Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Lecture # 13

Continuous RVs – Part III

Last Time

- Expectation for Continuous RVs
- Normal Random Variable
- Linear Combination of Normal RV
- Areas under Normal pdfs

Today

- Percentiles/Quantiles for Continuous RVs
- Linear Combination of *Several* Normal RVs
- Friends of Normal Distribution

Recall: Def. of Cumulative Distribution Function (CDF)

$$
\begin{aligned}
F(x_0) &\equiv P(X \leq x_0) \\
&= \int_{-\infty}^{x_0} f(x)\,dx \text{ for Continuous RVs}
\end{aligned}
$$

Percentiles/Quantiles for Continuous RVs

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

Plug in a probability p, get out the value of x_0 such that $F(x_0) = p$

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

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

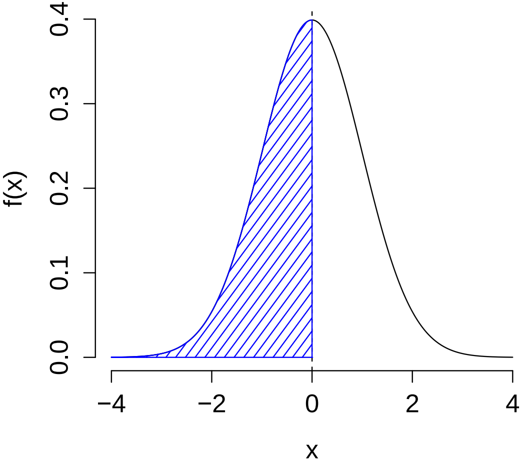
The median of a continuous random variable is $Q(0.5)$, i.e. the value of x_0 such that

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

What is the median of a standard normal RV?

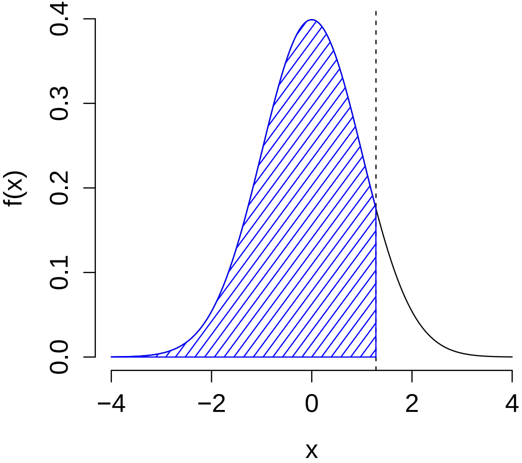
What is the median of a standard normal RV?

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



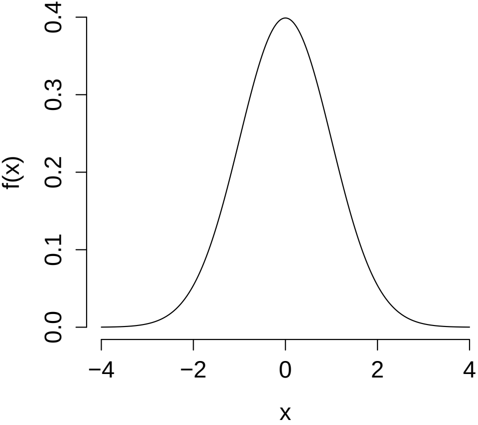
90th Percentile of a Standard Normal

`qnorm(0.9)` ≈ 1.28



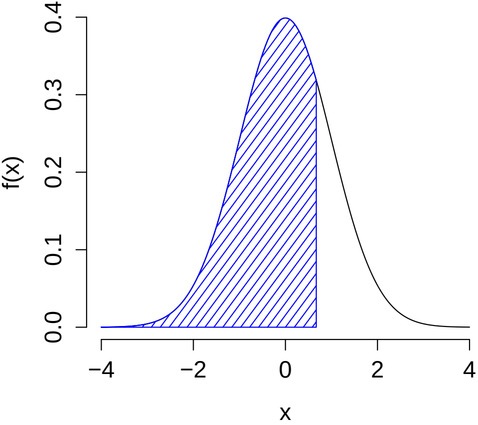
Using Quantile Function to find Symmetric Intervals

