

UPPSALA UNIVERSITET

FÖRELÄSNINGSANTECKNINGAR

Sannolikhetssteori 1

Rami Abou Zahra

Inlämningsdatum
October 4, 2022

CONTENTS

1. Repetition - (K2.1)	2
1.1. Mängdlära	2
1.2. Begrepp	2
2. Regler för sannolikheter - (K2.2)	3
2.1. Kolmogorovs Axiom	3
2.2. A^c	5
2.3. $B-A$	5
3. Tolkning av sannolikheter	6
3.1. Sannolikhetsmåttet P	6
4. Betingade sannolikheten $P(A B)$	8
4.1. Oberoende utsagor	9
5. Sammanfattning K2	13
5.1. Komplement och additionssatsen	13
5.2. Sannolikhet på utfallsrum	13
5.3. Betingning	14
5.4. Oberoende	14
5.5. Lagen om total sannolikhet	16
6. Slumpvariabler	17
6.1. Viktiga slumpvariabler	20
7. Sammanfattning K3	23
7.1. Definition av Slumpvariabel	23
7.2. Fördelningsfunktioner	24
8. Medelvärde	26
8.1. Egenskaper för väntevärden	30
8.2. Kovarians	34
8.3. Mer om kontinuerliga sannoliketsrum	36
9. Lektion 21/9	40
10. Kort introduktion till mätteori	42
10.1. Minneslösa diskreta variabler	51

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 = (Ω, 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 specificerad 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 fuffigt med denna definition är att det blir 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 $P : 2^\Omega \rightarrow \mathbb{R}$ som uppfyller:

- $P(A) \geq 0 \quad \forall A \in 2^\Omega$
- $P(\Omega) = 1$
- $P(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty P(A_i)$ 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:

$$\Omega = \{kr, kl\} \times \{kr, kl\} = \{(kr, kr), (kr, kl), (kl, kr), (kl, kl)\}$$

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

Exempel:

Singla slant n gånger: $\Omega = \{kr, kl\}^n$

$$P(\{x\}) = \frac{1}{2^n} \quad \forall x \in \Omega, \quad P(A) = \sum_{x \in A} \frac{1}{2^n}$$

$$P(\text{exakt } k \text{ krona}) = \sum_{xx \text{ innehåller } k \text{ kronor}} \frac{1}{2^n} = \binom{n}{k} \left(\frac{1}{2^n}\right)$$

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 $P(\{6\}) = \frac{1}{3}$

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

Exempel:

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

Eftersom $\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$ 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 slingan"

Exempel:

Vad är sannolikheten att tärningen hamnar på en sexa på n :te slingan?

Jo, $P(\{x\}) = \underbrace{\left(\frac{5}{6}\right)^{n-1}}_{\text{alla andra siffror}} \cdot \frac{1}{6}$

Exempel:

Slumpa ett reellt tal mellan 0 och 1:

$\Omega = (0, 1) \subseteq \mathbb{R}$, då är $P(A) = \text{längden på intervallet } A = 1$

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

Vill man räkna ut unionen av sannolikheten summerar man sannolikheterna:

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

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

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

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

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

Exempel:

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

Vi sätter $P(A) = \int_A f(x) dx$

Exempel:

Tag enhetskvadraten $\Omega = [0, 1]^2$, $P(A) = \text{arean}$. 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:

$$\begin{aligned} \exists A \subseteq \Omega : \\ \sum_{x \in A} P(\{x\}) = 1 \end{aligned}$$

Definition/Sats 2.2: Kontinuerliga Sannolikhetsrum

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

2.2. A^c .

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

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

2.3. $B-A$.

$$x \in B \setminus A \Rightarrow x \in B \wedge x \notin A$$

$$\Rightarrow x \in B \wedge x \in A^c$$

$$x \in B \cap A^c$$

$$P(B) = P(A \cap B) + P(B \setminus A) \Rightarrow P(B \setminus A) = P(B) - P(A \cap B)$$

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

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

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

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:

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

□

Definition/Sats 3.2

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

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

Lemma 3.1: De Morgans lagar

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

Bevis 3.2: Bevis av Lemma

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

□

Bevis 3.3: Bevis av sats

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

$$\begin{aligned} P\left(\bigcap_{i=1}^{\infty} A_i\right) &= 1 - P\left(\left(\bigcap_{i=1}^{\infty} A_i\right)^c\right) = 1 - P\left(\bigcup_{i=1}^{\infty} A_i^c\right) \\ &\Rightarrow 1 - \lim_{n \rightarrow \infty} P(A_i^c) = \lim_{n \rightarrow \infty} (1 - P(A_i^c)) = \lim_{n \rightarrow \infty} P(A_n) \end{aligned}$$

□

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 $\Omega = \{1, 2, 3, 4, \dots\}$, $P(\{n\}) = \frac{1}{2^n}$

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 $= \{1\}^c = \{2, 3, 4, 5, \dots\}$

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 $P(2|B) = \frac{1}{2}$ 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 $Q : 2^B \rightarrow \mathbb{R}$ (för något $B \in \Omega$) och $Q(A) = \frac{P(A)}{P(B)}$ = betingade sannolikheten

Mer generellt kan vi definiera $Q : 2^\Omega \rightarrow \mathbb{R}$ genom $Q(A) = \frac{P(A \cap B)}{P(B)}$ (med andra ord, den betingade sannolikheten)

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

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

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

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, \dots disjunkta. Då är $B \cap A_1, B \cap A_2, \dots$ också disjunkta.

