PODSTAWOWE ROZKŁADY SKOKOWE

Rozpatrujemy jednowymiarowe rozkłady skokowe.

Przypomnienie.
Zmienna losowa ma rozkład skokowy (dyskretny) gdy ma skończony lub przeliczalny zbiór wartości.
Rozkłady skokowe najczęściej określamy przez podanie funkcji prawdopodobieństwa.

Najprostsza (ze względu na liczebność zbioru wartości) funkcja prawdopodobieństwa ma jednoelementowy zbiór wartości.
Rozkład jednopunktowy
Określamy:
P(X = c) = 1 gdzie c ustalona liczba.
Wtedy

Wartość oczekiwana

$$EX = c;$$

Wariancja

$$D^2X = 0$$

(tylko ten rozkład ma zerową wariancję !!!)

Problem.
Na płycie układu cyfrowego jest 16 punktów kontrolnych. W czterech z nich jest stan 1 w pozostałych stan 0. Wybieramy losowo jeden z 16-tu punktów, jakie jest prawdopodobieństwo trafienia na stan 1?

Problemy tego typu (obserwujemy jedno z dwóch zdarzeń) modelujemy rozkładem dwupunktowym.
Rozkład dwupunktowy (zerojedynkowy)
Niech $p \in (0, 1)$ będzie ustaloną liczbą. Określamy:

$$P(X = 0) = q,\ P(X = 1) = p \text{ ; gdzie } q = 1 - p.$$

Umowa:
0 - porażka, q - prawdopodobieństwo porażki,
1 - sukces, p - prawdopodobieństwo sukcesu,

Wartość oczekiwana

$$EX = p,$$

Wariancja

$$D^2X = pq$$

Momenty zwykłe

$$m_2 = m_3 = m_4 = p$$

Momenty centralne

$$\mu_3 = pq(q - p),\ \ \mu_4 = pq(1 - 3pq)$$

Współczynnik asymetrii

$$a = \frac{q - p}{\sqrt{pq}}$$

Kurtoza

$$k = \frac{1}{pq} - 3$$

Funkcja charakterystyczna

$$\varphi(t) = q + pe^{it},$$

Rozwiązanie Problemu:
Szukane prawdopodobieństwo to prawdopodobieństwo sukcesu p. Obliczamy p stosując klasyczną definicję prawdopodobieństwa p = 4/16 = 0,25.

Rozkład ten jest wykorzystywany np. w statystycznej kontroli jakości. Można np. przyjąć, że $X = 0$ gdy wyrób dobry, $X = 1$ gdy wyrób jest wadliwy, wtedy $p = P(X = 1)$ traktujemy jako wadliwość wyrobu.

Problem.
Prawdopodobieństwo uszkodzenia komputera przed upływem gwarancji wynosi 0,2. Firma zakupiła 6 jednakowych komputerów. Obliczyć prawdopodobieństwo, że przed upływem gwarancji 2 komputery ulegną uszkodzeniu. Jaka jest najbardziej prawdopodobna liczba uszkodzonych komputerów przed upływem gwarancji?.
Zakładamy, że komputery te są eksploatowane w tych samych warunkach i psują się niezależnie.
Problemy tego typu (powtarzamy wielokrotnie, niezależnie doświadczenie w tych samych warunkach - **próby Bernoulliego**) modelujemy rozkładem dwumianowym.

Rozkład dwumianowy
Dla danych $p \in (0, 1)$, $n \in N$ określamy funkcję prawdopodobieństwa

$$\boxed{P(X = k) = \binom{n}{k} p^k q^{n-k}} \qquad \text{gdzie } q = 1 - p \qquad \text{(wzór Bernoulliego)}$$

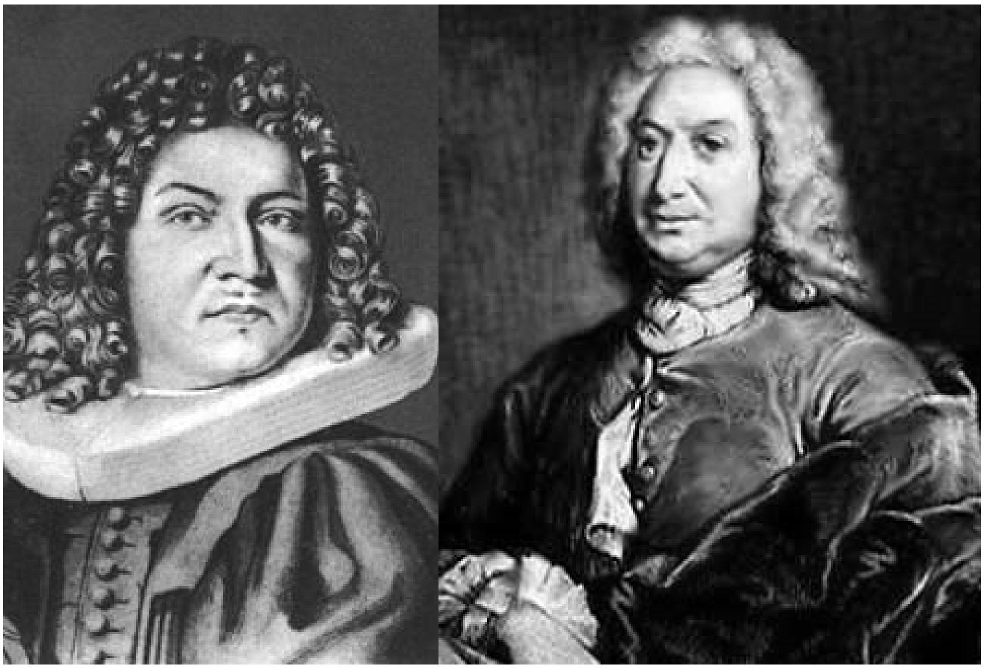
$$k = 0, 1, 2, \ \ldots, n.$$

Zauważmy, że gdy $n = 1$ to rozkład dwumianowy jest rozkładem zerojedynkowym.
Sprawdzenie poprawności:

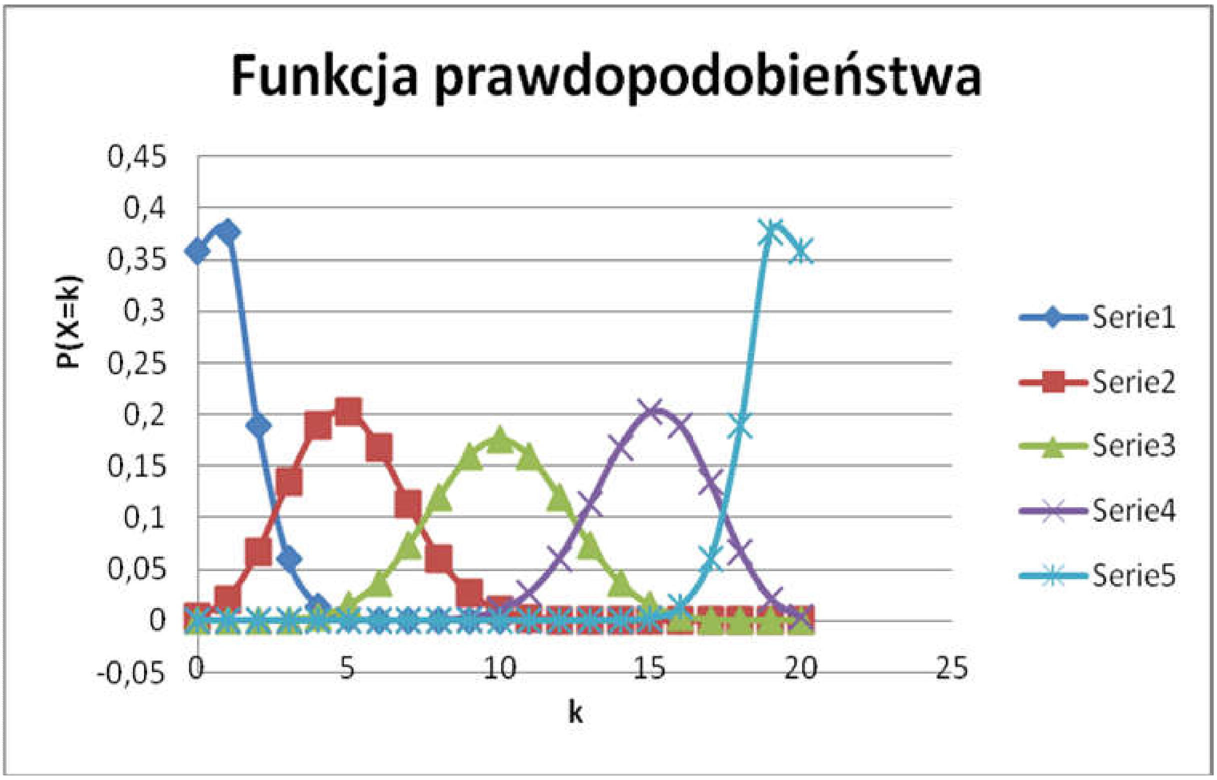
$$\sum_{k=0}^{n} P(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1$$

Jeśli przyjmiemy, że n oznacza liczbę niezależnych doświadczeń z których każde kończy się jednym z dwóch wyników: „sukcesem" (z prawdopodobieństwem p w każdym doświadczeniu) lub „porażką" i zmienna losowa X oznacza liczbę „sukcesów" to powyższy wzór wyznacza **prawdopodobieństwo uzyskania dokładnie *k* sukcesów w *n* doświadczeniach** (próbach).

Własność
Rozkład dwumianowy jest sumą n niezależnych rozkładów zerojedynkowych o tym samym parametrze p.



Jakub Bernoulli (1654 - 1705) - szwajcarski matematyk i fizyk.



Wykresy funkcji prawdopodobieństwa rozkładu dwumianowego dla n = 20.
Dla p = 0,05 - Seria 1,
Dla p = 0,25 - Seria 2,
Dla p = 0,5 - Seria 3,
Dla p = 0,75 - Seria 4,
Dla p = 0,95 - Seria 5,

Wartość oczekiwana

$$EX = np,$$

Wariancja

$$D^2X = npq$$

Momenty zwykłe

$$m_2 = np((n-1)p+1)$$

$$m_3 = np((n-1)(n-2)p^2 + 3(n-1)p+1)$$

$$m_4 = np((n-1)(n-2)(n-3)p^3 + 6(n-1)(n-2)p^2 + 7(n-1)p+1)$$

Momenty centralne

$$\mu_3 = npq(q-p), \quad \mu_4 = npq(1+3pq(n-2))$$

Współczynnik asymetrii

$$a = \frac{q-p}{\sqrt{npq}}$$

Kurtoza

$$k = \frac{1-6pq}{npq} + 3$$

Funkcja charakterystyczna

$$\varphi(t) = \left(q + pe^{it}\right)^n,$$

Własność

Dla rozkładu dwumianowego zachodzi zależność rekurencyjna

$$P(X=k+1) = \frac{n-k}{k+1}\frac{p}{q}P(X=k) \qquad P(X=0) = q^n$$

Rozwiązanie Problemu:

Przyjmujemy, że sukcesem jest uszkodzenie komputera przed upływem gwarancji.

