



# Komputerowa analiza danych doświadczalnych

Wykład 7  
7.04.2017

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

*Semestr letni 2016/2017*



# Centralne twierdzenie graniczne - przypomnienie

## Sploty

## Pobieranie próby, estymatory



# Centralne twierdzenie graniczne

# Centralne twierdzenie graniczne

- Dlaczego rozkład normalny jest tak ważny w rachunku prawdopodobieństwa i statystyce?
- Mówi o tym **centralne twierdzenie graniczne** (*ang. central limit theorem*) – jedno z najważniejszych twierdzeń rachunku prawdopodobieństwa:
  - jeżeli zmienne losowe  $X_i$  są zmiennymi niezależnymi o jednakowych wartościach średnich  $a$  i odchyleniach standardowych  $b$ , to **rozkład normalny** ma zmienna:

$$X = \lim_{n \rightarrow \infty} \sum_{i=1}^n X_i \quad E(X) = na, \quad \sigma^2(X) = nb^2$$

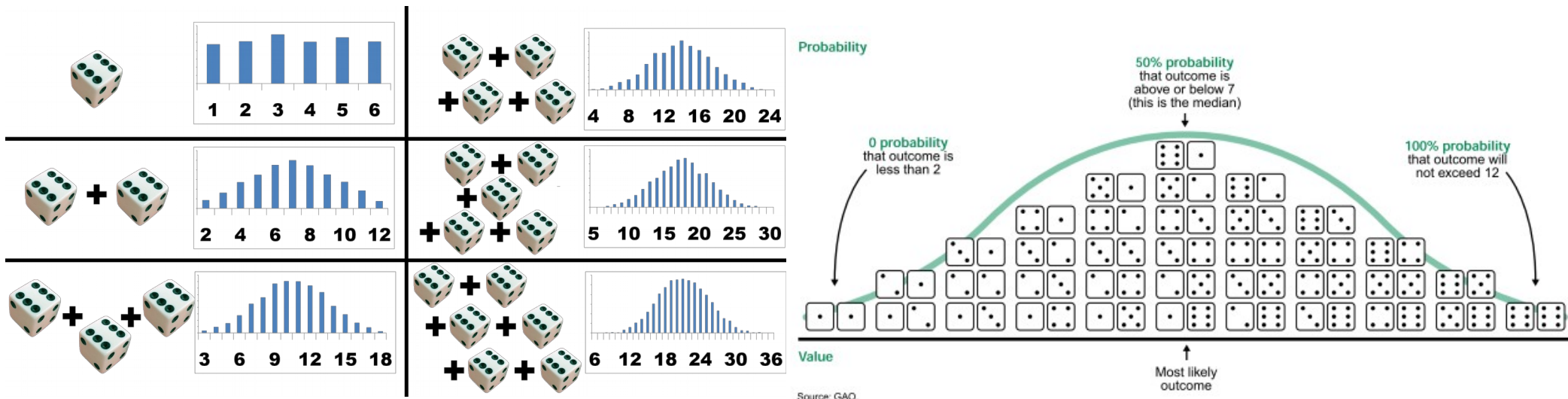
- ponadto, zmienna  $\xi = \frac{1}{n} X = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i$  ma rozkład normalny z:

$$E(\xi) = a, \quad \sigma^2(\xi) = b^2/n$$

- Innymi słowy – mając  $n$  niezależnych zmiennych o jednakowym (**dowolnym!**) rozkładzie, to ich suma dla dużych  $n$  zbiega do rozkładu normalnego

# Centralne twierdzenie graniczne - przykład 1

- Wyobraźmy sobie eksperyment polegający na rzucie kostką (kostkami) i obserwowaniu całkowitej liczby oczek:
  - kolejne rzuty kostką (kostkami) są niezależne
  - jeśli rzucamy kostką jednokrotnie (albo 1 kostką), to prawdopodobieństwo uzyskania danej wartości jest jednakowe
  - jeśli rzucamy kostką dwukrotnie (albo 2 kostkami), to prawdopodobieństwo uzyskania sumy oczek nie jest już jednakowe
  - jeśli rzucimy kostką  $n$ -krotnie ( $n$ -kostkami) → rozkład normalny





# Sploty

# Suma zmiennych losowych jako splot

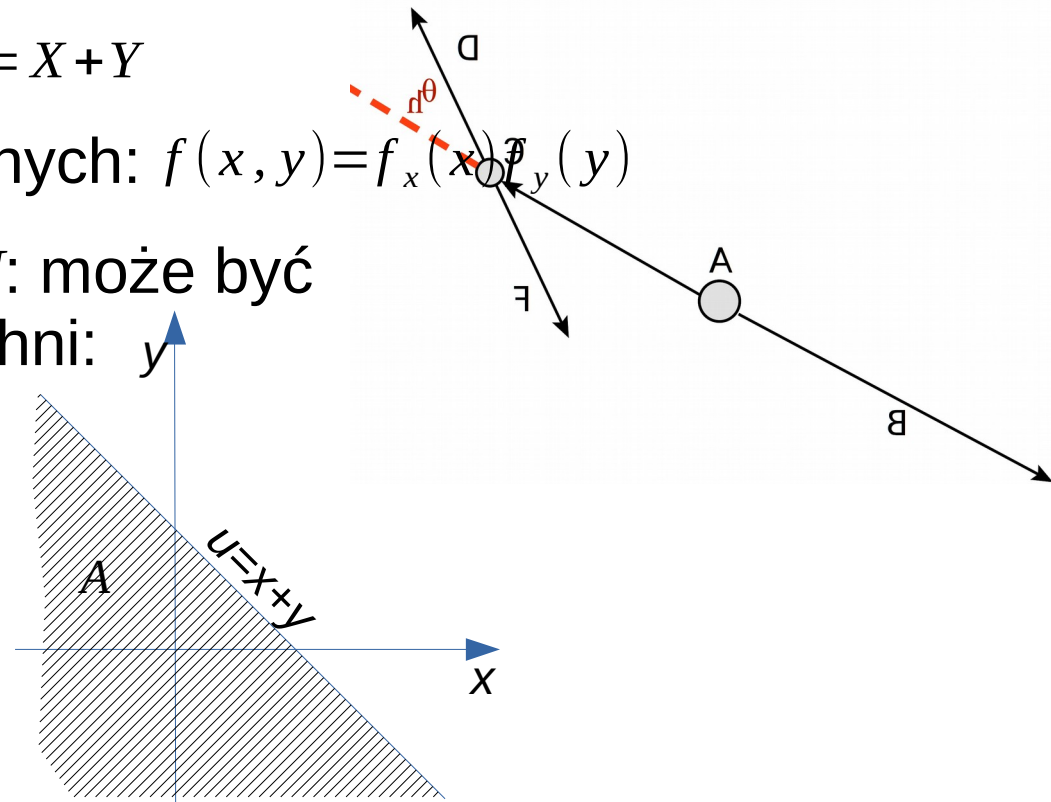
- W doświadczeniach eksperymentalnych bardzo często mamy do czynienia z sumą dwóch zmiennych losowych
- Na przykład – rozpad cząstek nietrwałych opisany jest pewnym kątem rozpadu, wynikającym ze statystycznego charakteru zjawiska fizycznego, zaś niepewność jego pomiaru z niedokładności przyrządu. Obserwowany rozkład jest **splotem** dwóch rozkładów

