Kapittel 2

Hva er statistikk?

I statistikk beskjeftiger vi oss med observasjoner. Et sett observasjoner som er skaffet til veie på en spesifisert måte, utgjør et **utvalg**. Den sentrale oppgaven er å lære av erfaring ved å studere slike utvalg.

Det utvalget av observasjoner som vi studerer, er vanligvis bare en delmengde av en større mengde mulige (eller potensielle) observasjoner. **Populasjonen** er den samlingen av mulige observasjoner som utvalget er en del av. En populasjon kan være **endelig** (bestå av et endelig antall observasjoner) eller **uendelig** (bestå av et uendelig antall observasjoner). Antall observasjoner i en endelig populasjon blir kalt for **populasjonens størrelse**.

Det er vanligvis *populasjonen* vi prøver å lære noe om. Det synes derfor opplagt at vi vet hvilken populasjon vi ønsker å studere. Men det er også viktig at vi gjør oss opp en mening om hvilken populasjon det utvalget vi sitter med, *i virkeligheten* er hentet fra. Og det er ikke alltid like lett.

Eksempel 2.1 *Skjegg og fløyel i Bergen by*

Rannveig studerte ved Universitetet i Bergen i 1968. Hun ville finne ut hvor stor andel av mennene i byen som både hadde skjegg og gikk i fløyelsbukse. Derfor stilte hun seg opp på et gatehjørne og registrerte i 15 minutter skjegg- og buksetilstanden hos mennene som passerte: ja (både skjegg og fløyelsbukse), nei (ikke sant at både skjegg og fløyelsbukse), nei, ja, ja, nei, ja, ...

Risikerte Rannveig å ende opp med et galt svar?

Løsning: Ja. Rannveig ønsket å lære noe om populasjonen av svar på spørsmålet «Både skjegg og fløyelsbukse?» for samtlige menn i Bergen. Men hvis hun var uheldig med valg av gatehjørne og tidspunkt, kan de fleste forbipasserende ha vært studenter som kom fra et møte om marxismen-leninismens filosofiske grunnlag. Dette var i studentrevolusjonens, 68-ernes tid, og uniformen til de mannlige radikale studentene var skjegg, brun fløyelsbukse, grønn skulderveske og – om været tilsa det – grønn vindjakke. ▲

Generalisering av kunnskap fra et utvalg til en populasjon kaller vi **inferens**. Hensikten med en utvalgsundersøkelse er inferens til den populasjonen vi er interessert i. Men hvis utvalget i virkeligheten er fra en annen populasjon, er det denne siste vi lærer noe om. Hvis vi ikke vet at dette skjer, trekker vi enten gale konklusjoner, eller vi er «heldige» og trekker riktige konklusjoner på tross av metodefeil. Problemer oppstår hvis utvalget ikke er representativt for den populasjonen vi tror vi studerer. Vi sier i så fall at vi har **utvalgsskjevhet**.

Men hva betyr det egentlig at et utvalg er representativt? Vi kan si at et utvalg er **representativt** i den grad de egenskapene vi studerer, fordeler seg i utvalget på samme måte som i den populasjonen vi er interessert i. Og vi kan si at vi har utvalgsskjevhet i den grad det ikke forholder seg slik.

Eksempel 2.2 *Promillekjøring på Lanzarote*[1]

Politiet på Lanzarote gjorde 3061 alkotester i løpet av 2000. Tjue av bilførerne hadde *ikke* alkohol i blodet. Hvilken populasjon kan vi generalisere til?

Løsning: To populasjoner er av interesse. Den ene er populasjonen av alkoholkonsentrasjon i blodet hos bilførerne på Lanzarote. Den andre er populasjonen av alkoholkonsentrasjon i blodet hos den delmengden av bilførerne på Lanzarote som politiet finner grunn til å teste.

Vi skjønner at vi kan foreta inferens fra utvalget til den siste populasjonen. Men hvis vi ikke er oppmerksom på at to populasjoner er involvert, kan vi komme til å foreta inferens til den første og påstå: Under én prosent av bilførerne på Lanzarote kjører i edru tilstand. Dette blir selvfølgelig helt galt dersom politiet på Lanzarote kun tester bilførere som kjører mistenkelig. Hvis politiet ville estimere andelen edru bilførere på Lanzarote, kunne de gi alle bilførere samme sannsynlighet for å bli stoppet og testet.

Nå er nok politiet mer opptatt av trafikksikkerhet enn av forskning. Men hvis de ville forske, måtte de gjøre seg opp en mening om hva «gi alle bilførere samme sannsynlighet» skulle bety: Skulle alle med førerkort ha samme sannsynlighet for å bli testet? Skulle alle som på et gitt tidspunkt kjører bil, ha samme sannsynlighet? Det er mye å ta stilling til. ▲

Et spørsmål som melder seg etter å ha lest eksempel 2.2, er om vi kan påstå at andelen edru bilførere er lavere enn én prosent i den populasjonen politiet finner grunn til å teste. Andelen edru førere i *utvalget* er rett nok under en prosent, men rettferdiggjør dette inferens til populasjonen? Svaret avhenger – som alltid – blant annet av i hvilken grad utvalget er representativt for populasjonen.

Etter det vi nå har sagt, er det lett å tro at hovedmålet i statistikk er å skaffe seg representative utvalg. Slik er det også, men bare til en viss grad. For hvordan skal vi kunne vite om et utvalg er representativt? Vi kjenner jo ikke populasjonen og kan derfor ikke kontrollere ved å sammenlikne med den. Løsningen ligger i at vi i tillegg til å prøve å begrense dårlig representativitet også prøver å *kontrollere* den. Vi prøver å trekke våre utvalg på en slik måte at det går an å regne på sannsynligheten for ulike grader av avvik mellom fordelingen av de egenskapene vi er opptatt av i utvalgene og i de (ukjente) populasjonene som utvalgene skal representere. Ut fra disse beregningene vurderer vi så påliteligheten til de slutningene som vi trekker om populasjonene. Men for å kunne gjennomføre og forstå slike beregninger må vi beherske sannsynlighetsregning. Sannsynlighet og sannsynlighetsregning er tema for neste kapittel.

---

[1] VG Nett, 29. januar 2001, *20 av 3061 bilister var edru.*

Kapittel 3

Sannsynlighet

3.1 Deterministiske og stokastiske modeller

I teknisk utviklingsarbeid bruker en ofte matematiske modeller til å beskrive ulike sammenhenger. Matematiske modeller kan deles inn i to klasser, deterministiske og stokastiske.

1. En **deterministisk** modell beskriver sammenhenger som er entydige, uten uforutsigbar variasjon.

2. En **stokastisk**, eller **sannsynlighetsteoretisk**, modell beskriver sammenhenger som i vesentlig grad preges av uforutsigbar eller tilfeldig variasjon.

