Problem Set 4

John Hancock

April 25, 2017

Contents

1 References and License

We are answering questions in the material from MIT OpenCourseWare course 18.05, Introduction to Probability and Statistics.
In this document we are answering questions Orloff and Bloom ask in [7].
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 use documentation in [**logicNot**], [**proofs**], [**bars**], [**packageClash**], [**curlyBrace**], [**cases**] to write the LaTeXsource code for this document.

2 Estimate Error

The first question Orloff and Bloom ask in [7] is about an accountant that rounds his cacluations (entries) to the nearest dollar. We assume the accountant has made 300 calculations. Orloff and Bloom want us to estimate the probability that the total error is greater than five dollars.
We use the central limit theorem [4] and techniques for estimating probability that Orloff and Bloom show in [4] in order to find this estimate.
In order to apply the central limit theorem, we first define a random variable X_i. X_i is the error the accountant makes on her $i^t h$ calculation. Orloff and Bloom tell us that X_i is uniformly distributed on $[-0.5, 0.5]$.
We also need the mean μ, and standard deviation σ of X_i in order to make our estimate.
In [6] Orloff and Bloom state that a uniformly distributed random variable on $[a, b]$ has the distribution function $f(x) = \frac{1}{a-b}$.
In [5] Orloff and Bloom define the mean $E(X)$ of a continuous random variable X with pdf $f(x)$ to be:

$$E(X) = \int_a^b x f(x)\, dx \tag{1}$$

For this problem, $f(x) = \frac{1}{-0.5-0.5} = -1$. Therefore we apply Orloff and Bloom's definition of the mean value of a continuous random variable to find that the mean value of X_i is

$$E(X_i) = \int_{-0.5}^{0.5} -x\, dx. \tag{2}$$

We use the power rule for integrals from [1] to find the antiderivative of the function above that we must integrate in order to find the mean value of X_i. The antiderivative of $g(x) = -x$ is $\frac{-x^2}{2}$.
We replace the integral on the right hand side of the equation above with its antiderivative:

$$E(X_i) = \left.\frac{-x^2}{2}\right|_{-0.5}^{0.5}. \tag{3}$$

And we evaluate the antiderivative over the interval $[-0.5, 0.5]$:

$$E(X_i) = \frac{-\left(-0.5^2\right)}{2} - \frac{-\left(0.5^2\right)}{2} \tag{4}$$

Now we do some arithmetic to simplify the right hand side of the equation above:

$$E(X_i) = \frac{-1}{8} - \frac{-1}{8} = 0. \tag{5}$$

In order to find the standard deviation of X_i, we use a property of variance from [5], for a continuous random variable X:

$$\mathrm{Var}\left(X\right) = E\left(X^2\right) - E\left(X\right)^2. \tag{6}$$

We apply the same reasoning to find $E\left(X_i^2\right)$ that we use to find $E\left(X_i\right)$:

$$E\left(X_i^2\right) = \int_{-0.5}^{0.5} -\left(x^2\right)\, dx. \tag{7}$$

This implies:

$$E\left(X_i^2\right) = \left.\frac{-x^3}{3}\right|_{-0.5}^{0.5}. \tag{8}$$

Which implies

$$E\left(X_i^2\right) = \frac{-\left(-0.5^3\right)}{3} - \frac{-\left(0.5^3\right)}{3} \tag{9}$$

The right hand side of the equation above simplifies to:

$$E\left(X_i^2\right) = \frac{-\left(-1\right)}{24} - \frac{-1}{24} \tag{10}$$

Therefore the variance of X_i is $\frac{1}{12}$.
Orloff and Bloom ask us to estimate the probability of the size of the error the accountant makes after 300 calculations. So, we define a random variable S to be the sum of 300 values of the X_i. Therefore S is the total erorr that the accountant makes after 300 calculations.
In order use the central limit theorme to estimate the probability that a random variable is in a range we need to know its mean and standard deviation.
Thereofre we need to know the mean of S. We start with:

$$E\left(S\right) = E\left(\sum_{i=1}^{n} X_i\right). \tag{11}$$

We use a property of expected value from [5] to find that the mean value

$$E\left(S\right) = \sum_{i=1}^{3} 00E\left(X_i\right). \tag{12}$$

Above we found that $E\left(X_i\right) = 0$. Therefore, by the equation above, $E\left(S\right) = 0$. Mean and expected value are synonymous, and to use the notation Orloff and Bloom use in their treatment of the central limit theorem we write the mean of S, $\mu_s = 0$.

