

Föreläsning 9: Diskret sannolikhetssteori · 1MA020

Vilhelm Agdur¹

16 februari 2023

¹ vilhelm.agdur@math.uu.se

Vi introducerar den diskreta sannolikhetssteorin, vilket är den som behandlar samma sorts objekt som kombinatoriken.

Varför har vi ett avsnitt om diskret sannolikhetssteori² i en kurs om kombinatorik?

Diskret sannolikhetssteori och kombinatorik studerar samma klass av objekt – diskreta strukturer – så områdena överlappar. Ofta är vad man ser i början när man lär sig om sannolikhetssteori olika problem vars lösning kan sammanfattas som “översätt till ett problem med att räkna någonting, lös det kombinatorikproblemet, och översätt tillbaka till en sannolikhet”.

Så det är en anledning till att prata om diskret sannolikhetssteori i denna kurs – vi kan få många exempel, och de exemplen är ofta mer praktiskt tillämpbara än motsvarande kombinatorikproblem. Så när man börjat tröttna på överdrivet abstrakta exempel, eller klämkäcka exempel om glasskiosker, kan sannolikhetssteorin komma som en frisk fläkt.

Men det är så klart inte så att fälten bara överlappar i ena riktningen – det är precis lika sant att det finns många kombinatoriska problem där den enklaste och vackraste lösningen använder sannolikhetssteori. Detta kallas för den *probabilistiska metoden*³ – i dess vanligaste form visar vi att något kombinatoriskt objekt måste existera genom att vi visar att ett slumpmässigt valt objekt kan ha egenskapen. I många fall känner vi inte till något konkret exempel på ett sådant objekt – bara att det måste existera.

Men låt oss börja med att definiera vad vi egentligen menar med diskret sannolikhet.

Händelser och sannolikheter

Definition 1. Ett sannolikhetsrum (Ω, μ) består av en mängd Ω och en funktion $\mu : \Omega \rightarrow [0, 1]$, sådana att

- Ω är icke-tomt och ändlig eller uppräkneligt oändlig⁴,
- det gäller att

$$\sum_{\omega \in \Omega} \mu(\omega) = 1.$$

Mängden Ω kallas för *utfallsrum*, och elementen ω i Ω alltså för *utfall*.⁵ Funktionen μ kallar vi för vårt sannolikhetsmått.

² I kursplanen kallat “klassisk sannolikhetssteori”, vilket jag tolkar som en mer tvetydig term för “diskret sannolikhetssteori”.

³ På engelska *the probabilistic method* – det finns en utsökt bok med just denna titel av Noga Alon och Joel Spencer som utforskar just detta ämne.

Den lämpar sig definitivt inte som första bok om varken sannolikhetssteori eller kombinatorik, men har man läst någon kurs i vardera ämne och uppnått lite matematisk mognad är den nog ett bra men utmanande val av bok.

⁴ Det vill säga, antingen är den ändlig, eller så kan vi numrera alla dess element $1, 2, 3, \dots$. Detta är skillnaden mellan diskret och kontinuerlig sannolikhetssteori – i kontinuerlig sannolikhetssteori tillåter vi oss överuppräknliga mängder.

I den kontinuerliga sannolikhetssteorin behöver man alltså kunna “räkna” saker som är fler än heltalen – det vill säga ta integraler. Som ni kanske är medvetna är det inte helt okomplicerat att definiera vad det ens betyder att ta en integral i allmänhet. Alltså stannar vi i den trevliga diskreta världen, där vi har summor istället för integraler.

⁵ För den som är van med programmering, och vet hur slumpgeneratorer i datorn fungerar, bör man tänka på ω som *seed* till slumpgeneratoren. Det är oftast inte ett intressant objekt i sig, men det bestämmer allt som slumpmässigt händer.

Definition 2. En *händelse* A är en delmängd till Ω . Dess *sannolikhet* ges av

$$\mathbb{P}(A) = \sum_{\omega \in A} \mu(\omega).$$

Notera här att vi definierar sannolikheter för *händelser*, **inte för utfall**.⁶

Exempel 3. Låt oss formulera de mest uppenbara exemplen av slump i vår nya terminologi.

