Economics 103 – Statistics for Economists

Rossa O'Keeffe-O'Donovan
(heavily indebted to Francis J. DiTraglia)

University of Pennsylvania

Lecture # 11

Continuous RVs II: The Normal RV and friends

Today: What we will cover

The Normal Random Variable

- Standard normal, $N(0,1)$
- Other normal random variables, $N(\mu, \sigma^2)$
- Linear combinations of independent normal RVs
- R functions

Friends of Normal RV

- Definitions, pdfs, intuition
- F, chi-squared, t

Standard Normal Random Variable: $N(0, 1)$

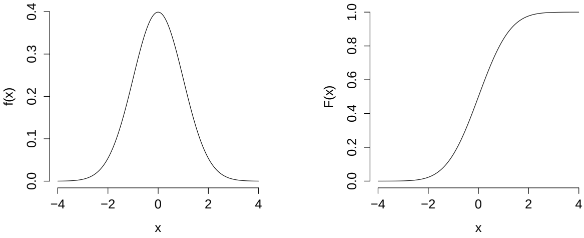
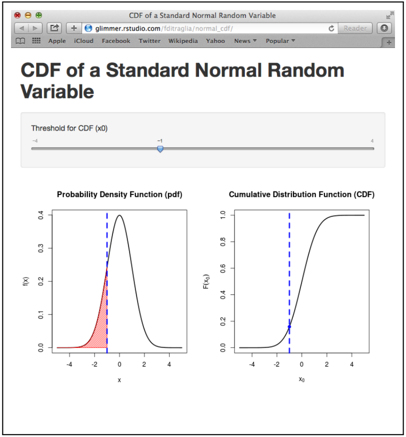


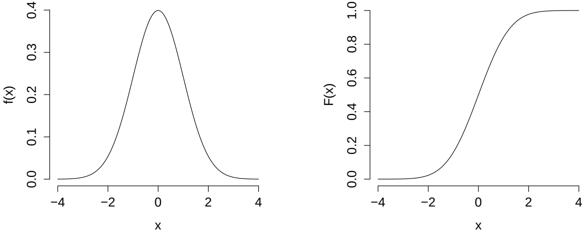
Figure: Standard Normal PDF (left) and CDF (Right)

- Notation: $X \sim N(0, 1)$
- Symmetric, Bell-shaped, $E[X] = 0$, $Var[X] = 1$
- Support Set $= (-\infty, \infty)$

https://fditraglia.shinyapps.io/normal_cdf



Standard Normal Random Variable: $N(0,1)$



- There is no closed-form expression for the $N(0,1)$ CDF.
- For Econ 103, don't need to know formula for $N(0,1)$ PDF.
  - For those interested: $f(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2}x^2)$
- You *do need* to know the R commands. . .

R Commands for the Standard Normal RV

`dnorm` – Standard Normal PDF

- To remember: `d` = density, `norm` = normal
- Example: `dnorm(1)` gives height of $N(0, 1)$ PDF at 0.1, $f(1)$.

`pnorm` – Standard Normal CDF

- To remember: `p` = probability, `norm` = normal
- Example: `pnorm(0.5)` $= P(X \leq 0.5)$ if $X \sim N(0, 1)$.

`rnorm` – Simulate Standard Normal Draws

- To remember: `r` = random, `norm` = normal.
- Example: `rnorm(10)` makes ten iid $N(0, 1)$ draws.

$\Phi(x_0)$ Denotes the $N(0,1)$ CDF

You will sometimes encounter the notation $\Phi(x_0)$ for the CDF of a standard normal RV, so that $\Phi(x_0) = P(X \leq x_0)$ if $X \sim N(0,1)$

It means the same thing as `pnorm`(x_0) but it is not an R command.

Confusingly enough, sometimes you will see $\phi(x)$ referring to the pdf of the standard normal random variable (Φ is an upper case 'phi' while ϕ is a lower case 'phi')

The $N(\mu, \sigma^2)$ Random Variable

Idea

Take a linear function of the $N(0,1)$ RV.

