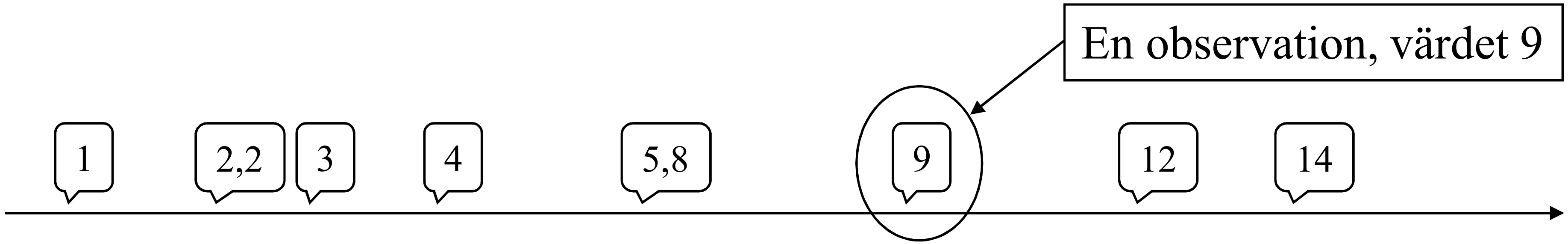
Diskreta stokastiska variabler

Definitioner:

- Utfallet i ett slumpmässigt försök i form av ett reellt tal, betraktat innan försöket utförts, kallas för stokastisk variabel eller slumpvariabel (ofta betecknad ξ, η)
- Ett resultat av försöket (utfall av slumpvariabeln) kallas för observerat värde eller observation (ofta betecknat x eller y)
- En stokastisk variabel som kan anta endast ett ändligt eller uppräkningsbart antal värden, så kallas diskret
  - kan vara oändligt många



Samtliga möjliga utfall (utfallsrummet)

Beskrivning av en diskret stokastisk variabel

En diskret stokastisk variabel, ξ, beskrivs med dess sannolikhetsfördelning

En sannolikhetsfördelning ska innefatta dels alla värden, x, den kan anta och dels sannolikheten för respektive värde, $P(\xi=x)$

Exempel: tärningskast

Utfallsrum: $\{1, 2, 3, 4, 5, 6\}$; $P(\xi=x) = 1/6$ för alla x

Exempel: slantsingling (ej högkant)

Utfallsrum: $\{0 = (krona), 1 = (klave)\}$; $P(\xi=x) = ½$ för alla x

Jämför: $A = krona \Leftrightarrow \xi = 0$

Sannolikhetsfördelning, sannolikhets- och fördelningsfunktion

- Låt ξ vara en diskret stokastisk variabel som antar värdena: $x_1 < x_2 < \ldots < x_k < \ldots$

- Sannolikhetsfunktionen (frekvensfunktionen) till ξ, $p(x_k)$, definieras som

$$p(x_k) = P(\xi = x_k)$$

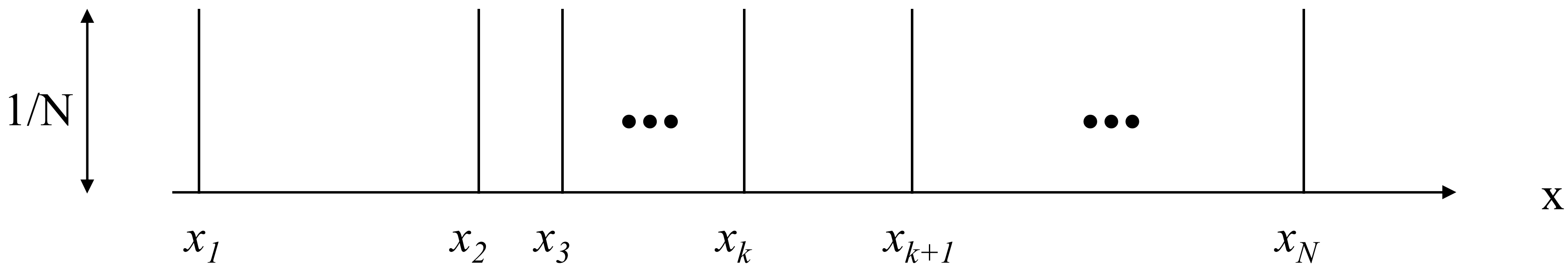
- Fördelningsfunktionen till ξ, $F(x_k)$, definieras som

