Kapitel 1

Sandsynlighedsfordelinger og stokastiske variable

1.1. Sandsynlighed.

"Sandsynligheden for, i et slag med to terninger, at få mindst 11 øjne, er $3/36 = 8.3\%$"

"Sandsynligheden for at barnet bliver en pige er ca. $1/2$"

"Sandsynligheden for at der, i en klasse med 25 børn, ikke findes to med samme fødselsdag, er 43.13%"

"Sandsynligheden for at en 20-årig mand kommer til at opleve sin 50 års fødselsdag er 93.6%"

"Sandsynligheden for at finde liv på Mars er under 0.1%"

Fælles for udsagn som disse er, at de angiver et tal mellem 0 og 1 (eller et antal procent mellem 0 og 100) som *sandsynligheden* for, at en eller anden hændelse vil indtræffe. For at indkredse, hvad sandsynlighed er for noget, ser vi lidt nærmere på nogle af disse udsagn.

Det første er det enkleste. Det er almindeligt accepteret, at de 36 mulige udfald (1,1), (1,2), ..., (6,5), (6,6) er lige sandsynlige, og netop 3 af dem giver et samlet antal øjne som er ≥ 11. Resultatet af et enkelt slag er helt uforudsigeligt, men hvis man slår mange gange, vil omkring 8.3% af slagene give 11 eller 12. Anderledes sagt, hvis vi lader $\#(\text{sum} \geq 11)$ betegne antallet af slag blandt N, som resulterer i mindst 11 øjne, så er

$$\lim_{N\to\infty} \frac{\#(\text{sum} \geq 11)}{N} = \frac{3}{36}.$$

En størrelse, der som $\#(\text{sum} \geq 11)/N$ angiver antallet af gange, en bestemt hændelse er indtruffet, i forhold til det samlede antal gange, den kunne være indtruffet, kaldes en *relativ hyppighed* eller en *frekvens.* Det ligger i begrebet sandsynlighed, således som det bruges i daglig tale, at en sandsynlighed er en størrelse, der approksimerer den relative hyppighed af en hændelse i en lang serie af ens udførte forsøg. Eller, om man vil,

En sandsynlighed er en relativ hyppighed
i en uendelig serie af forsøg.

Fra et matematisk synspunkt viser det sig desværre, at denne fortolkning ikke umiddelbart kan bruges som definition af begrebet sandsynlighed. Vi vil i stedet – som i alle andre introduktioner til sandsynlighedsregningen – tage udgangspunkt i de formelle regneregler for sandsynligheder, som man mere eller mindre bevidst underforstår, når man taler om dem. Herved får man en matematisk ramme, som viser sig egnet til beskrivelse af fænomener, der involverer tilfældighed. En stærk bekræftelse på denne teoris gyldighed får vi senere, når fortolkningen af sandsynligheder som grænseværdier for relative hyppigheder dukker op igen næsten af sig selv, i form af en matematisk sætning (de store tals lov, se kapitel 4).

Men selv om vi således begynder med at se bort fra det, der med et fint udtryk kaldes *sandsynlighedsregningens frekvensfortolkning*, er denne fortolkning stadig væsentlig for relevansen af de matematiske modeller, vi konstruerer til beskrivelse af virkelige fænomener. Det kan illustreres ved det sidste af udsagnene på side 1, det som vedrører eksistensen af liv på Mars. Udsagnet (som i øvrigt er frit opdigtet, lige som alle de andre) er nærmest medtaget for at illustrere, hvorledes ordet “sandsynlighed” undertiden bruges på en helt anden måde end i sandsynlighedsregningen og dens anvendelser. Hvis udsagnet om forekomsten af liv på Mars skal tillægges en præcis mening, må det forudsætte et rimeligt stort antal undersøgelser af Mars-lignende planeter; eller et tilsvarende materiale af laboratorieforsøg over adskillige millioner år; eller, i det mindste, et detaljeret kendskab til de processer, der frembringer og udvikler liv, sammen med kendskab til sandsynligheder for, at de finder sted under kemiske og fysiske forhold som på Mars. Alt dette er naturligvis absurd i øjeblikket, og det er klart, at et udsagn som dette blot skal forstås som “Jeg anser det for helt usandsynligt, at der er liv på Mars”.

