MIT Introduction to Statistics 18.05 Problem Set 2

John Hancock

March 9, 2017

Contents

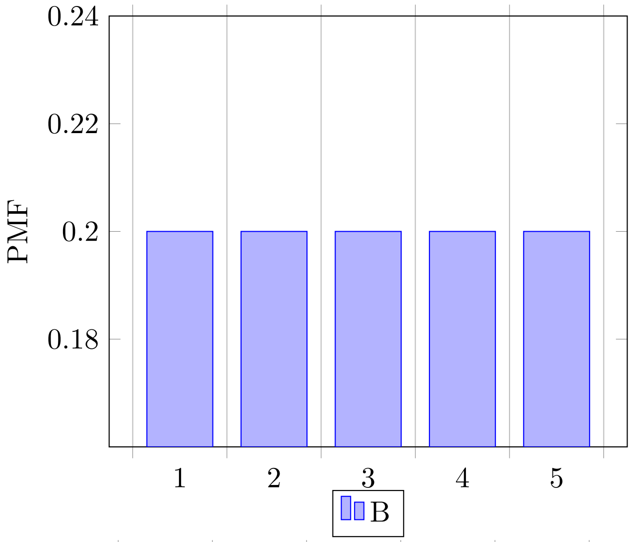
1 References and License

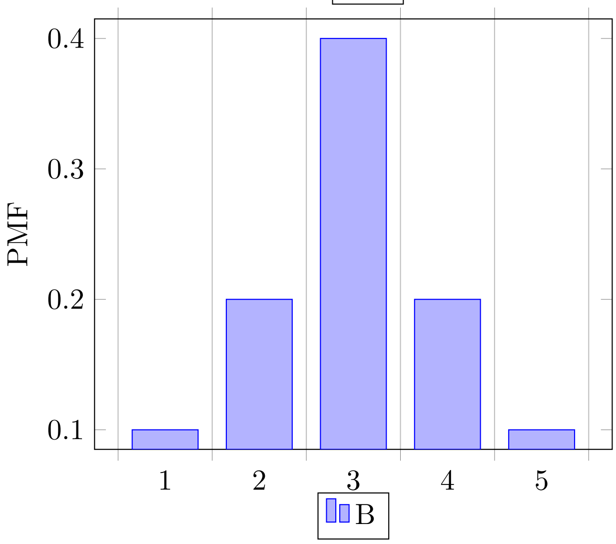
We are answering questions in the material from MIT OpenCourseWare course 18.05, Introduction to Probability and Statistics.
In this document we are answering questions Orloff and Bloom ask in [4].
Please see the references section for detailed citation information.

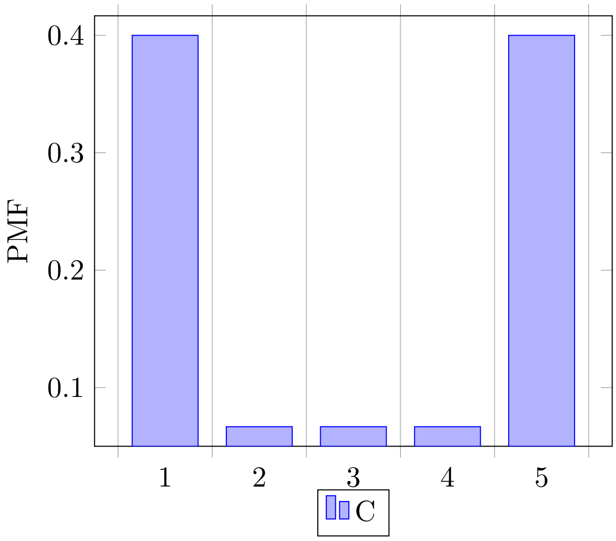
The material for the course is licensed under the terms at `http://ocw.mit.edu/terms`.
We use documentation in [8], [1], [7] to write LaTeXsource code for this document.

2 Order variables by size of standard deviation

In [4] Orloff and Bloom give us bar charts of three random variables and their probability mass functions:







In [5] Orloff and bloom state that the correct order of random variables by decreasing order of standard deviation is C, A, B.
We disagree with this answer.
The value of A is constant. Therefore the variance of A is zero. Hence, the standard deviation of A is also zero. Since zero is the minimum value of a standard deviation, Orloff and Bloom's answer must be incorrect.
We agree that the standard deviation of C is the largest, but B must have a positive standard deviation greater than zero. Therefore the order of these random variables by order of descending standard deviation is C, B, A.

