به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

آمار و احتمال مهندسی

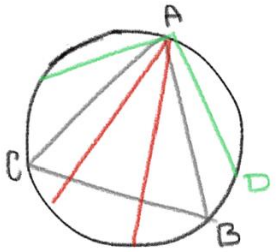
گزارش پروژه‌ی نهائی

نام و نام خانوادگی	فاطمه جلیلی
شماره دانشجویی	810199398
تاریخ ارسال گزارش	1400/11/2

فهرست گزارش سوالات

سوال اول: پارادوکس برتراند

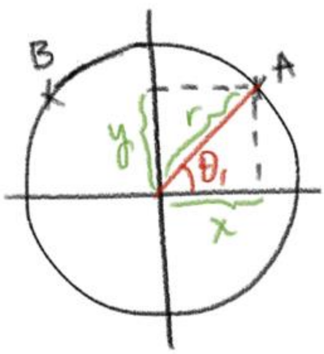
1-1-الف :



IMG 1 راه حل اول

زمانی طول وتر AD مطابق شکل 1 بزرگ تر از اضلاع مثلث متساوی الاضلاع می شود که نقطه ی D بین نقاط B,C روی دایره قرار بگیرد بنابراین برای محاسبه احتمال بایستی طول وتر BC را بر محیط دایره تقسیم کنیم ،از آن جایی که کمان BC رو به روی زاویه ی 60 درجه اس بنابراین طول آن برابر $\frac{\pi}{3}$ است و از آن جایی که محیط دایره برابر π است پس احتمال بزرگ تر بود طول وتر AD از ضلع مثلث برابر $\frac{1}{3}$ بدست می آید

1-1-ب :



IMG 2 شبیه سازی راه حل اول

از آنجایی که شعاع دایره برابر 1 است می توان مطابق شکل 2 طول و عرض نقاط رندوم را به ترتیب با استفاده از $cos\theta_1$ و $sin\theta_1$بدست آورد ، سپس این طول و عرض های نقاط 1 را در آرایه های x_1, y_1 و طول و عرض نقاط 2 ام را در x_2, y_2 ذخیره می کنیم

سپس برای بدست آوردن طول وتر کشیده از آنجایی که طول پاره خط بین دو نقطه با مختصاد x_1, y_1 و x_2, y_2 در مختصاد کارتزین مطابق رابطه ی $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ بدست می آید با جایگذاری در این فرمول و با استفاده از sqrt در کتابخانه math این حاصل را بدست می آوریم و طول وتر ها را در آرایه ای به نام distance ذخیره می کنیم

در مرحله ی بعدی برای رسم دایره با استفاده از کتابخانه matplotlib.circle با دادن مختصاد مرکز دایره که در اینجا مبدا مختصاد است ، شعاع دایره ، رنگ آن ، توخالی بودن و یا پر بودن آن و همچنین ضخامت آن به عنوان ورودی دایره را رسم می کنیم

برای رسم وتر ها و مثلث از آن جایی که از پیش مختصاد طول و عرض وتر ها را مطابق آنچه توضیح داده شد بدست آوردیم و در یک آرایه ذخیره کردیم با دادن این 4 آرایه (طول و عرض نقاط شروع و پایان

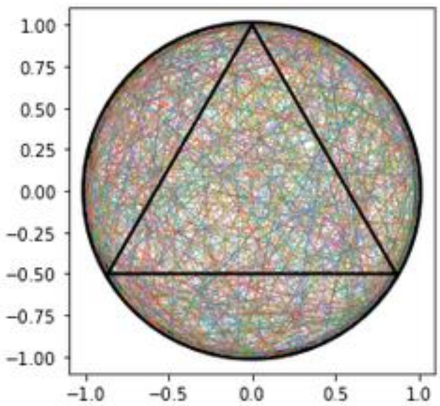
وتر ها) به عنوان ورودی به plt.plot به این صورت که طول ها را در یک پرانتز به عنوان ورودی اول و عرض ها را هم در یک پرانتز به عنوان ورودی دوم به تابع می دهیم و ضخامت آن ها را هم بسیار کم بر می گزینیم تا به دلیل تعداد زیاد وتر ها وضوح کافی را داشته باشیم و این چنین نباشد که داخل دایره پر شود

و در انتها با استفاده از plt.show همه ی آنچه به تفکیک رسم کردیم را در یک نمودار نمایش می دهیم.

