Continuous Distribution

Normal and Exponential distribution

Md. Fazlul Karim Patwary
IIT, JU

Continuous distribution

- A ***continuous probability distribution*** refers to
  - the range of all possible values that a continuous random value can assume,
  - together with the associated probabilities.

- The probability distribution of a continuous random variable is often called a ***probability density function***, or simply a ***probability function***.

Continuous distribution

If a random variable X may have values in a subset S of R^n then it is continuous random variable and for a range of values it has probability.

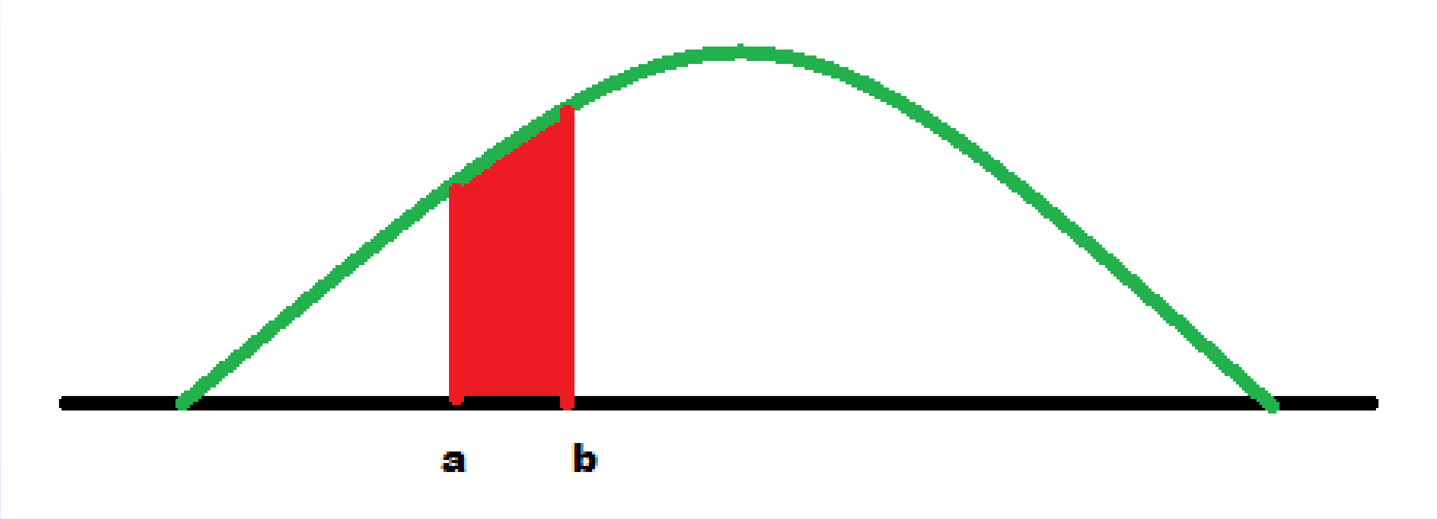
Unlike discrete probability, where we can calculate probability of a value, it expresses probability of a range of values:

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

Area under the curve from the range **a** to **b**

Continuous distribution

Area under the curve from the range **a** to **b**



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

For any fixed value of x, the probability is zero

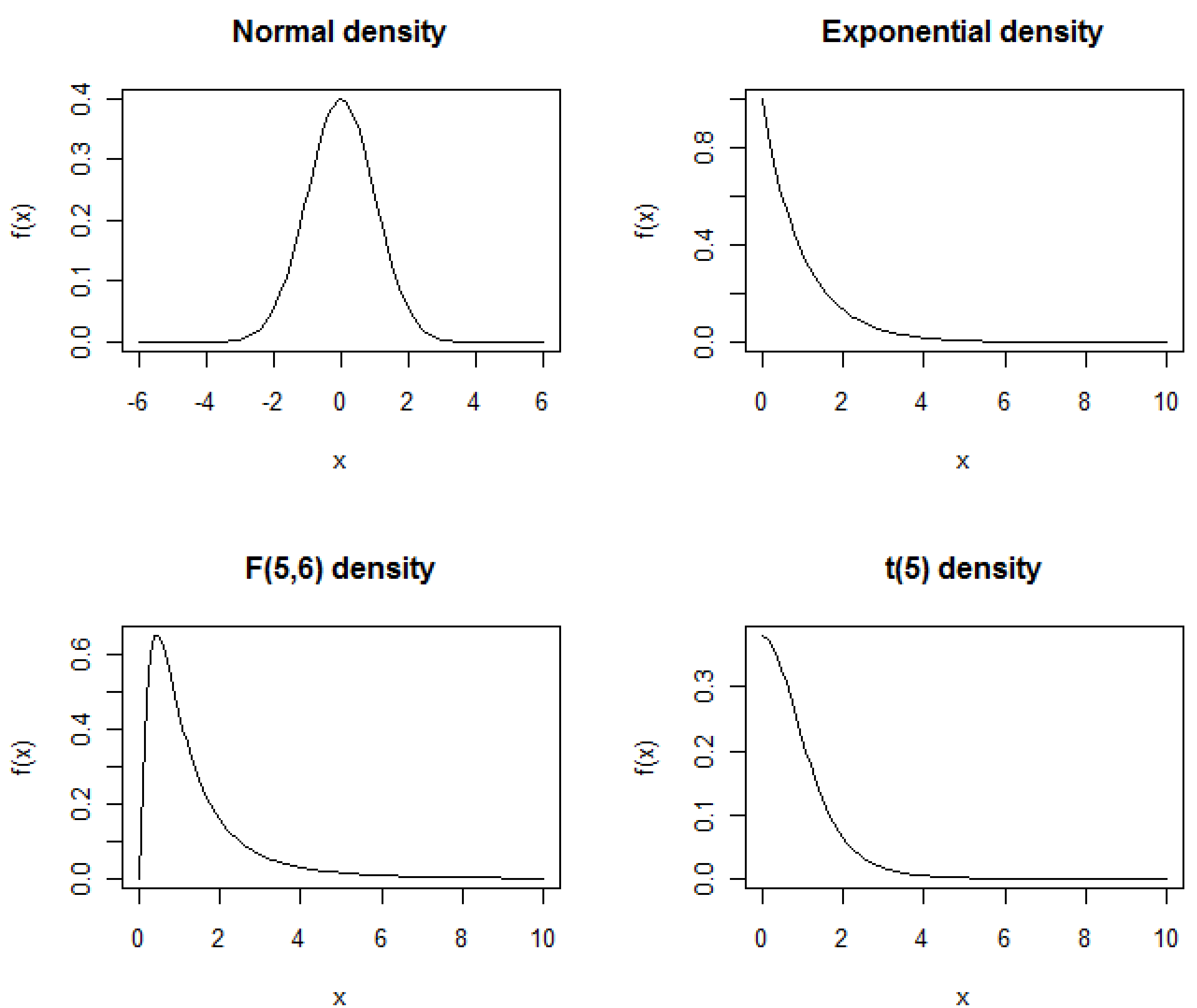
i.e. P(X=a) = *f*(X=a) = 0; for each x in S

Continuous distribution

Popular and useful continuous distributions are:

- **Normal distribution**
- **Exponential distribution**
- Fisher's F distribution
- Student's t distribution

