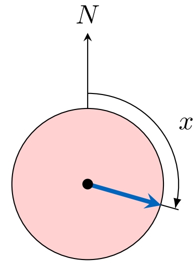
CONTINUOUS RANDOM VARIABLES

Consider a random experiment where a spinner is spun, and the continuous random variable X represents the angle spun, measured in degrees over the interval $[0, 360)$.



The probability that X assumes any precise value, such as $x = 125.333\ldots$, is zero due to the infinite number of possible outcomes within a continuous range. However, we can calculate the probability that X lies within an interval, such as $[90, 180]$. We can think of this as summing the probabilities over tiny subintervals within this range:

$$\begin{aligned} P(90 \leq X \leq 180) &= \sum_{x \in [90,180]} P(x \leq X < x + dx) \\ &= \sum_{x \in [90,180]} \frac{P(x \leq X < x + dx)}{dx} \cdot dx \end{aligned}$$

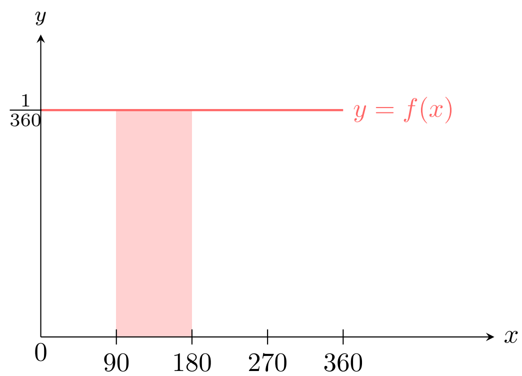
Here, dx represents an infinitesimally small interval, and the ratio, $\frac{P(x \leq X < x+dx)}{dx}$ is the probability per unit length. We define this as the probability density function, $f(x) = \frac{P(x \leq X < x+dx)}{dx}$. By the definition of integration, this summation becomes:

$$P(90 \leq X \leq 180) = \int_{90}^{180} f(x)\, dx$$

For a uniform spinner, the probability is evenly distributed across all angles, so the pdf is constant: $f(x) = \frac{1}{360}$ for $x \in [0, 360)$. Thus:

$$\begin{aligned} P(90 \leq X \leq 180) &= \int_{90}^{180} \frac{1}{360}\, dx \\ &= \left[\frac{x}{360}\right]_{90}^{180} \\ &= \frac{180}{360} - \frac{90}{360} \\ &= \frac{1}{4} \end{aligned}$$

Hence, the probability of spinning an angle between 90 and 180 degrees is $\frac{1}{4}$.



A DEFINITIONS

A.1 CONTINUOUS RANDOM VARIABLE

Definition Continuous Random Variable

A random variable is **continuous** if its set of possible values is an entire interval of real numbers. A continuous random variable can take any value within its range, meaning there is an infinite number of possibilities.

A.2 PROBABILITY DENSITY FUNCTION

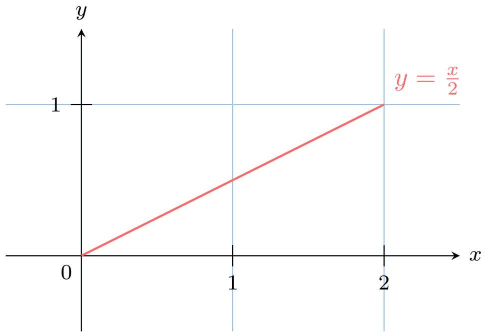
A probability density function describes the likelihood of a continuous random variable taking values within a specific range. Unlike discrete random variables, where probabilities are assigned to individual outcomes, continuous variables use the pdf to compute probabilities over intervals via integration.

Definition Probability Density Function

A function f is a **probability density function** (pdf) on the interval $[a, b]$ if:

- $f(x) \geq 0$ for all $x \in [a, b]$ (non-negative everywhere),
- $\int_a^b f(x)\,dx = 1$ (the total area under the curve equals 1).

Ex: The random variable X takes values on $[0, 2]$ with density $f(x) = \frac{x}{2}$.



Verify that f is a probability density function on $[0, 2]$.

