# Math 362: Mathematical Statistics II

Le Chen
le.chen@emory.edu

Emory University
Atlanta, GA

Last updated on April 13, 2021

2021 Spring

# Chapter 7. Inference Based on The Normal Distribution

**Def. Sampling distributions**

Distributions of *functions of random sample* of given size.
statistics / estimators

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

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

**Aim:** Determine distributions for

| | |
|---|---|
| Sample variance $S^2 := \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$ | *Chi square distr.* |
| $T := \dfrac{\overline{Y} - \mu}{S/\sqrt{n}}$ | *Student t distr.* |
| $\dfrac{S_1^2}{\sigma_1^2} \Big/ \dfrac{S_2^2}{\sigma_2^2}$ | *F distr.* |

**Thm 7.3.1.** Let $U = \sum_{i=1}^{m} Z_j^2$, 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,

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

**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,

$$\begin{aligned} F_{Z^2}(u) &= \mathbb{P}\left(Z^2 \le u\right) \\ &= \mathbb{P}\left(-\sqrt{u} \le Z \le \sqrt{u}\right) \\ &= 2\mathbb{P}(0 \le Z \le \sqrt{u}) \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{2\pi} e^{-z^2/2} \mathrm{d}z \end{aligned}$$

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

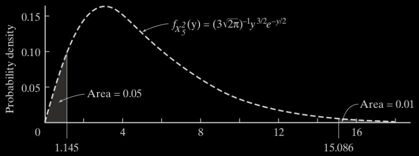
$$\begin{aligned} f_{Z^2}(u) &= \frac{\mathrm{d}}{\mathrm{d}u} F_{Z^2}(u) \\ &= \frac{2}{\sqrt{2\pi}} \frac{1}{2\sqrt{u}} e^{-u/2} \\ &= \frac{1}{\sqrt{2}\Gamma(1/2)} u^{(1/2)-1} e^{-u/2}, \end{aligned}$$

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

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

# Chi Square Table

| | *p* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| df | .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 |




$$\mathbb{P}(\chi_5^2 \le 1.145) = 0.05 \iff \chi^2_{0.05,5} = 1.145$$

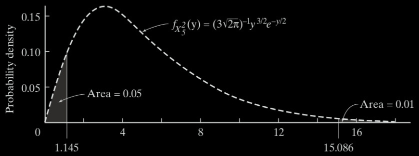
$$\mathbb{P}(\chi_5^2 \le 15.086) = 0.99 \iff \chi^2_{0.99,5} = 15.086$$

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

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

# Chi Square Table

| | p | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| df | .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 |




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

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

```
> scipy.stats.chi2.cdf(1.145, 5)
[1]: 0.04995622155207728
> scipy.stats.chi2.cdf(15.086, 5)
[1]: 0.9899988752378142
```

```
> scipy.stats.chi2.ppf(0.05, 5)
[1]: 1.1454762260617692
> scipy.stats.chi2.ppf(0.99, 5)
[1]: 15.08627246938899
```

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

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

(b) $\dfrac{(n-1)S^2}{\sigma^2} = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 \quad \sim \quad \text{Chi Square}(n-1).$

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

$$\overline{Y} = \frac{Y_1 + Y_2}{2}, \qquad Y_1 - \overline{Y} = \frac{Y_1 - Y_2}{2}, \qquad Y_2 - \overline{Y} = \frac{Y_2 - Y_1}{2}$$

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

**(a)** It is equivalanet 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)** $\frac{(n-1)S^2}{\sigma^2} = \left(\frac{Y_1 - Y_2}{\sqrt{2}\sigma}\right)^2$ and $\frac{Y_1 - Y_2}{\sqrt{2}\sigma} \sim N(0,1)$ ... □

