به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

آمار و احتمال مهندسی

پروژه‌ی نهایی

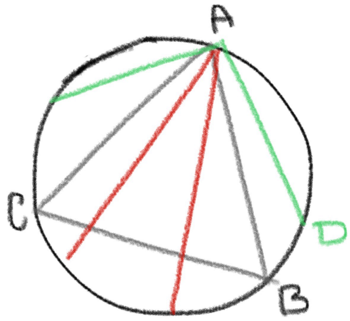
طراحان:	امیرعلی سلطانی تهرانی - شیوا شاکری
تاریخ ویرایش نسخه	۸ دی ۱۴۰۰
مهلت تحویل گزارش	۲۲ دی ۱۴۰۰

فهرست مطالب

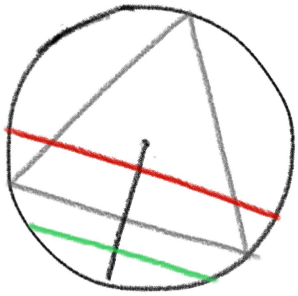
۱ پارادوکس برتراند

در برخی مسائل احتمالاتی، برای محاسبه‌ی احتمال یک پیشامد می‌توان از روش‌های متفاوت استفاده کرد که نتایج متفاوت و درستی را نیز به دنبال دارند؛ یکی از این مسائل را جوزف برتراند در سال ۱۸۸۹ مطرح کرد. مسأله‌ای که او بیان کرد را می‌توان با سه زاویه‌ی دید متفاوت حل کرد و به سه پاسخ متفاوت رسید. برای درک این پارادوکس، ابتدا مسأله را تعریف کرده و سپس به شبیه‌سازی راه‌حل‌های متفاوت آن مسائل دیگر می‌پردازیم.

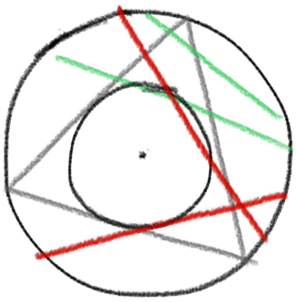
تعریف مساله. فرض کنید به‌طور تصادفی یک وتر را در دایره انتخاب کنیم. احتمال اینکه طول این وتر، بزرگتر از طول یک ضلع مثلث‌متساوی‌الاضلاع محاط در آن دایره باشد، چقدر است؟



شکل ۱: راه‌حل اول



شکل ۲: راه‌حل دوم



شکل ۳: راه‌حل سوم

۱.۱ راه‌حل اول: باتوجه به تقارن، برای رسم یک وتر تصادفی، می‌توان ابتدا دو نقطه‌ی تصادفی روی محیط دایره انتخاب کرد و آن‌ها را به‌هم وصل کرد تا وتر بین این دو نقطه حاصل شود. نقطه اول را A و نقطه‌ی دوم را D در نظر بگیرید. فرض کنید A یکی از رئوس مثلث متساوی‌الاضلاع محاط باشد؛

۱.۱.الف به‌صورت حل دستی احتمال اینکه وتر AD بزرگتر از طول ضلع مثلث ABC باشد را به‌دست آورید.

۱.۱.ب برای شبیه‌سازی این راه حل، ابتدا ۱۰۰۰ بار، دو متغیر تصادفی یکنواخت در بازه‌ی $[0, 2\pi]$، که معرف زاویه‌ی شعاع مربوط به دو نقطه‌ی تصادفی روی محیط یک دایره با شعاع واحد و مرکز مبدا هستند را تعریف کنید. (شکل ۴) مختصات نقاط انتهایی، ابتدایی و طول وتر آن را برحسب متغیرهای تصادفی بدست آورید . سپس با استفاده از کتابخانه‌ی $Matplotlib$ دایره‌ی مورد نظر و وترهای تولید شده را رسم کنید. با بدست آوردن اندازه‌ی ضلع مثلث متساوی‌الاضلاع محاط در این دایره، نسبت تعداد وترهای بلندتر از یک ضلع مثلث به تعداد کل وترها را به‌دست آورید.

۲.۱ راه‌حل دوم: باتوجه به تقارن، برای رسم یک وتر تصادفی، نقطه‌ای تصادفی روی محیط دایره انتخاب کرده و آن را به مرکز دایره وصل می‌کنیم. به این طریق توانسته‌ایم یک شعاع تصادفی از دایره انتخاب کنیم. سپس نقطه‌ای تصادفی از روی این شعاع انتخاب می‌کنیم. وتری وجود دارد که این شعاع در این نقطه، عمودمنصف آن است.

