MIT Introduction to Statistics 18.05 Slides 4 - Questions

John Hancock

February 23, 2017

Contents

1 References and License

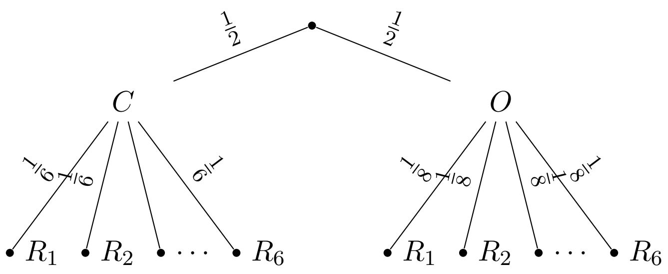
We are answering questions in the material from MIT OpenCourseWare course 18.05, Introduction to Probability and Statistics.
Please see the references section for detailed citation information.
The material for the course is licensed under the terms at `http://ocw.mit.edu/terms`.
We are answering the questions that Orloff and Bloom ask in [3].
We use documentation in [6] to write LaTeXsource code for this document.

2 Conditional Probability of Unknown Die

The first question Orloff and Bloom give in [3] is:

> 1. The Randomizer holds the 6-sided die in one fist and the 8-sided die in the other.
> 2. The Roller selects one of the Randomizers fists and covertly takes the die.
> 3. The Roller rolls the die in secret and reports the result to the table.
>
> Given the reported number, what is the probability that the 6-sided die was chosen?

Note: we needed to see the solution in [4] in order to write the answer to this question.
We have two cases.
The first case is the Roller reports a 7 or an 8. Then the probability that the 6-sided die was chosen is 0.
The second case is the Roller reports a number with a value from 1 to six.
We draw a probability tree to get started on a solutionn. We refer to the section titled, "Shorthand vs. precise trees," in [1] for guidance on drawing the tree below.
C is the event that the Roller selects the cube shaped 6-sided die.
O is the event that the Roller selects the octohedron shaped 8-sided die.
$R1$, $R2$, …, $R6$ are the events that the Roller reports one, two, or so on to six.



Since the probabilities on all edges in the tree connected to C are $\frac{1}{6}$, and the probabilities on all edges in the tree connected to O are $\frac{1}{8}$, we can calculate $P(C \mid R_1)$, and the result will be the same for any of the other leaf nodes in the tree above. This is because the calculation will involve the same numbers $\frac{1}{2}$, $\frac{1}{6}$, and $\frac{1}{8}$, and the same operations on these numbers.
Using the tree above, we can calculate $P(R_1 \mid C)$.
Now, we use Bayes' theorem in [1] to calculate $P(C \mid R_1)$

$$P\left(C \mid R_1\right) = \frac{P\left(R_1 \mid C\right) P\left(C\right)}{P\left(R_1\right)} \tag{1}$$

Now, we apply definitions for values on various parts of probability trees using the section titled "Shorthand vs. precise trees," in [1] to obtain values for the numerator and denominator on the righthand side of 1.
From the probability tree,

$$P\left(R_1 \mid C\right) = \frac{1}{6} \tag{2}$$

$P\left(C\right) = \frac{1}{2}$. Note: we are assuming the Roller uses either die with equal probability.
We apply Bayes rule [1] and the Law of Total Probability [1] to compute $P\left(R_1\right)$.

$$P\left(R_1\right) = P\left(R_1 \cap C\right) + P\left(R_1 \cap O\right) = P\left(R_1 \mid C\right) P\left(C\right) + P\left(R_1 \mid O\right) P\left(O\right) = \left(\frac{1}{6}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{8}\right)\left(\frac{1}{2}\right) \tag{3}$$

Now we have values for the numerator and denominator of the right hand side of 1.

$$P\left(C \mid R_1\right) = \frac{\left(\frac{1}{6}\right)\left(\frac{1}{2}\right)}{\left(\frac{1}{6}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{8}\right)\left(\frac{1}{2}\right)} \tag{4}$$

3 Concept Questions for CDF and PMF

In [3] Orloff and Bloom give us a table for a random variable X:

values of X	1	3	5	7
cdf $F\left(a\right)$	0.5	0.75	0.9	1

3.1 $P\left(X \leq 3\right)$

$P\left(X \leq 3\right) = 0.75$. We know this because of the definition of cdf, and because Orloff and Bloom give us the value of the cdf of X for $X \leq 3$.

