آمار و احتمال مهندسی

نیم‌سال اول ۹۸ــ۹۹

گردآورندگان: امیرحسین کریمی، آرمین مرادی

دانشکده‌ی مهندسی کامپیوتر

تمرین شماره‌ی ۵ کوورایانس و هم‌بستگی، ویژگی‌های امیدریاضی، قضیه‌ی حد مرکزی و اعداد بزرگ

کوواریانس و هم‌بستگی

مسئله‌ی ۱. مناسب برای ۴ تا ۹۹ سال

روز تولد ۱۰۰ نفر را در نظر بگیرید. فرض کنید روزهای تولد این افراد از هم مستقل هستند و با احتمال یکسانی می‌توانند هر یک از ۳۶۵ روز سال باشند.

الف

کوواریانس تعداد افرادی که در ۱ فروردین به دنیا آمده‌اند و تعداد افرادی که در ۲ فروردین به دنیا آمده‌اند را به دست آورید.

ب

از حاصل کوواریانس به‌دست‌آمده چه نتیجه‌ای می‌توان گرفت؟

حل.

الف

متغیر تصادفی X_j را تعداد افرادی که در روز jاُم سال به دنیا آمده‌اند در نظر بگیرید. پس باید حاصل $Cov(X_1, X_2)$ را به دست آوریم.
همین‌طور متغیر تصادفی‌های شاخص A_j و B_j را به صورت زیر تعریف می‌کنیم:

$$B_j = \begin{cases} 1 & \text{تولد نفر } j\text{اُم دو فروردین بوده} \\ 0 & \text{در غیر اینصورت} \end{cases} \qquad A_j = \begin{cases} 1 & \text{تولد نفر } j\text{اُم یک فروردین بوده} \\ 0 & \text{در غیر اینصورت} \end{cases}$$

پس داریم:

$$E[X_1X_2] = E\Big[(\sum_i A_i)(\sum_j B_j)\Big] = E\Big[\sum_{i,j} A_iB_i\Big] = 100 \times 99(\frac{1}{365})^2$$

$$Cov(X_۱, X_۲) = E[X_۱X_۲] - E[X_۱]E[X_۲] = ۱۰۰ \times ۹۹(\frac{۱}{۳۶۵})^۲ - (\frac{۱۰۰}{۳۶۵})^۲ = -\frac{۱۰۰}{۳۶۵^۲}$$

ب

از مقدار کوواریانس نتیجه‌ای نمی‌توان گرفت ولی علامت آن اهمیت دارد. علامت منفی آن به ما می‌گوید که اگر تعداد تولدها در یک روز بالا برود، احتمالا تعداد تولدها در روز دیگری از سال کاهش می‌یابد. این با شهود ما نیز تطابق دارد. $\triangleright$

مسئله‌ی ۲. * لاک‌پشت و بچه‌هایش

یک لاک‌پشت دریایی به تعداد $N \sim Pois(\lambda)$ در شن‌های ساحل تخم می‌گذارد. از هر تخم لاک‌پشت، به طور مستقل، با احتمال p بچه‌لاک‌پشتی متولد می‌شود. تعداد بچه‌لاک‌پشت‌هایی که متولد می‌شوند را X در نظر بگیرید، پس $X|N \sim Bin(N, p)$ (این یعنی به شرط دانستن مقدار N، متغیر تصادفی X از توزیع برنولی با پارامتری برابر با مقدار N پیروی می‌کند).
همین‌طور وقتی یک بچه‌لاک‌پشت سر از تخم بیرون می‌آورد، به طور غریزی شیب سرازیری ساحل و انعکاس نور ماه و ستارگان بر روی آب را دنبال می‌کند تا به دریا برسد. مدت زمانی که طول می‌کشد تا یک بچه‌لاک‌پشت حرکت کند تا به دریا برسد را $T \sim Exp(\mu)$ در نظر بگیرید.

الف

هم‌بستگی بین X و N را بیابید. (پاسخ شما باید تابعی از p باشد و λ حذف خواهد شد)

ب

اگر دو بچه‌لاک‌پشت به طور هم‌زمان سر از تخم بیرون بیاورند، زمان رسیدن آنها به ساحل را $T_۱$ و $T_۲$ در نظر بگیرید. حاصل $Cov(min(T_۱, T_۲), max(T_۱, T_۲))$ را به دست آورید. آیا درست است بگوییم $Cov(min(T_۱, T_۲), max(T_۱, T_۲)) = Cov(T_۱, T_۲)$؟

حل.

