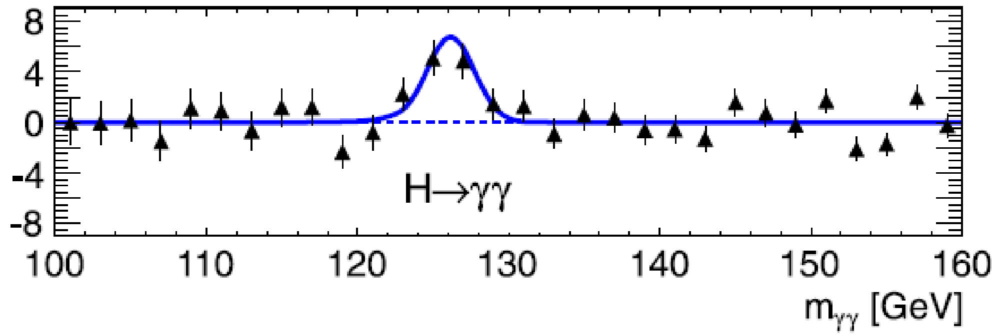
EP122
Measurement Techniques and Calibration

Topic 3

Introductory Probability



http://www.gantep.edu.tr/~bingul/ep122

Department of Engineering Physics

University of Gaziantep

Feb 2016

Content

PART I

BASIC SET THEORY

Set Theory

In mathematics, a well defined collection of objects is called the **set**.

Examples:

$A = \{1,2,3,4\}$	=> finite set
$M = \{\text{apple, banana, orange}\}$	=> finite set
$R = \{x \mid x \text{ is a river on Earth}\}$	=> finite set
$N = \{0,1, 2, 3, 4, ...\}$	=> infinite set
$P = \{2, 4, 8, ...\}$	=> infinite set
$K = \{x \mid 2<x<5, x \text{ is a real}\}$	=> infinite set

Notation:

$p \in A$ — *p* is elements of *A*

$A \subset B$ — *A* is subset of *B*

U — Universal set

ϕ — Empty set

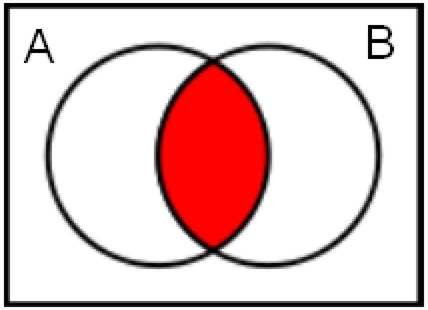
For any set A

$$\phi \subset A \subset U$$

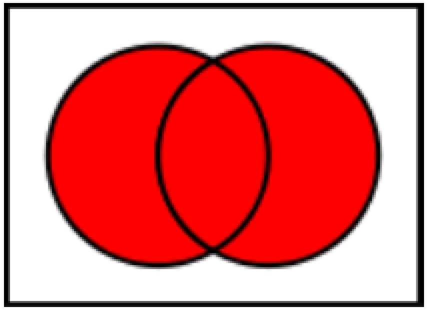
Set Operations:

Venn Diagram

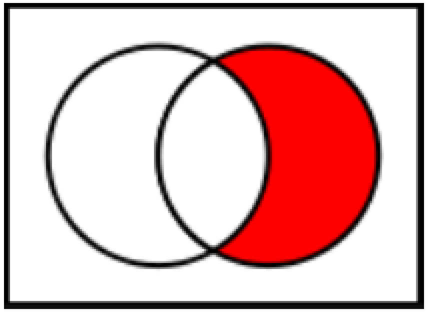
Intersection $A \cap B$



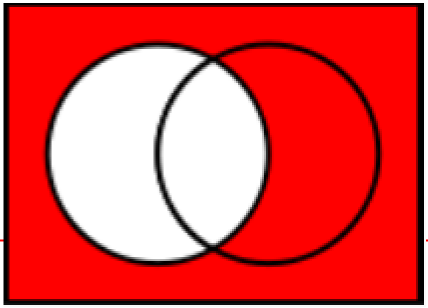
Union $A \cup B$



Difference $B - A$



Complement A^C



Example

Let

A ={1, 2, 3, 4}, B = {3, 4, 5, 6} and U={1, 2, 3, 4, 5, 6, 7,8 …}

A U B = {1, 2, 3, 4, 5, 6}

A n B = {3, 4}

A \ B = {1, 2}

A^c = {5, 6, 7, 8, ...}

PART II

BASIC PROBABILITY

Probability (=Olasılık)

Historically, the probability theory began with study of games of chance, such as roulette and cards.

The probability is the study of random or non-deterministic experiments

If a coin is tossed in the air, then

it is certain that the coin will come down

but

it is not certain that a head will appear.

Relative Frequency (=Göreli Sıklık)

Suppose we repeat an experiment of tossing a die.

Let

s be the number of times a “six” appears

n be the number of tosses

Then the ratio s/n becomes <u>stable</u> in the long run:

$$f = \frac{s}{n}$$

f approaches a limit as n -> ∞

This stability is the basis of probability theory!

Example

Here is the result obtained from a computer simulation for tossing of a coin and observing frequency of head!

HEAD

TAIL

n	s	f = s/n
10	4	0.4000000
100	41	0.4100000
1,000	476	0.4760000
10,000	5059	0.5059000
100,000	49942	0.4994200
1,000,000	500351	0.5003510
10,000,000	4998906	0.4998906
100,000,000	50006417	0.5000641
1,000,000,000	500000839	0.5000084

The result approaches a limit as $n \to \infty$

Example

Here is the result obtained from a computer simulation for tossing of a die and observing frequency of six!

n	s	f = s/n
10	3	0.3000000
100	19	0.1900000
1,000	186	0.1860000
10,000	1659	0.1659000
100,000	16748	0.1674800
1,000,000	166705	0.1667050
10,000,000	1667210	0.1667210
100,000,000	16666290	0.1666629
1,000,000,000	166666653	0.1666666

The result approaches a limit as $n \to \infty$

Probability Theory

The probability *p* of an event *A* is defined as follows:

If *A* occurs in *s* ways out of a total *n* equally likely ways then

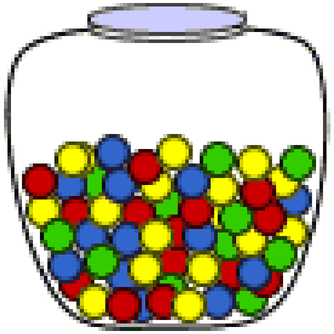
$$p = P(A) = \frac{s}{n}$$

* Tossing a coin: Head occurs 1 way out of 2 => p = 1/2
* Tossing a die: Six occurs 1 way out of 6 => p = 1/6
* Tossing a die: Even number occurs 3 ways out of 6 => p = 3/6

Probability is the measure of how likely an event is

Example

A glass jar contains 6 red, 5 green, 8 blue and 3 yellow marbles. If a single marble is chosen at random from the jar, what is the probability of choosing a red marble? a green marble? a blue marble? a yellow marble?



Outcomes: The possible outcomes of this experiment are red, green, blue and yellow.

Probabilities:

$$P(\text{red}) = \frac{\text{\# of ways to choose red}}{\text{total \# of marbles}} = \frac{6}{22} = \frac{3}{11}$$

$$P(\text{green}) = \frac{\text{\# of ways to choose green}}{\text{total \# of marbles}} = \frac{5}{22}$$

$$P(\text{blue}) = \frac{\text{\# of ways to choose blue}}{\text{total \# of marbles}} = \frac{8}{22} = \frac{4}{11}$$

$$P(\text{yellow}) = \frac{\text{\# of ways to choose yellow}}{\text{total \# of marbles}} = \frac{3}{22}$$

Sample Space (=Örneklem Uzayı)

The set *S* of all possible outcomes of some given experiment is called "sample space"

*** Tossing a coin:** S = {H, T}

*** Tossing two coins:** S = {HH, HT, TH, TT}

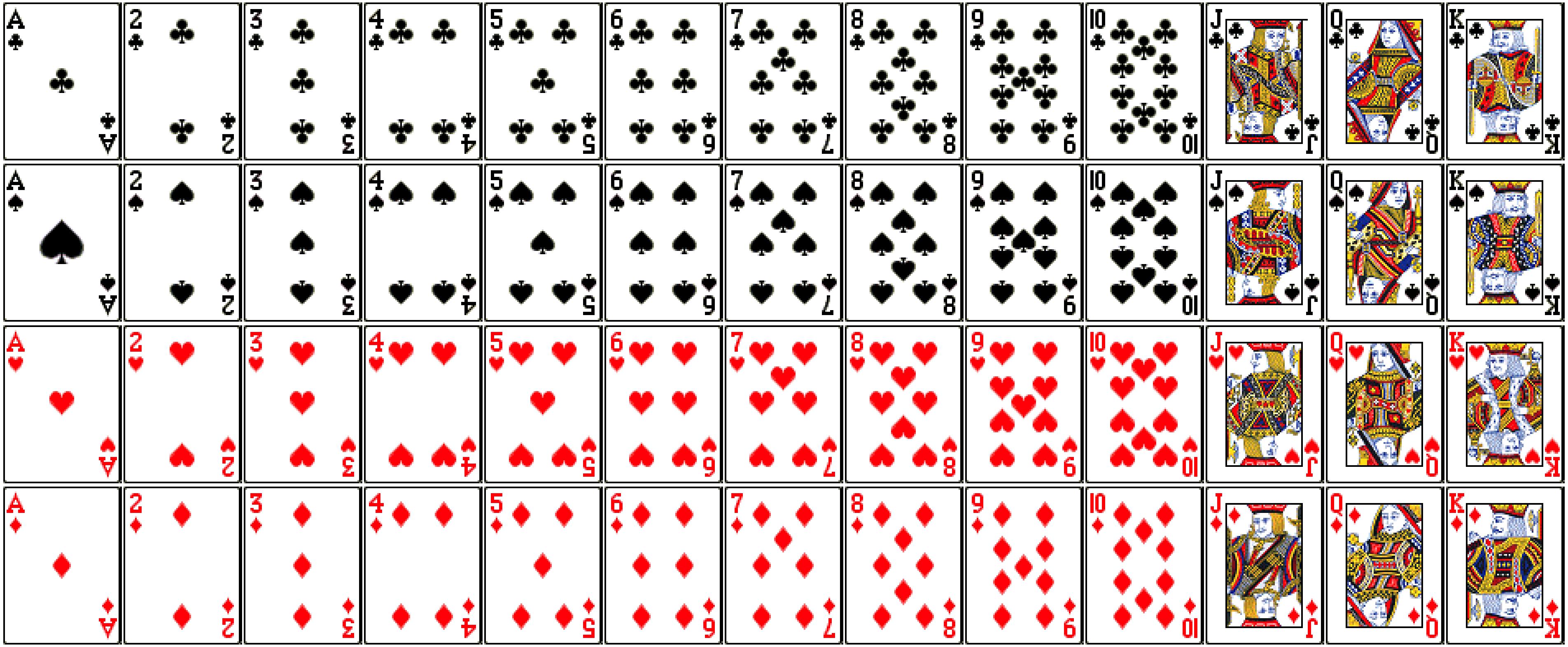
*** Tossing a die**:

S = {1, 2, 3, 4, 5, 6}

*** For two dice the outcomes are**

S = {11,12,13,14,15,16,
21,22,23,24,25,26,
31,32,33,34,35,36,
41,42,43,44,45,46,
51,52,53,54,55,56,
61,62,63,64,65,66}

Selecting a card randomly from a shuffle pack of playnig cards, the possible outcomes are:



Axioms of Probability

Let *S* be sample space and *A* and *B* are two events.

