King Abdullah University of Science and Technology

جامعة الملك عبدالله للعلوم والتقنية

Calculus: Differentiation and Its Application

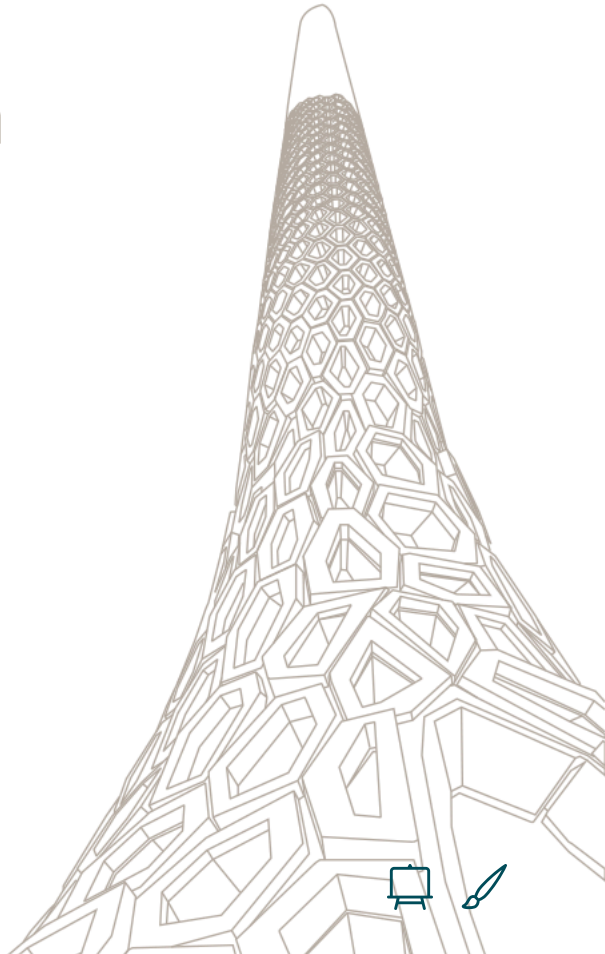
A statistical perspective

Haziq Jamil

Research Specialist
BAYESCOMP @ CEMSE-KAUST

`https://haziqj.ml/uitm-calculus` | PDF

June 13, 2026



(Almost) Everything you ought to know...

...about calculus in the first year

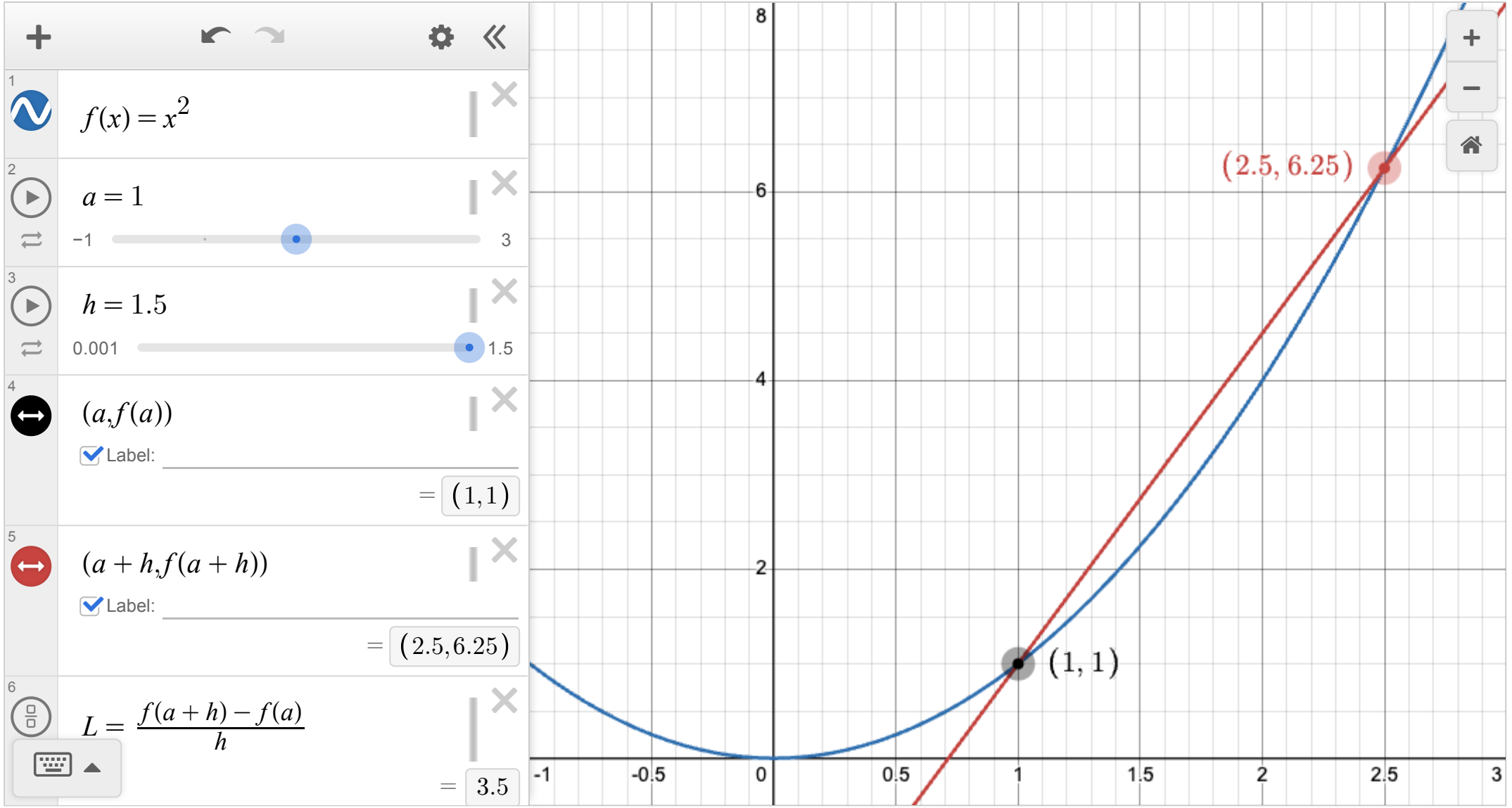
Let $f : \mathcal{X} \to \mathbb{R}$ be a *real-valued* function defined on an input set $\mathcal{X}$.

