

# Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Lecture # 12

# Continuous Distributions – Part II

# The Most Important RV of All

# Normal Random Variable

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

Parameters:  $\mu = E[X]$ ,  $\sigma^2 = \text{Var}(X)$

Support:  $(-\infty, +\infty)$

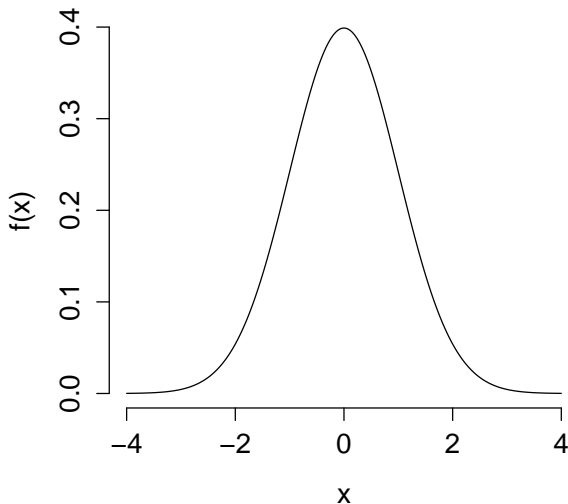
Probability Density Function

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

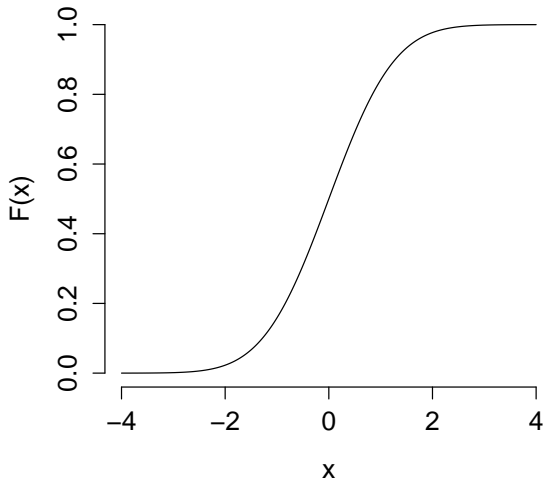
No Explicit Formula for CDF (use computer instead)

$$F(x_0) = \int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\} dx$$

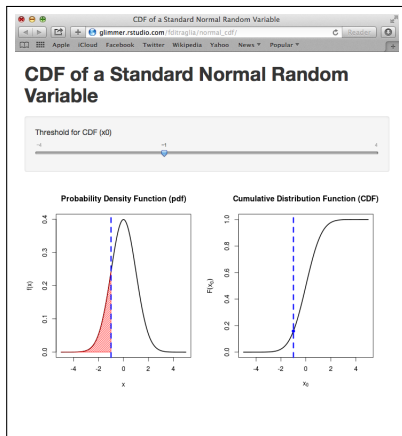
## Normal PDF Centered at the Mean (Here $\mu = 0$ , $\sigma^2 = 1$ )



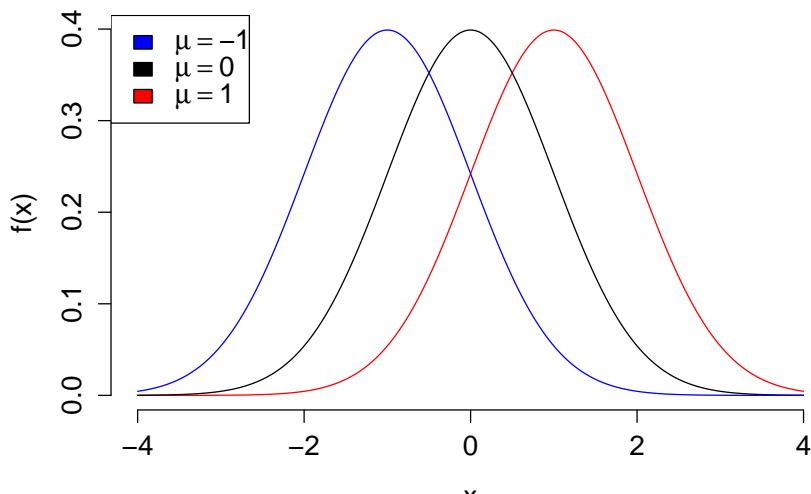
## Normal CDF ( $\mu = 0, \sigma^2 = 1$ )



