7.5 Applications of the Binomial Distribution

MATHEMATICAL TECHNIQUES

Suppose that the allele A for height is dominant, meaning that plants with genotypes AA and Aa are tall, while those with genotype aa are short. If an Aa plant is crossed with another Aa plant, $3/4$ of the offspring should be tall. Assuming that the other conditions for the binomial distribution are met, find the probabilities of the following.

- **EXERCISE 7.5.1**
  
  Exactly 3 out of 4 offspring are tall.

- **EXERCISE 7.5.2**
  
  Exactly 6 out of 8 offspring are tall.

- **EXERCISE 7.5.3**
  
  Fewer than 1 out of 4 offspring are tall.

- **EXERCISE 7.5.4**
  
  Fewer than 2 out of 8 offspring are tall.

When there are more than two outcomes of a trial, the distribution of all possibilities is described by the **multinomial distribution**. Consider an additive pair of alleles A and a, where an offspring of a cross between two Aa individuals is tall (with genotype Aa) with probability 0.25, is intermediate (with genotype Aa) with probability 0.5, and is short (with genotype aa) with probability 0.25. Count up all possible ways for the following to happen and find the associated probabilities.

- **EXERCISE 7.5.5**
  
  Out of 3 offspring, 1 is tall, 1 is intermediate, and 1 is short.

- **EXERCISE 7.5.6**
  
  Out of 4 offspring, 1 is tall and 2 are intermediate.

- **EXERCISE 7.5.7**
  
  Out of 4 offspring, 2 are tall and 2 are intermediate.

- **EXERCISE 7.5.8**
  
  Out of 4 offspring, 1 is tall, 2 are intermediate, and 1 is short.

There is a formula for the multinomial distribution (studied in 6.15.5–6.15.8) describing probabilities when there are more than two outcomes of a trial. Suppose there are 3 possible outcomes of each trial, numbered 1 through 3, with probabilities $p_1$, $p_2$, and $p_3$. If there are $n$ trials, the probability that there are exactly $k_1$ outcomes of the first type, $k_2$ outcomes of the second type, and $k_3$ outcomes of the third type is

$$M(k_1, k_2, k_3; n, p_1, p_2, p_3) = \frac{n!}{k_1!k_2!k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}$$

Consider the genetics described in 6.15.5–6.15.8 describing an additive pair of alleles A and a. Use the formula for the multinomial distribution to compute the following probabilities and compare with the earlier result.

- **EXERCISE 7.5.9**
  
  Out of 3 offspring, 1 is tall, 1 is intermediate, and 1 is short (as in exercise 6.15.5).

- **EXERCISE 7.5.10**
  
  Out of 3 offspring, 1 is tall and 2 are intermediate (as in exercise 6.15.6).

- **EXERCISE 7.5.11**
  
  Out of 4 offspring, 2 are tall and 2 are intermediate (as in exercise 6.15.7).

- **EXERCISE 7.5.12**
  
  Out of 4 offspring, 1 is tall, 2 are intermediate, and 1 is short (as in exercise 6.15.8).

APPLICATIONS

Suppose that the alleles A and a for height are additive, meaning that plants with genotype AA are tall, plants with genotype Aa are intermediate, and those with genotype aa are short. If an Aa plant is crossed with another Aa plant, $1/4$ of the offspring should be tall, $1/2$ should be intermediate, and $1/4$ should be short. Assuming
that the other conditions for the binomial distribution are met, find the probabilities of the following. Suppose such a cross produces 6 offspring.

- **EXERCISE 7.5.13**
  Find the expectation and the mode of the number of tall offspring.

- **EXERCISE 7.5.14**
  Find the expectation and the mode of the number of intermediate offspring.

- **EXERCISE 7.5.15**
  Find the probability that the number of tall offspring is less than or equal to the mode.

- **EXERCISE 7.5.16**
  Find the probability that the number of intermediate offspring is less than or equal to the mode.