Formal Definition

$N(\mu, \sigma^2) \equiv \mu + \sigma X$ where $X \sim N(0,1)$ and μ, σ are constants.

Properties of $N(\mu, \sigma^2)$ RV

- Parameters: Expected Value $= \mu$, Variance $= \sigma^2$
- Symmetric and bell-shaped.
- Support Set $= (-\infty, \infty)$
- $N(0,1)$ is the special case where $\mu = 0$ and $\sigma^2 = 1$.

Expected Value: μ *shifts* PDF

all of these have $\sigma = 1$

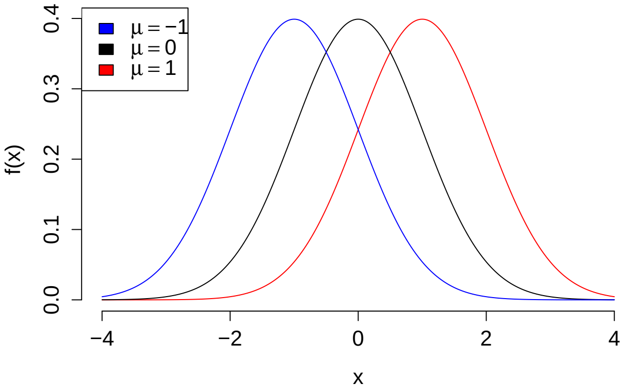


Figure: Blue $\mu = -1$, Black $\mu = 0$, Red $\mu = 1$

Standard Deviation: σ *scales* PDF

all of these have $\mu = 0$

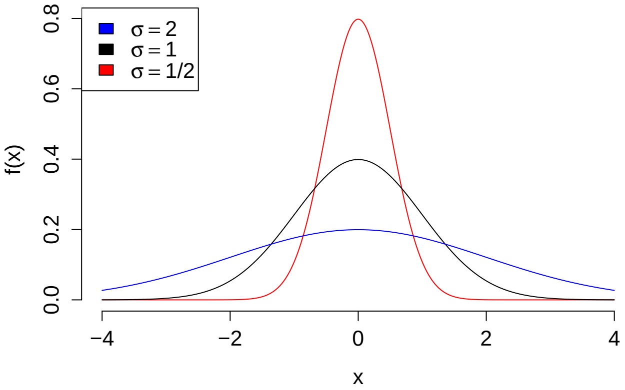


Figure: Blue $\sigma^2 = 4$, Black $\sigma^2 = 1$, Red $\sigma^2 = 1/4$

Linear Function of Normal RV is a Normal RV

Suppose that $X \sim N(\mu, \sigma^2)$. Then if a and b constants,

$$\boxed{a + bX \sim N(a + b\mu, b^2\sigma^2)}$$

Important

- Remember, for *any* RV, X:
  - $E[a + bX] = a + bE[X]$
  - $Var(a + bX) = b^2 Var(X)$
- Key point: linear transformation of normal is still normal!
- Linear transformation of Binomial is *not* Binomial!

Example

Suppose $X \sim N(\mu, \sigma^2)$ and let $Z = (X - \mu)/\sigma$. What is the distribution of Z?

(a) $N(\mu, \sigma^2)$

(b) $N(\mu, \sigma)$

(c) $N(0, \sigma^2)$

(d) $N(0, \sigma)$

(e) $N(0, 1)$

(e)

Linear Combinations of *Multiple Independent* Normals

Let $X \sim N(\mu_x, \sigma_x^2)$ independent of $Y \sim N(\mu_y, \sigma_y^2)$. Then if a, b, c are constants:

$$\boxed{aX + bY + c \sim N(a\mu_x + b\mu_y + c, a^2\sigma_x^2 + b^2\sigma_y^2)}$$

Important

- Result assumes independence
- Particular to Normal RV
- Extends to more than two Normal RVs

Suppose $X_1, X_2, \sim$ iid $N(\mu, \sigma^2)$

Let $\bar{X} = (X_1 + X_2)/2$. What is the distribution of $\bar{X}$?