برای بدست آوردن احتمال بزرگ تر بودن وترهای رندوم نسبت به اضلاع مثلث متساوی الاضلاع از آنجایی که طول ضلع مثلث محاط در دایره برابر$\sqrt{3}$ است ، در یک حلقه for چک می کنیم که آیا اندازه هر کدام از اعضای آرایه ی distance بزرگ تر از $\sqrt{3}$ هست یا خیر و در این صورت یکی به متغیری به نام success اضافه می کنیم و در پایان نسبت اندازه ی این متغیر به تعداد دفعات شبیه سازی یعنی 1000 را پرینت می کنیم که همان احتمال مورد نظر است

احتمال در شبیه سازی برابر 0.322 بدست می آید که نزدیک مقدار تئوری محاسبه شده است :

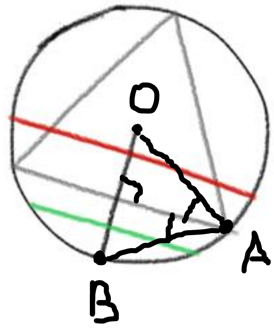
```
probability in fisrt method using 1000 times of simulation is : 0.322
```



نتیجه نهایی 1

1-2-الف :

از آن جایی که مطابق شکل وتر های عمود به شعاع که بالای ضلع مثلث هستند بزرگ تر از ضلع مثلث متساوی الاضلاع و وتر هایی که پایین تر از آن هستند کوچک تر از آن هستند و همچنین از آنجایی که ضلع مثلث شعاع عمود بر آن را نصف می کند (چرا که مثلث AOB متساوی الاضلاع است و دو زاویه مشخص شده در شکل برابر هستند)

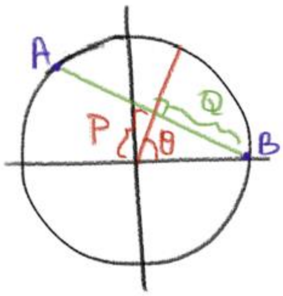


IMG 3 راه حل دوم

پس به احتمال ½ وتر بالای شعاع قرار دارد و بزرگ تر از ضلع مثلث است و به احتمال ½ هم کوچک تر از ضلع مثلث است

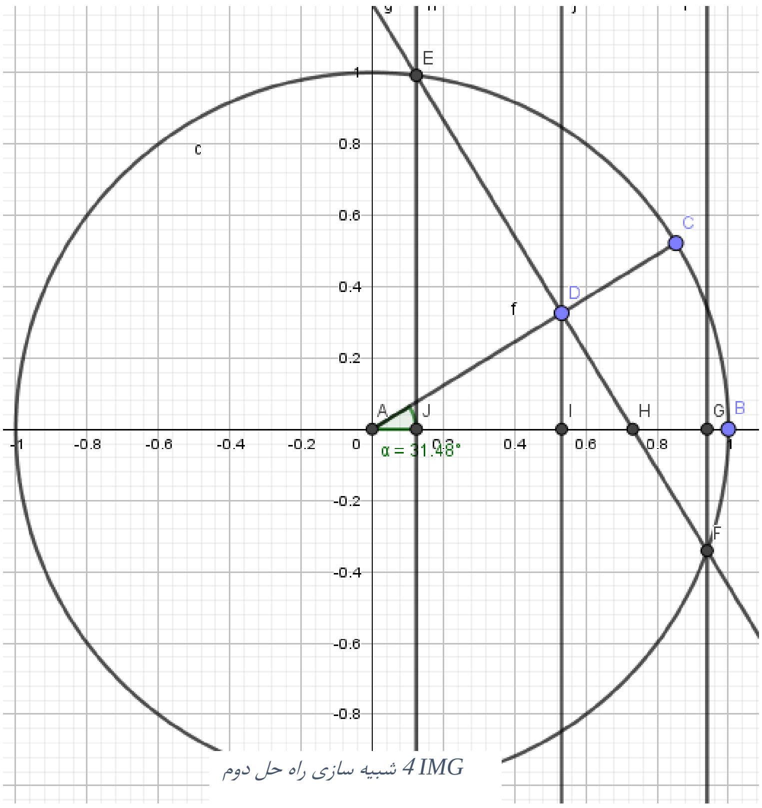
1-2-ب:

مطابق شکل می توان مقدار Q را با توجه به فیثاغورث از رابطه ی $\sqrt{1-P^2}$ بدست آورد



برای بدست آوردن نقاط ابتدایی و انتهایی وتر ها نیاز به کمی محاسبات داریم ، مطابق شکل داریم :

برای بدست آوردن طول نقطه ی اول یعنی F می توان این طول یعنی AG را به صورت مجموع AL+LG نوشت ، از آن جایی که زاویه ی DAL برابر زاویه ی LDA برابر زاویه ی HFG با استفاده از ویژگی های مثلث قائم الزاویه و برابر زاویه ی تتا است



IMG 4 شبیه سازی راه حل دوم

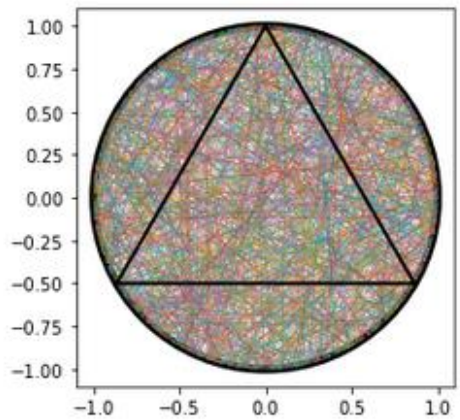
می توان طول AL را به صورت $pcos(\theta)$ و طول LG یعنی مجموع HL+HG را به صورت $DHsin\theta + HFsin\theta = DFsin\theta = Qsin\theta$ نوشت پس در کل طول نقطه ی اول برابر $pcos(\theta) + Qsin\theta$ بدست می آید به طور کاملا مشابه با استفاده از نوشتن نوشتن طول نقطه ی دوم به صورت AL-AJ طول نقطه ی دوم برابر $pcos(\theta) - Qsin\theta$ بدست می آید و همچنین با نوشتن عرض نقاط به صورت تفریق و جمع دو پاره خط که مقادیر آن ها قابل محاسبه هستند می توان مشابها عرض نقطه ی اول را به صورت $psin\theta - Qcos\theta$ و عرض نقطه ی دوم را به صورت $psin\theta + Qcos\theta$ نوشت

