Statistikk: uke 2:1

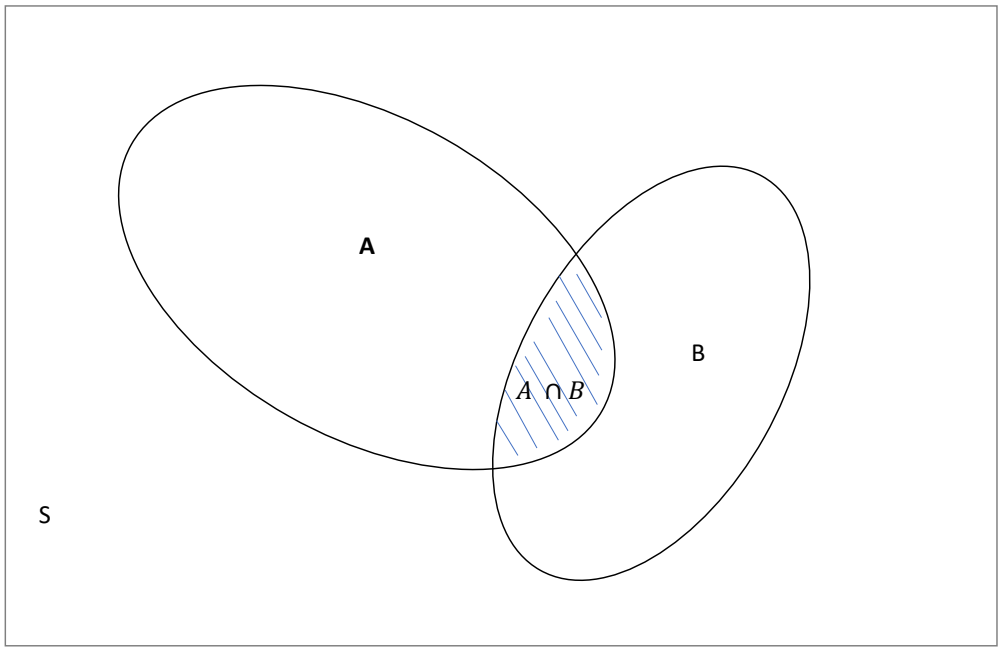
Jo Eidsvik

Matematiske fag, NTNU

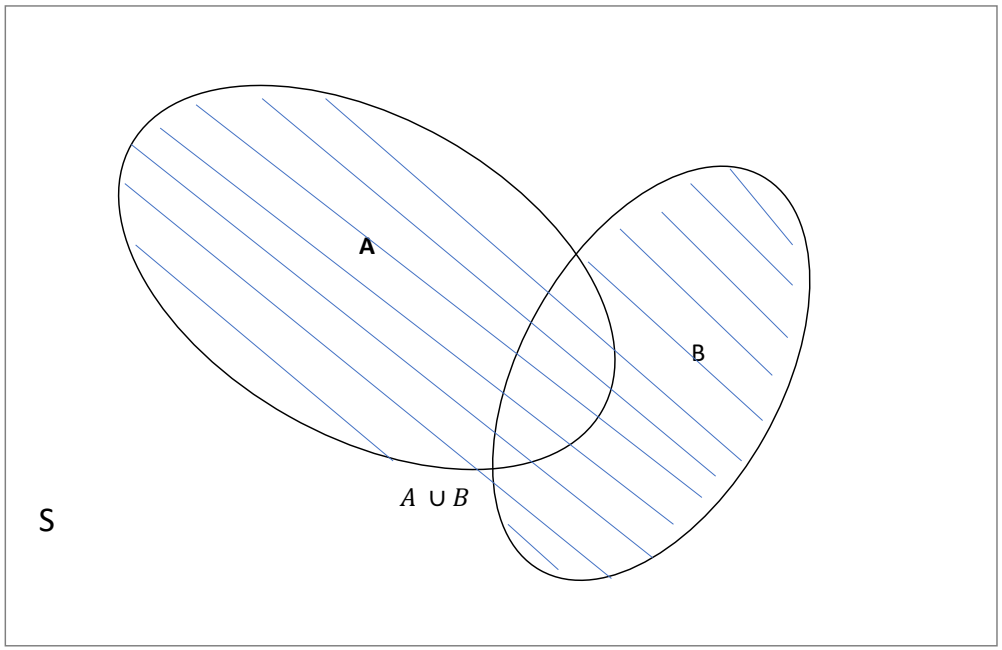
Basis for sannsynlighetsregning

- Utfallsrom S = alle mulige enkelthendelser (Sample space)
- Begivenheter eller hendelser A, B = delmengder av utfallsrommet (Events)
- Sannsynlighet for hendelse A er $P(A)$ (Probability)

Venn diagram - snitt

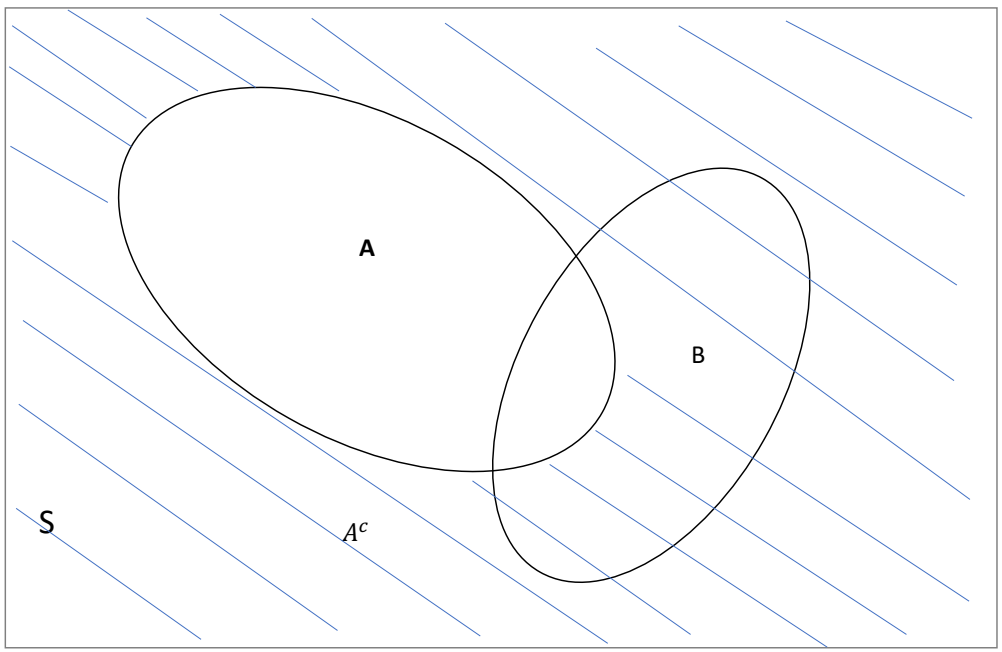


Venn diagram - union



Venn diagram - komplement

A^c er komplimentet av A, dvs alle elementer i S som ikke er i A.



Definisjon Sannsynlighet

Sannlighet kan defineres

1. $P(S) = 1$, $P(\emptyset) = 0$, der $\emptyset$ er den tomme mengde.
2. $0 \leq P(A) \leq 1$
3. $P(A \cup B) = P(A) + P(B)$ hvis A og B er disjunkte, dvs $A \cap B = \emptyset$.

Utledning av andre sannsynligheter

- $P(A^c) = 1 - P(A)$
- $P(A) = P(A \cap B) + P(A \cap B^c)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Kombinatorikk - avansert telling

- Mulige ordninger av $(1, 2, 3)$ er $3!$.
- Mulige kombinasjoner for valg av 2 blant 5 er $\binom{5}{2} = \frac{5!}{(5-2)!2!}$.

