

# Mathematical Statistics Homework 2

Joshua Ingram

3/1/2021

## 1. Longest Run

a.

Let  $X$  be the longest run of heads or tails. If a fair coin is tossed four times, then the sample space  $S$  is as follows:

Toss 1	Toss 2	Toss 3	Toss 4
H	H	H	H
T	H	H	H
H	T	H	H
T	T	H	H
H	H	T	H
T	H	T	H
H	T	T	H
T	T	T	H
H	H	H	T
T	H	H	T
H	T	H	T
T	T	H	T
H	H	T	T
T	H	T	T
H	T	T	T
T	T	T	T

Given  $S$ , the possible values that  $X$  may take on are  $k = 1, 2, 3, 4$ . We find the induced probabilities  $P(X = k)$  for all  $k$  as

$$P(X = 1) = P(\{HTHT, THTH\}) = \frac{2}{16} = 0.125$$

$$P(X = 2) = P(\{HTHH, TTHH, HHTH, HTTH, THHT, TTHT, HHTT, THTT\}) = \frac{8}{16} = 0.5$$

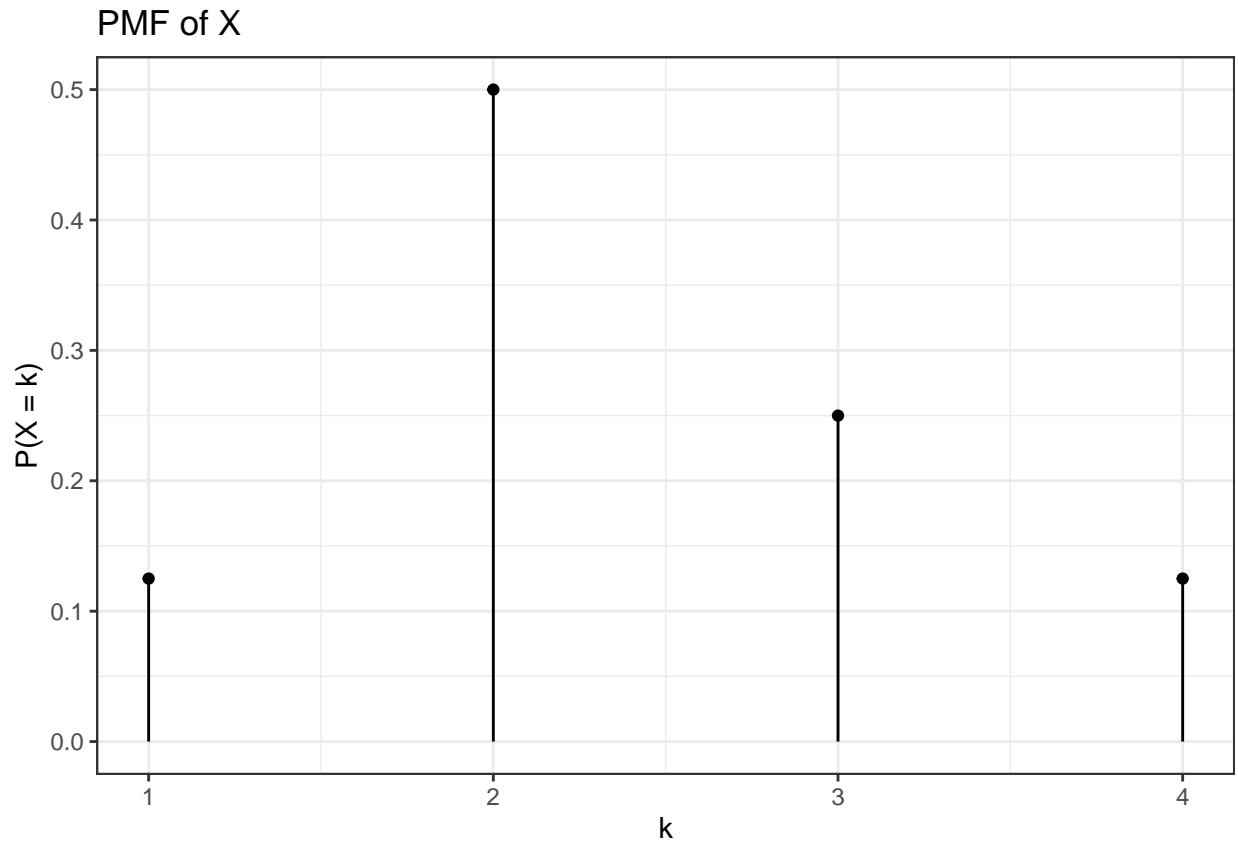
$$P(X = 3) = P(\{THHH, TTTH, HTTT, HHHT\}) = \frac{4}{16} = 0.25$$

