

# *Spis treści*

---

## Część I. Rachunek Prawdopodobieństwa

---

<b>1. Przestrzeń probabilistyczna</b>	3
1.1 Przykłady wstępne	3
1.2 Zdarzenia i prawdopodobieństwo – przypadek skończony	9
1.3 Przestrzeń probabilistyczna	12
1.4 Miara Lebesgue’a	18
1.5 Dystrybuanta	21
<b>2. Całka</b>	27
2.1 Całka Riemanna	27
2.2 Problem z definicją prawdopodobieństwa	31
2.3 Całka Lebesgue’a	33
2.4 Własności całki	36
2.5 Gęstość	39
2.6 Dystrybuanta	41
<b>3. Zmienne losowe</b>	47
3.1 Definicja zmiennej losowej	47
3.2 Rozkład zmiennej losowej	48
3.3 Gęstość	50
3.4 Przykłady finansowe	54
3.5 Wartość oczekiwana i wariancja	56
3.5.1 Przypadek dyskretny	56
3.5.2 Własności	59

3.5.3	Przypadek ogólny .....	61
3.5.4	Obliczanie .....	62
<b>4.</b>	<b>Zależności .....</b>	<b>69</b>
4.1	Niezależność .....	69
4.2	Zależność .....	77
<b>5.</b>	<b>Warunkowa wartość oczekiwana .....</b>	<b>83</b>
5.1	Model dwumianowy .....	83
5.2	Podziały i informacja .....	87
5.3	Warunkowa wartość oczekiwana – definicja .....	92
5.4	Własności warunkowej wartości oczekiwanej .....	96
5.5	Definicja w przypadku ogólnym .....	102
5.6	Własności* .....	104

---

## Część II. Procesy stochastyczne

---

<b>6.</b>	<b>Martyngały .....</b>	<b>111</b>
6.1	Filtracja .....	111
6.2	Definicja i przykłady .....	116
6.3	Własności .....	125
<b>7.</b>	<b>Ciągi zmiennych losowych .....</b>	<b>131</b>
7.1	Zbieżność ciągów zmiennych losowych .....	131
7.2	Centralne twierdzenie graniczne .....	140
<b>8.</b>	<b>Proces Wienera i Poissona .....</b>	<b>147</b>
8.1	Skalowane błędzenie przypadkowe .....	147
8.2	Definicja i własności procesu Wienera .....	157
8.3	Konstrukcja procesu Wienera .....	163
8.4	Process Poissona .....	166

Część I

## Rachunek Prawdopodobieństwa



# 1

## *Przestrzeń probabilistyczna*

### 1.1 Przykłady wstępne

Kluczowa własność rynków finansowych to niepewność dotycząca przyszłych cen rozmaitych wielkości takich jak ceny akcji i związanych z nimi indeksów giełdowych, wysokość stóp procentowych, kursy wymiany walut, ceny surowców i ogólnie wszelkie ceny. Nasze zadanie to zbudowanie modelu matematycznego opisującego ten właśnie aspekt rzeczywistości.

Rozważmy jako przykład ceny akcji. Załóżmy, że dzisiejsza cena pewnej akcji jest znana, powiedzmy, że jest to  $S(0) = 10$ . Jesteśmy zainteresowani ceną w pewnym przyszłym momencie i niech to będzie chwila  $T > 0$  a cenę oznaczamy oznaczamy przez  $S(T)$ . To czego nie wiemy to jak rozwinie się sytuacja na rynkach pomiędzy chwilą dzisiejszą a chwilą przyszłą.

Przypuśćmy, że zmiany cen zachodzą etapami, w każdym kroku mogą pójść w górę lub w dół o pewną zadaną wielkość (po angielsku tzw. *tick*). Przypuśćmy, że dla naszej akcji jest to 0.50. Czyli po jednym kroku mamy możliwe ceny 9.50 i 10.50, po dwóch krokach cena może być jedną z liczb 9, 10, 11, a po 20 krokach mamy do dyspozycji w naszym modelu ceny 0, 1, 2, ..., 19, 20.

Zbór wszystkich możliwych cen oznaczamy przez  $\Omega$  i nazywamy zbiorem wyników (lub, co jest przyjętą praktyką, zdarzeń elementarnych).

#### Uwaga 1.1

Kwestia nomenklatury jest ważna, gdyż źle dobrane terminy utrudniają rozu-

mienie istoty rzeczy. Niestety, rachunek prawdopodobieństwa w tym względzie ma spore defekty. Trudno walczyć z ogólnie przyjętymi terminami do których jesteśmy przyzwyczajeni, a w dodatku są one stosowane przez innych. Niemniej, podejmiemy próbę poprawienia terminów w pewnym zakresie. "Zdarzenie elementarne" jest pierwszym przykładem tradycyjnego terminu, który jest zły, choćby dlatego że jest za długi, a ponadto jak zobaczymy będzie używane pojęcie "zdarzenia" które będzie obejmować pewną liczbę "zdarzeń elementarnych" a to jest konstrukcja językowa logicznie niepoprawna.

Elementy zbioru  $\Omega$  będą oznaczane przez  $\omega$ .

Czyli w naszym, przykładzie mamy szereg zbiorów  $\Omega$  zależnie od liczby etapów. Musimy więc za każdym razem najpierw ustalić liczbę kroków. Niech to będzie najpierw 2 czyli  $\Omega = \{9, 10, 11\}$ .

Następny etap to badanie tzw. zdarzeń, czyli podzbiorów zbioru  $\Omega$ . Na przykład niech  $A = \{10, 11\}$ . Interpretacja jest prosta gdyż to zdarzenie można opisać słowami mówiąc, że "cena nie spadła". Zdarzenie  $B = \{11\}$  odpowiada sytuacji, że "cena wzrosła". Ogólnie, zdarzenie to podzbiór  $A \subset \Omega$  przy czym rozważamy też przypadki skrajne: zdarzenie puste  $A = \emptyset$ , oraz pełne  $A = \Omega$ , mało interesujące, ale wygodnie je również dopuścić.

Najbardziej interesują nas odpowiedzi na pytania typu: jakie jest prawdopodobieństwo że cena nie spadła, jakie jest prawdopodobieństwo, że cena wzrosła i inne podobne. Takie pytanie możemy sformułować matematycznie i jednolicie w następującej formie:

Jakie jest prawdopodobieństwo że cena  $\omega$  należy do zbioru  $A$ ?

Odpowiedź na to pytanie to liczba z przedziału  $[0, 1]$  i oznaczamy ją przez  $P(A)$ . Rozważamy więc funkcję, która przypisuje zdarzeniom liczby co zapisujemy tak

$$P : A \mapsto P(A).$$

Wprowadzając tradycyjne oznaczenie na rodzinę wszystkich podzbiorów zbioru  $\Omega$

$$2^\Omega = \{A \subset \Omega\}$$

możemy napisać

$$P : 2^\Omega \rightarrow [0, 1]$$

i wtedy  $2^\Omega$  jest dziedziną funkcji  $P$ .

Wprowadzamy naturalne założenia. Najpierw, raz na zawsze robimy ustalenie dotyczące dwóch szczególnych zdarzeń, pustego i pełnego, zakładając, że

$$\begin{aligned} P(\emptyset) &= 0, \\ P(\Omega) &= 1. \end{aligned}$$

Następnie, przyjmujemy że dla zdarzeń  $A, B$  nie mających wspólnych elementów (wyników, czyli w obecnym kontekście, cen) co zapisujemy  $A \cap B = \emptyset$ , zachodzi własność tzw **addytywności**

$$P(A \cup B) = P(A) + P(B).$$

Przykładamy tu intuicję związaną z mierzeniem pól: pole dużej działki to suma pól jej części składowych. Ogólniej, zakładamy że własność addytywności zachodzi dla dowolnej skończonej liczby zdarzeń parami rozłącznych. W tym kontekście, jasne jest, że dla wyznaczenia  $P(A)$  dla dowolnego  $A$  wystarczy znać prawdopodobieństwa zbiorów jedno-elementowych. Zaczniemy od przykładu

### Przykład 1.1

$\Omega = \{9, 10, 11\}$ ,  $P(\{9\}) = P(\{10\}) = P(\{11\}) = \frac{1}{3}$  i wtedy  $P(\{10, 11\}) = \frac{2}{3}$  zgodnie z addytywnością.

Ogólnie:

### Definicja 1.1

Jeśli  $\Omega$  jest zbiorem skończonym, mającym  $n$  elementów, prawdopodobieństwo zadane następującym wzorem dla zbiorów jedno-elementowych

$$P(\{\omega\}) = \frac{1}{n}$$

nazywamy jednostajnym.

