Udfald og sandsynlighed

Hvor statistik essentielt handler om at analysere et givet datasæt af observationer eller hændelser der allerede er sket, handler sandsynlighed om at kunne opstille en prognose ift. en fremtidig hændelse. Dette gør at begrebet *sandsynlig* nok er et af de vigtigste begreber inden for anvendt matematik, hvor det ofte går ud på at opstille matematiske modeller, der kan levere resultater til at kunne fremsætte en prognose om fremtidige hændelser.

For ud over at kunne bestemme sandsynligheden for et bestemt udfald af en tilfældig proces, såsom at vinde i et lotteri eller at slå et bestemt terningekast, udgør sandsynlighedsregning ligeledes en god værktøjskasse til at kunne vurdere anvendeligheden af en given videnskabelig metode. En videnskabelige metode, der kan levere et resultat og derved en prognose, som med stor sandsynlighed er korrekt ift. den virkelige hændelse, er selvsagt en anvendelig metode.

Sandsynlighed kan også betragtes som en størrelse, der beskriver om hvorvidt noget er forudsigelig (*deterministisk*) eller uforudsigeligt (*stokastisk/probalistisk*).

I nogle tilfælde kan sandsynligheder udledes rent deduktivt, f.eks. kan sandsynligheden for udfaldet af et terningekast udledes ved at betragte den geometriske symmetri i en terning til at alle udfald er lige sandsynlige. I andre tilfælde er det system som betragtes så komplekst, at sandsynligheden skal undersøges induktivt ved brug af statistik. I sådanne tilfælde skal sandsynlighed betragtes som grænseværdien af den statistiske analyse hvor antallet af observationer, dvs. datasættets størrelse, er uendelig stor. Dette princip kaldes også for *store tals lov*.

Sandsynlighedsfelt

Den matematiske struktur til at beskrive sandsynligheder minder meget om strukturen for statistike metoder hvor forskellen hovedsageligt skal findes i måden hvorpå de forskellige værdier skal fortolkes.

Ligesom man i statistik har nogle dataværdier der tilhører en observationsmængde, dvs. de værdier man kan observere, har man i sandsynlighed en udfaldsmængde der angiver hvilke værdier der kan være udfaldet af et stokastisk eksperiment. Hvis man f.eks. betragter hvilke mulige udfald, der kan ske ved at kaste en 6-sidet terning, vil man kunne udlede at udfaldet u må tilhøre mængden:

$$u \in \left\{ \boxdot, \boxdot, \boxdot, \boxdot, \boxdot, \boxdot \right\}$$

Udfaldsmængden kaldes også et *udfaldsrum*.

Ligesom der for statistiske datasæt kunne tilskrives en frekvens til hver mulig dataværdi i observationssættet, kan der for stokastiske eksperimenter tilskrives en sandsynlighed til hver mulig udfald. Dette gøres ved brug af sandsynlighedsfunktionen P hvor man indsætter et helt udsagn. Hvis man f.eks. vil skrive at der er $\frac{1}{6}$ sandsynlighed for at udfaldet af et terningkast er $\boxdot$, dvs. $u = \boxdot$, skriver man:

$$P\big(u = \boxdot\big) = \tfrac{1}{6}$$

Analogien mellem frekvens og sandsynlighed gælder også for dens sum. Ligesom i en statistisk fordeling hvor summen af frekvenserne skal give 1 eller 100% skal summen af sandsynlighederne for alle udfald også give 1 eller 100%. I eksemplet med terningkastet giver dette:

$$P\big(u = \boxdot\big) + P\big(u = \boxdot\big) + ... + P\big(u = \boxdot\big) = 1$$

Ydermere gælder der ligesom frekvenser, at sandsynligheder ikke kan have en negativ værdi, hvilket sammen med sætningen om summen sandsynligheder også medfører, at en sandsynlighed er en værdi mellem 0 og 1:

$$0 \leq P\big(u\big) \leq 1$$

Generelt set gælder følgende sætning:

> ***Sandsynlighedsfelt***
> Et sandsynlighedsfelt er en given mængde af udfald med et tilhørende sandsynlighed for hvert udfald:
>
> $$u_1, u_2, ..., u_n$$
> $$p_1, p_2, ..., p_n \qquad \text{hvor} \quad p_i = P(u = u_i)$$
>
> Sandsynligheden for et udfald u er en talværdi der er større eller lig med 0 og mindre eller lig med 1:
>
> $$0 \leq p_i \leq 1$$
>
> Lægges sandsynlighederne for alle n mulige udfald sammen giver summen 1 (eller 100%):
>
> $$p_1 + p_2 + ... + p_n = 1$$

Sandsynligheder kan, ligesom statistiske frekvenser, opskrives ved en tabel og afbilledes grafisk ved et søjlediagram. For et eksempel med hvor der er 6 forskellige udfald $u_1, u_2, ..., u_6$ kunne tabellen for sandsynlighedsfeltet f.eks. se ud som følger:

u	u_1	u_2	u_3	u_4	u_5	u_6
p	10%	10%	20%	30%	10%	20%

Det tilsvarende søjlediagram for sandsynlighedsfeltet vil derfor være:

Symmetrisk sandsynlighedsfelt

I tilfældet at sandsynlighederne for alle udfald er lige store siger man at man har et *symmetrisk sandsynlighedsfelt*.

I eksemplet med et terningkast af en 6-sidet terning vil man, hvis terningen er "ærlig", netop observere at sandsynlighederne for alle udfald er lige store, nemlig $\frac{1}{6}$ hvilket opstillet i en tabel kan vises ved:

u	$\boxdot$	$\boxdot$	$\boxdot$	$\boxdot$	$\boxdot$	$\boxdot$
p	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Det tilsvarende søjlediagram er:

Man har altså her et symmetrisk sandsynlighedsfelt. Grunden til at sandsynligheden netop er $\frac{1}{6}$ kommer af, at der er 6 mulige udfald. Idet summen af sandsynlighederne skal give 1 og at denne skal deles i 6 lige store sandsynligheder vil hver sandsynlighed være $\frac{1}{6}$.

Generelt set gælder følgende sætning:

> ***Symmetrisk sandsynlighedsfelt***
> Et sandsynlighedsfelt er *symmetrisk* når sandsynligheden for alle udfald er lige store:
>
> $$p_1 = p_2 = ... = p_n$$
>
> Sandsynligheden for hvert udfald p_i kan ydermere beregnes direkte ud fra antallet af forskellige udfald n ved:
>
> $$p_i = \frac{1}{n} \quad \text{for} \quad i \in 1, 2, ..., n$$
>
> Talværdien n er ligeledes størrelsen af udfaldsmængden $n = |\mathbb{U}|$.