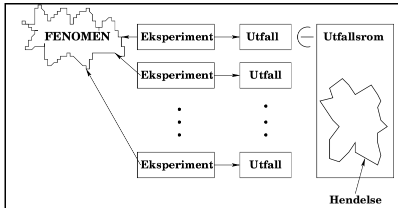
Kapittel 2: Sannsynlighet [2.6-2.8]

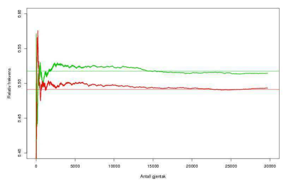
TMA4240 Statistikk (F2 og E7)

2.6, 2.7, 2.8: Betinget sannsynlighet [23.august 2004]

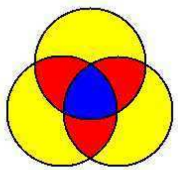
Oppsummering fra 2.1-2.5



	ordnet	ikke-ordnet
med tilbakelegg.	n^r	ikke pensum
uten tilbakelegg.	$\frac{n!}{(n-r)!}$	$\binom{n}{r} = \frac{n!}{r!(n-r)!}$



Sannsynlighet: egenskaper, gunstige vs. mulige, relativ frekvens



Sannsynlighet for mer enn en hendelse

2.6 Betinget sannsynlighet

DEF 2.9: Den betingede sannsynligheten for B, gitt A, skrives $P(B|A)$, og er definert som

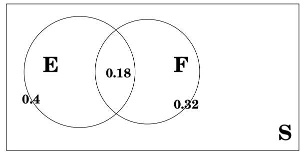
$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{hvis } P(A) > 0$$

TEO 2.13: Hvis både hendelsene A og B kan inntreffe i et eksperiment, så er

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

Biltilsynet: utslipp fra bil

- Biltilsynet har satt tall på følgende hendelser:
  - E: overstige hydrokarbongrensen, P(E)=0.32
  - F: overstige karbonoksydgrensen, P(F)=0.4
  - E∩F: overstige både hydrokarbon- og karbonoksydgrensen, P(E ∩ F)=0.18



- Velg ut en tilfeldig bil i Norge, og register utslipp

  a) Hva er sannsynligheten for at bilen du undersøker overskrider karbonoksydgrensen, gitt at du allerede vet at den overskrider hydrokarbongrensen?

  b) Hva er sannsynligheten for at bilen du undersøker overskrider hydrokarbongrensen, gitt at du allerede vet at den overskrider karbonoksydgrensen?

Uavhengige hendelser

DEF 2.10: To hendelser A og B er **uavhengige** hvis og bare hvis

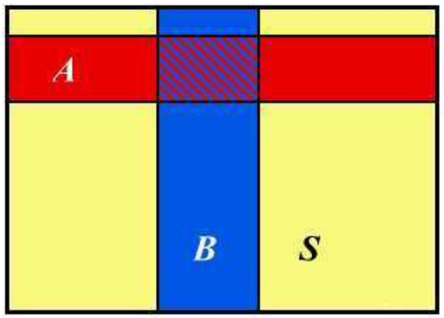
$$P(B|A) = P(B) \quad \text{eller} \quad P(A|B) = P(A)$$

Ellers, så er A og B **avhengige** hendelser.

TEO 2.14: To hendelser A og B er uavhengige hvis og bare hvis

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

Venndiagram for uavhengige hendelser



Multihendelser

TEO 2.15: Hvis hendelsene $A_1, A_2, ..., A_k$ kan inntreffe i et eksperiment, så er

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) =$$
$$P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1})$$