Now we turn our attention to finding the variance and standard deviaion of S.
In [5] Orloff and Bloom state that the variance of the sum of independent random varialbes is the sum of their variances. We assume the collection of X_i are independent.
This assumption allows us to write that the variance of S,

$$\mathrm{Var}\,(S) = \sum_{i=1}^{3} 00\frac{1}{12} = 25. \tag{13}$$

Becausee standard deviation is the square root of variance, the standard deviation of S, σ_S is 5.
Orloff and Bloom are asking us to compute the probability that the total error the accountant makes after 300 calculations is more than 5$. The total error the accountant makes might be a positive or negative value, so we need to estimate the probability that $S < -5$\$ or $S > 5$. However, this probability is equal to $1 - P\,(-5 \leq S \leq 5)$. We state this relationship with the equation:

$$P\,(|S| < 5) = 1 - P\,(-5 \leq S \leq 5)\,. \tag{14}$$

The central limit theorem states that standardized S approximately follows the normal distribution $N\,(0, 1)$.
We standardize S, and apply the central limit theorem like Orloff and Bloom do in [4] to get the approximation:

$$P\,(-5 \leq S \leq 5) = P\left(-\frac{5}{5} \leq \frac{S - \mu_S}{\sigma_S} \leq \frac{5}{5}\right) = P\left(-1 \leq \frac{S-0}{5} \leq 1\right) \approx P\,(-1 \leq Z > 1) \tag{15}$$

Note: the equations above are legal because S is a continuous random variable, and therefore we compute the eprobability that S is in a given interval with and integral. In the equations above we are using the property of integration that states a constant times the integral of a function is the integral of the constant times that function [1].
The rule of thumb [4] tells us that $P\,(-1 \leq Z \leq 1) \approx 0.68$. We use equation 14 to obtain our estimate that the probability that the total error the accountant makes after 300 calculations is more than 5$ is 0.32.

3 Difference of Dice

The second question Orloff and Bloom have is on the discrete events. Orloff and Bloom define two discrete random variables X, and Y. X and Y are the values we roll using two six-sided dice. They then define the event A as the event where the difference $Y - X$ is greater than or equal to 2.

The event A is a set of outcomes. Each member of A is an outcome of an event where we roll two dice, and the difference between the value we roll with the first die, and the value we roll with the second die is greter than or equal to 2.
We arrange the possible differences of X and Y in a table.
Thus, we represent all possible outcomes the event where we roll two dice, and then subtracting the value we roll with the first die from the value we roll with the second die as a cell in the table below.

X/Y	1	2	3	4	5	6
1	0	-1	-2	-3	-4	-5
2	1	0	-1	-2	-3	-4
3	2	1	0	-1	-2	-3
4	3	2	1	0	-1	-2
5	4	3	2	1	0	-1
6	5	4	3	2	1	0

Note: there is a probabiltiy of $\frac{1}{36}$ for any outcome we represent as a cell in the table above.
We have colored any square that represents an outcome in event A in blue.
By inspection 10 out of the 36 squares in the table above represent events in A. These events are disjoint, so we we can sum their probabilities to compute the probability of A:

$$P(A) = 10 \times \frac{1}{36} = \frac{5}{18}. \tag{16}$$

4 Continuous event

The next task Orloff and Bloom have for involves a continuous joint distribution (X, Y). The distribution is defined on $[0, 1] \times [0, 1]$. The probability density function of the joint distribution is $f(x, y) = 1$.
The first part of the task Orloff and Bloom give for this continuous joint distribution is for us to visualize the event $X > Y$.
The event $X > Y$ is half of a cube. The cube has corners at the coordinates: $(0, 0, 0))$, $(0, 0, 1)$, $(0, 1, 0)$, $(0, 1, 1)$, $(1, 0, 0)$, $(1, 0, 1)$, $(1, 1, 0)$, and $(1, 1, 1)$.
The event that $X > Y$ is the half of the cube with corners at $(0, 0, 0)$, $(0, 0, 1)$, $(1, 0, 0)$, $(1, 0, 1)$, $(1, 1, 0)$, and $(1, 1, 1)$.
The volume of the cube is 1. Therefore the event $X > Y$ has probability 0.5.

