Komputerowa analiza danych doświadczalnych

Wykład 5
24.03.2017

dr inż. Łukasz Graczykowski
lgraczyk@if.pw.edu.pl

Semestr letni 2016/2017

Metody Monte Carlo

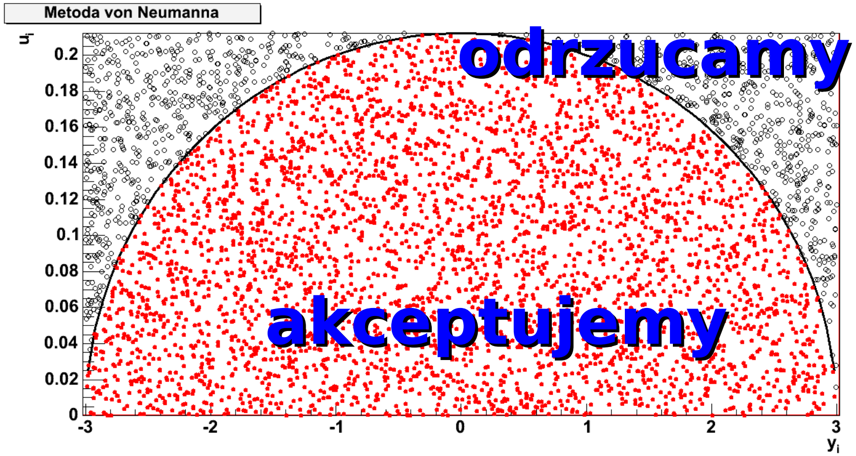
Najważniejsze rozkłady prawdopodobieństwa

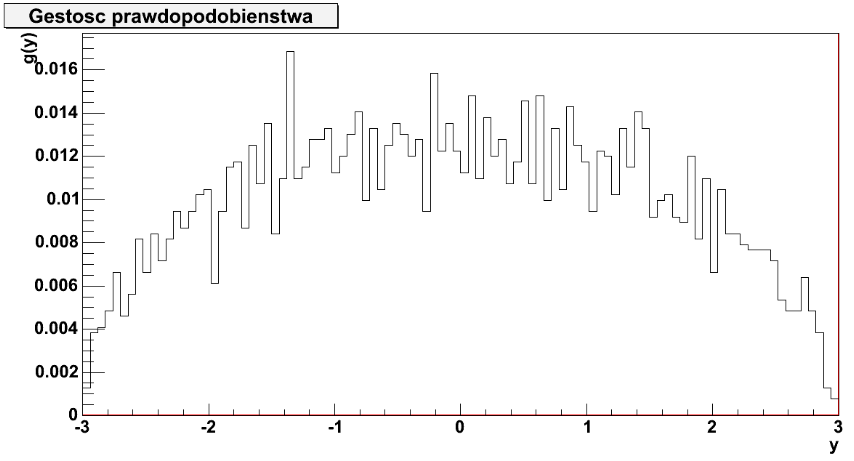
Metoda akceptacji-odrzuceń von Neumanna

Metoda (akceptacji) von Neumanna

- Jak to działa?
  - generujemy parę liczb z rozkładu jednorodnego: $(y_i, u_i) \quad a \leqslant y_i \leqslant b,\ 0 \leqslant u_i \leqslant d$
  - rozważamy krzywą: $u = g(y)$ oraz funkcję stałą: $u = d, d \geqslant g_{max}$
  - sprawdzamy, czy $u_i < g(y_i)$
  - jeśli warunek jest spełniony, <u>akceptujemy</u> liczbę y_i,
    jeśli nie - odrzucamy
  - zaakceptowane wartości y_i podlegają rozkładowi $g(y)$
  - rozkład $g(y)$ nie musi być unorm.
  - wydajność metody:

$$E = \frac{\int_a^b g(y)\,dy}{(b-a)\,d} \approx \frac{N_{accept}}{N_{all}}$$

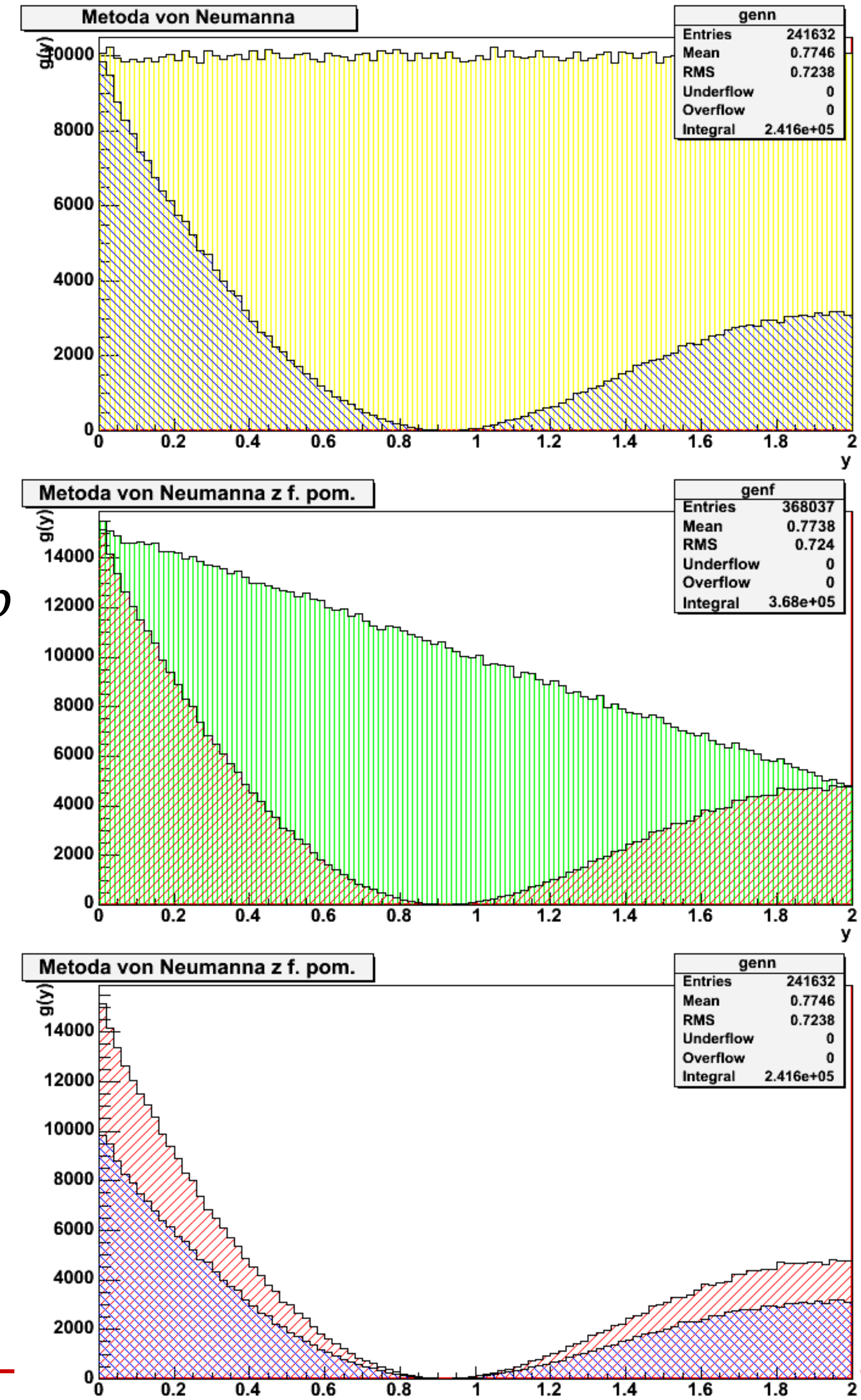




Metoda von Neumanna z funkcją pomocniczą

- Wydajność metody von Neumanna można poprawić, jeśli odpowiednio <u>zawężimy obszar losowania</u>:
  - wprowadzamy funkcję pomocniczą $s(y)$, z której "łatwo" wygenerować zmienne losowe (np. metodą odwrotnej dystrybuanty), i która spełnia warunek: $g(y) \leqslant c \cdot s(y), \quad a < y < b$
  - generujemy liczbę losową y_i z rozkładu $s(y)$ na przedziale $a < y_i < b$ oraz liczbę u_i z rozkładu jednorodnego na przedziale $0 < u_i < 1$
  - <u>odrzucamy</u> liczbę y_i, jeżeli: $u_i \geqslant \frac{g(y_i)}{c \cdot s(y_i)}$
  - wydajność metody:

$$E = \frac{\int_a^b g(y)\,dy}{c\int_a^b s(y)\,dy}$$

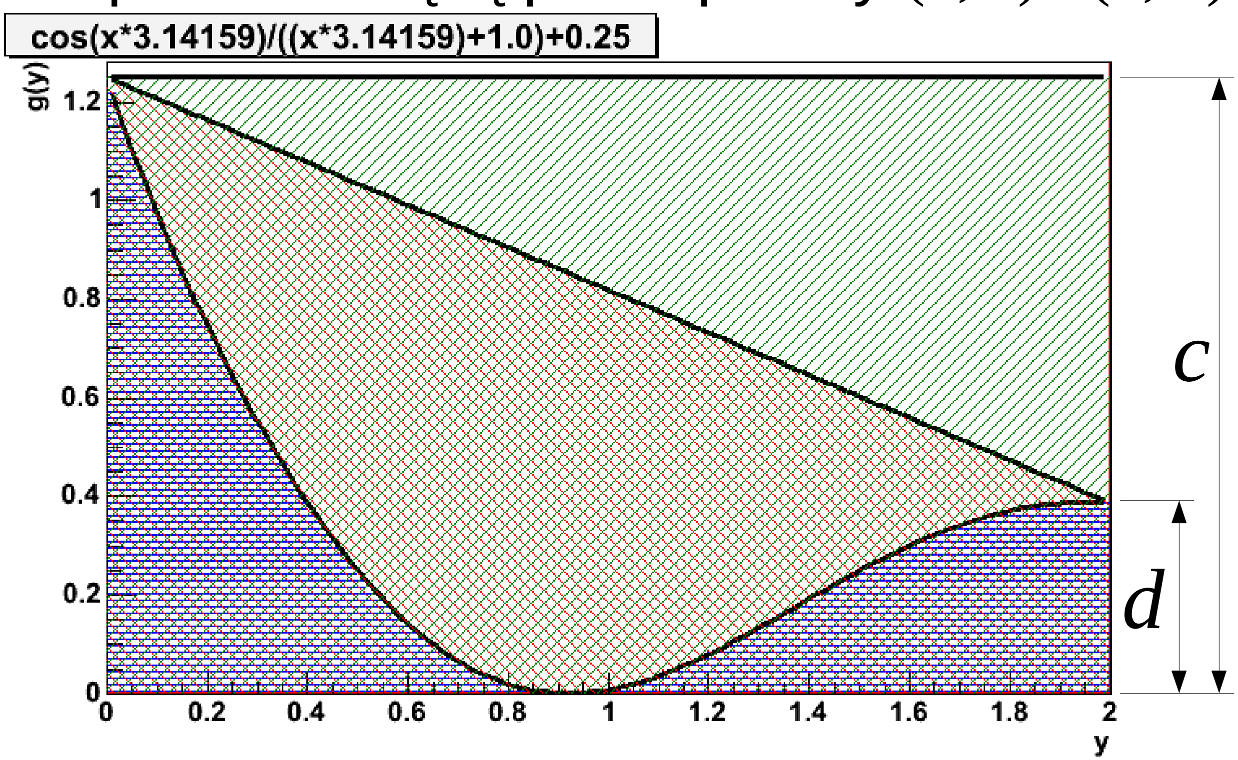


Metoda von Neumanna z fun. pom. – przykład

- Rozważmy funkcję gęstości postaci:

$$g(y)=\cos(\pi x)/(\pi x+1)+1/4,\ \ 0\leqslant y\leqslant 2$$

- Funkcja ta, w przedziale od 0 do 2, ma dwa maksima: $g(0)=c\,,\ \ g(2)=d$
- W zwykłej metodzie von Neumanna wybieramy prostą: $u_{max}=c$
- Tutaj możemy łatwo wybrać funkcję pomocniczę $s(y)$ jako prostą przechodzącą przez pubkty $(0,\,c)$ i $(2,\,d)$



