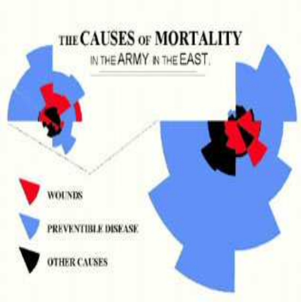
SF1901: Sannolikhetslära och statistik
Föreläsning 1.

Jan Grandell & Timo Koski

19.01.2016

KTH Matematik

Många tänker på tabeller [1] när de hör ordet "statistik". Här avses dock med *statistik* läran om hur man från observationer eller analyser under osäkerhet drar slutsatser och beskriver dessa slutsatser på ett korrekt sätt.



[1] Florence Nightingale (i bilden till höger) var även en framstående statistiker, tabellen till vänster har uppfunnits av henne

Lärandemål:

- aritmetiskt medelvärde, standardavvikelse, relativ frekvens
- utfall, utfallsrum, händelse, omöjlig (tom) händelse
- union av två händelser, snitt av två händelser
- komplementhändelse, De Morgans regler
- sannolikhet (snl), räknereglerna för sannolikhet
- Ex. på snl: $\frac{\text{antalet gynnsamma fall}}{\text{totalantalet fall}}$, kombinatorik

KTH Matematik

En studie från Google Inc. (2011)

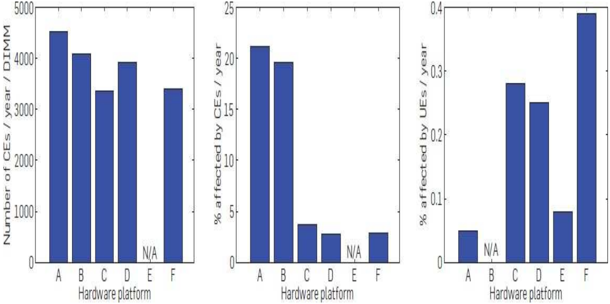
Errors in dynamic random access memory (DRAM) are a common form of hardware failure in modern compute clusters. Failures are costly both in terms of hardware replacement costs and service disruption. In this we analyze measurements of memory errors in Google's fleet of commodity servers over a period of 2.5 years. The collected data covers multiple vendors, DRAM capacities and technologies, and comprises many millions of DIMM (= Dual In-line Memory Module) days.

The goal is to answer questions such as the following: How common are memory errors in practice? What are their statistical properties? How are they affected by external factors, such as temperature and utilization, and by chip-specific factors, such as chip density, memory technology and DIMM age?

KTH Matematik

En studie från Google Inc. (2011)

Frequency of errors: The average number of correctable errors (CEs) per year per DIMM (left), the fraction of DIMMs that see at least one CE in a given year (middle) and the fraction of DIMMs that see at least one uncorrectable error (UE) in a given year (right). Platforms C, D, and F use SECDED, while platforms A, B, and E rely on error protection based on chipkill.



Tabell: ogrupperade data

Man undersökte 35 tändsticksaskar och noterade för varje ask hur många tändstickor den innehöll. Följande värden $(x_i, i = 1, \ldots, 35)$ erhölls:

51	52	49	51	52	51	53
52	48	52	50	53	49	50
51	53	51	52	50	51	53
53	55	50	49	53	50	51
51	52	48	53	50	49	51

Detta är ett exempel på *ogrupperade* data och **VARIATION**.

Hur många tändstickor tillverkas i Sverige ?

Varje dag tillverkar Swedish Match cirka 5 miljoner tändsticksaskar, vilket motsvarar omkring 250 miljoner tändstickor.

Ställer man askarna på varandra, räcker en minuts produktion till en pelare högre än Eiffeltornet i Paris. Lägger man askarna på rad på E4:an, räcker en dags produktion från Jönköping till Stockholm. Stickornas sammanlagda längd skulle räcka till Australien.

KTH Matematik

Frekvenser

En stor ogrupperad datamängd är svår att överskåda. Resultatet i exemplet ovan kan sammanfattas i en *frekvenstabell* av *grupperade* data.

- *absoluta frekvenserna* f_i för de olika förekommande värdena
- de *relativa frekvenserna* $p_i = f_i/n$. (Här avser $i = 1$ lägsta klassen, $i = 2$ nästa klass o.s.v..) $n =$ antalet data (=35).

KTH Matematik

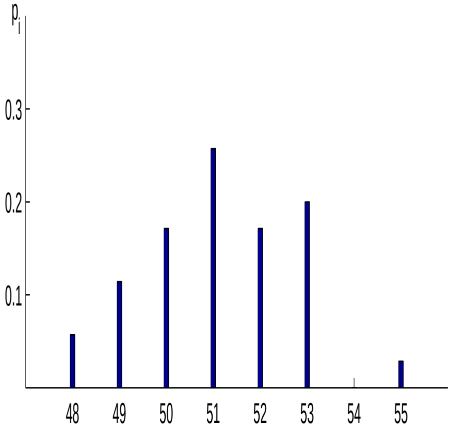
Frekvenstabell: grupperade data

Table : Frekvenstabell för antal tändstickor i tändsticksaskar.

