1 Chebyshev's Inequality vs. Central Limit Theorem

Let n be a positive integer. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with the following distribution:

$$\mathbb{P}[X_i = -1] = \frac{1}{12}; \qquad \mathbb{P}[X_i = 1] = \frac{9}{12}; \qquad \mathbb{P}[X_i = 2] = \frac{2}{12}.$$

(a) Calculate the expectations and variances of X_1, $\sum_{i=1}^n X_i$, $\sum_{i=1}^n (X_i - \mathbb{E}[X_i])$, and

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}.$$

(b) Use Chebyshev's Inequality to find an upper bound b for $\mathbb{P}[|Z_n| \geq 2]$.

(c) Can you use b to bound $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$?

(d) As $n \to \infty$, what is the distribution of Z_n?

(e) We know that if $Z \sim \mathcal{N}(0,1)$, then $\mathbb{P}[|Z| \leq 2] = \Phi(2) - \Phi(-2) \approx 0.9545$. As $n \to \infty$, can you provide approximations for $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$?

Solution:

(a) $\mathbb{E}[X_1] = -1/12 + 9/12 + 4/12 = 1$, and

$$\operatorname{var} X_1 = \frac{1}{12} \cdot 2^2 + \frac{9}{12} \cdot 0^2 + \frac{2}{12} \cdot 1^2 = \frac{1}{2}.$$

Using linearity of expectation and variance (since $X_1, \ldots, X_n$ are independent), we find that $\mathbb{E}[\sum_{i=1}^n X_i] = n$ and $\operatorname{var}(\sum_{i=1}^n X_i) = n/2$.

Again, by linearity of expectation, $\mathbb{E}[\sum_{i=1}^n X_i - n] = n - n = 0$. Subtracting a constant does not change the variance, so $\operatorname{var}(\sum_{i=1}^n X_i - n) = n/2$, as before.

Using the scaling properties of the expectation and variance, $\mathbb{E}[Z_n] = 0/\sqrt{n/2} = 0$ and $\operatorname{var} Z_n = (n/2)/(n/2) = 1$.

(b)

$$\mathbb{P}[|Z_n| \geq 2] \leq \frac{\operatorname{var} Z_n}{2^2} = \frac{1}{4}$$

(c) $1/4$ for both, since $\mathbb{P}[Z_n \geq 2] \leq \mathbb{P}[|Z_n| \geq 2]$ and $\mathbb{P}[Z_n \leq -2] \leq \mathbb{P}[|Z_n| \geq 2]$.

(d) By the Central Limit Theorem, we know that $Z_n \to \mathcal{N}(0,1)$, the standard normal distribution.

(e) Since $Z_n \to \mathcal{N}(0,1)$, we can approximate $\mathbb{P}[|Z_n| \geq 2] \approx 1 - 0.9545 = 0.0455$. By the symmetry of the normal distribution, $\mathbb{P}[Z_n \geq 2] = \mathbb{P}[Z_n \leq -2] \approx 0.0455/2 = 0.02275$.

It is interesting to note that the CLT provides a much smaller answer than Chebyshev. This is due to the fact that the CLT is applied to a particular kind of random variable, namely the (scaled) sum of a bunch of random variables. Chebyshev's inequality, however, holds for any random variable, and is therefore weaker.

2 Markov Chain Basics

A Markov chain is a sequence of random variables X_n, $n = 0,1,2,\ldots$. Here is one interpretation of a Markov chain: X_n is the state of a particle at time n. At each time step, the particle can jump to another state. Formally, a Markov chain satisfies the Markov property:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i), \tag{1}$$

for all n, and for all sequences of states $i_0, \ldots, i_{n-1}, i, j$. In other words, the Markov chain does not have any memory; the transition probability only depends on the current state, and not the history of states that have been visited in the past.

(a) In lecture, we learned that we can specify Markov chains by providing three ingredients: $\mathcal{X}$, P, and π_0. What do these represent, and what properties must they satisfy?

