Probability Theory

James D Emery

Last Edit: 7/4/2014

Contents

1 Introduction

The theory of probability is based on the concept of a random experiment. An experiment is random when the outcome is not known a priori. Thus, if one flips a coin, it may land on heads or tails. We do not know beforehand which outcome will happen. If we were to flip a coin 5 times in a row, we might get an outcome such as THTTHH, meaning the first flip gives heads, the second tails and so on. The set of all possible outcomes of an experiment is called the sample space. If an experiment is repeated a large number of times we may assign a probability to every point in the sample space. Thus if we flip a coin twice in a row, the sample space is the set, $\{HH, HT, TH, TT\}$. If we do this, say a thousand times. We could find HH occurring 243 times, HT occurring 253 times, TH occurring 249 times, and TT occurring 255 times. Then we can assign a probability to an event or an outcome according to its frequency. Thus the probability of HH is 243/1000, and so on. We expect that if we were to repeat the experiment a very large number of times that each outcome would get a probability very close to .25 . We expect this because if the flip is completely random then each of the four outcomes is equally likely. Now each subset of a sample space can be assigned a probability by simply assigning to the subset the sum of the probabilities of all of the points it contains.

We may model probability abstractly by assigning a probability measure

μ to a given sample space S. Clearly we must have

$$\mu(S) = 1$$

and if A and B are disjoint then

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

A random variable X is a real valued function defined on a sample space. The probability measure might be formulated in terms of a special random variable, where the sample space itself is considered to be a subset of the real numbers. Then the probability of a subset A might be written as

$$\Pr(X \in A) = \mu(A) = \int_A f(x)dx.$$

where $f(x)$ us called the probability density function. This formulation also may occur in higher dimensions where say the special random variables might be say X, Y ,and Z. There is usually a duality in probability theory, namely we think in terms of an abstract measure space, and secondly we think in terms of a random experiment. So we ask, "What is the probability that $X \in A$." We may mean $\mu(A)$, but we also may mean that an experiment, physical or otherwise, is performed, and the outcome is a number. If repeated an "infinite" number of times, there would be a relative frequency of $\mu(A)$ of the number being in the subset A.

Central to the theory of probability is the concept of independence. and independent events.

Suppose the probability of an event A is $\mu(A)$ and the probability of an event B is $\mu(B)$. Then the conditional probability of event B given that event A has occurred is

$$\mu(B|A) = \frac{\mu(B \cap A)}{\mu(A)}.$$

To explain this consider the case of two coin flips. The sample space is

$$\{HH, HT, TH, TT\}$$

What is the probability of a tail occurring given that a head has occurred? Let A be the event that a head has occurred. Then

$$A = \{HH, HT, TH\}$$

$$\mu(A) = .75$$

Let B be the event that a tail has occurred. Then

$$B = \{HT, TH, TT\}$$

$$\mu(B) = .75$$

Then the sample space for the conditional probability is

$$A = \{HH, HT, TH\}$$

And the event of a tail in this sample space is

$$B' = \{HT, TH\} = A \cap B$$

Hence the conditional probability is clearly

$$\frac{2}{3}$$

That is

$$\mu(B|A) = \frac{\mu(A \cap B)}{\mu(A)} = \frac{2}{3}.$$

Two events are independent iff the probability of B does not depend on A, thus

$$\mu(B|A) = \mu(B).$$

In that case

$$\mu(B) = \mu(B|A) = \frac{\mu(B \cap A)}{\mu(A)}.$$

So for independent events

$$\mu(B \cap A) = \mu(B)\mu(A).$$

In the case of two coin flips, let A be the occurrence of a head on the first flip. Let B be the occurrence of a tail on the second flip. Then A and B are independent and

$$\mu(A \cap B) = \mu(A)\mu(B) = (.5)(.5) = .25.$$

Consider an urn containing m black balls and n red balls. Let A be the event of selecting a black ball on the first draw and B the event of selecting a red ball on the second draw. Whether A and B are independent depends upon whether the first drawn ball is replaced. We have

$$\mu(B|A) = \frac{\mu(A \cap B)}{\mu(A)}.$$

Clearly

$$\mu(A) = \frac{m}{m+n}$$

If the first drawn ball is replaced, then

$$\mu(B|A) = \frac{n}{m+n} = \mu(B),$$

and the events are independent so that

$$\mu(A \cap B) = \left[\frac{m}{m+n}\right]\left[\frac{n}{m+n}\right].$$

However, if the first drawn ball is not replaced then

$$\mu(A) = \frac{m}{m+n},$$

but

$$\mu(B|A) = \frac{n}{m+n-1}.$$

So the probability of drawing a black ball and then a red ball is

$$\mu(A \cap B) = \mu(A)\mu(B|A) = \left[\frac{m}{m+n}\right]\left[\frac{n}{m+n-1}\right].$$

Let X be a random variable

$$X : S \to \Re$$

The measure of a subset A of $\Re$ is

$$\Pr(X \in A) = \mu(X^{-1}(A)).$$

Suppose the subset is I a small interval of length Δx containing x , then we can form a kind of derivative

$$\frac{\mu(X^{-1}(I)}{dx} = \lim_{\Delta x \to 0} \frac{\mu(X^{-1}(I)}{\Delta x} = f(x)$$

So we may write

$$\Pr(X \in A) = \int_A f(x)dx.$$

The function $f(x)$ is called the probability density function, or pdf, for the random variable X. Similarly for two random variables X and Y there is a joint pdf $f(x, y)$ so that

$$\Pr(X \in A, Y \in B) = \int_A \int_B f(x, y)dxdy.$$

In terms of the joint pdf $f(x, y)$, we have

$$\Pr(X \in A) = \int_A \int_{-\infty}^{\infty} f(x, y)dxdy$$

$$= \int_A f_1(x)dx,$$

where

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)dy.$$

And

$$\Pr(Y \in B) = \int_{-\infty}^{\infty} \int_B f(x, y)dxdy$$

$$= \int_B f_2(y)dy,$$

where

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)dx.$$

The functions $f_1(x), f_2(y)$ are called the marginal pdf's. If $f(x, y)$ is a product of a function in x and a function in y, then

$$f(x, y) = f_1(x)f_2(y).$$

In this case the random variables X and Y are stochastically independent. Indeed

$$\mu(X^{-1}(A) \cap Y^{-1}(B)) = \Pr(X \in A, Y \in B)$$

$$= \int_A \int_B f(x, y)dxdy$$

$$
\begin{aligned}
&= \int_A \int_B f_1(x) f_2(y) dx dy \\
&= \int_A f_1(x) dx \int_B f_2(y) dy. \\
&= \Pr(X \in A) \Pr(Y \in B) \\
&= \mu(X^{-1}(A)) \mu(Y^{-1}(B)).
\end{aligned}
$$

We can always write

$$f(x, y) = f_1(x) g(x, y)$$

The function $g(x, y)$ which we might write as $f(y|x)$ is called the conditional pdf. If X and Y are independent then we must have $g(x, y) = f_2(y)$. For otherwise we could find special sets A and B so that

$$\mu(X^{-1}(A) \cap Y^{-1}(B))$$

is not equal to

$$= \mu(X^{-1}(A)) \mu(Y^{-1}(B)),$$

which would contradict the independence of X and Y.

2 Expectation, Moments, Mean, and Variance

Suppose we have an abstract probability measure space A with probability measure m. Then

$$\int_A dm = m(A) = 1.$$

The integral here is an abstract integral defined on a measure space with say Lebesgue measure. For more on such integrals see a book on measure theory, or a book on real analysis. Given a function g defined on set A, the expected value of g is defined as

$$E(g) = \int_A g dm.$$

In the case of a continuous distribution on the real line with pdf $f(x)$, where the ordinary Riemann integral exists, this becomes

$$E(g) = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

In the case of a discrete distribution say on the natural numbers with the probability of k equal to p_k this becomes

$$E(g) = \sum_{k=1}^{\infty} g_k p_k.$$

We have similar expressions for the expectation for such cases as a continuous distribution in n space or say a finite distribution of possible poker hands. In the case of gambling we talk about the related concept of the expected winnings. The nth moment of a continuous distribution on the real line is the expectation of the nth power of x,

$$E(x^n) = \int_{-\infty}^{\infty} x^n f(x) dx.$$

The first moment is called the mean or average

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

The variance is the expectation of $(x - \mu)^2$

$$\sigma^2 = E((x - \mu)^2) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

From its definition we see that the expectation is linear. That is,

$$E(g + h) = E(g) + E(h),$$

and

$$E(cg) = cE(g),$$

where c is a constant. By expanding $(x - \mu)^2$ one sees that

$$\sigma^2 = E((x - \mu)^2) = E(x^2) - 2\mu E(x) + E(\mu^2) = E(x^2) - \mu^2.$$

If we consider the probability to be a weight, then the variance is like the moment of inertia in mechanics, and the mean is like the center of mass.

3 Bayesian Statistics

We may have an outcome event A due to several random causes $A_1, A_2, ..., A_m$. Bayesian methods may allow us to assess the likelihood of these causes. Let $A_1, A_2, ..., A_m$ be disjoint subsets of the sample space. Let A be a subset of the union of these sets

$$A \subset \cup_{i=1}^{m} A_i.$$

Then the probability of an outcome in A is equal to a sum of the weighted conditional probabilities. We have

$$A = \cup_{i=1}^{m} (A_i \cap A).$$

Hence the probability measure of A is

$$P(A) = \sum_{i=1}^{m} P(A_i \cap A)$$

$$= \sum_{i=1}^{m} P(A|A_i)P(A_i).$$

Then for each i, Bayes formula for the ith conditional probability is

$$P(A_j|A) = \frac{P(A \cap A_j)}{P(A)}$$

$$= \frac{P(A \cap A_j)}{P(A)} \frac{P(A_j)}{P(A_j)}$$

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

$$= \frac{P(A|A_j)P(A_j)}{\sum_{i=1}^{m} P(A_i \cap A)}$$

$$= \frac{P(A|A_j)P(A_j)}{\sum_{i=1}^{m} P(A|A_i)P(A_i)}$$

By comparing all of the conditional probabilities $P(A_i|A)$, for $i = 1, 2, ..., m$, one may make a decision about the most likely cause of A, or or about the most likely symptom associated with A.

Example 1 (See Hogg and Craig, p54, Problem 2.8)

Urn 1 contains 3 red chips and 7 blue chips. Urn 2 contains 6 red chips and 4 blue chips. An urn is selected and then a chip removed. Let A_1 be the event that urn 1 is selected and A_2 the event that urn 2 is selected. Let A be the event that a red chip is removed. We have

$$P(A_1) = 1/2$$

$$P(A_2) = 1/2$$

$$P(A|A_1) = 3/10$$

$$P(A|A_2) = 6/10.$$

(a) What is the probability of A? This probability is

$$P(A) = \sum_{i=1}^{m} P(A|A_i)P(A_i) = \frac{3}{10}\frac{1}{2} + \frac{6}{10}\frac{1}{2} = \frac{9}{20}.$$

(b) Given that the chip removed is red, what is the probability that it was drawn from the second urn? This probability is given by Bayes' formula as

$$P(A_2|A) = \frac{P(A_2)P(A|A_2)}{P(A_1)P(A|A_1) + P(A_2)P(A|A_2)} = \frac{2}{3}.$$

Let us compute these probabilities by brute force by counting equally likely events. Let us number the chips in urn 1 from 1 to 10, where chips 1,2,3 are red. Let us number the chips in urn 2 from 1 to 10 where chips 1,2,3,4,5,6 are red. Then the sample space is the set of points

$$\{(m, n) : 1 \le m \le 2, 1 \le n \le 10\}.$$

Each point has probability 1/20. Let us count using a computer program: Here is the program

```
import java.io.*;
//Bayes example
public class bayes{
 public static void main(String args[]){
  boolean flush = true;
  int urn;
  int chip;
  int color;
  int a;
  int a2bara;

  PrintWriter out = null;
```

```
  File data = new File("out.txt");
  try{
   out = new PrintWriter(new BufferedWriter(new FileWriter(data)),flush);
  }
   catch(IOException e){
   System.out.println(e);
  }
  a=0;
  a2bara=0;
  for(int i=0;i<2;i++){
   urn=i+1;
   for(int j=0;j<10;j++){
    chip=j+1;
    if( urn == 1){
     if( chip <= 3){
      color=1;
     }
     else{
      color=2;
     }
    }
    else{
     if( chip <= 6){
      color=1;
     }
     else{
      color=2;
     }
    }
    if(color == 1){
     System.out.println(" (" + urn + "," + chip + ") red");
     a=a+1;
     if(urn == 2){
      a2bara= a2bara+1;
     }
    }
    else{
     System.out.println(" (" + urn + "," + chip + ") blue ");
    }
   }

  }
  System.out.println(" Probability of red = " + a/20.);
  System.out.println(" Probability of urn 2 given chip is red = " + ((double)a2bara)/a);
  out.close();
 }
}
```

Here is the program output:

```
 (1,1) red
 (1,2) red
 (1,3) red
 (1,4) blue
 (1,5) blue
 (1,6) blue
 (1,7) blue
```

```
(1,8) blue
(1,9) blue
(1,10) blue
(2,1) red
(2,2) red
(2,3) red
(2,4) red
(2,5) red
(2,6) red
(2,7) blue
(2,8) blue
(2,9) blue
(2,10) blue
Probability of red = 0.45
Probability of urn 2, given chip is red = 0.6666
```

Example 2, Drug Testing There is a review of the book **More Sex is Safer Sex**, in the Notices of the American Mathematical Society, June/July 2008. This book is related to the best seller called **Freakonomics**. These books deal with topics that involve Bayesian Statistics. The review is titled **Economics and Common Sense**, and the author is Gil Kalai. He debunks some of the results given in these two books. There is a discussion of a nonintuitive result in an AIDS test, given in **More Sex is Safer Sex**, which Kalai disputes.

