Econ 103 – Statistics for Economists

Chapter 4 and 5: Continuous Random Variables

Mallick Hossain

University of Pennsylvania

Why Are Continuous RVs Important?

- Continuous RVs will be the foundation upon which we will build the statistical portion of the course
- Normal, chi-squared, t-, and F-distributions are all continuous random variables
- By learning the discrete RVs, you have developed an intuition about random variables that will translate into intuition about continuous RVs.
- This is a good chance to solidify your understanding of RVs.
- If you have an intuitive understanding of continuous RVs (especially the normal distribution), the rest of the semester will seem easy.

Continuous RVs – What Changes?

1. Probability Density Functions replace Probability Mass Functions (aka Probability Distributions)
2. Integrals Replace Sums

Everything Else is Essentially Unchanged!

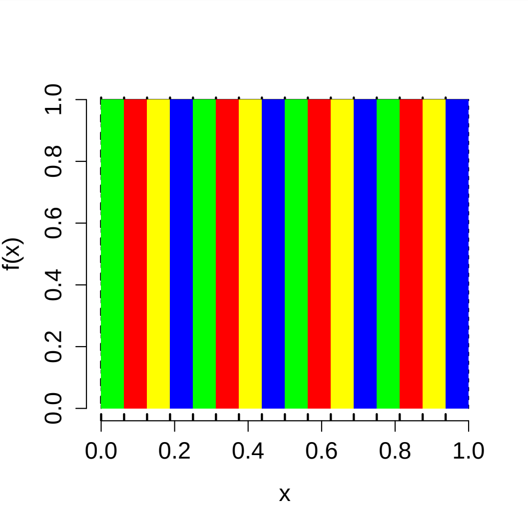
Twister

What is the probability of “Yellow?”

What is the probability of “Right Hand Blue?”

What is the probability that the spinner lands in any *particular* place?

From Twister to Density – Probability as *Area*



PDFs and CDFs for Continuous RVs

For continuous RVs, probability will be the area of *intervals* because individual points have zero probability.

Probability Density Function (PDF)

For a continuous random variable X,

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

where $f(x)$ is the *probability density function* for X.

Extremely Important

For any realization x, $P(X = x) = 0 \neq f(x)$!

Properties of PDFs

1. $\int_{-\infty}^{\infty} f(x)\, dx = 1$
2. $f(x) \geq 0$ for all x
3. $f(x)$ is *not* a probability and can be greater than one!
4. $P(X \leq x_0) = F(x_0) = \int_{-\infty}^{x_0} f(x)\, dx$

Expectation and Variance

Expectation of Continuous RVs

- Recall that for discrete RVs, $E[X] = \sum_{i=1}^{n} x_i p(x_i)$
- For continuous RVs, expected value is similar

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

- The integral replaces the sum!
- The definition holds for functions of the RV as well

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

What about all those rules for expected value?

- The only difference between expectation for continuous versus discrete is how we do the *calculation*.
- Sum for discrete; integral for continuous.
- All *properties* of expected value continue to hold!
- Includes linearity, shortcut for variance, etc.

Variance of Continuous RV

- Recall that for discrete RVs, $Var[X] = E[(X - E[X])^2]$
- For continuous RVs, expected value is the same, just with the different formula for expectation

$$\begin{aligned} Var[X] &= E[(X - E[X])^2] \\ &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \end{aligned}$$

- Shortcut formula still holds!

$$Var(X) = E[X^2] - (E[X])^2$$

We're Won't Say More About These, But Just So You're Aware of Them...

Joint Density

$$P(a \leq X \leq b \cap c \leq Y \leq d) = \int_c^d \int_a^b f(x, y)\, dxdy$$

Marginal Densities

$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$

Independence in Terms of Joint and Marginal Densities

$f_{XY}(x, y) = f_X(x) f_Y(y)$

Conditional Density

$f_{Y|X} = \frac{f_{XY}(x,y)}{f_X(x)}$

Uniform Random Variable

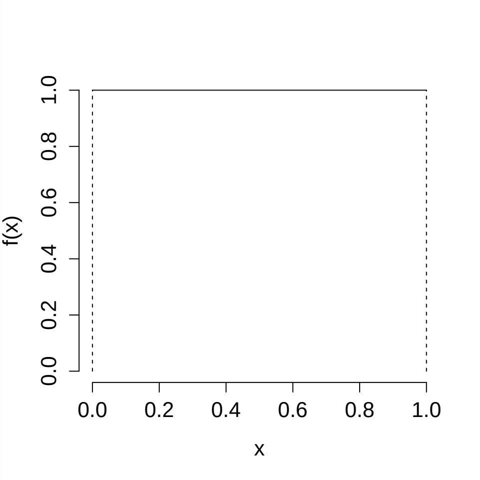
$X \sim$ **Uniform**$(0, 1)$

A Uniform(0, 1) RV is equally likely to take on *any value* in the range $[0, 1]$ and never takes on a value outside this range.

