Betinget sannsynlighet, total sannsynlighet og Bayes' setning:

En introduksjon med forslag til klasseaktivitet og prosjektoppgave

Ørnulf Borgan
Matematisk institutt
Universitetet i Oslo

Formål

I grunnkurset i matematikk i videregående skole skal elvenene "ha en intuitiv forståelse av uavhengighet og betinget sannsynlighet" (mål 5d), og de skal "kunne bruke addisjonssetningen og produktsetningen" (mål 5e). I videregående kurs I skal elevene "kjenne begrepene uavhengighet og betinget sannsynlighet, og kunne bruke Bayes' setning på to hendelser" (mål 6a i 2MX og mål 5a i 2MZ). Formålet med dette notatet er å gi lærerne en bedre bakgrunn for å oppfylle disse målene i læreplanene (avsnittene 1-6) og å komme med en idé til klasseaktivitet (avsnitt 7) og mulige prosjektoppgaver (avsnitt 8).

1. En intuitiv forståelse av betinget sannsynlighet og uavhengighet

For å få en *intuitiv forståelse* av hva vi mener med betinget sannsynlighet og avhengige og uavhengige hendelser, tar vi utgangspunkt i et enkelt eksempel. Eksemplet er motivert av spillet "redblack" som omtales i avsnitt 7.

Eksempel 1

Fra en kortstokk velger vi fire røde kort (ruter, hjerter) og to svarte kort (kløver, spar) og legger i en bunke. Vi trekker tilfeldig ett kort fra bunken og ser hvilken farge det har. Uten å legge kortet tilbake trekker vi ett kort til. Vi betrakter hendelsene A = "første kort er rødt" og B = "andre kort er svart". Det er klart at $P(A) = \frac{4}{6} = \frac{2}{3}$.

Hvis A har inntruffet, er det tre røde og to svarte kort igjen i bunken når det andre kortet trekkes. Sannsynligheten for B er da $\frac{2}{5}$. Siden denne sannsynligheten forutsetter at A har inntruffet, sier vi at $\frac{2}{5}$ er den *betingede sannsynligheten* for B gitt A. Vi skriver $P(B \mid A) = \frac{2}{5}$.

Den komplementære hendelsen til A er $\overline{A}$ = "første kort er svart". Vi har $P(B \mid \overline{A}) = \frac{1}{5}$.

I eksempel 1 er det *intuitivt* klart hva vi mener med den betingede sannsynligheten for B gitt A og den betingede sannsynligheten for B gitt $\overline{A}$. Det kommer av at vi ser på en hendelse knyttet til andre trekning når vi har gitt en hendelse knyttet til første trekning. I eksempel 1 er det de betingede sannsynlighetene som er "intuitivt opplagte". Det er ikke like klart hva den *ubetingede* sannsynligheten for B er. Det vender vi tilbake til i eksempel 5.

Eksempel 1 gir oss også en intuitiv forståelse av hva det betyr at to hendelser er avhengige.

Vi har i eksemplet at $P(B \mid A) = \frac{2}{5}$ og $P(B \mid \overline{A}) = \frac{1}{5}$. Hvis hendelsen *A* inntreffer, er sannsynligheten for *B* lik $\frac{2}{5}$. Hvis *A* ikke inntreffer, er sannsynligheten for *B* lik $\frac{1}{5}$. Det om hendelsen *A* inntreffer eller ikke gir oss informasjon som endrer sannsynligheten for hendelsen *B*. Derfor er *A* og *B* er *avhengige* hendelser.

For å gi en intuitiv forståelse av hva det betyr at to hendelser er uavhengige, endrer vi litt på eksemplet.

> **Eksempel 2**
> Som i eksempel 1 har vi en bunke med fire røde og to svarte kort. Vi trekker tilfeldig ett kort fra bunken og ser hvilken farge det har. Vi legger kortet tilbake i bunken og trekker tilfeldig ett kort til. Vi betrakter igjen hendelsene *A* = "første kort er rødt" og *B* = "andre kort er svart". Vi har fortsatt at $P(A) = \frac{4}{6} = \frac{2}{3}$.
> Men siden vi legger det første kortet tilbake i bunken før vi trekker det neste, er det nå fire røde og to svarte kort i bunken når det andre kortet trekkes. Og dette gjelder uansett om *A* inntraff eller ikke. Dette betyr at $P(B \mid A) = \frac{2}{6}$ og $P(B \mid \overline{A}) = \frac{2}{6}$. Disse betingede sannsynlighetene er dermed begge lik den ubetingede sannsynligheten $P(B)$.

Om hendelsen *A* inntreffer eller ikke gir i eksempel 2 ingen informasjon som endrer sannsynligheten for hendelsen *B*. Derfor er *A* og *B* er *uavhengige* hendelser.

2. Definisjon av betinget sannsynlighet

I eksempel 1 er det intuitivt klart hva vi skal forstå med betinget sannsynlighet. Det skyldes at vi ser på hva som skjer i andre trekning når vi vet resultatet av den første. Men det er ikke alltid like klart hva betinget sannsynlighet skal bety. For situasjonen i eksempel 1, kan vi for eksempel spørre:

- Hva er den betingede sannsynligheten for at begge kortene er røde gitt at minst ett av dem er rødt?
- Hva er den betingede sannsynligheten for at det første kortet er rødt gitt at det andre er svart?

