APMA 1650 – Homework 4

Due Monday, July 18, 2016

Homework is due during class or by 3:45 pm in the homework drop box in 182 George St.
Show all of your work used in deriving your solutions.

1. Let X be a continuous random variable which has a probability density $f(x)$ given by

$$f(x) = \begin{cases} cx^3 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of c which makes this a valid probability density function.

We integrate the density from 0 to 1 and set the integral equal to 1:

$$\begin{aligned} 1 &= \int_0^1 cx^3 dx \\ &= c\frac{x^4}{4}\Big|_0^1 \\ &= \frac{c}{4} \end{aligned}$$

From this we see that $c = 4$.

(b) What is $\mathbb{P}(0 \le X \le \frac{1}{4})$?

To find this probability, we integrate the density from 0 to 1/4.

$$\begin{aligned} \mathbb{P}(0 \le X \le \frac{1}{4}) &= \int_0^{1/4} 4x^3 dx \\ &= x^4\Big|_0^{1/4} \\ &= \frac{1}{4^4} \end{aligned}$$

(c) Find the expected value and variance of X.

$$\begin{aligned} \mathbb{E}(X) &= \int_0^1 x f(x) dx \\ &= \int_0^1 x 4x^3 dx \\ &= \int_0^1 4x^4 dx \\ &= 4\frac{x^5}{5}\Big|_0^1 \\ &= \frac{4}{5} \end{aligned}$$

1

To find the variance, we use the Magic Variance Formula.

$$\mathbb{E}(X^2) = \int_0^1 x^2 f(x)dx$$
$$= \int_0^1 x^2 4x^3 dx$$
$$= \int_0^1 4x^5 dx$$
$$= 4\frac{x^6}{6}\Big|_0^1$$
$$= \frac{4}{6} = \frac{2}{3}$$

By the Magic Variance Formula,

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$
$$= \frac{2}{3} - \left(\frac{4}{5}\right)$$
$$= \frac{2}{3} - \frac{16}{25}$$
$$= \frac{2}{75}$$

(d) Find the median of X.
First we find the CDF.

$$F(x) = \int_0^x f(t)dt$$
$$= \int_0^x 4t^3 dt$$
$$= t^4\Big|_0^x$$
$$= x^4$$

To find the median, set the CDF equal to $1/2$. Let m be the median. Then

$$1/2 = F(m)$$
$$= m^4$$

Thus $m = (1/2)^{1/4}$.

2. Suppose we are playing a game of darts. We throw darts at a circular dartboard with radius 1. Suppose the darts land uniformly at random on the dartboard, i.e. the probability of landing in a subset of the dartboard is equal to the area of the subset divided by the area of the dartboard.

(a) What is the probability that the dart hits the exact center of the dartboard?

The center is a single point with area of 0, so the probability is 0.

(b) What is the probability that the dart lands closer to the center than to the edge of the dartboard?

To land closer to the center than the edge, the dart must land in a circle of radius $1/2$, which has area $\pi(1/2)^2 = \pi/4$. Since the dartboard has area of $\pi 1^2 = \pi$, we divide these to get a probability of $1/4$.

(c) For $a < b < 1$, what is the probability that the dart lies at a distance between a and b of the center of the dartboard?

To land between a distance a and b of the center, we need to land in a ring with inner radius a and outer radius b. This ring has area $\pi b^2 - \pi a^2$. Dividing by the area of the circle π once again, the probability that the dart lands in this ring is $b^2 - a^2$.

3. A company that manufactures and bottles juice uses a machine that automatically fills 16-ounce bottles. There is some variation in the amount of liquid dispensed by the machine. The amount of juice dispensed has been observed to follow a normal distribution with mean of 16 ounces and standard deviation of 0.6 ounces.

(a) What is the probability that the machine dispenses more than 17 ounces?

Let X be the amount of liquid dispensed by the machine. Then $X \sim$ Normal(16, 0.6). Converting to the standard normal distribution,

$$\mathbb{P}(X > 17) = \mathbb{P}\left(Z > \frac{17 - 16}{0.6}\right) = \mathbb{P}(Z > 1.67)$$
$$= 1 - \mathbb{P}(Z \leq 1.67) = 1 - 0.9525 = 0.0475$$

where we used the Z-table to get the probabilities in the second line.

