# EXERCISES – CONCEPTUAL

# CHAPTER _ 5

# RESAMPLING METHODS

1. Using basic statistical properties of the variance, as well as singlevariable calculus, derive (5.6). In other words, prove that α given by (5.6) does indeed minimize Var(αX + (1 − α)Y ).

For properties of variance – click here

5.1 Properties of variance →

i) $Var(X+Y) = Var(X) + Var(Y) + 2\,cov(X,Y)$

ii) $Var(aX+b) = a^2\,Var(X)$

iii) $Cov(a_1X+b_1, a_2Y+b_2) = a_1a_2\,Cov(X,Y)$

$$Var(\alpha X+(1-\alpha)Y) = Var(\alpha X) + Var((1-\alpha)Y) + 2\,Cov(\alpha X,(1-\alpha)Y)$$
$$= \alpha^2\,Var(X) + (1-\alpha)^2\,Var(Y) + 2.\alpha.(1-\alpha).\,Cov(X,Y)$$

for $Var(\alpha X+(1-\alpha)Y)$ to be min, $\dfrac{d(\alpha X+(1-\alpha)Y)}{d\alpha} = 0$

$$0 = 2\alpha\,Var(X) - 2(1-\alpha)\,Var(Y) + 2\left[(1-\alpha)Cov(X,Y) - \alpha\,Cov(X,Y)\right]$$

$$0 = \alpha\,Var(X) - Var(Y) + \alpha\,Var(Y) + Cov(X,Y) - 2\alpha\,Cov(X,Y)$$

$$-\left[\alpha\left(Var(X) + Var(Y) - 2\,cov(X,Y)\right)\right] = -(Var(Y) - Cov(X,Y))$$

$$\alpha = \frac{Var(Y) - cov(X,Y)}{Var(X) + Var(Y) - 2\,cov(X,Y)}$$

2. 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.

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

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

(d) When $n = 5$, what is the probability that the jth observation is in the bootstrap sample? (e) When $n = 100$, what is the probability that the jth observation is in the bootstrap sample?

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

f) When $n = 10,000$, what is the probability that the jth observation is in the bootstrap sample?

$\equiv$ $X \rightarrow$ original, $Z \rightarrow$ bootstrap.

a.) Probability of any observation to be selected in bootstrap is $\frac{1}{n}$. Since selection is based on performed with replacement.

$$P(z_1 = x_j) = \frac{1}{n}$$

$$P(z_1 \neq x_j) = 1 - \frac{1}{n} = \frac{n-1}{n}$$

b.) Since second is independent of the first.

$$P(z_2 \neq x_j) = 1 - \frac{1}{n} = \frac{n-1}{n}$$

c.) $P(x_j \notin Z) = P(x_j \neq z_1) * P(x_j \neq z_2) * P(x_j \neq z_3) \cdots P(x_j \neq z_n)$

$$= \left(1 - \frac{1}{n}\right) * \left(1 - \frac{1}{n}\right) \cdot \underline{\text{n times}} \left(1 - \frac{1}{n}\right)$$

$$= \left(1 - \frac{1}{n}\right)^n$$

d.) For $n = 5$,

$_j \notin Z) = \left(1 - \frac{1}{5}\right)^5 = (1 - 0.2)^5 = (0.8)^5 = 0.32768$

$$P(x_j \in Z_5) = 1 - 0.32768 = 0.67232$$

$\boxed{n = 100}$

$$P(X_j \in Z) = 1 - P(x_j \notin Z) = 1 - \left(1 - \frac{1}{100}\right)^{100} = 0.6339$$

$n = 1000$

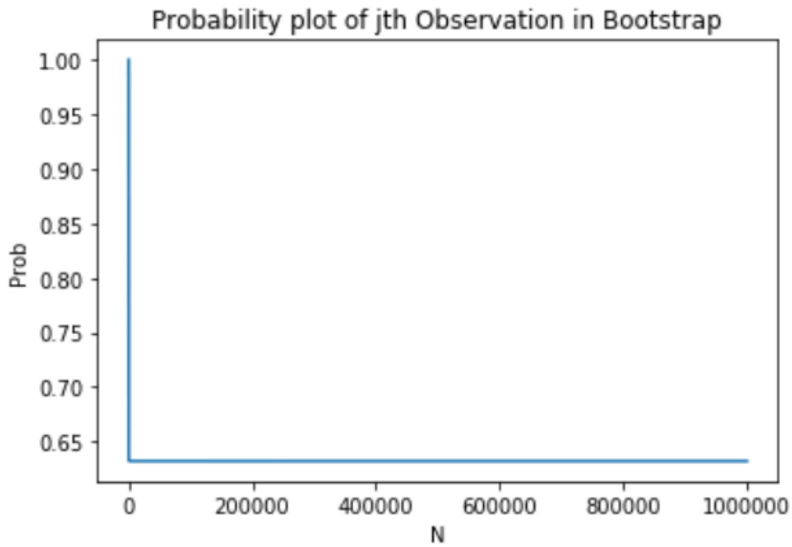
$$P(x_j \in Z) = 1 - P(x_j \notin Z) = 1 - \left(1 - \frac{1}{1000}\right)^{1000} = 0.6323$$

(g) Create a plot that displays, for each integer value of n from 1 to 100, 000, the probability that the jth observation is in the bootstrap sample. Comment on what you observe.