- Rozważmy zmienną losową:  $U = X + Y$

- Zakładamy niezależność zmiennych:  $f(x, y) = f_x(x) f_y(y)$

- Wtedy dystrybuanta zmiennej  $U$ : może być wyznaczona jako pole powierzchni:

$$\begin{aligned} F(u) &= P(U \leq u) = P(X + Y \leq u) = \\ &= \iint_A f_x(x) f_y(y) dx dy \\ &= \int_{-\infty}^{\infty} f_x(x) dx \int_{-\infty}^{u-x} f_y(y) dy \\ &= \int_{-\infty}^{\infty} f_y(y) dy \int_{-\infty}^{u-y} f_x(x) dx \end{aligned}$$



# Suma zmiennych losowych jako splot

- Z dystrybuanty wyznaczamy funkcję gęstości zmiennej  $U$ :

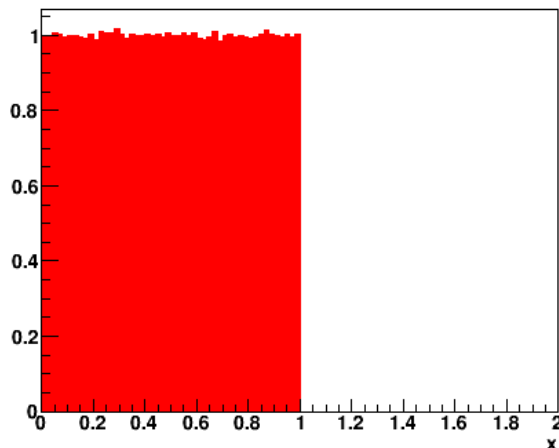
$$f(u) = \frac{dF(u)}{du} = \int_{-\infty}^{\infty} f_x(x) f_y(u-x) dx = \int_{-\infty}^{\infty} f_y(y) f_x(u-y) dy \equiv (f_x * f_y)(u)$$

- Funkcja  $f(u)$  tak zdefiniowana jest **splotem** funkcji  $f_x(x)$  i  $f_y(y)$
- Powyższy wzór będzie prawdziwy również wówczas, jeżeli zmienne  $X$  i  $Y$  są zdefiniowane tylko w pewnym zwartym obszarze (wtedy ustalamy odpowiednie – węższe i skończone, granice całkowania)
- Rozpatrzmy przypadek splotu dwóch rozkładów jednorodnych:

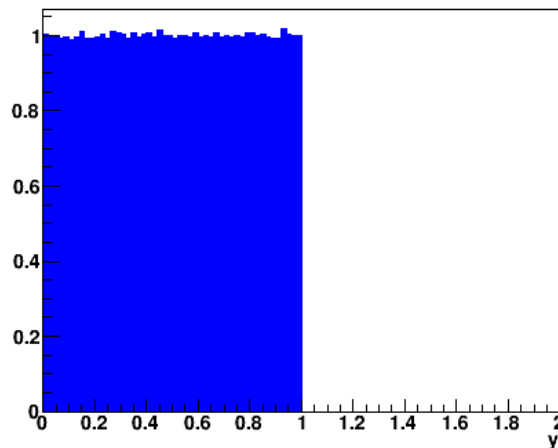
$$f_x(x) = \begin{cases} 1, & 0 \leq x < 1 \\ 0, & \text{w przeciwnym razie} \end{cases}$$

$$f_y(y) = \begin{cases} 1, & 0 \leq y < 1 \\ 0, & \text{w przeciwnym razie} \end{cases}$$

Rozkład jednostajny



Rozkład jednostajny



$$\begin{aligned} f(u) &= \int_0^1 f_x(x) f_y(u-x) dx = \\ &= \int_0^1 f_y(u-x) \end{aligned}$$



# Suma zmiennych losowych jako splot

- Rozpatrzmy przypadek splotu dwóch rozkładów jednorodnych:

$$f_x(x) = \begin{cases} 1, & 0 \leq x < 1 \\ 0, & \text{w przeciwnym razie} \end{cases} \quad f_y(y) = \begin{cases} 1, & 0 \leq y < 1 \\ 0, & \text{w przeciwnym razie} \end{cases}$$

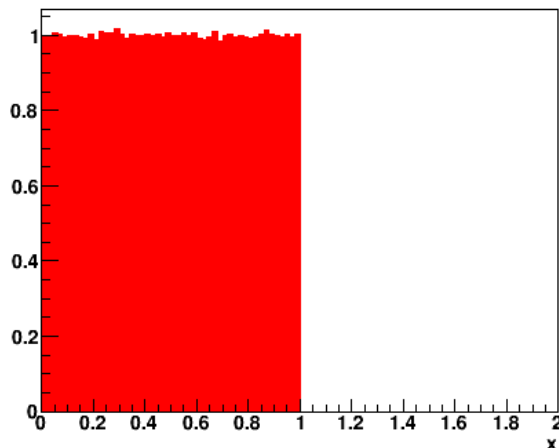
$$f(u) = \int_0^1 f_x(x) f_y(u-x) dx = \int_0^1 f_y(u-x) dx \quad \begin{matrix} v = u-x \\ dv = -dx \end{matrix} \Rightarrow f(u) = - \int_u^{u-1} f_y(v) dv = \int_{u-1}^u f_y(v) dv$$

- Zmienna  $u$  zmienia się od 0 do 2, zatem rozważmy 2 przypadki:

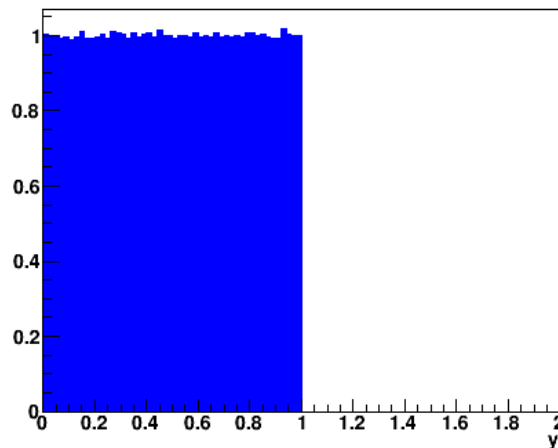
$$(a) \quad 0 \leq u < 1: f_1(u) = \int_0^u f_y(v) dv = \int_0^u 1 dv = u$$

$$(b) \quad 1 \leq u < 2: f_2(u) = \int_{u-1}^1 f_y(v) dv = \int_{u-1}^1 1 dv = 2 - u$$

