Stat 310 Homework 1 Key

Chapter 1, problems 2, 18, 20, 22, 24, 48, 64, 72, 73, 75. Due 9/9/99.

1.2. Two six-sided dice are thrown sequentially, and the face values that come up are recorded.

a) List the sample space. Well, noting the outcomes as (first die, second die), the 36 elements of the sample space are:

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

b) List the elements that make up the following events: (1) $A =$ the sum of the two values is at least 5

			14	15	16
		23	24	25	26
	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

(2) $B =$ the value of the first die is higher than the value of the second

21				
31	32			
41	42	43		
51	52	53	54	
61	62	63	64	65

(3) $C =$ the value of the first die is 4.

41 42 43 44 45 46.

c) List the elements of the following events: (1) $A \cap C$

41 42 43 44 45 46.

(2) $B \cup C$

21					
31	32				
41	42	43	44	45	46
51	52	53	54		
61	62	63	64	65	

(3) $A \cap (B \cup C)$

	32				
41	42	43	44	45	46
51	52	53	54		
61	62	63	64	65	

1.18. A lot of n items contains k defectives, and m are selected randomly and inspected. How should the value of m be chosen so that the probability that at least one defective item turns up is .90? Apply your answer to

a) $n = 1000$, $k = 10$. First off, we convert the problem so that instead of trying to compute the probability of one or more defects, we are trying to find 1 minus the probability of zero defects. This is given by

$$1 - \frac{\binom{n-k}{m}\binom{k}{0}}{\binom{n}{m}}$$

which we can interpret as choosing m of the $n-k$ good items and none of the k bad ones. This can also be viewed as

$$1 - \frac{\binom{n-m}{k}\binom{m}{0}}{\binom{n}{k}}$$

which we can interpret as identifying k of the $n-m$ nonchosen items as bad and none of the m chosen items as bad. Mathematically,

$$\frac{\binom{n-k}{m}\binom{k}{0}}{\binom{n}{m}} = \frac{(n-k)!}{(n-k-m)!m!}\frac{k!}{k!}\frac{m!(n-m)!}{n!} = \frac{(n-m)!}{(n-k-m)!k!}\frac{m!}{m!}\frac{k!(n-k)!}{n!} = \frac{\binom{n-m}{k}\binom{m}{0}}{\binom{n}{k}}$$

so the values are equivalent. Now, neither of these formulations is extremely well-suited to numerical evaluation, but they do suggest useful approximations. Suppose that we start drawing items from the bin. Then the chance that we will get no defectives is

$$\frac{n-k}{n} * \frac{n-k-1}{n-1} * \frac{n-k-2}{n-2} * \ldots * \frac{n-k-m+1}{n-m+1}.$$

If we assume that n and is sufficiently large that these ratios do not change over the m terms, we can write this as

$$\left(\frac{n-k}{n}\right)^m.$$

This is the "infinite population" assumption and it gets better and better as n and k increase as long as k/n is held fixed. For this problem, this approximation suggests

$$\begin{aligned} P(1 \text{ or more defects}) &\geq 0.9 \\ 1 - P(\text{no defects}) &\geq 0.9 \\ P(\text{no defects}) &\leq 0.1 \\ \left(\frac{n-k}{n}\right)^m &\leq 0.1 \\ m &\approx \frac{\log(0.1)}{\log(990/1000)} = 229.1. \end{aligned}$$

This gives a coarse approximation, as the ratio does not stay that constant over such a large fraction of the total extent $m/n = .2291$. Using the other approach, we get an approximation of Now suppose that we start identifying defective items both drawn from and remaining in the bin. Then the chance that we drew no defectives is

$$\frac{n-m}{n} * \frac{n-m-1}{n-1} * \frac{n-m-2}{n-2} * \ldots * \frac{n-k-m+1}{n-k+1}.$$

