Grundläggande matematisk statistik

Diskreta fördelningar

Uwe Menzel, 2018

uwe.menzel@matstat.org

www.matstat.org

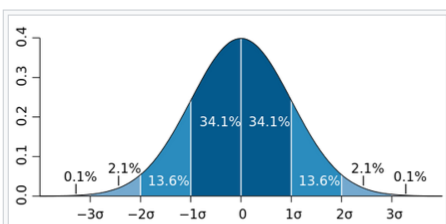
1

Begrepp ”fördelning”

- Hur beter sig en variabel slumpmässigt ?
- En slumpvariabel (s.v.) har en viss fördelning, d.v.s. en viss sannolikhetsfunktion/täthetsfunktion och fördelningsfunktion.
- Desssa funktioner kännetecknar slumpprocessen entydigt.

Sannolikhetsfördelning är inom sannolikhetsteori, statistik och matematisk statistik, en beskrivning (ofta i form av en funktion) av sannolikheterna för utfallen i ett utfallsrum.

Sannolikhetsfördelningar, ibland bara "fördelningar", förekommer i både diskreta och kontinuerliga utfallsrum och kallas därför ibland *diskret fördelning* eller *kontinuerlig fördelning*, för att ange typen av utfallsrum.



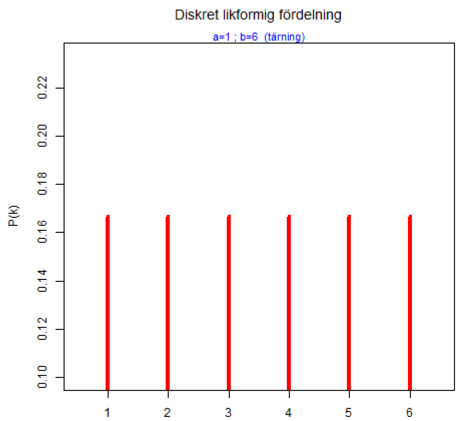
Normalfördelningen är en mycket vanligt förekommande sannolikhetsfördelning i statistiska modeller

Exempelvis är en likformig fördelning en fördelning där alla utfall är lika sannolika, vilket är fallet till exempel vid en dragning av ett nummer i en lottorad: där är alla utfall i det diskreta utfallsrummet [1, 2, 3, ... 34, 35] lika sannolika med sannolikheten 1/35.

Wikipedia: https://sv.wikipedia.org/wiki/Sannolikhetsf%C3%B6rdelning

2

Diskret likformig fördelning



$X \sim U(a, b)$ kodbeteckning

$\Omega = \{a, a+1, a+2, \dots, b\}$

$N = b - a + 1$ antalet utfall

$$p_X(k) = \frac{1}{N}$$

$$E(X) = \frac{a+b}{2}$$

$$V(X) = \frac{(b-a+2)\cdot(b-a)}{12}$$

Exempel tärning: slumpvariabel $X =$ ögontalet

$a = 1; b = 6; N = 6;\ p_X(k) = {}^1/_6\ ; E(X) = 3.5$

www.matstat.org

3

Binomialfördelningen

”bi”

Slumpförsök: **två** alternativa utfall: A, A^* (så kallad Bernoulli-försök)

$$\left.\begin{array}{l} P(A) = p \\ P(A^*) = 1 - p \end{array}\right\}$$ känt! p betecknas ofta som ”sannolikhet att lyckas”

Exempel 1: mynt utfall A: krona, utfall A^*: klave $\Big\}$ $P(A) = P(A^*) = p = \frac{1}{2}$

Exempel 2: tärning utfall A: kastar en femma, utfall A^*: kastar något annat $\Big\}$ $P(A) = p = \frac{1}{6}$, $P(A^*) = 1 - p = \frac{5}{6}$

Vi gör n oberoende repetitioner av detta försök!
Vi söker sannolikheten att det blir k gånger händelse A $(k \le n)$

Exempel 1, mynt: kastar 10 gånger och söker slh. att det blir 6 ggr. krona
$(n = 10; k = 6, p = {}^1/_2)$

Exempel 2, tärning: kastar 20 gånger och söker slh. att få 3 femmor
$(n = 20; k = 3; p = {}^1/_6)$

www.matstat.org

4

Binomialfördelningen

$X \sim Bin(n, p)$

Vi gör n oberoende Bernoulli-försök:

	A	A^*	summa
sannolikhet	p	$1-p$	1
antal	k	$n-k$	n

Slumpvariabel X: antalet A-utfall (”lyckade”) bland totalt n (Bernoulli-) försök

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

Vilken sannolikhetsfunktion har slumpvariabeln X ?

$$P(X=k) = p_X(k) = ?$$

www.matstat.org

5

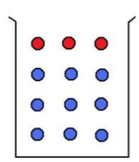
Binomialfördelningen

$X \sim Bin(n, p)$ $\quad P(X=k) = p_X(k) = ?$

Vi kan försöka hitta sannolikhetsfunktionen med hjälp av ett exempel:

Urna:

3 röda kulor
7 blåa kulor



Slumpförsök:

- ta en kula, registrera dess färg och lägg tillbaka
- händelse A: röd (”lyckad”) $P(A) = p = 0.3$
- händelse A^*: blå kula $P(A^*) = 1 - p = 0.7$
- upprepa detta $n = 4$ gånger → 4 Bernoulli försök

Slumpvariabel *X*: antalet röda (av totalt 4 dragningar)

”och”

$$P(X=4) = P[(A_1 = röd) \cap (A_2 = röd) \cap (A_3 = röd) \cap (A_4 = röd)]$$

$$= P(A_1 = röd) \cdot (A_1 = röd) \cdot (A_1 = röd) \cdot (A_1 = röd) \quad \text{pga. oberoendet}$$

$$= p \cdot p \cdot p \cdot p = p^4 = 0.3^4$$

www.matstat.org

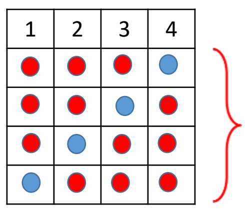
6

Binomialfördelningen

$$P(X=3) = 4 \cdot \overset{\text{tre ggr. röd}}{p^3} \cdot \underbrace{(1-p)}_{\text{en gång blå}}$$

Varför faktor 4 ?