For å svare på disse spørsmålene (og andre spørsmål som er av større praktisk interesse) trenger vi en definisjon av betinget sannsynlighet. Vi bruker et eksempel til å motivere definisjonen.

Eksempel 3

Tabellen nedenfor stammer fra Statistisk sentralbyrå. Den gir en oversikt over inngåtte ekteskap i 1997 etter brudens og brudgommens alder. Vi betrakter et tilfeldig valgt brudepar fra dette året, og ser på hendelsene A = "brudgommen er 20-24 år" og B = "bruden er 25-29 år". Vi finner:

$$P(A)=\frac{2263}{23815}=0,095 \qquad P(B)=\frac{8684}{23815}=0,365 \qquad P(A\cap B)=\frac{404}{23815}=0,017$$

Merk spesielt at sannsynligheten er 36,5% for at bruden er 25-29 år gammel.

Aktuelle befolkningstall 9/98

3 **Inngåtte ekteskap etter brudens og brudgommens alder. 1997**

Brudgommens alder	I alt	Brudens alder											
		15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-
Ekteskap i alt	**23,815**	**527**	**4,935**	**8,684**	**4,669**	**2,241**	**1,244**	**767**	**450**	**188**	**55**	**32**	**23**
15 - 19	106	55	41	4	5	-	-	-	-	1	-	-	-
20 - 24	2,263	251	1,537	404	49	16	2	2	2	-	-	-	-
25 - 29	7,728	137	2,408	4,265	745	124	41	5	1	2	-	-	-
30 - 34	6,132	61	690	2,873	1,970	435	79	13	9	2	-	-	-
35 - 39	3,078	14	156	762	1,184	710	201	40	9	2	-	-	-
40 - 44	1,756	7	56	233	464	513	353	102	25	2	1	-	-
45 - 49	1,162	2	27	96	160	264	294	234	73	10	2	-	-
50 - 54	882	-	7	38	68	135	181	248	163	35	4	3	-
55 - 59	400	-	11	4	17	33	70	85	97	66	15	1	1
60 - 64	160	-	2	3	6	9	16	29	44	35	9	5	2
65 - 69	73	-	-	2	1	1	4	7	17	21	12	6	2
70 -	75	-	-	-	-	1	3	2	10	12	12	17	18

Hvis vi vet at brudgommen er 20-24 år gammel, hva er da sannsynligheten for at bruden er 25-29 år? Med andre ord, hva er $P(B\mid A)$? Det er 2263 brudepar hvor brudgommen er 20-24 år, og av disse er det 404 hvor bruden er 25-29 år gammel. Derfor har vi at

$$P(B\mid A)=\frac{404}{2263}=0,179$$

Hvis vi vet at brudgommen er 20-24 år, er sannsynligheten 17,9% for at bruden er 25-29 år gammel. (Dette viser, ikke overraskende, at A og B er avhengige hendelser.)

Vi merker oss at vi kan skrive

$$P(B\mid A)=\frac{404}{2263}=\frac{\frac{404}{23815}}{\frac{2263}{23815}}=\frac{P(A\cap B)}{P(A)}$$

Den sammenhengen vi fant i eksempel 3 mellom $P(B|A)$, $P(A\cap B)$ og $P(A)$ gir en motivasjon for definisjonen:

Definisjon

Den betingede sannsynligheten for B gitt A er gitt ved

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \qquad (1)$$

så sant $P(A) > 0$.

Merk at definisjonen gir (ved å bytte om "rollene" til *A* og *B*):

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad (2)$$

så sant $P(B) > 0$.

La oss vise at den *formelle definisjonen* (1) er i overensstemmelse med den *intuitive forståelsen* fra eksempel 1.

Eksempel 4

Som i eksempel 1 har vi en bunke med fire røde og to svarte kort. Vi trekker tilfeldig ett kort fra bunken og ser hvilken farge det har. Uten å legge kortet tilbake trekker vi ett kort til.
Vi betrakter som før hendelsene *A* = "første kort er rødt" og *B* = "andre kort er svart".
Vi vil bruke definisjonen (1) til å finne den betingede sannsynligheten for *B* gitt *A*.
Vi må da først bestemme sannsynlighetene[1] for hendelsene $A \cap B$ og *A*.
Vi gjør dette ved å beregne antall mulige utfall når vi trekker to kort fra en kortstokk uten tilbakelegging og hvor mange av disse som er gunstige for de to hendelsene.
Vi kan trekke to kort på 6·5 = 30 mulige måter. Av disse er 4·2 = 8 gunstige for hendelsen $A \cap B$ mens 4·5 = 20 er gunstige for hendelsen *A*.
Dermed har vi $P(A \cap B) = \frac{8}{30}$ og $P(A) = \frac{20}{30}$. Definisjonen (1) av betinget sannsynlighet gir at

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{8}{30}}{\frac{20}{30}} = \frac{2}{5}$$

Dette stemmer med det vi fant intuitivt i eksempel 1.

I eksempel 2 la vi det første kortet tilbake i bunken før vi trakk det andre. Framgangsmåten i eksempel 4 kan brukes også i denne situasjonen til å finne formelt de betingede sannsynlighetene vi fant i eksempel 2. Gjør dette selv.