Vi vill räkna följande:

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

Notera:

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

Vi får då:

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

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 $A \cap B \subseteq A_2 \cap B \subseteq A_2 \cap B \subseteq \dots$ så gäller $P(\bigcup_{i=1}^{\infty} A_i|B) = \lim_{n \rightarrow \infty} P(A_n|B)$

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.

Ytterligare 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:

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

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

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

$\Rightarrow A$ och B är oberoende

Exempel:

Låt Ω = Uppsalas vuxna befolkning.

Låt A = {Man} B = {Bruna ögon} C = {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 $P(A) > 0, P(B) > 0, P(A^c) > 0, P(B^c) > 0$. $Q(A) = P(A|B)$ är ett sannolikhetsmått, då gäller:

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

Det vill säga:

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

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 händer 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:

$$\begin{aligned} A \subseteq A \cup B &\Rightarrow 1 = P(A) \leq P(A \cup B) \leq 1 \Rightarrow P(A \cup B) = 1 \\ P(A \cup B) &= \underbrace{P(A)}_1 + \underbrace{P(B)}_1 - P(A \cap B) \\ P(B) - P(A \cap B) &= 0 \Leftrightarrow P(A \cap B) = P(B) = P(B)P(A) \end{aligned}$$

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

Definition/Sats 4.6: Oberoende i flera variabler

$S \subseteq 2^\Omega$ är oberoende om $A_1, \dots, A_n \in S \Rightarrow P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot \dots \cdot P(A_n)$

Exempel:

Säg att vi har en mängd $\{A, B, C\}$, mängden är oberoende 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 viktig, ty om vi antar de 3 andra likheterna (parvis oberoende) är helt annat än full oberoende.

Exempel:

Låt $\Omega = \{1, 2, 3, 4\}$, $P(n) = \frac{1}{4}$ samt $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{2, 3\}$

Först och främst, $P(A) = \frac{1}{2} = P(B) = P(C)$

$P(A \cap B) = \frac{1}{4} = P(B \cap C) = P(A \cap C) \Rightarrow$ parvis oberoende

Om vi kollar sista grejen man måste kolla för oberoende, $P(A \cap B \cap C) = P(\emptyset) = 0$, men $P(A)P(B)P(C) = \frac{1}{8} \neq 0$, 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 $1, \dots, n$ 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)
- $P(A_{n+1}|A_n) = \frac{365-n}{365}$ 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 $P(A_{n+1}|A_n) = \frac{P(A_{n+1})}{P(A_n)}$

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

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

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

Exempel:

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

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

Strategin här går ut på att skriva om uppgiften i matte-termer.

Ω = klassen, A = de som gjorde inlämningsuppgifterna

B = de som inte gjorde inlämningsuppgifterna = A^c

C = de som klarade tentamen

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

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:

$$P(A|C) = \frac{P(C|A)P(A)}{P(C)} = \frac{0.9 \cdot 0.8}{0.86} \approx 0.837$$

Man kan tänka på det på följande sätt:

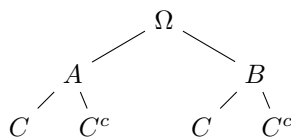


FIGURE 1.

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!

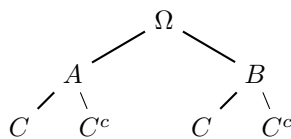


FIGURE 2.

Definition/Sats 4.7: Lagen om total sannolikhet

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

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 $P : \Omega \rightarrow \mathbb{R}$, 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 definierar en funktion $Q(A) = P(A|C) = \frac{P(A \cap C)}{P(C)}$ 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 $Q(A \cup B) = Q(A) + Q(C)$

$$\begin{aligned} Q(A \cup B) &= P(A \cup B|C) = \frac{P(\overbrace{((A \cup B) \cap C)}^{= P((A \cap C) \cup (B \cap C))})}{P(C)} \\ &= \frac{P((A \cap C) \cup (B \cap C))}{P(C)} \stackrel{add.}{=} \frac{P(A \cap C) + P(B \cap C) - \overbrace{P(A \cap B)}^{=0}}{P(C)} \\ &= Q(A) + Q(C) \end{aligned}$$

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 $P(A \cap B \cap C)$:

$$\begin{aligned} P(A \cap B \cap C) &= P(\underbrace{(A \cap B)}_{=Q} \cap C) = P(C \cap \underbrace{(A \cap B)}_{=Q}) \\ \Rightarrow P(C|Q) &= \frac{P(C \cap Q)}{P(Q)} = P(C|A \cap B) = \frac{P(C \cap A \cap B)}{P(A \cap B)} \\ \Rightarrow P(Q|C) &= \frac{P(Q \cap C)}{P(C)} = P(A \cap B|C) = \frac{P(A \cap B \cap C)}{P(C)} \\ \Rightarrow P(C|A \cap B) &= P(A \cap B|C) \end{aligned}$$

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:

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

Det är faktiskt så vi definierar oberoende:

Definition/Sats 5.2: Oberoende händelser

Två händelser A och B sägs vara *oberoende* om:

$$P(A|B) = P(A) \text{ förutsatt att } P(B) > 0$$

$$P(B|A) = P(B) \text{ förutsatt att } P(A) > 0$$

Anmärkning:

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

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

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

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

Svagare oberoende:

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 $\{A_1, \dots\}$ 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 $\{A_{i_1}, \dots, A_{i_k}\}$ med $i_1 < \dots < i_k$, gäller att $P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$

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 vara 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 oberoende).

