För att beskriva utvecklingen för en viss variabel över tiden är det ofta mer överskådligt att använda relativa *index* än absoluta variabelvärden. Ett *enkelt index* relaterar variabelvärdet för en viss period till värdet som gällde för en viss basperiod. När index baseras på flera variabelvärden för samma period talar man om *aggregerade index*. Dessa kan baseras på enkel summering av variabelvärdena i sig eller på ihopvägning av dem. När variabelvärdena är priser kan de *vägas* ihop med hjälp av respektive produkts sålda kvantitet för att beräkna ett index. *Laspeyres index* baserar denna ihopvägning på basperiodens kvantiteter medan *Paasche index* använder aktuell periods kvantiteter.

2 Sannolikhetslära

2.1 Inledning

Ordet *sannolikhet* används av de flesta av oss i dagligt tal. Vi säger t.ex. att ”med stor sannolikhet vinner Sverige fotbollsmatchen” eller att ”det är mycket liten sannolikhet att man får sju rätt på Lotto”. Med sannolikhet avser vi då ett slags mått på hur troligt det är att en viss företeelse kommer att äga rum. Begreppet symboliserar alltså den osäkerhet kring vad som kommer att hända i en viss situation. I statistisk analys är sannolikhet ett fundamentalt begrepp. Statistiska metoder är verktyg som kan användas för att göra bedömningar av hur sannolika olika företeelser är. I detta kapitel ska vi därför lägga grunden för resten av boken genom att beskriva hur man använder sig av sannolikhetslärans grundläggande delar.

Enligt Svenska Akademiens ordbok (SAOB) har begreppet sannolikhet flera olika betydelser, där en av dem är ”numeriskt värde uttryckt i tal mellan 0 och 1 som representerar möjligheten av en viss händelses inträffande”. Ju närmare 1 sannolikheten för ett visst fenomen är, desto troligare är det att fenomenet faktiskt kommer att inträffa. Om man kastar en tärning så är det samma sannolikhet för varje sida på tärningen att komma upp. En av de sex sidorna på tärningen har fyra prickar. Sannolikheten för att få en 4:a när man kastar en tärning är därför en sjättedel. Detta brukar skrivas

P(Tärningskastet resulterar i en 4:a) = 1 / 6 = 0,1667

Det engelska ordet för sannolikhet är *probability*, vilket är anledningen till att P används för att symbolisera sannolikheter. Att ett tärningskast resulterar i att man får just en 4:a kan kallas för en *händelse*.

Man kan inte få något annat än en 1:a, en 2:a, en 3:a, en 4:a, en 5:a eller en 6:a när man kastar en tärning. Vi säger därför att *utfallsrummet* av ett tärningskast är {1, 2, 3, 4, 5, 6}. Ett utfallsrum är alltså en beskrivning av de sätt på vilket ett osäkert skeende kan utfalla. Det osäkra skeendet kallas för *experiment*.

> Ett experiment är en skeende som leder till ett av flera olika möjliga utfall.
>
> Ett utfall är resultatet av ett experiment.
>
> En sannolikhet är ett mått på hur troligt det är att ett experiment får ett visst utfall.
>
> Ett utfallsrum är de olika möjliga utfall som ett experiment kan leda till.

Exempel 2-1

Kajsa singlar slant med en enkrona. Vad är utfallsrummet och de olika utfallens sannolikheter?

Lösning

Kajsa kan få antingen krona eller klave som resultat av slantsinglingen. Experimentets utfallsrum är alltså {krona, klave}. Eftersom utfallsrummet består av två olika utfall som är exakt lika troliga så är sannolikheten för vart och ett av dem lika med 0,5, eller, om man så vill, 50 %. Formellt uttryckt har vi alltså

$$P(\text{krona}) = 0{,}5$$

och

$$P(\text{klave}) = 0{,}5$$

Anta att man kastar två tärningar. Resultatet av tärningskastningen styrs av slumpen, och var och en av tärningarna kan visa en 1:a, en 2:a, en 3:a, en 4:a, en 5:a eller en 6:a. Det finns således $6 \cdot 6 = 36$ olika möjliga utfall av denna tärningskastning, vart och ett med samma sannolikhet. Om man räknar det sammanlagda antalet prickar på de båda tärningarna så är det tydligt att många av utfallen leder till samma händelse,

d.v.s. samma sammanlagda antal prickar. Om vi slår en 3:a med tärning 1 och en 4:a med tärning två så ger det samma antal prickar som om vi skulle ha slagit en 5:a med tärning 1 och en 2:a med tärning 2. De 36 olika möjliga utfallen är försökets utfallsrum. Utfallsrummet framgår av figur 2.1.