X – liczba uszkodzonych komputerów przed upływem gwarancji,

$$P(X=2) = \binom{6}{2}0{,}2^2 0{,}8^4 = 15 \cdot 0{,}04 \cdot 0{,}4096 = 0{,}24576$$

Funkcję prawdopodobieństwa zmiennej losowej X można przedstawić w tabelce:

x_k	0	1	2	3	4	5	6
p_k	0,2621	**0,3932**	0,2458	0,0819	0,0154	0,0015	0,0001

Zauważmy, że najbardziej prawdopodobną liczba uszkodzonych komputerów jest 1.

Przykład

Rzucamy 4 razy kostką sześcienną. Jakie jest prawdopodobieństwo, że w co najmniej 3 rzutach liczba oczek będzie podzielna przez 3?

Szukane prawdopodobieństwo to

P($X \geq 3$) = $P(X = 3) + P(X = 4)$, gdzie „sukcesem" jest uzyskanie 3 lub 6 oczek, więc $p = 1/3$.

Zatem

$$P(X=3)=\binom{4}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^1=4\cdot\frac{2}{81}=\frac{8}{81}$$

$$P(X=4)=\binom{4}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^0=1\cdot\frac{1}{81}=\frac{1}{81}$$

$$P(X\ge 3)=P(X=3)+P(X=4)=\frac{8}{81}+\frac{1}{81}=\frac{1}{9}$$

Przykład

Obliczymy wartość oczekiwaną rozkładu dwumianowego.

$$EX=\sum_{k=0}^{n}k\binom{n}{k}p^k q^{n-k}=\sum_{k=1}^{n}k\frac{n!}{k!(n-k)!}p^k q^{n-k}=$$

$$=np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}=np(p+q)^{n-1}=np$$

II sposób.

Rozkład dwumianowy jest sumą n niezależnych rozkładów zerojedynkowych X_i o tym samym parametrze p zatem

$$EX=E\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}EX_i=np$$

Analogicznie obliczymy wariancję rozkładu dwumianowego (korzystamy z niezależności).

$$D^2X=D^2\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}D^2X_i=npq$$

Problem.

Rzucamy symetryczną monetą. Obliczyć prawdopodobieństwo, że w trzech pierwszych rzutach wypadnie orzeł a w czwartym rzucie reszka.

Jest to problem wyznaczenia prawdopodobieństwa liczby prób poprzedzających pierwszy sukces, problemy tego typu modelujemy **rozkładem geometrycznym**.

Rozkład geometryczny

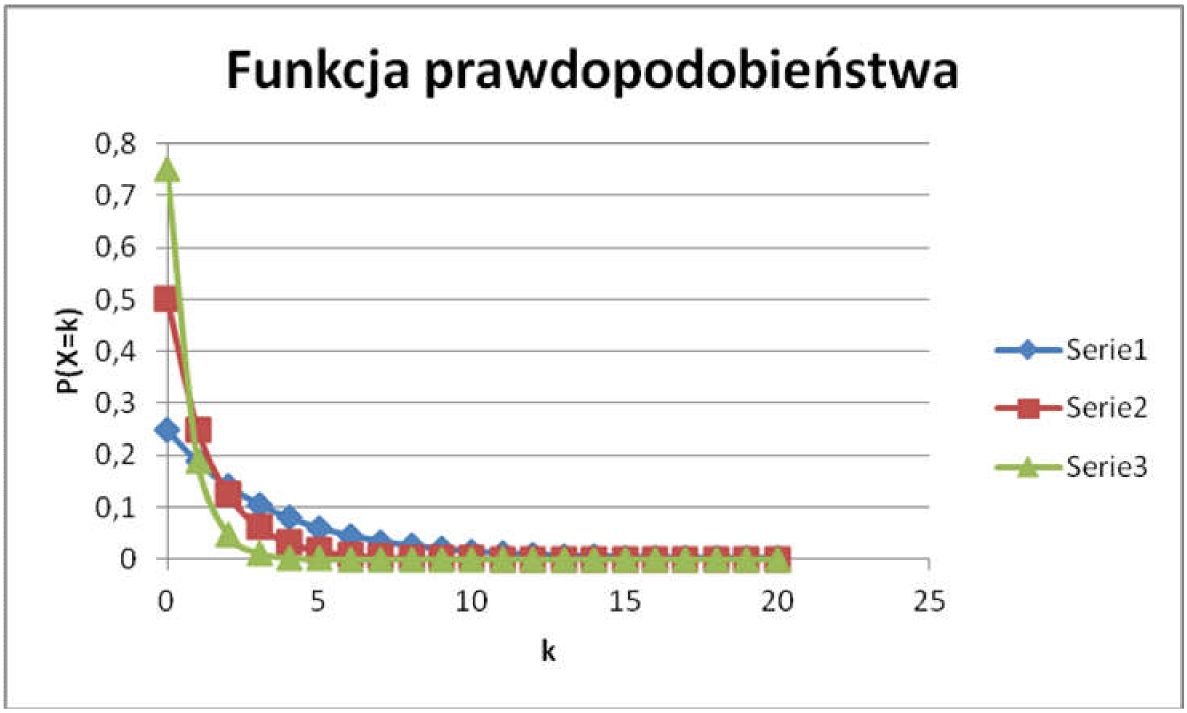
X - liczba prób Bernoulliego **poprzedzających** pierwszy sukces

$$\boxed{P(X=k)=pq^k}$$

q = 1 - p k = 0, 1, 2, ...

