Probability Distribution

- Frequency distribution (from data)
- Graphical representation of frequency distribution – Histogram
- Theoretical distribution – probability distribution

Class interval	Frequency	Relative frequency	Cumulative relative frequency
70 – 90	2	0.0250	0.0250
90 – 110	3	0.0375	0.0625
110 – 130	6	0.0750	0.1375
130 –150	14	0.1750	0.3125
150 –170	22	0.2750	0.5875
170 – 190	17	0.2125	0.8000
190 – 210	10	0.1250	0.9250
210 – 230	4	0.0500	0.9750
230 – 250	2	0.0250	1.0000

Random Variable

- The variable that associates a number with the outcome of a random experiment
- Denoted by upper case letter, such as *X.*
- Discrete/Continuous random variable
- Discrete Probability Distribution
  - Binomial Distribution
  - Poisson Distribution
- Continuous Probability Distribution
  - Exponential Distribution
  - Normal Distribution

- **Probability Distribution Function**: probability that a random variable assumes a particular value
  - Denoted as $P(X = x)$

x	0	1	2
P(X=x)	0.04	0.32	0.64

  - $\sum P(X = x) = 1$

- **Cumulative Distribution Function**
  - Denoted as $P(X \leq x)$

Example 1

Molded parts are classified by a length measurement as follows.

Length	4.9	5.0	5.1	5.2	5.3	5.4	5.5	5.6
Number of parts	0	3	10	21	30	18	16	2

(a) If the random variable X is the length of a randomly selected part, determine the probability distribution function.

(b) What is $P(X \leq 5.2)$?

(c) What is $P(X = 5.4)$?

(d) What is $P(5.0 \leq X \leq 5.5)$?

Mean or Expected Value,

$$E(X) = \sum x \, P(X = x)$$

Variance of a random variable X,

$$Var(X) = \sum [x - E(X)]^2 \, P(X = x)$$

Discrete Probability Distribution

Binomial Distribution

- Process
  - Each trial has two possible outcomes
  - Probability of any outcome remains fixed, say probability of success be *p*, then probability of failure = 1 – *p.*
  - Trials are independent and each trial is repeated *n* times

- X = random variable of the experiment; X = 0, 1, 2,…n
- $P(X = x) = {}^{n}C_{x}\, p^{x}\, (1-p)^{n-x}$; parameters: n, p
- $E(x) = np$
- $V(x) = npq$

Example 1

The random variable X has a binomial distribution with $n = 10$ and $p = 0.5$

Determine the following probabilities.

a) $P(X = 5)$

b) $P(X \leq 2)$

c) $P(X \geq 9)$

d) $P(3 \leq X < 5)$

e) $P(X \geq 3)$

Example 2

Some field representative of the EPA are doing spot checks of water pollution in rivers. Historically, 8 out of 10 such tests produce favorable results (i.e., no pollution). If the field group is going to perform 6 tests, find the chances of getting exactly three favorable results from this group of tests.

HINT:

X = number of tests showing favourable results

p = 0.8, n = 6

P(X = 3)

Example 3

Five employees are required to operate a chemical process; the process cannot be started until all the 5 work stations are manned. Employee records indicate that there is 30% chance of any employee being late and we know that they all come to work independently of each other. Management is interested in knowing the probabilities of 0, 1, 2, 3, 4 or 5 employees being late, so that a decision concerning the number of backup personnel can be made. Determine the probability distribution illustrating this situation.

HINT:

X = number of employees who are late

$p = 0.3$, $n = 5$

P(X = 0), P(X = 1), …….., P(X = 5)

Poisson Distribution

- Discrete distribution
- Random variable, $X = 0, 1, 2, \ldots\ldots$
- $P(X=x) = (\lambda^x e^{-\lambda}) / x!$
- Parameter, λ = mean number of occurrences per interval of time
- $E(X) = \lambda$
- $V(X) = \lambda$

Example 4

Suppose X has a Poisson distribution with mean of 0.4. Determine the following probabilities.

a) $P(X = 0)$
b) $P(X \leq 2)$
c) $P(X = 4)$
d) $P(2 \leq X < 7)$
e) $P(X \geq 3)$

Example 5

From the past record a bank has observed that 12 persons on an average arrives daily in a bank for taking loans. What is the probability that given a particular day

a) exactly 3 customers will arrive for taking loans?

b) at least 2 customers will arrive for taking loans?

