UPPSALA UNIVERSITET

Föreläsningsanteckningar

Sannolikhetsteori 1

Rami Abou Zahra

Inlämningsdatum
September 21, 2022

Contents

1. Repetition - (K2.1)

1.1. Mängdlära.

Tips för hela kursen! Rita venndiagram

1.2. Begrepp.

- Om A och B är disjunkta säger vi att de är **oförenliga**, dvs $A \cap B = \emptyset$
- A, B och C är disjunkta om $A \cap B = \emptyset$ och $A \cap C = \emptyset$ och $B \cap C = \emptyset$
- $\lambda \subseteq 2^{\Omega}$ är disjunkta om $A \cap B = \emptyset$ för alla $A, B \in \lambda$
- Sannolikhetsrum $= (\Omega, P)$
- $x \in \Omega$: x är ett element/utfall i Ω
- $A \subseteq \Omega$:A är en delmängd/händelse till Ω
- $2^{\Omega} = \{A : A \subseteq \Omega\}$, kallas även för potensmängden
- Ω är vår grundmängd/utfallsrum

Definition/Sats 1.1: Utfall, händelser, utfallsrum

Resultatet av ett slumpförsök kallas ett *utfall.* Mängden av möjliga utfall från ett visst slumpförsök kallas *utfallsrum.* En viss specifierad mängd utfall kallas för en *händelse.* Från detta följer det att ett utfall även är en händelse, precis som hela utfallsrummet också är en händelse

Det är viktigt här att notera att vi arbetar med mängder, och element i mängder behöver nödvändigtvis inte vara tal.

Det som är även viktigare att inse är att i den klassiska definitionen man kanske sett på högstadiet/gymnasiet så hade alla utfall "samma vikt", det vill säga om vi har totalt 2 möjliga utfall så har varje utfall en sannolikhet på $\frac{1}{2}$ att inträffa. Det som är fiffigt med denna definition är att det bilr enkelt att inse att sannolikheten för att hela utfallsrummet skall ske är 1 (vilket är något vi bevarar i vår definition), men verkligheten är lite annorlunda. Det kanske är så att sannolikheten för ett utfall faktiskt är $\frac{1}{6}$ men det andra är $\frac{5}{6}$.
Vi definierar följande:

Definition/Sats 1.2: Likformig sannolikhetsfördelning

Ett slumpexperiment med ändligt utfallsrum sägs ha *Likformig sannolikhetsfördelning* om alla utfall har samma sannolikhet.

Det vi noterar från definitionen ovan är att om vi har n:st utfall så kommer sannolikheten för varje utfall att vara $\frac{1}{n}$. Detta kallade vi för "klassiska" sannolikheten eftersom det är den man kanske klassiskt stött på, men faktum är att vi faktiskt definierar klassisk sannolikhet som just det:

Definition/Sats 1.3: Klassiska sannolikhetsdefinitionen

För ett slumpexperiment med ändligt utfallsrum och med likformig sannolikhetsfördelning gäller att sannolikheten för en händelse är lika med antalet utfall i händelsen dividerat med antalet utfall i utfallsrummet, dvs antalet gynnsamma utfall dividerat med antalet möjliga utfall.
Om händelsen A innehåller $n(A)$ utfall och utfallsrummet har $n(\Omega)$ utfall gäller alltså att:

$$P(A) = \frac{n(A)}{n(\Omega)}$$

2. Regler för sannolikheter - (K2.2)

2.1. **Kolmogorovs Axiom.**

Ett **sannolikhetsmått** är en funktion $P : 2^\Omega \to \mathbb{R}$ som uppfyller:

- $P(A) \geq 0 \quad \forall A \in 2^\Omega$
- $P(\Omega) = 1$
- $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=0}^{\infty} P(A_i)$ om A_i är parvis disjunkta

Exempel:

Singla slant är det klassiska exemplet, där har vi 2 möjliga utfall (krona eller klave).
Utfallsrummet Ω är mängden $\{krona, klave\}$
Ett rimligt antagande är att sannolikheten att landa på krona är $\frac{1}{2}$ och samma för klave, dvs $P(\{krona\}) = \frac{1}{2}$ och $P(\{klave\}) = \frac{1}{2}$

$P(\Omega) = 1$, $P(\emptyset) = 0$

Exempel:

Singla slant 2 gånger
Utfallsrummet bör rimligtvis vara kopplad till föregående exempel:

$\Omega = \{kr, kl\}\text{x}\{kr, kl\} = \{(kr, kr), (kr, kl), (kl, kr), (kl, kl)\}$

$P(\{x\}) = \frac{1}{4}$, $P(\text{minst en krona}) = P\left(\{(kr, kr), (kr, kl), (kl, kr)\}\right) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$

Exempel:

Singla slant n gånger: $\Omega = \{kr, kl\}^n$
$P(\{x\}) = \frac{1}{2^n} \quad \forall x \in \Omega, \quad P(A) = \sum_{x \in A} \frac{1}{2^n}$
$P(\text{exakt } k\text{st krona}) = \sum_{xx \text{ innehåller } k \text{ kronor}} \frac{1}{2^n} = \binom{n}{k}\left(\frac{1}{2^n}\right)$

Exempel:

Tärningskast är återigen ett klassiskt exempel, då är $\Omega = \{1, 2, 3, 4, 5, 6\}$

Är det en normal tärning så är sannolikheten för varje kast $\frac{1}{6}$, $P(\{x\}) = \frac{1}{6}$

Antag att jag har en riggad tärning sådant att ettan är ombytt till en sexa. Då kommer följande gälla: $P(\{1\}) = 0$ och $P(\{6\}) = \dfrac{1}{3}$

Sannolikheter ska man tänka som proportioner, som associerar en vikt till varje delmängd

Exempel:

Låt $\Omega = \mathbb{N}_+$, $P(\{n\}) = \dfrac{1}{2^n}$

Eftersom $\sum_{n=1}^{\infty} \dfrac{1}{2^n} = 1$ gäller det att $P(\Omega) = 1$

Kopplar vi detta exempel till verkligheten så kan detta vara ”hur stor är sannolikheten att slinga krona n gånger” eller ”sannolikheten att slinga krona för första gången på n:te slinglen”

Exempel:

Vad är sannolikheten att tärningen hamnar på en sexa på n:te slinglen?
Jo, $P(\{x\}) = \underbrace{\left(\dfrac{5}{6}\right)^{n-1}}_{\text{alla andra siffror}} \cdot \dfrac{1}{6}$

Exempel:

Slumpa ett reellt tal mellan 0 och 1:
$\Omega = (0,1) \subseteq \mathbb{R}$, då är $P(A) =$ längden på intervallet $A = 1$

Notera att det inte spelar roll om det är ett öppet eller slutet intervall

Vill man räkna ut unionen av sannolikheten summerar man sannolikheterna:

$$P\left(\left[\frac{1}{2}, \frac{1}{4}\right] \cup \left(\frac{3}{4}, \frac{7}{8}\right)\right) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$$

Vad är då sannolikheten att vi slumpar ett rationellt tal mellan $(0,1)$? Vi får inte glömma att $\mathbb{Q}$ är uppräknelig:

$$P(\mathbb{Q} \cap (0,1)) = P\left(\bigcup_{q \in \mathbb{Q} \cap (0,1)} \{q\}\right) = \sum_{q \in \mathbb{Q} \cap (0,1)} P(\{q\}) = 0$$

Hur ser $P(\text{irrationellt tal})$ ut?

$$P(\mathbb{Q}^c \cap (0,1))$$

$$\underbrace{(\mathbb{Q} \cap (0,1)) \cup (\mathbb{Q}^c \cap (0,1))}_{\text{disjunkta}} = \Omega$$

$$1 = P(\Omega) = \underbrace{P(\mathbb{Q} \cap (0,1))}_{=0} + P(\mathbb{Q}^c \cap (0,1)) \Rightarrow P(\text{irrationellt tal}) = 1 - 0 = 1$$

Exempel:

Ta en Riemann-integrerbar funktion $f : [0,1] \to \mathbb{R}$ så $\int_0^1 f(x)dx = 1$.

Vi sätter $P(A) = \int_A f(x)dx$

Exempel:

Tag enhetskvadraten $\Omega = [0,1]^2$, $P(A) =$ arean. Slumpa ett tal i kvadraten

Definition/Sats 2.1: Diskreta Sannolikhetsrum

Sannolikhetsrummet (Ω, P) kallas för **diskret** om det finns en uppräknelig delmängd $A \subseteq \Omega$ så att:

$$P(B) = \sum_{x \in B \cap A} P(\{x\})$$

Alternativ beskrivning:

$$\exists A \subseteq \Omega :$$

$$\sum_{x \in A} P(\{x\}) = 1$$

Definition/Sats 2.2: Kontinuerliga Sannolikhetsrum

Icke-diskreta sannoliketsrum (förutom blandade osv, men vi kommer inte arbeta med dessa ändå)

2.2. A^c.

Med komplementet menar vi $x \in A^c \Leftrightarrow x \in A$ där $(x \in \Omega, \quad A \subseteq \Omega)$

$P(A \cup A^c) = P(\Omega) = 1 = P(A) + P(A^c) \Rightarrow P(A^c) = 1 - P(A)$

2.3. B-A.

$$\begin{gathered} x \in B \setminus A \Rightarrow x \in B \wedge x \notin A \\ \Rightarrow x \in B \wedge x \in A^c \\ x \in B \cap A^c \\ P(B) = P(A \cap B) + P(B \setminus A) \Rightarrow P(B \setminus A) = P(B) - P(A \cap B) \\ \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B) \end{gathered}$$

3. Tolkning av sannolikheter

Om vi tar exemplet att singla slant. Vad betyder det att sannolikheten är $\frac{1}{2}$?
Man kan tolka det som att "det finns 2 fall, och båda har lika stor chans att inträffa"
Eller en mer data-inriktad tolkning, det vill säga om man singlar slant 100ggr, kommer ungefär hälften av kasten resultera i krona eller klave.

