PROBABILITY, STATISTICS, AND RANDOM PROCESSES EE 351K

INDEX

PROBABILITY

Mathematically, the probability of an outcome is equal to the number of possible positive outcomes divided by the total possible outcomes (*size* of the sample space).

$$P(\text{positive outcome}) = \frac{\text{number of possible positive outcomes}}{\text{total number of possible outcomes}}$$

For example, if there are 5 balls in a box and 3 are green, the probability of choosing a green ball is

$$P(\text{choosing green ball}) = \frac{3}{5}$$

SINGLE COIN TOSS

There is an equal probability that the outcome will be heads or tails.

$$P(H) = P(T) = \frac{1}{2}$$

DOUBLE COIN TOSS

There are four possible outcomes with equal probability.

$$P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$$

The probability of getting at least one heads would be the sum of the outcomes providing that result:

$$P(HH) + P(HT) + P(TH) = \frac{3}{4}$$

Alternatively, the probability of getting at least one heads could be thought of as one minus the probability of not getting at least one heads:

$$1 - P(TT) = \frac{3}{4}$$

MULTIPLE COIN TOSS

There are 2^n possible outcomes to a multiple coin toss (when considering order of the results).

The probability of getting at least one heads:

$$P(\text{getting at least one heads}) = \left(2^n - 1\right)\frac{1}{2^n}$$

$$P(\text{getting at least one heads}) = 1 - \frac{1}{2^n}$$

The probability of getting exactly 2 heads:

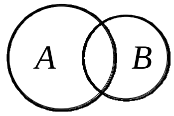
$$P\begin{pmatrix}\text{getting exactly} \\ \text{two heads}\end{pmatrix} = \frac{\overbrace{n}^{\substack{\text{number of possible} \\ \text{positions of first heads}}} \overbrace{(n-1)}^{\substack{\text{number of possible} \\ \text{positions of second heads}}}}{\underbrace{2}_{\substack{\text{order of occurance} \\ \text{is not a factor}}}} \cdot \underbrace{\frac{1}{2^n}}_{\substack{\text{inverse} \\ \text{outcome}}}$$

The probability of getting exactly 3 heads:

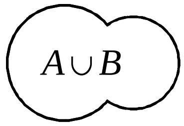
$$P\begin{pmatrix}\text{getting exactly} \\ \text{three heads}\end{pmatrix} = \frac{\overbrace{n}^{\substack{\text{number of possible} \\ \text{positions of first heads}}} \overbrace{(n-1)}^{\substack{\text{number of possible} \\ \text{positions of second heads}}} \overbrace{(n-2)}^{\substack{\text{number of possible} \\ \text{positions of third heads}}}}{\underbrace{2 \cdot 3}_{\substack{\text{number of ways to} \\ \text{order 3 items}}}} \cdot \underbrace{\frac{1}{2^n}}_{\substack{\text{inverse} \\ \text{outcome}}}$$

SET PROPERTIES

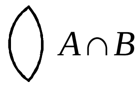
Given two sets A and B



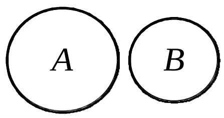
The **union** of two sets refers to that which is in set A **or** set B. In terms of area, it is the sum of the areas minus the common area.



The **intersection** of two sets refers to that which is in set A **and** set B. In terms of area, it is the common area found in both A and B.



Two sets with no common elements are called **disjoint**. More than two such sets are called **mutually exclusive**.



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

$$P(A) + P(\overline{A}) = 1$$

$\overline{A}$ means everything that is not in A

INCLUSION-EXCLUSION PRINCIPLE

Expanding on a theorem presented in the previous box, the probability that at least one event contained within a group of (possibly) intersecting sets of events occurs is the sum of the probabilities of each event minus the sum of the probabilities of all 2-way intersections, plus the sum of the probabilities of all 3-way intersections, minus the sum of all 4-way intersections, etc.

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{1 \le i < j \le n} P(E_i \cap E_j)$$

$$+ \sum_{1 \le i < j < k \le n} P(E_i \cap E_j \cap E_k) - \cdots$$

E = an event, in this case

W SAMPLE SPACE

The set of all possible outcomes. (Means the same as *probability space*, I think.) For example, if a coin is tossed twice, the sample space is

$$\Omega = \{(\mathrm{H}, \mathrm{H}), (\mathrm{H}, \mathrm{T}), (\mathrm{T}, \mathrm{H}), (\mathrm{T}, \mathrm{T})\}$$

Note that the sample space is not a number; it is a collection or set of results. This is frequently a source of confusion; the *size* of the sample space is a number, but the sample space itself is not a number. See Probability, p2.

W INFINITE SAMPLE SPACE

If a coin is tossed until it turns up heads, then the sample space for possible outcomes is

$$\Omega = \{1, 2, 3, \ldots\}$$

PERMUTATION

A permutation is a mapping of a finite set onto itself. For example if we have the set A = {a,b,c}, there are 3! possible permutations. One of them is

$$\begin{pmatrix} a & b & c \\ c & b & a \end{pmatrix}$$

In other words, there are n! ways to arrange n objects. However, in the case of a **cyclic permutation**, there are (n-1)! ways to arrange n objects. An example of a *cyclic permutation* would be the seating of n people at a round table.

ORDERING/COMBINATIONS 1

n different objects can be ordered in n! ways or *permutations*. What about the case when some objects are identical? For example, consider the letters in the name

KONSTANTOPOULOS

There are 15 letters in the word. This includes 4 Os, 2 Ns, 2 Ts, and 2 Ss. How many 15-letter combinations can be formed with these letters given that some letters are identical?

$$\text{number of words} = \frac{15!}{4!2!2!2!}$$

ORDERING/COMBINATIONS 2

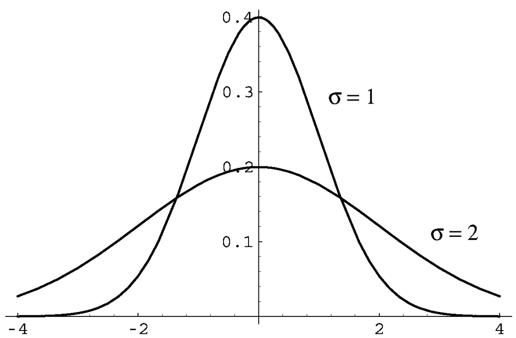
In how many ways can n identical objects be arranged in m containers?

$$\binom{n+m-1}{n-1} = \frac{(n+m-1)!}{(n-1)! \times (m-n)!}$$

A problem that appeared in the textbook was, how may ways can 6 identical letters be put in 3 mail boxes? The answer is 28.

NORMAL RANDOM VARIABLE

A normal random variable has a Gaussian density function centered at the expectation μ. The figure below shows plots of the density functions of two normal random variables, centered at the common expectation of 0. The plot having the sharper peak is for the special case of a standard normal random variable determined by and expectation of 0 and deviation of 1. A normal random variable does not necessarily have an expectation of zero.



Plot of the density functions for normal random variables with expectation μ = 0

Z STANDARD NORMAL RANDOM VARIABLE p.213

A standard normal random variable has the parameters expectation $\mu = 0$ and deviation $\sigma = 1$ (see Normal Density Function p.8).

A normal (i.e. Gaussian density) random variable with parameters μ and σ can be written in terms of the standard normal random variable:

$$\boxed{X = \sigma Z + \mu}$$

The process of changing a normal random variable to a standard normal random variable is called *standardization*. If X has a normal distribution with parameters μ and σ, then the standardized version of X is

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

INDEPENDENT EVENTS p.139

Two events A and B are called independent if the outcome of one does not affect the outcome of the other. Mathematically, (for a particular probability assignment/distribution) two events are independent if the probability of both events occurring is equal to the product of their probabilities.

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

For example, the outcome of the first roll of a die does not affect the second roll. The independence of two events can be lost if the probabilities are not even, e.g. an unfair coin or die.

Independence can also be expressed in terms of a conditional probability. The probability of A given B is still A.

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

And if this is true, then it is also true that

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

UNEVEN PROBABILITY

If we assign the probability n to the outcome of heads of an unfair coin, then

$$P(\text{H}) = n$$

$$P(\text{T}) = 1 - n$$

The probabilities of a double coin toss will be

TT: $(1-n)^2$

TH: $n(1-n)$

HT: $n(1-n)$

HH: n^2

BINOMIAL COEFFICIENT p.95

On my TI-86 calculator, the command to do this is ***a* nCr *b***. The **nCr** function is found under **MATH/PROB**.

$$\binom{a}{b} = \frac{a!}{(a-b)!\,b!}$$

$\binom{52}{10}$ is read "52 choose 10" and stands for the number of combinations of 10 there are in a pool of 52 units.

Example Problem: what are the chances that there will be 4 aces among 10 cards picked from a deck of 52?

$$\frac{\binom{48}{6}\times\binom{4}{4}}{\binom{52}{10}} = \frac{\frac{48!}{(48-6)!\,6!}\times\frac{4!}{(4-4)!\,4!}}{\frac{52!}{(52-10)!\,10!}} = \frac{\frac{48!}{42!\,6!}\times\frac{4!}{0!\,4!}}{\frac{52!}{42!\,10!}}$$

What this is saying is, "From 48 non-aces choose 6, then from 4 aces choose 4". The product is the number of possible combinations of 10 that can contain 4 aces. "Now divide this amount by all of the possible combinations of 10 cards out of 52." Note that **0!** = **1**, so we have

$$= \frac{\frac{48!}{42!\,6!}}{\frac{52!}{42!\,10!}} = \frac{48!\,42!\,10!}{42!\,6!\,52!} = 0.000776$$

MULTINOMIAL COEFFICIENT

A binomial coefficient becomes multinomial when there is more than one type of object to choose or there is more than one location to choose objects from.

