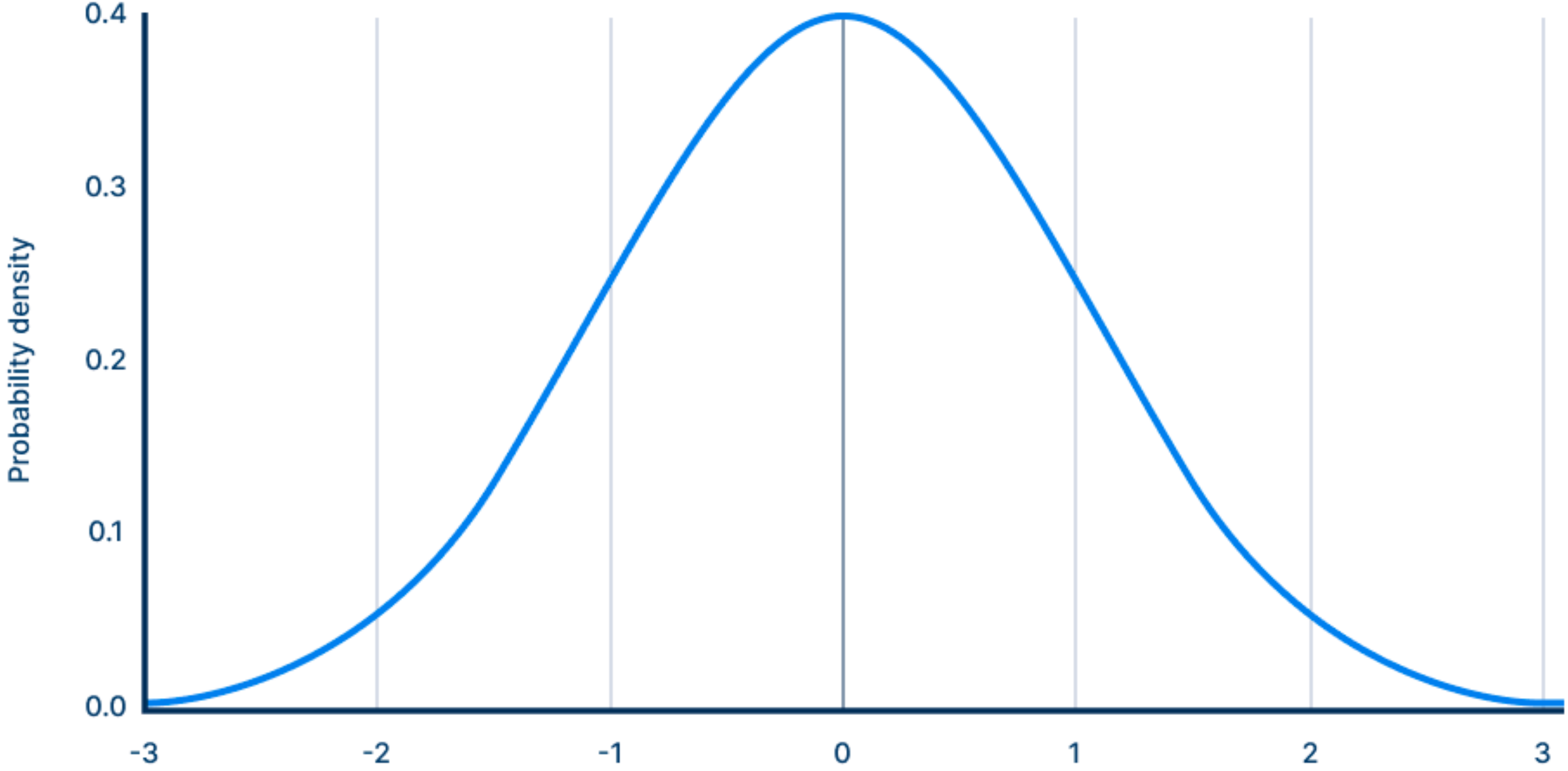
Lesson 018
The Normal Distribution

Friday, October 20 & Monday, October 23

"Do you think that the midterm will be graded on a curve?"

