CS 70 Discrete Mathematics and Probability Theory
Spring 2021 **HW 11**

# 1 Poisson Coupling

(a) Let $X$, $Y$ be discrete random variables taking values in $\mathbb{N}$. A common way to measure the "distance" between two probability distributions is known as the total variation norm, and it is given by

$$d(X,Y) = \frac{1}{2}\sum_{k=0}^{\infty} |\mathbb{P}(X=k) - \mathbb{P}(Y=k)|.$$

Show that

$$d(X,Y) \le \mathbb{P}(X \ne Y). \tag{1}$$

[*Hint*: Use the Law of Total Probability to split up the events according to $\{X = Y\}$ and $\{X \ne Y\}$.]

(b) Show that if $X_i, Y_i$, $i \in \mathbb{Z}_+$ are discrete random variables taking values in $\mathbb{N}$, then $\mathbb{P}(\sum_{i=1}^n X_i \ne \sum_{i=1}^n Y_i) \le \sum_{i=1}^n \mathbb{P}(X_i \ne Y_i)$. [*Hint*: Maybe try the Union Bound.]

Notice that the LHS of (1) only depends on the *marginal* distributions of $X$ and $Y$, whereas the RHS depends on the *joint* distribution of $X$ and $Y$. This leads us to the idea that we can find a good bound for $d(X,Y)$ by choosing a special joint distribution for $(X,Y)$ which makes $\mathbb{P}(X \ne Y)$ small.

We will now introduce a coupling argument which shows that the distribution of the sum of independent Bernoulli random variables with parameters $p_i$, $i = 1,\dots,n$, is close to a Poisson distribution with parameter $\lambda = p_1 + \cdots + p_n$.

(c) Let $(X_i,Y_i)$ and $(X_i,Y_j)$ be independent for $i \ne j$, but for each $i$, $X_i$ and $Y_i$ are *coupled*, meaning that they have the following discrete distribution:

$$\begin{aligned}
\mathbb{P}(X_i = 0, Y_i = 0) &= 1 - p_i, & \\
\mathbb{P}(X_i = 1, Y_i = y) &= \frac{e^{-p_i} p_i^y}{y!}, & y = 1,2,\dots, \\
\mathbb{P}(X_i = 1, Y_i = 0) &= e^{-p_i} - (1 - p_i), & \\
\mathbb{P}(X_i = x, Y_i = y) &= 0, & \text{otherwise.}
\end{aligned}$$

Recall that all valid distributions satisfy two important properties. Argue that this distribution is a valid joint distribution.

(d) Show that $X_i$ has the Bernoulli distribution with probability $p_i$.

(e) Show that $Y_i$ has the Poisson distribution with parameter $\lambda = p_i$.

(f) Show that $\mathbb{P}(X_i \neq Y_i) \leq p_i^2$.

(g) Finally, show that $d(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i) \leq \sum_{i=1}^n p_i^2$.

**Solution:**

(a) One has

$$\begin{aligned} d(X,Y) &= \frac{1}{2}\sum_{k=0}^{\infty} |\mathbb{P}(X=k) - \mathbb{P}(Y=k)| \\ &= \frac{1}{2}\sum_{k=0}^{\infty} |\mathbb{P}(X=k, X=Y) + \mathbb{P}(X=k, X\neq Y) - \mathbb{P}(Y=k, X=Y) \\ &\qquad - \mathbb{P}(Y=k, X\neq Y)|. \end{aligned}$$

Note that the event $\{X=k, X=Y\}$ is the same as $\{Y=k, X=Y\}$ (they both equal the event $\{X=Y=k\}$). Hence, these terms cancel and we have

$$\begin{aligned} d(X,Y) &= \frac{1}{2}\sum_{k=0}^{\infty} |\mathbb{P}(X=k, X\neq Y) - \mathbb{P}(Y=k, X\neq Y)| \\ &\leq \frac{1}{2}\left(\sum_{k=0}^{\infty} \mathbb{P}(X=k, X\neq Y) + \sum_{k=0}^{\infty} \mathbb{P}(Y=k, X\neq Y)\right) = \frac{1}{2}\left(\mathbb{P}(X\neq Y) + \mathbb{P}(X\neq Y)\right). \end{aligned}$$

