DATA 605 Week 7 Homework

Ilya Kats

October 15, 2017

Problem 1

Let $X_1, X_2, ..., X_n$ be n mutually independent random variables, each of which is uniformly distributed on the integers from 1 to k. Let Y denote the minimum of the X_i's. Find the distribution of Y.

SOLUTION

Number of possible combinations of X_i's is k^n (choosing n values out of k options with replacement).

Consider number of combinations with at least one 1. It is equal to all combinations (k^n) minus all combinations with values between 2 and k ($(k-1)^n$). So $P(Y=1) = \frac{k^n-(k-1)^n}{k^n}$.

Consider number of combinations with at least one 2 and no 1. It is euqal to all combinations (k^n) minus all combinations with at least one 1 (see above: $k^n-(k-1)^n$) and minus all combinations with values between 3 and k ($(k-2)^n$). So $P(Y=2) = \frac{k^n-(k^n-(k-1)^n)-(k-2)^n}{k^n} = \frac{k^n-k^n+(k-1)^n-(k-2)^n}{k^n} = \frac{(k-1)^n-(k-2)^n}{k^n}$.

Similarly considering combinations without 1 or 2 and with at least one 3,

$$\begin{aligned}
P(Y=3) &= \frac{k^n-(k^n-(k-1)^n)-((k-1)^n-(k-2)^n)-(k-3)^n}{k^n} \\
&= \frac{k^n-k^n+(k-1)^n-(k-1)^n+(k-2)^n-(k-3)^n}{k^n} \\
&= \frac{(k-2)^n-(k-3)^n}{k^n}
\end{aligned}$$

.

More generally, we can see that $P(Y=a) = \frac{(k-a+1)^n-(k-a)^n}{k^n}$.

SIMULATION

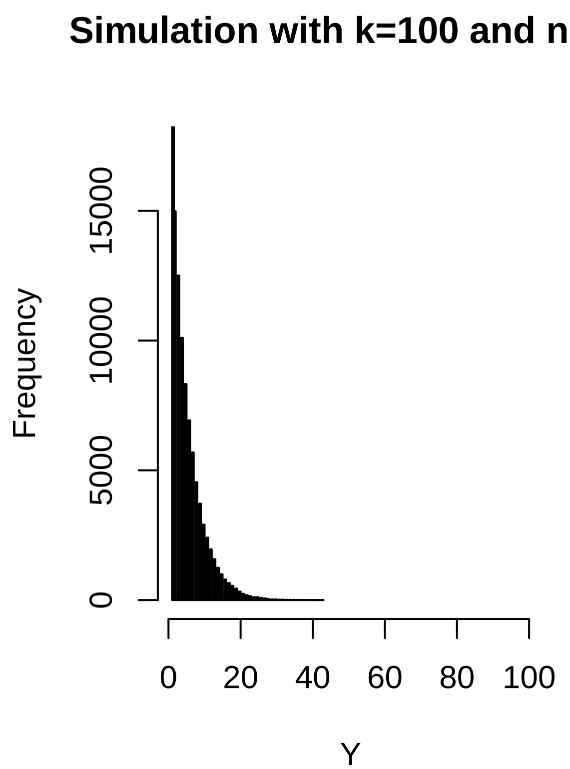
Set up a function to run simulated trials.

