

## Solutions to Assignment #14

1. [Problem 6.2.6 on page 238 in Allman and Rhodes]. Suppose a family has six children.

- (a) What is the probability that four are boys and two are girls?

*Solution:* Assume that having a boy or having a girl are equally likely events for this family. It then follows that the probability that a child is a boy is  $p = \frac{1}{2}$ . If  $X$  is the number of boys in the family, then  $X$  has a binomial distribution with parameters  $p = \frac{1}{2}$  and  $N = 6$ . Hence,

$$P[X = n] = \binom{N}{n} p^n (1-p)^{N-n} = \binom{6}{n} \left(\frac{1}{2}\right)^6 \quad (1)$$

for  $n = 0, 1, \dots, 6$ .

The probability of four boys and two girls is then

$$P[X = 4] = \binom{6}{4} \frac{1}{2^6} = \frac{15}{64}. \quad \square$$

- (b) Give the probability distribution for  $X$ .

*Solution:* This was given in equation (1).  $\square$

- (c) What is the expected number of boys in the family?

*Solution:*  $E(X) = pN = \frac{1}{2} \cdot 6 = 3$ .  $\square$

- (d) What is the probability that the family has four or more girls?

*Solution:* This is the same as the probability that the family two or fewer than two boys:

$$P[X \leq 2] = P[X = 0] + P[X = 1] + P[X = 2].$$

Thus, by the probability distribution function of  $X$  in (1),

$$\begin{aligned}P[X \leq 2] &= \binom{6}{0} \frac{1}{2^6} + \binom{6}{1} \frac{1}{2^6} + \binom{6}{2} \frac{1}{2^6} \\&= \frac{1}{64} [1 + 6 + 15] \\&= \frac{22}{64} \\&= \frac{11}{32}\end{aligned}$$

2. [Problem 6.2.16 on page 240 in Allman and Rhodes]. In humans, the hereditary Huntington disease is caused by a dominant mutation. Onset of Huntington disease occurs in midlife, between 35 and 44 years of age typically, and the progressive disorder leads eventually to death. Suppose that in a married couple one individual carries the allele for Huntington disease, and that the couple has four children.

- (a) What is the probability that none of the children will develop Huntington disease?

*Solution:* To do this problem first we model the number of children,  $X$ , that develops the disease by a binomial random variable with parameters  $p$  and  $N = 4$ , where  $p$  is the probability that an offspring carries the Huntington allele,  $H$ . If we assume the genotype of the parents are  $Hh$  and  $hh$ , respectively, then  $p$  is the probability that an offspring has the genotype  $Hh$ . Thus,

$$p = P[\text{allele } H \text{ from 1}^{\text{st}} \text{ parent}] \cdot P[\text{allele } h \text{ from 2}^{\text{nd}} \text{ parent}] = \frac{1}{2} \cdot 1 = \frac{1}{2}.$$

It then follows that

$$P[X = n] = \binom{4}{n} \left(\frac{1}{2}\right)^4 \quad (2)$$

for  $n = 0, 1, 2, 3, 4$ .

Thus, the probability that none of the children will develop Huntington disease is

$$P[X = 0] = \binom{4}{0} \frac{1}{2^4} = \frac{1}{16}. \quad \square$$

- (b) What is the probability that at least one of the children will develop Huntington disease?

*Solution:* The event [at least one of the children will develop Huntington] is the complement of the event [none of the children will develop Huntington]. Thus, by the previous part,

$$P[X \geq 1] = 1 - P[X = 0] = 1 - \frac{1}{16} = \frac{15}{16} \quad \text{or } 93.75\% \quad \square.$$

- (c) what is the probability that three or more of the children will develop Huntington disease?

*Solution:* This is  $P[X \geq 3] = P[X = 3] + P[X = 4]$ . Thus, by (2),

$$P[X \geq 3] = \binom{4}{3} \frac{1}{16} + \binom{4}{4} \frac{1}{16} = 4 \cdot \frac{1}{16} + \frac{1}{16} = \frac{5}{16} \quad \text{or } 31.25\% \quad \square.$$

3. [Problem 6.2.18 on pages 240 and 241 in Allman and Rhodes]. The goal of this exercise is to derive the formula

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (3)$$

for counting combinations of  $n$  things take  $k$  at a time.

