

TMS136

Föreläsning 1

Varför?

- Om vi gör mätningar vill vi modellera och kvantifiera de osäkerheter som obönhörligen finns
- Om vi handlar med värdepapper vill vi modellera och kvantifiera de risker som finns
- Om vi forskar fram nya mediciner vill vi säkerställa deras effekter



"Data don't make any sense,
we will have to resort to statistics."

Statistik

- När vi här pratar om statistik menar vi inte NHL:s skytteliga, Premier League-tabellen eller vilket batting average en baseball-spelare har
- Vi menar snarare slutsatser om populationer baserade på stickprov
- Vi har alltså en grupp av objekt vi vill uttala oss om där vi kikar på några av objekten och baserat på vad vi finner för dessa extrapolerar vi resultaten till hela gruppen samtidigt som vi har koll på vilken osäkerhet extrapolationen har

Statistik

- I fallen med mätosäkerheter, risker vid värdepappershantering och säkerställande av medicinens effekter vill vi baserat på stickprov och/eller historiska data bilda oss uppfattningar om vilka mekanismer som styr mätningarna, värdepappren och medicinernas effekter
- Vi kommer att modellera dessa mekanismer som slumpvariabler...
- Innan vi når fram till slumpvariabler behöver vi lite sannolikheteoretiska grunder

Slumpförsök

- Med *slumpförsök* (*random experiment*) menar vi försök som upprepade gånger utförs på samma sätt men som kan få olika utfall
- Enkla exempel är slantsingling och tärningskast



Utfallsrum

- *Utfallsrum (sample space)* är mängden av alla möjliga utfall för ett slumpförsök. Vi betecknar utfallsrummet S
- Vid slantsingling är utfallsrummet

$$S = \{Krona, Klave\}$$

- Vid tärningskast är utfallsrummet

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

Sample Space for Choosing a Card from a Deck

Ace	2	3	4	5	6	7	8	9	10	Jack	Queen	King
												
Ace	2	3	4	5	6	7	8	9	10	Jack	Queen	King
												
Ace	2	3	4	5	6	7	8	9	10	Jack	Queen	King
												
Ace	2	3	4	5	6	7	8	9	10	Jack	Queen	King
												

Kontinuerliga och diskreta utfallsrum

- Vi säger att utfallsrummet är *diskret (discrete)* om vi kan skapa en lista (möjligen oändligt lång) av de möjliga utfallen
- Vi säger att utfallsrummet är *kontinuerligt (continuous)* om det innehåller ett intervall (ändligt eller oändligt) av de reella talen

Exempel

- När vi singlar slant, kastar tärning eller drar kort ur en kortlek är utfallsrummet diskret
- Om vi knäcker en 15 cm lång penna kan brottpunkten hamna var som helst mellan 0 och 15 cm (i alla fall om vi tänker oss att vi kan mäta hur noggrant som helst) och därmed är utfallsrummet kontinuerligt



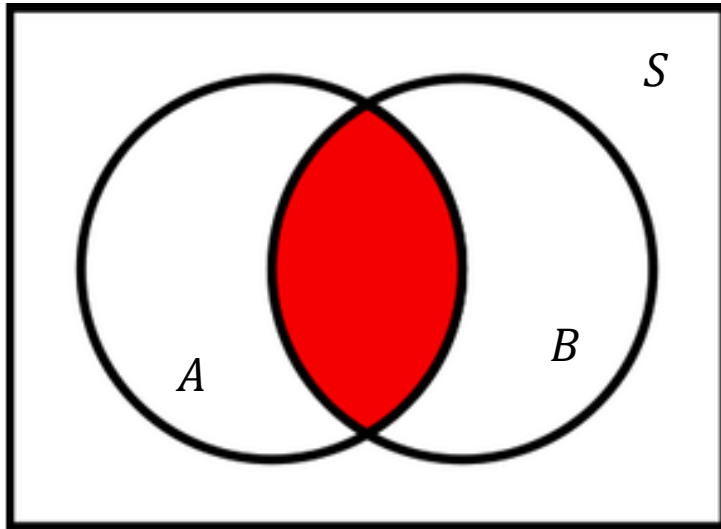
Händelse

- En *händelse (event)* är en delmängd (subset) av utfallsrummet
- Vi tärningskast har vi till exempel händelsen "udda" $\{1,3,5\}$
- När vi knäcker pennan har vi till exempel händelsen att brottpunkten hamnar mellan tre och sju cm $\{(3,7)\}$

