آمار و احتمالات مهندسی

تمرین سوم - متغیر تصادفی، میانگین و واریانس

محمد و ثمر

تاریخ تحویل ۱۴۰۱/۰۸/۲۷

سؤال ۱.

برای هر یک از متغیرهای زیر نوع توزیع را مشخص کنید.

الف) یک برنامه‌ی نظارت شبکه به طور مداوم درخواست‌ها را از طریق یک شبکه ارسال می‌کند. گاهی اوقات این درخواست‌ها به دلیل مشکلات شبکه با شکست مواجه می‌شوند و باید دوباره ارسال شوند. متغیر تصادفی A، تعداد عملیات ارسال مجدد را در یک ساعت نشان می‌دهد.

ب) یک مهندس ایمنی در حال بررسی حوادث صنعتی در یک کارخانه است. او فکر می‌کند که ۴۰ درصد حوادث، ناشی از عدم رعایت دستورالعمل‌ها توسط کارکنان است. برای آزمایش این تئوری، گزارش‌های حوادث به‌طور تصادفی انتخاب می‌شوند تا زمانی که گزارشی پیدا شود که حادثه‌ی ذکرشده‌اش ناشی از عدم رعایت روش‌های ایمنی کارمند است. متغیر تصادفی B تعداد گزارش تا قبل از رسیدن به مورد عدم رعایت دستورالعمل‌ها است.

پ) فرض کنید یک درمان جدید برای نوعی سرطان ارائه می‌شود که ۱۰ درصد احتمال موفقیت دارد. ۵۰۰ بیمار به طور آزمایشی دارو را مصرف کرده‌اند. متغیر تصادفی C تعداد بیماران بهبود یافته است.

ت) یک حافظه‌ی خالی از رم کامپیوتر ایجاد می‌شود. از بین n مکان حافظه، k عدد اختصاص داده شده و بقیه آزاد هستند؛ همچنین هر دو مکان یک احتمال مساوی برای اختصاص داده شدن داشته‌اند. متغیر تصادفی D نشان‌دهنده‌ی طول یک بلوک پیوسته از حافظه آزاد است.

ث) قضیه‌ی اعداد اول بیان می‌کند که برای n به اندازه‌ی کافی بزرگ، احتمال آنکه یک عدد صحیح با $n^۲$ رقم، عدد اول باشد برابر است با $\frac{1}{n^۲}$. فرض کنید می‌خواهیم یک عدد اول بزرگ با $n^۲$ رقم پیدا کنیم و این کار را با انتخاب یک عدد $n^۲$ رقمی و سپس چک کردن اول بودنش انجام می‌دهیم. متغیر تصادفی E تعداد اعداد صحیحی است که تا قبل از رسیدن به جواب امتحان می‌کنیم.

ج) فروشگاه لباس فروشی که در شهری کوچک و سرد واقع شده است، از ساعت ۶ تا ۲۰ باز است. مدیر این فروشگاه می‌خواهد برای باز بودن فروشگاه در نوبت شب (ساعت ۲۰ تا ۶) تصمیم بگیرد. برای این کار به مدت یک هفته فروشگاه را در نوبت شب نیز باز نگه داشته و میزان فروش در نوبت شب را محاسبه می‌کند. متغیر تصادفی F، تعداد فروش در نوبت شب است.

پاسخ .

الف) دارای توزیع پوآسون است؛ در واقع ارسال مجدد پیشامد نادری است که در بازه‌ی مشخص یک ساعته اتفاق می‌افتد.
ب) دارای توزیع هندسی است؛ ما تعداد گزارش‌ها تا رسیدن به اولین موفقیت (یعنی اولین مورد ناشی از عدم رعایت دستورالعمل) را می‌شماریم.
پ) دارای توزیع دوجمله‌ای است؛ هر بیمار با احتمالی بهبود پیدا کرده یا نکرده است. ما تعداد بهبودیافتگان را می‌شماریم.
ت) دارای توزیع هندسی است؛ در واقع تعداد مکان‌های آزاد از اولین مکان آزاد تا رسیدن به اولین مکان اختصاص‌یافته مد نظر است.
ث) دارای توزیع هندسی است؛ در واقع ما تعداد اعداد صحیح تا رسیدن به اولین موفقیت (یعنی اول بودن) را می‌شماریم.
ج) دارای توزیع پوآسون است؛ در واقع فروش لباس در نوبت شب اتفاقی با احتمال کم است که در بازه‌ی مشخص ساعات نوبت شب رخ

