1 Racing Candidate

Suppose that 40\% of a large population are in favor of candidate A. A pollster selects a random sample of 20 people and asks them about their opinion about the candidate. Let \( X \) be the number of people in favor of candidate A.

a) What is the sample space of \( X \), what is the distribution of \( X \)?

The sample space, or the image of \( X \) is the set of all possible values of \( X \). The possible values for \( X \) are the integers from zero through 20: \( \text{Im}(X) = \{0, \ldots, 20\} \)

\( X \) has a Binomial distribution with parameters \( n = 20 \) and success rate \( p = P(\text{person in favor of A}) = 0.4, \ i.e. \)

\[ X \sim B_{20, 0.4} \]

b) What is the probability that 10 people in the sample are in favor of candidate A?

\[ P(X = 10) = \binom{20}{10} 0.4^{10} 0.6^{10} = 0.117. \]

c) What is the probability that between 6 and 10 people in the sample are in favor of candidate A? (limits are included)

\[ P(6 \leq X \leq 10) = P(X \leq 10) - P(X \leq 5) = B_{20, 0.4}(10) - B_{20, 0.4}(5) = 0.8725 - 0.1256 = 0.7469. \]

\[ P(X \geq 11) = 1 - P(X \leq 10) = 1 - 0.8725 = 0.1275. \]

e) What is the expected value of \( X \), what is the variance?

\[ E[X] = 20 \cdot 0.4 = 8, \ Var[X] = npq = 20 \cdot 0.4 \cdot 0.6 = 4.8 \]

2 Snow in October

After October 1st the probability that a blizzard will occur on any particular day in the Midwest of Northern America is 0.1. To simplify the problem, assume for the following questions that October has infinitely many days.

a) What is the probability that there won’t be a blizzard in the first 10 days of October?

The probability that there will not be a blizzard is \( 0.9^{10} = 0.349. \)
b) What is the probability that the first blizzard will occur on October 20?

\[ P(\text{no blizzard in nineteen days, blizzard on 20th}) = 0.9^{19} \cdot 0.1 = 0.0135. \]

c) What is the expected date for the first blizzard (starting October 1st)?

For the expected value, we need some random variable \(X\) together with its probability mass function. One way to define \(X\) is \(X := \text{"date of first blizzard"}\) or, since it's not easy to compute with dates \(X = \#\text{ of days until first blizzard including day of blizzard} \). This random variable has a geometric distribution with \(p = 0.1\), we write \(X \sim \text{Geo}_{0.1}\).

For a geometric random variable we already know the expected value:

\[ E[X] = \frac{1}{p} = \frac{1}{0.1} = 10. \]

The expected date for the first blizzard would therefore be Oct 10th.

d) How does the distribution change when October only has 31 days? with infinitely many days, \(X\) has a geometric distribution; when we are only looking at the first 31 days, Let \(Y\) be the variable where October has 31 days, as it should. \(Y\) can be 31 at most, on top of that we have the event, that there is no blizzard in October. The probability for no snow storm in October is \(0.9^{31}\), and \(P(Y = k) = P(X = k)\) for any value of \(k\) of less than 32. From that we can compute the expected value, the variance and the distribution ...

3 Closing Windows, Baron p.78

About ten percent of users do not close Windows properly. Suppose that Windows is installed in a public library that is used by random people in a random order.

(a) On average, how many users of this computer do not close Windows properly before someone does? State a random variable and its distribution (with all parameters) for full credit.

Let \(X\) be the number of users that do not close Windows properly before someone does including that person. Then \(X\) has a geometric distribution with a success rate of \(p = 0.9\). The expected value of \(X\) is given as \(1/p = 1.1111\). Define a random variable \(Y\) to be the number of users that do not close windows properly before someone does, not including the final person who closes windows properly. If \(X = 1, then Y = 0; if X = 2, then Y = 1; if X = 3, then Y = 2, etc. In general, if \(X = k\), then \(Y = k - 1\), so we can write \(Y\) as a function of \(X\): \(Y = X - 1\). Therefore, \(E[Y] = E[X] - 1 = 1.1111\). The random variable \(Y\) has a modified geometric distribution with \(p = .9\).

(b) What is the probability that exactly 8 of the next 10 users will close Windows properly? State a random variable, give its image and distribution.