(b) What is the probability that the machine dispenses between 15.5 and 16.5 ounces?

$$\mathbb{P}(15.5 \leq X \leq 16.5) = \mathbb{P}\left(\frac{15.5 - 16}{0.6} \leq Z \leq \frac{16.5 - 16}{0.6}\right) = \mathbb{P}\left(-0.83 \leq Z \leq 0.83\right))$$
$$= 0.7967 - 0.2033 = 0.5934$$

(c) Suppose you wish to be 95% confident that the machine dispenses between 15.5 and 16.5 ounces. What standard deviation must your machine have for this to be the case?

Here we can just use the 68-95-99.7 rule. For 95% confidence, we want the range to be 2 standard deviations on either side of the mean, so since two standard deviations are 0.5, the machine must have a standard deviation of 0.25.

4. Yet another time, you find yourself the quality control manager for the Acme Widget Company. You are concerned about the number of defective widgets being produced by one of your factories.

(a) You have determined that the number of defective widgets produced per day by your factory has a mean of 10. You claim that the probability that your factory produces 15 or more defective widgets per day is less than or equal to 0.5. Is this claim justified? Explain your answer mathematically.

Let X be the number of defective widgets produced per day. Since we only know the mean, the best estimate we can get comes from Markov's Inequality:

$$\mathbb{P}(X \geq 15) \leq \frac{\mathbb{E}(X)}{15} = \frac{10}{15} = \frac{2}{3}$$

The claim is not justified since the best bound we can get is $2/3$.

(b) In addition, you have determined that the number of defective widgets produced per day by your factory has a variance of 5. You claim that the probability that your factory produces 15 or more defective widgets per day is less than or equal to 0.1. Is this claim justified? Explain your answer mathematically.

Here we know the variance as well, so we can use Chebyshev's Inequality. Recalling that Chebyshev's Inequality gives us a bound on deviation from the mean:

$$\mathbb{P}(|X - 10| \geq 5) \leq \frac{Var(X)}{5^2} = \frac{5}{5^2} = \frac{1}{5} = 0.20$$

We are interested in $\mathbb{P}(X \geq 15)$ which we know must be less than or equal to the probability above. Thus the best bound we can get is:

$$\mathbb{P}(X \geq 15) \leq \mathbb{P}(|X - 10| \geq 5) \leq 0.20$$

which is not good enough. Thus the claim is not justified.

(c) In addition, you have determined that the number of defective widgets produced per day by your factory has a distribution which is symmetric about its mean. You claim that the probability that your factory produces 15 or more defective widgets per day is less than or equal to 0.1. Is this claim justified? Explain your answer mathematically.

Since the distribution is symmetric about its mean, $\mathbb{P}(X \geq 15) = \frac{1}{2}\mathbb{P}(|X - 10| \geq 5)$, thus using the result of Chebyshev's Inequality above,

$$\mathbb{P}(X \geq 15) = \frac{1}{2}\mathbb{P}(|X - 10| \geq 5) \leq \frac{1}{2}0.20 = 0.10$$

Thus in this case the claim is justified.

(d) Finally, you have have determined that the number of defective widgets produced per day by your factory is approximately a normal distribution (with the mean and variance above). You claim that the probability that your factory produces 15 or more defective widgets per day is less than or equal to 0.02. Is this claim justified? Explain your answer mathematically.

To check this claim, we convert to the standard normal distribution. The variance is 5, so the standard deviation is $\sqrt{5}$. Thus we have:

$$\begin{aligned}
\mathbb{P}(X \geq 15) &= \mathbb{P}\left(Z \geq \frac{15 - 10}{\sqrt{5}}\right) \\
&= \mathbb{P}\left(Z \geq \frac{5}{\sqrt{5}}\right) \\
&= \mathbb{P}\left(Z \geq \sqrt{5}\right) \\
&= \mathbb{P}(Z \geq 2.24) \\
&= 1 - \mathbb{P}(Z \leq 2.24) \\
&= 1 - 0.9875 \\
&= 0.0125
\end{aligned}$$

where we used the Z-table to find our probability above. Thus the claim is justified. The take-home message from this problem is that the more information we have about a random variable, the better bound we can get on the probability of an outlier value.

5. In the previous homework we gave an example of approximating a binomial distribution by a Poisson distribution. In this problem, we will approximate a binomial distribution by a normal distribution. From looking at the pmfs of the binomial distribution (in class and in the notes), for large values of n the histogram of the binomial pmf looks "bell-shaped". It turns out that the normal distribution is a good approximation for the binomial approximation for large n. It also works for small n as long as p is not too far from 0.5. For convenience, we will let $q = 1 - p$.