Let us formulate a similar nonintuitive result in drug testing. Let us suppose that employees are being tested for the use of opium. Suppose that the test is wrong 5 percent of the time. Suppose this is both for a false positive and for a false negative. So suppose the event of being an opium user is O. The event of not being an opium user, that is of being free of opium use, is F. The event of testing positive for opium use is P and the event of testing negative for opium is N. Suppose there are few opium users and that the probability of O is 1 percent. So then we have

$$Pr(O) = .01$$

$$Pr(F) = .99$$

These are usually called prior probabilities. We have also the 5 percent test error probabilities

$$Pr(P|O) = .95$$

$$Pr(P|F) = .05$$

Now suppose you test positive for opium use. What is the probability that you are actually a user? Since the error of the test is 5 percent, we intuitively think that it is 95 per cent certain that you are an opium user. However, the probability to be evaluated is $Pr(O|P)$, which is given by Baye's law as

$$Pr(O|P) = \frac{Pr(P|O)Pr(O)}{Pr(P)}$$

$$= \frac{Pr(P|O)Pr(O)}{Pr(P \cap O) + Pr(P \cap F)}$$

$$= \frac{Pr(P|O)Pr(O)}{Pr(P|O)Pr(O) + Pr(P|F)Pr(F)}$$

$$= \frac{(.05)(.01)}{(.95)(.01) + (.05)(.99)} = .16102$$

So the probability is small that a positive test means that you are a user. This might make one a little reluctant to take a drug test. This result happens because there are so many drug non user candidates for experiencing a test error.

In the case of the AIDS test mentioned in the review, the reviewer criticizes this result. In the AIDS case one suspects that those who take such a test have reason to believe that they may have AIDS, so a relatively large percentage of the test takers may in fact have AIDS, and so are not a sample from the general population.

4 Discrete Probability Distributions: The Binomial Distribution

A discrete probability distribution is one in which the random variable X takes on discrete values, meaning noncontinuous separated values. A special case is a finite distribution. Suppose we throw a die and proclaim success if a 1 turns up, and a failure otherwise. Then the probability of success is $p = 1/6$. The probability of failure is $5/6$. Hence our sample space consists of two points $\{S, F\}$ and we assign our probability measure to be $m(S) = 1/6$ and $m(F) = 5/6$. Suppose we repeat this experiment 5 times. So an outcome might be SFSSF. So the probability of this event would be

$$p(1-p)pp(1-p) = p^3(1-p)^2 = 25/7776.$$

Suppose another trial gives SSSFF. Again we have 3 successes and so the probability of this is again

$$p^3(1-p)^2 = 25/7776.$$

Now what is the total probability of 3 successes in 5 trials. Let us identify our outcome of 3 success with the times when an S occurs. Thus our first event could be written as $\{1, 3, 4\}$ meaning the first throw, the third throw, and the fourth through were successes. Similarly the second event could be represented as $\{1, 2, 3\}$. So the number of 3 successes in 5 trials will be the number of ways of choosing 3 out of 5, namely the number of combinations of 5 things taken 3 at a time

$$C_3^5 = \frac{5!}{3!(5-3)!} = 10$$

Hence the total probability of 3 successes in 5 trials is

$$C_3^5 p^3(1-p)^2 = 250/7776.$$

More generally suppose we repeat our independent throws n times, then the probability of k successes is

$$C_k^n p^k (1-p)^{(n-k)}.$$

The binomial distribution is the distribution in the sample space of $S = \{0, 1, 2, 3, ..., n\}$, where the probability of $k \in S$ is

$$C_k^n p^k (1-p)^{(n-k)}.$$

This is the probability of k successes in n trials. The probability of the whole sample space itself Prob(S) must be 1.

Notice that by the binomial theorem

$$1 = (p + (1-p))^n = \sum_{k=0}^{n} C_k^n p^k (1-p)^{(n-k)},$$

which gives the binomial distribution its name. If we were to flip a coin n times and count the number of heads, then we would have a binomial distribution where the random variable X would be the number of heads and the value of p would be 1/2.

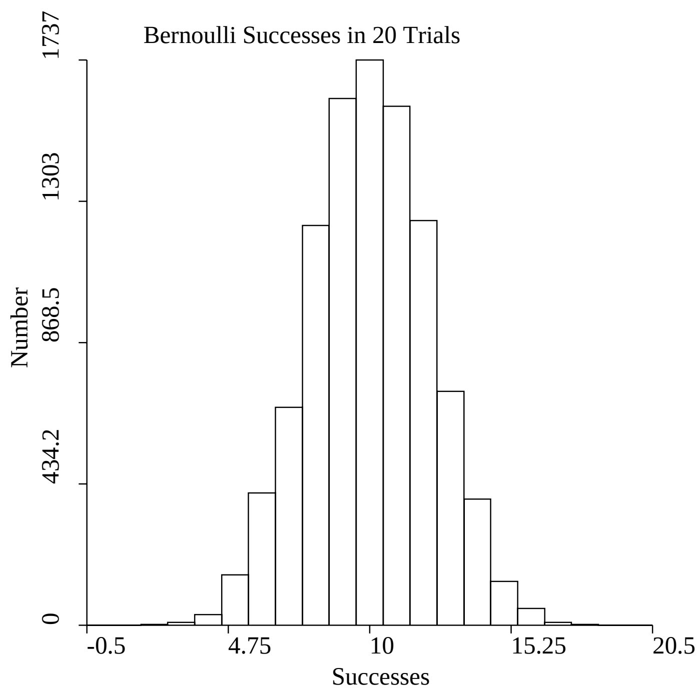

Figure 1: A sample from the binomial distribution, the number of successes in 20 trials with probability of success 1/2. The data was generated by program **binomialsample.ftn**, and 10000 samples were generated. The histogram was created by program **histogram.ftn**.

Using the moment genating function, one sees that the mean of the binomial distribution is

$$\mu = np,$$

and the variance is

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

Consider the case $p = 1/2$ and $n = 20$, then the probability of eight successes in 20 trials would be

$$C_k^n p^k (1-p)^{(n-k)} = \frac{20!}{8!12!}(1/2)^{20} = \frac{125970}{1048576} = .1201343536$$

Thus in 10,000 trials one should get about

$$(.1201)(10000) = 1201$$

cases of 8 successes.

Using program **binomialsamp.ftn**, we do calculations for such a sample:

```
 n= number of trials
 p= probability of success
 Enter n and p [10,.5]
  20 .5
 Enter the number of points in the sample [10000]
  10000
 Enter the file to hold the sample [a.txt]
  a.txt
Sample mean =   10.007600
Sample standard deviation =   2.2272041
Theoretical mean =   10.000000
Theoretical sdev =   2.2360680
Number of   0 successes       0
Number of   1 successes       0
Number of   2 successes       2
Number of   3 successes       8
Number of   4 successes      33
Number of   5 successes     155
Number of   6 successes     407
Number of   7 successes     670
Number of   8 successes    1229
Number of   9 successes    1619
Number of  10 successes    1737
Number of  11 successes    1595
Number of  12 successes    1244
Number of  13 successes     719
Number of  14 successes     387
Number of  15 successes     134
Number of  16 successes      51
Number of  17 successes       8
Number of  18 successes       2
Number of  19 successes       0
Number of  20 successes       0
```

