Math 224 **Spring 2018**
Homework 5 **Drew Armstrong**

---

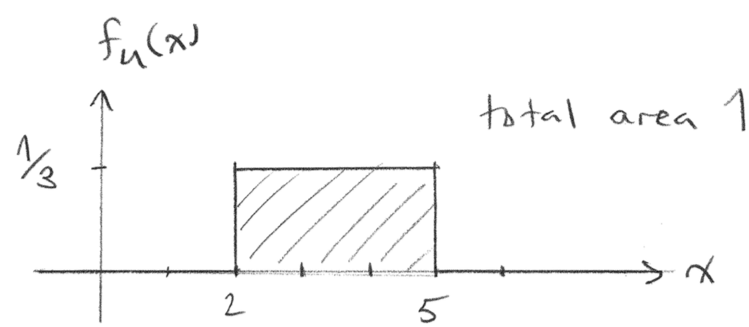
1. Let U be the uniform random variable on the interval $[2,5]$. Compute the following:

$$P(U=0), \quad P(U=3), \quad P(0<U<3), \quad P(3<U<4.5), \quad P(3\leq U\leq 4.5).$$

The pdf of U is defined as follows:

$$f_U(x) = \begin{cases} 1/3 & 2\leq x\leq 5, \\ 0 & \text{otherwise.} \end{cases}$$

Here is a picture:



Each desired probability is the area of a certain region under the curve. Fortunately each region is a rectangle (sometimes with width or height equal to zero) so we don't need to compute any integrals. First we have a couple of rectangles of zero width:

$$\begin{aligned} P(U=0) &= (\text{base})(\text{height}) = 0\cdot 0 = 0, \\ P(U=3) &= (\text{base})(\text{height}) = 0\cdot 1/3 = 0, \end{aligned}$$

In general we recall that $P(U=k)=0$ for any k. Next we have a region with two different heights:

$$\begin{aligned} P(0<U<3) &= P(0<U\leq 2) + P(2\leq U<3) \\ &= (\text{base})(\text{height}) + (\text{base})(\text{height}) \\ &= 1\cdot 0 + 1\cdot (1/3) = 1/3. \end{aligned}$$

Finally, we have

$$P(3<U<4.5) = (\text{base})(\text{height}) = 1.5\cdot(1/3) = 1/2$$

and

$$\begin{aligned} P(3\leq U\leq 4.5) &= P(U=3) + P(3<U<4.5) + P(U=4) \\ &= 0 + P(3<U<4.5) + 0 \\ &= 1/2. \end{aligned}$$

2. Let X be a continuous random variable with pdf defined as follows:

$$f_X(x) = \begin{cases} c \cdot x^2 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute the value of the constant c.
(b) Find the mean $\mu = E[X]$ and standard deviation $\sigma = \sqrt{\text{Var}(X)}$.
(c) Compute the probability $P(\mu - \sigma \le X \le \mu + \sigma)$.
(d) Draw the graph of f_X, showing the interval $\mu \pm \sigma$ in your picture.

(a) To find c we use the fact that the total area under a pdf equals 1. Thus we have

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f_X(x)\,dx \\ &= \int_0^1 c \cdot x^2\,dx \\ &= c \cdot \left.\frac{x^3}{3}\right|_0^1 = \frac{c}{3}, \end{aligned}$$

and it follows that $c = 3$.
(b) By definition, the first moment is

$$\begin{aligned} \mu = E[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx \\ &= \int_0^1 x \cdot 3x^2\,dx \\ &= 3 \cdot \left.\frac{x^4}{4}\right|_0^1 = \frac{3}{4}. \end{aligned}$$

Then the second moment is

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 \cdot f_X(x)\,dx \\ &= \int_0^1 x^2 \cdot 3x^2\,dx \\ &= 3 \cdot \left.\frac{x^5}{5}\right|_0^1 = \frac{3}{5}, \end{aligned}$$

