Tema 1: Hendelser, sannsynlighet, kombinatorikk Kapittel 2.1 - 2.7

ST1101 (Gunnar Taraldsen)

2019-01-12 17:19

Sentrale definisjoner og regneregler

Definisjoner: Stokastisk forsøk, utfallsrom, hendelser (snitt, union, komplement, disjunkte hendelser, venndiagram), sannsynlighet, betinget sannsynlighet, uavhengige hendelser, kombinatorikk, uniform sannsynlighetsmodell

wiki.math.ntnu.no/tma4245/tema/begreper/events

Regneregler sannsynlighet: addisjonssetningen, komplementærsetningen, betinget sannsynlighet, multiplikasjonssetningen, bayes regel, setningen om total sannsynlighet

wiki.math.ntnu.no/tma4245/tema/regler/probability

Regneregler kombinatorikk: telle antall mulige utfall; multiplikasjonssetningen, ordnet utvalg - trekning med tilbakelegging, ordnet utvalg - trekning uten tilbakelegging, ikke-ordnet utvalg - trekning uten tilbakelegging. Inndeling i grupper.

wiki.math.ntnu.no/tma4245/tema/regler/combinatorics

Stokastisk forsøk, utfallsrom, hendelser

Et **stokastisk forsøk** er et forsøk som:

1. Kan tenkes repetert uendelig mange ganger.
2. Har en veldefinert mengde av mulige utfall.

Et mulig utfall av det stokastiske forsøket kalles et **enkeltutfall**.

Et **utfallsrom** S er en mengde som inneholder alle mulige enkeltutfall.

Utfallsrommet kan være endelig eller uendelig.

En **hendelse** A er en delmengde av utfallsrommet, $A \subset S$, men alle delmengder er ikke nødvendigvis hendelser. Familien av hendelser inneholder den tomme mengden (den umulige hendelsen), komplementet til enhver hendelse (den motsatte hendelsen) og enhver tellbar union av hendelser (det at minst en av hendelsene inntreffer). Familien av hendelser er dermed en σ-algebra utstyrt med operasjonene union (eller), snitt (og) og komplement (det motsatte).

Operasjoner på hendelser

La $A \subset S$ og $B \subset S$ være to hendelser.

Snittet av A og B er hendelsen som inneholder alle utfall som er i både A og B .

$$A \cap B$$

A og B er **disjunkte hendelser** hvis de ikke har noen utfall til felles.

$$A \cap B = \emptyset$$

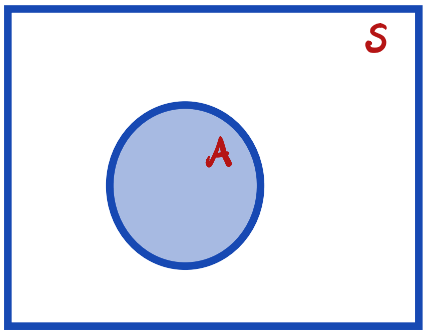
Unionen av A og B er hendelsen som inneholder alle utfall som er i A eller B eller både A og B.

$$A \cup B$$

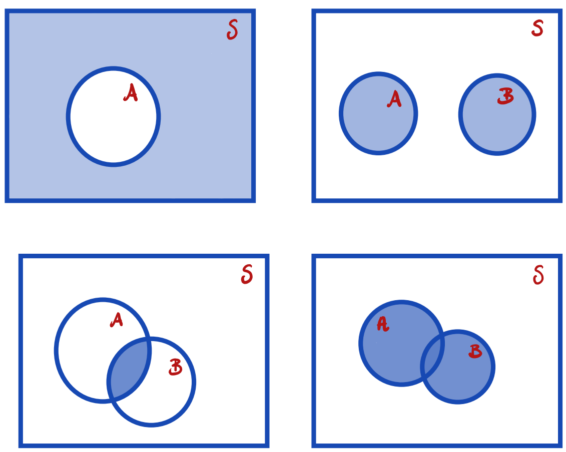
Komplementet til en hendelse A er hendelsen som inneholder alle utfall i S som ikke er i A.

$$A^c \subset S$$

Venndiagram



Hva viser venndiagrammene?