From the Figure, which shows a histogram of the simulation of this binomial example, we see that the distribution is centered at the mean 10, and the shape of the distribution is beginning to look like a normal distribution. This figure was produced using programs **binomialsamp.ftn** and **histogram.ftn**. In fact one can show that the binomial distribution for large n, is approximated in some sense by the normal distribution. One should also get some insight into why a random variable, that is equal to a sum of n independent random variables, tends to be normally distributed. Program **binomialsamp.ftn** works by using a random number generator that returns a random number x between 0 and 1. In this case if $x \leq p$, it is counted a success. Here is a listing of the program

```
c binomialsamp.ftn write a sample from a binomial distribution to a file
c 3/17/09
      implicit real*8(a-h,o-z)
      parameter (np=100000)
      dimension x(np)
      integer s(5000)
      character*30 fname
      dimension a(10)
      nf=0
      write(*,*)' n= number of trials '
      write(*,*)' p= probability of success '
      write(*,*)' Enter n and p [10,.5] '
      call readr(nf, a, nr)
      if(nr .eq. 2)then
       n=a(1)
       p=a(2)
      else
       n=10
       p=.5
      endif
      write(*,*)' Enter the number of points in the sample [10000] '
      call readr(nf, a, nr)
      if(nr .eq. 1)then
       ns=a(1)
      else
       ns=10000
      endif
      write(*,*)' Enter the file to hold the sample [a.txt] '
      read(*,'(a)')fname
      if(lenstr(fname) .eq. 0)then
       fname='a.txt'
      endif
      nf1=2
      open(nf1,file=fname,status='unknown')
      zero=0.
      iran=6789
      do i=1,ns
       s(i)=0
      enddo
      do i=1,ns
```

```
       call bsamp(iran,n,p,k)
       x(i)=k
       s(k+1)=s(k+1)+1
      write(nf1,'(1x,i6)')k
      enddo
      call meansdv(x,ns,am,sdv)
      write(*,'(a,g15.8)')' Sample mean = ',am
      write(*,'(a,g15.8)')' Sample standard deviation = ',sdv
      write(*,'(a,g15.8)')' Theoretical mean = ',n*p
      write(*,'(a,g15.8)')' Theoretical sdev = ',sqrt(n*p*(1-p))
      m=n+1
      do i=1,m
       write(*,'(a,i3,a,i8)')' Number of ',i-1,' successes ',s(i)
      enddo
      end
c+ bsamp  binomial random variate, k successes in n trials of probability p
      subroutine bsamp(iran,n,p,k)
      implicit real*8(a-h,o-z)
c  Input:
c  iran   seed on first input, next random integer on output
c          1 <= jran < 121500
c
c  Output:
c  k      number of successes

      one=1.
      k=0
      do i=1,n
       call randj(iran,r)
       if(r .le. p)then
        k=k+1
       endif
      enddo
      return
      end

c+ randj  congruential random number generator
        subroutine randj(jran,r)
        implicit real*8(a-h,o-z)
c  parameters
c   jran=seed on input, next random integer on output
c          1 <= jran < 121500
c   r=real random number between 0. and 1.
c  (see table in book 'numerical recipes')
c  (period is 121500, i.e. repeats after 121500 calls)
c   works for 32 bit integers
      data im,ia,ic /121500,2041,25673/
      a=im
      jran=jran*ia+ic
      jran=mod(jran,im)
c     r=mod(jran*ia+ic,im)/(real(im)
      r=jran/a
      return
      end
c+ meansdv  mean and standard deviation of array.
      subroutine meansdv(x,n,amean,sdv)
      implicit real*8(a-h,o-z)
```

```
c mean and standard deviation of x.
c n, number of values in x.
      dimension x(*)
      amean=0.
      do i=1,n
       amean=amean+x(i)
      enddo
      amean=amean/n
      var=0.
      do i=1,n
       var=var+(x(i)-amean)**2
      enddo
      var=var/float(n-1)
      sdv=sqrt(var)
      return
      end

c+ readr read a row of numbers and return in double precision array
      subroutine readr(nf, a, nr)
      implicit real*8(a-h,o-z)
c Input:
c nf    unit number of file to read
c       nf=0 is the standard input file (keyboard)
c Output:
c a     array containing double precision numbers found
c nr    number of values in returned array,
c     or 0 for empty or blank line,
c       or -1 for end of file on unit nf.
c   Numbers are separated by spaces.
c   Examples of valid numbers are:
c   12.13 34 45e4 4.78e-6 4e2,5.6D-23,10000.d015
c requires subroutine valsub and function lenstr
c a semicolon and all characters following are ignored.
c This can be used for comments.
c modified 6/16/97 added semicolon feature
      dimension a(*)
      character*200  b
      character*200  c
      character*1 d
        c=' '
      if(nf.eq.0)then
         read(*,'(a)',end=99)b
      else
         read(nf,'(a)',end=99)b
      endif
      nr=0
      lsemi=index(b,';')
      if(lsemi .gt. 0)then
       if(lsemi .gt. 1)then
        b=b(1:(lsemi-1))
       else
        return
       endif
      endif
      l=lenstr(b)
      if(l.ge.200)then
        write(*,*)' error in readr subroutine '
```

```
         write(*,*)' record is too long '
       endif
       do 1 i=1,l
        d=b(i:i)
        if (d.ne.' ') then
         k=lenstr(c)
         if (k.gt.0)then
            c=c(1:k)//d
         else
            c=d
         endif
        endif
        if( (d.eq.' ').or.(i.eq.l)) then
         if (c.ne.' ') then
           nr=nr+1
            call valsub(c,a(nr),ier)
           c=' '
         endif
        endif
1      continue
       return
99     nr=-1
       return
       end
c+ valsub converts string to floating point number (r*8)
       subroutine valsub(s,v,ier)
       implicit real*8(a-h,o-z)
c      examples of valid strings are: 12.13 34 45e4 4.78e-6 4E2
c      the string is checked for valid characters,
c      but the string can still be invalid.
c       s-string
c       v-returned value
c       ier- 0 normal
c            1 if invalid character found, v returned 0
c
       logical p
       character s*(*),c*50,t*50,ch*15
       character z*1
       data ch/'1234567890+-.eE'/
       v=0.
       ier=1
       l=lenstr(s)
       if(l.eq.0)return
       p=.true.
       do 10 i=1,l
       z=s(i:i)
       if((z.eq.'D').or.(z.eq.'d'))then
           s(i:i)='e'
       endif
       p=p.and.(index(ch,s(i:i)).ne.0)
 10    continue
       if(.not.p)return
       n=index(s,'.')
       if(n.eq.0)then
           n=index(s,'e')
           if(n.eq.0)n=index(s,'E')
           if(n.eq.0)n=index(s,'d')
```

```
          if(n.eq.0)n=index(s,'D')
          if(n.eq.0)then
             s=s(1:l)//'.'
          else
            t=s(n:l)
            s=s(1:(n-1))//'.'//t
          endif
          l=l+1
      endif
      write(c,'(a30)')s(1:l)
      read(c,'(g30.23)')v
      ier=0
      return
      end
c+ lenstr   nonblank length of string
      function lenstr(s)
c  length of the substring of s obtained by deleting all
c  trailing blanks from s.  thus the length of a string
c  containing only blanks will be 0.
      character   s*(*)
      lenstr=0
      n=len(s)
      do 10 i=n,1,-1
      if(s(i:i) .ne. ' ')then
         lenstr=i
         return
      endif
10    continue
      return
      end
```

5 Belief Functions in Decision Theory

Belief functions are usually based on Bayesian methods. (To be expanded)

6 The Error Function

We have

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

This can be calculated by considering

$$I = \int_0^{\infty} e^{-x^2} dx,$$

and

$$I^2 = \int_0^{\infty} \int_0^{\infty} e^{-x^2} e^{-y^2} dx dy.$$

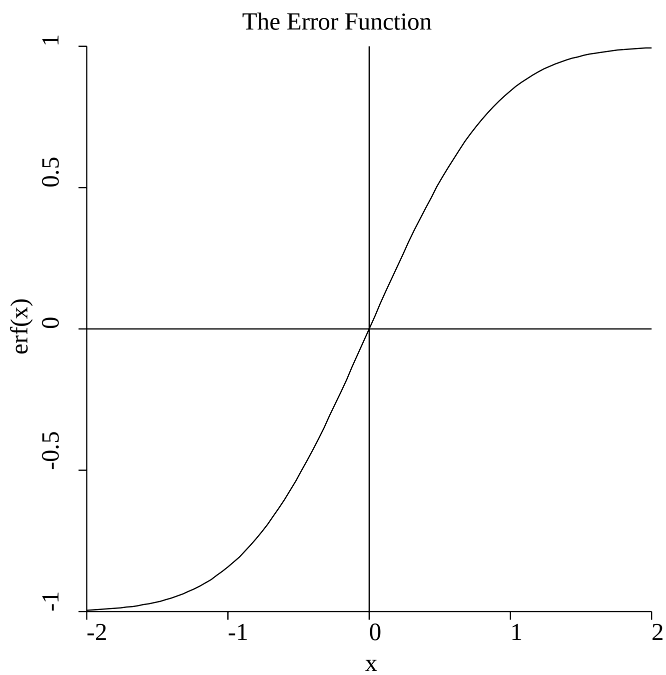

Figure 2: The Error function erf(x) is defined as $\frac{2}{\sqrt{\pi}}\int_0^x e^{-x^2}dx$.

Changing to polar coordinates, we find that

$$I^2 = \frac{\pi}{4},$$

$$I = \frac{\sqrt{\pi}}{2},$$

The error function is defined as

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du,$$

so

$$erf(\infty) = 1.$$

We have by the definition of the integral, for all x,

$$\frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du = -\frac{2}{\sqrt{\pi}}\int_x^0 e^{-u^2}du$$

and in particular for $-x$

$$erf(-x) = \frac{2}{\sqrt{\pi}}\int_0^{-x} e^{-u^2}du = -\frac{2}{\sqrt{\pi}}\int_{-x}^0 e^{-u^2}du.$$

And notice that because of the symmetric nature of

$$e^{-u^2},$$

that

$$\frac{2}{\sqrt{\pi}}\int_{-x}^0 e^{-u^2}du = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du.$$

So

$$\begin{aligned} erf(-x) &= \frac{2}{\sqrt{\pi}}\int_0^{-x} e^{-u^2}du \\ &= -\frac{2}{\sqrt{\pi}}\int_{-x}^0 e^{-u^2}du \\ &= -\frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du \\ &= -erf(x). \end{aligned}$$

So

$$erf(-x) = -erf(x).$$

That is the error function is an odd function. We only need compute the error function for nonnegative values. When the argument x is negative, the error function is then

$$erf(x) = -erf(-x).$$

The error function can be computed in various ways. One accurate, but not terribly efficient method is to use numerical integration. The function erf(x), which is in library emerylib.ftn, is done using Romberg integration. A program using this function is given below. Romberg integration is the method of using the trapezoid method together with Richardson extrapolation. This is possible according to the Euler-Maclaurin Summation Formula. See my Numerical Analysis (numanal.tex) and the book by Dahlquist and Bjorck **Numerical Methods** Prentice-Hall, 1974, chapter seven.

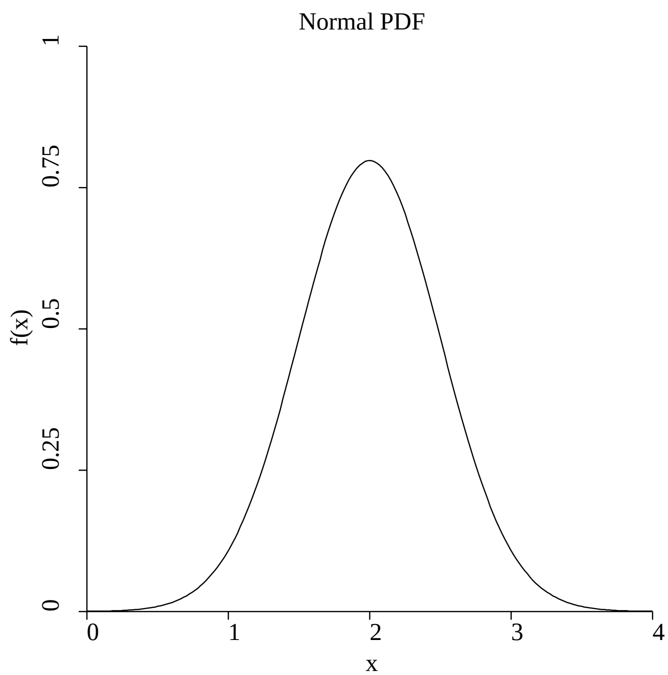

Figure 3: The Normal pdf with mean $\mu = 2$ and variance $\sigma^2 = 1$.

7 The Normal Distribution

The pdf (probability density function) of the normal distribution with mean μ and variance σ^2 is

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

We have

$$\begin{aligned}
&\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/(2\sigma^2)}dx \\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2}\sqrt{2}\sigma dy \\
&= \sqrt{2}\sigma\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2}dy \\
&= 1,
\end{aligned}$$

where we have used the substitution

$$y = \frac{x-\mu}{\sigma\sqrt{2}}.$$

The standard normal distribution has mean $\mu = 0$ and variance $\sigma^2 = 1$. The distribution function is

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$$

The standard normal distribution can be expressed using the error function $erf(x)$. So let $z = y/\sqrt{2}$, then $dz\sqrt{2} = dy$ and

$$F(x) = \frac{\sqrt{2}}{\sqrt{2\pi}} \int_{-\infty}^{x/\sqrt{2}} e^{-z^2} dz$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x/\sqrt{2}} e^{-z^2} dz$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{0} e^{-z^2} dz + \frac{1}{\sqrt{\pi}} \int_{0}^{x/\sqrt{2}} e^{-z^2} dz$$

$$= \frac{1}{\sqrt{\pi}} \int_{0}^{\infty} e^{-z^2} dz + \frac{1}{\sqrt{\pi}} \int_{0}^{x/\sqrt{2}} e^{-z^2} dz$$

$$= \frac{1}{2} erf(\infty) + \frac{1}{2} erf(x/\sqrt{2})$$

$$= \frac{1}{2}(1 + erf(x/\sqrt{2}))$$

If

$$y = \frac{1}{2}(1 + erf(x/\sqrt{2}))$$

then

$$erf(x/\sqrt{2}) = 2y - 1,$$

so

$$x = \sqrt{2}(erf)^{-1}(2y-1)$$

So the inverse of the normal distribution $F(x)$ is

$$F^{-1}(y) = \sqrt{2}(erf)^{-1}(2y-1)$$

If X has a normal distribution with mean μ and variance σ^2, then

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

has the standard normal distribution. Conversely, if Y has the standard normal distribution then

$$X = \mu + \sigma Y$$

has a normal distribution with mean μ and variance σ^2.

8 The Normal Distribution and the Inverse Normal Distribution in Matlab

Matlab and Octave have the error function erf(x) and its inverse erfinv(x). From above we can compute the standard normal distribution function and its inverse from erf(x) and its inverse erfinv(x). So we can create a function script, an m-file for each of these functions. The first one is called ndist.m and is simply the two lines

```
function v = ndist(x)
 v=.5*(1+erf(x/sqrt(2)));
```

The second is called ndistinv.m and is

```
function v = ndistinv(y)
v = sqrt(2.)*erfinv(2.*y-1);
```

To use these script functions we must change the working directory of matlab to the directory that contains these scripts, that is these files ndist.m and ndistinv.m. Then for example if we type in the matlab command

ndistinv(.4)

we will obtain the number

-.2533

This means that the interval from $(-\infty, -.2533)$ has a .4 probability.

9 Grading On the Curve

So suppose one has a set of scores on a test and the mean is $\mu = 55$ and the standard deviation is 30. Suppose the scores are approximately normally distributed. Suppose we want to break up the range of scores into 5 intervals so that there will be about a 20 percent chance of a score falling in each interval.

