# 1 Just One Tail, Please

Let $X$ be some random variable with finite mean and variance which is not necessarily non-negative. The *extended* version of Markov's Inequality states that for a non-negative function $\phi(x)$ which is monotonically increasing for $x > 0$ and some constant $\alpha > 0$,

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[\phi(X)]}{\phi(\alpha)}$$

Suppose $\mathbb{E}[X] = 0$, $\mathrm{Var}(X) = \sigma^2 < \infty$, and $\alpha > 0$.

(a) Use the extended version of Markov's Inequality stated above with $\phi(x) = (x+c)^2$, where $c$ is some positive constant, to show that:

$$\mathbb{P}(X \geq \alpha) \leq \frac{\sigma^2 + c^2}{(\alpha+c)^2}$$

(b) Note that the above bound applies for all positive $c$, so we can choose a value of $c$ to minimize the expression, yielding the best possible bound. Find the value for $c$ which will minimize the RHS expression (you may assume that the expression has a unique minimum).

We can plug in the minimizing value of $c$ you found in part (b) to prove the following bound:

$$\mathbb{P}(X \geq \alpha) \leq \frac{\sigma^2}{\alpha^2 + \sigma^2}.$$

This bound is also known as Cantelli's inequality.

(c) Recall that Chebyshev's inequality provides a two-sided bound. That is, it provides a bound on $\mathbb{P}(|X - \mathbb{E}[X]| \geq \alpha) = \mathbb{P}(X \geq \mathbb{E}[X] + \alpha) + \mathbb{P}(X \leq \mathbb{E}[X] - \alpha)$. If we only wanted to bound the probability of one of the tails, e.g. if we wanted to bound $\mathbb{P}(X \geq \mathbb{E}[X] + \alpha)$, it is tempting to just divide the bound we get from Chebyshev's by two.

(i) Why is this not always correct in general?

(ii) Provide an example of a random variable $X$ (does not have to be zero-mean) and a constant $\alpha$ such that using this method (dividing by two to bound one tail) is not correct, that is, $\mathbb{P}(X \geq \mathbb{E}[X] + \alpha) > \frac{\mathrm{Var}(X)}{2\alpha^2}$ or $\mathbb{P}(X \leq \mathbb{E}[X] - \alpha) > \frac{\mathrm{Var}(X)}{2\alpha^2}$.

Now we see the use of the bound proven in part (b) - it allows us to bound just one tail while still taking variance into account, and does not require us to assume any property of the random variable. Note that the bound is also always guaranteed to be less than 1 (and therefore at least somewhat useful), unlike Markov's and Chebyshev's inequality!

(d) Let's try out our new bound on a simple example. Suppose $X$ is a positively-valued random variable with $\mathbb{E}[X] = 3$ and $\mathrm{Var}(X) = 2$.

  (i) What bound would Markov's inequality give for $\mathbb{P}[X \geq 5]$?
  (ii) What bound would Chebyshev's inequality give for $\mathbb{P}[X \geq 5]$?
  (iii) What bound would Cantelli's Inequality give for $\mathbb{P}[X \geq 5]$? (*Note*: Recall that Cantelli's Inequality only applies for zero-mean random variables.)

**Solution:**

(a) Note that $\sigma^2 = \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2]$. Using the inequality presented in the problem, we have:

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[(X+c)^2]}{(\alpha+c)^2} = \frac{\mathbb{E}[X^2 + 2cX + c^2]}{(\alpha+c)^2} = \frac{\mathbb{E}[X^2] + 2c\,\mathbb{E}[X] + c^2}{(\alpha+c)^2} = \frac{\sigma^2 + c^2}{(\alpha+c)^2}$$

(b) We set the derivative with respect to $c$ of the above expression equal to 0, and solve for $c$.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}c}\frac{\sigma^2 + c^2}{(\alpha+c)^2} &= 0 \\
\frac{2c(\alpha+c)^2 - 2(\alpha+c)(\sigma^2+c^2)}{(\alpha+c)^4} &= 0 \\
2c(\alpha+c)^2 - 2(\alpha+c)(\sigma^2+c^2) &= 0 \\
\alpha c^2 + (\alpha^2 - \sigma^2)c - \sigma^2\alpha &= 0 \\
c &= \frac{\sigma^2}{\alpha}
\end{aligned}$$

To get the last step we use the quadratic equation and take the positive solution.

