Introduktion till sannolikhetsteori

MA1508

Sannolikhet

Vi kommer i huvudsak jobba med diskret sannolikhet, där det bara finns ändligt många möjliga utfall. T ex är en vanlig situation att vi har en informationskälla som slumpmässigt genererar ett element från en ändlig mängd $\mathcal{U}$. Låt elementen i $\mathcal{U}$ vara $u_1, u_2, \ldots, u_r$ och låt p_i vara sannoliketen att u_i är det valda elementet. Vi kommer ibland kalla $\mathcal{U}$ för ett alfabet, och elementen i $\mathcal{U}$ kan kallas för symboler. Ett annat namn för $\mathcal{U}$ är utfallsrum.

Om man hanterar situationer med oändligt många möjliga utfall blir det mesta sig likt, men ibland får man besvärliga teknikaliteter som man måste hantera. Vi slipper alltså det genom att hålla oss till ändliga utfallsrum, men ni kan känna till att vissa av våra satser måste omformuleras om vi tillåter andra fall.

Jag hoppas ni håller med mig att det måste gälla att $p_i \geq 0$ för $i \in \{1, 2, \ldots, r\}$. Jag hoppas att vi också är överrens om att

$$\sum_{i=1}^{r} p_i = 1.$$

Låt oss gå igenom lite elementära räknelagar för sannolikheter.

Exempel 1. Anta $\mathcal{U} = \{a, b, c, d\}$ och att $(p_0, p_1, p_2, p_3) = (0.5, 0.2, 0.2, 0.1)$. Vad är sannolikheten att a eller d genereras av källan? Det måste vara $p_0 + p_3 = 0.6$. $\triangle$

Mer allmänt, om det finns ett antal ömsesidigt uteslutande möjligheter, så måste sannolikheten att en av dem inträffar vara lika med summman av sannolikheterna för de olika utfallen.

Vi kan använda mängdläranotation för att uttrycka det. Vi har ett ett utfallsrum $\mathcal{U}$ av möjliga saker som kan hända, och så har vi delmängder $A_1, \ldots, A_n$ till $\mathcal{U}$. Med $P(A_i)$ betecknar vi sannolikheten att slumpen väljer ett element från A_i. (Man talar om de olika delmängderna till $\mathcal{U}$ som *händelser*. Om $\mathcal{U}$ representerar möjliga utfall från ett tärningskast så kan en händelse t ex vara att det blir ett udda antal prickar som kommer upp.) Om det gällller att $A_i \cap A_j = \emptyset$ för alla i, j som uppfyller att $i \neq j$ så gäller det också att

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i).$$

Exempel 2. Anta vi har samma källa som i förra exemplet.

Låt oss nu anta att vi låter källan generera två symboler i följd, och att de olika symbolerna genereras *oberoende* av varandra. Vad är sannolikheten att strängen ac genereras? Jag påstår att den sannolikheten är $p_1 \cdot p_3 = 0.1$. △

I själva verket gör vi den matematiska definitionen att två händelser, A och B, är *oberoende* om $P(A \cap B) = P(A)P(B)$. Notera att $A \cap B$ motsvarar att både A och B inträffar.

Exempel 3. Låt oss anta vi kastar en röd och en blå tärning. De är vanliga sexsidiga tärningar. Vi har ett utfallsrum som består av vektorer som $(1, 2)$, där vi låter första elementet var resultatet från den röda tärningen. Sammanlagt innehåller utfallsrummet 36 olika element. Varje element i utfallsrummet har sannolikheten $\frac{1}{36}$.

Låt A vara händelsen att den röda tärningen visar 2. Så $A = \{(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6)\}$. Låt B vara händelsen att den blå tärningen visar 3. Är händelserna A och B oberoende?

Låt oss kolla det noggrant.

$$P(A \cap B) = P(\{(2, 3)\}) = \frac{1}{36}.$$

$$P(A) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}.$$

$$P(B) = \frac{1}{6}.$$

Eftersom $P(A)P(B) = \frac{1}{36}$ så är händelserna oberoende.