Sprawdzenie poprawności

$$\sum_{k=0}^{\infty}P(X=k)=\sum_{k=0}^{\infty}pq^k=\frac{p}{1-q}=1$$



Wykresy funkcji prawdopodobieństwa rozkładu geometrycznego.
Dla p = 0,25 - Seria 1,
Dla p = 0,5 - Seria 2,
Dla p = 0,75 - Seria 3,

Wartość oczekiwana

$$EX = q/p,$$

Wariancja

$$D^2X = q/p^2$$

Momenty zwykłe

$$m_2 = \frac{q(q+1)}{p^2}$$

$$m_3 = \frac{q(q^2+4q+1)}{p^3}$$

$$m_4 = \frac{q(q^3+11q^2+11q+1)}{p^4}$$

Momenty centralne

$$\mu_3 = \frac{q(q+1)}{p^3}, \quad \mu_4 = \frac{q(q^2+7q+1)}{p^4}$$

Współczynnik asymetrii

$$a = \frac{1+q}{\sqrt{q}}$$

Kurtoza

$$k = \frac{p^2}{q} + 9$$

Funkcja charakterystyczna

$$\varphi(t) = \frac{p}{1-qe^{it}},$$

Własność
Dla rozkładu geometrycznego zachodzi zależność rekurencyjna

$$P(X=k+1) = qP(X=k) \qquad P(X=0)=p$$

Przykład
Obliczymy wartość oczekiwaną i wariancję rozkładu geometrycznego.

Dla szeregu geometrycznego mamy $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$

Po zróżniczkowaniu tej równości otrzymamy (1) $\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$

Po kolejnym zróżniczkowaniu otrzymamy (2) $\sum_{k=1}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$

Korzystając z równości (1) mamy

$$EX = \sum_{k=0}^{\infty} kpq^k = pq\sum_{k=1}^{\infty} kq^{k-1} = pq\frac{1}{(1-q)^2} = \frac{q}{p}$$

Korzystając z równości (2) mamy

$$E\left(X^2\right) = \sum_{k=0}^{\infty} k^2pq^k = pq^2\sum_{k=1}^{\infty} k(k-1)q^{k-2} + pq\sum_{k=1}^{\infty} kq^{k-1} = pq^2\frac{2}{(1-q)^3} + \frac{q}{p} = \frac{2q^2+qp}{p^2}$$

Wtedy $D^2X = E\left(X^2\right) - (EX)^2 = \frac{2q^2+qp}{p^2} - \frac{q}{p} = \frac{q^2+qp}{p^2} = \frac{q}{p^2}$

Rozwiązanie Problemu:
Przyjmujemy, że sukcesem jest wyrzucenie reszki.
X - liczba rzutów **poprzedzających** pierwszy sukces
$P(X=3) = 0{,}5^3 0{,}5 = 0{,}5^4 = 0{,}0625$

Przykład
Odbieramy wielokrotnie sygnał binarny z satelity. Prawdopodobieństwo, że sygnał będzie zakłócony wynosi 0,01. Jakie jest prawdopodobieństwo, że pierwszy zakłócony sygnał odbierzemy wśród 200 początkowych odebranych sygnałów.
Przyjmujemy, że sukcesem jest odebranie zakłóconego sygnału.
X - liczba niezakłóconych sygnałów poprzedzających pierwszy zakłócony sygnał.

$$P(X \le 199) = \sum_{k=0}^{199} 0{,}99^k 0{,}01 = 0{,}01\frac{0{,}99^{199}-1}{0{,}99-1} = 1-0{,}99^{199} = 0{,}864667$$

Problem.
System poczty elektronicznej WAT ma średnio 3 awarie na kwartał. Obliczyć prawdopodobieństwo, że w ciągu roku będzie 10 awarii tego systemu.

Zakładając długotrwałą jednorodność średniej liczby awarii możemy modelować ten problem **rozkładem Poissona**.

Rozkład Poissona
Dla $\lambda > 0$ określamy funkcję prawdopodobieństwa

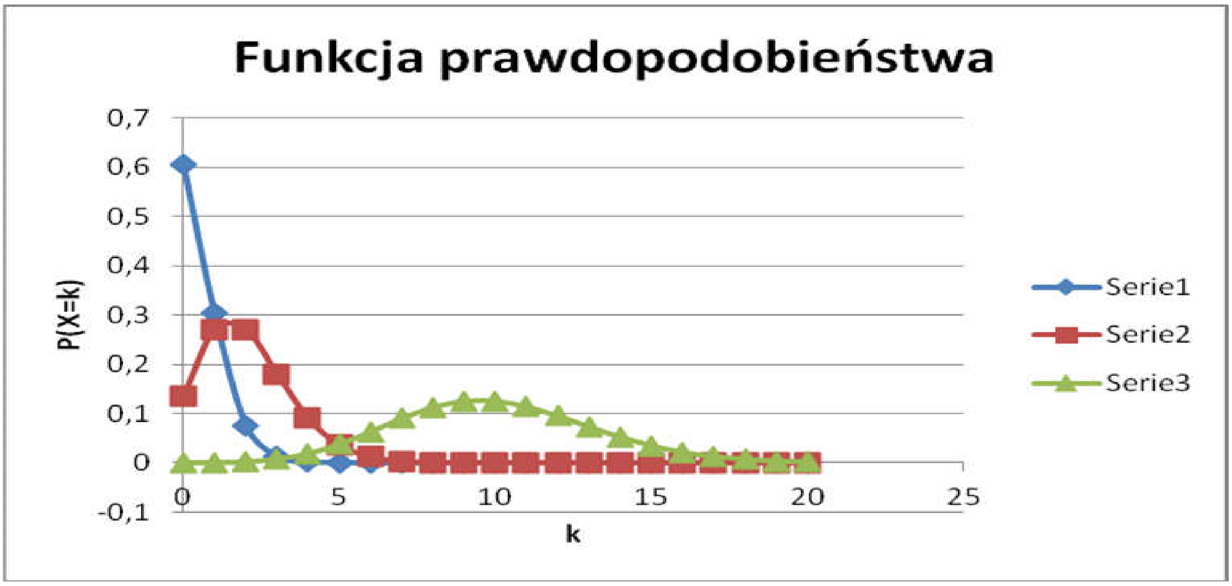
$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \ldots$$

(wartości tych prawdopodobieństw zawiera tablica rozkładu Poissona)

Siméon Denis Poisson (1781–1840), francuski mechanik teoretyk, fizyk i matematyk. W matematyce zajmował się całkami oznaczonymi, równaniami różnicowymi i różniczkowymi oraz teorią prawdopodobieństwa.