Multiple object types:

$$\binom{n}{n_1\ n_2 \cdots n_k} = \frac{n!}{n_1!\,n_2!\cdots n_k!}$$

$\binom{n}{n_1\ n_2 \cdots n_k}$ is read "n choose n_1 of type 1, n_2 of type 2 … and n_k of type k".

Multiple object locations:

For example, the number of ways we can arrange n objects in k boxes

$$\binom{n+k-1}{n} = \frac{(n+k-1)!}{n!\,(k-1)!}$$

DISTRIBUTION FUNCTIONS

DISTRIBUTION FUNCTION p.19

A distribution function is a function that describes probabilities as a function of outcomes. That is, for every possible outcome, the distribution function gives the probability. There are many different types of distributions functions; determining which one applies to a particular situation is often difficult so that in teaching this subject the topic may be avoided entirely with the advice offered that "you need to work a lot of problems in order to develop a sense of which distribution function to use."

When we take the derivative of a distribution function, the result is the **density function (p8)**.

m(w) DISCRETE UNIFORM DISTRIBUTION FUNCTION p.19,367

The function assigning probabilities to a finite set of equally likely outcomes. For example, the distribution function of a fair double coin toss is

$$m(\mathrm{H,\,H}) = m(\mathrm{H,\,T}) = m(\mathrm{T,\,H}) = m(\mathrm{T,\,T}) = \frac{1}{4}$$

The distribution function for the roll of a die is

$$m_i = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

In general, the discrete uniform distribution function is

$$P(X=x) = \begin{cases} 1/(l-k+1), & x = k, k+1, k+2, \ldots, l \\ 0, & \text{otherwise} \end{cases}$$

$$\mu = \frac{k+l}{2} \qquad \sigma^2 = \frac{(l-k)(l-k+2)}{12}$$

Generating Function: $g(t) = \dfrac{e^{tk} - e^{t(l+1)}}{(l-k+1)(1-e^t)}$

The distribution function may be an infinite series. For example, if a coin is tossed until the first time heads turns up then the distribution function would look like:

$$\sum_{\omega=0}^{\infty} m(\omega) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$$

The sum of a distribution function is the sum of the probabilities and is equal to one. See also Specific Generating Functions p19.

μ = center of the density, average value, expected value
σ^2 = variance
k = the lowest value in the sample space
l = the highest value in the sample space
$g(t)$ = generating function

CUMULATIVE DISTRIBUTION FUNCTION

p.61

When X is a continuous real-valued random variable, the cumulative distribution function of X is

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

In other words, it is the probability of a positive result over a range of initial occurrences. The cumulative distribution function is useful for finding the *density function* when a random variable that is a function of another random variable is involved. The density function is the differential of the cumulative distribution function.

X = random variable: the observation of an experimental outcome
x = a variable representing a particular outcome
$f(x)$ = density function
t = a dummy variable of integration

$F_X(x)$ CUMULATIVE NORMAL DISTRIBUTION FUNCTION

The derivative of the normal density function (p.8). The function has parameters μ (expected value) and σ (standard deviation). F_X must be computed using numerical integration; there are tables of values for this function in Appendix A of the textbook.

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\,du$$

X = random variable for the number of occurrences in a given interval in time, area, length, etc.
x = a particular outcome.
μ = center of the density, average value, expected value
σ = a positive value measuring the *spread* of the density, standard deviation

$F(x,y)$ JOINT CUMULATIVE DISTRIBUTION FUNCTION p.165

The example below is for two random variables and may be extended for additional variables.

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

BERNOULLI TRIALS PROCESS

p.96,233,261

A Bernoulli trials process is a sequence of n chance experiments such that 1) each experiment has two possible outcomes and 2) the probability p of success of each experiment is the same and is unaffected by the knowledge of previous experiments. Examples of Bernoulli trials are flipping coins, opinion poles, and win/lose betting. See also Negative Binomial Distribution p7.

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

p = probability of a successful outcome

$b(n,p,k)$ BINOMIAL DISTRIBUTION FUNCTION p.184

A function assigning probabilities to a finite set of trials where there are two possible results per trial, not necessarily of equal probability. The binomial distribution produces a bell-shaped curve. When the parameter n is large and the parameter p is small, the *Poisson Distribution* is a useful approximation may be used instead. The expectation is $E(X) = np$.

$$b(n, p, k) = \binom{n}{k} p^k q^{n-k}$$ where

$\binom{n}{k}$ accounts for the ways the result can be ordered

$$\mu = np \qquad \sigma^2 = npq \qquad \phi(x) = \left(q + pe^x\right)^n$$

In the case of **equal probabilities (*p* = 0.5)**, the function reduces to

$$b(n, 0.5, k) = \binom{n}{k} 0.5^n$$

In the case of **no successful outcomes (*k* = 0)**, the function reduces to

$$b(n, p, 0) = q^n$$

n = number of trials or selections
p = probability of success
q = probability of failure ($1-p$)
k = number of successful outcomes
μ = center of the density, average value, expected value
σ^2 = variance
$\phi(x)$ = density function

For example, if I can guess a person's age with a 70% success rate, what is the probability that out of ten people, I will guess the ages of exactly 8 people correctly?

$$b(10, .7, 8) = \binom{10}{8} .7^8 .3^{10-8} = 0.233$$

See also Specific Generating Functions p19.

MULTINOMIAL DISTRIBUTION

This problem involves more than one type of random variable. For example, a box contains M green balls and N red balls. If we choose k balls, what is the probability that m are green and n are red?

$$P\left(\text{of selecting } m,n \text{ balls}\right)=\frac{\binom{M}{m}\binom{N}{k-m}}{\binom{M+N}{k}}$$

This example was also used for *hypergeometric distribution*. See section 5.1.

JOINT DISTRIBUTION FUNCTION p.141

A joint distribution function describes the probabilities of outcomes involving multiple random variables. If the random variables are mutually independent then the joint distribution function is the product of the individual distribution functions of the random variables.

$$F_{X,Y}\left(x,y\right)=F_X\left(x\right)F_Y\left(y\right)$$

GEOMETRIC DISTRIBUTION p.184

The geometric distribution applies to a series of dual-outcome events (such as coin tosses) where p is the probability of success on any given event, q is the probability of failure, j is the event number, and T is a random variable that stands for the event that produces the first success. The geometric distribution function has the *memoryless property*, p8. See also Specific Generating Functions p19.

$$\boxed{P\left(T=j\right)=q^{j-1}p}$$

$$\mu=\frac{1}{p} \qquad \sigma^2=\frac{q}{p^2} \qquad g\left(t\right)=\frac{pe^t}{1-qe^t}$$

T = the first successful event in the series
j = the event number 1, 2, 3, etc.
p = the probability that any one event is successful
q = the probability that an event is not successful, $1-p$
μ = center of the density, average value, expected value
σ^2 = variance
$g(t)$ = generating function

NEGATIVE BINOMIAL DISTRIBUTION

p.186

Negative binomial distribution is a more general form of geometric distribution. A new variable k is introduced representing the number of successful outcomes in x attempts. For geometric distribution, $k = 1$.

$$u\left(x,k,p\right)=P\left(X=x\right)=\binom{x-1}{k-1}p^k q^{x-k}$$

This seems to describe the **Bernoulli Trials Process** which is a sequence of x chance experiments such that 1) each experiment has 2 possible outcomes and 2) the probability p of success is the same for each experiment and is not affected by knowledge of previous outcomes.

X = random variable: the observation of an experimental outcome
x = the number of attempts
k = the number of successful outcomes
p = the probability that any one event is successful
q = the probability that an event is not successful, $1 - p$

POISSON DISTRIBUTION p.187

An approximation of a discrete probability distribution. The Poisson distribution is used as an approximation to the binomial distribution when the parameters n and p are large and small, respectively. It is also used in situations where it may not be easy to interpret or measure the parameters n and p. See also Specific Generating Functions p19.

$$P\left(X=k\right)\approx\frac{\lambda^k}{k!}e^{-\lambda}$$

$$\mu=\lambda \qquad \sigma^2=\lambda \qquad g\left(t\right)=e^{\lambda\left(e^t-1\right)}$$

X = random variable: the observation of an experimental outcome
λ = rate of occurrence, i.e. the number of positive outcomes expected over a given period. This might be the product of the probability and the number of trials.
k = number of positive outcomes
μ = center of the density, average value, expected value
σ^2 = variance
$g(t)$ = generating function

EXPONENTIAL DISTRIBUTION p.205

The exponential distribution function is the integral of the exponential density function. It represents the probability that an event will take place *before* time t. See Exponential Density Function on page 9.

$$F(t) = 1 - e^{-\lambda t}$$

Expectation: $\mu = \frac{1}{\lambda}$ **Variance:** $\sigma^2 = \frac{1}{\lambda^2}$

Deviation: $\sigma = \frac{1}{\lambda}$ **Generating Fnct.:** $g(t) = \frac{\lambda}{\lambda - t}$

λ = rate of occurrence, a parameter
t = time, units to be specified

DENSITY FUNCTIONS

$f(x)$ DENSITY FUNCTION p.59

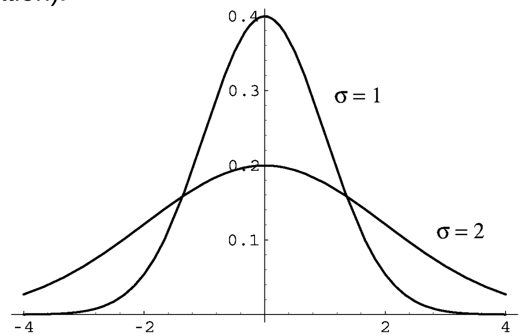
The density function is the derivative of the *distribution function* $F(x)$ (see p5). The integral of a density function over its entire interval is equal to one. So by integrating a density function over a particular interval we determine the probability of an outcome falling within that interval.

$$P(a \le X \le b) = \int_a^b f(x)\,dx, \quad f(x) = F'(x)$$

The density function has no negative values. A function that does not integrate to one may be *normalized* by dividing the function by its integral.

