Statistikk: uke 1:2

Jo Eidsvik

Matematiske fag, NTNU

Sannsynlighet

Sannsynlighet kan defineres

1. Som tellinger, der vi ser til historien. Gunstige over mulige.
2. Som en definisjon med nyttige regneregler.

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)

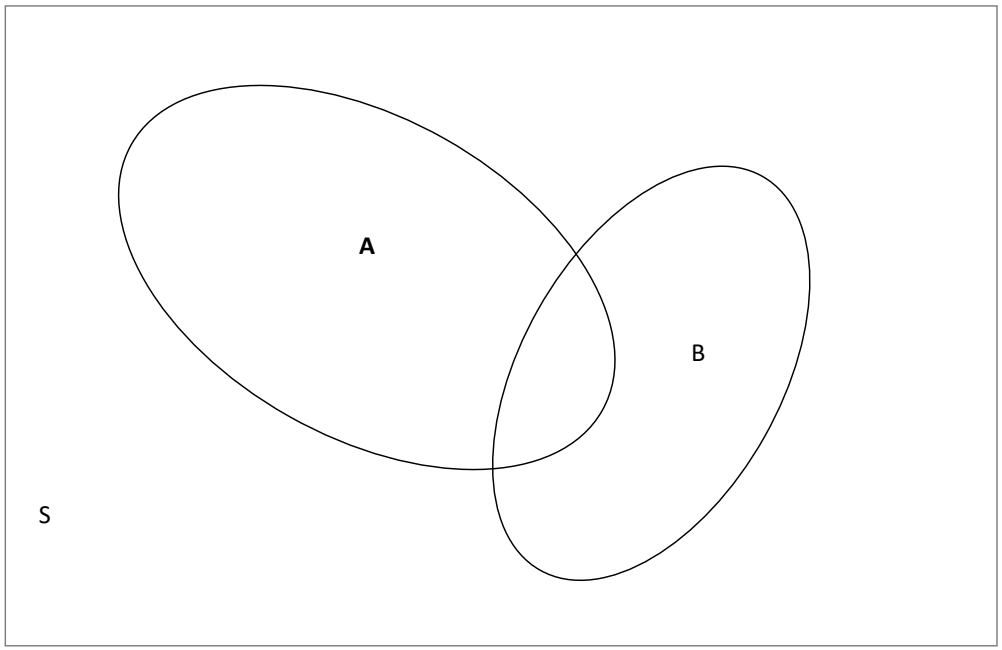
Utfallsrom

- Terningkast $S = \{1, 2, 3, 4, 5, 6\}$.
- Antall ulykker langs en veistrekning $S = \{0, 1, 2, \ldots\}$.
- Tid på 100 meter: $S = (0, \infty)$.

Hendelser

- Terningkast $A = \{5, 6\}$, $B = \{1\}$.
- Antall ulykker langs en veistrekning $A = \{0\}$, $B = \{0, 1, 2\}$.
- Tid på 100 meter: $A = (0, 10.00)$, $B = (9.90, 10.10)$.