الف

تعداد تخم لاکپشت‌هایی را که منجر به تولد بچه لاکپشت نمی‌شوند، Y در نظر بگیرید. در اینجا در کمال تعجب می‌فهمیم که X و Y از هم مستقل‌اند!!! از طرفی با نگاه به رابطه‌ی $X + Y = N$ به نظر می‌آید که این دو متغیر تصادفی شدیداً به هم وابسته باشند، ولی این زمانی درست است که

ما مقدار N را بدانیم.

$$Pr(X=i,Y=j)=Pr(X=i,Y=j|N=i+j)Pr(N=i+j)$$

$$=\frac{(i+j)!}{i!j!}p^iq^j\frac{e^{-\lambda}\lambda^{i+j}}{(i+j)!}=\frac{(\lambda p)^i}{i!}\frac{(\lambda q)^j}{j!}e^{-\lambda}=\frac{(\lambda p)^i}{i!}\frac{(\lambda q)^j}{j!}e^{-(p+q)\lambda}$$

$$=\left(e^{-\lambda p}\frac{(\lambda p)^i}{i!}\right)\left(e^{-\lambda q}\frac{(\lambda q)^j}{j!}\right)=Pr(X=i)Pr(Y=j)$$

می‌بینیم که $X\sim Pois(\lambda p)$ و $Y\sim Pois(\lambda q)$ و این دو از هم مستقل هستند. اما ناراحت نباشید؛ این فقط یک مورد خاص برای توزیع پوآسون بود.

$$Cov(N,X)=Cov(X+Y,X)=Cov(X,X)+Cov(Y,X)=Var(X)+0=\lambda p$$

$$Corr(N,X)=\frac{\lambda p}{SD(N)SD(X)}=\frac{\lambda p}{\sqrt{\lambda}\sqrt{\lambda p}}=\sqrt{p}$$

ب

$$T_1,T_2\sim Exp(\mu)\longrightarrow E[T_1]=E[T_2]=\frac{1}{\mu}$$

$$E[min(T_1,T_2)max(T_1,T_2)]=E[T_1T_2]=E[T_1]E[T_2]=\frac{1}{\mu^2}$$

$$min(T_1,T_2)\sim Exp(\mu+\mu)\longrightarrow E[min(T_1,T_2)]=\frac{1}{2\mu}$$

$$min(T_1,T_2)+max(T_1,T_2)=T_1+T_2\longrightarrow E[min(T_1,T_2)]+E[max(T_1,T_2)]=E[T_1]+E[T_2]$$

$$\frac{1}{2\mu}+E[max(T_1,T_2)]=\frac{1}{\mu}+\frac{1}{\mu}=\frac{2}{\mu}\longrightarrow E[max(T_1,T_2)]=\frac{3}{2\mu}$$

$$Cov(min(T_1,T_2),max(T_1,T_2))=\frac{1}{\mu^2}-\frac{1}{2\mu}\frac{3}{2\mu}=\frac{1}{4\mu^2}$$

دیدیم که $Cov(min(T_1,T_2),max(T_1,T_2))$ با $Cov(T_1,T_2)$ برابر نبود زیرا حاصل کوواریانس T_1 و T_2 برابر با صفر است (T_1 و T_2 از هم مستقل‌اند). $\triangleright$

مسئله‌ی ۳. سفر به اعماق کنگو

متین در سفرش به کنگو، اطلاعاتی از گوریل‌های آن منطقه جمع‌آوری کرد. این اطلاعات شامل دسته‌بندی ۷۰۰ گوریل از نظر قد و وزن آنهاست که در جدول زیر مشاهده می‌کنید.
به جای وزن گوریل‌ها، متغیر تصادفی X را در نظر بگیرید که می‌تواند سه مقدار ۰، ۱ و ۲ داشته باشد. به جای قد گوریل‌ها نیز متغیر تصادفی Y را در نظر بگیرید که می‌تواند دو مقدار ۰ و ۱ داشته باشد.

