Announcements

Topics:

In the Probability and Statistics module:

- **Section 10:** The Binomial Distribution
- **Section 13:** Continuous Random Variables
- **Section 14:** The Normal Distribution

To Do:

- Read sections 10, 13, and 14 in the "Probability and Statistics" module
- Work on Assignments and Suggested Practice Problems assigned on the webpage under the SCHEDULE + HOMEWORK link

Continuous Random Variables

Definition:

A random variable that takes on a *continuum* of values is called a continuous random variable.

Continuous Random Variables

Example:

Distributions of Lengths of Boa Constrictors

The boa constrictor is a large species of snake that can grow to anywhere between 1 m and 4 m in length.

Let L be the continuous random variable that measures the length of a snake.

$$L : S \rightarrow [1,4]$$

Continuous Random Variables

The lengths of 500 boas are recorded below:

Table 13.1

Length range (m)	Frequency	Relative frequency
$[1, 1.5)$	20	0.04
$[1.5, 2)$	58	0.116
$[2, 2.5)$	122	0.244
$[2.5, 3)$	180	0.36
$[3, 3.5)$	86	0.172
$[3.5, 4)$	34	0.068

Note: relative frequency = frequency/500 = probability

Continuous Random Variables

<u>Histogram for Probability Mass</u>:

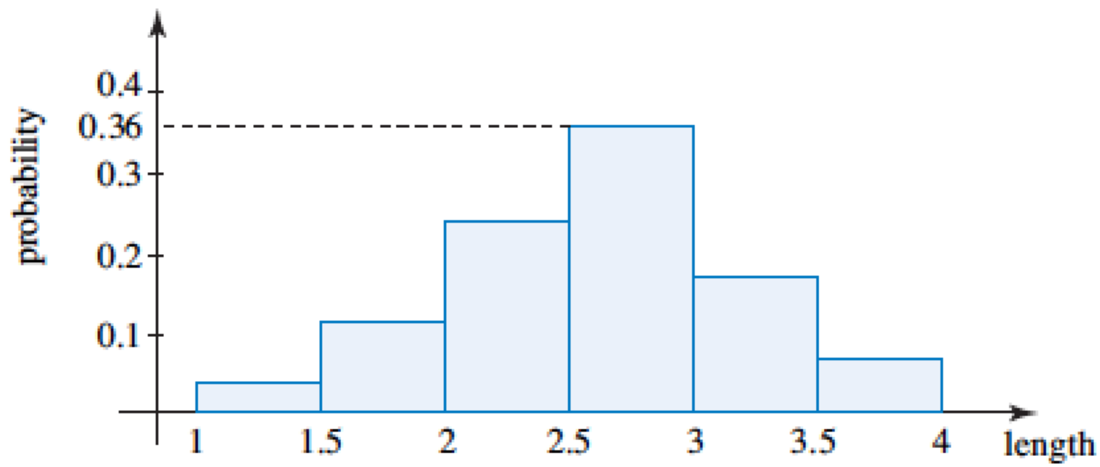


FIGURE 13.1

Histogram: the heights represent the probability

The probability that a randomly selected boa is between 2.5 m and 3 m in length is the <u>height</u> of the rectangle over [2.5, 3), i.e., 0.36.

Continuous Random Variables

To draw a histogram representing probability **density**, we re-label the vertical axis so that the probability that *L* belongs to an interval is the <u>area</u> of the rectangle above that interval.

Continuous Random Variables

Note:

$$\frac{\text{probability } \textbf{mass}}{\textbf{length} \text{ of interval}} = \text{probability } \textbf{density}$$

Continuous Random Variables

For example, consider the interval [2.5, 3). The probability that L falls in this range is 0.36.

