

Kursen består av två delar/ämnen

Sannolikhetsteori

Matematiska modeller
för slumpmässiga
fenomen

Statistik

- i) dra slutsatser från data
- ii) Testa hypoteser

(ca 60 % av kursen)

Sannolikhetsteori:

Vår modell: ("tänk att vi genomför
ett experiment")

i) Utfallsrummet

$S = \{\text{alla möjliga utfall av vårt experiment}\}$

ii) Ett element $w \in S$ är ett specifikt
utfall

iii) En händelse $A \subset S$ är en samling
utfall och utgör en delmängd av S

iv) En funktion $P: A \subset S \rightarrow [0, 1]$
som vi tolkar som sannolikheten
för A

Anm: Ibland används $\subset$ enbart för
en strikt delmängd, medan

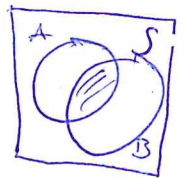
$\subseteq$ används i fall likhet är tillåtet. Här innebär $A \subset S$ att A kan vara S , dvs likhet är tillåtet.

Ex: Tärningskast: $S = \{1, 2, 3, 4, 5, 6\}$.

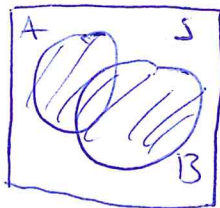
Händelse $A = \{1\}$, $B = \{1, 3, 5\}$.

Lite mängdlära

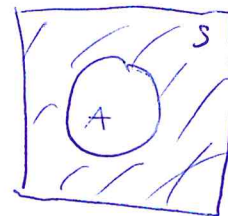
i) $A \cap B = \{w \in S : w \in A \text{ och } w \in B\}$



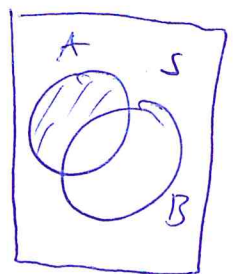
ii) $A \cup B = \{w \in S : w \in A \text{ och/eller } w \in B\}$



iii) $A^c = \{w \in S : w \notin A\}$

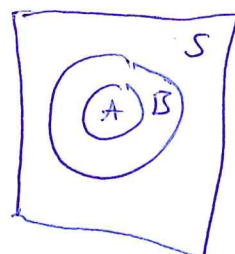


iv) $A \setminus B = \{w \in S : w \in A, w \notin B\}$



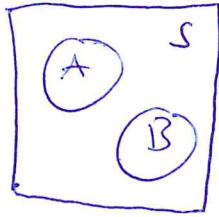
v) Vi skriver $A \subset B$ om

$$\forall w \in A \Rightarrow w \in B$$



vi) Om $A \cap B = \emptyset$ sägs A, B vara

disjunkta (disjoint, mutually exclusive)



Funktionen P kallas för ett sannolikhetsmått och uppfyller följande axiom:

1/ $0 \leq P(A) \leq 1 \quad \forall \quad A \subset S$

2/ $P(S) = 1$

3/ Om $A_1, A_2, A_3, \dots$ är disjunkta gäller att $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$

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

vi) ovan

Sats (räkne regler för P):

Låt $A, B \subset S$ vara godtyckliga.

Vi har att

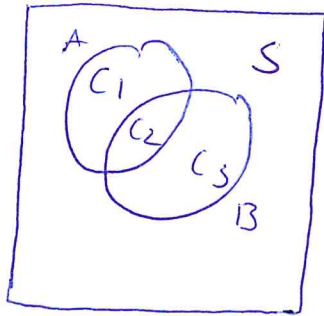
i) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$ii) \quad P(A^c) = 1 - P(A)$$

$$iii) \quad \text{Om } A \subset B \Rightarrow P(A) \leq P(B)$$

B : (öv 12, kap 2)

i)



$$C_1 = A \setminus B$$

$$C_2 = A \cap B$$

$$C_3 = B \setminus A$$

$$A \cup B = C_1 \cup C_2 \cup C_3$$

OBS: C_1, C_2 och C_3
är disjunkta

$\forall i$ har att

$$P(A \cup B) + P(A \cap B) = P(C_1 \cup C_2 \cup C_3) + P(C_2)$$

$$= P(C_1 \cup C_2) + P(C_3) + P(C_2)$$

$\uparrow$
axiom 3)

$$\downarrow$$

$$= P(C_1 \cup C_2) + P(C_3 \cup C_2) = P(A) + P(B)$$

$$ii) \quad A \cup A^c = S \Rightarrow \underset{\substack{\uparrow \\ \text{ax 2)}}}{1} = P(S) = P(A \cup A^c) = \underset{\substack{\uparrow \\ \text{ax 3)}}}{P(A) + P(A^c)} = P(A) + P(A^c)$$

