Practice Problems II - Some Solutions
some more will come later.

- **Question 1:** This is sampling (from 7 grades) with replacement (5 times). $|S| = 7^5$.

  (a) By counting we get: $f_X(1) = 7/7^5$; $f_X(2) = \binom{7}{2}[2^5 - 2]/7^5$; $f_X(3) = [\binom{5}{3}(7)_3 + \binom{7}{2}\binom{5}{2}\binom{3}{2}5]/7^5$; $f_X(4) = \binom{5}{2}(7)_4/7^5$; $f_X(5) = (7)_5/7^5$. Therefore $E(X) = \sum_{j=1}^{5} j\text{Prob}\{X = j\} = \sum_{j=1}^{5} j f_X(j) = 63217/7^5 = 3.761349438$.

  (b)

  $$\phi_X(s) = \sum_{j=1}^{5} \text{Prob}\{X = j\} s^j = \sum_{j=1}^{5} f_X(j) s^j = [7s + 630s^2 + 5250s^3 + 8400s^4 + 2520s^5]/7^5.$$

  Now it is routine to compute $E(X) = \phi'(1)$ and $V(X) = \phi''(1) + \phi'(1) - [\phi'(1)]^2$.

  Using indicators (let X_A be the indicator of the event an A was received (in one of the courses), X_B the indicator for a B was received, etc. Its very easy to compute the expectation and variance for each of them. This gives a much easier way to compute $E(X)$ than what we did above, using the frequency function and then the generating function. However it doesnt work for the variance because the indicators are not independent. Make sure you understand this.

- **Question 2a:** W_1 is the wait for a success (remove a bug) where the probability is $\mathcal{P} = 1/3$). We are asked for the probability $W_1 < 10$. Its the geometric probability

  $$\frac{1}{3} \sum_{n=1}^{9} (\frac{2}{3})^{n-1}.$$

  W_3 the wait for three successes. Its the negative binomial probability

  $$(\frac{1}{3})^3 \sum_{n=3}^{9} \binom{n-1}{2} (\frac{2}{3})^{n-3}.$$

- **Question 2d:** Its 6.

- **Question 5a:** We have $n = 1000$ Bernoulli trials where success is HEAD. If the coin were fair, we would expect $n\mathcal{P} = 500$ successes, so from the given data, we have $t = 200$ too few. Tchebycheff's inequality shows this event has probability at most $V(S_{1000})/t^2 = \frac{100(1/2)(1/2)}{200^2} = 1/160$ ($V(S_n) = n\mathcal{P}(1 - \mathcal{P})$). We are $159/160 = 99.375\%$ confident the coin is biased against HEADS (there were too few).

- **Question 5b:** We have n Bernoulli trials where Success on a trial is "TAIL". We are told that the number of successes is $S_n = n/3$. Recall that $E(S_n) = n\mathcal{P}$ and $V(S_n) = n\mathcal{P}(1 - \mathcal{P})$, so if the coin were fair, we expect $E(S_n) = n/2$ Tails and $V(S_n) = n/4$.

  Tchebycheff's inequality gives

  $$P(|S_n - n/2| \geq t) \leq \frac{V(S_n)}{t^2} = \frac{n/4}{t^2};$$

We are told that $S_n = n/3$, which means that we observed a deviation of $t = n/6$ from the expected. According to the above, a deviation this large (or larger) would occur with probability at most $n/(4t^2) = 9/n)$. The smaller this quantity, the more we believe the coin is biased against Tails.

In fact 1 minus the value of the right-hand side of the above inequality (in our case it is $1 - 9/n)$) is the *confidence* for our conclusion, and we want this to be .95 (or 95 percent). Solving $1 - 9/n) > .95$ reveals that once $n > 180$ are 95 percent confident that we saw too few Tails to believe the coin is fair.

- **Question 5c:** The logic here is similar to the above. S_{6000} is the number of threes that showed in 6000 tosses of a die. If the die were fair $E(S_{6000}) = 1000$ and $V(S_{6000}) = 5000/6$. Tchebycheff's inequality gives

$$P(|S_{6000} - 1000| \geq t) \leq \frac{5000)}{6t^2}$$

and to be 99 percent confident the deviation is large, we want the right hand side to be less than .01. This implies $t^2 > 5000/.06$, or that the deviation from the expectation is $t > 288.675....$ The required answer is $1000 + 289$, where we added because the coin is biased *in favor* of THREE, so we need to observe MORE than expected.

Ballot Theorem and Binary Trees

- **Question 1:** There are $\binom{2n}{n}/(n+1)$ rooted binary trees with n nodes, so 429 for $n = 7$.

  If there are three nodes in the left sub-tree there are also three in the right. Each of these is one of the 5 possible 3-node trees, so 25 of the 7-node trees have 3 nodes on the left (cartesian product principle).