We find

```
ndistinv(.2) = -0.8416
ndistinv(.4) = -.2533
ndistinv(.6) = .2533
ndistinv(.8) = 0.8416
```

Of course these values could also be found approximately from a table of the standard normal distribution function values in a mathematical handbook.

Then our breakpoints would be

$$\mu - .8416\sigma = 29.75$$

$$\mu - .2533\sigma = 47.4$$

$$\mu + .2533\sigma = 62.6$$

$$\mu + .8416\sigma = 80.2$$

So a score lower than 29.75 is an F, a score between 29.75 and 47.4 is a D, a score between 47.4 and 62.6 is a C, a score between 62.6 and 80.2 is a B, and a score greater than 80.2 is an A.

This is the so called "grading on the curve." One is assuming that test scores are normally distributed. In one sense they can't be because there is a minimum score and a maximum score, which is not true of a normal distribution. The distribution of scores depends heavily on the design of the test. The program histogram.ftn is useful for determining whether a set of numbers is normally distributed, and could be used to experiment with grade ranges. Here is an example of running the program histogram.ftn. The default file name was accepted by entering return. This particular file contained 100 points that was generated by a normal variate program with specified mean 50 and standard deviation 5.

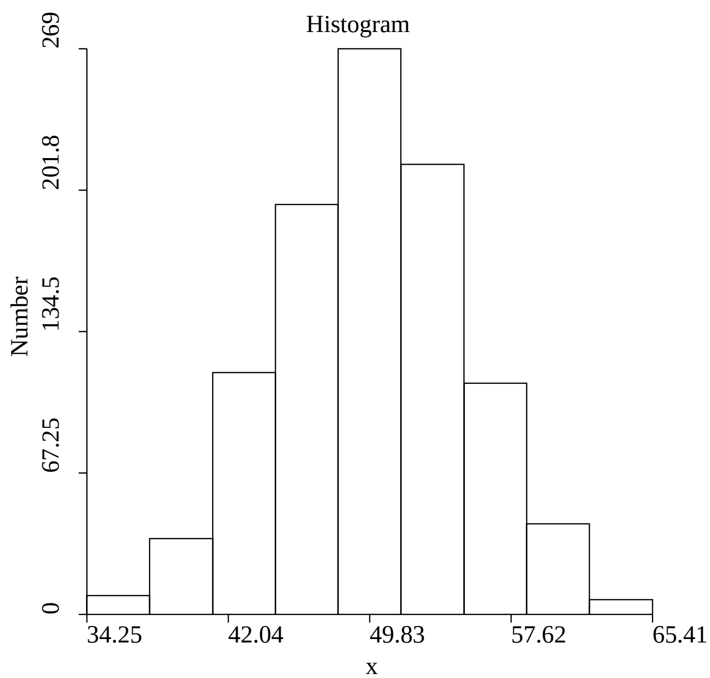

Figure 4: A histogram created by program histogram.ftn from a data file containing 100 scores.

```
Enter the file name for the data [x.txt]
 Sample mean =   49.834085
 Sample standard deviation =   5.1930853
 xmn= =   29.329064
 xmx= =   64.605969
 2 sigma= (  44.641000    ,  55.027171    )
 4 sigma= (  39.447915    ,  60.220256    )
 6 sigma= (  34.254829    ,  65.413341    )
 Enter bin range [xmin,xmax]
 Enter number of equally spaced bins [10]
 number bins=          9
 v(1)=        9.000000000000000
 number of points      =        1000
 points placed in bins =         998
 bin   1 (  34.25    ,  37.71    )   9.000     percentage=  .9000
 bin   2 (  37.71    ,  41.17    )   36.00     percentage=  3.600
 bin   3 (  41.17    ,  44.64    )   115.0     percentage=  11.50
 bin   4 (  44.64    ,  48.10    )   195.0     percentage=  19.50
 bin   5 (  48.10    ,  51.56    )   269.0     percentage=  26.90
 bin   6 (  51.56    ,  55.02    )   214.0     percentage=  21.40
 bin   7 (  55.02    ,  58.49    )   110.0     percentage=  11.00
 bin   8 (  58.49    ,  61.95    )   43.00     percentage=  4.300
 bin   9 (  61.95    ,  65.41    )   7.000     percentage=  .7000
 Wrote eg plot file q.eg
 Use these commands to make a postscript
 plot file with axis and labels:
 pltax q.eg p.eg x Number Histogram
 eg2ps p.eg p.ps
```

10 Computing the erf(x) function and the Normal Distribution Function

There are several methods to do this. One way is to do numerical integration. Romberg integration will compute the erf(x) function accurately, if not efficiently. Here is a program containing subroutines for doing this.

```
C:\je\ftn>type erftest.ftn
c erftest.ftn
      implicit real*8 (a-h,o-z)
      write(*,'(a,g22.14)')' erf(1) = .84270079294971 '
      write(*,'(a,g22.14)')' erf(2) = .99532226501895 '
      x1=-6.
      x2=6.
      n=11
      do i=1,n
       x=(i-1)*(x2-x1)/(n-1) + x1
       v=erf(x)
       write(*,'(a,g22.14,a,g22.14)')' x= ',x,' erf(x)= ',v
      enddo
      end
c+ erf  compute a value of the erf function (error function)
      function erf(x)
```

```
      implicit real*8 (a-h,o-z)
c parameters
c Input:
c    x value in the domain of the erf function
c Output
c     Returns the computed value of the error function
c     if x >= 0,
c     erf(x)= (2/pi) int_0^\infty \exp(-u^2) du
c     if x < 0 , erf(x) = -erf(|x|)
c     erf(0)= 0, erf(\infty) = 1
c     if x < 0, erf(x) is defined as -erf(|x|)
      external erfdnf
      zero=0.
      xmax=5.6
      a=0.
      b=abs(x)
      if((b .gt. zero ) .and. (b .lt. xmax))then
       rel=1.0e-12
       ab =1.0e-12
       call rmberg(erfdnf,a,b,rel,ab,v,ier)
      else
       if(b .eq. zero)v=0.
       if(b .ge. xmax)v=1.
      endif
      erf=v
      if(x .lt. zero)erf=-v
      end


c+ erfdnf the density function defining the erf function (error)
      function erfdnf(x)
       implicit real*8 (a-h,o-z)
c      pi=3.14159265358979d0
       sqrtpi=1.77245385090552d0
       f=exp(-x*x)
       f=2.*f/sqrtpi
       erfdnf=f
       return
      end


c+ rmberg romberg integration
      subroutine rmberg(f,a,b,rel,ab,s,ier)
      implicit real*8 (a-h,o-z)
c     beautified 5/15/96
c parameters
c  f-external function to be integrated: f(x)
c a,b-integration interval
c rel-relative convergence condition: convergence if
c     abs((s(i)-s(i-1))/s(i)) .lt. rel
c  ab-absolute convergence condition: convergence if
c     abs((s(i)-s(i-1)) .lt. ab
c   s-calculated value of integral
c ier-return parameter: ier=0 normal, ier=1 no convergence.
      external f
      dimension tbl(15,15)
      zero=0.
```

```
      ier=0
      n=15
      do i=1,n
       m=2**(i-1)+1
       do j=1,i
        if(j .eq. 1)then
         call trapez(f,a,b,m,tbl(i,1))
        endif
        if(j.ne.1)then
         d=(tbl(i,j-1)-tbl(i-1,j-1))/(4.**(j-1)-1)
        endif
        if(j.ne.1)then
         tbl(i,j)=tbl(i,j-1)+d
        endif
        s=tbl(i,j)
        if((j .ne. 1) .and. (i .ge. 4))then
         if(abs(d).lt. ab)then
          return
         endif
         re=rel
         if(s .ne. zero)then
          re=d/s
         endif
         if(abs(re) .lt. rel)then
          return
         endif
        endif
       enddo
      enddo
      ier=1
      return
      end

c+ trapez   integration by the trapezoid rule
      subroutine trapez(f,a,b,n,v)
       implicit real*8 (a-h,o-z)
c      beautified 5/15/96
c  parameters
c    f-external function to be integrated
c  a,b-integration interval
c    n-interval divided into n-1 pieces
c    v-value returned for integral
      v=0.
      i=1
      do while ( i .le. n)
       x=(i-1)*(b-a)/(n-1)+a
       y=f(x)
       if(i .eq. 1 .or. i .eq. n)then
        y=y/2
       endif
       v=v+y
       i=i+1
      enddo
      h=(b-a)/(n-1)
      v=v*h
      return
      end
```

11 The Inverse of the Normal Distribution Function

If

$$y = \frac{1}{2}(1 + erf(x/\sqrt{2}))$$

then

$$erf(x/\sqrt{2}) = 2y - 1,$$

so

$$x = \sqrt{2}(erf)^{-1}(2y - 1)$$

So the inverse of the normal distribution $F(x)$ is

$$F^{-1}(y) = \sqrt{2}(erf)^{-1}(2y - 1)$$

We can compute the inverse of the normal distribution function numerically. We can use the bisection method, or we could use Newton's method.

12 Sample Mean and Variance, Program *meansdev.c*

See also the section below for finding a normal distribution and plotting it from a data sample using program *meansdev.c*.

The sample mean of a set of n random values is

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The sample variance is

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

$$= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\mu + \mu^2)\right)$$

$$= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{2\mu}{n}\sum_{i=1}^{n}x_i + \mu^2\right)$$

$$= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\mu^2 + \mu^2\right)$$

$$= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2 - \mu^2\right)$$

$$= \frac{n}{n-1}\left(V^2 - \mu^2\right),$$

where

$$V^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2.$$

The divisor $n-1$ is used rather than n, because this makes σ^2 an unbiased estimator of the variance of the random variable X from which the n samples are selected.

If V_n^2 is a value for a set of n samples, then we can compute V_{n+1}^2 from V_n^2. We have

$$V_{n+1}^2 = \frac{1}{n+1}\sum_{i=1}^{n+1}x_i^2$$

$$= \frac{n}{n+1}\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2 + \frac{x_{n+1}^2}{n}\right)$$

$$= \frac{n}{n+1}\left(V_n^2 + \frac{x_{n+1}^2}{n}\right)$$

$$= \frac{n}{n+1}V_n^2 + \frac{x_{n+1}^2}{n+1}.$$

Similarly,

$$\mu_{n+1} = \frac{n}{n+1}\mu_n + \frac{x_{n+1}}{n+1}.$$

Computing in this way may reduce roundoff error when the sample size n is large and the sums get extremely large.

Here is a C program for computing the mean and standard deviation of data contained in a file.

```
//meansdev.c mean and standard deviation of a set of points.
#include <stdio.h>
#include <math.h>
#include <string.h>
main (int argc,char** argv){
 FILE *in;
 char s[255];
 double x;
 double a[200];
 double min;
 double max;
 double mean;
 double meanss;
 double var;
 double sdev;
 double meanp;
 double meanssp;
 double ss;
 int n=0;
 int i;
 if(argc < 2){
  printf("meansdev.c, James Emery, Version 2/19/2009.\n");
  printf("Computes the mean and  and standard deviation of a set of numbers,\n");
  printf("and the number range. See probabilitytheory.pdf by James Emery.\n");
  printf("The data file contains numbers, one number per line. \n");
  printf("Usage: meansdev datafile\n");
  return(1);
 }
 in=fopen(argv[1],"r");
 while(fgets(s,200,in) != NULL){

  x=atof(s);
  a[n]=x;
  n++;
  if(n == 1){
   mean=x;
   meanss=x*x;
   ss=x*x;
   min=x;
   max=x;
  }
  else{
   mean=((n-1)*meanp/(n)) + x/(n);
   meanss =((n-1)*meanssp/(n)) + x*x/(n);
   ss=ss+x*x;
   if(x < min){
    min=x;
   }
   if(x > max){
    max=x;
   }
  }
  //printf(" n= %d x= %15.10g mean= %15.10g meanss= %15.10g \n",n,x,mean,meanss);
  //printf("  ss/n= %15.10g \n",ss/n);
  meanp=mean;
  meanssp=meanss;
 }
```

```
 var=n*(meanss -mean*mean)/(n-1);
 sdev=sqrt(var);
 printf(" number of points= %d \n",n);
 printf(" mean= %15.10g  \n",mean);
 printf(" sdev= %15.10g \n",sdev);
 printf("  min= %15.10g \n",min);
 printf("  max= %15.10g \n",max);

 /*
 var=0.;
 for(i=0;i<n;i++){
  printf(" i=%d x=%15.10g \n",i,a[i]);
  var=var+(a[i]-mean)*(a[i]-mean);
 }
 var=var/(n-1);
 sdev=sqrt(var);
 printf(" n= %d  sdev= %15.10g \n",n,sdev);
 */
 return(0);
}
```

