

Math 362: Mathematical Statistics II

Le Chen

le.chen@emory.edu

Emory University
Atlanta, GA

Last updated on May 3, 2021

2021 Spring

Chapter 7. Inference Based on The Normal Distribution

§ 7.1 Introduction

§ 7.2 Comparing $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.3 Deriving the Distribution of $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.4 Drawing Inferences About μ

§ 7.5 Drawing Inferences About σ^2

Chapter 7. Inference Based on The Normal Distribution

§ 7.1 Introduction


§ 7.2 Comparing $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.3 Deriving the Distribution of $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.4 Drawing Inferences About μ


§ 7.5 Drawing Inferences About σ^2



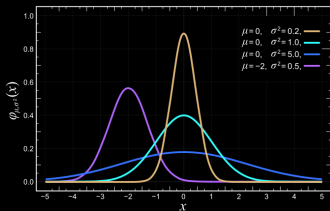
Carl Friedrich Gauss 
discovered the normal
distribution in 1809 as a way to
rationalize the [method of least
squares](#).

(1777-1855)



Marquis de Laplace proved 
the [central limit theorem](#) in
1810, consolidating the
importance of the normal
distribution in statistics.

(1749-1827)



Notation	$\mathcal{N}(\mu, \sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
CDF	$\frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) \right]$
Quantile	$\mu + \sigma\sqrt{2} \operatorname{erf}^{-1}(2p - 1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2
MAD	$\sigma\sqrt{2/\pi}$
Skewness	0
Ex. kurtosis	0
Entropy	$\frac{1}{2} \log(2\pi e\sigma^2)$
MGF	$\exp(\mu t + \sigma^2 t^2/2)$
CF	$\exp(i\mu t - \sigma^2 t^2/2)$
Fisher information	$\mathcal{I}(\mu, \sigma) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}$ $\mathcal{I}(\mu, \sigma^2) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) \end{pmatrix}$
Kullback-Leibler divergence	$D_{\text{KL}}(\mathcal{N}_0 \parallel \mathcal{N}_1) = \frac{1}{2} \left\{ \left(\frac{\sigma_0}{\sigma_1} \right)^2 + \frac{(\mu_1 - \mu_0)^2}{\sigma_1^2} - 1 + 2 \ln \frac{\sigma_1}{\sigma_0} \right\}$

https://en.wikipedia.org/wiki/Normal_distribution

Test for normal parameters (one sample test)

Let Y_1, \dots, Y_n be a random sample from $N(\mu, \sigma^2)$.

Prob. 1 Find a test statistic Λ in order to test $H_0 : \mu = \mu_0$ v.s. $H_1 : \mu \neq \mu_0$.

When σ^2 is known: $\Lambda = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$

When σ^2 is unknown: $\Lambda = ? \quad \Lambda \stackrel{?}{=} \frac{\bar{Y} - \mu_0}{s/\sqrt{n}} \sim ?$

Prob. 2 Find a test statistic Λ in order to test $H_0 : \sigma^2 = \sigma_0^2$ v.s. $H_1 : \sigma^2 \neq \sigma_0^2$.

Prob. 1 Find a test statistic for $H_0 : \mu = \mu_0$ v.s. $H_1 : \mu \neq \mu_0$, with σ^2 unknown

Sol. Composite-vs-composite test with:

$$\omega = \{(\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\}$$

$$\Omega = \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$$

The MLE under the two spaces are:

$$\omega_{\theta} = (\mu_{\theta}, \sigma_{\theta}^2) : \quad \mu_{\theta} = \mu_0 \quad \text{and} \quad \sigma_{\theta}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mu_0)^2 \quad (\text{Under } \omega)$$

$$\Omega_{\theta} = (\mu_{\theta}, \sigma_{\theta}^2) : \quad \mu_{\theta} = \bar{y} \quad \text{and} \quad \sigma_{\theta}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (\text{Under } \Omega)$$

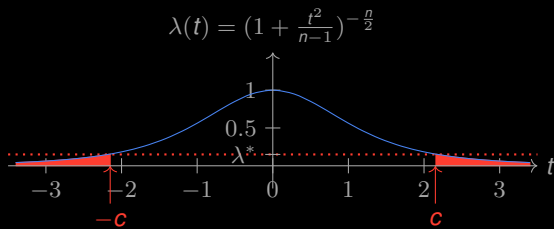
$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n} \exp \left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \mu}{\sigma} \right)^2 \right)$$

$$L(\omega_e) = \dots = \left[\frac{ne^{-1}}{2\pi \sum_{i=1}^n (y_i - \mu_0)^2} \right]^{n/2}$$

$$L(\Omega_e) = \dots = \left[\frac{ne^{-1}}{2\pi \sum_{i=1}^n (y_i - \bar{y})^2} \right]^{n/2}$$

