Report Project 2

Course code: IX1501
Date: 2023-10-09

Sushil Kc, sushilkc@kth.se
Sebastian Rone, sebgro@kth.se

Task 2: Bootstrapping for an estimation of probability

Summery

Task

We have a sample consisting of 10 observations as follows : 56, 101, 78, 67, 93, 87, 64, 72, 80, 69
The sample is considered to be a realization of 10 independent and identically distributed (i . i . d .) random
variables from an unknown probability distribution with unknown mean
μ . For given constraints a = −5 and b = 5 , we are interested in estimating the following probability :

In[ ]:=
$$p = P\left(a < \frac{\sum_{i=1}^{n} X_i}{n} - \mu < b\right)$$

Your task is as follows :

[Task 1] Explain how bootstrapping method can be used to estimate p . Include a pseudo code in your explanation .

[Task 2] Estimate p with bootstrapping method . (Hint : what is the most reasonable estimator of μ?Use it in the place of μ)

Result

Task1

1. We start with the original sample data.
2. We generated 1000 random bootstrap samples by sampling with replacement from the original data. In Mathematica we can use a function called RandomChoice.
3. For each bootstrap sample, we calculate the mean.
4. We store the means of all bootstrap samples.
5. Finally, we calculate the difference between the mean value of all the bootstrap samples and the original mean value to see if it’s value is between the designated interval.

With this method we can estimate the p, the probability that a random variable' s deviation from μ falls within[-5, 5] .

Task2

Using the bootstrapping method we get around 76 % probability that a random variable' s deviation from μ falls within[-5, 5] . First we did tests with one trial with 1000 number of samples in lists. We got slightly different probability each time ranging from 74 to 78%. Then did the same tests 10 times and took the average and we got around 76 % probability almost all the times.

Method

Code

In[33]:=
```
ClearAll["`*"]
```

To calculate the mean value (μ) of a sample, we simply add up all the observations in the sample and then divide by the number of observations . In our case, we have a sample consisting of 10 observations : 56, 101, 78, 67, 93, 87, 64, 72, 80, 69

To find the mean (μ) : $\mu = (56 + 101 + 78 + 67 + 93 + 87 + 64 + 72 + 80 + 69)/10$

$\mu = 767/10$

$\mu = 76.7$

So, the mean value of our sample is 76.7

In[34]:=
```
x = {56, 101, 78, 67, 93, 87, 64, 72, 80, 69};(** Our sample **)
```

In[35]:=
```
mainMean = N[Mean[x]] (** the mean value of our orginal sample**)
```

Out[35]=
```
76.7
```

In[36]:=
```
NumberOfSamples = 1000;
```

In[37]:=
```
sampleList = RandomChoice[x, {NumberOfSamples, 10}]
(**number of virtual samples out of the original sample.This method is
based on the principle of random sampling with replacement **)
```

Out[37]=
```
{{72, 67, 87, 87, 93, 101, 101, 78, 69, 67},
 {93, 80, 78, 56, 101, 78, 56, 101, 87, 56}, {67, 72, 64, 101, 56, 93, 56, 72, 72, 72},
 {101, 78, 56, 72, 101, 67, 64, 80, 64, 72}, ... 992 ...,
 {56, 93, 69, 87, 64, 56, 78, 56, 69, 72}, {69, 101, 72, 101, 78, 93, 87, 78, 80, 78},
 {67, 72, 80, 78, 56, 64, 69, 67, 78, 101}, {64, 93, 101, 78, 87, 64, 69, 101, 93, 101}}
```

Size in memory: 0.1 MB | Show more | Show all | Iconize | Store full expression in notebook

In[38]:=
```
totalMeanList = {};
```

In[39]:=
```
meanCount = 0;
```

In[40]:=
```
(**test withe one trial with 1000 number of samples in lists **)
```

In[41]:=
```
For[i = 1, i ≤ NumberOfSamples, i++,
 meanForASample = Mean[sampleList[[i]]];
 meandiff = meanForASample - mainMean;
 AppendTo[totalMeanList, meanForASample];
 If[meandiff > -5 && meandiff < 5, meanCount++]
]
```

In[42]:=
```
N[meanCount / NumberOfSamples] (**Avarage mean value of the virtual samples**)
```

Out[42]=
```
0.736
```

In[43]:=
```
prob = {};
```

In[44]:=

In[45]:=
```
(**tests with 10 trials with 1000 number of
 samples in lists and diffrent list every trails **)
```

In[46]:=
```
For[numT = 1, numT ≤ 100, numT++,
 meanCount = 0;
 sampleList = RandomChoice[x, {NumberOfSamples, 10}];
For[i = 1, i ≤ NumberOfSamples, i++,
   meanForASample = Mean[sampleList[[i]]];
   meandiff = meanForASample - mainMean;
   AppendTo[totalMeanList, meanForASample];
   If[meandiff > -5 && meandiff < 5, meanCount++]
  ] ×
AppendTo[prob, meanCount / NumberOfSamples];
]
```

In[47]:=
```
N[prob] (** Average mean value of the 10
 virtual samples with 1000 sample each in the lists  **)
```

Out[47]=
```
{0.755, 0.766, 0.776, 0.794, 0.788, 0.767, 0.731, 0.761, 0.755, 0.756, 0.772,
 0.764, 0.756, 0.766, 0.739, 0.755, 0.76, 0.758, 0.745, 0.753, 0.778, 0.752,
 0.748, 0.768, 0.765, 0.756, 0.783, 0.745, 0.752, 0.746, 0.75, 0.752, 0.758,
 0.74, 0.772, 0.763, 0.769, 0.757, 0.765, 0.766, 0.772, 0.759, 0.775, 0.764,
 0.775, 0.786, 0.744, 0.783, 0.765, 0.737, 0.754, 0.742, 0.773, 0.759, 0.756,
 0.764, 0.774, 0.79, 0.787, 0.768, 0.77, 0.777, 0.744, 0.785, 0.76, 0.761, 0.749,
 0.736, 0.739, 0.773, 0.757, 0.774, 0.787, 0.756, 0.754, 0.758, 0.758, 0.762,
 0.767, 0.781, 0.774, 0.779, 0.768, 0.769, 0.777, 0.761, 0.781, 0.776, 0.764,
 0.757, 0.764, 0.766, 0.773, 0.748, 0.776, 0.783, 0.768, 0.765, 0.775, 0.743}
```

In[48]:=
```
N[Total[prob] / 100] (** Average mean value**)
```

In[49]:=
```
N[Mean[prob]]
```

Out[49]=
```
0.76344
```