This may not appear to be the same as what we got using the other approach, so consider the case of $n = 1000$, $k = 3$, $m = 5$. The chance of no defectives is

$$\frac{997}{1000} * \frac{996}{999} * \frac{995}{998} * \frac{994}{997} * \frac{993}{996} = \frac{995}{1000} * \frac{994}{999} * \frac{993}{998}$$

so we see that when cancellation is included the two expressions are equal. Now, as to the approximation suggested by the second approach, If we assume that n is sufficiently large that these ratios do not change over the k terms, we can write this as

$$\left(\frac{n-m}{n}\right)^k.$$

In this case, we know that $k = 10$. Since the validity of the approximations relies on the ratios remaining approximately constant, *the better approximation is the one involving the smaller number of terms.* This approximation suggests

$$\begin{aligned}
P(\text{no defects}) &\leq 0.1 \\
\left(\frac{n-m}{n}\right)^k &\leq 0.1 \\
10 \log\left(\frac{1000-m}{1000}\right) &\leq \log(0.1) \\
\log(1000-m) &\leq \log(1000) + 0.1\log(0.1) \\
m &\approx 1000 - \exp\left[\log(1000) + 0.1\log(0.1)\right] \\
m &\approx 205.67.
\end{aligned}$$

Since this approximation involved only 10 terms, not over 200, we expect it to be better. Simulation using MATLAB, as below,

```
n = 1000;
k = 10;
nterms = 300;
zed = zeros(nterms,1);
zed(1) = (n-k)/n;
for(i = 2:nterms)
zed(i) = zed(i-1)*(n-k-i+1)/(n-i+1);
end
plot(zed)
```

gives `zed(204) = 0.1009` and `zed(205) = 0.0997` so the exact answer is $m = 205$.

b) $n = 10000$, $k = 100$. In this case, we have increased n and k but kept the ratio of the two the same, so the independence approximation (the first one) gets better. The math is the same as in part a), so let's take a look at the results:

$$\begin{aligned} \left(\frac{n-k}{n}\right)^m &\leq 0.1 \\ m &\approx \frac{\log(0.1)}{\log(9900/10000)} = 229.1, \end{aligned}$$

exactly as before. The second approximation though, gives

$$\begin{aligned} \left(\frac{n-m}{n}\right)^k &\leq 0.1 \\ m &\approx n - \exp\left[\log(n) + k^{-1}\log(0.1)\right] \\ m &\approx 10000 - \exp\left[\log(10000) + 0.01\log(0.1)\right] \\ m &\approx \frac{\log(0.1)}{\log(9900/10000)} = 227.6278, \end{aligned}$$

so we see that the two answers are now virtually in agreement. Again, as the second approximation uses only 100 terms, not 200 plus, we expect it to be slightly better. Updating the values of n and k in the MATLAB program given above and trying it again gives `zed(226) = 0.1005` and `zed(227) = 0.0995`, so the exact answer is $m = 227$. Note that if we up n and k by another factor of 100, the first approximation is now better, and the two approximations are now 229.1 and 230.23 - the exact routine gives `zed(229) = 0.1001` and `zed(230) = 0.0991`. As the population size goes to infinity, the number that we need to draw to achieve a given level of accuracy goes to a limit!

1.20. A deck of 52 cards is shuffled thoroughly, and n cards are turned up. What is the probability that the four aces are all next to each other?

We attack this problem by breaking it down into simpler problems. For instance, what is the probability that the first four cards off the deck are the four aces? This is given by

$$\frac{4}{52} * \frac{3}{51} * \frac{2}{50} * \frac{1}{49}.$$

The probability that the second through fifth cards are the four aces is given by

$$\frac{48}{52} * \frac{4}{51} * \frac{3}{50} * \frac{2}{49} * \frac{1}{48}$$

which reduces to the first probability found when we note that the numerator of the first term cancels the denominator of the last term. Thus, we can write the probability of the four aces occurring together as the probability that they will be in positions one through four, plus the probability that they will be in positions two through five (these two events are disjoint, so the probabilities add), plus the probability of positions three through six, and so on through the probability of positions forty-nine through fifty-two. Summing, this gives

$$49\frac{4}{52} * \frac{3}{51} * \frac{2}{50} * \frac{1}{49} = \frac{49}{\binom{52}{4}}$$

the total number of sets of indices corresponding to four sequential positions over the the total number of all possible sets of indexes correponding to the positions of the four aces.

2.22. A deck of cards is shuffled thoroughly, and n cards are turned up. What is the probability that a face card turns up? For what value of n is this probability about .5?

This is very similar in nature to problem 1.18. We interpret the probability that a face card turns up to mean that at least one shows up, so we work with the complement and find the probability that no face cards turn up. There are 12 face cards, so the chance of no face cards showing up in n cards is

$$\frac{40}{52} * \frac{39}{51} * \dots * \frac{40-n+1}{52-n+1}.$$

An approximation (see 1.18) suggests $n = \log(.5)/\log(40/52) = 2.6419$, so this will not be too onerous to do by hand. The probabilities are

n	P(no face cards)	P(at least one)
1	0.7692	0.2308
2	0.5882	0.4118
3	0.4471	0.5529

so we pick $n = 3$ so that the chance of at least one face card exceeds one half.