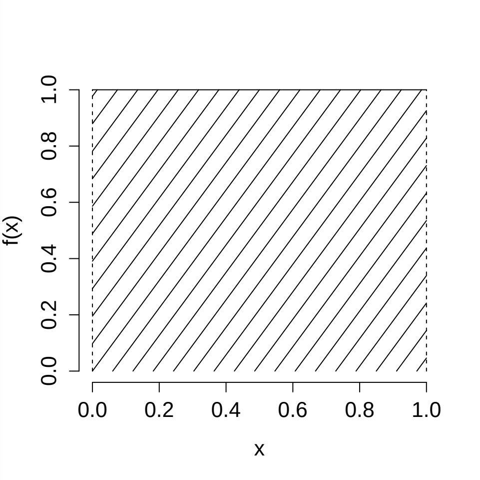
Uniform PDF

$f(x) = 1$ for $0 \leq x \leq 1$, zero elsewhere.

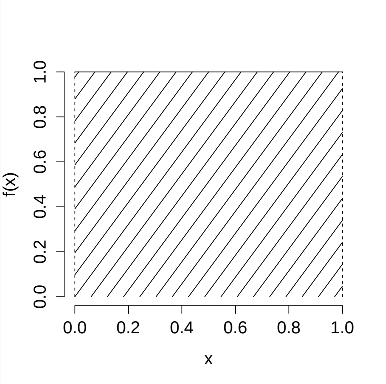
Uniform$(0, 1)$ PDF



What is the area of the shaded region?

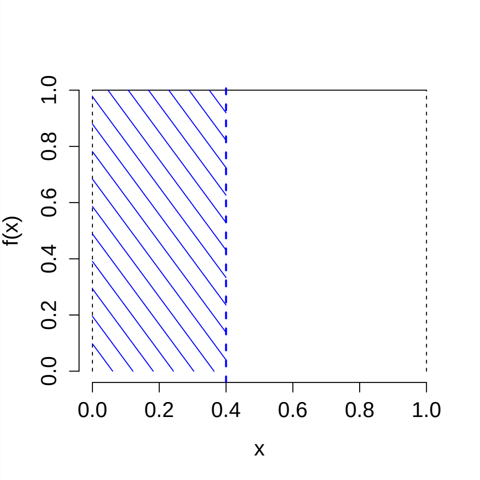


What is the area of the shaded region?

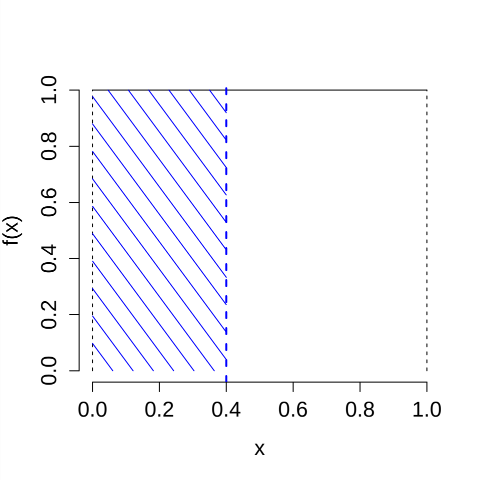


$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 1\,dx = x\big|_0^1 = 1 - 0 = 1$$

What is the area of the shaded region?



$F(0.4) = P(X \leq 0.4) = 0.4$



Relationship between PDF and CDF

Integrate pdf $\rightarrow$ CDF

$$F(x_0) = P(X \leq x_0) = \int_{-\infty}^{x_0} f(x)\,dx$$

Differentiate CDF $\rightarrow$ pdf

$$f(x) = \frac{d}{dx} F(x)$$

This is just the First Fundamental Theorem of Calculus.

