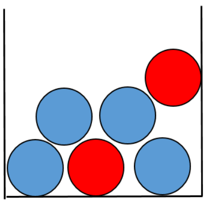
Binomisk Sannsynlighetsfordeling



- Prosedyre:
  - Utfør n ganger:
    - Trekk en tilfedig kule
    - registrer farge
    - legg kula tilbake
- Da er $X =$"Antall røde kuler" binomisk fordelt

$$X \sim \text{Binom}(n, p)$$

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 1, 2, \ldots$$

med $p = \frac{\text{antall røde kuler}}{\text{antall kuler}}$

- $E[X] = np$ og $Var[X] = np(1-p)$

Tables of the Binomial Cumulative Distribution

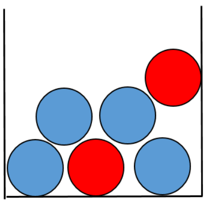
The table below gives the probability of obtaining at most x successes in n independent trials, each of which has a probability p of success. That is, if X denotes the number of successes, the table shows

$$P(X \leq x) = \sum_{r=0}^{x} C_r^n p^r (1-p)^{n-r}$$

$p=$		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
$n=2$	$x=0$	0.9801	0.9604	0.9409	0.9216	0.9025	0.8836	0.8649	0.8464	0.8281	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9999	0.9996	0.9991	0.9984	0.9975	0.9964	0.9951	0.9936	0.9919	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n=3$	$x=0$	0.9703	0.9412	0.9127	0.8847	0.8574	0.8306	0.8044	0.7787	0.7536	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9997	0.9988	0.9974	0.9953	0.9928	0.9896	0.9860	0.9818	0.9772	0.9720	0.9393	0.8960	0.8438	0.7840	0.7183	0.6480	0.5748	0.5000
	2	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9995	0.9993	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n=4$	$x=0$	0.9606	0.9224	0.8853	0.8493	0.8145	0.7807	0.7481	0.7164	0.6857	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9994	0.9977	0.9948	0.9909	0.9860	0.9801	0.9733	0.9656	0.9570	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	1.0000	1.0000	0.9999	0.9998	0.9995	0.9992	0.9987	0.9981	0.9973	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n=5$	$x=0$	0.9510	0.9039	0.8587	0.8154	0.7738	0.7339	0.6957	0.6591	0.6240	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
	1	0.9990	0.9962	0.9915	0.9852	0.9774	0.9681	0.9575	0.9456	0.9326	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
	2	1.0000	0.9999	0.9997	0.9994	0.9988	0.9980	0.9969	0.9955	0.9937	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
	3	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n=6$	$x=0$	0.9415	0.8858	0.8330	0.7828	0.7351	0.6899	0.6470	0.6064	0.5679	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.9985	0.9943	0.9875	0.9784	0.9672	0.9541	0.9392	0.9227	0.9048	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
	2	1.0000	0.9998	0.9995	0.9988	0.9978	0.9962	0.9942	0.9915	0.9882	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
	3	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9995	0.9992	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6563
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Hypergeometrisk Sannsynlighetsfordeling

- Prosedyre:
  - Utfør n ganger:
    - Trekk en tilfedig kule
    - registrer farge
    - legg kula til side



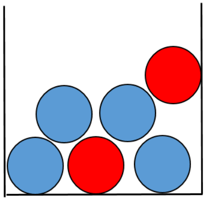
- Da er $X =$"Antall røde kuler" hypergeometrisk fordelt

$$X \sim \text{HyperGeo}(N, n, k)$$

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

med $N =$ "tot. antall kuler", $n =$"antall forsøk", $k =$"antall røde kuler"

Negativ Binomisk Sannsynlighetsfordeling



- Prosedyre:
  - Utfør inntil k røde:
    - Trekk en tilfedig kule
    - registrer farge
    - legg kula tilbake
- Da er $X =$"Antall trekk" negativ binomisk fordelt

$$X \sim \text{NegBin}(k, p)$$

med $p = \frac{\text{antall røde kuler}}{\text{antall kuler}}$

Idag

- Poisson prosess og poisson fordeling
- Kontinuerlig sannsynlighets fordeling
    - Normal (Gaussisk)

Poisson prosess

Antagelser:

1. Antall hendelser i disjunkte intervaller er uavhengige

Poisson prosess

Antagelser:

1. Antall hendelser i disjunkte intervaller er uavhengige
2. Sannsynlighet for en hendelse i en lite interval er proportional med lengde av interval

$$P(''X = 1''\text{i intervallet}(t, t + \Delta t)) = \lambda \Delta t + o(\Delta t)$$

Poisson prosess

Antagelser:

1. Antall hendelser i disjunkte intervaller er uavhengige
2. Sannsynlighet for en hendelse i en lite interval er proportional med lengde av interval

$$P(''X = 1''\text{i intervallet}(t, t + \Delta t)) = \lambda \Delta t + o(\Delta t)$$