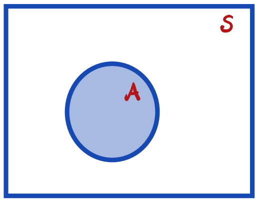
Hendelsen A^c, union av disjunkte hendelser A og B, snitt av hendelser A og B, union av hendelser A og B.

Sannsynlighet

La S være et ufallsrom, og la $A \subset S$ og $B \subset S$ være hendelser.

- Sannsynligheten for at hendelsen A inntreffer: $P(A)$
- Sannsynligheten for at hendelsen A *ikke* inntreffer: $P(A^c)$
- Sannsynligheten for at både A og B inntreffer: $P(A \cap B)$
- Sannsynligheten for at A eller B eller begge inntreffer: $P(A \cup B)$

$P(A)$ = Areal av hendelsen A i Venndiagrammet



Hva er egentlig funksjonen P?

Sannsynlighetsmål P

Et sannsynlighetsmål P på utfallsrommet S er en funksjon som oppfyller Kolmogorovs aksiomer:

1. $P(A) \geq 0$ for enhver hendelse A.
2. $P(\emptyset) = 0$.
3. Dersom $A_1, A_2, \ldots$ er parvis disjunkte hendelser ($A_i \cap A_j = \emptyset, \forall i \neq j$), så er

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \quad (\textbf{tellbar additivitet})$$

4. $P(S) = 1$

Merk: Kolmogorovs aksiomer er *nødvendige* og *tilstrekkelige* for at funksjonen P skal være et sannsynlighetsmål på utfallsrommet S. Fra disse aksiomene kan vi utlede en rekke regneregler og teoremer for P. Dersom kravet $P(S) = 1$ utelates, så er P et mål. Mål kan brukes til å måle masse, volum, areal, lengde og mye mye mer og i disse tilfellene må $P(S) = \infty$ kunne tillates.

Teorem 2.3.1 - 2.3.6

Teorem 2.3.1 - Komplementærsetningen: $P(A^c) = 1 - P(A)$

Teorem 2.3.2: $P(\emptyset) = 0$

Teorem 2.3.3: Hvis $A \subset B$, så er $P(A) \leq P(B)$

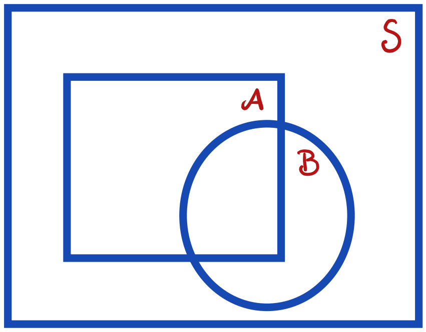
Teorem 2.3.4: For enhver hendelse A, så er $P(A) \leq 1$

Teorem 2.3.5: Hvis $A_1, A_2, \ldots, A_n$ er parvis disjunkte hendelser, så er $P(A_1 + \cdots + A_n) = P(A_1) + \cdots + P(A_n)$

Teorem 2.3.6 - Addisjonssetningen: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Betinget sannsynlighet

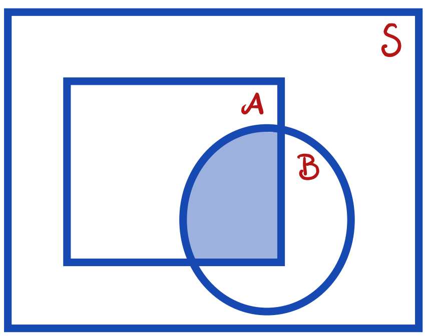
La $A \subset S$ og $B \subset S$ være to hendelser i et utfallsrom S.



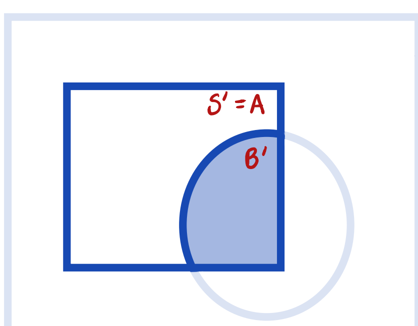
Anta at vi vet at A har skjedd. Hva er da sannsynligheten for at B intreffer?

$$P(B \mid A)$$

$P(A \cap B)$ er sannsynligheten for at utfall som både er i A og B inntreffer, men $P(A \cap B)$ tar ikke hensyn til at vi vet at A har skjedd.



