# Problem Set 7

Due: Wednesday, November 23, by 11:59pm

## Instructions

**Solutions format, collaboration policy, and late policy.** See PSet 1 for further details. The same requirements and policies still apply. Also follow the typesetting instructions from the prior PSets.

**Solutions submission.** You must submit your solution via Gradescope. In particular:

- Submit a *single* PDF file containing the solution to all tasks in the homework. Each numbered task should be solved on its own page (or pages). Follow the prompt on Gradescope to link tasks to your pages.
- Do not write your name on the individual pages – Gradescope will handle that.
- We encourage you to typeset your solution. The homepage provides links to resources to help you doing so using LaTeX. If you do use another tool (e.g., Microsoft Word), we request that you use a proper equation editor to display math (MS Word has one). For example, you should be able to write $\sum_{i=1}^{n} x^i$ instead of x^1 + x^2 + ... + x^n. You can also provide a handwritten solution, as long as it is on a single PDF file that satisfies the above submission format requirements. It is your responsibility to make sure handwritten solutions are readable – we will *not* grade unreadable write-ups.

## Task 1 – Sticks [20 pts]

Consider a stick of length 1. We break the stick at a position sampled uniformly along its length and throw away the shorter part. We use the random variable $X$ to represent the length of the part we keep and random variable $Z \in [0, 1]$ to denote where we broke it. We do this a second time on the piece of stick we kept, breaking it at a random point along its length, and keeping the longer part. We use the random variable $Y$ to denote the length of the part we keep after the second break.

**a)** Let $x \in \mathbb{R}$. For which values of $Z$ is $X \leqslant x$?

**b)** Compute $F_X(x) = P(X \leqslant x)$ using what you showed in part (a).

**c)** Compute $f_X$ from $F_X$. What distribution from the Zoo is $f_X$?

**d)** Conditioned on $X = x$, what is the distribution of $Y$? (Hint: Everything here is a re-scaling of what went on in the first round. You don't need to rederive everything again.)

**e)** Use the Law of Total Expectation to compute $\mathbb{E}[Y]$.

## Task 2 – Knitting Requires Concentration [20 pts]

Bob is slowly knitting a blanket, made of 100 squares. It takes an average of 1 hour for Bob to knit a square, with a standard deviation of 0.4 hours. The time to knit each square is independent. (You should treat time as continuous for this problem.)

**a)** What is the expectation of the total time to knit the blanket?

**b)** What is the variance of the total time to knit the blanket?

**c)** Bob will have 150 hours to knit between now and when he needs the blanket to be finished to stay warm at a football game. Use Markov's Inequality to give a *lower bound* on the probability that Bob finishes the blanket before the game.

**d)** Can we improve the lower bound from **c)** using Chebyshev's inequality? If so, what bound do you get?

## Task 3 – Chebyshev and Pairs of Events [20 pts]

Chebyshev's inequality can be used to bound the tails of random variables that are sums of random variables that are not independent, provided that pairs of random variables are not too positively correlated with each other. Let random variable $X = \sum_{i=1}^{n} X_i$ be the sum of indicator variables $X_1, \ldots, X_n$ for events $A_1, \ldots, A_n$. Suppose that $\mathbb{E}[X_i] = \mathbb{P}(A_i) = p$ for each $i$ and that for every $i \neq j$, we have $\mathbb{E}[X_i X_j] = \mathbb{P}[A_i \cap A_j] \leqslant p^2$.

**a)** Use linearity of expectation to show that $\mathbb{E}[(X_1 + X_2)^2] \leqslant 2p + 2p^2$.

**b)** Use linearity of expectation and the fact that $(\sum_{i=1}^{n} X_i)^2 = \sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} X_i X_j$ to show that $\mathbb{E}[X^2]$ is at most $np + n(n-1)p^2$. Then compute an upper bound on $\mathrm{Var}(X)$.

**c)** Use Chebyshev's inequality to show that $\mathbb{P}(X \leqslant pn/4) \leqslant \frac{16(1-p)}{9pn}$.

## Task 4 – Chernoff Bound [20 pts]

A certain city is experiencing a terrible city-wide fire. The city decides that it needs to put its firefighters all across the city to ensure that the fire can be put out. The city is conveniently arranged into a $100 \times 100$ grid of city blocks. Each block can be identified by two integers $(a, b)$ where $1 \leqslant a \leqslant 100$ and $1 \leqslant b \leqslant 100$. The city only has $1000$ firefighters, so it decides to send each firefighter to a uniformly random block location, independent of each other (i.e., multiple firefighters can end up at the same location). The city wants to make sure that every $30 \times 30$ sub-grid of blocks (corresponding to grid points $(a, b)$ with $A \leqslant a \leqslant A + 29$ and $B \leqslant b \leqslant B + 29$ for valid $A, B$) gets more than $10$ firefighters (sub-grids can overlap).

**a)** Use the Chernoff bound (in particular, the version presented in class) to compute an upper bound on the probability that a single subgrid gets at most $10$ firefighters.

**b)** Use the fact that the probability of a union of events is at most the sum of their probabilities, together with the result from above to calculate an upper bound on the probability that the city fails to meet its goal.

## Task 5 – Is this Chern-ough? [20 pts]

A quality control engineer wants to estimate the expected fraction of original voltage that is lost after an hour of continuously running a mechanical rabbit for randomly chosen batteries her company produces. Suppose that the true fractional voltage loss for randomly chosen batteries is distributed according to some unknown random variable $B$. She knows for sure that $\mathbb{E}[B] \geqslant 0.01$.

To do this she will run $n$ independent trials and measure the fractional loss $X_i$ of the original voltage at the end of hour for batteries $i = 1, \ldots, n$, compute the sum $X = \sum_{i=1}^{n} X_i$, and output the average loss, $\overline{X} = X/n$ as her estimate of $\mathbb{E}[B]$.

Using only the version of the Chernoff bound that we are have given in CSE 312, how many independent trials $n$ will she need to *guarantee* that with probability 99%, the average fractional loss $\overline{X}$ that she measures is at most twice the true expected loss fraction $\mathbb{E}[B]$?

## Task 6 – Extra Credit: I just want something to happen

Sometimes we have a non-negative integer-valued random variable $Y$ that is at most $n$ and is not always 0 and we want to get a lower bound on the probability that it is bigger than 0, $\mathbb{P}(Y \geqslant 1) = 1 - \mathbb{P}(Y = 0)$. (This often is of interest when $Y$ corresponds to the number of events that happen from some collection of $n$ possibilities.)

**a)** What does Chebyshev's inequality give as a lower bound for this probability in terms of $\mathbb{E}[Y]$ and $\mathbb{E}[Y^2]$? (Simplify your answer as much as possible.)

**b)** In order for the lower bound above to be non-trivial (bigger than 0), how much larger can $\mathbb{E}[Y^2]$ be than $\mathbb{E}[Y]^2$?

**c)** Prove the following bound, which is non-trivial for all values of $\mathbb{E}[Y^2]$ and $\mathbb{E}[Y]$?

$$\mathbb{P}(Y \geqslant 1) \geqslant \frac{\mathbb{E}[Y]^2}{\mathbb{E}[Y^2]}$$

Hint: Try to bound $\mathbb{E}[Y^2] \cdot \mathbb{P}(Y \geqslant 1)$. The Cauchy-Schwarz inequality is helpful here. It states that for real numbers $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ we have

$$\left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right) \geqslant \left(\sum_{i=1}^{n} a_i b_i\right)^2.$$