A1. $0 \leq P(A) \leq 1$

A2. $P(S) = 1$

A3. If *A* and *B* are mutually exclusive events (ayrık olaylar)

$$P(A \cup B) = P(A) + P(B)$$

Theorems of Probability

T1. $P(\phi) = 0$ (probability of impossible event is zero)

T2. If A^c is the complement of *A*, then $P(A^c) = 1 - P(A)$

T3. If *A* and *B* are <u>any</u> two events:

$$P(A - B) = P(A) - P(A \cap B)$$

T4. If *A* and *B* are <u>any</u> two events:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) - P(A \cap B \cap C)$$

Example

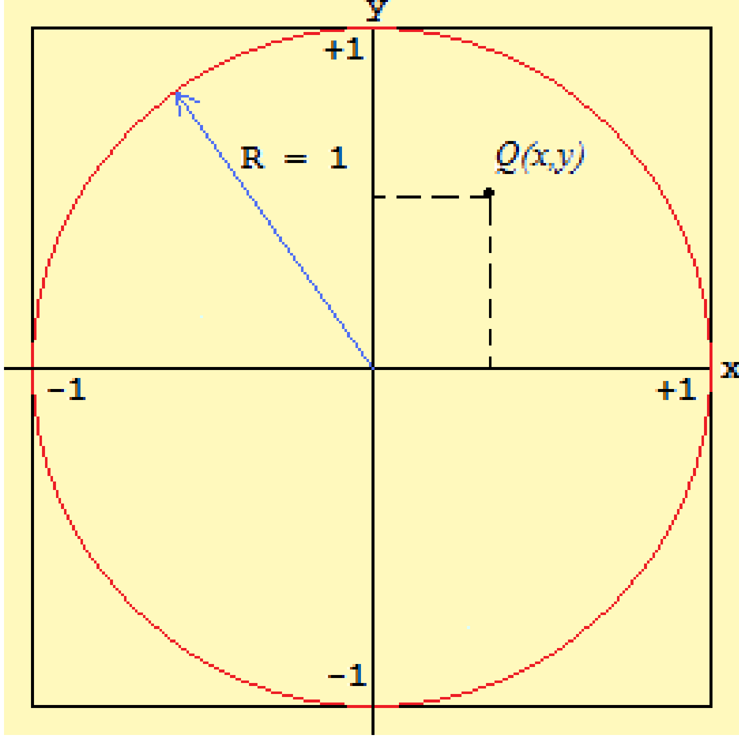
(a) What is the sample space for choosing 1 letter at random from the word DIVIDE?

(b) What is the probability of selecting the letter V from the word DIVIDE?

Example

A point Q is selected randomly in a square whose side is 2 cm. A circle is drawn tangent to the edges of the square.

Find the probability of the point being inside the circle.



PART III

RANDOM VARIABLES

Random Variable (=Raslantı Değişkeni)

- A Random Variable (RV) is a set of values that occur randomly and have associated probabilities.
- RV can be discrete (kesikli) or continues (sürekli).
- Probability mass function (pmf) describes the distribution of the discrete probabilities
- Probability distribution function (pdf) describes the distribution of the continues probabilities.

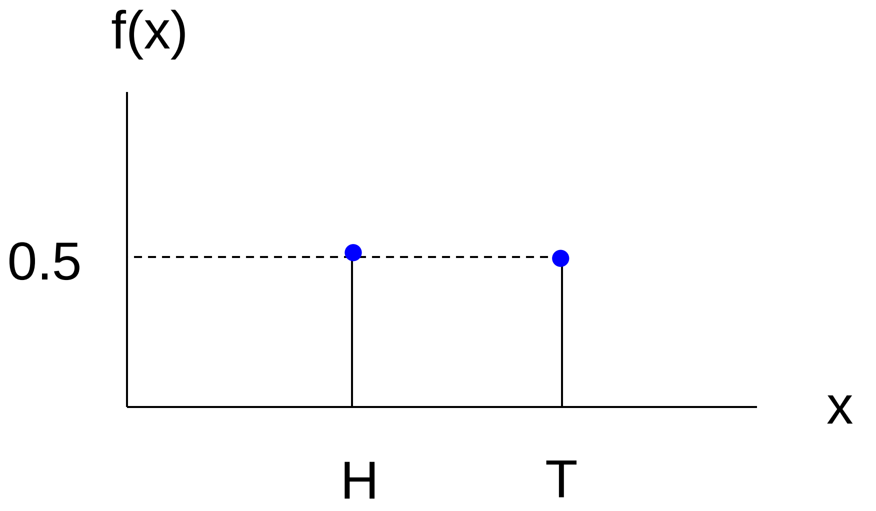
Discrete RV Examples:

pmf

Tossing a coin:

X = {H,T}

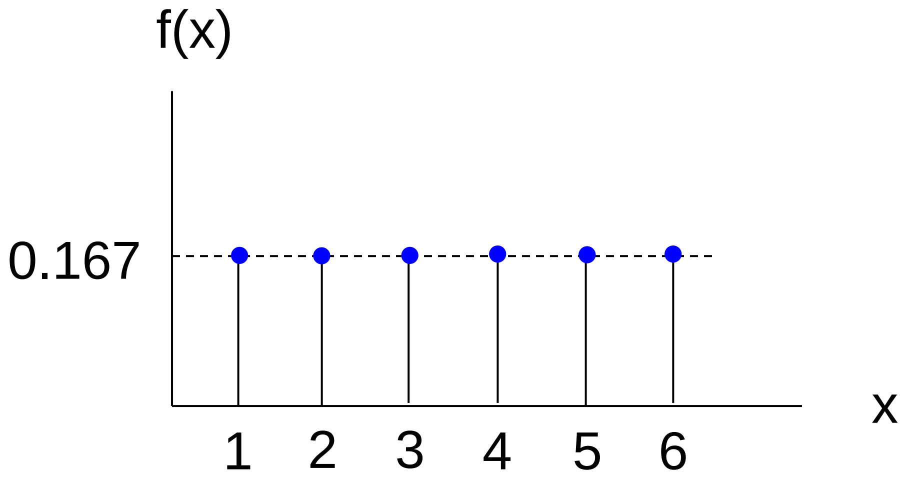
f(x) = {1/2, 1/2}



Tossing a die:

X = {1,2,3,4,5,6}

f(x) = {1/6,1/6,1/6,1/6, 1/6, 1/6}

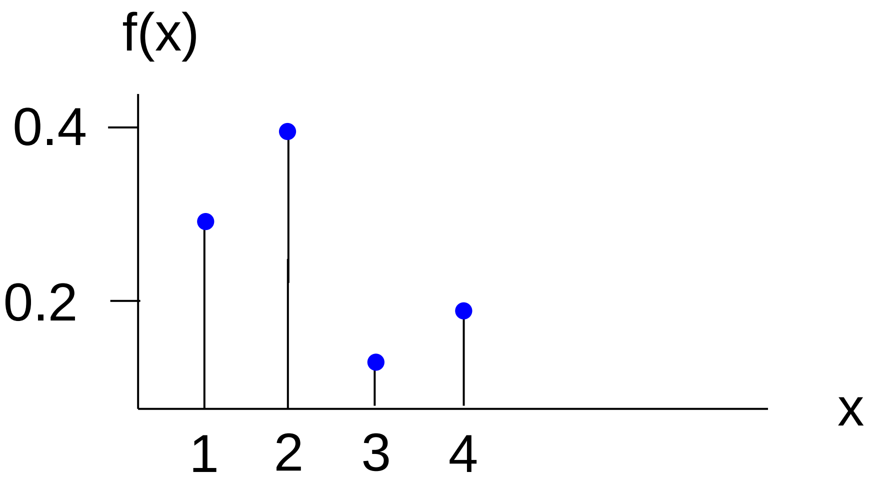


Discrete RV Examples:

X = {1,2,3,4}

f(x) = {0.3, 0.4, 0.1, 0.2}

pmf



Continues RV Examples:

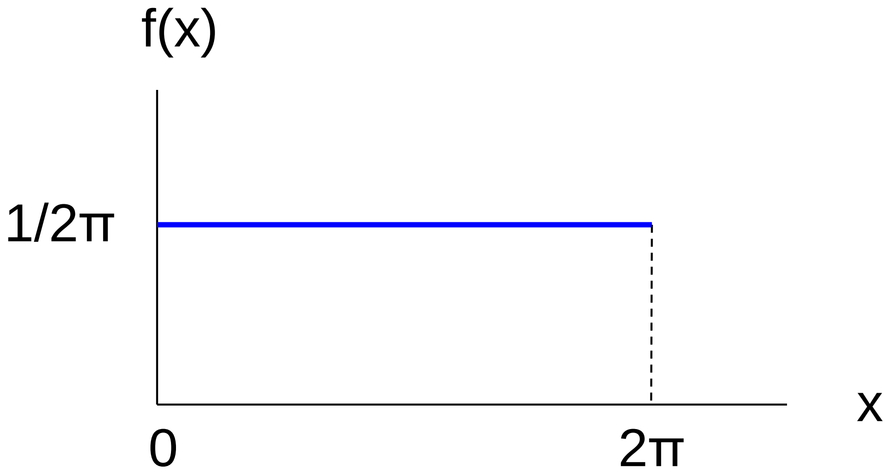
pdf

Random angle

in the range [0, 2π]

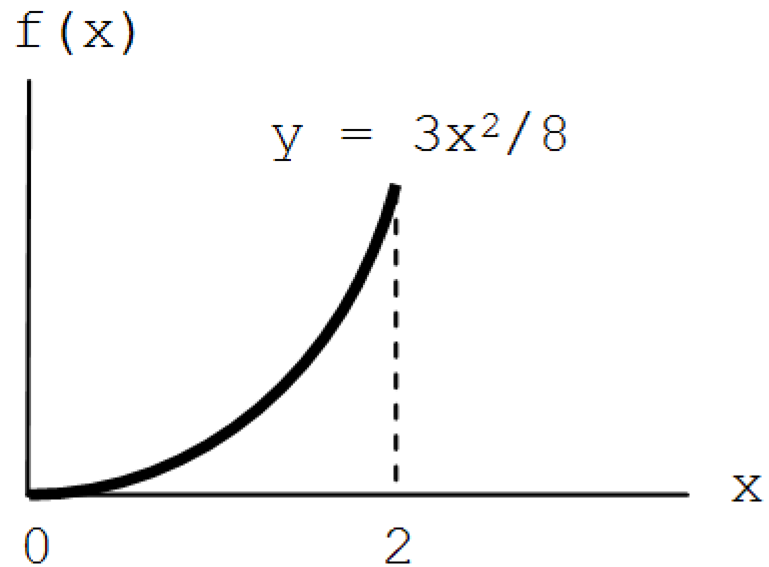
X = [0, 2π]

f(x) = 1/2π



X = [0, 2]

$f(x) = 3x^2/8$



Properties of Random Variables

Discreate RV

$$f(x_i) \geq 0$$

$$\sum_i f(x_i) = 1$$

$$\sum_{i=a}^{b} f(x_i) = P(a \leq x \leq b)$$

Continues RV

$$f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

$$\int_a^b f(x)dx = P(a \leq x \leq b)$$

Expectation Values

Expectation Value or Mean Value of RV is denoted by:

$$E\,[\,X\,] \qquad \text{or} \qquad < X > \qquad \text{or} \qquad \bar{x}$$

and defined by:

Discrete RV: $$E[X] = \sum_i x_i \; f(x_i)$$

Continues RV: $$E[X] = \int_{-\infty}^{+\infty} x \; f(x) dx$$

Mean of the Squares

It is defined as:

Discrete RV: $$E[X^2] = \sum_i x_i^2 \ f(x_i)$$

Continues RV: $$E[X^2] = \int_{-\infty}^{+\infty} x^2 \ f(x)dx$$

Note that (**RMS**: Root Mean Square)

$$RMS = \sqrt{E[X^2]}$$

Variance

Variance is defined as:

Discrete RV: $$\sigma^2 = \sum_i (x_i - \bar{x})^2 \; f(x_i)$$

Continues RV: $$\sigma^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 \; f(x) dx$$

One can also prove that

$$\begin{aligned} \sigma^2 &= E[X^2] - (E[X])^2 \\ &= \langle X^2 \rangle - \langle X \rangle^2 \end{aligned}$$

Standard Deviation

Square root of variance is called **standard deviation**:

$$\sigma = \sqrt{\sigma^2}$$

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

$$= \sqrt{<X^2> - <X>^2}$$

There are some applications of this equation in Quantum Mechanics

Example

$$X = \{1, 2, 3, 4, 5, 6\}$$

$$f(x) = \{1/6, 1/6, 1/6, 1/6, 1/6, 1/6\}$$

Find

(a) Expectation value of X, (b) RMS and (c) standard deviation

Example

Suppose a variable X can take the values 1, 2, 3, or 4.