Det er åpenbart at vi søker en stokastisk modell når vi skal beskrive et fenomen hvor uforutsigbarhet og tilfeldig variasjon står sentralt. Men også når vi studerer prosesser som er deterministiske, kan den underliggende lovmessigheten være tildekket av tilfeldig variasjon som må behandles sannsynlighetsteoretisk.

3.2 Stokastiske forsøk

Eksempel 3.1 *Akslinger*

I faget *Maskindeler*[1] målte tre studentgrupper diameteren på 26 akslinger av samme slag. Målingene er gitt i tabell 3.1.

Tabellen viser at måleresultatene varierer både fra aksling til aksling og fra studentgruppe til studentgruppe. Et måleresultat avhenger blant annet av (1) egenskapene hos den aksling som måles, (2) egenskapene hos måleredskapet, (3) hvor på akslingen det måles, og (4) hvordan det måles. Forholdene (3) og (4) kan igjen henge sammen med (5) hvem som måler.

[1] Høgskolelektor L.P. Bryne, Høgskolen i Ålesund, *Laboratorieoppgave for maskiningeniørstudenter 1997*.

Eksemplet illustrerer at et måleresultat kan bli påvirket av tilfeldig variasjon. Et poeng er også at en ikke nødvendigvis er på jakt etter én «eneste rette» verdi for det som måles. Det interessante for en bruker av akslinger vil være i hvilken grad dimensjonene varierer ut over toleransegrensene for den aktuelle anvendelsen. ▲

Tabell 3.1 Tre gruppers målinger av akslingdiametere (mm).

Aksling nr.	Gruppe 1	Gruppe 2	Gruppe 3
1	24.981	24.980	24.980
2	24.981	24.978	24.972
3	24.969	24.968	24.956
4	24.975	24.976	24.974
5	24.968	24.967	24.975
6	24.973	24.971	24.976
7	24.951	24.949	24.977
8	24.963	24.970	24.959
9	24.974	24.974	24.971
10	24.980	24.976	24.974
11	24.975	24.976	24.976
12	24.956	24.954	24.967
13	24.972	24.974	24.971
14	24.972	24.995	24.969
15	24.980	24.983	24.980
16	24.957	24.956	24.957
17	24.963	24.965	24.968
18	24.966	24.966	24.961
19	24.957	24.959	24.955
20	24.960	24.963	24.956
21	24.968	24.978	24.967
22	24.984	24.984	24.976
23	24.976	24.991	24.979
24	24.994	24.975	24.974
25	24.978	24.998	24.977
26	24.958	24.991	24.964

Da studentene i eksempel 3.1 målte diameteren på en aksling, gjennomførte de et stokastisk forsøk. Et **stokastisk forsøk** har følgende egenskaper:

1 Forsøket kan gjentas under prinsipielt like betingelser så ofte vi ønsker.

2 Resultatet av hver gjennomføring av forsøket er uforutsigbart.

3 Til forsøket er det spesifisert en endelig eller uendelig mengde resultater som er slik at en gjennomføring av forsøket med sikkerhet vil gi étt og bare étt av disse resultatene. Elementene i denne mengden skal vi kalle **utfall**, og mengden skal vi kalle forsøkets **utfallsrom**.[2]

Vi kan ofte velge mellom ulike mulige utfallsrom for et forsøk.

Eksempel 3.2 *Terningkast*

Når du kaster en terning, blir mengden $\{1, 2, 3, 4, 5\}$ ubrukbar som utfallsrom fordi 6 ikke er med. Mengden $\{1, 2, 3, 4, 5, 6, \text{Partall}\}$ blir ubrukbar fordi det finnes kastresultat som svarer til mer enn étt opplistet utfall. Men mengden

$$S_1 = \{1, 3, 5, \text{Partall}\}$$

kan brukes som utfallsrom fordi hvert tenkelig kastresultat svarer til étt og bare étt opplistet utfall. Likevel melder spørsmål seg: Kan det være riktig å betrakte elementet *Partall* i S_1 som et utfall? Består ikke *Partall* i virkeligheten av de tre utfallene 2, 4 og 6? Forholdet er at vi i praksis *selv bestemmer* hva vi skal oppfatte som et utfall. Beslutningen tar vi i det øyeblikk vi spesifiserer utfallsrommet. Det betyr at 2, 4 og 6 ikke er utfall i S_1. Terningen kan selvfølgelig vise 2, men i så fall er det utfallet *Partall* som har inntruffet. Betrakt forsøkene

Forsøk A: Terningkast med utfallsrom $S_A = \{1, 3, 5, \text{Partall}\}$ og
Forsøk B: Terningkast med utfallsrom $S_B = \{1, 2, 3, 4, 5, 6\}$.

I forsøk A blir *Partall* et utfall fordi *Partall* er listet opp i forsøkets utfallsrom S_A.

I forsøk B blir resultatet *Partall* ikke et utfall. Grunnen er at *Partall* ikke er listet opp i utfallsrommet S_B. Utsagnet «Det ble et partall» signaliserer i forsøk B at vi fikk et av utfallene 2, 4 og 6. ▲

I forsøk B i eksempel 3.2 svarte *Partall* til delmengden $\{2, 4, 6\}$ av utfallsrommet. En slik delmengde blir kalt en *hendelse*. Vi definerer:

En **hendelse** i et stokastisk forsøk er en delmengde av forsøkets utfallsrom.

Vi sier at en hendelse **inntreffer** dersom et av utfallene i hendelsen inntreffer.

Vi sier at en samling hendelser er **gjensidig utelukkende**, **gjensidig ekskluderende** eller **disjunkte**, dersom høyst én av hendelsene kan inntreffe ved hver gjennomføring av forsøket.

[2] En begrensning av et forsøks utfallsrom til utfall som virkelig *kan* inntreffe, vil ofte være umulig: Hvis vi skal måle høyden til et tilfeldig menneske, må utfallsrommet inneholde alle høyder som kan framkomme. Men vi kjenner ikke nedre og øvre grense for mulige høyder.

Problemet kan løses ved å inkludere alle positive høyder i utfallsrommet. Men prisen for en slik løsning blir at utfallsrommet vil inneholde utfall som åpenbart aldri kan inntreffe.

For en samling $A_1, A_2, \dots, A_n$ av hendelser som kan inntreffe ved gjennomføringen av et stokastisk forsøk, gjelder følgende ekvivalenser:

Hendelsene $A_1, A_2, \dots, A_n$ er *gjensidig utelukkende.*

$\Leftrightarrow$ Det finnes ikke noe utfall av forsøket som er med i mer enn én av hendelsene $A_1, A_2, \dots, A_n$.

