UPPSALA UNIVERSITET

Föreläsningsanteckningar

Sannolikhetsteori 1

Rami Abou Zahra

Inlämningsdatum
October 18, 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:

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

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 $(x \in \Omega, \quad A \subseteq \Omega)$

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

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:

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

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty}(P(B_1) + P(B_2) + P(B_3) + \cdots + P(B_n))$$
$$\Leftrightarrow \lim_{n\to\infty}(P(A_1) + (P(A_2) - P(A_1)) + (P(A_3) - P(A_2)) + \cdots + (P(A_n) - P(A_{n-1})))$$
$$\Leftrightarrow \lim_{n\to\infty} P(A_n)$$

$\square$

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

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

□

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 $P(A) > 0, P(B) > 0, P(A^c) > 0, P(B^c) > 0Q(A) = P(A|B)$ 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. □

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

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

□

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.
$\Omega =$ 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:

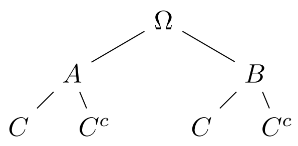

Figure 1.

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!

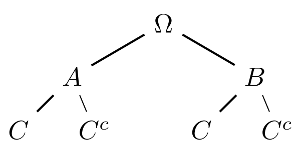

Figure 2.

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:

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

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 obereonde!

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 obereonde:

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

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

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

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. Sammanfattning K3

7.1. Definition av Slumpvariabel.

Vi påminner oss om definitionen av ett slumpförsök:

Definition/Sats 7.1: Slumpförsök

Ett *slumpförsök* på ett utfallsrum Ω består av ett försök som resulterar i ett av utfallen ($x \in \Omega$) i utfallsrummet.

Man vet ej på förhand vilka av utfallen som kommer inträffa, slumpförsök beskrivs genom att tala om vad sannolikheten att händelser inträffar i rummet är.

I vårt utfallsrum så har vi en samling händelser som kan inträffa, $\omega \in \Omega$. Genom att definiera ett sannolikhetsmått så kan vi mäta sannolikheten att dessa händelser inträffar, men i någon mening så säger de inte mycket *om* själva händelsen. Vi kan få ett numeriskt värde av $P(\{\omega\}) \geq 0$, men ω kan mycket väl vara en detaljrik händelse.

I vissa fall (rätt många) är det mer intressant att undersöka sannolikheten att ett numeriskt värde (som är associerat till händelser) inträffar, mer än att undersöka sannolikheten att händelser inträffar. Låt oss kika på exempel:

Exempel:
Du är en virolog som precis tagit fram vaccinet mot Covid-19. Innan du får använda det, måste du ta fram hur effektivt vaccinet är. Du testar ditt vaccin på sjuke Kalle, Anna, och Lina. Vi kan då definiera ett utfallsrum Ω som består av händelserna att personerna fortfarande är sjuka efter ha tagit vaccinet. Ω består då av tupler av personer som fortfarande är sjuka:

$$\Omega = \{(Kalle), (Anna), (Lina), (Kalle, Anna), (Kalle, Lina), \cdots\}$$

men som virolog så är detta inte av intresse.
Vad som är av intresse är givetvis *hur många* det var som fortfarande var sjuka efter de tagit ditt vaccin. Notera exempelvis, att effektiviteten av vaccinet inte ändras om det är $(Kalle, Anna)$ som fortfarande är sjuka eller om det är $(Kalle, Lina)$, trots att det är olika händelser i utfallsrummet.

Exempel:
Vi vill undersöka sannolikheten att summan av två tärningar visar 7 efter de kastas. Vårt utfallsrum består då av tupler av möjliga kombinationer och summor, exempelvis $(2, 3) \in \Omega$ som säger att ögonsumman är 5.
Detta blir snabbt opraktiskt att skriva ut eftersom vi skrivet ut att alla följande händelser inträffar:
(5,2), (2,5), (4,3), (3,4), (6,1), (1,6)

Därför definierar en så kallad *slumpvariabel*:

Definition/Sats 7.2: Slumpvariabel

En *Slumpvariabel* (även kallad *stokastisk variabel*) är en funktion som avbildar utfallsrummet på de reella talen

Betecknas $X : \Omega \to \mathbb{R}$

Det är ett otroligt ointuitivt namn för vad en slumpvariabel faktiskt är. Det är absolut inte en variabel, det är en funktion, och prefixed "slump" känns udda att ha med, men det för sent att ändra på namnet så vi får helt enkelt leva med det.
Kuriosa: Prefixed "slump" kommer troligtvis från att det är definierat på slumpförsök.

Anmärkning:

- Slumpvariabel sägs vara *diskret* om slumpvariabeln antar ändligt/uppräkneligt många värden
- Slumpvariabel behöver inte vara surjektiv

I första exemplet då du var en virolog, kan en slumpvariabel exempelvis vara antalet som *inte* tillfrisknade
I andra exemplet kan slumpvariabeln vara summan av ögonen på de båda tärningarna.

7.2. **Fördelningsfunktioner.**

Nu när vi har definierat vad en slumpvariabel är, vill vi tillämpa det på vår räkning. Vi vet att en slumpvariabels "output" är ett reellt tal, vi kan därmed undersöka sannolikheten att slumpvariabeln antar värden.
Vi kan också undersöka hur en slumpvariabel varierar genom att betrakta sannolikheten att slumpvariabeln antar högst ett värde lilla x, dvs $P(X(\omega) \leq x)$. Brukar betecknas $P(X \leq x)$

Detta, råkar även sammanfalla med definitionen av en så kallad *fördelningsfunktion*:

Definition/Sats 7.3: Fördelningsfunktion

Fördelningsfunktionen till en reellvärd slumpvariabel X är funktionen givet av:

$$F_X(x) = P(X \leq x) = P(\{\omega \mid X(\omega) \leq x\})$$

Fördelningsfunktionen kan komma att betecknas $P_X(x)$ framöver

Exempel:
Du ska bli betald att kasta tärning enligt följande:

Ögon	Utbetalning
1	1kr
2	2kr
3	2kr
4	4kr
5	4kr
6	4kr

Notera att vi har en slumpvariabel X som skickar "antal ögon" från utfallsrummet till ett reellt tal, dvs $X(6 \text{ ögon}) = 4$

Låt oss beräkna fördelningsfunktionen till slumpvariabeln X.
Notera att vi täcker en ganska liten del av de reella talen, faktumet är att vi täcker ett ändligt diskret intervall $\{1, 2, 4\}$. Detta betyder att får slumpvariabel är en *diskret slumpvariabel.*
Vi kan ställa oss frågan vad som händer när vi hamnar utanför intervallet, vad är sannolikheten att vår slumpvariabel $= 0$? Vi undersöker alltså:

$$P_X(x) = 0 \Leftrightarrow P(X \leq x) \quad \forall x < 1$$

Eftersom 0 inte finns i vår värdemängd för X är sannolikheten 0 att $X = 0$, oavsett händelse.

Vi undersöker vad sannolikheten att få ut 1kr är, dvs:

$$P_X(1) = 1/6 \Leftrightarrow P(X \leq 1) = \frac{1}{6}$$

Denna händelse inträffar antingen om $X < 1$ (som hade sannolikheten 0) eller om $X = 1$ som har sannolikheten $1/6$

Vi undersöker fördelningsfunktionen i $x = 2$. Notera att i tabellen är det 2 rader av 6 som har utbetalning 2kr.
Sannolikheten att $X = 2$ blir förstås $2/6$, men detta är inte värdet på fördelningsfunktionen i $x = 2$ just för att fördelningsfunktionen hade ett "mindre än eller lika med" tecken, alltså måste vi även addera sannolikheten att vi får ut 1kr, vilket ger oss totalt $3/6$:

$$P_X(2) = P(X \leq 2) = P(\{1, 2, 3\}) = \frac{3}{6}$$

Egenskaper hos fördelningsfunktionen:

- $0 \leq P_X(x) \leq 1 \quad \forall x$
- $\lim_{x \to -\infty} P_X(x) = 0$
- $\lim_{x \to \infty} P_X(x) = 1$
- $P(a < X \leq b) = P_X(b) - P_X(a)$
- $P(X > a) = 1 - P_X(a)$
- $P(X < a) = \lim_{h \to 0^+} P_X(a - h)$

Definition/Sats 7.4

Om $a \leq b$ så gäller att:

$$F_X(b) - F_X(a) = P(a < X \leq b)$$

Bevis 7.1

Vi vet att $F_X(b)$ är sannolikheten att $X \leq b$. Om $X \leq b$ så kan 2 fall inträffa:

- $X \leq a$
- $a < X \leq b$

Detta är oförenliga händelser (oberoende i någon mening), då kan vi använda additionssatsen (Kolmogorovs 3dje axiom):

$$F_X(b) = \underbrace{P(X \leq a)}_{F_X(a)} + P(a < X \leq b)$$

$$\Leftrightarrow F_X(b) - F_X(a) = P(a < X \leq b)$$

$\square$

Exempel:
Beräkna $P(51.5 < X \leq 56)$ givet att $P(X \leq 56) = 0.93$ och $P(X \leq 51.5) = 0.13$

Lösning:
Vi använder Sats 7.4:

$$P(51.5 < X \leq 56) = \underbrace{F_X(51.5)}_{P(X \leq 51.5)} - \underbrace{F_X(56)}_{P(X \leq 56)}$$

$$= 0.93 - 0.13 = 0.8$$

För att beskriva när slumpvariabeln *antar* ett värde inför vi så kallade *sannolikhetsfunktioner*. Detta begrepp kanske känns bekant sedan tidigare, och det stämmer, det har definierats tidigare, men förhoppningsvis förvirrar detta inte alltför mycket.

Definition/Sats 7.5: Sannolikhetsfunktion för slumpvariabel

Sannolikhetsfunktionen p_X för en diskret slumpvariabel X definieras av:

$$p_X(x) = P(X = x) = P(X \text{ antar värdet } x)$$

Exempel:
I exemplet ovan med utbetalning av att kasta tärning, hade istället $p_X(2) = 2/6$ eftersom i två av utfallen blir vi betalda 2kr

Egenskaper hos sannolikhetsfunktionen för slumpvariabler:

- $0 \leq p_X(k) \leq 1 \quad \forall k$
- $\sum_k p_X(k) = 1$
- $P(a \leq X \leq b) = \sum_{\{k: a \leq k \leq b\}} p_X(k)$
- $P(X \leq a) = \sum_{\{k: k \leq a\}} p_X(k)$
- $P(X > a) = \sum_{\{k: k > a\}} p_X(k) = 1 - \sum_{\{k: k \leq a\}} p_X(k)$

Anmärkning:
Säg att vi har en slumpvariabel X och vi vill räkna $P_X(x)$, men vår slumpvariabel X antar bara några värden under x. X antar värdet 0 överallt där den inte är definierad. Vi har främst betraktat diskreta fördelningsfunktioner, det vore konstigt att inte nämna något om kontinuerliga fördelningsfunktioner förstås

Anmärkning:
Fördelningsfunktionen kan beräknas ur sannolikhetsfunktionen enligt följande:

$$F_X(x) = \sum_{j \leq x} p_X(j)$$

På motsvarande sätt kan vi faktiskt också räkna sannolikhetsfunktionen ur fördelningsfunktionen! Om vi "låser in" X i ett intervall $a - 1 < X \leq a$, så kanske det blir mer uppenbart:

$$p_X(a) = P(X = a) \overset{\text{diskret.}}{=} P(a - 1 < X \leq a) = F_X(a) - F_X(a - 1)$$

Den nogranne kanske undrar vad som händer om vi stoppar in $a = 0$, då definierar vi följande:

$$p_X(0) = F_X(0)$$

Således får vi följande formel för att räkna $p_X(a)$:

$$p_X(a) = \begin{cases} F_X(0),\ a = 0 \\ F_X(a) - F_X(a-1),\ a \neq 0 \end{cases}$$

Exempel:
Enligt föregående anmärkning ser vi att vi kan räkna fördelningsfunktionen genom att summera sannolikhetsfunktionen, för att synliggöra i en tabell blir det:

k	$p_X(k)$	$F_X(k)$
0	0.05	0.05
1	0.10	0.15
2	0.20	0.35
3	0.30	0.65
4	0.20	0.85
5	0.10	0.95
6	0.05	1.00

7.3. **Kontinuerliga slumpvariabler.**

De slumpvariabler vi har betraktat har varit diskreta, det vill säga att de antar ändliga/uppräkneliga reella värden. Vi vill nu undersöka hur det ser ut när vår slumpvariabel kan anta ouppräknelig mängd värden, såsom ett delintervall till de reella talen.