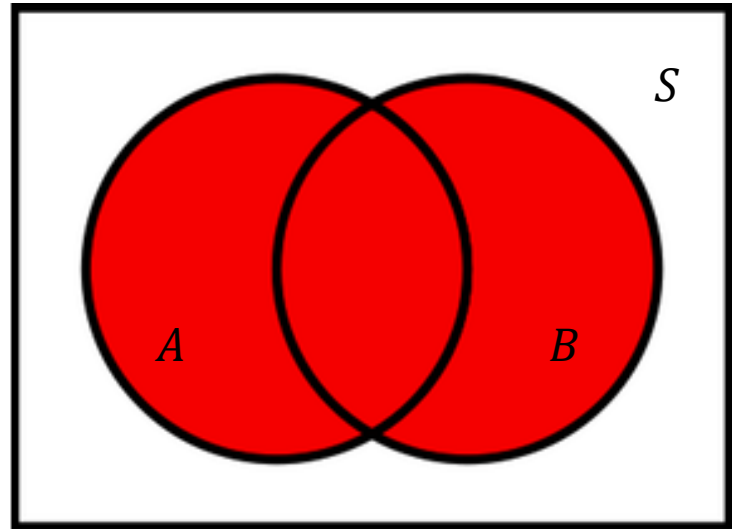
Operationer på mängder/event

- För två händelser A och B introducerar vi operationerna:
- *Union* $A \cup B$ som betyder att A eller B eller båda inträffar (när vi säger "eller" menar vi inte "antingen eller")
- *Snitt (intersection)* $A \cap B$ som betyder att både A och B inträffar

Venn-diagram



$A \cap B$ är röd



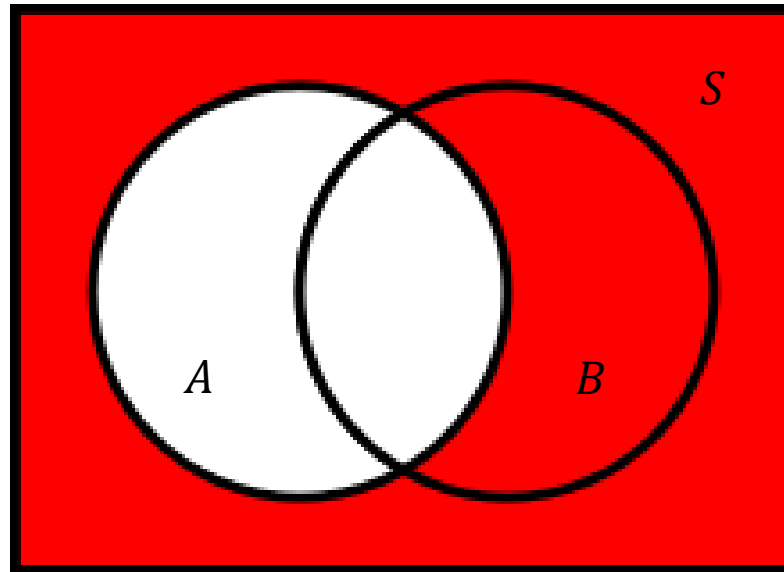
$A \cup B$ är röd

Exempel

- Vi låter A vara händelsen att vi vinner på v75 och B vara händelsen att vi vinner på stryktipset
- Då är $A \cup B$ händelsen att vi vinner på åtminstone ett av spelen
- Och $A \cap B$ är händelsen att vi vinner på både v75 och stryktipset

