Sannolikhetsteori

Sannolikhetsläran kan se som teorin om *slumpmässiga försök*

Med ett *slumpmässigt försök* menas ett försök vars resultat inte säkert kan förutsägas

Klassiska exempel: Slå en tärning, drag 5 kort ur en kortlek

Definitioner:

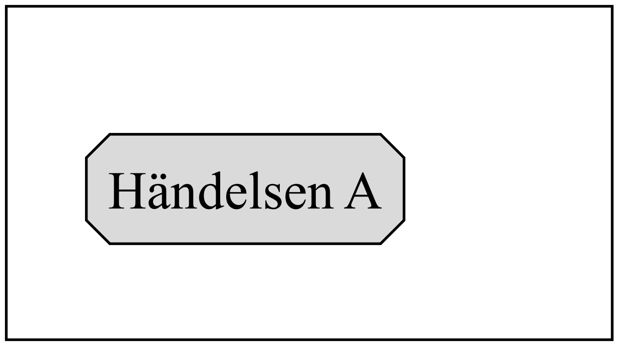
- Resultatet av ett slumpmässigt försök kallas ett **utfall**
- Mängden av alla möjliga utfall av ett slumpmässigt försök kallas **utfallsrum**, (betecknas med Ω)
- En samling utfall kallas *händelse (A, B, C,….),* delmängd av
- Om ett utfallsrum, Ω, är ändligt eller uppräkneligt sägs Ω vara ett *diskret utfallsrum*.
- Om *Ω* inte är ändligt eller uppräkneligt sägs Ω vara ett *kontinuerligt utfallsrum*.

Utfallsrum, händelse och komplement

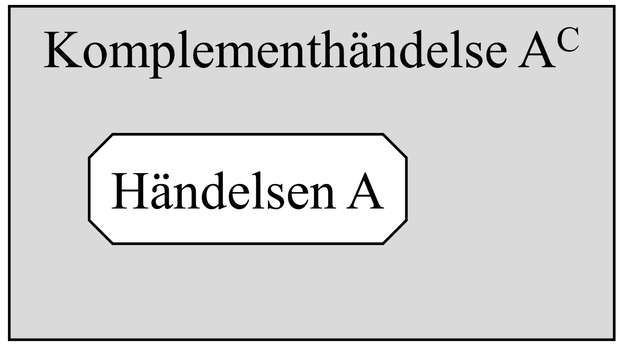
Då man undersöker händelser kan man med fördel använda mängdlärans symboler.

Grundmängd

Delmängd



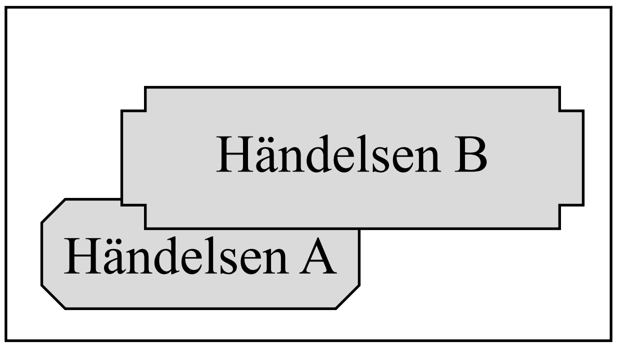
Komplementet till A



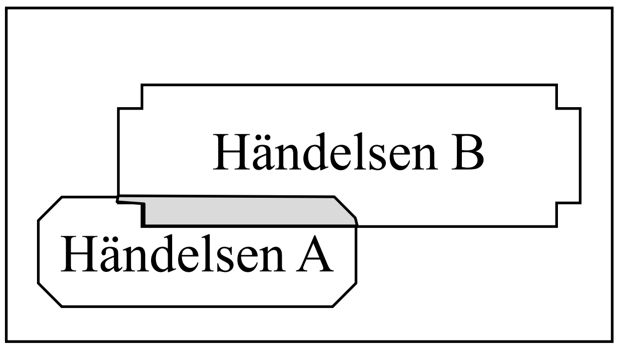
Unions-, snitt- och disjunkta händelser

Unionshändelse, $A \cup B$	**Minst en av händelserna A och B inträffar,** **A eller B inträffar**
Snitthändelse, $A \cap B$	**både A och B inträffar**
Disjunkta händelser	**A och B kan ej inträffa samtidigt, $A \cap B = \emptyset$** **($\emptyset$ = tomma mängden)**