Now, we want this value to be the <u>area</u> of the rectangle over [2.5, 3), so

probability density (height) = 0.36/(3-2.5) = 0.72

Continuous Random Variables

<u>Histogram for Probability Density</u>:

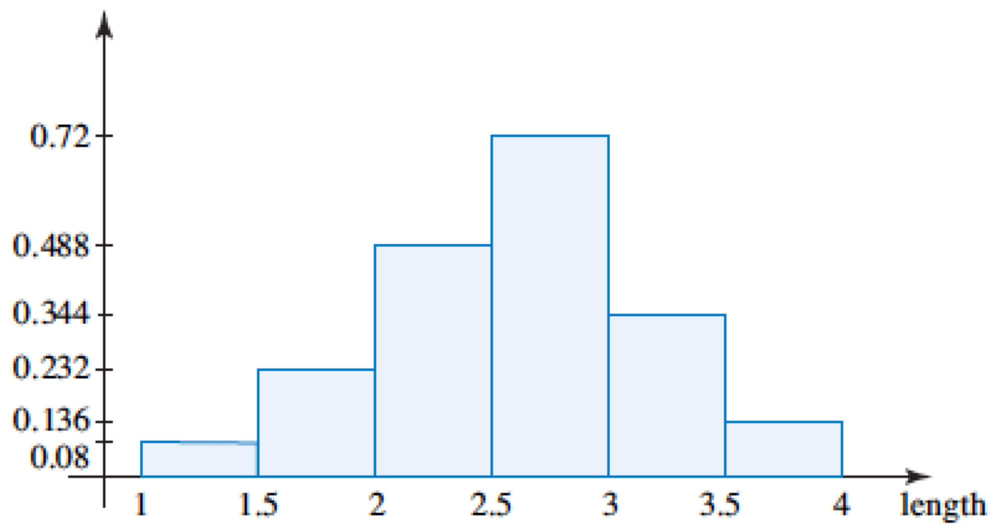


FIGURE 13.2

Histogram: the areas represent the probability

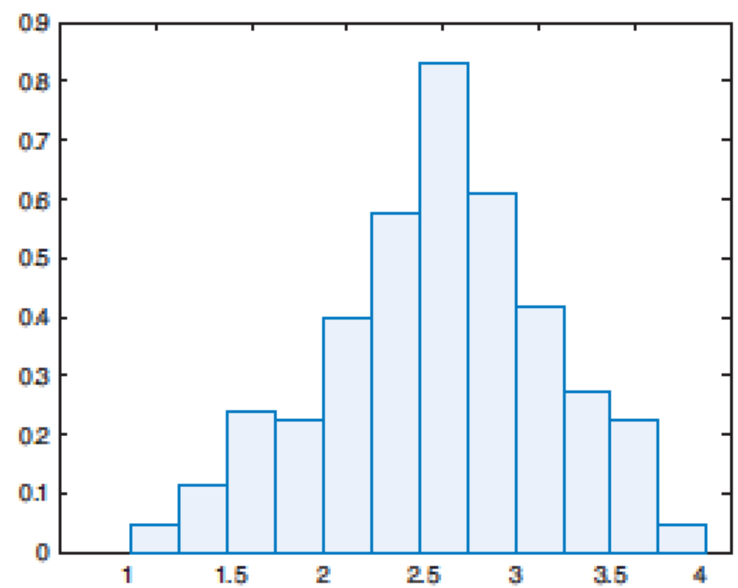
The probability that a randomly selected boa is between 2.5 m and 3 m in length is the <u>area</u> of the rectangle above [2.5, 3), i.e. 0.36.

Continuous Random Variables

To get a more precise probability mass (or density) function, we divide [1,4] into smaller subintervals:

Table 13.2

Length range (m)	Frequency	Relative frequency
$[1, 1.25)$	6	0.012
$[1.25, 1.5)$	14	0.028
$[1.5, 1.75)$	30	0.06
$[1.75, 2)$	28	0.056
$[2, 2.25)$	50	0.1
$[2.25, 2.5)$	72	0.144
$[2.5, 2.75)$	104	0.208
$[2.75, 3)$	76	0.152
$[3, 3.25)$	52	0.104
$[3.25, 3.5)$	34	0.068
$[3.5, 3.75)$	28	0.056
$[3.75, 4)$	6	0.012



Continuous Random Variables

As we continue to increase the number of subintervals, we obtain a more and more refined histogram.