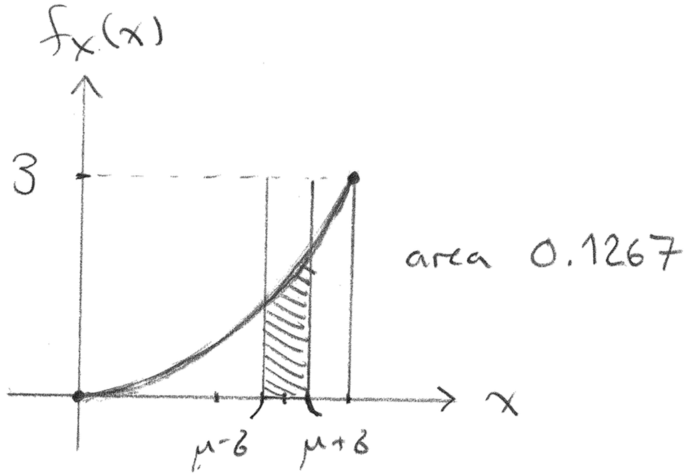
and hence

$$\begin{aligned} \sigma^2 = \text{Var}(X) &= E[X^2] - E[X]^2 = (3/5) - (3/4)^2 = 3/80, \\ \sigma &= \sqrt{3/80} = 0.1936. \end{aligned}$$

(c) We have

$$\begin{aligned} P(\mu - \sigma \le X \le \mu + \sigma) &= P(0.5564 \le X \le 0.9436) \\ &= \int_{0.5564}^{0.9436} 3x^2\,dx \\ &= 3 \cdot \left.\frac{x^3}{3}\right|_{0.5564}^{0.9436} = (0.9436)^3 - (0.5564)^3 = 66.80\%. \end{aligned}$$

(d) Here is a picture:



3. Let Z be a standard normal random variable, which is defined by the following pdf:

$$n(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}.$$

Let $\Phi(z)$ be the associated cdf (cumulative density function), which is defined by

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} n(x)\,dx.$$

Use the attached table to compute the following probabilities:

(a) $P(0 < Z < 0.5)$,
(b) $P(Z < -0.5)$,
(c) $P(Z > 1),\ P(Z > 2),\ P(Z > 3)$.
(d) $P(|Z| < 1),\ P(|Z| < 2),\ P(|Z| < 3)$,

(a)

$$P(0 < Z < 0.5) = \Phi(0.5) - \Phi(0) = 0.6915 - 0.5000 = 19.15\%.$$

(b)

$$P(Z < -0.5) = \Phi(-0.5) = 1 - \Phi(0.5) = 1 - 0.6915 = 30.85\%.$$

(c)

$$\begin{aligned} P(Z > 1) &= 1 - P(Z < 1) = 1 - \Phi(1) = 1 - 0.8413 = 15.87\%, \\ P(Z > 2) &= 1 - P(Z < 2) = 1 - \Phi(2) = 1 - 0.9772 = 2.28\%, \\ P(Z > 3) &= 1 - P(Z < 3) = 1 - \Phi(3) = 1 - 0.9987 = 0.13\%. \end{aligned}$$

(d) For any positive c we have

$$P(|Z| < c) = P(-c < Z < c) = \Phi(c) - \Phi(-c) = \Phi(c) - [1 - \Phi(c)] = 2 \cdot \Phi(c) - 1.$$

Thus we have

$$
\begin{aligned}
P(|Z| < 1) &= 2 \cdot \Phi(1) - 1 = 2 \cdot 0.8413 - 1 = 68.26\%, \\
P(|Z| < 2) &= 2 \cdot \Phi(2) - 1 = 2 \cdot 0.9772 - 1 = 95.44\%, \\
P(|Z| < 3) &= 2 \cdot \Phi(3) - 1 = 2 \cdot 0.9987 - 1 = 99.74\%.
\end{aligned}
$$

4. Continuing from Problem 3, use the attached table to find numbers $c, d \in \mathbb{R}$ solving the following equations:

(a) $P(Z > c) = P(|Z| > d) = 2.5\%$,
(b) $P(Z > c) = P(|Z| > d) = 5\%$,
(c) $P(Z > c) = P(|Z| > d) = 10\%$.