5 Random variables with pdf

The questions in this section are regarding a joint distribution of two continuous random variables X, and Y.

Orloff and Bloom state that (X, Y) has values in $[0,1] \times [0,1]$.
They also state that the pdf for the join distribution is $\frac{3}{2}\left(x^2 + y^2\right)$.

5.1 Valid probability density function

The first thing that Orloff and Bloom tell us to do with this joint distribution is show that f is a valid probability density function (pdf).
In [2] Orloff and Bloom state that a joint pdf must satisfy two properties:

1. $0 \leq f(x, y)$
2. The total probability is 1

Proof. First we show that for any $(x, y) \in [0,1] \times [0,1]$, $0 \leq f(x, y)$.
If $(x, y) \in [0,1] \times [0,1]$, then $0 \leq \frac{3}{2}\left(x^2 + y^2\right)$.
Now we will compute

$$\int_0^1 \int_0^1 \frac{3}{2}\left(x^2 + y^2\right)\, dy\, dx. \tag{17}$$

First we integrate the expression above with respect to y:

$$\int_0^1 \int_0^1 \frac{3}{2}\left(x^2 + y^2\right)\, dy\, dx = int_0^1 \frac{3}{2}\left(x^2 y + fracy^2 2\right)\Big|_0^1 dx. \tag{18}$$

Now we evaluate the resulting antiderivative over the interval $[0,1]$:

$$int_0^1 \frac{3}{2}\left(x^2 y + fracy^2 2\right)\Big|_0^1 dx = int_0^1 \frac{3}{2}\left(x^2 + frac12\right)\, dx \tag{19}$$

Now we integrate with respect to x:

$$int_0^1 \frac{3}{2}\left(x^2 + frac12\right)\, dx = int_0^1 \frac{3}{2}\left(\frac{x^3}{3} + fracx2\right)\Big|_0^1, \tag{20}$$

and evaluate the antiderivative over the interval $[0,1]$:

$$int_0^1 \frac{3}{2}\left(\frac{x^3}{3} + fracx2\right)\Big|_0^1 = \frac{3}{2}\left(\frac{1}{3} + frac12\right). \tag{21}$$

Now we do some arithmetic to simplify the right hand side of the equation above:

$$\frac{3}{2}\left(\frac{1}{3} + frac12\right) = \frac{3}{2}\left(\frac{2}{3}\right) = 1. \tag{22}$$

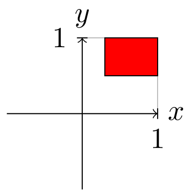
Therefore,

$$\int_0^1 \int_0^1 \frac{3}{2}\left(x^2 + y^2\right)\, dy\, dx = 1. \tag{23}$$

We have shown that $f(x, y)$ satisfies the two properties that Orloff and Bloom state a pdf must satisfy in [2]. Therefore $f(x, y)$ is a valid pdf. $\square$

5.2 Visualize event

The next task Orloff and Bloom give us is to visualize the event $A = `X > 0.3 and Y > 0.5'$.
Because (X, Y) takes values on $[0, 1] \times [0, 1]$, we visualize the event A as The square with corners $(0, 0.5)$, $(0, 1)$, $(1, 1)$, $(0.3, 0.5)$.
Here is a plot of the region:



Now we calculate the probability of A.
We use the definition of the probability of a joint distribution function from [2], and we apply the joint pdf of X, and Y to this defintition:

$$P(A) = \int_{0.5}^{1} \int_{0.3}^{1} \frac{3}{2} \left(x^2 + y^2\right) \, dx \, dy. \tag{24}$$

First we find the antiderivative of the pdf with respect to x:

$$P(A) = \int_{0.5}^{1} \frac{3}{2} \left(\frac{x^3}{3} + xy^2\right) \Bigg|_{x=0.3}^{1} dy. \tag{25}$$