We see that the factor of $1/2$ disappears and we are left with

$$d(X,Y) \leq \mathbb{P}(X \neq Y). \tag{2}$$

(b) Note that the event $\{\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\} \subseteq \{\exists i\; X_i \neq Y_i\}$, since if the two summations $\sum_{i=1}^n X_i$ and $\sum_{i=1}^n Y_i$ are different, then there must be at least one term which is different between the summations. Now, we can write

$$\mathbb{P}\left(\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\right) \leq \mathbb{P}(X_i \neq Y_i \text{ for some } i) = \mathbb{P}\left(\bigcup_{i=1}^n \{X_i \neq Y_i\}\right).$$

Now, we apply the Union Bound to the term on the right to obtain

$$\mathbb{P}\left(\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\right) \leq \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i). \tag{3}$$

(c) We need to verify that the probabilities sum to 1. Indeed,

$$\begin{aligned} \mathbb{P}(X_i=0, Y_i=0) + \mathbb{P}(X_i=1, Y_i=0) + \sum_{y=1}^{\infty} \mathbb{P}(X_i=1, Y_i=y) &= e^{-p_i} + \sum_{y=1}^{\infty} \frac{e^{-p_i} p_i^y}{y!} \\ &= \cancel{e^{-p_i}} + 1 - \cancel{e^{-p_i}} = 1. \end{aligned}$$

Also, the probabilities are non-negative, since $e^{-p_i} \geq 1 - p_i$ always.

(d) We know that $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 0, Y_i = 0) = 1 - p_i$, and that $\mathbb{P}(X_i = x) = 0$ for any $x \notin \{0, 1\}$. Then $\mathbb{P}(X_i = 0) + \mathbb{P}(X_i = 1) = 1$, so $\mathbb{P}(X_i = 1) = p_i$. This is a sufficient approach, but to be fully explicit, we can verify through direct calculation that $\mathbb{P}(X_i = 1) = p_i$:

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 1, Y_i = 0) + \sum_{y=1}^{\infty} \mathbb{P}(X_i = 1, Y_i = y) = e^{-p_i} - (1 - p_i) + \sum_{y=1}^{\infty} \frac{e^{-p_i} p_i^y}{y!}$$
$$= \cancel{e^{-p_i}} - 1 + p_i + 1 - \cancel{e^{-p_i}} = p_i.$$

Hence, $X_i$ has the Bernoulli distribution with probability of success $p_i$.

(e) We see that $\mathbb{P}(Y_i = 0) = \mathbb{P}(X_i = 0, Y_i = 0) + \mathbb{P}(X_i = 1, Y_i = 0) = e^{-p_i}$, and for $y = 1, 2, \ldots$ we have

$$\mathbb{P}(Y_i = y) = \mathbb{P}(X_i = 1, Y_i = y) = \frac{e^{-p_i} p_i^y}{y!}.$$

This is indeed the Poisson distribution with rate $\lambda = p_i$.

(f) We can recognize that $\mathbb{P}(X_i \neq Y_i) = 1 - \mathbb{P}(X_i = Y_i)$:

$$\begin{aligned}
\mathbb{P}(X_i \neq Y_i) &= 1 - \mathbb{P}(X_i = Y_i) = 1 - \mathbb{P}(X_i = 0, Y_i = 0) - \mathbb{P}(X_i = 1, Y_i = 1) \\
&= 1 - (1 - p_i) - \frac{e^{-p_i} p_i^1}{1!} \\
&= p_i - e^{-p_i} p_i \\
&= p_i(1 - e^{-p_i}) \leq p_i^2.
\end{aligned}$$

In the last line, we are using $1 - e^{-p_i} \leq p_i$. Note that this follows from $e^{-p_i} \geq 1 - p_i$ by rearranging the inequality.

