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 $x_1, \ldots x_n$ be iid realizations from probability mass function $p_X(\mathrm{x}\,; \theta)$ (if X discrete) or density $f_X(\mathrm{x}\,; \theta)$ (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:
  $$L\left(x_1, \ldots, x_n \mid \theta\right) = \prod_{i=1}^{n} p_X\left(x_i \mid \theta\right)$$
  If X is continuous:
  $$L\left(x_1, \ldots, x_n \mid \theta\right) = \prod_{i=1}^{n} f_X\left(x_i \mid \theta\right)$$
- **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).
  $$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} L\left(x_1, \ldots, x_n \mid \theta\right) = \arg\max_{\theta} \ln L\left(x_1, \ldots, x_n \mid \theta\right)$$
- **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:
  $$\ln L\left(x_1, \ldots, x_n \mid \theta\right) = \sum_{i=1}^{n} \ln p_X\left(x_i \mid \theta\right)$$
  If X is continuous:
  $$\ln L\left(x_1, \ldots, x_n \mid \theta\right) = \sum_{i=1}^{n} \ln f_X\left(x_i \mid \theta\right)$$
- **Bias**: The bias of an estimator $\hat{\theta}$ for a true parameter θ is defined as $\text{Bias}\left(\hat{\theta}, \theta\right) = \mathbb{E}[\hat{\theta}] - \theta$. An estimator $\hat{\theta}$ of θ is unbiased iff $\text{Bias}\left(\hat{\theta}, \theta\right) = 0$, 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 Klee has a collection of i.i.d. samples, $x_1, \ldots, x_n$, from a Poisson(θ) random variable, where θ is unknown. Find the MLE of θ.

3. Independent Shreds, You Say?

Jean is given 100 independent samples $x_1, x_2, \ldots, x_{100}$ from Bernoulli(θ), where θ is unknown. (Each sample is either a 0 or a 1). These 100 samples sum to 30. She 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 $y_1, y_2, \ldots y_n$ 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?