Concentration Inequalities Intro

Markov's Inequality: For any nonnegative random variable X and $t > 0$,

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

Chebyshev's Inequality: For any random variable X and $c > 0$,

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq c] \leq \frac{\text{Var}(X)}{c^2}.$$

Law of Large Numbers: Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2. We have the following:

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu$$

$$\text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n}.$$

Applying Chebyshev's inequality on the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$, we have that

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right] \leq \frac{\sigma^2}{n\varepsilon^2}$$

which means that as $n \to \infty$, the probability of the sample mean deviating from the true mean by any $\varepsilon > 0$ approaches zero.

1 Probabilistic Bounds

Note 17 A random variable X has variance $\text{Var}(X) = 9$ and expectation $\mathbb{E}[X] = 2$. Furthermore, the value of X is never greater than 10. Given this information, provide either a proof or a counterexample for the following statements.

(a) $\mathbb{E}[X^2] = 13$.

(b) $\mathbb{P}[X = 2] > 0$.

(c) $\mathbb{P}[X \geq 2] = \mathbb{P}[X \leq 2]$.

(d) $\mathbb{P}[X \leq 1] \leq 8/9$.

(e) $\mathbb{P}[X \geq 6] \leq 9/16$.

Solution:

(a) TRUE. Since $9 = \text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - 2^2$, we have $\mathbb{E}[X^2] = 9 + 4 = 13$.

(b) FALSE. It is not necessary for a random variable to be able to take on its mean as a value. As one possible counterexample, construct a random variable X that satisfies the conditions in the question but does not take on the value 2.

A simple example would be a random variable that takes on 2 values, where $\mathbb{P}[X = a] = \mathbb{P}[X = b] = 1/2$, and $a \neq b$ with both $a, b \leq 10$. The expectation must be 2, so we have $a/2 + b/2 = 2$. The variance is 9, so $\mathbb{E}[X^2] = 13$ (from Part ((a))) and $a^2/2 + b^2/2 = 13$. Solving for a and b, we get $\mathbb{P}[X = -1] = \mathbb{P}[X = 5] = 1/2$ as a counterexample.

(c) FALSE. The median of a random variable is not necessarily the mean, unless it is symmetric. As one possible counterexample, construct a random variable X that satisfies the conditions in the question but does not have an equal chance of being less than or greater than 2.

A simple example would be a random variable that takes on 2 values, where $\mathbb{P}[X = a] = p, \mathbb{P}[X = b] = 1 - p$. Here, we use the same approach as part (b) except with a generic p, since we want $p \neq 1/2$. The expectation must be 2, so we have $pa + (1 - p)b = 2$. The variance is 9, so $\mathbb{E}\left[X^2\right] = 13$ and $pa^2 + (1 - p)b^2 = 13$. Solving for a and b, we find the relation $b = 2 \pm 3/\sqrt{x}$, where $x = (1 - p)/p$. Then, we can find an example by plugging in values for x so that $a, b \leq 10$ and $p \neq 1/2$. One such counterexample is $\mathbb{P}[X = -7] = 1/10, \mathbb{P}[X = 3] = 9/10$.

(d) TRUE. Let $Y = 10 - X$. Since X is never exceeds 10, Y is a non-negative random variable. By Markov's inequality,

$$\mathbb{P}[10 - X \geq a] = \mathbb{P}[Y \geq a] \leq \frac{\mathbb{E}[Y]}{a} = \frac{\mathbb{E}[10 - X]}{a} = \frac{8}{a}.$$

Setting $a = 9$, we get $\mathbb{P}[X \leq 1] = \mathbb{P}[10 - X \geq 9] \leq 8/9$.

As a side note, if we were to try Chebyshev's inequality instead, noting that

$$\mathbb{P}[X \leq 1] + \mathbb{P}[X \geq 3] = \mathbb{P}[|X - 2| \geq 1] = \mathbb{P}[|X - \mathbb{E}[X]| \geq 1],$$

we'd get

$$\mathbb{P}[X \le 1] \le \mathbb{P}[X \le 1] + \mathbb{P}[X \ge 3] = \mathbb{P}[|X-2| \ge 1] \le \frac{(}{\text{Var}}(X))1 = 9,$$

which is an unhelpful bound.

(e) TRUE. Chebyshev's inequality says $\mathbb{P}[|X - \mathbb{E}[X]| \ge a] \le \text{Var}(X)/a^2$. If we set $a = 4$, we have

$$\mathbb{P}[|X-2| \ge 4] \le \frac{9}{16}.$$

Now we observe that $\mathbb{P}[X \ge 6] \le \mathbb{P}[|X-2| \ge 4]$, because the event $X \ge 6$ is a subset of the event $|X-2| \ge 4$.

