ECS 132 Spring 2024: Assignment 3
Solution

This problem set covers:

- Conditional probability, problems 1-2.
  Problem 1 should be straightforward and 2 is more involved.
- Discrete random variables and the Bernoulli, Binomial and Geometric distributions, problems 3-5.
  Problem 3 should be very simple, and 4 and 5 more involved.

1 Problem

Suppose the university has designed a e-mail spam filter that attempts to identify spam by looking for commonly occurring phrases in spam. E-mail analysis has shown that 80% of email is spam. Suppose that 10% of the spam email contain the phrase "Large inheritance", whereas this phrase is only used in 1% of non-spam emails. Suppose a new email is received containing the phrase "Large inheritance", what is the probability that it is spam?

Answer

We define the following 2 events

- S: event that an e-mail is spam
- F: event that the email has the phrase "Large inheritance"

Using Bayes' Rule we find

$$\begin{aligned} P(S|F) &= \frac{P(F|S)P(S)}{P(F)} \\ &= \frac{P(F|S)P(S)}{P(F|S)P(S) + P(F|S^c)P(S^c)} \\ &= \frac{0.1 \times 0.8}{0.1 \times 0.8 + 0.01 \times 0.2} \\ &\approx 0.9756 \end{aligned}$$

2 Problem

A binary communication channel carries data as one of two types of signals denoted by 0 and 1. Owing to noise, a transmitted 0 is sometimes received as a 1 and a transmitted 1 is sometimes received as a 0. For a given channel, assume a probability of 0.94 that a transmitted 0 is correctly received as a 0 and a probability 0.91 that a transmitted 1 is received as a 1. Further assume a probability of 0.45 of transmitting a 0. If a signal is sent, determine

1. Probability that a 1 is received.
2. Probability that a 0 is received.
3. Probability that a 1 was transmitted given that a 1 was received.
4. Probability that a 0 was transmitted given that a 0 was received.
5. Probability of an error (this means either a transmitted 0 was received as a 1, or a transmitted 1 was received as a 0).

Answer

Define the following events:

- T_0 : A 0 is transmitted
- R_0 : A 0 is received
- T_1 : $\overline{T_0}$
- R_1 : $\overline{R_0}$

Then the statements given in the problem formulation can be translated to the following:

- $P(R_0|T_0) = 0.94$
- $P(R_1|T_1) = 0.91$
- $P(T_0) = 0.45$

From those statements we can directly deduce:

- $P(R_1|T_0) = 0.06$
- $P(R_0|T_1) = 0.09$
- $P(T_1) = 0.55$

This provides all the information we need.

1) The law of total probability tells us that:

$$P(R_1) = P(R_1|T_0)P(T_0) + P(R_1|T_1)P(T_1) = (0.06)(0.45) + (0.91)(0.55) = 0.5275$$

2) The law of total probability tells us that:

$$P(R_0) = P(R_0|T_0)P(T_0) + P(R_0|T_1)P(T_1) = (0.94)(0.45) + (0.09)(0.55) = 0.4725$$

3)

$$P(T_1|R_1) = \frac{P(T_1)P(R_1|T_1)}{P(R_1)} = \frac{(0.55)(0.91)}{0.5275} = 0.949$$

4)

$$P(T_0|R_0) = \frac{P(T_0)P(R_0|T_0)}{P(R_0)} = \frac{(0.45)(0.94)}{0.4725} = 0.895$$

5) Probability of an error:

$$\begin{aligned} P(\text{ Error}) &= P(R_0 \cap T_1) + P(R_1 \cap T_0) \\ &= P(R_0|T_1)P(T_1) + P(R_1|T_0)P(T_0) \\ &= (0.09)(0.55) + (0.06)(0.55) = 0.0765 \end{aligned}$$

3 Problem

You have two sensors who can be active or asleep in 3 different time slots. Each sensor is independent and active with probability p=1/2 in each slot, and the slots are independent of one another. Consider the following:

- Event A_i = both sensors are active in slot i.
- Events A_i and A_j are independent if $i \neq j$.
- Event B = the two sensors are active in all three time slots.
- The random variable $X = 1$ if event B occurs, and $X = 0$ otherwise.

