1 Inequality Practice

(a) X is a random variable such that $X > -5$ and $\mathbb{E}[X] = -3$. Find an upper bound for the probability of X being greater than or equal to -1.

(b) You roll a die 100 times. Let Y be the sum of the numbers that appear on the die throughout the 100 rolls. Use Chebyshev's inequality to bound the probability of the sum Y being greater than 400 or less than 300.

Solution:

(a) We want to use Markov's Inequality, but we remember that Markov's only works with non-negative random variables. Then, we define a new random variable $Y = X + 5$, where Y is always non-negative, so we can use Markov's on Y. By linearity of expectation, $\mathbb{E}[Y] = -3 + 5 = 2$. So, $\mathbb{P}[Y \geq 4] \leq 2/4 = 1/2$.

(b) Let Y_i be the number on the die for the ith roll, for $i = 1, \ldots, 100$. Then, $Y = \sum_{i=1}^{100} Y_i$. By linearity of expectation, $\mathbb{E}[Y] = \sum_{i=1}^{100} \mathbb{E}[Y_i]$.

$$\mathbb{E}[Y_i] = \sum_{j=1}^{6} j \cdot \mathbb{P}[Y_i = j] = \sum_{j=1}^{6} j \cdot \frac{1}{6} = \frac{1}{6} \cdot \sum_{j=1}^{6} j = \frac{1}{6} \cdot 21 = \frac{7}{2}$$

Then, we have $\mathbb{E}[Y] = 100 \cdot (7/2) = 350$.

$$\mathbb{E}[Y_i^2] = \sum_{j=1}^{6} j^2 \cdot \mathbb{P}[Y_i = j] = \sum_{j=1}^{6} j^2 \cdot \frac{1}{6} = \frac{1}{6} \cdot \sum_{j=1}^{6} j^2 = \frac{1}{6} \cdot 91 = \frac{91}{6}$$

Then, we have

$$\operatorname{var}(Y_i) = \mathbb{E}[Y_i^2] - \mathbb{E}[Y_i]^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12},$$

Since the Y_is are independent, and therefore uncorrelated, we can add the $\operatorname{var}(Y_i)$s to get $\operatorname{var}(Y) = 100(35/12)$.

Putting it all together, we use Chebyshev's to get

$$\mathbb{P}[|X - 350| \geq 50] \leq \frac{100(35/12)}{50^2} = \frac{7}{60}.$$

2 Tightness of Inequalities

(a) Show by example that Markov's inequality is tight; that is, show that given $k > 0$, there exists a discrete non-negative random variable X such that $\mathbb{P}(X \geq k) = \mathbb{E}[X]/k$.

(b) Show by example that Chebyshev's inequality is tight; that is, show that given $k \geq 1$, there exists a random variable X such that $\mathbb{P}(|X - \mathbb{E}[X]| \geq k\sigma) = 1/k^2$, where $\sigma^2 = \operatorname{var} X$.

(c) Show that there is no non-negative discrete random variable $X \neq 0$, that takes values in some finite set $\{v_1, \ldots, v_N\}$, such that for all $k > 0$, Markov's inequality is tight; that is, $\mathbb{P}(X \geq k) = \mathbb{E}[X]/k$.

Solution:

(a) In the proof of Markov's Inequality ($\mathbb{P}[X \geq \alpha] \leq \frac{\mathbb{E}[X]}{\alpha}$), the first time we lose equality is at this step:

$$\mathbb{E}[X] = \sum_a (a \cdot \mathbb{P}[X = a]) \geq \sum_{a \geq \alpha} (a \cdot \mathbb{P}[X = a])$$

We get an inequality because we drop all $a \cdot \mathbb{P}[X = a]$ terms where $a < \alpha$. Thus, we can only maintain equality if all of these dropped terms were actually 0. This would mean either $a = 0$ or $\mathbb{P}[X = a] = 0$ for an $a > 0$, which means X can put probability on 0, but should put no probability on any other value $< \alpha$.

The next time we lose equality in the proof is the step following the one above:

$$\sum_{a \geq \alpha} (a \cdot \mathbb{P}[X = a]) \geq \alpha \cdot \sum_{a \geq \alpha} \mathbb{P}[X = a]$$

We get an inequality because we treat all $a \geq \alpha$ in the summation as just α, so we can pull out the α term. The only way for us to maintain equality is if we never have to substitute α for some larger a. This tells us that X should not put probability on any value $> \alpha$.

Both of these facts drive the intuition behind our example: that X can only take values 0 and α.

Let X be the random variable which is 0 with probability $1 - p$ and k with probability p, where $k > 0$. Then, $\mathbb{E}[X] = kp$, and Markov's inequality says

$$\mathbb{P}(X \geq k) \leq \frac{\mathbb{E}[X]}{k} = \frac{kp}{k} = p,$$

which is tight.

