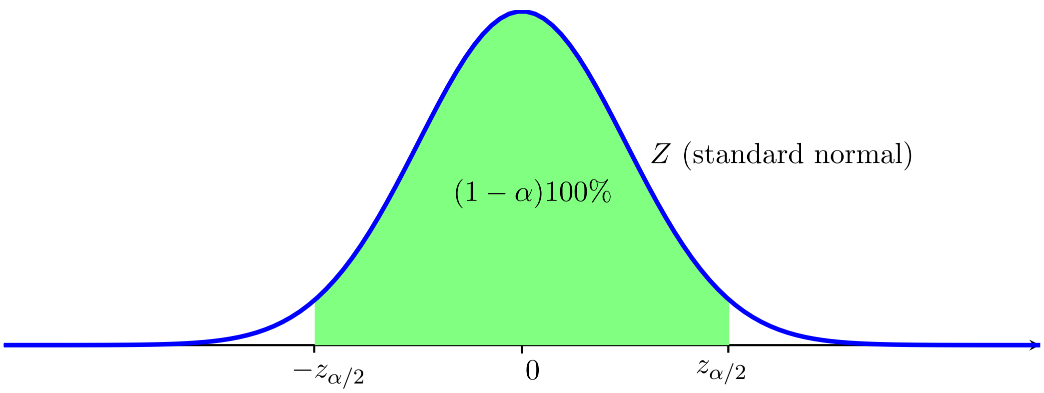
1. The t-distribution

Review of §5.3. Suppose $X \sim \text{Normal}(\mu, \sigma^2)$ with **known** σ and unknown μ. The maximum likelihood estimator for μ is the sample mean $\widehat{\mu} = \overline{X}$.

By **CLT**, we know that the **sample mean** $\overline{X} \sim \text{Normal}(\mu, \sigma^2/n)$. Then,

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



Then,

$$-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$$

We solve μ,

$$\mu = \overline{x} \pm z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right)$$

The $100(1-\alpha)\%$ **confidence interval** for μ

$$\overline{x} - z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right) \leq \mu \leq \overline{x} + z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right)$$

Suppose $X \sim \text{Normal}(\mu, \sigma^2)$ with **unknown** σ and unknown μ.

In §5.4, we calculated an **unbiased** estimate and estimator for the population variance σ^2, called **sample variance**:

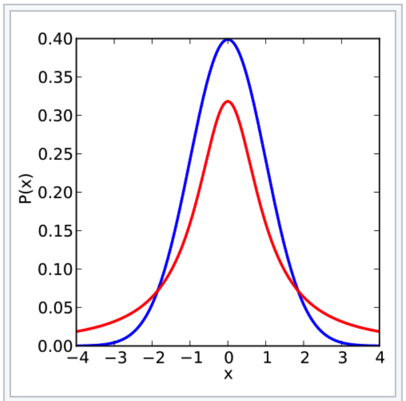
$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} \text{ and } S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

However, the distribution

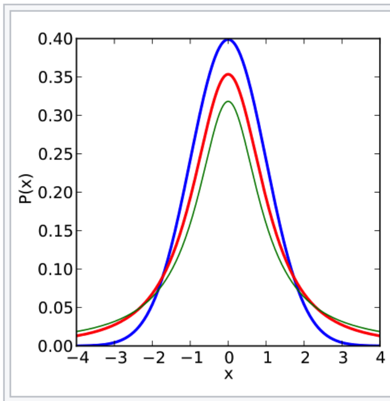
$$T_{n-1} = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

is **NOT** the standard normal distribution. This distribution T_{n-1} is called the **(Student's) t-distribution** with $n-1$ **d**egrees of **f**reedom (**df**).

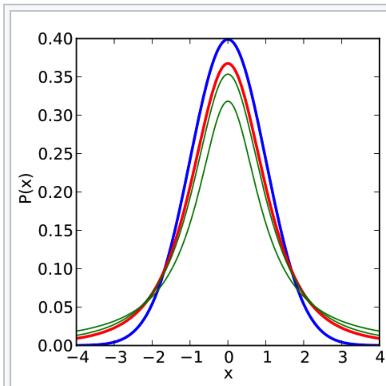
As n increases, T_{n-1} approaches the standard normal distribution Z.



