Chapter 3C

Continuous Random Variables

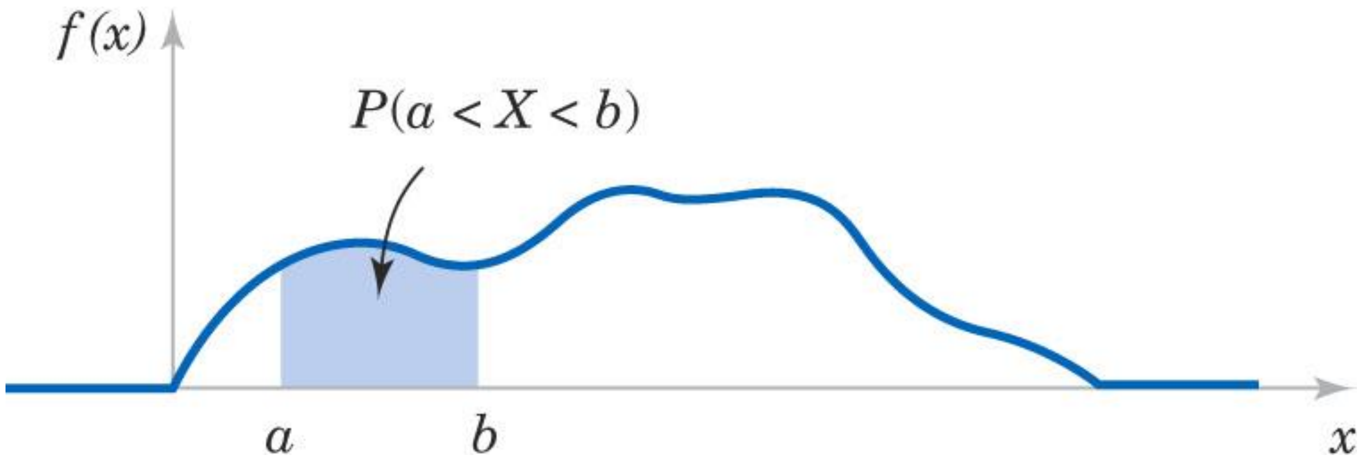
Continuous vs. Discrete RVs

- A discrete random variable is a *rv* with a finite (or countably infinite) range. They are usually integer counts.
  - Number of scratches on a surface.
  - Proportion of defective parts among 100 tested.
  - Number of transmitted bits received in error.

- A continuous random variable is a *rv* with an interval (either finite or infinite) of real numbers for its range. Its precision depends on the measuring instrument.
  - Electrical current and voltage.
  - Physical measurements, e.g., length, weight, time, temperature, pressure.

Probability Density Function

- Density functions, in contrast to mass functions, distribute probability continuously along an interval.
- A probability density function $f(x)$ describes the probability distribution of a continuous random variable.



Area under a curve

For a continuous, non-negative function, the mean value theorem of calculus says that $\int_a^b f(x)\,dx = (b-a)\,f(y)$ for some y in $[a,b]$.

If $b-a=0$, then $P(X=x)=0$.

This implies that if X is a continuous random variable, for any x_1 and x_2,

$$P(x_1 \le X \le x_2) = P(x_1 < X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X < x_2) \qquad (4\text{-}2)$$

From another perspective:

As x_1 approaches x_2, the area or probability becomes smaller and smaller.

As x_1 actually becomes x_2, the area or probability becomes zero.

Probability Density Function (PDF)

For a continuous random variable X,

a probability density function is a function such that:

(1) $f(x) \geq 0$ means that the function is always non-negative.

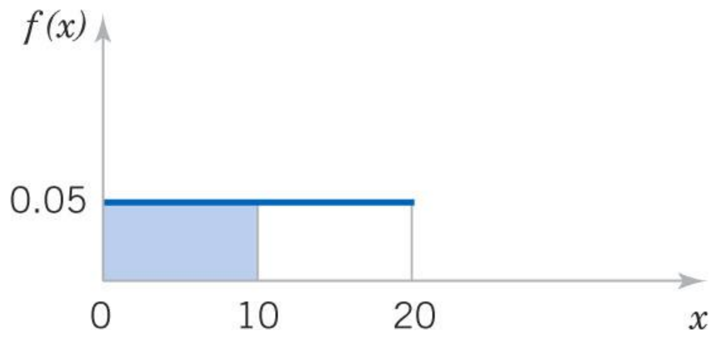
(2) $\int_{-\infty}^{\infty} f(x)dx = 1$

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

(4) $f(x) = 0$ means there is no area exactly at x.