$$P(X = 4) = P(\{HHHH, TTTT\}) = \frac{2}{16} = 0.125$$

Just to check that these are correct, these probabilities should sum to 1.

$$P(\cup_{k=1}^4 k) = 0.125 + 0.5 + 0.25 + 0.125 = 1$$

Now, we provide the Probability Mass Function (PMF) of the random variable  $X$ :



**b.**

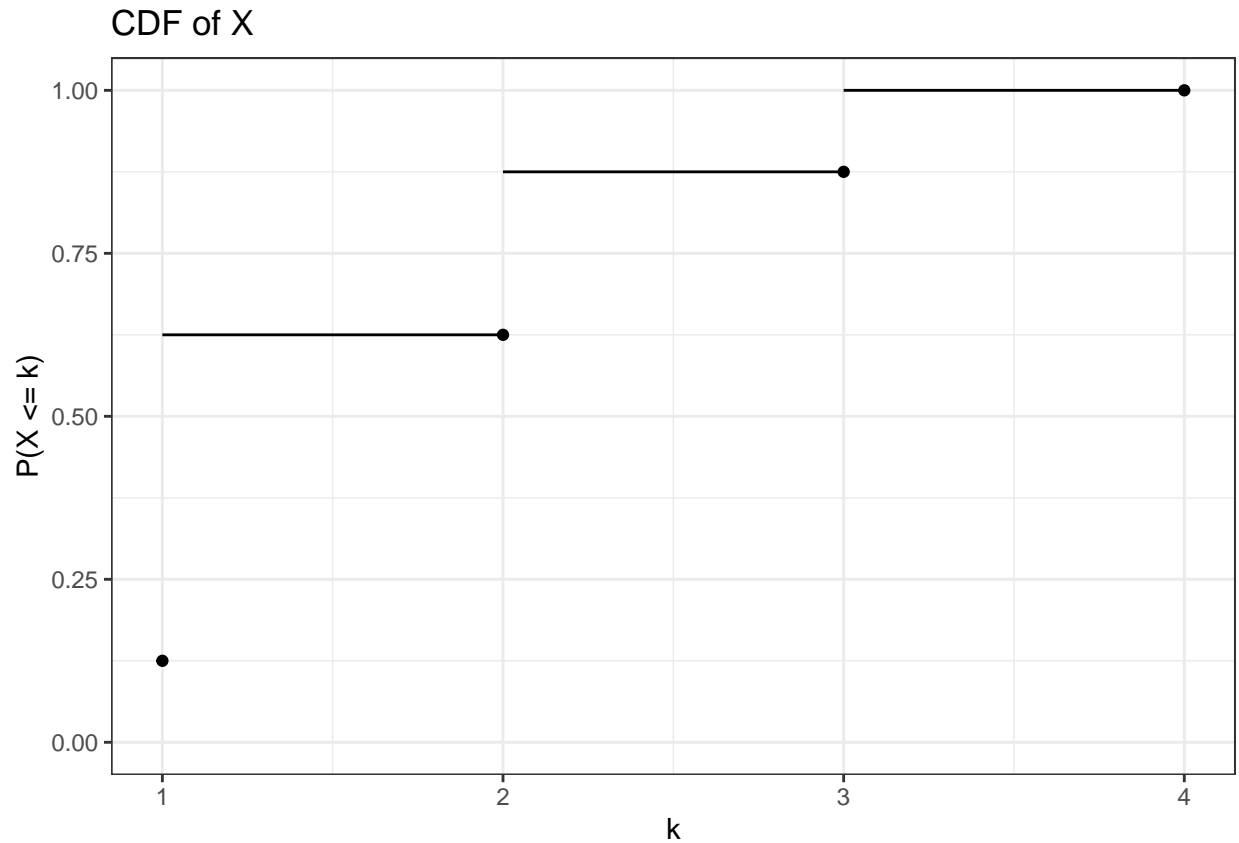
The Cumulative Distribution Function  $F_X(x)$  of  $X$  is as follows