$\Leftrightarrow$ Ethvert par av hendelsene $A_1, A_2, \dots, A_n$ har tomt snitt, $A_i \cap A_j = \emptyset$ for $i \neq j$.

Oppgave 3.1 *Utfallsrom*

Vi skal kaste en terning. Kan følgende mengde være et utfallsrom?

(a) $S_1 = \{1, 2, 3, 4, 5, 6\}$?
(b) $S_2 = \{2, 3, 4, 6, \text{oddetall}\}$?
(c) $S_3 = \{2, 4, \text{oddetall}\}$?
(d) $S_4 = \{2, 4, 6, \text{oddetall}\}$?

Oppgave 3.2 *Utfall og hendelser*

Et stokastisk forsøk består i å kaste en terning og registrere om antall øyne blir 2, 4, 6 eller et oddetall. Utfallsrommet er altså $S = \{2, 4, 6, \text{oddetall}\}$. Blir da

(a) resultatet 4 et utfall?
(b) mengden $\{4\}$ en hendelse?
(c) mengden $\{2, 6\}$ et utfall?
(d) mengden $\{2, 6\}$ en hendelse?
(e) resultatet *oddetall* et utfall?
(f) resultatet *partall* et utfall?
(g) resultatet *partall* $= \{2, 4, 6\}$ en hendelse?
(h) utfallsrommet S en hendelse?
(i) den tomme mengden $\emptyset$ en hendelse?

Oppgave 3.3 *Gjensidig utelukkende hendelser*

Vi skal kaste en terning. Blir da hendelsene

(a) $\{1, 3\}$ og $\{2, 3, 6\}$ gjensidig utelukkende?
(b) $\{1, 3\}$ og $\{2, 4, 6\}$ gjensidig utelukkende?

3.3 Sannsynlighet

Det er vanskelig å definere begrepet sannsynlighet på en måte som både er begripelig for en leser uten bakgrunn i matematikk, og som samtidig holder mål som del av et fundament for en presis teori.

Den tilnærmingen til sannsynlighetsbegrepet som vi synes passer best i et ingeniørstudium, har vi funnet i Erling Sverdrups bok *Lov og Tilfeldighet I*, Universitetsforlaget 1964, sidene 7 og 8. Framstillingen i vårt avsnitt 3.3.1 bærer preg av lån fra Sverdrup.

3.3.1 Sannsynlighet som relativ frekvens

En ingeniør møter ofte situasjoner hvor en prosess blir gjentatt under tilnærmet identiske betingelser, men hvor resultatet like fullt varierer. Ofte er han ved hver gjentakelse interessert i om et spesielt fenomen blir observert. Dersom prosessen gjentas n ganger og fenomenet observeres i a av disse gjentakelsene, sier vi at fenomenet forekommer med **frekvens**[3] a og **relativ frekvens** a/n.

Eksempel 3.3 *Masseproduksjon*

Et firma produserer elektriske motstander. Ukontrollerte forhold i produksjonen fører til at resistansen til enkelte motstander faller utenfor spesifiserte grenser. Det har ikke vært mulig å oppdage signaler som kan være til hjelp i forsøk på å forutsi når defekte motstander vil bli produsert.

Motstandene inspiseres maskinelt og sorteres i to klasser, *intakte* og *defekte.* For å studere den relative frekvensen av defekte motstander vil kvalitetsingeniør Davy Aa benytte utvalgsundersøkelser. Han er usikker på hvor store utvalg han bør velge, og bestemmer seg for å prøve med ulike størrelser. Men før han sammenlikner konsekvensene av å bruke forskjellige realistiske utvalgsstørrelser, ønsker han å gjøre et lite stykke forskning: Hva skjer hvis utvalget blir veldig stort? Han innhenter tillatelse fra firmaets ledelse og setter i gang: Ti ganger trekker han et tilfeldig utvalg på 10 motstander, ti ganger et på 1000 motstander og ti ganger et på 100 000 motstander. Hver gang kontrollerer han samtlige motstander. Resultatet sammenfatter han i tabellen[4] nedenfor.

Trekning	Utvalgsstr. 10		Utvalgsstr. 1000		Utvalgsstr. 100 000	
nummer	Ant. def.	% def.	Ant. def.	% def.	Ant. def.	% def.
1	1	10	38	3.8	3462	3.46
2	0	0	36	3.6	3564	3.56
3	0	0	30	3.0	3451	3.45
4	0	0	33	3.3	3512	3.51
5	0	0	37	3.7	3561	3.56
6	0	0	35	3.5	3458	3.46
7	2	20	39	3.9	3520	3.52
8	0	0	35	3.5	3440	3.44
9	0	0	36	3.6	3467	3.47
10	0	0	44	4.4	3439	3.44

Se nøye på tallene i tabellens tredje, femte og sjuende kolonne: Variasjonen i prosentvis andel defekte ser ut til å avta med økende utvalgsstørrelse. Aa gjetter på at en fortsatt økning av utvalgsstørrelsen vil resultere i en stabilisering av defektandelen i nærheten av 3.5 %.

Men er denne tendensen til stabilisering ved økende utvalgsstørrelse repeterbar? Aa gjentar forsøket og sammenfatter resultatene i tabellen som følger. Når

[3] Noen bruker ordet *hyppighet* i stedet for ordet *frekvens*.

[4] I denne tabellen er «Ant. def.» en forkortelse for «Antall defekte» og «% def.» en forkortelse for «Prosentvis andel defekte».

du studerer den, vil du se at det siste forsøket viser samme tendens som det første.

Trekning nummer	Utvalgsstr. 10 Ant. def.	% def.	Utvalgsstr. 1000 Ant. def.	% def.	Utvalgsstr. 100 000 Ant. def.	% def.
1	0	0	37	3.7	3528	3.53
2	0	0	36	3.6	3493	3.49
3	1	10	45	4.5	3533	3.53
4	0	0	41	4.1	3606	3.61
5	2	20	44	4.4	3556	3.56
6	0	0	31	3.1	3494	3.49
7	0	0	27	2.7	3579	3.58
8	0	0	33	3.3	3439	3.44
9	1	10	29	2.9	3438	3.44
10	1	10	26	2.6	3549	3.55

▲

Dersom vi over tid studerer når defekte enheter dukker opp i en produksjonsprosess, vil det enten

1 være mulig å finne signaler som kan være til hjelp ved vurdering av om neste produserte enhet vil bli defekt, eller det vil

2 ikke være mulig å finne slike signaler.