Area under a curve Example 1

- Let the continuous random variable X denote the current measured in a thin copper wire in milliamperes (mA). Assume that the range of X is $0 \le x \le 20$ and $f(x) = 0.05$. What is the probability that a current is less than 10mA?



$$P(X < 10) = \int_0^{10} 0.05dx = 0.5$$

Another example,

$$P(5 < X < 20) = \int_5^{20} 0.05dx = 0.75$$

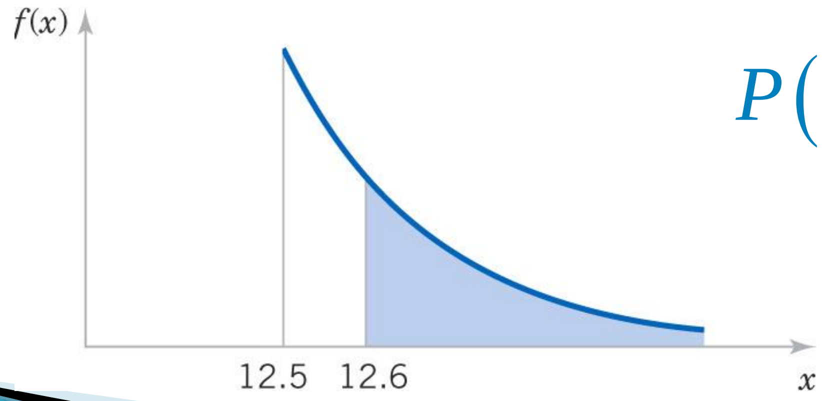
Note: This is actually the Uniform distribution

Area example 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 mm. Random disturbances to the process result in larger diameters. Historical data shows that the distribution of X can be modeled by:

$$f(x) = 20e^{-20(x-12.5)} \text{ where } x \geq 12.5 \text{ mm.}$$

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



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

Cumulative distribution functions

The cumulative distribution function

of a continuous random variable X is,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du \quad \text{for} \quad -\infty < x < \infty$$

PDFs vs. CDFs

- The probability density function (PDF) is the derivative of the cumulative distribution function (CDF).

$$\text{Given } F(x),\ f(x) = \frac{dF(x)}{dx} \text{ as long as the derivative exists.}$$

- Conversely, the cumulative distribution function (CDF) is the integral of the probability density function (PDF).

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\,du \quad \text{for} \quad -\infty < x < \infty \qquad \text{(4-3)}$$

CDF to PDF

- The time until a chemical reaction is complete (in milliseconds, ms) is approximated by this CDF:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-0.01x} & \text{for x} \geq 0 \end{cases}$$

- What is the PDF?

$$f(x) = \frac{dF(x)}{dx} = \frac{d}{dx}\begin{cases} 0 \\ 1 - e^{-0.01x} \end{cases} = \begin{cases} 0 & \text{for } x < 0 \\ 0.01e^{-0.01x} & \text{for x} \geq 0 \end{cases}$$

- What proportion of reactions is complete within 200 ms?
  - Use the CDF:
  - Confirm with the PDF

$$P(X < 200) = F(200) = 1 - e^{-2} = 0.8647$$

Mean and Variance of a CRV

Suppose X is a continuous random variable with probability density function $f(x)$. The mean or expected value of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)\,dx \qquad (4\text{-}4)$$

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

Expected value of a power of a random variable

If X is a continuous random variable

with a probability density function $f(x)$,

$$E\left[h(x)\right]=\int_{-\infty}^{\infty} h(x)f(x)\,dx$$

Allows us to use the same variance formula as for discrete:

$$V[X]=E[X^2]-(E[X])^2$$

Mean and Variance Example

- For the copper wire current example, the PDF is:

$f(x) = 0.05$ for $0 \le x \le 20$.

Find the mean and variance.

$$E(X) = \int_0^{20} x \cdot f(x)\,dx = \left.\frac{0.05x^2}{2}\right|_0^{20} = 10$$

$$V(X) = \int_0^{20} (x-10)^2 \cdot f(x)\,dx = \left.\frac{0.05(x-10)^3}{3}\right|_0^{20} = 33.33$$

Try $V[X]=E[X^2]-(E[X])^2$

Mean and Variance Example

- For the drilling operation example, find the mean and variance of *X* using integration by parts. Recall that $f(x) = 20e^{-20(x-12.5)}dx$ for x ≥ 12.5.

$$E(X) = \int_{12.5}^{\infty} x f(x)\,dx = \int_{12.5}^{\infty} x 20e^{-20(x-12.5)}dx \text{ and integrating by parts}$$

$$= -xe^{-20(x-12.5)} - \left.\frac{e^{-20(x-12.5)}}{20}\right|_{12.5}^{\infty} = 12.5 + 0.05 = 12.55 \text{ mm}$$

$$V(X) = \int_{12.5}^{\infty} (x-12.55)^2 f(x)\,dx = 0.0025 \text{ mm}^2 \text{ and } \sigma = 0.05 \text{ mm}$$

Try $V[X]=E[X^2]-(E[X])^2$

Some Special Continuous Distributions

- Sometimes we can use specific distributions to model situations

- Some important distributions we will cover are:
  - Uniform
  - Exponential
  - Normal

Comparing Distributions

- In a previous example the **Poisson** distribution defined a random variable as the number of flaws along a length of wire (flaws per mm).