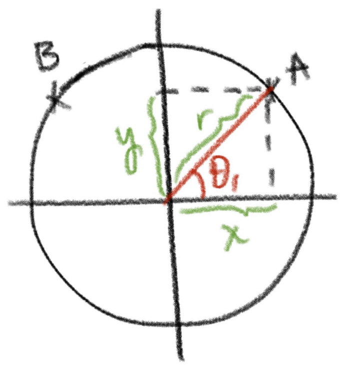
۲.۱.الف به‌صورت حل دستی احتمال اینکه این وتر بزرگتر از طول ضلع مثلث ABC باشد را به‌دست آورید.

۲.۱.ب برای شبیه‌سازی این قسمت نیز باید به تعداد ۱۰۰۰ بار، دو متغیر تصادفی یکنواخت تعریف کنید که یک متغیر تصادفی زاویه (θ) ، و دیگری شعاع تصادفی ($P \in (0, r)$)، که نسبتی از شعاع واحد است. باتوجه به شکل ۵، طول Pو Q را به‌دست آورید. سپس مختصات نقاط انتهایی، ابتدایی و طول وتر آن را برحسب متغیرهای تصادفی بدست آورید. سپس با استفاده از کتابخانه‌ی $Matplotlib$ دایره‌ی مورد نظر و وترهای تولید شده را رسم کنید و نسبت تعداد وترهای بلندتر از یک ضلع مثلث به تعداد کل وترها را محاسبه کنید.

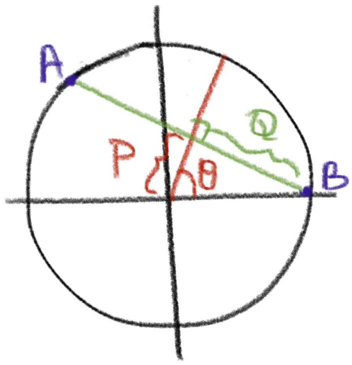
۳.۱ راه‌حل سوم: برای رسم یک وتر تصادفی، نقطه‌ای تصادفی داخل دایره انتخاب می‌کنیم و به مرکز دایره وصل می‌کنیم. سپس وتر عمود بر این خط در نقطهٔ انتخابی را رسم می‌کنیم.

۳.۱.الف به‌صورت حل دستی احتمال اینکه این وتر بزرگتر از طول ضلع مثلث ABC باشد را به‌دست آورید.

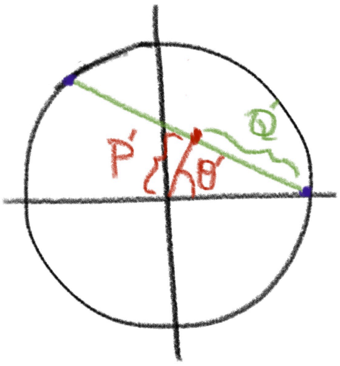
۳.۱.ب برای شبیه‌سازی این قسمت، یک فرآیند پواسون همگن روی دایره انجام می‌شود تا یک نقطه را به‌صورت یکنواخت روی دایره انتخاب کنید. باید یک متغیر تصادفی (θ') مانند قبل در نظر بگیرید. برای تولید متغیر تصادفی دیگر (P')، یک متغیر تصادفی یکنواخت در بازه‌ی واحد ایجاد کنید و سپس جذر آن‌را در شعاع ضرب کنید. توضیح دهید چرا باید از متغیر تصادفی یکنواخت ایجاد شده جذر بگیرید؟ سپس مقادیر P'و Q' را به‌دست آورید. سپس مختصات نقاط انتهایی، ابتدایی و طول وتر آن را برحسب متغیرهای تصادفی بدست آورید. سپس با استفاده از کتابخانه‌ی $Matplotlib$ دایره‌ی مورد نظر و وترهای تولید شده را رسم کنید و نسبت تعداد وترهای بلندتر از یک ضلع مثلث به تعداد کل وترها را محاسبه کنید.