Some examples

Function	Derivative
$f(x) = x^2$	$f'(x) = 2x$
$f(x) = \sum_n a_n x^n$	$f'(x) = \sum_n n a_n x^{n-1}$
$f(x) = \sin(x)$	$f'(x) = \cos(x)$
$f(x) = \cos(x)$	$f'(x) = -\sin(x)$
$f(x) = e^x$	$f'(x) = e^x$
$f(x) = \ln(x)$	$f'(x) = \frac{1}{x}$

Taylor series is a powerful tool for approximating functions using derivatives.

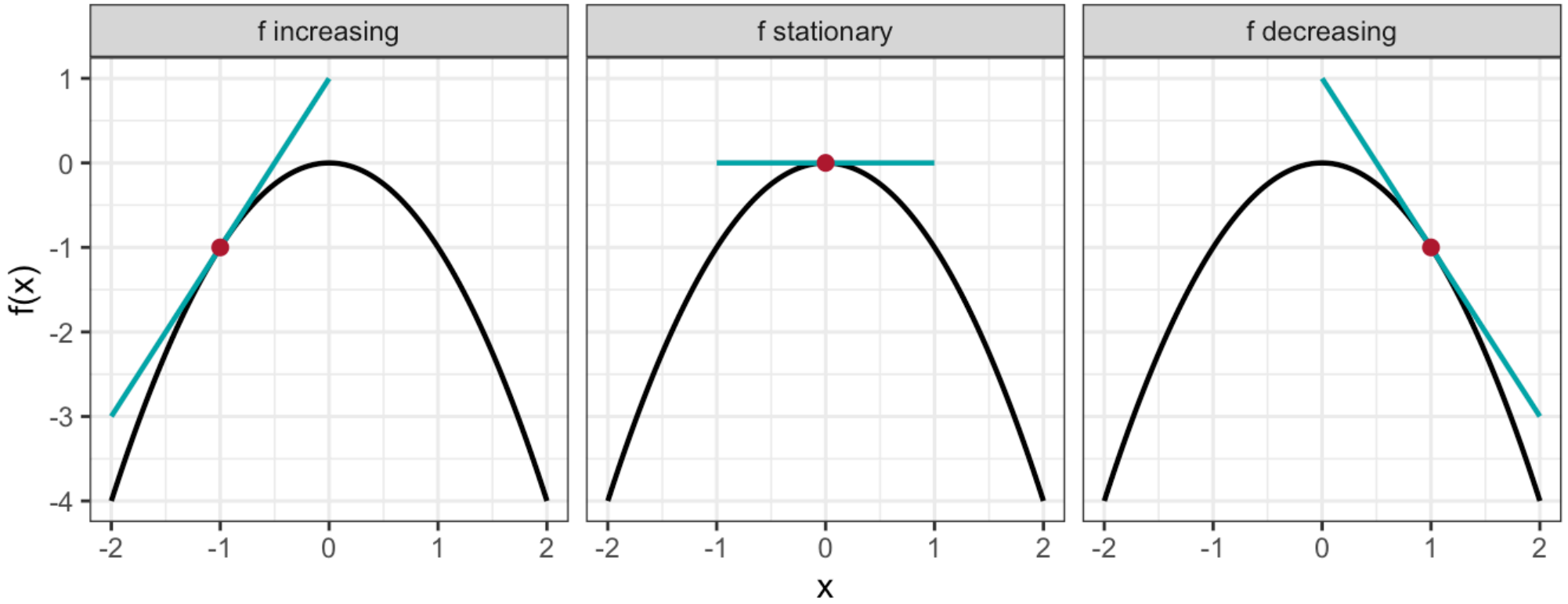
Graphically...



