

Rachunek prawdopodobieństwa jest dyscypliną matematyczną o cechach pokrewnych do tych jakie mają na przykład geometria i mechanika teoretyczna. W każdej dyscyplinie musimy starannie wyróżnić trzy aspekty teorii:

- formalną treść logiczną,
- podłoże intuicyjne,
- zastosowania.

**Formalna treść logiczna.** Z aksjomatycznego punktu widzenia matematyka zajmuje się jedynie relacjami pomiędzy obiektami niezdefiniowanymi. Na przykład geometria nie troszczy się o to czym są „w rzeczywistości” punkt i linia prosta. W geometrii są to pojęcia pierwotne (a więc niezdefiniowane) a pewniki (aksjomaty) geometrii ustalają relacje między nimi: np. dwa punkty wyznaczają jedną prostą. Podobnie grę w szachy można jedynie określić poprzez podanie zbioru przepisów i nie ma sensu mówić o definicji czy prawdziwej naturze pionka.

**Podłoże intuicyjne.** W przeciwieństwie do szachów aksjomaty geometrii i mechaniki dotyczą istniejącego podłoża intuicyjnego. Intuicja geometryczna jest na tyle silna, że ma skłonność wybiegać przed logiczne rozumowanie. Wiadomo, że intuicję można ćwiczyć i rozwijać. Początkujący szachista porusza się ostrożnie przypominając sobie ruchy, podczas gdy doświadczony gracz ocenia sytuację na jeden rzut oka i często nie potrafi racjonalnie uzasadnić swojej intuicji. Podobnie intuicje matematyczne rozwija się wraz z doświadczeniem i można wyrobić sobie poczucie dla takich pojęć jak na przykład przestrzeń czterowymiarowa,  $n$ -wymiarowa, nieskończenie wymiarowa. Wydaje się, że zbiorowa intuicja ludzkości czyni pewne postępy. Na przykład wprowadzone przez Newtona pojęcie pola sił, czy wprowadzona przez Maxwella koncepcja fal elektromagnetycznych nie budzą dzisiaj żadnych emocji a wprowadzane były jako „nie dające się pomyśleć” i „sprzeczne z intuicją”. Podobnie współczesny student nie docenia trudności z którymi musiała borykać się teoria prawdopodobieństwa w początkowej fazie swego rozwoju. Dzisiaj, gdy środki masowego przekazu zasypują nas wynikami badań opinii publicznej każdy przyswoił sobie intuicyjne pojęcie prawdopodobieństwa.

**Zastosowania.** Pojęcia geometrii i mechaniki utożsamia się zazwyczaj z pewnymi obiektami fizycznymi. Pojęcie ciała sztywnego jest jednym z podstawowych pojęć mechaniki pomimo, że nie istnieją w praktyce ciała sztywne. W zależności od celu teorii zaniedbujemy czasami strukturę atomową materii i traktujemy słońce jako kulę materii ciągłej, a czasem jako punkt materialny. W zastosowaniach abstrakcyjne modele matematyczne służą jako środki opisu rzeczywistości i różne modele mogą opisywać tę samą sytuację empiryczną.

Pojęciem pierwotnym w rachunku prawdopodobieństwa jest: przestrzeń zdarzeń elementarnych  $\Omega$ , której elementy nazywamy zdarzeniami elementarnymi oznaczając je przez  $\omega$ . W zagadnieniach

praktycznych przez przestrzeń zdarzeń elementarnych rozumie się zbiór niepodzielnych(nierozkładalnych) wyników eksperymentu czy obserwacji.

**Przykład (ToTo).** Z 49 liczb losujemy 6 liczb.  $\Omega = \{\omega = \{\omega_1, \dots, \omega_6\} : 1 \leq \omega_i \leq 49, i \neq j \Rightarrow \omega_i \neq \omega_j\}$ .

**Przykład.** Niech T oznacza spóźnienie określonego studenta na wykład trwający 90 min. Wówczas  $\Omega = [0, 90]$ . Dla dwóch studentów  $\Omega = [0, 90] \times [0, 90]$ .

Podzbiory danej przestrzeni zdarzeń elementarnych będziemy traktować jako zdarzenia losowe. Należy jednak zaznaczyć, że nie każdy podzbiór przestrzeni  $\Omega$  może być traktowany jako zdarzenia losowe, gdyż nie dla każdego podzbioru zbioru  $\Omega$  można określić sensownie jego prawdopodobieństwo. Rozkład prawdopodobieństwa będący rzeczywistą funkcją zbioru podobnie jak pole czy objętość musi mieć swoją dziedzinę, którą będzie właśnie rodzina zdarzeń losowych.

**Formalizacja.** Niech  $\Omega$  będzie przestrzenią zdarzeń elementarnych. Przez  $2^\Omega$  oznaczamy rodzinę wszystkich podzbiorów właściwych i niewłaściwych przestrzeni  $\Omega$ .

**Def.** Rodzinę  $\mathcal{S} \subset 2^\Omega$  podzbiorów przestrzeni  $\Omega$  nazywamy  $\sigma$ -ciałem jeżeli spełnia ona warunki:

$$(S_1) \quad \Omega \in \mathcal{S}$$

$$(S_2) \quad A \in \mathcal{S} \Rightarrow \bar{A} = \Omega - A \in \mathcal{S}$$

$$(S_3) \quad A_1, A_2, \dots \in \mathcal{S} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{S}$$

Z definicji widać, że  $\sigma$ -ciało jest niepustą rodziną podzbiorów zamkniętą ze względu na dopełnienie i przeliczalne sumowanie. Z praw de Morgana wynika, że jest to rodzina zamknięta ze względu na przeliczalne iloczyny. Oczywiście zbiór  $\emptyset$  również należy do  $\sigma$ -ciała.

Parę  $(\Omega, \mathcal{S})$  nazywamy przestrzenią mierzalną.

**Def.** Rozkładem prawdopodobieństwa w przestrzeni mierzalnej  $(\Omega, \mathcal{S})$  nazywamy rzeczywistą funkcję

$P$  określoną na  $\sigma$ -ciele  $\mathcal{S}$  czyli  $P: \mathcal{S} \rightarrow \mathbb{R}$  spełniającą następujące warunki (tzw. postulaty

Kołmogorowa):

$$(P_1) \quad \forall A \in \mathcal{S} \quad P(A) \geq 0$$

$$(P_2) \quad P(\Omega) = 1$$

$$(P_3) \quad \forall A_1, A_2, \dots \in \mathcal{S} \quad (A_i \cap A_j = \emptyset \text{ dla } i \neq j) \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Warunek  $P_3$  jest zwany warunkiem przeliczalnej addytywności funkcji zbioru  $P$ .

Trójkę  $(\Omega, \mathcal{S}, P)$  gdzie  $\Omega$  jest przestrzenią zdarzeń elementarnych,  $\mathcal{S}$   $\sigma$ -ciałem podzbiorów przestrzeni  $\Omega$  (spełniającym  $\mathbf{S}_1$ ,  $\mathbf{S}_2$  i  $\mathbf{S}_3$ ) a  $P$  rozkładem prawdopodobieństwa na rodzinie  $\mathcal{S}$  zdarzeń losowych (spełniającym  $\mathbf{P}_1$ ,  $\mathbf{P}_2$  i  $\mathbf{P}_3$ ) nazywamy **przestrzenią probabilistyczną**.

### Ogólne własności rozkładu prawdopodobieństwa w przestrzeniach probabilistycznych.

1.  $P(\emptyset) = 0$
2.  $\forall A, B \in \mathcal{S} \quad A \subset B \Rightarrow P(A) \leq P(B)$  (monotoniczność miary)
3.  $\forall A \in \mathcal{S} \quad P(\bar{A}) = 1 - P(A)$
4.  $\forall A, B \in \mathcal{S} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5.  $\forall A_1, \dots, A_n \in \mathcal{S} \quad P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) +$   
 $+ \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n+1} P(A_1 \cap A_2 \dots \cap A_n)$  (wzór włączeń i wyłączeń)
6.  $\forall A_1, \dots, A_n \in \mathcal{S} \quad P\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j)$  (oszacowanie od dołu)
7.  $\forall A_1, \dots, A_n \in \mathcal{S} \quad P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$  (skończona subaddytywność)
8.  $\forall A_1 \subset A_2 \subset A_3, \dots \in \mathcal{S} \quad P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n)$  (ciągłość z dołu)
9.  $\forall A_1 \supset A_2 \supset A_3, \dots \in \mathcal{S} \quad P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n)$  (ciągłość z góry)
10.  $\forall A_1, A_2, \dots \in \mathcal{S} \quad P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$  (przeliczalna subaddytywność)
11.  $\forall i \quad P(A_i) = 0 \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = 0$
12.  $\forall i \quad P(A_i) = 1 \Rightarrow P\left(\bigcap_{i=1}^{\infty} A_i\right) = 1$

Ad. 8 Definiujemy nowy ciąg zbiorów rozłącznych  $B_1=A_1, B_n=A_n - A_{n-1}$ . Widać, że  $A_n = \bigcup_{i=1}^n B_i$ . Stąd

$$P(A_n) = \sum_{i=1}^n P(B_i). \text{ Wobec tego } P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) = \lim_{n \rightarrow \infty} P(A_n).$$

Ad. 9 Jest to bezpośredni wniosek z 1, 5 i praw de Morgana.

**Uwaga.** Własności 8 i 9 nazywamy własnościami ciągłości miar probabilistycznych (rozkładów prawdopodobieństwa

### Przykłady $\sigma$ -ciał.

- $S = 2^\Omega$
- $S = \{\emptyset, \Omega\}$
- Iloczyn mnogościowy dowolnej ilości  $\sigma$ -ciał jest  $\sigma$ -ciałem (Łojasiewicz str. 73)
- $\sigma$ -ciałem generowanym przez rodzinę zbiorów  $A$  nazywamy najmniejsze  $\sigma$ -ciało zawierające rodzinę zbiorów  $A$ . Jest to część wspólna wszystkich  $\sigma$ -ciał zawierających rodzinę  $A$ . Oznaczamy to  $\sigma$ -ciało przez  $\sigma(A)$ .
- $\sigma$ -ciało generowane przez podział  $\Omega = \bigcup_{i=1}^{\infty} B_i$ ,  $B_i \cap B_j = \emptyset$  dla  $i \neq j$  to  $\sigma$ -ciało, którego elementami są sumy co najwyżej przeliczalne zbiorów rodziny  $\{B_i\}_{i=1}^{\infty}$ .
- **Zadanie.** Niech  $A \subset \Omega$  i  $B \subset \Omega$ . Znaleźć  $\sigma(\{A, B\})$ . Widać, że  $\sigma(\{A, B\}) = \sigma(\{A \cap B, A - B, B - A, \overline{A \cup B}\})$  i zbiory  $A \cap B, A - B, B - A, \overline{A \cup B}$  stanowią podział. Tworząc wszystkie możliwe sumy zbiorów  $A \cap B, A - B, B - A, \overline{A \cup B}$  (łącznie ze zbiorem pustym) otrzymamy 16 elementowe  $\sigma$ -ciało.
- Niech  $(X, d)$  będzie przestrzenią metryczną. Borelowskim  $\sigma$ -ciałem  $\mathcal{B}(X)$  nazywamy  $\sigma$ -ciało generowane przez rodzinę zbiorów otwartych. W  $R^n$   $\sigma$ -ciało zbiorów borelowskich jest identyczne z  $\sigma$ -ciałem generowanym przez rodzinę przedziałów, bo każdy zbiór otwarty w  $R^n$  jest przeliczalną sumą przedziałów domkniętych (mianowicie wszystkich przedziałów domkniętych zawartych w danym zbiorze otwartym o wierzchołkach wymiernych).  
Każdy zbiór liniowy (w  $R$ ) otwarty może być przedstawiony jako suma skończonej lub przeliczalnej ilości przedziałów otwartych rozłącznych i przedstawienie to jest jednoznaczne z dokładnością do kolejności zbiorów- Hartman, Mikusiński str. 11.  
Każdy zbiór otwarty  $R^n$  można przedstawić jako przeliczalną sumę przedziałów domkniętych o rozłącznych wnętrzach Hartman, Mikusiński str. 112.
- Problem konstruowania  $\sigma$ -ciał od wewnątrz (**odpowiedź negatywna**) poprzez kolejne dołączanie przeliczalnych sum i ich dopełnień - Billingsley-str. 39.

- Niech  $(X_1, \mathcal{S}_1)$  będzie przestrzenią mierzalną a  $f: X_1 \rightarrow X_2$  odwzorowaniem zbioru  $X_1$  na  $X_2$ . Funkcja  $f$  indukuje w  $X_2$   $\sigma$ -ciało  $\mathcal{S}_2$  składające się z podzbiorów  $X_2$ , których przeciwobrazy należą do  $\mathcal{S}_1$ . Jeżeli  $B \subset X_2$ , to  $f^{-1}(B) = \{x \in X_1 : f(x) \in B\} \subset X_1$ . Wiadomo, że przeciwobraz zachowuje wszystkie operacje teoriomnogościowe np.  $f^{-1}(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} f^{-1}(A_i)$ .

### Przykłady konstrukcji przestrzeni probabilistycznej.

- I.  $\Omega = \{\omega_1, \omega_2, \dots\}$  jest zbiorem co najwyżej przeliczalnym. Jako rodzinę zdarzeń losowych przyjmujemy rodzinę wszystkich możliwych podzbiorów zbioru  $\Omega$ , czyli  $\mathcal{S} = 2^\Omega = \mathcal{P}(\Omega)$ . Rozkład prawdopodobieństwa jest w pełni określony przez podanie prawdopodobieństw dla wszystkich zdarzeń elementarnych:

$$P(\{\omega_i\}) = p_i \geq 0, \quad \sum_{i=1}^{\infty} p_i = 1. \quad \text{Wówczas } P(A) = \sum_{\{i: \omega_i \in A\}} p_i$$

Przypadek szczególny :  $\Omega = \{\omega_1, \dots, \omega_N\}$ ,

$$\forall i : P(\{\omega_i\}) = \frac{1}{N},$$

$$A \in \Sigma = 2^\Omega \quad P(A) = \frac{\#A}{\#\Omega}$$

prowadzi do klasycznej definicji (Laplace'a) rozkładu prawdopodobieństwa.

- II. Niech  $\Omega$  będzie dowolnym niepustym zbiorem a  $\omega_0 \in \Omega$  dowolnym punktem zwanym atomem. Na rodzinie  $\mathcal{S} = 2^\Omega$  określamy rozkład prawdopodobieństwa następującym wzorem

$$P(A) = \begin{cases} 1, & \text{gdy } \omega_0 \in A \\ 0, & \text{gdy } \omega_0 \notin A \end{cases}.$$

Rozkład ten nazywamy atomową miarą Diraca skupioną w punkcie  $\omega_0$  i oznaczamy  $\delta_{\omega_0}(\cdot)$ .

- III. Niech  $\Omega$  będzie dowolnym zbiorem z wyróżnionym ciągiem punktów  $\omega_1, \omega_2, \dots$  zwanych atomami. Z ciągiem tym kojarzymy ciąg liczbowy  $p_i \geq 0$ ,  $\sum_{i=1}^{\infty} p_i = 1$ . Na rodzinie  $\mathcal{S} = 2^\Omega$  określamy rozkład prawdopodobieństwa następującym wzorem  $P(A) = \sum_{i=1}^{\infty} p_i \delta_{\omega_i}(A)$ . Miarą zbioru jest więc suma miar

atomów wchodzących w skład zbioru. Funkcję zbioru  $\sum_{i=1}^{\infty} \delta_{o_i}(\cdot)$  nazywamy miarą liczącą. Miara licząca nie jest oczywiście miarą probabilistyczną. Miara licząca danego zbioru jest równa liczbie atomów wchodzących w skład zbioru i nie jest oczywiście miarą skończoną.

IV.  $\Omega=R$ . W tym przypadku nie możemy przyjąć za rodzinę zdarzeń losowych rodziny wszystkich podzbiorów, gdyż na tak szerokiej klasie zbiorów nie da się skonstruować rozkładu prawdopodobieństwa o pożądanym własnościach np. nie posiadającego atomów. Jako rodzinę zdarzeń losowych  $\mathcal{S}$  przyjmujemy rodzinę  $\mathcal{A}(R)$  zbiorów borelowskich, czyli  $\sigma$ -ciało generowane przez rodzinę zbiorów otwartych w  $R$ . Jest to również  $\sigma$ -ciało generowane przez rodzinę przedziałów postaci  $[a, b)$ . Określimy więc najpierw rozkład prawdopodobieństwa na rodzinie przedziałów (nie jest ona  $\sigma$ -ciałem) za pomocą pomocniczej funkcji punktu zwanej dystrybuantą a następnie rozszerzymy skonstruowany rozkład na  $\mathcal{A}(R)$ .

**Def.** Rzeczywistą funkcję  $F$  zmiennej rzeczywistej nazywamy **dystrybuantą**, jeżeli spełnia ona warunki:

1.  $F$  jest niemalejąca,
2.  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow \infty} F(x) = 1$ ,
3.  $F$  jest lewostronnie ciągła.

Mając dystrybuantę określamy prawdopodobieństwo (miarę) przedziału wzorem  $P([a,b))=F(b)-F(a)$  a następnie przedłużamy zdefiniowaną funkcję na tzw.  $\sigma$ -ciało zbiorów mierzalnych zawierające  $\mathcal{A}(R)$  za pomocą :

- ogólnej procedury rozszerzania miary Caratheodory'ego (Łojasiewicz str 93, Dziubiński str. 214 Billingsley str. 168) - która w dużym skrócie polega na zdefiniowaniu za pomocą funkcji zbioru  $P$  miary zewnętrznej  $P^*$  na rodzinie wszystkich podzbiorów i zdefiniowaniu  $\sigma$ -ciała z wykorzystaniem  $P^*$ . Miara  $P^*$  obcięta do skonstruowanego  $\sigma$ -ciała jest już rozkładem prawdopodobieństwa (zupełnym). Miarę tę można obciąć do  $\sigma$ -ciała  $\mathcal{A}(R)$  i uzyskać poszukiwane jednoznaczne rozszerzenie miary  $P$  na  $\mathcal{A}(R)$ . Miara ta nie jest zupełna, bo podzbiór zbioru borelowskiego miary 0 nie musi być zbiorem borelowskim. Można ją uzupełnić standardową procedurą uzupełniania miary.
- rozszerzyć  $P$  na klasę zbiorów otwartych wykorzystując możliwość przedstawienia zbiorów otwartych w postaci przeliczalnych sum przedziałów a następnie zdefiniować klasę zbiorów mierzalnych w następujący sposób: zbiór  $A$  jest  $P$  mierzalny jeżeli dla każdego  $\varepsilon$  istnieją dwa zbiory otwarte  $U$  i  $V$  takie, że  $A \subset U$ ,  $U-A \subset V$  i  $P(V) < \varepsilon$ . Zdefiniowana klasa zbiorów mierzalnych zawiera w sobie  $\mathcal{A}(R)$ . Dalsze postępowanie jest identyczne jak w powyższym przypadku.

**Uwaga.** Ponieważ  $\{x_0\} = \bigcap_{n=1}^{\infty} [x_0, x_0 + \frac{1}{n})$  z ciągłości miary probabilistycznej wynika, że

$$P(\{x_0\}) = \lim_{n \rightarrow \infty} P([x_0, x_0 + \frac{1}{n})) = \lim_{n \rightarrow \infty} (F(x_0 + \frac{1}{n}) - F(x_0)) = F(x_0^+) - F(x_0).$$

Jeżeli  $x_0$  jest punktem ciągłości dystrybuanty  $F$ , to  $P(\{x_0\})=0$ . Z przeliczalnej addytywności wynika, że jeżeli  $F$  jest funkcją ciągłą, to prawdopodobieństwo dla dowolnego zbioru przeliczalnego np. dla zbioru liczb wymiernych jest równe 0. W związku z powyższym **z faktu, że prawdopodobieństwo zdarzenia jest równe 0 nie wynika, że jest to zdarzenie niemożliwe.**

Jeżeli dystrybuanta jest funkcją ciągłą, to odpowiadający jej rozkład prawdopodobieństwa nazywamy rozkładem ciągłym. Jeżeli dystrybuanta spełnia mocniejszy warunek tzw. **absolutnej ciągłości** to istnieje jednoznacznie (z dokładnością do zbiorów miary 0) wyznaczona nieujemna

funkcja  $f \geq 0$  zwana funkcją gęstości i  $F(x) = \int_{-\infty}^x f(t)dt$ . W dalszych rozważaniach przez rozkłady

ciągłe rozumiemy rozkłady absolutnie ciągłe (w odróżnieniu od rozkładów osobliwych).

Rozkłady takie na prostej zadane są więc funkcją gęstości w następujący sposób:  $P(A) = \int_A f(x)dx$

### Dystrybuanta osobliwa Cantora

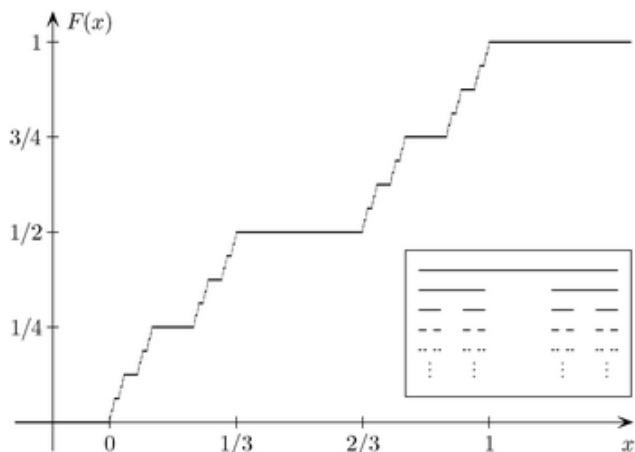
Przedstawmy przedział  $[0,1]$  w postaci  $[0,1] = A \cup B$ , gdzie

$$A = (\frac{1}{3}, \frac{2}{3}) \cup [(\frac{1}{9}, \frac{2}{9}) \cup (\frac{7}{9}, \frac{8}{9})] \cup [(\frac{1}{27}, \frac{2}{27}) \cup (\frac{7}{27}, \frac{8}{27}) \cup (\frac{19}{27}, \frac{20}{27}) \cup (\frac{25}{27}, \frac{26}{27})] \cup \dots \cup [(\frac{1}{3^n}, \frac{2}{3^n}) \cup (\frac{7}{3^n}, \frac{8}{3^n}) \cup \dots \cup (\frac{3^n-2}{3^n}, \frac{3^n-1}{3^n})] \cup \dots = A_{11} \cup [A_{21} \cup A_{22}] \cup [A_{31} \cup A_{32} \cup A_{33} \cup A_{34}] \cup \dots \cup [A_{n1} \cup A_{n2} \cup \dots \cup A_{n2^{n-1}}] \cup \dots$$

$$B = [0,1] - A$$

Określmy na zbiorze  $R - B$  funkcję  $F(x) = \begin{cases} 0, & \text{dla } x \leq 0 \\ \frac{2k-1}{2^n}, & \text{dla } x \in A_{nk} \\ 1, & \text{dla } x \geq 1 \end{cases}$ .