(b) The proof of Chebyshev's Inequality ($\mathbb{P}[|X - \mathbb{E}[X]| \geq \alpha] \leq \frac{\operatorname{var}(X)}{\alpha^2}$) comes from an application of Markov's Inequality to the variable $Y = (X - \mathbb{E}[X])^2$ being $\geq \alpha^2$. The only ways we can lose equality in the proof of Chebyshev's is if we lose equality in the application of Markov! Therefore, we need the variable Y to satisfy the conditions from Part (a) that ensure the application of Markov will be tight. To recap, we would need Y to only take values 0 and α^2. Thus, $(X - \mathbb{E}[X])$ can take on the values $\{-\alpha, 0, \alpha\}$.

Let

$$X = \begin{cases} -a & \text{with probability } k^{-2}/2 \\ a & \text{with probability } k^{-2}/2 \\ 0 & \text{with probability } 1-k^{-2} \end{cases}$$

for $a > 0$. Note that $\operatorname{var} X = a^2k^{-2}$, so $k\sigma = a$, so Chebyshev's inequality gives

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq k\sigma) = \mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{1}{k^2},$$

which is tight.

(c) In Part (a), we already built some intuition for when Markov is tight. For a formal proof, suppose $X \neq 0$ is a random variable which makes Markov's inequality tight for all $k > 0$.

First, we consider the case when $\mathbb{E}[X] = 0$. If we let v be the smallest positive element of $\{v_1, \dots, v_N\}$, then $\mathbb{P}(X \geq v) = 0$, which means that X takes on no positive values; since $\mathbb{E}[X] = 0$, this implies that $\mathbb{P}(X = 0) = 1$, but that was ruled out by our assumption that $X \neq 0$.

Next, we consider when $\mathbb{E}[X] \neq 0$, that is, $\mathbb{E}[X] > 0$. By choosing $0 < k < \mathbb{E}[X]$, we find that $\mathbb{P}(X \geq k) = \mathbb{E}[X]/k$, but the RHS is > 1, which is impossible.

3 Working with the Law of Large Numbers

(a) A fair coin is tossed and you win a prize if there are more than 60% heads. Which is better: 10 tosses or 100 tosses? Explain.

(b) A fair coin is tossed and you win a prize if there are more than 40% heads. Which is better: 10 tosses or 100 tosses? Explain.

(c) A coin is tossed and you win a prize if there are between 40% and 60% heads. Which is better: 10 tosses or 100 tosses? Explain.

(d) A coin is tossed and you win a prize if there are exactly 50% heads. Which is better: 10 tosses or 100 tosses? Explain.

Solution:

(a) 10 tosses. By LLN, the sample mean should have higher probability to be close to the population mean as n increases. Therefore the average proportion of coins that are heads should be closer to 0.50, and has a lower chance of being greater than 0.60 if there are 100 tosses (compared with 10 tosses).

(b) 100 tosses. Again, by LLN, the sample mean should have higher probability to be close to the population mean as n increases. Therefore the average proportion of coins that are heads should be closer to 0.50, and has a lower chance of being smaller than 0.40 if there are 100 tosses. A lower chance of being smaller than 0.40 is the desired result.

(c) 100 tosses. Again, by LLN, the average proportion of coins that are heads should be closer to 0.50, and has a lower chance of being both smaller than 0.40 if there are 100 tosses. Similarly, there is a lower chance of being larger than 0.60 if there are 100 tosses. Lower chances of both of these events is desired if we want the fraction of heads to be between 0.4 and 0.6.

(d) 10 tosses. Compare the probability of getting equal number of heads and tails between $2n$ and $2n+2$ tosses.

$$\begin{aligned}
\mathbb{P}[n \text{ heads in } 2n \text{ tosses}] &= \binom{2n}{n}\frac{1}{2^{2n}} \\
\mathbb{P}[n+1 \text{ heads in } 2n+2 \text{ tosses}] &= \binom{2n+2}{n+1}\frac{1}{2^{2n+2}} = \frac{(2n+2)!}{(n+1)!(n+1)!}\cdot\frac{1}{2^{2n+2}} \\
&= \frac{(2n+2)(2n+1)2n!}{(n+1)(n+1)n!n!}\cdot\frac{1}{2^{2n+2}} \\
&= \frac{2n+2}{n+1}\cdot\frac{2n+1}{n+1}\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} < \left(\frac{2n+2}{n+1}\right)^2\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} \\
&= 4\binom{2n}{n}\cdot\frac{1}{2^{2n+2}} = \binom{2n}{n}\frac{1}{2^{2n}} = \mathbb{P}[n \text{ heads in } 2n \text{ tosses}]
\end{aligned}$$

As we increment n, the probability will always decrease. Therefore, the larger n is, the less probability we'll get exactly 50% heads. $\square$

Note: By Stirling's approximation, $\binom{2n}{n}2^{-2n}$ is roughly $(\pi n)^{-1/2}$ for large n.