Answer these questions:

1. What is $P(X = 1)$?
2. To what family of distributions does the distribution of X belong? Provide this answer as X $\sim$ distribution_name(parameters). See the lecture notes for examples.
3. What is the value of $E(X)$?
4. What is the value of $Var(X)$?

Answer:

1. $X = 1$ corresponds to the event that both sensors are active in all three time slots. Recall HW2, Problem 4 where we derived this probability. We start by considering the probability that both sensors are active in a slot i which is $P(A_i) = 1/4$ since $S = \{(0,0),(0,1),(1,0),(1,1)\}$ and only one of those 4 states corresponds to both sensors on. The probability the sensors are both active in three **independent** slots is $P(A_1)\ P(A_2)\ P(A_3) = (1/4)^3 = (1/64)$. So
$$P(X = 1) = 1/64$$
2. This is a single trial with one of two outcomes: either they are both on in three slots or they are not. The former (which we call "success") happens with $p = 1/64$. So
$$X \sim \text{Bern}(p = 1/64)$$
3. For a Bernoulli distribution $E(X) = p = \frac{1}{64}$.
4. For a Bernoulli distribution $Var(x) = p(1-p) = \frac{1}{64} \cdot \frac{63}{64} = \frac{63}{4096} = 0.0154$.

4 Problem

The probability that a patient recovers from a rare blood disease is 0.4 and a total of 10 people are known to have contracted this disease. Let X denote the random variable which corresponds to the number of patients who survive the disease. Assume that the patient's recoveries are all independent of one another.

1. What is the equation for the probability that X = k of the ten people survive?
2. To what family of distributions does the distribution of X belong? Provide this answer as X $\sim$ distribution_name(parameters). See the lecture notes for examples.

For the rest of the problem, use R to do the things listed below. You can use the built-in functions for this family of distributions. Include your code in the pdf writeup you submit to Gradescope and submit your code to Canvas as a .R or .ipynb file.

3. Plot the probability mass function (pmf) of X.
4. Plot the cumulative distribution function (cdf) of X.
5. What is the probability that at least 8 survive, i.e., $P\{X \geq 8\}$?
6. What is the probability that 3 to 8 survive, i.e., $P\{3 \leq X \leq 8\}$?

Answers

1. $P(X = k) = \binom{10}{k} p^k (1-p)^{10-k} = \binom{10}{k} (0.4)^k (0.6)^{10-k}$.
2. The binomial family, X $\sim$ Binom(n=10, p=0.4).
3. Plot the probability mass function (pmf) of X. We can use R to get all the values and plot them.

```
n <- 10
k <- seq(0,n)
p <- 0.4
pr <- dbinom(k, n, p)
pr
[1] 0.0060466176 0.0403107840 0.1209323520 0.2149908480 0.2508226560
[6] 0.2006581248 0.1114767360 0.0424673280 0.0106168320 0.0015728640
[11] 0.0001048576
barplot(pr, names.arg=k, ylab = "p(i)", xlab = "i")
```

