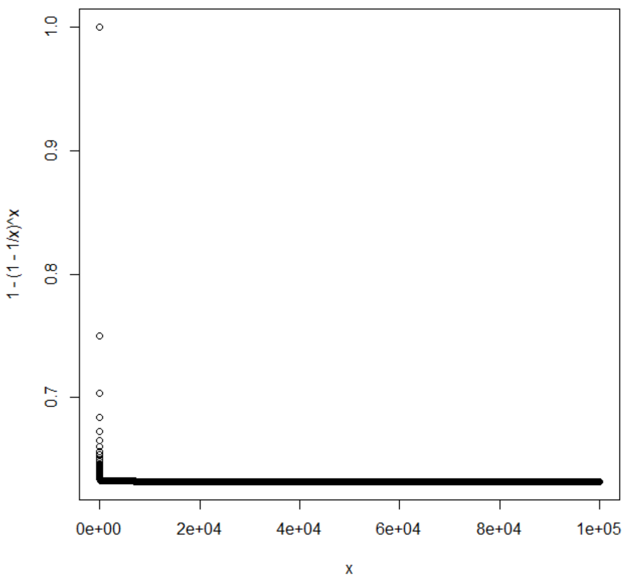
Problem 2

a) 1 – (1/n)
   The probability that the it is the jth observation is 1/n since there are n rows. Hence, the probability that it is not is 1 – 1/n
b) Since bootstrapping is sampling with replacement, the probability is still 1 – 1/n
c) Once again bootstrapping is sampling with replacement. The probability that it is not there is in an iteration is 1 – 1/n. In n iterations, the probability is (1 – 1/n)^n
d) Based on the above explanation, $1 - (1 - 0.2)^5 = 0.67$
e) $1 - (1 - 0.01)^{100} = 0.63$
f) $1 - (1 - 0.001)^{1000} = 0.63$
g) x = c(1:100000)
   plot(x, 1 - (1 - 1/x)^x)

   the value bottoms out at 0.63




```
h) > store=rep (NA , 10000)
> for (i in 1:10000) {
store[i]=sum(sample (1:100 , rep =TRUE)==4) >0
}
> mean(store)
```

I remember that as n gets larger, $(1 + (x/n))^n = e^x$

In this case, x =-1 and 1 – 1/e = 0.63

Problem 3

a) We divide the data set into n/k sets. We train a model on n/k – 1 sets while one of the sets will act as test set. This test set is altered continuously with each iteration.

b) In validation set method, we use a static test set. This can be highly variable. Also, in the validation set method, we have to divide a part of the sample just for testing purposes where as in K fold CV, we get to train our model on a larger sample

   K fold CV is better than LOOCV because it is computationally less intensive and also MSE has lesser variance since it is averaged. In the case of LOOCV, the variance of MSE is high because we are essentially training on the same training set. As the book has mentioned, k =5 to k = 10 often generate values that optimize the bias variance trade off