Rachunek prawdopodobieństwa i statystyka

Wykład 6

Warunkowe wartości oczekiwane

Wartością oczekiwaną dyskretnej zmiennej losowej k pod warunkiem, że zmienna losowa m przyjmuje wartość *m* nazywamy wielkość:

$$\mathcal{E}[\mathbf{k} \,|\, \mathbf{m}=m] \equiv \mu_{\mathrm{k}}(m) \equiv \sum_k k\,\mathrm{P}(\mathbf{k}=k \,|\, \mathbf{m}=m) = \frac{1}{\mathrm{P}_{\bullet,\mathrm{m}}}\sum_k k\,\mathrm{P}_{\mathrm{km}}$$

Wartością oczekiwaną ciągłej zmiennej losowej x pod warunkiem, że zmienna losowa y przyjmuje wartość *y* nazywamy wielkość:

$$\mathcal{E}[\mathbf{x} \,|\, \mathbf{y}=y] \equiv \mu_{\mathrm{x}}(y) \equiv \int_{-\infty}^{+\infty} x \cdot f(x \,|\, y)\,dx = \frac{1}{f_2(y)}\int_{-\infty}^{+\infty} x \cdot f(x,y)\,dx$$

Podobnie określamy wartość oczekiwaną dyskretnej zmiennej losowej m pod warunkiem, że zmienna losowa k przyjmuje wartość *k*:

$$\mathcal{E}[\mathbf{m} \,|\, \mathbf{k}=k] \equiv \mu_{\mathrm{m}}(k) \equiv \sum_m m\,\mathrm{P}(\mathbf{m}=m \,|\, \mathbf{k}=k) = \frac{1}{\mathrm{P}_{k,\bullet}}\sum_m m\,\mathrm{P}_{\mathrm{km}}$$

oraz wartość oczekiwaną ciągłej zmiennej losowej y pod warunkiem, że zmienna losowa x przyjmuje wartość *x*:

$$\mathcal{E}[\mathbf{y} \,|\, \mathbf{x}=x] \equiv \mu_{\mathrm{y}}(x) \equiv \int_{-\infty}^{+\infty} y \cdot f(y \,|\, x)\,dy = \frac{1}{f_1(x)}\int_{-\infty}^{+\infty} y \cdot f(x,y)\,dy$$

Linie regresji I-go rodzaju

Linią regresji I-go rodzaju zmiennej losowej $\mathbf{k}$ ($\mathbf{x}$) względem zmiennej losowej $\mathbf{m}$ ($\mathbf{y}$) nazywamy zbiór punktów o współrzędnych (k,m) w przypadku zmiennych dyskretnych i (x,y) w przypadku zmiennych ciągłych, spełniających równania:

$$k = \mu_k(m) \qquad\qquad x = \mu_x(y)$$

Analogicznie, **linią regresji I-go rodzaju** zmiennej losowej $\mathbf{m}$ ($\mathbf{y}$) względem zmiennej $\mathbf{k}$ ($\mathbf{x}$) nazywamy, odpowiednio dla zmiennych dyskretnych i ciągłych, zbiory punktów (k,m) lub (x,y), spełniające równania:

$$m = \mu_m(k) \qquad\qquad y = \mu_y(x)$$

Własności:

Średnie odchylenie kwadratowe zmiennej losowej $\mathbf{x}$ od pewnej funkcji $g(y)$ jest najmniejsze, gdy ta funkcja z p-twem 1 jest równa μ_x(y):

$$\mathcal{E}\left[\left(\mathbf{x}-\mu_x(y)\right)^2\right] = \min_g \mathcal{E}\left[\left(\mathbf{x}-g(y)\right)^2\right]$$

Średnie odchylenie kwadratowe zmiennej losowej $\mathbf{y}$ od pewnej funkcji $f(x)$ jest najmniejsze, gdy ta funkcja z p-twem 1 jest równa μ_y(x):

$$\mathcal{E}\left[\left(\mathbf{y}-\mu_y(x)\right)^2\right] = \min_g \mathcal{E}\left[\left(\mathbf{y}-f(x)\right)^2\right]$$

Linie regresji II-go rodzaju

Linią regresji II-go rodzaju nazywamy zadaną krzywą $y = h(x; a,b,\ldots)$ gdy spełnia ona warunek:

$$\mathcal{E}\left[\left(\mathrm{y}-h(x;a,b,\ldots)\right)^2\right]=\min(a,b,\ldots)$$

