a Happy Face **AND** Light Colored



3 Happy Faces which are light in color

Special Events

Null event

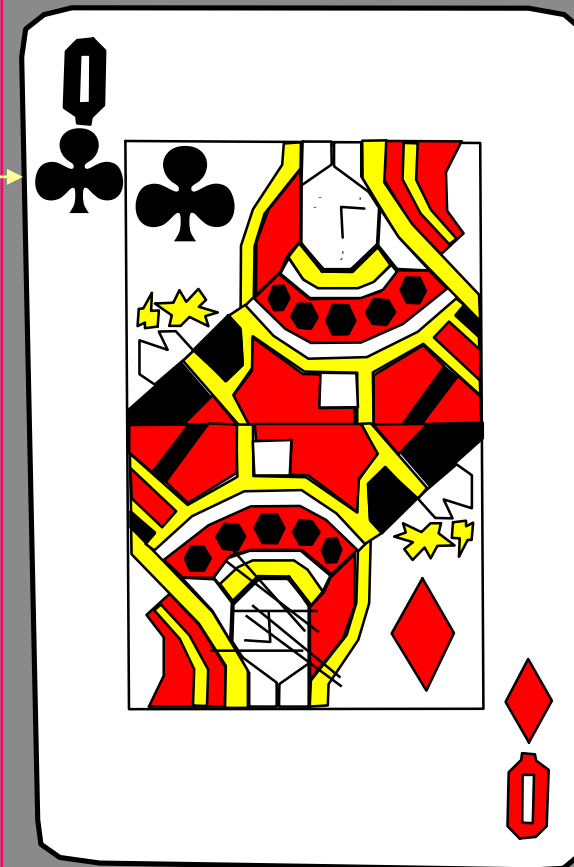
Club & diamond on
1 card draw

Complement of event

For event A,

All events not in A: A'

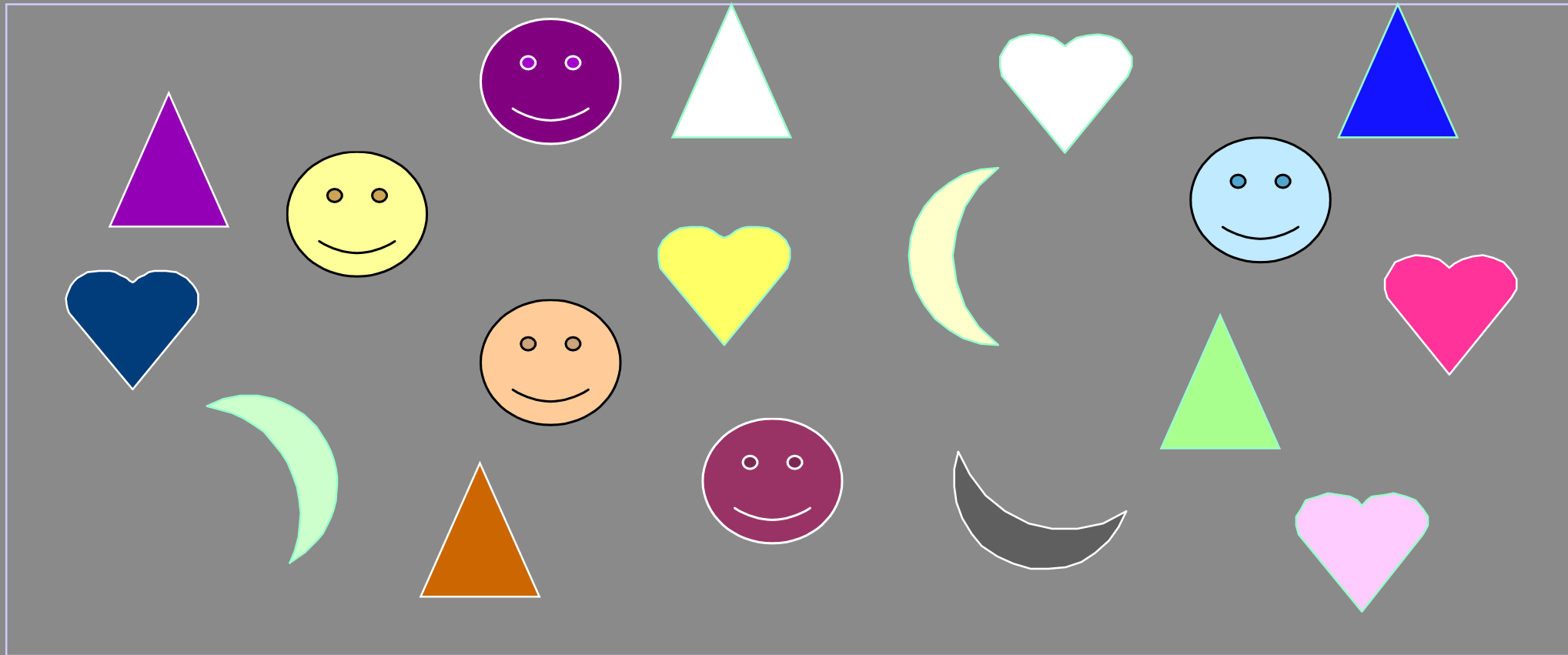
Null Event



Dependent or Independent Events

The Event of a Happy Face *GIVEN* it is Light Colored

E = Happy Face | Light Color



3 Items: 3 Happy Faces *Given* they are Light Colored

Contingency Table

A Deck of 52 Cards

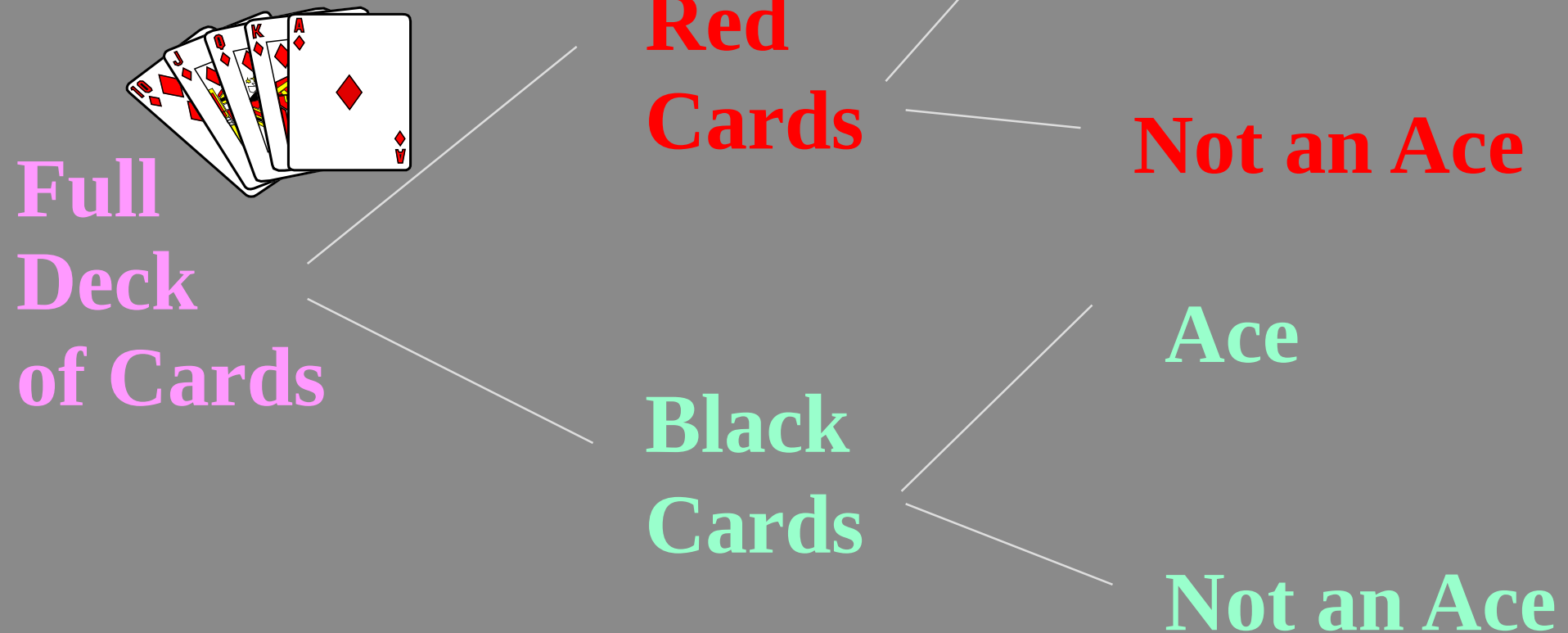
Red Ace

	Ace	Not an Ace	Total
Red	2	24	26
Black	2	24	26
Total	4	48	52

Sample Space

Tree Diagram

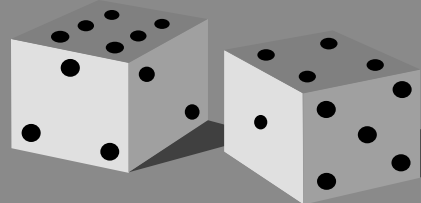
Event Possibilities



Computing Probability

- The Probability of an Event, E:

$$P(E) = \frac{\text{Number of Event Outcomes}}{\text{Total Number of Possible Outcomes in the Sample Space}}$$

$$= \frac{X}{T} \quad \text{e.g. } P(\text{ ) = 2/36$$

(There are 2 ways to get one 6 and the other 4)

- Each of the Outcome in the Sample Space equally likely to occur.