Kom ihåg exemplet med de tre elektroniska komponenterna: det finns 4 möjligheter att anordna de 4 försöken. De lyckade försöken kan befinner sig på positionerna $\{1,2,3\}$, $\{1,2,4\}$, $\{1,3,4\}$ eller $\{2,3,4\}$ (lyckade försök = röda punkter):



4 möjligheter för ordningsföljden

www.matstat.org

7

Binomialfördelningen

$P(X=2) = \binom{4}{2} \cdot p^2 \cdot (1-p)^2$ 2 röda, 2 blåa, $\binom{4}{2}$ möjligheter att anordna (att välja 2 av 4)

$P(X=1) = \binom{4}{1} \cdot p \cdot (1-p)^3$ 1 röd, 3 blåa, $\binom{4}{1}$ möjligheter att anordna (att välja 1 av 4)

$P(X=0) = \binom{4}{0} \cdot (1-p)^4$ 0 röda, 4 blåa, $\binom{4}{0} = 1$ möjlighet att anordna dem

allmänt: $\boxed{P(X=k) = p_X(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad k = 0, 1, 2, \ldots, n}$

Om en slumpvariabel X har en sådan sannolikhetsfunktion säger man att X är binomialfördelad

- slumpvariabel X: antalet ”lyckade”
- $P(X=k)$: sannolikhet att slumpvariabeln X antar värdet k, dvs. att få k ”lyckade” försök
- n: totala antalet (Bernoulli-)försök
- $p = P(A)$: sannolikhet att ”lyckas” i ett enda (Bernoulli-) försök

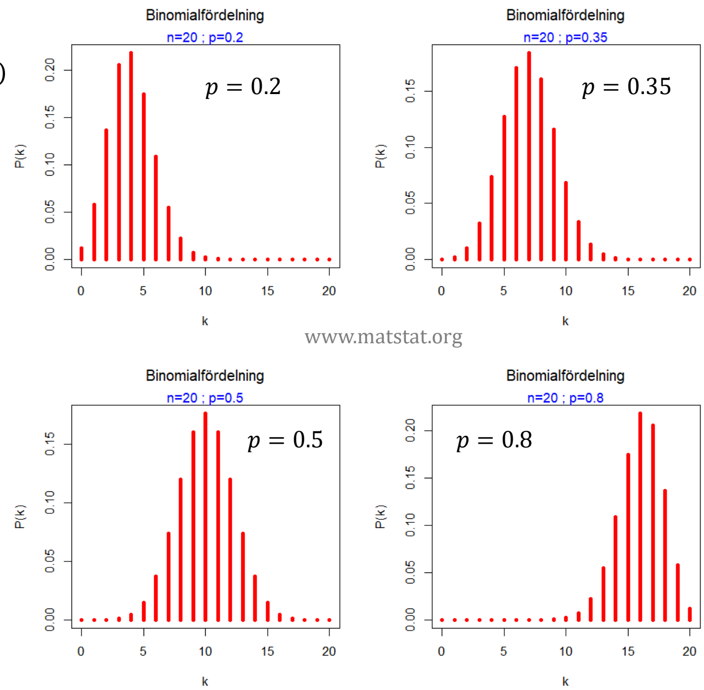
Kodbeteckning: $X \sim Bin(n, p)$ Utfallsrum: $\Omega_X = \{0, 1, 2, \ldots, n\}$

www.matstat.org

8

$X \sim Bin(n, p)$

Stolpdiagram, olika p:



9

Binomialfördelningen

<u>Exempel:</u> Myntkast, 20 ggr. med $P(\text{krona}) = 0.5$.

Sannolikhet att få krona 8 ggr.?:

Slumpvariabel X: antal kronor i 20 kast

$X \sim Bin(20, 0.5)$

$$P(X = 8) = p_X(8) = \binom{20}{8} \cdot 0.5^8 \cdot 0.5^{12} = 0.12$$

$P = dbinom(8, size = 20, prob = 0.5)$

www.matstat.org

10

Binomialfördelningen

Normering:

$$\sum_{k=0}^{n} p_X(k) = \sum_{k=0}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = 1$$ bevis med binomialteoremet

(se appendixet)

Väntevärde: $E(X) = \sum_{k=0}^{n} k \cdot p_X(k) = \ ... = n \cdot p$

<u>Exempel</u> myntkast: $p = {}^1/_2\,; n = 10 \ \rightarrow \ E(X) = 5$

Väntevärde, varians, standardavvikelse:

$$\boxed{\begin{array}{lcl} E(X) & = & n \cdot p \\ V(X) & = & n \cdot p \cdot (1-p) \\ D(X) & = & \sqrt{V(X)} \end{array}}$$

www.matstat.org

11

Binomialfördelningen

Fördelningsfunktion:

$$\boxed{F_X(a) = P(X \leq a) = \sum_{k=0}^{a} p_X(k)}$$

a: reellt tal
kan inte anges formelmässigt

Fördelningsfunktionen för $Bin(n, p)$ finns tabellerad

- Jogreus s. 340, Blom et al. s. 403
- bara för vissa p
- bara för $n = 2, 3, \ \ldots, 20$