Kortspill, Lotto, $+$ $++$

Poker

Hand	Probability
One pair	0.42
Two pairs	0.048
Three-of-a-kind	0.021
Straight	0.0039
Flush	0.0020
Full house	0.0014
Four-of-a-kind	0.00024
Straight flush	0.000014
Royal flush	0.0000015

Bursdag samme dag i en klasse

k	$P(A) = P$(at least two have same birthday)
15	0.253
22	0.475
23	0.507
40	0.891
50	0.970
70	0.999

Annet spill : Craps

Craps er et terningspill der man triller 2 terninger. Summen av terningene teller. Mulig sum er 2, 3,...,12.

- Direkte seier: Kaster 7 eller 11 i første kast.
- Indirekte seier: Kaster 4,5, 6, 8, 9, 10 i første kast. Kaster deretter andre resultat, ulikt det første, inntil man får samme tall igjen (vinner), eller man får 7 (taper).

Hva er sannsynligheten for å vinne ?

P(vinne i Craps)

$P(\text{vinne i Craps}) = P(\text{direkte seier 7} \cup \text{direkte seier 11} \cup \text{vinne 4-er} \cup \text{vinne 5-er} \cup \ldots$

Disjunkte hendelser.

$P(\text{vinne i Craps}) = P(\text{direkte seier 7}) + P(\text{direkte seier 11}) + P(\text{vinne 4-er}) + \ldots$

Vi kan regne ut hver enkelt sannsynlighet.

$P(\text{vinne i Craps}) = 0.49$

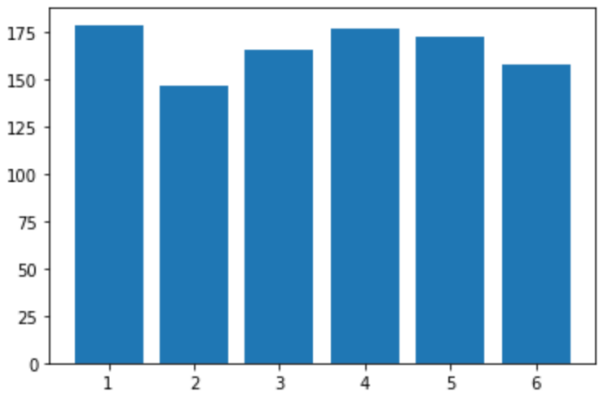
Datamaskin - simulering

Datamaskin kan brukes til å trekke variable med gitte sannsynligheter.
Da kan vi simulere forsøk og anslå sannsynligheter.
Simuler terningkast i Python:

```
import numpy as np

x=np.random.randint(1,6+1)

print(x)
```

```
6
```

```
x = np.random.randint(1,6+1,size=1000)

import matplotlib.pyplot as plt

plt.bar(*np.unique(x, return_counts=True))
```



Myntkast 300 ganger

Psykologilærer spør elever om å kaste et kronestykket 300 ganger. Og notere resultat av sekvens. (Langvarig lekse!)
Han kan vurdere om personer har jukset ved å se lengde av lengste sekvens av etterfølgende 'kron' eller 'mynt'.

Myntkast 300 ganger - Python

Vanskelig å bruke kombinatorikk til lengste sekvens.

