

1 Stokastiske Variable

En væsentlig del af Monte Carlo Metodens teori og praksis bygger på Stokastiske Variable. For at forstå Stokastiske Variable kræves der en forståelse for nogle simple begreber indenfor statistik. Disse begreber er udfaldsrum, delmængde, sandsynlighedsfunktion samt sandsynlighedsfelt. De nævnte begreber er også essentielle inden for statistik generelt.

1.1 Udfaldsrum og delmængder

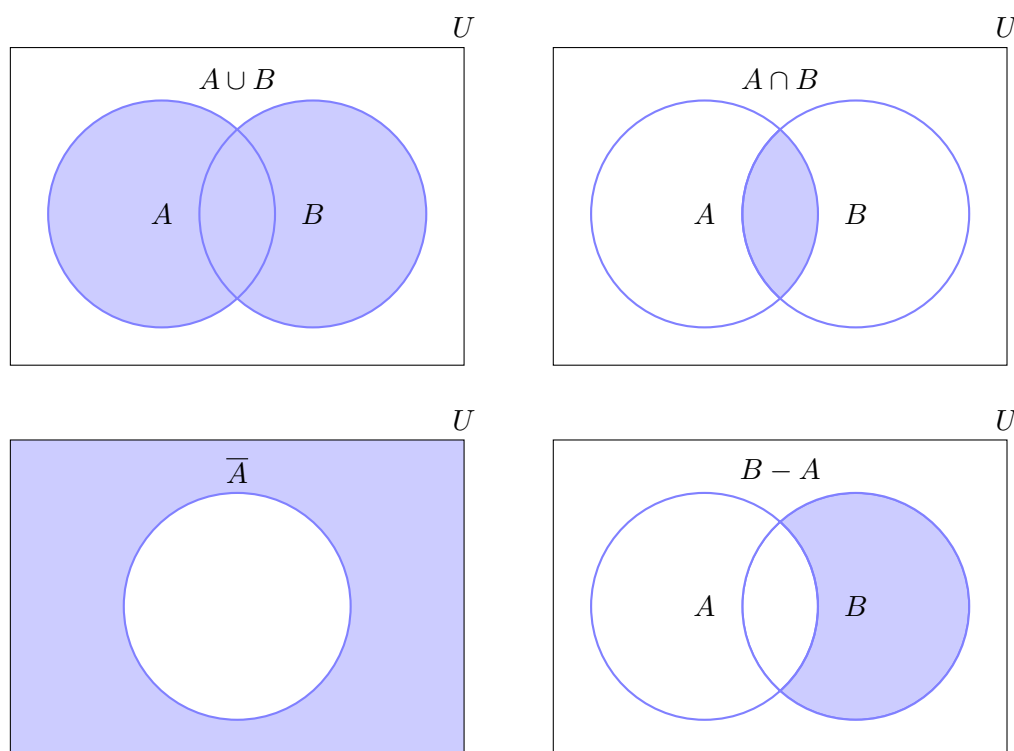
Der findes to typer af udfaldsrum, henholdsvis et diskret og kontinuert udfaldsrum. Forskellen på disse to er at der i et diskret udfaldsrum kun er et givent antal mulige udfald. Et eksempel på et diskret udfaldsrum kunne være en sekssidet terning. Denne sekssidet terning vil have et udfaldsrum U der indeholder elementerne $1, 2, 3, 4, 5, 6, 7$ dette kan skrives som $U = \{1, 2, 3, 4, 5, 6\}$. Altså vi har tale om et udfaldsrum med længden seks, hvilket skrives $|U| = 6$ eller som n . Hvorimod et kontinuert udfaldsrum kunne være højden på elever i en klasse, udfaldsrummet kunne altså være defineret som $U = [150cm; 210cm]$. Udfaldsrummet er kontinuert da elevers 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 $|U| \notin \mathbb{R}$. Derfor kaldes det også et endeligt og et ikke-endeligt udfaldsrum [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 operatorer 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})

Betegnelserne er de samme, som vi bruger i mængdelæren, og vi taler derfor om foreningsmængden $A \cup B$, fællesmængden $A \cap B$, mængdedifferensen $A \setminus B$ samt komplementmængden A^c [SC]. Nedenstående figurer også kaldet venn-diagrammer, illustreres overstående mængder og deres operatorer.



