PROBLEMS

SECTION 2.2. Probability Mass Functions

Problem 1. The MIT soccer team has 2 games scheduled for one weekend. It has a 0.4 probability of not losing the first game, and a 0.7 probability of not losing the second game, independent of the first. If it does not lose a particular game, the team is equally likely to win or tie, independent of what happens in the other game. The MIT team will receive 2 points for a win, 1 for a tie, and 0 for a loss. Find the PMF of the number of points that the team earns over the weekend.

Problem 2. You go to a party with 500 guests. What is the probability that exactly one other guest has the same birthday as you? Calculate this exactly and also approximately by using the Poisson PMF. (For simplicity, exclude birthdays on February 29.)

Problem 3. Fischer and Spassky play a chess match in which the first player to win a game wins the match. After 10 successive draws, the match is declared drawn. Each game is won by Fischer with probability 0.4, is won by Spassky with probability 0.3, and is a draw with probability 0.3, independent of previous games.

(a) What is the probability that Fischer wins the match?

(b) What is the PMF of the duration of the match?

Problem 4. An internet service provider uses 50 modems to serve the needs of 1000 customers. It is estimated that at a given time, each customer will need a connection with probability 0.01, independent of the other customers.

(a) What is the PMF of the number of modems in use at the given time?

(b) Repeat part (a) by approximating the PMF of the number of customers that need a connection with a Poisson PMF.

(c) What is the probability that there are more customers needing a connection than there are modems? Provide an exact, as well as an approximate formula based on the Poisson approximation of part (b).

Problem 5. A packet communication system consists of a buffer that stores packets from some source, and a communication line that retrieves packets from the buffer and transmits them to a receiver. The system operates in time-slot pairs. In the first slot, the system stores a number of packets that are generated by the source according to a Poisson PMF with parameter λ; however, the maximum number of packets that can be stored is a given integer b, and packets arriving to a full buffer are discarded. In the second slot, the system transmits either all the stored packets or c packets (whichever is less). Here, c is a given integer with $0 < c < b$.

(a) Assuming that at the beginning of the first slot the buffer is empty, find the PMF of the number of packets stored at the end of the first slot and at the end of the second slot.

(b) What is the probability that some packets get discarded during the first slot?

Problem 6. The Celtics and the Lakers are set to play a playoff series of n basketball games, where n is odd. The Celtics have a probability p of winning any one game, independent of other games.

(a) Find the values of p for which $n = 5$ is better for the Celtics than $n = 3$.

(b) Generalize part (a), i.e., for any $k > 0$, find the values for p for which $n = 2k + 1$ is better for the Celtics than $n = 2k - 1$.

Problem 7. You just rented a large house and the realtor gave you 5 keys, one for each of the 5 doors of the house. Unfortunately, all keys look identical, so to open the front door, you try them at random.

(a) Find the PMF of the number of trials you will need to open the door, under the following alternative assumptions: (1) after an unsuccessful trial, you mark the corresponding key, so that you never try it again, and (2) at each trial you are equally likely to choose any key.

(b) Repeat part (a) for the case where the realtor gave you an extra duplicate key for each of the 5 doors.

Problem 8. Recursive computation of the binomial PMF. Let X be a binomial random variable with parameters n and p. Show that its PMF can be computed by starting with $p_X(0) = (1-p)^n$, and then using the recursive formula

$$p_X(k+1) = \frac{p}{1-p} \cdot \frac{n-k}{k+1} \cdot p_X(k), \qquad k = 0, 1, \ldots, n-1.$$

Problem 9. Form of the binomial PMF. Consider a binomial random variable X with parameters n and p. Let k^* be the largest integer that is less than or equal to $(n+1)p$. Show that the PMF $p_X(k)$ is monotonically nondecreasing with k in the range from 0 to k^*, and is monotonically decreasing with k for $k \geq k^*$.

Problem 10. Form of the Poisson PMF. Let X be a Poisson random variable with parameter λ. Show that the PMF $p_X(k)$ increases monotonically with k up to the point where k reaches the largest integer not exceeding λ, and after that point decreases monotonically with k.

Problem 11.* The matchbox problem – inspired by Banach's smoking habits. A smoker mathematician carries one matchbox in his right pocket and one in his left pocket. Each time he wants to light a cigarette, he selects a matchbox from either pocket with probability $p = 1/2$, independent of earlier selections. The two matchboxes have initially n matches each. What is the PMF of the number of remaining matches at the moment when the mathematician reaches for a match and discovers that the corresponding matchbox is empty? How can we generalize to the case where the probabilities of a left and a right pocket selection are p and $1 - p$, respectively?

Solution. Let X be the number of matches that remain when a matchbox is found empty. For $k = 0, 1, \ldots, n$, let L_k (or R_k) be the event that an empty box is first discovered in the left (respectively. right) pocket while the number of matches in the right (respectively, left) pocket is k at that time. The PMF of X is

$$p_X(k) = \mathbf{P}(L_k) + \mathbf{P}(R_k), \qquad k = 0, 1, \ldots, n.$$

Viewing a left and a right pocket selection as a "success" and a "failure," respectively, $\mathbf{P}(L_k)$ is the probability that there are n successes in the first $2n - k$ trials, and trial $2n - k + 1$ is a success. or

$$\mathbf{P}(L_k) = \frac{1}{2}\binom{2n-k}{n}\left(\frac{1}{2}\right)^{2n-k}, \qquad k = 0, 1, \ldots, n.$$

By symmetry. $\mathbf{P}(L_k) = \mathbf{P}(R_k)$. so

$$p_X(k) = \mathbf{P}(L_k) + \mathbf{P}(R_k) = \binom{2n-k}{n}\left(\frac{1}{2}\right)^{2n-k}, \qquad k = 0, 1, \ldots, n.$$

In the more general case, where the probabilities of a left and a right pocket selection are p and $1 - p$, using a similar reasoning. we obtain

$$\mathbf{P}(L_k) = p\binom{2n-k}{n}p^n(1-p)^{n-k}, \qquad k = 0, 1, \ldots, n,$$

and

$$\mathbf{P}(R_k) = (1-p)\binom{2n-k}{n}p^{n-k}(1-p)^n, \qquad k = 0, 1, \ldots, n,$$

which yields

$$\begin{aligned} p_X(k) &= \mathbf{P}(L_k) + \mathbf{P}(R_k) \\ &= \binom{2n-k}{n}\left(p^{n+1}(1-p)^{n-k} + p^{n-k}(1-p)^{n+1}\right), \qquad k = 0, 1, \ldots, n. \end{aligned}$$

Problem 12.* Justification of the Poisson approximation property. Consider the PMF of a binomial random variable with parameters n and p. Show that asymptotically, as

$$n \to \infty, \qquad p \to 0,$$

while np is fixed at a given value λ, this PMF approaches the PMF of a Poisson random variable with parameter λ.

Solution. Using the equation $\lambda = np$, write the binomial PMF as

$$\begin{aligned} p_X(k) &= \frac{n!}{(n-k)!\,k!}p^k(1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k}\cdot\frac{\lambda^k}{k!}\cdot\left(1-\frac{\lambda}{n}\right)^{n-k}. \end{aligned}$$

Fix k and let $n \to \infty$. We have, for $j = 1, \ldots, k$,

$$\frac{n-k+j}{n} \to 1, \quad \left(1 - \frac{\lambda}{n}\right)^{-k} \to 1, \quad \left(1 - \frac{\lambda}{n}\right)^{n} \to e^{-\lambda}.$$

Thus, for each fixed k, as $n \to \infty$ we obtain

$$p_X(k) \to e^{-\lambda} \frac{\lambda^k}{k!}.$$

SECTION 2.3. Functions of Random Variables

Problem 13. A family has 5 natural children and has adopted 2 girls. Each natural child has equal probability of being a girl or a boy, independent of the other children. Find the PMF of the number of girls out of the 7 children.

Problem 14. Let X be a random variable that takes values from 0 to 9 with equal probability 1/10.

(a) Find the PMF of the random variable $Y = X \bmod(3)$.

(b) Find the PMF of the random variable $Y = 5 \bmod(X+1)$.

Problem 15. Let K be a random variable that takes, with equal probability $1/(2n+1)$, the integer values in the interval $[-n, n]$. Find the PMF of the random variable $Y = \ln X$, where $X = a^{|K|}$, and a is a positive number.

SECTION 2.4. Expectation, Mean, and Variance

Problem 16. Let X be a random variable with PMF

$$p_X(x) = \begin{cases} x^2/a, & \text{if } x = -3, -2, -1, 0, 1, 2, 3, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find a and $\mathbf{E}[X]$.

(b) What is the PMF of the random variable $Z = \big(X - \mathbf{E}[X]\big)^2$?

(c) Using the result from part (b), find the variance of X.

(d) Find the variance of X using the formula $\text{var}(X) = \sum_x \big(x - \mathbf{E}[X]\big)^2 p_X(x)$.

Problem 17. A city's temperature is modeled as a random variable with mean and standard deviation both equal to 10 degrees Celsius. A day is described as "normal" if the temperature during that day ranges within one standard deviation from the mean. What would be the temperature range for a normal day if temperature were expressed in degrees Fahrenheit?

Problem 18. Let a and b be positive integers with $a \leq b$, and let X be a random variable that takes as values, with equal probability, the powers of 2 in the interval $[2^a, 2^b]$. Find the expected value and the variance of X.

Problem 19. A prize is randomly placed in one of ten boxes, numbered from 1 to 10. You search for the prize by asking yes-no questions. Find the expected number of questions until you are sure about the location of the prize, under each of the following strategies.

(a) An enumeration strategy: you ask questions of the form "is it in box k?".

(b) A bisection strategy: you eliminate as close to half of the remaining boxes as possible by asking questions of the form "is it in a box numbered less than or equal to k?".

Solution. We will find the expected gain for each strategy, by computing the expected number of questions until we find the prize.

(a) With this strategy, the probability 1/10 of finding the location of the prize with i questions, where $i = 1, \ldots, 10$, is 1/10. Therefore, the expected number of questions is

$$\frac{1}{10}\sum_{i=1}^{10} i = \frac{1}{10} \cdot 55 = 5.5.$$

(b) It can be checked that for 4 of the 10 possible box numbers, exactly 4 questions will be needed, whereas for 6 of the 10 numbers, 3 questions will be needed. Therefore, with this strategy, the expected number of questions is

$$\frac{4}{10} \cdot 4 + \frac{6}{10} \cdot 3 = 3.4.$$

Problem 20. As an advertising campaign, a chocolate factory places golden tickets in some of its candy bars, with the promise that a golden ticket is worth a trip through the chocolate factory, and all the chocolate you can eat for life. If the probability of finding a golden ticket is p, find the mean and the variance of the number of candy bars you need to eat to find a ticket.

Problem 21. St. Petersburg paradox. You toss independently a fair coin and you count the number of tosses until the first tail appears. If this number is n, you receive 2^n dollars. What is the expected amount that you will receive? How much would you be willing to pay to play this game?

Problem 22. Two coins are simultaneously tossed until one of them comes up a head and the other a tail. The first coin comes up a head with probability p and the second with probability q. All tosses are assumed independent.

(a) Find the PMF, the expected value, and the variance of the number of tosses.

(b) What is the probability that the last toss of the first coin is a head?

Problem 23.

(a) A fair coin is tossed repeatedly and independently until two consecutive heads or two consecutive tails appear. Find the PMF, the expected value, and the variance of the number of tosses.

(b) Assume now that the coin is tossed until we obtain a tail that is immediately preceded by a head. Find the PMF and the expected value of the number of tosses.

SECTION 2.5. Joint PMFs of Multiple Random Variables

Problem 24. A stock market trader buys 100 shares of stock A and 200 shares of stock B. Let X and Y be the price changes of A and B. respectively. over a certain time period. and assume that the joint PMF of X and Y is uniform over the set of integers x and y satisfying

$$-2 \leq x \leq 4, \qquad -1 \leq y - x \leq 1.$$

(a) Find the marginal PMFs and the means of X and Y.

(b) Find the mean of the trader's profit.

Problem 25. A class of n students takes a test consisting of m questions. Suppose that student i submitted answers to the first m_i questions.

(a) The grader randomly picks one answer, call it (I, J), where I is the student ID number (taking values $1, \ldots, n$) and J is the question number (taking values $1, \ldots, m$). Assume that all answers are equally likely to be picked. Calculate the joint and the marginal PMFs of I and J.

(b) Assume that an answer to question j. if submitted by student i, is correct with probability p_{ij}. Each answer gets a points if it is correct and gets b points otherwise. Calculate the expected value of the score of student i.

Problem 26. PMF of the minimum of several random variables. On a given day. your golf score takes values from the range 101 to 110. with probability 0.1, independent of other days. Determined to improve your score, you decide to play on three different days and declare as your score the minimum X of the scores X_1, X_2, and X_3 on the different days.

(a) Calculate the PMF of X.

(b) By how much has your expected score improved as a result of playing on three days?

Problem 27.* The multinomial distribution. A die with r faces, numbered $1, \ldots, r$. is rolled a fixed number of times n. The probability that the ith face comes up on any one roll is denoted p_i, and the results of different rolls are assumed independent. Let X_i be the number of times that the ith face comes up.

(a) Find the joint PMF $p_{X_1, \ldots, X_r}(k_1, \ldots, k_r)$.

(b) Find the expected value and variance of X_i.

(c) Find $\mathbf{E}[X_i X_j]$ for $i \neq j$.

Solution. (a) The probability of a sequence of rolls where, for $i = 1, \ldots, r$, face i comes up k_i times is $p_1^{k_1} \cdots p_r^{k_r}$. Every such sequence determines a partition of the set of n rolls into r subsets with the ith subset having cardinality k_i (this is the set of rolls for which the ith face came up). The number of such partitions is the multinomial coefficient (cf. Section 1.6)

$$\binom{n}{k_1,\ldots,k_r} = \frac{n!}{k_1!\cdots k_r!}.$$

Thus, if $k_1 + \cdots + k_r = n$,

$$p_{X_1,\ldots,X_r}(k_1,\ldots,k_r) = \binom{n}{k_1,\ldots,k_r} p_1^{k_1}\cdots p_r^{k_r},$$

and otherwise, $p_{X_1,\ldots,X_r}(k_1,\ldots,k_r) = 0$.

(b) The random variable X_i is binomial with parameters n and p_i. Therefore, $\mathbf{E}[X_i] = np_i$, and $\text{var}(X_i) = np_i(1 - p_i)$.

(c) Suppose that $i \neq j$, and let $Y_{i,k}$ (or $Y_{j,k}$) be the Bernoulli random variable that takes the value 1 if face i (respectively, j) comes up on the kth roll, and the value 0 otherwise. Note that $Y_{i,k}Y_{j,k} = 0$, and that for $l \neq k$, $Y_{i,k}$ and $Y_{j,l}$ are independent, so that $\mathbf{E}[Y_{i,k}Y_{j,l}] = p_ip_j$. Therefore,

$$\begin{aligned}
\mathbf{E}[X_iX_j] &= \mathbf{E}\big[(Y_{i,1} + \cdots + Y_{i,n})(Y_{j,1} + \cdots + Y_{j,n})\big] \\
&= n(n-1)\mathbf{E}[Y_{i,1}Y_{j,2}] \\
&= n(n-1)p_ip_j.
\end{aligned}$$

Problem 28.* The quiz problem. Consider a quiz contest where a person is given a list of n questions and can answer these questions in any order he or she chooses. Question i will be answered correctly with probability p_i, and the person will then receive a reward v_i. At the first incorrect answer, the quiz terminates and the person is allowed to keep his or her previous rewards. The problem is to choose the ordering of questions so as to maximize the expected value of the total reward obtained. Show that it is optimal to answer questions in a nonincreasing order of $p_iv_i/(1 - p_i)$.

Solution. We will use a so-called interchange argument, which is often useful in sequencing problems. Let i and j be the kth and $(k+1)$st questions in an optimally ordered list

$$L = (i_1,\ldots,i_{k-1},i,j,i_{k+2},\ldots,i_n).$$

Consider the list

$$L' = (i_1,\ldots,i_{k-1},j,i,i_{k+2},\ldots,i_n)$$

obtained from L by interchanging the order of questions i and j. We compute the expected values of the rewards of L and L', and note that since L is optimally ordered, we have

$$\mathbf{E}[\text{reward of } L] \geq \mathbf{E}[\text{reward of } L'].$$

Define the *weight* of question i to be

$$w(i) = \frac{p_iv_i}{1 - p_i}.$$

We will show that any permutation of the questions in a nonincreasing order of weights maximizes the expected reward.

If $L = (i_1, \ldots, i_n)$ is a permutation of the questions, define $L^{(k)}$ to be the permutation obtained from L by interchanging questions i_k and i_{k+1}. Let us first compute the difference between the expected reward of L and that of $L^{(k)}$. We have

$$\mathbf{E}[\text{reward of } L] = p_{i_1} v_{i_1} + p_{i_1} p_{i_2} v_{i_2} + \cdots + p_{i_1} \cdots p_{i_n} v_{i_n},$$

and

$$\begin{aligned}\mathbf{E}\big[\text{reward of } L^{(k)}\big] = {} & p_{i_1} v_{i_1} + p_{i_1} p_{i_2} v_{i_2} + \cdots + p_{i_1} \cdots p_{i_{k-1}} v_{i_{k-1}} \\ & + p_{i_1} \cdots p_{i_{k-1}} p_{i_{k+1}} v_{i_{k+1}} + p_{i_1} \cdots p_{i_{k-1}} p_{i_{k+1}} p_{i_k} v_{i_k} \\ & + p_{i_1} \cdots p_{i_{k+2}} v_{i_{k+2}} + \cdots + p_{i_1} \cdots p_{i_n} v_{i_n}.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbf{E}\big[\text{reward of } L^{(k)}\big] - \mathbf{E}[\text{reward of } L] &= p_{i_1} \cdots p_{i_{k-1}} \big(p_{i_{k+1}} v_{i_{k+1}} + p_{i_{k+1}} p_{i_k} v_{i_k} \\ &\qquad - p_{i_k} v_{i_k} - p_{i_k} p_{i_{k+1}} v_{i_{k+1}}\big) \\ &= p_{i_1} \cdots p_{i_{k-1}} (1 - p_{i_k})(1 - p_{i_{k+1}}) \big(w(i_{k+1}) - w(i_k)\big).\end{aligned}$$

Now, let us go back to our problem. Consider any permutation L of the questions. If $w(i_k) < w(i_{k+1})$ for some k, it follows from the above equation that the permutation $L^{(k)}$ has an expected reward larger than that of L. So, an optimal permutation of the questions must be in a nonincreasing order of weights.

Let us finally show that any two such permutations have equal expected rewards. Assume that L is such a permutation and say that $w(i_k) = w(i_{k+1})$ for some k. We know that interchanging i_k and i_{k+1} preserves the expected reward. So, the expected reward of any permutation L' in a non-increasing order of weights is equal to that of L, because L' can be obtained from L by repeatedly interchanging adjacent questions having equal weights.

Problem 29.* The inclusion-exclusion formula. Let $A_1, A_2, \ldots, A_n$ be events. Let $S_1 = \{i \,|\, 1 \le i \le n\}$, $S_2 = \{(i_1, i_2) \,|\, 1 \le i_1 < i_2 \le n\}$, and more generally, let S_m be the set of all m-tuples $(i_1, \ldots, i_m)$ of indices that satisfy $1 \le i_1 < i_2 < \cdots < i_m \le n$. Show that

$$\begin{aligned}\mathbf{P}\left(\cup_{k=1}^n A_k\right) = {} & \sum_{i \in S_1} \mathbf{P}(A_i) - \sum_{(i_1, i_2) \in S_2} \mathbf{P}(A_{i_1} \cap A_{i_2}) \\ & + \sum_{(i_1, i_2, i_3) \in S_3} \mathbf{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \cdots + (-1)^{n-1} \mathbf{P}\left(\cap_{k=1}^n A_k\right).\end{aligned}$$

Hint: Let X_i be a binary random variable which is equal to 1 when A_i occurs, and equal to 0 otherwise. Relate the event of interest to the random variable $(1 - X_1)(1 - X_2) \cdots (1 - X_n)$.

Solution. Let us express the event $B = \cup_{k=1}^n A_k$ in terms of the random variables $X_1, \ldots, X_n$. The event B^c occurs when all of the random variables $X_1, \ldots, X_n$ are zero, which happens when the random variable $Y = (1-X_1)(1-X_2) \cdots (1-X_n)$ is equal to 1.

Note that Y can only take values in the set $\{0, 1\}$, so that $\mathbf{P}(B^c) = \mathbf{P}(Y = 1) = \mathbf{E}[Y]$. Therefore,

$$\begin{aligned}\mathbf{P}(B) &= 1 - \mathbf{E}\big[(1 - X_1)(1 - X_2)\cdots(1 - X_n)\big] \\ &= \mathbf{E}[X_1 + \cdots + X_n] - \mathbf{E}\left[\sum_{(i_1,i_2)\in S_2} X_{i_1}X_{i_2}\right] + \cdots + (-1)^{n-1}\mathbf{E}[X_1\cdots X_n].\end{aligned}$$

We note that

$$\mathbf{E}[X_i] = \mathbf{P}(A_i), \qquad \mathbf{E}[X_{i_1}X_{i_2}] = \mathbf{P}(A_{i_1} \cap A_{i_2}),$$
$$\mathbf{E}[X_{i_1}X_{i_2}X_{i_3}] = \mathbf{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}), \qquad \mathbf{E}\big[X_1X_2\cdots X_n\big] = \mathbf{P}(\cap_{k=1}^n A_k),$$

etc., from which the desired formula follows.

Problem 30.* Alvin's database of friends contains n entries, but due to a software glitch, the addresses correspond to the names in a totally random fashion. Alvin writes a holiday card to each of his friends and sends it to the (software-corrupted) address. What is the probability that at least one of his friends will get the correct card? *Hint:* Use the inclusion-exclusion formula.

Solution. Let A_k be the event that the kth card is sent to the correct address. We have for any k, j, i,

$$\mathbf{P}(A_k) = \frac{1}{n} = \frac{(n-1)!}{n!},$$
$$\mathbf{P}(A_k \cap A_j) = \mathbf{P}(A_k)\mathbf{P}(A_j \mid A_k) = \frac{1}{n}\cdot\frac{1}{n-1} = \frac{(n-2)!}{n!},$$
$$\mathbf{P}(A_k \cap A_j \cap A_i) = \frac{1}{n}\cdot\frac{1}{n-1}\cdot\frac{1}{n-2} = \frac{(n-3)!}{n!},$$

etc., and

$$\mathbf{P}(\cap_{k=1}^n A_k) = \frac{1}{n!}.$$

Applying the inclusion-exclusion formula,

$$\begin{aligned}\mathbf{P}\left(\cup_{k=1}^n A_k\right) = &\sum_{i\in S_1}\mathbf{P}(A_i) - \sum_{(i_1,i_2)\in S_2}\mathbf{P}(A_{i_1}\cap A_{i_2}) \\ &+ \sum_{(i_1,i_2,i_3)\in S_3}\mathbf{P}(A_{i_1}\cap A_{i_2}\cap A_{i_3}) - \cdots + (-1)^{n-1}\mathbf{P}\left(\cap_{k=1}^n A_k\right),\end{aligned}$$

we obtain the desired probability

$$\begin{aligned}\mathbf{P}(\cup_{k=1}^n A_k) &= \binom{n}{1}\frac{(n-1)!}{n!} - \binom{n}{2}\frac{(n-2)!}{n!} + \binom{n}{3}\frac{(n-3)!}{n!} - \cdots + (-1)^{n-1}\frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!}.\end{aligned}$$

When n is large, this probability can be approximated by $1 - e^{-1}$.

SECTION 2.6. Conditioning

Problem 31. Consider four independent rolls of a 6-sided die. Let X be the number of 1s and let Y be the number of 2s obtained. What is the joint PMF of X and Y?

Problem 32. D. Bernoulli's problem of joint lives. Consider $2m$ persons forming m couples who live together at a given time. Suppose that at some later time, the probability of each person being alive is p, independent of other persons. At that later time, let A be the number of persons that are alive and let S be the number of couples in which both partners are alive. For any survivor number a, find $\mathbf{E}[S \mid A = a]$.

Problem 33.* A coin that has probability of heads equal to p is tossed successively and independently until a head comes twice in a row or a tail comes twice in a row. Find the expected value of the number of tosses.

Solution. One possibility here is to calculate the PMF of X, the number of tosses until the game is over, and use it to compute $\mathbf{E}[X]$. However, with an unfair coin, this turns out to be cumbersome, so we argue by using the total expectation theorem and a suitable partition of the sample space. Let H_k (or T_k) be the event that a head (or a tail, respectively) comes at the kth toss, and let p (respectively, q) be the probability of H_k (respectively, T_k). Since H_1 and T_1 form a partition of the sample space, and $\mathbf{P}(H_1) = p$ and $\mathbf{P}(T_1) = q$, we have

$$\mathbf{E}[X] = p\mathbf{E}[X \mid H_1] + q\mathbf{E}[X \mid T_1].$$

Using again the total expectation theorem, we have

$$\mathbf{E}[X \mid H_1] = p\mathbf{E}[X \mid H_1 \cap H_2] + q\mathbf{E}[X \mid H_1 \cap T_2] = 2p + q\big(1 + \mathbf{E}[X \mid T_1]\big),$$

where we have used the fact

$$\mathbf{E}[X \mid H_1 \cap H_2] = 2$$

(since the game ends after two successive heads), and

$$\mathbf{E}[X \mid H_1 \cap T_2] = 1 + \mathbf{E}[X \mid T_1]$$

(since if the game is not over, only the last toss matters in determining the number of additional tosses up to termination). Similarly, we obtain

$$\mathbf{E}[X \mid T_1] = 2q + p\big(1 + \mathbf{E}[X \mid H_1]\big).$$

Combining the above two relations, collecting terms, and using the fact $p + q = 1$, we obtain after some calculation

$$\mathbf{E}[X \mid T_1] = \frac{2 + p^2}{1 - pq},$$

and similarly

$$\mathbf{E}[X \mid H_1] = \frac{2 + q^2}{1 - pq}.$$

Thus,

$$\mathbf{E}[X] = p \cdot \frac{2 + q^2}{1 - pq} + q \cdot \frac{2 + p^2}{1 - pq},$$

and finally, using the fact $p + q = 1$,

$$\mathbf{E}[X] = \frac{2 + pq}{1 - pq}.$$

In the case of a fair coin ($p = q = 1/2$), we obtain $\mathbf{E}[X] = 3$. It can also be verified that $2 \leq \mathbf{E}[X] \leq 3$ for all values of p.

Problem 34.* A spider and a fly move along a straight line. At each second, the fly moves a unit step to the right or to the left with equal probability p, and stays where it is with probability $1 - 2p$. The spider always takes a unit step in the direction of the fly. The spider and the fly start D units apart, where D is a random variable taking positive integer values with a given PMF. If the spider lands on top of the fly, it's the end. What is the expected value of the time it takes for this to happen?

Solution. Let T be the time at which the spider lands on top of the fly. We define

A_d: the event that initially the spider and the fly are d units apart.

B_d: the event that after one second the spider and the fly are d units apart.

Our approach will be to first apply the (conditional version of the) total expectation theorem to compute $\mathbf{E}[T \mid A_1]$, then use the result to compute $\mathbf{E}[T \mid A_2]$, and similarly compute sequentially $\mathbf{E}[T \mid A_d]$ for all relevant values of d. We will then apply the (unconditional version of the) total expectation theorem to compute $\mathbf{E}[T]$.

We have

$$A_d = (A_d \cap B_d) \cup (A_d \cap B_{d-1}) \cup (A_d \cap B_{d-2}), \qquad \text{if } d > 1.$$

This is because if the spider and the fly are at a distance $d > 1$ apart, then one second later their distance will be d (if the fly moves away from the spider) or $d - 1$ (if the fly does not move) or $d - 2$ (if the fly moves towards the spider). We also have, for the case where the spider and the fly start one unit apart,

$$A_1 = (A_1 \cap B_1) \cup (A_1 \cap B_0).$$

Using the total expectation theorem, we obtain

$$\begin{aligned}\mathbf{E}[T \mid A_d] = \mathbf{P}(B_d \mid A_d)&\mathbf{E}[T \mid A_d \cap B_d] \\ &+ \mathbf{P}(B_{d-1} \mid A_d)\mathbf{E}[T \mid A_d \cap B_{d-1}] \\ &+ \mathbf{P}(B_{d-2} \mid A_d)\mathbf{E}[T \mid A_d \cap B_{d-2}], \qquad \text{if } d > 1,\end{aligned}$$

and

$$\mathbf{E}[T \mid A_1] = \mathbf{P}(B_1 \mid A_1)\mathbf{E}[T \mid A_1 \cap B_1] + \mathbf{P}(B_0 \mid A_1)\mathbf{E}[T \mid A_1 \cap B_0], \quad \text{if } d = 1.$$

It can be seen based on the problem data that

$$\mathbf{P}(B_1 \mid A_1) = 2p, \qquad \mathbf{P}(B_0 \mid A_1) = 1 - 2p,$$

$$\mathbf{E}[T \mid A_1 \cap B_1] = 1 + \mathbf{E}[T \mid A_1], \qquad \mathbf{E}[T \mid A_1 \cap B_0] = 1,$$

so by applying the formula for the case $d = 1$, we obtain

$$\mathbf{E}[T \mid A_1] = 2p\big(1 + \mathbf{E}[T \mid A_1]\big) + (1 - 2p),$$

or

$$\mathbf{E}[T \mid A_1] = \frac{1}{1-2p}.$$

By applying the formula with $d = 2$, we obtain

$$\mathbf{E}[T \mid A_2] = p\mathbf{E}[T \mid A_2 \cap B_2] + (1-2p)\mathbf{E}[T \mid A_2 \cap B_1] + p\mathbf{E}[T \mid A_2 \cap B_0].$$

We have

$$\mathbf{E}[T \mid A_2 \cap B_0] = 1,$$
$$\mathbf{E}[T \mid A_2 \cap B_1] = 1 + \mathbf{E}[T \mid A_1],$$
$$\mathbf{E}[T \mid A_2 \cap B_2] = 1 + \mathbf{E}[T \mid A_2],$$

so by substituting these relations in the expression for $\mathbf{E}[T \mid A_2]$, we obtain

$$\begin{aligned}\mathbf{E}[T \mid A_2] &= p\big(1 + \mathbf{E}[T \mid A_2]\big) + (1-2p)\big(1 + \mathbf{E}[T \mid A_1]\big) + p \\ &= p\big(1 + \mathbf{E}[T \mid A_2]\big) + (1-2p)\left(1 + \frac{1}{1-2p}\right) + p.\end{aligned}$$

This equation yields after some calculation

$$\mathbf{E}[T \mid A_2] = \frac{2}{1-p}.$$

Generalizing, we obtain for $d > 2$,

$$\mathbf{E}[T \mid A_d] = p\big(1 + \mathbf{E}[T \mid A_d]\big) + (1-2p)\big(1 + \mathbf{E}[T \mid A_{d-1}]\big) + p\big(1 + \mathbf{E}[T \mid A_{d-2}]\big).$$

Thus, $\mathbf{E}[T \mid A_d]$ can be generated recursively for any initial distance d, using as initial conditions the values of $\mathbf{E}[T \mid A_1]$ and $\mathbf{E}[T \mid A_2]$ obtained earlier.

Finally, the expected value of T can be obtained using the given PMF for the initial distance D and the total expectation theorem:

$$\mathbf{E}[T] = \sum_d p_D(d)\mathbf{E}[T \mid A_d].$$

Problem 35.* Verify the expected value rule

$$\mathbf{E}\big[g(X,Y)\big] = \sum_x \sum_y g(x,y)p_{X,Y}(x,y),$$

using the expected value rule for a function of a single random variable. Then, use the rule for the special case of a linear function, to verify the formula

$$\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y],$$

where a and b are given scalars.

Solution. We use the total expectation theorem to reduce the problem to the case of a single random variable. In particular, we have

$$\begin{aligned}\mathbf{E}\big[g(X,Y)\big] &= \sum_y p_Y(y)\mathbf{E}\big[g(X,Y)\,|\,Y=y\big] \\ &= \sum_y p_Y(y)\mathbf{E}\big[g(X,y)\,|\,Y=y\big] \\ &= \sum_y p_Y(y)\sum_x g(x,y)p_{X|Y}(x\,|\,y) \\ &= \sum_x\sum_y g(x,y)p_{X,Y}(x,y),\end{aligned}$$

as desired. Note that the third equality above used the expected value rule for the function $g(X,y)$ of a single random variable X.

For the linear special case, the expected value rule gives

$$\begin{aligned}\mathbf{E}[aX+bY] &= \sum_x\sum_y (ax+by)p_{X,Y}(x,y) \\ &= a\sum_x x\sum_y p_{X,Y}(x,y) + b\sum_y y\sum_x p_{X,Y}(x,y) \\ &= a\sum_x xp_X(x) + b\sum_y yp_Y(y) \\ &= a\mathbf{E}[X] + b\mathbf{E}[Y].\end{aligned}$$

Problem 36.* The multiplication rule for conditional PMFs. Let X, Y, and Z be random variables.

(a) Show that

$$p_{X,Y,Z}(x,y,z) = p_X(x)p_{Y|X}(y\,|\,x)p_{Z|X,Y}(z\,|\,x,y).$$

(b) How can we interpret this formula as a special case of the multiplication rule given in Section 1.3?

(c) Generalize to the case of more than three random variables.

Solution. (a) We have

$$\begin{aligned}p_{X,Y,Z}(x,y,z) &= \mathbf{P}(X=x, Y=y, Z=z) \\ &= \mathbf{P}(X=x)\mathbf{P}(Y=y, Z=z\,|\,X=x) \\ &= \mathbf{P}(X=x)\mathbf{P}(Y=y\,|\,X=x)\mathbf{P}(Z=z\,|\,X=x, Y=y) \\ &= p_X(x)p_{Y|X}(y\,|\,x)p_{Z|X,Y}(z\,|\,x,y).\end{aligned}$$

(b) The formula can be written as

$$\mathbf{P}(X=x, Y=y, Z=z) = \mathbf{P}(X=x)\mathbf{P}(Y=y\,|\,X=x)\mathbf{P}(Z=z\,|\,X=x, Y=y),$$

which is a special case of the multiplication rule.

(c) The generalization is

$$p_{X_1,\dots,X_n}(x_1,\dots,x_n)$$
$$= p_{X_1}(x_1)p_{X_2|X_1}(x_2\,|\,x_1)\cdots p_{X_n|X_1,\dots,X_{n-1}}(x_n\,|\,x_1,\dots,x_{n-1}).$$

Problem 37.* Splitting a Poisson random variable. A transmitter sends out either a 1 with probability p, or a 0 with probability $1-p$, independent of earlier transmissions. If the number of transmissions within a given time interval has a Poisson PMF with parameter λ, show that the number of 1s transmitted in that same time interval has a Poisson PMF with parameter $p\lambda$.

Solution. Let X and Y be the numbers of 1s and 0s transmitted, respectively. Let $Z = X + Y$ be the total number of symbols transmitted. We have

$$\begin{aligned}\mathbf{P}(X=n, Y=m) &= \mathbf{P}(X=n, Y=m\,|\,Z=n+m)\mathbf{P}(Z=n+m)\\ &= \binom{n+m}{n}p^n(1-p)^m\cdot\frac{e^{-\lambda}\lambda^{n+m}}{(n+m)!}\\ &= \frac{e^{-\lambda p}(\lambda p)^n}{n!}\cdot\frac{e^{-\lambda(1-p)}\big(\lambda(1-p)\big)^m}{m!}.\end{aligned}$$

Thus,

$$\begin{aligned}\mathbf{P}(X=n) &= \sum_{m=0}^{\infty}\mathbf{P}(X=n, Y=m)\\ &= \frac{e^{-\lambda p}(\lambda p)^n}{n!}e^{-\lambda(1-p)}\sum_{m=0}^{\infty}\frac{\big(\lambda(1-p)\big)^m}{m!}\\ &= \frac{e^{-\lambda p}(\lambda p)^n}{n!}e^{-\lambda(1-p)}e^{\lambda(1-p)}\\ &= \frac{e^{-\lambda p}(\lambda p)^n}{n!},\end{aligned}$$

so that X is Poisson with parameter λp.

SECTION 2.7. Independence

Problem 38. Alice passes through four traffic lights on her way to work, and each light is equally likely to be green or red, independent of the others.

(a) What is the PMF, the mean, and the variance of the number of red lights that Alice encounters?

(b) Suppose that each red light delays Alice by exactly two minutes. What is the variance of Alice's commuting time?

Problem 39. Each morning, Hungry Harry eats some eggs. On any given morning, the number of eggs he eats is equally likely to be 1, 2, 3, 4, 5, or 6, independent of what he has done in the past. Let X be the number of eggs that Harry eats in 10 days. Find the mean and variance of X.

Problem 40. A particular professor is known for his arbitrary grading policies. Each paper receives a grade from the set $\{A, A-, B+, B, B-, C+\}$, with equal probability, independent of other papers. How many papers do you expect to hand in before you receive each possible grade at least once?

Problem 41. You drive to work 5 days a week for a full year (50 weeks), and with probability $p = 0.02$ you get a traffic ticket on any given day, independent of other days. Let X be the total number of tickets you get in the year.

(a) What is the probability that the number of tickets you get is exactly equal to the expected value of X?

(b) Calculate approximately the probability in (a) using a Poisson approximation.

(c) Any one of the tickets is \$10 or \$20 or \$50 with respective probabilities 0.5, 0.3, and 0.2, and independent of other tickets. Find the mean and the variance of the amount of money you pay in traffic tickets during the year.

(d) Suppose you don't know the probability p of getting a ticket, but you got 5 tickets during the year, and you estimate p by the sample mean

$$\hat{p} = \frac{5}{250} = 0.02.$$

What is the range of possible values of p assuming that the difference between p and the sample mean $\hat{p}$ is within 5 times the standard deviation of the sample mean?

Problem 42. **Computational problem.** Here is a probabilistic method for computing the area of a given subset S of the unit square. The method uses a sequence of independent random selections of points in the unit square $[0, 1] \times [0, 1]$, according to a uniform probability law. If the ith point belongs to the subset S the value of a random variable X_i is set to 1, and otherwise it is set to 0. Let $X_1, X_2, \ldots$ be the sequence of random variables thus defined, and for any n, let

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

(a) Show that $\mathbf{E}[S_n]$ is equal to the area of the subset S, and that $\text{var}(S_n)$ diminishes to 0 as n increases.

(b) Show that to calculate S_n, it is sufficient to know S_{n-1} and X_n, so the past values of X_k, $k = 1, \ldots, n-1$, do not need to be remembered. Give a formula.

(c) Write a computer program to generate S_n for $n = 1, 2, \ldots, 10000$, using the computer's random number generator, for the case where the subset S is the circle inscribed within the unit square. How can you use your program to measure experimentally the value of π?

(d) Use a similar computer program to calculate approximately the area of the set of all (x, y) that lie within the unit square and satisfy $0 \leq \cos \pi x + \sin \pi y \leq 1$.

Problem 43.* Suppose that X and Y are independent, identically distributed, geometric random variables with parameter p. Show that

$$\mathbf{P}(X = i \,|\, X + Y = n) = \frac{1}{n-1}, \qquad i = 1, \ldots, n-1.$$

Solution. Consider repeatedly and independently tossing a coin with probability of heads p. We can interpret $\mathbf{P}(X = i \,|\, X + Y = n)$ as the probability that we obtained a head for the first time on the ith toss given that we obtained a head for the second time on the nth toss. We can then argue, intuitively, that given that the second head occurred on the nth toss, the first head is equally likely to have come up at any toss between 1 and $n-1$. To establish this precisely, note that we have

$$\mathbf{P}(X = i \,|\, X + Y = n) = \frac{\mathbf{P}(X = i,\ X + Y = n)}{\mathbf{P}(X + Y = n)} = \frac{\mathbf{P}(X = i)\mathbf{P}(Y = n - i)}{\mathbf{P}(X + Y = n)}.$$

Also

$$\mathbf{P}(X = i) = p(1-p)^{i-1}, \qquad \text{for } i \geq 1,$$

and

$$\mathbf{P}(Y = n - i) = p(1-p)^{n-i-1}, \qquad \text{for } n - i \geq 1.$$

It follows that

$$\mathbf{P}(X = i)\mathbf{P}(Y = n - i) = \begin{cases} p^2(1-p)^{n-2}, & \text{if } i = 1, \ldots, n-1, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, for any i and j in the range $[1, n-1]$, we have

$$\mathbf{P}(X = i \,|\, X + Y = n) = \mathbf{P}(X = j \,|\, X + Y = n).$$

Hence

$$\mathbf{P}(X = i \,|\, X + Y = n) = \frac{1}{n-1}, \qquad i = 1, \ldots, n-1.$$

Problem 44.* Let X and Y be two random variables with given joint PMF, and let g and h be two functions of X and Y, respectively. Show that if X and Y are independent, then the same is true for the random variables $g(X)$ and $h(Y)$.

Solution. Let $U = g(X)$ and $V = h(Y)$. Then, we have

$$\begin{aligned} p_{U,V}(u, v) &= \sum_{\{(x,y) \,|\, g(x)=u,\, h(y)=v\}} p_{X,Y}(x, y) \\ &= \sum_{\{(x,y) \,|\, g(x)=u,\, h(y)=v\}} p_X(x) p_Y(y) \\ &= \sum_{\{x \,|\, g(x)=u\}} p_X(x) \sum_{\{x \,|\, h(y)=v\}} p_Y(y) \\ &= p_U(u) p_V(v), \end{aligned}$$

so U and V are independent.

Problem 45.* Variability extremes. Let $X_1, \ldots, X_n$ be independent random variables and let $X = X_1 + \cdots + X_n$ be their sum.

(a) Suppose that each X_i is Bernoulli with parameter p_i, and that $p_1, \ldots, p_n$ are chosen so that the mean of X is a given $\mu > 0$. Show that the variance of X is maximized if the p_i are chosen to be all equal to μ/n.

(b) Suppose that each X_i is geometric with parameter p_i, and that $p_1, \ldots, p_n$ are chosen so that the mean of X is a given $\mu > 0$. Show that the variance of X is minimized if the p_i are chosen to be all equal to n/μ. [Note the strikingly different character of the results of parts (a) and (b).]

Solution. (a) We have

$$\text{var}(X) = \sum_{i=1}^{n} \text{var}(X_i) = \sum_{i=1}^{n} p_i(1 - p_i) = \mu - \sum_{i=1}^{n} p_i^2.$$

Thus maximizing the variance is equivalent to minimizing $\sum_{i=1}^{n} p_i^2$. It can be seen (using the constraint $\sum_{i=1}^{n} p_i = \mu$) that

$$\sum_{i=1}^{n} p_i^2 = \sum_{i=1}^{n} (\mu/n)^2 + \sum_{i=1}^{n} (p_i - \mu/n)^2,$$

so $\sum_{i=1}^{n} p_i^2$ is minimized when $p_i = \mu/n$ for all i.

(b) We have

$$\mu = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} \frac{1}{p_i},$$

and

$$\text{var}(X) = \sum_{i=1}^{n} \text{var}(X_i) = \sum_{i=1}^{n} \frac{1 - p_i}{p_i^2}.$$

Introducing the change of variables $y_i = 1/p_i = \mathbf{E}[X_i]$, we see that the constraint becomes

$$\sum_{i=1}^{n} y_i = \mu,$$

and that we must minimize

$$\sum_{i=1}^{n} y_i(y_i - 1) = \sum_{i=1}^{n} y_i^2 - \mu,$$

subject to that constraint. This is the same problem as the one of part (a), so the method of proof given there applies.