Two or More Random Variables

Frequency of applications during a given week

Number of Applications					
Interest Rate	5	6	7	8	Total
7.00%	3	4	6	2	15
7.50%	2	4	3	1	10
8%	3	1	1	0	5
Total	8	9	10	3	30

Two or More Random Variables

Joint Probability Distribution

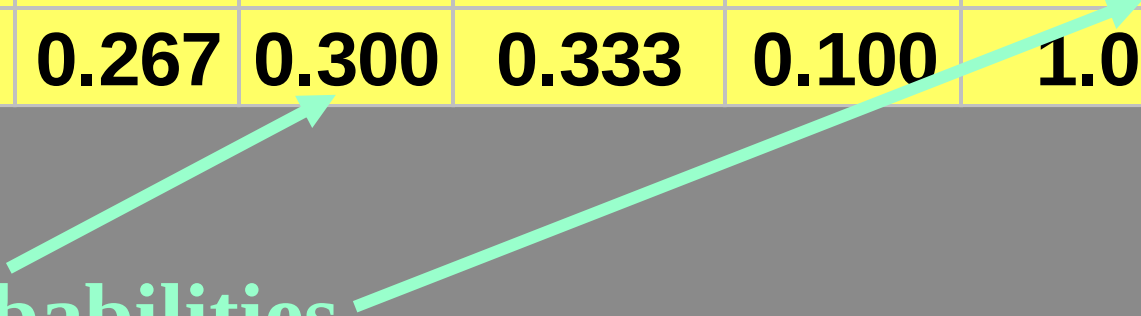
Number of Applications					
Interest Rate	5	6	7	8	Total
7.00%	0.100	0.133	0.200	0.067	0.500
7.50%	0.067	0.133	0.100	0.033	0.333
8.00%	0.100	0.033	0.033	0.000	0.167
Total	0.267	0.300	0.333	0.100	1.000

Two or More Random Variables

Joint Probability Distribution

Number of Applications					
Interest Rate	5	6	7	8	Total
7.00%	0.100	0.133	0.200	0.067	0.500
7.50%	0.067	0.133	0.100	0.033	0.333
8.00%	0.100	0.033	0.033	0.000	0.167
Total	0.267	0.300	0.333	0.100	1.000

Marginal Probabilities



Computing Joint Probability

The Probability of a *Joint Event*, *A and B*:

$$P(A \text{ and } B)$$

$$= \frac{\text{Number of Event Outcomes from both A and B}}{\text{Total Number of Possible Outcomes in Sample Space}}$$

e.g. $P(\text{Red Card and Ace})$

$$= \frac{2 \text{ Red Aces}}{52 \text{ Total Number of Cards}} = \frac{1}{26}$$

Joint Probability Using Contingency Table

Event	Event		Total
	B_1	B_2	
A_1	$P(A_1 \text{ and } B_1)$	$P(A_1 \text{ and } B_2)$	$P(A_1)$
A_2	$P(A_2 \text{ and } B_1)$	$P(A_2 \text{ and } B_2)$	$P(A_2)$
Total	$P(B_1)$	$P(B_2)$	1

Joint Probability

Marginal (Simple) Probability

Computing Compound Probability

The Probability of a *Compound Event*, **A or B**:

$$P(A \text{ or } B) = \frac{\text{Numer of Event Outcomes from Either A or B}}{\text{Total Outcomes in the Sample Space}}$$

e.g.

P(Red Card *or* Ace)

$$= \frac{4 \text{ Aces} + 26 \text{ Red Cards} - 2 \text{ Red Aces}}{52 \text{ Total Number of Cards}} = \frac{28}{52} = \frac{7}{13}$$

Compound Probability

Addition Rule

$$P(\textcolor{red}{A}_1 \text{ or } \textcolor{blue}{B}_1) = P(\textcolor{red}{A}_1) + P(\textcolor{blue}{B}_1) - P(\textcolor{red}{A}_1 \text{ and } \textcolor{blue}{B}_1)$$

Event	Event		Total
	$\textcolor{blue}{B}_1$	B_2	
$\textcolor{red}{A}_1$	$P(\textcolor{red}{A}_1 \text{ and } \textcolor{blue}{B}_1)$	$P(\textcolor{red}{A}_1 \text{ and } B_2)$	$P(\textcolor{red}{A}_1)$
A_2	$P(A_2 \text{ and } \textcolor{blue}{B}_1)$	$P(A_2 \text{ and } B_2)$	$P(A_2)$
Total	$P(\textcolor{blue}{B}_1)$	$P(B_2)$	1

For **Mutually Exclusive** Events: $P(A \text{ or } B) = P(A) + P(B)$

Computing Conditional Probability

The Probability of **Event A** given that Event B has occurred:

$$P(\mathbf{A} \mid \mathbf{B}) = \frac{P(A \text{ and } B)}{P(B)}$$

e.g.

$$P(\mathbf{Red Card} \text{ given that it is an Ace}) = \frac{2 \text{ Red Aces}}{4 \text{ Aces}} = \frac{1}{2}$$

Conditional Probability Using Contingency Table

Conditional Event: Draw 1 Card. Note Kind & Color

Type	Color		Total
	Red	Black	
Ace	2	2	4
Non-Ace	24	24	48
Total	26	26	52

Revised
Sample
Space

$$P(\text{Ace} \mid \text{Red}) = \frac{P(\text{Ace AND Red})}{P(\text{Red})} = \frac{2 / 52}{26 / 52} = \frac{2}{26}$$

Conditional Probability and Statistical Independence

Conditional Probability:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

Multiplication Rule:

$$\begin{aligned} P(A \text{ and } B) &= P(A \mid B) \cdot P(B) \\ &= P(B \mid A) \cdot P(A) \end{aligned}$$

Conditional Probability and Statistical Independence *(continued)*

Events are Independent:

$$P(A \mid B) = P(A)$$

$$\text{Or, } P(B \mid A) = P(B)$$

$$\text{Or, } P(A \text{ and } B) = P(A) \cdot P(B)$$

Events A and B are *Independent* when the probability of one event, A is *not affected* by another event, B.

Discrete Probability Distribution Example

Event: Toss 2 Coins.

Count # Tails.



Probability distribution

Values

probability

0

$1/4 = .25$

1

$2/4 = .50$

2

$1/4 = .25$

Discrete Random Variable

Summary Measures

Expected value (The mean)

Weighted average of the probability distribution

$$\mu = E(X) = \sum x_i p(x_i)$$

In slow-moving product demand example,
the expected value is :

$$E(X) = 0 \times 0.1 + 1 \times .2 + 2 \times .4 + 3 \times .3 = 1.9$$

The average demand on the long run is 1.9

Discrete Random Variable Summary Measures

Variance

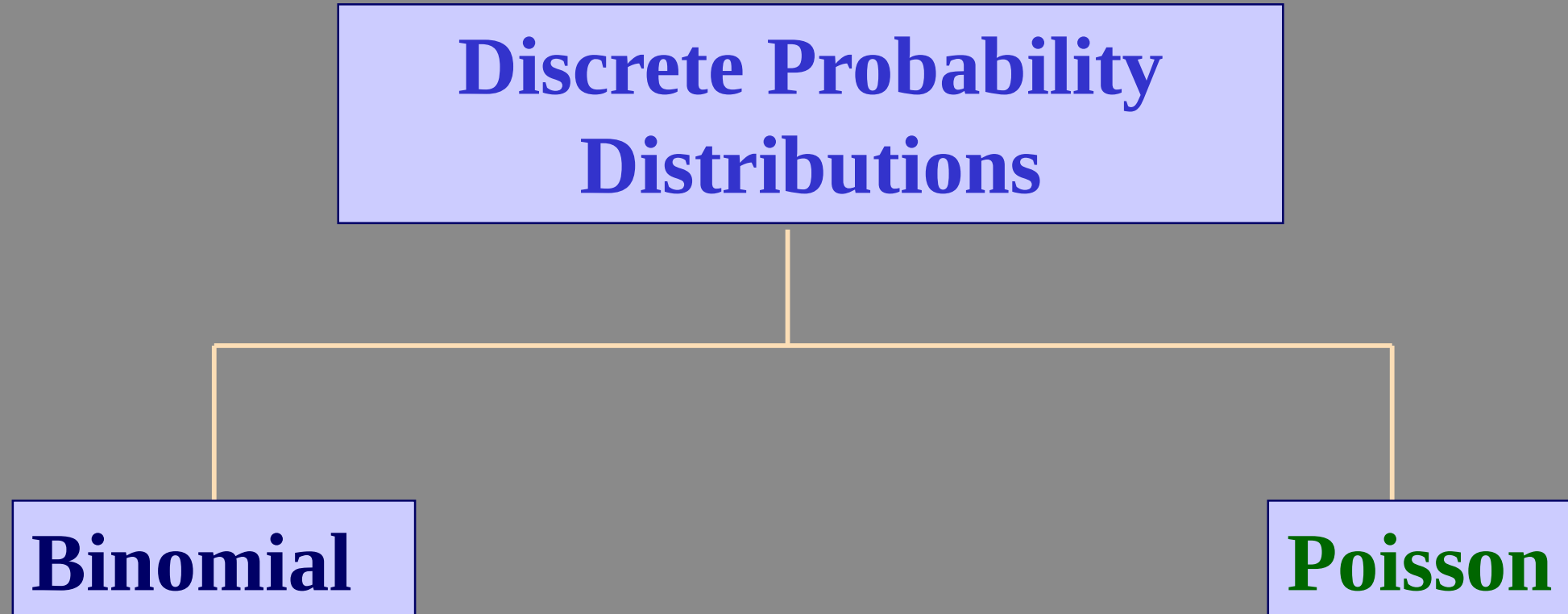
Weighted average squared deviation about mean

$$\text{Var}[X] = \sigma^2 = E[(x_i - E(X))^2] = \sum (x_i - E(X))^2 p(x_i)$$

For the Product demand example, the variance is:

$$\begin{aligned}\text{Var}[X] = \sigma^2 = & (0 - 1.9)^2(.1) + (1 - 1.9)^2(.2) + \\ & (2 - 1.9)^2(.4) + (3 - 1.9)^2(.3) = .89\end{aligned}$$

Important Discrete Probability Distribution Models



Bernoulli Distribution

- Two possible mutually exclusive outcomes with constant probabilities of occurrence

“Success” ($x=1$) or “failure” ($x=0$)

Example : Response to telemarketing

The probability mass function is

- $p(x) = p$ if $x=1$
- $P(x) = 1 - p$ if $x=0$

Where p is the probability of success

Binomial Distribution

- **‘N’ identical trials**
 - **Example: 15 tosses of a coin, 10 light bulbs taken from a warehouse**
- **2 mutually exclusive outcomes on each trial**
 - **Example: Heads or tails in each toss of a coin, defective or not defective light bulbs**

Binomial Distributions

- **Constant Probability for each Trial**
 - **Example:** Probability of getting a tail is the same each time we toss the coin and each light bulb has the same probability of being defective
- **2 Sampling Methods:**
 - **Infinite Population Without Replacement**
 - **Finite Population With Replacement**
- **Trials are Independent:**
 - **The Outcome of One Trial Does Not Affect the Outcome of Another**