می‌دهد. (البته لازم به ذکر است که برای مثال فروش یک فروشگاه در راه‌آهن یا فرودگاه حتی در شب نیز به حدی است که دیگر نمی‌توان آن را کم دانست و درنتیجه استفاده از توزیع پوآسون برای آن مناسب نمی‌باشد.)

سؤال ۲.

در دوران همه‌گیری ویروس کرونا، افراد تمایل به دوری از جمعیت‌های زیاد دارند و دوست دارند به خصوص در فضاهای سربسته مانند سالن‌های سینما کنار جمعیت کمتری باشند. مدیر یک مجموعه تفریحی که سینما هم دارد، این نکته را با مسئول فروش بلیت‌ها در میان می‌گذارد. مسئول فروش سانس‌های مختلف را چک می‌کند. می‌گوید : "به طور مثال برای ساعت ۱۷، هر یک از ۴ سالن مجموعه، به ترتیب ۲۰، ۲۳، ۱۷ و ۱۰۰ تماشاچی دارند. در نتیجه به طور میانگین در هر سالن ۴۰ تماشاچی هستند و با توجه به ۱۲۰ نفره بودن سالن‌ها این تعداد یک سوم ظرفیت سالن است و جمعیت زیاد نیست. برای بقیه‌ی سانس‌ها هم حدودا به همین شکل است." اما آن شب تعداد زیادی تماس با قسمت انتقادات و پیشنهادات مجموعه گرفته می‌شود و از جمعیت زیاد سالن سینما در دوران همه‌گیری ویروس کرونا در سانس‌هایی از جمله سانس ساعت ۱۷ شکایت می‌شود.
الف) آیا مسئول فروش در اعلام تعداد میانگین جمعیت هر سالن و پیش‌بینی احساس تماشاچیان اشتباه کرده است؟ اگر خیر، دلیل احساس تماشاچیان چیست و اگر بله، تعداد میانگین درست برای سانس ساعت ۱۷ چقدر است؟
ب) احتمالا تماس‌های مربوط به سانس ساعت ۱۷ بیشتر از تماشاچیان کدام سالن‌ها بوده است؟
پ) به نظر شما با چه ترکیب جمعیتی در سالن‌های سینما، با همین روش استدلال و میانگین‌گیری مسئول فروش، نظر تماشاچیان مانند پیش‌بینی مسئول فروش می‌بود؟ (منظور این است که در چه صورتی اگر مسئول فروش پیش‌بینی شلوغ یا خلوت بودن سالن‌ها را می‌کرد، تماشاچیان هم همان احساس را می‌کردند؟ همچنین در این بخش صرفا استفاده از مفهوم هم کافی است اما در صورت نیاز می‌توان از فرمول‌ها هم برای توضیح بهتر پاسخ کمک گرفت.)

پاسخ .

الف) در واقع مسئول فروش مسئله را از نگاه تماشاچیان ندیده است، بلکه از نگاه سالن‌ها دیده است! یعنی حساب کرده که به طور میانگین هر سالن چند تماشاچی دارد، نه اینکه هر تماشاچی چند نفر در کنار خود احساس می‌کند. با توجه به تعداد تماشاچیان در هر سالن، تماشاچیان حس می‌کنند که جمعیت به اندازه‌ای است که در سالن خود آن‌هاست و از باقی سالن‌ها اطلاعی ندارند. همچنین شایان ذکر است که به خاطر زیاد بودن تعداد تماشاچیان در سالن‌های با جمعیت زیاد، احتمال اینکه کسی احساس شلوغی کند هم بیشتر می‌شود. (یعنی به طور مثال در سانس ساعت ۱۷، $\frac{۱۰۰}{۱۲۰}$ مجموع کل جمعیت در ۴ سالن احساس می‌کنند که به خاطر حضور ۹۹ نفر در اطراف خودشان در یک سالن ۱۲۰ نفره، در فضای شلوغی هستند.) پس در میانگین هم این عوامل تاثیر خواهند گذاشت و روش میانگین‌گیری متفاوت است. برای محاسبه‌ی میانگین بهتر داریم:

$$E[X] = ۲۰ \times \frac{۲۰}{۱۶۰} + ۲۳ \times \frac{۲۳}{۱۶۰} + ۱۷ \times \frac{۱۷}{۱۶۰} + ۱۰۰ \times \frac{۱۰۰}{۱۶۰} = ۷۰٫۱۱۲۵$$