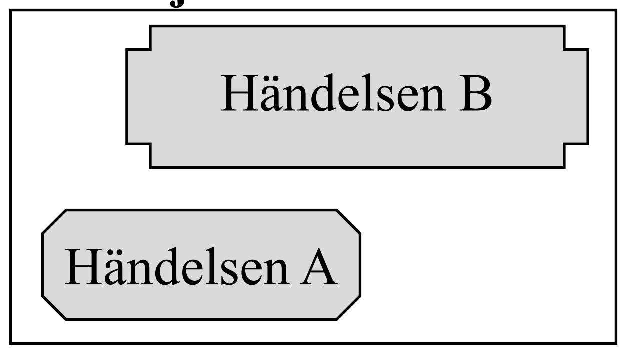
Unionshändelse



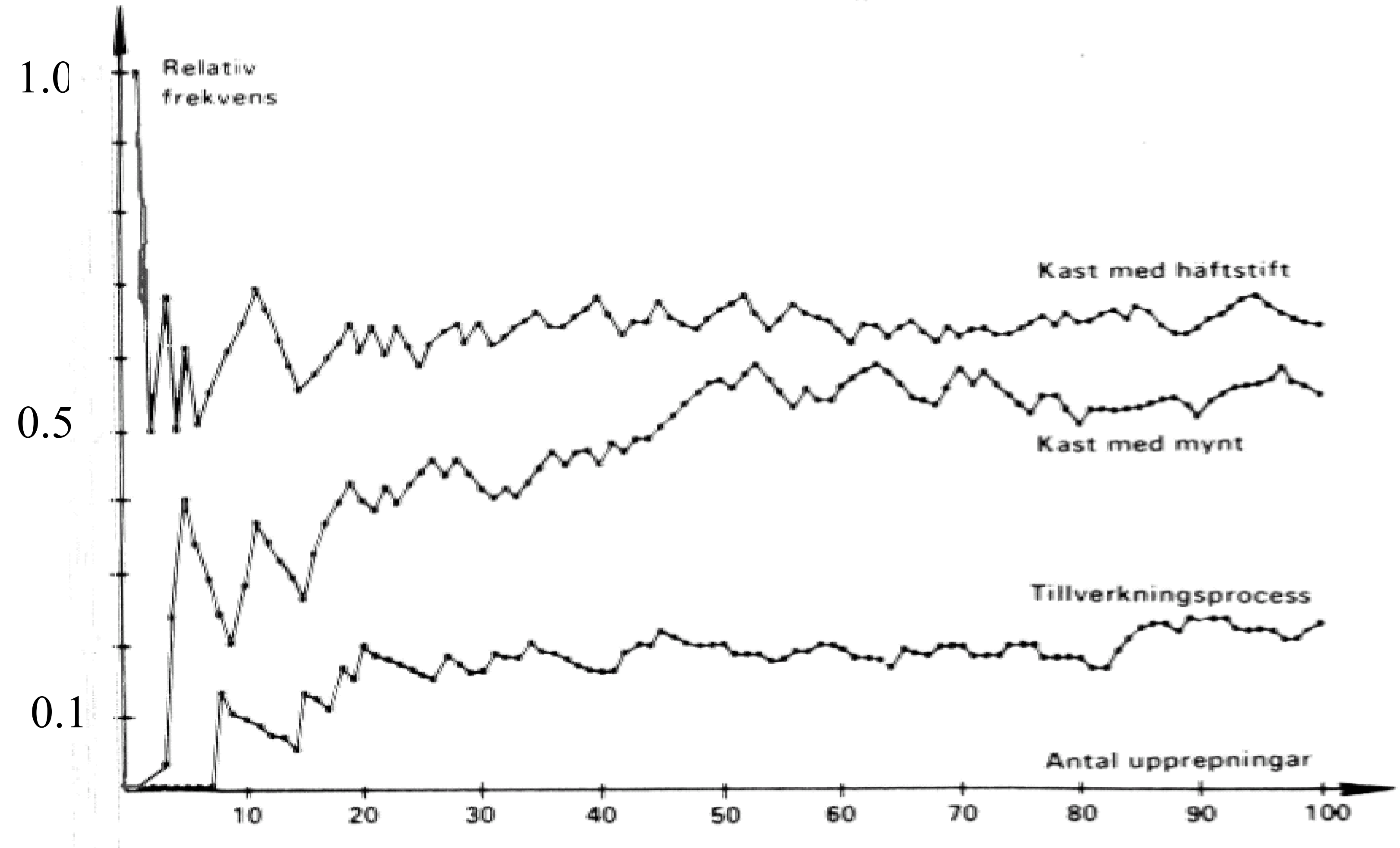
Snitthändelse



Disjunkta händelser



Sannolikhet som relativ frekvens



Definition av sannolikhet

Kolmogorovs axiomsystem

- En funktion $\boldsymbol{P}$, som till varje händelse $\boldsymbol{A}$ i utfallsrummet Ω, ordnar ett reellt tal $\boldsymbol{P(A)}$, är ett sannolikhetsmått, ($\boldsymbol{P(A)}$ kallas sannolikheten för $\boldsymbol{A}$), om $\boldsymbol{P}$ har följande egenskaper (sannolikhetsteorins axiomsystem):
  - 1. $\mathbf{0} \leq \boldsymbol{P(A)} \leq \mathbf{1}$, för alla $\boldsymbol{A}$
  - 2. $\boldsymbol{P(\Omega)} = \mathbf{1}$
  - 3. $\boldsymbol{P(A \cup B) = P(A) + P(B)}$, om $\boldsymbol{A}$ och $\boldsymbol{B}$ är disjunkta

Klassiska (sats 4)

- Det finns $\boldsymbol{m}$ möjliga utfall med lika sannolikhet. Om händelse $\boldsymbol{A}$ innefattar $\boldsymbol{g}$ av utfallen, så blir sannolikheten för händelsen $\boldsymbol{A}$
  - $\boldsymbol{P(A) = g/m}$

Räkneregler för sannolikheter

Sats 1 (Komplementsatsen)

- $P(A) = 1 - P(A^C) \leftrightarrow P(A^C) = 1 - P(A)$

Sats 2 (Additionssatsen, 2 händelser)

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

Sats 3

- $P(A \cup B) \leq P(A) + P(B)$ (Booles olikhet)

Kombinatorik
Multiplikationsprincipen

- Totalt ska k stycken operationer (moment) utföras. Den första operationen kan utföras på n_1 sätt, den andra på n_2 sätt och den i:te på n_i sätt. Då antalet sätt de k operationerna kan utföras på

$$n_1 \times n_2 \times \ldots \times n_k$$

Operation 1 n_1 olika sätt	Operation 2 n_2 olika sätt	…	Operation k n_k olika sätt

Urval utan/med återläggning och utan/med hänsyn till ordningen

- Dragning ***med återlägg*** av $\boldsymbol{n}$ element ur ***N*** ***med hänsyn till ordningen*** kan ske på $\boldsymbol{N^n}$ olika sätt.
- Dragning ***med återlägg*** av *n* element ur *N* ***utan hänsyn till ordningen*** kan ske på $\binom{\boldsymbol{N+n-1}}{\boldsymbol{n}}$ olika sätt.
- Dragning ***utan återlägg*** av *n* element ur *N* ***med hänsyn till ordningen*** kan ske på ***N(N-1)….(N-n+1)*** = $\boldsymbol{N_{(n)}}$ olika sätt. *(Permutationer)*
- Dragning ***utan återlägg*** av *n* element ur *N* ***utan hänsyn till ordningen*** kan ske på $\binom{\boldsymbol{N}}{\boldsymbol{n}}$ sätt. (*Kombinationer)*

Betingad sannolikhet

Definition:

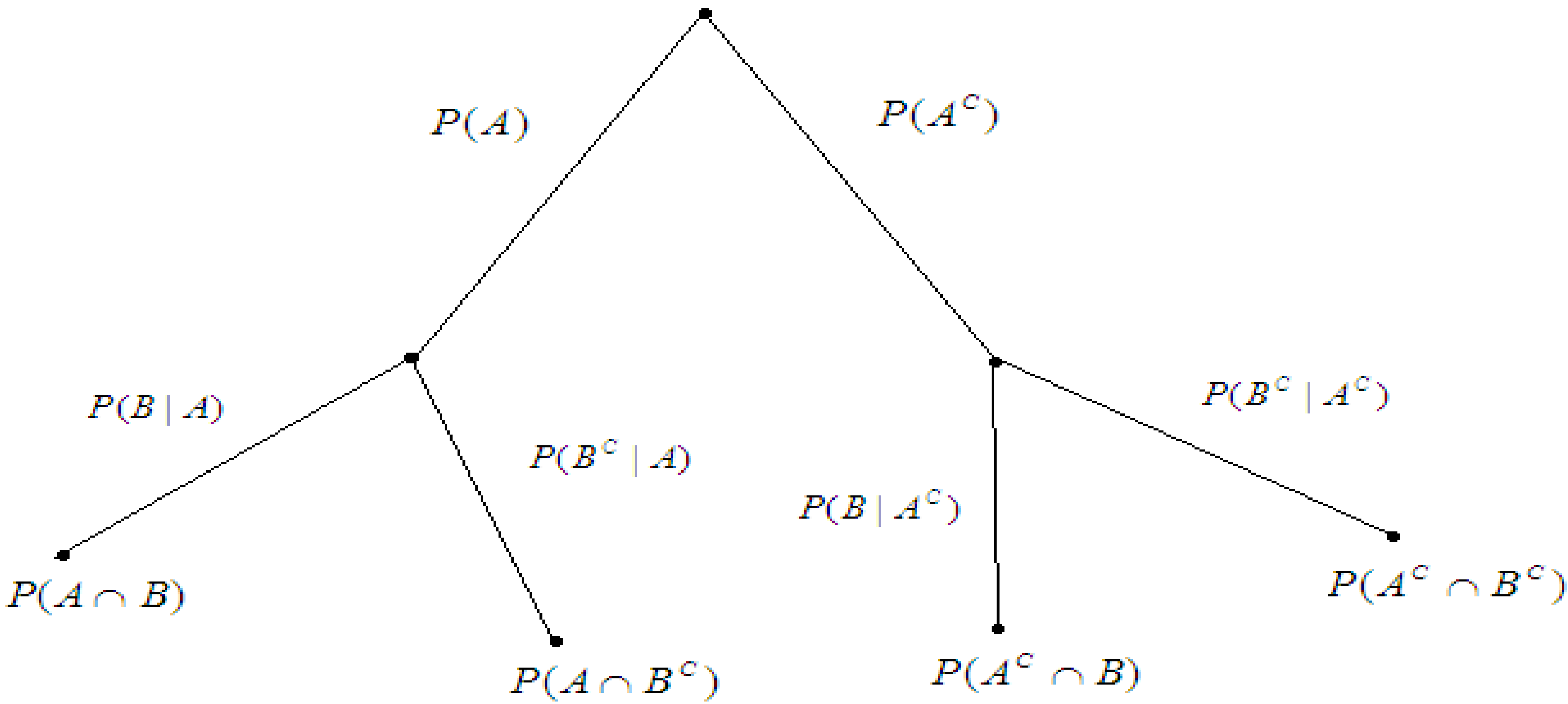
Med den betingade sannolikheten ***P(A|B)*** menas sannolikheten för händelsen A givet att händelsen B har inträffat:

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

definitionen kan skrivas om som

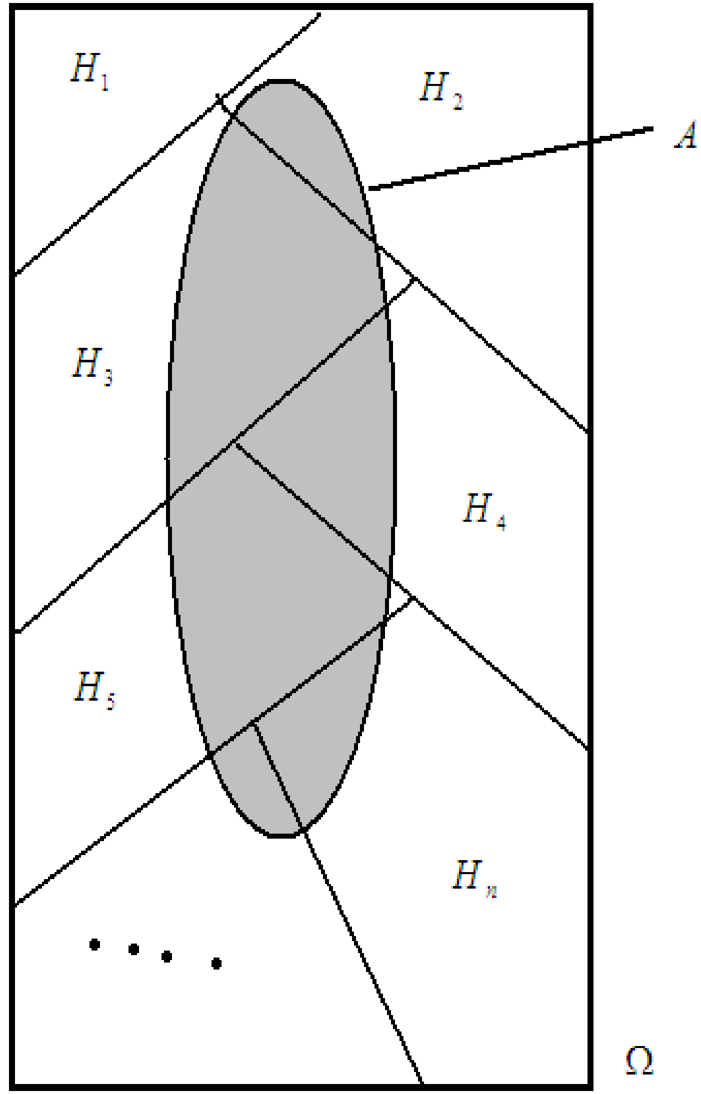
$$\boldsymbol{P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)}$$