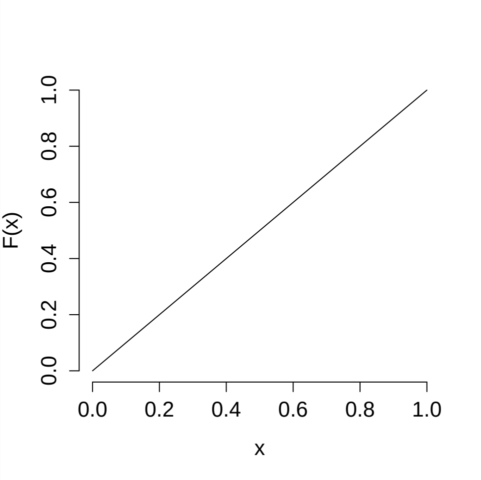
Example: Uniform$(0,1)$ RV

Integrate the pdf, $f(x) = 1$, to get the CDF

$$F(x_0) = \int_{-\infty}^{x_0} f(x)\, dx = \int_0^{x_0} 1\, dx = x|_0^{x_0} = x_0 - 0 = x_0$$

$$F(x_0) = \begin{cases} 0, x_0 < 0 \\ x_0, 0 \leq x_0 \leq 1 \\ 1, x_0 > 1 \end{cases}$$

Uniform$(0, 1)$ CDF

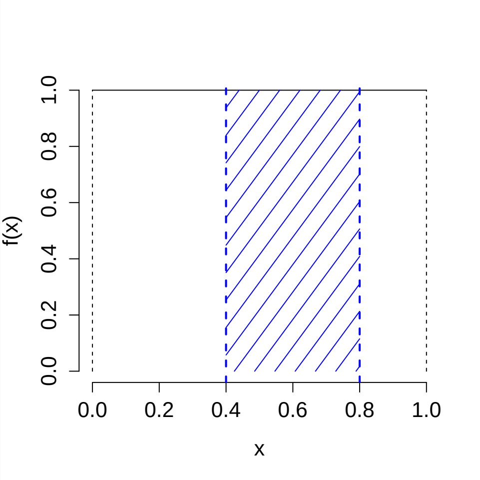


Example: Uniform$(0, 1)$ RV

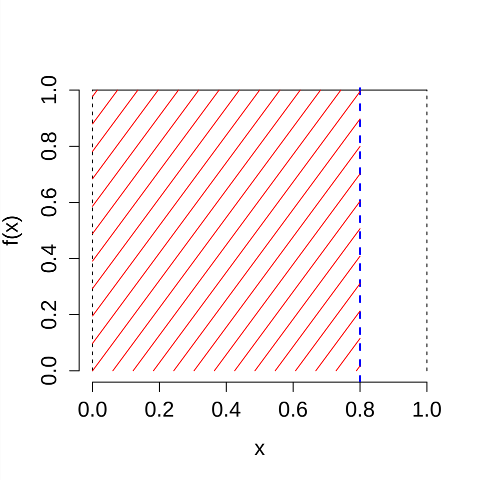
Differentiate the CDF, $F(x_0) = x_0$, to get the pdf

$$\frac{d}{dx}F(x) = 1 = f(x)$$

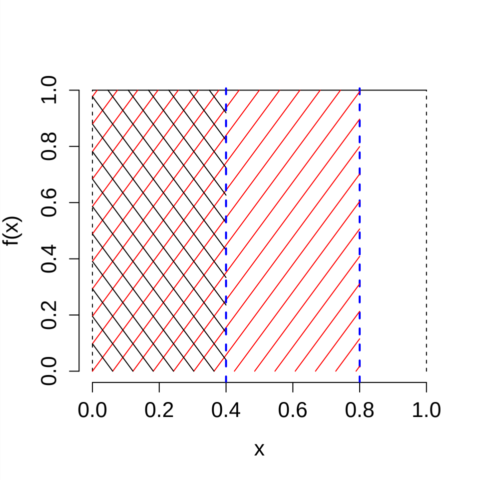
What is $P(0.4 \leq X \leq 0.8)$ if $X \sim \text{Uniform}(0, 1)$?



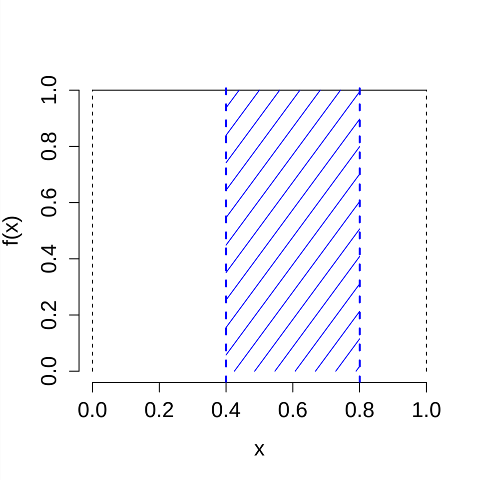
$F(0.8) = P(X \leq 0.8)$



$F(0.8) - F(0.4) = ?$



$F(0.8) - F(0.4) = P(0.4 \leq X \leq 0.8) = 0.4$



Key Idea: Probability of Interval for Continuous RV

$$\boxed{P(a \leq X \leq b) = \int_a^b f(x)\,dx = F(b) - F(a)}$$

This is just the Second Fundamental Theorem of Calculus.

