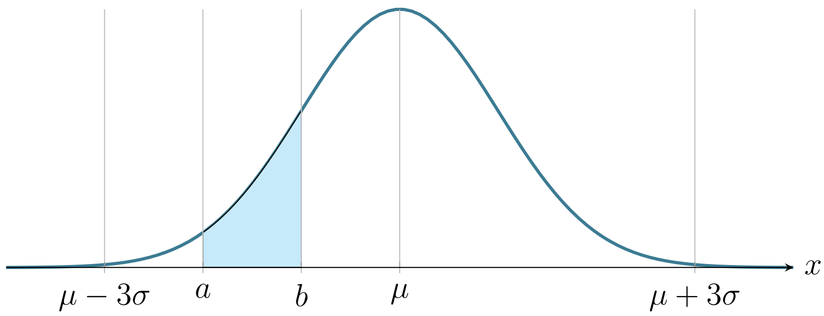
5 Normal distribution

Unlike the binomial distribution, which takes integer values, the normal distribution takes real values, and we visualize it with symmetrical bell-shaped curve.

There are two important numbers for the normal distribution: the **mean** μ and the **variance**, σ^2, where σ is the standard deviation.

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

We visualize probabilities, such as $P(a \leq X \leq b)$ below, as the area under the curve.



5.1 "Three sigma" rule

A convenient heuristic for the normal distribution is that nearly all values ($\approx 99.7\%$) takes values between $\mu - 3\sigma$ and $\mu + 3\sigma$.

Example: let X represents the percentage marks a student got on a test.
If the mean is 68 and variance is 100, then nearly all students scored between 38 and 98.
If the mean is 11 with standard deviation 5, then a normal distribution will not be a good model for X as a normal model will imply that the marks will be between -4 and 26, but percentage marks cannot be negative.

5.2 Probability of equality

Because we are now working with real numbers, $P(X = a) = 0$ for all a. This means that $P(X \leq a) = P(X < a)$ for a normal distribution.

Compare this to the situation for a binomial random variable. Because of this behavior, "NORMPDF" has a different meaning from what we are used to and we will not be pressing that in our calculator for our syllabus.

5.3 Normcdf

Try to use "NORMCDF" to calculate the following probabilities, where $X \sim N(5,7)$:

- $P(2 < X \leq 6)$
- $P(X < 4)$
- $P(X \geq 3)$

Answers: 0.519, 0.353, 0.775

5.4 Invnorm

Try to use "INVNORM" to calculate b for the following, where $X \sim N(5,7)$

- $P(X < b) = 0.7$
- $P(X > b) = 0.4$

Answers: 6.39, 5.67

5.5 Standardization

For unknown mean and/or variance, we apply the "standardization" technique, where we transform any normal random variable we have to the standard normal distribution Z.

$$\boxed{Z = \frac{X - \mu}{\sigma}, \quad Z \sim N(0,1)}$$

Example: Let $X \sim N(\mu, 7)$ and $P(X \leq 3) = 0.2$. Then $P\left(Z \leq \frac{3-\mu}{\sqrt{7}}\right) = 0.2$ and we can then proceed to find μ using invnorm.

5.6 Symmetry

For tougher, non-standard questions, symmetry and visualization of the normal curve can be a useful strategy.

5.7 Linear combinations

Recall the expectation and variance formulas for distributions such as $2X - 3Y$, $X_1 + X_2$ and $2X$ in our DRV topic. We will be using them extensively in this topic.

For example, let the weight of an apple be modeled by $X \sim N(5, 7)$ and the weight of an orange by $Y \sim N(8, 2^2)$.

Then the total weight of an apple and orange is $X + Y \sim N(13, 11)$.

To get the probability that an apple weighs more than an orange, we want to find $P(X > Y) = P(X - Y > 0)$. $X - Y \sim N(-3, 7 + 4)$.

To get the total weight of two apples, we have $X_1 + X_2 \sim N(10, 7 + 7) = N(10, 2 \times 7)$.

To get twice the weight of an apple, we have $2X \sim N(10, 2^2 \times 7)$.

5.8 *Sample mean (to be covered in the next topic of sampling)

$$\text{Total: } X_1 + \cdots + X_n \sim N(n\mu, n\sigma^2)$$

$$\text{Sample mean: } \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$