Sannolikhet och statistik

Binomialfördelningen

VT 2009

Uwe.Menzel@math.uu.se

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

Bernoulli-Experiment

Johann Bernoulli

Slh. = Sannolikhet

- Ett slumpförsök har 2 alternativa utfall: A, A*
- Vi känner slh:a för båda utfallen: P(A)=*p och P(A*)=1-p*
  - mynt: P(*krona*) = p = 0.5 och P(*vapen*) = 1 - p = 0.5
  - tärning: P(femma) = 1/6 och P(ingen femma)=1 – p = 5/6
- Vi upprepar detta försök flera (n) gånger.
- *Att beräkna:* slh. att det blir k gånger utfall A
  - slh. att få 5 ggr. krona i 10 myntkast (n=10 k=5 p=0.5)
  - slh. att få 10 femmor i 100 tärningskast (n=100 k=10 p=1/6)

Bernoulli-Experiment

utfall	A	A*
slh.	p	1 - p
antal	k	n - k

Slumpvariabel X: antalet utfall A

$\Omega_X = \{ 0, 1, 2, \ldots n \}$

$p_X(k)$ = ???

Det här vill man helst veta för alla k som kan förekomma, dvs. för hela utfallsrummet Ω

OBS !! Blanda inte ihop:

- p är slh. att lyckas i ett enda försök
- $p_X(k)$ är slh. att lyckas k gånger i n försök

Exempel: 5 myntkast (n=5)

$$P_X(k=5) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{32} = 0.03125$$

s.v. X = antal framsidor

$$P_X(k=4) = 5 \cdot \frac{1}{32} = 0.15625$$

osv. ...

Exempel: 5 myntkast (n=5)

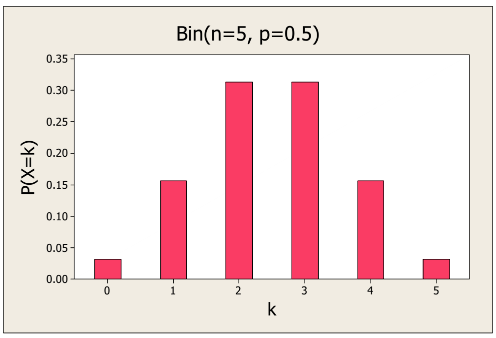
$$P_X(k=3) = \binom{5}{3} \cdot p^3 \cdot (1-p)^2 = 10 \cdot \frac{1}{2^3} \cdot \frac{1}{2^2} = \frac{10}{32} = 0.3125$$

$$P_X(k=2) = \binom{5}{2} \cdot p^2 \cdot (1-p)^3 = 10 \cdot \frac{1}{2^2} \cdot \frac{1}{2^3} = \frac{10}{32} = 0.3125$$

$$P_X(k=1) = \binom{5}{1} \cdot p^1 \cdot (1-p)^4 = 5 \cdot \frac{1}{2^1} \cdot \frac{1}{2^4} = \frac{5}{32} = 0.15625$$

$$P_X(k=0) = \binom{5}{0} \cdot p^0 \cdot (1-p)^5 = 1 \cdot 1 \cdot \frac{1}{2^2} = \frac{1}{32} = 0.03125$$

Sannolikhetsfunktion (5 myntkast)



x	P(X = k)
0	0.03125
1	0.15625
2	0.31250
3	0.31250
4	0.15625
5	0.03125

Binomial.MPJ

Allmän beräkning av sannolikhetsfunktionen

utfall	A	A*
slh.	p	1 - p
antal	k	n - k

k = 0,1, 2, ... n

$$X \in Bin(n,p)$$

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

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

En slumpvariabel X med ovanstående sannolikhetsfunktion kallas binomialfördelad.
Man skriver X ~ Bin(n,p) *n = antalet försök* *p = slh. ”att lycka” (A kommer upp)*

Fördelningsfunktionen för Bin(n,p)

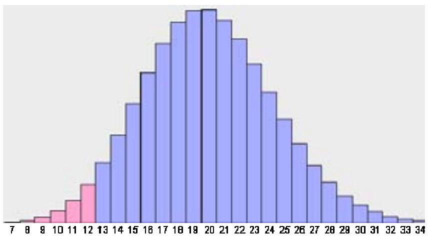
$$F_X(x) = P(X \le x) = \sum_{k=0}^{x} p_X(k)$$