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

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

Likformig fördelning

- Det finns *N* stycken lika sannolika utfall
  - $P(\xi = x_i) = 1/N$
  - N kallas för den likformiga fördelningens parameter



N stycken lika sannolika utfall

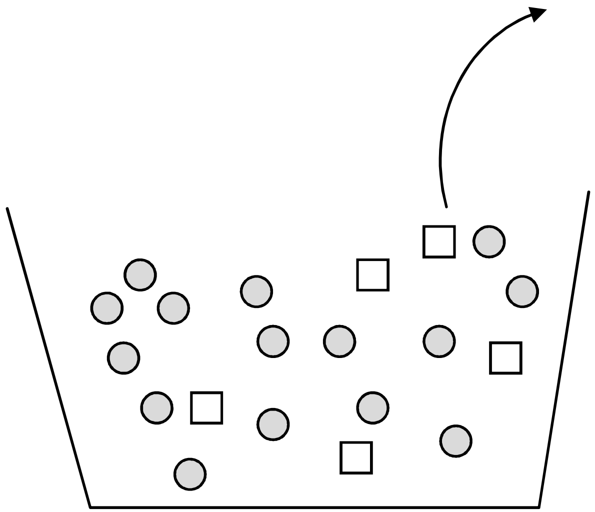
Exempel: antal ögon vid ett tärningskast

Hypergeometrisk fördelning

◆ En mängd innehåller totalt N element, av vilka Np är av speciellt slag (andelen speciella är p). Välj slumpmässigt, **utan återlägg**, ett urval av n element.
Låt ξ betecknar antalet speciella element i urvalet.
ξ är då hypergeometriskt fördelad

Välj n stycken
utan återläggning,
ξ är antalet speciella

Totalt N element
Np är speciella



□ - speciell

Hypergeometrisk fördelning

Med den klassiska definitionen av sannolikhet fås:

$$P(\xi = k)\,\frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}}$$

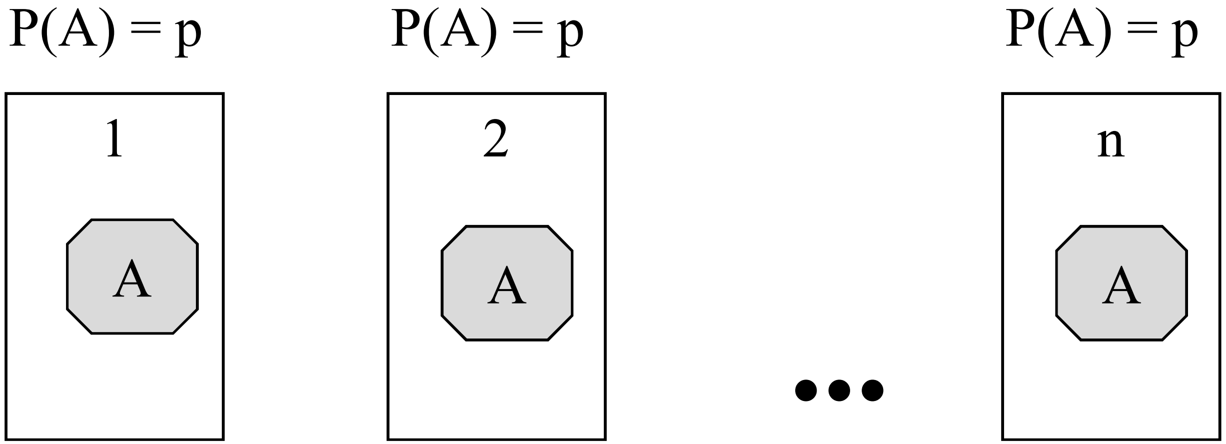
x är ett heltal sådant att $0 \leq k \leq Np$
och $0 \leq n\text{-}k \leq N\text{-}Np$

$$\boxed{\xi \epsilon Hyp(N; n; p)}$$

Binomialfördelning

- Ett försök består av n **oberoende upprepningar** av delförsök. A är en speciell händelse som inträffar med **samma sannolikhet** p i varje delförsök. Slumpvariabeln ξ betecknar antalet gånger händelsen A inträffar i hela försöket. ξ är då binomialfördelad