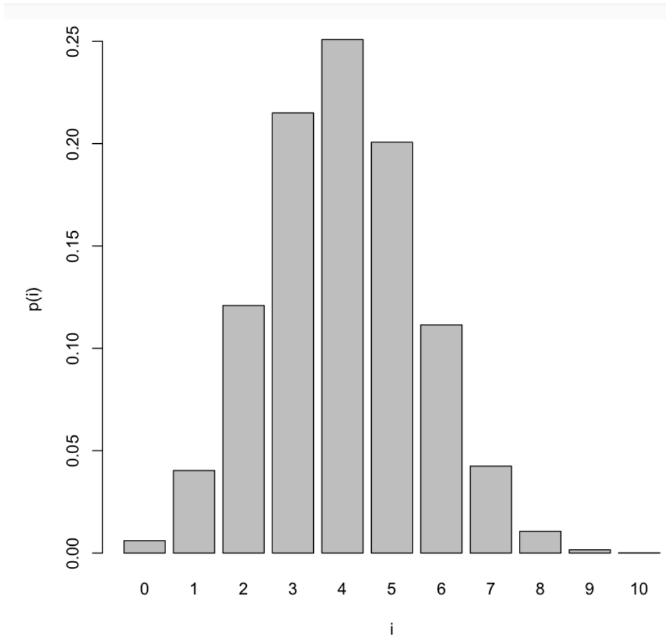

Figure 1: The probability mass function of Binomial random variable with parameter $n = 10$ and $p = 0.4$.

4. Plot the cumulative distribution function (cdf) of X.

```
pr <- pbinom(k, n, p)
[1] 0.006046618 0.046357402 0.167289754 0.382280602 0.633103258 0.833761382
[7] 0.945238118 0.987705446 0.998322278 0.999895142 1.000000000
barplot(pr, names.arg=k, ylab = "P(X <= i)", xlab = "i")
```

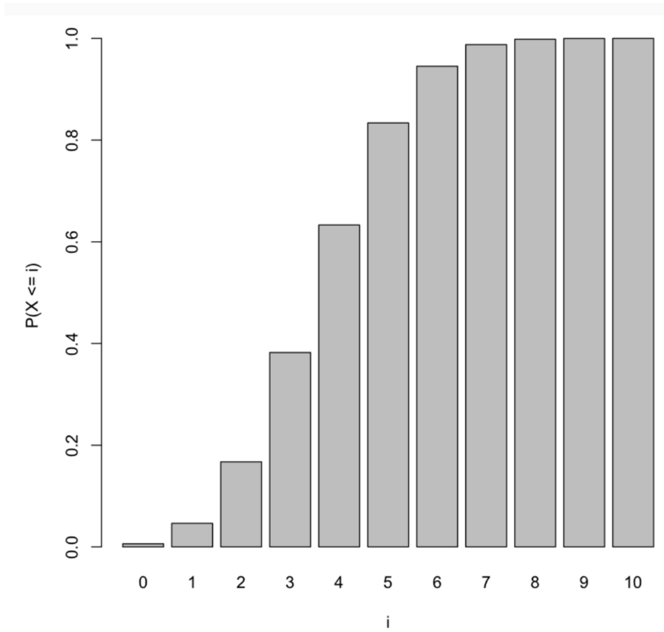

Figure 2: The cumulative distribution function of Binomial random variable with parameter $n = 10$ and $p = 0.4$.

5. What is the probability that at least 8 survive, i.e., $P\{X \geq 8\}$?

   $P\{X \geq 8\} = 1 - P\{X \leq 7\}$. Using R, we can do the following

```
pr = pbinom(7, n, 0)
ans <- 1 - pr
ans
0.01229455
```

6. What is the probability that 3 to 8 survive, i.e., $P\{3 \leq X \leq 8\}$? The required probability is

$$\begin{aligned} P\{3 \leq X \leq 8\} &= P\{X = 3\} + P\{3 < X \leq 8\} \\ &= P\{2 < X \leq 8\} \\ &= \text{pbinom}(8, n, p) - \text{pbinom}(2, n, p) \end{aligned}$$

   Again, we can use R

```
pbinom(8, n, p) - pbinom(2, n, p)
```

5 Problem

Consider the following program statement consisting of a **while** loop

$$\text{w}hile \ \neg \text{B} \ do \ S$$

Assume that the Boolean expression B takes the value true with probability p and the value false with probability q. Assume that the successive test on B are independent.

1. Find the probability that the loop will be executed k times.
2. Find the expected number of times the loop will be executed.
3. Considering the same above assumptions, suppose the loop is now changed to

$$\text{r}epeat \ \text{S} \ until \ B$$

What is the expected number of times that the repeat loop will be executed?