Hence,

$$\begin{aligned}
\lambda &= \frac{L(\omega_{\theta})}{L(\Omega_{\theta})} = \left[\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \mu_0)^2} \right]^{n/2} = \cdots = \left[1 + \frac{n(\bar{y} - \mu_0)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right]^{-n/2} \\
&= \left[1 + \frac{1}{n-1} \left(\frac{\bar{y} - \mu_0}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} / \sqrt{n}} \right)^2 \right]^{-n/2} \\
&= \left[1 + \frac{1}{n-1} \left(\frac{\bar{y} - \mu_0}{s / \sqrt{n}} \right)^2 \right]^{-n/2} \\
&= \left[1 + \frac{t^2}{n-1} \right]^{-n/2}, \quad t = \frac{\bar{y} - \mu_0}{s / \sqrt{n}}
\end{aligned}$$



$$\lambda \in (0, \lambda^*] \quad \Leftrightarrow \quad |t| \geq c.$$

Finally, the test statistic is

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}}$$

$$\text{with } \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad \text{and} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

The critical region takes the form: $|t| \geq c$.

Question: Find the exact distribution of T .

Prob. 2 Find a test statistic for $H_0 : \sigma^2 = \sigma_0^2$ v.s. $H_1 : \sigma^2 \neq \sigma_0^2$, with μ unknown

Sol. Composite-vs-composite test with:

$$\begin{aligned}\omega &= \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\} \\ \Omega &= \{(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}\end{aligned}$$

The MLE under the two spaces are:

$$\omega_{\theta} = (\mu_{\theta}, \sigma_{\theta}^2) : \quad \mu_{\theta} = \bar{y} \quad \text{and} \quad \sigma_{\theta}^2 = \sigma_0^2 \quad (\text{Under } \omega)$$

$$\Omega_{\theta} = (\mu_{\theta}, \sigma_{\theta}^2) : \quad \mu_{\theta} = \bar{y} \quad \text{and} \quad \sigma_{\theta}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (\text{Under } \Omega)$$

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n} \exp \left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \mu}{\sigma} \right)^2 \right)$$

$$L(\omega_e) = (2\pi\sigma^2)^{-n} \exp \left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \bar{y}}{\sigma_0} \right)^2 \right)$$

$$L(\Omega_e) = \dots = \left[\frac{ne^{-1}}{2\pi \sum_{i=1}^n (y_i - \bar{y})^2} \right]^{n/2}$$

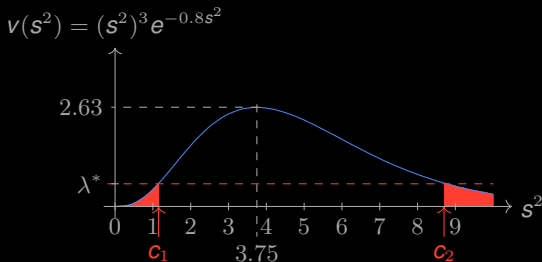
Hence,

$$\begin{aligned}
\lambda &= \frac{L(\omega_{\theta})}{L(\Omega_{\theta})} = \left[\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n\sigma_0^2} \right]^{n/2} \exp \left(-\frac{1}{2} \sum_{i=1}^n \left(\frac{y_i - \bar{y}}{\sigma_0} \right)^2 + \frac{n}{2} \right) \\
&= \left[\frac{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}{\frac{n}{n-1} \sigma_0^2} \right]^{n/2} \exp \left(-\frac{n-1}{2\sigma_0^2} \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 + \frac{n}{2} \right) \\
&= \left[\frac{\mathbf{s}^2}{\frac{n}{n-1} \sigma_0^2} \right]^{n/2} \exp \left(-\frac{n-1}{2\sigma_0^2} \mathbf{s}^2 + \frac{n}{2} \right)
\end{aligned}$$

\Downarrow

$$\lambda(\mathbf{s}^2) = \left[\frac{\mathbf{s}^2}{\frac{n}{n-1} \sigma_0^2} \right]^{n/2} \exp \left(-\frac{n-1}{2\sigma_0^2} \mathbf{s}^2 + \frac{n}{2} \right) \iff \nu(\mathbf{s}^2) = (\mathbf{s}^2)^{\frac{n}{2}} e^{-\lambda \mathbf{s}^2}$$

By setting $n = 6$ and $\lambda = 0.8$, we see ...



This suggests that the critical region should be of the form in terms of S^2 :

$$(0, c_1) \cup (c_2, \infty)$$

For convenience, we put $\alpha/2$ mass on each tails of S^2 :

Find c_1 and c_2 such that

$$\int_0^{c_1} f_{S^2}(z) dz = \int_{c_2}^{\infty} f_{S^2}(z) dz = \frac{\alpha}{2}.$$

Finally, the test statistic is

$$\boxed{S^2 = \frac{1}{n-1} \sum_{i=1}^n \left(Y_i - \bar{Y} \right)^2} \quad \text{with} \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

Question: Find the exact distribution of S^2 .