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



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

A inträffar k gånger vid de n försöken

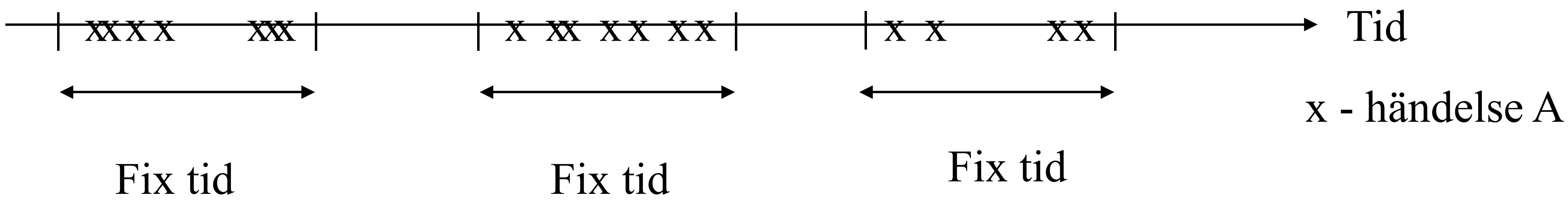
Poissonfördelning

- Betrakta händelser A som inträffar slumpmässiga i tiden och oberoende av varandra. Slumpvariabeln ξ betecknar antalet händelser A under ett tidsintervall av fix längd. ξ är då poissonfördelad