$f_X(x)$ NORMAL DENSITY FUNCTION p.212

If a large number of mutually independent random variables is considered, the normal density function is a close approximation. The normal density function has parameters μ (expected value) and σ (standard deviation).



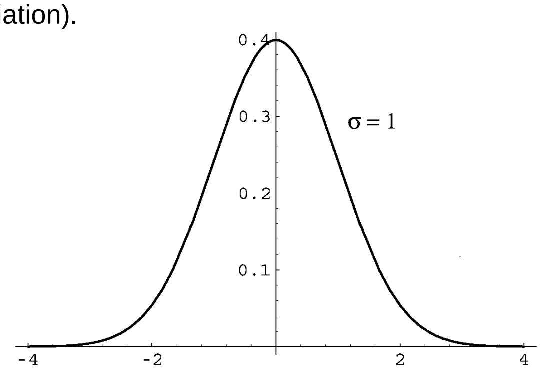
Plot of the normal density function for m = 0

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

X = random variable: the observation of an experimental outcome
x = a particular outcome.
μ = center of the density, average value, expected value
σ = the standard deviation, a positive value measuring the *spread* of the density

$\phi(x)$ STANDARD NORMAL DENSITY FUNCTION p.325

The case of the normal density function with parameters $\mu = 0$ (expected value) and $\sigma = 1$ (standard deviation).



Plot of the standard normal density function

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

$f(w)$ CONTINUOUS UNIFORM DENSITY

p.205

Uniform density means that the probability of an outcome is equally weighted over the interval of consideration. *Continuous* means that there are an infinite number of possible outcomes (as opposed to a *discrete* number). Consider random values on the interval $[a,b]$. The (uniform) density function is

$$f(\omega) = \frac{1}{b-a}$$

The mean value or expectation of an experiment having uniform density is then

$$\mu = \frac{b-a}{2}$$

$f(x,y)$ JOINT DENSITY FUNCTION p.165

Where X and Y are continuous random variables, the joint density function is

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x\,\partial y}, \quad \text{where} \quad f(x) = F'(x)$$

The joint density function satisfies the following equation:

$$F(x,y) = \int_{-\infty}^{t}\int_{-\infty}^{u} f(t,u)\,dt\,du$$

The joint density function can involve more than two variables and looks like

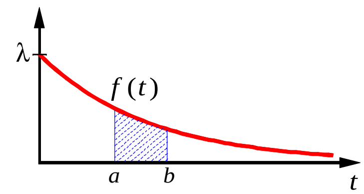
$$f(x_1, x_2, \cdots, x_n) = \frac{\partial^n F(x_1, x_2, \cdots, x_n)}{\partial x_1\,\partial x_2 \cdots \partial x_n}$$

EXPONENTIAL DENSITY FUNCTION p.205

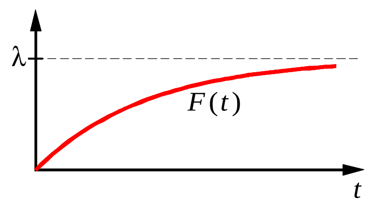
The exponential density has the parameter λ. The function is often used to describe an expected lifetime where the parameter λ is the **failure rate**. A higher value of λ means the failure is likely to be sooner. The exponential density function has the *memoryless property*, p8. The exponential density function will be on the exam.

Example: the probability that a light bulb burns out after t hours. The total area under the curve is one; the area from the interval $[a,b]$ is the probability of failure during that time.

Exponential density function: $f(t) = \lambda e^{-\lambda t}$



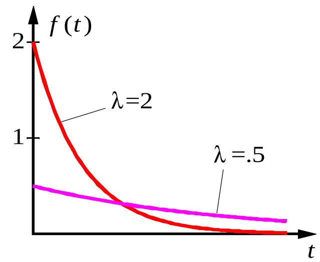
Probability distribution function: $F(t) = 1 - e^{-\lambda t}$



e.g. for a random variable T

$$P(T > x) = e^{-\lambda x} \qquad P(T \le x) = 1 - e^{-\lambda x}$$

Two exponential density functions are plotted at right with $\lambda = 2$ (high failure rate) and $\lambda = 0.5$ (low failure rate). Note that in each case at right the area under the curve is 1. Both curves extend to infinity.



Expectation: $\mu = \dfrac{1}{\lambda}$

Variance: $\sigma^2 = \dfrac{1}{\lambda^2}$ **Deviation:** $\sigma = \dfrac{1}{\lambda}$

Generating Function: $g(t) = \dfrac{\lambda}{\lambda - t}$

RELIABILITY p.205

The reliability is the probability that an event will take place after a given amount of time.

$$\boxed{\text{reliability} = \int_T^{\infty} f(t)\,dt}$$

For example, from the previous light bulb example, the probability that the bulb will last more than T hours is its reliability.

$$\text{reliability} = \int_T^{\infty} \lambda e^{-\lambda t}\,dt = e^{-\lambda T}$$

λ = failure rate
t = time [s]

DE MORGAN'S LAWS

Augustus De Morgan, British mathematician 1806-1871.

$$P\left(\overline{A \cup B}\right) = P\left(\overline{A} \cap \overline{B}\right)$$

$$P\left(\overline{A \cap B}\right) = P\left(\overline{A} \cup \overline{B}\right)$$

From this we can get

$$A - (B \cap C) = A \cap \left(\overline{B \cup C}\right) = A \cap \left(\overline{B} \cap \overline{C}\right)$$
$$= \left(A \cap \overline{B}\right) \cap \left(A \cap \overline{C}\right) = (A - B) \cap (A - C)$$

POKER PROBABILITIES

The probability of being dealt a certain poker hand can be described as the product of the probability of getting one specific set of cards satisfying the requirement times the number of possible sets that would satisfy the requirement.

For example, what is the probability of getting 3 of a kind?

The probability of getting one specific hand satisfying this requirement is

$$\frac{1}{\binom{52}{5}} = \frac{1}{2{,}598{,}960} = 384.77 \times 10^{-9}$$

How many ways can you have 3 of a kind? Consider that you have 3 2s and 2 non-2s that are not a pair. Given that there are 4 suits, there are 4 ways to have 3 2s. There are 48 ways to have the 4th card and 44 ways to have the remaining card. So the number of ways you can have 3 2s is 4×48×44=8448. Multiply that by the 13 different numerical values in a deck of cards to get the total number of 5-card hands that contain 3 of a kind (109,824). So the probability is getting 3 of a kind is

$$\frac{1}{\binom{52}{5}} \times 4 \times 48 \times 44 \times 13 = 0.04226$$

$P(A|B)$ CONDITIONAL PROBABILITY

$P(A|B)$ reads, "the probability that event A will occur given that event B has occurred." Since we know that B has occurred, the sample space now consists of only those outcomes in which B has occurred.

$$\boxed{P(A\,|\,B) = \frac{P(A \cap B)}{P(B)}}$$

For example, let X be the outcome of rolling a die once. Let A be the event $\{X = 6\}$ and B be the event $\{X > 4\}$.
$P(A) = 1/6$. But if the die has been rolled and we are told that B has occurred, then we can only have a 5 or a 6 so

$$P(A\,|\,B) = \frac{1/6}{1/3} = \frac{1}{2}$$

$f(x|E)$ CONTINUOUS CONDITIONAL DENSITY FUNCTION p.162

The formula for continuous conditional density is

$$f(x|E) = \begin{cases} f(x)/P(E), & \text{if } x \in E \\ 0, & \text{if } x \notin E \end{cases}$$

For example, if we know that a spinner has stopped with its pointer in the upper half of a circle, $0 \le x \le ½$, then the conditional density is

$$f(x|E) = \begin{cases} 1/(1/2), & \text{if } 0 \le x \le 1/2 \\ 0, & \text{if } 1/2 < x < 1 \end{cases} = \begin{cases} 2, & \text{if } 0 \le x \le 1/2 \\ 0, & \text{if } 1/2 < x < 1 \end{cases}$$

$f(x)$ = the density function for random variables X_i

E = an event with positive probability that gives some evidence about which hypotheses are correct

$P(E)$ = the probability of event E occurring

BAYES' THEOREM

Bayes' theorem is useful if we know $P(A|B)$ and want to find $P(B|A)$.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

This was not found in our textbook.

