2/2/2021. LECTURE 1.

INTRODUCTION.

Consider a problem when we wish to make inference about a parameter θ given observations, or data, x.

Classical Setting (likelihood based inference)

Data is treated as if it is random even after it has been observed

Parameter is viewed as a fixed unknown constant. Consequently, no probability distribution can be attached to the parameter.

Bayesian approach.

Parameters, having not yet been observed, are treated as random variables. Therefore, they possess a probability distribution.

Data having been observed, is treated as fixed.

Example.

Consider n independent Bernoulli trials in which we observe x, the number of times an event occurs. We are interested in making inferences about θ, the probability of the event occurring in a single trial.

Classical approach to this problem.

Prior to observing the data, the probability of observing x was:

$$(f_\theta(x)) \leftarrow f(x|\theta) = P(X = x \mid \theta) = \binom{n}{x} \theta^{x} (1-\theta)^{n-x}$$

$$x \in \{0, 1, \ldots, n\}$$

This is a function of (the future) x for known θ.

If we know x, we could treat this as a function of θ, $L(\theta)$, the Likelihood function.

One strategy is to use the value of θ which maximises the likelihood.

The mle is $\frac{x}{n}$ with corresponding estimator $\frac{X}{n} = T(X)$.

Properties of the estimator (bias, consistency, ...) depend upon the sampling distribution of $T(X)$ given θ

i.e. distributions derived from $f(\underset{\text{random}}{x} \mid \underset{\text{fixed}}{\theta})$

[The distribution of $X \mid \theta$ e.g. $X \mid \theta \sim Bin(n, \theta)$]

* Such an approach can lead to nonsensical answers. *

Example.

Suppose in the Bernoulli example, we wish to estimate θ^2.

An intuitive estimator (and also the mle) is

$$\left(\frac{X}{n}\right)^2$$

This is biased.

$$\begin{aligned}
\mathbb{E}(X^2 \mid \theta) &= Var(X \mid \theta) + \mathbb{E}^2(X \mid \theta) \\
&= n\theta(1-\theta) + n^2\theta^2 \\
&= n\theta + n(n-1)\theta^2 \\
&= \mathbb{E}(X \mid \theta) + n(n-1)\theta^2
\end{aligned}$$

i.e. $$\theta^2 = \frac{\mathbb{E}(X^2 \mid \theta) - \mathbb{E}(X \mid \theta)}{n(n-1)}$$ (assuming $n > 1$)

$$= \mathbb{E}\left(\frac{X^2 - X}{n(n-1)} \,\middle|\, \theta\right)$$

Then, for $n > 1$, $\dfrac{X^2 - X}{n(n-1)} = \dfrac{X(X-1)}{n(n-1)}$ is an unbiased estimator of θ^2.

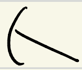
Suppose we observe $x = 1$ (i.e. see a success and so I <u>know</u> $\theta \neq 0$ and thus that $\theta^2 \neq 0$). Our unbiased estimate of θ^2 is 0.

- We estimate a chance as zero even though the event has occurred.

Now let us consider two types of Bernoulli trials.

1. Toss a coin n times and observe x heads, parameter θ_c

2. Toss a drawing pin n times and observe x pin-ups

 pin up ,  pin down

parameter θ_p.

The maximum likelihood estimates for θ_c and θ_p are identical (x/n) and share the same properties.

Classical statistics treats these two situations identically.

Is this sensible ?

I have lots of experience of tossing coins and these are well known to have propensities near to $1/2$.

I have some (prior) knowledge about θ_c.

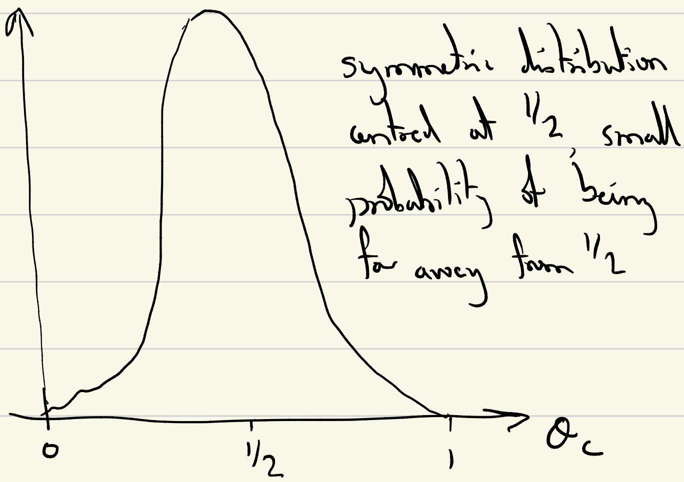
Equally, I have little knowledge about the properties of drawing pins

I don't really know much about θ_p.

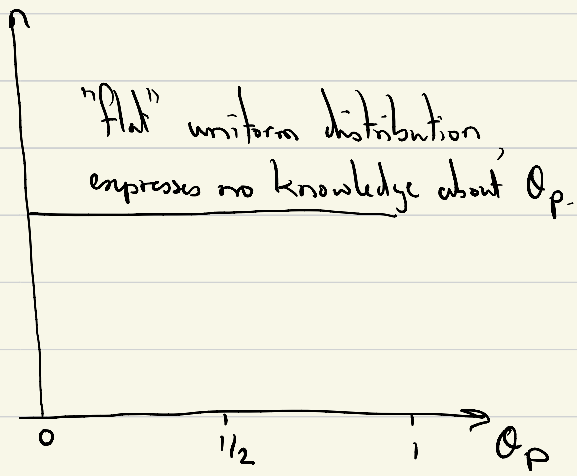
Shouldn't I take these differences into account somehow?

Classical approach provides NO SCOPE for taking into account different prior knowledge. It treats both θ_c and θ_p as fixed unknown constants.

Bayesian approach: reflect prior knowledge about θ_c and θ_p by specifying probability distributions for them.



Coins



Drawing Pins