

Section 9: Maximum Likelihood Estimation

Review of Main Concepts

- **Realization/Sample:** A realization/sample x of a random variable X is the value that is actually observed.
- **Likelihood:** Let x_1, \dots, x_n be iid realizations from probability mass function $p_X(x; \theta)$ (if X discrete) or density $f_X(x; \theta)$ (if X continuous), where θ is a parameter (or a vector of parameters). We define the likelihood function to be the probability of seeing the data.

If X is discrete:

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n p_X(x_i; \theta)$$

If X is continuous:

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta)$$

- **Maximum Likelihood Estimator (MLE):** We denote the MLE of θ as $\hat{\theta}_{\text{MLE}}$ or simply $\hat{\theta}$, the parameter (or vector of parameters) that maximizes the likelihood function (probability of seeing the data).

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta} L(x_1, \dots, x_n; \theta) = \arg \max_{\theta} \ln L(x_1, \dots, x_n; \theta)$$

- **Log-Likelihood:** We define the log-likelihood as the natural logarithm of the likelihood function. Since the logarithm is a strictly increasing function, the value of θ that maximizes the likelihood will be exactly the same as the value that maximizes the log-likelihood.

If X is discrete:

$$\ln L(x_1, \dots, x_n; \theta) = \sum_{i=1}^n \ln p_X(x_i; \theta)$$

If X is continuous:

$$\ln L(x_1, \dots, x_n; \theta) = \sum_{i=1}^n \ln f_X(x_i; \theta)$$

- **Bias:** The bias of an estimator $\hat{\theta}$ for a true parameter θ is defined as $\text{Bias}(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta$. An estimator $\hat{\theta}$ of θ is unbiased iff $\text{Bias}(\hat{\theta}, \theta) = 0$, or equivalently $\mathbb{E}[\hat{\theta}] = \theta$.
- **Steps to find the maximum likelihood estimator, $\hat{\theta}$:**
 - (a) Find the likelihood and log-likelihood of the data.
 - (b) Take the derivative of the log-likelihood
 - (c) Set it to 0 to find a candidate for the MLE, $\hat{\theta}$. (note: at this step, we change from the θ to the $\hat{\theta}$ because in this step we are solving for the *maximum* likelihood estimator for θ)
 - (d) Take the second derivative and show that $\hat{\theta}$ indeed is a maximizer, that $\frac{\partial^2 L}{\partial \theta^2} < 0$ at $\hat{\theta}$. Also ensure that it is the global maximizer: check points of non-differentiability and boundary values.

1. Content Review

- (a) **True or False:** The Log-Likelihood gives a slightly different estimate, but because it is close enough and easier to compute we use it for our estimate of θ .

- (b) **True or False:** $\hat{\theta}$ is the true parameter and θ is our estimate.
- (c) **True or False:** An estimator is unbiased if $\text{Bias}(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta = 0$ or equivalently $\mathbb{E}[\hat{\theta}] = \theta$
- (d) You flip a coin 10 times and observe HHHTHHTHHH (8 heads, 2 tails). What is the MLE of θ , where θ is the true probability of seeing tails?
- ☐ $\hat{\theta} = .2$
- ☐ $\hat{\theta} = .25$
- ☐ $\hat{\theta} = .8$
- ☐ $\hat{\theta} = .3$

0. Lemonade Stand

Suppose I run a lemonade stand, which costs me \$100 a day to operate. I sell a drink of lemonade for \$20. Every person who walks by my stand either buys a drink or doesn't (no one buys more than one). If it is raining, n_1 people walk by my stand, and each buys a drink independently with probability p_1 . If it isn't raining, n_2 people walk by my stand, and each buys a drink independently with probability p_2 . It rains each day with probability p_3 , independently of every other day. Let X be my profit over the next week. In terms of n_1, n_2, p_1, p_2 and p_3 , what is $\mathbb{E}[X]$?

2. Mystery Dish!

A fancy new restaurant has opened up which features only 4 dishes. The unique feature of dining here is that they will serve you any of the four dishes randomly according to the following probability distribution: give dish A with probability 0.5, dish B with probability θ , dish C with probability 2θ , and dish D with probability $0.5 - 3\theta$

Each diner is served a dish independently. Let x_A be the number of people who received dish A, x_B the number of people who received dish B, etc, where $x_A + x_B + x_C + x_D = n$. Find the $\hat{\theta}$, the maximum likelihood estimator for θ .

3. A Red Poisson

Suppose that x_1, \dots, x_n are i.i.d. samples from a $\text{Poisson}(\theta)$ random variable, where θ is unknown. Find the MLE for θ .

4. Independent Shreds, You Say?

You are given 100 independent samples x_1, x_2, \dots, x_{100} from $\text{Bernoulli}(\theta)$, where θ is unknown. (Each sample is either a 0 or a 1). These 100 samples sum to 30. You would like to estimate the distribution's parameter θ . Give all answers to 3 significant digits.

- (a) What is the maximum likelihood estimator $\hat{\theta}$ of θ ?
- (b) Is $\hat{\theta}$ an unbiased estimator of θ ?

