CS 70 Discrete Mathematics and Probability Theory
Fall 2017 Kannan Ramchandran and Satish Rao

HW 11

Sundry

Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. (In case of homework party, you can also just describe the group.) How did you work on this homework? Working in groups of 3-5 will earn credit for your "Sundry" grade.

Please copy the following statement and sign next to it:

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up. (*signature here*)

1 Proof with Indicators

Let $n \in \mathbb{Z}_+$. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and let $A_1, \ldots, A_n$ be events. Prove that $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \mathbb{P}(A_i \cap A_j) \geq 0$.

Solution:

We write the summation with indicators. Let X_i be the indicator for event A_i, $i = 1, \ldots, n$. Then,

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \mathbb{P}(A_i \cap A_j) = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \mathbb{E}[X_i X_j] = \mathbb{E}\Big[\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j X_i X_j\Big]$$
$$= \mathbb{E}\Big[\Big(\sum_{i=1}^n \alpha_i X_i\Big)\Big(\sum_{j=1}^n \alpha_j X_j\Big)\Big] = \mathbb{E}\Big[\Big(\sum_{i=1}^n \alpha_i X_i\Big)^2\Big] \geq 0.$$

2 Balls and Bins

Throw n balls into m bins, where m and n are positive integers. Let X be the number of bins with exactly one ball. Compute $\operatorname{var} X$.

Solution:

Let X_i be the indicator that bin i has exactly one ball, for each $i = 1, \ldots, m$. One has

$$\mathbb{E}[X_i] = \binom{n}{1} \cdot \left(\frac{1}{m}\right)^1 \left(\frac{m-1}{m}\right)^{n-1} = \frac{n}{m}\left(\frac{m-1}{m}\right)^{n-1}$$

and for $j \in \{1, \ldots, n\}$, $j \neq i$,

$$\mathbb{E}[X_i X_j] = \binom{n}{1}\binom{n-1}{1}\left(\frac{1}{m}\right)^1\left(\frac{1}{m}\right)^1\left(\frac{m-2}{m}\right)^{n-2} = \frac{n(n-1)}{m^2}\left(\frac{m-2}{m}\right)^{n-2}$$

so that

$$\operatorname{var} X = m \cdot \frac{n}{m}\left(\frac{m-1}{m}\right)^{n-1} + m(m-1) \cdot \frac{n(n-1)}{m^2}\left(\frac{m-2}{m}\right)^{n-2} - m^2\left[\frac{n}{m}\left(\frac{m-1}{m}\right)^{n-1}\right]^2.$$

3 Portfolio Optimization

Suppose that there are n assets, where n is a positive integer. For each unit dollar invested in asset i, for $i = 1, \ldots, n$, with probability p_i the value of the asset will grow by α_i to $1 + \alpha_i$, and with probability $1 - p_i$ the value of the asset will shrink by α_i to $1 - \alpha_i$. Let the proportion of money invested in asset i be w_i (so that $\sum_{i=1}^n w_i = 1$), and let X_i be a random variable denoting the final value of the ith asset per unit dollar. Then $X = w_1 X_1 + \cdots + w_n X_n$ is the total value. For simplicity, assume that the outcomes of the different assets are independent.

(a) Compute the expectation $\mathbb{E}[X]$. What values of w_i maximize this quantity?

(b) Compute the variance $\operatorname{var} X$. What values of w_i minimize this quantity?

Solution:

(a) Let Y_i be a Bernoulli random variable with parameter p_i. Then $X_i = 1 + \alpha_i(2Y_i - 1)$. Hence for each asset i we have $\mathbb{E}[X_i] = 1 + \alpha_i(2\mathbb{E}[Y_i] - 1) = 1 + \alpha_i(2p_i - 1)$. By linearity, we get $\mathbb{E}[X] = \sum_{i=1}^n w_i(1 + \alpha_i(2p_i - 1))$. To maximize $\mathbb{E}[X]$, we should concentrate all weights on the asset with the largest expected return, i.e., $w_i = 1$ for $i = i^*$ and $w_i = 0$ otherwise, where $i^* \in \arg\max_{i=1,\ldots,n}\{\alpha_i(2p_i - 1)\}$.

