Chapter 8 - Probability

Statisticians are concerned with the presentation and analysis of data that occur in a planned study. Statistical tools are used in quality control, and you will be learning to use some of them in this course. The purpose of this module is to teach you some basic concepts from statistics that you will need. These notes are organized to first present some probability theory (Chapter 8), the underlying mathematical language of statistics, and then move on to the relevant statistical ideas (Chapter 9). A list of reference books is included at the end of these notes. Additional tutorial material will be provided separately.

Basic Definitions

When reading the following definitions, it is helpful to think about the process of conducting a statistical experiment, which includes the measurement of data and the subsequent analysis of it.

A *population* is defined to be the totality of the observations relevant to a study. Populations possess a certain characteristic of interest.

Example: The heights of residents of a certain city, or the set of all 50,000 cans of Brand A soup made in July are two examples of populations.

An *experiment* is any process that generates a set of raw data.

Example: The measurement of heights of people, or the measurement of weights of cans of soup, are both experiments.

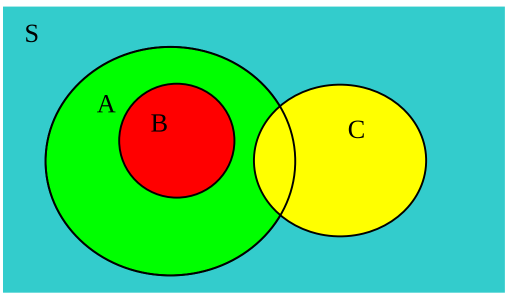
The *sample space* is the set of all possible outcomes of an experiment; it is denoted S. An *element* of a sample space is called an outcome. A sample is a subset of a population.

Example: S = {500 cans of soup made in July}, a selections of cans.
S = {H,T}, the possible outcomes of a coin toss.

A *parameter* is a characteristic of a population. A *statistic* is a characteristic of a sample.

Example: The average weight of all 50,000 cans of Brand A soup made in July is a parameter. The average weight of a sample of 500 cans of Brand A soup is a statistic.

An *event* is a subset of a sample space. Events consist of certain occurances of interest. A *simple event* consists of precisely one element; other events are *compound*.



Events in a sample space (Venn diagram).

An event is said to *occur* if when the experiment is conducted, the outcome belongs to the event. In the Venn diagram, the subsets *A,B,C* are events in the sample space *S*.

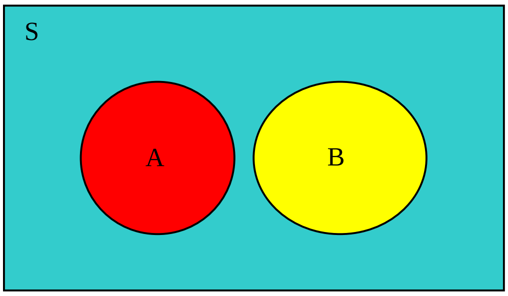
Example: *S = {HH,HT,TH,TT}*, the possible outcomes of the toss of two coins.
E = {HH,HT,TH}, the event that at least one *H* occurred.

Recall the definitions of and notation for set *union*, *intersection*, and *complement*:

$$A \cup B = A + B, \quad A \cap B = AB, \quad A' = A^c = \bar{A}$$

Two events *A* and *B* are *mutually exclusive* if they contain no elements in common (their intersection is the empty set):

$$A \cap B = \varnothing$$



Mutually exclusive events.

Example: *S = {HH,HT,TH,TT}*, the possible outcomes of the toss of two coins.
A = {HH,HT}, *B = {TT}* are mutually exclusive.

Events *A* and *B* are *independent* if the outcome of one has no influence on the outcome of the other.

Counting

A *permutation* is an arrangement of all or part of a set of objects. The number of permutations of *n* distinct objects is *n!* (*n* factorial; *n!=n.(n-1).(n-2)…2.1*, *0!=1*).

The number of permutations of *n* distinct objects taken *r* at a time is

$$P_r^n = \frac{n!}{(n-r)!}$$

The number of combinations of *n* distinct objects taken *r* at a time, without regard to order, is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Example: Consider the set {a,b,c,d}. *(i)* There are *4!=24* ways of arranging these four distinct objects: (abcd), (abdc), …, etc. *(ii)* There are *4!/(4-2)!=12* ways of arranging these four objects two at a time: (ab), (ac), (ad), (ba),…, etc. *(iii)* There are *4!/[(2!(4-2)!]=6* ways of combining these four objects taken two at a time irrespective of order: (ab), (ac), (ad), (bc), (bd), (cd).