Na zbiorze  $B$  (czyli dla  $t \in B$ ) funkcję  $F$  określamy wzorem  $F(t) = \sup_{x < t, x \in A} F(x)$ .



Tak określona funkcja jest dystrybuantą ciągłą i jej pochodna jest równa 0 dla każdego  $x \in A$ , czyli prawie wszędzie.

Oczywiście  $F(x)=P((-\infty,x))$ . Jeżeli rozkład prawdopodobieństwa na prostej jest czysto atomowy to odpowiada mu dystrybuanta schodkowa o skokach zlokalizowanych w atomach. Wartości skoków są równe prawdopodobieństwom przypisanym poszczególnym atomom.

**Tw. Lebesgue'a.** Każdą dystrybuantę można przedstawić w postaci  $F=c_1F_{\text{dyskr.}}+c_2F_{\text{a. cg.}}+c_3F_{\text{osobl.}}$ ,  
 $c_i \geq 0, c_1+c_2+c_3=1$ .

**Przykład.** Prom kursuje pomiędzy przystaniami  $A$  i  $B$  odległymi o  $k$  km. Prawdopodobieństwo znajdowania się promu w  $A$  jest równe 0.1 a w  $B$  równe 0.2. Prom płynie ze stałą prędkością pomiędzy  $A$  i  $B$ . Niech  $X$  oznacza odległość promu od  $A$ . Znaleźć rozkład (dystrybuantę) i dokonać dekompozycji na część ciągłą i dyskretną.

V.  $\Omega=R^n$ . W tym przypadku podobnie jak w poprzednim jako rodzinę zdarzeń losowych  $\mathcal{S}$  przyjmujemy

rodzinę  $\mathcal{A}(R^n)$  zbiorów borelowskich.  $\mathcal{A}(R^n)$  generowane przez rodzinę zbiorów otwartych jest również generowane przez rodzinę przedziałów w  $R^n$  (np. postaci  $[a,b)$ ). Rozkład prawdopodobieństwa na rodzinie wielowymiarowych przedziałów określamy za pomocą dystrybuanty a następnie rozszerzymy skonstruowany rozkład na  $\mathcal{A}(R^n)$ . Z uwagi na prostotę ograniczymy dalsze rozważania do  $n=2$ .

**Def.** Rzeczywistą funkcję  $F(x,y)$  zmiennych rzeczywistych  $x,y$  nazywamy **dystrybuantą** jeżeli spełnia ona warunki:

1.  $F$  jest niemalejąca ze względu na każdą ze zmiennych z osobna,
2.  $\forall y \lim_{x \rightarrow -\infty} F(x,y) = 0, \forall x \lim_{y \rightarrow -\infty} F(x,y) = 0, \lim_{\substack{x \rightarrow \infty \\ y \rightarrow \infty}} F(x,y) = 1,$
3.  $F$  jest lewostronnie ciągła ze względu na każdą ze zmiennych z osobna,
4.  $\forall x_1 \leq x_2, y_1 \leq y_2 \quad F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1) \geq 0.$

Mając dystrybuantę definiujemy  $P([x_1,x_2) \times [y_1,y_2)) = F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)$  a następnie rozszerzamy rozkład na  $\mathcal{A}(R^2)$  tak jak w poprzednim przypadku.

**Uwaga:** Funkcja  $F(x_1,x_2) = \begin{cases} 1, & \text{dla } x_1, x_2 \geq 0, \max(x_1, x_2) \geq 1 \\ 0, & \text{w pozostałych przypadkach} \end{cases}$  spełnia 1,2 i 3 a nie spełnia 4.

Wówczas np.  $P([0.5, 2) \times [0.5, 2)) = 1-1-1+0 = -1$  - sprzeczność. Warunek 4 jest więc istotny. Jeżeli dystrybuanta jest absolutnie ciągła, to  $P(A) = \iint_A f(x,y) dx dy$



## Prawdopodobieństwo geometryczne- uwagi - przykłady

1. Obliczyć prawdopodobieństwo, że pierwiastki równania  $x^2 + 2ax + b = 0$  są rzeczywiste dodatnie, jeżeli  $(a,b)$  jest losowo wybranym punktem prostokąta  $\{(a,b): |a| < 2, |b| < 1\}$ .
2. Kawalek drutu o długości 20 cm zgięto pod kątem prostym w przypadkowo wziętym punkcie. Następnie zgięto drut jeszcze w dwóch punktach, tak by utworzyła się ramka prostokątna o obwodzie 20 cm. Jakie jest prawdopodobieństwo, że pole ramki nie przekroczy  $21 \text{ cm}^2$ ?
3. Zadanie Buffona. Płaszczyznę podzielono prostymi równoległymi odległymi o  $2a$ . Na płaszczyznę tę rzucamy w sposób przypadkowy odcinek o długości  $2l < 2a$ . Jakie jest prawdopodobieństwo, że odcinek przetnie jedną z prostych?
4. Pani  $X$  i pani  $Y$  idąc z domu do biura mają do przebycia pewien wspólny odcinek drogi  $AB$  z tym, że przebywają go w przeciwnych kierunkach. Pani  $X$  przybywa do punktu  $A$  zaś pani  $Y$  do  $B$  w przypadkowym momencie czasu pomiędzy godziną  $7^{30}$  i  $7^{45}$  i idzie ze stałą prędkością. Każda z pań przechodzi odcinek  $AB$  w ciągu 5 min. Obliczyć prawdopodobieństwo p. spotkania się pań  $X$  i  $Y$ .
5. Odcinek o długości 10 cm został podzielony w sposób losowy na 3 części. Obliczyć prawdopodobieństwo, że z tych części można zbudować trójkąt.

Uzupełnienie

Niech będzie dana przestrzeń probabilistyczna  $(\Omega, \mathcal{S}, P)$  i zdarzenia losowe  $A_1, \dots, A_n \in \mathcal{S}$ . Zdarzenie

$\bigcup_{i=1}^n A_i$  oznacza, że zaszło przynajmniej jedno ze zdarzeń  $A_1, \dots, A_n$ .

Dowodzi się, że

$$(*) \quad P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n+1} P(A_1 \cap A_2 \dots \cap A_n)$$

Uwaga ; Dowód metodą indukcji matematycznej jest uciążliwy. Elegancki lub i bardzo kształcący dowód uzyskuje się metodą włączeń i wyłączeń - (zobacz Jakubowski J, Sztencel R. Rachunek prawdopodobieństwa dla prawie każdego wyd II str 43)

## Przykłady zastosowań

- **Sekretarka wkłada losowo 10 tomów akt do 3 szuflad. jakie jest prawdopodobieństwo, że co najmniej jedna szuflada będzie pusta**

Rozwiązanie I. Szuflady i akta są rozróżnialne. Numerujemy szuflady i akta. Każdej teczce z aktami przypisujemy numer szuflady do której ta teczka trafia. Oznaczmy  $\omega_k$ ,  $k=1, \dots, 10$  numer szuflady do której trafia  $k$ -ta teczka.

$$\Omega = \{\omega = (\omega_1, \dots, \omega_{10}) : \omega_k \in \{1, 2, 3\} \quad k = 1 \div 10\}.$$

Zdarzeniami elementarnymi są 10-cio elementowe wariacje z powtórzeniem ze zbioru 3 elementowego.

Oznaczmy teraz  $A_i$  -  $i$ -ta szuflada jest pusta. Zdarzenie  $\bigcup_{i=1}^3 A_i$  oznacza, że przynajmniej jedna szuflada jest pusta. Z wzoru (\*)

$$P\left(\bigcup_{i=1}^3 A_i\right) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3) = \\ = 3\left(\frac{2}{3}\right)^{10} - 3\left(\frac{1}{3}\right)^{10} + 0 = \frac{2^{10}-1}{3^9}$$

Rozwiązanie II. Szuflady są rozróżnialne a akta są nierozróżnialne.

$\Omega = \{\omega = (\omega_{(1)}, \dots, \omega_{(10)}) : \omega_{(1)} \leq \omega_{(2)} \leq \dots \leq \omega_{(10)}\}$   $\omega_k \in \{1, 2, 3\}$   $k = 1 \div 10$ . Mamy tu do czynienia z rozmieszczeniem 10-ciu nierozróżnialnych kul (akt) w 3 rozróżnialnych szufladach Zdarzeniami elementarnymi są tu 10-cio elementowe kombinacje z powtórzeniem ze zbioru 3-elementowego. Obliczymy najpierw prawdopodobieństwo zdarzenia przeciwnego, że żadna szuflada nie jest pusta. Wkładamy więc po jednej teczce do każdej szuflady a pozostałe 7 teczek rozmieszczamy dowolnie w 3 szufladach na  $C_7^{7+3-1} = C_7^9$  sposobów. Stąd  $P\left(\bigcup_{i=1}^3 A_i\right) = 1 - \frac{C_7^9}{C_{10}^{12}}$ .

- **Sekretarka wkłada losowo  $n$  listów do  $n$  kopert. Jakie jest prawdopodobieństwo, że choć jeden list dotrze do adresata.**

Tu musimy potraktować listy i koperty jako rozróżnialne. Numerujemy więc listy (kule) i koperty (szuflady). Każdemu listowi przypisujemy numer koperty do której trafia. Oznaczmy  $\omega_i, i=1, \dots, n$  numer koperty do której trafia  $i$ -ty list. Zdarzeniami elementarnymi są permutacje zbioru 10-cio elementowego  $\Omega = \{\omega = (\omega_1, \dots, \omega_{10}) : \omega_i = 1 \div 10, i = 1 \div 10, i \neq j \Rightarrow \omega_i \neq \omega_j\}$

Oznaczmy teraz  $A_i$  -  $i$ -ty list trafia do (właściwego) adresata ( $\omega_i = i$ , czyli w permutacji

$\{\omega = (\omega_1, \dots, \omega_{10})$  na  $i$ -tej pozycji jest koincydencja). Zdarzenie  $\bigcup_{i=1}^n A_i$  oznacza, że przynajmniej jeden list trafia do adresata. Zauważmy, że

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}, P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}, P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \frac{(n-k)!}{n!}. \text{ Wobec tego ze wzoru (*)}$$

$$P\left(\bigcup_{i=1}^n A_i\right) = \binom{n}{1} \frac{(n-1)!}{n!} - \binom{n}{2} \frac{(n-2)!}{n!} + \binom{n}{3} \frac{(n-3)!}{n!} + \dots + (-1)^{n+1} \frac{1}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} + \dots + (-1)^{n+1} \frac{1}{n!} \approx 1 - e^{-1} \approx 0.6321$$

Z wzoru Taylora mamy  $\left| P\left(\bigcup_{i=1}^n A_i\right) - (1 - e^{-1}) \right| \leq \frac{1}{(n+1)!}$ , co oznacza, że już dla niewielkich  $n$

prawdopodobieństwo  $P\left(\bigcup_{i=1}^n A_i\right)$  praktycznie nie zależy od  $n$ .

Oznaczmy  $p_{0,n} = 1 - P\left(\bigcup_{i=1}^n A_i\right) = P\left(\left(\bigcup_{i=1}^n A_i\right)'\right) = \frac{1}{2!} - \frac{1}{3!} + \dots + (-1)^n \frac{1}{n!} \approx e^{-1}$  prawdopodobieństwo, że żaden spośród  $n$  listów nie dotrze do adresata (w permutacji nie ma żadnej koincydencji).

*Problem.* Jakie jest prawdopodobieństwo zdarzenia  $B$ , że dokładnie  $k$  spośród  $n$  listów dotrze do adresata (inaczej w permutacji mamy dokładnie  $k$  koincydencji).

Aby dokładnie  $k$  spośród  $n$  listów dotarło do adresata (dokładnie  $k$ -koincydencji) wybieramy  $k$  spośród  $n$  listów które trafiają do właściwych kopert na  $C_n^k = \binom{n}{k}$  sposobów i zliczamy wszystkie permutacje  $n-k$  elementowe w których żaden element nie jest na swoim miejscu (0 koincydencji). Oznaczmy przez  $n_{0,n-k}$

liczbę tych permutacji. Z początkowego fragmentu rozumowania wynika, że  $p_{0,n-k} = \frac{n_{0,n-k}}{(n-k)!}$ . Stąd

$$n_{0,n-k} = p_{0,n-k} \cdot (n-k)!$$

$$P(B) = \frac{\binom{n}{k} n_{0,n-k}}{n!} = \frac{\binom{n}{k} p_{0,n-k} \cdot (n-k)!}{n!} = \frac{1}{k!} \left[ \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots + (-1)^{n-k} \frac{1}{(n-k)!} \right]$$

- Z talii 52 kart wybrano 13. Jakie jest prawdopodobieństwo że

a) brak będzie przynajmniej jednego koloru,

b) brak będzie dokładnie jednego koloru.

Odp. Oznaczmy przez  $A_1, \dots, A_4$  zdarzenia brak trefli, kar, kierów i pików. Oznaczmy

$$p_1 = P(A_1) = \frac{C_{39}^{13}}{C_{52}^{13}}, \quad p_2 = P(A_1 \cap A_2) = \frac{C_{26}^{13}}{C_{52}^{13}}, \quad p_3 = P(A_1 \cap A_2 \cap A_3) = \frac{C_{13}^{13}}{C_{52}^{13}}$$

Ad a) Z wzoru włączeń i wyłączeń mamy prawdopodobieństwo braku co najmniej 1 koloru

$$b_1 = P\left(\bigcup_{i=1}^4 A_i\right) = 4p_1 - 6p_2 + 4p_3$$

Ad b) podobnie jak w pkt a obliczamy prawdopodobieństwo braku co najmniej 2 kolorów

$b_2 = P(A_1 \cap A_2 \cup \dots \cup A_3 \cap A_4) = 6p_2 - 12p_3 + 4p_3$  (3 spośród 15 iloczynów  $(A_i \cap A_j) \cap (A_k \cap A_l)$  są puste a 4 iloczyny  $(A_i \cap A_j) \cap (A_k \cap A_l) \cap (A_m \cap A_n)$  spośród 20 są niepuste i mają prawdopodobieństwa równe  $p_3$ ). Stąd

prawdopodobieństwo braku dokładnie 1 koloru jest równe  $d_1 = b_1 - b_2 = 4p_1 - 12p_2 + 12p_3$ . Oczywiście  $d_0 = 1 - b_1$ ,

$$d_2 = b_2 - b_3, \quad d_3 = b_3 = 4p_3$$

### **Prawdopodobieństwo warunkowe.**

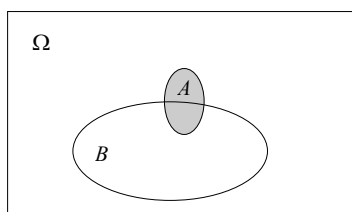
Niech  $(\Omega, \mathcal{S}, P)$  będzie przestrzenią probabilistyczną i niech  $B \in \mathcal{S}$  będzie dowolnym ustalonym zdarzeniem o niezerowym prawdopodobieństwie  $P(B) > 0$ .

Def. Prawdopodobieństwem warunkowym zdarzenia  $A \in \mathcal{S}$  pod warunkiem  $B$  nazywamy liczbę

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Def. Funkcję  $P(\cdot | B): \mathcal{S} \rightarrow \mathbb{R}$  nazywamy warunkowym rozkładem prawdopodobieństwa.

Z definicji widać że  $P(A \cap B) = P(A|B) P(B)$



Informacja, że zaszło zdarzenie  $B$  powoduje zastąpienie wyjściowego rozkładu rozkładem warunkowym w którym cała masa prawdopodobieństwa jest skupiona w  $B$ . Zdarzenia rozłączne z  $B$  muszą mieć prawdopodobieństwo 0. Jeżeli zdarzenia  $A_1$  i  $A_2$  są zawarte w  $B$  to stosunek ich prawdopodobieństw powinien być zachowany. Dla każdego zdarzenia  $A$  rozkład warunkowy przypisuje więc przeskalowane prawdopodobieństwo części wspólnej  $A \cap B$ .

### **Niezależność zdarzeń.**

Niezależność jest kluczowym pojęciem, odróżniającym rachunek prawdopodobieństwa od innych działów matematyki. Wiąże się z prawdopodobieństwem warunkowym w ten sposób, że zdarzenia  $A$  i  $B$  są niezależne, jeżeli wiedza o zajściu zdarzenia  $B$  nie ma wpływu na prawdopodobieństwo zajścia zdarzenia  $A$ . W przypadku, gdy  $P(B) > 0$  powyższy warunek jest równoważny warunkowi

$P(A|B) = P(A)$ , z którego wynika warunek  $P(A \cap B) = P(A)P(B)$ . Ostatni warunek ma sens nawet wówczas, gdy  $P(B) = 0$  i symetrycznie traktuje ona zdarzenia  $A$  i  $B$ . Motywuje to następującą definicję.

**Def.** Zdarzenia  $A$  i  $B$  są niezależne jeżeli  $P(A \cap B) = P(A)P(B)$ .

Nie zawsze na podstawie słownego opisu zdarzeń można zorientować się, czy są one niezależne. Świadczy o tym poniższy przykład.

**Przykład.** Wybieramy jedną rodzinę spośród wszystkich rodzin mających  $n$  dzieci. Niech  $A$  oznacza zdarzenie, że w wybranej rodzinie jest co najwyżej jedna dziewczynka, natomiast zdarzenie  $B$  polega na tym że w rodzinie są dzieci obojga płci.

Zdarzeniami elementarnymi są  $n$ -elementowe ciągi np.  $(DC...DDC)$  uporządkowane wg. starszeństwa dzieci). Oczywiście  $P(A) = \frac{n+1}{2^n}$  ( $n$  ciągów z jedną literką  $D$  i 1 ciąg samych  $C$ ). Natomiast

$P(B) = \frac{2^n - 2}{2^n}$  (ze zbioru  $2^n$  ciągów wykluczamy 2 ciągi samych  $D$  i samych  $C$ ). Natomiast

$P(A \cap B) = \frac{n}{2^n}$  (dokładnie jedna dziewczynka). Warunek niezależności  $P(A \cap B) = P(A)P(B)$  jest

równoważny  $\frac{n+1}{2^n} \frac{2^n - 2}{2^n} = \frac{n}{2^n} \Leftrightarrow n+1 = 2^{n-1} \Leftrightarrow n = 3$ . Okazuje się, że zdarzenia mające

identyczny opis słowny raz są niezależne ( $n = 3$ ) a innym razem są zależne ( $n \neq 3$ ).

Dla większej liczby zdarzeń definicja jest bardziej skomplikowana.

**Def.** Zdarzenia  $A_1, \dots, A_n$  są niezależne gdy  $P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$  dla wszystkich ciągów wskaźników  $(i_1, \dots, i_k)$ , gdzie  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ ,  $k = 2, 3, \dots, n$ .

Poniższe przykłady wskazują, że nie można zredukować liczby warunków ( $2^n - n - 1$ ). Można jedynie nieco je przeformułować.

**Przykład.** Dla  $n = 3$  z niezależności parami zdarzeń  $A_1, A_2$  i  $A_3$  nie wynika ich niezależność. Wystarczy przyjąć  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$  i  $P(\{\omega_i\}) = \frac{1}{4}$  oraz  $A_i = \{\omega_1, \omega_{i+1}\}$   $i=1, 2, 3$ . Oczywiście  $P(A_i) = \frac{1}{2}$ ,  $P(A_i \cap A_j) = P(\{\omega_1\}) = \frac{1}{4} = P(A_i)P(A_j)$  dla  $i \neq j$ . Ale

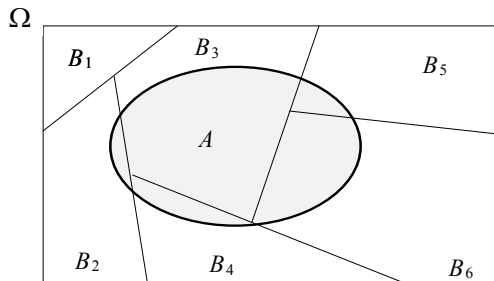
$$P(A_1 \cap A_2 \cap A_3) = P(\{\omega_1\}) = \frac{1}{4} \neq P(A_1)P(A_2)P(A_3).$$

Można także pokazać, że  $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$  nie pociąga za sobą niezależności parami

**Przykład.** Rzucamy kostką w kształcie ośmiościanu foremnego którego ścianki są ponumerowane cyframi od 1 do 8. Niech  $A_1 = \{1, 2, 3, 4\}$ ,  $A_2 = A_3 = \{1, 5, 6, 7\}$ . Widać, że  $P(A_i) = \frac{1}{2}$  i  $P(A_1 \cap A_2 \cap A_3) = P(\{1\}) = \frac{1}{8} = P(A_1)P(A_2)P(A_3)$  natomiast żadne dwa spośród zdarzeń  $A_1, A_2$  i  $A_3$  nie są niezależne.