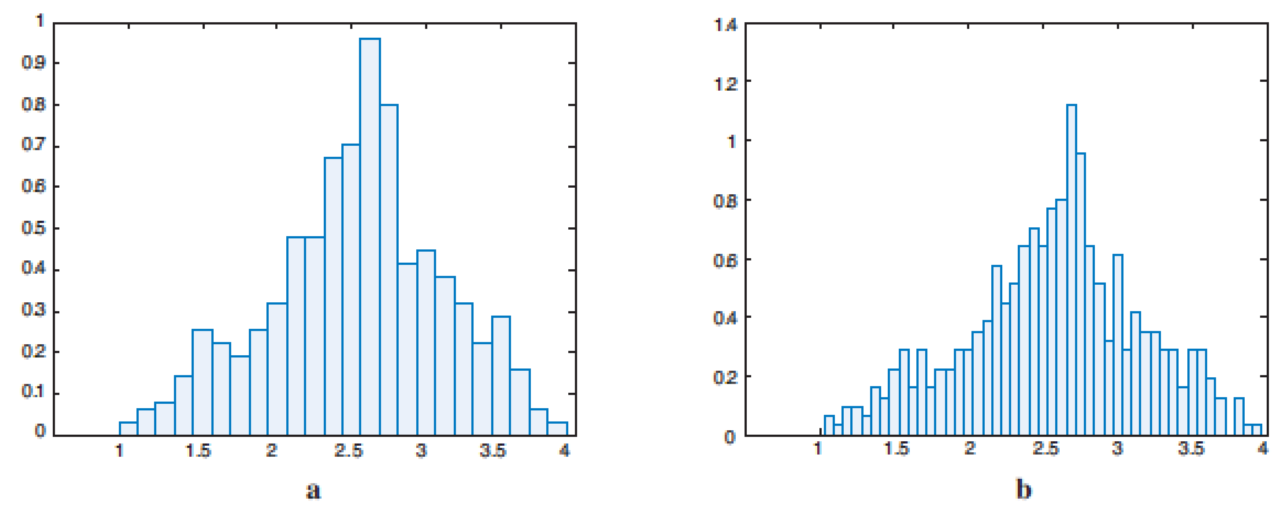


FIGURE 13.4

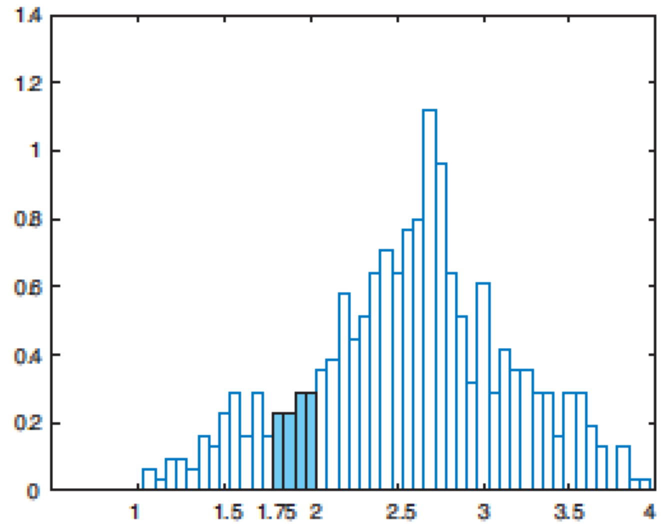
Histograms based on 24 and 48 subintervals

Continuous Random Variables

Riemann Sum:

FIGURE 13.5

Probability of boa length between 1.75 m and 2 m



The probability that a randomly chosen boa is between 1.75 m and 2 m in length is the sum of the areas of the rectangles over the interval [1.75, 2).

Continuous Random Variables

To obtain the probability density **function**, we let the length of the intervals approach 0 and the number of rectangles approach ∞.

