THEORETICAL DISTRIBUTIONS

THEORETICAL DISTRIBUTIONS

Discrete probability distributions and

Continuous probability distributions

Discrete probability distributions

- Binomial distribution
- Poisson distribution
- Geometric distribution
- Negative binomial distribution

- A fixed number of observations (trials), n
 - e.g., 15 tosses of a coin; 20 patients; 1000 people surveyed
- A binary random variable
 - e.g., head or tail in each toss of a coin; defective or not defective light bulb
 - Generally called "success" and "failure"
 - Probability of success is p, probability of failure is
 1 p
- Constant probability for each observation
 - e.g., Probability of getting a tail is the same each time we toss the coin

Definition: Binomial distribution

Suppose that n independent experiments, or trials, are performed, where n is a fixed number, and that each experiment results in a "success" with probability p and a "failure" with probability l-p. The total number of successes, X, is a binomial random variable with parameters n and p.

We write: **X ~ Bin (n, p)** {reads: "X is distributed binomially with parameters n and p}

And the probability that X=r (i.e., that there are exactly r successes) is:

$$P(X=r) = \binom{n}{r} p^r (1-p)^{n-r}$$

Example:1

Take the example of 5 coin tosses. What's the probability that you flip exactly 3 heads in 5 coin tosses?

Solution:

One way to get exactly 3 heads: HHHTT

What's the probability of this <u>exact</u> arrangement? $P(heads)x P(heads) x P(heads) x P(tails) x P(tails) = (1/2)^3 x (1/2)^2$

Another way to get exactly 3 heads: THHHT Probability of this exact outcome = $(1/2)^1 x (1/2)^3 x (1/2)^1 = (1/2)^3 x (1/2)^2$

In fact, $(1/2)^3 x (1/2)^2$ is the probability of each unique outcome that has exactly 3 heads and 2 tails.

So, the overall probability of 3 heads and 2 tails is: $(1/2)^3 x (1/2)^2 + (1/2)^3 x (1/2)^2 + (1/2)^3 x (1/2)^2 +$ for as many unique arrangements as there are—but how many are there?

$\binom{5}{3}$	ways to arrange 3 heads in 5 trials	
C = 51/3121 = 10		

Outcome	Probability	
ТНННТ	$(1/2)^3 x (1/2)^2$	
НННТТ	$(1/2)^3 x (1/2)^2$	
ТТННН	$(1/2)^3 x (1/2)^2$	
HTTHH	$(1/2)^3 x (1/2)^2$	
ННТТН	$(1/2)^3 x (1/2)^2$	
THTHH	$(1/2)^3 x (1/2)^2$	
HTHTH	$(1/2)^3 x (1/2)^2$	
ННТНТ	$(1/2)^3 x (1/2)^2$	
THHTH	$(1/2)^3 x (1/2)^2$	
HTHHT	$(1/2)^3 x (1/2)^2$	
10 arrangements $x (1/2)^3 x (1/2)^2$		

The probability
of each unique
outcome (note:
they are all
equal)

∴ P(3 heads and 2 tails) =
$$\binom{5}{3}$$
 x P(heads)³ x P(tails)²

$$=10 \times (\frac{1}{2})^{5} = 31.25\%$$

Example:2 The probability that a radio manufactured by a company will be defective is 1/10. If 15 of such radio's are inspected, find the probability that (a) Exactly 3 defectives; (b) At least 1 defective; and (c) None will be defective.

Solution:

Given data : n = 15; p = 0.1; and q = 0.9

Since lot size is unknown theoretically correct distribution is binomial.

- (a) Exactly 3 defectives = P(3)=15C3(0.1)3(0.9)12=0.1285
- (b)at least one defective

The required probability =
$$P(r \ge 1) = P(1) + P(2) + = 1 - P(r < 1)$$

This can be written as = $1 - P(0) = 1 - \{15C0(0.1)0 \cdot (0.9)15\}$
= $1 - \{1 \times 0.2058\} = 0.7941$

(c)None will be defective= P(0) = 0.2059.

Example:3

The probability that a bulb produced by a factory will fuse after 10 days is 0.05. Find the probability out of 5 such bulbs

(a) None; (b) Not more than one; (c) Greater than one and (d) At least one; will fuse after 400 days of use.

Solution:

Given data : P = 0.05; n = 5; and q = 0.95

- (a) None will fuse is P(r = 0) = 0.7738
- (b) Not more than one will fuse is P(r < 1) = 0.9974
- (c) More than one will fuse is $P(r > 1) = 1 \{P(0) + P(1)\} = 0.0226$
- (d) At least one will fuse is P(r > 1) = 1 P(0) = 0.2262.