3 Compute variance and Standard Deviation

In citeslides5 Orloff and Bloom ask us to compute the variance and standard deviation of the following random variable X.

values of X, x_i	1	2	3	4	5
PMF $p(x_i)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{4}{10}$	$\frac{2}{10}$	$\frac{1}{10}$

From the table and the definition of expected value, we compute

$$E(X) = \sum_{i=1}^{n} p(x_i)\, x_i \tag{1}$$

$$\sum_{i=1}^{n} p(x_i)\, x_i = \frac{1}{10}1 + \frac{2}{10}2 + \frac{4}{10}3 + \frac{2}{10}4 + \frac{1}{10}5 = \frac{27}{10} = 3.0 \tag{2}$$

In [6] Orloff and Bloom show that $\mathrm{Var}(X) = E\left(X^2\right) - E(X)^2$
Substituting the value of X^2 into the definition of expected value of a discrete random variable give us:

$$E\left(X^2\right) = \sum_{i=1}^{n} p\left(x_i\right) x_i^2 \tag{3}$$

And,

$$\sum_{i=1}^{n} p\left(x_i\right) x_i^2 = \frac{1}{10}1 + \frac{2}{10}4 + \frac{4}{10}9 + \frac{2}{10}16 + \frac{1}{10}25 = \frac{93}{10} = 10.2 \tag{4}$$

Therefore $\mathrm{Var}\left(X\right) = 10.2 - \left(3\right)^2 = 10.2 - 9 = 1.2$.
In [6] Orloff and Bloom give the definition of the standard deviation of a random variable $\sigma\left(X\right) = \sqrt{\mathrm{Var}\left(X\right)}$
Therefore the standard deviation of X is $\sqrt{3.05} \approx 1.746$

4 Variance of Bernoulli random variable

The next question Orloff and Bloom ask in the lecture 5 slides is for a proof that if $X \sim \mathrm{Bernoulli}\left(p\right)$, then $\mathrm{Var}\left(X\right) = p\left(1-p\right)$.
Orloff and Bloom prove this in [6].

5 Variance of a binomial random variable

Next, Orloff and Bloom ask for a proof that the variance of a random variable $X \sim \mathrm{binomial}\left(n, p\right) = np\left(1-p\right)$.
Orloff and Bloom also prove this in [6].

6 Variance of a sum

In this section Orloff and Bloom pose the question:
Suppose $X_1, X_2, \ldots, X_n$ are all independent random variables with $\sigma = 2$. Define a new random variable, $\bar{X}$ that is the average of $X_1, X_2, \ldots, X_n$.
They ask, "What is the standard deviation of $\bar{X}$?"
We know from [6] that, for two independent random variables X, and Y, $\mathrm{Var}\left(X+Y\right) = \mathrm{Var}\left(X\right) + \mathrm{Var}\left(Y\right)$
To extend this property to a sum of more than two independent random variables, we let $Y = Z + W$, where Z, and W are independent random variables.
Then $\mathrm{Var}\left(Z+W\right) = \mathrm{Var}\left(Z\right) + \mathrm{Var}\left(Z\right)$, and $\mathrm{Var}\left(X+Y\right) = \mathrm{Var}\left(X\right) + \mathrm{Var}\left(Z\right) + \mathrm{Var}\left(W\right)$.
We continue to rewrite the last term in the sum of variances until we have an expression on the right hand side of the sum that is the sum of variances of the independent random variables whose sum we wish to know the variance of.

$\bar{X}$ is the average of the random variables $X_1, X_2, \ldots, X_n$, so:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{5}$$

We apply the variance function to both sides of the equation above:

$$\mathrm{Var}\left(\bar{X}\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \tag{6}$$

In [6] Orloff and Bloom show that for constants a, b:

$$\mathrm{Var}\left(aX + b\right) = a^2 \mathrm{Var}\left(X\right) \tag{7}$$

Therefore

$$\mathrm{Var}\left(\bar{X}\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \tag{8}$$

Recall what we showed regarding extending the property of variance to the sum of multiple independent random variables. Because it is true, we can write

$$\mathrm{Var}\left(\bar{X}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) \tag{9}$$

Orloff and Bloom give us that $\sigma(X_i) = 2$, so

$$\mathrm{Var}\left(\bar{X}\right) = \frac{1}{n^2}\sum_{i=1}^{n} (4) \tag{10}$$

We evaluate the sum:

$$\mathrm{Var}\left(\bar{X}\right) = \frac{1}{n^2} 4n \tag{11}$$

We simplify the right hand side of the equation above:

$$\mathrm{Var}\left(\bar{X}\right) = \frac{4}{n} \tag{12}$$

Since the standard deviation is defined as the square root of the variance, we apply this definition to arrive at the answer to the question:

$$\sigma\left(\bar{X}\right) = \frac{2}{\sqrt{n}} \tag{13}$$

7 Continuous random variable

In this section we answer three questions on a continuous random variable X where Orloff and Bloom give us that X has a range $[0, 2]$, and the probability density function of X is $f(x) = cx^2$.