Exempel:
Låt X vara en slumpvariabel. X talar om energin i en viss molekyl i ett givet medium. Eftersom X kan anta alla värden i ett givet intervall, så är X en kontinuerlig slumpvariabel.

Hur kan vi då ta fram villkor som gör att vi kan särskilja mellan kontinuerliga och diskreta slumpvariabler? Om vi har en diskret slumpvariabel är det relativt enkelt att summera händelser, detta sker naturligt med $\sum$, det är ”mekaniskt” möjligt att göra det för hand, givet ändliga värden.

För uppräkneliga värden blir detta lite svårare, men vi blickar tillbaks till envariabelanalysen, specifikt iden bakom integraltestet. Vi kan tänka på det som ”summerar vi oändligt små skillnader, så integrerar vi”.

Definition/Sats 7.6: Kontinuerlig slumpvariabel och täthetsfunktion

En slumpvariabel X sägs vara *kontinuerlig* om det finns en funktion $f_X(x)$ sådant att det för alla mängder A gäller att:

$$P(X \in A) = \int_A f_X(x)dx$$

Givet en kontinuerlig slumpvariabel X, beskriver täthetsfunktionen sannolikheten att X antar värde i mängden A. I det endimensionella fallet är detta A ett intervall, då ser det ut på följande (mer bekanta) form:

$$P(a < X \leq b) = \int_a^b f_X(x)dx$$

Anmärkning:
Detta gäller för *kontinuerliga* slumpvariabler X. För diskreta, är det bara att summera händelserna:

$$P(a < X \leq b) = \sum_{k=a}^{b} p_X(k)$$

Här är p_X sannolikhetsfunktionen för X

8. 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 8.1: Stora talens lag

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

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

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 8.2: 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 $\dfrac{X_1 + \cdots + X_n}{n} \to \sum x P_X(x)$
(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 $E(X)$:

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

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 8.3: 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 8.1: Law of the unconcious statistician

$$E(g(X)) = \sum_{y:g(X)=y} yP(g(X)=y) = \sum_{y:g(X)=y}\sum_{x:g(x)=y} P(X=x)$$

$$= \sum_{y:g(X)=y}\sum_{x:g(x)=y}\underbrace{y}_{=g(x)} P(X=x)$$

$$= \sum_{y:g(X)=y}\sum_{x:g(x)=y} g(x)P(X=x)$$

$$= \sum_{x} g(x)P(X=x) = \sum_{x} g(x)P_X(x)$$

□

Definition/Sats 8.4

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 8.5: 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 8.2

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

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

□

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 8.6

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

Bevis 8.3

Vi visar detta på liknande sätt som tidigare:

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

□

Definition/Sats 8.7: 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 8.8: 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 $f_Y(y) \quad y \in [1,5]$
$f_Y(y) = c\dfrac{1}{y}$. 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 $1 \leq t \leq 5 \Rightarrow P(Y \leq t) = \int_{-\infty}^{t} f_Y(y)dy = \int_{1}^{t} f_Y(y)dy = \dfrac{\ln(t)}{\ln(5)}$

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 $Var(X)$:
$E(X) = ap + b(1-p)$

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

8.1. **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 $-\infty$)
- $E(cX) = cE(X) \quad c \in \mathbb{R}$
- $|E(X)| \le E(|(X)|)$ (Ye Olde' Triangelolikheten)
- $X \ge 0 \Rightarrow E(X) \ge 0$
- $X \le Y \Rightarrow E(X) \le E(Y)$

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

Bevis 8.4

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

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

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

Så:

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

□

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

Bevis 8.5

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

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

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

□

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

Bevis 8.6

Uppenbarligen gäller:

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

□

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 8.7

Detta följer från

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

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

□

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 8.8

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 8.9

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 8.10

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 8.9: 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 8.11: 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 8.12

Vi ser:

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

□

Definition/Sats 8.10: Chebyshevs olikhet

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

Bevis 8.13: 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 8.11: 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 $\overline{X_n}$ för medelvärdet:

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

För $\varepsilon > 0$ har vi:

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

Bevis 8.14: Annorlunda Stora talens lag

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

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

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

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

□

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

8.2. **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 8.12: 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

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

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 8.13

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

Bevis 8.15

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

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

$\square$

8.3. **Mer om kontinuerliga sannoliketsrum.**

Standardexemplet är att slumpa ett reellt tal mellan 0 och 1. Alla möjliga utfall är givetvis talen mellan 0 och 1, så vårat utfallsrum är $\Omega = [0, 1]$
Detta intervall har längd 1. Säg att vi har $0 \leq a \leq b \leq 1$, vad är då sannolikheten att $P(x \in [a, b]) = b - a$ (längden av intervallet), eftersom längden ger hur stor del av intervallet $[0, 1]$ som utgörs av $[0, 1]$

Samma gäller för öppna intervall, samt halvöppna intervall
Givetvis finns specialfallet $P(\{a\}) = P([a, a]) = a - a = 0$

Från Kolmogorovs axiom samt att $\mathbb{Q}$ är en uppräknelig mängd följer det att $P(\mathbb{Q} \cap [0, 1]) = P(\bigcup_{q \in \mathbb{Q} \cap [0,1]} [q, q]) = \sum_{q \in \mathbb{Q} \cap [0,1]} P([q, q]) \sum 0 = 0$

För ett irrationellt tal $P([0, 1 \backslash \mathbb{Q}]) = \underbrace{P([0, 1])}_{\text{1-0=1}} - \underbrace{P[0, 1 \cap \mathbb{Q}]}_{0} = 1 - 0 = 1$

8.3.1. *Cantormängden.* Låt $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$
Då blir $P(C_1) = P([0, \frac{1}{3}]) + P([\frac{2}{3}, 1]) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$

Vi låter nu $C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$
$P(C_2) = 2^2 \cdot \frac{1}{3^2}$

Och så fortsätter vi att dela ner intervallen i tredjedelar$\cdots$
Notera att vi även dubblar antal intervall för varje indelning. Alltså blir $P(C_n) = 2^n \cdot \frac{1}{3^n} = \left(\frac{2}{3}\right)^n$

Cantormängden definieras som $\bigcap_{n=1}^{\infty} = C_n$
Frågan är om det finns något finns kvar i mängden, för vi delar in i mindre och mindre delar.
Exempelvis ligger $0 \in C$ samt ändpunkterna på delintervallen

Mer generellt; tag $x \in [0, 1]$, med decimalutveckling $x = 0.x_1x_2x_3 \cdots$ i bas 3, dvs $x = \sum_{n=1}^{\infty} = \frac{x_n}{3^n}$
Om $x_i \in \{0, 2\}$ så är $x \in \bigcap_{n=1}^{\infty} C_n$. Det följer att denna mängd är ouppräknelig

Man kan ställa sig frågan, vad är då $P(C)$?
$P(C_n) = \frac{2^n}{3^n}$. Vi vet att $C_n \subseteq \cdots C_3 \subseteq C_2 \subseteq C_1$
Om Kolmogorovs axiom håller måste vi har $P(C) = P\left(\bigcap_{n=1}^{\infty} C_n\right) = \lim_{n\to\infty} P(C_n) = \lim_{n\to\infty} \left(\frac{2}{3}\right)^n = 0$

Exempel:
Om vi ändrar lite på exemplet, det vill säga slumpa tal mellan 0 och 2. Då är $\Omega = [0, 2]$ där
$P(A) = \frac{\text{längden av } A}{2}$

Definition/Sats 8.14: Teaser: Kontinuerligt likformig slumpvariabel

En slumpvariabel $X : \Omega \to \mathbb{R}$ kallas *kontinuerligt likformig kontinuerligt* om:

$$\exists I = [a, b], a < b$$

$$\text{Om } a \leq c \leq d \leq b : P\left(X \in [c, d]\right) = \frac{\overbrace{d-c}^{||[c,d]||}}{\underbrace{b-a}_{||[a,b]||}}$$

Om slumpvariabeln X är likformig fördelad på ett intervall (a, b) säger vi $X \sim U(a, b)$

Definition/Sats 8.15: Absolut kontinuerlig fördelning

En slumpvariabel $X : \Omega \to \mathbb{R}^n$ kallas *absolut kontinuerlig* om det finns en Riemann integrerbar funktion $f : \mathbb{R}^n \to \mathbb{R}$ så att:

$$P\left(X \in A\right) = \int_A f_X(x)dx$$

För den endimensionella slumpvariabeln X gäller följande:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Vi måste även ha:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

för att Kolmogorovs andra axiom ska uppfyllas.
En sådan funktion $f_X(x)$ kallas för en täthetsfunktion.

Om $X \sim U(a, b)$, vad är då f_X?
Vi antar $a \leq c < d \leq b$

$$a \leq c \leq d \Rightarrow P\left(X \in [c, d]\right) = \frac{a-c}{b-a} = \int_c^d \frac{1}{b-a}dx$$

$$P(X \in [a, b]) = \frac{b-a}{b-a} = 1 = \int_a^b \frac{1}{b-a}dx$$

$$\Leftrightarrow f_X(x) = \begin{cases} \frac{1}{b-a}, a < x \leq b \\ 0, \text{ annars} \end{cases}$$

Exempel:
En endimensionell slumpvariabel $X : \Omega \to \mathbb{R}$ kallas normalfördelad om:

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \qquad \mu \in \mathbb{R}, \sigma > 0$$

Vi skriver $X \sim N(\mu, \sigma^2)$

Definition/Sats 8.16: Fördelningsfunktion

Fördelningsfunktionen till en slumpvariabel X är $F_X(x) = P(X \leq x)$
För kontinuerliga funktioner är vi främst intresserade över ett intervall:

$$\begin{aligned} P(a \leq X \leq b) &= P(\{X \leq b\} \setminus \{X \leq a\}) = P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a) \end{aligned}$$

Om X är absolutkontinuerlig så är:

$$P(a \leq X \leq b) = F_X(b) - F_X(a)$$

X kallas *kontinuerlig* om $P(X = x) = 0 \quad \forall x \in \mathbb{R}$
Om X är kontinuerlig så är $P(a \leq X \leq b) = P(a < XX < b) = F_X(b) - F_X(a)$

Exempel:
Säg att vi har $X \sim N(0,1)$, vi får då:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\dfrac{x^2}{2}}$$

Detta kallas för *standardiserad normalfördelning*
Vi skriver φ för fördelningsfunktionen atill $N(0,1)$, med andra ord:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\dfrac{x^2}{2}} \quad x \in \mathbb{R}$$

Om $X \sim N(,\mu,\sigma^2)$, så är

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{\sigma^2}} dx \\ &\int_{\dfrac{a-\mu}{\sigma}}^{\dfrac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\dfrac{y^2}{2}} dy \\ a \leq X \leq b \Leftrightarrow &\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma} \\ \text{Så } P(a \leq X \leq b) &= P(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}) \\ &= \varphi\left(\frac{b-\mu}{\sigma}\right) - \varphi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

Sätt $a' = \dfrac{a-\mu}{\sigma}$, $b' = \dfrac{b-\mu}{\sigma}$ får vi:

$$P(a' \leq \frac{X-\mu}{\sigma} \leq b') = \int_{a'}^{b'} \frac{1}{\sqrt{2\pi}} e^{-\dfrac{y^2}{2}} dy$$

Med andra ord är $X \sim N(\mu,\sigma^2) \Leftrightarrow \dfrac{X-\mu}{\sigma} \sim N(0,1)$

Definition/Sats 8.17: Centrala gränsvärdessatsen

Om vi tar en massa slumpvariabler $X_1, X_2, \dots \in L^2$ (de är alla oberoende), har väntevärdet μ med varians $\sigma^2 > 0$ och är lika fördelade (kan vara Bernoulli fördelade, Hypergeometrisk fördelade etc).