Let \(Y\) denote the number of people who do close Windows properly (out of 10 users). Then \(Y \sim B_{10,0.9}\) and

\[ P(Y = 8) = \binom{10}{8} 0.9^8 0.1^2 = 0.1937. \]

4 Deaths from Horse Kicks

After analyzing data from the Prussian cavalry for a period of 20 years, statisticians came to the conclusion that the number of deaths from horse kicks follows a Poisson distribution with \(\lambda = 0.6\). Let \(X\) be the number of deaths from horse kicks in 20 years. What is the probability that
a) no deaths occurred?

\[ X \text{ has a Poisson distribution.} \]

\[ P(X = 0) = 0.6^0 e^{-0.6} / 0! = e^{-0.6} = 0.5488 \]

b) one or two deaths occurred?

\[ P(1 \leq X \leq 2) = P(X = 1) + P(X = 2) = 0.6^1 e^{-0.6} / 1! + 0.6^2 e^{-0.6} / 2! = 0.4281 \]

c) 3 or more deaths occurred?

\[ P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.9769 = 0.0231 \]
d) more than 3 deaths occurred?

\[ P(X > 3) = 1 - P(X \leq 3) = 1 - 0.9966 = 0.0034 \]

5 Which distribution is it?

Some jobs submitted for processing on a particular CPU have fatal programming errors, while others do not. Suppose that the long run fraction of jobs with fatal programming errors is \( p = 0.05 \).

a) Find (under appropriate distribution assumptions) the probability that among the next 10 jobs submitted there are less than 3 with fatal errors. \( X := \text{"\# of jobs with fatal errors\" has a binomial distribution with } n = 10, p = 0.05. \)

\[ P(X \leq 2) = B_{10,0.05}(2) = 0.9885 \]

b) Find (under appropriate model assumptions) the probability that among the next 25 jobs submitted there are less than 5 with fatal errors.

\( X \sim B_{25,0.05} \) but we can use the Poisson approximation with \( \lambda = 25 \cdot 0.05 = 1.25. \)

\[ P(X < 5) = P(X \leq 4) = B_{25,0.05}(4) \approx P_{01.25}(4) \text{table look-up } 0.9912 + 0.9906 \frac{2}{2} = 0.9909. \]

c) One begins monitoring the processing of jobs and lets

\( X = \text{the total number of jobs processed until the first with a fatal error.} \)

Find (again under appropriate distribution assumptions) \( P(X \leq 30) \)

\( X \) now has a geometric distribution with \( p = 0.05. \)

The probability mass function is

\[ p_X(k) = \frac{0.95^{k-1} \cdot 0.05}{k \text{ good ones} \cdot \text{fatal job}}. \]
Then,

\[ P(X \leq 30) = 1 - P(\text{at least 30 good jobs}) = 1 - 0.95^{30} = 0.7854. \]

Another correct answer is to interpret \( X \) as the number of jobs processed before the first one with a fatal error. With this interpretation, \( X \) has a modified geometric distribution with \( p = 0.05 \). In other words, \( X = Y - 1 \), where \( Y \) has a geometric distribution with \( p = 0.05 \). Then,

\[ P(X \leq 30) = P(Y \leq 31) = 1 - 0.95^{31} = 0.796. \]

d) Monitor the processing of jobs again and define

\[ Y = \text{the total number of jobs processed until the 2nd with a fatal error}. \]

Find the probability mass function for \( Y \).

\[
p_Y(k) = P(k \text{ jobs, 1 fatal among the first } k-1, \text{ one fatal job in the end}) = \frac{(k-1) \cdot 0.95^{k-2} \cdot 0.05 \cdot 0.05}{1 \cdot k-1} = \binom{k-1}{1} \cdot 0.95^{k-2} \cdot 0.05^2.
\]

6 **Homicides in Chicago**

It has been suggested that the number of homicides that happen during one month in Des Moines follow a Poisson distribution with parameter \( \lambda = 3 \).

(a) What is the probability that there occur less than 4 homicides in Des Moines in two months? Let \( X \) be the number of homicides in Des Moines in two months. Then \( X \sim Po_6 \) and

\[ P(X < 4) = P(X \leq 3) = Po_6(3) = 0.1512. \]

(b) In a given month the number of homicides in Chicago has been 24. Does this indicate that Chicago is as safe as Des Moines? Explain why or why not.

*There are many acceptable answers, depending on your interpretation. Here is a great answer from Dr. Hofmann.*

In order to make a comparison between Chicago or DSM, we will have to adjust rates to the same units - i.e. we could for example look at homicide rates per 100,000 people. Based on a single number for Chicago, we can’t draw any conclusion about its safety, but ... we can google for the FBI crime data by city and find out, that Chicago had a population of close to 3 Mio people in 2006 and 201 homicides in the first 6 months. This gives us a annual homicide rate in Chicago of \( 2 \cdot 201/30 = 13.4 \) per 100,000 people.

*Des Moines has roughly 200,000 people, which would correspond to annual homicide rate of 3/2\cdot12 = 18 per 100,000 people. BUT: I made up the above murder rate - the murder rate in Des Moines for 2006 is actually just 2 per 100,000!*

*So yes, Des Moines is safer than Chicago in terms of homicides.*