Econ 5023: Statistics for Decision Making

Univariate Statistics (XI): Special, Continuous Parametric Distributions

Le Wang

Other useful, parametric distributions:

1. Log-normal Distribution
2. Chi-squared (χ^2) distribution (with **k** degrees of freedom)
3. Student-t distribution (with **k** degrees of freedom)
4. F-distribution (with $\mathbf{k_1}$ and $\mathbf{k_2}$ degrees of freedom)

All these distributions can be constructed from a Normal Distribution!

Summary of Relationships among Four Distributions

1. $Z = N(0,1) \xrightarrow{\sum_{j=1}^{k} Z_j} \chi_k^2$
2. $t_k \xrightarrow{(t_k)^2} F_{1,k}$
3. $t_k \xrightarrow{k \to \infty} Z = N(0,1)$
4. $F_{k_1,k_2} \xrightarrow{k_2 \to \infty} k_1 \cdot \chi_{k_1}^2$

Things to pay attention when discussing a parametric distribution:

1. What is the probability mass/density function? The relationship between a potential value and the (relative) probability
2. How does this density function look like?
3. What are the features of this distribution? Moments (mean, variance, skewness and kurtosis)
4. What kind of things can be characterized by this distribution?

Note: In what follows, we will present the relationships among normal, Chi-square, t, and F distributions.

How we can be defined one distribution as a function of another. NOT the actual density or CDF functions, even though they are parametric.

Normal

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), \quad -\infty < y < \infty$$

Any normal variable is standardized by

$Z = \frac{Y-\mu}{\sigma} \sim N(0,1)$.

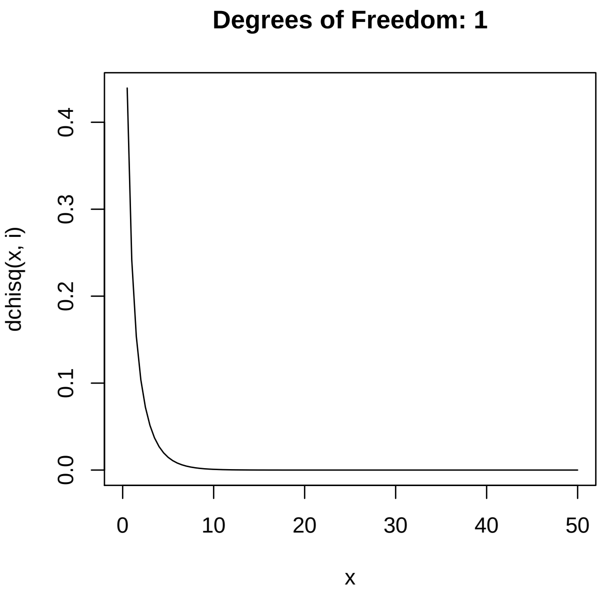
Chi-squared χ^2 with k degrees of freedom

$$Y = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

$Z_j \sim N(0,1), j = 1, \ldots, k$ is a standard normal variable.

Examples:

1. $Y = Z_1^2$ is distributed as χ_1^2
2. $Y = Z_1^2 + Z_2^2$ is distributed as χ_2^2
3. ...


Degrees of Freedom: 1
dchisq(x, i)
0.0
0.1
0.2
0.3
0.4
0
10
20
30
40
50
x

1. It is non-negative [**Question:** Why?]. So, your forecast for an outcome following a χ^2 distribution should never be negative.
2. Asymmetrical and skewed to the right.
3. The Larger the degrees of freedom, the less skewed the density becomes. In particular, skewness $= \sqrt{\frac{8}{k}}$

The χ^2 distributions arises from the need in estimtaion of variance. It is associated with many test statistics, as most of them are about the variances under alternative specifications.

It also leads to some other distributions, e.g., those involving both mean and variance!

Example: (Under Certain Assumptions)

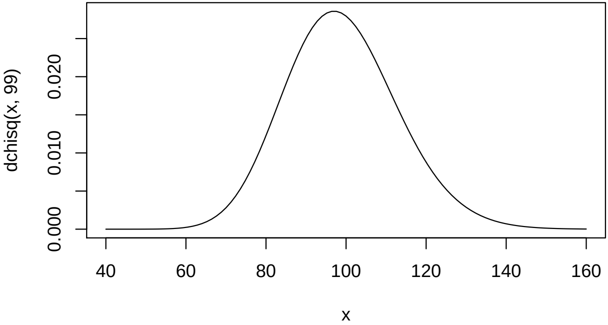
$$\frac{(N-1)\widehat{\sigma}^2}{\sigma^2} = \frac{(N-1)\frac{\sum(X_i-\overline{X})^2}{N-1}}{\sigma^2} \sim \chi^2_{N-1}$$

See the following link for a proof:
`https://onlinecourses.science.psu.edu/stat414/node/174`

Again, how to visualize this?