Då är medelvärdet $\overline{X_n} = \dfrac{X_1 + \cdots + X_n}{n} \approx N(\mu, \sigma^2)$ fördelad, i meningen att:

$$P\left(a \leq \frac{\overline{X_n} - \mu}{\sigma\sqrt{n}} \leq b\right) \to^{n\to\infty} \varphi(b) - \varphi(a)$$

$$= \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\dfrac{y^2}{2}} dy$$

$$\text{Kom ihåg: } E(\overline{X_n}) = E(\frac{1}{n}(X_1 + \cdots + X_n)) = \frac{n\mu}{n} = \mu$$

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

$$D(\overline{X_n}) = \frac{\sigma}{\sqrt{n}}$$

Övning:
Visa att om $E(X) = \mu \in \mathbb{R}$, $Var(X) = \sigma^2 > 0$, så är:

$$E\left(\frac{X-\mu}{\sigma}\right) = 0 \qquad Var\left(\frac{X-\mu}{\sigma}\right) = 1$$

Specialfall:
Om $X_i \sim Be(\frac{1}{2})$ (de Moivre's sats)
Generaliserar vi det till $X_i \sim Be(p)$ (de Moivre's-Laplace sats)

Betrakta X_1, X_2 som är oberoende slumpvariabler med $Be(\frac{1}{2})$:

$$X = \sum_{n=1}^{\infty} \frac{2X_n}{3^n} = \text{ decimalutveckling i bas 3}$$

Be är antingen 0 eller 1, $2X_n \in \{0, 2\}$ med båda sannolikheter $\dfrac{1}{2}$ att anta.
Då gäller $X \in C$ (Cantormängden)

Övning:
Visa att X ej är diskret, ej absolutkontinuerlig, men kontinuerlig.

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:

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

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

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

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

10. Kort introduktion till måtteori

Vi hade följande betecknkng $X \sim U(0,1)$ för att indikera att X är likformigt fördelat, med andra ord så har fördelningen sannolikhetsmått $P((a,b)) = b - a$ (om $(a,b) \subseteq [0,1]) = P([a,b]) = P([a,b))$.
Vi visade även att $P(\mathbb{Q}) = 0$ och att $P(C) = 0$ (där C är Cantormängden)

Men detta var bara några exempel, vad händer om vi har en godtycklig delmängd?
Dvs, vad är $P(A)$ för godtyckligt $A \subseteq [0,1]$?

Förslag: Om det är likformigt fördelning kan vi tänka oss att det ska vara längden av A (hur stor del av A täcker intervallet?). Vi kan kalla det för $P^*(A) = inf\,\{\sum_{n=1}^{\infty}(b_n - a_n) \text{ om } A \subseteq \bigcup(a_n, b_n)\}$
Detta kallas för det yttre måttet. Tyvärr kommer detta inte funka, utan vi kommer ha följande problem, $P^* : 2^{[0,1]} \to \mathbb{R}$ är inte ett sannolikhetsmått (uppfyller ej Kolmogorovs axiom) enligt Banach-Tarski problemet.

Vad vi inser är att vi bnorde sluta försöka hitta en delmängd. Vi skrotar de fula delmängderna och behåller den fina.

Definition/Sats 10.1

Ett mått på en mängd Ω är en funktion $\mu : 2^{\Omega} \to [0,\infty]$ som uppfyller:

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \underbrace{\sum_{n=1}^{\infty} \mu(A_n)}_{\infty \text{ om en term är } \infty} \qquad A_i \cap A_j = \emptyset \quad (i \neq j)$$

Notera att μ är ett sannolikhetsmått om $\mu(\Omega) = 1$

Vi vill definiera längden av delmängder, säg $A \subseteq \mathbb{R}$ (eller mer generellt, volym av $A \subseteq \mathbb{R}^n$)

Vi skriver det yttre måttet:

$$\mu^*(A) = inf\left\{\sum_{n=1}^{\infty}(b_n - a_n) \text{ om } \subseteq \bigcup(a_n, b_n)\right\} \quad A \subseteq \mathbb{R}$$

För $A \subseteq \mathbb{R}^n$:

$$\mu^*(A) = inf\left\{\sum_{n=1}^{\infty}(b_{k,1} - a_{k,1}) \cdots (b_{k,n} - a_{k,n}) \text{ om } A \subset \bigcup(a_{k,1}, b_{k,1}) \cdots (a_{k,n}, b_{k,n})\right\}$$

Problemet kvarstår, yttre måttet är inte ett mått och det finns inget mått, säg $\mu : 2^{\mathbb{R}} \to [0,\infty]$ så att om vi vill mäta intervall och vi vill att det ska vara $b - a$ och vi vill dessutom att måttet av $A + x$ ska vara samma A.
Det vill säga, vi vill att följande ska var uppfyllda:

$$(1) \quad \mu([a,b]) = b - a$$
$$(2) \quad \mu(A + x) = \mu(A)$$

μ^* uppfyller detta, men de uppfyller inte kraven från Definition 10.1

Iden är följande:
Hitta en stor nog samling delmängder $F \subseteq 2^{\mathbb{R}^n}$ så att $\mu : F \to [0,\infty]$ uppfyller Definition 10.1
F kommer vara följande:

$$B(\mathbb{R}^n) = \text{ minsta samling delmängder} F \subseteq 2^{\Omega} \text{ som uppfyller :}$$
$$(1) \quad (a_1, b_1)x \cdots x(a_n, b_n) \in F \qquad -\infty \leq a_i \leq b_i \leq \infty$$
$$(2) \quad A \in F \Rightarrow A^c \in F$$
$$(3) \quad A_1, A_2, \cdots \in F \Rightarrow \bigcup_{n=1}^{\infty} A_n \in F$$

Att det ens existerar ett ”minsta” i en mängd är inte alltid garanterat, men i detta fall gäller det (vi kikar på detta lite senare).
Om vi har (1)-(3), så kommer vi även ha exempelvis:

$$\begin{cases} (4) \quad A_1, A_2, \cdots, \in F \Rightarrow \bigcap_{n=1}^{\infty} (\bigcup_{n=1}^{\infty} A_n^c)^c \in F \\ (5) \quad B \backslash A = B \cap A^c \in F \text{ om } A, B \in F \\ \vdots \end{cases}$$

Vi vill mäta allt som vi kan bilda om vi tar union, komplement, snitt, differens osv.
Tar vi en mängd $A \in B(\mathbb{R}^n)$, så kallas de för *Borelmängder*

Definition/Sats 10.2: Specialfall av Caratheodorys sats

- $\mu^* : B(\mathbb{R}^n) \to [0, \infty]$ uppfyller Definition 10.1 (Lebesgue)
  Dessutom är μ^* den enda funktionen som tar Borelmängder och spottar ut någonting mellan 0 till ∞ så att $\mu^*((a, b)) = b - a$ (samma princip för $n \geq 1$)

  Man kan fråga sig hur stor är den här samlingen delmängder? Den är stor. Den är väldigt stor.
- (Caratheodory) Låt R vara mängden av delmängder på formen:
  $$I_1 \cup \cdots \cup I_m \qquad \text{för öppna intervall } \subseteq \mathbb{R}^n$$
  (Unionen av rektanglar)
  Om $\mu_0 : R \to [0, \infty]$ mäter delmängderna och uppfyller kraven för ett mått (Definition 10.1), så finns en unik $\mu : B(\mathbb{R}^n) \to [0, \infty]$ så att $\mu(A) = \mu_0(A)$ för $A \in R$

  Med andra ord, unionen av rektanglar bestämmer unikt vad måttet av allt annat är, vilket kanske inte är så förvånande om man tänker hur man approximerar exempelvis area (finare och finare rektanglar täcker arean). Detta påminner kanske lite om Riemann-integralen.

Exempel:
Om vi sätter $\mu_0((a, b)) = b - a$ får vi precis vad första punkten i Sats 10.2 säger.
Mer generellt, om vi sätter

$$\mu_0((a, b)) = \int_a^b f(x) dx$$

där $f \geq 0$ och Riemann-integrerbar, kommer också uppfylla Definition 10.1

Om $\int_{-\infty}^{\infty} f(x) dx = 1$ får vi ett sannolikhetsmått. Vi har exempelvis kollat på likformig fördelning:

$$f(x) = \begin{cases} \dfrac{1}{b - a}, x \in [a, b] \\ 0, x \notin [a, b] \end{cases}$$

Eller exempelvis:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\dfrac{(x - \mu)^2}{2\sigma^2}}$$

Sats 10.2 kommer bevisas i reell analys/integrationsteori. Detta är en maffig/enorm sats.

Vi ska nu definiera om Sannolikhetsrum så att vi inte mäter alla delmängder, utan vissa:

Definition/Sats 10.3: Sannolikhetsrum V2

Tag en delmängd $F \subseteq 2^{\Omega}$, denna kallas för en σ-algebra om:

- $\emptyset \in F$
- $A \in F \Rightarrow A^c \in F$
- $A_1, A_2, \cdots, \in F \Rightarrow \bigcup_{n=1}^{\infty} A_n \in F$

Ett sannolikhetsrum är trippeln (Ω, F, P) om $P : F \to [0,1]$ är ett sannolikhetsmått

Om vi då har en funktion μ som inte är definierad på hela sannolikhetsrummet utan på F så är det ett mått om måttet av unionen av A_n är summan av måtten:

$$\mu : F \to [0, \infty] \qquad m\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} m(A_n) \qquad A_i \cap A_j = \emptyset \quad i \neq j$$

Om $\mu(\Omega) = 1$ kallas μ ett *sannolikhetsmått*

Exempel:
Exempel på σ-algebra är $B(\mathbb{R}^n)$ eller 2^{Ω}

Det som är fiffigt är att vi kan specifiera vad sannolikheten till ett intervall är, och sedan vet vi att vi har ett sannolikhetsmått.

Exempel:
$(\mathbb{R}, B(\mathbb{R}), P)$ med $P((a,b)) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ ($N(0,1)$-normalfördelning) är ett sannolikhetsrum.
Då kommer Caratheodorys sats ge ett sannolikhetsmått på Borelmängderna (A):

$$B(\mathbb{R}^n) : P(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Detta är bara ett skrivsätt, vi kan exempelvis integrera över $\mathbb{Q}$ och få 0 (enligt samma argument som tidigare där vi visade att $P(\mathbb{Q}) = 0$)

En slumpvariabel kommer vi också definiera om:

Definition/Sats 10.4: Slumpvariabel V2

En slumpvariabel är en funktion $X : \Omega \to \mathbb{R}^n$ så att:
$$\{X\{\in A\}\} = \{\omega \in \Omega | X(\omega) \in A\} \in F \quad \forall A \in B(\mathbb{R}^n)$$
Framförallt om vi kan ge en sannolikhet till $\{a \leq X \leq b\} \in F$

Kontinuerlig om $\forall x \in \mathbb{R} \quad P(X = x) = 0$ (saknar atomer)
X kallas absolutkontinuerlig om det finns en funktion f_X så att $P(a \leq X \leq b) = \int_a^b f_X(x)dx \Leftrightarrow$ $F_X(t) = \underbrace{\int_{-\infty}^{t} f_X(x)dx}_{\text{täthetsfunktion}}$

En absolutkontinuerlig slumpvariabel bestäms unikt av f_X

Definition/Sats 10.5: Fördelning

$P(X \in A)$ är ett mått $Q : B(\mathbb{R}^n) \to [0,1]$

Fördelningen bestäms *unikt* av $P\overbrace{(a \leq X \leq b)}^{a<b\in\mathbb{R}}$

Det enda som krävs är att veta $P(a \leq X \leq b)$, kvittar Borelmängderna.
Den bestäms även unikt av $P(\underbrace{X}_{\text{fördeln. funk.}} \leq a)$

Fördelningsfunktioner F_X har följande egenskaper:

- Om $s \leq t \Rightarrow F_X(s) \leq F_X(t)$
- $\lim_{h\to 0^+} F_X(t+h) = F_X(t)$
- $\lim_{h\to 0^-} F_X(t+h) = P(X < t)$ (Motsäger sida 25?)
- $\lim_{t\to\infty} F_X(t) = 1$
- $\lim_{t\to -\infty} F_X(t) = 0$

