Statistik och Dataanalys I

Föreläsning 12 - Betingade sannolikheter och Bayes sats

Oskar Gustafsson

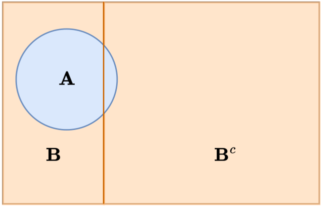
Statistiska institutionen
Stockholms universitet

Översikt

- **Betingad sannolikhet**
- **Lagen om total sannolikhet**
- **Bayes sats**

Betingad sannolikhet

- Covid:
  - A $= \{$positivt hemtest$\}$. B $= \{$har covid$\}$.
  - Intresse: $P(B|A) = P(\text{har covid}|\text{positivt hemtest})$
- Tecknet | läses 'givet' eller betingat på.
- Sannolikheten för att A inträffar **givet att B har inträffat**.

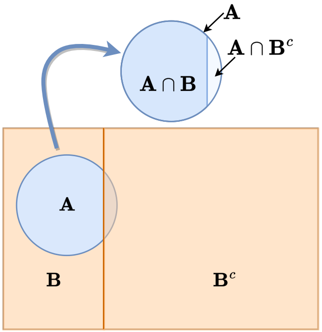


Betingad sannolikhet - världen krymper

- **Betingad sannolikhet**

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

- **Betinga på A** innebär att blå cirkeln blir **vårt nya utfallsrum**.
- Inget utanför blå cirkeln kan längre inträffa. "A is the new S".



Korstabeller och sannolikheter

Antal

Pets		Gender: Female	Male	Total
	Has cats	3412	2388	5800
	Has dogs	3431	3587	7018
	Has both	897	577	1474
	Total	7740	6552	14,292

Snittsannolikheter (Table %)

Pets		Gender: Female	Male	Total
	Has cats	23.9%	16.7%	40.6%
	Has dogs	24.0%	25.1%	49.1%
	Has both	6.3%	4.0%	10.3%
	Total	54.2%	45.8%	100%

Betingat på kön (Column %)

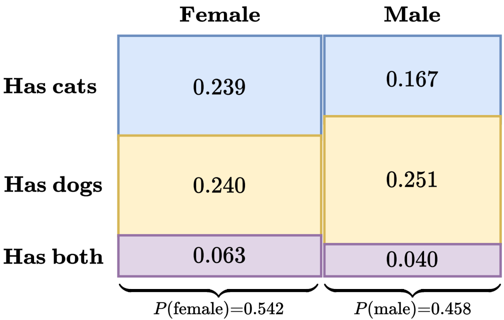
Pets		Gender: Female	Male	Total
	Has cats	44.1%	36.4%	40.6%
	Has dogs	44.3%	54.8%	49.1%
	Has both	11.6%	8.8%	10.3%
	Total	100%	100%	100%

Betingat på husdjur (Row %)

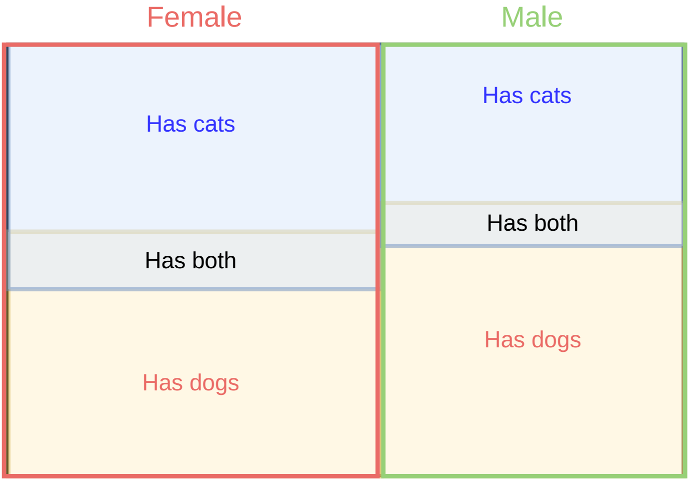
Pets		Gender: Female	Male	Total
	Has cats	58.8%	41.2%	100%
	Has dogs	48.9%	51.1%	100%
	Has both	60.9%	39.1%	100%
	Total	54.2%	45.8%	100%

Korstabell och mosaic-plott

		Gender		
		Female	Male	Total
Pets	Has cats	23.9%	16.7%	40.6%
	Has dogs	24.0%	25.1%	49.1%
	Has both	6.3%	4.0%	10.3%
	Total	54.2%	45.8%	100%