[1] Vi kan ikke bruke produktsetningen fra grunnkurset til å bestemme sannsynligheten for $A \cap B$. Produktsetningen ble i grunnkurset begrunnet ved et intuitivt resonnement. Når vi gir en formell matematisk framstilling av dette stoffet, er

Vi kan også bruke framgangsmåten i eksempel 4 til å finne de betingede sannsynlighetene vi spurte om i starten av dette avsnittet. Vi illustrerer framgangsmåten for det siste spørsmålet.

Eksempel 5

Vi har en bunke med fire røde og to svarte kort. Som i eksemplene 1 og 4 trekker vi tilfeldig ett kort fra bunken og ser hvilken farge det har. Uten å legge kortet tilbake trekker vi ett kort til. Som før ser vi på hendelsene A = "første kort er rødt" og B = "andre kort er svart".
Vi vil bestemme den betingede sannsynligheten for at det første kortet er rødt gitt at det andre er svart, dvs. $P(A \mid B)$.

Vi må først finne $P(B)$. Vi kan få et svart kort andre gang på to måter: Vi kan trekke et rødt kort første gang og et svart kort andre gang, eller vi kan trekke et svart kort både første og andre gang. Dette betyr at hendelsen B er en union av de to disjunkte hendelsene $A \cap B$ og $\overline{A} \cap B$.
Addisjonssetningen for disjunkte hendelser gir derfor at

$$P(B) = P(A \cap B) + P(\overline{A} \cap B) = \frac{4 \cdot 2}{6 \cdot 5} + \frac{2 \cdot 1}{6 \cdot 5} = \frac{10}{30} = \frac{2}{6}$$

Merk at dette er det samme som sannsynligheten for at det første kortet er svart!

Ved å bruke (2) får vi nå

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{8}{30}}{\frac{10}{30}} = \frac{4}{5} = 0,80$$

Hvis det andre kortet er svart, er sannsynligheten 80% for at det første kortet var rødt.

I eksempel 5 fant vi at hvis det andre kortet vi trekker er svart, så er sannsynligheten $\frac{4}{5}$ for at det første kortet vi trakk var rødt. Hva betyr egentlig dette?
For å forstå det, minner vi om at *sannsynlighet er relativ frekvens i det lange løp*.
I eksempel 1 fant vi $P(A) = \frac{2}{3}$. Det betyr at hvis vi om og om igjen trekker tilfeldig to kort fra en bunke med fire røde og to svarte kort, så vil vi i det lange løp få et rødt kort første gang i omtrent $\frac{2}{3} = 67\%$ av trekningene.

---

produktsetningen en konsekvens av definisjonen (1) av betinget sannsynlighet (jfr. avsnitt 3). Så hvis vi bruker produktsetningen her, lager vi et resonnement som "går i ring."

I eksempel 5 fant vi $P(A \mid B) = \frac{4}{5} = 80\%$. Det betyr at hvis vi bare teller de trekningene hvor det andre kortet er svart, så vil vi i det lange løp *i disse trekningene* få et rødt kort første gang i omtrent 80% av trekningene.

Denne måten å fortolke sannsynlighet og betinget sannsynlighet gjelder generelt:

- *P(A) svarer til den relative frekvensen av A når vi gjentar et forsøk mange ganger.*
- *P(A|B) svarer til den relative frekvensen av A når vi bare teller de forsøkene hvor B inntreffer.*

3. Avhengige og uavhengige hendelser

I eksempel 1 fant vi at $P(B \mid A) = \frac{2}{5}$, mens vi i eksempel 5 fant at $P(B) = \frac{1}{3}$.

Det at hendelsen A = "første kort er rødt" inntreffer, endrer sannsynligheten for hendelsen B = "andre kort er svart". I eksempel 1 er A og B *avhengige* hendelser.

I eksempel 2 er sannsynligheten $\frac{1}{3}$ for B = "andre kort er svart" både når vi betinger med A = "første kort er rødt" og når vi ikke gjør det. Sannsynligheten for B blir ikke endret av at A inntreffer. I eksempel 2 er A og B er *uavhengige* hendelser.

Dette gir definisjonen

Definisjon

Hvis $P(B|A) = P(B)$ er A og B uavhengige hendelser.

Hvis $P(B|A) \neq P(B)$ er A og B avhengige hendelser.

Ut fra definisjonen kan vi vise:

Resultat[2]

Anta at $P(A) > 0$ og $P(B) > 0$. Da er følgende tre utsagn ekvivalente

- $P(B|A) = P(B)$ (3)
- $P(A|B) = P(A)$ (4)
- $P(A \cap B) = P(A) \cdot P(B)$ (5)

[2] Resultatet viser at det ikke spiller noen rolle hvilken at utsagnene (3)-(5) vi bruker som definisjon av uavhengighet. I skolematematikken er det nok vanligst å bruke (3) [eventuelt (4)] som definisjon. I mer "avanserte" framstillinger er det vanlig å bruke (5) som definisjon på at hendelsene A og B er uavhengige. En slik definisjon er ikke så intuitiv som den vi har gitt, men den har den fordelen at den er symmetrisk i A og B og at den gjelder selv om (minst) en av hendelsene har sannsynlighet lik null.