Jest dość jasne, że przyjęcie prawdopodobieństwa jednostajnego nie daje dobrego modelu cen akcji po 2 krokach. Można twierdzić, że cena 10 powinna być wyróżniona gdyż może się pojawić w wyniku dwóch scenariuszy, ruchu w górę a potem, w dół lub na odwrót, podczas gdy każda cena skrajna, tylko w wyniku jednego. Następująca modyfikacja powyższego przykładu to oddaje. Wprowadzamy upraszczające oznaczenie (chodzi tylko o redukcję liczby nawiasów we wzorach)

$$P(\{\omega\}) = p_\omega.$$

Założenie, że  $P(\Omega) = 1$  wraz z addytywnością daje ograniczenie, mówiące, że suma wszystkich prawdopodobieństw poszczególnych wyników sumuje się do 1:

$$\sum_{\omega \in \Omega} p_\omega = 1.$$

Aby funkcja  $P$  miała wartości w przedziale  $[0, 1]$  musimy też założyć

$$p_\omega \geq 0.$$

### Przykład 1.2

$\Omega = \{9, 10, 11\}$ ,  $p_9 = \frac{1}{4}$ ,  $p_{10} = \frac{1}{2}$ ,  $p_{11} = \frac{1}{4}$  i wtedy

$$P(\{10, 11\}) = \frac{3}{4}.$$

Ten przykład ilustruje ogólną metodę stosowaną w sytuacji gdy zbiór  $\Omega$  jest skończony:

$$P(A) = \sum_{\omega \in A} p_{\omega}$$

czyli korzystamy z addytywności rozbijając zbiór  $A$  na jednoelementowe części.

Rozwińmy tę ideę do bardziej skomplikowanej sytuacji. W powyższym przykładzie przyjęliśmy, że każdy scenariusz ma wagę  $0.5^2 = \frac{1}{4}$  i wybór  $p_{10}$  to  $2 \cdot 0.5^2 = \frac{1}{2}$  odpowiadający liczbie scenariuszy.

### Przykład 1.3

$\Omega = \{0, 1, 2, 3, \dots, 19, 20\}$ , waga scenariusza obejmującego 20 kroków to  $0.5^{20}$  i przyjmujemy

$$p_n = \binom{20}{n} \frac{1}{2^{20}} = \frac{20 \cdot 19 \cdot \dots \cdot (20 - n)}{1 \cdot 2 \cdot \dots \cdot n} \frac{1}{2^{20}}.$$

Jest to przykład tzw. modelu dwumianowego.

Liczby  $p_n$  są zilustrowane na Rysunku 1.1. Ceny końcowe powstały przez dodawanie i odejmowanie kwoty 0.50 stąd taki podpis.

Zdrowy rozsądek podpowiada, że ten model nie jest dobry. Rozszerzenie go o choć jeden krok doprowadzi do pojawienia się ujemnych cen. Wprowadzamy więc modyfikację dla poszczególnych kroków. Pierwszy krok prowadzący od ceny 10 do możliwych 9.50 i 10.50 oznacza zmianę o 5%. Zakładamy, że kolejne zmiany są właśnie takie, czyli zamiast dodawać i odejmować liczbę 0.50 mnożymy ceny na każdym kroku przez 1.05 lub 0.95. Wtedy

$$\Omega = \{10 \cdot 1, 05^n \cdot 0, 95^{20-n} : n = 0, 1, 2, 3, \dots, 19, 20\}$$

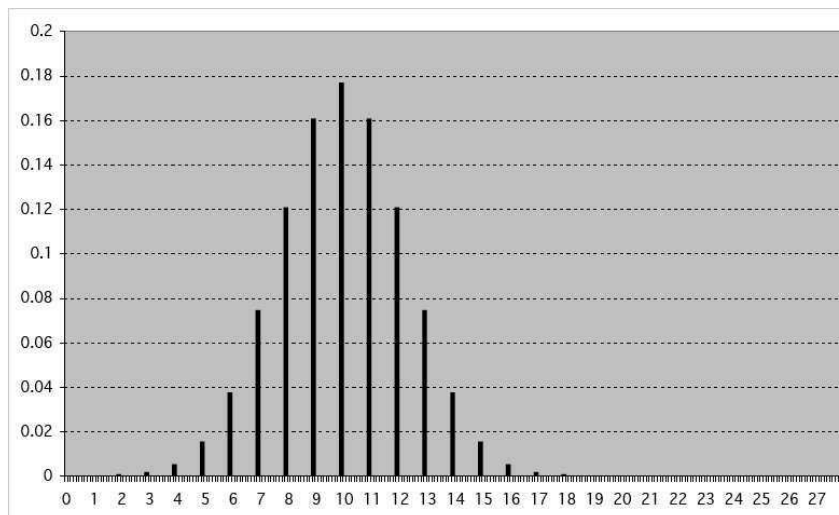
Ceny w takiej postaci będą zawsze dodatnie, bez względu na liczbę kroków.

Bierzemy prawdopodobieństwa tak jak poprzednio, dla ceny o numerze  $n$  kładąc

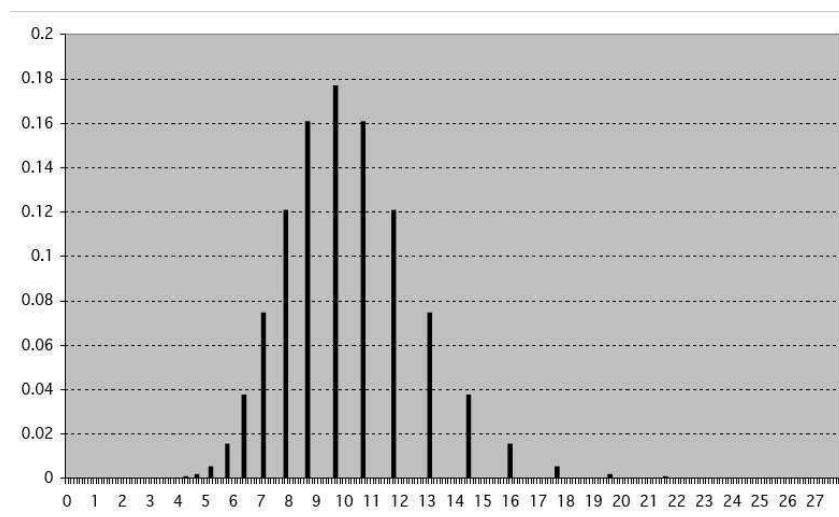
$$p_n = \binom{20}{n} \frac{1}{2^{20}}$$

i otrzymujemy sytuację zilustrowaną na Rysunku 1.2. Widać subtelną różnicę w umiejscowieniu pionowych słupków których wysokość pokazuje wielkość prawdopodobieństwa, więc jest taka sama w obu przypadkach.





**Rysunek 1.1** Zmiany addytywne cen akcji



**Rysunek 1.2** Multiplikatywne zmiany cen

### Uwaga 1.2

Trzeba podkreślić, że tej chwili nie prowadzimy żadnych analiz, czy obliczeń, tylko rozważamy na chybił trafił różne możliwości wyborów prawdopodobieństw zgodnych z wprowadzanymi ograniczeniami matematycznymi a kryterium ich sensowności jest podstawowa intuicja związana z funkcjonowaniem rynku. Oczywiście, kryterium sensowności modelu musi być bardziej wyrafinowane i będzie to osobny szeroki temat.

Pojęcie prawdopodobieństwa dwumianowego będzie wielokrotnie wykorzystywane i zasługuje na ogólne sformułowanie. Stosuje się ono do sytuacji, gdzie powtarzamy  $N$  razy eksperyment typu: sukces - porażka. Oznaczmy przez  $\Omega$  zbiór wyników, a wynik to liczba sukcesów, więc

$$\Omega = \{0, 1, 2, \dots, N-1, N\}.$$

Niech  $k \in \Omega$ . Liczba sposobów uzyskania  $k$  sukcesów dana jest wzorem wywodzącym się z kombinatoryki

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

gdzie  $n! = 1 \times 2 \times 3 \times \dots \times (n-1) \times n$ , a  $0! = 1$ .

