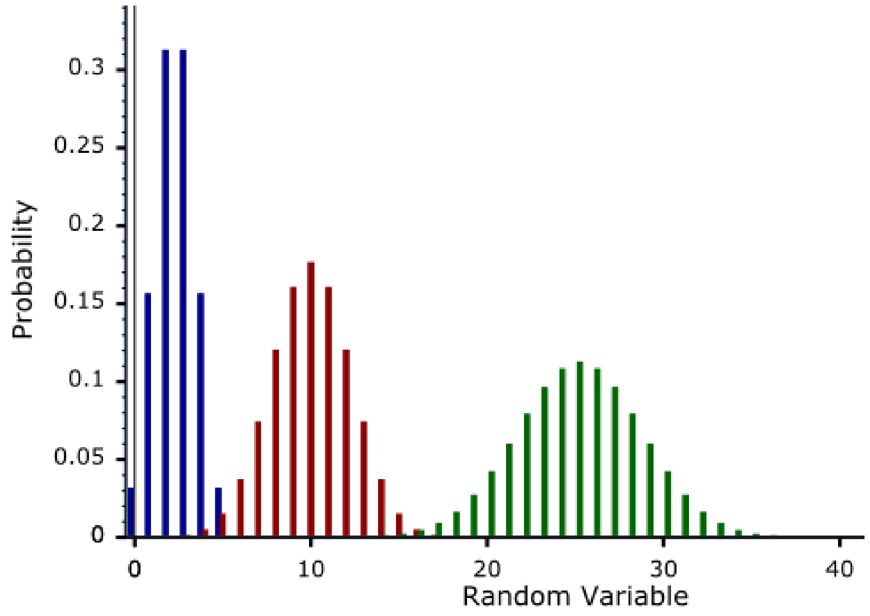
Random variables and
probability distributions



Random Variables (1)

- Random variable is a *numerical summary* of random outcomes.
- A random variable is a variable that takes on numerical values realized by the outcomes in the sample space generated by a *random experiment*.

> A **random variable** X associates a numerical value with each outcome of an experiment.

- The word "random" serves as a reminder of the fact that, beforehand, we *do not know* the outcome of an experiment or its *associated value* of X.
- We use *capital letters* like X and Y *to denote random variables* and *lowercase letters* like x and y to denote their *specific values*.

Example: Consider X to be the *number of heads* obtained in *three tosses* of a coin.

a. List the numerical values of X and the corresponding elementary outcomes.

b. Identify all of the elementary outcomes that produce the value x

Random Variables (2)

Answer:

a. First, X is a variable since the *number of heads* in three tosses of a coin can have any of the values *0, 1, 2, or 3*. Second, this *variable is random* in the sense that the value that would occur in a given instance *cannot be predicted with certainty*. We can, though, make a list of the elementary outcomes and the associated values of X.

Outcome	Value of X
HHH	3
HHT	2
HTH	2
HTT	1
THH	2
THT	1
TTH	1
TTT	0

b. Scanning our list, we now identify the events (i.e., the collection of the elementary outcomes) that correspond to the distinct values of X.

Numerical Value of X as an Event		Composition of the Event
$[X = 0]$	=	{TTT}
$[X = 1]$	=	{HTT, THT, TTH}
$[X = 2]$	=	{HHT, HTH, THH}
$[X = 3]$	=	{HHH}

Random Variables (3)

Random variables can be *discrete* or *continuous*.

A. Discrete random variable

- **Probability distribution of a discrete random variable**
- Mean (expected value) and standard deviation of a probability distribution

B. Continuous random variable

- Probability model for a continuous random variable
- Normal distribution
- Standard normal distribution
- Probability calculations with normal distributions

Probability distribution of a discrete random variable (1)

- The *probability distribution* of a random variable is a representation of the *possibilities* for all possible outcomes.

> The **probability distribution** or, simply the **distribution,** of a discrete random variable X is a list of the distinct numerical values of X along with their associated probabilities.
>
> Often, a formula can be used in place of a detailed list.

- The *probability* that a particular value x_i occurs is denoted by $f(x_i)$.

Form of a Discrete Probability Distribution

Value of x	Probability $f(x)$
x_1	$f(x_1)$
x_2	$f(x_2)$
.	.
.	.
.	.
x_k	$f(x_k)$
Total	1

Probability distribution of a discrete random variable (2)

Example:

If X represent the *number of heads* obtained in three tosses of a fair coin, find the probability distribution of X.

Answer: The distinct values of X are $x = 0, 1, 2, and\ 3$. Their probabilities would be,

The Probability Distribution of X,
the Number of Heads in Three Tosses of a Coin

Value of X	Probability
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$
Total	1

Probability distribution of a discrete random variable (3)

The **probability distribution** of a discrete random variable X is described as the function

$$f(x_i) = P[X = x_i]$$

which gives the probability for each value and satisfies:

1. $f(x_i) \geq 0$ for each value x_i of X
2. $\sum_{i=1}^{k} f(x_i) = 1$

- A probability distribution or the probability function describes the manner in which the ***total probability 1*** gets apportioned to the individual values of the random variable.

Probability distribution of a discrete random variable (4)

Example:

The following table describes the number of homework assignments *due next week* for a randomly selected set of students taking at least 14 credits. Determine the probability that;

a) X is equal to or larger than 2 and

b) X is less than or equal to 2.