For general positive c and d we have

$$
\begin{aligned}
P(Z > c) &= \alpha \\
1 - P(Z < c) &= \alpha \\
1 - \Phi(c) &= \alpha \\
\Phi(c) &= 1 - \alpha
\end{aligned}
$$

and

$$
\begin{aligned}
P(|Z| > d) &= \alpha \\
P(Z < -d) + P(Z > d) &= \alpha \\
\Phi(-d) + 1 - \Phi(d) &= \alpha \\
[1 - \Phi(d)] + 1 - \Phi(d) &= \alpha \\
2 \cdot [1 - \Phi(d)] &= \alpha \\
\Phi(d) &= 1 - \frac{\alpha}{2}.
\end{aligned}
$$

(a) Using a reverse table look-up gives

$$P(Z > c) = 2.5\% \quad \Rightarrow \quad \Phi(c) = 97.5\% \quad \Rightarrow \quad c = 1.96$$

and

$$P(|Z| > d) = 2.5\% \quad \Rightarrow \quad \Phi(d) = 98.75\% \quad \Rightarrow \quad d = 2.24.$$

(b) Using a reverse table look-up gives

$$P(Z > c) = 5\% \quad \Rightarrow \quad \Phi(c) = 95\% \quad \Rightarrow \quad c = 1.65$$

and

$$P(|Z| > d) = 5\% \quad \Rightarrow \quad \Phi(d) = 97.5\% \quad \Rightarrow \quad d = 1.96.$$

(c) Using a reverse table look-up gives

$$P(Z > c) = 10\% \quad \Rightarrow \quad \Phi(c) = 90\% \quad \Rightarrow \quad c = 1.28$$

and

$$P(|Z| > d) = 10\% \quad \Rightarrow \quad \Phi(d) = 95\% \quad \Rightarrow \quad d = 1.65.$$

5. Let $X \sim N(\mu, \sigma^2)$ be a normal random variable with mean μ and variance σ^2. Let $\alpha, \beta \in \mathbb{R}$ be any constants such that $\alpha \neq 0$ and consider the random variable

$$Y = \alpha X + \beta.$$

(a) Show that $E[Y] = \alpha\mu + \beta$ and $\mathrm{Var}(Y) = \alpha^2\sigma^2$.
(b) Show that Y has a normal distribution $N(\alpha\mu + \beta, \alpha^2\sigma^2)$. In other words, show that for all real numbers $y_1 \leq y_2$ we have

$$P(y_1 \leq Y \leq y_2) = \int_{y_1}^{y_2} \frac{1}{\sqrt{2\pi\alpha^2\sigma^2}} \cdot e^{-[y-(\alpha\mu+\beta)]^2/2\alpha^2\sigma^2}\, dy.$$

[Hint: For all $x_1 \leq x_2$ you may assume that

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-(x-\mu)^2/2\sigma^2}\, dx.$$

Now use the substitution $y = \alpha x + \beta$.]

It follows from this problem that $Z = (X-\mu)/\sigma = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$ has a **standard** normal distribution. That is extremely useful.

(a) By general properties of E and Var we have

$$E[Y] = E[\alpha X + \beta] = \alpha E[X] + \beta = \alpha\mu + \beta$$

and

$$\mathrm{Var}(Y) = \mathrm{Var}(\alpha X + \beta) = \alpha^2 \mathrm{Var}(X) = \alpha^2\sigma^2.$$

(b) To show that Y is normal we want to show for all real numbers $y_1 \leq y_2$ that

$$(?) \qquad P(y_1 \leq Y \leq y_2) = \int_{y=y_1}^{y=y_2} \frac{1}{\sqrt{2\pi\alpha^2\sigma^2}} \cdot e^{-(y-\alpha\mu-\beta)^2/2\alpha^2\sigma^2}\, dy.$$