- The **exponential** distribution defines a random variable as the interval between flaws (mm's between flaws – the inverse).

Exponential Distribution

- The random variable X that equals the distance between successive events of a Poisson process with mean number of events $\lambda > 0$ per unit interval is an exponential random variable with parameter λ.

$$f(x) = \lambda e^{-\lambda x} \text{ and } F(x) = 1 - e^{-\lambda x}$$

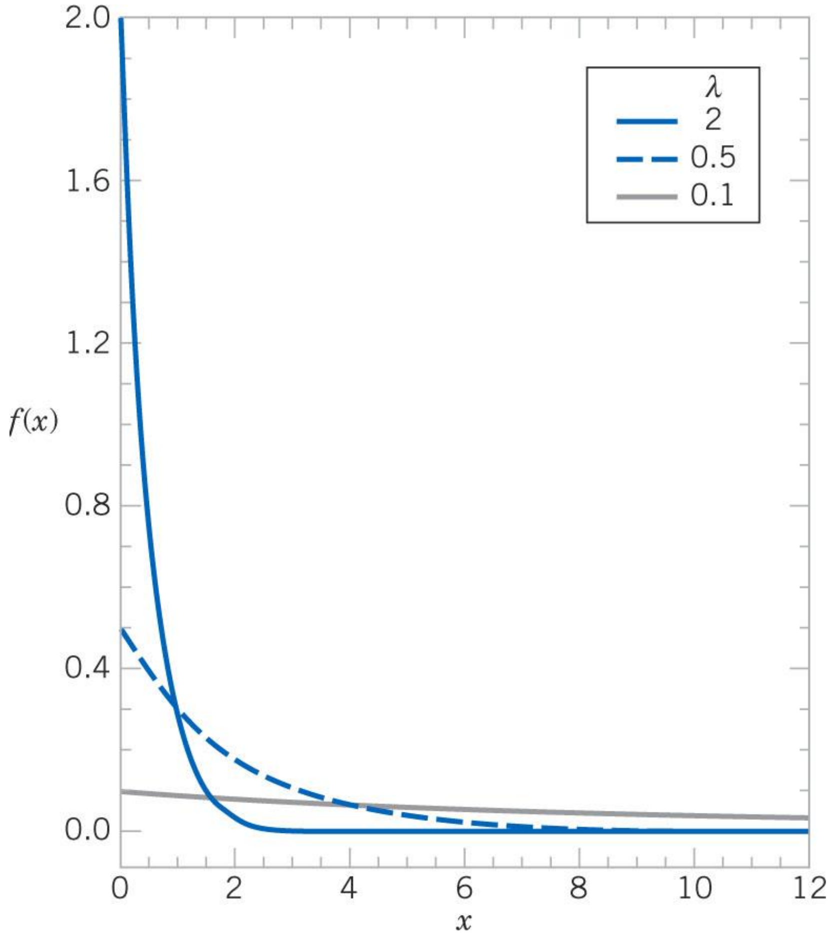
- Measures:

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

Exponential Distribution graph

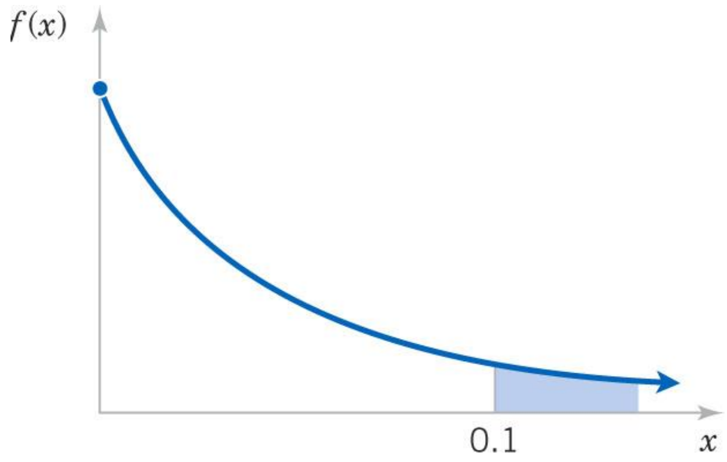
- The y-intercept of the exponential probability density function is λ.

- The random variable is non-negative and extends to infinity.



Exponential Example 1

- In a large corporate computer network, user log-ons to the system can be modeled as a Poisson process with a mean of 25 log-ons per hour. What is the probability that there are no log-ons in the next 6 minutes (0.1 hour)? Let X denote the time in hours from the start of the interval until the first log-on.