Probability

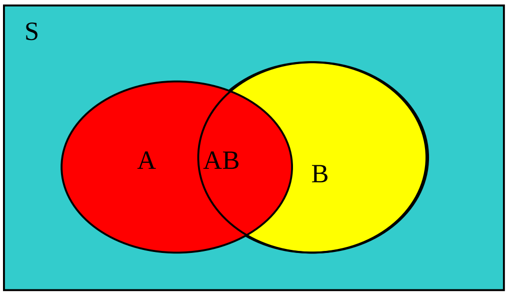
The *probability P(A)* of an event *A* is a real number satisfying

$$0 \le P(A) \le 1, \qquad P(\varnothing) = 0, \qquad P(S) = 1$$

If an experiment can result in *N equally likely* outcomes, then the probability of an event *A* consisting of *n* simple events, then

$$P(A) = \frac{n}{N}$$

Example: *S = {HH,HT,TH,TT}*, the possible outcomes of the toss of two coins.
A = {HH,HT,TH}. Then *P(A)=3/4*.



Probability of union of two events.

For any two events *A* and *B*,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Example: *S = {HH,HT,TH,TT}*, the possible outcomes of the toss of two coins.
A = {HH,TT}, *B = {HT,TT}*. Then

$$P(A \cup B) = 1/2 + 1/2 - 1/4 = 3/4$$

If *A* and *B* are *mutually exclusive* events, then

$$P(A \cup B) = P(A) + P(B)$$

Example: *S = {HH,HT,TH,TT}*, the possible outcomes of the toss of two coins.
A = {HH}, *B = {TT}*. Then

$$P(A \cup B) = 1/4 + 1/4 = 1/2$$

If *A* and *A'* are *complementary events*, i.e.

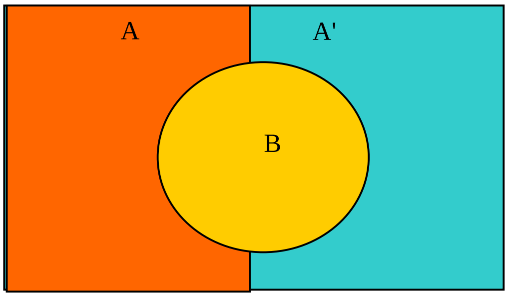
$$A \cup A' = S, \qquad A \cap A' = \varnothing$$

then

$$P(A') = 1 - P(A)$$

Sometimes we need the probability of an event *B* when we know that an event *A* has occurred. If we know that *A* has occurred, then we know that not all outcomes in *B* can occur, and so the probability needs to be revised. The *conditional probability P(B|A)* is the probability of *B* occurring given that *A* has occurred:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \quad if \quad P(A) > 0$$



Conditional probability.

Example: $S = \{HH,HT,TH,TT\}$, the possible outcomes of the toss of two coins. $A = \{HH,HT\}$, $B = \{HT\}$. Then

$$P(B \mid A) = \frac{1/4}{1/2} = 1/2$$

so that if *A* has occurred, the likelihood of *B* has increased (from 1/4 to 1/2).

If both *A* and *B* can occur in an experiment, then

$$P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$$

Independence means that $P(B|A)=P(B)$, and $P(A|B)=P(A)$. Thus events *A* and *B* are independent if and only if

$$P(A \cap B) = P(A)P(B)$$

Bayes' Rule:

Consider a *partition*

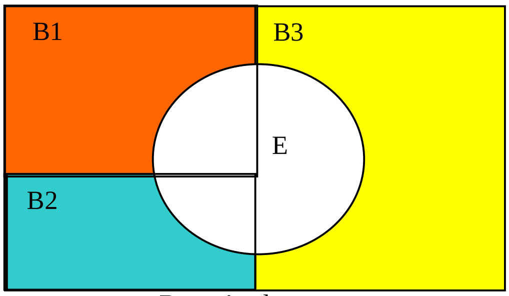
$$\{B_1, B_2, \ldots, B_n\}$$

(mutually exclusive events covering completely the sample space) of a sample space S. Then any event E can be expressed as

$$E = \bigcup_{i=1}^{n} B_i$$

and so

$$P(E) = \sum_{i=1}^{n} P(E \cap B_i)$$



Bayes' rule

Baye's rule states that

$$P(B_i \mid E) = \frac{P(B_i)P(E \mid B_i)}{\sum_{j=1}^{n} P(B_j)P(E \mid B_j)}$$