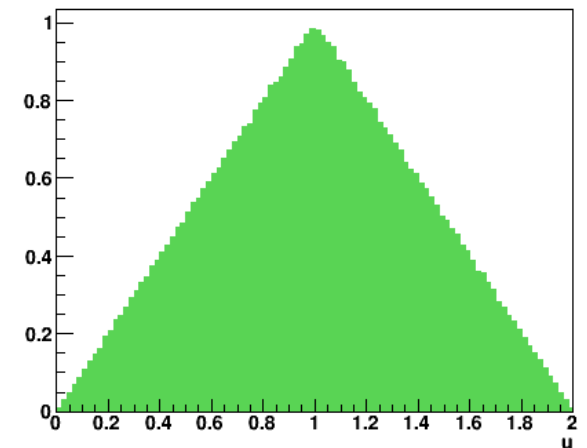
Rozkład jednostajny



Rozkład jednostajny



Splot 2 rozkładów jednostajnych



# Suma zmiennych losowych jako splot

- Rozpatrzmy przypadek splotu dwóch rozkładów jednorodnych:

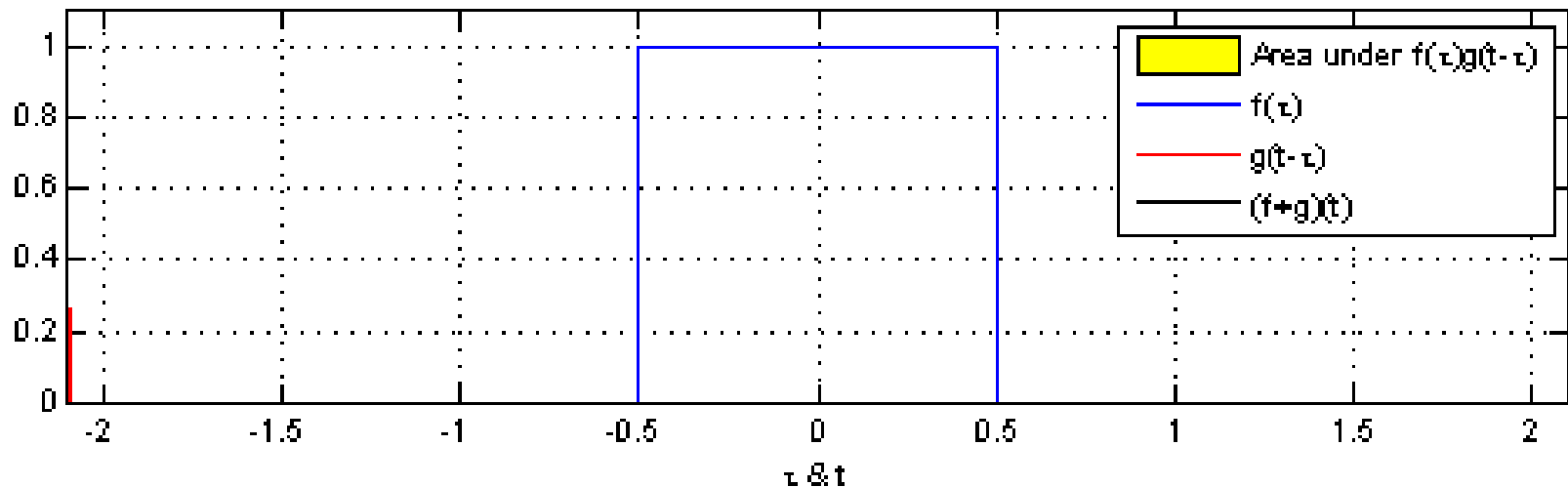
$$f_x(x) = \begin{cases} 1, & 0 \leq x < 1 \\ 0, & \text{w przeciwnym razie} \end{cases} \quad f_y(y) = \begin{cases} 1, & 0 \leq y < 1 \\ 0, & \text{w przeciwnym razie} \end{cases}$$

$$f(u) = \int_0^1 f_x(x) f_y(u-x) dx = \int_0^1 f_y(u-x) dx \quad \begin{matrix} v = u-x \\ dv = -dx \end{matrix} \Rightarrow f(u) = - \int_u^{u-1} f_y(v) dv = \int_{u-1}^u f_y(v) dv$$

- Zmienna  $u$  zmienia się od 0 do 2, zatem rozważmy 2 przypadki:

$$(a) \quad 0 \leq u < 1: f_1(u) = \int_0^u f_y(v) dv = \int_0^u 1 dv = u$$

$$(b) \quad 1 \leq u < 2: f_2(u) = \int_{u-1}^1 f_y(v) dv = \int_{u-1}^1 1 dv = 2 - u$$



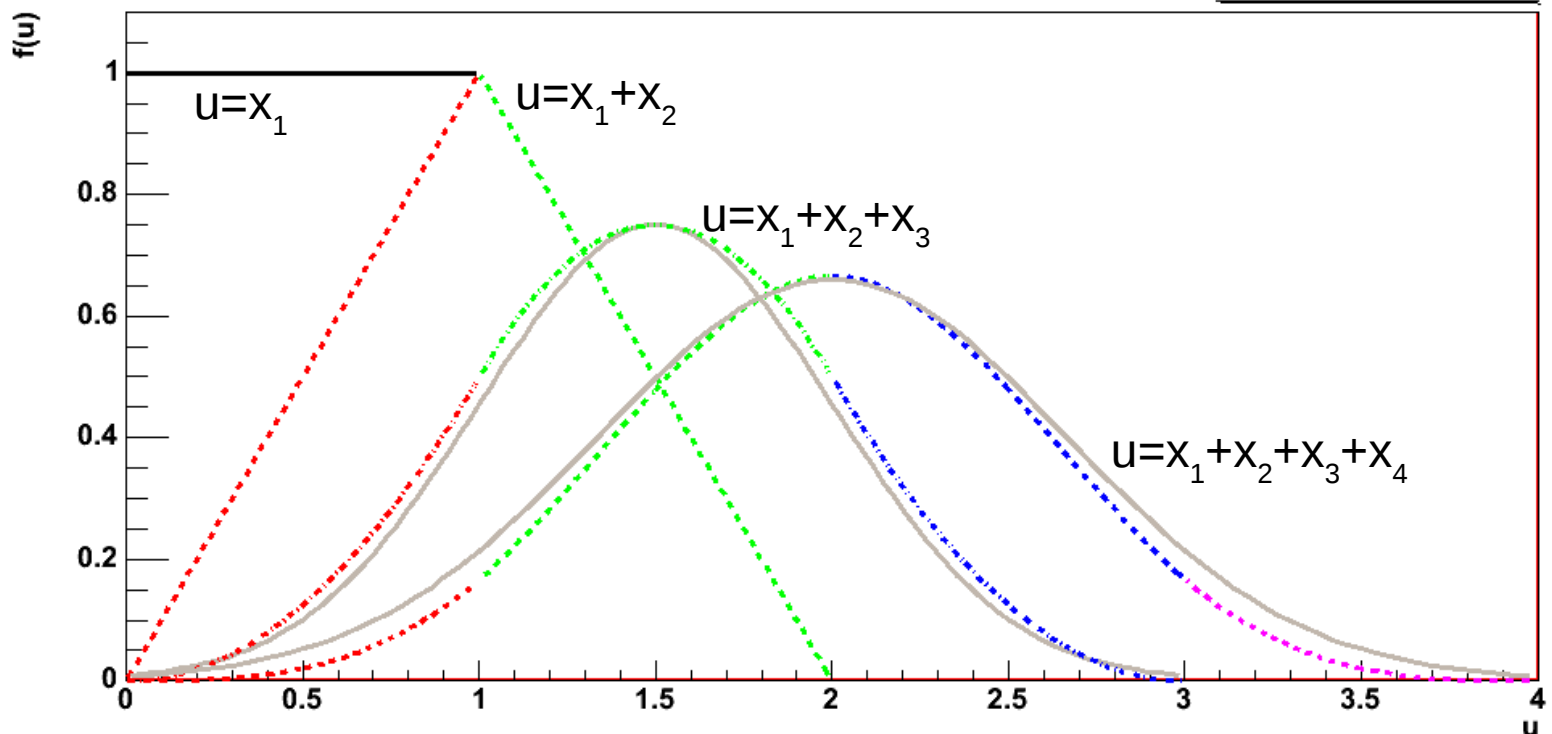
[https://en.wikipedia.org/wiki/Convolution#/media/File:Convolution\\_of\\_box\\_signal\\_with\\_itself2.gif](https://en.wikipedia.org/wiki/Convolution#/media/File:Convolution_of_box_signal_with_itself2.gif)

