# MIT Introduction to Statistics 18.05 Class 7 Slides - Solutions

John Hancock

April 20, 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 [8].
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 [1], [9], [10], [3] to write the LaTeXsource code for this document.

## 2 Estimate Error

The first question Orloff and Bloom ask in [8] is about an accountant that rounds his cacluations (entries) to the nearest dollar. We assume the ac-

countant 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 [5] and techniques for estimating probability that Orloff and Bloom show in [5] 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 $i^th$ calculation. Orloff and Bloom tell us that $X_i$ is uniformly distributed on $[-0.5, 0.5]$.
We also need the mean $\mu$, and standard deviation $\sigma$ of $X_i$ in order to make our estimate.
In [7] Orloff and Bloom state that a uniformly distributed random variable on $[a, b]$ has the distribution function $f(x) = \frac{1}{a-b}$.
In [6] 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 [2] 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 [6], for a continuous random variable $X$:

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

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 [6] 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 [6] 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 $S$,

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

Becausee standard deviation is the square root of variance, the standard deviation of $S$, $\sigma_S$ 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 $1 - P\left(-5 \leq S \leq 5\right)$. We state this relationship with the equation:

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

The central limit theorem states that standardized $S$ approximately follows the normal distribution $N\left(0, 1\right)$.
We standardize $S$, and apply the central limit theorem like Orloff and Bloom do in [5] to get the approximation:

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

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 [2].
The rule of thumb [5] tells us that $P\left(-1 \leq Z \leq 1\right) \approx 0.68$. 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 [4] Orloff and Bloom state that a joint pdf must satisfy two properties:

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

*Proof.* First we show that for any $(x, y) \in [0,1] \times [0,1]$, $0 \leq f\left(x, y\right)$.
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:

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

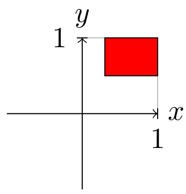
Therefore,

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

We have shown that $f\left(x, y\right)$ satisfies the two properties that Orloff and Bloom state a pdf must satisfy in [4]. Therefore $f\left(x, y\right)$ 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 [4], 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:

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

We use a calculator once more to find that $P\left(A\right) = 0.9745$. 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 [8]

### 5.3 Cumulative distribution function

The cumulative distribution function (cdf), $F\left(x, y\right)$ of the pdf for the joint distribution of $X$, and $Y$ the integral of its probability density function $f\left(x, y\right)$ that Orloff and Bloom give for this problem [4].
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 $y$.

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

Now we integrate with respect to $u$:

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

Next, we integrate with respect to v:

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

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

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

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

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

### 5.4 Marginal PDF

In order to

## References


[1] Latex.org user alainremillard. *Logical Not Symbol*. Available at `http://www.math.csusb.edu/notes/logic/lognot/node2.html` (2012/3/16).

[2] Michael Dougherty. *Chapter 6 Basic Integration*.

[3] tex.StackExchange.com user Keks Dose and Community. *LaTeX Error: Option clash for package xcolor (for table) [duplicate] (Answer)*. Available at `https://tex.stackexchange.com/questions/87197/latex-error-option-clash-for-package-xcolor-for-table` (2017/4/13).

[4] 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).

[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_Reading6b.pdf` (Spring 2014).

[6] 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).

[7] 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).

[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_class7slides.pdf` (Spring 2014).

[9] ShareLatex.com. *Theorems and proofs*. Available at `https://www.sharelatex.com/learn/Theorems_and_proofs` (2017).

[10] tex.StackExchange.com users Svend Tvsekaeg and David Carlisle. *Absolute Value Symbols (Answer)*. Available at `https://tex.stackexchange.com/questions/43008/absolute-value-symbols` (2015/5/30).