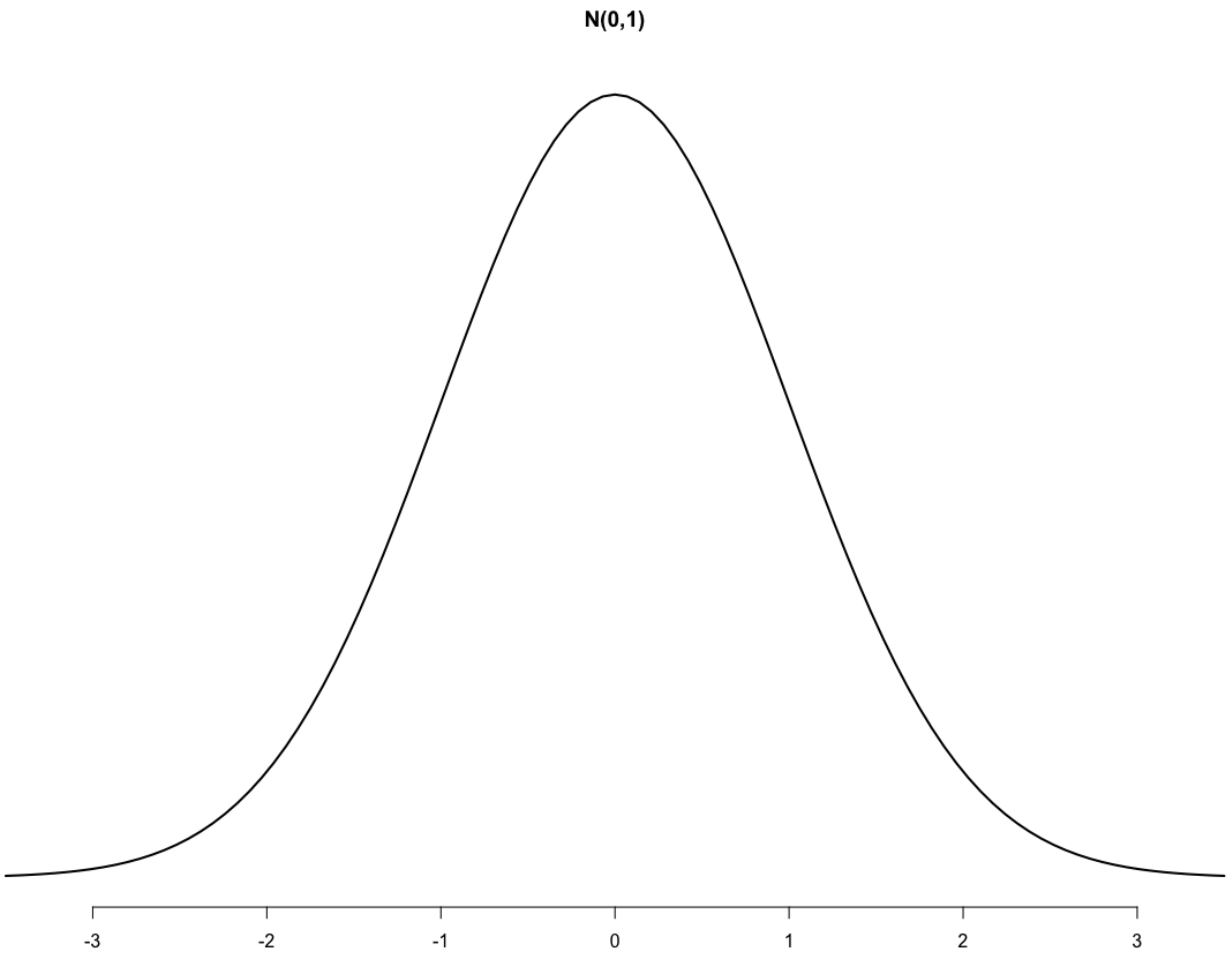

Probability density
0.4
0.3
0.2
0.1
0.0
-3
-2
-1
0
1
2
3

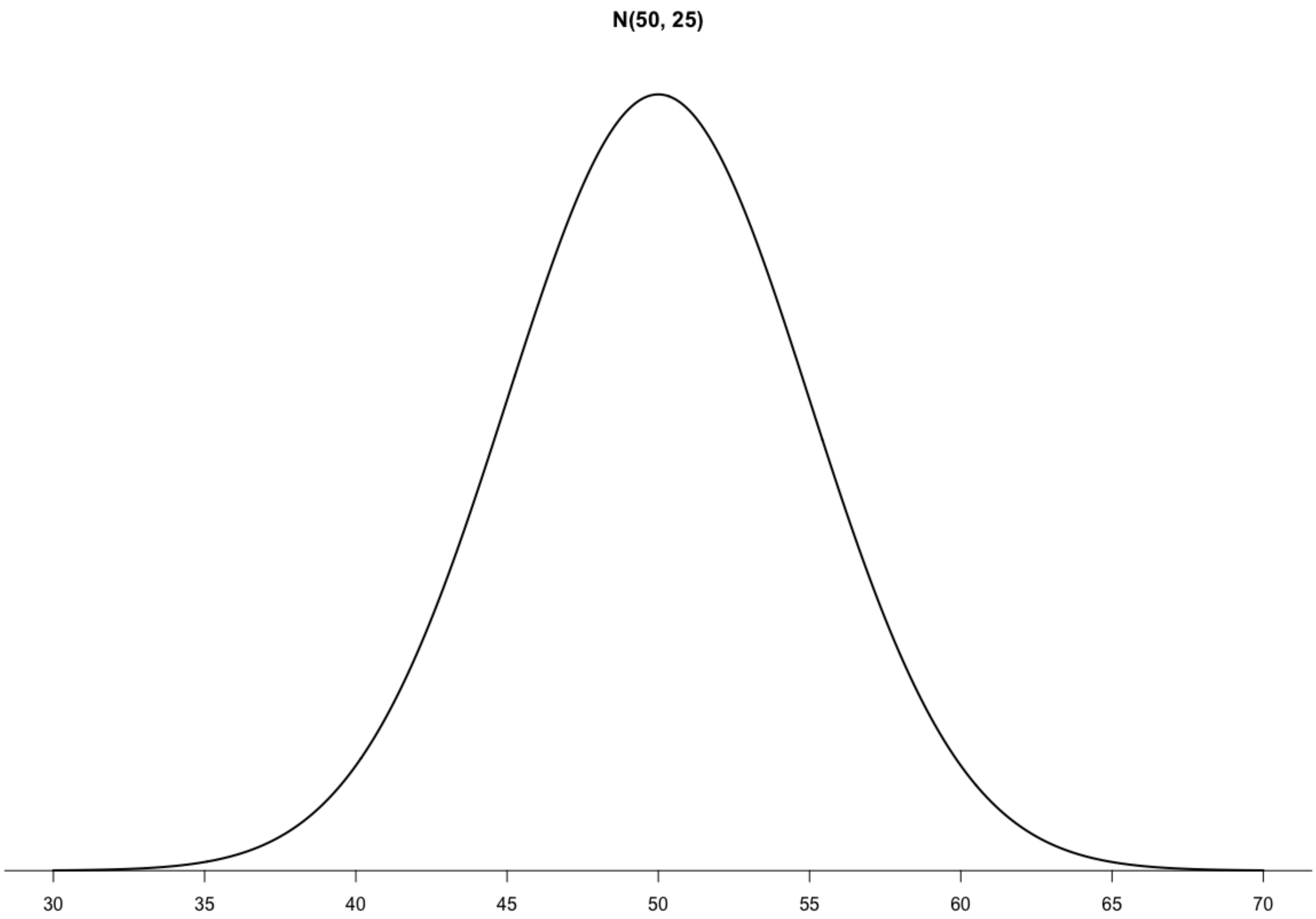
The Normal Distribution

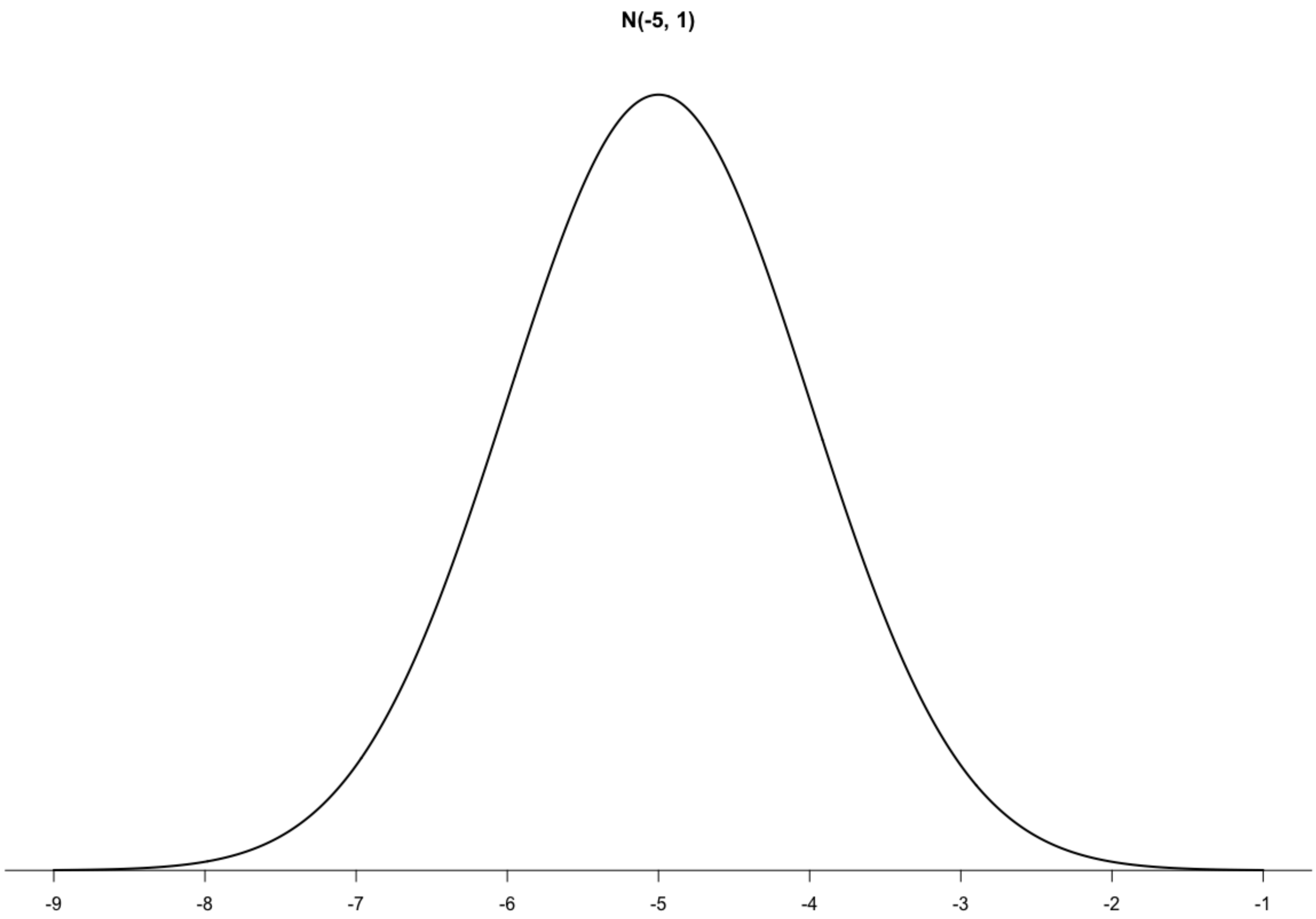
- The **normal distribution** is the single most important distribution in all of statistics and probability.
- Also referred to as the **Gaussian distribution**.
- It characterizes many natural phenomena (heights, weights, reaction times, etc.)
- It also characterizes much "limiting behaviour" (more on this later)

The Normal Distribution

- It is a symmetric distribution, parameterized by its mean and variance.
- μ is the mean of the distribution.
- σ^2 is the variance of the distribution.
- We write $X \sim N(\mu, \sigma^2)$
  - Thus, $E[X] = \mu$ and $\text{var}(X) = \sigma^2$.


N(0,1)
-3
-2
-1
0
1
2
3


N(50, 25)
30
35
40
45
50
55
60
65
70


N(-5, 1)
-9
-8
-7
-6
-5
-4
-3
-2
-1

The Normal PDF

- There is no closed form expression for the CDF.
- The PDF is defined on $X \in (-\infty, \infty)$ and is

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

- This can be integrated for probabilities (in theory).

The Normal Distribution: Probability Calculations

- The form of the normal PDF makes calculations challenging, in general.
- Instead of using the PDF directly, we typically rely on external tools.
  - Probability tables: allow you to look up probability values for the **standard normal,** $N(0,1)$.
  - Software allows you to compute normal probabilities for any normal distribution.