Kom ihåg: $p_X(k) = F_X(k) - F_X(k-1)$ kan användas för att räkna ut $p_X(k)$ med hjälp av en tabell för fördelningsfunktionen.

www.matstat.org

12

Tabell:

$X \sim Bin(n, p)$

Tabell 6. Binomialfördelningen

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

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

13

Binomialfördelningen

<u>Exempel</u>: $X \sim Bin(5, 0.2)$, dvs. $n = 5$ och $p = 0.2$

Vad är sannolikheten att det blir 3 ”lyckade” (dvs. 3 ggr. händelsen med slh. p)?

a) $p_X(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$ måste beräknas för $n = 5$; $k = 3$; $p = 0.2$

$$p_X(3) = \binom{5}{3} \cdot 0.2^3 \cdot 0.8^2 = \frac{5!}{3! \cdot 2!} \cdot 0.2^3 \cdot 0.8^2 = \frac{4 \cdot 5}{2} \cdot 0.2^3 \cdot 0.8^2 = \underline{\underline{0.0512}}$$

b) **eller** med hjälp av tabellen över fördelningsfunktionen:

$$p_X(3) = F_X(3) - F_X(2) = 0.99328 - 0.94208 = \underline{\underline{0.0512}}$$ (se tabell ovan)

stora n, k : tabell är vanligtvis enklare och snabbare (om en tabell med lämpligt p finns)

c) ännu enklare: R

dbinom : sannolikhetsfunktion (”probability density”)
pbinom : fördelningsfunktion (”probability distribution”)

dbinom(x, size, prob)

3 5 0.2

dbinom(3, 5, 0.2)=$\underline{\underline{0.0512}}$

www.matstat.org

14

Binomialfördelningen

Fördelning för komplementära händelser:

$P(A) \ = p \qquad s.v.X: antalet\ A \qquad X \sim Bin(n,p)$

$P(A^*) = 1-p \qquad s.v.Y:\ antalet\ A^* \qquad Y \sim Bin(n,1-p)$

$$X + Y = n$$

antal A — antal A* — antal försök

Sammanhang mellan sannolikhetsfunktionerna för X och Y:

$$\boxed{P(X = k) = P(Y = n-k)}$$

Om det blev k ggr. A-händelsen, så blev det $(n-k)$ gånger A*-händelsen.

<u>Exempel</u> mynt: 10 kast, 3 kronor → 7 klave

www.matstat.org

15

Binomialfördelningen

Fördelning för komplementära händelser:

$P(A) \ = p \qquad s.v.X: antalet\ A \qquad X \sim Bin(n,p)$

$P(A^*) = 1-p \qquad s.v.Y:\ antalet\ A^* \qquad Y \sim Bin(n,1-p)$

Sammanhang mellan sannolikhetsfunktionerna för X och Y:

$$F_X(k) = P(X \le k) = P(n-Y \le k) = P(-Y \le k-n) = P(Y \ge n-k)$$
$$= 1 - P(Y \le n-k-1) = 1 - F_Y(n-k-1)$$

$$\Rightarrow \boxed{F_X(k) = 1 - F_y(n-k-1)}$$

$$\Rightarrow \boxed{P(X \le k) = P(Y \ge n-k)}$$

www.matstat.org

16

Binomialfördelningen

Sammanhang mellan sannolikhetsfunktionerna för X och Y:

Exempel: $n = 5$; $k = 2$

X	Y
0	5
1	4
2	3
3	2
4	1
5	0

$$P(X \leq k) = P(Y \geq n - k)$$

$$\Downarrow$$

$$P(X \leq 2) = P(Y \geq 3)$$

www.matstat.org

17

Binomialfördelningen

Sammanhang mellan sannolikhetsfunktionerna för X och Y:

Exempel: $X \sim Bin(n, p)$ med $P(A) = p = 0.7$; $n = 5$ $\rightarrow$ $X \sim Bin(5, 0.7)$

Söker: $F_X(1)$

a) använd $F_X(k) = 1 - F_y(n - k - 1)$ och tabell

OBS! Inga värden för $p = 0.7$ i tabellen, därför måste vi använda $F_Y(k)$

$$F_X(1) = 1 - F_Y(\underset{n}{5} - \underset{k}{1} - 1) = 1 - F_Y(3) \quad \text{där } Y \sim Bin(5, 0.3)$$

$$= 1 - 0.969 = \underline{\underline{0.031}}$$

b) använd $F_X(a) = \sum_{k=0}^{a} p_X(k)$

$$F_X(1) = p_X(0) + p_X(1) = \binom{5}{0} \cdot 0.7^0 \cdot 0.3^5 + \binom{5}{1} \cdot 0.7^1 \cdot 0.3^4 = \underline{\underline{0.031}}$$

www.matstat.org

18

Binomialfördelningen

Additionssats för binomialfördelningar:

$$\left.\begin{array}{l} X \sim Bin(n_1, p) \\ Y \sim Bin(n_2, p) \end{array}\right\} \quad X + Y \sim Bin(n_1 + n_2, p)$$

OBS! Slumpvariablerna X och Y måste ha samma parameter p !

