آمار و احتمال مهندسی

نیم‌سال اول ۱۴۰۳-۱۴۰۲

مدرس: دکتر نجفی

دانشکده‌ی مهندسی کامپیوتر

مهلت تحویل: روز امتحان پایان‌ترم

تمرین عملی

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص‌شده است.
- هم‌کاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.

مسئله‌ی ۱. فتح کلیمانجارو

شایان که یک کوهنورد ماهر است قصد دارد از کوه کلیمانجارو بالا برود. از آنجا که شایان اولین نفری است که امسال این قله را می‌زند، قصد دارد در مسیر پرچم‌هایی قرار دهد تا کوهنوردان بعدی مسیر را گم نکنند. اگر پای کوه را ارتفاع ۰ و قله را در ارتفاع ۱ در نظر بگیریم، شایان این ارتفاع را به $n =$ ۱۰۰۰۰۰ قسمت مساوی تقسیم می‌کند. سپس شروع به بالا رفتن از کوه می‌کند. هر موقع که به انتهای یک قسمت می‌رسد به احتمال $p =$ ۰/۲ و مستقل از تصمیمات قبلی یک پرچم در این مکان قرار می‌دهد.

ما نیاز داریم که توزیع فاصله بین پرچم‌های متوالی را بدانیم. از آنجا که تعداد پرچم ها نسبتاً زیاد است و قدرت تحلیل ما محدود، نیاز داریم که شما حرکت شایان را شبیه سازی کنید و سپس توزیع فاصله‌های پرچم‌های متوالی را روی نموداری ترسیم کنید.

الف)

- جایگاه پرچم‌ها را با توجه به توزیعی که از آن پیروی می‌کنند شبیه سازی کنید.
- فاصله جفت پرچم‌های متوالی را به‌دست آورید و در آرایه‌ای ذخیره کنید.
- هیستوگرام فواصل را با توابع پایه R ترسیم کنید.
- هیستوگرام فواصل را با استفاده از ggplot رسم کنید.
- هیستوگرام لگاریتمی فواصل را رسم کنید.
- نمودار رسم شده به چه توزیعی شباهت دارد؟ علت پیروی از این توضیع را شرح دهید.

ب) حال که شایان در تلاش نخست توانست به راحتی قله این کوه را فتح کند، حرکتش بسیار سریع شده است. او این بار ۵۰۰۰۰ پرچم را با توزیع uniform در ارتفاع‌ها (بازه ۰ تا ۱) قرار می‌دهد. مکان پرچم‌ها را بر حسب ارتفاع مرتب کرده و سپس مشابه قبل فاصله پرچم های متوالی را به‌دست آورده و هیستوگرام این فواصل را ترسیم کنید.

- فاصله بین جفت پرچم های متوالی را به‌دست آورید.
- هیستوگرام فواصل را با توابع پایه R ترسیم کنید.
- با استفاده از ggplot فواصل را ترسیم کنید.
- نمودار رسم شده به چه توزیعی شباهت دارد؟ آیا به لحاظ شکل تابع توزیع مشابه با حالت الف) شد؟ آیا می‌توانید توجیهی برای این اتفاق بیابید؟

مسئله‌ی ۲. تماشای آنلاین ویدئو

تصور کنید می‌خواهیم یک ویدیو را به صورت آنلاین تماشا کنیم. تماشای بدون وقفه ویدئو مستلزم دانلود با نرخ ثابت ۱ مگابایت بر ثانیه است، در غیر این صورت ویدئو دچار وقفه شده و شروع به بافر می‌کند. به عبارت دیگر، بافر هنگامی رخ می‌دهد که تمامی بخش‌های دانلود شده را تماشا کرده باشیم و بخش بعدی ویدیو هنوز دانلود نشده باشد، در این صورت باید کمی صبر کنیم تا بخش‌های بعدی دانلود شوند تا بتوانیم دوباره به تماشای ویدیو ادامه دهیم. یک «تجربه خوشایند» در تماشای یک ویدئوی آنلاین را به این صورت تعریف می‌کنیم که در حین تماشا و در هیچ زمانی از ویدئو وقفه‌ای رخ ندهد. همچنین، زمان را در این سوال به صورت گسسته و با واحد ۱ ثانیه در نظر بگیرید.