The probabilities associated with each outcome are described by the following table:

Outcome	**1**	**2**	**3**	**4**
Probability	**0.1**	**0.3**	**0.4**	**0.2**

Find

(a) $P(X = 2 \text{ or } X = 3)$

(b) $1 - P(X = 1)$

Example

The following table gives the probability distribution of, X, the number of telephones in a randomly selected home in a certain community.

x:	**0**	**1**	**2**	**3**	**4**
f(x):	**0.021**	**0.412**	**0.283**	**0.188**	**0.096**

One home is selected randomly.

The probability that it will have:

(a) no telephone is 0.021

(b) fewer than two telephones is 0.433

(c) at least three telephones is 0.284

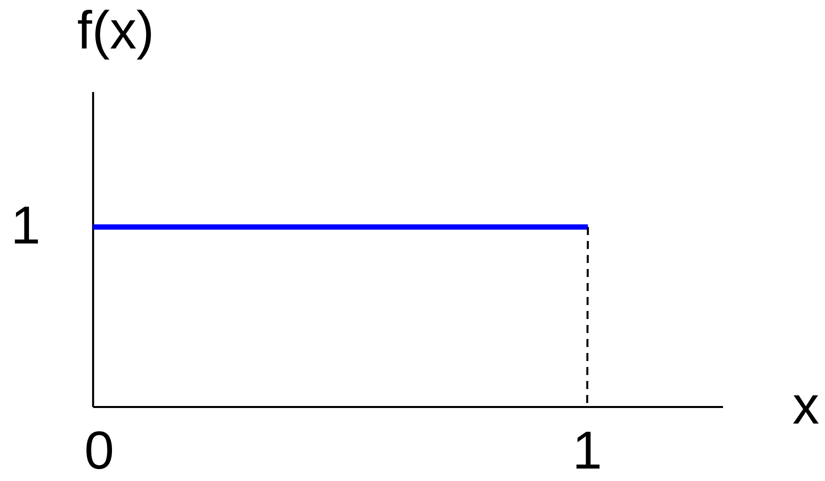
(d) one or two telephones is 0.695

Example

Consider the uniform pdf.

X = [0,1]

f(x) = 1



Find

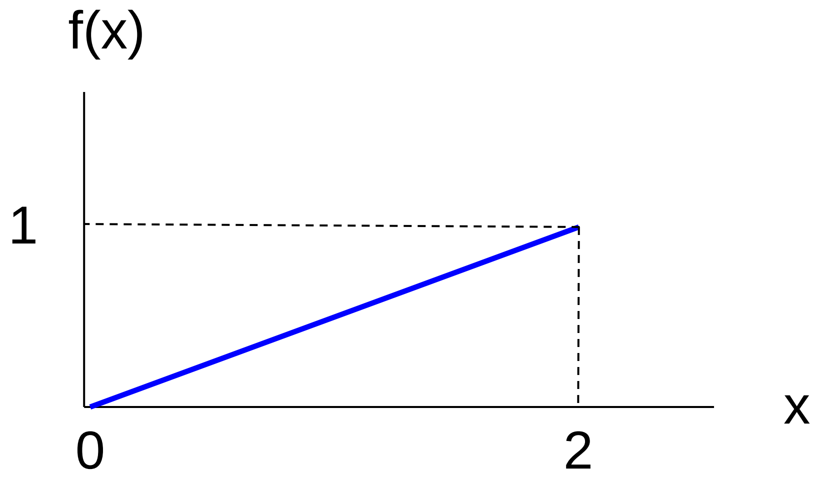
(a) mean, (b) RMS and (c) standard deviation

Example

Consider the triangular pdf.

X = [0,2]

f(x) = kx



Find

(a) the value k (b) mean and (c) standard deviation

PART IV

SPECIAL DISTRIBUTION FUNCTIONS

Binomial Distribution Function

The binomial distribution function specifies the number of times (*k*) that an event occurs in *n* independent trials where *p* is the probability of the event occurring in a single trial.

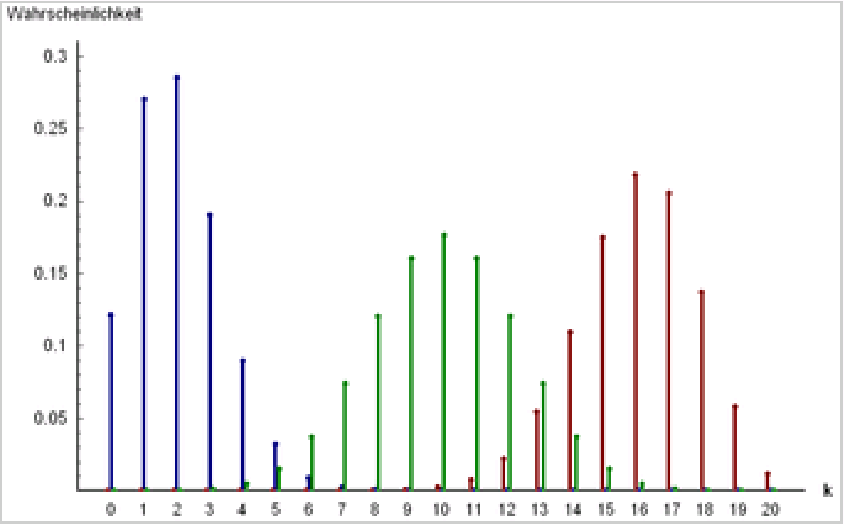
$$p_{binom}(n,k,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\text{mean}: \quad \langle k \rangle = np$$

$$\text{std.dev}: \quad \sigma = \sqrt{np(1-p)}$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$



Binomial distribution for $n = 20$
$p = 0.1$ (blue), $p = 0.5$ (green) and $p = 0.8$ (red)

Example: A coin is tossed 3 times.

If you call heads a success, draw pmf for $k = 0, 1, 2, 3$.

Example: A coin is tossed 6 times.

The probability of getting exactly four heads:

$$P = \frac{6!}{4!\,(6-4)!}\left(\frac{1}{2}\right)^{4}\left(1-\frac{1}{2}\right)^{6-4} = 0.234375$$

The probability of getting at least four heads:

$$P = \binom{6}{4}\left(\frac{1}{2}\right)^{4}\left(1-\frac{1}{2}\right)^{6-4} + \binom{6}{5}\left(\frac{1}{2}\right)^{5}\left(1-\frac{1}{2}\right)^{6-5} + \binom{6}{6}\left(\frac{1}{2}\right)^{6}\left(1-\frac{1}{2}\right)^{0}$$

$$= 0.34375$$

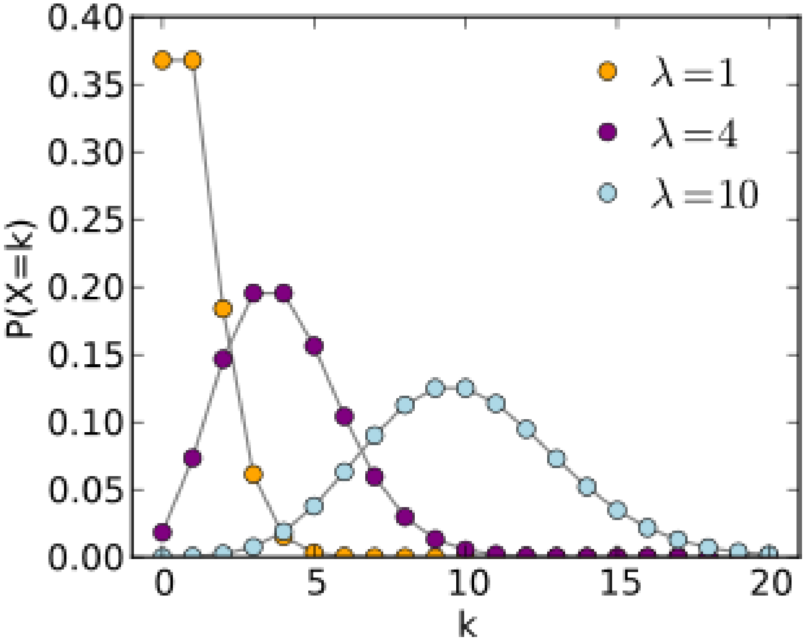
Poisson Distribution Function

In the binomial equation, if the probability *p* is so small then the distribution of events can be approximated by the Poisson distribution.

$$p_{poisson}(k,\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$\text{mean}: \quad \lambda = np$$

$$\text{std.dev}: \quad \sigma = \sqrt{\lambda}$$



$\lim_{p \to 0}$ Binomial Distribution = Poisson Distribution

Example: *Birthday problem*

Probability of one person to have birthday in any day is 1/365 = 0.00274. Calculate the probability that 4 people share a birthday in a group of 1000 people.

Mean: $\lambda = 1000*0.00274 = 2.734$

Probability: $p = \exp(-2.734)*2.734^4/4! = 0.151$

Example

Suppose 1% of items made by a factory are defective. Find the probability that 6 defective items in a sample of 300 items.

The Gaussian or Normal Distribution Function

In Statistics, if the number of events is very large (n>20), then the Gaussian (normal) distribution function may be used to describe nearly all events.

The Gaussian distribution is a continuous Random Variable of the form:

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

$$\text{mean}: \quad \mu$$

$$\text{std.dev}: \quad \sigma$$

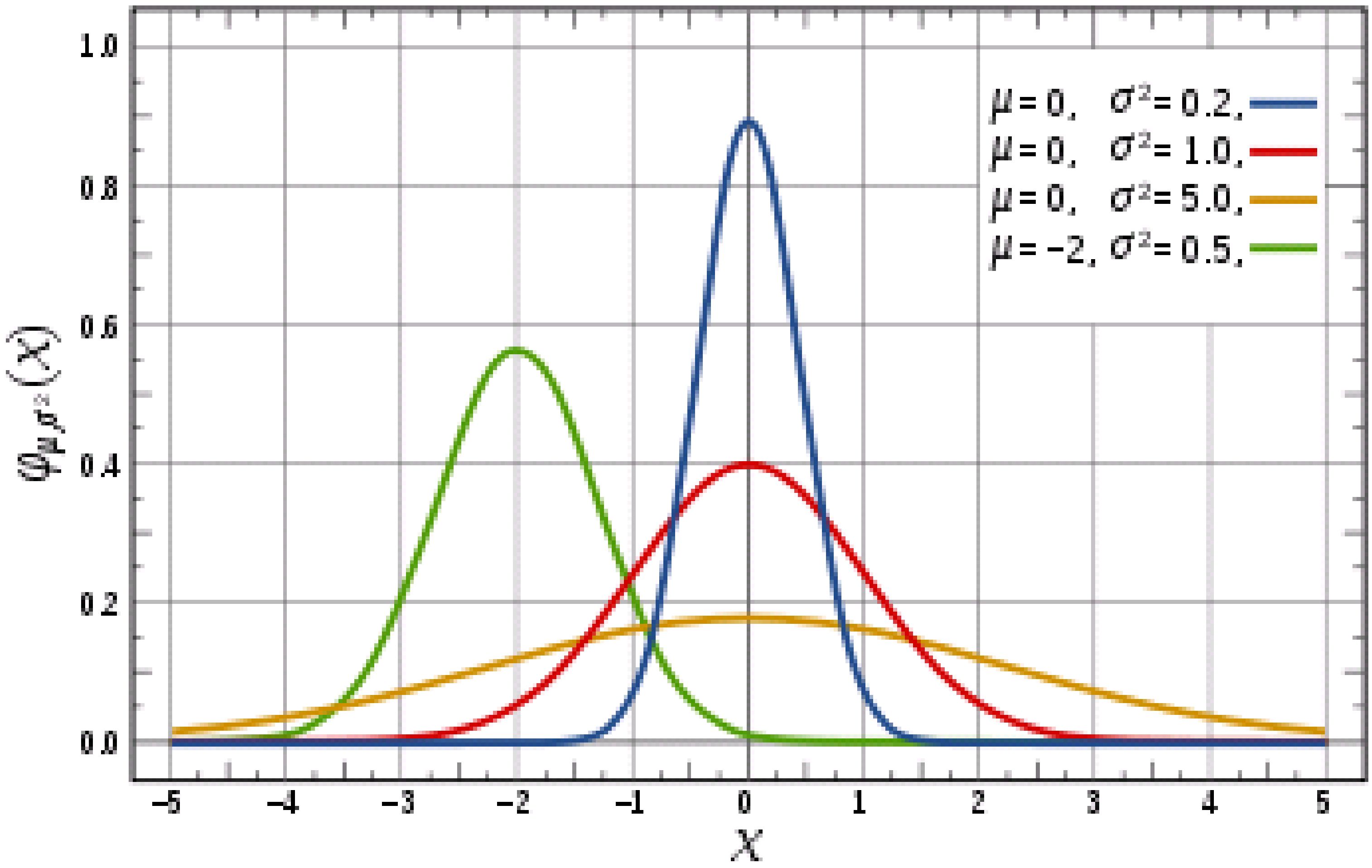
$$p_{gauss}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

μ = mean

σ = standart deviation

π = 3.141593

e = 2.718281



Properties of Gaussian Function

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

$$p(x) \geq 0$$

$$\int_{-\infty}^{+\infty} p(x)dx = 1$$

$$E[X] = \int_{-\infty}^{+\infty} xp(x)dx = \mu$$

$$\int_{a}^{b} p(x)dx = P(a \leq x \leq b)$$

$$\int_{-\infty}^{+\infty} (x-\mu)^2 p(x)dx = \sigma^2$$

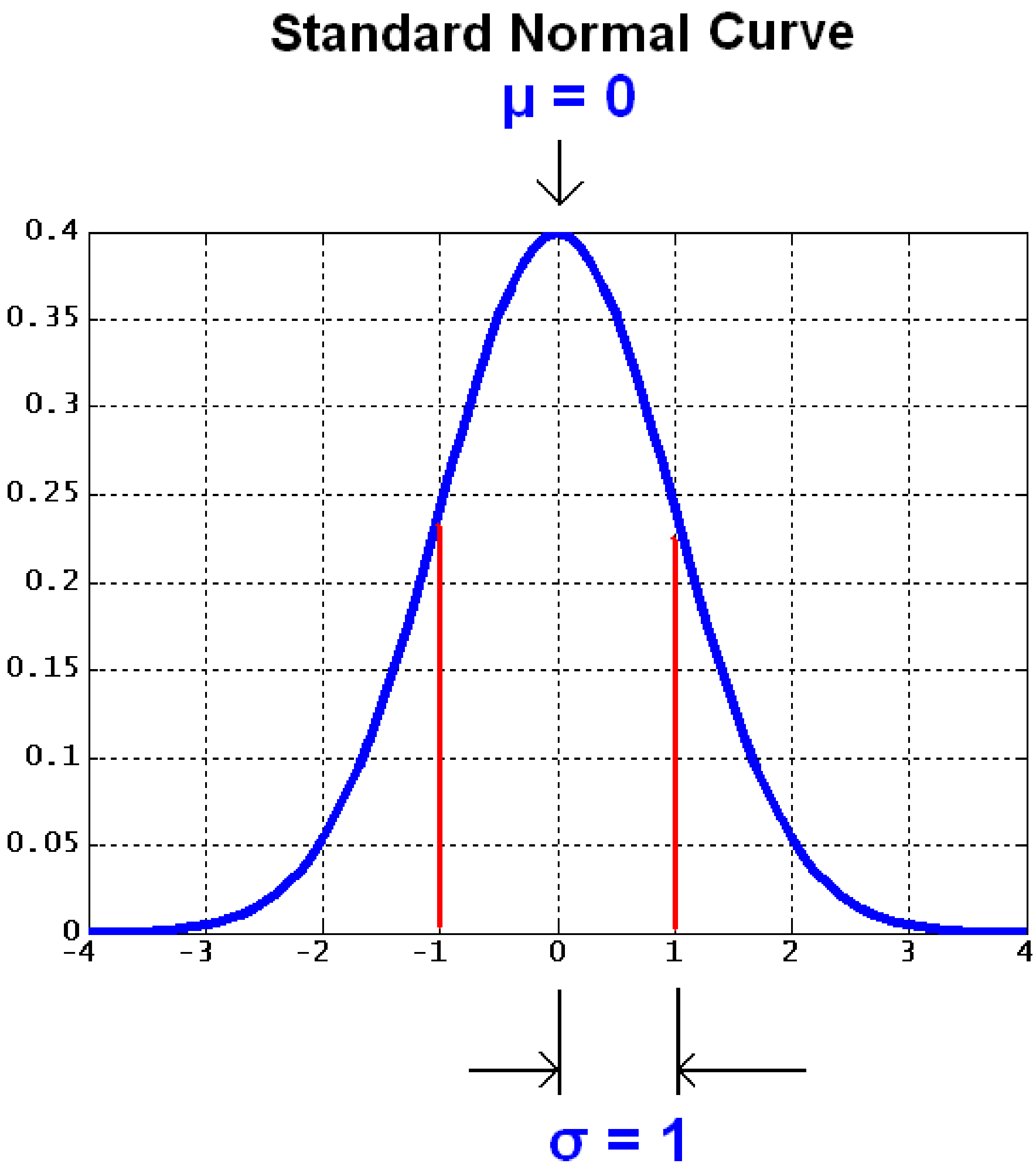
Standard Normal Curve

The normal distribution function for

$\mu = 0$ and $\sigma = 1$

is called the **standard normal distribution function**.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \approx 0.4 \exp(-x^2/2)$$

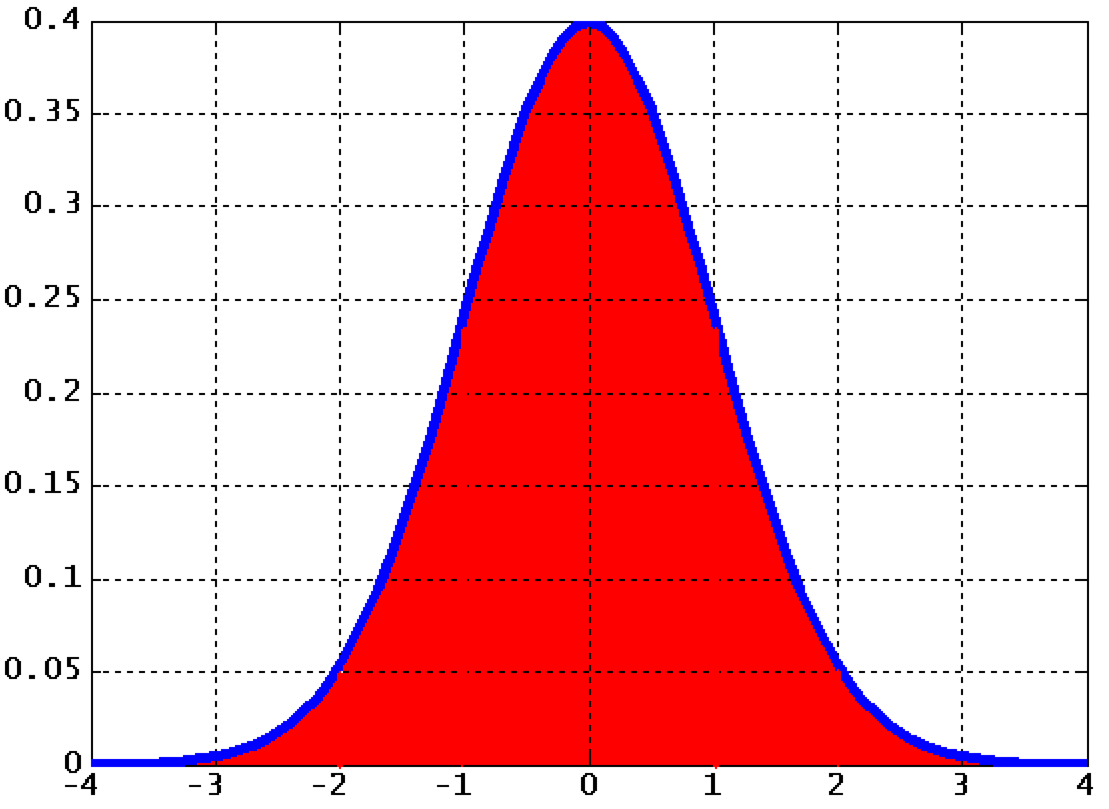

Standard Normal Curve
μ = 0
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-4
-3
-2
-1
0
1
2
3
4
σ = 1

Area Under the Curve

Total area under the standard normal curve is 1.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

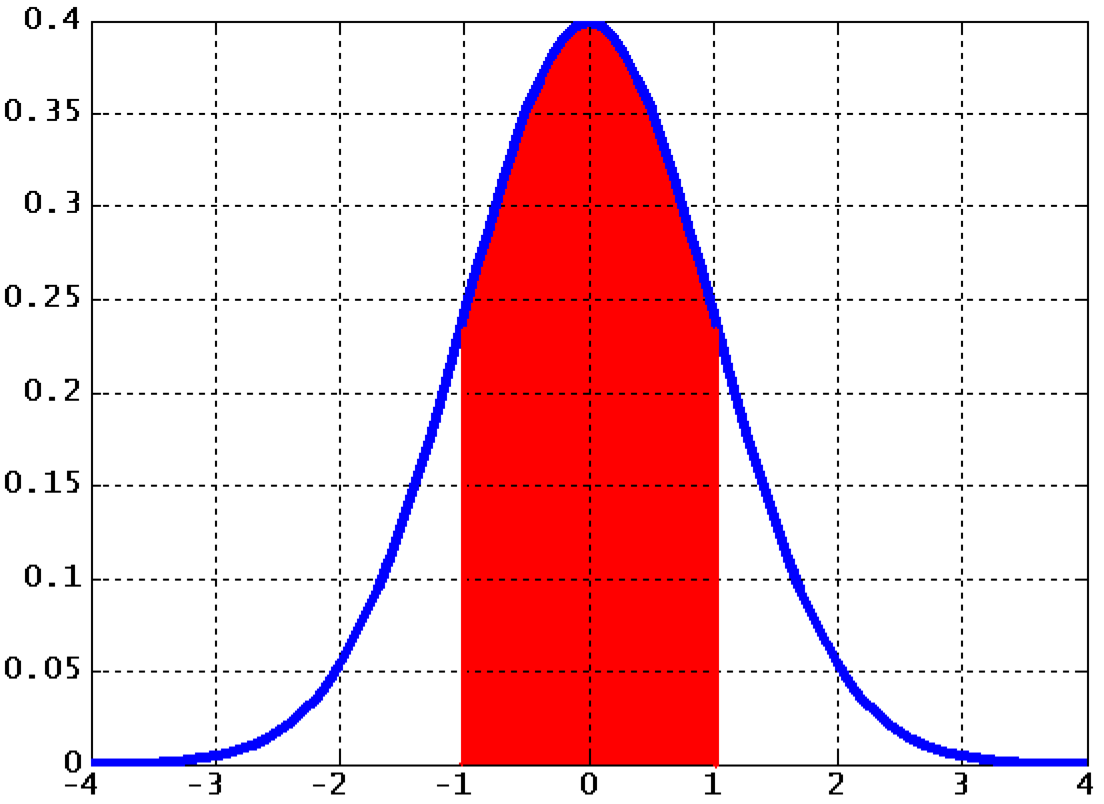


Area under the standard normal curve between [-1, 1] is:

$$\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.6827$$

This corresponds

+- 1 sigma

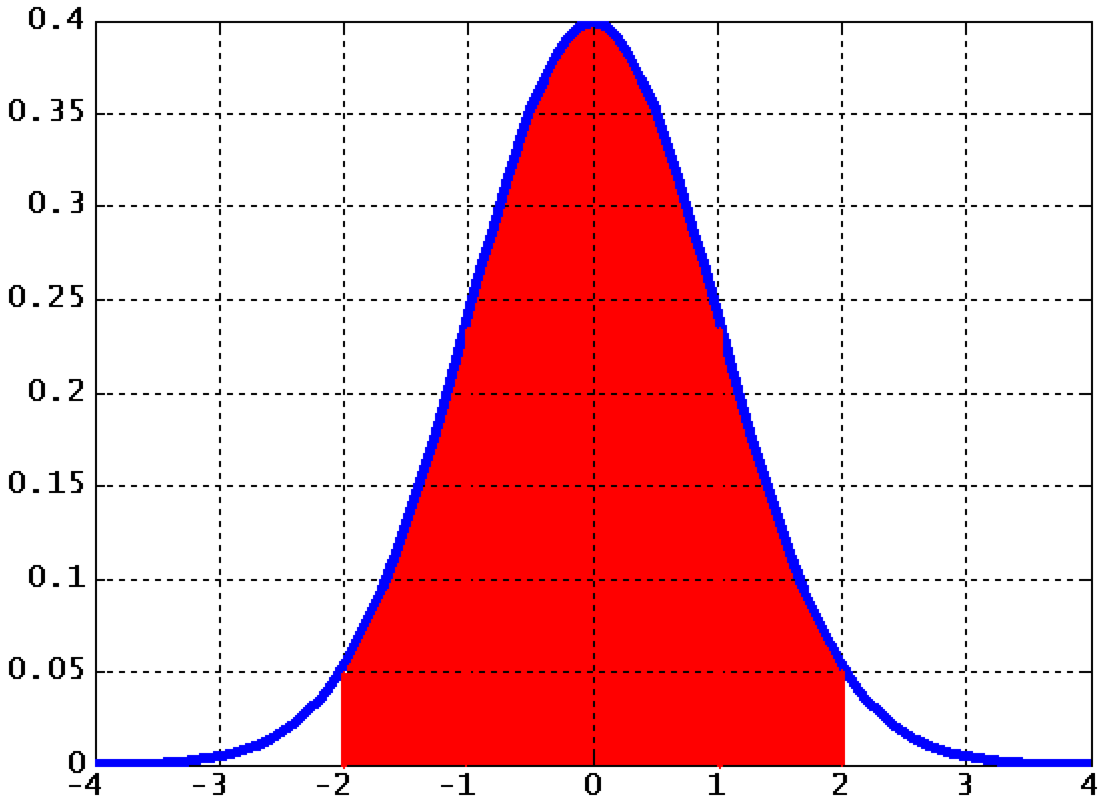


Area under the standard normal curve between [-2, 2] is:

$$\int_{-2}^{2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.9545$$

This corresponds

+- 2 sigma

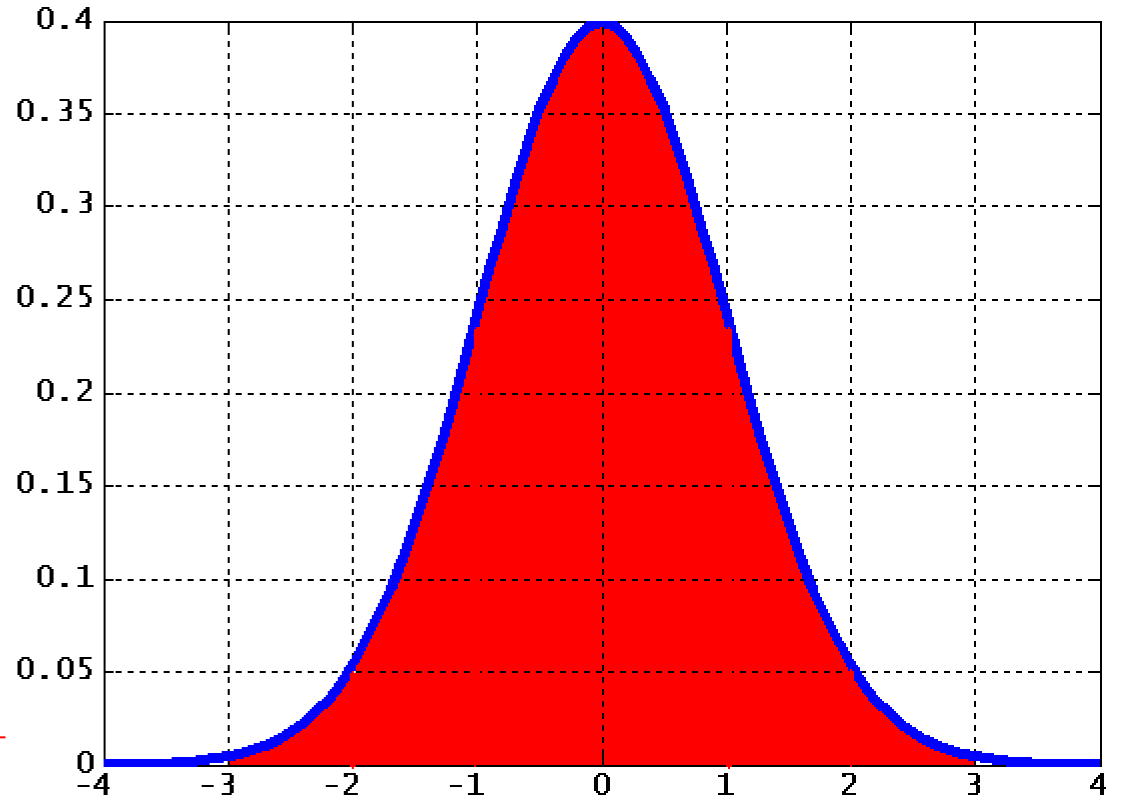


Area under the standard normal curve between [-3, 3] is:

$$\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.9973$$

This corresponds

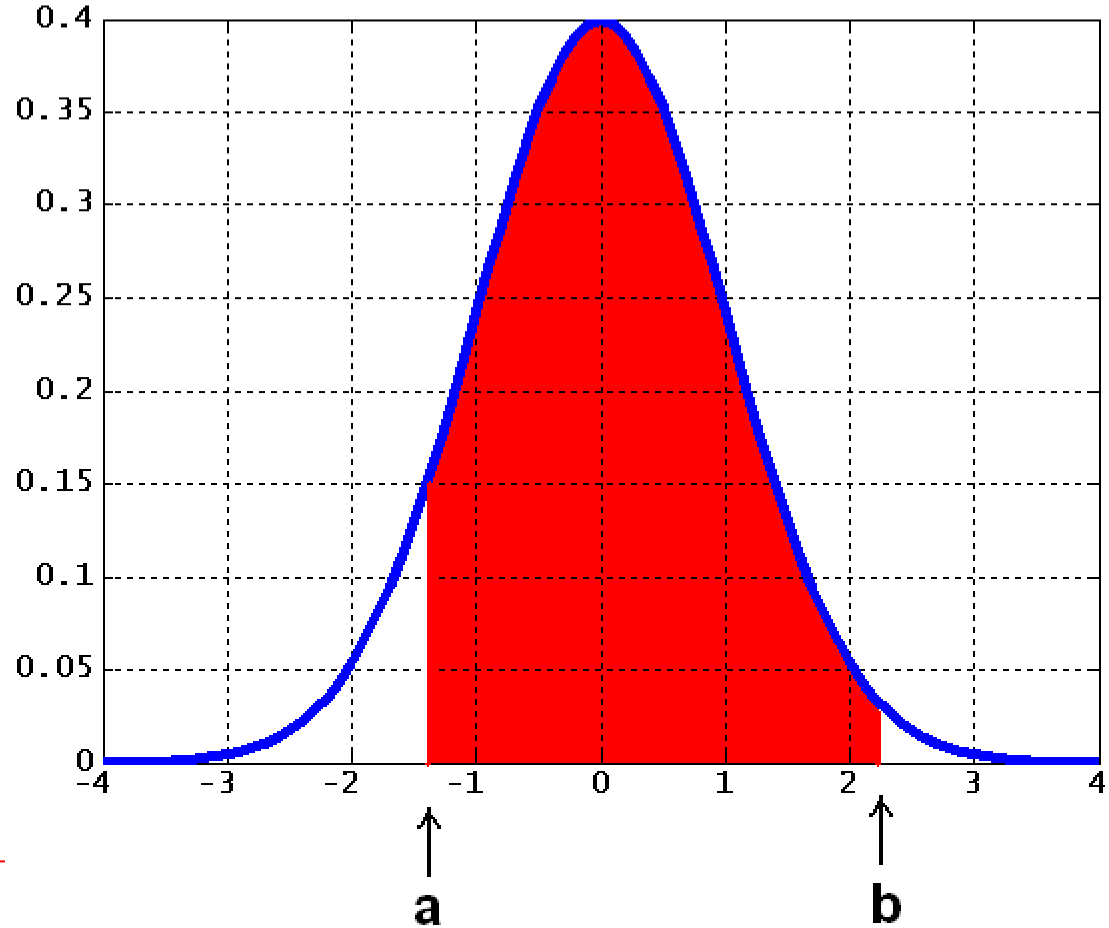
+- 3 sigma

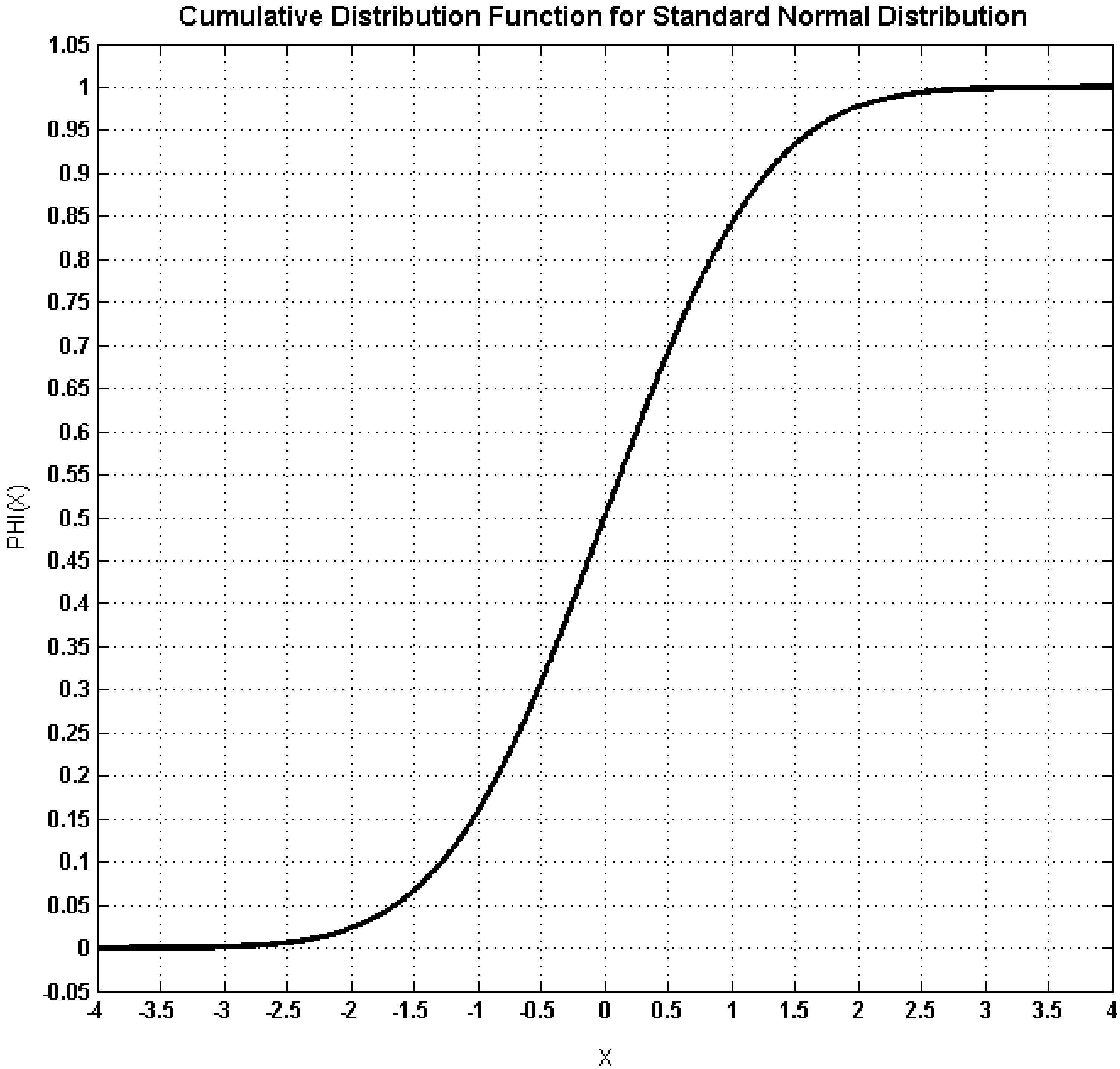


Area under the standard normal curve between [a, b] is:

$$\int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \Phi(b) - \Phi(a)$$

The values of the function phi(x) can be taken from a table or from the figure on next page.




Cumulative Distribution Function for Standard Normal Distribution
PHI(X)
X
1.05
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4

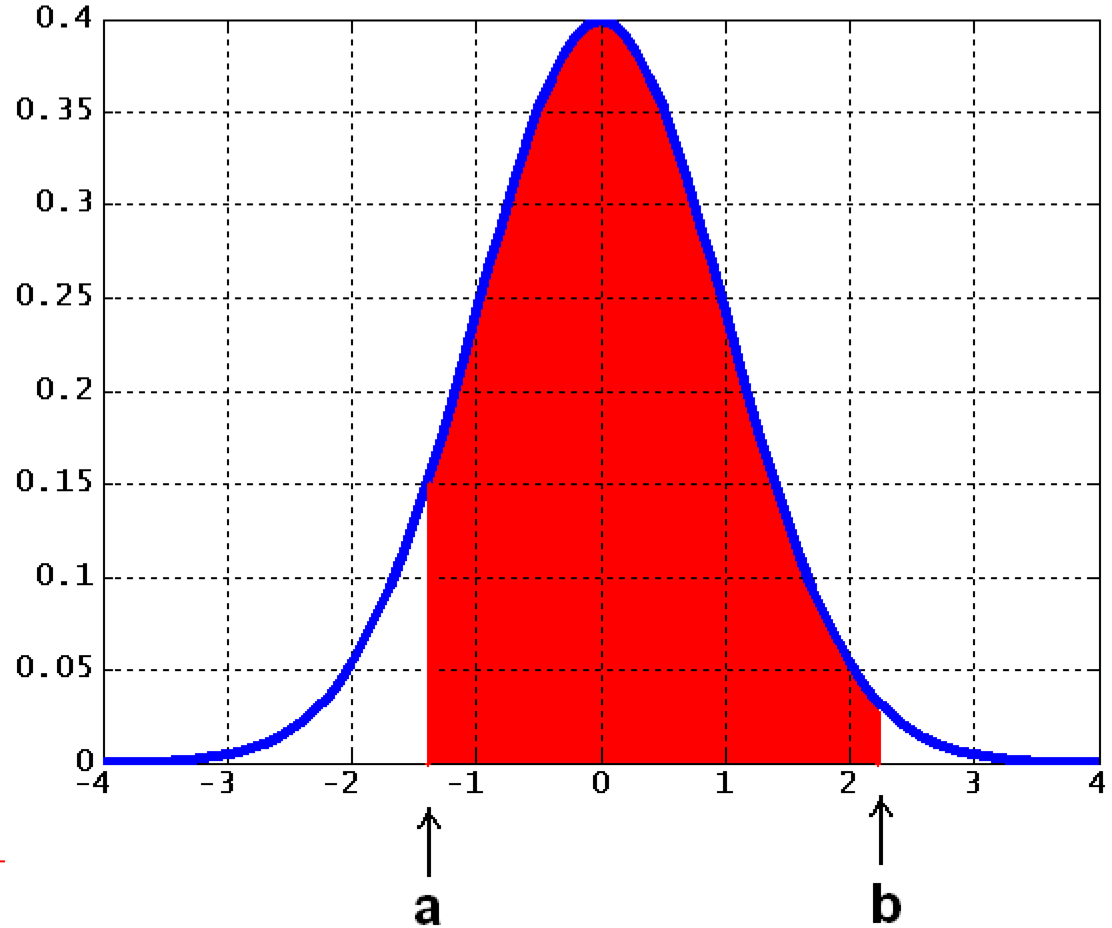
Example:

$$\frac{1}{\sqrt{2\pi}}\int_{-1.2}^{2.3} e^{-x^2/2}dx = \Phi(2.3) - \Phi(-1.2) = 0.99 - 0.12 = 0.87$$

From CDF figure (previous page)

$$\Phi(2.3) \approx 0.99$$

$$\Phi(-1.2) \approx 0.12$$



Alternatively one can use Octave's **`normcdf(x)`** function:

$$\int_{-1.2}^{2.3} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \Phi(2.3) - \Phi(-1.2) = 0.8742$$

```
octave:> normcdf(2.3)
ans = 0.9893
octave:> normcdf(-1.2)
ans = 0.1151
octave:> normcdf(2.3)-normcdf(-1.2)
ans = 0.8742
```

Example: Evaluate the integral

$$\frac{1}{\sqrt{2\pi}}\int_{0}^{1.4} e^{-x^2/2}\,dx$$

Example: Evaluate the integral

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{1.1} e^{-x^2/2}dx$$

Example: Evaluate the integral

$$\frac{1}{\sqrt{2\pi}}\int_{-0.5}^{0.5} e^{-x^2/2}\,dx$$

Example: Let X be a random variable with the standard normal distribution. Find:

(a) $P(0<X<1.4)$

(b) $P(X<-1.1)$

(c) $P(|X|<0.5)$

(d) $P(X>3)$

(e) $P(X>5)$

If a random varibale *X* is normally distributed, we compute the probability that *X* lies between *a* and *b* as follows.

1. First we change *a* and *b* into standard units:

$$a^* = \frac{a-\mu}{\sigma} \qquad b^* = \frac{b-\mu}{\sigma}$$

2. Then we compute the probability from

$$P(a < X < b) = P(a^* < X^* < b^*)$$

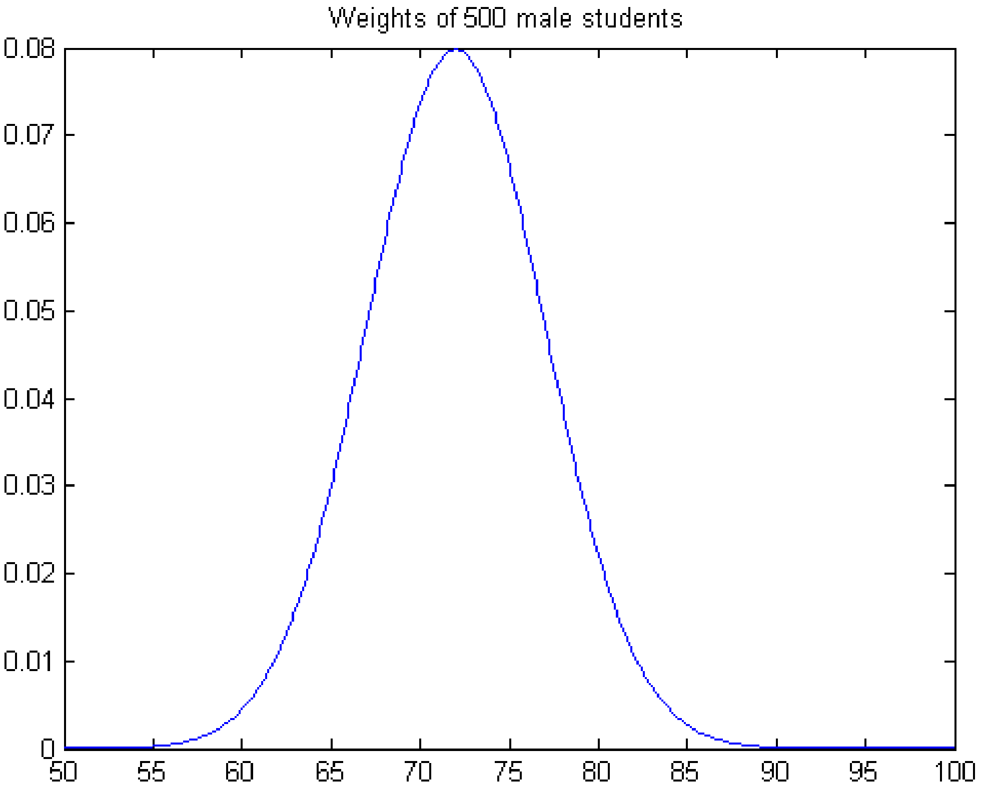
= area under the standard normal curve between a and b

Example: Suppose the temperature during June is normally distributed with mean 20 ^{o}C and standard deviation 3.33 ^{o}C. Find the probability *p* that the temperature is between 21.11 ^{o}C and 26.66 ^{o}C (Answer: 0.3479)

Example: Mean weight of 500 male students at a certain university is 72 kg and the standard deviation is 5 kg. Assuming that the weights are normally distributed, find how many students weigh:

(a) between 66 and 75 kg (Answer: 305)

(b) more than 80 kg (Answer: 27)