13 Calculating Normal Distribution Probabilities

Suppose a random variable X has a normal distribution with mean μ and variance σ^2. Let us calculate the probability that $x_1 < X < x_2$. We know that the random variable

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

has the standard normal distribution with mean zero and variance one. The standard normal distribution function $F(z)$ is tabulated in tables. Recall that $F(z)$ is the probability that Z is in the set $-\infty < Z < z$. So the probability P of the event $x_1 < X < x_2$, is the probability of

$$\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_1 - \mu}{\sigma}$$

or of

$$z_1 < Z < z_2$$

where

$$z_1 = \frac{x_1 - \mu}{\sigma}$$

and

$$z_2 = \frac{x_2 - \mu}{\sigma}.$$

So for example let us calculate the probability P, that X lies between $x_1 = \mu - \sigma$ and $x_2 = \mu + \sigma$. Computing we find

$$z_1 = \frac{x_1 - \mu}{\sigma} = -1$$

and

$$z_2 = \frac{x_2 - \mu}{\sigma} = 1$$

So the probability of this event is

$$P = F(1) - F(-1)$$

By the symmetry of the normal distribution we have for $z < 0$

$$F(z) = 1 - F(-z)$$

Hence

$$F(-1) = 1 - F(1).$$

So our probability is

$$P = F(1) - F(-1) = F(1) - (1 - F(1)) = 2F(1) - 1$$

From a table of the standard normal distribution function we find that

$$F(1) = .8413$$

Thus

$$P = 2(.8413) - 1 = .68260$$

14 The Moment Generating Function

The moment generating function is defined as the expectation of the exponential function,

$$M(t) = E[\exp(tx)].$$

For a continuous distribution with pdf (probability density function) $f(x)$, we have

$$M(t) = \int_{-\infty}^{\infty} \exp(tx) f(x) dx.$$

This is related to the Laplace Transform of $f(x)$, which is usually written as

$$L(s) = \int_0^\infty exp(-sx)f(x)dx,$$

where in full generality s is a complex variable. There is also a double sided Laplace Transform, with the integral lower limit being $-\infty$ (see Van Der Pol and Bremmer). For an advanced treatment of the Laplace Transform, see Widder. The mean and the variance of a distribution can be obtained from values of the derivative of $M(t)$. That is for example the mean is

$$\mu = M'(0).$$

and the variance is

$$\sigma^2 = M''(0) - \mu^2.$$

Example 1. Consider the pdf of the gamma distribution

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-x/\beta),$$

where Γ is the gamma function, which is defined by

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy,$$

and that for an integer n we have

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

After a change of variable we find that (Hogg and Craig, p93)

$$M(t) = \frac{1}{(1-\beta t)^\alpha} \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1} \exp(-y) dy$$

$$= \frac{1}{(1-\beta t)^\alpha}.$$

We find that

$$\mu = \alpha\beta$$

and

$$\sigma^2 = \alpha\beta^2.$$

Example 2. Consider now the normal distribution with pdf

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

The moment generating function is

$$M(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(xt) \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(xt + \frac{-(x-\mu)^2}{2\sigma^2}\right) dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2 - 2x(\sigma^2 t + \mu) + \mu^2}{2\sigma^2}\right) dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-(\sigma^2 t + \mu))^2 - (\sigma^2 t + \mu)^2 + \mu^2}{2\sigma^2}\right) dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-(\sigma^2 t + \mu))^2 - \sigma^4 t^2 + 2\sigma^2 t\mu}{2\sigma^2}\right) dx$$

$$= \exp(\sigma^2 t^2/2 + t\mu) \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-(\sigma^2 t + \mu))^2}{2\sigma^2}\right) dx$$

Letting

$$y = \frac{x - (\sigma^2 t + \mu)}{\sigma},$$

and

$$dx = \sigma dy,$$

this becomes

$$M(t) == \exp(\sigma^2 t^2/2 + t\mu) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy$$

$$= \exp(\sigma^2 t^2/2 + t\mu).$$

Then

$$M'(t) = \exp(\sigma^2 t^2/2 + t\mu)(\sigma^2 t + \mu)$$

and

$$M''(t) = \exp(\sigma^2 t^2/2 + t\mu)(\sigma^2 t + \mu)^2 + \exp(\sigma^2 t^2/2 + t\mu)\sigma^2.$$

Thus the mean is

$$M'(0) = \mu,$$

and the variance is

$$M''(0) - \mu^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2.$$

There is a problem with the moment generating function, it does not exist for all distributions. The characteristic function given in the next section plays a role similar to the moment generating function, and is more general.

15 The Characteristic Function

The characteristic function of a distribution is the expectation of e^{itx}. It is related to the Fourier Transform just as the Moment Generating function is related to the Laplace Transform.

The Fourier transform of the function f is defined as (Goldberg, **The Fourier Transform**)

$$g(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt.$$

By the Fourier integral theorem

$$f(t) = \int_{-\infty}^{\infty} g(\omega)e^{i\omega t}d\omega.$$

Example . Let

$$f(t) = \left\{ \begin{array}{cc} 2e^{-3t} & t \geq 0 \\ 0 & t < 0 \end{array} \right\}$$

Then

$$g(\omega) = \frac{2}{3 + i\omega}$$

The characteristic function is defined by

$$\phi(t) = E[e^{itx}]$$

$$= \int_{-\infty}^{\infty} e^{itx} f(x)dx.$$

$$\phi'(t)) = \int_{-\infty}^{\infty} ixe^{itx} f(x)dx.$$

So we can compute the mean and variance from first and second derivatives of the characteristic function.

$$E[X] = -i\phi'(0)$$

and

$$E[X^2] = -\phi''(0)$$

Example. Given a normal distribution with variance σ^2 and mean 0,

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-u^2/(2\sigma^2))du.$$

The characteristic function is

$$\phi(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} \exp(-x^2/(2\sigma^2))dx$$
$$= \exp(-\sigma^2 t^2/2))$$

(Lamperti page 60). Hence

$$\phi'(t) = -\sigma^2 t \exp(-\sigma^2 t^2/2))$$
$$\phi''(t) = -\sigma^2 \exp(-\sigma^2 t^2/2)) + \sigma^4 t^2 \exp(-\sigma^2 t^2/2))$$

Hence

$$E[X] = -i\phi'(0) = 0$$

and

$$E[X^2] = -\phi''(0) = \sigma^2.$$

The characterize function may be used to prove central limit theorems. See Lamperti.

16 The Central Limit Theorem

If each random variable X_i has mean μ and standard deviation σ, and

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

then the distribution of random variable

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

approaches the standard normal distribution as $n \to \infty$.

17 The Generation of a Normally Distributed Random Variable

Suppose X has the uniform distribution on the interval $[0, 1]$. The mean is

$$\mu = \int_0^1 x dx = \frac{1}{2}.$$

The variance is

$$\sigma^2 = \int_0^1 (x-\mu)^2 dx$$

$$= \int_0^1 (x^2 - 2x\mu + \mu^2) dx$$

$$= \left[\frac{x^3}{3} - x^2\mu + \mu^2 x \right]_0^1$$

$$= \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}.$$

Hence by the Central Limit theorem

$$Z_n = (\bar{X} - 1/2)\sqrt{12n},$$

has an approximately normal distribution.
Proposition If Z has a standard normal distribution, then

$$X = \mu + \sigma Z,$$

has a normal distribution with mean μ and variance σ.
Proof. The distribution function of X is

$$F(a) = Pr(\mu + \sigma Z \leq a)$$

$$= Pr(\sigma Z \leq a - \mu)$$

$$= Pr\left(Z \leq \frac{a-\mu}{\sigma}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(a-\mu)/\sigma} \exp(-t^2/2) dt.$$

Let

$$t = \frac{x-\mu}{\sigma},$$

$$dt = \frac{dx}{\sigma}.$$

Then the integral becomes

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{a} \exp(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2)dx.$$

18 The Inverse Function Method of Generating A Random Variate

Suppose a random variable X has distribution function F, that is

$$F(a) = Pr(X \leq a),$$

where $F(-\infty) = 0$, and $F(\infty) = 1$. In general F is a monotone increasing function, so that it has an inverse F^{-1}. Let Y be a uniformly distributed random variable on the interval $[0, 1]$. Let

$$Z = F^{-1}(Y).$$

Then let G be the distribution function of X. We have

$$G(a) = Pr(Z \leq a)$$

$$= Pr(F^{-1}(Y) \leq a)$$

$$= Pr(FF^{-1}(Y) \leq F(a))$$

$$Pr(Y \leq F(a))$$

$$= F(a).$$

The last equality follows because Y has the uniform distribution. We have shown that Z has the distribution function F.

19 The Polar Method of Generating A Normal Random Sample

One may compute

$$\int_{-\infty}^{\infty} e^{-t^2/2} dt = \sqrt{2\pi},$$

by computing the two dimensional product of two such intervals after a change to polar coordinates. This can be exploited to compute the inverse functions of two variables and then using two uniformly distributed random variables to generate two normally distributed random variables. See Knuth, **The Art of Computer Programming**, V2, p105. Here is a resulting program:

```
c+ emgaus  normal random sample
      function emgaus(iseed,amean,stddev)
      dimension v(2)
c
c  parameters:
c  iseed-seed for the random number generator jerand.
c        this is an integer between 1 and 2147483647.
c        a random seed can be set by calling jerand with
c        ns=1 (see the remarks in jerand).
c  amean-mean of the normal distribution to be sampled.
c  stddev-standard deviation of the normal distribution to
c          be sampled.
c
c     Reference: D. E. Knuth, The Art of Computer Programming,
c     volume 2, page 104. This is the polar method for
c     generating a normal sample.
c
10    call jerand(iseed,2,0,v)
      v(1)=2*v(1)-1.
      v(2)=2.*v(2)-1.
      s=v(1)*v(1)+v(2)*v(2)
      if(s.ge.1)go to 10
      x=v(1)*sqrt(-2.*alog(s)/s)
      emgaus=amean+stddev*x
      return
      end
```

20 Generating A Uniform Random Sample

See **Numerical Recipes** Chapter seven.

```
c randnum.ftn test of random number generator
c computes the number of times the random number
c falls in each of 100 bins
        implicit real*8(a-h,o-z)
        dimension m(100)
        do 5 i=1,100
         m(i)=0
5       continue
```

```
        jran=1
       do 10 i=1,121501
        call randj(jran,r)
        k=r*100 + 1
        m(k)=m(k)+1
        if(jran .eq. 1)then
         write(*,*)' jran =1 at ', i
        endif
10     continue
       do 20 i=1,100
        write(*,*)i,m(i)
20     continue
       end
c+ randj  simple congruential random number generator
        subroutine randj(jran,r)
        implicit real*8(a-h,o-z)
c  parameters
c   jran=seed on input, next random integer on output
c         1 <= jran <121500
c   r=real random number between 0. and 1.
c  (see table in book 'numerical recipes')
c  (period is 121500, i.e. repeats after 121500 calls)
c   works for 32 bit integers
      data im,ia,ic /121500,2041,25673/
      a=im
      jran=jran*ia+ic
      jran=mod(jran,im)
c     r=mod(jran*ia+ic,im)/(real(im)
      r=jran/a
      return
      end
```

21 Stochastic Processes

22 The Poisson Process, the Poisson Distribution, and the Exponential Distribution

A Poisson Process is a stochastic process which for example would model the occurance of lightning strikes or radioactive emission of particles. Here we derive probability distributions for such a process.

We want to calculate the probability of n points falling in an interval of length τ on the real line, where the average number of points per unit interval is λ. This probability distribution is called the discrete poisson distribution defined on the set $\{0, 1, 2, 3, 4, ...\}$.