Klass i	Absolut frekvens f_i	Relativ frekvens (%) $100\, p_i$
48	2	5.7
49	4	11.4
50	6	17.1
51	9	25.7
52	6	17.1
53	7	20.0
54	0	0.0
55	1	2.9
S:a	35	100.0

KTH Matematik

Större åskådlighet får man genom ett *stolpdiagram* med de relativa frekvenserna p_i inritade.



Lägesmått (Variation)

Låt allmänt $x_1, \ldots, x_n$ vara de data som skall bearbetas. Som lägesmått används ofta (aritmeriska) *medelvärdet*

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

I exemplet blir $\sum_{j=1}^{35} x_j = 1789$ och $\bar{x} = 1789/35 = 51.1143$.

Spridningsmått (Variation)

Som spridningsmått används ofta *(stickprovs)variansen*

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$$

eller *(stickprovets) standardavvikelse*

$$s = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2},$$

Kvadratrotsutdragningen motiveras av att samma enhet då erhålles som för de givna värdena.

KTH Matematik

I kursens formelsamling hittar man

$$s^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})^2$$

$$= \frac{1}{n-1} \left(\sum_{j=1}^{n} x_j^2 - n\bar{x}^2 \right)$$

I exemplet blir $\sum_{j=1}^{35} x_j^2 = 91533$ och $\bar{x}^2 = 2612.7$ och standardavvikelsen blir $s = \sqrt{\frac{1}{34}(91533 - 35 \cdot 2612.67)} \approx 1.62$.

KTH Matematik

Med *kovariansen* mellan x- och y-värdena i en datamängd $(x_1, y_1), (x_2, y_2) \ldots, (x_n, y_n)$ menas

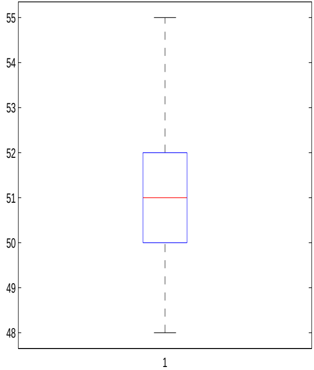
$$c_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

och med *korrelationskoefficienten* menas

$$r = \frac{c_{xy}}{s_x s_y},$$

där s_x och s_y är stickprovsstandardavvikelserna för x- respektive y-data.

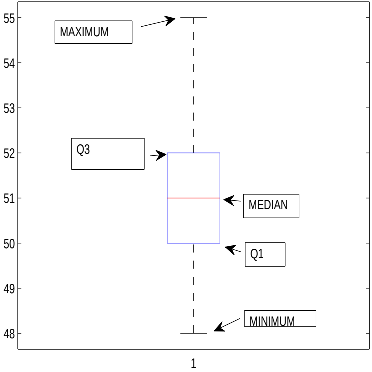
BOXPLOT för tändsticksdata



51	52	49	51	52	51	53
52	48	52	50	53	49	50
51	53	51	52	50	51	53
53	55	50	49	53	50	51
51	52	48	53	50	49	51

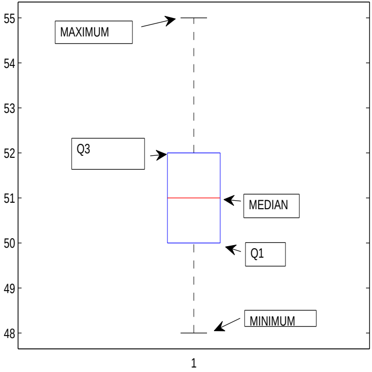
BOXPLOT a.k.a box-and-whisker diagram

A **boxplot** *is a graph of a data set that consists of a line from the minimum value to the maximum value and a box with lines drawn at the first quartile Q_1, the median and the third quartile Q_3*



BOXPLOT

första kvartilen $Q_1 =$ *25% av värdena ligger under, medianen* $=$ *mittpunkten, 50 % ligger under, the tredje kvartilen* $Q_3 =$ *75% ligger under*



New York Times 2009: Big Data

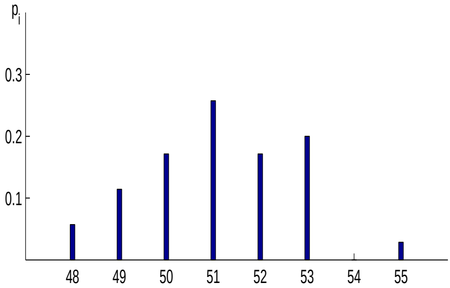
statistics . . . uses powerful computers and sophisticated mathematical models to hunt for meaningful patterns and insights in vast troves of data. The applications are as diverse as improving Internet search and online advertising, culling gene sequencing information for cancer research and analyzing sensor and location data to optimize the handling of food shipments.

mathematical model = sannolikhetsmodell (slumpmodell)

KTH Matematik

Sannolikheten för händelsen "$50 \leq$ antalet tändstickor i en ask ≤ 53"

$$p_{50} + p_{51} + p_{52} + p_{53} = 0.171 + 0.257 + 0.171 + 0.20 \approx 0.80$$



Vi skriver **sannolikheten för händelsen** "$50 \leq$ antalet tändstickor i en ask ≤ 53"

$$P\left(50 \leq \text{antalet tändstickor i en ask} \leq 53\right) = 0.80$$

Detta är en sannolikhet som baserar sig på observationer. On inget ändras i produktionsprocessen förväntar vi oss att om vi får en ny tänsticksask och checkar antalet tändstickor i den, så har vi 80% chans att händelsen "$50 \leq$ antalet tändstickor i en ask ≤ 53" inträffar.