- först n_1 oberoende Bernoulli-försök med slh. p
- sedan n_2 oberoende Bernoulli-försök med slh. p

samma sak som $n_1 + n_2$ Bernoulli-försök på en gång!

www.matstat.org

19

Binomialfördelningen

<u>Exempel</u>: Blom s. 187

Ett (Bernoulli-) försök lyckas med sannolikhet $p = 0.8$.
Man utför 12 oberoende försök.

a) Vilken fördelning har slumpvariabeln $X =$ antalet lyckade försök ?

$X \sim Bin(12, 0.8)$

b) Vilken fördelning har slumpvariabeln $Y =$ antalet misslyckade försök ?

$Y \sim Bin(12, 0.2)$

OBS! akta "≤"-tecknet

c) Ange $P(2 \leq Y \leq 4)$

tabell

$$P(2 \leq Y \leq 4) = P(1 < Y \leq 4) = F_X(4) - F_X(1) = 0.927 - 0.275 = \underline{\underline{0.652}}$$

www.matstat.org

20

Binomialfördelningen

Exempel: Blom s. 187

Ett (Bernoulli-) försök lyckas med sannolikhet $p = 0.8$.
Man utför 12 oberoende försök.

d) Ange sannolikheten att antalet lyckade försök överstiger 7 men ej 10

$P(7 < X \leq 10) = F_X(10) - F_X(7)$ $\quad$ $p = 0.8$ finns dock inte i tabellen, använd Y

$$= 1 - F_Y(12 - 10 - 1) - 1 + F_Y(12 - 7 - 1) = F_Y(4) - F_Y(1)$$

$$= 0.927 - 0.275 = \underline{\underline{0.652}}$$

OBS!: det är detsamma som i del c), för att

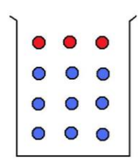
$X \in \{8, 9, 10\}$ är likvärdigt med $Y \in \{2, 3, 4\}$

X	Y	Σ
8	4	12
9	3	12
10	2	12

21

Hypergeometrisk fördelning

Urna:
- 3 röda kulor
- 7 blåa kulor



Slumpförsök:
- dra en kula, lägg inte tillbaka (skillnad till Bin!)
- dra totalt n kulor

Om kulorna **inte** läggs tillbaka:

$$\left.\begin{array}{l} P(röd\ i\ 2{:}a \mid röd\ i\ 1{:}a) = \dfrac{2}{9} \\ P(röd\ i\ 2{:}a \mid blå\ i\ 1{:}a) = \dfrac{3}{9} \end{array}\right\}$$

De enskilda dragningarna är **inte** längre oberoende → **inte** någon binomialfördelning

Slumpvariabel X = antalet röda bland de n dragna

$$P(X = k) = p_X(k) = ?$$

www.matstat.org

22

Hypergeometrisk fördelning

Anmärkning: Vi hade $P(röd\ i\ 2{:}a \mid röd\ i\ 1{:}a) = \frac{2}{9}$

$P(röd\ i\ 2{:}a \mid blå\ i\ 1{:}a) = \frac{3}{9}$

Detta betyder dock **inte** att den totala sannolikheten att dra en röd kula i 2:a dragningen skiljer sig från sannolikheten att dra en röd kula i 1:a dragningen (som man kanske kunde tänka):

Kom ihåg den **totala sannolikheten (**föreläsning **F1)**:

$$P(röd\ i\ 2{:}a) = P(röd\ i\ 2{:}a \mid röd\ i\ 1{:}a) \cdot P(röd\ i\ 1{:}a) + P(röd\ i\ 2{:}a \mid blå\ i\ 1{:}a) \cdot P(blå\ i\ 1{:}a)$$

$$= \frac{2}{9} \cdot \frac{3}{10} + \frac{3}{9} \cdot \frac{7}{10} = \frac{6+21}{90} = \frac{27}{90} = \frac{3}{10} = \underline{\underline{P(röd\ i\ 1{:}a)}}$$

Slumpvariabeln X är **inte binomialfördelad** eftersom de enskilda dragningarna är **icke oberoende** (som erfordras för binomialfördelningen)!

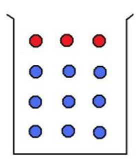
www.matstat.org

23

Hypergeometrisk fördelning

<u>Urna</u>:

- 3 röda kulor
- 7 blåa kulor



<u>Slumpförsök</u>:

- dra en kula, lägg inte tillbaka (skillnad till Bin!)
- dra totalt n kulor

Slumpvariabel X = antalet röda bland de n dragna $\quad \Omega_X = \{0, 1, 2, 3\}$

Sannolikhetsfunktionen erhålls med hjälp av den klassiska sannolikhets-definitionen: $P = {}^{g}/_{m}$

Låt oss anta att vi har 3 röda, 7 blåa (se bild ovan), vi drar totalt 5 kulor, och vill veta sannolikheten för att dra 2 röda:

$$P(X=2) = \frac{g}{m} = \frac{\binom{3}{2} \cdot \binom{7}{3}}{\binom{10}{5}}$$

$\binom{3}{2}$: antalet möjligheter att dra 2 bland de 3 röda

$\binom{7}{3}$: antalet möjligheter att dra 3 bland de 7 blåa
(om det blev 2 röda av 5 dragna, så måste 3 blåa ha dragits)

$\binom{10}{5}$: antalet möjligheter att dra 5 kulor bland 10

www.matstat.org

24

Hypergeometrisk fördelning

Sannolikhetsfunktion:

Slumpvariabel X = antalet "röda" bland de n dragna

$\Omega_X = \{0, 1, 2, \ldots, \min(n, m)\}$

$$\boxed{P(X = k) = \frac{\binom{m}{k} \cdot \binom{N-m}{n-k}}{\binom{N}{n}}}$$

N : totala antalet "kulor"
n: antalet dragna
m: antalet "röda"