Mean and Variance of Binomial Distribution

- If $X \sim B(n, p)$ (that is, X is a binomially distributed random variable), then the expected value of X is
- E[X]=npand the variance is
 - Var[X]=np(1-p)

Definition: Poisson distribution

The Poisson distribution models counts, such as the number of new cases of SARS that occur in women in New England next month.

The distribution tells you the probability of all possible numbers of new cases, from 0 to infinity.

If X = # of new cases next month and $X \sim Poisson(\lambda)$, then the probability that X = k (a particular count) is:

$$p(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Poisson distribution

Example: 1

A small life insurance company has determined that on the average it receives 6 death claims per day. Find the probability that the company receives at least seven death claims on a randomly selected day.

$$P(x \ge 7) = 1 - P(x \le 6) = 0.393697$$

Example: 2

The number of traffic accidents that occurs on a particular stretch of road during a month follows a Poisson distribution with a mean of 9.4. Find the probability that less than two accidents will occur on this stretch of road during a randomly selected month.

$$P(x < 2) = P(x = 0) + P(x = 1) = 0.000860$$

Poisson distribution

Example:3

Suppose that a rare disease has an incidence of 1 in 1000 person-years. Assuming that members of the population are affected independently, find the probability of k cases in a population of 10,000 (followed over 1 year) for k=0,1,2The expected value (mean) $=\lambda = .001*10,000 = 10$ 10 new cases expected in this population per year .Then

$$P(X = 0) = \frac{(10)^{0} e^{-(10)}}{0!} = .0000454$$

$$P(X = 1) = \frac{(10)^{1} e^{-(10)}}{1!} = .000454$$

$$P(X = 2) = \frac{(10)^{2} e^{-(10)}}{2!} = .00227$$

Poisson Mean and Variance

Mean

$$\mu = \lambda$$

Variance and Standard Deviation

$$\sigma^2 = \lambda$$

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

For a Poisson random variable, the variance and mean are the same!

 λ = expected number of hits in a given time period where

Definition: Geometric Distribution

In a series of Bernoulli trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until the first success. Then X is a geometric random variable with parameter 0 and

$$f(x) = (1 - p)^{x-1}p$$
 $x = 1, 2, ...$ (3-9)

Example:1

The probability that a wafer contains a large particle of contamination is 0.01. If it is assumed that the wafers are independent, what is the probability that exactly 125 wafers need to be analyzed before a large particle is detected?

Let X denote the number of samples analyzed until a large particle is detected. Then X is a geometric random variable with p = 0.01. The requested probability is

$$P(X = 125) = (0.99)^{124}0.01 = 0.0029$$

Example:2

Over a very long period of time, it has been noted that on Friday's 25% of the customers at the drive-in window at the bank make deposits. What is the probability that it takes 4 customers at the drive-in window before the first one makes a deposit?

Solution:

This problem is a geometric distribution problem with π = 0.25. Let x = number of customers at the drive-in window before a customer makes a deposit. The desired probability is

$$p(4) = (.75)^{4-1}(.25) = 0.0117$$

Example:3

From past experience it is known that 3% of accounts in a large accounting population are in error.

What is the probability that 5 accounts are audited before an account in error is found?

Solution:

$$P(X = 5) = P(1st \ 4 \ correctly \ stated) P(5th \ in \ error)$$

= $(0.97)^4 (0.03)$

What is the probability that the first account in error occurs in the first five accounts audited?

Solution:

$$P(X \le 5) = 1 - P(First 5 correctly state)$$

= 1 - (0.97)⁵

The Mean and Standard Deviation of a Geometric Random Variable

If X is a geometric random variable with probability of success p on each trial, then the **mean**, or **expected value**, of the random variable, that is, the expected number of trials required to get the first success, is $\mu = 1/p$. The variance of X is $(1 - p)/p^2$.

Definition: Negative Binomial Distribution

A generalization of a geometric distribution in which the random variable is the number of Bernoulli trials required to obtain r successes results in the **negative binomial distribution**.

In a series of Bernoulli trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until r successes occur. Then X is a **negative binomial random variable** with parameters $0 and <math>r = 1, 2, 3, \ldots$, and

$$f(x) = {x-1 \choose r-1} (1-p)^{x-r} p^r \qquad x = r, r+1, r+2, \dots$$
 (3-11)

Example:1

Victor is a basketball player. He is a 60% free throw shooter. During the season, what is the probability that Victor makes his second free throw on his fourth shot.

Solution: Step 1:

The probability of success (P) is 0.60 (because, Victor is a 60% free throw shooter. That means his probability of making a free throw is 0.60).

And the number of trials (x) is 4, and the number of successes (r) is 2.

$$=> P = 0.60, 1 - P = 1 - 0.60 = 0.40$$

 $x = 4, r = 2$

Step 2:

To solve this problem, we enter these values into the negative binomial formula.

