SF1914/SF1916: Sannolikhetsteori och statistik
Föreläsning 2
Grundläggande sannolikhetsteori, betingade sannolikheter, oberoende händelser

Tatjana Pavlenko

30 augusti, 2018

KTH Matematik

Sannolikhetsgrunder (repetition)

- **Slumpförsöket** är en experiment som kan upprepas om och om igen och där resultatet inte kan på förhand avgöras.
- Resultatet av ett slumpmässigt försök kallas för **utfall**, betecknas med ω. Mängden av möjliga utfall kallas **utfallsrum**, bet. med Ω, det gäller att $\omega \in \Omega$.
- **Händelse** är uppsättning intressanta utfall. Bet. med $A, B, C, \ldots$. För en händelse A gäller det att $A \subset \Omega$, dvs A är *delmängd* av Ω.
- Sedan presenterades några viktiga händelser, Venndiagram samt operationer från grundläggande mängdlära.

Kolmogorovs axiomsystem (rep.)

- Sannolikhetsmått $P(A)$ för varje $A \subseteq \Omega$.
- Axiomatiska uppbyggnaden av sannolikhetslära. "Grundbegriffe", (1933) av A.N. Kolmogorov.
- Sannolikhetsmåttet P ska uppfylla följande axiom

Ax. 1: För varje händelse A gäller det att $0 \leq P(A) \leq 1$.

Ax. 2: För hela Ω gäller att $P(\Omega) = 1$

Ax. 3: Om $A_1, A_2, \ldots,$ är en följd av av parvis oförenliga händelser så gäller att

$$P\left(A_1 \cup A_2 \cup \ldots\right) = P(A_1) + P(A_2) + \cdots$$

Överensstämmer med frekvenstolkningen av $P(A)$.

- Masstolkning av $P(A)$:

  lägg ut massan av 1 på Ω. Då är $P(A) = \text{massan på A}$.

KTH Matematik

Likformig sannolikhetsfördelning (rep.)

- Ω består av m stycken lika möjliga utfall (utfallsrummet är diskret),

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\},$$

dvs. $p_i = P(\omega_i) = 1/m$ för alla $i = 1, \ldots, m$ och enl. *Ax. 3* får vi *den klassiska sannolikhetsdefinitionen.*

- Betrakta $A \subset \Omega$ som innehåller g utfall. Då gäller

$$P(A) = \sum_{\omega_i \in A} P(\omega_i) = \sum_{\omega_i \in A} \frac{1}{m} = \frac{\text{ant. för A gynsamma utfall}}{\text{ant. möjliga utfall}} = \frac{g}{m}.$$

- *Sats* (Multiplikationsprincipen). *Antag att åtgärd i kan utföras på a_i olika sätt där $i = 1, 2, \ldots, n$, dvs n st olika åtgärd föreligger. I så fall finns det totalt*

$$a_1 \cdot a_2 \cdot a_3 \cdots a_n$$

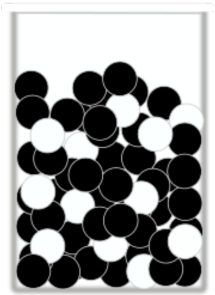
sätt att utföra de n åtgärderna.

- Med hjälp av multiplikationsprincipen kan man bestämma antalet sätt att välja ut k st element bland n distinkta. *Sats*:

	Med återläggn.(Må)	Utan återläggn.(Uå)
Med ordningshänsyn (Mo)	n^k	$\frac{n!}{(n-k)!}$
Utan ordningshänsyn (Uo)	$\binom{n+k-1}{k}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

- Urnmodeller. Ex på tavla.

KTH Matematik



FIGUR: En klassisk urnmodell är ett av statistikerns favoritobjekt som används för sannolikhetsberäkningar. I samband med likformiga sannolikhetsfördelningar finns många praktiska problem som kan lösas genom att återföra problemen till dragning av föremål från urnor.

I en urna finns s svarta och v vita kulor. Man drar slumpmässigt n kulor ur urnan.

Hur stor är sannolikhet att k vita kulor erhålls vid dragningen?

- Antag att dragning är *utan* återläggning. Då blir det sökta sannolikhet (see Blom, avsn, 2.5, del a))

$$\frac{\binom{v}{k}\binom{s}{n-k}}{\binom{v+s}{n}}$$

- Antag nu att dragning var *med* återläggning. Nu får vi (see Blom, avsn, 2.5, del b))

$$\binom{n}{k}\left(\frac{v}{v+s}\right)^k\left(\frac{s}{v+s}\right)^{n-k}.$$

Betingad sannolikhet

- Hur påverkar information om att en händelse inträffar sannolikheterna för att andra händelser gör det?
- Antag att vi vet att B har inträffat. Vad är sannolikhet av någon annan händelse A giver att B har inträffat? Bet. $P(A|B)$.
- Ex. på tavla: spamfiltrering.
- Definition. Låt A och B vara två händelser, $P(B) > 0$. Uttrycket

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

kallas *den betingade sannolikheten för A givet att B hat inträffat*.