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 $(1, 1), (1, 2), \dots, (6, 6)$. 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 $X : \Omega \rightarrow \mathbb{R}$. 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 $X = \text{längd}$, och ta en annan slumpvariabel $Y = \text{skostorlek}$, och sist men inte minst $Z = \text{ålder}$

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 $\Omega = \{H, T\}^n$. Detta är ett exempel på en klassisk sannolikhet, det vill säga $P(\omega) = \frac{1}{2^n} \quad \forall \omega \in \Omega$

Vi kan då definiera en slumpvariabel $X = \text{antalet krona (heads)}$, då kanske det hade sett ut på följande sätt om $n = 3$ och funktionen på följden hade sett ut på följande:

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

En annan slumpvariabel vi kan skapa är $Y = \text{antalet klave} = n - X$

En annan slumpvariabel vi kan skapa är följande:

$$X_1 = \left. \begin{array}{l} 1, \text{ första slanten hamnar på krona} \\ 0, \text{ annars} \end{array} \right\}$$

$$X_i = \left. \begin{array}{l} 1, \omega_i = H \\ 0, \omega_i = T \end{array} \right\}$$

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

En grej slumpvariabler är bra till är att beskriva händelser.

Exempel: Samma sannolikhetsrum och $X : \Omega \rightarrow \mathbb{R}$

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

Ett annat exempel vi kan ta är $\{\omega : X(\omega) \geq 2\} = \text{Minst 2 krona}$. 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 $P(X < 2) = P(\{X = 0\} \cup \{X = 1\}) = P(X = 0) + P(X = 1) = \frac{1}{2^n} + \frac{n}{2^n}$

Vi får då $\Rightarrow P(X \geq 2) = 1 - \frac{1}{2^n} - \frac{n}{2^n}$

Inga konstigheter, bara lite kombinatorik, hävdar föreläsaren.

Vi kan generalisera begreppet slumpvariabler:

Definition/Sats 6.2

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

Isåfall kan vi skriva $X = (X_1, \dots, X_n)$ 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 sannolikhetsrum och samma definition av X_i . Då är X_1, \dots, X_n en n -dimensionell slumpvariabel

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

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

Skriver vi $X^{-1}(A)$ ($A \in \mathbb{R}$) 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 följande:

- $P(X^{-1}(A)) \geq 0 \quad \forall A \subseteq \mathbb{R}$ (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 $A_1, A_2, \dots \subseteq \mathbb{R}$. 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, \dots är disjunkta, så måste vi kolla att följande gäller:

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

Definition/Sats 6.3: Diskreta slumpvariabler

Vi säger att en slumpvariabel $X : \Omega \rightarrow \mathbb{R}^n$ är en diskret slumpvariabel om sannolikhetsmättet $P \circ X^{-1}$ 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 $\{x_1, x_2, \dots\} \subseteq \mathbb{R}^n$ så att $P(x \in \{x_1, x_2, \dots\}) = 1$

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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ så att:

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

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 $= \frac{1}{2}$ som är samma sak som $P(X = 0)$, alltså gäller följande:

$$P_{X_i}(X) = \begin{cases} 1/2, & x = 0 \text{ eller } x = 1 \\ 0 & \text{annars} \end{cases}$$

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

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

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

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:

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

Vi antar att $\{x_1, \dots, x_n\} \in \{0, 1\}$.

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

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

Definition/Sats 6.6: Oberoende slumpvariabler

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

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

$$P(X_{i_1} \in A_{i_1}, \dots, X_{i_m} \in A_{i_m}, X_{j_1} \in \mathbb{R}, \dots, X_{j_{n-m}} \in \mathbb{R})$$

$$\Rightarrow P(X_{i_1} \in A_{i_1}) \cdots P(X_{i_m} \in A_{i_m}) \underbrace{P(X_{j_1} \in \mathbb{R})}_{=1} \cdots \underbrace{P(X_{j_{n-m}} \in \mathbb{R})}_{=1}$$

$$\text{Antag } P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n) \quad (2)$$

Då gäller (1) \Leftrightarrow (2)

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

Definition/Sats 6.7

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

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

Bevis 6.1: Bevis av föregående sats

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

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

Exempel:

Låt $\Omega = \{\text{Uppsalas befolkning}\}$, $X = \text{längd}$, $Y = \text{vikt}$, $Z = \text{antal syskon}$

I detta exempel så är X, Y beroende, men X, Z är oberoende samtidigt är Y, Z beroende

Vi tänker oss att X_1, \dots, 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:

$$P_x(1) = P \quad P_x(0) = 1 - P \quad P \in [0, 1]$$

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 $P(\omega) = P(\omega_1, \dots, \omega_n) = P^k(1-p)^{n-k}$ 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, \dots, 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, \dots, X_n med samma fördelning som X .

Bevis 6.2: Existens av oberoende slumpvariabler

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

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 binomialfördelat om $P_X(k) = \binom{n}{k} P^k (1-P)^{n-k}$ för $k = 0, 1, \dots, n$

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

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

Definition/Sats 6.11

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

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

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

Vi tänker på $\text{Bin}(n, P)$ som följande:

Upprepade slumpförsök n gånger. Vinst med sannolikhet $P \in [0, 1]$ och förlust med sannolikhet $1-P = q$. $X \sim \text{Bin}(n, p)$ räknar antalet vinster.