# Suma zmiennych losowych jako splot

- Analogicznie będzie z sumą trzech zmiennych losowych:

$$f(u) = \begin{cases} 1/2 u^2, & 0 \leq u < 1 \\ 1/2 (-2u^2 + 6u - 3), & 1 \leq u < 2 \\ 1/2 (u-3)^2, & 2 \leq u < 3 \end{cases}$$

- Zgodnie z CTG – im więcej rozkładów w splocie, tym bardziej rozkład sumy przypomina rozkład Gaussa:



# Sploty z rozkładem normalnym

- Przykład: Mierzmy zmienną  $X$  opisaną gęstością prawdopodobieństwa  $f_x(x)$ . Pomiar obarczony jest niepewnością  $Y$  mającą rozkład normalny. Wynik jest zatem sumą zmiennych losowych:  $U = X + Y$
- Gęstość prawdopodobieństwa zmiennej  $U$  wynosi wtedy:

$$f(u) = \int_{-\infty}^{\infty} f_x(x) f_y(u-x) dx = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} f_x(x) \exp\left(\frac{-(u-x)^2}{2\sigma^2}\right) dx$$

- Problem: eksperymentalnie otrzymujemy funkcję  $f(u)$ , ale tak naprawdę interesuje nas  $f_x(x)$ . Jak ją wyznaczyć?
  - w ogólnym przypadku jest to niemożliwe
  - można tego dokonać dla pewnej ograniczonej klasy funkcji  $f(u)$
  - najczęściej posługujemy się tutaj metodami Monte Carlo

# Sploty z rozkładem normalnym - przykład 1

- Przykład: Splot rozkładu jednostajnego z rozkładem normalnym (o średniej równej 0)
- W tym przypadku możliwe jest rozwiązanie analityczne. Korzystamy ze wzorów:

$$f(x) = \frac{1}{b-a}; x \in \langle a, b \rangle \quad g(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2} \quad h(u) = \int_{-\infty}^{\infty} f(x)g(u-x)dx$$

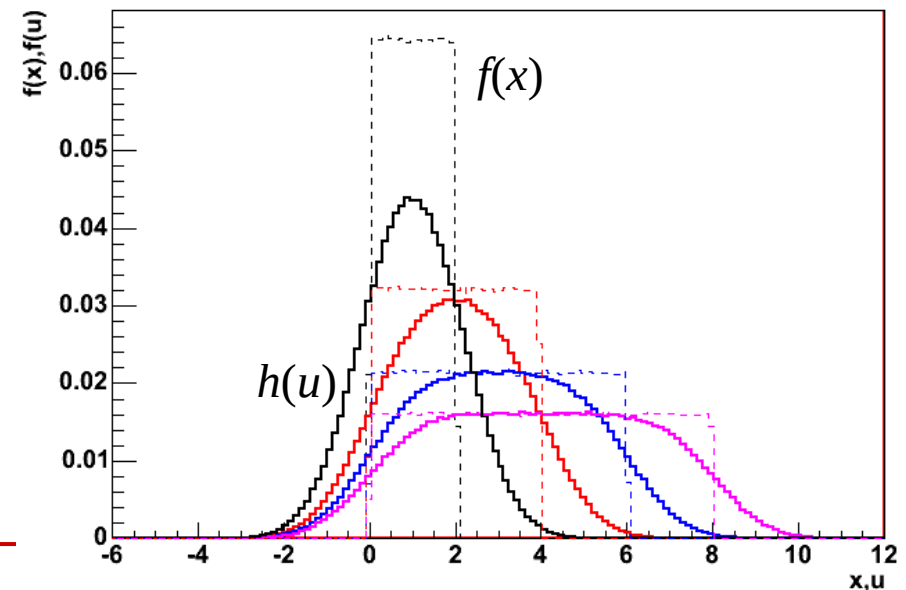
$$f(x) = 0; x \in \mathbb{R} \setminus \langle a, b \rangle$$

- Wtedy, wprowadzając zmienną  $v = (x-u)/\sigma$  otrzymujemy:

$$h(u) = \frac{1}{b-a} \frac{1}{\sqrt{2\pi}\sigma} \int_a^b \exp\left(-\frac{(u-x)^2}{2\sigma^2}\right) dx = \frac{1}{b-a} \frac{1}{\sqrt{2\pi}} \int_{(a-u)/\sigma}^{(b-u)/\sigma} \exp\left(-\frac{1}{2}v^2\right) dv$$

- Zaś uwzględniając dystrybuantę rozkładu normalnego:

$$h(u) = \frac{1}{b-a} \left( \Phi_0\left(\frac{b-u}{\sigma}\right) - \Phi_0\left(\frac{a-u}{\sigma}\right) \right)$$