Chapter 7. Inference Based on The Normal Distribution

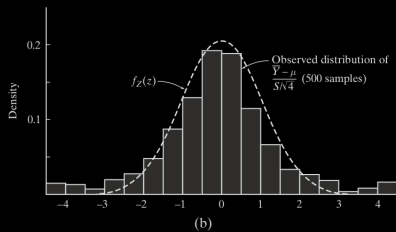
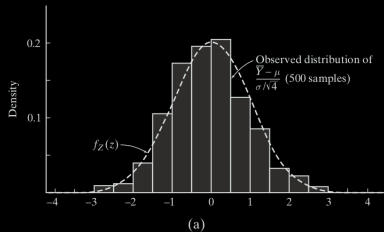
§ 7.1 Introduction

§ 7.2 Comparing $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.3 Deriving the Distribution of $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.4 Drawing Inferences About μ

§ 7.5 Drawing Inferences About σ^2





Statistician William Sealy Gosset, known as "Student"



Ref. Student's t distribution comes from William Sealy Gosset's 1908 paper in *Biometrika* under the pseudonym "Student".

Gosset worked at the Guinness Brewery in Dublin, Ireland, and was interested in the problems of small samples – for example, the chemical properties of barley where sample sizes might be as few as 3.

- V1 One version of the origin of the pseudonym is that Gosset's employer preferred staff to use pen names when publishing scientific papers instead of their real name, so he used the name "Student" to hide his identity.
- V2 Another version is that Guinness did not want their competitors to know that they were using the t -test to determine the quality of raw material

Chapter 7. Inference Based on The Normal Distribution

§ 7.1 Introduction

§ 7.2 Comparing $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.3 Deriving the Distribution of $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.4 Drawing Inferences About μ

§ 7.5 Drawing Inferences About σ^2

Def. **Sampling distributions**

Distributions of functions of random sample
statistics / estimators

E.g. A random sample of size n from $N(\mu, \sigma^2)$ with σ^2 known.

Sample mean $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \sim N(\mu, \sigma^2/n)$

Aim: Determine distributions for

Sample variance $S^2 := \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$ | *Chi square distr.*

$T := \frac{\bar{Y} - \mu}{S/\sqrt{n}}$ | *Student t distr.*

$\frac{S_1^2}{\sigma_1^2} / \frac{S_2^2}{\sigma_2^2}$ | *F distr.*

Thm 7.3.1. Let $U = \sum_{j=1}^m Z_j^2$, where Z_j are independent $N(0, 1)$ normal r.v.s. Then

$$U \sim \text{Gamma}(\text{shape}=m/2, \text{rate}=1/2).$$

namely,

$$f_U(u) = \frac{1}{2^{m/2}\Gamma(m/2)} u^{\frac{m}{2}-1} e^{-u/2}, \quad u \geq 0.$$

Def 7.3.1. U in Thm 7.3.1 is called **chi square distribution** with m dgs of freedom.

Proof. We first consider the case when $m = 1$. In this case,

$$\begin{aligned}
 F_{Z^2}(u) &= \mathbb{P}(Z^2 \leq u) \\
 &= \mathbb{P}(-\sqrt{u} \leq Z \leq \sqrt{u}) \\
 &= 2\mathbb{P}(0 \leq Z \leq \sqrt{u}) \\
 &= \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{u}} e^{-z^2/2} dz
 \end{aligned}$$

Differentiating both sides of the above eq. in order to obtain the pdf:

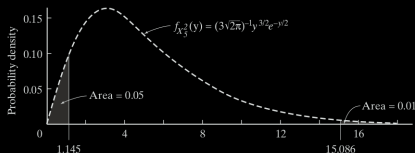
$$\begin{aligned}
 f_{Z^2}(u) &= \frac{d}{du} F_{Z^2}(u) \\
 &= \frac{2}{\sqrt{2\pi}} \frac{1}{2\sqrt{u}} e^{-u/2} \\
 &= \frac{1}{\sqrt{2}\Gamma(1/2)} u^{(1/2)-1} e^{-u/2},
 \end{aligned}$$

which is the pdf of a gamma distribution with $r = \lambda = 1/2$.

Then adding m independent copies of gamma distributions gives another gamma distribution with $r = m/2$ and $\lambda = 1/2$ (See Theorem 4.6.4). \square

Chi Square Table

df	p							
	.01	.025	.05	.10	.90	.95	.975	.99
1	0.000157	0.000982	0.00393	0.0158	2.706	3.841	5.024	6.635
2	0.0201	0.0506	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.336	26.217



$$\mathbb{P}(\chi_5^2 \leq 1.145) = 0.05 \iff \chi_{0.05,5}^2 = 1.145$$

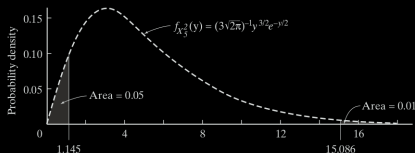
$$\mathbb{P}(\chi_5^2 \leq 15.086) = 0.99 \iff \chi_{0.99,5}^2 = 15.086$$

```
1 > pchisq(1.145, df = 5)
2 [1] 0.04995622
3 > pchisq(15.086, df = 5)
4 [1] 0.9899989
```

```
1 > qchisq(0.05, df = 5)
2 [1] 1.145476
3 > qchisq(0.99, df = 5)
4 [1] 15.08627
```