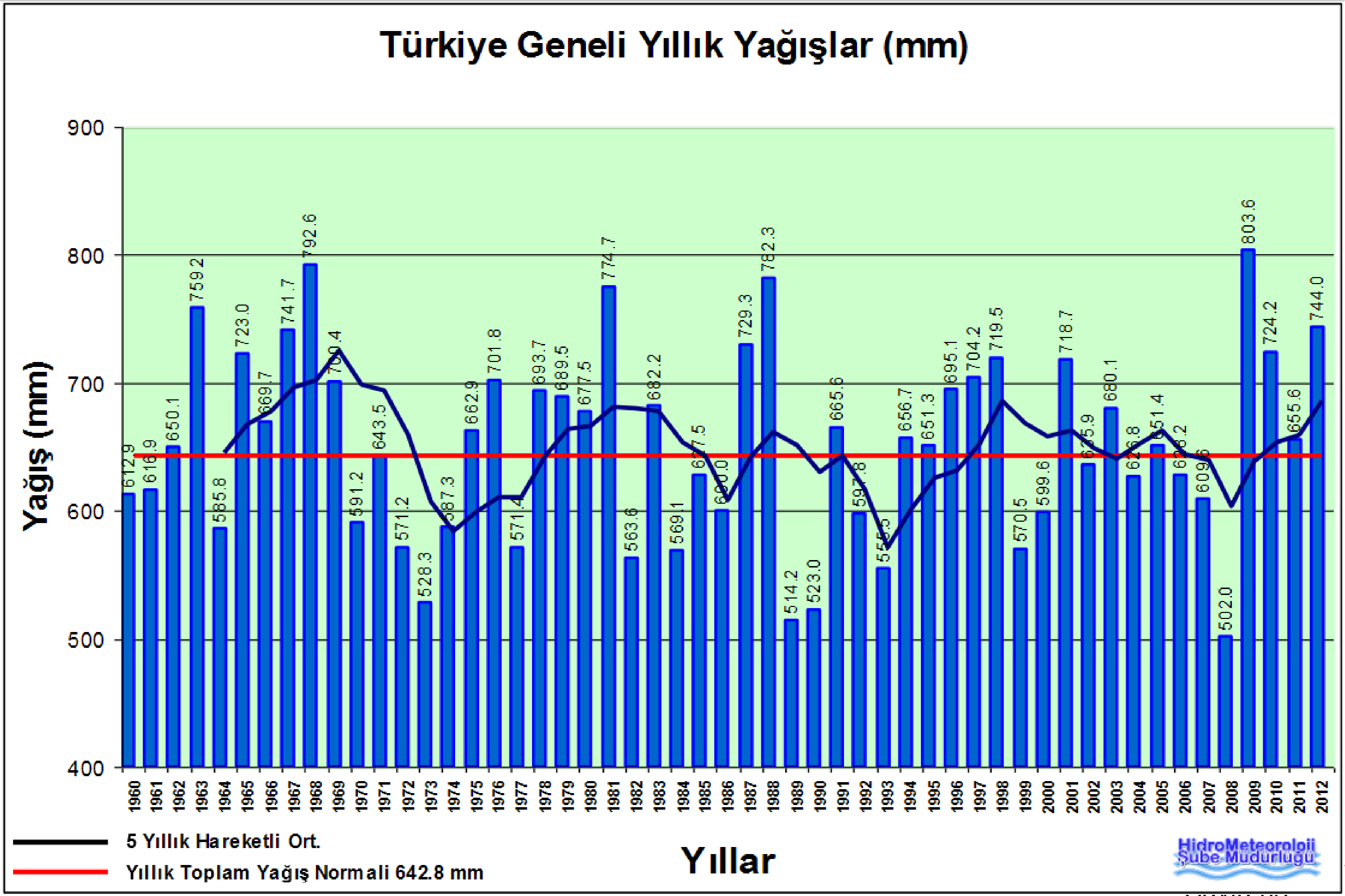
Example Normal Distributions

- Here we will examine some interesting **real data** whose values are distributed normally.

- For each example, histogram of the data is fitted to a Gaussian Function indicated by a blue line.

- All data files can be found at: http://www1.gantep.edu.tr/~bingul/ep122/data

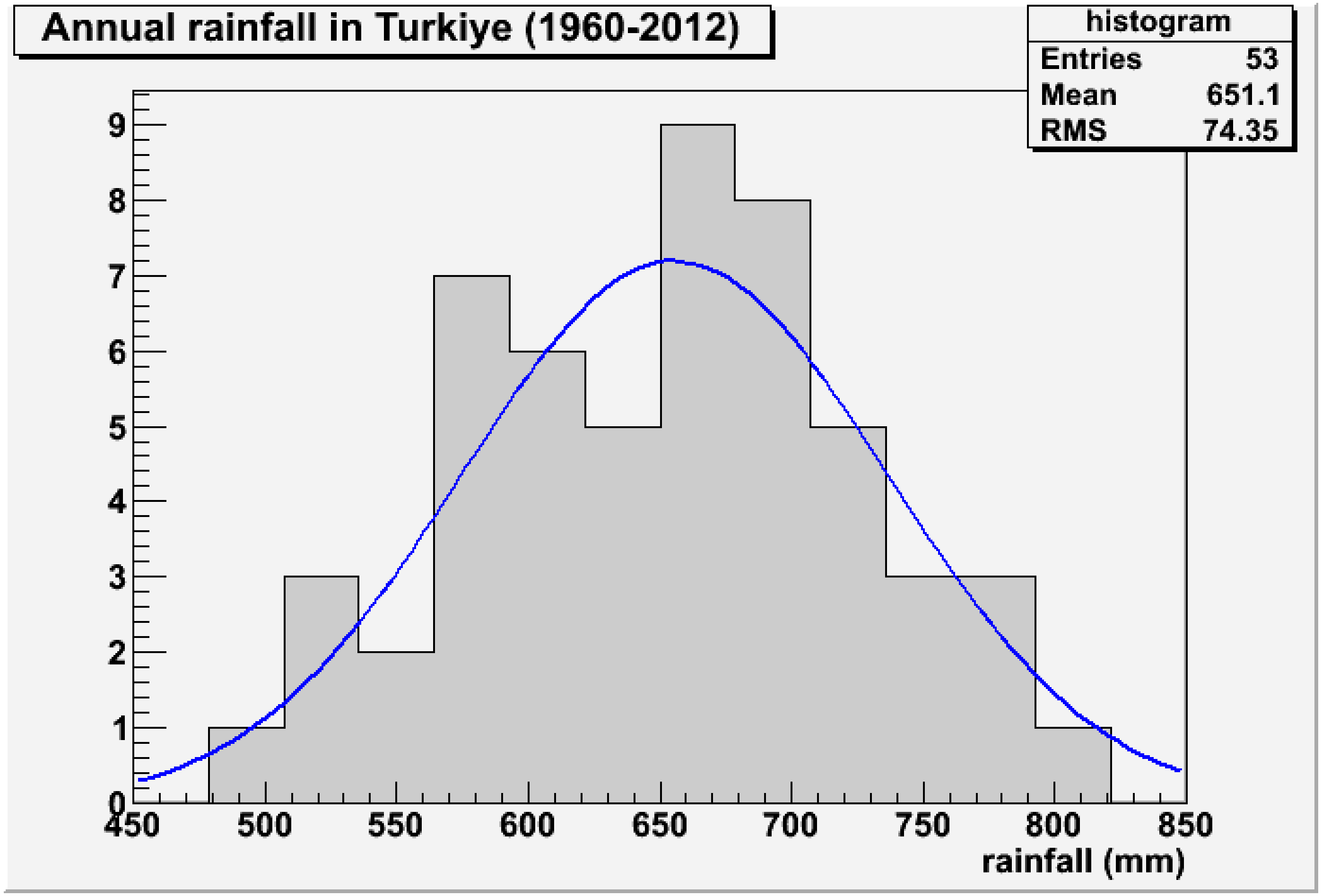
Annual Rainfall (1960-2012)



Annual Rainfall (1960-2012)

Mean : <x> = 651.10 mm

Std. Dev. : σ = 74.35 mm

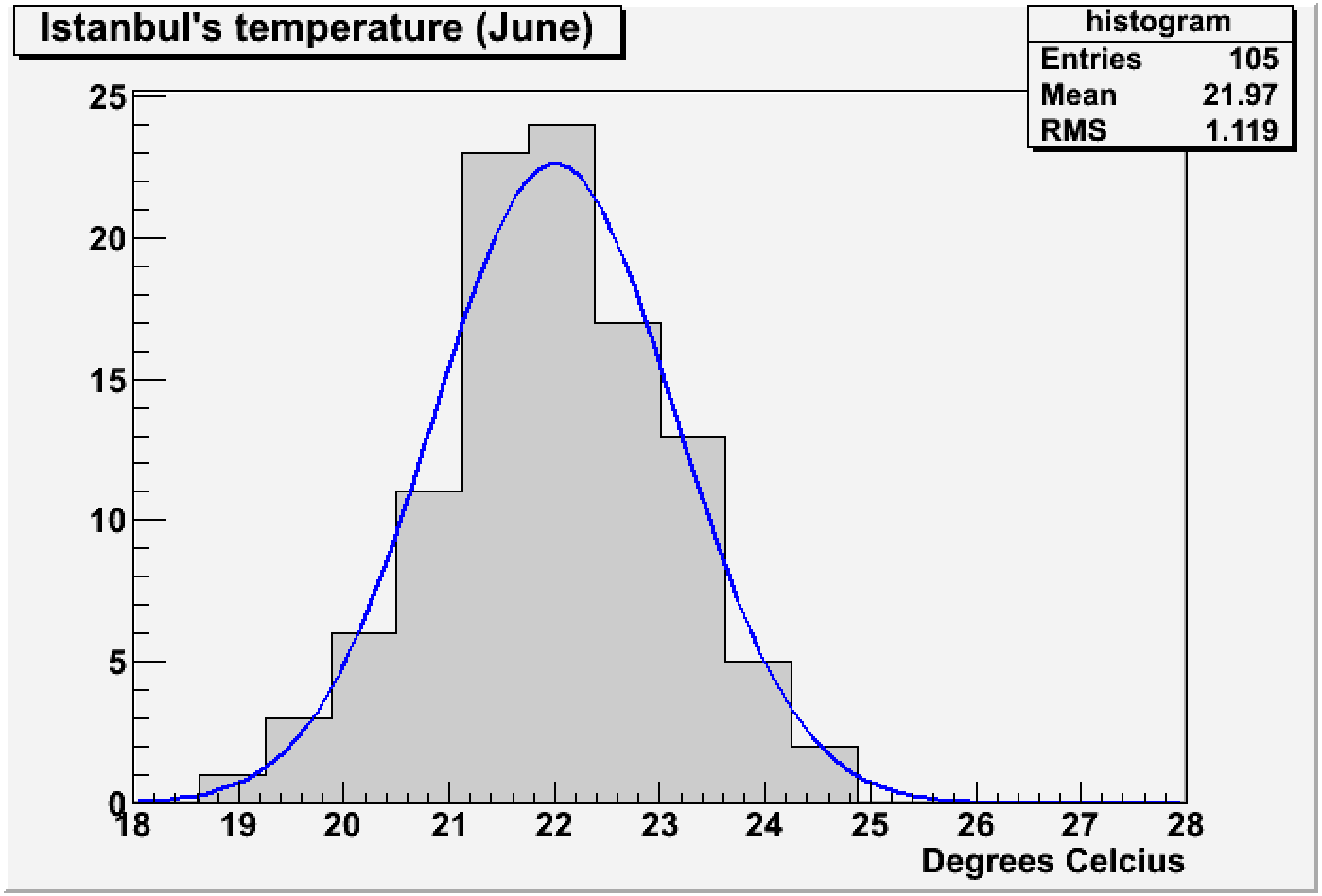


Data: http://www.mgm.gov.tr

Air temperature in Istanbul for the last 105 years.