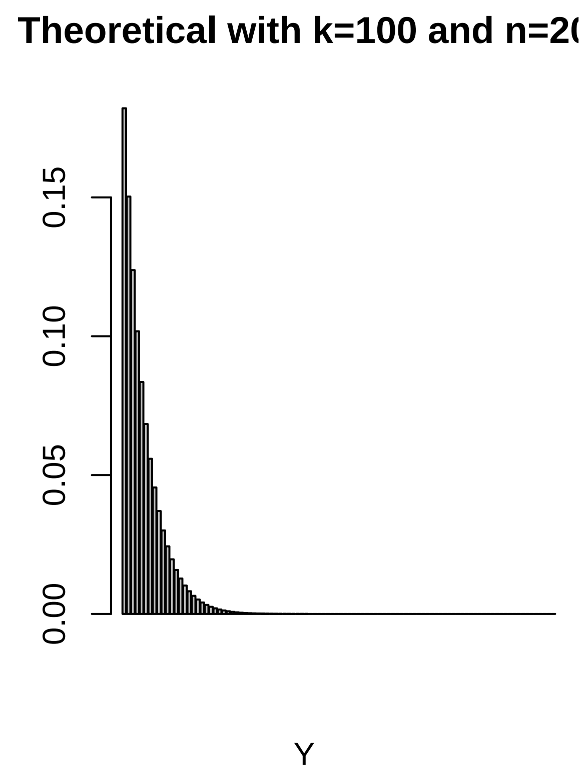
```r
problem1sim <- function(k,n,trials=100000) {
  Y<-rep(0,trials)
  for (i in 1:trials) {
    x<-sample.int(k,size=n,replace=TRUE)
    Y[i]<-min(x)
  }
  return(Y)
}
```

Plot distribution of simulated trials and theoretical probability distribution for several values of k and n.

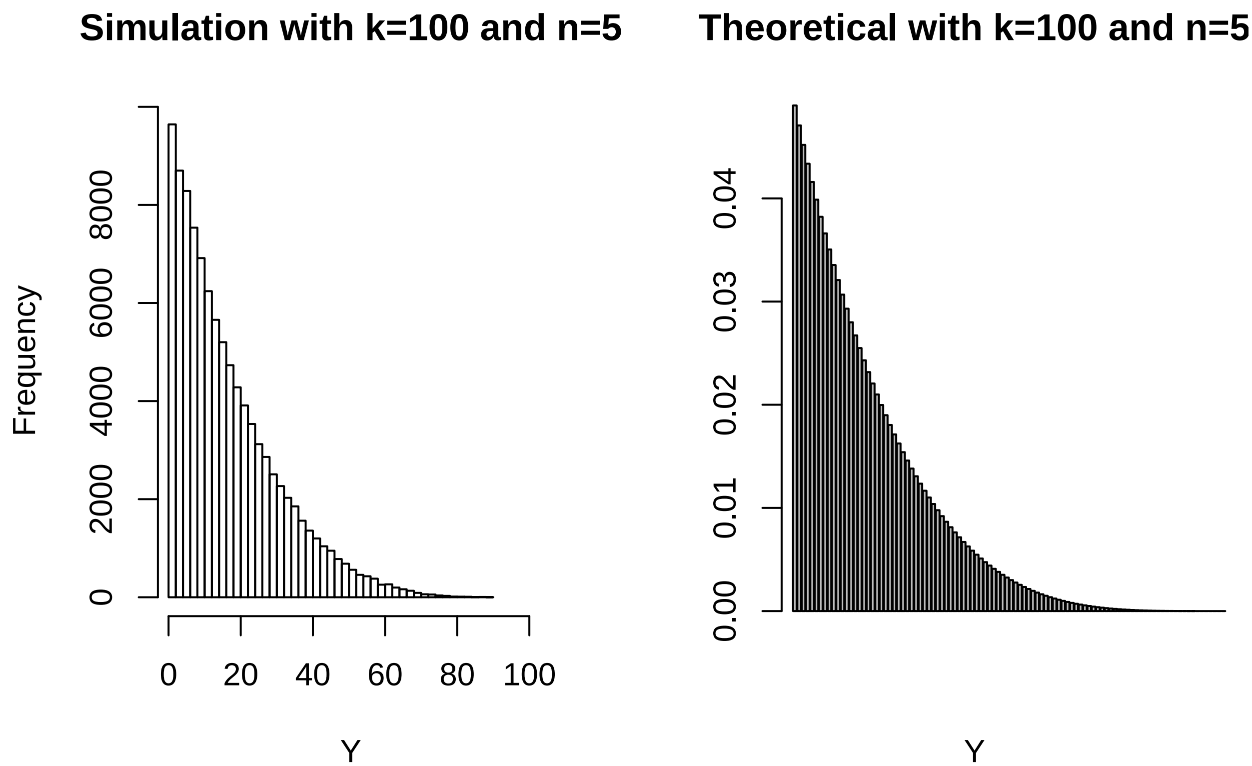
```r
# Run 1
par(mfrow=c(1,2))
k<-100
n<-20
hist(problem1sim(k,n),breaks=60,
```

```r
     main=paste("Simulation with k=",k," and n=",n,sep=""),
     xlab="Y",xlim=c(1,k))
pY<-((k-1:k+1)^n-(k-1:k)^n)/k^n
barplot(pY,main=paste("Theoretical with k=",k," and n=",n,sep=""),
        xlab="Y",xlim=c(1,k))
```