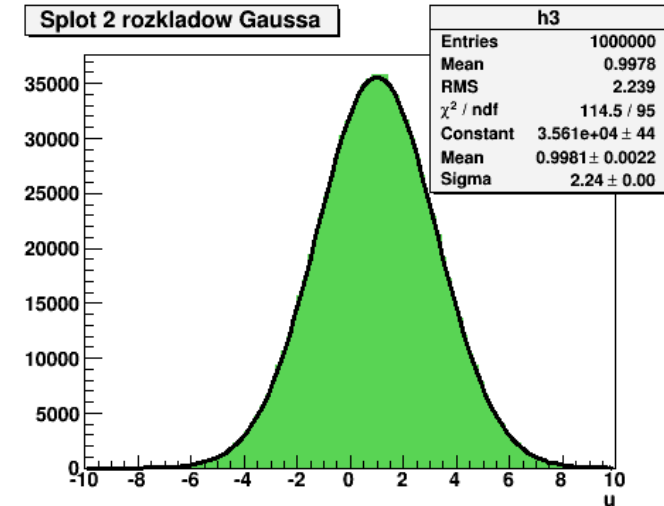
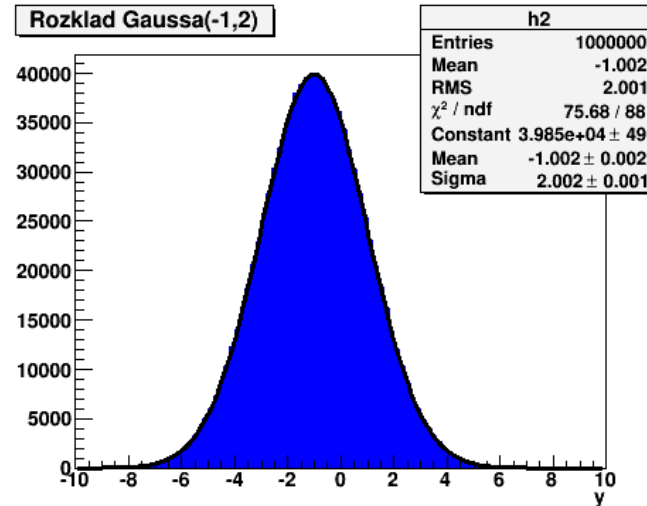
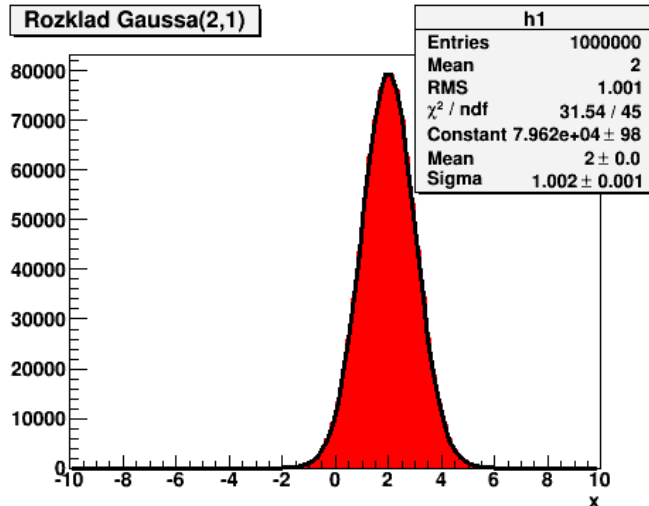


# Sploty z rozkładem normalnym - przykład 2

- Przykład: Splot dwóch rozkładów normalnych – dodawanie niepewności “w kwadracie”
- Splot dwóch rozkładów normalnych o wartościach średnich równych 0 i wariancjach  $\sigma_x$ ,  $\sigma_y$  ma postać rozkładu normalnego:

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{u^2}{2\sigma^2}\right), \quad \sigma^2 = \sigma_x^2 + \sigma_y^2$$

- Widzimy, że **wariancje się dodają** (odchylenia std. dodają się w kwadracie)
- Jeśli średnie rozkładów różne od 0 – **wartości oczekiwane również się dodają**



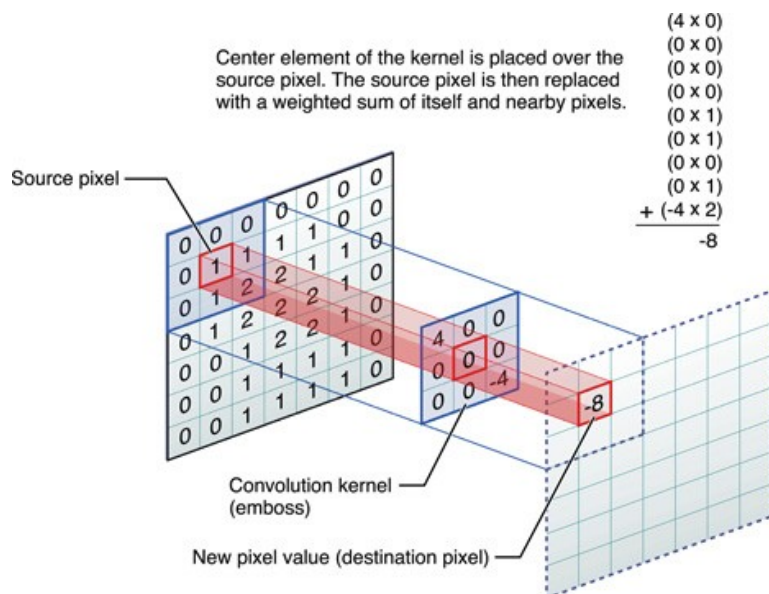
# Zastosowanie splotów

- Cyfrowe przetwarzanie obrazów
- Akustyka
- Muzyka elektroniczna
- W fizyce gdzie się pojawia superpozycja
- W planowaniu radioterapii (rozkłady dawki)



<https://upload.wikimedia.org/wikipedia/en/2/24/Lenna.png>  
Playboy 1972 – standardowy obrazek w grafice komput.

	-2	-1	0
	-1	1	1
	0	1	2



Original



Emboss

<https://developer.apple.com/library/content/documentation/Performance/Conceptual/vImage/ConvolutionOperations/ConvolutionOperations.html>

# Zastosowanie splotów

- Bardzo ważnym zastosowaniem splotów są badania farmakokinetyczne leków – koncentracja leku w osoczu krwi w czasie jest splotem funkcji absorpcji leku oraz jego eliminacji

- The absorption rate  $r_{abs}$  that results in plasma concentration  $c(t)$  may be estimated by solving following eq.