En kan i virkelige produksjonsprosesser ofte oppleve situasjoner som beskrevet under punkt 1. Defekte enheter kan for eksempel hope seg opp fordi justeringer av produksjonsutstyr er kommet ut av lage. Vi sier at produksjonsprosessen er **ute av kontroll**.

Dersom en på den annen side er fri for årsaker til systematisk forekomst av defekte enheter, er situasjonen som beskrevet under punkt 2. Vi sier da at produksjonsprosessen er **under kontroll**. I så fall vil det først være når vi deler et stort antall observasjoner opp i grupper og beregner den relative frekvensen av defekte i hver gruppe, at vi oppdager at

- denne relative frekvensen er rimelig stabil, og at
- stabiliteten blir tydeligere til større vi gjør gruppene.

Det er nærliggende å tro at den omtalte stabiliteten avdekker en underliggende egenskap ved produksjonsprosessen, og at den relative frekvensen vil nærme seg en entydig bestemt grenseverdi etter som utvalgsstørrelsen vokser. Dersom det forholder seg slik, er det denne grenseverdien som vi intuitivt skal forestille oss som **sannsynligheten** for at en produsert enhet vil vise seg å være defekt. Vi

Sannsynlighet som relativ frekvens

Anta at en hendelse kan inntreffe ved gjennomføringen av et forsøk som kan gjentas et stort antall ganger under tilnærmet like betingelser. Da skal vi forestille oss hendelsens **tilnærmede sannsynlighet** som hendelsens **relative frekvens** i løpet av et stort antall slike forsøk.

Sannsynlighetens nøyaktige verdi skal vi forestille oss som den grenseverdien hendelsens relative frekvens ville nærme seg dersom vi kunne la antall gjentakelser av forsøket gå mot uendelig.

3.3.2 Sannsynlighet ved symmetribetraktning

Iblant kan en sannsynlighet bestemmes ved teoretiske betraktninger alene. en skal gi eksempler på dette, er det vanlig å snakke om såkalte *rettfe* mynter og terninger. En **rettferdig mynt** er en mynt som er tosidet, som kan bli stående på høykant etter et kast og hvor hver av de to sidene har sa sannsynlighet for å vende opp etter et kast. En **rettferdig n-sidet ter** er en «terning» med n sider og hvor hver av sidene har samme sannsynli for å bli resultatet av et kast. Begrepet «n-sidet terning» er en generalis av det vanlige begrepet «terning», som betegner en kubisk formet seks spillebrikke hvor sidene er merket fra 1 til 6. Når vi ikke presiserer antall på en terning, går vi ut fra at den er sekssidet og tradisjonelt merket.

Eksempel 3.4 *Terningkast med rettferdig terning*

Anta at vi har en rettferdig terning. Hva blir sannsynligheten for at et kast vise (a) fem øyne, (b) et partallig antall øyne?

Svar på (a): En rettferdig terning har ingen hukommelse som gjør at resultatene vil variere systematisk. Samtidig oppfører den seg symmetrisk i forstand at de ulike sidene i det lange løp vil vende opp i en like stor and kastene.

Med seks terningsider følger det at den andelen ganger hver side har v opp, bør nærme seg 1/6 når antall kast går mot uendelig. Sannsynlighete fem øyne blir derfor 1/6.

Svar på (b): I det lange løp vil vi få to øyne i 1/6 av kastene, fire i 1/ kastene og seks i 1/6 av kastene. Andel kast hvor hendelsen *Partall* inntr må derfor bli $1/6 + 1/6 + 1/6 = 3/6 = 1/2$. Siden sannsynlighet er lik re frekvens i det lange løp, følger det at sannsynligheten for hendelsen *Partal* 1/2.

Oppgave 3.4 *Kortlek* 1

En godt blandet kortstokk inneholder 26 røde og 26 svarte kort.
(a) Hva blir sannsynligheten for å ende opp med rød farge, dersom du tre et kort tilfeldig?

(b) Et tilfeldig trukket kort legges på bordet med fargesiden ned – uten at du har sett fargesiden. Du blir deretter bedt om å angi sannsynligheten for at kortet er rødt. Hva bør du svare?

Oppgave 3.5 *Kortlek 2*

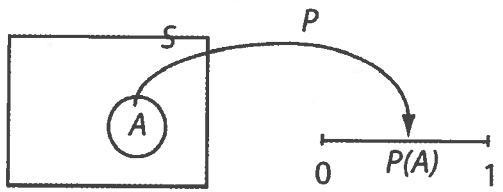
En godt blandet kortstokk inneholder totalt 51 kort, deriblant ett av typen sparess. (Ett kort er kommet bort.)
(a) Bestem sannsynligheten for å få sparess, hvis du trekker et kort tilfeldig.
(b) Et tilfeldig trukket kort legges på bordet med fargesiden ned – uten at du har sett fargesiden. Du blir deretter bedt om å angi sannsynligheten for at det er sparess som ligger der. Hva bør du svare?
(c) Et tilfeldig trukket kort legges på bordet med fargesiden ned – uten at du har sett fargesiden. En medstudent vil vedde hundre kroner på at det er et annet kort enn sparess som ligger der, og spør om du vil vedde ti kroner mot. Er tilbudet godt for deg?

3.3.3 Sannsynlighet som funksjon

Hvis A er en hendelse, lar vi $P(A)$ betegne sannsynligheten for at A skal inntreffe.[5] I tråd med vår oppfatning av sannsynlighet som relativ frekvens i det lange løp, kan vi da betrakte sannsynligheten P som en entydig avbildning

$$A \xrightarrow{P} P(A)$$

av hendelsene i utfallsrommet på en delmengde av det lukkede intervallet $[0, 1]$. Se figur 3.1. Vi kan altså oppfatte sannsynlighet som en funksjon P med definisjonsmengde lik mengden av alle hendelser i utfallsrommet.



Figur 3.1 Sannsynligheten P som avbildning av hendelsen A.

3.3.4 Matematisk definisjon av sannsynlighet

Dersom S er utfallsrommet til et stokastisk forsøk og sannsynlighet følger vanlige regneregler for andeler, vil sannsynlighetsfunksjonen P tilfredsstille følgende tre krav:

1 $0 \leq P(A) \leq 1$ for enhver hendelse $A \subseteq S$.

[5]Bokstaven P kommer fra det engelske ordet *probability*, som betyr sannsynlighet.

2 $P(S) = 1$.

3 **Addisjonsregelen for gjensidig utelukkende hendelser**
Hvis $A_1, A_2, \ldots$ er en endelig eller uendelig sekvens av gjensidig utelukkende hendelser, vil

$$\boxed{P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots.} \tag{3.1}$$

Vi skal heretter gå ut fra at sannsynlighet er en funksjon som alltid tilfredsstiller disse tre kravene.