Det finns däremot tolkningar via Kolmogorovs axiom, det vill säga:

- $P(A) = p$ betyder att A utgör p enheter av utfallsrummet Ω
- Om vi upprepat slumpar ett $x \in \Omega$ så kommer tillslut $x \in A$ inträffa med frekvens p (**stora talens lag**)

3.1. Sannolikhetsmåttet P.

Uppfyller följande:

- $P(A^c) = 1 - P(A)$
- $P(\emptyset) = P(\Omega^c) = 1 - P(\Omega) = 1 - 1 = 0$
- A, B disjunkta gäller $P(A \cup B) = P(A \cup B \cup \emptyset \cdots)$ (ty axiomet säger att vi skall ha oändliga disjunkta par, vi kan därför fylla ut med oändligt många tomma mängder) $\Rightarrow P(A) + P(B)$
- $P(B \setminus A) = P(B) - P(A \cap B)$
- Om $A \subseteq B$ så gäller $A \cap B = A$ och $P(B \setminus A) = P(B) - P(A)$
- Om $P(B \setminus A) \geq 0$ så $A \subseteq B \Rightarrow P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - \underbrace{P(A \cap B)}_{\geq 0}$
- $P(A \cup B) \leq P(A) + P(B)$ (Booles olikhet)

Definition/Sats 3.1

Om $A_1 \subseteq A_2 \subseteq \cdots \subseteq \Omega$ så gäller

$$P(\bigcup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} P(A_n)$$

Kallas även för att sannolikhetsmåttet är kontinuerligt ovanifrån

Bevis 3.1: Bevis av föregående sats

$$\underbrace{A_1}_{B_1}, \underbrace{A_2 \setminus A_1}_{B_2}, \underbrace{A_3 \setminus A_2}_{B_3} \cdots \underbrace{A_{n+1} \setminus A_n}_{B_{n+1}}$$

B_i är disjunkta, och följande gäller:

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty}(P(B_1) + P(B_2) + P(B_3) + \cdots + P(B_n))$$

$$\Leftrightarrow \lim_{n\to\infty}(P(A_1) + (P(A_2) - P(A_1)) + (P(A_3) - P(A_2)) + \cdots + (P(A_n) - P(A_{n-1})))$$

$$\Leftrightarrow \lim_{n\to\infty} P(A_n)$$

□

Definition/Sats 3.2

Låt $A_3 \subseteq A_2 \subseteq A_1 \cdots \subseteq \Omega$:

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} P(A_n)$$

Lemma 3.1: De morgans lagar

- $\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c$
- $\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c$

Bevis 3.2: Bevis av Lemma

$$\begin{aligned} x \in \left(\bigcup_{i=1}^{\infty} A_i\right)^c &\Leftrightarrow x \notin \bigcup_{i=1}^{\infty} A_i \\ &\Leftrightarrow x \notin A_i \quad \forall i \\ &\Leftrightarrow x \in A_i^c \quad \forall i \\ &\Leftrightarrow x \in \bigcap_{i=1}^{\infty} A_i^c \end{aligned}$$

$\square$

Bevis 3.3: Bebis av sats

Vi har $A_1^c \subseteq A_2^c \subseteq A_3^c \subseteq \cdots$:

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = 1 - P\left(\left(\bigcap_{i=1}^{\infty} A_i\right)^c\right) = 1 - P\left(\bigcup_{i=1}^{\infty} A_i^c\right)$$
$$\Rightarrow 1 - \lim_{n\to\infty} P(A_i^c) = \lim_{n\to\infty} (1 - P(A_i^c)) = \lim_{n\to\infty} P(A_n)$$

$\square$

4. Betingade sannolikheten $P(A|B)$

Definition/Sats 4.1: Betingade sannolikheten

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \text{Sannolikheten för } A \text{ givet } B \text{ förutsatt att } P(B) > 0 \text{ och } P(A) > 0$$

Detta är sannolikheten att $x \in A$ givet att $x \in B$

Exempel:

Låt $\Omega = \{1, 2, 3, 4, \cdots\}$, $P(\{n\}) = \dfrac{1}{2^n}$

Detta sade vi kunde representera antalet slantsinglingar som krävs för att landa på krona (eller klave)
Säg nu att vi sätter det här B = första försöket landar på klave $= \{1\}^c = \{2, 3, 4, 5, \cdots\}$

Vi förväntar oss att $P(1|B) = 0$ (B gäller, alltså att vi har fått klave på första försöket, men då gäller det att det inte finns någon chans att vi får krona på första försöket)

Med motiveringen över gäller $P(2|B) = \dfrac{1}{2}$ och följande:

$$P(n|B) = \frac{1}{2^{n-1}} = \frac{P(\{n\} \cap B)}{P(B)} = \frac{P(\{n\})}{1/2} = 2P(n) = 2 \cdot \frac{1}{2^n} = \frac{1}{2^{n-1}}$$

Vi kan definiera ett sannolikhetsmått $Q : 2^B \to \mathbb{R}$ (för något $B \in \Omega$) och $Q(A) = \dfrac{P(A)}{P(B)}$ = betingade sannolikheten
Mer generellt kan vi definiera $Q : 2^\Omega \to \mathbb{R}$ genom $Q(A) = \dfrac{P(A \cap B)}{P(B)}$ (med andra ord, den betingade sannolikheten)

För att visa att Q är ett sannolikhetsmått måste vi visa att den uppfyller Kolmogorovs axiom:

- $P(A) \geq 0 \quad \forall A \in 2^\Omega$
- $P(\Omega) = 1$
- $P\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=0}^\infty P(A_i)$ om A_i är parvis disjunkta

Detta kommer inte vara så svårt, om vi visar det för $Q : 2^\Omega \to \mathbb{R}$ så har vi visat det för $Q : 2^B \to \mathbb{R}$.
Vi visar första axiomet:

$$Q(A) = \frac{P(A \cap B)}{P(B)} \geq 0 \quad \forall A \in 2^\Omega$$

Andra axiomet:

$$Q(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

Tredje axiomet:
Antag $A_1, A_2, \cdots$ disjunkta. Då är $B \cap A_1, B \cap A_2, \cdots$ också disjunkta.
Vi vill räkna följande:

$$Q\left(\bigcup_{i=1}^\infty A_i\right) = \frac{P\left(\left(\bigcup_{i=1}^\infty A_i\right) \cap B\right)}{P(B)}$$

Notera:

$$\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B = \bigcup_{i=1}^{\infty} (A_i \cap B) \text{ ty följande:}$$

$$\begin{aligned} x \in \left(\bigcup_{i=1}^{\infty} A_i\right) \cap B &\Rightarrow x \in A_i \text{ för något } i \text{ och } x \in B \\ &\Leftrightarrow x \in A_i \cap B \text{ för något } i \\ &\Leftrightarrow x \in \bigcup_{i=1}^{\infty} (A_i \cap B) \end{aligned}$$

Vi får då:

$$Q\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{P(\bigcup_{i=11}^{\infty} A_i \cap B)}{P(B)} = \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} = \underbrace{\frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)}}_{Q(A_i)} = \sum_{i=1}^{\infty} Q(A_i)$$

Nu följer det till exempel att:

- $P(A^c|B) = 1 - P(A|B)$
- $P(A \cup C|B) = P(A|B) + P(C|B) - P(A \cap C|B)$
- Om $A \cap B \subseteq A_2 \cap B \subseteq A_2 \cap B \subseteq \cdots$ så gäller $P\left(\bigcup_{i=1}^{\infty} A_i|B\right) = \lim_{n\to\infty} P(A_n|B)$

4.1. **Oberoende utsagor.**

Antag att $P(A) > 0$ och $P(B) > 0$. Vi säger att A och B är **oberoende** om $P(A|B) = P(A)$ och $P(B|A) = P(B)$

Anmärkning:
$P(B|A) = P(B) \Leftrightarrow P(A|B) = P(A)$. Kan bevisas genom Bayes sats.
Ytterliggare något att notera är att oberoende är ej en ekvivalensrelation ty den är ej transitiv.

Exempel:
Singla slant 2ggr, $\Omega = \{kr, kl\}^2$.
Vi ansätter A = första försöket ger krona $= \{(kr, kr), (kr, kl)\}$
Vi ansätter B = andra försöket ger krona $= \{(kl, kr), (kr, kr)\}$

Vi får då följande:

$$P(A) = \frac{1}{2} = P(B) \qquad P(A \cap B) = P(kr, kr) = \frac{1}{4}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{1/2} = \frac{1}{2} = P(A)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{1/4}{1/2} = \frac{1}{2} = P(B)$$

$$\Rightarrow A \text{ och } B \text{ är oberoende}$$

Exempel:
Låt Ω = Uppsalas vuxna befolkning.
Låt $A = \{\text{Man}\}$ $\quad B = \{\text{Bruna ögon}\}$ $\quad C = \{\text{Över 170cm}\}$

Avgör vilka som är oberoende

Definition/Sats 4.2: Bayes sats

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

Bevis 4.1: Bays sats

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

$\square$

Definition/Sats 4.3

Om $A\&B$ är oberoende så $\Leftrightarrow$ $A^c\&B$ oberoende $\Leftrightarrow$ $A\&B^c$ oberoende $\Leftrightarrow$ $A^c\&B^c$ oberoende

Bevis 4.2

Antag $A\&B$ är oberoende. Antag även att $P(A) > 0, P(B) > 0, P(A^c) > 0, P(B^c) > 0 Q(A) = P(A|B)$ är ett sannolikhetsmått, då gäller:

$$Q(A^c) = 1 - Q(A)$$

Det vill säga:

$$P(A^c|B) = 1 - \underbrace{P(A|B)}_{P(A)} = 1 - (P(A)) = P(A^c) \Rightarrow A^c\&B \text{ är oberoende}$$

Alla andra riktningar/implikationer följer på samma vis. $\square$

Definition/Sats 4.4: Enkel liten sats

$A\&B$ är oberoende $\Leftrightarrow P(A\cap B) = P(A)P(B)$

Bevis 4.3: Enkel liten sats

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

$\square$

Definition/Sats 4.5: Oberoende (part 2)

Detta är definitionen av oberoende vi i princip alltid kommer använda:
A och B är oberoende om $P(A\cap B) = P(A)P(B)$

Anmärkning:
Vad hännder om $P(A)$ eller $P(B)$ är 0?
Antag att $P(A) = 0$, eftersom $A\cap B \subseteq A \Rightarrow 0 \leq P(A\cap B) \leq P(A) = 0$
Detta ger då att $P(A\cap B) = 0 = P(A\cap B) = 0\cdot P(B)$
Men då betyder det att A och B alltid är oberoende om $P(A) = 0$

Anmärkning:
Vad händer om $P(A) = 1$?

Rimligtvis borde $P(A\cap B) = 1\cdot P(B) = P(B)$. Detta sker:

$$A\subseteq A\cup B \Rightarrow 1 = P(A) \leq P(A\cup B) \leq 1 \Rightarrow P(A\cup B) = 1$$
$$\underbrace{P(A\cup B)}_{1} = \underbrace{P(A)}_{1} + P(B) - P(A\cap B)$$
$$P(B) - P(A\cap B) = 0 \Leftrightarrow P(A\cap B) = P(B) = P(B)P(A)$$

Om $P(A) = 1$ så är A och B alltid oberoende, alltså kan vi utöka Sats 4.3 till godtyckliga hänelser A och B

Definition/Sats 4.6: Oberoende i flera variabler

$S\subseteq 2^{\Omega}$ är obereonde om $A_1,\cdots,A_n\in S \Rightarrow P(A_1\cap\cdots\cap A_n) = P(A_1)\cdot\cdots\cdot P(A_n)$

Exempel:
Säg att vi har en mängd $\{A,B,C\}$, mängden är obereonde om $P(A\cap B) = P(A)P(B)$ samt $P(A\cap C) = P(A)P(C)$ samt $P(B\cap C) = P(B)P(C)$ och $P(A\cap B\cap C) = P(A)P(B)P(C)$
Sista likheten är vitkig, ty om vi antar de 3 andra likheterna (parvis oberoende) är helt annat än full oberoende.

Exempel:
Låt $\Omega = \{1,2,3,4\}$, $P(n) = \dfrac{1}{4}$ samt $A = \{1,2\}$, $B = \{1,3\}$, $C = \{2,3\}$
Först och främst, $P(A) = \dfrac{1}{2} = P(B) = P(C)$
$P(A\cap B) = \dfrac{1}{4} = P(B\cap C) = P(A\cap C) \Rightarrow$ parvis oberoende
Om vi kollar sista grejen man måste kolla för obereonde, $P(A\cap B\cap C) = P(\emptyset) = 0$, men $P(A)P(B)P(C) = \dfrac{1}{8} \neq 0$, alltså ej oberoende i alla variabler.