قد/وزن	سبک	متوسط	سنگین
کوتاه	۱۷۰	۷۰	۳۰
بلند	۸۵	۱۹۰	۱۵۵

الف

با استفاده از از توزیع توأم و توزیع حاشیه‌ای بررسی کنید که آیا وزن و قد گوریل‌ها از هم مستقل هستند یا نه؟

ب

حاصل $Corr(X,Y)$ را به دست آورید و نتیجه را با قسمت الف مقایسه کنید.

حل.

الف

توزیع توأم و توزیع‌های حاشیه‌ای X و Y را به دست می‌آوریم.

X/Y	۰	۱	۲	توزیع حاشیه‌ای Y
۰	۱۷۰/۷۰۰	۷۰/۷۰۰	۳۰/۷۰۰	۲۷۰/۷۰۰
۱	۸۵/۷۰۰	۱۹۰/۷۰۰	۱۵۵/۷۰۰	۴۳۰/۷۰۰
توزیع حاشیه‌ای X	۲۵۵/۷۰۰	۲۶۰/۷۰۰	۱۸۵/۷۰۰	۱

اگر X و Y از هم مستقل باشند، رابطه‌ی $Pr(X=x, Y=y) = Pr(X=x)Pr(Y=y)$ باید برقرار باشد.
اگر $Pr(X=۰, Y=۰) \stackrel{?}{=} Pr(X=۰)Pr(Y=۰)$ را چک کنیم:

$$\frac{۱۷۰}{۷۰۰} \stackrel{?}{=} \frac{۲۵۵}{۷۰۰} \cdot \frac{۲۷۰}{۷۰۰} \Longrightarrow ۱۱۹۰۰ \neq ۶۸۸۵۰$$

می‌بینیم که وزن و قد گوریل‌ها از هم مستقل نیستند.

ب

$$E[X] = \frac{۲۶۰}{۷۰۰} + ۲ \cdot \frac{۱۸۵}{۷۰۰} = \frac{۹}{۱۰} \quad , \quad E[Y] = \frac{۴۳۰}{۷۰۰} = \frac{۴۳}{۷۰}$$

$$E[X^۲] = \frac{۲۶۰}{۷۰۰} + ۴ \cdot \frac{۱۸۵}{۷۰۰} = \frac{۱۰}{۷} \quad , \quad E[Y^۲] = \frac{۴۳}{۷۰}$$

$$Var(X) = \frac{۱۰}{۷} - (\frac{۹}{۱۰})^{۲} = \frac{۴۳۳}{۷۰۰} \quad , \quad Var(Y) = \frac{۴۳}{۷۰} - (\frac{۴۳}{۷۰})^{۲} = \frac{۱۱۶۱}{۴۹۰۰}$$

$$E[XY] = \frac{۱۹۰}{۷۰۰} + ۲ \cdot \frac{۱۵۵}{۷۰۰} = \frac{۵}{۷} \longrightarrow Cov(X,Y) = E[XY] - E[X]E[Y] = \frac{۱۱۳}{۷۰۰}$$

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\frac{۱۱۳}{۷۰۰}}{\sqrt{\frac{۴۳۳}{۷۰۰}}\sqrt{\frac{۱۱۶۱}{۴۹۰۰}}} \approx ۰/۴۲۲$$

مقدار هم‌بستگی نیز به این اشاره دارد که وزن و قد گوریل‌ها از هم مستقل نیستند. ▷

خطی‌بودن امید ریاضی

مسئله‌ی ۴. * ساقی

آقای توالی شکسته‌بند محل است ک به دلیل مهارتش در بهبود شکستگی ساق پا به او ساقی می‌گویند. اما موفقیت اتفاقی نیست! او برای بهبود کامل شکستگی ساق پا یک فرمول بسیار پیچیده دارد.
آقای توالی برای اینکه بتواند کار خود را به خوبی انجام دهد باید زرده‌ی تخم ۵ گونه پرنده را با قیر آسفالت قاطی کند و به خورد مریض بدهد. اگر حتی یکی از ۵ نوع تخم پرنده در معجون وجود نداشته باشد، مریض دچار دل‌شکستگی می‌شود.
تخم ۵ گونه پرنده را نمی‌توان از روی ظاهر تشخیص داد و خریدن هر تخم پرنده و تشخیص نوع آن یک ساعت طول می‌کشد.