فرض کنید که سرعت دانلود ما در هر ثانیه یک عدد تصادفی با توزیع یونیفورم بین ۰/۶ و ۱/۴ مگابایت بر ثانیه و مستقل از ثانیه‌های پیشین باشد. در این صورت‌احتمال یک تجربه خوشایند در تماشای یک ویدئو با طول مشخص چقدر است؟

الف) تابعی بنویسید تا با دریافت طول ویدئو به عنوان ورودی، احتمال تماشای بدون وقفه ویدئو را تخمین بزند. به منظور تخمین احتمال، می‌توانید تماشای ویدئو را به تعداد ۱۰۰۰۰ بار شبیه‌سازی کرده و سپس نسبت تعداد تجربیات خوشایند به کل آزمایشات را به جای احتمال در نظر بگیرید.

ب) احتمال یک تجربه خوشایند را به صورت تابعی از طول ویدئو رسم کنید. طول ویدئو را بین ۲ تا ۲۰۰ ثانیه با گام‌های به طول ۱ ثانیه در نظر بگیرید. برای احتمال حداقل ۱۰۰۰۰ شبیه‌سازی انجام دهید تا نمودار همواری مشاهده کنید.

برای کاهش احتمال توقف‌های ناخوشایند در حین تماشای ویدئو می‌توانیم فرآیند بافرینگ را جلوتر آغاز کرده و پس از چند ثانیه شروع به تماشای کلیپ کنیم - کاری که تقریباً همه انجام می‌دهند. به این کار headstart گفته می‌شود. در این روش ابتدا کمی صبر می‌کنیم تا بخشی از ویدیو دانلود شود و سپس به تماشای آن می‌پردازیم. تابعی که در قسمت قبل نوشتید را با افزودن یک متغیر ورودی جدید (تعداد ثانیه‌های headstart) به گونه‌ای تغییر دهید تا امکان پشتیبانی از این استراتژی را داشته باشد. دقت کنید که در طول زمان headstart نرخ دانلود هر ثانیه مشابه قبل از یک توزیع یونیفرم پیروی می‌کند.

ج) در این قسمت فرض کنید که سرعت دانلود ما در هر ثانیه یک عدد تصادفی با توزیع یونیفورم بین ۰/۲ و ۱/۵ مگابایت بر ثانیه و مستقل از ثانیه‌های پیشین باشد. طول ویدئو را همواره ۱۰۰ ثانیه در نظر بگیرید. نمودار احتمال تماشای بدون وقفه را نسبت به مقادیر مختلف برای زمان headstart رسم کنید. می‌توانید مقادیر زمان headstart را بین ۰ تا ۴۰ ثانیه با گام‌های ۱ ثانیه پوشش دهید.

د) با توجه به نتیجه‌ی کدهای فوق در مورد اثرات طول ویدیو و headstart بر احتمال تماشای بدون وقفه توضیح دهید.

مسئله‌ی ۳. تفت (تف؟!) دادن سیب‌زمینی!

به شما n قطعه خرد شده از سیب زمینی داده شده که برای سادگی فرض کنید به شکل مکعب‌های متنظم و هم‌شکل هستند. می‌خواهید تمام سیب زمینی‌ها را به طور کامل و صحیح سرخ کنید: یعنی تمامی وجوه هر قطعه سیب‌زمینی کاملاً سرخ شوند، و در عین حال هیچ وجهی از هیچ قطعه‌ای از آنان نیز نسوخته باشد. برای اینکار یک عدد ماهیتابه در اختیار دارید که می‌توانید تمامی سیب زمینی‌هایتان را در آن قرار داده و روی اجاق گاز حرارت بدهید.

برای سادگی فرض کنید که وقتی وجه خاصی از یک سیب‌زمینی به سمت کف ماهیتابه است، سایر ۵ وجه آن هیچ حرارتی نمی‌بینند. همچنین سیب‌زمینی‌ها روی یکدیگر قرار نمی‌گیرند، و روی پخت یکدیگر اثری نمی‌گذارند. همچنین می‌دانیم که حداقل زمان لازم برای پخت کامل یک وجه از هر سیب‌زمینی ۳۰ ثانیه، و حداکثر زمان قابل تحمل برای هر وجه پیش از سوختن ۹۰ ثانیه است.