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

By integrating the PDF

$$= 1 - F\left(0.1\right) = 0.082$$

Or use the CDF

Exponential Example 2

- Continuing with the last example, what is the probability that the time until the next log-on is between 2 and 3 minutes (0.033 & 0.05 hours)?

$$P\left(0.033 < X < 0.05\right) = \int_{0.033}^{0.05} 25e^{-25x}$$

$$= -e^{=25x}\Big|_{0.033}^{0.05} = 0.152$$ By integrating the PDF

$$= F\left(0.05\right) - F\left(0.033\right) = 0.152$$ Or use the CDF

Exponential Example 2

- Continuing, what is the interval of time such that the probability that no log-on occurs during the interval is 0.90?

$$P\left(X > x\right) = e^{-25x} = 0.90, \quad -25x = \ln\left(0.90\right)$$

Solving the CDF for x

$$x = \frac{-0.10536}{-25} = 0.00421 \text{ hour} = 0.253 \text{ minute}$$

- What is the mean and standard deviation of the time until the next log-in?

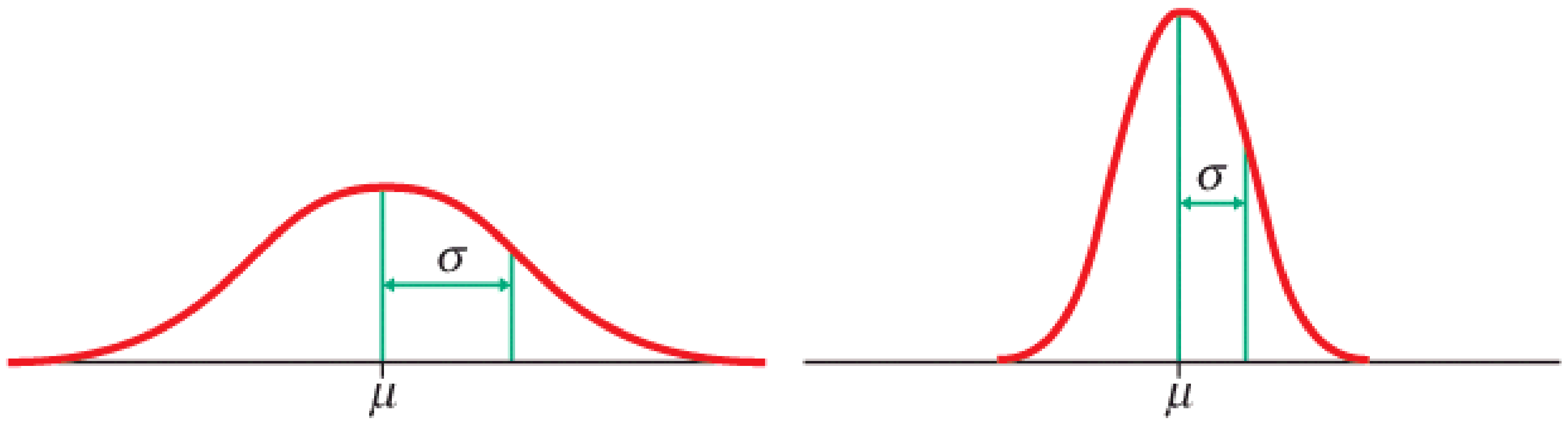
$$\mu = \frac{1}{\lambda} = \frac{1}{25} = 0.04 \text{ hour } = 2.4 \text{ minutes}$$

$$\sigma = \frac{1}{\lambda} = \frac{1}{25} = 0.04 \text{ hour } = 2.4 \text{ minutes}$$

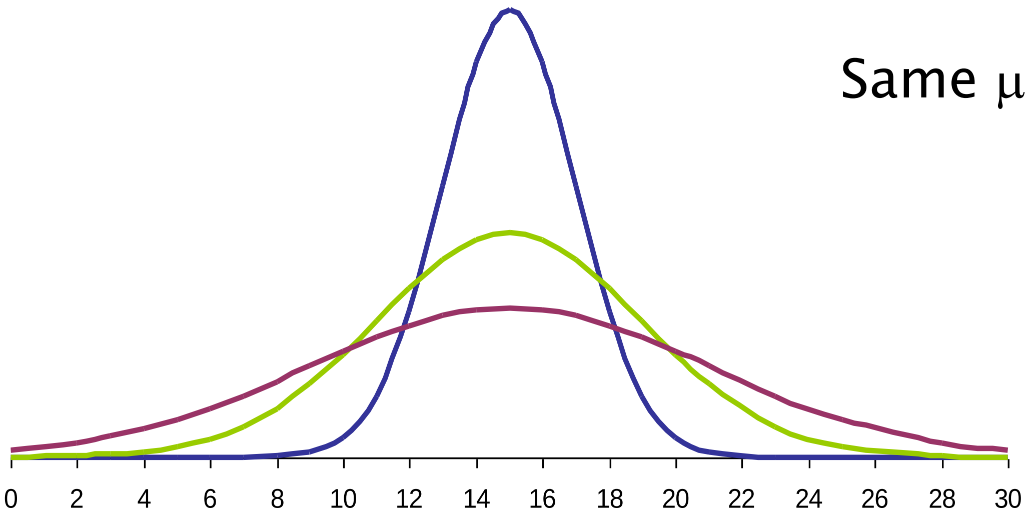
The Normal Distribution

- The fundamental distribution underlying most of inferential statistics is the *Normal (Gaussian) distribution*.