Betingad sannolikhet



Satsen om total sannolikhet och Bayes sats (sats 5 & 6)

Om händelserna $H_1, H_2, \ldots\ldots, H_n$ är parvis oförenliga och tillsammans fyller hela Ω



dvs $H_i \cap H_j = \Phi$ och $\bigcup_{i=1}^{n} H_i = \Omega$

då gäller för varje händelse A att

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A|H_i) \text{ (total sannolikhet)}$$

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{j=1}^{n} P(H_j)P(A|H_j)} \text{ (Bayes sats)}$$

Oberoende händelser

Definition

Två händelser är oberoende av varandra om

$$\boldsymbol{P(A \cap B) = P(A)P(B)}$$

Exempel på oberoende händelser: flera kast med en tärning.
Om B har inträffat påverkar detta inte sannolikheten för att A ska inträffa.

Om ***A*** och ***B*** är oberoende gäller

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

OBS!! Förväxla inte oberoende och disjunkta händelser!!

Sannolikhet oberoende upprepade försök

- Ett försök upprepas n oberoende gånger. A är en händelse som inträffar med sannolikheten p i ett försök. Sannolikheten att händelsen A inträffar k gånger under de n försöken är

$$P(\text{A inträffar k gånger}) = \binom{n}{k} p^k (1-p)^{n-k}$$

A inträffar k gånger vid de n försöken

Stokastiska variabler i en dimension

Definition (”mjuka varianten”)

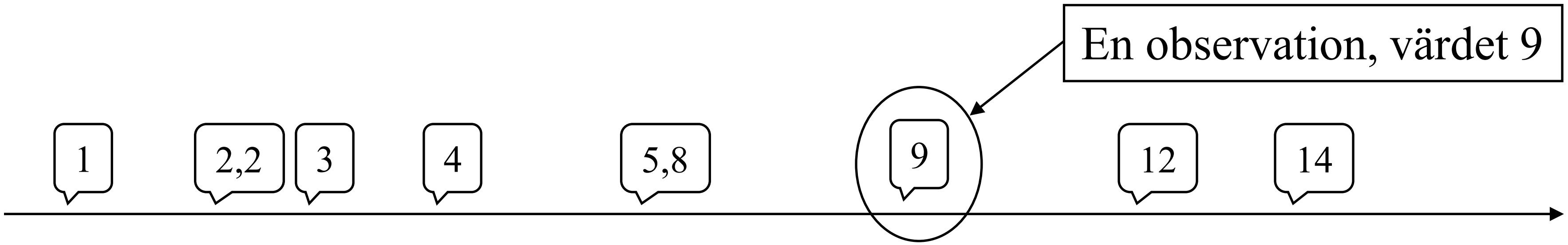
- Utfallet i ett slumpmässigt försök i form av ett reellt tal, betraktat <u>*innan*</u> försöket utförts, kallas för *stokastisk variabel* (ofta betecknad X, Y, Z osv). Resultatet sedan försöket utförts kallas *observerat värde* på en stokastiska variabeln (ofta betecknat x, y, z) .

Definition (matematisk)

- En *stokastisk variabel, X,* är en reellvärd funktion definierad på ett utfallsrum (Ω). $X\colon \Omega \to \mathbb{R}$

Det finns två typer av stokastiska variabler

- En ***diskret*** stokastisk variabel kan anta ett uppräkningsbart antal värden.
  - **kan vara oändligt många**



Samtliga möjliga utfall (utfallsrummet) för någon diskret stokastisk variabel.

- En ***kontinuerlig*** stokastisk variabel kan anta alla värden i ett intervall.
  - **alla kontinuerliga utfallsrum är oändliga**

Sannolikhetsfördelning, sannolikhets- och fördelningsfunktion

Låt X vara en diskret stokastisk variabel som antar värdena:

$x_1 < x_2 < \ldots < x_k < \ldots$

Definitioner

- Med *sannolikhetsfunktionen* (frekvensfunktionen) till X, $p(x_k)$, menas

$$p(x_k) = P(X=x_k)$$

- Med *fördelningsfunktionen till X*, $F(x_k)$, menas

$$F(xk)=P(X \leq x_k)=\sum_{i=1}^{k} P(X=x_k)$$

Det gäller också att: $P(X= x_k) = F(x_k) - F(x_{k-1})$

Likformig fördelning

- Det finns N stycken lika sannolika utfall av ett slumpmässigt försök, slumpvariabeln X *betecknar utfallets ordningsnummer.* X blir då *likformigt fördelad* $P(X=k) = 1/N$, $k= 1,2,3,4,5,\ldots,N$

  Ex: Kasta en symmetrisk tärning en gång.

  Låt X = antal prickar, $\Omega = \{1,2,3,4,5,6\}$ och

  antag att alla utfall är lika sannolika.

  Då blir X likformigt fördelad över Ω.

