Sannolikhet och statistik I

Johan Jonasson

Mars 2019

Några inledande småsaker:

- Titta regelbundet på hemsidan: http://www.math.chalmers.se/Stat/Grundutb/CTH/mve302/1819
- Det finns en detaljerad lista med lärandemål på hemsidan.
- Tidigare år har TM jobbat på egen hand med ett Matlabhäfte. Det gäller inte längre, men häftet finns på hemsidan och innehåller en del saker som är av värde för framtiden för både TM och Kf.

Finns slump? Ja, i kvantvärlden tycks det vara så, men det vi betraktar som slump är oftast i huvudsak mest osäkerhet; skeenden som är för komplexa för att man ska kunna ha kontroll över dem, men som i princip kan vara deterministiska.

Exempel.

- Hur kommer ett tärningsslag att utfalla? I princip förutsägbart, men i praktiken oförutsägbart.
- Hur kommer svenska folket att rösta i valet 2022?
- Finns liv på andra planeter?
- Dog farao Tutankhamon av en genetisk sjukdom?
- Härstammar fåglar från dinosaurier?

Vad är sannolikhet?

Förslag: Vad relativa frekvensen konvergerar mot. Till exempel, singla slant n gånger. Låt X_n vara antalet gånger du får klave. Rimligt att anse att sannolikheten att få klave vid kast med detta mynt är

$$\lim_{n \to \infty} \frac{X_n}{n}.$$

Två problem:

- Existerar gränsvärdet?
- Kräver upprepbarhet.

En del slumpfenomen är inte ens i princip upprepbara, men vi vill ändå använda sannolikhetsteori på dem. Några exempel.

- Vad är sannolikheten att Barcelona slår Chelsea i morgon?
- Vad är sannolikheten att det regnar på fredag?
- Vad är sannolikheten den här ultraljudsbilden visar en cancertumör?

Då är sannolikhet snarare en grad av tro, som man vill sätta siffror på. Den kan uppdateras av ny information, t.ex. det regnar idag". Viktigt inom AI!

Sannolikhetsteorin *säger inget* om vad sannolikhet är, bara hur den måste uppföra sig.

Man utgår alltid från en modell, precis som i annan matematik. Modellen har man goda skäl att tro på eller intressera sig för (sannolikhetsteori) eller vill testa (statistikteori).

Sannolikhetsteori

Ett försök vars utfall beror på slumpen kallas för ett *slumpförsök*.

Mängden S av alla möjliga utfall, u, kallas för försökets *utfallsrum*. Mer bildligt: Om vårt slumpförsök är ett lotteri, så är S tunnan med alla lotter.

Exempel. Singla slant. De möjliga utfallen är krona och klave. Boken skriver T och H för tails och heads. Med de beteckningarna får vi

$$S = \{H, T\}.$$

Exempel. Välj ett tal på måfå mellan 0 och 1.

$$S = [0, 1].$$

Exempel. Hur länge håller LED-lampan?

$$S = [0, \infty).$$

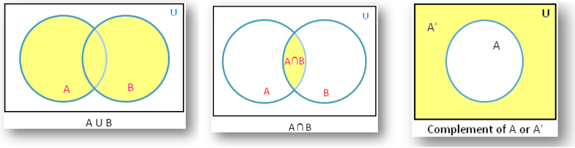
Definition: En delmängd A till S kallas för en *händelse*.

Exempel. Betrakta ett tärningskast. $S = \{1, 2, 3, 4, 5, 6\}$.

- $A = \{1, 3, 5\} = \{u \in S : u \text{ udda}\} = \{\text{udda utfall}\}$.
- $B = \{1, 2\} = \{u \in S : u \leq 2\} = \{\text{vi får högst en tvåa}\}$.

Om slumpförsöket får ett utfall $u \in A$, säger vi att "A inträffar".

Att säga att $A \cup B$ inträffar är då samma sak som att säga att A eller B inträffar, att säga att $A \cap B$ inträffar är samma sak som att säga att både A och B inträffar och att säga att A^c inträffar är samma sak som att säga att A inte inträffar.



Figur: Mängdoperationer

Sannolikhetsmått

Man skriver $\mathcal{P}(S)$ för mängden av alla delmängder till S.

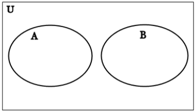
Definition: Om $\mathbb{P} : \mathcal{P}(S) \to \mathbb{R}$ uppfyller

(i) $0 \leq \mathbb{P}(A) \leq 1$ för alla A,

(ii) $\mathbb{P}(S) = 1$,

(iii) $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ för alla följder av parvis disjunkta mängder $A_1, A_2, \ldots$ (dvs sådana att $A_i \cap A_j = 0$ då $i \neq j$),

kallas $\mathbb{P}$ för ett *sannolikhetsmått*.