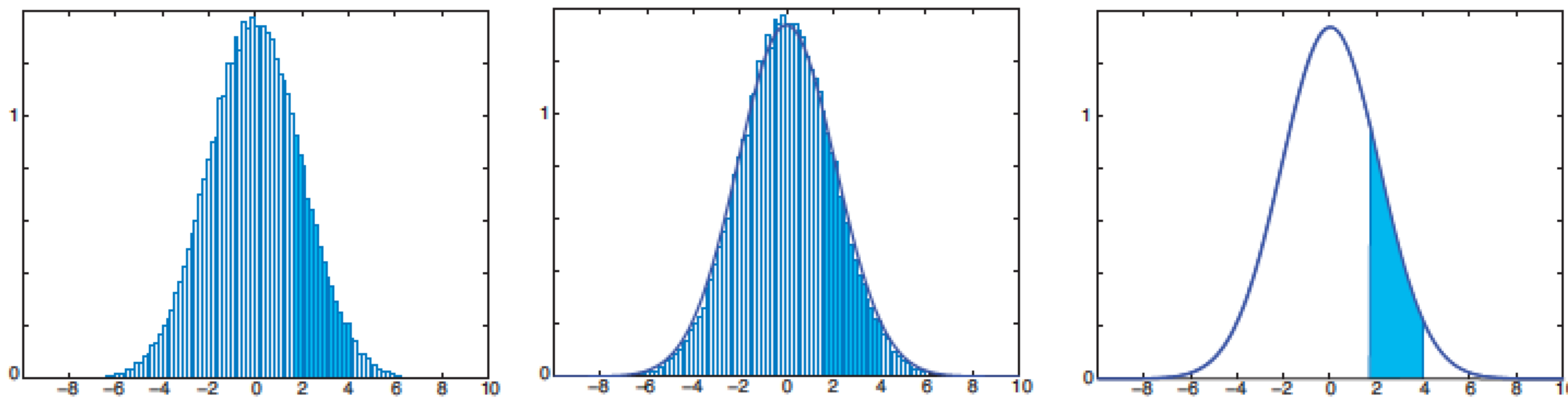


FIGURE 13.6

From a histogram to a density function

Probability Density Functions

Definition: Defining Properties of a PDF

Assume that the interval *I* represents the range of a continuous random variable *X*. A function *f(x)* can be a probability density function if

(1) $f(x) \geq 0 \quad \text{for all } x \in I.$

(2) $\int_I f(x)dx = 1.$

Probability Density Functions

Example:

Show that $f(x) = \dfrac{2}{\pi(1+x^2)}$

could be a probability density function for some continuous random variable on $[0, \infty)$.

Calculating Probabilities

For a continuous random variable, we calculate the probability that a random variable belongs to an *interval* of real numbers.

The probability that an outcome *X* is between *a* and *b* is the area under the graph of *f(x)* on [*a*,*b*]:

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

Calculating Probabilities

The probability that an outcome is *equal* to a particular value is zero.

$$P(a \le X \le a) = \int_a^a f(x)dx = 0$$

For this reason, **including or excluding the endpoints of an interval does not affect the probability**, i.e.,

$$P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a < X < b)$$

Calculating Probabilities

Example #32:
The distance between a seed and the plant it came from is modelled by the density function

$$f(x) = \frac{2}{\pi(1+x^2)}$$

where *x* represents the distance (in metres), $x \in [0,\infty)$.

What is the probability that a seed will be found farther than 5 m from the plant?

Cumulative Distribution Function

Definition:

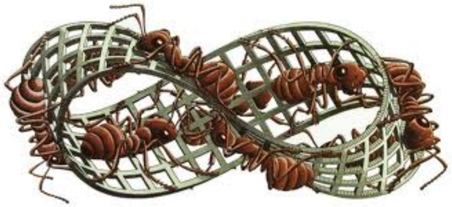
Suppose that *f(x)* is a probability density function defined on an interval [*a,b*]. The function *F(x)* defined by

$$F(x) = P(X \le x) = \int_a^x f(t)\,dt$$

for all *x* in [*a,b*] is called a cumulative distribution function of *f(x)*.