الف

اگر آقای ساقی هربار مجبور باشد بطور اتفاقی یکی از ۵ نوع تخم پرنده را خریداری کند، به او کمک کنید تا امیدریاضی زمان لازم برای جمع آوری هر ۵ نوع را به‌دست آورد. .

ب

اگر آقای توالی بخواهد ۲۴ ساعت تمام نخوابد و هر ساعت یک تخم پرنده را امتحان کند، به طور میانگین چند نوع متمایز تخم پرنده به دست خواهد آورد؟

حل. توالی بررسی تخم‌ها را با هر وقتی که تخم جدیدی یافت می‌شود جدا می‌کنیم و هر بار پس از دیده شدن نوع جدید، متغیر تصادفی جدیدی تعریف می‌کنیم. از خطی بودن امید ریاضی بهره می‌گیریم.
در امتحان اول حتما نوع جدیدی تخم پرنده خواهیم داشت.

الف

$$X = X_۱ + X_۲ + X_۳ ... + X_۵$$
$$E[X] = E[X_۱ + ... + X_۵] = \sum_{i=۱}^{۵} E[X_i]$$
$$E[X_۱] = ۱$$
$$E[X_۲] = N \times \frac{۱}{N - ۱} = \frac{۵}{۴}$$
$$E[X_i] = \frac{N}{N - (i - ۱)} ... E[X] = N \sum_{i=۱}^{۵} \frac{۱}{i}$$
$$E[X] = ۵ \times (۱ + \frac{۱}{۲} + ... + \frac{۱}{۵})$$
$$E[X] = \frac{۱۳۷}{۱۲}$$

ب

X_i را یک می‌گیریم وقتی در n امتحان (۲۴ امتحان) تخم نوع i دیده‌شده‌باشد.

$$P(X_i = ۰) = (۱ - p)^n$$
$$P(X_i = ۱) = ۱ - (۱ - p)^n$$
$$p = ۱/۱۰$$
$$P(X - i = ۱) = ۱ - (\frac{۹}{۱۰)^{۲۴}}) = ۰/۹۲$$
$$E[X] = \sum_{i=۱}^{۱۰} E[X_i]$$
$$E[X] = \sum_{i=۱}^{۱۰} ۱ \times P(X_i = ۱) + \sum_{i=۱}^{۱۰} ۰ \times P(X_i = ۰)$$
$$E[X] = \sum_{i=۱}^{۱۰} ۱ \times P(X_i = ۱)$$
$$E[X] = ۱۰ - \sum_{i=۱}^{۱۰} (۱ - \frac{۱}{۱۰})^{۲۴}$$
$$E[X] = ۱۰ - ۱۰۰/۹^{۲۴} E[X] = ۱۰ - ۰/۸ \approx ۹/۲$$

▷

مسئله‌ی ۵. * در جنگل مِیمانم

به دلیل آتش‌سوزی‌های بی‌سابقه در جنگل آمازون، میمون‌ها قصد مهاجرت هرچه سریع‌تر به سوی آفریقا را دارند... اما مشکل آنجاست که فقط یک قایق دارند. برای اینکه حق هیچ میمونی ضایع نشود، رئیس قبیله تصمیم می‌گیرد همه را به صف بایستاند. او اسم هر میمون را درون نارگیلی می‌نویسد و سپس همه‌ی نارگیل‌ها را پشت یک وانت می‌اندازد. میمون‌ها به ترتیب از اول صف یکی یکی نارگیل برمی‌دارند و نام داخل آن را می‌خوانند. میمونی که نامش خوانده شود سوار قایق شده و از شر آتش رهایی می‌یابد!

الف

به طور میانگین انتظار می‌رود اگر ۱۰۰ میمون داشته باشیم، چند میمون اسم خودشان را درون نارگیلی که برمی‌دارند ببینند؟ این رابطه را برای n میمون نیز بدست آورید.

ب

برای n میمون امید ریاضی جفت میمون‌هایی که هرکدام اسم دیگری را برمی‌دارد محاسبه کنید.

ج

میمون‌ها را در صف به ترتیب با اعداد ۱ تا n شماره‌گذاری می‌کنیم و حاصل نارگیل‌برداری میمون iم را a_i می‌نامیم. اگر b_i را $max\{a_1, \cdots, a_i\}$ تعریف کنیم، امید ریاضی تعداد اعضای منحصربه‌فرد $\{b_1, b_2, \cdots, b_n\}$ را به‌دست آورید.