Alternatively, we can compute $\mathbb{P}(X_i \neq Y_i)$ directly:

$$\mathbb{P}(X_i \neq Y_i) = \mathbb{P}(X_i = 1, Y_i = 0) + \mathbb{P}(X_i = 1, Y_i \geq 2) = e^{-p_i} - (1 - p_i) + \sum_{y=2}^{\infty} \frac{e^{-p_i} p_i^y}{y!}$$
$$= e^{-p_i} - (1 - p_i) + 1 - e^{-p_i} - p_i e^{-p_i} = p_i(1 - e^{-p_i}) \leq p_i^2.$$

(g) Thanks to the inequalities we have proved, we can write down

$$d\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} Y_i\right) \underbrace{\leq}_{(2)} \mathbb{P}\left(\sum_{i=1}^{n} X_i \neq \sum_{i=1}^{n} Y_i\right) \underbrace{\leq}_{(3)} \sum_{i=1}^{n} \mathbb{P}(X_i \neq Y_i) \leq \sum_{i=1}^{n} p_i^2.$$

The last inequality is from Part (f).

This is known as Le Cam's Theorem. It provides precise bounds on how far the sum of independent Bernoulli random variables is from a Poisson distribution.

## 2 Combining Distributions

Let $X \sim Pois(\lambda), Y \sim Pois(\mu)$ be independent. Prove that the distribution of $X$ conditional on $X+Y$ is a binomial distribution, e.g. that $X|X+Y$ is binomial. What are the parameters of the binomial distribution?

*Hint*: Recall that we can prove $X|X+Y$ is binomial if it's PMF is of the same form

**Solution:**

$$\begin{aligned}
P(X=k|X+Y=n) &= \frac{P(X=k \cap X+Y=n)}{P(X+Y=n)} = \frac{P(X=k \cap Y=n-k)}{P(X+Y=n)} \\
&= \frac{\frac{\lambda^k e^{-\lambda}}{k!} \times \frac{\mu^{n-k} e^{-\mu}}{(n-k)!}}{\frac{(\lambda+\mu)^n e^{-(\lambda+\mu)}}{n!}} \\
&= \frac{n!}{k!(n-k)!} \times \frac{e^{-\lambda} e^{-\mu}}{e^{-(\lambda+\mu)}} \times \frac{\lambda^k \mu^{n-k}}{(\lambda+\mu)^n} \\
&= \binom{n}{k} (\frac{\lambda}{\lambda+\mu})^k (\frac{\mu}{\lambda+\mu})^{n-k} \\
&= \binom{n}{k} (\frac{\lambda}{\lambda+\mu})^k (1-\frac{\lambda}{\lambda+\mu})^{n-k}
\end{aligned}$$

Hence, it is a binomial distribution with $Z \sim \text{Bin}(n, \frac{\lambda}{\lambda+\mu})$.

## 3 Double-Check Your Intuition Again

(a) You roll a fair six-sided die and record the result $X$. You roll the die again and record the result $Y$.

(i) What is $\text{cov}(X+Y, X-Y)$?

(ii) Prove that $X+Y$ and $X-Y$ are not independent.

For each of the problems below, if you think the answer is "yes" then provide a proof. If you think the answer is "no", then provide a counterexample.

(b) If $X$ is a random variable and $\text{Var}(X) = 0$, then must $X$ be a constant?

(c) If $X$ is a random variable and $c$ is a constant, then is $\text{Var}(cX) = c\,\text{Var}(X)$?

(d) If $A$ and $B$ are random variables with nonzero standard deviations and $\text{Corr}(A,B) = 0$, then are $A$ and $B$ independent?

(e) If $X$ and $Y$ are not necessarily independent random variables, but $\text{Corr}(X,Y) = 0$, and $X$ and $Y$ have nonzero standard deviations, then is $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$?

(f) If $X$ and $Y$ are random variables then is $\mathbb{E}(\max(X,Y)\min(X,Y)) = \mathbb{E}(XY)$?