- Normal distributions are a family of symmetric, bell shaped density curves defined by a mean μ and a standard deviation σ: denoted $N(\mu, \sigma)$. The inflection points are $\mu \pm \sigma$.

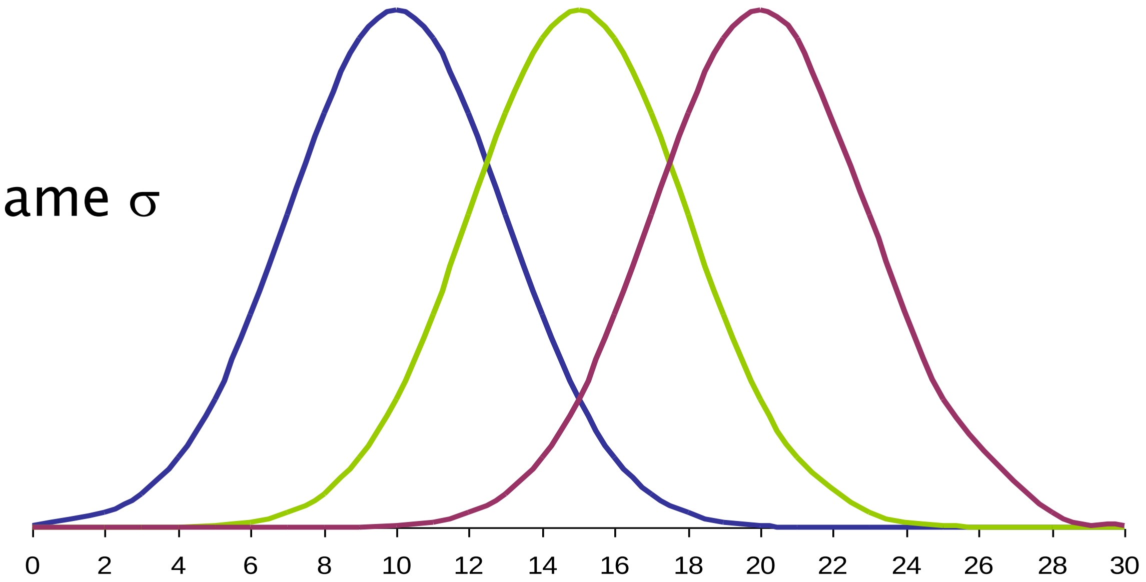


A Family of Normal Density Curves



Same μ different σ

Different μ same σ



The Normal Distribution

- A continuous random variable X defined by it's parameters μ, and σ (where $-\infty<\mu<\infty$ and $\sigma>0$)
- Denoted: N(μ, σ^2)

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

- Measures:
  - $E[X]=\mu$ and $V[X]=\sigma^2$

Properties of the Normal Distribution

1. The curve is symmetric about the mean (i.e. area under the curve to the left of the mean is equal to the area under the curve to the right of the mean).

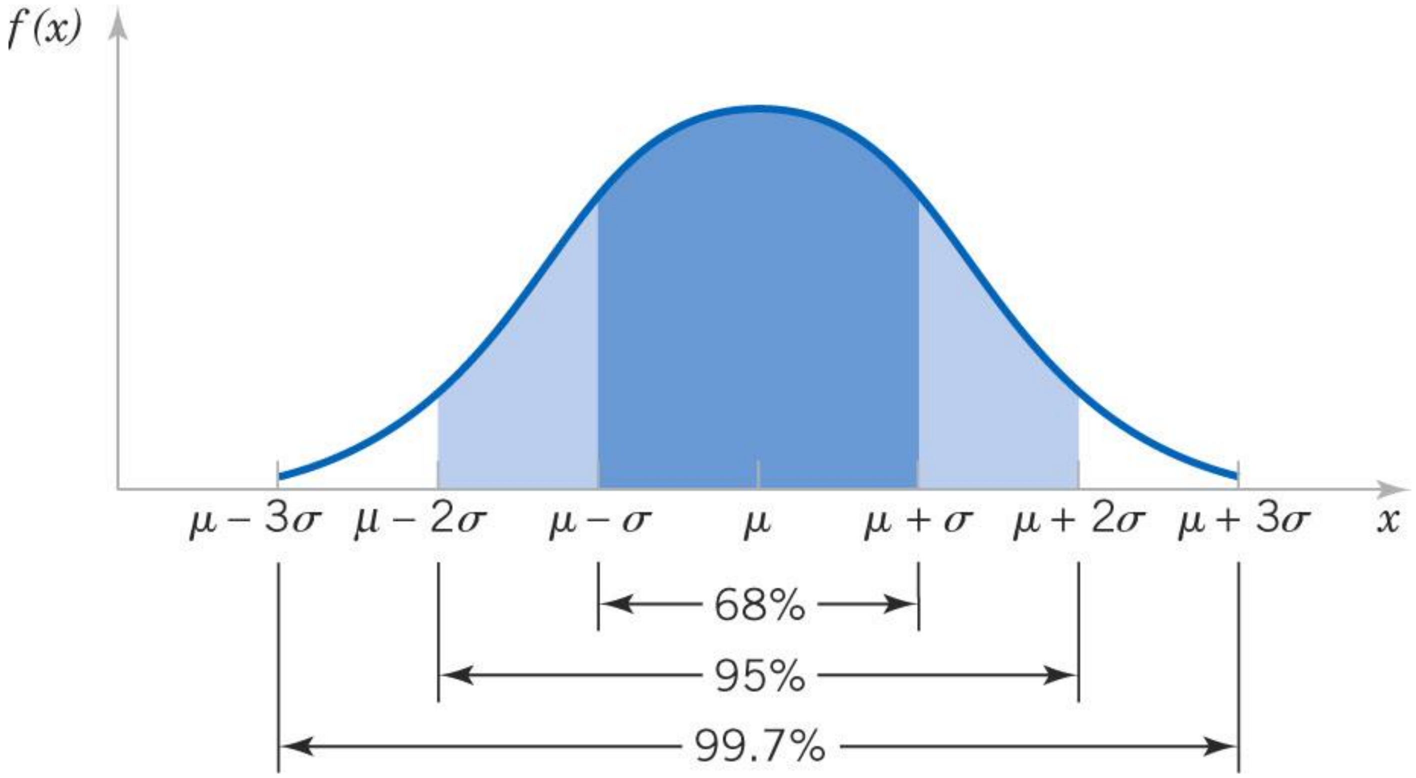
2. The mean = median = mode. So, the highest point of the curve is at $x = \mu$.