Answer

a) The event $[X \geq 2]$ is composed of $[X = 2]$, $[X = 3]$ and $[X = 4]$. Thus,

$$P[X \geq 2] = f(2) + f(3) + f(4)$$

$$= 0.40 + 0.25 + 0.10 = \mathbf{0.75}$$

b) Similarly,

$$P[X \leq 2] = f(0) + f(1) + f(2)$$

$$= 0.02 + 0.23 + 0.40 = \mathbf{0.65}$$

A Probability Distribution for Number of Homework Assignments Due Next Week

Value x	Probability $f(x)$
0	.02
1	.23
2	.40
3	.25
4	.10

A. Discrete random variable

- ✓ **Probability distribution of a discrete random variable**
- ✓ **Mean (expected value) and standard deviation of a probability distribution**

B. Continuous random variable

- Probability model for a continuous random variable
- Normal distribution
- Standard normal distribution
- Probability calculations with normal distributions

Mean (expected value) of a probability distribution (1)

- The mean of a random variable X is also called its **expected value** and, alternatively, denoted by $E(X)$.

> The **mean** of X or **population mean**
>
> $$\begin{aligned} E(X) &= \mu \\ &= \sum (\text{Value} \times \text{Probability}) = \sum x_i f(x_i) \end{aligned}$$
>
> Here the sum extends over all the distinct values x_i of X.

Example: A construction company submits bids for two projects. Listed here are the *profit* and the *probability of winning each project*. Assume that the outcomes of the *two bids are independent*.

a) List the possible outcomes (win/not win) for the two projects and find their probabilities.

b) Let X denote the company's total profit out of the two contracts. Determine the probability distribution of X.

c) If it costs the company \$2000 for preparatory surveys and paperwork for the two bids, what is the expected net profit?

	Profit	Chance of winning bid
Project A	\$175,000	0.50
Project B	\$220,000	0.65

Mean (expected value) of a probability distribution (2)

Answers:

a) Lets denote "win" by W and "not win" by N, and attach subscripts A or B to identify the project. Listed here are the possible outcomes and calculation of the corresponding probabilities.

For instance, $P(W_A N_B) = P(W_A)P(N_B)$, by independence

$$= 0.50x0.35 = 0.175$$

Outcome	Probability
$W_A W_B$	$0.50 \times 0.65 = 0.325$
$W_A N_B$	$0.50 \times 0.35 = 0.175$
$N_A W_B$	$0.50 \times 0.65 = 0.325$
$N_A N_B$	$0.50 \times 0.35 = 0.175$

b) and c) The amounts of profit (X) for the various outcomes are listed below.

Outcome	Profit (\$) X
$W_A W_B$	$175,000 + 220,000 = 395,000$
$W_A N_B$	$175,000$
$N_A W_B$	$220,000$
$N_A N_B$	0

Mean (expected value) of a probability distribution (3)

Answer…cont'ed

Present the probability distribution of X and calculate $E(X)$.

x	$f(x)$	$xf(x)$	
0	0.175	0	
175,000	0.175	30,625	
220,000	0.325	71,500	
395,000	0.325	128,375	
Total		230,500	$= E(X)$

$$Expected\ net\ profit = E(X) - cost = \$230{,}500 - \$2{,}000 = \mathbf{\$228,500}$$

Variance and standard deviation of probability distribution (1)

- The variance of X is abbreviated as $\mathrm{Var}(X)$ and is also denoted by σ^2.

$$\mathrm{Var}(X) = \sum (Deviation)^2 \, x \, (Probability) = \sum (x_i - \mu)^2 f(x_i)$$

- The **standard deviation** of X is the **positive square root** of the variance and is denoted by $sd(X)$ or σ (a Greek lower-case sigma).

Variance and Standard Deviation of X

$$\sigma^2 = \mathrm{Var}(X) = \sum (x_i - \mu)^2 f(x_i)$$

$$\sigma = \mathrm{sd}(X) = +\sqrt{\mathrm{Var}(X)}$$

Alternative Formula for Hand Calculation

$$\sigma^2 = \sum x_i^2 f(x_i) - \mu^2$$

Variance and standard deviation of probability distribution (2)

Example:

Upon examination of the claims records of 280 policy holders over a period of five years, an insurance company makes an empirical determination of the probability distribution of X =number of claims in five years.

a) Calculate the expected value of X.

b) Calculate the standard deviation of X.

x	$f(x)$
0	.315
1	.289
2	.201
3	.114
4	.063
5	.012
6	.006

x	$f(x)$	$xf(x)$	$x^2 f(x)$
0	0.315	0	0
1	0.289	0.289	0.289
2	0.201	0.402	0.804
3	0.114	0.342	1.026
4	0.063	0.252	1.008
5	0.012	0.060	0.300
6	0.006	0.036	0.216
Total		1.381	3.643

Answers:

(a) & (b) The expectation and standard deviation of X are calculated in the left side table.

$E(X)$ or $\mu = \mathbf{1.381}$

$var(X)$ or $\sigma^2 = 3.643 - (1.381)^2 = \mathbf{1.736}$

Standard deviation of X is $\sigma = \sqrt{1.736} = \mathbf{1.318}$

A. Discrete random variable

- **Probability distribution of a discrete random variable**
- **Mean (expected value) and standard deviation of a probability distribution**

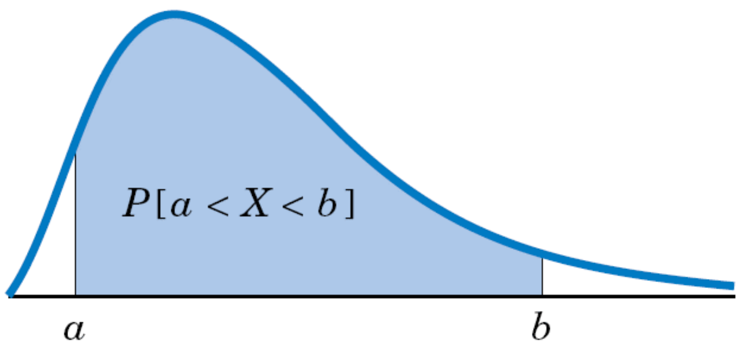
B. Continuous random variable

- **Probability model for a continuous random variable**
- Normal distribution
- Standard normal distribution
- Probability calculations with normal distributions

Probability model for a *continuous* random variable (1)

- The probability distribution of a continuous random variable can ideally *assume any value* in an interval.