(g) If $X$ and $Y$ are independent random variables with nonzero standard deviations, then is

$$\operatorname{Corr}(\max(X,Y),\min(X,Y)) = \operatorname{Corr}(X,Y)?$$

**Solution:**

(a) (i) $\operatorname{cov}(X+Y,X-Y) = \operatorname{cov}(X,X) + \operatorname{cov}(X,Y) - \operatorname{cov}(Y,X) - \operatorname{cov}(Y,Y) = \operatorname{cov}(X,X) - \operatorname{cov}(Y,Y) = 0$

(ii) Observe that $\mathbb{P}(X+Y=7, X-Y=0) = 0$ because if $X-Y=0$, then the sum of our two dice rolls must be even. However both $\mathbb{P}(X+Y=7)$ and $\mathbb{P}(X-Y=0)$ are nonzero so $\mathbb{P}(X+Y=7, X-Y=0) \neq \mathbb{P}(X+Y=7)\cdot\mathbb{P}(X-Y=0)$

(b) Yes. If we write $\mu = \mathbb{E}[X]$, then $0 = \operatorname{Var}(X) = \mathbb{E}[(X-\mu)^2]$ so $(X-\mu)^2$ must be identically 0 since perfect squares are non-negative. Thus $X = \mu$.

(c) No. We have $\operatorname{Var}(cX) = \mathbb{E}[(cX - \mathbb{E}[cX])^2] = c^2\,\mathbb{E}[(X-\mathbb{E}[X])^2] = c^2\operatorname{Var}(X)$ so if $\operatorname{Var}(X) \neq 0$ and $c \neq 0$ or $c \neq 1$ then $\operatorname{Var}(cX) \neq c\operatorname{Var}(X)$. This does prove that $\sigma(cX) = c\sigma(X)$ though.

(d) No. Let $A = X+Y$ and $B = X-Y$ from part (a). Since $A$ and $B$ are not constants then part (b) says they must have nonzero variances which means they also have nonzero standard deviations. Part (a) says that their covariance is 0 which means they are uncorrelated, and that they are not independent.

Recall from lecture that the converse is true though.

(e) Yes. If $\operatorname{Corr}(X,Y) = 0$, then $\operatorname{cov}(X,Y) = 0$. We have $\operatorname{Var}(X+Y) = \operatorname{cov}(X+Y,X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{cov}(X,Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$.

(f) Yes. For any values $x,y$ we have $\max(x,y)\min(x,y) = xy$. Thus, $\mathbb{E}[\max(X,Y)\min(X,Y)] = \mathbb{E}[XY]$.

(g) No. You may be tempted to think that because $(\max(x,y),\min(x,y))$ is either $(x,y)$ or $(y,x)$, then $\operatorname{Corr}(\max(X,Y),\min(X,Y)) = \operatorname{Corr}(X,Y)$ because $\operatorname{Corr}(X,Y) = \operatorname{Corr}(Y,X)$. That reasoning is flawed because $(\max(X,Y),\min(X,Y))$ is not always equal to $(X,Y)$ or always equal to $(Y,X)$ and the inconsistency affects the correlation. It is possible for $X$ and $Y$ to be independent while $\max(X,Y)$ and $\min(X,Y)$ are not.

For a concrete example, suppose $X$ is either 0 or 1 with probability $1/2$ each and $Y$ is independently drawn from the same distribution. Then $\operatorname{Corr}(X,Y) = 0$ because $X$ and $Y$ are independent. Even though $X$ never gives information about $Y$, if you know $\max(X,Y) = 0$ then you know for sure $\min(X,Y) = 0$.

More formally, $\max(X,Y) = 1$ with probability $3/4$ and 0 with probability $1/4$, and $\min(X,Y) = 1$ with probability $1/4$ and 0 with probability $3/4$. This means

$$\mathbb{E}[\max(X,Y)] = 1 * 3/4 + 0 * 1/4 = 3/4$$

and

$$\mathbb{E}[\min(X,Y)] = 1*1/4 + 0*3/4 = 1/4.$$

Thus,

$$\text{cov}(\max(X,Y),\min(X,Y)) = \mathbb{E}[\max(X,Y)\min(X,Y)] - 3/16 = 1/4 - 3/16 = 1/16 \neq 0.$$

We conclude that $\text{Corr}(\max(X,Y),\min(X,Y)) \neq 0 = \text{Corr}(X,Y)$.