- Aby otrzymać wzór $s(y)$ rozważamy układ równań:

$$c=a\cdot 0+b$$
$$d=a\cdot 2+b$$

- Z czego wzór na $s(y)$:

$$s(y)=\frac{d-c}{2}\,y+c$$

- **Jak otrzymać wartość losową z tego rozkładu?**

Metoda von Neumanna z fun. pom. – przykład

- **Metodą odwrotnej dystrybuanty!**
- Liczymy dystrybuantę:

$$S(y)=\frac{d-c}{4}y^2+cy$$

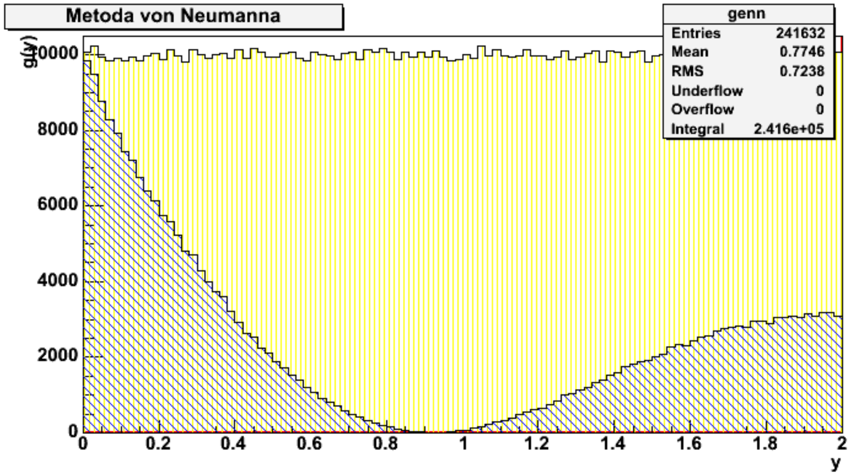
- Oraz jej funkcję odwrotną:

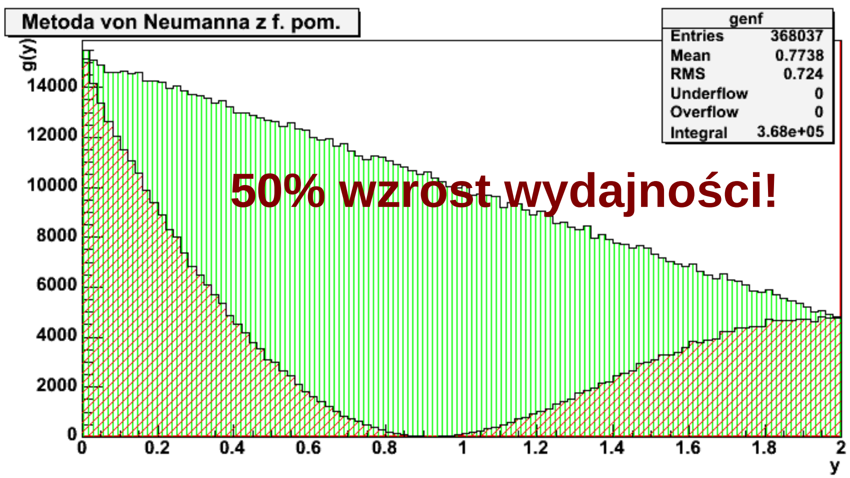
$$y=S^{-1}(x)=2\frac{c^2-\sqrt{xc(d-c)+(d-c)^2}}{c(c-d)}$$

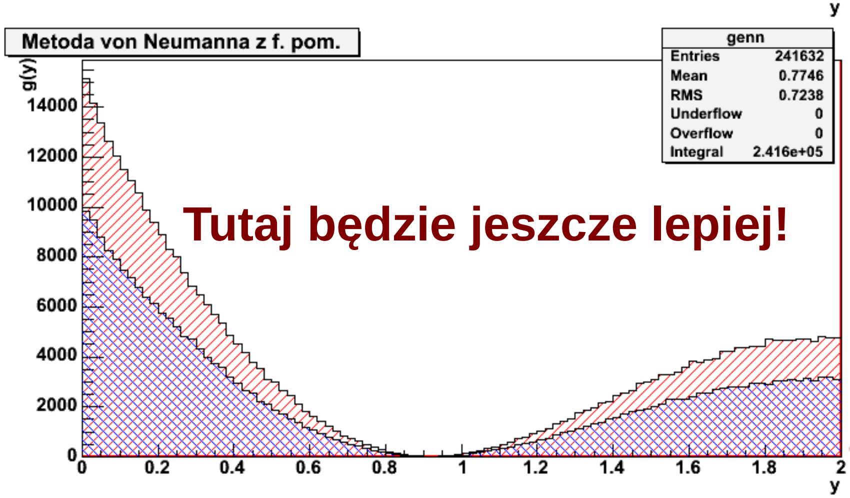
- Losujemy wartość x_i z rozkładu jednorodnego w granicach:

$$S(0)=0,\ \ S(2)=d+c$$

- I wstawiamy ją do wzoru na odwrotną dystrybuantę by otrzymać y_i z rozkł. $s(y)$
- Losujemy pomocniczą wartość u_i z rozkładu jednorodnego $0<u_i<1$
- Sprawdzamy warunek akceptacji y_i: $u_i<\frac{g(y_i)}{s(y_i)}$







Generacja liczb o rozkładzie normalnym

- Jak pamiętamy, **rozkład normalny nie ma analitycznej formy dystrybuanty**
- Do generowania liczb z rozkładu normalnego o $\hat{x}=0,\ \sigma=1$ (standardowego) służy **metoda Box'a-Muller'a**

$$f(z)=\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-z^2}{2}\right)$$

$$z=\frac{x-\hat{x}}{\sigma}$$

transformacja dowolnego rozkł. norm. do standardowego

- Generujemy parę liczb (u_1,u_2) z rozkładów jednorodnych (0,1) i dokonujemy zamiany zmiennych:

$$v_1=2u_1-1 \qquad v_2=2u_2-1$$

- Obliczamy: $s=v_1^2+v_2^2$
- Gdy $s\geqslant 1$ <u>odrzucamy parę</u>
- Otrzymujemy <u>dwie</u> liczby pseudolosowe opisane rozkładem normalnym standardowym:

$$x_1=v_1\sqrt{-(2/s)\ln s} \qquad x_2=v_2\sqrt{-(2/s)\ln s}$$

Całkowanie metodą Monte Carlo

- Jak już zauważylismy, pole powierzchni pod rozpatrywaną krzywą w stosunku do pola prostokąta, z którego losujemy dwie liczby pseudolosowe, ma się (w przybliżeniu) do siebie tak jak liczba par zaakceptowanych do odrzuconych:
$$\frac{\int_a^b g(y)\,dy}{(b-a)\,d} \approx \frac{N_{accept}}{N_{all}}$$
- Co pozwala na <u>przybliżone</u> obliczenie wartości <u>całki oznaczonej</u>:

$$\int_a^b g(y)\,dy \approx \frac{N_{accept}}{N_{all}}(b-a)\,d$$

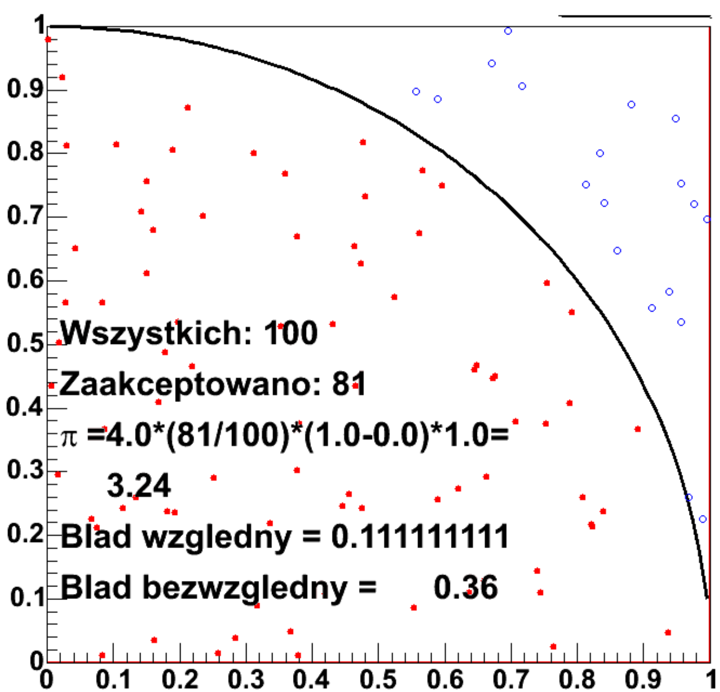
- W ten sposób można obliczyć **dowolną** całkę oznaczoną poprzez prostą generację dwóch liczb z rozkładu jednorodnego. W wersji n-wymiarowej oczywiście możemy to zrobić dla dowolnej liczby zmiennych losowych (i obliczać całki wielowymiarowe)
- Względna dokładność obliczenia całki: $\frac{\Delta I}{I} = \frac{1}{\sqrt{N_{wszystkie}}}$

Całkowanie metodą Monte Carlo – przykład

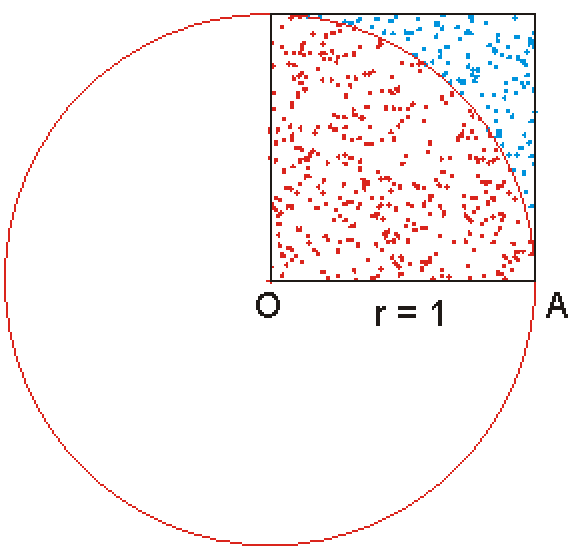
- Najpopularniejszy przypadek to wykorzystanie metody Monte Carlo do obliczenia wartości liczby π
- W tym celu rozpatrzmy ćwiartkę okręgu o jednostkowym promieniu. Funkcja opisująca tę ćwiartkę to:

$$g(y)=\sqrt{(R^2-y^2)};\;\; 0\leqslant y\leqslant 1;0\leqslant y\leqslant 1$$

- Pole ćwiartki jednostkowego okręgu to: $I=\int_0^1 g(y)dy=\pi/4 \Rightarrow \boxed{\pi=4\cdot I}$
- Wartość całki obliczamy metodą Monte Carlo: $I\approx\frac{N_{accept}}{N_{all}}(b-a)d$



wszystko przypomina rzucanie lotkami (darts)