حال که مختصاد ابتدایی و انتهایی وتر ها را بدست آوردیم طول و عرض های نقاط 1 را در آرایه های x_1, y_1 و طول و عرض نقاط 2 ام را در x_2, y_2 ذخیره می کنیم

از اینجا به بعد درست مشابه راه حل اول پیش می رویم یعنی طول وتر ها را مطابق آنچه توضیح داده شد محاسبه و ذخیره می کنیم ، دایره و وتر ها را رسم می کنیم و در آخر احتمال را محاسبه می کنیم و آن را پرینت می کنیم

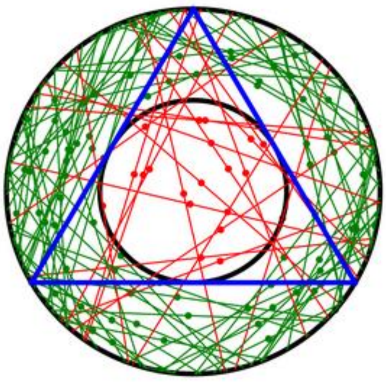
احتمال در شبیه سازی برابر 0.507 بدست می آید که نزدیک مقدار تئوری محاسبه شده است :

```
probability in second method using 1000 times of simulation is : 0.507
```



نتیجه نهایی 2

1-3-الف:



IMG 4 راه حل سوم

دایره ای محاط بر مثل متساوی الاضلاع درون مثلث رسم می کنیم در این روش اگر نقطه ای که انتخاب می کنیم داخل این دایره باشد وتر رسم رسم بزرگ تر از ضلع مثلث خواهد بود چرا که به مرور که از مرکز دور می شویم وتر های موازی کوچک تر شده تا جایی که به دایره مماس شده و با ضلع مثلث برابر می شود و از آن جا به بعد کوچک تر از ضلع مثلث خواهد بود ، می دانیم شعاع دایره ی محاطی مثلث متساوی الاضلاع نصف شعاع دایره ی محیطی آن است چرا که میانه ها که در اینجا ارتفاع نیز هستند یدیگر را به نسبت 1 به 2 قطع می کنند پس احتمال قرار گیری نقطه درون دایره کوچک برابر نسبت مساحت دایره ی کوچک به مساحت دایره بزرگ یعنی مربع نسبت شعاع های آن ها برابر ¼ خواهد بود.

1-3-الف:

شبیه سازی این راه حل دقیقا مطابق راه حل قبلی است منتها با این تفاوت که اینبار از جذر P استفاده می کنیم چرا که طبق گفته ی سوال نحوه ی انتخاب نقطه روی مساحت دایره یکنواخت است به این معنا که میانگین فاصله ی بین نقاط در هر فاصله ای از مرکز برابر باشد یعنی به نسبتی که شعاع دایره های هم مرکز با دایره اصلی بیش تر می شود باید تعداد نقاط روی محیط آن ها نیز بیش تر شود ، یعنی pdf یک خط صاف به صورت cx است ، از آنجایی که مساحت زیر منحنی pdf باید برابر 1 باشد و x از 0 تا شعاع دایره یعنی 1 افزایش پیدا می کند پس داریم :

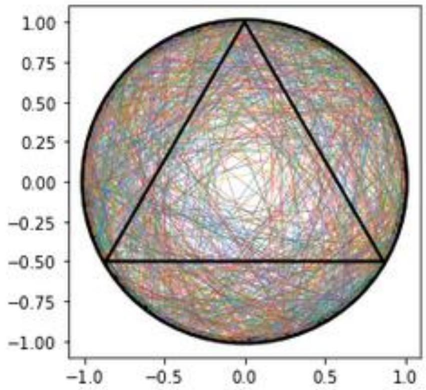
$$\int_0^1 cx = \frac{c}{2} = 1 \rightarrow c = 2 \rightarrow f_X(x) = 2x \rightarrow F_X(x) = \int f_X(x) = x^2$$

یعنی به عبارتی $y = x^2$ پس برای بدست آوردن xباید از y جذر بگیریم

حال کد را دقیقا مطابق راه حل دوم می نویسیم ولی برای p از جذر متغیر های یکنواخت استفاده می کنیم و طول و عرض نقاط ابتدایی و انتهایی مطابق محاسبات توضیح داده شده بدست می آوریم و تعریف می کنیم ، طول وتر ها را مطابق آنچه توضیح داده شد محاسبه و ذخیره می کنیم ، دایره و وتر ها را رسم می کنیم و در آخر احتمال را محاسبه می کنیم و آن را پرینت می کنیم