Let's program it. Suppose $N = 100$. First, let's look at what $\chi^2_{100-1=99}$ will look like.

```
curve(dchisq(x,99), 40, 160)
```



$$\frac{(N-1)\widehat{\sigma}^2}{\sigma^2}$$

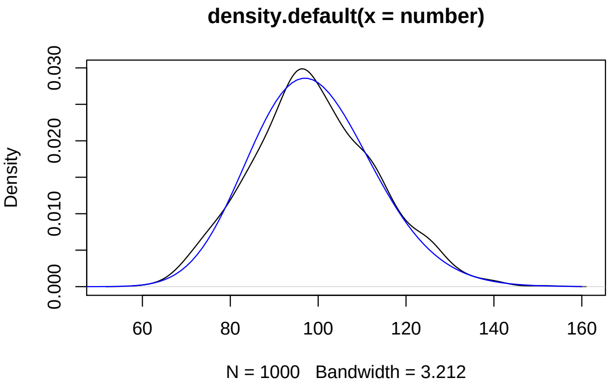
```r
# Set seed to ensure reproducibility
set.seed(123456)

# Initiate an object to store simulated values
number <- double(0)

# Use Loops to Simulate 1000 values
for (i in 1:1000){
  # Draw a random sample from the standard normal (mean 0 
  sample <- rnorm(100)

  # Calculate the number using the formula above
  # And then store it
  number[i] <- (100-1)*var(sample)/1
}
```

```r
plot(density(number))
curve(dchisq(x,99), 40, 160, col='blue', add=T)
```



Let's accumulate what we know!

1. $\frac{\overline{X}-\mu}{\sigma/\sqrt{N}} \sim N(0,1)$
2. $\frac{(N-1)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-1} \to \frac{\widehat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2_{N-1}}{(N-1)}$

Student t distribution with k degrees of freedom

$$t_k = \frac{Z}{\sqrt{\chi_k^2/k}}$$

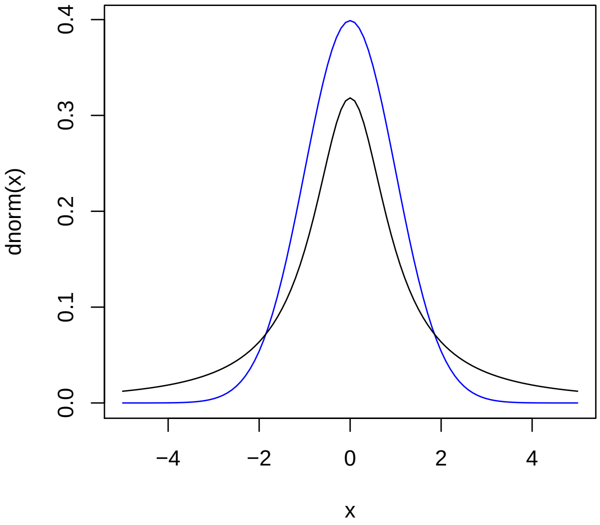
where Z is a standard normal, and χ_k^2 is a chi-squared variable with k degrees of freedom. It will be similarly defined below.

F distribution with k_1, k_2 degrees of freedom

$$F_{k_1,k_2} = \frac{\frac{\chi_{k_1}^2}{k_1}}{\frac{\chi_{k_2}^2}{k_2}}$$

$$(t_k)^2 = \left(\frac{Z}{\sqrt{\chi_k^2/k}}\right)^2 = \frac{Z^2}{\chi_k^2/k} = \frac{Z^2/1}{\chi_k^2/k} = \frac{\chi_1^2/1}{\chi_k^2/k} = F_{1,k}$$

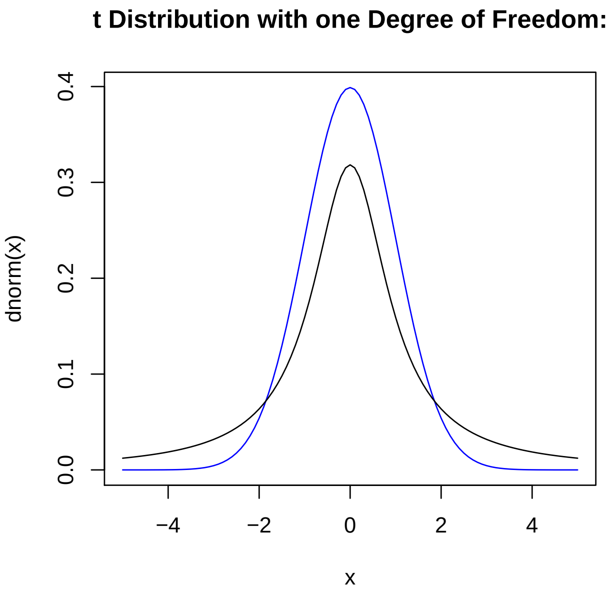
t Distribution with Degrees of Freedom: 1