Vi kan se ett tärningskast som att det har utfallsrum

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

och sannolikhetsmått μ som skickar varje utfall på $\frac{1}{6}$, alltså $\mu(\omega) = \frac{1}{6}$ för alla ω .

Vad är sannolikheten att vårt tärningskast ger oss ett udda tal? Jo, vad vi frågar efter är sannolikheten för händelsen $U = \{1, 3, 5\}$, vilken vi beräknar som

$$\mathbb{P}(U) = \sum_{\omega \in U} \mu(\omega) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Om vi singlar slant har vi istället utfallsrum $\Omega = \{\text{krona}, \text{klave}\}$, och vårt sannolikhetsmått μ är lika med $\frac{1}{2}$ för bägge utfallen.

Sannolikheten att vi får krona är alltså sannolikheten av *händelsen* $\{\text{krona}\}$, och ges av

$$\sum_{\omega \in \{\text{krona}\}} \mu(\omega) = \mu(\text{krona}) = \frac{1}{2}.$$

Vi kan också använda våra kunskaper från tidigare föreläsningar för att lösa mer invecklade problem.

Exempel 4. Antag att en grupp av n stycken försupna studenter går på en efterfest. När festen till slut är över är alla överraskande nog kapabla att gå hem, men ingen är nykter nog att känna igen sin egen jacka, så de bara tar en slumpmässig jacka på vägen ut.

Vad är sannolikheten att *ingen* student kommer hem med sin egen jacka?

Vi får fundera ett ögonblick på hur vi formaliserar det här problemet. Vi kan skriva tilldelningen av jackor till studenter som en permutation av längd n ur alfabetet av jackor.

Om vi numrerar studenterna och jackorna, så att student ett kom dit i jacka ett, student två i jacka två, och så vidare, så kan vi betrakta utfallet som en permutation av $[n]$. Alltså kan vi sätta $\Omega = S_n$, alltså mängden av sådana permutationer.

Hur skall vi tänka för att lista ut vad sannolikheten för varje given permutation är? Vi kan resonera på ett komplicerat sätt med olika

⁶ Detta beror på att vi i den kontinuerliga sannolikhetsteorin inte längre kan definiera μ som en funktion från utfall till reella tal, utan måste definiera den som ett genuint *mått*, alltså en funktion från *händelser* (=delmängder) till utfall. Precis som vi inte kan ta integralen av en funktion i en enda punkt, utan tar integraler över intervall.

Ibland kan vi komma att vara slarviga och prata om sannolikheter för enskilda utfall, men det korrekta sättet att skriva är alltid sannolikheten för händelsen $\{\omega\}$, inte för utfallet ω .

ordningar de kan gå ut i, och varje student tar varje kvarvarande jacka med samma sannolikhet, för att få svaret, eller så kan vi resonera på ett enkelt sätt.

Eftersom alla studenterna är förpackade för att kunna se skillnad på jackor är problemet helt symmetriskt – det finns ingen anledning till varför något specifikt utfall skulle vara mer sannolikt än något annat, eftersom studenterna inte kan se skillnad på utfallen oavsett.⁷ Alltså måste sannolikheten för varje utfall vara lika, och för att de skall summera till 1 måste de alltså vara $\frac{1}{n!}$.

⁷ Jackorna har temporärt gjorts osärskiljbara av övermåga drickande.