: dhyper

$k \leq m$

antalet röda bland de dragna

antalet röda i urnan

$n - k \leq N - m$

antalet blåa bland de dragna

antalet blåa i urnan

Om ovanstående restriktioner inte är uppfyllda blir sannolikheten noll. Detta säkerställs automatiskt i formlerna eftersom t. ex. $\binom{m}{k} = 0$ om $k > m$.

www.matstat.org

25

Hypergeometrisk fördelning

Kodbeteckning: $X \sim Hyp(N, n, m)$

N: så många finns totalt

n: så många dras

m: så många "röda" finns

Naturligtvis gäller detta inte bara för kulor som dras ur en urna. Fördelningen gäller allmänt för "dragning" utan återläggning bland två sorters föremål. (se svensk Wikipedia artikel "hypergeometrisk fördelning")

Väntevärde, varians, standardavvikelse:

Låt s.v. X vara hypergeometriskt fördelad, dvs. $X \sim Hyp(N, m, n)$. Det gäller:

$$E(X) = n \cdot p$$

$$V(X) = n \cdot p \cdot (1 - p) \cdot \frac{N - n}{N - 1}$$

$$D(X) = \sqrt{V(X)}$$

$p = \frac{m}{N}$ (andelen "röda/lyckade")

www.matstat.org

26

Hypergeometrisk fördelning

Approximation med binomialfördelningen:

$\frac{n}{N} < 0.1$ tumregel, drar inte mer än 10% av alla

$Hyp(N, m, n) \rightarrow Bin\left(n, \frac{m}{N}\right)$ $p = \frac{m}{N}$ (andelen "röda/lyckade")

t.ex. $N = 10000\,; n = 10 \rightarrow \frac{N-n}{N-1} = \frac{9990}{9999} \approx 1$

$\Rightarrow$

$V(X) = n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1} \approx n \cdot p \cdot (1-p)$ ... som för $Bin(n, p)$

www.matstat.org

27

Poisson-fördelningen

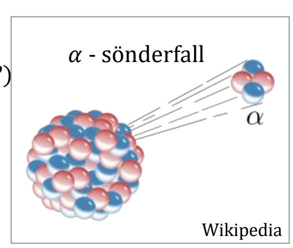
☆ = händelse ("event")

Förekomst och förutsättningar:

- Poisson-fördelningen beskriver sannolikheten att ett visst antal slumpmässiga händelser inträffar i en bestämd tids- eller längdintervall.
- Händelserna (events) måste inträffa helt slumpmässigt (i tid eller rum), dvs. om en (eller flera) händelse(r) har inträffat, så påverkar detta inte antalet händelser som inträffar under ett senare intervall ("minneslöshet")

Exempel för sådana händelser:

- förfrågorna som kommer till en server (oberoende i tid?)
- olyckor på E4:an (oberoende i tid/rum?)
- mejl man får (oberoende i tid?)
- kunder som kommer till en kassa
- radioaktivt sönderfall (oberoende i tid!!)



www.matstat.org

28

Poisson-fördelningen $X \sim Po(\mu)$

Sannolikhetsfunktion: slumpvariabel $X =$ antalet händelser i ett tids-/rumsintervall

$$p_X(k) = \frac{\mu^k}{k!}e^{-\mu} \qquad k = 0, 1, 2, 3, \ldots$$ ... där μ måste vara dimensionslös

$$p_X(k) = \frac{(\lambda t)^k}{k!}e^{-\lambda t} \qquad k = 0, 1, 2, 3, \ldots$$ ... där λ har dimensionen [1/s]

$$p_X(k) = \frac{(\lambda l)^k}{k!}e^{-\lambda l} \qquad k = 0, 1, 2, 3, \ldots$$ ... där λ har dimensionen[1/m]

Väntevärde, varians, standardavvikelse:

$$\begin{aligned} E(X) &= \mu \\ V(X) &= \mu \\ D(X) &= \sqrt{\mu} \end{aligned}$$

”genomsnitt” för tids/längd-intervallet

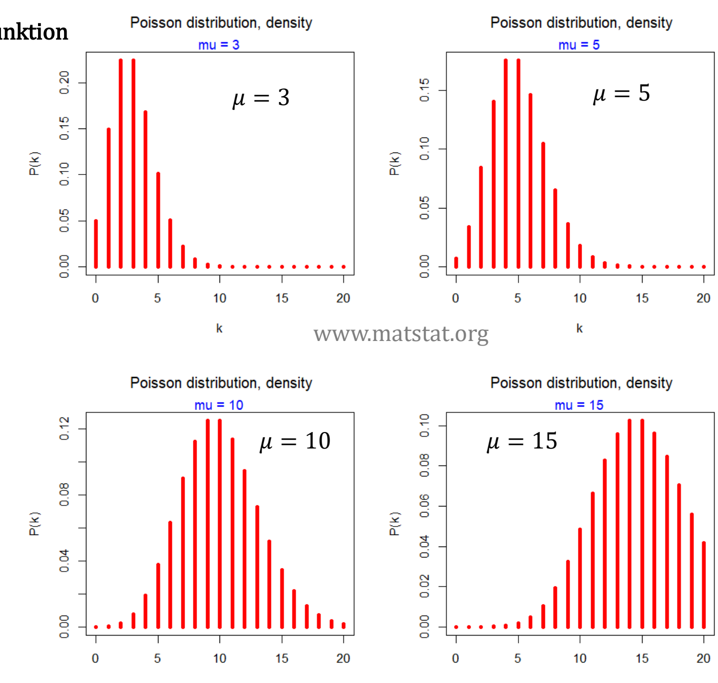
www.matstat.org

29

Sannolikhetsfunktion

$X \sim Po(\mu)$

Olika μ:



www.matstat.org

30

Poisson-fördelningen $X \sim Po(\mu)$

Normering: slumpvariabel $X =$ antalet händelser i ett visst tids-/rumsintervall

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{\mu^k}{k!} \cdot e^{-\mu} = e^{-\mu} \cdot \underbrace{\sum_{k=0}^{\infty} \frac{\mu^k}{k!}}_{= e^{\mu} \text{ (serieutveckling)}} = e^{-\mu} \cdot e^{\mu} = \underline{\underline{1}} \quad ✓$$

