باسمه تعالی

آمار و احتمال مهندسی

نیم‌سال اول ۹۵-۹۴

دکتر مطهری

دانشکده مهندسی کامپیوتر

تمرین سوم | متغیر تصادفی | زمان تحویل: ۹۴/۸/۱۶

مسایل

۱ مساله اول

طبق قانون بنفورد در هر مجموعه داده عادی، که به میزان لازم گستردگی داشته باشد، اولین رقم هر داده به احتمال ٪۳۰ برابر یک و به احتمال ٪۱۸ برابر دو می‌باشد. به طور عمومی این قانون توزیع احتمالی رقم اول را به صورت زیر ارائه می‌دهد:

$$\mathbb{P}\{D=i\} = \log_{۱۰}\left(\frac{j+۱}{j}\right) \qquad for \quad j \in \{۱, ۲, \cdots, ۹\}$$

- تحقیق کنید که این توزیع احتمال معتبر است.
- در قسمت‌های بعدی سعی می‌کنیم تا حدی این قانون را توجیه کنیم. فرض کنید داده‌ها به صورت نماد علمی $X \cdot ۱۰^N$ نوشته شده باشند که $۱ \leq X < ۱۰$.

  هم‌چنین فرض کنید جزء غیر صحیح لگاریتم داده، یک متغیر تصادفی یکنواخت در بازه صفر تا یک باشد. با استفاده از این اطلاع، تابع چگالی توزیع احتمال متغیر X را بیابید.
- با استفاده از نتیجه قسمت قبل، قانون بنفورد را به دست آورید.

۲ مساله دوم

شایان و فاطمه در حال بازی گل یا پوچ با یکدیگر می‌باشند. شایان با احتمال $p_۱$ گل را انتخاب می‌کند و فاطمه با احتمال $p_۲$. برنده‌ی بازی اولین شخصی است که مکان گل را در دستان دیگری درست حدس بزند. اگر بازی با گل در دستان فاطمه آغاز شود، چقدر احتمال دارد که شایان برنده شود؟

۳ مساله سوم

برای متغیر تصادفی گسسته X که فضای نمونه آن اعداد طبیعی است، نشان دهید:

$$\mathbb{E}\{X\} = \sum_{k=۱}^{\infty} \mathbb{P}\{X \geq K\}$$

۴ مساله چهارم

با شروع باران در یک ناحیه شمالی، به طور میانگین ۲۰ قطره باران در هر $cm^۲$ در دقیقه به زمین برخورد می‌کند. فرض کنید تعداد قطرات باران برخوردی در یک ناحیه $۵cm^۲$ را با متغیر تصادفی X نمایش دهیم.

- یک توزیع مناسب برای این متغیر انتخاب نمایید و دلیل این انتخاب را توضیح دهید.
- با کمک پاسخ قسمت قبل، احتمال این را بیابید که در چنین ناحیه‌ای در مدت سه ثانیه، هیچ قطره‌ی بارانی به زمین برخورد نکند.

۵ مساله پنجم

شایان و فاطمه در شهری زندگی می‌کنند که ۱۰۰۰ نفر جمعیت دارد و هر کدام از آن‌ها تعداد ۵۰ نفر از این افراد را می‌شناسند. فرض کنید آشنایان هر کدام از شایان و فاطمه، به طور کاملا تصادفی و یکنواخت توزیع شده باشند و آشنایان شایان و فاطمه از هم مستقل باشند. یعنی دانستن این که فردی با فاطمه آشنا است، هیچ تاثیری در احتمال آشنایی او با شایان ندارد.

طبیعتا ممکن است هردوی آن‌ها فکر کنند که چون کمتر از ٪۵ جمعیت را می‌شناسند، احتمال بسیار کمی دارد که اصلا دوست مشترکی داشته باشند.

- میانگین تعداد دوستان مشترک شایان و فاطمه را بیابید.
- توزیع متغیر تصادفی تعداد دوستان مشترک شایان و فاطمه را بیابید.

۶ مساله ششم

فرض کنید X و Y دو متغیر تصادفی با توزیع $Pois(\lambda)$ باشند. هم‌چنین این دو متغیر مستقل نباشند و حتی با همدیگر برابر باشد. در این صورت آیا می‌توان ادعا کرد که حاصل‌جمع این دو، یک متغیر تصادفی با توزیع پواسون و نرخ ۲λ است؟ از انجام هرگونه محاسبه ریاضی برای حل این سوال بپرهیزید.