ب) با توجه به نکات پاسخ قسمت الف، احتمالا بیشتر تماس‌ها از طرف تماشاچیان سالن ۱۰۰ تایی سانس ساعت ۱۷ بوده است.
پ) اگر تعداد تماشاچیان در سالن‌ها نزدیک هم باشد، چه از نگاه سالن‌ها (روش میانگین‌گیری مسئول فروش بلیت) و چه از نگاه تماشاچیان (روش میانگین‌گیری پاسخ قسمت الف)، تقریبا میانگین تماشاچیان یکسان می‌شود. اما در مثال‌هایی مانند مثال ما که تعداد تماشاچیان در بعضی سالن‌ها با هم تفاوت زیادی دارد یا پراکندگی پاسخ‌ها زیاد است، این دو روش میانگین‌گیری با توجه به نکاتی که در پاسخ قسمت الف گفته شد به نتایج مختلفی می‌انجامند.

سؤال ۳.

یک جفت سکه را بطور همزمان پرتاب می‌کنیم. احتمال شیر آمدن هر سکه p هست.

الف)اگر این دو را همزمان با هم n بار پرتاب کنیم و X مجموع تعداد پرتاب‌هایی باشد که نتایج این سکه با هم متفاوت‌اند، X از چه توزیع احتمالی پیروی می‌کند؟

ب) فرض کنید اگر نتایج این دو سکه متفاوت باشند، شما ۱ دلار جایزه می‌گیرید و در غیر آن صورت ۱ دلار از دست می‌دهید. اگر Y کل سود شما باشد، Y را برحسب X بدست آورید. امید ریاضی و واریانس Y چقدر است؟

پاسخ .

الف) احتمال متفاوت بودن نتیجه پرتاب دو سکه معادل است با:

$$P(HT) + P(TH) = ۲p(۱ - p)$$

می‌دانیم اگر یک آزمایش تصادفی با دو خروجی موفقیت و شکست را n بار تکرار کنیم، متغیر تصادفی X که برابر تعداد موفقیت‌ها در n آزمایش است، از توزیع دوجمله‌ای پیروی می‌کند. بنابراین X یک توزیع دوجمله‌ای به شکل زیر می‌باشد:

$$X \sim Bin(n, ۲p(۱ - p))$$

ب)

ابتدا رابطه بین Y و X را پیدا می‌کنیم:

$$Y = (۱) \times (X) + (-۱) \times (n - X) = X - (n - X) = ۲X - n$$

می‌دانیم میانگین توزیع $X \sim Bin(n, p)$ ب رابر np است؛ پس داریم:

$$E[X] = n \times ۲p(۱ - p)$$

$$E[Y] = E[۲X - n] = ۲E[X] - n = ۴np(۱ - p) - n$$

هم‌چنین می‌دانیم واریانس توزیع $X \sim Bin(n, p)$ برابر npq $(q = ۱ - p)$ است؛ در نتیجه داریم:

$$Var[X] = n \times ۲p(۱ - p) \times [۱ - ۲p(۱ - p)] = ۲np - ۶np^۲ + ۸p^۳ - ۴p^۴$$

$$Var[Y] = Var[۲X - n] = ۴ \times Var[X] = ۴n \times ۲p(۱ - p) \times [۱ - ۲p(۱ - p)] = ۸np - ۲۴np^۲ + ۳۲p^۳ - ۱۶p^۴$$

سؤال ۴.

