1 Introduction

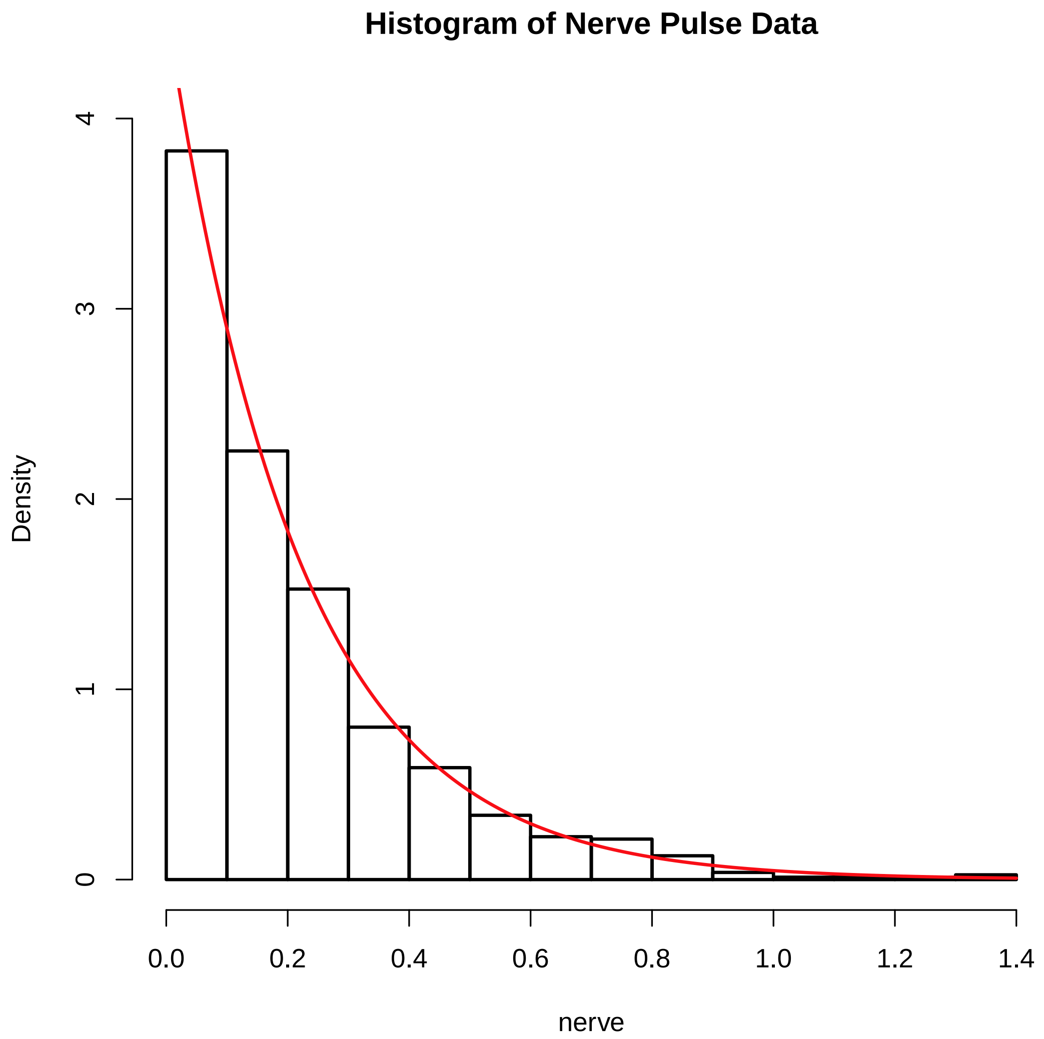
In the later chapters when we study the application of probability theory to statistical inference, we will often consider a random sample of size n from a given distribution.

We say that $X_1, \ldots, X_n$ is a random sample from a distribution with pdf/pmf f_θ if $X_1, \ldots, X_n$ are independent random variables each having pdf/pmf f_θ.

In statistics, θ represents an unknown parameter that specifies the specific distribution within a family of distributions. We collect the random sample $X_1, \ldots, X_n$ and from the sample, we compute one or more statistics, $h(X_1, \ldots, X_n)$, to carry out inference concerning θ.

To obtain the properties of the resulting statistical procedures, we need to obtain the distribution of $Y = h(X_1, \ldots, X_n)$ (often called *sampling distribution* of Y). When $h(X_1, \ldots, X_n)$ is a sum (or mean), we can use techniques from Chapter 3 to obtain its distribution.

Example 43 from Chapter 6 Cox and Lewis (1966) reported 799 waiting times between successive pulses along a nerve fiber. The data appear in the following histogram:



To model these data, we suppose that $X_1, \ldots, X_n$ is a random sample from an exponential (λ) distribution.

In Chapter 6 we will obtain $\hat{\lambda} = n / \sum_{i=1}^{n} X_i$ as the *maximum likelihood estimator* of λ. This and other methods of statistical inference concerning λ will involve functions of the sum, $Y = \sum_{i=1}^{n} X_i$.

We can use the moment generating function to obtain the distribution of Y:

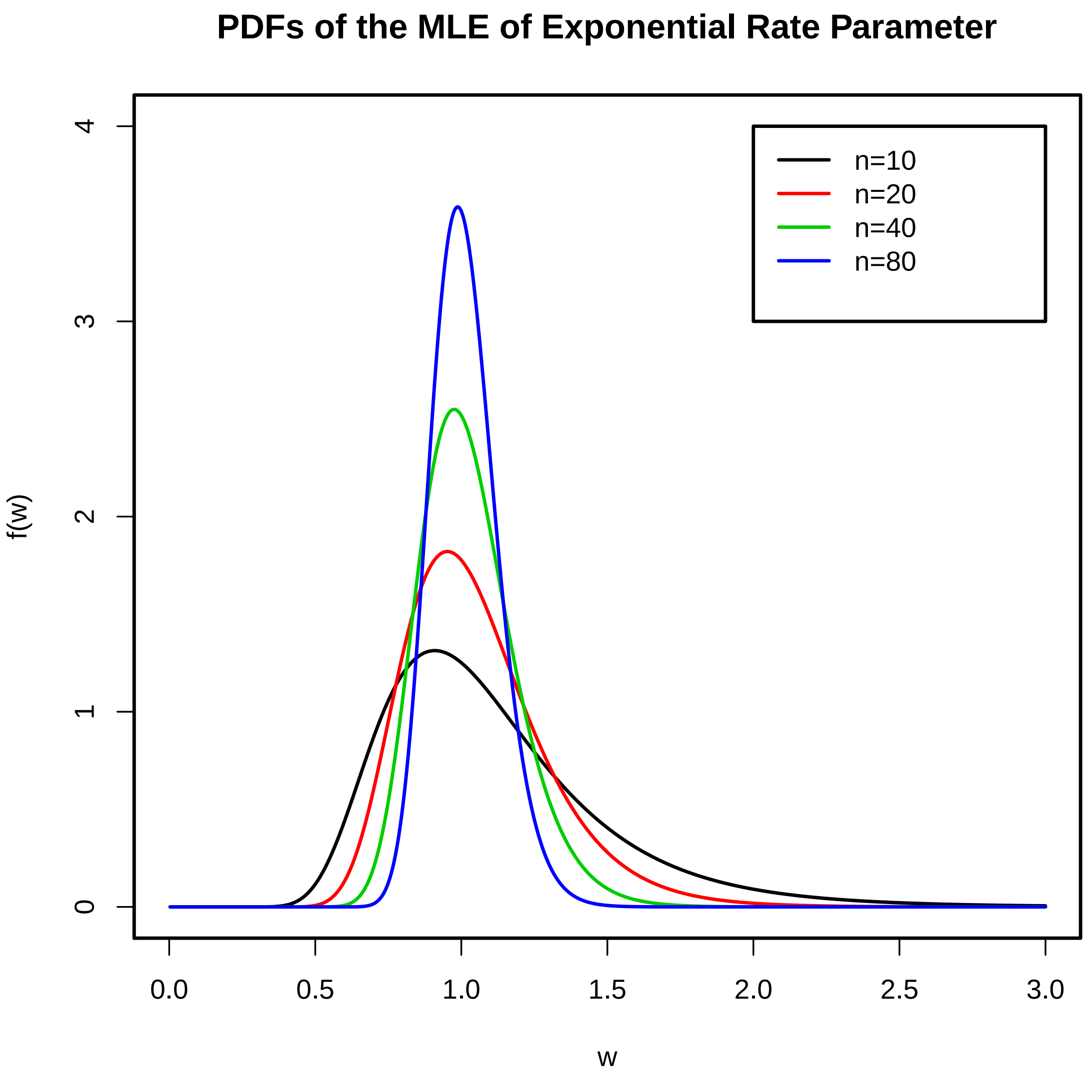
$$\begin{aligned} M_Y(s) &= M_{X_1+\cdots+X_n}(s) = E\{\exp[s(X_1 + \cdots + X_n)]\} \\ &= \prod_{i=1}^{n} E[e^{sX_i}] = \left(\frac{\lambda}{\lambda - s}\right)^n \end{aligned}$$