Notera att F_X är kontinuerlig i t m $P(X = 1) = 0$
Framförallt är F_X en kontinuerlig funktion $\Leftrightarrow X$ är kontinuerlig slumpvariabel
Framförallt om F_X är deriverbar så är $f_X = F'_X$, så är f_X en täthetsfunktion (gäller även om F_X är deriverbar föutom i ändligt många punkter) X är absolutkontinuerlig $\Leftrightarrow$ det finns en funktion $f_X \geq 0$ så att $\int_a^b f_X(x)dx = F_X(b) - F_X(a) = P(a \leq X \leq b)$

DIY Sammanfattning:
Vi kan inte alltid ge en sannolikhet till varje delmängd, så vi gör det bara i en σ-algebra ($F \subseteq 2^\Omega$)
På reella talen använder vi Borelmängder, och, det här är det viktigaste, bestäms sannolikhetsmåttet $P : F \to [0,1]$ unikt av $P((a,b))$ för ($a < b \in \mathbb{R}$). Det bestäms också unikt av fördelningsfunktionen som var, säg $F(x) = P((-\infty, x])$ (exempelvis $P((a,b)) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\sigma > 0, \mu \in \mathbb{R}$)

Exempel:
Likformig fördelning på ett slutet intervall $[a,b]$. Då brukar vi skriva en sådan slumpvariabel som $X \sim U(a,b)$
Täthetsfunktionen har vi kikat på tidigare och blev:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, \; x \in [a,b] \\ 0, \; x \notin [a,b] \end{cases}$$

Fördelningsfunktionen har vi också räknat på tidigare:

$$F_X(t) = \begin{cases} 0, \ x \leq a \\ \dfrac{x-a}{b-a}, \ x \in [a,b] \\ 1, \ x \geq b \end{cases}$$

Exempel:
(Exponentialfördelning) $X \sim Exp(\beta) \quad (\beta > 0)$, vi skriver:

$$f_X(x) = \begin{cases} \beta e^{-\beta x}, \ x > 0 \\ 0, \ x < 0 \end{cases}$$

Integralen behöver vara lika med 1 för att Kolmogorovs axiom skall gälla. Detta är upp till läsaren att verifiera. Vi finner fördelningsfunktionen:

$$F_X(t) = \int_{-\infty}^{t} f_X(x)dx \begin{cases} 0, \ t \leq 0 \\ \int_0^t \beta e^{-\beta x} dx = -e^{-\beta x}|_0^t = 1 - e^{-\beta t} \end{cases}$$

Anmärkning:
Stor lutning i normalfördelning ger stor varians, dvs liten förflyttning ger stor varians

Exempel:

$X \sim N(\mu, \sigma^2)$ hade täthetsfunktion $f_X(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$

Definition/Sats 10.6: Överkurs

Om F uppfyller egenskaperna 1,2,3 under Sats 10.5, så är F en fördelningsfunktion till någon slumpvariabel.

Mer generellt, om $F_1, F_2, \cdots$ uppfyller 1,2,3 så finns oberoende slumpvariabler $X_1, X_2, \cdots$ så att $F_k = F_{X_k}$

Egenskaper: Absolutkontinuerliga slumpvariabler $X, Y : \Omega \to \mathbb{R}$

- X, Y oberoende $\Leftrightarrow f_{X,Y}(x,y) = f_X(x)f_Y(y)$
- Väntevärdet $E(X)$ definieras av:
$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx \quad \text{om integralen absolutkonvergent eller om den är } \infty \text{ eller } -\infty$$
- $E(g(X)) = \int g(x)f_X(x)dx$ om integralen är definierad
- $X \in L^p \Leftrightarrow E(|X|^p) = \int |X|^p f_X(x)dx < \infty$
- $Var(X) = E(X - E(X))^2 = \int (x - E(X))^2 f_X(dx) = E(X)^2 - (E(X))^2$ om $X \in L^2$
- $Cov(X,Y) = E(X - E(X))(Y - E(Y)) = E(XY) - E(X)E(Y)$ om $X, Y \in L^2$
- $E(g(X,Y)) = \int\int g(x,y)f_{X,Y}(x)dxdy$ om (X,Y) är Absolutkontinuerliga
- $E(aX + bY) = aE(X) + bE(Y)$, $Var(cX) = c^2Var(X)$, $Var(X+Y) = Var(X) + Var(Y)$ om $\underbrace{E(XY) = E(X)E(Y)}_{\text{okorrelerade slumpvariabler}}$, kovariansen är bilinjär osv gäller även för Absolutkontinuerliga slumpvariabler
- $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
- Markovs olikhet är bevarad: $P(X \geq a) \leq \dfrac{E(X)}{a}$ om $a > 0$
- Stora talens lag: $P(|\overline{X_n} - \mu| > \varepsilon) \overset{n\to\infty}{\to} \quad \forall \varepsilon > 0$
- Faltningsformlerna: Om X, Y är oberoende så:
$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$$
- Betingad fördelning: $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_(y)}$ om $f_Y(y) > 0 \; \forall y$
- Betingat väntevärde: $E(X|Y = y) = \int x f_{X|Y}(x|y)dx$
- $E(E(X|Y)) = E(X)$

Exempel:
Tag en exponentialfördelad variabel $X \sim Exp(\beta)$, vad är $E(X)$?

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^{\infty} x\beta e^{-\beta x}dx = \frac{1}{\beta}$$

Övning:
Visa att $Var(X) = \dfrac{1}{\beta^2}$

Exempel:
Tag en likformigt fördelad variabel och räkna $E(X)$

Likformig fördelad var om $f_X(x) = \begin{cases} \dfrac{1}{b-a}, \; x \in [a,b] \\ 0, \; x \notin [a,b] \end{cases}$:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx = \int_a^b \frac{x}{b-a}dx = \frac{x^2}{2(b-a)}|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

Vi passar på att räkna ut $Var(x)$:

$$E(X)^2 = \int_a^b x^2 f_X(x)dx = \frac{x^2}{3(b-a)}|_a^b = \frac{b^3 - a^3}{3(b-a)}$$

$$(E(X))^2 = \left(\frac{b+a}{2}\right)^2$$

$$Var(X) = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2 = \frac{4(b^2 + ab + a^2) - 3(b^2 + 2ab + a^2)}{12} = \frac{b^2 - 2ab + a^2}{12}$$

$$= \frac{(b-a)^2}{12}$$ $\Leftarrow$ Beror bara på $b - a$, variansen bryr sig inte om vi förflyttar intervallet

Övning:
Om $X \sim N(\mu, \sigma^2)$ så är $E(X) = \mu$ och $Var(X) = \sigma^2$.
Tips: Utnyttja att $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$
Använd Faltningsformlerna för att visa att om $X \sim N(0,1), Y \sim N(0,1)$ och X, Y obereonde, så är $X + Y \sim N(0,2)$ fördelad
Mer generellt, om $X \sim N(\mu_1, \sigma_1^2)$ och $Y \sim N(\mu_2, \sigma_2^2)$ $(X, Y$ oberoende) så är $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

Om $X \sim Exp(\lambda)$ och $Y \sim Exp(\lambda)$ $(\lambda > 0)$ och X, Y obereonde, vad är täthetsfunktionen till $X + Y$?
Vi använder faltningsformlerna:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} \overbrace{f_X(x)}^{>0 \Leftrightarrow x>0} \underbrace{f_Y(z-x)}_{x<z} dx$$

$$X \sim Exp(\lambda) \Leftrightarrow F_X(t) = \begin{cases} 1 - e^{-\lambda t}, \ t > 0 \\ 0, \ t \le 0 \end{cases}$$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, \ t > 0 \\ 0, \ t \le 0 \end{cases} \qquad E(X) = \frac{1}{\lambda} \qquad Var(X) = \frac{1}{X^2}$$

$$\Rightarrow f_{X+Y} = \begin{cases} 0, \ z \ge 0 \\ \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx, \ z > 0 \end{cases}$$

$$= \lambda^2 \int_0^z e^{-\lambda x} e^{-\lambda z + \lambda x} dx = \lambda^2 \int_0^z e^{-\lambda x} e^{-\lambda z} e^{\lambda x} dx$$

$$= \lambda^2 z e^{-\lambda z}$$

$$\Rightarrow f_{X+Y}(z) = \begin{cases} \lambda^2 z e^{-\lambda z}, \ z > 0 \\ 0, \ z \le 0 \end{cases}$$

$X + Y$ är alltså $\Gamma(2, \lambda)$

Övning:
Visa att om $X_1, \cdots, X_n \sim Exp(\lambda)$ är obereonde så är $X_1 + \cdots + X_n \sim \Gamma(n, \lambda)$
Detta går ut på samma sätt, visa med induktion

Övning:
Om $X \sim Exp(\lambda_1)$ och $Y \sim Exp(\lambda_2)$ och $\lambda_1 \neq \lambda_2$ och X, Y obereonde, vad är då täthetsfunktionen f_{X+Y}?

Övning:
Om $X \sim Exp(\lambda_1)$ och $Y \sim Exp(\lambda_2)$ och X, Y obereonde, vad är då fördelningen till $Z = \min\{X, Y\}$?
Vad som menas är $Z(\omega) = \min\{X(\omega), Y(\omega)\}$ (minsta av de)

Lösning:
Tricker är att kolla på fördelningsfunktionen. Vad är sannolikheten att $P(Z > z) = P(\min\{X, Y\} > z) = 1 - F_Z(z)$:

$$\min\{X, Y\} > z \Leftrightarrow X > z \quad \& Y > z$$

$$P(Z > z) = P(X > z, \ Y > z) \overset{ober.}{=} P(X > z) P(Y > z)$$

$$= (1 - F_X(z))(1 - F_Y(z)) = \begin{cases} 1, \ z \le 0 \\ e^{-\lambda_1 z} e^{-\lambda_2 z}, \ z > 0 \end{cases}$$

$$= \begin{cases} e^{-(\lambda_1 + \lambda_2) z}, \ z > 0 \\ 1, \ z \le 0 \end{cases}$$

$$\Rightarrow F_Z(t) = \begin{cases} 1 - e^{-(\lambda_1 + \lambda_2) t}, \ t > 0 \\ 0, \ t \le 0 \end{cases}$$

$$\Rightarrow \min\{X, Y\} \sim Exp(\lambda_1 + \lambda_2)$$

Övning:
Om $X \leq 0$ är $Exp(\lambda_2)$-fördelad ($\lambda > 0$), vad är täthetsfunktionen till $\sqrt{X}$?

Lösning:

$$F_{\sqrt{X}}(t) = P(\sqrt{X} \leq t) = \begin{cases} 0, \ t \leq 0 \\ P(X \leq t^2), \ t > 0 \end{cases}$$

$$= \begin{cases} F_X(t^2), \ t > 0 \\ 0, \ t \leq 0 \end{cases} \quad \Rightarrow f_{\sqrt{X}}(t) = \begin{cases} \frac{d}{dt} F_X(t^2), \ t > 0 \\ 0, \ t \leq 0 \end{cases}$$

$$= \begin{cases} 2t f_X(t^2), \ t > 0 \\ 0, \ t \leq 0 \end{cases} \quad = \begin{cases} 2\lambda t e^{-\lambda t^2}, \ t > 0 \\ 0, \ t \leq 0 \end{cases}$$

Övning:
Beskriv fördelningen till X^2 och X^3 om $X \sim Exp(\lambda)$ eller om $X \sim N(0,1)$

Om $X \sim Exp(\lambda)$ så gäller (betingad):

$$P(X > s + t \mid X > s) = P(X > t) \qquad (s, t \leq 0)$$

(Minneslöshet kallas detta även)

Lösning:

$$P(X > s + t \mid X > s) = \frac{P(X > s + t \,\&\, X > s)}{P(X > s)}$$

$$= \frac{P(X > s + t)}{P(X > s)} = \frac{1 - F_X(s + t)}{1 - F_X(s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F_X(t) = P(X > t)$$

Exempel på minnelösa slumpvariabler
Om $X =$ tiden det tar innan nästa kund kommer in på ICA så är det minneslöst. Slumpvariabeln är fördelad på samma sätt även om vi hoppar fram i tiden.

