Inequalities, LLN, CLT, Markov Chains

TFs: Justin Zhu

Practice Problems

Example 1. Sample Mean.
For i.i.d. r.v.s. $X_1, X_2, \dots X_n$ with mean μ and variance σ^2, give a value of n (a specific number) that will ensure that there is at least a 99% chance that the sample mean will be within 2 standard deviations of the true mean μ.

Solution
Let $\bar{X}$ be the sample mean, which is equal to $\frac{X_1+\ldots+X_n}{n}$. Saying that there is at least a 99% chance that $|\bar{X}_n - \mu| < 2\sigma$ is the same as saying that there is at most a 1% chance for $|\bar{X}_n - \mu| > 2\sigma$. Thus, we want to calculate an n such that:

$$P\left(|\bar{X}_n - \mu| > 2\sigma\right) < 0.01$$

Applying Chebyshev's inequality, we get the following:

$$P(|\bar{X}_n - \mu| > 2\sigma) \leq \frac{\mathrm{Var}\bar{X}_n}{(2\sigma)^2} = \frac{\frac{\sigma^2}{n}}{4\sigma^2} = \frac{1}{4n}$$

Therefore, if we choose $n = 25$, we get the desired inequality.

Example 2. Practicing Inequalities.
Fill each inequality below with either $=, \leq, \geq$ or ?. and explain. In all instances below, assume that X and Y are positive random variables, although not necessarily independent. Assume that the expected values exist.

(a) $E(X^4)$ ______ $\sqrt{E(X^2)E(X^6)}$

(b) $P(|X+Y| > 2)$ ______ $\frac{1}{16}E((X+Y)^4)$

(c) $\sqrt{E(X)+50}$ ______ $E(\sqrt{X+50})$

(d) $E(Y|10X)$ ______ $E(Y|X)$

(e) $E(\cos(X))$ ______ $\cos(E(X))$

(f) $SD(X) + SD(Y)$ ______ $SD(X+Y)$

Solution

(a) $E(X^4) \boxed{\leq} \sqrt{E(X^2)E(X^6)}$
Note that by Cauchy-Schwarz inequality, we have that

$$E(X^4)^2 = E(XX^3)^2 \leq E(X^2)E(X^6)$$

(b) $P(|X+Y| > 2) \boxed{\leq} \frac{1}{16}E((X+Y)^4)$
By the Markov Inequality, we have that

$$P(|X+Y| > 2) = P((X+Y)^4 > 2^4) \leq \frac{1}{2^4}E((X+Y)^4) = \frac{1}{16}E((X+Y)^4)$$

(c) $\sqrt{E(X)+50} \boxed{\geq} E(\sqrt{X+50})$
By Jensen's inequality, because the square root function is a concave function, it follows that first applying the square root function and then taking an expectation yields a small value than first taking an expectation and then taking the square root.

(d) $E(Y|10X) \boxed{=} E(Y|X)$
Knowing 10X and X is being given the exact same information, so both should definitely be equal! As a note, try not to confuse yourself and believe that being given 10X increases the value of X at all. You would literally be given the quantity 10X, and from that you can obtain X.

(e) $E(\cos(X)) \boxed{?} \cos(E(X))$
Because the cosine function is neither concave nor convex, there is no definitive answer.

(f) $SD(X) + SD(Y) \boxed{\geq} SD(X+Y)$
If we square both sides, note that we have

$$\text{Var}(X) + \text{Var}(Y) + 2\sqrt{\text{Var}(X)\text{Var}(Y)} \_\_\_ \text{Var}(X+Y)$$

But note that $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$, and so we have

$$\text{Var}(X) + \text{Var}(Y) + 2\sqrt{\text{Var}(X)\text{Var}(Y)} \_\_\_ \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$$

And in this case, note that $\sqrt{\text{Var}(X)\text{Var}(Y)} \geq \text{Cov}(X,Y)$ since correlation is always less than or equal to 1, and so it follows that our answer is $\geq$.

Example 3. Gamma CLT.

(a) Explain why a Gamma random variable with parameters (n, λ) is approximately Normal when n is large.

(b) Let $X_n \sim \text{Gamma}(n, \lambda)$. Determine a and b such that

$$\frac{X_n - a}{b} \to N(0,1)$$

as $n \to \infty$

Solution

(a) Let $X_n = Y_1 + Y_2 + \ldots + Y_n$. Then, we can have $X_n \sim \text{Gamma}(n, \lambda)$ and Y_i's be i.i.d. $\text{Expo}(\lambda)$. By the central limit theorem, since X_n is the sum of i.i.d. random variables, it converges to a normal distribution as $n \to \infty$.