Some useful facts of t- distribution

1. Symmetric [skewness $= 0$]
2. Fat tail [kurtosis > 3] (very useful for modelling financial variables)
3. As $k \to \infty$, $t_k \to N(0, 1)$

Fat tail further examined



Fat tail further examined: Let's examine the actual "probability" of extreme values (those at both ends)

```
# A large positive value
dt(3,1)

## [1] 0.03183099

dnorm(3)

## [1] 0.004431848

# A large negative value
dt(-3,1)

## [1] 0.03183099

dnorm(-3)

## [1] 0.004431848
```

$$t_k = \frac{Z}{\sqrt{\frac{\chi^2_k}{k}}}$$

As k increases, $t_k \to Z$. Intuitively, you can think that the numeratore is multiplied by k. The standard normal variable prevails as k increases.

$t_1 = 1$

```
pt(1,1)

## [1] 0.75

pt(1,10)

## [1] 0.8295534

pt(1,100)

## [1] 0.8401379

pt(1,1000)

## [1] 0.8412238

pt(1,10000)

## [1] 0.8413326
```

$Z = 1$

```
pnorm(1)

## [1] 0.8413447
```

Example: What is this?

$$\frac{\overline{X} - \mu}{\widehat{\sigma}/\sqrt{N}}$$

Example: What is this?

$$\begin{aligned}\frac{\overline{X}-\mu}{\widehat{\sigma}/\sqrt{N}} &= \frac{(\overline{X}-\mu)}{(\sigma/\sqrt{N})}\bigg/\frac{(\widehat{\sigma}/\sqrt{N})}{(\sigma/\sqrt{N})} \\ &= \frac{(\overline{X}-\mu)}{(\sigma/\sqrt{N})}\bigg/\frac{\widehat{\sigma}}{\sigma}\end{aligned}$$

Let's accumulate what we know!

1. $\frac{\overline{X}-\mu}{\sigma/\sqrt{N}} \sim N(0,1)$
2. $\frac{(N-1)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-1} \to \frac{\widehat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2_{N-1}}{(N-1)}$

$$
\begin{aligned}
\frac{\overline{X}-\mu}{\widehat{\sigma}/\sqrt{N}} &= \frac{(\overline{X}-\mu)}{(\sigma/\sqrt{N})}\Big/\frac{(\widehat{\sigma}/\sqrt{N})}{(\sigma/\sqrt{N})} \\
&= \frac{(\overline{X}-\mu)}{(\sigma/\sqrt{N})}\Big/\frac{\widehat{\sigma}}{\sigma} \\
&= \frac{N(0,1)}{\sqrt{\frac{\chi^2_{N-1}}{(N-1)}}} \\
&\sim t_{N-1}
\end{aligned}
$$

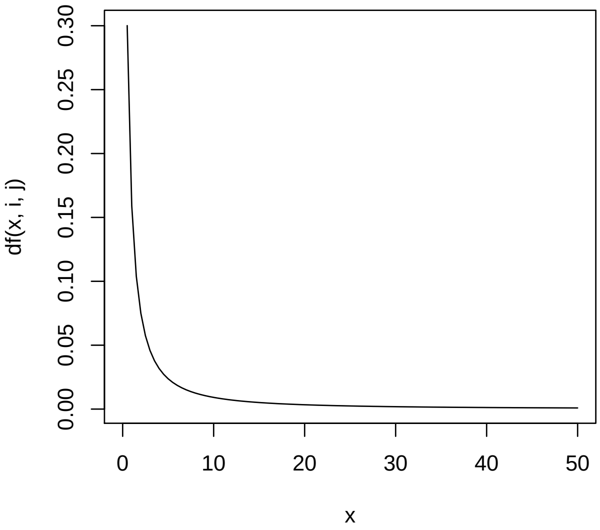
1. $\frac{\overline{X}-\mu}{\sigma/\sqrt{N}} \sim N(0,1)$
2. $\frac{(N-1)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-1} \rightarrow \frac{\widehat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2_{N-1}}{(N-1)}$
3. $t_k = \frac{Z}{\sqrt{\frac{\chi^2_k}{k}}}$

What else do we know? $N \to \infty$, the degrees of freedom for the t_{N-1} distribution increase... it becomes the standard normal distribution!

Let's accumulate what we know!

1. $\frac{\overline{X}-\mu}{\sigma/\sqrt{N}} \sim N(0,1)$
2. $\frac{(N-1)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-1} \rightarrow \frac{\widehat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2_{N-1}}{(N-1)}$
3. $\frac{\overline{X}-\mu}{\widehat{\sigma}/\sqrt{N}} \sim t_{N-1}$

F Distribution with Degrees of Freedom: 1 and 1



1. It is non-negative [**Question:** Why?]. So, your forecast for an outcome following a F distribution should never be negative.
2. Asymmetrical and skewed to the right.

