Massachusetts Institute of Technology
6.042J/18.062J, Spring '02: Mathematics for Computer Science
Professor Albert Meyer and Dr. Radhika Nagpal
revised April 27, 2002, 997 minutes

Solutions to 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.

Solution. Let R be the score of a student chosen at random. Apply Markov's Bound to $R - 30$:

$$\Pr(R \geq 180) = \Pr(R - 30 \geq 150) \leq \frac{\mathrm{E}\left[R - 30\right]}{150} = \frac{60}{150} = \frac{2}{5}.$$

So at most $\frac{2}{5} \cdot 120 = 48$ students scored greater than or equal to 180. ■

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

Solution. Of the 120 students, 48 score 180 and 72 score 30. The mean is $\frac{48 \cdot 180 + 72 \cdot 30}{120} = 90$, as required. ■

(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.

Solution. Apply Markov's Bound to $100 - R$:

$$\Pr(R \leq 50) = \Pr(100 - R \geq 50) \leq \frac{\mathrm{E}\left[100 - R\right]}{50} = \frac{10}{50} = \frac{1}{5}.$$

So at most $\frac{1}{5} \cdot 120 = 24$ students scored 50 or less. ■

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?

Solution. The probability of having a single child (a boy) is $\frac{1}{2}$, the probability of having two children (a girl and then a boy) is $\frac{1}{4}$, and so forth; the probability of having k children is $\frac{1}{2^k}$. Therefore, the expected number of children is:

$$\begin{aligned} \mathrm{E}[\text{children}] &= \sum_{k=1}^{\infty} k \cdot \frac{1}{2^k} \\ &= 2 \end{aligned}$$

Now, let $C = 2$ be the expected number of children up to and including the first boy. We can compute the variance as follows:

$$\mathrm{Var}[C] = \mathrm{E}\left[C^2\right] - \mathrm{E}[C]^2 = \sum_{k=1}^{\infty} k^2 \cdot \left(\frac{1}{2}\right)^k - 2^2 = \frac{\frac{1}{2} + (\frac{1}{2})^2}{(1 - \frac{1}{2})^3} - 4 = 6 - 4 = 2$$

The sum is computed by differentiating the formula for the sum of an infinite geometric sequence.
■

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?

Solution. Let H be the number of hearts in your hand, and let X_i be the indicator random variable for the event that the ith card in your hand is a heart. Then $H = \sum_{i=1}^{13} X_i$. So

$$\mathrm{E}[H] = \sum_{i=1}^{13} \mathrm{E}[X_i] = \sum_{i=1}^{13} \Pr(i\text{th card is a heart}) = \sum_{i=1}^{13} \frac{1}{4} = \frac{13}{4}.$$

To compute variance, we first compute $\mathrm{E}\left[H^2\right]$, following the solution to Problem 4c.

$$\mathrm{E}\left[H^2\right] = \mathrm{E}\left[\left(\sum_{i=1}^{13} X_i\right)^2\right] = \sum_i \mathrm{E}[X_i] + \sum_i \sum_{j \neq i} \mathrm{E}[X_i X_j]$$

Note that $X_i^2 = X_i$, so $\mathrm{E}\left[X_i^2\right] = 1/4$. When $i \neq j$, $X_i X_j = 1$ if cards i and j are both hearts, and 0 otherwise. The probability that both cards are hearts is $\binom{13}{2}/\binom{52}{2}$, so $\mathrm{E}[X_i X_j] = \binom{13}{2}/\binom{52}{2} = 3/51$. In the summation above, there are 13 $\mathrm{E}\left[X_i^2\right]$ terms, and $13 \cdot 12 = 156$ terms of the form $\mathrm{E}[X_i X_j]$ with $i \neq j$. Thus,

$$\mathrm{Var}[H] = \mathrm{E}\left[H^2\right] - \mathrm{E}[H]^2 = 13 \cdot \frac{1}{4} + 156 \cdot \frac{3}{51} - (\frac{13}{4})^2 \approx 1.864.$$

■

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

Solution. Let N denote the number of suits in a hand. Let X_s be the indicator random variable for the event that there is a spade in the hand. Define X_h, X_d, X_c analogously for the three remaining suits: hearts, diamonds, and clubs. Then $N = X_s + X_h + X_d + X_c$.

The expectation of each of the indicator variables is the probability that a hand contains the corresponding suit. Thus, for $i \in \{s, h, d, c\}$

$$\mathrm{E}[X_i] = 1 - \frac{\binom{52-13}{13}}{\binom{52}{13}} \approx 0.9872,$$

because $\binom{52-13}{13}$ is the number of hands that are missing the particular suit, and $\binom{52}{13}$ is the total number of hands.

$$\mathrm{E}[N] = \mathrm{E}[X_s + X_h + X_d + X_c] = \mathrm{E}[X_s] + \mathrm{E}[X_h] + \mathrm{E}[X_d] + \mathrm{E}[X_c] \approx 4 \cdot 0.9872 = 3.9488.$$