To do this calculation we start with a finite line of length t in place of the real line, and then let t go to infinity. So let there be a long interval on the real line of length t. Let there be a short subinterval of length τ. N points are placed randomly in the long interval. The probability of a single point successfully falling in the short interval is the ratio of the two segments $p = \tau/t$. The probability of n successes in N Bernoulli trials is

$$C_n^N p^n (1-p)^{N-n},$$

where C_n^N is the number of combinations of N things taken n at a time

$$C_n^N = \frac{N!}{n!(N-n!} = \frac{N}{1}\frac{N-1}{2}\frac{N-2}{3}...\frac{N-n+1}{n}.$$

Let the average number of points found in a unit length be

$$\lambda = \frac{N}{t}.$$

Now we shall keep λ, the average number of points per unit interval constant, while letting N and t go to infinity. So we have

$$C_n^N (\frac{\tau}{t})^n (1-\frac{\tau}{t})^{N-n} = C_n^N (\frac{\tau\lambda}{N})^n (1-\frac{\tau\lambda}{N})^{N-n}$$

$$= N(N-1)(N-2)...(N-n+1)\frac{(\tau\lambda)^n}{N^n n!}(1-\frac{\tau\lambda}{N})^{N-n}$$

$$= (1-\frac{1}{N})(1-\frac{2}{N})...(1-\frac{(n-1)}{N})\frac{(\tau\lambda)^n}{n!}(1-\frac{\tau\lambda}{N})^{N-n}$$

$$= (1-\frac{1}{N})(1-\frac{2}{N})...(1-\frac{(n-1)}{N})\frac{(\tau\lambda)^n}{n!}\frac{(1-\frac{\tau\lambda}{N})^N}{(1-\frac{\tau\lambda}{N})^n}.$$

Then as N goes to infinity this becomes

$$\frac{(\tau\lambda)^n}{n!}e^{-\lambda\tau}.$$

So this is the probability of n points falling in an interval of length τ where the average number of points per unit length is λ. This is the discrete poisson distribution defined on the set $\{0, 1, 2, 3, 4, ...\}$.

Notice that

$$\sum_{n=0}^{\infty}\frac{(\tau\lambda)^n}{n!}e^{-\lambda\tau} = e^{-\lambda\tau}\sum_{n=0}^{\infty}\frac{(\tau\lambda)^n}{n!} = e^{-\lambda\tau}e^{\lambda\tau} = 1.$$

Let a random variable X be the distance between succeeding points. Then the probability that X is greater than τ is the probability that there are no points in the interval of length τ. That is

$$\Pr(X > \tau) = \frac{(\tau\lambda)^n}{n!} e^{-\lambda\tau},$$

where $n = 0$. So

$$\Pr(X > \tau) = e^{-\lambda\tau}.$$

Then the distribution function of X is

$$F(\tau) = \Pr(X \le \tau) = 1 - e^{-\lambda\tau}.$$

X is said to have the exponential distribution. The pdf is the derivative

$$\frac{dF}{d\tau} = \lambda e^{-\lambda\tau}.$$

Let U be the uniform distribution in the interval $[0, 1]$. Let

$$u = 1 - e^{-\lambda\tau},$$

Then

$$\begin{aligned} e^{-\lambda\tau} &= 1 - u \\ -\lambda\tau &= \ln(1-u) \\ \tau &= -\frac{\ln(1-u)}{\lambda} \end{aligned}$$

Let

$$X = -\frac{\ln(1-u)}{\lambda}$$

Then

$$\begin{aligned} \Pr(X \le \tau) &= \Pr(\{u : X(u) \le \tau\}) \\ &= \Pr(u \le 1 - e^{-\lambda\tau}) \\ &= 1 - e^{-\lambda\tau} \\ &= F(\tau) \end{aligned}$$

That is X has the exponential distribution.

Certain conditions must be met for a process to have the Poisson distribution. For example, if certain events occurred in a five minute interval, one interval per hour, then the process would not have a Poisson distribution because the average number of events would not be uniform throughout the whole hour or day or whatever. Hogg and Craig **Introduction to Mathematical Statistics**, second edition, page 87, gives postulates for a Poisson process, as we have described above.

23 Markov Chains

A Markov Process is like a finite state machine, which has a finite number of states, but where the transition to a next state depends on the current state and a transition probability matrix instead of on the current state and some set of inputs. A probability vector for the probabilities for being in any of the k states, after n steps in the chain, is given by an nth power of the of the transision matrix.

For an elementary introduction to Markov chains see the freshman level book **Introduction to Finite Mathematics** by John J Kemeny, J. Laurie Snell, and Gerald L. Thompson, Prentice-Hall, 1957.

At a higher level see **A First Course in Probability** Sheldon Ross, 3rd edition, 1988.

24 The Gamma Distribution

The gamma distribution is the probability model for waiting times and is related to the Poisson process. The exponential distribution is a special case of the gamma distribution. Recall from a previous section that the exponential random variable is the time between events of a poisson process. See pages 91-94 of Hogg and Craig.

We derive the gamma distribution from the Poisson distribution following Parsen p261. The waiting time for the rth event in a series of events having the Poisson probability function at the rate of λ events per unit time has p.d.f.

$$f_{\lambda,r}(t) = \frac{\lambda}{(r-1)!}(\lambda t)^{r-1}e^{-\lambda t}, t \geq 0.$$

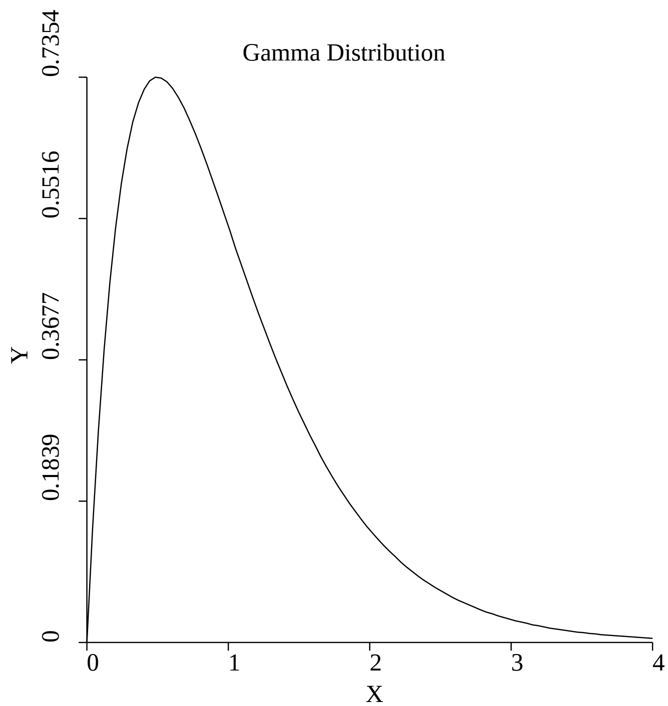

Figure 5: The Gamma Distribution with mean $\mu = 1$ and variance $\sigma^2 = 1/2$. The Gamma parameters are $r = 2$, and $\lambda = 2$.

To prove this, let T_r be the random variable giving the time of the rth event. Let $F_r(t)$ be the distribution function for T_r. Then $F_r(t)$ is the probability that the time T_r of the rth occurrence will be less than or equal to t. Then $1 - F_r(t)$ is the probability that there is either 1 occurrence in time t, or 2 occurrences in time t, .., or $r-1$ occurrences in time t. This is a sum of Poisson probabilities

$$1 - F_r(t) = \sum_{k=0}^{r-1} \frac{1}{k!}(\lambda t)^k e^{-\lambda t}.$$

Then

$$F_r(t) = 1 - \sum_{k=0}^{r-1} \frac{1}{k!}(\lambda t)^k e^{-\lambda t}$$
$$= 1 - e^{-\lambda t} \sum_{k=0}^{r-1} \frac{1}{k!}(\lambda t)^k$$

We take the derivative to get the p.d.f. for T_r. We have

$$f_{\lambda,r}(t) = \frac{dF_r(t)}{dt} = \lambda e^{-\lambda t} \sum_{k=0}^{r-1} \frac{1}{k!}(\lambda t)^k - e^{-\lambda t}\mu \sum_{k=1}^{r-1} \frac{1}{(k-1)!}(\lambda t)^{k-1}$$
$$= \frac{\lambda}{(r-1)!}(\lambda t)^{r-1} e^{-\lambda t}.$$

Recall that the gamma function is

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy,$$

and that for an integer n we have

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

Thus letting $y = \lambda t$ we have

$$\int_0^\infty f_{\lambda,r}(t)dt = \frac{\lambda}{(r-1)!} \int_0^\infty (\lambda t)^{r-1} e^{-\lambda t} dt$$
$$= \frac{1}{(r-1)!} \int_0^\infty y^{r-1} e^{-y} dy$$

$$= \frac{\Gamma(r)}{(r-1)!} = 1.$$

So far we have treated the gamma distribution when r is an integer. We can generalize it. Recall that the gamma function is given by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

We see that

$$\Gamma(1) = 1,$$

and if n is an integer then

$$\Gamma(n+1) = n\Gamma(n).$$

We have

$$\Gamma(n) = (n-1)!, \Gamma(0) = \infty.$$

We have

$$\Gamma(1/2) = \sqrt{\pi}, \Gamma(m + 1/2) = \frac{1 \cdot 3 \cdot ... \cdot (2m-1)}{2^m}\sqrt{\pi}.$$

The formula

$$\Gamma(n+1) = n\Gamma(n),$$

This allows us to extend the definition to negative numbers.

If we write the gamma p.d.f, following Parsen on page 261, where r is the degrees of freedom, and λ is the events per unit distance, we write

$$f_{r,\lambda}(x) = \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1} e^{-\lambda x}.$$

Hogg and Craig uses a bit different notation with $\lambda = 1/\beta$ and $r = \alpha$. On page 93 of Hogg and Craig it is shown using moment generating functions that the mean of a gamma random variable X, after changing to is Parzen notation is

$$E(X) = \frac{r}{\lambda}(\text{Parzen notation}).$$

Likewise the variance is

$$\sigma^2 = \frac{r}{\lambda^2}.$$

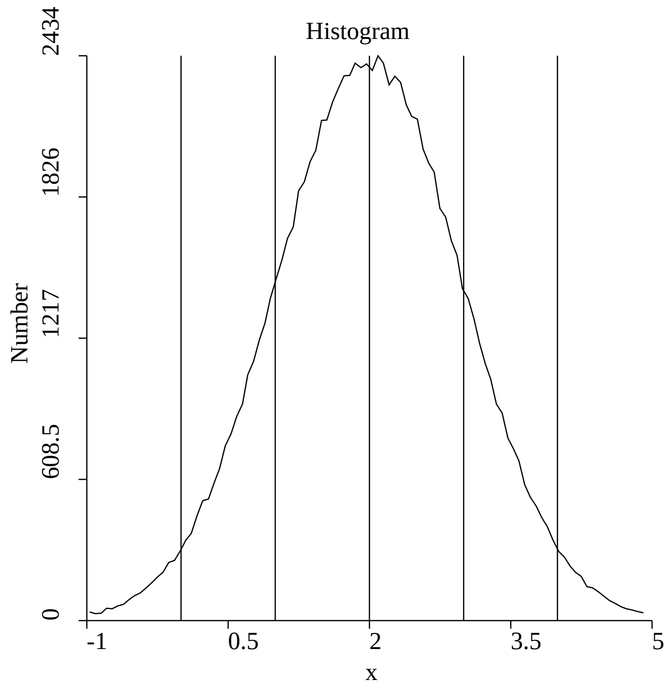

Figure 6: A test of normal random variate generation. This test was done with **normdist.ftn**, which is a Fortran program that generates 100,000 points from a normal distribution with mean $\mu = 2$ and standard deviation $\sigma = 1$. This plot is a histogram with 100 bins in the interval of length $6\sigma = 6$.