Å andra sidan om C är händelsen att summa av tärningarna är 12 och D är sannolikheten att den röda tärningen visar 6 så är C och D inte oberoende händelser. Ni kan kontrollera att $P(C \cap D) = \frac{1}{36}$ men $P(C)P(D) = \frac{1}{36} \cdot \frac{1}{6}$.
△

Betingad sannolikhet

Låt A och B vara två händelser, båda med positiv sannolikhet. Vi definierar den *betingade sannolikheten* för A givet B, som

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

Den betingade sannolikheten $P(A|B)$ ger oss sannolikheten för att A inträffat, om vi vet att B inträffat.

Exempel 4. Anta vi kastar en vanlig tärning. Låt A vara händelsen att en etta resulterar och låt B vara händelsen att tärningen visar ett udda tal. Då är $P(A \cap B) = \frac{1}{6}$ och $P(B) = \frac{1}{2}$, så

$$P(A|B) = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Känns det som ett intuitivt rimligt svar? △

Vi kan koppla begreppet betingad sannoliket till begreppet oberoende händelser.

Proposition. *Om A och B är oberoende händelser med positiv sannolikhet så är*

$$P(A|B) = P(A).$$

Bevis. Om A och B är oberoende kan vi räkna såhär:

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

□

Betingade sannolikheter är viktiga i många sammanhang inom vår kurs. T ex anta vi studerar textkomprimering. En första modell kan vara att anta att vi har en källa som genererar symboler, $x_1, x_2, \ldots$ från alfabetet a, b,... , å,ä,ö med någon viss sannolikhet för varje bokstav, och att de olika symbolerna genereras oberoende av varandra. T ex tänker vi oss att sannolikheten $P(x_1 = \text{ä})) = 0.017$. Men antagandet om oberoende medför att $P(x_1 = \text{ä} \wedge x_2 = \text{ä}) = 0.017^2$. En bättre modell kan ta hänsyn till att $P(x_2 = \text{ä}|x_1 = \text{ä})$ är väldigt liten, nästan 0 i svenska språket.

Vi nämner nu en viktigt formel för betingade sannolikheter.

Sats (Bayes' sats)**.** *Om A och B är händelser med positiv sannolikhet så är*

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

Bevis. Vi vet att

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

och att

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

Så $P(A \cap B) = P(B|A)P(A)$.

Sätt in detta i uttrycket för $P(A|B)$:

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

□

Exempel 5. Anta att 2% av befolkningen har en viss sjukdom. Det finns ett test för sjukdomen alltid ger rätt svar för sjuka individer, men som också ger ett falskt positivt utslag för 10% av de som är friska. Om en slumpmässigt vald person från befolkningen testas, vad är sannolikheten att personen verkligen är sjuk om testet ger ett positivt resultat?

Detta är en klassisk typ av problem för att applicera Bayes' sats. Låt A vara händelsen att person är sjuk och B vara händelsen att testet ger ett positivt utslag. Med A^c menar vi händelsen att personen inte är sjuk. Enligt förutsättningarna gäller att $P(A) = 0.02$. Formuleringerna om testet betyder att $P(B|A) = 1$ och $P(B|A^c) = 0.1$. Vi kan räkna ut att

$$P(B) = P(A \cap B) + P(A^c \cap B) = P(A)P(B|A) + P(A^c)P(B|A^c) =$$

$$= 0.02 \cdot 1 + 0.98 \cdot 0.1 = 0.118$$

Vi vill räkna $P(A|B)$. Enligt Bayes' sats är

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{1 \cdot 0.02}{0.118} \approx 0.169.$$

Så bara runt 17% av de som testar positivt är verkligen sjuka. △

Stokastiska variabler

Låt oss introducera ett begrepp med ett lustigt namn. Med en *stokastisk variabel* (på engelska heter det Random Variable) menar vi bara en funktion på ett ufallsrum. För vår del kommer stokastiska variabler oftast att vara reellvärda.

Exempel 6. Låt oss anta en informationskälla genererar en av fyra symboler a,b,c, d. Utfallsrummet kan allså antas vara $\{a, b, c, d\}$. En av många stokastiska variabler på detta uftallsrum är funktionen X sådan att $X(a) = 1$, $X(b) = 2$, $X(c) = 3$ och $X(d) = 4$.