۲ نامزد انتخابات کارگروه آموزشی دانشکده در حال پخش کردن برگه‌های تبلیغات خود در بین ۲۰۰۰ نفر از دانشجویان هستند. با توجه به محدودیت زمان و هزینه، هر نامزد توانسته در بخش انتشارات تنها ۲۰۰ برگه تبلیغ پرینت بگیرد؛ سپس به طور تصادفی و با احتمال یکسان آن‌ها را بین ۲۰۰ نفر از ۲۰۰۰ دانشجو پخش کرده است. فرض کنید دانستن اینکه چه کسانی تبلیغ نامزد اول را دریافت کرده‌اند، هیچ اطلاعاتی در مورد افرادی که تبلیغ نامزد دوم را دریافت کرده‌اند نمی‌دهد. شایان ذکر است که از دید دو نامزد انتخابات، بعید است که یک دانشجو برگه‌ی تبلیغ هردو را گرفته باشد و با اطلاعات کافی از هر دو نامزد رای بدهد.

الف) توزیع تعداد دانشجویانی که توانسته‌اند بعد از دریافت و مطالعه‌ی هر دو تبلیغ و با اطلاعات کافی رای بدهند از چه نوع است؟چرا؟ (متغیر تصادفی تعداد دانشجویان با این شرایط را X در نظر بگیرید.)

ب) تابع جرمی احتمال X را بیابید.

پ) امید ریاضی X را با استفاده از متغیرهای شاخص بیابید.

پاسخ .

الف) تعداد ۲۰۰ نفر از ۲۰۰۰ دانشجو، تبلیغ نامزد اول را دریافت کرده‌اند. همچنین می‌توان گفت که ۲۰۰ تبلیغ نامزد دوم نیز، بین x نفر از بین ۲۰۰ دانشجویی که تبلیغ نامزد اول را دریافت کرده‌اند به علاوه‌ی $x - ۲۰۰$ نفر از بین ۱۸۰۰ دانشجویی که تبلیغ نامزد اول را دریافت نکرده‌اند پخش می‌شود. در واقع چون تعداد تبلیغات هر نامزد مقدار ثابت ۲۰۰ است، اینکه هر فرد تبلیغ را دریافت کند یا نکند مستقل از فرد دیگر نیست

و گویی ما یک نمونه‌گیری بدون جایگذاری داریم. پس توزیع ما از نوع فوق هندسی است.

ب) اگر x تعداد افرادی باشد که هر دو تبلیغ را دریافت کرده‌اند، با توجه به استدلال پاسخ بخش الف باید ابتدا از ۲۰۰ دانشجویی که تبلیغ نامزد اول را دیده‌اند تعداد x نفر را انتخاب کنیم و سپس از ۱۸۰۰ دانشجوی دیگر تعداد $x - ۲۰۰$ نفر را انتخاب کنیم تا در نهایت ۲۰۰ تبلیغ نامزد دوم پخش شده باشد. ما تعداد راه‌های ممکن این دو انتخاب را در هم ضرب می‌کنیم تا تعداد کل راه‌هایی که نامزد دوم می‌تواند تبلیغ پخش کند را بدست آوریم. در نهایت حاصل را بر کل تعداد راه‌هایی که نامزد دوم می‌تواند بین کل ۲۰۰۰ دانشجو تبلیغ پخش کند تقسیم می‌کنیم تا احتمال بدست آید.

$$P(X=x) = \begin{cases} \frac{\binom{۲۰۰}{x}\binom{۲۰۰۰-۲۰۰}{۲۰۰-x}}{\binom{۲۰۰۰}{۲۰۰}} & , \quad ۰ \leq x \leq ۲۰۰ \\ ۰ \quad , & otherwise \end{cases}$$

پ) متغیر تصادفی شاخص I_jرا برای شخص jام بدین صورت تعریف می‌کنیم که آیا این شخص هردو تبلیغ را دریافت کرده یا خیر. احتمال اینکه دانشجوی خاصی تبلیغ هر دو نامزد را دریافت کرده باشد $(\frac{۲۰۰}{۲۰۰۰})^۲$ است؛ زیرا دریافت تبلیغ هریک از دو نامزد اتفاقات مستقلی هستند. پس با تعریف $X = \sum_{j=۱}^{n} I_j$ و با توجه به خطی بودن رابطه‌ی امید ریاضی، داریم :

$$E[X] = E[\sum_{j=۱}^{n} I_j] = \sum_{j=۱}^{۲۰۰۰} E[I_j] = \sum_{j=۱}^{۲۰۰۰} P(I_j = ۱) = ۲۰۰۰ \times (\frac{۲۰۰}{۲۰۰۰})^۲ = ۲۰$$

لازم به توضیح است که برای متغیر شاخص داریم: $E[I] = ۱ \times P(I=۱) + ۰ \times P(I=۰) = P(I=۱)$.