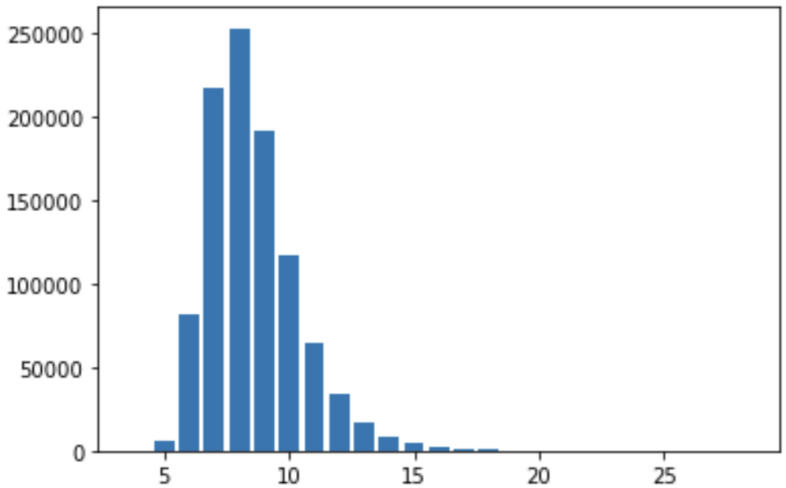
Lett å simulere mange ganger i Python.

```
import numpy as np

Len=300
x=np.random.randint(1,2+1,Len)

print(x)

lmax=0
i=0
l=1
while i<Len-1:
    if x[i+1]==x[i]:
        l=l+1
    else:
        l=1
    if l>lmax:
        lmax=l;
    i=i+1


print(lmax)
```

```
[1 1 2 1 2 1 1 1 2 2 1 1 2 2 2 1 1 2 1 1 2 2 2 2 1 1 1 2 1 1 1 1 2 2 2 2 1
 1 1 1 1 1 2 1 1 1 2 2 2 1 2 2 2 1 2 2 1 1 2 2 1 2 1 2 1 2 1 1 1 1 2 2 2 1
 2 2 2 2 2 2 2 1 2 2 1 2 2 2 1 2 2 2 1 2 1 2 1 2 2 1 1 2 2 2 2 2 2 1 2 2 2
 1 1 1 2 2 1 2 2 1 1 1 2 2 2 1 2 2 1 1 2 1 1 1 1 2 1 2 2 1 1 1 1 2 2 2 2 1
 1 2 1 2 2 2 1 1 2 1 2 2 2 2 2 2 1 2 2 1 1 2 1 1 1 1 2 2 1 2 1 2 1 1 2 2 2
 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 2 1 2 2 1 2 2 1 1 1 2 1 2 1
 1 1 2 2 2 1 1 2 1 2 2 2 1 1 2 1 1 1 2 2 1 1 1 2 1 2 2 2 1 2 1 2 1 2 2 2 2
 1 1 2 2 1 1 2 2 2 1 1 2 1 1 1 2 2 2 1 2 1 1 1 2 1 2 1 2 2 1 1 2 1 2 2 1 1
 2 1 2 1]
9
```

Myntkast 300 ganger - Resultat

Lett å simulere mange ganger i Python (1.000.000 ganger): Kun 18 ganger < 5. Usannsynlig!




```
min(lvec)
```

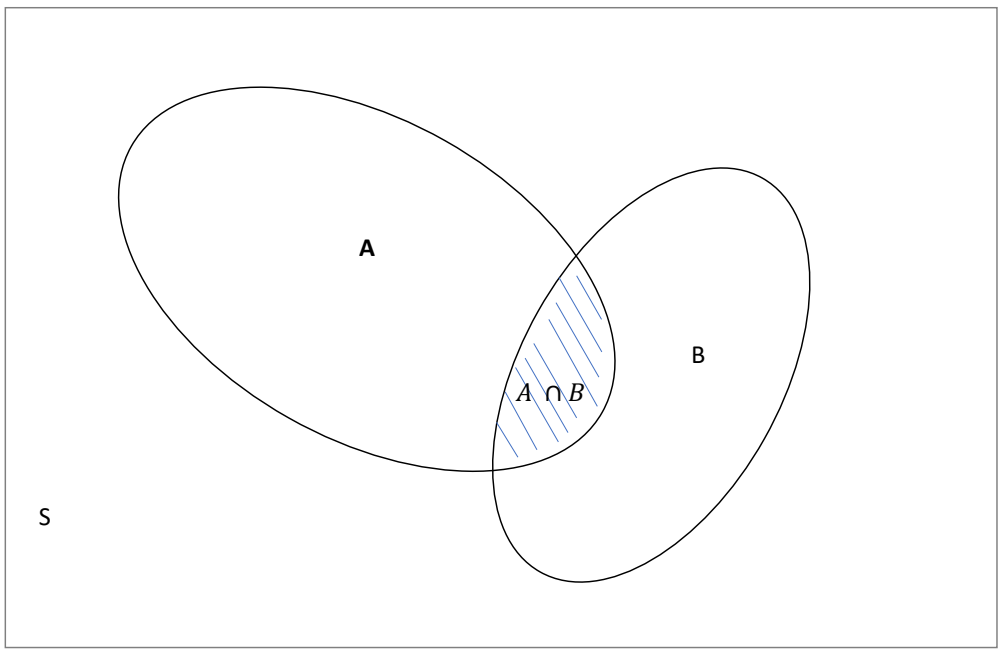
```
4
```

```
sortlvec=np.sort(lvec)
print(sortlvec[1:50])
```

```
[4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
 5 5 5 5 5 5 5 5 5 5 5 5]
```