Anmärkning:
Om A,B,C är parvis oberoende så är inte A,B,C nödvändigtvis oberoende, men om vi lägger till att A och $B\cap C$ är oberoende, så är A,B,C oberoende.
Detta gäller eftersom $P(A\cap(B\cap C)) = P(A)P(B\cap C) = P(A)P(B)P(C)$

Exempel:
Det är 22 personer i klassrummet, vad är sannolikheten att alla i klassrummet har olika födelsedagar? Vi kommer behöva göra några antaganden för att göra det här lite lättare för oss.
Vi betecknar A_n = person $1,\cdots,n$ har olika födelsedagar. Det vi söker är A_{22} (22 är en speciell siffra för det här problemet).
Antaganden:

- Antag att $P(A_1) = 1$ (uppenbart att en person har samma födelsedag som en person)
- $P(A_{n+1}|A_n) = \dfrac{365-n}{365}$ lika stor sannolikhet att födas på alla dagar (inga skottår i vår miljö)

Notera, $A_{n+1}\subseteq A_n \Rightarrow A_n\cap A_{n+1} = A_{n+1}$ samt $P(A_{n+1}|A_n) = \dfrac{P(A_{n+1})}{P(A_n)}$

Vi har då $P(A_22) = P(A_{22}|A_{21})P(A_{21}) = P(A_{22}|A_{21})P(A_{21}|A_{20})P(A_{20}$
$= \cdots = \underbrace{P(A_{22}|A_{21})}_{\frac{344}{365}}\cdots = \dfrac{364!}{343!365^{21}} \approx 0.52$

Detta var för $P(A_{22})$, för $P(A_n) = \dfrac{364!}{(365-n)!365^{n-1}}$

Vi sade även att 22 var ett speciellt tal, detta ty $P(A_{23}) \approx 0.49$, alltså där vi bryter 50 procent steget.

Exempel:

Antag att 80 procent av klassen gjorde inlämningsuppgifterna. Av de som gjorde inlämningsuppgifterna, så klarade 90 procent tentamen. Av de som inte gjorde inlämningsuppgifterna klarade 70 procent tentamen.

- Hur stor andel klarade tentamen?
- Hur stor andel av de som klarade tentamen hade gjort inlämningsuppgifterna?

Strategin här går ut på att skriva om uppgiften i matte-termer.
Ω = klassen, A = de som gjorde inlämningsuppgifterna
B = de som inte gjorde inlämningsuppgifterna = A^c
C = de som klarade tentamen

Det vi har givet är att $P(A) = 0.8$, samt att $P(B) == P(A^c) = 0.2$, $P(C|A) = 0.9$, $P(C^c|B) = 0.7 = P(C^c|A^c)$

Vi söker $P(C)$. Vi vet även att $A \cup B$ samt att $A \cap B = \emptyset$ (disjunkta).
Vi får då att $(A \cap C) \cup (B \cap C) = C$ samt $(A \cap C) \cap (B \cap C) = \emptyset$
Vi skriver om $P(C) = P((C \cap A) \cup (C \cap B)) = \underbrace{P(A \cap C)}_{P(C|A)P(A)} + \underbrace{P(B \cap C)}_{P(C|B)P(B)}$
$\Rightarrow 0.9 \cdot 0.8 + 0.7 \cdot 0.2 = 0.86 = P(C)$

Nästa uppgift söker efter $P(A|C)$. Här kan vi använda Bayes sats:

$$P(A|C) = \frac{P(C|A)P(A)}{P(C)} = \frac{0.9 \cdot 0.9}{0.86} \approx 0.837$$

Man kan tänka på det på följande sätt:

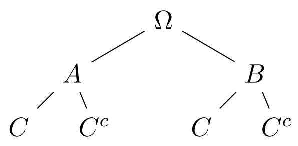

Figure 1.

Från högstadiet kanske vi minns att om vi vill veta sannolikheten att $C - A - C$ och $C - B - C$ inträffar så multiplicerar vi $P(C)P(A)P(C)$ och adderar produkten $P(C)P(B)P(C)$, men detta är ju precis det vi har ägnat föreläsningen åt!

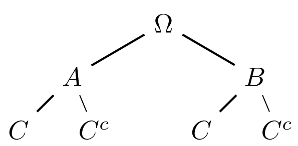

Figure 2.

Definition/Sats 4.7: Lagen om total sannolikhet

Antag att $A_1, \cdots, A_n$ är disjunkta och $B \subseteq \bigcup_{i=1}^{n}$.
Då är:

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Specialfall: $A \cap A^c \Rightarrow P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$

5. Sammanfattning K2

5.1. **Komplement och additionssatsen.**

Om A och B är godtyckliga händelser i utfallsrummet Ω så gäller följande:

- $P(A^c) = 1 - P(A)$
- $P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

5.2. **Sannolikhet på utfallsrum.**

Vi vill definiera en funktion som tar varje händelse i vårt utfallsrum och tilldelar den ett värde mellan 0 till 1 som talar om hur sannolikt det är att denna händelse inträffar.
Denna reellvärda funktion $P : \Omega \to \mathbb{R}$, kallar vi för *sannolikhetsfunktionen* på utfallsrummet.
Vi kan däremot inte kalla funktionen för ett sannolikhetsfunktion om den inte uppfyller följande axiom:

- $0 \leq P \leq 1$
- $P(\Omega) = 1$
- Om $A \cap B = \emptyset$ så gäller $P(A \cup B) = P(A) + P(B)$

Dessa axiom, kallas för *Kolmogorovs axiom.*

En funktion P som är definierad på delmängder till utfallsrummet Ω som också uppfyller Kolmogorovs axiom kallas för ett *sannolikhetsmått* på Ω

Ur detta ska vi se vad som händer om vi definierar betingning som en sannolikhetsfunktion, uppfyller den axiomen?

Vi fixerar en händelse C i vårt utfallsrum och defnierar en funktion $Q(A) = P(A|C) = \dfrac{P(A \cap C)}{P(C)}$ där $P(C) > 0$, vi verifierar om detta är ett sannolikhetsmått på Ω genom att kolla om axiomen uppfylls:

Första axiomet: Detta följer ur att $P(C) > 0$ och att $P \in [0, 1]$. Då kan inte bråket hamna utanför intervallet

Andra axiomet: Vi testar att stoppa in hela Ω i funktionen:

$$Q(\Omega) = P(\Omega|C) = \frac{P(\Omega \cap C)}{P(C)} = \frac{P(C)}{P(C)} = 1$$

Tredje axiomet: Här kommer vi nog behöva använda lite mängdlära, specifikt saker från komplement och additionssatsen samt distributiva lagar.
Givet att $A \cap B = \emptyset$ vill vi visa att detta betyder att $Q(A \cup B) = Q(A) + Q(C)$

$$Q(A \cup B) = P(A \cup B|C) = \frac{P\overbrace{((A \cup B) \cap C)}^{= P((A \cap C) \cup (B \cap C))}}{P(C)}$$

$$= \frac{P((A \cap C) \cup (B \cap C))}{P(C)} \overset{add.}{=} \frac{P(A \cap C) + P(B \cap C) - \overbrace{P\overbrace{(A \cap B)}^{= \emptyset}}^{=0}}{P(C)}$$

$$= Q(A) + Q(C)$$

Från detta, följer faktiskt följande:

- $P(A^c|C) = 1 - P(A|C)$
- $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$ —C

5.3. **Betingning.**

Givetvis kan faktumet att en annan händelse har inträffat påverka sannolikheten att en annan händelse inträffar, detta kallas för *betingning*, där man undersöker sannolikheten för att en händelse A inträffar, givet att en händelse B inträffar.
Uttallas även A betingat B och skrivs $P(A|B)$

Exempel:
Antag att vi har en kortlek (52 kort, 4st av dessa 52 är ess osv) och vi ska dra två kort från en kortlek. Låt A = händelsen att vi drar ett ess vid första draget och B = händelsen att vi drar ett ess vid andra draget, vad är då $P(B|A)$?

Om A har inträffat har vi inte längre 52 kort, utan 51 (vi har nämligen dragit ett) och vi har inte längre 4 ess, utan 3, alltså har vi en chans på $\frac{3}{51}$ givet att A har inträffat, vilket vi skriver på följande: $P(B|A) = \frac{3}{51}$

Definition/Sats 5.1: Betingad sannolikhet

Antag $P(A) > 0$. Den *betingade sannolikheten* för händelsen B givet att händelsen A har inträffat skrivs $P(B|A)$ och definieras som

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

Från detta följer det givetvis att $P(B \cap A) = P(A)P(B|A)$.
Coolt faktum! Eftersom snitt-operatorn är kommutativ, så innebär det faktiskt följande: $P(A|B) = P(B|A)$

Låt oss undersöka vad som händer som vi betraktar $P(A \cap B \cap C)$:

$$P(A \cap B \cap C) = P(\underbrace{(A \cap B)}_{=Q} \cap C) = P(C \cap \underbrace{(A \cap B)}_{=Q})$$

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

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

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

5.4. **Oberoende.**

Med betingning har vi undersökt hur sannolikheten påverkas av andra händelser, exempelvis hur sannolikheten att dra ett ess påverkas av att dra ett annat kort. När man studerar slumpexperiment är det ofta av intresse att veta om händelserna beror av varandra eller inte, eftersom de kan möjligen påverka slutsatserna av detta slumpexperiment.

Informellt säger vi att två händelser är *oberoende* om de inte har med varandra att göra.

Exempel:
Låt L = att vinna på lotto en viss dag, R = att det regnar i Stockholm samma dag
Eftersom dessa händelser inte har något med varandra att göra, så säger vi att dessa är *oberoende*. Det vi formellt vill formulera, är att sannolikheten för att L inträffar är densamma även om R inträffar (och vice versa).
Använder vi notationen från betingning, så uttrycker vi det på följande sätt:

$$P(L|R) = P(L) \qquad P(R|L) = P(R)$$

Det är faktiskt så vi definierar oberoende:

Definition/Sats 5.2: Oberoende händelser

Två händelser A och B sägs vara *oberoende* om:
$P(A|B) = P(A)$ förutsatt att $P(B) > 0$
$P(B|A) = P(B)$ förutsatt att $P(A) > 0$

Anmärkning:
Vi sade tidigare att betingade händelser kommuterar ($P(A|B) = P(B|A)$), detta gäller även här förutsatt att sannolikheten för vardera händelser är > 0, men från detta följer det ju att $P(B) = P(A)$. Från detta följer det då att det räcker att verifiera att $P(A|B) = P(A)$ för att visa att både A och B är oberoende!

Låt oss undersöka vidare, eftersom vi vet hur vi kan uttrycka $P(A|B)$, så bör vi kunna hitta ett uttryck för $P(A \cap B)$ förutsatt att A och B är oberoende:

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

Intressant! Givetvis antas $P(A)$ och $P(B)$ vara > 0

Svagare obereonde:
En svagare variant av oberoende är att titta på par av oberoende händelser i utfallsrummet. Att händelser är parvis oberoende innebär **inte** att mängden av dessa händelser är fullständigt oberoende, man måste nämligen undersöka alla par och se till att även de är oberoende.

