Sannolikhet och statistik

Binomialfördelning

HT 2008

Uwe.Menzel@math.uu.se

http://www.math.uu.se/~uwe/

Exempel: p = 0.05

Exempel: Låda med 1000 säkringar, 50 är felaktiga

$P\,(\text{fel}) = 0.05$ och $P\,(\text{inget fel}) = 0.95$ för varje försök

Tar **5 stycken**: slh. att jag får precis 1 felaktig, 2 felaktiga osv.?

Slumpvariabel X = antalet felaktiga av de 5: $\Omega = \{0, 1, 2, 3, 4, 5\}$

$$P_X\,(1) = \binom{5}{1}(0.05)^1\,(0.95)^4 = 0.203$$

$$P_X\,(2) = \binom{5}{2}(0.05)^2\,(0.95)^3 = 0.021$$

$$P_X\,(3) = \binom{5}{3}(0.05)^3\,(0.95)^2 = 0.00113$$

Exempel: allmän p

Abstraktion: Låda med (många) säkringar, andel p felaktiga

$P\,(\text{fel}) = p$ och $P\,(\text{inget fel}) = 1 - p$ för varje försök

Tar **n stycken**: slh. att jag får precis k felaktiga?

Slumpvariabel X = antalet felaktiga av n: $\Omega = \{0, 1, \ldots n\}$

$$P_X\,(k) = \binom{n}{k} p^k\,(1-p)^{n-k}$$

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

Förekomst av Bin(n,p)

Upprepar ett försök **n** gånger

Varje enskilt försök kan utfalla på två alternativa sätt:

A	A^*
p	$1-p$

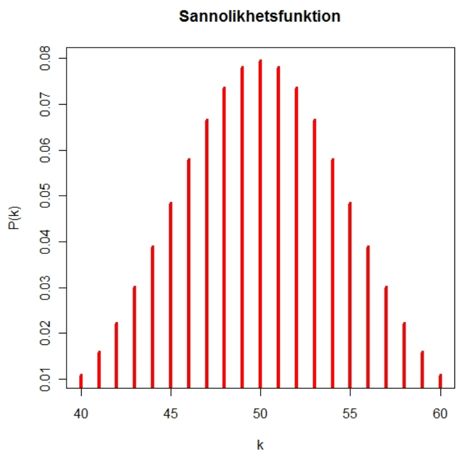
Slh. att få utfall A exakt **k** gånger, alltså A^* $(n-k)$ gånger?

A	A^*
k	$n-k$

Diskret s.v. X med $\Omega = \{0, 1, 2, \ldots, n\}$

$$p_X\,(k) = \binom{n}{k} p^k\,(1-p)^{n-k} \qquad k = 0, 1, \ldots n$$

Sannolikhetsfunktion för binomialfördelningen




```
sum(dbinom(0:100, 100, 0.5))  # måste vara ett
x<-40:60
X<-dbinom(x, 100, 0.5)
plot(x,X,type="h",lwd=4,col=2,xlab="k",ylab="P(k)",main="Sannolikhetsfunktion")
```

Fler exempel för binomialfördelning

Myntkast: antalet kronor i 10 kast;
$n = 10, \quad p = 1/2, \ k = 0, 1, \ldots 10$

Antal 6:or i 100 tärningskast;
$n = 100, \quad p = 1/6, \ \ (1-p) = 5/6, \ k = 0, 1, \ldots 100$

Urna med (många!) vita och svarta kulor, slh. att dra k svarta om totalt n blir dragna; $p = \dfrac{\text{antalet svarta}}{\text{antalet kulor}}, \ k = 0, 1, \ldots n$

Opinion: antalet parti-A väljare bland 1000 som frågades

- om det finns bara två partier A och B
- om man känner andelen av A väljare ($= p$) i hela befolkningen
- annars: maximum-likelihood eller dylikt

Fördelningsfunktion

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots n$$

$$F_X(k) = \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j}$$

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

Tabell för fördelningsfunktionen för Bin(n,p)

Tabell 6. Binomialfördelningen

$P(X \leq x)$ där $X \in \mathrm{Bin}(n, p)$.
För $p > .5$ utnyttja att $P(X \leq x) = P(Y \geq n - x)$ där $Y \in \mathrm{Bin}(n, 1-p)$