To see this, we will use use the fact that X is normal to obtain[1]

$$\begin{aligned}
P(y_1 \leq Y \leq y_2) &= P(y_1 \leq \alpha X + \beta \leq y_2) \\
&= P(y_1 - \beta \leq \alpha X \leq y_2 - \beta) \\
&= P\left(\frac{y_1-\beta}{\alpha} \leq X \leq \frac{y_2-\beta}{\alpha}\right) \\
(*) \qquad &= \int_{x=(y_1-\beta)/\alpha}^{x=(y_2-\beta)/\alpha} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-(x-\mu)^2/2\sigma^2}\, dx.
\end{aligned}$$

Then to show that the expressions $(*)$ and $(?)$ are equal we will make the substitution

$$\begin{aligned}
y &= \alpha x + \beta, \\
x &= (y-\beta)/\alpha, \\
dy &= \alpha \cdot dx.
\end{aligned}$$

[1]In the third line here we will assume that $\alpha > 0$. The proof for $\alpha < 0$ is exactly the same except that it will switch the limits of integration.

Finally, we observe that

$$
\begin{aligned}
\int_{y=y_1}^{y=y_2} \frac{1}{\sqrt{2\pi\alpha^2\sigma^2}} \cdot e^{-(y-\alpha\mu-\beta)^2/2a^2\sigma^2}\, dy &= \int_{x=(y_1-\beta)/\alpha}^{x=(y_2-\beta)/\alpha} \frac{1}{\sqrt{2\pi\alpha^2\sigma^2}} \cdot e^{-(\alpha x+\cancel{\beta}-\alpha\mu-\cancel{\beta})^2/2\alpha^2\sigma^2}\, \alpha \cdot dx \\
&= \int_{x=(y_1-\beta)/\alpha}^{x=(y_2-\beta)/\alpha} \frac{\cancel{\alpha}}{\sqrt{2\pi\cancel{\alpha^2}\sigma^2}} \cdot e^{-\cancel{\alpha^2}(x-\mu)^2/2\cancel{\alpha^2}\sigma^2}\, dx \\
&= \int_{x=(y_1-\beta)/\alpha}^{x=(y_2-\beta)/\alpha} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-(x-\mu)^2/2\sigma^2}\, dx
\end{aligned}
$$

as desired.

6. The average weight of a bag of chips from a certain factory is 150 grams. Assume that the weight is normally distributed with a standard deviation of 12 grams.

(a) What is the probability that a given bag of chips has weight greater than 160 grams?
(b) Collect a random sample of 10 bags of chips and let Y be the number that have weight greater than 160 grams. Compute the probability $P(Y \leq 2)$.

(a) Let X be the weight of a random bag of chips. We have assumed that $X \sim N(\mu = 150, \sigma^2 = 144)$. To compute the probability $P(X > 160)$ we first standardize and then look up the answer in a table of z-scores:

$$
\begin{aligned}
P(X > 160) &= P(X - 150 > 10) \\
&= P\left(\frac{X-150}{12} > 0.83\right) \\
&= 1 - P\left(\frac{X-150}{12} \leq 0.83\right) \\
&= 1 - \Phi(0.83) = 1 - 0.7967 = 20.33\%.
\end{aligned}
$$

(b) Now suppose that 10 bags are selected at random and let Y be the number with weight greater than 160 grams. We can think of each bag of chips as a coin flip and from part (a) we know that $P(H) = 20.33\%$. Thus for any k we have

$$P(Y = k) = \binom{10}{k}(0.2033)^k(0.7967)^{10-k}.$$

My computer tells me that

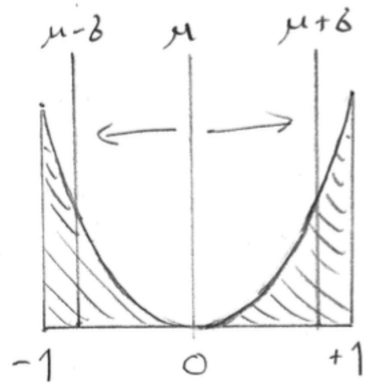
$$P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2) = 66.78\%.$$