25 Test of Normal Random Variate Generation

We test a normal random variate generator using the Fortran program **normdist.ftn**. Here is a listing of the program:

```
c normdist.ftn normal random samples test
c 10/11/96
      implicit real*8(a-h,o-z)
      parameter (np=100000)
      dimension x(np)
      dimension xl(100)
      dimension v(100)
```

```
      zero=0.
      sigma=1.
      amean=2.
      n=np
      iran=6789
      do i=1,n
       call nsamp(iran,amean,sigma,r)
       x(i)=r
c      write(*,'(i5,g15.8)')i,v
      enddo
      call meansdv(x,n,am,sdv)
      write(*,'(a,g15.8)')'Sample mean = ',am
      write(*,'(a,g15.8)')'Sample standard deviation = ',sdv
      xmn=amean-3.*sigma
      xmx=amean+3.*sigma
      nl=100
      call hstgrm(x,n,xmn,xmx,nl,xl,v,vm)
      open(1,file='q.gi',status='unknown')
      write(1,'(a)')'v-1 1 -1 1'
      write(1,'(a,4(g15.8,1x))')'w',xmn,xmx,zero,vm
      do i=1,(nl-1)
       xm=(i-1)*(xmx-xmn)/nl+xmn+(xmx-xmn)/(2.*nl)
       if(i .eq. 1)then
        write(1,'(a,2(g15.8,1x))')'m',xm,v(i)
       else
        write(1,'(a,2(g15.8,1x))')'d',xm,v(i)
       endif
      enddo
c     draw sigma lines
      write(1,'(a,2(g15.8,1x))')'m',(am-2.*sigma),zero
      write(1,'(a,2(g15.8,1x))')'d',(am-2.*sigma),vm

      write(1,'(a,2(g15.8,1x))')'m',(am-1.*sigma),zero
      write(1,'(a,2(g15.8,1x))')'d',(am-1.*sigma),vm

      write(1,'(a,2(g15.8,1x))')'m',(am-0.*sigma),zero
      write(1,'(a,2(g15.8,1x))')'d',(am-0.*sigma),vm

      write(1,'(a,2(g15.8,1x))')'m',(am+1.*sigma),zero
      write(1,'(a,2(g15.8,1x))')'d',(am+1.*sigma),vm

      write(1,'(a,2(g15.8,1x))')'m',(am+2.*sigma),zero
      write(1,'(a,2(g15.8,1x))')'d',(am+2.*sigma),vm

      write(*,*)' Use the next command, and one of the '
      write(*,*)' next commands to make a plot: '
      write(*,'(a)')' pltax q.gi p.gi x Number Histogram'
      write(*,'(a)')' pltgpr'
      write(*,'(a)')' pltvga p.gi'
      write(*,'(a)')' pltgl p.gi'
      write(*,'(a)')' eg2ps p.gi p.ps'
      end
c+ nsamp  normal random sample
      subroutine nsamp(iran,amean,sigma,x)
      implicit real*8(a-h,o-z)
      dimension v(2)
c  Input:
```

```
c  iran   seed on first input, next random integer on output
c          1 <= jran < 121500
c  amean  mean of the normal distribution to be sampled.
c  sigma  standard deviation of the normal distribution to
c         be sampled, that is, sigma=sqrt(variance).
c  Output:
c  x      normal random variable
c     Reference: D.E. Knuth, the art of computer programming,
c     volume 2, page 104. This is the polar method for
c     generating a normal sample.
c
      one=1.
      s=2.
      do while(s .ge. one)
       call randj(iran,r)
       v(1)=2*r-1.
       call randj(iran,r)
       v(2)=2.*r-1.
       s=v(1)*v(1)+v(2)*v(2)
      enddo
      x=v(1)*sqrt(-2.* log(s)/s)
      x=amean + sigma*x
      return
      end
c+ hstgrm  histogram of a sample.
      subroutine hstgrm(x,n,xmn,xmx,nl,xl,v,vm)
c 10/11/96 modification of old subroutine
        implicit real*8(a-h,o-z)
c Input:
c x       sample vector.
c n       number of points in the sample.
c xmn,xmx interval definition.
c nl      number of levels dividing the interval.
c         (number of bins is nl-1)
c Output:
c xl      vector of nl levels.
c v       vector of length nl-1, the number of sample points in each bin.
c vm      the maximum value found in v.
c
      dimension x(*),xl(*),v(*)
      b=(xmx-xmn)/(nl-1)
      do i=1,nl
       xl(i)=(i-1)*b+xmn
       v(i)=0.
      enddo
      vm=0.
      do i=1,n
       j=(x(i)-xmn)/b+1.5
       if((j .ge. 1).and.(j .le. nl))then
        v(j)=v(j)+1
        if(v(j) .gt. vm)then
         vm=v(j)
        endif
       endif
      enddo
      return
      end
```

```
c+ randj  congruential random number generator
        subroutine randj(jran,r)
        implicit real*8(a-h,o-z)
c  parameters
c   jran=seed on input, next random integer on output
c          1 <= jran < 121500
c   r=real random number between 0. and 1.
c  (see table in book 'numerical recipes')
c  (period is 121500, i.e. repeats after 121500 calls)
c   works for 32 bit integers
      data im,ia,ic /121500,2041,25673/
      a=im
      jran=jran*ia+ic
      jran=mod(jran,im)
c     r=mod(jran*ia+ic,im)/(real(im)
      r=jran/a
      return
      end
c+ meansdv  mean and standard deviation of array.
      subroutine meansdv(x,n,amean,sdv)
      implicit real*8(a-h,o-z)
c mean and standard deviation of x.
c n, number of values in x.
      dimension x(*)
      amean=0.
      do i=1,n
       amean=amean+x(i)
      enddo
      amean=amean/n
      var=0.
      do i=1,n
       var=var+(x(i)-amean)**2
      enddo
      var=var/float(n-1)
      sdv=sqrt(var)
      return
      end
```

26 Determining a Normal Distribution by Sampling, Using Program *meansdev.c*

The program reads a file of points, computes μ and σ for a normal distribution and generates an eg plot of the the normal curve. To get the final postscript file the programs **pltax.c** and **eg2ps.c** are employed. The data for the plot in the figure is:

```
2.513
2.505
2.497
2.514
```

```
2.498
2.503
2.4789
2.537
2.497
2.513
```

To see some information on running the program, type the program name with no parameters:

```
meansdev.c, James Emery, Version 12/31/2009.
Computes the mean and  and standard deviation of a set of numbers,
and the number range. See probabilitytheory.pdf by James Emery.
The data file contains numbers, one number per line.
The program also generates an eg plot file called p.eg.
Add labeled axes with pltax.c, and convert to Postscript with eg2ps.c.
Usage: meansdev datafile
```

The output of the program is:

```
number of points= 10
 mean=         2.50559
 sdev=    0.0152451413
  min=          2.4789
  max=           2.537
 i=0 x=          2.513 dev=        0.00741
 i=1 x=          2.505 dev=       -0.00059
 i=2 x=          2.497 dev=       -0.00859
 i=3 x=          2.514 dev=        0.00841
 i=4 x=          2.498 dev=       -0.00759
 i=5 x=          2.503 dev=       -0.00259
 i=6 x=         2.4789 dev=       -0.02669
 i=7 x=          2.537 dev=        0.03141
 i=8 x=          2.497 dev=       -0.00859
 i=9 x=          2.513 dev=        0.00741
```

```
//meansdev.c mean and standard deviation of a set of points.
#include <stdio.h>
#include <math.h>
#include <string.h>
int  ncrvplt( char* fname, double mu, double sigma, double xmn, double xmx, int n);
int main (int argc,char** argv){
 FILE *in;
 char s[255];
 double x;
 double a[200];
 double min;
 double max;
 double mean;
 double meanss;
 double var;
 double sdev;
 double meanp;
 double meanssp;
 double ss;
```

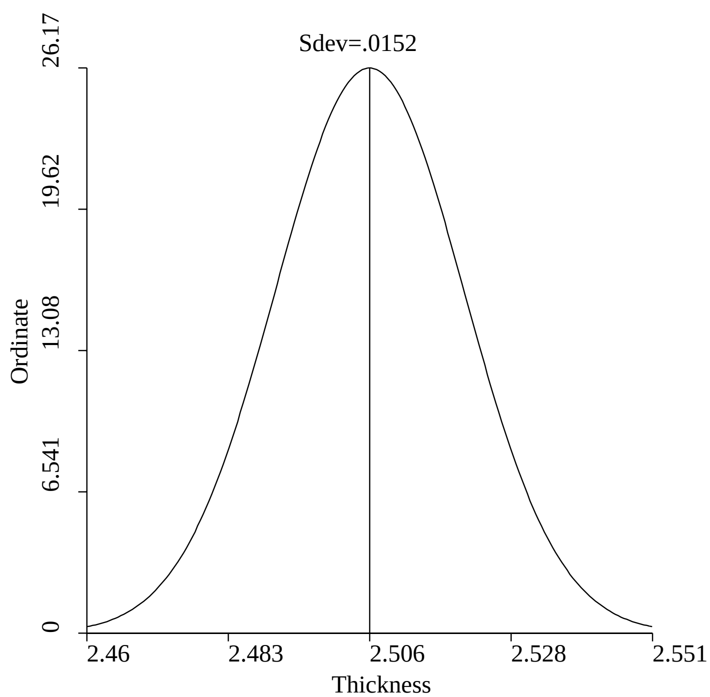

Figure 7: A Normal Distribution Curve computed by program **meansdev.c**

```
double xmn,xmx;
int n=0;
int np;
int i;
if(argc < 2){
 printf("meansdev.c, James Emery, Version 2/19/2009.\n");
 printf("Computes the mean and  and standard deviation of a set of numbers,\n");
 printf("and the number range. See probabilitytheory.pdf by James Emery.\n");
 printf("The data file contains numbers, one number per line. \n");
 printf("Usage: meansdev datafile\n");
 return(1);
}
in=fopen(argv[1],"r");
while(fgets(s,200,in) != NULL){

 x=atof(s);
 a[n]=x;
 n++;
 if(n == 1){
  mean=x;
  meanss=x*x;
  ss=x*x;
  min=x;
  max=x;
 }
 else{
  mean=((n-1)*meanp/(n)) + x/(n);
  meanss =((n-1)*meanssp/(n)) + x*x/(n);
  ss=ss+x*x;
  if(x < min){
   min=x;
  }
  if(x > max){
   max=x;
  }
 }
 //printf(" n= %d x= %15.10g mean= %15.10g meanss= %15.10g \n",n,x,mean,meanss);
 //printf("  ss/n= %15.10g \n",ss/n);
 meanp=mean;
 meanssp=meanss;
}
var=n*(meanss -mean*mean)/(n-1);
sdev=sqrt(var);
printf(" number of points= %d \n",n);
printf(" mean= %15.10g  \n",mean);
printf(" sdev= %15.10g \n",sdev);
printf("  min= %15.10g \n",min);
printf("  max= %15.10g \n",max);

for(i=0;i<n;i++){
 printf(" i=%d x=%15.10g dev=%15.10g \n",i,a[i],a[i]-mean);
}
//var=var/(n-1);
//sdev=sqrt(var);
//printf(" n= %d  sdev= %15.10g \n",n,sdev);

xmn=mean-3.*sdev;
```

```
 xmx=mean+3.*sdev;
 np=200;
 ncrvplt("p.eg",mean,sdev,xmn,xmx,np);
 return(0);
}
//c+ ncrvplt normal curve plot
int  ncrvplt( char* fname, double mu, double sigma, double xmn, double xmx, int n){
 FILE *out;
 double ymn,ymx;
 double x,y;
 double c;
 double pi=3.14159265358979;
 int i;
 out=fopen(fname,"w");
 ymn=0.;
 ymx=0.;
 c = 1./(sigma*sqrt(2.*pi));
 for(i=0;i < n; i++){
  x= i*(xmx-xmn)/(n-1) + xmn;
  y= c*exp(-pow(mu-x,2)/(2.*pow(sigma,2)));
  if(ymx < y)ymx=y;
 }
 fprintf(out,"v -1 1 -1 \n");
 fprintf(out,"w %15.10g %15.10g %15.10g %15.10g\n",xmn,xmx,ymn,ymx);
 for(i=0;i < n; i++){
  x= i*(xmx-xmn)/(n-1) + xmn;
  y= c*exp(-pow(mu-x,2)/(2.*pow(sigma,2)));
  if(i > 0){
   fprintf(out,"d %15.10g %15.10g \n",x,y);
  }
  else{
   fprintf(out,"m %15.10g %15.10g \n",x,y);
  }
 }
 fprintf(out,"m %15.10g %15.10g \n",xmn,ymn);
 fprintf(out,"d %15.10g %15.10g \n",xmx,ymn);
 fprintf(out,"m %15.10g %15.10g \n",mu,ymn);
 fprintf(out,"d %15.10g %15.10g \n",mu,ymx);
 return(0);
}
```

27 Probability in Physics

Assume a 'gas' consisting of only 2 particles, and suppose there are three particle states, labelled 1,2,3. Suppose the particles are distinguishable, namely one is named A and a second is named B. Here is a table of the possible arrangement of the 2 particles in the three states:

1	2	3
AB	...	...
...	AB	...
...	...	AB
A	B	...
B	A	...
A	...	B
B	...	A
...	A	B
...	B	A

This is the Maxwell-Boltzmann case, any number of particles can be in any state and the particles are distinguishable. There are 9 rows in the table. So there are a total of $3^2 = 9$ possible states for the gas.