Continuous distribution



R codes

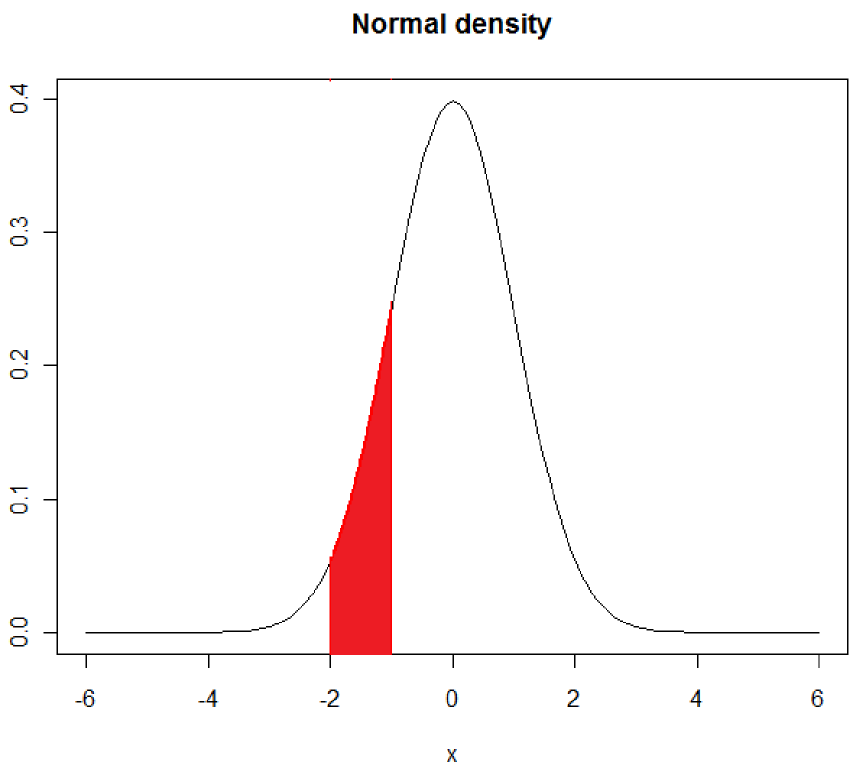
```
# R code to draw plots
par(mfrow=c(2,2))
plot(function(x) dnorm(x), -6, 6,main = "Normal density", ylab="f(x)")
plot(function(x) dexp(x), 0, 10,main = "Exponential density",ylab="f(x)")
plot(function(x) df(x,5,6), 0, 10,main = "F(5,6) density",ylab="f(x)")
plot(function(x) dt(x,5), 0, 10,main = "t(5) density",ylab="f(x)")

par(mfrow=c(1,1))
plot(function(x) dnorm(x), -6, 6,main = "Normal density", ylab="f(x)")
curve(dnorm(x), add=TRUE, from=-2,to=-1,col="red",lwd=2)
abline(v=-2:-1,col = "red")
```

Normal distribution

- Carl Friedrich Gauss proposed this distribution
- Called Gaussian distribution
- A Bell shaped curve:
  - Curve concentrated in center and decreases on either side: data has less tendency to produce unusual extreme value.
  - Symmetric: probability of deviation from mean are comparable in either direction
- Curve is mesokurtic

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad -\infty \leq x \leq \infty$$

Normal distribution



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

R code: Let a = -2 and b = -1

```
pnorm(-1, mean=0,sd=1) - pnorm(-2,mean=0,sd=1)
[1]   0.1359051
```

Standard Normal distribution

- A normal distribution with mean 0 and standard deviation 1 is called **standard normal distribution**.

- Any normal distribution can be transformed into a standard normal distribution:

If X has normal distribution with mean $\mu=22$ and standard deviation $\sigma=5$ then

$$z=(X-\mu)/\sigma$$

has a standard normal distribution.

Why normal distribution is important?

- The normal distribution is the most commonly used of all probability distributions in statistical analysis.

- Many distributions actually found in nature and industry are normal.

  i.e. IQs, weights, and heights of a large number of people and the variations in dimensions of a large number of parts produced by a machine.

Why normal distribution is important?

- The normal distribution often can be used to approximate other distributions, such as the binomial and the Poisson distributions.

- Distributions of sample means and proportions are often normal, regardless of the distribution of the parent population

Example: Normal distribution

The lifetime of light bulbs is known to be normally distributed with ***μ = 100 h and σ = 8 h.*** *What* is the probability that a bulb picked at random will have a lifetime between 110 and 120 burning hours?

Sol. We have to find **P(110 < *X* < *120)=?***

We can transform X (normal variate) to Z (standard normal variate)

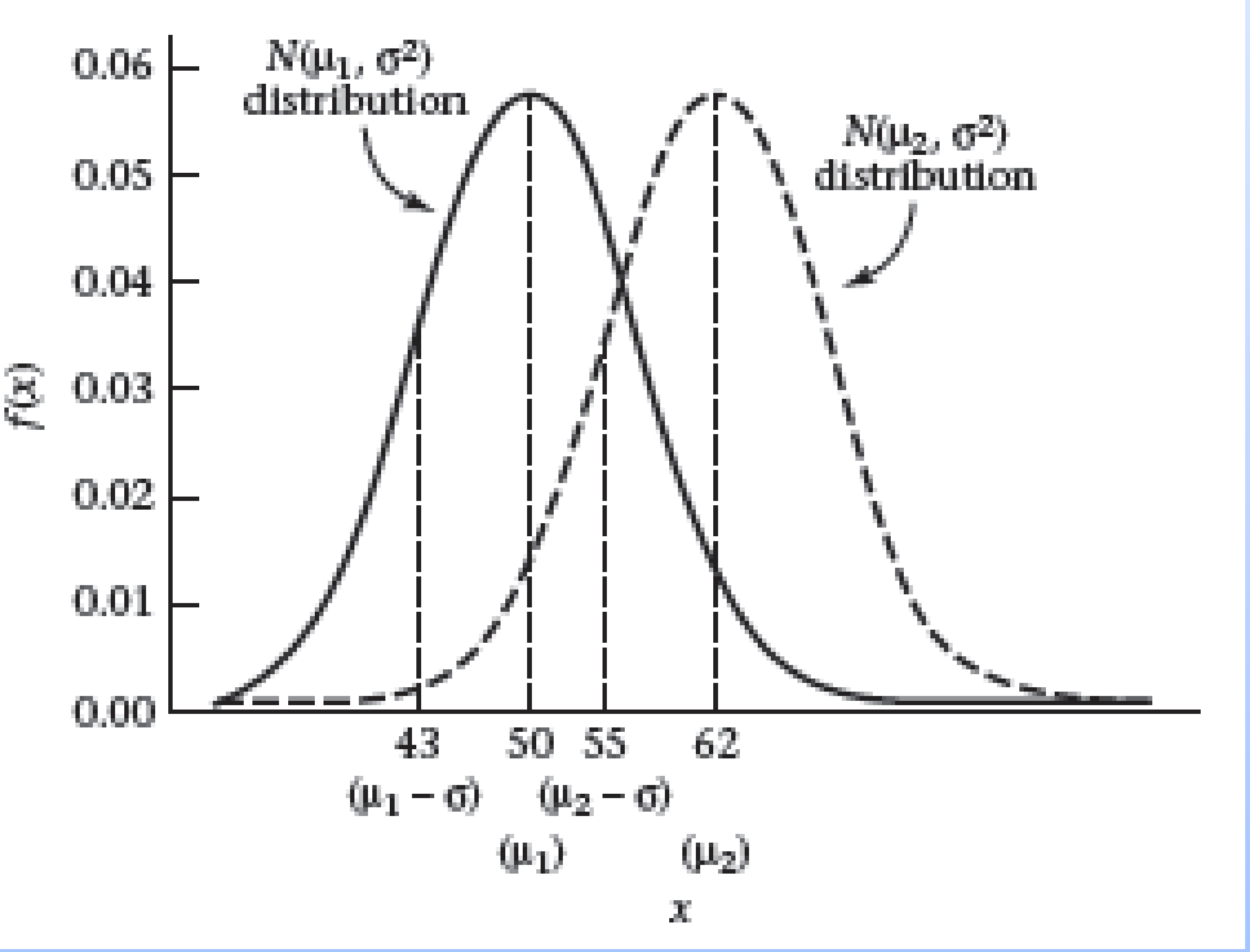
$$Z_1 = \frac{(x_1 - \mu)}{\sigma} = \frac{(110-100)}{8} = \cancel{-1.25} \quad 1.25$$

$$Z_2 = \frac{(x_2 - \mu)}{\sigma} = \frac{(120-100)}{8} = \cancel{1.5} \quad 2.5$$