### Tw. (o prawdopodobieństwie całkowitym)

Jeżeli  $\Omega = \bigcup_{i=1}^{\infty} B_i$ ,  $i \neq j \Rightarrow B_i \cap B_j = \emptyset$ ,  $P(B_i) > 0 \forall i$  to,  $P(A) = \sum_{i=1}^{\infty} P(A | B_i)P(B_i)$



Dowód :

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P\left(\bigcup_{i=1}^{\infty} (A \cap B_i)\right) = \sum_{i=1}^{\infty} P(A \cap B_i) = \\ &= \sum_{i=1}^{\infty} P(A | B_i)P(B_i) \end{aligned}$$

Z powyższego wyniku następujący wzór Bayesa:

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{\sum_i P(A | B_i)P(B_i)}$$

**Zadanie.** Trzech myśliwych strzeliło jednocześnie do dzika. Dzika zabiła jedna kula. Obliczyć prawdopodobieństwo, że dzika trafił drugi myśliwy, jeżeli myśliwi strzelają niezależnie a prawdopodobieństwa trafienia do celu dla poszczególnych myśliwych są równe 0.2, 0.4, 0.6. (Odp. 0.277).

### Paradoks Simpsona

Firmy A i B prowadzą takie same szkolenia. Dokonano analizy ilość osób kończących szkolenia z oceną pozytywną, z podziałem na płeć:

- mężczyźni:
  - firma A: 210 osób zdało, 190 – nie zdało (zdawalność 52,5%)
  - firma B: 30 osób zdało, 70 – nie zdało (zdawalność 30,0%)
- kobiety:
  - firma A: 590 osób zdało, 10 – nie zdało (zdawalność 98,3%)
  - firma B: 870 osób zdało, 30 – nie zdało (zdawalność 96,7%)

Zarówno dla kobiet jak i mężczyzn oddzielnie firma A kształci lepiej. Jednak dokonując analizy całościowej:

- firma A: 800 osób zdało, 200 – nie zdało (zdawalność 80%)
- firma B: 900 osób zdało, 100 – nie zdało (zdawalność 90%)

wynika, że firma B jest lepsza!

Może się więc zdarzyć, że

$$P(Z | A \cap M) > P(Z | B \cap M),$$

$$P(Z | A \cap K) > P(Z | B \cap K),$$

a  $P(Z | A) < P(Z | B)$ .

Rzeczywiście

$$\begin{aligned}
P(Z | A) &= \frac{P(Z \cap A)}{P(A)} = \frac{P((Z \cap A \cap M) \cup (Z \cap A \cap K))}{P(A)} = \frac{P((Z \cap A \cap M))}{P(A)} + \frac{P((Z \cap A \cap K))}{P(A)} = \\
&= \frac{P((Z \cap A \cap M))}{P(A \cap M)} \frac{P(A \cap M)}{P(A)} + \frac{P((Z \cap A \cap K))}{P(A \cap K)} \frac{P(A \cap K)}{P(A)} = \\
&= P(Z | A \cap M)P(M | A) + P(Z | A \cap K)P(K | A)
\end{aligned}$$

Podobnie

$$\begin{aligned}
P(Z | B) &= \frac{P(Z \cap B)}{P(B)} = \frac{P((Z \cap B \cap M) \cup (Z \cap B \cap K))}{P(B)} = \frac{P((Z \cap B \cap M))}{P(B)} + \frac{P((Z \cap B \cap K))}{P(B)} = \\
&= \frac{P((Z \cap B \cap M))}{P(B \cap M)} \frac{P(B \cap M)}{P(B)} + \frac{P((Z \cap B \cap K))}{P(B \cap K)} \frac{P(B \cap K)}{P(B)} = \\
&= P(Z | B \cap M)P(M | B) + P(Z | B \cap K)P(K | B)
\end{aligned}$$

Widać, że  $P(Z | A)$  jest średnią ważoną prawdopodobieństw  $P(Z | A \cap M)$  i  $P(Z | A \cap K)$  z wagami  $P(M | A)$  i  $P(K | A)$ .

Podobnie  $P(Z | B)$  jest średnią ważoną prawdopodobieństw  $P(Z | B \cap M)$  i  $P(Z | B \cap K)$  z wagami  $P(M | B)$  i  $P(K | B)$ .

$$\text{U nas } P(M | A) = \frac{4}{10} \quad P(K | A) = \frac{6}{10} \quad P(M | B) = \frac{1}{10} \quad P(K | B) = \frac{9}{10}$$

$$P(Z | A) = \frac{210}{400} \frac{4}{10} + \frac{590}{600} \frac{6}{10} = \frac{8}{10} \quad \text{i} \quad P(Z | B) = \frac{30}{100} \frac{1}{10} + \frac{870}{900} \frac{9}{10} = \frac{9}{10}.$$

Ważona średnia pary liczb mniejszych od ich odpowiedników w drugiej parze może być większa od ważonej średniej (z innymi wagami niż poprzednio) liczb drugiej pary.

## Zmienne losowe.

W różnych zagadnieniach praktycznych przestrzeni zdarzeń elementarnych  $\Omega$  może mieć różne postacie. Aby ujednoczyć sposób opisu dokonujemy odwzorowania przestrzeni  $\Omega$  w  $R$  a w bardziej złożonych przypadkach w  $R^n$  za pomocą funkcji zwanej zmienną losową. Jeżeli zbiór wartości tej funkcji jest podzbiorem  $R$ , to mówimy o jednowymiarowej zmiennej losowej. Jeżeli zbiór wartości tej funkcji jest podzbiorem  $R^n$ , to mówimy o  $n$ -wymiarowej zmiennej losowej. Szczególnie interesować nas będą 1 i 2 wymiarowe zmienne losowe.

**Def.** Niech  $(\Omega, \mathcal{S}, P)$  będzie przestrzenią probabilistyczną. Jednowymiarową zmienną losową nazywamy rzeczywistą funkcję  $X: \Omega \ni \omega \rightarrow X(\omega) \in R$  spełniającą warunek:

$$(m) \quad \forall x \in R \quad \{\omega : X(\omega) < x\} \in \mathcal{S}.$$

Warunek (m) zwany warunkiem mierzalności umożliwia przeniesienie rozkładu prawdopodobieństwa z  $(\Omega, \mathcal{S}, P)$  do przestrzeni mierzalnej  $(R, \mathcal{A}(R))$  poprzez zdefiniowanie dystrybuanty w  $R$  za pomocą standardowej formuły  $F_X(x) = P^X((-\infty, x]) = P(\{\omega : X(\omega) \leq x\})$ . Dystrybuantę  $F_X$  nazywamy dystrybuantą zmiennej losowej  $X$ . Przestrzeń mierzalna  $(R, \mathcal{A}(R))$  wyposażona w rozkład prawdopodobieństwa  $P^X$  staje się przestrzenią probabilistyczną  $(R, \mathcal{A}(R), P^X)$  stowarzyszoną z  $(\Omega, \mathcal{S}, P)$ .

### Jednowymiarowe dyskretne zmienne losowe.

Jednowymiarową zmienną losową  $X$  nazywamy zmienną dyskretną jeżeli zbiór  $\mathcal{X}$  jej wartości jest skończony lub przeliczalny.

### Wybrane rozkłady jednowymiarowych dyskretnych zmiennych losowych.

- **Rozkład jednopunktowy** -  $\mathcal{X} = \{x_0\}$  ;  $P\{\omega : X(\omega) = x_0\} = 1$
- **Rozkład dwupunktowy** -  $\mathcal{X} = \{x_1, x_2\}$  ;  $P\{\omega : X(\omega) = x_1\} = p$ ,  $P\{\omega : X(\omega) = x_2\} = q = 1 - p$ ,  $p \in (0, 1)$
- **Rozkład dwumianowy**  $\mathcal{B}(n, p)$  -  $\mathcal{X} = \{0, 1, \dots, n\}$  ;  $P\{\omega : X(\omega) = k\} = \binom{n}{k} p^k q^{n-k}$  - pojawia się

wówczas, gdy w niezmiennych warunkach wykonujemy  $n$  razy eksperyment w wyniku którego może pojawić się zdarzenie  $A$  zwane sukcesem z prawdopodobieństwem  $p$  lub zdarzenie przeciwne zwane porażką z prawdopodobieństwem  $q = 1 - p$  (schemat dwumianowy, próby Bernoulliego). Rozkład dwumianowy podaje prawdopodobieństwo  $k$  sukcesów w  $n$  próbach Bernoulliego.

- **Rozkład Poissona**  $\mathcal{P}(\lambda)$  -  $\mathcal{X} = \{0, 1, \dots\}$  ;  $P\{\omega : X(\omega) = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$



**Tw. Poissona.** Niech  $X_n$  będzie ciągiem zmiennych losowych o rozkładach dwumianowych  $B(n, p_n)$ .

Jeżeli  $\lim_{n \rightarrow \infty} np_n = \lambda > 0$ , to  $\lim_{n \rightarrow \infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ .

Rzeczywiście oznaczając  $np_n = \lambda_n \rightarrow \lambda$

$$P(X_n = k) = \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} = \\ 1\left(1 - \frac{1}{n}\right)\dots\left(1 - \frac{k-1}{n}\right) \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^n \left(1 - \frac{\lambda_n}{n}\right)^{-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}.$$

Oszacowanie błędu aproksymacji rozkładu dwumianowego rozkładem Poissona podaje poniższa nierówność prawdziwa dla każdego zbioru borelowskiego  $B$

$$|P_{B(n, p_n)}(B) - P_{\mathcal{P}(\lambda)}(B)| \leq \frac{\lambda^2}{n}$$

Z powyższego twierdzenia wynika, że rozkład Poissona można uważać za graniczny przypadek rozkładu dwumianowego. Aproksymacja rozkładu dwumianowego rozkładem Poissona jest **wystarczająco dokładna**, gdy  $n$  jest duże,  $p$ -małe a  $np$ -średnie. Często interpretuje się rozkład Poissona jako rozkład zmiennej losowej, która przybiera wiele różnych wartości ale z małymi prawdopodobieństwami. Z tego powodu rozkład Poissona nazywa się czasami **prawem małych liczb**. Rozkład Poissona modeluje ilość zgłoszeń w systemach masowej obsługi. Jest również punktem wyjścia do konstrukcji rozkładów złożonych użytecznych np. w matematyce ubezpieczeniowej.

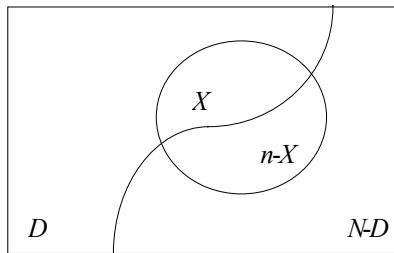
Przykłady

- 1 Po terenie miasta jeździ 1000 samochodów. Prawdopodobieństwo wezwania pogotowia technicznego przez jeden samochód wynosi  $p=0.002$ . Obliczyć prawdopodobieństwo  $P(A)$  wezwania pogotowia przez którykolwiek z samochodów zakładając, że wezwania są zdarzeniami niezależnymi. Podać wynik dokładny i przybliżony uzyskany z aproksymacji rozkładu dwumianowego rozkładem Poissona. Oszacować teoretycznie błąd tego przybliżenia i sprawdzić jego dokładność w rozważanym przypadku.

(Odp. dokł.  $P(A)=1-0.998^{1000}=0.864935$ ; przybl.  $P(A)=1-e^{-2}=0.864665$ , oszacowanie błędu: błąd  $\leq 0.004$ .)

- 2 Tekst broszury zawiera 100000 znaków. W trakcie pisania każdy znak może zostać błędnie wprowadzony z prawdopodobieństwem 0.001. Z kolei redaktor znajduje każdy z błędów z prawdopodobieństwem 0.9, po czym tekst wraca do autora, który znajduje każdy z pozostałych błędów z prawdopodobieństwem 0.5. Jaka jest szansa, że po obu korektach broszura będzie zawierała nie więcej niż 3 błędy. (Odp. Przybliżenie rozkładem Poissona z parametrem  $\lambda = 5 = 0.001 \cdot 0.1 \cdot 0.5$   $p=i\text{Poisson}(3;5)=0.26503 \pm 0.00025$ )

**Rozkład hipergeometryczny-  $H(N,D,n)$  - (Rohatgi str. 335)**



W urnie jest  $D$  kul białych i  $N-D$  czarnych. Z urny losujemy  $n \leq N$  kul. Niech  $X$  będzie zmienną losową oznaczającą ilość białych kul wśród  $n$  wylosowanych.  $X$  może przybierać wartości ze zbioru  $\mathcal{A} = \{\max(0, n+D-N), \dots, \min(D, n)\}$ . Rozkład prawdopodobieństwa zmiennej  $X$  zwany rozkładem hipergeometrycznym jest zadany wzorem :

$$P(X = x) = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}} \quad x \in [\max\{0, n+D-N\}, \min\{D, n\}].$$

Rozkład hipergeometryczny ma zastosowanie do szacowania licznosci populacji zwierząt np. ryb w stawie

**Wybrane rozkłady jednowymiarowych ciągłych zmiennych losowych.**

Jednowymiarowe rozkłady ciągłe zadane są funkcjami gęstości (względem miary Lebesgue'a)

- **Rozkład jednostajny  $U(a,b)$**  -  $f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b) \\ 0, & x \notin (a, b) \end{cases}$

- **Rozkład normalny  $N(m, \sigma)$**  -  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$

**Standaryzacja:**  $X \sim N(m, \sigma) \Leftrightarrow \frac{X-m}{\sigma} \sim N(0,1)$  . Dystrybuanta standaryzowanego rozkładu normalnego jest dostępna w tablicach lub w postaci procedur komputerowych .

- **Rozkład wykładniczy  $\mathcal{E}(\lambda)$**  -  $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$  . Rozkład wykładniczy ma szerokie

zastosowania w teorii niezawodności do modelowania czasu życia elementu i w systemach masowej obsługi do modelowania czasu oczekiwania na usługę. Niech  $T$  będzie zmienną losową oznaczającą czas życia elementu. Niech  $F(t)$  oznacza dystrybuantę zmiennej losowej  $T$ . W analizie czasu życia wprowadza się dwie użyteczne funkcje:

\* funkcję przeżycia (survival function)  $S(t) = 1 - F(t) = P(T \geq t)$  , która określa prawdopodobieństwo, że element przeżyje co najmniej  $t$

\* funkcję awaryjności ( hazard function, failure rate)  $\lambda(t) = \frac{f(t)}{S(t)}$ , którą interpretujemy jako gęstość warunkową prawdopodobieństwa wystąpienia awarii pod warunkiem, że element przeżył  $t$ .

Rzeczywiście  $P(T \in [t, t + \Delta t) | T \geq t) = \frac{P(T \in [t, t + \Delta t))}{P(T \geq t)} \approx \frac{f(t)\Delta t}{S(t)} = \lambda(t)\Delta t$ . Funkcja  $\lambda(t)$  jednoznacznie

wyznacza gęstość, gdyż  $\lambda(t) = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)}$ . Rozwiązując równanie różniczkowe otrzymujemy

$S(t) = e^{-\int_0^t \lambda(u) du}$  a następnie  $f(t) = \lambda(t)e^{-\int_0^t \lambda(u) du}$  dla  $t > 0$ . Widać, że jeśli funkcja awaryjności jest

stała, to czas życia elementu ma rozkład wykładniczy. Stałość funkcji awaryjności oznacza, że element nie starzeje się, co w wielu realnych sytuacjach jest nie do przyjęcia. Przyjmując jako

funkcję awaryjności funkcję potęgową  $\lambda(t) = \frac{\alpha}{\beta} t^{\alpha-1}$  otrzymujemy funkcję gęstości rozkładu

Weibulla  $f(t) = \frac{\alpha}{\beta} t^{\alpha-1} e^{-\frac{t^\alpha}{\beta}}$ ,  $t > 0$ .

**Rozkład gamma**  $\mathcal{G}(a, p)$ ,  $a, p > 0$  -  $f(x) = \begin{cases} \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax}, & x > 0 \\ 0, & x \leq 0 \end{cases}$ , gdzie  $\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt$ . Rodzina

rozkładów gamma jest szeroką rodziną rozkładów obejmującą m.in. rodzinę rozkładów wykładniczych.

### Twierdzenia o dodawaniu

$$\left. \begin{array}{l} X_1 \sim \mathcal{B}(n_1, p) \\ X_2 \sim \mathcal{B}(n_2, p) \\ X_1, X_2 \text{ niezależne} \end{array} \right\} \Rightarrow X_1 + X_2 \sim \mathcal{B}(n_1 + n_2, p) \parallel$$

$$\left. \begin{array}{l} X_1 \sim \mathcal{P}(\lambda_1) \\ X_2 \sim \mathcal{P}(\lambda_2) \\ X_1, X_2 \text{ niezależne} \end{array} \right\} \Rightarrow X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$$

$$\left. \begin{array}{l} X_1 \sim N(m_1, \sigma_1) \\ X_2 \sim N(m_2, \sigma_2) \\ X_1, X_2 \text{ niezależne} \end{array} \right\} \Rightarrow X_1 + X_2 \sim N(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$

$$\left. \begin{array}{l} X_1 \sim \mathcal{G}(a, p_1) \\ X_2 \sim \mathcal{G}(a, p_2) \\ X_1, X_2 \text{ niezależne} \end{array} \right\} \Rightarrow X_1 + X_2 \sim \mathcal{G}(a, p_1 + p_2)$$

## Proces Poissona

Rozważmy rodzinę zmiennych losowych  $\{N_t; t > 0\}$  indeksowaną ciągłym parametrem  $t \in [0, \infty)$ .

Rodzinę  $\{N_t; t > 0\}$  nazywamy procesem stochastycznym. Funkcję  $N(\cdot, \omega) : R_+ \ni t \rightarrow R$  nazywamy realizacją procesu albo trajekcją. Niech  $\{N_t; t > 0\}$  będzie procesem zliczającym interesujące nas zdarzenia tzn.

$$N_t = \text{liczba zdarzeń w przedziale } [0, t].$$

Zakładamy dodatkowo, że  $P(N_0 = 0) = 1$ .

Niech  $T_1, T_2, \dots$  będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie wykładniczym  $\mathcal{E}(\lambda)$ , który interpretujemy jako czasy pomiędzy kolejnymi zdarzeniami. Wówczas

$$N_t = \begin{cases} 0 & , \text{ dla } T_1 > t \\ \sup_n \{T_1 + T_2 + \dots + T_n \leq t\} & , \text{ dla } T_1 \leq t \end{cases}$$

jest tzw. **jednorodnym procesem Poissona** na półprostej  $R_+ = [0, \infty)$ .

Oznaczając przez  $S_n = T_1 + T_2 + \dots + T_n$  moment pojawienia się  $n$ -tego sygnału (zdarzenia) możemy łatwo wyznaczyć rozkład zmiennej losowej  $S_n$ . Rzeczywiście zmienna  $S_n$  jest sumą niezależnych zmiennych losowych o rozkładzie  $\mathcal{E}(\lambda) = \mathcal{G}(\lambda, 1)$ . Stosując twierdzenie o dodawaniu dla rozkładu

$\mathcal{G}(\lambda, 1)$  otrzymujemy natychmiast że zmienna losowa  $S_n$  ma rozkład  $\mathcal{G}(\lambda, n)$

o funkcji gęstości

$$g_n(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} \mathbf{1}_{(0, \infty)}(t)$$

i dystrybuancie

$$G_n(t) = \int_0^t g_n(u) du = 1 - \int_t^\infty g_n(u) du = (1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!}) \mathbf{1}_{(0, \infty)}(t).$$

Ten szczególny przypadek rozkładu gamma jest znany jako rozkład Erlanga.

Ławo zauważyć, że

$$\begin{aligned} P(N_t = k) &= P(S_k \leq t, S_{k+1} > t) = P(S_k \leq t) + P(S_{k+1} > t) - P(\{S_k \leq t\} \cup \{S_{k+1} > t\}) = \\ &= G_k(t) + (1 - G_{k+1}(t)) - 1 = e^{-\lambda t} \frac{(\lambda t)^k}{k!}. \end{aligned}$$

Zatem zmienna losowa  $N_t$  ma rozkład Poissona z parametrem  $\lambda t$ . Oczywiście  $P(N_0 = 0) = 1$  i proces ma prawostronnie ciągle trajektorie. Ponadto proces Poissona jest procesem przyrostach niezależnych, tzn. dla dowolnych chwil  $0 = t_0 < t_1 < t_2 < \dots < t_n$  zmienne losowe

$N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$  są niezależne i  $N_t - N_s$  (dla  $t > s$ ) ma rozkład Poissona z parametrem  $\lambda(t - s)$ .

