Q1. Using basic statistical properties of the variance, as well as single variable calculus, derive (5.6). In other words, prove that α given by (5.6) does indeed minimize $\mathrm{Var}(\alpha X + (1-\alpha)Y)$.

We have

$$\mathrm{Var}(\alpha X + (1-\alpha)Y) = \alpha^2\sigma_X^2 + (1-\alpha)^2\sigma_Y^2 + 2\alpha(1-\alpha)\sigma_{XY}.$$

We now take the fist derivative of $\mathrm{Var}(\alpha X + (1-\alpha)Y)$ *relative to* α *and we get*

$$\frac{\partial}{\partial\alpha}\mathrm{Var}(\alpha X + (1-\alpha)Y) = 2\alpha\sigma_X^2 - 2\sigma_Y^2 + 2\alpha\sigma_Y^2 + 2\sigma_{XY} - 4\alpha\sigma_{XY}.$$

We now seek critical points by equalling the last expression to 0,

$$2\alpha\sigma_X^2 - 2\sigma_Y^2 + 2\alpha\sigma_Y^2 + 2\sigma_{XY} - 4\alpha\sigma_{XY} = 0,$$

which implies that

$$\alpha = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}.$$

It remains to check that this point is in fact a minimum, this is equivalent to prove that the second derivative is positive,

$$\frac{\partial^2}{\partial\alpha^2}\mathrm{Var}(\alpha X + (1-\alpha)Y) = 2\sigma_X^2 + 2\sigma_Y^2 - 4\sigma_{XY} = 2\mathrm{Var}(X - Y) \geq 0.$$

–

Q2. We will now derive the probability that a given observation is part of a bootstrap sample. Suppose that we obtain a bootstrap sample from a set of n observations.

(a) What is the probability that the first bootstrap observation is not the jth observation from the original sample ? Justify your answer.

$1 - 1/n.$

(b) What is the probability that the second bootstrap observation is not the jth observation from the original sample ?

$1 - 1/n.$

(c) Argue that the probability that the jth observation is not in the bootstrap sample is $(1 - 1/n)^n$.

As bootstrapping sample with replacement, we have that the probability that the jth observation is not in the bootstrap sample is the product of the probabilities that each bootstrap observation is not the jth observation from the original sample

$$(1 - 1/n) \cdot \cdot \cdot (1 - 1/n) = (1 - 1/n)^n$$

as these probabilities are independant.

(d) When $n = 5$, what is the probability that the jth observation is in the bootstrap sample ?

We have

$$P\ (\textit{jth obs in bootstrap sample}) = 1 - (1 - 1/5)^5 = 0.672.$$

(e) When n = 100, what is the probability that the jth observation is in the bootstrap sample ?

We have

$$P\ (\textit{jth obs in bootstrap sample}) = 1 - (1 - 1/100)^{100} = 0.634.$$

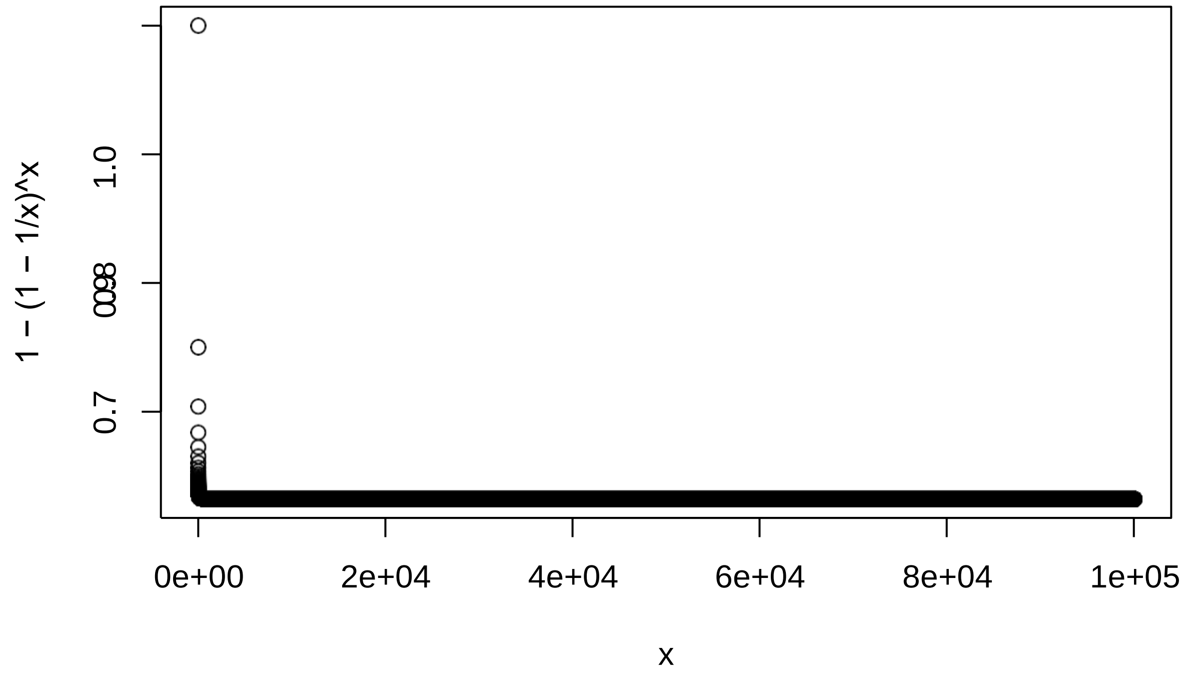
(f) When n = 10000, what is the probability that the jth observation is in the bootstrap sample ?