Formally, a combination of  $n$  things taken  $k$  at a time is an *unordered* subset of the set of  $n$  objects consisting of  $k$  elements. However, it is better to think of it more concretely as follows. Imagine a box  $n$  balls with the numbers  $1, 2, 3, \dots, n$  printed on them. Pick  $k$  of the balls and place them in a row in the order in which you picked them. Then, since the *order* in which you picked them does not matter, put them in a bag. What you end up with in the bag is a combination.

The number of different bags of balls is  $\binom{n}{k}$ .

- (a) When you pick the first ball out of the box, how many different choices could you make for it? When you pick the second ball, why are there  $n - 1$  different choices for it? For the  $l^{\text{th}}$  ball, why are there  $n - l + 1$  choices?

*Solution:* Since there are  $n$  different balls in the ball to begin with, there are  $n$  different choices for the first pick. Once this one is chosen and taken out of the box, there remain  $n - 1$  different choices in it, and so there are  $n - 1$  different choices for the second pick. Similarly, for the third pick there would be  $n - 2$  choices; for the fourth pick,  $n - 3$  choices. Continuing in this fashion, we see that for the  $l$  pick there would be  $n - (l - 1)$  or  $n - l + 1$  choices.  $\square$

- (b) Why does part (a) indicate that, when the  $k$  balls are all in a row, there are  $n(n - 1)(n - 2) \cdots (n - k + 1)$  possible choices you might have made? (The count of these *ordered* choices is sometimes called a *permutation*.)

*Solution:* For each choice in the first pick there are  $n - 1$  choices in the second pick; thus, there are  $n(n - 1)$  choices for the first two balls. For each one of these there are  $n - 2$  choices for the third pick, and so there are  $n(n - 1)(n - 2)$  possibilities for the first three balls. Continuing in this fashion we then see that there are  $n(n - 1)(n - 2) \cdots (n - k + 1)$  for choosing  $k$  balls.  $\square$

- (c) Several different *ordered choices* might lead to the same collections of balls in the bag (i.e., to the same *combination*), so the answer to part (b) is bigger than the number of combinations. To see how much bigger, it's easiest to imagine having the balls in the bag, and (going backwards in

time) putting them back in order in a row. Using reasoning similar to that used in parts (a) and (b), explain why there are  $k!$  choices of ways this could be done.

*Solution:* When there are  $k$  balls in the bag, there are  $k$  different choices for the first ball in the row. Once this first ball is picked, for each of these choices, there are  $k - 1$  different ways of picking the second ball put of the bag; thus, there are  $k(k - 1)$  choices for the first to balls in the row. Similarly there are  $k(k - 1)(k - 2)$  for the first three balls in the row. Hence, continuing in this fashion, there are  $k(k - 1)(k - 2) \cdots 2 \cdot 1 = k!$  choices for putting back the balls from the bag in a row.  $\square$

(d) Using parts (b) and (c), conclude that

$$\binom{n}{k} = \frac{n(n - 1)(n - 2) \cdots (n - k + 1)}{k!}. \quad (4)$$

*Solution:* The total number of different bags of  $k$  balls chosen out of  $n$  balls in the box is equal to the number of ways of getting the  $k$  balls out of the box in the first place and putting them in the order of pick in a row divided by the different ways of ordering the  $k$  balls in the row, namely the number of permutations of the  $k$  balls in the row. Since there are  $n(n - 1)(n - 2) \cdots (n - k + 1)$  of picking  $k$  balls out of the box and putting them in row by part (b), the result follows.  $\square$

(e) Explain why the formula in (4) can be written as the formula in (3).

*Solution:* Observe that  $n! = n(n - 1)(n - 2) \cdots (n - k + 1) \cdot (n - k)!$ , so that

$$\frac{n!}{(n - k)!} = n(n - 1)(n - 2) \cdots (n - k + 1).$$

It then follows that

$$\frac{n(n - 1)(n - 2) \cdots (n - k + 1)}{k!} = \frac{n!}{k!(n - k)!}. \quad \square$$

4. [Problems 6.2.19 and 6.2.20 on pages 241 and 242 in Allman and Rhodes].

