UC Berkeley
Department of Electrical Engineering and Computer Science

EE 126: Probablity and Random Processes

Problem Set 4: Solutions
Fall 2007

Issued: Thursday, September 20, 2007 **Due:** Friday, September 28, 2007

Reading: Bertsekas & Tsitsiklis, §2.1 — 2.5

Problem 4.1

Each of n files is saved independently onto either hard drive A (with probability p) or hard drive B (with probability $1-p$). Let A be the total number of files selected for drive A and let B be the total number of files selected for drive B.

1. Determine the PMF, expected value, and variance for random variable A.
2. Evaluate the probability that the first file to be saved ends up being the only one on its drive.
3. Evaluate the probability that at least one drive ends up with a total of exactly one file.
4. Evaluate the expectation and the variance for the difference, $D = A - B$.
5. Assume $n \geq 2$. Given that both of the first two files to be saved go onto drive A, find the conditional expectation, variance and probability law for random variable A.

Solution:

(a) We recognize that A is a binomial random variable (sum of n independent Bernoulli random variables).

$$p_A(a) = \begin{cases} \binom{n}{a} p^a (1-p)^{n-a} & a = 0, 1, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

So, we have $\mathbb{E}[A] = np$, $\text{var}(A) = np(1-p)$.

(b) Let L denote the desired event, i.e. L: the first file saved ends up being the only one on its drive. Let FA: the first file is saved on drive A. Let FB: the first file is saved on drive B. We apply the law of total probability and obtain

$$\begin{aligned} \mathbb{P}(L) &= \mathbb{P}(L|FA) \cdot \mathbb{P}(FA) + \mathbb{P}(L|FB) \cdot \mathbb{P}(FB) \\ &= p \cdot (1-p)^{n-1} + (1-p) \cdot p^{n-1} \end{aligned}$$

(c) Let EA: drive A ends up with exactly one file. Let EB: drive B ends up with exactly one file. Finally, let E denote the desired event - at least one drive ends up with a total of exactly one file. Notice that $E = EA \cup EB$. In general, we know that, $\mathbb{P}(E) = \mathbb{P}(EA) + \mathbb{P}(EB) - \mathbb{P}(EA \cap EB)$. Now when $n \neq 2$, $\mathbb{P}(EA \cap EB) = 0$. We distinguish two cases:

case (1): $n \neq 2$**.** We have $\mathbb{P}(E) = \mathbb{P}(EA) + \mathbb{P}(EB) = \binom{n}{1}p(1-p)^{n-1} + \binom{n}{1}(1-p)p^{n-1}$, since both A and B are binomial random variables (with different "success" probabilities).

case (2): $n = 2$**.** We have $\mathbb{P}(E) = \mathbb{P}(EA) + \mathbb{P}(EB) - \mathbb{P}(EA \cap EB) = 2p(1-p) + 2(1-p)p - 2p(1-p) = 2p(1-p)$.

(d) By linearity of expectation, we have $\mathbb{E}[D] = \mathbb{E}[A] - \mathbb{E}[B]$. Both A and B are binomial, and we have $\mathbb{E}[D] = np - n(1-p) = n(2p-1)$. Now since A and B are not independent (i.e. A and B are dependent) we cannot simply add their variances. Notice however that $A + B = n$, hence $B = n - A$. Therefore, $\mathrm{var}(D) = \mathrm{var}(A - B) = \mathrm{var}(A - n + A) = \mathrm{var}(2A - n) = 4\,\mathrm{var}(A) = 4np(1-p)$.

