# Lecture 1B
# MTH6102: Bayesian Statistical Methods

Eftychia Solea

Queen Mary University of London

2023

# Today's agenda

Today's lecture will cover:

- Review
- Continue with MLE
- Assessing uncertainty in classical statistics.

# Frequentist statistics vs Bayesian statistics

- Frequentist statistics: Defines probability as long-term frequency in a repeatable random experiment.
- Bayesian statistics uses probability to quantify degree of belief in hypotheses/parameters.

- In frequentist statistics, $\theta$ is fixed. In Bayesian $\theta$ is random variable.

- Frequentists put probability distributions on (random, repeatable, experimental) data given a parameter, while Bayesians put probability distributions on everything (parameters and data).

- Only the likelihood has meaning in both frequentist and Bayesian statistics.

# Review: MLE

- The maximum likelihood estimate (MLE) is a way to estimate the value of a parameter of interest.

- The idea of MLE is to choose the parameter value such that the observed data have the biggest probability.

- **Definition:** Given the observed data $y$, the maximum likelihood estimate (MLE) for the parameter $\theta$ is the value of $\theta$ that maximises the likelihood function defined as

$$\mathcal{L}(\theta|y) = p(y \mid \theta), \quad \theta \in \Theta,$$

  where $\Theta$ is the range of values of $\theta$.
- That is, the MLE is the value of $\theta$ for which the observed data $y$ is most likely.

# Review: MLE

- If is often easier to work with the natural log of the likelihood function
- For short this is simply called the log likelihood, defined as

$$\ell(\theta; y) = \log \mathcal{L}(\theta|y).$$

- Since $\ln(x)$ is an increasing function, the maxima of the likelihood and log likelihood coincide.

# Review: MLE

**Binomial example.**

A coin is flipped $n$ times. Given that there were $k$ heads, find the maximum likelihood estimate for the probability $q$ of heads on a single toss.

- Let $Y$ be the number of heads. Then, $Y$ is binomially distributed $Y \sim Bin(n, q)$.
- Data: The data is the result of the experiment. In this case it is $y = k$ heads. Given $k$, the likelihood is

$$\mathcal{L}(q|k) = p(k \mid q) = \binom{n}{k} q^k (1-q)^{n-k}, \quad q \in (0,1).$$

- Note the likelihood is a function of $q$

# Review: MLE

**Binomial example**

- The log likelihood is

$$\ell(q;k) = \log \binom{n}{k} + k\log(q) + (n-k)\log(1-q).$$

- To find the MLE, $\hat{q}$, we use calculus. Take the derivative of the log likelihood function and set it to 0, to obtain

$$\frac{d}{dq}\ell(q;k) = \frac{k}{q} - \frac{n-k}{1-q} = 0.$$

- Solving this for $q$ we get

$$\hat{q} = \frac{k}{n}.$$

- This is an extreme point in the interior of the domain $0 < q < 1$.

**Binomial example**

- Note: Our goal is to find a global maximum!
- We check that the critical point is indeed a maximum with the second derivative

$$\frac{d^2}{dq^2}\ell(q;k) = -\frac{k}{q^2} - \frac{n-k}{(1-q)^2} < 0 \quad \forall\, 0 < q < 1.$$

- Then, $\hat{q}$ is indeed a global maximum since it is the only critical point in the interior and is a maximum

**Binomial example**

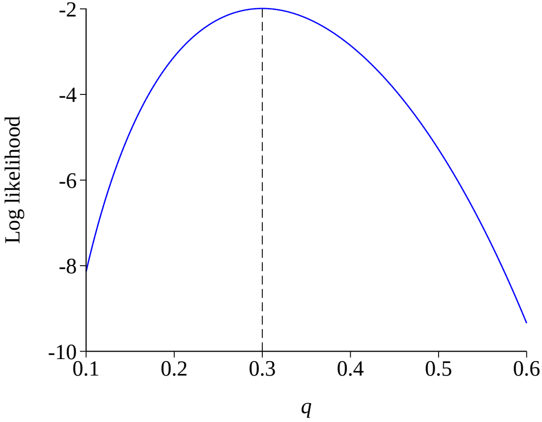
- For the binomial example, as $q \to 0$ or $1$, $\ell \to -\infty$.
- So the stationary point must be a global maximum.
- Hence the MLE is $\hat{q} = \frac{Y}{n}$.