Chi Square Table

df	p							
	.01	.025	.05	.10	.90	.95	.975	.99
1	0.000157	0.000982	0.00393	0.0158	2.706	3.841	5.024	6.635
2	0.0201	0.0506	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.336	26.217



$$\mathbb{P}(\chi_5^2 \leq 1.145) = 0.05 \iff \chi_{0.05,5}^2 = 1.145$$

$$\mathbb{P}(\chi_5^2 \leq 15.086) = 0.99 \iff \chi_{0.99,5}^2 = 15.086$$

1	> scipy.stats.chi2.cdf(1.145, 5)	1	> scipy.stats.chi2.ppf(0.05, 5)
2	[1]: 0.04995622155207728	2	[1]: 1.1454762260617692
3	> scipy.stats.chi2.cdf(15.086, 5)	3	> scipy.stats.chi2.ppf(0.99, 5)
4	[1]: 0.9899988752378142	4	[1]: 15.08627246938899

Thm 7.3.2. Let Y_1, \dots, Y_n be a random sample from $N(\mu, \sigma^2)$. Then

(a) S^2 and \bar{Y} are independent.

$$(b) \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2 \sim \text{Chi Square}(n-1).$$

Proof. We will prove the case $n = 2$.

$$\bar{Y} = \frac{Y_1 + Y_2}{2}, \quad Y_1 - \bar{Y} = \frac{Y_1 - Y_2}{2}, \quad Y_2 - \bar{Y} = \frac{Y_2 - Y_1}{2}$$

$$S^2 = \dots = \frac{1}{2} (Y_1 - Y_2)^2$$

(a) It is equivalent to show $Y_1 + Y_2 \perp Y_1 - Y_2$. Since they are normal, it suffices to show that

$$\mathbb{E}[(Y_1 + Y_2)(Y_1 - Y_2)] = \mathbb{E}[Y_1 + Y_2]\mathbb{E}[Y_1 - Y_2]$$

$$(b) \frac{(n-1)S^2}{\sigma^2} = \left(\frac{Y_1 - Y_2}{\sqrt{2}\sigma} \right)^2 \text{ and } \frac{Y_1 - Y_2}{\sqrt{2}\sigma} \sim N(0, 1) \dots$$

□

Def 7.3.2. If $U \sim \text{Chi Square}(n)$ and $V \sim \text{Chi Square}(m)$, and $U \perp V$, then

$$F := \frac{V/m}{U/n}$$

follows the **(Snedecor's) F distribution** with m and n degrees of freedom.

Thm 7.3.3. Let $F_{m,n} = \frac{V/m}{U/n}$ be an F r.v. with m and n degrees of freedom. Then

$$f_{F_{m,n}}(w) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma(m/2)\Gamma(n/2)} \times \frac{w^{m/2-1}}{(n + mw)^{(m+n)/2}}, \quad w \geq 0$$

Equivalently,

$$f_{F_{m,n}}(w) = B(m/2, n/2)^{-1} \left(\frac{m}{n}\right)^{\frac{m}{2}} w^{\frac{m}{2}-1} \left(1 + \frac{m}{n}w\right)^{-\frac{m+n}{2}}$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

Recall

Thm 3.8.4 Let X and Y be independent continuous random variables, with pdf $f_X(x)$ and $f_Y(y)$, respectively.

Assume that X is zero for at most a set of isolated points.

Then $W = Y/X$ follows a distribution with pdf:

$$f_W(w) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(wx) dx.$$

Thm 3.8.2 Suppose X is a continuous random variable and $a \neq 0$.

Then $Y = aX + b$ follows a distribution with pdf:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Proof. Let us first find the pdf for $W := V/U$. By Theorem 7.3.1,

$$f_V(v) = \frac{1}{2^{m/2}\Gamma(m/2)} v^{(m/2)-1} e^{-v/2},$$

$$f_U(u) = \frac{1}{2^{n/2}\Gamma(n/2)} u^{(n/2)-1} e^{-u/2}.$$

Then by Theorem 3.8.4, we see that the pdf of W is

$$\begin{aligned} f_W(w) &= \int_{-\infty}^{\infty} |u| f_U(u) f_V(uw) du \\ &= \int_0^{\infty} u \frac{1}{2^{n/2}\Gamma(n/2)} u^{(n/2)-1} e^{-u/2} \frac{1}{2^{m/2}\Gamma(m/2)} (uw)^{(m/2)-1} e^{-uw/2} du \\ &= \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} w^{(m/2)-1} \int_0^{\infty} u^{\frac{n+m}{2}-1} e^{-\frac{1+w}{2}u} du \end{aligned}$$

Then by the change of variables, $y = \frac{1+w}{2}u$, we see that

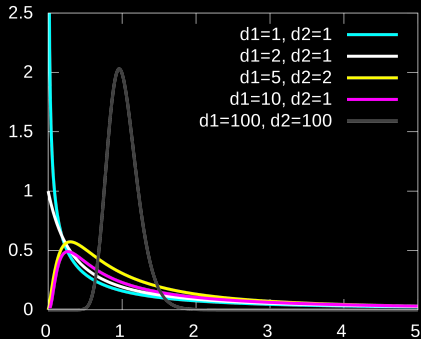
$$\begin{aligned} f_W(w) &= \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} w^{(m/2)-1} \left(\frac{2}{1+w} \right)^{\frac{n+m}{2}} \int_0^\infty y^{\frac{n+m}{2}-1} e^{-y} dy \\ &= \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} w^{(m/2)-1} \left(\frac{2}{1+w} \right)^{\frac{n+m}{2}} \Gamma\left(\frac{n+m}{2}\right) \end{aligned}$$

where the last equality is due to the definition of the Gamma function.

