

Problem Set 6
Fall 2020

1. Consistent Estimator

Let i.i.d. $X_n \sim \mathcal{N}(0, \sigma^2)$ for $n \geq 1$ and σ^2 unknown. Find some estimator Z_n in terms of X_1, \dots, X_n such that Z_n converges in probability to σ^2 .

Remark: these estimators are called consistent, referring to the idea that the estimator will converge in probability to the correct value as the number of samples approaches infinity.

2. Convergence in Probability to a Random Variable

Consider X as a Bernoulli random variable with $p = \frac{1}{2}$. Let X_n be a sequence with $X_n = (1 + \frac{1}{n})X$. Show that X_n converges in probability to X .

3. Random Walk on the Cube

Consider the symmetric random walk on the vertices of the 3-dimensional unit cube where two vertices are connected by an edge if and only if the line connecting them is an edge of the cube. In other words, this is the random walk on the graph with 8 nodes each written as a string of 3 bits, so that the vertex set is $\{0, 1\}^3$, and where two vertices are connected by an edge if and only if their corresponding bit strings differ in exactly one location.

This random walk is modified so that the nodes 000 and 111 are made absorbing.

- (a) What are the communicating classes of the resulting Markov chain? For each class, determine its period, and whether it is transient or recurrent.
- (b) For each transient state, what is the probability that the modified random walk started at that state gets absorbed in the state 000?

4. Confidence Interval Comparisons

In order to estimate the probability of a head in a coin flip, p , you flip a coin n times, where n is a positive integer, and count the number of heads, S_n . You use the estimator $\hat{p} = S_n/n$.

- (a) You choose the sample size n to have a guarantee

$$\mathbb{P}(|\hat{p} - p| \geq \epsilon) \leq \delta.$$

Using Chebyshev Inequality, determine n with the following parameters. Note that you should not have p in your final answer.

- (i) Compare the value of n when $\epsilon = 0.05, \delta = 0.1$ to the value of n when $\epsilon = 0.1, \delta = 0.1$.
- (ii) Compare the value of n when $\epsilon = 0.1, \delta = 0.05$ to the value of n when $\epsilon = 0.1, \delta = 0.1$.
- (b) Now, we change the scenario slightly. You know that $p \in (0.4, 0.6)$ and would now like to determine the smallest n such that

$$\mathbb{P}\left(\frac{|\hat{p} - p|}{p} \leq 0.05\right) \geq 0.95.$$

Use the CLT to find the value of n that you should use. *Recall that the CLT states that the sum of IID random variables tends to a normal distribution with the sample mean and variance as it's parameters for n large enough.*

5. Two-Population Sampling

We are conducting a public opinion poll to determine the fraction p of people who will vote for Mr. Whatshisname as the next president. We ask N_1 college-educated and N_2 non-college-educated people, where N_1 and N_2 are positive integers. We assume that the votes in each of the two groups are i.i.d. Bernoulli(p_1) and Bernoulli(p_2), respectively in favor of Whatshisname. In the general population, the percentage of college-educated people is known to be q .

- (a) What is a 95% confidence interval for p , using an upper bound for the variance?
- (b) How do we choose N_1 and N_2 subject to $N_1 + N_2 = N$ to minimize the width of that interval? (You may ignore the constraint that N_1 and N_2 must be integers.)

6. Metropolis-Hastings Algorithm

In this problem we introduce the **Metropolis-Hastings Algorithm**, which is an example of **Markov Chain Monte Carlo (MCMC)** sampling. In the lab this week, you will implement Metropolis-Hastings and explore its performance.

Suppose that π is a probability distribution on a finite set \mathcal{X} . Assume that we can compute π up to a normalizing constant. Specifically, assume that we can efficiently calculate $\tilde{\pi}(x)$ for any $x \in \mathcal{X}$, where $\pi(x) = \tilde{\pi}(x) / \sum_{x' \in \mathcal{X}} \tilde{\pi}(x')$. The normalizing constant $1 / \sum_{x' \in \mathcal{X}} \tilde{\pi}(x')$ is called the **partition function** in some contexts, and it can be difficult to compute if \mathcal{X} is very large.

Instead of computing π directly, we will use $\tilde{\pi}$ to design an algorithm to *sample* from the distribution π . We can then approximate π if we take enough samples. The idea behind MCMC methods is to design a Markov chain whose stationary distribution is π ; then, we can “run” the chain until it is close to stationarity, and then collect samples from the chain.

Initialize the chain with $X_0 = x_0$, where x_0 is picked arbitrarily from \mathcal{X} . Let $f : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$ be a **proposal distribution**: for each $x \in \mathcal{X}$, $f(x, \cdot)$ is a probability distribution on \mathcal{X} . (In the lab, you will look at what the desirable properties of a proposal distribution are.) If the chain is at state $x \in \mathcal{X}$, the chain makes a transition according to the following rule:

- Propose the next state y according to the distribution $f(x, \cdot)$.
- Accept the proposal with probability

$$A(x, y) = \min\left\{1, \frac{\pi(y) f(y, x)}{\pi(x) f(x, y)}\right\}.$$

- If the proposal is accepted, then move the chain to y ; otherwise, stay at x .

Assume that the proposal distribution f is chosen to make the chain irreducible.

- (a) Explain why the Markov chain can be simulated efficiently, even though π cannot be computed efficiently.
- (b) The key to showing why Metropolis-Hastings works is to look at the **detailed balance equations**. Suppose we have a finite irreducible Markov chain on a state space \mathcal{X} with transition matrix P . Show that if there exists a distribution π on \mathcal{X} such that for all $x, y \in \mathcal{X}$,

$$\pi(x)P(x, y) = \pi(y)P(y, x),$$

- (c) then π is the stationary distribution of the chain. If these equations hold, then the Markov chain is called **reversible** because it turns out that the equations imply that the chain looks the same going forwards as backwards.
- (d) Now return to the Metropolis-Hastings chain. Use detailed balance to argue that π is the stationary distribution of the chain.
- (e) If the chain is aperiodic, then the chain will converge to the stationary distribution. If the chain is not aperiodic, we can force it to be aperiodic by considering the **lazy chain**: on each transition, the chain decides not to move with probability $1/2$ (independently of the propose-accept step). Explain why the lazy chain is aperiodic, and explain why the stationary distribution is the same as before.