P(110 < *X* < *120)=P*(1.25 < *Z* < *2.5)= 0.0994* *(Using R code)*

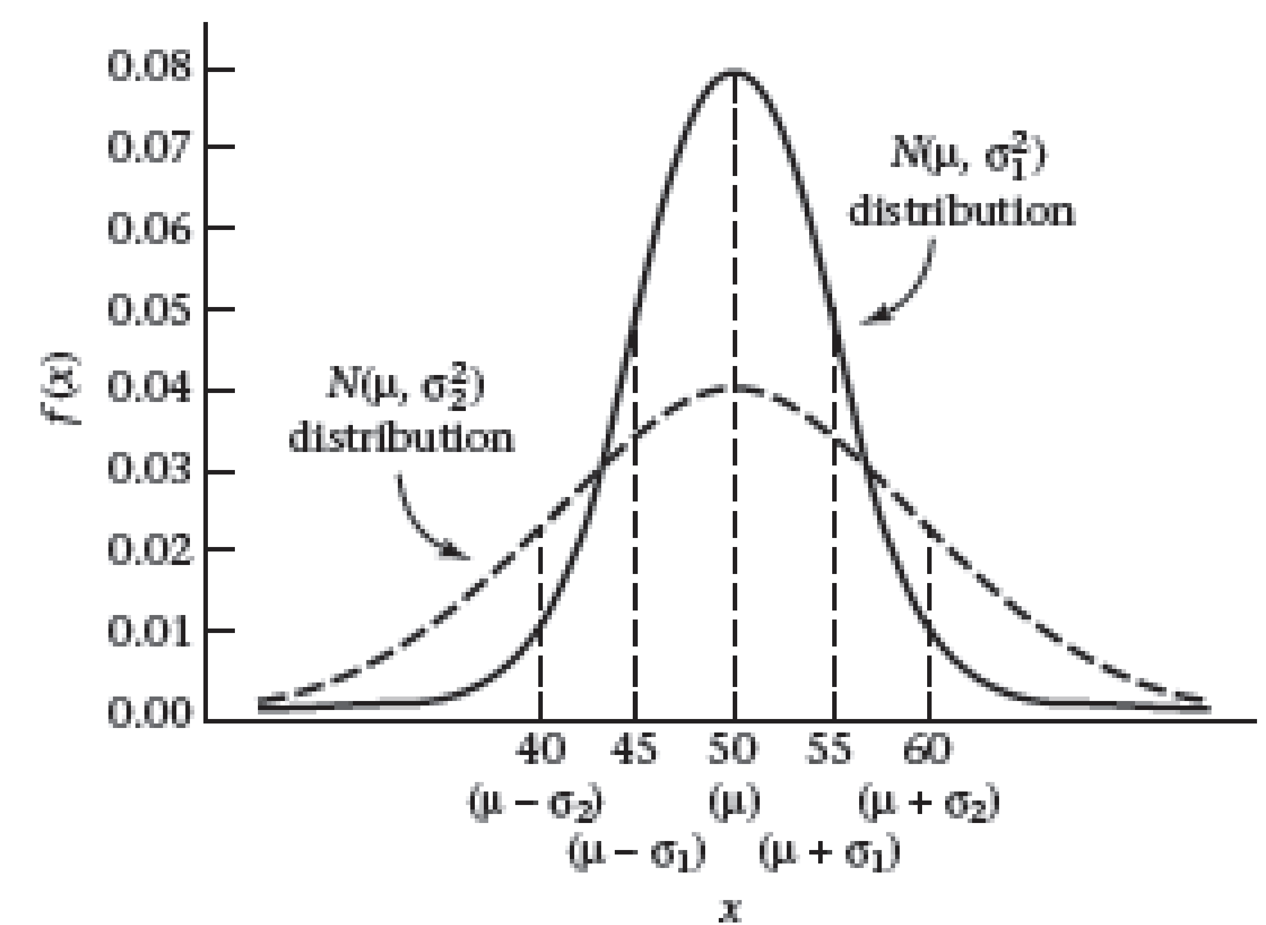
Some Definition

- A normal distribution with mean μ and variance σ2 will generally be referred to as an N(μ,σ2) distribution.