Hvis hendelsene $A_1, A_2, ..., A_k$ er uavhengige, så er

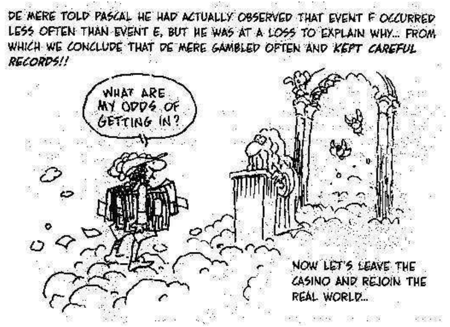
$$P(A_1 \cap A_2 \cap \cdots \cap A_k) =$$
$$P(A_1) \cdot P(A_2)P(A_3) \cdots P(A_k)$$

Tilbake til deMere

- de Mere mente (fra empiriske data) at 1) var større enn 2):
  1. minst en sekser på fire kast med en terning, eller
  2. minst en dobbel-sekser på 24 kast med to terninger?
- Ved bruk at komplementærsetningen og setning for sannsynlighet for uavhengige hendelser kan vi regne ut sannsynligheten for 1) og 2).

deMere i mål

- P(minst en sekser på fire kast med en terning)=$1-(\frac{5}{6})^4=0.5177$
- P(minst en dobbelt-sekser på 24 kast)=1-$(\frac{35}{36})^{24}$=0.4914.



Idrettsutøvere

- Idrettsutøvere kan deles inn i tre kategorier:
  - De som doper seg nå (2%)
  - De som har dopet seg tidligere (14 %)
  - De som aldri har dopet seg (84 %)
- La sannsynligheten for positiv dopingtest for de tre gruppene våre hhv. 80 %, 6 % og 3 %.

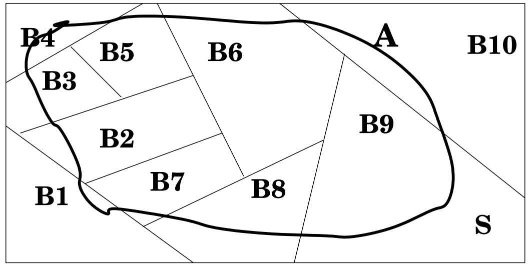
a) Hva er sannsynligheten for at en tilfeldig valgt idrettsutøver gir positiv test?

b) Hva er sannsynligheten for at en idrettsutøver som avlegger positiv dopingtest, virkelig er dopet?

Total sannsynlighet

TEO 2.16:Total sannsynlighet Hvis hendelsene $B_1, B_2, \ldots, B_k$ gir en partisjon (oppdeling) av utfallsrommet S, slik at $P(B_i) \neq 0$ for $i = 1, \ldots, k$, da har vi for hver hendelse A i S

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$



Bayes regel

TEO 2.17: Bayes regel Hvis hendelsene $B_1, B_2, \ldots, B_k$ gir en partisjon (oppdeling) av utfallsrommet S, slik at $P(B_i) \neq 0$ for $i = 1, \ldots, k$, da har vi for hver hendelse A i S hvor $P(A) \neq 0$ at

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)}$$

for $r = 1, 2, \ldots, k$.

Sykdom og test

- Defi ner to hendelser:
  - S= syk person
  - T= positiv test
- Sensitivitet: Sannsynligheten for å oppdage at en person er syk. Mer presist: Sannsynligheten for at testen slår ut positivt, gitt at personen er syk. $P(T|S)$
  - HIV-test: 98.0%
  - Mammografi : 98.0%
- Spesifi sitet: Sannsynligheten for å oppdage at en person er frisk. Mer presist: Sannsynligheten for at testen slår ut negativt, gitt at personen er frisk. $P(T'|S')$
  - HIV-test: 99.5%
  - Mammografi : 95.0%
- Interessante størrelser:
  - Sannsynligheten for at en person med positiv test virkelig er syk.
  - Sannsynligheten for at en person med negativ test virkelig er frisk.