We recognize this as the mgf of a gamma (n, λ) distribution.

We can use methods of Chapter 2 to obtain the exact distribution of $W = n/Y$ with pdf

$$f_W(w) = \frac{(n\lambda)^n}{\Gamma(n) w^{n+1}} e^{-\lambda n/w}, \quad w > 0.$$

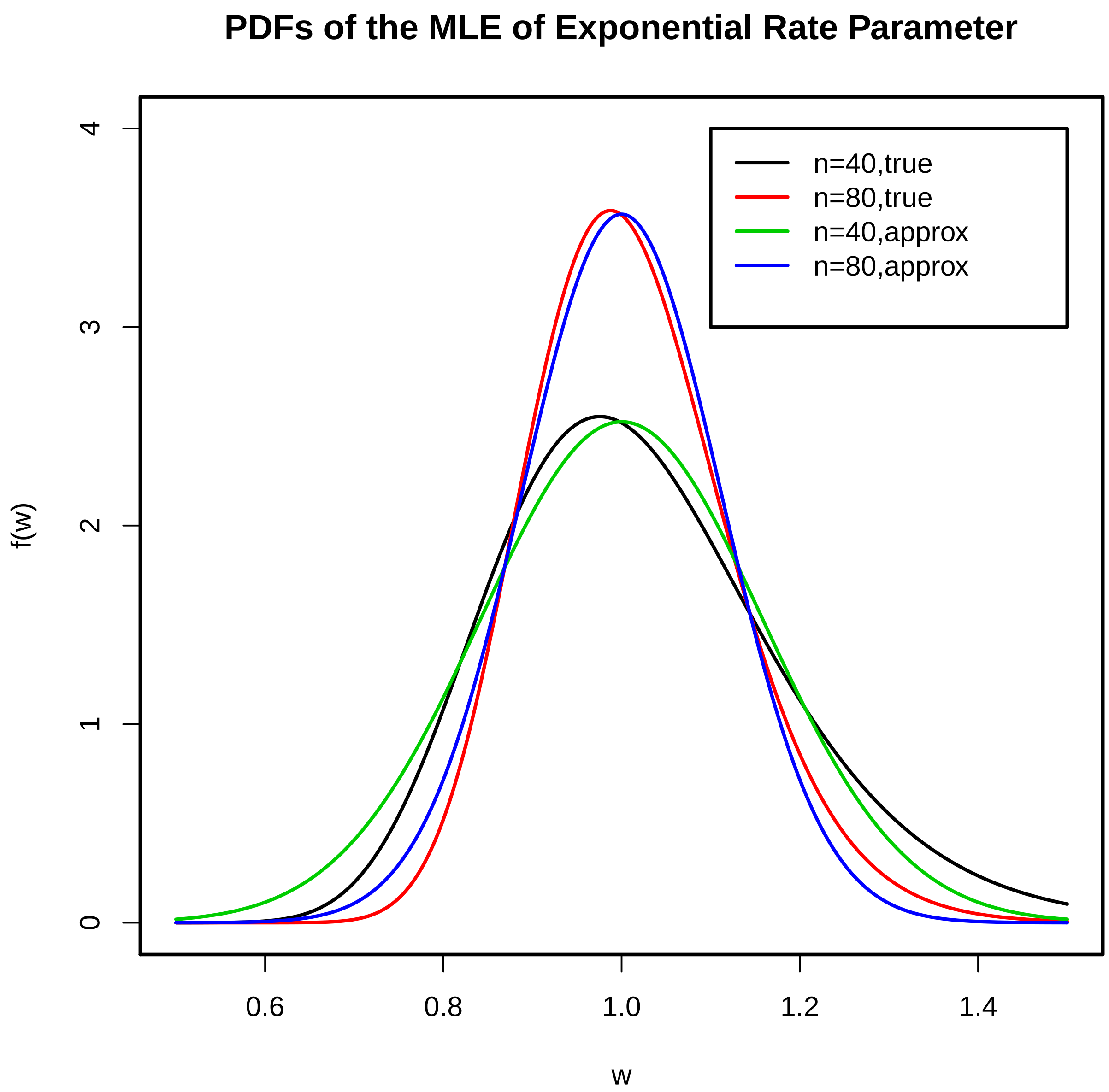
Here are plots of the pdf of the mle when $\lambda = 1$ various values of n.



Remarks: For the above example of the distribution of the mle for the rate parameter for the exponential distribution, the derivation of the distribution was relatively straight-forward. This will not be the case for many problems of interest, and we may want to pursue other approaches.

We will find that in certain problems, we can approximate the sampling distribution of a statistic using the normal distribution. In this chapter we will see this for the sample mean using a result called the *Central Limit Theorem*. Later we will see that a similar result holds for the maximum likelihood estimate. Below is a plot of the exact distribution of the mle for $n = 40$ and $n = 80$ for the mle of the exponential rate parameter. The approximating normal pdf is also plotted.

Also, notice that the spread of the distributions is decreasing as the sample size increases. This implies that the probability that the estimator is close to the true value should increase with sample size. We will see in the current chapter that this holds for the sample mean. In a later chapter we will see that a similar result hold for mles.


PDFs of the MLE of Exponential Rate Parameter
n=40,true
n=80,true
n=40,approx
n=80,approx
f(w)
w
0
1
2
3
4
0.6
0.8
1.0
1.2
1.4

The limiting behavior of sums (or sample means) of independent random variables provides useful and interesting results for approximating the distribution of the sum (or sample mean).

- The Weak Law of Large Numbers tells us that a sample mean will be very close (with high probability) to the population mean for a large enough sample.
- The Central Limit Theorem enables us to use the normal distribution to approximate the sampling distribution of the sample mean for large samples.

2 Convergence in Probability and the Law of Large Numbers

We first recall Chebychev's inequality and consider an application of it to the distribution of the sample mean.

Chebychev's inequality: Suppose X is a random variable with finite variance σ_X^2, and let $\mu_X = E(X)$. Then for each $a > 0$

$$P(|X - \mu_X| \geq a) \leq \frac{\sigma_X^2}{a^2}.$$

Application: Estimation of μ using $\bar{X}_n$

Suppose that $X_1, \ldots, X_n$ is a random sample from a distribution with mean μ_X and variance σ_X^2. Consider the problem of estimating μ_X by using the sample mean.

Let

$$\bar{X}_n = (\text{or } M_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + \cdots + X_n)$$

be the sample mean. Then

$$E(\bar{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{n\mu_X}{n} = \mu_X.$$

and

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{n\sigma_X^2}{n^2} = \frac{\sigma_X^2}{n}.$$

By Chebychev's Theorem, we have

$$P\left(|\bar{X}_n - \mu_X| \geq a\right) \leq \frac{\sigma_X^2}{n} \cdot \frac{1}{a^2}.$$

This tells us that when n is sufficiently big, there is only a very small probability that $\bar{X}_n$ will differ from μ by more than a.

