Assignment 4 (Due on Tuesday, May 5)

Instructions

As usual, please upload one file on Blackboard including your answers/solutions to problems. For the R problems, include your input and output. A random subset of problems will be graded but you need to do all the problems (unless you trust your luck (probably you shouldn't)).

R functions you may need

We will use R for problems regarding binomial and Poisson distributions. In addition to R commands we have learnt previously, the following may help. In order to get help for any of the below functions, you can use the `help` function in R. For example, with typing `help(dbinom)`, you can get help about the `dbinom` function.

Binomial distribution

Suppose X is a binomial random variable with parameters n and p. (Think of X as the number of successes you will have in n experiments, where the success probability is p for each experiment, independently of each other.) For $0 \leq j \leq n$,

$$P(X = j) = \binom{n}{j} p^j (1-p)^{n-j}.$$

- dbinom(x, size, prob) (probability density function) gives the probability $P(X = x_i)$ for all x_i in the vector x.
- pbinom(x, size, prob) (cumulative distribution function) gives the probability $P(X \leq x_i)$ for all x_i in the vector x.
- qbinom(q, size, prob) (quantile function) gives the smallest integer m_i for which $P(X \leq m_i) \geq q_i$ for all q_i in the vector q.

  Here the parameters are:

- x is a vector of nonnegative integers (can be a single number).
- q is a vector of probabilities.
- size (n) is the number of trials (experiments).
- prob (p) is the probability of success of each trial.

Example 1. Suppose we are tossing a coin 5 times and counting the number of heads, which is given by the binomial random variable with parameters 5 (size) and 0.5 (success rate p).

1. If we type `dbinom(0:5, 5, 0.5)`, we see

```
> dbinom(0:5, 5, 0.5)
[1] 0.03125 0.15625 0.31250 0.31250 0.15625 0.03125
```

which shows

$$
\begin{aligned}
P(X = 0) &= 0.03125 \\
P(X = 1) &= 0.15625 \\
P(X = 2) &= 0.31250 \\
P(X = 3) &= 0.31250 \\
P(X = 4) &= 0.15625 \\
P(X = 5) &= 0.03125
\end{aligned}
$$

Recall that 0:5 gives you the vector (0,1,2,3,4,5).

2. If we type `pbinom(0:5, 5, 0.5)`, we see

```
> pbinom(0:5, 5, 0.5)
[1] 0.03125 0.18750 0.50000 0.81250 0.96875 1.00000
```

These numbers show

$$
\begin{aligned}
P(X \leq 0) &= 0.03125 \\
P(X \leq 1) &= 0.18750 \\
P(X \leq 2) &= 0.50000 \\
P(X \leq 3) &= 0.81250 \\
P(X \leq 4) &= 0.96875 \\
P(X \leq 5) &= 1
\end{aligned}
$$

3. If we type `qbinom(c(0.1,0.7), 5, 0.5)`, we see

```
> qbinom(c(0.1,0.7), 5, 0.5)
[1] 1 3
```

This means

$$
\begin{aligned}
P(X \leq 0) < 0.1 \quad &\text{and} \quad P(X \leq 1) \geq 0.1 \\
P(X \leq 2) < 0.7 \quad &\text{and} \quad P(X \leq 3) \geq 0.7
\end{aligned}
$$

(The last line says that getting 2 or fewer heads has a chance strictly smaller than 0.7 but getting 3 or fewer heads has chance at least 0.7)

Poisson distribution

Poisson distribution has one parameter, usually denoted as μ or λ. If Y is a Poisson random variable with parameter μ, then μ is the mean value of Y and for $k = 0, 1, 2, \ldots$

$$P(Y = k) = e^{-\mu} \frac{\mu^k}{k!}$$

Density, cumulative distribution, and quantile functions are given as

- dpois(x, lambda, log = FALSE) (probability density function) gives the probability $P(Y = x_i)$ for all x_i in the vector x.
- ppois(q, lambda, lower.tail = TRUE, log.p = FALSE) (cumulative distribution function) gives the probability $P(Y \leq q_i)$ for all q_i in the vector q.

- qpois(p, lambda, lower.tail = TRUE, log.p = FALSE) (quantile function) gives the smallest integer m_i for which $P(Y \leq m_i) \geq p_i$ for all p_i in the vector p. (This is when lower.tail= TRUE. If lower.tail= FALSE, then it gives the largest integer m_i for which $P(Y \geq m_i) \geq p_i$ for all p_i in the vector p)

These are the analogs of dbinom, pbinom, and qbinom functions discussed above. The parameters are

- x: vector of (non-negative integer) quantiles.
- q: vector of quantiles.
- p: vector of probabilities.
- lambda (our mu=μ): vector of (non-negative) means.
- log, log.p: logical; if TRUE, probabilities p are given as log(p). [Don't worry about this one for this lab]
- lower.tail: logical; if TRUE (default), probabilities are $P(Y \leq x)$, otherwise, $P(X > x)$.

Note that the default values for lower.tail and log/log.p are TRUE and FALSE, respectively. So if you don't need to change the default values, you don't need to include them in your functions.

Example 2. Suppose X is a Poisson distribution with parameter (mean) 12.