احتمال در شبیه سازی برابر 0.247 بدست می آید که نزدیک مقدار تئوری محاسبه شده است :

```
probability in third method using 1000 times of simulation is : 0.247
```



نتیجه نهایی 3

سوال دوم: تخمین عدد اویلر

1) شخص تعدادی متغیر های تصادفی با توزیع نرمال را باهم جمع می کنیم تا حاصل بزرگ تر از 1 شود و حداقل این تعداد مورد سوال است ، در حالت کلی تر می خواهیم ببینیم حداقل چه تعداد متغیر تصادفی یکنواخت U_i را باید باهم جمع کنیم تا حاصل بزرگ تر از x شود ؛ یعنی :

$$N(x) = \min\{n: \sum_{i=1}^{n} U_i > x\}$$

فرض کنیم :

$$m(x) = E[N(x)] = \int_0^1 E[N(x)|U_1 = y]dy$$

داریم :

$$E[N(x)|U_1 = y] = \begin{cases} 1 & y > x \\ 1 + m(x-y) & y < x \end{cases}$$

با جایگذاری عبارت سوم در دوم داریم:

$$m(x) = \int_0^x m(x-y)dy = 1 + \int_0^x m(u)du \overset{\frac{d}{dx}}{\rightarrow} m'(x) = m(x)$$

$$\rightarrow \ln\big(m(x)\big) = x + c \rightarrow m(x) = ke^x$$

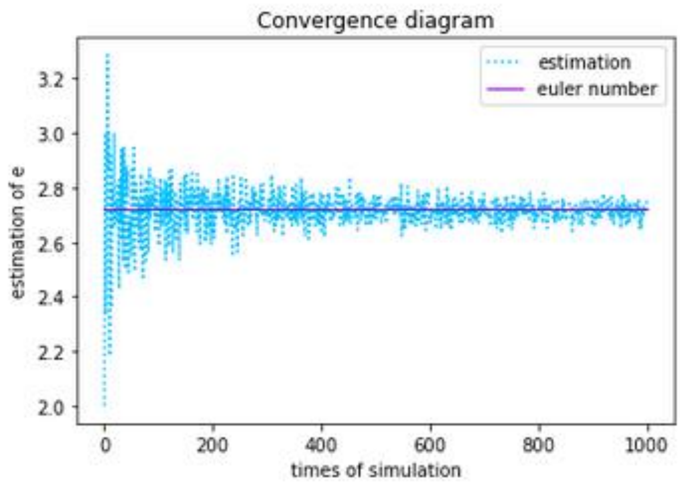
از آن جایی که $m(0) = 1$ پس با جایگذاری $k = 1$ است پس $m(x) = e^x$ پس مقدار خواسته شده که مطابق تعریف $N(1)$ نام گذاری شد برابر e بدست می آید

2) برای شبیه سازی تابعی به نام estimate تعریف می کنیم که با استفاده از آن می توانیم به تعداد ورودی دلخواه این روند را شبیه سازی کنیم و میانگین تعداد دفعات تولید توزیع یکنواخت را تا حاصل مجموع بزرگ تر از 1 شود را بدست آوریم

به این ترتیب که در یک حلقه ی for تا زمانی که مجموع اعداد تولید شده از 1 کم تر باشد عدد رندوم تولید می کنیم و مجموع را با عدد رندوم جدید جمع می کنیم تا بالاخره حاصل از 1 بزرگ تر شود و در این حلقه while هر بار تعداد دفعاتی که حاصل بزرگ تر از 1 شود را در متغیری ذخیره می کنیم و مجموع آن ها را بر تعداد یعنی ورودی تابع که تعداد دفعات شبیه سازی است تقسیم می کنیم تا میانگین تعداد دفعات تولید توزیع یکنواخت را تا حاصل مجموع بزرگ تر از 1 شود را بدست آوریم و آن را به عنوان خروجی تابع در نظر می گیریم

حال به ازای ورودی 1 تا 1000 حاصل تابع estimate را بدست می آوریم و به عنوان متغیر y نمودار در نظر می گیریم و نمودار همگرایی را با استفاده از کتابخانه matplotlib رسم می کنیم همچنین مقدار دقیق عدد e را هم به عنوان نمودار دوم یکجا رسم می کنیم تا نمودار همگرایی بدست آمده قابل مقایسه با مقدار دقیق باشد ، عنوان نمودار و محور های x,y را هم به طور مناسب تعیین می کنیم و نمودار نهایی را نمایش می دهیم که به صورت شکل زیر خواهد بود :

که همانطو که مشخص است هر چه تعداد دفعات شبیه سازی بیش تر می شود مقدار تخمین e دقیق تر خواهد بود



نتیجه نهایی 4

سوال سوم: کبریت باناخ

1)