Na przykład, niech  $N = 4$  i wtedy

$$\binom{4}{0} = 1, \quad \binom{4}{1} = 4, \quad \binom{4}{2} = 6, \quad \binom{4}{3} = 4, \quad \binom{4}{4} = 1$$

### Definicja 1.2

Niech  $p \in (0, 1)$ . Prawdopodobieństwo dwumianowe na  $\Omega = \{0, 1, 2, \dots, N-1, N\}$  dane jest wzorem

$$p_k = \binom{N}{k} p^k (1-p)^{N-k} \quad \text{dla } k \in \Omega.$$

Znany wzór Newtona

$$\sum_{k=0}^N \binom{N}{k} p^k (1-p)^{N-k} = 1$$

pozwała stwierdzić, że istotnie mamy tu prawdopodobieństwo gdyż liczby dodają się do jedności.

## 1.2 Zdarzenia i prawdopodobieństwo – przypadek skończony

Popryżej rozpatrywaliśmy skończony zbiór  $\Omega$  reprezentujący wszystkie przyszłe ceny jakiegoś waloru a dla dowolnych jego podzbiorów określaliśmy ich prawdopodobieństwa. Okazuje się, że z punktu widzenia przyszłych praktycznych zastosowań celowe jest podejście ogólniejsze, bardziej abstrakcyjne, oparte na aksjomatach, co brzmi paradoksalnie, ale jak się przekonamy, jest to prawda.

Kluczowy aksjomat to addytywność. Przypomnijmy, że dla rozłącznych zbiorów  $A, B$  pozwala on obliczyć prawdopodobieństwo ich sumy  $A \cup B$  jeśli znane są prawdopodobieństwa składowych. Stosowanie tej zasady nie przysparzało kłopotów gdyż funkcja  $P$  była określona na wszystkich podzbiorach zbioru  $\Omega$ . Okazuje się, że wskazane jest dopuszczenie innej sytuacji, gdy  $P$  jest określona tylko na niektórych zbiorach. Pojawia się pytanie co należy założyć aby powstała logiczna matematyczna konstrukcja. Motywacja poniższej definicji jest taka: aby sformułować zasadę addytywności występujące w niej pojęcia muszą mieć sens, w szczególności chodzi o prawdopodobieństwo sumy musi mieć sens jeśli mają je prawdopodobieństwa składowych. Innymi słowy, jeśli  $A, B$  należą do dziedziny funkcji  $P$  ich suma  $A \cup B$  też musi tam należeć co gwarantuje że ma sens napisanie  $P(A \cup B)$ .

Dziedzina prawdopodobieństwa będzie oznaczana literą  $\mathcal{F}$  a jej elementy będą nazywane zdarzeniami. Na  $\mathcal{F}$  nakładamy pewne warunki.

### Definicja 1.3

Rodzinę  $F$  podzbiorów zbioru wyników  $\Omega$  która spełnia warunki

1.  $\emptyset \in F$ ,
2. jeśli  $A, B \in F$  to  $A \cup B \in F$ ,
3. jeśli  $A \in F$  to  $\Omega \setminus A \in F$ ,

nazywamy ciałem zbiorów (zdarzeń).

Odnotujmy pewne proste własności ciała zbiorów.

### Twierdzenie 1.1

Niech  $\mathcal{F}$  będzie ciałem zbiorów. Wtedy

**Proposition 1.2**

1.  $\Omega \in \mathcal{F}$ .
2. Dla dowolnego  $n$ , jeśli  $A_i \in \mathcal{F}$ ,  $i = 1, 2, \dots, n$ , to

$$\bigcup_{i=1}^n A_i = A_1 \cup \dots \cup A_n \in \mathcal{F}.$$

**Dowód**

1. Możemy oczywiście zapisać  $\Omega = \Omega \setminus \emptyset$  oraz skorzystać z definicji ciała. Po pierwsze  $\emptyset \in \mathcal{F}$  z własności 1, a następnie z własności 3 mamy  $A \in \mathcal{F}$  implikuje  $\Omega \setminus A \in \mathcal{F}$  co stosujemy dla  $A = \emptyset$  dostając jako konkluzję fakt  $\Omega \in \mathcal{F}$ .

2. Dla  $n = 2$  dowodzony fakt to postulat 2 w definicji ciała.

Niech  $n = 3$  i weźmy dowolne  $A_1, A_2, A_3 \in \mathcal{F}$ . Kładziemy:

$$\begin{aligned} A &= A_1, \\ B &= A_2 \cup A_3. \end{aligned}$$

Na podstawie własności 2 z definicji ciała mamy  $B \in \mathcal{F}$ . Teraz

$$A_1 \cup A_2 \cup A_3 = A \cup B \in \mathcal{F}$$

znów na podstawie własności 2 korzystając z tego że zarówno  $A$  jak i  $B$  są elementami  $\mathcal{F}$ . Działając według tego schematu uzyskamy tezę kolejno dla dowolnego  $n = 4, 5, \dots$  powtarzając rozumowanie stosowną liczbę razy. Precyzyjny dowód wymaga zastosowania indukcji matematycznej ale biegli w jej arkanach już na pewno widzą, że powyższy argument pozwoli wykonać kluczowy dla tej metody, tzw. krok indukcyjny.  $\square$

**Twierdzenie 1.3**

Jeśli  $\mathcal{F}$  jest ciałem,  $A, B \in \mathcal{F}$  to  $A \cap B \in \mathcal{F}$ .

**Dowód**

Ponieważ  $A, B \in \mathcal{F}$  więc  $\Omega \setminus A, \Omega \setminus B \in \mathcal{F}$  na podstawie postulatu 3 w definicji ciała. Teraz mamy  $\Omega \setminus A, \Omega \setminus B \in \mathcal{F}$  więc na podstawie postulatu 2 mamy wniosek dotyczący ich sumy  $(\Omega \setminus A) \cup (\Omega \setminus B) \in \mathcal{F}$ . Można łatwo pokazać następujący wzór, tzw. prawo de Morgana

$$(\Omega \setminus A) \cup (\Omega \setminus B) = \Omega \setminus (A \cap B).$$

Daje on następujący wniosek:  $\Omega \setminus (A \cap B) \in \mathcal{F}$  gdyż zbiór z lewej strony tam należy. Uzupełnienie zbioru z  $\mathcal{F}$  jest w  $\mathcal{F}$  więc  $\Omega \setminus [\Omega \setminus (A \cap B)] \in \mathcal{F}$  ale

$$\Omega \setminus [\Omega \setminus (A \cap B)] = A \cap B$$

co kończy dowód.  $\square$

### Twierdzenie 1.4

Dla każdego  $n$ , jeśli  $A_1, A_2, \dots, A_n \in \mathcal{F}$  to

$$A_1 \cap A_2 \cap \dots \cap A_n = \bigcap_{i=1}^n A_i \in \mathcal{F}$$

### Dowód

Możemy rozumować podobnie jak dla sumy zauważywszy, że  $A_1 \cap A_2 \cap A_3 = (A_1 \cap A_2) \cap A_3$  co daje wynik dla trzech zbiorów i pozwala na rozciągnięcie kolejno wyniku na dowolną liczbę, jak i daje klucz do indukcji matematycznej.  $\square$

### Definicja 1.4

Prawdopodobieństwo na skończonym zbiorze  $\Omega$  to funkcja przypisująca zdarzeniom liczby:  $P : \mathcal{F} \rightarrow [0, 1]$ , która spełnia warunki

1.  $P(\emptyset) = 0$ ,  $P(\Omega) = 1$ ,
2. Addytywność: jeśli  $A_k \cap A_j = \emptyset$  dla  $k \neq j$ . to

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

Poniższe twierdzenie mówi, że powyższy warunek addytywności wystarczy założyć dla dwóch zbiorów.

### Twierdzenie 1.5

Jeśli  $P(A \cup B) = P(A) + P(B)$  dla zdarzeń spełniających  $A \cap B = \emptyset$ , to

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k)$$

o ile  $A_k$  są parami rozłączne, czyli  $A_k \cap A_j = \emptyset$  jeśli  $k \neq j$ .