(a) $N(\mu, \sigma^2/2)$

(b) $N(0, 1)$

(c) $N(\mu, \sigma^2)$

(d) $N(\mu, 2\sigma^2)$

(e) $N(2\mu, 2\sigma^2)$

(a)

Where does the Empirical Rule come from?

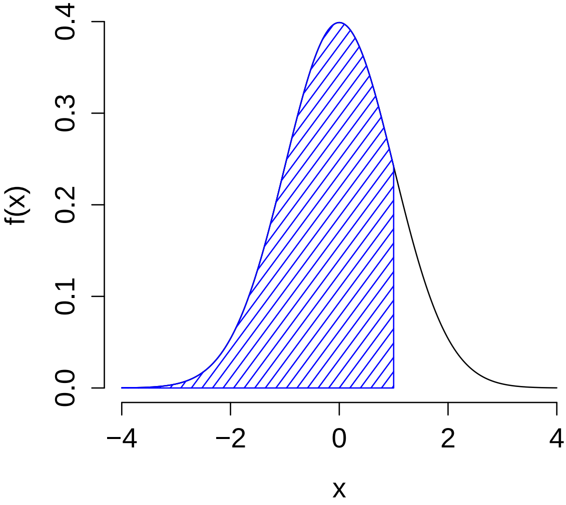
Empirical Rule

Approximately 68% of observations within $\mu \pm \sigma$

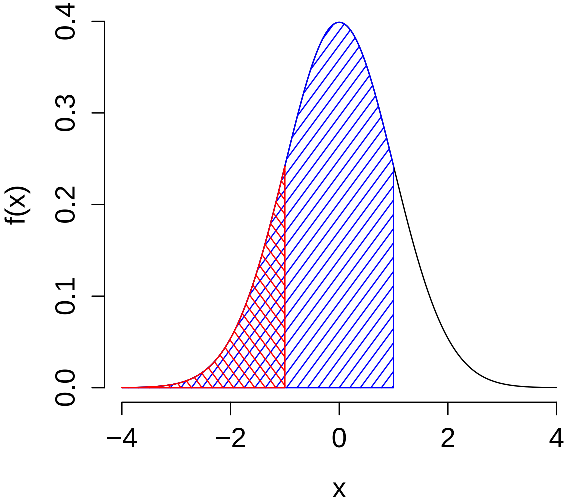
Approximately 95% of observations within $\mu \pm 2\sigma$

