CSE 312

Section 9: Maximum Likelihood Estimation

Review

- **Weak Law of Large Numbers** (not covered): Let $X_1, \ldots, X_n$ 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 $\epsilon > 0$, $\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$.
  - Basically, as we sample more, the sample mean should converge to the population mean.
- **Realization/Sample**: A realization/sample x of a random variable X is the value that is actually observed.

Review

- **Likelihood**: Let $x_1, \dots, x_n$ be iid samples from pmf $p_X(x;\theta)$ where θ are the distribution's parameters. We define the likelihood function to be the probability of seeing the data given the parameters as

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

  - The continuous case just replaces a pmf with a pdf.
  - Used in statistical learning
  - Bar notation and semicolon notation mean the same thing (semicolon attempts to clear up confusion about conditional probability vs parameters)
    - Re-typesetting takes a long time in Office equations—sorry
  - Sometimes (most times), you'll also see pmfs represented with p and pdfs represented with f—this class used f for both to emphasize they are analogous.

Review (cont.)

- **Maximum Likelihood Estimator**: We denote the MLE of θ as $\hat{\theta}$, the parameters that maximize the likelihood function. Expressed mathematically, we have

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(x_1, \dots, x_n; \theta)$$

  - Note that we often take the argmax of the *(natural) log likelihood*. This is equivalent to the above since the log function is monotone increasing. This is better for numerical stability: we often get floating pointer underflow when multiplying small numbers together, so taking the log lets us add them instead (think back to Naïve Bayes)

Review (cont.)

- **Bias**: The bias of an estimator $\hat{\theta}$ for a ***true*** parameter θ is defined as $E[\hat{\theta}] - \theta$. An estimator is unbiased iff $E[\hat{\theta}] = \theta$.
- **Steps to find the MLE**:
  - (a) Find the likelihood and log-likelihood of the data
  - (b) Take the derivative of the log-likelihood wrt θ and set it to 0 to find $\hat{\theta}$.
  - (c) Take the second derivative and show that $\hat{\theta}$ is a global maximizer (calc 2: second-derivative test)

Problem 1a

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.

- True
- False

Problem 1b

When doing MLE, $\hat{\theta}$ is the true parameter and θ is our estimate.

- True
- False

Problem 1c

An estimator is unbiased if $\text{Bias}(\hat{\theta}, \theta) = E[\hat{\theta}] - \theta = 0$ or equivalently $E[\hat{\theta}] = \theta$

- True
- False

Problem 2

- 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 student is assigned a grad independently. Let x_A be the number of people who received an A, etc., so $x_A + x_B + x_C + x_D = n$. Find the MLE for θ.

Problem 6

- 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. Your research has shown that the number of sightings on this species depends on the number of monkeys living in the reserve, θ_1, and the 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:

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

a) What is the likelihood function?

b) What is the log-likelihood function?

c) What is the partial derivative of the log-likelihood function with respect to θ_1?

d) What is the partial derivative of the log-likelihood function with respect to θ_2?

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

Problem 4b

- You are 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. You would like to estimate the distribution's parameter θ.

+ Is $\hat{\theta}$ an unbiased estimator of θ?

Problem 9

- 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 $Bin(2, 0.8)$, and if $X = 1$, Y follows a Binomial distribution $Bin(2, 0.3)$. The probabilities for X are given by $P(X = 0) = 0.8$ and $P(X = 1) = 0.2$. Find $Cov(X, Y)$.