Answer:

- $f(x) = \frac{x}{2} \geq 0$ for all $x \in [0, 2]$, since $x \geq 0$.
- Compute the total area:

$$\begin{aligned}\int_0^2 f(x)\,dx &= \int_0^2 \frac{x}{2}\,dx \\ &= \left[\frac{x^2}{4}\right]_0^2 \\ &= \frac{2^2}{4} - 0 = 1\end{aligned}$$

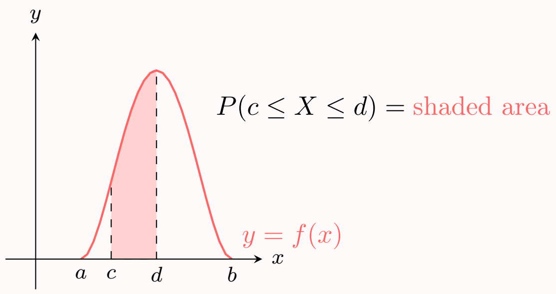
Since both conditions hold, $f(x) = \frac{x}{2}$ is a valid pdf on $[0, 2]$.

Definition Density of a Continuous Random Variable

A random variable X with values on $[a, b]$ has a density f, if the probability that X lies between c and d ($c, d \in [a, b]$) is:

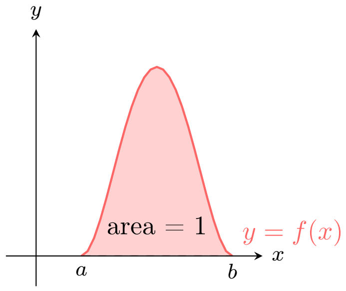
$$P(c \leq X \leq d) = \int_c^d f(x)\,dx$$

This represents the area under the curve $y = f(x)$ from $x = c$ to $x = d$.



Remark

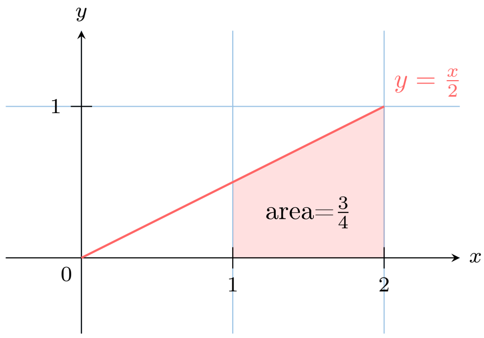
- Since $f(x) \geq 0$, $P(c \leq X \leq d) \geq 0$.
- Since $\int_a^b f(x)\,dx = 1$, $P(a \leq X \leq b) = 1$.



Ex: The random variable X with values on $[0,2]$ has density $f(x) = \frac{x}{2}$. Find $P(1 \leq X \leq 2)$.

Answer:

$$\begin{aligned} P(1 \leq X \leq 2) &= \int_1^2 \frac{x}{2}\, dx \\ &= \left[\frac{x^2}{4}\right]_1^2 \\ &= \frac{2^2}{4} - \frac{1^2}{4} \\ &= 1 - \frac{1}{4} \\ &= \frac{3}{4} \end{aligned}$$



A.3 EXPECTATION

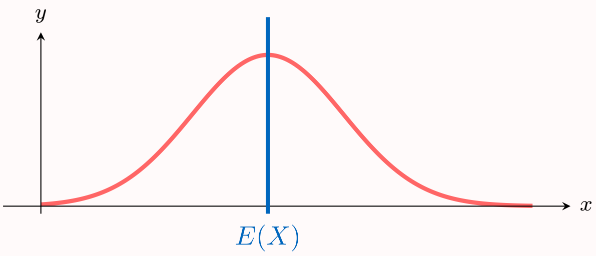
The expectation (or expected value) of a continuous random variable is the "average" value it would take if the experiment were repeated infinitely. It represents the center of the distribution and is calculated as a weighted average, where the pdf $f(x)$ provides the weighting:

$$\begin{aligned} E(X) &= \sum_{x\in[a,b]} xP(x \leqslant X < x + \mathrm{d}x) \\ &= \sum_{x\in[a,b]} x\frac{P(x \leqslant X < x + \mathrm{d}x)}{\mathrm{d}x}\mathrm{d}x \\ &= \int_a^b x f(x)\, \mathrm{d}x \end{aligned}$$