$$P(\xi = x) = e^{-\lambda}\frac{\lambda^x}{x!}, x = 0, 1, 2, ...$$

$$\boxed{\xi \in Po(\lambda)}$$

  – λ är genomsnittligt antal händelser A under intervallet

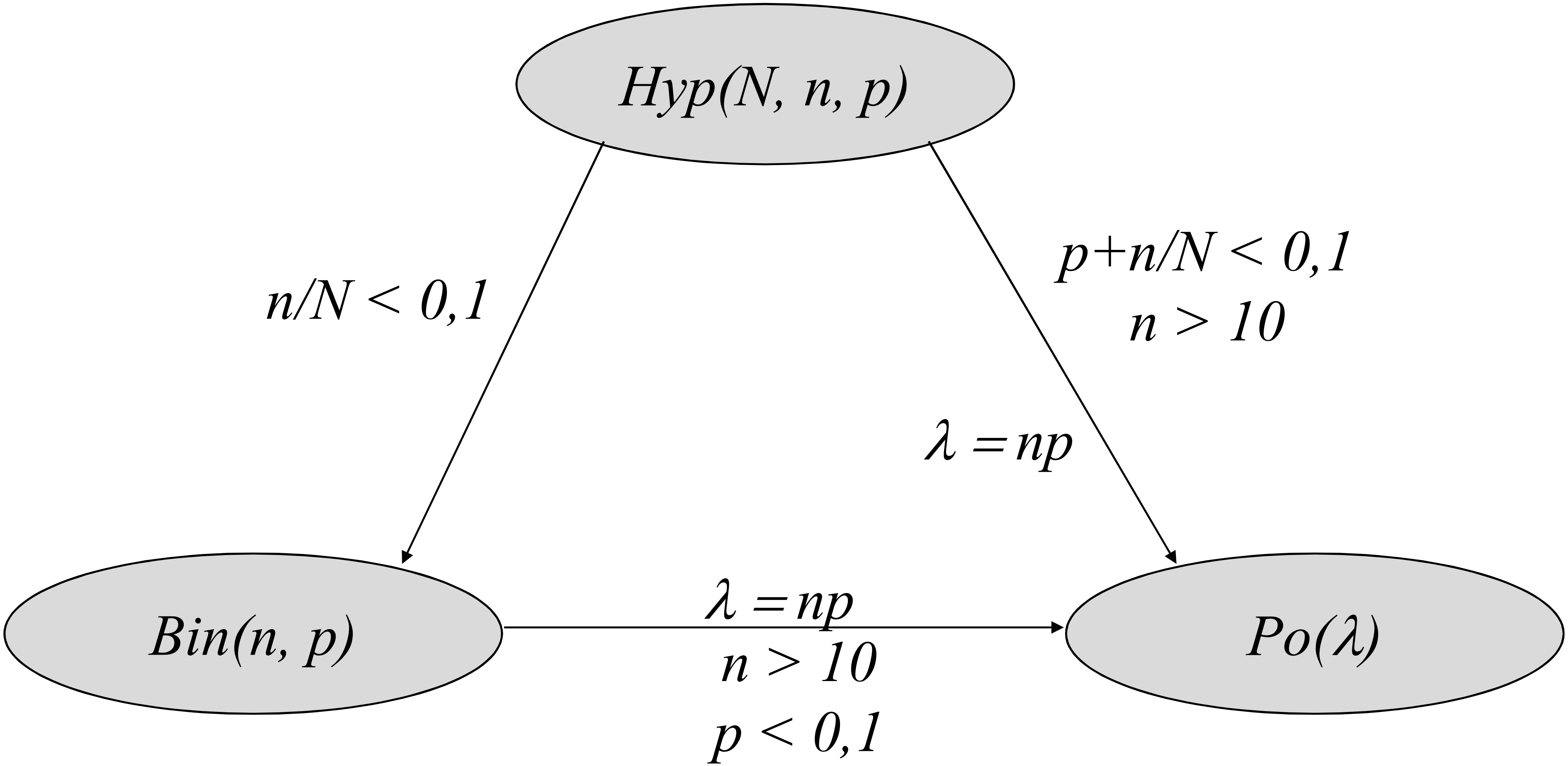


Diskreta fördelningar med *Mathematica*

- Om $\boldsymbol{\xi \epsilon Hyp(N; n; p)}$
  $P(\xi = x)$ = ***PDF[HypergeometricDistribution[n,Np,N],x]***
  $F(x) = P(\xi \leq x)$ =
  = ***CDF[HypergeometricDistribution[n,Np,N],x]***
- Om $\boldsymbol{\xi \epsilon Bin(n; p)}$
  $P(\xi = x)$ = ***PDF[BinomialDistribution[n,p],x]***
  $F(x) = P(\xi \leq x)$ = ***CDF[BinomialDistribution[n,p],x]***
- Om $\boldsymbol{\xi \epsilon Po(\lambda)}$
  $P(\xi = x)$ = ***PDF[PoissonDistribution[λ],x]***
  $F(x) = P(\xi \leq x)$ = ***CDF[BinomialDistribution[λ],x]***