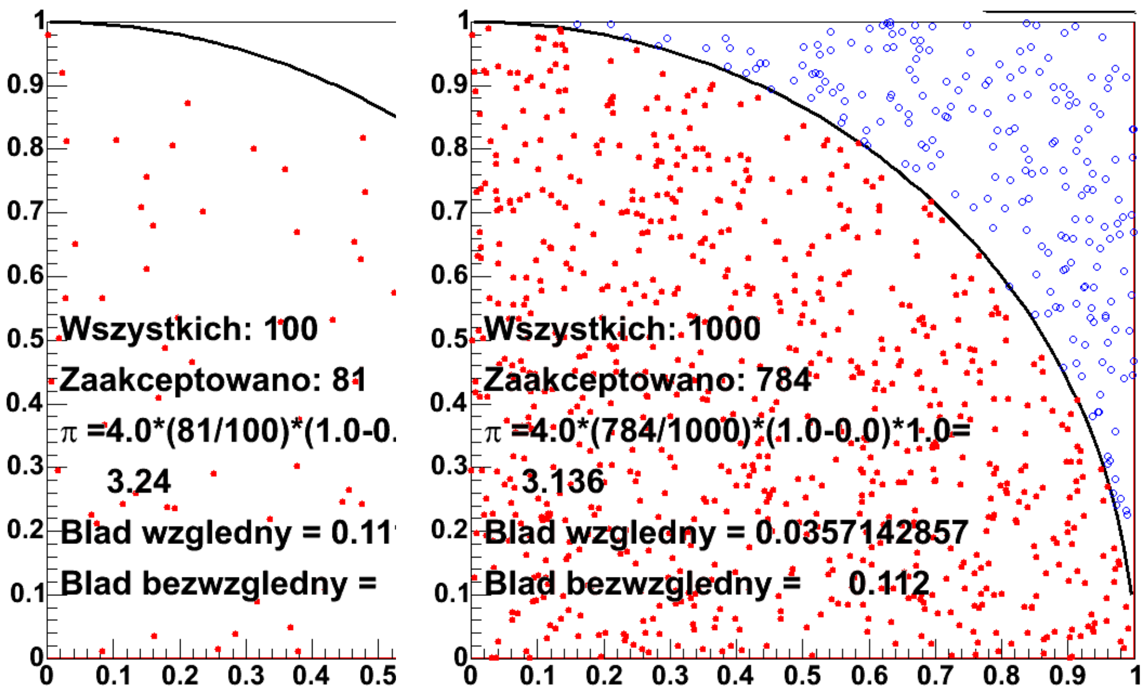


Całkowanie metodą Monte Carlo – przykład

- Najpopularniejszy przypadek to wykorzystanie metody Monte Carlo do obliczenia wartości π
- W tym celu rozpatrzmy ćwiartkę okręgu o jednostkowym promieniu. Funkcja opisująca tę ćwiartkę to:

$$g(y)=\sqrt{(R^2-y^2)};\quad 0\leqslant y\leqslant 1;0\leqslant y\leqslant 1$$

- Pole ćwiartki jednostkowego okręgu to: $I=\int_0^1 g(y)\,dy=\pi/4\Rightarrow\pi=4\cdot I$
- Wartość całki obliczamy metodą Monte Carlo: $I\approx\frac{N_{accept}}{N_{all}}(b-a)d$

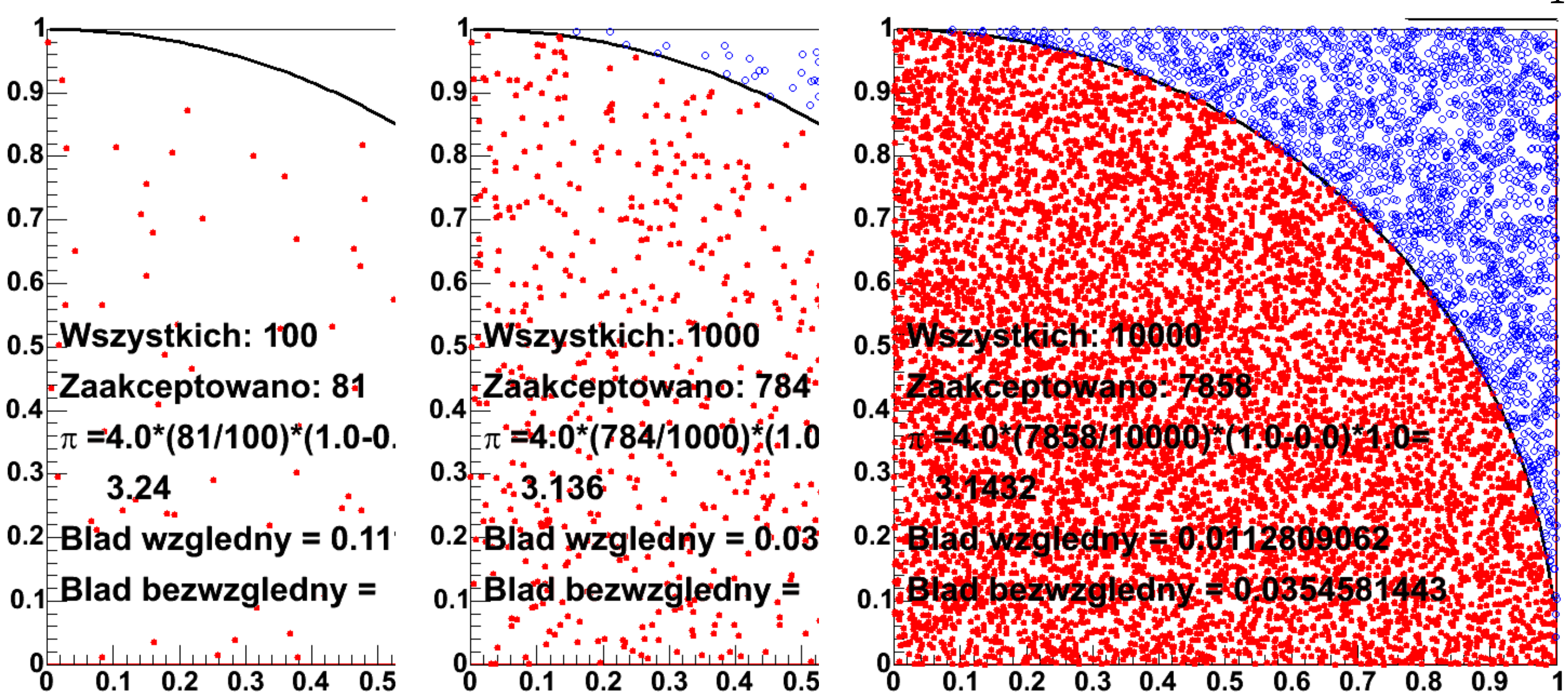


Całkowanie metodą Monte Carlo – przykład

- Najpopularniejszy przypadek to wykorzystanie metody Monte Carlo do obliczenia wartości π
- W tym celu rozpatrzmy ćwiartkę okręgu o jednostkowym promieniu. Funkcja opisująca tę ćwiartkę to:

$$g(y)=\sqrt{(R^2-y^2)};\;\; 0\leqslant y\leqslant 1;0\leqslant y\leqslant 1$$

- Pole ćwiartki jednostkowego okręgu to: $I=\int_0^1 g(y)\,dy=\pi/4 \Rightarrow \pi=4\cdot I$
- Wartość całki obliczamy metodą Monte Carlo: $I\approx\frac{N_{accept}}{N_{all}}(b-a)d$

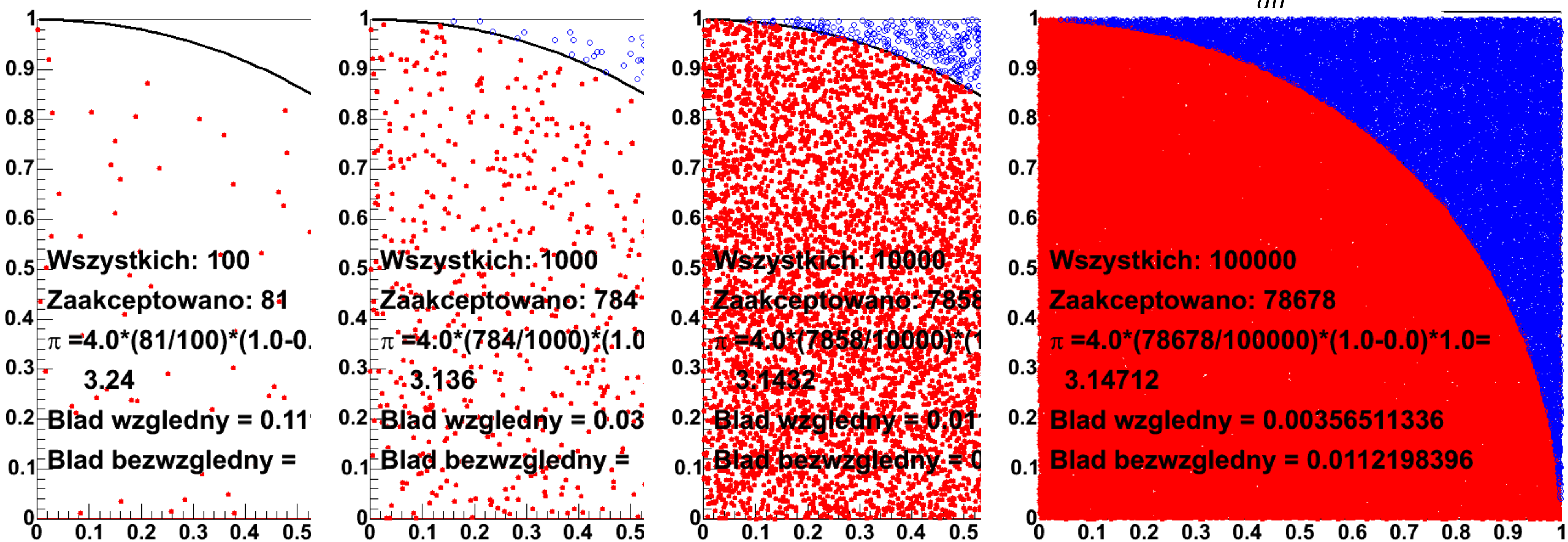


Całkowanie metodą Monte Carlo – przykład

- Najpopularniejszy przypadek to wykorzystanie metody Monte Carlo do obliczenia wartości π
- W tym celu rozpatrzmy ćwiartkę okręgu o jednostkowym promieniu. Funkcja opisująca tę ćwiartkę to:

$$g(y)=\sqrt{(R^2-y^2)};\ \ 0\leqslant y\leqslant 1;0\leqslant y\leqslant 1$$

- Pole ćwiartki jednostkowego okręgu to: $I=\int_0^1 g(y)dy=\pi/4 \Rightarrow \pi=4\cdot I$
- Wartość całki obliczamy metodą Monte Carlo: $I\approx\frac{N_{accept}}{N_{all}}(b-a)d$

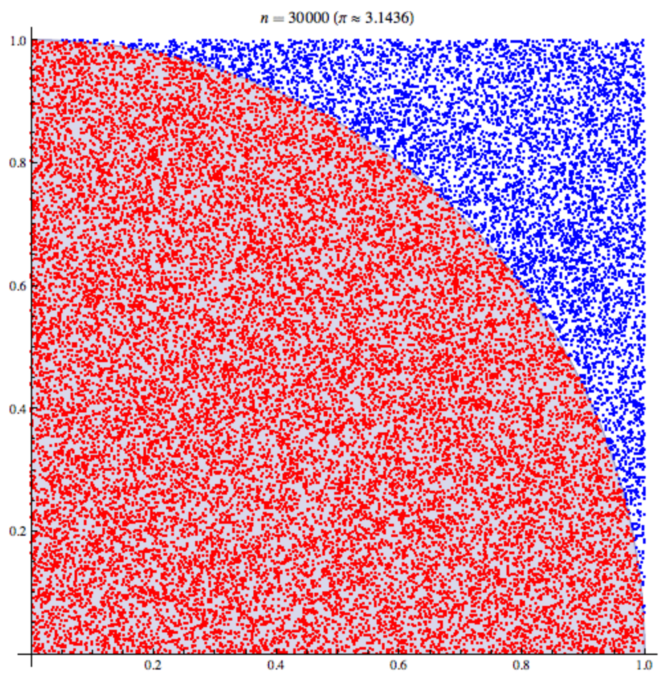


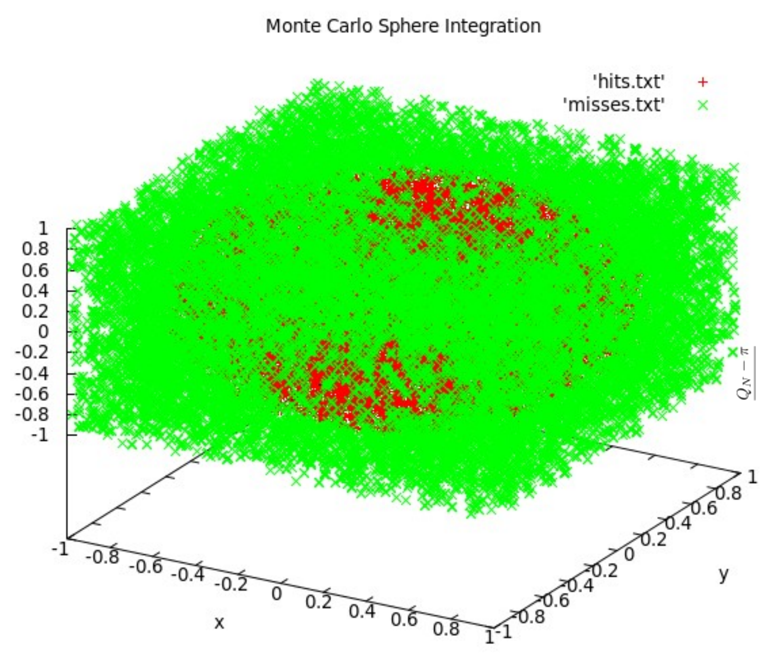
Całkowanie metodą Monte Carlo – przykład

