

**KOMPENDIUM FOR
FORKURS I MATEMATIKK
FOR
MASTERSTUDIET I
INFORMASJONSSIKKERHET
VED HØGSKOLEN I GJØVIK
SOMMEREN 2004**

av Hans Engenes

18. august 2004

Innhold

1	Tallteori	3
1.1	Innledning	3
1.2	Heltall, heltallsdivisjon og restdivisjon	4
1.3	Divisorer, perfekte tall og primtall	5
1.4	Største felles divisor, Euklids algoritme og Aritmetikkens Fundamentalsetning	7
1.5	Kongruens og regning modulo m	14
2	Algebraiske strukturer	20
2.1	Innledning	20
2.2	Grupper	21
2.3	Ringer	25
2.4	Kropper	30
2.5	Polynomringer	31
3	Sannsynlighetsregning	34
3.1	Innledning	34
3.2	Sannsynlighetsmodeller	34
3.3	Betinget sannsynlighet og uavhengighet	37
3.4	Stokastiske variable	39
3.5	Binomiske fordelinger og normalfordelinger	43
4	Kombinatorikk	49
5	Elementær mengdelære	58
5.1	Grunnbegrepene	58
5.2	Den tomme mengden og andre spesielle mengder	59
5.3	Listeform, løsningsmengder og symbolikken $\{x \in A \mid \dots\}$	59
5.4	Delmengder og potensmengder	61
5.5	Union, snitt, differens og komplement	63

1 Tallteori

1.1 Innledning

Ordet "tallteori" brukes vanligvis om den delen av matematikken som dreier seg om begreper, problemstillinger og løsningsmetoder knyttet til hele tall, oftest både positive og negative. Av og til trekkes andre tallbegreper inn, slik som rasjonale, reelle og komplekse tall, men det skal vi ikke gjøre i dette kompendiet. Tallteorien har historiske røtter flere tusen år bakover. Den har blitt utviklet hovedsakelig gjennom ren og skjær nysgjerrighet, og av at en god del dyktige matematikere har funnet det uimotståelig at en del tilsynelatende enkle spørsmål har vært urimelig vanskelige å besvare. Mange "enkle spørsmål" er fortsatt ubesvart — noen slike er omtalt i dette kompendiet. Tanken om anvendelser utenfor matematikken har nok vært fjern for alle som har bidratt innen dette feltet, inntil for ca. 30–40 år siden. En av det 20. århundrets aller største tallteoretikere, briten G. H. Hardy, skrev en gang i 1930-årene at for tallteorien kunne man være trygg på at den aldri ville bli utsatt for praktiske anvendelser. Men da Hardy døde, i 1947, var datarevolusjonen allerede i gang (det tok riktignok noen år før noen skjønnte det), og tallteorien var på vei til å bli "anvendt matematikk". De senere årene har behovet for mest mulig sikker kryptering gjort tallteorien til et stadig viktigere verktøy. Dette kapittelet presenterer et lite utdrag fra det som gjerne kalles "elementær tallteori", et uttrykk som ikke betyr at det bare dreier seg om enkle ting. Det å forstå et begrep eller en påstand (setning, teorem) er ofte slett ikke så lett, og det vil være viktig å se sammenhenger mellom begrepene, påstandene og begrunnelsene (forklaringene, bevisene). De bevisene som er tatt med i dette kompendiet er ment som hjelp for leserne til å *innse* (men hva er det egentlig å *innse* noe?) at det som påstås i teoremet er sant. Av og til kan denne opplevelsen komme direkte, ved første gjennomlesing, men det er nok mer typisk at en må gjennomgå beviset flere ganger. Ofte er det behov for en tydelig markering av at "her er beviset slutt". Til det brukes her tegnet \triangle , plassert ytterst til høyre. Kapittelet inneholder en del oppgaver. Gi disse en sjanse, dvs. prøv å løse dem! Diskuter dem med medstudenter, og spør gjerne læreren.

1.2 Heltall, heltallsdivisjon og restdivisjon

Heltallene danner en uendelig mengde¹:

$$\mathbf{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}.$$

De danner også et *regnesystem*, der en fritt kan addere (beregne summer), subtrahere (beregne differenser) og multiplisere (beregne produkt), og til en viss grad (men langt fra fritt) kan dividere (beregne kvotienter). Regneoperasjonene har en del viktige egenskaper, også kalt de grunnleggende regneregulene.

Definisjon 1.2.1 Vi sier at et heltall a *går opp i* (eller *dividerer*, eller er en *divisor* for) et annet heltall b — med symboler: $a \mid b$ (les: ” a går opp i b ”, eller ” b er delelig med a ”) — dersom det fins et heltall q slik at $aq = b$.

Oppgave 1.2.1 Finn ut om noen av tallene 2, 3, 5, 10, 543, 547 og 551 går opp i 31407. Bruk gjerne kalkulator. Kunne du klart dette uten kalkulator? Hvordan?

Merknad 1.2.1 Noen fakta om $a \mid b$:

$$a \mid a \text{ for alle } a \in \mathbf{Z}.^2$$

Hvis $a \mid b$ og $b \mid c$, da følger det at $a \mid c$.

Hvis $a \mid b$ og $a \mid c$, da følger det at $a \mid (bx + cy)$ for alle $x, y \in \mathbf{Z}$.

Hvis $a \mid b$ og $b \mid a$, da er enten $a = b$ eller $a = -b$.

Oppgave 1.2.2 Er dette korrekte påstander? Er de opplagte? Kan de begrunnes? Hvordan?

Hvis a og b er heltall med $a > b > 0$, og vi har en samling på a stk. individer (f.eks. personer, kroner, sandkorn eller andre interessante ting), da kan vi begynne å gruppere samlingen i smågrupper med b stk. i hver. Kanskje vi kan danne mange slike (adskilte) grupper, eller kanskje bare noen få, men før eller senere — etter å ha dannet q ” b -grupper” — må den situasjonen oppstå at det ikke er nok individer igjen til å danne flere ” b -grupper”. Det må bety at antallet gjenværende individer, r , er et heltall som er *mindre enn* b . Antall individer i de q ” b -gruppene” er qb . Hvis $r = 0$, da er $a = qb$, og det betyr at $b \mid a$. Men uansett så er konklusjonen at

$$a = qb + r, \quad 0 \leq r < b,$$

¹Se side 58 ang. mengder og mengdelære.

²Symbolet ” \in ” leses ”som er element i mengden”, eller bare ”element i”. Dette er forklart på side 5.1.

og vi sier da at q er *heltallskvotienten* og r er *resten* ved divisjonen $a : b$. Merk at disse to er *entydig* bestemt av a og b , for hvis

$$a = q_1b + r_1 = q_2b + r_2, \quad 0 \leq r_1 \leq r_2 < b,$$

da følger det at $q_1b - q_2b = r_2 - r_1$, dvs. $(q_1 - q_2)b = r_2 - r_1$, der $0 \leq r_2 - r_1 < b$, og det eneste tallet i dette spennet som b går opp i er 0, og derfor har vi både at $r_2 - r_1 = 0$ og $(q_1 - q_2)b = 0$, som gir at $r_2 = r_1$ og $q_2 = q_1$ (siden $b \neq 0$).

1.3 Divisorer, perfekte tall og primtall

Definisjon 1.3.1 De *positive divisorene* til et heltall a er de positive heltallene som går opp i a .

Eksempel 1.3.1 De positive divisorene til 108 er: 1, 2, 3, 4, 6, 9, 12, 18, 27, 36, 54 og 108, mens de positive divisorene til 109 er 1 og 109.

Oppgave 1.3.1 Sjekk at påstandene i eksempelet over stemmer, og finn alle de positive divisorene til 105 og 112. Bruk gjerne kalkulator.

Definisjon 1.3.2 De *ekte divisorene* til et heltall a er de positive divisorene til a bortsett fra a selv.

Eksempel 1.3.2 De ekte divisorene til 108 er: 1, 2, 3, 4, 6, 9, 12, 18, 27, 36, og 54, mens 1 er den eneste ekte divisor til 109.

Noen heltall er lik summen av sine ekte divisorer. Dette gjelder f.eks. 6 ($=1+2+3$) og 28 ($=1+2+4+7+14$). Slike tall kalles *perfekte*. De har blitt studert siden oldtiden, ihvertfall siden Pythagoras' tid, ca. 500 f.Kr. Etter iherdig innsats fra mange i løpet av 2500 år har en funnet 39 perfekte tall. De to neste, etter 28, er 496 og 8128, og det største kjente er $2^{88}(2^{89} - 1)$. Alle de 39 som er kjent, er partall³. Det er fremdeles ukjent om det fins perfekte oddetall.

³Det har vært kjent siden Euklids tid — og bevist i hans bok *Elementer* ca. år 300 f.Kr. — at a er et perfekt partall hvis a kan skrives som et slikt produkt: $a = 2^{m-1}(2^m - 1)$ der $m \geq 2$ og $2^m - 1$ er primtall (se Definisjon 1.3.3 på side 6). Den sveitsiske matematikeren Leonhard Euler (ansett som den mest produktive matematiker gjennom tidene) viste at *alle* perfekte partall er av denne typen (det ble først kjent i 1849, 66 år etter hans død). Det er ikke vanskelig å vise at $2^m - 1$ ikke er primtall hvis m ikke er det, så m må være primtall hvis $a = 2^{m-1}(2^m - 1)$ skal kunne være et perfekt partall. Tallet $2^m - 1$ kalles *Mersenne-tall nr. m* og skrives ofte M_m , mens primtall av typen $2^p - 1$, der p er primtall, kalles *Mersenne-primtall*. M_m kan være sammensatt selv om m er primtall — f.eks. er $M_{11} = 2^{11} - 1 = 23 \cdot 89$. (Sjekk dette!) Jakten på virkelig store primtall har stort sett handlet om utvikling av metoder for å identifisere Mersenne-primtall, dvs. for å fastslå om $2^p - 1$ er primtall når p er det. Det er ukjent om det fins uendelig mange Mersenne-primtall, og bare 39 er hittil identifisert (disse gir dermed de 39 hittil kjente perfekte tallene).

Et langt viktigere begrep i tallteorien er:

Definisjon 1.3.3 Heltall > 1 med 1 som eneste ekte divisor kalles *primtall*. Heltall > 1 som ikke er primtall kalles *sammensatte tall*.

Primtallene har også blitt studert siden oldtiden, enda mer intenst enn perfekte tall, og mye er kjent om dem. Flere viktige spørsmål om primtall er imidlertid uavklarte, og noen av disse er viktige for moderne informatikk. De 10 første primtallene er: 2, 3, 5, 7, 11, 13, 17, 19, 23 og 29. Merk at 2 er det eneste primtallet som er partall, da alle andre partall har 2 som ekte divisor. Det er lett å finne en god del flere primtall, men etterhvert blir de mer og mer "sjeldne", dvs. at det er en tendens til økende mellomrom mellom dem (men merk Merknad 1.3.1 nedenfor), og det er ikke lett å identifisere riktig store primtall. Det er derimot forholdsvis lett å innse at det eksisterer uendelig mange primtall, dvs. at for ethvert heltall a fins primtall som er $\geq a$. Dette var kjent allerede på Euklids tid, omkring 300 f.Kr., og vi setter det opp som et "teorem" (dvs. en påstand som er bevist):

Teorem 1.3.1 Det fins uendelig mange primtall, dvs. for ethvert heltall a fins primtall som er $\geq a$.

Bevis: Først merker vi oss at ethvert heltall > 1 har minst én divisor som er > 1 , og den minste av disse er garantert et primtall (**Oppgave:** Er *det* så opplagt?), så ethvert heltall > 1 er delelig med minst ett primtall. Så tenker vi oss at vi har en endelig serie med primtall: p_1, p_2, \dots, p_n . Vi skal påvise at dette ikke kan være *alle* primtall. Dette innser vi slik: Tallet $a = p_1 p_2 \cdots p_n + 1$ har minst én primtallsdivisor, men ingen av primtallene i serien p_1, p_2, \dots, p_n går opp i a , fordi hvert av dem vil gi rest lik 1 når a divideres med det. Vi vet at a har minst én primtallsdivisor, og derfor fins minst ett primtall som ikke er med i serien p_1, p_2, \dots, p_n . Konklusjonen er at ingen endelig primtallsserie inneholder alle primtall, og det betyr at de (primtallene) er uendelig i antall. \triangle

Merknad 1.3.1 Påstanden ovenfor, at det er en tendens til økende mellomrom mellom primtallene, er upresis, for primtallenes fordeling i tallrekken er tildels uforutsigbar. En viktig hovedtendens er riktignok kjent, på den måten at antallet primtall som er $\leq a$ — ofte betegnet med symbolet $\pi(a)$ — for store tall a er omtrent lik $a/2$ dividert med antallet siffer i a (se fotnote⁴), og dette innebærer

⁴Dette anslaget over antall primtall $\leq a$ er en forenklet utgave av det såkalte *primtalls-teoremet*. Mer presist sier dette at forholdet mellom $\pi(a)$ og verdien av $a/\ln(a)$ nærmer seg 1 når $a \rightarrow \infty$. Dette ble forutsagt av tyskeren Carl Friedrich Gauss (kanskje den mest bety-

en ”tendens til økende mellomrom”. Likevel er det mange (ingen vet om det er uendelig mange) tilfelle av primtall med ”nesten-minimalt” mellomrom, dvs. med mellomrom lik 2 (det fins to primtall som har mellomrom lik 1 — hvilke er det, og hvorfor er dette det eneste eksempelet med mellomrom lik 1?). Her er noen eksempler med mellomrom 2: 11 og 13, 29 og 31, 71 og 73, 521 og 523. Som sagt, ingen vet om det fins uendelig mange slike *primtallspar*. Et annet ”klassisk” åpent primtallsproblem er *Goldbachs hypotese*, som sier (dvs. påstår, uten bevis) at ethvert partall > 2 kan skrives som en sum av to primtall. Matematikeren Goldbach fremmet denne påstanden i et brev til den mer kjente matematikeren Euler i 1742, og den er altså fremdeles uavklart.

1.4 Største felles divisor, Euklids algoritme og Aritmetikkens Fundamentalsetning

Siden 1 er en divisor for ethvert heltall, så har to heltall, a og b , alltid minst én felles divisor. Og siden antallet divisorer er endelig, så har a og b alltid en *største felles divisor*, som er omtalt i følgende definisjon:

Definisjon 1.4.1 $\gcd(a,b)$ (etter det engelske uttrykket *greatest common divisor*) er den største felles divisor for heltallene a og b .

Oppgave 1.4.1 Finn $\gcd(108,72)$, f.eks. ved å finne alle divisorene til både 108 og 72 (eller på annen måte). Hvordan kan resultatet brukes til å forkorte brøken $\frac{72}{108}$?

Merknad 1.4.1 Hvis c er en felles divisor for a og b , og q og r er heltall slik at $a = qb + r$, da er c også en divisor for r , fordi $r = a - qb$. Og hvis d er en felles divisor for b og r , da er den også en divisor for a , så likheten $a = qb + r$ innebærer at fellesdivisorene for a og b er nøyaktig de samme som fellesdivisorene for b og r , og da følger det at $\gcd(a,b) = \gcd(b,r)$. Hvis $r = 0$, dvs. hvis $a = qb$, da innebærer dette at $\gcd(a,b) = \gcd(b,r) = \gcd(b,0) = b$.

delige matematiker siden Newton) i 1793, og bevist først i 1896 av franskmannen Hadamard og belgieren de la Vallée-Poussin. De brukte begge, uavhengig av hverandre, funksjonsteoretiske metoder i sine bevis. Selv om dette viktige teoremet dermed var fastslått, ble det sett på som viktig og oppsiktsvekkende da det i 1949 kom et rent tallteoretisk bevis. Det var produsert av nordmannen Atle Selberg, født i 1917 i Langesund.

Definisjonen av $\gcd(a,b)$ kan tyde på at vi må finne alle divisorene til a og b før vi kan fastslå $\gcd(a,b)$. Dette er svært tidkrevende, men det har ganske lenge (ca. 2300 år, minst) vært kjent at det finnes en langt raskere metode. Den kalles

Euklids algoritme: Vi starter med to heltall, a og b , med $0 < b < a$. Da kan vi finne (ved "vanlig divisjon") heltall q_1 og r_1 med $0 \leq r_1 < b$ slik at

$$a = q_1 b + r_1 .$$

Hvis $r_1 = 0$, da følger det at $a = q_1 b$, og da er $b = \gcd(a,b)$. Hvis $r_1 > 0$, da kan vi finne heltall q_2 og r_2 med $0 \leq r_2 < r_1$ slik at

$$b = q_2 r_1 + r_2 .$$

Hvis $r_2 = 0$, da følger det at $b = q_2 r_1$, og da er $r_1 = \gcd(b,r_1) = \gcd(a,b)$. Hvis $r_2 > 0$, da kan vi finne heltall q_3 og r_3 med $0 \leq r_3 < r_2$ slik at

$$r_1 = q_3 r_2 + r_3 .$$

Hvis $r_3 = 0$, da følger det at ...

Denne prosessen er slik at $b > r_1 > r_2 > r_3 > \dots$, og $r_i \geq 0$ for alle i . Da må vi før eller senere oppnå at $r_i = 0$, og da er, iflg. Merknad 1.4.1 ovenfor, $r_{i-1} = \gcd(r_{i+1}, r_i) = \dots = \gcd(b, r_1) = \gcd(a, b)$.

(**Oppgave:** Innse at denne metoden alltid fører fram til $\gcd(a,b)$.)

Eksempel 1.4.1 Vi prøver ut Euklids algoritme med tallene $a = 777\,067\,673$ og $b = 26\,480\,567$. Divisjon (bruk gjerne kalkulator) viser at a da er lik $29 \cdot b$ pluss en rest r , som vi finner slik: $r = a - 29 \cdot b = 9\,131\,230$, og vi er i gang:

$$777\,067\,673 = 29 \cdot 26\,480\,567 + 9\,131\,230 ,$$

$$26\,480\,567 = 2 \cdot 9\,131\,230 + 8\,218\,107 ,$$

$$9\,131\,230 = 1 \cdot 8\,218\,107 + 913\,123 ,$$

$$8\,218\,107 = 9 \cdot 913\,123 + 0 .$$

Konklusjon: $\gcd(777\,067\,673, 26\,480\,567) = 913\,123$.

Merk at vi ikke har funnet alle divisorene til a og b her, og at metoden likevel garanterer (med forbehold om regnefeil — men det kan sjekkes!) at $\gcd(777\,067\,673, 26\,480\,567) = 913\,123$.

Oppgave 1.4.2 Forkort brøken $\frac{26\,480\,567}{777\,067\,673}$ så mye som mulig.

Oppgave 1.4.3 Finn $\gcd(108, 72)$, $\gcd(713, 841)$ og $\gcd(1024, 729)$ ved Euklids algoritme.

Definisjon 1.4.2 Hvis a og b er slik at $\gcd(a,b)=1$ — som er laveste mulige verdi for $\gcd(a,b)$ — da sier vi at a og b er *relativt primiske*.

Eksempel 1.4.2 Hvis du har gjort Oppgave 1.4.3 over, så har du eksempler. Hvis ikke, så gjør den nå.

Merknad 1.4.2 Hvis to heltall, a og b , er slik at 1 kan skrives som en (heltallig) lineær kombinasjon av a og b , dvs. hvis det fins heltall m og n — ikke nødvendigvis positive — slik at

$$ma + nb = 1 ,$$

da må a og b være relativt primiske, for enhver felles divisor c er jo også en divisor for $ma + nb$, dvs. for 1, og da er $c = 1$. At det motsatte gjelder — at ethvert par av relativt primiske tall kan gi 1 som en lineær kombinasjon — er kanskje ikke så opplagt, men det følger av Euklids algoritme:

Teorem 1.4.1 Hvis $\gcd(a,b)=1$, da fins heltall m og n slik at $ma + nb = 1$.

Bevis: Iflg. forutsetningen her vil Euklids algoritme føre til en ”sluttrest” lik 1, slik:

$$a = q_1b + r_1 ,$$

$$b = q_2r_1 + r_2 ,$$

$$r_1 = q_3r_2 + r_3 ,$$

$$r_2 = q_4r_3 + r_4 ,$$

$$\vdots$$

$$r_{i-2} = q_i r_{i-1} + r_i ,$$

$$r_{i-1} = q_{i+1} r_i + 1 .$$

Her kan den siste likningen ”løses m.h.p. 1”, og så kan vi sette inn for r_i ved å bruke likningen over, osv. til vi har fått uttrykt 1 ved a og b . \triangle

Vi illustrerer dette med et eksempel:

Eksempel 1.4.3 Euklids algoritme med $a = 851$ og $b = 361$ gir:

$$851 = 2 \cdot 361 + 129 ,$$

$$361 = 2 \cdot 129 + 103 ,$$

$$129 = 1 \cdot 103 + 26 ,$$

$$103 = 3 \cdot 26 + 25 ,$$

$$26 = 1 \cdot 25 + 1 .$$

Innsettinger ”nedenfra og opp”, kombinert med sammentrekninger og forenklinger, her gir:

$$\begin{aligned} 1 &= 26 - 1 \cdot 25 = 26 - 1 \cdot (103 - 3 \cdot 26) = 26 - 103 + 3 \cdot 26 = 4 \cdot 26 - 103 = \\ &= 4 \cdot (129 - 1 \cdot 103) - 103 = 4 \cdot 129 - 5 \cdot 103 = 4 \cdot 129 - 5 \cdot (361 - 2 \cdot 129) = \\ &= 14 \cdot 129 - 5 \cdot 361 = 14 \cdot (851 - 2 \cdot 361) - 5 \cdot 361 = 14 \cdot 851 - 33 \cdot 361 . \end{aligned}$$

Euklids algoritme har flere viktige konsekvenser:

Teorem 1.4.2 Hvis p er et primtall, og en divisor for ab , da er p en divisor for a eller b (eller begge).

