

Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Lecture # 9

Continuous Distributions – Part II

Last Time: Continuous RVs, Probability As Area

Probability Density Function (pdf)

- ▶ $\int_a^b f(x) dx = P(a \leq X \leq b)$
- ▶ $f(x) \geq 0$ for all x in the support
- ▶ $f(x) \neq P(X = x)$, can be greater than one

Cumulative Distribution Function

- ▶ $F(x_0) \equiv P(X \leq x_0) = \int_{-\infty}^{x_0} f(x) dx$
- ▶ First Fundamental Theorem of Calculus: $f(x) = F'(x)$

Last Time: Uniform(0, 1) RV

Intuition

Equally likely to take on any value on its support: $[0, 1]$

Probability Density Function

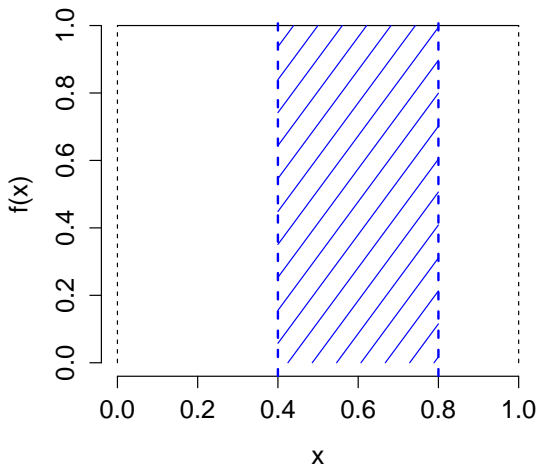
$f(x) = 1$ for $x \in [0, 1]$, zero otherwise

Cumulative Distribution Function

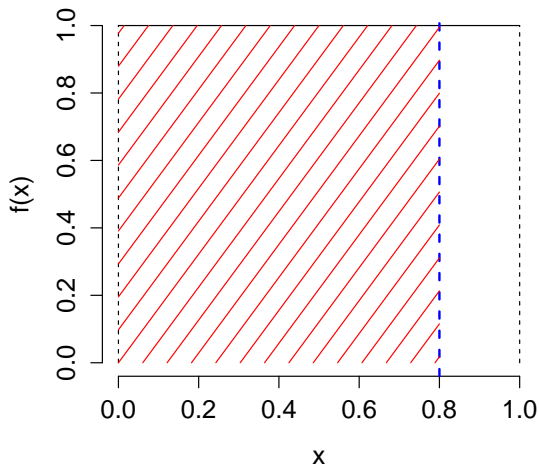
$$F(x_0) = \begin{cases} 0, & x_0 < 0 \\ x_0, & 0 \leq x_0 \leq 1 \\ 1, & x_0 > 1 \end{cases}$$

Key Idea: Probability of Intervals

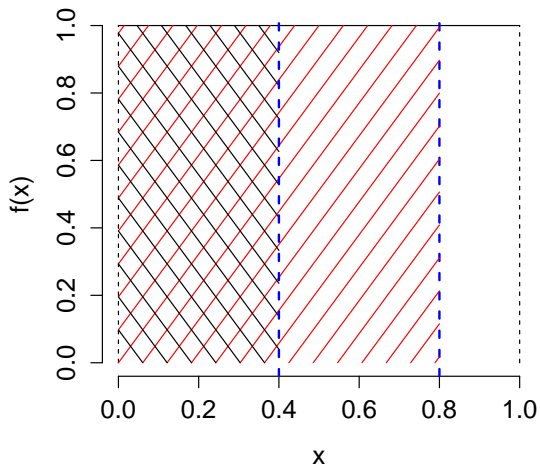
What is $P(0.4 \leq X \leq 0.8)$ if $X \sim \text{Uniform}(0, 1)$?