**Przykład.** Klienci przychodzą do banku zgodnie z procesem Poissona o intensywności  $\lambda$  klientów na godzinę. Wiadomo, że e ciągu pierwszej godziny bank odwiedziło 2 klientów. Obliczyć prawdopodobieństwo, że

- Pierwszy klient przyszedł w ciągu pierwszych 30 minut a drugi w ciągu następnych 30 minut
- Pierwszy klient przyszedł w ciągu pierwszych 15 minut a drugi w ciągu ostatnich 15 minut
- Żaden klient nie przyszedł w ciągu pierwszych 15 minut.

$$P(N_{1/2} = 1, N_1 - N_{1/2} = 1 | N_1 = 2) = \frac{P(N_{1/2} = 1, N_1 - N_{1/2} = 1, N_1 = 2)}{P(N_1 = 2)}$$

$$\text{Ad a)} \quad = \frac{P(N_{1/2} = 1, N_1 - N_{1/2} = 1)}{P(N_1 = 2)} = \frac{P(N_{1/2} = 1)P(N_1 - N_{1/2} = 1)}{P(N_1 = 2)} = \frac{\frac{\lambda}{2} e^{-\lambda/2} \cdot \frac{\lambda}{2} e^{-\lambda/2}}{\frac{\lambda^2}{2!} e^{-\lambda}} = \frac{1}{2}$$

$$P(N_{1/4} = 1, N_1 - N_{3/4} = 1 | N_1 = 2) = \frac{P(N_{1/4} = 1, N_{3/4} - N_{1/4} = 0, N_1 - N_{3/4} = 1)}{P(N_1 = 2)} =$$

$$\text{Ad b)} \quad \frac{P(N_{1/4} = 1)P(N_{3/4} - N_{1/4} = 0)P(N_1 - N_{3/4} = 1)}{P(N_1 = 2)} = \frac{\frac{\lambda}{4} e^{-\lambda/4} \cdot e^{-\lambda/2} \cdot \frac{\lambda}{4} e^{-\lambda/4}}{\frac{\lambda^2}{2!} e^{-\lambda}} = \frac{1}{8}$$

$$P(N_{1/4} = 0 | N_1 = 2) = \frac{P(N_{1/4} = 0, N_1 - N_{1/4} = 2)}{P(N_1 = 2)} = \frac{P(N_{1/4} = 0)P(N_1 - N_{1/4} = 2)}{P(N_1 = 2)}$$

$$\text{Ad c)} \quad = \frac{e^{-\lambda/4} \cdot \frac{(\frac{3\lambda}{4})^2}{2!} e^{-3\lambda/4}}{\frac{\lambda^2}{2!} e^{-\lambda}} = \frac{9}{16}$$

**Przykład.** Niech  $\{N_t; t > 0\}$  będzie procesem jednorodnym procesem Poissona o intensywności  $\lambda$ . Niech  $S_n$  oznacza moment zajścia  $n$ -tego zdarzenia. Wyznaczyć dystrybuantę i gęstość rozkładu warunkowego  $S_4$  gdy wiadomo, że  $N_1 = 2$ .

**Uwaga.** Wiadomo, że  $\{S_4 > t\} = \{N_t \leq 3\}$ . Rzeczywiście czwarte zdarzenie pojawi się po chwili  $t$ , czyli w przedziale  $(t, \infty)$ , gdy w przedziale czasu  $[0, t]$  zaobserwujemy co najwyżej 3 zdarzenia.

Dla  $t > 1$  mamy

$$\begin{aligned} F_{S_4|N_1=2}(t) &= P(S_4 \leq t | N_1 = 2) = 1 - P(S_4 > t | N_1 = 2) = 1 - P(N_t \leq 3 | N_1 = 2) = \\ &= 1 - \frac{P(N_t \leq 3, N_1 = 2)}{P(N_1 = 2)} = 1 - \frac{P(N_1 = 2, N_t - N_1 \leq 1)}{P(N_1 = 2)} = 1 - \frac{P(N_1 = 2)P(N_t - N_1 \leq 1)}{P(N_1 = 2)} = \\ &= 1 - P(N_t - N_1 \leq 1) = 1 - \left(1 + \frac{\lambda(t-1)}{1!}\right)e^{-\lambda(t-1)}. \end{aligned}$$

Stąd  $f_{S_4|N_1=2}(t) = \lambda^2(t-1)e^{-\lambda(t-1)}\mathbf{1}_{(1,\infty)}(t)$ .

## **Wartość oczekiwana, wariancja i inne momenty zmiennej losowej.**

**Def.** Wartością oczekiwaną zmiennej losowej  $X$  określonej na przestrzeni  $(\Omega, \mathcal{S}, P)$  nazywamy **skończoną** całkę zmiennej  $X$  względem rozkładu  $P$  (o ile całka istnieje) (istnieje, gdy całka z  $|X|$  jest skończona) tzn.

$$E(X) = \int_{\Omega} X(\omega) P(d\omega).$$

Wartość oczekiwana jest odpowiednikiem pojęcia środka (jednostkowej) masy i jest traktowana jako tzw. parametr położenia zmiennej losowej. Wartość oczekiwana nie musi istnieć np. rozkład Cauchy'ego o funkcji gęstości  $f(x) = \frac{1}{\pi(1+x^2)}$ . Czasami istnieje więc potrzeba używania innych parametrów położenia np. **mediany**.

Niech  $X$  będzie zmienną losową określoną na przestrzeni  $(\Omega, \Sigma, P)$  a  $g$  pewną funkcją rzeczywistą taką, że  $g(X)$  jest zmienną losową ( $g$  musi być funkcją borelowską).

**Def.** Wartością oczekiwaną zmiennej losowej  $g(X)$  nazywamy (o ile istnieje) **skończoną** całkę

$$E(g(X)) = \int_{\Omega} (gX)(\omega) P(d\omega) = \int_{\Omega} g(X(\omega)) P(d\omega) = \{ \text{zmiana zmiennych } x = X(\omega) \} = \int_R g(x) dP^X$$

**Def.** Momentem zwykłym rzędu  $k$  nazywamy liczbę  $m_k = E(X^k)$  (o ile istnieje). Np.  $m_1 = E(X)$   $m_2 = E(X^2)$ .

**Def.** Momentem centralnym rzędu  $k$  nazywamy liczbę (o ile istnieje)  $\mu_k = E(X - E(X))^k$ .

Łatwo zauważyć, że  $\mu_1 = E(X - E(X)) = 0$ . Drugi moment centralny  $\mu_2 = E(X - E(X))^2$  nazywamy **wariancją** zmiennej losowej  $X$  i oznaczamy  $V(X)$ . Wariancja zmiennej losowej jest odpowiednikiem pojęcia momentu bezwładności jednostkowej masy względem środka masy i jest miarą rozrzutu wartości zmiennej losowej wokół wartości oczekiwanej.

Bezpośrednio z własności całki wg. miary (rozkładu prawdopodobieństwa) wynika, że

- $E(aX+b) = aE(X)+b$
- $V(aX+b) = a^2V(X)$
- $V(X) = \mu_2 = m_2 - m_1^2$

W tabeli 1 podano wartości oczekiwane i wariancje dla wcześniej rozważanych rozkładów prawdopodobieństwa.

**Tabela 1.**

Rozkład	Postać rozkładu	$E(X)$	$V(X)$
jednopunktowy	$P\{\omega: X(\omega)=x_0\}=1$	$x_0$	0
dwupunktowy	$P\{\omega: X(\omega)=x_1\}=p, P\{\omega: X(\omega)=x_2\}=q=1-p$	$px_1+qx_2$	$pq(x_1-x_2)^2$
dwumianowy - $B(n,p)$	$P\{\omega: X(\omega)=k\}=\binom{n}{k}p^kq^{n-k}, k=0,1,\dots,n$	$np$	$npq$
Poissona - $P(\lambda)$	$P\{\omega: X(\omega)=k\}=\frac{\lambda^k}{k!}e^{-\lambda}, k=1,2,\dots$	$\lambda$	$\lambda$
hipergeometryczny $H(N,D,n)$	$P(X=x)=\frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$ $x \in [\max\{0, n+D-N\}, \min\{D, n\}]$ .	$\frac{nD}{N}$	$\frac{nD}{N} \frac{N-D}{N} \frac{N-n}{N-1}$
jednostajny - $U(a,b)$	$f(x)=\begin{cases} \frac{1}{b-a}, & x \in (a,b) \\ 0, & x \notin (a,b) \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
normalny - $N(m,\sigma)$	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$	$m$	$\sigma^2$
wykładniczy - $\mathcal{E}(\lambda)$	$f(x)=\begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
gamma $\mathcal{G}(a,p)$	$f(x)=\begin{cases} \frac{a^p}{\Gamma(p)}x^{p-1}e^{-ax}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	$\frac{p}{a}$	$\frac{p}{a^2}$

Przykłady obliczania  $E(X)$  i  $V(X)$

- rozkład dwumianowy

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

Wiadomo, że  $\forall p, q \in R \quad (p+q)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k}$  (\*)

Wzór (\*) przedstawi wielomian dwóch zmiennych  $p$  i  $q$  (czyli funkcję różniczkowalną). Różniczkując cząstkowo (\*) względem  $p$  otrzymujemy

$$\forall p, q \in R \quad n(p+q)^{n-1} = \sum_{k=0}^n k \binom{n}{k} p^{k-1} q^{n-k}. \text{ Mnożąc obie strony przez } p \text{ otrzymujemy}$$

$$\forall p, q \in R \quad np(p+q)^{n-1} = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} (**)$$



W szczególności dla  $p+q=1$   $E(X) = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = np$ .

Różniczkując (\*\*) względem  $p$  mamy

$$\forall p, q \in R \quad n(p+q)^{n-1} + np(n-1)(p+q)^{n-2} = \sum_{k=0}^n k^2 \binom{n}{k} p^{k-1} q^{n-k}.$$

Mnożąc przez  $p$   $\forall p, q \in R$   $np(p+q)^{n-1} + np^2(n-1)(p+q)^{n-2} = \sum_{k=0}^n k^2 \binom{n}{k} p^k q^{n-k}$ . Stąd dla

$$p+q=1 \quad E(X^2) = \sum_{k=0}^n k^2 \binom{n}{k} p^k q^{n-k} = np + np^2(n-1) = np + n^2 p^2 - np^2.$$

Wobec tego  $V(X) = E(X^2) - E^2(X) = np + n^2 p^2 - np^2 - n^2 p^2 = np - np^2 = np(1-p) = npq$

- Rozkład Poissona

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

Wiadomo z kursu analizy matematycznej, że  $\forall \lambda \in R$   $e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$ . Powyższy szereg potęgowy

można różniczkować wyraz po wyrazie więc  $\forall \lambda \in R$   $e^{\lambda} = \sum_{k=0}^{\infty} k \frac{\lambda^{k-1}}{k!}$ . Mnożąc obustronnie przez

$$\lambda \text{ otrzymujemy } \forall \lambda \in R \quad \lambda e^{\lambda} = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} \text{ (#).}$$

Różniczkując (#) mamy  $\forall \lambda \in R$ :  $e^{\lambda} + \lambda e^{\lambda} = \sum_{k=0}^{\infty} k^2 \frac{\lambda^{k-1}}{k!}$  i stąd

$$\forall \lambda \in R: \quad \lambda e^{\lambda} + \lambda^2 e^{\lambda} = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} \text{ (##)}$$

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

Wobec tego  $V(X) = E(X^2) - E^2(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$

- Rozkład normalny

$$E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

Wiadomo, że  $\forall (m, \sigma) \in R \times R_+$   $\frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 1$  Mamy więc w półpłaszczyźnie  $R \times R_+$

określoną stałą funkcję  $G(m, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 1$ . Różniczkujemy tą funkcję cząstkowo

$$\frac{\partial}{\partial m}(G(m, \sigma)) = \frac{\partial}{\partial m} \left( \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx \right) = 0. \text{ Wykorzystując możliwość zmiany kolejności}$$

różniczkowania i całkowania (jedno z tw dotyczących zmiany kolejności limesowania) otrzymujemy

$$\forall (m, \sigma) \in R \times R_+ \quad \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \frac{(x-m)}{\sigma^2} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 0 \Rightarrow \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} (x-m) e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 0 \Rightarrow$$

$$\frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx - m \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 0 \Rightarrow E(X) - m = 0 \Rightarrow E(X) = m$$

Różniczkując cząstkowo  $\frac{\partial}{\partial m}$  uzyskaną w powyższym rozumowaniu tożsamość

$$\forall (m, \sigma) \in R \times R_+ \quad \frac{\partial}{\partial m} \left( \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} (x-m) e^{-\frac{(x-m)^2}{2\sigma^2}} dx \right) = \frac{\partial}{\partial m} (0) = 0 \text{ uzyskujemy}$$

$$-\frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx + \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \frac{(x-m)^2}{\sigma^2} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{\partial}{\partial m} (0) = 0 \quad / \cdot \sigma^2 \Rightarrow$$

$$-\sigma^2 \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx + \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 0 \Leftrightarrow -\sigma^2 + V(X) = 0 \Leftrightarrow V(X) = \sigma^2$$

- **Rozkład gamma**  $\mathcal{G}(a, p)$ ,  $\forall a, p > 0$   $\frac{a^p}{\Gamma(p)} \int_0^{\infty} x^{p-1} e^{-ax} dx = 1$

$$E(X) = \frac{a^p}{\Gamma(p)} \int_0^{\infty} x \cdot x^{p-1} e^{-ax} dx = \frac{a^p}{\Gamma(p)} \int_0^{\infty} x^{(p+1)-1} e^{-ax} dx = \frac{a^p}{\Gamma(p)} \frac{\Gamma(p+1)}{a^{p+1}} \left[ \frac{a^{p+1}}{\Gamma(p+1)} \int_0^{\infty} x^{(p+1)-1} e^{-ax} dx \right] = \frac{p}{a}$$

Podobnie

$$E(X^2) = \frac{a^p}{\Gamma(p)} \int_0^{\infty} x^2 \cdot x^{p-1} e^{-ax} dx = \frac{a^p}{\Gamma(p)} \int_0^{\infty} x^{(p+2)-1} e^{-ax} dx = \frac{a^p}{\Gamma(p)} \frac{\Gamma(p+2)}{a^{p+2}} \left[ \frac{a^{p+2}}{\Gamma(p+2)} \int_0^{\infty} x^{(p+2)-1} e^{-ax} dx \right] = \frac{p(p+1)}{a^2}$$

$$\text{Stąd } V(X) = E(X^2) - E^2(X) = \frac{p(p+1)}{a^2} - \frac{p^2}{a^2} = \frac{p}{a^2}$$

Na momentach oparte są użyteczne parametry opisowe rozkładu zmiennej losowej:

- $\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$  współczynnik asymetrii lub skośności (*skewness*)
- $\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3$  współczynnik spłaszczenia (*kurtosis*)

### Nierówność Czebyszewa.

**Tw.** Jeżeli zmienna losowa  $X$  ma (skończoną) wariancję, to  $\forall \varepsilon > 0 \quad P(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$ .

**Dowód.**  $V(X) = \int_{\Omega} (X(\omega) - E(X))^2 P(d\omega) =$

$$= \int_{\{\omega: |X(\omega) - E(X)| \geq \varepsilon\}} (X(\omega) - E(X))^2 P(d\omega) + \int_{\{\omega: |X(\omega) - E(X)| < \varepsilon\}} (X(\omega) - E(X))^2 P(d\omega) \geq$$

$$\int_{\{\omega: |X(\omega) - E(X)| \geq \varepsilon\}} (X(\omega) - E(X))^2 P(d\omega) \geq \varepsilon^2 \int_{\{\omega: |X(\omega) - E(X)| \geq \varepsilon\}} P(d\omega) = \varepsilon^2 P(|X - E(X)| \geq \varepsilon).$$

Korzystając z wzoru na prawdopodobieństwo zdarzenia przeciwnego otrzymujemy

$$\forall \varepsilon > 0 \quad P(|X - E(X)| < \varepsilon) \geq 1 - \frac{V(X)}{\varepsilon^2}.$$

**Przykład.** Dla zmiennej losowej o rozkładzie normalnym  $N(m, \sigma)$  oszacować  $P(|X - m| < 3\sigma)$

- za pomocą nierówności Czebyszewa -  $P(|X - m| < 3\sigma) \geq 1 - \frac{\sigma^2}{9\sigma^2} = 0.8889$
- wykorzystując postać rozkładu -  $P(|X - m| < 3\sigma) = 0.9973$ .

**Reguła 3 $\sigma$ .** Dla rozkładu **normalnego** praktycznie wszystkie wartości zmiennej losowej zawarte są w przedziale  $(m - 3\sigma, m + 3\sigma)$ .

### Kwantyle

Z uwagi na to, że momenty zmiennej losowej nie muszą istnieć (np. rozkład Cauchy'ego) istnieje potrzeba użycia innych parametrów opisowych, które tej wady nie mają.

**Def.** Kwantylem rzędu  $p$  zmiennej losowej o dystrybuancie  $F$  nazywamy każdą liczbę  $x_p$  spełniającą warunek  $F(x_p) \leq p \leq F(x_p^+)$ .

Jeżeli dystrybuanta jest ciągła i monotoniczna, to kwantyl jest wyznaczony jednoznacznie przez równość  $F(x_p) = p$ . W przypadku, gdy dystrybuanta przyjmuje wartość  $p$  na pewnym przedziale, to kwantylem rzędu  $p$  jest każda liczba z tego przedziału. Kwantyl rzędu 0.5 nazywamy **medianą**. Mediana jest konkurencyjnym w stosunku do wartości oczekiwanej parametrem położenia. Kwantyl

rzędu 0.25 (tzn.  $x_{0,25}$ ) nazywamy **dolnym kwartylem** a kwantyl rzędu 0.75 (tzn.  $x_{0,75}$ ) **górnym kwartylem**. Różnicę pomiędzy górnym i dolnym kwartylem ( tzn.  $x_{0,75}-x_{0,25}$ ) nazywamy **odległością międzykwartylową**, która jest konkurencyjnym w stosunku do odchylenia standardowego ( $SD(X)=\sqrt{V(X)}$ ) parametrem rozproszenia.

**Przykład.** Punkt materialny  $M$  porusza się ze stałą prędkością po okręgu o promieniu  $r$ . Niech  $P$  będzie ustalonym punktem okręgu a  $X$  odległością punktu  $M$  od punktu  $P$ . Znaleźć  $E(X)$  i  $V(X)$ .

**Rozw.** Niech  $O$  będzie środkiem koła i niech  $\Phi$  oznacza kąt pomiędzy  $OP$  i  $OM$ . Z uwagi na stałą prędkość kątową  $\Phi$  ma rozkład jednostajny na przedziale  $[0,2\pi)$ . Odległość  $X$  (losowa) pomiędzy punktami  $P$  i  $M$  jest związana z kątem  $\Phi$  wzorem  $X = 2r \sin(\frac{\Phi}{2})$ . Przyjmijmy  $\Psi = \frac{\Phi}{2}$ . Oczywiście  $\Psi = \frac{\Phi}{2} \sim U([0, \pi])$  i  $X = 2r \sin(\Psi)$ . Dla  $x \in [0, 2r)$  wyznaczamy wartość dystrybuanty

$$F_X(x) = P(X < x) = P(2r \sin(\Psi) < x) = P(\sin(\Psi) < \frac{x}{2r}) = P(\Psi \in [0, \arcsin \frac{x}{2r}) \cup (\pi - \arcsin \frac{x}{2r}, \pi)) = \frac{2}{\pi} \arcsin \frac{x}{2r}$$

Funkcja gęstości jest postaci  $f_X(x) = F_X'(x) = \frac{2r}{\pi \sqrt{4r^2 - x^2}} \mathbf{1}_{[0, 2r)}(x)$

Wartość oczekiwaną można obliczyć dwoma sposobami

$$E(X) = E(2r \sin \Psi) = 2r \int_0^{\pi} \sin t \left(\frac{1}{\pi} dt\right) = \frac{4r}{\pi}, \quad E(X) = \int_0^{2r} x f_X(x) dy = \frac{4r}{\pi}. \text{ Podobnie można obliczyć}$$

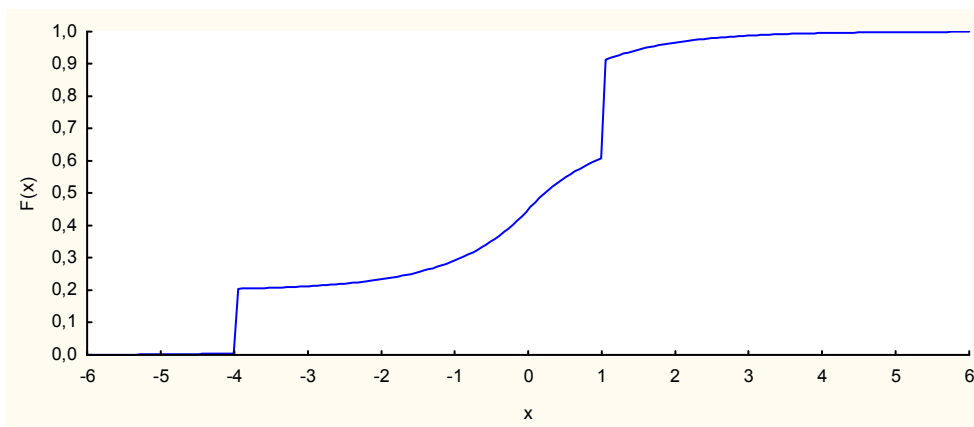
$$E(X^2) = 2r^2 \text{ i } V(X) = E(X^2) - E^2(X) = 2r^2 - \left(\frac{4r}{\pi}\right)^2 = \left(2 - \frac{16}{\pi^2}\right)r^2$$

**Przykład.** Dany jest rozkład zmiennej losowej  $X$  postaci  $P^X = 0,2 \delta_{\{-4\}} + 0,3 \delta_{\{1\}} + f \cdot l$ , gdzie  $f(x) = ae^{-|x|}$  a  $l$  oznacza miarę Lebesgue'a na  $R$ .

a. Wyznaczyć dystrybuantę tego rozkładu i naszkicować jej wykres

$$\int_R dP^X = 1 \Rightarrow 0,2 + 0,3 + a \int_R e^{-|x|} dx = 1 \Rightarrow a = \frac{1}{4}.$$

$$F(x) = \begin{cases} \frac{1}{4} \int_{-\infty}^x e^{-|t|} dt ; x \leq -4 \\ \frac{2}{10} + \frac{1}{4} \int_{-\infty}^x e^{-|t|} dt ; -4 < x \leq 1 \\ \frac{5}{10} + \frac{1}{4} \int_{-\infty}^x e^{-|t|} dt ; x > 1 \end{cases} \quad F(x) = \begin{cases} \frac{1}{4} e^x ; x \leq -4 \\ \frac{2}{10} + \frac{1}{4} e^x ; -4 < x \leq 0 \\ \frac{45}{100} + \frac{1}{4} (1 - e^{-x}) ; 0 < x \leq 1 \\ \frac{75}{100} + \frac{1}{4} (1 - e^{-x}) ; x > 1 \end{cases} \text{ (wersja lewostr. cg.)}$$