Exempel:

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

Exempel:

Dra 10 lotter. Varje lott har vinstchans på 10%. $X = \text{antal vinster} \sim \text{Bin}(10, 0.1)$.

$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.9^{10} \approx 65\%$

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 $X \sim \text{Bin}(n, P)$ så är $n - X \sim \text{Bin}(n, 1 - P)$

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

Definition/Sats 6.12

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

Bevis 6.3

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

$$\begin{aligned}
 &= \sum_{j=0}^k P(X = j, Y = k - j) = \sum_{j=0}^k P(X = j)P(Y = k - j) \\
 &= \sum_{j=0}^k \binom{n_1}{j} P^j (1 - P)^{n_1 - j} \binom{n_2}{k - j} P^{k - j} (1 - P)^{n_2 - (k - j)} \\
 &= \sum_{j=0}^k \binom{n_1}{j} \binom{n_2}{k - j} P^k (1 - P)^{n_1 + n_2 - k} \\
 &= \left(\sum_{j=0}^k \binom{n_1}{j} \binom{n_2}{k - j} \right) P^k (1 - P)^{n_1 + n_2 - k} \\
 &= \sum_{j=0}^k \binom{n_1}{j} \binom{n_2}{k - j} = \binom{n_1 + n_2}{k} \\
 &\Rightarrow X + Y \sim \text{Bin}(n_1 + n_2, P)
 \end{aligned}$$

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

□

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

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 $\text{Hyp}(\underbrace{n_1 + n_2}_{\text{antalet lotter}}, \underbrace{k}_{\text{dragningar}}, \underbrace{n_1}_{\text{vinstlotter}})$ -fördelad

Exempel:

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

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 $\text{Bin}(5, \frac{13}{52})$

Hypergeometrisk fördelning är alltså binomialfördelad fast utan återlägg.

Om $X \sim \text{Hyp}(N, n, m)$ så får vi med återlägg $\text{Bin}(n, \frac{m}{N})$.

Om N är mycket större än antalet dragningar n , så är $\text{Hyp}(N, n, m) \approx \text{Bin}(n, \frac{m}{N})$

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:

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

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

$$\text{Betecknas } X : \Omega \rightarrow \mathbb{R}$$

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/uppräknligt 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 vår 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:

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

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 7.4: 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

8. MEDELVÄRDE

Vi börjar med ett exempel, myntkastet såklart där $\Omega = \{H, T\}^N$ och N är väldigt stort.

Vi definierar sannolikhetsmättet på rummet som $P(\omega) = \frac{1}{2^n}$.

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 $\frac{n}{2}$ krona. Detta är en frekvenstolkning.

Med andra ord, förväntar vi oss följande:

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

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

Definition/Sats 8.1: Stora talens lag

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

$$P\left(\left|\frac{x_1 + \dots + x_n}{n} - E(X_i)\right| > \varepsilon\right) \xrightarrow[n \rightarrow \infty]{} 0 \text{ för något tal } E(X_i)$$

För diskreta slumpvariabler är:

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

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/medelvärde* 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 $X_1 + \dots + X_n = \sum x \cdot nP_X(x)$ och $\frac{X_1 + \dots + X_n}{n} \rightarrow \sum xP_X(x)$
(Vi summerar över alla X med positiv sannolikhet)

Vi kan definiera $E(X) = \sum xP_X(x)$ 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:

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

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

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 $E(X)$:

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

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 unconscious statistician

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

Bevis 8.1: Law of the unconscious statistician

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

□

Definition/Sats 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 $E(|X|^p) < \infty$

$$L^p(\Omega) = \left\{ X : \Omega \rightarrow \mathbb{R} : E(|X|^p) < \infty, X \text{ är diskret} \right\}$$

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 $E : L^1 \rightarrow \mathbb{R}$

Bevis 8.2

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

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

□

Exempel:

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

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

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

Anmärkning:

Alla $X \sim \text{Bin}(n, p)$ kan inte skrivas $X = X_1 + \dots + X_n$ där X_1, \dots, X_n är Bernoulli-fördelade!

Men, vi vet att det finns $X_1, \dots, X_n \sim \text{Be}(p)$ som är oberoende och $X_1 + \dots + X_n \sim \text{Bin}(n, p)$, 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:

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

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 definierar 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 $P(0.5 < Y \leq 0.9)$. $\underbrace{P(Y \leq 0.9)}_{0.81} = P(\{Y \leq 0.5\} \cup \{0.5 < Y \leq 0.9\}) = \underbrace{P(Y \leq 0.5)}_{0.25} +$

$\Rightarrow 0.81 - 0.25 = 0.56$.

Eftersom de är disjunkta kan vi summera sannolikheterna.