7. Let $X_1, X_2, \ldots, X_{15}$ be independent and identically distributed (iid) random variables. Suppose that each X_i has pdf defined by the following function:

$$f(x) = \begin{cases} \frac{3}{2} \cdot x^2 & \text{if } -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute $E[X_i]$ and $\mathrm{Var}(X_i)$.
(b) Consider the sum $Y = X_1 + X_2 + \cdots + X_{15}$. Use part (a) to compute $E[Y]$ and $\mathrm{Var}(Y)$.
(c) The Central Limit Theorem says that Y is approximately normal. Use this fact to estimate the probability $P(-0.3 \leq Y \leq 0.5)$.

(a) Here is a graph of the pdf of each individual X_i:



Since the distribution is symmetric about zero, we conclude without doing any work that $\mu = E[X_i] = 0$ for each i. To find σ, however, we need to compute an integral. For any i, the variance of X_i is given by

$$\begin{aligned}
\sigma^2 = \mathrm{Var}(X_i) &= E\left[X_i^2\right] - E[X_i]^2 \\
&= E[X_i^2] - 0 \\
&= \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx - 0 \\
&= \int_{-1}^{1} x^2 \cdot \frac{3}{2}x^2\, dx \\
&= \frac{3}{2}\int_{-1}^{1} x^4\, dx \\
&= \frac{3}{2} \cdot \frac{x^5}{5}\bigg|_{-1}^{1} = \frac{3}{2}\cdot\frac{1}{5} - \frac{3}{2}\cdot\frac{(-1)^5}{5} = \frac{6}{10} = \frac{3}{5}.
\end{aligned}$$

(b) It follows that Y has mean and variance given by

$$\mu_Y = E[Y] = E[X_1] + E[X_2] + \cdots + E[X_{15}] = 0 + 0 + \cdots + 0 = 0$$

and

$$\sigma_Y^2 = \mathrm{Var}(Y) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_{15}) = \frac{3}{5} + \frac{3}{5} + \cdots + \frac{3}{5} = 15 \cdot \frac{3}{5} = 9,$$

By the Central Limit Theorem, the sum Y is approximately normal and hence $(Y - \mu_Y)/\sigma_Y = Y/3$ is approximately standard normal. We conclude that

$$\begin{aligned}
P(-0.3 \le Y \le 0.5) &= P\left(\frac{-0.3}{3} \le \frac{Y}{3} \le \frac{0.5}{3}\right) \\
&= P\left(-0.1 \le \frac{Y}{3} \le 0.17\right) \\
&\approx \Phi(0.17) - \Phi(-0.1) \\
&= \Phi(0.17) - [1 - \Phi(0.1)] \\
&= \Phi(0.17) + \Phi(0.1) - 1 \\
&= 0.5675 + 0.5398 - 1 = 10.73\%.
\end{aligned}$$

8. Suppose that $n = 48$ seeds are planted and suppose that each seed has a probability $p = 75\%$ of germinating. Let X be the number of seeds that germinate and use the Central

Limit Theorem to estimate the probability $P(35 \le X \le 40)$ that between 35 and 40 seeds germinate. Don't forget to use a continuity correction.

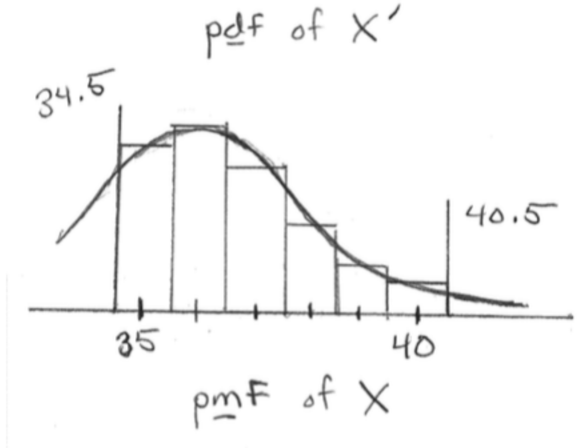
We observe that X is a binomial random variable with the following pmf:

$$P(X = k) = \binom{48}{k}(0.75)^k(0.25)^{48-k}.$$

My laptop tells me that the exact probability is

$$P(35 \le X \le 40) = \sum_{k=35}^{40} P(X = k) = \sum_{k=35}^{40} \binom{48}{k}(0.75)^k(0.25)^{48-k} = 63.74\%.$$