حل.

الف

I_i را به صورت زیر تعریف می‌کنیم:

$$I_i = \begin{cases} 1 & \text{اگر میمون } i \text{ ام نارگیل خود را بردارد} \\ 0 & \text{o.w.} \end{cases}$$

حال امید ریاضی مجموع I_i ها را حساب می‌کنیم:

$$E[\sum_{i=1}^{n} I_i] = \sum_{i=1}^{n} E[I_i] = \sum_{i=1}^{n} \frac{1}{n} = 1$$

همانطور که از رابطه مشخص است، تعداد میمون‌هایی که اسم خود را برمی‌دارند مستقل از تعداد کل آنها و برابر یک است.

ب

I_{ij} را به صورت زیر تعریف می‌کنیم:

$$I_{ij} = \begin{cases} ۱ & \text{اگر میمون } i \text{ ام و میمون } j \text{ ام نارگیل یکدیگر را بردارند} \\ ۰ & \text{o.w.} \end{cases}$$

حال امید ریاضی مجموع I_{ij} ها را حساب می‌کنیم::

$$E[\sum_{i=۱}^{n}\sum_{j=i}^{n} I_{ij}] = \sum_{i=۱}^{n}\sum_{j=i}^{n} E[I_{ij}] = \frac{n(n-۱)}{۲} \times \frac{۱}{n(n-۱)} = \frac{۱}{۲}$$

دوباره جواب نهایی همیشه برابر نیم بوده و به تعداد وابستگی ندارد.

ج

برای هر k تای اول ترتیب متغیر نشانگر I_k را به صورت زیر تعیین می‌کنیم:

$$I_k = \begin{cases} ۱ & if\ max_{۱\leqslant i\leqslant k-۱}(a_۱, a_۲, \ldots, a_i) \neq max_{۱\leqslant i\leqslant k}(a_۱, a_۲, \ldots, a_i) \\ ۰ & \text{o.w.} \end{cases}$$

بدین معنی که نشانگر هنگامی یک می‌شود که عضو k ام ترتیب، بزرگترین عضو تا آنجا باشد که احتمال آن برابر $\frac{۱}{k}$ است.

$$E[\sum_{k=۱}^{n} I_k] = \sum_{k=۱}^{n} E[I_k] = \sum_{k=۱}^{n} \frac{۱}{k}$$

$\triangleright$

شرطی‌بودن امید ریاضی

مسئله‌ی ۶. * کامپیوت.

توزیع توأم متغیر X و Y به صورت زیر است:

$$f_{XY}(x,y) = \begin{cases} a(۳x+y), & ۰ < x, y < ۱ \\ ۰, & \text{o.w.} \end{cases} \quad (۱)$$

الف

ثابت a را بدست آورید.

ب

احتمال $P(X < 0.5|Y > 0.5)$ را محاسبه کنید.

ج

$E[X|Y > 0.5]$ و $Var(X|Y = 0.5)$ را محاسبه کنید.

حل.

الف

می‌دانیم مجموع احتمال‌ها در فضای موجود باید یک بشود. از انتگرال کمک می‌گیریم:

$$\iint_{XY\in[0,1]} a(3x+y) = 1$$

پس از محاسبه، a برابر با $\frac{1}{2}$ به‌دست می‌آید.

ب

$$P(x < 0.5|y > 0.5) = \frac{P(x < 0.5, y > 0.5)}{P(y < 0.5)}$$

$$P(x < 0.5|y > 0.5) = \frac{\int_0^{0.5}\int_{0.5}^{1} f(x,y)dxdy}{\int_0^{0.5}\int_{0.5}^{1} f(x,y)dxdy} = \frac{\frac{1}{8}}{\frac{7}{16}} = \frac{2}{7}$$

ج

در ابتدا $f_Y(y)$ را ب دست می‌آوریم زیرا در ادامه از آن در هر دو قسمت استفاده خواهیم کرد.