To compute variance, we compute $\mathrm{E}\left[N^2\right]$ directly, and then $\mathrm{Var}[N] = \mathrm{E}\left[N^2\right] - \mathrm{E}[N]^2$. We first calculate $\Pr(N = i)$ for $i = 1, 2, 3, 4$.

$$\Pr(N = 1) = \frac{4}{\binom{52}{13}} \approx 6.299 \times 10^{-12},$$

because there are four possible hands of all one suit out of $\binom{52}{13}$ possible hands of 13 cards.

$$\Pr(N = 2) = \frac{\binom{4}{2}\left[\binom{26}{13} - 2\right]}{\binom{52}{13}} \approx 9.827 \times 10^{-5},$$

because there are $\binom{4}{2}$ ways to choose which two suits. There are $\binom{26}{13}$ ways to choose 13 cards from the 26 possible cards of those two suits. Two of those $\binom{26}{13}$ hands actually contain only one suit.

$$\Pr(N = 3) = \frac{\binom{4}{3}\left[\binom{39}{13} - \binom{3}{2}\binom{26}{13} + 3\right]}{\binom{52}{13}} \approx 0.05097,$$

because there are $\binom{4}{3}$ ways to choose the three suits, and $\binom{39}{13}$ ways to choose 13 cards from the 39 cards of those three suits. But that includes hands with only one or two suits. So, using Inclusion-Exclusion, we subtract the $\binom{3}{2}\binom{26}{13}$ ways to choose two of the three suits and then choose 13 cards from the 26 cards of those suits. This subtracts out the three hands with only one suit twice, so we add these back.

Of course,

$$\Pr(N = 4) = 1 - (\Pr(N = 1) + \Pr(N = 2) + \Pr(N = 3)) \approx 0.9489.$$

Thus,

$$\begin{aligned} \mathrm{E}\left[N^2\right] &= 1 \cdot \Pr(N = 1) + 4 \cdot \Pr(N = 2) + 9 \cdot \Pr(N = 3) + 16 \cdot \Pr(N = 4) \\ &\approx 0 + 0.0004 + 0.4587 + 15.1824 = 15.6415 \end{aligned}$$

This allows us to complete the calculation of the variance:

$$\mathrm{Var}\left[N\right] \approx 15.6415 - (3.9488)^2 \approx 0.0485.$$

■

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.

Solution. The Weak Law of Large Numbers tells us that the proportion of heads will approach $1/2$ if the fair coin was picked, and it will approach $3/4$ if the other coin was picked. But it does not tell us anything about which of these two numbers it will approach, as we have no information about which coin is picked. ■

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

Solution. To guess which coin was picked, set a threshold t between $1/2$ and $3/4$. If the proportion of heads is less than the threshold, guess it was the fair coin; otherwise, guess the biased coin. Let the random variable H_n be the number of heads in the first n flips. We need to flip the coin enough times so that $\Pr(H_n/n > t) \leq 0.05$ if the fair coin was picked, and $\Pr(H_n/n < t) \leq 0.05$ if the biased coin was picked. A natural threshold to choose is $5/8$, exactly in the middle of $1/2$ and $3/4$.

H_n is the sum of independent Bernoulli variables, which each have variance $1/4$ for the fair coin and $3/16$ for the biased coin. Using Chebyshev's Inequality for the fair coin,

$$\begin{aligned}
\Pr\left(\frac{H_n}{n} > \frac{5}{8}\right) &= \Pr\left(\frac{H_n}{n} - \frac{1}{2} > \frac{5}{8} - \frac{1}{2}\right) = \Pr\left(H_n - \frac{n}{2} > \frac{n}{8}\right) \\
&= \Pr\left(H_n - \mathrm{E}[H_n] > \frac{n}{8}\right) \leq \Pr\left(|H_n - \mathrm{E}[H_n]| > \frac{n}{8}\right) \\
&\leq \frac{\text{Var}[H_n]}{(n/8)^2} = \frac{n/4}{n^2/64} = \frac{16}{n}
\end{aligned}$$

For the biased coin, we have

$$\begin{aligned}
\Pr\left(\frac{H_n}{n} < \frac{5}{8}\right) &= \Pr\left(\frac{3}{4} - \frac{H_n}{n} > \frac{3}{4} - \frac{5}{8}\right) = \Pr\left(\frac{3n}{4} - H_n > \frac{n}{8}\right) \\
&= \Pr\left(\mathrm{E}[H_n] - H_n > \frac{n}{8}\right) \leq \Pr\left(|H_n - \mathrm{E}[H_n]| > \frac{n}{8}\right) \\
&\leq \frac{\text{Var}[H_n]}{(n/8)^2} = \frac{3n/16}{n^2/64} = \frac{12}{n}
\end{aligned}$$

We are 95% confident if these are at most 0.05, which is satisfied if $n \geq 320$.

Because the variance of the biased coin is less that of the fair coin, we can do slightly better if make our threshold a bit bigger, to about 0.634, which gives 95% confidence with 279 coin flips.

Because H_n has a binomial distribution, we can get a much better bound using the estimates from Lecture 21, giving 95% confidence when $n > 42$. ■

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?