Som nästa steg i vår räkning får vi fundera på vad händelsen att ingen student går hem med sin egen jacka är. Det betyder, i vår formulering av utfallen som permutationer av n , att $\omega_i \neq i$ för alla i , alltså att ω är ett derangemang. Så vår händelse är mängden av derangemang, vilka vi ju redan räknat i en tidigare föreläsning att den har storlek

$$n! \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

Så vi kan räkna ut att

$$\begin{aligned} \mathbb{P}(\text{Ingen har sin egen jacka}) &= \mathbb{P}(\{\omega \in S_n : \omega(i) \neq i \forall i\}) \\ &= \sum_{\omega \in S_n : \omega(i) \neq i \forall i} \mu(\omega) \\ &= \sum_{\omega \in S_n : \omega(i) \neq i \forall i} \frac{1}{n!} \\ &= \frac{1}{n!} |\{\omega \in S_n : \omega(i) \neq i \forall i\}| \\ &= \frac{1}{n!} n! \sum_{k=0}^n \frac{(-1)^k}{k!} = \sum_{k=0}^n \frac{(-1)^k}{k!} \end{aligned}$$

vilket vi känner igen som de första n termerna i Taylorutvecklingen av e^{-1} , så sannolikheten att ingen får med sig sin egen jacka är, för stora nog n , ungefär 36.8%.

Betingad sannolikhet

Ett av de allra mest användbara verktygen för att resonera om sannolikheter är *betingad sannolikhet*. Det låter oss utföra argument av stilen “under antagandet att A är sant så är sannolikheten för B det här, så eftersom vi vet sannolikheten av A är sannolikheten för B detta”. Vi bryter alltså ned ett potentiellt svårt problem i enklare beståndsdelar.

Rent konkret gör vi detta med följande definition:

Definition 5. Givet två händelser A och B , sådana att $\mathbb{P}(B) > 0$, definierar vi *sannolikheten för A givet B* som

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Vi kan tolka detta som att vi “zoomat in” på bara B , och skapat oss ett nytt sannolikhetsmått $\mu|_B$, som ges av

$$\mu|_B : B \rightarrow [0, 1] : x \mapsto \frac{\mu(x)}{\sum_{x \in B} \mu(x)}$$

så att

$$\mathbb{P}(A | B) = \mu|_B(A \cap B).$$

Så termen $\frac{1}{\mathbb{P}(B)} = \frac{1}{\sum_{x \in B} \mu(x)}$ är helt enkelt där för att få sannolikheterna i $\mu|_B$ att summera till ett.

Funderar man lite grann på saken bör det kännas rimligt att detta mäter sannolikheten för en händelse, givet att vi redan vet att händelsen B inträffar – vi räknar ju bara på utfallen där B inträffat.

Definition 6. Vi säger att två händelser A och B är *oberoende* ifall

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B).$$

Namnet motiveras av att detta är samma sak som att $\mathbb{P}(A | B) = \mathbb{P}(A)$ – att få veta huruvida B inträffade ger oss alltså ingen information alls om ifall A inträffade, vår skattning av sannolikheten för det är helt oförändrad.

Lemma 7 (Lagen om total sannolikhet). *Det är alldeles uppenbart från vår definition att*⁸

$$\mathbb{P}(A | B) \mathbb{P}(B) = \mathbb{P}(A \cap B).$$

Detta generaliserar enkelt till att, om vi har en samling B_1, B_2, \dots, B_k av disjunkta händelser sådana att

$$A \subseteq \bigcup_{i=1}^k B_i,$$

alltså sådana att närhelst A inträffar så inträffar exakt en av händelserna B_i , så är

$$\mathbb{P}(A) = \sum_{i=1}^k \mathbb{P}(A | B_i) \mathbb{P}(B_i).$$

I en enkel form får vi alltså för alla par av händelser A och B att

$$\mathbb{P}(A) = \mathbb{P}(A | B) \mathbb{P}(B) + \mathbb{P}(A | B^c) \mathbb{P}(B).$$

Exempel 8. Antag att vi har en urna som innehåller hundra glaskulor, som kan vara av glas eller sten, och kan vara antingen röda, gröna, eller blå. Vi drar upp en slumpmässig kula ur vår urna.