آقای باناخ هر بار به طور رندوم دست در یکی از جیب های خود می کند یعنی احتمال برداشتن کبریت از هر یک از جیب ها ½ است ، این کار را تا زمانی ادامه می دهد که دست در جیبش کند و ببیند که خالی است یعنی بایستی قبلا N کبریت از این جیب برداشته باشد و از آن جایی که فرضا k کبریت در جیب دیگر باقی مانده است پس آقای باناخ به نوعی 2N-k+1 انتخاب برای اینکه از کدام جیب کبریت بردارد داشته است (به علاوه 1 به خاطر انتخاب آخر است که می بیند جیب خالی است یعنی N+1 بار انتخاب مربوط به جیبی که خالی شده و N-k انتخاب مربوط به جیبی که هنوز k کبریت در آن باقی مانده است)

بنابراین این یک توزیع برنولی منفی است که داریم :

(فرضا k تعداد کبریت های مانده در جیب راست باشد)

$$P_R(K = k) = \binom{2N - K}{N} \left(\frac{1}{2}\right)^{2n+1-k}$$

به همین صورت به طور مشابه برای جیب چپ همین احتمال (P_L) وجود دارد پس برای بدست آوردن P کل که مربوط به هر دو جیب است باید حاصل فوق را ضرب در دو کنیم :

$$P(K=k)=\binom{2N-K}{N}\left(\frac{1}{2}\right)^{2n-k}$$

2) امید توزیع به دست آمده را می توان با استفاده از تقریب Stirling's approximation که بیان می کند :