[Problems 6.2.19]

The binomial distribution received its name because of its relationship to the expression  $(x + y)^n$ , a power of a binomial. In fact, the numbers  $\binom{n}{k}$  are often called the *binomial coefficients*, because they give the coefficients in the expansion for  $(x + y)^n$ . That is

$$(x + y)^n = \binom{n}{0}y^n + \binom{n}{1}xy^{n-1} + \dots + \binom{n}{k}x^k y^{n-k} + \dots + \binom{n}{n}x^n. \quad (5)$$

- (a) Check formula for  $n = 2, 3$ , and 4 using (3).

*Solution:* For  $n = 2$ ,

$$(x + y)^2 = y^2 + 2xy + x^2.$$

Observe that  $\binom{2}{1} = 2$ .

For  $n = 3$ ,

$$(x + y)^3 = y^3 + 3xy^2 + 3x^2y + x^3.$$

Observe that  $\binom{3}{1} = 3$  and  $\binom{3}{2} = 3$ .

For  $n = 4$ ,

$$(x + y)^4 = y^4 + 4xy^3 + 6x^2y^2 + 4x^3y + x^4.$$

Observe that  $\binom{4}{1} = 4$  and  $\binom{4}{3} = 4$ . For the coefficient in the middle, compute

$$\binom{4}{2} = \frac{4!}{2!2!} = 2 \cdot 3 = 6. \quad \square$$

- (b) By thinking of  $(x + y)^n$  as a product of  $n$  copies of  $(x + y)$ , explain why this product will produce a term  $x^k y^{n-k}$  for each way we can choose  $k$  of the copies. Explain why this justifies formula (5).

*Solution:* Write  $(x + y)^n = (x + y)(x + y)(x + y) \cdots (x + y)$ . When we expand this expression, we get monomials of the form

$$x^k y^{n-k}.$$

These come about by picking an  $x$  from  $k$  of the binomials  $(x + y)$ , and  $n - k$  factors of  $y$  from the remaining ones. There are  $\binom{n}{k}$  ways of making these choices. Thus, the terms

$$\binom{n}{k} x^k y^{n-k}$$

appear in the expansion of  $(x + y)^n$  as stated in formula (5).  $\square$

- (c) What is the sum  $\sum_{i=0}^3 \binom{3}{i}$ ?  $\sum_{i=0}^4 \binom{4}{i}$ ? Give a formula for  $\sum_{i=0}^n \binom{n}{i}$ .

*Solution:*  $\sum_{i=0}^3 \binom{3}{i} = \sum_{i=0}^3 \binom{3}{i} (1)^i (1)^{3-i} = (1 + 1)^3$ , by formula (5) with

$x = y = 1$  and  $n = 3$ . It then follows that  $\sum_{i=0}^3 \binom{3}{i} = 2^3 = 8$ .

Similarly,  $\sum_{i=0}^4 \binom{4}{i} = 2^4 = 16$  and  $\sum_{i=0}^n \binom{n}{i} = 2^n$ .  $\square$

[Problems 6.2.20]

Suppose a trial has two possible independent outcomes: a success or a failure. If the probability of a success is  $p$  then the number of successes,  $X$ , in  $n$  trials has a binomial distribution with parameters  $p$  and  $n$ . Show that  $E(X) = np$ ; that is the expected value for the number of successes in  $n$  trials is  $np$ . Proceed as follows:

- (a) Express the expected value of  $X$  as a sum of factorials and powers of  $p$  and  $q$ , where  $q = 1 - p$ .

*Solution:*  $E(X) = \sum_{i=0}^n i \cdot P[X = i]$ , where

$$P[X = i] = \binom{n}{i} p^i q^{n-i} \quad \text{for } i = 0, 1, 2, \dots, n.$$

Thus,

$$E(X) = \sum_{i=1}^n i \cdot \frac{n!}{i!(n-i)!} p^i q^{n-i}, \quad (6)$$

since the term for  $i = 0$  does not contribute to the sum.  $\square$

(b) Show that

$$i \cdot \frac{n!}{i!(n-i)!} p^i q^{n-i} = pn \cdot \frac{(n-1)!}{(i-1)!(n-i)!} p^{i-1} q^{(n-1)-(i-1)}. \quad (7)$$

*Solution:* Factor an  $np$  and cancel  $i$  in the numerator with the  $i$  in  $i!$  in the denominator to get

$$i \cdot \frac{n!}{i!(n-i)!} p^i q^{n-i} = np \cdot \frac{(n-1)!}{(i-1)!(n-i)!} p^{i-1} q^{n-i},$$

and observe that  $n - i = (n - 1) - (i - 1)$ .  $\square$

(c) Use part (b) to factor  $pn$  from the expression in part (a). Then, use equation (5) to complete the problem.