**Binomial example**

$n = 40$

$k = 12$

- $\hat{q} = \frac{k}{n} = 0.3$



# Board example: Coins

A coin is taken from a box containing three coins, which gives heads with probability $q = 1/3$, $q = 1/2$, and $q = 2/3$. The mystery coin is tossed 80 times, resulting in 49 heads and 31 tails.

- What is the likelihood of this data for each type of coin? Which coin gives the maximum likelihood?
- Now suppose that we have a single coin with unknown probability $q$ of landing heads. Find the likelihood and log likelihood functions given the same data. What is the MLE for $q$?

Work from scratch. Set the problem by defining random variables and pmf.

# An example with continuous data

**Example: Light bulbs**

- The time until failure for a type of light bulb is exponentially distributed with parameter $\lambda$.
- We tested $n$ bulbs and observe independently failure times $t = (t_1, \ldots, t_n)$.
- The unknown parameter is $\lambda$.
- Find the likelihood function and the log likelihood function
- Find the MLE for $\lambda$

# Board example: Light bulbs

Suppose 5 bulbs are tested and have lifetimes of $2, 3, 1, 3, 4$ years, respectively.

- Find the MLE of $\lambda$.

Work from scratch. Set the problem by defining random variables and pmf.

# MLE: Normal example

- $Y_1, \ldots, Y_n$ be an i.i.d $\sim N(\mu, \sigma^2)$.
- There are two unknown parameters.
- So $\theta$ is a vector, $\theta = (\mu, \sigma)$.
- **Exercise:** Find the likelihood function and the MLE of $\theta$.

# Assessing uncertainty

- Suppose we want to estimate a population parameter $\theta$.
- In frequentist statistics the idea is to design an estimator $\hat{\theta}$, where an estimator is a function of the data.
- Sample statistics or estimators vary from sample to sample (they will not match the parameter exactly)
- We usually want to assess the uncertainty in any estimate.

# Assessing uncertainty

- **KEY QUESTIONS**: For a given sample statistic, what are plausible values for the population parameter? How much uncertainty surrounds the sample statistic?
- **KEY ANSWER:** It depends on how much the statistic varies from sample to sample!
- In frequentist statistics, two common summaries of the uncertainty are:
  - the standard error;
  - a confidence interval.
- The quantify the uncertainty in $\hat{\theta}$ due to random variation in the data we might have observed.

# Sampling distribution

- Frequentist statistics uses the idea of the sampling distribution.
- If we could repeatedly generate data from a certain model, we would get a distribution of values for $\hat{\theta}$.
- This is the sampling distribution for $\hat{\theta}$.

A sampling distribution is the distribution of sample statistics computed for different samples of the same size from the same population.

- A sampling distribution shows us how the sample statistic varies from sample to sample

# Standard error

The standard error of $\hat{\theta}$ is the standard deviation of the sampling distribution

- It quantifies the spread or the variability of the sampling distribution.
- So this is the simplest summary of the uncertainty in $\hat{\theta}$.
- It measures how much the statistic varies from sample to sample and quantifies the uncertainty in $\hat{\theta}$ due to random variation in the data we might have observed.

# History of likelihood

- The use of likelihood in frequentist statistics was mainly developed by Ronald Fisher.
- "On an Absolute Criterion for Fitting Frequency Curves"
- Published in 1912, while he was a maths undergraduate.
- Later papers more fully developed the theory.