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

$$iii) \quad \text{Om } A \subset B \Rightarrow B = A \cup (B \setminus A)$$

$$\Rightarrow P(B) = P(A \cup (B \setminus A)) \underset{\substack{\uparrow \\ \text{ax 3)}}}{=} P(A) + P(B \setminus A) \underset{\substack{\uparrow \\ \text{ax 1)}}}{\geq} P(A) \quad \square$$

Divisionsregeln: Om alla utfall i S är lika sannolika så gäller att

$$P(A) = \frac{\# \text{ utfall i } A}{\# \text{ utfall i } S} = \frac{|A|}{|S|}$$

Tal: 5 familjer har två barn:

1	2	3	4	5
P	P	P	F	F
F	F	F	F	F

a) välj en familj slumpmässigt och låt

$A = \{ \text{den valda familjen har två flickor} \}$

b) välj en flicka slumpmässigt och låt

$B = \{ \text{den valda flickans familj har två flickor} \}$

L: a) $P(A) = \frac{|A|}{|S|} = \frac{2}{5}$

b) $P(B) = \frac{|B|}{|S|} = \frac{4}{7} //$

Lite kombinatorik:

Ett experiment genomförs i k steg.

Låt $N_\ell = \#$ möjliga utfall i steg ℓ .

Vi får då att $|S| = N_1 \cdot N_2 \cdots N_k$

Låt $A = \{a_1, a_2, \dots, a_n\}$ vara en mängd objekt ("kortlek"). En permutation av A är en lista (listning) av objekten i A .

Ex: $A = \{a_1, a_2, a_3\}$ perm: (a_1, a_2, a_3) ,
 (a_1, a_3, a_2) , (a_3, a_1, a_2) etc

I) På hur många sätt kan vi permutera A ?

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

II) På hur många sätt kan vi välja k av objekten i A och lägga i ordning (dvs hur många listor av längd k kan vi skapa?)?

$$n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

III) På hur många sätt kan vi välja k av objekten i A om vi ej tar hänsyn till ordning?

(vi säger att vi kombinerar de k objekten, tänk pokerhand)

listor är $\frac{n!}{(n-k)!}$. En specifik

lista $(\underbrace{a_1, a_2, a_3, \dots, a_k}_{k \text{ objekt}})$ kan permuteras på $k!$ olika sätt. Alla dessa permutationer innehåller samma k objekt.

$$\Rightarrow \# \text{ kombinationer} = \frac{\# \text{ listor}}{\# \text{ perm.}}$$

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

$\binom{n}{k}$ kallas för en binomialkoefficient.

Tal: En löpartävling har 52 deltagare.

På hur många sätt kan topp-5 listan se ut?

L: $52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 = \frac{52!}{47!} //$

Tal: Hur många pocher händer finns det?

L: Vi väljer 5 kort ur en

lek om 52 $\Rightarrow \binom{52}{5} = \frac{52!}{47! \cdot 5!} //$

Ex: Vid en fartkontroll är hastighetsbegr. 50 km/h. 32 % av alla bilar hör fortare än 50 km/h och 4 % av alla bilar hör fortare än 70 km/h.

Vad är andelen forthörare som hör över 70 km/h?

$$\frac{0.04}{0.32} = 0.125 \quad (12.5 \%)$$

Tolkning: Sannolikheten att en slumpmässigt vald bil hör över 70 km/h givet att den hör "50" är 0.125.

Def: Låt $A, B \subset S$ s.a. $P(B) > 0$.

