

MIT Introduction to Statistics 18.05 Reading 6A

Think Questions

John Hancock

April 6, 2017

Contents

1	References and License	1
2	Questions about X	2
2.1	Value of c	2
2.2	Mean, variance, and standard deviation of X	2
2.2.1	Mean of X	2
2.2.2	Variance of X	3
2.2.3	Standard Deviation of X	4
2.2.4	Median value of X	4
2.3	Standard Deviation of copies	4
2.4	PDF of function of X	5
3	Histograms	6
3.1	Equal bin widths of 0.5	6
3.1.1	Frequency Histogram	6
3.1.2	Density Histogram	6
3.2	Unequal bin widths	7
3.2.1	Frequency Histogram	7
3.2.2	Density Histogram	8
4	Election	8
4.1	Central Limit Theorem	8

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].

We use material in [2] to write the L^AT_EXcode for this document.

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>.

2 Questions about X

In this section we answer questions Orloff and Bloom ask in [5] regarding a random variable X .

Orloff and Bloom specify that X is defined on $[0, 1]$, and the pdf of X is cx^2 .

2.1 Value of c

Orloff and Bloom ask us to calculate the value of c . We will use rules and properties for integration from [1] in order to calculate the value for c .

We know

$$\int_0^1 cx^2 dx = 1. \quad (1)$$

Therefore

$$c \int_0^1 x^2 dx = 1. \quad (2)$$

The anti-derivative of x^2 is $\frac{x^3}{3} + C$, so we can replace the integral in the equation above with:

$$c \left(\frac{x^3}{3} \Big|_0^1 \right) = 1. \quad (3)$$

We then evaluate the anti-derivative over the interval $[0, 1]$ to obtain:

$$c \left(\frac{1^3}{3} \right) = 1. \quad (4)$$

This implies $c = 3$.

2.2 Mean, variance, and standard deviation of X

2.2.1 Mean of X

We use the definition of mean value that Orloff and Bloom give in [5]. The mean value of X is

$$\mu = \int_0^1 x (3x^2) dx. \quad (5)$$

We multiply the terms in the polynomial in the integral above to get:

$$\mu = \int_0^1 (3x^3) dx. \quad (6)$$

We replace the integral above with its anti-derivative:

$$\mu = \frac{3x^4}{4} \Big|_0^1. \quad (7)$$

We evaluate the anti-derivative over the closed interval $[0, 1]$ to find the value of the mean of X :

$$\mu = \frac{3}{5} - \frac{18}{16} + \frac{27}{48}. \quad (8)$$

2.2.2 Variance of X

We use the definition of the variance of a continuous random variable in [5] to compute the variance of X .

The definition of Variance Orloff and Bloom give in [5]:

$$\text{Var}(X) = E\left((X - \mu)^2\right). \quad (9)$$

We use the values for c and μ that we find above to find:

$$\text{Var}(X) = \int_0^1 x^2 3 \left(x - \frac{3}{4}\right)^2 dx. \quad (10)$$

Now we multiply some of the factors in the polynomial in the integral above to get:

$$\text{Var}(X) = \int_0^1 x^2 3 \left(x^2 - \frac{6x}{4} + \frac{9}{16}\right) dx. \quad (11)$$

We continue multiplying factors:

$$\text{Var}(X) = \int_0^1 3x^4 - \frac{18x^3}{4} + \frac{27x^2}{16} dx. \quad (12)$$

Now we replace the integral above with its anti-derivative:

$$\text{Var}(X) = \frac{3x^4}{5} - \frac{18x^4}{16} + \frac{27x^3}{48} \Big|_0^1. \quad (13)$$

And, we evaluate the anti-derivative over the interval $[0, 1]$:

$$\text{Var}(X) = \frac{3}{5} - \frac{18}{16} + \frac{27}{48} = \frac{3}{5} - \frac{18}{16} + \frac{9}{16}. \quad (14)$$

Now we simplify the expression above further:

$$\text{Var}(X) = \frac{3}{5} - \frac{18}{16} + \frac{9}{16} = \frac{3}{5} - \frac{9}{16} = \frac{48}{80} - \frac{45}{80} = \frac{3}{80}. \quad (15)$$

Therefore the variance of X is $\frac{3}{80}$.