Komplement

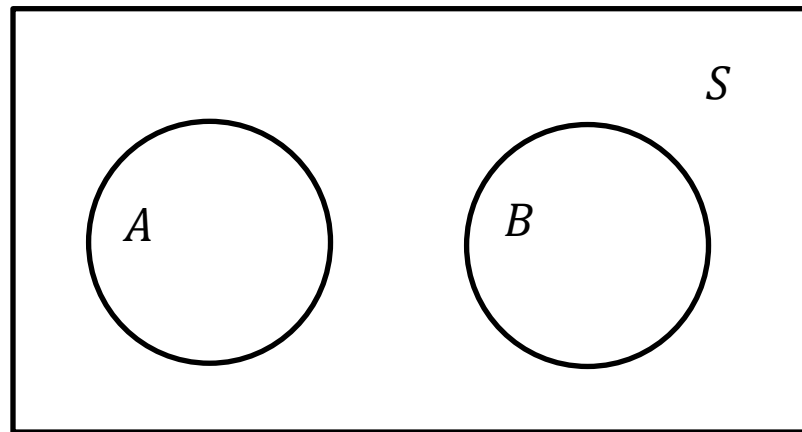
- *Komplementet (the complement) A' till en händelse A är händelsen att A inte inträffar*



A' är **röd**

Disjunkta eller ömsesidigt uteslutande händelser

- Om "snittet är tomt" skriver vi $A \cap B = \emptyset$ och säger att A och B är *disjunkta* eller *ömsesidigt uteslutande* (*disjoint* or *mutually exclusive*)



- Om så är fallet kan alltså inte A och B inträffa "samtidigt"
- Observera att $A \cap A' = \emptyset$ alltid! (Det kan inte både regna och vara uppehåll)

Multiplikationsregeln

- Om ett experiment består av k steg som oberoende av varandra kan utfalla på $n_1, \dots, n_k$ sätt kan hela experimentet få

$$n_1 \cdot \dots \cdot n_k$$

olika utfall

- Lite längre fram kommer vi göra en skarpare definition av oberoende

Exempel

- Ett experiment består av delexperimenten kasta en vanlig tärning, singla en slant och snurra ett roulettehjul
- Att kasta en tärning kan utfalla på sex olika sätt, att singla en slant kan utfalla på två olika sätt och att snurra ett roulettehjul kan utfalla på 37 olika sätt
- Eftersom delexperimenten är oberoende av varandra (utfallet av det ena påverkar inte utfallet av det andra) kan alltså hela experimentet utfalla på $6 \cdot 2 \cdot 37 = 444$ olika sätt

Att räkna...

- Den klassiska definitionen av sannolikhet säger ju att sannolikheten är "antalet gynnsamma utfall delat på antalet möjliga utfall"
- Det är så klart viktigt att kunna räkna hur många gynnsamma och möjliga utfall det finns i olika situationer...

Permutationer

- Om vi har bokstäverna a, b, c kan dessa ordnas

$abc, acb, bac, bca, cab, cba$

- Det finns alltså $6 = 3 \cdot 2 \cdot 1$ möjliga permutationer
- Vi kan tänka oss att bokstäverna ligger i en låda och vi drar en i taget utan återläggning och lägger dem på ett bord från vänster till höger
- Den första bokstaven är en av tre möjliga, den andra är en av två möjliga och till slut finns en kvar...

Permutationer, generellt

- n (olika) objekt kan utan återläggning dras i

$$n! = n(n - 1) \dots 2 \cdot 1$$

olika ordningar

- Observera att detta följer från multiplikationsregeln

Permutationer av delmängder

- Om det finns n olika objekt och vi drar r av dessa utan återläggning kan detta göras på

$$P_r^n = n(n-1) \cdots (n-r+1) = \frac{n!}{(n-r)!}$$

(olika) sätt

- Vi har 5 olika böcker i en låda och vill ställa 3 (godtyckligt valda) av dessa i bokhyllan. Det finns då $\frac{5!}{(5-3)!} = \frac{5!}{2!} = 5 \cdot 4 \cdot 3 = 60$ möjliga "bokuppställningar"

Om det finns dubletter då?

- Om vi har $n = n_1 + \dots + n_r$ objekt där n_1 är av typ 1, n_2 är av typ 2 o s v kan dragningen utan återläggning se ut på

$$\frac{n!}{n_1! \cdots n_r!}$$

olika sätt

- Bokstäverna A,B,B,A kan ordnas på $\frac{4!}{2!2!} = 6$ sätt...