Finally, by Theorem 3.8.2, we see that $F = \frac{V/m}{U/n} = \frac{n}{m}W$ follows a distribution with pdf

$$\begin{aligned} f_F(y) &= \frac{m}{n} f_W\left(\frac{m}{n}y\right) \\ &= \frac{m}{n} \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} \left(\frac{m}{n}y\right)^{(m/2)-1} \left(\frac{2}{1+\frac{m}{n}y}\right)^{\frac{n+m}{2}} \Gamma\left(\frac{n+m}{2}\right) \\ &= \cdots \quad y \geq 0. \end{aligned}$$

□

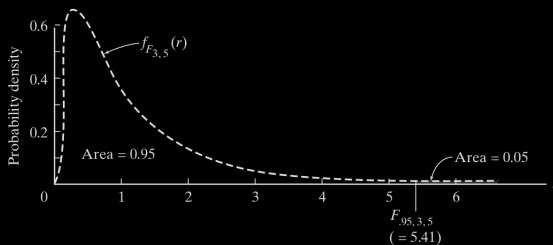


```

1 # Draw F density
2 x=seq(0,5,0.01)
3 pdf= cbind(df(x, df1 = 1, df2 = 1),
4 df(x, df1 = 2, df2 = 1),
5 df(x, df1 = 5, df2 = 2),
6 df(x, df1 = 10, df2 = 1),
7 df(x, df1 = 100, df2 = 100))
8 matplot(x,pdf, type = "l")
9 title("F with various dgrs of freedom")

```

F- Table



$$\mathbb{P}(F_{3,5} \leq 5.41) = 0.95 \quad \Longleftrightarrow \quad F_{0.95, 3, 5} = 5.41$$

```
1 | > pf(5.41, df1 = 3, df2 = 5)
2 | [1] 0.9500093
```

```
1 | > qf(0.95, df1 = 3, df2 = 5)
2 | [1] 5.409451
```

```
1 | > scipy.stats.f.cdf(5.41, 3, 5)
2 | [1] 0.9500092950699683
```

```
1 | > scipy.stats.f.ppf(0.95, 3, 5)
2 | [1] 5.40945131805649
```


Def 7.3.3. Suppose $Z \sim N(0, 1)$, $U \sim \text{Chi Square}(n)$, and $Z \perp U$. Then

$$T_n = \frac{Z}{\sqrt{U/n}}$$

follows the **Student's t-distribution** of n degrees of freedom.

Remark $T_n^2 \sim F$ -distribution with 1 and n degrees of freedom.

Thm 7.3.4. The pdf of the Student t of degree n is

$$f_{T_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \times \left(1 + \frac{t^2}{n}\right)^{-\frac{n+2}{2}}, \quad t \in \mathbb{R}.$$

Proof. Note that $T_n^2 = \frac{Z^2}{U/n}$ follows an $F(1, n)$ distribution. Hence,

$$f_{T_n^2}(t) = \frac{n^{\frac{n}{2}} \Gamma(\frac{n+1}{2})}{\Gamma(\frac{1}{2}) \Gamma(\frac{n}{2})} t^{-\frac{1}{2}} \frac{1}{(n+t)^{\frac{n+1}{2}}}, \quad t > 0.$$

Therefore,

$$F_{T_n}(t) = \mathbb{P}(T_n \leq t) = \mathbb{P}(-\infty < T_n \leq 0) + \mathbb{P}(0 \leq T_n \leq t).$$

The term $\mathbb{P}(-\infty < T_n \leq 0)$ is a constant which will disappear upon differentiation.

Notice that

$$\begin{aligned} \{T_n^2 \leq t^2\} &= \{-t \leq T_n \leq t\} = \{-t \leq T_n \leq 0\} \cup \{0 \leq T_n \leq t\} \\ &= \left\{-t\sqrt{U/n} \leq Z \leq 0\right\} \cup \left\{0 \leq Z \leq t\sqrt{U/n}\right\} \end{aligned}$$

By symmetry of the distribution of Z ,

$$\mathbb{P}\left(-t\sqrt{U/n} \leq Z \leq 0\right) = \mathbb{P}\left(0 \leq Z \leq t\sqrt{U/n}\right)$$

Therefore,

$$\begin{aligned}\mathbb{P}\left(T_n^2 \leq t^2\right) &= \mathbb{P}\left(-t\sqrt{U/n} \leq Z \leq 0\right) + \mathbb{P}\left(0 \leq Z \leq t\sqrt{U/n}\right) \\ &= 2\mathbb{P}\left(0 \leq Z \leq t\sqrt{U/n}\right) \\ &= 2\mathbb{P}(0 \leq T_n \leq t).\end{aligned}$$