### Dowód

Intuicja jest prosta: stosując wielokrotnie własność dla dwóch zbiorów otrzymamy ją dla dowolnej ich ilości po skończonej liczbie przejść. Możemy to zapisać symbolicznie

$$\begin{aligned}
 P\left(\bigcup_{k=1}^n A_k\right) &= P\left(A_1 \cup \bigcup_{k=2}^n A_k\right) \\
 &= P(A_1) + P\left(\bigcup_{k=2}^n A_k\right) \quad \text{stosując addytywność dwóch zdarzeń} \\
 &= P(A_1) + P\left(A_2 \cup \bigcup_{k=3}^n A_k\right) \\
 &= P(A_1) + P(A_2) + P\left(\bigcup_{k=3}^n A_k\right) \quad \text{stosując addytywność dwóch zdarzeń} \\
 &= \dots \\
 &= \sum_{k=1}^n P(A_k).
 \end{aligned}$$

Dzięki temu, że zdarzenia są parami rozłączne, mamy rozłączność na każdym etapie, np.  $A_1 \cap \bigcup_{k=2}^n A_k = \emptyset$ ,  $A_2 \cap \bigcup_{k=3}^n A_k = \emptyset$ , itd. (Ścisły dowód wymaga zastosowania indukcji matematycznej, a wtedy addytywność dwóch zdarzeń pozwoli przeprowadzić dowód kroku indukcyjnego.)  $\square$

## 1.3 Przestrzeń probabilistyczna

Słabość powyższych modeli polega między innymi na arbitralnym ograniczeniu zbioru dopuszczalnych przyszłych cen. To ograniczenie to wynikające z wyboru zbioru  $\Omega$  założenie że mamy tylko pewien skończony zbiór cen. Jasne jest ustalenie dziś liczby dopuszczalnych cen nie jest realistyczne. Powinniśmy być elastyczni i dopuścić zbiory o większej liczbie elementów. Aby być uniwersalnym, jedynym rozwiązaniem jest dopuszczenie nieskończonego zbioru możliwych cen, gdyż taki model zawrze w sobie dowolne skończone przypadki.

Nieskończoność ma wiele postaci i zaczynamy od najprostszej, gdzie  $\Omega = \{1, 2, 3, \dots\}$  czyli jest zbiorem liczb naturalnych. W kontekście finansowym zakładamy, że cena jest wyrażona jako liczba pewnych jednostek, na przykład groszy. Dopuszczenie sytuacji że nie ma ograniczenia na jej wysokość jest uzasadnione – nawet jeśli cena jest ogromna to może wzrosnąć o jeden grosz, więc jakiegokolwiek ograniczenie byłoby sztuczne.

Rozważanie zbiorów nieskończonych wymaga bardziej wyrafinowanych narzędzi. Tak jak w sytuacji zbioru skończonego, budujemy prawdopodobieństwo zakładając, że mamy liczby  $p_n$  dla każdego  $n \in \Omega$  i są to liczby z przedziału  $[0,1]$ . Aby zdefiniować prawdopodobieństwo dla dowolnego zdarzenia zgodnie z ideą wzoru

$$P(A) = \sum_{\omega \in A} p_\omega$$

musimy dopuścić zbiory nieskończone, więc modyfikujemy warunek addytywności: dla parami rozłącznych zdarzeń  $A_k$  postulujemy

$$P(A_1 \cup A_2 \cup \dots) = P\left(\bigcup_{k=1}^{\infty} A_k\right) = P(A_1) + P(A_2) + \dots = \sum_{k=1}^{\infty} P(A_k).$$

Różnica polega na zastąpieniu sumy skończonej, sumą szeregu.

### Przykład 1.4

Przypomnijmy, że rozpoczęliśmy nasze rozważania od analizy zachowania się cen w okresie pomiędzy chwilą obecną 0 a przyszłą  $T$  zakładając że w tym okresie zachodzi pewna liczba zmian cen zgodnie z pewnym schematem, addytywnym lub multiplikatywnym, teraz trzymamy się tego ostatniego. Tych zmian była określona liczba (2 czy 20). Teraz dopuścimy sytuację, że tych zmian może być bardzo wiele. Liczba zmian jest dziś nieznana, będzie skończona, ale nie możemy ograniczyć jej z góry. Dopuszczamy więc dowolną liczbę naturalną  $N$  jako liczbę zmian i zakładamy, że prawdopodobieństwo  $N$  zmian wynosi

$$q_N = \frac{\lambda^N e^{-\lambda}}{N!}.$$

Proponowany wzór, jak zobaczymy później ma szerokie zastosowania, a teraz zaakceptujemy go jako dający ciąg malejący liczb których suma wynosi jeden:

$$\sum_{N=0}^{\infty} q_N = \sum_{N=0}^{\infty} \frac{\lambda^N e^{-\lambda}}{N!} = e^{-\lambda} \sum_{N=0}^{\infty} \frac{\lambda^N}{N!} = e^{-\lambda} e^{\lambda} = 1.$$

Aby nie komplikować, założymy, że zmiany w górę i w dół są jednakowo prawdopodobne i zachodzą w sposób multiplikatywny o 5%. Więc przy  $N$  zmianach cena w chwili  $T$  może wynieść

$$S(T) = 10 \times 1.05^n \times 0.95^{N-n}$$

z prawdopodobieństwem

$$p_n = q_N \binom{N}{n} \frac{1}{2^N}.$$

Model jest naturalny choć dość skomplikowany - losowa jest ilość zmian  $N$  jak i liczba ruchów w górę  $n$ .

### Przykład 1.5

Inny przykład gdzie w naturalny sposób pojawia się zbiór nieskończony to badanie chwili gdy coś nieprzewidzianego może się zdarzyć. Mierzac czas liczbą kroków i biorąc  $\Omega = \{1, 2, 3, \dots\}$  przypuśćmy, że ktoś rzuca niesymetryczną monetą z prawdopodobieństwem  $p$  wyrzucenia orła. Jeśli  $n \in \Omega$  oznacza liczbę reszek zanim orzeł pojawi się po raz pierwszy, to naturalne jest przyjąć

$$p_n = p(1-p)^{n-1}$$

i łatwo pokazać że suma  $p_n$  daje 1.

Z punktu widzenia budowy gładkiej teorii, założenie o addytywności prawdopodobieństwa z przeliczalną liczbą zdarzeń okazuje się optymalne. Istnieją przykłady, które pokazują, że nie można wymagać by  $P$  było określone dla wszystkich podzbiorów zbioru  $\Omega$ . Te fakty znajdują odbicie w następujących definicjach kładących ogólnie przyjęte fundamenty pod teorię prawdopodobieństwa.

### Definicja 1.5

Sigma ciało na zbiorze  $\Omega$  to rodzina  $F$  podzbiorów zbioru  $\Omega$  spełniająca

1.  $\emptyset \in F$
2. jeśli  $A \in F$  to  $\Omega \setminus A \in F$
3. jeśli  $A_n \in F$ ,  $n = 1, 2, 3, \dots$  to  $\bigcup_{n=1}^{\infty} A_n \in F$ .

### Definicja 1.6

Odzworowanie  $P : F \rightarrow [0, 1]$  nazywamy miarą prawdopodobieństwa jeśli

1.  $P(\emptyset) = 0$ ,  $P(\Omega) = 1$
2. jeśli  $A_i \cap A_j = \emptyset$  dla  $i \neq j$ ,  $A_i \in F$ , to (przeliczalna addytywność)

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

### Definicja 1.7

Trójka  $(\Omega, F, P)$  jest przestrzenią probabilistyczną jeśli  $F$  jest sigma ciałem, a  $P$  miarą prawdopodobieństwa na  $F$ .



Tak jak w przypadku skończonym, nakładamy na  $\mathcal{F}$  warunki związane z postulatem addytywności, teraz rozciągniętym na nieskończone ciągi zdarzeń, dzięki którym  $\mathcal{F}$  pełni skutecznie rolę dziedziny odwzorowania  $P$

Odnotujmy najpierw oczywisty fakt:

### Twierdzenie 1.6

Sigma ciało jest ciałem.

### Dowód

Dla dowodu punktu 3 (postulaty 1 i 2 są identyczne dla obu pojęć) bierzmy  $A, B \in \mathcal{F}$  i konstruujemy ciąg  $A_1 = A, A_2 = B = A_3 = A_4 = \dots$  i mając  $A_n \in \mathcal{F}$  dostajemy  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$  ale  $\bigcup_{n=1}^{\infty} A_n = A \cup B$ .  $\square$

Prawdopodobieństwo ma wiele użytecznych własności:

### Twierdzenie 1.7

1. Dla zbiorów parami rozłącznych

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

2.  $P(\Omega \setminus A) = 1 - P(A)$ .
3. Jeśli  $A \subset B$  to  $P(A) \leq P(B)$  (monotoniczność).

### Dowód

**1.** Skończona addytywność to własność prostsza niż przeliczalna addytywność, więc wydaje się oczywiste że powinna zchodzić, ale wymaga jednak uzasadnienia, które powinno być proste. Biorąc  $A_{n+1} = A_{n+1} = \dots = \emptyset$ , ich prawdopodobieństwo to zero więc suma szeregu redukuje się do sumy skończonej i tak samo suma nieskończona zbiorów redukuje się do sumy skończonej. Dzięki temu obie strony postulatu przeliczalnej addytywności redukują się do odpowiednich wyrażeń dla skończonego przypadku. Dowód wystarczyło przeprowadzić dla dwóch zbiorów gdyż jak wiemy to już wystarcza dla dowolnej ich skończonej liczby.  $\square$

**2.** Wzór  $P(\Omega \setminus A) = 1 - P(A)$  wynika ze skończonej addytywności zastosowanej do  $A_1 = 1, A_2 = \Omega \setminus A$ .

