In-Class Problems — 11

Problem 1. One hundred twenty students take the 6.042 final exam. The mean on the exam is 90 and the lowest score was 30. You have no other information about the students and the exam, e.g., you should not assume that the final is worth 100 points.

(a) State the best possible upper bound on the number of students who scored at least 180.

(b) Give an example set of scores which achieve your bound.

(c) If the maximum score on the exam was 100, give the best possible upper bound on the number of students who scored *at most* 50.

Problem 2. Waiting for Godot A couple plans to have children until they have a boy, whom they'd like to name Godot. What is the expected number of children that they have, and what is the variance?

Problem 3. Suppose you are playing the game "Hearts" with three of your friends. In Hearts, all the cards are dealt to the players, in this case the four of you will each have 13 cards.

(a) What is the expectation and variance of the number of hearts in your hand?

(b) What is the expectation and variance of the number of suits in your hand?

Problem 4. We have two coins: one is a fair coin and the other is a coin that produces heads with probability 3/4. One of the two coins is picked, and this coin is tossed n times.

(a) Does the Weak Law of Large Numbers allow us to *predict* what limit, if any, is approached by the expected proportion of heads that turn up as n approaches infinity? Briefly explain.

(b) How many tosses suffice to make us 95% confident which coin was chosen? Explain.

Problem 5. [5 points]

We want to determine the percentage r of people who intend to watch the final episode of *Seinfeld*. Our polling method is the same as that discussed in lecture: we sample the opinion of n people chosen uniformly and at random, with replacement.

How many people should be polled so that r is within 3 points of the true percentage, with probability at least .96?

Problem 6. Recall than in the Chinese Appetizer Problem R is number of people that get the same appetizer after a spin of the Lazy Susan.

(a) Compute the variance of R.

(b) Show that the upper bound that you get on $\Pr(R \geq N)$ from the strong form of Chebyshev's inequality is tight.

Problem 7. Suppose you flip a fair coin 100 times. The coin flips are all mutually independent. What is an upper bound on the probability that the number of heads is at least 70 . . .

(a) . . . according to Markov's Inequality?

Solution: The expected number of heads of 50. So the probability that the number of heads is at least 70 is at most $50/70 = 0.71$.

(b) . . . according to Chebyshev's Inequality?

Solution: Let X_i be the random variable whose value is 1 if the ith coin flip is heads. Then $\text{Var}[X_i] = 1/2 - (1/2)^2 = 1/4$. So $\text{Var}[X_1 + \cdots + X_100] = 100/4 = 25$. The variance of the number of heads is $100/4 = 25$, so the standard deviation is 5. So 70 is four times the standard deviation from the mean. Since the distribution is symmetric, the probability is at most $\frac{1}{2} \cdot \frac{1}{4^2} = \frac{1}{32}$

(c) . . . according to Chernoff's Bound?

Solution: We apply Chernoff's bound with $c = 70/50 = 1.4$. This gives us that $\alpha = \ln(1.4) + 1/1.4 - 1 = 0.05076$ and that the probability is at most $e^{-0.05076 \cdot 1.4 \cdot 50} = 0.0286$.

Problem 8. Give upper bounds for the following probabilities using the Markov, Chebyshev, or Chernoff inequality, where appropriate. If more than one inequality applies, use the inequality that gives you the best bounds.

(a) Suppose you play the following game: Flip 5 independent and fair coins. Score one point if the difference between the number of heads and tails is at least 3. What is the probability that your average score is greater than 0.9 if you play the game 100 times?

(b) Let a and b be two numbers picked independently and uniformly at random from the set $\{0, 1, 2, 3, 4\}$. For $i = 0, 1, 2, 3, 4$, define the random variable X_i to be equal to $ai + b \bmod p$. Let $Y = \frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)$. What is the probability that Y is greater than 3.5? *Hint*: Observe that X_i is distributed uniformly on the set $\{0, 1, 2, 3, 4\}$.