Mer formellt säger vi att en mängd händelser $\{A_1, \cdots\}$ sägs vara *parvis oberoende* om för alla par (i, j) (där $i \neq j$), gäller att $P(A_i \cap A_j) = P(A_i)P(A_j)$
Mängden sägs vara *fullständigt oberoende* om det för alla $k \geq 2$ och alla delmängder $\{A_{i_1}, \cdots, A_{i_k}\}$ med $i_1 < \cdots < i_k$, gäller att $P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k})$

Exempel:
Antag att vi har två normala tärningar (6 sidor), en röd och en svart. Vi låter A vara händelsen att vi slår ett udda tal på den röda tärningen, och B vara händelsen att vi slår ett udda tal på den svarta tärningen. Låt nu C var händelsen att summan av den röda och svarta tärningen är udda.
Avgör om händelserna är parvis och eller fullständigt oberoende.

Det lättaste är att avgöra om händelserna är fullständigt oberoende, så vi kollar det först. Då vill vi kolla $P(A \cap B \cap C)$, vilket översatt till ord blir "sannolikheten att A, B, C inträffar". Att slå udda på båda tärningar är inte osannolikt, men tyvärr kan då inte summan bli udda eftersom udda+udda = jämnt. Alltså måste $P(A \cap B \cap C) = 0$.
Om det skulle vara så att händelserna är fullständigt oberoende, så skulle även $P(A \cap B \cap C) = 0 = P(A)P(B)P(C)$, men eftersom $P(A), P(B)$ och $P(C)$ har sannolikhet > 0, så motsäger detta att sannolikheten $= 0$, alltså är de ej fullständigt oberoende.

Vi undersöker nu om de är parvis oberoende. A och B är oberoende eftersom resultatet från A inte påverkar B alls. Det gäller nu att undersöka om (A, C) samt (B, C) är oberoende, men enligt anmärkningen ovan gäller det faktiskt bara att undersöka om A och C är oberoende, så följer det att B och C är oberoende (eftersom A och C är obereonde).
Vi vill kolla att $P(C|A) = P(C)$. Givet att A är ett udda tal, så måste alltså vi slå ett jämnt tal från B för att C ska gälla. Att slå ett jämnt tal har sannolikheten $\frac{1}{2}$, alltså är $P(C|A) = \frac{1}{2}$. Vi måste nu visa att $P(C) = \frac{1}{2}$:
Betrakta alla slagningar som par, vi får då $(1, 1), (1, 2), \cdots, (6, 6)$. Det är $6 \cdot 6 = 36$st par. Hur många av dessa par har ett udda och ett jämnt tal? Rimligtvis 18 av de! Alltså är sannolikheten att C inträffar $\frac{18}{36} = \frac{1}{2}$
Detta var ju dock precis $P(C|A)$, alltså har vi visat att $P(C|A) = P(C)$ vilket betyder att händelserna parvis är oberoende.

5.5. **Lagen om total sannolikhet.**

Premisserna går ut på att det ibland är lättare att beräkna en betingad sannolikhet än att direkt räkna sannolikheten.
Målet är att hitta en "sluten formel" för att räkna $P(B)$ betingat andra händelser i utfallsrummet. Vi undersöker:

FIGURE 3. Initialt

6. Slumpvariabler

Definition/Sats 6.1: Slumpvariabel

En *slumpvariabel* är en funktion $X : \Omega \to \mathbb{R}$. Till varje utfall $\omega \in \Omega$ associeras en *observation* $X(\omega) \in \mathbb{R}$

Exempel:
Vi tar vårt favoritexempel där $\Omega = \{\text{Uppsalas befolkning}\}$
Vi kan då låta X =längd, och ta en annan slumpvariabel Y = skostorlek, och sist men inte minst Z = ålder
Då hade $X(\text{Markus}) = 173$ och $Y(\text{Markus}) = 40$ och $Z(\text{Markus}) = 25$

Exempel:
Vi kan ta vår andra favorit, singla slant n gånger. Istället för krona klave, skriver vi $\{H, T\}$ för heads och tails.
Då är $\Omega = \{H, T\}$. Detta är ett exempel på en klassisk sannolikhet, det vill säga $P(\omega) = \dfrac{1}{2^n} \quad \forall \omega \in \Omega$
Vi kan då definiera en slumpvariabel X = antalet krona (heads), då kanske det hade sett ut på följande sätt om $n = 3$ och funktionen på följden hade sett ut på följande:

$$X(H, T, H) = 2 \qquad X(T, T, T) = 0$$

En annna slumpvariabel vi kan skapa är Y = antalet klave $= n - X$
En annan slumpvariabel vi kan skapa är följande:

$$X_1 = \begin{matrix} 1, & \text{första slanten hamnar på krona} \\ 0, & \text{annars} \end{matrix} \Bigg\}$$

$$X_i = \begin{matrix} 1, \omega_i = H \\ 0, \omega_i = T \end{matrix} \Bigg\}$$

$$\Rightarrow X = \sum_{i=1}^{n} X_i$$

En grej slumpvariabler är bra till är att beskriva händelser.

Exempel: Samma sannoliketsrum och $X : \Omega \to \mathbb{R}$

Då är $\{\omega : X(\omega) = 2\}$ = Antalet krona är exakt $P(\{\omega : X(\omega) = 2\}) = \dfrac{\binom{n}{2}}{2^n}$

Ett annat exempel vi kan ta är $\{\omega : X(\omega) \geq 2\}$ = Minst 2 krona. Vi vill nu hitta $P(X \geq 2)$.
Om vi lägger på följande: $P(\{X \geq 2\} \cup \{X < 2\})$ som är disjunkta och vi kan därmed summera utfallen $= P(X \geq 2) + P(X < 2) = 1$
Vi kan skriva om $P(X < 2) = P(\{X = 0\} \cup \{X = 1\}) = P(X = 0) + P(X = 1) = \dfrac{1}{2^n} + \dfrac{n}{2^n}$
Vi får då $\Rightarrow P(X \geq 2) = 1 - \dfrac{1}{2^n} - \dfrac{n}{2^n}$

Inga konstigheter, bara lite kombinatorik, hävdar föreläsaren.

Vi kan generalisera begreppet slumpvariabler:

Definition/Sats 6.2

En n-dimensionell slumpvariabel är en funktion $X : \Omega \to \mathbb{R}^n$

Isåfall kan vi skriva $X = (X_1, \cdots, X_n)$ där X_i är slumpvariabel
Vi kommer inte använda flerdimensionella slumpvariabler så mycket, men de kommer behövas för att uttrycka vissa händelser när vi har flera samtidigt.

Exempel:
Samma sannoliketsrum och samma definition av X_i. Då är $X_1, \cdots, X_n$ en n-dimensionell slumpvariabel

Det är viktigt att komma ihåg att dessa slumpvariabler måste vara definierade på samma sannoliketsrum.

Vi skriver till exempel $P(X = a)$ för $P(\{\omega : X(\omega) = a\})$.

Vi skriver även till exempel $P(a < X \leq b) = P(\{\omega : X(\omega) \in (a, b]\})$

Om vi skriver $P(X_1 \in A_1, \cdots, X_n \in A_n)$ menar vi att vi tar sannolikheten för snittet av alla, dvs $P(\{\omega : X_1(\omega) \in A_1\} \cap \cdots \cap \{\omega : X_n(\omega) \in A_n\})$

Skriver vi $X^{-1}(A) \qquad (A \in \mathbb{R})$ definierar vi detta genom $\omega \in X^{-1}(A) \Leftrightarrow X(\omega) \in A$. Kallas även för urbilden av A under X. Vi skriver $P(X \in A)$ för $P(X^{-1}(A))$

$P \circ X^{-1}$ definierar ett sannolikhetsmått på $\mathbb{R}$. Med andra ord $(P \circ X^{-1})(A) = P(X^{-1}(A)) = P(X \in A)$
Om det är ett sannolikhetsmått så ska Kolmogorovs axiom gälla, detta måste vi verifiera vilket vi gör enligt föjande:

- $P(X^{-1}(A)) \geq 0 \qquad \forall A \subseteq \mathbb{R}$ (detta gäller eftersom $X^{-1}(A) \subseteq \Omega$)
- $P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$
- Först notera att $X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B)$. Detta följer eftersom om vi tar ett element $\omega \in X^{-1}(A \cap B)$ så betyder det att $X(\omega) \in A$ och $X(\omega) \in B$
  Att säga det är samma sak som att säga $\omega \in X^{-1}(A)$ och $\omega \in X^{-1}(B) \Leftrightarrow \omega \in X^{-1}(A) \cap X^{-1}(B)$

  Så om $A \cap B = \emptyset$ så kommer $X^{-1}(A) \cap X^{-1}(B) = X^{-1}(A \cap B) = X^{-1}(\emptyset) = \emptyset$. Vi kan nu relatera disjunkta händelser i $\mathbb{R}$ till disjunkta händelser i Ω
  Tag nu en oändlig följd av händelser $A_1, A_2, \cdots \subseteq \mathbb{R}$. Då gäller
  $$X^{-1}\left(\bigcup_{i=1}^{\infty} A_i\right) = \bigcup_{i=1}^{\infty} X^{-1}(A_i)$$
  Så om $A_1, A_2 \cdots$ är disjunkta, så måste vi kolla att följande gäller:

$$P \circ X^{-1}\left(\bigcup_{i=1}^{\infty} A_i\right) = P(X^{-1}\left(\bigcup_{i=1}^{\infty} A_i\right)) = P\left(\bigcup_{i=1}^{\infty} \underbrace{X^{-1}(A_i)}_{\text{disjunkta}}\right) = \sum_{i=1}^{\infty} P(X^{-1}(A_i)) = \sum P \circ X^{-1}(A_i)$$

Definition/Sats 6.3: Diskreta slumpvariabler

Vi säger att en slumpvariabel $X : \Omega \to \mathbb{R}^n$ är en diskret slumpvariabel om sanholikhetsmåttet $P \circ X^{-1}$ är diskret.

Alternativt, X kallas diskret om det finns en Sannolikhetsfunktion $P_x(x)$ så att $P(X \in A) = \sum P_x(x)$

Vad betyder det att måttet var diskret? Jo, det betyder att det finns en uppräknelig mängd $\{x_1, x_2, \cdots, \} \subseteq \mathbb{R}^n$ så att $P(x \in \{x_1, x_2 \cdots\}) = 1$
Från tidigare föreläsningar vet vi att vissa av dessa utfall måste ha positiv sannolikhet.

Definition/Sats 6.4: Sannolikhetsfunktionen

$$P_X(x) = P(X = x)$$

Definition/Sats 6.5: Kontinuerlig/absolutkontinuerlig slumpvariabel

En *Kontinuerlig/absolutkontinuerlig slumpvariabel* X har en Riemann-integrerbar funktion $f : \mathbb{R}^n \to \mathbb{R}$ så att:

$$P(X \in A) = \int_A f(x)dx \qquad A \subseteq \mathbb{R}^n$$

Fördelningen (måttet Q) till en diskret slumpvariabel X bestäms unikt av Sannolikhetsfunktionen $P(X)$. Om X är kontinuerlig, så $P(X = x) = 0 \quad \forall x \in \mathbb{R}^n$, dvs inte definierad unikt.
X bestäms unikt av fördelningsfunktionen $F_X(x) = P(X \leq x)$

Exempel:
Låt $\Omega = \{H, T\}^n$ och slumpvariabeln X_i som den är definierad ovan.
Vi vill nu hitta $P(X = 1)$, vilket gäller om singlingen är krona $= \dfrac{1}{2}$ som är samma sak som $P(X = 0)$, alltså gäller följande:

$$P_{X_i}(X) = \left.\begin{matrix} 1/2,\ x = 0 \text{ eller } x = 1 \\ 0 \text{ annars} \end{matrix}\right\}$$

Tar vi X till att vara antalet krona (som tidigare), så letar vi efter $P(X = k)$. Vi kan börja med att undersöka vad $P(X = 0)$:

$$P(X = 0) = \frac{1}{2^n} \qquad P(X = k) = \frac{\binom{n}{k}}{2^n} \qquad (k = 0, 1, \cdots, n)$$

$$P_x(x) = \begin{cases} \dfrac{\binom{n}{x}}{2^n}, & x = 0, \cdots, n \\ 0, \text{ annars} \end{cases}$$

För en diskret slumpvariabel så bestäms *fördelningen* till X (sannolikhetsmåttet $P \circ X^{-1}$) unikt genom Sannolikhetsfunktionen.