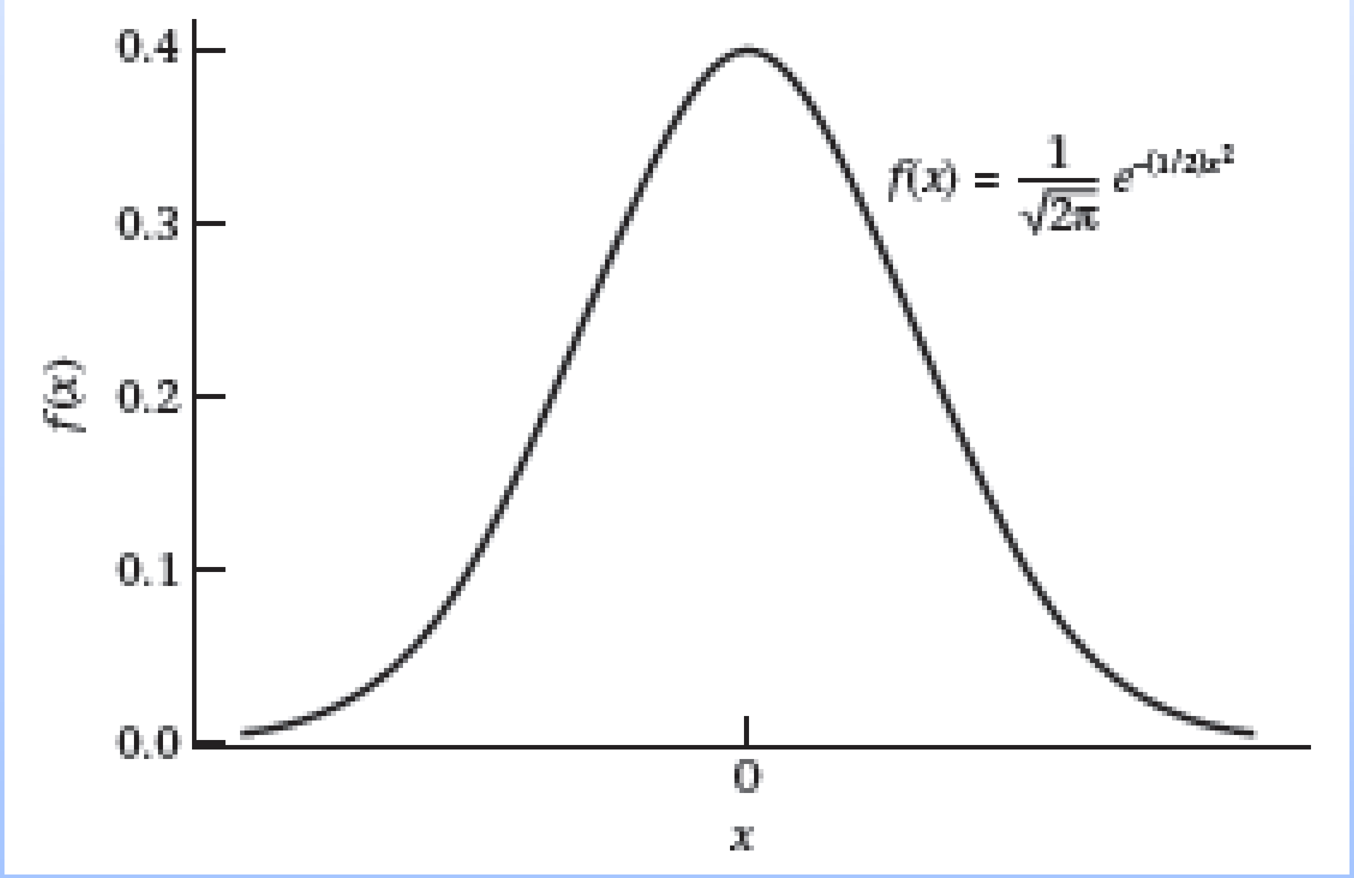
Some Definition

- Two normal distributions with the same means and different variances.

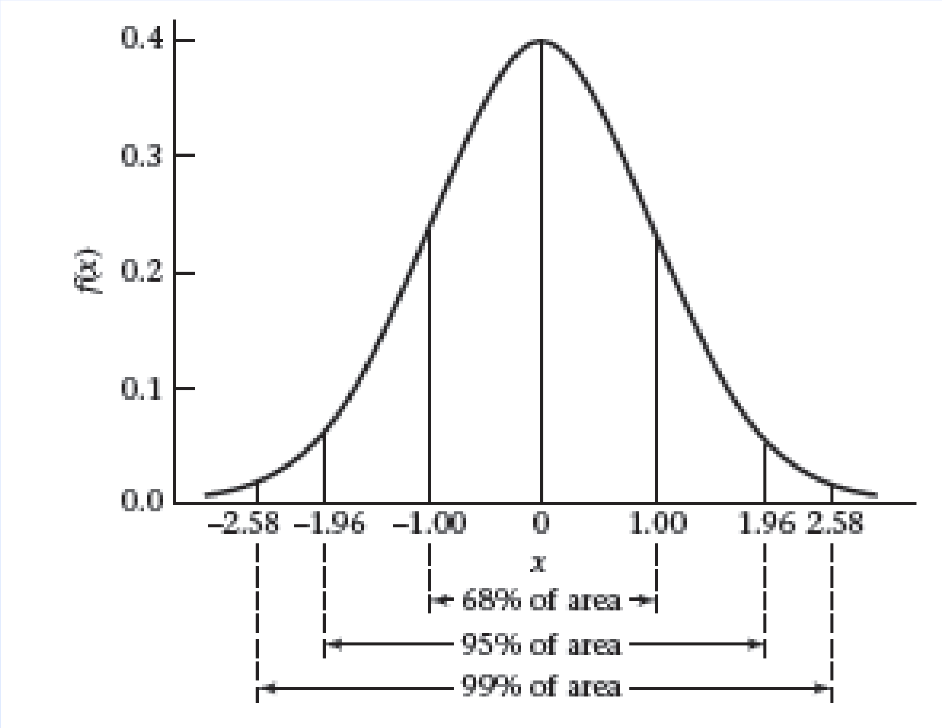


Some Definition

- A normal distribution with mean 0 and variance 1 is called a standard normal distribution. This distribution is also called an *N(0,1) distribution.*



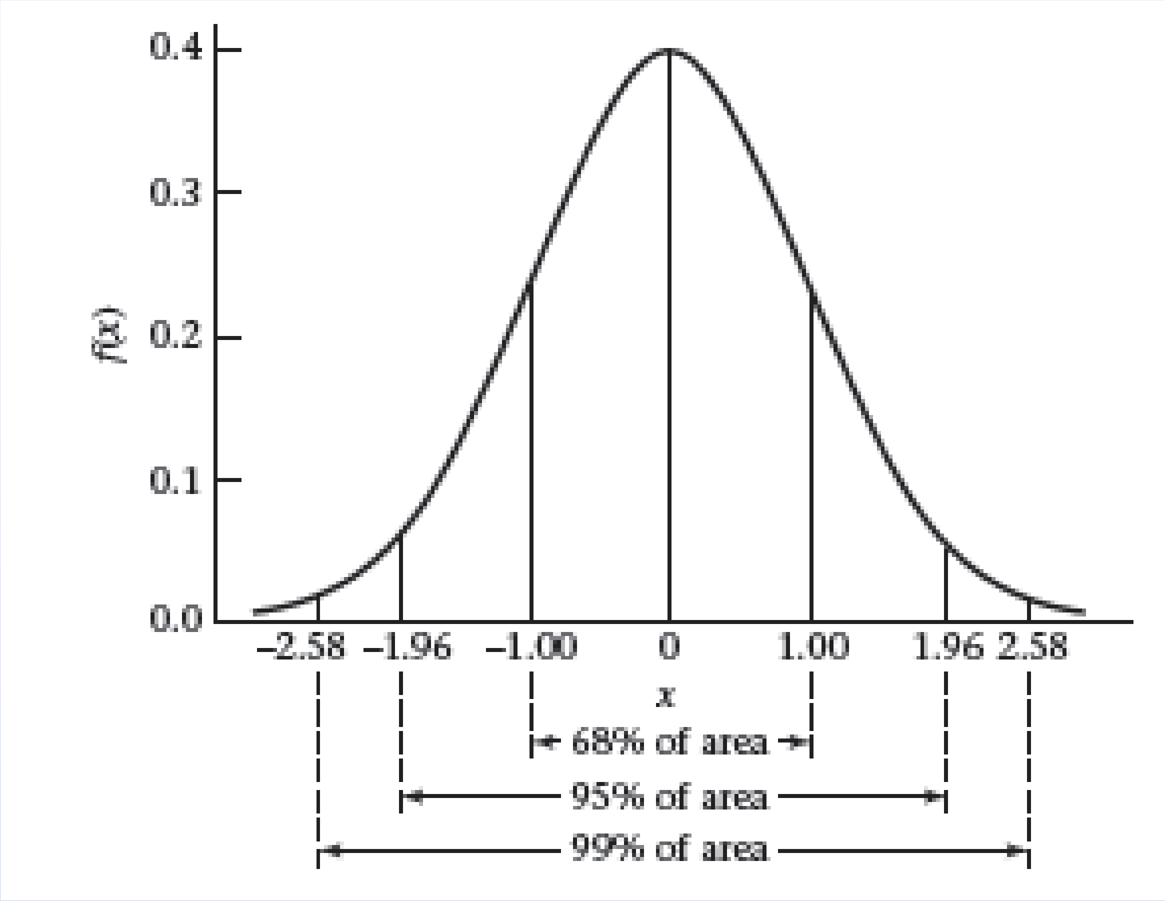
Some Definition



- About 68% of the area under the standard normal density lies between +1 and −1.

