# دانشگاه تهران، دانشکده مهندسی برق و کامپیوتر
## آمار و احتمال مهندسی

تمرین اول ـ متغیرهای تصادفی، واریانس و توزیع های گسسته

طراح: مهدی جمال خواه

سوپروایزر: نسا عباسی

تاریخ تحویل: ۷ آبان ۱۴۰۲

### ۱. برد قطعی — ۲۰ نمره

یک بازی را درنظر بگیرید که اگر شما $x$ دلار شرط ببندید، با احتمال $\frac{۱}{۲}$ ، $x$ دلارتان را می بازید و با احتمال $\frac{۱}{۲}$ ، $x$ دلارتان را بعلاوه $x$ دلار دیگر خواهید برد. شما با این استراتژی وارد بازی می شوید: بار اول یک دلار شرط می بندید و اگر باختید، بار بعد دو برابر سری قبل شرط می بندید و همین طور ادامه می دهید تا درنهایت برنده شوید؛ یا اینکه پول کافی برای ادامه بازی نداشته باشید.(ما این حالت رو باخت کامل می‌نامیم) فرض کنید که $N = ۲^k$ دلار پول دارید و وارد این بازی میشوید:

الف) اگر برای شما باخت کامل اتفاق افتد چند دلار را از دست داده‌اید؟ (۵نمره)

ب) اگر برنده شوید چند دلار خواهید برد؟ (۵نمره)

ج) چقدر احتمال دارد که شما هرگز برنده نشوید؟(یعنی باخت کامل اتفاق بیافتد.) (۵نمره)

د) همان استراتژی را در نظر بگیرید با این تفاوت که اگر بردید دوباره از اول بازی را شروع می کنید، چقدر احتمال دارد که پول شما دو برابر شود؟(قبل از اینکه به باخت کامل برسید) (۵نمره)

**پاسخ :**

الف) درنهایت $k$ مرحله میتوانیم بازی کنیم.

| | | | | | | |
|---|---|---|---|---|---|---|
| مرحله | ۱ | ۲ | ۳ | ۴ | $\dots$ | $k$ |
| میزان شرط بندی | ۱ | ۲ | ۴ | ۸ | $\dots$ | $۲^{k-۱}$ |

$$\text{میزان باخت} = ۱ + ۲ + ۴ + ۸ + \dots + ۲^{k-۱} = ۲^k - ۱$$

ب) فرض کنید در $i$ امین مرحله برنده شوید درنتیجه طبق مورد قبل تا مرحله $(i - ۱)$ ام $(۲^i - ۱)$ دلار از دست داده اید و در مرحله $i$ ام $۲^i$ دلار برنده می شوید، درنتیجه سود شما برابر است با:

$$\text{میزان سود درصورت برنده شدن} = ۲^i - ۲^i + ۱ = ۱$$

ج) این بازی معادل یک توزیع هندسی با پارامتر $p = \frac{۱}{۲}$ می باشد و ما باید احتمال این را بدست اوریم که برد در مرحله ای بیش تر از $k$ رخ بدهد.

$$X \sim Geo(\frac{۱}{۲}) \longrightarrow P(X > k) = \left(\frac{۱}{۲}\right)^k = \frac{۱}{۲^k} = \frac{۱}{N}$$

یعنی هر چه پول اولیه شما بیش تر باشد احتمال باخت کامل شما کمتر است.

د) هر بار فقط یک دلار می توانید برنده شوید پس باید $N$ بار برنده شوید:

$$P(\text{برد در یک مرحله}) = ۱ - \frac{۱}{N}$$
$$P(\text{برد در } N \text{ مرحله}) = (۱ - \frac{۱}{N})^N$$

**مطالعه بیش تر:** $(N \to \infty)$

$$P(\text{برد در یک مرحله}) = \lim_{N\to\infty} (۱ - \frac{۱}{N}) \approx ۱$$
$$P(\text{برد در } N \text{ مرحله}) = \lim_{N\to\infty} (۱ - \frac{۱}{N})^N = \frac{۱}{e} = ۰/۳۶۷$$

اگر شما بی نهایت پول داشته باشید احتمال برد شما تقریبا یک خواهد شد اما احتمال دو برابر شدن پول شما بیش تر از ۰/۳۶۷ نخواهد شد! این استراتژی به Martingale Strategy معروف است.