Example: Three politicians have been nominated for the office of Prime Minister. The probability that Mr Smith will be elected is 0.3, the probability that Ms Jones will be elected is 0.5, and the probability that Ms Williams will be elected is 0.2. If Mr Smith is elected, the probability of a tax increase is 0.8. If Ms Jones or Ms Williams are elected, the probabilities for a tax increase are 0.1 and 0.4, respectively. If after the election taxes were increased, what is the probability that Ms Williams was elected Prime Minister? *Solution*: Consider the following events

E: the politician elected increased taxes
B1: Mr Smith is elected
B2: Ms Jones is elected
B3: Ms Williams is elected

Now

$$P(B_1 \cap E) = P(B_1)P(E \mid B_1) = 0.3 \cdot 0.8 = 0.24$$
$$P(B_2 \cap E) = P(B_2)P(E \mid B_2) = 0.5 \cdot 0.1 = 0.05$$
$$P(B_3 \cap E) = P(B_3)P(E \mid B_3) = 0.2 \cdot 0.4 = 0.08$$

and so, using Bayes' rule,

$$P(B_3 \mid E) = \frac{0.08}{0.24 + 0.05 + 0.08} = 0.216$$

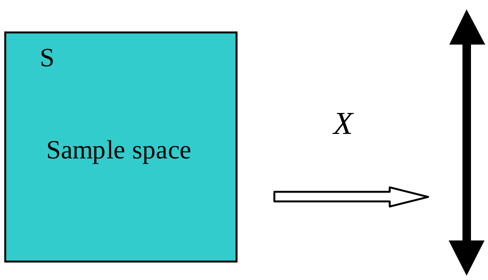
Thus the probability that Ms Williams was elected given that taxes increased after the election is 0.216.

Random Variables

Statisticians are often interested in numerical values associated with outcomes of experiments. This provides a convenient summary of information available in an experiment.

A *random variable* is a real-valued function defined on the sample space. Capital letters such as X are used to denote random variables, while lower case letters x correspond to values that the random variable can take.

Random variables can be *discrete* or *continuous,* depending on whether the range is discrete or continuous. These typically come from count or measured data, respectively.



Example: *S = {HH,HT,TH,TT},* the possible outcomes of the toss of two coins. X = number of heads (*X(HH)=2, X(HT)=1, X(TH)=1, X(TT)=0*) This is a discrete random variable.

Example: S = {500 cans of soup made in July}, X = weight of each can. This is a continuous random variable.

A *probability distribution* is used to model the (idealized) randomness of a random variable. We will give separate definitions in the discrete and continuous cases.

Discrete Probability Distributions

Let X be a discrete random variable, taking discrete values labeled x. The quantity

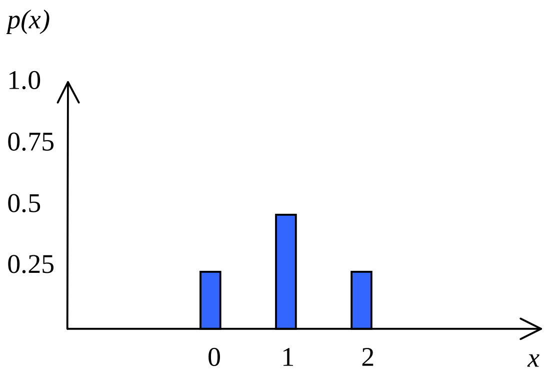
$$p(x) = P(X = x)$$

is called a *discrete probability distribution,* and it must satisfy the following conditions

$$\sum_x p(x) = 1, \quad p(x) \geq 0$$

Note that the sum is over a finite (e.g. $1,2,\ldots,N$) or countable, infinite, range (e.g. $1,2,\ldots$.).

Example: $S = \{HH,HT,TH,TT\}$, the possible outcomes of the toss of two coins. X = number of heads, range $x = 0,1,2$. $p(0) = 1/4$, $p(1) = 1/2$, $p(2) = 1/4$.



The *cumulative distribution function* (CDF) is defined by

$$F(x) = P(X \leq x) = \sum_{t \leq x} p(t)$$

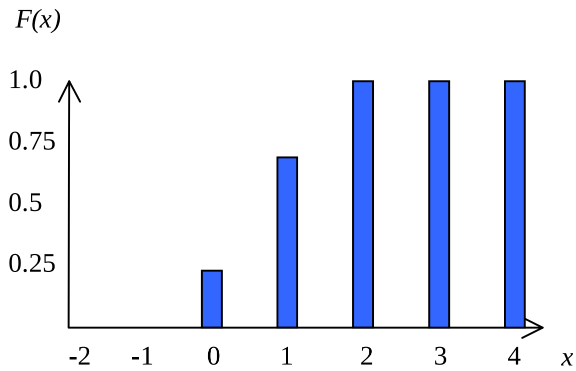
The CDF is a nondecreasing function satisfying

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to +\infty} F(x) = 1$$

Example: $S = \{HH,HT,TH,TT\}$, the possible outcomes of the toss of two coins. X = number of heads, range x = 0,1,2.

$$F(x) = 0,\ x < 0, \quad F(0) = 0.25$$

$$F(1) = 0.75, \quad F(x) = 1,\ x \geq 2$$



Note that if either *p(x)* or *F(x)* is known, one can readily evaluate probabilities of events of interest.