Consider one set of 10 islands, each which has a 0.2 chance of switching from empty to occupied, and a 0.1 chance of switching from occupied to empty equations 6.3 and 6.4. The equilibrium fraction occupied is 2/3. Consider a second set of 10 islands which are all occupied if the weather is good (probability 2/3) and all empty if the weather is bad (probability 1/3). Compute the following for both sets of islands.

- **EXERCISE 7.5.17**
  The probability that exactly 7 out of 10 are occupied.

- **EXERCISE 7.5.18**
  The probability that all 10 are occupied.

- **EXERCISE 7.5.19**
  Find the mean number of islands occupied.

- **EXERCISE 7.5.20**
  Find the variance of the number of islands occupied.

- **EXERCISE 7.5.21**
  Find the mode of the number of islands occupied.

- **EXERCISE 7.5.22**
  Sketch the probability distribution for each set of islands. Why does the second set fail to follow the binomial distribution?

- **EXERCISE 7.5.23**
  Consider the mutate genes described in exercise 6.2.19, where a gene has a 1.0% chance of mutating each time a cell divides, and a 1.0% chance of correcting the mutation. Suppose that 4 genes start out normal. Find the probability that there are more than 2 mutants after 1 division, after 2 divisions, and after a long time.

- **EXERCISE 7.5.24**
  Consider the lemmings described in exercise 6.2.20, where a lemming has a probability 0.2 of jumping off the cliff each hour and a probability 0.1 of crawling back up. Suppose that 5 lemmings start at top. Find the probability that more of the lemmings are at the bottom after 1 hour, after 2 hours, and after a long time.

- **EXERCISE 7.5.25**
  Starting with 5 molecules, each leaving with probability 0.8 per minute, compute and graph the probability distribution describing the number remaining at the following times.

  - **EXERCISE 7.5.26**
    1 minute.
  - **EXERCISE 7.5.27**
    2 minutes.
  - **EXERCISE 7.5.28**
    5 minutes.
  - **EXERCISE 7.5.29**
    10 minutes.

- **EXERCISE 7.5.26**
  Starting with 5 molecules, each leaving with probability 0.8 per minute never to return, find and graph the following probabilities as functions of time.

  - **EXERCISE 7.5.29**
    Exactly 1 remains. At what time is this probability a maximum?

  - **EXERCISE 7.5.30**
    Exactly 2 remain. At what time is this probability a maximum?
Suppose that 10 independent experiments are run in which 5 molecules begin inside a cell, and leave with probability 0.8 each minute and never return.

- **EXERCISE 7.5.31**
  Using the results in exercise 6.15.25, find the probability that exactly 3 out of 10 such experiments have exactly 4 molecules after one minute.

- **EXERCISE 7.5.32**
  Using the results in exercise 6.15.27, find the probability that exactly 5 out of 10 such experiments have exactly 1 molecule after five minutes.

- 40 molecules begin inside a cell. Each leaves independently with probability 0.8 each minute.
  
  - **EXERCISE 7.5.33**
    Find the expected number remaining inside as a function of time.
  
  - **EXERCISE 7.5.34**
    Find the mode at \( t = 1, t = 2 \) and \( t = 3 \). Indicate whether the mode is equal to, greater than, or less than the mean.
  
  - **EXERCISE 7.5.35**
    Compute the variance. At what time is it a maximum?
  
  - **EXERCISE 7.5.36**
    Find the coefficient of variation of the number remaining inside as a function of time. Is it an increasing or a decreasing function?

- Figure 6.3b, illustrating stochastic immigration, was generated by adding 2 individuals with probability 0.5, and 0 with probability 0.5 for 100 generations. The results in the figure show final populations of 106 and 96.
  
  - **EXERCISE 7.5.37**
    How can these results be described in terms of the binomial distribution?
  
  - **EXERCISE 7.5.38**
    What is the expected number after 100 generations?
  
  - **EXERCISE 7.5.39**
    What is the variance after 100 generations?
  
