

Föreläsning 2

Dragning utan återläggning

Sid 7 prentes

$$P(h \text{ vita}) = \frac{\binom{V}{h} \binom{S}{n-h}}{\binom{V+S}{n}}$$

Vilka vita vi tar
påverkas ej av vilka svart
vi tar och tvärtom

$$P(h_1 \text{ av sort 1 } \cap h_2 \text{ av sort 2 } \dots) = \frac{\binom{V_1}{h_1} \binom{V_2}{h_2} \dots \binom{V_r}{h_r}}{\binom{N}{n}}$$

$V_1 + \dots + V_r = N$
 $h_1 + \dots + h_r = n$

Dragning med återläggning ~~multinomial~~

$$\binom{N}{h} p^h (1-p)^{n-h} = \binom{N}{h} \frac{V^h S^{n-h}}{(V+S)^n}$$

multinomial
formel
för lösningen

$$\frac{N!}{h_1! h_2! \dots h_r!} \left(\frac{V_1}{N}\right)^{h_1} \left(\frac{V_2}{N}\right)^{h_2} \dots$$

$$N = V_1 + \dots + V_r$$

Kap 2

2.6

Betingad sannolikhet

sannolikheten för B betingad A skrivs $P(B|A)$

	Röhre	icke-röhre	
MÅN	20	80	100
kvinnor	50	100	150
	70	180	250

A - Mån B - röhre

$$P(A) = \frac{100}{250} = 0.4$$

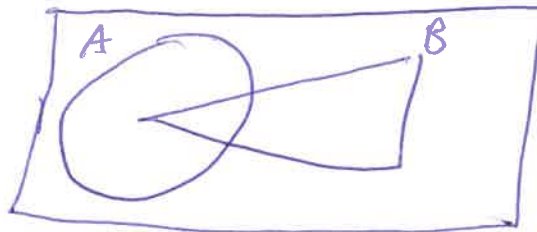
$$P(B) = \frac{70}{250} = 0.28$$

$$P(A \cap B) = \frac{20}{250} = 0.08$$

$$P(B|A) = \frac{20}{100} = \frac{\frac{20}{250}}{\frac{100}{250}} = \frac{P(A \cap B)}{P(A)}$$

Detta är betingningsformeln

Se även Venn diagram



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

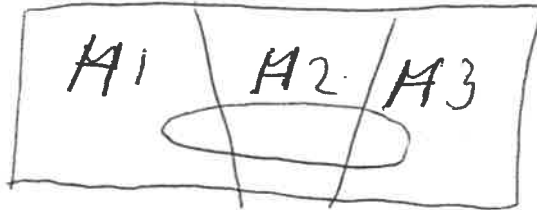
Kap 2 föreläsning 2 sid 2

Visa betingningsformeln

~~Räkna 2.27~~

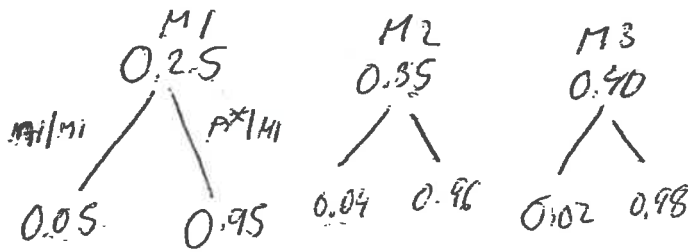
exempel 2.17

Rita Venn diagram för
sedan formeln
sedan exemplet
sedan formeln igen



$$P(A) = P(A \cap M1) + P(A \cap M2) + P(A \cap M3)$$

$$= P(A|M1) \cdot P(M1) + P(A|M2) \cdot P(M2) + P(A|M3) \cdot P(M3)$$



$$= 0,0345$$

P(A) =

Detta är lagen om total sth

ex 2.19
$$P(H_1|A) = \frac{P(A|H_1) \cdot P(H_1)}{\sum_i P(A|H_i) \cdot P(H_i)} = 0,36$$

Definiera över utgående från betingning

~~gör ex 2.20~~
gör ex 2.23

gör ex 2.20
$$P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}$$

om över

En person utsätter sig för en olycksrisk

på $\frac{1}{1000}$ 1000 gånger

$$P(\text{minst en olycka}) = 1 - P(\text{ingen olycka}) = 1 - \left(1 - \frac{1}{1000}\right)^{1000} \approx 0,63$$

Räkna ~~2.27~~

Exempel 2.17 Flermaskinstilverkning

I en fabrik tillverkas 25 % av enheterna vid maskin 1, 35 % vid maskin 2 och 40 % vid maskin 3. Av produktionen är respektive 5 %, 4 % och 2 % defekt. Man blandar enheterna och sänder dem till kunderna. Hur stor är sannolikheten att en slumpmässigt vald enhet är felaktig?