5. Y Me?

Let y_1, y_2, \dots, y_n be i.i.d. samples of a random variable with density function

$$f_Y(y; \theta) = \frac{1}{2\theta} \exp\left(-\frac{|y|}{\theta}\right).$$

Find the MLE for θ in terms of $|y_i|$ and n .

6. Bird Watching

You are an ornithologist studying a rare species of birds in a nature reserve. Over a period of 50 days, you record the number of sightings of this bird (you see x_1, x_2, \dots, x_{50} birds on each day). Your research has shown that the number of sightings on this species depends on the average number of monkeys in the reserve, θ_1 , and the average number of eagles in the reserve, θ_2 . After years of studying this rare species in other environments, you've found the number of birds observed on a particular day follows the following distribution:

$$p_X(k) = \frac{1}{k!} (\theta_1^k \cdot e^{-\theta_1} \cdot \theta_2^k \cdot e^{-3\theta_2})$$

Find the MLE for θ_1 and θ_2 (i.e., find $\hat{\theta}_1$ and $\hat{\theta}_2$).

- (a) What is the likelihood function?
- (b) What is the log-likelihood function?
- (c) We want to find values of θ_1 and θ_2 that maximize the likelihood function. To do this, we will take the partial derivative with respect to each of these parameters and solve for the values that make them both zero. First, take the partial derivative of the likelihood function with respect to θ_1 .

(d) Now, take the partial derivative with respect to θ_2 .

(e) Set both these partial derivatives to 0, and solve for $\hat{\theta}_1$ and $\hat{\theta}_2$.

7. A biased estimator

In class, we showed that the maximum likelihood estimate of the variance θ_2 of a normal distribution (when both the true mean μ and true variance σ^2 are unknown) is what's called the *population variance*. That is

$$\hat{\theta}_2 = \left(\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\theta}_1)^2 \right)$$

where $\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n x_i$ is the MLE of the mean. Is $\hat{\theta}_2$ unbiased?

8. It Means Nothing

Suppose x_1, x_2, \dots, x_n are samples from a normal distribution whose mean is known to be μ but the variance is unknown. How does the maximum likelihood estimator for the variance differ from the maximum likelihood estimator when both mean and variance are unknown? Which if any is unbiased?

9. Covariance Connection

Let X be the network connection status, where $X = 0$ represents a stable connection and $X = 1$ represents an unstable connection. Let Y be the number of successes in data transmission, taking values in the set $\{0, 1, 2\}$. If $X = 0$, Y follows a Binomial distribution $\text{Bin}(2, 0.8)$, and if $X = 1$, Y follows a Binomial distribution $\text{Bin}(2, 0.3)$. The probabilities for X are given by $P(X = 0) = 0.8$ and $P(X = 1) = 0.2$. Find $\text{Cov}(X, Y)$. (note that we don't know that X and Y are independent here!)

10. Trinomial Distribution

A generalization of the Binomial model is when there is a sequence of n independent trials, but with three outcomes, where $\mathbb{P}(\text{outcome } i) = p_i$ for $i = 1, 2, 3$ and of course $p_1 + p_2 + p_3 = 1$. Let X_i be the number of times outcome i occurred for $i = 1, 2, 3$, where $X_1 + X_2 + X_3 = n$. Find the joint PMF $p_{X_1, X_2, X_3}(x_1, x_2, x_3)$ and specify its value for all $x_1, x_2, x_3 \in \mathbb{R}$.

11. Do You “Urn” to Learn More About Probability?

Suppose that 3 balls are chosen without replacement from an urn consisting of 5 white and 8 red balls. Let $X_i = 1$ if the i -th ball selected is white and let it be equal to 0 otherwise. Give the joint probability mass function of

- (a) X_1, X_2
- (b) X_1, X_2, X_3

12. Successes

Consider a sequence of independent Bernoulli trials, each of which is a success with probability p . Let X_1 be the number of failures preceding the first success, and let X_2 be the number of failures between the first 2 successes. Find the joint pmf of X_1 and X_2 . Write an expression for $E[\sqrt{X_1 X_2}]$. You can leave your answer in the form of a sum.

13. Continuous joint density

The joint density of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} xe^{-(x+y)} & x > 0, y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

and the joint density of W and V is given by

$$f_{W,V}(w,v) = \begin{cases} 2 & 0 < w < v, 0 < v < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent? Are W and V independent?

14. Trapped Miner

A miner is trapped in a mine containing 3 doors.

- D_1 : The 1st door leads to a tunnel that will take him to safety after 3 hours.
- D_2 : The 2nd door leads to a tunnel that returns him to the mine after 5 hours.
- D_3 : The 3rd door leads to a tunnel that returns him to the mine after a number of hours that is Binomial with parameters $(12, \frac{1}{3})$.

At all times, he is equally likely to choose any one of the doors. What is the expected number of hours for this miner to reach safety?

15. 3 points on a line

Three points X_1, X_2, X_3 are selected at random on a line L (continuous independent uniform distributions). What is the probability that X_2 lies between X_1 and X_3 ?