Minitab Demonstration

If $X \sim N(\mu, \sigma^2)$, what is the probability that $X = 2$?

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

0%

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

0%

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

0%

0

0%

1

0%

If $X \sim N(\mu, \sigma^2)$, what is the probability that $X \leq 2$?

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

0%

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

0%

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

0%

0

0%

1

0%

If $X \sim N(\mu, \sigma^2)$, what is the probability density at $X = 2$?

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

0%

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

0%

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

0%

0

0%

1

0%

If $X \sim N(\mu, \sigma^2)$, what is the probability that $X \geq 2$?

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

0%

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

0%

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

0%

0

0%

1

0%

Standardization

- If $X \sim N(\mu, \sigma^2)$, then $aX + b$ remains a normal distribution.
- $E[aX + b] = aE[X] + b = a\mu + b$.
- $\text{var}(aX + b) = a^2\text{var}(X) = a^2\sigma^2$

$$Z = \frac{X - \mu}{\sigma}$$

Standardization

- We will have $Z \sim N(0,1)$.
- Note:

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

- Probability statements about X can be made into probability statements about Z.
- This is called **standardization**.

The Standard Normal

- If we have $Z \sim N(0,1)$ then:
  - We denote the PDF of Z as $\varphi(z)$.
  - We denote the CDF of Z as $\Phi(z)$.
  - We have $E[Z] = 0$ and $\text{var}(Z) = 1$.
- We can go back to arbitrary $X \sim N(\mu, \sigma)$ using $X = \sigma Z + \mu$.

Suppose that $X \sim N(2,4)$. If we wish to solve $P(X \leq 3)$, which probability is equivalent?

$P(Z \leq 3)$ with $Z = \frac{X-2}{2}$.

0%

$P(Z \leq 0.5)$ with $Z = \frac{X-2}{2}$.

0%

$P(Z \leq 0.5)$ with $Z = \frac{X-2}{4}$.

0%

$P(Z \leq 3)$ with $Z = \frac{X-2}{4}$.

0%

Suppose that $X \sim N(0, 4)$. If we wish to solve $P(-2 \leq X \leq 3)$, which probability is equivalent?

$P(-2 \leq Z \leq 3)$ with $Z = \frac{X}{2}$.

0%

$P(-2 \leq Z \leq 3)$ with $Z = \frac{X}{4}$.

0%

$P(-1 \leq Z \leq 1.5)$ with $Z = \frac{X}{2}$.

0%

$P(-1 \leq Z \leq 1.5)$ with $Z = \frac{X}{4}$.

0%

Suppose that $X \sim N(-2, 9)$. If we wish to solve $P(X \geq -3)$, which probability is equivalent?

$P(Z \geq -\frac{5}{3})$ with $Z = \frac{X-2}{3}$.

0%

$P(Z \geq -\frac{5}{3})$ with $Z = \frac{X+2}{3}$.

0%

$P(Z \geq -\frac{1}{3})$ with $Z = \frac{X+2}{9}$.