Example: Uniform(0,1) Random Variable

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x \cdot 1\,dx \\
&= \left.\frac{x^2}{2}\right|_0^1 = 1/2 - 0 = 1/2
\end{aligned}
$$

Example: Uniform(0,1) Random Variable

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_0^1 x^2 \cdot 1\, dx \\ &= \left. \frac{x^3}{3} \right|_0^1 = 1/3 \end{aligned}$$

Example: Uniform$(0, 1)$ RV

$$
\begin{aligned}
Var(X) &= E\left[(X - E[X])^2\right] = E[X^2] - (E[X])^2 \\
&= 1/3 - (1/2)^2 \\
&= 1/12 \\
&\approx 0.083
\end{aligned}
$$

Much More Complicated Without the Shortcut Formula!

$$
\begin{aligned}
Var(X) &= E\left[(X - E[X])^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx \\
&= \int_0^1 (x - 1/2)^2 \cdot 1\, dx = \int_0^1 (x^2 - x + 1/4)\, dx \\
&= \left(\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4}\right)\Big|_0^1 = 1/3 - 1/2 + 1/4 \\
&= 4/12 - 6/12 + 3/12 = 1/12
\end{aligned}
$$

Check-In

So where does that leave us?

What We've Accomplished

We've covered all the basic properties of RVs on this Handout.

Where are we headed next?

Next up is the most important RV of all: the normal RV. After that it's time to do some statistics!

How should you be studying?

If you *master* the material on RVs (both continuous and discrete) and in particular the normal RV the rest of the semester will seem easy. If you don't, you're in for a rough time...

The Normal Distribution

The Normal RV

Figure 1: Standard Normal RV (PDF)

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

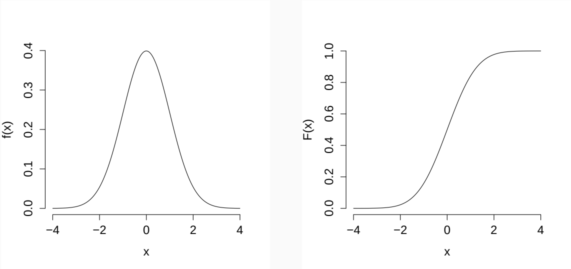
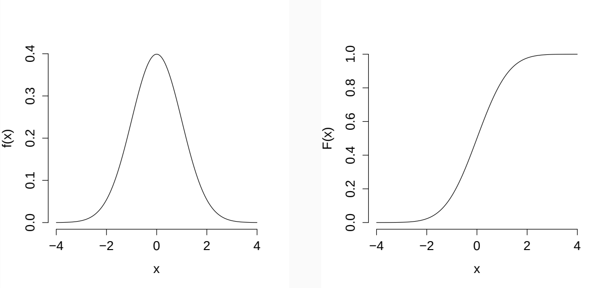


Figure 2: Standard Normal PDF (left) and CDF (Right)

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

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.
- You *do need* to know the R commands...

R Commands for the Standard Normal RV

`dnorm` – Standard Normal PDF

- Mnemonic: `d` $=$ density, `norm` $=$ normal
- Example: `dnorm(0)` gives height of $N(0,1)$ PDF at zero.

`pnorm` – Standard Normal CDF

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

`rnorm` – Simulate Standard Normal Draws

- Mnemonic: `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 $\Phi(x_0)$. It means the same thing as `pnorm`(x_0) but it's not an R command.