HINT:

X = number of customers taking loan in a particular day

$\lambda = 12$

a) $P(X = 3)$ b) $P(X \geq 2)$

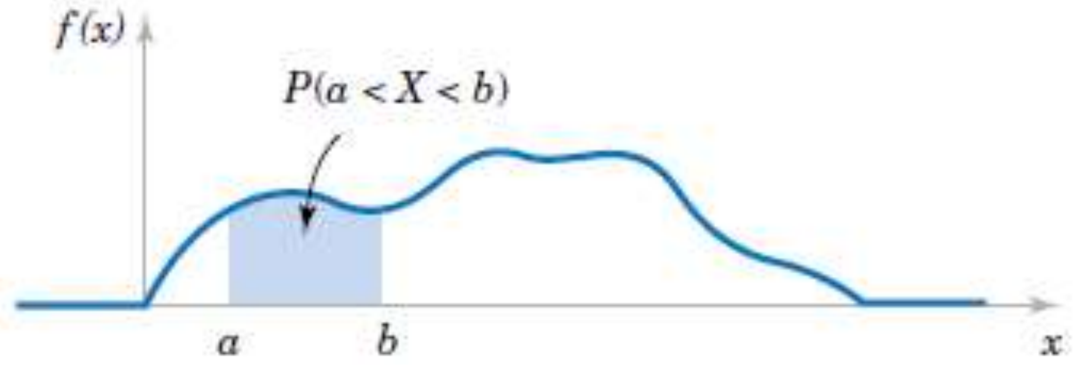
When $n \geq 20$ and $p \leq 0.05$, the Poisson distribution can be used as binominal approximation to avoid tedious job of calculations.

Use $\lambda = np$

$$P(X=x) = [(np)^x e^{-(np)}] / x!$$

Continuous Distribution function

probability density function $f(x)$ is used to describe the probability distribution of a **continuous random variable** X.



For a continuous random variable X, a **probability density function** is a function such that

(1) $f(x) \geq 0$

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

(3) $P(a \leq X \leq b) = \int_{a}^{b} f(x)\, dx =$ area under $f(x)$ from a to b

for any a and b

The **cumulative distribution function** of a continuous random variable X is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\, du$$

The **mean** or **expected value** of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

The **variance** of X, denoted as $V(X)$ or σ^2, is

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

Let the continuous random variable X denote the diameter of a hole drilled in a sheet metal component. The target diameter is 12.5 millimeters. Most random disturbances to the process result in larger diameters. Historical data show that the distribution of X can be modeled by a probability density function $f(x) = 20e^{-20(x-12.5)}$, $x \geq 12.5$.

If a part with a diameter larger than 12.60 mm is scrapped, what proportion of parts is scrapped?

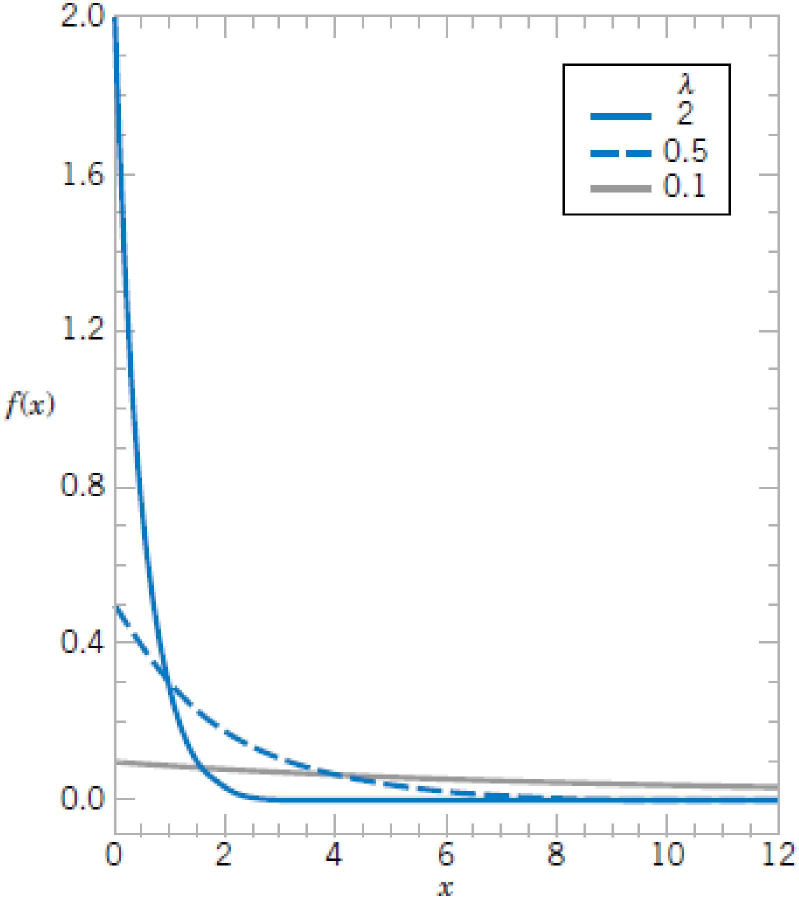
$$P(X > 12.60) = \int_{12.6}^{\infty} f(x)\,dx = \int_{12.6}^{\infty} 20e^{-20(x-12.5)}\,dx = -e^{-20(x-12.5)}\Big|_{12.6}^{\infty} = 0.135$$