I enkelte framstillinger blir vårt krav 1 erstattet av det svakere kravet $1^*: 0 \leq P(A)$. Denne forenklingen er mulig fordi utsagnet $P(A) \leq 1$ kan utledes av de tre kravene 1*, 2 og 3.

I matematisk statistikk blir sannsynlighet gjerne *definert* som en funksjon som tilfredsstiller kravene 1*, 2 og 3. Disse tre kravene blir kalt **Kolmogorovs aksiomer**.

3.4 Sammensatte hendelser

Før vi bruker addisjonsregelen for gjensidig utelukkende hendelser, må vi forsikre oss om at de involverte hendelsene virkelig *er* gjensidig utelukkende. For glemmer vi det, kan det fort bære galt av sted. Følgende eksempel illustrerer dette.

Eksempel 3.5 *Minst en femmer på sju terningkast*

Du kaster en rettferdig terning sju ganger. Hva blir sannsynligheten for å få fem øyne minst en gang?

Løsning: I første kast blir sannsynligheten for å få fem øyne lik $1/6$. Det samme blir sannsynligheten for å få fem øyne i andre, tredje, ..., sjuende kast. Siden hendelsene «Fem øyne i første kast», «Fem øyne i andre kast» osv. ikke gjensidig utelukker hverandre, blir det galt å beregne sannsynligheten for å få fem øyne minst en gang i løpet av de sju kastene som

$$\text{NB: Gal beregning: } \underbrace{1/6 + 1/6 + \cdots + 1/6}_{\text{sju ganger}} = 7/6 > 1.$$

Hvordan du på rett måte kan bestemme denne sannsynligheten, skal vi vise i eksempel 3.13 på side 40. ▲

Subtraksjonsregelen

$$\boxed{P(A \backslash B) = P(A) - P(A \cap B).} \tag{3.2}$$

(e) mottatte signal 0 skal være rett overført,
(f) sendte signal skal bli rett overført,
(g) mottatte signal skal være rett overført.

Oppgave 3.52 *Mer linjestøy*

Denne oppgaven dreier seg om sending av signaler over samme linje som ble omtalt i oppgave 3.51. Opplysninger som er gitt der, gjelder fremdeles.

Et dataprogram skal sendes som en fil med totalt 431 498 tegn av type 0 og 801 177 tegn av type 1. Bestem sannsynligheten for at programmet skal bli rett mottatt.

Kapittel 4

Stokastiske variabler

Ingeniøren ønsker å måle, sammenlikne, vurdere og oppsummere ved hjelp av tall. Observasjoner som i utgangspunktet ikke har form av tall, oversetter han ofte til tall – slik at han kan anvende metoder som er utviklet for å håndtere tall. Til slik oversettelse bruker vi i statistikk noen spesielle funksjoner som blir kalt stokastiske variabler.

En **stokastisk variabel** er en funksjon som avbilder hvert av et forsøks mulige utfall på et reelt tall. Dersom et stokastisk forsøk gir et utfall som av en stokastisk variabel avbildes på det reelle tallet x, sier vi at x er en **realisert verdi** av den stokastiske variabelen. I stedet for «realisert verdi» nyttes ofte begrepet **observert verdi**.

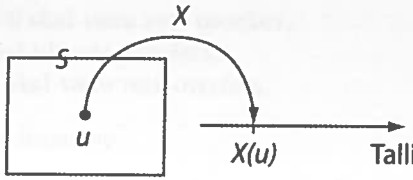
La oss nå betrakte et stokastisk forsøk med utfallsrom S. Hvis en stokastisk variabel X avbilder et utfall u i S på det reelle tallet $X(u)$, er det vanlig å skrive

$$X = X(u). \tag{4.1}$$

I likning (4.1) oppfatter vi X på høyre side av likhetstegnet som den funksjonen vi i vår definisjon gav navnet stokastisk variabel. Funksjonen X er tegnet inn over den krumme avbildningspilen på figur 4.1. Funksjonsverdien $X(u)$ på høyre side av likhetstegnet i (4.1) er det reelle tallet som funksjonen X avbilder utfallet u på. Symbolet X på venstre side av likhetstegnet oppfatter vi som en variabel. Denne variabelen blir tilordnet verdien $X(u)$, som på figur 4.1 er tegnet inn under tallinja til høyre. Vi ser at når X er en stokastisk variabel, blir symbolet «X» brukt i dobbel betydning, både (1) som variabel, på venstre side i (4.1), og (2) som funksjon, på høyre side i (4.1).

Stokastiske variabler blir betegnet med store bokstaver, ofte X eller Y, mens realiserte verdier av disse variablene blir betegnet med tilsvarende små bokstaver (ofte x eller y). De realiserte verdiene blir betraktet som konstanter. Andre små bokstaver, som a, b og c, blir også brukt til å representere reelle tall som den stokastiske variabelen kan eller eventuelt ikke kan bli lik.

Dersom Y er en stokastisk variabel og y er et reelt tall, skal vi med hendelsen $\{Y \leq y\}$ mene mengden av alle utfall som fører til at Y får en verdi mindre enn



Figur 4.1 Stokastisk variabel.

eller lik y. Vi definerer altså hendelsen $\{Y \leq y\}$ ved

$$\{Y \leq y\} = \{u \in S \colon Y(u) \leq y\}.$$

Hendelsen $\{Y \leq y\}$ skal vi lese «hendelsen at Y er mindre enn eller lik y», eller enklere «hendelsen at Y er høyst lik y».

På samme måte definerer vi «hendelsen at Y er større enn a og høyst lik b» ved

$$\{a < Y \leq b\} = \{u \in S \colon a < Y(u) \leq b\}.$$

I praksis vil vi ofte utelate parentesen i uttrykk som $\{a < Y \leq b\}$ og i stedet skrive «hendelsen $a < Y \leq b$».

For en stokastisk variabel X definerer vi **fordelingsfunksjonen** F_X ved

$$F_X(x) = P(X \leq x) \quad \text{for } x \in R.$$

Når det ikke skaper fare for misforståelser, utelater vi indeksen og skriver F for fordelingsfunksjonen.

En fordelingsfunksjon er en voksende ikke-negativ funksjon som er definert på mengden av reelle tall, som nærmer seg 0 når argumentet[1] går mot $-\infty$ og som nærmer seg 1 når argumentet nærmer seg ∞. En fordelingsfunksjon F må altså tilfredsstille kravene

$$\boxed{\begin{array}{lrcll} (1) & F(x) & \geq & 0 & \text{for alle } x, \\ (2) & F(a) & \leq & F(b) & \text{for } a < b, \\ (3) & \lim\limits_{x \to -\infty} F(x) & = & 0, & \\ (4) & \lim\limits_{x \to \infty} F(x) & = & 1. & \end{array}}$$

For en stokastisk variabel X og reelle tall a og b hvor $a \leq b$, gjelder alltid

$$\boxed{\begin{aligned} P(X > a) &= 1 - P(X \leq a) = 1 - F(a), \\ P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a). \end{aligned}}$$

[1] Husk: Hvis F er en funksjon av den frie variabelen x, blir x kalt for funksjonens argument. Funksjonens argument må tilhøre funksjonens definisjonsmengde. I vårt tilfelle: $x \in D_F = R$.