Sprawdzenie poprawności

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$



Wykresy funkcji prawdopodobieństwa rozkładu Poissona.
Dla $\lambda = 0{,}5$ - Seria 1,
Dla $\lambda = 2$ - Seria 2,
Dla $\lambda = 10$ - Seria 3,

Wartość oczekiwana

$$EX = \lambda,$$

Wariancja

$$D^2X = \lambda$$

Momenty zwykłe

$$m_2 = \lambda + \lambda^2$$
$$m_3 = \lambda + 3\lambda^2 + \lambda^3$$
$$m_4 = \lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4$$

Momenty centralne

$$\mu_3 = \lambda\,,\quad \mu_4 = \lambda + 3\lambda^2$$

Współczynnik asymetrii

$$a = \frac{1}{\sqrt{\lambda}}$$

Kurtoza

$$k = \frac{1}{\lambda} + 3$$

Funkcja charakterystyczna

$$\varphi(t) = e^{\lambda\left(e^{it}-1\right)},$$

Przykład
Obliczymy wartość oczekiwaną i wariancję rozkładu Poissona.

$$EX = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$$

$$E\left(X^2\right)=\sum_{k=0}^{\infty}k^2\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\sum_{k=1}^{\infty}k\frac{\lambda^k}{(k-1)!}=\lambda e^{-\lambda}\sum_{j=0}^{\infty}(j+1)\frac{\lambda^j}{j!}=\lambda(\lambda+1)$$

Wtedy $D^2X=E\left(X^2\right)-(EX)^2=\lambda(\lambda+1)-\lambda^2=\lambda$

Własność

Dla rozkładu Poissona zachodzi zależność

$$P(X=k+1)=\frac{\lambda}{k+1}P(X=k) \qquad P(X=0)=e^{-\lambda}$$

Własność

dla $\lambda > 9$ rozkład Poissona można przybliżać rozkładem N(λ, $\sqrt{\lambda}$), zachodzi wtedy

$$P(X=k)\approx\Phi\left(\frac{k+0{,}5-\lambda}{\sqrt{\lambda}}\right)-\Phi\left(\frac{k-0{,}5-\lambda}{\sqrt{\lambda}}\right)$$

gdzie Φ - dystrybuanta rozkładu N(0, 1)

Rozwiązanie Problemu:

W ciągu roku średnio będzie $\lambda = 12$ awarii.

X - liczba awarii w ciągu roku.

$P(X=10)=\frac{12^{10}}{10!}e^{-10}=0{,}347$ (odczyt z tablicy rozkładu Poissona)

Rozkład Poissona (możliwość odczytu w tablicy) może dla dużych n (praktycznie $n \geq 30$) i małych p (praktycznie $p \leq 0{,}2$) przybliżać rozkład dwumianowy (**przybliżenie Poissona**)

$$\binom{n}{k}p^kq^{n-k}\approx\frac{\lambda^k}{k!}e^{-\lambda} \qquad \text{gdzie } \lambda=n\cdot p$$

Uzasadnienie.

$$\binom{n}{k}p^kq^{n-k}=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}=\frac{n(n-1)(n-2)...(n-k+1)}{k!}p^k\frac{(1-p)^n}{(1-p)^k}$$

podstawiamy $p=\frac{\lambda}{n}$

$$\binom{n}{k}p^kq^{n-k}=\frac{n(n-1)(n-2)...(n-k+1)}{k!}\frac{\lambda^k}{n^k}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}=$$

$$\frac{n(n-1)(n-2)...(n-k+1)}{n^k}\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

Gdy spełnione są założenia, to
- pierwszy i ostatni czynnik dążą do 1,
- trzeci czynnik dąży do $e^{-\lambda}$,

Stąd $\binom{n}{k} p^k q^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$

Przykład

W pudełku jest 400 żarówek. Jakie jest prawdopodobieństwo, że wśród nich jest 5 żarówek wadliwych, jeśli wadliwość produkcji takich żarówek wynosi 0,5%?
Jaka jest najbardziej prawdopodobna liczba uszkodzonych żarówek w tym pudełku?

Zastosujemy przybliżenie Poissona, $\lambda = n \cdot p = 400 \cdot 0{,}005 = 2$.
W tablicy rozkładu Poissona (tablica I) odczytamy, że:

$$P(X = 5) = 0{,}0361$$

Również w tablicy rozkładu Poissona odczytamy, że najbardziej prawdopodobna liczba uszkodzonych żarówek w tym pudełku to 1 lub 2 (dla obu tych liczb prawdopodobieństwo jest równe 0,2707).

Rozkład Poissona - tablica wartości funkcji prawdopodobieństwa.