But what *is* a derivative?

The derivative of a function tells you:

- 🚀 How fast the function is *changing* at any point
- 📐 The **slope** of the tangent line at that point

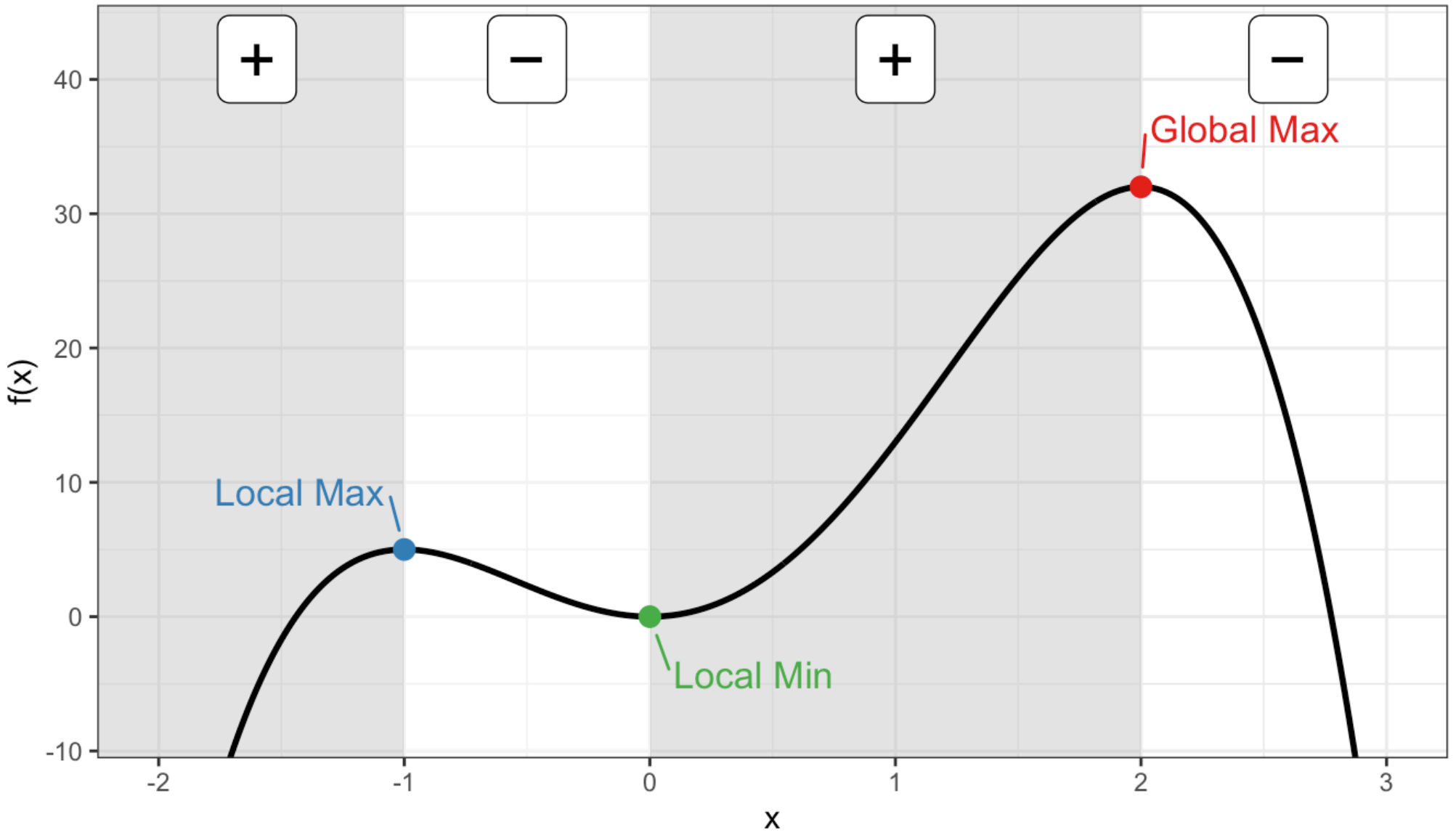


The concept of optimisation

- When f is some kind of a "reward" function, then the value of x that maximises f is highly of interest. Some examples:
  - 💰 **Profit maximisation**: Find the price that maximises profit.
  - 🧬 **Biological processes**: Find the conditions that maximise growth or reproduction rates.
  - 👷 **Engineering**: Find the design parameters that maximise strength or efficiency.
- Derivatives help us find so-called *critical values*: Solve $f'(x) = 0$.

Example 1 Find the maximum of $f(x) = -3x^4 + 4x^3 + 12x^2$.

Graphically...



How do we know if it's a maxima or minima?