## ۲. نامه رسان	۱۵ نمره

یک نامه رسان باید $n$ نامه را پخش کند. او می خواهد بداند که اگر به جای اینکه هر نامه را دقیقا به گیرنده آن بدهد، به صورت کاملا تصادفی پخش کند(یعنی همه‌ی $n!$ با شانس یکسانی اتفاق می افتد.):

الف) به طور میانگین چند نفر نامه‌ی خود را دریافت میکنند؟ (۵نمره)

ب) چقدر احتمال دارد **دقیقا** $(n-۱)$ نفر نامه‌ی خود را دریافت کنند؟ (۵نمره)

ج) چقدر احتمال دارد که هیچ کس نامه‌ی خود را دریافت نکند؟ (۵نمره)

**پاسخ :**

الف) متغیر تصادفی شاخص $A_i$ را در نظر بگیرید، این متغیر یک خواهد بود اگر که شخص $i$ ام نامه خودش را دریافت کند و در غیر اینصورت صفر می باشد.
اگر متغیر تصادفی $N$ تعداد افرادی را نشان بدهد که نامه خود را دریافت می کنند داریم:

$$N = A_۱ + A_۲ + \ldots + A_n = \sum_{i=۱}^{n} A_i$$

هدف محاسبه امید ریاضی $N$ است با توجه به خطی بودن امید ریاضی داریم:

$$E[N] = E[A_۱] + E[A_۲] + \ldots + E[A_n] = \sum_{i=۱}^{n} E[A_i]$$

اما احتمال اینکه $A_i$ یک باشد چقدر است؟ اگر شخص $i$ ام نامه خودش را دریافت کند مابقی نامه ها به $(n-۱)!$ پخش می شوند و تعداد کل حالت ها $n!$ می باشد.

$$P(A_i = ۱) = \frac{(n-۱)!}{n!} = \frac{۱}{n} \to E[A_i] = \frac{۱}{n}$$

در نتیجه داریم:

$$E[N] = \underbrace{\frac{۱}{n} + \frac{۱}{n} + \ldots + \frac{۱}{n}}_{n \text{ تا}} = n(\frac{۱}{n}) = ۱$$

دقت کنید امید ریاضی همیشه خاصیت خطی دارد **حتی اگر متغیر ها به هم وابسته باشند.**

ب) اگر $(n-1)$ نفر نامه خود را دریافت کنند پس قطعا آن یک نفر دیگر هم نامه خودش را دریافت کرده است پس چنین چیزی امکان پذیر نیست و احتمال آن **صفر** است.

ج) متمم آن را بدست می‌آوریم.
با استفاده از اصل شمول و عدم شمول داریم:

$$\begin{aligned} P(N>0) &= P(A_1 \cup A_2 \cup \ldots \cup A_n) \\ &= \sum_i P(A_i) - \sum_{i<j} P(A_i \cup A_j) + \sum_{i<j<k} P(A_i \cup A_j \cup A_k) - \ldots \\ &= \binom{n}{1}\frac{(n-1)!}{n!} - \binom{n}{2}\frac{(n-2)!}{n!} + \binom{n}{3}\frac{(n-3)!}{n!} - \ldots \\ &= \sum_{i=1}^{n} \binom{n}{i} \frac{(-1)^{i+1}(n-i)!}{n!} \\ &= \sum_{i=1}^{n} \frac{(-1)^{i+1}}{i!} \end{aligned}$$

$$\begin{aligned} P(N=0) = 1 - P(N>0) = 1 - \sum_{i=1}^{n}\frac{(-1)^{i+1}}{i!} &= 1 + \sum_{i=1}^{n}\frac{(-1)^{i}}{i!} \\ &= \sum_{i=0}^{n}\frac{(-1)^{i}}{i!} \end{aligned}$$

به تفاوت بدست اوردن امید ریاضی و تابع جرمی احتمال $N$ توجه کنید که چگونه در دومی وابستگی $A_i$ ها مشکل ساز شد و راه حل را طولانی کرد.