Nearly all observations within $\mu \pm 3\sigma$

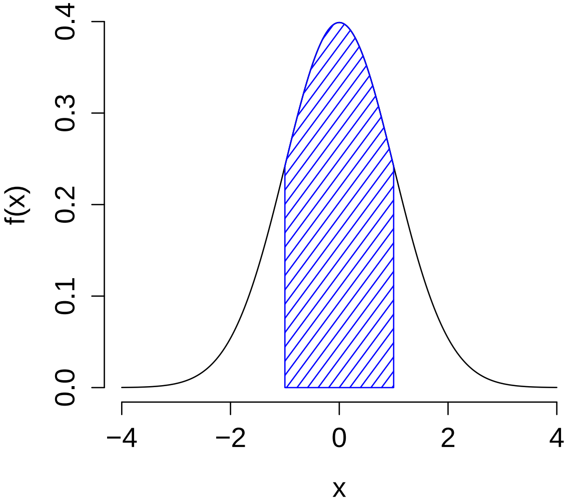
`pnorm(1)` ≈ 0.84



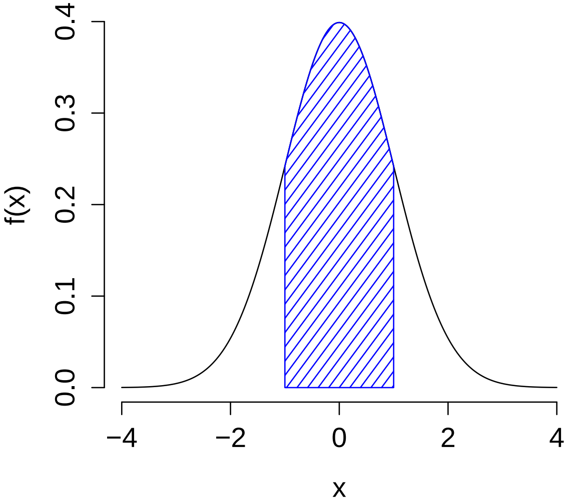
`pnorm(1) - pnorm(-1)` $\approx 0.84 - 0.16$



`pnorm(1) - pnorm(-1)` ≈ 0.68



Middle 68% of $N(0,1) \Rightarrow$ approx. $(-1,1)$



Suppose $X \sim N(0, 1)$

$$
\begin{aligned}
P(-1 \le X \le 1) &= \texttt{pnorm(1) - pnorm(-1)} \\
&\approx 0.683
\end{aligned}
$$

$$
\begin{aligned}
P(-2 \le X \le 2) &= \texttt{pnorm(2) - pnorm(-2)} \\
&\approx 0.954
\end{aligned}
$$

$$
\begin{aligned}
P(-3 \le X \le 3) &= \texttt{pnorm(3) - pnorm(-3)} \\
&\approx 0.997
\end{aligned}
$$

And (this one is new):

$$
\begin{aligned}
P(-1.64 \le X \le 1.64) &= \texttt{pnorm(1.64) - pnorm(-1.64)} \\
&\approx 0.899
\end{aligned}
$$

What if $X \sim N(\mu, \sigma^2)$?

$$
\begin{aligned}
P(X \leq a) &= P(X - \mu \leq a - \mu) \\
&= P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) \\
&= P\left(Z \leq \frac{a - \mu}{\sigma}\right)
\end{aligned}
$$

Where Z is a standard normal random variable, i.e. $N(0, 1)$.

Which of these equals $P\left(Z \le (a-\mu)/\sigma\right)$ if $Z \sim N(0,1)$?

(a) `pnorm(a)`

(b) $1 -$ `pnorm(a)`

(c) `pnorm(a)`$/\sigma - \mu$

(d) `pnorm`$\left(\frac{a-\mu}{\sigma}\right)$

(e) None of the above.

(d)

Probability *Above* a Threshold: $X \sim N(\mu, \sigma^2)$

$$
\begin{aligned}
P(X \geq b) &= 1 - P(X \leq b) = 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= 1 - P\left(Z \leq \frac{b-\mu}{\sigma}\right) \\
&= 1 - \texttt{pnorm}((b-\mu)/\sigma)
\end{aligned}
$$

Where Z is a standard normal random variable.

Probability of an Interval: $X \sim N(\mu, \sigma^2)$

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\ &= P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) \\ &= \texttt{pnorm}((b-\mu)/\sigma) - \texttt{pnorm}((a-\mu)/\sigma) \end{aligned}$$

Where Z is a standard normal random variable.

Suppose $X \sim N(\mu, \sigma^2)$

What is $P(\mu - \sigma \leq X \leq \mu + \sigma)$?

$$
\begin{aligned}
P(\mu - \sigma \leq X \leq \mu + \sigma) &= P\left(-1 \leq \frac{X - \mu}{\sigma} \leq 1\right) \\
&= P\left(-1 \leq Z \leq 1\right) \\
&= \texttt{pnorm(1)} - \texttt{pnorm(-1)} \\
&\approx 0.68
\end{aligned}
$$

Percentiles/Quantiles for Continuous RVs

Quantile Function $Q(p)$ is the inverse of CDF $F(x_0)$

$Q(p)$, gives the value of x_0 such that $F(x_0) = p$

- (i.e. the value of x_0 such that proportion p of the time the realization of the RV is less than x_0)

$$Q(p) = F^{-1}(p)$$

In other words:

$$Q(p) = \text{the value of } x_0 \text{ such that } \int_{-\infty}^{x_0} f(x)\,dx = p$$

Inverse exists as long as $F(x_0)$ is *strictly increasing*.