Cumulative Distribution Function

Example #30 (modified):
Suppose that the lifetime of an insect is given by the probability density function



$$f(t) = 0.2e^{-0.2t}$$

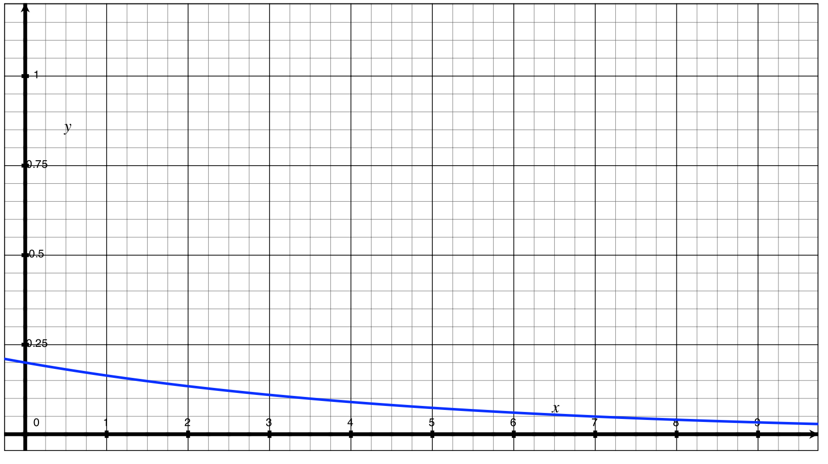
where t is measured in days, $t \in [0,\infty)$.

(a) Determine the corresponding cumulative distribution function, *F(t)*.
(b) Find the probability that the insect will live between 5-7 days.

Cumulative Distribution Function

Example #30 (modified):

probability density function f(t)



cumulative distribution function F(t)

Cumulative Distribution Function

Properties of the CDF:

Assume that *f* is a probability density function, defined and continuous on an interval *[a,b]*. The left end *a* could be a real number or negative infinity; the right end *b* could be a real number or infinity. Denote by *F* the associated cumulative distribution function. Then

(1) $0 \le F(x) \le 1$ for all $x \in [a,b]$.

(2) $F(x)$ is continuous and non-decreasing.

(3) $\lim_{x \to a} F(x) = 0$ and $\lim_{x \to b} F(x) = 1$.

The Mean and the Variance

Definition:
Let *X* be a continuous random variable with probability density function *f(x)*, defined on an interval *[a,b]*.

The mean (or the expected value) of *X* is given by

$$\mu = E(X) = \int_a^b x\, f(x)\, dx$$

The variance of *X* is

$$\mathrm{var}(X) = E\left[(X-\mu)^2\right] = \int_a^b (x-\mu)^2 f(x)\, dx$$

The Mean and the Variance

Example #24:
Consider the continuous random variable X given by the probability density function

$$f(x) = 0.3 + 0.2x \quad \text{for} \quad 0 \leq x \leq 2.$$

Find the probability that the values of X are at least one standard deviation above the mean.

The Normal Distribution

The normal distribution is the most important continuous distribution as it can be used to model many phenomena in a variety of fields.

Many measurements for large sample sizes are said to be ‘normally distributed’.

For example, heights of trees, IQ scores, and duration of pregnancy are all normally distributed measurements.

The Normal Distribution

Definition:

A continuous random variable *X* has a normal distribution (or is distributed normally) with mean μ and variance σ^2, denoted by $X \sim N(\mu, \sigma^2)$, if its probability density function is

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

where $x \in (-\infty, \infty)$.

The Normal Distribution

The graph of the probability density function of the normal distribution (also known as the Gaussian distribution) is a bell-shaped curve.

Properties of the Normal Distribution Density Function

Theorem:
The probability density function *f(x)* of the normal distribution satisfies the following properties:

(a) f(x) is symmetric with respect to the vertical line $x = \mu$.

(b) *f(x)* is increasing for $x < \mu$ and decreasing for $x > \mu$. It has a local (also global) maximum value $1/\sigma\sqrt{2\pi}$ at $x = \mu$.

(c) The inflection points of *f(x)* are $x = \mu \pm \sigma$.

