In the name of God

استاد : دکتر تیموری

دانشجو : توحید حقیقی سیس

شماره دانشجویی : 830598021

موضوع : تمرین نهم

تمرین اول :

12. To estimate θ, we generated 20 independent values having mean θ. If the successive values obtained were

102,	112,	131,	107,	114,	95,	133,	145,	139,	117
93,	111,	124,	122,	136,	141,	119,	122,	151,	143

how many additional random variables do you think we will have to generate if we want to be 99 percent certain that our final estimate of θ is correct to within ± 0.5?

بر اساس توضیحات کتاب به صورت زیر عمل میکنیم :

$$
\begin{aligned}
E[(\overline{X}-\theta)^2] &= \mathrm{Var}(\overline{X}) \quad (\text{since } E[\overline{X}]=\theta) \\
&= \mathrm{Var}\left(\frac{1}{n}\sum_{1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) \quad (\text{by independence}) \\
&= \frac{\sigma^2}{n} \quad (\text{since } \mathrm{Var}(X_i)=\sigma^2)
\end{aligned}
$$

Variance($S^۲$) = ۲۸۲٫۲۳۹۴۷۳۶۸

S = ۱۶٫۷۹۹

$\longrightarrow Z_{\alpha/۲} \times \frac{S}{\sqrt{n}} = ۰٫۵$

$\longrightarrow Z_{۵/۱۰۰} = ۲٫۵۷۵۸$

$\longrightarrow ۲٫۵۷۵۸ \times \frac{۱۶٫۷۹۹}{\sqrt{n}} = ۰٫۵$

$\longrightarrow \sqrt{n} = ۸۶٫۵۴۶۷$

$\boxed{\longrightarrow n = ۷۴۹۱}$

تمرین دوم :

13. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having unknown mean μ. For given constants $a < b$, we are interested in estimating $p = P\{a < \sum_{i=1}^{n} X_i/n - \mu < b\}$.

(a) Explain how we can use the bootstrap approach to estimate p.
(b) Estimate p if $n = 10$ and the values of the X_i are 56, 101, 78, 67, 93, 87, 64, 72, 80, and 69. Take $a = -5, b = 5$.

In the following three exercises $X_1, \ldots, X_n$ is a sample from a distribution whose variance is (the unknown) σ^2. We are planning to estimate σ^2 by the sample variance $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$, and we want to use the bootstrap technique to estimate $\mathrm{Var}(S^2)$.

به منظور برآورد پارامترهای جامعه آماری، از یک نمونه آماری استفاده می‌کنیم. گاهی حجم نمونه برای برآورد چنین پارامتری کم است. برای افزایش دقت و همچنین برآورد خطای برآوردگر پارامتر جامعه، از روش‌های »بازنمونه‌گیری (Re-sampling) استفاده می‌شود. «بوت استرپ (Bootstrap) و جک نایف(Jackknife) ، از جمله روش‌های پرکاربرد در این زمینه هستند.

این روش در گروه روش‌های آمار ناپارامتری و تکنیک‌های بازنمونه‌گیری طبقه‌بندی می‌شود و به منظور برآورد پارامتر جامعه آماری با استفاده از نمونه‌گیری با جایگذاری به کار می‌رود. بوت‌استرپ همچنین برای محاسبه فاصله اطمینان برای برآوردگر کاربرد دارد. فرض کنید می‌خواهید در یک مسئله یادگیری ماشین، مهارت مدل تولید شده را اندازه‌گیری کنید. برای این کار با استفاده از روش بوت‌استرپ می‌توان یک فاصله اطمینان ایجاد کرد و برای پیش‌بینی داده‌های جدید از آن کمک گرفت.

همانطور که دیده شد، در روش جک نایف برای n مشاهده، nزیرنمونه به اندازه n-1 برای برآورد پارامتر جامعه آماری ایجاد شد. ولی در روش بوت استرپ زیرنمونه‌ها، بوسیله بازنمونه‌گیری با جایگذاری از نمونه اصلی تولید می‌شوند. با توجه به اینکه تعداد نمونه اصلی برابر با n است می‌توان بی‌نهایت زیرنمونه با اندازه n با جایگذاری ایجاد کرد. در روش بوت استرپ تعداد بازنمونه‌گیری با Nb و تعداد مشاهدات در هر زیرنمونه نیز با nbمشخص می‌شود. واضح است که Nbو nbممکن است از nبزرگتر، مساوی یا کوچکتر باشند زیرا نمونه‌گیری با جایگذاری است و در حقیقت حجم نمونه اصلی را می‌توان بسیار بزرگ در نظر گرفت.