	Tärning 2					
Tärning 1	**1:a**	**2:a**	**3:a**	**4:a**	**5:a**	**6:a**
1:a	2	3	4	5	6	7
2:a	3	4	5	6	7	8
3:a	4	5	6	7	8	9
4:a	5	6	7	8	9	10
5:a	6	7	8	9	10	11
6:a	7	8	9	10	11	12

Figur 2.1: Utfallsrummet vid kast med två tärningar

En händelse är formellt en *delmängd* av ett utfallsrum. Anta att vi vid tärningskastningen definierar händelsen A som ”det sammanlagda antalet prickar är 7”. Vi kan enkelt avläsa från figur 2.2 att 6 av de 36 olika utfallen motsvarar händelsen A. Sannolikheten för händelsen A kan alltså skrivas $P(\mathrm{A}) = 6/36 = 1/6 = 0{,}1667$.

	Tärning 2					
Tärning 1	**1:a**	**2:a**	**3:a**	**4:a**	**5:a**	**6:a**
1:a	2	3	4	5	6	7
2:a	3	4	5	6	7	8
3:a	4	5	6	7	8	9
4:a	5	6	7	8	9	10
5:a	6	7	8	9	10	11
6:a	7	8	9	10	11	12

Figur 2.2: Händelsen A

Vi inför nu några nya begrepp i sammanhanget. *Komplementet* till en händelse är samma sak som att händelsen inte inträffar. Komplementet till händelsen A är alltså att ”det sammanlagda antalet prickar inte är

7”. Vi betecknar komplementhändelsen till A med $\overline{\mathrm{A}}$. Vi inser att sannolikheten alltid är 1 att en händelse antingen inträffar eller inte inträffar, d.v.s.

$P(\mathrm{A}) + P(\overline{\mathrm{A}}) = 1,$

och således kan vi lätt beräkna

$P(\overline{\mathrm{A}}) = 1 - P(\mathrm{A}) = 1 - 0{,}1667 = 0{,}8333.$

Vi kan också se i figur 2.2 att just 30 av de 36 utfallen motsvarar händelsen $\overline{\mathrm{A}}$, så vi kunde också ha räknat fram denna sannolikhet som

$P(\overline{\mathrm{A}}) = 30/36 = 5/6 = 0{,}8333.$

Det faktum att sannolikheten alltid är 1 att en händelse antingen inträffar eller inte inträffar ger oss den mycket användbara *subtraktionsregeln* $P(\overline{\mathrm{A}}) = P(\mathrm{A}) - 1$.

En händelse är en delmängd av ett utfallsrum.

Komplementet till en händelse är att händelsen inte inträffar.

Subtraktionsregeln: $P(\overline{\mathrm{A}}) = P(\mathrm{A}) - 1$

Anta att händelsen B innebär att tärning 2 visar minst 5 prickar. Vi inser lätt att 12 av de 36 möjliga utfallen motsvarar denna händelse, och således att

$P(\mathrm{B}) = 12/36 = 1/3 = 0{,}3333.$

Vi kan nu definiera *unionen* av två händelser som att minst en av de båda händelserna inträffar. Unionen mellan två händelser A och B skrivs A∪B. Sannolikheten för att minst en av händelserna A och B inträffar motsvaras av 16 av de 36 möjliga utfallen, vilket framgår av figur 2.3. Vi har alltså

$P(\mathrm{A} \cup \mathrm{B}) = 16/36 = 4/9 = 0{,}4444.$

	Tärning 2					
Tärning 1	**1:a**	**2:a**	**3:a**	**4:a**	**5:a**	**6:a**
1:a	2	3	4	5	6	7
2:a	3	4	5	6	7	8
3:a	4	5	6	7	8	9
4:a	5	6	7	8	9	10
5:a	6	7	8	9	10	11
6:a	7	8	9	10	11	12