Thus, the probability that Victor will make his second successful free throw on his fourth shot is 0.1728.

Example:2

What is the probability that all three servers fail within five requests? The probability is $P(X \le 5)$ and

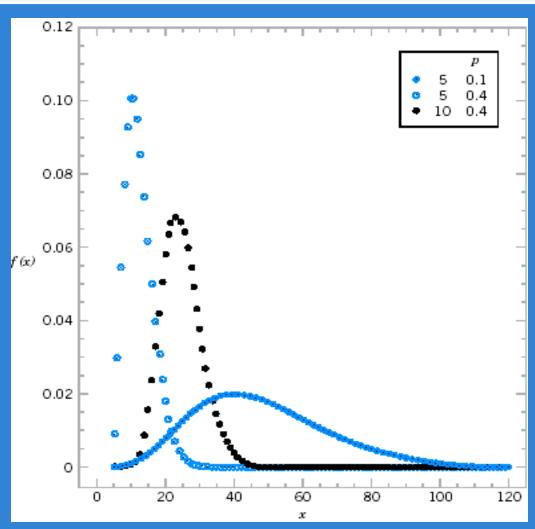
$$P(X \le 5) = P(X = 3) + P(X = 4) + P(X = 5)$$

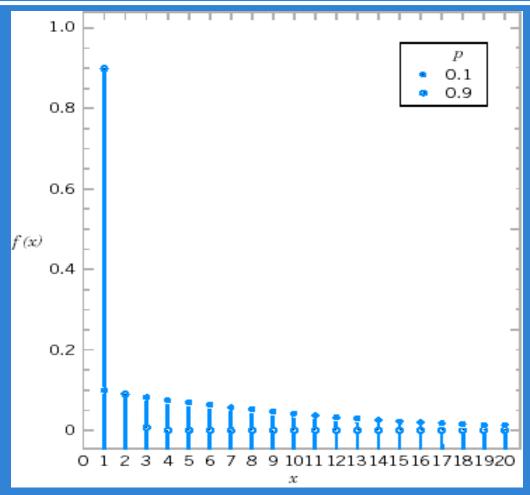
$$= 0.0005^{3} + {3 \choose 2} 0.0005^{3} (0.9995) + {4 \choose 2} 0.0005^{3} (0.9995)^{2}$$

$$= 1.25 \times 10^{-10} + 3.75 \times 10^{-10} + 7.49 \times 10^{-10}$$

$$= 1.249 \times 10^{-9}$$

Negative binomial distributions for selected values of the parameters r and p.





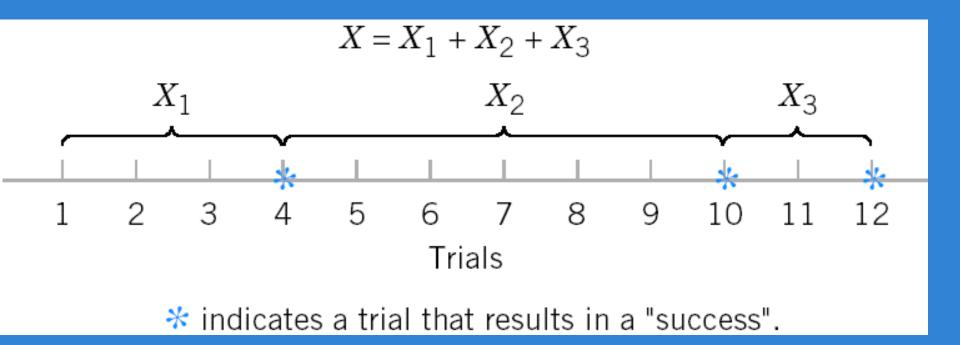
Geometric distributions for selected values of the parameter *p*.

Mean and variance of Negative Binomial Distribution

If X is a negative binomial random variable with parameters p and r,

$$\mu = E(X) = r/p$$
 and $\sigma^2 = V(X) = r(1-p)/p^2$ (3-12)

Geometric and Negative Binomial Distributions



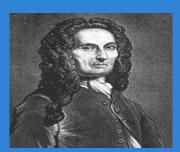
Negative binomial random variable represented as a sum of geometric random variables.

Continuous probability distributions

- Normal distribution
- Exponential distribution

The Normal Distribution

 Discovered in 1733 by de Moivre as an approximation to the binomial distribution when the number of trails is large Derived in 1809 by Gauss



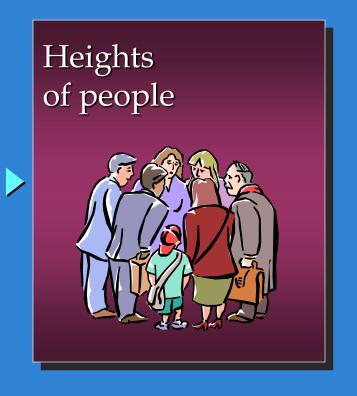
- Importance lies in the Central Limit Theorem, which Moivre (1667-states that the sum of a large number of independent 1754) random variables (binomial, Poisson, etc.) will approximate a normal distribution
 - Example: Human height is determined by a large number of factors, both genetic and environmental, which are additive in their effects.
 Thus, it follows a normal distribution.



