Problem Set 11

Due: 2:35pm in lecture, Tuesday, December 5, 2000.

Be sure to indicate your TA's name clearly on the solutions your turn in.

Self-study Read: Rosen Sections 4.4 Self-study problems (not to be handed in): 13, x, x, x.

Reading: Lecture notes; Rosen 4.4, 4.5

Policy on collaboration If you collaborated with someone on the problem sets, you should mention their names on top of the problem set. If you did not collaborate with anyone, you should explicitly say so. As a reminder, the write-up of your solutions should be entirely your own.

Problem 1 (15 points) Let random variables X and Y be the outcomes of two dice rolls. Assume that the dice are fair, 6-sided, and independent. Each problem part below lists two random variables derived from X and Y. State whether or not these two random variables are independent and briefly sketch a justification for your answer.

(a) X and Y^2

These random variables are independent. For all x, y,

$$\begin{aligned}
\Pr(X = x \cap Y^2 = y) &= \Pr(X = x \cap Y = \sqrt{y}) \\
&= \Pr(X = x) \cdot \Pr(Y = \sqrt{y}) \\
&= \Pr(X = x) \cdot \Pr(Y^2 = y)
\end{aligned}$$

The second equation uses the independence of X and Y.

(b) $X + Y$ and $X - Y$

These random variables are not independent. For example:

$$\begin{aligned}
\Pr(X+Y=2) &= \frac{1}{36} \\
\Pr(X-Y=0) &= \frac{1}{6} \\
\Pr(X+Y=2 \cap X-Y=0) &= \frac{1}{36} \\
&\neq \frac{1}{36}\cdot\frac{1}{6}
\end{aligned}$$

(c) $\lfloor \frac{X-1}{2} \rfloor + \lfloor \frac{Y-1}{2} \rfloor$ and $(X \bmod 2) + (Y \bmod 2)$
These random variables are independent. The quantities $\lfloor \frac{X-1}{2} \rfloor$ and $(X \bmod 2)$ are independent, and so are the analogous expressions involving Y.

$$\begin{aligned}
\Pr\left(\left\lfloor \frac{X-1}{2} \right\rfloor = a\right) &= \frac{1}{3} && a \in \{0,1,2\} \\
\Pr\left(X \bmod 2 = b\right) &= \frac{1}{2} && b \in \{0,1\} \\
\Pr\left(\left\lfloor \frac{X-1}{2} \right\rfloor = a \;\cap\; X \bmod 2 = b\right) &= \frac{1}{6} && a \in \{0,1,2\}, b \in \{0,1\} \\
&= \frac{1}{3}\cdot\frac{1}{2}
\end{aligned}$$

Thus, the above expressions involve four mutually independent random variables; two are summed in the first expression and two are summed in the second. Consequently, the two expressions are independent.

Problem 2 **(15 points)** Let S and T be independent random variables with a countable range R. Let g be a function with domain and range R. Prove that S and $g(T)$ are independent.

We must prove that for all $x, y \in R$,

$$\Pr(S = x \cap g(T) = y) = \Pr(S = x) \cdot \Pr(g(T) = y).$$

Let $g^{-1}(y)$ be the set of all values that the function g maps to y. Then the event that $g(T) = y$ is identical to the event that $T \in g^{-1}(y)$.

$$\begin{aligned}
\Pr(S = x \cap g(T) = y) &= \Pr\left(S = x \cap T \in g^{-1}(y)\right) \\
&= \Pr\left(S = x \cap \left(\bigcup_{z \in g^{-1}(y)} T = z\right)\right) \\
&= \Pr\left(\bigcup_{z \in g^{-1}(y)} (S = x \cap T = z)\right) \\
&= \sum_{z \in g^{-1}(y)} \Pr(S = x \cap T = z) \\
&= \sum_{z \in g^{-1}(y)} \Pr(S = x) \cdot \Pr(T = z) \\
&= \Pr(S = x) \cdot \sum_{z \in g^{-1}(y)} \Pr(T = z) \\
&= \Pr(S = x) \cdot \Pr(T \in g^{-1}(y)) \\
&= \Pr(S = x) \cdot \Pr(g(T) = y)
\end{aligned}$$

The third step uses the fact that set intersection distributes over set union. The independence of S and T is used in the fifth step.

Problem 3 **(5 points)** Let the random variable R be equal to the number of heads that come up when four mutually independent, fair coins are flipped.

(a) Compute the probability distribution function for R. That is, determine $\Pr(R = k)$ for all k.

$$\begin{aligned}
\Pr(R = 0) &= \frac{1}{16} \\
\Pr(R = 1) &= \frac{4}{16} \\
\Pr(R = 2) &= \frac{6}{16} \\
\Pr(R = 3) &= \frac{4}{16} \\
\Pr(R = 4) &= \frac{1}{16}
\end{aligned}$$

(b) Compute the cumulative distribution function for R. That is, find $\Pr(R \leq k)$ for all k.

$$\begin{aligned}
\Pr(R \leq 0) &= \frac{1}{16} \\
\Pr(R \leq 1) &= \frac{5}{16} \\
\Pr(R \leq 2) &= \frac{11}{16} \\
\Pr(R \leq 3) &= \frac{15}{16} \\
\Pr(R \leq 4) &= \frac{16}{16}
\end{aligned}$$

Problem 4 (12 points)

Let random variables X and Y be the outcomes of two dice rolls. Assume that the dice are fair, 6-sided, and independent. Each problem part below lists two random variables derived from X and Y. State whether or not these two random variables are independent and justify your answer.

(a) $(X \bmod 3) + (Y \bmod 3)$ and $(X \bmod 2) + (Y \bmod 2)$

Independent. $X \bmod 3$ and $X \bmod 2$ are independent, which you can see by drawing a 3x2 table of values of $X \bmod 3$ versus values of $X \bmod 2$. Obviously, $Y \bmod 3$ and $Y \bmod 2$ are independent as well. Furthermore, it is true in general that if A is independent of B and A' is independent of B' then $f(A, B)$ is independent of $g(A', B')$ for any functions f and g. Here the two functions are addition over integers.

(b) $((X + Y) \bmod 2)$ and $((X * Y) \bmod 2)$

These random variables are not independent. $(X+Y \bmod 2) = 1$ iff X and Y have different parity. $(X * Y \bmod 2) = 1$ iff both X and Y have odd parity. Therefore, if $(X * Y \bmod 2) = 1$ then $(X + Y \bmod 2) = 0$, and hence the two events depend on each other.

Problem 5 (15 points)

Consider the process of tossing a sequence of n independent fair coins $X_1, X_2, \ldots, X_n$. Let V be the first time two consecutive heads occur in the sequence, i.e. $(X_{V-1}, X_V) = (H, H)$ and $(X_{i-1}, X_i) \neq (H, H)$ for all $2 \leq i < V$. If the sequence of n coin tosses does not contain two consecutive heads, then we assign $V = n + 1$.

Prove that V has distribution

$$\Pr(V = k) = \frac{\mathsf{Fib}(k-2)}{2^k} \qquad (2 \leq k \leq n)$$

where $\mathsf{Fib}(k)$ is the k-th Fibonacci number (assume $\mathsf{Fib}(0) = \mathsf{Fib}(1) = 1$.)

Let's count the sequences of coin tosses of length k such that the only pair *HH* that occurs in the sequence is at the end. Let $H(k)$ be the number of such sequences that start with H, and let $T(k)$ be the number of such sequences that start with T. Observe that $H(2) = 1$ and $T(2) = 0$. Similarly, $H(3) = 0$ and $T(3) = 1$. Given a valid sequence *HS* of length k that starts with H, the sequence *THS* is also valid and starts with T. Given a valid sequence *TS* of length k that starts with T, the sequences HTS and TTS are also valid and start with H and T respectively. From this we deduce the recurrence relations

$$\begin{aligned} H(k+1) &= T(k) \\ T(k+1) &= H(k) + T(k). \end{aligned}$$

The total number of valid sequences of length k is $T(k) + H(k)$, hence it satisfies the following recurrence

$$\begin{aligned} T(k+1) + H(k+1) &= (T(k) + H(k)) + T(k) \\ &= (T(k) + H(k)) + (T(k-1) + H(k-1)). \end{aligned}$$

This is the same recurrence relation that defines the Fibonacci numbers, and with the given initial conditions we can conclude that the total number of valid sequences of length k is $\mathsf{Fib}(k-2)$.

Now the total number of sequences of length n such that the first pair of heads occurs after k-th toss is $\mathsf{Fib}(k-2)2^{n-k}$, since we don't care what the coin tosses are in the last $n-k$ places.

The sample space in this problem is the set of all sequences of coin tosses of length n, so its size is 2^n. Therefore the probability

$$P(V = k) = \frac{\mathsf{Fib}(k-2)2^{n-k}}{2^n} = \frac{\mathsf{Fib}(k-2)}{2^k},$$

as desired.

Question for the curious (not graded): What is the cumulative distribution function for V? It is a similar expression involving $\mathsf{Fib}(k + c_1)$ and 2^{k+c_2} for some small c_1, c_2.

The equation is $\Pr\{V \le k\} = 1 - \frac{\mathsf{Fib}(k+1)}{2^k}$, and you can show it by induction: The base case: $\Pr\{V \le 2\} = 1/4 = 1 - 3/4 = 1 - \frac{\mathsf{Fib}(3)}{2^2}$. And the inductive case holds too.