  - **EXERCISE 7.5.40**
    Write the probability of exactly 106 in terms of the binomial distribution.

- Unbeknownst to the experimenter, a cell contains two different types of molecule, one which is inside with probability \( p_1 \) and the other which is inside with probability \( p_2 \). Suppose there are 2 of each type of molecule.
  
  - **EXERCISE 7.5.41**
    Suppose \( p_1 = 0 \) and \( p_2 = 1 \). Find and graph the probability distribution for the total number inside. Find the expectation and the variance.
  
  - **EXERCISE 7.5.42**
    Suppose \( p_1 = 0.25 \) and \( p_2 = 0.75 \). Find and graph the probability distribution for the total number inside. Find the expectation and the variance (this can be written as the the sum of two binomial random variables).
  
  - **EXERCISE 7.5.43**
    Compare with the results if \( p_1 = p_2 = 0.5 \).
  
  - **EXERCISE 7.5.44**
    It turns out that all four molecules are different, and that \( p_1 = 0, p_2 = 0.25, p_3 = 0.75 \) and \( p_4 = 1 \). Find and graph the probability distribution for the total number inside. Find the expectation and the variance.
Chapter 7

Answers

7.5.1. \( b(3; 4, 0.75) = \binom{4}{3} 0.75^3 \cdot 0.25 = 4 \cdot 0.75^3 \cdot 0.25^1 = 0.422 \).

7.5.3. We must add the probability that 1 is tall and the probability that 0 are tall. \( b(1; 4, 0.75) = \binom{4}{1} 0.75^1 \cdot 0.25^3 = 4 \cdot 0.75^1 \cdot 0.25^3 = 0.0469 \). \( b(0; 4, 0.75) = \binom{4}{0} 0.75^0 \cdot 0.25^4 = 1 \cdot 0.75^0 \cdot 0.25^4 = 0.0039 \). The total is 0.0508.

7.5.5. If we list the possible orders, we find 6 \(\{\text{TIS, TSI, ITS, IST, STI, SIT}\}\). The probability of each of these is \(0.25^1 \cdot 0.5^1 \cdot 0.25^3 = 0.0312\), for a total of \(6 \cdot 0.0312 = 0.1872\).

7.5.7. As with the binomial distribution, there are 6 possible ways to choose 2 tall and 2 intermediates out of 4 \(\{\text{TITI, TITT, TITT, ITIT, ITIT, ITTT}\}\). The probability of each of these is \(0.25^2 \cdot 0.5^2 = 0.0156\), for a total of \(6 \cdot 0.0156 = 0.0936\).

7.5.9. In this case, \(n = 3, k_1 = 1, k_2 = 1, k_3 = 1, p_1 = 0.25, p_2 = 0.5\) and \(p_3 = 0.25\). Then the probability is

\[
M(1, 1; 3, 0.25, 0.5, 0.25) = \frac{3!}{1!1!1!} 0.25^1 0.5^1 0.25^1 = 0.1872.
\]

This matches the answer in exercise 6.15.5.

7.5.11. In this case, \(n = 4, k_1 = 2, k_2 = 2, k_3 = 0, p_1 = 0.25, p_2 = 0.5\) and \(p_3 = 0.25\). Then the probability is

\[
M(2, 2; 4, 0.25, 0.5, 0.25) = \frac{4!}{2!2!0!} 0.25^2 0.5^2 0.25^0 = 0.0936.
\]

This matches the answer in exercise 6.15.7.

7.5.13. The expectation is \(np = 1.5\). The mode is the smallest integer bigger than \(np - 1 + p = 1.5 - 1 + 0.25 = 0.75\), so is equal to 1.

7.5.15. This is \(b(0, 6, 0.25) + b(1, 6, 0.25)\). \(b(1, 6, 0.25) = \binom{6}{1} 0.25^1 \cdot 0.75^5 = 6 \cdot 0.25^1 \cdot 0.75^5 = 0.356\). \(b(0; 6, 0.25) = \binom{6}{0} 0.25^0 \cdot 0.75^6 = 1 \cdot 0.25^0 \cdot 0.75^6 = 0.178\). The total is 0.534.