Hypergeometrisk, binomial- och poissonfördelning

Approximationsregler



Kontinuerliga stokastiska variabler

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

Definition

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

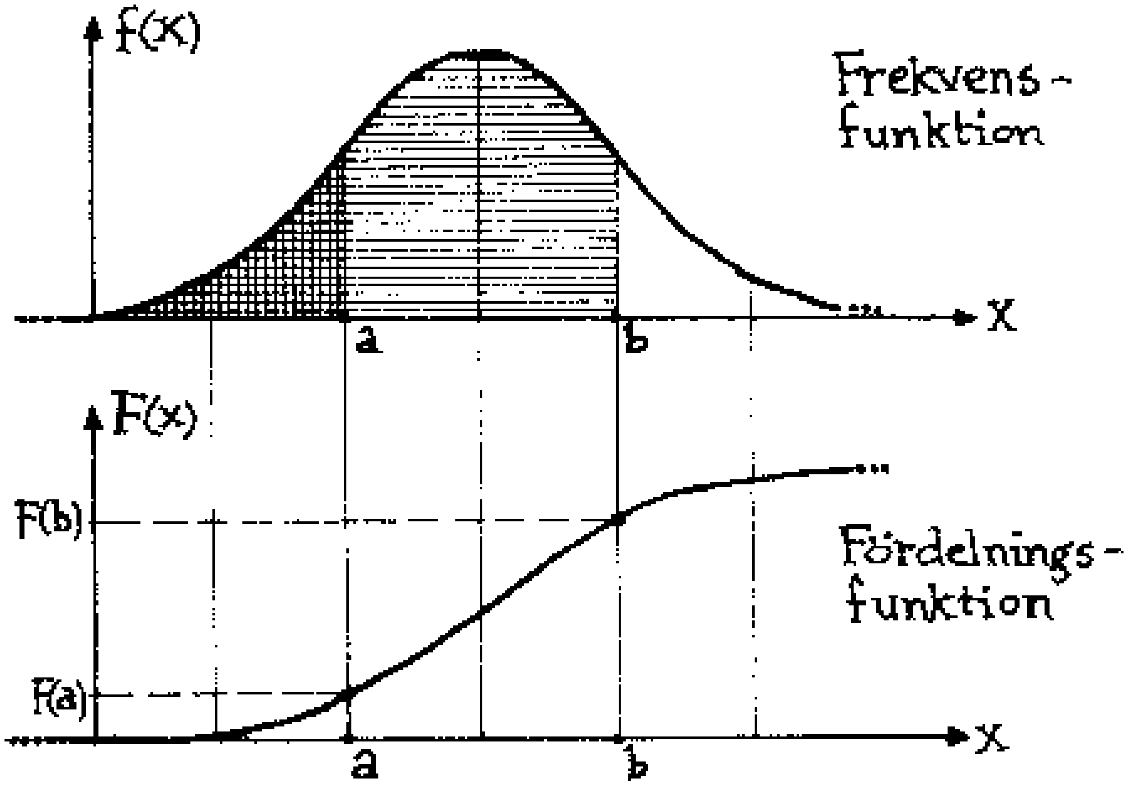
Frekvens- och fördelningsfunktion

- Frekvensfunktionen definieras genom

$$P(\xi \le x) = \int_{-\infty}^{x} f(t)\,dt$$

- Fördelningsfunktionen definieras som

$$F(x) = P(\xi \le x) = \int_{-\infty}^{x} f(t)\,dt$$



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

Frekvens- och fördelningsfunktion

Egenskaper:

$$f(x) \geq 0 \quad \text{och} \quad \int_{-\infty}^{\infty} f(t)\,dt = 1$$

$$F(x) = P(\xi \leq x) = \int_{-\infty}^{x} f(t)\,dt \quad \text{och} \quad F'(x) = f(x)$$

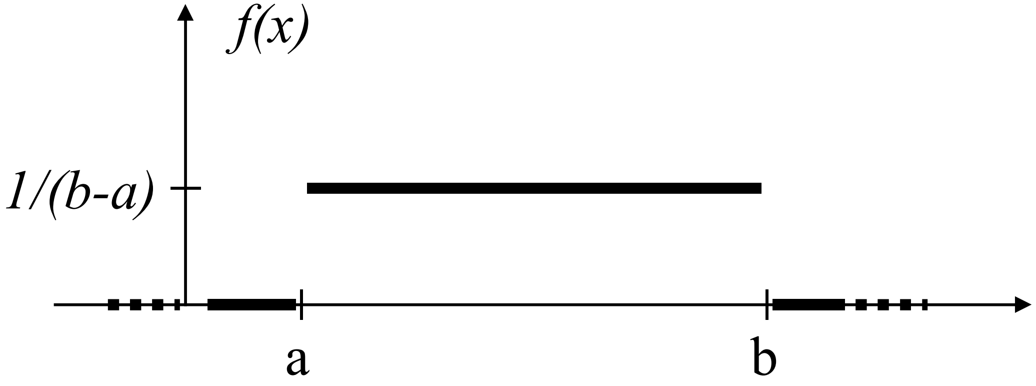
$$P(a < \xi \leq b) = \int_{a}^{b} f(t)\,dt = F(b) - F(a)$$

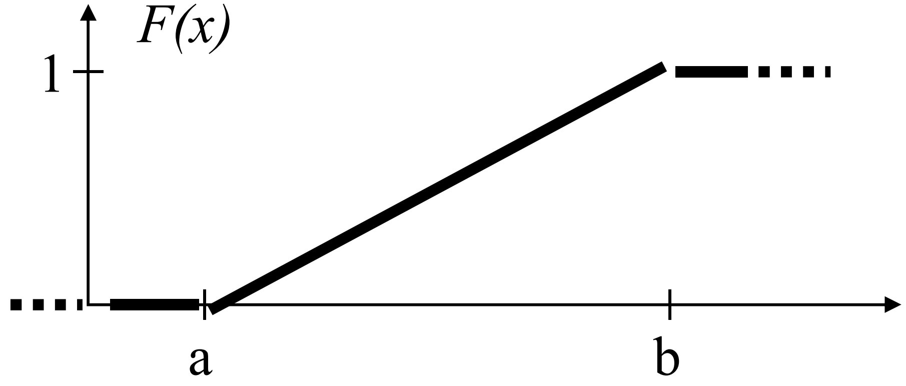
$$P(\xi > x) = \int_{x}^{\infty} f(t)\,dt = 1 - F(x)$$

$$P(\xi = x) = 0, \; \textit{för} \;\; \textit{alla} \;\; x$$

Rektangelfördelning (likformig)