Solution. Solution

$\epsilon = .03, \delta = .04$

$$\begin{aligned}
\delta \geq 2F_{1/2}((1/2-\epsilon)n) \quad & \Rightarrow \quad .05 \geq 2F_{1/2}(0.49n) \\
& \Rightarrow \quad 0.025 \geq \frac{0.51}{0.02} f_{1/2}(0.49n) \\
& \Rightarrow \quad 0.025 \geq \frac{25.5}{\sqrt{2n\pi(.49)(.51)}} 2^{(H(.49)-1)n} \\
& \Rightarrow \quad 0.0012285 \geq \frac{2^{(H(.49)-1)n}}{\sqrt{n}} \\
& \Rightarrow \quad 0.0012285 \geq \frac{2^{-0.000288574n}}{\sqrt{n}} \\
& \Rightarrow \quad 0.0012285 \geq \frac{(.9998)^n}{\sqrt{n}}
\end{aligned}$$

Choosing $n = 10388$ will do it. ■

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 $\text{Var}[X_1 + \cdots + X_1 00] = 100/4 = 25$. 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.

Solution. Get a reliable test by repeatedly asking randomly chosen psychologists to do the diagnosis and computing the fraction, f, of "normal" diagnoses. Let $s > 1/2$ be the probability of a "normal" diagnosis on a sociopathic patient. Diagnose normal if f is closer to $1/2$ than to s.

By diagnosing independently enough times, there is small probability that f differs from the expected value of a diagnosis by very much (the expected value is either $1/2$ or s depending on whether or not the patient is normal—we don't know which). By the Law of Large Numbers, by Chebyshev's bound, or by the estimation of the binomial distribution as used in the Gallup Poll example in Lecture Notes, if we diagnose enough times, we can make this probability as small as desired, so the probability of a correct diagnosis can be as close to 1 as desired.

■

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

Solution. Not really. It figures that some sociopaths would be particularly good liars and be able to fool most psychologists. Conversely, there figure to be normal people whose anxious or defensive manner makes a sociopathic diagnosis by many psychologists more likely. Also, there may be some specially expert psychologists whose diagnoses are more reliable than average; this implies if one expert diagnoses "normal," the diagnosis is likely to be correct, making it more likely that another expert will also correctly diagnose "normal," so their diagnoses will not be independent of each other. ■

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

Solution. $X_i = 1$ with probability $1/n$ and 0 otherwise. Thus $X_i^2 = 1$ with probability $1/n$ and 0 otherwise. So $\mathrm{E}\left[X_i^2\right] = 1/n$.

■

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

Solution. The probability that X_i and X_j are both 1 is $1/n \cdot 1/(n-1) = 1/n(n-1)$. Thus $X_i X_j = 1$ with probability $1/n(n-1)$, and is zero otherwise. So $\mathrm{E}\left[X_i X_j\right] = 1/n(n-1)$.

■

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

Solution.

$$\begin{aligned} \mathrm{E}\left[S_n^2\right] &= \sum_i \mathrm{E}\left[X_i^2\right] + \sum_i \sum_{j \neq i} \mathrm{E}\left[X_i X_j\right] \\ &= n \cdot \frac{1}{n} + n(n-1) \cdot \frac{1}{n(n-1)} \\ &= 2. \end{aligned}$$

■

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

Solution. Solution:

$$\begin{aligned} \mathrm{Var}\left[S_n\right] &= \mathrm{E}\left[S_n^2\right] - \mathrm{E}^2\left[S_n\right] \\ &= 2 - (n(1/n))^2 \\ &= 2 - 1 \\ &= 1. \end{aligned}$$

■

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

Solution. Solution:

$$\begin{aligned}
\Pr(S_n \geq 11) &= \Pr(S_n - \mathrm{E}[S_n] \geq 11 - \mathrm{E}[S_n]) \\
&= \Pr(S_n - \mathrm{E}[S_n] \geq 10) \\
&\leq \frac{\mathrm{Var}[S_n]}{10^2} = .01
\end{aligned}$$

Note that the X_i's are Bernoulli variables but are *not* independent, so S_n does not have a Bernoulli distribution and the Bernoulli estimates from Lecture Notes 21 do not apply.

■

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.)

Solution. Solution:

We use Markov's inequality $\Pr(X \geq c) \leq E[X]/c$, with $c = 235$, to obtain the upper bound $200/235 \approx .85$. ■

(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.

Solution. Solution:

We use the one-sided Chebyshev inequality $\Pr(X \geq E[X] + c\sigma_X) \leq \frac{1}{1+c^2}$ with $c = 5$ to get an upper bound of $1/26$. ■

(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 $1/7$th 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).

Solution. Solution:

We use the Chernoff bound. Among 1000 students, let N be the number with at least 235 credits. So $\mathrm{E}[N] = (1/7)700 + (1/2)300 = 250$ and

$$\begin{aligned}\Pr\{N \geq 300\} &= \Pr\{X \geq (1.2)250\} \leq e^{-(\ln 1.2 + 1/(1.2) - 1)(1.2)250} = e^{-(\ln 1.2 - .1666)300} \\ &= e^{-(.182 - .1666)300} \leq e^{-4.6} \approx .01.\end{aligned}$$

■

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%?