> The **probability density function** $f(x)$ describes the distribution of probability for a continuous random variable. It has the properties:
>
> 1. The total area under the probability density curve is 1.
> 2. $P[a \leq X \leq b]$ = area under the probability density curve between a and b.
> 3. $f(x) \geq 0$ for all x.



> With a continuous random variable, the probability that $X = x$ is **always** 0. It is only meaningful to speak about the probability that X lies in an interval.

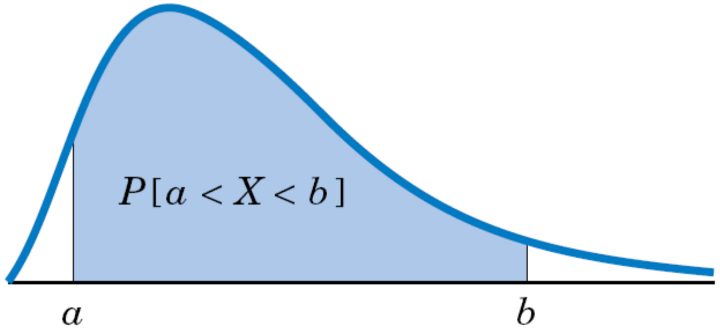
- A single point x, being an interval with a width of 0, supports 0 area, so $P[X = x] = 0$

Probability model for a *continuous* random variable (2)

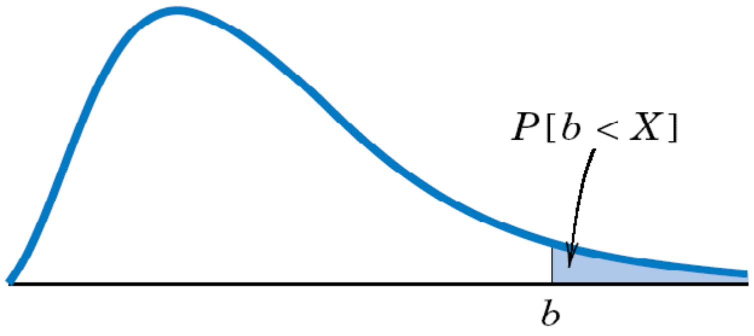
- When determining the probability of an interval a to b, we need not be concerned if either or both endpoints are included in the interval. Since the probabilities of $X = a$ and $X = b$ are both equal to 0,

$$P[a \le X \le b] = P[a < X \le b] = P[a \le X < b] = \mathrm{P}[a < X < b]$$

- $\mathbf{P}[\boldsymbol{a} < \boldsymbol{X} < \boldsymbol{b}]$ =(Area to left of $\boldsymbol{b}$)−(Area to the left of $\boldsymbol{a}$)

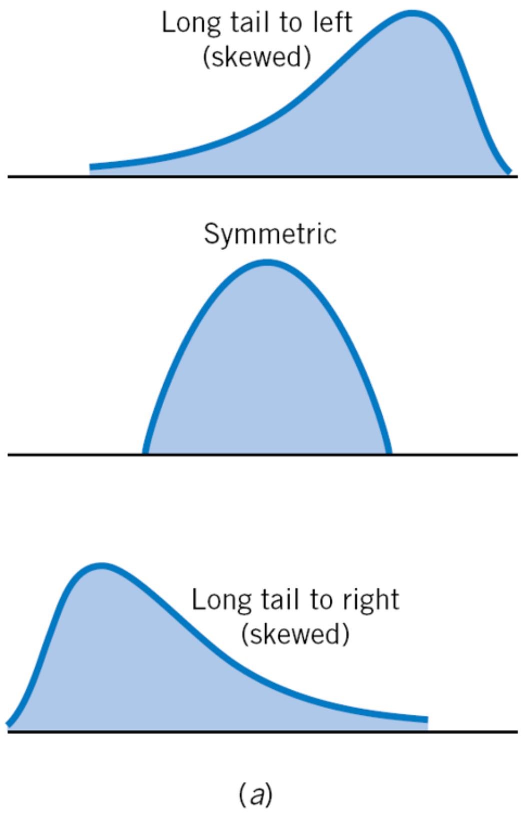


- $\mathbf{P}[\boldsymbol{b} < \boldsymbol{X}] = \mathbf{1} -$ (Area to left of $\boldsymbol{b}$)

