CS 70 Discrete Mathematics and Probability Theory
Fall 2020

HW 12

1 Graph

Consider a random graph (undirected, no multi-edges, no self-loops) on n nodes, where each possible edge exists independently with probability p. Let X be the number of isolated nodes (nodes with degree 0).

(a) What is $\text{Var}(X)$?

Solution:

(a) Define X_i as the indicator when node i is isolated. Since $Var(X) = E[X^2] - E[X]^2$, and $E[X]$ is $n(1-p)^{n-1}$, it remains to calculate $E[X^2]$.

$$\begin{aligned}
E[X^2] &= E[(X_1 + \ldots + X_n)^2] \\
&= E[(\sum_{i=1}^{n} X_i^2) + (\sum_{i \neq j} X_i X_j)]. \\
&= \sum_{i=1}^{n} (1-p)^{n-1} + \sum_{i \neq j} (1-p)^{2n-3} \\
&= n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}
\end{aligned}$$

Putting it all together, $Var(X) = E[X^2] - (E[X])^2 = n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3} - n^2(1-p)^{2n-2}$.

2 Whitening

Let X and Y be two random variables, with $\text{Var}(X) > 0, \text{Var}(Y) > 0$. Show that it is possible to construct $\tilde{X} = aX + bY$ and $\tilde{Y} = cX + dY$, where a, b, c, d are scalars to be chosen subject to the constraint $ad - bc \neq 0$, such that $\text{cov}(\tilde{X}, \tilde{Y}) = 0$.

You may find it unnecessary to transform Y, that is, you only need to solve for a, b to get $\text{cov}(\tilde{X}, Y) = 0$.

Solution:

The covariance between $\tilde{X}$ and $\tilde{Y}$ is given by

$$\text{cov}(\tilde{X}, \tilde{Y}) = ac\,\text{Var}(X) + bd\,\text{Var}(Y) + (ad + bc)\,\text{cov}(X, Y).$$

Our goal is to make this quantity 0. This is an underdetermined equation in a,b,c,d. We can start by choosing $a = d = 1$ and $c = 0$. This simplifies the equation to $b\,\mathrm{Var}(Y) + \mathrm{cov}(X,Y) = 0$, from which we get

$$b = -\frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(Y)}.$$

Hence a suitable transformation is given by

$$\tilde{X} = X - \frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(Y)} Y$$

and $\tilde{Y} = Y$ (note that the condition $ad - bc \neq 0$ is indeed satisfied – this condition was imposed to avoid trivial solutions).

3 Probabilistic Bounds

A random variable X has variance $\mathrm{Var}(X) = 9$ 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) $\mathbb{E}\left[X^2\right] = 13$.

(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 $9 = \mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - 2^2$, 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. 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 $a \neq b$. 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 (**??**)) 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. 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 $\mathbb{P}[X \leq 1] = \mathbb{P}[10 - X \geq 9] \leq 8/9$.

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

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

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

4 Subset Card Game

Jonathan and Yiming are playing a card game. Jonathan has $k > 2$ cards, and each card has a real number written on it. Jonathan tells Yiming (truthfully), that the sum of the card values is 0, and that the sum of squares of the values on the cards is 1. Specifically, if the card values are $c_1, c_2, \ldots, c_k$, then we have $\sum_{i=1}^{k} c_i = 0$ and $\sum_{i=1}^{k} c_i^2 = 1$. Jonathan and Yiming also agree on a positive target value of α.

The cards are then going to be dealt randomly in the following fashion: for each card in the deck, a fair coin is flipped. If the coin lands heads, then the card goes to Yiming, and if the coin lands tails, the card goes to Jonathan. Note that it is possible for either player to end up with no cards/all the cards.

A player wins the game if the sum of the card values in their hand is at least α, otherwise it is a tie.

(a) Prove that the probability that Yiming wins is at most $\frac{1}{8\alpha^2}$.

(b) Find a deck of k cards and target value α where the probability that Yiming wins is exactly $\frac{1}{8\alpha^2}$.

Solution:

(a) Let I_i be the indicator random variable indicating whether or not card i goes to Yiming. Define $S = \sum_{i=1}^{k} c_i I_i$ as the value of Yiming's hand. Then, we see that $\mathbb{E}[S] = \sum_{i=1}^{k} c_i \cdot \frac{1}{2} = 0$ and

$$\begin{aligned}
\mathrm{Var}(S) &= \sum_{i=1}^{k} \mathrm{Var}(c_i I_i) \qquad \text{(due to independence) of } I_i \\
&= \sum_{i=1}^{k} c_i^2 \,\mathrm{Var}(I_i)
\end{aligned}$$

We know that I_i is a Bernoulli random variable, so its variance is $\frac{1}{4}$. Thus, we see that $\text{Var}(S) = \frac{1}{4}$.

