

# HW4

Zunqiu Wang

9/16/2021

Q1 a

```
x <- 1:6
y <- 1:6
names(x) <- x
names(y) <- y
pair <- function(x,y) {
  z <- paste(x,y)
  return(z)
}
space <- outer(x, y, pair)

result.pair <- c()

result <- function(df) {
  for (i in 1:nrow(df)) {
    for (j in 1:ncol(df)) {
      if (j == i+1) {
        result.pair[i] <- df[i,j]
      }
    }
  }
  return(result.pair)
}
result(space)
```

```
## [1] "1 2" "2 3" "3 4" "4 5" "5 6"
```

```
length(result(space))
```

```
## [1] 5
```

```
prob <- length(result(space)) / (nrow(space) * ncol(space))
prob
```

```
## [1] 0.1388889
```

$$P(\text{sequential pair}) = \frac{5}{36}$$

Q1 b

$$P(A) = P(\text{hitting inner circle}) = \frac{2}{3}$$

$$P(B) = P(\text{hitting bulls eye}) = 5\%$$

$$P(A|B) = 1$$

$$P(A \cap B|A) = \frac{P(B)P(A|B)}{P(A)} = \frac{5\% * 1}{\frac{2}{3}} = 0.075$$

Q1 c

$$P(T = +|D) = 0.95$$

$$P(D) = 0.001$$

$$P(T = +|ND) = 0.05$$

$$P(ND) = 1 - P(D) = 1 - 0.001 = 0.999$$

$$P(D|T = +) = \frac{P(T = +|D)P(D)}{P(T = +)} = \frac{0.95 * 0.001}{P(T = +|D)P(D) + P(T = +|ND)P(ND)} = \frac{0.95 * 0.001}{0.95 * 0.001 + 0.05 * 0.999} = 0.019$$

```
sensitivity=0.95
Pd=0.001
FPR=0.05
Prob_D <- function(sensitivity,Pd,FPR) {
  return((sensitivity*Pd)/((sensitivity*Pd)+(FPR*(1-Pd))))
}
Prob_D(sensitivity,Pd,FPR)
```

```
## [1] 0.01866405
```

Q1 e

$$P(D) = 0.0001$$

$$P(D|T = +) = \frac{P(T = +|D)P(D)}{P(T = +)} = \frac{0.95 * 0.0001}{P(T = +|D)P(D) + P(T = +|ND)P(ND)} = \frac{0.95 * 0.0001}{0.95 * 0.0001 + 0.05 * 0.9999} = 0.0019$$

```
sensitivity=0.95
Pd=0.0001
FPR=0.05
Prob_D <- function(sensitivity,Pd,FPR) {
  return((sensitivity*Pd)/((sensitivity*Pd)+(FPR*(1-Pd))))
}
Prob_D(sensitivity,Pd,FPR)
```

```
## [1] 0.001896586
```

Q1 e

```
# # The probability of chance to have disease is proportional
# to total population so testing rare disease largely depends
# on sample of population. In reality, it is challenging to get
# everyone tested, and to keep false positive rate as low as possible is hard to achieve.
```

Q2 a

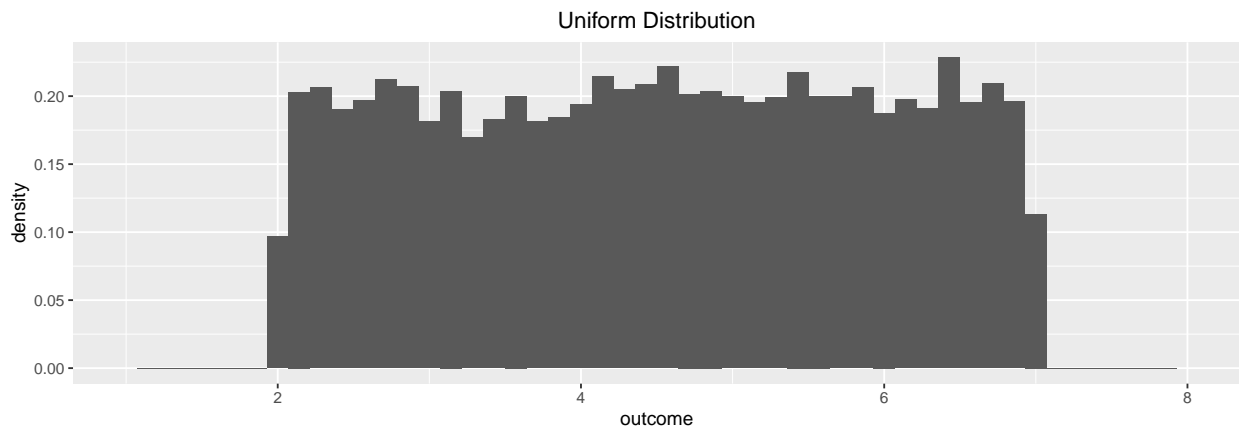
```
set.seed(1234)
die1 <- 1:20
out <- sample(die1, 10000, replace=TRUE) # generate distribution
length(which(out<=10)) # get the count
```

```
## [1] 5007
```

```
# sum(out<=10)
```

Q2 b

```
library(ggplot2)
unif.distr <- runif(10000,2,7)
ggplot(data.frame(unif.distr), aes(x=unif.distr)) +
  geom_histogram(aes(y=..density..), bins = 50) +
  xlim(1, 8) +
  labs(title = "Uniform Distribution", x = "outcome", y = "density") +
  theme(plot.title = element_text(hjust = 0.5))
```