Binomial Probability Distribution Function

$$P(X) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

$P(X)$ = probability that X successes given a knowledge of n and p

X = number of 'successes' in sample, ($X = 0, 1, 2, \dots, n$)

p = probability of each 'success'

n = sample size

Tails in 2 Tosses of Coin

<u>X</u>	<u>$P(X)$</u>
0	$1/4 = .25$
1	$2/4 = .50$
2	$1/4 = .25$

Binomial Distribution Characteristics

Mean

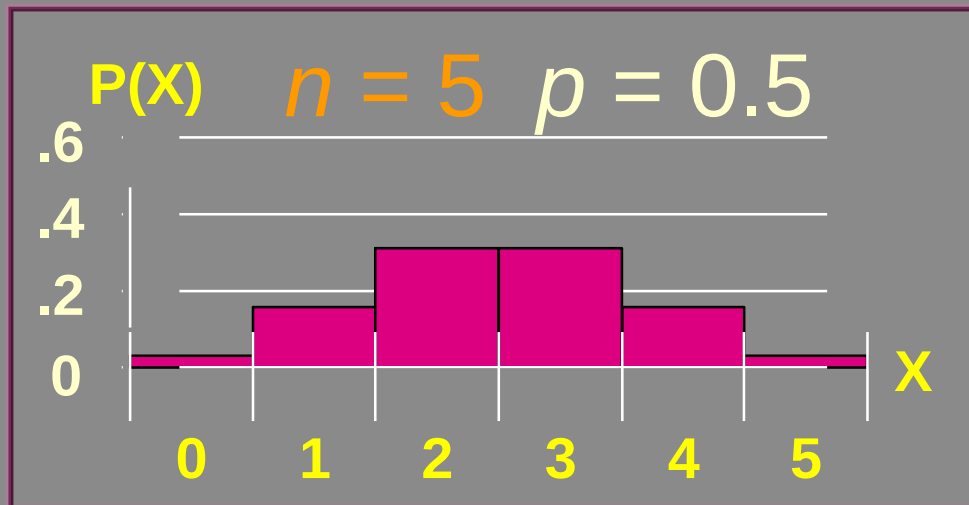
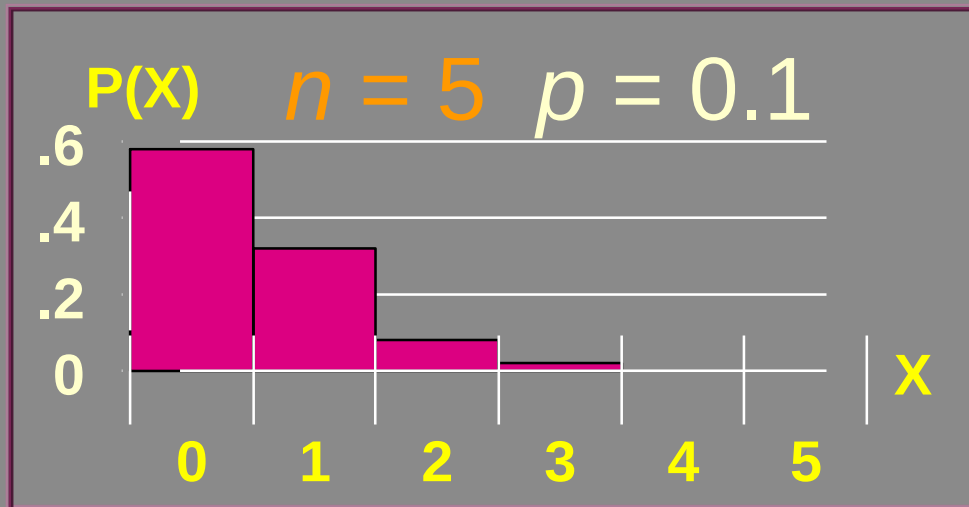
$$\mu = E(X) = np$$

$$\text{e.g. } \mu = 5 (.1) = .5$$

Standard Deviation

$$\sigma = \sqrt{np(1-p)}$$

$$\begin{aligned} \text{e.g. } \sigma &= \sqrt{5(.5)(1 - .5)} \\ &= 1.118 \end{aligned}$$



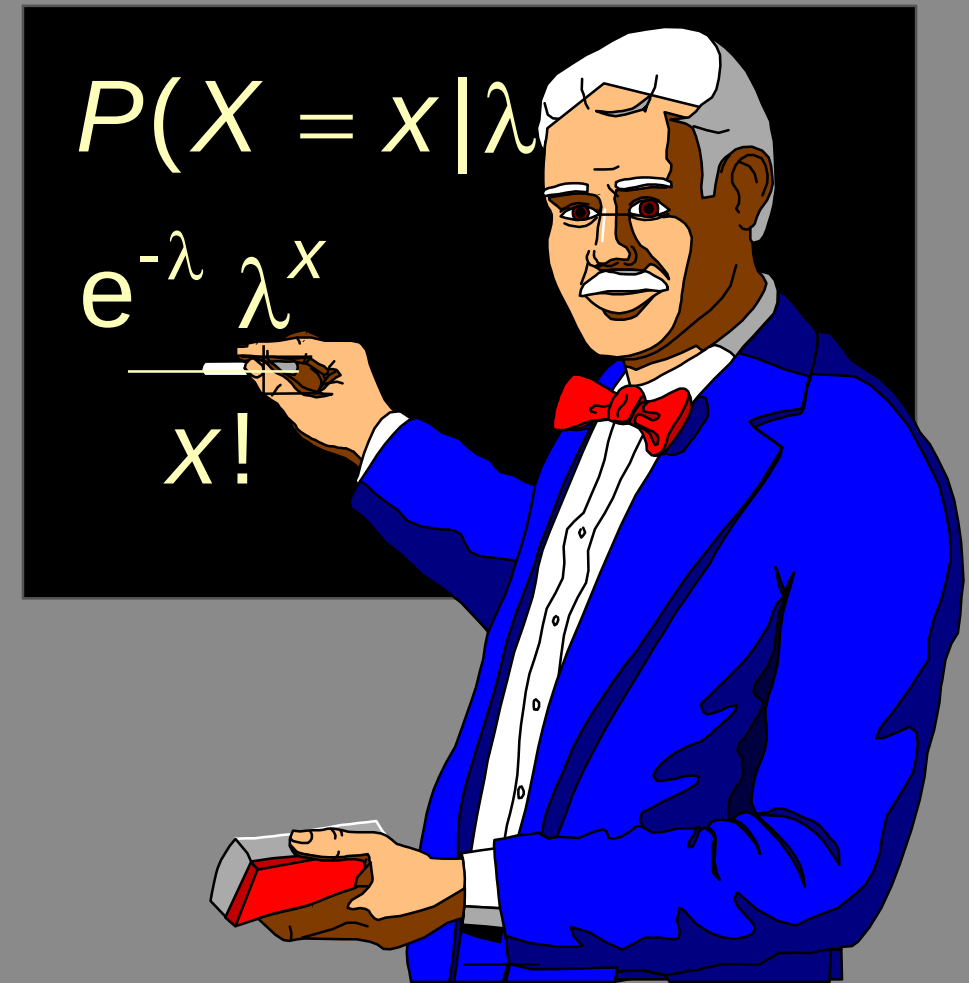
Poisson Distribution

Poisson process:

- Discrete events in an 'interval'
 - The probability of one success in an interval is stable
 - The probability of more than one success in this interval is 0
- Probability of success is Independent from interval to Interval

Examples:

- # Customers arriving in 15 min
- # Defects per case of light bulbs



Poisson Distribution Function

$$P(X) = \frac{e^{-\lambda} \lambda^X}{X!}$$

$P(X)$ = probability of X successes given λ

λ = expected (mean) number of 'successes'

e = 2.71828 (base of natural logs)

X = number of 'successes' per unit

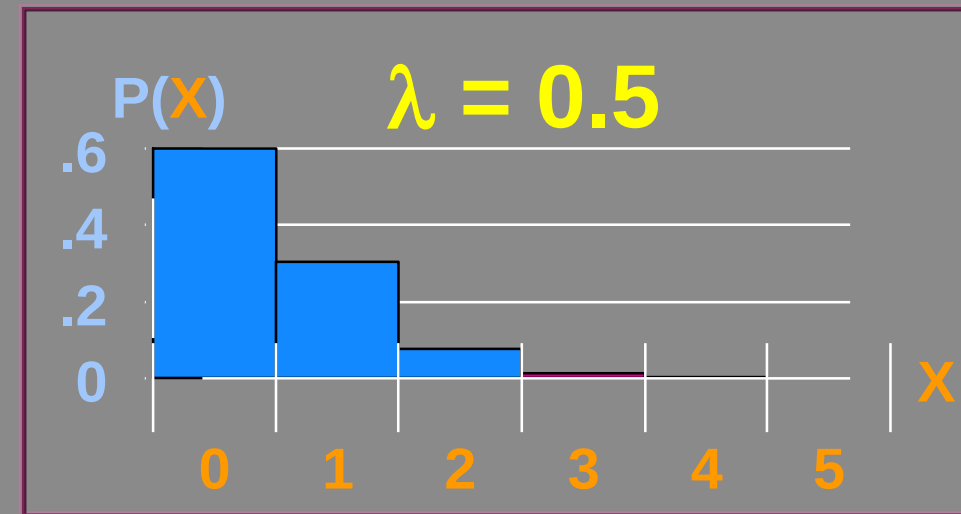
e.g. Find the probability of 4 customers arriving in 3 minutes when the mean is 3.6

$$P(X) = \frac{e^{-3.6} 3.6^4}{4!} = .1912$$

Poisson Distribution Characteristics

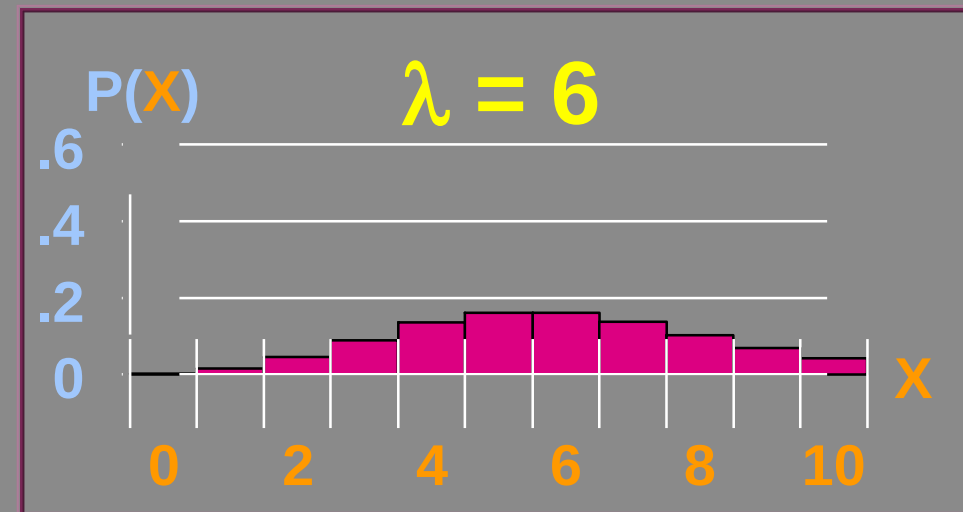
Mean

$$\begin{aligned}\mu &= E(X) = \lambda \\ &= \sum_{i=1}^N X_i P(X_i)\end{aligned}$$