Definition/Sats 8.8: Egenskaper hos fördelningsfunktioner

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

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 $f_Y(y)$ $y \in [1, 5]$

$f_Y(y) = c \frac{1}{y}$. Vi måste även ha att integralen $\int_{-\infty}^{\infty} f_Y(y) dy = 1$

Bestäm fördelningsfunktionen (primitiv funktion till täthetsfunktionen)

$P(Y \leq t) = \int_{-\infty}^t f_Y(y) dy$

Om $t \leq 1 \Rightarrow P(Y \leq t) = 0$

Om $1 \leq t \leq 5 \Rightarrow P(Y \leq t) = \int_{-\infty}^t f_Y(y) dy = \int_1^t f_Y(y) dy = \frac{\ln(t)}{\ln(5)}$

Exempel:

Med tvåpunktsfördelning menas att $P_X(a) = p$ och $P_X(b) = 1 - p$ (notera att detta är $Be(p)$ om $a = 1$ och $b = 0$)

Beräkna $E(X)$ och $Var(X)$:

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

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

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 väntevärdena är definierade (vi tillåter inte att ena är ∞ och den andra $-\infty$)
- $E(cX) = cE(X)$ $c \in \mathbb{R}$
- $|E(X)| \leq E(|X|)$ (Ye Olde' Triangelolikheten)
- $X \geq 0 \Rightarrow E(X) \geq 0$
- $X \leq Y \Rightarrow E(X) \leq E(Y)$

Proposition:

Om $q > p$ så är $L^q \subseteq L^p$

Bevis 8.4

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

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

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

Så:

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

□

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 $p \geq 1$):

$$\begin{aligned} |X + Y|^P &\leq (|X| + |Y|)^P \leq (2 \max\{|X|, |Y|\})^P \leq 2^P \max\{|X|, |Y|\} \leq 2^P (|X|^P + |Y|^P) \\ &\Rightarrow E(|X + Y|^P) \leq 2^P (E(|X|^P) + E(|Y|^P)) < \infty \end{aligned}$$

□

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 definierades som $E(X - E(X))^2$. Ett mått på hur mycket variabeln avviker från väntevärdet.

Standardavvikelsen definierade vi som $D(X) = \sqrt{\text{Var}(X)}$. En grej vi kan notera direkt är att $\text{Var}(X)$ alltid är positiv, alltså alltid definierad.

Proposition:

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

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

Bevis 8.7

Detta följer från

$$\begin{aligned} E(X - E(X))^2 &= E(X^2 - \underbrace{2XE(X)}_{\text{ändlig}} + (EX)^2) = E(X^2) - \underbrace{2E(X)E(X)}_{2(E(X))^2} + (E(X))^2 \\ &\Rightarrow E(X^2) - (E(X))^2 \end{aligned}$$

□

Vi säger att X & Y är *okorrelerade* om $E(XY) = E(X)E(Y)$ för $X, Y \in L^2$

Notera, oberoende \Rightarrow okorrelerade, men inte tvärtom!

Exempel:

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

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

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

Då är X & X^2 okorrelerade, men X och X^2 kan ju inte vara oberoende!

$$P(X = 0, X^2 = 0) = P(X = 0) = \frac{1}{3} \neq P(X = 0)P(X^2 = 0) = \frac{1}{3} * \frac{1}{3}$$

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:

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

□

Anmärkning:

Vad är $Var(cX)$?:

$$\begin{aligned} Var(cX) &= E(cX)^2 - (E(cX))^2 \\ E(c^2X^2) - (cE(X))^2 &= c^2E(X^2) - c^2(E(X))^2 = c^2Var(X) \end{aligned}$$

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 $E((X_1 + X_2)Y)$:

$$\begin{aligned} &= E(X_1Y + X_2Y) = E(X_1Y) + E(X_2Y) = E(X_1)E(Y) + E(X_2)E(Y) \\ &\Rightarrow (E(X_1) + E(X_2))E(Y) = E(X_1 + X_2)E(Y) \end{aligned}$$

□

Proposition:

Om X_1, \dots, X_n är parvis okorrelerade, så $Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$

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 $P(|X| \geq a) \leq \frac{E(|X|)}{a}$ för $a > 0$
Ju större a är, desto mindre borde mängden $P(|X| \geq a)$ vara.

Bevis 8.11: Markovs olikhet

$$\begin{aligned}
 |X| &= |X| 1_{X \geq a} + |X| 1_{X < a} \\
 E(|X|) &= E(|X| 1_{X \geq a}) + E(\underbrace{|X| 1_{X < a}}_{\geq 0}) \geq E\left(\underbrace{|X| 1_{X \geq a}}_{\geq a}\right) \\
 &\leq E(a 1_{X \geq a}) = aE(1_{X \geq a}) = a(1 * P(X \geq a) + 0 * P(X < a)) = aP(X \geq a)
 \end{aligned}$$

Alltså $E(|X|) \geq aP(X \geq a)$ □

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

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 $X_1, X_2, \dots \in L^2$ (oändlig följd av okorrelerade slumpvariabler)

Antag även att $E(X_1) = E(X_2) = \dots = \mu \in \mathbb{R}$ och $Var(X_1) = Var(X_2) = \dots = \sigma^2 \in \mathbb{R}$

Vi skriver \bar{X}_n för medelvärde:

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

För $\varepsilon > 0$ har vi:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0$$

Bevis 8.14: Annorlunda Stora talens lag

$$\begin{aligned}
E(\bar{X}_n) &= E\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \frac{E(X_1) + \dots + E(X_n)}{n} = \frac{n\mu}{n} = \mu \\
\text{Var}(\bar{X}_n) &= \text{Var}\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \frac{1}{n^2}\text{Var}(X_1 + \dots + X_n) \\
&\Rightarrow \frac{1}{n^2}(\overbrace{\text{Var}(X_1)}^{\sigma^2} + \dots + \overbrace{\text{Var}(X_n)}^{\sigma^2}) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n} \\
\text{Chebyshevs olikhet sade } P\left(\left|\bar{X}_n - \underbrace{E(\bar{X}_n)}_{\mu}\right| \geq \varepsilon\right) &\leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0
\end{aligned}$$

□

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 $(X, Y) \in L^2$ (väntevärdena av kvadraterna är ändliga) betecknas:

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

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

Egenskaper:

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

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 $\text{Cov}(X, X) = \text{Var}(X) = 0$?