And now we find the antiderivative of the pdf with respect to y:

$$P(A) = \frac{3}{2} \left(\frac{x^3 y}{3} + x \frac{y^3}{3}\right) \Bigg|_{x=0.3}^{1} \Bigg|_{y=0.5}^{1}. \tag{26}$$

We apply the distributive rule for multiplication to find that the equation above is true if, and only if:

$$P(A) = \frac{1}{2} \left(x^3 y + xy^3\right) \Bigg|_{x=0.3}^{1} \Bigg|_{y=0.5}^{1}. \tag{27}$$

Now we evaluate the antiderivatives over the intervals we specifiy in order to compute the probability:

$$P(A) = \left(\frac{1}{2} \left((1)^3 (1) + (1) (1)^3\right)\right) - \left(\frac{1}{2} \left((0.3)^3 (0.5) + (0.3) (0.5)^3\right)\right). \tag{28}$$

The equation above simplifies to:

$$P(A) = 1 - \left(\frac{1}{2} \left((0.3)^3 (0.5) + (0.3) (0.5)^3\right)\right). \tag{29}$$

We use a calculator to simplify the equation above further to:

$$P(A) = 1 - \left(\frac{1}{2}(0.0135 + 0.0375)\right). \tag{30}$$

We use a calculator once more to find that $P(A) = 0.9745$. Note: this differs from Orloff and Bloom's solution, however, it seems they mistakenly use 4xy for the pdf of the joint distribution of X, and Y in [7]

5.3 Cumulative distribution function

The cumulative distribution function (cdf), $F(x, y)$ of the pdf for the joint distribution of X, and Y the integral of its probability density function $f(x, y)$ that Orloff and Bloom give for this problem [2].
We replace vairables x, and y with u, and v, so that we may use x, and Y as the upper limits of our integral. Doing so yeilds a cdf in terms of x, and y.

$$F(X, Y) = \int_0^y \int_0^x \frac{3}{2}\left(u^2 + v^2\right) du\, dv. \tag{31}$$

Now we integrate with respect to u:

$$F(X, Y) = \int_0^y \frac{3}{2}\left(\frac{u^3}{3} + uv^2\right)\Bigg|_{u=0}^{x} dv. \tag{32}$$

Next, we integrate with respect to v:

$$F(X, Y) = \frac{3}{2}\left(\frac{uv^3}{3} + \frac{uv^3}{3}\right)\Bigg|_{u=0}^{x}\Bigg|_{v=0}^{y}. \tag{33}$$

We evaluate the antiderivative at the limits of integration specified in the equation above to obtain:

$$F(X, Y) = \left(\frac{3}{2}\left(\frac{x^3 y}{3} + \frac{xy^3}{3}\right)\right) - \left(\frac{3}{2}\left(\frac{(0)(0)^3}{3} + \frac{(0)(0)^3}{3}\right)\right) \tag{34}$$

Finally we use the distributive rule for multiplication as well as some arithmetic to simplify the equation above:

$$F(X, Y) = \frac{1}{2}\left(x^3 y + xy^3\right) \tag{35}$$

Note: this result disagrees with the one that Orloff and Bloom give in the [8] for this question. The reult Orloff and Bloom give is:

$$F(X, Y) = \frac{x^3 y}{3} + \frac{xy^3}{3}. \tag{36}$$

However, it is consistent with the function Orloff and Bloom use for $F_X(x)$ in the fifth part of this problem in [8].

5.4 Marginal PDF

In order to find the marginal pdf $f_X(X)$ of $f(x,y)$ we integrate out the variable y [2].

$$f(x,y) = \int_0^1 \frac{3}{2}\left(x^2+y^2\right)\,dy. \tag{37}$$

We replace f with its antiderivative, with a note to integrate over the closed interval $[0,1]$:

$$f_X(X) = \frac{3}{2}\left(x^2 y + \frac{y^3}{3}\right)\Bigg|_{y=0}^{1}. \tag{38}$$

Now we replace y in the equation above with the limits of integration:

$$f_X(X) = \left(\frac{3}{2}\left(x^2(1) + \frac{(1)^3}{3})\right)\right) - \left(\frac{3}{2}\left(x^2(0) + \frac{(0)^3}{3})\right)\right) \tag{39}$$