W szczególności **prostą regresji II-go rodzaju** nazywamy linię prostą $y = ax+b$ spełniającą warunek:

$$\mathcal{E}\left[(\mathrm{y}-a\,\mathrm{x}-b)^2\right]=\mathcal{E}\left[\left((\mathrm{y}-\mu_\mathrm{y})-a(\mathrm{x}-\mu_\mathrm{x})+\mu_\mathrm{y}-a\mu_\mathrm{x}-b\right)^2\right]=$$

$$=\sigma_\mathrm{y}^2+a^2\sigma_\mathrm{x}^2-2a\,\mathrm{cov}[\mathrm{x},\mathrm{y}]+\left(\mu_\mathrm{y}-a\mu_\mathrm{x}-b\right)^2=\sigma_\mathrm{y}^2+a^2\sigma_\mathrm{x}^2-2a\sigma_\mathrm{x}\sigma_\mathrm{y}\rho+\left(\mu_\mathrm{y}-a\mu_\mathrm{x}-b\right)^2=\min(a,b)$$

Szukamy minimum ze względu na parametry a i b:

$$\begin{cases}\dfrac{\partial}{\partial a}\mathcal{E}\left[(\mathrm{y}-a\,\mathrm{x}-b)^2\right]=2a\sigma_\mathrm{x}^2-2\sigma_\mathrm{x}\sigma_\mathrm{y}\rho-2\mu_\mathrm{x}\left(\mu_\mathrm{y}-a\mu_\mathrm{x}-b\right)=0\\[2ex]\dfrac{\partial}{\partial b}\mathcal{E}\left[(\mathrm{y}-a\,\mathrm{x}-b)^2\right]=-2\left(\mu_\mathrm{y}-a\mu_\mathrm{x}-b\right)=0\end{cases}\Rightarrow\begin{cases}a=\rho\dfrac{\sigma_\mathrm{y}}{\sigma_\mathrm{x}}\\[2ex]b=\mu_\mathrm{y}-a\mu_\mathrm{x}\end{cases}$$

Prosta regresji II-go rodzaju zmiennej y względem zmiennej x ma postać:

$$y=\rho\frac{\sigma_\mathrm{y}}{\sigma_\mathrm{x}}x-\rho\frac{\sigma_\mathrm{y}}{\sigma_\mathrm{x}}\mu_\mathrm{x}+\mu_\mathrm{y}$$

Podobnie znajdujemy prostą regresji II-go rodzaju zmiennej x względem y:

$$x=\rho\frac{\sigma_\mathrm{x}}{\sigma_\mathrm{y}}y-\rho\frac{\sigma_\mathrm{x}}{\sigma_\mathrm{y}}\mu_\mathrm{y}+\mu_\mathrm{x}\qquad\Rightarrow\qquad y=\frac{1}{\rho}\frac{\sigma_\mathrm{y}}{\sigma_\mathrm{x}}x-\frac{1}{\rho}\frac{\sigma_\mathrm{y}}{\sigma_\mathrm{x}}\mu_\mathrm{x}+\mu_\mathrm{y}$$

Linie regresji I-go rodzaju

<u>Przykład:</u> Linie regresji I-go rodzaju dla dwuwymiarowego rozkładu normalnego.

$$\mathcal{N}\left(x, y; \mu_x, \mu_y, \sigma_x, \sigma_y\right) =$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\left(1-\rho^2\right)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]\right\}$$

Rozkłady brzegowe to jednowymiarowe rozkłady normalne:

$$\mathcal{N}_1\left(x; \mu_x, \sigma_x\right) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{\left(x-\mu_x\right)^2}{2\sigma_x^2}\right) \qquad \mathcal{E}[\mathrm{x}] = \mu_x \qquad \mathcal{V}[\mathrm{x}] = \sigma_x^2$$

$$\mathcal{N}_2\left(y; \mu_y, \sigma_y\right) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{\left(y-\mu_y\right)^2}{2\sigma_y^2}\right) \qquad \mathcal{E}[\mathrm{y}] = \mu_y \qquad \mathcal{V}[\mathrm{y}] = \sigma_y^2$$

Rozkłady warunkowe i warunkowe wartości oczekiwane:

$$g(y \mid x) = \frac{\mathcal{N}\left(x, y; \mu_x, \mu_y, \sigma_x, \sigma_y\right)}{\mathcal{N}_1\left(x; \mu_x, \sigma_x\right)} = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2\sigma_y^2\left(1-\rho^2\right)}\left[y - (\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x))\right]^2\right)$$

$$g(x \mid y) = \frac{\mathcal{N}\left(x, y; \mu_x, \mu_y, \sigma_x, \sigma_y\right)}{\mathcal{N}_2\left(y; \mu_y, \sigma_y\right)} = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2\sigma_x^2\left(1-\rho^2\right)}\left[x - (\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y))\right]^2\right)$$

Linie regresji I-go rodzaju

Linie regresji I-go rodzaju to proste o równaniach:

$$y = \mathcal{E}[y \mid x] = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

$$x = \mathcal{E}[x \mid y] = \mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y) \quad \Rightarrow \quad y = \mu_y + \frac{1}{\rho}\frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

Uwaga: W przypadku dwuwymiarowego rozkładu normalnego, linie regresji I-go rodzaju są jednocześnie prostymi regresji II-go rodzaju.