Udsagnet om sandsynligheden for, at en 20-årig mand kommer til at opleve sin 50 års fødselsdag, illustrerer andre fortolkningsproblemer, som kan opstå, når model og virkelighed konfronteres. Udregningen af sandsynligheden 93.6% er baseret på “Dødelighedstavle for årene 1984–85” i Statistisk Årbog 1987. De der angivne tal er beregnet ud fra størrelserne af de grupper, der fremkommer ved opdeling af befolkningen efter køn og alder, samt antal døde i disse grupper i optællingsåret. Udsagnet vedrører således en dansk “gennemsnitsmand”, og har ikke uden videre gyldighed for en person, som man ved noget mere om (f.eks. en sportstrænet ikke-ryger eller en overvægtig alkoholiker). Desuden forudsættes det ved udregningen, at dødeligheden i aldersgrupperne vil være den samme fremover som i 1984–85, så f.eks. en netop 40-årigs chance for at blive 41 er den samme i 2004–5 som i 1984–85. Udsagnet er således langt mere kompliceret, og skal forsynes med langt flere forbehold, end udsagnet vedrørende resultatet af et slag med to terninger. Alligevel er det klart, at det har en sandheds- og nytteværdi (f.eks. for forsikringsbranchen), som går langt ud over, hvad man kan tillægge udsagnet om liv på Mars.

1.2. Udfaldsrum og sandsynlighedsfunktion.

Vi begynder så småt på den matematiske formalisme. For at kunne tale om sandsynligheder, vil vi forudsætte, at der er givet en mængde E, kaldet *udfaldsrummet*, hvis elementer er de mulige udfald af det forsøg eller hændelsesforløb, vi taler om. I tilfældet med de to terninger er det naturligt at vælge mængden

$$E = \{(1,1),(1,2),\ldots,(6,5),(6,6)\} = \{1,\ldots,6\} \times \{1,\ldots,6\}$$

bestående af 36 elementer, der på indlysende måde repræsenterer de 36 mulige udfald. I tilfældet med de 25 børn og deres fødselsdage kunne vi vælge mængden

$$E = \{1,\ldots,365\}^{25} = \underbrace{\{1,\ldots,365\} \times \cdots \times \{1,\ldots,365\}}_{\text{25 gange}}$$

idet et udfald opfattes som en liste bestående af 25 tal, der angiver hvert barns fødselsdag ved dagens nummer i kalenderåret.

Sandsynlighederne for de enkelte udfald kan herefter specificeres ved hjælp af en *sandsynlighedsfunktion* $p\colon E \to \mathbf{R}$, som til hvert udfald knytter dets sandsynlighed. I eksemplet med de to terningkast vil vi naturligvis definere

$$p(x) = \frac{1}{36} \quad \text{for alle} \quad x = (x_1, x_2) \in E.$$

I tilfældet med de 25 børn er det tilsvarende underforstået, at $p(x) = 365^{-25}$ for alle $x = (x_1, \ldots, x_{25}) \in E$ (hvilket vil blive begrundet senere).

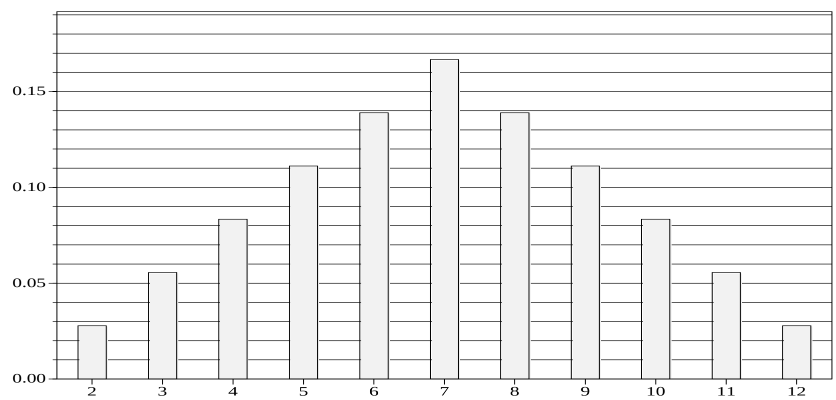
Sandsynlighedsfunktionens værdier kaldes *punktsandsynligheder.* De to ovennævnte tilfælde er karakteriserede ved, at alle punktsandsynligheder er ens, dvs. sandsynlighedsfunktionen er konstant. Det vil den ikke altid være. Men vi vil altid forudsætte, at den opfylder følgende to betingelser:

(p1) $$p(x) \geq 0,$$

(p2) $$\sum_{x \in E} p(x) = 1.$$

Eksempel 1.2.1. En sandsynlighedsfunktion på $E = \{2, 3, \ldots, 12\}$ kan defineres ved

$$\begin{aligned} p(2) = p(12) &= 1/36 \\ p(3) = p(11) &= 2/36 \\ p(4) = p(10) &= 3/36 \\ p(5) = p(9) \;\; &= 4/36 \\ p(6) = p(8) \;\; &= 5/36 \\ p(7) \qquad\quad &= 6/36 \end{aligned}$$



Sandsynlighedsfunktionen i eksempel 1.2.1.

(Dette er i øvrigt sandsynlighedsfunktionen, som beskriver fordelingen af summen af øjnene i et slag med to terninger. Herom senere).

I eksemplet med de 25 børn og deres fødselsdage kan vi begrunde valget af en konstant sandsynlighedsfunktion på følgende måde. Hvis alderskriterierne for optagelse i skolen er uændrede fra år til år, og hvis kriterierne for fordeling i klasser ikke har noget med alder at gøre, må børnenes fødselsdage være jævnt fordelt over kalenderåret. Dette forudsætter naturligvis, at antal fødte pr. dag ikke er årstidsafhængigt; hvilket det nok i virkeligheden er, men årstidsvariationen er så lille, at vi vælger at ignorere den. Vi forudsætter endvidere, at et barns fødselsdag falder uafhængigt af, hvornår de andre børn har fødselsdag. Dette er en antagelse om *stokastisk uafhængighed*, et begreb som vil blive diskuteret ret udførligt senere. Her indebærer det blandt andet, at vi ser bort fra forekomsten af tvillingepar i samme klasse. Endelig gør vi den (for visse årgange helt korrekte) antagelse, at et år har 365 dage. Kort sagt, vore antagelser om fordelingen af de 25 børns fødselsdage går ud på, at udfaldet lige så godt kunne være fremkommet ved en lodtrækning, hvor hvert barn trækker en kugle fra en kasse med 365 godt blandede kugler, nummereret fra 1 til 365. Da sandsynlighedsfunktionen naturligvis ikke må ændre sig, selv om vi ændrer nummereringen af kuglerne i én eller flere kasser, følger det, at den må være konstant.

Opgave 1.2.1. Hvad er sandsynligheden for, ved et slag med to terninger, at få

(a) sum ≥ 10 ?

(b) to ens ?

(c) mindst én sekser ?

(d) netop én sekser ?

Opgave 1.2.2. Hvad er sandsynligheden for, ved et slag med tre terninger, at få summen 10 ?

Opgave 1.2.3. To personer vælges tilfældigt.

(a) Hvad er sandsynligheden for at de har samme fødselsdag ?

(b) Hvad er sandsynligheden for at de har forskellige fødselsdage ?

Opgave 1.2.4. Syv personer vælges tilfældigt.

(a) Hvad er sandsynligheden for at de alle er født på en søndag ?

(b) Hvad er sandsynligheden for at de er født på samme ugedag ?

(c) Hvad er sandsynligheden for at de er født på hver sin ugedag ?

Opgave 1.2.5. (Sandsynligheden for at få en hånd med 0 honnørpoint i bridge). Et spil kort blandes, og tretten kort trækkes. Hvad er sandsynligheden for hverken at få billedkort eller esser ?

1.3. Sandsynlighedsfordelinger.

Definition. Ved en *sandsynlighedsfordeling* eller et *sandsynlighedsmål* på en endelig mængde E forstås en afbildning

$$P \colon \mathcal{D}(E) \to \mathbf{R}$$

fra mængden $\mathcal{D}(E)$ af delmængder af E ind i den reelle akse, med følgende egenskaber:

$$\text{(P1)} \qquad P(A) \geq 0 \;\text{ for alle }\; A \subseteq E$$

$$\text{(P2)} \qquad P(E) = 1$$

$$\text{(P3)} \qquad P(A \cup B) = P(A) + P(B) \;\text{ for }\; A, B \subseteq E,\; A \cap B = \emptyset$$

Delmængder af E kaldes i denne forbindelse *hændelser*, og $P(A)$ fortolkes som sandsynligheden for at hændelsen A indtræffer. Bemærk, at vi skelner mellem udfald og hændelser; et udfald er jo et element i E.