شکل ۴: شبیه‌سازی راه‌حل اول



شکل ۵: شبیه‌سازی راه‌حل دوم



شکل ۶: شبیه‌سازی راه‌حل سوم

۲ تخمین عدد اویلر

در این بخش می‌خواهیم به سراغ تخمین یکی از پرکاربردترین اعدادی که تابحال در زندگی خود داشتید، برویم. احتمالاً تا الان در بسیاری از موارد با عدد اویلر e کلنجار رفته‌اید.این عدد ثابتی است که مقداری حدود 2.71828 دارد. این ثابت اولین بار توسط ژاکوب برنولی در سال ۱۶۸۳ کشف شد. رابطه‌ای که در آن برنولی توانست این ثابت را پیدا کند، به شکل زیر بود.

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$

همان‌طوری هم که قطعاً می‌دانید این عدد کاربردهای بسیار زیادی در ریاضیات، حساب دیفرانسیل و حتی توزیع‌های آماری که با آن‌ها آشنا هستید، دارد.

حال به روش تخمینی که ما در نظر داریم، می‌پردازیم.

شخصی که علاقه‌ی زیادی به توزیع یک‌نواخت دارد ادعا کرده است که اگر پشت سیستمی بشیند و بارها توزیع یک‌نواخت در بازه $[0, 1]$ ایجاد کند و اعداد ایجادشده از این توزیع را با هم جمع کند تا بزرگ‌تر از یک شوند؛ به طور میانگین این شخص می‌بایست $e = 2.71828$ بار توزیع یک‌نواخت ساخته تا مجموع این اعداد از یک بیش‌تر شوند. حال می‌خواهیم ببینیم که حدس این شخص درست هست یا خیر.

شکل ۷: ژاکوب برنولی

شکل ۸: لئونارد اویلر

۱. ابتدا به صورت توضیحی یا ریاضی نشان دهید که این عدد معادل با عدد اویلر خواهد بود.

۲. برای چند مقدار n دلخواه آزمایش این شخص را انجام دهید و خروجی را به صورت نمودار یا عدد گزارش دهید. خروجی باید به‌گونه‌ای نشان داده شود که با عدد اویلر نیز مقایسه شده باشد.

۳ کبریت باناخ

استفان باناخ یکی از ریاضی‌دانان مشهور قرن بیستم است. این ریاضی‌دان اهل لهستان بوده و در ناحیه‌ای از شوروی که امروزه اوکراین است، خاک شده است. مساله‌ای که قرار است آن را حل کنیم، نه توسط جناب باناخ مطرح شده و حتی ایشان مساله را حل نیز نکرده‌اند بلکه یکی از همکارهای شوخ‌طبع ایشان به اسم هوگو اشتاین‌هاوس اولین بار این مساله را در جمع خودمانی آن‌ها مطرح کرده است. مساله به شرح زیر است.
از آن‌جایی که آقای باناخ علاقه‌ی زیادی به کبریت داشتند همیشه در جیب‌های خودشان کبریت می‌گذاشتند. یک روز آقای اشتاین‌هاوس می‌پرسند که اگر باناخ یک جعبه کبریت در جیب چپ و یک جعبه کبریت دیگر در جیب راست خود داشته باشد و در طول روز به صورت تصادفی و هم‌شانس دست‌های خود را در جیب‌های خود بکند و کبریتی را آتیش بزند؛ احتمال اینکه کبریت‌های یکی از جعبه‌ها به اتمام برسد و در جعبه دیگر دقیقاً k کبریت باقی مانده باشد، چقدر است؟ البته آقای اشتاین‌هاوس این فرض را کرده بودن که در هر دو جعبه دقیقاً N کبریت باشد.

شکل ۹: استفان باناخ

شکل ۱۰: هوگو اشتاین‌هاوس

۱. روابط احتمالی مساله بالا را نوشته و تا جایی که ممکن است ساده کنید.

۲. سعی کنید با استفاده از تقریب‌هایی که بلد هستید، عبارتی برای تعداد کبریت‌های جعبه پر بدست آورید. (امتیازی)

۳. با استفاده از شبیه‌سازی هیستوگرام مقادیر کبریت در جعبه پر را برای $N = 20$ ، $N = 40$ و $N = 60$ بدست آورده و با مقدار تقریبی مقایسه کنید.

۴ محاسبه انتگرال

یکی از مهم‌ترین اپراتورها در حساب دیفرانسیل، اپراتور انتگرال است. در بسیاری از موارد حل تحلیلی انتگرال‌ها بسیار وقت‌گیر و حتی نشدنی هست پس در این مواقع سراغ حل عددی این انتگرال‌ها می‌رویم. به عنوان مثال فرض کنید می‌خواهید مقدار انتگرال زیر را محاسبه کنید.