Linie regresji I-go rodzaju

<u>Przykład:</u> Chcemy umieć przewidzieć wzrost pewnego gatunku żyta na podstawie koncentracji fosforu w glebie. W tym celu badany cztery koncentracje 2, 4, 8, 16 ppm fosforu, i obserwujemy wzrost roślin przy każdej z nich począwszy od nasiona aż do momentu kwitnienia. Następnie dokonujemy pomiaru suchej masy każdej z roślin. Wyniki pomiarów przedstawia tabelka.

Roślina	Fosfor [ppm]	Masa [g]
1	2	4.1
2	2	3.8
3	2	4.0
4	2	3.9
5	4	5.2
6	4	4.9
7	4	5.0
8	4	4.8
9	8	5.7
10	8	5.9
11	8	6.0
12	8	6.2
13	16	11.7
14	16	8.9
15	16	10.1
16	16	10.3

Zmienne losowe: x - koncentrację fosforu, y - masa roślin.

$$\mu_x \equiv \mathcal{E}[x] = \frac{1}{16}(4\times 2 + 4\times 4 + 4\times 8 + 4\times 16) = 7.5$$

$$\mu_y \equiv \mathcal{E}[y] = \frac{1}{16}(4.1 + 3.8 + 4.0 + 3.9 + 5.2 + 4.9 + 5.0 + 4.8 + 5.7$$

$$+5.9 + 6.0 + 6.2 + 11.7 + 8.9 + 10.1 + 10.3) = \frac{100.5}{16} \cong 6.28$$

$$\mathcal{E}[x^2] = \frac{1360}{16} = 85 \qquad \mathcal{E}[y^2] = \frac{727.49}{16} = 45.468 \qquad \mathcal{E}[xy] = \frac{957.6}{16} = 59.85$$

$$\mathrm{cov}[x,y] = \mathcal{E}[xy] - \mathcal{E}[x]\mathcal{E}[y] = 59.85 - 7.5\times 6.28 = 12.75$$

$$\mathcal{D}[x] = \sqrt{\mathcal{E}[x^2] - (\mathcal{E}[x])^2} = \sqrt{85 - 7.5^2} = 5.36$$

$$\mathcal{D}[y] = \sqrt{\mathcal{E}[y^2] - (\mathcal{E}[y])^2} = \sqrt{45.468 - 6.28^2} = 2.46$$

$$\rho = \frac{\mathrm{cov}[x,y]}{\mathcal{D}[x]\mathcal{D}[y]} = 0.967$$

$$y = 0.44x + 2.96$$

Ważniejsze rozkłady p-twa

Rozkłady zmiennej losowej dyskretnej:

- rozkład płaski (jednostajny)
- rozkład dwupunktowy (Bernoulliego)
- rozkład dwu- i wielomianowy
- rozkład ujemny dwumianowy (Pascala)
- rozkład geometryczny
- rozkład hipergeometryczny
- rozkład Poissona

Rozkłady zmiennej losowej ciągłej:

- rozkład płaski (jednostajny)
- rozkład wykładniczy i Erlanga
- rozkład normalny (Gaussa)
- rozkład Studenta
- rozkład χ^2
- rozkład $\mathcal{F}$ Snedecora-Fishera

Rozkład płaski - zmienna dyskretna

W rozkładzie płaskim (jednostajnym, stałym, jednorodnym) każdy spośród n możliwych wyników losowania jest równie prawdopodobny:

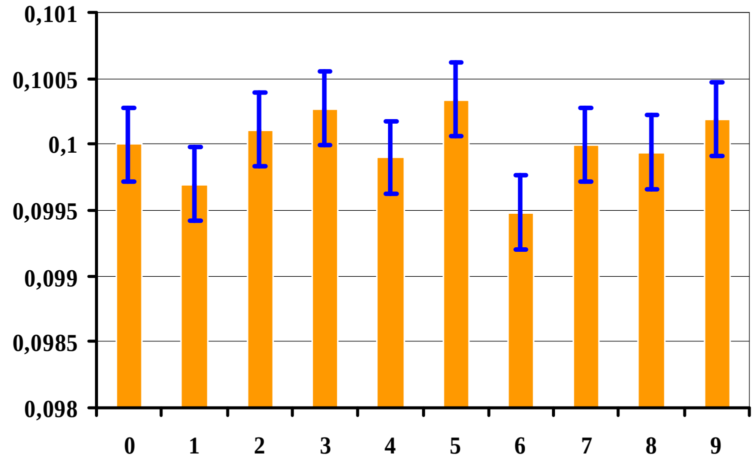
$$\boxed{P_k(n) = \frac{1}{n} \quad \text{dla} \quad k = 1, 2, \ldots, n}$$