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

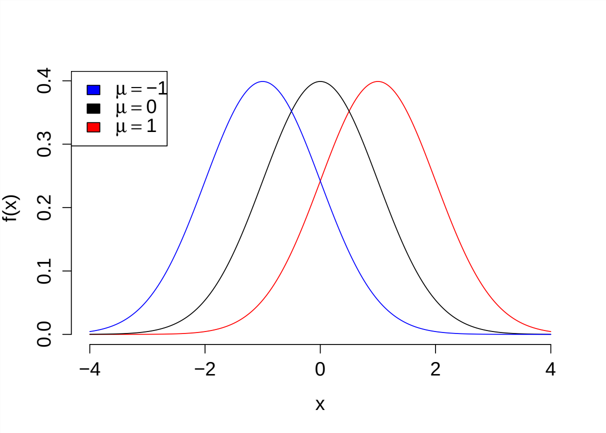


Figure 3: Blue $\mu = -1$, Black $\mu = 0$, Red $\mu = 1$

Standard Deviation: σ *scales* PDF

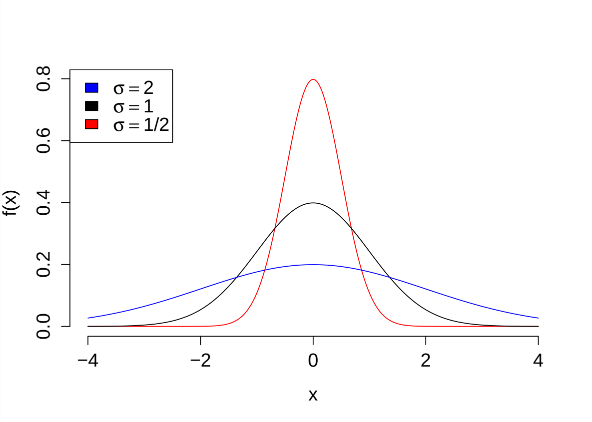


Figure 4: Blue $\sigma^2 = 4$, Black $\sigma^2 = 1$, Red $\sigma^2 = 1/4$

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

- For *any* RV X, $E[a + bX] = a + bE[X]$ and $Var(a + bX) = b^2 Var(X)$.
- 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) $N(0, 1)$

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

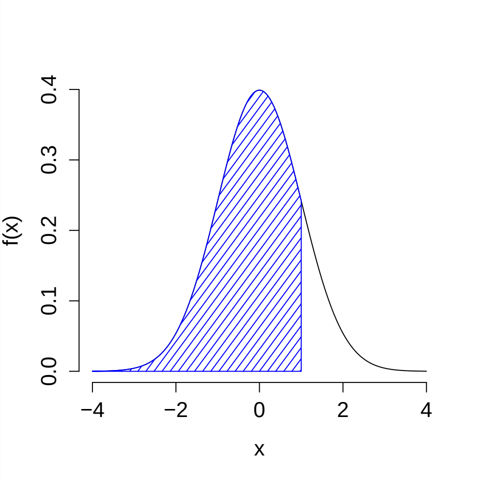
(e) $N(2\mu, 2\sigma^2)$

Where does the Empirical Rule come from?

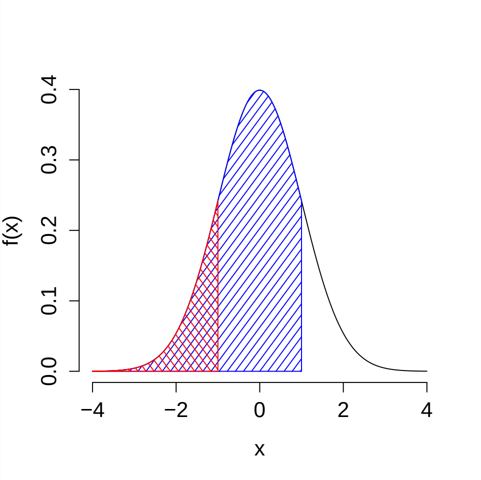
Empirical Rule

Approximately 68% of observations within $\mu \pm \sigma$
Approximately 95% of observations within $\mu \pm 2\sigma$
Nearly all observations within $\mu \pm 3\sigma$

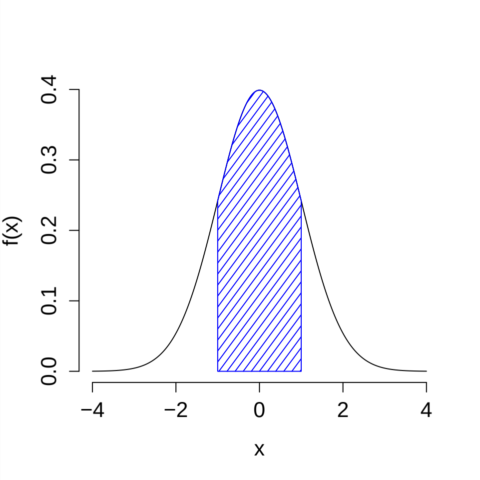
`pnorm(1)`≈ 0.84



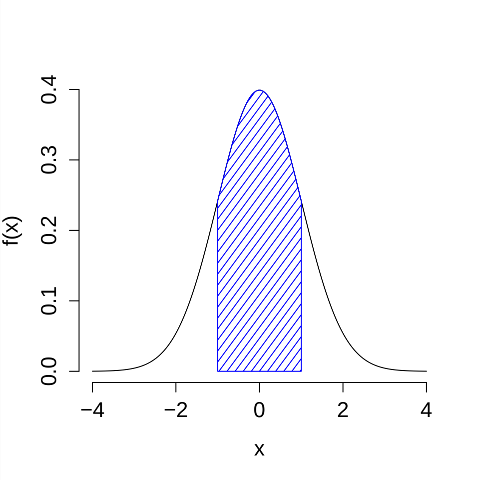
`pnorm(1) - pnorm(-1)` $\approx 0.84 - 0.16$



