Discussion 5

10 Oct 23

Announcements

- Midterm on Friday, please contact Prof Pawel/TAs for special assistance

Population vs Sample

Population

- The universe of possible data for a specified object, such as people, places, objects, etc.
- Not observable.
- Example: People (or IP addresses) who have visited or will visit a website.

Sample

- A subset of the population that is observable.
- Example: People who visited the website on a specific day.

The Challenge: Inference from Sample to Population

- Making inferences about the population based on sample data is a common challenge in statistics and parameter estimation.
- Example: Election Polling
- Imagine you want to predict the outcome of a national election (population) based on a survey of 1,000 registered voters (sample).
- The challenge is to use the survey results to draw accurate conclusions about the entire voting population.

Introduction to parameter estimation

- Used to make inferences about unknown population parameters based on sample data
- Involves the estimation of values such as population means, variances, proportions, or other key characteristics that describe a population.

i.i.d

- Independent: Data points are independent if the occurrence or value of one data point does not affect the occurrence or value of another. In other words, they are unrelated.
- Identically Distributed: Data points are identically distributed if they all follow the same probability distribution with the same parameters.
- Example: flipping a coin multiple times

Point estimation

- Aims to find a single value that represents the most likely value of the parameter.
- $\hat{\theta}$ (point estimate) represents the estimated value of the population parameter θ.
- Example: Suppose you want to estimate the mean income (μ) of all households in a city. You collect a random sample of 200 households and calculate the sample mean income ($\bar{x}$) to be $55,000. In this case, $55,000 is your point estimate for the mean income of all households in the city (μ).
- Point estimate varies from sample to sample.

Bias and variance

Bias measures the difference between the expected value of the estimator and the true value of the parameter.

Variance measures how much the estimator can vary as a function of the data sample.

Bias-Variance Trade-Off:

- Methods with lower bias tend to have higher variance and vice versa.
- Ideally, low variance and low bias is preferred

Steps to find MLE

1. Find the likelihood function (joint probability)
2. Find the log-likelihood function
3. Find the derivative of log-likelihood function
4. Put it equal to 0 and solve

Likelihood function

Mathematically, it is represented as L(θ) = f(x|θ), where:

- L(θ): Likelihood function.
- f(x|θ): Probability density function (PDF) or probability mass function (PMF) of the data, given θ.

L(θ) is typically calculated as the product (for independent observations) or joint distribution (for dependent observations) of the data points, given θ.

It is simply the joint probability.

$$p_{\text{model}}(X; \theta) = \prod_{i=1}^{m} p_{\text{model}}(x^{(i)}; \theta).$$

Property	Logarithm base b	Natural Log (base e)	Examples
1. Product Property	$\log_b(xy) = \log_b x + \log_b y$	$\ln(xy) = \ln x + \ln y$	$\log_3(10) = \log_3(5 \cdot 2)$ $= \log_3 5 + \log_3 2$
2. Quotient Property	$\log_b\left(\frac{x}{y}\right) = \log_b x - \log_b y$	$\ln\left(\frac{x}{y}\right) = \ln x - \ln y$	$\log_3\left(\frac{5}{7}\right) = \log_4 5 - \log_4 7$
3. Powers Property	$\log_b x^p = p\log_b x$	$\ln x^p = p\ln x$	$\ln 27 = \ln 3^3 = 3\ln 3$
4. Root Property	$\log_b \sqrt[p]{x} = \frac{1}{p}\log_b x$	$\ln \sqrt[p]{x} = \frac{1}{p}\ln x$	$\log_2 \sqrt[3]{y} = \frac{1}{3}\log_2 y$
5. Inverse Property	$\log_b b^x = x$ or $b^{\log_b x} = x$	$\ln e^x = x$ or $e^{\ln x} = x$	$\log_3 3^4 = 4$
6. Identity Property	$\log_b b = 1$	$\ln e = 1$	$\log_{\sqrt{4}} \sqrt{4} = 1$
7. Zero Property	$\log_b 1 = 0$	$\ln 1 = 0$	$\log_4 1 = 0$
8. Change of base Property	$\log_b x = \frac{\log_a x}{\log_a b}$	$\ln x = \frac{\log_a x}{\log_a e}$	$\log_5 6 = \frac{\log 6}{\log 5}$
9. Equality Property	If $\log_b x = \log_b y$ then $x = y$	If $\ln x = \ln y$ then $x = y$	$\log_5 x = \log_5 6$ $x = 6$
10. Reciprocal Property	$\log_b \frac{1}{x} = -\log_b x$	$\ln \frac{1}{x} = -\ln x$	$\ln \frac{1}{5} = -\ln 5$