Bevis: Anta at primtallet p *ikke* er en divisor for a . Siden p ikke har andre ekte divisorer enn 1 følger det at $\gcd(a,p)=1$, og derfor fins det heltall m og n slik at $ma + np = 1$. Multiplikasjon med b gir da: $mab + npb = b$, så hvis $p \mid ab$, da er p divisor for begge ledd i uttrykket for b , og da følger det at $p \mid b$. \triangle

Teorem 1.4.3 Hvis p er et primtall, og en divisor for $a_1 a_2 \cdots a_k$, da er p en divisor for minst én av faktorene a_i .

Bevis: Se på produktet $a_1 a_2 \cdots a_k$ som et produkt av to faktorer: $a_1(a_2 \cdots a_k)$, og bruk Teorem 1.4.2 på dette. Bruk så denne teknikken omigjen m.h.p. produktet $a_2 a_3 \cdots a_k$, osv. \triangle

Teorem 1.4.4 (Aritmetikkens fundamentalsetning) Ethvert heltall > 1 har en éntydig primtallsfaktoriserings.

Bevis: Først skal vi se at ethvert heltall > 1 har minst én primtallsfaktoriserings. Som bemerket i beviset for Teorem 1.3.1 har ethvert heltall $a > 1$ minst én primtallsdivisor, p_1 . Da er a/p_1 et positivt heltall. Hvis $a/p_1 = 1$, da er $a = p_1$. Hvis $a/p_1 > 1$, da har a/p_1 minst én primtallsdivisor, p_2 . Da er $(a/p_1)/p_2 = a/(p_1p_2)$ et positivt heltall. Hvis $a/(p_1p_2) = 1$ da er $a = p_1p_2$. Hvis $a/(p_1p_2) > 1$ da har $a/(p_1p_2)$ minst én primtallsdivisor, p_3 , osv. Vi har da en prosess med avtakende heltallige kvotienter, og den er nødt til å stoppe før eller senere, dvs. vi må for en eller annen verdi av i oppnå at $a/(p_1p_2 \cdots p_i) = 1$, og da er $a = p_1p_2 \cdots p_i$. Dermed vet vi at a har minst én primtallsfaktoriserings.

Så skal vi se at slik primtallsfaktoriserings er éntydig. Dette betyr at to ”konkurrerende” slike:

$$a = p_1p_2 \cdots p_i = q_1q_2 \cdots q_j ,$$

der både alle p 'ene og alle q 'ene er primtall, må være ”essensielt like”, dvs. inneholde de samme primtallsfaktorene, hver av dem like mange ganger. For å se at slik entydighet alltid foreligger, anta at det *ikke* var tilfellet, dvs. at vi har

$$p_1p_2 \cdots p_i = q_1q_2 \cdots q_j$$

der en av primtallsfaktorene, f.eks. p_k , opptrer flere ganger på (f.eks.) venstre side enn på høyre. Kanskje den ikke forekommer i det hele tatt på høyre side, og hvis den gjør det (i et visst antall eksemplarer, men færre enn på venstre side), da kan vi forkorte med p_k så mange ganger at den blir borte fra høyre-siden. Vi tenker oss at det allerede er gjort, dvs. vi kan uten å miste gyldighet for resonnementet anta at p_k er en faktor på venstre side og ikke på høyre side. Dette er en selv-motsigelse, for p_k er jo da en divisor for $q_1q_2 \cdots q_j$, og iflg. Teorem 1.4.3 skulle den da være en divisor for minst én av faktorene q_1, q_2, \dots, q_j , og det er umulig, for disse er jo primtall og ingen av dem er lik p_k . \triangle

Vi har vært inne på bruken av symbolet $\pi(a)$ for antallet primtall $\leq a$. Antallet positive heltall $\leq a$ som er relativt primiske til a er også en viktig størrelse, viktig nok til at den også har blitt tildelt et særskilt symbol, med et særskilt navn, som fremgår av følgende definisjon, der den greske bokstaven ϕ (”fi”) inngår:

Definisjon 1.4.3 *Eulers fi-funksjon*, betegnet med $\phi(a)$ (les: ”fi-av-a”) for positive heltall a , står for antallet positive heltall $\leq a$ som er relativt primiske til a .

Eksempel 1.4.4 $\phi(1) = 1$, fordi det eneste positive heltallet ≤ 1 er 1, og dette er relativt primisk til 1.

$\phi(2) = 1$, fordi 1 er relativt primisk til 1, mens 2 er det ikke.

$\phi(8) = 4$, fordi 1, 3, 5 og 7, og ingen andre tall ≤ 8 , er relativt primiske til 8.

$\phi(9) = 6$, fordi 1, 2, 4, 5, 7 og 8 er relativt primiske til 9, mens 3, 6 og 9 ikke er det.

$\phi(12) = 4$, fordi 1, 5, 7 og 11, og ingen andre tall ≤ 12 , er relativt primiske til 12.

Oppgave 1.4.4 Sjekk at påstandene i Eksempel 1.4.4 stemmer, og beregn $\phi(20)$, $\phi(24)$ og $\phi(29)$. Bruk gjerne kalkulator.

Følgende teorem uttrykker en viktig egenskap for Eulers fi-funksjon:

Teorem 1.4.5 Eulers fi-funksjon er en *multiplikativ* funksjon, i den forstand at $\phi(ab) = \phi(a)\phi(b)$ for alle par (a, b) av relativt primiske positive heltall.

Dette teoremet er bevist i de fleste lærebøker i tallteori. Vi tar ikke med beviset her.

I Teorem 1.4.5 finner vi definisjonen av multiplikative funksjoner. Slike funksjoner har en egenskap som vi uttrykker i neste teorem:

Teorem 1.4.6 Hvis f er en multiplikativ funksjon, og primtallsfaktoriseringen av a er: $a = p_1^{e_1} p_2^{e_2} \cdots p_s^{e_s}$ (dvs. at p_1, p_2, \dots, p_s er innbyrdes ulike primtall, og e_1, e_2, \dots, e_s er positive heltall), da gjelder:

$$f(a) = f(p_1^{e_1}) f(p_2^{e_2}) \cdots f(p_s^{e_s}).$$

Bevis: Siste primtallspotensfaktor, $p_s^{e_s}$, er relativt primisk med produktet av de foregående, $p_1^{e_1} p_2^{e_2} \cdots p_{s-1}^{e_{s-1}}$ (ja, den er vel det??), så den multiplikative egenskapen for f gir at:

$$f(a) = f((p_1^{e_1} p_2^{e_2} \cdots p_{s-1}^{e_{s-1}}) \cdot p_s^{e_s}) = f(p_1^{e_1} p_2^{e_2} \cdots p_{s-1}^{e_{s-1}}) \cdot f(p_s^{e_s}).$$

Den første faktoren i dette siste produktet er av samme form som det foregående uttrykket for $f(a)$, så den multiplikative egenskapen for f gjør at denne faktoren også kan faktoriseres, slik:

$$f(a) = f(p_1^{e_1} p_2^{e_2} \cdots p_{s-1}^{e_{s-1}}) \cdot f(p_s^{e_s}) = f(p_1^{e_1} p_2^{e_2} \cdots p_{s-2}^{e_{s-2}}) \cdot f(p_{s-1}^{e_{s-1}}) \cdot f(p_s^{e_s}),$$

osv. til vi har fullført den faktoriseringen som er uttrykt i teoremet. △

Vi skal videre se at $\phi(a)$ er spesielt enkel å beregne når a er et primtall eller en primtallspotens, og ved å kombinere dette med Teoremene 1.4.5 og 1.4.6 vil vi komme fram til en viktig formel for $\phi(a)$ generelt (Teorem 1.4.9 nedenfor).

Teorem 1.4.7 For alle primtall p har vi at $\phi(p) = p - 1$.

Bevis: At p er primtall betyr jo nettopp at *alle* de $p - 1$ positive heltallene $< p$ er relativt primiske til p . Tallet p selv er selvsagt ikke relativt primisk til seg selv, så derfor er antallet positive heltall $\leq p$ som er relativt primiske til p lik $p - 1$.

△

Teorem 1.4.8 For alle primtallspotenser p^s har vi at

$$\phi(p^e) = p^e - p^{e-1} = p^e \left(1 - \frac{1}{p}\right).$$

Bevis: Antallet positive heltall $\leq p^e$, totalt, er selvsagt lik p^e . De av disse som *ikke* er relativt primiske til p^e er de som er delelige med p , dvs. de som kan skrives på formen kp der k er et positivt heltall $\leq p^{e-1}$, så antallet slik er nøyaktig lik $\leq p^{e-1}$. Antallet positive heltall $\leq p^e$ som er relativt primiske til p^e er derfor gitt ved differensen $p^e - p^{e-1}$.

△

Teorem 1.4.9 Hvis a er et positivt heltall med primtallsfaktoriseringen: $a = p_1^{e_1} p_2^{e_2} \cdots p_s^{e_s}$ (dvs. at p_1, p_2, \dots, p_s er innbyrdes ulike primtall, og e_1, e_2, \dots, e_s er positive heltall), da gjelder:

$$\phi(a) = a \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_s}\right).$$

Bevis: Iflg. Teoremene 1.4.5 og 1.4.6 er

$$\phi(a) = \phi(p_1^{e_1} p_2^{e_2} \cdots p_s^{e_s}) = \phi(p_1^{e_1}) \phi(p_2^{e_2}) \cdots \phi(p_s^{e_s}).$$

Teorem 1.4.8 gir en formel for hver av faktorene her, og innsetting gir: er

$$\begin{aligned} \phi(a) &= p_1^{e_1} \left(1 - \frac{1}{p_1}\right) \cdot p_2^{e_2} \left(1 - \frac{1}{p_2}\right) \cdots p_s^{e_s} \left(1 - \frac{1}{p_s}\right) = \\ &= p_1^{e_1} p_2^{e_2} \cdots p_s^{e_s} \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_s}\right) = a \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_s}\right). \end{aligned}$$

△

1.5 Kongruens og regning modulo m

Vi forutsetter her at m er et positivt heltall. Vi tenker oss at tallinjen er ”kveilet opp” rundt en sirkel som er akkurat så stor at tallene $0, 1, 2, \dots, m - 1$ blir jevnt fordelt rundt sirkelen, mens m er det første positive tallet som ”treffer” 0. Alle heltallene vil da ”treffe” et av tallene $0, 1, 2, \dots, m - 1$. Definisjon 1.5.1 nedenfor legger opp til at vi skal oppfatte to heltall, a og b , som ”like m.h.p. m ” (men vi skal i stedet si ”kongruente modulo m ”) dersom de ”treffer” hverandre ved en slik ”oppkveiling”.

Definisjon 1.5.1 Vi sier at heltallene a og b er *kongruente modulo m* , og skriver dette symbolsk slik:

$$a \equiv b \pmod{m},$$

dersom $m \mid (a - b)$. Tallet m kalles kongruensens *modulus*.

Eksempel 1.5.1 $15 \equiv 3 \pmod{12}$, siden 12 går opp i $15 - 3 = 12$. (Tenk på urskiven, og det at ”klokken 3” er identisk med ”klokken 15” i et 12-timers-system.

$$64 \equiv 24 \pmod{8}, \text{ siden } 8 \text{ går opp i } 64 - 24 = 40.$$

$$-6 \equiv -33 \pmod{9}, \text{ siden } 9 \text{ går opp i } -6 - (-33) = 27.$$

$$56 \equiv 0 \pmod{7}, \text{ siden } 7 \text{ går opp i } 56 - 0 = 0.$$

Merknad 1.5.1 (i) $a \equiv b \pmod{m}$ hvis og bare hvis a og b gir samme rest ved deling med m .

(ii) *Refleksivitet*: $a \equiv a \pmod{m}$ for alle heltall a .

(iii) *Symmetri*: Hvis $a \equiv b \pmod{m}$, så er også $b \equiv a \pmod{m}$.

(iv) *Transitivitet*: Hvis $a \equiv b \pmod{m}$ og $b \equiv c \pmod{m}$, så er også $a \equiv c \pmod{m}$.

(v) Hvis $a \equiv b \pmod{m}$ og $c \equiv d \pmod{m}$, så er også

$$a + c \equiv b + d \pmod{m} \text{ og } ac \equiv bd \pmod{m}.$$

Oppgave 1.5.1 Er dette korrekte påstander? Er de opplagte? Kan de begrunnes? Hvordan?

De tre egenskapene refleksivitet, symmetri og transitivitet ((ii), (iii) og (iv) i Merknad 1.5.1 over) betyr at kongruens modulo m er en *ekvivalensrelasjon*, og dette betyr igjen at mengden \mathbf{Z} av alle hele tall blir oppdelt i *ekvivalensklasser*, slik: a og b er i samme ekvivalensklasse hvis og bare hvis $a \equiv b \pmod{m}$. Ekvivalensklassen til a modulo m betegnes med $[a]_m$ (dette er altså mengden av alle de hele tall som er kongruente med a modulo m). **Oppgave:** Innse at hver ekvivalensklasse inneholder nøyaktig ett av tallene $0, 1, 2, \dots, m - 1$.

Pkt. (v) i Merknad 1.5.1 ovenfor sikrer at vi kan regne med ekvivalensklasser som om de var tall, ved å bruke fritt valgte tall fra hver ekvivalensklasse, slik:

$[a]_m + [b]_m$ betyr $[a + b]_m$, og $[a]_m \cdot [b]_m$ betyr $[ab]_m$. Hver ekvivalensklasse kan representeres ved ett (hvilket som helst) av tallene i den, men som regel brukes tallene $0, 1, 2, \dots, m - 1$ som "hovedrepresentanter" for sine ekvivalensklasser. Dermed kan "regning med ekvivalensklasser" overføres til "regning med tallene $0, 1, 2, \dots, m - 1$ ", slik:

Definisjon 1.5.2 Hvis a, b og c alle er blant tallene $0, 1, 2, \dots, m - 1$, så skal $a +_m b = c$ bety at $[c]_m = [a + b]_m$ (dvs. at $c \equiv a + b \pmod{m}$), og $a \cdot_m b = c$ bety at $[c]_m = [ab]_m$ (dvs. at $c \equiv ab \pmod{m}$). Mengden $\mathbf{Z}_m = \{0, 1, \dots, m - 1\}$ blir dermed et *regnesystem*, med addisjon og multiplikasjon. Hvis det er klart at en mener "regning modulo m ", for en bestemt m , da skriver en ofte $a + b$ og ab (eller $a \cdot b$) i stedet for $a +_m b$ og $a \cdot_m b$.

Oppgave 1.5.2 Innse at "hovedrepresentanten" for $a +_m b$ — når a og b er blant tallene $0, 1, 2, \dots, m - 1$ — er lik den resten en får ved å dele summen $a + b$ (beregnet ved "vanlig addisjon") med m , og at det tilsvarende gjelder for multiplikasjon modulo m .

Eksempel 1.5.2 I \mathbf{Z}_2 er $0 + 0 = 0, 0 + 1 = 1 + 0 = 1, 1 + 1 = 0, 0 \cdot 0 = 0, 0 \cdot 1 = 1 \cdot 0 = 0$ og $1 \cdot 1 = 1$.

I \mathbf{Z}_3 er $1 + 1 = 2, 1 + 2 = 2 + 1 = 0, 2 + 2 = 1$ og $2 \cdot 2 = 1$.

I \mathbf{Z}_{12} er $7 + 9 = 4, 11 + 11 = 10, 7 \cdot 9 = 3$ og $11 \cdot 11 = 1$.

Oppgave 1.5.3 Sjekk at påstandene i eksemplene ovenfor er korrekte.

Oppgave 1.5.4 Beregn følgende:

$4 + 5$ modulo 7 (dvs. beregn summen $4 + 5$ i \mathbf{Z}_7),

$9 + 11$ modulo 12 (hva er klokken 11 timer etter kl. 9?),

$9 + 11$ modulo 20,

$4 \cdot 5$ modulo 7,

$9 \cdot 11$ modulo 12, og

$9 \cdot 11$ modulo 20.

Det er ikke så vanskelig å innse at "vanlige regneregler" gjelder for addisjon og multiplikasjon i \mathbf{Z}_m , inklusive de spesielle egenskapene som 0 og 1 har: Addisjon er assosiativ og kommutativ i \mathbf{Z}_m , dvs.: $a + (b + c) = (a + b) + c$ og $a + b = b + a$, multiplikasjon er assosiativ og kommutativ i \mathbf{Z}_m , dvs.: $a(bc) = (ab)c$ og $ab = ba$, multiplikasjon er distributiv over addisjon i \mathbf{Z}_m , dvs.: $a(b + c) = ab + ac$, 0 er virkningsløs (nøytral) m.h.p. addisjon i \mathbf{Z}_m , dvs.: $a + 0 = a$, og 1 er virkningsløs (nøytral) m.h.p. multiplikasjon i \mathbf{Z}_m , dvs.: $1 \cdot a = a$.

Oppgave 1.5.5 Innse at regning i \mathbf{Z}_m har de nevnte egenskapene.

Men hva med subtraksjon og divisjon (eller delelighet) i \mathbf{Z}_m ? Subtraksjon er det enkleste, for det er ikke så vanskelig å innse at det for a og b i $\{0, 1, \dots, m-1\}$ fins én og bare en c i $\{0, 1, \dots, m-1\}$ som oppfyller kravet $c + b = a$. Denne c kalles da — naturlig nok — differensen $a - b$ i \mathbf{Z}_m .

Oppgave 1.5.6 Hvordan kan en beregne differensen $a - b$ i \mathbf{Z}_m ? Beregn $5 - 11$ modulo 12 (dvs. i \mathbf{Z}_{12}). Finn ut om svaret du har fått er rett, ved å ”sette prøve”.

Divisjon og delelighet behandler vi på liknende måte, og vi tar det en del nøyere. Hvis det skal ha mening å si, om et tall c i \mathbf{Z}_m , at det er lik en kvotient a/b , da må c være slik at $cb = a$, og vi må være sikre på at det ikke kan være mer enn én c i \mathbf{Z}_m med denne egenskapen. Dette betyr at vi må forutsette $b \neq 0$, for mange (faktisk alle) $c \in \mathbf{Z}_m$ er slik at $c0 = 0$. Men selv om $b \neq 0$, kan vi ha flertydighet, for f.eks. i \mathbf{Z}_{12} er både $2 \cdot 4$ og $5 \cdot 4$ lik 8 (**Oppgave:** Er det flere løsninger i \mathbf{Z}_{12} for likningen $x \cdot 4 = 8$?). Problemet med flertydige kvotienter i \mathbf{Z}_{12} henger sammen med at 12 er et sammensatt tall. I \mathbf{Z}_p , når p er et primtall, har vi éntydige kvotienter, og faktisk *eksisterer* alle kvotienter a/b med $b \neq 0$. Dette er innholdet i neste teorem:

Teorem 1.5.1 Hvis p er et primtall, så fins det, for alle $a, b \in \mathbf{Z}_p$ med $b \neq 0$, én og bare en $c \in \mathbf{Z}_p$ slik at $cb = a$ i \mathbf{Z}_p , dvs. slik at $cb \equiv a \pmod{p}$.

Bevis: Først viser vi at det *fins* slik c : Da p er primtall, og dermed relativt primisk til b , fins det iflg. Teorem 1.4.1 heltall m og n slik at $mp + nb = 1$. Vi ganger med a på begge sider, og får: $mpa + nab = a$. Siden det første leddet, mpa , er delelig med p , betyr dette at $nab \equiv a \pmod{p}$. Derfor vil $c = na$ være slik at $cb \equiv a \pmod{p}$. Så viser vi at slik c er *éntydig*: Hvis c_1 og c_2 begge har den aktuelle egenskapen, dvs. hvis både $c_1b \equiv a \pmod{p}$ og $c_2b \equiv a \pmod{p}$, da følger det at $c_1b \equiv c_2b \pmod{p}$ (ved Merknad 1.5.1, pkt. (iii) og (iv)), dvs. at $c_1b - c_2b = (c_1 - c_2)b$ er delelig med p . Da p er primtall, betyr dette at $c_1 - c_2$ eller b er delelig med p (ved Teorem 1.4.2). Men b er *ikke* delelig med p , for det eneste tallet i \mathbf{Z}_p som er delelig med p er 0, og vi har forutsatt at $b \neq 0$ i \mathbf{Z}_p . Derfor må $c_1 - c_2$ være delelig med p , og det betyr at $c_1 \equiv c_2 \pmod{p}$, dvs. at $c_1 = c_2$ i \mathbf{Z}_p . \triangle

Merk at Teorem 1.5.1 sier at ”tallsystemet” \mathbf{Z}_p — der p er et primtall — har en viktig egenskap, som tallsystemene \mathbf{Q} og \mathbf{R} har, men ikke \mathbf{Z} , nemlig at en kan dele med alt unntatt 0, dvs. at kvotienter a/b fins i \mathbf{Z}_p når $b \neq 0$.

Oppgave 1.5.7 For noen forskjellige primtall p etter eget valg (f.eks. 2, 3, 5, 7 eller andre), sett opp en fullstendig "gangetabell" for \mathbf{Z}_p , som viser hva ab er lik i \mathbf{Z}_p for alle a og b i \mathbf{Z}_p . Bruk så tabellen til å finne ut hva $1/a$ (inversen til a) er i \mathbf{Z}_p , for alle $a \neq 0$ i \mathbf{Z}_p .

Hva med kvadratrøtter i \mathbf{Z}_m ? Vi vet at likningen $x^2 = 2$ ikke har noen løsninger i \mathbf{Z} eller i \mathbf{Q}^5 , men i \mathbf{R} har den to løsninger, nemlig $\pm\sqrt{2}$. I \mathbf{Z}_m skal vi stille spørsmålet slik: Hvor mange løsninger har likningen

$$x^2 = a$$

i \mathbf{Z}_m , og hvordan kan vi finne dem?

