**Variance and Standard Deviation**

The mean of a RV X gives the “center “ of X’s distribution, the balancing point for the region below the graph of X’s probability density function. F

The variance and standard deviation of X are measures of the “spread” of X’s distribution.

If X is a continuous RV with pdf f(x), then

**variance** =  and

**standard deviation** = 

**Example** Let X have a uniform distribution on the interval [0,4]. Then,

f(x) =  and  = 2. So, .

> f<-function(x) 1/4\*(x>=0&x<=4)

> g<-function(x) (x-2)^2\*f(x)

> integrate(g,0,4)

1.333333 with absolute error < 1.5e-14

**Alternate Way to Compute Variance**



> xxf<-function(x) x^2\*f(x)

> integrate(xxf,0,4)

5.333333 with absolute error < 5.9e-14

**Median**

The median of X is the value m such that P(X ≤ m) = P(X ≥ m) = ½.

**Example:** Let X have the pdf f(x) =. Find the median of X.

**Quantiles/Percentiles**

Quantiles/percentiles are a generalization of the median. If 0 ≤ q ≤ 1, then the qth quantile/percentile is the value kq such that P(X ≤ kq) = q. The median is the 50th percentile for X.

**Example**: Find the 20th percentile of X in the example above.

> f<-function(x) 2\*x\*(x<=1&x>=0)

> integrate(f,0,.447)

0.199809 with absolute error < 2.2e-15

**Important Families of Distributions**

**Uniform Distributions**

A uniform distribution is characterized by the interval on which it is defined. If the interval is [a,b], then the pdf is f(x) =.

R has uniform distributions “built in”.

**dunif(x,min,max) computes the pdf at x**

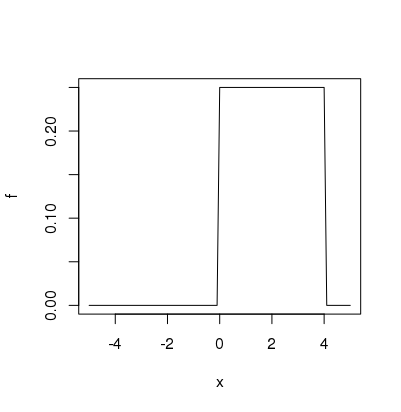
**punif(x,min,max) computes the value of the cdf at x**

**qunif(q, min, max) computes the qth quantile/percentile**

**runif(n,min,max) generates a random sample of n observations from the distribution.**

> f<-function(x) dunif(x,0,4)

> plot(f,-5,5)



P(1 ≤ X ≤ 3) > punif(3,0,4)-punif(1,0,4) [1] 0.5

P(X ≤ q) = .9 > qunif(.90,0,4) [1] 3.6

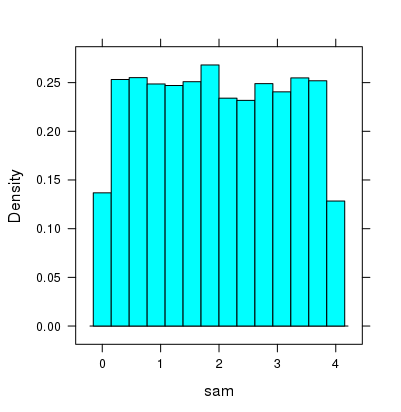
> runif(10,0,4)

[1] 2.9644012 1.1257994 2.5204828 0.2590472 1.4020343 1.8388996 2.0192951

[8] 1.1466988 1.9799251 0.3009602

> sam<-runif(10000,0,4)

> histogram(sam)



**Exponential Distributions**

If the pdf for X has the form f(x) =, then f has an exponential distribution with rate parameter > 0.

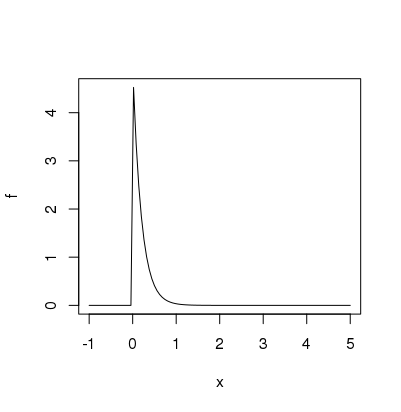
Models waiting times: X = time until the next click on a Geiger counter (radioactive decay) or X = time until a device stops functioning.

**Memoryless Property** For any a,b ≥0, P(X > a+b | X > a) = P(X > b)

**Exponential distributions in R: dexp(x,rate), pexp(x,rate), qexp(q,rate), rexp(n,rate)**

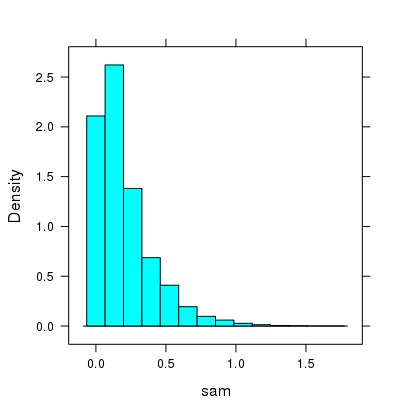
> f<-function(x) dexp(x,5)

> plot(f,-1,5)