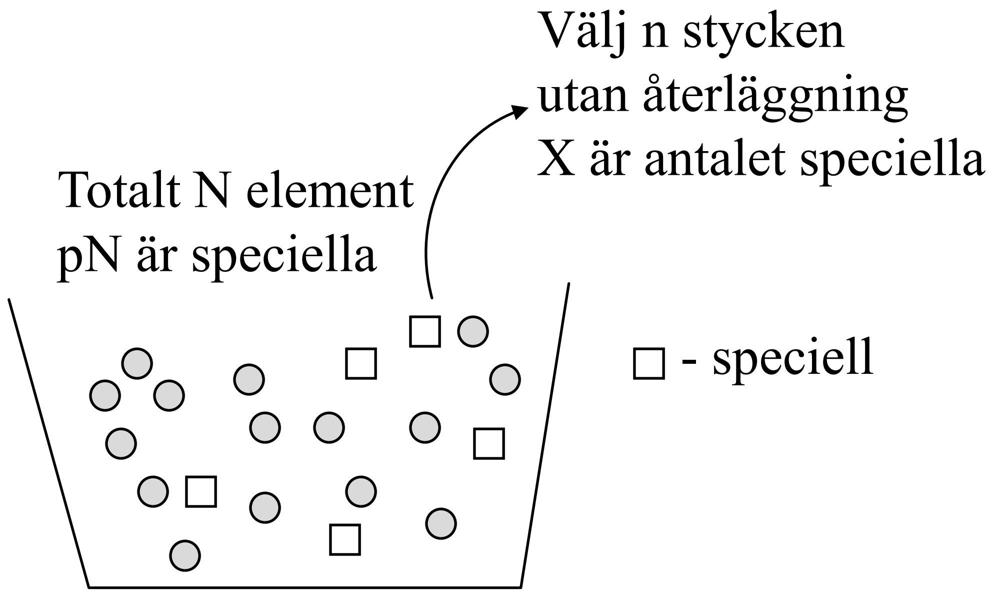
$$\boxed{X \in U(N)}$$

Hypergeometrisk fördelning

- En mängd innehåller totalt N element, av vilka Np är av speciellt slag (andelen speciella är p). Välj slumpmässigt, utan återläggning, ett urval av n element. Slumpvariabeln *X betecknar antalet speciella element i urvalet*. *X* blir då *hypergeometriskt fördelad*.

$$P(X=x)=\frac{\binom{Np}{x}\binom{N-Np}{n-x}}{\binom{N}{n}}$$

x är ett heltal sådant att $0 \le x \le Np$ *och* $0 \le n\text{-}x \le N\text{-}Np$



$$\boxed{X \in Hyp(N, n, p)}$$

Geometrisk fördelning eller ffg-fördelning

- Ett försök består av n upprepningar av oberoende delförsök, (ex: <u>dragning med återlägg</u>). A är en speciell händelse som inträffar med sannolikheten p i varje delförsök. Slumpvariabeln *X betecknar antalet genomförda delförsök tills händelsen A inträffar i för första gången*, *X* blir då *geometriskt fördelad* (eller *ffg-fördelad*.)

$$P(X = x) = p(1 - p)^{x-1} \, , x = 1,2,3,4,........$$

$$\boxed{X \in Geo(p)}$$

Binomialfördelning

- Ett försök består av n upprepningar av oberoende delförsök, (ex: dragning med återlägg). A är en speciell händelse som inträffar med samma sannolikhet, p, i varje delförsök. Slumpvariabeln *X betecknar antalet gånger händelsen A inträffar i hela försöket*. X blir då *binomialfördelad*

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, ..., n$$

A inträffar x gånger vid de n försöken

$$\boxed{X \in Bin(n, p)}$$

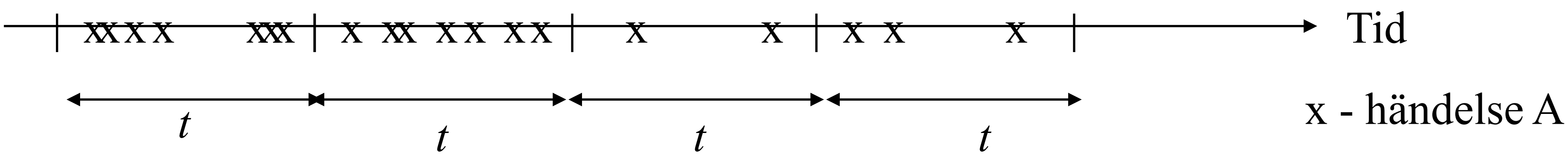
Poissonfördelning

- Betrakta händelser *A* som slumpmässig och oberoende av varandra inträffar i tiden. Slumpvariabeln *X betecknar antalet gånger A inträffar under ett tidsintervall av fix längd.*
  X blir då *poissonfördelad*

$$P(X = x) = e^{-\mu} \frac{\mu^x}{x!} \qquad x = 0, 1, \ldots$$

μ – kallas intensitetsparameter

μ anger genomsnittligt antal händelser A under tiden t



$$\boxed{X \in Po(\mu)}$$

Kontinuerliga stokastiska variabler

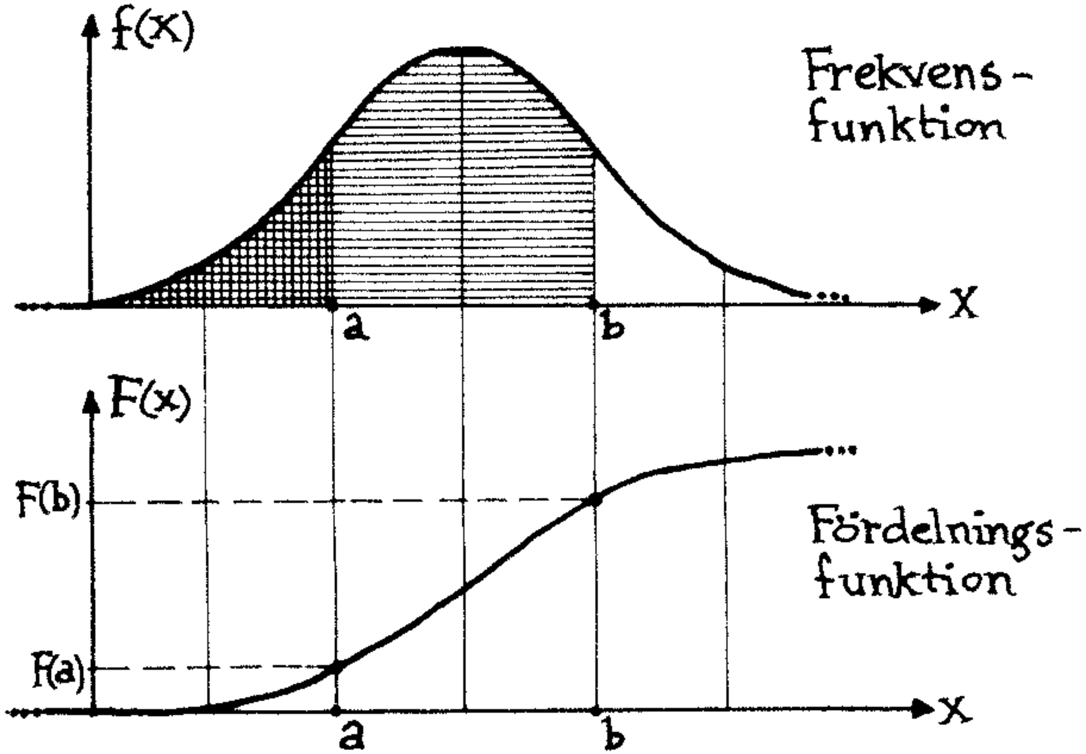
- En kontinuerlig stokastisk variabel kan anta alla värden i ett intervall. Tex: $\Omega = R$, eller $\Omega = \{x : 0 < x < 1\}$
- Utfallen ligger oändligt tätt vilket medför att inget utfall kan antas med positiv sannolikhet.
- Fördelningsfunktionen får följande utseende $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$ (*)

Definition

- Om det finns en funktion *f(x)* så att (*) gäller sägs *X* vara en kontinuerlig stokastisk variabel och *f(x)* kallas täthetsfunktion (frekvensfunktion).