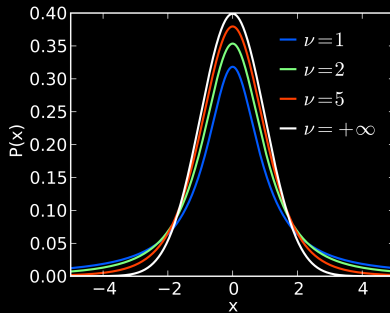
Hence,

$$F_{T_n}(t) = \textit{const.} + \frac{1}{2}\mathbb{P}\left(T_n^2 \leq t^2\right)$$

Finally, differentiation gives the density:

$$f_{T_n}(t) = \frac{d}{dt}F_{T_n}(t) = \frac{d}{dt}\frac{1}{2}F_{T_n^2}(t^2) = t \cdot f_{T_n^2}(t^2) = \dots.$$

□



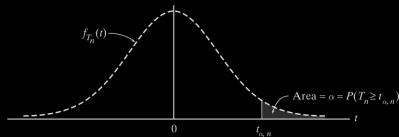
```

1 # Draw Student t-density
2 x=seq(-5,5,0.01)
3 pdf= cbind(dt(x, df = 1),
4           dt(x, df = 2),
5           dt(x, df = 5),
6           dt(x, df = 100))
7 matplot(x,pdf, type = "l")
8 title("Student's t-distributions")

```

t Table

df	α						
	.20	.15	.10	.05	.025	.01	.005
1	1.376	1.963	3.078	6.3138	12.706	31.821	63.657
2	1.061	1.386	1.886	2.9200	4.3027	6.965	9.9248
3	0.978	1.250	1.638	2.3534	3.1825	4.541	5.8409
4	0.941	1.190	1.533	2.1318	2.7764	3.747	4.6041
5	0.920	1.156	1.476	2.0150	2.5706	3.365	4.0321
6	0.906	1.134	1.440	1.9432	2.4469	3.143	3.7074
\vdots			\vdots				
30	0.854	1.055	1.310	1.6973	2.0423	2.457	2.7500
∞	0.84	1.04	1.28	1.64	1.96	2.33	2.58



$$\mathbb{P}(T_3 > 4.541) = 0.01 \iff t_{0.01, 3} = 4.541$$

1	<code>> 1-pt(4.541, df=3)</code>	1	<code>> alpha = 0.01</code>
2	<code>[1] 0.009998238</code>	2	<code>> qt(1-alpha, df = 3)</code>
		3	<code>[1] 4.540703</code>

1	<code>> 1 - scipy.stats.t.cdf(4.541, 3)</code>	1	<code>> scipy.stats.t.ppf(1-0.01, 3)</code>
2	<code>[1] 0.00999823806449407</code>	2	<code>[1] 4.540702858698419</code>

Thm 7.3.5. Let Y_1, \dots, Y_n be a random sample from $N(\mu, \sigma^2)$. Then

$$T_{n-1} = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim \text{Student's } t \text{ of degree } n-1.$$

Proof.

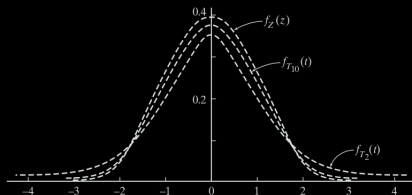
$$\frac{\bar{Y} - \mu}{S/\sqrt{n}} = \frac{\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}}$$

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \quad \perp \quad \frac{(n-1)S^2}{\sigma^2} \sim \text{Chi Square}(n-1)$$

By Def. 7.3.3 ...

□

As $n \rightarrow \infty$, Students' t distribution will converge to $N(0, 1)$:



Thm 7.3.6. $f_{T_n}(x) \rightarrow f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ as $n \rightarrow \infty$, where $Z \sim N(0, 1)$.

Proof By Stirling's formula:

$$\Gamma(z) = \sqrt{\frac{2\pi}{z}} \left(\frac{z}{e}\right)^z (1 + O(1/z)) \quad \text{as } z \rightarrow \infty$$

$$\Rightarrow \lim_{n \rightarrow \infty} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} = \frac{1}{\sqrt{2\pi}}$$

.....

□

Chapter 7. Inference Based on The Normal Distribution

§ 7.1 Introduction

§ 7.2 Comparing $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.3 Deriving the Distribution of $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.4 Drawing Inferences About μ

§ 7.5 Drawing Inferences About σ^2

Let Y_1, \dots, Y_n be a random sample from $N(\mu, \sigma^2)$.

Question Find a test statistic Λ in order to test $H_0 : \mu = \mu_0$ v.s. $H_1 : \mu \neq \mu_0$.