`pnorm(1) - pnorm(-1)`≈ 0.68



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



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

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

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

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

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 $P(Z \leq (a-\mu)/\sigma)$ 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.

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

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

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*.

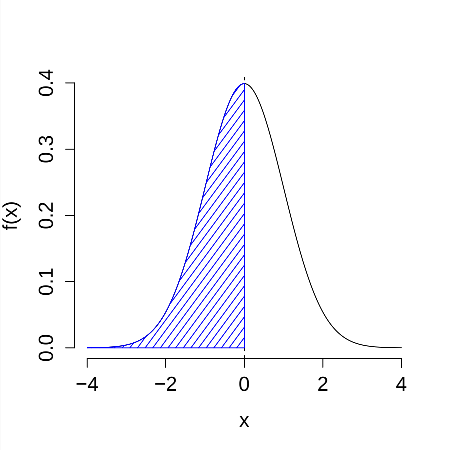
Example: Median

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

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

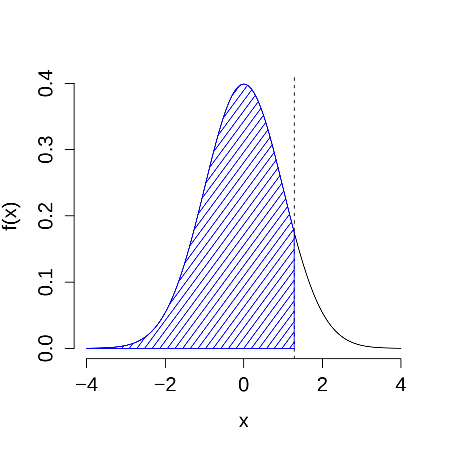
What is the median of a standard normal RV?

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



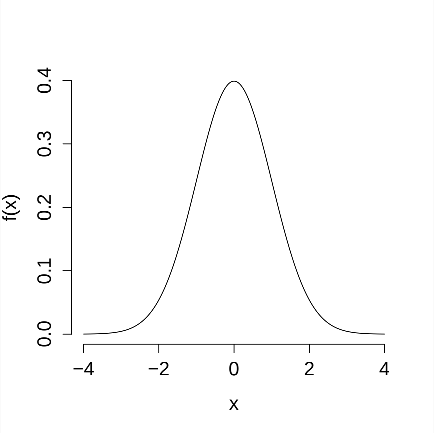
90th Percentile of a Standard Normal

`qnorm(0.9)`≈ 1.28



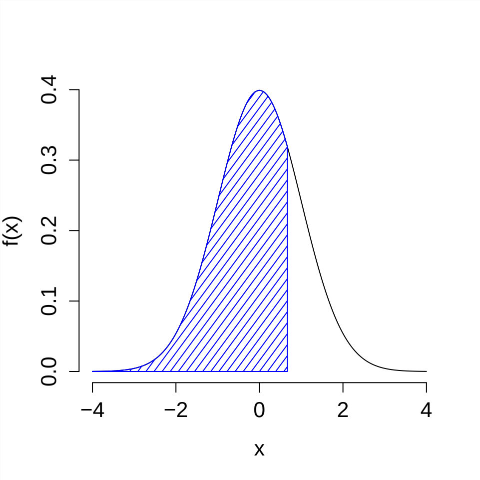
Using Quantile Function to find Symmetric Intervals