3.2 $P\left(X = 3\right)$

We know

$$F\left(X \leq 3\right) = P\left(X = 3\right) + P\left(X = 1\right) \tag{5}$$

Therefore

$$P\left(X = 3\right) = F\left(X \leq 3\right) - P\left(X = 1\right) = 0.75 - 0.5 = 0.25 \tag{6}$$

4 Sum of Binomial Random Variables

4.1 Sum of Binomial Random Variables with Same Heads Probability

In [3] Orloff an Bloom pose the question that if $X \sim \text{binomial}(n, p)$ and $Y \sim \text{binomial}(m, p)$, what distribution does $X + Y$ follow?
In [2] Orloff and Bloom give the definition of a random variable that folllows a binomial distribution.
Therefore we apply the definition of a binomial distribution to the binomial random variables Orloff and Bloom give us in this section:
X is the number of heads in n independent Bernoulli trials, and Y is the number of heads in m independent Bernoulli trials.
For both X and Y we are given that the Bernoulli trials have the same probablity for success, p.
Therefore $X + Y$ is the number of successes in $n + m$ independent Bernoulli trials with a probability of success p.
Therefore

$$(X + Y) \sim \text{binomial}(n + m, p) \tag{7}$$

4.2 Sum of Binomial Random Variables with Different Heads Probability

We arrive at the answer by process of elimination.
In [3] Orloff and Bloom give us two random variables: X, and Z, where $X \sim \text{binomial}(n, p)$, and $Z \sim \text{binomial}(n, q)$.
Orloff and Bloom then ask us which distribution $X + Z$ follows, and they give us four options to choose from.
The first option is $\text{binomial}(n, p + q)$. This cannot be correct because $X + Z$ is a sum of the number of successes in $n + n$ independent Bernoulli trials.
The second option is $textbinomial\,(n, pq)$. This also cannot be correct because $X + Z$ is a sum of the number of successes in $n + n$ independent Bernoulli trials.
The third option is $\text{binomial}(2n, p + q)$. This cannot be correct because there is a counterexample.
We construct the counterexample: suppose $X \sim \text{binomial}\left(n, \frac{2}{3}\right)$, and $Z \sim \text{binomial}\left(n, \frac{2}{3}\right)$.
Then, if the third option were correct, $X + Z \sim \text{binomial}\left(2n, \frac{4}{3}\right)$. No probability can be greater than 1, so the third option cannot be correct.
This leaves us with the final option of, "other."

5 Number of Successes Before Second Failure

In [3] Orloff and Bloom ask us to describe the pmf of a random variable X where X is the number of successes before the second failure of a sequence of independent Bernoulli trials.
Let ω be a sequence of trials that fits the description of a sequence of trials that Orloff and Bloom give in this question.
Let Ω be the set of all ω.
We assume ω has $n+2$ trials, where n of the trials are successful, and two of the trials are failures.
Orloff and Bloom implicitly state that all the sequences of trials end in a failure because they are asking for the number of successes before the second failure.
Therefore the $(n+2)^{nd}$ element of ω is the second failure.
We can partition Ω into $n+1$ disjoint subsets containing one element each, where each subset has the first failure in a different position.
Therefore we can apply the Law of Total Probability [1] to comupte the probability of the first failure occuring in any of the $n+1$ positions in ω to be $(n+1)(1-p)$, where p is the probability of a successful Bernoulli trial in ω.
There are n independent successful trials in ω, with probability p, one unsuccessful independent trial in ω with probability $(n+1)(1-p)$, and one final unsuccesful independent trial in ω with probability $(1-p)$.
We know from [1] that the probability of the union of these independent events is equal to the product of the probabilities of the events. Therefore

$$p(\omega) = p^n (n+1)(1-p)^2 \tag{8}$$

6 Forgetful Geometric Random Variables

In [3] Orloff and Bloom as us to show that for a random variable X that follows a geometric distribution of independent Bernoulli trials with probability p.