Eksempel 1.5.3 Vi beregner først alle andrepotenser x^2 , med $x \neq 0$, i \mathbf{Z}_8 : $1^2 = 1$, $2^2 = 4$, $3^2 = 9 = 1$ (**Oppgave:** Hvorfor er $9 = 1$ i \mathbf{Z}_8 ?), $4^2 = 16 = 0$ (**Oppgave:** Hvorfor er $16 = 0$ i \mathbf{Z}_8 ?), $5^2 = 25 = 1$ (**Oppgave:** Hvorfor er $25 = 1$ i \mathbf{Z}_8 ?), $6^2 = 36 = 4$ (**Oppgave:** Hvorfor er $36 = 4$ i \mathbf{Z}_8 ?), $7^2 = 49 = 1$ (**Oppgave:** Hvorfor er $49 = 1$ i \mathbf{Z}_8 ?). Vi ser her at likningen $x^2 = 1$ har hele fire løsninger i \mathbf{Z}_8 : 1, 3, 5 og 7. Likningen $x^2 = 4$ har to løsninger: 2 og 6, mens $x^2 = 0$ også har to løsninger: 0 og 4. Likningene $x^2 = 2$, $x^2 = 3$, $x^2 = 5$, $x^2 = 6$ og $x^2 = 7$ har alle null løsninger.

I \mathbf{Z}_{10} har vi følgende andrepotenser: $1^2 = 1$, $2^2 = 4$, $3^2 = 9$, $4^2 = 16 = 6$ (**Oppgave:** Hvorfor er $16 = 6$ i \mathbf{Z}_{10} ?), $5^2 = 25 = 5$ (**Oppgave:** Hvorfor er $25 = 5$ i \mathbf{Z}_{10} ?), $6^2 = 36 = 6$ (**Oppgave:** Hvorfor er $36 = 6$ i \mathbf{Z}_{10} ?), $7^2 = 49 = 9$ (**Oppgave:** Hvorfor er $49 = 9$ i \mathbf{Z}_{10} ?), $8^2 = 64 = 4$ (**Oppgave:** Hvorfor er $64 = 4$ i \mathbf{Z}_{10} ?), $9^2 = 81 = 1$ (**Oppgave:** Hvorfor er $81 = 1$ i \mathbf{Z}_{10} ?).

Vi ser her at likningen $x^2 = 1$ har to løsninger i \mathbf{Z}_{10} : 1 og 9. Likningen $x^2 = 4$ har også to løsninger: 2 og 8, mens $x^2 = 5$ har én løsning: 5. Likningen $x^2 = 6$ har to løsninger: 4 og 6, og det samme har $x^2 = 9$: 3 og 7, mens $x^2 = 0$ kun har den trivielle løsningen: $x = 0$. Likningene $x^2 = 2$, $x^2 = 3$, $x^2 = 7$ og $x^2 = 8$ har alle null løsninger i \mathbf{Z}_{10} .

⁵Bevis for at tallet 2 ikke har noen rasjonal kvadratrot:

La p/q være et helt tilfeldig rasjonalt tall, ferdig forkortet (dvs. at p og q ikke har noen felles faktor > 1). Da har heller ikke p^2 og q^2 noen felles faktor, for disse har jo de samme primtallfaktorene som h.h.v. p og q , men i dobbelt antall. Da er brøken $p^2/q^2 = (p/q)^2$ også ferdig forkortet. Hvis denne brøken skal være lik det hele tallet 2, da må q^2 være lik 1, dvs. $q = 1$, så det rasjonale tallet p/q må ha vært et helt tall. Men ikke noe helt tall har 2.-potens lik 2, for $1^2 = 1$, som er < 2 , og $n > 1 \Rightarrow n^2 > 2$. (**Oppgave:** Innse at den sistnevnte implikasjonen gjelder for hele tall.) Konklusjonen er at p/q ikke kan ha 2.-potens lik 2. Dermed er påstanden bevist.

I \mathbf{Z}_{11} har vi følgende andrepotenser: $1^2 = 1$, $2^2 = 4$, $3^2 = 9$, $4^2 = 16 = 5$ (**Oppgave:** Hvorfor er $16 = 5$ i \mathbf{Z}_{11} ?), $5^2 = 25 = 3$ (**Oppgave:** Hvorfor er $25 = 3$ i \mathbf{Z}_{11} ?), $6^2 = 36 = 3$ (**Oppgave:** Hvorfor er $36 = 3$ i \mathbf{Z}_{11} ?), $7^2 = 49 = 5$ (**Oppgave:** Hvorfor er $49 = 5$ i \mathbf{Z}_{15} ?), $8^2 = 64 = 9$ (**Oppgave:** Hvorfor er $64 = 9$ i \mathbf{Z}_{11} ?), $9^2 = 81 = 4$ (**Oppgave:** Hvorfor er $81 = 4$ i \mathbf{Z}_{11} ?), $10^2 = 100 = 1$ (**Oppgave:** Hvorfor er $100 = 1$ i \mathbf{Z}_{11} ?).

Vi ser her at likningen $x^2 = 1$ har to løsninger i \mathbf{Z}_{11} : 1 og 10. Likningen $x^2 = 3$ har også to løsninger: 5 og 6, og det samme har $x^2 = 4$: 2 og 9, og $x^2 = 5$: 4 og 7, og $x^2 = 9$: 3 og 8, mens $x^2 = 0$ kun har den trivielle løsningen: $x = 0$. Likningene $x^2 = 2$, $x^2 = 6$, $x^2 = 7$, $x^2 = 8$ og $x^2 = 10$ har alle null løsninger i \mathbf{Z}_{11} .

Definisjon 1.5.3 Hvis p er et primtall, og $a \in \mathbf{Z}_p$ er slik at likningen $x^2 = a$ har minst én løsning i \mathbf{Z}_p , da sier vi at a er et *kvadratisk residy* for p .

Eksempel 1.5.4 Ifølge det siste av de tre eksemplene i Eksempel 1.5.3, er 0, 1, 3, 4, 5 og 9 kvadratiske residyer for 11, mens 2, 6, 7, 8 og 10 ikke er det.

Oppgave 1.5.8 Finn de kvadratiske residyer for 8 og til 10.

Til avslutning tar vi med — foreløpig uten bevis⁶ — et resultat med viktige anvendelser i forbindelse med kryptering. Teoremet skyldes den franske juristen (av utdannelse) og matematikeren (av interesse) Pierre de Fermat (1601-65):

Teorem 1.5.2 ("Fermats lille teorem"⁷) Hvis p er et primtall, og a et heltall som ikke er delelig med p , da gjelder:

$$a^{p-1} \equiv 1 \pmod{p},$$

eller, med andre ord: $a^{p-1} - 1$ er delelig med p .

⁶Et bevis kommer i neste kapittel, side 25.

⁷Navnet er valgt for å skille teoremet fra det såkalte "Fermats store teorem", som sier at det for heltallig $n > 2$ ikke fins heltall x , y og z med $x^n + y^n = z^n$. Med $n = 2$ får vi en likning som vi kjenner igjen fra Pythagoras' setning, og som har mange heltallige løsninger, f.eks. $(x, y, z) = (3, 4, 5)$ og $(x, y, z) = (5, 12, 13)$. Lister av slike "Pythagoreiske tripler" er funnet på Babylonske leirtavler fra ca. 1700 f.Kr., så interessen for slike er ganske gammel! Fermat skrev i margin på en bok han hadde at han hadde funnet et bevis for at heltallsløsninger er umulig hvis $n > 2$, men at det ikke var plass i margin til å skrive inn beviset. Denne påstanden var en uoverkommelig utfordring for utallige matematikere frem til 1995, da påstanden endelig ble bevist av engelskmannen Andrew Wiles. Det er nok få som tror at Fermat virkelig hadde et gyldig bevis for 350 år siden, selv om han er ansett som en av de aller største i matematikkens historie.

Vi kan lett innse at $a^{p-1} - 1$ *ikke* er delelig med p hvis a er det, for a^{p-1} og $a^{p-1} - 1$ kan ikke begge være delelige med p . Vi kan også innse at følgende teorem er en konsekvens av Fermats lille teorem:):

Teorem 1.5.3 Hvis p er et primtall, så gjelder

$$a^p \equiv a \pmod{p}$$

for *ethvert* positivt heltall a .

Bevis: Hvis a *ikke* er delelig med p , da følger påstanden fra Teorem 1.5.2, fordi $a \cdot a^{p-1} \equiv a \pmod{p}$ følger fra $a^{p-1} \equiv 1 \pmod{p}$ ved Merknad 1.5.1, pkt. (v). Hvis a *er* delelig med p , da er også påstanden i Teorem 1.5.3 korrekt, fordi både a^p og a da er lik 0 i \mathbf{Z}_p . △

Fermats lille teorem kan også være til hjelp ved beregning av inverser (se Oppgave 1.5.7) i \mathbf{Z}_p :

Teorem 1.5.4 Hvis p er et primtall, og a et positivt heltall som ikke er delelig med p , da er a^{p-2} (eller egentlig den resten en får ved å dele a^{p-2} med p) inversen til a i \mathbf{Z}_p .

Bevis: $a \cdot a^{p-2} = a^{p-1} = 1$ i \mathbf{Z}_p . △

2 Algebraiske strukturer

2.1 Innledning

I algebra er det viktig å være bevisst på regneoperasjonenes egenskaper. Her er noen eksempler:

I tallsystemene \mathbf{Z} (de hele tallene), \mathbf{Q} (de rasjonale tallene) og \mathbf{R} (de reelle tallene) har *addisjon* følgende egenskaper:

Assosiativitet: $a + (b + c)$ er lik $(a + b) + c$ for alle tall a , b og c .

Kommutativitet: $a + b$ er lik $b + a$ for alle tall a og b .

Eksistens av nøytralt element: Det fins et tall (vi kaller det "null", og skriver det "0") som ikke endrer noe tall ved addisjon: $a + 0$ er lik a for alle tall a .

Eksistens av inverser: For ethvert tall a fins det et tall b som er slik at $a + b$ er lik det nøytrale elementet. For hver a er det alltid bare én slik b , og vi skriver den $-a$.

I de endelige tallsystemene \mathbf{Z}_m har "addisjon modulo m " de samme fire egenskapene (jf. side 15).

Hva om vi erstatter addisjon (+) med multiplikasjon (\cdot) ovenfor? Da gjelder de tre første egenskapene (bortsett fra at det nøytrale element er 1 og ikke 0). Eksistens av inverser gjelder *ikke*. (Hva er det som svikter her?)

Her er en annen slags regneoperasjon:

En *permutasjon* på mengden $\mathbf{Z}_m = \{0, 1, 2, \dots, m-1\}$ er en funksjon p fra \mathbf{Z}_m til \mathbf{Z}_m som ikke gjentar noen verdier, dvs. at $p(i) \neq p(j)$ for alle i og j i \mathbf{Z}_m med $i \neq j$. Slike permutasjoner kan vi tenke på som rangeringer av tallene i \mathbf{Z}_m , dvs. som oppstillinger der $p(0)$ settes på første plass, $p(1)$ på andre plass osv.

Vi definerer en "regneoperasjon" på mengden $P(\mathbf{Z}_m)$ av alle permutasjoner på \mathbf{Z}_m slik (p , q og r betyr her "vilkaarlige permutasjoner på \mathbf{Z}_m "): $p \circ q$ skal bety den permutasjonen som fremkommer ved:

$$(p \circ q)(i) = p(q(i)) \quad \text{for alle } i \in \mathbf{Z}_m .$$

Dette er en assosiativ operasjon, for funksjonsverdiene til både $p \circ (q \circ r)$ og $(p \circ q) \circ r$, for ethvert tall $i \in \mathbf{Z}_m$, er gitt ved: $p(q(r(i)))$, og derfor er $p \circ (q \circ r)$ og $(p \circ q) \circ r$ samme permutasjon. Operasjonen har et "nøytralt element", nemlig den permutasjonen p_0 som ikke flytter på noen tall: $p_0(i) = i$ for alle $i \in \mathbf{Z}_m$, og inverser fins: Hvis $p \in P(\mathbf{Z}_m)$, og vi definerer q ved at $q(i)$ skal være det tallet $j \in \mathbf{Z}_m$ som er slik at $p(j) = i$, da blir $p \circ q = q \circ p = p_0$. Operasjonen "o" er derimot *ikke* kommutativ, for $p \circ q$ og $q \circ p$ kan godt være forskjellige permutasjoner. (**Oppgave:** Sjekk riktigheten av alle disse påstandene!)

I dette kapitlet skal vi fokusere på regneoperasjonenes formelle (algebraiske) egenskaper, og bruke dette til å trekke konklusjoner som vil gjelde for mange ulike eksempler, og til å gjennomskue både likheter og ulikheter mellom eksempler.

2.2 Grupper

Definisjon 2.2.1 En *gruppe* er en mengde (gruppas *grunnmengde*) med en assosiativ regneoperasjon og et nøytralt element og med inverser for alle elementer.

I generell ("abstrakt") sammenheng vil vi ofte skrive G for grunnmengden, $*$ for regneoperasjonen, e for det nøytrale elementet, og a^{-1} for inversen til ethvert element a i G (men når regneoperasjonen er (eller kalles) addisjon, vil det være mer naturlig å skrive 0 i stedet for e , og $-a$ i stedet for a^{-1}). En gruppe betegnes ofte med et par, slik: $(G, *)$. Hvis regneoperasjonen er underforstått, omtaler vi ofte en gruppe bare med symbolet for dens grunnmengde, f.eks. G . Vi skal videre skrive a^n (for heltall $n > 0$) for $a * a * \dots * a$ med n "faktorer", a^{-n} (for heltall $n > 0$) for $a^{-1} * a^{-1} * \dots * a^{-1}$ med n "faktorer", og a^0 for e (men når regneoperasjonen er — eller kalles — addisjon, vil det være mer naturlig å skrive na i stedet for a^n , og $-na$ i stedet for a^{-n}).⁸

Kravene til en gruppe ("gruppe-aksiomene") kan skrives slik:

$$x * (y * z) = (x * y) * z \text{ for alle } x, y \text{ og } z,$$

$$x * e = e * x = x \text{ for alle } x, \text{ og}$$

$$x * x^{-1} = x^{-1} * x = e \text{ for alle } x.$$

Merk at inverser er éntydige, for hvis x, y og z i en gruppe er slik at $x * y = e$ og $z * x = e$, da følger det at

$$y = e * y = (z * x) * y = z * (x * y) = z * e = z .$$

Definisjon 2.2.2 En gruppe $(G, *)$ kalles *abelsk* hvis $*$ er en kommutativ operasjon.⁹

Noen **eksempler** har vi allerede sett på, i innledningen:

⁸Det er en innarbeidet skikk at ordene "addisjon" og "sum", og tegnet "+", bare brukes i abelske grupper (jf. Definisjon 2.2.2 nedenfor). Ordene "multiplikasjon" og "produkt", og tegnet "*" (eller ".", eller "sammenstilling uten regnetegn", slik som i vanlig algebra), derimot, brukes mer allment. I neste avsnitt skal vi se på algebraiske systemer med *to* regneoperasjoner, ofte kalt "addisjon" og "multiplikasjon", som skal oppfylle visse formelle krav. Slike systemer kalles *ringer*.

⁹Niels Henrik Abel (1802-1829) var død før gruppebegrepet ble innført i matematikken, men han var svært nær ved å identifisere dette som et viktig begrep.

$(\mathbf{Z}, +)$, $(\mathbf{Q}, +)$, $(\mathbf{R}, +)$, $(\mathbf{Z}_m, +)$ og $(P(\mathbf{Z}_m), \circ)$ er grupper. De tre førstnevnte er abelske, den sistnevnte er det ikke.

Her er noen flere eksempler:

Eksempel 2.2.1 Hvis vi skriver \mathbf{Q}^* (hvv. \mathbf{R}^*) for mengden av alle de rasjonale (hvv. reelle) tall som er ulik 0, og lar ” \cdot ” bety multiplikasjon (som vanlig), da er (\mathbf{Q}^*, \cdot) og (\mathbf{R}^*, \cdot) grupper.

Oppgave 2.2.1 Sjekk at den siste påstanden i Eksempel 2.2.1 er rett. Er disse gruppene abelske? Hvorfor er ikke de hele tallene $\neq 0$, med multiplikasjon, tatt med som et eksempel? Er (\mathbf{Z}_m, \cdot) en gruppe for noen verdier av m ?

Neste eksempel beskriver grupper som har betydning for tallteorien:

Eksempel 2.2.2 De tallene i $\mathbf{Z}_m = \{0, 1, \dots, m-1\}$ som er relativt primiske til m (jf. Definisjon 1.4.2 på side 9) danner en delmengde av \mathbf{Z}_m :

$$\mathbf{Z}_m^* = \{x \in \mathbf{Z}_m \mid \gcd(x, m) = 1\} .$$

Med ”multiplikasjon modulo m ” er dette en abelsk gruppe, fordi:

(1) \mathbf{Z}_m^* inneholder det (multiplikative) nøytrale elementet 1, siden $\gcd(1, m) = 1$,

(2) $x, y \in \mathbf{Z}_m^* \Rightarrow xy \in \mathbf{Z}_m^*$, for hvis $\gcd(x, m) = \gcd(y, m) = 1$, da fins, iflg. Teorem 1.4.1 på side 9, heltall a, b, c og d slik at $ax + bm = cy + dm = 1$. Av dette følger at $(ax + bm)(cy + dm) = 1$, dvs. $acxy + (adx + bcy + bdm)m = 1$ (gang ut de to parentesene!), og da er $\gcd(xy, m) = 1$ iflg. Merknad 1.4.2 på side 9.

(3) For hver $x \in \mathbf{Z}_m^*$ fins $y \in \mathbf{Z}_m^*$ med $yx = 1 \pmod{m}$ (dvs. at x har en multiplikativ invers i \mathbf{Z}_m^*), for når $\gcd(x, m) = 1$ da fins heltall a og b slik at $ax + bm = 1$, og ved å sette y lik det tallet i \mathbf{Z}_m som er lik $a \pmod{m}$, får vi: $yx = (a + km)x = ax + kxm = 1 - bm + kxm = 1 + (kx - b)m$, dvs. $yx = 1 \pmod{m}$. (**Oppgave:** Innse at y må være i \mathbf{Z}_m^* , dvs. at $\gcd(y, m) = 1$.)

Merknad 2.2.1 Når m er primtall, da er $\mathbf{Z}_m^* = \{1, \dots, m-1\}$.

Vi trenger noen flere generelle begreper for å komme videre:

Definisjon 2.2.3 Vi forutsetter her at $(G, *)$ er en gruppe. $(G, *)$ kalles *endelig* hvis G er en endelig mengde, og da sier vi at antall elementer i G er gruppas *orden*. Hvis H er en delmengde av G , med $e \in H$, $x \in H \Rightarrow x^{-1} \in H$ og

$x, y \in H \Rightarrow x * y \in H$, da er $(H, *)$ også en gruppe, og den kalles en *undergruppe* av $(G, *)$. Hvis $a \in G$ er slik at ethvert element i G er lik en heltallspotens av a , da sier vi at G er en *syklisk* gruppe, *generert* av a (og a kalles en *generator* for G). Hvis $a \in G$ er slik at $a^n = e$ for et heltall $n > 0$, da sier vi at a har *orden* lik det minste slike heltall n . Hvis $a \in G$ er slik at $a^n \neq e$ for alle heltall $n > 0$, da sier vi at a har *uendelig orden*.

Her er noen eksempler:

Gruppen $(\mathbf{Z}_m, +)$ har orden m . Den er syklisk, generert av 1.

Gruppen (\mathbf{Z}_m^*, \cdot) har orden $\phi(m)$ (jf. Definisjon 1.4.3 på side 11). Den er syklisk for visse verdier av m , men ikke for alle (se oppgave 2.2.2 nedenfor).

$(\mathbf{Z}, +)$ er en syklisk undergruppe (generert av 1) av $(\mathbf{Q}, +)$, som videre er en undergruppe av $(\mathbf{R}, +)$. De to sistnevnte gruppene er ikke sykliske. (Er det opplagt? Kan det begrunnes? Hvordan?)

Oppgave 2.2.2 Sett opp komplette "gangetabeller" for (\mathbf{Z}_8^*, \cdot) og (\mathbf{Z}_9^*, \cdot) , og finn derved ut at nøyaktig én av disse er syklisk, og finn ut hvilke elementer som er generator for denne.

Vi har vært inne på flere stikkord for egenskaper som grupper, og deres elementer, kan ha: En gruppe kan være abelsk eller ikke-abelsk (Def. 2.2.2), av endelig eller uendelig orden (Def. 2.2.3), syklisk eller ikke-syklisk (Def. 2.2.3). Et element i en gruppe kan ha endelig eller uendelig orden (Def. 2.2.3), og det kan være en generator for gruppen. Slike stikkord kan bidra til at en kan bli "kjent" med ulike grupper. En annen måte å bli kjent med en gruppe på er å finne ut hvilke undergrupper den har, og hvilke egenskaper disse undergruppene har. Dette kan være vanskelig å finne ut, så det er viktig med teoremer o.l. som sier noe om hva slags undergrupper en gitt gruppe kan ha. Her er et slikt:

Teorem 2.2.1 En undergruppe av en syklisk gruppe er selv syklisk.

Bevis: Hvis $(G, *)$ er en syklisk gruppe, generert av elementet a , og H er en undergruppe av G , da er enten H triviell: $H = \{e\}$ (og da er den syklisk, generert av det nøytrale elementet e), eller så inneholder H minst ett element $b \neq e$. Siden a genererer G , og $b \in G$, er $b = a^n$ for et heltall $n \neq 0$. Siden både $b = a^n$ og $b^{-1} = a^{-n}$ da er i H (H er jo en undergruppe), så H inneholder minst én *positiv* potens av a .

Sett N lik det minste heltallet > 0 som gjør at $a^N \in H$.

Påstand: Da er a^N en generator for H , dvs. at ethvert element i H er en heltallspotens av a^N , og dermed er H syklisk.

Vi begrunner denne påstanden:

Hvis $c \in H$, da er $c \in G$, og da er c lik en heltallspotens av a : $c = a^m$ (G er jo syklisk, generert av a).

Hvis $m = 0$, da er $c = a^0 = e = (a^N)^0$.

Hvis $m > 0$, da er $m \geq N$ (siden N er det minste heltallet > 0 som gjør at $a^N \in H$). La r være den resten en får ved divisjonen $m : N$ (jf. Merknad 1.2 på side 5). Da er $m = kN + r$, der k er et heltall og $0 \leq r < N$, og vi får at $a^r = a^{m-kN} = a^m * (a^N)^{-k} \in H$. Siden N lik det *minste* heltallet > 0 som gjør at $a^N \in H$, kan ikke r være > 0 , så det følger at $r = 0$, og dermed er $c = a^m = a^{kN} = (a^N)^k$.