Standard Deviation

$$\sigma = \sqrt{\lambda}$$

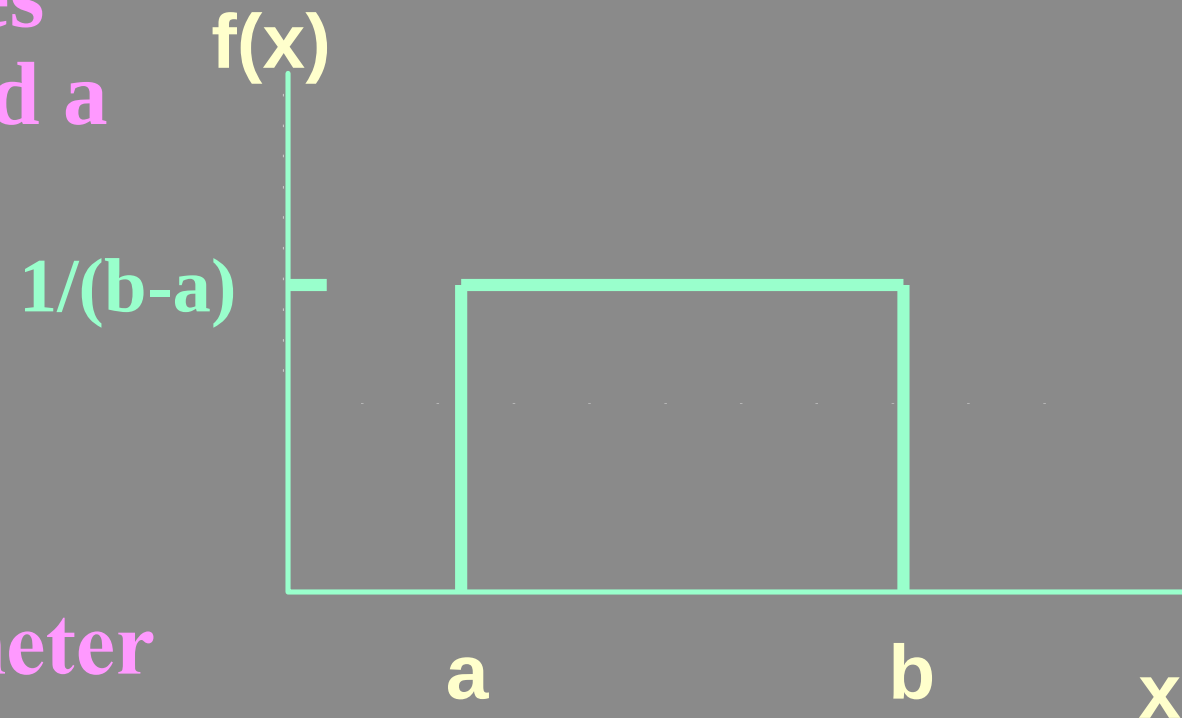


Continuous Probability Distributions

- Uniform
- Triangular
- Normal
- Exponential

The Uniform Distribution

- Equally Likely chances of occurrences of RV values between a maximum and a minimum
- Mean = $(b+a)/2$
- Variance = $(b-a)^2/12$
- 'a' is a location parameter
- 'b-a' is a scale parameter
- no shape parameter



The Uniform Distribution

Probability Density Function

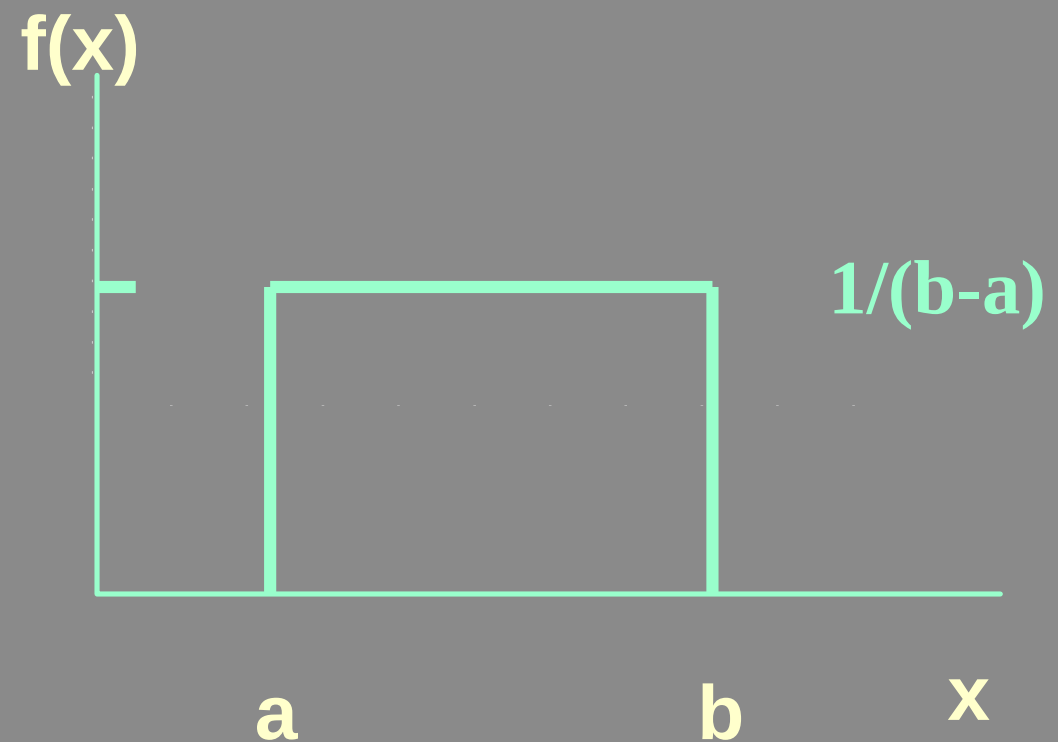
$$f(x) = \frac{1}{b-a} \quad \text{if } a \leq x \leq b$$