n	x \ p	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
2	0	.90250	.81000	.72250	.64000	.56250	.49000	.36000	.25000
	1	.99750	.99000	.97750	.96000	.93750	.91000	.84000	.75000
3	0	.85737	.72900	.61412	.51200	.42188	.34300	.21600	.12500
	1	.99275	.97200	.93925	.89600	.84375	.78400	.64800	.50000
	2	.99987	.99900	.99662	.99200	.98438	.97300	.93600	.87500
4	0	.81451	.65610	.52201	.40960	.31641	.24010	.12960	.06250
	1	.98598	.94770	.89048	.81920	.73828	.65170	.47520	.31250
	2	.99952	.99630	.98802	.97280	.94922	.91630	.82080	.68750
	3	.99999	.99990	.99949	.99840	.99609	.99190	.97440	.93750
5	0	.77378	.59049	.44371	.32768	.23730	.16807	.07776	.03125
	1	.97741	.91854	.83521	.73728	.63281	.52822	.33696	.18750
	2	.99884	.99144	.97339	.94208	.89648	.83692	.68256	.50000
	3	.99997	.99954	.99777	.99328	.98438	.96922	.91296	.81250
	4	1.00000	.99999	.99992	.99968	.99902	.99757	.98976	.96875
6	0	.73509	.53144	.37715	.26214	.17798	.11765	.04666	.01562
	1	.96723	.88574	.77648	.65536	.53394	.42017	.23328	.10938
	2	.99777	.98415	.95266	.90112	.83057	.74431	.54432	.34375
	3	.99991	.99873	.99411	.98304	.96240	.92953	.82080	.65625
	4	1.00000	.99995	.99960	.99840	.99536	.98906	.95904	.89063
	5	1.00000	1.00000	.99999	.99994	.99976	.99927	.99590	.98438
7	0	.69834	.47830	.32058	.20972	.13348	.08235	.02799	.00781
	1	.95562	.85031	.71658	.57672	.44495	.32942	.15863	.06250
	2	.99624	.97431	.92623	.85197	.75641	.64707	.41990	.22656
	3	.99981	.99727	.98790	.96666	.92944	.87396	.71021	.50000
	4	.99999	.99982	.99878	.99533	.98712	.97120	.90374	.77344
	5	1.00000	.99999	.99993	.99963	.99866	.99621	.98116	.93750
	6	1.00000	1.00000	1.00000	.99999	.99994	.99978	.99836	.99219
8	0	.66342	.43047	.27249	.16777	.10011	.05765	.01680	.00391
	1	.94276	.81310	.65718	.50332	.36708	.25530	.10638	.03516
	2	.99421	.96191	.89479	.79692	.67854	.55177	.31539	.14453
	3	.99963	.99498	.97865	.94372	.88618	.80590	.59409	.36328
	4	.99998	.99957	.99715	.98959	.97270	.94203	.82633	.63672
	5	1.00000	.99998	.99976	.99877	.99577	.98871	.95019	.85547
	6	1.00000	1.00000	.99999	.99992	.99962	.99871	.99148	.96484
	7	1.00000	1.00000	1.00000	1.00000	.99998	.99993	.99934	.99609
9	0	.63025	.38742	.23162	.13422	.07508	.04035	.01008	.00195
	1	.92879	.77484	.59948	.43621	.30034	.19600	.07054	.01953
	2	.99164	.94703	.85915	.73820	.60068	.46283	.23179	.08984

Väntevärde, varians, standardavvikelse

Låt $X \in Bin\,(n, p)$:

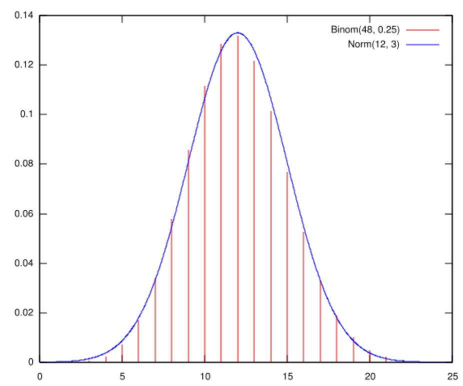
$$
\begin{aligned}
E\,(X) &= \sum_{k=0}^{n} k \cdot p_X\,(k) = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k\,(1-p)^{n-k} = \cdots = np \\
V\,(X) &= E\,(X^2) - (E\,(X))^2 = \cdots = np\,(1-p) \\
D\,(X) &= \sqrt{np\,(1-p)}
\end{aligned}
$$

Normalapproximation för $np\,(1-p) \geq 10$

För stora n är X ungefär normalfördelad med binomialfördelningens väntevärde och standardavvikelse:

$$
\begin{aligned}
& X \in Bin\,(n, p) \quad \text{med } n \cdot p \cdot (1-p) \geq 10 \\
\Longrightarrow\ & X \in N\left(np, \sqrt{np\,(1-p)}\right)
\end{aligned}
$$

Normalapproximation för $np\,(1-p) \geq 10$



Figur: Sats från Moivre-Laplace; n =48 , p=0.25

Normalapproximation för $np\,(1-p) \geq 10$

För stora n är X ungefär normalfördelad med binomialfördelningens väntevärde och standardavvikelse:

$$X \in N\left(np, \sqrt{np\,(1-p)}\right)$$

$$
\begin{aligned}
P\,(a < X \leq b) &\approx F_X\,(b) - F_X\,(a) \quad \text{(för normalfördelning)} \\
&= \Phi\left(\frac{b - np}{\sqrt{np\,(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np\,(1-p)}}\right)
\end{aligned}
$$

Exempel för normalapproximation: Massproduktion av byggelement. Ett byggelement blir defekt med $p = 0.1$. En byggmästare köper 1000 stycken.