BAYES' FORMULA p.145

This is a famous formula but we will rarely use it. If the number of hypotheses is small, a simple tree measure calculation is done; if the number of hypotheses is large, then we use a computer.

$$P(H_i|E) = \frac{P(H_i)P(E|H_i)}{\sum_{k=1}^{m} P(H_k)P(E|H_k)}$$

Bayes probabilities are used for medical diagnosis. Given a set of test results and the probabilities for test outcomes, what are the probabilities that the patient has the disease?

H_i = a set of pairwise disjoint events called *hypotheses*

E = an event that gives some evidence about which hypotheses are correct

$P(H_i)$ = a set of probabilities called *prior probabilities*

$P(H_i|E)$ = conditional probabilities called *posterior probabilities*

BAYES' INVERSE PROBLEM

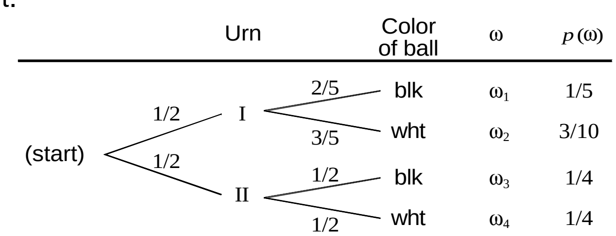
Bayes proposed to find the conditional probability that the unknown probability p lies between a and b, given m successes in n trials.

$$P(a \le p < b \mid m \text{ successes in } n \text{ trials}) = \frac{\int_a^b x^m (1-x)^{n-m}\,dx}{\int_0^1 x^m (1-x)^{n-m}\,dx}$$

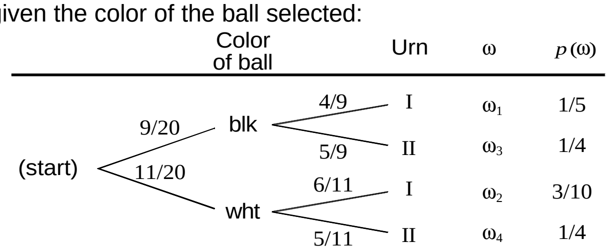
The computation of the integrals is too difficult for exact solution except for small values of j and n.

PROBABILITY TREE

Example: Urn I contains 2 black balls and three white balls, urn II contains 1 black ball and 1 white ball. The following tree shows the probabilities involved in selecting an urn and random and selecting a ball from it.



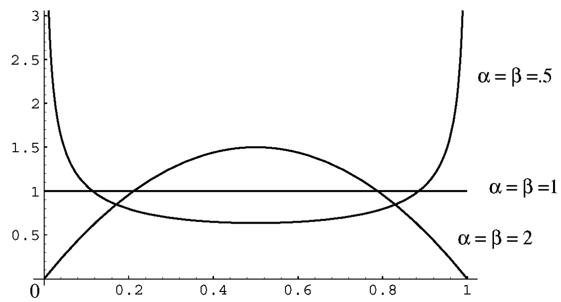
The **reverse tree** gives the probabilities of the urn chosen given the color of the ball selected:

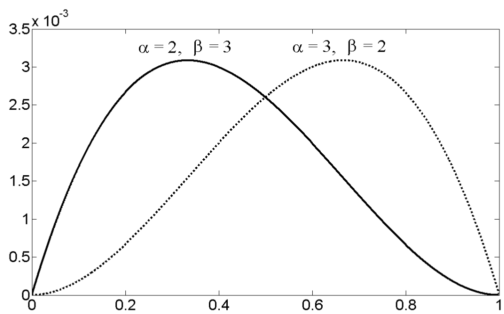


$B(\alpha,\beta,x)$ BETA DENSITY FUNCTION p.168

A density function having positive parameters α and β. When both parameters are equal to one, the beta density is the uniform density. When they are both greater than one, the function is bell-shaped; when they are both less than one, the function is U-shaped. A beta density function can be used to fit data that does not fit the Gaussian curve of a normal density function (p8).

Beta Density Functions





$$B(\alpha,\beta,x)=\begin{cases}\left[\dfrac{1}{B(\alpha,\beta)}\right]x^{\alpha-1}(1-x)^{\beta-1}, & \text{if } 0\le x\le 1\\ 0, & \text{otherwise}\end{cases}$$

Beta function: $$B(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

Given α and β, the probability of an event being successful is

$$P(\text{success})=\frac{\alpha}{\alpha+\beta}$$

If α and β are integers:

$$B(\alpha,\beta)=\frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$$

MEMORYLESS PROPERTY p.206

The memoryless property applies to the exponential density function and the geometric distribution function.

$$P\big(T>(r+s)\,|\,(T>r)\big)=P(T>s)$$

EXPECTATION

$E(X)$, μ EXPECTED VALUE OF DISCRETE RANDOM VARIABLES p.225

The expected value, also known as the *mean* and sometimes identified as μ, is the sum of the product of each possible outcome and its probability. It is the *center of mass* in a distribution function. If a large number of experiments are undertaken and the results are averaged, the value obtained should be close to the *expected value*. The formula for the expected value is

$$\mu=E(X)=\sum_{x\in\Omega}xm(x)$$

provided the sum converges. For example, the expected value for the roll of a die is 1(1/6)+2(1/6)+3(1/6)+4(1/6)+5(1/6)+6(1/6)=3.5.

If the probability p is the same for each of n possible outcomes, then the expected value is just

$$\mu=E(X)=np$$

For a *uniform density* problem (p.9), the expected value is

$$\mu=E(X)=\frac{b-a}{2}$$

X = numerically-valued discrete random variable: the observation of an experimental outcome
$m(x)$ = discrete distribution function
Ω = the sample space

$E(X)$, μ EXPECTED VALUE OF CONTINUOUS RANDOM VARIABLES

p.268

The expected value, also known as the *mean* and sometimes identified as μ, is the *center of mass* in a distribution function. If a large number of experiments are undertaken and the results are averaged, the value obtained should be close to the *expected value*. The formula for the expected value is

$$\mu = E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

provided that $\int_{-\infty}^{+\infty} |x|\, f(x)\, dx$ is finite.

Otherwise the expected value does not exist. Note that the limits of integration may be reduced provided they include the sample space.

For an *exponential density* $f(t) = \lambda e^{-\lambda t}$ (p.9), the expected value is

$$\mu = \frac{1}{\lambda}$$

X = random variable: the observation of an experimental outcome
$f(x)$ = the density function for random variable X.

$E(\phi(X))$ EXPECTATION OF A FUNCTION

p.229

If X and Y are two random variables and Y can be written as a function of X, then the expected value of Y can be computed using the distribution of X.

$$E(\phi(X)) = \sum_{x \in \Omega} \phi(x)\, m(x)$$

again, with the provision that the sum converges.

X = numerically-valued discrete random variable with sample space Ω
$\phi(X)$ = a real-valued function of the random variable X with domain Ω
Ω = the sample space

PROPERTIES OF EXPECTATION p.268, 394

If X is a real valued random variable with $E(X) = \mu$, then

$$E(X^2) = V(X) + \mu^2 \qquad E(X^n) = \int_{-\infty}^{+\infty} x^n f_X(x)\, dx$$

If X and Y are two random variables with finite expected values, then

$$E(X+Y) = E(X) + E(Y)$$

If X is a random variable and c is a constant

$$E(cX) = cE(X)$$

If X and Y are *independent*

$$E(X \cdot Y) = E(X)E(Y)$$

$f_X(x)$ = density function for the random variable X

MARKOV INEQUALITY

The probability that an outcome will be greater than or equal to some constant k is less than or equal to the expected value divided by that constant.

$$P(X > k) \le \frac{E(X)}{k}$$

For example, if the expected height of a person is 5.5 feet, then the Markov inequality states that the probability that a person is more than 11 feet tall is no more than ½. This example demonstrates the *looseness* of the Markov inequality. A more meaningful inequality is the *Chebyshev inequality*, which is a special case of Markov's inequality (p. 16).

VARIANCE

$V(X)$, σ^2 VARIANCE OF DISCRETE RANDOM VARIABLES p.257

Variance is a measure of the deviation of an outcome from the expected value. The variance is found by taking the difference between the expected value and each possible outcome, squaring that difference, multiplying that square by the probability of the outcome, and then summing these for each possible outcome. The *expected value* is more useful as a prediction when the outcome is not likely to deviate too much from the expected value.

$$\sigma^2 = V(X) = E\left((X-\mu)^2\right) = \sum_x (x-\mu)^2 m(x)$$

For discrete random variables, the variance can be found by a couple of methods:

Method 1: $\sum_x (x-\mu)^2 m(x)$ Find the expected value μ. Subtract μ from each possible outcome. Square each of these results. Multiply each result by its probability and then sum all of these.

For example, the variance of the roll of a die is $[(1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2](1/6)=35/12$.

Method 2: $E(X^2)-\mu^2$ Multiply the probability of each outcome by the square of the outcome. Sum the results to get $E(X^2)$. Find the expected value μ. Subtract the square of μ from $E(X^2)$.

For example, for the roll of a die, $E(X^2)=1\left(\frac{1}{6}\right)+4\left(\frac{1}{6}\right)+9\left(\frac{1}{6}\right)+16\left(\frac{1}{6}\right)+25\left(\frac{1}{6}\right)+36\left(\frac{1}{6}\right)=\frac{91}{6}$, then $E(X^2)-\mu^2=\frac{91}{6}-\left(\frac{7}{2}\right)^2=\frac{35}{12}$.

The variance of a Bernoulli Trials process is npq.

