Maximum Likelihood Estimation
Parameter estimation:
Given independent samples \( x_1, x_2, \ldots, x_n \) from a distribution \( f(X|\theta) \), estimate \( \theta \).

Example: Given outcomes \( HHTHTH \) of independent flips of a coin, estimate \( \theta = P(\text{Heads}) \).

\( P(X|\theta) \): Probability of generating \( X \) given parameter (or model) \( \theta \).
Viewed as a function of \( X \), with \( \theta \) fixed, this is a probability.
Viewed as a function of \( \theta \), with \( X \) fixed, this is a likelihood and often written \( L(X|\theta) \).

Maximum likelihood estimation:
What \( \hat{\theta} \) maximizes \( L(x_1, x_2, \ldots, x_n | \theta) \) =
\[ \prod_{i=1}^{n} f(x_i | \theta) \] .

Approach: \( \frac{\partial}{\partial \theta} L(x_1, x_2, \ldots, x_n | \theta) = 0 \) and solve for \( \hat{\theta} \).
More often: \( \frac{\partial}{\partial \theta} \ln L(x_1, x_2, \ldots, x_n | \theta) = 0 \)

Because \( \ln L \) is monotonically increasing with \( L \),
the maximum of \( \ln L \) will be at the same value of \( \theta \) as the maximum of \( L \).

The benefit of \( \ln \) is that it turns \( \prod \) into a \( \sum \), avoiding the chain rule when you differentiate.
\( y > x \Rightarrow \ln y > \ln x \): monotonically increasing
Apply MLE to $\text{Ber}(\theta)$, independent
$\theta$ is the probability of heads, $n$ flips of the coin $x_1, x_2, \ldots, x_n$, yielding $m_0$ tails and $m_1$ heads, where $m_0 + m_1 = n$.

$L(x_1, x_2, \ldots, x_n | \theta) = \theta^{m_1} (1-\theta)^{m_0}$

$\ln L(x_1, x_2, \ldots, x_n | \theta) = m_1 \ln \theta + m_0 \ln (1-\theta)$

$\Rightarrow \frac{\partial \ln L(x_1, x_2, \ldots, x_n | \theta)}{\partial \theta} = \frac{m_1}{\theta} + \frac{-m_0}{1-\theta} = 0$

Let $\hat{\theta}$ be the solution to this equation.

$m_1 (1-\hat{\theta}) - m_0 \hat{\theta} = 0$

$m_1 = (m_0 + m_1) \hat{\theta}$

$\hat{\theta} = \frac{m_1}{m_0 + m_1}$

A good estimate of $\theta$ is the fraction of flips that came up heads, $m_1 / n$.

Is $\hat{\theta}$ a maximum, minimum, or --- of $\ln L$?

$\frac{\partial^2 \ln L(x_1, x_2, \ldots, x_n | \theta)}{\partial \theta^2} = - \frac{n_1}{\theta^2} - \frac{n_0}{(1-\theta)^2} < 0$

for all values of $\theta$.

$\Rightarrow \ln L$ is concave downward everywhere.

So $\hat{\theta} = \hat{\theta}$ is a global maximum.
\( X_i \sim N(\theta_1, \theta_2) \), both \( \theta_1 = \mu \) and \( \theta_2 = \sigma^2 \) are unknown.

Find \( \hat{\theta}_1, \hat{\theta}_2 \) that maximize

\[ L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2). \]