Wartość oczekiwana: $\mathcal{E}[\mathrm{k}] = \sum_{k=1}^{n} k \frac{1}{n} = \frac{1}{n} \sum_{k=1}^{n} k = \frac{1}{n} \frac{1}{2} n(n+1) = \boxed{\frac{1}{2}(n+1)}$

Wariancja:

$$\mathcal{V}[\mathrm{k}] = \mathcal{E}[\mathrm{k}^2] - (\mathcal{E}[\mathrm{k}])^2 = \sum_{k=1}^{n} k^2 \frac{1}{n} - (\mathcal{E}[\mathrm{k}])^2 = \frac{1}{n} \frac{1}{6} n(n+1)(2n+1) - \frac{1}{4}(n+1)^2 = \boxed{\frac{n^2-1}{12}}$$

Przykład: Rozkład cyfr w liczbie π jest zgodny z modelem rozkładu płaskiego.



Przykład: Pomiar za pomocą miernika cyfrowego. Jako błąd zaokrąglenia warto przyjąć dyspersję rozkładu płaskiego, a nie błąd wynikający z klasy przyrządu.

Rozkład płaski - zmienna ciągła

W **rozkładzie płaskim** (jednostajnym, stałym, jednorodnym) każda wartość z przedziału $[a,b]$ może być wylosowana z jednakowym prawdopodobieństwem:

$$\boxed{f(x) = \frac{1}{b-a} \quad \text{dla} \quad a \le x \le b}$$

Wartość oczekiwana: $\mathcal{E}[\mathrm{x}] = \int_a^b \frac{x}{b-a}\,\mathrm{d}\,x = \frac{1}{b-a}\left[\frac{1}{2}x^2\right]\Bigg|_a^b = \frac{1}{2}\frac{b^2-a^2}{b-a} = \boxed{\frac{1}{2}(a+b)}$

Wariancja:

$$\mathcal{V}[\mathrm{x}] = \mathcal{E}[\mathrm{x}^2] - (\mathcal{E}[\mathrm{x}])^2 = \int_a^b \frac{x^2}{b-a}\,\mathrm{d}\,x - (\mathcal{E}[\mathrm{x}])^2 = \frac{1}{b-a}\left(\frac{1}{3}x^3\right)\Bigg|_a^b - \frac{1}{4}(a+b)^2 = \boxed{\frac{(b-a)^2}{12}}$$

<u>Przykład:</u> Niech zmienna losowa x ma rozkład jednostajny na przedziale [0,1]. Wówczas zmienna losowa t określona równaniem

$$\mathrm{x} = \lambda \int_0^t \exp(-\lambda t')\,dt' = 1 - \exp(-\lambda t)$$

czyli $\quad \mathrm{t} = -\frac{1}{\lambda}\ln(1-x) = -\tau \ln(1-x) \quad$ albo $\quad \mathrm{t} = -\tau \ln(x)$

ma rozkład wykładniczy $\quad \mathcal{E}(t;\lambda) = \lambda \exp(-\lambda t)$

Rozkład dwupunktowy (Bernoulliego)

Eksperyment Bernoulliego ma dwa możliwe wyniki „sukces” i „porażka”, które wzajemnie się wykluczają i występują losowo.

Próbą Bernoulliego nazywamy ustaloną sekwencję n powtórzeń eksperymentu Bernoulliego. W każdej próbie p-two sukcesu p i porażki $q = 1 - p$ są identyczne, kolejne próby są niezależne: wyniki dotychczasowych prób nie mają wpływu na wyniki kolejnych prób.

Rozkład dwupunktowy (Bernoulliego) ma postać:

$$\mathbf{P}_k(p) = p^k (1-p)^{1-k} \qquad \text{gdzie } k = 0,1$$

Wartość oczekiwana: $\mathcal{E}[\mathrm{k}] = \sum_{k=0}^{1} k \cdot \mathbf{P}_k = 0 \cdot (1-p) + 1 \cdot p = p$

Wariancja: $\mathcal{V}[\mathrm{k}] = \sum_{k=0}^{1} (\mathrm{k} - p)^2 \cdot \mathbf{P}_k = \sum_{k=0}^{1} \mathrm{k}^2 \cdot \mathbf{P}_k - p^2 = 0 \cdot (1-p) + 1 \cdot p - p^2 = p(1-p)$

Przykład: Urna zawiera 10 kul czerwonych i 10 kul czarnych.