Som vi skal se, er det at angive en sandsynlighedsfordeling på E præcis det samme som at angive en sandsynlighedsfunktion. Men først noterer vi os følgende umiddelbare konsekvenser af definitionen:

(1) $P(\emptyset) = 0$.

Thi ifølge (P3) ovenfor er $P(\emptyset) = P(\emptyset) + P(\emptyset)$.

(2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *for vilkårlige* $A, B \subseteq E$.

Denne regneregel (som, især i ældre litteratur, kaldes *additionssætningen*) har (P3) som specialtilfælde. Den bevises således: Da $A \setminus B$ og $A \cap B$ er disjunkte og tilsammen udgør A, følger det af (P3) at

$$P(A) = P(A \setminus B) + P(A \cap B).$$

Da $A \setminus B$ og B er disjunkte og tilsammen udgør $A \cup B$ er tilsvarende

$$P(A \cup B) = P(A \setminus B) + P(B).$$

Ved at trække disse to relationer fra hinanden og flytte lidt rundt på leddene får man umiddelbart (2).

(3) *For parvis disjunkte mængder* $A_1, \ldots, A_n$ *er*

$$P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n).$$

Dette bevises ved induktion efter n. For $n = 2$ er påstanden ækvivalent med (P3). Induktionstrinnet (altså beviset for, at påstanden er gyldig for $n + 1$ parvis disjunkte hændelser, når den antages gyldig for n) forløber således: Ifølge (P3) er

$$P(A_1 \cup \cdots \cup A_n \cup A_{n+1}) = P(A_1 \cup \cdots \cup A_n) + P(A_{n+1}).$$

Men ifølge induktionsantagelsen er dette lig med

$$P(A_1) + \cdots + P(A_n) + P(A_{n+1}),$$

hvilket skulle vises.

(4) $P(E \setminus A) = 1 - P(A)$.

Dette følger umiddelbart af (P2) og (P3).

Lad nu $p \colon E \to \mathbf{R}$ være en sandsynlighedsfunktion, og definer en afbildning $P \colon \mathcal{D}(E) \to \mathbf{R}$ ved

$$P(A) = \sum_{x \in A} p(x).$$

Da vil P være en sandsynlighedsfordeling. (P1) er jo trivielt opfyldt, (P2) følger af umiddelbart af forudsætningen (p2), og (P3) følger af at vi for disjunkte mængder A og B netop har

$$\sum_{x \in A} p(x) + \sum_{x \in B} p(x) = \sum_{x \in A \cup B} p(x).$$

Omvendt kan vi, ud fra en sandsynlighedsfordeling P, definere en funktion $p\colon E \to \mathbf{R}$ ved $p(x) = P(\{x\})$. Denne funktion vil være en sandsynlighedsfunktion. (p1) er jo trivielt opfyldt, og (p2) følger ved anvendelse af (3) ovenfor, idet E netop er disjunkt foreningsmængde af delmængderne $\{x\}$, $x \in E$. Hvis vi ud fra denne funktion p konstruerer en sandsynlighedsfordeling som beskrevet ovenfor, får vi åbenbart den sandsynlighedsfordeling P som vi startede med.

Sammenfattende har vi således

Sætning 1.3.1. *Mellem sandsynlighedsfunktioner* $p\colon E \to \mathbf{R}$ *og sandsynlighedsfordelinger* $P\colon \mathcal{D}(E) \to \mathbf{R}$ *er der en entydig korrespondance, bestemt ved*

$$P(A) = \sum_{x \in A} p(x)$$

og

$$p(x) = P(\{x\}).$$

Ligefordelinger. Vi har set to eksempler på sandsynlighedsfordelinger, givet ved en konstant sandsynlighedsfunktion p. Sådanne fordelinger kaldes *ligefordelinger*. Generelt kan vi definere *ligefordelingen på en endelig mængde* E ved sin sandsynlighedsfunktion

$$p(x) = 1/|E|,$$

idet vi med $|M|$ betegner antallet af elementer i en (endelig) mængde M. Den tilhørende sandsynlighedsfordeling er åbenbart givet ved