The *expected value* (or *mathematical expectation*) of a discrete random variable *X* is defined by

$$E[X] = \sum_x xp(x)$$

The summation is over the range of values of the random variable.

Example: $S = \{HH,HT,TH,TT\}$, the possible outcomes of the toss of two coins. *X* = number of heads. Then

$P(\text{one head}) = F(1)-F(0) = 0.75 - 0.25 = 0.5 = p(1) = P(X=1).$

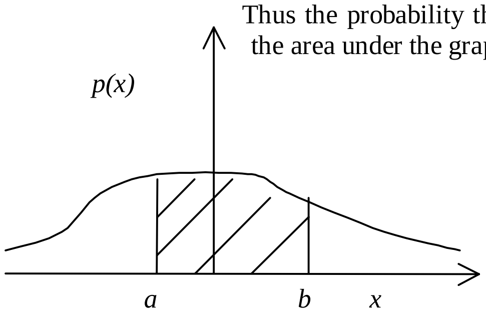
$E[X] = 0\cdot 0.25 + 1\cdot 0.5 + 2\cdot 0.25 = 1$

Continuous Distributions

Let *X* be a continuous random variable, taking continuous values labeled *x*. The quantity *p(x)* is called a *probability density* function if

$$P(a < X < b) = \int_a^b p(x)dx, \quad p(x) \geq 0, \quad \int_{-\infty}^{+\infty} p(x)dx = 1$$

Thus the probability that *X* takes values between points *a* and *b* is given by the area under the graph of *p(x)* between these points.



We will assume that *p(x)* is a real valued integrable function, with no impulses. Then for any point *a*,

$$P(X = a) = 0$$

and

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

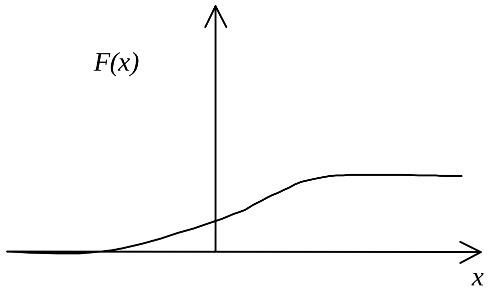
So the probability that *X* takes *exactly* the value *a* is zero; however, we have the interpretation

$$p(a)\Delta x = P(a - \Delta x/2 < X < a + \Delta x/2)$$

This expression gives the probability that X takes a value near a.

The *cumulative distribution function* (CDF) of a continuous random variable X with density $p(x)$ is given by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} p(t)dt$$



The CDF is a nondecreasing function satisfying

$$\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to+\infty} F(x) = 1$$

Note that if either $p(x)$ or $F(x)$ is known, one can readily evaluate probabilities:

$$P(a < X < b) = F(b) - F(a) = \int_{a}^{b} p(x)dx$$

The *expected value* (or *mathematical expectation*) of a continuous random variable X is defined by

$$E[X] = \int_{-\infty}^{+\infty} xp(x)dx$$

Example: Let X be a continuous random variable with density $p(x)=1$ if x lies between 0 and 1, and $p(x)=0$ otherwise. So X is *uniformly* distributed on the interval [0,1]. Then $E[X]=0.5$ and

$$F(x) = 0,\ x \le 0, \quad F(x) = x,\ x \in [0,1], \quad F(x) = 1,\ x \ge 1$$

From the definitions, it is evident that

$$p(x) = \frac{dF(x)}{dx}$$

That is, the probability density is the derivative of the CDF.

Further Definitions Associated with Distributions/Random Variables

You will have noticed some similarities and differences between discrete and continuous random variables and their associated distributions. Expectation and CDF are common to both discrete and continuous cases, while calculating expectations and probabilities involves sums and integrals, respectively (these are the same operations in a more abstract sense).

The *mean* of a random variable X is the real number defined by

$$\mu = E[X]$$

This, and many other concepts, applies to both discrete and continuous cases.