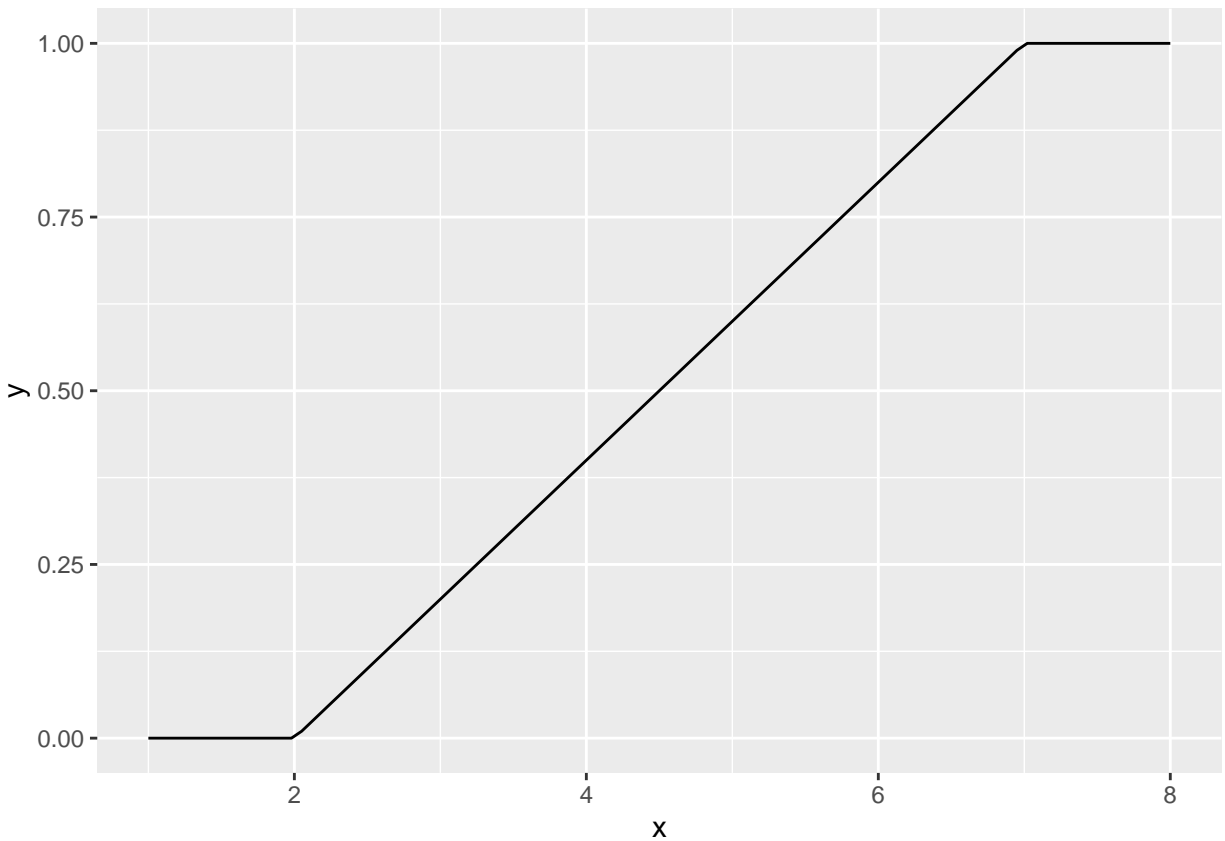
Q2 c

```
uniform.fun.pdf <- function(x) {
  ifelse(x >= 2 & x <= 7, 1, 0)
}
dunif(1:8,2,7) # equivalent using built-in function to get pdf
```

```
## [1] 0.0 0.2 0.2 0.2 0.2 0.2 0.2 0.0
```

Q2 d

```
uniform.fun.cdf <- function(x) {
  ifelse(x >= 2 & x <= 7, (x-2)/(7-2), ifelse(x < 2, 0, 1))
}
ggplot(data.frame(x=c(1, 8)), aes(x)) +
  stat_function(fun=uniform.fun.cdf)
```



```
prob.unif.cdf <- uniform.fun.cdf(3.2) - uniform.fun.cdf(1.5)
punif(3.2,2,7)-punif(1.5,2,7) #equivalent using built in function to calculate cdf
```

```
## [1] 0.24
```

Q3 a

```
pbinom(500, 10000, 0.05)
```

```
## [1] 0.511895
```

```
set.seed(1234)
die1 <- 1:20
out <- sample(die1, 10000, replace=TRUE)
n <- sum(out==20)
```

Q3 b

```
set.seed(1111)
rbinom(1,100,1/100) # equivalent
```

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

Q3 c

```
drop.2.more.prob <- 1-ppois(1,lambda = 1) # two or more times  
ppois(1,1, lower.tail = FALSE) # equivalent for upper tail
```

```
## [1] 0.2642411
```

Q3 d

```
mean <- 70  
sd <- 10  
upper.85.prob <- 1-pnorm(85, mean, sd) # 1 minus lower tail to get upper tail  
between.50.60.prob <- pnorm(60, mean, sd) - pnorm(50, mean, sd)  
# difference (cumulative probability) between a range
```