Problem 22, Ross, Probability Models

One way to investigate the number of trials that need to be performed until the same outcome occurs k consecutive times is to solve the probability problem proposed in Problem 22. Another way to study this is to write an R program and run a simulation.

In the discussion about the digits of π being uniformly distributed, below the statement of the problem, there is a hypotheses test that is discussed. Some questions that come up from this discussion are

1. The author states that that $SD(N)$ is approximatedly $E[N]$. Is this in fact the case?
2. Related to this one might wonder what the distribution of N looks like?
3. Following the issue of what the distribution of N, is using a standarization argument appropriate?

Run the following R program to investigate these issues. Also consider the following problems:

1. Flip a fair coin many time, what is the distribution of the number of performaces to get 3 heads or 3 tails in a row? Is 10 flips consistent with what is expected?
2. Flip a fair coin many time, what is the distribution of the number of performaces to get 5 heads or 5 tails in a row? Is 50 flips consistent with what is expected?
3. Roll a fair die many time, what is the distribution of the number of performaces to get 2 of the same number of dots in a row? Is 50 rolls consistent with what is expected?

Plot the simuated distribution of N for each of the above.

Considering the distribution of N, is the using standarization a reasonable way to test whether a specific number of rolls is expected?

```
# Simulation of Problem 22, Ross Probability Models

p <- 1          # number of equally likely values (0,1,...,p)
m <- p+1        # number of possible outcomes
k <- 3          # number of values in a run

B <- 1000

N <- numeric(B) # vector of simulated values
X <- matrix(NA,nrow=B,ncol=k)
                # matrix of runs

for(i in 1:B){
x <- sample(0:p,k,replace=TRUE)
#y <- x # used to check program, N = length(y)
N[i] <- k       # count length of sequence
repeat{
        u <- unique(x)          # check if all the same
        if(length(u)==1){
```

```
                X[i,] <- x      # store final run
                break           # if length is 1, all the same, stop
        }
        new <- sample(0:p,1,replace=TRUE)       # new last value
        x <- c(x[2:k],new)      # move window
        #y <- c(y,new)
        N[i] <- N[i] + 1        # incrment count
}
}

N       # length of sequence in ending with a run of k
x       # the run
#y      # sequence
#length(y)

# Plot the distribution of N.

hist(N)
N.mean <- mean(N)
N.mean
N.var <- var(N)
N.var
N.sd <- sqrt(N.var)
N.sd

# For p = 1, m = 2, k = 3, test if N of 10 is consistent with
# E[N] = (m^k - 1)/(m - 1)

value <- 10

N.E <- (m^k - 1)/(m - 1)
N.SD <- N.E      # according to the book

z = (value - N.E)/N.SD
z

p_value <- 2*(1-pnorm(abs(z)))
p_value

z.obs <- (value - N.mean)/N.sd
z.obs

p_value.obs <- 2*(1-pnorm(abs(z.obs)))
p_value.obs
```