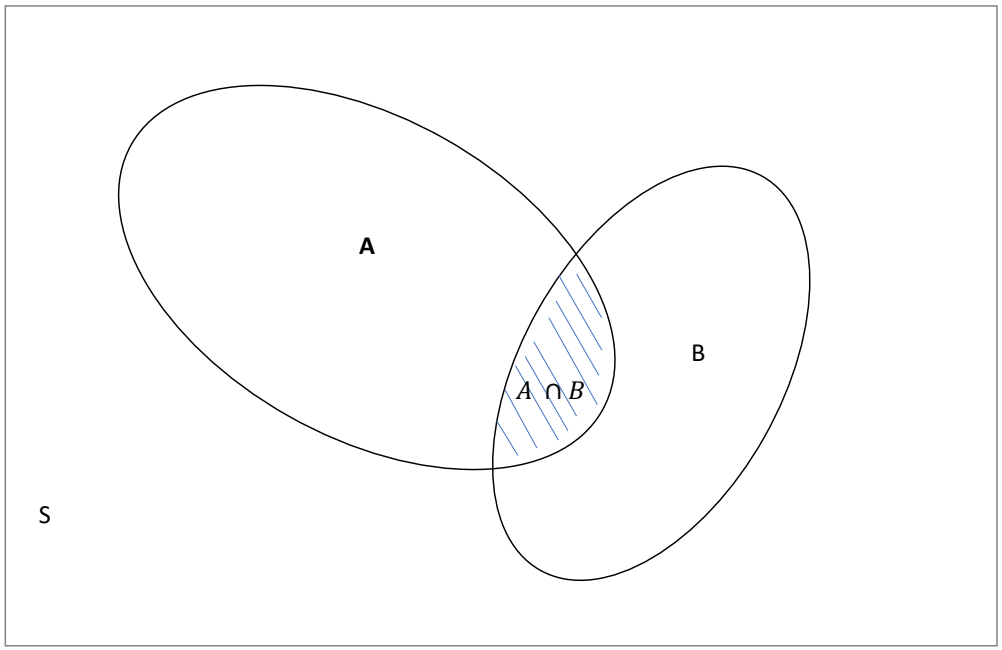
Venn diagram



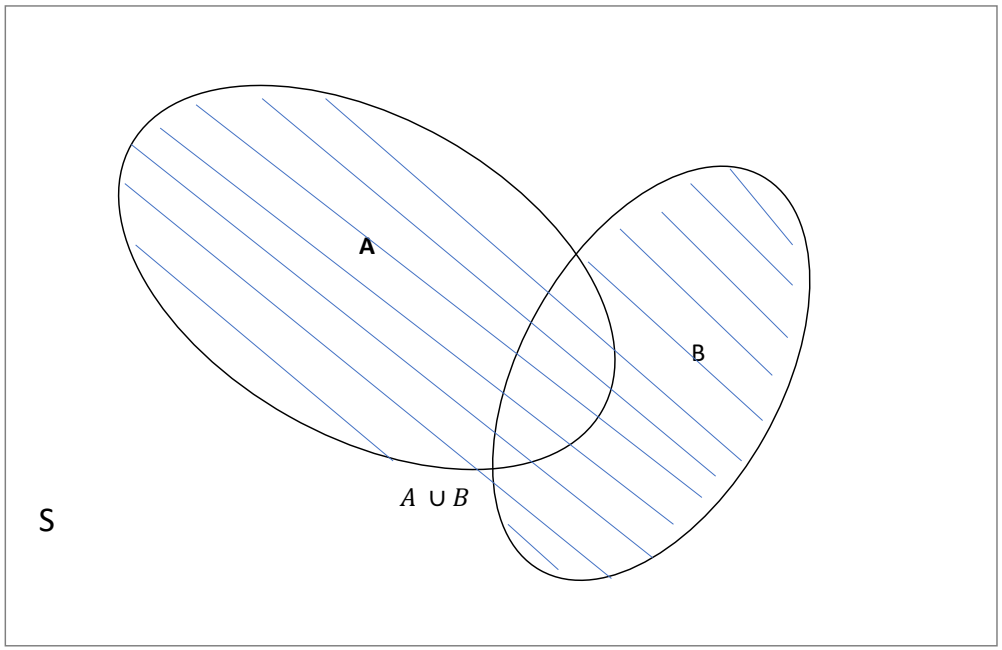
Mengder

$A \cap B$ (snitt) er enkelthendelser som er i A og i B.
$A \cup B$ (union) er enkelthendelser som er i A eller i B eller i begge.
A^c (komplement) er enkelthendelser som **ikke** er i A.