(b) If we specify $\mathcal{X}$, P, and π_0, we are implicitly defining a sequence of random variables X_n, $n = 0,1,2,\ldots$, that satisfies (**??**). Explain why this is true.

(c) Calculate $\mathbb{P}(X_1 = j)$ in terms of π_0 and P. Then, express your answer in matrix notation. What is the formula for $\mathbb{P}(X_n = j)$ in matrix form?

Solution:

(a) $\mathcal{X}$ is the set of states, which is the range of possible values for X_n. In this course, we only consider finite $\mathcal{X}$.

P contains the transition probabilities. $P(i,j)$ is the probability of transitioning from state i to state j. It must satisfy $\sum_{j \in \mathcal{X}} P(i,j) = 1 \; \forall i \in \mathcal{X}$, which says that the probability that *some* transition occurs must be 1. Also, the entries must be non-negative: $P(i,j) \geq 0 \; \forall i,j \in \mathcal{X}$. A matrix satisfying these two properties is called a stochastic matrix.

Note that we allow states to transition to themselves, i.e. it is possible for $P(i,i) > 0$.

π_0 is the initial distribution, that is, $\pi_0(i) = \mathbb{P}(X_0 = i)$. Similarly, we let π_n be the distribution of X_n. Since π_0 is a probability distribution, its entries must be non-negative and $\sum_{i \in \mathcal{X}} \pi_0(i) = 1$.

(b) The sequence of random variables X_n, $n = 0, 1, 2, \ldots$, is defined in the following way:

- X_0 has distribution π_0, i.e. $\mathbb{P}(X_0 = i) = \pi_0(i)$.
- X_{n+1} has distribution given by

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i) = P(i, j),$$

for all $n = 0, 1, 2, \ldots$.

It is important to realize the connection between the Markov property (**??**) and the transition matrix P. P contains information about the transition probabilities in one step. If the Markov property did not hold, then P would not be enough to specify the distribution of X_{n+1}. Conversely, if we only specify P, then we are implicitly assuming that the transition probabilities do not depend on anything other than the current state. Note that this convention is different from what EE16A uses, if you have taken that class/are taking it right now.

(c) By the Law of Total Probability,

$$\mathbb{P}(X_1 = j) = \sum_{i \in \mathscr{X}} \mathbb{P}(X_1 = j, X_0 = i) = \sum_{i \in \mathscr{X}} \mathbb{P}(X_0 = i)\mathbb{P}(X_1 = j \mid X_0 = i) = \sum_{i \in \mathscr{X}} \pi_0(i)P(i, j).$$

If we write $\pi_1(j) = \mathbb{P}(X_1 = j)$ and π_0 as row vectors, then in matrix notation we have

$$\pi_1 = \pi_0 P.$$

The effect of a transition is right-multiplication by P. After n time steps, we have

$$\pi_n = \pi_0 P^n.$$

At this point, it should be mentioned that many calculations involving Markov chains are very naturally expressed with the language of matrices. Consequently, Markov chains are very well-suited for computers, which is one of the reasons why Markov chain models are so popular in practice.

3 Playing Blackjack

You are playing a game of Blackjack where you start with \$100. You are a particularly risk-loving player who does not believe in leaving the table until you either make \$400, or lose all your money. At each turn you either win \$100 with probability p, or you lose \$100 with probability $1 - p$.

(a) Formulate this problem as a Markov chain i.e. define your state space, transition probabilities, and determine your starting state.

(b) Find the probability that you end the game with \$400.

Solution:

(a) Since it is only possible for us to either win or lose \$100, we define the following state space $\chi = \{0, 100, 200, 300, 400\}$. The following are the transition probabilities:

$$\mathbb{P}(0,0) = \mathbb{P}(400,400) = 1$$
$$\mathbb{P}(i, i+100) = p \text{ for } i \in \{100, 200, 300\}$$
$$\mathbb{P}(i, i-100) = 1-p \text{ for } i \in \{100, 200, 300\}$$

(b) We want to find the probability that we are "absorbed" by state 400 before we are absorbed by state 0. We can calculate this probability by leveraging the memoryless property of Markov Chains. Define a_i as the probability of reaching state 400 before 0 starting at state i.