The *variance* of a random variable is defined by

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = E[X^2] - \mu^2$$

If X and Y are two random variables, the *joint distribution function* is defined by

$$F(x,y) = P(X \le x, Y \le y)$$

This is a function of two variables, and the definition generalizes to n random variables. In the continuous case, the joint density $p(x,y)$ is defined by

$$F(x,y) = \iint_{s \le x, t \le y} p(s,t)\,ds\,dt$$

while in the discrete case

$$F(x,y) = \sum_{s \le x} \sum_{t \le y} p(s,t)$$

It is important to know that X and Y are independent if and only if

$$p(x,y) = p_X(x) p_Y(y)$$

Examples of Discrete Distributions

There are a number of common discrete distributions that are used in statistics and quality control. Numerical values for distributions can be found in statistical tables and in software packages such as *Maple*.

Uniform Distribution

The uniform distribution is defined for any finite sample space S so that each outcome has equal probability (see above).

Bernoulli Distribution

The Bernoulli distribution is defined for experiments that have two possible outcomes, “success” (1), or “failure” (0). It is defined by

$$p(0) = P(X = 0) = 1 - p, \quad p(1) = P(X = 1) = p$$

where p is the probability of success.

Binomial Distribution

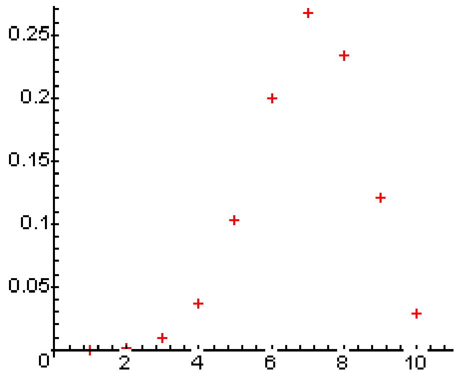
Suppose now that we repeat the Bernoulli success-fail experiment n times, independently, with probability of each success p. Let X denote the number of successes. This is called a *binomial* random variable with parameters *(n,p)*. The associated distribution is given by

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

The mean and variance of the binomial *(n,p)* distribution are

$$\mu = np, \quad \sigma^2 = np(1-p)$$

Binomial (10,0.7)



Example: Suppose we toss a fair coin four times, each toss independent of the other tosses. What is the probability that two heads and two tails are obtained? To solve this, let X equal the number of heads (successes), so that X is binomial with parameters *(n=4, p=1/2)*. Then

$$P(X = 2) = \binom{4}{2}(1/2)^2(1-1/2)^2 = 3/8 = 0.375$$

stats[statevalf,pf,binomiald[4,0.5]](2);

.3750

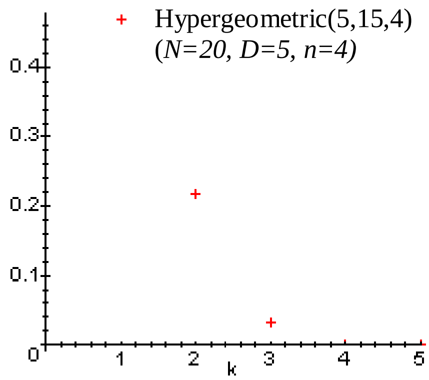
Geometric Distribution

Consider again the success-failure experiments, with probability of success p. Let X be the number of trials required until the first success. Then X takes values $1,2,\ldots,\infty$, and has distribution

$$p(n) = P(X = n) = p(1-p)^{n-1}, \qquad n = 1,2,\ldots$$

Hypergeometric Distribution

Sometimes sampling from a population is done without replacement, and the outcomes can be categorized as success or failure. Suppose we have a population of size N which contains D "successess" and N-D "failures". Select a sample of size n. Then the probability of selecting k "successes" is given by the *hypergeometric* distribution



$$p(k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$$

This is called a hypergeometric experiment and the associated random variable X is called a hypergeometric random variable. The mean and variance are given by

$$\mu = \frac{nD}{N}, \quad \sigma^2 = \frac{nD}{N}(1-\frac{D}{N})(\frac{N-n}{N-1})$$

Example: Suppose that a batch of 20 transistors contains 5 bad ones. What is the probability of a quality control inspector finding 3 bad transistors in a sample of 4 transistors? *Solution:* Let X be the number of "successes", i.e. the number of bad transistors. Now $N=20$, $D=5$, $n=4$ and $k=3$. Therefore $P(X=3)=0.031$.

```
stats[statevalf,pf, hypergeometric[5, 15, 4]](3);
```