Figur 2.3: Unionen A∪B

Vi definierar *snittet* mellan två händelser som att båda händelserna inträffar. Snittet mellan två händelser A och B skrivs A∩B, och detta innebär alltså att det sammanlagda antalet prickar är 7, samtidigt som tärning 2 visar minst 5 prickar. Sannolikheten för att händelserna A och B samtidigt inträffar motsvaras av 2 av de 36 utfallen, vilket framgår av figur 2.4. Det gäller således att

$P(\mathrm{A} \cap \mathrm{B}) = 2/36 = 1/18 = 0{,}0555.$

	Tärning 2					
Tärning 1	**1:a**	**2:a**	**3:a**	**4:a**	**5:a**	**6:a**
1:a	2	3	4	5	6	7
2:a	3	4	5	6	7	8
3:a	4	5	6	7	8	9
4:a	5	6	7	8	9	10
5:a	6	7	8	9	10	11
6:a	7	8	9	10	11	12

Figur 2.4: Snittet A∩B

Vi kan notera att ömsesidigt uteslutande händelser inte har något snitt. Om t.ex. händelsen C definieras som att ”det sammanlagda antalet prickar är 4” så kan A och C aldrig inträffa samtidigt, vilket innebär att de är ömsesidigt uteslutande, eller *disjunkta*. För två disjunkta händelser A och C gäller alltså att $P(\mathrm{A} \cap \mathrm{B}) = 0$.

Unionen mellan två händelser innebär att minst en av de båda händelserna inträffar.

Snittet mellan två händelser innebär att båda händelserna inträffar.

Disjunkta händelser är ömsesidigt uteslutande.

I det allmänna fallet kan man också notera att $P(\text{A}) + P(\text{B})$ inte blir lika med $P(\text{A}\cup\text{B})$. Det beror på att de utfall som är gemensamma för händelserna A och B, alltså snittet mellan händelserna, i så fall skulle räknas med två gånger. Man måste alltså subtrahera snittet om man vill uttrycka sannolikheten för en union med hjälp av händelsernas enskilda sannolikheter. Detta är den s.k. *unionregeln*.

Unionregeln:

$$P(\text{A}\cup\text{B}) = P(\text{A}) + P(\text{B}) - P(\text{A}\cap\text{B})$$

Exempel 2-2

Verifiera att unionregeln stämmer för händelserna A och B i det tidigare exemplet med de två tärningarna.

Lösning

Vi har sannolikheterna för händelserna A och B:

$$P(\text{A}) = 0{,}1667$$

och

$$P(\text{B}) = 0{,}3333.$$

Vi känner också till att

$$P(\text{A}\cap\text{B}) = 0{,}0555.$$

Med unionregeln kan vi då beräkna

$$P(\text{A}\cup\text{B}) = 0{,}1667 + 0{,}3333 - 0{,}0555 = 0{,}4444,$$

vilket stämmer med den tidigare observationen i figur 2.3 att 16 av de 36 utfallen motsvarar minst en av händelserna A och B.

2.2 Betingade sannolikheter

När man analyserar sannolikheten för en viss händelse är man ofta intresserad av att variera förutsättningarna för händelsen. Anta t.ex. att man i en viss undersökning har frågat 100 personer, 60 män och 40 kvinnor, om deras inställning i en viss samhällelig fråga. Resultatet av undersökningen framgår av korstabellen nedan.

	Män	**Kvinnor**	**Totalt**
Positiv	42	26	68
Negativ	18	14	32
Totalt	60	40	100

Vi låter händelsen A motsvara att en slumpmässigt vald person från dessa 100 personer är positivt inställd och händelsen B att den slumpmässigt valda personen är en man. Vi inser då lätt att

$$P(\text{A}) = 68/100 = 0{,}68$$

och att

$$P(\text{B}) = 60/100 = 0{,}6.$$

Men anta nu att vi redan känner till att den valda personen är en man, hur sannolikt är det då att han är positivt inställd? Uppenbarligen är det andelen positivt inställda bland männen som efterfrågas. Andelen män i undersökningen är

$P(\mathrm{B}) = 0{,}6$

och andelen som samtidigt är män och positiva är

$P(\mathrm{A} \cap \mathrm{B}) = 42/100 = 0{,}42$.