$$F_X(1) = P(X \leq 1) = 0.125$$

$$F_X(2) = P(X \leq 2) = 0.625$$

$$F_X(3) = P(X \leq 3) = 0.875$$

$$F_X(4) = P(X \leq 4) = 1$$



c.

$$E_X[X] = \sum_{k=1}^4 k * P(X = k) = (1 * 0.125) + (2 * 0.5) + (3 * 0.25) + (4 * 0.125) = 2.375$$

d.

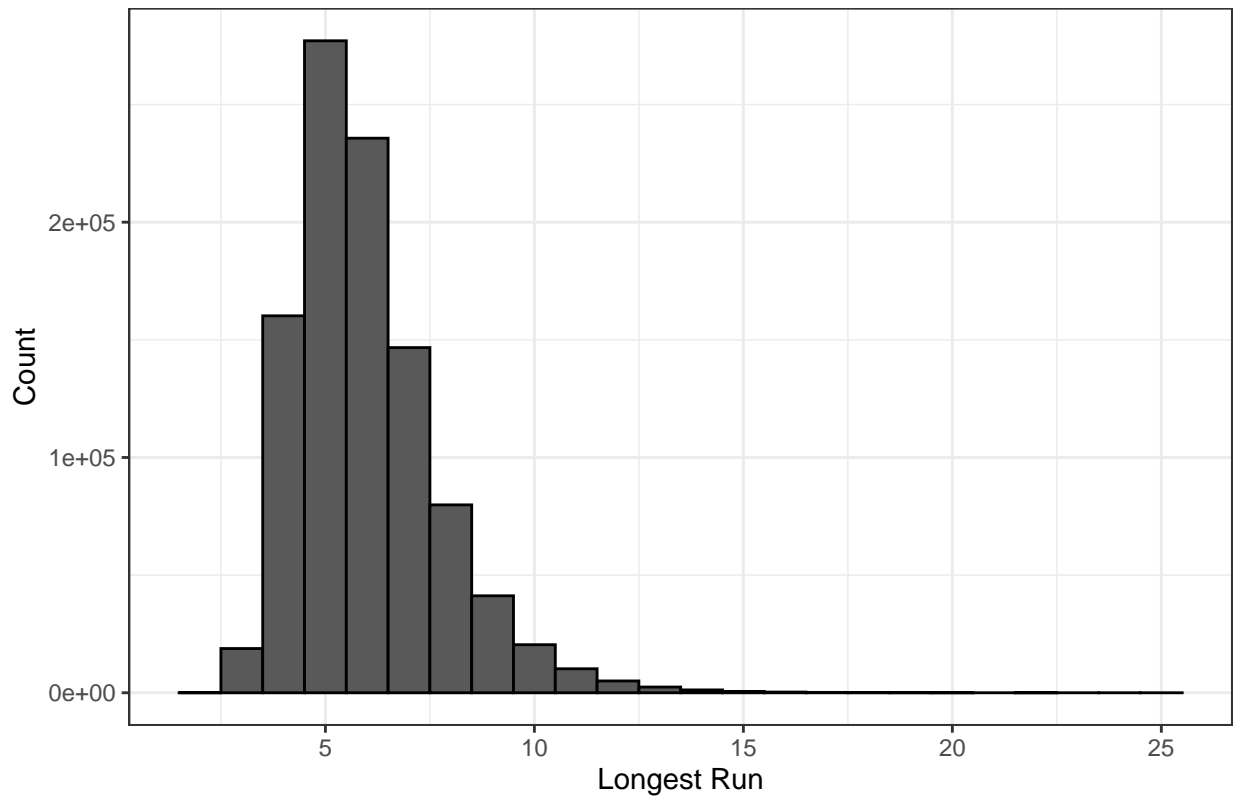
The following function returns the longest run of a value in a sequence.

```
# function that returns longest run in a sequence of two items
longestrun <- function(seq){
  currentrun <- 1
  longestrun <-1
  for(i in 1:(length(seq)-1)){
    if(seq[i+1] == seq[i]){
      currentrun <- currentrun + 1
      if(currentrun > longestrun) longestrun <- currentrun
    }
    else currentrun <- 1
  }
  return(longestrun)
}
```