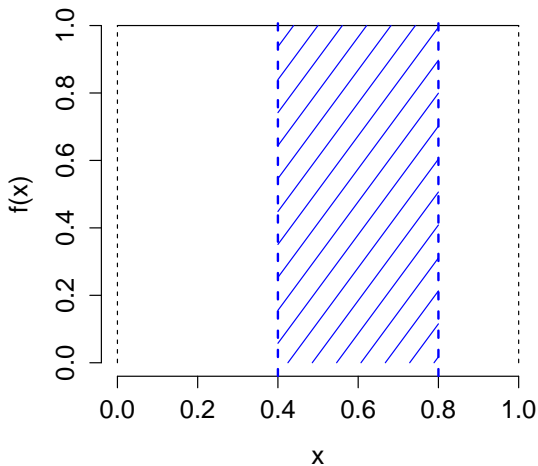
$$F(0.8) = P(X \leq 0.8)$$



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



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



Probability of Interval for Continuous RV

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

This is just the Second Fundamental Theorem of Calculus.

Expected Value for Continuous RVs

$$\int_{-\infty}^{\infty} xf(x) dx$$

Remember: Integrals Replace Sums!

Example: Uniform(0,1) Random Variable



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

Expected Value of a Function of a Continuous RV

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

Example: Uniform(0, 1) RV

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

Once we have defined expected value for continuous RVs, we can use everything we know about variance, covariance, etc. from discrete RVs!

Variance of Continuous RV

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

where

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

Shortcut formula still holds for continuous RVs!

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

Example: Uniform(0,1) Random Variable



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

Much More Complicated Without the Shortcut Formula!

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

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

Joint Density

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

Marginal Densities

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

Independence in Terms of Joint and Marginal Densities

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

Conditional Density

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

We've now covered everything on the
[Random Variables Handout](#)

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

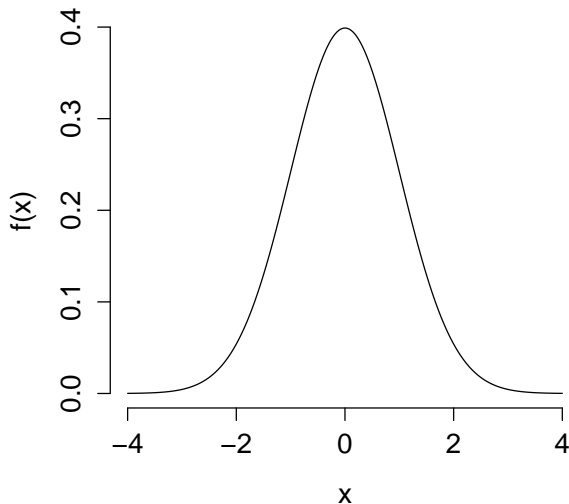
Probability Density Function

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

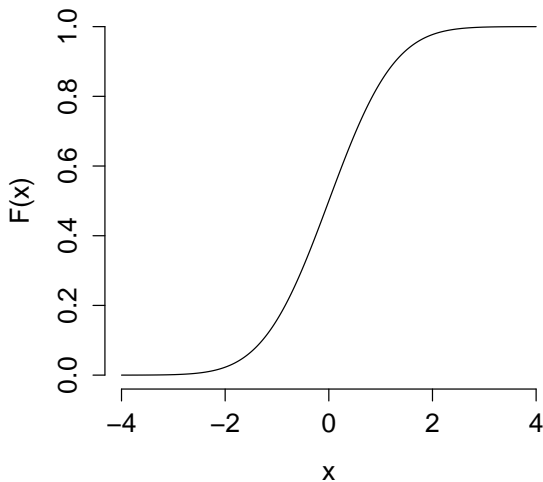
No Explicit Formula for CDF (use computer instead)

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

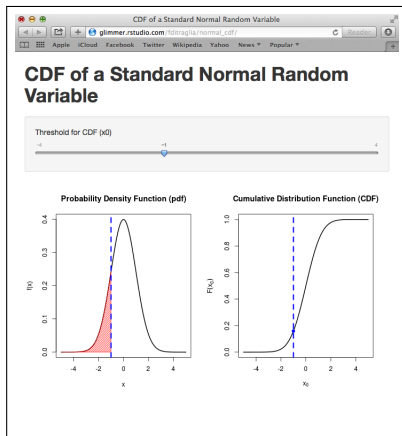
Normal PDF Centered at the Mean (Here $\mu = 0$, $\sigma^2 = 1$)