(e) Let C: both of the first two files to be saved go onto drive A. Consider the Bernoulli random variable X_i with parameter $p, i = 3, \ldots, n$. Notice that $\mathbb{E}[A|C] = \mathbb{E}[2 + \sum_{i=3}^{n} X_i] = 2 + \mathbb{E}[\sum_{i=3}^{n}] = 2 + (n-2)p$.
This is because each of the remaining $n-2$ trials still have success probability p. Now $\mathrm{var}(A|C) = \mathrm{var}(2 + \sum_{i=3}^{n} X_i) = \mathrm{var}(\sum_{i=3}^{n} X_i) = \sum_{i=3}^{n}(X_i)$ (by independence).
So, $\mathrm{var}(A|C) = (n-2)p(1-p)$. Finally, we desire $p_{A|C}(a)$.
Notice that in our new universe $A = 2 + \sum_{i=3}^{n} X_i$. The number of successes in the remaining trials is binomial. We just shift the PMF so that, for instance, $p_{A|C}(2) = \mathbb{P}(0 \text{ successes in } n-2 \text{ trials})$. So we have

$$p_{A|C}(a) = \begin{cases} \binom{n-2}{a-2}p^{a-2}(1-p)^{n-a} & a = 2, 3, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

Problem 4.2

The probability that any particular bulb will burn out during its K^{th} month of use is given by the PMF for K,

$$p_K(K_0) = \frac{1}{5}\left(\frac{4}{5}\right)^{K_0 - 1}, K_0 = 1, 2, 3, \ldots$$

Four bulbs are life-tested simultaneously. Determine the probability that

1. None of the four bulbs fails during its first month of use.
2. Exactly two bulbs have failed by the end of the third month.
3. Exactly one bulb fails during each of the first three months.
4. Exactly one bulb has failed by the end of the second month, and exactly two bulbs are still working at the start of the fifth month.

Solution: 1. $1 - p_K(1)$ is the probability that one bulb doesn't burn out at the first month. Since all 4 burn-out events are independent, the probability that all of them don't burn out at the first month is

$$(1 - p_K(1))^4 = (4/5)^4 = 0.4096$$

2. Let X_i be the indicator that the i-th buld burn out in during first three months. $P(X_i = 1) = p_K(1) + p_K(2) + p_K(3)$. $X = \sum_{i=1}^{4} X_i$ is a binomial random variable.

$$P(X = 2) = \binom{4}{2}(p_K(1)+p_K(2)+p_K(3))^2(1-p_K(1)-p_K(2)-p_K(3))^2 = \frac{6 \times 61^2 \times 64^2}{125^4} = 0.3746$$

3. 4 bulbs can be partitioned into 4 groups. Each group has exactly one bulb. There are 4! ways of partitions. Group 1 has the bulb that burns out at the first month. The second group has the one that fails at the second month and the third one burns out at the third month. The last one survives in the first three months.

$$\binom{4}{1,1,1,1} p_K(1)p_K(2)p_K(3)(1 - p_K(1) - p_K(2) - p_K(3)) = 0.05$$

4. Again, we partition 4 bulbs into 3 groups. First two groups have one bulb each. The last group has 2 bulbs. There are $4!/(1!1!2!) = 12$ ways of partitions. The bulb in the first group fails during the first 2 months. And the bulb in the second group burns out at the third or fouth month. And the last 2 bulbs survives during the first 4 months.

$$\binom{4}{1,1,2}(p_K(1) + p_K(2))(p_K(3) + p_K(4))(1 - p_K(1) - p_K(2) - p_K(3) - p_K(4))^2 = 0.167$$

Problem 4.3

A particular circuit works if all ten of its component devices work. Each circuit is tested before leaving the factory. Each working circuit can be sold for k dollars, but each nonworking circuit is worthless and must be thrown away. Each circuit can be built with either *ordinary* devices or *ultra-reliable* devices. An ordinary device has a failure probability of $q = 0.1$ while an ultra-reliable device has a failure probability of $q/2$, independent of any other device. However, each ordinary device costs \$1 whereas an ultra-reliable device costs \$3.

Should you build your circuit with ordinary devices or ultra-reliable devices in order to maximize your expected profit $E[R]$? Keep in mind that your answer will depend on k.