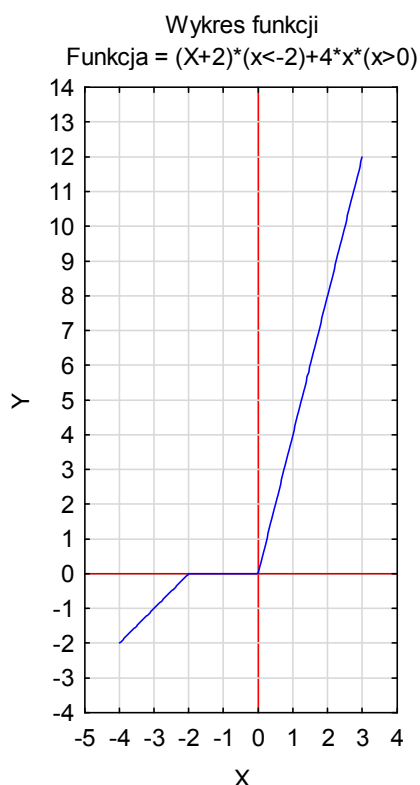


b. Obliczyć  $P(X > 0)$  i  $P(-0,5 < X < 2)$

$$P(X > 0) = 1 - P(X \leq 0) = 1 - F(0_+) = 0,55$$

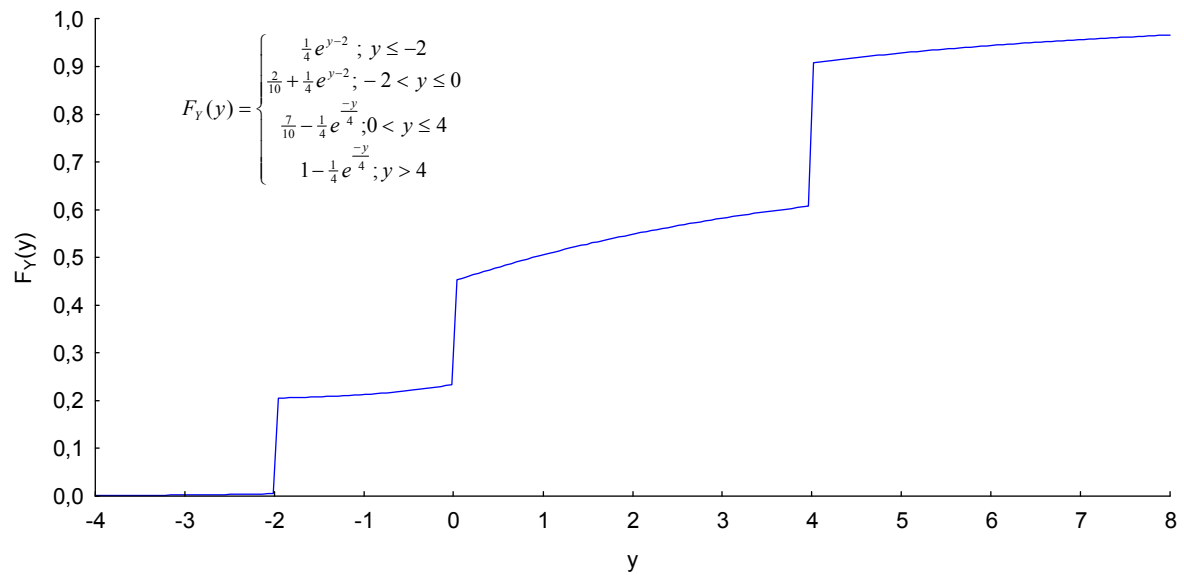
$$P(-0,5 < X < 2) = F(2) - F(-0,5_+) = \frac{8}{10} - \frac{1}{4}(e^{-2} + e^{-0,5}) = 0,6145$$

c. Niech  $Y = \begin{cases} X + 2, & \text{dla } X < -2 \\ 0, & \text{dla } -2 \leq X \leq 0 \\ 4X, & \text{dla } X > 0 \end{cases}$ . Wyznaczyć rozkład zmiennej losowej  $Y$ .



Najpierw wyznaczymy dystrybuantę zmiennej losowej  $Y$ .

$$F_Y(y) = P(Y < y) = \begin{cases} F(y-2); y \leq 0 \\ F(\frac{y}{4}); y > 0 \end{cases} = \begin{cases} \frac{1}{4}e^{-y-2}; y \leq -2 \\ \frac{2}{10} + \frac{1}{4}e^{-y-2}; -2 < y \leq 0 \\ \frac{7}{10} - \frac{1}{4}e^{-\frac{y}{4}}; 0 < y \leq 4 \\ 1 - \frac{1}{4}e^{-\frac{y}{4}}; y > 4 \end{cases}$$



Stąd uzyskujemy rozkład  $P^Y$

$$P^Y = 0,2\delta_{\{-2\}} + 0,25(1 - e^{-2})\delta_{\{0\}} + 0,3\delta_{\{4\}} + \left( \frac{1}{4}e^{y-2}\mathbf{1}_{(-\infty, 0]}(y) + \frac{1}{16}e^{-\frac{y}{4}}\mathbf{1}_{(0, \infty)}(y) \right) \cdot l$$

## Wielowymiarowe zmienne losowe (wektory losowe)

Z uwagi na zakres zastosowań i prostotę rozważań ograniczymy nasze zainteresowanie do dwuwymiarowych zmiennych losowych. Niech  $(\Omega, \mathcal{S}, P)$  będzie przestrzenią probabilistyczną.

**Def.** Dwuwymiarową zmienną losową nazywamy parę rzeczywistych funkcji określonych na  $\Omega$

$(X, Y): \Omega \ni \omega \rightarrow (X(\omega), Y(\omega)) \in \mathbb{R}^2$  spełniającą warunek:

$$(m) \quad \forall (x, y) \in \mathbb{R}^2 \quad \{\omega: X(\omega) < x, Y(\omega) < y\} \in \mathcal{S}.$$

Warunek (m) zwany warunkiem mierzalności umożliwia przeniesienie rozkładu prawdopodobieństwa z  $(\Omega, \mathcal{S}, P)$  do przestrzeni mierzalnej  $(\mathbb{R}^2, \mathcal{A}(\mathbb{R}^2))$  poprzez zdefiniowanie dystrybuanty w  $\mathbb{R}^2$  za pomocą standardowej formuły  $F_{X,Y}(x,y) = P(\{\omega: X(\omega) < x, Y(\omega) < y\})$ . Dystrybuantę  $F_{X,Y}$  nazywamy dystrybuantą zmiennej losowej  $(X, Y)$ . Przestrzeń mierzalna  $(\mathbb{R}^2, \mathcal{A}(\mathbb{R}^2))$  wyposażona w rozkład prawdopodobieństwa  $P^{XY}$  staje się przestrzenią probabilistyczną  $(\mathbb{R}^2, \mathcal{A}(\mathbb{R}^2), P^{XY})$  stowarzyszoną z  $(\Omega, \mathcal{S}, P)$ .

Z rozkładu dwuwymiarowej zmiennej losowej  $(X, Y)$  natychmiast uzyskujemy rozkłady zmiennych  $X$  i  $Y$ . Są to tzw. rozkłady brzegowe. Ich dystrybuanty dane są wzorami:

$$F_X(x) = P(\{\omega: X(\omega) < x\}) = \lim_{y \rightarrow \infty} P(\{\omega: X(\omega) < x, Y(\omega) < y\}) = \lim_{y \rightarrow \infty} F_{X,Y}(x,y),$$

$$F_Y(y) = P(\{\omega: Y(\omega) < y\}) = \lim_{x \rightarrow \infty} P(\{\omega: X(\omega) < x, Y(\omega) < y\}) = \lim_{x \rightarrow \infty} F_{X,Y}(x,y).$$

Zmienne losowe  $X$  i  $Y$  są **niezależne** wtedy i tylko wtedy, gdy  $F_{X,Y}(x,y) = F_X(x) F_Y(y)$ .

Jeżeli  $(X, Y)$  jest **dwuwymiarową dyskretną zmienną losową** zmienną losową tzn. zbiór jej wartości (czyli zbiór par  $(x_i, y_j)$ ) jest skończony lub przeliczalny, to rozkład prawdopodobieństwa może być zadany za pomocą **funkcji rozkładu** dwuwymiarowej zmiennej losowej

$$p_{ij} = P(\{\omega: X(\omega) = x_i, Y(\omega) = y_j\}).$$

Z powyższej funkcji natychmiast otrzymujemy funkcje rozkładu jednowymiarowych zmiennych losowych:

$$p_{i\bullet} = P(\{\omega: X(\omega) = x_i\}) = \sum_j p_{ij}, \quad p_{\bullet j} = P(\{\omega: Y(\omega) = y_j\}) = \sum_i p_{ij}.$$

Warunek niezależności można zapisać w postaci:

$$\text{dyskretne zmienne losowe } X \text{ i } Y \text{ są niezależne} \Leftrightarrow \forall i, j \quad p_{ij} = p_{i\bullet} p_{\bullet j}.$$

W przypadku, gdy dyskretna zmienna losowa  $(X,Y)$  przyjmuje skończoną ilość wartości rozkład podaje się w postaci tzw. tablicy wielodzielczej.

$x \setminus Y$	$y_1$	.	.	$y_s$	$p_{i\bullet}$
$x_1$	$p_{11}$	.	.	$p_{1s}$	$p_{1\bullet}$
.	.	.	.	.	.
.	.	.	.	.	.
$x_r$	$p_{r1}$	.	.	$p_{rs}$	$p_{r\bullet}$
$p_{\bullet j}$	$p_{\bullet 1}$	.	.	$p_{\bullet s}$	1

Łatwo zauważyć, że dla dowolnego ustalonego  $1 \leq j \leq s$  ciąg liczb  $p_{ij} = \frac{p_{ij}}{p_{\bullet j}}$ ,  $i=1, \dots, r$  może być interpretowany jako rozkład prawdopodobieństwa, bo  $p_{ij} \geq 0$  i  $\sum_i p_{ij} = 1$ . Jest to rozkład warunkowy

$$p_{ij} = \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)} = P(X = x_i | Y = y_j). \text{ Podobnie definiuje się inne rozkłady warunkowe.}$$

Dwuwymiarowa zmienna losowa  $(X,Y)$  jest **zmienną losową zmienną losową typu ciągłego**, jeżeli istnieje nieujemna funkcja  $f_{X,Y}(x,y) \geq 0$  spełniająca warunek  $\iint_{R^2} f_{X,Y}(x,y) dx dy = 1$  zwana funkcją gęstości (względem miary Lebesgue'a) a dystrybuanta wyraża się wzorem  $F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u,v) du dv$ . Funkcja gęstości jest więc z dokładnością do zbioru miary (L) 0 pochodną dystrybuanty:

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}.$$

Funkcje gęstości rozkładów brzegowych są postaci

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx.$$

Warunek niezależności można zapisać w postaci:

$$\text{ciągłe zmienne losowe } X \text{ i } Y \text{ są niezależne} \Leftrightarrow f_{X,Y}(x,y) = f_X(x) f_Y(y) \text{ prawie wszędzie (L)}$$

**Def.** Dystrybuantą warunkową zmiennej losowej  $X$  ze względu na zdarzenie  $B$  ( $P(B) > 0$ ) nazywamy funkcję

$$F_{X|B}(x) = \frac{P(\{\omega: X(\omega) < x\} \cap B)}{P(B)}, \quad x \in R.$$



Jeżeli zdarzenie  $B$  jest określone za pomocą zmiennej losowej  $Y$  w następujący sposób

$B = \{\omega: y - \Delta y \leq Y(\omega) < y + \Delta y\}$ , to przechodząc do granicy z  $\Delta y \rightarrow 0$  dystrybuantę warunkowego rozkładu zmiennej  $X$  przy ustalonej wartości zmiennej  $Y=y$

$$\begin{aligned} F_{X|Y=y}(x) &= F_{X|y}(x) = \lim_{\Delta y \rightarrow 0^+} \frac{P(\{\omega: X(\omega) < x, y - \Delta y \leq Y(\omega) < y + \Delta y\})}{P(\{\omega: y - \Delta y \leq Y(\omega) < y + \Delta y\})} = \\ &= \lim_{\Delta y \rightarrow 0^+} \frac{F_{(X,Y)}(x, y + \Delta y) - F_{(X,Y)}(x, y - \Delta y)}{F_Y(y + \Delta y) - F_Y(y - \Delta y)} = \lim_{\Delta y \rightarrow 0^+} \frac{\int_{-\infty}^x \int_{y-\Delta y}^{y+\Delta y} f_{X,Y}(u, v) du dv}{\int_{y-\Delta y}^{y+\Delta y} f_Y(v) dv} = \\ &= \int_{-\infty}^x \frac{f_{X,Y}(u, y)}{f_Y(y)} du \quad \text{zakładając, że } f_Y(y) > 0. \end{aligned}$$

Funkcja gęstości rozkładu warunkowego zmiennej  $X$  pod warunkiem  $Y=y$  wyraża się wzorem

$$f_{X|Y=y}(x) = f_{X|y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad \text{dla } y \text{ spełniających warunek } f_Y(y) > 0 \text{ a dystrybuanta}$$

$$F_{X|Y=y}(x) = \int_{-\infty}^x f_{X|y}(u) du.$$

Podobnie  $f_{Y|X=x}(y) = f_{Y|x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$ ,  $F_{Y|X=x}(y) = \int_{-\infty}^y f_{Y|x}(v) dv$ .

Z powyższych definicji wynika, że dla **ciągłej** dwuwymiarowej zmiennej losowej  $(X, Y)$  prawdziwy jest wzór:

$$\forall x \in R \quad \forall B \in \mathcal{B}(R) \quad P(\{\omega: X(\omega) < x, Y(\omega) \in B\}) = \int_B F_{X|y}(x) f_Y(y) dy.$$

Dla dwuwymiarowej **dyskretnej** zmiennej losowej  $(X, Y)$  powyższy wzór przybiera postać:

$$\forall x \in R \quad \forall B \in \mathcal{B}(R) \quad P(\{\omega: X(\omega) < x, Y(\omega) \in B\}) = \sum_{\{j: y_j \in B\}} F_{X|y_j}(x) p_{\bullet j}.$$

Wzory te można zapisać łącznie w postaci

$$\forall x \in R \quad \forall B \in \mathcal{B}(R) \quad P(\{\omega: X(\omega) < x, Y(\omega) \in B\}) = \int_B F_{X|Y=y}(x) P_Y(dy) = \int_{Y^{-1}(B)} F_{X|Y(\omega)}(x) P(d\omega)$$

We wzorze tym można wyraźnie dostrzec wersję znanego wzoru na prawdopodobieństwo całkowite.

## Dwuwymiarowa zmienna losowa o rozkładzie normalnym $N(m_x, m_y, \sigma_x, \sigma_y, \rho)$ .

Rozkład normalny  $N(m_x, m_y, \sigma_x, \sigma_y, \rho)$  dwuwymiarowej zmiennej losowej  $(X, Y)$  jest zadany funkcją gęstości względem dwuwymiarowej miary Lebesgue'a

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-m_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-m_x}{\sigma_x}\right)\left(\frac{y-m_y}{\sigma_y}\right) + \left(\frac{y-m_y}{\sigma_y}\right)^2}, \quad \sigma_x > 0, \sigma_y > 0, -1 < \rho < 1.$$

Po prostych rachunkach otrzymujemy:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}, \quad \text{czyli}$$

rozkłady brzegowe dwuwymiarowej zmiennej losowej o rozkładzie  $N(m_x, m_y, \sigma_x, \sigma_y, \rho)$  są jednowymiarowymi rozkładami normalnymi  $N(m_x, \sigma_x)$  i  $N(m_y, \sigma_y)$ . Ponadto

$$f_{X|Y=y}(x) = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho^2}} e^{-\frac{(x-m_x-\rho\frac{\sigma_x}{\sigma_y}(y-m_y))^2}{2\sigma_x^2(1-\rho^2)}} \sim N\left(m_x + \rho\frac{\sigma_x}{\sigma_y}(y-m_y), \sigma_x\sqrt{1-\rho^2}\right)$$
$$f_{Y|X=x}(y) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{(y-m_y-\rho\frac{\sigma_y}{\sigma_x}(x-m_x))^2}{2\sigma_y^2(1-\rho^2)}} \sim N\left(m_y + \rho\frac{\sigma_y}{\sigma_x}(x-m_x), \sigma_y\sqrt{1-\rho^2}\right),$$

czyli rozkłady warunkowe również są normalne z wartością oczekiwaną będącą liniową funkcją warunku i wariancją niezależną od warunku mniejszą niż wariancja w odpowiednim rozkładzie bezwarunkowym. Z postaci funkcji gęstości dwuwymiarowego rozkładu normalnego wynika, że składowe  $X$  i  $Y$  są niezależne wtedy i tylko wtedy, gdy  $\rho=0$ .

## Momenty wielowymiarowej zmiennej losowej

Niech  $(X, Y)$  będzie dwuwymiarową zmienną losową a  $g: R^2 \rightarrow R$  funkcją borelowską. Wówczas  $g(X, Y)$  jest jednowymiarową zmienną losową dla której możemy zdefiniować wartość oczekiwaną (o ile istnieje) wzorem

$$E(g(X, Y)) = \iint_{\Omega} g(X(\omega), Y(\omega))P(d\omega),$$

który w przypadku, gdy  $(X, Y)$  jest dyskretną zmienną losową przybiera postać

$$E(g(X, Y)) = \sum_{i,j} g(x_i, y_j) p_{ij},$$

a w przypadku ciągłym

$$E(g(X, Y)) = \iint_{R^2} g(x, y) f_{X,Y}(x, y) dx dy.$$

Oczywiście warunkiem koniecznym i wystarczającym istnienia  $E(g(X,Y))$  jest sumowalność  $g(X,Y)$ .

Def. **Momentem zwykłym** rzędu  $k+l$  dwuwymiarowej zmiennej losowej  $(X,Y)$  nazywamy liczbę

$$m_{kl}=E(X^k Y^l) \quad (\text{o ile istnieje}).$$

W szczególności

$$m_{10}=E(X), m_{01}=E(Y), m_{20}=E(X^2), m_{02}=E(Y^2), m_{11}=E(XY).$$

Def. **Momentem centralnym** rzędu  $k+l$  dwuwymiarowej zmiennej losowej  $(X,Y)$  nazywamy liczbę

$$\mu_{kl}=E((X-E(X))^k (Y-E(Y))^l) \quad (\text{o ile istnieje}).$$

W szczególności

$$\mu_{10}=0, \mu_{01}=0, \mu_{20}=V(X), \mu_{02}=V(Y).$$

**Moment mieszany**  $\mu_{11}=E((X-E(X))(Y-E(Y)))$  nazywamy **kowariancją** zmiennych  $X$  i  $Y$ , którą oznaczamy  $Cov(X,Y)=\mu_{11}$ . Bezpośredni rachunek pokazuje, że  $\mu_{11}=m_{11}-m_{10}m_{01}$ .

Łatwo zauważyć, że jeżeli  $X$  i  $Y$  są niezależne, to  $Cov(X,Y)=0$ .

Def. **Współczynnikiem korelacji** zmiennych losowych  $X$  i  $Y$  nazywamy liczbę

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}}.$$

**Przykład.** Jeżeli  $(X,Y) \sim N(m_x, m_y, \sigma_x, \sigma_y, \rho)$ , to  $\rho(X,Y)=\rho$ .

**Lemat.**  $|Cov(X,Y)| \leq \sqrt{V(X)V(Y)}$

**Dowód.**  $V(X+\alpha Y) = E[(X-E(X)+\alpha(Y-E(Y)))]^2 = E[(X-E(X))^2 + 2\alpha(X-E(X))(Y-E(Y)) + \alpha^2(Y-E(Y))^2] =$   
 $=V(X)+2\alpha Cov(X,Y)+\alpha^2 V(Y) \geq 0$  dla każdego  $\alpha$ . Stąd  $\Delta = 4 Cov^2(X,Y)-4V(X)V(Y) \leq 0$  co jest równoważne tezie lematu.

Z lematu wynika

**Tw.**  $|\rho(X,Y)| \leq 1$

Ponadto prawdziwe jest następujące

**Tw.** Jeżeli  $|\rho(X,Y)|=1$ , to istnieją takie stałe  $\alpha, \beta, \gamma$ , że  $P(\{\omega : \alpha X(\omega)+\beta Y(\omega)=\gamma\})=1$  tzn. z prawdopodobieństwem 1 pomiędzy zmiennymi  $X$  i  $Y$  zachodzi zależność liniowa.

**Dowód.** Wyznaczając wariancję kombinacji liniowej  $\alpha X + \beta Y$  jak w lemacie wnioskujemy z równości  $|\rho(X,Y)|=1$ , że istnieją takie  $\alpha, \beta$ , że  $\alpha^2 + \beta^2 > 0$  i  $V(\alpha X + \beta Y) = 0$ , co oznacza, że zmienna losowa  $\alpha X + \beta Y$  ma rozkład jednopunktowy skupiony np. w punkcie  $\gamma$  co kończy dowód.

Z powyższych faktów otrzymujemy

**Tw.** Jeżeli zmienne losowe  $X$  i  $Y$  są niezależne, to  $\rho(X,Y)=0$ .

Twierdzenie **odwrotne nie jest prawdziwe** (np. rozkład jednostajny na kole).

Współczynnik korelacji jest miarą stochastycznej liniowej zależności pomiędzy zmiennymi losowymi.

Dla  $n$  wymiarowej zmiennej losowej  $\mathbf{X}=(X_1, \dots, X_n)$  którą traktować będziemy jako wektor kolumnowy

$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$  odpowiednikami wartości oczekiwanej i wariancji są wektor wartości oczekiwanych

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{bmatrix}$$

i macierz kowariancyjna

$$V(\mathbf{X}) = \begin{bmatrix} V(X_1) & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & V(X_2) & \dots & Cov(X_2, X_n) \\ \vdots & \vdots & \dots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \dots & V(X_n) \end{bmatrix}.$$

Łatwo zauważyć, dla dowolnego  $n$  wymiarowego wektora  $\mathbf{l}$  mamy

$$E(\mathbf{l}^T \mathbf{X}) = \mathbf{l}^T E(\mathbf{X}) \text{ i } V(\mathbf{l}^T \mathbf{X}) = E(\mathbf{l}^T (\mathbf{X} - E(\mathbf{X})) (\mathbf{X} - E(\mathbf{X}))^T \mathbf{l}) = \mathbf{l}^T V(\mathbf{X}) \mathbf{l}.$$

Z ostatniej równości wynika, że macierz kowariancyjna jest nieujemnie określona. Stąd natychmiast dostajemy nierówności  $\det V(\mathbf{X}) \geq 0$  a w szczególności  $Cov^2(X,Y) - V(X)V(Y) \leq 0$ .