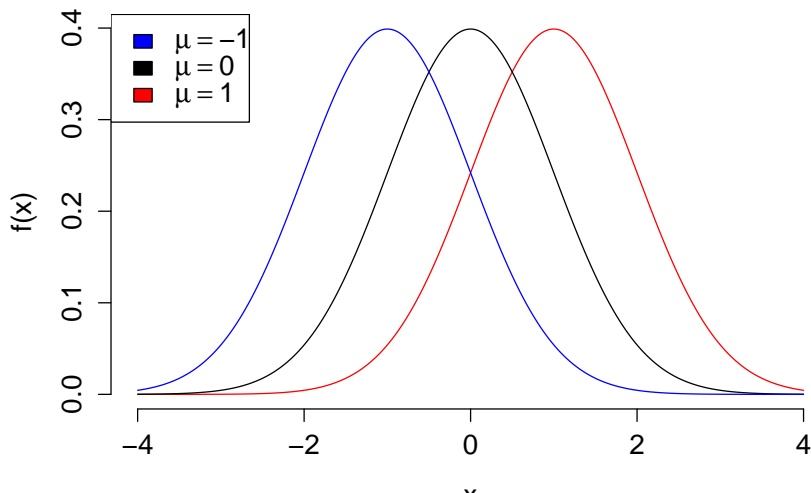
Normal CDF ($\mu = 0, \sigma^2 = 1$)



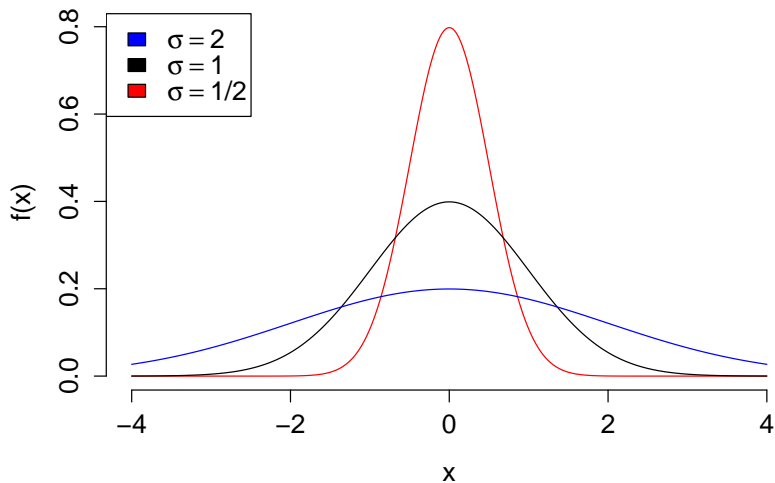
http://glimmer.rstudio.com/fditraglia/normal_cdf/



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?