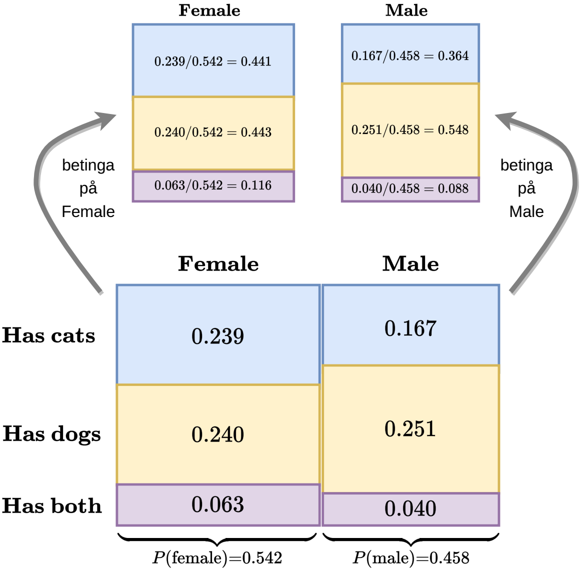
Venndiagram



Korstabell och betingad sannolikhet

Pets	Gender		
	Female	**Male**	**Total**
Has cats	44.1%	36.4%	40.6%
Has dogs	44.3%	54.8%	49.1%
Has both	11.6%	8.8%	10.3%
Total	**100%**	**100%**	**100%**

Korstabell och betingad sannolikhet



Allmänna multiplikationsregeln

Allmänna multiplikationsregeln. För händelser A och B

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

- Terminologi:

$$\underbrace{P(A \cap B)}_{\text{snittsannolikhet}} = \underbrace{P(A)}_{\text{marginell sannolikhet}} \cdot \underbrace{P(B|A)}_{\text{betingad sannolikhet}}$$

- A och B är **oberoende händelser** om (och endast om)

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

Oberoende händelser - variant.
A och B är oberoende om (och endast om)

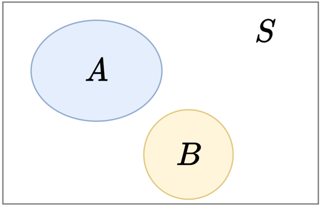
$$P(B|A) = P(B)$$

Oberoende händelser

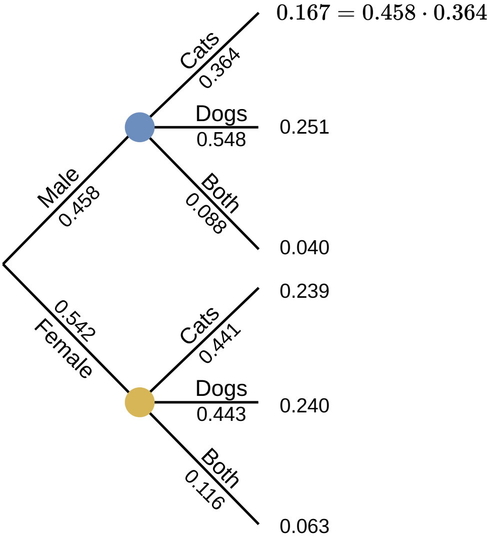
Oberoende händelser - variant.
A och B är oberoende om (och endast om)

$$P(B|A) = P(B)$$

- Oberoende händelser - vetskapen om att A har inträffat påverkar inte sannolikheten för B.
- **Oberoende $\neq$ Disjunkta**. Disjunkta händelser kan ju inte inträffa samtidigt!



Betingade sannolikheter bäst i form av träd



Snitt- och marginella sannolikheter bäst i tabell

simultansannolikheter (joint) kön och husdjur

Pets	Gender: Female	Male	Total
Has cats	23.9%	16.7%	40.6%
Has dogs	24.0%	25.1%	49.1%
Has both	6.3%	4.0%	10.3%
Total	54.2%	45.8%	100%

marginalsannolikheter husdjur

marginalsannolikheter kön

Bayes sats

- **Allmänna multiplikationsregeln**

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

- **Betingad sannolikhet**

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

- **Bayes sats** 😍

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

- **Bayes vänder betingningar**: beräkna $P(B|A)$ från $P(A|B)$.