.03095975232

Poisson Distribution

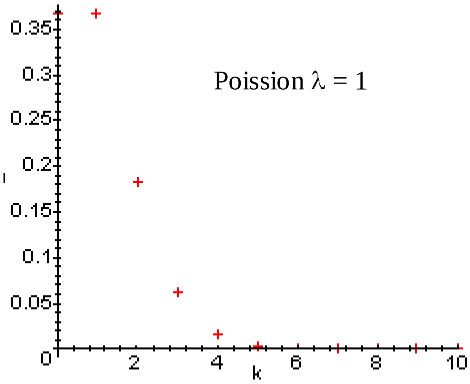
A *Poisson* random variable X records, say, the number of successes occurring during a given time interval. The Poisson distribution with parameter $\lambda>0$ is given by

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

The mean and variance are both equal to λ:

$$\mu = \lambda, \quad \sigma^2 = \lambda$$

Example: Suppose that the number of phone calls received per hour *X* is Poisson with parameter λ=1. Calculate the probability that there is at least one phone call per hour. *Solution:*



$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1} = 0.633$$

stats[statevalf,pf,poisson[1]](0);

.3678794412

The Poisson distribution is a good approximation to the binomial distribution when *n* is large and *p* is small:

$$P(X_{binomial} = i) \approx e^{-\lambda} \frac{\lambda^i}{i!} \qquad (n \text{ large}, p \text{ small})$$

Examples of Continuous Distributions

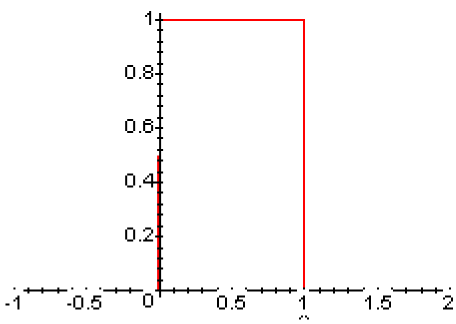
Here are some continuous distributions commonly used in statistics and quality control.

Uniform Distribution

The uniform distribution is defined for any finite interval *(a,b)* by the density function

$$p(x) = \frac{1}{b-a},\ x \in [a,b], \quad p(x) = 0,\ x \notin [a,b]$$

Uniform [0,1]



Exponential Distribution

The probability density of the exponential distribution is

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & if \quad x \geq 0 \\ 0 & if \quad x < 0 \end{cases}$$

The exponential distribution is often used in reliability analysis and queuing problems. The mean and variance are

$$\mu = \frac{1}{\lambda}, \quad \sigma^2 = \frac{1}{\lambda^2}$$

Gamma Distribution

Recall that the *gamma function* is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

The gamma function satisfies the recursion

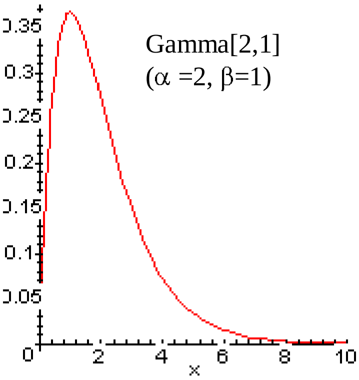
$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1), \quad \Gamma(1) = 1$$

and if n is an integer,

$$\Gamma(n) = (n-1)!$$

The *gamma distribution* is defined by the density function

$$p(x) = \begin{cases} \lambda e^{-\lambda x} \dfrac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & if \quad x \geq 0 \\ 0 & if \quad x < 0 \end{cases}$$



Gamma[2,1]
(α =2, β=1)

The mean and variance are given by

$$\mu = \frac{\alpha}{\lambda}, \quad \sigma^2 = \frac{\alpha}{\lambda^2}$$

The exponential distribution is a special case of the gamma distribution with α=1.

Normal or Gaussian Distribution

The *normal distribution* is one of the most widely used distributions in statistical quality control. It is also called the *Gaussian distribution.* The probability density of a normal random variable is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

Here, the parameters denote the mean and variance:

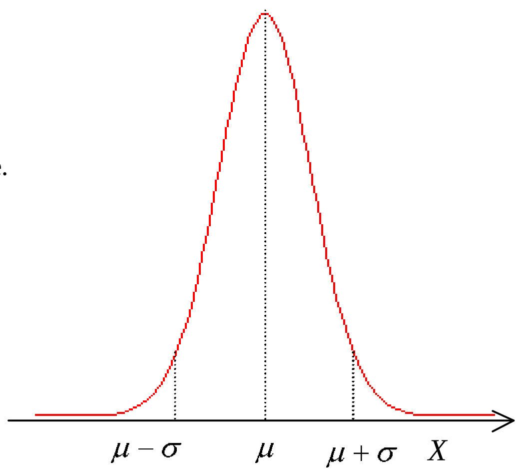
$$\mu = E[X], \quad \sigma^2 = E[X^2] - (E[X])^2$$

The shape of the normal distribution is a bell shaped curve.