Exempel:
Tag X_i från tidigare. Vad är då Sannolikhetsfunktionen för den flerdimensionella slumpvariabeln?
Vi söker alltså:

$$P_{X_1, \cdots, X_n}(x_1, \cdots, x_n) = P(X_1 = x_1, \cdots, X_n = x_n)$$

Vi antar att $\{x_1, \cdots, x_n\} \in \{0, 1\}$.
Men vad betyder det att någon av inputen är 0? Det som är viktigt att notera är att alla händelser i detta fall är oberoende, då kan vi göra

$$P(X_1 = x_1) \cdots P(X_n = x_n) = \frac{1}{2^n}, \qquad P_{\bar{X}}(\bar{x}) = \begin{cases} 1/2^n, x_i \in \{0, 1\} & \forall i \\ 0 \text{ annars} \end{cases}$$

Definition/Sats 6.6: Oberoende slumpvariabler

$X_1, \cdots, X_n$ är oberoende om för varje $A_1, \cdots, A_n \subseteq \mathbb{R}$ så är $\{x_1 \in A_1\}, \cdots, \{X_n \in A_n\}$ oberoende.

Med andra ord, så är sannolikheten (1) $P(X_{i_1} \in A_{i_1}, \cdots, X_{i_m} \in A_{i_m}) = P(X_{i_1} \in A_{i_1}) \cdots P(X_{i_m} \in A_{i_m})$
Detta går att skriva om:

$$P(X_{i_1} \in A_{i_1}, \cdots, X_{i_m} \in A_{i_m}, X_{j_1} \in \mathbb{R}, \cdots, X_{j_{n-m}} \in \mathbb{R})$$
$$\Rightarrow P(X_{i_1} \in A_{i_1}) \cdots P(X_{i_m} \in A_{i_m}) \underbrace{P(X_{j_1} \in \mathbb{R})}_{=1} \cdots \underbrace{P(X_{j_{n-m}}\mathbb{R})}_{=1}$$
$$\text{Antag } P(X_1 \in A_1, \cdots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n) \quad (2)$$

Då gäller (1) $\Leftrightarrow$ (2)

Det räcker alltså att kolla $P(X_1 \in A_1, \cdots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n)$ för att visa obereonde.

Definition/Sats 6.7

För diskreta slumpvariabler $X_1, \cdots, X_n$, har vi oberoende omm:

$$P_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = P_{X_1}(x_1) \cdots P_{X_n}(x_n)$$
$$\Leftrightarrow \{X_1 \in A_1\}, \cdots, \{X_n \in A_n\} \text{ är oberoende för varje } A_1, \cdots, A_n \subseteq \mathbb{R}^n$$

Bevis 6.1: Bevis av föregående sats

Riktningen $\Rightarrow$ är självklar (sätt $A_1 = \{X_1 = x_1\}, \cdots, \{X_n = x_n\}$)

Andra håller är mindre självklar. Vi vill visa $P(X_1 \in A_1, \cdots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n)$ för alla delmängder $A_1, \cdots, A_n$ $\square$

Exempel:
Låt $\Omega = \{\text{Uppsalas befolkning}\}$, X = längd, Y = vikt, Z = antal syskon
I detta exempel så är X, Y beroende, men X, Z är oberoende samtidigt är Y, Z beroende

Vi tänker oss att $X_1, \cdots, X_n$ är oberoende om de värden de antar inte ”påverkar varandra”

6.1. **Viktiga slumpvariabler.**

Definition/Sats 6.8: Bernoulli-fördelning

Vi säger att X är Bernoulli-fördelad om:
$$P_x(1) = P \qquad P_x(0) = 1 - P \qquad P \in [0, 1]$$

Exempel:
Gamla goda exemplet med singla slant är Bernoulli-Fördelad med $P = \frac{1}{2}$

Vi skriver $X \sim Be(P)$ eller $X \in Be(P)$

Exempel:
Singla slant exempel, fast $P(\omega) = P(\omega_1, \cdots, \omega_n) = P^k(1-p)^{n-k}$ för något $P \in [0, 1]$ där k är antalet krona.
Om vi definierar X_i som 1 om krona och 0 om klave, så blir $X_1, \cdots, X_n$ oberoende och $Be(P)$-fördelade.

Observation:
Lika fördelad är inte samma sak som lika! $P_X = P_Y$ medför inte att $X = Y$

Definition/Sats 6.9: Existens

Om X är diskret med Sannolikhetsfunktion P_X, så finns det oberoende slumpvariabler $X_1, \cdots, X_n$ med samma fördelning som X.

Bevis 6.2: Existens av oberoende slumpvariabler

Låt $A = \{x : P_X(x) > 0\}$, $\Omega = A^n$, $X_i(\omega) = X_i(\omega_1, \cdots, \omega_n) = \omega_i$.

Definiera $P(\omega) = P_{X_1}(\omega_1) P_{X_2}(\omega_2) \cdots P_{X_n}(\omega_n)$
Då följer $P(X_i = \omega_i) = P_X(\omega_i)$ $\square$

Definition/Sats 6.10: Binomialt fördelat

Vi säger att X är binomialfördelad om $P_X(k) = \binom{n}{k} P^k(1-P)^{n-k}$ för $k = 0, 1, \cdots, n$

Vi skriver $X \sim Bin(n, p)$

Detta kommer från att summan av Bernoulli-fördelade variabler blir precis binomialt fördelade.

Definition/Sats 6.11

Om $X_1, \cdots, X_n \sim Be(P)$ och oberoende så är $X = X_1 + \cdots + X_n \sim Bin(n, P)$

Detta följer ur $P(X_1 + \cdots + X_n = k) = P(X_i = 1 \text{ för } k\text{st } i \text{ och } X_i = 0 \text{ för } n-k\text{st } i)$

$$= \binom{n}{k} \underbrace{P(X_1 = 1, \cdots, X_k = 1, X_{k+1} = 0, \cdots, X_n = 0)}_{\underbrace{P(X_1 = 1) \cdots P(X_k = 1)}_{P^k} \cdots \underbrace{P(X_n = 0)}_{(1-P)^{n-k}}} = \binom{n}{k} P^k(1-P)^{n-k}$$

Vi tänker på $Bin(n, P)$ som följande:
Upprepade slumpförsök n gånger. Vinst med sannolikhet $P \in [0, 1]$ och förlust med sannolikhet $1-P = q$.
$X \sim Bin(n, p)$ räknar antalet vinster.

Exempel:

$$\{H, T\}^n,\ X = \text{antal } H = X_1 + \cdots + X_n,\ X_i = \begin{cases} 1, \omega_i = H \\ 0, \omega_i = T \end{cases} \quad \Rightarrow X \sim Bin(n.p)$$

Exempel:
Dra 10 lotter. Varje lott har vinstchans på 10%. $X = \text{antal vinster} \sim Bin(10, 0.1)$.
$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.9^{10} \approx 65\%$

Kom ihåg! Säg att vi vill räkna sannolikheten att vi har minst 5 vinster ($P(X \geq 5)$). Se sida 474 i boken. Där finns tabell över binomialfördelningar.

Notera!
Säg att $X \sim Bin(n, P)$ så är $n - X \sim Bin(n, 1-P)$

$$P(n - X = k) = P(X = n - k) = \binom{n}{n-k} P^{n-k}(1-P)^{n-(n-k)} = \binom{n}{k}(1-P)^k P^{n-k}$$

Definition/Sats 6.12

Om $X \sim Bin(n_1, P)$ och $Y \sim Bin(n_2, P)$ är oberoende, så är $X + Y \sim Bin(n_1 + n_2, P)$

Bevis 6.3

Vi vill hitta $P(X+Y=k)$:

$$=\sum_{j=0}^{k} P(X=j, Y=k-j) = \sum_{j=0}^{k} P(X=j)P(Y=k-j)$$

$$\sum_{j=0}^{k} \binom{n_1}{j} P^j (1-P)^{n_1-j} \binom{n_2}{k-j} P^{k-j}(1-P)^{n_2-(k-j)}$$

$$=\sum_{j=0}^{k} \binom{n_1}{j}\binom{n_2}{k-j} P^k (1-P)^{n_1+n_2-k}$$

$$=\left(\sum_{j=0}^{k} \binom{n_1}{j}\binom{n_2}{k-j}\right) P^k (1-P)^{n_1+n_2-k}$$

$$\sum_{j=0}^{k} \binom{n_1}{j}\binom{n_2}{k-j} = \binom{n_1+n_2}{k}$$

$$\Rightarrow X+Y \sim Bin(n_1+n_2, P)$$

Det följer att $P(j) = \dfrac{\binom{n_1}{j}\binom{n_1}{k-j}}{\binom{n_1+n_2}{k}}$ $\quad j = 0, \cdots, k$ är en sannolikhetsfunktion

□

Om X har fördelning P så skriver vi $X \sim Hyp(n_1, n_2, k)$, eller $X \sim Hyp(n_1+n_2, k, n_1)$, eller $X \sim Hyp(n_1+n_2, k, \underbrace{\frac{n_1}{n_1+n_2}}_{\text{vinstchans}})$.

Kallas för *Hypergeometrisk fördelning*

Intuition: Tänk n_1 som vinstlotter, och n_2 som lotter utan vinst. Dra k lotter. Då är X = antal vinstlotter, så kommer $Hyp(\underbrace{n_1+n_2}_{\text{antalet lotter}}, \underbrace{k}_{\text{dragningar}}, \underbrace{n_1}_{\text{vinstlotter}})$-fördelad

Exempel:
Givet en kortlek (52 kort) där 13st är hjärter. Dra 5 kort. Då är antal hjärter vi drar hypergeometriskt fördelad enligt $Hyp(52, 5, 13)$

Exempel:
Givet samma kortlek som föregående exempel. Dra 5 kort fast med återlägg (dra kort, kolla vad det är, lägga tillbaks i högen). Antalet hjärter är binomialfördelad där parametrarna blir $Bin(5, \frac{13}{52})$

Hypergeometrisk fördelning är alltså binomialfördelad fast utan återlägg.
Om $X \sim Hyp(N, n, m)$ så får vi med återlägg $Bin(n, \frac{m}{N})$.
Om N är mycket större än antalet dragningar n, så är $Hyp(N, n, m) \approx Bin(n, \frac{m}{N})$

7. Medelvärde

Vi börjar med ett exempel, myntkastet såklart där $\Omega = \{H, T\}^N$ och N är väldigt stort.
Vi definierar sannolikhetsmåttet på rummet som $P(\omega) = \dfrac{1}{2^n}$.
Vi har även de stokastiska variablerna som spottar ut vad vi får på det i:te kastet,
$X_i(\omega) = \begin{cases} 1, \omega_i = H \\ 0, \omega_i = T \end{cases}$

Om vi gör n myntkast och n är stort, förväntar vi oss att ha 50% H och 50% T eller alternativt formulerat ca $\dfrac{n}{2}$ krona. Detta är en frekvenstolkning.