Betinget sannsynlighet

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



Betinget - betyr det noe

			Hjemme				Borte			
Plass	Lag	Kamper	V	U	T	Mål	V	U	T	Mål
1	Bodø/Glimt	19	7	1	2	27 - 9	6	2	1	22 - 8
2	Viking	19	6	2	1	24 - 12	7	1	2	23 - 14
3	Brann	18	6	2	1	18 - 12	5	1	3	15 - 14
4	Tromsø	19	5	1	3	12 - 9	5	2	3	18 - 18
5	Rosenborg	18	5	4	0	14 - 4	3	2	4	10 - 17
6	Sandefjord Fotball	18	8	0	1	26 - 7	1	1	7	9 - 19
7	KFUM	19	4	3	3	19 - 12	3	3	3	12 - 10
8	Molde	19	5	1	4	15 - 11	3	2	4	12 - 12
9	Vålerenga	19	4	2	3	18 - 13	4	1	5	13 - 17
10	Fredrikstad	18	6	1	3	17 - 10	1	4	3	6 - 10
11	Sarpsborg 08	18	2	3	3	14 - 13	3	4	3	14 - 15
12	HamKam	19	3	3	3	11 - 15	2	3	5	10 - 16
13	Bryne	19	4	3	3	15 - 16	1	2	6	8 - 16
14	Kristiansund	18	2	3	4	9 - 14	3	2	4	11 - 19
15	Strømsgodset	19	1	0	8	7 - 21	2	1	7	16 - 23
16	Haugesund	19	1	2	7	6 - 25	0	0	9	4 - 24

P(vinn|H)	P(vinn|B)	P(vinn)
0.7000	0.6667	0.6842
0.6667	0.7000	0.6842
0.6667	0.5556	0.6111
0.5556	0.5000	0.5263
0.5556	0.3333	0.4444
0.8889	0.1111	0.5000
0.4000	0.3333	0.3684
0.5000	0.3333	0.4211
0.4444	0.4000	0.4211
0.6000	0.1250	0.3889
0.2500	0.3000	0.2778
0.3333	0.2000	0.2632
0.4000	0.1111	0.2632
0.2222	0.3333	0.2778
0.1111	0.2000	0.1579
0.1000	0	0.0526

Uavhengig?

In 1964, a woman shopping in Los Angeles had her purse snatched by a young, blond female wearing a ponytail. The thief fled on foot but was seen shortly thereafter getting into a yellow automobile driven by a black male who had a mustache and a beard. A police investigation subsequently turned up a suspect, one Janet Collins, who was blond, wore a ponytail, and associated with a black male who drove a yellow car and had a mustache. An arrest was made.

Not having any tangible evidence, and no reliable witnesses, the prosecutor sought to build his case on the unlikelihood of Ms. Collins and her companion sharing these characteristics and not being the guilty parties. First, the bits of evidence that were available were assigned probabilities. It was estimated, for example, that the probability of a female wearing a ponytail in Los Angeles was $\frac{1}{10}$. Table 2.7.1 lists the probabilities quoted for the six "facts" agreed on by the victim and the eyewitnesses (38).

TABLE 2.7.1

Characteristic	Probability
Yellow automobile	$\frac{1}{10}$
Man with a mustache	$\frac{1}{4}$
Woman with a ponytail	$\frac{1}{10}$
Woman with blond hair	$\frac{1}{3}$
Black man with beard	$\frac{1}{10}$
Interracial couple in car	$\frac{1}{1000}$