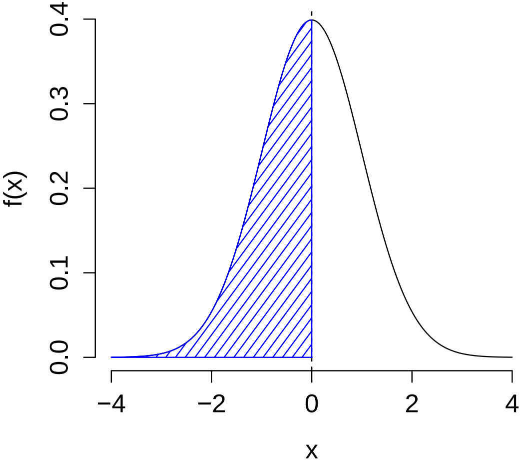
Example: Median

The median of a continuous random variable is $Q(0.5)$, i.e. the value of x_0 such that

$$\int_{-\infty}^{x_0} f(x)\, dx = 1/2$$

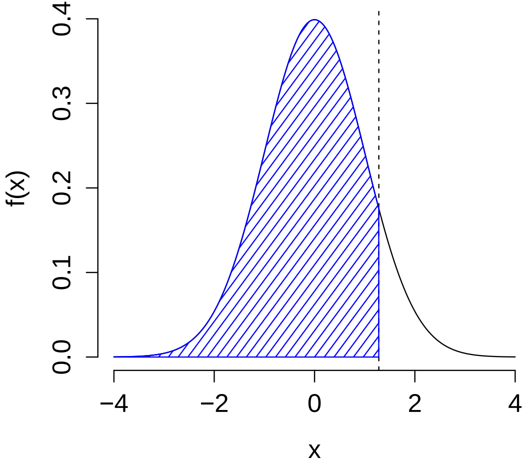
What is the median of a standard normal RV?

By symmetry, $Q(0.5) = 0$. R command: `qnorm(0.5)`



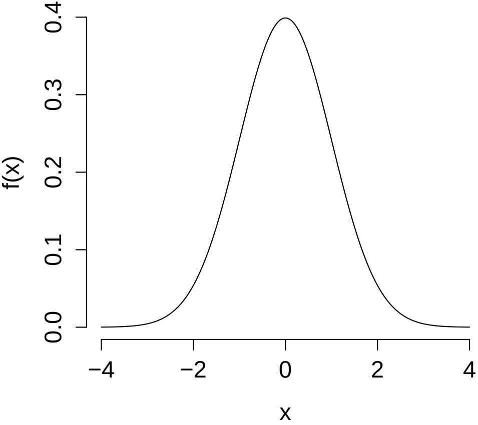
90th Percentile of a Standard Normal

`qnorm(0.9)` ≈ 1.28



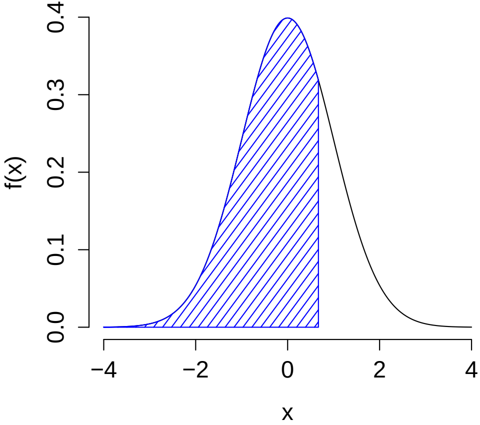
Using Quantile Function to find Symmetric Intervals