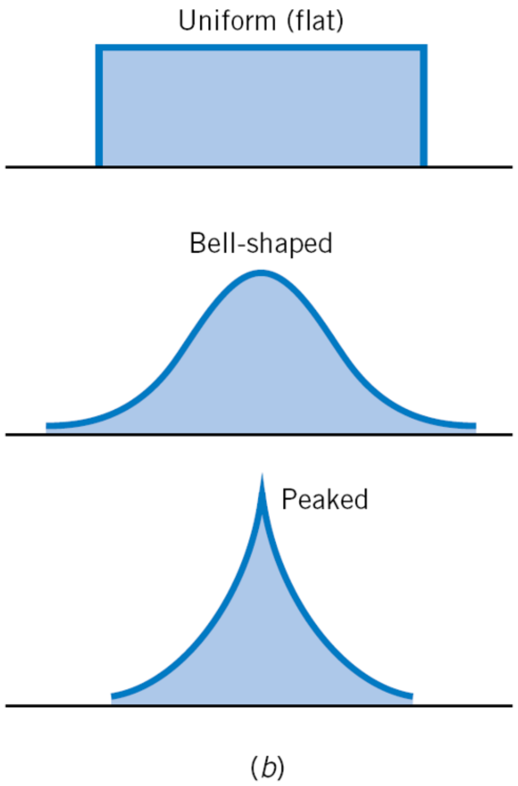


Features of a continuous distribution (1)

(a) Symmetry and deviations from symmetry



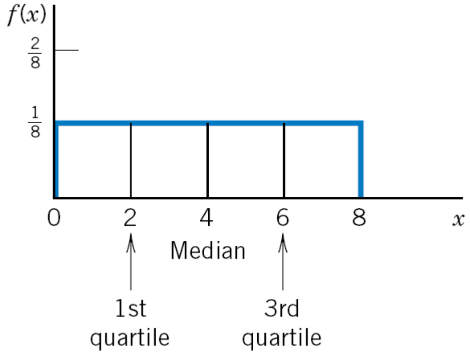
(b) Different peakedness

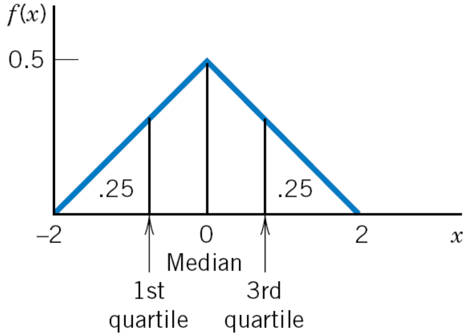


Features of a continuous distribution (2)

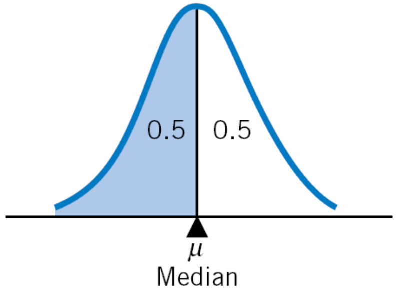
The population **100*p*-th percentile** is an x value that supports area p to its left and $1 - p$ to its right.

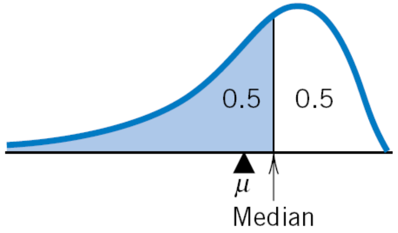
$$\begin{aligned}\text{Lower (first) quartile} &= \text{25th percentile}\\ \text{Second quartile (or median)} &= \text{50th percentile}\\ \text{Upper (third) quartile} &= \text{75th percentile}\end{aligned}$$

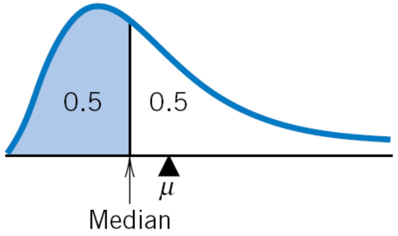




- Mean is the ***balance point*** and median is the point of ***equal division*** of the probability mass.







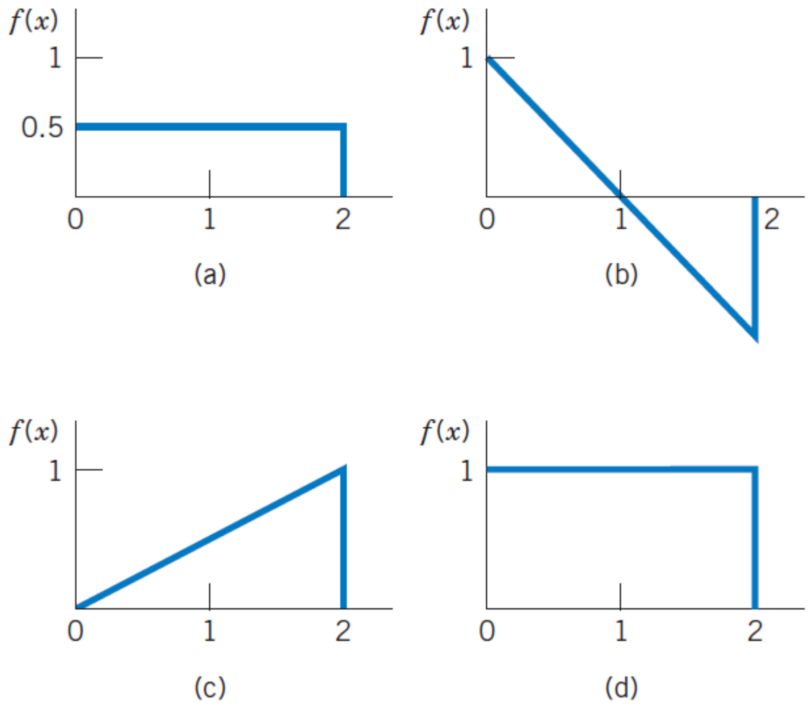
The **standardized variable**

$$Z = \frac{X - \mu}{\sigma} = \frac{\text{Variable} - \text{Mean}}{\text{Standard deviation}}$$

has mean 0 and sd 1.