7.1 The value of c

In order to calculate the value of c, we use the property of probability density functions $f(x)$ Orloff and Bloom give in [3]:

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \tag{14}$$

In [3], Orloff and Bloom give us a note that if we know the range of the continuous random variable X, then in practice we do not integrate over $[-\infty, \infty]$, but over the range of X, instead.
Therefore, in the context of what Orloff and Bloom tell us about X, and $f(x)$ for this problem:

$$\int_{0}^{2} f(x)\, dx = 1 \tag{15}$$

Since $f(x) = cx^2$, we know:

$$\int_{0}^{2} cx^2 dx = 1 \tag{16}$$

The integral of a constant times a function is the constant times the integral of the function [2].
Therefore:

$$c \int_{0}^{2} x^2 dx = 1 \tag{17}$$

Evaluating the integral we obtain:

$$c \left(\frac{x^3}{3} \bigg|_0^2 \right) = 1 \tag{18}$$

We substitute the values for x:

$$c \left(\frac{2^3}{3} - \frac{0^3}{3} \right) = 1 \tag{19}$$

This implies

$$c = \frac{1}{\left(\frac{8}{3}\right)} = \frac{3}{8} \tag{20}$$

7.2 The CDF

Let the cumulative distribution function, CDF, of $f(x)$, be function $F(x)$. Orloff and Bloom give the definition of a cumulative distribution function in [3]

$$F(b) = P(X \leq b) = \int_{-infty}^{b} f(x)\,dx \tag{21}$$

Where $f(x)$ is the probability density function of a continuous random variable.
We apply this definition to $f(x)$, for the definition of $f(x)$ that Orloff and Bloom give us in this problem.
Hence, the CDF of $f(x)$ is:

$$\int_{-infty}^{b} cx^2 dx = \frac{3}{8}\frac{x^3}{3}\bigg|_{-infty}^{b} \tag{22}$$

Note that we solved for c in the previous section.
However, also in [3], Orloff and bloom point out that if a random variable is defined on an interval other than $[-infty, infty]$, then a robablity mass function for that random variable will be defined to have a value of 0 on those intervals where the random variable is not defined.
Therefore we write the CDF as

$$F(b) = \begin{cases} 0 & \text{if } b \leq 0 \\ \frac{3}{8}\frac{x^3}{3}\Big|_{0}^{b} & \text{if } 0 \leq b \leq 2 \\ 1 & \text{if } b \geq 2 \end{cases}$$

7.3 Probability of X over an interval

In this subsection, Orloff and Bloom as us to calculate $P(1 \leq X \leq 2)$
We apply the definition of a random variable from [3] to calcualte this probability. We can apply this definition because, for this problem, Orloff and Bloom give us that X is a continuous random variable.

$$P(1 \leq X \leq 2) = \int_{1}^{2} \frac{3}{8}x^2 dx \tag{23}$$

We replace the integral in the right hand side of the equation above with its anti-derivative:

$$P(1 \leq X \leq 2) = \frac{3}{8}\frac{x^3}{3}\bigg|_{1}^{2} \tag{24}$$

Now we evaluate the anti-derivative over the interval $[1, 2]$:

$$P(1 \leq X \leq 2) = \frac{3}{8}\left(\frac{x^2}{3} - \frac{x^1}{3}\right) = \frac{3}{8}\frac{7}{3} = \frac{7}{8} \tag{25}$$

8 Given a CDF

In this section Orloff and Bloom give us a CDF F with range $[0, b]$, and define F as $F(y) = \frac{y^2}{9}$.