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



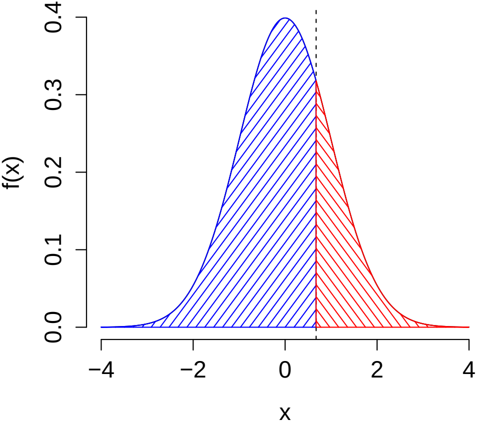
qnorm(0.75) ≈ 0.67

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



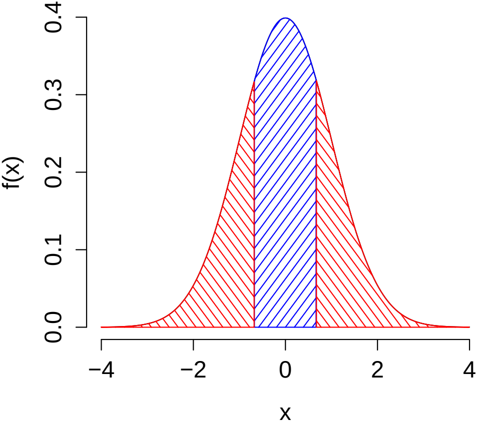
`qnorm(0.75)` ≈ 0.67

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



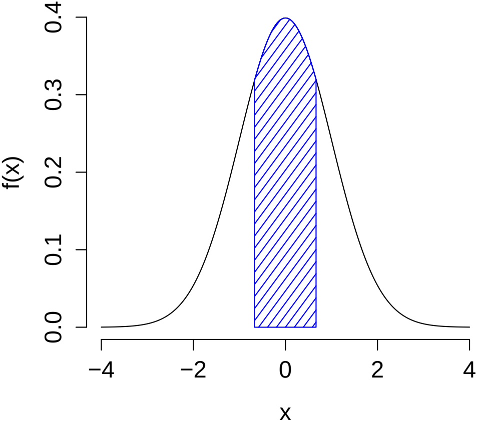
`pnorm(0.67)-pnorm(-0.67)`≈?

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 $P(-c \leq X \leq c) = 0.5$?



68% Central Interval for Standard Normal

Suppose X is a standard normal random variable. What value of c ensures that $P(-c \leq X \leq c) \approx 0.68$?

95% Central Interval for Standard Normal

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

R Commands for *Arbitrary* Normal Distributions

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`)`

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:

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

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:

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

Suppose X has mean μ_x variance σ_x^2 and is independent of Y, which has mean μ_y variance σ_y^2. Let a, b be constants.

What is $E[aX + bY]$?

Suppose X has mean μ_x variance σ_x^2 and is independent of Y, which has mean μ_y variance σ_y^2. Let a, b be constants.

What is $E[aX + bY]$?

$E[aX + bY] = aE[X] + bE[Y] = a\mu_x + b\mu_y$

What is $Var(aX + bY)$?

Suppose X has mean μ_x variance σ_x^2 and is independent of Y, which has mean μ_y variance σ_y^2. Let a, b be constants.

What is $E[aX + bY]$?

$E[aX + bY] = aE[X] + bE[Y] = a\mu_x + b\mu_y$

What is $Var(aX + bY)$?

$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) = a^2\sigma_x^2 + b^2\sigma_y^2$

By independence.

Now suppose $X \sim N(\mu_x, \sigma_x^2)$ independent of $Y \sim N(\mu_y, \sigma_y^2)$. Let a, b be constants.

What is $E[aX + bY]$?

Now suppose $X \sim N(\mu_x, \sigma_x^2)$ independent of $Y \sim N(\mu_y, \sigma_y^2)$. Let a, b be constants.

What is $E[aX + bY]$?

$E[aX + bY] = aE[X] + bE[Y] = a\mu_x + b\mu_y$

What is $Var(aX + bY)$?