Den efterfrågade sannolikheten blir då

$P(\mathrm{A} \cap \mathrm{B}) \,/\, P(\mathrm{B}) = 0{,}42/0{,}60 = 0{,}7$.

Alternativt kan vi säga att av 60 män är det 42 som är positiva, vilket även det ger den sökta sannolikheten $42/60 = 0{,}7$. Sannolikheten för att en viss händelse inträffar givet att en viss annan händelse inträffar kallas för en *betingad sannolikhet*, vilket betecknas med $P(\mathrm{A}|\mathrm{B})$.

> Räkneregel för betingad sannolikhet:
>
> $P(\mathrm{A}|\mathrm{B}) = P(\mathrm{A} \cap \mathrm{B}) \,/\, P(\mathrm{B})$

Exempel 2-3

Sannolikheten att en viss aktie går upp på börsen är 0,8 om företagets VD uttalar sig i media samma dag. Sannolikheten att aktien går upp samma dag som VD:n uttalar sig i media är 0,24. Hur sannolikt är det då att VD:n inte uttalar sig i media en viss dag?

Lösning

Vi låter A symbolisera händelsen ”aktien går upp” och B ”VD:n uttalar sig i media”. Det faktum att aktien går upp givet att VD:n uttalar sig i media betecknar vi då med A|B. Vi får veta att

$P(\mathrm{A}|\mathrm{B}) = 0{,}8$.

Att det en viss dag inträffar både att aktien går upp och att VD:n uttalar sig uttrycker vi med $\mathrm{A} \cap \mathrm{B}$, och vi vet att

$P(\mathrm{A} \cap \mathrm{B}) = 0{,}24$.

Räkneregeln för betingad sannolikhet säger att

$P(\mathrm{A}|\mathrm{B}) = P(\mathrm{A} \cap \mathrm{B}) \,/\, P(\mathrm{B})$,

vilket lätt kan arrangeras om till

$P(\mathrm{B}) = P(\mathrm{A} \cap \mathrm{B}) \,/\, P(\mathrm{A}|\mathrm{B})$,

vilket här ger oss att

$P(\mathrm{B}) = 0{,}24/0{,}8 = 0{,}3$.

Att VD:n inte uttalar sig i media är komplementhändelsen till att han uttalar sig i media, och betecknas alltså med $\overline{\mathrm{B}}$. Med hjälp av subtraktionsregeln kan vi beräkna

$P(\overline{\mathrm{B}}) = 1 - \mathrm{P}(\mathrm{B}) = 1 - 0{,}3 = 0{,}7$.

Sannolikheten att VD:n inte uttalar sig i media en viss dag är alltså 0,7.

2.3 Oberoende händelser

I fallet med betingade sannolikheter gällde att sannolikheten för den ena händelsen var *beroende* av huruvida den andra händelsen har inträffat. Det motsatta fallet är när händelser är *oberoende* av varandra. Om man t.ex. singlar slant två gånger så är resultatet av den andra singlingen oberoende av resultatet av den första singlingen. I båda fallen har händelsen ”klave kommer upp” sannolikheten 0,5. Det finns uppenbarligen fyra tänkbara utfall av att singla en slant två gånger: krona-krona, krona-klave, klave-krona och klave-klave. Alla utfallen har samma sannolikhet, vilken alltså måste vara $1/4 = 0{,}25$. Skulle man singla slant tre gånger finns 8 möjliga utfall, och varje sådant utfall måste då ha sannolikheten $1/8 = 0{,}125$, och så vidare.

Den allmänna princip med vilken man beräknar sannolikheter för utfall för oberoende händelser kallas *multiplikationsprincipen* eller *produktregeln*, vilken säger att om två händelser är oberoende så får vi sannolikheten för att båda händelserna inträffar samtidigt genom att multiplicera sannolikheterna för de enskilda händelserna.

Multiplikationsprincipen för oberoende händelser:

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

Multiplikationsprincipen gäller för vilket antal händelser som helst. Exempelvis är 5 händelser A, B, C, D och E oberoende om $P(A \cap B \cap C \cap D \cap E) = P(A) \cdot P(B) \cdot P(C) \cdot P(D) \cdot P(E)$. Principen gäller även för komplement, vilket innebär att om A och B är oberoende händelser så kommer även $\overline{A}$ och $\overline{B}$ att vara oberoende händelser.