$$E(X) = \int_{12.5}^{\infty} xf(x)\,dx = \int_{12.5}^{\infty} x\,20e^{-20(x-12.5)}\,dx = 12.55$$

$$V(X) = \int_{12.5}^{\infty} (x - 12.55)^2 f(x)\,dx = 0.0025.$$

Exponential distribution

If the random variable X has an exponential distribution with parameter λ,



$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \le x < \infty$$

$$F(x) = P(X \le x) = 1 - e^{-\lambda x}$$

$$\mu = E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{\lambda^2}$$

EXAMPLE: In a large corporate computer network, user log-ons to the system can be modelled as a Poisson process with a mean of 25 log-ons/hour. What is the probability that there are no logons in an interval of 6 minutes?

- Let *X* denote the time in hours from the start of the interval until the first log-on.
- Then, *X* has an exponential distribution with λ = 25 log-ons/hour.
- We are interested in the probability that *X* exceeds 6 minutes.
- Because λ is given in log-ons/hour, we express all time units in hours. That is, 6 minutes = 0.1 hour.
- The probability requested is

$$P(X > 0.1) = \int_{0.1}^{\infty} 25e^{-25x}\,dx = e^{-25(0.1)} = 0.082$$

$$P(X > 0.1) = 1 - F(0.1) = e^{-25(0.1)}$$

Normal Probability Distribution

- Most useful and frequently encountered continuous probability distribution
- Random variables like
  - Monthly rate of return for a particular stock
  - Weekly sales
  - Height of persons
  - Length of any component/part manufactured
  - Tensile strength of steel samples
- X = normal random variable; $-\infty < x < +\infty$