Siden A har skjedd, så minsker det mulige utfallsrommet fra S til $S' = A$. Den betingede sannsynligheten for B gitt A er da sannsynligheten for hendelsen $B' = A \cap B$ der B' er en hendelse i S'.



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

Multiplikasjonssetningen

Betinget sannsynlighet:

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

Multiplikasjonssetningen:

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

og

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

Uavhengige hendelser

La A og B være to hendelser i S. Hendelsene er **uavhengige** dersom

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

For B slik at $P(B) > 0$, medfører dette at

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

Informasjonen 'B har skjedd' endrer ikke sannsynligheten for at A inntreffer.

Hvis vi vet at hendelser er uavhengige er det enkelt å regne ut sannsynligheten for snittet.

Hvis $P(A|B) \neq P(A)$ så er A og B avhengige.

Observerer at $P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B)$ når det antas at $P(A) > 0$ og $P(B) > 0$.

Snittet av mer enn to hendelser

For hendelsene A, B og C er

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

Merk: Rekkefølge er ikke viktig,

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

Multiplikasjonssetningen for hendelser $A_1, A_2, \ldots, A_n$:

$$\begin{aligned} P(A_1 \cap A_2 \cap \ldots \cap A_n) = & P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \\ & \cdot P(A_{n-1}|A_1 \cap A_2 \cap \ldots \cap A_{n-2}) \cdot \\ & \cdots P(A_2|A_1)P(A_1) \end{aligned}$$

Uavhengige hendelser

La A og B være to hendelser i S. Hendelsene er uavhengige dersom

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

Hendelsene A, B og C og er uavhengige dersom

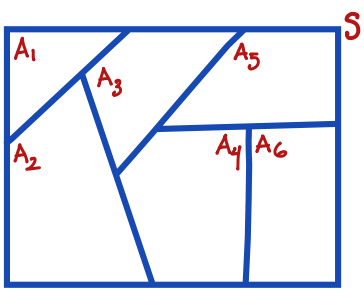
1. $P(A \cap B \cap C) = P(A)P(B)P(C)$
2. $P(A \cap B) = P(A)P(B)$
3. $P(A \cap C) = P(A)P(C)$
4. $P(B \cap C) = P(B)P(C)$

Hendelsene $A_1, A_2, \ldots, A_n$ er uavhengige dersom

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}),$$

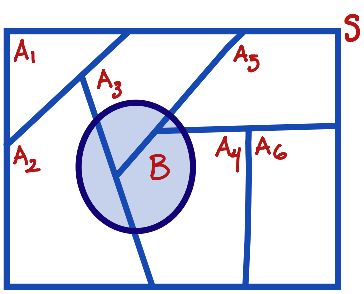
for ethvert sett av indekser $i_1, i_2, \ldots, i_k$ fra $\{1, \ldots, n\}$

Partisjon av utfallsrommet



La $A_1, A_2, \ldots, A_n$ være parvis disjunkte hendelser i utfallsrommet S ($A_i \cap A_j = \emptyset \ \forall i \neq j$) slik at $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n = S$.

Vi sier at $A_1, A_2, \ldots, A_n$ er en **partisjon av utfallsrommet** S.



La $A_1, A_2, \ldots, A_n$ være en partisjon av utfallsrommet S og la B være en hendelse i S.

Setningen om total sannsynlighet:

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Bayes regel:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

Eksempel fra genetikk (eksamensoppgave 2010)

For et spesielt gen finnes det to genvarianter a og A.

Mulige genotyper for en person er da aa, aA og AA.

Genvariant a er sjelden; 0.01% av alle personer i en populasjon har genotype aa, 1.98% har genotype aA og 98.01% har genotype AA.

Genvariant a disponenerer for en spesifikk sykdom. Sannsynligheten for at sykdommen kommer til utrykk blant personer med henholdvis genotype aa, aA og AA er 0.6, 0.02 og 0.01.

a) Hva er sannsynligheten for at en tilfeldig valgt person har sykdommen?

b) Hva er sannsynligheten for at en person har genotype aa, gitt at personen har sykdommen? Hva er sannsynligheten for at en person har genotype AA, gitt at personen har sykdommen?