$$P(A) = |A|/|E|.$$

Eksempel 1.3.1. Betragt situationen med de 25 børn og deres fødselsdage (altså $E = \{1, \ldots, 365\}^{25}$, $P =$ ligefordelingen), og lad $A \subseteq E$ betegne den hændelse, at de 25 børn har forskellige fødselsdage. Elementerne i A kan optælles på følgende måde. Vi begynder med at dele op efter første barns fødselsdag x_1. For hvert af de 365 mulige valg af x_1 kan x_2 ($=$ andet barns fødselsdag) åbenbart vælges på 364 måder. For hvert valg af x_1 og x_2 (indbyrdes forskellige) kan x_3 vælges på 363 måder, osv. osv. Alt i alt fås

$$|A| = 365 \times 364 \times \cdots \times 341,$$

idet barn nr. 25 åbenbart har $341 = 365 - 24$ muligheder. Sandsynligheden for, at de 25 børn har 25 forskellige fødselsdage er således

$$P(A) = \frac{|A|}{|E|} = \frac{365 \times 364 \times \cdots \times 341}{365^{25}} = 0.4313.$$

OPGAVE 1.3.1. Tabellér sandsynligheden for, at k tilfældigt valgte personer har forskellige fødselsdage, for $k = 1, \ldots, 30$.

OPGAVE 1.3.2. Hvad er, i et slag med fem terninger, sandsynligheden for at få mindst én sekser? (Vink: det er lettere at udregne sandsynligheden for den komplementære hændelse; benyt (4)).

OPGAVE 1.3.3*. Lad P være en sandsynlighedsfordeling på E. Vis regnereglen

$$\begin{aligned} P(A \cup B \cup C) = &P(A) + P(B) + P(C) \\ &- P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C). \end{aligned}$$

(Vink: benyt (2) nogle gange).

OPGAVE 1.3.4. Hvad er, i et slag med tre terninger, sandsynligheden for at mindst to terninger viser det samme? (Vink: Lad A_{12} betegne den hændelse at terning nr. 1 og 2 viser det samme. Definer A_{13} og A_{23} tilsvarende, og benyt opgave 1.3.3).

OPGAVE 1.3.5*. Lad P være en sandsynlighedsfordeling på E.

(a) Vis at hvis $A \subseteq B$, så er $P(A) \leq P(B)$.

(b) Vis at $P(A_1 \cup \cdots \cup A_n) \leq P(A_1) + \cdots + P(A_n)$. (Vink: Vis først resultatet for $n = 2$, og benyt så induktion efter n).

1.4. Stokastiske Variable.

Man kan i princippet formulere hele sandsynlighedsregningen ved hjælp af sandsynlighedsfordelinger. Men notationen bliver let for tung. Betragt f.eks. følgende præcisering af resultatet vedrørende et slag med to terninger:

Lad P betegne ligefordelingen på $E = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$. Da er

$$P\left(\{(x_1, x_2) \mid x_1 + x_2 \geq 11\}\right) = \frac{3}{36}.$$

Vi vil fremover formulere dette lidt mere kortfattet:

Lad (X_1, X_2) være ligefordelt på $E = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$. Da er

$$P(X_1 + X_2 \geq 11) = \frac{3}{36}.$$

Den sidste formulering indebærer en vedtagelse om, at (X_1, X_2) betegner et tilfældigt udfald, som i dette tilfælde f.eks. kunne være resultatet af et slag med to terninger. Vi tillader os så at skrive hændelsen $\{(x_1, x_2)|x_1+$

$x_2 \geq 11\}$ på den simplere form $X_1 + X_2 \geq 11$. Sandsynlighedsmålet kaldes P, med mindre andet er nævnt. I denne sammenhæng kaldes (X_1, X_2) en *stokastisk variabel*. Vi anvender normalt store bogstaver for stokastiske variable.

Indtil videre er der blot tale om en forenklet sprogbrug. Ovenstående konvention er sammenlignelig med de konventioner, man kender fra matematisk analyse og fysik omkring det almindelige variabelbegreb. Ligesom en *variabel* i matematisk analyse er en størrelse der *varierer* (i modsætning til en konstant) er en *stokastisk variabel* X en størrelse, der varierer *stokastisk* (dvs. tilfældigt), således at X falder i delmængden $A \subseteq E$ med sandsynlighed $P(A)$.