Frekvens- och fördelningsfunktion

- Frekvensfunktionen, *f(x)* (täthetsfunktionen)



- Fördelningsfunktionen

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

f(x) kan ses som fördelning av sannolikhetsmassa.
Sannolikheten för utfall inom x och x+Δx är ungefär f(x)Δx

Täthets-(Frekvens-), *f(x)* och fördelningsfunktion, *F(x)*

$$f(x) \geq 0 \; och \; \int_{-\infty}^{\infty} f(x)dt = 1$$

$$\boxed{\begin{array}{l} 0 \leq P(A) \leq 1 \\ \\ P(\Omega) = 1 \end{array}}$$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \; och \; F'(x) = f(x)$$

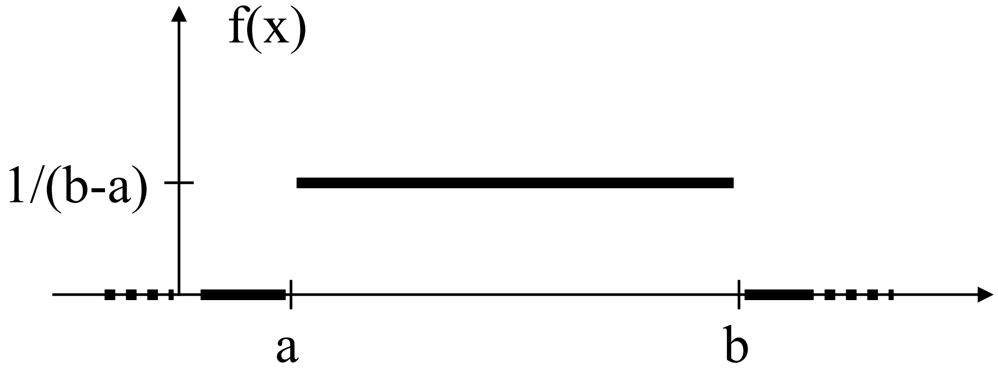
$$P(a < X \leq b) = \int_{a}^{b} f(x)dx = F(b) - F(a)$$

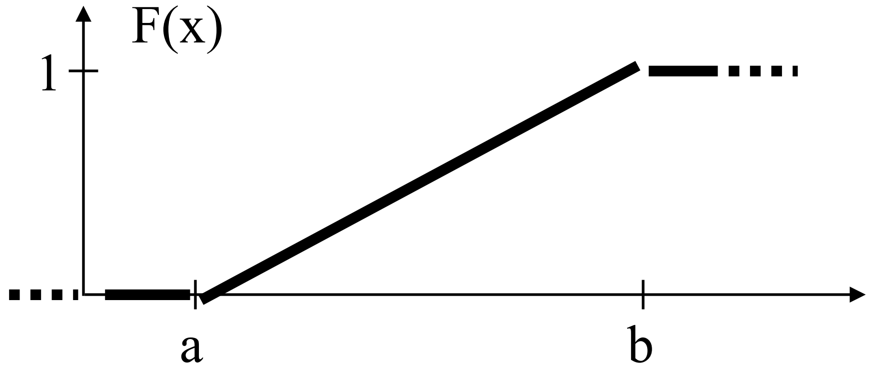
$$P(X > x) = \int_{x}^{\infty} f(t)dt = 1 - F(x)$$

$P(X = 0) = 0$, för alla x

Likformig fördelning

- Om frekvensfunktionen för en s.v. X är konstant i ett intervall, och noll utanför sägs X vara *likformigtfördelad över intervallet*





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

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \dfrac{(x-a)}{(b-a)} & a < x < b \\ 1 & x \geq b \end{cases}$$