  The answer is 68: The left sub-tree can have $j \in \{0, 1, 2, 3, 4, 5, 6\}$ nodes, and the right will have $6 - j$. When $j = 0$ exactly four of the 6-node sub-trees will have height 2 (so the whole tree has height 3). When $j = 1$ exactly six of the 5-node sub-trees will have height 2 (so again, the whole tree has height 3). When $j = 2$ there are two possible 2-node left sub-trees and for each, six of the 4-node right sub-tree have height 2 (again, the whole tree has height 3), so twelve of the trees with two nodes on the left and four nodes on the right have height 3. The case $j = 3$ is tricky. Of the five 3-node trees, all but one have height 2: of the 25 pairings of the 3-node trees on the left and right we just have to avoid the case where both have height 1 (so the whole tree would have height only 2). Thus there are twenty-four 7-node trees of height 3 having 3 nodes on the left. Finally the cases $j = 4, 5, 6$ are the same as $j = 0, 1, 2$, reversing left and right. (this is harder than I first thought).

- **Question 2:** The probability that a random tree has height 3 is 68/429. The conditional probability, given three nodes on the left is 24/25.

- **Question 3:** Let S_j be the number of $5 bills in the cash register after the j^{th} ticket is sold; $S_0 = 0$. If the j^{th} person pays with a $5 bill $X_j = 1$; if she pays with a $10 bill, the cashier needs to give change of $5, so $X_j = -1$. $S_j = X_1 + \cdots + X_j$. At the end, after $m = 100$ customers, $S_{100} =$ the number of 5 bills minus the number of 10 bills, so $S_{100} = k = 50$. By the ballot theorem, of the $\binom{100}{75}$ paths from $(0, 0)$ to $(100, 50)$, $k/m = 1/2$ of them are positive.

- **Question 4:** Of the $\binom{100}{75}$ paths frorm $(0, 0)$ to $100, 50)$, $\frac{k+1}{n_H+1} = \frac{51}{76}$ of them are non-negative.

- **Question 5:** (see answer for question 9) $\frac{(52)(101}{(77)(76)}\binom{100}{75} = .8974709501\binom{100}{75}$

- **Question 9:** In the random walk on the integers let $m = n_H + n_T$ and $S_m = X_1 + \cdots + X_m = n_H - n_T$. We assume $S_m = k \geq -1$ and want to count the paths from $(0, 0)$ to (m, k) for which $S_j \geq -1$; i.e. they never go below $y = -1$ (candidate A never trails by more than 1). We will start from $(-2, -2)$ and assume we get two heads to take us to $(0, 0)$. From here the path will be "good" if it never descends below $y = -1$. So pretending that $y = -2$ is the $x-$axis (it *is* from the point of view of the start at $(-2, -2)$) we will use the ballot theorem to count the number of "positive" paths from $(-2, -2)$ to (m, k), i.e., paths that never go below $y = -1$. This walk has $m^* = m + 2$ steps, $n_H^* = n_H + 2$ Heads, $n_T^* = n_t$ Tails, and an excess of $k^* = k + 2$ Heads over Tails. Then by the Ballot Theorem, there are

$$\frac{k^*}{m^*}\binom{m^*}{k^*} = \frac{k+2}{m+2}\binom{m+2}{n_H+2} = \frac{(k+2)(m+1)}{(n_H+2)(n_H+1)}\binom{m}{n_H}$$

  good paths, and the fraction of ALL paths that are good is $\frac{(k+2)(m+1)}{(n_H+2)(n_H+1)}$. In question 7 this fraction is $\frac{(52)(151)}{(102)(101)} = .7621821006$.

- **Question 10:** There are $\binom{m}{n_H}$ paths from $(0, 0)$ to (m, k). We have $m = n_H + n_T = 12$ and $k = n_H - n_T = 2$, so $n_H = 7$ and $n_T = 5$. Therefore there are 792 paths.

- Question 11: By the Ballot Theorem there is a fraction $k/m = 2/12$ of them that are positive (so 132) and $(k+1)/(n_H+1) = 3/8$ of them that are non-negative (so 297).
- Question 12: By the same reasoning there are $\binom{6}{3} = 20$ paths from $(0,0)$ to $(6,0)$ (let A be this set of paths) and $\binom{6}{4} = 15$ paths from $(6,0)$ to $(12,2)$ (let B be this set of paths). For each path in A and each path in B we can "join" the pair to get a path from $(0,0)$ to $(12,2)$ that passes through $(6,0)$. Therefore 300 of the 792 paths that pass through $(6,0)$.
- **Question 13:** NONE. S_m and m must have the same parity - after an even number of tosses (12) you cant have an odd profit (5).