Venn diagram - snitt

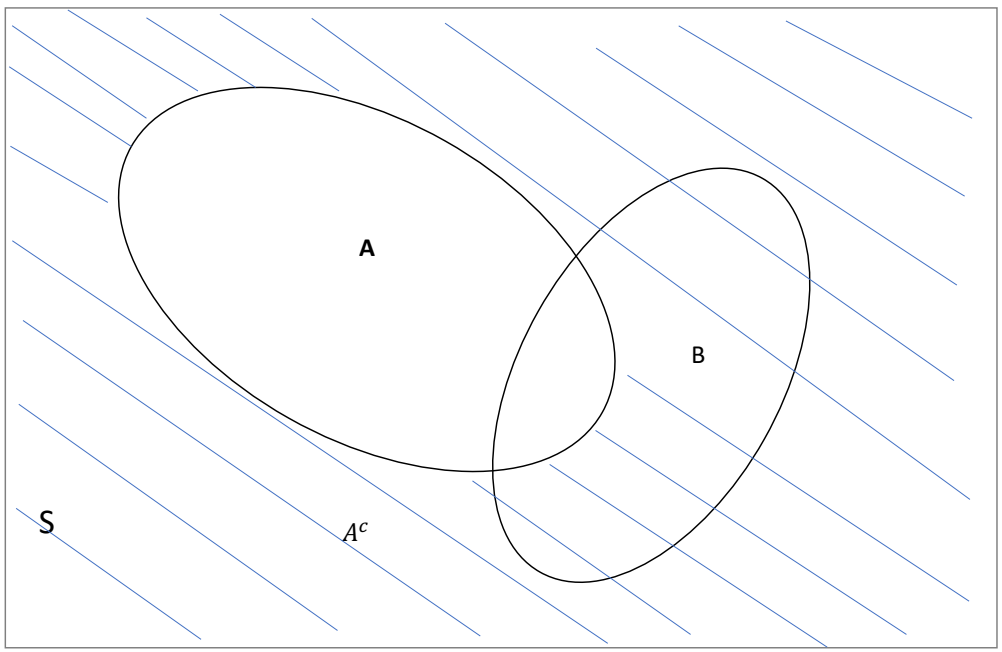


Venn diagram - union



Venn diagram - komplement

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



Snitt og union av hendelser

- Terningkast $A = \{5, 6\}$, $B = \{1\}$. $A \cap B = \emptyset$, $A \cup B = \{1, 5, 6\}$.
- Antall ulykker langs en veistrekning $A = \{0\}$, $B = \{0, 1, 2\}$.
  $A \cap B = \{0\} = A$, $A \cup B = \{0, 1, 2\} = B$.
- Tid på 100 meter: $A = (0, 10.00)$, $B = (9.90, 10.10)$.
  $A \cap B = (10, 10.10)$, $A \cup B = (0, 10.10)$.

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 $A \cap B = \emptyset$.

Sannsynlighet for snitt og union av hendelser

- Kan vi regne ut $P(A^c)$ fra $P(A)$?
- Kan vi regne ut $P(A \cup B)$ fra $P(A)$, $P(B)$ og $P(A \cap B)$?
- Kan vi regne ut $P(A \cap B^c)$ fra $P(A)$ og $P(A \cap B)$?

Utledning av andre sannsynligheter

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

Finne sannsynligheter - case

Et firma har fått klage om at de tilsetter for mange NTNU studenter og for mange menn. De har bestemt at det (i andel / sannsynlighet) skal være 25 % menn fra NTNU, 30 % utenfra NTNU og 60 % kvinner.
Hva er P(kvinne utenfor NTNU)?

Finne sannsynligheter - regning

Tolkning av tekst:
A =tilsatt er kvinne
B =tilsatt er utenfra NTNU
Vi har oppgitt: $P(A) = 0.6$, $P(B) = 0.3$, $P(A^c \cap B^c) = 0.25$.
Finner $P(A \cap B)$.