And we simplify the equation above to give the marginal probability in a more agreeable form:

$$f_X(X) = \frac{3}{2}\left(x^2 + \frac{1}{3}\right) \tag{40}$$

Now that we have the marginal probability, we can answeer the second part of the question, which is to find the marginal probability $P(X<0.5)$.
$P(X<0.5)$ is the integral of $f_X(X)$ over the interval $[0,0.5]$ [2].

$$P(X<0.5) = \int_0^{0.5} \frac{3}{2}\left(x^2+\frac{1}{3}\right)\,dx. \tag{41}$$

We replace $f_X(X)$ with its antiderivative, evaluated over the interval $[0,0.5]$:

$$P(X<0.5) = \frac{3}{2}\left(\frac{x^3}{3}+\frac{x}{3}\right)\Bigg|_0^{0.5}. \tag{42}$$

And we evaluate the antiderivative at the limits of integration:

$$P(X<0.5) = \left(\frac{3}{2}\left(\frac{0.5^3}{3}+\frac{0.5}{3}\right)\right) - \left(\frac{3}{2}\left(\frac{0^3}{3}+\frac{0}{3}\right)\right) \tag{43}$$

The equation above simplfies to:

$$P(X<0.5) = \frac{3}{2}\left(\frac{1}{24}+\frac{1}{6}\right), \tag{44}$$

which further simplifies to:

$$P(X<0.5) = \frac{5}{16}. \tag{45}$$

5.5 Marginal cdf

In this section we will do as Orloff and Bloom ask, and find the marginal cumulative distribution function (cdf) $F_X(X)$, and $P(X < 0.5)$.
In [2] Orloff and Bloom state that, "If X and Y jointly take values on $[a, b] \times [c, d]$, then

$$F_X(X) = F(x, d), F_Y(Y) = F(b, y)." \tag{46}$$

We must compute $P(X < 0.5)$, so we use the marginal cdf $F(x, 1)$.
Previously in this problem, we found that the cdf for the joint distribution we are dealing with is

$$F(X, Y) = \frac{1}{2}\left(x^3 y + x y^3\right) \tag{47}$$

Therefore

$$F(x, 1) = \frac{1}{2}\left(x^3 + x\right) \tag{48}$$

We use the definition of joint cdf from [2] to write

$$P(X < 0.5) = F(0.5, 1). \tag{49}$$

Now, we can replace the right hand side of the equation above with the right hand side of equation 48, where we have applied the value 0.5 as the value for x in the right hand side of equation 48:

$$P(X < 0.5) = \frac{1}{2}\left(0.5^3 + 0.5\right). \tag{50}$$

We use arithmetic to simplify the right hand side of the equation above to find:

$$P(X < 0.5) = \frac{5}{16}. \tag{51}$$

5.6 Cdf of discrete joint distribution

In the last part of this question, Orloff and Bloom give a new, discrete joint distribution for us to investigate. They give us the following table, that defines the joint distribution:

X/Y	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

Note: the value in the $i, j^t h$ entry in the table above is the probability that $X = i$, and $Y = j$.
Orloff and Bloom ask us to find the joint cdf $F(3.5, 4)$.
$F(3.5, 4)$ is the sum of all probabilities where $X \leq 3.5$, and $Y \leq 4$.
Therefore $F(3.5, 4)$ is the sum of the probabilities in the shaded entries in the table below:

X/Y	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

There are 12 shaded entries in the table above, and they all have a value of probability $\frac{1}{36}$. Therefore the sum of the probabilities in all the shaded squares is $12 \times \frac{1}{36} = \frac{1}{3}$. Hence,

$$P(3.5, 4) = \frac{1}{3}. \tag{52}$$

We skip the next two questions on indpendence of discrete joint distributions because we have already answered similar questions previously.
The next question from [7] that we answer is one on independence of continuous random variables of different probability mass functions.
Before we delve into the particulars of this problem, it behooves us to note the definition of independence for continous jointly distributed random variables from [2]: $f(x, y) = f_X(x) f_Y(y)$.
Orloff and Bloom go on to state that jointly distributed random variables are independent if we, "…can factor the joint pdf or cdf as the product of a function of X and a function of y."
For this problem, we are given 3 joint pdf's, and we assume that the joint pdf's are defined over suitable regions such that the integrals over the regions equal 1. The three joint pdf's are:

1. $f(x, y) = 4x^2y^3$.
2. $f(x, y) = \frac{1}{2}\left(x^3y + xy^3\right)$.
3. $f(x, y) = 6e^{-3x-2y}$

We factor the first pdf. Let $f(x) = 4x^2$, $g(y) = 2y^3$. Then

$$4x^2y^3 = f(x) g(y). \tag{53}$$

.
We can factor the first pdf as a function of x, and a function of y, so x, and y are independently distributed with the first pdf.

Lemma 1. $a^{x+y} = a^x a^y$

$$a^{x+y} = \underbrace{aaa\ldots a}_{x+y\ a\text{'s}}. \tag{54}$$

We use the associative property of multiplication to state that the above is true if and only if:

$$a^{x+y} = \underbrace{aaa\ldots a}_{x\ a\text{'s}}\underbrace{aaa\ldots a}_{y\ a\text{'s}} \tag{55}$$

$$\underbrace{aaa\ldots a}_{x\ a\text{'s}}\underbrace{aaa\ldots a}_{y\ a\text{'s}} = a^x a^y. \tag{56}$$

Therefore

$$a^{x+y} = a^x a^y. \tag{57}$$

We factor the third pdf. Let $f(x) = e^{-3x}$, and let $g(y,) = e^{-2y}$.
The lemma above implies that $e^{-3x}e^{-2y} = e^{-3x-2y}$. Therefore we can factor $e^{-3x-2y} = f(x)g(y)$. Hence x and y are independently distributed in the distribution with the third pdf.
If random variables X and Y follow the joint distribution with the second pdf, then they cannot be indpendent. We factor the pdf:

$$\frac{1}{2}\left(x^3y + xy^3\right) = \frac{1}{2}(xy)\left(x^2+y^2\right) = \frac{1}{2}(xy)(x+y)(x-y). \tag{58}$$

We suppose any factorization of $f(x,y)$ would involve some product of the factors of $f(x,y)$ that we find above. However, each of these factors involves x, and y.
Supposition aside, in [2] Orloff and Bloom state that this function will not factor into a product of functions where one function only involves x, and the other only involves y.
Therefore we cannot factor $f(x,y)$ as the product of two functions $f(x)$, and $g(y)$.

6 Covariance of coin flips

Orloff and Bloom give us the following problem: We flip a fair coin 3 times. We let X be the number of heads in the first 2 flips, and we let Y be the number of heads in the last two flips.
The problem Orloff and Bloom give us is to calculate the covariance of X and Y.
The expected value $E(X)$, of X is:

$$E(X) = \frac{1}{4}\times 0 + \frac{1}{2}\times 1 + \frac{1}{4}times2 = 1. \tag{59}$$

Similarly the expected value $E(Y)$ of Y is also 1.
We use the technique from [3] and write X and Y as the sum of independent Bernoulli random variables X_1, X_2, and X_3, where X_1 is the outcome of the Bernoulli trial of flipping the coin for the first time, and so on.
We use the property of covariance from [3]

$$\operatorname{Cov}(X_1 + X_2, Y) = \operatorname{Cov}(X_1, Y) + \operatorname{Cov}(X_2, Y). \tag{60}$$

We substitute $Y = X_2 + X_3$

$$\operatorname{Cov}(X_1 + X_2, X_2 + X_3) = \operatorname{Cov}(X_1, X_2 + X_3) + \operatorname{Cov}(X_2, X_2 + X_3). \tag{61}$$

Now we use the identity:

$$\operatorname{Cov}(X, Y) = E(XY) - \mu_X \mu_Y. \tag{62}$$

We use the commutative property of multiplication:

$$\operatorname{Cov}(X, Y) = E(YX) - \mu_Y \mu_X. \tag{63}$$

Therefore

$$\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X). \tag{64}$$

We apply the identity above to equation 60 to get:

$$\operatorname{Cov}(X_1 + X_2, X_2 + X_3) = \operatorname{Cov}(X_2 + X_3, X_1) + \operatorname{Cov}(X_2 + X_3, X_2). \tag{65}$$