Second derivative test: Measure the **change in slope** around a point x, i.e. $f''(\hat{x}) = \frac{d}{dx}\left(\frac{df}{dx}(x)\right) = \frac{d^2f}{dx^2}(x)$.

Second derivative test

From Example 1, the second derivative is given by

$$\begin{aligned} f''(x) &= \frac{d}{dx}\left(-12x^3 + 12x^2 + 24x\right) \\ &= -36x^2 + 24x + 24 \end{aligned}$$

Plug in the critical points:

- $x = -1$: $f''(-1) = -36 - 24 + 24 = -36 < 0$, hence local maximum.
- $x = 0$: $f''(0) = 0 + 0 + 24 = 24 > 0$, hence local minimum.
- $x = 2$: $f''(2) = -144 + 48 + 24 = -72 < 0$, hence local maximum.

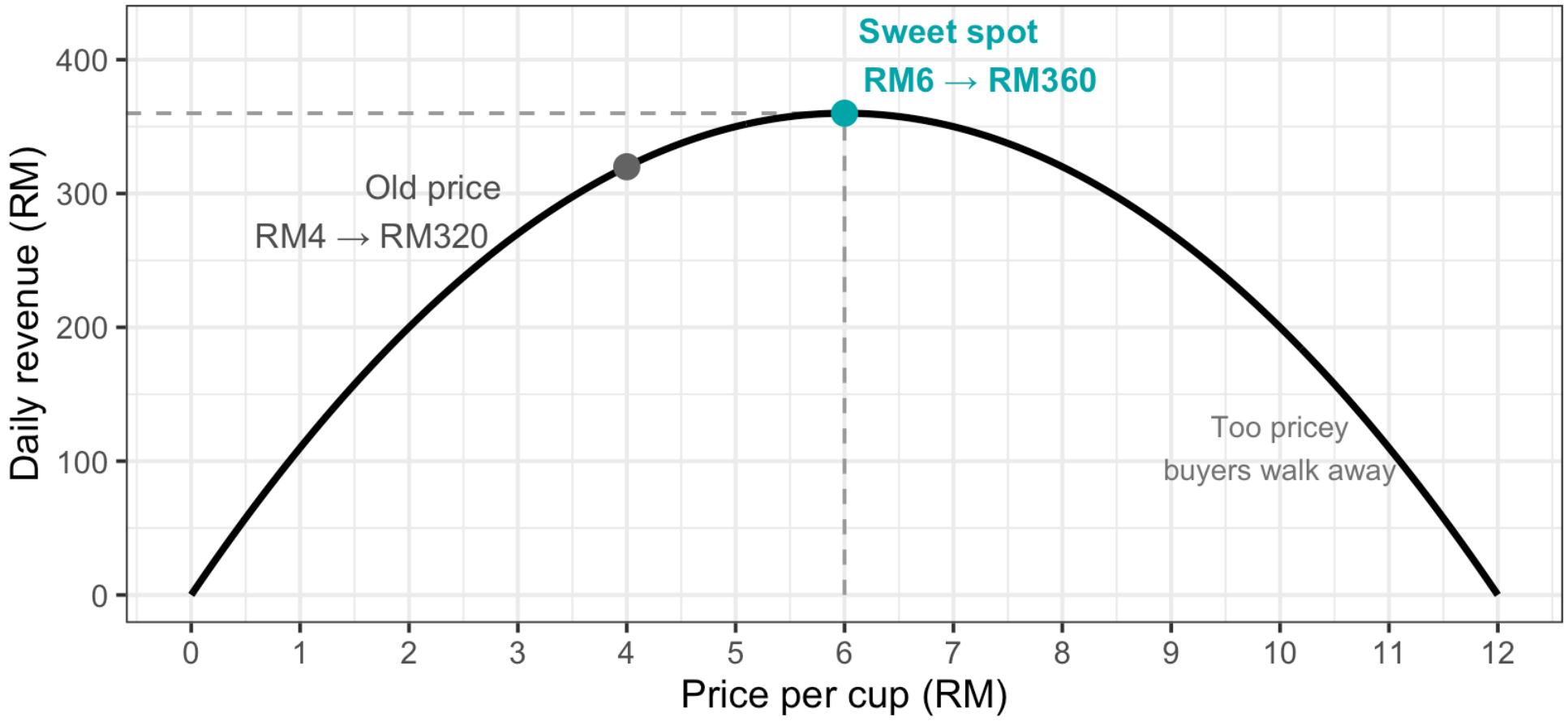
Tip

Often it is not enough to just differentiate once to find optima. Differentiate twice to classify critical points.

Optimisation in the real world

Example 2 (Pricing a cup of fries 🍟) Your Potato Story outlet sells 80 cups a day at RM4 each. Market testing shows that every RM1 you add to the price loses 10 cups a day. What price brings in the most money?