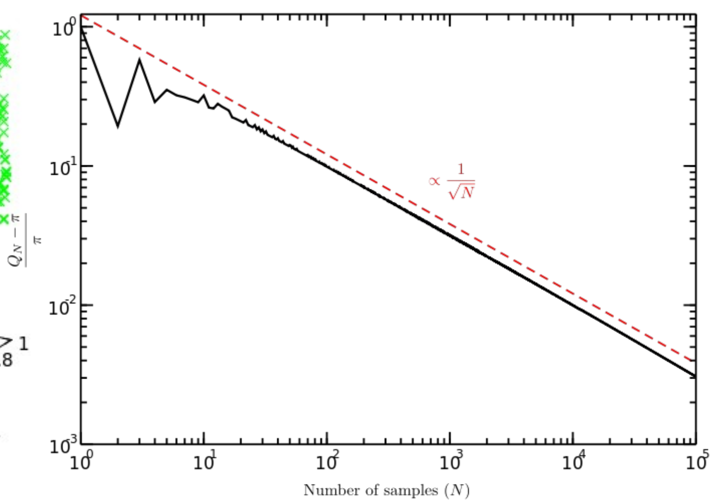
- Najpopularniejszy przypadek to wykorzystanie metody Monte Carlo do obliczenia wartości π
- W tym celu rozpatrzmy ćwiartkę okręgu o jednostkowym promieniu. Funkcja opisująca tę ćwiartkę to:

$$g(y)=\sqrt{(R^2-y^2)};\;\; 0\leqslant y\leqslant 1;\, 0\leqslant y\leqslant 1$$

- Pole ćwiartki jednostkowego okręgu to: $I=\int_0^1 g(y)\,dy=\pi/4\Rightarrow\pi=4\cdot I$
- Wartość całki obliczamy metodą Monte Carlo: $I\approx\frac{N_{accept}}{N_{all}}(b-a)d$







Najważniejsze rozkłady prawdopodobieństwa

Rozkład dwumianowy

- W <u>Polsce</u> znany również jako rozkład Bernoulliego (*ang. binomial distribution*) – w innych krajach może oznaczać inny rozkład
- Rozważmy proste doświadczenie – rzut monetą:
  - w wyniku rzutu możemy otrzymać dwa wykluczające się wyniki
  - zatem przestrzeń zdarzeń elementarnych: $E = A + \bar{A}$
  - możemy zdefiniować prawdopodobieństwa:

$P(A) = p$

$P(\bar{A}) = 1 - p = q$

- Wynik doświadczenia może być zmienną losową X_i, która przybiera wartość 1 lub 0 w zależności od tego, czy zaszło zdarzenie A lub $\bar{A}$
- Jeśli powtórzymy wielokrotnie doświadczenie, to otrzymamy rozkład zmiennej losowej $X = X_1 + X_2 + \ldots X_n$

Rozkład dwumianowy

- Z rachunku prawdopodobieństwa wiemy, że jeżeli przestrzeń zdarzeń elementarnych $E=A_1+A_2+\ldots+A_n$ i zdarzenia są niezależne, to:

$$P(A_1 A_2 \ldots A_n)=P(A_1)P(A_2)\ldots P(A_n)$$

- Z tego wynika, że prawdopodobieństwo, że k pierwszych doświadczeń (z n) da wynik zdarzenia A a pozostałe n-k dadzą wynik zdarzenia $\bar{A}$, wynosi:

$$P(A^k \bar{A}^{n-k})=P(A^k)P(\bar{A}^{n-k})=p^k q^{n-k}$$

- Zgodnie z kombinatoryką, pojawienie się k razy zdarzenia A w n doświadczeniach realizuje się na "n po k" sposobów: różniących się kolejnością zdarzeń A i $\bar{A}$

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

- Prawdopodobieństwo wystąpienia k razy zdarzenia A i n-k razy zdarzenia $\bar{A}$ w n doświadczeniach, w dowolnej kolejności, wynosi:

$$P(k)=W_k^n=\binom{n}{k}p^k q^{n-k}; \quad q=1-p$$

- Tak zdefiniowany rozkład nazywamy **rozkładem dwumianowym**

Rozkład dwumianowy

- Policzmy wartość oczekiwaną i wariancję rozkładu dwumianowego
- Dla pojedynczego doświadczenia X_i (zmiennej losowej, która może przyjąć wartość 1 lub 0):

$$E(X)=\sum_{i=1}^{n} x_i P(X=x_i)$$

$$E(X_i)=1\cdot P(X_i=1)+0\cdot P(X_i=0) \quad \boxed{E(X_i)=1\cdot p+0\cdot q=p}$$

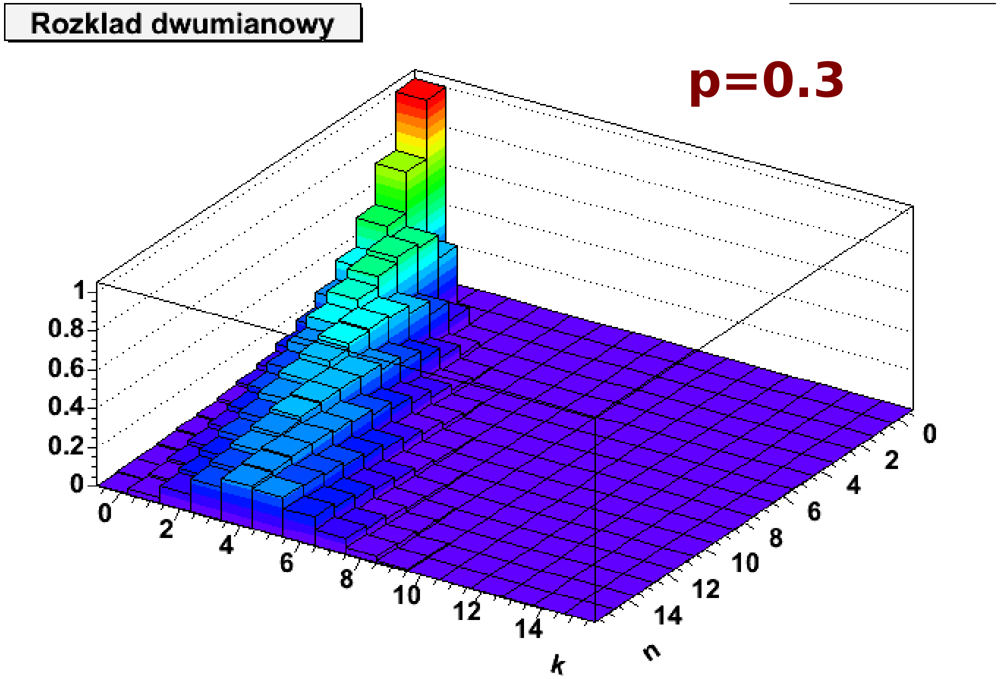
$$\sigma^2(X_i)=E\left((x_i-p)^2\right)=(1-p)^2 p+(0-p)^2 q=\boxed{pq}$$

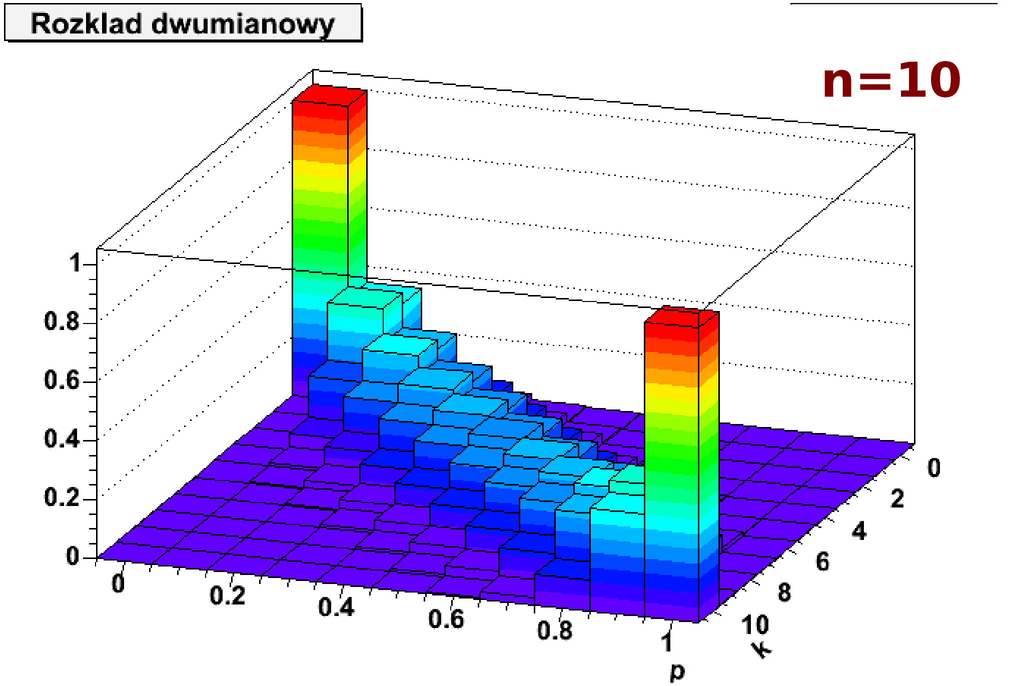
- Z własności warotśći oczekiwanej: $E(X=X_1+X_{2\ldots}+X_n)=\sum_{i=1}^{n} E(X_i)=np$
- Zakładając niezależność zmiennych (zerowe kowariancje) otrzymamy z kolei: $\sigma^2(X)=npq$
- Dla 2 zdarzeń losowych:

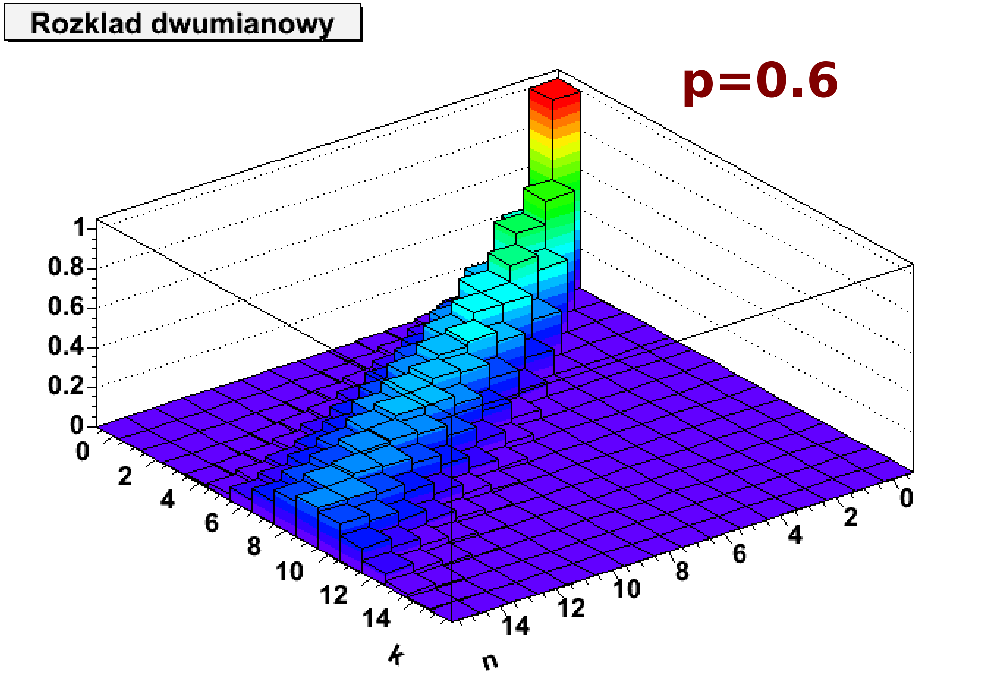
$$\sigma^2(X)=\binom{2}{2} p^2(2-2\mathrm{p})^2+\binom{2}{1} pq(1-2\mathrm{p})^2+\binom{2}{0} q^2(0-2\mathrm{p})^2=$$

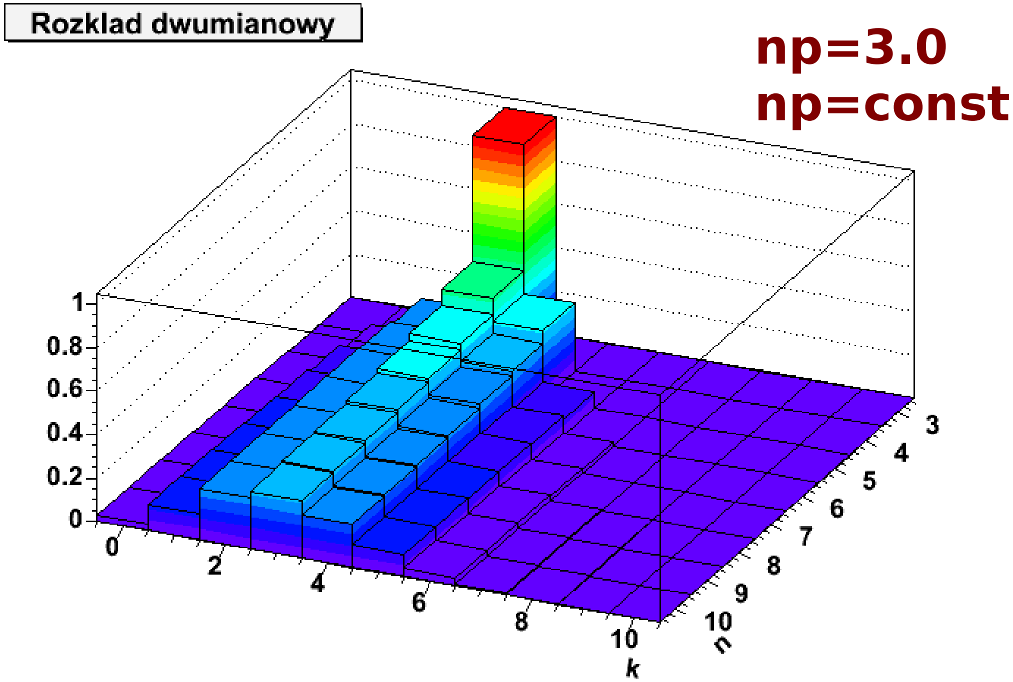
$$2p^2(4-8\mathrm{p}+4\mathrm{p}^2)+2(p-2\mathrm{p}^2)+(1-2\mathrm{p}+p^2)4\mathrm{p}^2=2\mathrm{p}(1-p)=2\mathrm{pq}$$

Rozkład dwumianowy – rysunek





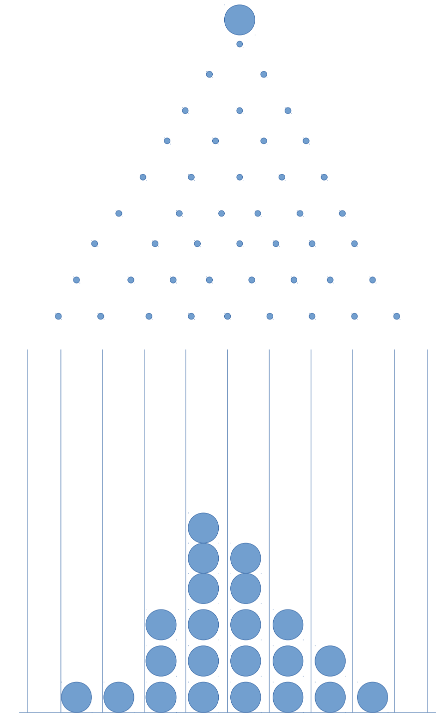




Rozkład dwumianowy - tablica Galtona

- Innym przykładem realizacji rozkładu dwumianowego jest **tablica (deska) Galtona**:
  - mamy n rzędów kołeczków
  - kuleczka może przesunąć się w lewo (z prawdopod. p=0,5) lub w prawo (q=0.5)
  - kuleczka przesunie się k razy w lewo i n-k razy w prawo
  - każde przesunięcie jest niezależne
  - zatem dla jednej konkretnej konfiguracji (drogi) “spadku” kulki prawdopodobieństwo: $p^k q^{n-k}$
  - jeśli mamy różne konfiguracje przesunięć:

$$P(k)=W_k^n=\binom{n}{k}p^k q^{n-k};\ \ q=1-p$$



deska Galtona na Wydziale Fizyki PW

http://www.if.pw.edu.pl/~pluta/pl/tgak.jpg

Rozkład dwumianowy - inne przykłady z życia

$$P(k)=W_k^n=\binom{n}{k}p^k q^{n-k}; \quad q=1-p$$

1) n – ilość studentów na 3 roku fizyki
p – prawdopodobieństwo zaliczenia KADD'ów
k – ilość osób, które przedmiot zaliczyły

2) n – liczba dzieci urodzonych w 2015 roku
p – prawdopodobieństwo, że urodzi się dziewczynka (=0,5)
k – ilość urodzonych dziewczynek

Rozkład wielomianowy - uogólnienie

- Jeśli przestrzeń zdarzeń elementarnych: $E = A_1 + A_2 + \ldots + A_l$
- Zdarzenia się wzajemnie wykluczają: $P(A_j) = p_j, \quad \sum_{j=1}^{l} p_j = 1$
- To prawdopodobieństwo zajścia k_j razy zdarzenia A_j:

$$W^n_{k_1, k_2, \ldots, k_l} = \frac{n!}{\prod_{j=1}^{l} k_j!} \prod_{j=1}^{l} p_j^{k_j}, \qquad \sum_{j=1}^{l} k_j = n$$

- Taki rozkład nazywamy **rozkładem wielomianowym**
- Jeśli zdefiniujemy zmienne losowe X_{ij} równe 1, gdy wynikiem i-tego doświadczenia jest zdarzenie A_j, lub równe 0 w przeciwnym razie, oraz

$$X_j = \sum_{i=1}^{n} X_{ij}$$

- Wtedy wartość oczekiwana i elementy macierzy kowariancji:

$$E(X_j) = \widehat{x}_j = n\,p_j \qquad\qquad c_{ij} = np_i(\delta_{ij} - p_j)$$

- Dokładniejsze wyprowadzenie:
https://pl.wikibooks.org/wiki/Statystyka_matematyczna/Twierdzenie_o_rozk%C5%82adzie_wielomianowym

Częstość i prawo wielkich liczb

- W rzeczywistości nie znamy prawodpodobieństw zdarzeń (np. p_j w rozkł. wielomianowym) – wyznaczamy je eksperymentalnie
- **Częstość** wystąpienia zdarzenia A_j w n doświadczeniach będzie określona wzorem: $H_j = \frac{1}{n} X_j$
- Częstość jest zmienną losową, dla ktorej (przy n próbach):

$$E(H_j) = \hat{h}_j = E\left(\frac{x_j}{n}\right) = p_j \qquad \sigma^2(H_j) = \sigma^2\left(\frac{X_j}{n}\right) = \frac{1}{n^2}\sigma^2(X_j) = \frac{1}{n} p_j(1-p_j)$$

- <u>Wartość oczekiwana częstości jest równa jego prawdopodobieństwu</u>. Iloczyn $p_j(1\text{-}p_j)$ jest zawsze mniejszy od 1/4, więc <u>standardowe odchylenie częstości jest mniejsze</u> niż $1/\sqrt{n}$. Jest to **prawo wielkich liczb**
- Przeprowadzenie n prób umożliwia pomiar prawdopodobieństwa zdarzenia A_j, <u>kwadrat niepewności jest wtedy odwrotnie proporcjonalny do</u> n. Jest to tzw. **niepewność statystyczna**

Rozkład hipergeometryczny

- W urnie jest N kul – k białych i N-K czarnych
- W n próbach wyciągamy (bez zwracania) k kul białych i n-k=l kul czarnych. Jakie jest prawdopodobieństwo wyciągnięcia k kul białych?
- Wylosowanie kolejnej kulki zmienia proporcje kul białych do czarnych i wpływa na wynik kolejnego losowania – <u>rozkład dwumianowy nie ma tu zastosowania</u>. Mamy jednak:
  - liczba możliwości wylosowania n z N kulek: $\binom{N}{n}$
  - prawdopodobieństwo takiego zdarzenia: $1/\binom{N}{n}$
  - możliwość wylosowania k spośród K białych i l spośród L czarnych kulek wynoszą: $\binom{K}{k}$ $\binom{L}{l}$
  - prawdopodobieństwo szukane wynosi zatem: $W_k = \frac{\binom{K}{k}\binom{L}{l}}{\binom{N}{n}}$
- Analogicznie jak w rozkładzie dwumianowym, definiujemy zmienną losową: $X = \sum_{i=1}^{n} X_i$

Rozkład hipergeometryczny

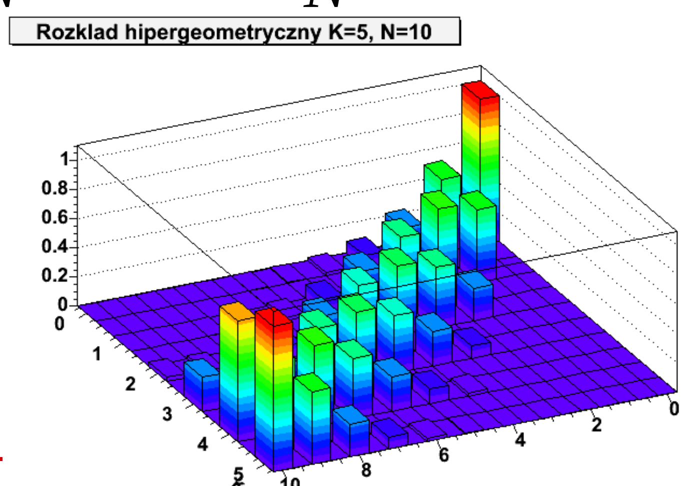
- Analogicznie jak w rozkładzie dwumianowym, definiujemy zmienną losową: $X=\sum_{i=1}^{n} X_i$
  - X_i przyjmuje wartość 1 dla białych i 0 dla czarnych wylosowanych kul