Väntevärde :

Subst.: $m = k - 1$

$$E(X) = \sum_{k=0}^{\infty} k \cdot p_X(k) = \sum_{k=0}^{\infty} k \cdot \frac{\mu^k}{k!} \cdot e^{-\mu} = e^{-\mu} \cdot \sum_{k=1}^{\infty} \frac{\mu \cdot \mu^{k-1}}{(k-1)!}$$

$$= e^{-\mu} \cdot \mu \cdot \underbrace{\sum_{m=0}^{\infty} \frac{\mu^m}{m!}}_{= e^{\mu} \text{ (serieutveckling)}} = e^{-\mu} \cdot \mu \cdot e^{\mu} = \underline{\underline{\mu}}$$

www.matstat.org

31

Poisson-fördelningen $X \sim Po(\mu)$

Fördelningsfunktion: slumpvariabel $X =$ antalet händelser i ett visst tids-/rumsintervall

$$F_X(a) = P(X \le a) = \sum_{k=0}^{a} p_X(k) = \sum_{k=0}^{a} \frac{\mu^k}{k!} \cdot e^{-\mu}$$ tabell!

Additionsegenskap:

$$\left.\begin{array}{l} X \sim Po(\mu_1) \\ Y \sim Po(\mu_2) \end{array}\right\} \quad X + Y \sim Po(\mu_1 + \mu_2)$$

OBS!: differensen $X - Y$ är ej Poisson-fördelad!

www.matstat.org

32

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

Fördelningsfunktion:

Tabell:

$X \sim Po(\mu)$

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
5	.99941	.99850	.99680	.99396	.98962	.98344	.97509	.96433	.95096
6	.99992	.99975	.99938	.99866	.99743	.99547	.99254	.98841	.98283
7	.99999	.99996	.99989	.99974	.99944	.99890	.99802	.99666	.99467
8	1.00000	1.00000	.99998	.99995	.99989	.99976	.99953	.99914	.99851
9			1.00000	.99999	.99998	.99995	.99990	.99980	.99962
10				1.00000	1.00000	.99999	.99998	.99996	.99991
11						1.00000	1.00000	.99999	.99998
12								1.00000	1.00000

x \ μ	2.8	3.0	3.2	3.4	3.6	3.8	4.0	4.2	4.4
0	.06081	.04979	.04076	.03337	.02732	.02237	.01832	.01500	.01228

Poisson-fördelningen

Skalering:

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

μ måste vara dimensionslös !

Ibland anges:

$\mu = \lambda \cdot t$ (om det handlar om en process i tiden) eller
$\mu = \lambda \cdot l$ (om det handlar om en process i rummet)

med enheter: $[\lambda] = \frac{1}{s}$ resp. $[\lambda] = \frac{1}{m}$

$$\Rightarrow \quad p_X(k) = \frac{(\lambda \cdot t)^k}{k!} \cdot e^{-\lambda \cdot t}$$

www.matstat.org

Poisson-fördelningen

Exempel: inkommande e-post (Alm s. 89)

En student får i genomsnitt 1.5 e-brev per timme. Vi antar att antalet e-brev per tidsintervall är Poisson-fördelad. S.v. X: antal e-brev i tidsintervall.

$$X \sim Po(\mu) \qquad \mu = \lambda \cdot t \qquad \lambda = \frac{1.5}{h}$$

a) Sannolikheten att det inte kommer något brev under en timme?:

$$t = 1h \qquad \mu = \lambda \cdot t = \frac{1.5}{h} \cdot 1\text{h} = 1.5 \qquad k = 0 \ \text{(inget brev)}$$

$$p_X(0) = \frac{\mu^0}{0!} \cdot e^{-1.5} = e^{-1.5} \cong \underline{\underline{0.2231}}$$

R: dpois(0, 1.5)

b) Sannolikheten att studenten får mer än 3 e-brev under två timmar?:

$$t = 2h \qquad \mu = \lambda \cdot t = \frac{1.5}{h} \cdot 2\text{h} = 3$$

(väntevärdet för 2 timmar!)

$$P(X > 3) = 1 - P(X \leq 3) = 1 - F_X(3) \ \text{där} \ X \sim Po(3)$$

tabell

$$= 1 - 0.6472 = \underline{\underline{0.3528}}$$

R: ppois(3, 3) ger $F_X(3)$

www.matstat.org

35

Poisson-fördelningen

Exempel: inkommande e-post (fortsättning)

En student får i genomsnitt 1.5 e-brev per timme. Vi antar att antalet e-brev per tidsintervall är Poisson-fördelad. S.v. X: antal emissioner i tidsintervall.

$$X \sim Po(\mu) \qquad \mu = \lambda \cdot t \qquad \lambda = \frac{1.5}{h}$$

c) Slh. att studenten får högst tio e-brev under en arbetsdag på 8 timmar?:

$$t = 8h \qquad \mu = \lambda \cdot t = \frac{1.5}{h} \cdot 8\text{h} = 12$$

(väntevärdet för 8 timmar!)