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

λ \ k	0	1	2	3	4	5	6	7	8	9	10	11
0,1	0,9048	0,0905	0,0045	0,0002	0,0000							
0,2	8187	1637	0164	0011	0001	0,0000						
0,3	7408	2222	0333	0033	0003	0000						
0,4	6703	2681	0536	0027	0007	0001	0,0000					
0,5	6065	3033	0758	0126	0016	0002	0000					
0,6	5488	3293	0988	0198	0030	0004	0000					
0,7	4966	3476	1217	0284	0050	0007	0001	0,0000				
0,8	4493	3595	1438	0383	0077	0012	0002	0000				
0,9	4066	3659	1646	0494	0111	0020	0003	0000				
1,0	3679	3679	1839	0613	0153	0031	0005	0001	0,0000			
1,5	2231	3347	2510	1255	0471	0141	0035	0008	0001	0,0000		
2,0	1353	2707	2707	1804	0902	0361	0120	0034	0009	0002	0,0000	
2,5	0821	2052	2565	2138	1336	0668	0278	0099	0031	0009	0002	0,0000
3,0	0498	1494	2240	2240	1680	1008	0504	0216	0081	0027	0008	0002
3,5	0302	1057	1850	2158	1888	1322	0771	0385	0169	0066	0023	0007
4,0	0183	0733	1465	1954	1954	1563	1042	0595	0298	0132	0053	0019
4,5	0111	0500	1125	1687	1898	1708	1281	0824	0463	0232	0104	0043
5,0	0067	0337	0842	1404	1755	1755	1462	1044	0653	0363	0181	0082
6,0	0025	0149	0446	0892	1339	1606	1606	1377	1033	0688	0413	0225
7,0	0009	0064	0223	0521	0912	1277	1490	1490	1304	1014	0710	0452
8,0	0003	0027	0107	0286	0573	0916	1221	1396	1396	1241	0993	0722
9,0	0001	0011	0050	0150	0337	0607	0911	1171	1318	1318	1186	0970
10,0	0000	0005	0023	0076	0189	0378	0631	0901	1126	1251	1251	1137

λ \ k	12	13	14	15	16	17	18	19	20	21	22	23	24
0,1													
0,2													
0,3													
0,4													
0,5													
0,6													
0,7													
0,8													
0,9													
1,0													
1,5													
2,0													
2,5													
3,0	0,0001	0,0000											
3,5	0002	0001	0,0000										
4,0	0006	0002	0001	0,0000									
4,5	0016	0006	0002	0001	0,0000								
5,0	0034	0013	0005	0002	0000								
6,0	0113	0052	0022	0009	0003	0,0001	0,0000						
7,0	0264	0142	0071	0033	0014	0006	0002	0,0001	0,0000				
8,0	0481	0296	0169	0090	0045	0021	0009	0004	0002	0,0001	0,0000		
9,0	0728	0504	0324	0194	0109	0058	0029	0014	0006	0003	0001	0,0000	
10,0	0948	0729	0521	0347	0217	0128	0017	0037	0019	0009	0004	0002	0,0001

Problem.
Gra polega na skreśleniu 6 liczb spośród 49. Uczestnik gry wygrywa gdy wśród wylosowanych przez organizatora 6 liczb są co najmniej 3 wytypowane przez gracza. Obliczyć prawdopodobieństwo, że gracz wygra?.

Problemy tego typu (losowanie bez zwracania gdy wśród elementów jest pewien wyróżniony podzbiór) modelujemy rozkładem **hipergeometrycznym**.

Rozkład hipergeometryczny
Dla danej liczby obiektów N z których M ma określoną własność losujemy n elementów bez zwracania. X - liczba wylosowanych obiektów o określonej własności określamy funkcję prawdopodobieństwa

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

gdzie n, N, M to liczby całkowite nieujemne, $M \le N$, $n \le N$,
$\max(0, M + n - N) \le k \le \min(M, n)$

Sprawdzenie poprawności:
Korzystamy z twierdzenia Vandermonda

$$\sum_{k=0}^{n}\binom{a}{k}\binom{b}{n-k}=\binom{a+b}{n}$$

wtedy

$$\sum_{k=0}^{n} P(X=k) = \sum_{k=0}^{n} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}} \sum_{k=0}^{n} \binom{M}{k}\binom{N-M}{n-k} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1$$

Wartość oczekiwana

EX = Mn/N,

Wariancja

$$D^2X = \frac{nM(N-M)(N-n)}{N^2(N-1)}$$

Momenty zwykłe

$$m_2 = \frac{nM}{N}\left[\frac{(n-1)(M-1)}{N-1} + 1\right]$$

$$m_3 = \frac{nM}{N}\left[\frac{(n-1)(n-2)(M-1)(M-2)}{(N-1)(N-2)} + \frac{3(n-1)(M-1)}{N-1} + 1\right]$$

$$m_4 = \frac{nM}{N}\left[\frac{(n-1)(n-2)(n-3)(M-1)(M-2)(M-3)}{(N-1)(N-2)(N-3)} + \right.$$

$$\left. + \frac{6(n-1)(n-2)(M-1)(M-2)}{(N-1)(N-2)} + \frac{7(n-1)(M-1)}{N-1} + 1\right]$$

Momenty centralne

$$\mu_3 = \frac{nM(N-M)(N-2M)(N-n)(N-2n)}{N^3(N-1)(N-2)},$$

$$\mu_4 = \frac{nM(N-M)(N-n)}{N^4(N-1)(N-2)(N-3)}\left(N^3(N+1) - 6nN^2(N-n) + 3M(N-M)(n(N-n)(N+6) - 2N^2)\right)$$

Współczynnik asymetrii

$$a = \frac{(N-2M)(N-2n)\sqrt{N-1}}{(N-2)\sqrt{nM(N-M)(N-n)}}$$

Kurtoza

$$k = \frac{N^2(N-1)}{n(N-2)(N-3)(N-n)}\left(\frac{N(N+1) - 6n(N-n)}{M(N-M)} + \frac{6n(N-n)(5N-6)}{N^2(N-1)} - 6\right) + 3$$

Własność

Dla rozkładu hipergeometrycznego zachodzi zależność

$$P(X=k) = P(X=k-1)\frac{(n+1-k)(M+1-k)}{k(N-M-n+k)}$$

Rozwiązanie Problemu:

W tym przypadku n = 6, N = 49, M =6.

X – liczba odgadniętych liczb,

wtedy

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

Prawdopodobieństwa dla k od 3 do 6 można przedstawić w tabelce:

k	3	4	5	6
p_k	0,017650403867	0,000968619724	0,000018449900	0,000000071511

Suma tych prawdopodobieństw wynosi około 0,018637545002 i jest to szukane prawdopodobieństwo.