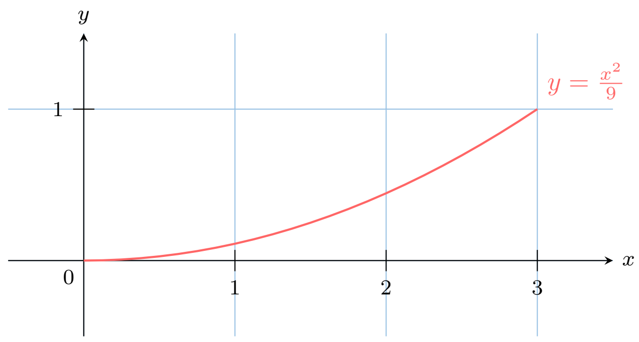
Definition **Expectation**

For a continuous random variable X with density f on $[a,b]$, the **expected value** is

$$E(X) = \int_a^b x f(x)\, dx.$$



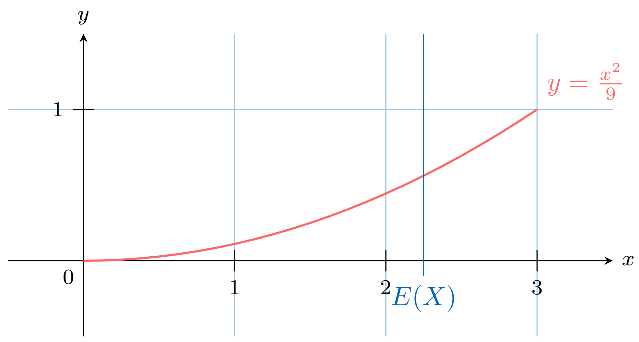
Ex: The random variable X with values on $[0, 3]$ has density $f(x) = \frac{x^2}{9}$:



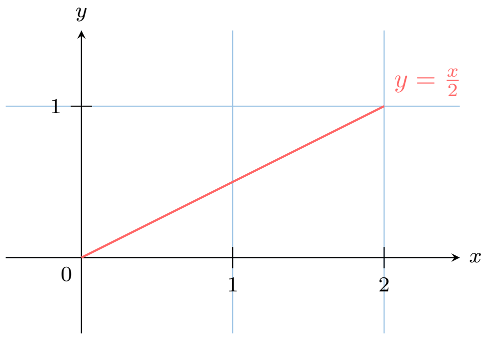
Find $E(X)$.

Answer: Compute $E(X)$:

$$\begin{aligned} E(X) &= \int_0^3 x \cdot \frac{x^2}{9}\, dx \\ &= \int_0^3 \frac{x^3}{9}\, dx \\ &= \left[\frac{x^4}{36}\right]_0^3 \\ &= \frac{3^4}{36} - 0 \\ &= 2.25 \end{aligned}$$



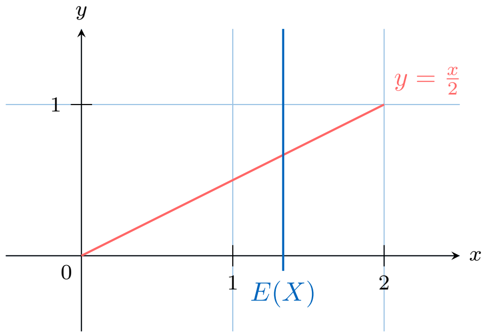
Ex: The random variable X with values on $[0, 2]$ has density $f(x) = \frac{x}{2}$:



Find $E(X)$.

Answer: Compute $E(X)$:

$$\begin{aligned}
E(X) &= \int_0^2 x \cdot \frac{x}{2}\,dx \\
&= \int_0^2 \frac{x^2}{2}\,dx \\
&= \left[\frac{x^3}{6}\right]_0^2 \\
&= \frac{2^3}{6} - 0 \\
&= \frac{8}{6} \\
&= \frac{4}{3}
\end{aligned}$$



A.4 VARIANCE

The variance of a continuous random variable measures the spread of its values around the expected value if the experiment were repeated infinitely. It quantifies the distribution's dispersion and can be calculated as for a discrete random variable:

$$\begin{aligned}
V(X) &= \sum_{x\in[a,b]} (x - E(X))^2 P(x \le X < x + dx) \\
&= \sum_{x\in[a,b]} (x - E(X))^2 \frac{P(x \le X < x + dx)}{dx} \cdot dx \\
&= \int_a^b (x - E(X))^2 f(x)\,dx
\end{aligned}$$

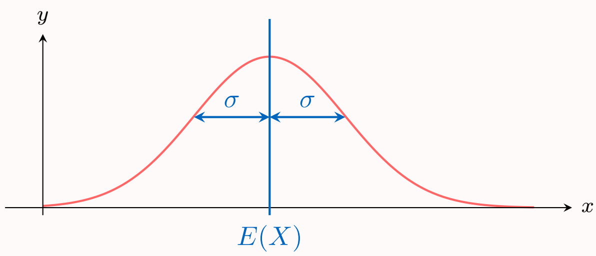
Definition Variance and Standard Deviation

For a continuous random variable X with density f on $[a, b]$, the **variance** is

$$V(X) = \int_a^b (x - E(X))^2 f(x)\,dx.$$

The **standard deviation** is

$$\sigma = \sqrt{V(X)}.$$



Proposition Computational Formula for Variance

A more convenient formula for variance computation is:

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

Ex: The random variable X with values on $[0,2]$ has density $f(x) = \frac{x}{2}$.
Find $V(X)$.