Til slutt: Hvis $m < 0$, da er $c = a^m = (a^{-1})^{-m}$, og samme resonnement som vi nettopp har gjennomført gjelder med a^{-1} i stedet for a og $-m$ i stedet for m .

Dette fullfører begrunnelsen for påstanden, og dermed for beviset for teoremet. \triangle

Et viktig resultat om endelige grupper (ikke nødvendigvis abelske) og deres undergrupper sier at en undergruppes orden alltid går opp i den større gruppas orden:

Teorem 2.2.2 (Lagrange)¹⁰ Hvis G er en endelig gruppe av orden n , og H er en undergruppe av G , av orden m , da er n delelig med m .

Bevis: Vi skal vise at G kan deles inn i adskilte (disjunkte) delmengder som hver har m elementer.

For hver $a \in G$, la aH være mengden av alle G -elementer som kan skrives på formen $a * b$, med $b \in H$.

Påstand 1: Delmengdene aH , med $a \in G$, utfyller tilsammen hele G .

Begrunnelse: For hver $a \in G$ gjelder at $a = a * e$, og $e \in H$. Derfor er $a \in aH$, så hver $a \in G$ er med i en av delmengdene.

Påstand 2: Delmengdene aH , med $a \in G$, er enten identiske eller adskilte (disjunkte). **Begrunnelse:** Anta at a_1 og a_2 i G er slik at a_1H og a_2H har et felles element, c , dvs. at det fins $b_1 \in H$ og $b_2 \in H$ slik at $c = a_1 * b_1 = a_2 * b_2$.

Vi skal se at da må a_1H og a_2H være samme delmengde.

For det første: Hver $d \in a_1H$ er også i a_2H , for $d = a_1 * b = c * b_1^{-1} * b = a_2 * b_2 * b_1^{-1} * b$, som er i a_2H fordi $b_2 * b_1^{-1} * b$ er i H .

For det andre: Hver $d \in a_2H$ er også i a_1H , for $d = a_2 * b = c * b_2^{-1} * b = a_1 * b_1 * b_2^{-1} * b$, som er i a_1H fordi $b_1 * b_2^{-1} * b$ er i H .

¹⁰Joseph-Louis Lagrange (1736–1813), dominerende i fransk matematikk i siste halvdel av 1700-tallet, og — i likhet med Abel — en forløper for gruppeteorien.

Påstand 3: Alle delmengdene aH har precis m elementer hver.

Begrunnelse: For fastholdt a fremkommer hvert element i aH slik: $a * b$ med $b \in H$, og ulike b 'er i H gir ulike elementer i aH , fordi

$$a * b_1 = a * b_2 \Rightarrow a^{-1} * a * b_1 = a^{-1} * a * b_2 \Rightarrow e * b_1 = e * b_2 \Rightarrow b_1 = b_2 .$$

Påstandene 1, 2 og 3 betyr at delmengdene aH , med $a \in G$, alle har m elementer, og at de utfyller hele G uten overlappinger. Derfor er n delelig med m . \triangle

Følgende er en direkte konsekvens av Lagranges teorem, fordi ethvert element av orden m genererer en undergruppe av orden m :

Teorem 2.2.3 Hvis G er en endelig gruppe av orden n , og a er et element i G , av orden m , da er n delelig med m .

Vi skal så se at Lagranges teorem kan brukes til å bevise *Fermats lille teorem*, som vi var innom i kapittelet om tallteori, på side 18. Vi gjentar teoremet her:

Teorem 2.2.4 ("Fermats lille teorem") Hvis p er et primtall, og a et heltall som ikke er delelig med p , da gjelder:

$$a^{p-1} \equiv 1 \pmod{p} ,$$

eller, med andre ord: $a^{p-1} - 1$ er delelig med p .

Bevis: Gruppen $\mathbf{Z}_p^* = \{1, 2, \dots, p-1\}$ (jf. Merknad 2.2.1), med multiplikasjon modulo p , har orden $p-1$, så hvis m er ordenen til et tall $a \in \mathbf{Z}_p^*$, da er $p-1$ delelig med m , dvs. at $p-1 = mq$ for et eller annet heltall q (her bruker vi Teorem 2.2.3). Da følger det at $a^{p-1} = a^{mq} = (a^m)^q = 1^q = 1$ i \mathbf{Z}_p^* (hvordan vet vi at $a^m = 1$ i \mathbf{Z}_p^* ?), dvs. at $a^{p-1} - 1$ er delelig med p . Når dette gjelder for alle tall $a \in \mathbf{Z}_p^*$, dvs. for alle tall $a \in \mathbf{Z}_p$ unntatt 0, da gjelder det for alle tall a , generelt, unntatt de som er 0 modulo p , dvs. unntatt de som er delelige med p .

\triangle

2.3 Ringer

Mens grupper har bare én regneoperasjon, skal vi nå se på algebraiske strukturer der vi har et samspill av *to* regneoperasjoner:

Definisjon 2.3.1 En *ring* er en mengde (ringens *grunnmengde*, R) med to assosiative operasjoner, vanligvis omtalt som *addisjon* og *multiplikasjon*, med regnetegn hhv. $+$ og $*$, på en slik måte at

(1) $(R, +)$ er en abelsk gruppe (det nøytrale elementet for denne betegnes da "0"),

(2) det fins et nøytralt element for " $*$ " (dette betegnes "1", så kravet til dette er at $x * 1 = 1 * x = x$ for alle $x \in R$), og

(3) multiplikasjonen er *distributiv* over addisjonen (det betyr at $x * (y + z)$ skal være lik $x * y + x * z$, og $(y + z) * x$ skal være lik $y * x + z * x$, for alle $x, y, z \in R$).

Hvis $(R, +, *)$ er en ring der $x * y = y * x$ for alle $x, y \in R$, da kalles den en *kommutativ* ring.

Som vanlig er det ikke symbolene (her: R , $+$ og $*$) som er viktige — de vil i eksempler skiftes ut med andre symboler som av ulike grunner kan passe bedre —, men det er de formelle egenskapene til regneoperasjonene som er viktige. Disse er uttrykt i Definisjon 2.3.1 over, og disse egenskapene vil være fellestrekk for de ulike eksemplene. Utover disse felles egenskapene vil eksemplene være ulike m.h.p. andre spørsmål knyttet til regneoperasjonene, og det er ofte slike ulikheter vi vil ønske å fokusere på. Her er noen eksempler:

Merknad 2.3.1 I en del lærebøker kreves det ikke at alle ringer skal ha nøytralt element for " $*$ ", men det krever vi altså her.

Eksempel 2.3.1 Tallmengdene \mathbf{Z} , \mathbf{Q} og \mathbf{R} , alle med operasjonene $+$ og \cdot , er ringer. Det er også \mathbf{Z}_m , for alle positive heltall m . Alle disse er dessuten kommutative ringer. (Er dette noe vi egentlig vet fra før? Er det opplagte fakta? Har vi vært innom det som trengs for å underbygge disse påstandene? Hvis så, hvor?)

Neste eksempel dreier seg om *matriser*, som er viktige i den delen av matematikken som kalles *lineær algebra*. Matriser vil ikke bli brukt i det etterfølgende, så dette eksempelet kan godt spares til en senere gjennomlesing.

Eksempel 2.3.2 Mengden $M_2(\mathbf{R})$ av alle reelle 2×2 -matriser, dvs. av alle slike "firkant-oppstillinger" av reelle tall: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, der addisjon og multiplikasjon er definert slik:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} w & x \\ y & z \end{bmatrix} \stackrel{\text{def}}{=} \begin{bmatrix} a + w & b + x \\ c + y & d + z \end{bmatrix}$$

$$\text{og } \begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} w & x \\ y & z \end{bmatrix} \stackrel{\text{def}}{=} \begin{bmatrix} aw + by & ax + bz \\ cw + dy & cx + dz \end{bmatrix},$$

er en ring med $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ som additivt nøytralt element, og $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ som multiplikativt nøytralt element. (**Oppgave:** Sjekk at $M_2(\mathbf{R})$ er en ring!)

$M_2(\mathbf{R})$ er ikke en kommutativ ring, for vi har f.eks. at

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} * \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 + 2 \cdot 7 & 1 \cdot 6 + 2 \cdot 8 \\ 3 \cdot 5 + 4 \cdot 7 & 3 \cdot 6 + 4 \cdot 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix},$$

$$\text{mens } \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} * \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 5 \cdot 1 + 6 \cdot 3 & 5 \cdot 2 + 6 \cdot 4 \\ 7 \cdot 1 + 8 \cdot 3 & 7 \cdot 2 + 8 \cdot 4 \end{bmatrix} = \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix}.$$

Oppgave 2.3.1 Regn ut $\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} * \begin{bmatrix} 3 & 2 \\ 4 & 2 \end{bmatrix}$ og $\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} * \begin{bmatrix} 3 & 4 \\ 4 & 3 \end{bmatrix}$ i ringen $M_2(\mathbf{R})$ (jf. Eksempel 2.3.2 ovenfor).

Merknad 2.3.2 Innse at svarene på Oppgave 2.3.1 ovenfor betyr at vi i ringer kan ha $a * b = a * c$ samtidig som $b \neq c$ og $a \neq 0$. Men hvis a har en *multiplikativ invers* i ringen, dvs. hvis det fins et element d som er slik at $a * d = d * a = 1$, da kan vi resonnerer slik:

$$a * b = a * c \Rightarrow d * a * b = d * a * c \Rightarrow 1 * b = 1 * c \Rightarrow b = c.$$

Dette må bety, i sammenheng med matrisene i Oppgave 2.3.1 ovenfor, at $\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix}$ ikke har noen multiplikativ invers i ringen $M_2(\mathbf{R})$.

Problemet med eksistens av multiplikative inverser er viktig i forbindelse med ringer, så vi innfører et nytt ord i den forbindelse:

Definisjon 2.3.2 Et element i en ring $(R, +, *)$ som har en multiplikativ invers kalles en *enhet* i R . Mengden av alle enheter i R skrives R^* . Merk at $(R^*, *)$ er en gruppe. Den kalles *enhetsgruppa* til ringen $(R, +, *)$.

Oppgave 2.3.2 Hvordan passer symbolbruken i definisjonen over med den som ble brukt i Eksempel 2.2.1 og i Merknad 2.2.1?

Hva er enhetene i ringen $(\mathbf{Z}_m, +, \cdot)$? Det er de tallene $x \in \mathbf{Z}_m$ som er slik at det fins $y \in \mathbf{Z}_m = \{0, 1, \dots, m-1\}$ med $x \cdot y = 1$ modulo m , dvs. at det fins y slik at $xy - 1$ er delelig med m , dvs. at det fins y og q slik at $xy - 1 = qm$, dvs. at det fins y og q slik at $xy + (-q)m = 1$, dvs. at $\gcd(x, m) = 1$ (jf. Merknad 1.4.2 og Teorem 1.4.1 på side 9).

Enhetene i ringen $(\mathbf{Z}_m, +, \cdot)$ er altså de tallene i \mathbf{Z}_m som er relativt primiske med m . Antallet slik er gitt ved Eulers fi-funksjon, $\phi(m)$ (jf. Definisjon 1.4.3 på side 11). Enhetsgruppen \mathbf{Z}_m^* har derfor orden $\phi(m)$.

Merk at dette bekrefter at symbolbruken i Eksempel 2.2.2 på side 22 er i overensstemmelse med den i Definisjon ref142 ovenfor.

For hvilke m er \mathbf{Z}_m^* en syklisk gruppe? Dette er ikke enkelt å finne ut av, og vi gir svaret her uten bevis:

Teorem 2.3.1 \mathbf{Z}_m^* er en syklisk gruppe hvis og bare hvis m er enten 2 eller 4 eller en potens av et odde primtall, p^k , eller det dobbelte av en potens av et odde primtall, $2p^k$.

Merk at dette betyr at \mathbf{Z}_p^* er en syklisk gruppe når p er primtall. Da er også alle undergruppene sykliske (jf. Teorem 2.2.1 på side 23). Hver undergruppes orden må gå opp i $p-1$, som er ordenen til \mathbf{Z}_p^* (jf. Teorem 2.2.2, side 24). Hver undergruppe vil også være syklisk, generert av et element (tall) i \mathbf{Z}_p^* . Dette gir oss en mulighet til å identifisere alle undergruppene til \mathbf{Z}_p^* . Vi avslutter avsnittet om ringer ved å se nærmere på eksempelet \mathbf{Z}_{19}^* :

Eksempel 2.3.3 Undergruppene i \mathbf{Z}_{19}^* : Hver undergruppe i \mathbf{Z}_{19}^* er generert av et (eller flere) tall med orden som går opp i $19-1 = 18$, dvs. med orden lik 1, 2, 3, 6, 9 eller 18. Vi ser nærmere på ordenene til hvert tall i \mathbf{Z}_{19}^* :

1 har orden 1 (Det er vel opplagt?)

2 har orden 18 (og genererer dermed hele \mathbf{Z}_{19}^*), fordi 19 går opp i $2^{18} - 1$ (iflg. Teorem 2.2.4, side 25), og ikke i $2^k - 1$ for $k \in \{1, \dots, 17\}$ (dette må sjekkes for hver av de 17 verdiene av $2^k - 1$).

3 har også orden 18 (og genererer dermed hele \mathbf{Z}_{19}^*), fordi 19 går opp i $3^{18} - 1$ (iflg. Teorem 2.2.4, side 25), og ikke i $3^k - 1$ for $k \in \{1, \dots, 17\}$ (dette må sjekkes for hver av de 17 verdiene av $3^k - 1$).

4 har orden 9 (og genererer dermed en ekte undergruppe av \mathbf{Z}_{19}^*), fordi 19 går opp i $4^9 - 1 = 2^{19} - 1$, og ikke i $4^k - 1$ for $k \in \{1, \dots, 8\}$ (se kommentarene ovenfor om 2).

5 har også orden 9, fordi 19 går opp i $5^9 - 1 = 19 \cdot 102796$, og ikke i $5^k - 1$ for $k \in \{1, \dots, 8\}$.

6 har også orden 9, fordi 19 går opp i $6^9 - 1 = 19 \cdot 530405$, og ikke i $6^k - 1$ for $k \in \{1, \dots, 8\}$.

7 har orden 3 (og genererer dermed en ekte undergruppe av \mathbf{Z}_{19}^*), fordi 19 går opp i $7^3 - 1 = 19 \cdot 18$, og ikke i $7^k - 1$ for k lik 1 eller 2.

8 har orden 6 fordi 2 har orden 18 (se kommentarene ovenfor om 2).

9 har orden 9 fordi 3 har orden 18 (se kommentarene ovenfor om 3).

10 har orden 18 (og genererer dermed hele \mathbf{Z}_{19}^*), fordi 19 går opp i $10^{18} - 1$ og ikke i $10^k - 1$ for $k \in \{1, \dots, 17\}$ (alt dette må sjekkes numerisk).

11 har orden 3 (og genererer dermed en ekte undergruppe av \mathbf{Z}_{19}^*), fordi 19 går opp i $11^3 - 1 = 19 \cdot 70$, og ikke i $11^k - 1$ for k lik 1 eller 2.

12 har orden 6 (og genererer dermed en ekte undergruppe av \mathbf{Z}_{19}^*), fordi 19 går opp i $12^6 - 1 = 19 \cdot 157157$, og ikke i $12^k - 1$ for $k \in \{1, \dots, 5\}$.

13 har orden 18 (og genererer dermed hele \mathbf{Z}_{19}^*), fordi 19 går opp i $13^{18} - 1$, og ikke i $13^k - 1$ for $k \in \{1, \dots, 17\}$ (alt dette må sjekkes for de aktuelle verdiene av k).

14 har også orden 18 (og genererer dermed hele \mathbf{Z}_{19}^*), fordi 19 går opp i $14^{18} - 1$, og ikke i $14^k - 1$ for $k \in \{1, \dots, 17\}$ (alt dette må sjekkes for de aktuelle verdiene av k).

15 har også orden 18 (og genererer dermed hele \mathbf{Z}_{19}^*), fordi 19 går opp i $14^{18} - 1$, og ikke i $15^k - 1$ for $k \in \{1, \dots, 17\}$ (alt dette må sjekkes for de aktuelle verdiene av k).

16 har orden 9 (og genererer dermed en ekte undergruppe av \mathbf{Z}_{19}^*), fordi 19 går opp i $16^9 - 1 = 2^{36} - 1$, og ikke i $16^k - 1$ for $k \in \{1, \dots, 8\}$.

17 har også orden 9 (og genererer dermed en ekte undergruppe av \mathbf{Z}_{19}^*), fordi 19 går opp i $17^9 - 1 = 19 \cdot 6241467184$, og ikke i $17^k - 1$ for $k \in \{1, \dots, 8\}$.

18 har orden 2 (og genererer dermed en ekte undergruppe av \mathbf{Z}_{19}^*), fordi 19 går opp i $18^2 - 1 = 19 \cdot 17$, og ikke i $18^1 - 1 = 17$.

Av dette ser vi at \mathbf{Z}_{19}^* har:

1 undergruppe av orden 1 (det er selvsagt den trivielle undergruppen, $\{1\}$),

1 av orden 2 (den som er generert av 18, og som derfor er delmengden $\{1, 18\}$ av \mathbf{Z}_{19}^*),

høyst 2 av orden 3: en generert av 7 og en generert av 11, men nærmere undersøkelse viser at 7 og 11 genererer samme undergruppe, nemlig delmengden $\{1, 7, 11\}$ av \mathbf{Z}_{19}^* , (**Oppgave:** Vis at $7^2 = 11$ og at $11^2 = 7$ i \mathbf{Z}_{19}^* .)

høyst 2 av orden 6: en generert av 8 og en generert av 12, men nærmere undersøkelse viser at 8 og 12 genererer samme undergruppe, nemlig delmengden $\{1, 7, 8, 11, 12, 18\}$ av \mathbf{Z}_{19}^* , (**Oppgave:** Vis dette.)

høyst 6 av orden 9: en generert av 4, en av 5, en av 6, en av 9, en av 16 og en av 17, men nærmere undersøkelse viser at alle disse genererer samme undergruppe, nemlig delmengden $\{1, 4, 5, 6, 7, 9, 11, 16, 17\}$ av \mathbf{Z}_{19}^* , og

høyst 6 undergrupper av orden 18, men \mathbf{Z}_{19}^* har jo bare en undergruppe av orden 18, nemlig \mathbf{Z}_{19}^* selv, så vi får med "på kjøpet" at \mathbf{Z}_{19}^* har 6 generatorer: 2, 3, 10, 13, 14 og 15.

2.4 Kropper

Definisjon 2.4.1 En *kropp* (engelsk: *field*) er en kommutativ ring der alle elementer unntatt 0 er en enhet (dvs. har en multiplikativ invers).

Merknad 2.4.1 I enhver kropp gjelder følgende: Hvis $ab = 0$, da er enten $a = 0$ eller $b = 0$ (eller begge deler). Vi vet at dette ikke gjelder generelt i ringer, for f.eks. i \mathbf{Z}_4 er $2 \cdot 2 = 0$, selv om $2 \neq 0$, men hvis vi har $ab = 0$ i en kropp, og $a \neq 0$, da følger det at $b = a^{-1}ab = a^{-1}0 = 0$.

Eksempel 2.4.1 Ringene \mathbf{Q} (de rasjonale tallene) og \mathbf{R} (de reelle tallene) er kropper, men \mathbf{Z} (de hele tallene) er ikke en kropp. \mathbf{Z}_m er en kropp hvis og bare hvis m er et primtall (jf. Merknad 2.2.1).

Definisjon 2.4.2 En kropp har *karakteristikk 0* hvis 0 ikke er lik en endelig sum av 1'ere. For kropper som ikke har karakteristikk 0 er karakteristikken lik det minste antall 1'ere som har sum lik 0.

Eksempel 2.4.2 Kroppene \mathbf{Q} (de rasjonale tallene) og \mathbf{R} (de reelle tallene) har karakteristikk 0. Kroppen \mathbf{Z}_p (der p er et primtall) har karakteristikk p .

Merknad 2.4.2 Hvis en kropp har karakteristikk $m \neq 0$, da er m et primtall, for hvis m er et produkt av to positive heltall: $m = pq$, da kan "summen av m 1'ere skrives slik: $m1 = pq1 = p1 \cdot q1$ (er dette gyldig regning i ringer generelt??), og siden dette er lik 0, må enten $p1$ eller $q1$ være lik 0 (jf. Merknad 2.4.1 ovenfor). Siden m er det *minste* antall 1'ere som har sum lik 0, følger det at p og q ikke begge kan være $< m$. Altså er en av dem lik m , og da må den andre være lik 1. Av dette følger det at m er et primtall.

2.5 Polynomringer

Her skal X brukes som en formell variabel, som ikke i utgangspunktet skal tolkes som symbol for noe bestemt tall eller annet matematisk objekt. Vi skal ”regne” med potenser av X slik: $X^n X^m = X^{n+m}$, der vi forutsetter at eksponentene n og m er heltall ≥ 0 . Førstepotenser, X^1 , skrives vanligvis bare som X .

Definisjon 2.5.1 Med et *polynom* i X over en kommutativ ring R menes et sumuttrykk med endelig mange ledd, som hver er et ”produkt” (en sammenstilling) av et element i R (ofte betegnet med små bokstaver i starten av alfabetet: a, b osv., eller evt. nummerert: a_0, a_1, a_2 osv., etter behov) og en potens X^n av X , der n er heltall ≥ 0 som vanligvis varierer fra ledd til ledd. R -elementet kalles da en *koeffisient*, og n kalles *leddets grad*. I ledd av grad 0 sløyfer vi vanligvis X^0 , og skriver koeffisienten alene, som et *konstantledd*. (Hvis et polynom har flere ledd med samme grad, da kan disse trekkes sammen til ett, ved at ” X -delen” behandles som en ”felles faktor”, som kan ”settes utenfor”, f.eks. slik: $aX^5 + bX^5 + cX^5 = (a + b + c)X^5$. Vi vil vanligvis forutsette at dette alltid er gjort fullt ut, slik at leddene i et polynom har ulike grader.)