Problem 46.* Entropy and uncertainty. Consider a random variable X that can take n values, $x_1, \ldots, x_n$, with corresponding probabilities $p_1, \ldots, p_n$. The **entropy** of X is defined to be

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i.$$

(All logarithms in this problem are with respect to base two.) The entropy $H(X)$ provides a measure of the uncertainty about the value of X. To get a sense of this, note that $H(X) \geq 0$ and that $H(X)$ is very close to 0 when X is "nearly deterministic," i.e., takes one of its possible values with probability very close to 1 (since we have $p \log p \approx 0$ if either $p \approx 0$ or $p \approx 1$).

The notion of entropy is fundamental in information theory, which originated with C. Shannon's famous work and is described in many specialized textbooks. For example, it can be shown that $H(X)$ is a lower bound to the average number of yes-no questions (such as "is $X = x_1$?" or "is $X < x_5$?") that must be asked in order to determine the value of X. Furthermore, if k is the average number of questions required to determine the value of a string of independent identically distributed random variables $X_1, X_2, \ldots, X_n$, then, with a suitable strategy, k/n can be made as close to $H(X)$ as desired, when n is large.

(a) Show that if $q_1, \ldots, q_n$ are nonnegative numbers such that $\sum_{i=1}^{n} q_i = 1$, then

$$H(X) \leq -\sum_{i=1}^{n} p_i \log q_i,$$

with equality if and only if $p_i = q_i$ for all i. As a special case, show that $H(X) \leq \log n$, with equality if and only if $p_i = 1/n$ for all i. *Hint*: Use the inequality $\ln \alpha \leq \alpha - 1$, for $\alpha > 0$, which holds with equality if and only if $\alpha = 1$; here $\ln \alpha$ stands for the natural logarithm.

(b) Let X and Y be random variables taking a finite number of values, and having joint PMF $p_{X,Y}(x, y)$. Define

$$I(X,Y) = \sum_{x} \sum_{y} p_{X,Y}(x,y) \log \left(\frac{p_{X,Y}(x,y)}{p_X(x) p_Y(y)} \right).$$

Show that $I(X,Y) \geq 0$, and that $I(X,Y) = 0$ if and only if X and Y are independent.

(c) Show that

$$I(X,Y) = H(X) + H(Y) - H(X,Y),$$

where

$$H(X,Y) = -\sum_{x} \sum_{y} p_{X,Y}(x,y) \log p_{X,Y}(x,y),$$

$$H(X) = -\sum_{x} p_X(x) \log p_X(x), \qquad H(Y) = -\sum_{y} p_Y(y) \log p_Y(y).$$

(d) Show that

$$I(X,Y) = H(X) - H(X \mid Y),$$

where

$$H(X \mid Y) = -\sum_{y} p_Y(y) \sum_{x} p_{X|Y}(x \mid y) \log p_{X|Y}(x \mid y).$$

[Note that $H(X \mid Y)$ may be viewed as the conditional entropy of X given Y, that is, the entropy of the conditional distribution of X, given that $Y = y$, averaged over all possible values y. Thus, the quantity $I(X,Y) = H(X) - H(X \mid Y)$ is the reduction in the entropy (uncertainty) on X, when Y becomes known. It can be therefore interpreted as the information about X that is conveyed by Y, and is called the **mutual information** of X and Y.]

Solution. (a) We will use the inequality $\ln \alpha \le \alpha - 1$. (To see why this inequality is true, write $\ln \alpha = \int_1^\alpha \beta^{-1}\, d\beta < \int_1^\alpha d\beta = \alpha - 1$ for $\alpha > 1$, and write $\ln \alpha = -\int_\alpha^1 \beta^{-1}\, d\beta < -\int_\alpha^1 d\beta = \alpha - 1$ for $0 < \alpha < 1$.)

We have

$$-\sum_{i=1}^{n} p_i \ln p_i + \sum_{i=1}^{n} p_i \ln q_i = \sum_{i=1}^{n} p_i \ln\left(\frac{q_i}{p_i}\right) \le \sum_{i=1}^{n} p_i \left(\frac{q_i}{p_i} - 1\right) = 0,$$

with equality if and only if $p_i = q_i$ for all i. Since $\ln p = \log p \ln 2$, we obtain the desired relation $H(X) \le -\sum_{i=1}^n p_i \log q_i$. The inequality $H(X) \le \log n$ is obtained by setting $q_i = 1/n$ for all i.

(b) The numbers $p_X(x)p_Y(y)$ satisfy $\sum_x \sum_y p_X(x)p_Y(y) = 1$, so by part (a), we have

$$\sum_x \sum_y p_{X,Y}(x,y) \log\big(p_{X,Y}(x,y)\big) \ge \sum_x \sum_y p_{X,Y}(x,y) \log\big(p_X(x)p_Y(y)\big),$$

with equality if and only if

$$p_{X,Y}(x,y) = p_X(x)p_Y(y), \qquad \text{for all } x \text{ and } y,$$

which is equivalent to X and Y being independent.

(c) We have

$$I(X,Y) = \sum_x \sum_y p_{X,Y}(x,y) \log p_{X,Y}(x,y) - \sum_x \sum_y p_{X,Y}(x,y) \log\big(p_X(x)p_Y(y)\big),$$

and

$$\sum_x \sum_y p_{X,Y}(x,y) \log p_{X,Y}(x,y) = -H(X,Y),$$

$$\begin{aligned}
-\sum_x \sum_y p_{X,Y}(x,y) \log\big(p_X(x)p_Y(y)\big) &= -\sum_x \sum_y p_{X,Y}(x,y) \log p_X(x) \\
&\quad - \sum_x \sum_y p_{X,Y}(x,y) \log p_Y(y) \\
&= -\sum_x p_X(x) \log p_X(x) - \sum_y p_Y(y) \log p_Y(y) \\
&= H(X) + H(Y).
\end{aligned}$$

Combining the above three relations, we obtain $I(X,Y) = H(X) + H(Y) - H(X,Y)$.

(d) From the calculation in part (c), we have

$$\begin{aligned} I(X,Y) &= \sum_x \sum_y p_{X,Y}(x,y) \log p_{X,Y}(x,y) - \sum_x p_X(x) \log p_X(x) \\ &\quad - \sum_x \sum_y p_{X,Y}(x,y) \log p_Y(y) \\ &= H(X) + \sum_x \sum_y p_{X,Y}(x,y) \log\left(\frac{p_{X,Y}(x,y)}{p_Y(y)}\right) \\ &= H(X) + \sum_x \sum_y p_Y(y) p_{X|Y}(x \mid y) \log p_{X|Y}(x \mid y) \\ &= H(X) - H(X \mid Y). \end{aligned}$$