Solution:

Let R denote the profit resulting from building the circuit that requires ten component devices to not fail, where each device fails independently of the others with probability f. Letting I denote the income and C denote the cost, we can express this profit as $R = I - C$. For each option, that is the option to use either ordinary or ultra-reliable devices, C is constant but I is a random variable that has nonzero value k only if all ten component devices have not failed, which occurs with probability $(1 - f)^{10}$. Thus,

$$\mathbf{E}[R] = \mathbf{E}[I - C] = \mathbf{E}[I] - C = k(1 - f)^{10} - C \quad .$$

$$\begin{array}{rll} \textbf{Ordinary Option:} & C = 10, f = q = 0.10 & \Rightarrow \mathbf{E}[R] = k(0.90)^{10} - 10 \\ \textbf{Ultra-Reliable Option:} & C = 30, f = \frac{q}{2} = 0.05 & \Rightarrow \mathbf{E}[R] = k(0.95)^{10} - 30 \end{array}$$

We are interested in picking the option that gives maximum expected profit. Comparing the two expected profits we see that for $k \geq 80$ the cicuit with ultra-reliable components gives a better expected value and for $k < 80$ the circuit with ordinary components gives a better expected value.

Problem 4.4

Professor May B. Right often has her science facts wrong, and answers each of her students' questions incorrectly with probability 1/4, independently of other questions. In each lecture May is asked either 1 or 2 questions with equal probability.

1. What is the probability that May gives wrong answers to all the questions she gets in a given lecture?
2. Given that May gave wrong answers to all the questions she got in a given lecture, what is the probability that she got two questions?
3. Let X and Y be the number of questions May gets and the number of questions she answers correctly in a lecture, respectively. What is the mean and variance of X and the mean and the variance of Y?
4. Give a neatly labeled sketch of the joint PMF $p_{X,Y}(x, y)$.
5. Let $Z = X + 2Y$. What is the expectation and variance of Z?

For the remaining parts of this problem, assume that May has 20 lectures each semester and each lecture is independent of any other lecture.

1. The university where May works has a peculiar compensation plan. Each lecture May gets paid a base salary of \$1,000 plus \$40 for each question she answers and an additional \$80 for each of these she answers correctly. In terms of random variable Z, she gets paid $\$1000 + \$40Z$ per lecture. What is the expected value and variance of her *semesterly* salary?
2. Determined to improve her reputation, May decides to teach an additional 20-lecture class in her specialty (math), where she answers questions incorrectly with probability 1/10 rather than 1/4. What is the expected number of questions that she will answer wrong in a randomly chosen lecture (math or science).

Solution:

1. Use the total probability theorem by conditioning on the number of questions that May has to answer. Let A be the event that she gives all wrong answers in a given lecture, let B_1 be the event that she gets one question in a given lecture, and let B_2 be the event that she gets two questions in a given lecture. Then

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2).$$

From the problem statement, she is equally likely to get one or two questions in a given lecture, so $\mathbb{P}(B_1) = \mathbb{P}(B_2) = \frac{1}{2}$. Also, from the problem, $\mathbb{P}(A|B_1) = \frac{1}{4}$, and, because of independence, $\mathbb{P}(A|B_2) = (\frac{1}{4})^2 = \frac{1}{16}$. Thus we have

$$\mathbb{P}(A) = \frac{1}{4}\frac{1}{2} + \frac{1}{16}\frac{1}{2} = \frac{5}{32}.$$

2. Let events A and B_2 be defined as in the previous part. Using Bayes's Rule:

$$\mathbb{P}(B_2|A) = \frac{\mathbb{P}(A|B_2)\mathbb{P}(B_2)}{\mathbb{P}(A)}.$$

From the previous part, we said $\mathbb{P}(B_2) = \frac{1}{2}$, $\mathbb{P}(A|B_2) = \frac{1}{16}$, and $\mathbb{P}(A) = \frac{5}{32}$. Thus

$$\mathbb{P}(B_2|A) = \frac{\frac{1}{16}\frac{1}{2}}{\frac{5}{32}} = \frac{1}{5}.$$

As one would expect, given that May answers all the questions in a given lecture, it's more likely that she got only one question rather than two.