Låt i Sats 2.9 händelsen H_i vara "enheten tillverkas vid maskin i " och låt händelsen A vara "enheten är felaktig". Då finner man

$$P(A) = 0.25 \cdot 0.05 + 0.35 \cdot 0.04 + 0.40 \cdot 0.02 = 0.0345. \quad \square$$

Exempel 2.18 Användning av delfall

I en DNA-kedja om 8 baser blir det en mutation i två olika på måfå. Vi söker sannolikheten att dessa två finns intill varandra. Låt $H_1 =$ "den första mutationen är i nummer 1 eller 8" och $H_2 =$ "den första mutationen finns bland nr 2-7". Om $A =$ "de två mutationerna finns intill varandra" så ger Sats 2.9 och klassisk sannolikhetsdefinition att

$$P(A) = P(A|H_1)P(H_1) + P(A|H_2)P(H_2) = \frac{1}{7} \cdot \frac{2}{8} + \frac{2}{7} \cdot \frac{6}{8} = \frac{1}{4}.$$

Detta eftersom baserna i ändarna (dvs H_1) bara har en gränne medan de i mitten (dvs H_2) har 2 stycken grannar.

Vi hade kunnat få resultatet genom direkt användning av klassisk sannolikhetsdefinition genom att betrakta alla delmängder av storlek 2, varav 7 stycken är gynnsamma för A medan det finns $\binom{8}{2} = 28$ stycken möjliga. I så fall har vi betraktat de två valda utan hänsyn till ordning, men om vi tagit hänsyn till ordningen hade både antalet för A gynnsamma och antalet möjliga ökat med en faktor 2, varför sannolikheten blivit densamma. \square

Ofta vill man "vända" på en betingning. En vanlig varningstext på cigarettpaketet är "9 av 10 med strupcancer är rökare". Med $A =$ "rökare" och $B =$ "får strupcancer" kan detta tolkas som $P(A|B) = 9/10$. Mer intressant vore att få veta $P(B|A) =$ risken att få strupcancer om man är rökare och jämföra denna med $P(B|A^*) =$ risken att få strupcancer om man inte är rökare. Enligt (2.9) gäller att

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} \quad (2.12)$$

och man ser att man måste veta $P(A)$, dvs hur vanlig rökning är, samt $P(B)$, dvs hur vanlig strupcancer är.

Om vi i (2.12) låter $B = H_i$ och använder Sats 2.9 på $P(A)$ erhålls följande sats.

Sats 2.10 Bayes sats

Under samma villkor som i Sats 2.9 gäller

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{j=1}^n P(H_j)P(A|H_j)}. \quad (2.13)$$

Redan på 1700-talet angav den engelske prästen Bayes en formel liknande (2.13). Sällan har någon formel missbrukats så ofta av så många. Orsaken härtill är bl.a. att sannolikheterna $P(H_i)$ ofta är svåra att bestämma i praktiken. Ett lättvindigt sätt att då bli av med dem är helt enkelt att anta dem vara lika, varvid de kan förkortas bort ur formeln! Att detta förfaringssätt i regel inte är tillåtet framgår av följande exempel.

Exempel 2.19 Flermaskinstilverkning (forts. från Exempel 2.17)

Antag att en kund påträffar en felaktig enhet. Hur stor är sannolikheten att den har tillverkats vid maskin 1? Bayes sats ger svaret ($A =$ "defekt enhet")

$$P(H_1|A) = \frac{0.25 \cdot 0.05}{0.25 \cdot 0.05 + 0.35 \cdot 0.04 + 0.40 \cdot 0.02} = 0.36.$$

(Om vi hade sett sannolikheterna $P(H_i)$ lika, hade vi fått det felaktiga resultatet $P(H_1|A) = 0.05/(0.05 + 0.04 + 0.02) = 0.45$.) \square

Exempel 2.20 Diagnostiskt test

Ett diagnostiskt test för en viss sjukdom är sådant att om personen har sjukdomen ger testet ett positivt utslag med sannolikheten 0.9999 (kallas sensitiviteten hos testet). Om personen inte har sjukdomen ger testet ett negativt utslag med sannolikheten 0.995 (kallas specificiteten). Vi vill veta sannolikheten att personen har sjukdomen om testet gett ett positivt utslag.

Formuleringen av problemet är ofullständig ty man måste veta hur vanlig sjukdomen är. Låt sannolikheten för sjukdomen vara p (kallas prevalensen av sjukdomen). Vi undrar nu över vad sannolikheten (uttryckt i p) är att en person vars test ger positivt utslag verkligen har sjukdomen. Vi låter $H_1 =$ "personen har sjukdomen" och $H_2 = H_1^c =$ "personen har inte sjukdomen". Vi ser att $P(H_1) = p$ och $P(H_2) = 1 - p$. Låt $A =$ "testet ger positivt utslag". Vi har $P(A|H_1) = 0.9999$ och $P(A^*|H_2) = 0.995$. Vad vi söker är $P(H_1|A)$. Bayes sats ger resultatet

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{P(H_1)P(A|H_1) + P(H_2)P(A|H_2)} = \frac{p \cdot 0.9999}{p \cdot 0.9999 + (1 - p) \cdot 0.005}.$$

Om $p = 0.2$ blir $P(H_1|A) \approx 0.98$ men om $p = 0.001$ blir $P(H_1|A) \approx 0.17$. Resultatet har en viss relevans för frågan om man skall masstesta. I fallet $p = 0.2$ (högrisk-grupp) är nästan alla med positivt utslag sjuka, men om $p = 0.001$ (lågrisk-grupp) är huvuddelen av de med positivt utslag egentligen friska. \square