1 Stokastiske variabler

En væsentlig del af Monte Carlo Metodens teori og praksis bygger på stokastiske variabler. For at forstå stokastiske variabler kræves der en forståelse for simplere begreber indenfor statistik. Disse begreber er: udfaldsrum, delmængde, sandsynlighedsfunktion samt sandsynlighedsfelt. Disse begreber redegøres der for i nedenstående afsnit.

1.1 Udfaldsrum og delmængder

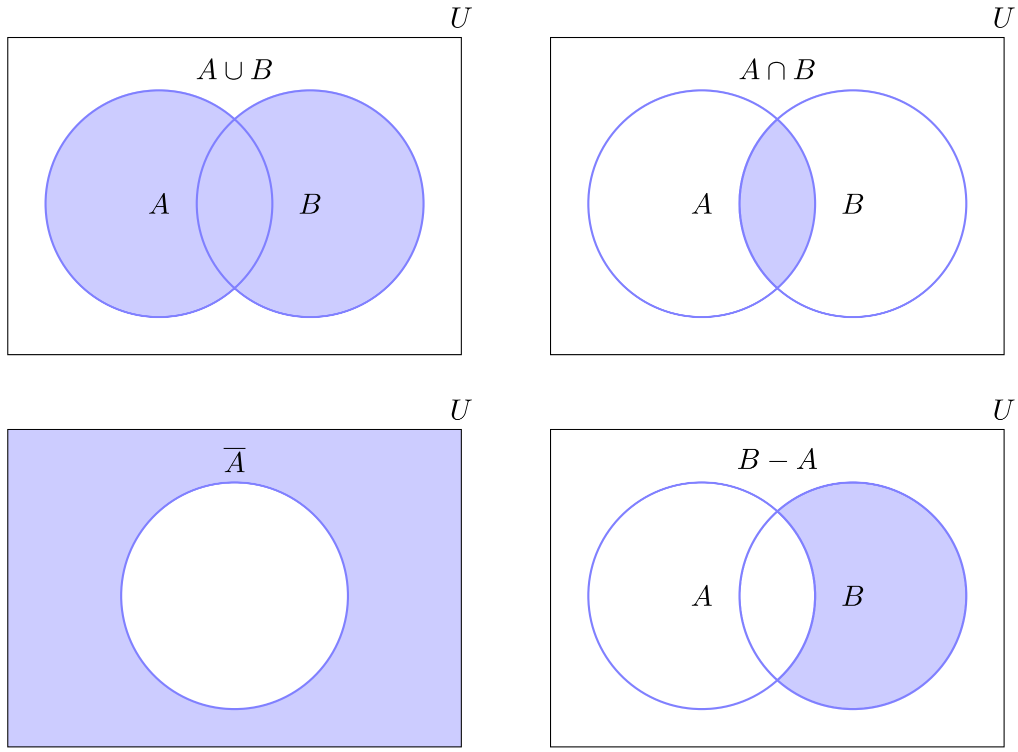
Der findes to typer af udfaldsrum, et diskret og kontinuert udfaldsrum. Forskellen på disse to er, at der i et diskret udfaldsrum kun er et givent antal mulige udfald. Et eksempel på et diskret udfaldsrum er en sekssidet terning. Denne sekssidet terning vil have et udfaldsrum U, der indeholder elementerne $1, 2, 3, 4, 5, 6$. Det kan skrives som $U = \{1, 2, 3, 4, 5, 6\}$. Altså er der tale om et udfaldsrum med seks elementer, hvilket skrives $|U| = 6$ eller som n. Et kontinuert udfaldsrum kunne eksempelvis være højden af elever i en klasse. I dette tilfælde kan udfaldsrummet være defineret som $U = [150cm; 210cm]$. Udfaldsrummet er kontinuert da elevers højde kan variere med et infinitesimal. En elev a kan være en uendelig lille stykke højere end elev b. Et sådan udfaldsrum indenholder ikke et endeligt antal elementer [**SC**].

En delmængde er en mængde af et udfaldsrum. Delmængden beskrives ofte med A eller B. Hvis vi endnu engang tager udgangspunkt i den sekssidet terning, kan en delmængde af ternings udfaldsrum være $A = \{1, 3\}$. En delmængde indeholder derved 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, at 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 samt 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. $A \backslash B$: 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 $A \backslash B$ samt komplementærmængden A^c [**SC**]. Nedenstående figurer, også kaldet venn-diagrammer, illustrerer overstående mængder og deres operatorer.



1.2 Sandsynlighedsfunktion

Sandsynlighedsfunktion denoteres P og beskriver sandsynligheden for et element i 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 additionsloven til at udlede den totale sandsynlighed for udfaldsrummet [**SC**].

$$P(1) + P(2) + P(3) = 1 \tag{1.1}$$
$$P(\{1,2\}) + P(3) = 1 \tag{1.2}$$

Desuden kan $\{1,2\}$ også beskrives som en delmængde af U på formen $A = \{1,2\}$. Herved gælder det, at $P(A) = P(1) + P(2)$.

Sandsynlighedsfunktion kan også være afhængig af en anden sandsynlighed. 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 de resterende 50% er med hvid chokolade. Lad denne sandsynlighed værende betegnet med P_c. Desuden har 70% af de is der har mørk chokolade, mandler og 50% af isene med hvide chokolade også mandler. Denne vil således være beskrevet som $P(u_i|P_c)$. Herved er sandsynligheden for udfaldet u_i afhængig af P_c.

1.3 Sandsynlighedsfelt

Et sandsynlighedsfelt findes på to former. Det endelige - også kaldet det diskrete sandsynlighedsfelt - samt det kontinuærte 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.

$$P(U) = 1 \tag{1.3}$$
$$P(u_1) + P(u_i) + \cdots + P(u_n) = 1 \tag{1.4}$$
$$P(u_1) = P(u_i) = \cdots = P(u_n) = \frac{1}{|U|} \tag{1.5}$$

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 hinanden. Altså kan $P(u_1) \neq P(u_2)$ [**SC**].

1.4 Stokastiske variabler

Der findes to typer af stokastiske variabler. Disse er begge denoteres med X og desuden er x_i 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 diskrete og kontinuerte stokastiske variabler er udfaldsrummet og de tilhørende regneregler. Et eksempel på en stokastiske variabel er Europæisk Roulette. Dette spil har et symmetrisk sandsynlighedsfelt med udfaldsrummet $U = \{0, 1, \ldots, 36\}$, hvilket indeholder $37 = |U|$ udfald. Det vil altså sige, at den kugle man smider ned i spillet kan beskrives som en stokastiske variabel X, og X har en lige stor sandsynlighed for at blive et element i udfaldsrummet, da der er tale om et symmetrisk sandsynlighedsfelt, hvilket betyder at $P(u_1) = P(u_i) = \cdots = P(u_n)$, 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 G, R og S. Vi kan beskrive sandsynligheden for, at den stokastiske variabel X ville være et element i delmængden 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 $\mathbb{E}[X] = 18$. Det kan udregnes på følgende måde:

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

Dette er kun tilfældet ved et symmetrisk sandsynlighedsfelt, hvis sandsynligheden for de forskellige udfald er forskellige fra hinanden, altså $P(u_1) \neq P(U_2)$. Her vil man kunne anvende følgende formel for den forventede værdi af X.

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

Stokastiske variabler på en computer er ofte repræsenteret som et pseudo-genereret tilfældigt tal, hvilket er et tilfældigt tal udvalgt af en computer. 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.