CSE 312: Foundations of Computing II
Quiz Section #9: Law of Large Numbers, Maximum Likelihood Estimation, Confidence Intervals

Review: Main Theorems and Concepts

Weak Law of Large Numbers (WLLN): Let $X_1, \ldots, X_n$ be iid random variables with common mean μ and variance σ^2. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean for a sample of size n. Then, for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| > \epsilon) = 0$. We say that $\bar{X}_n$ converges in probability to μ.

Strong Law of Large Numbers (SLLN): Let $X_1, \ldots, X_n$ be iid random variables with common mean μ and variance σ^2. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean for a sample of size n. Then, $\mathbb{P}(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$. We say that $\bar{X}_n$ converges almost surely to μ. The SLLN implies the WLLN, but not vice versa.

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 $p_X(\mathrm{x} \mid \theta)$ (if X discrete) or density $f_X(\mathrm{x} \mid \theta)$ (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:

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

If X is continuous:

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

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).

$$\hat{\theta}_{\text{MLE}} = \underset{\theta}{\operatorname{argmax}}\, L(x_1, \ldots, x_n \mid \theta) = \underset{\theta}{\operatorname{argmax}} \ln L(x_1, \ldots, x_n \mid \theta)$$

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:

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

If X is continuous:

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

Bias: The bias of an estimator $\hat{\theta}$ for a true parameter θ is defined as $\text{Bias}\left(\hat{\theta}, \theta\right) = \mathbb{E}\left[\hat{\theta}\right] - \theta$. An estimator $\hat{\theta}$ of θ is unbiased iff $\text{Bias}\left(\hat{\theta}, \theta\right) = 0$, or equivalently $\mathbb{E}\left[\hat{\theta}\right] = \theta$.

Steps to find the maximum likelihood estimator, $\hat{\theta}$:

1. Find the likelihood and log-likelihood of the data.

2. Take the derivative of the log-likelihood and set it to 0 to find a candidate for the MLE, $\hat{\theta}$.

3. Take the second derivative and show that $\hat{\theta}$ indeed is a maximizer, that $\frac{d^2L}{d\theta^2} < 0$ at $\hat{\theta}$. Also ensure that it is the global maximizer: check points of non-differentiability and boundary values.

Confidence Intervals: The probability that the MLE $\hat{\theta}$ of a parameter θ is equal to the true value of θ is 0. We say that $(\hat{\theta} - \Delta, \hat{\theta} + \Delta)$ is a $K\%$ confidence interval for θ if and only if $\mathbb{P}\left(\theta \in \left(\hat{\theta} - \Delta, \hat{\theta} + \Delta\right)\right) \geq K/100$.

Exercises

1. Let $f(x \mid \theta) = \theta x^{\theta-1}$ for $0 \leq x \leq 1$, where θ is any positive real number. Let $x_1, x_2, \ldots, x_n$ be i.i.d. samples from this distribution. Derive the maximum likelihood estimator $\hat{\theta}$.

2. Suppose $x_1, \ldots, x_n$ are iid realizations from density

$$f_X(x \mid \theta) = \begin{cases} \frac{\theta x^{\theta-1}}{3^\theta}, & 0 \leq x \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

   Find the MLE for θ.

3. Suppose $x_1, \ldots, x_{2n}$ are iid realizations from the Laplace density (double exponential density)

$$f_X(x \mid \theta) = \frac{1}{2} e^{-|x-\theta|}$$

   Find the MLE for θ. For this problem, you need not verify that the MLE is indeed a maximizer. You may find the **sign** function useful:

$$\operatorname{sgn}(x) = \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

4. You are given 100 independent samples $x_1, x_2, \ldots, x_{100}$ from Ber(p), where p is unknown. These 100 samples sum to 30. You would like to estimate the distribution's parameter p. Give all answers to 3 significant digits.

   (a) What is the maximum likelihood estimator $\hat{p}$ of p?

   (b) Is $\hat{p}$ an unbiased estimator of p?

   (c) Give your best approximation for the 95% confidence interval of p.

   (d) Give your best approximation for the 90% confidence interval of p.

   (e) Give three different reasons why your answers to (c) and (d) are only approximations.

   (f) Explain why it makes sense that the interval in (d) is bigger (or smaller, depending on your answers) than the interval in (c).

5. Suppose $X_1, \ldots, X_n$ are iid random variables from some distribution with unknown mean θ and known variance σ^2, and your estimate $\hat{\theta}$ for its mean θ is the sample mean $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$. For any α, construct a $100\,(1-\alpha)\,\%$ confidence interval (centered around the estimate $\hat{\theta}$) for the true parameter θ. You may assume n is "sufficiently large".

6. (a) Suppose $x_1, x_2, \ldots, x_n$ are samples from a normal distribution whose mean is known to be zero, but whose variance is unknown. What is the maximum likelihood estimator for its variance?

   (b) Suppose the 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?

7. (a) Suppose that $\hat{\theta}$ is a biased estimator for θ with $\mathbb{E}[\hat{\theta}] = \alpha\theta$, for some constant $\alpha > 0$. Find an unbiased estimator for θ and prove that it is unbiased.

   (b) In lecture, we saw that the maximum likelihood estimator for the population variance θ_2 of $N(\theta_1, \theta_2)$ is the sample variance

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

   where $\hat{\theta}_1$ is the sample mean. It can be shown that $\mathbb{E}[\hat{\theta}_2] = \frac{n-1}{n} \cdot \theta_2$, so that $\hat{\theta}_2$ is biased and always underestimates the variance θ_2. Use your result from part (a) to find an unbiased estimator of the variance θ_2.