To compute an approximation by hand we will use the de Moivre-Laplace Theorem, which says that X is approximately normal with mean $np = 36$ and variance $\sigma^2 = np(1-p) = 9$, i.e., standard deviation $\sigma = 3$. Let X' be a **continuous** random variable with $X' \sim N(36, 3^2)$. Here is a picture comparing the probability **mass** function of the discrete variable X to the probability **density** function of the continuous variable X':



The picture suggests that we should use the following continuity correction:[2]

$$P(35 \le X \le 40) \approx P(34.5 \le X' \le 40.5).$$

And then because $(X' - 36)/3$ is **standard** normal we obtain

$$\begin{aligned}
P(34.5 \le X' \le 40.5) &= P(-1.5 \le X' - 36 \le 4.5) \\
&= P\left(-0.5 \le \frac{X' - 36}{3} \le 1.5\right) \\
&= \Phi(1.5) - \Phi(-0.5) \\
&= \Phi(1.5) - [1 - \Phi(0.5)] \\
&= \Phi(1.5) + \Phi(0.5) - 1 = 0.9332 + 0.6915 - 1 = 62.47\%
\end{aligned}$$

Not too bad.

9. Suppose that a six-sided die is rolled 24 times and let X_i be the number that shows up on the i-th roll. Let $\overline{X} = (X_1 + X_2 + \cdots + X_{24})/24$ be the average of the numbers that show up.

(a) Assuming that the die is fair, compute the expected value and variance:

$$E\left[\overline{X}\right] \qquad \text{and} \qquad \mathrm{Var}\left(\overline{X}\right).$$

---

[2]If you don't do this then you will still get a reasonable answer, it just won't be as accurate.

(b) Assuming that the die is fair, use the Central Limit Theorem to estimate the probability $P\left(\overline{X} \geq 4\right)$.

(c) Suppose you roll an unknown six-sided die 24 times and get an average value of 4.

Is the die fair?

In other words: Let H_0 be the hypothesis that the die is fair. Should you reject this hypothesis at the 5% level of significance?

(a) Let X_i be the number that shows up on the i-th roll. Then each X_i is identically distributed with the following pmf:

k	1	2	3	4	5	6
$P(X_i = k)$	1/6	1/6	1/6	1/6	1/6	1/6

We compute from this table that

$$\begin{aligned}
E[X_i] &= (1+2+3+4+5+6)/6 = 7/2, \\
E[X_i^2] &= (1^2+2^2+3^2+4^2+5^2+6^2)/6 = 91/6, \\
\mathrm{Var}(X_i) &= E[X_i^2] - E[X_i]^2 = 91/6 - (7/2)^2 = 35/12.
\end{aligned}$$

Now it follows that

$$\begin{aligned}
E[\overline{X}] &= \frac{1}{24} \cdot (E[X_1] + \cdots E[X_{24}]) \\
&= \frac{1}{24} \cdot (3.5 + \cdots + 3.5) = \frac{1}{24} \cdot 24 \cdot 3.5 = 3.5
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}(\overline{X}) &= \frac{1}{24^2} \cdot (\mathrm{Var}(X_1) + \cdots \mathrm{Var}(X_{24})) \\
&= \frac{1}{24} \cdot \left(\frac{35}{12} + \cdots + \frac{35}{12}\right) = \frac{1}{24^2} \cdot 24 \cdot \frac{35}{12} = \frac{1}{24} \cdot \frac{35}{12},
\end{aligned}$$

and hence $\sigma = \sqrt{\frac{1}{24} \cdot \frac{35}{12}} = 0.3486$.