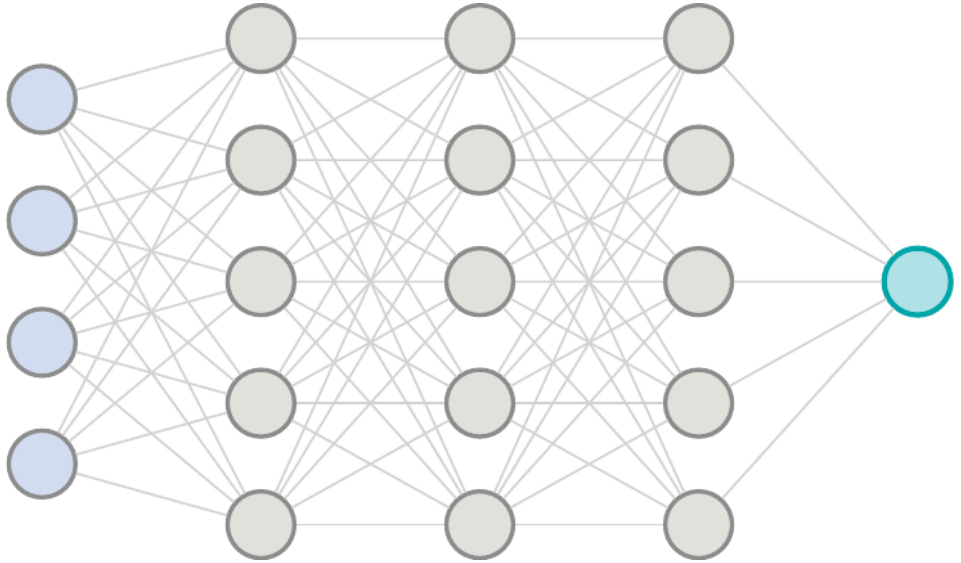
Finding the sweet spot 🍟



Optimisation in AI

input

hidden layer



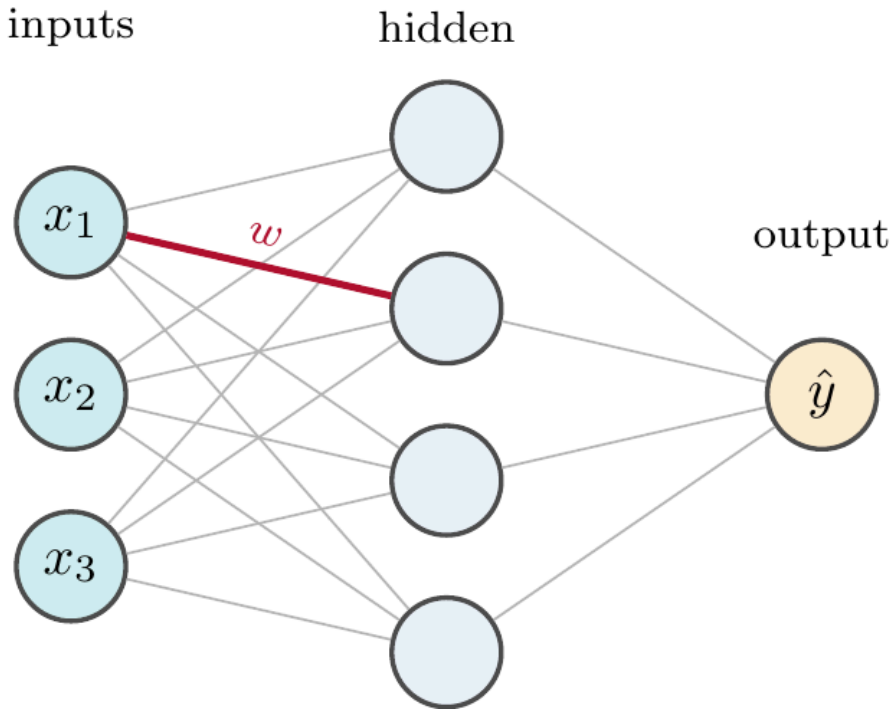
output

cat 🐱

Transformers, the tech behind LLMs, and Deep Learning: 3B1B
https://youtu.be/wjZofJX0v4M?si=IzEE6fzSlDLLpLsc

Inside the machine 🧠

We cracked the potato problem **by hand**. But what *is* an "AI"? Underneath, it is a network of simple units wired together:



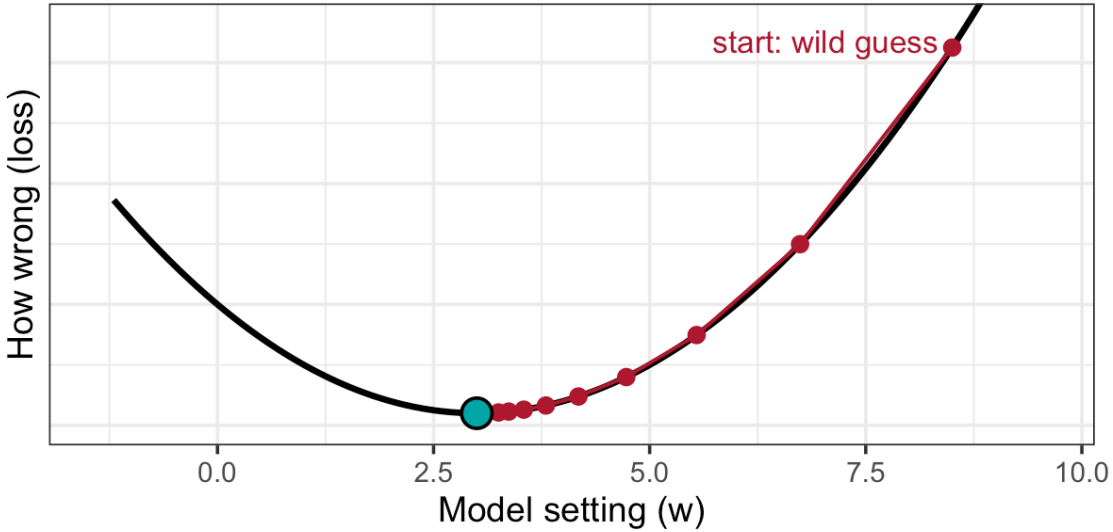
$$h = \sigma(\overbrace{w_1 x_1 + w_2 x_2 + \cdots + b}^{\text{a linear equation}})$$

An LLM is a **giant** version of this, with billions of arrows.

How does AI learn? 🤖

Follow the slope downhill

1. Start with a wild guess!
2. Measure how **wrong** it is, via $L(w)$, the *loss* function.
3. The slope points **uphill** → step the *opposite* way.
4. Repeat until you hit the bottom.



Summary so far

- 📈 Derivatives represent rate of change (slope) of a function $f : \mathcal{X} \to \mathbb{R}$.
- 🎯 Interested in optimising an *objective* function $f(x)$ representing some kind of "reward" or "cost".
- 🔍 Find critical points by solving $f'(x) = 0$.
- Use the second derivative test to classify critical points:
  - If $f''(x) < 0$, then f is concave down at x and x is a local maximum. ⛰️
  - If $f''(x) > 0$, then f is concave up at x and x is a local minimum. 🫧
  - If $f''(x) = 0$, then the test is inconclusive. 🤷

A statistical perspective

But what *is* statistics?

Statistics is a scientific subject that deals with the collection, analysis, interpretation, and presentation of data.

- **Collection** means designing experiments, questionnaires, sampling schemes, and also administration of data collection.
- **Analysis** means mathematically modelling, estimation, testing, forecasting.

See also: *The Art of Statistics: Learning from Data* by David Spiegelhalter.

Motivation

Demand isn't fixed—each passer-by is a 'gamble'. So **at today's price**, what's the chance a customer buys?

Measuring how p *changes* when you raise the price compares two such numbers — that's an **A/B test**. First, let's pin down one.

A probabilistic model

Each X_i is a *random variable* taking only two possible outcomes, i.e.

$$X_i = \begin{cases} 1 & \text{w.p.} \quad p \quad (\text{buys}) \\ 0 & \text{w.p.} \quad 1-p \quad (\text{walks on}) \end{cases}$$

This is known as a **Bernoulli** random variable.

Learning from data

p is unknown

If we do not know p, then it is not possible to calculate probabilities, expectations, variances… ☹

The likelihood function

Definition 2 Given a probability function $x \mapsto f(x \mid \theta)$ where x is a realisation of a random variable X, the *likelihood function* is $\theta \mapsto f(x \mid \theta)$, often written $\mathcal{L}(\theta) = f(x \mid \theta)$.

Parameteric statistical models

Assume that $X_i \sim f(x \mid \theta)$ independently for $i = 1, \ldots, n$. Here, functional form of f is known, but the parameter θ is unknown. Examples:

Name	$f(x \mid \theta)$	θ	**Remarks**
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	$p \in (0, 1)$	No. successes in n trials
Poisson	$\frac{\lambda^x e^{-\lambda}}{x!}$	$\lambda > 0$	Count data
Uniform	$\frac{1}{b-a}$ for $x \in [a, b]$	$a < b$	Equally likely outcomes
Exponential	$\lambda e^{-\lambda x}$ for $x \geq 0$	$\lambda > 0$	Waiting time
Normal	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\mu \in \mathbb{R},\ \sigma^2 > 0$	Bell curve

⚠ Everything we need to know about the distribution is captured by the parameter θ.