Med andra ord, förväntar vi oss följande:

$$X_1 + \cdots + X_n = \frac{n}{2} \text{ med stor sannolikhet}$$

Det här med "stor sannolikhet" är viktigt, eftersom man *tekniskt sett* kan dra krona krona krona $\cdots$.

Definition/Sats 7.1: Stora talens lag

Om $X_1, X_2, X_3, \cdots$ är obereonde och likafördelade slumpvariabler så har vi, för varje $\varepsilon > 0$:

$$P\left(\left|\frac{x_1 + \cdots + x_n}{n} - E(X_i)\right| > \varepsilon\right) \underbrace{\rightarrow}_{n \rightarrow \infty} 0 \text{ för något tal } E(X_i)$$

För diskreta slumpvariabler är:

$$E(X) = \sum_x x P_X(x) \text{ om summan är absolutkonvergent } (\exists \text{ vissa specialfall})$$

Förutsatt att summan ej beror på ordningen av termer (absolutkonvergent eller $X \geq 0$ eller $0 \geq X$).
En slags mittpunkt för sannoliketsrummet.

Definition/Sats 7.2: Väntevärdet/Medelvärde

Talet $E(X)$ kallas *väntevärdet/medelvärdet* till X

Tänk såhär, om n är stort, förväntar vi oss cirka $n \cdot p$st x om $P_X(x) = p$, så vi förväntar oss alltså $X_1 + \cdots + X_n = \sum x \cdot n P_X(x)$ och $\dfrac{X_1 + \cdots + X_n}{n} \rightarrow \sum x P_X(x)$
(Vi summerar över alla X med positiv sannolikhet)

Vi kan definiera $E(X) = \sum x P_X(x)$ om summan är ∞ för varje ordning av termer (samma för $-\infty$), exempelvis om $X \geq 0$

Exempel:
Säg att $X \sim Be(p)$, då är $P_X(1) = p$, $P_X(0) = 1 - p$, $E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$

Exempel:
$X \sim Hyp(N, n, m)$:

$$E(X) = \sum_{k=0}^{n} k P_X(k) = \sum_{k=0}^{n} k \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$$

Per definition har vi att $\binom{n}{k} = \frac{n!}{(n-k)!k!} = n\frac{(n-1)!}{k(n-k)!(k-1)!} = \frac{n}{k}\binom{n-1}{k-1}$. Då är $E(X)$:

$$= \sum_{k=1}^{n} k \frac{m}{k} \frac{n}{N} \frac{\binom{m-1}{k-1}\binom{N-m}{n-k}}{\binom{N-1}{n-1}} = m\frac{n}{N} \underbrace{\underbrace{\sum_{k=0}^{n-1} \frac{\binom{m-1}{k}\binom{N-m}{n-1-k}}{\binom{N-1}{n-1}}}_{Hyp(N-1,n-1,m-1)}}_{=1} = n\frac{m}{N}$$

Exempel:

Från envariabelanalys vet vi att $\sum_{n=1}^{\infty} \frac{1}{n^2}$ konvergerar mot c ($c = \frac{\pi^2}{6}$)

Sätt $P_X(n) = \frac{1}{cn^2} \quad \forall n \in \mathbb{N}_+$. Då är $E(X)$:

$$= \sum_{n=1}^{\infty} n \frac{1}{cn^2} = \sum_{n=1}^{\infty} \frac{1}{cn} = \infty$$

Definition/Sats 7.3: Law of the unconcious statistician

Givet en funktion $g : \mathbb{R}^n \to \mathbb{R}$ Vi har $E(g(X)) = \sum g(x) P_X(x)$

Bevis 7.1: Law of the unconcious statistician

$$\begin{aligned} E(g(X)) &= \sum_{y:g(X)=y} yP(g(X)=y) = \sum_{y:g(X)=y} \sum_{x:g(x)=y} P(X=x) \\ &= \sum_{y:g(X)=y} \sum_{x:g(x)=y} \underbrace{y}_{=g(x)} P(X=x) \\ &= \sum_{y:g(X)=y} \sum_{x:g(x)=y} g(x)P(X=x) \\ &= \sum_{x} g(x)P(X=x) = \sum_{x} g(x)P_X(x) \end{aligned}$$

□

Definition/Sats 7.4

Vi säger att $X \in L^1(\Omega)$ om $\sum |x| P_X(x) < \infty$.
Mer generellt skriver vi att $X \in L^p(\Omega)$ om $\underbrace{\sum |x|^p P_X(x)}_{E(|X|^p)} < \infty$

Med andra ord, $X \in L^p$ om $E(|X|^p) < \infty$

$L^P(\Omega) = \left\{X : \Omega \to \mathbb{R} : E(|X|^P) < \infty, X \text{ är diskret}\right\}$
L^1 = absolutkonvergent = ändligt väntevärde

Definition/Sats 7.5: Väntevärdet är linjärt

$E(aX + bY) = aE(X) + bE(Y) \quad \forall a,b \in \mathbb{R} \quad X,Y \in L^1$

Eftersom L^1 är ett vektorrum så är $E : L^1 \to \mathbb{R}$

Bevis 7.2

Vi sätter $g(x,y) = ax + by$. Då blir $g(X,Y) = aX + bY$

$$\begin{aligned} E(aX+bY) = E(g(X,Y)) &= \sum_{x,y} g(x,y)P_{X,Y}(x,y) = \sum (ax+by)P_{X,Y}(x,y) \\ &= a\sum_x \sum_y xP_{X,Y}(x,y) + b\sum_y \sum_x yP_{X,Y}(x,y) \\ &= \sum_x x \underbrace{\sum_y P_{x,y}(x,y)}_{=P(X=x,\, y\in\mathbb{R}) = P_X(x)} + b\sum_y y \underbrace{\sum_x P_{x,y}(x,y)}_{=P_Y(y)} \\ &= a\sum_x xP_X(x) + b\sum_y yP_X(y) = aE(X) + bE(Y) \end{aligned}$$

$\square$

Exempel:
Tag miljön för myntkast. Då var $X = X_1 + \cdots + X_n \sim Bin(n, \frac{1}{2})$

Mer generellt, om $X_1, \cdots, X_n \sim Be(p)$ är obereonde så är $X = X_1 + \cdots + X_n \sim Bin(n,p)$
Eftersom väntevärdet var en linjär operator och väntevärdet för $Be(p) = p$, så kommer $E(X) = np$.

Så om $X \sim Bin(n,p)$ så är $E(X) = np$

Anmärkning:
Alla $X \sim Bin(n,p)$ kan inte skrivas $X = X_1 + \cdots + X_n$ där $X_1, \cdots, X_n$ är Bernoulli-fördelade!

Men, vi vet att det finns $X_1, \cdots, X_n \sim Be(p)$ som är oberoende och $X_1 + \cdots + X_n \sim Bin(n,p)$, och alla binomialfördelade variabler har samma väntevärde. Enligt definitionen av väntevärdet är det enbart sannolikhetsfunktionen som bestämmer vad väntevärdet är.

Definition/Sats 7.6

Om X är obereonde och Y är obereonde, så är $E(XY) = E(X)E(Y)$

Bevis 7.3

Vi visar detta på liknande sätt som tidigare:

$$\begin{aligned} g(x,y) = xy \quad E(XY) &\sum_{x,y} xy \underbrace{P_{X,y}(x,y)}_{(\text{oberoende}) \Rightarrow P_X(x)P_Y(y)} \\ &= \sum_x xP_X(x) \underbrace{\sum_y yP_Y(y)}_{E(Y)} \\ &= E(X)E(Y) \end{aligned}$$

$\square$

Definition/Sats 7.7: Varians

Variansen av X definieras genom:

$$Var(X) = E((X - E(X))^2), \quad X \in L^2$$

Intuition:
$X - E(X)$ är skillnaden mellan vad vi observerar och vad medelvärdet är, så om sannolikhetsfördelningen är utspridd så kommer vi observera många grejer som avviker och ligger långt ifrån väntevärdet. Om sannolikheten är liten, borde skillnaden vara liten.
Om X avviker från $E(X)$ mycket så är variansen $Var(X)$ stor, om X ligger nära $E(X)$ så är $Var(X)$ litet.

Tänk på det som ett medelvärde på hur mycket medelvärdet avviker från väntevärdet (**RÄTTA OM FEL**)

Var kommer kvadraten ifrån då? Då måste vi kolla på standardavvikelsen som för X definieras genom:

$$D(X) = \sqrt{Var(X)} \quad X \in L^2$$

Varför inte $D(X) = E(|X - E(X)|)$? Skillnaden mellan det vi observerar och medelvärdet? (detta är medelavvikelsen från medelvärdet). Har inte detta mer tydligt betydelse då?

Svaret på varför vi inte definierar det på det sättet är att det är svårare att räkna på, belopp är jobbiga att räkna med. Kvadrater är lättare att räkna på, oavsett hur vi defnierar det så kommer det vara ett mått på hur mycket variabeln avviker från väntevärdet.

Detta går givetvis att mäta på många sätt, men vår definition är lätt att räkna på.

Både $Var(X)$ och $D(X)$ är spridningsmått (hur mycket variabeln sprider sig på $\mathbb{R}$) och generellt är $Var(X)$ lättare att räkna på.

Exempel:

Låt Y vara en slumpvariabel med fördelningsfunktionen $F_Y(t) = P(Y \leq t) = \begin{cases} 0, t < 0 \\ t^2, t \in [0,1] \\ 1, t > 1 \end{cases}$

Rita upp $F(t)$
Beräkna $P(Y \leq 0.5) = F_Y(0.5) = 0.5^2 = 0.25$
Beräkna $P(0.5 < Y \leq 0.9)$. $\underbrace{P(Y \leq 0.9)}_{0.81} = P(\{Y \leq 0.5\} \cup \{0.5 < Y < 0.9\}) = \underbrace{P(Y \leq 0.5)}_{0.25} +$
$\Rightarrow 0.81 - 0.25 = 0.56$.
Eftersom de är disjunkta kan vi summera sannolikheterna.

Definition/Sats 7.8: Egenskaper hos fördelningsfunktioner

$$P(X < a) = \lim_{h \to 0^+} F(a - h)$$

Exempel:
Vid en produktion vill vi tillverka kolvar med en viss diameter. Vi har dock inte absolut precision, felet kan beskrivas med en slumpvariabel Y = absolutfelet i diametern. Täthetsfunktionen till Y är omvänt proportionell mot absolutfelet.

