Theorem 5.4-2

If $X_1, X_2, \ldots, X_n$ are independent chi-square RVs with $r_1, r_2, \ldots, r_n$ degrees of freedom, respectively.

Then $Y = X_1 + X_2 + \cdots + X_n$ is $\chi^2(r_1 + r_2 + \cdots + r_n)$.

Proof.

For each X_i, its mgf is given by

$$M_{X_i}(t) = (1-2t)^{-r_i/2}, \qquad i = 1, \ldots, n. \qquad t < \frac{1}{2}.$$

Hence by the corollary for theorem 5.4-1, we have

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = (1-2t)^{-\frac{1}{2}(r_1 + \cdots + r_n)}$$

$$\Rightarrow Y \sim \chi^2(r_1 + \cdots + r_n)$$

Theorem 3.3-2

If the random variable X is $N(\mu, \sigma^2)$, $\sigma^2 > 0$, then the random variable

$$V = \frac{(X-\mu)^2}{\sigma^2} = Z^2 \text{ is } \chi^2(1).$$

Proof of theorem 3.3-2

Obviously, $Z = \dfrac{X-\mu}{\sigma}$ is $N(0,1)$. Since $V = Z^2$, the cdf for V is given by

$$G(v) = P(Z^2 \le v) = P(-\sqrt{v} \le Z \le \sqrt{v}). \qquad v \ge 0.$$

$$= \int_{-\sqrt{v}}^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 2\int_0^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$$= 2\int_0^{v} \frac{1}{\sqrt{2\pi}} e^{-y/2} d\sqrt{y} \qquad \leftarrow \text{Changing of variable with } z = \sqrt{y}$$

$$= \int_0^{v} \frac{1}{\sqrt{2\pi y}} e^{-y/2} dy.$$

By fundamental theorem of Calculus, its pdf is given by

$$g(v) = G'(v) = \frac{1}{\sqrt{2\pi v}} e^{-v/2} = \frac{1}{\sqrt{2\pi}} v^{1/2-1} e^{-v/2} = \frac{1}{2^{1/2}\Gamma(1/2)} v^{1/2-1} e^{-v/2}, \qquad 0 < v < \infty$$

Hence V follows $\chi^2(1)$.

The next two corollaries combine and extend the results of Theorems 3.3-2 and 5.4-2 and give one interpretation of degrees of freedom.

Corollary

Let $Z_1, \ldots, Z_n$ have standard normal distributions, $N(0,1)$. If $Z_1, \ldots, Z_n$ are independent, then $W = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi^2(n)$.

Proof. Since $Z_i \sim N(0,1)$, $Z_i^2 \sim \chi^2(1)$. $i = 1, 2, \ldots, n$.

By theorem 5.4-2 and 3.3-2, we have $W \sim \chi^2(n)$.

Corollary

If $X_1, \ldots, X_n$ are independent normal RVs with $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$. Then the distribution of

$$W = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2$$

is $\chi^2(n)$.

Proof. Since $X_i \sim N(\mu_i, \sigma_i^2)$, we have $\frac{X_i - \mu_i}{\sigma_i} \sim N(0,1)$.

$\Rightarrow Z_i = \frac{X_i - \mu_i}{\sigma_i} \sim N(0,1)$. And $Z_1, \ldots, Z_n$ are independent, by the above corollary, we have $W \sim \chi^2(n)$.

Chapter 5 Distributions of Functions of random variables

Section 5.5 Random function associated with Normal distribution

The similar idea works for summation of Gaussian RVs:

$$f(x_1, x_2, \ldots, x_n) = f(x_1) f(x_2) \cdots f(x_n)$$

Theorem 5.5-1

If $X_1, X_2, \ldots, X_n$ are n *mutually independent* normal variables with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \ \sigma_2^2, \ldots, \ \sigma_n^2$ respectively, then the linear function

$$Y = \sum_{i=1}^{n} c_i X_i$$

has the normal distribution

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

Proof.