$$f_Y(y) = int.۱\frac{۱}{۲}(۳x + y)dx$$
$$f_Y(y) = \frac{۳}{۴} + y$$

$$E(X|Y > ۰/۵) = \int_X xf(x|y > ۰/۵)$$
$$= \int_X x\frac{f(x, y > ۰/۵)}{f(y > ۰/۵)}$$
$$= \int_۰^۱ x\frac{\int_{۰/۵} ۱ f(x, y)dy}{\int ۰/۵۱ f(y)}$$
$$= \frac{۱۹}{۶}$$

$$Var(X^۲|Y = ۰/۵) = E[X^۲|Y = ۰/۵] - E[X|Y = ۰/۵]$$
$$= \int_۰^۱ x^۲\frac{f_{XY}(x, ۰/۵)}{f_Y(۰/۵)} - \int_۰^۱ x\frac{f_{XY}(x, ۰/۵)}{f_Y(۰/۵)} = \frac{۵۲}{۱۲۰}$$

▷

مسئله‌ی ۷. * نمایی‌شون

سه متغیر تصادفی $X_۱$، $X_۲$ و $X_۳$ از توزیع نمایی با میانگین $\frac{۱}{\lambda_i}$ پیروی می‌کنند.

الف

$E[X_۱ + X_۲ + X_۳|X_۱ > ۱, X_۲ > ۲, X_۳ > ۳]$ را بر حسب λ_iها بیابید.

ب

ثابت کنید $Pr(X_۱ < X_۲) = \frac{\lambda_۱}{\lambda_۱+\lambda_۲}$.

ج

$Pr(X_۱ = min(X_۱, X_۲, X_۳))$ را بیابید.

راهنمایی: سعی کنید احتمال خواسته‌شده را با استفاده از X_1 و $min(X_2, X_3)$ بیان کنید. توزیع کمینه‌ی چند متغیر نمایی را به یاد بیاورید و از نتایج قسمت ب استفاده کنید.

حل.

الف

با استفاده از خواص خطی بودن امید ریاضی، بی‌حافظه بودن و مستقل بودن متغیر تصادفی‌ها از هم، داریم:

$$E(X_1|X_1 > 1) + E(X_2|X_2 > 2) + E(X_3|X_3 > 3) = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} + 6$$

ب

از آنجایی که X_1 و X_2 از هم مستقل هستند، چگالی احتمال توأم آنها برابر است با:

$$Pr(X_1 = a, X_2 = b) = \lambda_1 \lambda_2 e^{-(\lambda_1 a + \lambda_2 b)}$$

چگالی احتمال توأم روی اعداد حقیقی مثبت تعریف شده و باید روی قسمتی انتگرال بگیریم که X_2 در $[0, \infty]$ و X_1 در $[0, X_2]$ مقدار میگیرند.

$$\begin{aligned} Pr(X_1 < X_2) &= \int_0^\infty \int_0^b Pr(X_1 = a, X_2 = b) da db \\ &= \int_0^\infty \lambda_2 e^{-\lambda_2 b} \int_0^b \lambda_1 e^{-\lambda_1 a} da db \\ &= \int_0^\infty \lambda_2 e^{-\lambda_2 b}(1 - e^{-\lambda_1 b}) db \\ &= \int_0^\infty \lambda_2 e^{-\lambda_2 b} db - \int_0^\infty \lambda_2 e^{-(\lambda_1 + \lambda_2) b} db \\ &= 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{\lambda_1}{\lambda_1 + \lambda_2} \end{aligned}$$

ج

احتمال خواسته شده همان $Pr(X_1 \leqslant min(X_2, X_3))$ است. همینطور می‌دانیم $min(X_2, X_3)$ از X_1 مستقل است و توزیع آن $Exp(\lambda_2 + \lambda_3)$ می‌باشد. پس:

$$Pr\big(X_1 = min(X_1, X_2, X_3)\big) = Pr\big(X_1 \leqslant min(X_2, X_3)\big) = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3}$$

$\triangleright$

قضیه‌ی حد مرکزی و اعداد بزرگ

مسئله‌ی ۸. * رضا و مارمولک‌ها

در شهر بنگلنگ هندوستان ۱۰۰۰ خانه وجود دارد. در این شهر به احتمال $\frac{۱}{۲}$ در خانه‌ای که درآن حضور دارید هیچ مارمولک سمی‌ای وجود نخواهد داشت. رضا جون[یور] ۱۱ ساله ، شاگرد آقای هفت‌لنگ است. آقای هفت‌لنگ برای پیدا کردن هر ۱۰ مارمولک ۰/۲۵ نمره اختصاص داده است. به رضا کمک کنید تا بتواند تعدادی مارمولک سمی جمع کند و نمره‌ی امتیازی دریافت کند.