HIV-test

- Hva er sannsynligheten for at en person med positiv HIV-test virkelig er HIV-smittet?
- Anta
  - $P(S) = 0,0005$ (Dagbladet febr 2003, 1900 smittet av HIV i Norge (av 4 000 000), dvs 0.5 promille.)
  - Sensitivitet $P(T|S)$= 0.98
  - Spesifi sitet $P(T'|S')$= 0.995

	syk	frisk	totalt
positiv test			
negativ test			
totalt			

Avhengighet av forekomst

- "Hva er sannsynligheten for at en person med positiv test virkelig er syk" er avhengig av P(S).
- Med sensitivitet $P(T|S)$= 0.98 og spesifi sitet $P(T'|S')$= 0.995.

P(S)	P(S|T)
$\frac{1}{10000}$	0.02
$\frac{1}{5000}$	0.04
$\frac{1}{1000}$	0.16
$\frac{1}{500}$	0.50
$\frac{1}{100}$	0.66

- Dette gir et problem ved masseundersøkelser. De fleste av personene med positiv prøve kan faktisk være friske.

TV-debatt

- Eksamensoppgave fra HiSør-Trøndelag 1998
- Et politisk spørsmål blir tatt opp i en TV-debatt. Et stykke ut i debatten stiller programlederen det samme spørsmålet til seerne. Vi ser heretter bare på de seerne som har en oppfatning om spørsmålet.
- De som mener *ja* oppfordres til å ringe et bestemt telefonnummer og de som mener *nei* et annet nummer.
- Vi antar i denne oppgaven at 70% av seerne mener *ja* og 30% mener *nei*.
- Vi antar videre at sannsynligheten for at en tilfeldig *ja*-seer ringer inn, er 0.05, og sannsynligheten for at en tilfeldig *nei*-seer ringer inn, er 0.10.
- La J være hendelsen at en seer mener *ja*, og R være hendelsen at han ringer.
- a) Formuler de fi re opplysningene som sannsynligheter (betingede og ubetingede) for J og R. Bestem P(R).
- b) Hvor stor andel av innringerne mener *ja*? Gir resultatet et riktig bilde av seernes oppfatning?

Tykkelse av veidekke

- Før en seksjon av en ny vei er godkjent for bruk inspiseres tykkelsen av veidekket ved hjelp av ultralyd.
- For et 8' (20 cm) dekke, gjøres dette hver $\frac{1}{10}$ mile (160 meter). Hver 160 meter seksjon vil aksepteres hvis den målte tykkelsen er minst 7.5' (19cm).
- Anta at 90% av seksjonenene følger forskriftene (de er faktisk tykkere enn 19cm)
- Men, en ultralydmåling er dessverre bare 80% sikker,
  - dvs. 80% sjanse for at målingen viser >=19cm gitt at tykkelsen i virkeligheten er >=19cm,
  - og at det er 80% sjanse for at målingen viser <19cm, gitt at tykkelsen i virkeligheten er <19cm.
- a) Hva er sannsynligheten for at en tilfeldig valgt seksjon av veien er laget etter forskriftene (dvs. >=19cm tykk) og blir godkjent for bruk (ultralydmåling >=19cm).
- b) Hva er sannsynligheten for at en tilfeldig valgt seksjon av veien ikke er laget etter forskriftene (dvs. <19cm tykk), men blir godkjent for bruk (ultralydmåling >19cm).

Tykkelse av veidekke (forts.)

- D="tykkelsen av dekket i valgt seksjon er >=19cm"
- T="ultralydmålingen er >=19cm"
- Anta at 90% av seksjonenene følger forskriftene (de er faktisk tykkere enn 19cm): dvs P(D)=0.9.
- Men, en ultralydmåling er dessverre bare 80% sikker: P(T|D)=0.8 og P(T'|D')=0.8
- c) Hva er sannsynligheten for at en seksjonen som er laget etter forskriftene (dvs. >19cm tykk) blir godkjent for bruk (ultralydmåling >19cm)?
- d) Hva er sannsynligheten for at en seksjonen som er godkjent for bruk (ultralydmåling >19cm), virkelig er laget etter forskriftene?