### **Warunkowa wartość oczekiwana i momenty warunkowe.**

Niech  $(X,Y)$  będzie dwuwymiarową zmienną losową. Przez  $P_{X|Y}$  oznaczymy warunkowy rozkład zmiennej losowej  $X$  pod warunkiem  $Y=y$ . Podobnie  $P_{Y|X}$  oznacza warunkowy rozkład zmiennej losowej  $Y$  pod warunkiem  $X=x$ . W zwykły sposób definiuje się wartości oczekiwane w powyższych rozkładach warunkowych jako całki względem odpowiednich warunkowych rozkładów prawdopodobieństwa i nazywa się je (o ile istnieją) warunkowymi wartościami oczekiwanymi

$$E(X|Y=y)=E(X|y)=\int_{\Omega} X(\omega)P_{X|y}(d\omega), \quad E(Y|X=x)=E(Y|x)=\int_{\Omega} Y(\omega)P_{Y|x}(d\omega).$$

Powyższe wzory przybierają postać

$$E(X|y)=\int_{-\infty}^{\infty} xf_{X|y}(x)dx, \quad E(Y|x)=\int_{-\infty}^{\infty} yf_{Y|x}(y)dy \quad \text{w przypadku ciągłym i}$$

$$E(X|y_j)=\sum_i x_i p(x_i | Y = y_j), \quad E(Y|x_i)=\sum_j y_j p(y_j | X = x_i) \quad \text{w przypadku dyskretnym.}$$

Analogicznie definiujemy warunkową wariancję i pozostałe momenty warunkowe. Z powyższych wzorów jasno wynika, że **warunkowa wartość oczekiwana jest funkcją warunku**.

### Regresja pierwszego i drugiego rodzaju

Niech  $(X, Y)$  będzie dwuwymiarową zmienną losową a  $E(Y|x)$  warunkową wartością oczekiwaną zmiennej losowej  $Y$  pod warunkiem  $X=x$ .

**Def.** Linia regresji pierwszego rodzaju zmiennej losowej  $Y$  względem zmiennej losowej  $X$  nazywamy wykres warunkowej wartości oczekiwanej  $E(Y|x)$  tzn. zbiór punktów  $(x, E(Y|x))$ .

**Def.** Prosta regresji drugiego rodzaju zmiennej losowej  $Y$  względem zmiennej losowej  $X$  nazywamy prostą  $y=ax+b$ , której parametru  $a$  i  $b$  spełniają warunek  $E(Y-aX-b)^2=\min$ .

Widać, że  $\psi(a, b)=E(Y-aX-b)^2=E(Y^2)+a^2E(X^2)+b^2-2aE(XY)-2bE(Y)+2abE(X)$ . Rozwiązując zagadnienie minimalizacji  $\psi(a, b)$  z warunku koniecznego istnienia ekstremum otrzymujemy układ

$$\begin{bmatrix} E(X^2) & E(X) \\ E(X) & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} E(XY) \\ E(Y) \end{bmatrix},$$

z którego otrzymujemy  $a = \frac{\text{Cov}(X, Y)}{V(X)} = \rho \sqrt{\frac{V(Y)}{V(X)}}$ ,  $b = E(Y) - aE(X)$  co po uwzględnieniu warunku wystarczającego prowadzi do równania prostej regresji drugiego rodzaju  $\frac{y-m_Y}{\sigma_Y} = \rho \frac{x-m_X}{\sigma_X}$ .

**Zadanie.** Dwuwymiarowa zmienna losowa  $(X, Y)$  ma rozkład o funkcji gęstości  $f(x, y)=cxy$ , dla  $0 \leq x \leq y \leq 1$ .

Wyznaczyć :

- stałą  $c$ ,

- współczynnik korelacji  $\rho_{(X, Y)}$  zmiennych  $X$  i  $Y$ . Czy zmienne  $X$  i  $Y$  są niezależne?

- linię regresji pierwszego rodzaju zmiennej  $Y$  względem  $X$  (wykres),

- prostą regresji drugiego rodzaju zmiennej  $Y$  względem  $X$  (wykres).

Rozkład prawdopodobieństwa zmiennej losowej  $(X, Y)$  skupiony jest na trójkącie  $T$ :  $\begin{cases} 0 \leq x \leq 1 \\ x \leq y \leq 1 \end{cases}$ .

a) Z warunku  $\iint_T f(x, y) dx dy = 1$  otrzymujemy  $c=8$ .

b) Dla  $0 \leq x \leq 1$   $f_X(x) = \int_x^1 8xy dy = 4x(1-x^2)$ , podobnie dla  $0 \leq y \leq 1$   $f_Y(y) = \int_0^y 8xy dx = 4y^3$ . Stąd

$$E(X) = \frac{8}{15}, E(Y) = \frac{4}{5}, E(X^2) = \frac{1}{3}, E(Y^2) = \frac{2}{3}, V(X) = \frac{11}{225}, V(Y) = \frac{2}{75}, E(XY) = 8 \int_0^1 \int_x^1 x^2 y^2 dx dy = \frac{4}{9}.$$

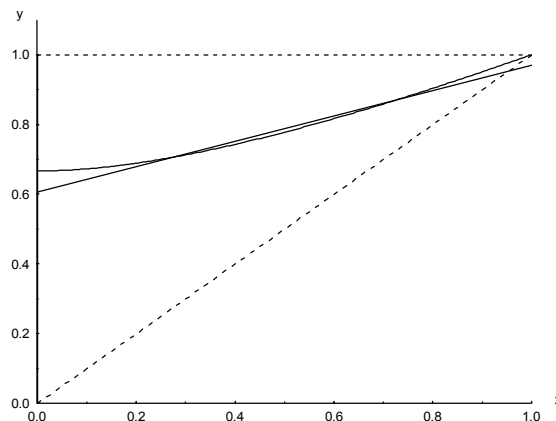
$$\text{Wobec tego } \rho(X, Y) = \frac{\frac{4}{9} - \frac{8}{15} \cdot \frac{4}{5}}{\sqrt{\frac{11}{225} \cdot \frac{2}{75}}} = \frac{4}{\sqrt{66}} = 0.4924$$

c) Dla dowolnego ustalonego  $0 \leq x < 1$  wyznaczamy funkcję  $f_{Y|x}(y) = \frac{f(x, y)}{f_X(x)} = \frac{2y}{1-x^2}$ ,  $x \leq y < 1$ , a następnie

$$E(Y|x) = \int_x^1 y f_{Y|x}(y) dy = \frac{2}{3} \frac{x^2 + x + 1}{x+1}$$

d) Prostą regresji II rodzaju wyznaczamy z ogólnego wzoru  $\frac{y-m_Y}{\sigma_Y} = \rho \frac{x-m_X}{\sigma_X}$ . Podstawiając odpowiednie

wartości otrzymujemy  $y = 0.6060 + 0.3636x$ . Poniższy rysunek przedstawia obie linie regresji na tle nośnika rozkładu dwuwymiarowej zmiennej losowej.



Bezpośrednim rachunkiem można sprawdzić, że

$$\begin{aligned} E(E(Y|X)) &= \int_{\Omega} \frac{X^2 + X + 1}{X + 1} dP = \int_{-\infty}^{\infty} \frac{x^2 + x + 1}{x + 1} f_X(x) dx = \int_0^1 \frac{x^2 + x + 1}{x + 1} 4x(1-x^2) dx = \\ &= \int_0^1 \frac{8}{3} x(1-x^3) dx = \frac{4}{5} = E(Y) \end{aligned}$$

$$V(E(Y|X)) = V(\varphi(X)) = 0,0065878, \quad V(\psi(X)) = 0,00646465$$

$$\frac{V(E(Y|X))}{V(Y)} = 0,2473 \quad \frac{V(\psi(X))}{V(Y)} = 0,2424$$

Powyższe ilorazy korelacyjne pokazują, że najlepsza średniokwadratowa aproksymacja zmiennej Y wyjaśnia ją w 24,7% , natomiast najlepsza średniokwadratowa liniowa aproksymacja zmiennej Y

wyjaśnia ją w 24,2%. W tym przypadku regresja liniowa jest prawie tak dobra jak optymalna regresja nieliniowa .

Jak zaznaczono wcześniej warunkowa wartość oczekiwana  $E(Y|x)$  jest funkcją warunku  $x$ . Traktując warunek  $x$  jako wartość zmiennej losowej  $X(\omega)$  dla pewnego  $\omega$  możemy traktować warunkową wartość oczekiwaną jako zmienną losową będącą funkcją zmiennej losowej  $X$  czyli warunku . Dokładniej

$$E(Y|X): \Omega \ni \omega \rightarrow E(Y|X)(\omega) = E(Y|X=x), \text{ gdy } X(\omega)=x$$

Warunkowa wartość oczekiwana  $E(Y|X)$  jest to zmienna losowa spełniająca dwa warunki

- jest ona stała na warstwicach warunkującej zmiennej losowej  $X$  czyli jest mierzalna względem podciała  $\sigma(X)$  generowanego przez zmienną losową  $X$

$$\bullet \forall A \in \sigma(X) \quad \int_A Y(\omega) dP = \int_A E(Y|X)(\omega) dP .$$

Uzasadnienie: Każdy zbiór  $A \in \sigma(X)$  jest przeciwobrazem przez zmienną losową  $X$  pewnego zbioru borelowskiego  $B \in \mathcal{B}(R)$  czyli  $A = X^{-1}(B)$ . Stąd

$$\begin{aligned} \int_A Y(\omega) dP &= \int_{X^{-1}(B)} Y(\omega) dP = \int_{\left\{ \begin{array}{l} x = X(\omega) \\ y = Y(\omega) \end{array} \right\}} y dP^{X,Y} = \iint_{(x,y) \in B \times R} y f_{X,Y}(x,y) dx dy = \int_B \left[ \int_R y \frac{f_{X,Y}(x,y)}{f_X(x)} dy \right] f_X(x) dx \\ &= \int_B E(Y|X=x) f_X(x) dx = \int_{X^{-1}(B)} E(Y|X)(\omega) dP = \int_A E(Y|X)(\omega) dP \end{aligned}$$

Z drugiego warunku przyjmując  $A = \Omega$  wynika natychmiast bardzo użyteczny wzór na iterowanie wartości oczekiwanych

$$\int_{\Omega} Y(\omega) dP = \int_{\Omega} E(Y|X)(\omega) dP \quad \text{czyli} \quad E(Y) = E[E(Y|X)] ,$$

który można uważać za probabilistyczną wersję twierdzenia Fubinię.

**Uwaga.** Warunkowa wartość oczekiwana  $E(Y|X)$  ma bardzo ważną własność. Wyrażenie  $E(Y-g(X))^2$ , które można traktować jako błąd średniokwadratowy aproksymacji zmiennej losowej  $Y$  poprzez zmienną losową  $g(X)$  przyjmuje wartość minimalną wówczas, gdy  $g(X) = E(Y|X)$  z prawdopodobieństwem 1. Rzeczywiście

$$\begin{aligned} E(Y-g(X))^2 &= E(Y - E(Y|X) + E(Y|X) - g(X))^2 = E(Y - E(Y|X))^2 + 2E((Y - E(Y|X))(E(Y|X) - g(X))) + \\ &\quad + E(E(Y|X) - g(X))^2 = E(Y - E(Y|X))^2 + E(E(Y|X) - g(X))^2, \end{aligned}$$

bo  $E((Y - E(Y|X))(E(Y|X) - g(X))) = E[E((Y - E(Y|X))(E(Y|X) - g(X)) | X)] = E[(E(Y|X) - g(X))E(Y - E(Y|X) | X)] = 0$ , gdyż  $E(Y - E(Y|X) | X) = 0$ .

Widać, że wyrażenie  $E(Y-g(X))^2$  przybiera najmniejszą wartość gdy  $E(E(Y|X) - g(X))^2 = 0$ , co ma miejsce wówczas, gdy  $g(X) = E(Y|X)$  z prawdopodobieństwem 1.

Porównując regresję pierwszego i drugiego rodzaju widzimy, że regresja pierwszego rodzaju minimalizuje błąd średniokwadratowy  $E(Y-g(X))^2$  aproksymacji zmiennej losowej  $Y$  przez zmienną losową  $g(X)$  w klasie wszystkich zmiennych losowych  $g(X)$  o skończonej wariancji a regresja liniowa drugiego rodzaju minimalizuje ten błąd w węższej klasie zmiennych losowych będących liniowymi (właściwie afinicznymi) funkcjami warunku. Do wyznaczenia regresji pierwszego rodzaju niezbędna jest znajomość łącznego rozkładu zmiennych  $X$  i  $Y$  a do wyznaczenia regresji liniowej drugiego rodzaju wystarczy znajomość momentów rzędu pierwszego i drugiego.

Powyższe rozważania można łatwo uogólnić na wielowymiarowe zmienna losowe. Ponadto w zagadnieniu regresji drugiego rodzaju można klasę funkcji liniowych zastąpić dowolną parametryczną klasą mierzalnych funkcji warunku. Wyznaczenie parametrów regresji drugiego rodzaju prowadzi do poszukiwania minimum funkcji nieliniowej (ze względu na parametry) i może być zrealizowane numerycznie za pomocą łatwo dostępnego oprogramowania.

### ***Wielowymiarowy rozkład normalny***

Wielowymiarowy rozkład normalny podobnie jak wcześniej omawiany dwuwymiarowy rozkład normalny może być zdefiniowany poprzez funkcję gęstości względem odpowiedniej wielowymiarowej miary Lebesgue'a , lecz sposób ten prowadzi często do uciążliwych rachunków. Wygodniej jest wprowadzić  $n$  wymiarowy rozkład normalny w oparciu o **twierdzenie Cramera-Wolda** głoszące, że rozkład  $n$  wymiarowej zmiennej losowej  $X$  jest całkowicie określony poprzez podanie jednowymiarowych rozkładów funkcji liniowych  $l^T X$  dla wszystkich  $l \in R^n$  (symbol  $^T$  oznacza transpozycję wektora traktowanego jako macierz kolumnową). To podejście może być także użyte do zdefiniowania rozkładu normalnego w nieskończenie wymiarowych przestrzeniach funkcyjnych Banacha czy Hilberta: wystarczy postulować aby funkcjonal liniowy miał jednowymiarowy rozkład normalny (Frechet 1951).

Niech  $X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$  będzie  $n$  wymiarową zmienną losową,  $m = E(X)$  wektorem wartości oczekiwanych a  $V = V(X)$  macierzą kowariancyjną.

**Def.** Zmienna losowa  $X$  ma  $n$  wymiarowy rozkład normalny  $N_n(m, V)$  wtedy i tylko wtedy, gdy dla każdego wektora  $l \in R^n$  jednowymiarowa zmienna losowa  $l^T X$  ma rozkład normalny  $N(l^T m, l^T V l)$ .

Z powyższej definicji łatwo wynikają następujące własności:

- jeżeli  $A$  jest macierzą typu  $(q, n)$  a  $n$  wymiarowa zmienna losowa  $X$  ma rozkład  $N_n(m, V)$  , to  $q$  wymiarowa zmienna losowa  $Y = AX$  ma rozkład normalny  $N_q(Am, AVA^T)$



- rozważmy  $p+q$  wymiarową zmienną losową  $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$  przy czym  $p$  pierwszych składowych tworzy  $p$  wymiarową zmienną  $\mathbf{X}_1$  a pozostałe  $q$  wymiarową zmienną losową  $\mathbf{X}_2$ . Dokonując odpowiedniego podziału na bloki możemy przedstawić wektor wartości oczekiwanych  $\mathbf{m}$  i macierz kowariancyjną  $\mathbf{V}$  w postaci blokowej  $\mathbf{m} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix}$   $\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}$ . Jeżeli  $\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N_{p+q} \left( \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \right)$ , to  $\mathbf{X}_1 \sim N_p(\mathbf{m}_1, \mathbf{V}_{11})$ ,  $\mathbf{X}_2 \sim N_q(\mathbf{m}_2, \mathbf{V}_{22})$   
 $\mathbf{X}_1 | \mathbf{X}_2 \sim N_p(\mathbf{m}_1 + \mathbf{V}_{12} \mathbf{V}_{22}^{-1} (\mathbf{X}_2 - \mathbf{m}_2), \mathbf{V}_{11} - \mathbf{V}_{12} \mathbf{V}_{22}^{-1} \mathbf{V}_{21})$   
 $\mathbf{X}_2 | \mathbf{X}_1 \sim N_q(\mathbf{m}_2 + \mathbf{V}_{21} \mathbf{V}_{11}^{-1} (\mathbf{X}_1 - \mathbf{m}_1), \mathbf{V}_{22} - \mathbf{V}_{21} \mathbf{V}_{11}^{-1} \mathbf{V}_{12})$
- jeżeli macierz kowariancyjna  $\mathbf{V}$   $n$  wymiarowej zmiennej losowej  $\mathbf{X}$  o rozkładzie  $N_n(\mathbf{m}, \mathbf{V})$  jest nieosobliwa, to istnieje gęstość tego rozkładu względem  $n$  wymiarowej miary Lebesgue'a i wyraża się wzorem  $f_n(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{V}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{V}^{-1} (\mathbf{x}-\mathbf{m})}$

**Zadanie.** Zmienna losowa  $(X, Y, Z)$  ma trójwymiarowy rozkład  $N(\mathbf{m}, \mathbf{V})$ , gdzie  $\mathbf{m} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$ ,  $\mathbf{V} = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & 4 \end{bmatrix}$ .

- Wyznaczyć : a) funkcję gęstości  $f(x, y, z)$ ,  
 b) współczynnik korelacji  $\rho_{(X, Z)}$ ,  
 c)  $P(-1 < X < 2 | Y - Z = 1)$ .

a) Ponieważ  $\mathbf{V}^{-1} = \begin{bmatrix} \frac{4}{15} & 0 & \frac{1}{15} \\ 0 & \frac{1}{4} & 0 \\ \frac{1}{15} & 0 & \frac{4}{15} \end{bmatrix}$  natychmiast otrzymujemy funkcję gęstości

$$f(\mathbf{x}, y, z) = \frac{1}{(2\pi)^{3/2} \sqrt{60}} e^{-\frac{1}{120}(16x^2 + 15y^2 + 16(z-2)^2 + 8x(z-2))}$$

b)  $\rho_{(X, Z)} = \frac{-1}{4}$

c) Z równości  $\begin{bmatrix} X \\ Y - Z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$  widać, że dwuwymiarowa zmienna losowa  $\begin{bmatrix} X \\ Y - Z \end{bmatrix}$  ma

dwuwymiarowy rozkład normalny o parametrach  $\begin{bmatrix} 0 \\ -2 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 1 & 8 \end{bmatrix}$ , skąd łatwo uzyskujemy warunkowy

rozkład zmiennej  $X$  pod warunkiem  $Y - Z = 1$  jako rozkład normalny o wartości oczekiwanej  $0 + 1 \frac{1}{8} (Y - Z + 2) = \frac{3}{8}$  i wariancji  $4 - \frac{1}{8} = \frac{31}{8}$  a stąd po standaryzacji uzyskujemy szukane

prawdopodobieństwo  $P(-1 < X < 2 | Y - Z = 1) = \Phi\left(\frac{11}{\sqrt{248}}\right) - \Phi\left(\frac{-13}{\sqrt{248}}\right) = 0.921$ .

## Twierdzenia graniczne

Niech  $X_1, X_2, \dots$  będzie ciągiem zmiennych losowych określonych na tej samej przestrzeni probabilistycznej  $(\Omega, \mathcal{S}, P)$ . Pojawia się naturalne pytanie: **Jak zdefiniować pojęcie granicy ciągu zmiennych losowych?** Ponieważ każda ze zmiennych losowych  $X_i$  jest (mierzalną) funkcją z  $\Omega \rightarrow \mathbb{R}$ , więc zbieżność ciągu zmiennych losowych  $X_1, X_2, \dots$  można rozumieć tak jak zbieżność ciągów funkcyjnych. Mówiąc o zbieżności punktowej ciągu funkcji rzeczywistych określonych na wspólnej dziedzinie, żądamy w każdym punkcie dziedziny zbieżności odpowiedniego ciągu liczbowego. W analizie matematycznej zbieżność punktowa jest w wielu przypadkach zbyt słaba i zastępuje się ją mocniejszą zbieżnością jednostajną. W rachunku prawdopodobieństwa żądanie zbieżności w każdym punkcie  $\omega$  przestrzeni  $\Omega$  jest zbyt ostre. Chcemy zdefiniować zbieżność ciągu zmiennych losowych nawet wówczas, jeżeli nie zachodzi ona dla pewnych punktów (zdarzeń elementarnych), które tworzą zbiór (zdarzenie losowe) o prawdopodobieństwie:

- równym 0
- bardzo małym dążącym do 0 gdy  $n \rightarrow \infty$

**Def.** (zbieżności wg. prawdopodobieństwa -  $P$ - $lim$ )

$$X_n \xrightarrow{P} X \Leftrightarrow \forall \varepsilon > 0 \lim_{n \rightarrow \infty} P(\{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\}) = 1$$

**Def.** (zbieżności z prawdopodobieństwem 1- prawie na pewno-  $p.n.$ - $lim$ )

$$X_n \xrightarrow{p.n.} X \Leftrightarrow P(\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1$$

Dla zmiennych losowych posiadających (skończone) momenty rzędu  $r$  wprowadzamy

**Def.** (zbieżności wg.  $r$ -średniej-  $L_r$ - $lim$ )

$$X_n \xrightarrow{L_r} X \Leftrightarrow \lim_{n \rightarrow \infty} E |X_n(\omega) - X(\omega)|^r = 0$$

**Def.** (zbieżności według dystrybuant – słaba zbieżność-  $d$ - $lim$ )

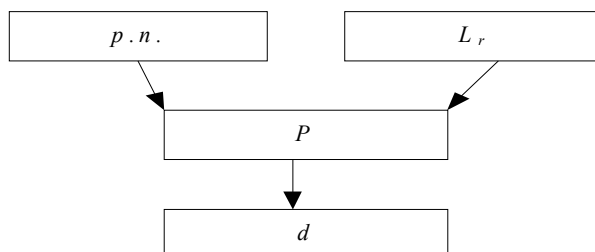
$$X_n \xrightarrow{d} X \Leftrightarrow \lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \text{ w każdym punkcie ciągłości dystrybuanty granicznej } F_X.$$

Interesującym faktem jest to, że ciąg dystrybuant może być zbieżny do funkcji, która nie ma własności dystrybuanty np. dla ciągu zmiennych losowych  $X_1, X_2, \dots$  o rozkładach jednopunktowych  $P(X_n = -n) = 1$  ciąg dystrybuant jest zbieżny do funkcji stałej  $F(x) \equiv 1$ , która nie jest dystrybuantą. Żądanie w definicji słabej zbieżności, aby granica ciągu dystrybuant była dystrybuantą jest więc istotne.