Answer:

- Compute $E(X)$:

$$\begin{aligned} E(X) &= \int_0^2 x \cdot \frac{x}{2}\, dx \\ &= \int_0^2 \frac{x^2}{2}\, dx \\ &= \left[\frac{x^3}{6}\right]_0^2 \\ &= \frac{2^3}{6} - 0 \\ &= \frac{8}{6} \\ &= \frac{4}{3} \end{aligned}$$

- Compute $\int_0^2 x^2 \cdot f(x)\, dx$:

$$\begin{aligned} \int_0^2 x^2 \cdot f(x)\, dx &= \int_0^2 x^2 \cdot \frac{x}{2}\, dx \\ &= \int_0^2 \frac{x^3}{2}\, dx \\ &= \left[\frac{x^4}{8}\right]_0^2 \\ &= \frac{2^4}{8} - 0 \\ &= \frac{16}{8} \\ &= 2 \end{aligned}$$

- Compute $V(X)$ using the alternative formula:

$$\begin{aligned} V(X) &= \int_0^2 x^2 \cdot f(x)\, dx - [E(X)]^2 \\ &= 2 - \left(\frac{4}{3}\right)^2 \\ &= 2 - \frac{16}{9} \\ &= \frac{18}{9} - \frac{16}{9} \\ &= \frac{2}{9} \end{aligned}$$

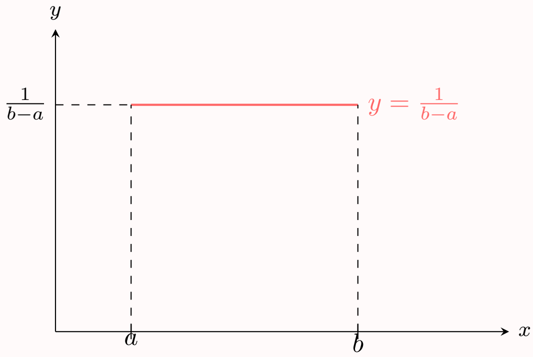
A.5 CONTINUOUS UNIFORM DISTRIBUTION

The continuous uniform distribution applies to events that are equally likely across an interval, such as the spinner example. The density is constant over the range.

Definition Continuous Uniform Distribution

A continuous random variable X follows a **continuous uniform distribution** on $[a, b]$ if its density is:

$$f(x) = \frac{1}{b-a} \quad \text{for} \quad a \leq x \leq b$$



Proposition Properties

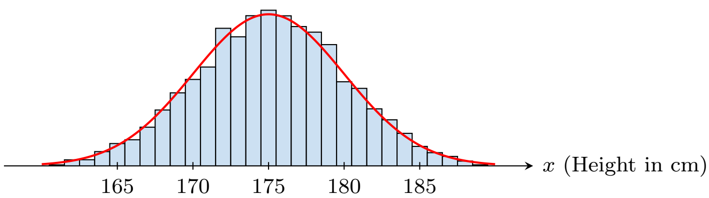
Let X be a continuous random variable following a continuous uniform distribution on $[a, b]$:

- for all $c, d \in [a, b] : P(c \leq X \leq d) = \frac{d-c}{b-a}$,
- $E(X) = \frac{a+b}{2}$.
- $V(X) = \frac{(b-a)^2}{12}$.

B NORMAL DISTRIBUTION

B.1 STANDARD NORMAL DISTRIBUTION

The normal distribution is a key continuous distribution in statistics, often used to model real-world phenomena (e.g., heights, test scores) due to the Central Limit Theorem. This theorem states that the sum or average of many independent random variables, under certain conditions, approximates a normal distribution as the sample size increases. The normal curve is bell-shaped, symmetric, and centered at its mean.
For example, we plot a histogram of the heights of boys at the university. The distribution represented by the histogram follows a bell-shaped curve, also known as a normal distribution.