X = numerically-valued discrete random variable
μ = the expected value of X, $E(X)$
$m(x)$ = discrete distribution function

$V(X)$, σ^2 VARIANCE OF CONTINUOUS RANDOM VARIABLES p.271

Variance is a measure of the deviation of an outcome from the expected value. The *expected value* is more useful as a prediction when the outcome is not likely to deviate too much from the expected value.

$$\sigma^2 = V(X) = E\left((X-\mu)^2\right) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx$$

Note that the limits of integration may be adjusted so long as they continue to include the sample space. If the integral fails to converge, the variance does not exist.

The variance of a uniform distribution on [0,1] is 1/12.

The variance of an exponential distribution is $1/\lambda^2$.

X = random variable: the observation of an experimental outcome
μ = the expected value of X, $E(X)$

$V(X)$, σ^2 PROPERTIES OF VARIANCE p.259

$$V(X+Y) = V(X) + V(Y)$$

$$V(cX) = c^2 V(X) \qquad V(X+c) = V(X)$$

$$V(X) = E(X^2) - \mu^2$$

$D(X)$, σ STANDARD DEVIATION p.257

The *standard deviation* of X is the square root of the *variance* and is sometimes written σ.

$$D(X) = \sqrt{V(X)} = \sqrt{E\left((X-\mu)^2\right)}$$

The standard deviation of a Bernoulli Trials process is

$$\sigma = \sqrt{npq}$$

X = random variable: the observation of an experimental outcome
$V(X)$ = the variance of X
μ = the expected value of X, $E(X)$

$\mathbf{cov}(X,Y)$ COVARIANCE p.280

The book doesn't go into detail about this. Covariance applies to both discrete and continuous random variables.

$$\operatorname{cov}(X,Y) = E\left[\left(X - \mu(X)\right)\left(Y - \mu(Y)\right)\right]$$
$$= E(XY) - \mu(X)\mu(Y)$$

Property of covariance:

$$V(X+Y) = V(X) + V(Y) + 2\operatorname{cov}(X,Y)$$

X = random variable: the observation of an experimental outcome

$V(X)$ = the variance of X

μ = the expected value of X, $E(X)$

$\mathbf{r}(X,Y)$ CORRELATION p.281

The book doesn't go into detail about this either. Correlation applies to continuous random variables. Another text calls this the correlation coefficient and has a separate function for discrete random variables which it calls correlation.

$$\rho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{V(X)V(Y)}}$$

X = random variable: the observation of an experimental outcome

$V(X)$ = the variance of X

μ = the expected value of X, $E(X)$

CONVOLUTION

SUM OF RANDOM VARIABLES p.285, 291

Discrete: Given $Z = X + Y$, where X and Y are independent discrete random variables with distribution functions $m_1(x)$ and $m_2(y)$, we can find the distribution function $m_3(z)$ of Z using convolution.

$$m_3 = m_1 * m_2$$
$$m_3(z) = \sum_k m_1(k) \cdot m_2(z-k)$$

Continuous: Given $Z = X + Y$, where X and Y are independent continuous random variables with density functions $f(x)$ and $g(y)$, we can find the density function $h(z)$ of Z using convolution. Note that we are talking density functions here where it was distribution functions where discrete random variables were concerned. Also note that the limits of integration may be adjusted for density functions that do not extend to infinity.

$$f(x) * g(y) = h(z) = \int_{-\infty}^{+\infty} f(z-y)\, g(y)\, dy$$
$$= \int_{-\infty}^{+\infty} g(z-x)\, f(x)\, dx$$

For more about the sum of random variables, see Properties of Generating Functions p20.

k = represents all of the integers for which the probabilities $m_1(k)$ and $m_2(z\text{-}k)$ exist. (In cases where the probability doesn't exist, the probability is zero.)

CONVOLUTION EXAMPLE

Suppose the distribution functions $m_1(x)$ and $m_2(y)$ for the discrete random variables X and Y are

$$m_1(x) = m_2(x) = \begin{pmatrix} 0 & 1 & 2 \\ 1/8 & 3/8 & 1/2 \end{pmatrix}$$

Given $Z = X + Y$, we can find the distribution function $m_3(z)$ of Z using convolution.

$$m_3(z) = m_1(x) * m_2(y) = \sum_k m_1(k) \cdot m_2(z-k)$$

For each possible value of k we have

$$P(z=0) = \frac{1}{8}\cdot\frac{1}{8} = \frac{1}{64}, \quad k = 0$$

$$P(z=1) = \frac{1}{8}\cdot\frac{3}{8} + \frac{3}{8}\cdot\frac{1}{8} = \frac{3}{32}, \quad k = 0, 1$$

$$P(z=2) = \frac{1}{8}\cdot\frac{1}{2} + \frac{3}{8}\cdot\frac{3}{8} + \frac{1}{2}\cdot\frac{1}{8} = \frac{17}{64}, \quad k = 0, 1, 2$$

$$P(z=3) = \frac{3}{8}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{3}{8} = \frac{3}{8}, \quad k = 1, 2$$

$$P(z=4) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}, \quad k = 2$$

Therefore

$$m_3(z) = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1/64 & 3/32 & 17/64 & 3/8 & 1/4 \end{pmatrix}$$

LAW OF LARGE NUMBERS

LAW OF LARGE NUMBERS p.305

Also called the *Law of Averages*, the law of large numbers is the *first fundamental theorem of probability*. It is sometimes called the *Weak Law of Large Numbers* to distinguish it from the *Strong Law of Large Numbers*. Probability may be viewed 1) *intuitively*, as the frequency at which an outcome occurs over the long run, and 2) *mathematically*, as a value of the distribution function for the random variable representing the experiment. The law of large numbers theorem shows that these two interpretations are consistent.

Let $X_1, X_2, \ldots X_n$ be an independent trials process with finite expected value $\mu = E(X_j)$ and finite variance $\sigma^2 = V(X_j)$. Let $S_n = X_1 + X_2 + \ldots + X_n$. Then for any $\varepsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \to 0 \quad \text{as } n \to \infty$$

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) \to 1 \quad \text{as } n \to \infty$$

Note that $\frac{S_n}{n}$ is the average of the individual outcomes, so $\left|\frac{S_n}{n} - \mu\right|$ is the average deviation.

In other words, if we conduct a lot of trials, the average result will be really close to the expected value.

μ = the expected value of X, $E(X)$
S_n = the sum of the random variables
n = the number of possible outcomes or the number of random variables
ε = any positive real number

CHEBYSHEV INEQUALITY p.305,316

Let X be a discrete random variable with expected value $\mu = E(X)$, and let $\varepsilon > 0$ be any positive real number. Then

$$P\left(|X - \mu| \geq \varepsilon\right) \leq \frac{V(X)}{\varepsilon^2}$$

In other words, the probability that the outcome differs from the expected value by an amount greater than or equal to the value ε is not greater than the variance divided by the square of ε.

X = random variable: the observation of an experimental outcome
μ = the expected value of X, $E(X)$
$V(X)$ = the variance of X
ε = any positive real number

CENTRAL LIMIT THEOREM

CENTRAL LIMIT THEOREM p.325

The *second fundamental theorem of probability* is the Central Limit Theorem. This theorem says that if S_n is the sum of n mutually independent random variables, then the distribution function of S_n is well-approximated by a certain type of continuous function known as the normal density function, given by the formula

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$ see p8.

σ = the deviation
σ^2 = the variance
μ = the expected value of X, $E(X)$

S_n* STANDARDIZED SUM OF S_n p.326

The standardized sum always has the expected value 0 and variance 1. A sum of variables is standardized by subtracting the expected number of successes and dividing by its standard deviation.

$$S_n^* = \frac{S_n - np}{\sqrt{jpq}} \quad \text{or} \quad S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

CENTRAL LIMIT THEOREM FOR BINOMIAL DISTRIBUTIONS p.328

For the binomial distribution $b(n,p,j)$ we have

$$\phi(x) = \lim_{n\to\infty} \sqrt{npq}\, b\left(n, p, \left\langle np + x\sqrt{npq}\right\rangle\right)$$

$\phi(x)$ = standard normal density
n = number of trials or selections
p = probability of success
q = probability of failure $(1-p)$

CENTRAL LIMIT THEOREM FOR BERNOULLI TRIALS p.330

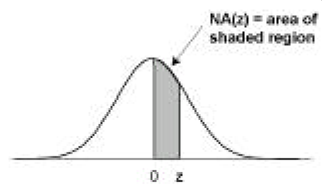
Where S_n is the number of successes in n Bernoulli trials (Bernoulli trials have 2 possible outcomes). Note that a^* and b^* are standardized values:

$$a^* = \frac{a - np}{\sqrt{npq}} \qquad b^* = \frac{b - np}{\sqrt{npq}}$$

$$\lim_{n\to\infty} P(a \le S_n \le b) = \int_{a^*}^{b^*} \phi(x)\,dx \;*$$

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

*For some reason, there is a big problem when performing this integration. The table of values in the next box are for areas under the curve of $\phi(x)$ and may be used as a close approximation instead of performing the integration. For example, for the integration from $a^* = -.2$ to $b^* = .3$, find the values of NA(z) for $z = .2$ and $z = .3$ in the table and add them together to get .1942. **Note** that in this case the values were added because they represented areas on each side of the mean (center). In the case where both values were on the same side of the mean (both have the same sign), a subtraction would have to take place to find the desired area. That is because NA(z) is the area bounded by z and the mean. Refer to the figure below.