Sannolikhet: empirisk

Sannolikheten

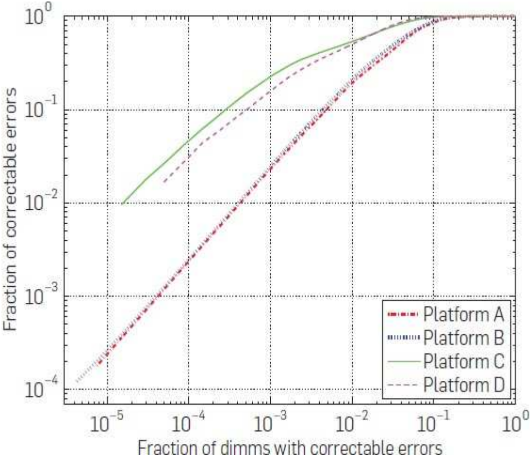
$$Pr\left(50 \leq \text{antalet tändstickor i en ask} \leq 53\ \right) = 0.80$$

baserade sig på 35 tändsticksaskar. De relativa frekvenserna kommer att ändras vi kollar nya tändsticksaskar, men de kommer att stabilisera sig.

KTH Matematik

Googles data säger att en given minnesmodul har approximativt 8%s sannolikhet för att bli drabbad av ett fel under ett givet år.

All platforms the top 20% of DIMMs with errors make up over 94% of all observed errors. The shape of the distribution curve provides evidence that it follows a *power-law distribution* (?). Intuitively, the skew in the distribution means that a DIMM that has seen a large number of errors is likely to see more errors in the future.



Sannolikhet

Vi postulerar att det finns ett tal som kallas sannolikheten för en händelse. Därmed menas helt enkelt att *man tilldelar varje händelse ett visst tal*. Om händelsen är A, betecknas talet med $P(A)$, *sannolikheten för* A. (P som i Probability.)

Beträffande talet $P(A)$ gäller allmänt, att man söker välja det så att den relativa frekvensen vid ett någorlunda stort antal försök kommer i närheten av $P(A)$. Om vi säger att $P(A) = 0.80$ kan vi ge detta uttalande den påtagliga men samtidigt vaga *frekvenstolkningen*: Vid ett stort antal försök blir den relativa frekvensen av händelsen A nog ungefär lika med 0.80.

KTH Matematik

Slumpförsök

Vi betraktar nu slumpförsök medelst allmänna beteckningar

- Varje möjligt resultat ω av ett slumpförsök kallas ett *utfall*, eller ett elementärt utfall.

KTH Matematik

Slumpförsök

Vi betraktar nu slumpförsök medelst allmänna beteckningar

- Varje möjligt resultat ω av ett slumpförsök kallas ett *utfall*, eller ett elementärt utfall.
- Mängden av alla *utfall*, eller resultat, kallar vi utfallsrummet och betecknar det med Ω.

KTH Matematik

Slumpförsök

Vi betraktar nu slumpförsök medelst allmänna beteckningar

- Varje möjligt resultat ω av ett slumpförsök kallas ett *utfall*, eller ett elementärt utfall.
- Mängden av alla *utfall*, eller resultat, kallar vi utfallsrummet och betecknar det med Ω.
- En händelse A är en mängd av utfall, dvs en delmängd av Ω, $A \subset \Omega$.

KTH Matematik

Ofta kan valet av utfallsrum bero på situationen eller den fråga vi vill studera.

Exempel: $\Omega =$ *de fem miljoner tändsticksaskarna producerade under en given dag.*
$\Omega =$ *de 35 miljoner tändsticksaskarna producerade under en sjudagars period.*

KTH Matematik

Exempel: tärningskast

Slumpförsok: kast av en tärning. Utfall: ω = antalet ögon

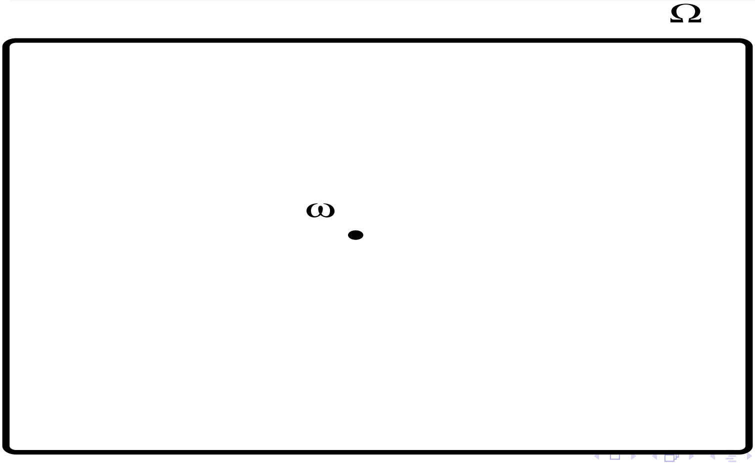
$$\Omega = \{\text{etta, tvåa, trea, fyra, femma, sexa}\}$$

Exempel på en händelse A = udda antal ögon = {etta, trea, femma }. Vi säger att *händelsen A inträffar*, om vi kastar en tärning (=genomför ett slumpförsok) och får etta eller trea eller femma. (Slut på exemplet.)