Kombinationer

- Vi såg att när vi pratar om permutationer så spelar ordningen på objekten roll
- Om vi till exempel drar fem kort på måfå ur en kortlek och vill bestämma sannolikheten att vi får "färg" spelar ordningen ingen roll
- Om vi inte bryr oss om ordningen utan bara frågar oss; På hur många sätt kan man (utan återläggning) välja r av n objekt pratar vi om *kombinationer*

Kombinationer

- Om fem lag (1,2,3,4,5) deltar i en turnering där "alla ska möta alla en gång" blir det 10 matcher:

(1,2),(1,3),(1,4),(1,5),(2,3),(2,4),(2,5),(3,4),(3,5),(4,5)

- Vi kan skriva detta som

$$C_2^5 = \binom{5}{2} = \frac{5!}{2! (5-2)!} = \frac{5!}{2! 3!} = \frac{5 \cdot 4}{2 \cdot 1} = 10$$

Kombinationer, generellt

- Om vi väljer r av n olika objekt finns det

$$C_r^n = \frac{n!}{r! (n - r)!}$$

olika kombinationer

- Om vi drar 5 av 52 kort i en vanlig kortlek kan detta göras på $C_5^{52} = 2598960$ olika sätt
- Färg kan fås på $4 \cdot C_5^{13} = 5148$ olika sätt
- Sannolikheten att få färg är alltså $\frac{5148}{2598960} = 0.002$

Exempel

- I en klass finns 13 tjejer och 12 killar
- En kommitté med tre tjejer och tre killar ska utses. Hur många olika kommittéer kan man skapa?
- Multiplikationsregeln ger

$$C_3^{13} \cdot C_3^{12} = \frac{13!}{3!10!} \cdot \frac{12!}{3!9!} = \frac{13 \cdot 12 \cdot 11}{3 \cdot 2 \cdot 1} \cdot \frac{12 \cdot 11 \cdot 10}{3 \cdot 2 \cdot 1} = 62920$$

Exempel, fortsättning

- Den ena killen är syskon med den ena tjejen. På hur många sätt kan kommittén se ut om båda syskonen är med?
- Om "tvingar in" syskonen finns det kvar att välja 2 tjejer bland 12 och 2 killar bland 11
- Multiplikationsprincipen ger

$$C_2^{12} \cdot C_2^{11} = \frac{12!}{2!10!} \cdot \frac{11!}{2!9!} = \frac{12 \cdot 11}{2 \cdot 1} \cdot \frac{11 \cdot 10}{2 \cdot 1} = 3630$$

Exempel, fortsättning

- Hur många olika kommittéer kan vi skapa om inte något av syskonen inte får vara med?
- Om syskonen inte får vara med ska 3 tjejer väljas bland 12 och 3 killar väljas bland 11
- Multiplikationsprincipen ger

$$C_3^{12} \cdot C_3^{11} = \frac{12!}{3!9!} \cdot \frac{11!}{3!8!} = \frac{12 \cdot 11 \cdot 10}{3 \cdot 2 \cdot 1} \cdot \frac{11 \cdot 10 \cdot 9}{3 \cdot 2 \cdot 1} = 36300$$

Sannolikheter

- För en händelse E skriver vi sannolikheten att E inträffar som

$$P(E)$$

- För en händelse E skriver vi sannolikheten att E inte inträffar som

$$P(E')$$

- För händelserna A och B skriver vi sannolikheten att åtminstone en av dem inträffar som

$$P(A \cup B)$$

- För händelserna A och B skriver vi sannolikheten att båda inträffar som

$$P(A \cap B)$$

Sannolikhetsaxiomen

- Grundstenarna på vilka sannolikheteasteorin byggs kallas *sannolikhetsaxiomen*
- $P(S) = 1$
- $0 \leq P(A) \leq 1$
- Om $A \cap B = \emptyset$ så $P(A \cup B) = P(A) + P(B)$

Följder av axiomen

- $P(A') = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Om A är en delmängd av B så $P(A) \leq P(B)$
- Det är en bra övning att verifiera dessa samband utifrån axiomen och grafiskt m h a Venn-diagram!