1.24. If n balls are distributed randomly into k urns, what is the probability that the last urn contains j balls?

We attack this problem via the time-honored practice of trying things for small values of n and trying to generalize. First off, consider the case $n = 1$. Here, j can only take on the values 0 and 1, and the probabilities are given by

$$\begin{aligned} P_{n=1}(j=0) &= \frac{k-1}{k} \\ P_{n=1}(j=1) &= \frac{1}{k}. \end{aligned}$$

To extend things to the $n = 2$ case, we view this as one of watching two tosses in a row and seeing whether or not they hit bin k. There are four possible outcomes here: (miss,miss), (miss,hit), (hit,miss), and (hit,hit). We can compute the probabilities associated with each of these from the results for the single toss case given above, as the results of one toss are independent of the others so the probability product rule applies. Here, j can take on the values 0, 1 and 2, with probabilities

$$\begin{aligned} P_{n=2}(j=0) &= (\text{miss}, \text{miss}) \\ &= \left(\frac{k-1}{k}\right)^2 \\ P_{n=2}(j=1) &= (\text{miss}, \text{hit}) + (\text{hit}, \text{miss}) \\ &= 2\left(\frac{1}{k}\right)\left(\frac{k-1}{k}\right) \end{aligned}$$

$$\begin{aligned} P_{n=2}(j=2) &= (\text{hit}, \text{hit}) \\ &= \left(\frac{1}{k}\right)^2. \end{aligned}$$

Extending this to the $n = 3$ case, we get

$$\begin{aligned} P_{n=3}(j=0) &= (\text{miss}, \text{miss}, \text{miss}) \\ &= \left(\frac{k-1}{k}\right)^3 \\ P_{n=3}(j=1) &= (\text{miss}, \text{miss}, \text{hit}) + (\text{miss}, \text{hit}, \text{miss}) + (\text{hit}, \text{miss}, \text{miss}) \\ &= 3\left(\frac{1}{k}\right)\left(\frac{k-1}{k}\right)^2 \\ P_{n=3}(j=2) &= (\text{miss}, \text{hit}, \text{hit}) + (\text{hit}, \text{miss}, \text{hit}) + (\text{hit}, \text{hit}, \text{miss}) \\ &= 3\left(\frac{1}{k}\right)^2\left(\frac{k-1}{k}\right) \\ P_{n=3}(j=3) &= (\text{hit}, \text{hit}, \text{hit}) \\ &= \left(\frac{1}{k}\right)^3. \end{aligned}$$

Generalizing, we see that each case the probability is given by

$$P(j) = \binom{n}{j}\left(\frac{1}{k}\right)^j\left(\frac{k-1}{k}\right)^{n-j},$$

so that the number of balls in bin k has a binomial distribution with parameters $(n, 1/k)$.

1.48. An urn contains three red and two white balls. A ball is drawn, and then it and another ball of the same color are placed back in the urn. Finally, a second ball is drawn.

a) What is the probability that the second ball drawn is white? Let D_1 denote the event that the first ball drawn is white, and D_2 the event that the second ball drawn is white. We want $P(D_2)$. To find this, we expand a few things. First,

$$P(D_2) = P(D_2 D_1) + P(D_2 D_1^c).$$

This simply says that something happened on the first draw. Now, the intersection probabilities themselves are not so easy to find, but we can reexpress those using the properties of conditional probability as

$$P(D_2) = P(D_2|D_1)P(D_1) + P(D_2|D_1^c)P(D_1^c).$$

At this point we can compute numbers, and we get

$$P(D_2) = \frac{3}{6} * \frac{2}{5} + \frac{2}{6} * \frac{3}{5} = \frac{2}{5}.$$

b) If the second ball drawn is white, what is the probability that the first ball drawn was red? Here we're being asked to find $P(D_1^c|D_2)$. We can rewrite this using the definition of conditional probability as

$$P(D_1^c|D_2) = \frac{P(D_2 D_1^c)}{P(D_2)}.$$

Using the expansions found in part a), we write this as

$$\begin{aligned} P(D_1^c|D_2) &= \frac{P(D_2|D_1^c)P(D_1^c)}{P(D_2|D_1)P(D_1)+P(D_2|D_1^c)P(D_1^c)} \\ &= \frac{1/5}{2/5} = \frac{1}{2}. \end{aligned}$$

1.64. If B is an event, with $P(B) > 0$, show that the set function $Q(A) = P(A|B)$ satisfies the axioms of a probability measure. Thus, for example,

$$P(A \cup C|B) = P(A|B) + P(C|B) - P(A \cap C|B).$$

We take the three axioms given on p.4 of the text.