(c) It is possible for one of the tails to contain more probability than the other. One example of a random variable which demonstrates this is $X$, where $\mathbb{P}(X = 0) = 0.75$ and $\mathbb{P}(X = 10) = 0.25$, with $\alpha = 7$. Here, $\mathbb{E}[X] = 2.5$ and $\mathrm{Var}(X) = 100 \cdot 0.25 \cdot 0.75$, so we have:

$$\mathbb{P}(X \geq \mathbb{E}[X] + 7) = 0.25 > \frac{\mathrm{Var}(X)}{2 \cdot 7^2} \approx 0.19$$

(d) Using Markov's: $\mathbb{P}(X \geq 5) \leq \frac{\mathbb{E}[X]}{5} = \frac{3}{5}$

Using Chebyshev's: $\mathbb{P}(X \geq 5) \leq \mathbb{P}(|X - \mathbb{E}[X]| \geq 2) \leq \frac{\mathrm{Var}(X)}{2^2} = \frac{1}{2}$

Using bound shown above (Cantelli's):
Since we have the condition that this bound applies to zero-mean random variables, let us define $Y = X - \mathbb{E}[X] = X - 3$. Note that $\mathrm{Var}(Y) = \mathrm{Var}(X)$.

Then we get: $\mathbb{P}(X \geq 5) = \mathbb{P}(Y \geq 2) \leq \frac{\mathrm{Var}(Y)}{2^2 + \mathrm{Var}(Y)} = \frac{1}{3}$.

We see that Cantelli's inequality (the bound from part (b)) does better than Chebyshev's, which does better than Markov's (note that having a smaller upper bound is better)! This is a good demonstration on how we might derive better bounds using Markov's inequality, if we know further information about the random variable like its variance.

# 2 Tightness of Inequalities

(a) Show by example that Markov's inequality is tight; that is, show that given some fixed $k > 0$, 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 some fixed $k \geq 1$, there exists a random variable $X$ such that $\mathbb{P}(|X - \mathbb{E}[X]| \geq k\sigma) = 1/k^2$, where $\sigma^2 = \mathrm{Var}(X)$.

**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 $\alpha$, so we can pull out the $\alpha$ term. The only way for us to maintain equality is if we never have to substitute $\alpha$ 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 $\alpha$.

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 ($\mathbb{P}[|X-\mathbb{E}[X]| \geq \alpha] \leq \frac{\text{Var}(X)}{\alpha^2}$) 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 $\alpha^2$. 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 $\text{Var}(X) = a^2k^{-2}$, 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.

# 3 Probabilistically Buying Probability Books

Chuck will go shopping for probability books for $K$ hours. Here, $K$ is a random variable and is equally likely to be 1, 2, or 3. The number of books $N$ that he buys is random and depends on how long he shops. We are told that

$$\mathbb{P}[N=n|K=k] = \begin{cases} \frac{c}{k} & \text{for } n = 1,\ldots,k \\ 0 & \text{otherwise} \end{cases}$$

for some constant $c$.

(a) Compute $c$.

(b) Find the joint distribution of $K$ and $N$.

(c) Find the marginal distribution of $N$.

(d) Find the conditional distribution of $K$ given that $N = 1$.

(e) We are now told that he bought at least 1 but no more than 2 books. Find the conditional mean and variance of $K$, given this piece of information.

(f) The cost of each book is a random variable with mean 3. What is the expectation of his total expenditure? *Hint:* Condition on events $N = 1,\ldots,N = 3$ and use the total expectation theorem.