Ett annat exempel kan vara tiden innan det ringer i mobilen, är också minneslöst

Däremot, om $Z =$ tiden jag lever är ej minneslöst, om jag lever till minst 100 år, så är sannolikheten att jag lever till 150år lite annorlunda än sannolikheten att jag lever till 50år.

Definition/Sats 10.7

Om $P(X > s + t \mid X > s) = P(X > t)$ för alla $s \geq 0$ och $t \geq 0$ så är X exponentialfördelad.

Bevis 10.1

Låt $\varphi(t) = P(X > t)$
Då kommer alltså $P(X > s + t \mid X > s) = P(X > t) \Leftrightarrow \varphi(s + t) = \varphi(s)\varphi(t)$

Vi ska försöka hitta φ som löser ovanstående ekvationen. Notera monotont ökande.

Sätt $q = \varphi(1)$, vad är då $\varphi(2)$?:

$$\varphi(2) = \varphi(1 + 1) = \varphi(1)\varphi(1) = q^2$$
$$n \in \mathbb{Z}_+ \Rightarrow \varphi(n) = \varphi(\underbrace{1 + \cdots + 1}_{n \text{ st}}) = \varphi(1)^n = q^n$$

Vi vet att φ tar heltalen som input, men vad är $\varphi(1/n)$?:

$$\varphi\left(\frac{1}{2}\right) = \sqrt{q} = q^{1/2}$$

På samma sätt är $\varphi(1)$:

$$\underbrace{\varphi\left(\frac{1}{n} + \cdots + \frac{1}{n}\right)}_{n \text{ termer}} = \varphi\left(\frac{1}{h}\right)^n \Rightarrow \varphi\left(\frac{1}{n}\right) = q^{1/n}$$

$$\frac{m}{n} \in \mathbb{Q}_{>0} \Rightarrow \varphi\left(\frac{m}{n}\right) = \varphi\left(\frac{1}{n}\right)^m = (q^{1/n})^m = q^{\frac{m}{n}}$$

Alltså $P(X > t) = q^t$ om $t \in \mathbb{Q}_{>0}$ så $F_X(t) = 1 - q^t$ för $t \in \mathbb{Q}_{>0}$

Sätt $q = e^{-\lambda}$ för $\lambda > 0 \Rightarrow F_X(t) = 1 - e^{-\lambda t}$

Vi har bara hittat funktionen för rationella tal, vad kan vi göra då?
Vi vet att fördelningsfunktionen är högerkontinuerlig och monotont ökande, så för $t > 0$ ($t \in \mathbb{R}$), ta rationella tal $q_n > t$ så att $q_n \to t$
Eftersom fördelningsfunktionen är högerkontinuerlig så måste $F_X(q_n) \to F_X(t)$:

$$1 - e^{-\lambda q_n} \overset{n \to \infty}{\to} 1 - e^{-\lambda t}$$

$F_X(t) = 1 - e^{-\lambda t}$ för $t \leq 0$, så $X \sim Exp(\lambda)$, men den är noll för alla negativa tal också $\square$

Exempel:
På ICA kommer det in 20 kunder per timme.
Om vi sätter X = tiden innan nästa kund kommer in, så är $X \sim Exp(\lambda)$, vad är då paramtern λ?

Väntevärdet kommer vara $E(X) = \frac{1}{20} = \frac{1}{\lambda} \Leftrightarrow \lambda = 20$
X är alltså exponentialfördelad med $\lambda = 20$

Antalet kunder som dyker upp inom ett visst tidsintervall följer något som kallas för *poisson*-fördelning

Mer nogrannt, säg att vi tal en massa slumpvariabler $X_1, X_2, \cdots$ som är normalfördelade med paramter λ och de alla är oberoende
Tar vi en variabel N_t så är antalet kunder som dyker upp i intervallet $[0, t]$
Då kommer N_t följa en poisson-fördelning med parameter λt

Vi skriver då $X \sim Po(\lambda t)$. Säger vi att $X \sim Po(\lambda)$ så är:

$$P_X(n) = \frac{\lambda^n}{n!} e^{-\lambda}$$

Om $X \sim Po(\lambda)$, vad är då $E(X)$?

$$\begin{aligned} E(X) &= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} e^{-\lambda} \\ &= \sum_{n=1}^{\infty} \frac{\lambda \lambda^{n-1}}{(n-1)!} e^{-\lambda} = \lambda \underbrace{\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda}}_{=1} = \lambda \end{aligned}$$

Övning:
Visa att $Var(X) = \lambda$

10.1. **Minneslösa diskreta variabler.**

Om $P(X > m + n \mid X > m) = P(X > n)$ för positiva heltal m, n och samma för X
Då kan vi igen sätta $\varphi(n) = P(X > n)$ (beviset blir kortare för vi betraktar enbart heltal)

Då kommer $\varphi(n)$ att vara $\varphi(1)^n$ enligt samma argument som tidigare.
Sätt då $\varphi(1) = q \in (0, 1)$, då är $\varphi(n) = q^n$, $P(X > n) = q^n$

Vad är då $P_X(n)$?

$$P(X = 1) = 1 - P(X > 1) = 1 - q^1 = 1 - q = p \in (0, 1)$$

Vad är då sannolikheten att $X = 2$?

$$P(X = 2) = 1 - P(X = 1) - P(X > 2) = \overbrace{1 - p}^{q} - q^2 = q(1 - q) = pq$$

Definition/Sats 10.8

Om vi vill hitta $P(X = n)$ är det pq^{n-1}

Bevis 10.2

Stämmer om $n = 1$, vi använder induktion efter det. Om det stämmer $\forall k < n$, vad händer då?

$$P(X = n) = 1 - \overbrace{P(X > n)}^{q^n} - P(X < n)$$

$$= 1 - q^n - \sum_{k=1}^{n-1} P(X = k) = 1 - q^n - \sum_{k=1}^{n-1} pq^{k-1} = 1 - q^n - p \underbrace{\sum_{k=0}^{n-2} q^k}_{\text{geometrisk summa}}$$

$$= 1 - q^n - p\frac{q^{n-1} - 1}{\underbrace{q - 1}_{-p}} = 1 - q^n - (1 - q^{n-1}) = q^{n-1} - q^n = q^{n-1} \underbrace{(1 - q)}_{p}$$

$$= pq^{n-1}$$

□

Exempel på en sådan är singla slant

Definition/Sats 10.9: Första gången fördelad

X kallas för *första gången fördelad* ($X \sim \text{ffg}(p)$) om $P(X = n) = p(1 - p)^{n-1}$

Exempel:
Om $X_1, X_2, \cdots \sim Be(p)$ och är oberoende, då är $X = \min\{n : X_n = 1\}$, med andra ord $X(\omega) = \min\{n : X_n(\omega) = 1\}$
Då är $X \sim \text{ffg}(p)$, varför det?

$$P(X = n) = P(X_1 = 0, \cdots, X_{n-1} = 0, X_n = 1) = \underbrace{P(X_1 = 0)}_{1-p} \cdots \underbrace{P(X_{n-1} = 0)}_{1-p} \underbrace{P(X_n = 1)}_{p}$$

$$= p(1 - p)^{n-1}$$

Övning:
Visa att $E(X) = \dfrac{1}{p}$ samt att $Var(X) = \dfrac{1-p}{p^2}$

Anmärkning:
X = antalet försök som krävs för att få ett lyckat utfall (med sannolikhet p)
$\sim$ffg(p) och då vet vi att $P(X = n)p^q n - 1 \quad n = 1, 2 \cdots$

Definition/Sats 10.10: $E(X) = \frac{1}{p}$

Om vi naivt skriver upp summan:

$$\sum npq^{n-1} = \frac{d}{dq}q^n = \cdots$$

Så går det att lösa ut analytiskt, men det har inte lika mycket med sannolikhet att göra och ger därmed inte så mycket förståelse.

Om vi istället betingar på första försöket:

$$P(X = k|X_1 = 1) = 1 \qquad \text{om } k = 1,\ 0 \text{ annars}$$

$$E(X|X_1 = 1) = \sum_{k=1}^{\infty} kP(X = k, X_1 = 1) = 1$$

Vad händer om vi betingar på de andra:

$$P(X = k, X_1 = 0) = \frac{P(X = k, X_1 = 0)}{\underbrace{P(X = 0)}_{1-p}} \qquad P(X = k, X_1 = 0) = \begin{cases} 0,\ k = 1 \\ p(1-p)^{k-1},\ k > 1 \end{cases}$$

$$\Rightarrow \frac{p(1-p)^{k-1}}{1-p} = \underbrace{p(1-p)^{k-2}}_{P_{X'(k)}}$$

$$\Rightarrow X' - 1 \sim \text{ffg}(p) \Rightarrow E(X|X_1 = 1) = E(X') = 1 + E(X)$$

$$\Rightarrow E(X|X_1 = k) = \begin{cases} 1,\ k = 1, & \text{med sannolikhet } p \\ 1 + E(X),\ k = 0 & \text{med sannolikhet } 1-p \end{cases}$$

$$\Rightarrow \underbrace{E(E(X|X_1))}_{E(X)} = \sum_k kP(E(X|X_1 = k)) = 1 \cdot p + (1 + E(X))(1-p)$$

$$E(X) = p + (1-p)(1 + E(X)) = \overbrace{p + (1-p)}^{1} + (1-p)E(X)$$

$$\Rightarrow 1 = \underbrace{(1 - (1-p))}_{p} E(X) \Rightarrow E(X) = \frac{1}{p}$$

På samma sätt visas $Var(x) = \frac{1-p}{p^2}$ (använd $Var(X) = E(Var(X|X_1)) + Var(E(X|X_1)))$

Anmärkning:
Det finns något som liknar ffg som kallas för *geometrisk fördelning*

Om Y är antalet misslyckade fall inann första lyckade utfall $= X - 1 \sim Geo(p)$
Då gäller att $P(Y = n) = pq^n \quad n = 0, 1, 2, \cdots$

Väntevärdet av en geometrisk fördelning:

$$E(Y) = E(X) - 1 = \frac{1}{p} - 1 = \frac{1-p}{p}$$

$$Var(Y) = Var(X - 1) = Var(X) = \frac{1-p}{p^2}$$

Definition/Sats 10.11: Negativ binomialfördelning

$X =$ anatlet försök som krävs att få rst lyckade utfall
Detta betcknas som $X \sim NegBin(r, p)$

Om vi vill få det $r$te lyckade utfallet på det $n$te försöket, ska vi få $r-1$ lyckat utfall på $n-1$ utfall. Hur många olika sätt kan detta ske på? Jo:

$$P(X = n) = \binom{n-1}{r-1} p^{r-1} q^{(n-1)-(r-1)} p = \binom{n-1}{r-1} p^r q^{n-r}$$

För att räkna väntevärdet kan det bli lite klurigt, men man kan skriv aom X som en summa:
$X = X_1 + \cdots + X_r,\ \underbrace{X_k}_{\text{ober}} =$ antalet försök som krävs att få lyckat utfall efter $k-1$ lyckade utfall $\sim$ ffg(p)

Anmärkning:
Här dyker det upp någon slags minneslöshet

Anmärkning:
$E(X) = E(X_1 + \cdots + X_r) = \overbrace{E(X_1)}^{1/p} + \cdots \overbrace{E(X_r)}^{1/p} = \frac{r}{p},\ Var(X) = r\frac{1-p}{p^2}$
Eftersom de är oberoende så kan man summera sannolikheterna.

11. Lektion

Uppgifter från blandade problem samt uppgifter från förra lektion som inte hann med

11.1. **308.**

Givet en 2-dimensionell slumpvariabel X, Y med täthetsfunktion $f_{X,Y}(x,y) = \begin{cases} cye^{-x}, \ 0 \leq x \leq y \leq 1 \\ 0, \ \text{annars} \end{cases}$

Bestäm c, och beräkna $f_X(x)$ och $f_Y(y)$. Är X, Y oberoende?