3. The curve has inflection points at $(\mu - \sigma)$ and $(\mu + \sigma)$.

4. The total area under the curve is equal to 1.

5. As x gets larger and larger (in either the positive or negative directions), the graph approaches but never reaches the horizontal axis.

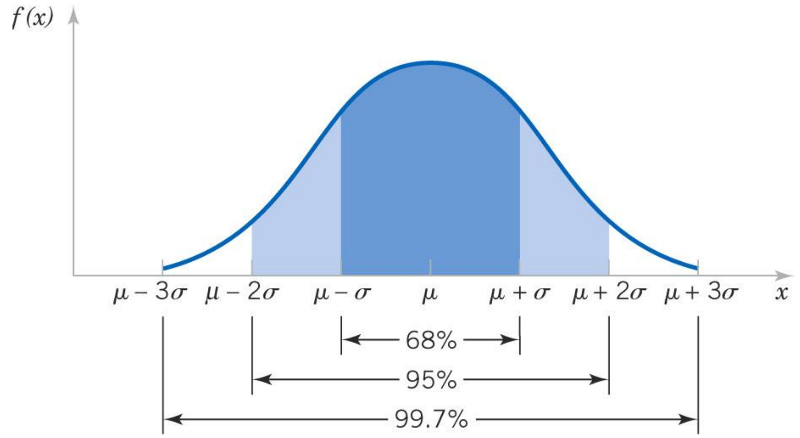
Empirical rule for Normal Distribution



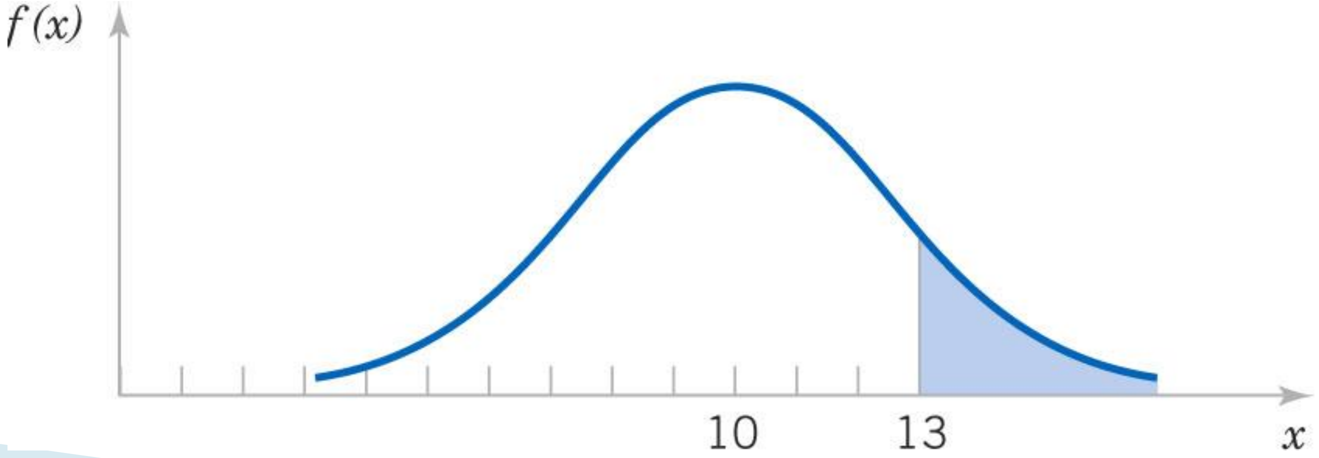
$$P(\mu - \sigma < X < \mu + \sigma) = 0.68$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.95$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997$$

Normal Example

- Assume that the current measurements in a strip of wire follows a normal distribution with a mean of 10 mA & a variance of 4 mA^2. Let X denote the current in mA.