3. Sannsynlighet for at det er mer enn er hendelse i en lite interval er neglisjerbart

$$P(''X \geq 2''\text{i intervallet}(t, t + \Delta t)) = o(\Delta t)$$

Poisson prosess

Antagelser:

1. Antall hendelser i disjunkte intervaller er uavhengige
2. Sannsynlighet for en hendelse i en lite interval er proportional med lengde av interval

$$P("X = 1" \text{i intervallet} (t, t + \Delta t)) = \lambda \Delta t + o(\Delta t)$$

3. Sannsynlighet for at det er mer enn er hendelse i en lite interval er neglisjerbart

$$P("X \geq 2" \text{i intervallet} (t, t + \Delta t)) = o(\Delta t)$$

En prosess som oppfyller 1), 2) og 3) kalles for Poisson prosess.

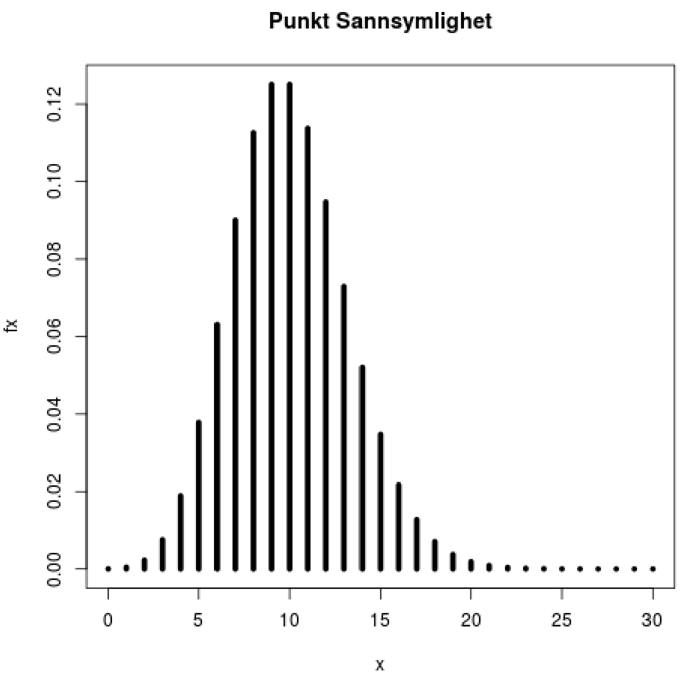
Poisson Sannsynlighetsfordeling

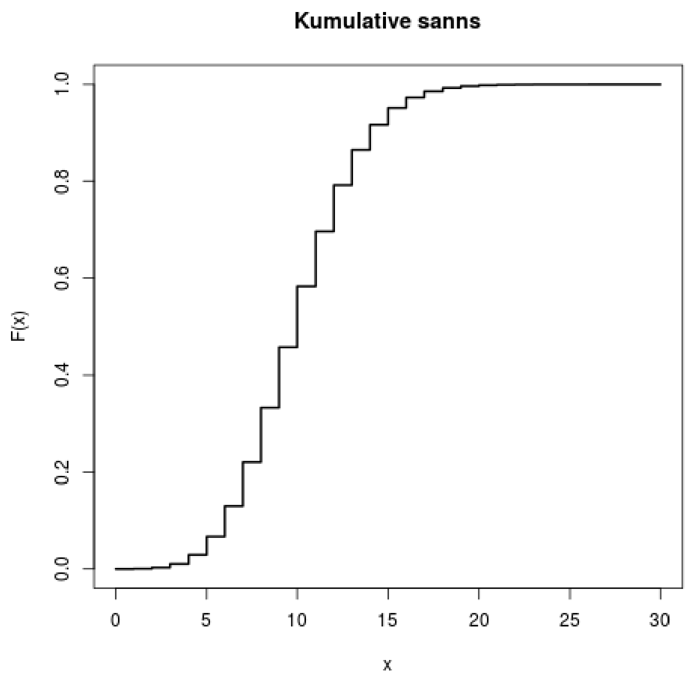
Hvis X betegner antall hendelser i $[0, t]$ for en Poisson prosess, så har X en Poisson fordeling:

$$X \sim \text{Poisson}(\lambda t)$$

$$P(X = x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

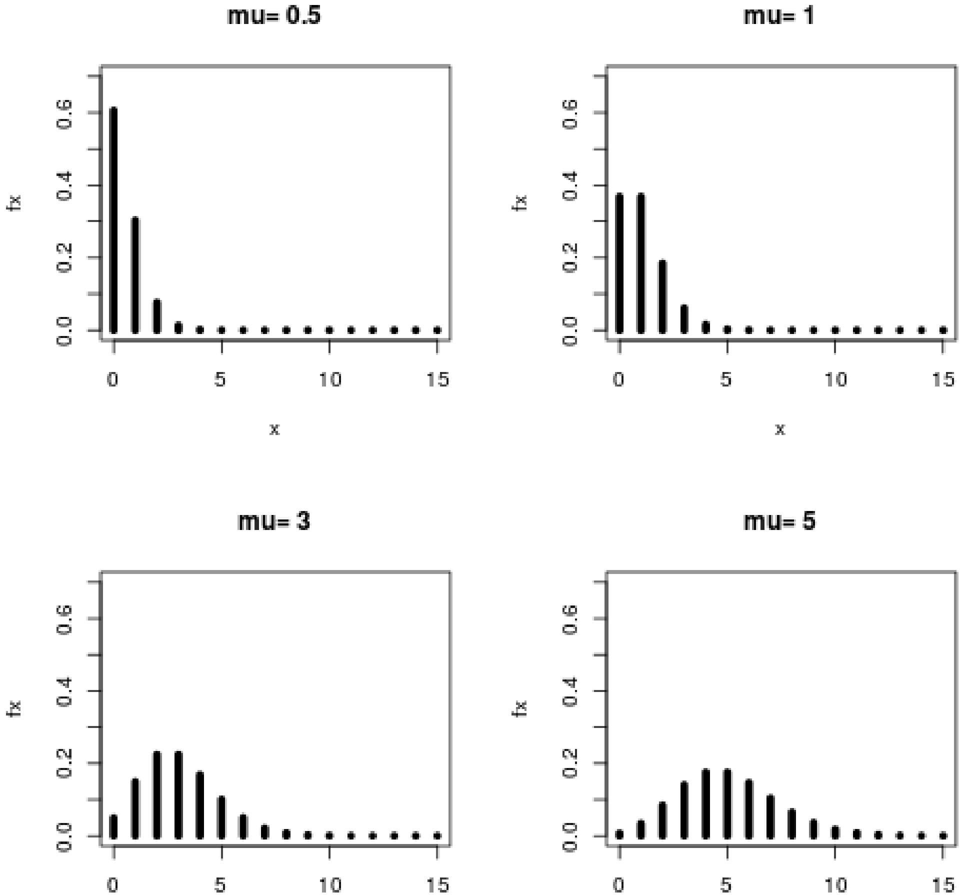
Poisson Fordeling - Eksempel 1





Poisson Fordeling

Parameter μ



Poisson Tabeller

Tables of the Poisson Cumulative Distribution

The table below gives the probability of that a Poisson random variable X with mean = λ is less than or equal to x. That is, the table gives

$$P(X \leq x) = \sum_{r=0}^{x} \lambda^r \frac{e^{-\lambda}}{r!}$$

$\lambda=$	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	11.0	10.0	12.0	14.0	15.0
$x=$ 0	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0174	0.0113	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008	0.0005	0.0002	0.0005	0.0001	0.0000	0.0000
2	0.0620	0.0430	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042	0.0028	0.0012	0.0028	0.0005	0.0001	0.0000
3	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149	0.0103	0.0049	0.0103	0.0023	0.0005	0.0002
4	0.2851	0.2237	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403	0.0293	0.0151	0.0293	0.0076	0.0018	0.0009
5	0.4457	0.3690	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885	0.0671	0.0375	0.0671	0.0203	0.0055	0.0028
6	0.6063	0.5265	0.4497	0.3782	0.3134	0.2562	0.2068	0.1649	0.1301	0.0786	0.1301	0.0458	0.0142	0.0076
7	0.7440	0.6728	0.5987	0.5246	0.4530	0.3856	0.3239	0.2687	0.2202	0.1432	0.2202	0.0895	0.0316	0.0180
8	0.8472	0.7916	0.7291	0.6620	0.5925	0.5231	0.4557	0.3918	0.3328	0.2320	0.3328	0.1550	0.0621	0.0374
9	0.9161	0.8774	0.8305	0.7764	0.7166	0.6530	0.5874	0.5218	0.4579	0.3405	0.4579	0.2424	0.1094	0.0699
10	0.9574	0.9332	0.9015	0.8622	0.8159	0.7634	0.7060	0.6453	0.5830	0.4599	0.5830	0.3472	0.1757	0.1185
11	0.9799	0.9661	0.9467	0.9208	0.8881	0.8487	0.8030	0.7520	0.6968	0.5793	0.6968	0.4616	0.2600	0.1848
12	0.9912	0.9840	0.9730	0.9573	0.9362	0.9091	0.8758	0.8364	0.7916	0.6887	0.7916	0.5760	0.3585	0.2676
13	0.9964	0.9929	0.9872	0.9784	0.9658	0.9486	0.9261	0.8981	0.8645	0.7813	0.8645	0.6815	0.4644	0.3632
14	0.9986	0.9970	0.9943	0.9897	0.9827	0.9726	0.9585	0.9400	0.9165	0.8540	0.9165	0.7720	0.5704	0.4657
15	0.9995	0.9988	0.9976	0.9954	0.9918	0.9862	0.9780	0.9665	0.9513	0.9074	0.9513	0.8444	0.6694	0.5681
16	0.9998	0.9996	0.9990	0.9980	0.9963	0.9934	0.9889	0.9823	0.9730	0.9441	0.9730	0.8987	0.7559	0.6641

Hva har vi gjort for de diskrete fordelingene?