Solution. $\Pr\{\geq 0.01\} = 1 - \Pr\{< 0.01\} \leq 1 - \left(\frac{1-0.01}{1-0.02}\right) 2^{-2000} \binom{2000}{0.01 \times 2000}$ ■

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

Solution. $\Pr\{\geq 600\} = 1 - \Pr\{< 600\}\ (n = 1000, \alpha = 3/5) \leq 1 - 2^{-1000} \binom{1000}{600} \frac{1-3/5}{1-6/5}$ ■

(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)?

Solution. g$\Pr\{\geq 200\} = 1 - \Pr\{< 200\}\ (n = 1000, \alpha = 1/5) \leq 1 - 2^{-1000} \binom{1000}{200} \frac{1-1/5}{1-2/5}$ ■

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) $\text{Var}\left[aX + bY\right]$

Solution. First, we compute $\text{Var}\left[X\right]$, which is also $\text{Var}\left[Y\right]$ since X and Y are identically distributed.

$$\begin{aligned}\text{Var}\left[X\right] = \text{E}\left[X^2\right] - \text{E}\left[X\right]^2 &= \sum_{i=1}^{n} i^2 \cdot \frac{1}{n} - \left(\sum_{i=1}^{n} i \cdot \frac{1}{n}\right)^2 \\ &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}\end{aligned}$$

We now use this to compute $\text{Var}\left[aX + bY\right]$:

$$\begin{aligned}\text{Var}\left[aX + bY\right] &= \text{Var}\left[aX\right] + \text{Var}\left[bY\right] && \text{because } aX \text{ and } bY \text{ are independent} \\ &= a^2\,\text{Var}\left[X\right] + b^2\,\text{Var}\left[Y\right] && \text{Theorem 4.4 in Lecture Notes 24} \\ &= (a^2 + b^2)\,\text{Var}\left[X\right] && \text{because } \text{Var}\left[X\right] = \text{Var}\left[Y\right] \\ &= (a^2 + b^2)\frac{n^2-1}{12} && \text{substitution from computation above}\end{aligned}$$

■

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

Solution. First we compute the probability that the maximum of X and Y is equal to i:

$$\begin{aligned}\Pr(\max(X,Y) = i) &= \Pr(X \le i \wedge Y \le i) - \Pr(X < i \wedge Y < i) \\ &= \Pr(X \le i)\Pr(Y \le i) - \Pr(X < i)\Pr(Y < i) \\ &= \frac{i}{n} \cdot \frac{i}{n} - \frac{i-1}{n} \cdot \frac{i-1}{n} \\ &= \frac{2i-1}{n^2}\end{aligned}$$

We now compute the expectation using these probabilities:

$$\begin{aligned}\text{E}\left[\max(X,Y)\right] &= \sum_{i=i}^{n} i \cdot \Pr(\max(X,Y) = i) = \sum_{i=i}^{n} i \cdot \frac{2i-1}{n^2} \\ &= \frac{1}{n^2} \cdot \sum_{i=1}^{n} (2i^2 - i) = \frac{1}{n^2} \cdot \left(2 \cdot \frac{n(2n+1)(n+1)}{6} - \frac{n(n+1)}{2}\right) \\ &= \frac{(n+1)(4n-1)}{6n}\end{aligned}$$

Alternatively, we could also compute it as follows:

$$\begin{aligned}
\mathrm{E}[\max(X,Y)] &= \sum_{i=0}^{\infty} \Pr(\max(X,Y) > i) = \sum_{i=0}^{n-1} (1 - \Pr(\max(X,Y) \le i)) \\
&= n - \sum_{i=0}^{n-1} \Pr(X \le i \wedge Y \le i) = n - \sum_{i=0}^{n-1} \Pr(X \le i) \Pr(Y \le i) \\
&= n - \sum_{i=0}^{n-1} \left(\frac{i}{n}\right)^2 = n - \frac{1}{n^2} \cdot \frac{n(n-1)(2n-1)}{6} \\
&= \frac{(n+1)(4n-1)}{6n}
\end{aligned}$$

■

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

Solution. Note that $\max(X,Y) + \min(X,Y) = X + Y$. Thus,

$$\begin{aligned}
\mathrm{E}[\min(X,Y)] &= \mathrm{E}[X] + \mathrm{E}[Y] - \mathrm{E}[\max(X,Y)] \\
&= \frac{n+1}{2} + \frac{n+1}{2} - \frac{(n+1)(4n-1)}{6n} \\
&= \frac{(n+1)(2n+1)}{6n}
\end{aligned}$$

■

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

Solution. We could compute this directly. However, note that $|X-Y| = \max(X,Y) - \min(X,Y)$. Thus,

$$\begin{aligned}
\mathrm{E}[|X-Y|] &= \mathrm{E}[\max(X,Y) - \min(X,Y)] \\
&= \mathrm{E}[\max(X,Y)] - \mathrm{E}[\min(X,Y)] \\
&= \frac{(n+1)(4n-1)}{6n} - \frac{(n+1)(2n+1)}{6n} \\
&= \frac{n^2-1}{3n}
\end{aligned}$$