Exempel 2-4

Produktionen av en viss elektronisk komponent kännetecknas av att i genomsnitt var tredje enhet är defekt och att varje komponents skick är oberoende av huruvida någon annan komponent fungerar. Vad är sannolikheten att man producerar fyra enheter i rad som inte är defekta?

Lösning

Låt A beteckna händelsen ”en producerad enhet är defekt”. Då har vi

$P(A) = 1/3 = 0{,}3333$.

Att en producerad enhet inte är defekt är komplementet till A, vilket innebär att

$P(\overline{A}) = 1 - 0{,}3333 = 0{,}6667$.

Att vi producerar fyra icke-defekta enheter i rad är då
$0{,}6667^4 = 0{,}1975$.

Exempel 2-5

Sannolikheten att ett företag i en viss bransch har kvinnlig VD är 0,4. 75 % av företagen i branschen är lönsamma. 35 % av företagen är både lönsamma och har kvinnlig VD. Kan det finnas ett beroende mellan lönsamhet och VD:s kön för de aktuella företagen?

Lösning

Om lönsamheten och VD:n kön var fullständigt oberoende faktorer skulle vi enligt multiplikationsprincipen ha sannolikheten $0{,}4 \cdot 0{,}75 = 0{,}3$ för att ett företag både är lönsamt och har kvinnlig VD.

Nu är denna sannolikhet tydligen 0,35, vilket visar att det kan finnas ett beroende.

Exempel 2-6

Sture singlar slant med en kompis. Myntet har precis visat ”klave” tre gånger i rad, och Sture tänker inför nästa singling därför satsa på ”krona” med motiveringen att sannolikheten för att det nu ska bli ”klave” en fjärde gång kan beräknas till $0{,}5 \cdot 0{,}5 \cdot 0{,}5 \cdot 0{,}5 = 0{,}0625$ då var och en av singlingarna är oberoende händelser. Håller du med Sture om detta?

Lösning

Visst är de olika singlingarna oberoende händelser, men det är också det enda vi håller med om. De föregående tre singlingarna har dock redan inträffat, och vi vet därför med sannolikheten 1 att var och en av dessa resulterade i klave. Det enda som är osäkert är det fjärde kastet, och sannolikheten för klave är 0,5 i detta kast.

2.4 Bayes teorem

Betingade sannolikheter för två händelser A och B kan, som vi såg tidigare, beräknas som

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

Ibland är det så att man har information om sannolikheten $P(B|A)$ när man söker sannolikheten $P(A|B)$. I det här avsnittet ska vi visa hur man kan använda den informationen för att lösa problemet. Utgångspunkten är räkneregeln för betingad sannolikhet,

$P(A|B) = P(A \cap B) / P(B)$,

som vi känner igen sedan tidigare. Vi ska nu framför allt utnyttja det uppenbara faktum att

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

Vi kan då konstatera att

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

Subtraktionsregeln säger att

$P(B) = P(B \cap A) + P(B \cap \overline{A})$,

vilket ger oss

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

Eftersom vi vet att

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

och att

$P(B \cap \overline{A}) = P(B|\overline{A}) \cdot P(\overline{A})$

så kan vi skriva

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\overline{A}) \cdot P(\overline{A})}.$$

Detta uttryck visar hur $P(A|B)$ beror på $P(B|A)$ och är känt som Bayes teorem.

Bayes teorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\overline{A}) \cdot P(\overline{A})}$$

Alternativa men ekvivalenta sätt att uttrycka Bayes teorem, som kan vara användbara i vissa situationer när andra sannolikheter är tillgängliga, är $P(A|B) = P(B|A) \cdot P(A) / P(B)$ och $P(A|B) = P(B|A) \cdot P(A) / [P(B \cap A) + P(B \cap \overline{A})]$

Exempel 2-7

På en frisörsalong för herrar arbetar två frisörer, Adam och Bertil. Adam har 90 % nöjda kunder medan Bertil har 80 %. Bertil arbetar dessutom långsammare, vilket leder till att 65 % av kunderna klipps av Adam. Om en slumpmässigt vald kund från salongen visar sig vara missnöjd, hur sannolikt är det att Adam har klippt kunden?