– Losujemy kolejno 10 kul, zwracając wylosowaną kulę do urny przed kolejnym losowaniem. Ten schemat odpowiada definicji próby Bernoulliego gdzie p=q=0.5

– Losujemy kolejno 10 kul, nie zwracając wylosowanych kul do urny. To nie są próby Bernoulliego, ponieważ p-two sukcesu p zmienia się podczas każdej próby.

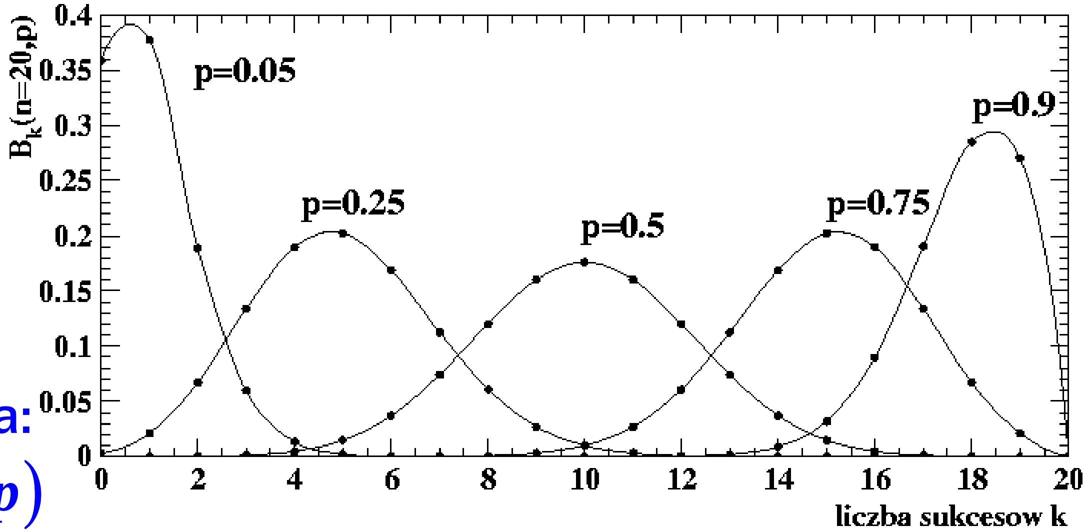
Rozkład dwumianowy

Rozkład dwumianowy to rozkład p-twa k sukcesów w n próbach Bernoulliego, czyli rozkład p-twa sumy n zmiennych losowych $k = k_1 + k_2 + \ldots + k_n$ każda z rozkładu dwupunktowego.

Rozkład p-twa dany jest przez:

$$\mathcal{B}_k(n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

gdzie $k = 0, 1, \ldots, n$



Wartość oczekiwana i wariancja:

$$\mathcal{E}[\mathrm{k}] = np \qquad \mathcal{V}[\mathrm{k}] = np(1-p)$$

Przykład: Pewna linia lotnicza stwierdziwszy, że 4% pasażerów którzy wykupią bilety, nie pojawia się na lotnisku, przyjęła politykę sprzedaży stu biletów na samolot, który ma tylko 98 miejsc. Jaka jest szansa na to, że wszyscy którzy przyjdą na lotnisko znajdą się w samolocie? $p = 0.96 \qquad n = 100$

$$\mathrm{P}(\mathrm{k} \le 98) = 1 - \mathcal{B}_{100}(n,p) - \mathcal{B}_{99}(n,p) = 1 - \binom{100}{100} \cdot p^{100} + \binom{100}{99} \cdot p^{99} \cdot (1-p) \cong 0.913$$

Rozkład wielomianowy

Eksperyment wielomianowy (uogólnienie eksperymentu dwumianowego) polega na:

1. wykonaniu n identycznych prób, przy czym w każdej próbie mamy do czynienia z j wzajemnie rozłącznymi i jedynie możliwymi zdarzeniami A_1, A_2, ... , A_j
2. p-two zajścia zdarzenia A_i wynosi p_i i nie zmienia się w kolejnych próbach, przy czym $\sum_i p_i = 1$.
3. kolejne próby są niezależne, tzn. rezultaty dotychczasowych prób nie mają wpływu na wyniki przyszłych prób.