We can simulate 1 million sequences of 50 coin flips to approximate the distribution of the longest run in 50 tosses and the expected value. We do this by the following:

```
set.seed(39)
sim_data <- unlist(lapply(1:1000000,
                        function(x)
                          longestrun(sample(c("H", "T"),
                                             size = 50,
                                             replace = TRUE,
                                             prob = c(0.5, 0.5))))))
sim_df <- data.frame(sim_data)
```

Distribution of Longest Runs for 1 Million Simulated 50 Tosses



```
## [1] 5.977351
```

Given the 1 million simulated 50 tosses, we receive the distribution shown above. We may find an estimate of the expected value by taking the mean of all the tosses, which we find to be 5.977351.

## 2. Textbook Exercises Chapter 1

### 1.38

Prove the following statements.

**a. If  $P(B) = 1$ , then  $P(A|B) = P(A)$  for any  $A$ .**

Given  $P(A|B) = \frac{P(A \cap B)}{P(B)}$ , then

$$P(A|B) = \frac{P(A \cap B)}{1} = P(A \cap B).$$

If  $A, B \subseteq S$ , where  $S$  is the sample space and  $P(B) = 1$ , then  $B = S$ . Since  $B = S$  and  $A \subseteq S$ ,  $A \subseteq B$ . Thus,  $P(A \cap S) = P(A \cap B) = P(A)$ . This means that

$$P(A|B) = P(A).$$

**b. If  $A \subset B$ , then  $P(B|A) = 1$  and  $P(A|B) = P(A)/P(B)$**

If  $A \subset B$ ,  $P(A \cap B) = P(A)$  and thus

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

and

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)}.$$

**c. If  $A$  and  $B$  are mutually exclusive, then ...**

If  $A$  and  $B$  are mutually exclusive, then  $P(A \cap B) = 0$  and

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B).$$

Additionally, given the mutual exclusivity of  $A$  and  $B$ ,  $P(A \cap (A \cup B)) = P(A)$ . Thus

$$P(A|A \cup B) = \frac{P(A)}{P(A) + P(B)}$$

**d.  $P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C)$**

Given

$$P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)},$$

we can rearrange this equality and solve for  $P(A \cap B \cap C)$ , showing

$$P(A \cap B \cap C) = P(A|B \cap C)P(B \cap C) = P(A|B \cap C)P(B|C)P(C)$$

### 1.39

A pair of events  $A$  and  $B$  cannot be simultaneously *mutually exclusive* and *independent*. Prove that if  $P(A) > 0$  and  $P(B) > 0$ , then:

**a. If  $A$  and  $B$  are mutually exclusive, they cannot be independent.**

If  $A$  and  $B$  are mutually exclusive and  $P(A) > 0$  and  $P(B) > 0$ , then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B).$$

Since  $P(A \cap B) = 0$  and  $P(A) > 0$  and  $P(B) > 0$ , the definition of independent events does not hold and

$$P(A \cap B) \neq P(A)P(B)$$

since  $P(A)P(B)$  should be non-zero.