Repetisjon og motivasjon

Et sannsynlighetsmål er en funksjon P som virker på hendelser i S og som oppfyller Kolmogorovs tre aksiomer.

En rekke egenskaper / regneregler for som må gjelde for funksjonen P kan utledes fra disse aksiomene.

Hvordan setter vi tallverdier for P? Matematisk må P oppfylle Kolmogorovs tre aksiomer, men P må også reflektere 'den virkelige verden'.

I statistikk knytter vi den matematiske og den virkelige verden sammen ved å anvende ulike sannsynlighetsmodeller.

Uniform sannsynlighetsmodell

I noen situasjoner ('stokastiske forsøk') er det rimelig å anta at alle enkeltutfall er like sannsynlige.

Eksempel: Terningspill, kortspill

Vi har en uniform sannsynlighetsmodell dersom

- Antall elementer i utfallsrommet er endelig $S = \{e_1, e_2, \ldots, e_m\}$
- Sannsynligheten for hvert enkeltutfall ($\{e_i\}, i = 1, \ldots, m$) er lik

$$P(\{e_1\}) = P(\{e_2\}) = \cdots P(\{e_m\}) = \frac{1}{m}$$

For enhver hendelse $A \subset S$ så er

$$P(A) = \frac{g}{m} = \frac{\text{antall gunstige enkeltutfall}}{\text{antall mulige enkeltutfall}}$$

der et 'gunstig' enkeltufall er et enkeltutfall e slik at $e \in A$, og et 'mulig' enkeltutfall er et enkeltutfall e slik at $e \in S$.

Kombinatorikk

Telleregler for å finne antall mulige og antall gunstige utfall.

Vi skiller mellom hendelser der rekkefølge er viktig eller irrelevant.

Multiplikasjonsregelen

Anta at k operasjoner skal utføres. Operasjon 1 kan utføres på n_1 måter, operasjon 2 kan utføres på n_2 måter, og så videre. Antall mulige sekvenser av k operasjoner er da

$$m = n_1 \cdot n_2 \cdots n_k$$

Observer: Dersom alle k operasjoner kan utføres på n måter så vil antall mulige sekvenser være $m = n^k$.

Telle permutasjoner

Hvis vi har n distinkte objekter kan vi danne $n!$ ulike permutasjoner (rekkefølger av lengde n)

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$$

Hvis vi har n distinkte objekter og vil danne permutasjoner av lengde r, så finnes det

$$n \cdot (n-1) \cdot (n-2) \cdots (n-r+1) = \frac{n!}{(n-r)!} = {}_nP_r$$

mulige permutasjoner.

$${}_nP_n = \frac{n!}{0!} = \frac{n!}{1} = n!$$

Telle kombinasjoner

Permutasjoner av A, B, C: ABC, ACB, BAC, BCA, CAB og CBA.

Hvis rekkefølge på bokstavene er irrelevant, er alle 6 permutasjoner like. Alle ordene er kombinasjoner av de tre bokstavene A, B og C.

Hvis vi har n distinkte objekter og vil danne *kombinasjoner* av lengde r, så finnes det

$$\frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!} = \binom{n}{r} = {}_nC_r$$

mulige kombinasjoner.

Urnemodell

Anta en samling av n distinkte kuler.

Det stokastiske forsøket går ut på å trekke r kuler fra urnen.

1. Trekning med tilbakelegging, ordnet utvalg:
   - n^r mulige trekninger
2. Trekning uten tilbakelegging, ordnet utvalg:
   - ${}_nP_r$ mulige trekninger (permuteringer)
3. Trekning uten tilbakelegging, ikke-ordnet utvalg:
   - ${}_nC_r = \binom{n}{r}$ mulige trekninger (kombinasjoner)

Urnemodell 1: Trekning med tilbakelegging, ordnet utvalg

Antall mulige utfall/rekkefølger finner vi ved multiplikasjonsregelen - hver trekning er en operasjon som kan utføres på n ulike måter:

$$n^r$$

Urnemodell 2: Trekning uten tilbakelegging, ordnet utvalg

Antall mulige utfall følger av multiplikasjonsregelen - trekning 1 er en operasjon som kan utføres på n ulike måter, trekning 2 er en operasjon som kan utføres på $n-1$ ulike måter, osv.