La oss vise resultatet. Anta at (3) holder. Da har vi at $P(B) = P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$. Ved å multiplisere med $P(A)$ på begge sider av dette uttrykket, får vi (5). Av (5) får vi videre at $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A) \cdot P(B)}{P(B)} = P(A)$, som viser at (4) holder. Omvendt kan vi fra (4) vise at (5) holder og deretter at (3) holder. Dette viser at utsagnene (3)-(5) er ekvivalente.

Eksempel 2 viser at hendelsene A = "første kort er rødt" og B = "andre kort er svart" er uavhengige når vi trekker med tilbakelegging. Dette svarer til vår intuitive oppfatning av hva uavhengighet er.

I mange situasjoner kan vi ikke kontrollere ved en beregning at to hendelser er uavhengige. Ofte må vi forutsette uavhengighet på grunnag av våre kunnskaper om den situasjonen vi studerer. Uavhengighet er da en forutsetning vi gjør om sannsynlighetsmodellen. For eksempel vil vi forutsette at barnas kjønn er uavhengige i en tobarnsfamilie som ikke har tvillinger.

4. Produktsetningen

Hvis vi multipliserer med $P(A)$ på begge sider av (1), får vi

Produktsetningen

La A og B være hendelser ved et forsøk. Da er

$$P(A \cap B) = P(A) \cdot P(B|A) \tag{6}$$

Ved å bruke produksetningen to ganger får vi:

$$\begin{aligned} &P(A \cap B \cap C) \\ &\quad = P(A \cap B) \cdot P(C|A \cap B) \\ &\quad = P(A) \cdot P(B|A) \cdot P(C \mid A \cap B) \end{aligned} \tag{7}$$

Ved et tilsvarende resonnement (formelt ved et induksjonsbevis) kan vi vise at produktsetningen gjelder for n hendelser.

Eksempel 6

Etter offentlig statistikk er sannsynligheten

- 94% for at en 65 år gammel kvinne skal bli minst 70 år
- 91% for at en 70 år gammel kvinne skal bli minst 75 år
- 84% for at en 75 år gammel kvinne skal bli minst 80 år

Hva er sannsynligheten for at en 65 år gammel kvinne skal bli minst 80 år?

Vi har en 65 år gammel kvinne og ser på hendelsene

A = "kvinnen blir minst 70 år"

B = "kvinnen blir minst 75 år"

C = "kvinnen blir minst 80 år"

Vi har $P(A) = 0{,}94$. Hvis hendelsen A inntreffer, er kvinnen blitt 70 år, så $P(B|A) = 0{,}91$. Hvis både hendelsene A og B inntreffer, har kvinnen blitt 75 år, så $P(C|A \cap B) = 0{,}84$.
Hvis kvinnen blir minst 80 år, må hun også bli 70 år og 75 år. Derfor er $C = A \cap B \cap C$.
Da gir (7):

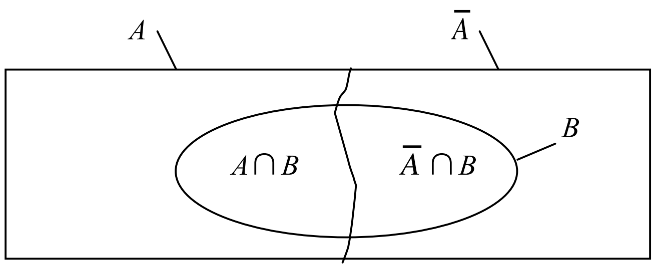
$$P(C) = P(A \cap B \cap C) = 0{,}94 \cdot 0{,}91 \cdot 0{,}84 = 0{,}72$$

Sannsynligheten er 72 % for at en 65 år gammel kvinne skal bli minst 80 år.

5. Total sannsynlighet

I eksempel 5 fant vi sannsynligheten for hendelsen B = "andre kort er svart" ved hjelp av sannsynlighetene for hendelsene A = "første kort er rødt" og $\overline{A}$ = "første kort er svart" og de betingede sannsynlighetene for B gitt A og B gitt $\overline{A}$.

Resonnementet i eksemplet har generell gyldighet.



Venndiagrammet viser at vi kan skrive B som en union av de disjunkte hendelsene $A \cap B$ og $\overline{A} \cap B$.
Addisjonssetningen for disjunkte hendelser gir derfor at

$$P(B) = P(A \cap B) + P(\overline{A} \cap B) \tag{8}$$

Av produktsetningen får vi

$$P(A \cap B) = P(A) \cdot P(B|A) \tag{9}$$

$$P(\overline{A} \cap B) = P(\overline{A}) \cdot P(B|\overline{A}) \tag{10}$$

Setter vi (9) og (10) inn i (8), får vi

Setningen om total sannsynlighet

La A og B være hendelser ved et forsøk. Da har vi at

$$P(B) = P(A) \cdot P(B|A) + P(\overline{A}) \cdot P(B|\overline{A}) \qquad (11)$$

La oss illustrere setningen om total sannsynlighet med et lite eksempel (motivert av spillet "redblack" som omtales i avsnitt 7).