**b. If  $A$  and  $B$  are independent, they cannot be mutually exclusive.**

If  $A$  and  $B$  are independent and  $P(A) > 0$  and  $P(B) > 0$ , then

$$P(A \cap B) = P(A)P(B)$$

is non-zero. Hence

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

If the events were to be mutually exclusive, then  $P(A \cup B) = P(A) + P(B)$ , but since they are independent this does not hold.

## 1.51

We can model  $X$  using the binomial distribution with  $n = 4$  trials and probability of success (choosing a defective microwave oven)  $p = 5/30 = 1/6$ . To find the PMF, we just find the probabilities for  $k = 0, 1, 2, 3, 4$  and for the CDF we take the cumulative probabilities. The binomial distribution is given by

$$P(X = k) = \binom{4}{k} p^k (1 - p)^{4-k}.$$

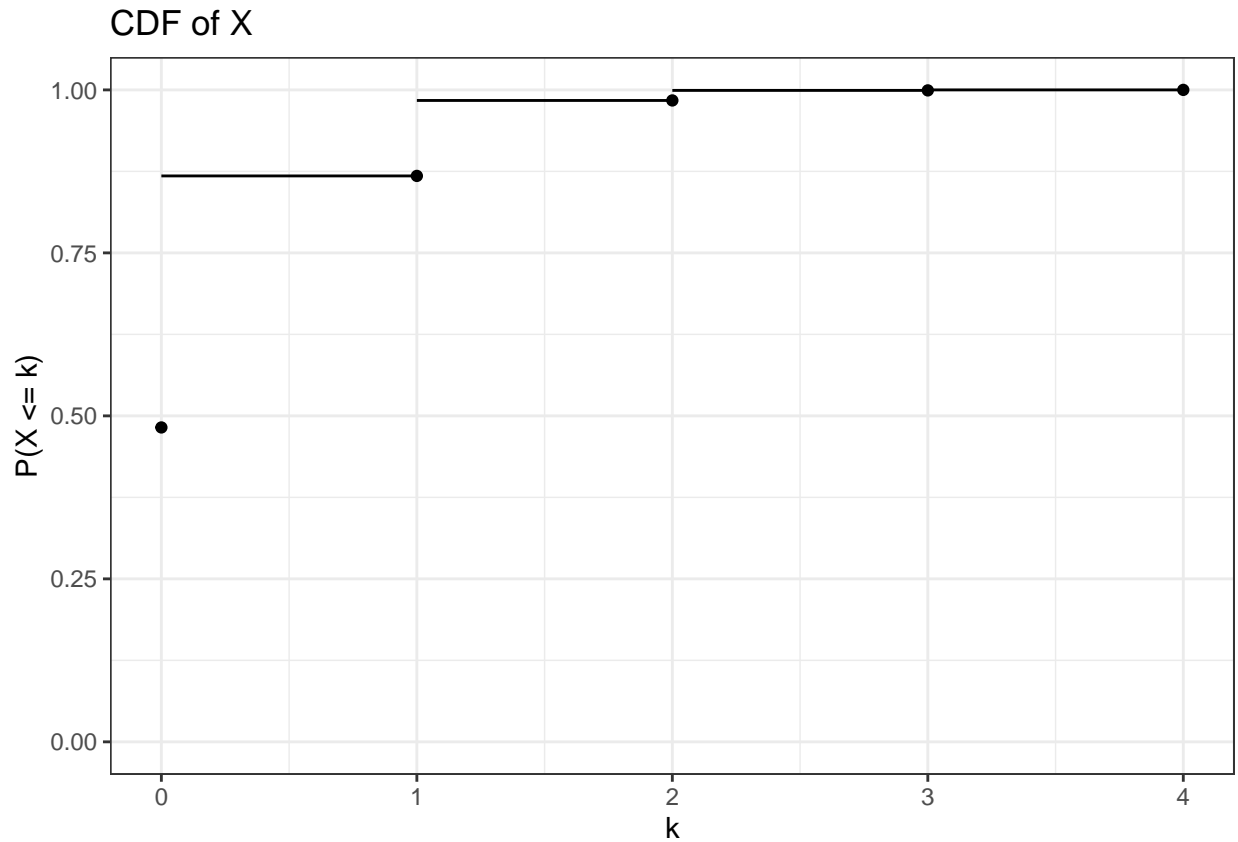
We use the built in `dbinom` and `pbinom` R functions to calculate the PMF and CDF.