Answer

1. Let random variable X denotes the times the loop will be executed. X is distributed according to geometric* and can take values $\{0, 1, 2, \ldots, \}$. X takes the value 0 if B takes the value True the first time. Similarly, X will take the value 1 if B takes value False the first time and then the value True on the second time. And so on. The pmf is given by

$$\begin{aligned} P(X = k) &= (1-p)^k p \qquad k \in \{0, 1, 2, \ldots, \} \\ &= q^k p \end{aligned}$$

Consequently, $X \sim \text{Geom}^*(p)$ and represents number of failure until success for independent Bernoulli trials.

2. You can go to the Jupyter notebook and see that it is stated as a fact that for $X \sim \text{Geom}^*(p)$, that

$$E[X] = \frac{1-p}{p} = \frac{q}{p}.$$

But here we want to provide the full derivation. First recall how to sum a geometric series where a is an arbitrary constant:

$$S = \sum_{k=0}^{\infty} aq^k = a \sum_{k=0}^{\infty} q^k$$

To achieve a closed-form solution we first multiply this by q:

$$qS = \sum_{k=0}^{\infty} aq^{k+1} = a \sum_{k=0}^{\infty} q^{k+1}$$

Subtracting the second from the first:

$$(1-q)S = a \left[\sum_{k=0}^{\infty} q^k - \sum_{k=0}^{\infty} q^{k+1} \right] = ax^0 = a$$

So we find:

$$S = \frac{a}{(1-q)} \text{ for } |q| < 1$$

Now consider the expectation value $X \sim \mathrm{Geom}^*(p)$. By definition:

$$\begin{aligned} E[X] &= \sum_{k=0}^{\infty} P(X=k)k \\ E[X] &= p\sum_{k=0}^{\infty} q^k k \end{aligned}$$

Note we can write the this in terms of a derivative of the geometric series:

$$\begin{aligned} E[X] &= p\sum_{k=0}^{\infty} q^k k = p\left(q\frac{d}{dq}S\right) \\ &= pq\frac{d}{dq}\left[\frac{1}{(1-q)}\right] = pq\frac{1}{(1-q)^2}. \end{aligned}$$

Recall that $p = (1-q)$ so the equation simplifies:

$$\begin{aligned} E[X] &= (1-q)q\frac{1}{(1-q)^2} \\ &= \frac{q}{(1-q)} \\ &= \frac{q}{p} \end{aligned}$$

3. Let random variable Y denote the times the repeat loop will be executed. Since the check is done at the end of the loop, the loop will be executed at least once. It is easy to see that the random variable Y takes value $\{1, 2, 3, \ldots\}$. $Y \sim$ Geom(p). In the Jupyter notebooks it is stated that for $Y \sim \mathrm{Geom}(p)$ that

$$E[Y] = \frac{1}{p}.$$

We provide the full derivation here.

$$\begin{aligned} P(Y=k) &= (1-p)^{k-1}p \qquad k \in \{1, 2, \ldots, \} \\ &= q^{k-1}p \end{aligned}$$

We can think of Y as the number of trials until success. Thus, when $X = k$, the first $k-1$ trials must be failure and the kth trial must be success.

$$
\begin{aligned}
E(Y) &= \sum_{k=1}^{\infty} kP(Y=k) \\
&= \sum_{k=1}^{\infty} kq^{k-1}p \\
&= p\sum_{k=1}^{\infty} kq^{k-1} \\
&= p\sum_{j=0}^{\infty} (j+1)q^{j} \quad \text{this by setting } j=k-1 \text{ in the above eq.} \\
&= p\left(\sum_{j=0}^{\infty} jq^{j} + \sum_{j=0}^{\infty} q^{j}\right). \\
&= p\left(\frac{q}{p^2} + \frac{1}{1-q}\right). \\
&= \frac{1}{p}
\end{aligned}
$$