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

Review/Mini-Lecture/Main Theorems and Concepts From Lecture

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

Strong Law of Large Numbers (SLLN): Let $X_1, \dots, X_n$ be iid random variables with common mean μ and variance σ^2. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean for a sample of size n. Then, $P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$. 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 $x_1, \dots x_n$ be iid realizations from mass function $p_X(x \mid \theta)$ (if X discrete) or density $f_X(x \mid \theta)$ (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:

$$L(x_1, \dots, x_n \mid \theta) = P\left(\bigcap_{i=1}^{n} \{X = x_i\} \mid \theta\right) = \prod_{i=1}^{n} p_X(x_i \mid \theta)$$

If X is continuous:

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

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

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

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:

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

If X is continuous:

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

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

Steps to find the maximum likelihood estimator, $\hat{\boldsymbol{\theta}}$:
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 MLE $\hat{\theta}$ of a parameter θ is wrong with probability 1. We say that: $(\hat{\theta} - \Delta, \hat{\theta} + \Delta)$ is a $100(1-\alpha)\%$ confidence interval for θ if and only if $P\left(\theta \in (\hat{\theta} - \Delta, \hat{\theta} + \Delta)\right) \geq 1 - \alpha$.

<u>Exercises</u>

1. Suppose $x_1, \dots, x_n$ are iid realizations from density

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

Find the MLE for θ.

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

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

Find the MLE for θ. You may find the **sign** function useful:

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

3. Suppose $X_1, \dots, X_n$ are iid rv's from some distribution with unknown mean θ and known variance σ^2, and your estimate $\hat{\theta}$ for its mean θ will be the sample mean $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$. For full generality, 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".