الف

احتمال اینکه رضا با یک‌بار گشتن همه‌ی هزار خانه بتواند بیشتر از نیم نمره بگیرد چقدر است؟

ب

باید حداقل چند خانه را بگردد تا احتمال اینکه در هر خانه به طور میانگین بیشتر از ۰/۳ و کمتر از ۰/۷ مارمولک پیدا کرده‌باشد بیشتر از ۰/۹ بشود؟

حل.

الف

برای آنکه رضا جون[یور] بتواند بیش از نیم نمره بگیرد، باید بیش از ۲۰ مارمولک پیدا کند. متغیر X_i را طوری تعریف می‌کنیم که اگر در خانه i مارمولک پیدا شد مقدار آن یک شود و در غیر آن صورت صفر. اگر Y را تعداد مارمولک‌های جمع شده بگیریم، خواهیم داشت:

$$Y = X_۱ + X_۲ + \ldots + X_{۱۰۰۰}$$
$$E[X_i] = ۰/۵ \quad Var(X_i) = ۰/۵(۱ - ۰/۵) = ۰/۲۵ = \sigma^۲$$
$$P(Y > ۲۰) = P(\frac{Y - ۵۰۰}{\sqrt{۱۰۰۰}\sigma} > \frac{۲۰ - ۵۰۰}{\sqrt{۱۰۰۰}\sigma} P(Y > ۲۰) \approx ۱ - \Phi(\frac{-۹۶}{\sqrt{۱۰}})$$

ب

همانند بخش اول عمل می‌کنیم.

$$\begin{aligned}
۰/۹ &= P(۰/۳ < \overline{X} < ۰/۷) \\
&= P(۰/۳ < \frac{Y}{n} < ۰/۷) \\
&= P(\frac{۰/۳n - ۰/۵n}{۰/۵\sqrt{n}} < \frac{Y - ۰/۵n}{۰/۵\sqrt{m}} < \frac{۰/۷n - ۰/۵n}{۰/۵\sqrt{n}}) \\
&\approx \Phi(۰/۴\sqrt{n}) - \phi(-۰/۴\sqrt{n}) \\
&= ۲\Phi(۰/۴\sqrt{n}) - ۱ \implies \Phi(۰/۴\sqrt{n}) = ۰/۹۵
\end{aligned}$$

با استفاده از جداول می‌توان مقدار $\Phi^{-۱}(۰/۹۵)$ را پیدا کرد.

$$\Phi^{-۱}(۰/۹۵) = ۱/۴ = ۰/۴\sqrt{n} \implies n \approx ۱۳$$

$\triangleright$

مسئله‌ی ۹. رفراندوم مرکزی

رفراندومی برگزار شده و می‌دانیم رای هر فرد به احتمال p آری است.

الف

اگر فرض کنیم $p = ۰/۵$، با استفاده از قضیه‌ی حد مرکزی تخمین بزنید از بین ۲۵ نفر، با چه احتمالی حداقل ۱۴ نفر رای آری می‌دهند؟

ب

اگر p مجهول و n تعداد افراد رای‌دهنده باشد، $\bar{X}_n$ را نسبت افراد موافق به کل افراد در نظر بگیرید. کمترین تعداد n چقدر باشد تا ۹۰% اطمینان داشته باشیم که اختلاف $\bar{X}_n$ از مقدار واقعی p به اندازه‌ی ۰/۰۱ است؟

حل.