**Solution:**

(a) For any $k$, we know that probabilities conditioned on $K=k$ must sum to 1, i.e

$$\sum_n \mathbb{P}[N=n|K=k]=1 \quad ,$$

so it must be that

$$1=\sum_{n=1}^{k}\mathbb{P}[N=n|K=k]=k\times\frac{c}{k}=c \quad .$$

Thus, $c=1$.

(b) The joint distribution specifies $\mathbb{P}[N=n\cap K=k]$ for all $n$ and $k$. Note that

$$\mathbb{P}[N=n\cap K=k]=\mathbb{P}[N=n|K=k]\mathbb{P}[K=k]$$

and we know $\mathbb{P}[N=n|K=k]$ and $\mathbb{P}[K=k]$ (it says all $k\in\{1,2,3\}$ are equally likely). We use this formula to calculate $\mathbb{P}[N=n\cap K=k]$ for each $n,k$ and list the result in a table:

| $n\setminus k$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | $\frac{1}{3}$ | $\frac{1}{6}$ | $\frac{1}{9}$ |
| 2 | 0 | $\frac{1}{6}$ | $\frac{1}{9}$ |
| 3 | 0 | 0 | $\frac{1}{9}$ |

Alternatively, we can define the joint distribution as a formula with specified domain. $\mathbb{P}[N=n,K=k]=\mathbb{P}[N=n\mid K=k]\mathbb{P}[K=k]=\frac{1}{k}\frac{1}{3}$ whenever it is nonzero. So,

$$\mathbb{P}[N=n,K=k]=\begin{cases}\frac{1}{3k} & k\in\{1,2,3\},n\in\{1,\dots,k\}\\ 0 & \text{otherwise}\end{cases}$$

(c) The marginal distribution of $N$ is given by the value of $\mathbb{P}[N=n]$, for each possible value of $n$. By the total probability rule,

$$\mathbb{P}[N=n]=\mathbb{P}[N=n\cap K=1]+\mathbb{P}[N=n\cap K=2]+\mathbb{P}[N=n\cap K=3] \quad .$$

Thus, we get

$$\mathbb{P}[N=n]=\begin{cases}\frac{1}{3}+\frac{1}{6}+\frac{1}{9} & \text{if } n=1\\ \frac{1}{6}+\frac{1}{9} & \text{if } n=2\\ \frac{1}{9} & \text{if } n=3\end{cases}=\begin{cases}\frac{11}{18} & \text{if } n=1\\ \frac{5}{18} & \text{if } n=2\\ \frac{2}{18} & \text{if } n=3\end{cases}$$

(d) By definition, $\mathbb{P}[K=k|N=1]=\frac{\mathbb{P}[K=k\cap N=1]}{\mathbb{P}[N=1]}$. The numerator comes from the joint distribution of $N$ and $K$ (part (b)), and the denominator comes from the marginal distribution of $N$ (part (c)). Plugging in, we get

$$\mathbb{P}[K=k|N=1]=\begin{cases}\frac{\frac{1}{3}}{\frac{11}{18}} & \text{if } k=1\\ \frac{\frac{1}{6}}{\frac{11}{18}} & \text{if } k=2\\ \frac{\frac{1}{9}}{\frac{11}{18}} & \text{if } k=3\end{cases}=\begin{cases}\frac{6}{11} & \text{if } k=1\\ \frac{3}{11} & \text{if } k=2\\ \frac{2}{11} & \text{if } k=3\end{cases}$$

(e) We first compute the distribution $\mathbb{P}[K=k|N=1\cup N=2]$ as we did in part (d):

$$\mathbb{P}[K=k|N=1\cup N=2]=\begin{cases}\frac{\frac{1}{3}}{\frac{16}{18}} & \text{if } k=1\\ \frac{\frac{1}{6}+\frac{1}{6}}{\frac{16}{18}} & \text{if } k=2\\ \frac{\frac{1}{9}+\frac{1}{9}}{\frac{16}{18}} & \text{if } k=3\end{cases}=\begin{cases}\frac{3}{8} & \text{if } k=1\\ \frac{3}{8} & \text{if } k=2\\ \frac{2}{8} & \text{if } k=3\end{cases}$$