Assuming we know σ^2, the last inequality can be used to determine a sample size, n, that will guarantee

$$P\left(|\bar{X}_n - \mu| < a_0\right) \geq p.$$

Set

$$1 - p = \frac{\sigma^2}{n} \cdot \frac{1}{a_0^2}$$

and solve for n. This gives

$$n = \frac{\sigma^2}{1-p} \cdot \frac{1}{a_0^2}.$$

Suppose, for example, that $\sigma = 4$ and we want to be 95% sure that the estimation error is no more than 1. Then

$$n = \frac{16}{0.05} \cdot \frac{1}{1^2} = 320.$$

2.1 The Weak Law of Large Numbers

Earlier we saw that Chebyshev's inequality implies that

$$P(|\bar{X}_n - \mu_X| \geq \varepsilon) \leq \frac{\sigma_X^2}{n} \cdot \frac{1}{\varepsilon^2}.$$

For any fixed value $\varepsilon > 0$, suppose we let the sample size increase indefinitely; i.e., let $n \longrightarrow \infty$. Then

$$0 \leq \lim_{n \to \infty} P(|\bar{X} - \mu_X| \geq \varepsilon) \leq \lim_{n \to \infty} \frac{\sigma_X^2}{n} \cdot \frac{1}{\varepsilon^2} = 0$$

Thus, no matter how small ε is, the probability that $\bar{X}_n$ is farther away from μ_X than ε goes to zero as the sample size goes to infinity. This result holds for any distribution with finite variance.

The above result is known as the weak law of large numbers. It is a special case of convergence in probability.

Definition: Let $Y_1, Y_2, \ldots$ be a sequence of random variables. The sequence $\{Y_i\}$ converges in probability to the constant b if for every number $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|Y_n - b| > \varepsilon) = 0.$$

We write $Y_n \xrightarrow{P} b$.

On the preceding page, we showed that

$$\bar{X}_n \xrightarrow{P} \mu.$$

Remark: See the text for a definition of convergence in probability of a sequence of random variables to a random variable Y.

Example 38   Let $X_1, X_2, \ldots$ be a sequence of Bernoulli (θ) rvs.

Let $Y_n = X_1 + \cdots + X_n$. Then $Y_n \sim$ Binomial (n, θ).

Later we will show that a "good" estimator of θ is

$$\hat{\theta}_n = Y_n/n.$$

We can use the Law of Large Numbers to say something about the large sample behavior of $\hat{\theta}_n$.

Since $E(X_i) = \theta$ and $\mathrm{Var}(X_i) = \theta(1-\theta)$, we can apply the Law of Large Numbers to conclude that

$$\hat{\theta}_n = \frac{Y_n}{n} = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} \theta.$$

Monte Carlo Approximation of an Integral

We can use the Weak Law of Large Numbers to justify using simulation to obtain a numerical approximation of an integral. The first step is to recognize whether the integral I can be written as an expectation of a function $h(\cdot)$ of a random variable X from a given distribution:

$$I = \int_{R_X} h(x) f_X(x) dx.$$

If so, we can use the following algorithm to approximate I:

1. Select a large positive integer n.
2. Generate $X_1, \ldots, X_n$ as a random sample from the distribution with pdf f_X.
3. Set $T_i = h(X_i)$.
4. Estimate I by $M_n = (T_1 + \cdots + T_n)/n$.

The WLLN implies that

$$M_n \xrightarrow{P} E(T_1) = E(h(X_1)) = \int_{R_X} h(x) f_X(x) dx = I.$$

3 Convergence in Distribution and the Central Limit Theorem

We used the mgf in Chapter 3 to show that a sum of independent normal random variables is normally distributed. As a special case, suppose that $X_1, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ distribution. Let $\bar{X}_n = \sum_{i=1}^n X_i/n$ be the sample mean. Then we know that

$$E(\bar{X}_n) = \mu \quad \text{and} \quad \mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Thus, since a linear combination of independent normal rvs is normally distributed,

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

A remarkable theorem and natural law is that this property remains approximately true under very general conditions *even when the distribution of each X_i is not normal.*

Later when we discuss statistical inference, we will see that hypothesis tests and confidence intervals concerning the mean μ of a normal population are based on the fact that we can standardize the sample mean

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.$$

We derive confidence intervals using the fact that

$$P\left(-Z_{(1+\gamma)/2} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < Z_{(1+\gamma)/2}\right) = \Phi(Z_{(1+\gamma)/2}) - \Phi(-Z_{(1+\gamma)/2}) = \gamma.$$

We base the rejection region of a test of hypothesis on the fact that

$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} > Z_{1-\alpha}\right) = 1 - \Phi(Z_{1-\alpha}) = \alpha.$$

We see from this that knowing the cdf of statistic (or at least, approximating it) will be very useful in statistical inference.

Definition: Let $X_1, X_2, \ldots$ be a sequence of random variables with cdfs $F_1, F_2, \ldots$, respectively. We say that X_n converges to X in distribution if

$$\lim_{n \to \infty} F_n(x) = F(x)$$

at all points x where F is continuous. We write $X_n \xrightarrow{D} X$.

Typically proving convergence in distribution using the definition is difficult. It is often much easier using moment generating functions.

Theorem: Let $X_1, X_2, \ldots$ be a sequence of random variables with cdfs $F_1, F_2, \ldots$ and mgfs $M_1, M_2, \ldots$, respectively. Suppose that X is a rv with cdf F and mgf M. If $M_n(s) \to M(s)$ for all s in an open interval containing zero, then X_n converges in distribution to X.

Example 39 Normal Approximation of the Poisson Distribution

Suppose that $X_n \sim$ Poisson (n). Let $Z_n = \frac{X_n - n}{\sqrt{n}}$. Since

$$M_{X_n}(s) = e^{n(e^s - 1)},$$

by the properties of mgfs we have

$$M_{Z_n}(s) = e^{-s\sqrt{n}} M_{X_n}(s/\sqrt{n}) = e^{-s\sqrt{n}} e^{n(e^{s/\sqrt{n}} - 1)}.$$

Take the logarithm and use the power series expansion $e^a = \sum_{k=0}^{\infty} \frac{a^k}{k!}$ to get

$$\begin{aligned} \log(M_{Z_n}(s)) &= -s\sqrt{n} + n(e^{s/\sqrt{n}} - 1) \\ &= -s\sqrt{n} + n(1 + s/\sqrt{n} + (s/\sqrt{n})^2/2! + \cdots - 1) \\ &= n\left(\frac{s^2}{2n} + \frac{s^3}{n^{3/2}3!} + \cdots\right) = \frac{s^2}{2} + R_n, \end{aligned}$$

where $R_n \longrightarrow 0$ as $n \longrightarrow \infty$.

From the previous slide,

$$\log(M_{Z_n}(s)) = \frac{s^2}{2} + R_n,$$

where $R_n \longrightarrow 0$ as $n \longrightarrow \infty$. Thus,

$$\lim_{n\to\infty} M_{Z_n}(s) = e^{s^2/2},$$

which is the mgf of a standard normal rv. Hence, Z_n converges in distribution to a standard normal rv.

Example 40 Normal Approximation to the Binomial Distribution

Let $X_n \sim$ binomial (n, θ). Then one can show that

$$\frac{\sqrt{n}(X_n - n\theta)}{\sqrt{n\theta(1-\theta)}} \xrightarrow{D} Z,$$

where $Z \sim N(0, 1)$. This is a classical result in probability theory with proofs due to DeMoivre and LaPlace. We will return to this result as an easy consequence of the Central Limit Theorem.

We can use the convergence of mgfs to prove a simple version of the central limit theorem which says we can approximate the cdf of the sample mean using the normal distribution.