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

Om $x \neq E(X)$ så måste $P_X(x) = 0 \Rightarrow P_X(E(X)) = 1$. Alltså om $\text{Var}(X) = 0$ 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 $\text{Cov}([X], [Y]) = \text{Cov}(X, Y)$. 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 $aX \in L^2$ ($a \in \mathbb{R}$). 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 $|Cov([X], [Y])| \leq Cov([X], [X])Cov([Y], [Y]) = \sqrt{Var([X])Var([Y])}$

Notera, ”ortogonala vektorer” ger att inre produkt är 0, vilket i vårt fall betyder att $Cov(X, Y) = 0$ vilket händer om X, Y är okorrelerade.

Definition/Sats 8.12: Korrelationskoefficienten

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

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 $[Y] = a[X]$ ($a \in \mathbb{R}$) eller om $[X] = 0$

Med andra 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:

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

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

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

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

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

□

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 vårt 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 $[a, b]$

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 $P([0, 1 \setminus \mathbb{Q}]) = \underbrace{P([0, 1])}_{1-0=1} - \underbrace{P[0, 1 \cap \mathbb{Q}]}_0 = 1 - 0 = 1$

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...

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 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 $x = 0.x_1x_2x_3 \dots$ 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 $C_n \subseteq \dots \subseteq C_3 \subseteq C_2 \subseteq C_1$

Om Kolmogorovs axiom håller måste vi ha $P(C) = P(\bigcap_{n=1}^{\infty} C_n) = \lim_{n \rightarrow \infty} P(C_n) = \lim_{n \rightarrow \infty} \left(\frac{2}{3}\right)^n = 0$

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 $X : \Omega \rightarrow \mathbb{R}$ kallas *kontinuerligt likformig* om:

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

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

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

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

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 $a \leq c < d \leq b$

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

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

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

Exempel:

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

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

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

Exempel:

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

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

Detta kallas för *standardiserad normalfördelning*

Vi skriver φ för fördelningsfunktionen till $N(0, 1)$, med andra ord:

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

Om $X \sim N(\mu, \sigma^2)$, så är

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

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

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

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

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 $\bar{X}_n = \frac{X_1 + \dots + X_n}{n} \approx N(\mu, \sigma^2)$ fördelad, i meningen att:

$$P\left(a \leq \frac{\bar{X}_n - \mu}{\sigma\sqrt{n}} \leq b\right) \xrightarrow{n \rightarrow \infty} \varphi(b) - \varphi(a)$$

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

$$\text{Kom ihåg: } E(\bar{X}_n) = E\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \frac{n\mu}{n} = \mu$$

$$Var(\bar{X}_n) = Var\left(\frac{1}{n}(X_1 + \dots + X_n)\right) = \frac{1}{n^2}(Var(X_1) + \dots) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$D(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

Övning:

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

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

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 $\frac{1}{2}$ att anta.

Då gäller $X \in C$ (Cantormängden)

Övning:

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

9. LEKTION 21/9

Övning 3.8.1

Betrakta den 2-dimensionella slumpvariabeln (X, Y) med sannolikhetsfunktion $P_{X,Y}(j, k) = c(j + k)$. Detta gäller för $i = 1, 2, 3$ och $k = 1, 2, 3$

Bestäm konstanten c och beräkna väntevärdet, variansen, och kovariansen. Beräkna även väntevärdet för X givet att $Y = 3$

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

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

Vi räknar väntevärdet:

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

Vi räknar variansen:

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

Vi räknar kovariansen:

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

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

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

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

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

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

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

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

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

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

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

Vi räknar det betingade väntevärdet:

$$\begin{aligned} E(X|Y = 3) &= \sum_{k=1}^3 kP(X = k|Y = 3) \\ &= \sum_{k=1}^3 k \frac{P(X = k, Y = 3)}{P(Y = 3)} = \sum_{k=1}^3 k \frac{P_{X,Y}(k, 3)}{P_Y(3)} \\ &= \frac{12}{5} \sum_{k=1}^3 k \frac{P_{X,Y}(k, 3)}{12} = \frac{12}{5} \frac{1}{36} \sum_{k=1}^3 k(k+3) \\ &= \frac{1}{15} (1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6) = \frac{32}{15} \end{aligned}$$

10. KORT INTRODUKTION TILL MÅTTEORI

Vi hade följande beteckning $X \sim U(0, 1)$ 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, $P^* : 2^{[0,1]} \rightarrow \mathbb{R}$ är inte ett sannolikhetsmått (uppfyller ej Kolmogorovs axiom) enligt Banach-Tarski problemet.