ب) در این قسمت ما باید به صورت رندوم ترکیب های ۱۰ تایی را انتخاب کنیم و میانگین آن را حساب میکنیم و از ۷ مقدار ان را کم میکنیم اگر بین ۵ و -۵ بود کانتر ان را یکی اضافه میکنیم این کار را ۱۰۰۰ یا بیشتر تکرا میکنیم و احتمال به دست می آوریم .

```
# Python program to get average of a list
def Average(lst):
    return sum(lst) / len(lst)

list_num=[56,101,78,67,93,78,64,72,80,69]

print(Average(list_num))

mean = 78.5
```

سورس ان به صورت زیر خواهد بود :

احتمال ۰.۷ به دست امده است

```python
import math
import numpy as np
import random

# Python program to get average of a list
def Average(lst):
    return sum(lst) / len(lst)

list_num=[56,101,78,67,93,78,64,72,80,69]
print("mean is : ")
print(Average(list_num))

r = 100;
X = [56,101,78,67,93,87,64,72,80,69]
n = len(X)

def randomvecgenerator(Y):
    r = []
    for i in range(0,len(Y)):
        t = random.randrange(0,len(Y))
        r.append(X[t])
    return r

#Bootstrap
counter=0
for i in range(r):
    idx = randomvecgenerator(X);
    mean_list=Average(idx)
    if((mean_list-78.5)<5):
        if((mean_list-78.5)>-5):
            counter=counter+1

print("-------------------------")
print(counter/r)
```

تمرین سوم :

14. If $n = 2$ and $X_1 = 1$ and $X_2 = 3$, what is the bootstrap estimate of $\mathrm{Var}(S^2)$?

این سوال رو با شبیه سازی با ۱۰۰ بار تکرار به صورت زیر انجام میشود .

```python
import math
import numpy as np
import random

r = 100;
X = [1,3]
n = len(X)

def randomvecgenerator(Y):
    r = []
    for i in range(0,len(Y)):
        t = random.randrange(0,len(Y))
        r.append(X[t])
    return r

# Estimator g
theta = np.var(X) * (len(X)/(len(X)-1))

#Bootstrap
MSE = 0;
for i in range(r):
    idx = randomvecgenerator(X);
    Yi = (np.var(idx) * (len(idx)/(len(idx)-1)) - theta) ** 2;
    MSE = MSE + Yi;


MSE = MSE / r


print('MSE for 100 simulations is:',MSE)
```

خروجی آن عدد ۲ خواهد بود .

ترکیب های ممکن برای ۱ و ۳ به صورت زیر است :

- ۱و۱
- ۱و۳
- ۳و۱
- ۳و۳

تمرین چهارم :

15. If $n = 15$ and the data are

$$5, 4, 9, 6, 21, 17, 11, 20, 7, 10, 21, 15, 13, 16, 8$$

approximate (by a simulation) the bootstrap estimate of $\mathrm{Var}(S^2)$.

```
import math
import numpy as np
import random

r = 100;
X = [5,4,9,6,21,17,11,20,7,10,21,15,13,16,8]
n = len(X)
def randomvecgenerator(Y):
    r = []
    for i in range(0,len(Y)):
        t = random.randrange(0,len(Y))
        r.append(X[t])
    return r

# Estimator g
theta = np.var(X) * (len(X)/(len(X)-1))

#Bootstrap
MSE = 0;
for i in range(r):
    idx = randomvecgenerator(X);
    Yi = (np.var(idx) * len(idx)/(len(idx)-1) - theta) ** 2;
    MSE = MSE + Yi;


MSE = MSE/r

print('MSE for 100 simulations is:',MSE)
```

خروجی آن به صورت زیر است :

MSE for 100 simulations is: 50.25622766439908