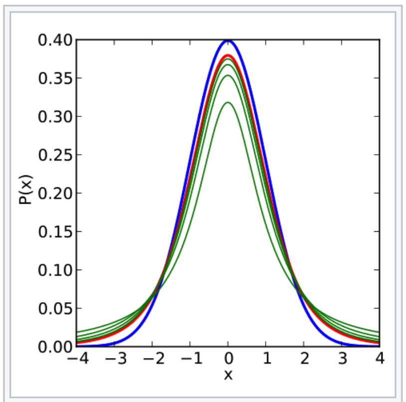
1 degree of freedom



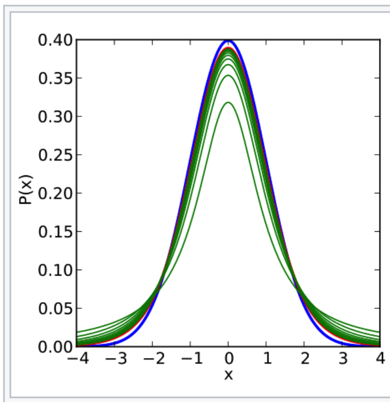
2 degrees of freedom



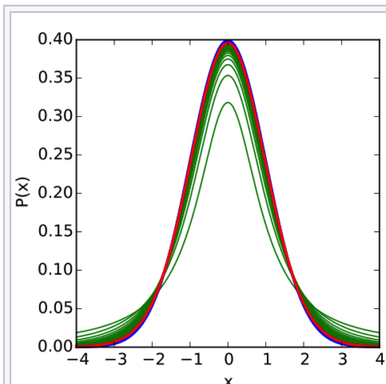
3 degrees of freedom



5 degrees of freedom

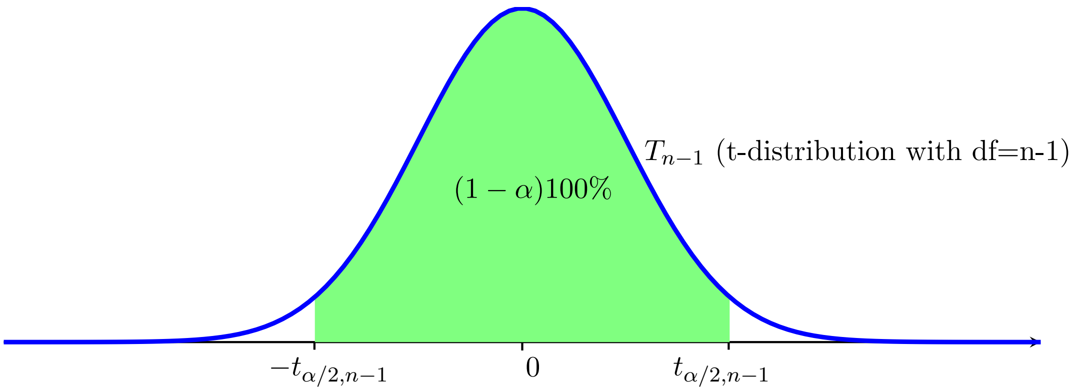


10 degrees of freedom



30 degrees of freedom

2. The confidence interval.



By t-distribution with degree of freedom **df**$=n-1$, the $100(1-\alpha)\%$ **confidence interval** for μ is

$$\overline{x} - t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right) \leq \mu \leq \overline{x} + t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right)$$

Similarly as $z_{\alpha/2} = \textbf{invNorm}(1-\alpha/2, 0, 1)$, we can use calculator to compute $t_{\alpha/2,n-1}$ as

$$t_{\alpha/2,n-1} = \textbf{invT}(1-\alpha/2, n-1)$$

Example 1. The grades in a test are known to follow a normal distribution, but the mean and variance are unknown. A random sample of 6 students produced the following scores: 68, 93, 82, 72, 65, 70. Use the t-distribution to find a 90% two-sided confidence interval for the mean score.

The Sample size $n=6$. The value $t_{\alpha/2,n-1} = t_{0.05,5} = \textbf{invT}(0.95, 5) \approx 2.015$.

The sample mean $\bar{x} = \dfrac{68+93+82+72+65+70}{6} = 75$.

The sample variance $s^2 = \dfrac{7^2+18^2+7^2+3^2+10^2+5^2}{5} = 111.2$. So $s \approx 10.55$

So, the 90% two-sided confidence interval for the mean score is $\mu = 75 \pm 2.015(\dfrac{10.55}{\sqrt{6}})$.

$66.321 \leq \mu \leq 83.679$.

Another way to calculate the confidence interval is by

$$\textbf{STAT} \rightarrow \textbf{TESTS} \rightarrow \textbf{TInterval}$$

If we have data, use **Data**, which are put in **L1:**

$$\textbf{STAT} \rightarrow \textbf{Edit}$$

If we have the sample average, use **Stats**.

Example 2. The grades in a test are known to follow a normal distribution, but the mean and variance are unknown. A random sample of 6 students produced the scores with mean 75 and standard deviation 10.55. Use the t-distribution to find a 90% two-sided confidence interval for the mean score.

3. One Sample t-Test

In §6.2, we learned hypothesis testing on the unknown mean of a normal distribution with **known** variance.

Now, suppose the variance is **unknown**, we use the sample variance S^2 and the t-distribution

$$T_{n-1} = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

To Test the Null Hypothesis $H_0 : \mu = \mu_0$ at the level of significance α, we define the **test statistic:**

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

1. (Right-sided) $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$.
Decision Rule: Reject H_0 if $t > t_{\alpha,n-1}$.

2. (Left-sided) $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$.
Decision Rule: Reject H_0 if $t < -t_{\alpha,n-1}$.

3. (Two-sided) $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.
Decision Rule: Reject H_0 if $t < -t_{\alpha/2,n-1}$ or $t > t_{\alpha/2,n-1}$.

Recall that $t_{\alpha,n-1}$ is calculated by

$$t_{\alpha,n-1} = \mathbf{invT}(1 - \alpha, n - 1)$$

Example 3. Suppose you want to test $H_0 : \mu = 12$ versus $H_1 : \mu > 12$ at the level of significance 0.05. The data is $\bar{y} = 13.2$ and $s = 2.5$ from a sample of 25.

Degree of freedom $df = n - 1 = 24$

$t_{\alpha,n-1} = \mathbf{invT}(0.95, 24) \approx 1.71$

The test statistics $t = \dfrac{13.2 - 12}{2.5/\sqrt{25}} = 2.4 > t_{\alpha,n-1}$. So, reject H_0.

We can also do this by Calculator:

STAT → TESTS → T-Test

Example 4. The grades in a test are known to follow a normal distribution, but the mean and variance are unknown. A random sample of 6 students produced the following scores: 68, 93, 82, 72, 65, 70. Suppose we want to test $H_0 : \mu = 67$ versus $H_1 : \mu > 67$ at the level of significance 0.05.

We can also do this by Calculator:

STAT → TESTS → T-Test

For this question, we need to use Data.

$t_{\alpha,n-1} = \mathbf{invT}(0.95, 5) \approx 2.02$