$$P(X = n+k \mid X \geq n) = P(X = k) \tag{9}$$

Proof. We assume without loss of generality that X is the count of the number of successful independent Bernoulli trials before the first failed Bernoulli trial.
We apply Bayes' theorem [1] to 9:

$$P(X = n+k \mid X \geq n) = \frac{P(X \geq n \mid X = n+k)\, P(X = n+k)}{P(X \geq n)} \tag{10}$$

If we are given that $X = n + k$, then we are certain that $X \geq n$. So, $P(X \geq n \mid X = n + k) = 1$.
Therefore we can rewrite equation 10:

$$P(X = n + k \mid X \geq n) = \frac{P(X = n + k)}{P(X \geq n)} \tag{11}$$

The denominator of the right hand side of 11 is the probability that in a series of n independent Bernoulli trials with probability p. we have n or more successes before the first failure. The probability of a series of Bernoulli trials having n or more successes before the first failure is the same as the probability of having exacttly n successes and no further successes and further successes. The probability of having further successes and no further successes is 1. The probability of n successful Bernoulli trials before the first failure is the probability mass function for the geometric distribution of a random variable with probability p. In [2] Orloff and Bloom show that this probability is $p^n(1-p)$. This is the denominator of the right hand side of equation 11.
The numerator of the right hand side of equation 11 is the probability of a series of $n + k$ successful Bernoulli trials before the first failed trial. So the numerator of the right hand side of 11 is equal to the probbability mass function for a geometric distribution with probability p. Hence the numerator of the right hand side of equation 11 is $p^{n+k}(1-p)$.
We substitute the values we find for the numerator and denominator of the right hand side of equation 11 to obtain:

$$P(X = n + k \mid X \geq n) = \frac{p^{n+k}(1-p)}{p^n(1-p)} = p^k \tag{12}$$

$\square$

This means that given n successes, the probability of k more successes is the same as the probability of k successes with no prior successes. We can state this another way, writing that prior successes have no influence over subsequent successes. It is as if the random variable that follows a geometric distribution has no memory of its previous values.

7 Expected Value as Payoff

In [3] Orloff and Bloom give us two scenarios, and ask us for a decision about what to do given each scenario.
The first scenario is that there is a game of chance where we have a 10% chance of winning 95 dollars, and a 90% chance of losing 5 dollars.
Orloff and Bloom then ask us whether or not we would play this game of chance.

We make our decision of whether or not to play based on the expected value of how much money we win. If the expected value of money we win is positive, then we decide to play. If the expected value of money we win is negative, meaning a loss, we decide not to play.
We use the definition of expected value Orloff and Bloom give in [5].
The expected value of how much money we win is $\frac{1}{10}95$ dollars$-\frac{9}{10}5$ dollars $=$ $\frac{50}{10}$dollars $= 5$ dollars.
The expected value of money we win is positive, so we decide to play.
The second scenario Orloff and Bloom give is that we have an opportunity to buy a lottery ticket that costs 5 dollars. If we buy the lottery ticket, we have a 10% chance to win 100 dollars, and a 90% chance to win 0 dollars. Orloff and Bloom then ask us whether or not we should buy a lottery ticket. If we buy a winning lottery ticket then our net gain is $100 - 5 = 95$ dollars. So, if we buy a winning lottery ticket, we have a 10% chance to gain 95 dollars.
If we buy a losing lottery ticket, then our net gain is $0 - 5 = -5$ dollars. Thus, we have a 90% chance of losing 5 dollars.
The expected value of how much money we win is therefore $\frac{1}{10}95$ dollars $-$ $\frac{9}{10}5$ dollars $=$ $\frac{50}{10}$dollars $= 5$ dollars.
This value is positive, so we decide to buy a lottery ticket.