Lindeberg and Lévy Central Limit Theorem. Let $X_1, \ldots, X_n$ be a random sample from a distribution having variance σ^2 (with $0 < \sigma^2 < \infty$) and mean μ. Let $\bar{X}_n$ be the sample mean, i.e.,

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

Then, for each real number z,

$$\lim_{n \to \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z \right) = \Phi(z).$$

Thus, the sequence $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}, n = 1, 2, \ldots$ converges in distribution to Z, a standard normal rv.

We can also express the Central Limit Theorem in terms of the sum, $S_n = X_1 + \cdots + X_n$, of the sequence of rvs $X_1, \ldots, X_n$ satisfying the assumptions of the Central Limit Theorem:

For each real number z,

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le z\right) = \Phi(z).$$

A less rigorous (but more revealing) way of stating the central limit theorem (CLT) is to say that, under the conditions of the theorem,

$\bar{X}_n$ is approximately distributed $N\left(\mu, \frac{\sigma^2}{n}\right)$ for all n sufficiently large.

or

S_n is approximately distributed $N\left(n\mu, n\sigma^2\right)$ for all n sufficiently large.

The CLT provides an explanation as to why many variables in practice tend to be normally distributed. Take for example IQ scores and certain measurement errors, which are approximately normally distributed.

- A person's intelligence is undoubtedly due to the combined influence of many different factors. Seems reasonable that IQ score is approximately the sum of a number of different quantities, each of which corresponds to a different factor influencing intelligence.
- In certain measurements such as astronomical measurements of position, it is reasonable to assume that light passing through many layers of the atmosphere which each impart an error to the measurement of position. Thus, the total error in measurement can be viewed as a sum of many small errors.

Example 41: Distribution of the Mean of Uniform Random Variables

Let $X_1, \ldots, X_n$ have a uniform $(0, 1)$ distribution. We let $\bar{X}_n$ be the sample mean. The central limit theorem implies that for *large* n, the distribution of $\bar{X}_n$ can be approximated using a normal distribution. The question is “What is large n?”

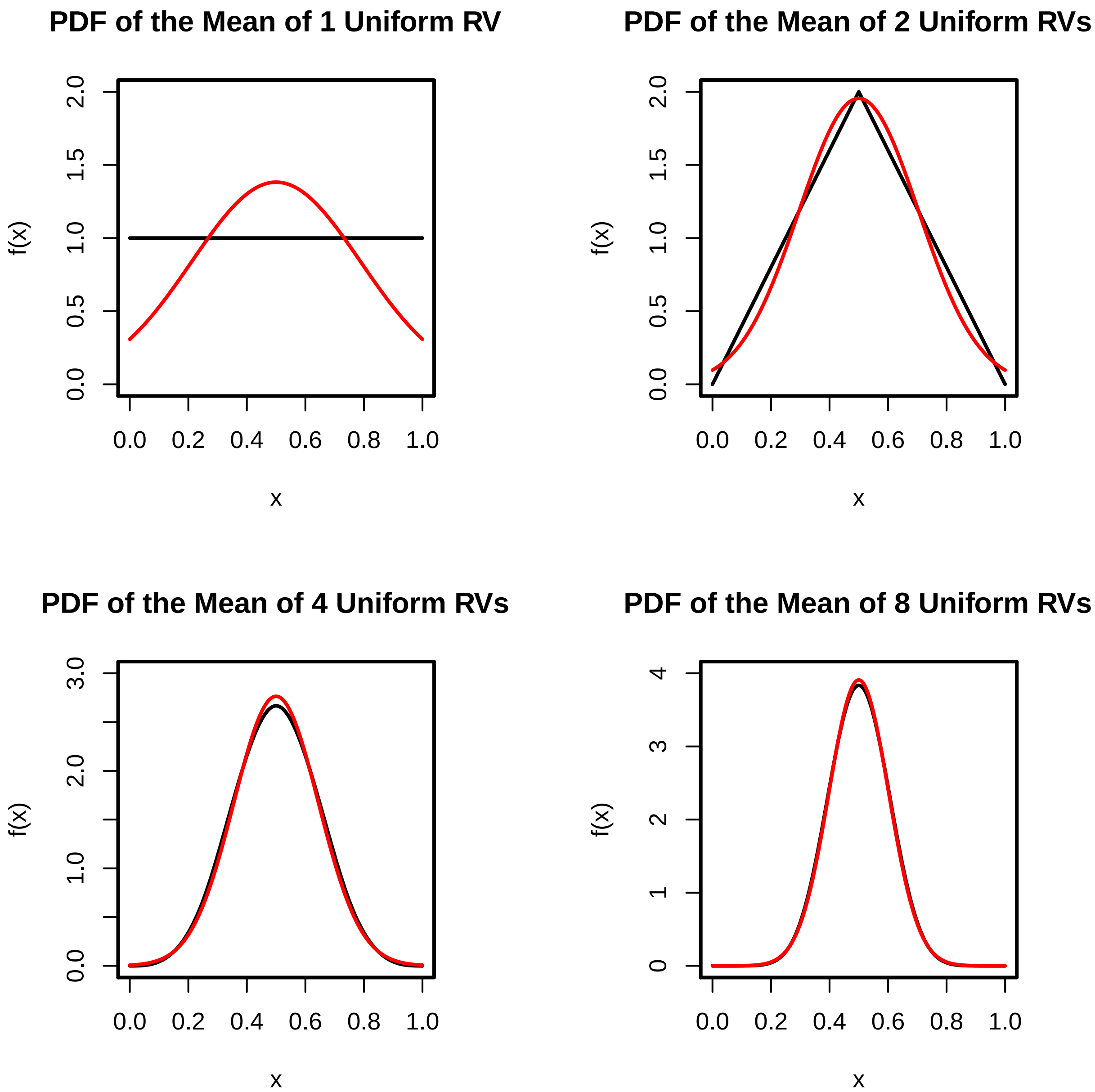
The following slide shows the exact distribution of $\bar{X}_n$ for $n = 1, 2, 4, 8$. We know that

$$\mu = \frac{1}{2} \quad \text{and} \quad \mathrm{Var}(X) = \frac{1}{12}$$

for the uniform$(0, 1)$ distribution. Then the mean and variance of $\bar{X}_n$ are

$$\mu = \frac{1}{2} \quad \text{and} \quad \sigma^2 = \frac{1}{n \cdot 12}.$$

The lines in red are the pdfs of the normal distribution with the above mean and variance.



PDF of the Mean of 1 Uniform RV
PDF of the Mean of 2 Uniform RVs
PDF of the Mean of 4 Uniform RVs
PDF of the Mean of 8 Uniform RVs
f(x)
x
0.0 0.5 1.0 1.5 2.0
0.0 1.0 2.0 3.0
0 1 2 3 4
0.0 0.2 0.4 0.6 0.8 1.0