Begrunnelse: Siden $\{X > a\} = \{X \leq a\}^c$, følger det ved komplementregelen at

$$P(X > a) = P(\{X \leq a\}^c) = 1 - P(X \leq a).$$

Tilsvarende følger det ved subtraksjonsregelen at

$$\begin{aligned} P(a < X \leq b) &= P(\{X \leq b\} \setminus \{X \leq a\}) \\ &= P(X \leq b) - P(\{X \leq a\} \cap \{X \leq b\}) \\ &= P(X \leq b) - P(X \leq a), \end{aligned}$$

hvor vi i siste overgang har brukt at $\{X \leq a\} \cap \{X \leq b\} = \{X \leq a\}$ når $a < b$. ■

Oppgave 4.1 *Regeltrim* 4

Bestem

(a) $P(X > 7)$ når $P(X \leq 7) = 0.83$,
(b) $P(7 < X \leq 8)$ når $P(X \leq 7) = 0.83$ og $P(X \leq 8) = 0.95$,
(c) $P(X \leq 9)$ når $P(3 < X \leq 9) = 0.33$ og $P(X \leq 3) = 0.15$,
(d) $P(X \leq 3)$ når $P(3 < X \leq 9) = 0.32$ og $P(X \leq 9) = 0.41$.

4.1 Diskrete stokastiske variabler

En mengde er *tellbar* hvis elementene i mengden kan nummereres. I statistikk og sannsynlighetsregning pleier vi å bruke ordet **diskret** i stedet for ordet tellbar. En **diskret mengde** er altså det samme som en tellbar mengde.

En stokastisk variabel blir kalt **diskret** dersom mengden av de verdier den kan anta, er tellbar.[2]

For en diskret stokastisk variabel X definerer vi **punktsannsynligheten**[3] p_X ved

$$p_X(x) = P(X = x) = P(\{u \in S \colon X(u) = x\}).$$

Punktsannsynligheten $p_X(x)$ er altså sannsynligheten for å få et utfall som gir den diskrete stokastiske variabelen X verdien x. Dersom X aldri kan få verdien x, blir $p_X(x) = 0$.

Når det er opplagt hvilken stokastisk variabel vi arbeider med, sløyfer vi indeksen og skriver p i stedet for p_X.

For punktsannsynligheten p til en diskret stokastisk variabel X gjelder:

$$0 \leq p(x) \leq 1 \text{ for ethvert reelt tall } x,$$

$$\textstyle\sum_{\text{alle } x} p(x) = 1,$$

$$F(x) = P(X \leq x) = \textstyle\sum_{y \leq x} p(y) \text{ og}$$

$$P(a < X \leq b) = \textstyle\sum_{a < y \leq b} p(y) = F(b) - F(a).$$

[2] Husk at en tellbar mengde kan være uendelig. Se eventuelt avsnitt 1.5 *Mengdelære* for en repetisjon.

[3] I enkelte lærebøker blir punktsannsynligheten til en diskret stokastisk variabel kalt variabelens *sannsynlighetstetthet*.

I oppsettet ovenfor betyr summasjon over «alle x» at vi summerer over alle de verdiene X kan anta. Tilsvarende betyr summasjonene over $y \leq x$ og $a < y \leq b$ at vi summerer over de verdiene som X kan anta i de aktuelle verdiområdene.

Med **sannsynlighetsfordelingen** til en diskret stokastisk variabel mener vi en forskrift som angir enten punktsannsynligheten eller verdien til fordelingsfunksjonen for enhver verdi variabelen kan anta.

Eksempel 4.1 *Terningkast*

Den stokastiske variabelen X får som verdi det tall som angir antall øyne ved kast med en rettferdig terning.
(a) Spesifiser punktsannsynligheten p.
(b) Spesifiser fordelingsfunksjonen F.
(c) Bestem sannsynlighetene $P(2 < X < 3)$, $P(3 < X \leq 5)$ og $P(\text{Partall})$.

Løsning: I løsningen som følger, må du huske at x representerer et reelt tall, mens X er en stokastisk variabel.
(a) Punktsannsynligheten p er gitt ved

$$\begin{aligned} p(1) &= P(X=1) = 1/6, \\ p(2) &= P(X=2) = 1/6, \\ &\vdots \\ p(6) &= P(X=6) = 1/6, \\ p(x) &= 0 \text{ for } x \notin \{1, 2, 3, 4, 5, 6\}. \end{aligned}$$

(b) Fordelingsfunksjonen F er gitt ved

$$\begin{aligned} F(x) &= 0 && \text{for } x < 1, \\ F(x) &= p(1) = 1/6 && \text{for } 1 \leq x < 2, \\ F(x) &= p(1) + p(2) = 1/3 && \text{for } 2 \leq x < 3, \\ &\vdots \\ F(x) &= p(1) + p(2) + \cdots + p(6) = 1 \text{ for } x \geq 6. \end{aligned}$$

(c) Sannsynlighetene blir

$$\begin{aligned} P(2 < X < 3) &= 0, \\ P(3 < X \leq 5) &= p(4) + p(5) = 1/6 + 1/6 = 1/3, \\ P(\text{Partall}) &= p(2) + p(4) + p(6) \\ &= 1/6 + 1/6 + 1/6 = 1/2. \end{aligned}$$

▲

Oppgave 4.2 *Rettferdig mynt*

Sidene på en rettferdig mynt er merket henholdsvis 1 og 2. Betrakt et stokastisk forsøk hvor utfallet er tallet på den siden som vender opp etter et kast med denne mynten. Spesifiser
(a) punktsannsynligheten p,
(b) fordelingsfunksjonen F.

Oppgave 4.3 *Firesidet terning*

Utfallet i et stokastisk forsøk er antall øyne X som er skjult etter et kast med en rettferdig firesidet terning, hvor sidene er merket henholdsvis 1, 2, 3 og 4.
(a) Spesifiser punktsannsynligheten p.
(b) Spesifiser fordelingsfunksjonen F.
(c) Bestem sannsynlighetene $P(X=3)$, $P(1 \leq X < 3)$ og $P(\text{Oddetall})$.