By theorem 5.4-1, we have

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(c_i t) = \prod_{i=1}^{n} \exp\left(u_i c_i t + \frac{1}{2} \sigma_i^2 c_i^2 t^2\right) = \exp\left[\left(\sum_{i=1}^{n} \mu_i c_i\right) t + \frac{1}{2}\left(\sum_{i=1}^{n} c_i^2 \sigma_i^2\right) t^2\right].$$

Example 1 (Page 201)

Let X_1 and X_2 be the number of pounds of butterfat produced by two cows.

Assume that $X_1 \sim N(693.2, 22820), \quad X_2 \sim N(631.7, 19205)$.

And moreover, X_1 and X_2 are independent. What's the probability for $P(X_1 > X_2)$?

Solution :

Let $Y = X_1 - X_2$.

Then $Y \sim N(693.2 - 631.7, 22820 + 19205) = N(61.5, 42025)$

$$\Rightarrow P(X_1 > X_2) = P(Y > 0) = P\left(\frac{Y - 61.5}{\sqrt{42025}} > \frac{0 - 61.5}{\sqrt{42025}}\right) = P(Z > -0.3) = 0.6179$$

Define by Z

Corollary for Theorem 5.5-1

If $X_1, X_2, \ldots, X_n$ are observations of a RV of size n from the normal distribution $N(\mu, \sigma^2)$, then the distribution of the ***sample mean***

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

Proof. Let $c_i = \frac{1}{n},\ \mu_i = \mu,\ \sigma_i^2 = \sigma^2,\ i = 1, 2, \ldots, n.$

Example 2 (Page 201)

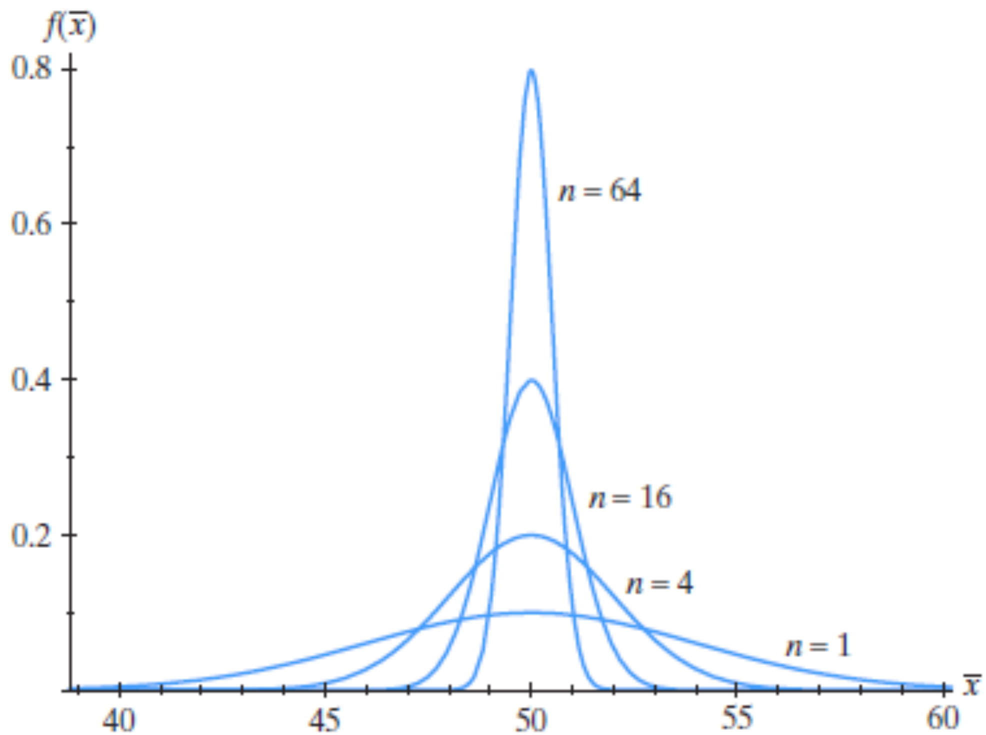
Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $N(50,16)$.

Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(50, \frac{16}{n})$.

To illustrate the effect of n, we graph of the pdf of $\overline{X}$ for $n = 1, 4, 16,$ and 64.