[http://glimmer.rstudio.com/fditraglia/normal\\_cdf/](http://glimmer.rstudio.com/fditraglia/normal_cdf/)

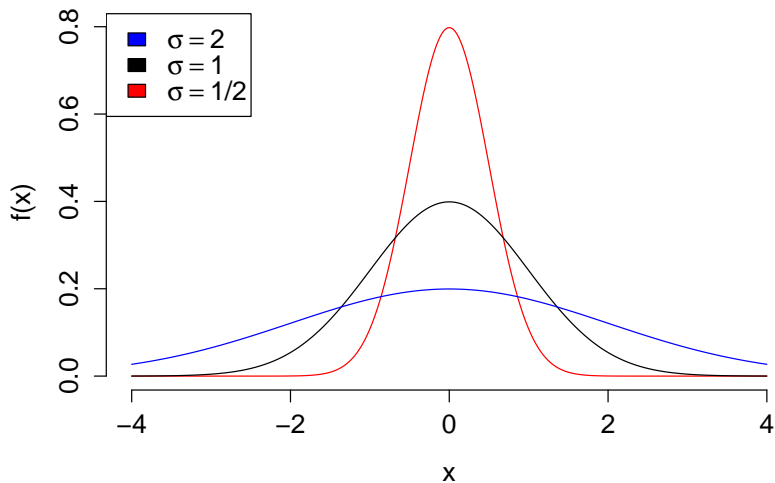


## Different Means, Same Variance





## Same Mean, Different Variances



## Linear Function of Normal RV is a Normal RV

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

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

### Important

- ▶ Using what we know about expectations of linear functions, no surprise what mean and variance are.
- ▶ Surprise is that the linear combination is *normal*
- ▶ Linear trans. does not preserve, e.g., Bernoulli or 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)  $N(0, 1)$



Figure : Standard Normal Distribution (PDF)

# Standard Normal Distribution: $N(0, 1)$



## Standard Normal Distribution: $N(0, 1)$

Mean = 0, Variance = Standard Deviation = 1

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Special symbol for Standard Normal CDF (no closed form):

$$\Phi(x_0) = \int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

R Command:  $\Phi(x_0) = \text{pnorm}()$

# Where does the Empirical Rule come from?

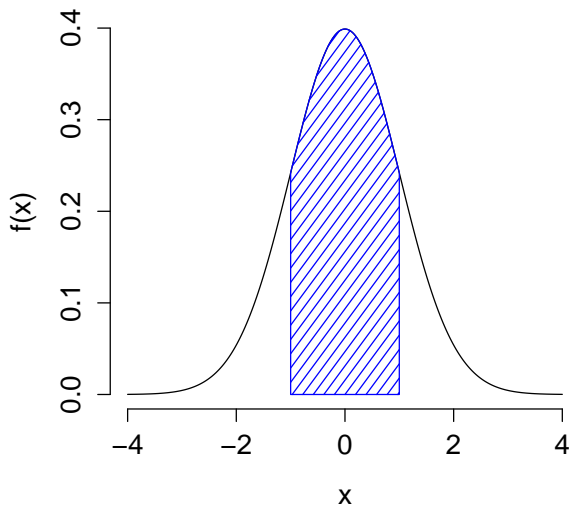
## Empirical Rule

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

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

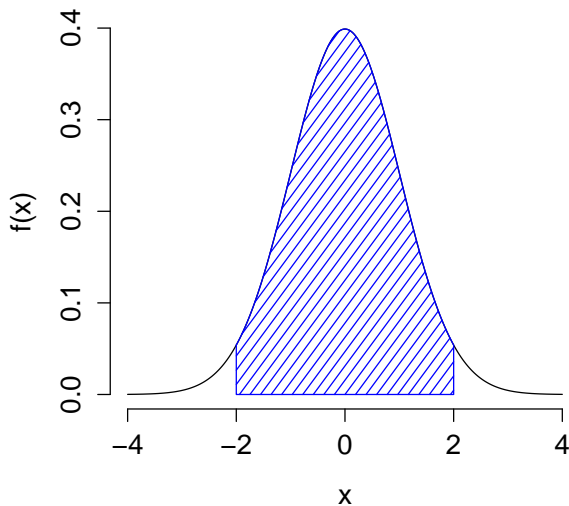
Nearly all observations within  $\mu \pm 3\sigma$