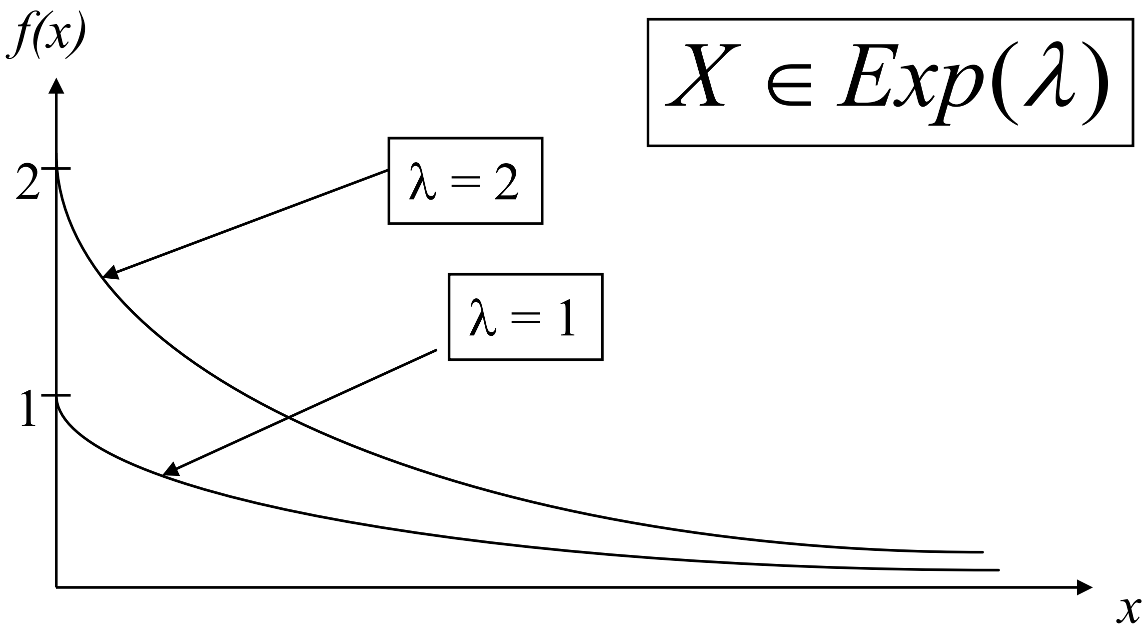
$$\boxed{X \in U(a,b)}$$

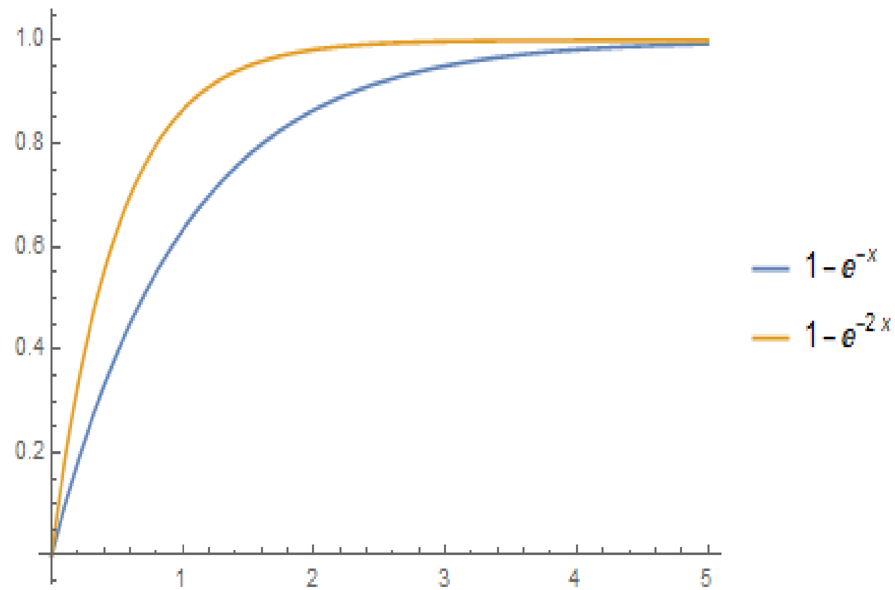
Exponentialfördelning

Om en s.v. X har täthetsfunktion enligt nedan sägs X vara *Exponentialfördelad* λ, $\lambda > 0$.

$$f_X(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases} \qquad F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

$$\boxed{X \in Exp(\lambda)}$$

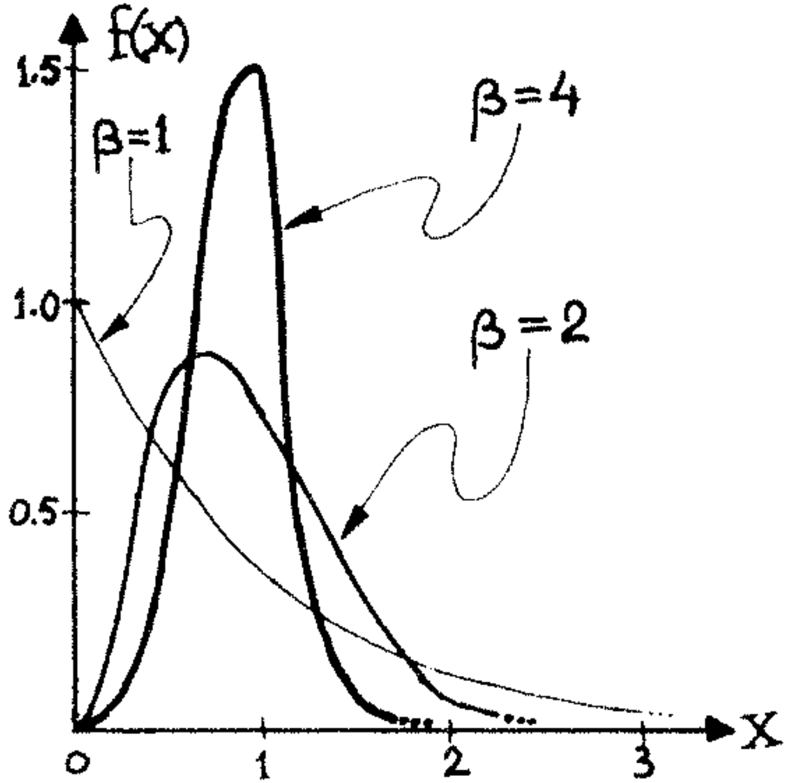




Radioaktivt sönderfall är typiskt exponentialfördelat
Livslängder på elektronikkomponenter är ofta exponentialfördelade

Weibullfördelningen

- En s.v. sägs vara Weibullfördelad om den har frekvens- och fördelningsfunktion enligt nedan ($c > 0$ & $\lambda > 0$)



$$f_X(x) = \begin{cases} 0, & x < 0 \\ \lambda c(\lambda x)^{c-1} e^{-(\lambda x)^c}, & x \geq 0 \end{cases}$$