$$Pr(-1 < X < 1) = .6827$$

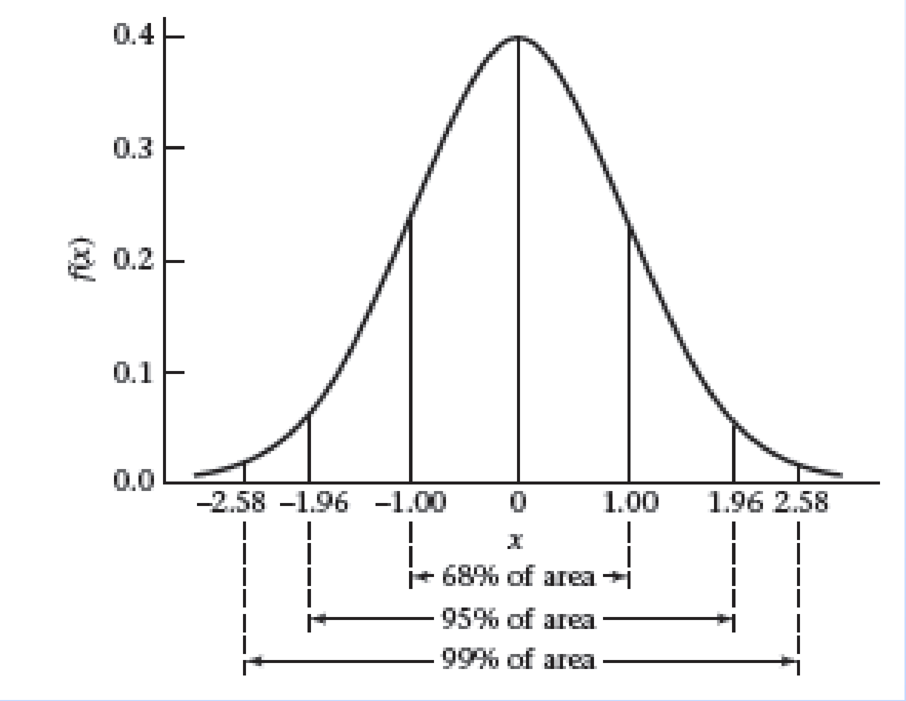
Some Definition



- About 95% of the area lies between +2 and −2, and

Pr(−1.96 < X < 1.96) = .95

Some Definition



- About 99% lies between +2.5 and −2.5.

$$Pr(-2.576 < X < 2.576) = .99$$

- Thus, the standard normal distribution slopes off very rapidly, and absolute values greater than 3 are unlikely

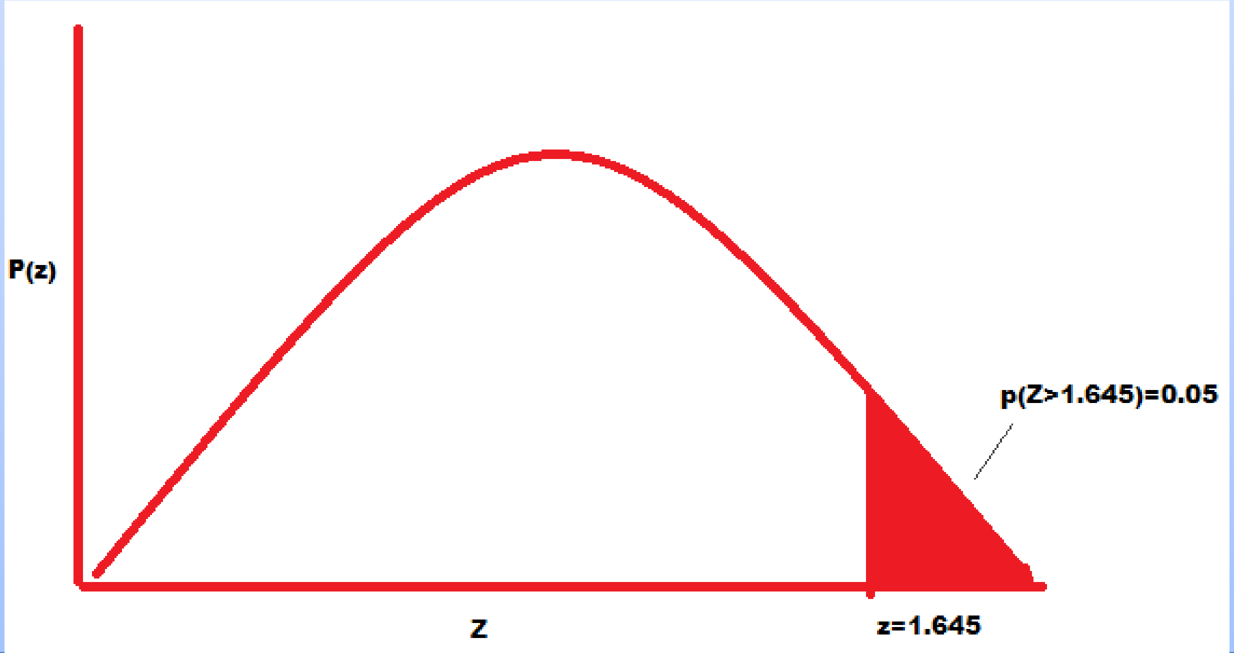
Percentile points for normally distributed variables

The lifetime of light bulbs is known to be normally distributed with ***μ = 100 h and σ = 8 h.*** *What* is the 95th percentile?

Sol. We have to find value of x = **a** for which **P(X>a*)=.05***

But we can have value z=1.64 for which **P(Z>z*)=.05***

Since we can transform z (normal variate) to X (standard normal variate)

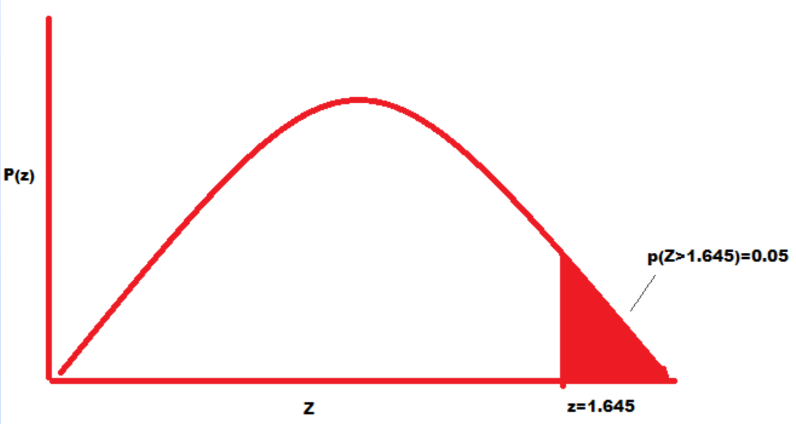


Percentile points for normally distributed variables

Back Calculation:

z=1.65 ; ***μ = 100 h and σ = 8 h.***

$z=(X- \mu)/ \sigma$



Hence, ***X = μ + z σ*** **= 113.12 h**

P(X > 113.12)=0.05 and 95^{th} percentile is 113.12

That is, ***95% bulbs will last 115.2 hours long.***

Percentile points for normally distributed variables

R code:

qnorm(.95, mean=100, sd=8) for x=a value

or *qnorm(.05)* for Z=z value

- Find 5th percentile.
- What is the interpretation of this result?

Normal approximation of binomial probabilities

When the number of observations or trials n is relatively large, the normal probability distribution can be used to approximate binomial probabilities.

Binomial: Mean = np, $SD = sqrt(np(1-p))$

If n >= 30 and np >= 5 and n(1-p)>=5 then we can transformed binomial X variate to standard normal (z) variate as:

Z=(X-np)/SD

Normal approximation of binomial probabilities

Example: For a large group of sales prospects, it has been observed that 20 percent of those contacted personally by a sales representative will make a purchase. If a sales representative contacts 30 prospects, What is the probability that 10 or more will make a purchase?

Sol. P = 0.02; n = 30; P(X>=10) = ?