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





Middle 95% of  $N(0, 1) \Rightarrow$  approx.  $(-2, 2)$



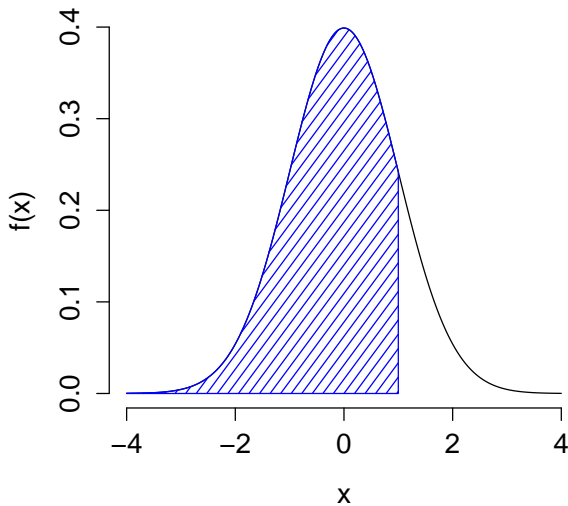
## More Formally...

$$\int_{-1}^1 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \approx 0.68$$

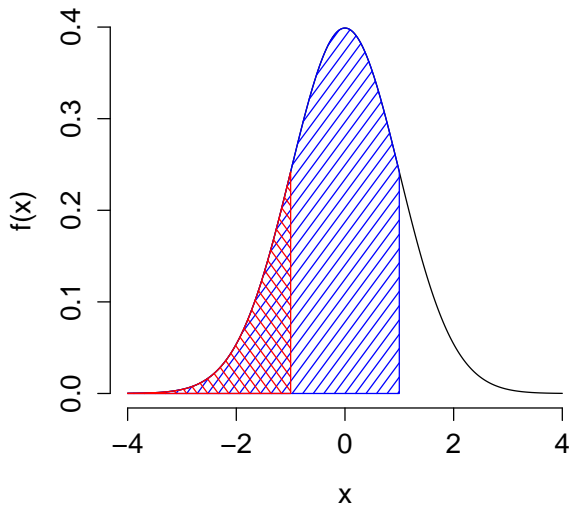
$$\int_{-2}^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \approx 0.95$$

But how do we know this?

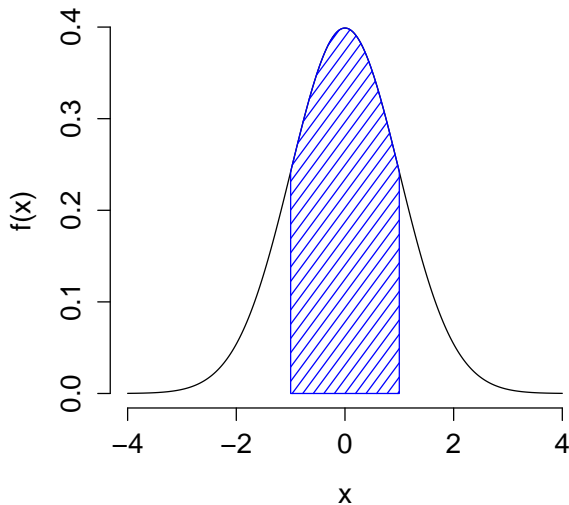
$$\Phi(1) = \text{pnorm}(1) \approx 0.84$$



$$\Phi(1) - \Phi(-1) = \text{pnorm}(1) - \text{pnorm}(-1) \approx 0.84 - 0.16$$



$$\Phi(1) - \Phi(-1) = \text{pnorm}(1) - \text{pnorm}(-1) \approx 0.68$$



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

$$\begin{aligned} P(-2 \leq X \leq 2) &= \Phi(2) - \Phi(-2) \\ &= \text{pnorm}(2) - \text{pnorm}(-2) \\ &\approx 0.95 \end{aligned}$$

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

$$\begin{aligned}P(-2 \leq X \leq 2) &= \Phi(2) - \Phi(-2) \\&= \text{pnorm}(2) - \text{pnorm}(-2) \\&\approx 0.95\end{aligned}$$

$$\begin{aligned}P(-3 \leq X \leq 3) &= \Phi(3) - \Phi(-3) \\&= \text{pnorm}(3) - \text{pnorm}(-3) \\&\approx 1\end{aligned}$$

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

$$P(X \leq a) =$$



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

$$P(X \leq a) = P(X - \mu \leq a - \mu)$$

$$=$$

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

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

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)  $\Phi(a)$
- (b)  $1 - \Phi(a)$
- (c)  $\Phi(a)/\sigma - \mu$
- (d)  $\Phi\left(\frac{a - \mu}{\sigma}\right)$
- (e) None of the above.