Now we can apply equation 60 again to the equation above:

$$\operatorname{Cov}(X_1 + X_2, X_2 + X_3) = \operatorname{Cov}(X_2, X_1) + \operatorname{Cov}(X_2, X_2) + \operatorname{Cov}(X_3, X_1) + \operatorname{Cov}(X_3, X_2) \tag{66}$$

The Bernoulli trials X_1, and X_3; and X_1 and X_2; and X_2, and X_3 are independent events. Therefore: $\operatorname{Cov}(X_2, X_1) = 0$, $\operatorname{Cov}(X_3, X_1) = 0$, and $\operatorname{Cov}(X_3, X_2) = 0$.
Therefore in order to compute $\operatorname{Cov}(X_1 + X_2, X_2 + X_3)$, We need only compute $\operatorname{Cov}(X_2, X_2)$.
In [3] Orloff and Bloom state that $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$.
X_2 is the toss of a fair coin, so it has expected value $0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}$, and variance $\frac{1}{2}\left(0 - \frac{1}{2}\right)^2 + \frac{1}{2}\left(1 - \frac{1}{2}\right)^2 = \frac{1}{4}$.
Therefore

$$\operatorname{Cov}(X_1 + X_2, X_2 + X_3) = \frac{1}{4}. \tag{67}$$

We defined $X = X_1 + X2$, and $Y = X_2 + X_3$, therefore

$$\operatorname{Cov}(X, Y) = \frac{1}{4}. \tag{68}$$

7 Correlation and Covariance of variables

7.1 Covariance

Orloff and Bloom ask us to consider flipping a fair coin $2n + 1$ times. They define the random variable X to be the number of heads in the first $n+1$ coim flips, and they define the random variable Y to be the number of heads in the last $n + 1$ flips. Orloff and Bloom then ask us to find the covariavariance $\text{Cov}(X, Y)$.
We use reasoning similar to the problem Orloff and Bloom gave us regarding 3 coin flips. We look at that problem as the case where $n = 3$.
Each coin flip is one independent Bernoulli trial with probability $\frac{1}{2}$.
We define a set of random variables first we define the abstract random variable X_i. We call X_i abstract because we have not given a particular value for i.

$$X_i = \begin{cases} 1 & \text{We flip the coin heads up on the } i^{th} \text{ toss.} \\ 0 & \text{We filp the coin tails up on the } i^{th} \text{ toss.} \end{cases} \tag{69}$$

Now we can define X and Y in terms of the X_i.

$$X = \sum_{i=1}^{n+1} X_i. \tag{70}$$

$$Y = \sum_{i=n}^{2n+1} X_i. \tag{71}$$

Therefore

$$\text{Cov}\,(X, Y) = \text{Cov}\left(\sum_{i=1}^{n+1} X_i, \sum_{j=n+1}^{2n} X_j \right). \tag{72}$$

We use equation 60 repeatedly to justify the statement that the previous equation is true if, and only if:

$$\text{Cov}\,(X, Y) = \sum_{i=1}^{n+1} \text{Cov}\left(X_i, \sum_{j=n+1}^{2n} X_j \right). \tag{73}$$

Now, since $\text{Cov}(X, Y) = \text{Cov}\,(Y, X)$,

$$\text{Cov}\,(X, Y) = \sum_{i=1}^{n+1} \text{Cov}\left(\sum_{j=n+1}^{2n} X_j, X_i \right). \tag{74}$$

And we apply equation 60 to the previous equation to obtain:

$$\operatorname{Cov}\left(X,Y\right)=\sum_{i=1}^{n+1}\sum_{j=n+1}^{2n}\operatorname{Cov}\left(X_j,X_i\right). \tag{75}$$

Each toss of a fair coin is an independent Bernoulli trial, so if $i\neq j$, then $\operatorname{Cov}(X_i,X_j)=0$ [3].
Therefore the only term in equation 75 that is not 0 is $\operatorname{Cov}(X_{n+1},X_{n+1})$.
We found in the previous problem involving three flips of a fair coin that the Covariance is $\frac{1}{4}$.
Therefore

$$\operatorname{Cov}\left(X,Y\right)=\frac{1}{4}. \tag{76}$$

7.2 Correlation

Orloff and Bloom define correlation in [3]. We denote correlation with the Greek letter ρ.