Exempel

- Sannolikheten att det regnar idag är 0.6, därför är sannolikheten att det inte regnar 0.4
- Sannolikheten att tärningen visar udda eller större än 3 är $1/2 + 1/2 - 1/6 = 5/6$
- Sannolikheten att tärningen visar mindre än 3 är $1/3$ och sannolikheten att den visar mindre än 4 är $1/2$

Flera händelser

- Sambandet $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ kallas ibland *additionsregeln* och kan generaliseras till att gälla fler än två händelser
- För tre händelser A, B, C får man (Venn-diagram)
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Ömsesidigt uteslutande eller disjunkta händelser

- Om det för händelserna $E_1, \dots, E_k$ gäller att $E_i \cap E_j = \emptyset$ för alla $i \neq j$ gäller även att

$$P(E_1 \cup \dots \cup E_k) = P(E_1) + \dots + P(E_k)$$

- Speciellt har vi (obs att $\cap$ "binder hårdare" än $\cup$)

$$\begin{aligned} P(A) &= P(A \cap B \cup A \cap B') \\ &= P(A \cap B) + P(A \cap B') \end{aligned}$$

Betingad sannolikhet

- I vissa sammanhang påverkas sannolikheten för en viss händelse av en annan händelse
- För två händelser A och B ges den *betingade sannolikheten* (*conditional probability*) för " A givet B " av

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

- Vi kan tänka på detta som att vi reducerar utfallsrummet till händelsen B

Exempel

- Låt R vara händelsen att det regnar och U vara händelsen att du har ett paraply
- Sannolikheten att det regnar är 0.6 och sannolikheten att du går med ett paraply och att det regnar är 0.3. Sannolikheten att du har ett paraply givet att det regnar är

$$P(U|R) = \frac{P(U \cap R)}{P(R)} = \frac{0.3}{0.6} = 0.5$$

Bayes sats

- Bayes sats (följer ur definitionen av betingad slh)

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

- Sannolikheten att det regnar givet att du har ett paraply är alltså

$$P(R|U) = \frac{P(U|R)P(R)}{P(U)}$$

Total sannolikhet

- Observera att man med hjälp av betingad sannolikhet kan skriva

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B') \\ &= P(A|B)P(B) + P(A|B')P(B') \end{aligned}$$

- Om det för händelserna $E_1, \dots, E_k$ gäller att $\bigcup_{i=1}^k E_i = S$ och $E_i \cap E_j = \emptyset$ för alla $i \neq j$ kan vi skriva

$$\begin{aligned} P(A) &= P(A \cap E_1) + \dots + P(A \cap E_k) \\ &= P(A|E_1)P(E_1) + \dots + P(A|E_k)P(E_k) \end{aligned}$$

Exempel

- I ett visst experiment finns två urnor innehållande svarta och vita bollar. Initialt innehåller den ena urnan fem svarta och sju vita bollar och den andra urnan sex svarta och åtta vita bollar. Två bollar dras på måfå från urnan med fem svarta och sju vita bollar och de två bollarna läggs i urnan som från början innehöll sex svarta och åtta vita bollar. Därefter dras tre bollar från urnan till vilken de två bollarna flyttades. Vad är sannolikheten att alla tre bollarna är vita?

Exempel forts.

- I det första steget kan tre olika saker ske; att man flyttar två svarta bollar, att man flyttar en svart och en vit eller att man flyttar två vita, Om vi kallar dessa händelser SS, SV, VV har vi att

$$P(SS) = \frac{\binom{5}{2}}{\binom{12}{2}} = \frac{5}{33}, \quad P(SV) = \frac{\binom{5}{1}\binom{7}{1}}{\binom{12}{2}} = \frac{35}{66},$$

$$P(VV) = \frac{\binom{7}{2}}{\binom{12}{2}} = \frac{7}{22}$$

Exempel forts.