Finne sannsynligheter - regning $P(A \cap B)$

$P(B^c) = 1 - P(B) = 0.7$
$B^c = [A \cap B^c] \cup [A^c \cap B^c]$
Regel for disjunkte hendelser:
$P(B^c) = P(A \cap B^c) + P(A^c \cap B^c)$
$P(A \cap B^c) = P(B^c) - P(A^c \cap B^c) = 0.7 - 0.25 = 0.45$
Tilsvarende: $A = [A \cap B^c] \cup [A \cap B]$
Regel for disjunkte hendelser:
$P(A) = P(A \cap B^c) + P(A \cap B)$
$P(A \cap B) = P(A) - P(A \cap B^c) = 0.6 - 0.45 = 0.15$

Sannsynligheter som gunstige / mulige

Om alle enkeltutfall i utfallsrommet er like sannsynlige, sÃ¥ kan sannsynligheter regnes ut

$$P(A) = \frac{\text{gunstige enkeltutfall for } A}{\text{mulige enkeltutfall i } S}$$

Kombinatorikk - avansert telling

- Kombinatorisk vinkling: gunstige / mulige. Avansert telling av enkeltutfall gir teller og nevner.
  Kortspill, Lotto, + mange andre anvendelser
- Sannsynlighet kan også finnes empirisk ved å telle hendelser over tid (annerledes enn kombinatorikk).
  Databaser, sannsynlighet for 'guttebarn' er 0.515. Google /Youtube 's/ Amazon's metoder er basert på 'telling' av tilfeller (struktur) i data.

Vinnersjanse i Poker

- Hva er sjansen for utdelt hus?
- Hva er sjansen for utdelt et par eller to par?

Annet spill : Craps



Vinnersjanse i spillet 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, 8, 9 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 ?

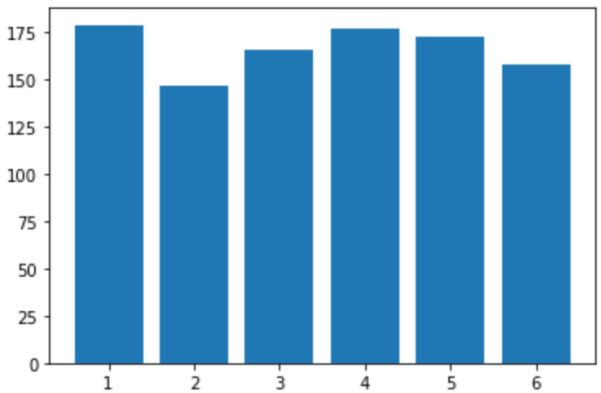
Datamaskin - simulering

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))
```



Datamaskin - bursdagsparadokset

Simuler bursdager i Python:

```python
import numpy as np

def simulate_birthday_paradox(num_people, num_trials, day_probabilities):
    """
    Simulates the birthday paradox with a non-uniform distribution of birthdays.

    Parameters:
    - num_people: Number of people in each trial
    - num_trials: Number of simulation trials
    - day_probabilities: A list or array of 365 probabilities summing to 1

    Returns:
    - Estimated probability of at least one shared birthday
    """
    assert len(day_probabilities) == 365, "Probability distribution must have 365 values"
    assert np.isclose(sum(day_probabilities), 1.0), "Probabilities must sum to 1"

    shared_count = 0

    for _ in range(num_trials):
        birthdays = np.random.choice(365, size=num_people, p=day_probabilities)
        if len(set(birthdays)) < num_people:
            shared_count += 1

    return shared_count / num_trials
```

```python
import matplotlib.pyplot as plt

# Example: Skewed distribution (e.g., more birthdays in summer months)
# Here summer months (days 150–200) have higher probability
day_probs = np.ones(365)
day_probs[150:200] *= 3  # Boost summer days p(summer)=3 times more likely
day_probs /= day_probs.sum()  # Normalize to sum to 1

plt.plot(day_probs)
plt.xlabel('Days')
plt.ylabel('Birth probability')
plt.show()

print(f"summer month (July) prob: {day_probs[175]*30:.3f}")
print(f"winter month (March) prob: {day_probs[75]*30:.3f}")
```