P(X >= 10 / n = 30, p = 0.20) = P(X = 10) + P(X = 11)+ ....

= 0.0355 + 0.0161 + 0.0064 + 0.0022 + 0:0007 +....

= 0.0610853

R code:

```
dbinom(10, 30, .2)+ dbinom(11, 30, .2)+dbinom(12, 30, .2)+dbinom(13, 30, .2)+
dbinom(14, 30, .2)+ dbinom(15, 30, .2)+dbinom(16, 30, .2)+dbinom(17, 30, .2)
```

Normal approximation of binomial probabilities

Using Normal approximation of Binomial

n=30 (>=30)

nq=30*0.2 = 6 (>=5)

nq=30*0.8 = 24 (>=5)

So we can approximate normal distribution

Mean = np = 30*0.2 = 6

Sd = sqrt(npq) = sqrt(30*.2*.8) = sqrt(4.8) = 2.19

P(X>=9.5) = ? Z = (x-mean)/sd= (9.5-6)/2.19= 1.60

So P(X>=9.5) = P(Z>1.60)

=**0.0548 (which is close to previous)**

(N.B. used 9.5 instead of 10 as a continuity correction)

R code: 1-pnorm(9.5, mean=6, sd=2.19) or 1-pnorm(1.6)

Normal approximation of binomial probabilities

Example: Suppose we want to compute the probability that between 50 and 75 of 100 white blood cells will be neutrophils, where the probability that any one cell is a neutrophil is .6. These values are chosen as proposed limits to the range of neutrophils in normal people, and we wish to predict what proportion of people will be in the normal range according to this definition.

Sol. Binomial probability function is $\binom{100}{k}(.6)^k(.4)^{100-k}$

The exact probability for this problem is $\sum_{k=50}^{75}\binom{100}{k}(.6)^k(.4)^{100-k}$

Normal approximation:

Mean = np = 100(.6) = 60, and

Variance is = np(1-p) = 100(.6)(.4) = 24.

Thus area between 49.5 and 75.5 for an *N(60,24) distribution:*

Pr(49.5 <= X <= 75.5) = Pr(-2.143 <= Z <= 3.164) = 0.983

Thus 98.3% of the people will be normal.

Normal approximation of binomial probabilities

Example: Suppose a neutrophil count is defined as abnormally high if the number of neutrophils is ≥ 76 and abnormally low if the number of neutrophils is ≤ 49. Calculate the proportion of people whose neutrophil counts are abnormally high or low.

Sol. Binomial probability function is $\binom{100}{k}(.6)^k(.4)^{100-k}$

The exact probability for this problem is

$$\sum_{k=76}^{100}\binom{100}{k}(.6)^k(.4)^{100-k}+\sum_{k=0}^{49}\binom{100}{k}(.6)^k(.4)^{100-k}$$

Normal approximation:

Mean = np = 100(.6) = 60, and

Variance is = np(1-p) = 100(.6)(.4) = 24.

Thus 1.7% of the people will be normal.

Normal approximation of poisson probabilities

When the mean λ of a Poisson distribution is relatively large, the normal probability distribution can be used to approximate Poisson probabilities.

A convenient rule is that such approximation is acceptable when λ = 10.0.

Mean= λ
Sd=sqrt(λ)

We can transformed binomial X variate to standard normal z variate as:

$$Z=(X- \lambda)/\mathrm{sqrt}(\lambda)$$

Normal approximation of poisson probabilities

Example: The average number of calls for service received by a machine repair department per 8-hr shift is 10.0. What is the probability that more than 15 calls will be received during a randomly selected 8-hr shift?

Using poisson distribution the exact probability is :

$P(X > 15/\lambda=10.0) = P(X = 16) + P(X = 17)+ \ldots\ldots$

$= 0.0217 + 0.0128 + 0.0071 + 0.0037 + 0.0019 + 0.0009 + 0.0004 + 0.0002 + 0.0001$

$= 0.04869345$

R code:
```
dpois(16, lambda=10) + dpois(17, lambda=10) + dpois(18, lambda=10) +
dpois(19, lambda=10) + dpois(20, lambda=10) + dpois(21, lambda=10) +
dpois(22, lambda=10) + dpois(23, lambda=10) + dpois(24, lambda=10)
```

Normal approximation of poisson probabilities

Using Normal distribution the probability is :

Z=(x-mean)/sd=(15.5-10)/sqrt(10)=1.74

P(X>15.5) = P(Z>1.74) = 0.0409

which is close to previous result.

R code:

```
1-pnorm(15.5, mean=10, sd=3.16) or 1-pnorm(1.74)
```

Normal approximation of poisson probabilities

R code:

Using z: 1-pnorm(1.6)

Using x: 1-pnorm(9.5, mean=6, sd=2.19)

Exponential distribution

If events occur in a Poisson process then length of time or space between successive events follows an ***exponential probability distribution***.

Since, time or space is continuous, such a measurement is a continuous random variable.

D

$$f(x) = \lambda e^{-\lambda x} \qquad x > 0 \; and \; \lambda > 0$$

Times Between Road Accidents

- The times between accidents for a 10-year period in a busy road can be modeled by the exponential distribution.

$$f(y) = \lambda e^{-\lambda y} \qquad y > 0 \; and \; \lambda > 0$$

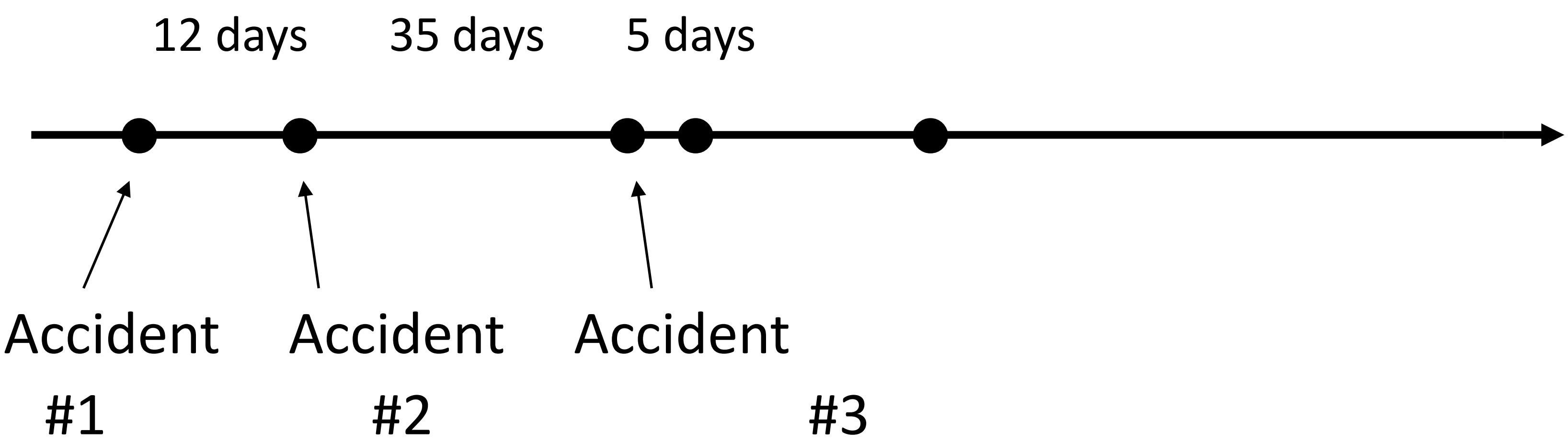
where λ is the accident rate

(the expected number of accidents per day in this case)

Example of time between accidents

Let X = the number of days between two accidents.