3. Monotoniczność wynika ze skończonej addytywności. Zdarzenia

$$\begin{aligned} A_1 &= A, & A_2 &= B \setminus A, \\ A_1 \cup A_2 &= B \end{aligned}$$

są rozłączne więc

$$P(A_1) + P(A_2) = P(A_1 \cup A_2)$$

czyli

$$P(A) + P(B \setminus A) = P(B)$$

ale  $P(B \setminus A) \geq 0$  więc dostajemy żadaną nierówność.

Czas na trochę bardziej skomplikowane własności.

### Twierdzenie 1.8

Prawdopodobieństwo ma następujące własności:

1. Subaddytywność:

$$P\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k).$$

2. Jeśli  $B_n \subset B_{n+1}$  to

$$P\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n \rightarrow \infty} P(B_n).$$

3. Przeliczalna subaddytywność:

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k).$$

4. Jeśli  $A_{n+1} \subset A_n$  to

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Uwaga: W własnościach 1,3 nie wymagamy aby zbiory były parami rozłączne.

### Dowód

1. Dla  $n = 2$ ,

$$P(A_1 \cup A_2) = P(A_1 \cup (A_2 \setminus A_1)) = P(A_1) + P(A_2 \setminus A_1) \leq P(A_1) + P(A_2)$$

korzystając z monotoniczności,  $(A_2 \setminus A_1 \subset A_2)$ . Powtarzając to wielokrotnie, a ściśle rzecz biorąc prowadząc dowód indukcyjny, otrzymujemy tezę. Krok indukcyjny: przejście z  $n$  do  $n+1$  jest identyczny jak powyższy krok pierwszy:

$$\begin{aligned} P\left(\bigcup_{k=1}^{n+1} A_k\right) &= P\left(A_{n+1} \cup \bigcup_{k=1}^n A_k\right) \\ &\leq P(A_{n+1}) + P\left(\bigcup_{k=1}^n A_k\right) \quad \text{jak dla } n=2 \\ &\leq P(A_{n+1}) + \sum_{k=1}^n P(A_k) \quad \text{z założenia indukcyjnego.} \end{aligned}$$

**2.** Wykorzystujemy rozmaite własności funkcji  $P$  a istota dowodu polega na rozłożeniu sumy wstępującej rodziny zbiorów na sumę zbiorów parami rozłącznych, które można sobie wyobrazić jako pierścienie

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} B_n\right) &= P\left(B_1 \cup \bigcup_{n=1}^{\infty} (B_{n+1} \setminus B_n)\right) \quad \text{gdyż } B_n \subset B_{n+1} \\ &= P(B_1) + \sum_{n=1}^{\infty} P(B_{n+1} \setminus B_n) \quad \text{zbiory są rozłączne} \\ &= P(B_1) + \lim_{m \rightarrow \infty} \sum_{n=1}^{m-1} P(B_{n+1} \setminus B_n) \quad \text{z definicji sumy szeregu} \\ &= \lim_{m \rightarrow \infty} P(B_m) \quad \text{gdyż } B_m = B_1 \cup \bigcup_{n=1}^{m-1} (B_{n+1} \setminus B_n). \end{aligned}$$

**3.** Definiujemy zbiry  $B_n = \bigcup_{k=1}^n A_k$  które są wstępujące,  $B_n \subset B_{n+1}$ , więc możemy zastosować punkt 2 co daje

$$\begin{aligned} P\left(\bigcup_{k=1}^{\infty} A_k\right) &= P\left(\bigcup_{k=1}^{\infty} B_k\right) \\ &= \lim P(B_n) \quad \text{korzystamy z punktu 2} \\ &= \lim P\left(\bigcup_{k=1}^n A_k\right) \\ &\leq \lim \sum_{k=1}^n P(A_k) \quad \text{na podstawie punktu 1} \\ &= \sum_{k=1}^{\infty} P(A_k). \end{aligned}$$

4. Oznaczamy

$$A = \bigcap_{n=1}^{\infty} A_n$$

a ponieważ zbiory  $A_n$  są zstępujące,  $A_{n+1} \subset A_n$ , to zbiory  $\Omega \setminus A_n$  są wstępujące i możemy zastosować punkt 2:

$$\begin{aligned} P(A) &= 1 - P\left(\bigcup_{n=1}^{\infty} (\Omega \setminus A_n)\right) \quad \text{korzystając z } P(A) = 1 - P(\Omega \setminus A) \\ &= 1 - \lim P(\Omega \setminus A_n) \quad \text{z punktu 2} \\ &= \lim P(A_n). \end{aligned}$$

□

## 1.4 Miara Lebesgue'a

Wracając do kwestii związanych z modelowaniem, rozważanie zbiorów skończonych lub ciągów jest kłopotliwe. Teoria jest dużo gładsza jeśli rozważyć przedziały jako dopuszczalne zakresy przyszłych cen, a w następnej fazie nawet wszystkie nieujemne liczby rzeczywiste. Wprawdzie można twierdzić, że wystarczą zbiory skończone bo wystarczy podawać ceny z dokładnością do jednego grosza, a mamy też ograniczenie górne, wszystkie pieniądze tego świata, ten punkt widzenia nie jest przekonujący. Mianowicie okazuje się, że pewne wielkości trzeba podawać bardzo dokładnie, na przykład kursy wymiany walut i odgórne ograniczenie na dokładność tego kwotowania byłoby niewskazane. Tak więc nie tylko względy elegancji teorii, ale również praktyczne powodują, że wskazane jest dopuszczanie przypadków

$$\Omega = [a, b], \quad \Omega = [0, \infty), \quad \Omega = (0, \infty).$$

Weźmy  $\Omega = [0, 1]$  co jest ważnym szczególnym przypadkiem. Z punktu widzenia modelowania cen, to jest całkiem dobry model jeśli przyjąć stosowną jednostkę. Jeśli na przykład 1 oznacza jeden milion, przedział jednostkowy pomieści wszystkie sensowne ceny wyrażone w sensownej walucie.

Chcemy wyposażać ten zbiór w sigma ciało oraz prawdopodobieństwo. Najprostsze rozwiązanie to rozszerzyć ideę prawdopodobieństwa jednostajnego, gdzie wszystkie wyniki są jednakowo prawdopodobne. Nie możemy tego zrobić bezpośrednio gdyż przyjmując  $P(\{\omega\}) = p$  doprowadzilibyśmy do nieskończonej wartości dla  $P(\Omega)$  i to wykorzystując jedynie jakiś ciąg jego elementów. Musimy więc mieć  $P(\{\omega\}) = 0$ , a jednostajność realizujemy wymagając aby

prawdopodobieństwo przedziałów o takiej samej długości było jednakowe, co sprowadza się do założenia

$$P([a, b]) = b - a$$

dla  $a < b$ ,  $[a, b] \subset [0, 1]$ . Mamy oczywiście  $P(\Omega) = 1$  oraz  $P(\{a\}) = P([a, a]) = 0$ . Poważny problem matematyczny to zdefiniowanie takiego prawdopodobieństwa  $P$  na jakimś sigma ciele.

Zaczynamy od zdefiniowania stosownego sigma ciała. Bierzemy najmniejsze możliwe sigma ciało zawierające przedziały i oznaczamy je przez  $\mathcal{B}_{[0,1]}$  a jego elementy nazywamy zbiorami borelowskimi (lub Borela). Formalnie,

$$\mathcal{B}_{[0,1]} = \bigcap \{ \mathcal{F} : \mathcal{F} \text{ jest sigma-ciałem takim, że dla } a, b \in [0, 1], [a, b] \in \mathcal{F} \}.$$

Uzasadnienie tego sformułowania jest następujące: przecięcie sigma ciał jest sigma ciałem (co wymaga dowodu, na szczęście prostego) a ponieważ przecinamy wszystkie sigma ciała zawierające przedziały, jako wynik dostaniemy najmniejsze możliwe, gdyż będzie się ono zawierać w każdym innym.