## 4 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 $\mathbb{E}[X] = 0$, $\text{Var}(X) = \sigma^2 < \infty$, 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). Plug in the minimizing value of $c$ to prove the following bound:

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

(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. Why is this not always correct in general? 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, $\mathbb{P}(X \geq \mathbb{E}[X] + \alpha) > \frac{\text{Var}(X)}{2\alpha^2}$ or $\mathbb{P}(X \leq \mathbb{E}[X] - \alpha) > \frac{\text{Var}(X)}{2\alpha^2}$.

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 $\mathrm{Var}(X) = 2$. What bound would Markov's inequality give for $\mathbb{P}[X \geq 5]$? What bound would Chebyshev's inequality give for $\mathbb{P}[X \geq 5]$? What about for the bound we proved in part (b)? (*Note*: Recall that the bound from part (b) 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. Plugging in this value for $c$ yields us the desired inequality.

This bound is also known as Cantelli's inequality.

(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.

## 5 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, and give a brief 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

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

Using Chebyshev's inequality:

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

(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

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

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

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

Finally, we get:

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

(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.

# 6 Practical Confidence Intervals

(a) It's New Year's Eve, and you're re-evaluating your finances for the next year. Based on previous spending patterns, you know that you spend \$1500 per month on average, with a standard deviation of \$500, and each month's expenditure is independently and identically distributed. As a college student, you also don't have any income. How much should you have in your bank account if you don't want to run out of money this year, with probability at least 95%?

(b) As a UC Berkeley CS student, you're always thinking about ways to become the next billionaire in Silicon Valley. After hours of brainstorming, you've finally cut your list of ideas down to 10, all of which you want to implement at the same time. A venture capitalist has agreed to back all 10 ideas, as long as your net return from implementing the ideas is positive with at least 95% probability.

Suppose that implementing an idea requires 50 thousand dollars, and your start-up then succeeds with probability $p$, generating 150 thousand dollars in revenue (for a net gain of 100 thousand dollars), or fails with probability $1-p$ (for a net loss of 50 thousand dollars). The success of each idea is independent of every other. What is the condition on $p$ that you need to satisfy to secure the venture capitalist's funding?

(c) One of your start-ups uses error-correcting codes, which can recover the original message as long as at least 1000 packets are received (not erased). Each packet gets erased independently with probability 0.8. How many packets should you send such that you can recover the message with probability at least 99%?

**Solution:**

(a) Let $T$ be the random variable representing the amount of money we spend in the year.

We have $T = \sum_{i=1}^{12} X_i$, where $X_i$ represents the spending in the $i$-th month. So, $\mathbb{E}[T] = 12 \cdot \mathbb{E}[E_1] = 18000$.

And, since the $X_i$s are independent, $\text{Var}(T) = 12 \cdot \text{Var}(X_1) = 12 \cdot 500^2 = 3,000,000$.

We want to have enough money in our bank account so that we don't finish the year in debt with 95% confidence. So, we want to keep some money $\varepsilon$ more than the mean expenditure such that the probability of deviating above the mean by more than $\varepsilon$ is less than 0.05.

Let's use Chebyshev's inequality here to express this.

$$\mathbb{P}(|T - \mathbb{E}[T]| \geq \varepsilon) \leq \frac{\text{Var}(T)}{\varepsilon^2} \leq 0.05$$

This gives us $\varepsilon^2 \geq \frac{3,000,000}{0.05}$. So, $\varepsilon \geq 7746$. This means that we want to have a balance of $\geq \mathbb{E}[T] + \varepsilon = 25746$.