- $P(8 < X < 12)$
  - About 68% (By Empirical Rule)
- $P(X < 12)$
  - About 84% (By Empirical Rule)
- $P(X > 13) = ?$
  - We can either integrate
  - OR….



Integrating the Normal PDF

- Recall the Normal PDF:

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

- In order to integrate an important identity is the Gaussian Integral:

$$\int_{-\infty}^{\infty} e^{-a(x+b)^2}\, dx = \sqrt{\frac{\pi}{a}}$$

Standard Normal Distribution

- A normal random variable with $\mu = 0$ and $\sigma^2 = 1$ is called a standard normal random variable and is denoted as Z.

$$Z \sim N(0,1)$$

- The cumulative distribution function of a standard normal random variable is denoted as:

$$\Phi(z) = P(Z \leq z) = F(z)$$

- Values can be found in the Z table

Standard Normal Distribution examples

- $P(Z \leq 1.58)$
  - 0.9429
- $P(Z \geq -1.58)$
  - 0.9429
- $P(-1.58 \leq Z \leq 1.58)$
  - 0.8858

Standardizing

- So when we standardize X using the formula

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

- The z-score (z-number) represents how many standard deviations x is into its distribution.

$$\text{In words}: z = \frac{(\text{value} - \text{mean})}{\text{standard deviation}}$$

Standardizing to find a probability

- Suppose X is a normal random variable with mean μ and variance σ^2 so, $X \sim N(\mu, \sigma^2)$

$$\text{Then, } P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P(Z \leq z)$$

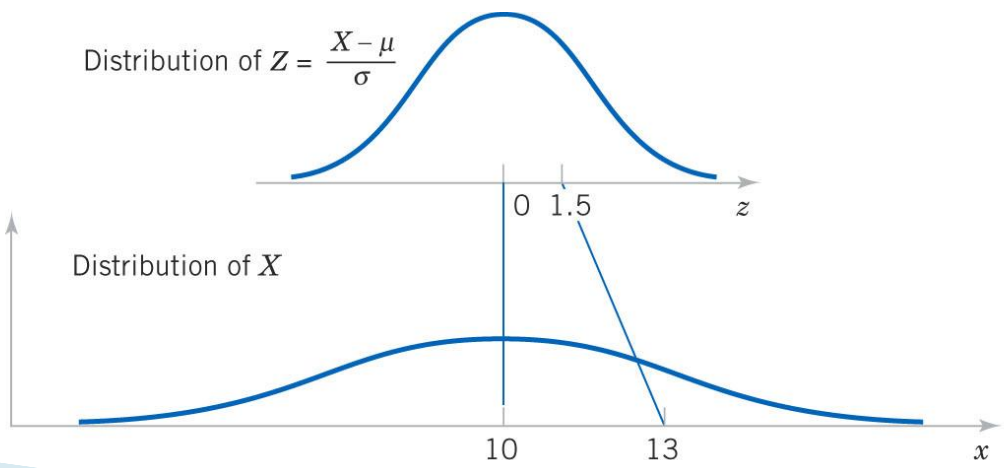
where Z is a standard normal random variable, and

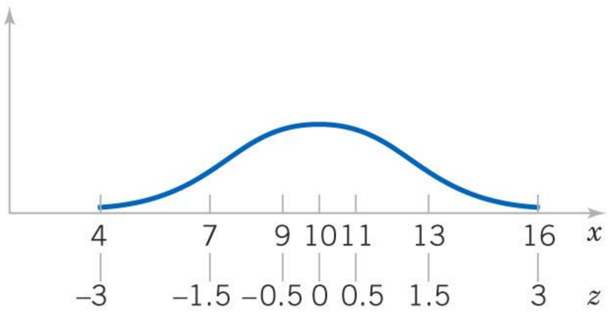
$$z = \frac{(x-\mu)}{\sigma}$$ is the z-value obtainedby standardizing X.

Standardizing Example

- Current measurements on a strip of wire are distributed normally with $\mu = 10$ mA and $\sigma^2 = 4$ mA. Let X be a random current reading.
  - Find $P(X > 13)$.