$$if\ n\to\infty : n! \sim\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

به صورت

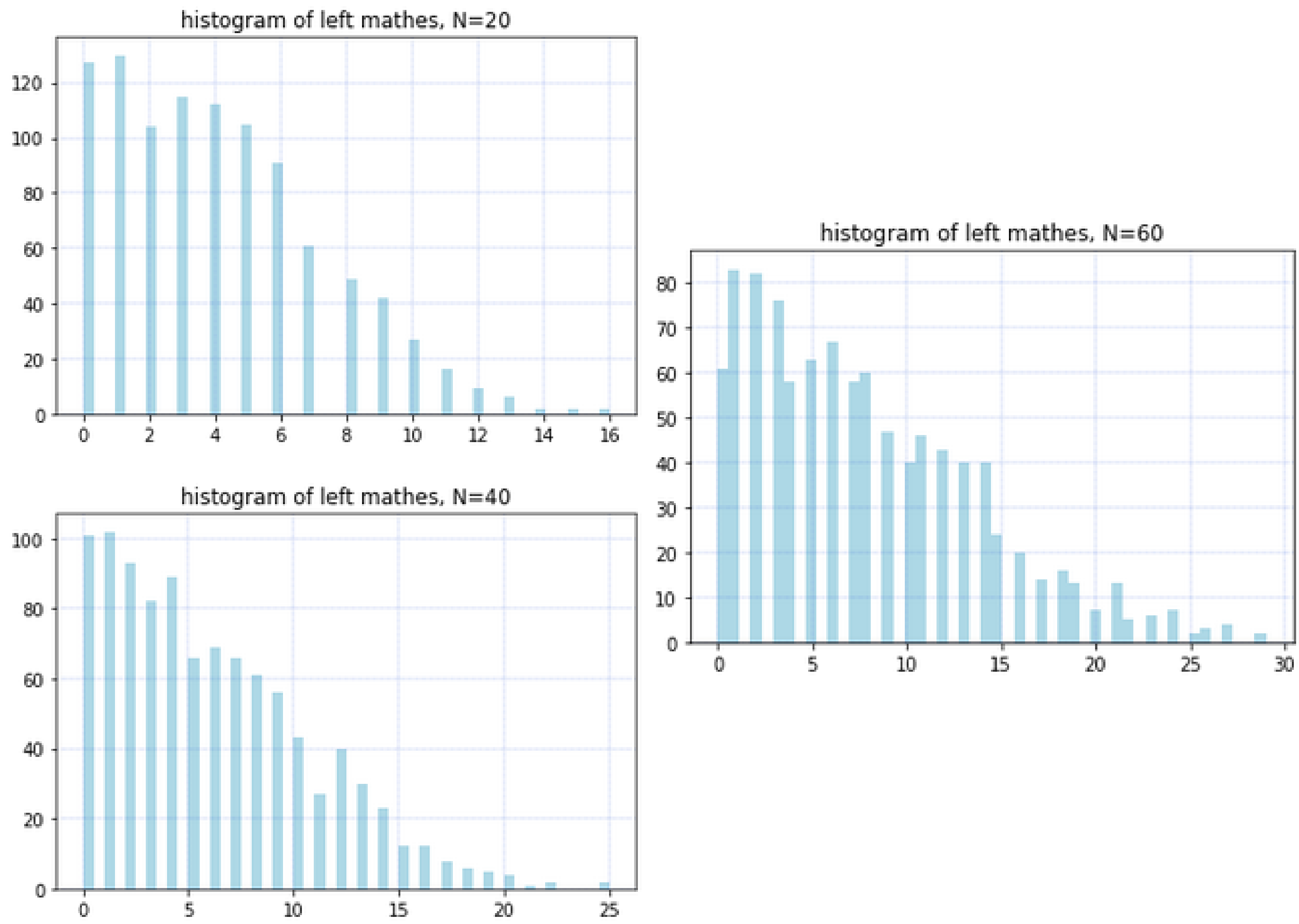
$$2\sqrt{\frac{N}{\pi}}-1$$

تقریب زد

3) برای شبیه سازی تابعی به نام estimate تعریف می کنیم که تعداد دفعات شبیه سازی را ورودی می گیرد ، دو متغیر که تعداد انتخاب ها از جیب راست و یا چپ را نشان می دهند تعریف می کنیم و در یک حلقه for تا زمانی که انتخاب ها از جیب چپ یا راست از N+1 کم تر است متغیر رندوم تولید می کنیم و یکی از جیب ها را انتخاب می کنیم و تعدا انتخاب های از آن جیب را افزایش داده تا به N+1 برسد در این زمان عضو i ام یک آرایه از پیش تعریف شده که همه عضو های آن را 0 گذاشتیم برابر تعداد سیگار های هر جیب منهای تعداد انتخاب از جیب دیگر می گذاریم و در آخر این آرایه را خروجی می دهیم

حال به ازای 3 مقدار N متفاوت داده شده هیستوگرام estimate(N) را رسم می کنیم و رنگ ، عنوان نمودار ، و مشخصات چارت بندی نمودار های را به دلخواه تعیین می کنیم و آن ها را رسم می کنیم

شکل سه هیستوگرام مطابق زیر خواهد بود :



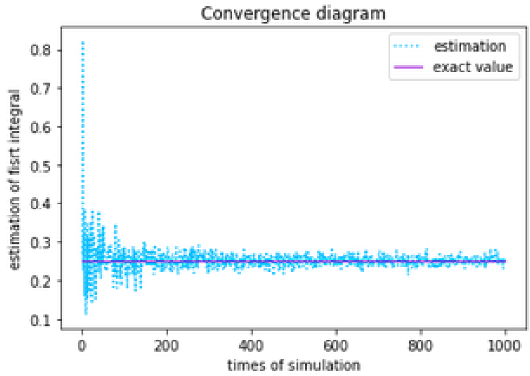
نتیجه نهایی 5

سوال چهارم: محاسبه انتگرال

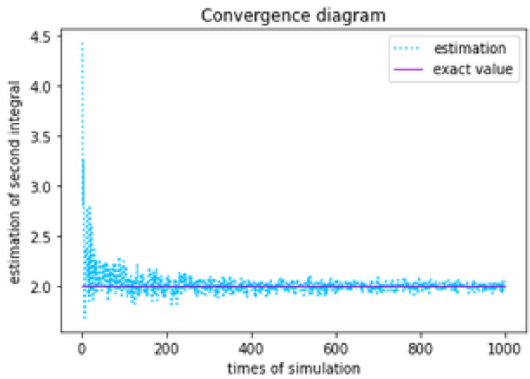
1)

برای محاسبه مقدار عددی انتگرال ها تابعی تعریف می کنیم که مقادیر بازه های انتگرال خود تابع و تعداد بار های انتخاب متغیر تصادفی یکنواخت برای بدست آوردن مقدار انتگرال در آن نقاط را به آن ورودی می دهیم و مقدار انتگرال را با استفاده فرمول داده شده بدست می آوریم و نمودار همگرایی مقادیر انتگرال ها را هم رسم ی کنیم به این صوت که برای x=1 تا 1000 مقدار تابع توصیف شده را با تکرار های 1 تا 1000 بار در تعداد بار های انتخاب متغیر تصادفی یکنواخت برای بدست آوردن مقدار انتگرال در آن نقاط محاسبه می کنیم و رسم می کنیم ، خروجی های انتگرال ها مطابق زیر خواهد بود :

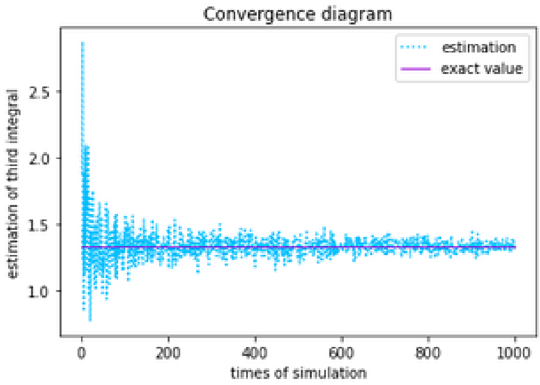
```
the first integral numerical value is  0.25384609625274096
```




```
the second integral numerical value is  2.026401576621315
```




```
the third integral numerical value is  1.3102287616238761
```



نتیجه نهایی 6

با توجه به اینکه انتگرال های داده شده ساده هستند مقادیر دقیق آن ها به سادگی به ترتیب برابر 0.25 و 2 و arctan(4) بدست می آید که برای مقایسه مقدار دقیق آن ها نیز به رنگ بنفش در نمودار های همگرایی رسم شده است

به این ترتیب خطای حاصل برای انتگرال های 1تا 3 با توجه به مقادیر بدست آمده که بالای هر نمودار نوشته شده به ترتیب برابر خواهد بود با 0.0038 ،0.026،00.015 که خطای کمی است و قابل صرف نظر است