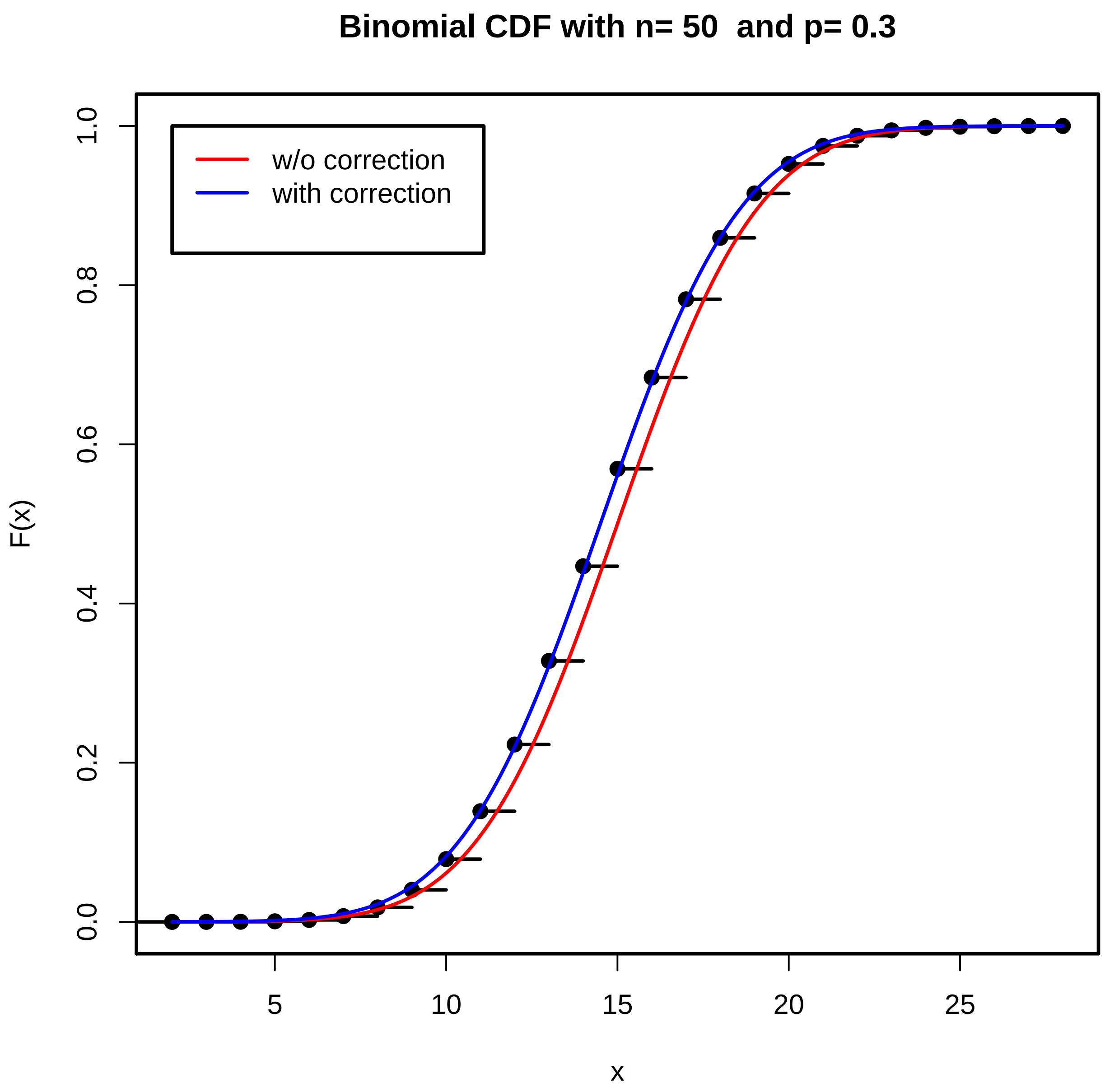
Example 40 (again): Normal Approximation of the Binomial Distribution

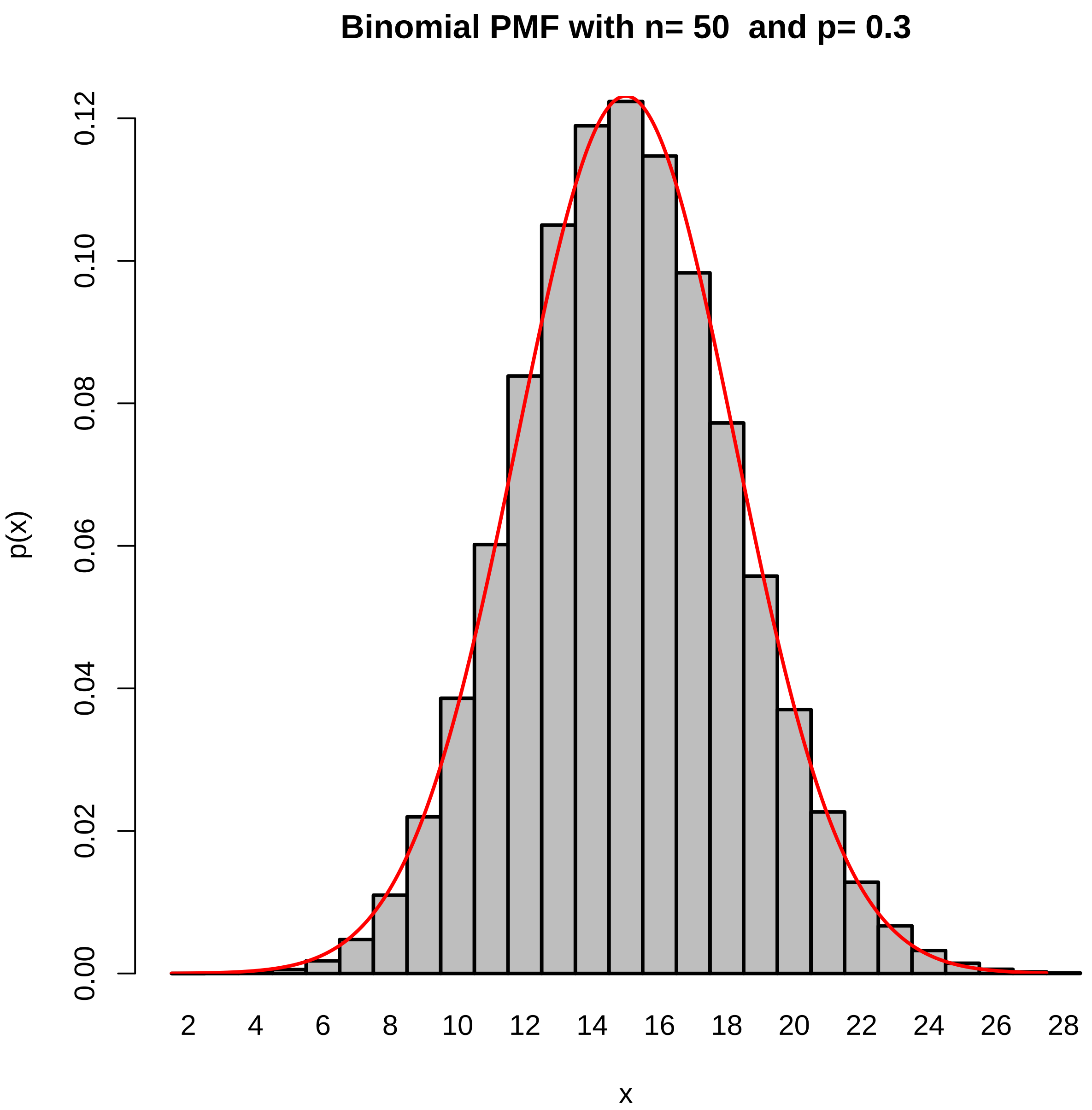
Recall Example 35 where we represented a binomial rv X with n trials and probability of success θ as a sum of Bernoulli rvs:

$$X = \sum_{i=1}^{n} X_i.$$

We write

$$\begin{aligned} P(X \le x) &= P(\bar{X}_n \le x/n) = P\left(\frac{\bar{X}_n - \theta}{\sqrt{\theta(1-\theta)/n}} \le \frac{x/n - \theta}{\sqrt{\theta(1-\theta)/n}} \right) \\ &\doteq \Phi\left(\frac{x - n\theta}{\sqrt{n\theta(1-\theta)}} \right) \end{aligned}$$


Binomial CDF with n= 50 and p= 0.3
w/o correction
with correction
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
25
x


Binomial PMF with n= 50 and p= 0.3
p(x)
0.00
0.02
0.04
0.06
0.08
0.10
0.12
2
4
6
8
10
12
14
16
18
20
22
24
26
28
x

Accuracy of the Monte Carlo Approximation

Recall that $X_1, \ldots, X_n$ are iid from a distribution with pdf f_x and $T_i = h(X_i),\ i = 1, ..., n$.

Then $E(T_i) = \int_{R_X} h(x) f_X(x) dx = I$.

Let $\sigma_T^2 = \mathrm{Var}(h(X_i))$.

Letting

$$M_n = \frac{1}{n} \sum_{i=1}^{n} T_i,$$

we have by the Central Limit Theorem,

M_n is approximately distributed $N\left(I, \frac{\sigma_T^2}{n}\right)$ for all n sufficiently large.