## ۳. سرعت تولید ۲۵ نمره

در یک شرکت ماشین سازی $n$ کارگر روی یک ماشین، کاری مستقل از دیگری انجام می دهند. کارگر $i$ ام با احتمال $p_i$ در یک روز کارش را با موفقیت انجام می دهد اما اگر موفق نشد فردا دوباره از اول آن کار را انجام خواهد داد. مدیر عامل شرکت می خواهد بداند که:

الف) احتمال اینکه ساخت یک ماشین $k$ روز طول بکشد چقدر است؟ (۱۵ نمره)

ب) به صورت میانگین ساخت یک ماشین چقدر طول خواهد کشید؟ (۱۰ نمره)

**پاسخ :**

قبل از آنکه مساله را حل کنیم باید یک رابطه را بدست اوریم:
**توزیع ماکزیمم:**
فرض کنید که $X_1, X_2, \ldots, X_n$ متغیر تصادفی مستقل از یکدیگر باشند.

$$X_m = max(X_1, X_2, \ldots, X_n)$$

$$\begin{aligned} F_{X_m}(x) &= P(X_m \le x) \\ &= P(max(X_۱, X_۲, \dots, X_n) \le x) \\ &= P(X_۱ \le x, X_۲ \le x, \dots, X_n \le x) \\ &= P(X_۱ \le x)P(X_۲ \le x)\dots P(X_n \le x) \end{aligned}$$

الف) متغیر های تصادفی زیر را در نظر می گیریم:
$X_i$ : تعداد روزی که طول می کشد تا کارگر $i$ ام کارش را انجام می دهد.
$X_m$ : تعداد روزی که طول می کشد تا ماشین ساخته شود.
زمانی کار ماشین تمام می شود که همه‌ی کارگر ها کارشان تمام شود یعنی زمان ساخت ماشین برابر است:

$$X_m = max(X_۱, X_۲, \dots, X_n)$$

در نتیجه طبق فرمول بالا داریم:

$$F_{X_m}(x) = P(X_۱ \le x)P(X_۲ \le x)\dots P(X_n \le x)$$

از طرفی $X_i$ ها توریع هندسی دارند:

$$X_i \sim Geo(p_i)$$

$$F_{X_i}(x) = P(X_i \le x) = \sum_{i=۱}^{x} P(X=i) = \sum_{i=۱}^{x} p_i(۱-p_i)^{i-۱} = ۱-(۱-p_i)^x$$

در نتیجه:

$$\begin{aligned} F_{X_m}(x) &= \prod_{i=۱}^{n}(۱-(۱-p_i)^x) \\ P(X_m = k) &= F_{X_m}(k) - F_{X_m}(k-۱) \\ &= \prod_{i=۱}^{n}\left(۱-(۱-p_i)^k\right) - \prod_{i=۱}^{n}(۱-(۱-p_i)^{k-۱}) \end{aligned}$$

ب)

$$\begin{aligned} E[X_m] &= \sum_{j=۰}^{\infty}(۱-F_{X_m(j)}) \\ &= \sum_{j=۰}^{\infty}\left(۱-\prod_{i=۱}^{n}(۱-(۱-p_i)^j)\right) \end{aligned}$$

## ۴. باگ پنتیوم — ۱۰ نمره

همان طور که به احتمال زیاد در درس برنامه نویسی پیشرفته دیده‌اید برای هر پروژه چند نمونه آزمایش ( test case ) روی برنامه شما اجرا می‌شود و بر اساس آن نمره شما تعیین خواهد شد. حال فرض کنید که برنامه شما باید دقیقا ۱۰۰ ویژگی مستقل از یکدیگر داشته باشد، اما شما می دانید که ۵ تا از آن ها باگ دارد. اگر ۱۰ تا مورد آزمایشی ( test case ) روی برنامه شما اجرا شود(هر کدام دقیقا یک ویژگی متفاوت را آزمایش می کنند) و به ازای هر تست که جواب اشتباه دهد ۲/۵ نمره از ۱۰۰ نمره شما کم می‌شود:
الف) چقدر احتمال دارد که نمره شما حداقل ۹۵ شود؟ (۵نمره)
ب) چقدر احتمال دارد که دستیار آموزشی متوجه تمام باگ های برنامه شما شود؟ (۵نمره)

**پاسخ :**

الف)
متغیر های تصادفی زیر را در نظر بگیرید:
$T$ : تعداد مورد های ازمایشی که جواب اشتباه داده‌اند
$S$ : نمره شما

$T$ توزیع فوق هندسی دارد:

$$\begin{aligned}
&T \sim HyperGeo(N = 100, k = 5, n = 10) \\
&S = 100 - 2.5T \\
&P(T = x) = \begin{cases} \frac{\binom{5}{x}\binom{95}{10-x}}{\binom{100}{10}} & 0 \leq x \leq 5 \\ 0 & otherwise \end{cases}
\end{aligned}$$

$$\begin{aligned}
P(S \geq 95) &= P(S = 100) + P(S = 97.5) + P(S = 95) \\
&= P(T = 0) + P(T = 1) + P(T = 2) \\
&= \frac{\binom{5}{0}\binom{95}{10} + \binom{5}{1}\binom{95}{9} + \binom{5}{2}\binom{95}{8}}{\binom{100}{10}} \\
&= 0.993
\end{aligned}$$