Time



Times Between Road Accidents

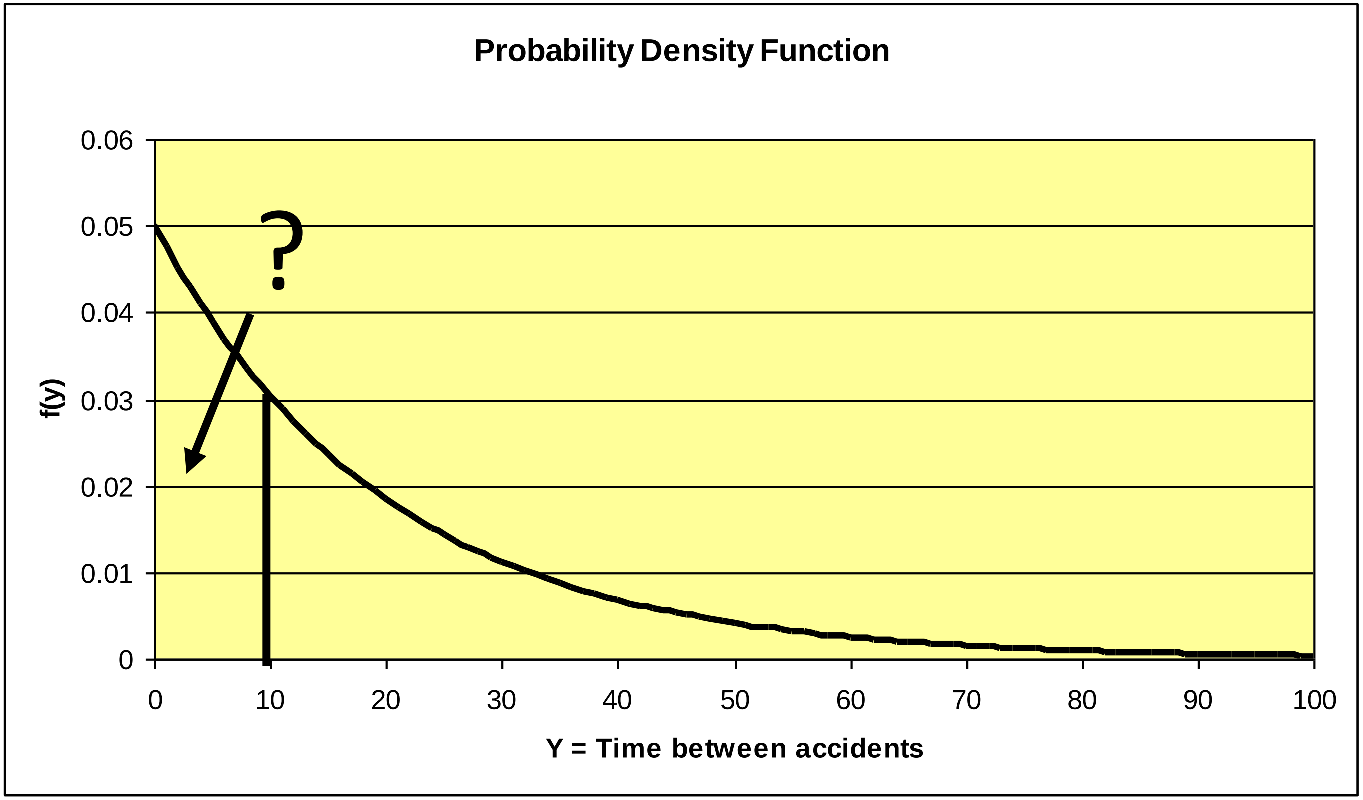
- Suppose in a 1000 day period there were 50 accidents.

$\lambda = 50/1000 = 0.05$ accidents per day

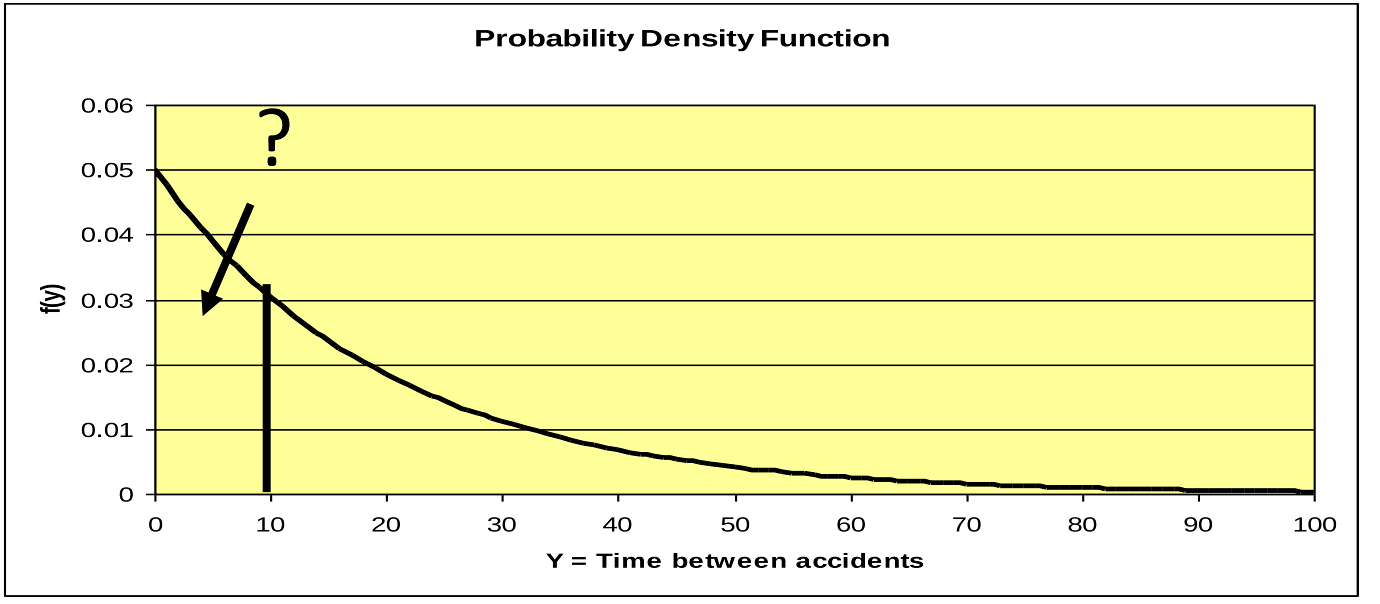
or

$1/\lambda = 1000/50 = 20$ days between accidents

What is the probability that this road will go less than 10 days between the next two accidents?