Figur: Disjunkta mängder.

Tag $A_1 = S$ och $A_2 = A_3 = \ldots = \emptyset$ i (iii) och få $\mathbb{P}(S) = \mathbb{P}(S) + \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) + \ldots$, vilket ger

$$\mathbb{P}(\emptyset) = 0.$$

Proposition 3.1: För alla händelser A och B gäller att

(a) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,

(b) $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$,

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

(d) $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$.

Bevis:

(a) Eftersom A och A^c är disjunkta och $A \cup A^c = S$ gäller enligt (ii) och (iii) att $1 = \mathbb{P}(S) = \mathbb{P}(A) + \mathbb{P}(A^c)$.

(b) Enligt (iii) gäller $\mathbb{P}(B) = \mathbb{P}((B \setminus A) \cup (A \cap B)) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)$.

(c) Enligt (iii) och sedan (b) gäller

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

(d) Om $A \subseteq B$ gäller $A \cap B = A$, så (b) $\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A) \geq \mathbb{P}(A)$ enligt (i).

Exempel. Risken för regn på lördag är 50%, risken för regn på söndag är 40%. Risken för regn båda dagarna är 30%. Vad är chansen till uppehåll hela helgen?

Vad är S? Inte så noga faktiskt och vi behöver inte ens specificera, men vi kan till exempel säga $S = \{(\text{regn},\text{regn}),(\text{regn},\text{uppehåll}),(\text{uppehåll},\text{regn}),(\text{uppehåll},\text{uppehåll})\}$.

Låt

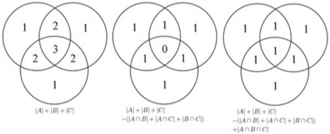
$$A = \{\text{regn på lördag}\}.$$
$$B = \{\text{regn på söndag}\}.$$

Vi söker $\mathbb{P}((A \cup B)^c)$.

$$\begin{aligned}\mathbb{P}((A \cup B)^c) &= 1 - \mathbb{P}(A \cup B)\\ &= 1 - (\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B))\\ &= 1 - (0.5 + 0.4 - 0.3) = 0.4.\end{aligned}$$

För tre händelser A, B och C gäller att

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ &\quad - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) \\ &\quad + \mathbb{P}(A \cap B \cap C).\end{aligned}$$

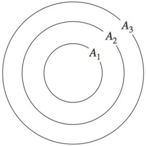


Figur: Inklusion-exklusionsprincipen.

Se gärna på den generella inklusion-exklusionsformel i boken, Proposition 1.5.

Kontinuitet hos sannolikheter

Antag att $A_1 \subseteq A_2 \subseteq \ldots$ och skriv $A = \bigcup_{n=1}^{\infty} A_n$. Då gäller enligt (iii) att (om vi låter $A_0 = \emptyset$)



Figur: Växande följd av händelser.

$$\begin{aligned}
\mathbb{P}(A) &= \sum_{n=1}^{\infty} \mathbb{P}(A_n \setminus A_{n-1}) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(A_n) - \mathbb{P}(A_{n-1}) \\
&= \lim_{N \to \infty} \sum_{n=1}^{N} \mathbb{P}(A_n) - \mathbb{P}(A_{n-1}) \\
&= \lim_{N \to \infty} \mathbb{P}(A_N).
\end{aligned}$$

Kortfattat: om $A_1 \subseteq A_2 \subseteq \ldots$ och $A = \bigcup_{n=1}^{\infty} A_n$ gäller att $\mathbb{P}(A_n) \to \mathbb{P}(A)$ då $n \to \infty$.

Om istället $A_1 \supseteq A_2 \supseteq \ldots$ och $A = \bigcap_{n=1}^{\infty} A_n$, gäller enligt det ovanstående att $\mathbb{P}(A_n^c) \to \mathbb{P}(A^c)$ och därmed också att $\mathbb{P}(A_n) \to \mathbb{P}(A)$ då $n \to \infty$.

Exempel. Välj ett tal på måfå i $[0, 1]$. Detta *betyder* att $S = [0, 1]$ och $\mathbb{P}(A) = \ell(A) =$ längden av A. Detta är ett sannolikhetsmått. (Vi kan faktiskt inte bevisa det eftersom det kräver en del måtteori att definiera längdmått.)

Låt till exempel $A_n = [0, 1/2 + 1/n]$. Då är $A = \cap_n A_n = [0, 1/2]$ och $\mathbb{P}(A_n) \to \mathbb{P}(A)$. (Detta kunde vi kollat direkt i det här fallet eftersom $\ell(A_n) = 1/2 + 1/n$ och $\ell(A) = 1/2$.)