△

I exemplet ovan specificerade jag inte ens vad sannolikhetsfördelningen på utfallsrummet var, för det behövdes inte för att definiera den stokastiska variabeln. Däremot är det viktigt för många räkningar med stokastiska variabler.

Exempel 7. Låt oss lägga till en sannolikhetsfördelning till föregående exempel. Vi säger att sannolikheten att symbolen a genereras är $p_a = 0.3$. För de andra gäller $p_b = 0.2$, $p_c = 0.1$ och $p_d = 0.4$.

Vi kan nu prata om sannolikheten att X antar olika värden. T ex är $P(X = 3) = 0.1$. Jag hoppas ni kan köpa notationen $P(X = 3)$, den är smidig.

△

Anta att vi har en stokastisk variabel, X, med del möjliga utfallen $x_1, x_2, \ldots, x_k \in \mathbb{R}$ och att sannolikheterna för utfallen är $p_1, p_2, \ldots, p_k$. (Så $P(X = x_1) = p_1$, osv.) Då kan vi introducera ett nytt begrepp med ett lustigt namn. *Väntevärdet* av X definieras som

$$\sum_{i=1}^{k} p_k x_k.$$

En beteckning för väntevärdet av X som ofta används är $E(X)$.

Exempel 8. Vi fortsätter med samma exempel som förut. Det gäller då att $E(X) = 1 \cdot 0.3 + 2 \cdot 0.2 + 3 \cdot 0.1 + 4 \cdot 0.4 = 2.6$. △

Idén med väntevärdet är att det ska motsvara vad man kan förvänta sig i genomsnitt för värde av den stokasiska variabeln. Så väntevärdet behöver inte vara ett värde den stokastiska variabeln ens kan anta, men den är ett genomsnittsvärde, på ett sätt vi kommer förklara noggrannare.

Exempel 9. Låt oss fortsätta med utfallsrummet $\{a, b, c, d\}$ och samma sannolikhetsfördelning som förut. Vi vill kommunicera vad källan producerar för symboler över en informationskanal som använder sig av binära sekvenser. Vi väljer därför en kodning, K, som omvandlar symbolerna till binära strängar. Den vi väljer uppfyller att $K(d) = 1$, $K(a) = 01$, $K(b) = 000$ och $K(c) = 001$.

Vi definierar $L(\alpha)$ som längden av $K(\alpha)$, där $\alpha \in \{a, b, c, d\}$. L är en reellvärd stokastisk variabel, så vi kan räkna dess väntevärde.

$$E(L) = 0.4 + 2 \cdot 0.3 + 3 \cdot 0.2 + 3 \cdot 0.1 = 1.9.$$

Notera att det finns en alternativ kod där vi använder 2 bitar till att koda alla symbolerna. I genomsnitt används uppenbarligen fler bitar med den koden. △

Vi nämner nu en intressant räknelag för väntevärden.

Proposition. *Om X och Y är stokastiska variabler så är*

$$E(X + Y) = E(X) + E(Y).$$

Om c är en konstant så är

$$E(cX) = cX.$$

Precis som händelser kan vara oberoende finns det en definition av oberoende för stokastiska variabler.

Definition. Två stokastiska variabler, X och Y, sägs vara *oberoende* om

$$= P(X = \alpha \wedge Y = \beta) = P(X = \alpha)P(Y = \beta)$$

för alla $\alpha, \beta \in \mathbb{R}$.

Proposition. *Om X och Y är oberoende stokastiska variabler så är*

$$E(XY) = E(X)E(Y).$$

Notera att oberoende krävs i propositionen ovan, men inte för att $E(X + Y) = E(X) + E(Y)$.

Exempel 10. Johan prospekterar för guld. Antalet gram guld han gräver upp under en dags är en stokastisk variabel, G, med väntevärde 0.9. Priset (i SEK/gram) han får när han säljer guldet är en stokastisk variabel, P, med väntevärde 1170. Om variabler kan antas vara oberoende har hans dagliga inkomst väntevärdet

$$E(GP) = 0.9 \cdot 1170 = 1053.$$

$\triangle$

Stora talens lag

Anta vi har en följd av oberoende stokastiska variabler $X_1, X_2, \ldots$ med samma fördelning och väntevärdet m. Vi kan t ex oss tänka att de stokastiska variabler modeller resultatet av tärningskast eller ett kvantmekaniskt experiment som upprepas många gånger.