Now, the mean will be

$$\mathbb{E}[K|N=1\cup N=2]=1\times\frac{3}{8}+2\times\frac{3}{8}+3\times\frac{2}{8}=\frac{15}{8}$$

and the variance will be

$$\begin{aligned}\mathrm{Var}\left(K|N=1\cup N=2\right)&=\mathbb{E}\left[K^2|N=1\cup N=2\right]-\mathbb{E}[K|N=1\cup N=2]^2\\&=1^2\times\frac{3}{8}+2^2\times\frac{3}{8}+3^2\times\frac{2}{8}-\left(\frac{15}{8}\right)^2\\&=\frac{39}{64}\\&\approx 0.61\end{aligned}$$

(f) Let $X$ be his total expenditure. Using the total expectation theorem, we have

$$\mathbb{E}[X]=\mathbb{E}[X|N=1]\mathbb{P}[N=1]+\mathbb{E}[X|N=2]\mathbb{P}[N=2]+\mathbb{E}[X|N=3]\mathbb{P}[N=3]$$

Since each book has an expected price of 3, $\mathbb{E}[X|N=n]=3\times n$, giving

$$\begin{aligned}\mathbb{E}[X]&=\mathbb{E}[X|N=1]\mathbb{P}[N=1]+\mathbb{E}[X|N=2]\mathbb{P}[N=2]+\mathbb{E}[X|N=3]\mathbb{P}[N=3]\\&=3\times\frac{11}{18}+6\times\frac{5}{18}+9\times\frac{2}{18}\\&=\frac{9}{2}\ .\end{aligned}$$

# 4 Law of Large Numbers

Recall that the *Law of Large Numbers* holds if, for every $\varepsilon>0$,

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{1}{n}S_n-\mathbb{E}\left[\frac{1}{n}S_n\right]\right|>\varepsilon\right)=0.$$

In class, we saw that the Law of Large Numbers holds for $S_n=X_1+\cdots+X_n$, where the $X_i$'s are i.i.d. random variables. This problem explores if the Law of Large Numbers holds under other circumstances.

Packets are sent from a source to a destination node over the Internet. Each packet is sent on a certain route, and the routes are disjoint. Each route has a failure probability of $p\in(0,1)$ and different routes fail independently. If a route fails, all packets sent along that route are lost. You can assume that the routing protocol has no knowledge of which route fails.

For each of the following routing protocols, determine whether the Law of Large Numbers holds when $S_n$ is defined as the total number of received packets out of $n$ packets sent. Answer **Yes** if the Law of Large Number holds, or **No** if not. Give a justification of your answer. (Whenever convenient, you can assume that $n$ is even.)

(a) **Yes** or **No**: Each packet is sent on a completely different route.

(b) **Yes** or **No**: The packets are split into $n/2$ pairs of packets. Each pair is sent together on its own route (i.e., different pairs are sent on different routes).

(c) **Yes** or **No**: The packets are split into 2 groups of $n/2$ packets. All the packets in each group are sent on the same route, and the two groups are sent on different routes.

(d) **Yes** or **No**: All the packets are sent on one route.

**Solution:**

(a) **Yes.** Define $X_i$ to be 1 if a packet is sent successfully on route $i$. Then $X_i$, $i = 1,\dots,n$ is 0 with probability $p$ and 1 otherwise. Since we have individual routes for each packet, we have a total of $n$ routes. The total number of successful packets sent is hence $S_n = X_1 + \cdots + X_n$. Since $S_n$ is a sum of i.i.d. Bernoulli random variables, $S_n \sim \text{Binomial}(n, 1-p)$.