Sounds familiar? χ^2!!

F and χ^2 statistics are really the same thing in that, after a normalization, chi-squared is the limiting distribution of the F as the denominator degrees (k_2) of freedom goes to infinity. The normalization is

$$\chi^2_{k_1} = k_1 * F_{k_1,k_2}, \quad k_2 \to \infty$$

$F_{2,71} = 2.05$

```
pf(2.05, 2, 71)

## [1] 0.8637133

pf(2.05, 2, 10000)

## [1] 0.871211

pf(2.05, 2, 100000)

## [1] 0.8712597

pf(2.05, 2, 1000000)

## [1] 0.8712646

pf(2.05, 2, 10000000)

## [1] 0.871265
```

$\chi^2_2 = 2.05 \times 2 = 4.1$

```
pchisq(4.1, 2)

## [1] 0.8712651
```

$F_{2,71} = 3$

```r
pf(3, 2, 71)
```

```
## [1] 0.9438628
```

```r
pf(3, 2, 10000)
```

```
## [1] 0.9501681
```

```r
pf(3, 2, 100000)
```

```
## [1] 0.9502085
```

```r
pf(3, 2, 1000000)
```

```
## [1] 0.9502125
```

```r
pf(3, 2, 10000000)
```

```
## [1] 0.9502129
```

$\chi^2_2 = 3 \times 2 = 6$

```r
pchisq(6, 2)
```

```
## [1] 0.9502129
```

Summary of Relationships among Four Distributions

1. $Z = N(0,1) \xrightarrow{\sum_{j=1}^{k} Z_j} \chi_k^2$
2. $t_k \xrightarrow{(t_k)^2} F_{1,k}$
3. $t_k \xrightarrow{k \to \infty} Z = N(0,1)$
4. $F_{k_1,k_2} \xrightarrow{k_2 \to \infty} k_1 \cdot \chi_{k_1}^2$

Log Normal Distribution (Optional)

The distributions of many economic and financial variables are lognormal instead of normal in their original forms.

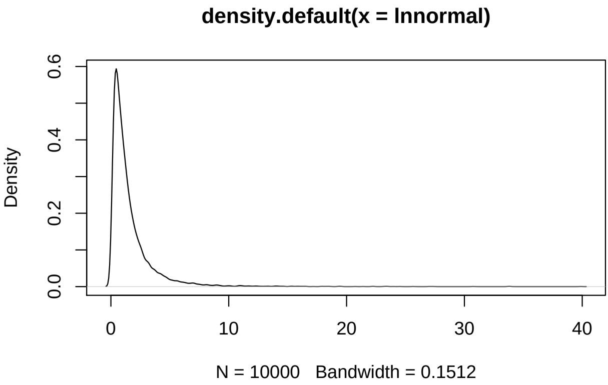
Normal

$$f(y)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), -\infty<y<\infty$$

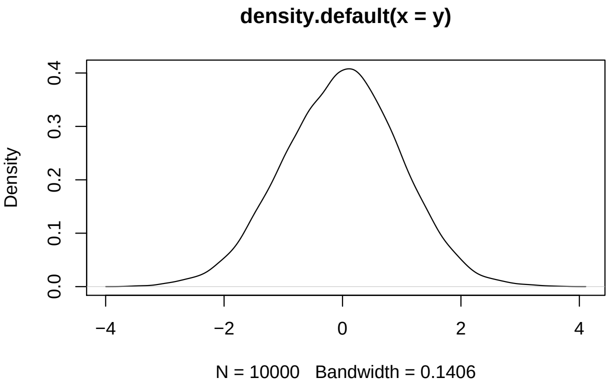
Log Normal Distribution

$$f(y)=\frac{1}{y\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right), 0<y<\infty$$

```
lnnormal<-rlnorm(10000)
plot(density(lnnormal))
```



```r
y<-log(lnnormal)
plot(density(y))
```



We often transform economic and financial variables by taking the log of the original variable of interest.

1. Many economic and financial variables grow exponentially, so their path is non-linear. Transforming these variables with a logarithm operation achieves linearity.
2. The transformation through logarithm operations changes the variables in concern from absolute terms to relative terms, so comparison can be made cross-sections and over time.
3. The logarithm transformation may help achieve stationarity in time series data, though this statement may be controversial.

Probably, one of the most convenient reasons for many economic and financial variables to follow lognormal distributions is the **non-negative constraint**.

That is, these variables can only take values that are greater than or equal to zero. Due partly to this, values closer to zero are compressed and those far away from zero are stretched out. Lognormal distributions possess these features.