The mean μ marks the center and peak of the curve, and the variance is the square *standard deviation σ.*

The distribution is often denoted *N(α,σ).*

Standard Normal *N(0,1)*



The proportion of the population which falls within the range μ+/-σ, i.e. within one standard deviation from the mean, is 68.26%. Also, 95.44% of the population lie within two standard deviations, μ+/-2σ, and 99.74% within three, μ+/-3σ. The standard deviation is a measure of the width of the curve; the smaller σ, the narrower the curve.

Calculation of probabilities using the normal distribution involves evaluation of areas under the normal density curve that cannot be done analytically. Many books have tables giving values for the CDF *F(x)* for a standardized normal distribution with mean 0 and variance 1; this is denoted *N(0,1).*The transformation

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

converts any given normal random variable *X* to a standard normal random variable *Z.*

Example: [Mitra] The length of a machined part is known to have a normal distribution with mean 100mm and standard deviation 2mm. What proportion of parts will have lengths above 103.3mm? *Solution:* Let *X* denote the length of the part. The standardized value of 103.3 corresponds to (103.3-100)/2=1.65. Then using a table for the standard normal distribution, or *Maple,*

$$P(X > 103.3) = P(Z > 1.65) = 1 - P(Z \le 1.65) = 1 - 0.9505 = 0.0495$$

stats[statevalf,cdf,normald](1.65);

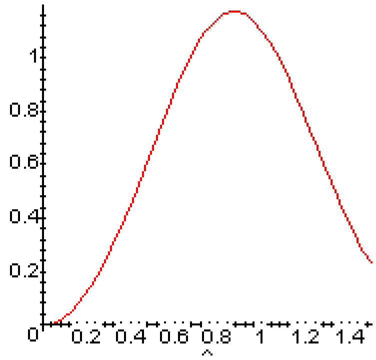
.9505285320

Weibull Distribution

The *Weibull distribution* is often used in reliability analyses to model the *time to failure* or *life length* of components. Such a continuous random variable T has a Weibull distribution with density

$$p(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} e^{-\left(\frac{t}{\alpha}\right)^{\beta}}, t > 0, \quad p(t) = 0, t \le 0$$

Weibull[3,1]
(α=1, β=3)



The parameters are α>0 and β>0, and when β=1, this reduces to the exponential distribution.

The mean and variance are

$$\mu = \alpha\Gamma(1+\frac{1}{\beta}), \quad \sigma^2 = \alpha^2\left\{\Gamma(1+\frac{2}{\beta}) - \left[\Gamma(1+\frac{1}{\beta})\right]^2\right\}$$

Example: [Mitra] The time to failure T of a cathode ray tube is modeled by a Wiebull distribution with parameters α=200 hours, β=1/3. What is the mean time to failure and what is the probability of the tube operating successfully for at least 800 hours? *Solution:*

$$\mu = 200 \cdot \Gamma(1+\frac{1}{1/3}) = 200 \cdot 6 = 1200\, hours$$

$$P(T > 800) = 1 - P(T \le 800) = 1 - 0.7955 = 0.2045$$

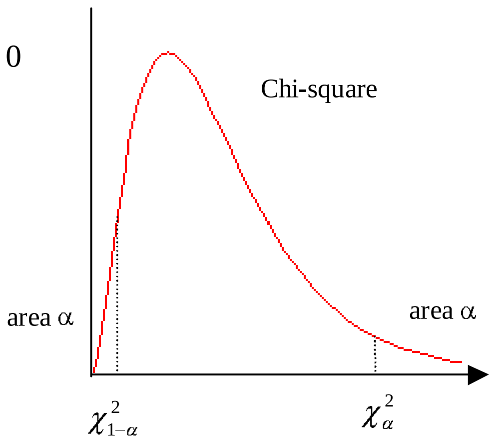
stats[statevalf,cdf,weibull[1/3,200]](800);

.7955437069

Chi-Square Distribution

The *Chi-square distribution* with ν *degrees of freedom* is an important special case of the gamma distribution (α=ν/2, λ=0.5). The density function is

$$p(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \; x > 0, \quad p(x) = 0, \; x \le 0$$



The points on the x-axis of the graph indicate where the area to the left and right, respectively, equal α.