$f(y) = 0.05e^{-0.05x}$



$$P(X \le 10) = F(10) = \int_0^{10} 0.05e^{-0.05y}dx$$

Recall: $\int e^u du = e^u$

$$F(10) = -e^{-0.05x} \Big|_0^{10} = 0.39$$

$$P(Y \leq 10) = F(10) = \int_0^{10} 0.05e^{-0.05y}\,dy$$

$$\text{Let } u = -0.05y$$

$$du = -0.05dy$$

If: $$\int e^u\,du = e^u$$

$$F(10) = -e^{-0.05y} \Big|_0^{10}$$

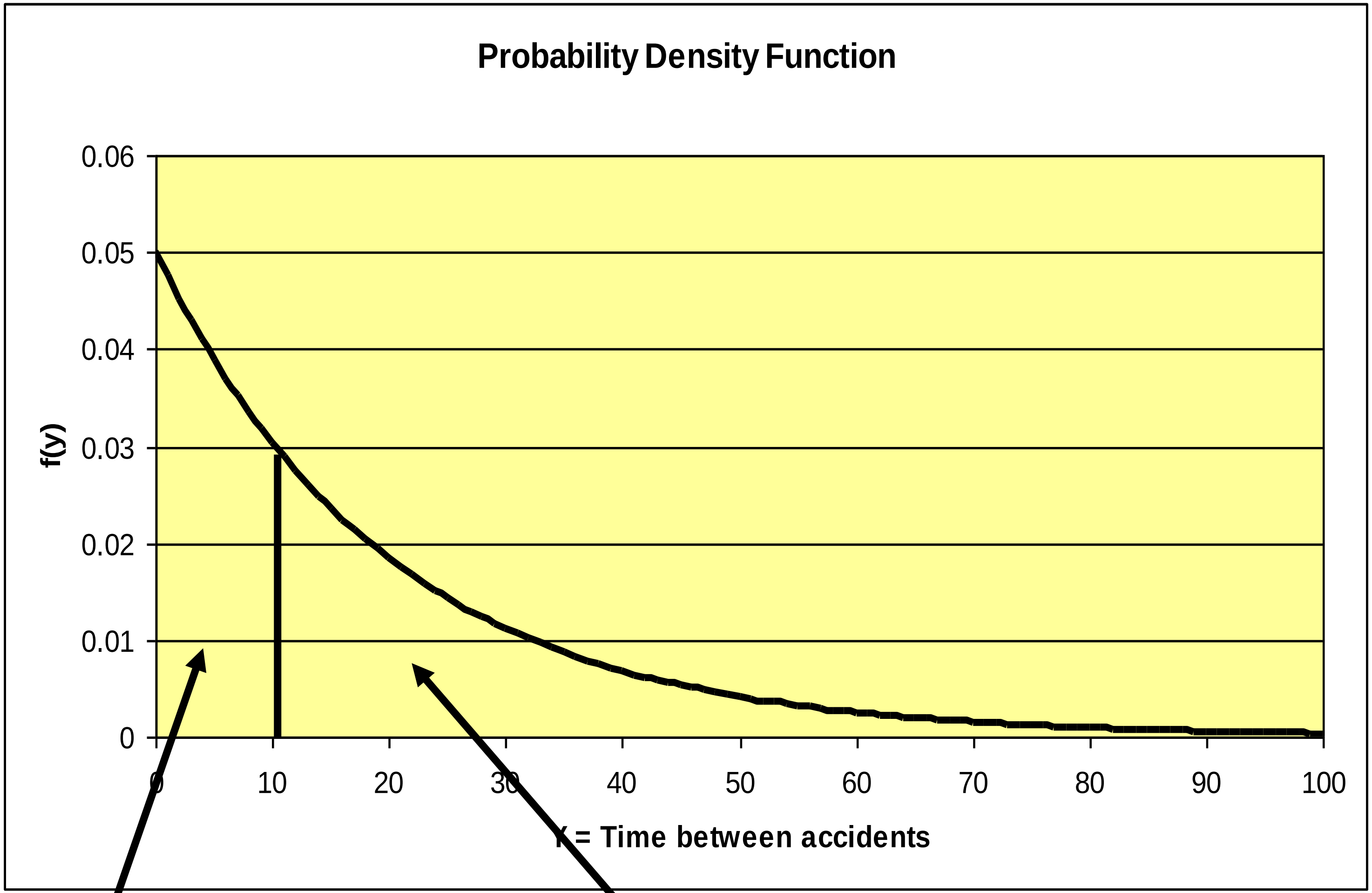
In General…

$$P(X \le x) = F(x) = \int_0^x \lambda e^{-\lambda t}\,dt$$

$$P(X \le x) = F(x) = -e^{-\lambda t}\,\Big|_0^x = 1 - e^{-\lambda x}$$

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

Exponential Distribution



$$1-e^{-\lambda y}$$

$$e^{-\lambda y}$$

Example: If the time to failure for an electrical component follows an exponential distribution with a mean time to failure of 1000 hours, what is the probability that a randomly chosen component will fail before 750 hours?

Hint: λ is the failure rate (expected number of failures per hour).

Mean and Variance for an Exponential Random Variable

$$E(X) = \mu = \int_0^{\infty} x\lambda e^{-\lambda x}\,dx = \frac{1}{\lambda}$$

$$Var(X) = \sigma^2 = \int_0^{\infty} x^2 \lambda e^{-\lambda x}\,dx - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Note: Mean = Standard Deviation

Example: The time between accidents at a factory follows an exponential distribution with a historical average of 1 accident every 900 days. What is the probability that that there will be more than 1200 days between the next two accidents?

Example: If the time between accidents follows an exponential distribution with a mean of 900 days, what is the probability that there will be less than 900 days between the next two accidents?

Relationship between Exponential & Poisson Distributions

- Recall that the Poisson distribution is used to compute the probability of a specific number of events occurring in a particular interval of time or space.
- Instead of the number of events being the random variable, consider the time or space between events as the random variable.

Exponential or Poisson Distribution?

- We model the number of industrial accidents occurring in one year.

- We model the length of time between two industrial accidents (assuming an accident occurring is a Poisson event).

- We model the time between radioactive particles passing by a counter (assuming a particle passing by is a Poisson event).

- We model the number of radioactive particles passing by a counter in one hour

Now let T = the time (or space) until the next Poisson event.

$$P(T > t) = e^{-\lambda t}$$

In other words, **the probability that the length of time (or space) until the next event is greater than some given time (or space), t, is the same as the probability that no events will occur in time (or space) t.**

Radioactive Particles

Question: The arrival of radioactive particles at a counter are Poisson events. So the number of particles in an interval of time follows a Poisson distribution. Suppose we average 2 particles per millisecond.

- What is the probability that no particles will pass the counter in the next 3 milliseconds?
- What is the probability that more than 3 milliseconds will elapse before the next particle passes?

Machine Failures

Question: If the number of machine failures in a given interval of time follows a Poisson distribution with an average of 1 failure per 1000 hours.

- what is the probability that there will be no failures during the next 2000 hours?

- What is the probability that the time until the next failure is more than 2000 hours?

Question: Number of failures in an interval of time follows a Poisson distribution. If the mean time to failure is 250 hours, what is the probability that more than 2000 hours will pass before the next failure occurs?

A. e^{-8}

B. $1 – e^{-8}$

C. $e^{-0.125}$

D. $1 – e^{-0.125}$

Problem and solution using R

Suppose the mean checkout time of a supermarket cashier is three minutes. Find the probability of a customer checkout being completed by the cashier in less than two minutes.

Mean=1/λ=1/3

$$P(T < 2) = 1 - e^{-1/3*2} = 0.48658$$

pexp(2, rate=1/3)=0.4865829

Memoryless property

A variable is memoryless with respect to t if, for all s with

$$\Pr(T > s + t \mid T > s) = \Pr(T > t) \text{ for all } s, t \geq 0.$$

Exponential distribution is memoryless distribution.

For example: P(*T* > 30 + 10/T>30)=P(T>10)

more than another 10 seconds before the first arrival, given that the first arrival has not yet happened after 30 seconds,

=

initial probability that we need to wait more than 10 seconds for the first arrival.

Again, if we waited for 30 seconds and the first arrival didn't happen (*T* > 30), probability that we'll need to wait another 10 seconds for the first arrival (*T* > 30 + 10) is the same as the initial probability that we need to wait more than 10 seconds for the first arrival (*T* > 10).

Solve conditional probability problems

- https://probabilityformula.org/conditional-probability-examples.html
- https://www.mathsisfun.com/data/probability-events-conditional.html

Solve Binomial probability problems

- https://www.mathsisfun.com/data/binomial-distribution.html