1. To find $P(X = 16)$, you can type

```
> dpois(16, lambda=12)  (or dpois(16, 12))
[1] 0.05429334
```

2. To find $P(X = 11)$, $P(X = 12)$, $P(X = 13)$ you can type

```
> dpois(11:13, lambda=12)  (or dpois(11:13, 12))
[1] 0.1143679 0.1143679 0.1055704
```

3. To find $P(X = 11) + P(X = 12) + P(X = 13)$ you can type

```
> ppois(13, lambda=12) - ppois(10, lambda=12)
[1] 0.3343062
```

or

```
> x= dpois(11:13, lambda=12)
> sum(x)
[1] 0.3343062
```

Remark. Suppose you have a binomial distributi with parameters n and p. If n is large and p is small, you can use the Poisson distribution with mean np to approximate the binomial distribution.

Problems

Part 1: Do the problems by hand

1. One card is randomly drawn from a standard deck of 52 cards. What is the probability that it is an ace or a spade?

2. A deck of cards is shuffled. True or false, and explain briefly:

   (a) The chance that the top card is the jack of clubs equals 1/52.

   (b) The chance that the bottom card is the jack of diamonds equals 1/52.

   (c) The chance that the top card is the jack of clubs or the bottom card is the jack of diamonds equals 2/52.

   (d) The chance that the top card is the jack of clubs or the bottom card is the jack of clubs equals 2/52.

   (e) The chance that the top card is the jack of clubs and the bottom card is the jack of diamonds equals $1/52 \times 1/52$.

   (f) The chance that the top card is the jack of clubs and the bottom card is the jack of clubs equals $1/52 \times 1/52$.

3. Suppose 10% of a population is affected by a disease. And suppose we have a test for this disease that is accurate 90% of the time. If an individual is tested positive, what is the probability that this individual is actually carrying this disease?

4. A pair of fair dice is tossed. Find the probability of getting at most a total of 5.

5. Suppose it is known that in a certain population, 10% of the people are color-blind. If a random sample of 12 people is drawn, find the probability that at least 2 of them are color-blind.

6. The ER of a hospital receives 3 patients in an hour on average. What is the probability that there are 5 visits to the ER in a given two-hour period?

Two fun problems (not to be graded).

(I) Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

(II) A family has two children. At least one child is a boy born on a Tuesday. What is the probability that the other one is also a boy?

Part 2: Use R to do the problems

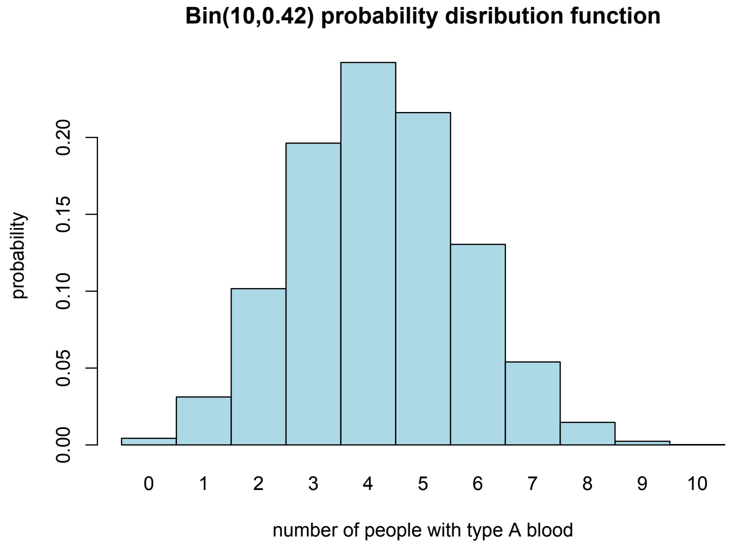
7. In the United States, 42% of the population has type A blood. Consider taking a sample of size 10. Let Y denote the number of people in the sample with type A blood. Find

   (a) $P(Y = 2)$

   (b) $P(2 \leq Y \leq 6)$

   (c) $P(Y > 3)$

   (d) Make a bar graph of this distribution. (You need the values $P(X = 0)$, $P(X = 1)$, $\ldots$, $P(X = 10)$.) Your plot should look like the following:



(In barplot function, you can use the arguments `space=0` and `names=0:10`)

8. An article in the Los Angeles Times (December 3, 1993) reports that 1 in 200 people carry the defective gene that causes inherited colon cancer. In a sample of 1000 individuals, we want to calculate the probability that a certain number of individuals carry the gene. We will use the Poisson distribution to approximate the binomial distribution. Find the following probabilities:

   (a) Between 5 and 8 (inclusive) carry the gene.

   (b) At least 8 carry the gene.

9. The number of people arriving for treatment at an emergency room can be modeled by a Poisson process with a rate parameter of 5 per hour. In other words, the mean number of people arriving in one hour is 5.

   (a) What is the probability that exactly four arrivals occur during a particular hour?

   (b) What is the probability that at least 4 people arrive during a particular hour?

   (c) What is the probability that at most 4 people arrive during a particular hour?

   (d) Since the mean number of people arriving during one hour is 5, the mean number of people arriving in half hour would be 2.5. Given this information, what is the probability that more than 5 people would arrive in any particular 45-minute period?