Karl F. Gauss (1777-1855)

- The <u>normal probability distribution</u> is the most important distribution for describing a continuous random variable.
- It is widely used in statistical inference.

It has been used in a wide variety of applications:





It has been used in a wide variety of applications:





Normal Probability Density Function

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

where:

 $\mu = \text{mean}$

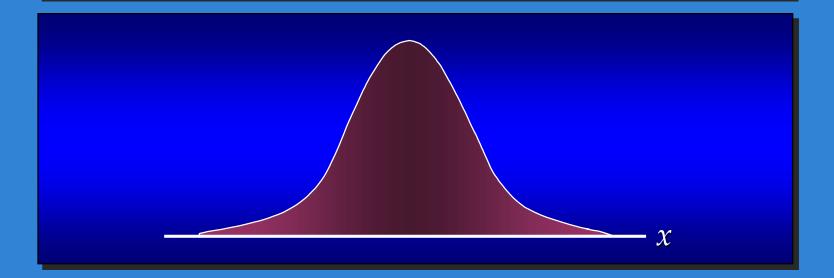
 σ = standard deviation

 $\pi = 3.14159$

e = 2.71828

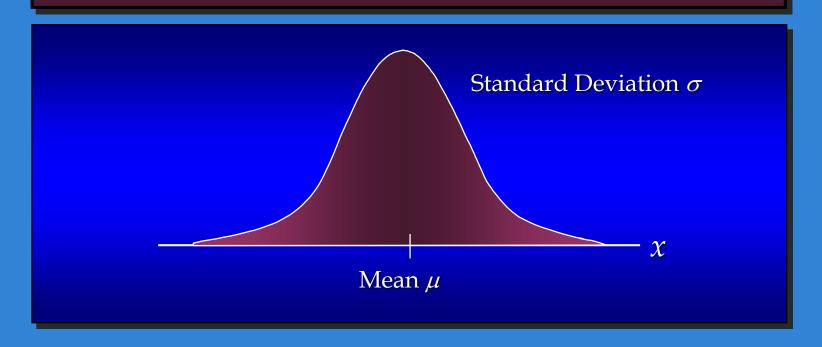
Characteristics

The distribution is <u>symmetric</u>; its skewness measure is zero.



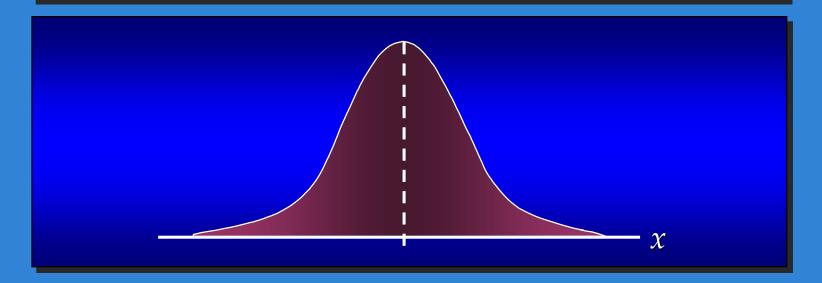
Characteristics

The entire family of normal probability distributions is defined by its $\underline{\text{mean}} \mu$ and its $\underline{\text{standard deviation}} \sigma$.



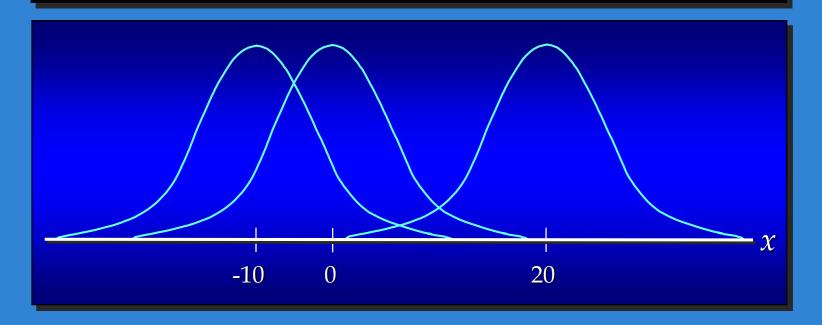
Characteristics

The <u>highest point</u> on the normal curve is at the <u>mean</u>, which is also the <u>median</u> and <u>mode</u>.