Venndiagram

Definition

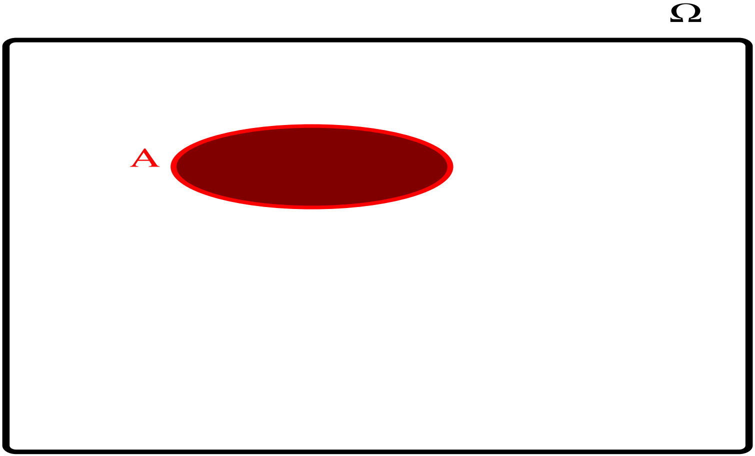
Mängden av alla utfall, eller resultat, kallar vi utfallsrummet och betecknar det med Ω.



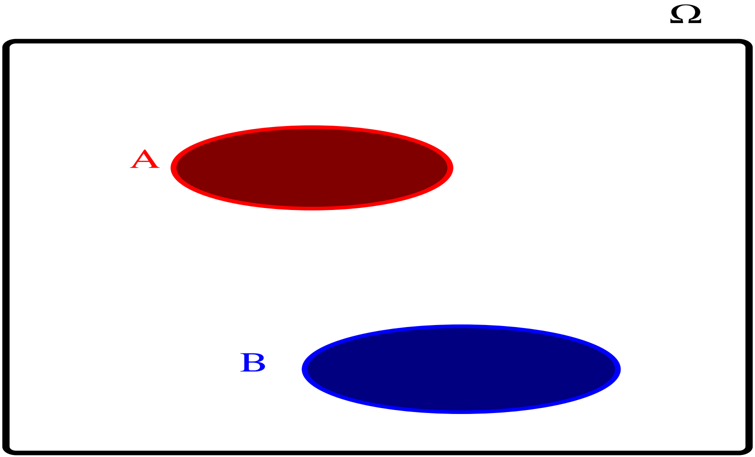
Venndiagram

Definition

En händelse A är en mängd av utfall, dvs en delmängd av Ω, $A \subset \Omega$.



Venndiagram; två händelser



Händelser $A \cap B$

Låt oss nu anta att vi är intresserade av två händelser A och B definierade på samma försök. Här är några exempel på vad som kan inträffa, och hur vi matematiskt kan uttrycka detta:

$$\text{"}A \text{ inträffar"}, \quad A$$

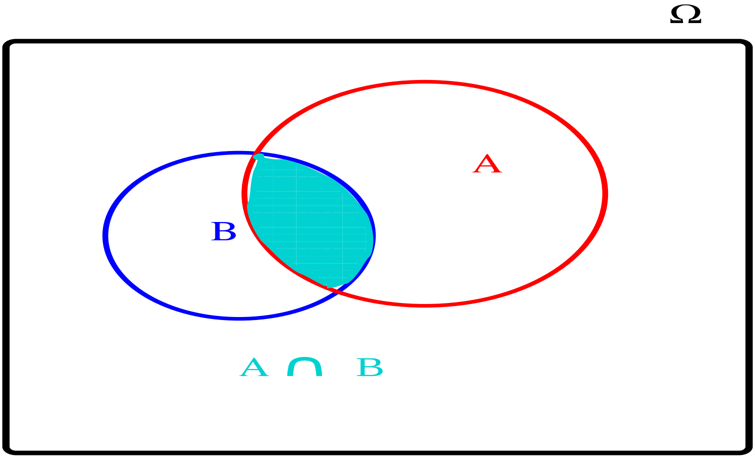
$$\text{"}A \text{ och } B \text{ inträffar"} \quad \text{eller} \quad \text{"}A \text{ snitt } B \text{ inträffar"}, \quad A \cap B$$

Exempel

$$\Omega = \{etta,\ tvåa,\ trea,\ fyra,\ femma,\ sexa\ \}$$

$A = udda\ antal\ ögon = \{etta,\ trea,\ femma\ \}$. $B = \{femma,\ sexa\ \}$, $A \cap B = \{femma\ \}$.