Distribution Function

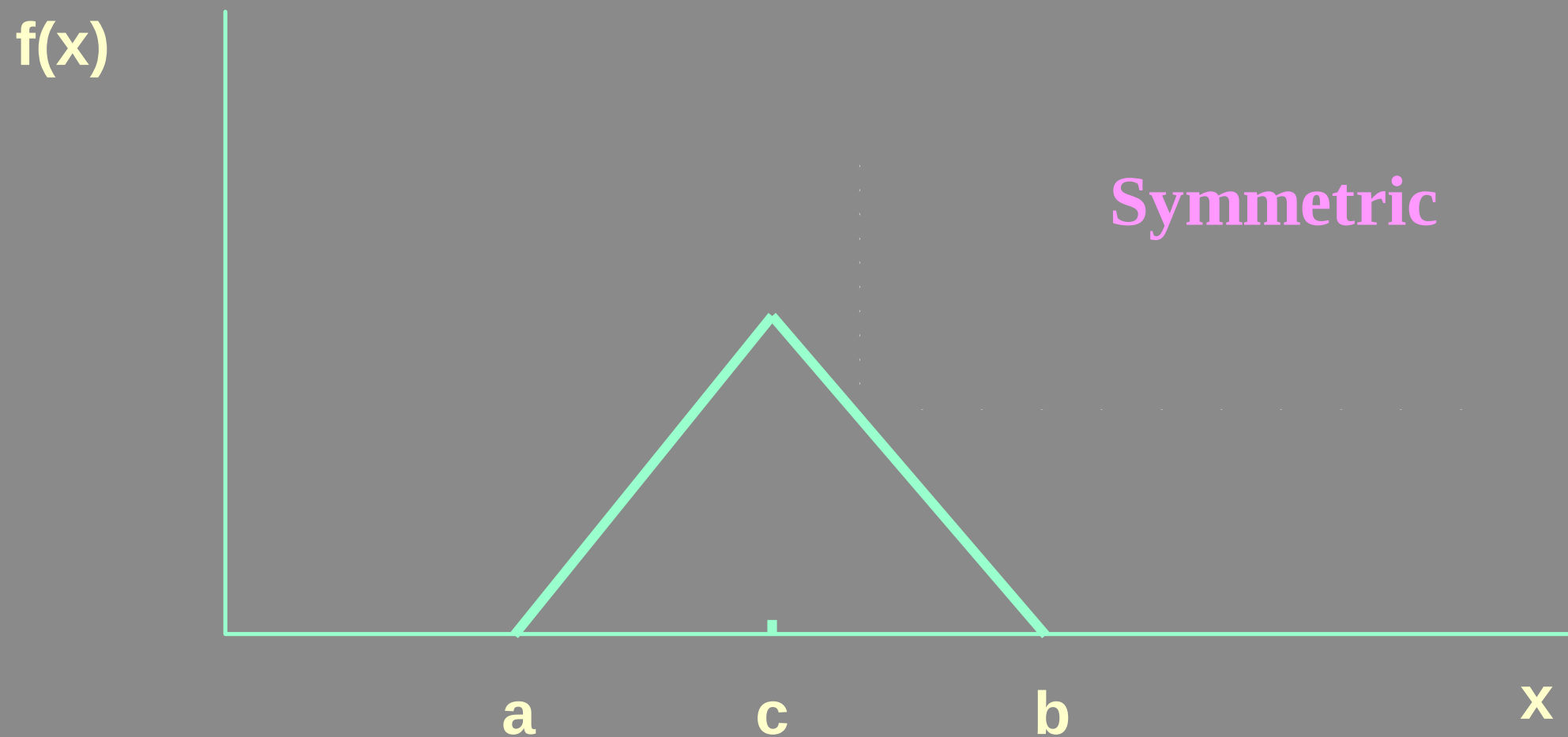
$$F(x) = 0 \quad \text{if } x < a$$

$$F(x) = \frac{x-a}{b-a} \quad \text{if } a \leq x \leq b$$

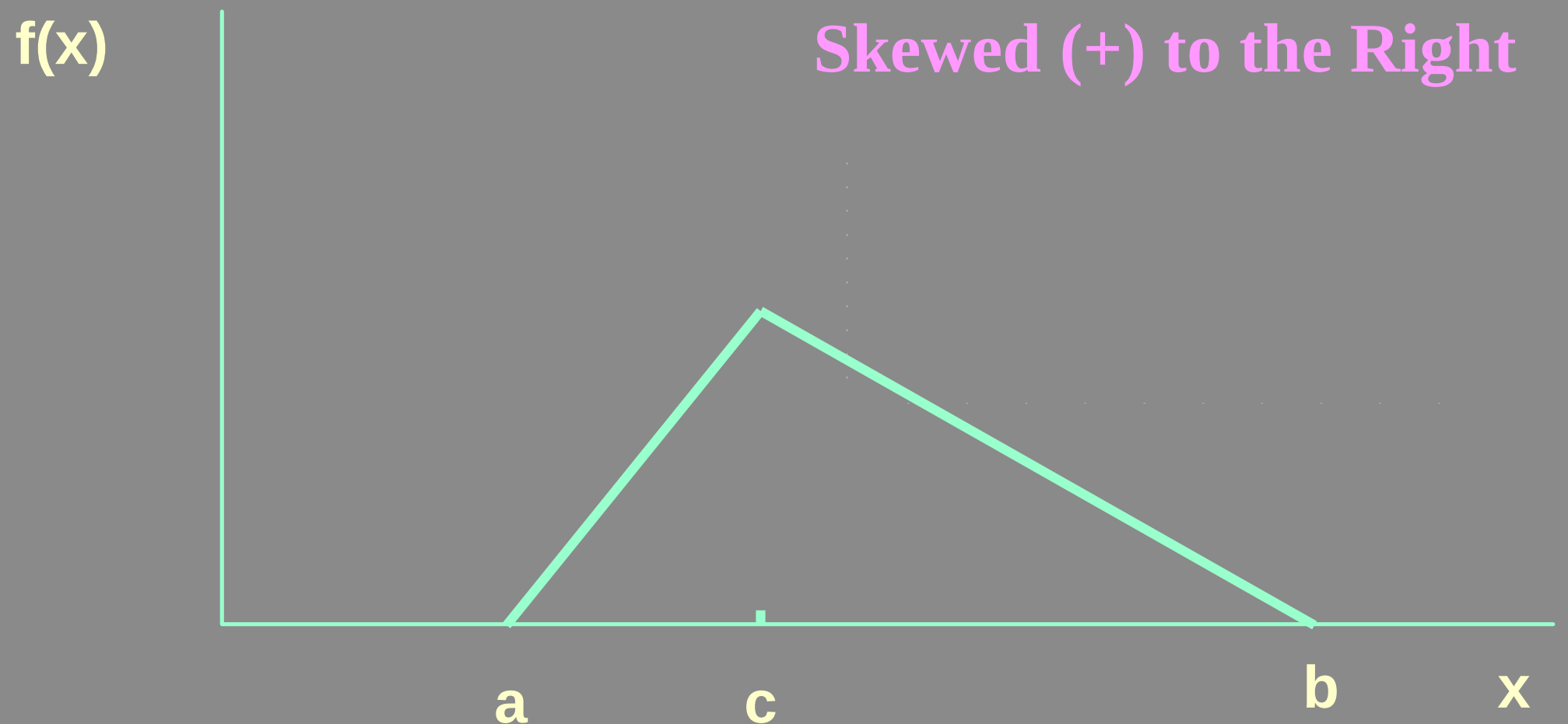
$$F(x) = 1 \quad \text{if } b < x$$



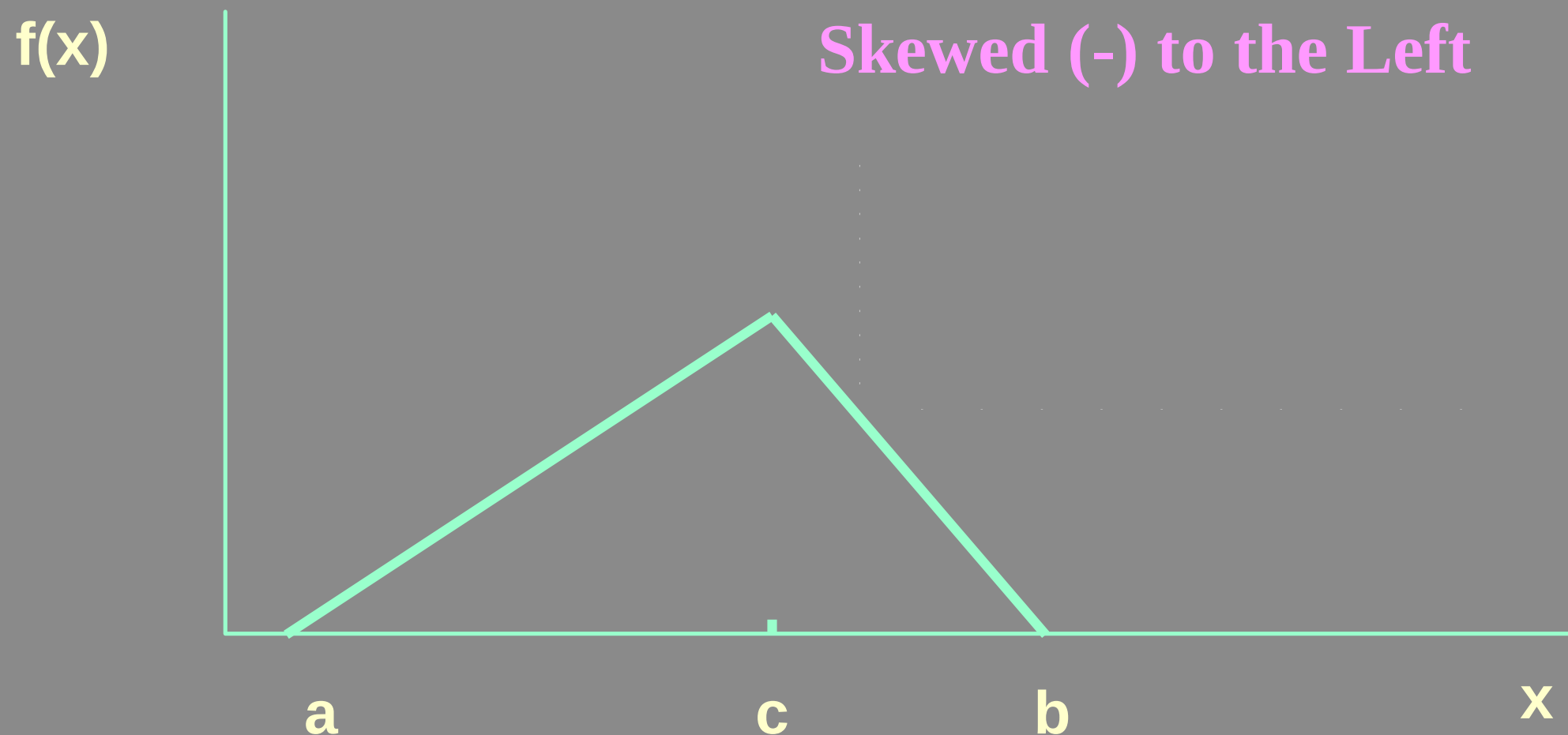
The Triangular Distribution



The Triangular Distribution



The Triangular Distribution



The Triangular Distribution

- Probability Distribution Function

$$f(x) = \frac{2(x-a)}{(b-a)(c-a)} \quad \text{if } a \leq x \leq c$$

$$f(x) = \frac{2(b-x)}{(b-a)(b-c)} \quad \text{if } c < x \leq b$$

$$f(x) = 0 \quad \text{otherwise}$$

The Triangular Distribution

- Distribution Function

$$F(x) = 0 \quad \text{if } x < a$$

$$F(x) = \frac{(x-a)^2}{(b-a)(c-a)} \quad \text{if } a \leq x \leq c$$

$$F(x) = 1 - \frac{(b-x)^2}{(b-a)(b-c)} \quad \text{if } c < x \leq b$$

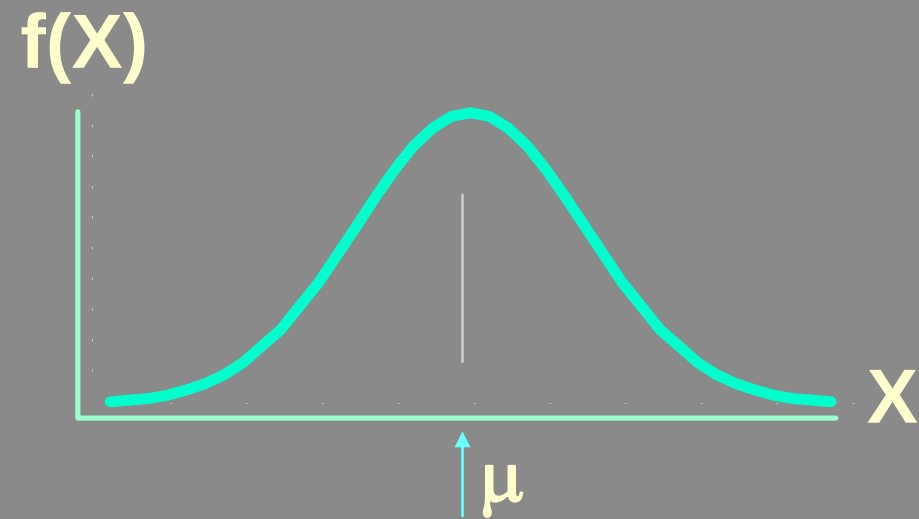
$$F(x) = 1 \quad \text{if } x > b$$

The Triangular Distribution

- **Parameters:** Minimum a, maximum b, most likely c
- Symmetric or skewed in either direction
- a location parameter
- (b-a) scale parameter
- c shape parameter
- Mean = $(a+b+c) / 3$
- Variance = $(a^2 + b^2 + c^2 - ab - ac - bc) / 18$
- Used as rough approximation of other distributions

The Normal Distribution

- ‘Bell Shaped’
- Symmetrical
- Mean, Median and Mode are Equal
- ‘Middle Spread’
Equals 1.33σ
- Random Variable has Infinite Range