Då är den betingade sannolikheten för A givet B

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

I vårt ex. $A = \{\text{en bil kör } > 70 \text{ km/h}\}$

$B = \{\text{bil kör } > 50 \text{ km/h}\}$

$$\Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} \quad (\text{ty } A \subset B.)$$

Tal: Vi kastar tärning 2 ggr.

Givet att ögonsumman = 4, vad är

sann. att vi fick minst en 3:a?

L: Vi har $S = \{(a, b) : 1 \leq a, b \leq 6\}$

$B = \{\text{ögonsumman} = 4\} = \{(1, 3), (2, 2), (3, 1)\}$

$A = \{\text{minst en 3:a}\} = \{(1, 3), (2, 3), \dots, (6, 3), (3, 1), \dots, (3, 6)\}$

$A \cap B = \{(1, 3), (3, 1)\}$

Vi får att $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{3/36} = \frac{2}{3} //$

"Att betinga på B är att göra B till det nya S."

Sats: Givet $B \subset S$ s.a. $P(B) > 0$,

låt P' vara funktionen som för varje $A \subset S$ tar värdet

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

Då är P' ett sannolikhetsmått.

"Varför? Alla räkneregler för P gäller för $P(\cdot | B)$ "

B: Vi måste verifiera att P' uppfyller de tre villkoren för ett sannolikhetsmått.

1/ visa att $P'(A) \in [0, 1] \quad \forall A \subset S$

$$P'(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0 \quad \text{ty } P(A \cap B) \geq 0$$

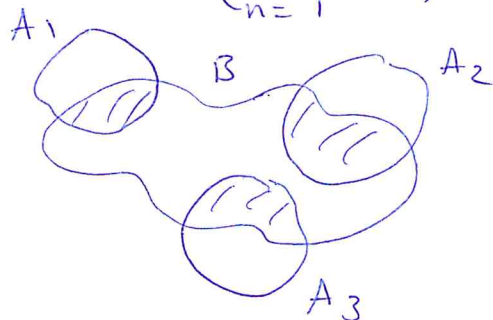
$$P'(A) = \frac{P(A \cap B)}{P(B)} \leq 1 \quad \text{ty } P(A \cap B) \leq P(B)$$

$$2/ \quad P'(S) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

3/ Låt $A_1, A_2, \dots$ vara disjunkta

Visa att $P'\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P'(A_n)$.

OBS: $\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B = \bigcup_{n=1}^{\infty} A_n \cap B$



$$\text{Låt } x \in \left(\bigcup_{n=1}^{\infty} A_n\right) \cap B$$

$$\Rightarrow x \in B \text{ och } x \in A_n \text{ för ett } n \Rightarrow x \in B \cap A_n$$

$$\Rightarrow x \in \bigcup_{n=1}^{\infty} (A_n \cap B)$$

Vi ser att

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \frac{\mathbb{P}\left(\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n \cap B\right)}{\mathbb{P}(B)}$$

$$= \frac{\sum_{n=1}^{\infty} P(A_n \cap B)}{P(B)} = \sum_{n=1}^{\infty} \frac{P(A_n \cap B)}{P(B)} =$$

$$= \sum_{n=1}^{\infty} P(A_n | B) = \sum_{n=1}^{\infty} P'(A_n) \quad \square$$

Tal: I en population är 20 % smittade av TBC. Ett snabbtest ger följande:

i) En smittad får positiv diagnos m.s. 0,9

$i\hat{c}/E_j$ " " negativ " " 0,7

Vald är sann. att en slumpmässig +
vald individ får positiv diagnos?

1: Steg 1: (Inför beteckningar)

Låt $M = \{ \text{valid person smittad} \}$

$$+ = \{ \text{" " f\"ar positiv diagnos} \}$$

$- = \angle$ " " " negativ " }

$$= (+)^c$$

Steg 2: (vad vet vi?)

$$P(+|M) = 0.9$$

$$P(-|M^c) = 0.7$$

$$P(M) = 0.2$$

Steg 3: (lösning) Vi söker $P(+)$

$$P(+)=P(+ \cap M)+P(+ \cap M^c)$$

$$=P(+|M)P(M)+P(+|M^c)P(M^c)$$

$$= \begin{cases} P(+|M^c) \stackrel{\text{Ann. } P^1}{=} 1 - P(-|M^c) = 1 - 0.7 = 0.3 \\ P(M^c) = 1 - P(M) = 1 - 0.2 = 0.8 \end{cases}$$

$$= 0.9 \cdot 0.2 + 0.3 \cdot 0.8 = 0.42 //$$

Ann: Första likheten gäller då

M och M^c är disjunkta.