- Frekvensfunktionen är konstant i ett intervall, och noll utanför



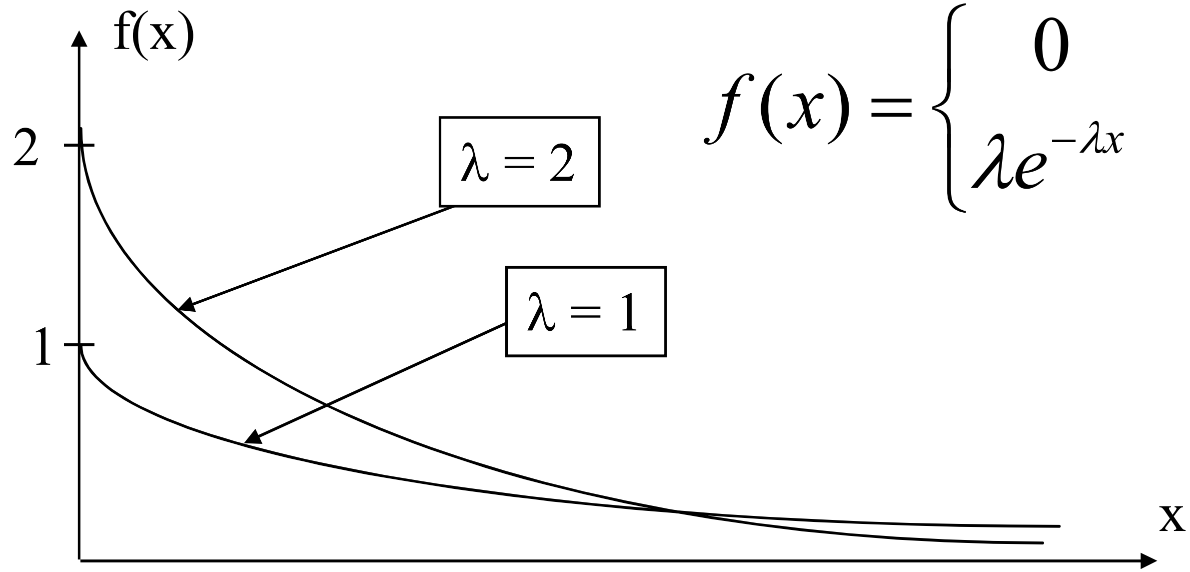


$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \textit{för övrigt} \end{cases} \qquad F(x) = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

$$\xi \in R(a,b) \text{ eller } \xi \in U(a,b)$$

Exponentialfördelning

- En stokastisk variabel är Exponentialfördelad om den har frekvens- och fördelningsfunktion enligt nedan