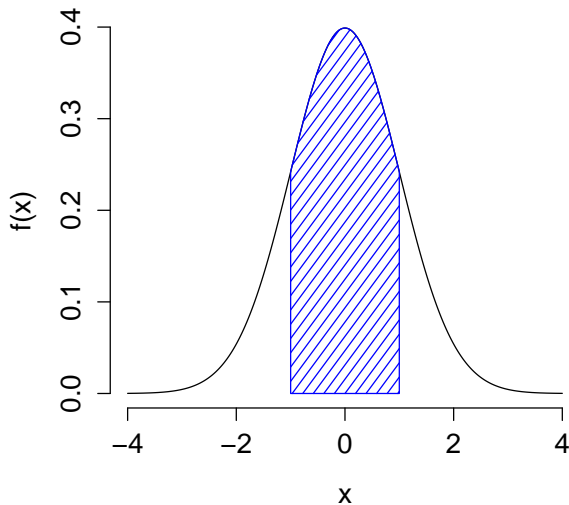
Empirical Rule

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

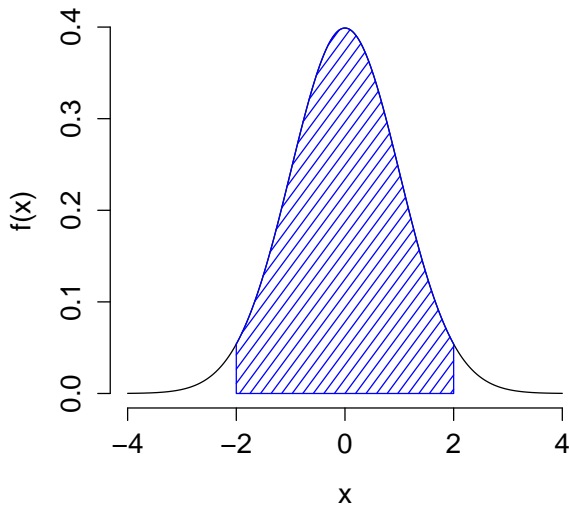
Approximately 95% of observations within $\mu \pm 2\sigma$

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. $(-2, 2)$



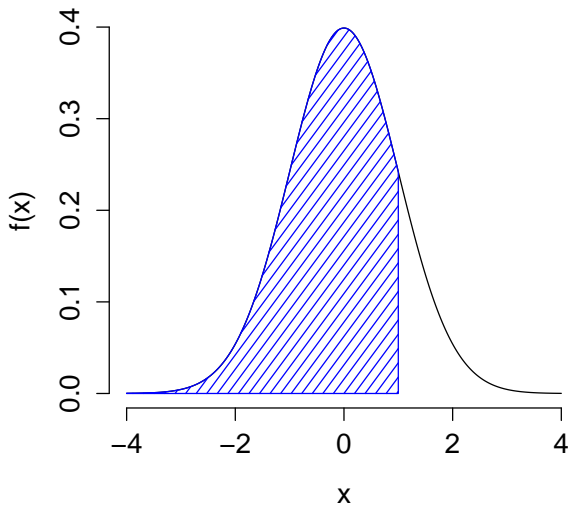
More Formally...

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

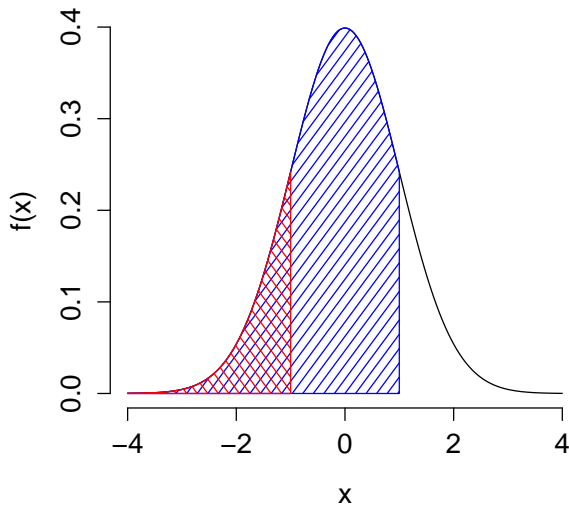
$$\int_{-2}^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \approx 0.95$$

But how do we know this?

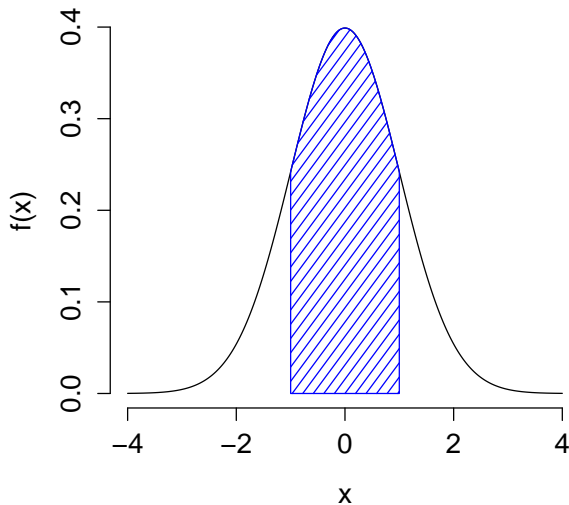
$$\Phi(1) = \text{pnorm}(1) \approx 0.84$$



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



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



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

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

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

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

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

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

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

Where Z is a standard normal random variable.

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

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

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

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