Def: $A_1, \dots, A_n$ är en partition av S

om i) de är disjunkta ii) $\bigcup_{k=1}^n A_k = S$.

Sats: Om $A_1, \dots, A_n$ är en partition av S har vi att

$$P(B) = \sum_{k=1}^n P(B \cap A_k) = \sum_{k=1}^n P(B|A_k)P(A_k)$$

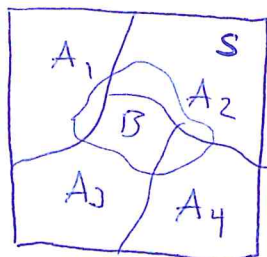
för alla $B \subset S$.

Den andra likheten gäller enbart

om $P(A_k) > 0 \quad \forall k$

(kallas LTP, Lagen om Total Sannolikhet)

B:



$$B = B \cap S = B \cap \left(\bigcup_{k=1}^n A_k \right)$$

$$= \bigcup_{k=1}^n \underbrace{(B \cap A_k)}_{\text{disjunkte!}}$$

$$\Rightarrow P(B) = P\left(\bigcup_{k=1}^n (B \cap A_k)\right) = \sum_{k=1}^n P(B \cap A_k)$$

"andra likheten är det av bet. sann" $\square$

Sats: (Bayes sats) Om $P(A), P(B) > 0$

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

B:

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

Sats: (Bayes sats generell)

Om $A_1, \dots, A_n$ är en partition av S så gäller att

$$P(A_k|B) = \frac{P(A_k) P(B|A_k)}{\sum_{j=1}^n P(B|A_j) P(A_j)}$$

B: övning! $\square$

Tal: (forts)

a) Givet att en person fått positiv diagnos, vad är sann. att hen är smittad?

L: Vi söker $P(M|+)$ och har

$$P(M|+) = P(+|M) \frac{P(M)}{P(+)} = 0.9 \cdot \frac{0.2}{0.42} = \dots = \frac{3}{7} //$$

b) Vad är sann. att testet visar rätt?

L: Låt $R = \{\text{testet är korrekt}\}$.

Vi har

$$\begin{aligned} P(R) &= P(R \cap +) + P(R \cap -) = P(M \cap +) + P(M^c \cap -) \\ &= P(+|M) P(M) + P(-|M^c) P(M^c) \\ &= 0.2 \cdot 0.9 + 0.8 \cdot 0.7 = 0.74 // \end{aligned}$$

Oberoende händelser

Intuitivt är A, B oberoende om

$$P(A|B) = P(A) \quad \text{och} \quad P(B|A) = P(B),$$

Problem: kräver att $P(A), P(B) > 0$.

Istället:

Def: $A, B \subset S$ är oberoende om

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

Följd: Om $P(B) > 0$ och A, B oberoende

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

Def: $A_1, A_2, \dots, A_n$ är (fullständig) oberoende om för varje ändlig

delmängd $\{i_1, i_2, \dots, i_n\} \subset \{1, 2, \dots\}$

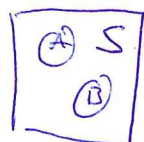
gäller att

$$P\left(\bigcap_{k=1}^n A_{i_k}\right) = \prod_{k=1}^n P(A_{i_k})$$

OBS!!! Oberoende är ej samma som

disjunkth! ober: $P(A \cap B) = P(A) P(B)$

disj: $P(A \cap B) = 0$



Tal: Antag att A och B är oberoende. Visa att A^c, B är oberoende.