2.2.3 Standard Deviation of X

The standard deviation of X is the square root of its variance [5]. Therefore the standard deviation of X is:

$$\sigma = \sqrt{\frac{3}{80}} \approx 0.194. \quad (16)$$

2.2.4 Median value of X

The median value of X is the 0.5 quantile of the cdf of X [5]. In the first part of this problem, we find that the pdf of X is $3x^2$. Therefore we must solve the equation:

$$\int_0^a 3x^2 dx = 0.5 \quad (17)$$

We replace the integral in the equation above with its anti-derivative:

$$\left. \frac{3x^3}{3} \right|_0^a = 0.5. \quad (18)$$

And, we evaluate the anti-derivative above over the interval $[0, a]$:

$$\frac{3a^2}{3} = 0.5. \quad (19)$$

Now it is a matter of doing some algebra to solve for a :

$$3a^2 = 0.5 \times 3. \quad (20)$$

This implies:

$$a^2 = 0.5. \quad (21)$$

Therefore, the median value of X is $\sqrt[3]{0.5} \approx 0.794$.

2.3 Standard Deviation of copies

For this part of the problem, Orloff and Bloom give us a set of random variables X_1, X_2, \dots, X_{16} that are independent, identically- distributed copies of X . They go on to define \bar{X} as the average of X_1, X_2, \dots, X_{16} .

Orloff and Bloom then ask us for the standard deviation, σ , of \bar{X} .

X_1, X_2, \dots, X_{16} are independent identically-distributed random variables. In [6] Orloff and Bloom show that, for two independent random variables X and Y , $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

\bar{X} is the average of X_1, X_2, \dots, X_{16} . Therefore

$$\bar{X} = \frac{X_1}{16} + \frac{X_2}{16} + \dots + \frac{X_{16}}{16} \quad (22)$$

We repeatedly apply the result on the sum of variances of random variables we quoted from [6] above to get

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{X_1}{16}\right) + \text{Var}\left(\frac{X_2}{16}\right) + \dots + \text{Var}\left(\frac{X_{16}}{16}\right). \quad (23)$$

In [5] Orloff and Bloom state that for a continuous random variable Z , $\text{Var}(aZ + b) = a^2\text{Var}(Z)$.

Therefore we may rewrite equation 23 as

$$\text{Var}(\bar{X}) = \frac{1}{16^2} (\text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_{16})). \quad (24)$$

For this problem, Orloff and Bloom give us that X_1, X_2, \dots, X_{16} are identical copies of X , so the variances of X_1, X_2, \dots, X_{16} are all equal to the variance of X .

We computed the variance of X above. $\text{Var}(X) = \frac{3}{80}$.

This implies that

$$\text{Var}(\bar{X}) = \frac{16}{16^2} \left(\frac{3}{80}\right). \quad (25)$$

Hence the standard deviation of \bar{X} is

$$\sigma_{\bar{X}} = \sqrt{\frac{3}{80 \times 16}} \approx 0.0484. \quad (26)$$

2.4 PDF of function of X

Orloff and Bloom ask us to let $Y = X^4$, and ask us to find the pdf of Y .

Let F_y be the cdf of Y . Then:

$$P(Y < y) = P(X^4 < y). \quad (27)$$

For $y \neq 0$, equation 27 is true if, and only if

$$P(Y < y) = P(X < \sqrt[4]{y}). \quad (28)$$

The cdf of X is x^3 . Furthermore, the cdf of X is defined on the interval $[0, 1]$ so equation 28 holds if, and only if:

$$P(Y < y) = x^3 \Big|_0^{\sqrt[4]{y}}. \quad (29)$$

We can evaluate equation 30 over the interval $[0, \sqrt[4]{y}]$; therefore, equation 30 is true if, and only if:

$$P(Y < y) = (\sqrt[4]{y})^3. \quad (30)$$

Equation 30 is true if, and only if, the cdf of Y is $(\sqrt[4]{y})^3$.

The pdf is the derivative of the cdf, therefore the pdf of Y is: $\frac{3}{4}y^{-\frac{1}{4}}$

3 Histograms

We use R-studio in order to draw the histograms for this list of numbers:

1, 1.2, 1.3, 1.6, 1.6, 2.1, 2.2, 2.6, 2.7, 3.1, 3.2, 3.4, 3.8, 3.9

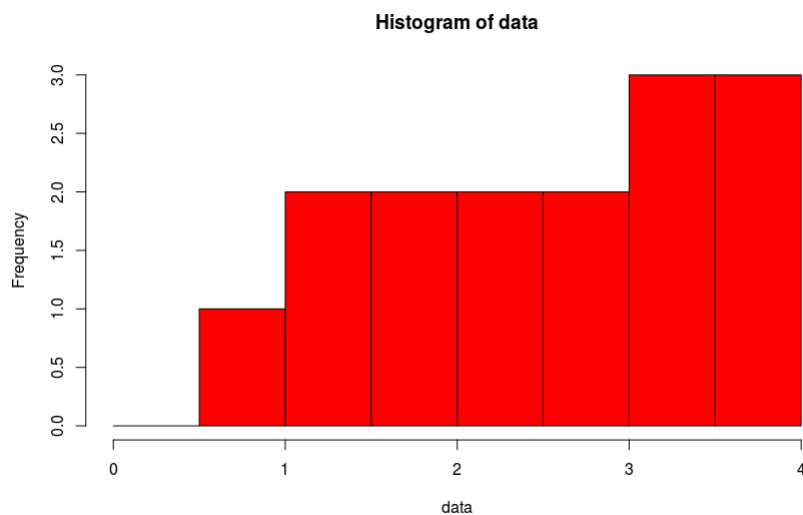
We rely on the examples in [3] in order to create the histograms below.

3.1 Equal bin widths of 0.5

3.1.1 Frequency Histogram

The R code we write to generate the image below is:

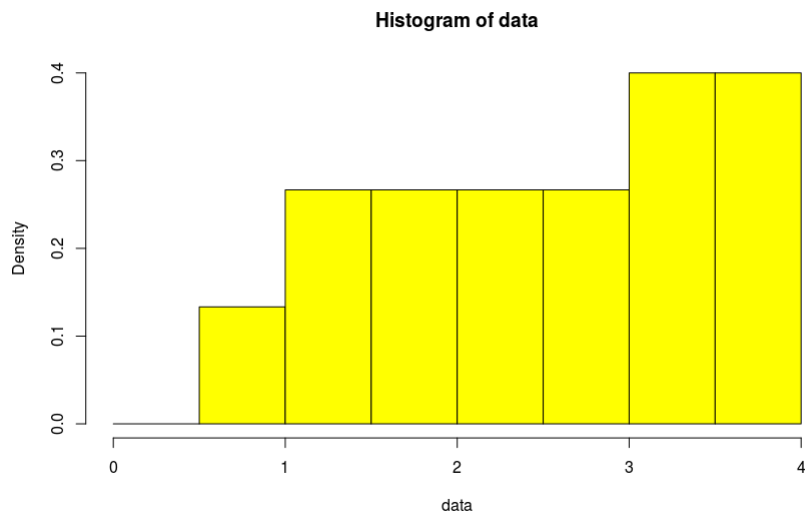
```
binwidth = .5;  
data = c(1, 1.2, 1.3, 1.6, 1.6, 2.1, 2.2, 2.6, 2.7, 3.1, 3.2, 3.4, 3.8, 3.9)  
bins = seq(0, 4, 0.5)  
hist(data, breaks=bins, col='red', freq=TRUE)
```



3.1.2 Density Histogram

The R code to generate the image below is:

```
binwidth = .5;  
data = c(1, 1.2, 1.3, 1.6, 1.6, 2.1, 2.2, 2.6, 2.7, 3.1, 3.2, 3.4, 3.8, 3.9)  
bins = seq(0, 4, 0.5)  
hist(data, breaks=bins, col='yellow', freq=FALSE)
```

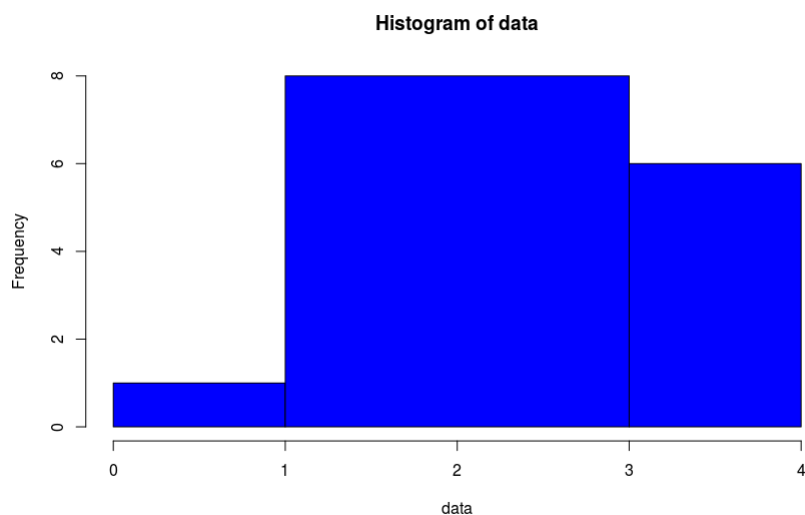


3.2 Unequal bin widths

3.2.1 Frequency Histogram

The R code we write to generate the image below is:

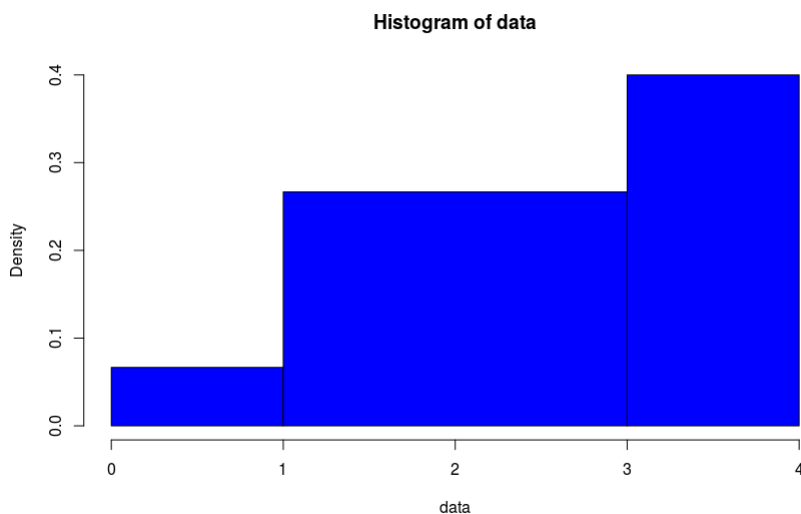
```
binwidth = .5;
data = c(1, 1.2, 1.3, 1.6, 1.6, 2.1, 2.2, 2.6, 2.7, 3.1, 3.2, 3.4, 3.8, 3.9)
bins = c(0,1,3,4)
hist(data, breaks=bins, col='blue', freq=TRUE)
```



3.2.2 Density Histogram

The R code we write to generate the image below is:

```
binwidth = .5;
data = c(1, 1.2, 1.3, 1.6, 1.6, 2.1, 2.2, 2.6, 2.7, 3.1, 3.2, 3.4, 3.8, 3.9)
bins = c(0,1,3,4)
hist(data, breaks=bins, col='blue', freq=FALSE)
```



4 Election

4.1 Central Limit Theorem

The first part of the board question that Orloff and Bloom ask us is a task to carefully write the central limit theorem.

The central limit theorem is:

Suppose $X_1, X_2, \dots, X_n, \dots$ are independent identically distributed random variables each having mean μ , and standard deviation σ . For each n let S_n denote the sum and let \bar{X}_n be the average of X_1, \dots, X_n .

$$S_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i \quad (31)$$

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{S_n}{n} \quad (32)$$

The properties of mean and variance show: $E(S_n) = n\mu$, $\text{Var}(S_n) = n\sigma^2$, $\sigma_{S_n} = \sqrt{n}\sigma$, $E(\bar{X}_n) = \mu$, $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$, $\sigma_{\bar{X}_n} = \frac{\sigma}{\sqrt{n}}$.

Since they are multiples of each other, S_n and \bar{X}_n have the same standardization

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (33)$$

Central Limit Theorem: for large n ,

$$X_n \approx N(\mu, \sigma^2 n), S_n \approx N(n\mu, n\sigma^2), Z_n \approx N(0, 1) \quad (34)$$

References

- [1] Michael Dougherty. *Chapter 6 Basic Integration*. Available at <http://faculty.swosu.edu/michael.dougherty/book/chapter06.pdf> (2012/11/20).
- [2] StackExchange.com user jamaicanworm. *Nice-looking p-th roots (in the question itself)*. Available at <http://faculty.swosu.edu/michael.dougherty/book/chapter06.pdf> (2012/03/22).
- [3] Jeremy Orloff and Jonathan Bloom. *Code for Studio 3*. Available at <https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/studio-resources/studio3.zip> (Spring 2014).
- [4] Jeremy Orloff and Jonathan Bloom. *Continuous Expectation and Variance, the Law of Large Numbers, and the Central Limit Theorem 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_class6slides.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. *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).