**Def 7.3.2.** If $U \sim$ Chi Square($n$) and $V \sim$ 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:

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

**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

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

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

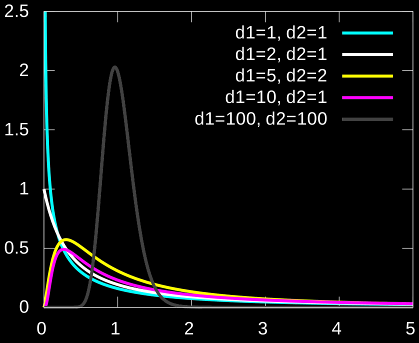
$$f_W(w) = \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} w^{(m/2)-1} \left(\frac{2}{1+w}\right)^{\frac{n+m}{2}} \int_0^\infty y^{\frac{n+m}{2}-1} e^{-y} \mathrm{d}y$$
$$= \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} w^{(m/2)-1} \left(\frac{2}{1+w}\right)^{\frac{n+m}{2}} \Gamma\left(\frac{n+m}{2}\right)$$

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

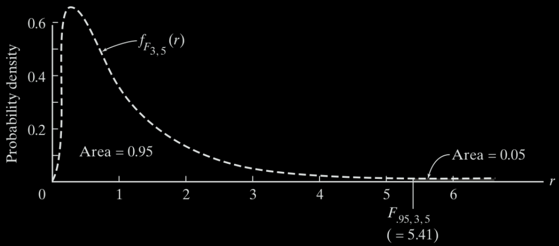
$$f_F(y) = \frac{m}{n} f_W\left(\frac{m}{n}y\right)$$
$$= \frac{m}{n} \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} \left(\frac{m}{n}y\right)^{(m/2)-1} \left(\frac{2}{1+\frac{m}{n}y}\right)^{\frac{n+m}{2}} \Gamma\left(\frac{n+m}{2}\right)$$
$$= \cdots \qquad y \geq 0.$$

□




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

# F- Table




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

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

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

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

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

**Def 7.3.3.** Suppose $Z \sim N(0,1)$, $U \sim$ 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

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

By symmetry of the distribution of $Z$,

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

Therefore,

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

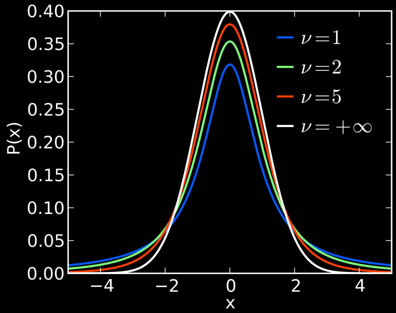
Hence,

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

Finally, differentiation gives the density:

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

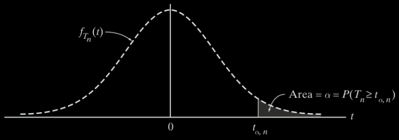
□




```
# Draw Student t−density
x=seq(−5,5,0.01)
pdf= cbind(dt(x, df = 1),
           dt(x, df = 2),
           dt(x, df = 5),
           dt(x, df = 100))
matplot(x,pdf, type = "l")
title("Student's t−distributions")
```

# t Table

| | $\alpha$ | | | | | | |
|---|---|---|---|---|---|---|---|
| df | .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 |
| ⋮ | | | ⋮ | | | | |
| 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-pt(4.541, df =3)
[1] 0.009998238
```

```
> alpha = 0.01
> qt(1-alpha, df = 3)
[1] 4.540703
```

```
> 1 - scipy.stats.t.cdf(4.541, 3)
[1] 0.009998238806449407
```

```
> scipy.stats.t.ppf(1-0.01, 3)
[1] 4.540702858698419
```

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

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

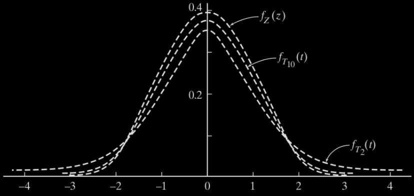
**Proof.**

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

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

By Def. 7.3.3 ... □

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




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

**Proof** By Stirling's formula:

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

$$\implies \quad \lim_{n\to\infty} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} = \frac{1}{\sqrt{2\pi}}$$

...... □