Observe that here, while we wanted to estimate $\mathbb{P}(T - \mathbb{E}[T] \geq \varepsilon)$, Chebyshev's inequality only gives us information about $\mathbb{P}(|T - \mathbb{E}[T]| \geq \varepsilon)$. But since

$$\mathbb{P}(|T - \mathbb{E}[T]| \geq \varepsilon) \geq \mathbb{P}(T - \mathbb{E}[T] \geq \varepsilon),$$

this is fine. We just get a more conservative estimate.

(b) For this question, to keep the numbers from exploding, let's work in thousands of dollars. Let $X_i$ be the profit made from idea $i$, and $T$ be the total profit made. We have $T = \sum_{i=1}^{10} X_i$.

Here, $\mathbb{E}[X_1] = 100p - 50(1-p) = 150p - 50$.

And $\text{Var}(X_1) = 150^2 p(1-p)$ as the distribution of $X_1$ is a shifted and scaled Bernoulli distribution. Using $\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2$ yields the same answer.

We have, $\mathbb{E}[T] = 10 \cdot \mathbb{E}[X_1]$. Similarly, $\text{Var}(T) = 10 \cdot \text{Var}(X_1)$.

Now, we want to bound the probability of T going below 0 by 0.05. In other words, we want $\mathbb{P}(T < 0) \leq 0.05$.

But, in order to apply Chebyshev's inequality, we need to look at deviation from the mean. We use the assumption that to get our funding we obviously need $\mathbb{E}[T] > 0$. Then:

$$\mathbb{P}(T < 0) \leq \mathbb{P}(T \leq 0 \cup T \geq 2\mathbb{E}[T]) = \mathbb{P}(|T - \mathbb{E}[T]| \geq \mathbb{E}[T]) \leq \frac{\text{Var}(T)}{\mathbb{E}[T]^2} \leq 0.05$$

Looking at just the last inequality, we have:

$$\frac{\text{Var}(T)}{\mathbb{E}[T]^2} = \frac{10 \cdot \text{Var}(X_1)}{100 \cdot \mathbb{E}[X_1]^2} = \frac{\text{Var}(X_1)}{10 \cdot \mathbb{E}[X_1]^2} \leq 0.05$$

$$\therefore \frac{\text{Var}(X_1)}{\mathbb{E}[X_1]^2} \leq 0.5$$

Now, substituting what we have for variance and expectation, we get the following:

$$-22500p^2 + 22500p \leq 0.5(150p - 50)^2$$

which gives us the quadratic:

$$33750p^2 - 30000p + 1250 \geq 0$$

The solutions for $p$ are $p \geq \frac{1}{9}(4 + \sqrt{13})$ and $p \leq \frac{1}{9}(4 - \sqrt{13})$. So $p \geq 0.845$ or $\leq 0.0438$.

The relevant solution here is to pick $p \geq 0.845$, since the other solution yields negative expectation (contradicting the earlier assumption of positive expectation).

(c) We want $k = 1000$ packets to get across without being erased. Say we send $n$ packets. Let $X_i$ be the indicator random variable representing whether the $i$th packet got across or not.

Let the total number of unerased packets sent across be $T$. We have $T = \sum_{i=1}^{n} X_i$ and we want $T \geq 1000$.

We want $\mathbb{P}(T < 1000) \leq 0.01$. Now, let's try to get this in a form so that we can use Chebyshev's inequality. We know that $\mathbb{E}[T] > 1000$, so we can say that

$$\begin{aligned}\mathbb{P}(T < 1000) &\leq \mathbb{P}\big(T \leq 1000 \cup T \geq \mathbb{E}[T] + (\mathbb{E}[T] - 1000)\big) \\ &= \mathbb{P}\big(|T - \mathbb{E}[T]| \geq (\mathbb{E}[T] - 1000)\big) \leq \frac{\text{Var}(T)}{(\mathbb{E}[T] - 1000)^2} \leq 0.01.\end{aligned}$$

What is $\mathbb{E}[T]$? $\mathbb{E}[T] = n\mathbb{E}[X_1] = n(1-p) = 0.2n$.