Niech zmienne losowe k_1, k_2,..., k_j zliczają liczby zajść poszczególnych zdarzeń w n próbach, przy czym $k_1 + k_2 + \ldots + k_j = n$. Funkcja p-twa rozkładu wielomianowego dana jest przez:

$$\mathcal{W}_{k_1 \ldots k_j}\left(n, p_1, \ldots, p_j\right) = \frac{n!}{k_1! k_2! \ldots k_j!} p_1^{k_1} p_2^{k_2} \ldots p_j^{k_j}$$

Wartości oczekiwane i wariancje: $\mathcal{E}\left[\mathbf{k}_i\right] = np_i$ $\quad \mathcal{V}\left[\mathbf{k}_i\right] = np_i\left(1 - p_i\right)$

Przykład: Urna zawiera 10 kul: cztery R, trzy G, trzy B. Losujemy 5 kul ze zwracaniem. Jaka jest szansa wylosowania $k_1 = 2$ kul R, $k_2 = 2$ kul G i $k_3 = 1$ kuli B?

$$P = \frac{5!}{2!2!1!}\left(\frac{4}{10}\right)^2\left(\frac{3}{10}\right)^2\left(\frac{3}{10}\right)^1 = 0.1296$$

Rozkład geometryczny

Rozkład geometryczny opisuje p-two pierwszego sukcesu w k-tej próbie Bernoulliego:

$$\mathcal{G}_k(p) = p(1-p)^{k-1}$$

Wartość oczekiwana:

$$\mathcal{E}[\mathrm{k}] = \sum_{k=1}^{\infty} k \cdot \mathcal{G}_k(p) = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = (1-q)\sum_{k=1}^{\infty}\frac{d}{dq}(q^k) = (1-q)\frac{d}{dq}\left(\sum_{k=1}^{\infty} q^k\right) =$$

$$= (1-q)\frac{d}{dq}\left(q\sum_{m=0}^{\infty} q^m\right) = (1-q)\frac{d}{dq}\left(\frac{q}{1-q}\right) = \frac{1}{1-q} = \frac{1}{p}$$

Wariancja:

$$\langle \mathrm{k}^2 \rangle = \sum_{k=1}^{\infty} k^2 \cdot \mathcal{G}_k(p) = p\sum_{k=1}^{\infty}(k-1+1)^2 q^{k-1} = \|m = k-1\| = p\sum_{m=0}^{\infty}(m+1)^2 q^m =$$

$$= p\sum_{m=0}^{\infty} m^2 q^m + 2p\sum_{m=0}^{\infty} mq^m + p\sum_{m=0}^{\infty}(1-p)^m = pq\sum_{m=1}^{\infty} m^2 q^{m-1} + 2pq\sum_{m=1}^{\infty} mq^{m-1} + \frac{p}{1-q} =$$

$$= q\langle \mathrm{k}^2 \rangle + 2\frac{q}{p} + 1 \quad \Rightarrow \quad \langle \mathrm{k}^2 \rangle = 2\frac{q}{p^2} + \frac{1}{p} \qquad \mathcal{V}[\mathrm{k}] = \langle \mathrm{k}^2 \rangle - \langle \mathrm{k} \rangle^2 = \frac{1-p}{p^2}$$

Przykład: Kontrola jakości. Niech p-two wadliwego wyrobu będzie równe $p=2{\cdot}10^{-6}$. Jaka jest szansa, że na pierwszy wadliwy wyrób natkniemy się wśród pierwszych 10000 skontrolowanych wyrobów?

$$P = \sum_{k=1}^{10000} p(1-p)^{k-1} = p\sum_{m=0}^{9999}(1-p)^m = p\frac{(1-p)^{10000}-1}{(1-p)-1} = 1-(1-p)^{10000} \cong 0.02$$

Rozkład ujemny dwumianowy

Rozkład ujemny dwumianowy (Pascala) opisuje p-two, że do uzyskania n sukcesów musimy wykonać $k \geq n$ prób według schematu Bernoulliego, przy czym ostatnia próba zakończona jest sukcesem:

$$\mathcal{U}_{k}(n,p) = \binom{k-1}{n-1} p^{n}(1-p)^{k-n}, \qquad n = 1,2,\ldots \qquad k = n, n+1, n+2, \ldots$$

Rozkład ujemny dwumianowy opisuje rozkład sumy niezależnych zmiennych losowych, każda z rozkładu geometrycznego $k = k_1 + k_2 + \ldots + k_n$:

$$\mathrm{P}_{k_1,k_2,\ldots,k_n}(p) = p(1-p)^{k_1-1} \cdot p(1-p)^{k_2-1} \cdot \ldots \cdot p(1-p)^{k_n-1} = p^{n}(1-p)^{k_1+k_2+\ldots+k_n-n}$$