Now similar to notation in the lecture notes, we define $A_n = S_n/n$ to be the fraction of successful packets sent, out of the $n$ packets. Moreover, for each $X_i$,

$$\mathbb{E}[X_i] = 1 - p$$

and

$$\text{Var}(X_i) = p(1-p).$$

Using Chebyshev's inequality:

$$\mathbb{P}[|A_n - \mathbb{E}[A_n]| > \varepsilon] = \mathbb{P}[|A_n - (1-p)| > \varepsilon] \le \frac{\text{Var}[A_n]}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty.$$

(b) **Yes.** Now we need $n/2$ routes for each pair of packets. Similarly to the previous question, we define $X_i$, $i = 1,\dots,n/2$ to be 0 with probability $p$ and 2 (packets) otherwise. Now the total number of packets is $S_n = X_1 + \cdots + X_{n/2}$ and the fraction of received packets is $A_n = S_n/n$.

Now for each $i = 1,\dots,n/2$,

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

and

$$\text{Var}(X_i) = 4p(1-p).$$

Thus,

$$\mathbb{E}[A_n] = \frac{\mathbb{E}[X_1] + \ldots + \mathbb{E}[X_{n/2}]}{n} = \frac{1}{n} \cdot \frac{n}{2} \cdot 2(1-p) = 1-p$$

and

$$\text{Var}[A_n] = \frac{1}{n^2}\left(\text{Var}[X_1] + \ldots + \text{Var}[X_{n/2}]\right) = \frac{1}{n^2} \cdot \frac{n}{2} 4p(1-p) = \frac{2p(1-p)}{n}.$$

Finally, we get:

$$\mathbb{P}[|A_n - \mathbb{E}[A_n]| > \varepsilon] = \mathbb{P}[|A_n - (1-p)| > \varepsilon] \leq \frac{2p(1-p)}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty.$$

(c) **No.** In this situation, we have that no packets get through with probability $p^2$, half the packets get through with probability $2p(1-p)$, and all the packets get through with probability $(1-p)^2$. This tells us that $\frac{1}{n}S_n$ is 0 with probability $p^2$, $\frac{1}{2}$ with probability $2p(1-p)$, and 1 with probability $(1-p)^2$. Since $\mathbb{E}\left[\frac{1}{n}S_n\right] = 1-p$, this gives us that

$$\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| = \begin{cases} 1-p & \text{with probability } p^2 \\ \left|p - \frac{1}{2}\right| & \text{with probability } 2p(1-p) \\ p & \text{with probability } (1-p)^2 \end{cases}$$

We now consider two cases: either $p = \frac{1}{2}$ or $p \neq \frac{1}{2}$. In the former case, we can take $\varepsilon = \frac{1}{4}$, and we'll have that

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right) &= \mathbb{P}\left(\frac{1}{n}S_n = 0 \cup \frac{1}{n}S_n = 1\right) \\ &= \frac{1}{2} \end{aligned}$$

In the latter case, we can take $\varepsilon = \frac{\min\left(1-p, \left|p-\frac{1}{2}\right|, p\right)}{2}$ and we'll have that

$$\mathbb{P}\left(\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right) = 1$$

Since neither of these probabilities converge to zero as $n \to \infty$, we have that the WLLN does not hold in either case.

(d) **No.** In this case, we have that no packets get through with probability $p$ and all the packets get through with probability $(1-p)$. Hence,

$$\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| = \begin{cases} 1-p & \text{with probability } p \\ p & \text{with probability } (1-p) \end{cases}$$

So if we take $\varepsilon = \frac{\min(p, 1-p)}{2}$, we have that

$$\mathbb{P}\left(\left|\frac{1}{n}S_n - \mathbb{E}\left[\frac{1}{n}S_n\right]\right| > \varepsilon\right) = 1$$

As in the previous part, because this does not converge to 0 as $n \to \infty$, we have that the WLLN does not hold.

For problems (c) and (d), you should've had the intuition that since the packets are automatically sent through 1 or 2 routes, increasing $n$ does not really help for LLN.