By Chebyshev, we see that $\mathbb{P}(|S| \geq \alpha) \leq \frac{1}{4\alpha^2}$. Now we need to make a symmetry argument, specifically that for each value of x, $\mathbb{P}(S = x) = \mathbb{P}(S = -x)$. This is true because for each outcome where Yiming gets x, Jonathan gets $-x$, since the sum of the card values is 0. However, we also know that the reverse outcome, where Jonathan gets Yiming's cards and vice versa, has the same probability of happening.

Since the distribution of S is symmetric around 0, we see that $\mathbb{P}(|S| \geq \alpha) = 2\mathbb{P}(S \geq \alpha)$, and plugging into our bound yields $\mathbb{P}(S \geq \alpha) \leq \frac{1}{8\alpha^2}$.

(b) We now need to appeal to the equality case of Chebyshev's inequality. Recall that the derivation of Chebyshev's inequality uses Markov's inequality on the quantity $(S - \mathbb{E}[S])^2$. Let's walk through the proof that $\mathbb{P}(S^2 \geq \alpha^2) \leq \frac{\mathbb{E}[S^2]}{\alpha^2}$ again:

$$\begin{aligned}
\mathbb{E}[S^2] &= \sum_v \mathbb{P}(S^2 = v) \cdot v \\
&= \sum_{0 \leq v < \alpha^2} \mathbb{P}(S^2 = v) \cdot v + \sum_{v \geq \alpha^2} \mathbb{P}(S^2 = v) \cdot v \\
&\geq \sum_{v \geq \alpha^2} \mathbb{P}(S^2 = v) \cdot v \\
&\geq \mathbb{P}(S^2 = \alpha^2) \cdot \alpha^2
\end{aligned}$$

In order for equality to hold, then equality must hold in both the third and fourth steps. We got the third step by saying that v is always at least 0, so we can drop them from the sum. Equality holds if it is not possible for S^2 to be anything strictly between 0 and α^2. We get the fourth line by observing that since $\mathbb{P}(S^2 = v) \cdot v \geq 0$ for all $v > \alpha^2$, we can also drop them from the sum. If we want equality to hold, then these values must also be 0, meaning that S^2 cannot take on values beyond α^2. This means that S^2 is either 0 or α.

If that's the case then the values of the cards can only be $-\alpha, 0$, or α, since it is possible for Yiming to get exactly one card. There also cannot exist two cards with value α, since otherwise Yiming could potentially end up with a hand value of $2\alpha \neq \alpha$. Thus, the deck must be of the form $(\alpha, -\alpha, 0, 0, \ldots, 0)$, and we pick $\alpha = \frac{1}{\sqrt{2}}$ to ensure that the sum of squares must be 1.

5 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 α 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 $\mathrm{Var}(X) = 2$. 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.

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

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.

6 Sum of Poisson Variables

Assume that you were given two independent Poisson random variables X_1, X_2. Assume that the first has mean λ_1 and the second has mean λ_2. Prove that X_1+X_2 is a Poisson random variable with mean $\lambda_1+\lambda_2$.

Hint: Recall the binomial theorem.

$$(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$$

Solution:

To show that $X_1 + X_2$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$, we have show that

$$\mathbb{P}[(X_1 + X_2) = i] = \frac{(\lambda_1 + \lambda_2)^i}{i!} \mathrm{e}^{-(\lambda_1 + \lambda_2)}.$$

We proceed as follows:

$$\begin{aligned}
\mathbb{P}[(X_1 + X_2) = i] &= \sum_{k=0}^{i} \mathbb{P}[X_1 = k, X_2 = (i-k)] = \sum_{k=0}^{i} \frac{\lambda_1^k}{k!} \mathrm{e}^{-\lambda_1} \cdot \frac{\lambda_2^{i-k}}{(i-k)!} \mathrm{e}^{-\lambda_2} \\
&= \mathrm{e}^{-\lambda_1} \mathrm{e}^{-\lambda_2} \sum_{k=0}^{i} \frac{1}{k!(i-k)!} \lambda_1^k \lambda_2^{i-k} = \frac{\mathrm{e}^{-\lambda_1} \mathrm{e}^{-\lambda_2}}{i!} \sum_{k=0}^{i} \frac{i!}{k!(i-k)!} \lambda_1^k \lambda_2^{i-k} \\
&= \frac{\mathrm{e}^{-(\lambda_1+\lambda_2)}}{i!} \sum_{k=0}^{i} \binom{i}{k} \lambda_1^k \lambda_2^{i-k} = \frac{\mathrm{e}^{-(\lambda_1+\lambda_2)}}{i!} (\lambda_1 + \lambda_2)^i
\end{aligned}$$

In the last line, we use the binomial expansion.