7.5.17. For the first set of islands, this is \(b(7; 10, 2/3) = 0.26\). The second set of islands are either all occupied or all empty, so the probability is 0.

7.5.19. For the first set of islands, the mean is \(np = 10 \cdot 2/3 = 6.67\). For the second set it is \(10 \cdot 2/3 + 0 \cdot \frac{1}{3} = 6.67\).

7.5.21. For the first set of islands, \(np - 1 + p = 6.67 + 1 + 0.67 = 6.33\). The mode is next integer, or 7. For the second set of islands, the mode is 10.

7.5.23. The probability that a gene is a mutant obeys the discrete-time dynamical system

\[ p_{t+1} = 0.99p_t + 0.01(1 - p_t) \]

with \(p_0 = 1\). Then \(p_1 = 0.99\) and \(p_2 = 0.9802\). We found that \(p_t\) approaches an equilibrium of 0.5 after a long time. At time 1, the number of mutants follows the binomial with \(n = 4\) and \(p = 0.99\). The probability of 2 mutants is 0.00059, of 3 mutants is tiny, and for 4 mutants is really tiny, for a total of about 0.00059. At time 2,
the number of mutants follows the binomial with \( n = 4 \) and \( p = 0.9802 \). The probability of 2 mutants is 0.0022, of 3 mutants is 0.00003, and for 4 mutants is tiny, for a total of about 0.0022. After a long time, half of the genes will be mutant, and the probability of 2 or more mutants becomes 0.6875.

**7.5.25.** At time 1, this is a binomial distribution with \( n = 5 \) and \( p = 0.8 \). The probabilities for 0 through 5 inside are 0.0003, 0.0064, 0.051, 0.205, 0.410 and 0.328.

![Probability distribution for t=1](image)

**7.5.27.** At time 5, this is a binomial distribution with \( n = 5 \) and \( p = 0.8^5 = 0.32 \). The probabilities for 0 through 5 inside are 0.137, 0.335, 0.326, 0.159, 0.0388, and 0.004.

![Probability distribution for t=5](image)

**7.5.29.** As a function of time, the probability that exactly 1 remains is \( b(1;5,0.8^t) = 5 \cdot 0.8^t(1-0.8^t)^4 \). We want to find the value of \( t \) that maximizes this expression. It is easier to replace \( 0.8^t \) with \( x \), and write the probability as \( f(x) = 5x(1-x)^4 \). Then \( f'(x) = 5(1-x)^4 - 20x(1-x)^3 \) which is equal to 0 when \( x = 1/5 = 0.2 \). Then \( t = \ln(0.2)/\ln(0.8) = 7.21 \).

![Probability exactly 1 remains](image)

**7.5.31.** The probability is \( b(3;10,0.410) = 0.205 \).

**7.5.33.** \( E(N_t) = 40 \cdot 0.8^t \).

**7.5.35.** The variance is \( 40 \cdot 0.8^t(1-0.8^t) \). This has a maximum at \( t = \ln(0.5)/\ln(0.8) = 3.106 \).

**7.5.37.** Let \( P \) be the random variable describing the number of immigrants. Because they arrive in pairs, \( P = 2I \), where \( I \) is a binomial random variable with \( p = 0.5 \) and \( n = 100 \).

**7.5.39.** \( \text{Var}(P) = \text{Var}(2I) = 2^2 \text{Var}(I) = 4 \cdot 100 \cdot 0.5 \cdot 0.5 = 100 \).

**7.5.41.** There must be two inside (both of the second type). The expectation is 2 and the variance is 0.
7.5.43. If both are the same, the total $N$ follows a binomial with $n = 4$ and $p = 0.5$. $E(N) = 2$ and $\text{Var}(N) = 0.5$. The average is the same, but the distribution is more spread out. When the molecules are different, it is more likely that exactly half will be inside.