.

Przykład
Obliczymy wartość oczekiwaną i wariancję rozkładu hipergeometrycznego.

$$EX=\sum_{k=0}^{n}k\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}=\frac{nM}{N}\sum_{k=1}^{n}\frac{\binom{M-1}{k-1}\binom{N-M}{n-k}}{\binom{N-1}{n-1}}=$$

$$=\frac{nM}{N}\sum_{j=1}^{n-1}\frac{\binom{M-1}{j}\binom{N-1-M+1}{n-1-j}}{\binom{N-1}{n-1}}=\frac{nM}{N}$$

Ostatnia równość wynika z przytoczonego wyżej twierdzenia Vandermonda.

Obliczymy wariancję rozkładu hipergeometrycznego.

$$D^2X=EX^2-(EX)^2=E(X(X-1))+EX-(EX)^2$$

$$E(X(X-1))=\sum_{k=0}^{n}k(k-1)\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}=\frac{n(n-1)M(M-1)}{N(N-1)}\sum_{k=2}^{n}\frac{\binom{M-2}{k-2}\binom{N-M}{n-k}}{\binom{N-2}{n-2}}=$$

$$=\frac{n(n-1)M(M-1)}{N(N-1)}\sum_{j=1}^{n-2}\frac{\binom{M-2}{j}\binom{N-M}{n-2-j}}{\binom{N-2}{n-2}}=\frac{n(n-1)M(M-1)}{N(N-1)}$$

Ostatnia równość wynika też z przytoczonego wyżej twierdzenia Vandermonda.
Wtedy

$$D^2X=\frac{n(n-1)M(M-1)}{N(N-1)}+\frac{nM}{N}-\left(\frac{nM}{N}\right)^2=\frac{nM(N-M)(N-n)}{N^2(N-1)}$$

Problem.
Rzucamy symetryczną kostką sześcienną. Rzuty wykonujemy aż wypadnie 6 oczek dwa razy (nie koniecznie kolejno). Obliczyć prawdopodobieństwo, że nie uda nam się to w pięciu rzutach a uda w szóstym rzucie.

Jest to problem wyznaczenia prawdopodobieństwa gdy wykonujemy ciąg prób Bernoulliego lecz nie narzucamy z góry liczby prób lecz liczbę pożądanych sukcesów, problemy tego typu modelujemy **rozkładem ujemnym dwumianowym**.

Rozkład ujemny dwumianowy.
Niech m - liczba pożądanych sukcesów.

X - liczba prób Bernoulliego **poprzedzających** m sukcesów.

Dla danych $p \in (0, 1)$, $m \in N$ określamy funkcję prawdopodobieństwa

$$\boxed{P(X=k) = \binom{k+m-1}{k} p^m q^k} \quad \text{gdzie } q = 1-p$$

$k = 0, 1, 2, \ ...$
Zauważmy, że gdy $m = 1$ to rozkład ujemny dwumianowy jest rozkładem geometrycznym.

Uwaga.
Określając uogólnienie symbolu Newtona

$$\binom{c}{k} = \frac{c(c-1)(c-2)....(c-k+1)}{k!} \qquad c \in R,\ \ k \in N \ \text{ oraz } \binom{c}{0} = 1$$

możemy wprost udowodnić, że $\binom{-m}{k} = (-1)^k \binom{k+m-1}{k}$

i zapisać funkcję prawdopodobieństwa rozkładu ujemnego dwumianowego w postaci

$$\boxed{P(X=k) = \binom{-m}{k} p^m (-q)^k}$$

Sprawdzenie poprawności:

Chcemy udowodnić, że $\sum\limits_{k=0}^{\infty} P(X=k) = \sum\limits_{k=0}^{\infty} \binom{k+m-1}{k} p^m q^k = \sum\limits_{k=0}^{\infty} \binom{-m}{k} p^m (-q)^k = 1$

Korzystając z twierdzenia dwumianowego

$$(x+y)^r = \sum_{k=0}^{\infty} \binom{r}{k} x^k y^{r-k} \text{, gdzie } r \in R,\ \ y > 0,\ \left|\frac{x}{y}\right| < 1$$

mamy $\qquad (1-q)^{-m} = \sum\limits_{k=0}^{\infty} \binom{-m}{k} (-q)^k$

i mnożąc obie strony przez p^m otrzymamy potrzebną równość.

Własność
Rozkład ujemny dwumianowy jest sumą m niezależnych rozkładów geometrycznych o tym samym parametrze p.

Wartość oczekiwana

$$EX = mq/p,$$

Wariancja

$$D^2X = mq/p^2$$

Momenty zwykłe

$$m_2 = \frac{mq(mq+1)}{p^2}$$

$$m_3 = \frac{mq(m^2q^2 + (3m+1)q + 1)}{p^3}$$

$$m_4 = \frac{mq(m^3q^3 + (6m^2 + 4m + 1)q^2 + (7m+4)q + 1)}{p^4}$$

Momenty centralne

$$\mu_3 = \frac{mq(q+1)}{p^3}, \quad \mu_4 = \frac{mq(q^2 + (3m+4)q + 1)}{p^4}$$

Współczynnik asymetrii

$$a = \frac{1+q}{\sqrt{mq}}$$

Kurtoza

$$k = \frac{p^2}{mq} + \frac{6}{m} + 3$$

Funkcja charakterystyczna

$$\varphi(t) = \left(\frac{p}{1 - qe^{it}}\right)^m,$$

Własność

Dla rozkładu ujemnego dwumianowego zachodzi zależność

$$P(X = k) = P(X = k-1)\frac{m-1+k}{k} \qquad P(X=0) = p^m$$

Przykład

Obliczymy wartość oczekiwaną rozkładu dwumianowego.