> sam<-rexp(10000,5)

> histogram(sam)



**Mean of X** > xf<-function(x) x\*dexp(x,5)

> integrate(xf,0,Inf)

0.2 with absolute error < 8.3e-05

**Variance of X**  > xxf<-function(x) x^2\*f(x)

> integrate(xxf,0,Inf)

0.08 with absolute error < 8.2e-06

Variance = .08-(.02)^2 = .04

Standard Deviation = sqrt(.04) = .2

**Normal Distributions (The Famous Bell-shaped Curves)**

The pdf has the form  This family of distributions has two parameters:

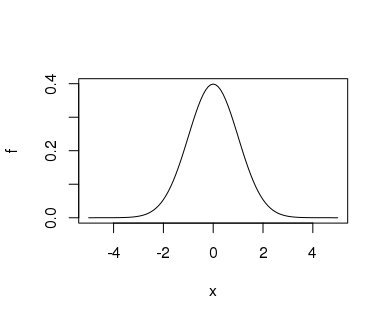
 = mean = standard deviation

If  = 0 and  = 1, the distribution is the **standard normal distribution**. In the standard normal case, we let Z represent the random variable. The standard normal distribution is often called the z-distribution. Many features have a normal distribution: IQ scores; SAT scores; heights of adult Frenchmen;….

**Normal Distributions in R: dnorm(x,mu, sigma), pnorm(x,mu,sigma), etc.**

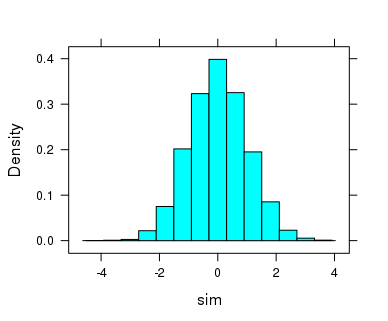
> f<-function(x) dnorm(x,0,1)

> plot(f,-5,5)



> sim<-rnorm(10000,0,1)

> histogram(sim)



**Example** The height of North American adult males has (approximately) a normal distribution with mean = 70 inches and sd = 3 inches.

What proportion of NA adult males are 6 ft = 72 inches or less. If X is the height of a randomly selected NAAM, we want P(X≤72). The cdf for X calculates that. F(72) = P(X≤72).

> pnorm(72, 70,3)

[1] 0.7475075

What is the height H such that 99% of all heights are ≤ H? Equivalently, what is the 99th percentile?

> qnorm(.99,70,3)

[1] 76.97904

**The 68, 95, 99.7 rule**

If X has a normal distribution with mean  and standard deviation , then







If mu = 5 and sigma = 2 If mu = 0 and sigma = 1

> pnorm(5,5,2)-pnorm(3,5,2) > pnorm(3,0,1)-pnorm(-3,0,1)

[1] 0.3413447 [1] 0.9973002

**Exercises 7**

1. Find, **by hand,** the exact values of the mean and variance for X if X has a uniform distribution on [1,9].
2. **Extra Credit:** Find, **by hand**, the mean, variance, and standard deviation for X if X has a uniform distribution on the interval [a,b].
3. IQ scores are normally distributed with mean = 100 and sd =15. Use R to answer the following. Include the R commands in your solution.
4. Carol has an IQ score of 128. What percent of people have IQ scores no higher than Carol’s IQ score?
5. Born in 1982, Christopher Hirata had recorded an IQ of 225. He is an outstanding example of child prodigy who had already completed his college level courses by the time he was 12. By the time he turned 16, he was working on a project with the NASA. He was the first one to win a gold medal in international Science Olympiad at the age of 13. He began attending classes for his PhD in astrophysics when he was just 18. At what percentile is his IQ score?
6. What is the 99th percentile for IQ scores?
7. What percent of people have IQ scores in the range 90 to 110?
8. Scores on an aptitude test are normally distributed with mean  and standard deviation . 16% of scores are below 20 and 2.5% of scores are above 110. Use the 68-95-99.7% rule to find  and .
9. Another important family of distributions is the set of **gamma distributions.** There are two parameters for the gammas:  = **shape** and = **rate**. If shape = 1, then gamma is an exponential distribution. The gamma pdf is given by , where C is the number that makes the total integral equal to 1. In R the pdf is given by dgamma(x,shape,rate), the cdf by pgamma(x,shape, rate), etc.
10. Plot the graphs of the following gamma pdfs

shape = 2 rate = 4

shape = 5 rate = 10

shape = 10 rate = 5

shape = 4 rate = 2

1. Based on the plots above, how would you describe the general shape of a gamma distribution?
2. Find the mean and variance of the gamma distribution when shape = 4 and rate = 3. **Use R to do this.**
3. Another important family of distributions is the set of **beta distributions**. There are two parameters for beta,  = shape1 and  = shape2. The pdf is for an appropriate C. In R the pdf is given by dbeta(x, shape1, shape2), etc.
4. Plot the graphs of the following beta pdf’s

shape1 = 2 shape2 = 2

shape1= 2 shape2 = .9

shape1 =4 shape2 = 2

shape1 = .9 shape2 = .85

b. Based on the plots above, what would you say about the shapes of a beta distribution.?