a = lower bound
b = upper bound
$\phi(x)$ = standard normal density function
n = number of trials or selections
p = probability of success
q = probability of failure $(1-p)$

TABLE OF VALUES FOR NA(0,*z*) p.331

The area under the normal density curve from 0 to z.

z	NA(*z*)	*z*	NA(*z*)	*z*	NA(*z*)	*z*	NA(*z*)
.0	.0000	1.0	.3413	2.0	.4772	3.0	.4987
.1	.0398	1.1	.3643	2.1	.4821	3.1	.4990
.2	.0763	1.2	.3849	2.2	.4861	3.2	.4993
.3	.1179	1.3	.4032	2.3	.4893	3.3	.4995
.4	.1554	1.4	.4192	2.4	.4918	3.4	.4997
.5	.1915	1.5	.4332	2.5	.4938	3.5	.4998
.6	.2257	1.6	.4452	2.6	.4953	3.6	.4998
.7	.2580	1.7	.4554	2.7	.4965	3.7	.4999
.8	.2881	1.8	.4641	2.8	.4974	3.8	.4999
.9	.3159	1.9	.4713	2.9	.4981	3.9	.5000

CENTRAL LIMIT THEOREM FOR THE SUM OF DISCRETE VARIABLES p.343

Where S_n is the sum of n discrete random variables:

$$\lim_{n\to\infty} P\left(a < \frac{S_n - n\mu}{\sqrt{n\sigma^2}} < b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}\,dx$$

Note that $\frac{S_n - n\mu}{\sqrt{n\sigma^2}} = S_n^*$. See Standardized Sum p.17.

Note also that a and b will have to be similarly standardized before applying the Table of Values for $\mathrm{NA}(z)$ that appears previous.

a = lower bound
b = upper bound
n = number of trials or selections

CENTRAL LIMIT THEOREM – GENERAL FORM p.343

Where S_n is the sum of n discrete random variables, and we assume that the deviation of this sum approaches infinity $s_n \to \infty$:

$$\lim_{n\to\infty} P\left(a < \frac{S_n - m_n}{s_n} < b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}\,dx$$

m_n = the mean of S_n
s_n = the deviation of S_n (square root of the variance)
a = lower bound
b = upper bound
n = number of trials or selections

APPROXIMATION THEOREM p.342

For n large:

$$P\left(S_n = j\right) \sim \frac{\phi\left(x_j\right)}{\sqrt{n\sigma^2}} = \frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{(j-n\mu)^2}{2n\sigma^2}}$$

where $x_j = \frac{(j - n\mu)}{\sqrt{n\sigma^2}}$

$\phi(x)$ = standard normal density
n = number of trials or selections
p = probability of success
μ = the expected value of X, $E(X)$
σ^2 = the variance

GENERATING FUNCTIONS

$g(t)$ GENERATING FUNCTIONS p.365

A generating function $g(t)$ produces the *moments* of a random variable X. The first moment of $g(t)$ is the mean; the variance may be determined from the first and second moments of $g(t)$; and knowledge of all of the moments determines the distribution function completely. So knowing the generating function provides more information than knowing the mean and variance only. The moments of $g(t)$ are its derivatives for $t = 0$. So $g(t)$ may be called the ***moment generating function*** for X.

Discrete: $$g(t) = E\left(e^{tX}\right) = \sum_{j=1}^{\infty} e^{tx_j} p\left(x_j\right)$$

Continuous: $$g(t) = E\left(e^{tX}\right) = \int_{-\infty}^{+\infty} e^{tx} f_X(x)\,dx$$

Uniform Density: $$g(t) = E\left(e^{tX}\right) = \frac{1}{b-a}\int_a^b e^{tx}\,dx$$

Note that the limits of integration are the range of the random variable X and are not necessarily infinite. Moments may also be calculated directly; see the next box.

t = just some variable we need in order to have a generating function
j = a counting variable (integer) for the dummy variable x
x = dummy variable, I think
$f_X(x)$ = density function for the random variable X

u_n MOMENTS p.366, 394

The moments are the derivatives of the generating function evaluated at $t = 0$. They describe the mean, variance, and distribution functions of a random variable. A moment is determined by differentiating the generating function n times and then setting $t = 0$. The moments of a generating function give useful information; for example the first moment ($n = 1$) is the *mean* of the random variable for $t = 0$.

$$\mu_n = E\left(X^n\right) = \frac{d^n}{dt^n} g(t)\bigg|_{t=0}$$

Discrete: $$\mu_n = \sum_{j=1}^{\infty} \left(x_j\right)^k P\left(X = x_j\right)$$

Continuous: $$\mu_n = E\left(X^n\right) = \int_{-\infty}^{+\infty} x^n f_X(x)\, dx$$

Mean: $\mu = \mu_1$ for $t = 0$

Variance: $\sigma^2 = \mu_2 - \mu_1^2$ for $t = 0$

Sanity check: $\mu_0 = 1$ for $t = 0$

t = just some variable we need in order to have a generating function
k = a dummy counting variable (integer) for the moment calculation
n = a counting variable (integer) for the moments, where $n = 1$ for the 1st moment, $n = 2$ for the 2nd moment, etc.

$g(t)$ SPECIFIC GENERATING FUNCTIONS p.366

Following are some distribution functions and their generating functions.

Uniform distribution for $1 \le j \le n$

$$p_X(j) = \frac{1}{n} \qquad g(t) = \frac{e^t\left(e^{nt} - 1\right)}{n\left(e^t - 1\right)}$$

$$\mu = (n+1)/2 \qquad \sigma^2 = \left(n^2 - 1\right)/12$$

Binomial distribution for $0 \le j \le n$

$$p_X(j) = \binom{n}{j} p^j q^{n-j} \qquad g(t) = \left(pe^t + q\right)^n$$

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

Geometric distribution for all j

$$p_X(j) = q^{j-1} p \qquad g(t) = \frac{pe^t}{1 - qe^t}$$

$$\mu = 1/p \qquad \sigma^2 = q/p^2$$

Poisson distribution with mean λ for all j

$$p_X(j) = \frac{e^{-\lambda}\lambda^j}{j!} \qquad g(t) = e^{\lambda\left(e^t - 1\right)}$$

$$\mu = \lambda \qquad \sigma^2 = \lambda$$

X = random variable: the observation of an experimental outcome
t = just some variable we need in order to have a generating function
j = a counting variable (integer) for the dummy variable x
x = dummy variable, I think

$h(z)$ ORDINARY GENERATING FUNCTION p.370

Here are the definitions of $h(z)$, but basically to get the ordinary generating function, find $g(t)$ and replace e^t by z, replace e^{2t} by z^2, etc., and leave everything else alone.

$$h(z) = g(\log z) = \sum_{j=0}^{n} z^j p(j)$$

z = just some variable we need in order to have a generating function
j = a counting variable (integer) for the dummy variable z
$p(j)$ = coefficient of z^j in $h(z)$

PROPERTIES OF GENERATING FUNCTIONS p.371

For $Y = X + a$: $g_Y(t) = E\left(e^{tY}\right) = E\left(e^{t(X+a)}\right)$
$= e^{ta}E\left(e^{tX}\right) = e^{ta}g_X(t)$

For $Y = bX$: $g_Y(t) = E\left(e^{tY}\right) = E\left(e^{tbX}\right)$
$= g_X(bt)$

For $Y = bX + a$: $g_Y(t) = E\left(e^{tY}\right) = E\left(e^{t(bX+a)}\right)$
$= e^{ta}E\left(e^{tbX}\right) = e^{ta}g_X(bt)$

For $X^* = \dfrac{X-\mu}{\sigma}$: $g_{x^*}(t) = e^{-\mu t/\sigma}g_X\left(\dfrac{t}{\sigma}\right)$

For $Z = X + Y$: $g_Z(t) = E\left(e^{tZ}\right) = E\left(e^{t(X+Y)}\right)$
$= E\left(e^{tX}\right)E\left(e^{tY}\right) = g_X(t)\,g_Y(t)$

therefore $g_Z(t) = g_X(t)\,g_Y(t)$

also $h_Z(z) = h_X(z)\,h_Y(z)$

For $t = 0$: $g(t) = 1$

p(*j*) COEFFICIENTS OF THE ORDINARY GENERATING FUNCTION p.370

This is defined by Taylor's formula:

$$p(j) = \frac{h^{(j)}(0)}{j!}$$

For example, if $h(z) = \frac{1}{4} + \frac{1}{2}z + \frac{1}{4}z^2$ then p has values of {1/4,1/2,1/4}.

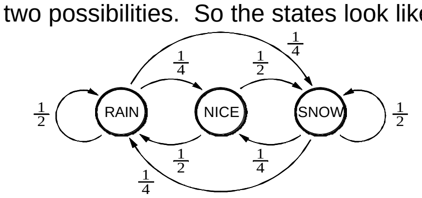
z = just some variable we need in order to have a generating function
j = a counting variable (integer) for the dummy variable z
$h(z)$ = ordinary generating function

MARKOV CHAINS

STATES p.405

A Markov chain is composed of various states with defined paths of movement between states and associated probabilities of movement along these paths. Permissible paths from one state to another are called ***steps***.

For example, let's say in the Land of Oz, there are never 2 nice days in a row. When they have a nice day, the following day will be rain or snow with equal probability. When they have snow or rain, there is a 50% chance that the following day will be the same and an equal chance of the other two possibilities. So the states look like this.