(b) The Central Limit Theorem tells us that $\overline{X}$ is approximately normal with mean 3.5 and standard deviation 0.3486. To compute the probability $P(\overline{X} > 4)$ we standardize then look up the answer in a table of z-scores:

$$\begin{aligned}
P(\overline{X} > 4) &= P(\overline{X} - 3.5 > 0.5) \\
&= P\left(\frac{\overline{X} - 3.5}{0.3486} > 1.43\right) \\
&= 1 - \Phi(1.43) \\
&= 1 - 0.9236 \\
&= 7.64\%.
\end{aligned}$$

(c) How surprising is this? In order to determine if the die is fair suppose we roll the die 24 times and let $\overline{X}$ be the average of the 24 numbers that show up. Let $\mu = E[\overline{X}]$ and

$\sigma^2 = \mathrm{Var}(\overline{X})$ so that $\overline{X}$ is approximately $N(\mu, \sigma^2)$. If the die is fair then we saw in parts (a) and (b) that $\mu = 3.5$ and $\sigma = 0.3486$. We will test the null hypothesis

$$H_0 = \text{“}\mu = 3.5\text{”}$$

against the two-sided alternative hypothesis

$$H_1 = \text{“}\mu \neq 3.5.\text{”}$$

At the 5% level of significance, the critical region for this test will be $|\overline{X} - 3.5| > c$ for some number such that

$$P(|\overline{X} - 3.5| > c) = 5\%.$$

Assuming that H_0 is true we know that $(\overline{X} - 3.5)/0.3486$ is approximately standard normal so we can solve for c by standardizing and then looking up in a table. We have

$$P\left(\left|\frac{\overline{X} - 3.5}{0.3486}\right| > \frac{c}{0.3486}\right) = 5\%$$

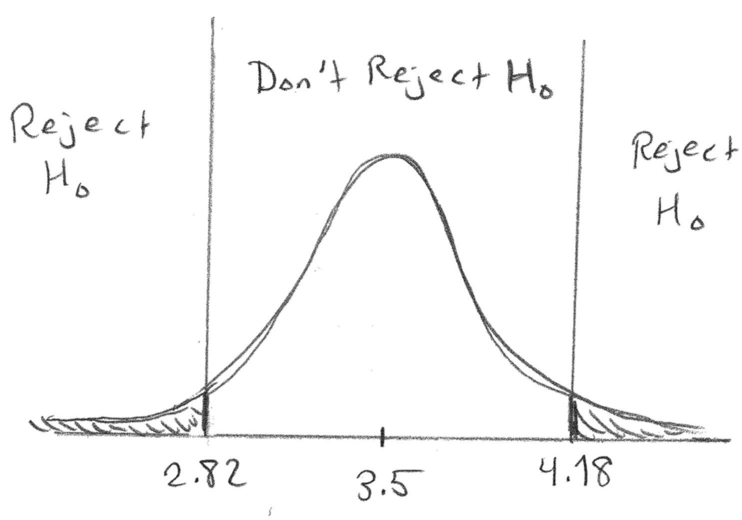
and then from Exercise 4(b) we know that

$$\frac{c}{0.3486} = 1.96 \quad \Rightarrow \quad c = 0.6834.$$

We will reject the null hypothesis when $|\overline{X} - 3.5| > 0.6834$, or

$$\begin{array}{ccccc} -0.6834 & < & \overline{X} - 3.5 & < & 0.6834 \\ 3.5 - 0.6834 & < & \overline{X} & < & 3.5 + 0.6834 \\ 2.82 & < & \overline{X} & < & 4.18. \end{array}$$

Here is a picture:



Finally, suppose we perform the experiment and get $\overline{X} = 4$. Since this is not in the critical region for the test, **we do not reject H_0.** In other words,

The die might be fair.