Features of a continuous distribution (3)

Example: Which of the functions sketched below could be a probability density function for a continuous random variable? Why or why not?



(a) The function is *non-negative* and the area of the *rectangle* is $0.5 \times 2 = 1$. It is a probability density function.

(b) Since $f(x)$ takes *negative values* over the interval from 1 to 2, it is *not* a probability density function.

(c) The function is *non-negative* and the area of the *triangle* is $\frac{1}{2} \times base \times height = \frac{1}{2} \times 2 \times 1 = 1$. It is a probability density function.

(d) The function is *non-negative*, but the area of the rectangle is $1 \times 2 = 2$ so it is *not* a probability density function.

Features of a continuous distribution (4)

Example…cont'ed

Determine the following probabilities from the curve $f(x)$ diagrammed in above *example (a)*.

$a)\quad P[\,0 < X\ < 0.5\,]$

Answer : $P[0\ <\ X\ <\ 0.5]\ =$ Area under $f(x)$ between the points 0 and $0.5 = 0.5 \times 0.5\ =\ \mathbf{0.25}$

$b)\quad P[\,0.5 < X\ < 1\,]$

Answer: $P[0.5\ <\ X\ < 1]\ =\ (1 -\ 0.5) \times 0.5\ =\ \mathbf{0.25}$

$c)\quad P[\,1 < X\ < 2\,]$

Answer: $P[1 <\ X\ <\ 2]\ =\ (2\ - 1) \times 0.5\ =\ \mathbf{0.5}$

$d)\quad P[\,X\ = 1\,]$

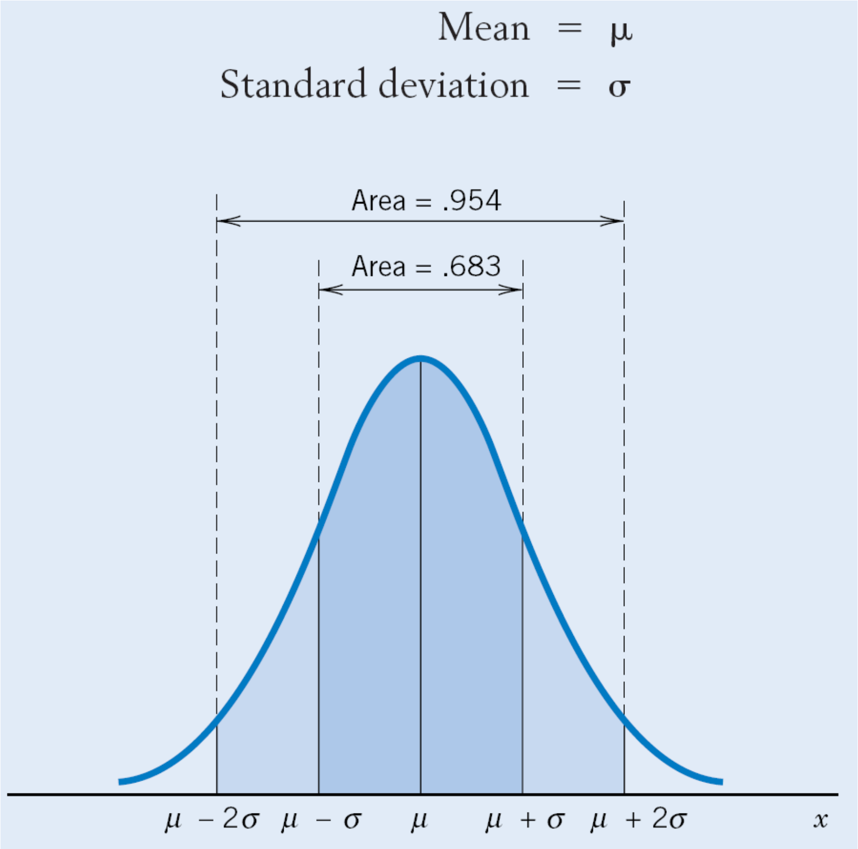
Answer: $\mathrm{P}[\mathrm{X}\ =\ 1]\ =\ \mathbf{0}$

A. Discrete random variable
   - ✓ **Probability distribution of a discrete random variable**
   - ✓ **Mean (expected value) and standard deviation of a probability distribution**

B. Continuous random variable
   - ✓ **Probability model for a continuous random variable**
   - ✓ **Normal distribution**
   - Standard normal distribution
   - Probability calculations with normal distributions

Normal distribution (1)

- The normal probability distribution represents a large family of distribution, each with *unique* specification for the parameters μ and σ^2.
- The normal distribution curve is ***symmetric about its mean μ***, which locates the peak of the bell.



- The probability of the interval extending,
  - *One sd* on each side of the mean:
  $$P[\mu - \sigma < X < \mu + \sigma] = 0.683$$
  - *Two sd* on each side of the mean:
  $$P[\mu - 2\sigma < X < \mu + 2\sigma] = 0.954$$
  - *Three sd* on each side of the mean:
  $$P[\mu - 3\sigma < X < \mu + 3\sigma] = 0.997$$
- The *curve never reaches* 0 for any value of x, but because the tail areas outside $(\mu - 3\sigma, \mu + 3\sigma)$ are very small, we usually terminate the graph at these points.