Case I. σ^2 is known:

$$\Lambda = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}}$$

Case II. σ^2 is unknown:

$$\Lambda = ? \quad \Lambda \stackrel{?}{=} \frac{\bar{Y} - \mu_0}{s/\sqrt{n}} \quad \sim \quad ?$$

Summary

A random sample of size n from
a normal distribution $N(\mu, \sigma^2)$

	σ^2 known	σ^2 unknown
Statistic	$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$	$T_{n-1} = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$
Score	$z = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}}$	$t = \frac{\bar{y} - \mu}{s/\sqrt{n}}$
Table	z_α	$t_{\alpha, n-1}$
100(1 - α)% C.I.	$\left(\bar{y} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{y} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$	$\left(\bar{y} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}, \bar{y} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \right)$
Test $H_0 : \mu = \mu_0$		
$H_1 : \mu > \mu_0$	Reject H_0 if $z \geq z_\alpha$	Reject H_0 if $t \geq t_{\alpha, n-1}$
$H_1 : \mu < \mu_0$	Reject H_0 if $z \leq -z_\alpha$	Reject H_0 if $t \leq -t_{\alpha, n-1}$
$H_1 : \mu \neq \mu_0$	Reject H_0 if $ z \geq z_{\alpha/2}$	Reject H_0 if $ t \geq t_{\alpha/2, n-1}$

Computing s from data

Step 1 $a = \sum_{i=1}^n y_i$

Step 2. $b = \sum_{i=1}^n y_i^2$

Step 3. $s = \sqrt{\frac{nb - a^2}{n(n-1)}}$

Proof.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{n(\sum_{i=1}^n y_i^2) - (\sum_{i=1}^n y_i)^2}{n(n-1)}$$

□

Case 7.4.1 How far apart are the bat and the insect when the bat first senses that insect is there?

Or, what is the effective range of a bat's echolocation system?

Table 7.4.1	
Catch Number	Detection Distance (cm)
1	62
2	52
3	68
4	23
5	34
6	45
7	27
8	42
9	83
10	56
11	40

Answer the question by construct a 95% C.I.

Sol. ...



```

1 # Case7_4_1.py
2 import numpy as np
3 import scipy.stats as st
4
5
6 # returns confidence interval of mean
7 def confIntMean(a, conf=0.95):
8     mean, sem, m = np.mean(a), st.sem(a), st.t.ppf((1+conf)/2., len(a)-1)
9     return mean - m*sem, mean + m*sem
10
11
12 def main():
13     alpha = 5
14     data = np.array([62, 52, 68, 23, 34, 45, 27, 42, 83, 56, 40])
15     lower, upper = confIntMean(data, 1-alpha/100)
16     print("""\
17
18     The {alpha}% confidence interval is ({lower:.2f},{upper:.2f})
19
20     """.format(**locals()))
21
22
23 if __name__ == "__main__":
24     main()

```

```

1 In [83]: run Case7_4_1.py
2
3     The 95% confidence interval is (36.21,60.51)

```

Eg. 7.4.2 Bank approval rates for inner-city residents v.s. rural ones.

Approval rate for rural residents is 62%.

Do bank treat two groups equally? $\alpha = 0.05$

Table 7.4.3			
Bank	Location	Affiliation	Percent Approved
1	3rd & Morgan	AU	59
2	Jefferson Pike	TU	65
3	East 150th & Clark	TU	69
4	Midway Mall	FT	53
5	N. Charter Highway	FT	60
6	Lewis & Abbot	AU	53
7	West 10th & Lorain	FT	58
8	Highway 70	FT	64
9	Parkway Northwest	AU	46
10	Lanier & Tower	TU	67
11	King & Tara Court	AU	51
12	Bluedot Corners	FT	59

Sol.

$$H_0 : \mu = 62 \quad v.s. \quad H_1 : \mu \neq 62.$$

Table 7.4.4						
Banks	n	\bar{y}	s	t Ratio	Critical Value	Reject H_0 ?
All	12	58.667	6.946	-1.66	± 2.2010	No

Table 7.4.5							
Banks	n	\bar{y}	s	t Ratio	Critical Value	Reject H_0 ?	
American United	4	52.25	5.38	-3.63	± 3.1825	Yes	
Federal Trust	5	58.80	3.96	-1.81	± 2.7764	No	
Third Union	3	67.00	2.00	+4.33	± 4.3027	Yes	

```

1 # Eg7_4_2.py
2 import numpy as np
3 import scipy.stats as st
4
5 data = np.array([59, 65, 69, 53, 60, 53, 58, 64, 46, 67, 51, 59])
6 alpha = 5
7 mean, sem = np.mean(data), st.sem(data)
8 n = len(data)
9 s = sem * np.sqrt(n)
10 cv = st.t.ppf(1-alpha/200., len(data)-1)
11 tRatio = (mean-62)/sem
12
13
14 print("""\
15     n={n}, sample mean={mean:.3f}, s={s:.3f}, t Ratio={tRatio:.2f}, Critical values
16         ={cv:.4f}
17     """).format(**locals())

```

```

1 In [113]: run Eg7_4_2.py
2
3     n=12, sample mean=58.667, s=6.946, t Ratio=-1.66, Critical values=2.2010

```


Chapter 7. Inference Based on The Normal Distribution

§ 7.1 Introduction

§ 7.2 Comparing $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.3 Deriving the Distribution of $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$

§ 7.4 Drawing Inferences About μ

§ 7.5 Drawing Inferences About σ^2

For a random sample of size n from $N(\mu, \sigma^2)$:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

\Downarrow

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2 \sim \text{Chi Square}(n-1)$$

$$\mathbb{P}\left(\chi_{\alpha/2, n-1}^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{1-\alpha/2, n-1}^2\right) = 1 - \alpha.$$

100(1 - α)% C.I. for σ^2 :

$$\left(\frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2}, \frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2} \right)$$

100(1 - α)% C.I. for σ :

$$\left(\sqrt{\frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2}}, \sqrt{\frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2}} \right)$$

Testing $H_0 : \sigma^2 = \sigma_0^2$

v.s.