In [14]:
```
X = np.arange(1,1000001)
def get_prob(X):
    return 1 - ( 1 - 1/X)**X

prob = get_prob(X)
plt.plot(X,prob)
plt.xlabel('N')
plt.ylabel('Prob')
plt.title('Probability plot of jth Observation in Bootstrap')
```

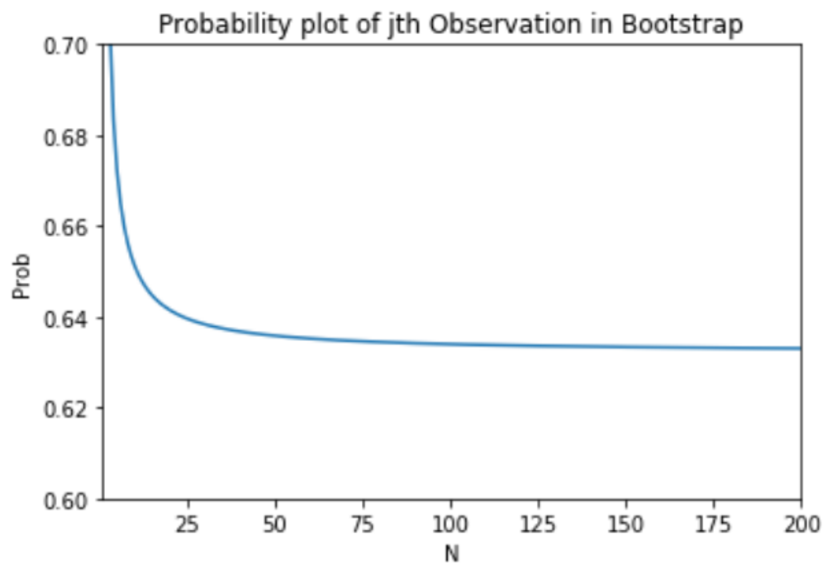
Out[14]:
```
Text(0.5, 1.0, 'Probability plot of jth Observation in Bootstrap')
```




**Although it may seem like its a sharp edge near x = 0, but zooming the graph, we can observe its a smooth curve.**

In [15]:
```
plt.plot(X,prob)
plt.ylim(0.6,0.7)
plt.xlim(1,200)
plt.xlabel('N')
plt.ylabel('Prob')
plt.title('Probability plot of jth Observation in Bootstrap')
```

Out[15]: `Text(0.5, 1.0, 'Probability plot of jth Observation in Bootstrap')`




(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.

In [36]:
```
num_samples = 1000
n = 100
count = 0
for i in range(num_samples):
    sample = np.random.randint(1,n+1,n)
    count += int(4 in sample)
print(count / num_samples)
```

```
0.611
```

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

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

**In k fold cross-validation, the whole observations are divided into k groups(folds), at given time, one fold is used for finding the error by a model which is trained on the remaining k-1 folds. This is repeated k times, such that each fold is once used as a validation set. The final error is the mean of the all the errors obtained from k folds.**

b) What are the advantages and disadvantages of k-fold crossvalidation relative to:

i. The validation set approach?

**Advantage –**
**Gives better estimate of test error as compared to validation set approach.**
**Validation set approach is highly variable, it depends on which the observations are seleceted in train an validation set, and hence repeating the same procedure with validation set approach may give us quite different results.**

**Disadvantage –**
**Computationally expensive, as a model is fitted k times, as compared to one time in validation set approach.**

ii. LOOCV?

**Advantage –**
**Computationally faster as compared to LOOCV, as model is fitted k times, whereas in LOOCV, it is fitted n times (n>k).**

**High variance in LOOCV.**

**Disadvantage –**
**LOOCV uses more observations for training than k fold, hence it results in less bias.**

4. 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 would use bootstrap for estimating the standard deviation of our prediction. Bootstrap selects random sample from the observations with repetition, and for each sample it would find the prediction. Using these observations than we would estimate the mean of predictions. From mean, we can find standard deviations of prediction.**