Przykłady zbiorów borelowskich to:  $\{\frac{1}{n}\} = \bigcup [\frac{1}{n}, \frac{1}{n}]$ ,  $\mathbb{Q} \cap [0, 1] = \bigcup [q_i, q_i]$ ,  $q_i \in [0, 1]$  wszystkie liczby wymierne z tego przedziału, co można uogólnić zauważając, że każdy zbiór przeliczalny jest borelowski jako suma zbiorów jednoelementowych.

Drugi krok to uzyskanie  $P$  określonego na  $\mathcal{B}_{[0,1]}$ . Mamy już  $P$  zdefiniowane na przedziałach jako długość. Dowód, że można przedłużyć  $P$  na całe  $\mathcal{B}_{[0,1]}$  jest długi i skomplikowany, więc przyjmujemy ten fakt bez dowodu:

### Twierdzenie 1.9

Istnieje  $P : \mathcal{B}_{[0,1]} \rightarrow [0, 1]$  spełniająca postulaty miary probabilistycznej.

Ta konstrukcja może być rozszerzona na cały zbiór liczb rzeczywistych  $\mathbb{R}$ . Mianowicie definiując zbiory borelowskie  $\mathcal{B}_{\mathbb{R}}$  analogicznie jak dla przedziału  $[0, 1]$  mamy następujący fakt:

### Twierdzenie 1.10

Istnieje funkcja  $m : \mathcal{B}_{\mathbb{R}} \rightarrow [0, +\infty]$  taka, że  $m([a, b]) = b - a$  oraz  $m$  jest przeliczalnie addytywna:

$$m\left(\bigcup B_n\right) = \sum m(B_n)$$

dla parami rozłącznych zbiorów borelowskich  $B_n$ .

### Uwaga 1.3

Jako wartość funkcji  $m$  dopuszczamy z konieczności symbol  $+\infty$ , gdyż przeliczalna addytywność wymusza nieskończoną miarę zbioru  $R$  zawierającego nieskończenie wiele przedziałów rozłącznych o długości jeden.

Funkcję  $m$  nazywamy **miarą Lebesgue’a**.

Teraz miara probabilistyczna na  $[0, 1]$  to ograniczenie miary  $m$  określonej na  $\mathbb{R}$  do przedziału  $[0, 1]$ . Ograniczając ją do innego przedziału otrzymujemy miarę, czyli funkcję przeliczalnie addytywną, ale niekoniecznie probabilistyczną, gdyż dla  $\Omega = [a, b]$ ,  $m([a, b])$  nie musi być jeden. Można temu łatwo zaradzić przyjmując

$$P(A) = \frac{1}{m([a, b])} m(A)$$

i już  $P([a, b]) = 1$ . Pomnożenie przez stałą aby zapewnić warunek  $P(\Omega) = 1$  nazywa się unormowaniem.

### Uwaga 1.4

Możliwe jest uogólnienie powyższych rozważań do wielu wymiarów. Z punktu widzenia praktyki istotny jest przypadek dwuwymiarowy. Myślimy o modelowaniu dwóch cen więc bierzemy  $\Omega = [0, 1] \times [0, 1]$ , podobnie jak przypadku jednowymiarowym rozważamy  $F = B_{[0, 1] \times [0, 1]}$  – najmniejsze sigma ciało zawierające prostokąty, a stosowna wersja konstrukcji miary Lebesgue’a daje jej dwuwymiarową realizację, miarę  $P$  przeliczalnie addytywną na zbiorach borelowskich, a na prostokątach równą ich polu.

Uzyskujemy możliwość modelowania cen, których zakres to dowolny przedział. Płacimy pewną cenę, którą jest abstrakcyjne pojęcie miary Lebesgue’a. Co gorsza, jak pamiętamy z dyskusji przypadków skończonych, prawdopodobieństwo jednostajne nie jest wystarczające. W przypadku skończonym wprowadziliśmy pewne wagi, różne dla różnych wyników. Podobnie postąpimy teraz, a przy okazji okaże się, że w praktyce poradzimy sobie bez miary Lebesgue’a, choć jej istnienie daje niezbędne podstawy matematyczne. Realizacja tego planu wiedzy przez wprowadzenie pojęcia całki, której jest poświęcony następny rozdział, a ten kończymy opisem ważnego narzędzia, które może zastąpić pojęcie miary.

## 1.5 Dystrybuanta

Miara jako funkcja określona na zbiorach jest trudnym pojęciem. Znacznie łatwiej zrozumieć funkcje określone na zbiorach liczbowych. Okazuje się, że można opisać miarę przy pomocy takiej funkcji.

### Definicja 1.8

Dystrybuanta:  $F : \mathbb{R} \rightarrow [0, 1]$  jest dana wzorem  $F(x) = P((-\infty, x])$ .

Wygodnie jest poprzedzić przykłady własnościami, gdyż ułatwią nam one poszczególne konstrukcje.

### Twierdzenie 1.11

1.  $F$  jest rosnąca:  $x \leq y$  implikuje  $F(x) \leq F(y)$ ,
2.  $F$  jest prawostronnie ciągła:  $\lim_{x \rightarrow a+} F(x) = F(a)$ ,
3.  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,
4.  $\lim_{x \rightarrow \infty} F(x) = 1$ .

### Dowód

1. Dla  $A = (-\infty, x]$ ,  $B = (-\infty, y]$  korzystamy z własności  $P$

$$A \subset B \text{ implikuje } P(A) \leq P(B)$$

co daje wynik gdyż jeżeli  $x \leq y$  to  $A \subset B$ .

2. Korzystamy z własności: jeśli  $A_{n+1} \subset A_n$  to  $P(\bigcap_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$  dla pewnego zstępującego ciągu przedziałów:

$$\begin{aligned} \lim_{x \rightarrow a+} F(x) &= \lim_{n \rightarrow \infty, x_n > a} F(x_n) \\ &= \lim_{n \rightarrow \infty, x_n > a} P((-\infty, x_n]) \quad \text{niech } A_n = (-\infty, x_n] \\ &= P\left(\bigcap_{n=1}^{\infty} A_n\right) \quad \text{ze wspomnianej własności} \\ &= P((-\infty, a]) = F(a) \end{aligned}$$

□

Dystrybuanta jest zdefiniowana przez miarę, ale istotne jest to, że można to odwrócić. Mając funkcję  $F$  spełniającą warunki 1-4 powyższego twierdzenia możemy skonstruować miarę  $P$  której  $F$  jest dystrybuantą. Robimy to w dwóch krokach:

- Definiujemy  $P$  na przedziałach:

$$P((a, b]) = F(b) - F(a)$$

Dla tak zdefiniowanej miary, jej dystrybuanta to dane  $F$ , mianowicie kładąc  $F_1(x) = P((-\infty, x])$  to  $P((a, b]) = P((-\infty, b] \setminus (-\infty, a]) = F_1(b) - F_1(a)$  dla każdych  $a, b$  więc  $F_1 = F$  po przejściu z  $a$  do  $-\infty$  gdzie  $F$  się zeruje z założenie, a  $F_1$  z definicji.

- Rozszerzamy  $P$  do zbiorów borelowskich co wymaga dowodu podobnego do dowodu istnienia miary Lebesgue'a, który oczywiście pomijamy.

Przykłady zaczniemy od bardzo prostej miary, chyba najprostszej możliwej.

### Definicja 1.9

Miara Diraca skoncentrowana w punkcie  $c$  jest dana wzorem

$$\delta_c(A) = \begin{cases} 1 & \text{jeśli } c \in A \\ 0 & \text{jeśli } c \notin A \end{cases}$$

Zauważmy, że:  $\delta_c(\{c\}) = 1$ . Dla konstrukcji dystrybuanty rozważamy dwa przypadki:

1.  $x < c$ , a wtedy  $c \notin (-\infty, x]$  więc

$$F(x) = \delta_c((-\infty, x]) = 0$$

2.  $x \geq c$ , a wtedy  $c \in (-\infty, x]$  więc

$$F(x) = \delta_c((-\infty, x]) = 1.$$

Możemy podsumować konstrukcję pisząc

$$F(x) = \begin{cases} 0 & \text{dla } x < c \\ 1 & \text{dla } x \geq c \end{cases}$$

Jeśli mając  $F$  chcielibyśmy odtworzyć miarę, zauważmy że dla  $a, b < c$ , oraz dla  $c \leq a, b$  różnica  $F(b) - F(a)$  będzie równa zero bo  $F$  jest stała. natomiast jeśli  $a < c \leq b$  to  $F(b) - F(a) = 1$ .



Zauważmy, że  $F$  ma skok w punkcie  $a$  od zera do jeden. Ma ona wszystkie cztery własności co oczywiście wynika z twierdzenia ale teraz możemy je zobaczyć, najlepiej samodzielnie szkicując jej wykres.