Eksempel 7

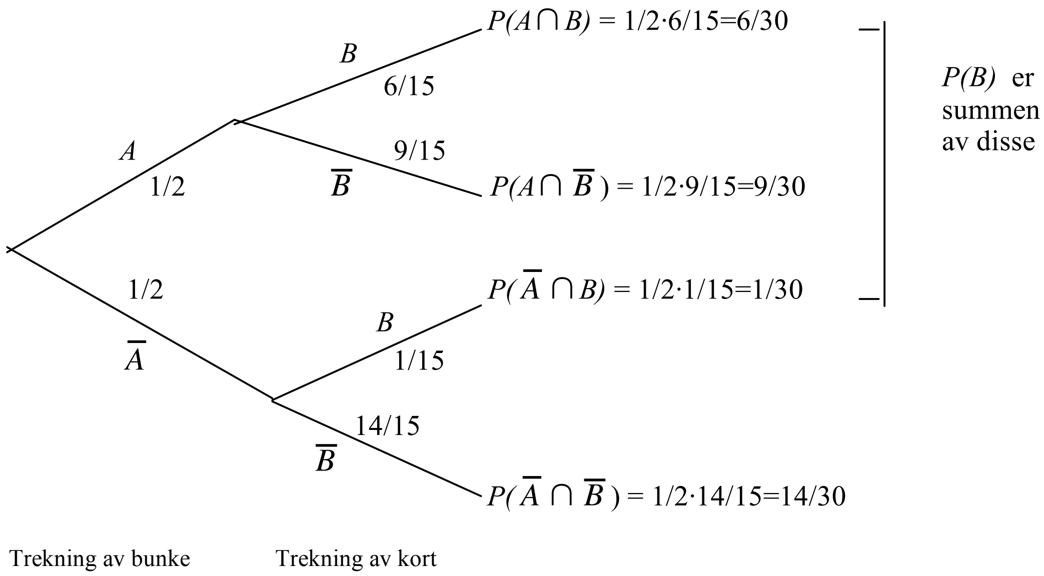
Fra en kortstokk velger vi ut fire røde kort og to svarte kort og legger i en bunke (bunke I). Vi velger også ut to røde kort og fire svarte kort og legger i en annen bunke (bunke II).
Vi velger tilfeldig en av bunkene og trekker to kort fra denne. Hva er sannsynligheten for at vi får to røde kort? La nå A = "vi trekker fra bunke I" og B = "vi trekker to røde kort".
Da er (jfr. produktsetningen)

$$P(A) = P(\overline{A}) = \frac{1}{2} \qquad P(B \mid A) = \frac{4}{6} \cdot \frac{3}{5} = \frac{6}{15} \qquad P(B \mid \overline{A}) = \frac{2}{6} \cdot \frac{1}{5} = \frac{1}{15}$$

Dermed gir (11)

$$P(B) = \frac{1}{2} \cdot \frac{6}{15} + \frac{1}{2} \cdot \frac{1}{15} = \frac{7}{30}$$

For å få oversikt, og til hjelp i beregningene, kan vi bruke sannsynlighetstreet på figuren nedenfor. Ved å gange sammen (de betingede) sannsynlighetene langs greinene av treet får vi sannsynlighetene til høyre på figuren; jfr. produktsetningen. Ved å legge sammen den øverste og den nest nederste sannsynligheten får vi *P(B)*; jfr setningen om total sannsynlighet.



I (11) delte vi opp utfallsrommet i to hendelser A og $\overline{A}$; jfr. venndiagrammet i starten av avsnittet. Setningen om total sannsynlighet gjelder også hvis vi deler opp utfallsrommet i mer enn to hendelser:

Den generelle setningen om total sannsynlighet

La $A_1, A_2, \ldots, A_n$ være n disjunkte hendelser, og anta at unionen av dem er hele utfallsrommet. Da har vi

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B \mid A_i) \tag{12}$$

Denne setningen bruker vi i avsnitt 7 når vi skal analysere spillet "redblack".

6. Bayes' setning

Vi har et forsøk med to hendelser A og B. Ved å kombinere (2) med (11) får vi:

Bayes' setning

La A og B være hendelser ved et forsøk. Da er

$$P(A \mid B) = \frac{P(A) \cdot P(B \mid A)}{P(A) \cdot P(B \mid A) + P(\overline{A}) \cdot P(B \mid \overline{A})} \tag{13}$$

Bayes' setning er nyttig i situasjoner hvor det er lett å bestemme $P(B|A)$ og $P(B|\overline{A})$, og vi ønsker å finne den "omvendte" betingede sannsynligheten $P(A|B)$.

La oss illustrere Bayes' setning med situasjonen i eksempel 7.

Eksempel 8

Vi har fire røde kort og to svarte kort i bunke I og to røde kort og fire svarte kort i bunke II.
Vi velger tilfeldig en av bunkene og trekker to kort fra denne. Hvis begge de to kortene vi trekker er røde, hva er da sannsynligheten for at vi har trukket fra bunke I? Vi spør om $P(A|B)$. Ved å bruke Bayes' setning får vi

$$P(A \mid B) = \frac{P(A) \cdot P(B \mid A)}{P(A) \cdot P(B \mid A) + P(\overline{A}) \cdot P(B \mid \overline{A})} = \frac{\frac{1}{2} \cdot \frac{6}{15}}{\frac{1}{2} \cdot \frac{6}{15} + \frac{1}{2} \cdot \frac{1}{15}} = \frac{6}{7}$$

Bayes' setning er sentral når en skal studere såkalt diagnostiske tester, det være seg graviditetstester, mammografiundersøkelser eller dopingtester. En lignende anvendelse er bruk av løgndetektor (som er lite brukt i Norge, men mer utbredt i USA):

Eksempel 9

En løgndetektor er et instrument som registrerer fysiologiske størrelser (bl.a. blodtrykk, puls og åndedrett) hos en person mens hun eller han svarer på spørsmål. Formålet er å bruke endringer i de fysiologiske størrelsene til å avgjøre om personen lyver eller ikke. I USA blir løgndetektor blant annet brukt i forbindelse med politiavhør av vitner og mistenkte.