- Givet händelserna SS, SV, VV har vi alltså efter "flytten" antingen åtta svarta och åtta vita bollar eller sju svarta och nio vita eller sex svarta och tio vita i den urnan till vilken bollarna flyttades. Om vi kallar händelsen att dra tre vita (ur urnan till vilken bollarna flyttades) VVV har vi då

$$P(VVV|SS) = \frac{\binom{8}{3}}{\binom{16}{3}} = \frac{1}{10}, \quad P(VVV|SV) = \frac{\binom{9}{3}}{\binom{16}{3}} = \frac{3}{20},$$

$$P(VVV|VV) = \frac{\binom{10}{3}}{\binom{16}{3}} = \frac{3}{14}$$

Exempel forts.

- Vi får alltså

$$\begin{aligned} P(VVV) &= P(VVV \cap SS) + P(VVV \cap SV) + P(VVV \cap VV) \\ &= P(VVV|SS)P(SS) + P(VVV|SV)P(SV) \\ &\quad + P(VVV|VV)P(VV) = \\ &\quad \frac{1}{10} \frac{5}{33} + \frac{3}{20} \frac{35}{66} + \frac{3}{14} \frac{7}{22} \approx 0.16 \end{aligned}$$

Bayes sats för flera händelser

- Om det för händelserna $E_1, \dots, E_k$ gäller att $\bigcup_{i=1}^k E_i = S$ och $E_i \cap E_j = \emptyset$ för alla $i \neq j$ har vi för händelsen A med $P(A) > 0$ att

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{P(A|E_1)P(E_1) + \dots + P(A|E_k)P(E_k)}$$

Oberoende händelser

- Två händelser A och B är *oberoende* (*independent*) om

$$P(A \cap B) = P(A)P(B)$$

- Låt A vara händelsen en tärning visar 6 i första kastet och B vara det andra kastet blir en sexa.
- Då är $P(A \cap B) = 1/36$ (ett av 36 möjliga utfall) och $P(A)P(B) = \frac{1}{6} \cdot \frac{1}{6} = 1/36$

Oberoende händelser

- Definitionen $P(A \cap B) = P(A)P(B)$ är ekvivalent med (de ekvivalenta sambanden)

$$P(A|B) = P(A) \text{ och } P(B|A) = P(B)$$

- Övning: Visa att $P(A \cap B) = P(A)P(B)$ ger att $P(A' \cap B') = P(A')P(B')$

OBS!

- Man ska inte blanda ihop disjunkta händelser med oberoende händelser!
- Disjunkta betyder att $P(A \cap B) = 0$
- Oberoende betyder att $P(A \cap B) = P(A)P(B)$
- Man ser alltså att två disjunkta händelser A och B sådana att $P(A) > 0$ och $P(B) > 0$ inte kan vara oberoende.
- På samma sätt kan inte två oberoende händelser med positiva sannolikheter vara disjunkta

Slumpvariabler eller stokastiska variabler

- En *stokastisk variabel* (*random variable*) X är en funktion från utfallsrummet till de reella talen

$$X: S \rightarrow \mathbb{R}$$

- Till exempel kan vi låta X vara förtjänsten i ett spel med slansingling där du vinner \$1 om myntet visar krona och förlorar \$1 om myntet visar klave

Stokastiska variabler

- I slantsinglingsspelet har vi att

$$P(X = 1) = P(X = -1) = 0.5$$

- I slantsinglingsspelet är X *diskret* eftersom X tar värden i en diskret mängd (enstaka punkter)
- En stokastisk variabel kallas *kontinuerlig* om den ta vilket som helst värde i ett visst intervall

Exempel

- Kontinuerliga: Elektrisk strömstyrka, längd, tryck, temperatur, tid...
- Diskreta: Antal bilar på en bro vid en viss tidpunkt, andelen defekta enheter i ett parti av produkter, antalet ettor bland hundra tärningskast...
- Man bör observera att man till exempel för elektrisk strömstyrka troligen har en viss mätnoggrannhet hos sin ampere-mätare och att detta borde göra strömstyrkan diskret. Men troligen kan instrumentet visa så många olika värden att det är behändigare att betrakta strömstyrkan som kontinuerlig...