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% hvorimod sandsynligheden for 3 er 20%. Desuden kan vi bruge additions loven til at udlede den totale sandsynlighed for udfaldsrummet [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 bliver det desuden gældende at $P(A) = P(1) + P(2)$.

Sandsynlighedsfunktion kan også være afhængig af en anden sandsynlighed, altså et eksempel på dette kunne være en kasse is. Kassen med is indeholder fire slags is, hvor 50% af isen er med mørk chokolade og den resterende 50% er med hvid chokolade, lad denne sandsynlighed være betegnet med P_c . Desuden er 70% af de der har mørk chokolade mandler og 50% af de med hvids chokolade har også mandler, lad denne sandsynlighed vil således være beskrevet som $P(u_i|P_c)$. Altså sandsynligheden for udfaldet u_i er afhængig af P_c

1.3 Sandsynlighedsfelt

Et sandsynlighedsfelt findes på to former, det endelige - også kaldet det diskrete sandsynlighedsfelt - samt det kontinuerte sandsynlighedsfelt også kaldet et ikke-endeligt sandsynlighedsfelt. Et 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, altså kan $P(u_1) \neq P(u_2)$ [SC].

1.4 Stokastiske variabler

Stokastiske variabler findes af to typer der begge denoteres med X , lad desuden x_i være et konkret udfald, disse to typer er henholdsvis diskret og kontinuert. En stokastiske variable er en variable som kan tage alle værdier i et givent udfaldsrum, med sandsynligheden P , altså $P(X = u_i) = P(u_i)$. Forskellen mellem diskret og kontinuert stokastiske variabler er deres udfaldsrum og deres tilhørende regneregler. Et eksempel på en stokastiske variabel kunne være Europæisk Roulette, dette spil har et symmetrisk sandsynlighedsfelt med udfaldsrummet $U = \{0, 1, \dots, 36\}$ der har indeholder $37 = |U|$ udfald. Det vil altså 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, da der er tale om et symmetrisk sandsynlighedsfelt. Altså $P(u_1) = P(u_i) = \dots = P(u_n)$, det betyder at $P(X = 1) = P(X = 4)$. Ydermere er roulettens udfaldsrum opdelt i tre farver grøn, rød og sort, disse kan betegnes som delmængderne henholdsvis G , R og S . Vi kan således beskrive sandsynligheden for at den stokastiske variabel X vil ligge i en delmængde således $P(X \in S) = P(X \in R) = \frac{18}{37}$ og $P(X \in G) = \frac{1}{37}$. Delmængden G kunne også have været defineret som $G = (R \cup S)^c$, altså de udfald som ikke falder ind under hverken R eller S , eller $G = 0$ [NM]

Lad desuden $\mathbb{E}[X]$ være den forventede værdi af X , den forventede værdi skal forstås næsten som et gennemsnit. I det ovenstående eksempel vil den forventede værdi X være $\mathbb{E}[X] = 18$, dette kan udregnes på følgende måde:

$$\mathbb{E}[X] = \sum_{i=1}^n \frac{u_i}{n} \quad (1.6)$$

Dette er kun tilfældet da det er et symmetrisk sandsynlighedsfelt, hvis sandsynligheden for de forskellige udfald er forskellige fra hinanden, altså $P(u_1) \neq P(u_2)$. Her ville man bruge følgende formel for den forventede værdi af X .

$$\mathbb{E}[X] = \sum_{i=1}^n u_i \times P(u_i) \quad (1.7)$$

Stokastiske Variabler på en computer er ofte repræsenteret som et pseudo-genereret tilfældigt tal, hvilket er et tilfældigt tal udvalgt af computeren. Disse tilfældigt genererede

tal bliver brugt til at udføre eksperimenter og simulering af Monte Carlo Metoden [**SBM**]. Dette vil komme til udtryk i det følgende afsnit, hvori det vil blive vist hvordan Monte Carlo Simulering gør brug af Stokastiske Variabler.