Rozkład ujemny dwumianowy jest sumą m niezależnych rozkładów geometrycznych X_i o tym samym parametrze p zatem

$$EX = E\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} EX_i = \frac{mq}{p}$$

Analogicznie obliczymy wariancję rozkładu ujemnego dwumianowego (korzystamy z niezależności).

$$D^2X = D^2\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} D^2X_i = \frac{mq}{p^2}$$

Rozwiązanie Problemu:

Prawdopodobieństwo sukcesu wynosi 1/6, chcemy uzyskać dwa sukcesy dopiero w szóstym rzucie.

Zatem $P(X=5) = \binom{5+2-1}{5}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^5 = 0{,}06698$

Przykład. Zadanie Banacha o zapałkach.
Pewien matematyk do zapalania papierosów posługiwał się dwoma pudełkami zapałek (jedno w lewej kieszeni i jedno w prawej kieszeni), wyciągając je na chybił-trafił. Po pewnym czasie stwierdził, że jedno jest puste. Jakie jest prawdopodobieństwo, że w drugim pudełku jest wtedy k zapałek, jeżeli początkowo w każdym pudełku było n zapałek?
Przyjmijmy, że sukcesem jest wybór lewej kieszeni, p = 0,5. Zatem n - k porażek poprzedza n + 1 sukces. To samo stosuje się do drugiej kieszeni, zatem szukane prawdopodobieństwo wynosi $2\binom{n-k+n+1-1}{n-k}0{,}5^{n+1}0{,}5^{n-k} = \binom{2n-k}{n}2^{-2n+k}$
Np. gdy n = 50 i k = 10 to szukane prawdopodobieństwo wynosi 0,048363.

Problem.
Spośród 100 ponumerowanych losów wybieramy jeden. Jakie jest prawdopodobieństwo, że zostanie wybrany los o numerze 13?.

Jest to problem wyznaczenia prawdopodobieństwa gdy losujemy jeden element bez preferencji, tzn. wybór każdego elementu jest tak samo prawdopodobny. Problemy tego typu modelujemy **rozkładem dyskretnym jednostajnym**.

Rozkład dyskretny jednostajny.

Funkcja rozkładu prawdopodobieństwa.

c, n - całkowite; n > 0

$$P(X=k)=\frac{1}{n} \quad k = c, c+1, c+2, ..., c+n-1$$

Parametry:

Wartość oczekiwana

$$EX = c + (n-1)/2;$$

Wariancja

$$D^2X = (n^2-1)/12$$

Odchylenie standardowe

$$DX = \sqrt{\frac{n^2-1}{12}}$$

Momenty zwykłe

$$m_2 = \frac{2n^2+3(2c-1)n+1}{6}+c(c-1)$$

$$m_3 = \frac{n^3+2(2c-1)n^2+(6c^2-6c+1)n}{4}+c(c^2-1{,}5c+0{,}5)$$

$$m_4 = \frac{6n^4+15(2c-1)n^3+10(6c^2-6c+1)n^2+30c(2c^2-3c+1)n-1}{30}+c^2(c^2-2c+1)$$

Momenty centralne

$$\mu_3 = 0$$

$$\mu_4 = \frac{(n^2-1)(3n^2-7)}{240}$$

Współczynnik asymetrii

$$a = 0$$

Kurtoza

$$k = 1{,}8 - 2{,}4/(n^2 - 1)$$

Funkcja charakterystyczna.

$$\varphi(t) = \frac{e^{ict}\left(1 - e^{\text{int}}\right)}{n\left(1 - e^{it}\right)}$$

Rozwiązanie Problemu:
X - numer wylosowanego losu. Zbiór wartości jest 100 elementowy więc każdy los może być wybrany z prawdopodobieństwem 0,01.
Zatem $P(X = 13) = 0{,}01$.

Pytania kontrolne.

1. Szczególnym przypadkiem rozkładu dwumianowego jest rozkład ....

2. Szczególnym przypadkiem rozkładu ujemnego dwumianowego jest rozkład ....

3. Rozkład dwumianowy jest symetryczny dla p = ...

4. Zbiór wartości zmiennej losowej o rozkładzie Poissona to zbiór ....

5. Funkcja prawdopodobieństwa rozkładu geometrycznego ma największą wartość dla k = ...

Pytania testowe.

1. Rozpatrujemy zdarzenie (3 < X < 5).
Dla którego z rozkładów Poissona z parametrem λ prawdopodobieństwo tego zdarzenia jest najmniejsze:

A λ = 2 B λ = 0,5 C λ = 1,5 D λ = 0,25

2. Funkcja prawdopodobieństwa zmiennej losowej X ma postać

$P(X = k) = \binom{25}{k}\left(\frac{1}{5}\right)^k\left(\frac{4}{5}\right)^{25-k}$ k = 0, 1, ..., 25. Wartość oczekiwana EX wynosi:

A 5 B 1/3 C 15 D 2/3

3. Rozpatrujemy zdarzenie (-1 < X < 1). Dla którego z rozkładów dwumianowych z parametrem $n = 10$ i p prawdopodobieństwo tego zdarzenia jest najmniejsze:

A p = 0,8 B p = 0,85 C p = 0,95 D p = 0,9

4. Rozpatrujemy zdarzenie (-1 < X < 1). Dla którego z rozkładów geometrycznych z parametrem p prawdopodobieństwo tego zdarzenia jest najmniejsze:

A p = 0,1 B p = 0,5 C p = 0,75 D p = 0,9

5. Wartość oczekiwana rozkładu dwumianowego z parametrem $n = 10$ wynosi 2. Wtedy dystrybuanta tego rozkładu jest równe:

A 0,2 B 1 C 1,6 D 20

L.Kowalski 24.10.16