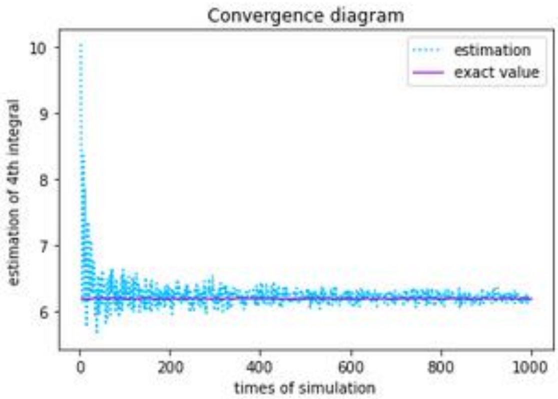
2) به طریق مشابه با استفاده از تابعی که تعریف کردیم مشابه قسمت قبل دو انتگرال زیر که حل تحلیلی ندارند را بدست می آوریم و نمودار همگرایی آن ها رسم می کنیم :

$$4th\ integral: \int_0^{\pi} \frac{sinx}{x} dx$$

$$5th\ integral: \int_0^{\pi} e^{sinx} dx$$

در جلوی هر نمودار مقداری که ولفرام برای این انتگرال می دهد هم برای مقایسه نوشته شده است و دیده می شود که خطا کم است و تخمین حاصل ، تخمین به نسبه خوبی است

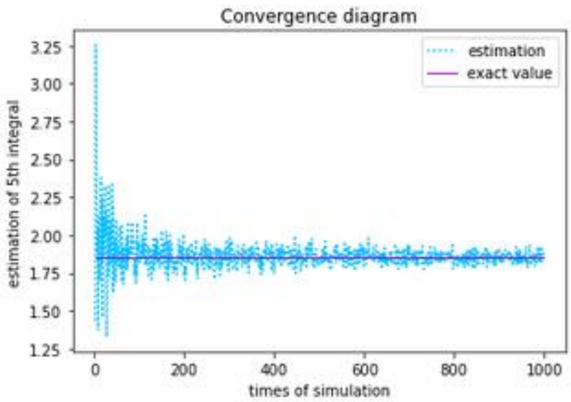
```
the 4th integral numerical value is  6.233619383471986
```



Definite integral

$$\int_0^{\pi} e^{\sin(x)}\, dx = \pi\,(\mathbf{L}_0(1) + I_0(1)) \approx 6.20876$$

```
the 5th integral numerical value is  1.896185156832818
```



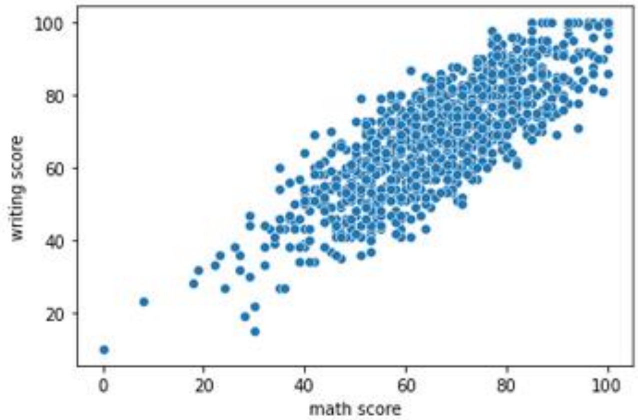
Definite integral

$$\int_0^{\pi} \frac{\sin(x)}{x}\, dx = \mathrm{Si}(\pi) \approx 1.85194$$

نتیجه نهایی 7

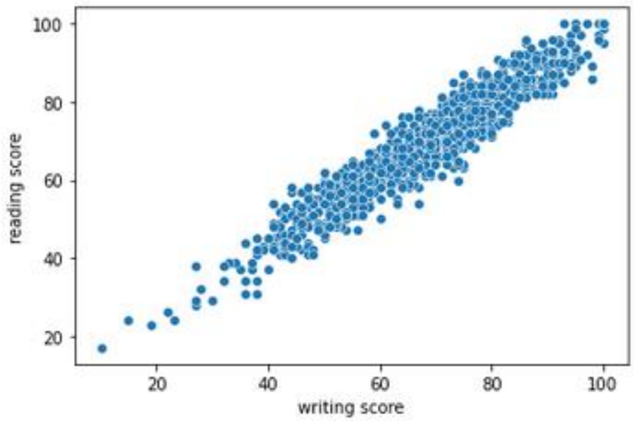
سوال پنجم: کار با داده

1-آ:



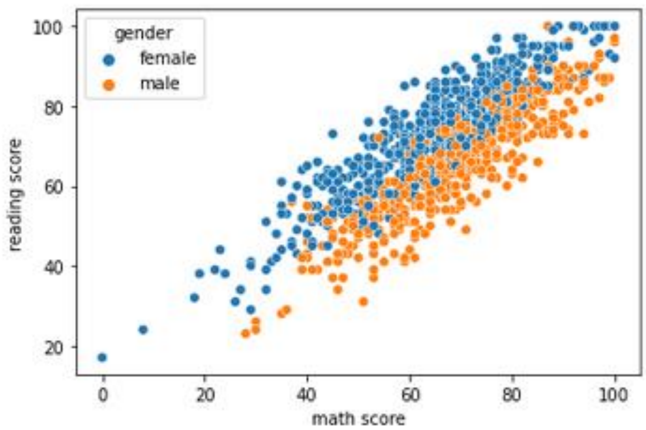
بله همانطور که دیده می شود با توجه به نمودار پراکندگی فرض درست است و نمره ریاضی و نمره نوشتار باهم نسبت مستقیم دارند