Vad vi inser är att vi borde 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 $\mu : 2^\Omega \rightarrow [0, \infty]$ som uppfyller:

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

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:

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

För $A \subseteq \mathbb{R}^n$:

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

Problemet kvarstår, yttre måttet är inte ett mått och det finns inget mått, säg $\mu : 2^{\mathbb{R}} \rightarrow [0, \infty]$ 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 vara 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

Idén är följande:

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

F kommer vara följande:

$B(\mathbb{R}^n) =$ minsta samling delmängder $F \subseteq 2^{\mathbb{R}^n}$ som uppfyller :

$$(1) \quad (a_1, b_1)x \cdots x(a_n, b_n) \in F \quad -\infty \leq a_i \leq b_i \leq \infty$$

$$(2) \quad A \in F \Rightarrow A^c \in F$$

$$(3) \quad A_1, A_2, \dots \in F \Rightarrow \bigcup_{n=1}^{\infty} A_n \in F$$

Att det ens existerar ett "minsta" i en mängd är inte alltid garanterat, men i detta fall gäller det (vi kitar på detta lite senare).

Om vi har (1)-(3), så kommer vi även ha exempelvis:

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

Vi vill mäta allt som vi kan bilda om vi tar union, komplement, snitt, differens osv.

Tar vi en mängd $A \in \mathcal{B}(\mathbb{R}^n)$, så kallas de för *Borelmängder*

Definition/Sats 10.2: Specialfall av Caratheodorys sats

- $\mu^* : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, \infty]$ 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:

$$I_1 \cup \dots \cup I_m \quad \text{för öppna intervall } \subseteq \mathbb{R}^n$$

(Unionen av rektanglar)

Om $\mu_0 : R \rightarrow [0, \infty]$ mäter delmängderna och uppfyller kraven för ett mått (Definition 10.1), så finns en unik $\mu : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, \infty]$ 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:

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

Eller exempelvis:

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

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:

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

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

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

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 specificera 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 $X : \Omega \rightarrow \mathbb{R}^n$ 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 $P(a \leq X \leq b) = \int_a^b f_X(x)dx \Leftrightarrow$

$$F_X(t) = \underbrace{\int_{-\infty}^t f_X(x)dx}_{\text{tätthetsfunktion}}$$

En absolutkontinuerlig slumpvariabel bestäms unikt av f_X

Definition/Sats 10.5: Fördelning

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

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

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 sannolikhetsmättet $P : F \rightarrow [0, 1]$ 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:

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

Exempel:

(Exponentialfördelning) $X \sim \text{Exp}(\beta)$ ($\beta > 0$), 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:

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

Anmärkning:

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

Exempel:

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

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, \dots uppfyller 1,2,3 så finns oberoende slumpvariabler X_1, X_2, \dots så att $F_k = F_{X_k}$

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

- 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
- $X \in L^p \Leftrightarrow E(|X|^p) = \int |X|^p f_X(x)dx < \infty$
- $Var(X) = E(X - E(X))^2 = \int (x - E(X))^2 f_X(x)dx = E(X)^2 - (E(X))^2$ om $X \in L^2$
- $Cov(X, Y) = E(X - E(X))(Y - E(Y)) = E(XY) - E(X)E(Y)$ om $X, Y \in L^2$
- $E(g(X, Y)) = \int \int g(x, y)f_{X,Y}(x, y)dxdy$ 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 $\underbrace{E(XY) = E(X)E(Y)}$, kovariansen är bilinjär osv gäller även för Absolutkontinuerliga slump-
okorrelerade slumpvariabler
variabler
- $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$
- Markovs olikhet är bevarad: $P(X \geq a) \leq \frac{E(X)}{a}$ om $a > 0$
- Stora talens lag: $P(|\bar{X}_n - \mu| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0$
- Falttningsformlerna: Om X, Y är oberoende så:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$$

- Betingad fördelning: $f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$ 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 $Var(X) = \frac{1}{\beta^2}$

Exempel:

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

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

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

Vi passar på att räkna ut $Var(x)$:

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

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

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

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

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 obereonde) 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:

$$\begin{aligned}
 f_{X+Y}(z) &= \int_{-\infty}^{\infty} \overbrace{f_X(x)}^{>0 \Leftrightarrow x > 0} \underbrace{f_Y(z-x)}_{x < z} dx \\
 X \sim Exp(\lambda) &\Leftrightarrow F_X(t) = \begin{cases} 1 - e^{-\lambda t}, & t > 0 \\ 0, & t \leq 0 \end{cases} \\
 f_X(x) &= \begin{cases} \lambda e^{-\lambda x}, & t > 0 \\ 0, & t \leq 0 \end{cases} \quad E(X) = \frac{1}{\lambda} \quad Var(X) = \frac{1}{\lambda^2} \\
 \Rightarrow f_{X+Y} &= \begin{cases} 0, & z \leq 0 \\ \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx, & z > 0 \end{cases} \\
 &= \lambda^2 \int_0^z e^{-\lambda x} e^{-\lambda z + \lambda x} dx = \lambda^2 \int_0^z e^{-\lambda z} e^{\lambda x} dx \\
 &= \lambda^2 z e^{-\lambda z} \\
 \Rightarrow f_{X+Y}(z) &= \begin{cases} \lambda^2 z e^{-\lambda z}, & z > 0 \\ 0, & z \leq 0 \end{cases}
 \end{aligned}$$

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

Övning:

Visa att om $X_1, \dots, X_n \sim Exp(\lambda)$ är obereonde så är $X_1 + \dots + X_n \sim \Gamma(n, \lambda)$