Aby uzyskać rozkład sumy $k = k_1 + k_2 + \ldots + k_n$ musimy zsumować łączny rozkład p-twa po takich wartościach $(k_1, k_2, \ldots, k_n)$ dla których suma jest stała i równa k:

$$\mathcal{U}_{k}(n,p) = \sum_{k_1+k_2+\ldots+k_n=k} p^{n}(1-p)^{k_1+k_2+\ldots+k_n-n} = \binom{k-1}{n-1} p^{n}(1-p)^{k-n}$$

Wartość oczekiwana i wariancja rozkładu ujemnego dwumianowego jako sumy n niezależnych zmiennych każda z rozkładu geometrycznego z tym samym parametrem p dane są przez:

$$\mathcal{E}[\mathrm{k}] = \frac{n}{p} \qquad \mathcal{V}[\mathrm{k}] = \frac{n(1-p)}{p^{2}}$$

Rozkład hipergeometryczny

Eksperyment hipergeometryczny polega na wylosowaniu bez zwracania n obiektów z populacji złożonej z N obiektów wśród których jest K obiektów posiadających pewną cechę i $N - K$ nie posiadających tej cechy.

Funkcja p-twa rozkładu hipergeometrycznego dana jest przez:

$$\mathcal{H}_k(N,K,n) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \qquad \begin{array}{l} N = 0,1,2,\ldots \\ n = 0,1,2,\ldots,N \\ K = 0,1,2,\ldots,N \\ 0 \le k \le \min(n,K) \end{array}$$

Wartość oczekiwana i wariancja:

$$\mathcal{E}[k] = np \qquad \mathcal{V}[k] = npq\frac{N-n}{N-1} \qquad \text{gdzie} \qquad p = \frac{K}{N}, \qquad q = 1-p$$

Gdy $N \to \infty,\ K \to \infty$ ale tak, że $\frac{K}{N} \to p,\ 0 < p < 1$ wtedy: $\mathcal{H}_k(N,K,n) \to \mathcal{B}_k(n,p)$

Przykład: Z partii 40 sztuk bezpieczników sprawdza się 4 losowo wybrane. Jeśli choć jeden jest wadliwy, to całą partię się odrzuca. Jakie jest p-two przyjęcia partii zawierającej 10% wadliwych bezpieczników?

$$P_{acc} = \mathcal{H}_0(N = 40, K = 4, n = 4) = \frac{\binom{4}{0}\binom{36}{4}}{\binom{40}{4}} = 0.6445$$

Rozkład wykładniczy

Rozważmy czas oczekiwania na przejazd samochodu. Brak samochodu w ciągu n kolejnych przedziałów czasowych Δt=t/n dany jest rozkładem dwumianowym:

$$Q_n = B_0(n,p) = (1-p)^n$$

Niech $p = \lambda\Delta t = \frac{\lambda t}{n}$ gdzie λ (tzw. intensywność) określa typową liczbę zdarzeń na jednostkę czasu:

$$p = \frac{\langle \mathrm{k} \rangle}{n} = \frac{\lambda t}{n} \quad \Rightarrow \quad \lambda = \frac{\langle \mathrm{k} \rangle}{t}$$

Przechodząc z n do nieskończoności $Q_n = \left(1 - \frac{\lambda t}{n}\right)^n \xrightarrow[n\to\infty]{} Q(t;\lambda) = \exp(-\lambda t)$

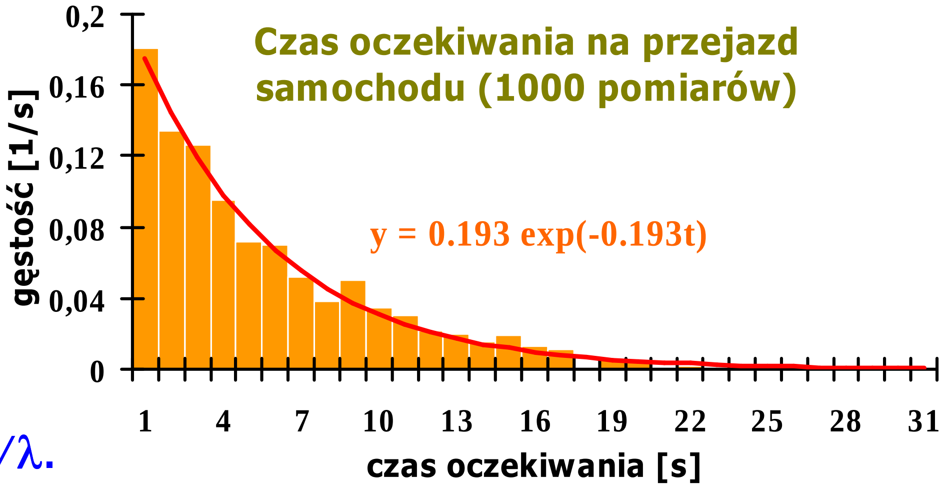
P-two sukcesu w przedziale $(0,\mathrm{t})$:

$$\mathrm{P}(0 \le t \le t;\lambda) \equiv \mathrm{F}(t;\lambda) = 1 - Q(t;\lambda) =$$
$$= 1 - \exp(-\lambda t)$$

