

Homework 3

I agree to abide by the Stern Code of Conduct (name & NYU ID) _____

Question 1

Suppose X_1, \dots, X_n are independent and identically distributed (i.i.d.) with the same distribution as X . Let $S_n = \sum_{i=1}^n X_i$, so $\bar{X}_n = \frac{1}{n}S_n$, this shorthand will be useful for parts (b) through (c).

a.

What do we need to know or assume about $E[X_1]$ to conclude that $\bar{X}_n \rightarrow \mu$ as $n \rightarrow \infty$?

Solution: Since $\bar{X}_n \rightarrow E[X]$ as $n \rightarrow \infty$, we need to assume $\mu = E[X]$.

b.

Suppose X can only be -1 or 1. What are the lowest and highest possible values of S_n ?

Solution: If every $X_i = -1$ then $S_n = -n$ is the lowest possible value. Similarly, $S_n = n$ is the highest possible value of S_n .

c.

Suppose $n = 99$, $\bar{X}_{99} = 0.01$, we are going to add one new observation X_{100} to the sample, and we know $-1 \leq X_{100} \leq 1$. What are the lowest and highest possible values for \bar{X}_{100} ?

Solution: If we compare the mean of the first n observations with the mean of the first $n + 1$, we see

$$\bar{X}_{n+1} = \frac{n}{n+1}\bar{X}_n + \frac{1}{n+1}X_{n+1}$$

Plugging in $n = 99$ and $\bar{X}_n = 0.01$, we have

$$\bar{X}_{n+1} = \frac{99}{100}0.01 + \frac{1}{100}X_{100} = 0.0099 + \frac{1}{100}X_{100}$$

Since $-1 \leq X_{100} \leq 1$, this number is between 0.0099 ± 0.01 or approximately 0 and 0.02.

Question 2

a.

At a casino there is a game where the player bets some amount of money X and then spins a wheel. If the wheel says they win, they get X dollars. Otherwise they lose and pay X dollars to the casino. I have the perfect gambling strategy to guarantee that I win this game. On the first round I bet X_1 . If I win, the payoff is X_1 dollars. If I lose, I go "double or nothing" and bet $X_2 = 2X_1$ on the second game. I repeat this strategy, doubling my bet every time until I win. If I win on the first round, my payoff will be X_1 . If I win on the n th round, my payoff will be $2^n X_1$. Aside from the practical fact that I may run out of money before I win, this sequence of payoffs/losses doesn't satisfy the assumptions of the law of large numbers. Which assumption(s) does it violate?

Solution: The law of large numbers applies to sequences of independent and identically distributed random variables. The sequence of payoffs/losses in this game are not identically distributed, since they double in size each time.

b.

Your roommate suggests rolling a standard 6-sided die to decide who does the cleaning chores each day. If the die comes up 3 or less, you win, and your roommate does the chores. If it's higher than 3, you do the chores. Let D represent the dice, so $D \in \{1, \dots, 6\}$, and let $X = 1$ if $D > 3$ and 0 otherwise. Let D_i, X_i represent the result on day i , so X_1, \dots, X_n are i.i.d. with the same distribution as X (and similarly for the D 's). What is $E[X]$? The proportion of times that you do the chores in n days is \bar{X}_n . What will that proportion be in the long run, as $n \rightarrow \infty$? Does this system seem fair to you?

Solution: $E[X] = P(D > 3) = 3/6 = 1/2$. In the long run $\bar{X}_n \rightarrow 1/2$ so I do the chores half of the time. This seems fair.

c.

Next, your roommate suggests a small change in the rules: keep track of the average of the dice $\bar{D}_n = \frac{1}{n} \sum_{i=1}^n D_i$, and then you only need to do the chores whenever this average is higher than 3. Based on this system, in the long run what proportion of the time will you be doing the chores?

Solution: Since $E[D] = 3.5$, we know $\bar{D}_n \rightarrow 3.5$ as $n \rightarrow \infty$. This means that eventually I will have to do the chores every day, so in the long run I do all the chores. That seems unfair.

d.

Continuing part (b), suppose you have had a streak of bad luck and had to do the chores every day for a week. Your roommate reassures you, "Don't worry, if you do the chores a bunch of times then it becomes more likely for me to do them a bunch of times, since it has to even out in the long run." Do you agree? Why or why not?

Solution: This is the "Gambler's fallacy." The dice rolls are *independent*, so it does not become any more or less likely for my roommate to do the chores based on previous days.

Question 3

Let X_1, X_2, \dots, X_n be i.i.d. with mean μ and variance σ^2 , so $\text{Var}(\bar{X}) = \sigma^2/n$.

Applying Chebyshev's inequality to the mean \bar{X} tells us that

$$P(|\bar{X} - \mu| \geq \sigma) \leq \frac{1}{n}$$

If we instead apply the central limit theorem, we would find

$$P(|\bar{X} - \mu| \geq \sigma) \approx 2\Phi(-\sqrt{n})$$

where Φ is the CDF of the standard normal distribution $N(0, 1)$. We can calculate this in R with the code `2*pnorm(-sqrt(n))`.

Part a.

Try both methods for $n = 5$ and for $n = 10$ and write the numbers you get below.

Solution:

Chebyshev:

```
c(1/5, 1/10)
```

```
## [1] 0.2 0.1
```

CLT: (easy solution)

```
c(2*pnorm(-sqrt(5)), 2*pnorm(-sqrt(10)))
```

```
## [1] 0.025347319 0.001565402
```

CLT: (fancy solution)

```
CLT <- function(n) 2*pnorm(-sqrt(n))  
CLT(c(5, 10))
```

```
## [1] 0.025347319 0.001565402
```

Part b.

Which method gives you a stronger conclusion about \bar{X} , Chebyshev or the CLT?

Solution: The CLT method says the probability that \bar{X} is close to the true mean is higher than the answer from the Chebyshev inequality. The Chebyshev inequality gives us less certainty, so the CLT is the one with a stronger conclusion.