Next, what is $\text{Var}(T)$? $\text{Var}(T) = n\text{Var}(X_1) = np(1-p) = 0.16n$.

Now, $\frac{\text{Var}(T)}{(\mathbb{E}[T] - k)^2} \leq 0.01 \implies 16n \leq (0.2n - 1000)^2$. This gives us the quadratic:

$$0.04n^2 - 416n + 1000000 \geq 0$$

Solving the last quadratic, we get $n \geq 6629$ or $n \leq 3774$. Since the second inequality doesn't make sense for our situation, our answer is $n \geq 6629$.

## 7 Balls in Bins Estimation

We throw $n > 0$ balls into $m \geq 2$ bins. Let $X$ and $Y$ represent the number of balls that land in bin 1 and 2 respectively.

(a) Calculate $\mathbb{E}[Y \mid X]$. [*Hint*: Your intuition may be more useful than formal calculations.]

(b) What are $L[Y \mid X]$ and $Q[Y \mid X]$ (where $Q[Y \mid X]$ is the best quadratic estimator of $Y$ given $X$)? [*Hint*: Your justification should be no more than two or three sentences, no calculations necessary! Think carefully about the meaning of the MMSE.]

(c) Unfortunately, your friend is not convinced by your answer to the previous part. Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(d) Compute $\text{Var}(X)$.

(e) Compute $\text{cov}(X,Y)$.

(f) Compute $L[Y \mid X]$ using the formula. Ensure that your answer is the same as your answer to part (b).

**Solution:**

(a) $\mathbb{E}[Y \mid X = x] = (n-x)/(m-1)$, because once we condition on $x$ balls landing in bin 1, the remaining $n-x$ balls are distributed uniformly among the other $m-1$ bins. Therefore,

$$\mathbb{E}[Y \mid X] = \frac{n-X}{m-1}.$$

(b) We showed that $\mathbb{E}[Y \mid X]$ is a linear function of $X$. Since $\mathbb{E}[Y \mid X]$ is the best *general* estimator of $Y$ given $X$, it must also be the best *linear* and *quadratic* estimator of $Y$ given $X$, i.e. $\mathbb{E}[Y \mid X]$, $L[Y \mid X]$, and $Q[Y \mid X]$ all coincide.

(c) Let $X_i$ be the indicator that the $i$th ball falls in bin 1. Then, $X = \sum_{i=1}^{n} X_i$, and by linearity of expectation, $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n/m$, since there are $n$ indicators and each ball has a probability $1/m$ of landing in bin 1. By symmetry, $\mathbb{E}[Y] = n/m$ as well.

(d) The number of balls that falls into the first bin is binomially distributed with parameters $n$ and $1/m$. Hence the variance is $n(1/m)(1-1/m)$.

(e) Let $X_i$ be as before, and let $Y_i$ be the indicator that the $i$th ball falls into bin 2.

$$\text{cov}(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{cov}(X_i, Y_j)$$

We can compute $\text{cov}(X_i, Y_i) = \mathbb{E}[X_i Y_i] - \mathbb{E}[X_i]\,\mathbb{E}[Y_i] = 0 - (1/m)(1/m) = -1/m^2$ (note that $\mathbb{E}[X_i Y_i] = 0$ because it is impossible for a ball to land in both bins 1 and 2). Also, we have $\text{cov}(X_i, Y_j) = 0$ because the indicator for the $i$th ball is independent of the indicator for the $j$th ball when $i \neq j$. Hence, $\text{cov}(X,Y) = n(-1/m^2) = -n/m^2$.

(f)

$$
\begin{aligned}
L[Y \mid X] &= \mathbb{E}[Y] + \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}(X - \mathbb{E}[X]) \\
&= \frac{n}{m} + \frac{-n/m^2}{n(1/m)(1-1/m)}\left(X - \frac{n}{m}\right) \\
&= \frac{n}{m} - \frac{1}{m-1}\left(X - \frac{n}{m}\right) \\
&= \frac{mn - n - mX + n}{m(m-1)} = \frac{n-X}{m-1}
\end{aligned}
$$