Każdy z typów zbieżności  $P$ - $lim$ ,  $p.n.$ - $lim$ ,  $L_r$ - $lim$  reprezentuje dla dostatecznie dużych  $n$  pewien rodzaj aproksymacji zmiennych losowych  $X_n$  i  $X$  jako funkcji  $\omega$ . Oznacza to, że rozkłady zmiennych  $X_n$  i  $X$  nie mogą

się zbyt różnić, a zatem powinna mieć miejsce aproksymacja według rozkładu. Z drugiej strony zbieżność  $d$ -lim zależy tylko od dystrybuant i odpowiednie zmienne losowe nie muszą być sobie bliskie jako funkcje  $\omega$ . W rzeczywistości słaba zbieżność nawet nie wymaga tego, aby zmienne losowe były określone na tej samej przestrzeni probabilistycznej.

Relacje pomiędzy różnymi rodzajami zbieżności przedstawia poniższy schemat



**Uwaga 1.** Korzystając z ciągłości miar probabilistycznych (wł. 8 i 9 str.3) można pokazać, że pojawiające się tam monotoniczne przejście do granicy można uogólnić zastępując indeks  $n$  indeksem ciągłym. Przypuśćmy, na przykład, że  $A_t \in \Sigma$  dla  $t > 0$ ,  $A_s \subset A_t$  dla  $s < t$  i  $A = \bigcup_{t>0} A_t$ . Jeżeli  $t_n \rightarrow \infty$ , to  $A_{t_n} \uparrow A$  i  $P(A_{t_n}) \uparrow P(A)$  z ciągłości miar probabilistycznych. Ale prawdopodobieństwa  $P(A_t)$  są niemalejące względem  $t$  i dlatego dla  $t \rightarrow \infty$  mają granicę równą  $\lim_{n \rightarrow \infty} P(A_{t_n})$ . Stąd zbieżność  $A_t \uparrow A$  implikuje zbieżność  $P(A_t) \uparrow P(A)$ . Ponadto z monotoniczności wynika  $\bigcup_{t>0} A_t = \bigcup_n A_{t_n} \in \Sigma$ , co dowodzi mierzalności granicznego zbioru (por. Billingsley str. 36)

**Uwaga 2.** Z uwagi 1 i definicji granicy wynika, że równoważność

$$P(\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1 \Leftrightarrow \forall \varepsilon > 0 \lim_{n \rightarrow \infty} P(\bigcap_{m=n}^{\infty} \{\omega : |X_m(\omega) - X(\omega)| < \varepsilon\}) = 1$$

z której w oczywisty sposób wynika, że zbieżność  $p.n.$  pociąga za sobą zbieżność według  $P$ .

**Uwaga 3.** Naśladując dowód nierówności Czebyszewa uzyskujemy nierówność  $E(|X-E(X)|^r) \geq \varepsilon^r P(|X-E(X)| < \varepsilon)$  z której natychmiast wynika, że zbieżność  $L_r$  pociąga za sobą zbieżność według  $P$ . Ponadto z nierówności Jensena  $E(g(X)) \geq g(E(X))$ ,  $g$ -funkcja wypukła) wynika, że  $r > s \Rightarrow$  zbieżność  $L_r$  implikuje zbieżność  $L_s$ .

**Uwaga 4.** Można pokazać, że zbieżność według prawdopodobieństwa do zera jest równoważna słabej zbieżności do zera, czyli zbieżności ciągu dystrybuant  $F_n$  do dystrybuanty  $F$  rozkładu jednopunktowego skupionego w zerze w każdym punkcie różnym od zera. Rzeczywiście, zakładając, że ciąg zmiennych losowych  $X_n$  jest zbieżny wg. prawdopodobieństwa do zera, otrzymujemy dla dowolnego  $\varepsilon > 0$

$$F_n(\varepsilon) = P(X_n < \varepsilon) \geq P(|X_n| < \varepsilon) \rightarrow 1, \text{ gdy } n \rightarrow \infty, F_n(-\varepsilon) = P(X_n < -\varepsilon) \leq P(|X_n| > \varepsilon) \rightarrow 0, \text{ gdy } n \rightarrow \infty,$$

co dowodzi słabej zbieżności ciągu  $X_n$ .

Podobnie zakładając zbieżność ciągu dystrybuant  $F_n$  do dystrybuanty  $F$  rozkładu jednopunktowego skupionego w zerze w każdym punkcie różnym od zera mamy dla dowolnego  $\varepsilon > 0$

$$P(|X_n| \geq \varepsilon) = P(X_n \geq \varepsilon) + P(X_n \leq -\varepsilon) \geq 1 - P(X_n < \varepsilon) + P(X_n < -\varepsilon/2) = 1 - F_n(\varepsilon) + F_n(-\varepsilon/2) \rightarrow 0, \text{ gdy } n \rightarrow \infty.$$

W ogólnym przypadku zbieżność według  $P$  jedynie implikuje zbieżność  $d$  (wniosek z tw. Slutskiego–por. Serfling str. 28). W przypadku słabej zbieżności nie można jej zdefiniować za pomocą zbieżności do zera ciągu odpowiednich różnic.

Podstawowymi zagadnieniami granicznymi są własności sum zmiennych losowych, gdy liczba składników dąży do nieskończoności, czyli własności szeregów zmiennych losowych. Rozważa się sumy

postaci  $a_n \sum_{k=1}^n X_k$  będące iloczynami dwu czynników, z których jeden dąży do zera a wariancja

drugiego dąży do nieskończoności. Wśród twierdzeń granicznych dotyczących powyższych sum szczególną rolę odgrywają te twierdzenia w których rozkładem granicznym jest rozkład jednopunktowy-

są to prawa wielkich liczb (**PWL**), oraz te w których rozkładem granicznym jest rozkład normalny- są to centralne twierdzenia graniczne (**CTG**). Pomocniczą i wielce użyteczną rolę odgrywają również prawa iterowanego logarytmu (**PIL**) charakteryzujące ekstremalne fluktuacje ciągu średnich lub sum częściowych, które tu nie będą omawiane.

### **Prawa wielkich liczb (PWL)**

Niech  $X_1, X_2, \dots$  będzie ciągiem zmiennych losowych określonych na tej samej przestrzeni probabilistycznej  $(\Omega, \Sigma, P)$  i posiadających (skończone) wartości oczekiwane  $E(X_i)$ .

Jeżeli  $\frac{1}{n} \sum_{i=1}^n (X_i - E(X_i)) \xrightarrow{P} 0$ , to mówimy, że dla ciągu  $X_1, X_2, \dots$  zachodzi słabe prawo wielkich liczb (**SPWL**).

Jeżeli  $\frac{1}{n} \sum_{i=1}^n (X_i - E(X_i)) \xrightarrow{p.n.} 0$ , to mówimy, że dla ciągu  $X_1, X_2, \dots$  zachodzi mocne prawo wielkich liczb (**MPWL**).

Bezpośrednio z nierówności Czebyszewa wynika następujące:

**SPWL Markowa.** Jeżeli dla ciągu  $X_1, X_2, \dots$  jest spełniony warunek (Markowa)

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} V \left( \sum_{i=1}^n X_i \right) = 0, \text{ to dla ciągu } X_1, X_2, \dots \text{ zachodzi SPWL.}$$

Jego szczególnym przypadkiem jest

**SPWL Czebyszewa.** Jeżeli zmienne losowe  $X_1, X_2, \dots$  są niezależne a ich wariancje są wspólnie ograniczone (tzn.  $\exists \sigma^2 \forall i V(X_i) < \sigma^2$ ), to dla ciągu  $X_1, X_2, \dots$  zachodzi SPWL.

Znacznie trudniejsze do udowodnienia (por. Billingsley) są następujące:

**MPWL Kołmogorowa.** Jeżeli  $X_1, X_2, \dots$  jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie (i.i.d) o skończonej wartości oczekiwanej, to dla ciągu  $X_1, X_2, \dots$  zachodzi MPWL.

**MPWL Kołmogorowa.** Jeżeli  $X_1, X_2, \dots$  jest ciągiem niezależnych zmiennych losowych o wariancjach spełniających warunek  $\sum_{n=1}^{\infty} \frac{V(X_n)}{n^2} < \infty$ , to dla ciągu  $X_1, X_2, \dots$  zachodzi MPWL.

Zastępując czynnik  $a_n = \frac{1}{n}$  wolniej zbieżnym do zera czynnikiem  $a_n = \frac{1}{\sqrt{n}}$  otrzymujemy przy różnych dodatkowych założeniach centralne twierdzenia graniczne z których szczególnie użyteczne w statystyce matematycznej jest następujące:

**Centralne twierdzenie graniczne CTG Lindeberga-Levy'ego.** Niech  $X_1, X_2, \dots$  jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie (i.i.d) z wartością oczekiwaną  $m$  i wariancją  $\sigma^2$ . Niech  $U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - m}{\sigma}$  i  $F_{U_n}(u) = P(U_n < u)$ . Wówczas

$$\forall u \quad \lim_{n \rightarrow \infty} F_{U_n}(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{x^2}{2}} dx$$

Uogólnienia podstawowego CTG Lindeberga-Levy'ego idą w różnych kierunkach:

- tw. Lapunowa zastępuje założenie identyczności rozkładów pewnymi warunkami na momenty
- CTG dla zależnych zmiennych losowych
- CTG w języku procesów stochastycznych (Donsker)

**Przykład.** Waga pasażerów samolotów jest pewną zmienną losową o wartości oczekiwanej  $\mu_1=70$  kg i odchyleniu standardowym  $\sigma_1=8$  kg. Także całkowity ciężar bagażu pasażera (tzn. łącznie z bagażem ręcznym) jest zmienną losową o wartości oczekiwanej  $\mu_2=21$  kg i odchyleniu standardowym  $\sigma_2=5$  kg. Zakładając, że powyższe zmienne losowe są niezależne obliczyć prawdopodobieństwo, że 292 osoby łącznie z bagażem nie ważą więcej niż 26500 kg.

**Oznaczenia:**

$X_i$  – waga  $i$ -tego pasażera

$Y_i$  – waga bagażu  $i$ -tego pasażera

$W_i = X_i + Y_i$  – łączna waga  $i$ -tego pasażera z bagażem

$X_i, Y_i, i=1, \dots, n$  niezależne

Z powyższych faktów wynika, że  $W_i, i=1, \dots, n$  jest ciągiem niezależnych zmiennych losowych o tym samym (nieznanym) rozkładzie z wartością oczekiwaną  $\mu = E(W_i) = \mu_1 + \mu_2$  i wariancją  $\sigma^2 = V(W_i) = \sigma_1^2 + \sigma_2^2$ .

$$P\left(\sum_{i=1}^{292} W_i < 26500\right) = P\left(\frac{1}{\sqrt{292}} \sum_{i=1}^{292} \frac{W_i - \mu}{\sigma} < \frac{1}{\sqrt{292}} \frac{26500 - 292\mu}{\sigma}\right) = P(U_n < -0.4546) = 0.32$$

## Literatura

1. Jakubowski J, Sztencel R. Rachunek prawdopodobieństwa dla prawie każdego, SCRIPT 2006
2. Jakubowski J, Sztencel R. Wstęp do rachunku prawdopodobieństwa, SCRIPT 2001
3. Billingsley P.: Prawdopodobieństwo i miara, PWN, 1987
4. Hartman S., Mikusiński J.: Teoria miary i całki Lebesgue'a. PWN, 1957
5. Plucińska A., Pluciński E.: Elementy probabilistyki. PWN, 1979
6. Rohatgi V., K.: Statistical inference, Wiley, 1984
7. Serfling R.: Twierdzenia graniczne statystyki matematycznej. PWN, 1991

## Dodatek 1. Wybrane porządkowe własności zbiorów liczbowych

$(N, \leq)$  jest dobrze uporządkowany  $\Leftrightarrow \forall_{A \subset N} \exists_{nm} A$

(tzn. w każdym podzbiorku zbioru  $N$  istnieje element najmniejszy)

$(Q, \leq)$  jest gęsto uporządkowany  $\Leftrightarrow \forall_{x,y \in Q} \exists_{z \in Q} : x < z < y$  (uwaga  $a < b \Leftrightarrow a \leq b$  i  $a \neq b$ )

(tzn. pomiędzy dwoma liczbami wymiernymi zawsze można znaleźć trzecią liczbę wymierną)

$(R, \leq)$  jest uporządkowany w sposób ciągły :

### Zasada ciągłości zbioru liczb rzeczywistych

- każdy ograniczony od góry podzbiór zbioru liczb rzeczywistych posiada kres górny
- każdy ograniczony od dołu podzbiór zbioru liczb rzeczywistych posiada kres dolny.

Definicje kresów podzbiorów zbioru  $R$

$$a = \sup A \Leftrightarrow \begin{cases} \forall_{x \in A} : x \leq a \\ \forall_{\varepsilon > 0} \exists_{y \in A} a - \varepsilon \leq y \end{cases} \quad a = \inf A \Leftrightarrow \begin{cases} \forall_{x \in A} : x \geq a \\ \forall_{\varepsilon > 0} \exists_{y \in A} a + \varepsilon \geq y \end{cases}$$

Zastosowanie zasady ciągłości do zdefiniowania potęgi o dowolnym wykładniku  $a \in R_+$ .

- potęga o wykładniku całkowitym  $a^n = \underbrace{a \cdots a}_{n\text{-razy}}$ ,  $a^{-n} = \frac{1}{a^n}$ ,  $a^0 = 1$
- potęga o wykładniku wymiernym  $a^{\frac{1}{n}} = \inf\{x \in Q : x > 0 \wedge x^n \geq a\}$ ;  $a^m = \left(a^{\frac{1}{n}}\right)^m$
- potęga o wykładniku rzeczywistym  
dla  $a > 1$   $a^\beta = \inf\{a^x : x \in Q \wedge x \geq \beta\}$ ; dla  $0 < a < 1$   $a^\beta = \left(\frac{1}{a}\right)^{-\beta}$

Uwagi o zasadzie indukcji matematycznej  $\{T(n_0) \wedge (\forall_{n \geq n_0} T(n) \Rightarrow T(n+1))\} \Rightarrow \forall_{n \geq n_0} T(n)$

Zasada indukcji matematycznej jest konsekwencją dobrego uporządkowania zbioru  $N$  liczb naturalnych, czyli w każdym zbiorze dobrze uporządkowanym prawdziwa jest zasada indukcji matematycznej.

**Dowód.** Załóżmy, że  $\{T(n_0) \wedge (\forall_{n \geq n_0} T(n) \Rightarrow T(n+1))\}$ , czyli że prawdziwy jest poprzednik implikacji w zasadzie indukcji matematycznej. Niech  $Z = \{n \geq n_0 : T(n) \text{ nie jest prawdziwa}\}$ . Jeżeli  $Z = \emptyset$ , to zachodzi zasada indukcji matematycznej. Jeżeli  $Z \neq \emptyset$ , to w  $Z$  istnieje element najmniejszy (dobre uporządkowanie  $Z$ ) powiedzmy  $m_0 \in Z$ . **Uwaga.**  $m_0 > n_0$  bo  $T(n_0)$  jest prawdziwe. Ponieważ  $m_0 - 1 \notin Z$  więc prawdziwe jest  $T(m_0 - 1)$  a z założenia otrzymujemy, że prawdziwe jest  $T(m_0)$ , czyli  $m_0 \notin Z$  – sprzeczność !

## Dodatek 2. Rodziny zbiorów i działania uogólnione

$X$  – ustalony zbiór (przestrzeń),  $T \neq \emptyset$  niepusty zbiór indeksów. Funkcja

$$T \ni t \rightarrow A_t \subset X$$

określa rodzinę zbiorów  $\{A_t\}_{t \in T}$  indeksowaną indeksami  $t$  ze zbioru indeksów  $T$ .

**Suma (unia) zbiorów rodziny  $\{A_t\}_{t \in T}$  :**

$$\bigcup_{t \in T} A_t = \{x \in X : \exists_{t \in T} x \in A_t\}$$

zbiór wszystkich tych elementów, które należą do przynajmniej jednego zbioru  $A_t$  z rodziny  $\{A_t\}_{t \in T}$

**Iloczyn** (przekrój-część wspólna) **zbiorów rodziny**  $\{A_t\}_{t \in T}$  :

$$\bigcap_{t \in T} A_t = \{x \in X : \forall_{t \in T} x \in A_t\} -$$

zbiór wszystkich tych elementów, które należą do wszystkich zbiorów  $A_t$  rodziny  $\{A_t\}_{t \in T}$ .

**Przykład.** a)  $\bigcup_{n=1}^{\infty} [1 + \frac{1}{n}, 2 + \frac{1}{n}] = (1, 3]$  ,      b)  $\bigcap_{n=1}^{\infty} (3 - \frac{3}{n}, 5 + \frac{1}{n}] = [3, 5]$  ,

c)  $\bigcup_{x \in (0,1)} [x^2 + x + 1, x + 4] = (1,5)$  ,      d)  $\bigcap_{x \in R} (-x^2 + 2x - 3, x^2 + 10] = (-2,10]$ .

Rozważmy ciąg zbiorów  $(A_n)_{n \in IN}$ , będących podzbiarami zbioru  $\Omega$ .

**Def.** Granicą górną  $\limsup_{n \rightarrow \infty} A_n$  ciągu zbiorów  $(A_n)_{n \in IN}$  nazywamy zbiór  $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$  .

Granicą dolną  $\liminf_{n \rightarrow \infty} A_n$  ciągu zbiorów  $(A_n)_{n \in IN}$  nazywamy zbiór  $\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$  .

Łatwo zauważyć, że

$$x \in \limsup_{n \rightarrow \infty} A_n \Leftrightarrow \forall n \in IN \exists m \geq n \ x \in A_m ,$$

co oznacza, że granica górna ciągu zbiorów jest zbiorem składającym się z elementów, które należą do nieskończenie wielu zbiorów ciągu  $(A_n)_{n \in IN}$ .

Podobnie

$$x \in \liminf_{n \rightarrow \infty} A_n \Leftrightarrow \exists n \in IN \ \forall m \geq n \ x \in A_m ,$$

co oznacza, że granica dolna ciągu zbiorów jest zbiorem składającym się z elementów, które należą do prawie wszystkich zbiorów ciągu  $(A_n)_{n \in IN}$ .

Bezpośrednio z definicji widać, że

$$\bigcap_{n=1}^{\infty} A_n \subset \liminf_{n \rightarrow \infty} A_n \subset \limsup_{n \rightarrow \infty} A_n \subset \bigcup_{n=1}^{\infty} A_n .$$

Jeżeli  $\liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n$ , to mówimy, że ciąg zbiorów  $(A_n)_{n \in IN}$  jest zbieżny do granicy

$$\lim_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n .$$

**Przykład.**  $A_n = (\frac{-1}{n}, 1 + \frac{(-1)^n}{n})$ ,  $\liminf_{n \rightarrow \infty} A_n = [0, 1)$ ,  $\limsup_{n \rightarrow \infty} A_n = [0, 1]$ . (koniec uzupełnień)

### Dodatek 3. Równoliczność zbiorów

**Def:**  $A \text{ rl } B \Leftrightarrow \exists f: A \xrightarrow{1:1} B$  (istnieje bijekcja jednego zbioru na drugi)

**Przykład.**  $\{a,b,c\}$  i  $\{1,2\}$  – nie są równoliczne

**Przykład.**  $A=N$  i  $B=\{x \in N : x = 2k - 1, k \in N\} = \{1,3,5,\dots\}$  – są równoliczne, bo istnieje bijekcja  $f(k) = 2k - 1$  zbioru  $A$  na zbiór  $B$ .

**Def:** Zbiór  $A$  jest **skończony** jeżeli jest równoliczny ze zbiorem  $\{1,\dots,n\}$  dla  $n \in N$ .

Zbiór  $A$  jest **nieskończony**  $\Leftrightarrow \sim(A \text{ jest skończony})$

#### Charakteryzacja zbiorów nieskończonych

Zbiór  $A$  jest **nieskończony**  $\Leftrightarrow \exists B \subset A, B \neq A: A \text{ rl } B$

(czyli zbiór nieskończony jest równoliczny ze swoim podzbiorem właściwym).

#### Twierdzenie Cantora –Bernsteina

$$\{A \text{ rl } B_0 \subset B \wedge B \text{ rl } A_0 \subset A\} \Rightarrow A \text{ rl } B$$

**Def:** Zbiór  $A$  jest **przeliczalny**  $\Leftrightarrow A \text{ rl } N$  (inaczej elementy zbioru przeliczalnego można ponumerować liczbami naturalnymi)

Zbiór  $A$  jest **co najwyżej przeliczalny**  $\Leftrightarrow A$  jest skończony lub przeliczalny

Zbiór  $A$  jest **nieprzeliczalny**  $\Leftrightarrow A$  jest nieskończony i  $\sim(A \text{ rl } N)$

**Fakty:** (metodą przekątniową można pokazać, że)

- Iloczyn kartezjański zbiorów przeliczalnych jest przeliczalny.

