

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 statistiske 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:

$$u \in \left\{ \begin{array}{c} \square, \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array}, \begin{array}{|c|c|} \hline \square & \begin{array}{|c|} \hline \square \\ \hline \end{array} \\ \hline \end{array}, \begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|} \hline \square \\ \hline \end{array} \\ \hline \end{array}, \begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array} \\ \hline \end{array}, \begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|} \hline \square \\ \hline \end{array} \\ \hline \end{array} \end{array} \right\}$$

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 terningekast er $\square$, dvs. $u = \square$, skriver man:

$$P(u = \square) = \frac{1}{6}$$

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 terningekastet giver dette:

$$P(u = \square) + P(u = \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array}) + \dots + P(u = \begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|} \hline \square \\ \hline \end{array} \\ \hline \end{array}) = 1$$

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:

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

Generelt set gælder følgende sætning:

Sandsynlighedsfelt

Et sandsynlighedsfelt er en given mængde af udfald med et tilhørende sandsynlighed for hvert udfald:

$$u_1, u_2, \dots, u_n$$

$$P(u = u_1), P(u = u_2), \dots, P(u = u_n)$$

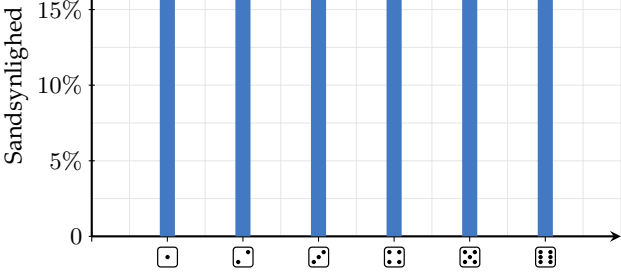
Sandsynligheden for et udfald u er en talværdi der er større eller lig med 0 og mindre eller lig med 1:

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

Lægges sandsynlighederne for alle n mulige udfald sammen giver summen 1 (eller 100%):

$$P(u = u_1) + P(u = u_2) + \dots + P(u = u_n) = 1$$

Sandsynligheder kan, ligesom statistiske frekvenser, afbildes grafisk ved et søjlediagram hvor udfaldsmængden plottes ift. sandsynlighederne for hvert udfald. For eksemplet med et terningekast af en 6-sidet terning vil sandsynlighedsfordelingen altså kunne afbildes ved nedenstående søjlediagram.



Symmetrisk sandsynlighedsfelt

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

I eksemplet med et terningekast 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	$\square$	$\begin{array}{ c c } \hline \square & \square \\ \hline \end{array}$	$\begin{array}{ c c } \hline \square & \begin{array}{ c } \hline \square \\ \hline \end{array} \\ \hline \end{array}$	$\begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c } \hline \square \\ \hline \end{array} \\ \hline \end{array}$	$\begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c c } \hline \square & \square \\ \hline \end{array} \\ \hline \end{array}$	$\begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c } \hline \square \\ \hline \end{array} \\ \hline \end{array}$
$P(u)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

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:

$$P(u = u_1) = P(u = u_2) = \dots = P(u = u_n)$$

Sandsynligheden for hvert udfald kan ydermere beregnes direkte ud fra antallet af forskellige udfald n ved:

$$P(u) = \frac{1}{n}$$

Talværdien n er ligeledes størrelsen af udfaldsmængden $n = |\mathbb{U}|$.

Stokastiske variable

Som det kan ses af eksemplet på et terningekast med en 6-sidet terning så er det ikke altid at udfaldene er tal. Dette kan selvfølgelig gøre selve analysen af et stokastisk eksperiment vanskeligt og man har derfor ofte brug for en variabel, der kvantificerer et udfald. En sådan variabel kaldes en stokastisk variabel.