Now consider the quantum mechanical cases where the particles are not distinguishable. First consider the Bose-Einstein case where any number of particles can be in the same state, because the wave functions symmetric and an interchange of the particles does not change the wave function. Then the two particles are each labelled A because they can not be distinguished. So the table becomes

1	2	3
AA	...	...
...	AA	...
...	...	AA
A	A	...
A	...	A
...	A	A

Now there are six rows in the table and $3 + 3 = 6$ states for the gas.

Next consider the Fermi-Einstein case. The wave functions are antisymmetric, so interchanging a pair of particles in the wave function changes the sign of the wave function. But the particles are not distinguishable so the interchange does not change the sign. The only way both of these things can happen is that the wave function is zero. That is, each state can be occupied by just one particle. This is the Pauli exclusion principle. Hence the table becomes

1	2	3
A	A	...
A	...	A
...	A	A

Now there are just 3 rows and 3 possible states for the gas.

In the case of m particles in n we can work out the three cases and the number of states.

This simple example comes from Reif, **Fundamentals of Statistical and Thermal Physics,** p333.

Now we can work out the expected number of particles in each state. This is the expected value in the sense of a random variable and a probability distribution.

28 Maxwell-Boltzmann Statistics

In statistical mechanics, Maxwell-Boltzmann statistics describes the statistical distribution of material particles over various energy states in thermal equilibrium, when the temperature is high enough and density is low enough to render quantum effects negligible. This is a classical distribution. $\frac{N_i}{N}$ is the proportion of the particles that are in state i.

$$\frac{N_i}{N} = \frac{g_i}{e^{\epsilon_i - \mu/kT}} = \frac{g_i e^{-\epsilon_i/kT}}{Z},$$

where

N_i is the number of particles in state i.

ϵ_i is the energy of the ith state.

g_i is the degeneracy of states.

μ is the chemical potential

κ is Boltzman's constant

T is absolute temperature

N is the total number of particles

Z is the partition function

$$Z = \sum_i g_i e^{-\epsilon_i/kT}.$$

The degeneracy of states g_i is the dimension of the eigenspace for the given energy eigenvalue. That is for a given eigenvalue there may be more than one linearly independent eigenvector, just as in the finite dimensional matrix or linear algebra case. So for example if the $g_i = 3$ then there are actually 3 distinct states for the energy level and so this must be counted.

29 Fermi-Dirac Statistics

Fermions are particles which are indistinguishable and obey the Pauli exclusion principle, i.e., no more than one particle may occupy the same quantum state at the same time. Fermions have half-integral spin. Statistical thermodynamics is used to describe the behavior of large numbers of particles. Number of particles in state i,

$$n_i = \frac{g_i}{e^{\epsilon_i - \mu/kT} + 1}$$

where g_i is the degeneracy of states. g_i is the dimension of the eigenspace for the given energy eigenvalue. μ is the chemical potential as introduced by Willard Gibbs. An example of a chemical potential is the fermi level in a semiconductor. For a single orbital the distribution would be

$$n_i = \frac{1}{e^{\epsilon_i - \mu/kT} + 1}$$

and takes the value between 0 and 1. That is, the orbital is occupied by at most 1 particle by the exclusion principle. Fermions are spin 1/2 particles, with antisymmetric wave functions. See Kittel Thermal Physics, or Margenau and Murphy for an old treatment. This distribution appeared in separate papers by Fermi and Dirac in 1926.

Fermi-Dirac statistics apply to fermions (particles that obey the Pauli exclusion principle.)

30 Bose-Einstein Statistics

Determines the statistical distribution of identical indistinguishable bosons over the energy states in thermal equilibrium. After Satyendra Nath Bose and his statistical theory of photons. Bose-Einstein Statistics applies to Bosons. Bosons, unlike fermions, are not subject to the Pauli exclusion principle: an unlimited number of particles may occupy the same state at the same time. Number of particles in state i,

$$n_i = \frac{g_i}{e^{\epsilon_i - \mu/kT} - 1}$$

n_i can be larger than 1, so for example many photons, which are Bosons, can occupy the same energy state. Bose-Einstein and Fermi-Dirac distributions differ only in the sign placed on the 1 in the denominator. The Bose-Einstein Condensate is a consequence of this distribution. In recent years Bose-Einstein condensates have been constructed by laser cooling.

In the classical limit of low temperatures all three distributions agree.

Feynman in volume III of his lectures has a section on a Boson gas derived from his Feynman diagram type of arguments with his amplitudes and so on.

31 The Random Walk

32 The Monte Carlo Method

See quadric.tex

33 Least Squares and Regression

See **Least Squares Approximation** , lsq.tex.
See **Regression**, regression.tex.

34 The Student's T Distribution

The pdf of the student's T distribution for ν degrees of freedom is

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

This distribution is similar to the normal distribution with mean 0 and variance 1, but the tails of the distribution have more probability, and the region near 0 has less. As the degrees of freedom ν go to infinity, the students's

T distribution converges to the standard normal distribution. Student is a pen name for William Sealy Gosset who published a derivation of it in 1908.

Let $x_1, ..., x_n$ be the numbers observed in a sample from a continuously distributed population with expected value μ. The sample mean and sample variance are respectively

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$$

and

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

The resulting t-value is

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}.$$

The t-distribution with $n - 1$ degrees of freedom is the sampling distribution of the t-value when the samples consist of independent identically distributed observations from a normally distributed population.

[10] Hogg Robert V, Craig Allen T **Introduction to Mathematical Statistics**, Second Edition, Macmillan, 1966. This book gives a derivation for the Student's T Distribution and some of its uses.

35 Appendix A, Related Documents

Statistics by James Emery
  stat.tex
Least Squares Approximation by James Emery
  lsq.tex
Regression
  regression.tex

36 Computer Programs

`meansdev.c`

```
corners.cpp tangents of point curve by least squares
corners.ftn tangents of point curve by least squares
flulsq.ftn least squares applied to fluorescence data
llsq.c linear least squares jan 18 1990, uses malloc
 dynamic memory allocation
llsq.ftn linear least squares
lscir.c least squares circle
lspoly.c least squares polynomial
lsq.ftn least squares
lsq94.ftn update of lsq.ftn least squares program
  outputs eg plot file, later version of lsq.ftn
lsqexp.ftn least squares exponential
lsqgen.ftn general linear least squares program and plot using functions sub.
lsql.c least squares line
lsqln.cm least squares line old comal program
lsqln.ftn least squares line
lsqplane.ftn least squares plane
lsqplt.pas least squares plot
lsqrat.ftn rational least squares
lsqsc.ftn least squares curve
lsqsc.c least squares space circle
lsqpln.c least squares plane
lsqscbp.cpp least squares fitting of circle in space, excluding bad points
lsqc3dgp.cpp least squares circle in 3 space determined by good points
levmarqd levenberg-marquardt nonlinear least squares example for gaussian
 function using numerical recipes functions
lsqfourier.ftn version of lsqgen.ftn for least squares approximation by trigonomet
rgpow.ftn regression for a power function
airperm.ftn regression for a power function
```

37 Calculation Examples

37.1 Birthdays

What is the probability of two or more common birthdays in an audience of n people? Let us neglect leap year and assume every year has 365 days.

The trick is to first calculate the probability that there are no common birthdays. The answer to the problem is one minus this. The number of possible birthday dates for the first person is 365, selecting one of these, the possibilities for the second person is 364. Having selected birthdays for the first $k-1$ persons, the possibility for the kth person is $365-k$. So the total number of ways that birthdays can be selected with no common birthday is

$$365(365-1)(365-2)....(365-(n-1)),$$

for n people in the audience. The possible ways of birthdays occuring in general is

$$365^n.$$

So the probability of no common birthdays q is the ratio of these two numbers,

$$q = \frac{365(365-1)(365-2)....(365-(n-1))}{365^n}.$$

Therefore the probability of one or more common birthdays is $p = 1 - q$,

$$p = 1 - \frac{365(365-1)(365-2)....(365-(n-1))}{365^n}.$$

Here is a computer program for calculating this probability for a list of audience sizes,

The Program birthdays.ftn:

```
c birthdays.ftn, the probability of two or more common birthdays
c in an audience of k people.
      implicit real*8(a-h,o-z)
      n=60
      do k=2,n
       q=1.0
       do i=0,k-1
        q=q*(365-i)/365
       end do
       p=1.0-q
       write(*,'(1x,a,i4,1x,a,g15.8)')'k=',k,'p=',p
      end do
      end
```

The List Produced by the program:

```
k=   2 p=  .27397260E-02
k=   3 p=  .82041659E-02
k=   4 p=  .16355912E-01
k=   5 p=  .27135574E-01
k=   6 p=  .40462484E-01
k=   7 p=  .56235703E-01
k=   8 p=  .74335292E-01
k=   9 p=  .94623834E-01
k=  10 p=  .11694818
k=  11 p=  .14114138
k=  12 p=  .16702479
k=  13 p=  .19441028
k=  14 p=  .22310251
k=  15 p=  .25290132
k=  16 p=  .28360401
k=  17 p=  .31500767
k=  18 p=  .34691142
k=  19 p=  .37911853
k=  20 p=  .41143838
k=  21 p=  .44368834
k=  22 p=  .47569531
k=  23 p=  .50729723
k=  24 p=  .53834426
k=  25 p=  .56869970
k=  26 p=  .59824082
k=  27 p=  .62685928
k=  28 p=  .65446147
k=  29 p=  .68096854
k=  30 p=  .70631624
k=  31 p=  .73045463
k=  32 p=  .75334753
k=  33 p=  .77497185
k=  34 p=  .79531686
k=  35 p=  .81438324
k=  36 p=  .83218211
k=  37 p=  .84873401
k=  38 p=  .86406782
k=  39 p=  .87821966
k=  40 p=  .89123181
k=  41 p=  .90315161
k=  42 p=  .91403047
k=  43 p=  .92392286
k=  44 p=  .93288537
k=  45 p=  .94097590
k=  46 p=  .94825284
k=  47 p=  .95477440
k=  48 p=  .96059797
k=  49 p=  .96577961
k=  50 p=  .97037358
k=  51 p=  .97443199
k=  52 p=  .97800451
k=  53 p=  .98113811
k=  54 p=  .98387696
k=  55 p=  .98626229
k=  56 p=  .98833235
k=  57 p=  .99012246
```

```
k=  58 p=  .99166498
k=  59 p=  .99298945
k=  60 p=  .99412266
```

So for example in an audience of 40 people, the probability of two or more common birthdays is almost 90 percent.

Of course for an audience of 366 people, independent of a probability calculation, two or more common birthdays is certain.

38 Bibliography

[1]Parzen Emanuel, **Modern Probability Theory and Its Applications**, Wiley, 1960.

[2]Hogg Robert V, Craig Allen T **Introduction to Mathematical Statistics**, Second Edition, Macmillan, 1966. (This book gives a derivation of the Students T distribution)

[3]Doob, J. L. **Stochastic Processes**, Wiley, 1953.

[4]Brunk, H. D. **Mathematical Statistics**, Blaisdell, 1965.

[5]Eisen, Martin **Introduction to Mathematical Probability Theory**, Prentice-Hall, 1969.

[6]Lamperti, John **Probability, An Introduction to the Mathematical Theory**, Prentice-Hall, 1969.

[7]Kolmogorov, A. N. **Foundations of the Theory of Probability**, Chelsea, New York, 1956 (translation of: Grundbegriffe der Wahrscheinlichkeitrechnung, which appeared in Ergebnisse Der Mathematik in 1933)

[8]Rainville Earl D. **The Laplace Transform: an Introduction**, Macmillan, 1965

[9]Van Der Pol, Balth., and Bremmer H. **Operational Calculus**, Cambridge University press, 1964

[10]Widder David Vernon **The Laplace Transform**, Princeton University Press, 1941

[11]Press William h, Teukolsky Saul A, Vetterling William, Flannery Brian P, **Numerical Recipes In Fortran 77**, 2nd Edition, 1996, (This book is available in several editions and versions).

[12] Knuth Donald E., **Seminumerical Algorithms, The Art of Computer Programming, V. 2**, Addison-Wesley, 1969, page 104.

[13] Ross Sheldon, **A First Course in Probability**, Macmillan, 3rd edition, 1988.