۷ مساله هفتم

به صورت کاملا تصادفی و یکنواخت، k توپ مختلف را در جعبه‌هایی به شماره ۱ تا n قرار می‌دهیم. چند توپ مختلف، ممکن است درون یک جعبه قرار بگیرند.

- میانگین تعداد جعبه‌های خالی را بیابید.
- میانگین تعداد توپ‌های قرار داده شده در سه جعبه اول را بیابید (فرض کنید $n > ۳$ می‌باشد).

شبیه سازی

۸ سوال اول

فرض کنید لامپ‌هایی داریم که طول عمر هر کدام از آن‌ها یک متغیر توانی با پارامتر $\lambda = ۱$ می‌باشد. فرض کنید در زمان صفر یک لامپ روشن می‌کنیم و به محض این که این لامپ سوخت، لامپ دیگری را جایگزین آن می‌کنیم و این کار را ادامه می‌دهیم. به این روش دنباله $t_۱, t_۲, t_۳, \ldots$ به دست می‌آید که t_i زمان سوختن لامپ iام را نمایش می‌دهد. با شبیه‌سازی دنباله‌های زیادی از این دست به سوالات زیر پاسخ دهد:

۱. این شبیه سازی را از زمان ۱ تا t بارها انجام دهید. در فایل "p11" میانگین تعداد لامپ‌های استفاده شده را برای $t = ۱۰, ۱۰۰, ۲۰۰, ۵۰۰, ۱۰۰۰$ بنویسید.

نکته: برای t های مختلف تعداد دفعات اجرای شبیه سازی را به گونه ای انتخاب کنید که میانگینی که به دست می‌آورید، در اجراهای مختلف خیلی تغییر نکند. به طور کلی هر وقت می‌خواهید عددی را گزارش کنید، سعی کنید تعداد تکرارها به حدی باشد که در اجراهای مختلف برنامه عدد محاسبه شده تغییرات زیادی نداشته باشد.

۲. با شبیه‌سازی تعداد زیادی لامپ، احتمال‌های زیر را محاسبه کنید:

(a) $\mathbb{P}[r > 1]$
(b) $\mathbb{P}[r > 2]$
(c) $\mathbb{P}[r > 3]$
(d) $\mathbb{P}[r > 2|r > 1]$
(e) $\mathbb{P}[r > 3|r > 1]$
(f) $\mathbb{P}[r > 3|r > 2]$
(g) $\mathbb{P}[r > 4|r > 1]$
(h) $\mathbb{P}[r > 4|r > 2]$
(i) $\mathbb{P}[r > 4|r > 3]$

که r بیانگر عمر لامپ است. با توجه به اعداد به دست آمده چه ویژگی مهمی در مورد عمر لامپ‌ها می‌توان گفت؟ احتمال‌های محاسبه شده را در فایل "p12" بنویسید.

۳. حال بازه‌ی زمانی $[۰, T_{۶۱}]$ را در نظر بگیرید. دقت داریم که در بار شبیه سازی این بازه متغیر است. موارد زیر را محاسبه کرده و در فایل "p12" بنویسید.

- امید ریاضی تعداد نقاط در $[۰, t_{۶۱}]$
- واریانس تعداد نقاط در $[۰, t_{۶۱}]$
- امید ریاضی تعداد نقاط در $[۰, \frac{t_{۶۱}}{۲}]$
- واریانس تعداد نقاط در $[۰, \frac{t_{۶۱}}{۲}]$
- امید ریاضی تعداد نقاط در $[\frac{t_{۶۱}}{۳}, \frac{۲t_{۶۱}}{۳}]$
- واریانس تعداد نقاط در $[\frac{t_{۶۱}}{۳}, \frac{۲t_{۶۱}}{۳}]$

با توجه به موارد محاسبه شده آیا می‌توان نتیجه‌ی جالبی گرفت؟!