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



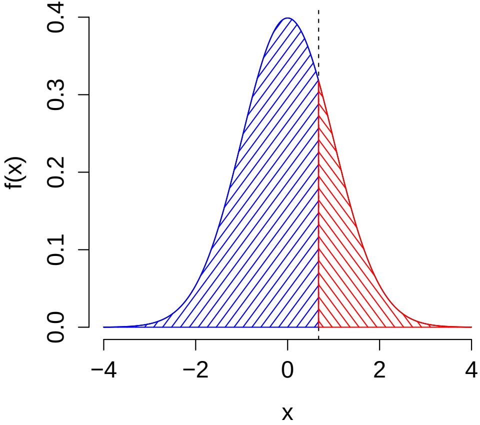
qnorm(0.75) ≈ 0.67

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



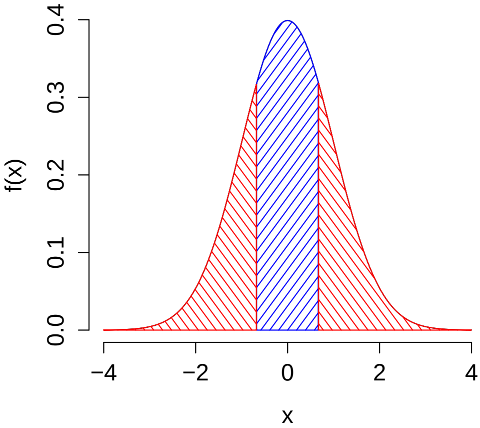
`qnorm(0.75)` ≈ 0.67

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



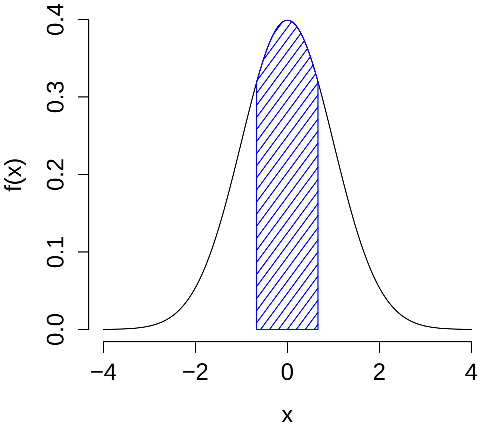
pnorm(0.67)-pnorm(-0.67)≈?

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



`pnorm(0.67)-pnorm(-0.67)`≈ 0.5

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



95% Central Interval for Standard Normal

Suppose X is a standard normal random variable. What value of c ensures that $P(-c \leq X \leq c) \approx 0.95$?

(see slide 20 for key numbers to remember)

R Commands for *Arbitrary* Normal RVs

Let $X \sim N(\mu, \sigma^2)$. Then we can use R to evaluate the CDF and Quantile function of X as follows:

CDF, $F(x)$	`pnorm(x, mean =` μ`, sd =` σ`)`
Quantile Function, $Q(p)$	`qnorm(p, mean =` μ`, sd =` σ`)`
Density, $f(x)$	`dnorm(x, mean =` μ`, sd =` σ`)`
Generate n random numbers	`rnorm(n, mean =` μ`, sd =` σ`)`

Notice that this means you don't have to transform X to a standard normal in order to find areas under its pdf using R.

Example from Homework: $X \sim N(0, 16)$

One Way:

$$
\begin{aligned}
P(X \geq 10) &= 1 - P(X \leq 10) = 1 - P(X/4 \leq 10/4) \\
&= 1 - P(Z \leq 2.5) = 1 - \Phi(2.5) = 1 - \texttt{pnorm(2.5)} \\
&\approx 0.006
\end{aligned}
$$

An Easier Way:

$$
\begin{aligned}
P(X \geq 10) &= 1 - P(X \leq 10) \\
&= 1 - \texttt{pnorm(10, mean = 0, sd = 4)} \\
&\approx 0.006
\end{aligned}
$$

Friends of the normal distribution

Functions of Independent RVs are Independent

If X and Y are independent random variables and g and h are functions, then the random variables $g(X)$ and $h(Y)$ are also independent.