■

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

Solution. We start with one of the standard formulas for variance:

$$\mathrm{Var}[|X-Y|] = \mathrm{E}\left[|X-Y|^2\right] - \mathrm{E}[|X-Y|]^2 = \mathrm{E}\left[(X-Y)^2\right] - \mathrm{E}[|X-Y|]^2$$

We could compute $\mathrm{E}\left[(X-Y)^2\right]$ directly, but note that $\mathrm{Var}\left[X-Y\right] = \mathrm{E}\left[(X-Y)^2\right] - \mathrm{E}\left[X-Y\right]^2$, and $\mathrm{E}\left[X-Y\right] = 0$. We can compute $\mathrm{Var}\left[X-Y\right]$ using the result in part (a). Thus,

$$\begin{aligned}
\mathrm{Var}\left[|X-Y|\right] &= \mathrm{Var}\left[X-Y\right] - \mathrm{E}\left[|X-Y|\right]^2 \\
&= (1^2 + (-1)^2) \cdot \frac{n^2-1}{12} - \left(\frac{n^2-1}{3n}\right)^2 \\
&= \frac{(n^2-1)(n^2+2)}{18n^2}
\end{aligned}$$

■

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) $\mathrm{E}[X] = 0$

Solution. True ■

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

Solution. False ■

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

Solution. True ■

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

Solution. True ■

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

Solution. True ■

(f) X, Y are independent variables.

Solution. False ■

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

Solution. False ■

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

Solution. False ■

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

Solution. True ■

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

Solution. False ■

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

Solution. False ■

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)

Solution. This problem can be modeled by the Polya Urn Scheme: Consider an urn with $n-1$ wrong keys and 1 right key. Each time the man chooses a key, he tries it and then places it back in the urn. We want to compute the waiting time, T, to pick the right key.

The probability of picking the right key on the first trial is $1/n$, and the probability of picking the wrong key is $(n-1)/n$. The probability stays the same on all subsequent trials since the picked key is always returned to the urn. Thus

$$P(T=k) = \left(\frac{n-1}{n}\right)^{k-1}\frac{1}{n}.$$

Note: this is the same as the waiting time for the first head in the Bernouli process with a biased coin. Let $p = 1/n$ and let $q = 1-p$. Then, the expectation of T is

$$E(T) = \sum_{k=1}^{\infty} kq^{k-1}p = p\sum_{k=1}^{\infty} kq^{k-1} = p\frac{1}{(1-q)^2} = \frac{1}{p} = n$$

The variance of T can be computed directly from the formula,

$$\text{Var}[T] = E(T^2) - E^2(T).$$

Again, by definition we have

$$E(T^2) = \sum_{k=1}^{\infty} k^2pq^{k-1}.$$

How do we evaluate this sum? We use the following trick. We know that

$$\sum_{k=1}^{\infty} kq^k = q\sum_{k=1}^{\infty} kq^{k-1} = \frac{q}{(1-q)^2}.$$

Moreover,

$$\begin{aligned}
\sum_{k=1}^{\infty} k^2q^{k-1} &= \sum_{k=1}^{\infty}\frac{d}{dq}kq^k \\
&= \frac{d}{dq}\sum_{k=1}^{\infty} kq^k \\
&= \frac{d}{dq}\frac{q}{(1-q)^2} \\
&= \frac{1}{(1-q)^2} + \frac{2q}{(1-q)^3} = \frac{1}{p^2} + \frac{2q}{p^3}.
\end{aligned}$$

Finally, the expectation $E(T^2)$ is

$$E(T^2) = p\left(\frac{1}{p^2} + \frac{2q}{p^3}\right) = \frac{1}{p} + \frac{2q}{p^2},$$

and the variance is

$$\text{Var}[T] = \frac{1}{p} + \frac{2q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} = n(n-1).$$

■

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

Solution. Again we look at this as a Polya Urn Process, but without replacement. The event that the waiting time for the right key, T, is k is the event that we pick the wrong key on the first trial, and we pick the wrong key on the second trial, etc, and we pick the right key on the k-th trial. Let K_i be the indicator random variable for the i-th trial, i.e. $K_i = 1$ if we pick the right key on the i-th trial, and 0 otherwise. Then

$$P(T = k) = P((K_1 = 0)\&(K_2 = 0)\&\cdots\&(K_{k-1} = 0)\&(K_k = 1)).$$

By the Multiplication Theorem we can compute

$$\begin{aligned} P(T = k) &= P(K_1 = 0)P(K_2 = 0|K_1 = 0)P(K_3 = 0|K_1 = 0, K_2 = 0)\cdots \\ &\qquad \cdots P(K_k = 1|K_1 = 0, \ldots, K_{k-1} = 0) \\ &= \frac{n-1}{n}\,\frac{n-2}{n-1}\,\frac{n-3}{n-2}\cdots\frac{n-k+1}{n-k+2}\,\frac{1}{n-k+1} \\ &= \frac{1}{n}. \end{aligned}$$

The expectation and variance are now easy to compute from the definitions.

$$E(T) = \sum_{k=1}^{n} kP(T = k) = \frac{1}{n}\sum_{k=1}^{n} k = \frac{n+1}{2}.$$

$$\begin{aligned} \text{Var}[T] &= E(T^2) - E^2(T) \\ &= \sum_{k=1}^{n} k^2 P(T = k) - \left(\frac{n+1}{2}\right)^2 \\ &= \frac{1}{n}\sum_{k=1}^{n} k^2 - \left(\frac{n+1}{2}\right)^2 \\ &= \frac{1}{n}\,\frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\ &= \frac{n^2-1}{12}. \end{aligned}$$

■

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?