Odnosimy ogólną własność dystrybucyjną:

### Twierdzenie 1.12

Jeśli  $F$  ma skok w  $c$  to  $P(\{c\}) > 0$

### Dowód

Dowód wykorzystuje prawostronną ciągłość  $F$

$$P(\{a\}) = \lim P((a - \frac{1}{n}, a]) = \lim_{x \rightarrow a, x < a} F(x) - F(a)$$

i jeśli funkcja ma skok, to ostatnie wyrażenie jest ściśle dodatnie.  $\square$

Weźmy teraz bardziej skomplikowaną miarę zbudowaną z miar Diraca

$$P = p_1 \delta_{c_1} + p_2 \delta_{c_2}$$

gdzie  $p_1 + p_2 = 1$ ,  $p_1, p_2 > 0$ ,  $c_1 < c_2$ . Badając trzy przypadki

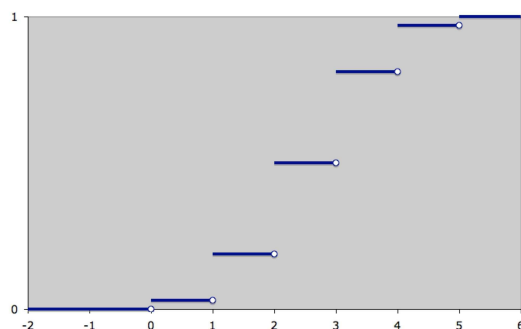
1.  $x < c_1$ ,  $P((-\infty, x]) = 0$  bo każda ze składowych jest zero
2.  $x \in [c_1, c_2)$ ,  $P((-\infty, x]) = p_1$  bo  $p_1 \delta_{c_1}((-\infty, x]) = p_1$  a drugi składnik znika bo  $c_2 \notin (-\infty, x]$
3.  $x \geq c_2$ ,  $P((-\infty, x]) = 1$  bo oba punkty  $c_1$  i  $c_2$  należą do  $(-\infty, x]$ , i obie miary Diraca dają jeden

dochodzimy do konkluzji

$$F(x) = \begin{cases} 0 & \text{dla } x < c_1 \\ p_1 & \text{dla } x \in [c_1, c_2) \\ 1 & \text{dla } x \geq c_2 \end{cases}$$

Często wykorzystywany szczególny przypadek to  $c_1 = -1$ ,  $c_2 = 1$ ,  $p_1 = p_2 = \frac{1}{2}$  czyli miara jednostajna w istocie zdefiniowana na zbiorze dwulementowym  $\{-1, 1\}$

Nawiązując do przykładów cen akcji, przy zmianach zachodzących w 5 krokach, z prawdopodobieństwem  $\frac{1}{2}$  ruchu ceny w górę w jednym kroku, wprowadzamy miarę mówiącą jakie jest prawdopodobieństwo  $n$  kroków w górę,  $n = 0, 1, 2, 3, 4, 5$ :



$$P = p_1\delta_0 + p_2\delta_1 + p_3\delta_2 + p_3\delta_3 + p_4\delta_4 + p_5\delta_5$$

gdzie  $p_n = \binom{5}{n} \frac{1}{2^n}$ . Rozważając stosowną liczbę przypadków możemy dojść do wniosku, że dystrybucja będzie miała wykres taki jak na Rysunku 1.5.

Kolejny przykład to  $P(A) = m(A \cap [0, 1])$  dla borelowskich zbiorów  $A$ , czyli w istocie zakładamy, że  $P$  jest miarą Lebesgue'a na przedziale  $[0, 1]$ . Konstrukcja dystrybucji wymaga rozważenia trzech przypadków

1.  $x < 0$  a wtedy  $P((-\infty, x]) = 0$  bo  $(-\infty, x] \cap [0, 1] = \emptyset$
2.  $x \in [0, 1]$  i mamy  $(-\infty, x] \cap [0, 1] = [0, x]$ ,  $P((-\infty, x]) = m([0, x]) = x$
3.  $x \geq 1$  a tutaj  $(-\infty, x] \cap [0, 1] = [0, 1]$ , i  $P((-\infty, x]) = m([0, 1]) = 1$

co daje

$$F(x) = \begin{cases} 0 & \text{dla } x < 0 \\ x & \text{dla } x \in [0, 1] \\ 1 & \text{dla } x \geq 1 \end{cases}$$

Zauważmy, że funkcja jest ciągła, co wymaga tylko zbadania jej w punktach 0, 1.

Ostatni przykład to tak zwana miara mieszana. Przypuśćmy, że chcemy modelować ceny akcji dopuszczając możliwość bankructwa firmy z pewnym prawdopodobieństwem a wtedy cena akcji spadnie do zera. Załóżmy, że chcemy uzyskać:  $P(S(T) = 0) = \frac{1}{10}$  a dla dodatnich cen przyjmujemy przedział  $(0, 1]$  z miarą Lebesgue'a ale z odpowiednim współczynnikiem aby dostać w wyniku miarę probabilistyczną

Dla  $A \subset [0, 1]$  kładziemy

$$P(A) = \begin{cases} \frac{1}{10} + \frac{9}{10}m(A) & \text{jeśli } 0 \in A \\ \frac{9}{10}m(A) & \text{jeśli } 0 \notin A \end{cases}$$

czyli

$$P = \frac{1}{10}\delta_0 + \frac{9}{10}m_{[0,1]}$$

gdzie  $m_{[0,1]}$  oznacza miarę Lebesgue'a ograniczoną do przedziału  $[0, 1]$ . Zauważmy, że

$$P([0, 1]) = \frac{1}{10} + \frac{9}{10}m_{[0,1]}([0, 1]) = 1$$

Budując dystrybuantę rozważamy przypadki

1.  $x < 0$ , wtedy tak jak poprzednio,  $F(x) = 0$
2.  $x = 0$  i mamy  $\delta_0((-\infty, 0]) = 1$  czyli  $F(0) = P((-\infty, 0]) = \frac{1}{10}$  gdyż drugi składnik jest zero:  $m_{[0,1]}((-\infty, 0]) = m(\{0\}) = 0$
3.  $x \in (0, 1)$  i wtedy pierwszy składnik daje  $\frac{1}{10}$  a drugi

$$\frac{9}{10}m_{[0,1]}([-\infty, x]) = \frac{9}{10}m([0, x]) = \frac{9}{10}x$$

czyli  $P((-\infty, x]) = \frac{1}{10} + \frac{9}{10}x$ .

4.  $x \geq 1$  a wtedy tak jak poprzednio dostajemy jeden

Zauważmy, że przypadki 2 i 3 można potraktować łącznie i dostajemy

$$F(x) = \begin{cases} 0 & \text{dla } x < 0 \\ \frac{1}{10} + \frac{9}{10}x & \text{dla } x \in [0, 1) \\ 1 & \text{dla } x \geq 1 \end{cases}$$

a otrzymana funkcja ma skok w punkcie 0 o wielkości  $\frac{1}{10}$  i jest ciągła poza tym punktem.



# 2

## Całka

### 2.1 Całka Riemanna

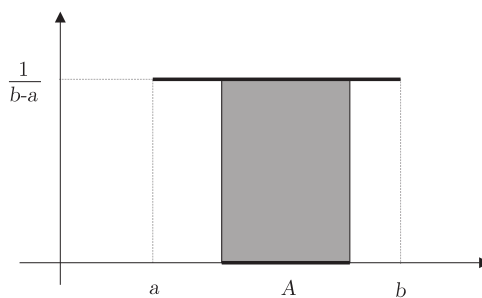
Przypomnijmy, że dla  $\Omega = [a, b]$  z sigma ciałem  $\mathcal{F} = \mathcal{B}_{[a,b]}$  zbiorów borelowskich wprowadziliśmy prawdopodobieństwo wzorem

