LECTURE 1: MAXIMUM LIKELIHOOD ESTIMATION

SET UP: You observe $\underline{X}$ (possibly a vector) from a parametric family of distributions: $\underline{X}$ has pdf (or pmf if discrete)

$$f(x;\theta)$$

where f is completely specified, although it depends on the parameter θ (θ possibly a vector)

EXAMPLE: $X = (X_1, \ldots, X_n)$ where each X_i has a Normal(μ, σ) density, and the X_i's are independent

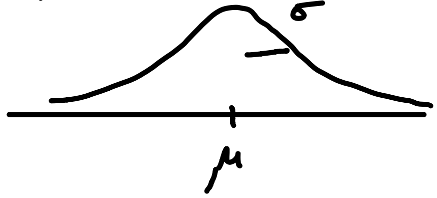
$$f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right)$$

The joint density of $(X_1, \ldots, X_n)$ is obtained as a product, by independence:

$$f_{\underline{X}}(x_1, \ldots, x_n) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

The parameters, μ, σ^2, are typically unknown

UNKNOWN, NOT OBSERVED	OBSERVED
μ, σ^2	$X_1, \ldots, X_n$



BASIC PROBLEM: ESTIMATE THE PARAMETER θ USING THE DATA $X_1, \ldots, X_n$

FIND A "GOOD" FUNCTION $T(X_1, \ldots, X_n)$ SO THAT $T(X_1, \ldots, X_n)$ IS A REASONABLE GUESS AT θ; SUCH FUNCTIONS ARE <u>ESTIMATORS</u>

A METHOD THAT OFTEN PRODUCES REASONABLE ESTIMATORS IS MAXIMUM LIKELIHOOD:

THE LIKELIHOOD PRINCIPLE SAYS TO FIND VALUE θ WHICH MAXIMIZES

$$L(\theta; x) = f(x; \theta)$$

HERE, the observed value x of X is held fixed, and the above is optimized over θ

<u>Example</u> $X_1, \ldots, X_n$ IID $N(\mu, \sigma^2)$
(Independent, Identically Distributed)

NOTE: maximizing $L(\theta)$ is equivalent to maximizing $\ell(\theta) = \log L(\theta)$, since log is monotone function

ℓ is often easier to work with, especially when f has a product form, as it does in the case of IID

$$\ell(\mu,\sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2 - \frac{n}{2}\log 2\pi$$

The final term is irrelevant since it does not depend on either μ or σ^2

Since everything is smooth, we can optimize by looking for critical points

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu)$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_i (x_i-\mu)^2$$

The simultaneous solutions are

$$\mu = \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}$$

$$\sigma^2 = \frac{1}{n}\sum_i (x_i-\bar{x})^2$$

Note that $\bar{X} \sim N(\mu, \sigma^2/n)$

It has many nice properties:

$\mathbb{E}[\bar{X}] = \mu$, also $\bar{X} \to \mu$ as $n \to \infty$

The first property is called <u>UNBIASED</u>

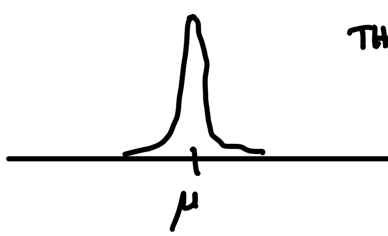
<u>DEFN</u> Suppose $X \sim f(x; \theta)$. Let $T = T(X)$ be an estimator of θ (That is, T is any function of X which does not depend on θ). T is called <u>UNBIASED</u> if

$$\mathbb{E}[T] = \theta$$

Note: The distribution of T depends on θ since T is a function of X, and $X \sim f(x; \theta)$. So $\mathbb{E}[T]$ is in general a function of θ. T is unbiased if this function is the identity.

This means the distribution of T is centered about the quantity we are trying to recover. However, T may still be spread out around θ.

<u>Ex</u> Normal as above. $\overline{X}$ is unbiased, and $Var(\overline{X}) = \sigma^2/n$, so the dispersion of the distribution about the unknown μ is shrinking as $n \to \infty$



THUS THE ACCURACY OF THE ESTIMATOR INCREASES WITH n

Exercise: Let $X_1, \dots, X_n$ be IID Poisson(λ). Find the Maximum Likelihood Estimator (MLE) of λ. Note

$$p(k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

BE CAREFUL:

Let $X_1, \dots, X_n$ be IID UNIF$[0, \theta]$

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$$= \frac{1}{\theta} I_{[0,\theta]}(x)$$

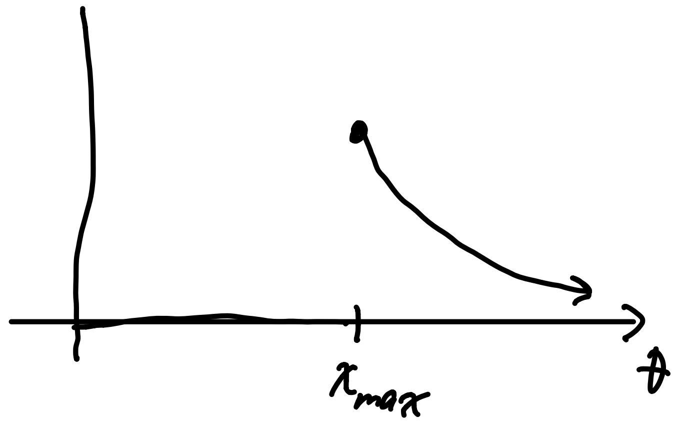
$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{1}{\theta} I_{[0,\theta]}(x_i)$$

$$= \theta^{-n} \prod_{i=1}^{n} I_{[0,\theta]}(x_i)$$

$$= \begin{cases} \theta^{-n} & \text{if } 0 \le x_1, \dots, x_n \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \theta^{-n} & \text{if } x_{max} \le \theta \\ 0 & \end{cases}$$

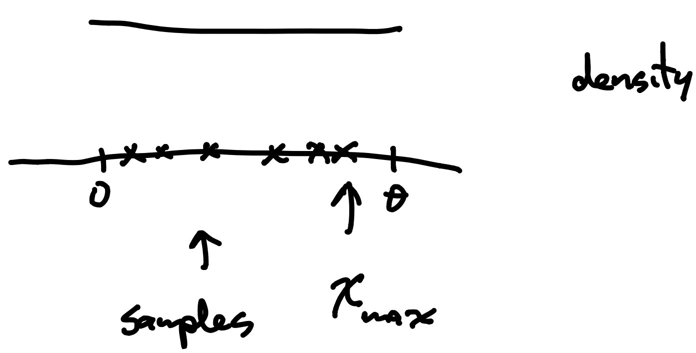
$$L(\theta) = \theta^{-n} I_{[x_{max}, \infty)}(\theta)$$



The value maximizing $L(\theta)$ is x_{max}

Exercise: Find c so that $c\,x_{max}$ is unbiased estimator of θ

Note that x_{max} itself is BIASED:



$x_{max} < \theta$ ALWAYS