$$P(X \leq 10) = F_X(10) \ \text{där} \ X \sim Po(12)$$

tabell

$$= \underline{\underline{0.3472}}$$

R: ppois(10, 12) beräknar $F_X(10)$

www.matstat.org

36

Poisson-fördelningen

Exempel: kärnfysikaliskt försök I (Blom s. 190, lätt modifierad)

Ett radioaktivt ämne emitterar genomsnittligt 8 partiklar per sekund. Antalet partiklar som emitteras per sekund kan anses som Poisson-fördelat. Slumpvariabel X: antal emissioner per tidsenhet.

$$X \sim Po(\mu) \qquad \mu = \lambda \cdot t \qquad \lambda = \frac{8}{s}$$

a) Sannolikhet att 7 partiklar emitteras under en viss sekund?:

$$t = 1s \qquad \mu = \lambda \cdot t = \frac{8}{s} \cdot 1s = 8 \qquad k = 7 \text{ (7 partiklar)}$$

$$P(X = 7) = p_X(7) = \frac{8^7}{7!} \cdot e^{-8} = \frac{2097152}{5040} \cdot 3.355 \cdot 10^{-4} \cong \underline{\underline{0.1395}}$$

eller:

$$P(X = 7) = p_X(7) = F_X(7) - F_X(6) = 0.453 - 0.3134 = \underline{\underline{0.1396}} \quad \text{tabell}$$

eller: dpois(7,8) → 0.1395865

www.matstat.org

37

Poisson-fördelningen

Exempel: kärnfysikaliskt försök I (fortsättning)

Ett radioaktivt ämne …

$$X \sim Po(\mu) \qquad \mu = \lambda \cdot t \qquad \lambda = \frac{8}{s}$$

b) Sannolikhet att högst 6 partiklar emitteras under en viss sekund?:

$$t = 1s \qquad \mu = \lambda \cdot t = \frac{8}{s} \cdot 1s = 8$$

$$P(X \le 6) = F_X(6) = \underline{\underline{0.3134}}$$

c) Sannolikhet att fler än 8 partiklar emitteras under en viss sekund?:

$$P(X > 8) = 1 - P(X \le 8) = 1 - F_X(8) \text{ där } X \sim Po(8) \quad \text{tabell}$$
$$= 1 - 0.5925 = \underline{\underline{0.4075}}$$

d) Sannolikhet att minst 6 och högst 10 partiklar emitteras under en viss sekund?:

$$P(6 \le X \le 10) = P(5 < X \le 10) = F_X(10) - F_X(5) \text{ där } X \sim Po(8) \quad \text{tabell}$$
$$= 0.8159 - 0.1912 = \cdots$$

$P = ppois(10{,}8) - ppois(5{,}8)$

www.matstat.org

38

Poisson-fördelningen

Exempel: kärnfysikaliskt försök II

Man vet att $X \sim Po$ och har mätt många intervaller (sekunder), så att man vet att slh. att inte få någon signal under en sekund är 0.07. Vilket är det vanligast förekommande antalet signaler? Slumpvariabel X: antal emissioner i en sekund.

$X \sim Po(\mu)$ $\quad$ $\mu = ??$ (vi försöker att beräkna μ först)

$$p_X(0) = 0.07 = \frac{\mu^0}{0!} \cdot e^{-\mu} = e^{-\mu} \;\rightarrow\; -\mu = \ln(0.07) \;\rightarrow\; \mu = 2.66$$

$$p_X(1) = \frac{\mu^1}{1!} \cdot e^{-\mu} = 2.66 \cdot e^{-2.66} = 0.186$$

$$p_X(2) = \frac{\mu^2}{2!} \cdot e^{-\mu} = \frac{2.66^2}{2} \cdot e^{-2.66} = \underline{\underline{0.247}} \;\Leftarrow\; \text{störst!}$$

$$p_X(3) = \frac{\mu^3}{3!} \cdot e^{-\mu} = \frac{2.66^3}{6} \cdot e^{-2.66} = 0.219$$

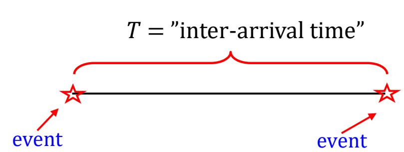
Svar: Det vanligast förekommande antalet emissioner är 2 per sekund.

www.matstat.org

39

Poisson-fördelningen

Väntetid (Inter-arrival time) för Poisson process: $X \sim Po(\lambda \cdot t)$



Kom ihåg att händelserna sker slumpmässigt (oberoende) i tid eller rum (T kan också vara en längd) Därför är T en slumpvariabel.

Hur är T fördelad?:

Sannolikheten att T är större än någon tid/längd t, alltså $P(T > t)$, är lika med sannolikheten att det inte blir någon händelse under t, alltså lika med $p_X(0)$ på intervallet t:

$$\Rightarrow \quad P(T > t) = p_X(0 \mid t)$$

$p_X(0 \mid t)$ = sannolikhet att det blir 0 händelser på intervallet t

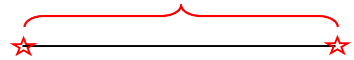
www.matstat.org

40

Poisson-fördelningen

Väntetid (Inter-arrival time) för Poisson process: $X \sim Po(\lambda \cdot t)$

$T =$ ”inter-arrival time”



$\Rightarrow$ $P(T > t) = p_X(0 \mid t)$ — $p_X(0 \mid t) =$ sannolikhet att det blir 0 händelser på intervallet t

$$p_X(k) = \frac{(\lambda \cdot t)^k}{k!} \cdot e^{-\lambda \cdot t}$$ slh. att det blir k händelser (”events”) på intervallet t

$P(T > t) = p_X(0) = e^{-\lambda \cdot t}$ $\Rightarrow$ $P(T \leq t) = 1 - e^{-\lambda \cdot t}$

$\Rightarrow$ $\boxed{F_T(t) = 1 - e^{-\lambda \cdot t}}$ — Detta är fördelningsfunktionen för den exponentiella fördelningen.

