Statistics 341

Fall 2008 - Homework Assignment #7

Due Friday, December 5

This assignment is worth a total of 95 points.

1. (49 pts) Roll two dice. Let Y_1 be the sum of the two dice and Y_2 be the number on the second dice rolled.

   (a) (10 pts) Find the joint probability distribution function for Y_1 and Y_2.
   The sample space for this experiment is below.

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

For each of the 36 outcomes we can determine the value of the random variables Y_1 and Y_2. They are

2	1	3	2	4	3	5	4	6	5	7	6
3	1	4	2	5	3	6	4	7	5	8	6
4	1	5	2	6	3	7	4	8	5	9	6
5	1	6	2	7	3	8	4	9	5	10	6
6	1	7	2	8	3	9	4	10	5	11	6
7	1	8	2	9	3	10	4	11	5	12	6

Each of the 36 outcomes has the same probability. The joint probability distribution function can be expressed by the table below. Each value in the table should be divided by 36.

	Y_2						
Y_1	1	2	3	4	5	6	Totals
2	1	0	0	0	0	0	1
3	1	1	0	0	0	0	2
4	1	1	1	0	0	0	3
5	1	1	1	1	0	0	4
6	1	1	1	1	1	1	5
7	1	1	1	1	1	1	6
8	0	1	1	1	1	1	5
9	0	0	1	1	1	1	4
10	0	0	0	1	1	1	3
11	0	0	0	0	1	1	2
12	0	0	0	0	0	1	1
Totals	6	6	6	6	6	6	36

(b) (6 pts; 3 pts each) Find the marginal probability distribution functions for Y_1 and Y_2.
The rows and columns in the table above are already summed. The marginal distributions of Y_1 and Y_2 are exactly what we would expect them to be. For Y_1, we get

y_1	2	3	4	5	6	7	8	9	10	11	12
$p_1(y_1)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

And for Y_2, we get

y_2	1	2	3	4	5	6
$p_2(y_2)$	1/6	1/6	1/6	1/6	1/6	1/6

(c) (4 pts) Are Y_1 and Y_2 independent? Explain your answer.
No. Look at the event $Y_1 = 2$ and $Y_2 = 2$. The value of $p(2, 2) = 0$, but $p_1(2) = 1/36$ and $p_2(y_2) = 1/6$.

(d) (3 pts) Find the conditional probability distribution function for Y_2 when $Y_1 = 7$.
When the sum of the two dice is 7, the distribution of Y_2 is

y_2	1	2	3	4	5	6
$p(y_2 \mid Y_1 = 7)$	1/6	1/6	1/6	1/6	1/6	1/6

(e) (3 pts) Find the conditional probability distribution function for Y_1 when $Y_2 = 3$.
When the second dice is 3, the distribution of Y_1 is

y_1	4	5	6	7	8	9
$p(y_1 \mid Y_2 = 3)$	1/6	1/6	1/6	1/6	1/6	1/6

(f) (12 pts; 3 pts for each of the 4 calculations) Find the covariance between Y_1 and Y_2.
For the covariance, we need the expected values of Y_1, Y_2, and Y_1Y_2. These values are

$$\begin{aligned} E(Y_1) &= \sum_{y_1} y_1 * p_1(y_1) \\ &= (1(2) + 2(3) + 3(4) + 4(5) + 5(6) + 6(7) \\ &\quad +5(8) + 4(9) + 3(10) + 2(11) + 1(12))/36 \\ &= 7 \end{aligned}$$

$$\begin{aligned} E(Y_2) &= \sum_{y_2} y_2 * p_2(y_2) \\ &= (1(1) + 2(1) + 3(1) + 4(1) + 5(1) + 6(1))/6 \\ &= 3.5 \end{aligned}$$

$$\begin{aligned} E(Y_1Y_2) &= \sum_{y_1}\sum_{y_2} y_1 y_2 p(y_1, y_2) \\ &= (1(2 + 3 + 4 + 5 + 6 + 7) \\ &\quad +2(3 + 4 + 5 + 6 + 7 + 8) \\ &\quad +3(4 + 5 + 6 + 7 + 8 + 9) \\ &\quad +4(5 + 6 + 7 + 8 + 9 + 10) \end{aligned}$$

$$
\begin{aligned}
&+5(6+7+8+9+10+11) \\
&+6(7+8+9+10+11+12))/36 \\
= \; & 987/36 = 27.4167
\end{aligned}
$$

$$
\begin{aligned}
Cov(Y_1, Y_2) &= E(Y_1Y_2) - E(Y_1)E(Y_2) \\
&= 27.4167 - 7(3.5) \\
&= 2.9167
\end{aligned}
$$

(g) (11 pts; 4 pts for each variance, and 3 pts for the correlation) Find the correlation between Y_1 and Y_2.

For the correlation, we need the variances of Y_1 and Y_2.

$$
\begin{aligned}
E(Y_1^2) &= \sum_{y_1} y_1^2 * p_1(y_1) \\
&= (1(2^2) + 2(3^2) + 3(4^2) + 4(5^2) + 5(6^2) + 6(7^2) \\
&\quad +5(8^2) + 4(9^2) + 3(10^2) + 2(11^2) + 1(12^2))/36 \\
&= 54.83333
\end{aligned}
$$

$$
\begin{aligned}
V(Y_1) &= E(Y_1^2) - (E(Y_1))^2 \\
&= 54.83333 - 7^2 \\
&= 5.83333
\end{aligned}
$$

$$
\begin{aligned}
E(Y_2^2) &= \sum_{y_2} y_2^2 * p_2(y_2) \\
&= (1^2(1) + 2^2(1) + 3^2(1) + 4^2(1) + 5^2(1) + 6^2(1))/36 \\
&= 91
\end{aligned}
$$

$$
\begin{aligned}
V(Y_2) &= E(Y_2^2) - (E(Y_2))^2 \\
&= 15.1667 - 3.5^2 \\
&= 2.9167
\end{aligned}
$$

$$
\begin{aligned}
Corr(Y_1, Y_2) &= \frac{Cov(Y_1, Y_2)}{\sqrt{(V(Y_1)V(Y_2)}} \\
&= \frac{2.9167}{\sqrt{5.8333 * 2.9167}} \\
&= 0.7071
\end{aligned}
$$