Venndiagram $A \cap B$



$$\text{"}A \text{ eller } B \text{ inträffar"} \quad \text{eller} \quad \text{"}A \text{ union } B \text{ inträffar"}, \quad A \cup B$$

Obs! $A \cup B$ betyder att minst en av A eller B inträffar, så $A \cap B$ kan mycket väl inträffa. I matematik betyder "eller" och/eller!

Händelser $A \cup B$

$A \cup B$ betyder att minst en av A eller B inträffar

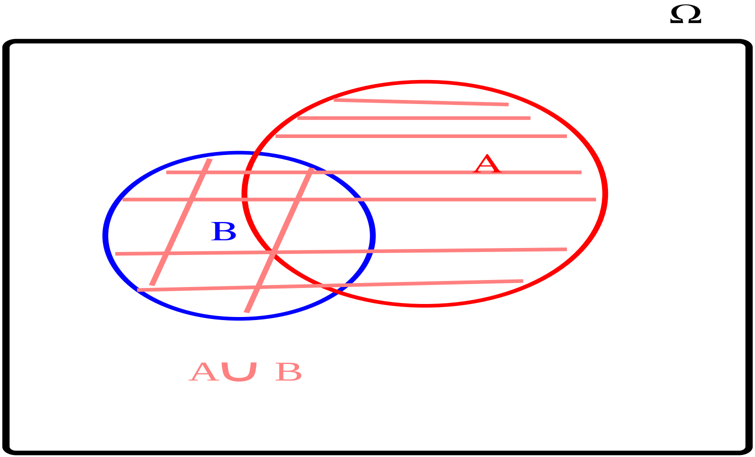
Exempel

$$\Omega = \{etta,\ tvåa,\ trea,\ fyra,\ femma,\ sexa\ \}$$

$A = udda\ antal\ ögon = \{etta,\ trea,\ femma\ \}$. $B = \{femma,\ sexa\ \}$,
$A \cup B = \{etta,\ trea,\ femma,\ sexa\ \}$.

KTH Matematik

Venndiagram $A \cup B$



Händelser A^*

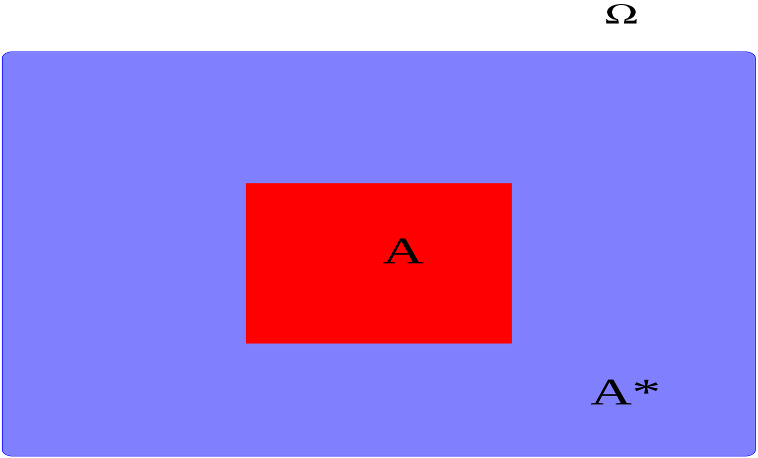
$$\text{"}A \text{ inträffar inte"}, \quad A^*.$$

Exempel

$$\Omega = \{\textit{etta, tvåa, trea, fyra, femma, sexa}\}$$

$A = \textit{udda antal ögon} = \{\textit{etta, trea, femma}\}$.
$A^* = \{\textit{tvåa, fyra, sexa}\} = \textit{jämnt antal ögon}$.

Venndiagram A^*



"tomma mängden" $\emptyset$

Om A och B utesluter varandra, dvs. omöjligt kan inträffa samtidigt, så säger vi att A och B är *disjunkta* eller oförenliga, dvs.

$$A \cap B = \emptyset$$

där $\emptyset$ är "tomma mängden" eller "den omöjliga händelsen".

$$\Omega^* = \emptyset$$

KTH Matematik

"tomma mängden" $\emptyset$

A och B utesluter varandra, dvs. omöjligt kan inträffa samtidigt.

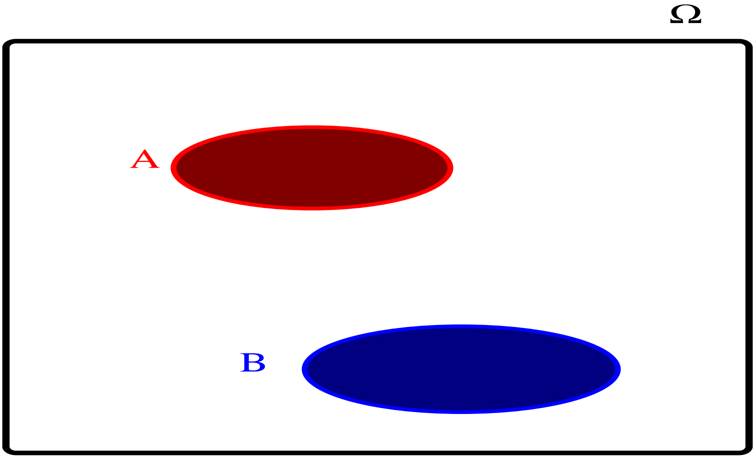
Exempel

$$\Omega = \{etta,\ tvåa,\ trea,\ fyra,\ femma,\ sexa\ \}$$

$A = \{etta,\ trea,\ femma\ \}$. $B = \{\ fyra,\ sexa\ \}$, $A \cap B = \emptyset$.