(d) $\lim_{x \to -\infty} f(x) = \lim_{x \to \infty} f(x) = 0$

Calculating Probabilities

If X is a normally distributed continuous random variable with mean μ and variance σ^2, then

$$P(a \leq X \leq b) = \int_a^b f(x)dx = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Calculating Probabilities

This integral cannot be evaluated without estimation techniques, such as using a Taylor polynomial to approximate $f(x)$.

To evaluate this integral, we reduce a general normal distribution to a special normal distribution, called the standard normal distribution, and then use tables of estimated values.

Standard Normal Distribution

Definition:

The standard normal distribution is the normal distribution with mean 0 and variance 1; in symbols, it is *N(0,1)*. Its probability density function is given by

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

for all $x \in (-\infty, \infty)$.

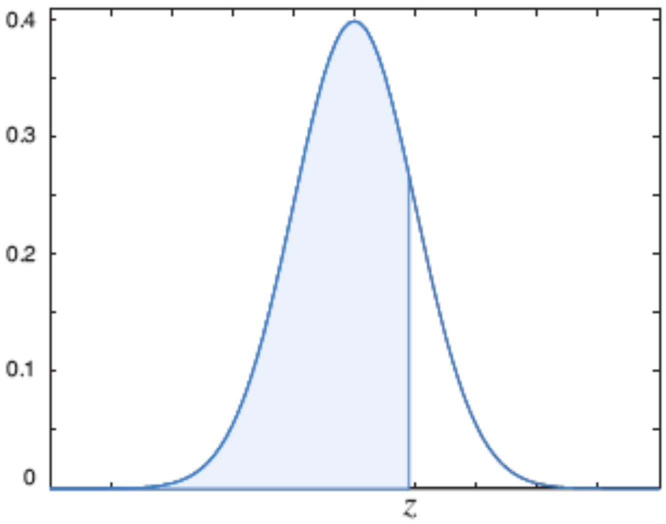
Standard Normal Distribution

We use the symbol Z to denote the continuous random variable that has the standard normal distribution; i.e., $Z \sim N(0,1)$.

Standard Normal Distribution

The cumulative distribution function of *Z* is given by

$$F(z) = \int_{-\infty}^{z} f(x)dx = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$



$*$ Note: $F(-z) = 1 - F(z)$

z	$F(z)$	z	$F(z)$	z	$F(z)$	z	$F(z)$
0	0.500000	1	0.841345	2	0.977250	3	0.998650
0.05	0.519938	1.05	0.853141	2.05	0.979818	3.05	0.998856
0.1	0.539828	1.1	0.864334	2.1	0.982136	3.1	0.999032
0.15	0.559618	1.15	0.874928	2.15	0.984222	3.15	0.999184
0.2	0.579260	1.2	0.884930	2.2	0.986097	3.2	0.999313
0.25	0.598706	1.25	0.894350	2.25	0.987776	3.25	0.999423
0.3	0.617911	1.3	0.903200	2.3	0.989276	3.3	0.999517
0.35	0.636831	1.35	0.911492	2.35	0.990613	3.35	0.999596
0.4	0.655422	1.4	0.919243	2.4	0.991802	3.4	0.999663
0.45	0.673645	1.45	0.926471	2.45	0.992857	3.45	0.999720
0.5	0.691462	1.5	0.933193	2.5	0.993790	3.5	0.999767
0.55	0.708840	1.55	0.939429	2.55	0.994614	3.55	0.999807
0.6	0.725747	1.6	0.945201	2.6	0.995339	3.6	0.999840
0.65	0.742154	1.65	0.950529	2.65	0.995975	3.65	0.999869
0.7	0.758036	1.7	0.955435	2.7	0.996533	3.7	0.999892
0.75	0.773373	1.75	0.959941	2.75	0.997020	3.75	0.999912
0.8	0.788145	1.8	0.964070	2.8	0.997445	3.8	0.999928
0.85	0.802337	1.85	0.967843	2.85	0.997814	3.85	0.999941
0.9	0.815940	1.9	0.971283	2.9	0.998134	3.9	0.999952
0.95	0.828944	1.95	0.974412	2.95	0.998411	3.95	0.999961
						4	0.999968