(b) For each asset i we have $\operatorname{var} X_i = \operatorname{var}(1 + 2\alpha_i Y_i - \alpha_i) = 4\alpha_i^2 \operatorname{var} Y_i = 4\alpha_i^2 p_i(1 - p_i)$. Since all assets are independent, by linearity we have $\operatorname{var} X = 4\sum_{i=1}^n w_i^2 \alpha_i^2 p_i(1 - p_i)$. To minimize $\operatorname{var} X$, write $w_n = 1 - w_1 - \cdots - w_{n-1}$ and then

$$\operatorname{var} X = 4(1 - w_1 - \cdots - w_{n-1})^2 \alpha_n^2 p_n(1 - p_n) + 4\sum_{i=1}^{n-1} w_i^2 \alpha_i^2 p_i(1 - p_i).$$

Differentiate with respect to w_i for $i = 1, \ldots, n-1$ and set the derivative to 0 to obtain

$$-8(1 - w_1 - \cdots - w_{n-1})\alpha_n^2 p_n(1 - p_n) + 8w_i \alpha_i^2 p_i(1 - p_i) = 0.$$

Sum these equations over $i = 1, \dots, n-1$ to obtain

$$-8(n-1)w_n\alpha_n^2 p_n(1-p_n) + 8\sum_{i=1}^{n-1}\alpha_i^2 p_i(1-p_i) = 0.$$

Solving, we see that

$$w_i \propto \frac{1}{\alpha_i^2 p_i(1-p_i)}.$$

To find the constant of proportionality, we need to enforce the condition that $\sum_{i=1}^n w_i = 1$. Let

$$\frac{1}{C} := \sum_{i=1}^n \frac{1}{\alpha_i^2 p_i(1-p_i)} = \frac{\sum_{i=1}^n \prod_{j\in\{1,\dots,n\}\setminus\{i\}} \alpha_j^2 p_j(1-p_j)}{\prod_{i=1}^n \alpha_i^2 p_i(1-p_i)}$$

so that

$$w_i = \frac{C}{\alpha_i^2 p_i(1-p_i)}.$$

4 Uniform Means

Let $X_1, X_2, \dots, X_n$ be n independent and identically distributed uniform random variables on the interval $[0,1]$ (where n is a positive integer).

(a) Let $Y = \min\{X_1, X_2, \dots, X_n\}$. Find $\mathbb{E}(Y)$. [*Hint*: Use the tail sum formula, which says the expected value of a nonnegative random variable is $\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x$. Note that we can use the tail sum formula since $Y \geq 0$.]

(b) Let $Z = \max\{X_1, X_2, \dots, X_n\}$. Find $\mathbb{E}(Z)$. [*Hint*: Find the CDF.]

Solution:

(a) To calculate $\mathbb{P}(Y > y)$, where $y \in [0,1]$, this means that each X_i is greater than y, for $i = 1, \dots, n$, so $\mathbb{P}(Y > y) = (1-y)^n$. We then use the tail sum formula:

$$\mathbb{E}(Y) = \int_0^1 \mathbb{P}(Y > y)\,\mathrm{d}y = \int_0^1 (1-y)^n\,\mathrm{d}y = -\frac{1}{n+1}(1-y)^{n+1}\Big|_0^1 = \frac{1}{n+1}.$$

Alternative Solution 1:
As explained above, $\mathbb{P}[Y \leq y] = 1-(1-y)^n$. This gives us the CDF, and if we take its derivative we'll get the probability density function $f(y) = n(1-y)^{n-1}$.