۴. برای $t = ۱۰۰$ زمان را به بازه‌های با طول ۴ تقسیم کنید و نمودار تعداد لامپ‌های سوخته در هر بازه را رسم کنید(برای این قسمت کافی است فقط یک بار شبیه سازی انجام شود). حال شبیه سازی را در فاصله‌ی زمانی ۰ تا $t = ۱۰^۶$ انجام دهید. این زمان را به بازه‌های ده تایی تقسیم کنید و تعداد لامپ‌های سوخته در هر بازه را محاسبه نمایید. هیستوگرام مربوط به تعداد لامپ‌های سوخته شده در هر بازه‌ی به طول ۱۰ را با دو پارامتر `break=30` و `freq=FALSE` رسم نمایید. حال کد زیر را اجرا نمایید و از مشاهدات خود در این بخش تمرین نتیجه گیری کنید.

```
points(0:30,dbinom(0:30,1000,0.01),type = "p",col="red")
```

۹ سوال دوم-امتیازی

یک مدل ساده و معروف برای بررسی ژن‌ها در جمعیت، مدل رایت-فیشر(Wright–Fisher) می‌باشد. در این مدل فرض می‌شود که در ابتدا یک جمعیت داریم که هر کدام از افراد آن ژنوم مخصوص به خود را دارد، یعنی در ابتدا n موجود و n ژنوم مختلف داریم. نسل بعدی طبق این مدل به این صورت است که تعداد افراد ثابت در نظر گرفته می‌شود و هر کدام از افراد نسل جدید ژنوم خود را با احتمال برابر و مستقل از دیگر افراد، از یکی از افراد نسل قبل می‌گیرد. به این صورت احتمال دارد که برخی از ژنوم‌ها در نسل بعدی مشاهده نشوند.

برای مثال فرض کنید در ابتدا دو نفر موجود هستند که دو ژنوم مختلف a و b را دارا می‌باشند. برای نسل بعدی چهار حالت متصور است. اول آنکه نفر اول ژنوم a و نفر دوم ژنوم b را داشته‌باشد. دوم آن‌که نفر اول ژنوم b و نفر دوم ژنوم a را داشته‌باشد. سوم آن‌که هر دو دارای ژنوم a باشند و چهارم آن‌که هر دو دارای ژنوم b باشند. هر کدام از این حالات به احتمال ۰/۲۵ رخ می‌دهد. علاوه بر این به احتمال ۰/۵ در نسل دوم تنها یک ژنوم باقی می‌ماند.

می‌توان ثابت کرد که با احتمال ۱ بعد از گذشت زمان کافی تنها یکی از ژنوم‌ها باقی خواهد ماند و هر کدام از ژنوم‌های اولیه با احتمال برابر ممکن است این ژنوم باقیمانده باشند.

۱. برنامه‌ای بنویسید که بتواند این مدل را برای جمعیت های مختلف شبیه‌سازی کند و با استفاده از آن میانگین زمان لازم برای آن‌که تنها یک ژنوم در جمعیت باقی بماند را محاسبه کند. این زمان را برای جمعیت‌های ۱۰و۱۰۰و۱۰۰۰و۱۰۰۰۰ نفری محاسبه کنید و به ترتیب در خط‌های فایل "p41" بنویسید. همچنین نمودار مناسبی برای جمعیت‌های ۱۰ تا ۱۰۰۰ نفری رسم کنید و در فایل گزارش خود قرار دهید.

۲. یک جمعیت هزار نفری را در نظر بگیرید.پیش از آنکه تنها یک ژنوم در جمعیت باقی بماند با احتمال خیلی زیاد دقیقا دو ژنوم وجود داشته‌است.(احتمال آن‌که در این روش در صورت وجود سه ژنوم یا بیشتر در جمعیت به یک جمعیت با تنها یک ژنوم برسیم بسیار اندک است) میانگین مدت زمان آن‌که در جمعیت تنها دو ژنوم وجود داشته‌است چقدر می‌باشد؟ میانگین مدت زمان آن‌که در جمعیت تنها سه، چهار، پنج، ده و یا بیست ژنوم موجود باشد را به ترتیب در خط‌های مختلف فایل "p42" بنویسید و هیستوگرام مقادیر مختلف آن را رسم کنید و در گزارش خود بیاورید. فکر می‌کنید این نمودارها بیانگر چه نوع توزیع‌هایی هستند؟