Solution. Let $\sigma^2 = \mathrm{Var}[X], \mu = \mathrm{E}[X]$. Then our estimator σ_e^2 is given by

$$\begin{aligned}
\sigma_e^2 &= \frac{\sum y_i^2}{n} - \left(\frac{\sum y_i}{n}\right)^2 \\
\mathrm{E}\left[\sigma_e^2\right] &= \mathrm{E}\left[\frac{\sum y_i^2}{n} - \left(\frac{\sum y_i}{n}\right)^2\right] \\
&= \frac{\sum \mathrm{E}\left[y_i^2\right]}{n} - \frac{\mathrm{E}\left[(\sum y_i)^2\right]}{n^2} \\
&= \frac{\sum(\sigma^2 + \mu^2)}{n} - \frac{\mathrm{Var}\left[\sum y_i\right] + \mathrm{E}^2\left[\sum y_i\right]}{n^2} \\
&= \frac{n(\sigma^2 + \mu^2)}{n} - \frac{n\sigma^2 + n^2\mu^2}{n^2} \\
&= \sigma^2\left(1 - \frac{1}{n}\right)
\end{aligned}$$

So this gives a biased estimator, but we can make it unbiased simply by multiplying by $\frac{n}{n-1}$.

$$\mathrm{E}\left[\frac{n\sigma_e^2}{n-1}\right] = \sigma^2$$

■

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 $p \in [0, 1/\sqrt{2}]$.

Solution. This part and the next were discussed in Notes 13.3, where Wald's Theorem is used to show that

$$\mathrm{E}[T] = \frac{1}{p}\left(\frac{1}{(1-p)^n} - 1\right).$$

The hint now tells us that

$$\mathrm{E}[T] \geq \frac{e^{pn} - 1}{p} = 10^{12} \cdot (e^{10} - 1),$$

so the expected *time* to complete the task is at least $e^{10} - 1$ seconds, or roughly 6 hours. Since the operation should only take 10 seconds in the absence of failures, this is quite a penalty. ■

(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.

Solution. Now we have to do 10 calculations each of which can be computed without regard to previous failures. So by the previous analysis, Alyssa expects the ith component to take

$$\mathrm{E}[T_i] \geq \frac{e^{pn} - 1}{p} = 10^{12}(e - 1)$$

operations (note that $n = 10^{12}$). So, the expected number of operations for the entire computation is

$$\mathrm{E}\left[\sum_{i=1}^{10} T_i\right] = \sum_{i=1}^{10} \mathrm{E}[T_i] = 10^{13}(e-1),$$

giving an expected total computing time of roughly 17 seconds. The 10 state saves take time $10 \cdot 10^{-4} = 10^{-3}$ seconds, which is negligible. ■

(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 $np \in [0, 1], n > 3$.

Solution. As before,

$$\begin{aligned} \mathrm{E}[T_i] &= \frac{1}{p}\left(\frac{1}{(1-p)^n} - 1\right) \\ &= \frac{1 + pn + \delta(pn)^2 - 1}{p} \\ &= n(1 + \delta pn) \end{aligned}$$

with $0 \le \delta < 2, n = 10^{13}/10000 = 10^9$. So, the expected number of operations for the entire computation, excluding state saves, is roughly 10^{13} (since $p = 10^{-12}$ and $n = 10^9$, we ignore the δpn term). This time the 10^4 state saves are not negligible, but only take one second, giving a total time of 11 seconds. ■

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.

Solution. According to Ben, the probability would be $(0.4)^{1000} = (2/5)^{1000}$. But the actual probability is $(1/6)^{600}(3/4)^{400}$, and we don't even need to evaluate these expressions to see that they must have different values. ■

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

Solution. Ben's reasoning would be ok if the event that a fly is large is independent of whether the next fly is large. That's not the case here: after the 99th fly in the first hour, we can predict whether the 100th fly will be large or small. The number, R, of flies caught in 10 hours is actually the sum of a random variable with distribution $f_{600,1/6}$ and another variable with distribution $f_{400,3/4}$, and R not only disagrees with Ben's model on the probability that all the flies will be caught, it does not even have a binomial distribution. ■

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

Solution. The expected number of flies caught is $600(1/6) + 400(3/4) = 400$, so by Markov, $\Pr\{R \geq 500\} \leq 400/500 = 0.8$. ■

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

Solution. The variance is $600(1/6)(5/6) + 400(3/4)(1/4) = 1900/12 \approx 158$, so the one-sided Chebyshev bound is

$$\Pr\{R - 400 \geq 100\} \leq \frac{1900/12}{1900/12 + 100^2} = 1900/121,900 \approx 1/64.$$

■

(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 $\sqrt{e} \approx 1.6$.)