$$P(max\ 3\ femmor) = P_X(0) + P_X(1) + P_X(2) + P_X(3)$$

Fördelningsfunktionen

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

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



Tabell över fördelningsfunktionen för binomialfördelningen

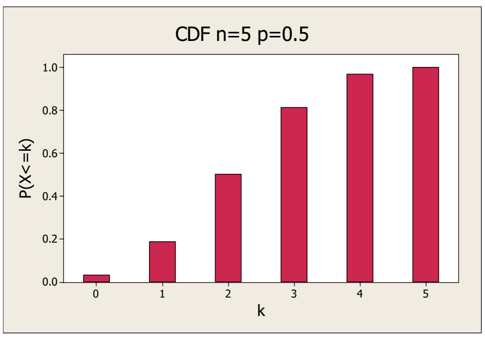
k

Tabell 6. Binomialfördelningen

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

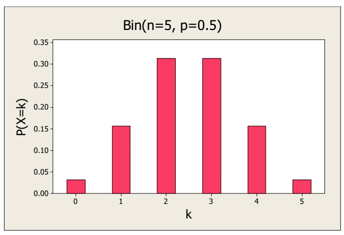
n	x	p 0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
2	0	.90250	.81000	.72250	.64000	.56250	.49000	.36000	.25000
	1	.99750	.99000	.97750	.96000	.93750	.91000	.84000	.75000
3	0	.85737	.72900	.61412	.51200	.42188	.34300	.21600	.12500
	1	.99275	.97200	.93925	.89600	.84375	.78400	.64800	.50000
	2	.99987	.99900	.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

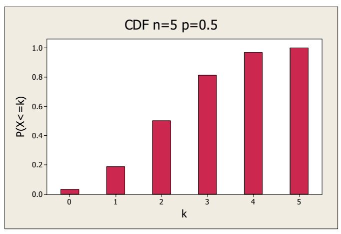
Fördelningsfunktion (5 myntkast)



x	P(X <= x)
0	0.03125
1	0.18750
2	0.50000
3	0.81250
4	0.96875
5	1.00000

Beräkning av en sannolikhet med hjälp av fördelningsfunktionen





$$P(2 < X \le 4) = F_X(4) - F_X(2)$$

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

Slumpvariabeln X är diskret → det är mycket viktigt att använda likhetstecknen korrekt.

Väntevärde, varians, standardavvikelse

Låt $X \in Bin(n, p)$:

$$E(X) = \sum_{k=0}^{n} k \cdot p_X(k) = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} = \cdots = np$$

$$V(X) = E(X^2) - (E(X))^2 = \cdots = np(1-p)$$

$$D(X) = \sqrt{np(1-p)}$$

Normalapproximation för Bin

Låt X vara en binomialfördelad slumpvariabel. Många försök görs (dvs. n är stor) → X är ungefär normalfördelad ... med binomialfördelningens väntevärde och varians.

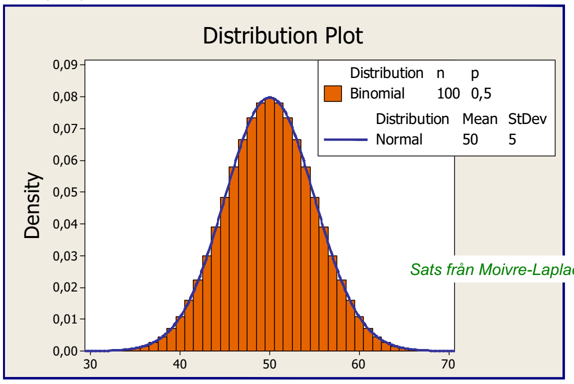
$$X \in Bin(n, p) \quad \text{med } n \cdot p \cdot (1-p) \geq 10$$

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

Sats från Moivre-Laplace

Normalapproximation för Bin

$X \in Bin(n,p)$ och n stor



Normalapproximation för Bin

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

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

$$P(a < X \leq b) \approx F_X(b) - F_X(a) \quad \text{(för normalfördelning)}$$

$$= \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right)$$

Normalapproximation för Bin

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 $\Omega = \{0, 1, \ldots 1000\}$

$$X \in Bin(n, p) = Bin(1000, 0.1) \quad \text{med} \quad np(1-p) = 90 > 10$$

$$X \in N\left(n \cdot p, \sqrt{np(1-p)}\right) = N\left(1000 \cdot 0.1, \sqrt{90}\right) = N(100, 9.487)$$

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

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 ..

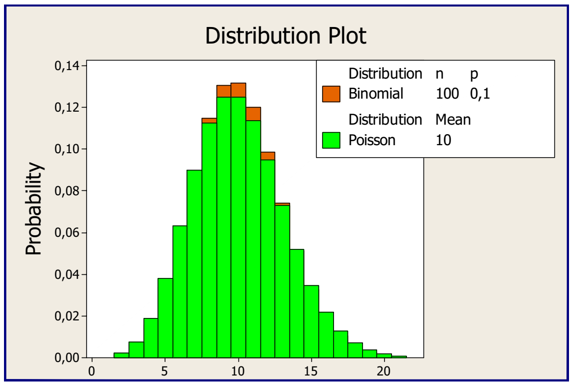
Poissonapproximation för p ≤ 0.1

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

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

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

Poissonapproximation för p ≤ 0.1



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	$n - k$

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

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

Tabell Poissonfördelning

Tabell 5. Poissonfördelningen
$P(X \le x)$ 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
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