- Beskrevet stokastisk forsøk
- utledet formel for punktsannsynlighet, $f(x) = P(X = x)$
- Utledet formel for $E[X]$ og $Var[X]$
- Regnet på eksempler
- Sett på sammenhenger mellom fordelinger:
  - hypergeometrisk $\approx$ binomisk når N er stor i forhold til n
  - binomisk $\approx$ poisson når n er stor og p er liten

Kontinuerlig Stokastisk Variabel

- Stok. var X kan ta en ikke tellbart antall verdier (typisk intervaller på tall linja)
- Sannsynlighet fordeling (sannsynlighetstetthet)

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

- Kumulative sannsynlighet fordeling

$$F(x) = P(X < x) = \int_{-\infty}^{x} f(u)du$$

- Forventningsverdi:

$$\mu = E[X] = \int_{-\infty}^{+\infty} x\ f(x)dx$$

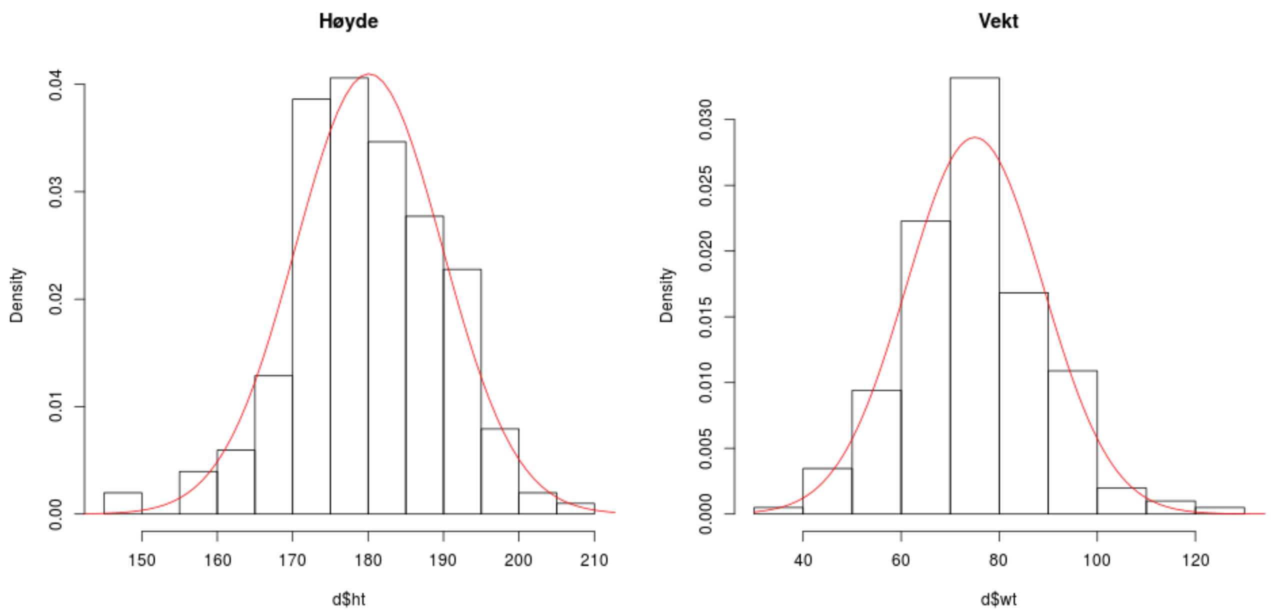
- Varians

$$\sigma^2 = Var[X] = E[(X-\mu)^2]$$

Normal Fordeling

- Historie
  - I 1733 DeMoivre avledet det som en tilnærming til en binomial fordeling
  - I 1783 Laplace brukte det for å beskrive fordeling av måle feil
  - I 1809 Gauss brukte det for å analysere astronomiske data
- Veldig mye brukt i statistikk
  - Fordeling av målefeil
  - Narturlig variasjoner
    - F.eks: Vekt til kvinner, temperatur, vekt ....
  - Lett å regne med
  - Mange fordeling lar seg tilnærme med en normal fordeling
  -

Vekt of Høyde til 202 atleter



Normal Fordeling