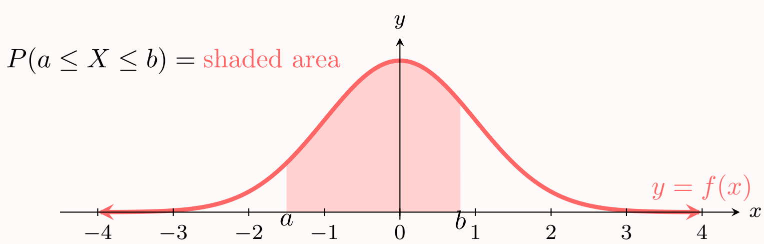


Definition Standard Normal Distribution

A continuous random variable X follows a **standard normal distribution** if it has a mean of 0 and a variance of 1. Its density is:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty$$

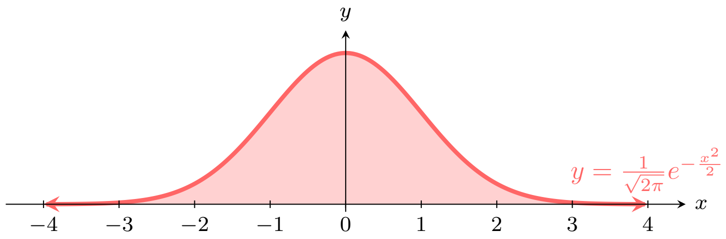
This is denoted $X \sim \mathcal{N}(0, 1)$.



Remark The total probability is 1:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx = 1$$

This is the area under the entire curve.



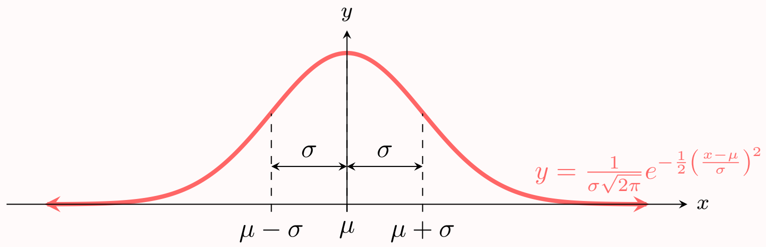
B.2 NORMAL DISTRIBUTION

Definition Normal Distribution

A continuous random variable X follows a **normal distribution** if its density is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

where μ is the **mean** and σ^2 is the **variance**. The graph is a **normal curve** (bell-shaped), denoted $X \sim \mathcal{N}(\mu, \sigma^2)$.



Note on Technology Use The integral of the normal distribution's probability density function cannot be expressed in terms of elementary functions. Therefore, calculating probabilities for a normal distribution, such as $P(a \leq X \leq b)$, must be done using a graphic display calculator (GDC) or statistical software.

Proposition Expectation and Variance

For $X \sim \mathcal{N}(\mu, \sigma^2)$:

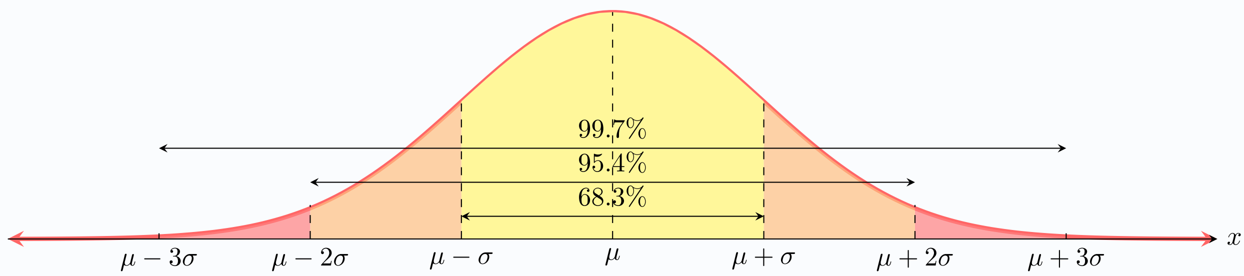
- $E(X) = \mu$,
- $V(X) = \sigma^2$.