(a) Let $X \sim$ Binomial(25, 0.4). Compute $\mathbb{P}(X = 10)$.

$$\mathbb{P}(X = 10) = \binom{25}{10}0.4^{10}0.6^{15} = 0.161158$$

(b) We will approximate X with a normal random variable. Let Y be a normal random variable with the same mean and variance as X. Compute $\mathbb{P}(9.5 \leq Y \leq 10.5)$. We will use this as an approximation for $\mathbb{P}(X = 10)$.

The mean and variance of X can be found from the probability distribution table. From that, we get:

$$\mathbb{E}(X) = np = 25(0.4) = 10$$
$$Var(X) = np(1 - p) = 25(0.4)(0.6) = 6$$

The standard deviation is $\sigma = \sqrt{6}$. We use the same parameters for Y, i.e. $Y \sim$ Normal$(10, \sqrt{6})$. Converting to the standard normal random variable, we have:

$$\mathbb{P}(9.5 \leq Y \leq 10.5) = \mathbb{P}\left(\frac{9.5 - 10}{\sqrt{6}} \leq Z \leq \frac{10.5 - 10}{\sqrt{6}}\right)$$
$$= \mathbb{P}\left(-0.20 \leq Z \leq 0.20\right)$$
$$= 0.5793 - 0.4207$$
$$= 0.1586$$

(c) Briefly explain why we used the event $(9.5 \leq Y \leq 10.5)$ to approximate the event $(X = 10)$.

The binomial distribution is discrete and the normal distribution is continuous, so we need a range of values in the normal distribution to approximate a value in the binomial distribution. The symmetric interval $[9.5, 10.5]$ centered about 10 is a reasonable approximation.

(d) Compute the relative error in your approximation. You may refer back to the previous problem set for the definition of relative error.

$$\text{Relative Error} \frac{|0.1612 - 0.1586|}{0.1612} = 0.016$$

This is between 1 and 2 percent, which is pretty good.

The general rule is that the normal approximation is "good enough" if

$$0 \leq p \pm 3\sqrt{\frac{pq}{n}} \leq 1$$

(e) Show that $0 \leq p \pm 3\sqrt{\frac{pq}{n}} \leq 1$ implies that

$$n \geq 9\left(\frac{p}{q}\right) \text{ and } n \geq 9\left(\frac{q}{p}\right)$$

Using the left inequality with the minus sign:

$$0 \leq p - 3\sqrt{\frac{pq}{n}}$$
$$p \geq 3\sqrt{\frac{pq}{n}}$$
$$p^2 \geq 9\frac{pq}{n}$$
$$n \geq 9\frac{q}{p}$$

6

Using the right inequality with the plus sign:

$$\leq p + 3\sqrt{\frac{pq}{n}} \leq 1$$

$$3\sqrt{\frac{pq}{n}} \leq 1 - p = q$$

$$9\frac{pq}{n} \leq q^2$$

$$n \geq 9\frac{p}{q}$$

(f) Using this rule, how large should n be to approximate a binomial distribution with $p = 0.4$, $p = 0.8$, $p = 0.9$, and $p = 0.99$? Was our approximation in part (b) justified according to this rule?

- For $p = 0.4$, $n \geq 9 \cdot 0.4/0.6 = 6$ and $n \geq 9 \cdot 0.6/0.4 = 13.5$, so need $n \geq 13.5$. Since it makes no sense for n to be a non-integer, we need $n \geq 14$.
- For $p = 0.8$, $n \geq 9 \cdot 0.8/0.2 = 36$ and $n \geq 9 \cdot 0.2/0.8 = 2.25$, so need $n \geq 36$. You really only need to check the first inequality since you can see it will give a larger bound.
- For $p = 0.9$, $n \geq 9 \cdot 0.9/0.1 = 81$ and $n \geq 9 \cdot 0.1/0.9 = 1$, so need $n \geq 81$.
- For $p = 0.99$, $n \geq 9 \cdot 0.99/0.01 = 891$ and $n \geq 9 \cdot 0.01/0.99 = 0.0909$, so need $n \geq 891$.

For part (b), we need $n \geq 14$ based on the above. Since $n = 25$ in our binomial distribution, we are justified in making our approximation.