The sharper the peak, the more concentrated in a small interval centered at μ.

When $n \to \infty$, $\overline{X} \sim N(50, 0)$.



Theorem 5.5-2

Let $X_1, \ldots, X_n$ be observations of a random sample of size n from the normal distribution $N(\mu, \sigma^2)$. Then

the sample mean, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

and the sample variance, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

are independent, and moreover,

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2$$

is $\chi^2(n-1)$.

Proof. The independence part is omitted which can be referred to section 6.7 on Page 294. Now we show the second part of the theorem:

Firstly, define $W = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$.

Proof of theorem 5.5-2(c.n.t.)

Note that $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\sum_{i=1}^{n}\left(\frac{(X_i-\overline{X})+(\overline{X}-\mu)}{\sigma}\right)^2$

$$=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2+\sum_{i=1}^{n}\frac{2}{\sigma^2}(X_i-\overline{X})(\overline{X}-\mu)+\sum_{i=1}^{n}\left(\frac{\overline{X}-\mu}{\sigma}\right)^2.$$

We observe that

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

$$=\frac{2}{\sigma^2}(X_i-\overline{X})\left(\sum_{i=1}^{n}\overline{X}-n\mu\right)=\frac{2}{\sigma^2}(X_i-\overline{X})\left(\sum_{i=1}^{n}\overline{X}-\sum_{i=1}^{n}\overline{X}\right)=0.$$

Thus $W=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2+\sum_{i=1}^{n}\left(\frac{\overline{X}-\mu}{\sigma}\right)^2.$

Now let $S^2:=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ and $Z:=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}.$

Then $W=\frac{(n-1)S^2}{\sigma^2}+Z^2$

Proof of theorem 5.5-2(c.n.t.)

We notice that

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \Rightarrow Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \Rightarrow Z^2 \sim \chi^2(1).$$

and

$$W = \frac{(n-1)S^2}{\sigma^2} + Z^2 \sim \chi^2(n).$$

Note that S^2 and $\overline{X}$ are independent, then S^2 and Z^2 are also independent.

$$\frac{(n-1)S^2}{\sigma^2} = W - Z^2$$

$$\Rightarrow E(e^{t\frac{(n-1)S^2}{\sigma^2}}) = E(e^{t(W-Z^2)})$$

$$= E(e^{tW})E(e^{-tZ^2})$$

$$= (1-2t)^{-1/2}(1-2t)^{1/2} = (1-2t)^{-(n-1)/2},$$

which means that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$

Remark

Combining corollary 5.4-3 and theorem 5.5-2,

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$$

$$\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 \sim \chi^2(n-1).$$

Example 3(Page 204)

Let X_1, X_2, X_3, X_4 be a random sample of size 4 from $N(76.4,\ 383)$. Then we have

$$\sum_{i=1}^{4}\frac{(X_i-76.4)^2}{383} \sim \chi^2(4)$$

$$\sum_{i=1}^{4}\frac{(X_i-\overline{X})^2}{383} \sim \chi^2(3).$$

We now prove a theorem that is the basis for some of the most important inferences in statistics.

Case: When $T \sim t(1)$, we have

$$f(t) = \frac{1}{\pi}\frac{1}{1+t^2},$$

which also refers to

T has a ***Cauchy distribution.***

Theorem 5.5-3 (Student's *t* distribution)

Let $T = \frac{Z}{\sqrt{U/r}}$, where $Z \sim N(0,\ 1)$, $U \sim \chi^2(r)$, Assume that

Z and U are independent. Then T is said to have a ***student's t distribution***

with pdf $f(t)$,

$$f(t) = \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\Gamma(r/2)}\frac{1}{(1+t^2/r)^{(r+1)/2}}, \qquad t \in (-\infty, +\infty).$$

Briefly, $T \sim t(r)$, where r is the degree of freedom.

Proof.