If. terningekastet kan man f.eks. vælge at den stokastiske variabel x skal angive antallet af øjne som terningen viser. I dette tilfælde vil man kunne opstille en tabel der viser værdien af den stokastiske variabel x svarende til udfaldet af terningekastet og sandsynligheden herfor.

u	$\square$	$\begin{array}{ c c } \hline \square & \square \\ \hline \end{array}$	$\begin{array}{ c c } \hline \square & \begin{array}{ c } \hline \square \\ \hline \end{array} \\ \hline \end{array}$	$\begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c } \hline \square \\ \hline \end{array} \\ \hline \end{array}$	$\begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c c } \hline \square & \square \\ \hline \end{array} \\ \hline \end{array}$	$\begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c } \hline \square \\ \hline \end{array} \\ \hline \end{array}$
x	1	2	3	4	5	6
$P(u)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Af tabellen kan man se at der er en en-til-en sammenhæng mellem udfaldet u og den stokastiske variabel x . Derfor er sandsynligheden for at den stokastiske variabel har værdien 1 den samme som sandsynligheden for at udfaldet er $\square$ og derfor givet ved:

$$P(x = 1) = P(u = \square) = \frac{1}{6}$$

Det er dog ikke altid at den stokastiske variabel direkte har en en-til-en sammenhæng med udfaldet som det var tilfældet med antallet af øjne x . Hvis man i stedet for definerer en anden stokastisk variabel y som giver talværdien 1 når udfaldet er enten $\square$ eller $\begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array}$ og talværdien 0 når udfaldet er enten $\begin{array}{|c|c|} \hline \square & \begin{array}{|c|} \hline \square \\ \hline \end{array} \\ \hline \end{array}$, $\begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|} \hline \square \\ \hline \end{array} \\ \hline \end{array}$ eller $\begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|} \hline \square \\ \hline \end{array} \\ \hline \end{array}$ kan man opstille tabellen:

u	$\square$	$\begin{array}{ c c } \hline \square & \square \\ \hline \end{array}$	$\begin{array}{ c c } \hline \square & \begin{array}{ c } \hline \square \\ \hline \end{array} \\ \hline \end{array}$	$\begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c } \hline \square \\ \hline \end{array} \\ \hline \end{array}$	$\begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c c } \hline \square & \square \\ \hline \end{array} \\ \hline \end{array}$	$\begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c c } \hline \begin{array}{ c } \hline \square \\ \hline \end{array} & \begin{array}{ c } \hline \square \\ \hline \end{array} \\ \hline \end{array}$
y	1	1	0	0	0	0
$P(u)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Her kan nu ses at der ikke længere er en en-til-en sammenhæng mellem udfaldet og den stokastiske variabel y , da der er flere udfald som giver samme talværdi for y . Sandsynligheden for at $y = 1$ skal nu beregnes ved at lægge sandsynlighederne for alle de udfald hvor $y = 1$ sammen:

$$P(y = 1) = P(u = \square) + P(u = \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array})$$

$$= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3} \approx 0.33 = 33\%$$

Der er altså 33% sandsynlighed for at den stokastiske variabel $y = 1$. Beregnes i stedet sandsynligheden for at $y = 0$, lægges sandsynlighederne for alle de udfald hvor $y = 0$ sammen:

$$P(y = 0) = P(u = \begin{array}{|c|c|} \hline \square & \begin{array}{|c|} \hline \square \\ \hline \end{array} \\ \hline \end{array}) + P(u = \begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|} \hline \square \\ \hline \end{array} \\ \hline \end{array}) + P(u = \begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array} \\ \hline \end{array}) + P(u = \begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|c|} \hline \begin{array}{|c|} \hline \square \\ \hline \end{array} & \begin{array}{|c|} \hline \square \\ \hline \end{array} \\ \hline \end{array})$$

$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6} = \frac{2}{3} \approx 0.67 = 67\%$$

Det skal her bemærkes at selvom der er tale om et symmetrisk sandsynlighedsfelt for selve udfaldet u er dette ikke nødvendig gældende for en stokastisk variabel som f.eks. y idet:

$$P(y = 0) \neq P(y = 1)$$