

Föreläsning 10: Diskret sannolikhetssteori, fortsättning · 1MA020

Vilhelm Agdur¹

23 februari 2023

¹ vilhelm.agdur@math.uu.se

Vi fortsätter att diskutera diskret sannolikhetssteori, 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 sannolikhetssteorin – 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 $X : \Omega \rightarrow V$. Givet varje utfall tar alltså vår slumpvariabel ett visst värde, och givet varje² delmängd $A \subseteq V$ blir $X \in A$ en händelse – specifikt är det händelsen

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

Det allra vanligaste fallet är när $V = \mathbb{R}$ eller någon delmängd till \mathbb{R} . I många introtexter om sannolikhetssteori *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, $X : \Omega \rightarrow \mathbb{R}$ 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

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

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.

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 $X : V \rightarrow V$ 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} ges av⁴

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

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

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

precis som vi förväntade oss.

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

Lemma 6. Det gäller att⁵

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

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 Notera att detta är en summa över *alla värden som X kan tänkas ta* – eftersom vi antagit att Ω är ändligt eller uppräknligt så kommer detta vara en summa över ändligt eller uppräknligt 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.

⁵ Det här fungerar bara för att vi har antagit att våra sannolikhetsrum är ändliga eller uppräknligt 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.⁶

Lemma 7 (Väntevärdets linjäritet). *Givet två slumpvariabler X och Y och två reella tal a och b gäller det att*

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

*Väntevärdet är alltså linjärt, som funktion från rummet av slumpvariabler in i \mathbb{R} .*⁷

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

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

□

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.⁸*

Det gäller att

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

Bevis. Per definition har vi att

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

□

⁶ Jag är så klart oerhört partisk, eftersom just gränslandet mellan kombinatorik och sannolikhetsteori är mitt område – men det är onekligen ett otroligt användbart resultat.

⁷ 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 $\langle X, Y \rangle = \mathbb{E}[XY]$, 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.

⁸ 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.

Sperners lemma

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

Definition 9. En samling \mathcal{F} av delmängder till $[n]$ kallas för en *anti-kedja* ifall det för varje par $F, G \in \mathcal{F}$ varken gäller att $F \subset G$ eller $G \subset F$.

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.⁹ Är det möjligt att hitta en ännu större genom att ha med delmängder av olika storlekar? Sperners lemma säger oss att svaret är nej.

⁹ 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 permutation σ av $[n]$, och låter I vara mängden av i sådana att

$$\{\sigma(1), \sigma(2), \dots, \sigma(i)\} \in \mathcal{F}.$$

Det är enkelt att se att I innehåller antingen noll eller ett element – om den innehöll både i och j , med $i < j$, vore ju

$$\{\sigma(1), \sigma(2), \dots, \sigma(i)\} \subset \{\sigma(1), \sigma(2), \dots, \sigma(i), \sigma(i+1), \dots, \sigma(j)\},$$

vilket skulle motsäga att \mathcal{F} är en antikedja.

Låt oss nu studera slumpvariabeln $X = |I|$. Att vi vet att I bara kan ha noll eller ett element ger oss omedelbart att $\mathbb{E}[X] \leq 1$,¹⁰ men låt oss studera detta väntevärde också på ett annat sätt.

Vi kan räkna att¹¹

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

Vad är sannolikheten att i ligger i I ? Att i ligger i I betyder att $\{\sigma(1), \sigma(2), \dots, \sigma(i)\} \in \mathcal{F}$, per definition. Så vad vi behöver förstå är den slumpmässiga mängden

$$U_i = \{\sigma(1), \sigma(2), \dots, \sigma(i)\}.$$

¹⁰ 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

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

□

¹¹ Här använder vi den kortare notationen $\mathbb{1}_{\{i \in I\}}$ för att beteckna $\mathbb{1}_{\{\omega \in \Omega: i \in I(\omega)\}}$.

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

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

$$\mathbb{P}(U_i \in \mathcal{F}_i) = \frac{|\mathcal{F}_i|}{\binom{[n]}{i}}.$$

Så samlar vi ihop vad vi har listat ut hittills i ett enda uttryck, och använder olikheten¹² $\binom{\lfloor \frac{n}{2} \rfloor}{i} \geq \binom{n}{i}$ för alla i , har vi att

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

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

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

vilket är precis Sperners lemma, som vi ville bevisa. \square

Räkneregler för slumpvariabler

Vi sammanfattar vad vi lärt oss om slumpvariabler hittills i följande räkneregler:¹³

Lemma 12. Om (Ω, μ) är något sannolikhetsrum, $A \subseteq \Omega$ någon händelse, och $X, Y : \Omega \rightarrow \mathbb{R}$ samt $Z : \Omega \rightarrow V$ 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 $\mathbb{E}[X] \leq C$.

¹² 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.

¹³ 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.

5. 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