Känna igen handskrivna siffror

- Förenkling: skilja på enbart 0:or och 1:or

- A ={vit pixel i mitten} och B ={siffran är en nolla}

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

$$P(B) = \frac{\text{antal bilder med nollor}}{\text{totalt antal bilder}}$$

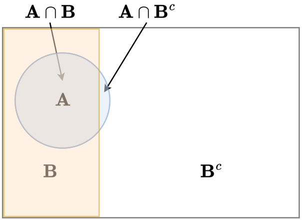
$$P(A) = \frac{\text{antal bilder med vit pixel i mitten}}{\text{totalt antal bilder}}$$

$$P(A|B) = \frac{\text{antal bilder med nolla som också har vit pixel i mitten}}{\text{antal bilder med nollor}}$$

Lagen om total sannolikhet

- Sannolikheten för varje händelse A kan delas upp som:

$$P(A) = P(A \cap B) + P(A \cap B^c)$$



- Allmänna multiplikationsregeln:

$$P(A \cap B) = P(A|B)P(B) \text{ och } P(A \cap B^c) = P(A|B^c)P(B^c)$$

- **Lagen om total sannolikhet**

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

Bayes sats - via lagen om total sannolikhet

- **Lagen om total sannolikhet**

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

- Bayes sats

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

- **Bayes sats** med lagen om total sannolikhet

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

Fungerar hemtest för Covid?

- **Bayes sats**

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

- Covid: A $= \{$pos$\}$. B $= \{$covid$\}$

$$P(\text{covid}|\text{pos}) = \frac{P(\text{pos}|\text{covid})P(\text{covid})}{P(\text{pos}|\text{covid})P(\text{covid}) + P(\text{pos}|\text{inte covid})P(\text{inte covid})}$$

- Notera: $P(\text{neg}|\text{inte covid}) = 1 - P(\text{pos}|\text{inte covid})$.

- **Prevalens**: $P(\text{covid})$ - andel med covid i populationen.
- **Sensitivitet**: $P(\text{pos}|\text{covid})$ - hur **känsligt** är testet för att upptäcka covid?
- **Specificitet**: $P(\text{neg}|\text{inte covid})$ - är testet **specifikt** för covid, eller reagerar det även på annat?

Fungerar hemtest för covid?

antigentest för detektering av Covid-19 är ett CE-certifierat test för självprovtagning som kan indikera pågående infektion av coronavirus. Snabbtest som utförs genom nästopsning i främre näsan. Med hög $\boxed{\text{specificitet på 99,20 \%}}$ samt hög $\boxed{\text{sensitivitet på 96,77 \%}}$. Passar för screening av symptomfria individer exempelvis på arbetsplatser.

- Sensitivitet: $P(\text{pos test}|\text{covid}) = 0.9677$
- Specificitet: $P(\text{neg test}|\text{inte covid}) = 0.9920$

Bayes sats

Den här widgeten låter dig undersöka hur tillförlitliga hemtest för Covid är genom att beräkna den betingade sannolikheten P(covid | positivt test) med hjälp av Bayes sats. Du kan också använda widgeten till andra problem, genom att ge händelserna covid och pos andra namn.

Event A:	cov
Event B:	pos
$P(\text{pos} \mid \text{cov})$	0.9677
$P(\text{not pos} \mid \text{not cov})$	0.992
$P(\text{cov})$	0.05

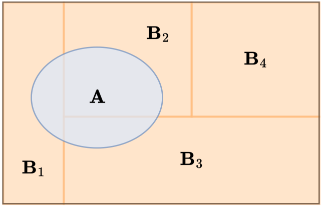
$\text{P}(\text{cov} \mid \text{pos}) = 0.8642$

Mattias Villani Bayes' theorem for events

Observable

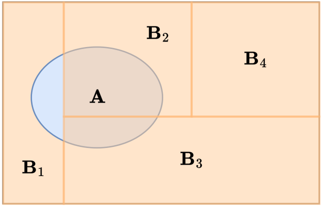
- Notera dock att vanligtvis är $P(\text{covid}|\text{symptom})$ större än prevalensen $P(\text{covid})$. Man testar sig pga symptom. Se slutet av denna notebook.

Lagen om total sannolikhet - allmän version



$$P(A) = \sum_{j=1}^{K} P(A|B_j)p(B_j)$$

Bayes sats - allmän version



$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{j=1}^{K} P(A|B_j)p(B_j)}$$

- Exempel: handskrivna siffor:
  - B_0 ={nolla}, B_1 = {etta}, B_2 = {tvåa}, ... , B_9 = {nia}.
  - A ={vit pixel i mitten}

Credits

Dessa slides skapades för kursen statistik och dataanalys 1 av Mattias Villani HT 2023, och har modifierats av Oscar Oelrich för statistik och dataanalys 1 VT 2024.