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



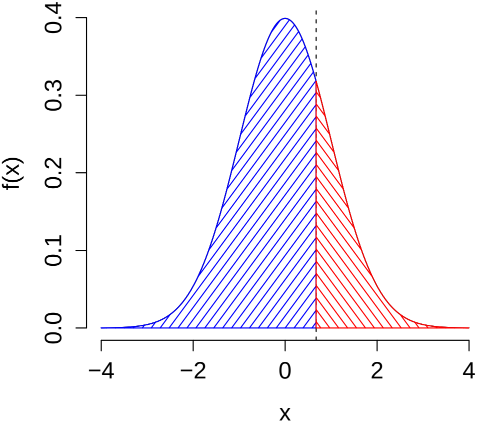
`qnorm(0.75)` ≈ 0.67

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



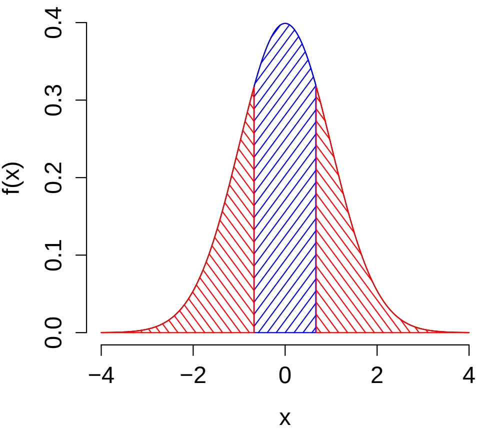
`qnorm(0.75)`≈ 0.67

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



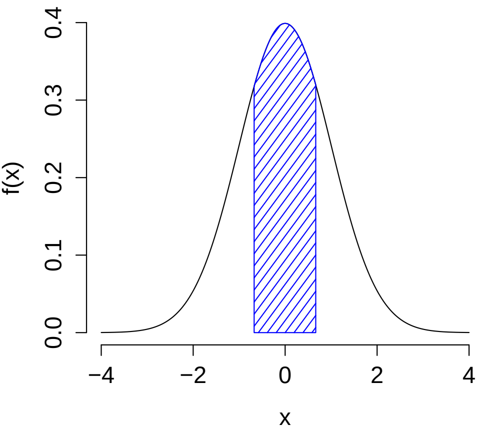
pnorm(0.67)-pnorm(-0.67)≈?

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 $P(-c \leq X \leq c) \approx 0.95$?

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`)`

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: $X \sim N(0, 16)$

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:

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

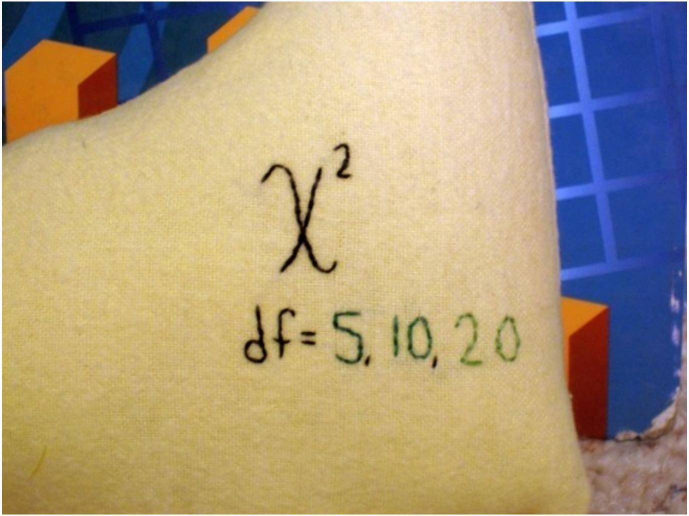
What's Next?

- We've discussed a lot about normal random variables
- Hopefully, you're starting to feel more confident about how to use them
- And maybe you're even starting to wonder if there's something more to the normal random variable
- What if...?

Chi Squared Distribution

Figure 5: PDF for χ^2-Distribution

Figure 6: χ^2 PDF – Halloween Edition



χ^2
df = 5, 10, 20

χ^2 Random Variable

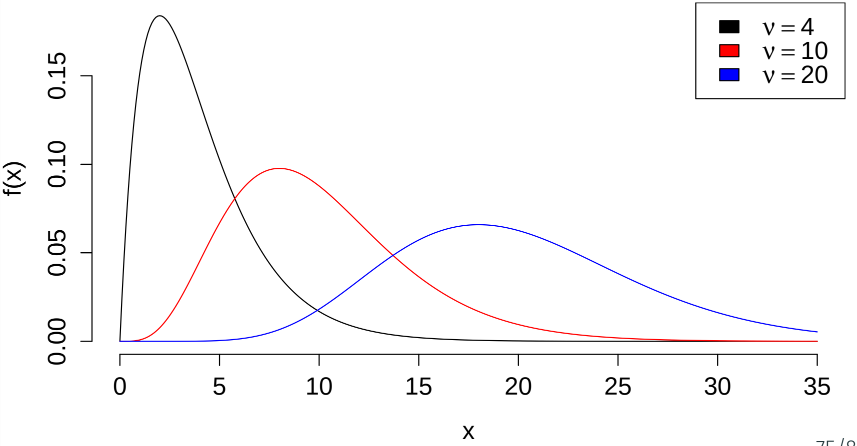
Let $X_1, \ldots, X_\nu \sim$ iid $N(0,1)$. Then,