Vi räknar först de marginella täthetsfunktionerna:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(xy)dy = \begin{cases} 0, \ \text{annars } x > 1, x < 0 \\ \int_x^1 cye^{-x}dy, \ 0 \leq x \leq 1 \end{cases}$$

$$\int_x^1 cye^{-x}dy = ce^{-x}\int_x^1 ydy = ce^{-x}\left(\frac{1}{2} - \frac{x^2}{2}\right)$$

$$\Rightarrow f_X(x) \begin{cases} c\left(\dfrac{1-x^2}{2}\right)e^{-x}, \ x \in [0,1] \\ 0, \ x \notin [0,1] \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \begin{cases} \int_0^y cye^{-x}dx, \ y \in [0,1] \\ 0, \ y \notin [0,1] \end{cases}$$

$$\int_0^y cye^{-x}dx = cy(1-e^{-y})$$

$$f_Y(y) = \begin{cases} cy(1-e^{-y}), \ y \in [0,1] \\ 0, \ y \notin [0,1] \end{cases}$$

Vi vet att vi har oberoende omm $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, vi multiplicerar f_X med f_Y och ser vad vi får:

$$f_X(x)f_Y(y) = \begin{cases} 0, \ (x,y) \notin [0,1]^2 \\ \underbrace{c^2\left(\dfrac{1-x^2}{2}\right)ye^{-x}(1-e^{-y})}_{0 \Longleftrightarrow x = \pm 1 \text{ eller } y = 0}, \ (x,y) \in [0,1]^2 \end{cases} \neq f_{X,Y}(x,y) = \begin{cases} cye^{-x}, \ 0 \leq x \leq y \leq 1 \\ 0, \ \text{annars} \end{cases}$$

Alltså är X, Y beroende.

Vi räknar konstanten c. Eftersom vi har täthetsfunktioner så måste de integreras till 1. Vi kan välja mellan att integrera f_Y eller f_X, vi väljer den lättaste dvs f_Y:

$$\int_{-\infty}^{\infty} f_Y(y)dy = 1$$

$$\int_{-\infty}^{\infty} = \int_0^1 cy(1-e^{-y})dy = c\int_0^1 y - ye^{-y}dy$$

$$= c\left(\int_0^1 ydy - \int_0^1 ye^{-y}dy\right) = c\left(\frac{1}{2} - \int_0^1 ye^{-y}dy\right)$$

$$\int_0^1 ye^{-y}dy \Rightarrow -2e^{-1} + 1 = 1 - \frac{2}{e}$$

$$\Rightarrow \int_{-\infty}^{\infty} f_Y(y)dy = c\left(\frac{1}{2} - \left(1 - \frac{2}{e}\right)\right) = c\left(\frac{2}{e} - \frac{1}{2}\right) = c\left(\frac{4-e}{2e}\right) = 1$$

$$\Rightarrow c = \frac{2e}{4-e}$$

11.2. **304.**

Vi har 4 glödlampor, 3st av typ A, 1st av typ B. Alla livslängder är oberoende.
Typ A har livslängd Exp-fördelad, med väntevärde $100h$, medan typ B har livslängd Exp-fördelad med väntevärde $200h$
Dra en slumpmässig glödlampa, vi noterar att den fungerar efter 200h. Vad är sannolikheten att vi har dragit lampa A?

Vi ska alltså betinga på att den fungerar efter $200h$, men vad är det vi ska betinga med?

Först har vi variabler $\overbrace{X_1, X_2, X_3}^{\text{typ } A} \sim Exp(\lambda_1)$ och $\overbrace{X_4}^{\text{typ } B} \sim Exp(\lambda_2)$
$E(X_1) = E(X_2) = E(X_3) = \dfrac{1}{\lambda_1} = 100 \Rightarrow \lambda_1 = \dfrac{1}{100}$
På samma sätt får vi $\lambda_2 \Rightarrow \dfrac{1}{200}$
Låt Y vara likformigt fördelad på $\{1,2,3,4\}$, med andra ord $P(Y = k) = \dfrac{1}{4}$ för $k = 1,2,3,4$
Vi definierar en variabel $Z = X_Y$ (vad som menas här är, $Z(\omega) = X_{Y(\omega)}(\omega)$)
Sannolikheten att vi har fått en A lampa kan besvaras genom:

$$P(Y \in \{1,2,3\}\,|Z \geq 200) = \frac{P(Y \in \{1,2,3\}\,\&Z \geq 200)}{P(Z \geq 200)}$$

Antag att X_1, X_2, X_3, X_4, Y är oberoende

$$Z(\omega) \geq 200\&Y(\omega) = 4 \Leftrightarrow X_4(\omega)\&Y(\omega) = 4P(Z \geq 200\&Y = 4) = P(X_4 \geq 200\&Y = 4)$$

$$= P(X_4 \geq 200)P(Y = 4) = \frac{1}{4}P(X_4 \geq 200) = \frac{1}{4}\left(1 - \underbrace{FX_4(200)}_{1 - e^{\frac{1}{200}200}}\right) = \frac{1}{4}e^{-1}$$

$$P(Z \geq 200\&Y = \{1,2,3\}) = P((\{Z \geq 200\} \cap \{Y = 1\}) \cup (\{Z \geq 200\} \cap \{Y = 2\}) \cup (\{Z \geq 200\} \cup \{Y = 3\}))$$
$$= P(Z \geq 200\&Y = 1) + P(Z \geq 200\&Y = 2) + P(Z \geq 200\&Y = 3)$$

$$= P(X_1 \geq 200\&Y = 1) + \cdots$$
$$P(X_1 \geq 200)P(Y = 1) \cdots$$
$$= \frac{1}{4}P(X_1 \geq 200) + \frac{1}{4}P(X_2 \geq 200) + \frac{1}{4}P(X_3 \geq 200) = \frac{3}{4}e^{-\frac{200}{100}} = \frac{3e^{-2}}{4}$$
$$P(Z \geq 200) = \sum_{k=1}^{4} P(Z \geq 200\&Y = k) = \sum_{k=1}^{4} P(X_k \geq 200)P(Y = k) = \frac{3e^{-2}}{4} + \frac{1}{4}e^{-1}$$
$$P(Y \in \{1,2,3\}\,|Z \geq 200) = \frac{\frac{3e^{-2}}{4}}{\frac{3e^{-2}}{4} + \frac{1}{4}e^{-1}} = \frac{3}{3+e}$$

11.3. **315.**

Antalet malariaparasiter per milliliter blod är $\begin{cases} X \sim N(3200, 1000^2) \text{ för vuxna} \\ Y \sim N(4000, 1000^2) \text{ för barn} \end{cases}$
Kraftig feber uppstår vid över 5000 per milliliter blod
Hur stor del av barnen/vuxna får kraftig feber?

Vi ska approximera lösningen genom tabellerna i boken eftersom vi inte kan exakt integrera täthetsfunktionen (s.483 tabell 4)

Vi behöver bara kolla på $N(0,1)$ eftersom $\dfrac{X-3200}{1000} \sim N(0,1) \sim \dfrac{Y-4000}{1000}$

$$P(Y > 5000) = P\left(\underbrace{\frac{Y-4000}{1000}}_{N(0,1)} > \underbrace{\frac{5000-4000}{1000}}_{=1}\right) = P\left(\frac{Y-4000}{1000} > 1\right) = 1-\Phi(1) = 1-\Phi(1.00)$$
$$= 1 - 0.8413 = 0.1587$$

Samma lösningsmetod för vuxna:

$$P(X > 5000) = P\left(\frac{X-3000}{1000} > \underbrace{\frac{5000-3200}{1000}}_{=1.8}\right) = 1-\Phi(1.8)$$
$$1 - 0.9641 = 0.0359$$

11.4. **3.11.2.**

$X_1, X_2, \cdots$ är oberoende slumpvariabler och $\sim Be(0.4)$, $Y = \sum_{i=1}^{20} X_i =$ antalet lyckade X_i.
Det följer även att $Y \sim Bin(20, 0.4)$. Medelvärdet $\overline{X_{20}} = \dfrac{1}{20}\sum_{i=1}^{20} X_i = \dfrac{Y}{20}$

Beräkna $P(\left|\overline{X_{20}} - 0.4\right| \geq 0.2)$ med hjälp av tabell 2 (s.474)
Uppskatta $P(\left|\overline{X_{20}} - 0.4\right| \geq 0.2)$ med hjälp av Chebyshevs olikhet

Det är inte X_i som är $\sim Bin$ utan Y, så vi skriver om sannolikheten:

$$P(\left|\overline{X_{20}} - 0.4\right| \geq 0.2) = P(\left|\underbrace{Y}_{\sim Bin(20,0.4)} - 8\right| \geq 4)$$
$$= P(Y \geq 12,\ Y \leq 4) = P(Y \geq 12) + \underbrace{P(Y \leq 4)}_{F_Y(4)} = (1 - P(Y < 12)) + F_Y(4)$$
$$= (1 - \underbrace{P(Y \leq 11)}_{F_Y(11)}) + F_Y(4) = 1 - F_Y(11) + F_Y(4)$$
$$1 - 0.9435 + 0.0510 = 0.0565 + 0.0510 = 0.1075$$

Chebyshevs olikhet: $P(|X - E(X)| \geq \varepsilon) \leq \dfrac{Var(X)}{\varepsilon^2}$ om $X \in L^2$

När vi ska uppskatta:

$$P\left(\left|\overline{X_{20}} - \overbrace{0.4}^{E(\overline{X_{20}})}\right| \geq 0.2\right) \leq \frac{Var(\overline{X_{20}})}{0.2} = 0.3$$

0.3 var en ganska dålig uppskattning av 0.1075, vilket kanske inte är så förvånande. Mer generellt, om man tar medelvärdet för något, så kommer det hoppa ur en $\dfrac{1}{n}$ från Var:

$$P(\left|\overline{X_n} - E(X)\right| \geq \varepsilon) \leq \frac{Var(X_i)}{n\varepsilon^2} \overset{n\to\infty}{\to} = O\left(\frac{1}{n}\right) \qquad \text{dålig konvergens, går mot 0 långsamt}$$

Ju högre L^p den ligger i desto bättre uppskattning kan vi göra

Anmärkning:
Markovs och Chebyshevs olikheter kommer på tentan!

Anmärkning:

$$Y_n \overset{P}{\to} 0 = P(|Y_n| \geq \varepsilon) \overset{n\to\infty}{\to} 0 \quad \forall \varepsilon > 0$$

12. Genererande funktioner till en slumpvariabel X

Vi vill studera fördelningen till slumpvariabler samt sannolikhetsmått. Om vi associerar sannolikhetsmått till någon funktion som gör att vi kan studera måttet så vore det noice. Det ska vi försöka göra.

Det finns 3st genererande funktioner som vi vill studera i denna kurs. Man skulle kunna säga att fördelningsfunktionen är en genererad funktion (diskontinuitet = atom, deriverbar = absolutkont. osv)

De 3 är följande:

- Momentgenererande funktionen (mgf), $\psi_X(t) = E(e^{tX})\ t \in \mathbb{R}$
- Karakteristiska funktionen, $\varphi_X(t) = E(e^{itX})\ t \in \mathbb{R}$

  Väntevärdet av en komplex slumpvariabel är följande:
  $$E(\cos(tX)) + iE(\sin(tx))$$
  Notera att Karakteristiska funktionen kanske får en att tänka på $1_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$

  Så är det ej.