B.3 EMPIRICAL RULE FOR NORMAL DISTRIBUTION

Proposition Empirical Rule for Normal Distribution

For $X \sim \mathcal{N}(\mu, \sigma^2)$, the probabilities for intervals centered at the mean are approximately:

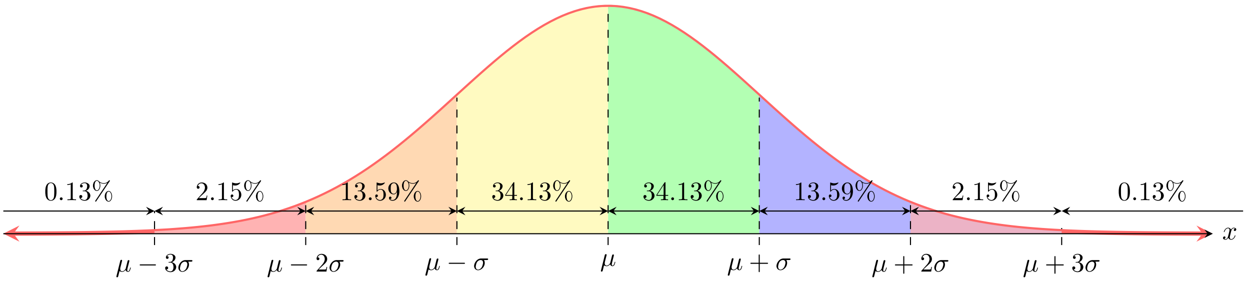
- $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 68.3\%$
- $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95.4\%$
- $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 99.7\%$



The Empirical Rule states that for a normal distribution, specific percentages of data lie within intervals around the mean. Due to the symmetry of the normal curve, we can break these down to find the probabilities for individual sections, each one standard deviation wide. For example, by symmetry, the area from μ to $\mu + \sigma$ is half of the total area from $\mu - \sigma$ to $\mu + \sigma$:

$$P(\mu \leq X \leq \mu + \sigma) = \frac{P(\mu - \sigma \leq X \leq \mu + \sigma)}{2} \approx 34.13\%$$

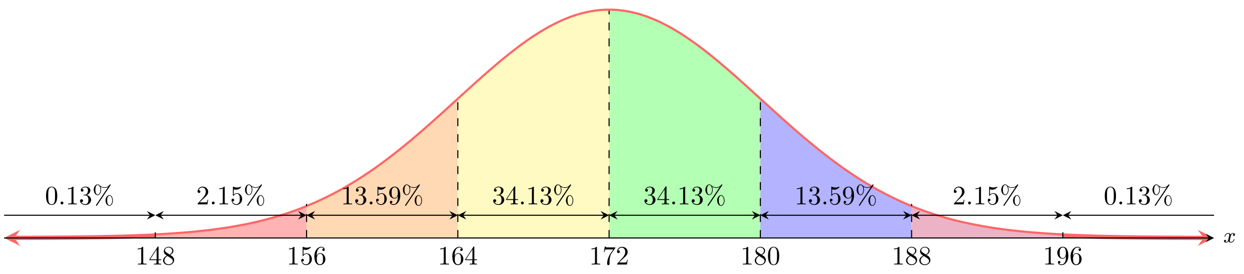
This gives us a detailed map of the normal distribution:



Ex: Students' heights at a school are normally distributed with mean $\mu = 172\,\text{cm}$ and standard deviation $\sigma = 8\,\text{cm}$.

1. Find the percentage of students with heights between 164 cm and 172 cm.
2. Find the percentage between 164 cm and 180 cm.
3. Find the percentage with heights above 196 cm.
4. Find the percentage with heights below 196 cm.
5. In a group of 500 students, how many are expected to have heights between 164 cm and 180 cm?

Answer: First, we label the distribution with the given mean and standard deviation:



1. $P(164 \leq X \leq 172) = P(\mu - \sigma \leq X \leq \mu) = 34.13\%$.
2. $P(164 \leq X \leq 180) = P(\mu - \sigma \leq X \leq \mu + \sigma) = 34.13\% + 34.13\% = 68.26\%$.
3. $P(X > 196) = P(X \geq \mu + 3\sigma) = 0.13\%$.
4. $P(X < 196) = 1 - P(X \geq 196) = 100\% - 0.13\% = 99.87\%$.
5. Expected number $= 68.26\% \times 500 = 0.6826 \times 500 \approx 341$ students.

B.4 QUANTILES

Definition Quantile

The ***p*-quantile** of a random variable X is the value x_p such that $P(X \leq x_p) = p$. For example, the 0.95-quantile (or 95th percentile) is the value $x_{0.95}$ below which 95% of the distribution's values lie.

Ex: Let $X \sim \mathcal{N}(7, 2^2)$. Find the value k such that $P(X \leq k) = 0.95$.

Answer: Using a calculator's inverse normal function with area$= 0.95$, $\mu = 7$, and $\sigma = 2$, we find $k \approx 10.29$.