3. We start by finding the PMF for X and Y. $p_X(x)$ is given from the problem statement:

$$p_X(x) = \begin{cases} \frac{1}{2} & \text{if } x \in \{1, 2\}, \\ 0 & \text{otherwise.} \end{cases}$$

The PMF for Y can be found by conditioning on X for each value that Y can take on. Because May can be asked at most two questions in any lecture, the range of Y is from 0 to 2. Thus for each value of Y, we find

$$p_Y(0) = \mathbb{P}(Y = 0|X = 0)\mathbb{P}(X = 1) + \mathbb{P}(Y = 0|X = 2)\mathbb{P}(X = 2) = \frac{1}{4}\frac{1}{2} + \frac{1}{16}\frac{1}{2} = \frac{5}{32},$$

$$p_Y(1) = \mathbb{P}(Y = 1|X = 1)\mathbb{P}(X = 1) + \mathbb{P}(Y = 1|X = 2)\mathbb{P}(X = 2) = \frac{3}{4}\frac{1}{2} + 2\frac{3}{4}\frac{1}{4}\frac{1}{2} = \frac{9}{16},$$

$$p_Y(2) = \mathbb{P}(Y = 2|X = 1)\mathbb{P}(X = 1) + \mathbb{P}(Y = 2|X = 2)\mathbb{P}(X = 2) = 0\frac{1}{2} + (\frac{3}{4})^2\frac{1}{2} = \frac{9}{32}.$$

Note that when calculating $\mathbb{P}(Y = 1|X = 2)$, we got $2\frac{3}{4}\frac{1}{4}$ because there are two ways for May to answer one question right when she's asked two questions: either she answers the first question correctly or she answers the second question correctly. Thus, overall

$$p_Y(y) = \begin{cases} 5/32 & \text{if } y = 0, \\ 9/16 & \text{if } y = 1, \\ 9/32 & \text{if } y = 2, \\ 0 & \text{otherwise.} \end{cases}$$

Now the mean and variance can be calculated explicitly from the PMFs:

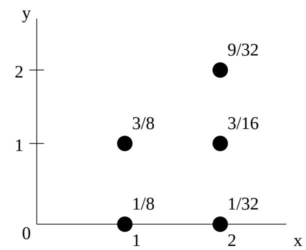
$$\mathbb{E}[X] = 1\frac{1}{2} + 2\frac{1}{2} = \frac{3}{2},$$

$$\text{var}(X) = (1 - \frac{3}{2})^2\frac{1}{2} + (2 - \frac{3}{2})^2\frac{1}{2} = \frac{1}{4},$$

$$\mathbb{E}[Y] = 0\frac{5}{32} + 1\frac{9}{16} + 2\frac{9}{32} = \frac{9}{8},$$

$$\text{var}(Y) = (0 - \frac{9}{8})^2\frac{5}{32} + (1 - \frac{9}{8})^2\frac{9}{16} + (2 - \frac{9}{8})^2\frac{9}{32} = \frac{27}{64}.$$

4. The joint PMF $p_{X,Y}(x,y)$ is plotted below. There are only five possible (x,y) pairs. For each point, $p_{X,Y}(x,y)$ was calculated by $p_{X,Y}(x,y) = p_X(x)p_{Y|X}(y|x)$.



5. By linearity of expectations,

$$\mathbb{E}[Z] = \mathbb{E}[X + 2Y] = \mathbb{E}[X] + 2\mathbb{E}[Y] = \frac{3}{2} + 2\frac{9}{8} = \frac{15}{4}.$$

Calculating $\text{var}(Z)$ is a little bit more tricky because X and Y are not independent; therefore we *cannot* add the variance of X to the variance of $2Y$ to obtain the variance of Z. (X and Y are clearly not independent because if we are told, for example, that $X = 1$, then we know that Y cannot equal 2, although normally without any information about X, Y could equal 2.)