Den høyeste graden som forekommer blant leddene i polynomet (egentlig: blant de leddene som har koeffisient $\neq 0$) kalles *polynomets grad*. Koeffisienten i leddet med høyest grad kalles polynomets *ledende koeffisient*. Polynomer med ledende koeffisient lik 1 kalles *moniske*.

Mengden av alle polynomer i X over R betegnes med $R[X]$. Denne er selv en kommutativ ring, når addisjon og multiplikasjon utføres på ”vanlig polynommanér”, som illustrert ved følgende eksempler:

$$(aX^2 + bX + c) + (dX^3 + eX^2 + fX + g) = dX^3 + (a + e)X^2 + (b + f)X + (c + g),$$

og

$$\begin{aligned} (aX^2 + bX + c)(dX^3 + eX^2 + fX + g) &= \\ &= aX^2(dX^3 + eX^2 + fX + g) + bX(dX^3 + eX^2 + fX + g) + c(dX^3 + eX^2 + fX + g) = \\ &= adX^5 + aeX^4 + afX^3 + agX^2 + bdX^4 + beX^3 + bfX^2 + bgX + cdX^3 + ceX^2 + \\ &cfX + cg = \\ &= adX^5 + (ae + bd)X^4 + (af + be + cd)X^3 + (ag + bf + ce)X^2 + (bg + cf)X + cg. \end{aligned}$$

Merk at all regning med X -potenser her er ”formell potensregning”, mens all regning med koeffisientene er addisjon og multiplikasjon slik disse skal utføres i ringen R .

Oppgave 2.5.1 Vis at $R[X]$ er en kommutativ ring.

I det følgende skal vi se nærmere på polynomringer over *kropper*, og spesielt over de endelige kroppene \mathbf{Z}_p (der p er et primtall). Vi starter med noen eksempler

på multiplikasjon i $\mathbf{Z}_p[X]$ for noen enkle (dvs. lave) primtall p . Merk at ethvert utregnet produkt også gir oss en *faktorisering* (av svaret).

Eksempel 2.5.1 I $\mathbf{Z}_2[X]$ har vi bare 0 og 1 som mulige koeffisienter, og $1+1=0$, så vi kan regne slik:

$$\begin{aligned}(X+1)(X+1) &= X^2 + (1+1)X + 1 = X^2 + 1, \\(X^2+X+1)(X+1) &= X^3 + (1+1)X^2 + (1+1)X + 1 = X^3 + 1, \\(X^2+X+1)(X^2+X+1) &= X^4 + (1+1)X^3 + (1+1+1)X^2 + (1+1)X + 1 = \\X^4 + X^2 + 1, \text{ og} \\(X^4+X+1)(X^3+X+1) &= X^7 + X^5 + (1+1)X^4 + X^3 + X^2 + (1+1)X + 1 = \\X^7 + X^5 + X^3 + X^2 + 1.\end{aligned}$$

Eksempel 2.5.2 I $\mathbf{Z}_3[X]$ har vi 0, 1 og 2 som mulige koeffisienter, og $1+1=2$, $1+2=0$, $2+2=1$ og $2 \cdot 2=1$ (er det virkelig slik??), så vi kan regne slik:

$$\begin{aligned}(X+1)(X+2) &= X^2 + (2+1)X + 2 = X^2 + 2, \\(X^2+X+1)(X+2) &= X^3 + (2+1)X^2 + (2+1)X + 2 = X^3 + 2, \\(X^2+X+1)(X^2+2X+2) &= X^4 + (2+1)X^3 + (2+2+1)X^2 + (2+2)X + 2 = \\X^4 + 2X^2 + X + 2, \text{ og} \\(X^4+X+2)(X^3+2X+1) &= X^7 + 2X^5 + (1+1)X^4 + 2X^3 + 2X^2 + (1+2 \cdot 2)X + 2 = \\X^7 + 2X^5 + 2X^4 + 2X^3 + 2X^2 + 2.\end{aligned}$$

Definisjon 2.5.2 Et polynom i $F[X]$ (der F er en kropp) kalles *irreducibelt over F* hvis det ikke har noen *ekte faktorisering*, dvs. hvis det ikke kan skrives som et produkt av to (eller flere) polynomer i $F[X]$ som begge (alle) har grad > 0 . (**Merk** at hver faktor i en ekte faktorisering har lavere grad enn det faktoriserte polynomet. **Oppgave:** Er det opplagt? Kan det vises? Hvordan?)

Merk at eksemplene ovenfor gir oss noen ekte faktoriseringer som avviker fra faktoriseringer i $\mathbf{R}[X]$, bl.a. følgende:

$$\begin{aligned}\text{I } \mathbf{Z}_2[X] \text{ er } X^2 + 1 &= (X+1)(X+1) \text{ og } X^3 + 1 = (X^2 + X + 1)(X+1), \text{ og} \\ \text{i } \mathbf{Z}_3[X] \text{ er } X^2 + 2 &= (X+1)(X+2) \text{ og } X^3 + 2 = (X^2 + X + 1)(X+2).\end{aligned}$$

Her har vi ”faktorisert” ved å snu på noen allerede utførte multiplikasjoner. Det er langt vanskeligere å faktorisere polynomer som vi ikke kjenner igjen som resultat av en kjent multiplikasjon, og å finne ut om et gitt polynom er irreducibelt. I $\mathbf{R}[X]$ fins det ingen generell faktoriseringmetode, men hvis F er en *endelig* kropp (som f.eks. $\mathbf{Z}_p[X]$), da kan vi i prinsippet gjennomgå alle mulige forslag til faktorisering, fordi det ”bare” er endelig mange muligheter. I praksis kan denne metoden være uoverkommelig tidkrevende, men her er et eksempel som illustrerer den:

Eksempel 2.5.3 Vi vil prøve å faktorisere det moniske polynomet $X^3 + X + 1$ i $\mathbf{Z}_2[X]$:

Hvis det fins noen ekte faktorisering av dette, da kan det skrives som et produkt av et 1.-grads og en 2.-grads polynom. Vi kan uten videre gå ut fra at hver av faktorene også er moniske. (**Oppgave:** Kan vi virkelig det? Kan det ikke finnes ekte faktoriseringer der de enkelte faktorene ikke er moniske?) Det fins ikke veldig mange 1.-grads moniske polynomer i $\mathbf{Z}_2[X]$ — faktisk er det bare to: X og $X + 1$. Av 2.-grads moniske polynomer i $\mathbf{Z}_2[X]$ fins det fire: X^2 , $X^2 + X$, $X^2 + 1$ og $X^2 + X + 1$ (er dette virkelig alle??). Dermed fins det bare åtte forskjellige produkter av et 1.-grads og en 2.-grads polynom i $\mathbf{Z}_2[X]$, og vi regner nå ut alle åtte:

$$X \cdot X^2 = X^3,$$

$$X(X^2 + X) = X^3 + X^2,$$

$$X(X^2 + 1) = X^3 + X,$$

$$X(X^2 + X + 1) = X^3 + X^2 + X,$$

$$(X + 1) \cdot X^2 = X^3 + X^2,$$

$$(X + 1)(X^2 + X) = X^3 + (1 + 1)X^2 + X = X^3 + X,$$

$$(X + 1)(X^2 + 1) = X^3 + X^2 + X + 1, \text{ og}$$

$$(X + 1)(X^2 + X + 1) = X^3 + (1 + 1)X^2 + (1 + 1)X + 1 = X^3 + 1.$$

Vi ser her at ingen av disse ga $X^3 + X + 1$ som svar. Dermed vet vi at dette polynomet ikke *kan* skrives som et produkt av et 1.-grads og et 2.-grads polynom. Derfor er $X^3 + X + 1$ irreducibelt over \mathbf{Z}_2 . Vi ser videre at $X^3 + X^2 + 1$ også er irreducibelt over \mathbf{Z}_2 , og at disse to er *alle* irreducible 3.-grads polynomer i $\mathbf{Z}_2[X]$.

Oppgave 2.5.2 Bruk denne metoden med komplett gjennom søking til å finne alle irreducible 4.-grads polynomer i $\mathbf{Z}_2[X]$, og alle irreducible 3.-grads polynomer i $\mathbf{Z}_3[X]$. Fins det noen irreducible 2.-grads polynomer i $\mathbf{Z}_2[X]$?

3 Sannsynlighetsregning

3.1 Innledning

Sannsynlighetsteorien gir ingen forklaring på hva sannsynlighet *er*, utover dette:

enhver sannsynlighet er et reelt tall som er ≥ 0 og ≤ 1 ,

samt en del regnetekniske "fakta" som er ansett som opplagte og nødvendige i forbindelse med *regning* med sannsynligheter. I neste avsnitt skal vi gå løs på dette, med mengdelæren som redskap. Bruk neste kapittel til repetisjon av mengdelæren, etter behov.

3.2 Sannsynlighetsmodeller

Her er noen eksempler på *stokastiske forsøk*:

1. Det skal gjøres en del kast med en vanlig spillterning. Hver gang skal resultatet (antall "øyne") noteres.
2. Det skal gjøres uttrekk av ett kort fra en godt stokket kortstokk. Ved tilbakelegging og omstokking gjøres dette en del ganger. Hver gang noteres hvilket kort som ble trukket.
3. Det skal gjøres kast med en vanlig mynt, om nødvendig flere ganger, inntil første gang den viser "kron". Antall kast som trengtes blir notert. Dette gjentas en del ganger.
4. UP setter opp en radarkontroll et sted der fartsgrensen er 50 km/time, og noterer farten til hver bil som passerer i en viss periode, målt i km/time.
5. I en klasse på HiG noterer læreren, etter hver time med klassen, navnet på den studenten som kom først inn til timen.

Blant disse fem eksemplene er det noen viktige felles trekk, og noen forskjeller. Det er likhetene som er viktigst, og først og fremst disse:

Alle har en viss uforutsigbarhet, på den måten at en ikke kan vite hva neste noterte resultat vil bli (det er derfor vi bruker uttrykket *stokastiske forsøk*), og alle har en viss forutsigbarhet, på den måten at en vet hva de *mulige* resultatene er: I (1) etter nummereringen ovenfor er det tallene 1, 2, ..., 6, i (2) er det de 52 kort-navnene kløver-to, ruter-to, ..., spar-ess, i (3) er det de positive hele tallene 1, 2, 3, ... (vi kan ikke sette noen øvre grense her, for hva skulle den være?), i (4) er det kanskje ikke helt klart, men hvis vi sier "de reelle tallene > 0 " da har vi ihvertfall fått med alle muligheter, og i (5) er det navnelisten for hele klassen.

Det vi her har omtalt som ”resultater” blir i sannsynlighetsregningen ofte kalt *utfall*. Mengden av alle de mulige utfallene i et stokastisk forsøk kalles derfor *utfallsrommet* til forsøket. Den betegnes oftest med den store bokstaven S .

Delmengder av S vil vi omtale som *hendelser*, ut fra tankegangen i følgende eksempler (igjen med henvisninger til de nummererte eksemplene ovenfor):

1. Eksempel på en hendelse: Neste kast gir et primtall. Dette svarer til delmengden $\{2, 3, 5\}$ av $S = \{1, 2, 3, 4, 5, 6\}$.
2. Eksempel på en hendelse: Neste trekk gir et bildekort. Dette svarer til delmengden $\{\text{kløver-knekt}, \dots, \text{spar-konge}\}$ av S .
3. Eksempel på en hendelse: I neste forsøk trengs flere enn 5 kast. Dette svarer til delmengden $\{6, 7, 8, \dots\}$ av $S = \{1, 2, 3, \dots\}$.
4. Eksempel på en hendelse: Neste bil holder lovlig fart. Dette svarer til delmengden $\langle 0, 50]$ av $S = \langle 0, \infty$.
5. Eksempel på en hendelse: I neste time er det en jente som kommer først. Dette svarer til delmengden av S bestående av alle jentenavn i klassen.

For et stokastisk forsøk kan en tanke seg at hver enkelt hendelse (delmengde av utfallsrommet S) har en viss sannsynlighet (et reelt tall ≥ 0 og ≤ 1). Hvordan en finner (beregner) slike ”hendelses-sannsynligheter” skal vi komme tilbake til, men først skal vi konsentrere oss om noen grunnleggende krav som alltid skal gjelde i sannsynlighetsregning:

Definisjon 3.2.1 En *sannsynlighetsmodell* (forkortes ofte til ”s.s.-modell”) for et stokastisk forsøk skal bestå av en mengde S (kalt modellens *utfallsrom*), samt reelle tall $P(A)$ knyttet til hver *hendelse* A (dvs. delmengde $A \subset S$) som oppfyller følgende krav:

- (i) $0 \leq P(A) \leq 1$ for alle $A \subset S$,
- (ii) $P(S) = 1$ og $P(\emptyset) = 0$ ¹¹, og
- (iii) hvis A og B er disjunkte¹² hendelser, da skal $P(A \cup B)$ være lik $P(A) + P(B)$.¹³

¹¹ \emptyset , den tomme mengden, er omtalt på side 59.

¹²Det betyr at $A \cap B = \emptyset$.

¹³Hvis S er en uendelig mengde, da må dette kravet forsterkes slik: Hvis A_1, A_2, A_3, \dots er parvis disjunkte hendelser, da skal $P(A_1 \cup A_2 \cup A_3 \cup \dots)$ være lik $P(A_1) + P(A_2) + P(A_3) + \dots$

Eksempel 3.2.1 Hvis vi i tilfellet med kast med én terning, med $S = \{1, 2, 3, 4, 5, 6\}$, for hver hendelse $A \subset S$ setter $P(A) = \frac{\text{antall tall i } A}{6}$, da har vi en s.s.-modell, som gjerne oppfattes som modellen for en ”rettferdig terning”. Denne s.s.-modellen innebærer f.eks. at

$$P(\text{primtall ved kast med én terning}) = P(\{2, 3, 5\}) = \frac{3}{6} = \frac{1}{2}.$$

Eksempel 3.2.2 Hvis vi i tilfellet med uttrekk av ett kort fra en godt stokket kortstokk, med S som beskrevet ovenfor, for hver hendelse $A \subset S$ setter $P(A) = \frac{\text{antall kort i } A}{52}$, da har vi en s.s.-modell, som gjerne oppfattes som modellen for et ”rettferdig kort-trekk”. Denne s.s.-modellen innebærer f.eks. at

$$P(\text{billedkort}) = P(\{\text{kløver-knekt}, \dots, \text{spar-konge}\}) = \frac{12}{52} = \frac{3}{13}.$$

Definisjon 3.2.2 For forsøk der S er en endelig mengde, med n elementer, vil formelen $P(A) = \frac{\text{antall elementer i } A}{n}$ gi opphav til en s.s.-modell, som kalles den *uniforme modellen* for forsøket. Ofte — men ikke alltid — oppfattes denne som ”naturlig” eller ”korrekt” for forsøket.

I situasjoner der det er korrekt å bruke en uniform modell, bli beregning av sannsynligheter ”reduert” til telling: For å beregne $P(A)$ må vi telle opp antall elementer i mengden A . Dette kan virke enkelt, men telling er ofte vanskelig. Læren om tellemetoder er en del av det området av matematikken som kalles *kombinatorikk*, og er tema for kapittelet med dette navnet. I neste eksempel bruker vi noe av dette:

Eksempel 3.2.3 Hvis en klasse med 23 jenter og 7 gutter velger en festkomité på 5 stk., ved ren tilfeldig trekning fra en boks med 30 navnelapper, hva er da sannsynligheten for at resultatet blir en ren jente-komité?

Her er det fornuftig å la utfallsrommet S bestå av alle mulige utvalg på 5 navn fra de 30 mulige. Siden rekkefølgen av de 5 som trekkes ut ikke har betydning for problemstillingen, regner vi utvalgene som *uordnede*, på den måten at vi i S bare har én liste med hvert 5-navns-utvalg (se side 52 om forskjellen på ordnede og uordnede utvalg). Hvor mange uordnede 5-navns-utvalg fins blant 30 mulige? Svaret på dette finner vi i Teorem 4.0.7 på side 53. Det er gitt ved en s.k. *binomial-koeffisient*, slik:

$$n = \binom{30}{5} = \frac{30 \cdot 29 \cdot 28 \cdot 27 \cdot 26}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 142506.$$

Antallet rene jentelister er gitt med samme formel, men med 23 i stedet for 30:

$$\text{Antall rene jentelister} = \binom{23}{5} = \frac{23 \cdot 22 \cdot 21 \cdot 20 \cdot 19}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 33649 .$$

Siden trekningen skjer ved ren tilfeldighet, slik at ingen er favorisert fremfor andre, må alle 5-navns-utvalg ha samme sannsynlighet, og det er derfor korrekt å bruke uniform modell. Derfor er sannsynligheten for at resultatet blir en ren jente-komit e like forholdet mellom de to tallene vi har funnet:

$$P(\text{Ren jente-komit e}) = \frac{33649}{142506} \approx 0,2361 .$$

Eksempel 3.2.4 Hva er sannsynligheten for     0 rette i vanlig Lotto?

Spillet g r ut p    sette kryss i 7 ruter av 28 mulige, i h p om at en treffer de samme som — eller flest mulig av — de samme rutene som trekningsmaskinen gir den etterf lgende l rdag. For   se likheten med forrige eksempel kan vi tenke oss at de 28 rutene p  kupongen har ulike person-navn: 7 guttenavn (disse representerer ”ukas rekke”, dvs. den vi h per   treffe med v re kryss) og 21 jentenavn. Hvis vi velger plasseringen av v re 7 kryss ved ren tilfeldighet, kan vi like gjerne tenke p  dette som det   trekke ut 7 av de 28 navnene tilfeldig. ”Antall rette” blir da lik antall guttenavn vi rekker ut. Sp rsm let i dette eksempelet kan dermed omskrives til: Hva er sannsynligheten for at vi trekker ut bare jentenavn? Metoden blir da den samme som i forrige eksempel, og vi f r:

$$P(0 \text{ rette}) = \frac{\binom{21}{7}}{\binom{28}{7}} = \frac{116280}{1184040} \approx 0,0982 .$$

(**Oppgave:** Er dette virkelig korrekt?)

3.3 Betinget sannsynlighet og uavhengighet

I sannsynlighetsregningen m  man v re bevisst p  betydningen av *informasjon*. Tenk deg f.eks. at du tar sjansen p  litt gambling, i en situasjon der en spill-leder utf rer kast med 3 vanlige terninger, og veddem lene dreier seg om summen av antall  yne som vises. Da vet vi at utfallsrommet m  v re: $S = \{3, 4, 5, \dots, 18\}$. Det er nok ikke korrekt   bruke en uniform modell her (med sannsynlighet 1/16 for hvert av de 16 mulige utfallene), for f.eks. sum=9 kan inntreffe p  mange flere ulike m ter enn sum=18, men en eller annen modell m  jo v re den korrekte, slik at hver hendelse $A \subset S$ har en viss sannsynlighet, $P(A)$. Tenk deg at du tar sjansen p    satse 50 kroner p  at neste kast gir ”sum > 10”, mot et l fte om

gevinst på 100 kroner hvis dette inntreffer. Så utføres kastet med de 3 terningene, men før du har fått se resultatet, sier spill-lederen: "Summen er et primtall. Du har valget mellom å trekke deg, og få igjen dine 50 kroner, eller å se hva summen ble." Hva bør du gjøre? Har spill-lederen gitt informasjon som reduserer sannsynligheten for "sum > 10"? Eller er den tvert imot blitt øket? Du får tenke deg om noen sekunder, og innser raskt at primtallene i S som er ≤ 10 er 3, 5 og 7, mens primtallene > 10 er 11, 13 og 17. Det er like mange på hver side, men betyr det at vannersjansen nå er "fifty/fifty", dvs. at $P(\text{sum} > 10)$ nå har blitt lik $1/2$? Nepe, men uansett hva som faktisk er fornuftig her, så kan vi resonnerer slik:

Hendelsen $A = \{11, 12, 13, 14, 15, 16, 17, 18\}$ har en viss sannsynlighet, $P(A)$, i utgangspunktet. Den informasjonen vi har fått er også knyttet til en hendelse, nemlig: $B = \{3, 5, 7, 11, 13, 17\}$. Informasjonen "utfallet er i B " uttrykker vi gjerne slik: " B er gitt", og "sannsynligheten for A etter at informasjonen er gitt" uttrykkes gjerne slik: "sannsynligheten for A , gitt B ", og skrives på symbolform slik: $P(A|B)$. For å finne ut mer om denne, kan vi tenke slik: Hvis A og B er disjunkte (dvs. at $A \cap B = \emptyset$), da vil det at B er gitt *utelukke* A , og vi må da ha $P(A|B) = 0$. Men vanligvis er A delvis innenfor og delvis utenfor B (slik som i spilleksempelet over). Den delen som er utenfor B er $A \setminus B$, og denne får sin sannsynlighet redusert til 0, mens den delen som er innenfor B , $A \cap B$, blir "oppgradert" på samme måte som B selv, som blir oppgradert fra $P(B)$ til 1, dvs. at dens (opprinnelige) sannsynlighet blir dividert med $P(B)$. Hvis vi reduserer sannsynlighetene til alle hendelser utenfor B til 0, og samtidig oppgraderer sannsynlighetene til alle hendelser innenfor B ved "divisjon med $P(B)$ ", da får vi faktisk en ny sannsynlighetsmodell, med nye (betingede) sannsynligheter, som kan uttrykkes i en enkelt formel, som vi ser i følgende definisjon:

Definisjon 3.3.1 Hvis A og B er hendelser i en s.s.-modell, da er *den betingede sannsynligheten* $P(A|B)$ (les: "sannsynligheten for A , gitt B ") definert ved:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} .$$

Vi ser at dette innebærer at

$$P(A \cap B) = P(A|B) \cdot P(B) .$$

I noen situasjoner er det slik at informasjon om hendelsen B ikke endrer sannsynligheten for A . Da sier vi at A er *uavhengig* av B . Dette uttrykker vi også i en definisjon:

Definisjon 3.3.2 Hendelsen A i en s.s.-modell er *uavhengig* av hendelsen B hvis

$$P(A|B) = P(A) , \quad \text{dvs. hvis} \quad P(A \cap B) = P(A) \cdot P(B) .$$

Eksempel 3.3.1 Ved kast med 2 vanlige terninger: en rød og en grønn, hva er sannsynligheten for "2 seksere"?