L: Vi måste visa att

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

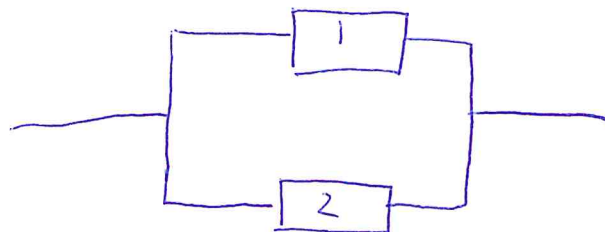
$$P(A^c \cap B) = P(B) - P(A \cap B)$$

$$\stackrel{\substack{\uparrow \\ A, B \text{ ober.}}}{=} P(B) - P(A) P(B) = P(B) (1 - P(A))$$

$$= P(B) P(A^c) \quad \square$$

Följd: Om A, B är ober. så är även A^c och B^c ober.

Tal: Parallellkoppling



Vi har att

- 1/ Komponenterna fungerar ober. av varandra
- 2/ Systemet fungerar om minst en komp. fung.

3/ 1 komp. 1 fungerar m.s. $\frac{1}{2}$
 " 2 " " $\frac{1}{3}$

Vad är sann. att systemet fungerar?

1: $A_1 = \{ \text{komp. 1 fungerar} \}$

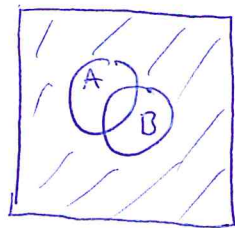
$A_2 = \{ \text{komp. 2 " } \}$

$A_1 \cup A_2 = \{ \text{systemet " } \}$

$$\Rightarrow P(\text{systemet fungerar}) = P(A_1 \cup A_2)$$

$$= 1 - P((A_1 \cup A_2)^c) = 1 - P(A_1^c \cap A_2^c)$$

$$= 1 - P(A_1^c) P(A_2^c) = 1 - (1 - \frac{1}{2})(1 - \frac{1}{3}) = \frac{2}{3}$$



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

//

Tal (reserv): Vi drar en pokerhand (5 kort) ur en standardkortlek.

a) Beräkna sann. att man får minst 2 hj. och 2 spader

1: Låt $H = \{\text{man får minst 2 hjärter}\}$
 $S_p = \{\text{" " " 2 spader}\}$

Vi söker $P(H \cap S_p)$. Är H och S_p oberoende? Intuitivt nej.

I stället: använder vi divisionsregeln:

$$P(H \cap S_p) = \frac{|H \cap S_p|}{1 \cdot 1} = \frac{\binom{13}{2} \binom{13}{2} \binom{26}{1}}{\binom{52}{5}}$$

$$\text{t.v. } 1 \cdot 1 = \binom{52}{5} \text{ och } |H \cap S_p| = \binom{13}{2} \binom{13}{2} \binom{26}{1} //$$

b) Vad är sann. för H givet S_p ?

2: Vi söker

$$P(H | S_p) = \frac{P(H \cap S_p)}{P(S_p)} = \left\{ P(S_p) = \frac{\binom{13}{2} \binom{39}{3}}{\binom{52}{5}} \right\}$$

$$\begin{aligned} &= \frac{\frac{\binom{13}{2} \binom{13}{2} \binom{26}{1}}{\binom{52}{5}}}{\frac{\binom{13}{2} \binom{39}{3}}{\binom{52}{5}}} = \frac{\binom{13}{2} \binom{13}{2} \binom{26}{1}}{\binom{13}{2} \binom{39}{3}} \\ &= \frac{\binom{13}{2} \binom{26}{1}}{\binom{39}{3}} \quad (\neq P(H)) // \end{aligned}$$

Slumpvariabler

En slumpvariabel (random variable, stochastic variable) är en funktion

$$X: S \rightarrow \mathbb{R}$$

Γ Kan göras mer generell, t.ex.

$$X = \{\text{färgen på draget kort ur kortlek}\}$$

$$X: S \rightarrow \{\text{spader, hjärter, ruter, klöver}\}$$

Ex: Vi kastar tärning 2 ggr, låt

$$X = \text{ögons summan. } S = \{(a, b) : 1 \leq a, b \leq 6\}$$

$$X(a, b) = a + b$$

$$\{X = 5\} = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$$

Vi skriver

$$\mathbb{P}(X=5) = \mathbb{P}(\{\omega \in S : X(\omega) = 5\})$$

Diskreta s.v.