To calculate $\text{var}(Z)$, first calculate the PMF for Z from the joint PDF for X and Y. For each (x,y) pair, we assign a value of Z. Then for each value z of Z, we calculate $p_Z(z)$ by summing over the probabilities of all (x,y) pairs that map to z. Thus we get

$$p_Z(z) = \begin{cases} 1/8 & \text{if } z = 1, \\ 1/32 & \text{if } z = 2, \\ 3/8 & \text{if } z = 3, \\ 3/16 & \text{if } z = 4, \\ 9/32 & \text{if } z = 6, \\ 0 & \text{otherwise.} \end{cases}$$

In this example, each (x,y) mapped to exactly one value of Z, but this does not have to be the case in general. Now the variance can be calculated as:

$$\text{var}(Z) = \frac{1}{8}(1 - \frac{15}{4})^2 + \frac{1}{32}(2 - \frac{15}{4})^2 + \frac{3}{8}(3 - \frac{15}{4})^2 + \frac{3}{16}(4 - \frac{15}{4}) + \frac{9}{32}(6 - \frac{15}{4})^2 = \frac{43}{16}.$$

6. For each lecture i, let Z_i be the random variable associated with the number of questions May gets asked plus two times the number May gets right. Also, for each lecture i, let D_i be the random variable $1000 + 40Z_i$. Let S be her semesterly salary. Because she teaches a total of 20 lectures, we have

$$S = \sum_{i=1}^{20} D_i = \sum_{i=1}^{20} 1000 + 40Z_i = 20000 + 40\sum_{i=1}^{20} Z_i.$$

By linearity of expectations,

$$\mathbb{E}[S] = 20000 + 40\mathbb{E}[\sum_{i=1}^{20} Z_i] = 20000 + 40(20)\mathbb{E}[Z_i] = 23000.$$

Since each of the D_i are independent, we have

$$\text{var}(S) = \sum_{i=1}^{20} \text{var}(D_i) = 20\text{var}(D_i) = 20\text{var}(1000 + 40Z_i) = 20(40^2\text{var}(Z_i)) = 36000.$$

7. Let Y be the number of questions she will answer wrong in a randomly chosen lecture. We can find $\mathbb{E}[Y]$ by conditioning on whether the lecture is in math or in science. Let M be the event that the lecture is in math, and let S be the event that the lecture is in science. Then

$$\mathbb{E}[Y] = \mathbb{E}[Y|M]\mathbb{P}(M) + \mathbb{E}[Y|S]\mathbb{P}(S).$$

Since there are an equal number of math and science lectures and we are choosing randomly among them, $\mathbb{P}(M) = \mathbb{P}(S) = \frac{1}{2}$. Now we need to calculate $\mathbb{E}[Y|M]$ and $\mathbb{E}[Y|S]$ by finding the respective conditional PMFs first. The PMFs can be determined in an manner analagous to how we calculated the PMF for the number of correct answers in part (c).

$$p_{Y|S}(y) = \begin{cases} \frac{1}{2}\frac{3}{4} + \frac{1}{2}(\frac{3}{4})^2 = 21/32 & \text{if } y = 0, \\ \frac{1}{2}\frac{1}{4} + \frac{1}{2}2\frac{1}{4}\frac{3}{4} = 5/16 & \text{if } y = 1, \\ \frac{1}{2}0 + \frac{1}{2}(\frac{1}{4})^2 = 1/32 & \text{if } y = 2, \\ 0 & \text{otherwise.} \end{cases}$$

$$p_{Y|M}(y) = \begin{cases} \frac{1}{2}\frac{9}{10} + \frac{1}{2}(\frac{9}{10})^2 = 171/200 & \text{if } y = 0, \\ \frac{1}{2}\frac{1}{10} + \frac{1}{2}2\frac{1}{10}\frac{9}{10} = 7/50 & \text{if } y = 1, \\ \frac{1}{2}0 + \frac{1}{2}(\frac{1}{10})^2 = 1/200 & \text{if } y = 2, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore

$$\mathbb{E}[Y|S] = 0\frac{21}{32} + 1\frac{5}{16} + 2\frac{1}{32} = \frac{3}{8},$$

$$\mathbb{E}[Y|M] = 0\frac{171}{200} + 1\frac{7}{50} + 2\frac{1}{200} = \frac{3}{20}.$$