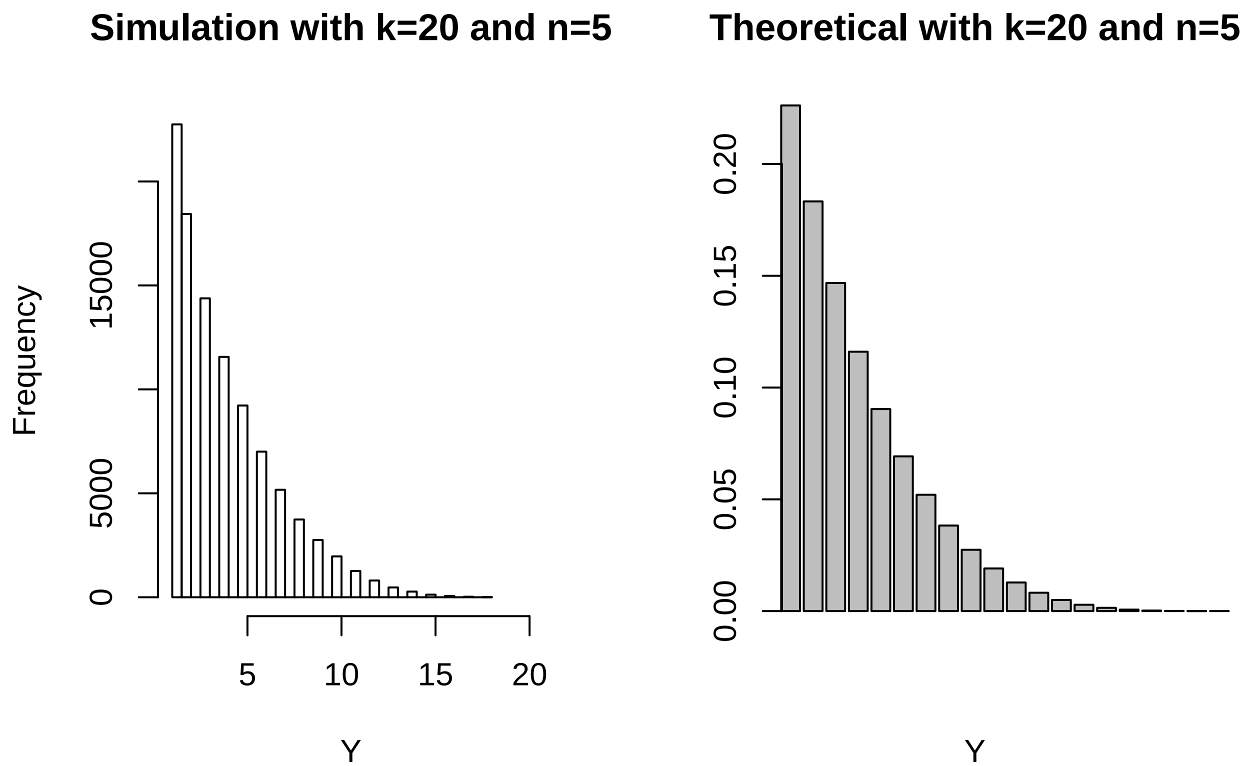




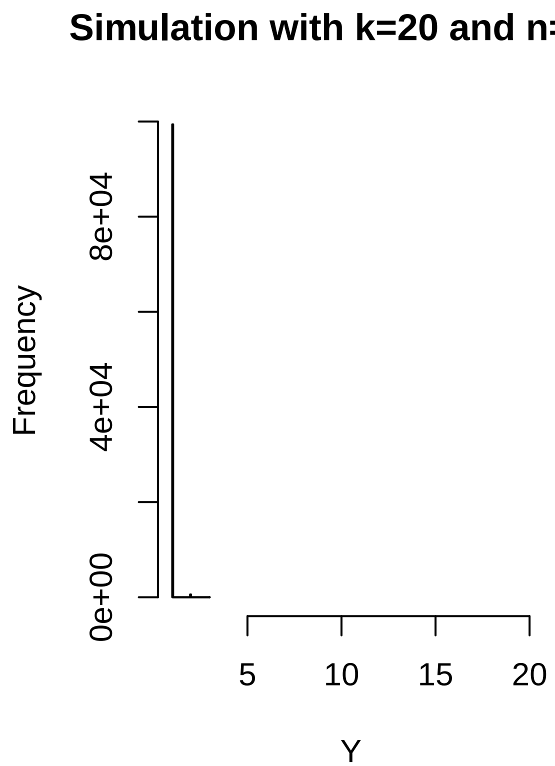

```r
# Run 2
par(mfrow=c(1,2))
k<-100
n<-5
hist(problem1sim(k,n),breaks=60,
     main=paste("Simulation with k=",k," and n=",n,sep=""),
     xlab="Y",xlim=c(1,k))
pY<-((k-1:k+1)^n-(k-1:k)^n)/k^n
barplot(pY,main=paste("Theoretical with k=",k," and n=",n,sep=""),
        xlab="Y",xlim=c(1,k))
```

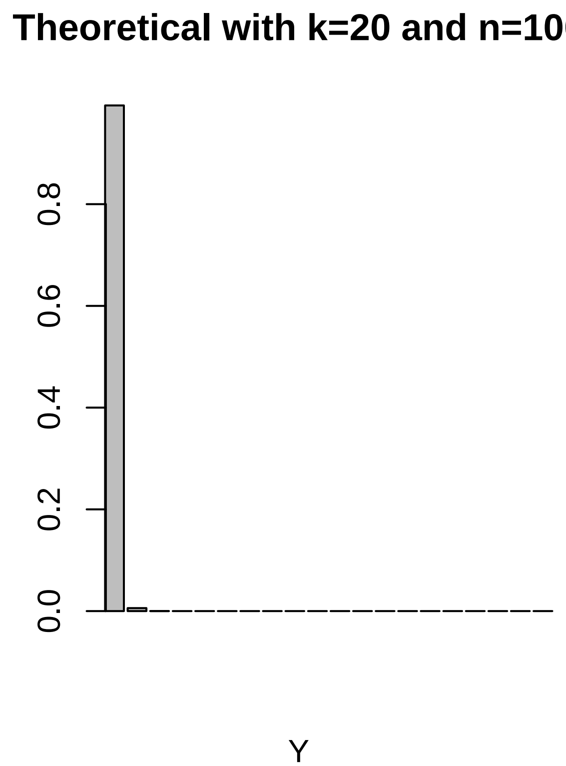



```r
# Run 3
par(mfrow=c(1,2))
k<-20
n<-5
hist(problem1sim(k,n),breaks=60,
     main=paste("Simulation with k=",k," and n=",n,sep=""),
     xlab="Y",xlim=c(1,k))
pY<-((k-1:k+1)^n-(k-1:k)^n)/k^n
barplot(pY,main=paste("Theoretical with k=",k," and n=",n,sep=""),
        xlab="Y",xlim=c(1,k))
```




```r
# Run 4
par(mfrow=c(1,2))
k<-20
n<-100
hist(problem1sim(k,n),breaks=60,
     main=paste("Simulation with k=",k," and n=",n,sep=""),
     xlab="Y",xlim=c(1,k))
pY<-((k-1:k+1)^n-(k-1:k)^n)/k^n
barplot(pY,main=paste("Theoretical with k=",k," and n=",n,sep=""),
        xlab="Y",xlim=c(1,k))
```





Problem 2

Your organization owns a copier (future lawyers, etc.) or MRI (future doctors). This machine has a manufacturer's expected lifetime of 10 years. This means that we expect one failure every ten years. (Include the probability statements and R Code for each part.).