Table 2: PMF

k	probs
0	0.4822531
1	0.3858025
2	0.1157407
3	0.0154321
4	0.0007716

Table 3: CDF

k	probs
0	0.4822531
1	0.8680556
2	0.9837963
3	0.9992284
4	1.0000000



## 1.52

Let  $X$  be a continuous random variable with pmf  $f(x)$  and cdf  $F(x)$ . For a fixed number  $x_0$ , define the function

$$g(x) = \begin{cases} f(x)/[1 - F(x_0)] & x \geq x_0 \\ 0 & x < x_0 \end{cases}.$$

Prove that  $g(x)$  is a pmf. (Assume that  $F(x_0) < 1$ .)

For a function to be a pmf, the integral across its domain must sum to 1 and the function must be nonnegative for all values in its domain. We prove this for  $g(x)$ .

Given  $F(x)$  is

$$F(x) = \int_{-\infty}^x = \frac{f(t)}{1 - F(x_0)} dt,$$

$\frac{1}{1-F(x_0)}$  is a constant and

$$F(x) = \int_{-\infty}^{x_0} 0 dt + \int_{x_0}^x \frac{f(t)}{1 - F(x_0)} dt.$$

If we take the limit as  $w \rightarrow \infty$ , the integral becomes

$$\lim_{w \rightarrow \infty} F(w) = \frac{1}{1 - F(x_0)} \int_{x_0}^w f(t) dt.$$

This then becomes

$$= \frac{1}{1 - F(x_0)} (\lim_{w \rightarrow \infty} F(w) - F(x_0)).$$

For a cdf  $F(x)$ , as  $x \rightarrow \infty$ ,  $F(x) \rightarrow 1$ . Thus,  $\lim_{w \rightarrow \infty} F(w) = 1$  and

$$\frac{1}{1 - F(x_0)} (\lim_{w \rightarrow \infty} F(w) - F(x_0)) = \frac{1}{1 - F(x_0)} (1 - F(x_0)) = 1.$$

Now, the  $g(x)$  must be nonnegative for all values in its domain. Given  $f(x)$  is a pmf,  $F(x_0) < 1$ , and  $g(x) = 0$  for  $x < x_0$ ,  $g(x)$  is nonnegative for all values in its domain.  $g(x)$  is a pmf.

## 1.54 part b

Determine the value of  $c$  that makes  $f(x)$  a PDF for

$$f(x) = ce^{-|x|}, \quad -\infty < x < \infty.$$

For all  $x$ ,  $f(x)$  will always be positive. Thus, we must find a value of  $c$  such that the integral of this function across its domain sums to 1.

$$\int_{-\infty}^{\infty} ce^{-|x|} dx = 1$$

First we should take the integral.

$$\begin{aligned} \int_{-\infty}^{\infty} ce^{-|x|} dx &= \int_{-\infty}^0 ce^{-(-x)} dx + \int_0^{\infty} ce^{-x} dx \\ &= \lim_{i \rightarrow -\infty} \int_i^0 ce^x dx + \lim_{j \rightarrow \infty} \int_0^j ce^{-x} dx \\ &= c(e^0 - \lim_{i \rightarrow -\infty} e^i) + c(\lim_{j \rightarrow \infty} -e^{-j} - (-e^0)) = c(1) + c(1) = 2c \end{aligned}$$

Then solve for  $c$



$$2c = 1 \Rightarrow c = \frac{1}{2}$$

$c$  must be  $\frac{1}{2}$  for  $f(x)$  to be a PDF.