Funkcja gęstości p-twa rozkładu wykładniczego ma postać:

$$\mathcal{E}(t;\lambda) = \frac{d\,\mathrm{F}(t;\lambda)}{dt} = \lambda \exp(-\lambda t)$$

gdzie t $\geqslant$ 0 oraz λ > 0. Czas życia to τ = 1/λ.



Rozkład wykładniczy

Wartość oczekiwana i wariancja:

$$\mathcal{E}[\mathrm{t}] = \lambda \int_0^{\infty} t \exp(-\lambda t)\, dt = \frac{1}{\lambda} = \tau \qquad \mathcal{V}[\mathrm{t}] = \lambda \int_0^{\infty} \left(t - \frac{1}{\lambda}\right)^2 \exp(-\lambda t)\, dt = \frac{1}{\lambda^2} = \tau^2$$

<u>Przykład:</u> („Brak pamięci") P-two braku samochodu w ciągu czasu T:

$$Q(T) = 1 - \mathrm{F}(T) = \exp(-\lambda T)$$

P-two, że samochód nie przejedzie jeszcze dodatkowo w czasie t jeśli nie przejechał w poprzedzającym czasie T to p-two warunkowe:

$$Q(t\,|\,T) = \mathrm{P}(\mathrm{t} > t + T\,|\,\mathrm{t} > T) = \frac{\mathrm{P}((\mathrm{t} > t + T) \cap (\mathrm{t} > T))}{\mathrm{P}(\mathrm{t} > T)} = \frac{\mathrm{P}(\mathrm{t} > t + T)}{\mathrm{P}(\mathrm{t} > T)} = \frac{Q(T + t)}{Q(T)} =$$

$$= \frac{\exp(-\lambda(t + T))}{\exp(-\lambda T)} = \exp(-\lambda t) = Q(t)$$

A więc $Q(t+T) = Q(t)Q(T)$ czyli pozostały okres życia nie zależy od przeszłości.

<u>Uwagi:</u>

- „brak pamięci" jest cechą wyłącznie funkcji wykładniczej (oraz y=0 i y=1)
- prosty matematycznie, ale nie intuicyjny (starzenie się człowieka, urządzeń)
- stosowany do opisu czasu życia urządzeń, których czas życia jest długi

Rozkład wykładniczy

Przykład: W pewnym szpitalu, w niedziele wieczorem pomiędzy 18:00 i 22:00 zgłasza się na izbę przyjęć średnio 5 pacjentów w ciągu godziny. Zakładając, że liczba zgłoszeń podlega rozkładowi Poissona, a czas pomiędzy kolejnymi zgłoszeniami rozkładowi wykładniczemu, znajdź:

- średni czas pomiędzy kolejnymi przyjęciami: $\tau = \frac{1}{\lambda} = \frac{1}{5} = 0.2 \text{ h}$
- p-two, że kolejne przyjęcie nastąpi w ciągu 15 minut od poprzedniego:

$$P(x \le 0.25) = \int_0^{0.25} \frac{1}{\tau}\exp\left(-\frac{t}{\tau}\right)dt = -\exp\left(-\frac{t}{0.2}\right)\Big|_0^{0.25} = 1 - \exp\left(-\frac{0.25}{0.2}\right) = 0.7135$$

- p-two, że czas pomiędzy kolejnymi przyjęciami będzie dłuższy niż 10 minut:

$$P(x > 0.1667) = \int_{0.1667}^{\infty} \frac{1}{\tau}\exp\left(-\frac{t}{\tau}\right)dt = -\exp\left(-\frac{t}{0.2}\right)\Big|_{0.1667}^{\infty} = \exp\left(-\frac{0.1667}{0.2}\right) = 0.4345$$

- p-two, że w dowolnej chwili czas do następnego przyjęcia będzie pomiędzy 15 i 25 minut:

$$P(0.25 < x < 0.4167) = \int_{0.25}^{0.4167} \frac{1}{\tau}\exp\left(-\frac{t}{\tau}\right)dt = -\exp\left(-\frac{t}{0.2}\right)\Big|_{0.25}^{0.4167} =$$

$$= -\exp\left(-\frac{0.4167}{0.2}\right) + \exp\left(-\frac{0.25}{0.2}\right) = 0.1620$$