Maximum Likelihood Estimation

Parameter estimation:

Given independent samples $x_1, x_2, \cdots, x_n$ from a distribution $f(x|\theta)$, estimate θ.

Ex: Given outcomes HHTHH of independent flips of a coin, estimate $\theta = P(\text{heads})$.

$P(x|\theta)$: Probability of generating x given parameter (or model) θ.

Viewed as a function of x, with θ fixed, this is a probability.

Viewed as a function of θ, with x fixed, this is a likelihood and often written $L(x|\theta)$.

Maximum likelihood estimation:

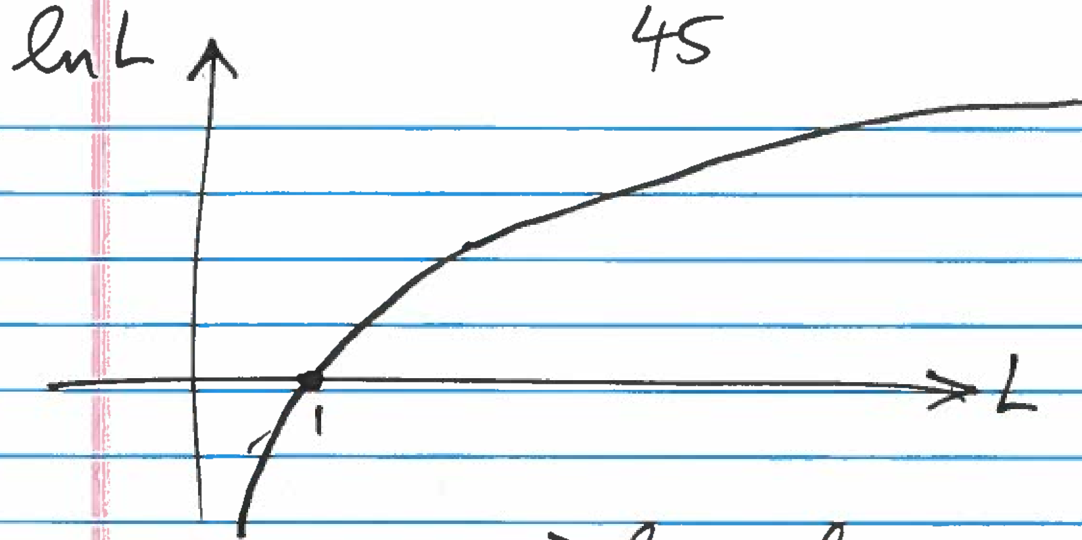
What θ maximizes $L(x_1, x_2, \cdots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$.

Approach: $\frac{\partial}{\partial\theta} L(x_1, x_2, \cdots, x_n|\theta) = 0$ and solve for θ.

More often: $\frac{\partial}{\partial\theta} \ln L(x_1, x_2, \cdots, x_n|\theta) = 0$ " " " "

Because $\ln L$ is monotonically increasing with L, the maximum of $\ln L$ will be at the same value of θ as the maximum of L.

The benefit of $\ln$ is that is turns $\prod$ into a Σ, avoiding the chain rule when you differentiate.



$y > x \Rightarrow \ln y > \ln x$: monotonically increasing

Apply MLE to Ber(θ) independent

θ is the probability of heads, n flips of the coin $x_1, x_2, \cdots, x_n$, yielding n_0 tails and n_1 heads, where $n_0 + n_1 = n$.

$$L(x_1, x_2, \cdots, x_n | \theta) = \theta^{n_1}(1-\theta)^{n_0} \quad \left[\binom{n}{n_1}?\right]$$

$$\ln L(x_1, x_2, \cdots, x_n | \theta) = n_1 \ln \theta + n_0 \ln(1-\theta)$$

$$\frac{\partial}{\partial\theta} \ln L(x_1, x_2, \cdots, x_n | \theta) = \frac{n_1}{\theta} + \frac{-n_0}{1-\theta} = 0$$

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

$$n_1(1-\hat{\theta}) - n_0\hat{\theta} = 0$$

$$n_1 = (n_0 + n_1)\hat{\theta}$$

$$\hat{\theta} = \frac{n_1}{n}$$

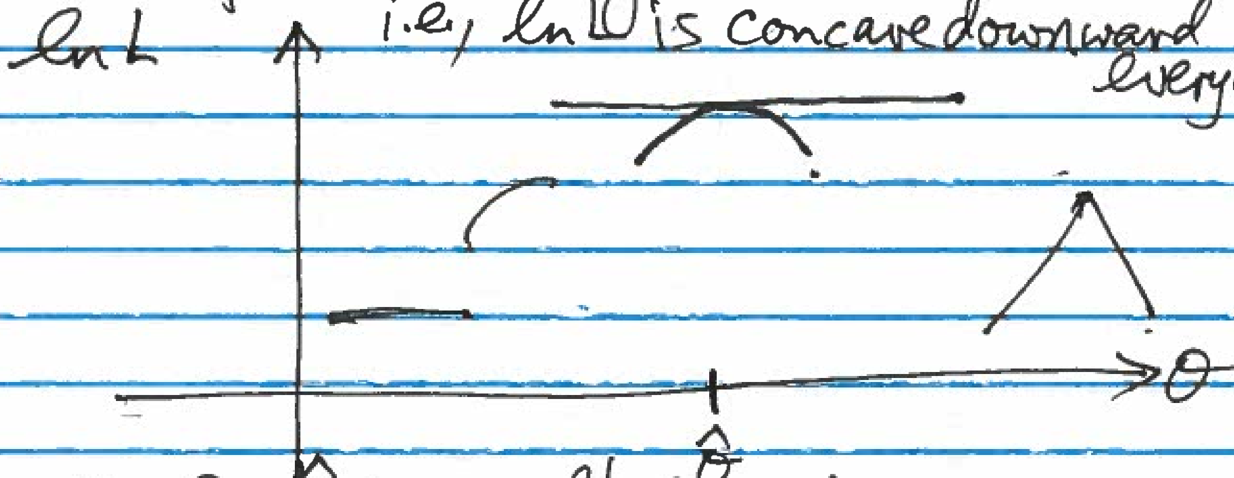
A good estimate of θ is the fraction of flips that came up heads, n_1/n.

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

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

for all values of θ with $0 \le \theta \le 1$.

i.e., $\ln L$ is concave downward everywhere



So $\theta = \hat{\theta}$ 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 $L(x_1, x_2, \cdots, x_n \mid \theta_1, \theta_2)$.