1-ب:



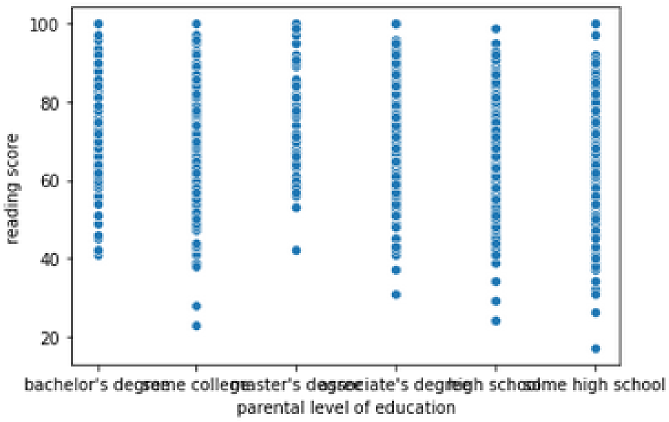
همانطور که دیده می شود با توجه به نمودار پراکندگی فرض غلط است و نمره خوانش و نمره نوشتار باهم نسبت مستقیم دارند

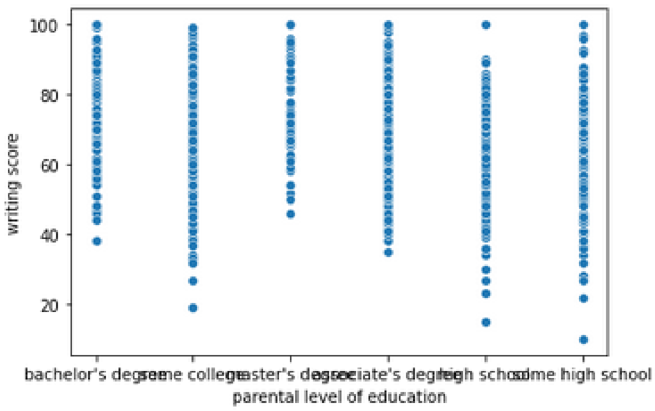
1-ج :



همانطور که دیده می شود در هر دو جنس نمره ریاضی و نمره خوانش رابطه ی مستقیم دارند ولی به طور کلی نمره خوانس دختر ها بالاتر از نمره خوانس پسرهاست ولی نمره ریاضی پسر ها بالا تر از نمره ریاضی دختر هاست

1-د:



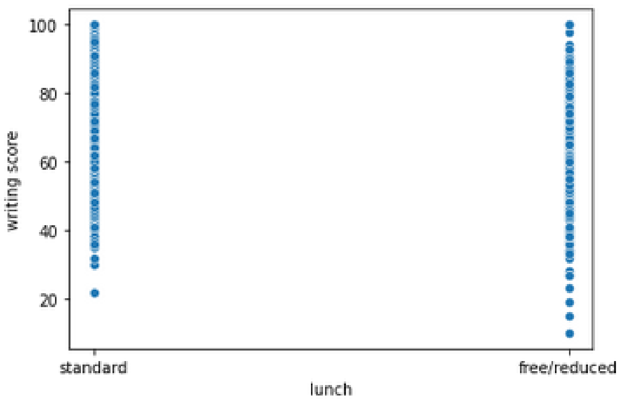


همانطور که دیه می شود مطابق نموداره های کشیده شده این چنین نیست و گاها فرزندانی که والدینشان تحصیلات بالاتری دارند نمره کم تری دارند مثل مقایسه بین

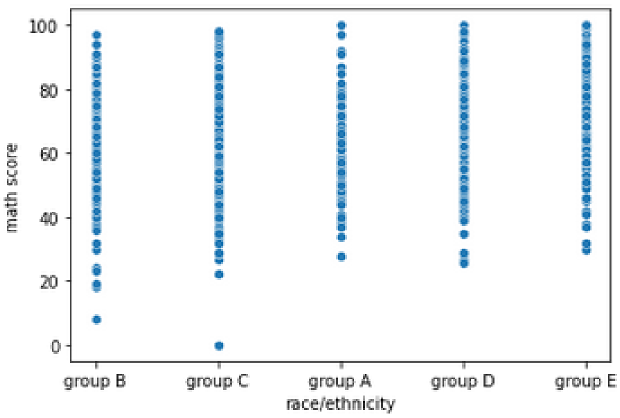
bachelor's degree ،

associate's degree

2)



کسانی که نمره خوانش بیش تری دارند به طور عمده دارای نهار استاندارد تری هستند



گروه E به طور میانگین نمرات ریاضی بالاتری نسبت به بقیه گروه ها دارند