KTH Matematik

Venndiagram; $A \cap B = \emptyset$



De Morgans regler

Sats

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Sats

$$(A \cap B)^* = A^* \cup B^*$$

KTH VETENSKAP OCH KONST

KTH Matematik

De Morgans regler

$$(A \cap B)^* = A^* \cup B^*$$

Exempel

$$\Omega = \{etta,\ tvåa,\ trea,\ fyra,\ femma,\ sexa\ \}$$

$$A = \{etta,\ trea,\ femma\ \}.\ B = \{femma,\ sexa\ \},\ A \cap B = \{\ femma\ \}$$

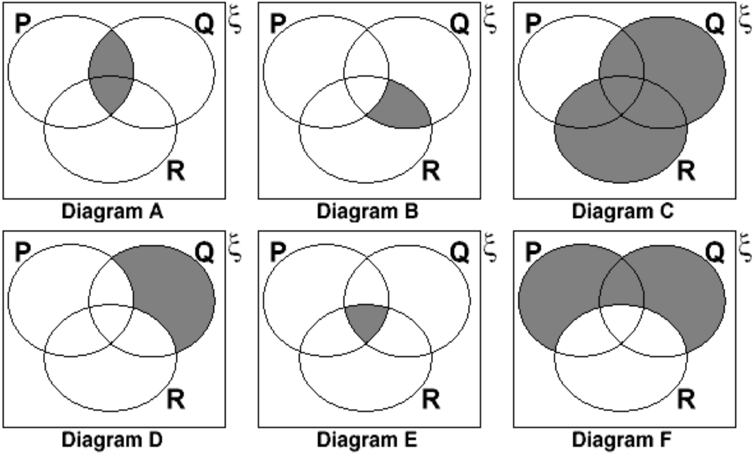
$$(A \cap B)^* = \{etta,\ tvåa,\ trea,\ fyra,\ sexa\ \}$$

$$A^* \cup B^* = \{tvåa,\ fyra,\ sexa\ \} \cup \{etta,\ tvåa,\ trea,\ fyra\ \}$$

$$= \{etta,\ tvåa,\ trea,\ fyra,\ sexa\ \}$$

KTH Matematik

Drill på Venndiagram



Drill på Venndiagram

a) Vilket diagram svarar mot $Q \cup R$?

b) Vilket diagram svarar mot $P \cap Q \cap R$?

c) Vilket diagram svarar mot $Q \cap R \cap P^*$?

d) Vilket diagram svarar mot $P \cap Q$?

e) Vilket diagram svarar mot $P^* \cap R^* \cap Q$?

f) Vilket diagram svarar mot $(P \cup Q) \cap R^*$?

g) Vilket diagram svarar mot $Q \cap (P \cup R)^*$?

Svaren på nästa slajd:

a) Diagram C
b) Diagram E
c) Diagram B
d) Diagram A
e) Diagram D
f) Diagram F
g) Diagram D

Varför har e) och g) samma svar?

Händelser

Har vi många händelser kan vi, precis som med summa- och produkt-tecken, använda ett förkortat skrivsätt:

$$\bigcup_1^n A_i = A_1 \cup A_2 \cup \ldots \cup A_n \quad \text{och} \quad \bigcap_1^n A_i = A_1 \cap A_2 \cap \ldots \cap A_n$$

Slumpförsök: tärningskast och $P(A)$

Låt oss säga att vi kastar en tärning, och är intresserade av händelsen

$$\{\text{vi får en sexa}\}.$$

Alla håller nog med om att, om det är en just tärning, att den sannolikheten är $\frac{1}{6}$. Symboliskt kan vi skriva

$$A = \{\text{vi får en sexa}\} \quad \text{och} \quad P(A) = \frac{1}{6}.$$

Är det överhuvudtaget meningsfullt att tala om sannolikheter, och om så är fallet, hur skall man tolka dessa?