$$c(t) = \int_0^t c_s(t-u)r_{abs}(u)du$$

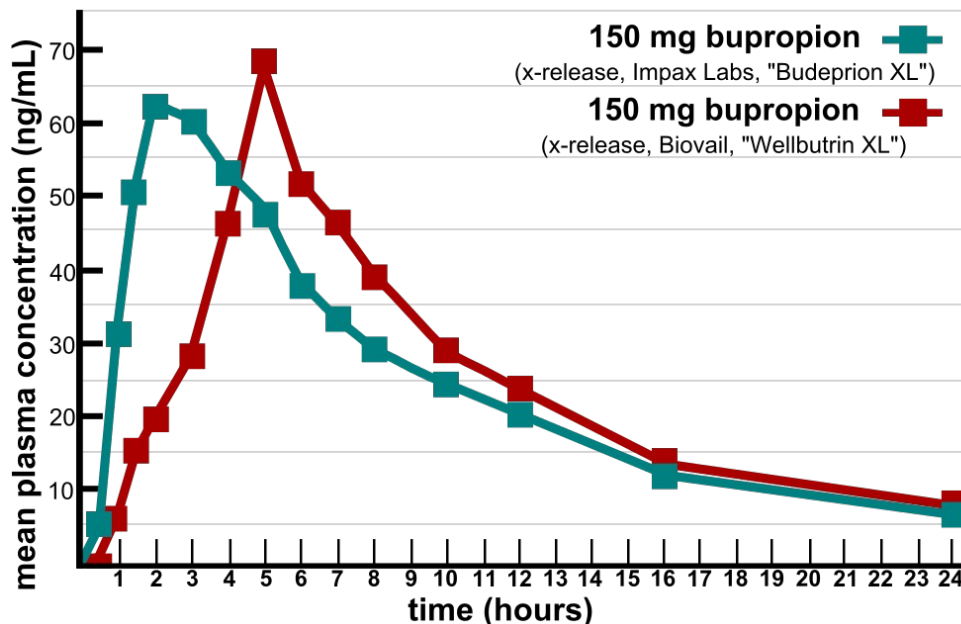
$c_s$  is the concentration time profile resulting from instantaneous absorption of a unit amount of drug which is typically absorbed from bolus IV injection or reference oral solution data

$c(t)$  is plasma conc. versus time profiles of tested formulation

$r_{abs}$  is the input rate of the oral solid dosage form in to the body

$u$  is the variable of integration

<https://www.slideshare.net/jaspreetguraya/in-vitro-in-vivo-correlation-ivivc>



[https://upload.wikimedia.org/wikipedia/commons/7/7d/Bupropion\\_bio\\_equivalency\\_comparison.svg](https://upload.wikimedia.org/wikipedia/commons/7/7d/Bupropion_bio_equivalency_comparison.svg)

[https://www.researchgate.net/publication/228486042\\_In\\_Vitro-In\\_Vivo\\_Correlation\\_IVIVC\\_and\\_Determining\\_Drug\\_Concentrations\\_in\\_Blood\\_from\\_Dissolution\\_Testing-A\\_Simple\\_and\\_Practical\\_Approach](https://www.researchgate.net/publication/228486042_In_Vitro-In_Vivo_Correlation_IVIVC_and_Determining_Drug_Concentrations_in_Blood_from_Dissolution_Testing-A_Simple_and_Practical_Approach)

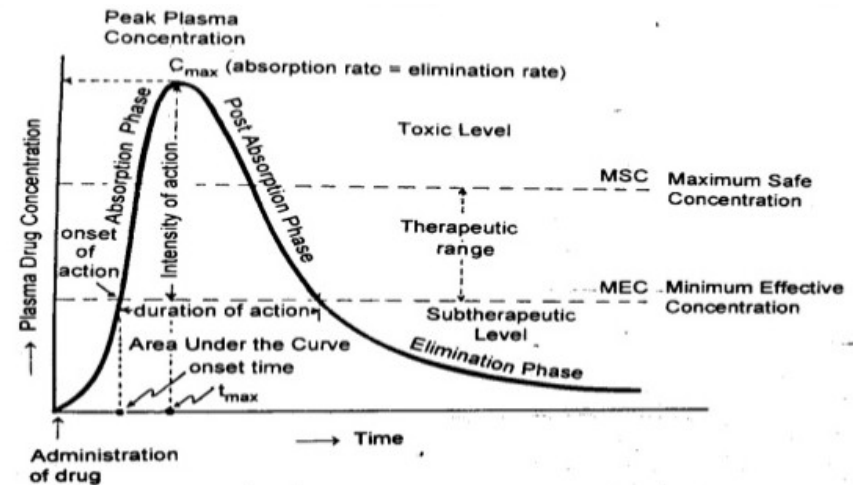
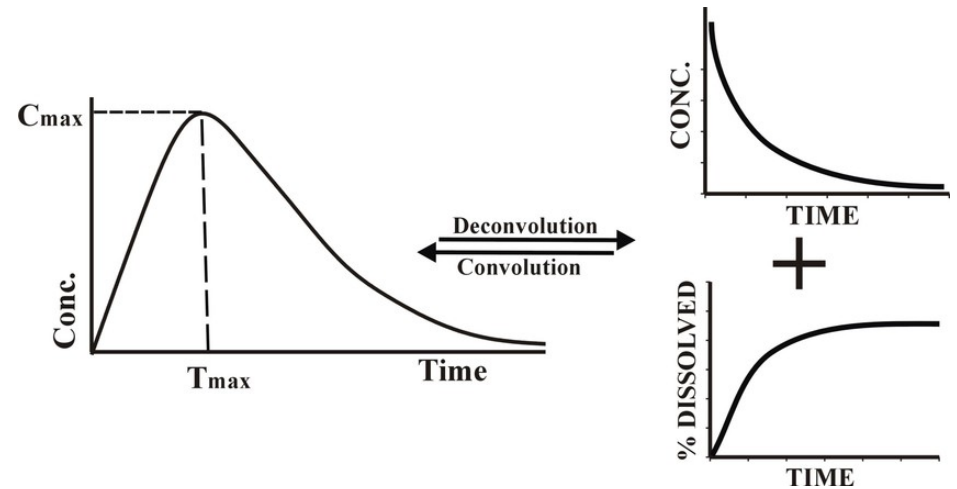


Fig. 9.1 A typical plasma concentration-time profile showing pharmacokinetic and pharmacodynamic parameters, obtained after oral administration of single dose of a drug.

