# Foundations of Data Science
# Examples Class 1

A 0/1 signal is being transmitted over a noisy channel. The transmitted signal at timeslot $i \in \{1, \dots, n\}$ is $x_i \in \{0, 1\}$, and furthermore we know that this signal starts at 0 and then flips to 1, i.e. there is a parameter $\theta \in \{1, \dots, n-1\}$ such that

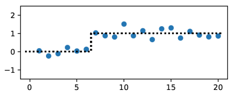
$$x_i = \begin{cases} 0 & \text{for } i \le \theta, \\ 1 & \text{for } i > \theta, \end{cases}$$

but the value of $\theta$ is unknown. The channel is noisy, and the received signal in timeslot $i$ is

assume independent unless told otherwise.

$$Y_i \sim x_i + \text{Normal}(0, \varepsilon^2)$$

where $\varepsilon$ is known.




(*i*) Given received signals $(y_1, \dots, y_n)$, find an expression for the log likelihood, $\log \text{lik}(\theta \mid y_1, \dots, y_n)$. [5 marks]

(*ii*) Give pseudocode for finding the maximum likelihood estimator $\hat{\theta}$. [3 marks]

Probability model

↓

Pr (observed data / params)

§1.4.1: standard random variables, and their densities

§1.4.2: to find Pr, differentiate the c.d.f.

§1.4.3: independence

Standard models:

§1.5 generative modelling

§1.6 supervised learning

≡

loglik (params | obs. data)

↓

code to maximize it. — §1.2 scipy.optimize.fmin, transforming variables

We want Pr (observed data / params)

$Y_i$ is a r.v.

$y_i$ is an observed value of this r.v.

what are parameters? unknown / known?

$\theta$ unknown parameter

$\varepsilon$ known / constant parameter

$n$ known.

$x_i$ is a function of $\theta$.

~~$x$~~

We want $\Pr_{Y_i}(y_i \mid \theta)$ and $\Pr(y_1, \dots, y_n \mid \theta)$

We're told $Y_i \sim x_i + N(0, \varepsilon^2) \sim N(x_i, \varepsilon^2)$

This is a standard distribution, and has density

$$\Pr_{Y_i}(y) = \frac{1}{\sqrt{2\pi\varepsilon^2}} e^{-(y - x_i)^2/2\varepsilon^2}$$

So

$$\Pr(y_1, \dots, y_n) = \Pr_{Y_1}(y_1) \times \dots \times \Pr_{Y_n}(y_n) \quad \text{assuming independence}$$

$$= \left(\frac{1}{\sqrt{2\pi\varepsilon^2}}\right)^n e^{-\sum_{i=1}^{n} (y_i - x_i)^2/2\varepsilon^2}$$

$$\text{loglik}(\theta \mid y_1, \dots, y_n) = -\frac{n}{2}\log(2\pi\varepsilon^2) - \frac{1}{2\varepsilon^2}\sum_{i=1}^{n}(y_i - x_i)^2$$

$$= -\frac{n}{2}\log(2\pi\varepsilon^2) - \frac{1}{2\varepsilon^2}\sum_{i=1}^{n}(y_i - x_i(\theta))^2$$

$$= \text{const} - \text{const} \times \sum (y_i - x_i(\theta))^2$$

$$x_i(\theta) = \begin{cases} 0 & \text{if } i \le \theta \\ 1 & \text{if } i > \theta \end{cases}$$

$$= 1_{i > \theta}$$

(*ii*) We want the m.l.e. $\hat{\theta}$, ie the value of $\theta$ that maximizes $\text{loglik}(\theta \mid y_1, \dots, y_n)$.

The question says $\theta \in \{1, \dots, n-1\}$, so let's iterate over all possible $\theta$ to find the m.l.e.

```
def loglik (θ, y⃗):
    s = 0
    for i in {1,...,n} :
        x_i = 0 if i ≤ θ else 1
        s = s − (y_i − x_i)²
    return s

θs = [(loglik (θ, y⃗), θ)  for θ in {1,...,n−1}]
(_, θ̂) = max (θs)
```