KTH Matematik

Slumpförsök: relativa frekvenser

Vi skall tolka detta som att om man kastar tärningen många gånger, så blir den relativa frekvensen 6or ungefär $\frac{1}{6}$. Allmänt sett, om vi har ett försök och en händelse A och gör försöket n gånger, så gäller

$$f_n(A) = \frac{\text{antalet gånger } A \text{ inträffar}}{n} \to P(A) \text{ då } n \text{ växer.}$$

Kolmogorovs axiomsystem (1933):

Ett *sannolikhetsmått* P är en funktion av händelser, sådan att:

(a) $0 \leq P(A) \leq 1$;

Kolmogorovs axiomsystem (1933):

Ett *sannolikhetsmått* P är en funktion av händelser, sådan att:

(a) $0 \leq P(A) \leq 1$;

(b) $P(\Omega) = 1$

KTH Matematik

Kolmogorovs axiomsystem (1933):

Ett *sannolikhetsmått* P är en funktion av händelser, sådan att:

(a) $0 \leq P(A) \leq 1$;

(b) $P(\Omega) = 1$

(c) om $A_1, A_2, \ldots$ är disjunkta händelser, så gäller

$$P\left(\bigcup_1^\infty A_i\right) = \sum_1^\infty P(A_i).$$

KTH Matematik

Kolmogorovs axiomsystem (1933):

Ett *sannolikhetsmått* P är en funktion av händelser, sådan att:

(a) $0 \leq P(A) \leq 1$;

(b) $P(\Omega) = 1$

(c) om $A_1, A_2, \ldots$ är disjunkta händelser, så gäller

$$P\left(\bigcup_1^\infty A_i\right) = \sum_1^\infty P(A_i).$$

- (a) och (b) kan ses som en kalibrering så att P stämmer med intuitionen (det blir lättare då) och (c) (som är det "viktiga" axiomet) betyder att P är ett mått.

KTH Matematik

Kolmogorovs axiomsystem: ett specialfall

Om $A_1, A_2, \ldots$ är disjunkta händelser, så gäller

$$P\left(\bigcup_{1}^{\infty} A_i\right) = \sum_{1}^{\infty} P(A_i).$$

Ett specialfall: Betrakta A och B sådana att $A \cap B = \emptyset$, då fås

$$P(A \cup B) = P(A) + P(B)$$

KTH Matematik

Regler för sannolikhetskalkyl (1)

Sats

$P(A^*) = 1 - P(A)$.

Bevis. Vi ska ge ett mycket formellt bevis, för att illustrera axiomsystemet: Eftersom A och A^* disjunkta och $A \cup A^* = \Omega$, så fås enligt (c) och (b) ovan

$$P(A) + P(A^*) = P(\Omega) = 1 \quad \Rightarrow \quad P(A^*) = 1 - P(A).$$

$\square$

Följdsats

$P(\emptyset) = 0$.

Regler för sannolikhetskalkyl (2)

Sats

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Bevis. Satsen följer med hjälp av Venn-diagram, och observationen att $P(A) + P(B)$ "mäter" $A \cap B$ två gånger. $\square$

KTH Matematik

Regler för sannolikhetskalkyl (3)

Om $A \cap B = \emptyset$, så fås $P(A \cap B) = P(\emptyset) = 0$, dvs.

$$P(A \cup B) = P(A) + P(B).$$

Detta följer av det ovan visade

Sats

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

KTH Matematik

Den klassiska sannolikhetsdefinitionen

Antag att Ω består av m (möjliga) elementarutfall $\omega_1, \ldots, \omega_m$, var och en med samma sannolikhet att inträffa, dvs

$$P(\omega_k) = \frac{1}{m} \qquad k = 1, \ldots, m.$$

Betrakta en händelse A, $A \subset \Omega$. Antag att A innehåller g (gynnsamma) elementarutfall. Då gäller

$$P(A) = \frac{g}{m}.$$

KTH Matematik

Problemet med den klassiska sannolikhetsdefinitionen, i mera komplicerade situationer, är att hitta en uppdelning av Ω i lika sannolika elementarutfall och att beräkna m och g. I många – de flesta – situationer är det inte alls möjligt att göra detta.

Kombinatoriska grundbegrepp

För att beräkna *m* och *g* behöver vi några kombinatoriska grundbegrepp. Vi inleder med en s.k. *multiplikationsprincip*.

Sats

Om en åtgärd kan utföras på n_1 och en annan (nästa) åtgärd kan utföras på n_2 sätt, så finns det $n_1 \cdot n_2$ sätt att utföra de båda åtgärderna.

KTH Matematik

Kombinatoriska grundbegrepp

Enligt multiplikationsprincipen finns det n^k olika sätt att plocka ut k st. av n st. föremål om varje föremål som har plockats ut stoppas tillbaka och om vi tar hänsyn till i vilken ordning de plockas ut.

Exempel

Antalet PIN-koder $= 10^4 = 10000$.

Enligt multiplikationsprincipen finns det n^k olika sätt att plocka ut k st. av n st. föremål om varje föremål som har plockats ut stoppas tillbaka och om vi tar hänsyn till i vilken ordning de plockas ut.

Exempel

Föremålen är $\heartsuit, \diamondsuit, \spadesuit$, *dvs.* $n = 3$. *Vi tar* $k = 2$. $\heartsuit\heartsuit$, $\heartsuit\diamondsuit$, $\heartsuit\spadesuit$, $\diamondsuit\diamondsuit$, $\diamondsuit\heartsuit$, $\diamondsuit\spadesuit$, $\spadesuit\spadesuit$, $\spadesuit\heartsuit$, $\spadesuit\diamondsuit$.