Bestäm täthetsfunktionen $f_Y(y) \quad y \in [1,5]$
$f_Y(y) = c\frac{1}{y}$. Vi måste även ha att integralen $\int_{-\infty}^{\infty} f_Y(y)dy = 1$

Bestäm fördelningsfunktionen (primitiv funktion till täthetsfunktionen)
$P(Y \leq t) = \int_{-\infty}^{t} f_Y(y)dt$
Om $t \leq 1 \Rightarrow P(Y \leq t) = 0$
Om $1 \leq t \leq 5 \Rightarrow P(Y \leq t) = \int_{-\infty}^{t} f_Y(y)dy = \int_{1}^{t} f_Y(y)dy = \frac{\ln(t)}{\ln(5)}$

Exempel:
Med tvåpunktsfördelning menas att $P_X(a) = p$ och $P_X(b) = 1 - p$ (notera att detta är $Be(p)$ om $a = 1$ och $b = 0$)
Beräkna $E(X)$ och $Var(X)$:
$E(X) = ap + b(1 - p)$

$$\begin{aligned} Var(X) &= E((X - E(X))^2) = E((X - (ap + b(1-p)))^2) \\ &= E(X^2 + 2XE(X) + (EX)^2) = E(X^2) \underbrace{-2(E(X)E(X)) + (E(X))^2}_{=(E(X))^2} \\ &\Rightarrow E(X^2) = \sum x^2 P_X(x) = a^2 P_X(a) + b^2 P_X(b) = a^2 p + b^2(1-p) \\ &\Rightarrow Var(x) = a^2 p + b^2(1-p) - (ap + b(1-p))^2 = p(1-p)(a-b)^2 \end{aligned}$$

7.1. **Egenskaper för väntevärden.**

- Väntevärdet av en konstant slumpvariabel, är inget annat än en konstant
- $E(X^p) = \sum x^p P_X(x)$ (här sätter vi $g(x) = x^p$)
- $E(|X|) = \sum |X| P_X(x)$ (låt $g(x) = |x|$)
- $E(X)$ är ändlig $\Leftrightarrow E|X| < \infty$
- $E(X + Y) = E(X) + E(Y)$ så länge väntevärderna är definierade (vi tillåter inte att ena är ∞ och den andra $-\infty$)
- $E(cX) = cE(X) \quad c \in \mathbb{R}$
- $|E(X)| \leq E(|(X)|)$ (Ye Olde' Triangelolikheten)
- $X \geq 0 \Rightarrow E(X) \geq 0$
- $X \leq Y \Rightarrow E(X) \leq E(Y)$

Proposition:
Om $q > p$ så är $L^q \subseteq L^p$

Bevis 7.4

Vi skriver $1_A(\omega) = \begin{cases} 1, \omega \in A \\ 0, \omega \notin A \end{cases}$

$|X|^P = |X|^P 1_{X \leq 1} + |X|^P 1_{X > 1}$:

$$\Rightarrow E(|X|^P) = \underbrace{E\left(\underbrace{\underbrace{|X|^P}_{\leq 1} 1_{x \leq 1}}_{\leq 1}\right)}_{\leq 1} + E(\underbrace{|X|^P}_{|X|^q} \underbrace{1_{x>1}}_{\leq 1})$$

Så:

$$E(|X|^q) < \infty \Rightarrow E(|X|^P) \leq 1 + E(|X|^q) < \infty$$

$\square$

Proposition:
Om $X, Y \in L^P \Rightarrow X + Y \in L^P$

Bevis 7.5

Notera att $|X| \leq \max\{|X|, |Y|\} \leq |X| + |Y|$.
Då gäller även följande (för $p \geq 1$):

$$|X+Y|^P \leq (|X| + |Y|)^P \leq (2\max\{|X|, |Y|\})^P \leq= 2^P \max\{|X|, |Y|\} \leq 2^P \left(|X|^P + |Y|^P\right)$$

$$\Rightarrow E(|X+Y|^P) \leq 2^P(E(|X|^P) + E(|Y|^P)) < \infty$$

$\square$

Proposition:
Om $X, Y \in L^2$ så $XY \in L^1$

Bevis 7.6

Uppenbarligen gäller:

$$|XY| \leq \frac{X^2 + Y^2}{2}$$

$\square$

Variansen av X defnierades som $E(X - E(X))^2$. Ett mått på hur mycket variabeln avviker från väntevärdet.
Standardavvikelsen definierade vi som $D(X) = \sqrt{Var(X)}$. En grej vi kan notera direkt är att $Var(X)$ alltid är positiv, alltså alltid defnierad.

Proposition:
$Var(X) < \infty \Leftrightarrow X \in L^2$. För $X \in L^2$ har vi:

$$Var(X) = E(X^2) - (E(X))^2$$

Bevis 7.7

Detta följer från

$$E(X - E(X))^2 = E(X^2 \underbrace{-2XE(X) + (EX)^2}_{\text{ändlig}}) = E(X^2) - \underbrace{-2E(X)E(X)}_{2(E(X))^2} + (E(X))^2$$

$$\Rightarrow E(X^2) - (E(X))^2$$

$\square$

Vi säger att $X \& Y$ är *okorrelerade* om $E(XY) = E(X)E(Y)$ för $X, Y \in L^2$
Notera, oberonde $\Rightarrow$ okorrelerade, men inte tvärtom!

Exempel:
$P_X(-1) = P_X(0) = P_X(1) = \frac{1}{3}$.

Då är $E(X) = -1 * \frac{1}{3} + 0 * \frac{1}{3} + 1 * \frac{1}{3} = 0$

Då är $E(X^3) = (-1)^3 * \frac{1}{3} + 0^3 * \frac{1}{3} + 1^3 * \frac{1}{3} = 0$

Då är $X \& X^2$ okorrelerade, men X och X^2 kan ju inte vara oberonde!
$P(X = 0, X^2 = 0) = P(X = 0) = \frac{1}{3} \neq P(X = 0)P(X^2 = 0) = \frac{1}{3} * \frac{1}{3}$

Varför bryr vi oss om okorrelerade variabler? Jo:

Proposition:
$Var(X+Y) = Var(X) + Var(Y) \Leftrightarrow X, Y$ är okorrelerade ($X, Y \in L^2$)

Bevis 7.8

Vi betraktar $Var(X+Y)$ som var $E(X+Y)^2 - (E(X+Y))^2$. Detta blir:

$$E(X^2 + 2XY + Y^2) - \big((E(X))^2 + 2E(X)E(Y) + (E(Y))^2\big)$$
$$\underbrace{\big(E(X^2) - (E(X))^2\big)}_{Var(X)} + \underbrace{\big(2E(XY) - 2E(X)E(Y)\big)}_{=0 \Leftrightarrow \text{X\&Y okorr.}} + \underbrace{\big(E(Y^2) - (E(Y))^2\big)}_{Var(Y)}$$

□

Anmärkning:
Vad är $Var(cX)$?:

$$Var(cX) = E(cX)^2 - (E(cX))^2$$
$$E(c^2X^2) - (cE(X))^2 = c^2E(X^2) - c^2(E(X))^2 = c^2Var(X)$$

Proposition:
Om X_1 och Y är okorrelerade och X_2 och Y är okorrelerade, så är $X_1 + X_2$ och Y okorrelerade.

Bevis 7.9

Vi har givet att $E(X_1Y) = E(X_1)E(Y)$ och $E(X_2Y) = E(X_2)E(Y)$.
Vi vill kolla vad $E((X_1 + X_2)Y)$:

$$= E(X_1Y + X_2Y) = E(X_1Y) + E(X_2Y) = E(X_1)E(Y) + E(X_2)E(Y)$$
$$\Rightarrow (E(X_1) + E(X_2))E(Y) = E(X_1 + X_2)E(Y)$$

□

Proposition:
Om $X_1, \cdots, X_n$ är parvis okorrelerade, så $Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n)$
Viktigaste specialfallet är när de är oberoende (ty det implicerar okorrelerade och vi kan då separera summorna).

Bevis 7.10

Vi kommer ihåg $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$. Om de är parvis okorrelerade bör ju även $X_1 + X_2$ och X_3 vara okorrelerade enligt Bevis 7.9. Men detta betyder att:

$$Var(X_1 + X_2 + X_3) = Var(X_1 + X_2) + Var(X_3) = Var(X_1) + Var(X_2) + Var(X_3)$$

Fortsätt med induktion □

Definition/Sats 7.9: Markovs olikhet

Om $X \in L^1$ (dvs $E(|X|) < \infty$) så är $P(|X| \geq a) \leq \dfrac{E(|X|)}{a}$ för $a > 0$
Ju större a är, desto mindre borde mängden $P(|X| \geq a)$ vara.

Bevis 7.11: Markovs olikhet

$$|X| = |X|\,1 > X \geq a + |X|\,1_{x<a}$$

$$E(|X|) = E(|X|\,1_{X\geq a}) + \underbrace{E(\underbrace{|X|\,1_{x<a}}_{\geq 0})}_{\geq 0} \geq E\left(\underbrace{|X|}_{\geq a}\,1_{X\geq a}\right)$$

$$\leq E(a1_{X\geq a}) = aE(1_{X\geq a}) = a(1 * P(X \geq a) + 0 * P(X < a)) = aP(X \geq a)$$

Alltså $E(|X|) \geq aP(X \geq a)$ □

En följd av detta är $X \in L^P \Rightarrow P(|X| \geq a) \leq \dfrac{E\,|X|^P}{a^P}$

Bevis 7.12

Vi ser:

$$P(|X| \geq a) = P(|X|^P \geq a^P) \leq \frac{E\,|X|^P}{a^P}$$

□

Definition/Sats 7.10: Chebyshevs olikhet

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2} \quad (\varepsilon > 0, X \in L^2)$$

Bevis 7.13: Chebyshevs olikhet

Sätt $P = 2$ i förra satsen:

$$P(|X| \geq \varepsilon) \leq \frac{E\,|X|^2}{\varepsilon^2}$$

Byt $|X|$ mot $|X - E(X)|$:

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{E(|X - E(X)|^2)}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}$$

□

Definition/Sats 7.11: Annorlunda Stora talens lag

Antag $X_1, X_2 \cdots \in L^2$ (oändlig följd av okorrelerade slumpvariabler)
Antag även att $E(X_1) = E(X_2) = \cdots = \mu \in \mathbb{R}$ och $Var(X_1) = Var(X_2) = \cdots = \sigma^2 \in \mathbb{R}$
Vi skriver $\bar{X_n}$ för medelvärdet:

$$\bar{X_n} = \frac{X_1 + \cdots + X_n}{n}$$

För $\varepsilon > 0$ har vi:

$$\lim_{n\to\infty} P\left(\left|\bar{X_n} - \mu\right| \geq \varepsilon\right) = 0$$

Bevis 7.14: Annorlunda Stora talens lag

$$E(\bar{X_n}) = E\left(\frac{1}{n}(X_1 + \cdots + X_n)\right) = \frac{E(X_1) + \cdots E(X_n)}{n} = \frac{n\mu}{n} = \mu$$

$$Var(\bar{X_n}) = Var\left(\frac{1}{n}(X_1 + \cdots + X_n)\right) = \frac{1}{n^2}Var(X_1 + \cdots + X_n)$$

$$\Rightarrow \frac{1}{n^2}(\overbrace{Var(X_1)}^{\sigma^2} + \cdots + \overbrace{Var(X_n)}^{\sigma^2}) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

$$\text{Chebyshevs olikhet sade } P\left(\left|\bar{X_n} - \underbrace{E(\bar{X_n})}_{\mu}\right| \geq \varepsilon\right) \leq \frac{Var(\bar{X_n})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \overbrace{\to}^{n\to\infty} = 0$$

□

Detta kallas oftast för "baby stora talens lag", det finns fler, men vi håller oss till denna i denna kurs.