I en studie av løgndetektoren fant en at:

- Hvis en person lyver, er det 90% sannsynlig at løgndetektoren vil avsløre dette.
- Hvis en person snakker sant, er det 15% sannsynlig at løgndetektoren likevel vil indikere at personen lyver.

Et vitne i en kriminalsak testes med en løgndetektor. Testen indikerer at vitnet lyver. Hva er sannsynligheten for at vitnet virkelig lyver hvis sannsynligheten for at vitne lyver (etter en subjektiv vurdering) settes til 25% ?

Vi betrakter hendelsene L = "vitnet lyver" og + = "testen indikerer at vitne lyver"
Av opplysningene over har vi

$$P(+|L) = 0{,}90 \qquad P(+|\overline{L}) = 0{,}15. \qquad P(+) = 0{,}25$$

Vi ønsker å finne den betingede sannsynligheten $P(L\,|+)$. Den finner vi av Bayes' setning:

$$P(L\,|+) = \frac{P(+)\cdot P(+\,|\,L)}{P(+)\cdot P(+\,|L) + P(\overline{L})\cdot P(+\,|\,\overline{L})} = \frac{0{,}25\cdot 0{,}90}{0{,}25\cdot 0{,}90 + 0{,}75\cdot 0{,}15} = 0{,}67$$

Det er 67% sannsynlig at vitnet lyver.

Bayes' setning gjelder også om vi deler opp utfallsrommet i mer enn to hendelser. Ved å kombinere (2) med (12) får vi

Bayes' setning på generell form

La $A_1, A_2, \ldots, A_n$ være n disjunkte hendelser, og anta at unionen av dem er hele utfallsrommet.
Da har vi

$$P(A_k\,|\,B) = \frac{P(A_k)\cdot P(B\,|\,A_k)}{\sum_{i=1}^{n} P(A_i)\cdot P(B\,|\,A_i)} \tag{14}$$

7. Redblack — et "sannsynlighetsspill"

Dette avsnittet beskriver et enkelt "sannsynlighetsspill" som kan brukes til å aktivisere klassen.

Gjennomføring og spilleregler

Spillet foregår på følgende måte: Elevene deles inn i grupper med to elever i hver gruppe. De to elevene i en gruppe skal annenhver gang være "spiller" og "spilleleder." Hver av elevene spiller et bestemt antall ganger (f.eks. 10 ganger) og poengene noteres for hver omgang (jfr. nedenfor). Den eleven i klassen som får flest poeng til sammen vinner spillet.

Hvert enkelt spill foregår på følgende måte: Spillelederen har to bunker med seks kort i hver. I den ene bunken er det 4 røde og 2 svarte kort, i den andre er det 2 røde og 4 svarte kort. Spillelederen legger bunkene i vifteform foran spilleren med baksiden opp. Spillerens oppgave er å gjette i hvilken av bunkene det er 4 røde og 2 svarte kort. Spilleren får (selvfølgelig!) bare ett forsøk på å gjette. Kortene må stokkes godt mellom hver omgang. (Hvis en bruker kort fra en vanlig kortstokk, må alle de tolv kortene stokkes sammen og deles på nytt inn i to bunker mellom hver omgang.)

Spilleren kan:

- Gjette uten å trekke noen kort. Hvis hun gjetter riktig får hun 10 poeng, hvis ikke får hun 0 poeng.
- Trekke to kort fra en av bunkene. Spilleren kan nå gjette uten å trekke flere kort. Hvis hun gjetter riktig får hun 8 poeng, hvis ikke får hun 0 poeng.
- Trekke ett kort til fra den samme bunken som de to kortene ble trukket fra. Nå må spilleren gjette. Hvis hun gjetter riktig, får hun 5 poeng, hvis ikke får hun 0 poeng.

Pedagogisk idé

Tanken med spillet er at elevene skal oppleve og få erfaring med at de får mer informasjon om hvilken bunke som inneholder de 4 røde kortene etter hvert som de trekker. Men det spørsmålet de (intuitivt) må ta standpunkt til, er om den ekstra informasjonen de får ved å trekke kort er verd den reduksjonen dette gir i poeng. Spillet gir videre opphav til en rekke problemstillinger læreren kan ta opp i klassen etter at elevene er ferdig med å spille. Og det gir mulighet for å ta opp noen perspektiver som peker ut over dette enkle spillet. Endelig er spillet lagd så enkelt at det er mulig for elevene (med hjelp fra læreren) å foreta en fullstendig analyse av spillet og avgjøre hva som er en "optimal strategi".