With one failure every ten years $p = 0.1$ and $q = 1 - p = 0.9$. In this scenario, a failure of the machine is considered *success* in probability distributions.

PART A

What is the probability that the machine will fail after 8 years? Provide also the expected value and standard deviation. Model as a geometric. (Hint: the probability is equivalent to not failing during the first 8 years?)

For geometric distribution, CDF $F_X(k) = P(X \leq k) = 1 - q^{k+1}$, where k is the number of failures before the first success (which is how R defines geometric distribution). Alternatively, $P(X > k) = 1 - P(X \leq k) = 1 - (1 - q^{k+1}) = q^{k+1}$. So for $k = 8$, $P(X > 8) = 0.9^9 \approx 0.3874$.

```
# Calculating P(X>8) using geometric distribution
pgeom(8, 0.1, lower.tail=FALSE)
```

```
## [1] 0.3874205
```

Expected number of years before the first machine failure is $E(X) = q/p = 0.9/0.1 = 9$.

Standard deviation $\sigma^2 = \sqrt{q/p^2} = \sqrt{0.9/0.1^2} \approx 9.4868$.

PART B

What is the probability that the machine will fail after 8 years? Provide also the expected value and standard deviation. Model as an exponential.

For exponential distribution, CDF $F_X(k) = P(X \leq k) = 1 - e^{-\lambda k}$, where λ is the rate parameter. For this example, $\lambda = 0.1$. $P(X > k) = 1 - P(X \leq k) = 1 - (1 - e^{-\lambda k}) = e^{-\lambda k}$. So for $k = 8$, $P(X > 8) = e^{-0.8} \approx 0.4493$.

```
# Calculating P(X>8) using exponential distribution
pexp(8, 0.1, lower.tail=FALSE)
```

```
## [1] 0.449329
```

Expected value is $E(X) = 1/\lambda = 1/0.1 = 10$.

Standard deviation $\sigma^2 = \sqrt{1/\lambda^2} = 1/\lambda = 10$.

PART C

What is the probability that the machine will fail after 8 years? Provide also the expected value and standard deviation. Model as a binomial. (Hint: 0 success in 8 years)

For binomial distribution, $P(X = k) = \binom{n}{k} p^k q^{n-k}$. Probability of a machine failure after 8 years is the same as probability of 0 *successes* after 8 trials. So for $k = 0$ and $n = 8$, $P(X = 0) = \binom{8}{0} 0.1^0 \times 0.9^{8-0} = 1 \times 1 \times 0.9^8 \approx 0.4305$.

```
# Calculating P(X=0) for n=8 using binomial distribution
pbinom(0,8,0.1,lower.tail=TRUE)
```

```
## [1] 0.4304672
```

Expected value and standard deviation will depend on number of years/trials tracked. Consider first 8 years.

Expected value $E(X) = np = 8 \times 0.1 = 0.8$.

Standard deviation $\sigma^2 = \sqrt{npq} = \sqrt{8 \times 0.1 \times 0.9} \approx 0.8485$.

PART D

What is the probability that the machine will fail after 8 years? Provide also the expected value and standard deviation. Model as a Poisson.

On average we observe $\lambda = 0.1$ machine failures per year. For Poisson distribution, $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$. Probabilty of a machine failure after 8 years is the same as probability of 0 *successess* after 8 intervals (similarly to the binomial distribution). $P(no\ failures\ in\ 8\ years) = P(X = 0)^8 = (\frac{e^{-0.1} \times 0.1^0}{0!})^8 = (e^{-0.1})^8 \approx 0.4493$

```
# Calculating P(X=0) for 8 intervals using Poisson distribution
ppois(0,0.1,lower.tail=TRUE)^8
```

```
## [1] 0.449329
```

For Poisson distribution, $E(X) = \sigma^2 = \lambda = 0.1$.