KTH Matematik

Nyttiga satser om betingade sannolikheter

- *Sats*: Lagen om total sannolikhet.

  Betrakta händelserna $H_1, \ldots, H_n$ som är parvis oförenliga och $\cup_{i=1}^{n} H_i = \Omega$. Då gäller för varje händelse A att

  $$P(A) = \sum_{i=1}^{n} P\left(A|H_i\right) P\left(H_i\right).$$

- Exempel på tavla.

KTH Matematik

- Bayes' formel. *"Konsten att vända en betingad sannolikhet".* . Hur kan man beräkna $P(B|A)$ om man känner $P(A|B)$? Ur definitionen för betingning får vi att sannolikheten för snitthändelsen kan beräknas på två sätt: $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.

- Från detta får man

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

- Om vi i formeln ovan låter $B = H_i$ och använder lagen om total sannolikhet på $P(A)$ erhålls följande

- *Sats*: Bayes' Sats.
  Under samma villkor som i lagen om total sannolikhet gäller att

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum_{j=1}^{n} P(A|H_j)P(H_j)} \quad \text{för} \quad i = 1, \ldots, n.$$

Figur: Thomas Bayes ($\sim 1702 - 1761$) var en engelsk matematiker, statistiker och presbyteriansk präst. Han är mest känd för att ha beskrivit ett matematiskt samband som senare av Richard Price formulerades om till *Bayes sats*.

För en viss typ av statistisk slutledningsprincip är Bayes sats fundamental vilket även antyds as dess namn *bayesiansk statistik*. Denna statistiska princip, som fått enormt uppsving sedan slutet av 1900-talet i och med datorernas ökade beräkningskapacitet, beskrivs i Blom bok, avsn. 11. s. 277.

Exempel: TBC-test.

Förekomsten av TBC-smitta i en viss delbefolkning är 20%. Låt S^+ beteckna händelse att personen verkligen är smittad, dvs $P(S^+) = 0.2$ och $P(S^-) = 0.8$. Det snabbtest man kan utföra för att testa förekomst av TBC-smitta är inte perfekt. Låt

- T^+ beteckna händelsen att personen testar positivt och
- T^-, pss negativt

Givet är

- $P(\text{en smittad person ger ett positivt test}) = P(T^+|S^+) = 0.9$ (och 0.1 att det blir negativt uttslag),
- $P(\text{en icke-smittad person ger ett negativt test}) = P(T^-|S^-) = 0.7$ (och 0.3 att det blir positivt utslag).

KTH Matematik

Exempel: TBC-test (forts.)

Fråga 1:

Vad är sannolikheten att en slumpmässigt utvald person ger ett positivt test?

Svar:

med hjälp av lagen om total sannolikhet får vi

$$P(T^+) = P(T^+|S^+)P(S^+) + P(T^+|S^-)P(S^-) =$$

$$0.9 \cdot 0.2 + 0.3 \cdot 0.8 = 0.42.$$

Obs!

Betingade sannolikheter också uppfyller kriterierna för att vara sannolikhetsmått, t ex

$$P(T^-|S^+) = 1 - P(T^+|S^+) = 0.1.$$

Exempel: TBC-test (forts.)

Fråga 2:

Vad är sannolikheten att en person som testar positivt verkligen är smittad?

Svar:

med hjälp av Bayes sats får vi

$$P(S^+|T^+) = \frac{P(T^+|S^+)P(S^+)}{P(T^+|S^+)P(S^+) + P(T^+|S^-)P(S^-)} =$$

$$\frac{0.9 \cdot 0.2}{0.9 \cdot 0.2 + 0.3 \cdot 0.8} \approx 0.429.$$

Tolkning:

Det är ca 43% sannolikhet att man verkligen är sjuk om testet givit positivt utfall.

Oberoende händelser

- Om händelsen B *inte påverkar* sannolikheten för att A inträffar så får vi $P(A|B) = P(A)$ och pss $P(B|A) = P(B)$. Uttryckt med hjälp av def. av betingade sannolikheter

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

- Detta leder till definition:

  Om $P(A \cap B) = P(A)P(B)$ sägs A och B vara oberoende.

- Obs! Oförenliga händelser är ej oberoende! På tavlan i mån av tid.

KTH Matematik

Oberoende händelser (forts.)

- Om A och B är oberoende är även A och B^*, A^* och B, samt A^* och B^* oberoende:

$$P(A \cap B^*) = P(A) - P(A \cap B) = P(A) - P(A)P(B)$$
$$= P(A)(1 - P(B)) = P(A)P(B^*).$$

- *Alla, ingen och någon*. Utvidgning till fler än två händelser. På tavlan i i mån av tid.

KTH Matematik