سؤال ۵.

هر یک از اعضای یک گروه n نفره، مستقل از یکدیگر، تاس سالمی را پرتاپ می‌کنند. به ازای هر k فردی که طی پرتاب تاس‌هایشان عدد مشابهی مشاهده کنند، k امتیاز به گروه داده می‌شود. به عنوان مثال در یک گروه ۱۱ نفره ($n = ۱۱$)، اگر پس از پرتاب تاس‌ها سه بازیکن عدد ۲، چهار بازیکن عدد ۵، یک بازیکن عدد ۱، یک بازیکن عدد ۳، یک بازیکن عدد ۴ و یک بازیکن عدد ۶ را مشاهده کنند، ۷ امتیاز (یعنی $۳ + ۴$ امتیاز) به گروه تعلق می‌گیرد. میانگین امتیاز گروه را بر حسب n حساب کنید.

پاسخ .

فرض کنید متغیر تصادفی X_i، تعداد افرادی را نمایش دهد که تاس آن‌ها i آمده است. با توجه به سالم بودن تاس، این تاس با احتمال $\frac{۱}{۶}$عدد i را نشان می‌دهد و با احتمال $\frac{۵}{۶}$ این عدد را نشان نمی‌دهد. پس نوع توزیع آن دوجمله‌ای است.

$$X_i \sim Bin(n,p) \quad , \quad p = \frac{۱}{۶} \quad \rightarrow \quad X_i \sim Bin(n, \frac{۱}{۶})$$

اگر امتیاز ناشی از تعداد افرادی که تاس آن‌ها i آمده است را با Y_i نشان دهیم، تنها زمانی امتیاز کسب می‌کنیم که بیش از یک بار این مقدار ظاهر شود، پس داریم:

$$Y_i = \begin{cases} X_i \quad , & X_i > ۱ \\ ۰ \quad , & X_i = ۰,۱ \end{cases}$$

می‌دانیم $E[X_i] = \sum_{k=۰}^{n} k \times P(X_i = k) = np$، پس با توجه به عبارت بالا و صفر بودن مقدار Y_i برای $X_i = ۰,۱$ ، باید این دو حالت را از امید ریاضی Y_i کم کنیم. داریم :

$$E[Y_i] = E[X_i] - (۱ \times P(X_i = ۱) + ۰ \times P(X_i = ۰)) = n \times (\frac{۱}{۶}) - (۱ \times \binom{n}{۱} (\frac{۱}{۶})^۱ (\frac{۵}{۶})^{n-۱} + ۰) = \frac{n}{۶}(۱ - (\frac{۵}{۶})^{n-۱})$$

امتیاز کل برابر است با:

$$Y = Y_۱ + Y_۲ + Y_۳ + Y_۴ + Y_۵ + Y_۶$$

با توجه به خطی بودن امید ریاضی داریم:

$$E[Y] = E[Y_۱ + ... + Y_۶] = ۶E[Y_i] = n(۱ - (\frac{۵}{۶})^{n-۱})$$

سؤال ۶.

فرض کنید که $X_۱, X_۲, ..., X_n$ متغیرهای تصادفی مستقل با توزیع یکنواخت در بازه (a, b) باشند.

$$X_i \sim Uniform(a, b) \quad , \quad a < b$$

الف) فرض کنید $Y = min(X_۱, X_۲, ..., X_n)$، تابع توزیع انباشته Y را بیابید.

ب) امید ریاضی متغیر تصادفی Y را محاسبه کنید.

پاسخ .