(i) Bell-shaped (ii) Symmetrical,

(iii) Single peak (iv) Mean = median = mode

Parameters

μ = mean of the random variable, x

σ = standard deviation of x

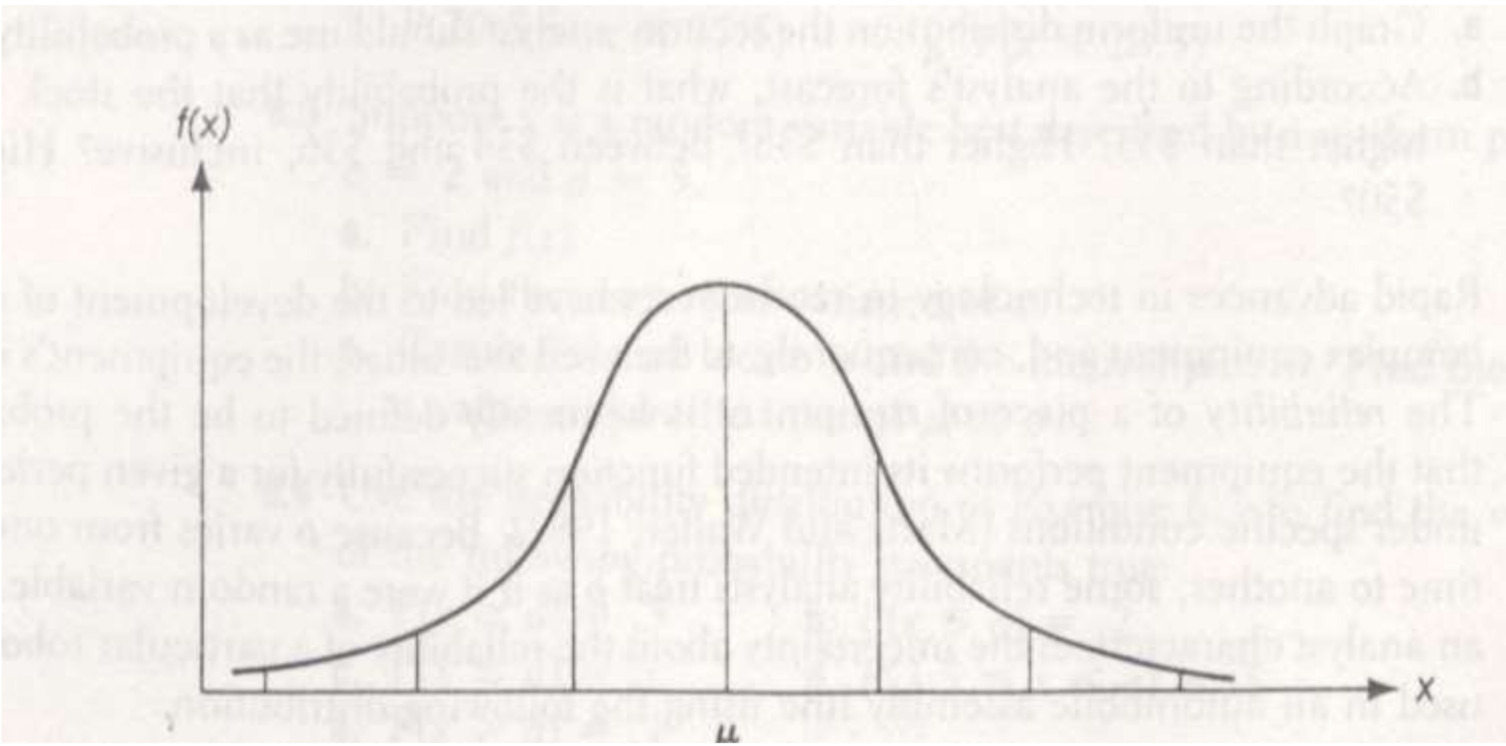


Figure 1: A normal probability distribution

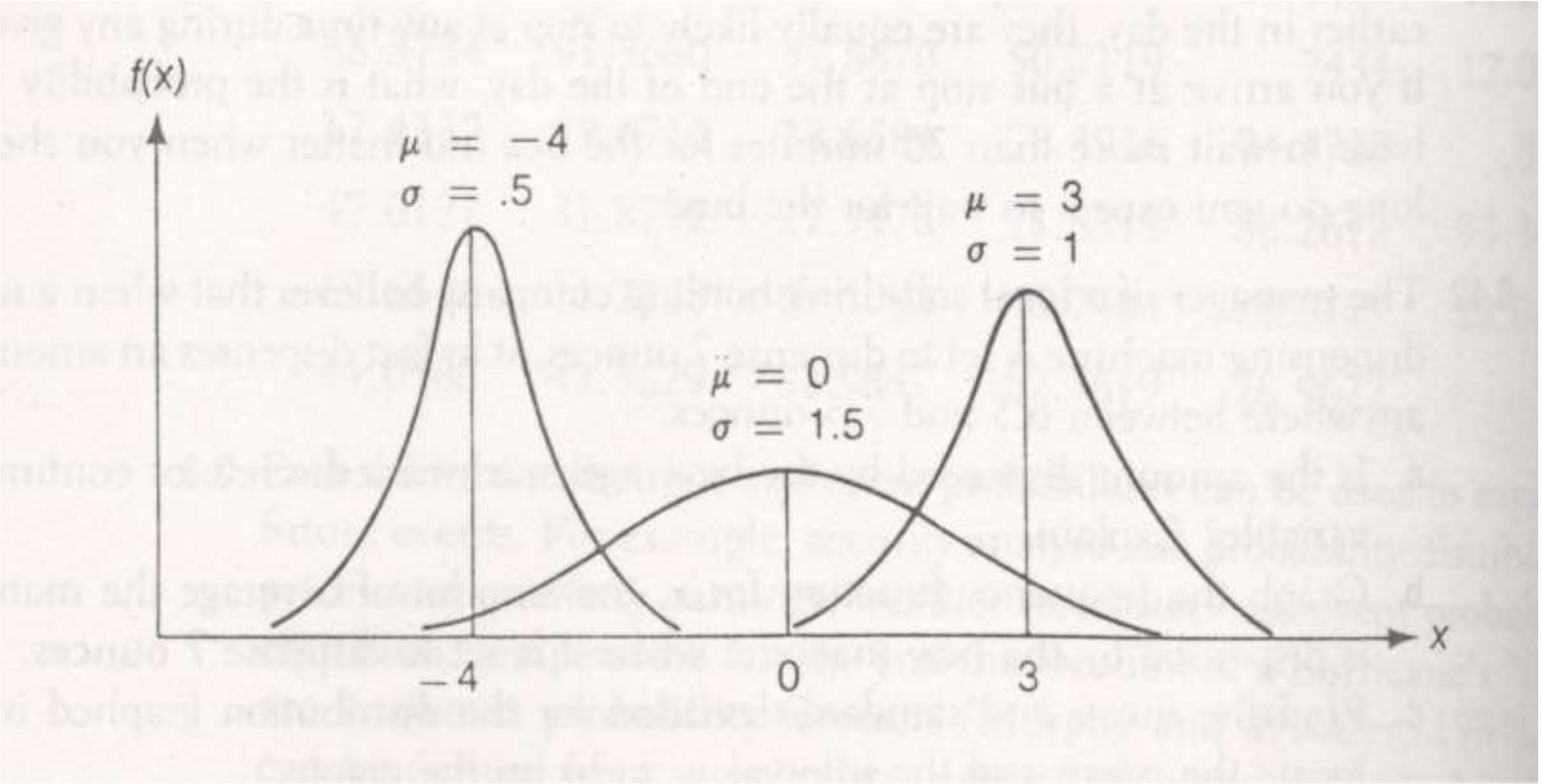
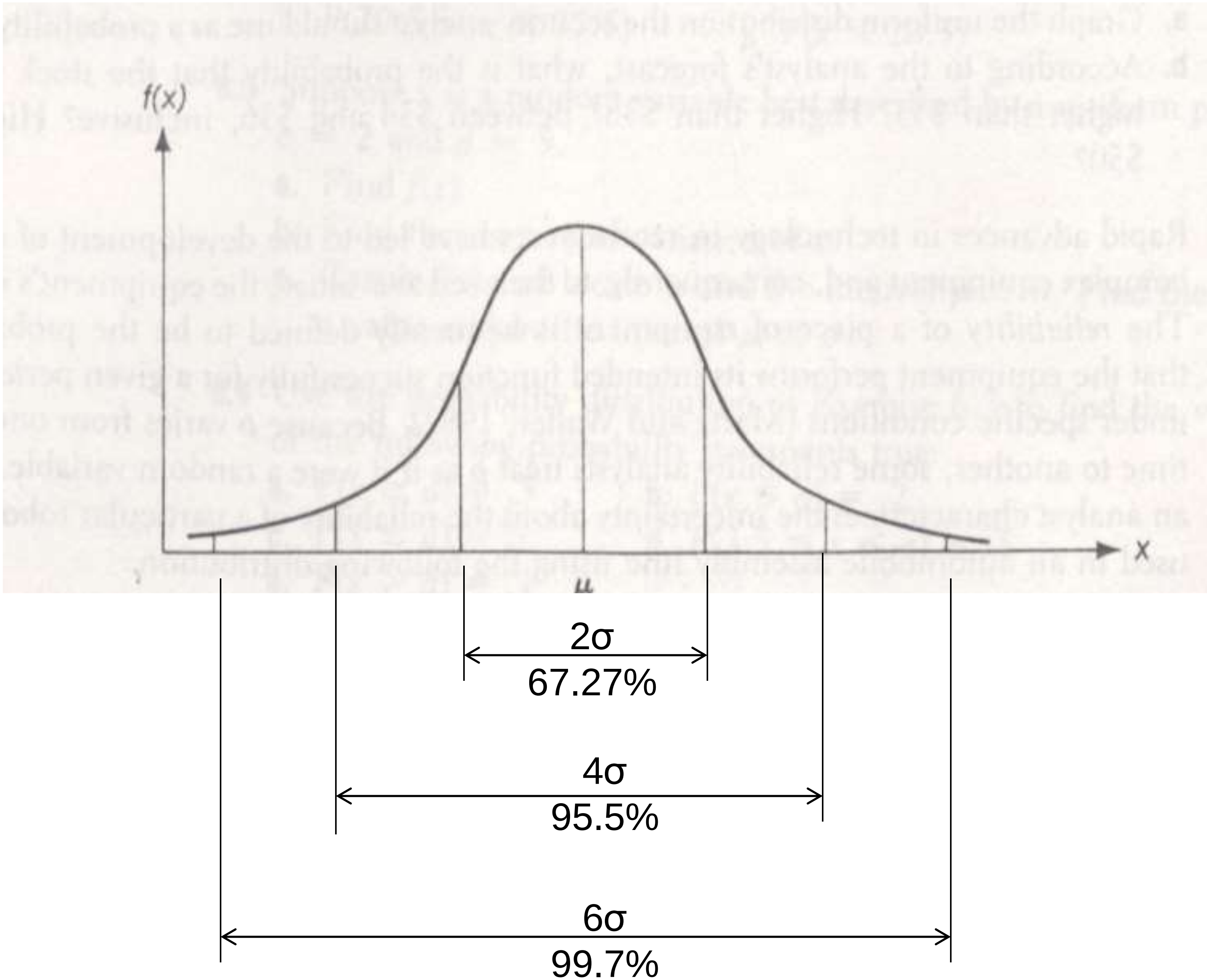