(b) We have reasoned above that X_n is approximately normal, but what are its parameters? The central limit theorem also tells us that

$$X_n \sim N\left(\frac{n}{\lambda}, \frac{n}{\lambda^2}\right)$$

In order to convert this normal distribution to a standard normal ($N(0,1)$), all we need to do is subtract the mean and divide by the standard deviation. Thus:

$$\frac{X_n - \frac{n}{\lambda}}{\frac{\sqrt{n}}{\lambda}} \sim N(0,1)$$

as $n \to \infty$.

Example 4. Two-State Markov Chain.
Suppose X_n is a two-state Markov chain with transition matrix

$$Q = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \end{array}$$

(a) Find the stationary distribution $\vec{s} = (s_0, s_1)$ of X_n by solving $\vec{s}Q = \vec{s}$.

(b) Show that this Markov Chain is reversible under the stationary distribution found in part (a)

(c) Let $Z_n = (X_{n-1}, X_n)$. Is Z_n a Markov chain? If so, what are the states and transition matrix?

Solution

(a) By solving $\vec{s}Q = \vec{s}$, we have that

$$s_0 = s_0(1-\alpha) + s_1\beta \text{ and } s_1 = s_0\alpha + s_1(1-\beta)$$

And by solving this system of linear equations, it follows that $\vec{s} = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$.

(b) To verify the validity of a stationary distribution for a chain, we just need to show that $s_i q_{ij} = s_j q_{ji}$, which is done if we can show that $s_0 q_{01} = s_1 q_{10}$. We have that

$$s_0 q_{01} = \frac{\alpha\beta}{\alpha+\beta} = s_1 q_{10}$$

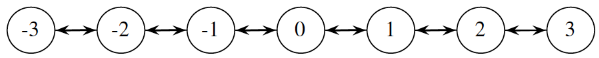
which satisfies our reversibility condition and verifies our stationary distribution from part (a).

(c) Yes, Z_n is a Markov Chain because conditional on Z_n, $Z_{n+1} \perp\!\!\!\perp Z_{n-1}$. This is because the components of Z_{n+1} and $Zn-1$ are either constants conditioned in Z_n or independent of each other given that X_n is a Markov Chain.

The states are given as $\{(0,0),(0,1),(1,0),(1,1)\}$. The transition matrix is given as

$$Q = \begin{array}{c c} & \begin{array}{cccc} (0,0) & (0,1) & (1,0) & (1,1) \end{array} \\ \begin{array}{c} (0,0) \\ (0,1) \\ (1,0) \\ (1,1) \end{array} & \begin{pmatrix} 1-\alpha & \alpha & 0 & 0 \\ 0 & 0 & \beta & 1-\beta \\ 1-\alpha & \alpha & 0 & 0 \\ 0 & 0 & \beta & 1-\beta \end{pmatrix} \end{array}$$

Example 5. Symmetrical Chain.
A Markov chain $X_0, X_1, X_2 \ldots$ with state space $\{3, 2, 1, 0, 1, 2, 3\}$ proceeds as follows. The chain starts at $X_0 = 0$. If X_n is not an endpoint (-3 or 3), then X_{n+1} is $X_n + 1$ or $X_n - 1$, each with probability $1/2$. Otherwise, the chain gets reflected off the endpoint, i.e., from 3 it always goes to 2 and from -3 it always goes to -2. A diagram of the chain is shown below.



(a) Is $|X_0|, |X_1|, |X_2|, \ldots$ a Markov Chain?

(b) Define the sign function $S(x)$ as follows:

$$S(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

Is $S(X_0), S(X_1), S(X_2), \ldots$ a Markov Chain?

(c) Find the stationary distribution of the original chain: $X_1, X_2, X_3 \ldots$

(d) Find a simple way to modify some of the transition probabilities q_{ij} for $i, j \in \{-3, 3\}$ to make the stationary distribution of the modified chain uniform over the states.

Solution

(a) Yes it is a Markov chain with state space $\{0, 1, 2, 3\}$. We need to check that this new chain satisfies the Markov property. If we are given that $|X_n| = 1$, then regardless of what $|X_{n-1}|, |X_{n-2}|, \ldots$ are, we know that there is a $1/2$ chance $|X_{n+1}|$ will be 0 and a $1/2$ chance that it will be 2. A similar argument can be made for all the states, and thus this chain satisfies the Markov property.