Men det bliver lidt mere indviklet. Vi vil nemlig også tillade følgende formulering:

Lad (X_1, X_2) være ligefordelt på $E = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$, og sæt $Y = X_1 + X_2$. Da er

$$P(Y \geq 11) = \frac{3}{36}.$$

Hermed har vi indført endnu en – tilsyneladende uskyldig – konvention, ifølge hvilken vi har lov til at give et udtryk (her $X_1 + X_2$) i den oprindelige variabel et nyt navn (her Y), og benytte dette navn i stedet for udtrykket. Men denne konvention dækker over noget mere. Vi vil nemlig også *opfatte Y som en stokastisk variabel i sig selv, med sit eget udfaldsrum* $\{2, 3, \ldots, 12\}$ *og sin egen fordeling* (som i øvrigt bliver fordelingen i eksempel 1.2.1).

Den sidste konvention kan opfattes som en indarbejdning i sprogbrugen af følgende matematiske resultat:

Sætning 1.4.1. *Lad P være en sandsynlighedsfordeling på E, og lad $t\colon E \to F$ være en afbildning ind i en anden endelig mængde F. Da vil den ved*

$$Q(B) = P\left(t^{-1}(B)\right)$$

definerede afbildning

$$Q\colon \mathcal{D}(F) \to \mathbf{R}$$

være en sandsynlighedsfordeling på F.

Den således konstruerede sandsynlighedsfordeling på F kaldes den *transformerede fordeling* og betegnes undertiden $Q = P \circ t^{-1}$ eller $Q = t(P)$.

Beviset for sætningen er helt elementært, og overlades til læseren. Det væsentlige er fortolkningen af den transformerede fordeling Q, som går ud på, at

Hvis X er en stokastisk variabel på E med fordeling P, så er $Y = t(X)$ en stokastisk variabel på F med fordeling Q.

Dette er den eneste rimelige definition af Y's fordeling Q. Man vil jo naturligt opfatte hændelsen $\{Y \in B\}$ som et andet udtryk for hændelsen $\{t(X) \in B\} = \{X \in t^{-1}(B)\}$, og derfor kræve at

$$Q(B) = P\left(t^{-1}(B)\right).$$

Men det betyder præcis, at Q må være den transformerede fordeling, som givet ved sætningen.

Alt dette kan let virke som en gang tomt ordkløveri. Men det er væsentligt, at man gør sig klart, hvad betydningen af disse konventioner for stokastiske variable går ud på, ellers vil man let blive forvirret senere hen. Notationen i forbindelse med stokastiske variable dækker over en uklarhed i den måde, hvorpå vi normalt taler om tilfældige eksperimenter. Når vi f.eks. i situationen med de to terningkast taler om "sandsynligheden for at få sum mindst 11" gør vi os næppe klart, om det er delmængden $\{(5,6),(6,5),(6,6)\}$ af det totale udfaldsrum $\{1,\dots,6\} \times \{1,\dots,6\}$ vi taler om, eller det er delmængden $\{11,12\}$ af variationsområdet $\{2,3,\dots,12\}$ for summen. Konventionen vedrørende fordelingen af en transformeret stokastisk variabel sikrer, at vi heller ikke behøver at gøre os dette klart. Summens fordeling er definitionsmæssigt fastsat, så det passer.

Vi skulle herefter være i stand til at tale nogenlunde frit om flere stokastiske variable på én gang, forudsat at de alle er afledt af (dvs. givet som funktioner af) samme "underliggende" variabel. Normalt vil det ikke være nødvendigt at indføre navne for fordelingerne af de enkelte variable, fordi alle hændelser kan føres tilbage til denne underliggende variabel. Således har vi, i tilfælde af en enkelt transformeret variabel $Y = t(X)$ med fordeling Q, allerede vedtaget, at udtrykkene $Q(Y \in B)$, $P(Y \in B)$, $P(t(X) \in B)$ og $P\left(X \in t^{-1}(B)\right)$ betyder det samme, og betegnelsen Q kan derfor normalt undværes.

Eksempel 1.4.1. Ser vi igen på det ligefordelte udfald (X_1, X_2) af et slag med to terninger, kan vi, foruden

$$Y = X_1 + X_2 \in \{2,3,\dots,12\}$$

se på følgende afledte variable:

$$\begin{aligned}
&X_1 \in \{1,\dots,6\} \quad \text{(antal øjne for første terning)},\\
&X_2 \in \{1,\dots,6\} \quad \text{(antal øjne for anden terning)},\\
&Z_1 = X_1 \wedge X_2 \in \{1,\dots,6\} \quad \text{(det mindste af de to antal)},\\
&Z_2 = X_1 \vee X_2 \in \{1,\dots,6\} \quad \text{(det største af de to antal)},\\
&Z = (Z_1, Z_2) \in \{(z_1,z_2) \mid z_1, z_2 \in \{1,\dots,6\}, z_1 \le z_2\}.
\end{aligned}$$

Vi har hidtil omtalt resultatet af et terningkast, som om det var muligt at skelne mellem de to terninger (de kunne f.eks. have hver sin farve).