(c) You are told that the average (mean) annual income is $10,000 in a particular population. Five people are picked independently and uniformly and at random from the population. What is the probability that their avearge (mean) annual income is greater than or equal to $1,000,000?

Problem 9. Psychological diagnosis of sociopaths is notoriously unreliable. For example, a random psychologist is as likely to diagnose a normal patient as sociopathic as he is to recognize that the patient is normal. A random psychologist is actually slightly *less* likely to correctly diagnose a real sociopath than not—sociopaths are really good liars.

(a) In these circumstances it might seem there was no reliable way to diagnose sociopaths, but in theory there is a way—if we have a large enough population of psychologists who reach their judgements independently. Explain how to obtain a reliable diagnosis in this case. Briefly explain why your method will be reliable, citing appropriate results from class. A qualitative answer is all that's required—you need not come up with any numerical estimates or growth rates.

(b) Is it plausible to assume that psychologists will make their diagnoses independently? Briefly explain.

Problem 10. The hat-check staff has had a long day, and at the end of the party they decide to return people's hats at random. Suppose that n people have their hats returned at random. Let $X_i = 1$ if the ith person gets his or her own hat back and 0 otherwise. Let $S_n = \sum_{i=1}^{n} X_i$, so S_n is the total number of people who get their own hat back. Show that

(a) $\mathrm{E}\left[X_i^2\right] = 1/n$.

(b) $\mathrm{E}\left[X_i X_j\right] = 1/n(n-1)$ for $i \neq j$.

(c) $\mathrm{E}\left[S_n^2\right] = 2$. *Hint:* Use (a) and (b).

(d) $\mathrm{Var}\left[S_n\right] = 1$.

(e) Using Chebyshev's Inequality, show that $\Pr(S_n \geq 11) \leq .01$ for any $n \geq 11$.

Problem 11. **(a)** Knowing only that the average graduating MIT student's total number of credits is 200, find a tight upper bound for the fraction of MIT students graduating with at least 235 credits. (Ignore the fact that students who graduate do so after at most 700 years; that is, assume that there is no theoretical upper bound on the possible number of credits a student may earn.)

(b) Assuming in addition only that the standard deviation of the total credits per graduating student is 7, give a tight bound on the fraction of students who can graduate with at least 235 credits.

(c) Different information about MIT credits is now provided by the Registrar. The standard deviation is not 7 after all. Instead, the registrar reports that 30% of past graduates are EECS majors and half of them had 235 or more credits. Only 1/7th of the non-EECS students graduated with that many credits. Use this information and the Chernoff bound to give a tight upper bound on the probability that among 1000 randomly chosen graduating students, at least 300 have 235 or more credits. (You may leave $\ln$ and e in your answer).

Problem 12. Use the lecture notes to compute upper bounds on the following probabilities:

(a) In a poll with 2000 people, what is the probability that the result is off by more than 1%?

(b) In 1000 independent coin flips, what is the probability of getting at least 600 heads?

(c) In a noisy communication channel with 10% error rate, what is the probability that out of 1000 transmitted bits, at least 200 bits are wrongly transmitted (*i.e.*, `1` is received as `0` and vice versa)?

Problem 13. Let X and Y be independent random variables taking on integer values in the range 1 to n uniformly. Compute the following quantities:

(a) $\text{Var}[aX + bY]$

(b) $\text{E}[\max(X, Y)]$

(c) $\text{E}[\min(X, Y)]$

(d) $\text{E}[|X - Y|]$

(e) $\text{Var}[|X - Y|]$.

Problem 14. Let X be a random variable whose value is an observation drawn uniformly at random from the set $\{-n, -n+1, \ldots, -2, -1, 0, 1, 2, \ldots, n-1, n\}$. Let $Y = X^2$. Then which of the following are true:

(a) $\text{E}[X] = 0$

(b) $\text{E}[Y] = 0$

(c) $\text{E}[Y] > \text{E}[X]$

(d) $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$

(e) $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$

(f) X, Y are independent variables.