Mean
Median
Mode

The Mathematical Model

$$f(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X-\mu)^2}$$

$f(X)$ = frequency of random variable X

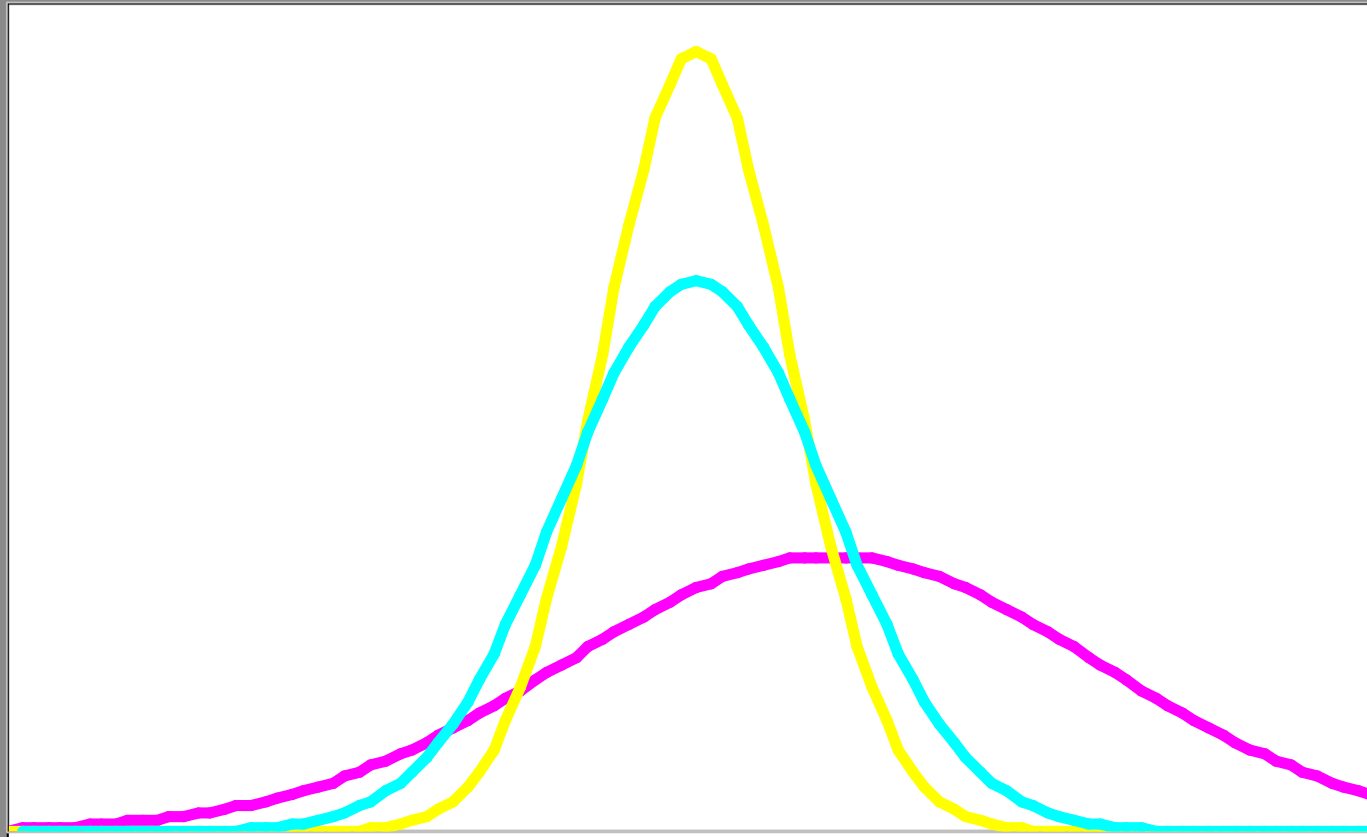
π = 3.14159; $e = 2.71828$

σ = population standard deviation

X = value of random variable ($-\infty < X < \infty$)

μ = population mean

Many Normal Distributions



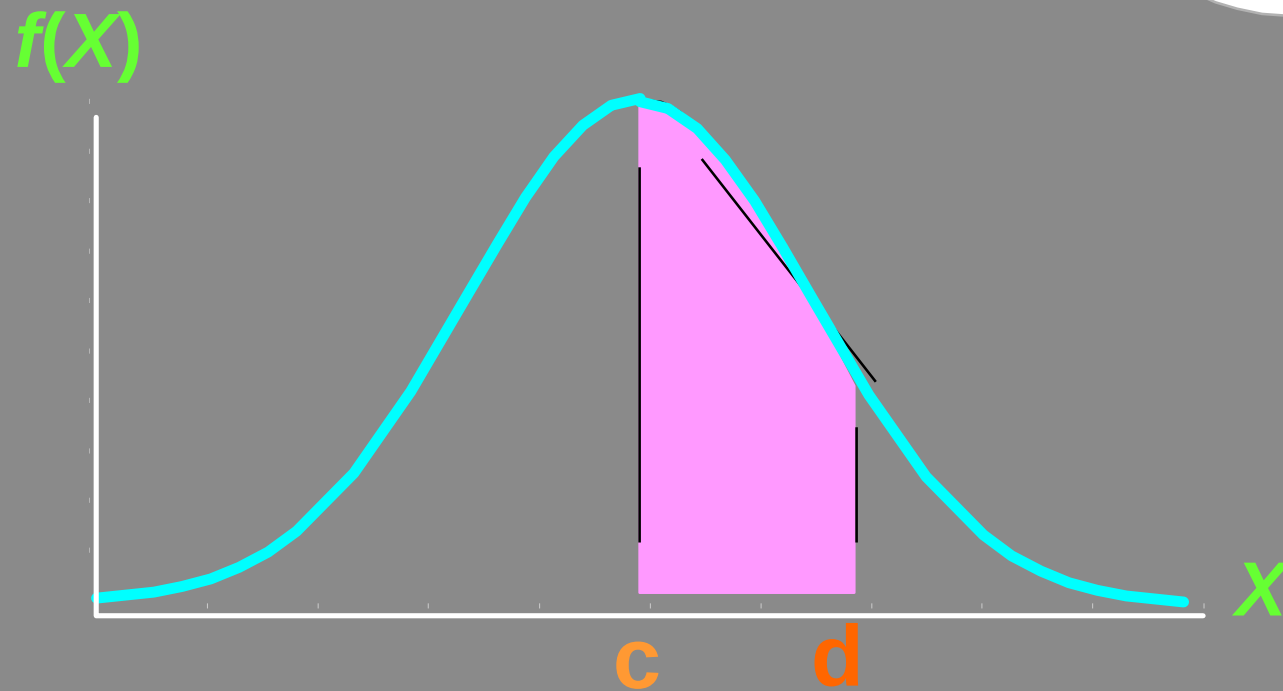
There are
an Infinite
Number

Varying the Parameters σ and μ , we obtain
Different Normal Distributions.

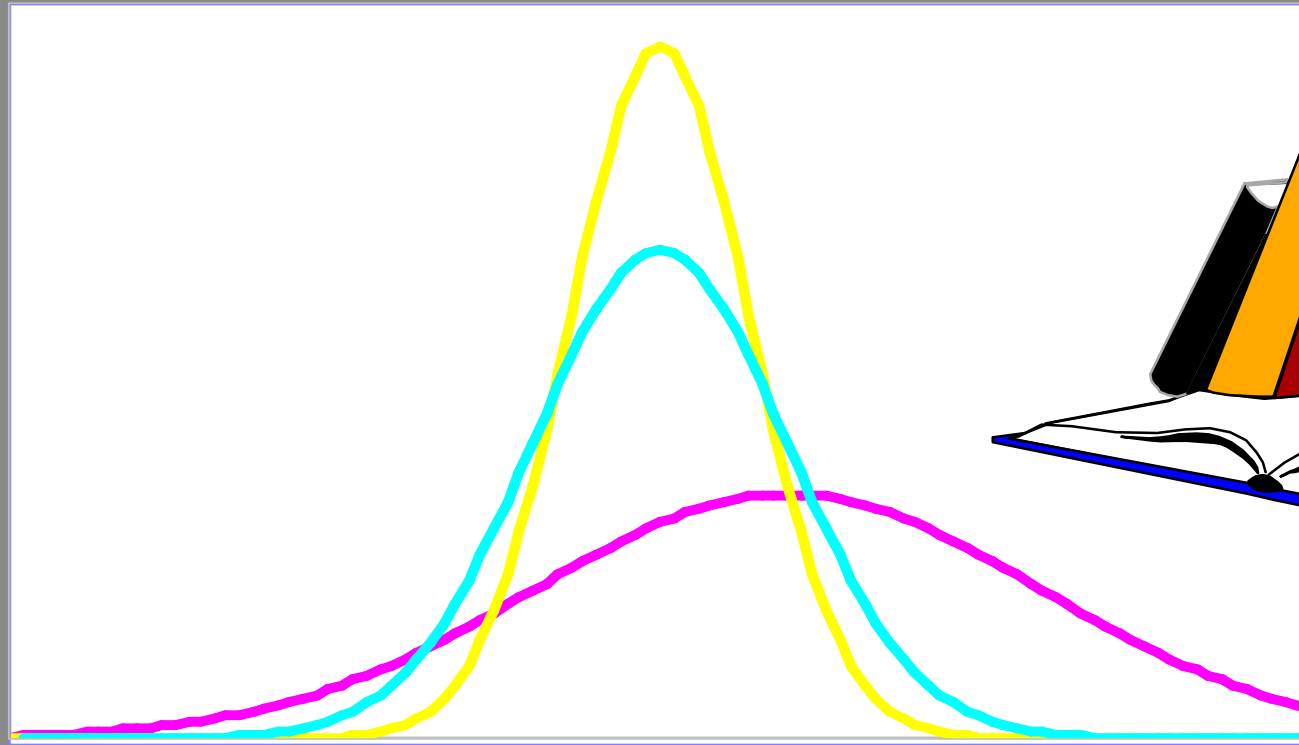
Normal Distribution: Finding Probabilities

Probability is the
area under the
curve

$$P(c \leq X \leq d) = ?$$



Which Table?



Each distribution
has its own table?

Infinitely Many Normal Distributions Means
Infinitely Many Tables to Look Up!