I almindelighed er de to terninger ikke til at kende fra hinanden, så den sidst definerede variable Z kan måske mere realistisk fortolkes som "udfaldet af et slag med to terninger". Bemærk at fordelingen af Z ikke er en ligefordeling. F.eks. er sandsynligheden for at få en treer og en firer jo

$$P(Z = (3,4)) = \frac{2}{36},$$

idet netop to udfald ((3,4) og (4,3)) i det oprindelige udfaldsrum fører til denne værdi af Z, medens sandsynligheden for at få to seksere er

$$P(Z = (6,6)) = \frac{1}{36},$$

idet denne hændelse svarer til netop ét udfald i det oprindelige udfaldsrum. Punktsandsynlighederne i fordelingen af Z ses at være

$$P(Z = (z_1, z_2)) = \begin{cases} 1/36 & \text{for } z_1 = z_2, \\ 2/36 & \text{for } z_1 < z_2. \end{cases}$$

Da alle de variable, vi her har indført, er givet som funktioner af (X_1, X_2), har det mening at tale om sandsynligheder for hændelser, der vedrører flere af dem. Vi kunne f.eks. være interesseret i sandsynligheden for hændelsen

$$\{Y \geq 5\} \cap \{Z_2 - Z_1 \leq 2\},$$

som er den hændelse, at summen er mindst 5 samtidig med at forskellen på de to øjenantal er højst 2. Vi kan udregne denne sandsynlighed ved optælling af elementerne i det oprindelige udfaldsrum, som tilfredsstiller disse betingelser. Men da hændelsen alene vedrører Z, idet den også kan skrives

$$\{Z_1 + Z_2 \geq 5\} \cap \{Z_2 - Z_1 \leq 2\},$$

er det også tilladt at udregne sandsynligheden ved summation af de relevante punktsandsynligheder i fordelingen af Z. Begge metoder skulle gerne give resultatet $18/36 = 1/2$ (Prøv!).

Opgave 1.4.1. Lad P være fordelingen på $E = \{1, \dots, 8\}$ med punktsandsynligheder

$$\begin{array}{llll} p(1) = 0.12 & p(2) = 0.08 & p(3) = 0.20 & p(4) = 0.11 \\ p(5) = 0.19 & p(6) = 0.14 & p(7) = 0.06 & p(8) = 0.10 \end{array}$$

Sæt $F = \{1, 2, 3, 4\}$, og definer $t \colon E \to F$ ved

$$\begin{array}{llll} t(1) = 1 & t(2) = 1 & t(3) = 1 & t(4) = 2 \\ t(5) = 3 & t(6) = 3 & t(7) = 4 & t(8) = 4 \end{array}$$

Opskriv sandsynlighedsfunktionen q for den transformerede fordeling $t(P)$.

Angiv en generel formel for sandsynlighedsfunktionen for en transformeret fordeling, udtrykt ved sandsynlighedsfunktionen for den oprindelige fordeling.

Opgave 1.4.2. For et slag med to terninger, opskriv sandsynlighedsfunktionen for fordelingen af det mindste af de to antal øjne (Z_1 i eksempel 1.4.1).

Opgave 1.4.3. En mønt kastes fire gange. Hvad er fordelingen af antallet af gange man får "krone"? (En fordeling angives normalt ved sin sandsynlighedsfunktion).

Opgave 1.4.4. Hvad er fordelingen af summen af øjnene i et kast med tre terninger? Tegn sandsynlighedsfunktionen (jvf. eksempel 1.2.1).

Opgave 1.4.5. Tre personer vælges tilfældigt. Hvad er fordelingen af antallet blandt disse, som er født på en søndag?

Opgave 1.4.6. Lad X være ligefordelt på E, og sæt $Y = t(X)$, hvor $t\colon E \to F$ er en given transformation. Under hvilke betingelser er Y også ligefordelt?