(g) $\mathrm{Var}[X] = 0$

(h) $\mathrm{Var}[Y] = 0$

(i) $\mathrm{Var}[Y] > \mathrm{Var}[X]$

(j) $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$

(k) $\mathrm{Var}[XY] = \mathrm{Var}[X]\,\mathrm{Var}[Y]$

Problem 15. A man has a set of n keys, one of which fits the door to his apartment. He tries the keys until he finds the correct one. Give the expected number and variance for the number of trials until success if

(a) he tries the keys at random (possibly repeating a key tried earlier)

(b) he chooses keys randomly from among those he has not yet tried.

Problem 16. An Unbiased Estimator

Suppose we are trying to estimate some physical parameter p. When we run our experiments and process the results, we obtain an estimator of p, call it p_e. But if our experiments are probabilistic, then p_e itself is a random variable which has a pdf over some range of values. We call the random variable p_e an *unbiased* estimator if $\mathrm{E}[p_e] = p$.

For example, say we are trying to estimate the height, h, of Green Hall. However, each of our measurements has some noise that is, say, Gaussian with zero mean. So each measurement can be viewed as a sample from a random variable X. The expected value of each measurement is thus $\mathrm{E}[X] = h$, since the probabilistic noise has zero mean. Then, given n independent trials, $x_1, ..., x_n$, an unbiased estimator for the height of Green Hall would be

$$h_e = \frac{x_1 + ... + x_n}{n},$$

since

$$\mathrm{E}[h_e] = \mathrm{E}\left[\frac{x_1 + ... + x_n}{n}\right] = \frac{\mathrm{E}[x_1] + ... + \mathrm{E}[x_n]}{n} = \mathrm{E}[x_1] = h.$$

Now say we take n independent observations of a random variable Y. Let the true (but unknown) variance of Y be $\mathrm{Var}[Y] = \sigma^2$. Then (see section 6.4 in the notes), we can define the following estimator σ_e^2 for $\mathrm{Var}[Y]$ using the data from our observations:

$$\sigma_e^2 = \frac{y_1^2 + y_2^2 + \ldots + y_n^2}{n} - \left(\frac{y_1 + y_2 + \ldots + y_n}{n}\right)^2.$$

Is this an unbiased estimator of the variance? In other words, is $\mathrm{E}\left[\sigma_e^2\right] = \sigma^2$? If not, can you suggest how to modify this estimator to make it unbiased?

Problem 17. Suppose we have a calculation that will require n operations to complete, and our computer has a mean time to failure of f operations where the failure process is memoryless. When the computer fails, it loses all state, so the calculation has to be restarted from the beginning.

(a) Louis Reasoner decides to just run the calculation over and over until it eventually completes without failing. If the calculation ever fails, Louis restarts the entire calculation. Give a lower bound on the time that Louis can expect to wait for his code to complete successfully on a typical gigahertz processor (10^{12} operations per second), with $n = 10^{13}$ and $f = 10^{12}$?

Hint: $e^{-(p-p^2)n} \leq (1-p)^n \leq e^{-pn}$ for all $p \in [0, 1/\sqrt{2}]$.

(b) Alyssa P. Hacker decides to divide Louis's calculation into 10 equal-length parts, and has each part save its state when it completes successfully. Saving state takes time s. When a failure occurs, Alyssa restarts the program from the latest saved state. How long can Alyssa expect to wait for her code to complete successfully on Louis's system? You can assume that $s < 10^{-4}$ seconds.

(c) Alyssa tries to further optimize the expected total computing time by dividing the calculation into 10 000 parts instead of 10. How long can Alyssa expect to wait for her code to complete successfully?

Hint: $1 + np \leq (1-p)^{-n} < 1 + np + 2n^2p^2$ for all $np \in [0, 1], n > 3$.