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 $P(Z > z) = P(\min\{X, Y\} > z) = 1 - F_Z(z)$:

$$\begin{aligned}
 \min\{X, Y\} > z &\Leftrightarrow X > z \quad \& Y > z \\
 P(Z > z) &= P(X > z, Y > z) \stackrel{ober.}{=} P(X > z)P(Y > z) \\
 &= (1 - F_X(z))(1 - F_Y(z)) = \begin{cases} 1, & z \leq 0 \\ e^{-\lambda_1 z} e^{-\lambda_2 z}, & z > 0 \end{cases} \\
 &= \begin{cases} e^{-(\lambda_1 + \lambda_2)z}, & z > 0 \\ 1, & z \leq 0 \end{cases} \\
 \Rightarrow F_Z(t) &= \begin{cases} 1 - e^{-(\lambda_1 + \lambda_2)t}, & t > 0 \\ 0, & t \leq 0 \end{cases} \\
 \Rightarrow \min\{X, Y\} &\sim Exp(\lambda_1 + \lambda_2)
 \end{aligned}$$

Övning:

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

Lösning:

$$\begin{aligned} F_{\sqrt{X}}(t) &= P(\sqrt{X} \leq t) = \begin{cases} 0, & t \leq 0 \\ P(X \leq t^2), & t > 0 \end{cases} \\ &= \begin{cases} F_X(t^2), & t > 0 \\ 0, & t \leq 0 \end{cases} \Rightarrow f_{\sqrt{X}}(t) = \begin{cases} \frac{d}{dt} F_X(t^2), & t > 0 \\ 0, & t \leq 0 \end{cases} \\ &= \begin{cases} 2t f_X(t^2), & t > 0 \\ 0, & t \leq 0 \end{cases} = \begin{cases} 2\lambda t e^{-\lambda t^2}, & t > 0 \\ 0, & t \leq 0 \end{cases} \end{aligned}$$

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):

$$P(X > s + t \mid X > s) = P(X > t) \quad (s, t \geq 0)$$

(Minneslöshet kallas detta även)

Lösning:

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

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 $P(X > s + t | X > s) = P(X > t)$ 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 $P(X > s + t | X > s) = P(X > t) \Leftrightarrow \varphi(s + t) = \varphi(s)\varphi(t)$

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

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

$$\begin{aligned}\varphi(2) &= \varphi(1 + 1) = \varphi(1)\varphi(1) = q^2 \\ n \in \mathbb{Z}_+ &\Rightarrow \varphi(n) = \varphi(\underbrace{1 + \dots + 1}_{n \text{ st}}) = \varphi(1)^n = q^n\end{aligned}$$

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 $\varphi(1)$:

$$\begin{aligned}\varphi\left(\underbrace{\frac{1}{n} + \dots + \frac{1}{n}}_{n \text{ termer}}\right) &= \varphi\left(\frac{1}{h}\right)^n \Rightarrow \varphi\left(\frac{1}{n}\right) = q^{1/n} \\ \frac{m}{n} \in \mathbb{Q}_{>0} &\Rightarrow \varphi\left(\frac{m}{n}\right) = \varphi\left(\frac{1}{n}\right)^m = (q^{1/n})^m = q^{\frac{m}{n}}\end{aligned}$$

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 $q_n \rightarrow t$

Eftersom fördelningsfunktionen är högerkontinuerlig så måste $F_X(q_n) \rightarrow F_X(t)$:

$$1 - e^{-\lambda q_n} \xrightarrow{n \rightarrow \infty} 1 - e^{-\lambda t}$$

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

Exempel:

På ICA kommer det in 20 kunder per timme.

Om vi sätter X = tiden innan nästa kund kommer in, så är $X \sim \text{Exp}(\lambda)$, 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 noggrant, säg att vi tar en massa slumpvariabler X_1, X_2, \dots som är normalfördelade med parameter 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 $X \sim \text{Po}(\lambda t)$. Säger vi att $X \sim \text{Po}(\lambda)$ så är:

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

Om $X \sim \text{Po}(\lambda)$, 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 $\text{Var}(X) = \lambda$

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?

$$\begin{aligned}
 P(X = n) &= 1 - \overbrace{P(X > n)}^{q^n} - P(X < n) \\
 &= 1 - q^n - \sum_{k=1}^{n-1} P(X = k) = 1 - q^n - \sum_{k=1}^{n-1} pq^{k-1} = 1 - q^n - p \underbrace{\sum_{k=0}^{n-2} q^k}_{\text{geometrisk summa}} \\
 &= 1 - q^n - p \underbrace{\frac{q^{n-1} - 1}{q - 1}}_{-p} = 1 - q^n - (1 - q^{n-1}) = q^{n-1} - q^n = q^{n-1} \underbrace{(1 - q)}_p \\
 &= pq^{n-1}
 \end{aligned}$$

□

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 $X_1, X_2, \dots \sim \text{Be}(p)$ 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?

$$\begin{aligned}
 P(X = n) &= P(X_1 = 0, \dots, X_{n-1} = 0, X_n = 1) = \underbrace{P(X_1 = 0)}_{1-p} \cdots \underbrace{P(X_{n-1} = 0)}_{1-p} \underbrace{P(X_n = 1)}_p \\
 &= p(1 - p)^{n-1}
 \end{aligned}$$

Övning:

Visa att $E(X) = \frac{1}{p}$ samt att $\text{Var}(X) = \frac{1-p}{p^2}$