Figure 2: Several normal probability distributions with different mean and standard deviation


f(x)
x
μ
2σ
67.27%
4σ
95.5%
6σ
99.7%

Probability density function

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

Cumulative Distribution Function

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Standard Normal distribution

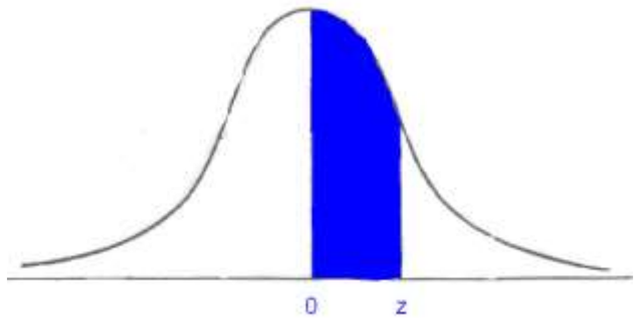
- Parameters

  mean of the random variable, $\mu = 0$

  standard deviation of, $\sigma = 1$

- Area under standard normal curve is 1
- Z-transformation

$$z = \frac{x - \mu}{\sigma}$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Example 1

Determine the following probabilities for the standard normal variable Z.

a) $P(Z < 1.32)$
b) $P(Z < 3)$
c) $P(Z > 1.45)$
d) $P(Z > -2.15)$
e) $P(-2.34 < Z < 1.76)$

Example 2

Determine the following probabilities for the standard normal variable Z.

a) $P(-1 < Z < 1)$
b) $P(-2 < Z < 2)$
c) $P(-3 < Z < 3)$
d) $P(Z > 3)$
e) $P(0 < Z < 1)$

Example 3

Determine the value of z that solves the following.

a) $P(Z < z) = 0.9$
b) $P(Z < z) = 0.5$
c) $P(Z > z) = 0.1$
d) $P(Z > z) = 0.9$
e) $P(Z < z) = 0.4$
f) $P(1 < Z < z) = 0.2$
g) $P(-2 < Z < z) = 0.3$
h) $P(-z < Z < z) = 0.8$

Example 4

Assume the random variable X is normally distributed with mean 10 and a variance of 4. Determine the following probabilities.

a) $P(X<13)$

b) $P(X>9)$

c) $P(6<X<14)$

d) $P(7<X<9)$

e) $P(11<X<13)$

Example 5

The compressive strength of samples of cement can be modeled by a normal distribution with mean of 6000 kg/cm^2 and a standard deviation of 100 kg/cm^2.

i) What is the probability that a sample's strength is less than 6250 kg/cm^2?

ii) What is the probability that a sample's strength is between 5800 kg/cm^2 and 5900 kg/cm^2 ?

iii) What strength is exceeded by 95% of the samples?