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

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

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

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



Which of these is  $P(X \geq b)$ ?

(a)  $\Phi(b)$

(b)  $1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$

(c)  $1 - \Phi(b)$

(d)  $1 - (\Phi(b)/\sigma - \mu)$

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

$$P(X \geq b) =$$

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

$$P(X \geq b) = 1 - P(X \leq b) =$$



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

$$P(X \geq b) = 1 - P(X \leq b) = 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right)$$

=

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

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

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

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

Where  $Z$  is a standard normal random variable.

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

$$P(a \leq X \leq b) =$$

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

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right)$$

=

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

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

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

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

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

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

Where  $Z$  is a standard normal random variable.



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



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

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

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

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



What is  $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$ ?

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

$$\begin{aligned}P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= P\left(-2 \leq \frac{X - \mu}{\sigma} \leq 2\right) \\&= P(-2 \leq Z \leq 2) \\&= \Phi(2) - \Phi(-2) \\&= \text{pnorm}(2) - \text{pnorm}(-2) \\&\approx 0.95\end{aligned}$$

## 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

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

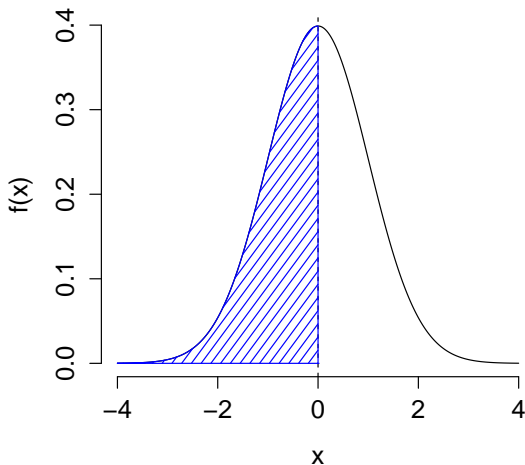
What is the median of a standard normal RV?



## What is the median of a standard normal RV?



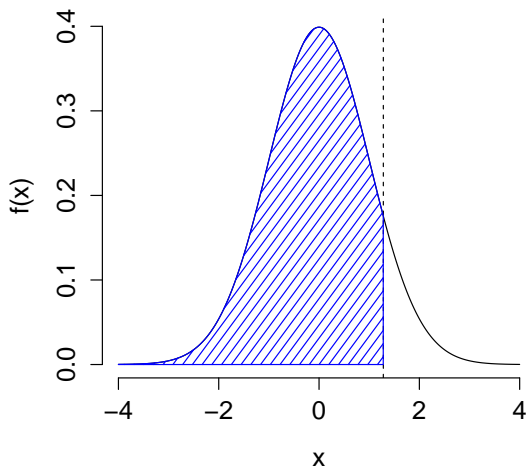
By symmetry,  $Q(0.5) = 0$ . R command: `qnorm()`