2. (46 pts) A playlist contains 15 songs, 5 songs from Artist 1, 5 songs from Artist 2, and 5 songs from Artist 3. Shuffle the playlist (randomly) and listen to the first 6 songs on the playlist. Let Y_1 be the number of songs from Artist 1 in the first 6 songs played and let Y_2 be the number of songs from Artist 2 in the first 6 songs played.

(a) (12 pts) Find the joint probability distribution function for Y_1 and Y_2.

To find the number of outcomes in each combination of Y_1 and Y_2, we will need to calculate

$$\binom{5}{y_1}\binom{5}{y_2}\binom{5}{6-y_1-y_2}$$

The table below contains these values. To get the probabilities, each number should be divided by 5005.

	Y_2						
Y_1	0	1	2	3	4	5	Totals
0	0	5	50	100	50	5	210
1	5	125	500	500	125	5	1260
2	50	500	1000	500	50	0	2100
3	100	500	500	100	0	0	1200
4	50	125	50	0	0	0	225
5	5	5	0	0	0	0	10
Totals	210	1260	2100	1200	225	10	5005

(b) (8 pts; 3 pts each and 2 pts for hypergeometric) Find the marginal probability distribution functions for Y_1 and Y_2. Both distributions are the same as one we have studied in the past. What is the distribution and parameters?

The marginal distributions of Y_1 and Y_2 are the same. Here is the table.

y_1 or y_2	0	1	2	3	4	5
$p_1(y_1)$ or $p_2(y_2)$	210/5005	1260/5005	2100/5005	1200/5005	225/5005	10/5005

This distribution is a hypergeometric distribution with $n = 6, N = 15, r = 5$.

(c) (4 pts) Are Y_1 and Y_2 independent? Explain your answer.

No, look at the outcome $y_1 = 0, y_2 = 0$. The joint probability distribution function is p(0,0) = 0, while the values of $p_1(0) = p_2(0) = 210/5005$.

(d) (3 pts) Find the conditional probability distribution function for Y_2 when $Y_1 = 3$.

y_2	0	1	2	3
$p(y_2 \mid Y_1 = 3)$	100/1200	500/1200	500/1200	100/1200

(e) (3 pts) Find the conditional probability distribution function for Y_1 when $Y_2 = 2$.

y_1	0	1	2	3	4
$p(y_1 \mid Y_2 = 2)$	50/2100	500/1200	1000/1200	500/1200	50/2100

(f) (9 pts; 3 pts for each calculation) Find the covariance between Y_1 and Y_2.

For the covariance, we need the expected values of Y_1, Y_2, and Y_1Y_2. The expected value of Y_1 and Y_2 will be the same. These values are

$$E(Y_1) = \sum_{y_1} y_1 * p_1(y_1)$$

$$
\begin{aligned}
&= (0(210) + 1(1260) + 2(2100) + 3(1200) + 4(225) + 5(10)) \\
&= 10010/5005 = 2
\end{aligned}
$$

$$
\begin{aligned}
E(Y_1Y_2) &= \sum_{y_1}\sum_{y_2} y_1y_2p(y_1, y_2) \\
&= (1(1(125) + 2(500) + 3(500) + 4(125)5(5)) \\
&\quad +2(1(500) + 2(1000) + 3(500) + 4(50)) \\
&\quad +3(1(500) + 2(500) + 3(100)) \\
&\quad +4(1(125) + 2(50)) \\
&\quad +5(5(1)))/5005 \\
&= 3.574129 \\
Cov(Y_1, Y_2) &= E(Y_1Y_2) - E(Y_1)E(Y_2) \\
&= 3.574129 - 2(2) \\
&= -0.4285714
\end{aligned}
$$

Note: You can get the expected value of Y_1 and Y_2 by using the fact that these are hypergeometric random variables with $r = 5, n = 5$ and $N = 15$.

(g) (7 pts; 4 pts for variance and 3 pts for correlation) Find the correlation between Y_1 and Y_2.

For the correlation, we need the values of the variance of Y_1 and Y_2. These two values will be the same.

$$
\begin{aligned}
E(Y_1^2) &= \sum_{y_1} y_1^2 * p_1(y_1) \\
&= (0^2(210) + 1^2(1260) + 2^2(2100) + 3^2(1200) + 4^2(225) + 5^2(10))/5005 \\
&= 24310/5005 = 4.857143
\end{aligned}
$$

$$
\begin{aligned}
V(Y_1) &= E(Y_1^2) - (E(Y_1))^2 \\
&= 4.857143 - 2^2 \\
&= 0.857143
\end{aligned}
$$

$$
\begin{aligned}
Corr(Y_1, Y_2) &= \frac{Cov(Y_1, Y_2)}{\sqrt{(V(Y_1)V(Y_2)}} \\
&= \frac{-0.4285714}{\sqrt{0.857143 * 0.857143}} \\
&= -0.5
\end{aligned}
$$

Note: You could also find the variance of Y_1 and Y_2 by using the fact these are hypergeometric random variables with $r = 5, n = 5$ and $N = 15$.