Now suppose $X \sim N(\mu_x, \sigma_x^2)$ independent of $Y \sim N(\mu_y, \sigma_y^2)$. Let a, b be constants.

What is $E[aX + bY]$?

$E[aX + bY] = aE[X] + bE[Y] = a\mu_x + b\mu_y$

What is $Var(aX + bY)$?

$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) = a^2\sigma_x^2 + b^2\sigma_y^2$

By independence.

Here's the Surprising Thing:

If X and Y are independent Normal Random Variables and a, b are constants, then $aX + bY$ is *also* a Normal Random Variable!

Linear Combinations of 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 Distribution
- 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) $N(\mu, 2\sigma^2)$

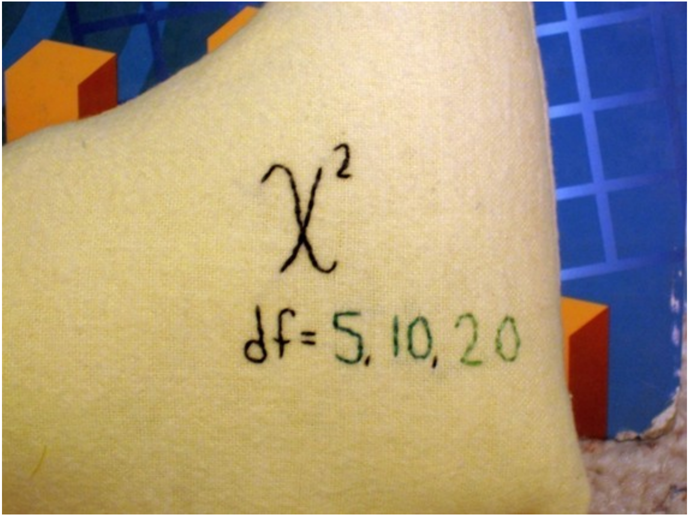
(e) $N(2\mu, 2\sigma^2)$

Figure : The Normal Distribution and Friends.

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.

Figure : PDF for χ^2-Distribution


χ^2
df = 5, 10, 20

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 is the *degrees of freedom*

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 is the *degrees of freedom*

Support $= (0, \infty)$

χ^2 PDFs

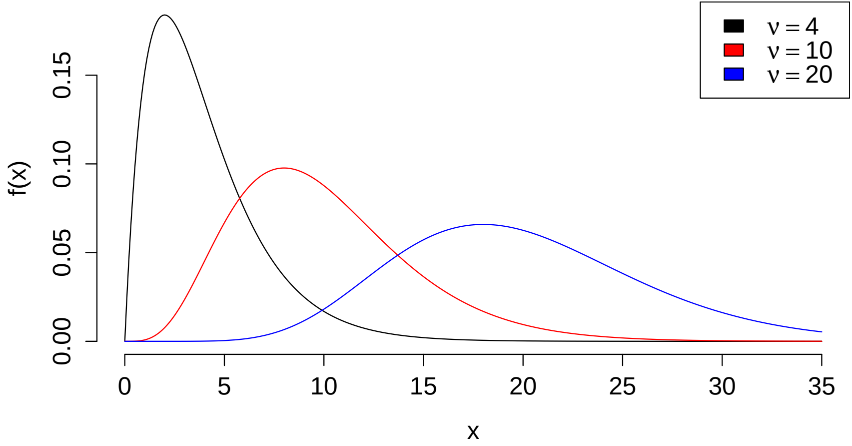
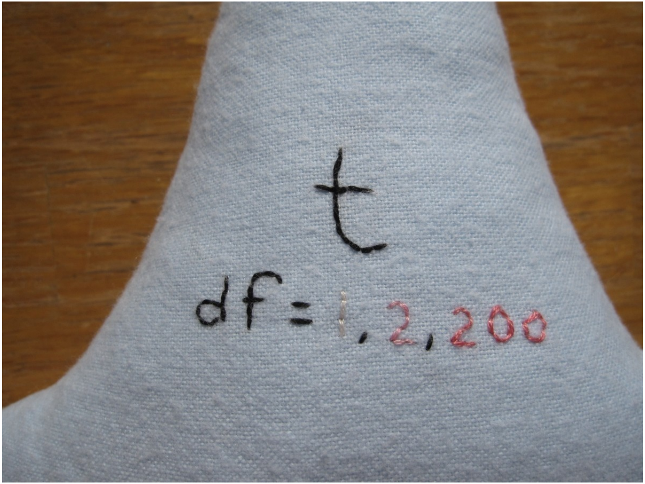


Figure : PDF for Student-t Distribution


t
df = 1, 2, 200

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.