Oppgave 4.4 *Uvanlig terning*

Sidene på en sekssidet terning er merket 1, 2, 3, 4, 5 og 6. Ved kast med denne terningen er $P(\text{Partall}) = 1/2$ og $p(1) = p(3) = 1/10$. Bestem $p(5)$.

Oppgave 4.5 *Fikset mynt*

En mynt er fikset, slik at sannsynligheten for krone ved et tilfeldig kast er 50 % større enn sannsynligheten for mynt. Bestem sannsynligheten for at resultatet av kastet skal bli mynt, dersom (a) mynten ikke kan bli stående på høykant, (b) sannsynligheten er 5 % for at mynten skal bli stående på høykant.

4.2 Kontinuerlige stokastiske variabler

Vi sier at en stokastisk variabel X med fordelingsfunksjon[4] F er **kontinuerlig**, dersom

1. fordelingsfunksjonen F er kontinuerlig,
2. **sannsynlighetstettheten** til X, definert ved $f(x) = \frac{d}{dx}F(x)$, er definert for alle eller alle unntatt et endelig antall verdier av x,
3. sannsynlighetstettheten f er stykkevis kontinuerlig.

Hvis de tre kravene ovenfor er tilfredsstilt, vil også

4. $f(x) \geq 0$ (fordi f er den deriverte av den ikke-avtagende funksjonen F),
5. $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)\, dx$,
6. $P(X \leq b) = F(b) = \int_{-\infty}^b f(x)\, dx$,
7. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

I en grafisk framstilling vil $F(a)$ framstå som arealet under f-grafen til venstre for den vertikale linja $x = a$. Se figur 4.2.

[4] Dersom flere stokastiske variabler er involvert, for eksempel X og Y, skiller vi mellom de respektive fordelingsfunksjonene ved å merke dem med tilsvarende indekser, for eksempel F_X og F_Y.

Forventningsverdien til en kontinuerlig stokastisk variabel X med sannsynlighetstetthet f er definert ved

$$\boxed{\mu_X = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx.} \tag{5.2}$$

Når det er opplagt hvilken stokastisk variabel vi tenker på, skriver vi μ i stedet for μ_X. Legg merke til følgende:

> Forventningsverdien er *ikke* en verdi vi forventer å få realisert ved gjennomføringen av ett enkelt forsøk.

Forventningsverdi kalles ofte *aritmetisk middelverdi*, eller bare *middelverdi*.

Eksempel 5.1 *Terningkast*

La X være en stokastisk variabel som angir antall øyne ved kast med en rettferdig terning. Bestem forventningsverdien μ.

Løsning: Forventningsverdien blir

$$\begin{aligned}\mu &= E(X) = \sum_{\text{alle } x} xp(x) \\ &= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5.\end{aligned}$$

Denne forventningsverdien kan *aldri* bli realisert ved et kast. ▲

Oppgave 5.1 *Fusketerning*

Ved kast med en spesiell fusketerning er sannsynligheten for å få seks øyne dobbelt så stor som sannsynligheten for å få ett øye. Sannsynlighetene for å få to, tre, fire og fem øyne er alle lik en seksdel. Bestem sannsynligheten for å få seks øyne og forventningsverdien for antall øyne ved kast med denne terningen.

5.2 Spredning, varians og standardavvik

Forventningsverdien til en stokastisk variabel uttrykker tilnærmet gjennomsnittet av variabelens verdier i en lang forsøksserie. Hvis variabelen svinger med store utslag rundt denne gjennomsnittsverdien, har vi **stor spredning**. Hvis variabelen svinger med små utslag, har vi **liten spredning**. De vanligste målene for spredning er *varians* og *standardavvik*.

Differansen $x - \mu$ mellom en observert verdi av X og forventningsverdien til X blir kalt observasjonens **avvik fra middelverdien**, mens $(x-\mu)^2$ kalles observasjonens **kvadratavvik fra middelverdien**.

Variansen til en stokastisk variabel er definert som forventningsverdien til variabelens kvadratavvik fra middelverdien:

$$\boxed{\sigma_X^2 = E\left[(X-\mu_X)^2\right] = \sum_{\text{alle } x} (x-\mu_X)^2\, p(x), \quad \text{hvis } X \text{ er diskret,}} \tag{5.3}$$

$$\boxed{\sigma_X^2 = E\left[(X-\mu_X)^2\right] = \int_{-\infty}^{\infty} (x-\mu_X)^2 f(x)\,dx, \quad \text{hvis } X \text{ er kontinuerlig.}} \tag{5.4}$$

Standardavviket til en stokastisk variabel er definert som kvadratroten av variabelens varians,

$$\text{standardavvik} = \sqrt{\text{varians}}. \tag{5.5}$$

Standardavviket til en stokastisk variabel X betegnes σ_X, eller bare σ dersom det er opplagt hvilken variabel vi tenker på. Variansen betegnes tilsvarende[1] σ_X^2 eller σ^2.

Eksempel 5.2 *Plotter*

Svein Mørk er en av tre skipsdesignere som er koblet opp mot samme plotter. Han klager over at plotteren nesten alltid er opptatt når han ber om utskrift. Den dataansvarlige vet at etterspørselen etter plotteren er gitt ved sannsynlig- Den dataansvarlige vet at etterspørselen etter plotteren er gitt ved sannsynlighetsfordelingen nedenfor. Her angir variabelen X antall designere som etterspø[r] plottertjenesten på et tilfeldig tidspunkt.

Etterspørsel, x	0	1	2	3
Sannsynlighet, $p(x)$	0.82	0.17	0.01	0.00

Beregn forventningsverdi, varians og standardavvik for X.

Løsning: Ved (5.1), (5.3) og (5.5) får vi

$$\begin{aligned}\mu &= \textstyle\sum_{\text{alle } x} xp(x) \\ &= 0 \times 0.82 + 1 \times 0.17 + \cdots + 3 \times 0.00 = 0.19, \\ \sigma^2 &= \textstyle\sum_{\text{alle } x} (x-\mu)^2\, p(x) \\ &= (0-0.19)^2 \times 0.82 + \cdots + (3-0.19)^2 \times 0.00 \\ &= 0.1739 = 0.17, \\ \sigma &= \sqrt{\sigma^2} = \sqrt{0.1739} = 0.42.\end{aligned}$$

[1] Noen skriver $Var(X)$ eller $VAR(X)$ for variansen og $Std(X)$ eller $STD(X)$ for standar[d]avviket.

Resten av oppgaven kan du løse når du repeterer, etter du har lært om diskrete sannsynlighetsfordelinger. Anta at alle de 10 000 misfarvede kulene ligger i en stor beholder. Bestem sannsynligheten for å få det utvalget som er spesifisert før deloppgave (d) ved å trekke kulene tilfeldig fra beholderen

(f) med tilbakelegging mellom hver trekning,

(g) uten tilbakelegging mellom hver trekning.