- Since Z and U are independent, the joint pdf of Z and U is given by

$$g(z,u) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\frac{1}{\Gamma(r/2)2^{r/2}}u^{r/2-1}e^{-u/2}, \qquad z \in (-\infty, +\infty), u \in (0, +\infty).$$

- The *cdf* of T is given by

$$F(t) = P(T \le t) = P(\frac{Z}{\sqrt{U/r}} \le t) = P(Z \le t\sqrt{U/r}) = \int_0^{\infty}\int_{-\infty}^{t\sqrt{u/r}} g(z,u)\,dz\,du$$

Proof of theorem 5.5-3 (c.n.t.)

$$F(t)=\frac{1}{\sqrt{2\pi}\Gamma(r/2)}\int_0^\infty\int_{-\infty}^{t\sqrt{u/r}}\frac{e^{-z^2/2}}{2^{r/2}}u^{r/2-1}e^{-u/2}dzdu$$

$$=\frac{1}{\sqrt{\pi}\Gamma(r/2)}\int_0^\infty\int_{-\infty}^{t\sqrt{u/r}}\frac{e^{-z^2/2}}{2^{(r+1)/2}}u^{r/2-1}e^{-u/2}dzdu$$

$$=\frac{1}{\sqrt{\pi}\Gamma(r/2)}\int_0^\infty\left[\int_{-\infty}^{t\sqrt{u/r}}\frac{e^{-z^2/2}}{2^{(r+1)/2}}dz\right]u^{r/2-1}e^{-u/2}du$$

- Hence the pdf of T is given by

$$f(t)=F'(t)=\frac{d}{dt}\left\{\frac{1}{\sqrt{\pi}\Gamma(r/2)}\int_0^\infty\left[\int_{-\infty}^{t\sqrt{u/r}}\frac{e^{-z^2/2}}{2^{(r+1)/2}}dz\right]u^{r/2-1}e^{-u/2}du\right\}$$

$$=\frac{1}{\sqrt{\pi}\Gamma(r/2)}\frac{d}{dt}\left\{\int_0^\infty\left[\int_{-\infty}^{t\sqrt{u/r}}\frac{e^{-z^2/2}}{2^{(r+1)/2}}dz\right]u^{r/2-1}e^{-u/2}du\right\}$$

$$=\frac{1}{\sqrt{\pi}\Gamma(r/2)}\int_0^\infty\frac{d}{dt}\left\{\left[\int_{-\infty}^{t\sqrt{u/r}}\frac{e^{-z^2/2}}{2^{(r+1)/2}}dz\right]u^{r/2-1}e^{-u/2}\right\}du$$

$$=\frac{1}{\sqrt{\pi}\Gamma(r/2)}\frac{1}{\sqrt{u/r}}\int_0^\infty\frac{d}{d\left(t\sqrt{u/r}\right)}\left\{\left[\int_{-\infty}^{t\sqrt{u/r}}\frac{e^{-z^2/2}}{2^{(r+1)/2}}dz\right]u^{r/2-1}e^{-u/2}\right\}du$$

Proof of theorem 5.5-3 (c.n.t.)

$$f(t) = \frac{1}{\sqrt{\pi}\Gamma(r/2)} \cdot \sqrt{u/r} \cdot \int_0^{\infty} \frac{d}{d\left(t\sqrt{u/r}\right)} \left\{ \left[\int_{-\infty}^{t\sqrt{u/r}} \frac{e^{-z^2/2}}{2^{(r+1)/2}} dz \right] u^{r/2-1} e^{-u/2} \right\} du$$

$$= \frac{1}{\sqrt{\pi}\Gamma(r/2)} \cdot \sqrt{u/r} \cdot \int_0^{\infty} \left[\frac{e^{-t^2(u/r)/2}}{2^{(r+1)/2}} \right] u^{r/2-1} e^{-u/2} du$$

$$= \frac{1}{\sqrt{\pi r}\Gamma(r/2)} \cdot \int_0^{\infty} \frac{u^{(r+1)/2-1}}{2^{(r+1)/2}} e^{-t^2(u/r)/2-u/2} du$$

$$= \frac{1}{\sqrt{\pi r}\Gamma(r/2)} \cdot \int_0^{\infty} \frac{u^{(r+1)/2-1}}{2^{(r+1)/2}} e^{-(u/2)\left(1+t^2/r\right)} du.$$

$$= \frac{1}{\sqrt{\pi r}\Gamma(r/2)} \cdot \frac{1}{(1+t^2/r)^{(r+1)/2-1}} \cdot \frac{1}{(1+t^2/r)} \int_0^{\infty} \frac{y^{(r+1)/2-1}}{2^{(r+1)/2}} e^{-y/2} du.$$