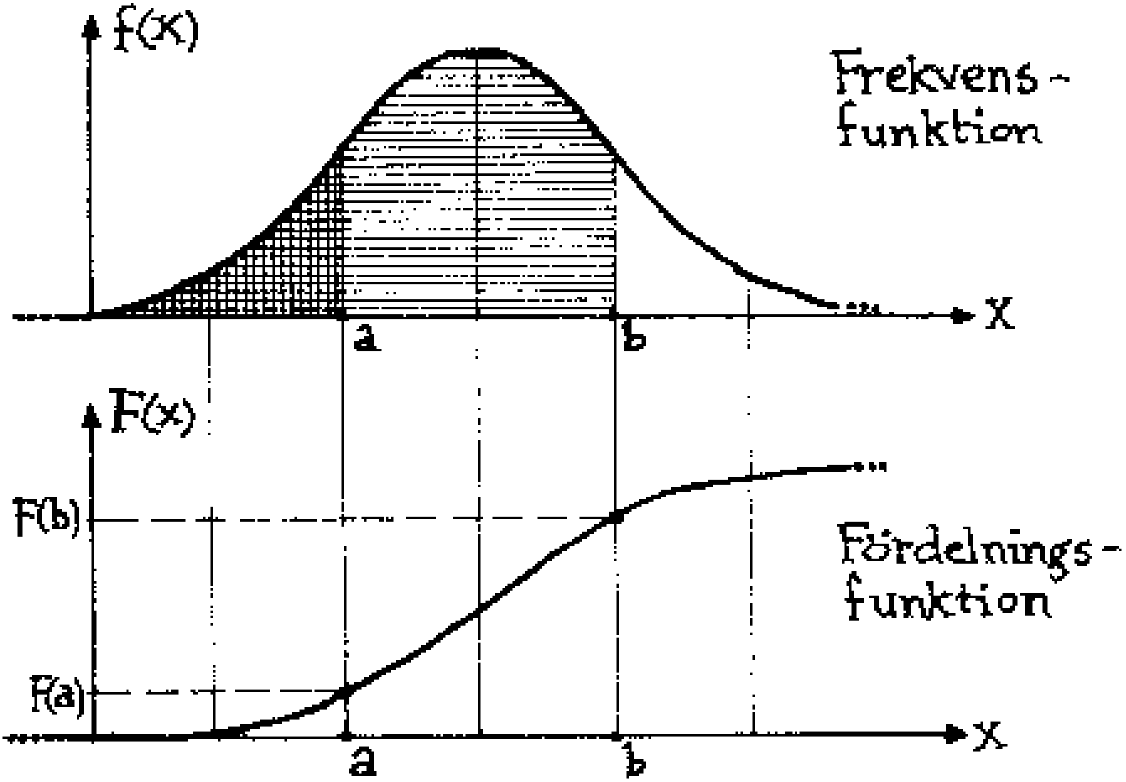
$$f(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases} \qquad F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

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

Exempel där exponentialfördelningen kan användas som modell:
Tiden mellan 2 α-partiklar vid radioaktivt sönderfall.
Livslängder på elektronikkomponenter

Normalfördelningen

- Normalfördelningen är vanligt förekommande
  - Den bestäms av två parametrar, väntevärde, μ, samt standardavvikelse, σ
  - $-\infty < \mu < \infty,\ \sigma > 0$
  - Utfallsrum: $-\infty < x < \infty$



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

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

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

Normalfördelningen

- För normalfördelningen är *F(x)* omöjlig att beräkna utan numeriska metoder (den går inte att lösa ut algebraiskt)

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

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

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

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

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

Normalfördelningen

Tabellen ger sannolikheten $\Phi(x) = P(\xi \leq x)$, där $\xi \in N(0,1)$.
För negativa x-värden använd relationen $\Phi(-x) = 1 - \Phi(x)$.

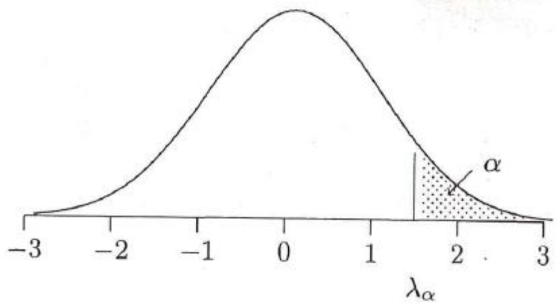
x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.791	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.997	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

Normalfördelningen (forts.)

Tabellen ger det λ_α-värde för vilket $P(\xi > \lambda_\alpha) = \alpha$, där $\xi \in N(0,1)$.

α	0.1	0.05	0.025	0.01	0.005	0.001
λ_α	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902

α	0.0005	0.0001	0.00005	0.00001	5.0×10^{-6}	1.0×10^{-6}
λ_α	3.2905	3.7190	3.8906	4.2649	4.4172	4.7534



Kontinuerliga fördelningar med *Mathematica*

- Om $\xi \epsilon\, \boldsymbol{U}(\boldsymbol{a}; \boldsymbol{b})$

  $F(x) = P(\xi \ \leq \ x) =$

  = CDF[UniformDistribution[{a,b}],x]

- Om $\boldsymbol{\xi \epsilon Exp}(\boldsymbol{\lambda})$

  $F(x) = P(\xi \ \leq \ x) =$

  =CDF[ExponentialDistribution[λ],x]

- Om $\boldsymbol{\xi \epsilon N}(\boldsymbol{\mu};\boldsymbol{\sigma})$

  $F(x) = P(\xi \ \leq \ x) =$

  = CDF[NormalDistribution[μ;σ],x]