الف

تعداد افراد موافق را قرار دهید $X \sim binomial(۲۵, ۰/۵)$. می‌دانیم:

$$E[X] = ۱۲/۵ \quad , \quad Var(X) = ۲۵ \cdot \frac{۱}{۴} = \frac{۲۵}{۴} \quad , \quad \sigma_X = \frac{۵}{۲}$$

پس از استاندارد کردن X و استفاده از قضیه‌ی حد مرکزی، داریم $Z = \frac{X-۱۲/۵}{۲/۵} \approx N(۰,۱)$، بنابراین:

$$Pr(X \geqslant ۱۴) = Pr\Big(\frac{X-۱۲/۵}{۲/۵} \geqslant \frac{۱۴-۱۲/۵}{۲/۵}\Big) = Pr(Z \geqslant ۰/۶) \approx \Phi(-۰/۶) = ۰/۲۷۴$$

ب

$$Pr(p - ۰/۰۱ \leqslant \bar{X}_n \leqslant p + ۰/۰۱) = ۰/۹$$

$$Pr(\bar{X}_n < p - ۰/۰۱) = Pr\big(X_n < n(p - ۰/۰۱)\big) = ۰/۰۵$$

$$Pr\Big(\frac{X_n - np}{\sqrt{np(۱-p)}} < \frac{-۰/۰۱n}{\sqrt{np(۱-p)}}\Big) = Pr\Big(Z < \frac{-۰/۰۱n}{\sqrt{np(۱-p)}}\Big) = ۰/۰۵$$

$$\frac{-۰/۰۱n}{\sqrt{np(p-۱)}} = \Phi^{-۱}(۰/۰۵)$$

$$\sqrt{n} = \Phi^{-۱}(۰/۰۵)\sqrt{p(۱-p)} \Longrightarrow n = ۱۰^{۴}\big(\Phi^{-۱}(۰/۰۵)\big)^{۲}p(۱-p)$$

$\triangleright$

مسئله‌ی ۱۰. کران‌های دوردست

متغیر تصادفی های $X_۱, X_۲, ..., X_{۲۵}$ i.i.d. هستند و به صورت زیر تعریف شده‌اند:

$$X_i = \begin{cases} ۱ & p = ۰/۶ \\ -۱ & q = ۰/۴ \end{cases}$$

تصحیح پیوستگی، تغییر کران‌ها برای یک متغیر تصادفی گسسته به کران‌ها برای یک متغیر تصادفی پیوسته است. برای مثال فرض کنید Z یک متغیر تصادفی گسسته باشد و بخواهیم $Pr(Z = ۵)$ را با توزیع پیوسته‌ای مثل نرمال تخمین بزنیم. برای این کار باید کرانی برای آن در نظر بگیریم، یعنی باید $Pr(۵-c \leqslant N \leqslant ۵+c)$ را محاسبه کنیم زیرا $Pr(N = ۵)$ برابر صفر است. قرار دهید $Y = X_۱ + X_۲ + ... + X_{۲۵}$. با استفاده از قضیه‌ی حد مرکزی و تصحیح پیوستگی، $Pr(۴ \leqslant Y \leqslant ۶)$ را تخمین بزنید.

حل. ابتدا میانگین و واریانس را برای X_i به دست می‌آوریم:

$$E[X_i] = (۰/۶)(۱) + (۰/۴)(-۱) = \frac{۱}{۵}$$

$$E[X_i^۲] = (۰/۶)(۱)^۲ + (۰/۴)(-۱)^۲ = ۱$$

$$Var(X_i) = ۱ - (\frac{۱}{۵})^{۲} = \frac{۲۴}{۲۵}$$

بنابراین:

$$E[Y] = ۲۵ \times \frac{۱}{۵} = ۵ \quad , \quad Var(Y) = ۲۵ \times \frac{۲۴}{۲۵} = ۲۴$$

$$\begin{aligned} Pr(۴ \leqslant Y \leqslant ۶) &= Pr(۳/۵ \leqslant Y \leqslant ۶/۵) \quad (\text{تصحیح پیوستگی}) \\ &= Pr(\frac{۳/۵ - ۵}{\sqrt{۲۴}} \leqslant \frac{Y - ۵}{\sqrt{۲۴}} \leqslant \frac{۶/۵ - ۵}{\sqrt{۲۴}}) \\ &= Pr(-۰/۳۰۶ \leqslant \frac{Y - ۵}{\sqrt{۲۴}} \leqslant ۰/۳۰۶) \\ &\approx \Phi(۰/۳۰۶) - \Phi(-۰/۳۰۶) \quad (\text{قضیه‌ی حد مرکزی}) \\ &= ۱ - ۲\Phi(-۰/۳۰۶) \approx ۰/۲۴ \end{aligned}$$

$\triangleright$

شاد باشید :)