*Solution:* Using (6) and (7) we obtain

$$\begin{aligned} E(X) &= \sum_{i=1}^n i \cdot \frac{n!}{i!(n-i)!} p^i q^{n-i} \\ &= pn \sum_{i=1}^n \frac{(n-1)!}{(i-1)!(n-i)!} p^{i-1} q^{(n-1)-(i-1)}. \end{aligned}$$

Now make the change of variables  $k = i - 1$  in the last sum to get

$$E(X) = pn \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} p^k q^{(n-1)-k}.$$

Thus, by (5) with  $x = p$ ,  $y = q$  and  $n - 1$  in place of  $n$ ,

$$E(X) = pn(p+q)^{n-1} = pn. \quad \square$$

5. [Problem 6.2.21 on page 242 in Allman and Rhodes]. The goal of this problem is to show that expected values of random variables are additive; that is, if  $X_1$  and  $X_2$  are two random variable defined on the same sample space,

$$E(X_1 + X_2) = E(X_1) + E(X_2). \quad (8)$$

Only the special case in which  $X_1$  and  $X_2$  are independent random variables will be considered. For simplicity, assume that both  $X_1$  and  $X_2$  take on only integer values between 1 and  $N$ .

- (a) Explain why the expected value of  $X_1 + X_2$  is

$$E(X_1 + X_2) = \sum_{i=1}^N \sum_{j=1}^N (i + j)P[X_1 = i]P[X_2 = j]. \quad (9)$$

*Solution:* Observe that  $X_1 + X_2 = m$  if and only if  $X_1 = i$  and  $X_2 = m - i$ . Thus

$$\begin{aligned} P[X_1 + X_2 = m] &= P[X_1 = i \text{ and } X_2 = m - i] \\ &= P[X_1 = i]P[X_2 = m - i \mid X_1 = i]. \end{aligned}$$

Now, since  $X_1$  and  $X_2$  are independent,

$$P[X_2 = m - i \mid X_1 = i] = P[X_2 = m - i].$$

Hence,

$$P[X_1 + X_2 = m] = P[X_1 = i]P[X_2 = m - i].$$

Possible values of  $m$  are of the form  $m = i + j$ , where  $i$  and  $j$  range from 1 to  $N$ . It then follows that

$$E(X_1 + X_2) = \sum_{\text{values of } m} mP[X_1 + X_2 = m] = \sum_{i=1}^N \sum_{j=1}^N (i + j)P[X_1 = i]P[X_2 = j]$$

and this is (9).  $\square$

- (b) Through algebra, show that the sum on the right-hand side of (9) can be written as

$$\sum_{i=1}^N iP[X_1 = i] \sum_{j=1}^N P[X_2 = j] + \sum_{j=1}^N jP[X_2 = j] \sum_{i=1}^N P[X_1 = i].$$

*Solution:* Multiplying out the summand in the sum in (9), we get that this can be written as the sum of two sums:

$$\sum_{i=1}^N \sum_{j=1}^N iP[X_1 = i]P[X_2 = j] + \sum_{i=1}^N \sum_{j=1}^N jP[X_1 = i]P[X_2 = j].$$

These can in turn be re-written as

$$\sum_{i=1}^N iP[X_1 = i] \sum_{j=1}^N P[X_2 = j] + \sum_{i=1}^N P[X_1 = i] \sum_{j=1}^N jP[X_2 = j],$$

which yields the result we wanted to show.  $\square$

- (c) What are  $\sum_{i=1}^N P[X_1 = i]$  and  $\sum_{j=1}^N P[X_2 = j]$ ? Use these results to conclude that (8) holds.

*Solution:*  $\sum_{i=1}^N P[X_1 = i]$  is the probability that  $X_1$  takes on at least one of its values, and this is 1. Similarly,  $\sum_{j=1}^N P[X_2 = j] = 1$ . It then follows from (9) and what we just proved in the previous part that

$$E(X_1 + X_2) = \sum_{i=1}^N iP[X_1 = i] + \sum_{j=1}^N jP[X_2 = j] = E(X_1) + E(X_2),$$

and this is (8).