Lösning

Vi låter A beteckna händelsen ”kunden har klippts av Adam”. Vidare låter vi B beteckna händelsen ”kunden är missnöjd”. Vi söker då sannolikheten att Adam har klippt kunden givet att kunden är missnöjd, d.v.s. $P(A|B)$. Vi vet att

$P(A) = 0{,}65$.

Således är

$P(\overline{A}) = 1 – 0{,}65 = 0{,}35$.

Vi vet att om Adam har klippt en kund så blir denne med 90 % sannolikhet nöjd, d.v.s.

$P(\overline{B}|A) = 0{,}9.$

Enligt subtraktionsregeln lämnar således 10 % av Adams kunder salongen missnöjda, d.v.s.

$P(B|A) = 0{,}1.$

Av Bertils kunder blir 80 % nöjda och således 20 % missnöjda, d.v.s.

$P(\overline{B}|\overline{A}) = 0{,}8$

och

$P(B|\overline{A}) = 0{,}2.$

Vi använder nu Bayes teorem för att beräkna $P(A|B)$.

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B|A)\cdot P(A) + P(B|\overline{A})\cdot P(\overline{A})} = \frac{0{,}1\cdot 0{,}65}{0{,}1\cdot 0{,}65 + 0{,}2\cdot 0{,}35} = 0{,}48$$

Om en slumpmässigt vald kund visar sig vara missnöjd så är sannolikheten att Adam har klippt honom alltså 0,48.

Exempel 2-8

Antag att en viss lögndetektor i reklamen påstås vara väldigt tillförlitlig, eftersom den ger utslag med sannolikheten 90 % om en person ljuger och med sannolikheten 5 % om personen inte ljuger. Antag vidare att 2 % av de personer som förhörs med hjälp av lögndetektor faktiskt ljuger. Hur sannolikt är det då att en person faktiskt ljuger om lögndetektorn ger utslag?

Domstolar i vissa länder använder lögndetektorer som stöd för juridisk prövning. Fundera på om du skulle vilja bli förhörd med hjälp av den aktuella lögndetektorn.

Lösning

Vi låter A beteckna händelsen ”personen ljuger” och B betecknar händelsen ”lögndetektorn ger utslag”. Vi söker sannolikheten att personen ljuger givet att lögndetektorn ger utslag, d.v.s. $P(A|B)$. Vi vet att

$P(A) = 0{,}02$

och således också att

$P(\overline{A}) = 1–0{,}02 = 0{,}98.$

Givet att en person ljuger ger lögndetektorn (korrekt) utslag med sannolikheten 0,9, d.v.s.

$P(B|A) = 0{,}9.$

Om en person inte ljuger ger lögndetektorn (felaktigt) utslag med sannolikheten 0,05, d.v.s.

$P(B|\overline{A}) = 0{,}05.$

Med Bayes teorem får vi då sannolikheten för $P(A|B)$.

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B|A)\cdot P(A) + P(B|\overline{A})\cdot P(\overline{A})} = \frac{0{,}9\cdot 0{,}02}{0{,}9\cdot 0{,}02 + 0{,}05\cdot 0{,}98} = 0{,}27$$

Sannolikheten att en person faktiskt ljuger när lögndetektorn påstår det är alltså bara 0,27. I de återstående 73 % av fallen är det så att den förhörda personen faktiskt talar sanning trots att lögndetektorn ger utslag för lögn.