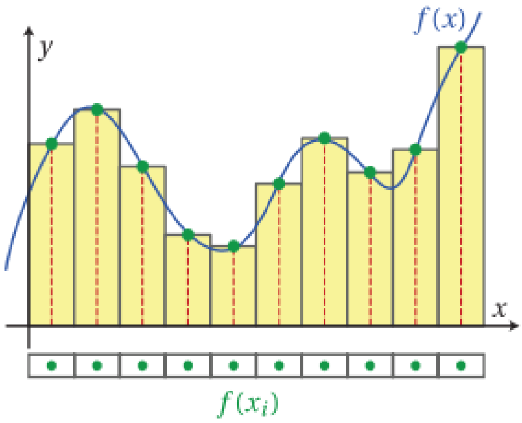
$$\int_0^1 \sqrt{1-x^4}\,dx$$

این انتگرال به صورت تحلیلی قابل محاسبه نیست ولی شاید در مواردی بخواهید مقدار این انتگرال را داشته باشید پس ناچار می‌بایست به سراغ روش‌های حل عددی بروید. یکی از این روش‌ها، استفاده از توزیع‌های آماری است، با توجه به اینکه در توابع تک‌متغیره ریاضی می‌توان انتگرال را به صورت مساحت زیر سطح نیز تعریف کرد و این مساحت را می‌توان به صورت مجموعی از مستطیل‌های کوچک که طول آن‌ها برابر با مقادیر تابع هستند نیز معرفی کرد پس با استفاده از توزیع یک‌نواخت به صورت شهودی می‌توان این مساحت را محاسبه کرد. پس در حالت کلی اگر بخواهیم انتگرال تابعی دلخواه به شکل زیر را محاسبه کنیم می‌توانیم از روش گفته‌شده ایده گرفته و روابط را بنویسیم.

$$F = \int_a^b f(x)\,dx$$

$$F^N \approx (b-a)\frac{1}{N-1}\sum_{i=0}^{N} f(x_i)$$

صحتی شهودی‌تر از این قضیه در شکل زیر قابل بیان است.



شکل ۱۱: محاسبه انگرال با توزیع یکنواخت

حال با استفاده از توضیحات ارائه‌شده به سوالات زیر پاسخ دهید.

۱. مقادیر انتگرال‌های زیر را بدست آورده و مقدار خطا از مقدار واقعی را بدست آورید. (مقدار واقعی را به صورت تحلیلی محاسبه کنید)

(آ) $\int_0^1 x^3\,dx$

(ب) $\int_0^{\pi} \sin(x)\,dx$

(ج) $\int_0^4 \frac{1}{1+x^2}\,dx$

۲. دو تابع دلخواه انتخاب کرده که انتگرال آن‌ها حل تحلیلی ندارند و به صورت عددی با روش گفته‌شده مقدار انتگرال آن‌ها را محاسبه کنید. نتیجه خود را با خروجی این انتگرال از سایت wolframalpha نیز مقایسه کنید.

۵ کار با داده

در این بخش قرار است که به بررسی یک مجموعه داده واقعی برویم. این مجموعه داده با نام Student Performance در اختیار شما قرار گرفته است. در این مجموعه داده هر سطر متناظر با یک دانش‌آموز است که بعضی از ویژگی‌های آن در ستون‌های مختلف نمایش داده شده است.

۱. با استفاده از رسم نمودار پراکندگی،فرض‌های زیر را به صورت شهودی مورد بررسی قرار داده و توضیحات لازم را ارائه دهید.

(آ) نمره ریاضی با نمره نوشتار نسبت مستقیم دارد.

(ب) نمره نوشتار با نمره خوانش نسبت عکس دارد.

(ج) نمودار پراکندگی نمره ریاضی و خوانش دخترها و پسرها را در یک نمودار رسم کرده و تحلیل کنید.

(د) میزان تحصیلات خانواده بر روی نمره خوانش و نوشتار دانش‌آموزان تاثیر دارد.

۲. سعی کنید دو فرض دیگر مشابه فرض‌های بالا را مطرح کرده و نمودار پراکندگی متناسب با فرض خود را رسم کنید. تحلیلی از دلیل انتخاب این فرض‌ها و توضیحات نتایج را ضمیمه کنید.