(h) Sammenlikn svarene i (f) og (g). Forklar det du ser.

Oppgave 6.20 *Ti lyspærer*

I en eske ligger det fire lyspærer av høy kvalitet og seks av lav. En pære av høy kvalitet har 70 % sannsynlighet for å lyse etter 9000 brukstimer, mens tilsvarende sannsynlighet for en pære av lav kvalitet er 30 %. Bestem sannsynligheten for at (a) en tilfeldig pære fra eska vil lyse etter 9000 brukstimer, (b) to tilfeldige pærer fra eska begge vil lyse etter 9000 brukstimer.

Kapittel 7

Diskrete fordelinger

Vi husker at en *diskret sannsynlighetsfordeling* angir sannsynligheten for at en diskret stokastisk variabel skal anta hver av sine mulige verdier. I dette kapitlet skal vi behandle:

- diskret uniform fordeling
- hypergeometrisk fordeling
- multivariat hypergeometrisk fordeling
- bernoullifordeling
- binomisk fordeling
- multinomisk fordeling
- poissonfordeling

7.1 Diskret uniform fordeling

En stokastisk variabel X er **diskret uniformt fordelt** dersom

1 det antall verdier X kan anta, er endelig, og

2 det er lik sannsynlighet for at X skal anta hver av disse verdiene.

Dersom de verdiene X kan anta, er

$$x_1, x_2, \ldots, x_n,$$

vil altså

$$p(x_1) = p(x_2) = \cdots = p(x_n). \tag{7.1}$$

Siden X er en funksjon og dermed en entydig avbildning, vil $x \neq y$ medføre at hendelsene $X = x$ og $X = y$ blir gjensidig utelukkende. All den stund X ve en gjennomføring av det aktuelle stokastiske forsøket er nødt til å anta en av sine mulige verdier, følger det ved addisjonssetningen for gjensidig utelukkende hendelser[1] at

$$P(X \in \{x_1, x_2, \ldots, x_n\}) = p(x_1) + \cdots + p(x_n) = 1. \tag{7.2}$$

Av (7.1) og (7.2) følger at den uniformt fordelte variabelen X får **punktsannsynlighet**

$$p(x_1) = p(x_2) = \cdots = p(x_n) = 1/n. \tag{7.3}$$

Eksempler på diskrete sannsynlighetsfordelinger dukker opp når vi kaster rettferdige mynter og terninger, og når vi foretar rettferdig trekning av kuler eller lapper fra en urne. I praktisk ingeniørarbeid kan problemstillinger ofte analyseres på nøyaktig samme måte som mynt-, terning- og urneproblemer.

Eksempel 7.1 *Rettferdig mynt*

Betrakt en rettferdig mynt som i hvert kast gir et av utfallene *mynt* eller *krone*. At mynten er rettferdig, betyr at sannsynligheten er den samme for hvert av de to utfallene, slik at $P(\text{mynt}) = P(\text{krone}) = 1/2$. Hvis vi definerer en diskret stokastisk variabel X ved $X(\text{mynt}) = 0$ og $X(\text{krone}) = 1$, følger det at $P(X = 0) = P(X = 1) = 1/2$, slik at X er diskret uniformt fordelt.

Vi får også en diskret uniform stokastisk variabel hvis vi definerer $X(\text{mynt}) = -3$ og $X(\text{krone}) = 117$. Det avgjørende er ikke hvilken verdi vi gir $X(\text{mynt})$ og $X(\text{krone})$, men at de to kravene i definisjonen av en diskret uniformt fordelt stokastisk variabel er tilfredsstilt. ▲

Eksempel 7.2 *Rettferdig terning*

Ved kast med en rettferdig terning er det naturlig å definere en stokastisk variabel X ved

$$X(\text{kastresultat}) = \text{antall øyne som terningen viser.}$$

Kravene til en diskret uniformt fordelt stokastisk variabel er tilfredsstilt. ▲

Eksempel 7.3 *Plasmaskjermer*

Vi betrakter produksjonen av en plasmaskjerm som et forsøk med utfallsrom $S = \{\text{intakt, defekt}\}$ og definerer den stokastiske variabelen X ved

$$X(\text{intakt}) = 0 \text{ og } X(\text{defekt}) = 1.$$

[1] For definisjon av *gjensidig utelukkende hendelser*, se side 1.5.

Punktsannsynligheten p vil da tilfredsstille kravet

$$p(0) + p(1) = P(\text{intakt}) + P(\text{defekt}) = 1.$$

Dersom $P(\text{intakt}) \neq P(\text{defekt})$, slik at $p(0) \neq p(1)$, vil X *ikke* være uniformt fordelt. ▲

Av eksempel 7.3 lærer vi:

> Ikke alle diskrete modeller med endelig utfallsrom er uniforme.

7.1.1 Sannsynlighet som «gunstige delt på mulige»

La X være en diskret uniformt fordelt stokastisk variabel med m mulige utfall, og la A være en hendelse i det utfallsrommet som er definisjonsmengde for X. Da sier vi at x er en **gunstig** X-verdi for hendelsen A dersom $x = X(u)$ for et utfall $u \in A$. Hvis vi lar $x_1, x_2, \ldots, x_g$ være de gunstige x-verdiene for A, følger det av (7.3) at

$$\begin{aligned} P(A) &= p(x_1) + p(x_2) + \cdots + p(x_g) \\ &= \underbrace{\frac{1}{m} + \frac{1}{m} + \cdots + \frac{1}{m}}_{g \text{ ganger}} = \frac{g}{m}. \end{aligned}$$

Eksempel 7.4 *Partall i kast med rettferdig terning*

Du skal kaste en rettferdig terning. Bestem sannsynligheten for at antall øyne skal bli et partall.

Løsning: Vi får

$$P(\text{Partall}) = \frac{g}{m} = \frac{3}{6} = \frac{1}{2}.$$

▲

Eksempel 7.5 *To hundre akslinger*

I et parti på 200 akslinger finnes det eksakt fire som ikke holder angitt minstekvalitet. Bestem sannsynligheten for at en tilfeldig aksling fra partiet skal holde angitt minstekvalitet.

Løsning: Siden hver aksling i partiet har samme sannsynlighet for å bli trukket ut, må sannsynligheten for å trekke en som holder angitt minstekvalitet, bli lik andelen slike akslinger i partiet, altså

$$P(\text{Holder angitt minstekvalitet}) = \frac{g}{m} = \frac{200-4}{200} = \frac{196}{200} = 0.98.$$

▲