- Sannolikhetsgenererade funktioner $g_X(s) = E(s^x)\ s \in \mathbb{R}_{\geq 0}$

  Notera att vi kan uttrycka momentgenererande funktioner med sannolikhetsgenererade funktioner:
  $$\psi_X(t) = g_X(e^t)$$

Anmärkning:
Om X är absolutkontinuerlig, då kommer den momentgenererande funktionen vara:

$$\psi_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx \qquad \text{(Laplacetransform)}$$

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)dx = \int \cos(tx) f_X(x)dx + i\int \sin(tx)dx \qquad \text{(Fouriertransform)}$$

$$g_X(s) = \sum_{n=0}^{\infty} P_X(n) \cdot s^n \text{ om } X \in \mathbb{N} \qquad \text{(Z-transform till } \{P_X(n)\}_{n=0}^{\infty})$$

Anmärkning:
Karakteristiska funktionen är komplexvärd, så vi kan ta absolutbelopp:

$$|\varphi_X(t)| \leq E(\left|e^{itX}\right|) = 1$$

Den karakteristiska funktionen är alltså begränsad, men komplexvärd

Anmärkning:
Den momentgenererande funktionen är reellvärd men inte alltid ändlig

Poängen med våra karakteristiska funktionen är att den ska beskriva X unikt

Definition/Sats 12.1

Om $\psi_X(t) = \psi_Y(t) = \psi(t)$, och om $\psi(t)$ är ändlig på ett intervall $(-\delta, \delta)$ för $\delta > 0$, då är $X = Y$

En intressant grej händer om vi tar n:te derivatan:

$$\psi^{(n)}(0) = E(X^n) \qquad n\text{:te momentet}$$

Notera att dessa inte beskriver unikt fördelningen till en slumpvariabel

Anmärkning:
Väntevärdet och variansen bestämmer inte fördelningen unikt

Definition/Sats 12.2

Om $\varphi_X(t) = \varphi_Y(t)$ $\forall t \in \mathbb{R}$, så kommer:

$$\Rightarrow F_X = F_Y$$

Anmärkning:
Karakteristiska är alltid ändlig och bestämmer alltid unbikt fördelningen. Den är däremot inte alltid deriverbar (eller n gånger deriverbar).

Säg att vi deriverar den momentgenererande funktionen:

$$\psi^{(k)}(0) = \frac{d^k}{dt^k}|_{t=0} \int e^{tx} f_X(x)dx$$

$$\text{Om vi kan byta plats} = \int x^k e^{tx} f_X(x)dx|_{t=0} = \int x^k f_X(x)dx = E(X^k)$$

12.1. **Egenskaper för mgf.**

Det är 2 egenskaper vi vill gå igenom:

- $\psi_{aX+b}(t) = E(e^{t(aX+b)}) = E(e^{taX}e^{tb}) = e^{tb}E(e^{taX}) = e^{tb}\psi(at)$
- X, Y oberoende, $\psi_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX}e^{tY}) \overset{ober.}{=} E(E^{tX})E(e^{tY}) = \psi_X(t)\psi_Y(t)$

Exempel:
Låt $X \sim N(0,1)$. Då är $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. Då kommer den momentgenererande funktionen till X vara:

$$E(e^{tX}) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} e^{tx}e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t^2/2}e^{-t^2/2+tx-x^2/2}dx$$

$$= \frac{e^{t^2/2}}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} \underbrace{e^{-\frac{(x-t)^2}{2}}}_{N(t,1)} dx}_{\sqrt{2\pi}} = e^{t^2/2}$$

$$\psi_X(t) = e^{t^2/2} \text{ om } X \sim N(0,1)$$

Vad händer om vi nu säger att $X \sim N(\mu, \sigma^2)$?

$$\Rightarrow \underbrace{\frac{X-\mu}{\sigma}}_{Y} \sim N(0,1)$$

$$\sigma Y + \mu = X \sim N(\mu, \sigma^2) \Rightarrow \psi_X(t) = \psi_{\sigma Y + \mu}(t) = e^{\mu t}e^{\frac{(\sigma t)^2}{2}} = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Definition/Sats 12.3

Om $X \sim N(\mu_1, \sigma_1^2)$ och $Y \sim N(\mu_2, \sigma_2^2)$, så är $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Bevis 12.1

Vi kollar på mgf för X, Y:

$$\begin{gathered}
\psi_{X+Y} = \psi_X(t)\psi_Y(t) = e^{\mu_1 t + (\sigma_1^2 t^2)/2} e^{\mu_2 t + (\sigma_2^2 t^2)/2} \\
= e^{(\mu_1+\mu_2)t + (\sigma_1^2+\sigma_2^2)t^2/2} = \psi_Z(t) \qquad Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \\
\Rightarrow F_{X+Y} = F_Z
\end{gathered}$$

Detta kunde göras utan faltning $\square$

Kuriosa:
Sannolikhetsgenererade funktionen heter så för att om vi deriverar sannolikhetsgenererade funktionen i 0 för en diskret slumpvariabel, så kommer första derivatan i 0 vara sannolikheten att $X = 1$, andra derivatan att $X = 2$ osv... Den kommer spotta ut sannolikhetsfunktionen, varpå namnet kommer.

13. Konvergens av slumpvariabler & centrala gränsvärdessatsen

13.1. Konvergens av slumpvariabler.

Vad menas? Massa olika grejer, det är inte så tydligt här vad som menas, exempelvis så kanske man pratar om *konvergens nästan överallt*. Då menar man:

$$\underbrace{P(X_n \to X)}_{P(\{\omega | X_n(\omega) \to X(\omega)\})} = 1$$

$$(X_n) \overset{a.s}{\to}$$

Punktvis konvergens, konvergens nästan överallt är den vanligaste typen av konvergens

Det finns en annan typ som heter *konvergens i sannolikhet* som är något vi förhoppningsvis känner igen:

$$P(|X_n - X| \geq \varepsilon) \overset{n\to\infty}{\to} 0 \quad \forall \varepsilon > 0$$

Ett exempel på konvergens i sannolikhet är *stora talens lag*, som säger att sannolikheten att medelvärdet konvergerar mot väntevärdet 0. Detta visade vi med Chebyshevs olikhet (när de var parvis okorrelerade). Den kallas därmed för den svaga stora talens lag, det finns en starkare variant som säger att medelvärdet konvergerar mot väntevärdet nästan överallt. (Visade detta i specialfall då vi kunde använda kovarians)

Det finns ett annat begrepp som kallas för *konvergens i medelvärde*: Då söker man efter att väntevärdet att $|X_n - X|^p \to 0$ när $n \to \infty$:

$$E\,|X_n - X|^p \overset{n\to\infty}{\to} 0$$

$$X_n \overset{L^p}{\to} X$$

Definition/Sats 13.1

En intressant grej som går fort att visa är att om vi har konvergens i medelvärdet så implicerar detta konvergens i sannolikhet.

Bevis 13.1

Kom ihåg Markovs olikhet:

$$\underbrace{P(|X| \geq t)}_{P(|X|^p) \geq t^p} \leq \frac{E(|X|)}{t}$$

$$\Rightarrow P(|X|^p \geq t^p) \leq \frac{E(|X|^p)}{t^p}$$

$$P(|X_n - X| \geq \varepsilon) \leq \frac{E(|X_n - X|^p)}{\varepsilon^p} \overset{n\to\infty}{\to} 0$$

□

Det finns fler konvergensbegrepp, såsom *konvergens i fördelning* (kanske den mest naturliga typen, för vi använder slumpvariabler för att räkna sannolikheter. Konvergens i fördelning hanterar precis detta)
Mer generellt säger den att fördelningsfunktionen till X_n går mot fördelningsfunktionen till X:

$$\underbrace{F_{X_n}(t)}_{P(X_n \leq t) \to P(X \leq t)} \to F_X(t) \qquad (X_n \overset{d}{\to} X)$$

Konvergensen gäller inte för alla t, detta gäller exempelvis om F_X inte är kontinuerlig. Vi är dock mest intresserade av normalfördelning, som är kontinuerlig, så vi behöver inte oroa oss alltför mycket.
Från detta följer det att:

$$\underbrace{P(a < X_N \leq b)}_{F_{X_n}(b) - F_{X_n}(a)} \overset{n\to\infty}{\to} \underbrace{P(a < X \leq b)}_{F_X(b) - F_X(a)}$$

Ytterliggare ett till konvergensbegrepp är *konvergens av momentgenererande funktioner*. Detta är precis som det låter, om vi tar den momentgenererande funktione av X_n så konvergerar den mot den momentgenererande funktionen till X:

$$\psi_{X_n}(t) \to \psi_X(t), \qquad t \in (-\delta, \delta), \quad \delta > 0$$

Definition/Sats 13.2

(Detta är en svår sats)

Om $\psi_{X_n} \to \psi_X$ så $\Rightarrow$ konvergens i fördelning ($X_n \xrightarrow{d} X$)

Konvergens av karakteristisk funktion är som det låter. Om vi tar en karakteristisk funktion för X_n och den konvergerar för X, så konvergerar den karakteristiska funktionen:

$$\varphi_{X_n}(t) \overset{n \to \infty}{\to} \varphi_X(t) \qquad \forall t \in \mathbb{R}$$

Ur denna följder det en sats:

Definition/Sats 13.3: Levys sats

Om $\varphi_{X_n} \to \varphi_X \Rightarrow F_{X_n} \to F_x$

Det finns *likformig konvergens* som ger konvergens i medelvärdet, det kommer inte gås igenom.

Definition/Sats 13.4: Centrala gränsvärdessatsen

Låt:

- $E(X_i) = \mu \in \mathbb{R}$
- $D(X_i) = \sigma > 0$

Om $X_1, X_2, \cdots$ är oberoende och lika fördelade slumpvariabler (i.i.d.r.v), och den momentgenererande funktionen mgf till X_i är ändlig på något intervall $(-\delta, \delta)$. Då kommer:

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

$$\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} X \sim N(0,1)$$

$$\Rightarrow P\left(a \leq \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}}\right) \overset{n \to \infty}{\to} \Phi(b) - \Phi(a)$$

Anmärkning:
Detta kommer på tentan!

Beviside:
Visa att $\psi_{\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}}} \to \psi_{N(0,1)}$

Definition/Sats 13.5

Om $X_1, X_2 \cdots$ är oberoende och likafördelade slumpvariabler och mgf är ändlig på $(-\delta, \delta)$, så har vi att:

$$\overline{X_n} \overset{d}{\mu}= E(X_1)$$

Påminn om vad den momentgenererande funktionen är samt vad derivatan var (momenten):

$$\psi_X''(0) = E(X^2) = \sigma^2 + \mu^2$$

$$\Rightarrow \psi_X(t) = 1 + \mu t + (\sigma^2 + \mu^2)\frac{t^2}{2!} + O(t^3)$$

Bevis 13.2

Bevis av föregående sats:

$$\psi_{\overline{X_n}}(t) = \psi_{\frac{1}{n}(X_1+\cdots+X_n)}(t) = \psi_{X_1+\cdots+X_n}(\frac{t}{n})$$

$$\Rightarrow \psi_{X_1}(\frac{t}{n})\cdots\psi_{X_n}(\frac{t}{n}) = \psi_{X_1}(\frac{t}{n})^n = (1 + \mu\frac{t}{n} + O(\frac{t^2}{n^2}))^n$$

$$\overset{n\to\infty}{\to} e^{\mu t} = \psi_\mu(t)$$

$$\Rightarrow \overline{X_n} \overset{d}{\to} \mu$$

$\square$

Uppgift 325:
1000 hushåll, sannolikheten för att ett visst hushåll är 0.6, 0 bilar är 0.3, och 2 bilar är 0.1.
Hur många parkeringsplatser ska minst planeras om sannolikheten för att alla bilar ska få plats är minst 0.9?

Lösning:
Varje hushåll har en fördelning för antalet bilar
En slumpvariabel för varje hushåll: $X_1, \cdots, X_{1000}$ är oberoende och likafördelade.
Sannolikhetsfunktionen:

$$P_{X_i}(0) = 0.3, \quad P_{X_i}(1) = 0.6, \quad P_{X_i}(2) = 0.1$$

$$X = \sum_{i=1}^{1000} X_i = \text{antalet bilar}$$

Vi vill lösa $P(X \leq N)$ där N är antal parkeringsplatser vi söker.

$$P(X \leq N) \geq 0.9 \Leftrightarrow P(X > N) \leq 0.1$$

Vi vill hitta N, vilket vi kan göra genom normalapproximation med något som heter *halvkorrektion.*
Halvkorrektion går ut på att, om vi vill approximera $P(X \leq N)$ (där X är diskret och heltalsvärd) så är $P(X \leq N) = P(X \leq N + 0.9) = P(X < N + 1) = P(X \leq N + 0.5)$
Vi byter till "mittpunkten" av intervallet av de tal som inte påverkar sannolikheten om man adderar (i vårat fall $[0, 1] = 0.5$)
Vi väljer då att korrigera $P(X \leq N + 0.5)$. Använd alltid halvkorrektion om det är diskreta variabler!