Solution.

$$\begin{aligned}
\Pr\{R \geq 500\} &= \Pr\{R \geq (5/4)400\} \\
&\leq \exp(-((5/4)\ln(5/4) - 5/4 + 1)400) \\
&= \exp(-(500\ln(5/4) - 500 + 400)) \\
&\approx \exp(-(500(0.223) - 100)) \\
&= \exp(-(111.5 - 100)) \\
&= e^{-11.5} \\
&= \frac{\sqrt{e}}{(e^3)^4} \\
&\approx \frac{1.6}{20^4} \\
&= \frac{1}{100,000}.
\end{aligned}$$

■

(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?

Solution. We know the expectation is 400 flies in 10 hours, so Markov's bound will hold because it only depends on the expectation and the nonnegativity of the number of flies. In the case of the Chernoff bound, we also need to know that the number of flies is a sum of independent Bernoulli variables; we are no longer given this, so Chernoff does not apply. To apply Chebyshev we need the variance, which we aren't given.

Actually, to apply Chebyshev, all we need is a bound on the variance, and there is one given that the expectation is 400 and R is nonnegative. The maximum possible variance for a nonnegative distribution with mean 400 occurs for the two-valued variable taking values 0 and 800 with equal probability. In this case the variance is 400^2. Plugging this value into the one-sided Chebyshev bound gives

$$\Pr\{R - 400 \geq 100\} \leq \frac{400^2}{400^2 + 100^2} = \frac{160,000}{170,000} = \frac{16}{17}$$

which is worse than the Markov bound.

This bound could be improved a bit further using the additional constraint that R must be a sum of two independent nonnegative integer-valued variables with expectations 100 and 300, respectively. In this case the maximum variance is $100^2 + 300^2 = 100,000$, yielding a one-sided Chebyshev bound of $100,000/(100,000 + 10,000) = 10/11$, still not as good as the Markov bound in this case. ■

(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?

Solution. Let S be the random variable equal to the number of small flies caught in 10 hours, and L the number of large flies. We are given that S has a Poisson distribution with rate $\lambda_S = 100$ and likewise for L with rate $\lambda_L = 300$. The total number of flies caught in 10 hours is $R ::= S + L$. We know that the arrival rate and mean of a Poisson distribution are the same, so $\mu_S = 100$ and $\mu_L = 300$. So $\mu_R = 100 + 300 = 400$ by linearity of expectation.

The Markov bound depends only on the expected number of flies caught. Since this is still 400, we have the same probability bound of 0.8 as in the previous case.

But since S and L are independent, we know from Notes 14 that R has a Poisson distribution with rate $\lambda_R = \lambda_S + \lambda_L = 400$. So not only is $\mu_R = 400$, but $\text{Var}[R] = 400$, so the one-sided Chebyshev bound becomes $400/(400 + 100^2) = 1/26$, a little larger than the previous case.

The standard Chernoff Bound *Theorem* does not apply because R is not a sum of Bernoulli variables. However, we know from Notes 14 that the standard Chernoff *bound* applies to the Poisson distribution even though it is not such a sum. Since the value of the Chernoff bound depends only on the expectation, the bound in this case remains the same $1/100,000$ as before. ■

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 $\Pr\{R \geq y\}$.

Solution.

- the Markov bound is μ_R/y, directly from (lost reference).
- the bound from the one-sided Chebyshev result is

$$\begin{aligned}\Pr\{R \geq y\} &= \Pr\{R - \mu_R \geq y - \mu_R\} \\ &\leq \frac{\sigma_R^2}{(y-\mu_R)^2 + \sigma_R^2}.\end{aligned}$$

- the bound from the Chernoff result is

$$\begin{aligned}\Pr\{R \geq y\} &= \Pr\left\{R \geq \frac{y}{\mu_R}\mu_R\right\} \\ &\leq \exp(-(\frac{y}{\mu_R}\ln\frac{y}{\mu_R} - \frac{y}{\mu_R} + 1)\mu_R) \\ &= \exp(-(y\ln\frac{y}{\mu_R} - y + \mu_R)).\end{aligned}$$

■

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

Solution. In this case $\mu_R = 1/2$ and $\sigma_R^2 = 1/4$, so

- Markov gives $\mu_R/y = 1/2$ which is exactly right,
- the bound from the one-sided Chebyshev result is

$$\frac{\sigma_R^2}{(y-\mu_R)^2 + \sigma_R^2} = \frac{1/4}{(1-(1/2))^2 + 1/4} = \frac{1}{2},$$

  and so is also exactly right,
- the bound from the Chernoff result is

$$\exp(-(\ln(1/(1/2)) - 1 + 1/2)) = e^{1/2 - \ln 2} = \sqrt{e}/2 > 0.83,$$

  and so is a large over-estimate.

■

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 $\Pr\{R > c\,\mathrm{E}[R]\}$ is tighter than the Chernoff bound, where R is positive and $c > 1$. (Providing an example is sufficient.)