Problem 18. The spider (remember her from the Tutor problem) is expecting guests and wants to catch 500 flies for her dinner. *Exactly* 100 flies pass by her web every hour. Exactly 60 of these flies are quite small and are caught with probability 1/6 each. Exactly 40 of the flies are big and are caught with probability 3/4 each. Assume all fly interceptions are mutually independent. Using this information, the methods from lecture can show that the poor spider has only about 1 chance in 100,000 of catching 500 flies within 10 hours.

Ben Bitdiddle knows he can get the best estimate using the approximations to the binomial distribution developed in Notes 12. He reasons that since 60% of the flies are small and 40% are large, the probability that a random fly will be caught is $0.6(1/6) + 0.4(3/4) = 0.4$, so he will use the approximation for the binomial cumulative distribution function, $F_{1000,0.4}$, to bound the probability that the spider catches at least 500 flies in 10 hours.

As usual, Ben hasn't got it quite right.

(a) According to Ben's reasoning, what is the probability that the spider will catch all 1000 flies that show up during the 10 hours? Show that this is not equal to the actual probability.

(b) How would you explain to Ben what is wrong with his reasoning?

(c) What would the Markov bound be on the probability that the spider will catch her quota of 500 flies?

(d) What would the Chebyshev bound be on the probability that the spider will catch her quota of 500 flies?

(e) What would the Chernoff bound be on the probability that the spider will catch her quota of 500 flies? (You can do this without a calculator knowing that $\ln 5/4 \approx 0.223$, $e^3 \approx 20$ and $\sqrt{e} \approx 1.6$.)

(f) Ben argues that he made his mistake because the description of the spider's situation is absurd: knowing the *expected* number of flies per hour is one thing, but knowing the *exact* number is far-fetched.

Which of the bounds above will hold if all we know is that the *expected* number of small flies caught per hour is 10 and of large flies is 30?

(g) Ben argues that we should model the spider's situation by assuming that the captures of large and small flies are independent Poisson processes with respective rates of 100 small flies captured per 10 hours and 300 large ones per 10 hours. Under these assumptions, what are the Markov, Chebyshev, and Chernoff bounds on the spider's probability of meeting her quota?

Problem 19. Let R be the sum of a finite number of Bernoulli variables.

(a) For $y \geq \mu_R$, write formulas in terms of y, μ_R and σ_R for the Markov, one-sided Chebyshev, and Chernoff bounds on $\Pr\{R \geq y\}$.

(b) Compare these bounds when R is a single unbiased Bernoulli variable and $y = 1$.

Consider the case where c is small, say $c = 1+\epsilon$ and $\epsilon < 1$. The Markov bound is $\frac{1}{1+\epsilon}$. If we further assume that $\text{Var}[R] = \text{E}[R]$, then the Chebyshev one-sided bound becomes

$$\Pr\{R - \text{E}[R] \geq \epsilon\,\text{E}[R]\} \leq \frac{\text{Var}[R]}{\text{Var}[R] + \epsilon^2\,\text{E}[R]} = \frac{1}{1+\epsilon^2}.$$

Since $\frac{1}{1+\epsilon} < \frac{1}{1+\epsilon^2}$, the Markov bound is tighter than the Chebyshev one-sided bound.

(c) Discuss when the Chebyshev one-sided bound on $\Pr\{R > c\,\text{E}[R]\}$ is tighter than the Chernoff bound, where R is positive and $c > 1$. (Providing an example is sufficient.)

Problem 20. Central Limit Theorem. Let B_n be a random variable with binomial distribution $f_{n,p}$.

(a) Write a formula defining the random variable, B_n^*, which is the "normalized" version of B_n (normalized to have mean 0 and standard deviation 1).

(b) Explain why

$$\lim_{n\to\infty} \Pr\{B_n^* < \beta\} = N(\beta),$$

where $N(\beta)$ is the normal distribution from the Central Limit Theorem (see Appendix).