$\leftarrow$ Changing of variable with $y = (1+t^2/r)u$

$$= \frac{1}{\sqrt{\pi r}\Gamma(r/2)} \cdot \frac{1}{(1+t^2/r)^{(r+1)/2}} \int_0^{\infty} \frac{y^{(r+1)/2-1}}{2^{(r+1)/2}} e^{-y/2} du.$$

$$= \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\Gamma(r/2)} \cdot \frac{1}{(1+t^2/r)^{(r+1)/2}} \int_0^{\infty} \frac{y^{(r+1)/2-1}}{\Gamma((r+1)/2)2^{(r+1)/2}} e^{-y/2} du.$$

Proof of theorem 5.5-3 (c.n.t.)

Note that the last integral is equal to one because the integrand is the pdf of a $\chi^2(r+1)$ RV.

Hence the pdf is given by

$$f(t) = \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\,\Gamma(r/2)} \cdot \frac{1}{(1+t^2/r)^{(r+1)/2}}, \qquad -\infty < t < +\infty.$$

$\square$.

Comparison

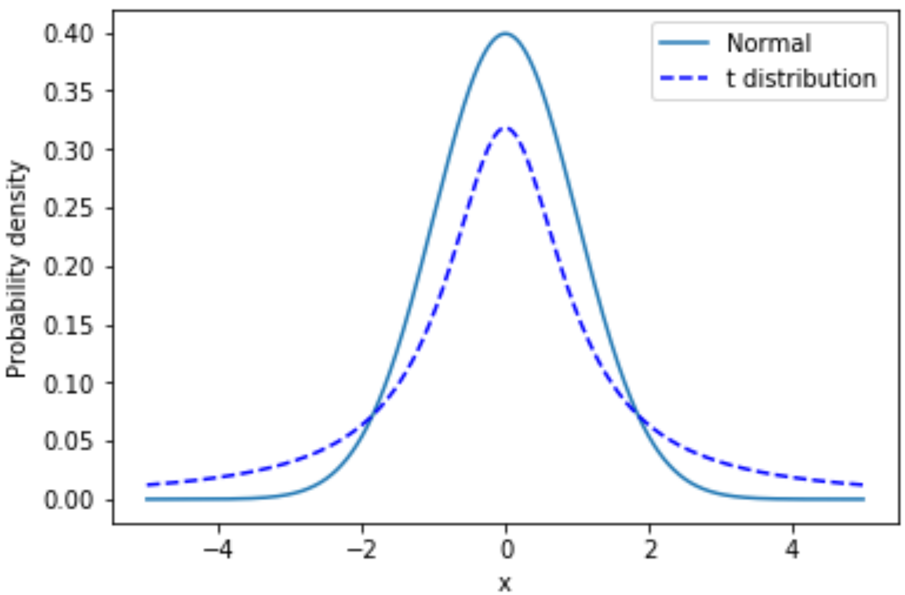
Student's t distribution is a ***heavy tailed distribution***, while normal distribution is a ***light tailed distribution***.