Kombinatoriska grundbegrepp: permutation

n st. föremål kan permuteras eller ordnas på

$$n! = n \cdot (n-1) \ldots 2 \cdot 1$$

olika sätt.

$$0! = 1$$

Kombinatoriska grundbegrepp: permutation

n st. föremål kan permuteras eller ordnas på

$$n! = n \cdot (n-1) \ldots 2 \cdot 1 \quad \text{utläses: n fakultet}$$

olika sätt.

$$0! = 1$$

Exempel

$n = 3$, $3! = 6$. $\heartsuit\diamondsuit\spadesuit$, $\heartsuit\spadesuit\diamondsuit$, $\diamondsuit\heartsuit\spadesuit$, $\diamondsuit\spadesuit\heartsuit$, $\spadesuit\heartsuit\diamondsuit$, $\spadesuit\diamondsuit\heartsuit$.

KTH Matematik

Kombinatoriska grundbegrepp: permutation

$$n \cdot (n-1) \ldots (n-k+1)$$

kallas kallas antalet permutationer av k element bland n (=antalet sätt at välja k element bland n utan återläggning och med hänsyn till ordningen). Vi har

$$n \cdot (n-1) \ldots (n-k+1) = \frac{n!}{(n-k)!}$$

KTH Matematik

Kombinatoriska grundbegrepp: permutation

$$n \cdot (n-1) \ldots (n-k+1)$$

kallas kallas antalet permutationer av k element bland n (=antalet sätt at välja k element bland n utan återläggning och med hänsyn till ordningen). Vi har

$$n \cdot (n-1) \ldots (n-k+1) = \frac{n!}{(n-k)!}$$

Exempel

$n = 3,\ k = 2.$ $\heartsuit\diamondsuit$, $\diamondsuit\heartsuit$, $\diamondsuit\spadesuit$, $\spadesuit\diamondsuit$, $\heartsuit\spadesuit$, $\spadesuit\heartsuit$.

Låt x vara antalet sätt att att plocka ut k st. av n st. föremål om vi ej tar hänsyn till i vilken ordning de plockas ut. Då gäller enligt multiplikationsprincipen att

$$\underbrace{n \cdot (n-1) \ldots (n-k+1)}_{k \text{ utan återläggning och med ordning}} = \underbrace{k!}_{\text{antalet sätt att ordna } k \text{ element}} \cdot x$$

$$\Rightarrow$$

$$x = \frac{n \cdot (n-1) \ldots (n-k+1)}{k!} = \frac{n!}{(n-k)!} \frac{1}{k!}$$

KTH Matematik

Kombinatoriska grundbegrepp: binomialkoefficienterna

Vi ger beteckningen $x = \binom{n}{k}$. Alltså: Det finns

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

olika sätt att plocka ut k st. av n om vi ej tar hänsyn till i vilken ordning de plockas ut.

Kombinatoriska grundbegrepp: binomialkoefficienterna

Vi ger beteckningen $x = \binom{n}{k}$. Alltså: Det finns

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

olika sätt att plocka ut k st. av n om vi ej tar hänsyn till i vilken ordning de plockas ut.

Exempel

$n = 3,\ k = 2.\ \heartsuit\diamondsuit,\ \heartsuit\spadesuit,\ \spadesuit\diamondsuit\ .$

Två urnmodeller: Dragning utan återläggning

I en urna finns kulor av två slag: v vita och s svarta. Drag n kulor ur urnan slumpmässigt och så att en kula som dragits inte stoppas tillbaka. dvs dragning utan återläggning.
Sätt $A =$ "Man får k vita kulor i urvalet".

Två urnmodeller: Dragning utan återläggning

Välj Ω: Alla uppsättningar om n kulor utan hänsyn till ordning.
Då fås:

$$m = \binom{v+s}{n} \quad \text{och} \quad g = \binom{v}{k}\binom{s}{n-k}$$

och således

$$P(A) = \frac{\binom{v}{k}\binom{s}{n-k}}{\binom{v+s}{n}}.$$

KTH Matematik

Två urnmodeller: Dragning med återläggning

Samma modell som i fallet med dragning utan återläggning, men kulorna stoppas tillbaka igen efter det att man observerat dess färg, och urnan skakas om för nästa dragning.
Välj Ω: Alla uppsättningar om n kulor med hänsyn till ordning:

$$m = (v + s)^n.$$

KTH Matematik

Två urnmodeller: Dragning med återläggning

Antag att vi valt ut k vita och $n-k$ svarta kulor. Dessa kan placeras på $\binom{n}{k}$ platser:

v	v		v	$\cdots\cdots$	v

Antal sätt att välja ut k vita $= v^k$. Antal sätt att välja ut $n-k$ svarta $= s^{n-k}$.

Två urnmodeller: Dragning med återläggning

Detta ger $g = \binom{n}{k} v^k s^{n-k}$ och således får vi

$$P(A) = \frac{\binom{n}{k} v^k s^{n-k}}{(v+s)^n} = \binom{n}{k} \left(\frac{v}{v+s}\right)^k \left(\frac{s}{v+s}\right)^{n-k}.$$

Kolmogorov[2]

[2] A.N.Kolmogorov, 1903 - 1987