En *kan* lage terninger som påvirker hverandre, f.eks. ved å lime dem sammen, magnetisere dem (hvis de er av jern) eller på andre måter, men det betyr at er noe "tull" med dem. "Vanlige" terninger påvirker ikke hverandre, og derfor kan vi forutsette at hendelser som kan beskrives med bruk av bare én og én av terningene er uavhengige av hverandre. Hvis vi lar A hendelsen "rød sekser" og B hendelsen "grønn sekser" (hver av disse har selvsagt(?) s.s. lik $1/6$), da får vi:

$$P(2 \text{ seksere}) = P(A \cap B) = P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} .$$

Oppgave 3.3.1 Innse at dette "gangeprinsippet" også gjelder ved kast med flere enn 2 terninger, slik at vi ved kast med n terninger har:

$$P(n \text{ seksere}) = \left(\frac{1}{6}\right)^n = \frac{1}{6^n} .$$

3.4 Stokastiske variable

I forbindelse med s.s.-modeller er man ofte interessert i variable størrelser, som for hvert enkelt utfall har en viss tallmessig verdi, som avhenger av utfallet. Slike størrelser har mye til felles med matematiske variable, slik vi møter dem i algebra, så det er vanlig å betegne dem med bokstaven X , eller Y , Z osv. De har også noe til felles med funksjoner, siden deres verdi avhenger av noe annet, nemlig av utfallet, så vi skriver deres verdier på funksjons-manér på den måten at $X(a)$, der $a \in S$, står for "verdien X for utfallet a ". Og siden vi i forbindelse med stokastiske forsøk ikke kan beregne hvilket utfall som vil inntreffe i neste forsøk, så kan vi heller ikke beregne hvilken verdi X vil få. Derfor sier vi at X er en *stokastisk variabel*. Eksempler følger, etter følgende definisjon:

Definisjon 3.4.1 Med en *stokastisk variabel* menes en funksjon, X , definert på et utfallsrom S , som for hvert utfall $a \in S$ har et reelt tall $X(a)$ som verdi.

Her er noen eksempler:

[1] La et stokastisk forsøk bestå av en serie på 5 kast med en vanlig terning, og la utfallene være lister $(x_1, x_2, x_3, x_4, x_5)$, der x_i representerer "resultatet i kast nr. i ". Utfallsrommet består dermed av alle slike lister av fem tall, alle fra tallmengden $\{1, 2, 3, 4, 5, 6\}$. La X være definert ved: "summen av antall øyne i de fem kastene". Det er det samme som å si at

$$X(x_1, x_2, x_3, x_4, x_5) = x_1 + x_2 + x_3 + x_4 + x_5 .$$

Det er da opplagt at alle X -verdier er hele tall, og at minste mulige verdi er 5 og største mulige verdi er 30. Verdimengden til X er derfor tallmengden $\{5, 6, 7, \dots, 30\}$.

- [2]] La et stokastisk forsøk bestå av at du trekker 13 kort tilfeldig, uten tilbakelegging, fra en godt stokket kortstokk. Utfallsrommet er mengden av alle lister av 13 ulike kort: $(x_1, x_2, \dots, x_{13})$. La X være ”antall ess”. De mulige X -verdiene er da alle heltall f.o.m. 0 t.o.m. 4, så verdimengden til X er $\{0, 1, 2, 3, 4\}$.
- [3]] La ett stokastisk forsøk bestå av at 1 av de 1400 HiG-studentene velges ut tilfeldig, og la $X(a)$ være ”vekten til student a , målt i kg.” Verdimengden til X er da kanskje litt uklar, men om vi bestemmer at den skal bestå av alle de reelle tallene ≥ 0 , da er vi sikre på at vi har fått med alle muligheter.

Hvilke muligheter har vi til å beregne sannsynligheter i disse tre situasjonene? Her er noen kommentarer om nettopp det:

I [1] er det viktig å være bevisst på hva vi legger i uttrykket ”en vanlig terning”. Egentlig burde vi heller si ”en ideell terning”, for terninger i virkelighetens verden er i beste fall tilnærmet ideelle. Uttrykket er ment å representere de ideelle forutsetningene som *definerer* en ”ideelt korrekt terning”. Mer konkret betyr dette at hvert av terningens seks mulige enkeltutfall skal ha sannsynlighet lik $1/6$, og at ulike enkeltkast ikke skal påvirke (avhenge av) hverandre. Vi *forutsetter* altså uavhengighet, og da gjelder *gange-prinsippet*, på samme måte som i Oppgave 3.3.1 ovenfor, ikke bare m.h.p. resultatserien ”bare seksere”, men m.h.p. enhver forhåndsbestilt resultatserie. Med serier på 5 kast vil dermed alle utfallene (det er 6^5 av dem ialt) få samme sannsynlighet, $(\frac{1}{6})^5$, og modellen blir uniform. Da blir sannsynlighetsregningen ”reduisert” til telling. F.eks. vil sannsynligheten for å få ”sum=6” være lik antall resultatserier med sum=6 dividert med 6^5 . Dette antallet er ikke så vanskelig å finne, for ”sum=6” på 5 kast kan bare skje på den måten at ett kast gir 2 og resten 1. Siden 2’eren kan opptre i hvilket som helst av de 5 kastene, blir antallet lik 5, og vi får at

$$P(X = 6) = \frac{5}{6^5} .$$

X -verdien 6 gir en nokså enkel optelling. Det er langt mer komplisert å telle ”antall mulige måter” for X -verdier nær midten av verdimengden til X . F.eks. vil det være nesten uoverkommelig å finne antallet ulike måter å få $X = 12$ (og dermed sannsynligheten for ”sum=12”), men prøv å beregne sannsynlighetene $P(X = 7)$ og $P(X = 8)$!

I [2], hvis vi baserer oss på *uordnede utvalg*, da fins ialt $\binom{52}{13} = \frac{52 \cdot 51 \cdot 50 \cdot \dots \cdot 40}{13 \cdot 12 \cdot 11 \cdot \dots \cdot 1}$ ulike mulige utvalg på 13 kort. Hvor mange av disse består av et gitt antall (x)

ess og resten $(13 - x)$ andre kort? Svaret på dette spørsmålet finner vi slik: Blant de 4 ess'ene kan vi plukke ut x stk. på $\binom{4}{x}$ ulike måter, og blant de 48 øvrige kortene kan vi plukke ut $13 - x$ stk. på $\binom{48}{13-x}$ ulike måter. Når hvert av de førstnevnte ess-utvalgene kan kombineres med hvilket som helst av de sistnevnte ikke-ess-utvalgene, da får vi ialt $\binom{4}{x} \cdot \binom{48}{13-x}$ ulike kombinasjoner. Dermed er

$$P(X = x) = \frac{\binom{4}{x} \cdot \binom{48}{13-x}}{\binom{52}{13}} \quad \text{for } x \in \{0, 1, 2, 3, 4\}.$$

F.eks. er sannsynligheten for ”nøyaktig ett ess” lik

$$P(X = 1) = \frac{\binom{4}{1} \cdot \binom{48}{12}}{\binom{52}{13}} \approx 0,4388.$$

I [3] ovenfor har vi ingen mulighet til å regne ut sannsynligheter for bestemte X -verdier. Verdimengden til X (alle reelle tall ≥ 0) er jo her en uendelig mengde — den kan ikke engang settes opp som en uendelig liste av enkeltverdier — og da kan vi ikke gå ut fra at sannsynligheten for en hendelse (f.eks. $P(70 \leq X \leq 75)$) er lik summen av sannsynlighetene til de utfall som inngår i hendelsen. For slike stokastiske variable (de kalles *kontinuerlig fordelt*, i motsetning til i [1] og [2], der de stokastiske variablene er *diskret fordelt*, jf. Definisjon 3.4.2 nedenfor) er vanligvis alle punktsannsynlighetene lik 0, dvs. at $P(X = x) = 0$ for alle x i verdimengden til X , og likevel kan hendelses-sannsynligheter som f.eks. $P(a \leq X \leq b)$ være > 0 når $a < b$. Dette kan vi få til, regneteknisk, ved hjelp av en såkalt *tetthetsfunksjon* for X , dvs. en funksjon $f(x)$, definert på hele tallinja, \mathbf{R} , som er slik at $P(a \leq X \leq b) = \int_a^b f(x)dx$ for alle reelle tall a og b . Det vil ofte være vanskelig å finne en slik tetthetsfunksjon som i en eller annen forstand er ”korrekt” for en gitt stokastisk variabel (f.eks. den i [3] ovenfor). Litt mer om dette kommer i neste avsnitt. For å sikre at formelen $P(a \leq X \leq b) = \int_a^b f(x)dx$ for $a \leq b$ gir sannsynlighetsverdier som oppfyller kravene i Definisjon 3.2.1 på side 35, er det nødvendig og tilstrekkelig at $f(x)$ er overalt ≥ 0 og at $\int_{-\infty}^{\infty} f(x)dx = 1$.

Definisjon 3.4.2 Hvis verdimengden til en stokastisk variabel X kan settes opp som en endelig eller uendelig liste av tall: $\{x_1, x_2, x_3, \dots\}$, da sier vi at X er *diskret fordelt*, og serien av *punktsannsynligheter* $P(X = x_1), P(X = x_2), P(X = x_3), \dots$ kalles *sannsynlighetsfordelingen* til X .

Hvis verdimengden til en stokastisk variabel X er et begrenset eller ubegrenset intervall¹⁴ på den reelle tallinja, da sier vi at X er *kontinuerlig fordelt*. Den funksjonen $f_X(x)$ (ofte skriver vi bare $f(x)$) som er slik at $P(a \leq X \leq b) = \int_a^b f_X(x) dx$ for alle a og b med $a \leq b$, kalles *sannsynlighetstettheten* til X .

Hvis vi kjenner sannsynlighetsfordelingen til en diskret fordelt stokastisk variabel — eller evt. sannsynlighetstettheten til en kontinuerlig fordelt stokastisk variabel, da kan vi i prinsippet¹⁵ besvare ethvert spørsmål om variabelen. Derfor konsentrerer man seg gjerne om s.s.-fordelingen, evt. s.s.-tettheten. Noen s.s.-fordelinger og s.s.-tettheter er spesielt viktige, på den måten at de kan uttrykkes ved matematiske formler, samtidig som de er relevante for mange praktiske anvendelser. Noen slike er behandlet i neste avsnitt.

For alle (vel, nesten alle) stokastiske variable X kan en definere (og ofte beregne) et såkalt sentralmål, eller tyngdepunkt, kalt *forventningsverdien* til X (denne er gjennomsnittlig X -verdi i det lange løp), og to spredningsmål, kalt *variansen* og *standardavviket* til X , som begge — på ulik, men nært beslektet måte — angir hvor stor spredning s.s.-fordelingen/s.s.-tettheten til X har bort fra forventningsverdien. Vi setter opp disse i to definisjoner:

Definisjon 3.4.3 For en diskret fordelt stokastisk variabel, X , er *forventningsverdien* til X (betegnet med $E(X)$ — ” E ” for ”expectation” — eller med μ_X , eller bare μ) lik summen av alle produkter av typen $x \cdot P(X = x)$, der x gjennomløper hele verdimengden V_X til X :

$$E(X) = \sum_{x \in V_X} x P(X = x) .$$

Hvis X er en kontinuerlig fordelt stokastisk variabel, med sannsynlighetstetthet $f_X(x)$, da er *forventningsverdien* til X gitt ved:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx .$$

Definisjon 3.4.4 For en diskret fordelt stokastisk variabel, X , er *variansen* til X (betegnet med $\text{Var}(X)$) lik summen av alle produkter av typen $(x - E(X))^2 \cdot P(X = x)$, der x gjennomløper hele verdimengden V_X til X :

$$\text{Var}(X) = \sum_{x \in V_X} (x - E(X))^2 P(X = x) .$$

¹⁴Merk at tallinja selv regnes som et ubegrenset intervall.

¹⁵Bare begrenset av matematiske vanskeligheter med å beregne summer og integraler.

Hvis X er en kontinuerlig fordelt stokastisk variabel, med sannsynlighetstetthet $f_X(x)$, da er *variansen* til X gitt ved:

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x) dx .$$

I begge tilfelle er *standardavviket* til X (betegnet med $SD(X)$ — for ”standard deviation” — eller med σ_X , eller bare σ) lik kvadratroten av variansen:

$$\sigma_X = \sqrt{\text{Var}(X)} .$$

I neste avsnitt skal vi komme nærmere inn på hvordan forventningsverdier, varianser og standardavvik kan regnes ut, ihvertfall for visse viktige, spesielle typer stokastiske variable.

3.5 Binomiske fordelinger og normalfordelinger

Definisjon 3.5.1 En *binomisk fordelt* stokastisk variabel er en variabel X som kan oppfattes som ”antall ja-svar” i et stokastisk forsøk som består av et fastsatt antall (n) ja/nei-spørsmål, der sannsynligheten for ”ja” er den samme (p) i alle de enkelte spørsmålene, og der hvert enkeltspørsmål er uavhengig av de andre. Verdimengden til X er da den endelige tallmengden $\{0, 1, 2, \dots, n\}$, så X er diskret fordelt.

Eksempel 3.5.1 Hvis spørsmålet ”er det en sekser?” besvares 10 ganger ved gjentatte kast med en vanlig terning, slik at X blir lik ”antall seksere på 10 kast”, da er X binomisk fordelt, med $n = 10$ og $p = 1/6$.

Teorem 3.5.1 Hvis X er binomisk fordelt med gitt n og p , da er s.s.-fordelingen til X gitt ved:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \text{for alle } x \in \{0, 1, 2, \dots, n\}.$$

Bevis: For hver $i \in \{1, 2, \dots, n\}$, la A_i være hendelsen ”det ble ”ja” på spørsmål nr. i ”. Vi finner først et opplegg for beregning av sannsynligheten for ”først x ja’er og deretter $n - x$ nei’er”:

Sannsynligheten for ”ja” i første spørsmål er lik $P(A_1) = p$ (det re jo forutsatt). Videre er sannsynligheten for ”ja” i første og andre spørsmål lik $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = p^2$ (iflg. Definisjon 3.3.1, for vi har jo også forutsatt uavhengighet). Videre (igjen p.g.a. uavhengighet) er sannsynligheten for ”ja” i første, andre og tredje spørsmål lik $P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2) \cdot P(A_3) = p^3$, osv. til vi er oppe i x ja’er. Sannsynligheten for disse, under ett, er altså p^x . Sannsynligheten for ”ja” i de x første spørsmålene og ”nei” i det neste er lik $P(A_1 \cap \dots \cap A_x \cap \overline{A_{x+1}}) = P(A_1 \cap \dots \cap A_x) \cdot P(\overline{A_{x+1}}) = p^x \cdot (1 - p)$, osv. til vi er oppe i x ja’er og resten nei’er. Sannsynligheten for ”ja” i de x første spørsmålene og ”nei” i de $n - x$ siste er dermed lik $p^x \cdot (1 - p)^{n-x}$. Men det er mange flere måter å få presis x ja’er og $n - x$ nei’er. Antallet måter vi kan markere de x ja-plassene i en resultatliste er $\binom{n}{x}$, og for hver slik markering vil vi få samme sannsynlighet, $p^x \cdot (1 - p)^{n-x}$, for ”ja på de markerte og nei på de umarkerte plassene”. (**Oppgave:** Er det virkelig slik??? Hvorfor??) Samlingen av alle disse måtene er hendelsen ” $X = x$ ”, og derfor er teoremets påstand korrekt. \triangle

Her er noen flere eksempler på bruk av binomiske fordelinger:

Sannsynligheten for presis to seksere i ett Yatzy-kast:

$$P(X = 2) = \binom{5}{2} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right)^{5-2} = 20 \cdot \frac{5^3}{6^5} \approx 0,3215.$$

Sannsynligheten for presis 5 ”kron” i kast med 10 mynter:

$$P(X = 5) = \binom{10}{5} \cdot \left(\frac{1}{2}\right)^5 \cdot \left(\frac{1}{2}\right)^{10-5} = 252 \cdot \frac{1}{2^{10}} \approx 0,2461.$$

Sannsynligheten for presis to billedkort (knekt, dame eller konge) ved 26 tilfeldige uttrekk fra en valig kortstokk, med tilbakelegginger og omstokking etter hvert trekk:

$$P(X = 2) = \binom{26}{2} \cdot \left(\frac{3}{13}\right)^2 \cdot \left(\frac{10}{13}\right)^{26-2} = 325 \cdot \frac{3^2 \cdot 10^{24}}{13^{26}} \approx 0,0319.$$

Men hva om ikke legger kortene tilbake i kortstokken etterhvert? Da vil ikke sannsynligheten for billedkort holde seg lik $\frac{12}{52} = \frac{3}{13}$, men vil avhenge av hvilke kort som etterhvert er trukket ut. F.eks., hvis de 12 første kortene som trekkes ut er billedkort, da er det ikke flere igjen, og sannsynligheten for billedkort i trekk nr.

13 vil bli lik 0. Vi kan likevel regne ut $P(X = 2)$, etter samme tankegang som i pkt. [2] på side 40, slik:

$$P(X = 2) = \frac{\binom{12}{2} \cdot \binom{40}{24}}{\binom{52}{26}} = \frac{66 \cdot 68852101650}{495918532948104} \approx 0,0084.$$

Vi ser at vi får andre verdier for punktsannsynlighetene her, så her er X ikke binomisk fordelt. Men siden vi har et metode, og kan sette opp en formel for punktsannsynlighetene, så har vi også her et navn på fordelingen: Vi sier at X her er *hypergeometrisk fordelt*.

Merknad 3.5.1 Vi har sett hvordan punktsannsynligheter beregnes for binomisk fordelte (og for hypergeometrisk fordelte) stokastiske variable. Ofte er det aktuelt å beregne sannsynligheter som omfatter flere — kanskje ganske mange — punktsannsynligheter, slik som $P(X \leq x_1)$, $P(X \geq x_1)$ eller $P(x_1 \leq X \leq x_2)$. Hvis X her er binomisk fordelt, da må slike sannsynligheter beregnes som summer av punktsannsynligheter, som beregnes én og én ved formelen i Teorem 3.5.1 over. Dette kan bli svært tidkrevende, som illustrert i eksempel ?? nedenfor, men under visse forutsetninger kan slike sannsynligheter beregnes tilnærmet, med god nøyaktighet, ved hjelp av en *normalfordelt* variabel som tilnærmer X . Dette kommer vi nærmere inn på nedenfor.

Eksempel 3.5.2 Sannsynligheten for at antall seksere er minst 15 og høyst 25 ved kast med 100 vanlige terninger er summen av 11 punktsannsynligheter for en binomisk fordelt X med $n = 100$ og $p = \frac{1}{6}$:

$$\begin{aligned} P(15 \leq X \leq 25) &= P(X = 15) + P(X = 16) + P(X = 17) + \cdots + P(X = 25) = \\ &= \sum_{x=15}^{25} P(X = x) = \sum_{x=15}^{25} \binom{100}{x} \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{100-x} = \\ &= \binom{100}{15} \left(\frac{1}{6}\right)^{15} \left(\frac{5}{6}\right)^{85} + \cdots + \binom{100}{25} \left(\frac{1}{6}\right)^{25} \left(\frac{5}{6}\right)^{75} \approx 0,7007. \end{aligned}$$

Forventningsverdi, varians og standardavvik kan lett beregnes hvis vi vet at en stokastisk variable er binomisk fordelt. Vi setter opp dette som et teorem, men tar ikke med noe bevis her:

Teorem 3.5.2 Hvis X er binomisk fordelt med gitt n og p , da er forventningsverdien til X gitt ved: $E(X) = np$, variansen er gitt ved: $\text{Var}(X) = np(1-p)$, og standardavviket er gitt ved: $\sigma = \sqrt{np(1-p)}$.

Eksempel 3.5.3 Hvis X er binomisk fordelt med $n = 100$ og $p = \frac{1}{6}$, da er forventningsverdien, variansen og standardavviket til X :

$$\mu_X = 100 \cdot \frac{1}{6} \approx 16,67, \quad \text{Var}(X) = 100 \cdot \frac{1}{6} \cdot \frac{5}{6} \approx 13,89, \quad \text{og} \quad \sigma_X = \sqrt{\text{Var}(X)} \approx 3,73.$$

Når det gjelder kontinuerlig fordelte stokastiske variable, skal vi bare se nærmere på en bestemt type, nemlig de som kalles *normalfordelte*. For hvert reelt tall μ og positivt reelt tall σ fins det en (og bare én) normalfordelt stokastisk variabel som har det førstnevnte som forventningsverdi og det sistnevnte som standardavvik. Den som har $\mu = 0$ og $\sigma = 1$ er av spesiell viktighet, så den har et spesielt navn — den kalles *standardnormalfordelingen*, og betegnes vanligvis med Z i stedet for X .

Vi har vært inne på at ”intervall-sannsynligheter” for kontinuerlig fordelte stokastiske variable X beregnes ved hjelp av en s.s.-tetthet $f_X(x)$ for X , slik: $P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(x) dx$. En kan sette opp matematiske uttrykk for s.s.-tetthetene til normalfordelte variable, men de er av en vanskelig sort, bl.a. på den måten at vanlige integrasjonsmetoder ikke strekker til. Dette er ikke noe stort problem i praksis, for både gode numeriske programmer, og tabeller med 4-sifret nøyaktighet, for beregning av slike sannsynligheter er lett tilgjengelige. Vi skal ikke bry oss noe mer om s.s.-tetthetene til normalfordelte variable enn at den for standardnormalfordelingen må tas med i følgende definisjon:

Definisjon 3.5.2 Den kontinuerlig fordelte stokastiske variabelen Z med s.s.-tetthet gitt ved: $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, $z \in \mathbf{R}$, kalles *standardnormalfordelt*¹⁶.