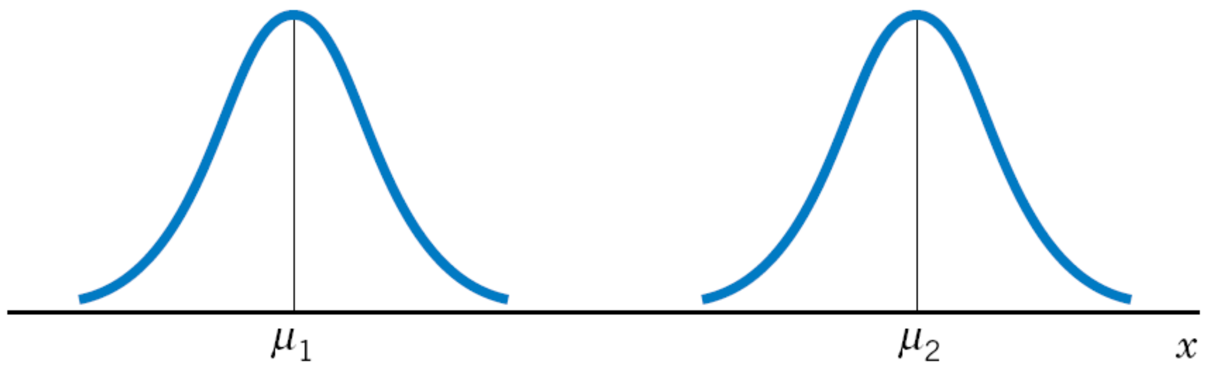
Normal distribution (2)

- If we know the mean and variance, we can define the normal distribution by using the following notation:

Notation

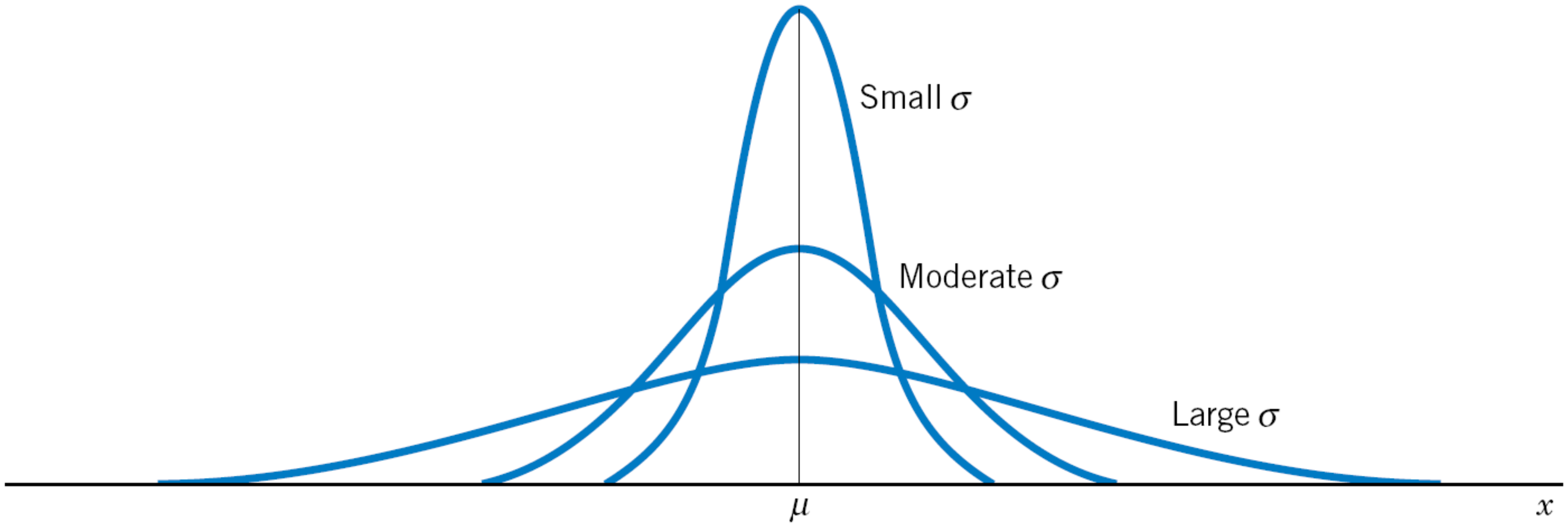
The normal distribution with a mean of μ and a standard deviation of σ is denoted by $N(\mu,\sigma)$.

- ***Change of mean*** from μ_1 to a larger value μ_2 ***merely slides*** the bell-shaped curve *along the axis* until a new center is established at μ_2. There is ***no change in the shape*** of the curve.



Normal distribution (3)

- A *different value for the* σ results in a different maximum height of the curve and *changes* the amount of the area in any fixed interval about μ. The position of the center does not change if only σ is changed.
- Decreasing σ increases the maximum height and the concentration of probability about μ.