Solution. Still consider the case where c is small, that is, $c = 1 + \epsilon$ and $\epsilon < 1$. Assume that $\mathrm{Var}[R] = \mathrm{E}[R] = 1$. The Chernoff bound is $e^{-(c \ln c - c + 1)}$. Numerical experiments show that

$$\frac{1}{1+\epsilon^2} < e^{-(c \ln c - c + 1)}.$$

Therefore, the Chebyshev one-sided bound on $\Pr\{R > c\,\mathrm{E}[R]\}$ is tighter than the Chernoff bound.
■

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).

Solution. The normalized version of any variable, R, is $(R - \mu_R)/\sigma_R$. But $\text{E}[B_n] = np$ and $\text{Var}[B_n] = np(1-p)$, so

$$B_n^* = \frac{B_n - np}{\sqrt{np(1-p)}}.$$

■

(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).

Solution. Note that B_n has the same distribution as the sum, S_n, of n mutually independent Bernoulli random variables, G_i, with mean p and variance $p(1-p)$. So the normalized version, B_n^* of B_n has the same distribution as the normalized version, T_n, of S_n. The Central Limit Theorem says that

$$\lim_{n\to\infty} \Pr\{T_n < \beta\} = N(\beta), \tag{1}$$

and since the distributions of T_n and B_n^* are the same, equation (1) holds for B_n^* as well. ■

(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.)

Solution. We have an unbiased binomial distribution for this case. The expected number of heads is 50 and the variance of the number of heads is $100/4 = 25$. Let S_{100} be the number of heads. We have

$$\begin{aligned}
\Pr\{30 < S_{100} < 70\} &= \Pr\{-4 < T_{100} < 4\} \\
&\approx N(4) - N(-4) \\
&= 2N(4) - 1 = 0.9999.
\end{aligned}$$

Note that to make this approximation practical n needs to be "sufficiently large". A rule often stated is that n is "sufficiently large" if $np \leq 5$ and $n(1-p) \leq 5$.

Since the distribution is symmetric, $\Pr\{30 < S_{100} < 70\} = 1 - 2\Pr\{S_{100} \geq 70\}$. Applying the generalized Chernoff bound derived in the lecture note, we have

$$\Pr\{S_{100} \geq 70\} \leq e^{-4^2/2}$$

where $4 = \frac{20}{5}$. Therefore,

$$\Pr\{30 < S_{100} < 70\} = 1 - 2\Pr\{S_{100} \geq 70\} \geq 1 - 2e^{-4^2/2} = 0.9993$$

which gives us a good lower bound on $\Pr\{S_{100} \geq 70\}$. ■

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* $\text{Var}[X_i] \leq b$ *for some $b \geq 0$ and all $i \geq 1$. Let*

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

Then for every $\epsilon > 0$,

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

(a) Prove this Theorem.

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

Solution. Essentially identical to the proof attached, except that $\text{Var}[G_i]$ gets replaced by b, and the equality becomes $\leq$ where the b is first used. ■

(b) Conclude

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

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

Solution. For fixed ϵ, the bound $b/(n\epsilon^2) \to 0$ as $n \to \infty$. ■

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

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

Choose a to minimize this last bound.

Solution. *Proof.*

$$\begin{aligned}
\Pr\{R - \mu_R \geq x\} &\leq \Pr\left\{(R - \mu_R + a)^2 \geq (x+a)^2\right\} && \text{(from the hint)} \\
&\leq \frac{\mathrm{E}\left[(R - \mu_R + a)^2\right]}{(x+a)^2} && \text{(Markov bound)} \\
&= \frac{\mathrm{E}\left[(R - \mu_R)^2\right] + 2a(\mathrm{E}[R] - \mu_R) + a^2}{(x+a)^2} && \text{(linearity of expectation)} \\
&= \frac{\mathrm{E}\left[(R - \mu_R)^2\right] + a^2}{(x+a)^2} && (\mathrm{E}[R] = \mu_R) \\
&= \frac{\mathrm{Var}[R] + a^2}{(x+a)^2} && \text{(def of variance).} \qquad (2)
\end{aligned}$$

Setting the derivative of (2) with respect to a to zero, we find that $a = \mathrm{Var}[R]/x$ minimizes (2). Plugging in this value for a, we have

$$\begin{aligned}
\Pr\{R - \mu_R \geq x\} &\leq \frac{\mathrm{Var}[R] + (\mathrm{Var}[R]/x)^2}{(x + \mathrm{Var}[R]/x)^2} && (a = \mathrm{Var}[R]/x \text{ in (2)}) \\
&= \frac{\mathrm{Var}[R]\,x^2 + \mathrm{Var}^2[R]}{(x^2 + \mathrm{Var}[R])^2} && (\text{multiply by } x^2/x^2) \\
&= \frac{\mathrm{Var}[R]\,(x^2 + \mathrm{Var}[R])}{(x^2 + \mathrm{Var}[R])^2} \\
&= \frac{\mathrm{Var}[R]}{x^2 + \mathrm{Var}[R]}.
\end{aligned}$$

□

■

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?