Massachusetts Institute of Technology Handout 34
6.042J/18.062J: Mathematics for Computer Science December 11, 1998
Professors Nancy Lynch and Srinivas Devadas

Problem Set 11 Solutions

Problem 6 (10 points) Several statements involving probability are listed below. Assume that A and B are events, S and T are random variables, and the underlying sample space is countable. Furthermore, assume that all expectations exist. For each statement, describe restrictions on A, B, S, and T sufficient to ensure that the statement is true. Make your restrictions as minimal as you can. No proofs are necessary.

(a) $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

This statement is true provided that the event $A \cap B$ has probability zero.

(b) $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

This statement is always true.

(c) $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$

This statement is true provided that events A and B are independent.

(d) $\Pr(A \cap B) < \Pr(A)$

This statement is true provided $\Pr(A \cap \overline{B}) > 0$.

(e) $\mathrm{Ex}(S + T) = \mathrm{Ex}(S) + \mathrm{Ex}(T)$

This statement is always true.

(f) $\mathrm{Ex}(S \cdot T) = \mathrm{Ex}(S) \cdot \mathrm{Ex}(T)$

This statement is true provided that random variables S and T are independent.

Problem 7 **(10 points)** Suppose that I roll n dice that are 6-sided, fair, and mutually independent. What is the expected value of the largest number that comes up?

Let the random variable M be the largest number that comes up. We will use the identity

$$\mathrm{Ex}(M) = \sum_{i=0}^{\infty} \Pr(M > i)$$

We must compute the probability of the event $M > i$; that is, the event that some die shows a value greater than i.

$$\begin{aligned}
\mathrm{Ex}(M > i) &= \sum_{i=0}^{5} \Pr(M > i) \\
&= \sum_{i=0}^{5} 1 - \Pr(\text{every die} \le i) \\
&= \sum_{i=0}^{5} 1 - (\Pr(\text{one die} \le i))^n \\
&= \sum_{i=0}^{5} 1 - \left(\frac{i}{6}\right)^n \\
&= 6 - \frac{1^n + 2^n + 3^n + 4^n + 5^n}{6^n}
\end{aligned}$$

Problem 8 **(10 points)** Suppose that I choose a listing of the numbers $1, 2, \ldots, n$ uniformly at random. What is the expected number of entries that are greater than all preceding entries? For example, in the listing 4, 2, 1, 5, 3, the numbers 4 and 5 are greater than all preceding entries.

Let i_k be an indicator for the event that the k-th entry is greater than the preceding $k-1$ entries. By linearity of expectation

$$\begin{aligned}
\text{Ex}(\#\text{ new maxima}) &= \sum_{k=1}^{n} \text{Ex}(i_k) \\
&= \sum_{k=1}^{n} 1 \cdot \Pr(i_k = 1) + 0 \cdot \Pr(i_k = 0) \\
&= \sum_{k=1}^{n} \Pr(i_k = 1)
\end{aligned}$$

All that remains is to compute $\Pr(i_k = 1)$, which is the probability that the k-th entry is greater than all preceding entries. Note that each of the first k entries is equally likely to be the largest. Therefore, $\Pr(i_k = 1) = \frac{1}{k}$. Substituting into the equation above gives:

$$\begin{aligned}
\text{Ex}(\#\text{ new maxima}) &= \sum_{k=1}^{n} \Pr(i_k = 1) \\
&= \sum_{k=1}^{n} \frac{1}{k} \\
&= H_n
\end{aligned}$$

Problem 9 **(15 points)** Let R be a random variable taking on non-negative integer values. Let r be the probability generating function for R; that is, define

$$r(x) \quad = \quad \Pr(R = 0) + \Pr(R = 1) \cdot x + \Pr(R = 2) \cdot x^2 + \Pr(R = 3) \cdot x^3 + \ldots .$$

(a) Write an expression involving r that is equal to the expected value of R.

$$\begin{aligned}
r'(x) &= 1 \cdot \Pr(R = 1) + 2 \cdot \Pr(R = 2) \cdot x + 3 \cdot \Pr(R = 3) \cdot x^2 + \ldots \\
r'(1) &= 1 \cdot \Pr(R = 1) + 2 \cdot \Pr(R = 2) + 3 \cdot \Pr(R = 3) + \ldots \\
&= \text{Ex}(R)
\end{aligned}$$

(b) Suppose that $\Pr(R = k) = \frac{2}{3^{k+1}}$ for all $k \geq 0$. Find a closed-form expression for r, the probability generating function for R.

$$\begin{aligned}
r(x) &= \frac{2}{3^1} + \frac{2}{3^2} \cdot x + \frac{2}{3^3} \cdot x^2 + \frac{2}{3^3} \cdot x^3 + \ldots \\
&= \frac{\frac{2}{3}}{1 - \frac{x}{3}} \\
&= \frac{2}{3 - x}
\end{aligned}$$

The first equation follows from the definition of the probability generating function r. The second equation uses the formula for the sum of an infinite geometric series, and the third equation is obtained by simplification.

(c) Using your solution to part (a), determine the expected value of the random variable R as defined in part (b).

$$\begin{aligned} r'(x) &= \frac{2}{(3-x)^2} \\ r'(1) &= \frac{2}{(3-1)^2} \\ &= \frac{1}{2} \end{aligned}$$

Problem 10 The World Series

New York and Atlanta have both made it to the World Series again in 2000. Assume that New York wins each game with probability $2/3$ independently of the outcomes of other games[1]. Let R be the random variable denoting the number of the 4th game won by New York, e.g., if NY wins the 1st, 2nd, 4th, and 6th games, then $R = 6$; let $R ::= 0$ if NY doesn't win four games.

Let's generalize slightly and assume the teams keep playing, possibly forever, until NY has indeed won 4 games. This makes R a random variable taking on all natural number values.

(a) For $n > 0$, give a simple closed form expression for $\Pr\{R = n\}$.

For $0 < n < 3$, we have $\Pr\{R = n\} = 0$. For $4 \leq n$, we have

$$\Pr\{R = n\} = \binom{n-1}{3} \cdot (2/3)^4 \cdot \frac{1}{3^{n-4}} = \frac{8(n-1)(n-2)(n-3)}{3^{n+1}}.$$

(b) What is the most likely value of R?

5

(c) What is the probability that NY never wins 4 games, i.e., that $R = 0$?

0

(d) What is the probability that New York really wins the 7 game Series?

$$\Pr\{4 \leq R \leq 7\} = \Pr\{R = 4\} + \Pr\{R = 5\} + \Pr\{R = 6\} + \Pr\{R = 7\} = 0.8267$$

[1]—they wish :-)

Problem 11 **(10 points)** Suppose that I choose a listing of the numbers $1, 2, \ldots, n$ uniformly at random. What is the expected number of entries that are greater than all preceding entries? For example, in the listing 4, 2, 1, 5, 3, the numbers 4 and 5 are greater than all preceding entries.

Let i_k be an indicator for the event that the k-th entry is greater than the preceding $k-1$ entries. By linearity of expectation

$$\begin{aligned} \text{Ex}(\#\text{ new maxima}) &= \sum_{k=1}^{n} \text{Ex}(i_k) \\ &= \sum_{k=1}^{n} 1 \cdot \Pr(i_k = 1) + 0 \cdot \Pr(i_k = 0) \\ &= \sum_{k=1}^{n} \Pr(i_k = 1) \end{aligned}$$

All that remains is to compute $\Pr(i_k = 1)$, which is the probability that the k-th entry is greater than all preceding entries. Note that each of the first k entries is equally likely to be the largest. Therefore, $\Pr(i_k = 1) = \frac{1}{k}$. Substituting into the equation above gives:

$$\begin{aligned} \text{Ex}(\#\text{ new maxima}) &= \sum_{k=1}^{n} \Pr(i_k = 1) \\ &= \sum_{k=1}^{n} \frac{1}{k} \\ &= H_n \end{aligned}$$

Problem 12 **(10 points)** There is a dinner party for $n \geq 2$ couples. The hostess chooses a seating arrangement for the $2n$ people around a circular table uniformly at random. What is the expected number of couples such that the husband and wife are sitting next to each other?

Let I_k be an indicator variable for the event that couple k is sitting together. Then the number of couples sitting together is $I_1 + I_2 + \ldots + I_n$. The probability that a couple is seated together is $2/(2n-1)$, since a wife is equally likely to occupy each of the $2n-1$ positions relative to her husband and there are 2 positions right next to him. Consequently, $\Pr(I_k = 1) = 2/(2n-1)$, and we can reason as follows:

$$\begin{aligned}
\mathrm{Ex}(\text{couples sitting together}) &= \mathrm{Ex}\left(\sum_{k=1}^{n} I_k\right) \\
&= \sum_{k=1}^{n} \mathrm{Ex}(I_k) \\
&= \sum_{k=1}^{n} \Pr(I_k = 1) \\
&= \sum_{k=1}^{n} \frac{2}{2n-1} \\
&= \frac{2n}{2n-1}
\end{aligned}$$

Thus, regardless of the size of the party, the expected number of couples seated together is always close to 1.