To obtain a region that has a high probability of containing the exact value of the integral, we use the normal approximation to obtain

$$P[I - 3\sigma_T/\sqrt{n} \leq M_n \leq I + 3\sigma_T/\sqrt{n}] \doteq 0.9973.$$

Thus, with high probability $|M_n - I| \leq 3\sigma_T/\sqrt{n}$.

In Chapter 6 we will extend this reasoning to obtain *approximate 99.7% confidence bounds on* I given by

$$[M - 3S/\sqrt{n}, M + 3S/\sqrt{n}],$$

where we use the sample variance of $T_1, \ldots, T_n$,

$$S^2 = \frac{\sum_{i=1}^{n}(T_i - M_n)^2}{n-1},$$

to estimate the unknown variance of $T_i = h(X_i)$, σ_T^2.

4 Normal Distribution Theory

The normal distribution plays an important role in statistical inference. We will now present some results related to the distribution of statistics computed from normal data. We also will introduce some commonly used distributions in statistical applications.

Distribution of a sum of independent normal rvs

Let $X_1, \ldots, X_n$ be independent random variables such that

$$X_i \sim N(\mu_i, \sigma_i^2), \quad i = 1, \ldots, n.$$

Let $Y = b + a_1 X_1 + \cdots + a_n X_n$. Then using the mgf as in Example 36, we obtain

$$Y \sim N\left(b + \sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

Distribution of the Sample Mean

If $X_1, \ldots, X_n$ forms a random sample from a $N(\mu, \sigma^2)$ distribution, then

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n).$$

Covariance of Linear Combinations of Normal Random Variables

Let $X_1, \ldots, X_n$ be independent random variables such that

$$X_i \sim N(\mu_i, \sigma_i^2), \quad i = 1, \ldots, n.$$

Define $U = a + \sum_{i=1}^{n} b_i X_i$ and $V = c + \sum_{j=1}^{n} d_j X_j$. Then

$$\mathrm{Cov}(U, V) = \sum_{i=1}^{n}\sum_{j=1}^{n} b_i d_j \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n} b_i d_i \sigma_i^2.$$

5 The Chi-Squared Distribution

The chi-squared distribution with n degrees of freedom (df) is a gamma distribution with $\alpha = n/2$ and $\lambda = 1/2$. It has pdf

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2}, \quad x > 0.$$

If X has a chi-squared distribution with n df, we write

$$X \sim \chi^2(n).$$

Properties of the Chi-Squared Distribution

- If $X \sim \chi^2(n)$, then $E(X) = n$ and $\mathrm{Var}(X) = 2n$.

- The mgf of a chi-squared rv X with n df is

$$M_X(s) = \left(\frac{1}{1-2s}\right)^{n/2}, \text{ for } s < 1/2.$$

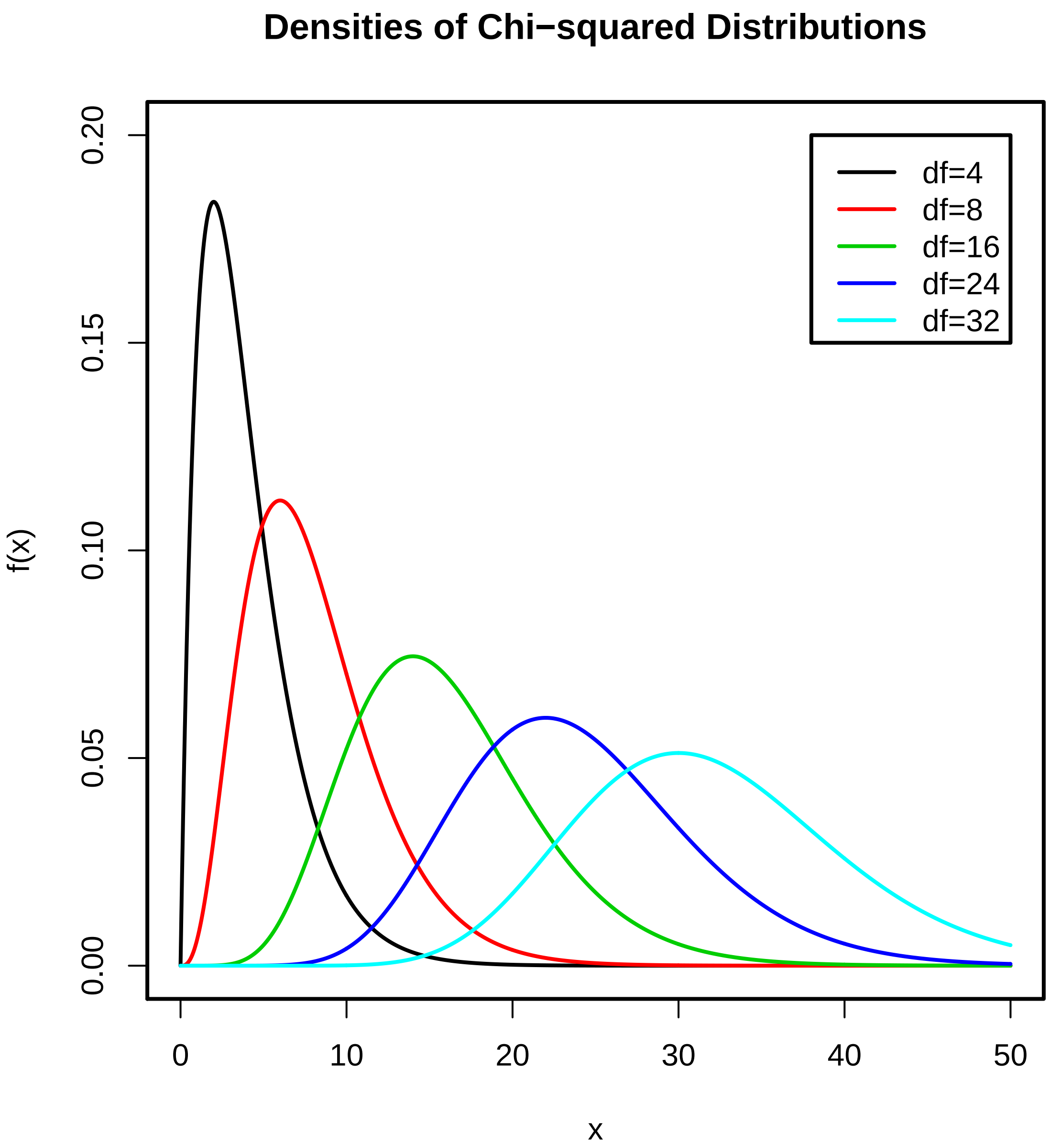
- If $X_1, \ldots, X_k$ are independent chi-squared rvs with dfs $n_1, \ldots, n_k$, respectively, then

$$X_1 + \cdots + X_k \sim \chi^2(n_1 + \cdots + n_k).$$

- If $Z \sim N(0,1)$, then $Z^2 \sim \chi^2(1)$. This was shown in Example 22.

- If $Z_1, \ldots, Z_n$ form a random sample from the $N(0,1)$ distribution, then

$$Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n).$$



Densities of Chi−squared Distributions
df=4
df=8
df=16
df=24
df=32
0.00
0.05
0.10
0.15
0.20
f(x)
0
10
20
30
40
50
x

6 The t Distribution

The t distribution is fundamental for statistical inference concerning the mean of a normal population. We will first define it rather abstractly and then see how it relates to sampling from a normal distribution.