A. Discrete random variable

- ✓ **Probability distribution of a discrete random variable**
- ✓ **Mean (expected value) and standard deviation of a probability distribution**

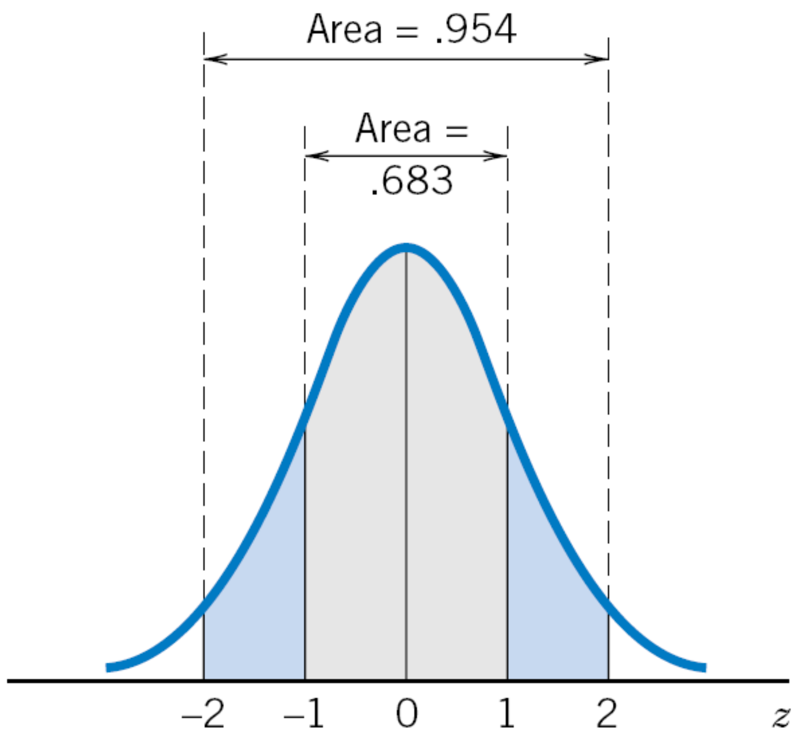
B. Continuous random variable

- ✓ **Probability model for a continuous random variable**
- ✓ **Normal distribution**
- ✓ **Standard normal distribution**
- Probability calculations with normal distributions

Standard normal distribution (1)

- The particular normal distribution that has a mean of 0 and a standard deviation of 1 is called the ***standard normal distribution***.

> The **standard normal distribution** has a bell-shaped density with
>
> $$\text{Mean } \mu = 0$$
> $$\text{Standard deviation } \sigma = 1$$
>
> The standard normal distribution is denoted by $N(0, 1)$.



Standard normal distribution (2)

- The following properties can be observed from the symmetry of the **standard normal curve about 0**.

  1. $P[Z \leq 0\,] = 0.5$

  2. $P[Z \; \leq -z] = 1 - P[Z \leq z] \; = P[\,Z \geq z\,]$

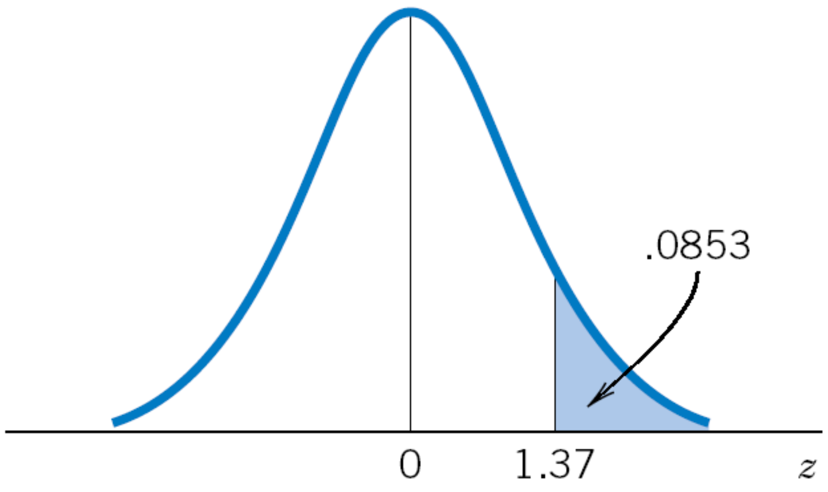
Example:

Find $P[Z \leq 1.37]$ ***and*** $P[Z > 1.37\,]$

Answer:

From the normal table, we see that the

probability or area to the left of 1.37 is $\mathbf{0.9147} = [Z \leq 1.37]$.



Due to [Z > 1.37] is the complement of $P[Z \leq 1.37]$
$[Z > 1.37\,] = 1 - P[Z \leq 1.37] = 1 - 0.9147 = \mathbf{0.0853}$

Standard normal distribution (3)

Example:

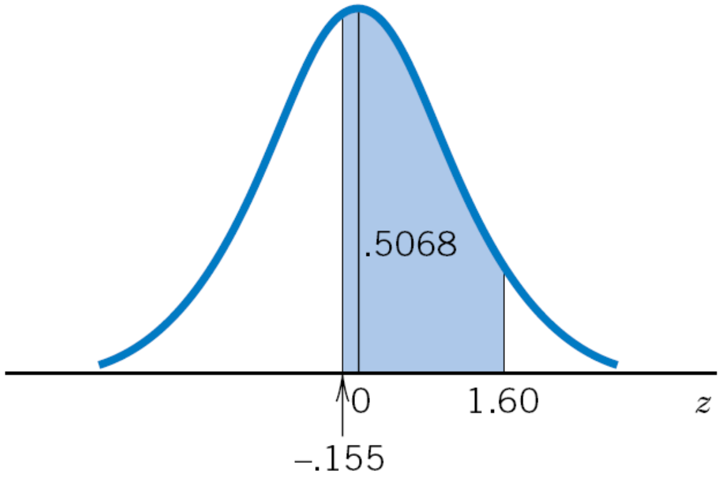
Calculate $P[-0.155 < Z < 1.60]$

Answer:

$P[Z < 1.60]$ =Area to left of 1.60 $= 0.9452$

$[Z \leq -0.155]$ =Area to left of -0.155 $= 0.4384$

$\therefore$ $P[-0.155 < Z < 1.60] = 0.9452 - 0.4384 = \mathbf{0.5068}$



Example:

Find $P[Z < -1.9 \; or \; Z > 2.1]$

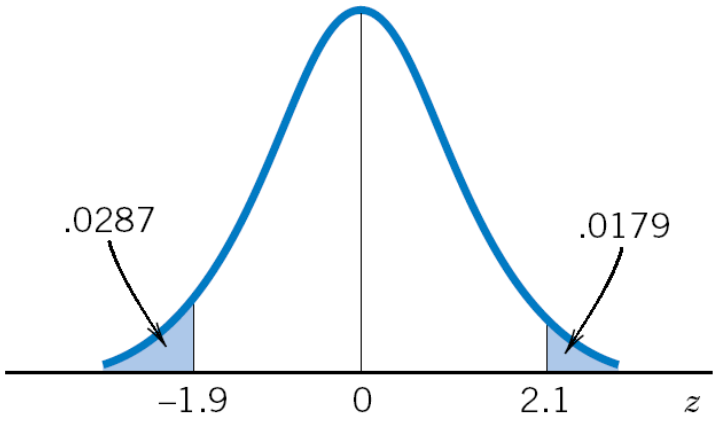
Answer:

The two events $Z < -1.9$ and $Z > 2.1$ are incompatible,

so we add their probabilities:

$$P[Z < -1.9 \; or \; Z > 2.1] = P[Z < -1.9] + P[Z > 2.1]$$

$$= 0.0287 + 0.0179 = \mathbf{0.0466}$$



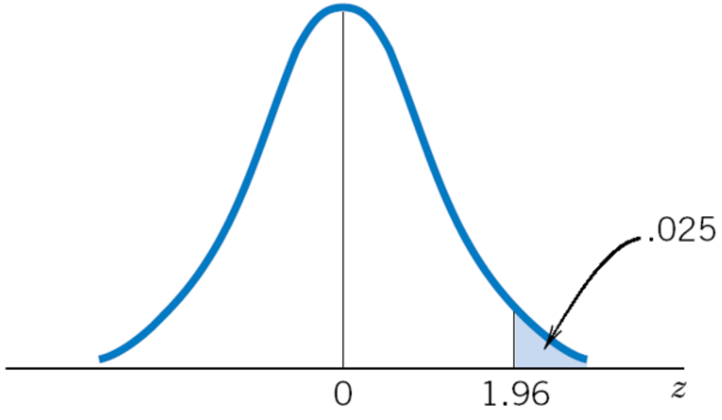
Standard normal distribution (4)

Example:

Locate the value of z that satisfies $P[Z > z] = 0.025$

Answer:

The total area is 1, the area to the left of z must be $1 - 0.0250 = 0.9750$. The marginal value with the tabular entry 0.9750 is $z = 1.96$



$P[Z > 1.96] = .025.$

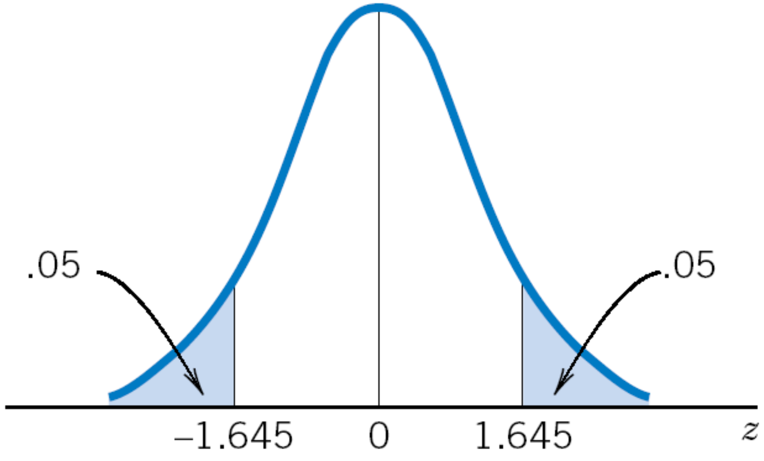
Example: Obtain the value of z for which $P[-z \leq Z \leq z] = 0.9$

Answer:

We observe from the symmetry of the curve that

$$P[Z < -z] = P[Z > z] = 0.05$$

From the normal table $z = 1.65$



A. Discrete random variable

- **Probability distribution of a discrete random variable**
- **Mean (expected value) and standard deviation of a probability distribution**

B. Continuous random variable

- **Probability model for a continuous random variable**
- **Normal distribution**
- **Standard normal distribution**
- **Probability calculations with normal distributions**

Probability calculations with normal distributions

- To find the probability that X lies in a given interval, convert the interval to the z scale and then calculate the probability by using the standard normal table.

If X is distributed as $N(\mu, \sigma)$, then the standardized variable

$$Z = \frac{X - \mu}{\sigma}$$

has the standard normal distribution.

If X is distributed as $N(\mu, \sigma)$, then

$$P[a \leq X \leq b] = P\left[\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right]$$

where Z has the standard normal distribution.

Example:

Given that X has the normal distribution $N(60, 4)$, find $P(55 \leq X \leq 63)$**.**

Answer: Here the standardized variable is $\frac{X-60}{4}$

$$x = 55 \text{ gives } z = \frac{55-60}{4} = -1.25,\ P(z \leq -1.25) = 0.1056$$

$$x = 63 \text{ gives } z = \frac{63-60}{4} = 0.75,\ \ P(z \leq 0.75) = 0.7734$$

so, the required probability is $0.7734 - 0.1056 = \mathbf{0.6678}$