$$n \cdot (n-1) \cdot (n-2) \cdots (n-r+1) = \frac{n!}{(n-r)!} = {}_nP_r$$

Urnemodell 3: Trekning uten tilbakelegging, ikke-ordnet utvalg

Se på en spesiell trekning (uten tilbakelegging). Denne sekvensen av objekter kan stokkes om på $r!$ mulige måter. Siden rekkefølge er irrelevant (ikke-ordnet utvalg) er antall kombinasjoner av r objekter gitt ved

$$\frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!} = \binom{n}{r} = {}_nC_r$$

Eksempel 1: Sprintfinale langrenn

I en sprint finale i langrenn er det 6 finalister

- Johannes
- Federico
- Dario
- Emil
- Finn
- Sergej

1. Hvor mange ulike 'paller' (gull, sølv og bronse) kan (teoretisk sett) oppstå?
2. Anslå sannsynligheten for at Johannes vinner sprintfinalen.

Eksempel 2: VM håndball 2017

I gruppespillet i VM i håndball 2017 var det 6 lag i hver gruppe. Gruppe B bestod av Argentina, Norge, Polen, Sverige, Tsjekkia og Ungarn. Etter at alle hadde spilt mot alle gikk de fire lagene med høyest poengsum videre til delfinaler.

1. Hvor mange kamper ble spilt i gruppe B?
2. Hvor mange kombinasjoner av fire lag kunne (teoretisk sett) gå videre til delfinaler?
3. Anslå sannsynligheten for at Norge går videre til delfinale.

Eksempel 3: Pokerhender - 3 like

Farge: Ruter, hjerter, spar, kløver

Rangering ('tall'): Ess (A), 2, 3, ..., 10, Knekt (J), Dame (D), Konge (K)

Tre like: Tre kort av samme rangering og to 'single'

Anta at en kortstokk med 52 kort stokkes godt, og 5 kort deles ut til deg.

Hva er sannsynligheten for å få 3 like?

Urnemodell med grupper

En urne inneholder n kuler av k distinkte typer, n_1 av type 1, n_2 av type 2, etc. Antall permuteringer av n kuler er da

$$\frac{n!}{n_1!n_2!\cdots n_k!}$$

Eksempel:

Urne med 5 hvite, 4 sorte, 3 røde kuler.

Hva er sannsynligheten for å trekke sekvensen

hvit - rød - hvit - sort?

Diskret sannsynlighetsmodell

Anta at utfallsrommet S er endelig eller tellbart uendelig.

Definer funksjonen p på enkeltutfall i S slik at

1. For alle enkeltutfall $e \in S$: $p(e) \geq 0$
2. $\sum_{e \in S} p(e) = 1$

Hvis vi definerer P slik at

$$P(A) = \sum_{e \in A} p(e)$$

så er P et *sannsynlighetsmål* (oppfyller Kolmogorovs aksiomer).

Uniform sannynlighetsmodell

$$p(e) = \frac{1}{m}, \quad P(A) = \frac{g}{m} = \frac{\text{antall gunstige enkeltutfall}}{\text{antall mulige enkeltutfall}}$$

er et spesialtilfelle for endelige utfallsrom $S = \{e_1, e_2, \ldots, e_m\}$.

Diskret sannsynlighetsmodell: Eksempel

En interesseorganisasjon for lungesyke regner med at 7% av alle voksene lider av en eller annen lungesykdom. Organisasjonen regner dessuten med at 90% av de som lider av en lungesykdom røyker, mens 25% av de som ikke har noen lungesykdom røyker.

Vi tenker oss at vi trekker ut en tilfeldig voksen person. L er hendelsen 'personen har en lungesykdom', og R er hendelsen 'personen røyker'.

Enkeltutfall:

- e_1: (røyker, har lungesykdom)
- e_2: (ikke-røyker, har lungesykdom)
- e_3: (røyker, har ikke lungesykdom)
- e_4: (ikke-røyker, har ikke lungesykdom)

$S = \{e_1, e_2, e_3, e_4\}$.