χ^2 Random Variable

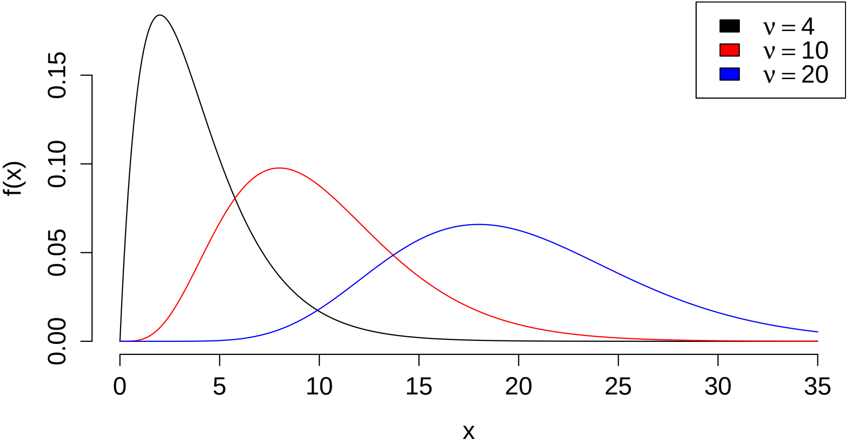
Let $X_1, \ldots, X_\nu \sim$ iid $N(0,1)$. Then,

$$\left(X_1^2 + \ldots + X_\nu^2\right) \sim \chi^2(\nu)$$

where the parameter ν ('nu') is the *degrees of freedom*

Support $= (0, \infty)$

χ^2 PDFs



Student-t Random Variable

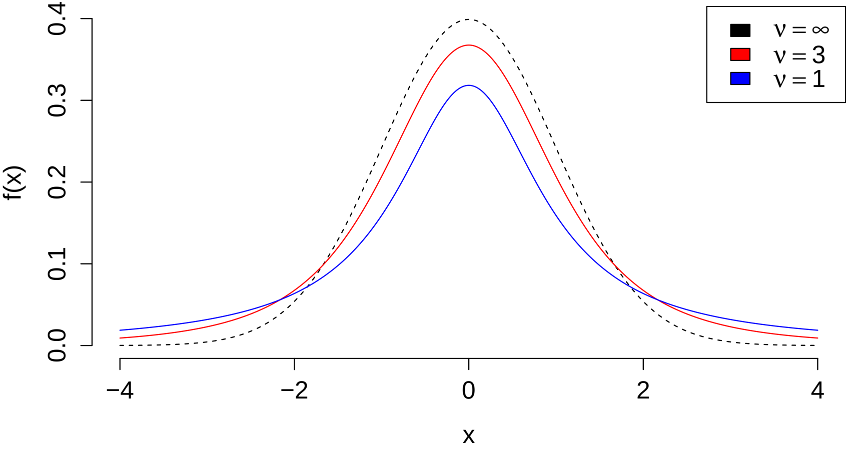
Let $X \sim N(0,1)$ independent of $Y \sim \chi^2(\nu)$. Then,

$$\frac{X}{\sqrt{Y/\nu}} \sim t(\nu)$$

where the parameter ν is the degrees of freedom.