Zbiory  $A$  i  $B$  są przeliczalne więc  $A = \{a_1, a_2, a_3, \dots\}$ ,  $B = \{b_1, b_2, b_3, \dots\}$

$$A \times B = \begin{matrix} (a_1, b_1)_1 & (a_1, b_2)_2 & (a_1, b_3)_4 & (a_1, b_4)_7 & \dots \\ (a_2, b_1)_3 & (a_2, b_2)_5 & (a_2, b_3)_8 & (a_2, b_4)_{12} & \dots \\ (a_3, b_4)_6 & (a_3, b_2)_9 & (a_3, b_3)_{13} & (a_3, b_4)_{18} & \dots \\ (a_4, b_1)_{10} & (a_4, b_2)_{14} & (a_4, b_3)_{19} & (a_4, b_4)_{25} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix}$$

$$f(i, j) = \frac{(i+j-1)(i+j-2)}{2} + i \text{ jest bijekcją } N \times N \text{ na } N \text{ ( znaleźć } f^{-1} \text{ - ćwiczenia)}$$

(Wskazówka: element  $(i,j)$  jest  $i$ -tym elementem na na  $(i+j-1)$ -szej przekątnej )

- $Q$  jest przeliczalny, (jako wniosek, bo  $Q \subset Z \times N$  (przel.) i  $N \subset Q$ )



- Suma (mnożośćowa) przeliczalnej ilości zbiorów przeliczalnych jest zbiorem przeliczalnym  
Dowód jak dla iloczynu kartezjańskiego
- $\mathbb{R}$  jest nieprzeliczalny ( $\mathbb{R}$  r/l  $(0,1)$  a  $(0,1)$  jest nieprzeliczalny- dowód nie wprost metodą przekątniową lub zasadą szufladkową Dirichleta)

Funkcja  $\arctg(\frac{x}{\pi} + \frac{1}{2})$  jest bijekcją zbioru  $\mathbb{R}$  na przedział  $(0,1)$ . Po przyjęciu umowy, że te rozwinięcia dziesiętne liczb wymiernych, które są skończone zastępujemy rozwinięciami nieskończonymi mamy wzajemnie jednoznaczność między liczbami rzeczywistymi z przedziału  $(0,1)$  a ciągami cyfr rozwinięcia dziesiętnego).

Dla dowodu nie wprost założmy, że wszystkie liczby rzeczywiste z przedziału  $(0,1)$  można ustawić w ciąg (ponumerować liczbami naturalnymi)

$$a_1 = 0, a_{1,1} a_{1,2} a_{1,3}, \dots$$

$$a_2 = 0, a_{2,1} a_{2,2} a_{2,3}, \dots$$

$$a_3 = 0, a_{3,1} a_{3,2} a_{3,3}, \dots$$

⋮

Rozważmy liczbę  $c = 0, c_1 c_2 c_3 \dots$  z przedziału  $(0,1)$  taką, że  $c_i \neq a_{i,i}$ . Liczba  $c$  różni się od każdej z liczb  $a_i$   $i$ -tą cyfrą rozwinięcia dziesiętnego, więc nie występuje w tym ciągu - sprzeczność!

## Dodatek 4. Całka względem rozkładu prawdopodobieństwa(względem miary). (Plucińska)

Całka względem rozkładu prawdopodobieństwa to całka względem miary probabilistycznej. Jej konstrukcja jest podobna do konstrukcji klasycznej całki Lebesgue'a znanej z kursu analizy matematycznej. Całka Lebesgue'a to całka względem miary Lebesgue'a (czyli nieujemnej, zerującej się na zbiorze pustym, przeliczalnie addytywnej funkcji zbioru, która przedziałowi przypisuje jego długość). W przypadku całki funkcji jednej zmiennej rzeczywistej całki Riemanna i Lebesgue'a można określić jako granice odpowiednich sum całkowych. Zasadnicza różnica pomiędzy całką Riemanna a całką Lebesgue'a sprowadza się do metody segregacji punktów na osi  $x$ . Metoda (L) polega na segregacji punktów na osi  $x$  według wartości, które funkcja na nich przyjmuje, podczas gdy w metodzie (R) grupuje się punkty w kolejne przedziały wykorzystując naturalny porządek na osi  $x$ . Przyrównywano metodę Riemanna do liczenia pieniędzy rozdzielonych byle jak na kupki, drogą przeliczania ich po kolei w każdej kupce i dodawania, a metodę Lebesgue'a do rozdzielania monet na kupki według ich wartości i zsumowania wartości poszczególnych kucek (**Hartman, Mikusiński**). Różnice między riemannowską a lebesgue'owską definicją całki można dostrzec, gdy funkcja jest przedziałami stała. Weźmy np. funkcję równą 1 w przedziałach  $[0, 1)$  i  $[4, 5)$  2 w przedziałach  $[1, 2)$  i  $[3, 4)$  i 3 w przedziale  $[2, 3)$ .

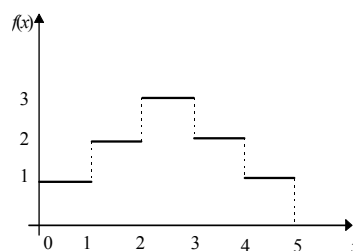
Obliczamy całkę (sumę) Riemanna (podział odcinka  $[0, 5]$  na osi  $x$

$$1 \cdot (1-0) + 2 \cdot (2-1) + 3 \cdot (3-2) + 2 \cdot (4-3) + 1 \cdot (5-4) = 9$$

Obliczamy całkę (sumę) Lebesgue'a (podział odcinka  $[1, 3+\delta]$  na osi  $y$

$$1 \cdot [(1-0) + (5-4)] + 2 \cdot [(2-1) + (4-3)] + 3 \cdot (3-2) = 9.$$

Całki te są sobie równe. Dowodzi się że gdy całka Riemanna istnieje, to istnieje też całka Lebesgue'a i obie całki są sobie równe. Ponieważ implikacja odwrotna jest fałszywa (funkcja Dirichleta) całka Lebesgue'a jest więc uogólnieniem całki Riemanna. Największa korzyść z wprowadzenia całki Lebesgue'a polega na swobodzie dokonywania przejść



granicznych. Jedną z największych trudności spotykanych w teorii Riemanna, polega na tym, że granica ciągu funkcji całkownych w sensie Riemanna (lub nawet ciągłych) może nie być całkowna w sensie Riemanna. W teorii Lebesgue'a trudność ta jest prawie całkowicie wyeliminowana, bo granice funkcji mierzalnych są zawsze mierzalne.

Niech  $(\Omega, \Sigma, P)$  będzie przestrzenią probabilistyczną a  $X: \Omega \rightarrow \mathcal{C} \subset \mathbb{R}$  zmienną losową określoną na tej przestrzeni.

Funkcją charakterystyczną (indykatorem) zbioru  $A$  nazywamy funkcję  $I_A(\omega) = \begin{cases} 1, & \text{dla } \omega \in A \\ 0, & \text{dla } \omega \notin A \end{cases}$ .

Zmienną losową nazywamy **prostą**, jeżeli zbiór  $\mathcal{C}$  jej wartości jest **skończony** to znaczy  $\mathcal{C} = \{x_1, \dots, x_n\}$ .

Każdą zmienną losową prostą można zapisać w postaci

$$X(\omega) = \sum_{k=1}^n x_k I_{A_k}(\omega),$$

gdzie  $A_k$  jest odpowiednio dobranym układem zupełnym zdarzeń tzn.  $\Omega = \bigcup_{k=1}^n A_k$ ,  $i \neq j \Rightarrow A_i \cap A_j = \emptyset$ .

### 1. Całka zmiennej losowej prostej.

Całkę zmiennej losowej prostej względem rozkładu (miary)  $P$  po zbiorze  $\Omega$  definiujemy wzorem

$$\int_{\Omega} X(\omega) P(d\omega) = \sum_{k=1}^n x_k P(\{\omega : X(\omega) = x_k\}) = \sum_{k=1}^n x_k p_k$$

a po zbiorze  $A \in \mathcal{S}$  wzorem

$$\int_A X(\omega) P(d\omega) = \int_{\Omega} X(\omega) I_A(\omega) P(d\omega).$$

## 2. Całka zmiennej losowej ograniczonej.

Jeżeli  $X$  jest **nieujemną** zmienną losową **ograniczoną**, to ma następujące własności:

**W1.** Istnieje niemalejący ciąg  $\{X_n\}$  zmiennych losowych prostych zbieżnych do zmiennej losowej  $X$ .

**W2.** Jeśli  $\{X_n'\}$  i  $\{X_n''\}$  są niemalejącymi ciągami zmiennych losowych prostych zbieżnych do tej samej zmiennej losowej  $X$ , to (Kołodziej str. 300)

$$\lim_{n \rightarrow \infty} \int_{\Omega} X_n' P(d\omega) = \lim_{n \rightarrow \infty} \int_{\Omega} X_n'' P(d\omega)$$

Całkę **nieujemnej ograniczonej** zmiennej losowej  $X$  względem rozkładu  $P$  po zbiorze  $\Omega$  definiujemy wzorem

$$\int_{\Omega} X(\omega) P(d\omega) = \lim_{n \rightarrow \infty} \int_{\Omega} X_n(\omega) P(d\omega),$$

gdzie  $\{X_n\}$  jest dowolnym niemalejącym ciągiem zmiennych losowych prostych zbieżnym do zmiennej losowej  $X$ . Aby zdefiniować całkę z **dowolnej** zmiennej losowej **ograniczonej** wykorzystujemy dekompozycję:

$X(\omega) = X_+(\omega) - X_-(\omega)$ , gdzie  $X_+(\omega) = \max\{X(\omega), 0\}$ ,  $X_-(\omega) = \max\{0, -X(\omega)\}$ . Stąd

$$\int_{\Omega} X(\omega) P(d\omega) = \int_{\Omega} X_+(\omega) P(d\omega) - \int_{\Omega} X_-(\omega) P(d\omega).$$

Analogicznie definiujemy całkę po zbiorze  $A$ . Całka zdefiniowana w powyższy sposób ma następujące własności:

**W3.** Jeżeli  $A_k$  stanowią zupełny układ zdarzeń, to

$$\int_{\Omega} X(\omega) P(d\omega) = \sum_k \int_{A_k} X(\omega) P(d\omega).$$

**W4.** Całka jest funkcjonałem liniowym

$$\int_{\Omega} (\alpha X_1(\omega) + \beta X_2(\omega)) P(d\omega) = \alpha \int_{\Omega} X_1(\omega) P(d\omega) + \beta \int_{\Omega} X_2(\omega) P(d\omega).$$

**W5.** Jeżeli  $X_1(\omega) \leq X_2(\omega)$  dla każdego  $\omega$ , to  $\int_{\Omega} X_1(\omega) P(d\omega) \leq \int_{\Omega} X_2(\omega) P(d\omega)$ .

**W6.** Wartość bezwzględna całki jest nie większa od całki z bezwzględnej wartości tzn.

$$\left| \int_{\Omega} X(\omega) P(d\omega) \right| \leq \int_{\Omega} |X(\omega)| P(d\omega).$$

## 3. Całka zmiennej losowej w ogólnym przypadku

Dla  $X \geq 0$  całkę definiujemy wzorem

$$\int_{\Omega} X(\omega) P(d\omega) = \lim_{n \rightarrow \infty} \int_{\{\omega: X(\omega) \leq n\}} X(\omega) P(d\omega)$$

o ile ta granica istnieje. Zmienną losową  $X$  przyjmującą dowolne wartości przedstawiamy w postaci  $X(\omega) = X_+(\omega) - X_-(\omega)$ , a następnie przyjmujemy

$$\int_{\Omega} X(\omega)P(d\omega) = \int_{\Omega} X_+(\omega)P(d\omega) - \int_{\Omega} X_-(\omega)P(d\omega)$$

ile obie całki po prawej stronie istnieją i przynajmniej jedna jest skończona. Jeżeli obie całki są skończone, to zmienną losową nazywamy **sumowalną**. ( $X$  sumowalna  $\Leftrightarrow \int_{\Omega} |X(\omega)| P(d\omega) < \infty$ ).

Bezpośrednio z definicji całki i definicji rozkładu  $P_X$  zmiennej losowej  $X$  wynika, że

$$\int_{\Omega} X(\omega)P(d\omega) = \int_{\mathfrak{R}} xP_X(dx)$$

Można wykazać, że jeśli  $X$  jest dyskretną zmienną losową, to

$$\int_{\Omega} X(\omega)P(d\omega) = \sum_k x_k P(\{\omega : X(\omega) = x_k\}) = \sum_k x_k p_k, \text{ o ile szereg jest zbieżny.}$$

a jeśli  $X$  jest zmienną losową (absolutnie) ciągłą o funkcji gęstości  $f$ , to

$$\int_{\Omega} X(\omega)P(d\omega) = \int_{\mathfrak{R}} x f(x) dx \quad \text{o ile całka istnieje.}$$

Jeżeli  $X$  jest zmienną losową o rozkładzie mieszanym  $P=c_1P_{\text{dyskr}}+c_2P_{\text{cg}}$ , to

$$\int_{\Omega} X(\omega)P(d\omega) = c_1 \int_{\Omega} X(\omega)P_{\text{dyskr}}(d\omega) + c_2 \int_{\Omega} X(\omega)P_{\text{cg}}(d\omega).$$

**Twierdzenie Lebesgue'a o zbieżności monotonicznej.** Jeżeli  $(X_n)$  niemalejącym ciągiem nieujemnych zmiennych losowych (czyli  $0 \leq X_n \uparrow X$ ), to  $\lim_{n \rightarrow \infty} \int_{\Omega} X_n dP = \int_{\Omega} X dP$

**Twierdzenie Lebesgue'a o zbieżności ograniczonej.** Jeżeli dla pewnej całkowalnej zmiennej losowej  $Y$  zachodzą nierówności  $|X_n| \leq Y$ ,  $n \in \mathbb{N}$ , to wszystkie zmienne losowe  $X_n$  są całkowalne, a ponadto, jeśli  $X_n \xrightarrow{p.n.} X$ , to  $X$  jest całkowalna i

$$\lim_{n \rightarrow \infty} \int_{\Omega} |X_n - X| dP = 0 \quad (\text{inaczej } X_n \xrightarrow{L_1} X)$$

oraz

$$\lim_{n \rightarrow \infty} \int_{\Omega} X_n dP = \int_{\Omega} X dP$$

**Zamiana zmiennych.** (Billingsley 214) Niech  $(\Omega, \Sigma)$  oraz  $(\Omega', \Sigma')$  będą przestrzeniami mierzalnymi i niech  $T: \Omega \rightarrow \Omega'$  będzie odwzorowaniem  $\Sigma/\Sigma'$  mierzalnym. Dla miary  $\mu$  na  $\sigma$ -ciele  $\Sigma$  definiujemy miarę  $\nu = \mu \circ T^{-1}$  na  $\sigma$ -ciele  $\Sigma'$  wzorem  $\nu(A') = \mu(T^{-1}(A')) = \mu(T^{-1}(A'))$ . Załóżmy, że  $g$  jest  $\Sigma'$  mierzalną funkcją rzeczywistą na zbiorze  $\Omega'$ ; stąd złożenie  $g \circ T$  jest  $\Sigma$  mierzalną funkcją rzeczywistą na zbiorze  $\Omega$

**Tw.** Jeżeli  $g$  jest funkcją całkowalną, to

$$\int_{\Omega'} g(\omega') \nu(d\omega') = \int_{\Omega} g(T(\omega)) \mu(d\omega), \quad \int_{A'} g(\omega') \nu(d\omega') = \int_{T^{-1}(A')} g(T(\omega)) \mu(d\omega).$$

**Dowód.** Dla  $g=I_A$  twierdzenie sprowadza się do wzoru definiującego transport miary. Na mocy liniowości całki twierdzenie jest prawdziwe dla nieujemnych funkcji prostych a z tw. o zbieżności

monotonicznej jest prawdziwe dla nieujemnych funkcji  $g$ . Dla funkcji całkwalnej rozkładamy ją na część dodatnią i ujemną i stosujemy wcześniej uzyskany wynik.

**Uzupełnienie. Całkowanie przez sprowadzenie do całki iterowanej.**

Niech  $g$  będzie funkcją określoną na podzbiórze  $A$  przestrzeni  $R^n = R^{n_1} \times R^{n_2}$  ( $n = n_1 + n_2$ ).

Przyjmijmy oznaczenia:

$$A(x_1) = \{x_2 \in R^{n_2} : (x_1, x_2) \in A\} \text{ dla każdego } x_1 \in R^{n_1},$$

$$A_1 = \{x_1 \in R^{n_1} : A(x_1) \neq \emptyset\}$$

Jeżeli  $x_1 \in A_1$ , to  $x_2 \rightarrow g(x_1, x_2)$  jest dobrze określoną funkcją na zbiorze  $A(x_1)$ , całkę tej funkcji (o ile istnieje) oznaczamy przez

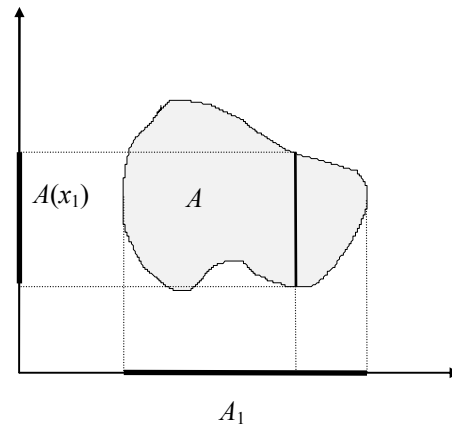
$$\bar{g} = \int_{A(x_1)} g(x_1, x_2) dx_2.$$

Jeżeli całka po prawej stronie istnieje przynajmniej prawie wszędzie na zbiorze  $A_1$ , to określona jest tym wzorem prawie wszędzie na zbiorze  $A_1$  funkcja  $\bar{g}$ . Całkę funkcji  $\bar{g}$  tzn.

$$\int_{A_1} \bar{g}(x_1) dx_1 = \int_{A_1} \left( \int_{A(x_1)} g(x_1, x_2) dx_2 \right) dx_1 \text{ nazywamy całką iterowaną.}$$

**Tw. (Fubniego)** Załóżmy, że zbiór  $A \subset R^n$  i jego rzut  $A_1$  na  $R^{n_1}$  są zbiorami mierzalnymi. Jeżeli funkcja  $g$  jest funkcją (L) całkwalną na  $A$ , to prawie dla każdego  $x_1 \in A_1$  funkcja  $x_2 \rightarrow g(x_1, x_2)$  jest (L) całkwalna na  $A(x_1)$  a określona prawie wszędzie na  $A_1$  funkcja  $\bar{g} = \int_{A(x_1)} g(x_1, x_2) dx_2$  jest (L) całkwalna na

$$A_1 \text{ i } \iint_A g(x_1, x_2) dx_1 dx_2 = \int_{A_1} dx_1 \int_{A(x_1)} g(x_1, x_2) dx_2.$$



**Dodatek 5.** Funkcja  $\Gamma$  Eulera jest dla  $x > 0$  definiowana wzorem

- $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$

Rozbijając całkę niewłaściwą na sumę całek  $\int_0^{\infty} t^{x-1} e^{-t} dt = \int_0^1 t^{x-1} e^{-t} dt + \int_1^{\infty} t^{x-1} e^{-t} dt$  można pokazać, że całka niewłaściwa występująca w definicji funkcji  $\Gamma$  jest zbieżna. Rzeczywiście dla  $t \in (0,1]$  i

dowolnego  $x > 0$  mamy  $0 \leq t^{x-1} e^{-t} \leq t^{x-1}$  i  $\int_0^1 t^{x-1} dt = \frac{1}{x} < \infty$ , natomiast dla  $t \in [1, \infty)$  i dowolnego

$x > 0$  prawdziwe jest oszacowanie  $0 \leq t^{x-1} e^{-t} \leq \frac{t^{[x]}}{e^t} \leq \frac{t^{[x]}}{t^{[x]+2}} = \frac{([x]+2)!}{t^2}$  i

$\int_1^{\infty} \frac{([x]+2)!}{t^2} dt = ([x]+2)!$ , więc całka  $\int_1^{\infty} t^{x-1} e^{-t} dt$  jest zbieżna.

Z definicji widać, że

- $\Gamma(1) = 1$

Całkując przez części łatwo pokazać, że

- $\Gamma(x+1) = x\Gamma(x)$

Rzeczywiście

$$\Gamma(x+1) = \int_0^{\infty} t^x e^{-t} dt = \left\{ \begin{array}{l} f(t) = t^x \quad g'(t) = e^{-t} \\ f'(t) = xt^{x-1} \quad g(t) = -e^{-t} \end{array} \right\} = -t^x e^{-t} \Big|_0^{\infty} + x \int_0^{\infty} t^{x-1} e^{-t} dt = x\Gamma(x).$$

Stąd  $\Gamma(n+1) = n!$

- $\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} t^{\frac{1}{2}-1} e^{-t} dt = \int_0^{\infty} \sqrt{t} e^{-t} dt = \int_0^{\infty} e^{-x^2} dx = 2 \int_0^{\infty} e^{-x^2} dx = 2 \frac{\sqrt{\pi}}{2} = \sqrt{\pi}$ , gdyż

$$I^2 = \int_0^{\infty} e^{-x^2} dx \int_0^{\infty} e^{-y^2} dy = \iint_{[0, \infty)^2} e^{-(x^2+y^2)} dx dy = \int_0^{\frac{\pi}{2}} d\phi \int_0^{\infty} e^{-r^2} r dr = \frac{\pi}{2} \int_0^{\infty} e^{-r^2} r dr = \frac{\pi}{4} \int_0^{\infty} e^{-u} du = \frac{\pi}{4}, \text{ a stąd}$$

$$I = \int_0^{\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}.$$

$\Gamma(x)$