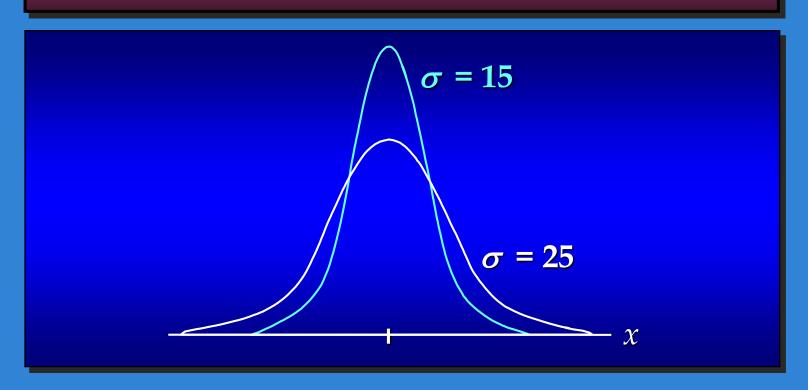
Characteristics

The mean can be any numerical value: negative, zero, or positive.



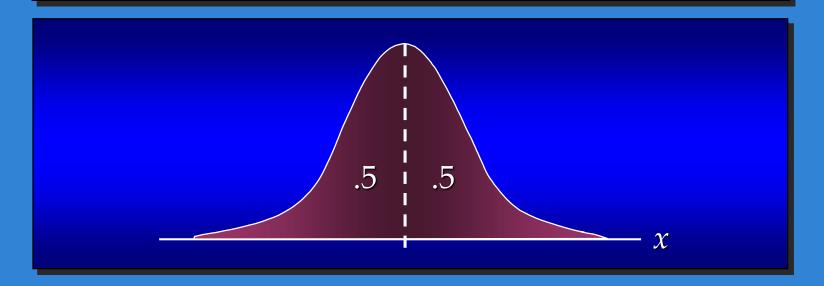
Characteristics

The standard deviation determines the width of the curve: larger values result in wider, flatter curves.



Characteristics

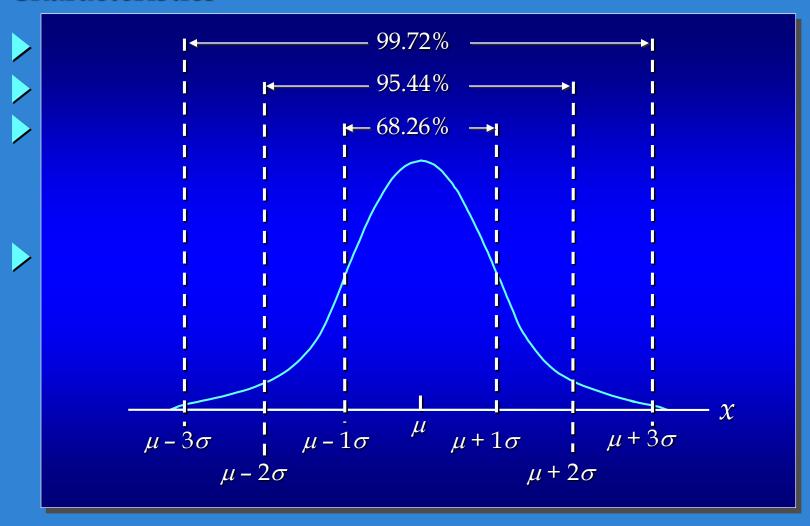
Probabilities for the normal random variable are given by <u>areas under the curve</u>. The total area under the curve is 1 (.5 to the left of the mean and .5 to the right).



Characteristics

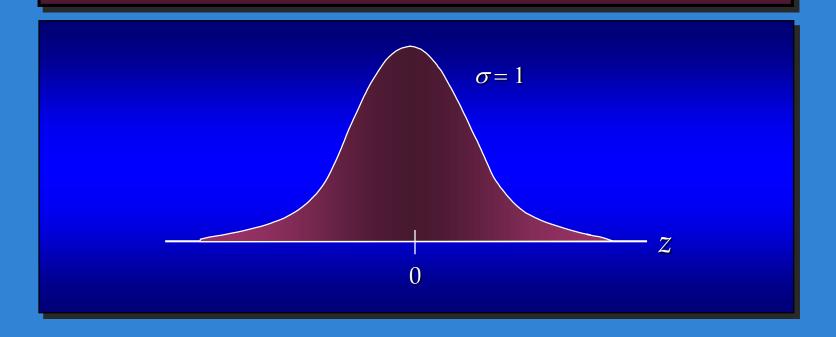
- 68.26% of values of a normal random variable are within +/-1 standard deviation of its mean.
- 95.44% of values of a normal random variable are within +/- 2 standard deviations of its mean.
- 99.72% of values of a normal random variable are within +/- 3 standard deviations of its mean.