8 Seating Arrangements

The last question Orloff and Bloom present in [3] is a question on seating arrangements.
First, Orloff and Bloom ask us to suppose that we have n people seated a certain way, and for some reason they all get up and randomly choose new seats, with any choice of seat being equally likely.
Orloff and Bloom then ask us for the expected value of the number of people that choose the same seat they were sitting in previously.
We will exploit the first property 3 of expected values of sums of random variables from [5].
This property is

$$E\left(X + Y\right) = E\left(X\right) + E\left(Y\right) \tag{13}$$

In [5] Orloff and Bloom state that equation 13 holds when X, and Y are varaibles on a sample space Ω.
Suppose $Y = X_1 + X_2$, where X_1 and X_2 are also random variables on Ω. Then we may apply equation 13 to obtain

$$E\left(X + Y\right) = E\left(X\right) + E\left(X_1\right) + E\left(X_2\right) \tag{14}$$

Now suppose $Y = X_1 + X_2 + \ldots + X_n$. We use the associative property of the $+$ operator to write $Y = X_1 + (X_2 + \ldots + X_n)$.

Then, by 13 $E(X+Y) = E(X) + E(X_1) + E(X_2 + X_3 + \ldots + X_n)$.
We can repeatedly apply 13 to show

$$E(X+Y) = E(X_1) + E(X_2) + \ldots + E(X_n) \tag{15}$$

Returning to the question at hand, let X_i be the random variable that has value 1 if the i^{th} person chooses the same seat she sat in originally, and 0 otherwise. The people are equally likely to choose any seat, so there is a $\frac{1}{n}$ probability that the i^{th} person will choose the same seat. Therefore the expected value of X_i is $1 \times \frac{1}{n} + 0 \times n - 1n$. That is to say, $E(X_i) = \frac{1}{n}$.
We let X be the random variable that is equal to the total number of people who choose the same seat. Then

$$X = \sum_{i=1}^{n} X_i \tag{16}$$

In order to answer this question we must know $E(X)$. Since X is defined as it is in equation 16, and equation 15 holds,

$$E(X) = \sum i = 1^n \frac{1}{n} = 1 \tag{17}$$

References

[1] Jeremy Orloff and Jonathan Bloom. *Conditional Probability, Independence and Bayes Theorem Class 3, 18.05, Spring 2014*. Available at `https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading3.pdf` (Spring 2014).

[2] Jeremy Orloff and Jonathan Bloom. *Discrete Random Variables Class 4, 18.05, Spring 2014 Jeremy Orloff and Jonathan Bloom*. Available at `https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading4a.pdf` (Spring 2014).

[3] Jeremy Orloff and Jonathan Bloom. *Discrete Random Variables; Expectation 18.05 Spring 2014 Jeremy Orloff and Jonathan Bloom*. Available at `https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/class-slides/MIT18_05S14_class4slides.pdf` (Spring 2014).

[4] Jeremy Orloff and Jonathan Bloom. *Discrete Random Variables; Expectation 18.05 Spring 2014 Jeremy Orloff and Jonathan Bloom*. Available at `https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/class-slides/MIT18_05S14_class4_slides.pdf` (Spring 2014).

[5] Jeremy Orloff and Jonathan Bloom. *Discrete Random Variables: Expected Value Class 4, 18.05, Spring 2014 Jeremy Orloff and Jonathan Bloom*. Available at https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading4b.pdf (Spring 2014).

[6] ShareLaTex. *Typesetting quotations*. Available at https://www.sharelatex.com/learn/Typesetting_quotations (Spring 2014).