TRANSITION MATRIX p.406

A transition matrix or **P-matrix** is an arrangement of all of the probabilities of moving between states. So p_{ij} is the probability of moving from state i to state j in one step.

For the example above, the transition matrix is

$$\overline{\mathrm{P}} = \begin{array}{c} \\ \text{rain} \\ \text{nice} \\ \text{snow} \end{array} \begin{array}{c} \begin{array}{ccc} \text{rain} & \text{nice} & \text{snow} \end{array} \\ \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array}$$

So the values in the first row represent the probabilities of the weather following a rainy day, etc. Notice that the rows each sum to 1 but the columns do not. We can use the terminology p_{12} to mean the probability of having a nice day (2) after a rainy day (1). We can read the result from element p_{12} of the matrix.

MATRIX POWERS p.406

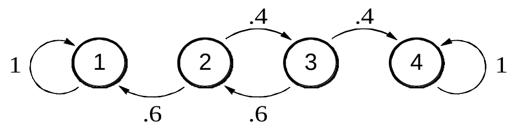
The above **P-matrix** raised to the second power gives us 2nd day probabilities, raised to a power of 3 gives us 3rd day probabilities, etc.

$$\overline{\mathrm{P}}^2 = \begin{pmatrix} .438 & .188 & .375 \\ .375 & .250 & .375 \\ .375 & .188 & .438 \end{pmatrix}$$

We use the notation $p_{12}^{(2)}$ to mean the probability of having a nice day 2 days after a rainy day, e.g. $p_{\text{rain nice}}^{(\#\text{days})} = (.188)$.

ABSORBING CHAINS p.415

A Markov chain is *absorbing* if there are one or more states from which it is not possible to leave and it is possible to get to one of these states from any state in the chain.



A state that is not absorbing (states 2 and 3 in this example) is called a ***transient state***. The P-matrix for the example above is

$$\overline{\mathrm{P}} = \begin{matrix} & 1 & 2 & 3 & 4 \\ 1 & 1 & 0 & 0 & 0 \\ 2 & .6 & 0 & .4 & 0 \\ 3 & 0 & .6 & 0 & .4 \\ 4 & 0 & 0 & 0 & 1 \end{matrix}$$

CANONICAL FORM p.417

Using the P-matrix in the previous box as an example, reorder the rows and columns so that the transient states are listed first.

$$\overline{\mathrm{P}} = \begin{matrix} & 2 & 3 & 1 & 4 \\ 2 & 0 & .4 & .6 & 0 \\ 3 & .6 & 0 & 0 & .4 \\ 1 & 0 & 0 & 1 & 0 \\ 4 & 0 & 0 & 0 & 1 \end{matrix}$$

Note that we have submatrices defined as

$$\overline{\mathrm{P}} = \left(\begin{array}{c|c} \overline{\mathrm{Q}} & \overline{\mathrm{R}} \\ \hline 0 & \overline{\mathrm{I}} \end{array} \right)$$ where

Q = a matrix to be used later to find the fundamental matrix N

R = a matrix to be used later to find the probability of absorption matrix B

0 = a matrix of zeros

I = an identity matrix

Note that in this particular example, the 4 matrices are all the same size, but this is not always the case.

FUNDAMENTAL MATRIX OF AN ABSORBING CHAIN p.418

The fundamental matrix of an absorbing chain, or the N-matrix, gives additional information. The value n_{ij} of the N-matrix is the expected number of times the chain will be in state j given that it begins in state i.

$$\overline{\mathrm{N}} = \left(\overline{\mathrm{I}} - \overline{\mathrm{Q}}\right)^{-1}$$

From our example P-matrix we have

$$N = \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & .4 \\ .6 & 0 \end{pmatrix} \right]^{-1} = \begin{matrix} & 2 & 3 \\ 2 & 1.32 & .526 \\ 3 & .789 & 1.32 \end{matrix}$$

Q = a submatrix extracted from the P-matrix canonical form and used to obtain the fundamental matrix

I = an identity matrix

TIME TO ABSORPTION p.419

The time to absorption or t-matrix, gives the number of expected steps before the chain is absorbed. In other words, given that the chain begins in state i, we can expect absorption to occur in t_i steps.

$$t = \overline{\mathrm{N}}c$$

From our example N-matrix of the previous box we have

$$t = \begin{pmatrix} 1.32 & .526 \\ .789 & 1.32 \end{pmatrix} \times \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{matrix} 2 \\ 3 \end{matrix} \overbrace{\begin{pmatrix} 1.84 \\ 2.11 \end{pmatrix}}^{\text{Number of steps to absorption}}$$

N = the fundamental matrix

c = a column matrix of ones

PROBABILITY OF ABSORPTION p.420

The probability of absorption or B-matrix, gives the probability that the chain is absorbed in state j given that it began in state i.

$$\overline{\mathrm{B}} = \overline{\mathrm{N}}\overline{\mathrm{R}}$$

From our example N-matrix of the previous box we have

$$\overline{\mathrm{B}} = \begin{pmatrix} 1.32 & .526 \\ .789 & 1.32 \end{pmatrix} \times \begin{pmatrix} .6 & 0 \\ 0 & .4 \end{pmatrix} = \begin{matrix} & 1 & 4 \\ 2 & .789 & .211 \\ 3 & .474 & .526 \end{matrix}$$

N = the fundamental matrix

R = a submatrix of the canonical form

REGULAR MARKOV CHAIN p.433

A Markov chain is called a *regular* chain if some power of the transition matrix has only positive elements. In other words, for some n, it is possible to go from any state to any state in exactly n steps. Every regular chain is also ergodic.

ERGODIC MARKOV CHAIN p.433

A Markov chain is called an *ergodic* chain if it is possible to go from every state to every other state (not necessarily in one move). Ergodic chains are sometimes called *irreducible*.

W FIXED PROBABILITY MATRIX p.434

As the transition matrix of a regular Markov chain is raised to a higher power, the result tends toward a matrix of common rows called the ***fixed probability matrix***. Sometimes you can use a calculator and raise P to a power of about 12 to see what it goes to as n gets large. Other times this doesn't work (all rows of W are not equal) and you have to use the method of Solving For w in the next box.

$$\overline{\mathrm{W}} = \lim_{n\to\infty} \overline{\mathrm{P}}^n$$

If we define w as one of the common rows of W, then

$$wP = w \quad \text{and} \quad Pc = c$$

w is called the ***fixed probability vector***. The elements of w will all be positive and will sum to one. The fact that all rows of W are the same means that the probability of arriving at a particular state after many steps is the same regardless of the starting point.

$$w\left(\overline{\mathrm{P}} - \overline{\mathrm{I}}\right) = 0$$

P = the transition matrix
c = a column matrix of ones
w = the fixed probability vector
I = an identity matrix

SOLVING FOR w p.436

Since $w\overline{\mathrm{P}} = w$ and $w_1 + w_2 + w_3 = 1$ (assuming a 3×3 P-matrix), we have 4 equations and 3 unknowns

$$w_1 + w_2 + w_3 = 1$$
$$\overline{\mathrm{P}}_{i1}\overline{\mathrm{w}} = p_{11}w_1 + p_{21}w_2 + p_{31}w_3 = w_1$$
$$\overline{\mathrm{P}}_{i2}\overline{\mathrm{w}} = p_{12}w_1 + p_{22}w_2 + p_{32}w_3 = w_2$$
$$\overline{\mathrm{P}}_{i3}\overline{\mathrm{w}} = p_{13}w_1 + p_{23}w_2 + p_{33}w_3 = w_3$$

P = the transition matrix
p_{ij} = the element from row i and column j of the transition matrix
w = the fixed probability vector
w_j = the element from column j of any row of limiting matrix W

Z FUNDAMENTAL MATRIX OF AN ERGODIC CHAIN p.456

As with absorbing chains, the fundamental matrix of an ergodic chain leads to useful information, but is found in a different way

$$\overline{\mathrm{Z}} = \left(\overline{\mathrm{I}} - \overline{\mathrm{P}} + \overline{\mathrm{W}}\right)^{-1}$$

P = the transition matrix
I = an identity matrix
W = the fixed probability matrix

MEAN FIRST PASSAGE MATRIX OF AN ERGODIC CHAIN p.459

The mean first passage matrix gives the expected number of steps from an initial state to a destination state. The mean first passage matrix is denoted by the letter M and is found one element at a time using the following formula

$$m_{ij} = \frac{z_{jj} - z_{ij}}{w_j}$$

m_{ij} = an element of the mean first passage matrix
z = an element of the fundamental matrix
w_j = an element of the fixed probability vector

SOME SAMPLE/CLASSIC PROBLEMS

MEDICAL PROBABILITIES

A drug is thought to be effective with probability x each time it is used. A beta function can be estimated to fit the probability and expected density (see Beta Density Function p12). A new probability can be given data from more recent trials and successes. Given α and β, and subsequent knowledge that there have been i successes in n new events, the probability of an event being successful is

$$P(\text{success}) = \frac{\alpha + i}{\alpha + \beta + n}$$

THE ENVELOPE PROBLEM

This is also called the *hat check* problem. n letters are randomly inserted into n addressed envelopes. What is the probability that no letter will be put into the correct envelope?