Svar: Väntetiden T är exponentiellt fördelad (se föreläsning **F5**).

www.matstat.org

41

Geometrisk fördelning

Man upprepar ett försök som ”lyckas” med sannolikhet p. Man fortsätter så länge tills man lyckas för första gången.
Slumpvariabel X: antal försök tills man lyckas för första gången.

Exempel: Man kastar ett mynt så länge tills man får krona. Vad är sannolikheten att man måste kasta 1 gång, 2 gånger, 3 gånger, … osv.

Sannolikhetsfunktion:

$$P(X = k) = p_X(k) = \underbrace{(1-p)^{k-1}}_{\text{misslyckas } (k-1) \text{ gånger}} \cdot p$$

lyckas en gång, i försök nummer k

Utfallsrum:

$$\Omega_X = \{1, 2, 3, 4, \ldots..\}$$

Kodbeteckning: $X \sim Ge(p)$

OBS: X definieras ibland som antalet misslyckade försök innan den ”lyckade” inträffar!

$\Rightarrow$ $p_X(k) = (1-p)^k \cdot p$ $\quad k = 0, 1, 2, \ldots$ se Wikipedia ”Geometric distribution”

www.matstat.org

42

Geometrisk fördelning $X \sim Ge(p)$

Väntevärde:

$$E(X) = \sum_{k=1}^{\infty} k \cdot p_X(k) = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \cdot p = p \cdot \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1}$$

substitution: $q = 1 - p$

$$E(X) = (1-q) \sum_{k=1}^{\infty} k \cdot q^{k-1} = (1-q) \sum_{k=0}^{\infty} k \cdot q^{k-1} = (1-q) \sum_{k=0}^{\infty} \frac{d}{dq} q^k$$

$$= (1-q) \frac{d}{dq} \sum_{k=0}^{\infty} q^k = (1-q) \frac{d}{dq} \left(\frac{1}{1-q} \right) = (1-q) \cdot \frac{-1}{(1-q)^2} \cdot (-1)$$

$$= \frac{1}{1-q} = \underline{\underline{\frac{1}{p}}}$$

$$\boxed{E(X) = \frac{1}{p}}$$

är p liten behöver man i snitt ett större antal försök för att lyckas för första gången

www.matstat.org

43

Geometrisk fördelning $X \sim Ge(p)$

Varians:

$$\boxed{V(X) = \frac{1-p}{p^2}}$$

p mindre $\rightarrow$ större varians

Standardavvikelse:

$$D(X) = \sqrt{V(X)}$$

Fördelningsfunktion:

se också appendixet

$P(X > k) =$ slh. att misslyckas bland de k första försöken $= (1-p)^k$

$$F_X(k) = P(X \le k) = 1 - P(X > k) = 1 - \underbrace{(1-p)}_{q}{}^k = 1 - q^k$$

$q = 1 - p$ (definition)

$$\boxed{F_X(k) = 1 - q^k}$$

www.matstat.org

44

Geometrisk fördelning

Exempel: tärningskast

Vad är sannolikheten att första femman kommer i det fjärde kastet?:

Slumpvariabel X: antal kast tills femman kommer $X \sim Ge(p)$

$p = \frac{1}{6}$ sannolikhet att lyckas = sannolikhet att kasta en femma

$$P(X = 4) = p_X(4) = \underbrace{(1 - p)^3}_{\text{misslyckas 3 ggr.}} \cdot p = \left(\frac{5}{6}\right)^3 \cdot \frac{1}{6} = \underline{\underline{0.096}}$$

www.matstat.org

45

Appendix

Diskreta fördelningar

Uwe Menzel, 2018
uwe.menzel@matstat.org
www.matstat.org

46

Geometrisk fördelning

$P(X > k) =$ slh. att misslyckas bland de k första försöken $= (1-p)^k$

$$\left.\begin{aligned} P(X > k) = {} & (1-p)^k \cdot p \\ & +(1-p)^k \cdot (1-p) \cdot p \\ & +(1-p)^k \cdot (1-p)^2 \cdot p \\ & +(1-p)^k \cdot (1-p)^3 \cdot p \end{aligned}\right\} \text{... allt som kan hända efter } k \text{ misslyckade försök}$$

$+ \; ...$

$q = 1 - p$ (definition)

$= (1-p)^k \cdot [p + (1-p) \cdot p + (1-p)^2 \cdot p + (1-p)^3 \cdot p + \; ...]$

$= (1-p)^k \cdot p \cdot [1 + q + q^2 + q^3 + q^4 + \; ....]$ geometrisk serie

$= (1-p)^k \cdot p \cdot \frac{1}{1-q} = (1-p)^k \cdot p \cdot \frac{1}{p} = (1-p)^k$ som förut

www.matstat.org

47

Binomialfördelning; Normering

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \cdot y^{n-k}$$ Binomialteorem

$\Downarrow$ sätter: $x = p$, $y = 1 - p$

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

www.matstat.org

48