0%

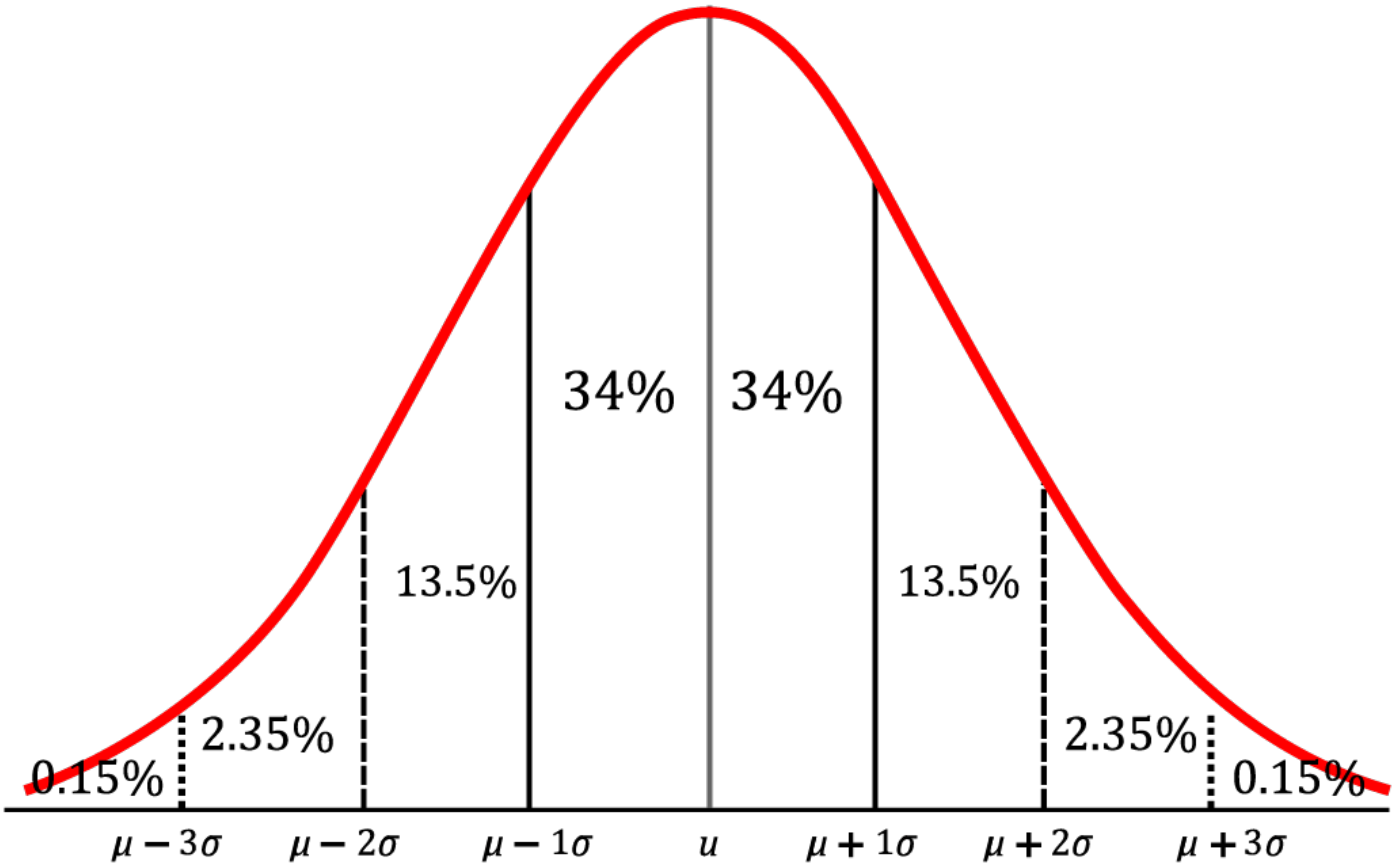
$P(Z \geq -\frac{1}{3})$ with $Z = \frac{X+2}{3}$.

0%

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051

The Empirical Rule

- If $X \sim N(\mu, \sigma^2)$ then nearly all density falls within $\mu \pm 3\sigma$.
  - $68\,\%$ of observations fall in the range $\mu \pm \sigma$.
  - $95\,\%$ of observations fall in the range $\mu \pm 2\sigma$.
  - $99.7\,\%$ of observations fall in the range $\mu \pm 3\sigma$.


34%
34%
13.5%
13.5%
2.35%
2.35%
0.15%
0.15%
μ − 3σ
μ − 2σ
μ − 1σ
u
μ + 1σ
μ + 2σ
μ + 3σ

$$X \sim N(10,\ 9)$$

$$P(X \geq 13) = 1 - P(X \leq 13)$$

$$P(X \leq \mu + \sigma) \approx 0.5 + 0.34 = 0.84$$

$$P(X \geq 13) \approx 1 - 0.84 = 0.16$$

Critical Values for the Normal Distribution

- Recall that we defined $\eta(p)$ to be the value such that $P(X \leq \eta(p)) = p$.
- If we take $Z \sim N(0,1)$, then $\eta(p)$ has $\Phi(\eta(p)) = p$.
- In this case we denote $\eta(p)$ as Z_p, and call it a **critical value** for Z.
- Note that $Z_p = -Z_{1-p}$. Why?

