Discussion. For S_n the sum of n digits the argument shows that the distribution of S_n is symmetric about $(4\frac{1}{2})n$ for every n. For odd n, say $n = 2m + 1$, this symmetry can be used just as above to identify a probability in the distribution of S_{2m+1} that is exactly 1/2:

$$P(S_{2m+1} \le 9m + 4) = P(S_{2m+1} \ge 9m + 5) = \frac{1}{2}$$

For odd n the histogram of S_n has bars of equal height at the integers $(4\frac{1}{2})n \pm 1/2$, $(4\frac{1}{2})n \pm 3/2, \ldots$, so the distribution splits perfectly into two equal halves. For even n the histogram of S_n has a bar exactly on the point of symmetry $(4\frac{1}{2})n$, and equal bars at $(4\frac{1}{2})n \pm 1$, $(4\frac{1}{2})n \pm 2, \ldots$. Then the distribution of S_n does not split into equal halves to the right and left of $(4\frac{1}{2})n$, because there is a lump of probability right on the point of symmetry which cannot be split in two. It can be shown that for even r the central probability $P[S_n = (4\frac{1}{2})n]$ is actually the largest individual probabilty ir the distribution of S_n. It will be seen in Section 3.3 that for large n the distributior of S_n follows a normal curve very closely. This is similar to what happens for large n to the binomial $(n, 1/2)$ distribution of $X_1 + \cdots + X_n$ for X_i picked at randorr from $\{0, 1\}$. It follows that as in the binomial case, for large even n the distributior of the sum of n digits has central term $P[S_n = (4\frac{1}{2})n]$ that converges to zero ver slowly, like a constant over $\sqrt{n}$. For very large $n = 2m$ this term can be ignored, sc

$$P(S_{2m} \le 9m) = P(S_{2m} \ge 9m) \approx \frac{1}{2}$$

The approximate probability $\frac{1}{2}$ is less than the true probability by

$$P(S_{2m} = 9m)/2 \sim c/\sqrt{m}$$

where the constant c can be shown using the normal approximation to be equal t $1/\sqrt{33\pi}$, and "$\sim$" means that the ratio of the two sides tends to 1 as $m \to \infty$. (Se Exercise 3.3.31).

Exercises 3.1

1. Let X be the number of heads in three tosses of a fair coin.

a) Display the distribution of X in a table. b) Find the distribution of $|X - 1|$.

2. Let X and Y be the numbers obtained in two draws at random from a box containir four tickets 1, 2, 3, and 4. Display the joint distribution table for X and Y:

a) for sampling with replacement; b) for sampling without replacement.

Calculate $P(X \le Y)$ from the table in each case.

3. Suppose a fair die is rolled twice. Let S be the sum of the numbers on the two rolls.

a) What is the range of S? b) Find the distribution of S.

4. Let X_1 and X_2 be the numbers obtained on two rolls of a fair die. Let $Y_1 = \max(X_1, X$ $Y_2 = \min(X_1, X_2)$. Display joint distribution tables for a) (X_1, X_2); b) (Y_1, Y_2).

5. Find the distribution of $X_1 X_2$ for X_1 and X_2 as in Exercise 4.

6. A fair coin is tossed three times. Let X be the number of heads on the first two tosses, Y the number of heads on the last two tosses.

   a) Make a table showing the joint distribution of X and Y.

   b) Are X and Y independent? c) Find the distribution of $X + Y$.

7. Let A, B, and C be events that are independent, with probabilities a, b, and c. Let N be the random number of events that occur.

   a) Express the event $(N = 2)$ in terms of A, B, and C. b) Find $P(N = 2)$.

8. A hand of five cards contains two aces and three kings. The five cards are shuffled and dealt one by one, until an ace appears.

   a) Display in a table the distribution of the number of cards dealt.

   b) Suppose that dealing is continued until the second ace appears. Again display the distribution of the number of cards dealt.

   c) Explain why the probabilities in the second table are just those in the first in a different order. (*Hint:* Think about dealing off the bottom of the deck!)

9. A box contains 8 tickets. Two are marked 1, two marked 2, two marked 3, and two marked 4. Tickets are drawn at random from the box without replacement until a number appears that has appeared before. Let X be the number of draws that are made. Make a table to display the probability distribution of X.

10. **Blocks of Bernoulli trials.** In $n + m$ independent Bernoulli (p) trials, let S_n be the number of successes in the first n trials, T_m the number of successes in the last m trials.

    a) What is the distribution of S_n? Why?

    b) What is the distribution of T_m? Why?

    c) What is the distribution of $S_n + T_m$? Why?

    d) Are S_n and T_m independent? Why?

11. **Binomial sums.** Let U_n have binomial(n, p) distribution and let V_m have binomial(m, p) distribution. Suppose U_n and V_m are independent.

    a) Find the distribution of $U_n + V_m$ without calculation by a simple argument that refers to the solution of Exercise 10.

    b) Compare the result of part a) to a calculation of $P(U_n + V_m = k)$ for $0 \le k \le n + m$ from the joint distribution of U_n and V_m, and hence prove the identity

$$\sum_{j=0}^{n} \binom{n}{j} \binom{m}{k-j} = \binom{n+m}{k}$$

    c) Derive the identity in part b) by a counting argument. [*Hint:* Classify the subsets of size k of $\{1, \ldots, n+m\}$ by how many elements of $\{1, \ldots, n\}$ they contain.]

    d) Derive the identity in part b) in another way by finding the coefficient of $p^k q^{n+m-k}$ in $(p+q)^{n+m} = (p+q)^n (p+q)^m$ in two different ways.

c) Simplify the sum $\sum_{j=0}^{n} \binom{n}{j}^2$.

