This tutorial comes with a spreadsheet which you can use as a guide to fill out the results of doing each part of the tasks listed.

Task 1. Retrieve your submissions from Homework 5 in Week 9 and compare solutions around the group. For Question 1(b)(ii), how did you calculate the "more than 5" probability? For Question 2(e), how does the standard deviation for A, the mean distance driven over three days, compare to the standard deviation for K_1, the distance on the first day? What about the standard devation of the mean distance driven over five days, ten days, or 100 days?

Task 2. (This is Question 117 from the textbook [Carlton & Devore 2017].)
If the side of a square X is random with the pdf $f_X(x) = x/8$, $0 < x < 4$, and Y is the area of the square, find the pdf of Y.

Task 3. A farm grows apples for sale to distributors. The price of apples depends on their weight and distributors will only take apples of a certain weight range. This year's apples from the farm have weights normally distributed with mean 170g and standard deviation 35g. They are sold to distributors in batches of 500.
Distributor A is looking for apples between 120g and 240g in weight: they check for this by taking a random sample of 10 apples from a batch and will accept the batch if at least 8 are in the target weight range.
Distributor B has a different policy. They are interested in any apples weighing over 100g: to test for this they sample 5 and reject the whole batch if even one of those is underweight.
Which distributor is more likely to accept a batch of apples from this farm?

Task 4. The Normal distribution can be used as an approximation for calculating probabilities from the Binomial distribution. This task explores how well approximations like this work in a simple case.
If $X \sim \text{Binom}(n, p)$ is a discrete random variable counting successes in n trials each with probability p then we have the following approximation.

$$P(X \leq x) \;=\; B(x; n, p) \;\approx\; \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right) \qquad x \in \{0, 1, \ldots, n\}$$

This is considered a good approximation in practice if $np \geq 10$ and $nq \geq 10$ where $q = (1 - p)$. Here you will look at a situation where both values are much less than 10.

(a) Suppose discrete random variable $X \sim \text{Binom}(4, 0.5)$ counts the number of heads thrown in four tosses of a fair coin: $X \in \{0, 1, 2, 3, 4\}$. What are the mean and standard deviation of the random variable X?

(b) Use the Normal approximation above to estimate the probability that X takes the value 2.

(c) Complete this to draw up a table of estimates for the probability that X takes each of the possible values 0 to 4.

(d) Now compute a table of the PMF for X showing the exact probability that it takes each of these five possible values.

(e) Calculate the table of errors: what percentage variation is there in each estimate compared to the exact value? Which values have the largest error?

Solution Notes

Task 1. See the solution guide for Homework 5. For the last part, as the number of days being considered goes up the expected value of the mean stays the same but the standard deviation becomes smaller and smaller.

Task 2. Random variable $Y = X^2$ and this transformation has inverse $h(y) = \sqrt{y}$ over the range taken by X. This inverse function is differentiable with $h'(y) = 1/(2\sqrt{y})$. This satisfies the requirements of the *Transformation Theorem* from §3.7 of the textbook giving PDF as follows.

$$f_Y(y) \;=\; f_x(h(y)) \cdot |h'(y)| \;=\; \frac{\sqrt{y}}{8} \cdot \left|\frac{1}{2\sqrt{y}}\right| \;=\; \frac{1}{16} \qquad 0 < y < 16$$

Thus we have a uniform distribution $Y \sim \mathrm{Unif}(0, 16)$.

$$f_Y(y) \;=\; \begin{cases} \frac{1}{16} & 0 < y < 16 \\ 0 & \text{otherwise} \end{cases}$$

Task 3. We take normal random variable $W \sim \mathrm{N}(170, 35)$ to represent the weight of an apple from this year's crop.

$$\begin{aligned} P(W \leq 100) \;&=\; \Phi\left(\frac{100-170}{35}\right) \;=\; \Phi(-2.00) \;=\; 0.0228 \\ P(W \leq 120) \;&=\; \Phi\left(\frac{120-170}{35}\right) \;=\; \Phi(-1.43) \;=\; 0.0764 \\ P(W \leq 240) \;&=\; \Phi\left(\frac{240-170}{35}\right) \;=\; \Phi(+2.00) \;=\; 0.9772 \end{aligned}$$

$$\begin{aligned} P(\text{Apple acceptable by A}) \;&=\; P(120 \leq W \leq 240) \;=\; P(W \leq 240) - P(W \leq 120) \;=\; 0.9008 \\ P(\text{Apple acceptable by B}) \;&=\; P(100 \leq W) \;=\; 1 - P(W \leq 100) \;=\; 0.9772 \end{aligned}$$

These calculations are done using the tables at the back of the textbook, pages 601 and 602. More precise calculations with machine assistance may differ in the last decimal place.

The experiment of sampling 5 or 10 apples introduces a binomial distribution. The size of the batch turns out to be irrelevant; it's also not relevant whether or not each apple is replaced after being tested. Page 99 of the textbook shows how to calculate binomial probabilities.

$$\begin{aligned} P(\text{Batch accepted by A}) \;&=\; B(2; 10, (1-0.9008)) \\ =\; \binom{10}{2} 0.9008^8 (1-0.9008)^2 &+ \binom{10}{1} 0.9008^9 (1-0.9008) + \binom{10}{0} 0.9008^{10} \;=\; 0.931 \\ P(\text{Batch accepted by B}) \;&=\; B(0; 5, (1-0.9772)) \;=\; \binom{5}{0} 0.9772^5 \;=\; 0.891 \end{aligned}$$

Distributor A is more likely to accept any given batch of apples (even though they are more likely to reject any individual apple). Again, variation in the last decimal place may happen with different methods but makes no difference to the final answer.

Task 4.

(a) $\mathrm{E}(X) = 2$ and $\mathrm{SD}(X) = 1$. Mean and standard deviation of a binomial random variable can be calculated using the Proposition from Section 2.4.3 of the textbook, page 101.

(b) Tables of values for Φ are in the back of the textbook.

$$\begin{aligned} P(X=2) \;=\; P(X \leq 2) - P(X \leq 1) \;&\approx\; \Phi\left(\frac{2+0.5-2}{1}\right) - \Phi\left(\frac{1+0.5-2}{1}\right) \\ &=\; \Phi(0.5) - \Phi(-0.5) \;=\; 0.6915 - 0.3085 \;=\; 0.3830 \end{aligned}$$

(c) This repeats the calculations for each possible value of x.

x	0	1	2	3	4
Approximate $P(X=x)$	0.0606	0.2417	0.3830	0.2417	0.6060

Note that these do not add up to 1. The tails of the normal distribution below -2.5 and above 2.5 don't contribute to any of these estimates.

(d) Section 2.4.1 in the textbook shows how to calculate binomial probabilities like this, in particular the theorem at the top of page 99.

x	0	1	2	3	4
Exact $P(X=x)$	0.0625	0.2500	0.3750	0.2500	0.0625

(e)

x	0	1	2	3	4
Error in estimate	-3.04%	-3.32%	2.13%	-3.32%	-3.04%

The largest error is in the shoulder values of 1 and 3. Even with this small value of n all errors are below 5%.