

Puzzle 2: Presidential election

In 1948, **Truman** ran against **Dewey** in close elections

A newspaper ran a phone poll of how people voted

Dewey won the poll decisively - newspaper declared:



On to the victory rally ...

... of Truman 😊

It's not δ 's fault:

$$\mathbb{P} [|E_{\text{in}} - E_{\text{out}}| > \epsilon] \leq \delta$$



The bias

In 1948, phones were expensive.

If the data is sampled in a biased way, learning will produce a similarly biased outcome.

Example: normal period in the market

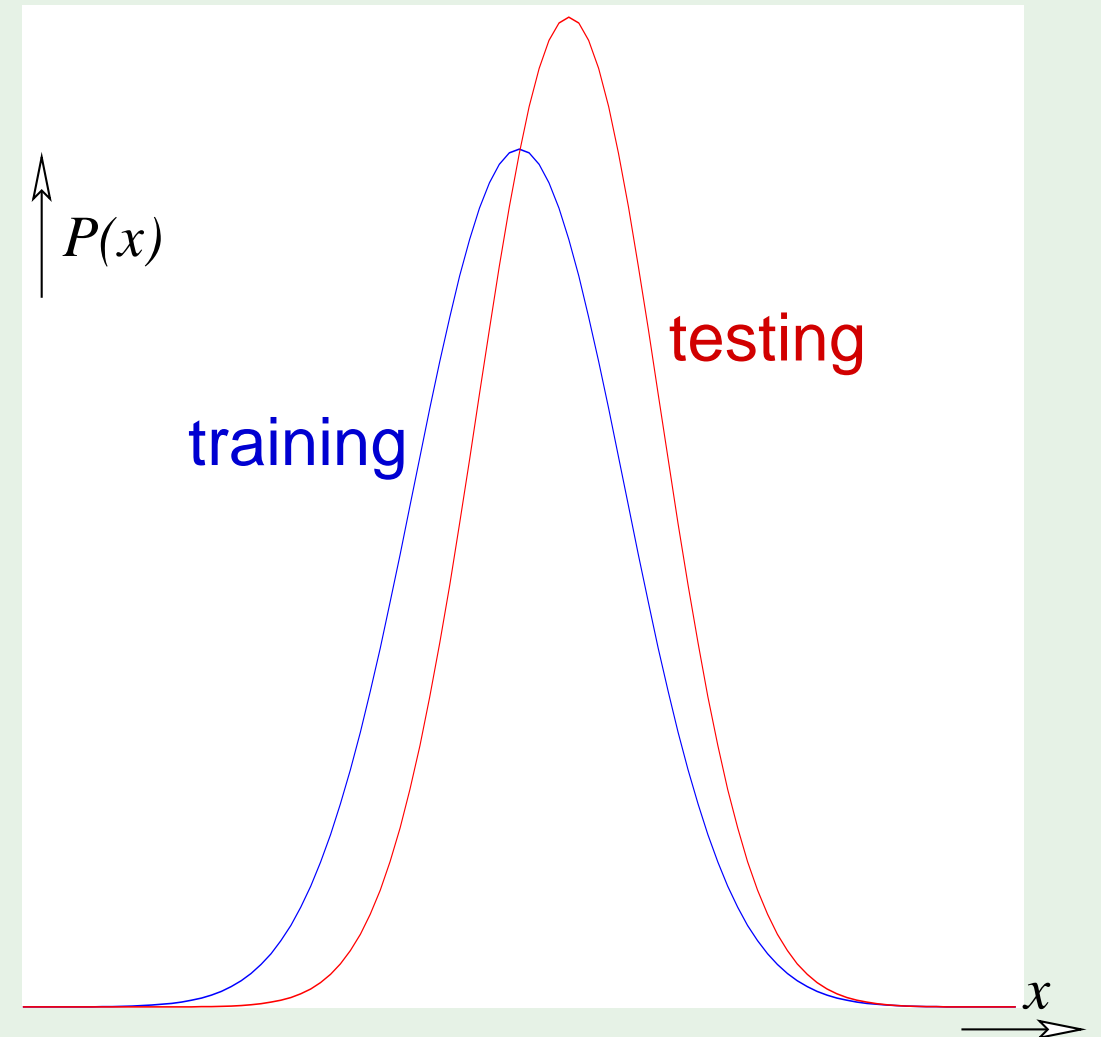
Testing: live trading in real market

Matching the distributions

Methods to match training and testing distributions

Doesn't work if:

Region has $P = 0$ in training, but $P > 0$ in testing



Puzzle 3: Credit approval

Historical records of customers

Input: information on credit application:

Target: profitable for the bank

age	23 years
gender	male
annual salary	\$30,000
years in residence	1 year
years in job	1 year
current debt	\$15,000
...	...