Common Critical Values

- $Z_{0.975} = 1.96$ and $Z_{0.025} = -1.96$.
  - Gives $P(-1.96 \leq Z \leq 1.96) = 0.95$.
- $Z_{0.95} = 1.645$ and $Z_{0.05} = -1.645$.
  - Gives $P(-1.645 \leq Z \leq 1.645) = 0.90$
- $Z_{0.995} = 2.58$ and $Z_{0.005} = -2.58$.
  - Gives $P(-2.58 \leq Z \leq 2.58) = 0.99$.

Critical Values of Arbitrary Normal Distributions

- If we know Z_p, and we want $\eta(p)$ for $X \sim N(0,1)$ we can use the same transformation.
- $\eta(p) = \sigma Z_p + \mu$.

$$
\begin{aligned}
P\left\{X \le \eta(p)\right\} &= P\left(Z \le \frac{\eta(p) - \mu}{\sigma}\right) \\
&= P\left(Z \le \frac{\sigma Z_p + \mu - \mu}{\sigma}\right) \\
&= P\left(Z \le Z_p\right) \\
&= p
\end{aligned}
$$

If we know that $Z_{0.975}$ is 1.96, then what is $\eta(0.975)$ for $X \sim N(-2, 5)$?

$(-2)(1.96) + \sqrt{5}$

0%

$\sqrt{5}(1.96) - 2$

0%

$\sqrt{5}(1.96) + 2$

0%

$(-2)(1.96) - \sqrt{5}$

0%

We know that $Z_{0.95} = 1.645$. If $\eta(p) = 2$, and $\mu = 1$, then what is σ^2?

$\left(\frac{2-1}{1.645}\right)^2 = 0.3695$ — 0%

$\left(\frac{2-1}{1.645}\right) = 0.6079$ — 0%

$\left(\frac{1.645-1}{2}\right)^2 = 0.1040$ — 0%

$\left(\frac{1.645-1}{2}\right)^2 = 0.3225$ — 0%

Normal Approximation to the Binomial

- Suppose that $X \sim \text{Bin}(n, p)$.
- We know that $E[X] = np$ and $\text{var}(X) = np(1 - p)$.
- If n is *sufficiently large*, then we get that $X \dot{\sim} N(np, np(1 - p))$.
- General rule of thumb $np \geq 10$ and $n(1 - p) \geq 10$.
- Why use the approximation?

Continuity Correction

- Since X is discrete, $\{X \leq 2\} = \{X \leq 2.5\}$.
- We need to use this information for **continuous approximations** to **discrete distributions**.
- If we wish to use $X \leq x$ we should consider $X \leq x + 0.5$ in the approximation.
- If we wish to use $X \geq x$ we should consider $X \geq x - 0.5$ in the approximation.

If we wish to solve $P(X \leq 2)$ using a continuous approximation, where W is approximating X, what probability do we compute for W?

$P(W \leq 2)$.

0%

$P(W \leq 2.5)$.

0%

$P(W \leq 1.5)$.

0%

$P(W \in [1.5, 2.5])$.

0%

If we wish to solve $P(X < 3)$ using a continuous approximation, where W is approximating X, what probability do we compute for W?

$P(W \leq 3)$. 0%

$P(W \leq 2.5)$. 0%

$P(W \leq 3.5)$. 0%

$P(W \leq 2)$. 0%

Example 1

- Suppose $X \sim \text{Bin}(100, 0.25)$. What is $P(X \leq 25)$?
- First, $X \dot{\sim} N(25, 18.75)$. Call this W.
- Instead of $\{W \leq 25\}$ we consider $\{W \leq 25.5\}$.
- $P\left(X \leq 25\right) \approx P\left(W \leq 25.5\right)$.
- $P(W \leq 25.5) = P\left(Z \leq \dfrac{25.5 - 25}{18.75}\right) = \Phi\left(\dfrac{2}{75}\right)$

Example 2

- Suppose $X \sim$ Bin $\left(50, \dfrac{1}{3}\right)$. What is $P(X > 25)$?
- First, $X \dot\sim N\left(\dfrac{50}{3}, \dfrac{100}{9}\right)$. Call this W.
- Instead of $\{W > 25\}$ we consider $\{W \geq 25.5\}$.
    - Note: $\{X > 25\} = \{X \geq 26\}$, so $\{W \geq 26\} \to \{W \geq 25.5\}$.
- $P\left(X > 25\right) \approx P\left(W \geq 25.5\right) = 1 - \Phi\left(\dfrac{25.5 - 50/3}{100/3}\right)$.