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

where the parameter ν is the *degrees of freedom*

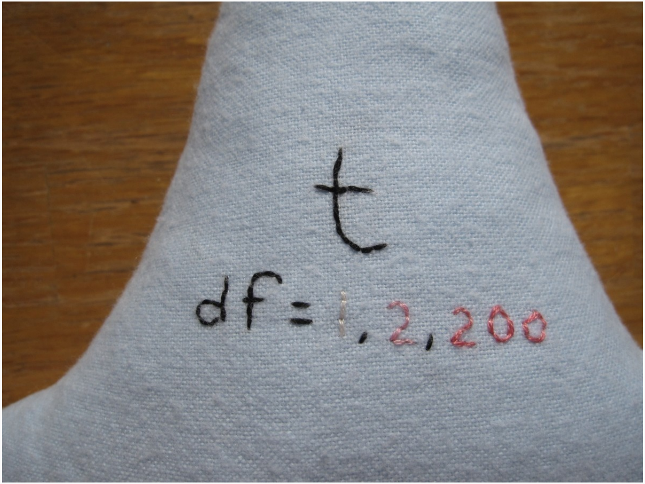
Support = $(0, \infty)$

χ^2 PDFs



t-Distribution

Figure 7: PDF for Student-t Distribution


t
df=1,2,200

Student-t Random Variable

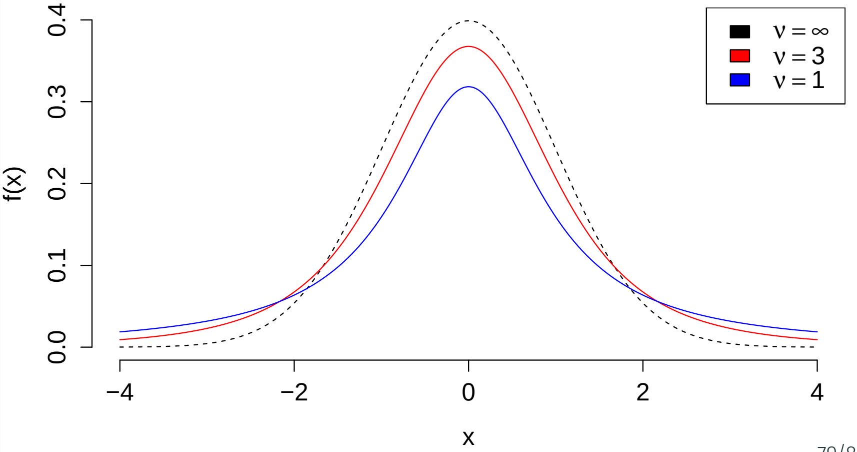
Let $X \sim N(0, 1)$ independent of $Y \sim \chi^2(\nu)$. Then,

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

where the parameter ν is the degrees of freedom.

- 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 Distribution

F Random Variable

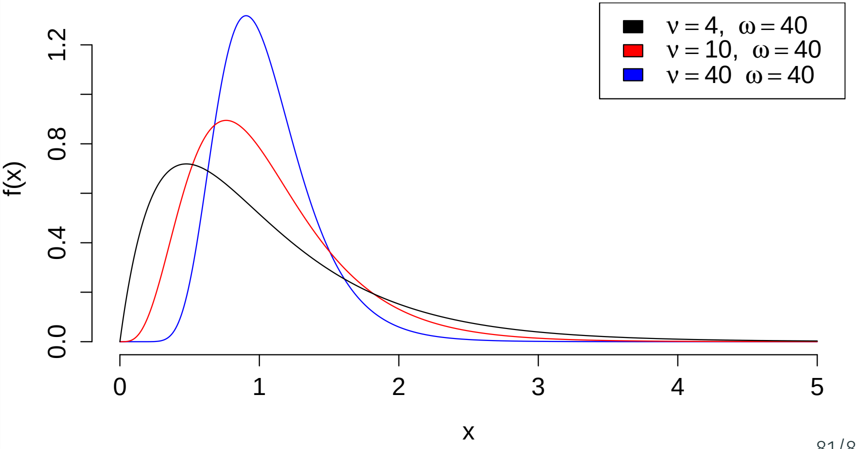
Suppose $X \sim \chi^2(\nu)$ independent of $Y \sim \chi^2(\omega)$. Then,

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

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

Support = $(0, \infty)$

F PDFs



R Commands

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.

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.

Practice

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?

`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