As a side note, we can't apply Markov's inequality here; as-is, X is not nonnegative, and if we did the same transformation $Y = 10 - X$ from before, we'd want an upper bound on $\mathbb{P}[10 - X \le 10 - 6] = \mathbb{P}[Y \le 4]$, which we cannot do with Markov's inequality; it only gives an upper bound on probabilities of the form $\mathbb{P}[Y \ge a]$.

2 Vegas

Note 17

On the planet Vegas, everyone carries a coin. Many people are honest and carry a fair coin (heads on one side and tails on the other), but a fraction p of them cheat and carry a trick coin with heads on both sides. You want to estimate p with the following experiment: you pick a random sample of n people and ask each one to flip their coin. Assume that each person is independently likely to carry a fair or a trick coin.

(a) Let X be the proportion of coin flips which are heads. Find $\mathbb{E}[X]$.

(b) Given the results of your experiment, how should you estimate p? (*Hint:* Construct an unbiased estimator for p using part (a). Recall that $\hat{p}$ is an unbiased estimator if $\mathbb{E}[\hat{p}] = p$.)

(c) How many people do you need to ask to be 95% sure that your answer is off by at most 0.05?

Solution:

(a) Let X_i be the indicator that the ith person's coin flips heads. Then $X = \frac{1}{n}\sum_{i=1}^{n} X_i$. Applying linearity, we have

$$\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \mathbb{E}[X_i].$$

By total probability,

$$\mathbb{E}[X_i] = p \cdot 1 + (1-p) \cdot \frac{1}{2} = \frac{1}{2}(p+1).$$

(b) We want to construct an estimate $\hat{p}$ such that $\mathbb{E}[\hat{p}] = p$. Then, if we have a large enough sample, we'd expect to get a good estimate of p. In other words, we measure X, the fraction of people whose coin flips heads. How can we use this observation to construct $\hat{p}$? From part (a), $\mathbb{E}[X] = \frac{1}{2}(p+1)$. By applying (reverse) linearity to isolate p, we find that

$$p = 2\,\mathbb{E}[X] - 1 = \mathbb{E}[2X - 1].$$

Thus, our estimator $\hat{p}$ should be $2X-1$.

(c) We want to find n such that $\mathbb{P}[|\hat{p}-p| \leq 0.05] > 0.95$. Another way to state this is that we want

$$P[|\hat{p}-p| > 0.05] \leq 0.05.$$

Notice that $\mathbb{E}[\hat{p}] = p$ by construction, so we can immediately apply Chebyshev's inequality on $\hat{p}$. What we get is:

$$\mathbb{P}[|\hat{p}-p| > 0.05] \leq \mathbb{P}[\hat{p}-p| \geq 0.05] \leq \frac{\text{Var}[\hat{p}]}{0.05^2}$$

If $\frac{\text{Var}(\hat{p})}{0.05^2} \leq 0.05$, then we have $\mathbb{P}[|\hat{p}-p| > 0.05] \leq 0.05$ as desired. So, we want n such that $\text{Var}(\hat{p}) \leq 0.05^3$.

$$\text{Var}(\hat{p}) = \text{Var}(2X-1) = 4\,\text{Var}(X) = \frac{4}{n^2}\,\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{4}{n}\,\text{Var}(X_1).$$

But X_i is an indicator (Bernoulli variable), so its variance is bounded by $\frac{1}{4}$ (note that $p(1-p)$ is maximized at $p = \frac{1}{2}$ to yield a value of $\frac{1}{4}$). Therefore we have

$$\text{Var}[\hat{p}] \leq \frac{4}{n}\frac{1}{4} = \frac{1}{n}.$$

So, we choose n such that $\frac{1}{n} \leq 0.05^3$, giving $n \geq \frac{1}{0.05^3} = 8000$.

3 Working with the Law of Large Numbers

Note 17

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

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

(c) A fair coin is tossed multiple times and you win a prize if there are between 40% and 60% heads. Which number of tosses would you prefer: 10 tosses or 100 tosses? Explain.

(d) A fair coin is tossed multiple times and you win a prize if there are exactly 50% heads. Which number of tosses would you prefer: 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. Intuitively, the more tosses we have, the harder it gets for *exactly* half of the tosses to be heads; more tosses gives more of a restriction. In extremes, compare the probability of getting exactly 1 head out of 2 tosses (this is 0.5), and the probability of getting *exactly* 500,000 heads out of a million tosses; the latter is much much smaller than 0.5, because we're targeting such a specific number.

More rigorously, we can 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. □

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

See `https://github.com/dingyiming0427/CS70-demo/` for a code demo.