2.5 Kombinatorik

Hur många olika ”ord” kan man konstruera med hjälp av bokstäverna A, B och C? Uppenbarligen är svaret sex stycken: ABC, ACB, BCA, BAC, CAB och CBA. Men kunde vi ha resonerat oss fram till det utan att pröva? Ja, det finns ju 3 ”positioner”, där det ska finnas exakt en bokstav på varje position. Till den första positionen kan vi välja vilken som helst av de tre bokstäverna. Till den andra positionen har vi sedan ett fritt val mellan de två resterande, och på den tredje platsen sätter vi avslutningsvis den sista bokstaven. Således måste det finnas $3\cdot 2\cdot 1 = 6$ olika ”ord” som kan bildas av bokstäverna A, B och C. Om vi hade utgått från de 7 bokstäverna A, B, C, D, E, F och G så blir det arbetsamt att pröva sig fram till hur många olika ord som kan bildas, men med samma logik som ovan kan vi lätt räkna ut att det måste bli $7{\cdot}6{\cdot}5{\cdot}4{\cdot}3{\cdot}2{\cdot}1 = 5\ 040$ ord. Generellt gäller alltså att om vi hade haft n olika bokstäver så hade vi kunnat sortera dessa på $n(n-1)(n-2)\cdots 1$ olika sätt. Denna produkt, $n(n-1)(n-2)\cdots 1$, kallas allmänt för *fakulteten* för n och betecknas $n!$. Vi har alltså till exempel $7! = 7{\cdot}6{\cdot}5{\cdot}4{\cdot}3{\cdot}2{\cdot}1 = 5\ 040$.

Fakulteten för ett tal n:

$$n! = n(n-1)(n-2)\cdots 1$$

Anta nu att vi har de 7 bokstäverna A, B, C, D, E, F och G, och att vi vill bilda ord om endast 4 av dessa 7 bokstäver. Hur många sådana ord finns det? Uppenbarligen finns det 7 sätt att välja ut den första bokstaven, 6 sätt att välja den andra, 5 sätt att välja den tredje samt 4 sätt att välja den fjärde. Tydligen finns alltså $7{\cdot}6{\cdot}5{\cdot}4 = 840$ sätt att bilda ord om 4 bokstäver från en mängd av 7 bokstäver. Med andra ord finns det 840 sätt att välja ut och rangordna 4 element från en mängd bestående av 7 element. Varje sådant sätt kallas för en *permutation*.

Generellt kan vi välja ut och rangordna r element från en mängd bestående av n element på $n!/(n-r)!$ olika sätt. I fallet med de 7 bokstäverna hade vi $n = 7$ och $r = 4$, och som vi såg tidigare blev antalet permutationer då

$$\frac{n!}{(n-r)!} = \frac{7!}{(7-4)!} = \frac{7!}{3!} = \frac{7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{3\cdot 2\cdot 1} = 7\cdot 6\cdot 5\cdot 4 = 840.$$

En permutation är ett möjligt sätt att välja ut och rangordna r element från en mängd bestående av n element.

Antalet permutationer:

$$n\mathbf{P}r = \frac{n!}{(n-r)!}$$

Om ordningen mellan de utvalda r elementen från mängden med n element inte spelar någon roll kallas ett sätt att välja ut dem i stället för en *kombination*. Kombinationer är mycket vanliga, och kombinationsprincipen är mycket viktig i den analytiska statistiken, vilket kommer att framgå längre fram i den här boken.

De flesta av oss inser att det är mycket låg sannolikhet för att man får 7 rätt på Lotto om man spelar en enkelrad. Men exakt vilken sannolikhet har egentligen denna händelse? När man spelar på Lotto handlar det om att gissa vilka 7 nummer av de totalt 35 möjliga som dras. Antalet permutationer kan då beräknas som

$$35\mathbf{P}7 = 35!/(35-7)!$$

Men är lottodragningen verkligen en permutation? Nej, det måste ju vara en kombination eftersom det inte spelar någon som helst roll i vilken ordning de 7 numren dras. Vi måste alltså justera för att ordningen mellan numren inte spelar någon roll. Antalet sätt att sortera 7 dragna nummer vet vi är $7! = 5\ 040$. Det måste alltså finnas 5 040 gånger fler sätt att välja ut 7 nummer av 35 möjliga om ordningen mellan de utvalda numren är relevant. Tydligen kan antalet olika sätt att välja 7 nummer av 35 möjliga om ordningen inte spelar roll beräknas som

$$\frac{35!/(35-7)!}{7!} = \frac{35!}{7!(35-7)!} = 6\ 724\ 520.$$

Man har alltså en chans på 6 724 520 att få 7 rätt på Lotto om man spelar en enkelrad. Eller mer formellt:

$$P(7 \text{ rätt på Lotto}) = 1/6724520 = 0{,}0000001487.$$

Generellt finns alltså $n!/(r!(n-r)!)$ sätt att välja ut r element från en mängd med n element, och varje sådant sätt kallas för en *kombination*.