Example: the MLE for p, by differentiation

We have $X \sim \mathbf{Bin}(n, p)$ — X cups sold out of n passers-by. Treat the pmf as a function of p and take logs:

Now compare two prices: an A/B test 🍟

Remember the A/B test we promised? To learn the *effect* of a price hike, run both prices at once and estimate each p by its sample proportion:

Group A — keep RM4

$n_A = 100$ pass by, $X_A = 70$ buy

$\hat{p}_A = 70/100 = \mathbf{0.70}$

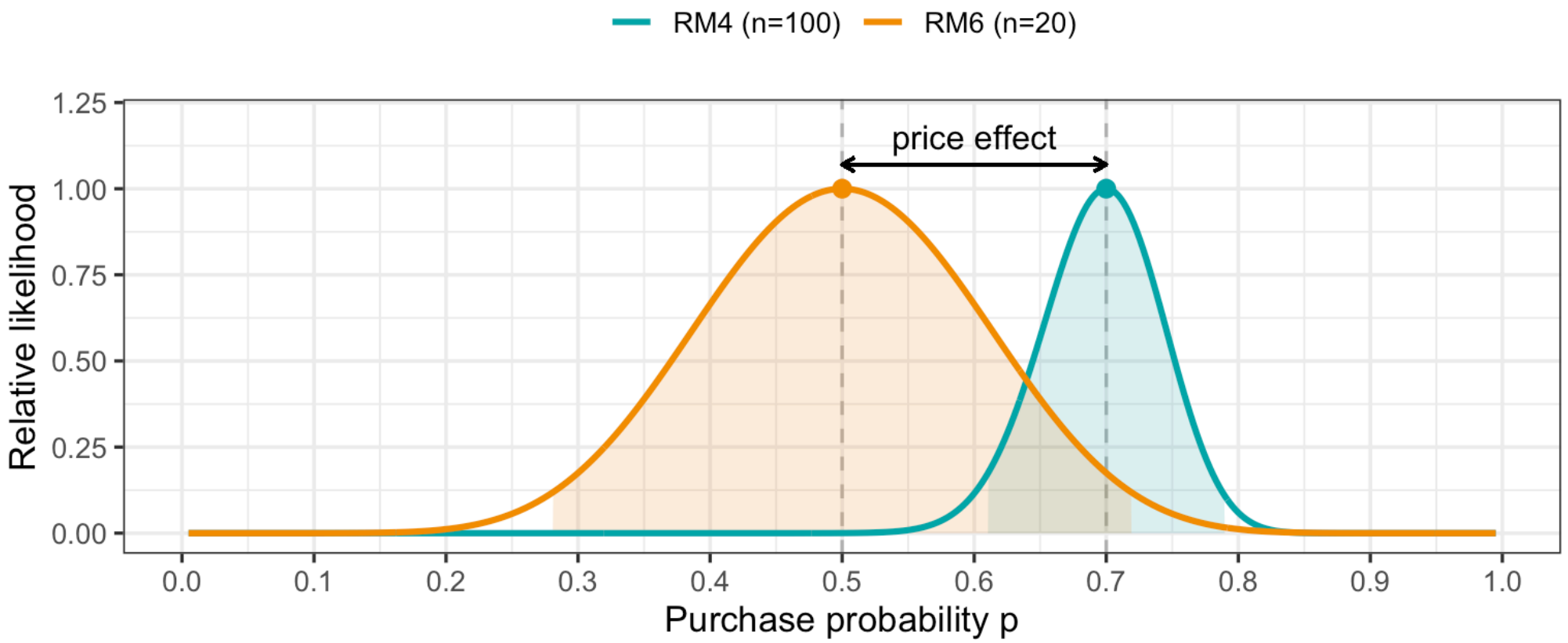
Group B — raise to RM6

$n_B = 20$ pass by, $X_B = 10$ buy

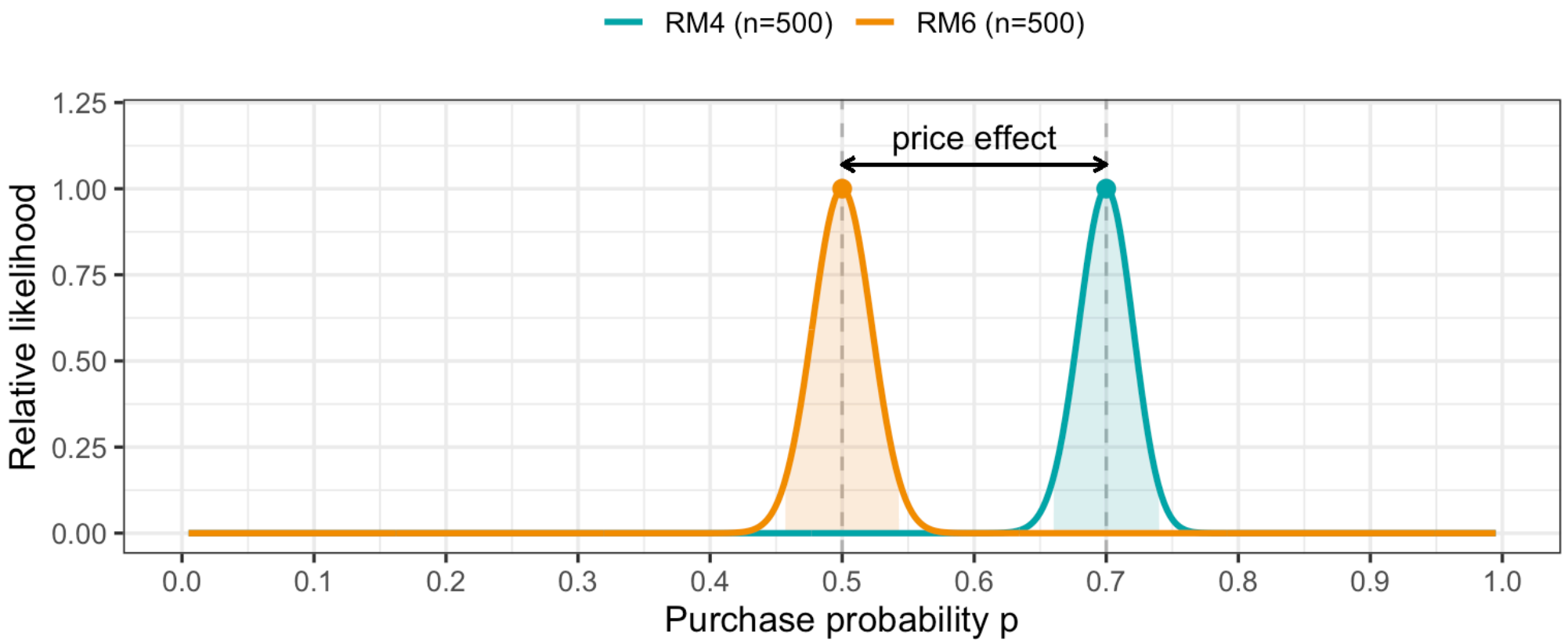
$\hat{p}_B = 10/20 = \mathbf{0.50}$

Per customer, $\mathrm{E}(X_i) = p$ (expected cups $= np$); on average you make one sale every $1/p$ passers-by.

Two prices, two likelihoods



Two prices, two likelihoods 🍟 (cont.)



More data = more confidence. With more data, both groups likelihood curves are sharp and trustworthy. The drop in buy-rate at RM6 is now *clearly* visible, and the 95% ci do not overlap. The price effect is (likely) real! 💰

How sharp is the peak? Differentiate again