الف) می‌دانیم تابع توزیع انباشته به صورت زیر تعریف می‌شود:

$$F_X(a) = Prob\{X \leqslant a\}$$

پس داریم:

$$F_Y(y) = P(Y \leqslant y) = ۱ - P(Y > y) = ۱ - P(min(X_۱, X_۲, ..., X_n) > y)$$

رخداد $min(X_۱, X_۲, ..., X_n) > y$ زمانی اتفاق می‌افتد که برای هر X_i داشته باشیم $X_i > y$. با توجه به اینکه متغیرهای تصادفی X_i از هم مستقل‌اند، خواهیم داشت:

$$F_Y(y) = ۱ - P(min(X_۱, X_۲, ..., X_n) > y) = ۱ - P(X_۱ > y)P(X_۲ > y)...P(X_n > y)$$
$$= ۱ - (P(X_i > y))^n$$

حال کافی است احتمال $P(X_i > y)$ را محاسبه کنیم.

$$P(X_i \leqslant y) = F_{X_i}(y) = \begin{cases} ۰ & y \leqslant a \\ \frac{y-a}{b-a} & a < y < b \\ ۱ & y \geqslant b \end{cases} \Longrightarrow P(X_i > y) = \begin{cases} ۱ & y \leqslant a \\ ۱ - \frac{y-a}{b-a} & a < y < b \\ ۰ & y \geqslant b \end{cases}$$

در نتیجه:

$$F_Y(y) = ۱ - (P(X_i > y))^n = \begin{cases} ۰ & y \leqslant a \\ ۱ - (۱ - \frac{y-a}{b-a})^n & a < y < b \\ ۱ & y \geqslant b \end{cases}$$

ب)

$$E[Y] = \int y f_Y(y)dy = \int y(\frac{dF_Y(y)}{dy})dy = \int_a^b y\frac{n}{b-a}(\frac{b-y}{b-a})^{n-۱}dy = \frac{n}{(b-a)^n}\int_a^b y(b-y)^{n-۱}dy$$

انتگرال جز به جز می‌گیریم:

$$E(Y)=\frac{n}{(b-a)^n}(\frac{-y}{b}(b-y)^n+\int\frac{1}{n}(b-y)^n dy)=\frac{n}{(b-a)^n}[\frac{-(b-y)^n(ny+b)}{n(n+1)}]_a^b$$
$$=\frac{n}{(b-a)^n}[0+\frac{(b-a)^n(na+b)}{n(n+1)}]=\frac{na+b}{n+1}$$

سؤال ۷.

تانک‌های ارتش یک کشور از ۱ تا N شماره‌گذاری شده‌اند. این کشور در یک جنگ، n عدد از تانک‌های خود را از دست می‌دهد و تانک‌ها به دست دشمن می‌افتد. دشمن درمی‌یابد که تانک‌های تصاحب‌شده شماره‌گذاری شده‌اند.
الف) دشمن چگونه می‌تواند تخمینی برای تعداد کل تانک‌های این کشور (یعنی N) به دست آورد؟
ب) اگر دشمن ۱۲ تانک تصاحب کند و بیشترین و کمترین شماره‌های تصاحب شده ۲ و ۱۱۷ باشند، تخمین تعداد کل تانک‌های این کشور چقدر است؟

پاسخ .
الف) فرض کنید $X_1, \ldots X_n$ شماره تانک‌های تصاحب شده باشند که به طور کاملا تصادفی از بین N تانک انتخاب شده‌اند. متغیر تصادفی $Y=max(X_i)$ را در نظر بگیرید. داریم :

$$P(Y=k)=\frac{\binom{k-1}{n-1}}{\binom{N}{n}}\quad,\quad k=n,n+1,\ldots,N$$

در واقع می‌توان گفت که اگر n تانک تصاحب شوند، تانکی که از بین این n تانک بیشترین شماره یعنی k را دارد حتما شماره‌ای بزرگتر یا مساوی n دارد. $(k=n,\ldots,N)$ در مورد احتمال هم می‌توان گفت که اگر بیشترین شماره در تانک‌های تصاحب‌شده برابر k باشد، باید $n-1$ تانک تصاحب‌شده‌ی دیگر را از بین تانک‌های با شماره‌ی کوچک‌تر یا مساوی $k-1$ انتخاب کنیم. در نهایت نیز حاصل این انتخاب را باید بر تعداد کل راه‌های ممکن برای انتخاب n تانک تصاحب‌شده از بین N تانک دشمن تقسیم می‌کنیم.
در ادامه بیش‌ترین شماره تانک تصاحب‌شده را می‌توان به عنوان تخمینی از $E[Y]$ در نظر گرفت. داریم :