Mean temperature: <x> = 21.97 ^{o}C

Std. Dev. : σ = 1.12 ^{o}C

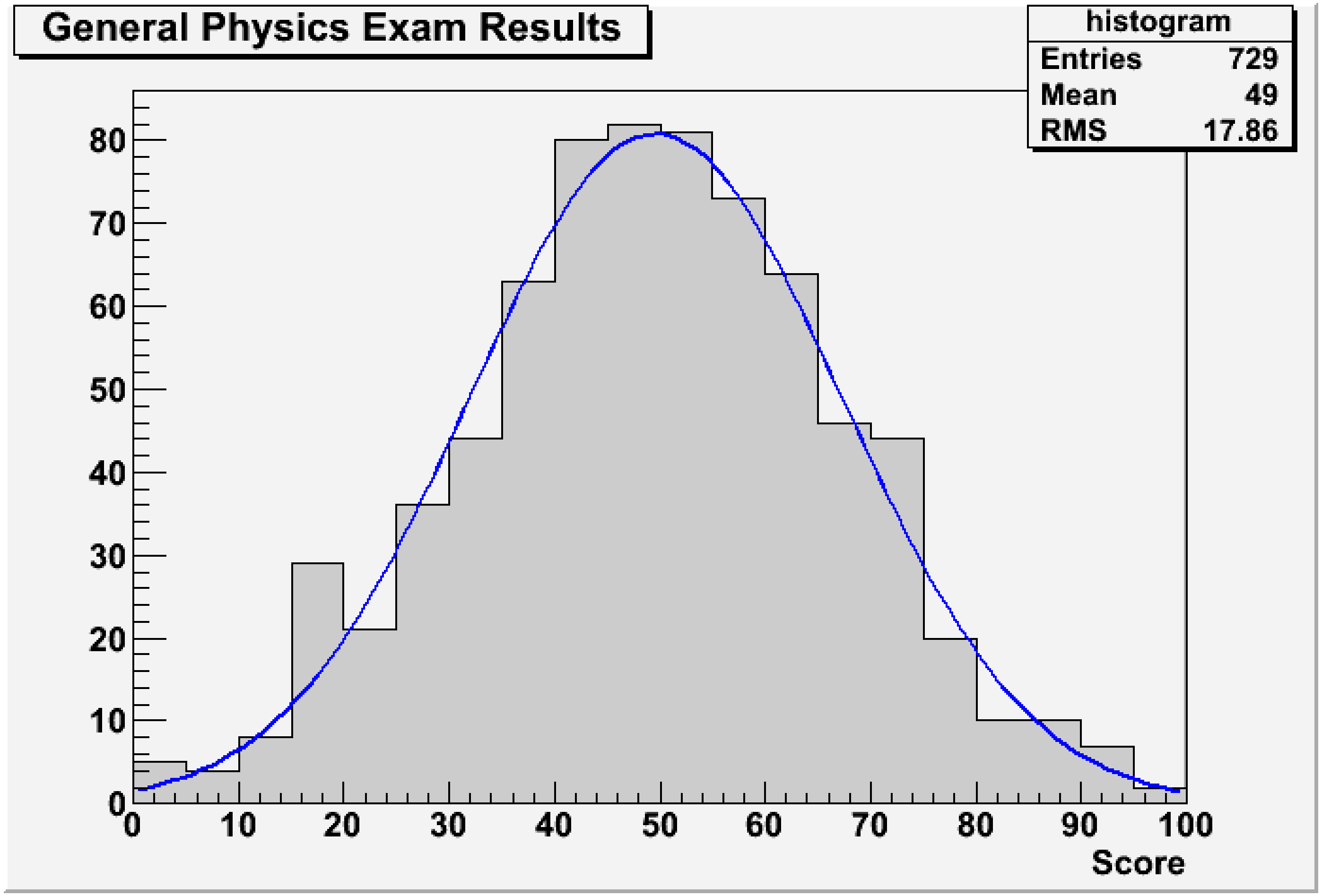


Data: http://data.giss.nasa.gov/tmp/gistemp/STATIONS/tmp_649170620000_14_0/station.txt

“EP106 General Physics II” Course exam results (2010)

Mean score: <x> = 49.0

Std. Dev. : σ = 17.9

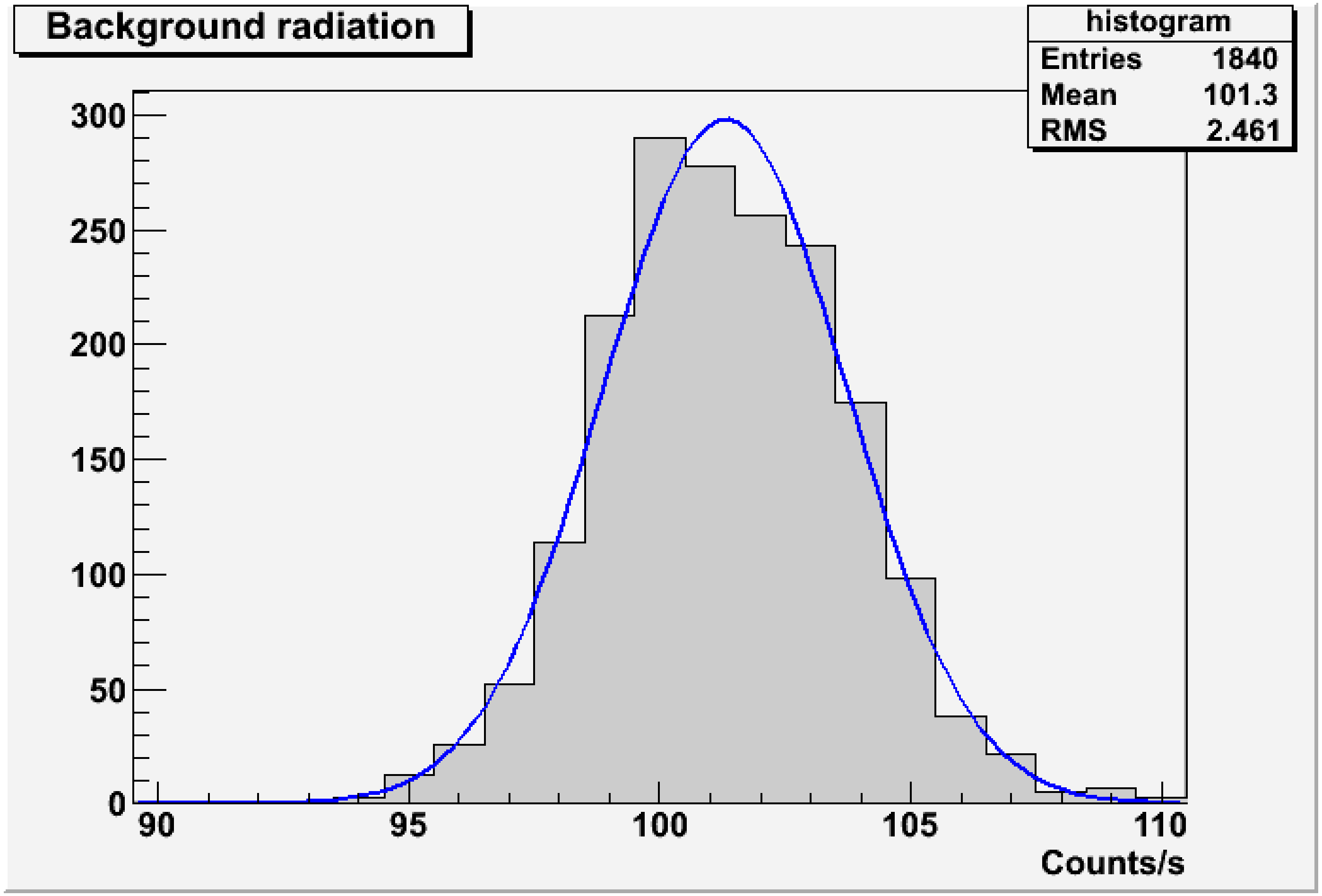


Data: http://www1.gantep.edu.tr/~physics/ep106/exam-statistics.php

Background Radiation in Gaziantep (2013)

Mean : $\langle x \rangle = 101.3$ counts / sec

Std. Dev. : $\sigma = 2.5$ counts / sec



Data is obtained by: Research Assistant Sadık Zuhur (University of Gazaintep)

Questions

1. Let
   U = {a,b,c,d,e,f,g}, A = {a,b,c,d,e}, B = {a,c,e,g}, C = {b,e,f,g}
   Find AnB, AUB, C\B, $(A \backslash B^c)^c$ and C^c n A.

2. Which of the following is a random experiment?
   (a) Tossing a coin.
   (b) Rolling a single 6-sided die.
   (c) Choosing a marble from a jar.
   (d) All of the above.
3. Which of the following is an outcome?
   (a) Rolling a pair of dice.
   (b) Landing on red.
   (c) Choosing 2 marbles from a jar.
   (d) None of the above.
4. Which of the following experiments does NOT have equally likely outcomes?
   (a) Choose a number at random from 1 to 7.
   (b) Toss a coin.
   (c) Choose a letter at random from the word SCHOOL.
   (d) None of the above.

5. What is the probability of choosing a vowel from the English alphabet?
6. A number from 1 to 11 is chosen at random. What is the probability of choosing an odd number?
7. What is the probability of choosing a king from a standard deck of playing cards?
8. What is the probability of choosing the letter i from the word probability?
9. What is the probability of choosing a jack or a queen from a standard deck of 52 playing cards?
10. What is the sample space for choosing a letter from the word mathematics?
11. What is the sample space for choosing a prime number less than 15 at random?
12. What is the probability that a single throw of a die will result in either 2 or 5?

13. A fair coin is tossed 5 times. Find the probability of getting (a) exactly four heads (b) at least two heads (c) no heads.

14. A fair dice is tossed 7 times. Find the probability of getting: (a) exactly four ONEs (b) at least four ONEs (c) no ONEs

15. A woman has 8 children, the probability of each child being female is 50%. What is the probability of being
(a) 4 children female (b) all children female.

16. A communication system contains 6 stations. Independent probability of each station being functional is 90%. If the system requires at least 4 stations to be functional what is the probability that the communication system is functional. (Answer: 0.9842)

17. Suppose 2% of the people on the average are left-handed. Find the probability of 3 or more left-handed among 100 people.
(Answer: 0.325)

18. Given the random variable and corresponding probability mass function (pmf)

$$X = \{1, 2, 3, 4, 5\}$$
$$f(x) = \{0.1, 0.3, 0.4, 0.1, 0.1\}$$

Calculate

(a) $\sum f(x_i)$

(b) $P(1<x<5)$

(c) $E[X]$

(d) $E[X^2]$

(e) RMS

(f) Variance

(g) standard deviation

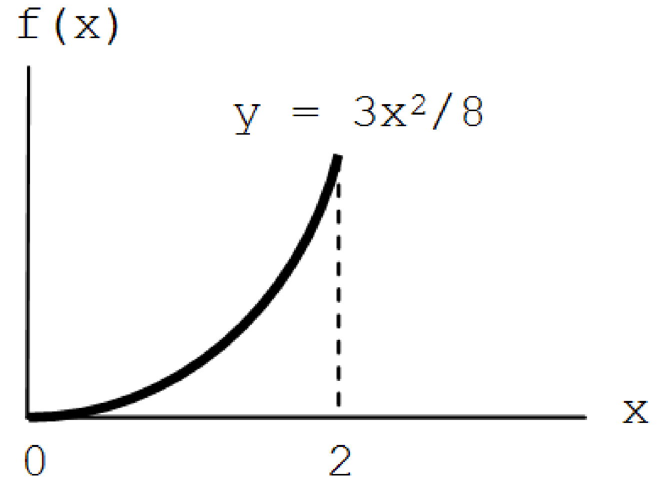
19. Given the probability density function $f(x) = 3x^2/8$.

Calculate

(a) Mean

(b) RMS

(c) $P(x>0.5)$



20. Given the pdf $f(x) = N\sin(x)$ for $0<x<\pi$.

(a) Compute the normalization constant N.

(b) Compute the expectation value and variance.

21. Given the pdf $f(x) = k\exp(-0.1x)$ for $0 < x <$ inf.

(a) Compute the normalization constant k.

(b) Find $P(x>2)$.

22. Evaluate the following integrals:

$$\frac{1}{\sqrt{2\pi}}\int_{1}^{2} e^{-z^2/2}\,dz$$

$$\frac{1}{2\sqrt{2\pi}}\int_{1}^{2} e^{-(z-1)^2/8}\,dz$$

23. Suppose the diameters, D, of screws manufactured by a company are normally distributed with mean 0.25 cm and standard deviation 0.02 cm. A screw is considered defective if its diameter $D < 0.22$ cm.

A sample of 250 screws are selected randomly.

Estimate the number of defective screws in this sample.

References

1. “Data Analysis for Physical Science Students”, L. Lyons, Cambridge University Press
2. “Probability and Statistics” - M. Spiegel et. al., Shaum
3. “Probability” - S. Lipshutz, Shaum
4. “Radiation Detection and Measurement”, G.F.Knoll, Wiley
5. http://www1.gantep.edu.tr/~andrew/eee283