Standard normal distribution:

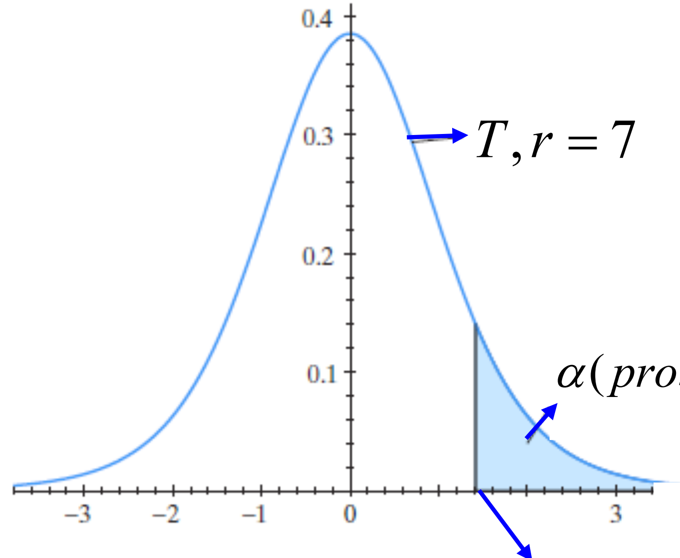
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-1/2x^2}.$$

Student's t distribution with $r = 1$:

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$



Furthermore, the right-tail probabilities of size α are denoted by $t_\alpha(r)$.



Construction of student's t distribution

Assume $X_1, X_2, \ldots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$.

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

$$U = \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{\sigma} \right)^2 \sim \chi^2(n-1)$$

$$= \frac{n-1}{\sigma^2} S^2$$

$$T = \frac{\dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}}{\sqrt{\left(\dfrac{n-1}{\sigma^2} S^2 \right) \Big/ (n-1)}} = \frac{\overline{X} - \mu}{S / \sqrt{n}} \sim t(n-1).$$

Notice that

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

$$T = \frac{Z}{\sqrt{U / r}} \sim t(r).$$

Section 5.6 Central limit theorem (CLT)

➢ Interpretation：

If $X_1, X_2, \ldots, X_n$ are observations of a RV of size n from the normal distribution $N(\mu, \sigma^2)$, then

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1).$$

The distribution of Z depends on we are dealing with sequences of distributions.

Theorem 5.6-1 (Central Limit Theorem))

If $\overline{X}$ is the sample mean of a random sample $X_1, X_2, \ldots, X_n$ of size n from a distribution with finite mean μ and finite positive variance σ^2, then the limit of the distribution of

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{\sqrt{n}}{\sigma}(\overline{X} - \mu) \sim N(0,1) \qquad \text{as } n \to \infty$$

equaivalently, $\overline{X} \sim N(\mu, \frac{1}{n}\sigma^2)$ when $n \to \infty$. $\Leftrightarrow \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$ when $n \to \infty$.

- Practical use of CLT is to approximate the cdf of $\bar{X}$ or $\sum_{i=0}^{n} X_i$.

Recall that if $Y \sim N(\mu, \sigma^2)$, then

$$P(a < Y < b) = P\left(\frac{a-\mu}{\sigma} < \frac{Y-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right),$$

$\Phi()$ is the *cdf* for $N(0,1)$.

Example 1

Let $\bar{X}$ be a sample mean of a random sample of size 25 froma distribution of size 25 from a distribution with mean 15 and variance 4.

Question : Compute $P(14.4 < \bar{X} < 15.6)$ approximately.

Solution :

By CLT, approximately, $\bar{X} \sim N(15, \frac{4}{25})$.

Approximately, $\frac{\bar{X}-15}{0.4} \sim N(0,1)$

Hence $P\left(\frac{14.4-15}{0.4} < \frac{\bar{X}-15}{0.4} < \frac{15.6-15}{0.4}\right)$

$$= \Phi(1.5) - \Phi(-1.5) = 0.9332 - (1-0.9332) = 0.8664.$$

Example 2

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from U(0, 1)

with pdf $f(x) = 1, x \in (0,1)$. For $U(0,1)$, $\mu = \dfrac{1}{2}$, $\sigma^2 = \dfrac{1}{12}$. Let $Y = \sum_{i=1}^{n} X_i$.

By CLT, $Y \sim N(\frac{n}{2}, \frac{n}{12})$ for ***large n.***

Since $X_1, X_2, \ldots, X_n$ are drawn from $U(0,1)$,

it's possible to derive the pdf of Y, $g(y)$.

Let's compare $N(\frac{n}{2}, \frac{n}{12})$ with the true pdf for $n = 2$ and $n = 4$: