Föreläsning 10: Slumpvariabler · 1MA020

Vilhelm Agdur[1]

[1] vilhelm.agdur@math.uu.se

27 februari 2023

Vi fortsätter att diskutera diskret sannolikhetsteori, och introducerar slumpvariabler och deras väntevärden.

Vi använder den teori vi byggt upp för att bevisa några fler resultat inom kombinatoriken.

Slumpvariabler

Hittills är vad vi har sett bara hälften av vad man intuitivt tänker ingår i sannolikhetsteorin – vi har diskuterat slumpmässiga *händelser*, som antingen inträffar eller inte, men vi har inte definierat slumpmässiga tal. Frågan om ifall det kommer att regna imorgon eller inte kan vi modellera i vår formalism, men inte frågan om hur många millimeter det kommer regna.

Definition 1. Givet ett sannolikhetsrum (Ω, μ) är en *slumpvariabel* X som tar värden i V en funktion $X : \Omega \to V$. Givet varje utfall tar alltså vår slumpvariabel ett visst värde, och givet varje[2] delmängd $A \subseteq V$ blir $X \in A$ en händelse – specifikt är det händelsen

$$\{\omega \in \Omega \mid X(\omega) \in A\} = X^{-1}(A).$$

[2] Detta är lite av en lögn i det allmänna fallet, eftersom det kan finnas *väldigt* skumma delmängder till V, men så länge vi tänker oss våra diskreta sannolikhetsrum är det sant.

Det allra vanligaste fallet är när $V = \mathbb{R}$ eller någon delmängd till $\mathbb{R}$. I många introtexter om sannolikhetsteori *definierar* man att slumpvariabler tar värden i $\mathbb{R}$ – men eftersom vi sysslar med kombinatorik kommer vi att vilja ha mer exotiska slumpvariabler, som slumpmässiga permutationer eller slumpmässiga mängder.

Exempel 2. Låt oss återbesöka vårt exempel med ett tärningskast. Vi konstaterade att vi kan ta $\Omega = \{1, 2, 3, 4, 5, 6\}$ och $\mu(\omega) = 1/6$ för alla $\omega \in \Omega$.

Vi kan naturligt betrakta vårt tärningskast som en slumpvariabel – i detta fall blir det en mycket enkel funktion, $X : \Omega \hookrightarrow \mathbb{R}$ skickar helt enkelt varje ω på sig självt.

Vårt tärningskast är ett specialfall av ett mer allmänt fenomen, som det kommer vara bekvämt att ha en terminologi för.

Definition 3. Givet en ändlig mängd V är ett *likformigt fördelat slumpmässigt element av* V en slumpvariabel X sådan att $\mathbb{P}(X = v) = \frac{1}{|V|}$ för varje $v \in V$.[3] Alla element av V är alltså lika sannolika. Vi kan skriva detta som

$$X \overset{\text{u}}{\in} V.$$

[3] Vill man göra detta fullständigt rigoröst i vår formalism kan man säga att X är definierad på sannolikhetsrummet (V, μ) där $\mu(v) = \frac{1}{|V|}$ för alla $v \in V$, och $X : V \to V$ är identitetsfunktionen.

Men det blir väldigt många abstrakta ord för att inte säga så mycket alls som vi inte redan sade när vi definierade X som att den blir lika med varje element i V med samma sannolikhet.

Detta innebär alltså att för varje mängd $W \subseteq V$ så blir

$$\mathbb{P}(X \in W) = \frac{|W|}{|V|}.$$

Om någon säger att "vi låter X vara en slumpmässig graf / träd / mängd / etc." utan att specificera hur X är fördelad menar de att den är likformig.

Vi vet att om vi slår vår tärning många gånger kommer vi i genomsnitt att få upp 3.5. Hur gör vi den intuitionen rigorös?

Definition 4. *Väntevärdet* av en slumpvariabel X som tar värden i $\mathbb{R}$[4] ges av[5]

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x).$$

Vi tar alltså summan över alla tänkbara värden x för X, multiplicerar x med sannolikheten att X faktiskt blir x,[6] och summerar. I specialfallet där X bara tar värden $0, 1, 2, \ldots$ blir alltså formeln

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\mathbb{P}(X = k).$$

Exempel 5. Så om vi åter tar exemplet med tärningskastet så blir alltså väntevärdet

$$\begin{aligned}\mathbb{E}[X] &= 1\mathbb{P}(X = 1) + 2\mathbb{P}(X = 2) + \ldots + 6\mathbb{P}(X = 6) \\ &= \frac{1+2+3+4+5+6}{6} = \frac{7}{2} = 3.5\end{aligned}$$

precis som vi förväntade oss.

Ibland är det mer användbart att skriva definitionen av väntevärde på en alternativ form:

Lemma 6. *Det gäller för varje slumpvariabel X som tar värden i $\mathbb{R}$ att*[7]

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mu(\omega).$$

Bevis. Vi kan skriva

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x \in X(\Omega)} x\mathbb{P}(X = x) \\ &= \sum_{x \in X(\Omega)} x \left(\sum_{\omega \in \Omega : X(\omega) = x} \mu(\omega) \right) \\ &= \sum_{x \in X(\Omega)} \sum_{\omega \in \Omega : X(\omega) = x} x\mu(\omega) \\ &= \sum_{x \in X(\Omega)} \sum_{\omega \in \Omega : X(\omega) = x} X(\omega)\mu(\omega) \\ &= \sum_{\omega \in \Omega} X(\omega)\mu(\omega).\end{aligned}$$

□

[4] Samma definition hade fungerat precis lika väl över de komplexa talen. För den mer matematiskt sinnade kan det nog vara intressant att fundera över vad vi verkligen behöver anta om rummet vår slumpvariabel tar värden i för att väntevärdet skall bli väldefinierat.

[5] Notera att detta är en summa över *alla värden som X kan tänkas ta* – eftersom vi antagit att Ω är ändligt eller uppräkneligt så kommer detta vara en summa över ändligt eller uppräkneligt många summander, vilket är okej.

Hade vi velat modellera en *kontinuerlig* slumpvariabel – som till exempel en normalfördelning, som nog många sett redan – som kan ta vilket reellt tal som helst som värde, hade vi behövt definiera detta som en integral, inte en summa. Att slippa ge definitioner som fungerar i dessa fall är en av anledningarna till varför vi begränsar oss till bara diskret sannolikhetsteori.

[6] Notera att när den här summan löper över oändligt många tal är det fullt möjligt att den inte konvergerar, trots att $\sum_{x \in X(\Omega)} \mathbb{P}(x = X) = 1$. Det finns slumpvariabler som alltid tar ändliga värden, men ändå har oändligt väntevärde.

[7] Det här fungerar bara för att vi har antagit att våra sannolikhetsrum är ändliga eller uppräkneligt oändliga, så vi kan skriva våra sannolikheter som summor. I det mer allmänna fallet hade vi behövt skriva en integral mot sannolikhetsmåttet, och det kräver betydligt mer avancerad analys än vad vi kan.

Eftersom vi definierat slumpvariabler som att de helt enkelt är funktioner från Ω kan vi göra all den algebra vi vanligen kan på funktioner in i $\mathbb{R}$. Till exempel är det, givet två slumpvariabler X och Y, helt väldefinierat att skriva $X + Y$, och det betyder precis vad vi förväntar oss att det skall betyda – vi slumpar ett X och ett Y och sedan adderar vi dem med varandra.

När vi nu har introducerat addition av slumpvariabler så kan vi bevisa vad som, i min mening, är en av de allra mest användbara satserna i hela matematiken.[8]

[8] Jag är så klart oerhört partisk, eftersom just gränslandet mellan kombinatorik och sannolikhetsteori är mitt område – men det är onekligen ett otroligt användbart resultat.

Lemma 7 (Väntevärdets linjäritet)**.** *Givet två slumpvariabler X och Y som tar värden i $\mathbb{R}$ och två reella tal a och b gäller det att*

$$\mathbb{E}\left[aX + bY\right] = a\mathbb{E}\left[X\right] + b\mathbb{E}\left[Y\right].$$

Väntevärdet är alltså linjärt, som funktion från rummet av slumpvariabler in i $\mathbb{R}$.[9]

[9] Detta sätt att formulera det skrapar lite på ytan av en väldigt djup teori – väntevärden är nämligen "bara" integraler mot sannolikhetsmått, och samlingen av funktioner från Ω in i $\mathbb{R}$ blir ju ett vektorrum. Vi kan ge det vektorrummet en inre produkt genom att skriva $\langle X, Y\rangle = \mathbb{E}\left[XY\right]$, och vi har börjat med funktionalanalys.

Men detta är ju en kurs i kombinatorik, så att utforska detta får vänta till en framtida kurs för er.

Bevis. Vi använder den alternativa formeln för väntevärde vi fann i Lemma 6 och skriver

$$\begin{aligned}\mathbb{E}\left[aX + bY\right] &= \sum_{\omega\in\Omega} (aX + bY)(\omega)\mu(\omega)\\ &= \sum_{\omega\in\Omega} (aX(\omega) + bY(\omega))\mu(\omega)\\ &= a\sum_{\omega\in\Omega} X(\omega)\mu(\omega) + b\sum_{\omega\in\Omega} Y(\omega)\mu(\omega)\\ &= a\mathbb{E}\left[X\right] + b\mathbb{E}\left[Y\right].\end{aligned}$$

□

För att göra det här verkligt användbart behöver vi konceptet med indikatorvariabler, som vi introducerade när vi bevisade inklusion-exklusion.

Proposition 8. *För en händelse A blir dess indikatorfunktion $\mathbb{1}_A$, som ges av att $\mathbb{1}_A(\omega) = 1$ om $\omega \in A$ och noll annars, en slumpvariabel.*[10]

Det gäller att

$$\mathbb{P}\left(A\right) = \mathbb{E}\left[\mathbb{1}_A\right].$$

[10] Det är ju en funktion från utfall till reella tal – per definition är det en slumpvariabel. Vi behöver bara känna igen att den är det.

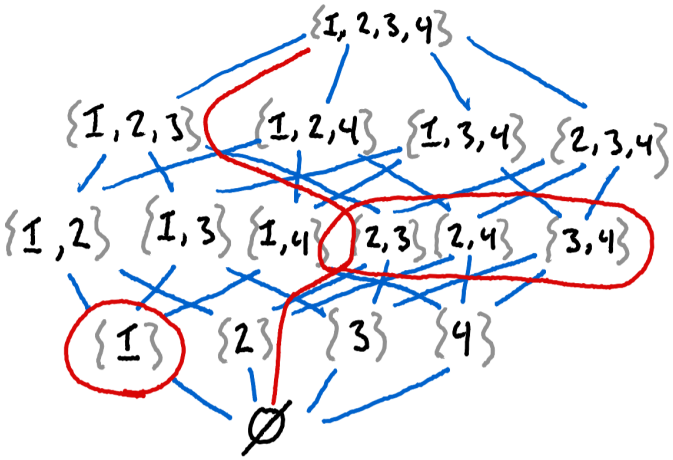
Bevis. Per definition har vi att

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_A\right] &= 0\cdot\mathbb{P}\left(\mathbb{1}_A = 0\right) + 1\cdot\mathbb{P}\left(\mathbb{1}_A = 1\right)\\ &= \sum_{\omega:\mathbb{1}_A(\omega)=1} \mu(\omega)\\ &= \sum_{\omega\in A} \mu(\omega) = \mathbb{P}\left(A\right).\end{aligned}$$

□

Sperners lemma

Låt oss nu ta vad vi har lärt oss och tillämpa det på ett faktiskt kombinatoriskt resultat.



Figur 1: En illustration av delmängdsgittret för mängden $\{1,2,3,4\}$. Längst upp ser vi hela mängden, längst ner är tomma mängden. De blå strecken indikerar att en mängd är en delmängd till en annan. Det röda strecket från $\emptyset$ till hela mängden är en maximal kedja, och samlingen av mängder inringade i rött bildar en antikedja.

Definition 9. En följd av icketomma delmängder

$$F_1 \subsetneq \ldots \subsetneq F_{k-1} \subsetneq F_k \subseteq [n]$$

till $[n]$ kallas för en *kedja*. Ifall $k = n$ kallar vi kedjan för en *maximal* kedja,[11] och då måste vi, eftersom vi kräver att varje mängd är en *strikt* delmängd till nästa, ha att $|F_i| = i$. I så fall låter vi också $F_0 = \emptyset$.

En samling $\mathcal{G}$ av delmängder till $[n]$ kallas för en *anti-kedja* ifall det för varje par $F, G \in \mathcal{G}$ varken gäller att $F \subseteq G$ eller $G \subseteq F$.

Denna definition illustreras i Figur 1.

[11] Att vi inte kan ha en längre kedja torde vara uppenbart.

Att varje maximal kedja består av precis n stycken mängder är uppenbart. Hur stor kan en anti-kedja vara? Ett enkelt sätt att skapa sig en sådan är att ta alla delmängder av storlek k till $[n]$ för något k – att dessa inte kan vara delmängder till varandra är uppenbart. Att det val av k som gör denna anti-kedja som störst blir $\lfloor \frac{n}{2} \rfloor$ är inte allt för svårt att se.[12] Är det möjligt att hitta en ännu större genom att ha med delmängder av olika storlekar? Sperners lemma säger oss att svaret är nej.

[12] Vi hade också kunnat välja $\lceil \frac{n}{2} \rceil$, det ger samma storlek.

Lemma 10 (Sperners lemma)**.** *För varje anti-kedja $\mathcal{F}$ i $[n]$ gäller det att*

$$|\mathcal{F}| \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}.$$

Bevis. Vi tar en likformigt slumpmässig maximal kedja

$$\emptyset = F_0 \subsetneq F_1 \subsetneq \ldots \subsetneq F_{n-1} \subsetneq F_n = [n],$$

och låter I vara mängden av i sådana att F_i ligger i vår anti-kedja $\mathcal{F}$.

Det är enkelt att se att I innehåller antingen noll eller ett element – en kedja och en anti-kedja kan ju omöjligen skära varandra i mer än en mängd.[13]

Låt oss nu studera slumpvariabeln $X = |I|$. Att vi vet att I bara kan ha noll eller ett element ger oss omedelbart att $\mathbb{E}[X] \leq 1$,[14] men låt oss studera detta väntevärde också på ett annat sätt.

Vi kan räkna med hjälp av väntevärdets linjäritet att[15]

$$\begin{aligned}\mathbb{E}[X] = \mathbb{E}[|I|] &= \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{\{i\in I\}}\right] \\ &= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{1}_{\{i\in I\}}\right] = \sum_{i=1}^{n} \mathbb{P}(i \in I).\end{aligned}$$

Vad är sannolikheten att i ligger i I? Att i ligger i I betyder att $F_i \in \mathcal{G}$, per definition. Så vad vi behöver förstå är den slumpmässiga mängden F_i.

Eftersom vi valde vår kedja som en likformigt slumpmässig kedja finns det ingen anledning till varför något tal skulle vara mer sannolikt än något annat att dyka upp i denna mängd. Så F_i är alltså, av symmetriskäl, ett likformigt slumpmässigt element ur $\binom{[n]}{i}$, mängden av delmängder av storlek i.

Vad är sannolikheten att F_i faller i $\mathcal{G}$? Jo, om vi låter $\mathcal{G}_i$ beteckna samlingen av element i $\mathcal{G}$ av storlek i vet vi att $\mathcal{G}_i \subseteq \binom{[n]}{i}$ och $F_i \overset{u}{\in} \binom{[n]}{i}$, så vi måste ha att

$$\mathbb{P}(F_i \in \mathcal{G}_i) = \frac{|\mathcal{G}_i|}{\binom{n}{i}}.$$

Så samlar vi ihop vad vi har listat ut hittills i ett enda uttryck, och använder olikheten[16] $\binom{n}{i} \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}$ för alla i, har vi att

$$1 \geq \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{P}(i \in I) = \sum_{i=1}^{n} \frac{|\mathcal{G}_i|}{\binom{n}{i}} \geq \sum_{i=1}^{n} \frac{|\mathcal{G}_i|}{\binom{n}{\lfloor \frac{n}{2} \rfloor}}$$

så multiplicerar vi bägge sidorna av detta med $\binom{n}{\lfloor \frac{n}{2} \rfloor}$ får vi att

$$\binom{n}{\lfloor \frac{n}{2} \rfloor} \geq \sum_{i=1}^{n} |\mathcal{G}_i| = |\mathcal{G}|$$

vilket är precis Sperners lemma, som vi ville bevisa. □

Caro-Weis sats

Antag att vi har en grupp av n personer på ett läger, och säg att person nummer i känner d_i andra personer sedan tidigare. Du vill bilda en mindre grupp av personer för en lära-känna-varandra-lek[17],

[13] Ifall vi hade både i och j i I, med $i < j$, hade vi ju haft att $F_i \subsetneq F_j$ med bägge i $\mathcal{F}$, vilket motsäger att $\mathcal{F}$ skulle vara en anti-kedja.

[14] Detta kan vi göra till ett allmänt lemma:

Lemma 11. *Om $X(\omega) \leq C$ för varje $\omega \in \Omega$ gäller det att $\mathbb{E}[X] \leq C$.*

Bevis. Vi kan räkna

$$\begin{aligned}\mathbb{E}[X] &= \sum_{\omega\in\Omega} X(\omega)\mu(\omega) \\ &\leq \sum_{\omega\in\Omega} C\mu(\omega) \\ &= C\sum_{\omega\in\Omega} \mu(\omega) = C.\end{aligned}$$

□

[15] Här använder vi den kortare notationen $\mathbb{1}_{\{i\in I\}}$ för att beteckna $\mathbb{1}_{\{\omega\in\Omega : i\in I(\omega)\}}$.

[16] Vi har strikt sett inte faktiskt bevisat den någon gång, men det bör vara någorlunda enkelt att övertyga sig själv om att den är sann.

[17] Eftersom du är en fruktansvärt ondskefull person. Ingen tycker om sådana lekar.

så du vill hitta en så stor grupp som möjligt av personer som *inte* känner varandra. Finns det någon garanti för hur stor du kan göra den mindre gruppen?

Den matematiska formaliseringen av det här är så klart i termer av grafer, så låt oss ge de rätta definitionerna först innan vi ger resultatet.

Definition 12. *Graden* av en nod v i en graf $G = (V_G, E_G)$ är antalet kanter noden har. Alltså

$$d_v = |\{e \in E_G \mid v \in e\}|.$$

Definition 13. En *oberoende mängd* $S \subseteq V_G$ i en graf $G = (V_G, E_G)$ är en mängd av noder sådana att det inte finns några kanter mellan något par av noder i S. Alltså, mängden

$$E(S) = \{\{u, v\} \in E_G \mid u, v \in S\}$$

är tom.

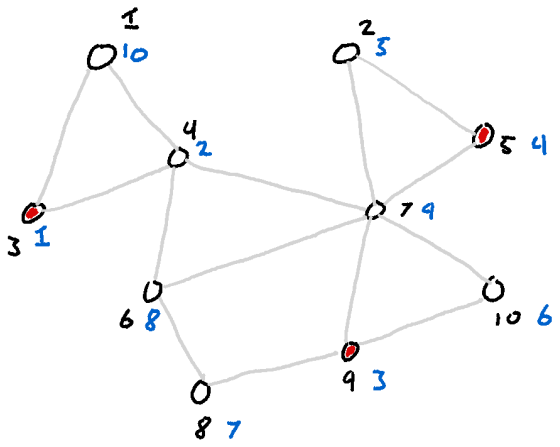
Teorem 14 (Caro-Wei)**.** *Varje graf G på n noder, där nod i har grad d_i, har alltid en oberoende mängd S sådan att*

$$|S| \geq \sum_{i=1}^{n} \frac{1}{d_i + 1}.$$

Bevis. Numrera noderna i G från 1 till n, och välj sedan likformigt slumpmässigt ett nytt sätt att etikettera G, så att nod i nu har etikett $\sigma(i)$. σ är alltså en likformig permutation av $[n]$.

Hur använder vi detta för att skapa vår oberoende mängd? Jo, beteckna mängden av grannar till i med $N(i)$ – alltså mängden av alla noder som har en kant till i. Om vi lägger in i i S får vi alltså inte ta med något annat element i $N(i)$ om S skall vara en oberoende mängd.[18]

[18] Tänk er det som att vi lägger ett pussel, där varje bit är "formad som" en $N(i)$, och vi inte får lov att välja två pusselbitar som överlappar. Målet är att lyckas lägga så många bitar som möjligt. (Det här går nog att omvandla till ett faktiskt spel – jag kräver inga royalties för idén.)



Figur 2: En illustration av vår konstruktion av den oberoende mängden S. Vår första etikettering av grafen är i svart, vår ometikettering σ i blått, och den resulterande oberoende mängden S markerad i rött.

Om vi nu låter

$$S = \{i \in V_G \mid \sigma(i) < \sigma(j) \quad \forall j \in N(i)\}$$

så hävdar vi att detta måste vara en oberoende mängd. Varför? Tänk för motsägelse att det fanns ett par i, j i S med en kant mellan sig. Eftersom i ligger i S måste alla is grannar ha högre etikett än i – specifikt så måste alltså $\sigma(i) < \sigma(j)$. Men det betyder ju att j har en granne med lägre etikett, så j kan inte ligga i S, och vi har en motsägelse.

Vi vill alltså förstå oss på mängden S. Låt A_i vara händelsen att i fick en lägre etikett än alla sina grannar i vår slumpmässiga ometikettering – vi har då av väntevärdets linjäritet att

$$\mathbb{E}\left[|S|\right] = \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{A_i}\right] = \sum_{i=1}^{n} \mathbb{P}\left(A_i\right).$$

Det räcker alltså för oss att förstå sannolikheterna för händelserna A_i. Eftersom σ var likformigt slumpmässig betyder det alltså att vi vill räkna antalet sätt att etikettera grafen sådana att i får en lägre etikett än alla sina grannar.

För att konstruera ett sådant sätt att etikettera G börjar vi med att välja vilka tal som skall stå på i och dess grannar – detta kan vi göra på $\binom{n}{d_i+1}$ sätt, eftersom vi skall ha d_i etiketter på dess grannar och en etikett på den själv.

Sedan väljer vi hur vi placerar dessa etiketter på i och $N(i)$ – vi måste så klart välja att placera det lägsta av talen på i, men de återstående d_i talen kan vi placera ut fritt[19], och alltså på $d_i!$ sätt.

Till slut väljer vi hur vi placerar resten av etiketterna på noderna utanför is grannskap – detta kan vi också göra helt fritt, så på $(n - d_i - 1)!$ sätt. Så totalt har vi sett att det finns

$$\binom{n}{d_i+1} d_i!(n-d_i-1)!$$

sätt att välja en etikettering av G sådan att i får en lägre etikett än alla sina grannar.

Så sannolikheten att en slumpmässig etikettering är sådan ges alltså av

$$\begin{aligned}
\mathbb{P}\left(A_i\right) &= \frac{1}{n!}\binom{n}{d_i+1} d_i!(n-d_i-1)! \\
&= \frac{1}{n!}\frac{n!}{(d_i+1)!(n-(d_i+1))!} d_i!(n-d_i-1)! \\
&= \frac{1}{d_i+1}
\end{aligned}$$

så

$$\mathbb{E}\left[|S|\right] = \sum_{i=1}^{n} \mathbb{P}\left(A_i\right) = \sum_{i=1}^{n} \frac{1}{d_i+1}.$$

[19] Det är här som vår teknik med väntevärdets linjäritet verkligen lönar sig – vi har kunnat zooma in bara på i, och behöver inte bry oss om vad som händer utanför just i. Hade vi inte gjort det hade vi kanske behövt bekymra oss om kanter mellan is grannar här, och inte kunnat placera ut etiketterna helt fritt.

Vi har alltså visat att vi *i genomsnitt* hittar en oberoende mängd av vår sökta storlek med denna metoden. Men det genomsnittliga värdet kan ju omöjligen vara mindre än *alla* specifika möjliga värden[20] – alltså måste det finnas något specifikt val av σ sådant att storleken på $S(\sigma)$ blir åtminstone detta. Alltså är vi klara. □

[20] Låt oss formulera detta som ett lemma och bevisa det:

Lemma 15. *För varje slumpvariabel X med $\mathbb{E}[X] = C$ måste det finnas åtminstone ett ω sådant att $X(\omega) \geq C$.*

Bevis. Antag för motsägelse att $X(\omega) < C$ för alla ω. Då kan vi räkna att

$$\begin{aligned} C = \mathbb{E}[X] &= \sum_{\omega\in\Omega} X(\omega)\mu(\omega) \\ &< \sum_{\omega\in\Omega} C\mu(\omega) \\ &= C\sum_{\omega\in\Omega} \mu(\omega) = C \end{aligned}$$

så $C < C$, en motsägelse. □

Första-moment-metoden

Hittills har vi sett ett sätt som väntevärden och sannolikheter kan samspela – om slumpvariabeln vi vill studera kan formuleras som antalet "ja" på en samling ja-nej-frågor så får vi att dess väntevärde är summan av sannolikheterna för ett "ja" på varje enskild fråga.

Detta låter oss alltså svara på en fråga om ett väntevärde genom att räkna ut en bunt sannolikheter. Hur gör vi om det vi verkligen är intresserade av är en sannolikhet?

Lemma 16 (Markovs olikhet)**.** *Om en icke-negativ[21] slumpvariabel X har väntevärde ν gäller det för varje $C > 0$ att[22]*

$$\mathbb{P}(X > C\nu) < \frac{1}{C}.$$

[21] Det vill säga, $X(\omega) \geq 0$ för alla $\omega \in \Omega$ – den tar aldrig ett negativt värde.

[22] Eller ekvivalent att

$$\mathbb{P}(X > C) \leq \frac{\nu}{C}.$$

Bevis. Vi kan räkna att[23]

$$\begin{aligned} \nu = \mathbb{E}[X] &= \sum_{\omega\in\Omega} X(\omega)\mu(\omega) \\ &\geq \sum_{\substack{\omega\in\Omega \\ X(\Omega)>C\nu}} X(\omega)\mu(\omega) \\ &> \sum_{\substack{\omega\in\Omega \\ X(\Omega)>C\nu}} C\nu\mu(\omega) \\ &= C\nu\mathbb{P}(X > C\nu) \end{aligned}$$

så om vi delar bägge sidor av detta med $C\nu$ får vi resultatet. □

[23] Var i beviset använder vi antagandet att X är ickenegativ?

Vad för slags problem kan man tänkas tillämpa det här verktyget på?

Definition 17. En Erdős-Renyi-graf (med parametrar n och p) är en slumpmässig graf G på n noder, där varje kant är med med sannolikhet p, oberoende av om varje annan kant är med. Vi skriver att $G \sim G_{n,p}$.

Om $p = \frac{1}{2}$ så kommer alltså $G_{n,p}$ helt enkelt vara en likformigt slumpmässig graf på n noder, men för det mesta kommer vi vara intresserade av fallet när p är en funktion av n.

Vi kan säga en hel del om den "lokala" strukturen av en $G_{n,p}$ bara med de verktyg vi har lärt oss hittills – alltså de egenskaper hos den

som vi kan avgöra om de gäller genom att studera den lokalt runt varje nod. Desto svårare blir det om vi ställer oss frågor om ifall den är, till exempel, sammanhängande.

Ett exempel på en lokal struktur är ifall vi har isolerade noder – om vi tänker oss det som att noderna är personer och kanterna är vänskapsrelationer är vi alltså intresserade av sannolikheten att inte ha några vänner.[24]

[24] Som vi alla vet går denna upp markant om man studerar matematik, men vår modell är inte sofistikerad nog att fånga detta.

Proposition 18. *Om $p > \frac{c\log(n)}{n}$ för något $c > 1$ så finns det asymptotiskt nästan säkert*[25] *inga isolerade noder*[26] *i $G_{n,p}$.*

[25] Vad sjutton betyder det? Det betyder att, om p_n är en följd sådan att $p_n > \frac{\log(n)}{n}$ för varje n, och G_n är en G_{n,p_n} för varje n, så går sannolikheten att G_n har en isolerad nod mot noll.

[26] Och vad är en isolerad nod? Det är en nod utan grannar.

Bevis. Beviset följer ett mönster som förhoppningsvis börjar bli bekant vid det här laget.

Låt $G \sim G_{n,p}$. Vi låter I_i vara händelsen att nod i är isolerad, och konstaterar att om I är mängden av isolerade noder i G så är

$$\mathbb{E}\left[|I|\right] = \sum_{i=1}^{n} \mathbb{P}\left(I_i\right)$$

så vad vi behöver räkna ut är sannolikheten att en viss given nod är isolerad.

Detta är en relativt enkel beräkning – det finns $n-1$ noder som i hade kunnat ha en kant till, och varje kant finns med sannolikhet p, oberoende av varje annan kant. Alltså är sannolikheten att inga av kanterna finns $(1-p)^{n-1}$, och vi har att

$$\mathbb{E}\left[|I|\right] = \sum_{i=1}^{n} \mathbb{P}\left(I_i\right) = n(1-p)^{n-1}.$$

Markovs olikhet ger oss nu, för varje $C > 0$, att

$$\mathbb{P}\left(|I| > C\mathbb{E}\left[|I|\right]\right) < \frac{1}{C}. \tag{1}$$

Hur omvandlar vi detta till det resultat vi vill ha? Om vi tar ett väldigt litet ϵ och låter $C = \frac{1-\epsilon}{\mathbb{E}[|I|]}$ så blir (1) till

$$\mathbb{P}\left(|I| > 1-\epsilon\right) < \frac{\mathbb{E}\left[|I|\right]}{1-\epsilon}.$$

Eftersom $|I|$ självklart enbart tar heltalsvärden är händelsen i vänster led precis samma händelse som händelsen att $|I| \geq 1$, det vill säga $I \neq \emptyset$.[27] Så om vi ersätter vänster led med detta får vi att

[27] Varje tänkbart värde på $|I|$ som är större än $1-\epsilon$ är också ≥ 1, och vice versa.

$$\mathbb{P}\left(I \neq \emptyset\right) < \frac{\mathbb{E}\left[|I|\right]}{1-\epsilon}$$

och här kan vi utan problem ta gränsvärdet $\epsilon \to 0$ och få[28]

[28] Notera att vi, när vi tar gränsvärdet här, måste ersätta $<$ med $\leq$ – för oss är det inget problem eftersom vi oavsett skall visa att ν går mot noll.

$$\mathbb{P}\left(I \neq \emptyset\right) \leq \mathbb{E}\left[|I|\right].$$

Så det enda som återstår att göra är att visa att $\mathbb{E}[|I|]$ går mot noll. Enligt satsens antaganden har vi att $p > c\frac{\log(n)}{n}$, så vi kan räkna att

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}[|I|] &= \lim_{n\to\infty} n(1-p)^{n-1} \\
&\leq \lim_{n\to\infty} n\left(1 - c\frac{\log(n)}{n}\right)^{n-1} \\
&= \lim_{n\to\infty} \frac{n}{1 - c\frac{\log(n)}{n}} \left(1 - \frac{c\log(n)}{n}\right)^{n} \\
&= \lim_{n\to\infty} n \underbrace{\frac{1}{1 - c\frac{\log(n)}{n}}}_{\to 1 \text{ när } n\to\infty} \left(\underbrace{\left(1 - \frac{c}{n/\log(n)}\right)^{\frac{n}{\log(n)}}}_{\to e^{-c} \text{ när } n\to\infty} \right)^{\log(n)} \\
&= \lim_{n\to\infty} n\left(e^{-c}\right)^{\log(n)} = \lim_{n\to\infty} n^{1-c}
\end{aligned}$$

och eftersom $c > 1$ går detta mot noll, såsom önskat.[29] □

Det finns några saker som är värda att anmärka på här. Om vi hade valt $c \leq 1$ hade inte vår räkning fungerat längre – och detta är inte ett sammanträffande eller ett resultat av att vi använde en svag metod.

I själva verket kan man visa att antalet isolerade noder faktiskt kommer gå mot oändligheten om $p < \frac{\log(n)}{n}$ – så vårt resultat är det bästa möjliga.

Vi nämnde innan att vi kan studera lokala problem som dessa enkelt, men att "globala" problem är svårare, och nämnde frågan om grafen är sammanhängande som ett exempel på en svår global fråga. I själva verket kan man visa att grafen kommer vara sammanhängande så snart vi inte längre har några isolerade noder – så detta resultat tillsammans med vad vi just visade visar alltså att en Erdős-Renyi-graf är sammanhängande så snart $p \geq \frac{c\log(n)}{n}$.

[29] I steget mellan rad fyra och fem går vi väldigt snabbt fram – egentligen hade man behövt kolla att

$$f(n,c) = \left(1 - \frac{c}{n/\log(n)}\right)^{\frac{n}{\log(n)}}$$

går mot e^{-c} snabbt nog att vi får lov att göra den substitutionen. Som tur är gäller det att $f(n,c) - e^{-c}$ är ungefär $\frac{\log(n)}{n}$ – specifikt

$$\frac{e^{-c} - f(n,c)}{\frac{\log(n)}{n}} \to \frac{c^2}{2e^c}$$

vilket är bra nog. Men detta är mycket mer analys än vad vi faktiskt vill göra i denna kurs.

Räkneregler för slumpvariabler

Vi sammanfattar vad vi lärt oss om slumpvariabler hittills i följande räkneregler:[30]

[30] Den här biten skippar vi på föreläsningen – den ligger här för att vara behjälplig som sammanfattning och när man gör övningarna. Den finns också i vår samling av formler och räkneregler.

Lemma 19. *Om (Ω, μ) är något sannolikhetsrum, $A \subseteq \Omega$ någon händelse, och $X, Y : \Omega \to \mathbb{R}$ samt $Z : \Omega \to V$ är slumpvariabler som tar värden i $\mathbb{R}$ och i någon godtycklig mängd V, så gäller att:*

1.
$$\mathbb{E}[X] = \sum_{x\in X(\Omega)} x\mathbb{P}(X = x) = \sum_{\omega\in\Omega} X(\omega)\mu(\omega).$$

2. *För alla $a, b \in \mathbb{R}$ så är*

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

Väntevärdet är alltså en linjär funktional.

3.

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A].$$

4. *Om $X(\omega) \leq C$ för varje ω, eller ekvivalent om $\mathbb{P}(X \leq C) = 1$, så är $\mathbb{E}[X] \leq C$.*

5. *Om $\mathbb{E}[X] = C$ så finns det åtminstone ett ω sådant att $X(\omega) \geq C$.*

6. *Markovs olikhet ger oss att, om $\mathbb{E}[X_n] \to 0$ för någon följd av ickenegativa slumpvariabler X_n som enbart tar heltalsvärden, så måste också $\mathbb{P}(X_n > 0) \to 0$.*

7. *Om Z är likformigt fördelad på V så gäller det för varje delmängd $W \subseteq V$ att*

$$\mathbb{P}(Z \in W) = \frac{|W|}{|V|}.$$

Övningar

Övning 1.

Definition 20. En familj $\mathcal{F}$ av delmängder till $[n]$ kallas för *skärande* om $A \cap B \neq \emptyset$ för alla par av A och B i $\mathcal{F}$.

Hur stor kan en skärande familj av mängder vara, om vi kräver att varje $A \in \mathcal{F}$ har storlek exakt k? Svaret ges av följande sats:

Teorem 21 (Erdős-Ko-Rado)**.** *För varje skärande familj $\mathcal{F}$ av delmängder av storlek k till $[n]$ gäller det, om $n \geq 2k$, att*

$$|F| \leq \binom{n-1}{k-1}.$$

Delfråga a: Hitta, för alla n och k, ett exempel på en skärande familj $\mathcal{F}$ av delmängder av storlek k till $[n]$ sådan att

$$|F| \leq \binom{n-1}{k-1}.$$

Delfråga b: Bevisa följande lemma:[31]

Lemma 22. *Antag att $n \geq 2k$, och låt $\mathcal{F} \subseteq \binom{[n]}{k}$ vara en skärande familj, och låt för varje $s \in \{0, 1, \ldots, n-1\}$*

$$A_s = \{s, s+1, \ldots, s+k-1\}$$

där additionen är modulo n. Då kan $\mathcal{F}$ innehålla högst k av mängderna A_s.

[31] Den här delen kräver ingen probabilistisk metod, det är bara ett direkt bevis. Sannolikhetsteorin kommer in i nästa delfråga.

Dra sedan slutsatsen från detta att samma lemma gäller även om vi tar någon permutation $\sigma \in S_n$ och låter

$$A_s = \{\sigma(s), \sigma(s+1), \ldots, \sigma(s+k-1)\}.$$

Delfråga c: Nu skall vi bevisa Erdős-Ko-Rado. Idén är att vi vill skapa en likformigt slumpmässig $A \overset{u}{\in} \binom{[n]}{k}$ och studera sannolikheten att denna ligger i $\mathcal{F}$ – det finns ett uppenbart uttryck för denna sannolikhet, och vi vill skapa A på ett sätt som gör att vi också kan använda vårt lemma från förra delfrågan för att begränsa den.

Beviset börjar alltså med "Tag en likformigt fördelad permutation $\sigma \overset{u}{\in} S_n$ och ett likformigt fördelat heltal $s \overset{u}{\in} \{0, 1, \ldots, n-1\}$". Skriv resten av beviset.

Övning 2. Bevisa följande proposition:

Proposition 23. *Antag att $v_1, v_2, \ldots, v_n$ är n stycken enhetsvektorer i $\mathbb{R}^n$, alltså $\|v_i\| = 1$ för alla i. Då finns det en följd $\eta_1, \eta_2, \ldots, \eta_n$, med $\eta_i = \pm 1$ för varje i, sådan att*

$$\left\| \sum_{i=1}^{n} \eta_i v_i \right\| \leq \sqrt{n}.$$

Övning 3. I denna övning skall vi bevisa följande resultat:

Teorem 24. *Låt $G = (V, E)$ vara någon graf, och antag att $|V| = n$ och $|E| = n\frac{d}{2}$ för något $d \geq 1$. Då finns det en oberoende mängd i G av storlek åtminstone $\frac{n}{2d}$.*

Idén för beviset är att vi tar en slumpmässig delmängd $S \subseteq V$ genom att ta med varje nod med sannolikhet $p = \frac{1}{d}$. Denna kommer så klart inte vara garanterad att vara en oberoende mängd – men om vi, för varje kant $\{u, v\}$ som kopplar ihop $u, v \in S$, tar bort u eller v ur S så blir den återstående mängden av noder oberoende.

Bevisa satsen genom att räkna ut väntevärdet av storleken på S och väntevärdet av antalet noder vi tvingas ta bort ur S, och se att vi kommer ha i genomsnitt $\frac{n}{2d}$ noder kvar.

Övning 4. Låt $p \in (0, 1)$ vara fixt. För varje $i \in \mathbb{N}$, låt

$$X_i = \begin{cases} 1 & \text{med sannolikhet } p \\ 0 & \text{annars,} \end{cases}$$

så att $X_1, X_2, X_3, \ldots$ blir en slumpmässig följd av nollor och ettor. Låt I vara det minsta I sådant att $X_I = 0$ – detta blir alltså ett slumpmässigt heltal.

Beräkna, för varje $i \in \mathbb{N}$, $\mathbb{P}(I = i)$. Räkna sedan ut $\mathbb{E}[I]$.

Övning 5. Antag att du samlar på Pokemonkort.[32] Vi föreställer

[32] Ersätt med ditt favorit-gacha med lootboxes om du vill ha ett mer samtida exempel.

oss en väldigt enkel modell för hur du får ett nytt kort – det finns n stycken distinkta kort totalt, och du kan köpa ett nytt kort åt taget. Det nya kortet du får är likformigt fördelat i samlingen av kort – varje kort är lika sannolikt.

När du köper ditt första kort är du garanterad att få ett kort du inte har innan. När du köper ditt andra kort är sannolikheten bara $\frac{1}{n}$ att du råkar få det kort du redan fick en gång – men när du redan har de flesta av korten kommer du oftast bara att få ett kort du redan äger, inte ett nytt, så du behöver köpa väldigt många paket för att gå från att ha en samling av $n-1$ kort till att ha en fullständig samling.

Låt T vara antalet gånger du behöver köpa ett nytt kort för att få en fullständig samling, om du börjar på noll. Beräkna $\mathbb{E}[T]$.[33]

Övning 6. Använd väntevärdets linjäritet för att ge ytterligare ett alternativt bevis för unionsbegränsningen.[34]

[33] Ledtråd: Lösningen på det här problemet använder lösningen på föregående problem.

[34] Ledtråd: Återigen algebra med indikatorvariabler. Hur uttrycker man $\mathbb{1}_{\{A\cup B\}}$ i termer av $\mathbb{1}_{\{A\}}$ och $\mathbb{1}_{\{B\}}$?