This implies that

$$\mathbb{E}[Y] = \frac{3}{20}\frac{1}{2} + \frac{3}{8}\frac{1}{2} = \frac{21}{80}.$$

Problem 4.5

Chuck will go shopping for probability books for K hours. Here, K is a random variable and is equally likely to be 1, 2, 3, or 4. The number of books N that he buys is random and depends on how long he shops. We are told that

$$p_{N|K}(n \mid k) = \frac{1}{k}, \qquad \text{for } n = 1, \ldots, k.$$

1. Find the joint PMF of K and N.
2. Find the marginal PMF of N.
3. Find the conditional PMF of K given that $N = 2$.
4. We are now told that he bought at least 2 but no more than 3 books. Find the conditional mean and variance of K, given this piece of information.

Solution:
We are given the following information:

$$p_K(k) = \begin{cases} 1/4, & \text{if } k = 1,2,3,4; \\ 0, & \text{otherwise} \end{cases}$$

$$p_{N|K}(n \mid k) = \begin{cases} 1/k, & \text{if } n = 1,\ldots,k; \\ 0, & \text{otherwise} \end{cases}$$

(a) We use the fact that $p_{N,K}(n,k) = p_{N|K}(n \mid k)p_K(k)$ to arrive at the following joint PMF:

$$p_{N,K}(n,k) = \begin{cases} 1/(4k), & \text{if } k = 1,2,3,4 \text{ and } n = 1,\ldots,k; \\ 0, & \text{otherwise} \end{cases}$$

(b) The marginal PMF $p_N(n)$ is given by the following formula:

$$p_N(n) = \sum_k p_{N,K}(n,k) = \sum_{k=n}^{4} \frac{1}{4k}$$

On simplification this yields

$$p_N(n) = \begin{cases} 1/4 + 1/8 + 1/12 + 1/16 = 25/48, & n = 1; \\ 1/8 + 1/12 + 1/16 = 13/48, & n = 2; \\ 1/12 + 1/16 = 7/48, & n = 3; \\ 1/16 = 3/48, & n = 4; \\ 0, & \text{otherwise.} \end{cases}$$

(c) The conditional PMF is

$$p_{K|N}(k \mid 2) = \frac{p_{N,K}(2,k)}{p_N(2)} = \begin{cases} 6/13, & k = 2; \\ 4/13, & k = 3; \\ 3/13, & k = 4; \\ 0, & \text{otherwise.} \end{cases}$$

(d) Let A be the event $2 \leq N \leq 3$. We first find the conditional PMF of K given A.

$$\begin{aligned}
p_{K|A}(k) &= \frac{\mathbb{P}(K=k,A)}{\mathbb{P}(A)} \\
\mathbb{P}(A) &= p_N(2)+p_N(3) = \frac{5}{12} \\
\mathbb{P}(K=k,A) &= \begin{cases} \frac{1}{8}, & k=2; \\ \frac{1}{12}+\frac{1}{12}, & k=3; \\ \frac{1}{16}+\frac{1}{16}, & k=4; \\ 0, & \text{otherwise} \end{cases} \\
p_{K|A}(k) &= \begin{cases} \frac{3}{10}, & k=2; \\ \frac{2}{5}, & k=3; \\ \frac{3}{10}, & k=4; \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