(c) Suppose you flip a fair coin 100 times. The coin flips are all mutually independent. Use the standard normal distribution to approximate the probability that the number of heads is between 30 and 70. Can you give a lower bound on this probability using the generalized Chernoff bound? (You can leave $N(\beta)$ and e in your answer.)

Problem 21. Generalized Law of Large Numbers.

The Weak Law of Large Numbers in Notes 13 was given for a sequence $G_1, G_2, \ldots$ of mutually independent random variables with the same mean and variance. We can generalize the Law to sequences of mutually independent random variables, possibly with different means and variances, as long as their variances are bounded by some constant. Namely,

Theorem. *Let $X_1, X_2, \ldots$ be a sequence of mutually independent random variables such that* $\mathrm{Var}[X_i] \le b$ *for some $b \ge 0$ and all $i \ge 1$. Let*

$$\overline{X}_n ::= (X_1 + X_2 + \cdots + X_n)/n, \qquad \mu_n ::= \mathrm{E}\left[\overline{X}_n\right].$$

Then for every $\epsilon > 0$,

$$\Pr\left\{\left|\overline{X}_n - \mu_n\right| \ge \epsilon\right\} \le \frac{b}{n\epsilon^2}.$$

(a) Prove this Theorem.

Hint: Adapt the proof (attached) of the Weak Law of Large Numbers from Notes 13.

(b) Conclude

Corollary (Weak Law of Large Numbers). *For every $\epsilon > 0$,*

$$\lim_{n\to\infty} \Pr\left\{\left|\overline{X}_n - \mu_n\right| \ge \epsilon\right\} = 0.$$

Problem 22. One-sided Chebyshev Bound. Prove the one-sided Chebyshev bound. *Hint:* $R - \mu_R \ge x$ implies $(R - \mu_R + a)^2 \ge (x + a)^2$ for all nonnegative $a \in \mathbb{R}$. By Markov,

$$\Pr\left\{(R - \mu_R + a)^2 \ge (x + a)^2\right\} \le \frac{\mathrm{E}\left[(R - \mu_R + a)^2\right]}{(x + a)^2}.$$

Choose a to minimize this last bound.

Problem 23. Let $R : S \to \mathbb{N}$ be a random variable.

(a) Let $R' = cR$ for some positive constant c. If $\sigma_R = s$, what is $\sigma_{R'}$?

(b) If $\mathrm{E}[R] = 1$, how large can $\mathrm{Var}[R]$ be?

(c) If R is always positive (nonzero), how large can $\mathrm{E}[1/R]$ be?

A Events

$$
\begin{aligned}
\Pr\left\{\bigcup_{n\in\mathbb{N}} A_n\right\} &= \sum_{n\in\mathbb{N}} \Pr\{A_n\} \text{ for pairwise disjoint } A_n \\
\Pr\left\{\overline{B}\right\} &= 1 - \Pr\{B\} \\
\Pr\{A \cup B\} &= \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\} \\
\Pr\{A \cup B\} &\le \Pr\{A\} + \Pr\{B\} \quad \text{[Boole's inequality]} \\
\Pr\{A\} &\le \Pr\{A \cup B\} \quad \text{[monotonicity]} \\
\Pr\{A \mid B\} &= \frac{\Pr\{A \cap B\}}{\Pr\{B\}}
\end{aligned}
$$

B Law of Total Probability

Let $B_0, B_1, \ldots$ be disjoint events whose union is the entire sample space. Then for all events A,

$$\Pr\{A\} = \sum_{i\in\mathbb{N}} \Pr\{A \cap B_i\}.$$

C Independence

Definition. Events A and B are independent iff

$$\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\}.$$

Events $A_0, A_1, A_2, \ldots$ are *mutually independent* iff for all subsets $J \subset \mathbb{N}$,

$$\Pr\left\{\bigcap_{i\in J} A_i\right\} = \prod_{i\in J} \Pr\{A_i\}.$$