Student-t Random Variable

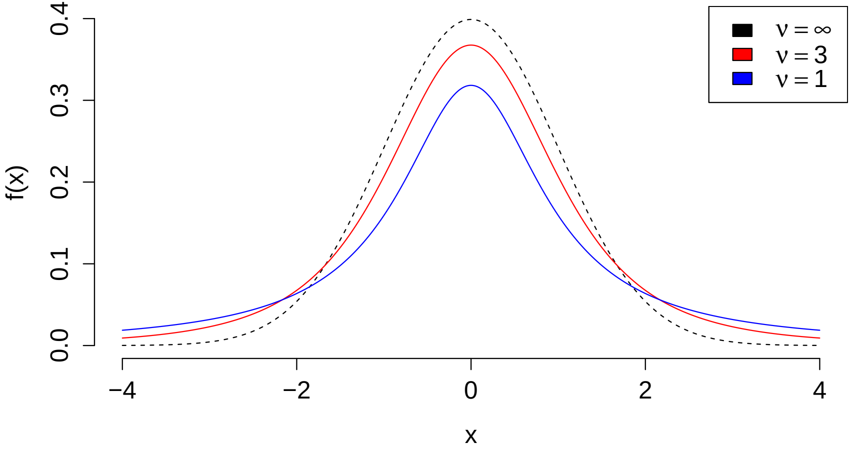
Let $X \sim N(0, 1)$ independent of $Y \sim \chi^2(\nu)$. Then,

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

where the parameter ν is the degrees of freedom.

- Support $= (-\infty, \infty)$
- As $\nu \to \infty$, $t \to$ 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

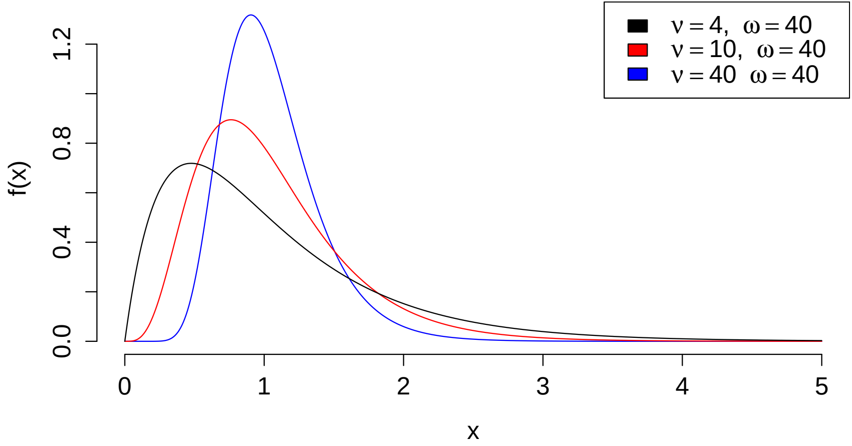
Suppose $X \sim \chi^2(\nu)$ independent of $Y \sim \chi^2(\omega)$. Then,

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

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

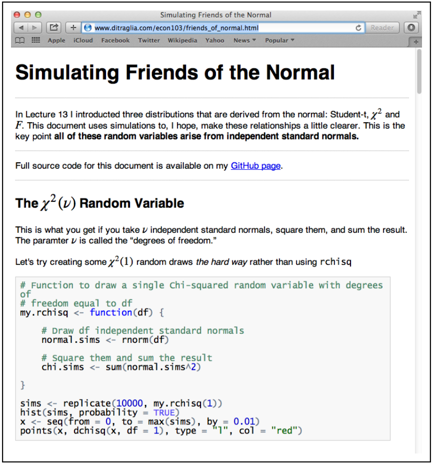
$F(x) = P(X \leq x)$ is the CDF, $Q(p) = F^{-1}(p)$ the Quantile Function