$$\rho=frac\operatorname{Cov}\left(X,Y\right)\sigma_X\sigma_Y. \tag{77}$$

X and X are both the number of heads in $n+1$ coin tosses, so their standard deviation will be the same.
Therefore we will determine σ_X, and we will also know what σ_Y is.
Let X_i have the same definition that it has in the previous section.
Then we use the definition of variance of a discrete random variable [9] to write:

$$\operatorname{Var}\left(X_i\right)=p\left(0\right)\left(0-\mu_{X_i}\right)^2+p\left(1\right)\left(1-\mu_{X_i}\right)^2. \tag{78}$$

X_i is equal to 1 if we toss a fair coin heads up, and 0 otherwis, so $\mu_{X_i}=\frac{1}{2}$, and $p\left(0\right)=p\left(1\right)=\frac{1}{2}$.
Therefore

$$\operatorname{Var}\left(X_i\right)=\frac{1}{2}\left(0-\frac{1}{2}\right)^2+\frac{1}{2}\left(1-\frac{1}{2}\right)^2. \tag{79}$$

We do the arithmetic on the right hand side of the previous equation to find:

$$\operatorname{Var}\left(X_i\right)=\frac{1}{4}. \tag{80}$$

Now we repeatedly use the first property of variance from [9] and the equation above to write:

$$\operatorname{Var}\left(\sum_{i=1}^{n+1}X_i\right)=\sum_{i=1}^{n+1}\operatorname{Var}\left(X_i\right). \tag{81}$$

Since $\mathrm{Var}(X_i) = \frac{1}{4}$,

$$\mathrm{Var}\left(\sum_{i=1}^{n+1} X_i\right) = \sum_{i=1}^{n+1} \frac{1}{4}. \tag{82}$$

Now we evaluate the sum above to find:

$$\mathrm{Var}\left(\sum_{i=1}^{n+1} X_i\right) = \frac{n+1}{4}. \tag{83}$$

Since X and Y are both the number of heads in $n+1$ tosses of a fair coin, $\mathrm{Var}(X) = \mathrm{Var}\,(Y) = \frac{n+1}{4}$, and $\sigma_X = \sigma_Y = \sqrt{\frac{n+1}{4}} = \frac{\sqrt{n+1}}{2}$.
Therefore

$$\rho = \frac{\frac{1}{4}}{\frac{\sqrt{n+1}}{2}\frac{\sqrt{n+1}}{2}} \tag{84}$$

The previous equation simplifies to :

$$\rho = \frac{1}{n+1}. \tag{85}$$

References

[1] Michael Dougherty. *Chapter 6 Basic Integration.* Available at `http://faculty.swosu.edu/michael.dougherty/book/chapter06.pdf` (2012/11/20).

[2] Jeremy Orloff and Jonathan Bloom. *Continuous Random Variables Class 5, 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_Reading5b.pdf` (Spring 2014).

[3] Jeremy Orloff and Jonathan Bloom. *Covariance and Correlation Class 7, 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_Reading7b.pdf` (Spring 2014).

[4] Jeremy Orloff and Jonathan Bloom. *Expectation, Variance and Standard Deviation for Continuous Random Variables Class 6, 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_Reading6b.pdf` (Spring 2014).

[5] Jeremy Orloff and Jonathan Bloom. *Expectation, Variance and Standard Deviation for Continuous Random Variables Class 6, 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_Reading6a.pdf (Spring 2014).

[6] Jeremy Orloff and Jonathan Bloom. *Gallery of Continuous Random Variables Class 5, 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_Reading5c.pdf (Spring 2014).

[7] Jeremy Orloff and Jonathan Bloom. *Joint Distributions, Independence Covariance and Correlation 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_class7slides.pdf (Spring 2014).

[8] Jeremy Orloff and Jonathan Bloom. *Joint Distributions, Independence Covariance and Correlation 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_class7_slides.pdf (Spring 2014).

[9] Jeremy Orloff and Jonathan Bloom. *Variance of Discrete Random Variables Class 5, 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_Reading5a.pdf (Spring 2014).