The Normal and the Standard Normal Distributions

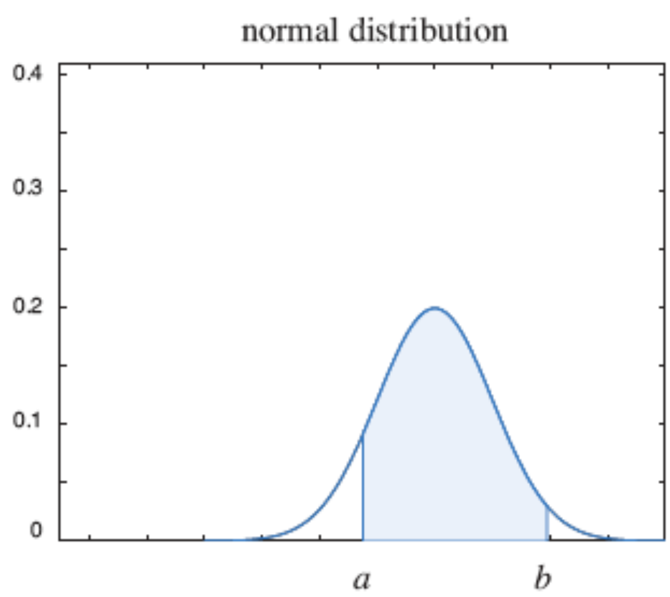
Theorem:

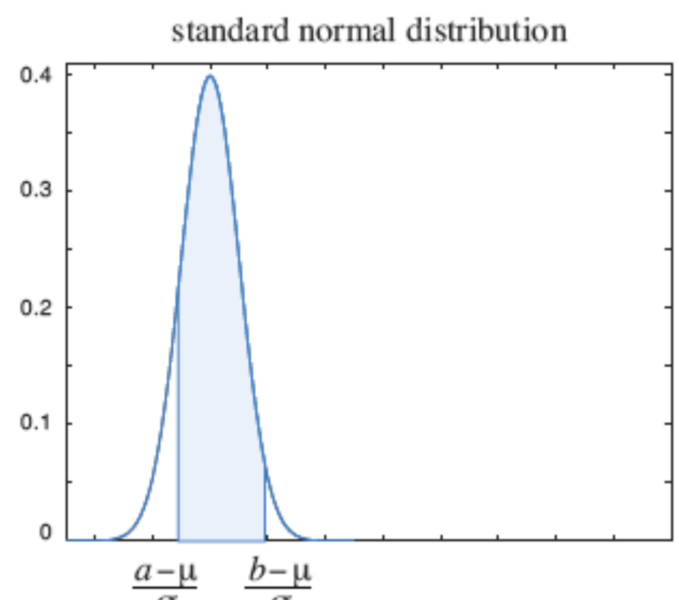
Assume that $X \sim N(\mu, \sigma^2)$. The random variable $Z = (X - \mu)/\sigma$ has the standard normal distribution, i.e., $Z \sim N(0,1)$.

So then $P(a \le X \le b) = P(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma})$.

The Normal and the Standard Normal Distributions

In words, the area under the normal distribution density function between *a* and *b* is equal to the area under the standard normal density function between $(a-\mu)/\sigma$ and $(b-\mu)/\sigma$.





The Normal and the Standard Normal Distributions

Example #10:

Let $X \sim N(-2,4)$; find $P(-3 \le X \le 1)$.

Example #30:

Let $X \sim N(2,144)$; find a value of x that satisfies $P(X > x) = 0.3$.

Application

Example:

Intelligence quotient (IQ) scores are distributed normally with mean 100 and standard deviation 15.

(a) What percentage of the population has an IQ score between 85 and 115?

(b) What percentage of the population has an IQ above 140?

(c) What IQ score do 90% of people fall under?

68-95-99.7 Rule

If X is a continuous random variable distributed normally with mean μ and standard deviation σ, then

$$P(\mu - \sigma \le X \le \mu + \sigma) = 0.683$$

$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = 0.955$$

$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 0.997$$