MLE Example

Say we have a random variable X that tells you the number of chocolates a person gets on their birthday.

We have a sample of say 50 students. So our random variables are:

$$X_i = \text{number of chocolates}$$

and we have:

$$X_1, X_2, \ldots X_{50}$$

Let's just assume that

$$\sum X_i = 150$$

This is to just keep things simple.

To find the mean of the number of chocolates owned by a person:

We have the parameter λ for a Poisson distribution.

For Poisson distribution:

$$p(X_i|\lambda) = \frac{e^{-\lambda}\lambda^{X_i}}{X_i!}$$

For Poisson distribution:

$$p(X_i|\lambda) = \frac{e^{-\lambda}\lambda^{X_i}}{X_i!}$$

So, our likelihood function becomes:

$$L(\lambda|X_1, X_2, \ldots X_{50}) = \prod_{i=1}^{50} \frac{e^{-\lambda}\lambda^{X_i}}{X_i!}$$

For Poisson distribution:

$$p(X_i|\lambda) = \frac{e^{-\lambda}\lambda^{X_i}}{X_i!}$$

So, our likelihood function becomes:

$$L(\lambda|X_1, X_2, \ldots X_{50}) = \prod_{i=1}^{50} \frac{e^{-\lambda}\lambda^{X_i}}{X_i!}$$

Taking the log likelihood,

$$l(\lambda|X_1, X_2, ... X_{50}) = log(\prod_{i=1}^{50} \frac{e^{-\lambda}\lambda^{X_i}}{X_i!})$$

$$= \sum_{i=1}^{50} [-\lambda + x_i log\lambda - log(x_i!)]$$

Now we drop the last term since it doesn't have lambda and it's a constant.

So here, we only have to worry about the first two terms.

For the first time, we are just adding lambda 50 times. And the second term is a factorial.

$$= -50\lambda + \sum_{i=1}^{50} x_i log\lambda$$

This is the log-likelihood.

Find the derivative of the log-likelihood with respect to λ:

$$\frac{d}{d\lambda} l(\lambda | X_1, X_2, \ldots X_{50}) = -50 + \sum_{i=1}^{50} \frac{x_i}{\lambda}$$

Set the derivative equal to 0 to find the MLE for λ:

$$0 = -50 + \sum_{i=1}^{50} \frac{x_i}{\lambda}$$

Now, isolate λ:

$$50 = \sum_{i=1}^{50} \frac{x_i}{\lambda}$$

To solve for λ, you can rearrange the equation:

$$\lambda = \frac{1}{50} \sum_{i=1}^{50} x_i$$

So, the maximum likelihood estimate for λ in the Poisson distribution is:

$$\hat{\lambda} = \frac{1}{50} \sum_{i=1}^{50} x_i$$

$$\hat{\lambda} = \frac{1}{50} \sum_{i=1}^{50} X_i$$

Substituting the given value $\sum X_i = 150$:

$$\hat{\lambda} = \frac{1}{50}(150) = 3$$

So, the Maximum Likelihood Estimate (MLE) for λ is $\hat{\lambda} = 3$. This means that the estimated average number of chocolates owned by a person is 3, based on your sample data of 50 students.