ب)

$$\begin{aligned}
P(T = 5) &= \frac{\binom{5}{5}\binom{95}{5}}{\binom{100}{10}} \\
&= 3.34 \times 10^{-6}
\end{aligned}$$

## ۵. ایستگاه مترو ۳۰ نمره

یک مدل از خط شماره ۱ مترو را در نظر بگیرید که به بالای شهر می رود: فرض کنید که این مترو یک برنامه زمانبندی ثابت و قطعی دارد و هر ۱۰ دقیقه یک قطار در هر ایستگاه وارد می‌شود و تمام مسافرانی را که می‌خواهند به سمت بالای شهر بروند، سوار می‌کند. فرض کنید تمام مسافران بدون تأخیر و بدون تغییر در برنامه زمانبندی، سوار و پیاده می‌شوند. ایستگاه ها از ۱ تا $m$ به صورت افزایشی از پایین به سمت بالای شهر شماره گذاری شده اند .فرض کنید ورود مسافران به سمت بالای شهر در هر ایستگاه بر اساس توزیع پواسون مستقل با نرخ $\lambda_i$ در دقیقه در ایستگاه $i$ ام باشد، $1 \leq i \leq m - 1$. (ایستگاه $m$ اخرین ایستگاه است و هیچ کس در این ایستگاه سوار نمی شود تا به بالای شهر برود). فرض کنید هر فرد که وارد ایستگاه $i$ ام می شود، مستقل از هر کس دیگری با احتمال $P_{i,j}$ در ایستگاه $j$ ام پیاده می شود پس به ازای هر $i$ داریم: $\sum_{j=i+1}^{m} P_{i,j} = 1$. یک سفر به سمت بالای شهر از این مترو را در نظر بگیرید که از ایستگاه اول شروع می‌شود(در ایستگاه اول نیز همان فرض‌ها برقرار است). $D_j$ را تعداد افرادی که در ایستگاه $j$ ام پیاده می‌شوند در نظر بگیرید( $2 \leq j \leq m$ ).
الف) میانگین و واریانس $D_j$ را بدست اورید. (۱۵نمره)
ب) $P(D_2 = 2, D_3 = 3)$ را محاسبه کنید. (۱۵نمره)

**پاسخ :**

الف) ایستگاه $i$ ام را در نظر بگیرید در این ایستگاه مسافران در بازه ۱۰ دقیقه‌ای با توزیع پواسون با نرخ $10\lambda_i$ وارد می‌شوند و با احتمال $P_{i,j}$ در ایستگاه $j$ ام پیاده می‌شوند. در نتیجه اگر $N_{i,j}$ تعداد مسافرانی را نشان دهد که در ایستگاه $i$ ام سوار و در ایستگاه $j$ ام پیاده میشوند،

این متغیر نیز توزیع پواسون دارد زیرا ویژگی های توزیع پواسون را دارد و نرخ آن نیز برابر است با $۱۰\lambda_i P_{i,j}$ زیرا هر کسی که وارد این ایستگاه می شود با احتمال $P_{i,j}$ در ایستگاه $j$ ام پیاده می شود پس به صورت میانگین $۱۰\lambda_i P_{i,j}$ نفر از ان ها در ایستگاه $j$ ام پیاده می شود و نرخ توزیع پوسوان همان میانگین آن است.

$$N_{i,j} \sim Poi(۱۰\lambda_i P_{i,j})$$

حال تعداد کسانی که در ایستگاه $j$ پیاده می شوند در یکی از ایستگاه های قبل از $j$ سوار شده اند یعنی $D_j$ مجموع چند توزیع پواسون خواهد بود پس خود نیز توزیع پواسون دارد زیرا ویژگی های توزیع پواسون را داراست.

$$D_j = \sum_{i=۱}^{j-۱} N_{i,j} = \sum_{i=۱}^{j-۱} Poi(۱۰\lambda_i P_{i,j}) \to D_j \sim Poi(\mu_j); \mu_j = \sum_{i=۱}^{j-۱} ۱۰\lambda_i P_{i,j}$$

میانگین و واریانس توزیع پواسون را نیز میدانیم:

$$E[D_j] = Var[D_j] = \mu_j = \sum_{i=۱}^{j-۱} ۱۰\lambda_i P_{i,j} \quad ۲ \leq j \leq m$$

ب) چون $D_j$ ها مجموع چند توزیع مستقل هستند پس خودشان نیز مستقل هستند.

$$P(D_۲ = ۲, D_۳ = ۳) = P(D_۲ = ۲)P(D_۳ = ۳) = \left(\frac{e^{-\mu_۲}\mu_۲^۲}{۲!}\right)\left(\frac{e^{-\mu_۳}\mu_۳^۳}{۳!}\right)$$

## ۶. Perfect shuffle　　　امتیازی ــ ۱۰ نمره

یک دسته کارت ۵۲ تایی داریم و می خواهیم که آن را بر بزنیم تا به توزیع یکنواخت برسیم، یعنی هر !۵۲ حالت ممکن با شانس یکسانی اتفاق بیفتد. برای اینکار از یک نوع بر زدن خاص استفاده میکنیم به این صورت که کارت بالایی را بر میداریم و با احتمال یکسان در یکی از ۵۲ جای ممکن قرار می‌دهیم (بین کارت‌ها، روی کارت بالایی و زیر کارت‌ها). چند بار باید این کار را انجام دهیم تا به توزیع یکنواخت برسیم؟

**پاسخ :**

اخرین کارت را در نظر بگیرید(از اینجا به بعد این کارت را $A$ می نامیم)

- کارت اول را بر میداریم و به صورت رندم در یکی از ۵۲ جای ممکن می گذاریم .چقدر احتمال دارد که این کارت زیر کارت $A$ قرار گیرد؟ یک حالت از ۵۲ حالت زیر کارت $A$ قرار دارد. پس احتمال ان $\frac{۱}{۵۲}$ است. یعنی به صورت میانگین باید ۵۲ بار اینکار را انجام دهیم تا اولین کارت زیر کارت $A$ قرار گیرد.
- حال با فرض اینکه یک کارت زیر کارت $A$ قرار دارد، چقدر احتمال دارد که کارت برداشته شده زیر کارت $A$ قرار گیرد؟ بله $\frac{۲}{۵۲}$ . چون ۲ حالت از ۵۲ حالت زیر کارت $A$ قرار دارد. یعنی به صورت میانگین باید $\frac{۵۲}{۲}$ بار اینکار را انجام داد تا کارت دوم نیز زیر کارت $A$ قرار گیرد.

$\vdots$

- حال با فرض اینکه ۵۰ کارت زیر کارت $A$ قرار دارد، چقدر احتمال دارد که کارت برداشته شده زیر کارت $A$ قرار گیرد؟ بله $\frac{۵۱}{۵۲}$ . چون ۵۱ حالت از ۵۲ حالت زیر کارت $A$ قرار دارد. یعنی به صورت میانگین باید $\frac{۵۲}{۵۱}$ بار اینکار را انجام داد تا کارت ۵۱ ام نیز زیر کارت $A$ قرار گیرد.
- در نهایت کارت $A$ را که اکنون در بالای همه‌ی کارت ها قرار دارد برمیداریم و در یکی از ۵۲ حالت قرار میدهیم.

در نتیجه اگر $X$ تعداد انجام این بر زدن تا رسیدن به توزیع یکنواخت در کارت ها باشد داریم:

$$E[X] = \frac{۵۲}{۱} + \frac{۵۲}{۲} + \ldots + \frac{۵۲}{۵۱} + \frac{۵۲}{۵۲} = ۲۳۵/۹۷۸ \approx ۲۳۶$$

دقت کنید که ما روشی را پیش گرفتیم که در نهایت تمام جای های کارت ها به صورت رندم انتخاب شده است. پیش از اینکه تمامی کارت ها زیر کارت $A$ قرار گیرند هنوز قسمتی از چینش قبلی وجود دارد یعنی همه‌ی !۵۲ جایگشت احتمال یکسانی ندارند.

**مطالعه بیش تر:** این اثبات مقدمه‌ای برای اثبات قضیه seven shuffle می باشد. این قضیه می‌گوید که اگر هفت بار riffle shuffle را روی یک دسته کارت ۵۲ تایی انجام دهیم به توزیع یکنواخت می‌رسیم یعنی هر !۵۲ حالت ممکن با شانس یکسانی اتفاق می افتد.