۳. در بعضی مواقع در نمودار پراکندگی از لغزش (jitter) استفاده می‌شود. درباره این پدیده و علت آن تحقیق کرده و در یکی از نمودارهای پراکندگی این مفهوم را پیاده‌سازی کنید.

۴. چند نمودار دیگر (به دلخواه) با استفاده از مجموعه داده رسم کرده و توضیحاتی را ارائه دهید. این نمودارها می‌تواند شامل نمودار میله‌ای، دایره‌ای و ... باشد.

۵. دو آزمون فرض در داده طراحی و پیاده‌سازی کنید. نتایج را تحلیل کرده و منطق جواب را توضیح دهید.

۶ توضیحات

دانشجویان عزیز حتما به نکات زیر توجه داشته باشند.

- پروژه به‌گونه‌ای طراحی شده که به دانش آماری فراتر از آن‌چه در این درس آموخته‌اید نیاز نداشته باشد وآن‌چه را که آموخته‌اید تثبیت و تفهیم می‌کند. به همین جهت انجام آن برای یادگیری درس اکیدا توصیه می‌شود.

- صرف‌نظر از رویکرد آموزشی این پروژه، آخرین نقطه‌ی جبران نمراتتان در این درس می‌باشد و بنا به سابقه‌ی چندساله، به اسکیل‌شدن نمرات امیدی نیست، در نتیجه از اهمیت این موقعیت غافل نشوید.

- شما می‌بایست علاوه بر کدهای پیاده‌شده، گزارشی تحلیلی از نتایج خود ارائه دهید. توجه داشته باشید که مفهوم گزارش پروژه با مفهوم توضیح کد متفاوت است در نتیجه در فایل گزارش، از درج کد جدا بپرهیزید.

- کدهای پایتون و آر خود را حتما در قالب دفترچه‌ی ژوپیتر بارگذاری کنید. دستیاران آموزشی موظف به اجرای کدهای شما نیستند.

- اسکریپت‌های خود را خوانا و تمیز بنویسید. طبیعتا این درس، درس برنامه‌نویسی نیست اما کد بسیار پیچیده و غیرقابل فهم نمره‌ی کامل را دریافت نمی‌کند. استفاده از توابع و نام‌های متغیرهای بامعنا به خوانایی کد می‌افزاید.

- گزارش‌کار، اولین و مهم‌ترین آیتم نمره‌دهی می‌باشد در نتیجه با صرف زمان مناسب، گزارشی تهیه کنید که بازتاب‌گر زحماتی باشد که برای انجام پروژه کشیده‌اید. استفاده‌ی صحیح از نیم‌فاصله، علائم نگارشی، گویا بودن جملات و پاراگراف‌بندی مناسب از جمله مواردیست که در نگاه اول جلب توجه می‌کند و نکاتی نظیر استفاده از زیرنویس برای تصاویر و بالانویس برای جداول، ارجاع‌دادن به روابط و تصاویر با شماره‌ی مربوط به هر کدام و ... از جمله خصوصیت‌های یک نوشته‌ی آکادمیک است. متن گزارش را با فونت B Nazanin و اندازه‌ی ۱۴ در قالب گزارش قرار داده شده روی سایت تایپ نمائید. از قرار دادن عکس از نوشته‌ی دست‌نویس خود در گزارش به‌شدت پرهیز کنید و روابط ریاضی را نیز تایپ کنید.

- با توجه به مفهوم امتیازی بودن پروژه، به شدت با موارد تقلب چه در کد و چه در گزارش برخورد خواهد شد.

- سعی می‌شود از برخی از دوستان از طریق تماس تصویری سوالاتی در قالب تحویل پروژه پرسیده شود. در نتیجه مشخص است که هر شخص باید به تمامی محتوایی که ارائه می‌دهد مسلط باشد.

- در نهایت یک فایل گزارش پی‌دی‌اف را در کنار دفترچه‌های ژوپیتر زیپ کرده و با نام <sid>-surname.zip در صفحه‌ی درس بارگذاری کنید.

- ابهامات خود در مورد سوالات و یا قالب گزارش در گروه تلگرامی درس مطرح کنید. در انتهای هر پیام طراحان را منشن کنید. سوالات در گروه پرسیده شده و همان‌جا پاسخ داده خواهند شد تا در دسترس همه‌ی دانشجوها قرار بگیرند.