<https://image.slidesharecdn.com/pharmacokineticmodels-140930004231-phapp01/95/pharmacokinetic-models-8-638.jpg?cb=1412037860>



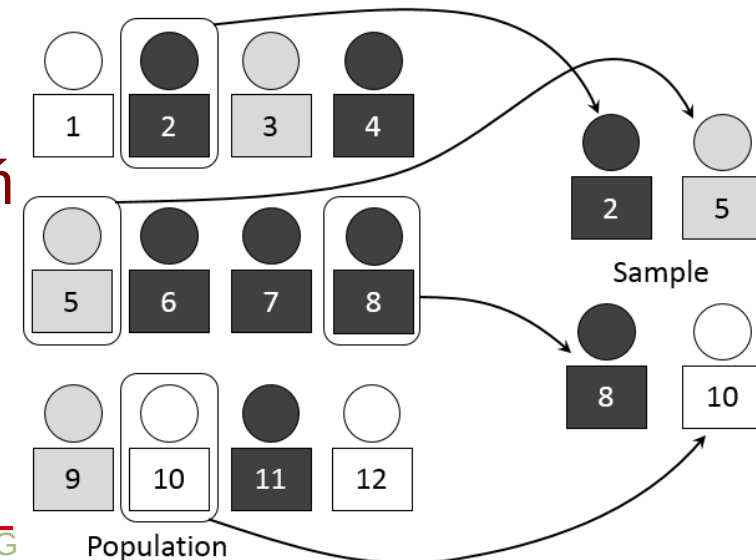


# Pobieranie próby

# Pobieranie próby

- W przypadku pomiarów eksperymentalnych najczęściej nie znamy rozkładu prawdopodobieństwa opisującego dany pomiar (np. parametru rozkładu Poissona w rozpadach promieniotwórczych, czy parametrów rozkładu Gaussa opisującego jakąś populację)
- Te parametry chcemy wyznaczyć doświadczalnie, nie jesteśmy jednak w stanie zebrać nieskończenie wiele pomiarów
- W konsekwencji jesteśmy zmuszeni **przybliżać rozkład gęstości za pomocą rozkładu częstości** (histogramu o skończonej liczbie wejść)
- **Próba** (*ang. sample*) nazywamy zespół doświadczeń wykonywanych w celu określenia kształtu (parametrów) poszukiwanego rozkładu:

- próba otrzymywana jest poprzez wybór elementów z (często nieskończonego) zbioru wszystkich możliwych doświadczeń (wszystkich możliwych pomiarów), zwanego **populacją generalną**
- próbę o  $n$  składnikach nazywamy próbą  $n$ -wymiarową



[https://en.wikipedia.org/wiki/Sampling\\_%28statistics%29#/media/File:Simple\\_random\\_sampling.PNG](https://en.wikipedia.org/wiki/Sampling_%28statistics%29#/media/File:Simple_random_sampling.PNG)

# Pobieranie próby

- Cała “sztuka” polega na odpowiednim wybraniu próby z populacji, by aproksymacja rozkładu gęstości była jemu jak najwierniejsza
- Załóżmy, że rozkład zmiennej losowej  $X$  opisywany jest funkcją  $f(x)$  – interesują nas wartości zmiennej  $X$  uzyskane przez poszczególne elementy próby
- Pobieramy  $l$  prób, każda o wymiarze  $n$ , i zaobserwowaliśmy następujące wartości zmiennej  $X$ :
  - 1. próba:  $X_1^{(1)}, X_2^{(1)}, \dots, X_n^{(1)}$
  - $\vdots$
  - $j$ -ta próba:  $X_1^{(j)}, X_2^{(j)}, \dots, X_n^{(j)}$
  - $\vdots$
  - $l$ -ta próba:  $X_1^{(l)}, X_2^{(l)}, \dots, X_n^{(l)}$
- Każdą próbę możemy przedstawić jako wektor ( $n$ -wymiarową zmienną losową):  $\mathbf{x}^{(j)} = (X_1^{(j)}, X_2^{(j)}, \dots, X_n^{(j)})$
- Wektor ma rozkład gęstości prawdopodobieństwa:
$$g(\mathbf{x}) = g(x_1, x_2, \dots, x_n)$$

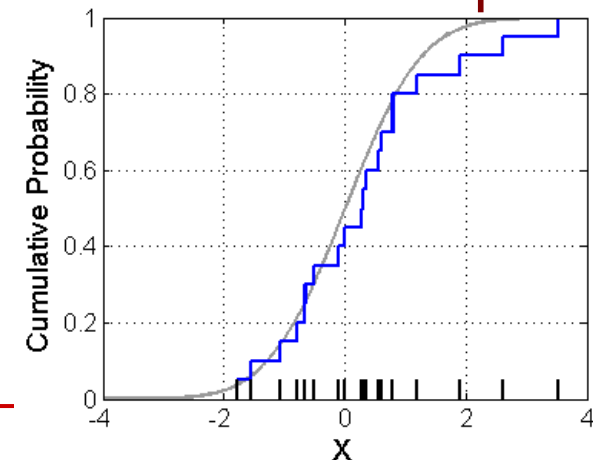
# Pobieranie próby

- Aby można było mówić o losowym pobieraniu próby:
  - zmienne  $X_i$  muszą być niezależne, czyli:  $g(\mathbf{x}) = g_1(x_1)g_1(x_2)\dots g_n(x_n)$
  - poszczególne rozkłady muszą być jednakowe i identyczne z rozkładem gęstości populacji:  $g_1(x_1) = g_2(x_2) = \dots = g_n(x_n) = f(x)$
- Należy podkreślić, że w rzeczywistym procesie pobierania próby często bardzo trudno jest zapewnić pełną losowość – nie ma tutaj jednej recepty jak to zrobić (należy starać się spełnić powyższe warunki)
- Teraz zdefiniujemy pojęcia, które charakteryzują próbę losową

# Pobieranie próby

- Teraz zdefiniujemy pojęcia, które charakteryzują próbę losową:
  - założmy, że mamy  $n$ -elementową próbę i odkładamy wyniki na osi liczb. Przez  $n_x$  oznaczmy taką liczbę wartości, które są mniejsze niż pewna stała  $x$ , czyli mamy spełnioną definicję dystrybuanty:  $X \leq x$
  - wielkość  $W_n(x) = n_x/n$  nazywamy **dystrybuantą empiryczną**
  - jest to funkcja schodkowa zwiększająca się o  $1/n$  dla każdej kolejnej wartości z próby; dla dużych  $n$  dąży do dystrybuanty
  - funkcję elementów próby (czyli zmiennej losowej  $X$ ) nazywamy **statystyką**
  - najważniejszym przykładem statystyki jest **średnia z próby** (*ang. sample mean*) zdefiniowana jako średnia z elementów próby:

$$\bar{X} = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$$



[https://upload.wikimedia.org/wikipedia/commons/1/17/Empirical\\_CDF.png](https://upload.wikimedia.org/wikipedia/commons/1/17/Empirical_CDF.png)

# Pobieranie próby – przykład

- Przykład – wzrost Polaków
- Niewątpliwie wzrost Polaków (zmienna losowa  $X$ ) podlega pewnemu rozkładowi  $f(x)$  z dystrybuantą  $F(x)$
- Pomiar wzrostu pojedynczego Polaka daje wartość  $x$
- Jeżeli stworzymy  $n$ -wymiarową próbę losową, tzn. wybierzemy  $n$  Polaków, to rozkład prawdopodobieństwa wyboru dla każdej z osób (od  $g_1(x_1)$  do  $g_n(x_n)$ ) jest taki sam jak dla całej populacji i równy  $f(x)$
- Dla każdej tak skonstruowanej próby możemy teraz policzyć jej  $W_n(x)$ . Oczywiście im większe będzie  $n$ , im więcej ludzi weźmiemy do naszej próby, tym rozkład wyliczony z próby będzie bliższy rozkładowi rzeczywiście istniejącemu w populacji
- **Zadaniem estymacji** jest znalezienie takiej statystyki (a więc funkcji określonej na wektorze  $X$ ), aby najlepiej przybliżała ona rzeczywistą wartość parametru opisującego rzeczywisty rozkład zmiennej losowej  $X$