Def: En s.v. är diskret om den endast kan anta ändligt eller uppräkneligt oändligt antal värden.

Anm: Vi kommer ofta betrakta

$$\underline{X}: S \rightarrow \mathbb{Z} \quad (\underline{X} \in \mathbb{Z})$$

Def: För en diskret s.v. $\underline{X}$

definieras täthetsfunktionen (tfn)

(frekvens-, sannolikhets-fn, density-function)

$f_{\underline{X}}(x)$ genom

$$f_{\underline{X}}(x) = \mathbb{P}(\underline{X} = x) \quad \text{för alla möjliga värden } x \text{ som } \underline{X} \text{ kan anta,}$$

Anm: ^{1/} Ofta skriver vi bara $f(x)$,

^{2/} Om $\mathbb{P}(\underline{X} = a) = 0$ används ofta

$$\text{att } f_{\underline{X}}(a) = 0.$$

Def: Om $\underline{X} \in \mathbb{R}$ definieras

fördelningsfunktionen (tfn, cdf = cumulative distribution function) av

$$F_{\underline{X}}(x) := \mathbb{P}(\underline{X} \leq x) \quad \forall x \in \mathbb{R},$$

Ex: En slant singlas 2 ggr.

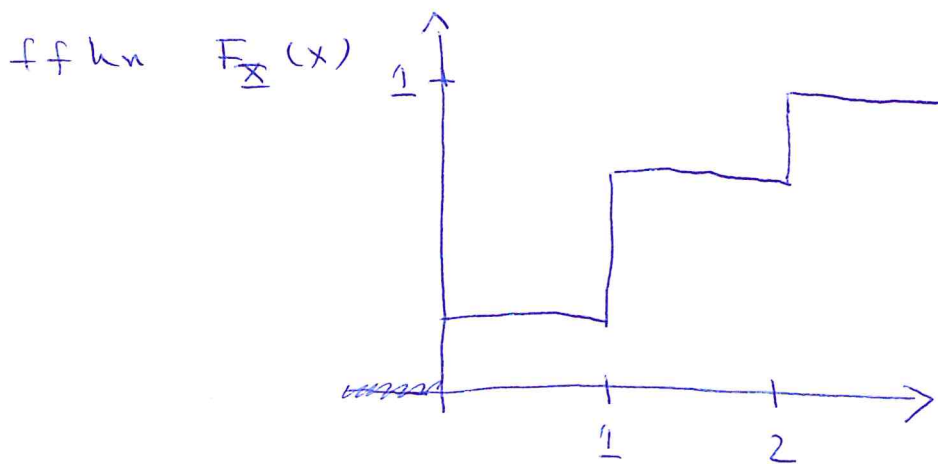
Om $\underline{X} = \#H$, hitta tfn och skissa

f.f.k.n.

2: $\bar{X}$ kan anta värdena 0, 1, 2.

Vi har att

$$P(\bar{X}=k) = \begin{cases} \frac{1}{4} & \text{om } k=0 \\ \frac{1}{2} & \text{om } k=1 \\ \frac{1}{4} & \text{om } k=2, \end{cases}$$



OBS:

$$\begin{aligned} F_{\bar{X}}(1.5) &= P(\bar{X} \leq 1.5) \\ &= P(\bar{X}=0 \vee \bar{X}=1) \\ &= P(\bar{X}=0) + P(\bar{X}=1). \end{aligned}$$

Sats: För en f.f.k.n. gäller alltid att

1/ $f_{\bar{X}}(x) \geq 0 \quad \forall x$

2/ $P(\bar{X} \in A) = \sum_{x \in A} f_{\bar{X}}(x) \quad \text{för } A \subset \mathbb{R}$

3/ $\sum_{\text{alla } x} f_{\bar{X}}(x) = 1$

B: 1/ $f_{\bar{X}}(x) = P(\bar{X}=x) \geq 0$

2/ $P(\bar{X} \in A) = P\left(\bigcup_{x \in A} \{\bar{X}=x\}\right) = \sum_{x \in A} P(\bar{X}=x)$

3/ Tag $A = S$ i 2/.

□