Vi vill approximera $P(X \leq N + 0.5)$. Då måste vi hitta väntevärdet:

$$0 \cdot 0.3 + 1 \cdot 0.6 + 2 \cdot 0.1 = 0.8 = E(X_i)$$

$$Var(X_i) = 0^2 \cdot 0.3 + 1^1 \cdot 0.6 + 2^2 \cdot 0.1 - 0.8^2 = 0.36$$

$$D(X_i) = \sqrt{(Var(X_i))} = 0.6$$

$$P\left(\underbrace{\frac{\frac{X}{1000} - 0.8}{0.6/\sqrt{1000}}}_{\approx N(0,1)} > \frac{\frac{N+0.5}{1000} - 0.8}{0.6/\sqrt{1000}}\right) \leq 0.1$$

$$\approx 1 - \Phi\left(\frac{\frac{N+0.5}{1000} - 0.8}{0.6\sqrt{1000}}\right)$$

$$P(Y > \lambda_{0.1}) = 0.1 \Rightarrow \lambda_{0.1} \approx 1.2816$$

$$\Rightarrow \frac{\frac{N+0.5}{1000} - 0.8}{0.6\sqrt{1000}} = 1.2816$$

$$\Rightarrow N = \left(\frac{1.2816 \cdot 0.6}{\sqrt{1000}} 00.8\right) \cdot 1000 - 0.5 \approx 823.82 \approx 824$$

Övning:
Om vi är snåla och tar 810 parkeringsplatser istället för 824, vad är då sannolikheten?

$$P\left(\frac{\frac{X}{1000} - 0.8}{0.6\sqrt{1000}} \leq \underbrace{\frac{\frac{810.5}{1000} - 0.8}{0.6\sqrt{1000}}}_{\approx 0.55}\right) \approx 0.7088 \approx 71\%$$

Detta är sannolikheten att vi har mindre bilar än parkeringsplatser.

Definition/Sats 13.6: Weirestrass approximationssats

För en kontinuerlig funktion $f : [0,1] \to \mathbb{R}$ (man kan låta den gå från A om man vill, men då måste A vara slutet), så finns en följd polynom p_n så att $p_n \to f$ likformigt.

$$\sup_{x\in[0,1]} |p_n(x) - f(x)| \overset{n\to\infty}{\to} 0$$

Beviside: [Bernstein, 1912]:
Går ut på att tag slumpvariabler $Y_1, Y_2, \cdots \sim Be(x)$. Dessa ska vara oberoende. Vi tar medelvärdet:

$$E(f(\overline{Y_n})), \text{ förväntas (enligt stora talens lag) } \to \underbrace{f(x)}_{E(Y_i)}$$

$$E(f(\overline{Y_n})) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) P_{\overline{Y_n}}(k) = \underbrace{\sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}}_{\text{Bernsteinpolynom}}$$

14. Anmärkning om formelsamling

På sida 4 står det approximationer med massa bubblor. Vi gör om den lite:

15.1. **Uppgift 1.**

Sannolikheten att alla har samma färg är kombinatorik. Antingen har vi plockat 4st vita, 4st blå eller 4st gröna. På hur många sätt kan man välja 4st gröna? Jo $\binom{8}{4}$st sätt att välja 4 gröna, $\binom{6}{4}$ 4st blå, osv osv.
Totalt har vi $\binom{8}{4}+\binom{6}{4}+\binom{4}{4}$ och totalt finns det 18st val får vi att sannolikheten är:

$$\frac{\binom{8}{4}+\binom{6}{4}+\binom{4}{4}}{\binom{18}{4}}$$

Sannolikheten att det finns minst en boll av alla färger bland de 4 plockade. Detta kan löses kombinatoriskt, eller så kan vi lösa det sannolikhetsteoretiskt.
Låt $A =$ minst en grön boll, $B =$ minst en blå boll, $C =$ minst en vit
Vi söker då $P(A \cap B \cap C)$. Notera att $P(A \cup B \cup C) = 1$.

$P(A \cup B) = 1 - P(\text{alla vita}) = 1 - \dfrac{\binom{4}{4}}{\binom{18}{4}}$

$P(A \cup C) = 1 - P(\text{alla blå}) = 1 - \dfrac{\binom{6}{4}}{\binom{18}{4}}$

$P(B \cup C) = 1 - \dfrac{\binom{8}{4}}{\binom{18}{4}}$

$P(A) = 1 - P(\text{alla blå och vit}) = 1 - \dfrac{\binom{10}{4}}{\binom{18}{4}}$

$P(B) = 1 - \dfrac{\binom{12}{4}}{\binom{18}{4}}$

$P(C) = 1 - \dfrac{\binom{14}{4}}{\binom{18}{4}}$

$\overbrace{P(A \cup B \cup C)}^{=1} = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap) + P(A \cap B \cap C)$

$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 1 - \dfrac{\binom{10}{4}}{\binom{18}{4}} + 1 - \dfrac{\binom{12}{4}}{\binom{18}{4}} - 1 + \dfrac{\binom{4}{4}}{\binom{18}{4}}$

$P(A \cap C)$ och $P(B \cap C)$ räknas ut på samma sätt.
Lös ut $P(A \cap B \cap C)$ och stoppa in värdena.

15.2. **Uppgift 2.**

Axiomen är triviala. Att F inträffar när E inträffar, så betyder det att $E \subseteq F$

$F = \underbrace{E \cup (F \backslash E)}_{\text{disjunkta}}$, då kan vi summera sannolikheterna:

$$P(F) = P(E) + \underbrace{P(F \backslash E)}_{\geq 0} \geq P(E) + 0 = P(E)$$

15.3. **Uppgift 3.**

$X \sim Bin(n, p) = E(X) = np, Var(X) = np(1-p)$ (finns i formelsamlingen)

$$E(Z) = 2E(X) + E(Y) = 2 \cdot 5 \cdot 0.1 + 10 \cdot 0.2 = 3$$

$$Var(Z) \overset{ober.}{=} Var(2X) + Var(Y) = 4Var(X) + Var(Y) = 4 \cdot 5 \cdot 0.1 \cdot (1 - 0.1) + 10 \cdot 0.2 \cdot 0.8 = 3.4$$

För att bestämma sannolikheten att $Z = 2$ kollar vi på vad det betyder att $Z = 2$:

$$Z = 2 \Leftrightarrow \{(0, 2), (1, 0)\}$$

Detta är disjunkta händelser:

$$P(X = 0 \& Y = 2) + P(X = 1 \& Y = 0) = P(X = 0)P(Y = 2) + P(X = 1)P(Y = 0)$$

$$0.9^5 \cdot \binom{10}{2} 0.2^2 (1 - 0.2)^{10-2} + 5 \cdot 0.1 \cdot 0.9^4 \cdot 0.8^{10}$$

15.4. **Uppgift 4.**

Sannolikheten att en slumpmässig planta ger en grön avkomma.
Vi vet att $P(A) = 1/9$, $P(B) = P(C) = 4/9$ samt
$P(\text{Grön}|A) = 1$, $P(\text{Grön}|B) = 3/4$
$P(\text{Grön}|C) = 9/16$
Vi använder lagen om total sannolikhet för att hitta $P(\text{Grön})$:

$$P(\text{Grön}) = P(\text{Grön}|A)P(A) + P(\text{Grön}|B)P(B) + P(\text{Grön}|C)P(C)$$
$$= \frac{25}{36}$$

Om avkomman är grön, vad är då sannolikheten att den kommer från A, B, C?
Vi söker alltså $P(A|\text{Grön})$ osv. Vi kan vända på betigningen:

$$P(A|\text{Grön}) = \frac{P(\text{Grön}|A)P(A)}{P(\text{Grön})}$$
$$= \frac{1 \cdot 1/9}{25/36} = \frac{4}{25}$$

15.5. **Uppgift 5.**

Centrala gränsvärdessatsen för att visa:

$$\lim_{n \to \infty} e^{-n} \sum_{k=1}^{n} \frac{n^k}{k!} = 0.5$$

Notera, detta ser lite ut som:

$$X_n \sim Po(n) \Rightarrow P(X_n = k) = \frac{e^{-n} n^k}{k!}$$

Det är som att vi summerar $Po(n)$, vi får då:

$$\lim_{n \to \infty} e^{-n} \sum_{k=1}^{n} \frac{n^k}{k!} = P(1 \leq X_n \leq n)$$

Om $Y_1, Y_2, \cdots \sim Po(1)$ (och oberoende), då kommer:

$$\sum_{k=1}^{n} Y_k \sim Po(n)$$

Om vi byter ut $k = 1$ mot 0:

$$\underbrace{e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}}_{f_X \text{ för } X\sim Po(n)} = P(\overbrace{Y_1+\cdots+Y_n}^{n\cdot\overline{Y_n}} \le n)$$

För en poisson n fördlening gäller att $E(X_n) = n$ och $Var(X_n) = n$:

$$\begin{aligned}
&= P(\overline{Y_n} \le 1) = P(\underbrace{\overline{Y_n} - 1}_{E(\overline{Y_n})} \le 0)\\
&= P\left(\frac{\overline{Y_n} - n}{D(\overline{Y_n}) \le 0}\right)\\
&\overset{CGS}{\Rightarrow} \Phi(0) = 0.5\\
&\lim_{n\to\infty} e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} = 0.5\\
e^{-n}\sum_{k=1}^{n}\frac{n^k}{k!} &= \left(e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}\right) - e^n \to 0.5
\end{aligned}$$

15.6. **Uppgift 7.**

Lösning till a:

$$\begin{aligned}
\psi_{X_1+X_2}(t) &\overset{ober.}{=} \psi_{X_1}(t)\psi_{X_2}(t) = (1-2t)^{-n_1/2}(1-2t)^{-n_2/2}\\
&= (1-2t)^{-(n_1+n_2)/2}
\end{aligned}$$

Kom ihåg att $E(X^k) = \psi^{(k)})(0)$, vi får:

$$\begin{aligned}
E(X) &= \frac{d}{dt}(1-2t)^{-n/2}|_{t=0} = \left(\frac{-n}{2}\right)(1-2t)^{-\frac{n}{2}-1}\cdot(-2)\\
&= n(1-2\cdot 0)^{\frac{-n}{2}-1} = n\\
E(X^2) &= \psi''_X(0) = \frac{d}{dt}n(1-2)^{-n/2-1}|_{t=0}\\
&\Rightarrow n(n+2)\\
Var(X) &= E(X^2) - (E(X))^2 = n(n+2) - n^2 = n^2 + 2n - n^2 = 2n
\end{aligned}$$

15.7. **Uppgift 8.**

Finns 2 metoder, man kan räkna ut fördelningsfunktionen och ta gränsvärdet och försöka lösa det, eller så tar man gränsvärdet av momentgenererande funktionen och försöker identifiera den med en fördelning

Vi vill hitta $\lim_{n\to\infty} F_{X_n}(t) = \lim_{n\to\infty} P(X_n \le t)$:

$$t \in \mathbb{N} P(X_n \le t) = 1 - P(X_n > t) = 1 - \left(1 - \frac{\lambda}{\lambda + n}\right)^{t+1}$$

$$P(X_n \le t) = 0,\ t < 0$$

$$t > 0 \Rightarrow P(X_n/n \le t) = 1 - \left(1 - \frac{\lambda}{\lambda + n}\right)^{\lfloor t \rfloor + 1}$$

$$P\left(\frac{X_n}{n} \le t\right) = 1 - \left(1 - \frac{\lambda}{\lambda + n}\right)^{\lfloor nt \rfloor + 1}$$

$$\Rightarrow \lim_{n\to\infty} \left(1 - \frac{\lambda}{\lambda + n}\right)^{\lfloor nt \rfloor + 1}$$

$$\Rightarrow \lim_{m\to\infty} \left(1 - \frac{\lambda}{m}\right)^{mt - m\lambda} \qquad m = \lambda + n$$

$$= \lim_{m\to\infty} \underbrace{\left(1 - \frac{\lambda}{m}\right)^{mt}}_{e^{-\lambda t}} \cdot \underbrace{\left(1 - \frac{\lambda}{m}\right)^{-\lambda t}}_{1}$$

Anmärkning:
$\lfloor t \rfloor$ = största heltal $\le t$