12. Grouping multinomial categories. Suppose that counts $(N_1, \ldots, N_m)$ are the numbers of results in m categories in n repeated trials. So $(N_1, \ldots, N_m)$ has multinomial distribution with parameters n and $p_1, \ldots, p_m$, as in the box above Example 7. Let $1 \le i < j \le m$. Answer the following questions with an explanation, but no calculation.

a) What is the distribution of N_i? b) What is the distribution of $N_i + N_j$?

c) What is the joint distribution of N_i, N_j, and $n - N_i - N_j$?

13. A box contains $2n$ balls of n different colors, with 2 of each color. Balls are picked at random from the box with replacement until two balls of the same color have appeared. Let X be the number of draws made.

a) Find a formula for $P(X > k)$, $k = 2, 3, \ldots$

b) Assuming n is large, use an exponential approximation to find a formula for k in terms of n such that $P(X > k)$ is approximately $1/2$. Evaluate k for n equal to one million.

14. In a World Series, teams A and B play until one team has won four games. Assume that each game played is won by team A with probability p, independently of all previous games.

a) For $g = 4$ through 7, find a formula in terms of p and $q = 1 - p$ for the probability that team A wins in g games.

b) What is the probability that team A wins the World Series, in terms of p and q?

c) Use your formula to evaluate this probability for $p = 2/3$.

d) Let X be a binomial $(7, p)$ random variable. Explain why $P(\text{A wins}) = P(X \ge 4)$ using an intuitive argument. Verify algebraically that this is true.

e) Let G represent the number of games played. What is the distribution of G? For what value of p is G independent of the winner of the series?

15. Let X and Y be independent, each uniformly distributed on $\{1, 2, \ldots, n\}$. Find:

a) $P(X = Y)$; b) $P(X < Y)$; c) $P(X > Y)$;

d) $P(\max(X, Y) = k)$ for $1 \le k \le n$;

e) $P(\min(X, Y) = k)$ for $1 \le k \le n$; f) $P(X + Y = k)$ for $2 \le k \le 2n$.

16. Discrete convolution formula. Let X and Y be independent random variables with non-negative integer values. Show that:

a) $P(X + Y = n) = \sum_{k=0}^{n} P(X = k)P(Y = n - k)$.

b) Find the probability that the sum of numbers on four dice is 8, by taking X to be the sum on two of the dice, Y the sum on the other two.

17. Let X be the number of heads in 20 fair coin tosses, Y a number picked uniformly at random from $\{0, 1, \ldots, 20\}$, independently of X. Let $Z = \max(X, Y)$.

a) Find a formula for $P(Z = k)$, $k = 0, \ldots, 20$.

b) Without calculating out $P(Z = k)$ exactly, sketch the histogram of Z, and explain its unusual shape.

18. Three dice are rolled.

a) What is the probability that the total number of spots showing is 11 or more? [*Hint*: No long calculations!]

b) Find a number m such that if five dice are rolled, the probability that the total number of spots showing is m or more is the same as this probability of 11 or more spots from three dice.

19. Sum of biased dice. Let S be the sum of numbers obtained by rolling two biased dice with possibly different biases described by probabilities $p_1, \ldots, p_6$, and $r_1, \ldots, r_6$, all assumed to be nonzero.

a) Find formulae for $P(S = k)$ for $k = 2, 7$, and 12.

b) Show that $P(S = 7) > P(S = 2)\frac{r_6}{r_1} + P(S = 12)\frac{r_1}{r_6}$.

c) Deduce that no matter how the two dice are biased, the numbers 2, 7, and 12 cannot be equally likely values for the sum. In particular, the sum cannot be uniformly distributed on the numbers from 2 to 12.

d) Do there exist positive integers a and b and independent non-constant random variables X and Y such that $X+Y$ has uniform distribution on the set of integers $\{a, a+1, \ldots, a+b\}$?

20. Pairwise independence. Let $X_1, \ldots, X_n$ be a sequence of random variables. Suppose that X_i and X_j are independent for every pair (i, j) with $1 \le i < j \le n$. Does this imply $X_1, \ldots, X_n$ are independent? Sketch a proof or counterexample.

21. Sequential independence. Let $X_1, \ldots, X_n$ be a sequence of random variables. Suppose that for every $1 \le m \le n-1$ the random sequence $(X_1, \ldots, X_m)$ is independent of the next random variable X_{m+1}. Does this imply $X_1, \ldots, X_n$ are independent? Sketch a proof or give a counterexample.

22. Suppose that random variables X and Y, each with a finite number of possible values, have joint probabilities of the form

$$P(X = x, Y = y) = f(x)g(y)$$

for some functions f and g, for all (x, y).

a) Find formulae for $P(X = x)$ and $P(Y = y)$ in terms of f and g.

b) Use your formulae to show that X and Y are independent.

23. Suppose X and Y are two random variables such that $X \ge Y$.

a) For a fixed number T, which would be greater, $P(X \le T)$ or $P(Y \le T)$?

b) What if T is a random variable?

24. Suppose a box contains tickets, each labeled by an integer. Let X, Y, and Z be the results of draws at random with replacement from the box: Show that, no matter what the distribution of numbers in the box,

a) $P(X + Y \text{ is even}) \ge 1/2$; b) $P(X + Y + Z \text{ is a multiple of } 3) \ge 1/4$.