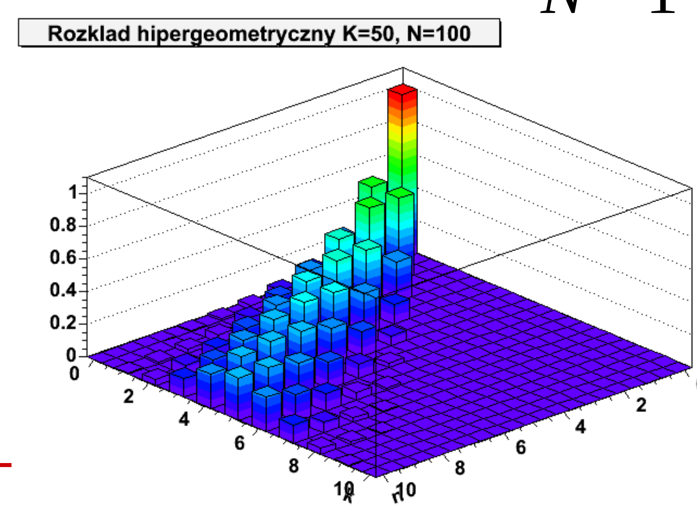
- Można pokazać, że (Brandt):

$$E(X)=n\frac{K}{N} \qquad \sigma^2(X)=\frac{nK(K-N)(N-n)}{N^2(N-1)}$$

- Dla $n \ll N$ rezultat kolejnego losowania niewiele wpływa na następne wyniki. Wtedy rozkłąd hipergeometrycznzy upodabnia się do dwumianowego:

$$p=\frac{K}{N}, \quad q=\frac{N-K}{N}, \quad E(X)=n\frac{K}{N}=np, \quad \sigma^2(X)=\frac{npq(N-n)}{N-1}$$





Rozkład Poissona

- Rozważmy rozkład dwumianowy:

$$P(k)=W_k^n=\binom{n}{k}p^k q^{n-k};\quad q=1-p$$

- dla $n\rightarrow\infty$ ale przy stałym $np=\lambda$ rozkład dwumianowy dąży do **rozkładu Poissona** (wyprowadzenie – Brandt):

$$\lim_{n\rightarrow\infty} W_k^n=f(k)=\frac{\lambda^k}{k!}e^{-\lambda} \qquad\qquad W_k^n=\binom{n}{k}p^k q^{n-k}$$

  – normalizacja:

$$\sum_{k=0}^{\infty} f(k)=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\dots\right)=e^{-\lambda}e^{\lambda}=1$$

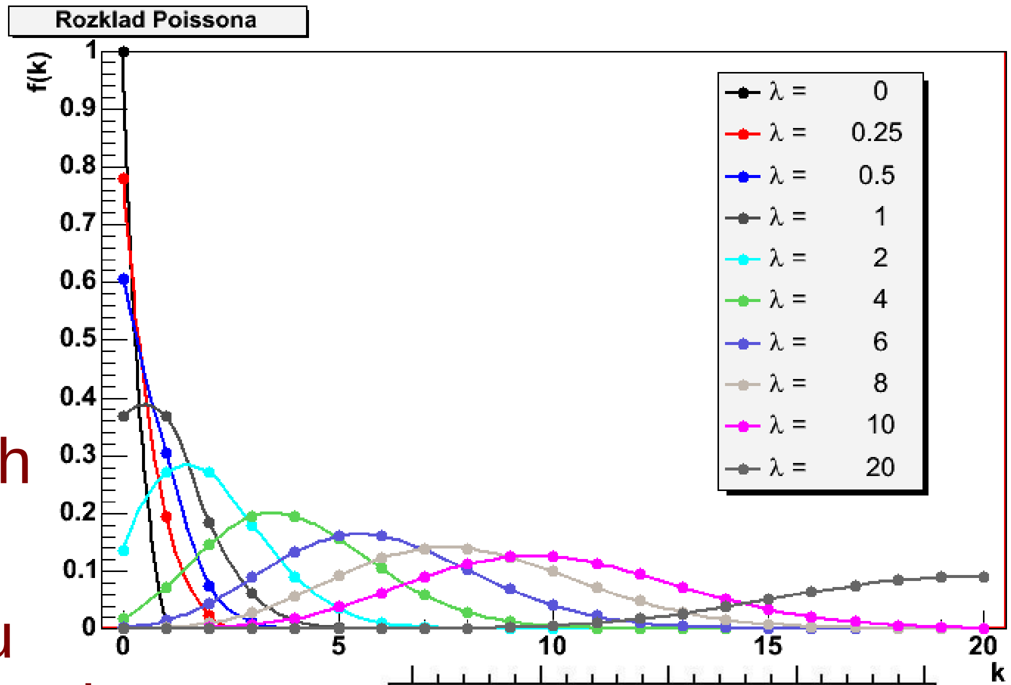
  – wartość oczekiwana: $E(K)=\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda\sum_{j=0}^{\infty}\frac{\lambda^j}{j!}e^{-\lambda}=\lambda$

  – wariancja: $\sigma^2(K)=E(K^2)-(E(K))^2=\lambda(\lambda+1)-\lambda^2=\lambda$

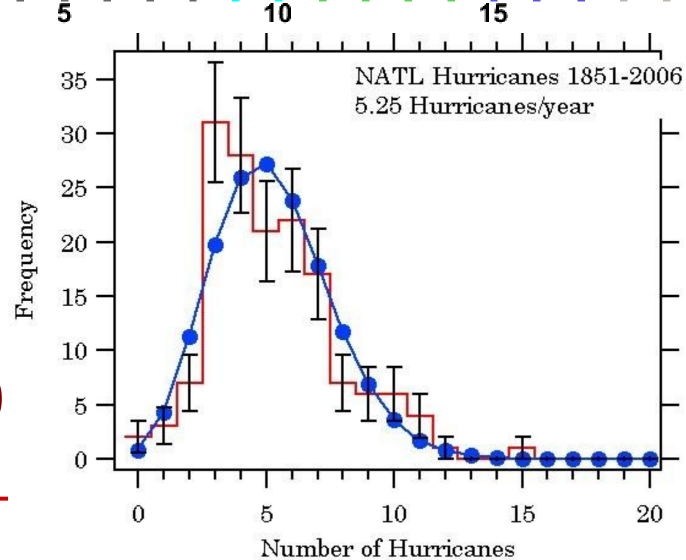
  – Skosność i wsp. asymetrii: $\mu_3=E\left(\left(k-\hat{k}\right)^3\right)=\lambda \qquad \gamma=\frac{\mu_3}{\sigma^3}=\frac{\lambda}{\lambda^{3/2}}=\lambda^{-1/2}$

Rozkład Poissona – przykłady

- Rozkład Poissona stosujemy wtedy, gdy mamy dużą liczbę niezależnych zdarzeń, z których tylko nieliczne mają interesującą nas własność (duże n, małe p w rozkł. dwumianowym)

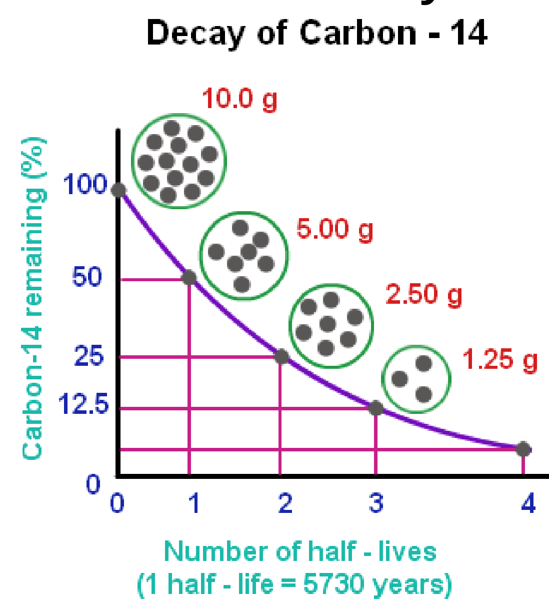


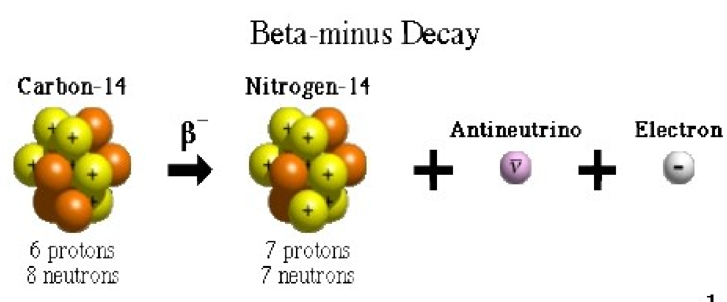
- Rozkład Poissona występuje tam, gdzie mamy zjawiska dyskretne, gdy prawdopodobieństwo wystąpienia zjawiska jest stałe w czasie lub przestrzeni:
  - liczba połączeń przychodzących do centrali na minutę
  - liczba mutacji w danym odcinku DNA po ekspozycji na pewną dawkę promieniowania
  - liczbę zabitych każdego roku przez kopnięcie konia w korpusie kawalerii w Prusach (Wikipedia)



Rozkład Poissona - rozpad promieniotwórczy

- Mamy jądro promieniotwórcze o czasie życia τ. Obserwujemy je w czasie $T \ll \tau$. Prawdopodobieństwo rozpadu jądra w tym czasie $W \ll 1$. Dzielimy czas T na n przedziałów, prawdopodobieństwo: $p=W/n$.
- Obserwujemy w czasie T źródło zawierające N jąder. Liczba przedziałów czasowych n_k, w których zaobserwowano $k=0, 1, 2, 3$ itd. rozpadów. Wtedy częstość $h(k) = n_k/n$.
- Doświadczalnie zaobserwowano, że dla $N \rightarrow \infty$ i dużych n rozkład $h(k)$ dąży do rozkładu Poissona, co stanowi bezpośredni dowód na niezależność i statystyczny charakter rozpadów promieniotwórczych (badania Rutherforda i Geigera).
- Analogicznie – częstość obserwowania k gwiazd w elemencie kąta bryłowego sfery niebieskiej lub k rodzynek w jednostkowym elemencie objętości keksu



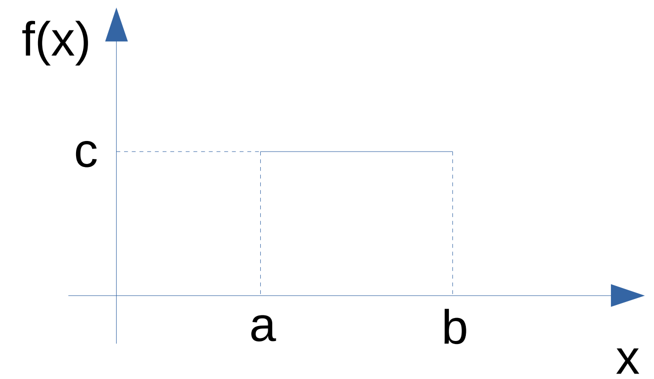


Rozkład jednostajny

- Gęstość prawdopodobieństwa:

$$f(x)=c\,;\,x\in\langle a,b\rangle$$

$$f(x)=0\,;\,x\in\mathbb{R}\setminus\langle a,b\rangle$$



- Współczynnik (normalizacja) c:

$$\int_{-\infty}^{\infty} f(x)\,dx=c\int_a^b dx=c(b-a)=1\Rightarrow c=\frac{1}{b-a}$$

$$f(x)=\frac{1}{b-a}\,;\,x\in\langle a,b\rangle$$

$$f(x)=0\,;\,x\in\mathbb{R}\setminus\langle a,b\rangle$$

- Dystrybuanta:

$$F(x)=0\,;\,x<a$$

$$F(x)=\frac{1}{b-a}\int_a^x dx'=\frac{x-a}{b-a}\,;\,x\in\langle a;b\rangle$$

$$F(x)=1\,;\,x>b$$

Wariancja: $\sigma^2(X)=E(X^2)-(E(X))^2$

$$E(X^2)=\frac{1}{b-a}\int_a^b x^2\,dx=\frac{(b^3-a^3)}{3(b-a)}=$$

$$=\frac{(b-a)(b^2+ba+a^2)}{3(b-a)}=\frac{b^2+ba+a^2}{3}$$

$$\sigma^2(X)=\frac{b^2+ba+a^2}{3}-\left(\frac{b+a}{2}\right)^2=$$

$$=\frac{b^2+ba+a^2}{3}-\frac{b^2+2ba+a^2}{4}=\boxed{\frac{(b-a)^2}{12}}$$

- Wartość oczekiwana:

$$E(X)=\hat{x}=\frac{1}{b-a}\int_a^b x\,dx=\frac{1}{2(b-a)}(b^2-a^2)=\frac{(b-a)(b+a)}{2(b-a)}=\boxed{\frac{b+a}{2}}$$