En stokastisk variabel X som fremkommer fra den standardnormalfordelte Z ved hjelp av et reelt tall μ og et positivt reelt tall σ , slik: $X = \sigma Z + \mu$, kalles *normalfordelt*¹⁷.

Bruken av symbolene μ og σ i definisjonen over er i overensstemmelse med tidligere bruk av disse. Dette blir fastslått i neste teorem, som vi tar med her uten bevis:

¹⁶Grafen til denne s.s.-tettheten kalles ofte *Gauss-kurven*. Merk at den er symmetrisk om 2.-aksen — det kan det bli bruk for i regning basert på tabellverdier.

¹⁷Denne sammenhengen mellom X og Z betyr at ”intervall-sannsynligheter” for X kan regnes om til ”intervall-sannsynligheter” for Z ved vanlig algebra, slik:

$$P(x_1 \leq X \leq x_2) = P(x_1 \leq \sigma Z + \mu \leq x_2) = P(x_1 - \mu \leq \sigma Z \leq x_2 - \mu) = P\left(\frac{x_1 - \mu}{\sigma} \leq Z \leq \frac{x_2 - \mu}{\sigma}\right).$$

Teorem 3.5.3 Hvis $X = \sigma Z + \mu$, der Z er standardnormalfordelt og σ er > 0 , da er μ lik forventningsverdien og σ er lik standardavviket til X .

Vi tar med et par eksempler på utregning av et par normalfordelingssannsynligheter, basert på tabellverdier. De fleste normalfordelingstabeller gir sannsynlighetsverdier av typen $P(Z \leq z)$, med fire desimalers nøyaktighet, for z -verdiene $\{0,00, 0,01, 0,02, \dots\}$, opp til $z = 3,99$ e.l. P.g.a. symmetri (se fotnote 16) vil det da fremgå at $P(Z \leq 0,00) = 0,5000$, og $P(Z \leq z)$ for negative z -verdier kan finnes slik: $P(Z \leq -z) = 1 - P(Z \leq z)$. Merk også at $P(x_1 \leq X \leq x_2) = P(X \leq x_2) - P(X < x_1)$ for enhver stokastisk variabel X , og at $P(X \leq x) = P(X < x)$ for enhver normalfordelt variabel. (**Oppgave:** Hvorfor er det slik???)

Det første eksempelet nedenfor dreier seg om en standardnormalfordelt variabel, mens det neste dreier seg om en annen normalfordeling:

Eksempel 3.5.4 Vi finner sannsynligheten for at en standardnormalfordelt variabel, Z , har verdi mellom $-0,5$ og $1,2$:

$$\begin{aligned} P(-0,5 \leq Z \leq 1,2) &= P(Z \leq 1,2) - P(Z \leq -0,5) = \\ &= P(Z \leq 1,2) - (1 - P(Z \leq 0,5)) = 0,8849 - (1 - 0,6915) = 0,5764, \end{aligned}$$

der vi på slutten har funnet to sannsynlighetsverdier fra tabell.

Eksempel 3.5.5 Vi finner sannsynligheten for at en normalfordelt variabel, X , med $\mu = 5,0$ og $\sigma = 3$, har verdi mellom $3,5$ og $8,6$, med metoden omtalt i fotnote 17:

$$\begin{aligned} P(3,5 \leq X \leq 8,6) &= P\left(\frac{3,5-5,0}{3} \leq Z \leq \frac{8,6-5,0}{3}\right) = \\ &= P(-0,5 \leq Z \leq 1,2) = 0,5764, \end{aligned}$$

der vi har utnyttet resultatet i det foregående eksempelet.

Til slutt tar vi med det som i denne sammenhengen er hovedgrunnen til vektleggingen på normalfordelinger:

Hvis X er binomisk fordelt, og n er "ganske stor", da er X tilnærmet normalfordelt, på følgende måte:

Hvis x_1 og x_2 er to tall i $\{0, 1, 2, \dots, n\}$ (= verdimengden til X), med $x_1 \leq x_2$, og Y er normalfordelt med $\mu = np$ og $\sigma = \sqrt{np(1-p)}$, da er

$$P(x_1 \leq X \leq x_2) \approx P(x_1 - 0,5 \leq Y \leq x_2 + 0,5).$$

Vi illustrerer dette ved å gå tilbake til Eksempel 3.5.2 på side 45. Vi vil nå beregne sannsynligheten for at antall seksere er minst 15 og høyst 25 ved kast

med 100 vanlige terninger, ved hjelp av tilnærming med en normalfordelt Y med $\mu = 100 \cdot \frac{1}{6} = 16,667$ og $\sigma = \sqrt{100 \cdot \frac{1}{6} \cdot 56} = 3,727$:

$$\begin{aligned} P(15 \leq X \leq 25) &\approx P(14,500 \leq Y \leq 25,500) = \\ &= P\left(\frac{14,500 - 16,667}{3,727} \leq Z \leq \frac{25,500 - 16,667}{3,727}\right) = \\ &= P(-0,58 \leq Z \leq 2,37) = P(Z \leq 2,37) - (1 - P(Z \leq 0,58)) = \\ &= 0,9911 - (1 - 0,7190) = 0,7101 . \end{aligned}$$

Som vi ser, blir resultatet temmelig likt det vi fikk i Eksempel 3.5.2, og her har vi unngått problemet med å beregne et stort antall punktsannsynligheter for X .

4 Kombinatorikk

Som kjent er det nyttig å kunne telle. Men telling er ikke alltid så trivielt som en ofte tenker seg. Tre krav er vesentlige for enhver vellykket opptelling: (1) Det må være fullstendig klart *hva* som skal telles (dvs. hva "individene" er), (2) en må være sikker på at alle "individer" blir med i tellingen, og (3) en må være sikker på at ingen "individer" telles mer enn en gang.

Ulike opptellingsproblemer krever ulike opptellingsmetoder, og en opptellingsmetode må kunne beskrives på en måte som etterkommer de tre kravene ovenfor. En vanlig metode er å gruppere elementene i den mengden A som skal telles, med tanke på at det kan være enklere å telle de enkelte gruppene hver for seg enn det var å telle hele mengden A direkte. Hvis vi attpå til kan ordne det slik at alle gruppene blir like store, da kan totalantallet beregnes som produktet av antall grupper og antallet innenfor en gruppe.

Eksempel 4.0.6 Vi tenker oss at m og n er to positive hele tall, og at mengden A består av alle tallpar (i, j) med $i \in \{1, 2, \dots, m\}$ og $j \in \{1, 2, \dots, n\}$, dvs.:

$$A = \{(i, j) \mid i \in \{1, 2, \dots, m\} \wedge j \in \{1, 2, \dots, n\}\}$$

Hver enkelt j -verdi kan da oppfattes som "merkelapp" for en gruppe, G_j , av A -elementer, slik:

$$G_j = \{(i, j) \mid i \in \{1, 2, \dots, m\}\}$$

Det er da opplagt (ja, det er vel det?) at A blir inndelt i n grupper, hver med m elementer, så antall elementer i A er lik mn .

Eksempel 4.0.7 Vi tenker oss at m , n og p er tre positive hele tall, og at mengden A består av alle talltripler (i, j, k) med $i \in \{1, 2, \dots, m\}$, $j \in \{1, 2, \dots, n\}$ og $k \in \{1, 2, \dots, p\}$, dvs.:

$$A = \{(i, j, k) \mid i \in \{1, 2, \dots, m\} \wedge j \in \{1, 2, \dots, n\} \wedge k \in \{1, 2, \dots, p\}\}$$

Hver enkelt k -verdi kan da oppfattes som "merkelapp" for en gruppe, G_k , av A -elementer, slik:

$$G_k = \{(i, j, k) \mid i \in \{1, 2, \dots, m\} \wedge j \in \{1, 2, \dots, n\}\}$$

Det er da igjen opplagt at A blir inndelt i p grupper, hver med mn elementer (det følger jo av forrige eksempel), så antall elementer i A er lik mnp .

Oppgave 4.0.1 Innse at tankegangen fra eksemplene over kan føres videre til følgende konklusjon: Hvis n, m_1, m_2, \dots, m_n alle er positive hele tall, da fins nøyaktig

$$m_1 \cdot m_2 \cdot \dots \cdot m_n$$

ulike serier av n tall, (i_1, i_2, \dots, i_n) , der $i_1 \in \{1, 2, \dots, m_1\}$, $i_2 \in \{1, 2, \dots, m_2\}$, \dots , $i_n \in \{1, 2, \dots, m_n\}$.

Oppgave 4.0.2 Innse at resultatet i oppgave 4.0.1 ovenfor også gjelder for antall ulike mulige serier av lengde n :

$$(a_1, a_2, \dots, a_n)$$

der a_1 er element i en eller annen mengde med m_1 elementer, a_2 er element i en eller annen mengde med m_2 elementer, \dots , a_n er element i en eller annen mengde med m_n elementer.

Ved å spesialisere oppgave 4.0.2 ovenfor til det tilfellet at $m_1 = m_2 = \dots = m_n$ (vi kan da bruke m for den felles verdien), da fremkommer følgende resultat:

Teorem 4.0.4 Hvis m og n er positive hele tall, og M er en mengde med nøyaktig m elementer, da har mengden

$$A = \{(a_1, a_2, \dots, a_n) \mid a_i \in M \text{ for alle } i \in \{1, 2, \dots, n\}\}$$

nøyaktig m^n elementer.

Mengden A i Teorem 4.0.4 ovenfor kalles ofte mengden av alle **ordnede n -utvalg fra M med tilbakelegging**, idet en tenker seg at ethvert A -element (a_1, a_2, \dots, a_n) kan oppnås ved at en først velger ut et tilfeldig M -element, a_1 , noterer dette og legger det tilbake i M før en velger et nytt tilfeldig element, a_2 , som noteres og legges tilbake før a_3 velges, osv. inntil n uttrekksresultater er notert i form av en ordnet serie (a_1, a_2, \dots, a_n) .

En annen spesialisering av oppgave 4.0.2 ovenfor oppnås ved at vi tenker oss at ordnede serier av typen (a_1, a_2, \dots, a_n) fremkommer ved **utvalg uten tilbakelegging**. Da velges første-elementet a_1 fritt fra mengden M , som altså har m elementer. Når så andre-elementet a_2 skal velges foretas trekningen i den reduserte mengden $M \setminus \{a_1\}$, som har $m - 1$ elementer. Så velges a_3 fra mengden $M \setminus \{a_1, a_2\}$, som har $m - 2$ elementer (fordi vi *vet* at $a_2 \neq a_1$), osv.

osv. til vi til slutt skal fastsette siste-elementet a_n ved fritt valg i mengden $M \setminus \{a_1, a_2, \dots, a_{n-1}\}$, som har $m - (n - 1)$ elementer (fordi vi vet at alle de allerede valgte elementene er ulike). Merk at dette forutsetter at $n \leq m$ (hvorfor det?). Iflg. oppgave 4.0.2 ovenfor gir dette følgende resultat:

Teorem 4.0.5 Hvis m og n er positive hele tall med $n \leq m$, og M er en mengde med nøyaktig m elementer, da fins nøyaktig

$$m \cdot (m - 1) \cdot (m - 2) \cdot \dots \cdot (m - (n - 1))$$

ulike ordnede n -utvalg fra M uten tilbakelegging.

I Teorem 4.0.5 ovenfor kan en, som et spesialtilfelle, ha at $n = m$. Dette er et viktig spesialtilfelle, for i en slik situasjon ”bruker en opp” hele mengden M hver gang et ordnet utvalg settes opp, slik at hvert utvalg rett og slett er en ordning av hele mengden M . Ofte brukes ordet **permutasjoner** om slike ordninger av en hel mengde. I dette spesialtilfellet blir antallet ordnede utvalg uten tilbakelegging (dvs. antall permutasjoner av M) lik produktet

$$m \cdot (m - 1) \cdot (m - 2) \cdot \dots \cdot 1$$

Dette uttrykket har man bruk for i flere andre sammenhenger, så det skrives vanligvis på forkortet form, slik: $m!$ (leses ” m -fakultet”, på engelsk ” m -factorial”). Vi vil også ha bruk for denne symbolikken med $m = 0$, og verdien av $0!$ er pr. definisjon satt lik 1. Dette kan oppfattes slik at vi regner den ”tomme sekvensen” $()$ som en permutasjon (den eneste) av den tomme mengden \emptyset , slik at ”antall permutasjoner av \emptyset ” er lik 1. Dermed har vi følgende definisjon:

$$m! \stackrel{\text{def}}{\Leftrightarrow} \begin{cases} 1 & \text{hvis } m = 0 \\ m \cdot (m - 1) \cdot (m - 2) \cdot \dots \cdot 1 & \text{hvis } m \in \{1, 2, 3, \dots\} \end{cases}$$

Som en direkte konsekvens av Teorem 4.0.5 ovenfor har vi da:

Teorem 4.0.6 Hvis $m \in \mathbf{N}$, og M er en mengde med nøyaktig m elementer, da fins nøyaktig $m!$ permutasjoner av mengden M .

Verdien av uttrykket $m!$ øker svært raskt med m . De første verdiene er enkle å regne ut manuelt, men etterhvert blir arbeidet uoverkommelig. Her er noen verdier:

$$0! = 1 \quad 1! = 1 \quad 2! = 2 \quad 3! = 6 \quad 4! = 24 \quad 5! = 120 \quad 6! = 720 \quad 7! = 5040$$

$$8! = 40\,320 \quad 9! = 362\,880 \quad 10! = 3\,628\,800 \quad 20! = 2\,432\,902\,008\,176\,640\,000$$

$$30! = 265\,252\,859\,812\,191\,058\,636\,308\,480\,000\,000$$

$$40! = 815\,915\,283\,247\,897\,734\,345\,611\,269\,596\,115\,894\,272\,000\,000\,000$$

I mange sammenhenger er det viktig å kunne telle **uordnede utvalg** av n individer fra en grunnmengde M med m individer ialt. Uordnede utvalg kan også noteres som ” n -strenger”: (a_1, a_2, \dots, a_n) , men en skal da se bort fra rekkefølgen, dvs. (for å ta et eksempel med $n = 3$) resultatene (a, b, c) , (a, c, b) , (b, a, c) , (b, c, a) , (c, a, b) og (c, b, a) skal oppfattes som identiske, og ikke telle som seks ulike muligheter. Når vi snakker om uordnede utvalg skal vi forutsette at utvalgene gjøres **uten tilbakelegging**, slik at en ikke får noen gjentakelser innen et enkelt utvalgsresultat. Dette betyr at en rett og slett velger ut en *delmengde* av M , med nøyaktig n elementer.

Hvor mange slike utvalg er mulig i M ? Svaret, foreløpig betegnet med C_n , kan finnes ved følgende resonnering:

Hvert uordnet n -utvalg fra M gir opphav til $n!$ ulike *ordnede* utvalg, iflg. Teorem 4.0.6 ovenfor. Dette betyr at alle de ordnede utvalgene blir gruppert i grupper à $n!$ stk., og antall grupper er lik C_n (= antall uordnede utvalg). Iflg. Teorem 4.0.5 ovenfor er det totale antallet *ordnede* n -utvalg fra M lik

$$m \cdot (m - 1) \cdot (m - 2) \cdot \dots \cdot (m - (n - 1))$$

som også må være lik produktet av antall grupper og gruppenes størrelse, dvs.:

$$C_n \cdot n! = m \cdot (m - 1) \cdot (m - 2) \cdot \dots \cdot (m - (n - 1))$$

$$\text{dvs.:} \quad C_n = \frac{m \cdot (m - 1) \cdot (m - 2) \cdot \dots \cdot (m - (n - 1))}{n!}$$

Dette brøkuttrykket må altså representere et helt tall. Det avhenger av både n og m , og betegnes med det spesielle symbolet $\binom{m}{n}$, som kalles en **binomialkoeffisient** (mer om grunnen til dette navnet nedenfor). Denne definisjonen forutsetter at $n \in \{1, 2, \dots, m\}$, men det er også nyttig å sette $\binom{m}{0} = 1$, ved at vi regner den tomme mengden \emptyset som et uordnet utvalg på 0 stk. fra mengden M . Dermed er den fullstendige definisjonen:

$$\binom{m}{n} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{hvis } n = 0 \\ \frac{m \cdot (m - 1) \cdot (m - 2) \cdot \dots \cdot (m - (n - 1))}{n!} & \text{hvis } n \in \{1, 2, \dots, m\} \end{cases}$$

Det som er gjort ovenfor leder til følgende

Teorem 4.0.7 Hvis mengden M har nøyaktig m elementer, og $n \in \{0, 1, 2, \dots, m\}$, da har M nøyaktig $\binom{m}{n}$ delmengder med nøyaktig n elementer.

Binomialkoeffisienter er viktig i mange sammenhenger. En viktig anvendelse, som også er opphavet til navnet ”binomialkoeffisient”, er denne:

Teorem 4.0.8 (Newtons¹⁸ binomialformel):

$$m \in \mathbf{N} \Rightarrow (a + b)^m = \sum_{n=0}^m \binom{m}{n} a^{m-n} b^n$$

Bevis: Hvis $(a + b)^m$ ganges helt ut, da får vi en sum av ledd som fremkommer ved at en for hver av de m faktorene $(a + b)$ ”velger” a eller b , slik at hvert enkelt ledd blir produktet av et visst antall (n) b 'er og resten $(m - n)$ a 'er, dvs. de enkelte ledd vil se slik ut: $a^{m-n} b^n$. Mange ledd vil bli like, så en kan gruppere og trekke sammen. Antallet ledd som er lik $a^{m-n} b^n$ er det samme som antallet ulike måter å velge ut n faktorer fra de m faktorene i produktet $(a + b)^m$ (det er de faktorene der vi bruker b — i de øvrige $m - n$ faktorene bruker vi a). Dette antallet er lik $\binom{m}{n}$, ifølge Teorem 4.0.7 ovenfor. Derfor blir summen, etter sammentrekking av like ledd, slik som angitt i Newtons binomialformel. \triangle

Mange ”lover” og regneregler gjelder for binomialkoeffisientene. En slik fremkommer enkelt ved at en setter $a = b = 1$ inn i Newtons binomialformel. Dette gir følgende:

Teorem 4.0.9
$$m \in \mathbf{N} \Rightarrow \sum_{n=0}^m \binom{m}{n} = 2^m$$

En enkel regneregler er ”symmetrisetningen”:

Teorem 4.0.10
$$\binom{m}{n} = \binom{m}{m-n} \text{ hvis } m \in \mathbf{N} \text{ og } n \in \{0, 1, 2, \dots, m\}$$

¹⁸Isaac Newton (1642–1727), engelsk matematiker, fysiker, teolog, mystiker, politiker, ... Av mange ansett som tidenes viktigste matematiker (og fysiker!).

Bevis: Om vi, i en liste over alle n -element-delmengder av en mengde M med m elementer, erstatter hver delmengde med dens komplement (jf. side 64) i M , da får vi en komplett liste over alle $m - n$ -element-delmengder av M . Antallet delmengder i de to listene blir derfor det samme. \triangle

En annen viktig regneregel for binomialkoeffisientene er følgende:

Teorem 4.0.11
$$\binom{m}{n} + \binom{m}{n+1} = \binom{m+1}{n+1}$$

hvis $m \in \mathbf{N}^+$ og $n \in \{0, 1, 2, \dots, m-1\}$.

Bevis: Binomialkoeffisienten på høyre side av likhetstegnet er lik antall delmengder med $n+1$ elementer i en mengde M' med $m+1$ elementer. Denne sistnevnte mengden kan oppfattes som $M \cup \{m^*\}$ der $M = M' \setminus \{m^*\}$ har m elementer (m^* kan være et hvilket som helst element i M').

Delmengdene i M' med $n+1$ elementer kan da inndeles i to grupper: De som inneholder m^* , og de som ikke inneholder m^* . Disse to gruppene er disjunkte (**Oppgave:** Innse dette!). De som tilhører den førstnevnte gruppen er fremkommet ved at delmengder av M med n elementer (slike fins det $\binom{m}{n}$ av) ”utstyres” med m^* som et tilleggselement. De som tilhører den sistnevnte gruppen er rett og slett delmengder av M med $n+1$ elementer, og slike fins det $\binom{m}{n+1}$ av. Derfor er setningen korrekt. \triangle

Teorem 4.0.11, sammen med identitetene $\binom{m}{0} = \binom{m}{m} = 1$, som gjelder for alle $m \in \mathbf{N}$ (selvsagt har en m -element-mengde nøyaktig én delmengde med 0 elementer, og nøyaktig én med m elementer), er grunnlaget for oppstilling av den såkalte **Pascals trekant** (etter den franske filosof og matematiker Blaise Pascal (1623–1662)), der binomialkoeffisientene i et ”trekantformet” skjema, med $\binom{0}{0} = 1$ på toppen, og med

$$\binom{m}{0} \quad \binom{m}{1} \quad \binom{m}{2} \quad \binom{m}{3} \quad \dots \quad \binom{m}{m-1} \quad \binom{m}{m}$$

som linje nr. m , slik:

Oppgave 4.0.3 Tenk deg at en type sko skal produseres i 10 størrelser og i 4 forskjellige farger. Hvor mange ulike typer av disse skoene må da produseres?

Oppgave 4.0.4 Tenk deg at en bilforhandler tilbyr en bestemt bilmodell med følgende valgmuligheter: 4 farger, 3 motorstørrelser, 5 typer setetrekk og 3 typer lydanlegg. Hvor mange forskjellige utgaver av denne bilmodellen må da produseres?

Oppgave 4.0.5 På hvor mange måter kan man fylle ut en vanlig tippekupong? (På hver av ialt 12 kamper skal en gjette enten "H" (hjemmeseier), "U" (uavgjort) eller "B" (borteseier).)

Oppgave 4.0.6 En gruppe på 20 personer har et internt lotteri med fem ulike premier, rangert fra 1. til 5. premie, og vinnerne trekkes ved trekning fra en eske med 20 navnelapper (en for hver person). Hvor mange ulike resultatlist er mulig, hvis trekningen utføres *med* tilbakelegging (dvs. hvis de tillater at en vinner kan få flere gevinster)? Hvor mange ulike resultatlist er mulig, hvis trekningen utføres *uten* tilbakelegging? (Se også oppgave 4.0.10 nedenfor.)