8.1 The value of b

In this section we solve for b in the cumulative distribution function Orloff and Bloom give us.
Because F is a cumulative distribution function, and it has range $[0, b]$, we can write the following equation:

$$F(b) - F(0) = 1 \tag{26}$$

Another way to write this is that we know the integral of the probability density function that F is defined to be, must have the value of 1 over the range for which it is defined. This is from the definition of a continuous random variable, so we are implicitly assuming that F is the CDF of some continous random variable.
If we substitute the definition of F that Orloff and Bloom give us for this problem, we have an equation in the unknown, b:

$$\frac{b^2}{9} - \frac{0^2}{9} = 1 \tag{27}$$

This implies

$$\frac{b^2}{9} = 1 \tag{28}$$

Hence, $b = 3$.

8.2 Find the PDF

In this section we will find the PDF of F.
This is the derivative of F. We differentiate F with respect to y to find the answer.

$$\frac{d}{dy}\frac{y^2}{9} = \frac{2y}{9} \tag{29}$$

9 Probability of a point

In this section, Orloff and Bloom ask us about the probability of a continuous random variable X at a point.
The first way Orloff and Bloom ask this is, "What is $P\left(a \leq X \leq a\right)$?" If $X \leq a$, and $a \leq X$, then $X = a$. Also, X is a continuous random variable, so by definition [3] there is some probablity density function f such that

$$P\left(a \leq X \leq a\right) = \int_{a}^{a} f\left(x\right) dx \tag{30}$$

Let F be the anti-derivative of f. Then

$$P\left(a \leq X \leq a\right) = F\left(a\right) - F\left(a\right) = 0 \tag{31}$$

9.1 Probability of a specific point

In this question, Orloff and Bloom are stating a special case of the previous question where $a = 0$. By what we show in the previuos section, $P\left(X = 0\right) = 0$.

9.2 Random variables attain values

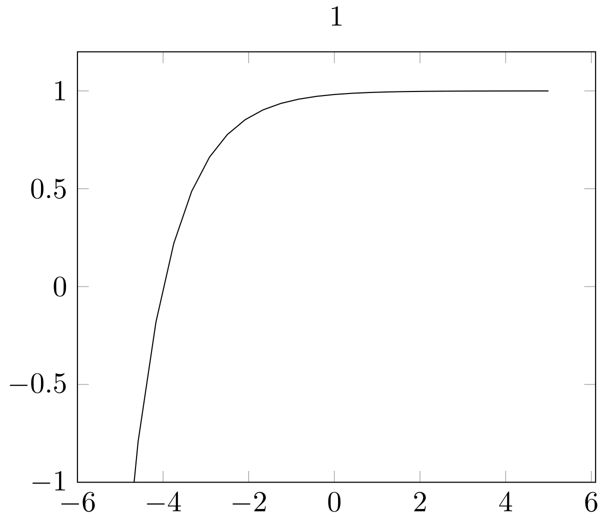
Orloff and Bloom ask, "Does $P\left(X = a\right) = 0$ mean X never equals a?"
The answer to this question is no. By the definition of a continuous random variable, there is a function f such that for any $c \leq d$

$$P\left(c \leq X \leq d\right) = \int_{c}^{d} f\left(x\right) dx \tag{32}$$

So, $P\left(X = a\right) = 0$ means that the area under the region of a curve that starts and ends at the same point is 0. If f is defined at a then it is possible that X can equal a.

10 Valid cumulative distribution functions

For this question, Orloff and Bloom present us with four plots of functions and ask us to identify the plots that are valid cumulative distribution functions.
Here are the plots:



References

[1] StackExchange.com user Niel de Beaudrap. *Math symbol question: Vertical bar for "evaluated at . . . "* Available at `http://tex.stackexchange.com/questions/40160/math-symbol-question-vertical-bar-for-evaluated-at` (accessed 2011/4/11).

[2] Paul Dawkins. *Proof of Various Integral Facts/Formulas/Properties.* Available at `https://www.sharelatex.com/learn/Pgfplots_package` (accessed 2017/3/8).

[3] 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).

[4] Jeremy Orloff and Jonathan Bloom. *Variance; Continuous Random Variables 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_class5slides.pdf` (Spring 2014).

[5] Jeremy Orloff and Jonathan Bloom. *Variance; Continuous Random Variables 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_class5_slides.pdf` (Spring 2014).

[6] 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).

[7] ShareLaTeX.com. *Pgfplots package.* Available at https://www.sharelatex.com/learn/Pgfplots_package (2017).

[8] ShareLaTeX.com. *Variance; Continuous Random Variables 18.05 Spring 2014 Jeremy Orloff and Jonathan Bloom.* Available at https://www.sharelatex.com/learn/Pgfplots_package (2017).