Rozkład wykładniczy

- Gęstość prawdopodobieństwa:

$$f(x)=\lambda e^{-\lambda x};\ x\geqslant 0;\ \lambda>0$$

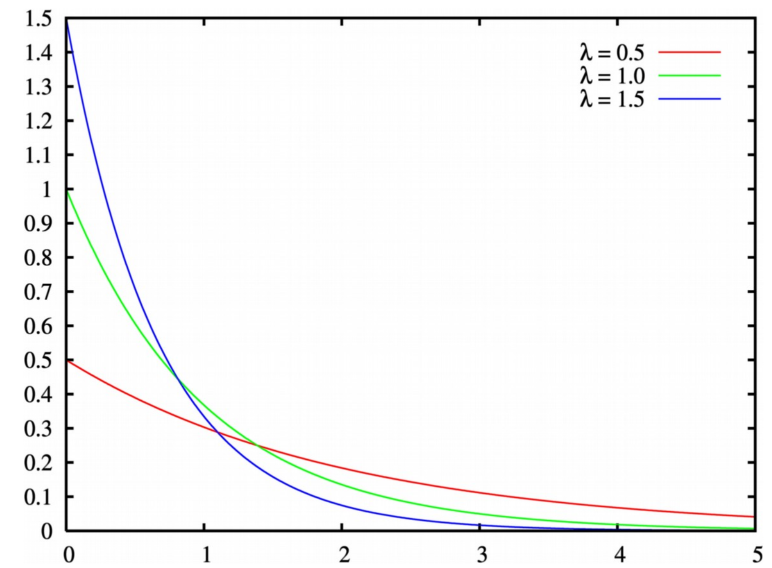
$$f(x)=0;\ x<0$$

- Dystrybuanta:

$$F(x)=0;\ x<0$$

$$F(x)=\int_0^x f(x)\,dx=\lambda\int_0^x e^{-\lambda x'}dx'=\left[\frac{-\lambda}{\lambda}e^{-\lambda x'}\right]_0^x$$

$$F(x)=1-e^{-\lambda x};\ x\geq 0$$

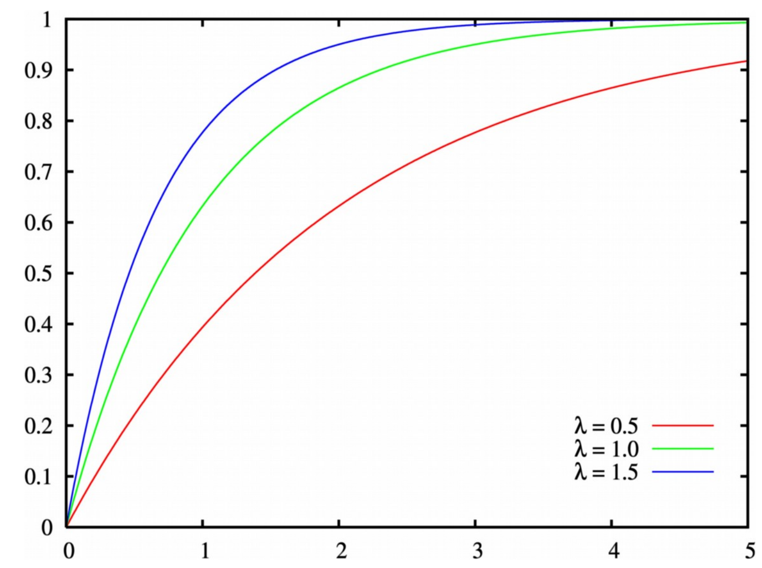


- Wartość oczekiwana:

$$E(x)=\hat{x}=\int_0^\infty x f(x)\,dx=\lambda\int_0^\infty e^{-\lambda x}x\,dx=\boxed{\frac{1}{\lambda}}$$

- Wariancja: $E(x^2)=\int_0^\infty x^2 f(x)\,dx=\frac{2}{\lambda^2}$

$$\sigma^2(x)=E(x^2)-(E(x))^2=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\boxed{\frac{1}{\lambda^2}}$$



Rozkład normalny standardowy

- Gęstość prawdopodobieństwa:

$$f(x) \equiv \varphi_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

  – rozkład o średniej 0 i wariancji 1

- Dystrybuanta nie ma postaci analitycznej (korzystamy z tabel)
- Rozkład jest unormowany:

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$$

- Jeśli wprowadzimy zmienną:

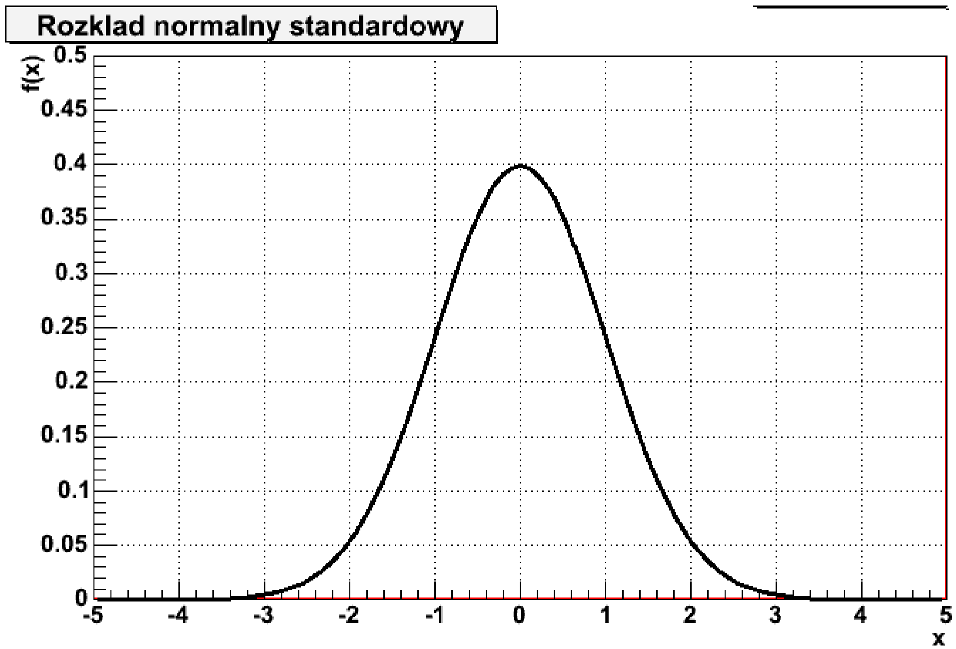
$$Y = (X - a)/b$$

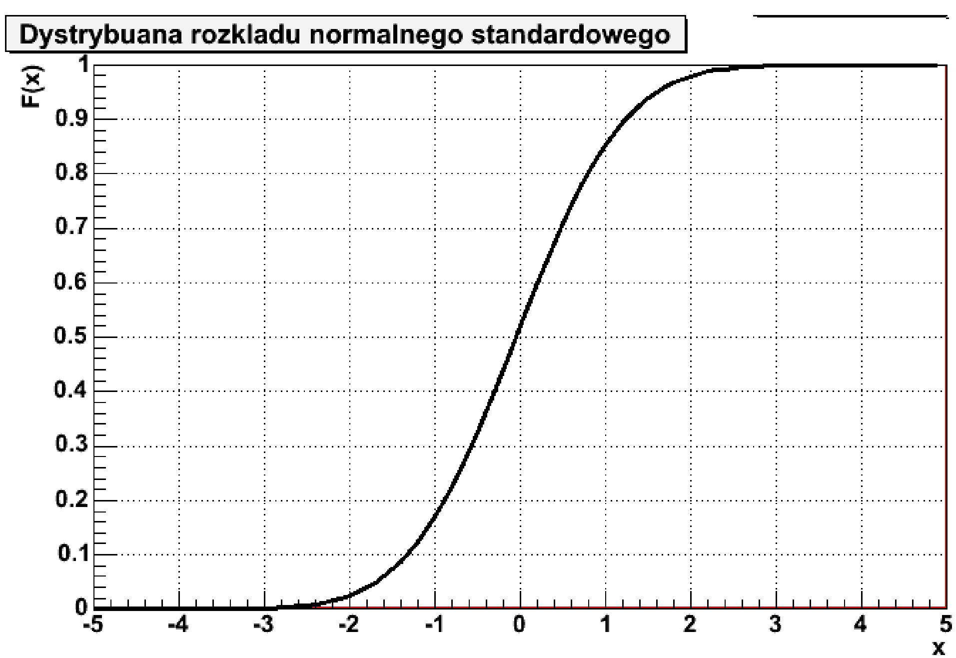
- Otrzymamy rozkład Gaussa:

$$f(y) \equiv \varphi(y) = \frac{1}{\sqrt{2\pi} b} e^{-(y-a)^2/2b^2}$$

  – średnia (przesunięcie): $\hat{y} = a$

  – wariancja (szerokość): $\sigma^2(Y) = b$





Rozkład normalny standardowy – własności

- Punkt przegięcia rozkładu:
  - standardowego $x=\pm 1$
  - Gaussa $x=a\pm b$
- Założmy, że znamy dystrybuantę:

$$F_0(x) \equiv \Phi_0(x) = P(X \leqslant x)$$

- Ze względu na asymetrię gęstości:

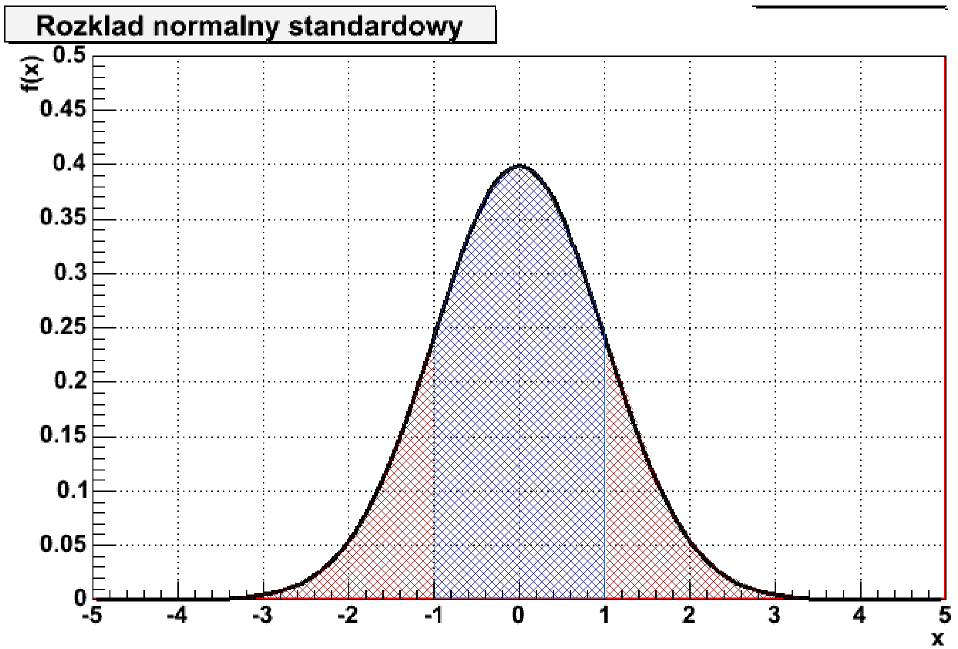
$$P(|X|>x) = 2\,\Phi_0(-|x|) = 2(1-\varphi_0(|x|))$$