Consider the following two independent rvs:

- $Z \sim N(0,1)$
- $U \sim \chi^2(n)$

Then the following rv is said to have a t distribution with n df (written as $t(n)$):

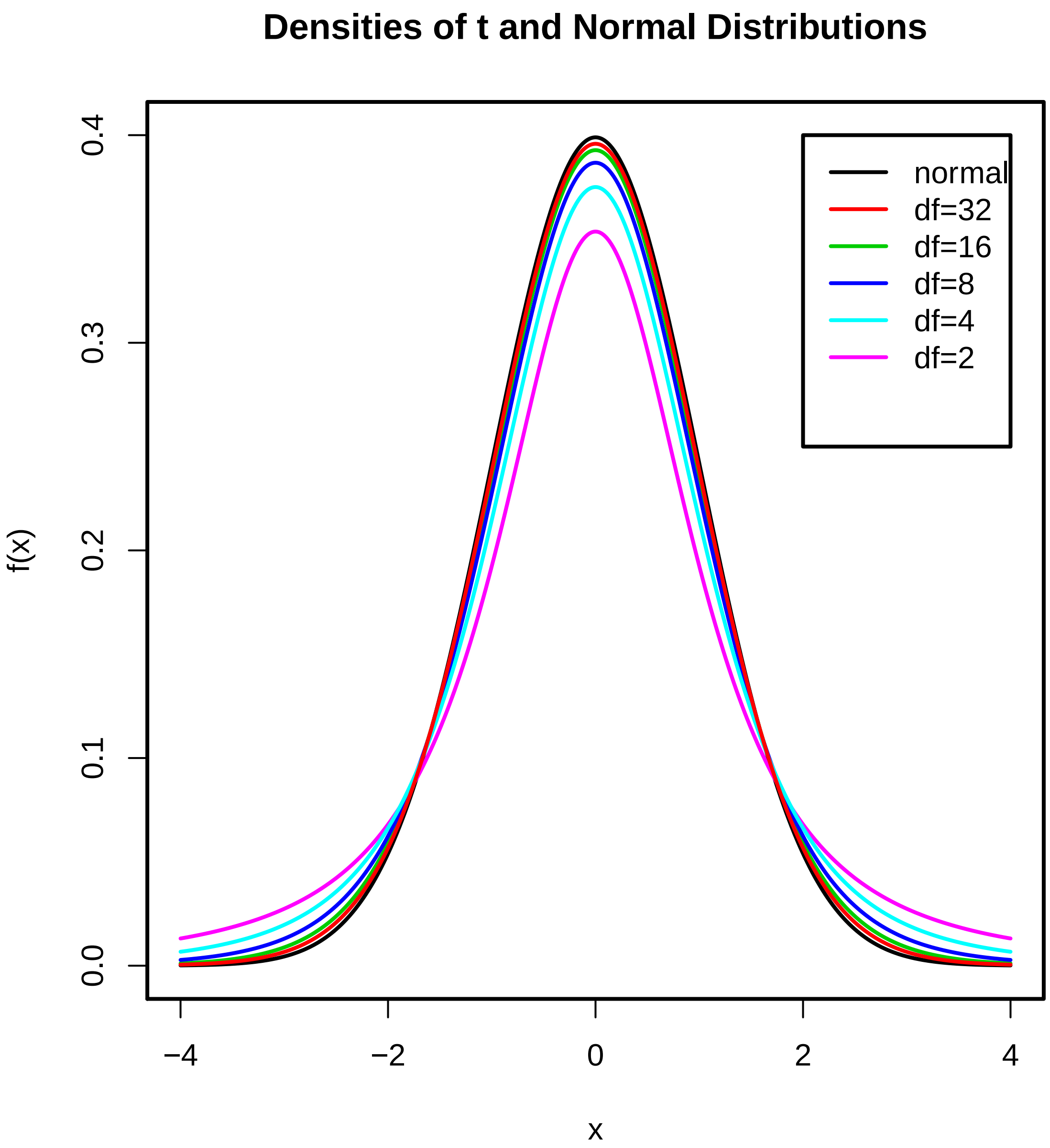
$$T = \frac{Z}{\left(\dfrac{U}{n}\right)^{1/2}}$$

It is a straight-forward exercise to derive the pdf of T:

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}, \text{ where } -\infty < t < \infty.$$

6.1 Properties of the t Distribution

- Each $t(n)$ pdf is symmetric about zero.
- For $n = 1$, the $t(1)$ distribution is the same as the Cauchy distribution.
- Each $t(n)$ pdf has a smaller peak and thicker tails than the standard normal pdf.
- As $n \to \infty$, the $t(n)$ pdf approaches the standard normal pdf.
- For $n > 1$, $E(T) = 0$.
- For $n > 2$, $\mathrm{Var}(T) = n/(n-2)$.


Densities of t and Normal Distributions
normal
df=32
df=16
df=8
df=4
df=2
f(x)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
x

7 The F Distribution

The F Distribution is used in inferences concerning variances of normal populations.

Consider the two following independent random variables:

- $U \sim \chi^2(m)$
- $V \sim \chi^2(n)$

The following random variable has the F Distribution with m and n degrees of freedom (written as $F(m, n)$):

$$W = \frac{U/m}{V/n}.$$

The pdf of the $F(m,n)$ Distribution can be shown to be

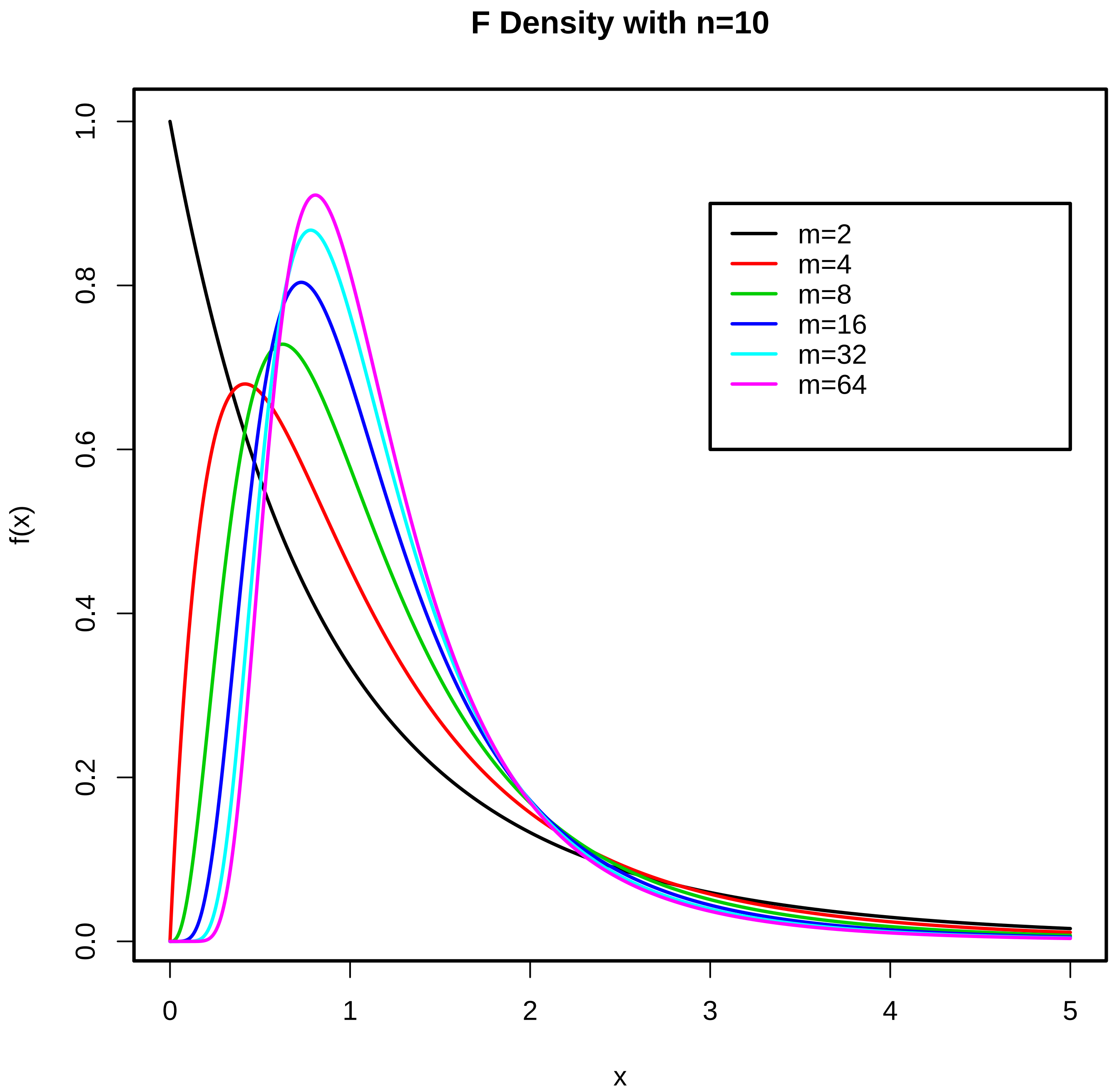
$$f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} w^{m/2-1} \left(1 + \frac{m}{n}w\right)^{-(m+n)/2}, \; w > 0.$$