	$F(x)$	$Q(p)$
$N(\mu, \sigma^2)$	`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`)`

Mnemonic: "p" is for Probability, "q" is for Quantile.

http://fditraglia.github.io/Econ103Public/Rtutorials/friends_of_normal.html

Source Code on my Github Page



Simulating Friends of the Normal

In Lecture 13 I introducted three distributions that are derived from the normal: Student-t, χ^2 and F. This document uses simulations to, I hope, make these relationships a little clearer. This is the key point **all of these random variables arise from independent standard normals.**

Full source code for this document is available on my GitHub page.

The $\chi^2(\nu)$ Random Variable

This is what you get if you take ν independent standard normals, square them, and sum the result. The paramter ν is called the "degrees of freedom."

Let's try creating some $\chi^2(1)$ random draws *the hard way* rather than using `rchisq`

```
# Function to draw a single Chi-squared random variable with degrees
of
# freedom equal to df
my.rchisq <- function(df) {

    # Draw df independent standard normals
    normal.sims <- rnorm(df)

    # Square them and sum the result
    chi.sims <- sum(normal.sims^2)

}

sims <- replicate(10000, my.rchisq(1))
hist(sims, probability = TRUE)
x <- seq(from = 0, to = max(sims), by = 0.01)
points(x, dchisq(x, df = 1), type = "l", col = "red")
```

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`)`

Mnemonic: "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 $Y_1 = X_1^2 + X_2^2$?

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 $\Rightarrow$ $Y_1 \sim \chi^2(2)$

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 $Y_2 = (Y_1/2)/(X_3^2)$?

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

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

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 $\Rightarrow$ $Y_2 \sim F(2,1)$

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 $Z = X_3/\sqrt{Y_1/2}$?

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 $\Rightarrow$ $Z \sim t(2)$

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

Let $Y = (X_1 - \mu)^2 + (X_2 - \mu)^2$. What is the distribution of Y/σ^2?

(a) $F(2, 1)$

(b) $\chi^2(2)$

(c) $t(2)$

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

(e) None of the above

$$Y_1 \sim \chi^2(2), \quad Y_2 \sim F(2,1), \quad Z \sim t(2)$$

What is the median of Y_1?

$Y_1 \sim \chi^2(2), \quad Y_2 \sim F(2,1), \quad Z \sim t(2)$

What is the median of Y_1?

`qchisq(0.5, df = 2)` ≈ 1.4

$$Y_1 \sim \chi^2(2), \quad Y_2 \sim F(2,1), \quad Z \sim t(2)$$

What is the median of Y_1?

`qchisq(0.5, df = 2)` ≈ 1.4

What is $P(Y_2 \leq 5)$?

$Y_1 \sim \chi^2(2), \quad Y_2 \sim F(2,1), \quad Z \sim t(2)$

What is the median of Y_1?

`qchisq(0.5, df = 2)` ≈ 1.4

What is $P(Y_2 \leq 5)$?

`pf(5, df1 = 2, df2 = 1)` ≈ 0.7

$Y_1 \sim \chi^2(2), \quad Y_2 \sim F(2,1), \quad Z \sim t(2)$

What is the median of Y_1?

`qchisq(0.5, df = 2)` ≈ 1.4

What is $P(Y_2 \leq 5)$?

`pf(5, df1 = 2, df2 = 1)` ≈ 0.7

What value of c gives $P(-c \leq Z \leq c) = 0.5$?

Use Symmetry (like normal)

$c =$ `qt(0.75, df = 2)` ≈ 0.8

$$Y_1 \sim \chi^2(2), \quad Y_2 \sim F(2,1), \quad Z \sim t(2)$$

What is the median of Y_1?

`qchisq(0.5, df = 2)` ≈ 1.4

What is $P(Y_2 \leq 5)$?

`pf(5, df1 = 2, df2 = 1)` ≈ 0.7

What value of c gives $P(-c \leq Z \leq c) = 0.5$?

Use Symmetry (like normal)

$c =$ `qt(0.75, df = 2)` ≈ 0.8

or equivalently $-c =$ `qt(0.25, df = 2)` ≈ -0.8