# Estymatory

- Typowy problem analizy danych: znamy (np. z prawa fizycznego) ogólną postać gęstości prawdopodobieństwa w danej populacji, należy “jedynie” wyznaczyć parametry tego rozkładu. Przykład:
  - mierzymy rozpad radioaktywny w czasie:  $N(t) = N_0(1 - \exp(-\lambda t))$
  - parametr  $\lambda$  wyznaczamy na podstawie próby – mierząc skończoną ilość razy ilość rozpadów w czasie → wynik nigdy nie będzie dokładny, bo próba jest skończona, mamy problem **estymacji parametrów**
  - poszukiwana wielkość uzyskiwana jest funkcją elementów próby (**statystyką**) i jest nazywana **estymatorem**:  $S = S(X_1, X_2, \dots, X_n)$
  - estymator jest **nieobciążony**, jeżeli niezależnie od liczebności próby jego wartość oczekiwana jest równa wartości estymowanego parametru:

$$E(S(X_1, X_2, \dots, X_n)) = \lambda, \text{ dla każdego } n$$

- estymator jest **zgodny**, jeżeli jego wariancja znika:

$$\lim_{n \rightarrow \infty} \sigma(S(X_1, X_2, \dots, X_n)) = 0$$

# Estymatory - wartość oczekiwana

- **Wartość średnia** ze wszystkich elementów próby jest zmienną losową (jest funkcją zmiennych losowych). Jej wartość oczekiwana (tej średniej):

$$E(\bar{X}) = \frac{1}{n} (E(X_1) + E(X_2) + \dots + E(X_n)) = E(X) = \hat{x}, \text{ dla każdego } n$$

- Wniosek: **wartość średnia (arytmetyczna) z próby to estymator nieobciążony wartości oczekiwanej** zmiennej  $X$  w populacji
- Możemy obliczyć wariancję wartości średniej:

$$\begin{aligned} \sigma^2(\bar{X}) &= E\{\bar{X} - E(\bar{X})\}^2 = E\left\{\left(\frac{X_1 + X_2 + \dots + X_n}{n} - \hat{x}\right)^2\right\} \\ &= \frac{1}{n^2} E\{[(X_1 - \hat{x}) + (X_2 - \hat{x}) + \dots + (X_n - \hat{x})]^2\} \end{aligned}$$

- Z uwagi na niezależność zmiennych kowariancje między zmiennymi  $X_i$  znikają, czyli ostatecznie:

$$\sigma^2(\bar{X}) = \frac{1}{n} \sigma^2(X) \quad \lim_{n \rightarrow \infty} \sigma^2(\bar{X}) = 0$$

- Wniosek: **wartość średnia (arytmetyczna) z próby jest również estymatorem zgodnym wartości oczekiwanej**



# Estymatory - wariancja

- Jak pamiętamy z definicji wariancji, nie jest ona zmienną losową
- Możemy wariancję przybliżyć przez średnią arytmetyczną odchyłeń kwadratowych od wartości średniej:

$$S'^2(X) = \frac{1}{n} \left( (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2 \right)$$

- Wartość oczekiwana tej wielkości:

$$\begin{aligned} E(S'^2(X)) &= \frac{1}{n} E \left\{ \sum_{i=1}^n (X_i - \bar{X})^2 \right\} = \frac{1}{n} E \left\{ \sum_{i=1}^n (X_i - \hat{x} + \hat{x} - \bar{X})^2 \right\} \\ &= \frac{1}{n} E \left\{ \sum_{i=1}^n (X_i - \hat{x})^2 + \sum_{i=1}^n (\hat{x} - \bar{X})^2 + 2 \sum_{i=1}^n (X_i - \hat{x})(\hat{x} - \bar{X}) \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ E((X_i - \hat{x})^2) - E((\bar{X} - \hat{x})^2) \right\} = \frac{1}{n} \left\{ n \sigma^2(X) - n \left( \frac{1}{n} \sigma^2(X) \right) \right\} \\ &= \frac{n-1}{n} \sigma^2(X) \end{aligned}$$

- Widać więc, że  $S'^2$  jest **estymatorem obciążonym** dla wariancji populacji mającym wartość oczekiwaną mniejszą niż  $\sigma^2(X)$

# Estymatory - wariancja

- Możemy jednak nieznacznie zmodyfikować definicję wariancji z próby i wprowadzić estymator:

$$s^2(X) = \frac{1}{n-1} \left( (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2 \right)$$

- Otrzymujemy **estymator nieobciążony wariancji populacji**

- Jeśli podstawimy ten wzór do wzoru:  $\sigma^2(\bar{X}) = \frac{1}{n} \sigma^2(X)$

- To otrzymamy **estymator wariancji wartości średniej**:

$$s^2(\bar{X}) = \frac{1}{n} s^2(X) = \frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2$$

- Zaś odpowiadające odchylenie standardowe (**niepewność średniej z próby**):

$$\Delta \bar{X} = \sqrt{s^2(\bar{X})} = s(\bar{X}) = \frac{1}{\sqrt{n}} s(X)$$

- Jaka jest zaś **niepewność wariancji z próby** (bez wyprowadzenia)?

$$\Delta S^2 = S^2 \sqrt{\frac{2}{n-1}}$$

- Odchylenie standardowe próby:  $S = \sqrt{S^2} = \frac{1}{\sqrt{n-1}} \sqrt{\sum_{i=1}^n (X_i - \hat{X})^2}$

# Estymatory - wariancja

- Podsumowując zatem **estymatory nieobciążone**:

- wartości oczekiwanej populacji → średnia z próby (**wynik doświadczenia**):

$$\bar{X} = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$$

- wariancji populacji – wariancja z próby (aproksymowana):

$$S^2(X) = \frac{1}{n-1} ((X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2)$$

- wariancji wartości średniej z próby (**patrz niepewność typu A**):

$$S^2(\bar{X}) = \frac{1}{n} S^2(X) = \frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2$$

- wariancji (aproksymowanej) wariancji z próby

$$\text{Var}(S^2) = S^4 \left( \frac{2}{n-1} \right)$$

- odchylenia standardowego wartości średniej z próby:

$$S = \sqrt{S^2(X)} = \frac{1}{\sqrt{n-1}} \sqrt{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}$$

- dalej możemy wyznaczać np. wariancję odchylenia std. próby i tak dalej (w nieskończoność)...



**KONIEC**