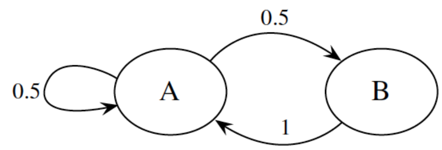
(b) We will show that this is not a Markov chain because it does not satisfy the Markov property. In general, when showing that something does or does not satisfy the Markov property, it is a good idea to ask yourself: "If I knew more than just the *one* previous state, would that help me make a more informed decision about what the next state is?".

In this case, if I told you that the last three states $S(X_{n-2}), S(X_{n-1}), S(X_n)$ were $0, 1, 1$, then we know that the chain is at 2, i.e. $X_n = 2$. This means that $P(S(X_{n+1}) = 0) = 0$ because we can't jump from 2 to 0. However, if I told you that the last two states $S(X_{n-1}), S(X_n)$ were 0 and 1, then you would know that $X_n = 1$ and say that $P(S(X_{n+1}) = 0) = 1/2$. Thus, the Markov property is not satisfied because even though in both situations $S(X_n) = 1$, we had different probabilities for $P(S(X_{n+1}) = 1)$.

(c) We see that this is simply a random walk on an undirected graph, so the stationary distribution is proportional to the degrees of each of the states. The degrees are: $(1, 2, 2, 2, 2, 2, 1)$, and so the stationary distribution is: $\vec{s} = (1/12, 1/6, 1/6, 1/6, 1/6, 1/6, 1/12)$

(d) If we could somehow get all the states to have the same degree, then the stationary distribution would be uniform across all states. Therefore, all we need to do is connect states 3 and -3 together, so that there is a $1/2$ chance at 3 to go to 2 or -3, and equivalently for -3.

Example 6. Not a Markov Chain.
A Markov chain has two states, A and B, with transitions as follows:



Suppose we do not get to observe this Markov chain, which we'll call $X_1, X_2, \ldots$. Instead, whenever the chain transitions from A back to A, we observe a 0, and whenever it changes states, we observe a 1. Let the sequence of 0's and 1's be called $Y_0, Y_1, Y_2, \ldots$ For example, if the X chain starts out as

$$A, A, B, A, B, A, A, \ldots$$

then the Y chain starts out as

$$0, 1, 1, 1, 1, 0, \ldots$$

(a) Show that $Y_0, Y_1, Y_2, \ldots$ is not a Markov Chain.

(b) In the past, when we have encountered processes that are not directly Markov chains, our remedy was to create a new state Z_n which would represent not only the current state X_n that the chain is on, but also remember the past m states. Thus, Z_n is not just one state, but a tuple of states that represent the chain's history:

$$Z_n = \left(X_n, X_{n-1}, X_{n-2} \ldots X_{n-(m-1)}\right)$$

Show that such a trick will not work for $Y_0, Y_1, Y_2 \ldots$. That is, no matter how large m is, $\{Z_n\}$ will never be a Markov Chain.

Solution

(a) We will show that knowing the chain outputted a 1 is not enough information to give the probability of the chain outputting 1 again on the next iteration. This is true because if the most recent sequence of outputs have been $0, 1$, then the chain is clearly at B in which case the probability of it outputting 1 next is 1. However, if the chain's outputs had been $0, 1, 1$, then we know that the chain is at state A in which case the probability of the next output being 1 (and hence the next state be B) is only $1/2$.

(b) The problem is saying that no finite amount of "memory" in this Markov chain is enough to be able to give a unique set of probabilities for the next output being 0 or 1, given a history of m outputs.

We notice that 1's in this Markov chain output always come in pairs. Assuming that you start at state A (which is valid because state A is recurrent), we see that outputting a 1 means that another 1 must follow.

Suppose again that we're in the situation where out last output has been 1. From the above fact that 1's always come in pairs, we can reason that it's not the number of 1's that have been outputted in the recent past that matters, but rather its parity (i.e even or odd). If a 0 is followed by an even number of 1's, then we know that the chain is at state A so the probability of outputting either a 1 or 0 is $1/2$. If the number of 1's following that first 0 is odd, then we know that the next output must be 1 with probability 1.

No amount of memory m can completely capture the parity of the number or 1's that came before. This doesn't mean that there doesn't exist some clever way to create states such that this sequence of 0's and 1's becomes a Markov chain. Rather, this conclusions says that this one particular technique (remembering the past m states) does not work in this case.