We have

$$P\ (\textit{jth obs in bootstrap sample}) = 1 - (1 - 1/10000)^{10000} = 0.632.$$

(g) Create a plot that displays, for each integer value of n from 1 to 100000, the probability that the jth observation is in the bootstrap sample. Comment on what you observe.

```
x <-1:100000
plot(x,1-(1-1/x)^x)
```



We may see that the plot quickly reaches an asymptote at about 0.632.

(h) We will now investigate numerically the probability that a bootstrap sample of size $n = 100$ contains the jth observation. Here $j = 4$. We repeatedly create bootstrap samples, and each time we record whether or not the fourth observation is contained in the bootstrap sample.

```
store <- rep(NA,10000) for (i
in1:10000) {
	store[i] <- sum(sample(1:100,rep =TRUE) ==4) >0
}
mean(store)
```

```
## [1] 0.635
```

Comment on the results obtained.

A known fact from calculus tells us that

$$\lim_{n \to \infty} (1 + x/n)^n = e^x.$$

If we apply this fact to our case, we get that the probability that a bootstrap sample of size n contains the jth observation converges to $1 - 1/e = 0.632$ as $n \to \infty$.

Q3. We now review k-fold cross-validation.

(a) Explain how k-fold cross-validation is implemented.

The k-fold cross validation is implemented by taking the n observations and randomly splitting it into k non-overlapping groups of length of (approximately) n/k. These groups acts as a validation set, and the remainder (of length n n/k) acts as a training set. The test error is then estimated by averaging the k resulting MSE estimates.

(b) What are the advantages and disadvantages of k-fold cross-validation relativeto: i.The

validation set approach ?

The validation set approach has two main drawbacks compared to k-fold cross-validation. First, the validation estimate of the test error rate can be highly variable (depending on precisely which observations are included in the training set and which observations are included in the validation set). Second, only a subset of the observations are used to fit the model. Since statistical methods tend to perform worse when trained on fewer observations, this suggests that the validation set error rate may tend to overestimate the test error ratefor the model fit on the entire data set.

ii.LOOCV ?

The LOOCV cross-validation approach is a special case of k-fold cross-validation in which $k = n$. This approach has two drawbacks compared to k-fold cross-validation. First, it requires fitting the potentially computationally expensive model n times compared to k-fold cross-validation which requires the model to be fitted only k times. Second, the LOOCV cross-validation approach may give approximately unbiased estimates of the test error, since each training set contains n 1 observations; however, this approach has higher variance than k-fold cross-validation (since we are averaging the outputs of n fitted models trained on an almost identical set of observations, these outputs are highly correlated, and the mean of highly correlated quantities has higher variance than less correlated ones). So, there is a bias-variance trade-off associated with the choice of k in k-fold cross-validation; typically using $k = 5$ or $k = 10$ yield test error rate estimates that suffer neither from excessively high bias nor from very high variance.

Q4. Suppose that we use some statistical learning method to make a prediction for the response Y for a particular value of the predictor X. Carefully describe how we might estimate the standard deviation of our prediction.

We may estimate the standard deviation of our prediction by using the bootstrap method. In this case, rather than obtaining new independant data sets from the population and fitting our model on those data sets, we instead obtain repeated random samples from the original data set. In this case, we perform sampling with replacement B times and then find the corresponding estimates and the standard deviation of those B estimates by using equation (5.8).