The probability that the first letter is put into the correct envelope is $1/n$. Given that the first has been placed in the proper envelope, the probability that the second one is put into the correct envelope is $1/(n-1)$ and so on. So the probability that all are put into the correct envelopes is the product of the individual probabilities or

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = \frac{1}{n!}$$

But the question was what is the probability that NO letter will be put into the correct envelope. To make a long story short, this turns out to be

$$P(\text{no letter in correct envelope}) = \frac{1}{2!} - \frac{1}{3!} + \cdots (-1)^n \frac{1}{n!}$$

THE BIRTHDAY PROBLEM

Given r people, what is the probability that there are at least two with the same birthday? It is easier to find the probability that no two will have the same birthday and subtract that from one.

Considering the first person, he could have any of 365 birthdays. Then the second person could only have one of 364 birthdays since one had been taken by the first person. The third person could have one of the 363 *unused* birthdays, etc. The sample space consists of all of the possible combinations of birthdays that the group could have.

$$P\begin{pmatrix}\text{some share}\\ \text{a birthday}\end{pmatrix} = 1 - \frac{365 \cdot 364 \cdot \ldots \cdot (365 - r + 1)}{365^r}$$

A CARD PROBLEM

A Gin hand of 10 cards is dealt. What is the probability that 4 cards belong to one suit, and there are 3 cards in each of two other suits?

$$\frac{\binom{4}{1}\binom{3}{2}\binom{13}{4}\binom{13}{3}\binom{13}{3}}{\binom{52}{10}} = 0.044$$

The equation reads, "from 4 suits choose 1 suit, from the remaining 3 suits choose 2 suits, from one suit of 13 cards choose 4 cards, from another suit choose 3 cards, and from another suit choose 3 cards. Divide the product of these by the number of 10-card hands possible from a deck of 52 cards."

GENERAL MATHEMATICAL

EULER'S EQUATION

$$e^{j\phi} = \cos\phi + j\sin\phi$$

TRIGONOMETRIC IDENTITIES

$$e^{+j\theta} + e^{-j\theta} = 2\cos\theta$$

$$e^{+j\theta} - e^{-j\theta} = j2\sin\theta$$

$$e^{\pm j\theta} = \cos\theta \pm \mathrm{j}\sin\theta$$

LOGARITHMS

$$\ln x = b \leftrightarrow e^b = x \qquad \ln x^y = y\ln x$$

$$\ln e^x = x \qquad e^{a\ln b} = b^a$$

$$\log_a x = y \leftrightarrow a^y = x$$

CALCULUS – L'HÔPITOL'S RULE

If the limit of $f(x)/g(x)$ as x approaches c produces the indeterminate form $0/0$, ∞/∞, or $-\infty/\infty$, then the derivative of both numerator and denominator may be taken

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \lim_{x\to c}\frac{f'(x)}{g'(x)}$$

provided the limit on the right exists or is infinite. The derivative may be taken repeatedly provided the numerator and denominator get the same treatment.

To convert a limit to a form on which L'Hôpital's Rule can be used, try algebraic manipulation *or* try setting y equal to the limit then take the natural log of both sides. The $\ln$ can be placed to the right of $\lim$. This is manipulated into fractional form so L'Hôpital's Rule can be used, thus getting rid of the $\ln$. When this limit is found, this is actually the value of $\ln y$ where y is the value we are looking for.

Other indeterminate forms (which might be convertible) are 1^∞, ∞^0, 0^0, $0\cdot\infty$, and $\infty-\infty$. Note that $0^\infty = 0$

CALCULUS - DERIVATIVES

$$\frac{d}{dx}\frac{u}{v} = \frac{v\cdot u' - u\cdot v'}{v^2}$$

$$\frac{d}{dx}a^x = a^x\ln a \qquad \frac{d}{dx}a^u = u'\cdot a^x\ln a$$

$$\frac{d}{dx}e^u = u'\cdot e^u$$

$$\frac{d}{dx}\ln x = \frac{1}{x} \qquad \frac{d}{dx}\ln u = \frac{u'}{u}$$

CALCULUS - INTEGRATION

$$\int dx = x + C \qquad \int x^n dx = \frac{x^{n+1}}{n+1} + C$$

$$\int e^u dx = \frac{1}{u'}\cdot e^u + C \qquad \int xe^x dx = (x-1)e^x + C$$

$$\int xe^{ax}dx = \frac{e^{ax}}{a^2}(ax-1) + C$$

$$\int_0^\infty x^n e^{-ax^2}dx = \begin{cases} \dfrac{\left[(n-1)/2\right]!}{2a^{(n+1)/2}} & \text{for odd } n \\ \dfrac{1\cdot 3\cdot 5\cdots(n-1)}{2^{(n/2)+1}a^{(n/2)}}\sqrt{\dfrac{\pi}{a}} & \text{for even } n \end{cases}$$

$$\int\frac{1}{x}dx = \ln|x| + C \qquad \int a^x dx = \frac{1}{\ln a}a^x + C$$

$$\int\sin^2 u\,du = \tfrac{1}{2}u - \tfrac{1}{4}\sin 2u + C$$

$$\int\cos^2 u\,du = \tfrac{1}{2}u + \tfrac{1}{4}\sin 2u + C$$

Integration by parts: $\int u\,dv = uv - \int v\,du$

FACTORIAL

$n!$ is the number of ways a collection of n objects can be ordered. See also Stirling Approximation.

STIRLING APPROXIMATION

Useful in calculating large factorials.

$$n! \simeq n^n e^{-n} \quad \text{or} \quad n! = n^n e^{-n}\sqrt{2\pi n}$$

e^x INFINITE SUM

Useful in the subject of probability.

$$e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!}$$

e^{-x} ANOTHER *e* THING

As $n \to \infty$, $\left(1-\frac{x}{n}\right)^n \to e^{-x}$

SERIES

$$\frac{n(n-1)}{2}=1+2+3+\cdots+(n-1)$$

$$\sqrt{1+x}\simeq 1+\frac{1}{2}x\,,\quad |x|\ll 1$$

$$\frac{1}{\sqrt{1+x}}\simeq 1-\frac{x}{2}+\frac{3x^2}{8}-\frac{5x^3}{16}+\frac{35x^4}{128}-\cdots,\quad -\tfrac{1}{2}<x<\tfrac{1}{2}$$

$$\frac{1}{1-x^2}\simeq 1+x^2+x^4+x^6+\cdots,\quad -\tfrac{1}{2}<x<\tfrac{1}{2}$$

$$\frac{1}{(1-x)^2}\simeq 1+2x+3x^2+4x^3+\cdots,\quad -\tfrac{1}{2}<x<\tfrac{1}{2}$$

$$\frac{1}{1+x}\simeq 1-x+x^2-x^3+\cdots,\quad -\tfrac{1}{2}<x<\tfrac{1}{2}$$

$$\frac{1}{1-x}\simeq 1+x+x^2+x^3+\cdots,\quad -1<x<1$$

$$\frac{x}{1-x}=x+x^2+x^3+x^4+\cdots,\quad -1<x<1$$

$$e^x=1+x+\frac{x^2}{2}+\frac{x^3}{3!}+\frac{x^4}{4!}+\cdots$$

$$\frac{x(x+1)}{2}=1+2+3+\cdots+n$$

$$\frac{x(x+1)(2x+1)}{6}=1^2+2^2+3^2+\cdots+n^2$$

BINOMIAL THEOREM

Also called binomial expansion. When m is a positive integer, this is a finite series of $m+1$ terms. When m is not a positive integer, the series converges for $-1<x<1$.

$$(1+x)^m=1+mx+\frac{m(m-1)}{2!}x^2+\cdots+\frac{m(m-1)(m-2)\cdots(m-n+1)}{n!}x^n+\cdots$$

QUADRATIC EQUATION

GIven the equation $ax^2+bx+c=0$.

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$$

LINEARIZING AN EQUATION

Small nonlinear terms are removed. Nonlinear terms include:

- variables raised to a power
- variables multiplied by other variables

Δ values are considered variables, e.g. Δt.

SPHERE

$\text{Area}=\pi d^2=4\pi r^2$ $\qquad$ $\text{Volume}=\tfrac{1}{6}\pi d^3=\tfrac{4}{3}\pi r^3$

GRAPHING TERMINOLOGY

With *x* being the horizontal axis and *y* the vertical, we have a graph of ***y* versus *x*** or ***y* as a function of *x***. The *x*-axis represents the **independent variable** and the *y*-axis represents the **dependent variable**, so that when a graph is used to illustrate data, the data of regular interval (often this is time) is plotted on the *x*-axis and the corresponding data is dependent on those values and is plotted on the *y*-axis.

GLOSSARY

derangement A permutation of elements in which the position of all elements have changed with respect to a reference permutation.

distribution function A distribution function assigns probabilities to each possible outcome. The sum of the probabilities is 1. The probability density function may be obtained by taking the derivative of the distribution function.

independent trials A special class of random variables. A sequence of random variables $X_1, X_2, \ldots X_n$ that are mutually independent and that have the same distribution is called a sequence of independent trials or an independent trials process.

median A value of a random variable for which all greater values make the distribution function greater than one half and all lesser values make it less than one half. Or, a value in an ordered set of values below and above which there is an equal number of values.

random variable A variable representing the outcome of a particular experiment. For example, the random variable X_1 might represent the outcome of two coin tosses. It's value could be HT or HH, etc.

stochastic Random; involving a random variable; involving chance or probability.

uniform distribution The probabilities of all outcomes are equal. If the sample space contains n discrete outcomes numbered 1 through n, then the uniform distribution function is $m(\omega)=1/n$.