En kombination är ett möjligt sätt att välja ut r element från en mängd bestående av n element där den inbördes ordningen mellan de r elementen inte är relevant.

Antalet kombinationer:

$$n\mathbf{C}r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Antalet kombinationer betecknas alltså med $n\mathbf{C}r$ (på de flesta miniräknare) eller med $\binom{n}{r}$ som ska uttalas ”n över r”.

Exempel 2-9

Sture ska baka olika 12 kakor, men han har inte bestämt sig i vilken ordning de ska bakas. Hur många olika alternativ finns det?

När kakorna är bakade ska Sture slumpmässigt välja ut och ge bort 3 av kakorna till sin mamma i födelsedagspresent. Hur sannolikt är det att de tre sist bakade kakorna blir de som hamnar i presentpaketet?

Av de återstående 9 kakorna ska Sture äta upp 4. På hur många olika sätt kan han konsumera dessa 4 kakor om han äter dem en och en?

Lösning

Det finns 12! = 479 001 600 olika ”ordningar” att baka de 12 kakorna i.

Antalet olika presentpaket är en fråga om kombinationer, eftersom ett paket har samma innehåll oavsett i vilken ordning de tre kakorna läggs i. Vi får då

$$\binom{12}{3} = \frac{12!}{3!(12-3)!} = 220 \text{ olika paket}$$

Sannolikheten är alltså $1/220 = 0{,}0045$ att de tre sista kakorna blir de som hamnar i paketet.

När det gäller ätandet av fyra av de återstående 9 kakorna spelar ordningen roll, eftersom Sture äter kakorna en och en. Det är då en fråga om permutationer. Vi får då

$$\frac{9!}{(9-4)!} = 3\ 024 \text{ olika sätt.}$$

2.6 Syntes

Det här kapitlet har behandlat begreppet *sannolikhet*. En sannolikhet är ett mått på hur troligt det är att ett skeende som kan resultera i flera olika möjliga *utfall* leder fram till ett visst utfall. En sannolikhet konkretiseras som ett numeriskt värde mellan 0 och 1, eller mellan 0 % och 100 %. Om sannolikheten för ett visst utfall är 75 % så betyder det att detta utfall inträffar 3 gånger av 4 i det långa loppet.

En beskrivning av de olika utfall som är möjliga kallas *utfallsrum*. När man till exempel kastar två tärningar består utfallsrummet av 36 olika utfall, eftersom varje tärning har 6 olika sidor. En *händelse* utgörs av en *delmängd* av ett utfallsrum. Om man kastar två tärningar och bara är intresserad av det totala antalet prickar så motsvarar de 36 utfallen endast 11 olika händelser eftersom många av utfallen innebär samma sammanlagda antal prickar. Flera olika utfall i utfallsrummet kan alltså motsvara samma händelse.

Komplementet till en händelse innebär att händelsen inte inträffar. Komplementet till händelsen ”tärningarna visar sammanlagt 7 prickar” är då att ”tärningarna visar inte sammanlagt 7 prickar”. Enligt *subtraktionsregeln* är sannolikheten alltid 1 att en händelse eller dess komplementhändelse inträffar.

Unionen mellan ett antal händelser innebär att minst en av händelserna inträffar, medan *snittet* mellan ett antal händelser innebär att samtliga händelser inträffar. När händelser inte har något snitt säger man att de är ömsesidigt uteslutande, eller *disjunkta*.

En *betingad sannolikhet* är sannolikheten för att en viss händelse inträffar, under förutsättning att en viss annan händelse inträffar. Om två sannolikheter är *oberoende* så har den ena händelsens eventuella inträffande ingen påverkan på sannolikheten för huruvida den andra händelsen faktiskt inträffar. För oberoende händelser kan *multiplikationsprincipen* användas för att bestämma sannolikheten för händelsernas snitt. Med hjälp av *Bayes teorem* kan man beräkna den betingade sannolikheten för två händelser utifrån information om den omvända betingade sannolikheten för de båda händelserna.

Ett sätt på vilket ett visst antal element kan väljas ut från en viss mängd element kallas för en *permutation* om det är så att ordningen mellan elementen är av betydelse. Om ordningen elementen emellan saknar betydelse så kallas sättet att välja ut dem för en *kombination*.