Låt oss defininera

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Det är medelvärdet av de första n stokastiska variablerna i följden.

Som vi kanske väntat blir $E(\bar{X}_n) = m$ vilket vi kan se med följande räkning:

$$E(\bar{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot nm = m.$$

Vi kan nu också förklara varför väntevärdet är ett så viktigt begrepp.

Sats (Stora talens lag)**.** *Anta vi har en följd av oberoende stokastiska variabler $X_1, X_2, \ldots$ med samma fördelning och väntevärdet m. Låt ϵ_1 och ϵ_2 vara positiva tal. Då finns det ett tal N sådant att om $n > N$ så gäller det att*

$$P\left(m - \epsilon_1 < \bar{X}_n < m + \epsilon_1\right) > 1 - \epsilon_2.$$

Om vi väljer ϵ_1 och ϵ_2 till att vara små tal så betyder $P\left(m - \epsilon_1 < \bar{X}_n < m + \epsilon_1\right) > 1 - \epsilon_2$ att $\bar{X}_n$ kommer ligga väldigt nära m med väldigt hög sannolikhet, bara n väljs tillräckligt stort.

Det är här vi kan tänka oss namnet väntevärde kommer ifrån. Vi kan vänta oss att medelvärdet från många försök kommer ligga nära väntevärdet. En noggrannare sats än stora talens lag säger fördelningen av $\bar{X}_n$ kommer vara nästan normalfördelad. För att presentera den satsen lite bättre ska vi introducera lite ytterligare terminologi.

Varians och standardavvikelse

Definition. Låt X vara en stokastisk variabel med väntevärde m. *Variansen* av X definieras som

$$\mathrm{Var}(X) = E((X - m)^2).$$

Exempel 11. Om $P(X = 0) = \frac{3}{4}$ och $P(X = 1) = \frac{1}{4}$ så är $E(X) = \frac{1}{4}$. För att förklara räkningarna av variansen defininerar vi $T = (X - \frac{1}{4})^2$. Om $X = 0$ blir $T = \frac{1}{16}$, vilket inträffar med sannolikheten $\frac{3}{4}$. Om $X = 1$ så blir $T = \frac{9}{16}$, vilket inträffar med sannolikheten $\frac{1}{4}$.

Så

$$\mathrm{Var}(X) = E(T) = \frac{1}{16} \cdot \frac{3}{4} + \frac{9}{16} \cdot \frac{1}{4} = \frac{12}{64} = \frac{3}{16}.$$

$\triangle$

Variansen av en stokastisk variabel är ett mått på hur mycket dess värden förväntas varierera runt väntevärdet. Man pratar också om *standardavvikelsen*, som är kvadratroten av variansen.

Exempel 12. Om $\mathrm{Var}(X) = 2$ så är standardavvikelsen för X lika med $\sqrt{2}$. $\triangle$

Det finns många tänkbara sätt att definiera ett sådant mått, variansen och standardavvikelsen är val som visat sig vara smidigt, men andra mått förekommer ibland i tillämpningar.

Notera att om c är en konstant så är $\mathrm{Var}(cX) = c^2 \,\mathrm{Var}(X)$. För oberoende stokastiska variabler finns det också en intressant räknelag när man tar summor.

Proposition. *Om X och Y är oberoende stokastiska variabler så är*

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Låt oss nu återgå till en följd av oberoende stokastiska variabler $X_1, X_2, \ldots$ med samma fördelning och väntevärdet m. Kalla variansen X_1 (och därmed även av $X_2, X_3, \ldots$) för w. Sätt $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Då gäller att

$$\mathrm{Var}(\bar{X}_n) = \frac{w}{n}.$$

Att variansen avtar med ökande n visar att fördelningen kommer bli allt mer koncentrerad runt väntevärdet, dvs med stor sannolikhet kommer $\bar{X}_n$ anta ett värde nära m. En sats som heter centrala gränsvärdessatsen säger att sannolikheten för att $\bar{X}_n$ antar olika värden nära m kan beskrivas väldigt bra med den så kallade normalfördelningen. (Som ni stött på i gymnasiet och förmodligen på BTH.)