- Support $= (-\infty, \infty)$
- As $\nu \to \infty$, $t(\nu) \to$ Standard Normal.
- Symmetric around zero, but mean and variance may not exist!
- Degrees of freedom ν control "thickness of tails"

Student-t PDFs



F Random Variable

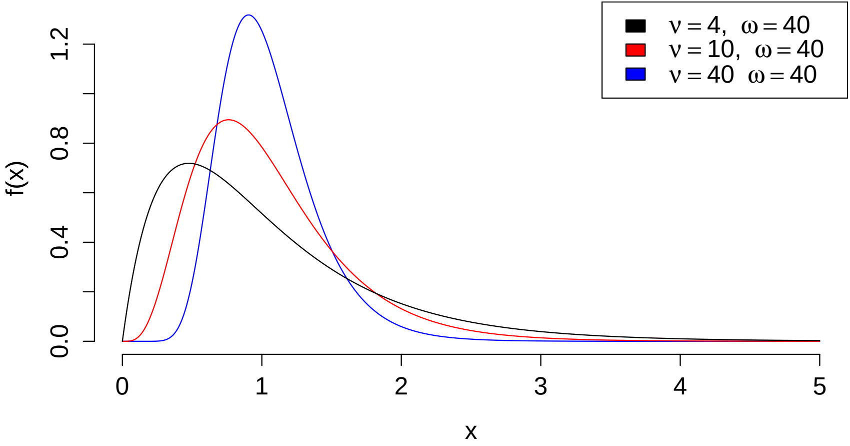
Suppose $X \sim \chi^2(\nu)$ independent of $Y \sim \chi^2(\omega)$. Then,

$$\frac{X/\nu}{Y/\omega} \sim F(\nu, \omega)$$

where ν is the numerator degrees of freedom and ω is the denominator degrees of freedom.

Support $= (0, \infty)$

F PDFs



R Commands – CDFs and Quantile Functions

$F(x) = P(X \leq x)$ is the CDF, $Q(p) = F^{-1}(p)$ the Quantile Function

	$F(x)$	$Q(p)$
$N(\mu, \sigma^2)$	`pnorm(x, mean =` μ`, sd =` σ`)`	`qnorm(p, mean =` μ`, sd =` σ`)`
$\chi^2(\nu)$	`pchisq(x, df =` ν`)`	`qchisq(p, df =` ν`)`
$t(\nu)$	`pt(x, df =` ν`)`	`qt(p, df =` ν`)`
$F(\nu, \omega)$	`pf(x, df1 =` ν`, df2 =` ω`)`	`qf(p, df1 =` ν`, df2 =` ω`)`

To remember: "p" is for Probability, "q" is for Quantile.

R Commands – PDFs and Random Draws

	$f(x)$	Make `n` iid Random Draws
$N(\mu, \sigma^2)$	`dnorm(x, mean =` μ`, sd =` σ`)`	`rnorm(n, mean =` μ`, sd =` σ`)`
$\chi^2(\nu)$	`dchisq(x, df =` ν`)`	`rchisq(n, df =` ν`)`
$t(\nu)$	`dt(x, df =` ν`)`	`rt(n, df =` ν`)`
$F(\nu, \omega)$	`df(x, df1 =` ν`, df2 =` ω`)`	`rf(n, df1 =` ν`, df2 =` ω`)`

To remember: "d" is for Density, "r" is for Random.

Example: $X_1, X_2, X_3 \sim$ iid $N(0,1)$

What is the distribution of $Y_1 = X_1^2 + X_2^2$?

Sum of squares of two indep. std. normals $\Rightarrow$ $Y_1 \sim \chi^2(2)$

What is the distribution of $Y_2 = (Y_1/2)/(X_3^2)$?

$Y_1 \sim \chi^2(2)$ and $X_3^2 \sim \chi^2(1)$

Hence $Y_2 =$ ratio of two indep. χ^2 RVs, each divided by its degrees of freedom $\Rightarrow$ $Y_2 \sim F(2,1)$

What is the distribution of $Z = X_3/\sqrt{Y_1/2}$?

Ratio of standard normal and square root of independent χ^2 RV divided by its degrees of freedom $\Rightarrow$ $Z \sim t(2)$