Oppgave 4.0.7 Skriv opp alle permutasjoner av mengden $A = \{a, b, c, d\}$ (dvs. skriv de fire elementene i alle mulige rekkefølger).

Oppgave 4.0.8 Hvor mange permutasjoner har mengden $B = \{a, b, c, d, e\}$?

Oppgave 4.0.9 På hvor mange ulike måter kan en klasse på 20 elever fordele seg i et klasserom med 20 pulter? Og hvis klasserommet hadde 25 pulter?

Oppgave 4.0.10 Som oppgave 4.0.6 ovenfor, men med den endringen at trekningen utføres *uten* tilbakelegging, og at lotteriet har fem *identiske* gevinster, slik at en regner to trekningslist som identiske hvis de inneholder de samme fem navnene.

Oppgave 4.0.11 Hvor mange delmengder av $C = \{a, b, c, d, e, f, g\}$ har nøyaktig 3 elementer? Hvor mange har høyst 3 elementer? Hvor mange har minst 3 elementer? For hvor mange delmengder av C er "antall elementer" $\in \{3, 4, 5\}$ (dvs. hvor mange delmengder har 3, 4 eller 5 elementer)?

Oppgave 4.0.12 (a) Hvor mange ulike "ord" på 3 bokstaver kan dannes av alfabetets 29 bokstaver, hvis en ikke stiller noen krav til mening eller lesbarhet?
(b) Hvor mange ulike "ord" på 3 *forskjellige* bokstaver kan dannes av alfabetets 29 bokstaver, hvis en ikke stiller noen krav til mening eller lesbarhet?
(c) Hvor mange ulike "ord" på 3 bokstaver kan dannes av alfabetets 29 bokstaver, hvis en ikke stiller noen krav til mening, men et visst krav til lesbarhet på den

måten at ingen vokal skal etterfølges av en vokal, og ingen konsonant etterfølges av en konsonant? (Alfabetet har 9 vokaler og 20 konsonanter.)

(d) Hvor mange ulike "ord" på 3 bokstaver kan dannes av alfabetets 29 bokstaver, hvis en ikke stiller noen krav til mening eller lesbarhet, men forutsetter at ingen bokstav brukes mer enn to ganger?

Oppgave 4.0.13 Som oppgave 4.0.12 over, men med "ord" på 4 bokstaver.

Oppgave 4.0.14 Som oppgave 4.0.12 over, men med "ord" på 5 bokstaver.

Oppgave 4.0.15 En vanlig LOTTO-kupong fylles ut ved at man krysser av 7 av tallene 1–28. På hvor mange ulike måter kan en slik kupong fylles ut?

Oppgave 4.0.16 En klasse med 14 jenter og 11 gutter skal velge en festkomité på seks personer. Hvor mange forskjellige komitéer kan bli resultatet av dette valget?

Oppgave 4.0.17 Som oppgave 4.0.16 over, men med krav om at komitéen skal bestå av tre gutter og tre jenter.

Oppgave 4.0.18 Som oppgave 4.0.16 over, men med krav om at komitéen skal ha minst to fra hvert kjønn.

5 Elementær mengdelære

5.1 Grunnbegrepene

Begrepet **mengde** må forstås sammen med begrepet **element**, først og fremst i forbindelse med uttrykksmåten "... er et element i ...", som f.eks. i følgende utsagn:

Tallet 223 er et element i mengden av alle naturlige tall.

Tallet $\sqrt{5}$ er et element i mengden av alle reelle tall.

Tallet $\sqrt{5}$ er ikke et element i mengden av alle rasjonale tall.

HiG er et element i mengden av alle norske høgskoler.

Julenissen er et element i mengden av alle nisser.

Tallet 3 er et element i løsningsmengden¹⁹ til likningen $x^2 = 9$.

Tallet 3 er et element i løsningsmengden til ulikheten $x^2 \leq 10$.

Tallparet (eller punktet, om en vil) $(3, -2)$ er et element i løsningsmengden til likningssystemet $x + y = 1$, $x - y = 5$.

Vi skal stort sett bruke store bokstaver (f.eks. A , B , ...) som navn på (symbol for) mengder, og vi skal bruke det spesielle symbolet \in som symbol for uttrykksmåten "er element i", slik at

$$a \in A$$

er den matematisk/symbolske utgaven av utsagnet " a er et element i mengden A ". Utsagnet " a er ikke et element i mengden A " skrives på symbolsk form slik: $a \notin A$.

Hva som helst kan opptre som elementer i mengder: Tall, punkter, figurer, nisser, funksjoner, statsministre, blyanter, galakser, mengder(!), Noe av det viktigste i begrepet "mengde" er at *enhver mengde skal være fullstendig identifisert ved sine elementer*, dvs.:

$$(1) \quad \boxed{A = B \text{ hvis og bare hvis } A \text{ og } B \text{ har nøyaktig de samme elementer.}}$$

Dette betyr at tilsynelatende ulike mengder kan vise seg å være identiske, som f.eks. følgende:

Eksempel 5.1.1 : $A =$ løsningsmengden til likningen $x^2 - 4 = 0$ $B =$ løsningsmengden til likningen $y^4 - y^2 - 12 = 0$ Både A og B består av tallene 2 og -2 , så her har vi ikke to mengder, men *en*, beskrevet på to forskjellige måter.

Oppgave 5.1.1 Innse at mengdene i eksempel 2.1.1 ovenfor er som beskrevet.

¹⁹Mer om "løsningsmengder" i avsnitt 2.3

5.2 Den tomme mengden og andre spesielle mengder

Liksom 0 er et viktig tallsymbol, selv om en bare skal drive med positive naturlige tall ("antall"), trenger vi å kunne snakke om "mengden uten elementer". Den kalles **den tomme mengden**, og betegnes med symbolet \emptyset , som dermed er definert slik:

$$A = \emptyset \stackrel{\text{def}}{\iff} A \text{ har ingen elementer.}$$

Merk at dette definerer en bestemt mengde (dvs. at vi ikke kan ha flere ulike tomme mengder), p.g.a. prinsippet (1) ovenfor, slik at vi har "hjemmel" for å bruke bestemt form ("den tomme mengden" i stedet for "en tom mengde"), og for å innføre et særskilt symbol for denne, slik vi har gjort.

Symbolene **N**, **Z**, **Q**, **R** og **C** skal vi knytte til følgende spesielle mengder:

N er mengden av alle naturlige tall, dvs. telle-tallene $0, 1, 2, 3, 4, \dots$.
(I en del lærebøker regnes 0 ikke med til de naturlige tallene.)

Z er mengden av alle hele tall, dvs. $\dots - 3, -2, -1, 0, 1, 2, 3, \dots$.

Q er mengden av alle rasjonale tall, dvs. de positive og negative brøktallene, inklusive de hele tallene.

R er mengden av alle reelle tall, dvs. alle tallene på tallinja.

C er mengden av alle komplekse tall²⁰.

Forøvrig skriver vi \mathbf{R}^2 for mengden av alle tallpar (x, y) , der x og y begge er reelle tall (ofte omtalt som "mengden av alle punkter i xy -planet", eller kort og godt " xy -planet"), og \mathbf{R}^3 for mengden av alle talltripler (x, y, z) , der x , y og z alle er reelle tall (ofte omtalt som "mengden av alle punkter i rommet", eller kort og godt "rommet").

5.3 Listeform, løsningsmengder og symbolikken $\{x \in A \mid \dots\}$

Klammeparenteser brukes når en skal angi en mengdes elementer direkte, f.eks. når mengden er så liten at den kan listes opp i sin helhet. Vi setter da komma mellom (symbolene for) de enkelte elementene, hvis det er mer enn ett av dem, som i eksemplene:

$\{-2, -1, 0, 1, 2\}$ er mengden av alle hele tall med absoluttverdi ≤ 2 .

$\{2, 3, 5, 7, 11, 13, 17, 19\}$ er mengden av alle primtall < 20 .

$\{a, e, i, o, u, y, \text{æ}, \text{ø}, \text{å}\}$ er mengden av alle vokaler i det norske alfabetet.

²⁰Vi vil ikke få bruk for kjennskap til komplekse tall.

Ofte blir slik listeform brukt uten at alle elementene er satt opp direkte, på den måten at noen elementer er erstattet med prikker eller liknende. Slik skrivemåte bør bare brukes når det er helt klart hva prikkene står for, slik at en kan føle seg trygg på at den som leser det en skriver oppfatter prikkene slik de er ment. Noen typiske eksempler er: $\{1, 2, 3, \dots, 20\}$ (mengden av alle heltall f.o.m. 1 t.o.m. 20).

$\{1, 1/2, 1/2^2, \dots, 1/2^{10}\}$ (mengden av alle tall av typen $1/2^n$ for heltallige n -verdier f.o.m. 0 t.o.m. 10) Slik skrivemåte kan også brukes i forbindelse med en del uendelige mengder, f.eks. slik:

$$\{1, 2, 3, \dots\} \quad (\text{mengden av alle heltall } \geq 1)$$

$$\{1, 2, 2^2, 2^3, \dots\} \quad (\text{mengden av alle tall av typen } 2^n \text{ for } n \in \mathbf{N})$$

Klammeparenteser brukes også i uttrykk av typen

$$\{x \in A \mid \dots\},$$

som leses slik: "Mengden av alle elementer x i A som oppfyller kravet \dots ". Her brukes " x " som en "element-variabel", A forutsettes å være en eller annen mengde, og " \dots " skal uttrykke et eller annet krav som x skal oppfylle. I stedet for " x " og " A " kan en bruke andre bokstaver, bare de er "ledige" og ikke bundet til annen bruk. Her er noen eksempler:

$$\{x \in \mathbf{N} \mid 7 \leq x \leq 10\} \quad (= \{7, 8, 9, 10\})$$

$$\{y \in \mathbf{N} \mid y \text{ er delelig med } 5\} \quad (= \{5, 10, 15, 20, \dots\})$$

$$\{a \in \mathbf{R} \mid \sin a = 0\} \quad (= \{\dots - 3\pi, -2\pi, -\pi, 0, \pi, 2\pi, 3\pi, \dots\})$$

Ofte er "grunnmengden" underforstått, slik at vi kan skrive $\{x \mid \dots\}$ i stedet for $\{x \in A \mid \dots\}$ (evt. med en annen bokstav enn " x ").

Hvis kravet " \dots " i mengdeuttrykket $\{x \in A \mid \dots\}$ er en likning eller ulikhet eller et system av en eller flere likninger/ulikheter i x , da kaller vi $\{x \in A \mid \dots\}$ **løsningsmengden** til likningen/ulikheten/systemet i grunnmengden A .

Eksempler:

$\{x \in \mathbf{R} \mid (x^2 - 2)(x - 3) = 0\} \quad (= \{-\sqrt{2}, \sqrt{2}, 3\})$ er løsningsmengden til likningen $(x^2 - 2)(x - 3) = 0$ i grunnmengden \mathbf{R} .

$\{x \in \mathbf{Z} \mid (x^2 - 2)(x - 3) = 0\} \quad (= \{3\})$ er løsningsmengden til likningen $(x^2 - 2)(x - 3) = 0$ i grunnmengden \mathbf{Z} .

$\{x \in \mathbf{R} \mid 1 < x^2 < 4\} \quad (= \langle -2, -1 \rangle \cup \langle 1, 2 \rangle)$ ²¹ er løsningsmengden til

²¹Mer om bruken av symbolet \cup (unionssymbolet) i avsnitt 2.5 nedenfor.

ulikhetssystemet $1 < x^2 < 4$ i grunnmengden \mathbf{R} .

Også likninger/ulikheter i to eller tre ukjente har løsningsmengder, som i eksemplene: vspace2mm

$\{(x, y) \in \mathbf{R}^2 \mid y = x^2 \text{ og } x = y^2\}$ ($= \{(0, 0), (1, 1)\}$) er løsningsmengden til likningssystemet $y = x^2$ og $x = y^2$ i grunnmengden \mathbf{R}^2 .

$\{(x, y, z) \in \mathbf{R}^3 \mid x^2 + y^2 + z^2 \leq 9\}$ ($=$ mengden av alle punkter i rommet med avstand til origo ≤ 3) er løsningsmengden til ulikheten $x^2 + y^2 + z^2 \leq 9$ i grunnmengden \mathbf{R}^3 .

5.4 Delmengder og potensmengder

Utsagnet $A \subset B$ leses ” A er en delmengde av B ”, og betyr at ethvert element i A også er element i B , dvs.:

$$A \subset B \quad \stackrel{\text{def}}{\Leftrightarrow} \quad x \in A \Rightarrow x \in B$$

Utsagnet ” A er ikke en delmengde av B ” skrives på symbolsk form slik: $A \not\subset B$.

Kjeder av delmengde-utsagn skal oppfattes som ”og”-kombinasjoner, på den måten at utsagnet $A \subset B \subset C$ betyr ” $A \subset B$ og $B \subset C$ ”. F.eks. er følgende en kombinasjon av tre utsagn om delmengder (og alle tre er korrekte, eller *sanne*, utsagn):

$$\mathbf{N} \subset \mathbf{Z} \subset \mathbf{Q} \subset \mathbf{R}$$

Merk også at utsagnene

$$\emptyset \subset A \quad \text{og} \quad A \subset A$$

er sanne for *enhver* mengde A . Det første er sant fordi \emptyset inneholder ingen elementer som ikke er i A (\emptyset inneholder jo ingen elementer overhodet). At det andre er sant, er opplagt (ja, det er vel det?).

Ved å bruke delmengdene i A som *elementer* i en ny mengde, får vi dannet *mengden av alle delmengder i A* , vanligvis kalt **potensmengden** til A . Den skal vi betegne med symbolet $\mathcal{P}(A)$. Definisjonen kan uttrykkes med symboler, slik:

$$B \in \mathcal{P}(A) \quad \stackrel{\text{def}}{\Leftrightarrow} \quad B \subset A$$

Eksempel 5.4.1 Potensmengden til $\{a, b, c\}$ har åtte elementer, og de er:

$$\emptyset, \quad \{a\}, \quad \{b\}, \quad \{c\}, \quad \{a, b\}, \quad \{a, c\}, \quad \{b, c\}, \quad \{a, b, c\}.$$

Oppgave 5.4.1 Skriv opp alle delmengdene (og beskriv dermed potensmengdene) til mengdene

$$M_1 = \{u\}, \quad M_2 = \{u, v\}, \quad M_3 = \{u, v, w\},$$

$$M_4 = \{u, v, w, x\} \quad \text{og} \quad M_5 = \{u, v, w, x, y\}.$$

Hva kan vi si om antall elementer i $\mathcal{P}(A)$ (dvs. om antall delmengder i A), hvis vi kjenner antall elementer i A ? For det første, hvis A er en uendelig mengde, da er også $\mathcal{P}(A)$ uendelig (det er vel opplagt?), og hvis A er en endelig mengde, da er også $\mathcal{P}(A)$ endelig (det er vel også opplagt?). For endelige mengder har dette spørsmålet et langt mer presist svar. Det kommer her:

Teorem 5.4.1 Hvis A er en endelig mengde, med n elementer, da har A nøyaktig 2^n delmengder, dvs. at potensmengden $\mathcal{P}(A)$ har 2^n elementer.

Bevis: Vi bruker induksjon. (Det som følger videre forutsetter ikke at dette beviset er forstått, så det kan om nødvendig utsettes til senere. Kapittel 8 gir en grundigere fremstilling av denne viktige bevisteknikken. Hovedpoenget i dette tilfellet ligger i spørsmålet *Hva skjer med antall delmengder når grunnmengden får ett element mer?* At dette spørsmålet er naturlig her fremgikk forhåpentlig av arbeidet med oppgave 5.4.1 ovenfor.)

Hvis $n = 0$, da er $A = \emptyset$, som har en eneste delmengde, nemlig \emptyset selv, så antall elementer i potensmengden $\mathcal{P}(A) = \mathcal{P}(\emptyset)$ er lik $1 = 2^0$.

Hvis vi forutsetter at n er et slikt tall at alle mengder med n elementer har precis 2^n delmengder (vi har alt sett at det fins minst ett slikt tall, nemlig 0), og vi har en mengde A med $n+1$ elementer, da har A minst ett element, og vi kan velge ut ett slikt og kalle det a . Hvis vi fjerner a fra A da gjenstår en delmengde A' av A med n elementer. Den har, ifølge vår forutsetning om n , precis 2^n delmengder. Enhver delmengde av A vil enten inneholde a eller ikke. De som *ikke* inneholder a er delmengder av A' , og slike er det altså 2^n av. De som inneholder a er det også 2^n av, fordi enhver delmengde av A som inneholder a fremkommer ved at en delmengde som *ikke* inneholder a "utvides" med tilleggsmengden $\{a\}$. Det totale antall delmengder av A er dermed lik $2^n + 2^n = 2 \cdot 2^n = 2^{n+1}$. (Hva synes du om dette opptellingsresonnementet? Tenk nøye over om du vil akseptere det!) Dette betyr at $n+1$ også har den egenskapen vi forutsatte at n hadde, dvs. enhver mengde med $n+1$ elementer har precis 2^{n+1} delmengder. Dette fullfører induksjonsbeviset, så konklusjonen er at påstanden gjelder for *alle* naturlige tall n . △

5.5 Union, snitt, differens og komplement

Unionen av to eller flere mengder (symbol: \cup) er ”sammenslåingen” av disse mengdene til *en* mengde. Det spiller ingen rolle om mengdene overlapper hverandre — den nye mengdens elementer er precis alle elementene i de mengdene en startet med:

$$x \in A \cup B \stackrel{\text{def}}{\Leftrightarrow} x \in A \text{ eller } x \in B$$

der ordet ”eller” som vanlig skal oppfattes på den ”inklusive” måten, dvs. i betydningen ”det ene eller det andre eller begge deler”.

Union-tegnet kan også brukes slik:

$$x \in A_1 \cup A_2 \cup \dots \cup A_n = \bigcup_{i=1}^n A_i \stackrel{\text{def}}{\Leftrightarrow} x \in A_1 \text{ eller } x \in A_2 \text{ eller } \dots \text{ eller } x \in A_n$$

eller, enda mer generelt:

$$x \in \bigcup_{A \in \mathcal{A}} A = \bigcup \mathcal{A} \stackrel{\text{def}}{\Leftrightarrow} x \in A \text{ for minst en } A \in \mathcal{A}$$

der vi har forutsatt at \mathcal{A} er en (endelig eller uendelig) mengde av mengder.

Noen eksempler:

$$A = \{\text{Per, Pål}\}, B = \{\text{Pål, Espen}\} \Rightarrow A \cup B = \{\text{Per, Pål, Espen}\}$$

$$A_i = \{0, 1, 2, \dots, i\} \text{ for } i \in \mathbf{N} \Rightarrow \bigcup_{i=0}^{23} A_i = A_{23} \text{ og } \bigcup_{i=0}^{\infty} A_i = \mathbf{N}$$

$$A = \bigcup \{\{a\} \mid a \in A\} \quad \text{for enhver mengde } A$$

Oppgave 5.5.1 Innse at påstandene i disse tre eksemplene er korrekte.

Snittet (eller **snittmengden**) av to eller flere mengder (symbol: \cap) er ”overlappingen” eller ”felleselementene” til disse mengdene, dvs.:

$$x \in A \cap B \stackrel{\text{def}}{\Leftrightarrow} x \in A \text{ og } x \in B$$

Mer generelt:

$$x \in A_1 \cap A_2 \cap \dots \cap A_n = \bigcap_{i=1}^n A_i \stackrel{\text{def}}{\Leftrightarrow} x \in A_1 \text{ og } x \in A_2 \text{ og } \dots \text{ og } x \in A_n$$

og enda mer generelt:

$$x \in \bigcap_{A \in \mathcal{A}} A = \bigcap \mathcal{A} \stackrel{\text{def}}{\Leftrightarrow} x \in A \text{ for alle } A \in \mathcal{A}$$

der vi har forutsatt at \mathcal{A} er en (endelig eller uendelig) mengde av mengder.

Noen eksempler:

$$A = \{\text{Per, Pål}\}, B = \{\text{Pål, Espen}\} \Rightarrow A \cap B = \{\text{Pål}\}$$

$$A_i = \{i, i+1, i+2, \dots\} \text{ for } i \in \mathbf{N} \Rightarrow \bigcap_{i=0}^{23} A_i = A_{23} \text{ og } \bigcap_{i=0}^{\infty} A_i = \emptyset$$

$$\bigcap \{A \setminus \{a\} \mid a \in A\} = \emptyset \text{ for enhver ikketom mengde } A.$$

Oppgave 5.5.2 Innse at påstandene i disse tre eksemplene er korrekte. (Vil det tredje eksempelet fortsatt være korrekt om vi stryker ordet "ikketom"?)

To mengder, A og B , sies å være **disjunkte** dersom $A \cap B = \emptyset$.

Differensen av to mengder (symbol: $A \setminus B$, leses "A minus B") er "det som gjenstår av den ene etter fjerning av de elementene som også hører til den andre mengden", dvs.:

$$x \in A \setminus B \stackrel{\text{def}}{\Leftrightarrow} x \in A \text{ og } x \notin B$$

$$\text{Et eksempel: } A = \{\text{Per, Pål}\}, B = \{\text{Pål, Espen}\} \Rightarrow A \setminus B = \{\text{Per}\}$$

Merk at union, snitt og differens er entydig bestemt av de mengdene som inngår. Som oftest er disse mengdene delmengder av en eller annen "naturlig grunnmengde", som f.eks. et tallsystem, planet \mathbf{R}^2 , rommet \mathbf{R}^3 e.l., men en trenger ikke angi grunnmengden i forbindelse med union, snitt eller differens. Annerledes er det med **komplement** (symbol: \overline{A}), som er avhengig av å bli oppfattet m.h.p. en bestemt grunnmengde:

$$\overline{A} \stackrel{\text{def}}{=} S \setminus A$$

der S er grunnmengden som A er en delmengde av.

Et eksempel med intervaller på tallinjen: $\overline{\langle -\infty, 5 \rangle} = [5, \infty)$.