

Section 8: 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 and set it to 0 to find a candidate for the MLE, $\hat{\theta}$.
 - (c) 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. 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 MLE for θ , $\hat{\theta}$.

2. 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 of θ .

3. 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 ?

4. 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 .

5. 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?

6. Faulty Machines

You are trying to use a machine that only works on some days. If on a given day, the machine is working it will break down the next day with probability $0 < b < 1$, and works on the next day with probability $1 - b$. If it is not working on a given day, it will work on the next day with probability $0 < r < 1$ and not work the next day with probability $1 - r$.

(a) In this problem we will formulate this process as a Markov chain. First, let $X^{(t)}$ be a variable that denotes the state of the machine at time t . Then, define a state space $\mathcal{S}$ that includes all the possible states that the machine can be in. Lastly, for all $A, B \in \mathcal{S}$ find $\mathbb{P}(X^{(t+1)} = A \mid X^{(t)} = B)$ (A and B can be the same state).

(b) Suppose that on day 1, the machine is working. What is the probability that it is working on day 3?

(c) As $n \rightarrow \infty$, what does the probability that the machine is working on day n converge to? To get the answer, solve for the *stationary distribution*.

7. Another Markov Chain

Suppose that the following is the transition probability matrix for a 4 state Markov chain (states 1,2,3,4).

$$M = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/3 & 0 & 0 & 2/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/5 & 2/5 & 2/5 & 0 \end{bmatrix}$$

- (a) What is the probability that $X^{(2)} = 4$ given that $X^{(0)} = 4$?
- (b) Write down the system of equations that the stationary distribution must satisfy and solve them.

8. Three Tails

You flip a fair coin until you see three tails in a row. Model this as a Markov chain with the following states:

- S : start state, which we are only in before flipping any coins.
- H : We see a heads, which means no streak of tails currently exists.
- T : We've seen exactly one tail in a row so far.
- TT : We've seen exactly two tails in a row so far.
- TTT : We've accomplished our goal of seeing three tails in a row, stop flipping, and stay there.

- (a) Write down the transition probability matrix.
- (b) Write down the system of equations whose variables are $D(s)$ for each state $s \in \{S, H, T, TT, TTT\}$, where $D(s)$ is the expected number of steps until state TTT is reached starting from state s . Solve this system of equations to find $D(S)$.
- (c) Write down the system of equations whose variables are $\gamma(s)$ for each state $s \in \{S, H, T, TT, TTT\}$, where $\gamma(s)$ is the expected number of heads seen before state TTT is reached. Solve this system to find $\gamma(S)$, the expected number of heads seen overall until getting three tails in a row.