Then

$$\mathbb{E}(Y) = \int_0^1 y \cdot n(1-y)^{n-1}\,\mathrm{d}y.$$

Perform a u substitution, where $u = 1-y$ and $\mathrm{d}u = -\mathrm{d}y$. We see:

$$\begin{aligned}\mathbb{E}(Y) &= n\cdot\int_0^1 -(1-u)\cdot u^{n-1}\,\mathrm{d}u = n\cdot\int_0^1 (u^n - u^{n-1})\,\mathrm{d}u = n\Big[\frac{u^{n+1}}{n+1} - \frac{u^n}{n}\Big]_{u=0}^1 \\ &= n\Big[\frac{(1-y)^{n+1}}{n+1} - \frac{(1-y)^n}{n}\Big]_{y=0}^1 = n\Big[0 - \Big(\frac{1}{n+1} - \frac{1}{n}\Big)\Big] = n\Big[\frac{1}{n} - \frac{1}{n+1}\Big] = \frac{1}{n+1}.\end{aligned}$$

Alternative Solution 2:

Consider adding another independent uniform variable X_{n+1}. $\mathbb{P}(X_{n+1} < Y)$ is just the probability that X_{n+1} is the minimum, which is $1/(n+1)$ by symmetry since all the X_i's are identical. It so happens that because X_{n+1} is a uniform variable on [0,1], this probability is equal to $\mathbb{E}(Y)$. Let f_Y denote the PDF of Y.

$$\begin{aligned}\mathbb{P}(X_{n+1} < Y) &= \int_0^1 \mathbb{P}(X_{n+1} < y \mid Y = y) f_Y(y)\,\mathrm{d}y \\ &= \int_0^1 \mathbb{P}(X_{n+1} < y) f_Y(y)\,\mathrm{d}y && \text{(by independence)} \\ &= \int_0^1 y f_Y(y)\,\mathrm{d}y && \text{(CDF of the uniform distribution)} \\ &= \mathbb{E}(Y).\end{aligned}$$

Alternative Solution 3:

Since $X_1,\ldots,X_n$ are i.i.d., their values split the interval $[0,1]$ into $n+1$ sections, and we expect these sections to be of equal length because they are uniformly distributed. Therefore, $\mathbb{E}(Y) = 1/(n+1)$, the position of the smallest indicator.

(b) We could use the tail sum formula, but it turns out that the CDF is in a form that makes it easy to take an integral. If $Z \le z$, where $z \in [0,1]$, each X_i must be less than z, which happens with probability z, so $\mathbb{P}[Z \le z] = z^n$. This gives us the CDF, and if we take its derivative we'll get the probability density function $f(z) = nz^{n-1}$. Then

$$\mathbb{E}(Z) = \int_0^1 z\cdot nz^{n-1}\,\mathrm{d}z = \int_0^1 nz^n\,\mathrm{d}z = \Big[n\cdot\frac{z^{n+1}}{n+1}\Big]_{z=0}^1 = \frac{n}{n+1}.$$

Alternative Solution:

As in the previous part, add another independent uniform random variable X_{n+1}. The probability $\mathbb{P}(X_{n+1} > Z)$ is just the probability that X_{n+1} is the maximum, which is $1/(n+1)$ by symmetry.

$$\begin{aligned}\mathbb{P}(X_{n+1} > Z) &= \int_0^1 \mathbb{P}(X_{n+1} > z \mid Z = z) f_Z(z)\,\mathrm{d}z = \int_0^1 \mathbb{P}(X_{n+1} > z) f_Z(z)\,\mathrm{d}z \\ &= \int_0^1 (1-z) f_Z(z)\,\mathrm{d}z = \int_0^1 f_Z(z)\,\mathrm{d}z - \int_0^1 z f_Z(z)\,\mathrm{d}z \\ \frac{1}{n+1} &= 1 - \mathbb{E}(Z) \\ \mathbb{E}(Z) &= \frac{n}{n+1}\end{aligned}$$

Alternative Solution 2:
Since $X_1, \ldots, X_n$ are i.i.d., their values split the interval $[0,1]$ into $n+1$ sections, and we expect these sections to be of equal length because they are uniformly distributed. The expectation of the smallest X_i is $1/(n+1)$, the expectation of the second smallest is $2/(n+1)$, etc. Therefore, $\mathbb{E}(Z) = n/(n+1)$, the position of the largest indicator.