Characteristics

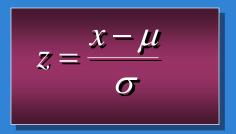


A random variable having a normal distribution with a mean of 0 and a standard deviation of 1 is said to have a <u>standard normal probability</u> distribution.

The letter z is used to designate the standard normal random variable.

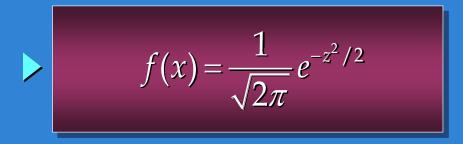


Converting to the Standard Normal Distribution



We can think of z as a measure of the number of standard deviations x is from μ .

Standard Normal Density Function



where:

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

 $\pi = 3.14159$
 $e = 2.71828$

Example: Pep Zone

Pep Zone sells auto parts and supplies including a popular multi-grade motor oil. When the stock of this oil drops to 20 gallons, a replenishment order is placed.

Example: Pep Zone

The store manager is concerned that sales are being lost due to stockouts while waiting for an order. It has been determined that demand during replenishment lead-time is normally distributed with a mean of 15 gallons and a standard deviation of 6 gallons. The manager would like to know the

probability of a stockout, P(x > 20).

Solving for the Stockout Probability

Step 1: Convert x to the standard normal distribution.

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

= (20 - 15)/6
= .83

Step 2: Find the area under the standard normal curve to the left of z = .83.

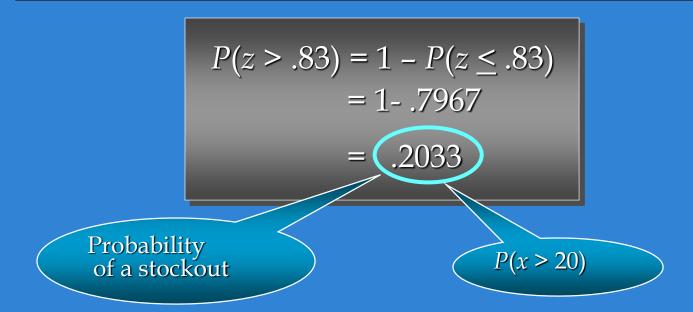
see next slide

Cumulative Probability Table for the Standard Normal Distribution

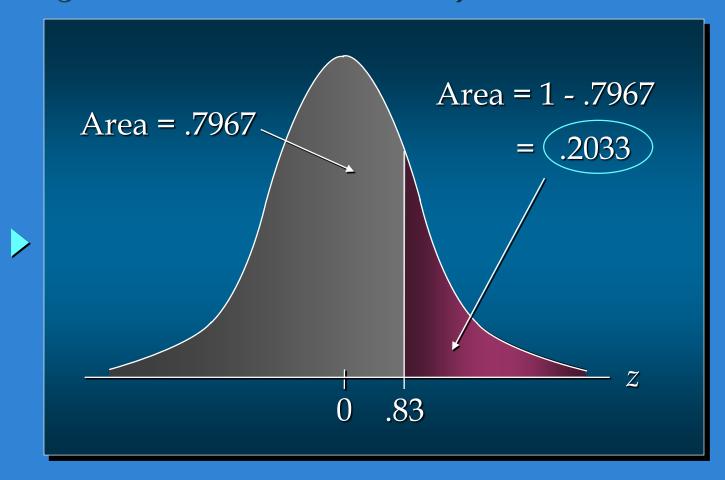
Z	.00	.01	.02	(.03)	.04	.05	.06	.07	.08	.09
•		•			•	•	•	•	•	•
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	7001	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
•	•	•	•	•		•	•	•	•	•

Solving for the Stockout Probability

Step 3: Compute the area under the standard normal curve to the right of z = .83.

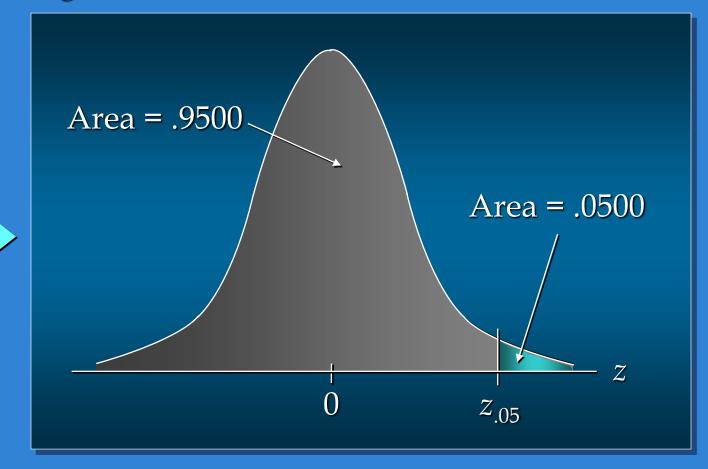


Solving for the Stockout Probability



If the manager of Pep Zone wants the probability of a stockout to be no more than .05, what should the reorder point be?