الف) ابتدا به سادگی استدلال کنید که با هیچ استراتژی از پختن سیب‌زمینی‌ها، یک پخت کامل و صحیح نمی‌تواند کمتر از ۱۸۰ ثانیه و بیشتر از ۵۴۰ ثانیه به طول بیانجامد.

حال فرض کنید که زمان پخت را میانگین دو مقدار فوق، یعنی ۳۶۰ ثانیه در نظر گرفته‌ایم. سیب‌زمینی‌ها را به صورت تصادفی به کف ماهیتابه ریخته و لذا برای هر کدام یک وجه تصادفی به سمت ماهیتابه قرار گرفته است. بالطبع در صورتی که در کل زمان پخت به همین شکل رهایشان کنیم، یکی از وجوهشان خواهد سوخت و مابقی خام می‌مانند. لذا در طی پخت آنان را f بار در بازه‌های مساوی تفت می‌دهیم. پس از هر بار تفت دادن، یکی از وجوه تصادفی هر سیب‌زمینی با احتمال‌های یکسان، مستقل از سایر سیب‌زمینی‌ها و همچنین وجوه قبلی، به سمت کف ماهیتابه قرار می‌گیرد. می‌تواند همان وجه قبلی نیز باشد. همچنین، برای سادگی فرض کنید که خود زمان لازم برای تفت دادن ناچیز است.

ب) فرض کنید ۲۰ $= n$. در این صورت، احتمال یک پخت کامل و موفق را بر حسب افزایش (تعداد تفت دادن‌ها) تخمین زده و رسم کنید. می‌توانید تعداد تفت دادن‌ها را بین ۴۰ تا ۲۵۰ بار (با گام‌های ۵) انتخاب کنید، اما انتخب نهایی با خودتان است. همچنین، هر شبیه‌سازی را آنقدر تکرار کنید که نمودار هموار و زیبایی حاصل شود.

توجه: مابقی قسمت‌های این سوال نمره ندارند و تصحیح نمی‌شوند، و صرفاً برای دانشجویان کنجکاوی است که قصد دارند بیشتر به این مسئله به ظاهر ساده! فکر کنند.

ج) فرض کنید که احتمال پخت کامل و سالم را با P نمایش دهیم که بالطبع تابعی از n و f است. با استدلال و به صورت کاملاً نظری نشان دهید:

$$P = [g(f)]^n$$

که در این رابطه تابع g فقط به f و سایر پارامترهای مسئله (غیر از n) مرتبط است.

د) قصد داریم تابع $g(f)$ را تخمین بزنیم. شبیه‌سازی بخش ب) را با ۱ $= n$ تکرار کنید، و مقادیر حاصله برای $P -$ ۱ را به صورت نیم‌لگاریتمی رسم کنید. شبیه به چه تابعی است؟ با استفاده از محاسبه شیب خط در حالت نیم‌لگاریتمی فوق، تابع $g(f)$ را تقریب بزنید. در این صورت یک تقریب برای کل P (حتی وقتی ۱ $\neq n$ نیز داریم).

ه) با فرض برقراری تقریب قسمت د)، به صورت نظری نشان دهید که با افزایش تعداد قطعات سیب‌زمینی n تعداد تفت‌های لازم f، به منظور رسیدن به یک احتمال مشخص (مثلاً ۹۹ درصد) برای پخت کامل و صحیح به صورت لگاریتمی افزایش پیدا می‌کند: یعنی . $f_{necessary} = O(\log n)$

و) اگر وقت و حوصله داشتید و هنوز هم کنجکاو بودید، نمودار قسمت ب) را برای مقادیر مختلف n تکرار کنید. اما به جای رسم احتمال پخت کامل و صحیح بر حسب f ، آن را بر حسب مقدار نرمالیزه شده $\frac{f}{\log n}$ رسم کنید. با افزایش n چه پدیده عجیبی را مشاهده می‌کنید؟ (به این رخداد در آمار و احتمال ابعاد بالا، اصطلاحاً «تغییر فاز» یا phase transition نیز گفته می‌شود.)

موفق باشید :)