5 Darts (Again!)

Alvin is playing darts. His aim follows an exponential distribution; that is, the probability density that the dart is x distance from the center is $f_X(x) = \exp(-x)$. The board's radius is 4 units.

(a) What is the probability the dart will stay within the board?

(b) Say you know Alvin made it on the board. What is the probability he is within 1 unit from the center?

(c) If Alvin is within 1 unit from the center, he scores 4 points, if he is within 2 units, he scores 3, etc. In other words, Alvin scores $\lfloor 5 - x \rfloor$, where x is the distance from the center. What is Alvin's expected score after one throw?

Solution:

(a) The CDF of an exponential is $\mathbb{P}[X \le x] = 1 - \exp(-x)$. Therefore,

$$\mathbb{P}[X \le 4] = 1 - \exp(-4).$$

(b) We are given that the dart must be within the board, which means that the dart is at least 4 units away from the center. We can use the definition of conditional probability:

$$\mathbb{P}[X \le 1 \mid X \le 4] = \frac{\mathbb{P}[X \le 1 \cap X \le 4]}{\mathbb{P}[X \le 4]} = \frac{\mathbb{P}[X \le 1]}{\mathbb{P}[X \le 4]} = \frac{1 - \exp(-1)}{1 - \exp(-4)}.$$

(c)

$$\begin{aligned}
\mathbb{E}[\text{score}] &= \int_0^1 4\exp(-x)\,\mathrm{d}x + \int_1^2 3\exp(-x)\,\mathrm{d}x + \int_2^3 2\exp(-x)\,\mathrm{d}x + \int_3^4 \exp(-x)\,\mathrm{d}x \\
&= 4\big(-\exp(-1) + 1\big) + 3\big(-\exp(-2) + \exp(-1)\big) + 2\big(-\exp(-3) + \exp(-2)\big) \\
&\quad + \big(-\exp(-4) + \exp(-3)\big) \\
&= 4 - \exp(-1) - \exp(-2) - \exp(-3) - \exp(-4).
\end{aligned}$$

6 Exponential Distributions: Lightbulbs

A brand new lightbulb has just been installed in our classroom, and you know the life span of a lightbulb is exponentially distributed with a mean of 50 days.

(a) Suppose an electrician is scheduled to check on the lightbulb in 30 days and replace it if it is broken. What is the probability that the electrician will find the bulb broken?

(b) Suppose the electrician finds the bulb broken and replaces it with a new one. What is the probablity that the new bulb will last at least 30 days?

(c) Suppose the electrician finds the bulb in working condition and leaves. What is the probability that the bulb will last at least another 30 days?

Solution:

(a) Let $X \sim \text{Exponential}(1/50)$ be the time until the bulb is broken. For an exponential random variable with parameter λ, the density function is $f_X(x) = \lambda\, \mathrm{e}^{-\lambda x}$ for $x > 0$. So in this case $\lambda = 1/50$. Thus we can integrate the density to find the probability that the lightbulb broke in the first 30 days:

$$\mathbb{P}[X < 30] = \int_0^{30} \left(\frac{1}{50} \cdot \mathrm{e}^{-x/50}\right) \mathrm{d}x = 1 - \mathrm{e}^{-30/50} = 1 - \mathrm{e}^{-3/5} \approx 0.451.$$

(b) The new bulb's waiting time Y is i.i.d. with the old bulb's. So the answer is

$$\mathbb{P}[Y > 30] = 1 - \mathbb{P}[Y < 30] = 1 - (1 - \mathrm{e}^{-3/5}) = \mathrm{e}^{-3/5} \approx 0.549.$$

(c) The bulb is memoryless, so the probability it will last 60 days given that it has lasted 30 days, is just the probability it will last 30 days:

$$\mathbb{P}[X > 60 \mid X > 30] = \mathbb{P}[X - 30 > 30 \mid X > 30] = \mathbb{P}[X > 30] = \mathrm{e}^{-3/5} \approx 0.549.$$