Spørsmål og problemstillinger

Etter at elevene har spilt, kan de (i grupper eller i diskusjon i klassen) prøve å svare på f.eks. følgende spørsmål. For enkelhets skyld kaller vi den bunken som har 4 røde og 2 svarte kort for bunke I og den med 2 røde og 4 svarte kort for bunke II.

- Hva er sannsynligheten for å gjette riktig uten å trekke noen kort?
- Gitt at en trekker to kort fra bunke I, hva er sannsynligheten for å få
  — to røde kort?
  — to svarte kort?
  — ett kort av hver farge?
- Hva blir sannsynlighetene ovenfor gitt at en trekker fra bunke II?
- Er hendelsene "trekker fra bunke I" og "to røde kort" avhengige eller uavhengige hendelser? Hva med hendelsene "trekker fra bunke I" og "ett kort av hver farge"?
- Hvis en trekker to røde kort, hva er da (den betingede) sannsynligheten for at en trekker fra bunke I? Hva er denne (betingede) sannsynligheten hvis en trekker to svarte kort? Hva er den hvis vi trekker ett kort av hver farge?
- Gitt at vi trekker to+ett kort fra bunke I, hva er da den betingede sannsynligheten for å
  — først to røde kort og så et rødt kort til?
  — først to røde kort og så et svart kort?
  — først to svarte kort og så et rødt kort?
  — først ett kort av hver farge og så et rødt kort?
  — først ett kort av hver farge og så et svart kort?
- Hva er den beste strategien en kan bruke når en spiller "redblack"? (Jfr. nedenfor.)

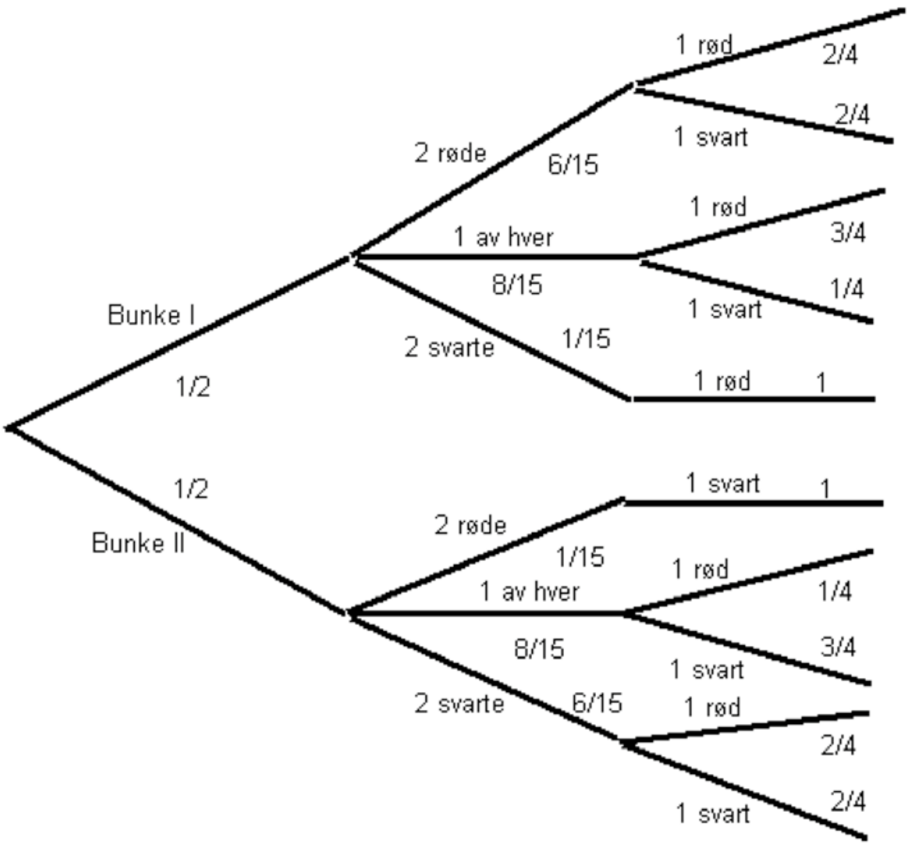
Selv om "redblack" er et enkelt spill, gir det noen perspektiver som læreren kan diskutere med elevene. Poenget i spillet er å trekke en konklusjon på grunnlag av usikker informasjon. Og dette er en problemstilling en ofte stilles ovenfor. I spillet skal en på grunnlag av den informasjonen de trukne kortene gir, prøve å gjette hvilken bunke som har 4 røde kort. Når en bruker en løgndetektor (jfr. eksempel 9), skal en på lignende måte prøve å avgjøre om personen lyver på grunnlag av den informasjon testen gir. Og ved en mammografiundersøkelse skal en på grunnlag av mammogrammet prøve å avgjøre om kvinnen har brystkreft eller ikke.

En analyse av spillet:

For å få oversikt over spillet og til hjelp i beregningene, kan vi benytte sannsynlighetstreet på neste side. Vi vil bruke det til å analysere vinnersjansene og forventet poengsum ved fire mulige spillestrategier.

Strategi I: Eleven gjetter uten å trekke noen kort. Sannsynligheten for å gjette riktig er da 50%, og forventet[3] poengsum er $10 \cdot 0{,}50 = 5$.