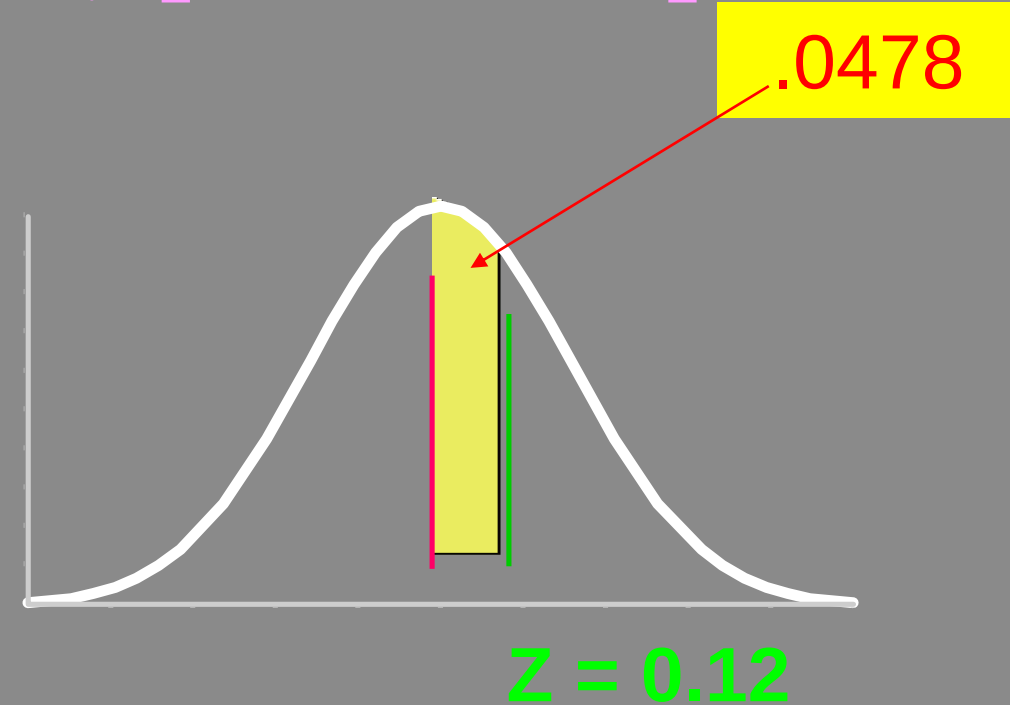
Solution (I): The Standardized Normal Distribution

Standardized Normal Distribution
Table (Portion)

$$\mu_z = 0 \quad \text{and} \quad \sigma_z = 1$$

Z	.00	.01	.02
0.0	.0000	.0040	.0080
0.1	.0398	.0438	.0478
0.2	.0793	.0832	.0871
0.3	.0179	.0217	.0255

Probabilities



Only One Table is Needed

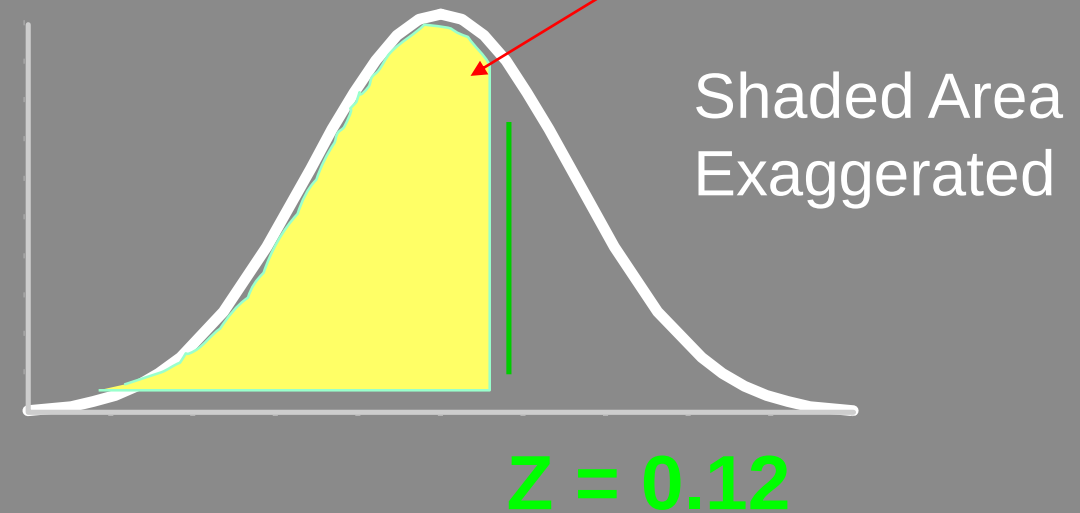
Solution (II): The Cumulative Standardized Normal Distribution

Cumulative Standardized Normal Distribution Table (Portion)

Z	.00	.01	.02
0.0	.50000	.50400	.50800
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.5179	.5217	.5255

$\mu_Z = 0$ and $\sigma_Z = 1$

.5478



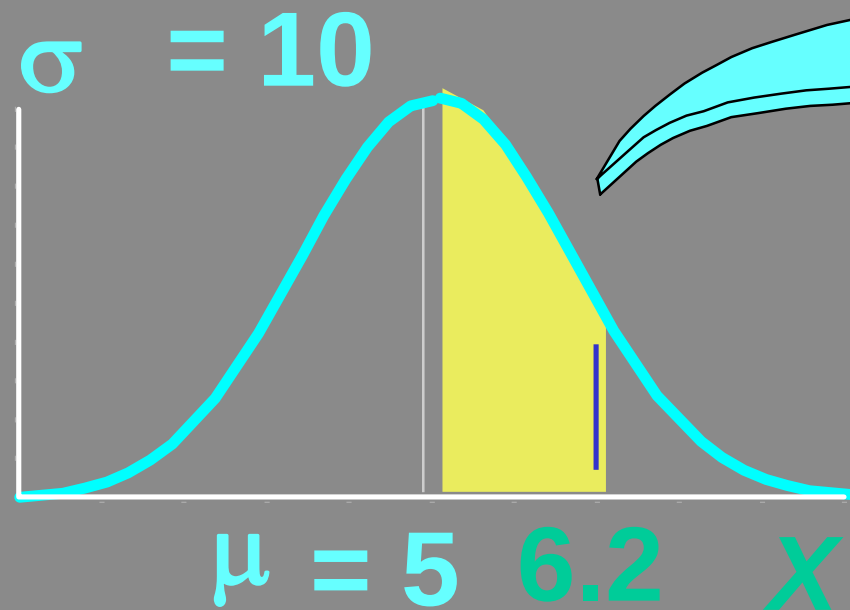
Probabilities

Only One Table is Needed

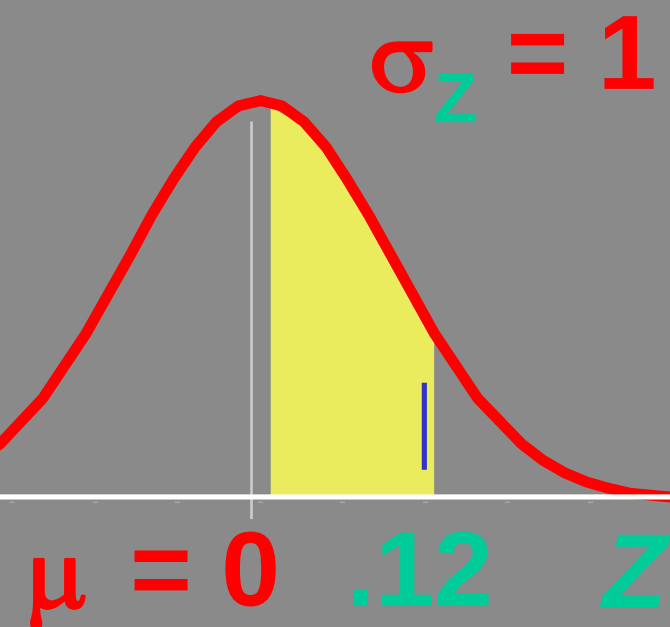
Standardizing Example

$$Z = \frac{X - \mu}{\sigma} = \frac{6.2 - 5}{10} = 0.12$$

Normal
Distribution



Standardized
Normal Distribution



Example:

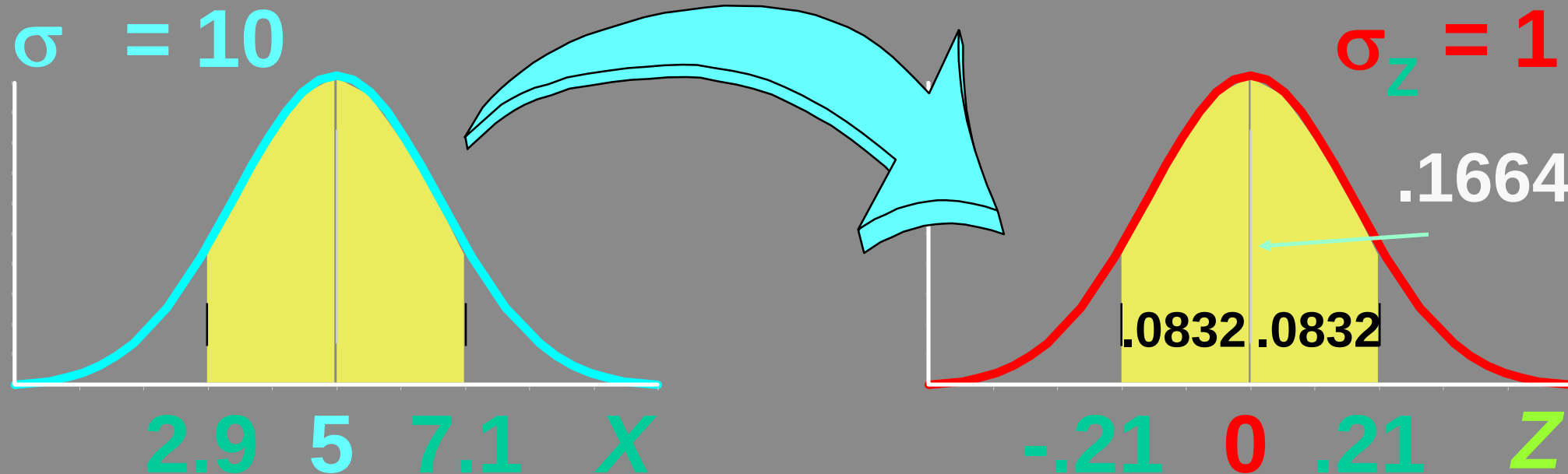
$$P(2.9 < X < 7.1) = .1664$$

$$z = \frac{x - \mu}{\sigma} = \frac{2.9 - 5}{10} = -.21$$

Normal
Distribution

$$z = \frac{x - \mu}{\sigma} = \frac{7.1 - 5}{10} = .21$$

Standardized
Normal Distribution

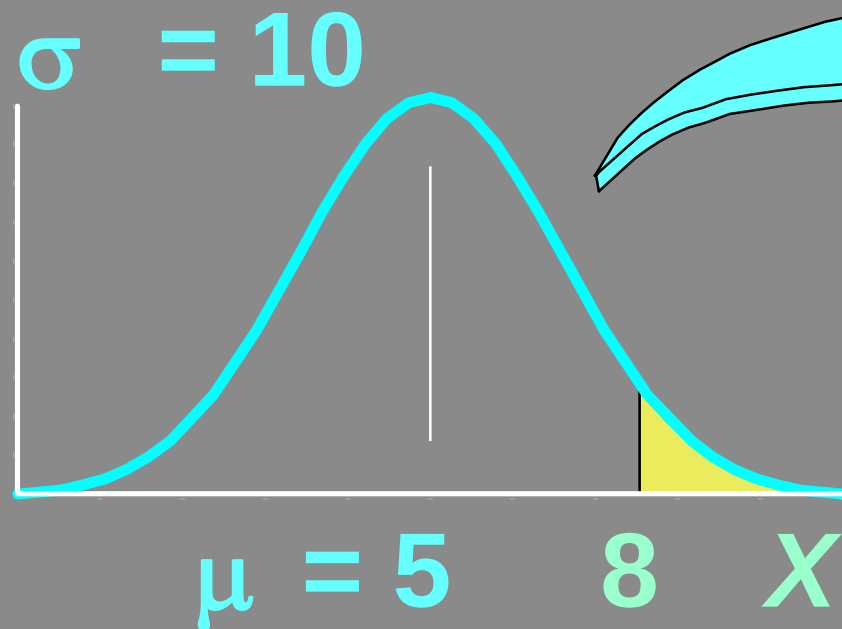


Shaded Area Exaggerated

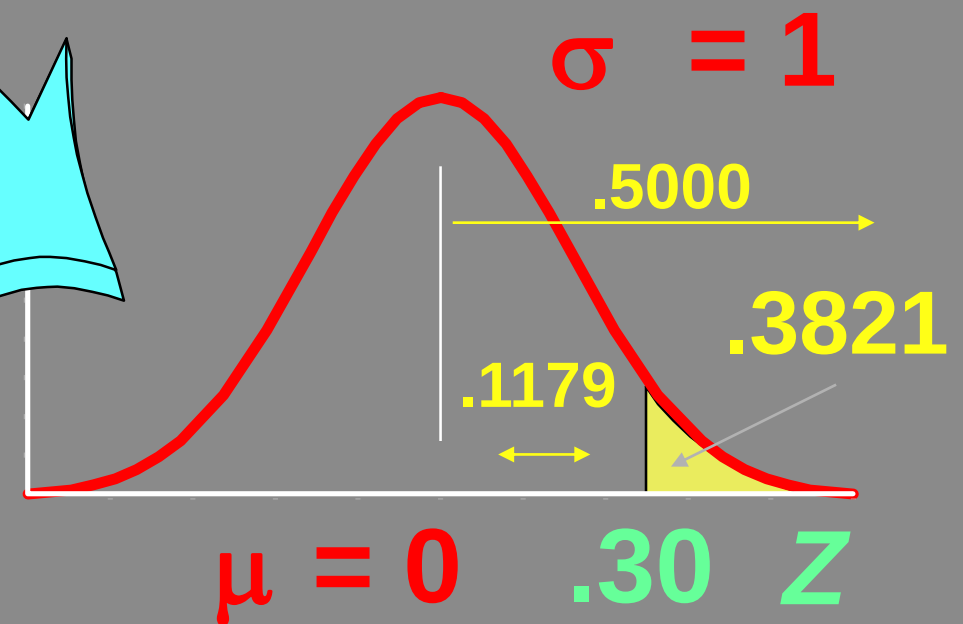
Example: $P(X \geq 8) = .3821$

$$Z = \frac{X - \mu}{\sigma} = \frac{8 - 5}{10} = .30$$

Normal
Distribution



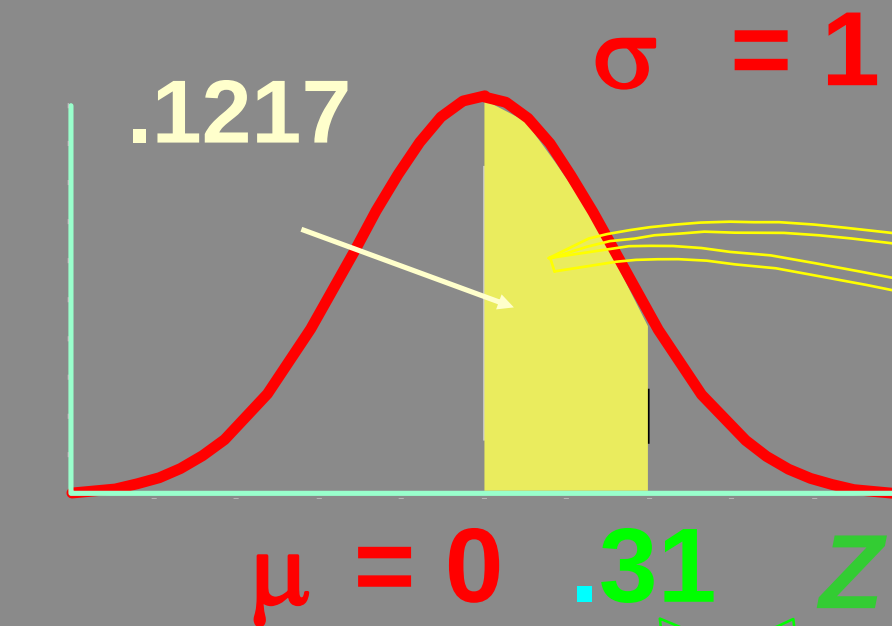
Standardized
Normal Distribution



Shaded Area Exaggerated

Finding Z Values for Known Probabilities

What Is Z Given
Probability = 0.1217?

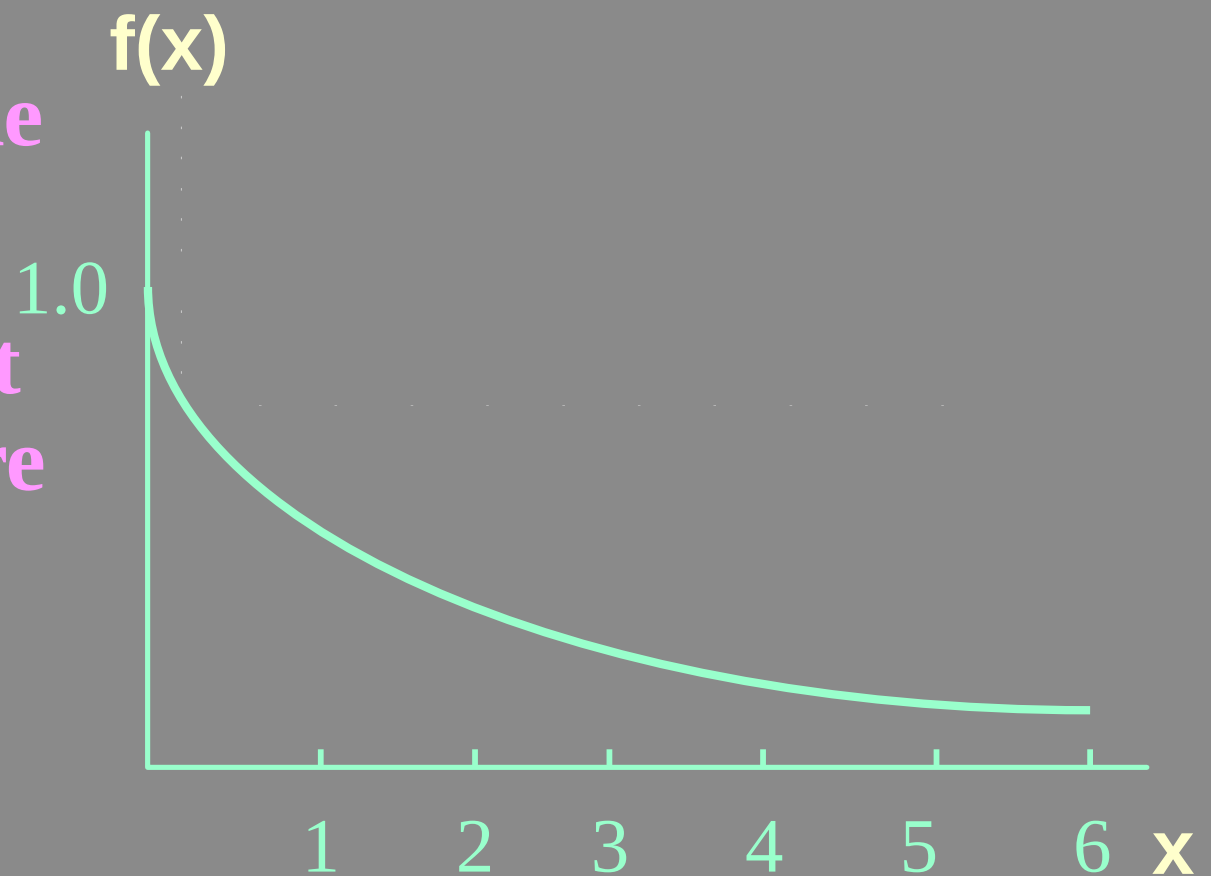


Standardized Normal
Probability Table (Portion)

Z	.00	.01	0.2
0.0	.0000	.0040	.0080
0.1	.0398	.0438	.0478
0.2	.0793	.0832	.0871
0.3	.1179	.1217	.1255

Exponential Distribution

- Models time between customer arrivals to a service system and the time to failure of machines
- *Memoryless* : the current time has no effect on future outcomes
- No shape or location parameter
- λ is the scale parameter



Exponential Distribution

Density : $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$

and

Distribution : $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$

$f(x)$ = frequency of random variable x

e = 2.71828

$1/\lambda$ = Mean of the exponential distribution

$(1/\lambda)^2$ = Variance of the exponential distribution

Sampling Distributions and Sampling Error

- How good is the sample estimate ?
- Multiple samples – Each with a mean
- Sampling distribution of the means
- As the sample size increases – Variance decreases

$$\text{Standard error of the mean} = \frac{\sigma}{\sqrt{n}}$$

- Sampling Distribution of the mean ?
 - For large n: Normal regardless of the population
 - Central Limit Theorem

CLT

distribution of sample means approximates a normal distribution as the sample size becomes larger, assuming that all samples are identical in size, and regardless of the population distribution shape

CLT

Given a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population.

All samples will follow an approximate normal distribution pattern, with all variances being approximately equal to the variance of the population, divided by each sample's size