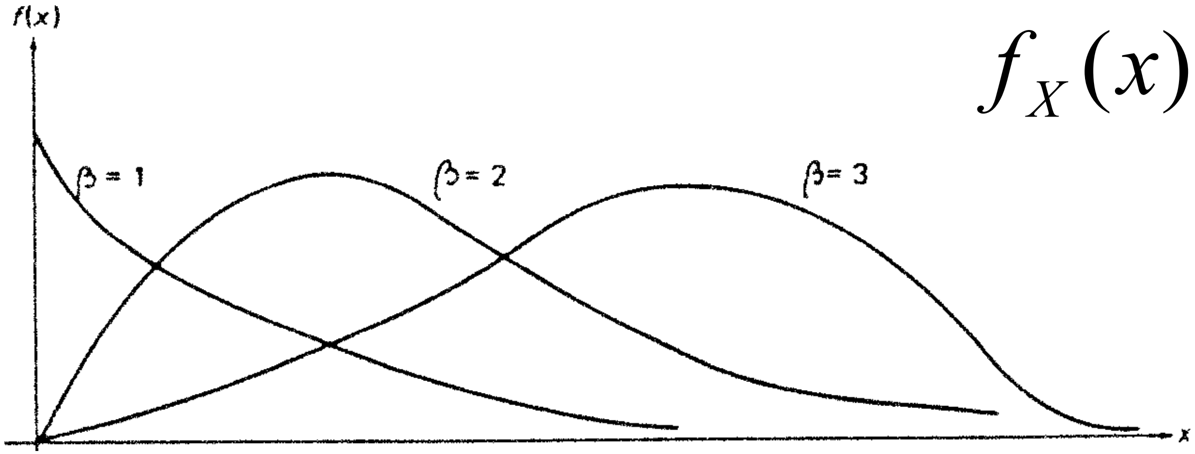
$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-(\lambda x)^c}, & x \geq 0 \end{cases}$$

$$\boxed{X \in Wei(\lambda, c)}$$

Weibullfördelningen är vanlig inom tillförlitlighetsteori

Gammafördelning

- En s.v. X sägs vara gammafördelad om den har täthetsfunktion enligt nedan ($c > 0$ & $\lambda > 0$)



$\lambda = 1$

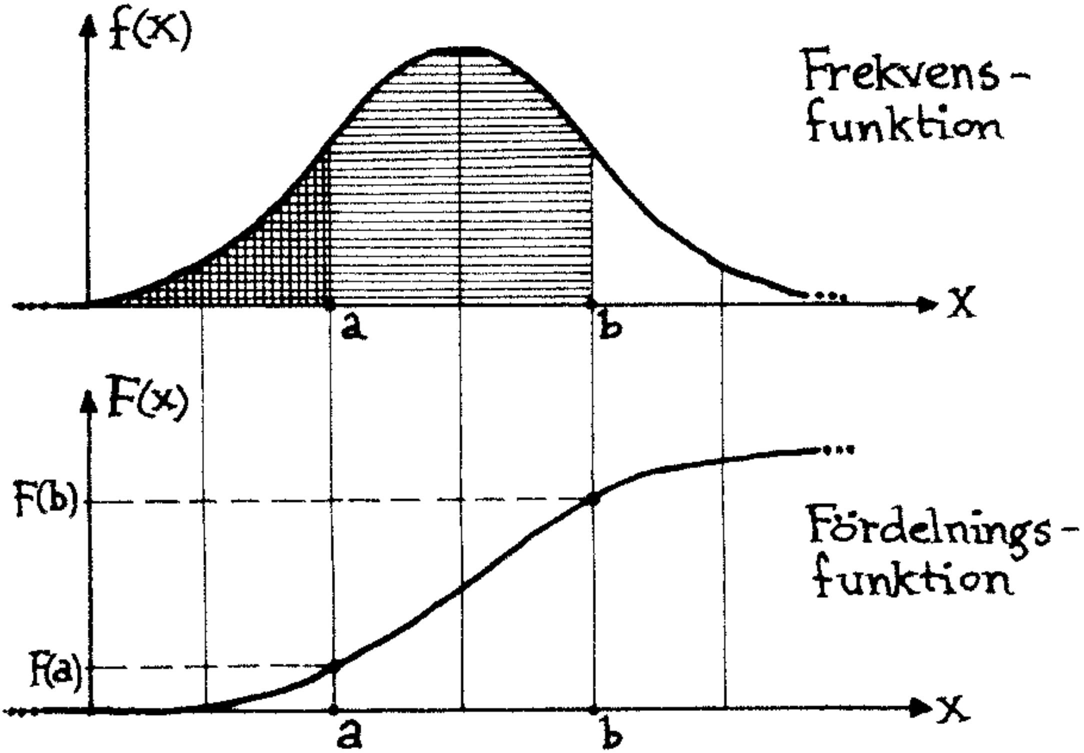
$$f_X(x) = \begin{cases} 0, & x < 0 \\ \dfrac{\lambda^c}{\Gamma(c)} x^{c-1} e^{-\lambda x}, & x \geq 0 \end{cases}$$

$$X \in \Gamma(\lambda, c)$$

$$\Gamma(c) = \int_0^\infty x^{c-1} e^{-x} dx$$

Normalfördelningen

- En s.v. X sägs vara normalfördelad om den har täthetsfunktion
  Den bestäms av två parametrar:
  Lägesparametern μ samt spridningsparametern σ



$$\boxed{X \in N(\mu, \sigma)}$$

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

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/(2\sigma^2)} dt$$

Normalfördelningen

- För normalfördelningen är *F(x)* bara möjlig att beräkna med numeriska metoder (den går inte att lösa algebraiskt)

  Därför finns tabeller för *N(0,1)*, vilken har fördelningsfunktionen

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$$\boxed{\text{Om } X \in N(\mu, \sigma) \text{ så gäller att } P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)}$$

$$\boxed{\frac{X-\mu}{\sigma} \in N(0,1)}$$

$$\boxed{\Phi(-x) = 1 - \Phi(x)}$$

Stokastiska variabler i två dimensioner

Definition

- En *stokastisk variabel, (X,Y),* är en reellvärd funktion definierad på ett utfallsrum (Ω). $(X,Y): \Omega \rightarrow \mathcal{R}$

Definition

- Funktionen

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

kallas *fördelningsfunktionen* för (X,Y).

- Om (X,Y) båda antar ett uppräkneligt antal värden sägs (X,Y) vara *diskret*.

Definition

Med den *simultana sannolikhetsfunktionen* för en diskret tvådimensionell stokastisk variabel (X,Y) menas

$$p_{X,Y}(j,k) = P(X = j, Y = k) \quad , j \in \Omega_X, k \in \Omega_Y$$

Den *marginella sannolikhetsfuktionen* för X fås ur

$$p_X(j) = \sum_{k=0}^{\infty} p_{X,Y}(j,k)$$

Dessutom gäller

$$\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} p_{X,Y}(x,y) = 1$$