Solving for the Reorder Point



Solving for the Reorder Point

Step 1: Find the *z*-value that cuts off an area of .05 in the right tail of the standard normal distribution.

Z	.00	.01	.02	.03 (.04	.05	.06	.07	.08	.09	
					•	•					
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441	
1.6	19452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545	
1.7	.9554	.9564	.9573	.9582	.9591	Q599	.9608	.9616	.9625	.9633	
1.8	.9641	.9649	.9656	.9664	.9671		9686	.9693	.9699	.9706	
1.9	.9713	.9719	.972	We look up the complement of the tail area (105 = .95)							
	•	•	tail area $(105 = .95)$								

Solving for the Reorder Point

Step 2: Convert $z_{.05}$ to the corresponding value of x.

$$x = \mu + z_{.05}\sigma$$
= 15 + 1.645(6)
= 24.87 or 25

A reorder point of 25 gallons will place the probability of a stockout during leadtime at (slightly less than) .05.

Solving for the Reorder Point

By raising the reorder point from 20 gallons to 25 gallons on hand, the probability of a stockout decreases from about .20 to .05.

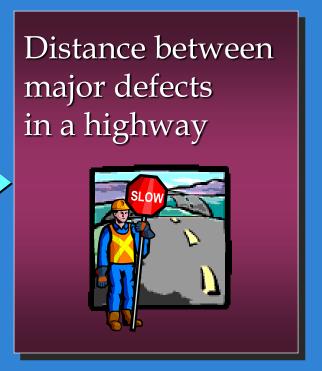
This is a significant decrease in the chance that Pep Zone will be out of stock and unable to meet a customer's desire to make a purchase.

The exponential probability distribution is useful in describing the time it takes to complete a task.

The exponential random variables can be used to describe:







Density Function

$$f(x) = \frac{1}{\mu} e^{-x/\mu} \quad \text{for } x \ge 0, \, \mu > 0$$

where:
$$\mu$$
 = mean e = 2.71828

Cumulative Probabilities



$$P(x \le x_0) = 1 - e^{-x_0/\mu}$$

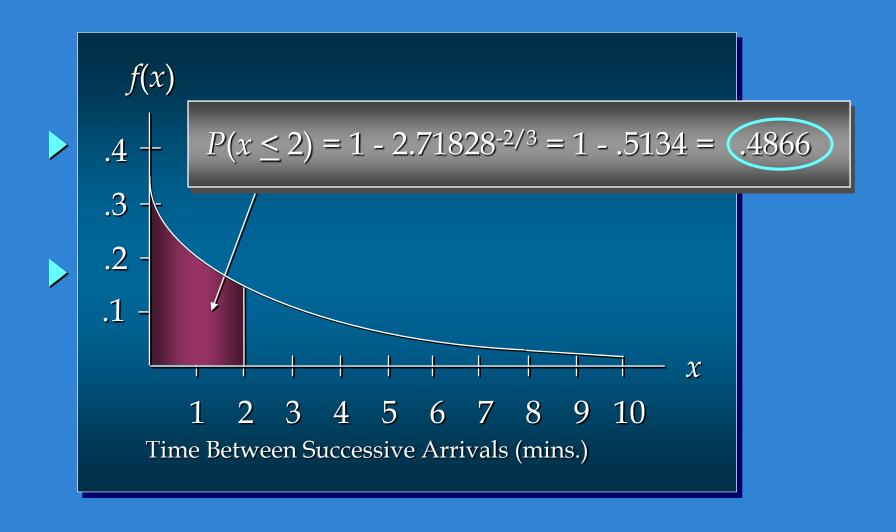
where:

 x_0 = some specific value of x

Example: 1(Al's Full-Service Pump)

The time between arrivals of cars at Al's full-service gas pump follows an exponential probability distribution with a mean time between arrivals of 3 minutes. Al would like to know the probability that the time between two successive arrivals will be 2 minutes or less.





- A property of the exponential distribution is that the mean, μ , and standard deviation, σ , are equal.
- Thus, the standard deviation, σ , and variance, σ^2 , for the time between arrivals at Al's full-service pump are:

$$\sigma = \mu = 3$$
 minutes

$$\sigma^2 = (3)^2 = 9$$

The exponential distribution is skewed to the right.

The skewness measure for the exponential distribution is 2.