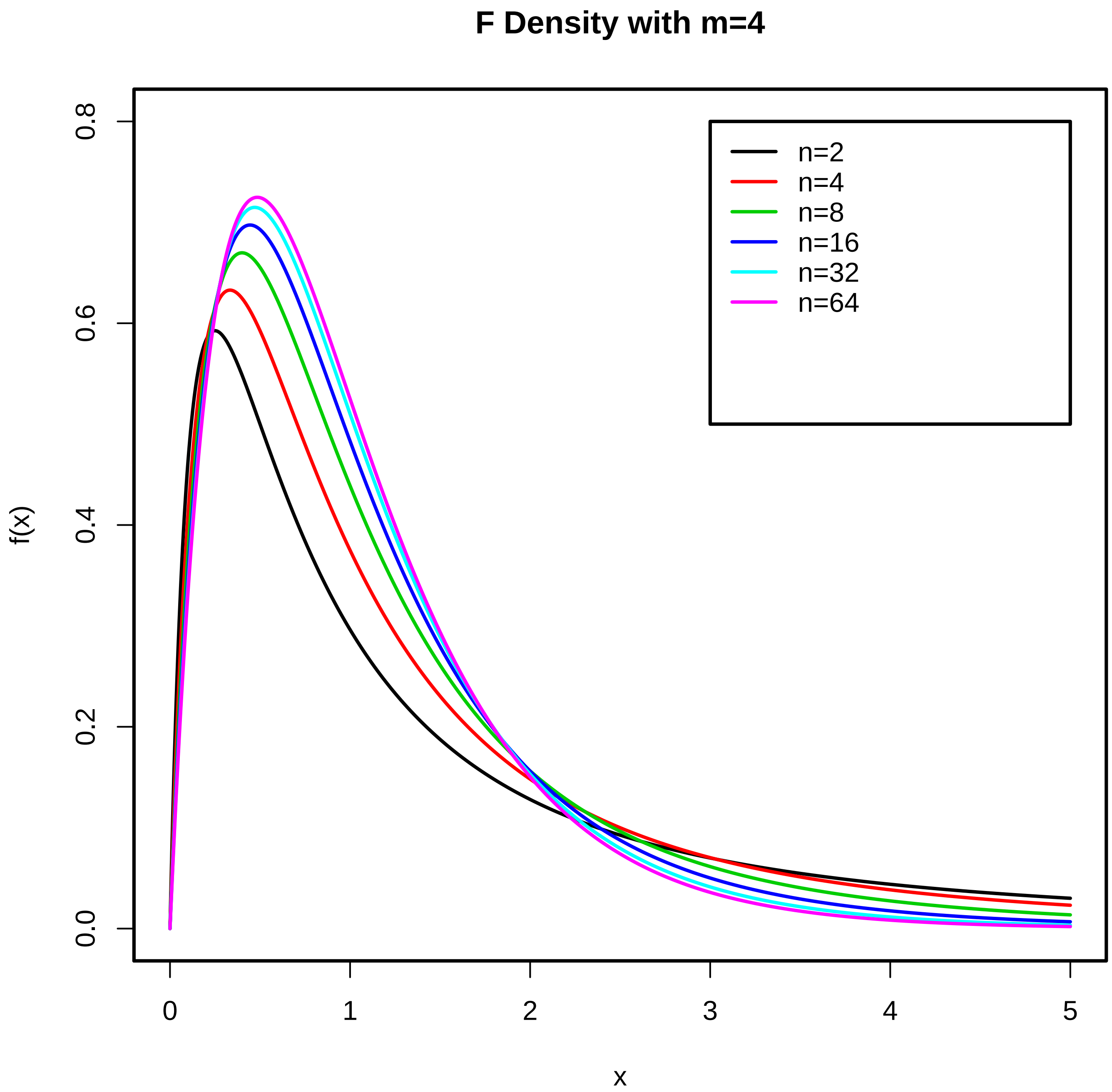
For $n > 2$,

$$E(W) = \frac{n}{n-2}.$$

Some Facts about the F Distribution

- If $W \sim F(m,n)$, then $1/W \sim F(n,m)$.
- If $T \sim t(n)$, then $T^2 \sim F(1,n)$.
- If $W \sim F(m,n)$, then $(m/n)W/(1+(m/n)W) \sim$ beta $(m/2, n/2)$.
- If $W_n \sim F(m,n)$, then $mW_n \xrightarrow{D} X$ where $X \sim \chi^2(m)$ as $n \to \infty$.


F Density with n=10
m=2
m=4
m=8
m=16
m=32
m=64
f(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5


F Density with m=4
n=2
n=4
n=8
n=16
n=32
n=64
f(x)
0.0
0.2
0.4
0.6
0.8
x
0
1
2
3
4
5

8 The Sampling Distribution of the Sample Mean and Sample Variance

A basic problem in statistics is inference for the mean and variance of a normal population. Suppose now that $X_1, \ldots, X_n$ form a random sample from a $N(\mu, \sigma^2)$ distribution. We will derive the following estimators for the population mean and variance:

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

We can obtain the mean and variance of these estimators using properties of normal random variables. For other aspects of inference, we need a more complete description of the sampling distribution of the sample mean and sample variance than just their means and variances.

Theorem: Suppose that $X_1, \ldots, X_n$ form a random sample from a $N(\mu, \sigma^2)$ distribution. Then the sample mean $\bar{X}$ and the sample variance S^2 are independent random variables satisfying

$$\bar{X} \sim N(\mu, \sigma^2/n) \quad \text{and} \quad \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1).$$

Remarks:

- The property of independence of the sample mean and sample variance characterizes the normal distribution. That means that if the sample mean and sample variance are independent, then $X_1, \ldots, X_n$ must be a random sample from the normal distribution.
- The proof of independence depends on the fact that $\bar{X}$ and the vector $(X_1 - X_2, X_1 + X_2 - 2X_3, \ldots, X_1 + \cdots + X_{n-1} - (n-1)X_n)$ are independent.
- The derivation of the sampling distribution of S^2 makes use of the fact that $(n-1)S^2/\sigma^2$ can be expressed as the sum of squares of $n-1$ independent $N(0,1)$ rvs.

8.1 The t Distribution and Random Samples from the Normal Distribution

We again consider a random sample $X_1, \ldots, X_n$ from a normal distribution. Define the following random variables:

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}, \quad \text{and} \quad U = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{\sigma^2}$$

We know that

$$Z \sim N(0,1) \quad \text{and} \qquad U \sim \chi^2(n-1).$$

We define the new rv:

$$T = \frac{Z}{\left(\dfrac{U}{n-1}\right)^{1/2}}$$

Then

$$T \sim t(n-1).$$

We then rewrite T in terms of $X_1, \ldots, X_n$:

$$
\begin{aligned}
T &= \frac{Z}{\left(\dfrac{U}{n-1}\right)^{1/2}} = \frac{\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\left(\dfrac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{(n-1)\sigma^2}\right)^{1/2}} \\
&= \frac{\sqrt{n}(\bar{X}_n - \mu)}{(S^2)^{1/2}} = \frac{\bar{X}_n - \mu}{S/\sqrt{n}}
\end{aligned}
$$

where

$$
S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2
$$

This result will be useful in deriving confidence intervals and hypothesis tests concerning the mean of a normal population.