(i) $P(\Omega) = 1$.

$$Q(\Omega) = P(\Omega|B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

(ii) If $A \subset \Omega$, then $P(A) \geq 0$.

$$Q(A) = P(A|B) = \frac{P(AB)}{P(B)}.$$

As $P(B) > 0$, the last term on the right will be ≥ 0 if $P(AB) \geq 0$. As $A, B \in \Omega$, $AB \in \Omega$, and hence $P(AB) \geq 0$ and $Q(A) \geq 0$.

(iii) If A_1 and A_2 are disjoint, then

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

Well,

$$\begin{aligned} Q(A_1 \cup A_2) &= P(A_1 \cup A_2|B) \\ &= \frac{P((A_1 \cup A_2) \cap B)}{P(B)} \\ &= \frac{P[(A_1 \cap B) \cup (A_2 \cap B)]}{P(B)} \\ &= \frac{P(A_1 \cap B) + P(A_2 \cap B) - P(A_1 \cap A_2 \cap B)}{P(B)} \\ &= \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} \\ &= Q(A_1) + Q(A_2). \end{aligned}$$

1.72. Suppose that n components are connected in series. For each unit, there is a backup unit, and the system fails only if both a unit and its backup fail. Assuming that all of the units are independent and fail with probability p, what is the probability that the system works? For $n = 10$ and $p = .05$, compare these results with those of Example F in Section 1.6.

Ok, a component works if either the initial unit or its backup works. The component fails only if both units fail. As the units are independent, the probability that a component fails is p^2 by the product rule, and the probability that a component works is $1 - p^2$. For the system to work, all n components must work. As these components function independently, the probability that the system works is $(1 - p^2)^n$

When $n = 10$ and $p = .05$, the probability that the system works is

$$[1 - (.05)^2]^1 0 = 0.9753,$$

which is considerably higher than the probability of 0.60 found for the system without backups.

1.73. A system has n independent units, each of which fails with probability p. The system fails only if k or more of the units fail. What is the probability that the system fails?

Well, this can be written as

$$\sum_{j=k}^{n} P(j \text{ units fail}).$$

As p is the probability that a unit fails, p^j is the probability that j specified units fail, $(1 - p)$ is the probability that a unit succeeds, $(1 - p)^{n-j}$ is the probability that $n - j$ specified units fail, and there are $\binom{n}{j}$ ways of choosing which j out of the n fail, so when combined this becomes

$$\sum_{j=k}^{n} \binom{n}{j} p^j (1 - p)^{n-j}.$$

This is one minus the cdf of a binomial random variable with parameters n and p.

1.75. This problem deals with an elementary aspect of a simple branching process. A population starts with one member; at time $t = 1$, it either divides with probability p or dies with probability $1 - p$. If it divides, then both of its children behave independently with the same two alternatives at time $t = 2$. What is the probability that there are no members in the third generation? For what value of p is this probability equal to .5?

Let D_1 be the event that the first individual dies at time $t = 1$, and let D_2 be the event that the population is zero at time $t = 2$. Then

$$P(D_2) = P(D_2|D_1)P(D_1) + P(D_2|D_1^c)P(D_1^c).$$

From the statement of the problem, $P(D_1^c) = p$ and $P(D_1) = 1 - p$. As a population cannot spontaneously appear, $P(D_2|D_1) = 1$. Now, if the first individual divided at time $t = 1$, there are two independent individuals presented with the option of dividing or dying at time $t = 2$. The chance that both will die (so that the population will be zero) is $(1-p)^2$. Putting it all together,

$$P(D_2) = 1 * (1-p) + (1-p)^2 * p = p^3 - 2p^2 + 1.$$

Solving for the value of p such that this probability is 0.5 is done numerically - I plotted the function using MATLAB;

```
zed = [0:0.001:1];
y = zed.^3 - 2*zed.^2 + 1;
plot(zed,y)
```

and found the root to be $p = 0.597$.