Example:2:

Let take a customer who goes for shopping and lets suppose the time he spends in the shop is exponentially distributed with a mean value 15 minutes. Then what will be the probability that the customer will spend more than 20 minutes in shopping? What is the probability that the customer will spend more than 20 minutes in the bank given that he is still in the shop after 15 minutes?

Solution:

Here mean is given as 15 minutes Now the mean of an exponential distribution is $1/\lambda$ So $1/\lambda = 10$

So $\lambda = 1/10$

First we need the probability that the customer will spend more than 20 minutes in shopping

So P
$$(X > 20) = 0.26$$

For the next part we need to find the probability that the customer will spend more than 20 minutes in the bank given that he is still in the shop after 15 minutes.

So
$$P(X > 20)$$
 at $X > 15 = P(X > 5) = 0.604$

Mean and Standard deviation of Exponential Distribution

• The mean or expected value of an exponentially distributed random variable X with rate parameter λ is given by

$$E[X]=1/\lambda$$

The standard deviation of X is given by

$$S.D[X]=1/\lambda$$

1.A random sample is to be selected from a lot of 15 articles, 4 of which are nonconforming.

What is the probability that the sample will contain less than 3 non-containing items.

- **2.** A lot of 200 capacitors is known to have 20% defective. If a sample of 50 have to be inspected from a lot what is probability of finding (a) Exactly one defective;
- (b)(b) More than 3 defectives; and (c) Less than 3 defectives.
- **3**.From a lot of 15 missiles 5 are selected at random and fired. If the lot contains 4defective missiles that will not fire. What is probability ability that
- (a) All will 5 fire; and
- (b) At most 2 will not fire.

- **4.** To avoid detection at customs, a travelers has planed 6 narcotic tablets in a bottle containing 9 vitamin pills that are similar in appearance. If the customs official selects the 3 of tablets at random for analysis what is the probability that the traveler will be arrested from illegal possession of narcotic pills.
- **5**. When a certain quality characteristic of a manufacturing product falls below its(LSL) the product is designated as class I defective. When product falls above (USL). The
- product is designated as defective class II. A sample of 5 articles is taken from a lot of 15articles that contains 2 defectives of class I and three defectives of class II. What is the probability that the sample will contain no defective of class I and exactly one defective of class II.

- **6.** A random sample of 20 is to be selected from a lot containing 200 articles, 12 of which are defective. Determine probability that the sample will contain Less than 2 defectives.
- **7.** A lot of 20 radio sets containing 3 defective sets. If 5 sets are selected at random
- what is probability that
- (a) Less than 3 defectives
- (b) (b) Exactly 2 defectives
- (c) Greater than one defective
- (d) Less than or equal to one defective and
- (e) Greater than 2 defective

- **8**. The probability that a casting produced by a certain foundry has blow holes of 0.004. Find the probability that less than 4 of the next 2000 items produced by the foundry has the blow holes.
- **9**. If an inventory study it was determined that average the demand for a particular item at a warehouse were made 5 items per day. What is the probability that on a given day this term is required :
- (a) More than 3 times ; (b) Not at all.
- **10**. A manufacturer know from experience that five out of 1000 shells is defective. What is the probability that in a batch of 600 :
- (a) Exactly 4 defective; and (b) More than 4 defectives n = 600.
- **11**. A Secretary makes 2 errors per page on the average. What is the probability that on the next page she makes :
- (a) 4 or more errors; and (b) No error

12. In a precision grinding of a complicated part is more economical to rework. Than top scrap. It is desired to establish a rework percentage of 1.25. Assume data follows Normal distributions. The standard deviation is 0.01 upper specification limits is 25.38.

Find the mean with this estimated mean what % of the terms are falling below 25.1 mm.

- **13**. If weights of 300 students are normally distributed. With a mean of 68 kg and variance of 9 kg how many students have their weights:

 (a) Greater than 72 kg; (b) Less than 64 kg; and (c) Between 64 and 72 kg
- **14**. A transistor radio operator on 3 size, 1.5 volts battery so that nominally it operates on 4.5 volts. Suppose actual voltage of single new battery is nominally distributed with mean 1.5 volts and standard deviation of 0.2 volts, the radio will not operate properly at the outset if voltage falls outside the range 4 to 5 volts.

What is the probability that radio will not operate properly.

References

- [1] A. Papoulis, Probability, Random Variables, and Stochastic Processes. New York: McGraw-Hill,1975.
- [2] A. Papoulis, Probability, Random Variables, and Stochastic Processes. New York: McGraw-Hill, third edition ed., 1991.
- [3] V. K. Rohatgi, An Introduction to Probability and Mathematical Statistics. New York: John Wiley and Sons, 1976.
- [4] W. Feller, An Introduction to Probability Theory and its Applications, Volume I. New York: Wiley, 1966.

THE END