Strategi II: Eleven trekker to kort og gjetter uansett hvilke kort han får. Hvis han får to røde kort, gjetter han at han trekker fra bunke I. Hvis han får to svarte kort, gjetter han at han at det er den andre bunken som er bunke I. Hvis han får ett kort av hver farge, gjetter han at han trekker fra bunke I. (Det blir samme resultat om han gjetter "tilfeldig" på en av de to bunkene hvis han får ett kort av hver farge.) Sannsynligheten for å gjette riktig med denne strategien kan vi finne ut fra sannsynlighetstreet. Formelt svarer dette til å bruke den generelle setningen (12) om total sannsynlighet



Valg av bunke Trekning av 2 kort Trekning av ett kort

Vi finner at sannsynligheten for å gjette riktig er

$$\frac{1}{2}\cdot\frac{6}{15}+\frac{1}{2}\cdot\frac{8}{15}+0+0+0+\frac{1}{2}\cdot\frac{6}{15}=\frac{2}{3}$$

Forventet poengsum med denne strategien er derfor $8\cdot\frac{2}{3}=5{,}3$.

[3] Forventningsverdi er noe elevene først lærer formelt i VKII. Men i en så enkel situasjon som den vi studerer her, er det "intuitivt opplagt" hva forventningen er.

Strategi III: Eleven trekker først to kort og så ett kort til (uansett hva de to første kortene er). Hvis 2 eller 3 av de kortene han har trukket er røde, gjetter han at han har trukket fra bunke I. Ellers gjetter han at det er den andre bunken som bunke I. Igjen kan vi beregne sannsynligheten for å gjette riktig ut fra sannsynlighetstreet, eller formelt ved å bruke (12).Vi finner at sannsynligheten for å gjette riktig nå er

$$\frac{1}{2}\cdot\frac{6}{15}\cdot\frac{2}{4}+\frac{1}{2}\cdot\frac{6}{15}\cdot\frac{2}{4}+\frac{1}{2}\cdot\frac{8}{15}\cdot\frac{3}{4}+0+0+0+0+\frac{1}{2}\cdot\frac{8}{15}\cdot\frac{3}{4}+\frac{1}{2}\cdot\frac{6}{15}\cdot\frac{2}{4}+\frac{1}{2}\cdot\frac{6}{15}\cdot\frac{2}{4}=\frac{4}{5}$$

Forventet poengsum med denne strategien er derfor $5\cdot\frac{4}{5}=\frac{20}{5}=4$.

Strategi IV: Spilleren trekker først to kort. Hvis han får to røde kort, gjetter han at han trekker fra bunke I. Hvis han får to svarte kort, gjetter han at han at det er den andre bunken som er bunke I. Hvis han får ett kort av hver farge, trekker han et kort til. Hvis det siste kortet han trekker er rødt (dvs. hvis 2 av de 3 kortene han har trukket er røde), gjetter han på at han trekker fra bunke I. Ellers gjetter han på at det er den andre bunken som er bunke I. Med denne strategien er sannsynligheten for at han gjetter riktig i andre omgang (og dermed får 8 poeng) lik

$$\frac{1}{2}\cdot\frac{6}{15}+\frac{1}{2}\cdot\frac{6}{15}=\frac{2}{5}$$

og sannsynligheten for at han gjetter riktig i tredje omgang (og dermed får 5 poeng) er lik

$$\frac{1}{2}\cdot\frac{8}{15}\cdot\frac{3}{4}+\frac{1}{2}\cdot\frac{8}{15}\cdot\frac{3}{4}=\frac{2}{5}$$

Forventet poengsum med denne strategien er derfor $8\cdot\frac{2}{5}+5\cdot\frac{2}{5}=5{,}2$.

Denne analysen viser oss at det er strategi II som er den optimale.

8. Forslag til prosjektoppgave

En mulig prosjektoppgave (stor oppgave) er å be elevene foreta en grundig analyse av spillet "redblack" og hvilken strategi som lønner seg i dette spillet. Dette forutsetter at læreren ikke har gjennomgått stoffet i forrige avsnitt i detalj.

En oppgave som har et "videre perspektiv" er knyttet til problemstillinger rundt diagnostisk testing.

Her er noen problemstillinger:

- Hvordan defineres betinget sannsynlighet og hva sier Bayes' setning?
- Hva er en diagnostisk test?
- Hva betyr begrepene *sensitivitet* og *spesifisitet* for en diagnostisk test?
- Et eksempel på en diagnostiske test er en mammografiundersøkelse. Finn litt ut om hvordan denne testen virker og hvor pålitelige den er. Angi også minst en annen diagnostisk test fra medisin eller andre områder og beskriv denne/disse.
- I det norske mammografiprogrammet innkalles kvinner i aldersgruppen 50-69 år (se http://www.kreftregisteret.no/ for nærmere informasjon). Diskuter grunnen til at en ikke også kaller inn yngre kvinner.

Litteratur:

Odd O. Aalen. *Innføring i statistikk med medisinske eksempler.* 2. utgave. Ad Notam Gyldendal. 1998. ISBN 82-417-0929-3.

Kapitlene 3 og 4 inneholder en enkel og leseverdig innføring i sannsynlighetsregning fram til og med Bayes' setning. Diagnostisk testing behandles spesielt i avsnittene 4.8 og 4.9.