(at the α level of significance)

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$H_1 : \sigma^2 < \sigma_0^2$:

Reject H_0 if

$$\chi^2 \leq \chi_{\alpha, n-1}^2$$

$H_1 : \sigma^2 \neq \sigma_0^2$:

Reject H_0 if

$$\chi^2 \leq \chi_{\alpha/2, n-1}^2 \text{ or}$$

$$\chi^2 \geq \chi_{1-\alpha/2, n-1}^2$$

$H_1 : \sigma^2 > \sigma_0^2$:

Reject H_0 if

$$\chi^2 \geq \chi_{1-\alpha, n-1}^2$$

E.g. 1. The width of a confidence interval for σ^2 is a function of n and S^2 :

$$W = \frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2} - \frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2}$$

Find the smallest n such that the average width of a 95% C.I. for σ^2 is no greater than $0.8\sigma^2$.

Sol. Notice that $\mathbb{E}[S^2] = \sigma^2$. Hence, we need to find n s.t.

$$(n-1) \left(\frac{1}{\chi_{0.025, n-1}^2} - \frac{1}{\chi_{0.975, n-1}^2} \right) \leq 0.8.$$

Trial and error (numerics on R) gives $n = 57$.

```

1 > # Example 7.5.1
2 > n=seq(45,60,1)
3 > l=qchisq(0.025,n-1)
4 > u=qchisq(0.975,n-1)
5 > e=(n-1)* (1/l-1/u)
6 > m=cbind(n,l,u,e)
7 > colnames(m) = c("n",
8 +                 "chi(0.025,n-1)",
9 +                 "chi(0.975,n-1)",
10 +                "error")
11 > m
12      n chi(0.025,n-1) chi(0.975,n-1) error
13 [1,] 45      27.57457      64.20146 0.9103307
14 [2,] 46      28.36615      65.41016 0.8984312
15 [3,] 47      29.16005      66.61653 0.8869812
16 [4,] 48      29.95620      67.82065 0.8759533
17 [5,] 49      30.75451      69.02259 0.8653224
18 [6,] 50      31.55492      70.22241 0.8550654
19 [7,] 51      32.35736      71.42020 0.8451612
20 [8,] 52      33.16179      72.61599 0.8355901
21 [9,] 53      33.96813      73.80986 0.8263340
22 [10,] 54      34.77633      75.00186 0.8173761
23 [11,] 55      35.58634      76.19205 0.8087008
24 [12,] 56      36.39811      77.38047 0.8002937
25
26 [13,] 57      37.21159      78.56716 0.7921414
27 [14,] 58      38.02674      79.75219 0.7842313
28 [15,] 59      38.84351      80.93559 0.7765517
29 [16,] 60      39.66186      82.11741 0.7690918

```

Case Study 7.5.2

Mutual funds are investment vehicles consisting of a portfolio of various types of investments. If such an investment is to meet annual spending needs, the owner of shares in the fund is interested in the average of the annual returns of the fund. Investors are also concerned with the volatility of the annual returns, measured by the variance or standard deviation. One common method of evaluating a mutual fund is to compare it to a benchmark, the Lipper Average being one of these. This index number is the average of returns from a universe of mutual funds.

The Global Rock Fund is a typical mutual fund, with heavy investments in international funds. It claimed to best the Lipper Average in terms of volatility over the period from 1989 through 2007. Its returns are given in the table below.

Year	Investment Return %	Year	Investment Return %
1989	15.32	1999	27.43
1990	1.62	2000	8.57
1991	28.43	2001	1.88
1992	11.91	2002	-7.96
1993	20.71	2003	35.98
1994	-2.15	2004	14.27
1995	23.29	2005	10.33
1996	15.96	2006	15.94
1997	11.12	2007	16.71
1998	0.37		

The standard deviation for these returns is 11.28%, while the corresponding figure for the Lipper Average is 11.67%. Now, clearly, the Global Rock Fund has a smaller standard deviation than the Lipper Average, but is this small difference due just to random variation? The hypothesis test is meant to answer such questions.

$$H_0 : \sigma^2 = (11.67)^2$$

versus

$$H_1 : \sigma^2 < (11.67)^2$$

Let $\alpha = 0.05$. With $n = 19$, the critical value for the chi square ratio [from part (b) of Theorem 7.5.2] is $\chi^2_{1-\alpha, n-1} = \chi^2_{0.95, 18} = 9.390$ (see Figure 7.5.3). But

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(19-1)(11.28)^2}{(11.67)^2} = 16.82$$

so our decision is clear: Do not reject H_0 .