The MLE solves $\ell'(p) = 0$. That first derivative — the slope of the log-likelihood — is the **score**:

$$U(p) = \ell'(p) = \frac{X}{p} - \frac{n - X}{1 - p}.$$

Average the curvature: Fisher information

Definition 3 (Fisher information) Under certain regularity conditions, the Fisher information is defined as

$$\mathcal{I}(p) = -\,\mathrm{E}\left[\frac{d^2}{dp^2}\ell(p)\right].$$

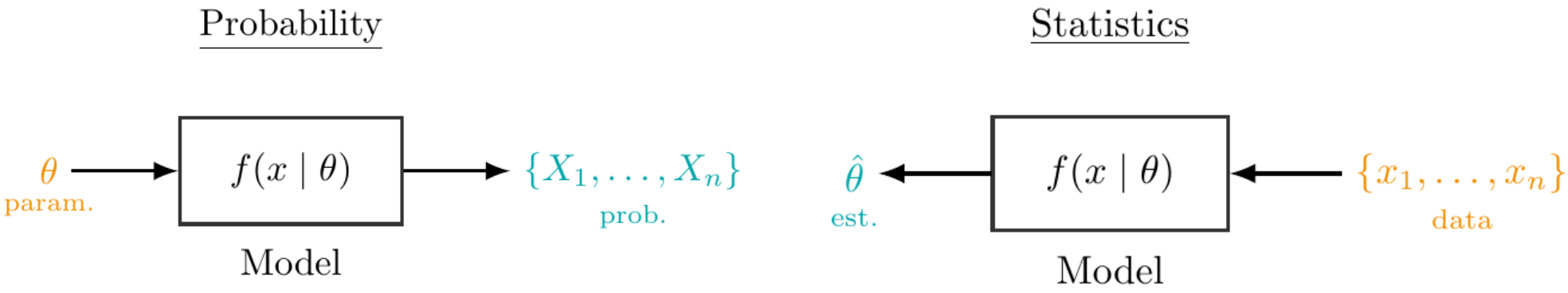
Information = confidence 🎯

Group	n	$\mathcal{I}(\hat{p}) = \frac{n}{\hat{p}(1-\hat{p})}$	$\mathrm{SE} = 1/\sqrt{\mathcal{I}}$
RM4 (sharp)	100	≈ 476	0.046
RM6 (wide)	20	$= 80$	0.112

Conclusions

Summary

Given a model, probability allows us to *predict* data. Statistics on the other hand, allows us to *learn* from data.



Where to go from here

Autodiff

Computers compute derivatives for us: nudge x a little and measure the change

```
fun <- function(x) x ^ 2
# f'(2) = 4
numDeriv::grad(fun, x = 2)
```

```
[1] 4
```

Deep learning goes further with *automatic differentiation* — the chain rule applied exactly, even for functions with **billions** of inputs.

شكراً جزيلاً

https://haziqj.ml/uitm-calculus