We also know that for $i \in \{100, 200, 300\}$, we have the following relation:

$$a_i = (1-p)a_{i-100} + pa_{i+100} \text{ for } i \in \{100, 200, 300\}$$

We also know that $a_0 = 0$, since if you are at state 0, then there is no chance that you end up at state 400. We also have $a_{400} = 1$ since if we are at state 400, then we have already succeeded in our goal to reach 400.

We have three unknowns $(a_{100}, a_{200}, a_{300})$ and three equations, and we can now solve this system of equations for a_{100}.

$$a_0 = 0, a_{400} = 1$$
$$\implies a_i = (1-p)a_{i-100} + pa_{i+100} \text{ for } i \in \{100, 200, 300\}$$
$$a_{100} = pa_{200}$$
$$a_{200} = (1-p)a_{100} + pa_{300} \implies a_{200}[1 - p(1-p)] = pa_{300}$$
$$\implies a_{200} = \frac{pa_{300}}{1-p(1-p)}$$
$$a_{300} = (1-p)a_{200} + p \implies a_{300} = \frac{(1-p)pa_{300}}{1-p(1-p)} + p$$
$$\implies a_{300} = \frac{p(1-p(1-p))}{1-2p(1-p)}$$
$$\implies a_{200} = \frac{p^2}{1-2p(1-p)}$$
$$\implies a_{100} = \frac{p^3}{1-2p(1-p)}$$

This problem is called Gambler's Ruin, where it is used to show that even if p is decently large, after playing a large number of games without stopping, you will end up at 0 dollars with high probability. Let's look at a nicer way to solve the recurrence relation that gives a somewhat more insightful answer to the problem.

Suppose we have states 0 through N, and you start at state k. You go up a state with probability p and go down with probability $1-p$. You win if you end up at state N, and lose if you end up at state 0.

Again, we can write the recurrence relation as

$$a_i = pa_{i+1} + (1-p)a_{i-1}$$

for $1 \leq i \leq N-1$. We also know that $a_N = 1$ and $a_0 = 0$. We can rewrite the recurrence relation into the following form:

$$(1-p)(a_i - a_{i-1}) = p(a_{i+1} - a_i) \Rightarrow a_{i+1} - a_i = \frac{1-p}{p}(a_i - a_{i-1})$$

Define $w = \frac{1-p}{p}$, which is often called the odds ratio, and define $b_i = a_{i+1} - a_i$. Note that this tells us $a_i = b_0 + b_1 + \cdots + b_{i-1}$. So, the recurrence we have derived is $b_i = w \cdot b_{i-1}$. This tells us that $b_i = w^i b_0$, and

$$a_i = b_0 + \cdots + b_{i-1} = (1 + w + w^2 + \cdots + w^{i-1})b_0$$

What is b_0? We can now use our information that $a_N = 1$, to see that $b_0 = \frac{1}{1+w+w^2+\cdots+w^{N-1}}$. Thus, we finally see that

$$a_i = \frac{1 + w + w^2 + \cdots + w^{i-1}}{1 + w + w^2 + \cdots + w^{N-1}} = \frac{w^i - 1}{w^N - 1}$$

where we used the geometric series formula in the last step: $1 + w + w^2 + \cdots + w^{i-1} = \frac{w^i-1}{w-1}$. Note that the formula only works if $w \neq 1$.

By the way, if you are interested in how to derive the geometric series, first write it like this:

$$S = 1 + w + w^2 + \cdots + w^{i-1}$$

multiply both sides by w, to get

$$wS = w + w^2 + \cdots + w^i$$

and subtracting these two equations will cancel most of the terms! We get:

$$(w-1)S = w^i - 1$$

solving for S yields $\frac{w^i-1}{w-1}$.