$$E[Y]=\sum_{k=n}^{N}kP(Y=k)=\sum_{k=n}^{N}k\frac{\binom{k-1}{n-1}}{\binom{N}{n}}=\frac{1}{\binom{N}{n}}\sum_{k=n}^{N}\frac{k!}{(n-1)!(k-n)!}=\frac{n}{\binom{N}{n}}\sum_{k=n}^{N}\binom{k}{n}$$

برای محاسبه‌ی $\sum_{k=n}^{N}\binom{k}{n}$ باید توجه کنیم که $\binom{k}{n}$ ضریب x^n در چندجمله‌ای $(1+x)^k$ است. در نتیجه $\sum_{k=n}^{N}\binom{k}{n}$ ضریب x^n در چندجمله‌ای $\sum_{k=n}^{N}(1+x)^k$ است. پس داریم :

$$\sum_{k=n}^{N}(1+x)^k=(1+x)^n\sum_{k=0}^{N-n}(1+x)^k=(1+x)^n\frac{(1+x)^{N-n+1}-1}{(1+x)-1}=\frac{1}{x}[(1+x)^{N+1}-(1+x)^n]$$

ضریب x^n در چندجمله‌ای $\frac{1}{x}[(1+x)^{N+1}-(1+x)^n]$، همان ضریب x^{n+1} است در چندجمله‌ای $(1+x)^{N+1}$ که برابر است با $\binom{N+1}{n+1}$. پس خواهیم داشت:

$$E[Y]=\frac{n\binom{N+1}{n+1}}{\binom{N}{n}}=\frac{n\frac{(N+1)!}{(n+1)!(N-n)!}}{\frac{N!}{n!(N-n)!}}=\frac{n(N+1)}{n+1}$$

برای تخمین زدن کل تانک‌های این کشور معادله $E[Y]=\frac{n(N+1)}{n+1}$ را برای N حل می‌کنیم که مقدار $N=\frac{n+1}{n}E[Y]-1$ بدست می‌آید.

ب) با توجه به رابطه‌ی بدست آمده داریم:

$$N \simeq \frac{۱۳}{۱۲} \times ۱۱۷ - ۱ \simeq ۱۲۶$$

سؤال ۸.

تمرین کامپیوتری سری سوم با موضوع «توزیع احتمال‌های گسسته» را می‌توانید از طریق این لینک [۱] دریافت کنید.

- یک کپی از فایل مذکور با نام CA3_SID در گوگل درایو خود تهیه کنید.
- در فایل خود بخش‌هایی که به وسیله مستطیل مشخص شده‌اند را با کدهای مناسب جایگزین کنید.
- سوالاتی که به زبان فارسی و رنگ سفید مطرح شده‌اند را در همان سلول پاسخ دهید.
- فایل کد خود را با ایمیل hesam.as.sa.as@gmail.com با دسترسی Editor به اشتراک بگذارید.
- لینک فایل پاسخ خود را در بخش متنی جایگاه آپلود این تمرین در سامانه ایلرن قرار دهید.
- دقت کنید در صورتی که لازم به ایجاد یک سلول جدید برای اجرای کد داشتید، اول سلول از R%% استفاده کنید تا سلول به عنوان کد R تشخیص داده شود.

هرگونه انتقاد، پیشنهاد یا نکته جانبی را می‌توانید از طریق یک سلول متنی در ابتدای فایل (قبل از سرفصل اصلی) به ما منتقل کنید.

پاسخ .

تکمیل شده‌ی فایل صورت سوال از طریق این لینک [۲] در دسترس است.

[1] TODO
[2] https://colab.research.google.com/drive/1M6ImcYvMjRYocanPYBTOycQk2cCHUXFU?usp=sharing