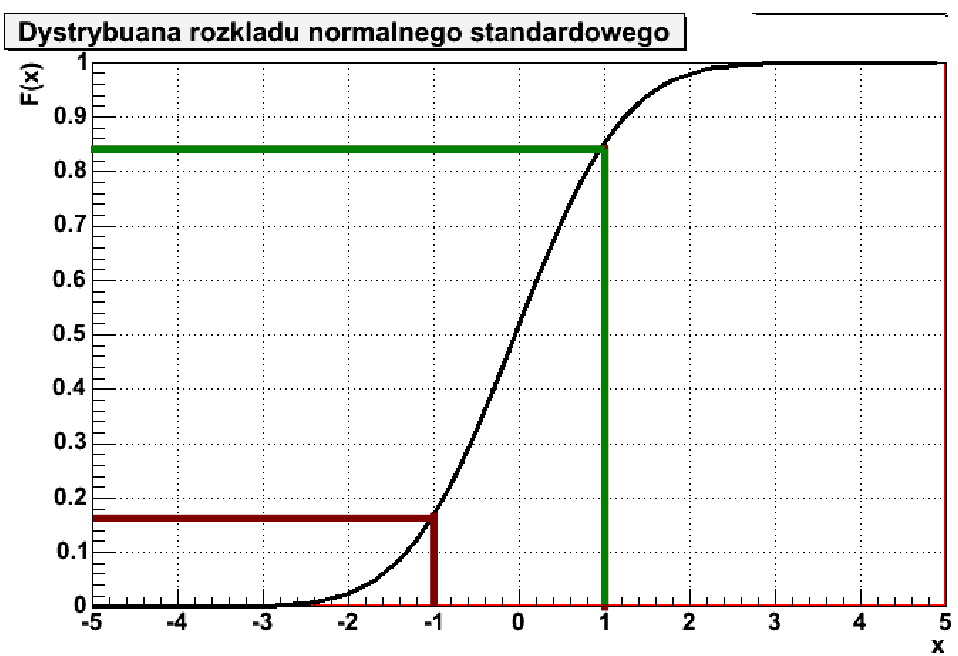
- Analogicznie, wewnątrz przedziału $2x$:

$$P(|X|\leqslant x) = 2\,\Phi_0(|x|) - 1$$

- Dystrybuantę r. norm. można uogólnić na r. Gaussa:

$$\Phi(y) = \Phi_0\left(\frac{x-a}{b}\right)$$





Rozkład normalny standardowy – własności

- Wtedy szczególnie interesujące jest obliczenie występowania zmiennej los. dla wielokrotności odchylenia standardowego:

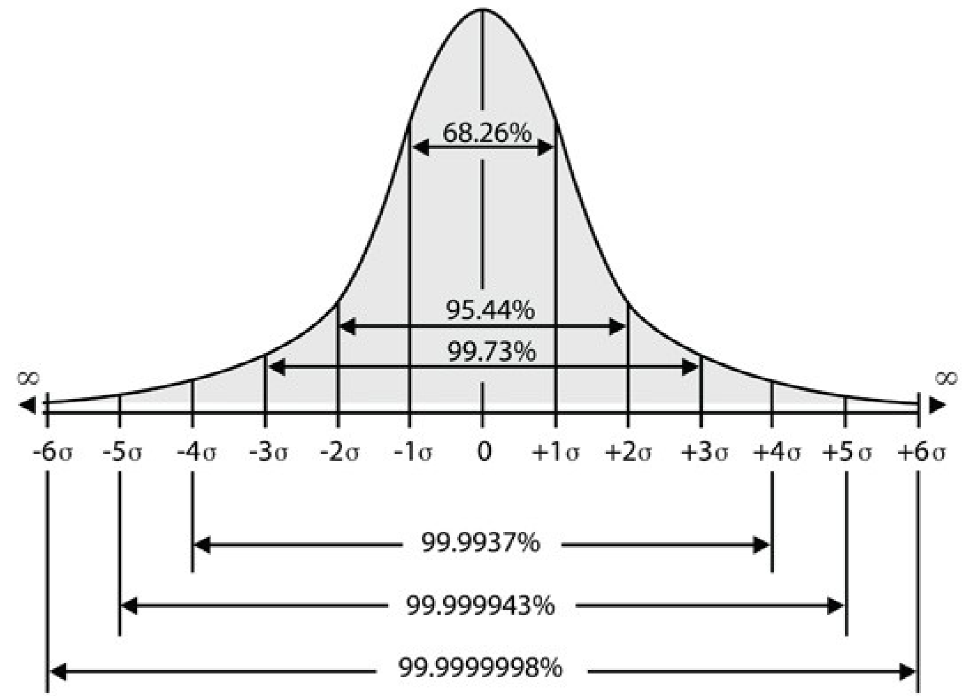
$$P(|Y-a|\leqslant n\sigma)=2\Phi_0\left(\frac{nb}{b}\right)-1=2\Phi_0(n)-1$$

- Otrzymamy wtedy:

$$P(|Y-a|\leqslant\sigma)=68{,}3\,\% \qquad P(|Y-a|>\sigma)=31{,}7\,\%$$

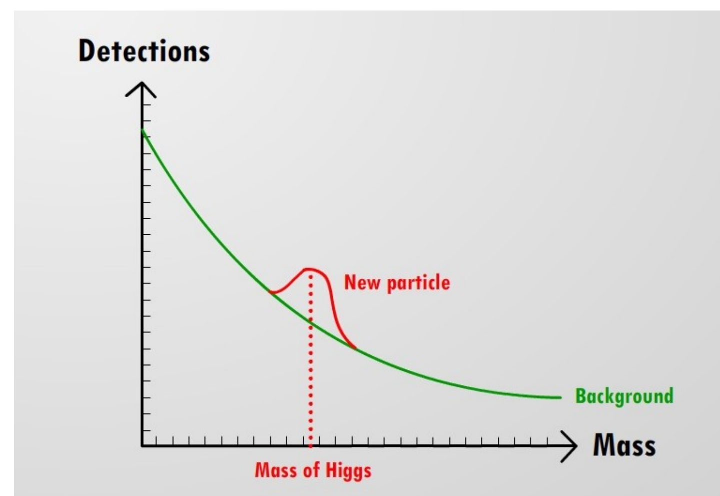
$$P(|Y-a|\leqslant 2\sigma)=95{,}4\,\% \qquad P(|Y-a|>2\sigma)=4{,}6\,\%$$

$$P(|Y-a|\leqslant 3\sigma)=99{,}8\,\% \qquad P(|Y-a|>3\sigma)=0{,}2\,\%$$



- Z Wykładu 1 pamiętamy, że **współczynnik rozszerzenia** niepewność typu A zwykle jest między 2 a 3 – tu widać dlaczego

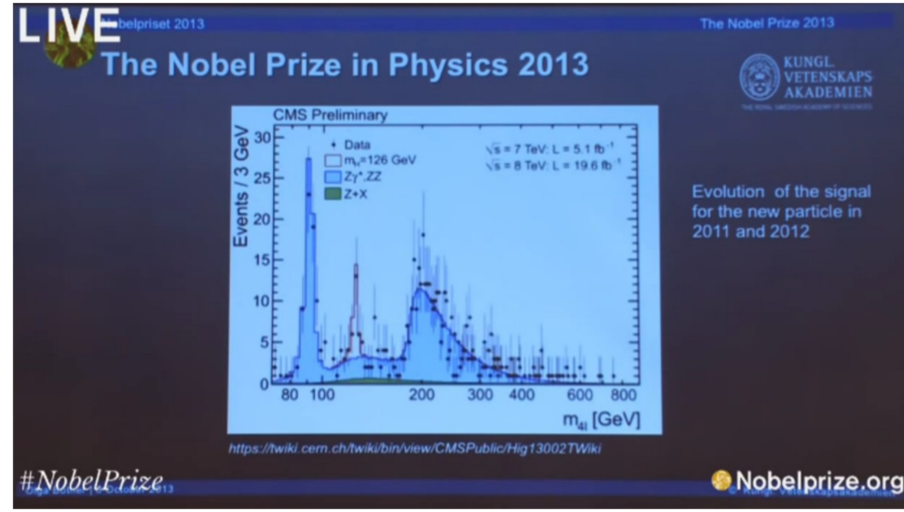
- W nauce przez odchylenie standardowe określamy również różnice w obserwowanym sygnalne eksperymentalnym w stosunku do sytuacji, gdy efektu fizycznego nie ma



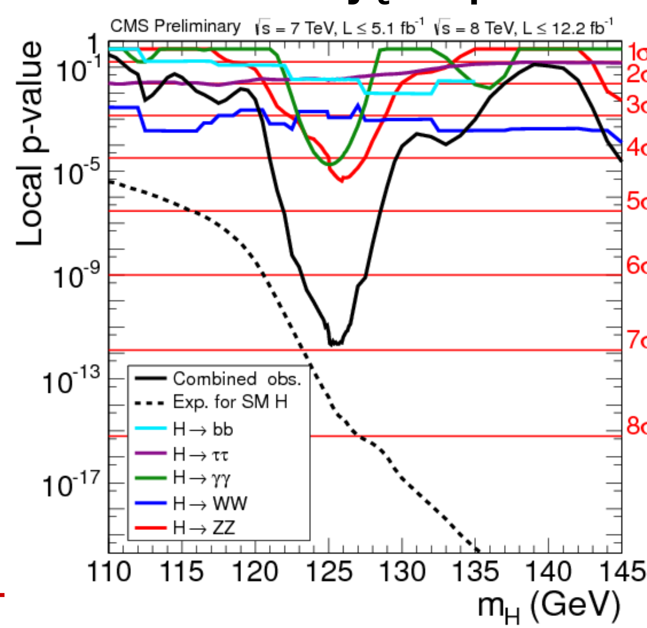
Wielokrotności sigma

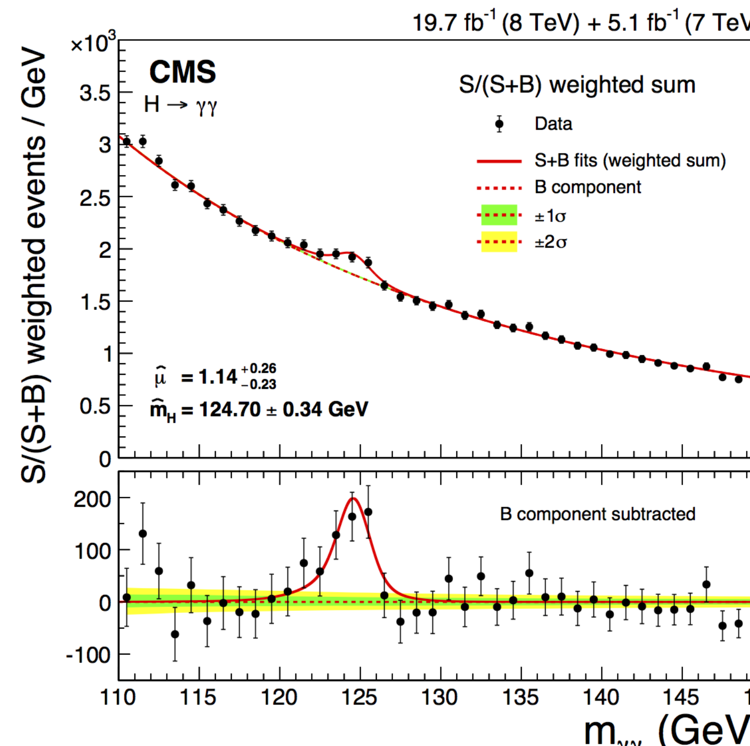
- Idealnym przykładem jest odkrycie bozonu Higgsa
- W fizyce cząstek przyjęło się, że dopiero mając **odchylenie 5σ można mówić o odkryciu**:

$$P(|Y-a|\leqslant 5\sigma)=99{,}99994\,\%$$



- Różnica na takim poziomie wymagała zebrania dużej ilości danych, stąd potwierdzenie jego istnienia zajęło ponad 3 lata





KONIEC