

1 Stokastiske Variable

For at forstå Stokastiske Variable kræves der en forståelse for nogle simple begreber inden for statistik. Disse begreber er udfaldsrum, delmængde, sandsynlighedsfunktion samt sandsynlighedsfelt. De nævnte begreber er essentielle inden for statistik.

1.1 Udfaldsrum og delmængder

Der findes to typer af udfaldsrum, henholdsvis et diskret og kontinuert udfaldsrum. Forskellen på disse to er at der i et diskret udfaldsrum kun er et givent antal mulige udfald. Et eksempel på et diskret udfaldsrum kunne være en sekssidet terning. Denne sekssidet terning vil have et udfaldsrum U der indeholder elementerne 1, 2, 3, 4, 5, 6, 7 dette kan også skrives som $U = \{1, 2, 3, 4, 5, 6\}$. Altså vi har tale om et udfaldsrum med længden seks, hvilket skrives $|U| = 6$. Hvorimod et kontinuert udfaldsrum kunne være højden på elever i en klasse, udfaldsrummet kunne være defineret som $U = [150cm; 210cm]$. Udfaldsrummet er kontinuert da elevens højde kan variere med et infinitesimal, altså elev A kunne være en uendeligt lille mængde højere end elev B . I sådan et udfaldsrum findes der ikke nogen længde af udfaldsrummet, altså $|U| \notin \mathbb{R}$ [SC].

En delmængde er en mængde af udfaldsrummet, det beskrives ofte med A eller B . Hvis vi tager udgangspunkt i den sekssidet terning igen, kunne en delmængde af dets udfaldsrum være $A = \{1, 3\}$. Altså en delmængde indeholder altså en eller flere elementer fra udfaldsrummet.

Ydermere vil et bestemt element i et udfaldsrum eller delmængde denoteres u_i eller a_i , altså for henholdsvis A og U . Det vil sige hvis vi har en mængde Q vil et bestemt element i Q denoteres som q_i . Desuden findes der en række notationer der beskriver delmængders relationer til hinanden og udfaldsrummet.

1. $A \cup B$: udfaldet ligger i enten A eller B, evt. i både A og B
2. $A \cap B$: udfaldet ligger i både A og B
3. $A \setminus B$: udfaldet ligger i A og ikke B
4. A^c : udfaldet ligger ikke i A (dette kan også denoteres \bar{A})

1.2 Sandsynlighedsfunktion

Sandsynlighedsfunktion denoteres P og beskriver sandsynligheden for et element i et udfaldsrummet U . For eksempel kunne $P(\{1, 2\}) = 0.8$ og $P(3) = 0.2$ hvor udfaldsrummet $U = \{1, 2, 3\}$. Dette betyder at sandsynligheden for 1 eller 2 er 80% hvor i mod sandsynligheden for 3 er 20%. Desuden kan vi bruge additions loven til at udlede den totale

sandsynlighed for udfaldsrummet $[\mathbf{SC}]$.

$$P(1) + P(2) + P(3) = 1 \quad (1.1)$$

$$P(\{1, 2\}) + P(3) = 1 \quad (1.2)$$

Desuden kan $\{1, 2\}$ også beskrives som en delmængde af U på formen $A = \{1, 2\}$, altså så det desuden er gældende at $P(A) = P(1) + P(2)$

1.3 Sandsynlighedsfelt

Et sandsynlighedsfelt findes på to former, det endelige eller også kaldet det diskrete sandsynlighedsfelt samt det kontinuere sandsynlighedsfelt. Det diskrete sandsynlighedsfelt kan denoteres (U, P) og er bestående af et udfaldsrum U og en sandsynlighedsfunktion P . Hvis der er tale om et symmetrisk sandsynlighedsfelt, vil følgende være gældende for P .

$$P(U) = 1 \quad (1.3)$$

$$P(u_1) + P(u_i) + \dots + P(u_n) = 1 \quad (1.4)$$

$$P(u_1) = P(u_i) = \dots = P(u_n) = \frac{1}{|U|} \quad (1.5)$$

De to udsagn udtrykker derved at sandsynligheden for $P(U_i)$ er i intervallet $[0; 1]$ samt at den samlede sandsynlighed for feltet er 1, hvor $1 = 100\%$. I et ikke symmetrisk sandsynlighedsfelt vil de to første regler af de overstående tre gælde, men sandsynligheden for de enkelte elementer i udfaldsrummet er ikke nødvendigvis lig.

1.4 Stokastiske variable

Stokastiske variable findes af to type der begge denoteres med X , disse to type er henholdsvis diskret og kontinuert. En stokastiske variable er en variable som kan tage alle værdier i udfaldsrummet, med sandsynligheden P . Altså $P(X = u_i) = P(u_i)$. Forskellen mellem diskret og kontinuere stokastiske variable er deres udfaldsrum og der tilhørende regne regler. Et eksempel på en stokastiske variable kunne være europæisk roulette, dette har et symmetrisk sandsynlighedsfelt med et udfaldsrum $U = \{0, 1, \dots, 36\}$ der har indeholder $37 = |U|$ udfald. Det vil sige at den kugle man smider ned i spillet kan beskrives som en stokastiske variable X og den har en lige stor sandsynlighed for at tage et udfald i udfaldsrummet. Altså $P(u_1) = P(u_i) = \dots = P(u_n)$, det betyder at $P(X = 1) = P(X = 4)$. Yder mere er roulettes udfaldsrum opdelt i tre farver grøn, rød og sort, lad disse værende betegnet som delmængderne henholdsvis G , R og S . Vi kan således beskrive sandsynligheden for at den stokastiske variable X vil ligge i en delmængde soledes $P(X \in S) = P(X \in R) = \frac{18}{37}$ og $P(X \in G) = \frac{1}{37}$. Delmængden G kunne også have været defineret som $G = (R \cup S)^c$, altså de udfalds som ikke falder ind under hverken R eller S , eller $G = 0$