The mean and variance are

$$\mu = \nu, \quad \sigma^2 = 2\nu$$

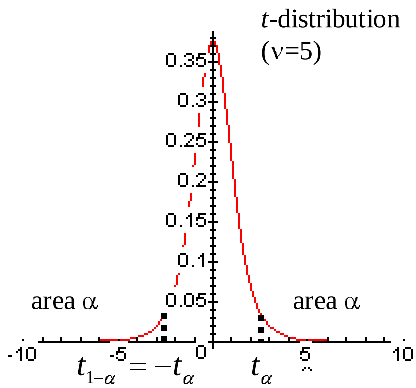
t-Distribution

The *student t-distribution* with ν *degrees of freedom* is the distribution of the random variable

$$T = \frac{Z}{\sqrt{V/\nu}}$$

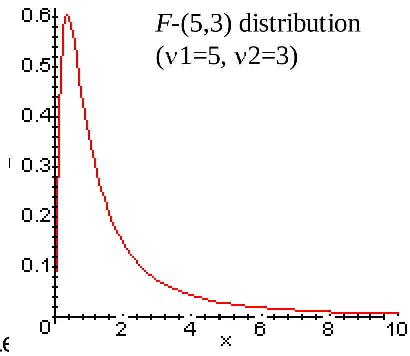
where *Z* is standard normal *N(0,1)* and *V* is independent of *Z* with a Chi-square distribution with ν degrees of freedom. It has density

$$p(t) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)\sqrt{\pi\nu}} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$



F-Distribution

The *F*-distribution is the distribution of the random variable



$$F = \frac{U / \nu_1}{V / \nu_2}$$

where U and V are independent chi-square random variables with $\nu 1$, $\nu 2$ degrees of freedom, respectively.

Law of Large Numbers and Central Limit Theorem

Two of the most important results in probability theory concern the nature of a large number of independent random variables. The central limit theorem explains the remarkable fact that many natural populations have distributions that are bell-shaped.

The *law of large numbers* states that if

$$X_1, X_2, \ldots$$

is a sequence of independent random variables with a common distribution, then

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to E[X_1]$$

as n tends to infinity. That is, *the average of a sequence of* *<u>independent indentically distributed (IID)</u>* *random variables converges to the common mean of that distribution.*

Example: Let E be an event and $P(E)$ the probability of occurrence. Let

$$X_i = \begin{cases} 1 & \text{if} \quad E \text{ occurs on ith trial} \\ 0 & \quad\quad \text{otherwise} \end{cases}$$

Then since $E[Xi]=P(E)$, the law of large numbers says

$$\frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\text{number of times } E \text{ occurs in } n \text{ trials}}{n} \to E[X_1] = P(E)$$

which means that the limiting proportion of time that event E occurs is just $P(E)$.

The *central limit theorem* states that under the above assumptions the distribution of

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal distribution *N(0,1)* as *n* tends to infinity. This gives a useful approximation that is used widely.

Example: Let *X* be binomial with parameters *(n,p)*. Then for large *n* the distribution of

$$\frac{X - E[X]}{\sqrt{Var(X)}} = \frac{X - np}{\sqrt{np(1-p)}}$$

is well approximated by the standard normal distribution *N(0,1)*.

Chapter 9 - Statistics

Statistics deals with the collection, classification, analysis, and making of inferences from data. *Descriptive statistics* is concerned with the description or summary of data, say in the form of tables and graphs, or in terms of numerical quantities derived from the data. *Inferential statistics* deals with methods by which one makes inferences or generalizations or conclusions about a population based on data from a sample.

Descriptive Statistics

Sampling

Statistical experiments result in a sample of data that is often recorded as numerical values. *Sampling* is the process of conducting an experiment that draws a sample from a population (recall that a population consists of the totality of all observations relevant to a study). Since it is not usually practical to sample a whole population, the sample should be chosen in a random fashion (i.e. independently and at random) and should adequately represent the population; this is needed if inferences drawn from the sample are to be accurate. Each observation in a population is considered to be the value of a random variable *X* with some (known or unknown) distribution. For instance, it may be reasonable to suppose that a certain population is normal, in which case we refer to it is a "normal population". Other populations may be described as a "binomial population", etc. An issue of major interest for statisticians is in arriving at conclusions concerning unknown population parameters (such as mean and variance). Knowledge of such parameters is then useful in quality control for evaluating product and process quality.

A *random sample* of size *n* from a population with density *p(x)* is a sequence of independent random variables each with density *p(x)*

$$X_1, X_2, \ldots, X_n$$