Om vi låter A vara händelsen att kulan vi drar är av glas, och B_r , B_g , och B_b vara händelserna att den är röd, grön, eller blå, så säger oss alltså lagen om total sannolikhet att

$$\mathbb{P}(A) = \mathbb{P}(A | B_r) \mathbb{P}(B_r) + \mathbb{P}(A | B_g) \mathbb{P}(B_g) + \mathbb{P}(A | B_b) \mathbb{P}(B_b).$$

⁸ Och som vår diskussion om hur vi använder betingad sannolikhet indikerade är detta den centrala egenskapen den har – den låter oss dela upp det potentiellt svåra problemet att förstå $A \cap B$ i de enklare problemen att förstå B och förstå A givet B .

Alltså: Om vi vet fördelningen mellan de olika färgerna i urnan (sannolikheterna för B_r , B_g , och B_b) och vi vet hur stor andel av varje given färg som är glaskulor, kan vi räkna ut hur stor andel av alla kulor som är av glas.⁹

Låt oss, innan vi går vidare, ge några basala räkneregler för sannolikheter, som sammanfattning av vad vi sett hittills:

Lemma 9. *Det gäller för alla händelser A och B att*

- per definition är $\mathbb{P}(A) = \sum_{\omega \in A} \mu(\omega)$,
- så $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
- och om A och B har tomt snitt, $A \cap B = \emptyset$, så är $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$,
- och om de inte nödvändigtvis har tomt snitt har vi att

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

- $\mathbb{P}(A \cap B) = \mathbb{P}(A | B) \mathbb{P}(B)$,
- och per definition är A och B oberoende precis när $\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B)$.

Unionsbegränsningar, med tillämpning på Ramseytalen

Vi hade kunnat ge hela detta avsnittet av kursen med enbart exempel om kulor av olika färger i olika urnor – av någon anledning är det det första exemplet som dyker upp i de flesta probabilisters huvud. Låt oss undvika det, och istället ge ett lite mer intressant exempel.

Vi börjar med att påminna oss om ett resultat vi bevisade tidigare, bara omklätt i probabilistisk skrud:

Lemma 10 (Inklusion-exklusion). *Det gäller, för varje samling av händelser A_1, A_2, \dots, A_n , att*

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{\substack{I \subseteq [n] \\ |I|=k}} \mathbb{P}\left(\bigcap_{i \in I} A_i\right).$$

Så specifikt har vi för varje par av händelser A och B att

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Bevis. Vi utelämnar det, eftersom det är så snarlikt till det kombinatoriska fallet vi redan bevisat. \square

Låt oss nu introducera ett oerhört potent lemma, som ändå är väldigt enkelt:

⁹ Som vi formulerat det här är det ju oerhört oöverraskande att vi kan göra det. Så är det – det är inget märkligt som pågår här. Vad som är överraskande är om något hur ofta lagen om total sannolikhet är användbar, vilket vi kommer se i senare exempel.

Lemma 11 (Unionsbegränsning). *Antag att vi har en samling av händelser A_1, A_2, \dots, A_k . Vi är intresserade av sannolikheten att någon av händelserna inträffar, alltså sannolikheten för deras union. Det gäller att*

$$\mathbb{P}\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{i=1}^k \mathbb{P}(A_i).$$

Bevis. Inklusion-exklusion ger oss att

$$\mathbb{P}\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k \mathbb{P}(A_i) - \left(\sum_{k=2}^n (-1)^k \sum_{\substack{I \subseteq [n] \\ |I|=k}} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) \right)$$

och vi kommer ihåg att de extra termerna, som vi stoppat in i ett minustecken, är en korrektion för att vi råkat räkna punkterna i snitten mellan A_i och A_j för många gånger, så alltså måste vi göra höger led större om vi stryker den korrektionen. Alltså har vi vår sökta olikhet. \square

Övningar

Övning 1. Ge ett bevis för Lemma 7.

Övning 2. Visa att

$$\sum_{i=1}^k \mathbb{P}(A_i) = \sum_{\omega \in \Omega} |\{i : \omega \in A_i\}| \mu(\omega).$$

Använd detta för att ge ett alternativt bevis för Lemma 11.