Statistical Models and Maximum Likelihood

Prof. Dennis Sun

Data 401

Statistical Models

In a statistical model, observed data is regarded as random, arising from some underlying random process.

Why might it make sense to model data as random?

• If we take a simple random sample of 1000 Americans to estimate the percentage who approve of the president, then the data is random by construction.
• Randomness can be used to model ignorance. For example, in a linear regression model for house prices:

\[
\text{House Price} = \beta_0 + \beta_1 \cdot (\text{Number of Bedrooms}) \\
+ \cdots \\
+ \beta_K \cdot (\text{Square Footage}) \\
+ \epsilon, \sim \text{Normal}(0, \sigma^2)
\]

the error term \( \epsilon \) is supposed to capture all the other factors that affect house price but are not in the model.
**Probability vs. Statistics**

![Diagram showing the relationship between population/statistics and sample/data, with Binomial distributions.]

**Typical Probability Question:** A coin with probability 0.5 of landing heads is tossed 100 times. What is the probability it lands heads 60 times?

**Typical Statistics Question:** A coin with unknown probability $p$ of landing heads is tossed 100 times. It lands heads 60 times. What is your estimate for $p$?

---

**Estimating $p$**

Recall from STAT 305:

A Binomial($n, p$) random variable $X$ has p.m.f.

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$  

What is the probability that you get 60 heads if you toss a fair ($p = 0.5$) coin 100 times?

Now suppose we don’t know $p$. The probability now becomes

**Idea:** To estimate $p$, choose the value that maximizes this probability!
Method of Maximum Likelihood

**Maximum likelihood** is a recipe for estimating unknown parameters from observed data:

1. Write down the p.m.f. or p.d.f. of the data.
2. Plug in any known parameters (e.g., \( n = 100 \)) and the observed data (e.g., \( x = 60 \)).
3. Now, this function should solely be a function of the unknown parameter. We call this function the likelihood, denoted \( L \).

\[
L(p) = \frac{100!}{60!(100-60)!} p^{60} (1 - p)^{100-60}
\]

4. Find the value of the unknown parameter that maximizes the likelihood. This value is the **maximum likelihood estimate**, or MLE.

Finding the MLE

**In-Class Exercise**

A coin with unknown probability \( p \) of landing heads is tossed 100 times. It lands heads 60 times. Use calculus to find the maximum likelihood estimate of \( p \).
More Practice with Maximum Likelihood

In-Class Exercise
The reading of a certain voltmeter is Normal(μ, σ = 0.5), where μ is the true voltage. Suppose you connect the voltmeter across a battery, and it reads 3.5V. What is your estimate of the voltage of the battery?

More Practice with Maximum Likelihood

In-Class Exercise
Suppose the time that I have to wait at the traffic light on Highland is modeled as Exponential(λ). This morning, I had to wait 0.5 minutes to cross Highland. What is your estimate of λ?