7.2. **Kovarians.**

Kovariansen av den 2-dimensionella slumpvariabeln $(X, Y) \in L^2$ (väntevärderna av kvadraterna är ändliga) betecknas:

$$\begin{aligned} Cov(X, Y) &= E\left((X - E(X))(Y - E(Y))\right) \\ &= E(XY - XE(Y) - YE(X) + E(X)E(Y)) \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) = Cov(X, Y) \end{aligned}$$

Ett slags spridningsmått/varians för det 2-dimensionella fallet.

Egenskaper:

- $Cov(X, X) = Var(X)$
- $Cov(X, Y) = Cov(Y, X)$
- $Cov(aX, Y) = aCov(X, Y) = Cov(X, aY)$
- $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$
- $Cov(a, X) = 0 = Cov(X, a)$
- $Cov(X, Y) = 0 \Leftrightarrow E(XY) = E(X)E(Y) \Leftrightarrow X, Y$ är okorrelerade
- X, Y oberonde $\Rightarrow$ okorrelerade $\Rightarrow Cov(X, Y) = 0$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- $Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$
- $Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n) + 2\sum_{i<j} Cov(X_i, X_j)$

Notera! Detta betyder alltså att Cov är en bilinjär funktion!
Notera! första 4 punkter påminner om en inre produkt i linjär algebra, men Cov är inte en inre produkt på L^2

När är $Cov(X, X) = Var(X) = 0$?

$$Var(X) = E(X - E(X))^2 = \sum_x \underbrace{(x - E(X))^2 P_X(x)}_{\text{positiva, alla termer =0}} = 0$$

Om $x \neq E(X)$ så måste $P_X(x) = 0 \Rightarrow P_X(E(X)) = 1$. Alltså om $Var(X) = 0$ så måste $X = E(X)$ med sannolikhet 1, med andra ord X vara konstant på en mängd med sannolikhet 1. (X är nästan konstant).

Definiera en ekvivalensrelation $\sim$ på L^2 genom $X \sim Y$ om $X - Y$ är konstant med sannolikhet 1 (nästan konstant). Det finns en delmängd med sannolikhet 1 och för den delmängden så spottar $X - Y$ en konstant.

Ekvivalensklasser: $[X] = \{Y : Y \sim X\}$
Vi kan definiera $[X] + [Y] = [X + Y]$ och $a[X] = [aX]$, samt $Cov([X], [Y]) = Cov(X, Y)$. Alla dessa är väldefinierade.

Vi skriver $L^2/\sim = \{[X] : X \in L^2\}$ (vektorrum)
Kom ihåg, $X, Y \in L^2 \Rightarrow X + Y \in L^2$, samt $aX \in L^2 \quad (a \in \mathbb{R})$. Då är L^2 också ett vektorrum.

Kovarians är väldefinierat på $L^2/\sim$ och blir nu en inre produkt på $L^2/\sim$
Från detta följer Cauchy-Schwarz olikhet, dvs $|Cov([X],[Y])| \leq Cov([X],[X])Xov([Y],[Y]) = \sqrt{Var([X])Var([Y])}$
Notera, "ortogonala vektorer" ger att inre produtken är 0, vilket i vårat fall betyder att $Cov(X,Y) = 0$ vilket händer om X, Y är okorrelerade.

Definition/Sats 7.12: Korrelationskoefficienten

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$
$$|\rho(X,Y)| \leq 1$$
$$\rho(X,Y) = 0 \Leftrightarrow X, Y \text{ okorrelerade}$$

När är $|\rho(X,Y)| = 1$?

$$\rho([X],[Y]) = 1 \Leftrightarrow |Cov([X],[Y])| = \sqrt{Var([X])Var([Y])}$$

Likhet gäller $\Leftrightarrow$ vektorerna är linjärt beroende, dvs $[Y] = a[X] \quad (a \in \mathbb{R})$ eller om $[X] = 0$
Med anda ord är $|\rho(X,Y)| = 1 \Leftrightarrow X - aY = b$ på en mängd med sannolikhet 1 för några $a, b \in \mathbb{R}$.

Tänk på ρ som något slags mått på hur beroende variablerna är.

Den betingade sannolikhetsfunktionen $P_{X|Y}(x|y)$ är defnierad av (givet att $P_Y(Y) > 0$):

$$P_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P_{X,Y}(x,y)}{P_Y(y)}$$

Lagen om total sannolikhet för detta fall blir:

$$P(X = x) = \sum_y \underbrace{P(X = x|Y = y)P(Y = y)}_{P(X = x, Y = y)}$$
$$P_X(x) = \sum_y P_{X|Y}(x|y)P_Y(y)$$

Kom ihåg: För varje y så att $P_Y(y) > 0$ så är $P_{X|Y}(x|y)$ en sannolikhetsfunktion. Vi säger inte till vilken slumpvariabel, men exempelvis till slumpvariabeln

$$X|Y = y = X|_{\{Y=y\}} : \quad \underbrace{\{Y = y\}}_{\text{har sannolikhetsmått } Q(A) = P(A|Y=y)} \quad \to \mathbb{R}$$

Väntevärdet $E(X|Y = y) = \sum x P_{X|Y}(x|y)$

Det betingade väntevärdet $E(X|Y)$ är slumpvariabeln $E(X|Y)(\omega) = E(X|Y = y)$ om $Y(\omega) = y$. Detta gäller $\forall \omega \in \Omega$

Definition/Sats 7.13

$E(E(X|Y)) = E(X)$

Bevis 7.15

Vi sätter $g(y) = E(X|Y = y)$. Då är $g(Y) = E(X|Y)$

$$E(E(X|Y)) = E(g(Y)) = \sum_y g(y)P_Y(y) = \sum_y E(X|Y = y)P_Y(y) \sum_y \sum_x xP_{X|Y}(x|y)P_Y(y)$$

$$= \sum_x \sum_y xP_{X|Y}(x|y)P_Y(y) = \sum_x x \underbrace{\sum_y P_{X|Y}(x|y)P_Y(y)}_{P_X(x)}$$

$$= \sum xP_X(x) = E(X)$$

$\square$

8. Lektion 21/9

Övning 3.8.1
Betrakta den 2-dimensionella slumpvariabeln (X, Y) med sannolikhetsfunktion $P_{X,Y}(j, k) = c(j + k)$. Detta gäller för $i = 1, 2, 3$ och $k = 1, 2, 3$
Bestäm konstanten c och beräkna väntevärdet, variansen, och kovariansen. Beräkna även väntevärdet för X givet att $Y = 3$

Vi börjar med att bestämma konstanten. Om det ska vara en sannolikhetsfunktion så betyder att om vi summerar alla sannolikheter så får vi 1:

$$\sum_{j=1}^{3}\sum_{k=1}^{3} P_{X,Y}(j,k) = 1$$
$$= c\sum_{j=1}^{3}\sum_{k=1}^{3}(j+k)$$
$$\sum_{k=1}^{3}(j+k) = j+1+j+2+j+3 = 3j+6$$
$$\sum_{j=1}^{3} 3j+6 = 3\sum_{j=1}^{3}(j+2) = 3(3+4+5) = 3\cdot 12 = 36$$
$$\Rightarrow c\cdot 36 = 1 \Leftrightarrow c = \frac{1}{36}$$

Vi räknar väntevärdet:

$$E(X) = \sum_{k=1}^{3} kP_X(k), \quad P_X(k) = ?$$
$$P_X(k) = P(X=k) = P(X=k, Y=1) + P(X=k, Y=2) + P(X=k, Y=3)$$
$$= P_{X,Y}(k,1) + P_{X,Y}(k,2) + P_{X,Y}(k,3)$$
$$= \frac{k+1}{36} + \frac{k+2}{36} + \frac{k+3}{36} = \frac{3k+6}{36} = \frac{k+2}{12}$$
$$P_X(1) = \frac{3}{12} = \frac{1}{4} = P_Y(1)$$
$$P_X(2) = \frac{4}{12} = \frac{1}{3} = P_Y(2)$$
$$P_X(3) = \frac{5}{12} = P_Y(3)$$
$$\Leftrightarrow E(X) = 1\cdot\frac{3}{12} + 2\cdot\frac{4}{12} + 3\cdot\frac{5}{12} = \frac{3+2\cdot 4+3\cdot 5}{12} = \frac{13}{6}$$

Vi räknar variansen:

$$Var(X) = E(X^2) - (E(X))^2$$
$$= \sum_{k=1}^{3} k^2 P_X(k) - \frac{13^2}{6^2}$$
$$\sum_{k=1}^{3} k^2 P_X(k) = 1^2\cdot\frac{3}{12} + 2^2\frac{4}{12} + 3^2\cdot\frac{5}{12} = \frac{3+4\cdot 4+9\cdot 5}{12} = \frac{64}{12}$$
$$Var(X) = \frac{64}{12} - \frac{13^2}{6^2} = \frac{23}{36}$$

Vi räknar kovariansen:

$$Cov(X,Y) = E(XY) - \underbrace{E(X)E(Y)}_{=\frac{13}{6}\cdot\frac{13}{6}\ \text{(symmetri)}}$$

$$E(XY) = \sum_{k:P_{X,Y}>0} kP(XY=k), \quad X,Y \in \{1,2,3\}$$

$$\Rightarrow XY \in \{1\cdot 1, 1\cdot 2, 1\cdot 3, 2\cdot 1, 2\cdot 2, 2\cdot 3, 3\cdot 1, 3\cdot 2, 3\cdot 3\} = \{1,2,3,4,6,9\}$$

$$P(XY=1) = P(X=1,Y=1) = P_{X,Y}(1,1) = \frac{1+1}{36} = \frac{1}{18}$$

$$P(XY=2) = P(X=1,Y=2) + P(X=2,Y=1) = \frac{1+2}{36} + \frac{2+1}{36} = \frac{1}{6}$$

$$P(XY=3) = P(X=1,Y=3) + P(X=3,Y=1) = \frac{1+3}{36} + \frac{3+1}{36} = \frac{2}{9}$$

$$P(XY=4) = \frac{4}{36} = \frac{1}{9}$$

$$P(XY=6) = \frac{5}{18}$$

$$P(XY=9) = \frac{1}{6}$$

$$E(XY) = \sum_{k\in\{1,2,3,4,6,9\}} kP(XY=k) = 1\cdot\frac{1}{18} + 2\cdot\frac{3}{18} + 3\cdot\frac{4}{18} + 4\cdot\frac{2}{18} + 6\cdot\frac{5}{18} + 9\cdot\frac{3}{18} = \frac{14}{3}$$

$$\Rightarrow Cov(X,Y) = \frac{14}{3} - \frac{13^2}{6^2} = \frac{-1}{36}$$

Vi räknar det betingade väntevärdet:

$$E(X|Y=3) = \sum_{k=1}^{3} kP(X=k|Y=3)$$

$$= \sum_{k=1}^{3} k\frac{P(X=k,Y=3)}{P(Y=3)} = \sum_{k=1}^{3} k\frac{P_{X,Y}(k,3)}{P_Y(3)}$$

$$= \frac{12}{5}\sum_{k=1}^{3} k\frac{P_{X,Y}(k,3)}{=}\frac{12}{5}\frac{1}{36}\sum_{k=1}^{3} k(k+3)$$

$$= \frac{1}{15}(1\cdot 4 + 2\cdot 5 + 3\cdot 6) = \frac{32}{15}$$