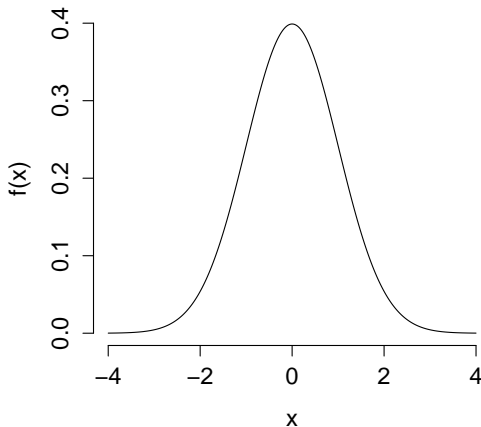
## 90th Percentile of a Standard Normal

$$\text{qnorm}(0.9) \approx 1.28$$



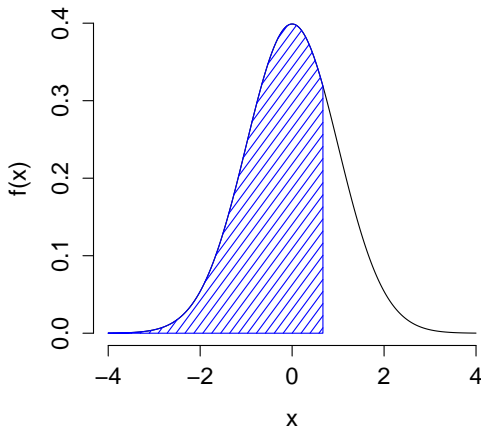
## Using Quantile Function to find Symmetric Intervals

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



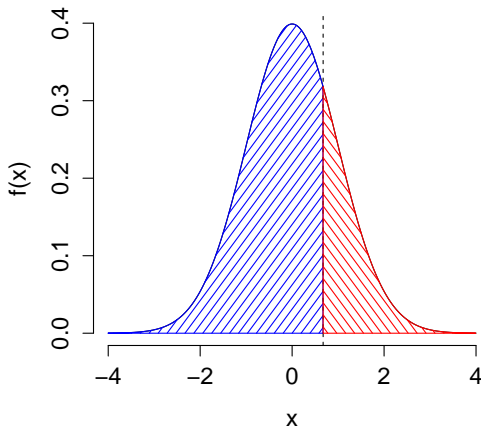
$$\text{qnorm}(0.75) \approx 0.67$$

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



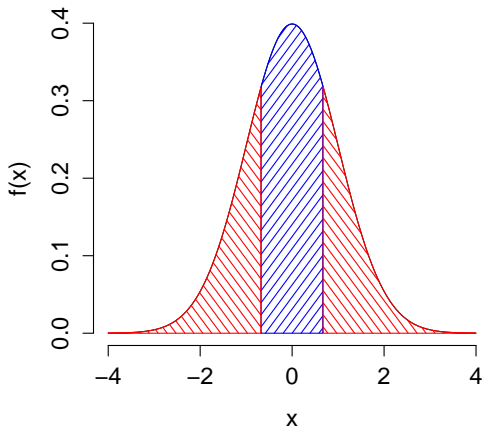
$$\text{qnorm}(0.75) \approx 0.67$$

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



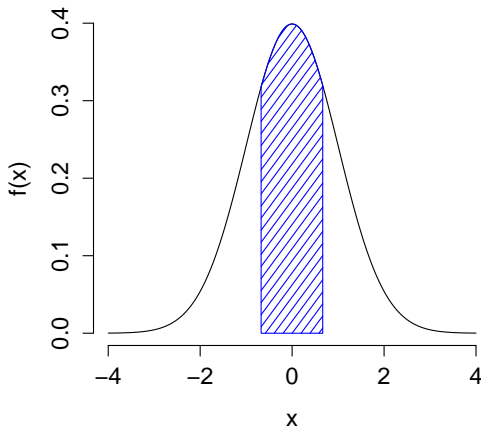
$$\text{pnorm}(0.67) - \text{pnorm}(-0.67) \approx ?$$

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



$$\text{pnorm}(0.67) - \text{pnorm}(-0.67) \approx 0.5$$

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 $F(x)$	<code>pnorm(x, mean = <math>\mu</math>, sd = <math>\sigma</math>)</code>
Quantile Function $Q(p)$	<code>qnorm(p, mean = <math>\mu</math>, sd = <math>\sigma</math>)</code>

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:

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

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

One Way:

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

An Easier Way:

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

Suppose  $X$  has mean  $\mu_x$  variance  $\sigma_x^2$  and is independent of  $Y$ , which has mean  $\mu_y$  variance  $\sigma_y^2$ . Let  $a, b$  be constants.

What is  $E[aX + bY]$ ?

Suppose  $X$  has mean  $\mu_x$  variance  $\sigma_x^2$  and is independent of  $Y$ , which has mean  $\mu_y$  variance  $\sigma_y^2$ . 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  $Var(aX + bY)$ ?

Suppose  $X$  has mean  $\mu_x$  variance  $\sigma_x^2$  and is independent of  $Y$ , which has mean  $\mu_y$  variance  $\sigma_y^2$ . 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.

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:

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

## Important

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

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



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