$P(X > 13) = P(Z > \underline{1.5}) = 1 - P(Z < \underline{1.5}) = 1 - \underline{0.9332} = \underline{0.0668}$





Standardizing Example (Between)

- From the previous example with μ = 10 and σ = 2 mA, what is the probability that the current measurement is between 9 and 11 mA?

$$P(9 < X < 11) = P\left(\frac{9-10}{2} < \frac{x-10}{2} < \frac{11-10}{2}\right)$$

$$= P(-0.5 < z < 0.5)$$

$$= P(z < 0.5) - P(z < -0.5)$$

$$= 0.69146 - 0.30854 = 0.38292$$

Using Excel	
0.38292	= NORMDIST(11,10,2,TRUE) - NORMDIST(9,10,2,TRUE)

Finding X-value from a probability ("Un-standardizing")

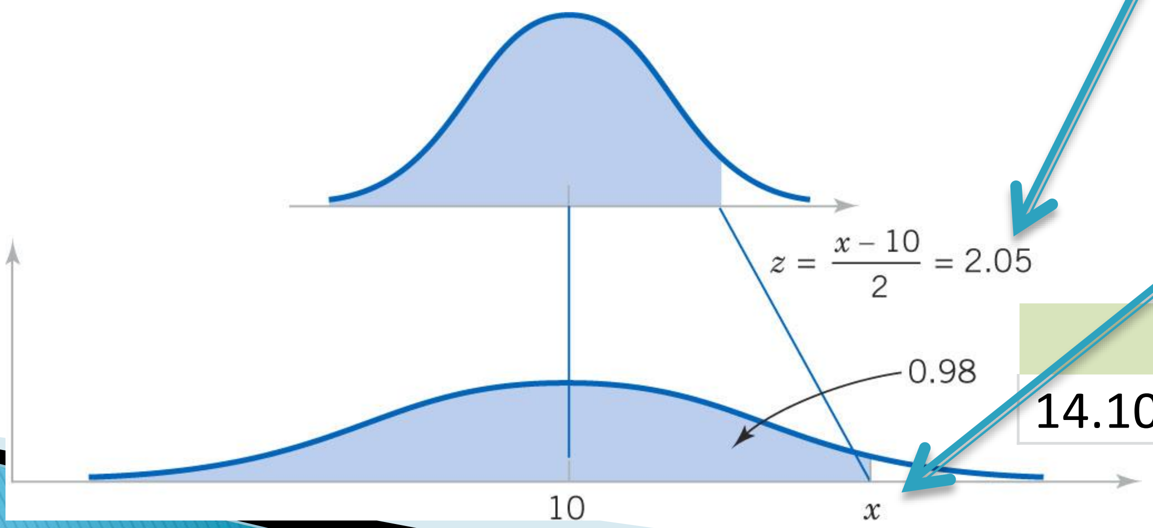
- Determine the value for which the probability that a current measurement is below this value is 0.98.
- We First need to find the Z-score that corresponds to the given area
- We can then can "un-standardize" the z scores ($x = \mu + \sigma z$)

$$P(X < x) = P\left(\frac{X-10}{2} < \frac{x-10}{2}\right)$$

$$= P\left(Z < \frac{x-10}{2}\right) = 0.98$$

$z = 2.05$ is the closest value.

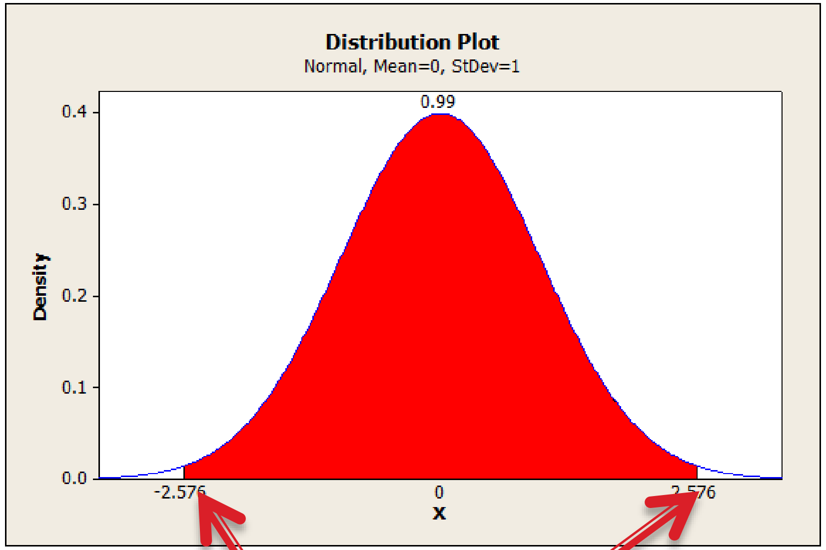
$x = 2(2.05) + 10 = 14.1$ mA.



Using Excel	
14.107	= NORMINV(0.98,10,2)

Signal Detection Example 2

- Determine the symmetric bounds about 0 that include 99% of all noise readings. We need to find x such that $P(-x < N < x) = 0.99$.
- First we need to find the z scores such that $P(-z < Z < z) = 0.99$



So $z_1 = -2.576$ and $z_2 = 2.576$

So once we find the z scores we can unstandardize z using the following formula:
$x = 0 + 0.45 * z$

$$x_1 = -2.576(0.45) + 0 = -1.16$$

$$x_2 = 2.576(0.45) + 0 = 1.16$$