Sannolikheten att han får minst 80, högst 120 felaktiga ?

Slumpvariabel X: antalet felaktiga han köpt $\Omega = \{0, 1, \ldots 1000\}$

$$\begin{aligned} X &\in Bin\,(n,p) = Bin\,(1000, 0.1) \quad \text{med} \quad np\,(1-p) = 90 > 10 \\ X &\in N\left(n \cdot p, \sqrt{np\,(1-p)}\right) = N\left(1000 \cdot 0.1, \sqrt{90}\right) = N\,(100, 9.487) \end{aligned}$$

$$\begin{aligned} P\,(\mathbf{79} < X \leq 120) &= F_X\,(120) - F_X\,(79) \\ &= \Phi\left(\frac{120-100}{9.487}\right) - \Phi\left(\frac{79-100}{9.487}\right) \\ &= \Phi\,(2.11) - \Phi\,(-2.21) \\ &= 0.9826 + 0.98645 - 1 = 0.969 \end{aligned}$$

Poissonapproximation om $p \leq 0.1$

Om p är litet kan man approximera binomialfördelningen med poissonfördelningen där $\mu = np$

$$\begin{aligned} X &\in Bin\,(n,p) \quad \text{med } p < 0.1 \\ \Longrightarrow X &\in Po\,(n \cdot p) \end{aligned}$$

$$p_X\,(k) \approx \frac{\mu^k}{k!} e^{-\mu} = \frac{(np)^k}{k!} e^{-np}$$

Normalapproximation: halvkorrektion

Istället för:

$$P\,(a < X \leq b) \approx \Phi\left(\frac{b-np}{\sqrt{np\,(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np\,(1-p)}}\right)$$

kan man räkna:

$$P\,(a < X \leq b) \approx \Phi\left(\frac{b+\mathbf{1/2}-np}{\sqrt{np\,(1-p)}}\right) - \Phi\left(\frac{a+\mathbf{1/2}-np}{\sqrt{np\,(1-p)}}\right)$$

för att det blir mera noggrannt ..

Exempel för Poissonapproximation: Massproduktion av detaljer. En tillverkad detalj blir defekt med $p = 0.005$. Förpackas utan kontroll i kartonger med 100 stycken.

Sannolikheten att en kartong innehåller mer än 3 dåliga detaljer ?

Slumpvariabel X: antalet dåliga detaljer per kartong

$A =$ dålig	$A^* =$ ok
$p = 0.005$	$1 - p = 0.995$
k	$n - k$

$$\begin{aligned} X \in Bin\,(n,p) &= Bin\,(100, 0.005) \quad p < 0.1 \\ X \in Po\,(n \cdot p) &= Po\,(100 \cdot 0.005) = Po\,(0.5) \quad \text{dvs.} \quad \mu = 0.5 \end{aligned}$$

$$\begin{aligned} P\,(X > 3) &= 1 - P\,(X \leq 3) = 1 - F_X\,(3) \quad \text{med} \quad X \in Po\,(0.5) \\ &= 1 - 0.99825 = 0.00175 \quad \text{tabell nästa sida} \end{aligned}$$

Poissonfördelning: tabell

Tabell 5. Poissonfördelningen
$P(X \leq x)$ där $X \in \text{Po}(\mu)$.

$\mu =$

x | μ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	.90484	.81873	.74082	.67032	.60653	.54881	.49659	.44933	.40657
1	.99532	.98248	.96306	.93845	.90980	.87810	.84420	.80879	.77248
2	.99985	.99885	.99640	.99207	.98561	.97688	.96586	.95258	.93714
$X =$ 3	1.00000	.99994	.99973	.99922	.99825	.99664	.99425	.99092	.98654
4		1.00000	.99998	.99994	.99983	.99961	.99921	.99859	.99766
5			1.00000	1.00000	.99999	.99996	.99991	.99982	.99966
6					1.00000	1.00000	.99999	.99998	.99996
7							1.00000	1.00000	1.00000

x | μ	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6
0	.36788	.30119	.24660	.20190	.16530	.13534	.11080	.09072	.07427
1	.73576	.66263	.59183	.52493	.46284	.40601	.35457	.30844	.26738
2	.91970	.87949	.83350	.78336	.73062	.67668	.62271	.56971	.51843
3	.98101	.96623	.94627	.92119	.89129	.85712	.81935	.77872	.73600
4	.99634	.99225	.98575	.97632	.96359	.94735	.92750	.90413	.87742

Figur: Del av tabellen för Poissonfördelningen