Because the conditional PMF of K given A is symmetric around $k=3$, we know $\boxed{\mathbb{E}[K \mid A]=3}$. We now find the conditional variance of K given A.

$$\begin{aligned}
var(K \mid A) &= \mathbb{E}[(K-\mathbb{E}[K \mid A])^2 \mid A] \\
&= \frac{3}{10}\cdot(2-3)^2+\frac{2}{5}\cdot 0+\frac{3}{10}\cdot(4-3)^2 \\
&= \boxed{\tfrac{3}{5}}
\end{aligned}$$

Problem 4.6

Every package of Bobak's favorite ramen noodles comes with a plastic figurine of one of the characters in Battlestar Galactica. There are c different character figurines, and each package is equally likely to contain any character. Bobak buys one package of ramen each day, hoping to collect one figurine for each character.

(a) In lecture, we saw that expectation is linear, so that $\mathbb{E}[f(X)] = f(\mathbb{E}[X])$ for any linear function $f(X) = aX + b$ of a random variable X. Prove that the expectation operator is also linear when applied to functions of multiple random variables: i.e., if $f(X_1,\ldots,X_n) = \sum_{i=1}^n a_i X_i + b$ for some real numbers $a_1, a_2, \ldots, a_n, b$, then $\mathbb{E}[f(X_1,\ldots,X_n)] = f(\mathbb{E}[X_1],\ldots,\mathbb{E}[X_n])$.

(b) Find the expected number of days which elapse between the acquisitions of the jth new character and the $(j+1)$th new character.

(c) Find the expected number of days which elapse before Bobak has collected all of the characters.

Solution:

(a)

$$\begin{aligned}
& \mathbb{E}[f(X_1,\ldots,X_n)] \\
= \; & \mathbb{E}[\sum_{i=1}^{n} a_i X_i + b] \\
= \; & \sum_{x_1,\ldots,x_n} (\sum_{i=1}^{n} a_i x_i + b) P_{X_1,\ldots,X_n}(x_1,\cdots,x_n) \\
= \; & \sum_{i=1}^{n} \Big(a_i \sum_{x_1,\ldots,x_n} x_i P_{X_1,\ldots,X_n}(x_1,\cdots,x_n) \Big) + b \Big(\sum_{x_1,\ldots,x_n} P_{X_1,\ldots,X_n}(x_1,\cdots,x_n) \Big) \\
= \; & \sum_{i=1}^{n} a_i \sum_{x_i} x_i \Big(\sum_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n} P_{X_1,\ldots,X_n}(x_1,\cdots,x_n) \Big) + b \\
= \; & \sum_{i=1}^{n} a_i \sum_{x_i} x_i P_{X_i}(x_i) + b \\
= \; & \sum_{i=1}^{n} a_i \mathbb{E}(X_i) + b \\
= \; & f(\mathbb{E}(X_1),\ldots,\mathbb{E}(X_n))
\end{aligned}$$

(b) Let X_j be the number of days between the acquisitions of the jth new character and the $(j+1)$th new character. There are $c-j$ new characters out of c characters. Every package may have one of $c-j$ new characters with probability $(c-j)/c$. Thus, X_j follows a geometric distribution with $p = (c-j)/c$. $E(X_j) = 1/p = c/(c-j)$

(c) The sum of $X_0, X_1, \ldots, X_{c-1}$ is the random variable denoting the number of days that all characters are collected. It is a linear function over c random variables. According to (a), its expectation is

$$\begin{aligned}
& \mathbb{E}(X_0 + \ldots + X_{c-1}) \\
= \; & \mathbb{E}(X_0) + \ldots + \mathbb{E}(X_{c-1}) \\
= \; & \frac{c}{c} + \frac{c}{c-1} + \ldots + \frac{c}{2} + \frac{c}{1} \\
= \; & c(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{c}) \\
\approx \; & c\ln(c)
\end{aligned}$$

For full credit, you only needed to specify the sum $\sum_{j=0}^{c-1} \frac{c}{c-j}$.