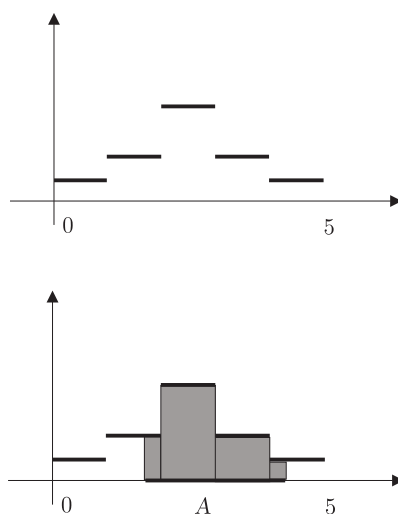
$$P(A) = \frac{1}{b-a}m(A)$$

Możemy je zilustrować graficznie jak na Rysunku 2.1

co pozwala zapisać prawdopodobieństwo w trochę sztuczny sposób ale jest to pożyteczne jako motywacja dalszych kroków: długość przedziału to całka funkcji stałej

$$P([c, d]) = \int_c^d \frac{1}{b-a} dx.$$





W tej interpretacji prawdopodobieństwo to pole pod wykresem pewnej funkcji zawarte nad pewnym przedziałem. Ogólnie, zbiór może być bardziej skomplikowany i wtedy mierzymy go miarą Lebesgue'a

Zastąpmy funkcję stałą przez funkcję  $f$  nieco bardziej skomplikowaną. Niech na przykład  $\Omega = [0, 5]$  i weźmy funkcję o wykresie jak na Rysunku 2.1 z tym że pole pod nią ma być jeden

$$\int_0^5 f(x)dx = 1.$$

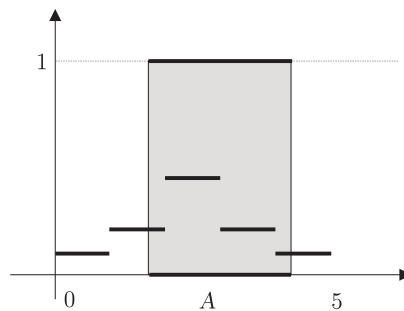
Teraz dla  $A \subset [0, 5]$  kładziemy  $P(A)$  równe polu pod wykresem wyznaczonym przez zbiór  $A$  jak na Rysunku 2.1

Aby obliczyć pole pod wykresem nad zbiorem  $A$  musimy wyeliminować część wykresu która jest poza tym zbiorem. Możemy to zrobić mnożąc naszą funkcję przez zero w punktach poza zbiorem  $A$ . Nad zbiorem  $A$  nie chcemy nic zmieniać, więc mnożymy ją przez jeden. W naturalny sposób pojawia się tak zwana funkcja charakterystyczna zbioru  $A$

$$\mathbb{I}_A(x) = \begin{cases} 1 & \text{gdy } x \in A \\ 0 & \text{gdy } x \notin A \end{cases}$$

Teraz możemy zapisać całkę po zbiorze  $A$  jako całkę po całym zbiorze ze stosownie zmodyfikowanej funkcji

$$P(A) = \int_0^5 \mathbb{I}_A(x)f(x)dx$$



co ilustruje Rysunek 2.1

Podstawowa idea oparta jest na rozbudowaniu wyżej zamieszczonych rysunków przez zwiększenie liczby i zagęszczenie schodków. Intuicja jest oparta na obliczeniu pola zawartego pod wykresem funkcji przyjmującej nieujemne wartości. To pole przybliżamy dzieląc je na paski, które odpowiadają tym słupkom.

Niech

$$f : [a, b] \rightarrow [0, \infty)$$

a przedział dzielimy siatką punktów pośrednich

$$a = c_1 < c_2 < \dots < c_n = b$$

na przykład na równe części:  $c_i = a + i \frac{b-a}{n}$ ,  $i = 0, 1, 2, \dots, n$ . Te kawałki nie są rozłączne ale ich przecięcia to co najwyżej pojedyncze punkty co możemy zignorować, gdyż nie będą miały wpływu na obliczane pole (pole linii to zero). Dla podziału równomiernego mamy

$$|c_{i+1} - c_i| = \frac{b-a}{n} \rightarrow 0$$

ale to co istotne dla ogólnych  $c_i$  to założenie, że te podziały się podobnie zagęszczają wraz ze wzrostem  $n$ . Formalnie, oznaczamy

$$\delta_n = \max_{i=1, \dots, n-1} |c_{i+1} - c_i|$$

i nazywamy tą liczbę średnicą podziału  $\{c_n\}$ . Zakładamy, że

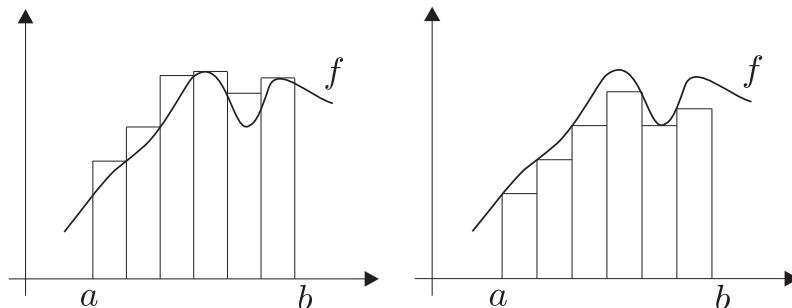
$$\delta_n \rightarrow 0 \quad \text{gdy } n \rightarrow \infty.$$

Definiujemy tak zwane sumy aproksymacyjne górne i dolne:

$$S_n = \sum_{i=1}^{n-1} \max_{x \in [c_i, c_{i+1}]} f(x)(c_{i+1} - c_i) \text{ sumy górne,}$$

$$s_n = \sum_{i=1}^{n-1} \min_{x \in [c_i, c_{i+1}]} f(x)(c_{i+1} - c_i) \text{ sumy dolne,}$$

co ilustruje poniższy rysunek:



Na rysunku mamy wykres funkcji ciągłej i do takich się na razie ograniczamy. Pożądana sytuacja ma miejsce gdy w granicy dostajemy pole pod wykresem. Odzwierciela to formalna definicja

### Definicja 2.1

Jeśli

$$\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} s_n$$

dla dowolnego ciągu podziałów odcinka  $[a, b]$ , to mówimy że  $f$  jest całkowalna w sensie Riemanna i oznaczamy

$$\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} s_n = \int_a^b f(x) dx.$$

Okazuje się, że dla ważnej klasy funkcji całka Riemanna zawsze istnieje.

### Twierdzenie 2.1

Jeśli  $f$  jest ciągła to jest ona całkowalna w sensie Riemanna.

Narzędziem do jej obliczania jest tak zwane podstawowe twierdzenie rachunku różniczkowego i całkowego, które pozwala obliczyć całkę jako przyrost tak zwanej funkcji pierwotnej:

### Twierdzenie 2.2

Jeśli istnieje  $F$  taka że  $F' = f$  to

$$\int_a^b f(x) dx = F(b) - F(a).$$



### Przykład 2.1

Niech  $f(x) = e^{-x}$ , wtedy  $F(x) = -e^{-x}$  i mamy

$$\int_0^n e^{-x} dx = (-e^{-n}) - (-e^0) = 1 - e^{-n}$$

Do zastosowań będzie nam potrzebna tak zwana całka niewłaściwa którą definiujemy następująco

$$\begin{aligned}\int_0^\infty f(x) dx &= \lim_{x_n \rightarrow \infty} \int_0^{x_n} f(x) dx, \\ \int_{-\infty}^\infty f(x) dx &= \lim_{x_n \rightarrow \infty, y_n \rightarrow -\infty} \int_{y_n}^{x_n} f(x) dx.\end{aligned}$$

### Przykład 2.2

Dla funkcji wykładniczej mamy

$$\int_0^\infty e^{-x} dx = \lim(1 - e^{-n}) = 1.$$

Jeśli  $f(x) = \lambda e^{-\lambda x}$ , to  $F(x) = -e^{-\lambda x}$ ,  $\lambda > 0$  i wtedy

$$\int_0^\infty \lambda e^{-\lambda x} dx = \lim(1 - e^{-\lambda n}) = 1.$$

## 2.2 Problem z definicją prawdopodobieństwa

Wykonaliśmy pierwszy krok w kierunku budowy prawdopodobieństwa na dowolnym przedziale z możliwością nierównomiernego rozkładu – będzie ono wyznaczone przez całki z nieujemnych funkcji  $f$  wzorem postaci

$$P(A) = \int \mathbb{I}_A(x) f(x) dx.$$

Matematycznie pojawia się jednak problem – całka Riemanna nie daje własności przeliczalnej addytywności i będziemy musieli ją nieco zmodyfikować. Najpierw przeanalizujemy przykład, który ten problem ilustruje.

Dla  $\Omega = [0, 1]$ , bierzemy  $f(x) = 3x^2$ , i dla  $[a, b] \subset [0, 1]$  definiujemy

$$P([a, b]) = \int_a^b 3x^2 dx = \int_0^1 \mathbb{I}_{[a, b]}(x) 3x^2 dx$$

Chcielibyśmy mieć własność przeliczalnej addytywności: jeśli  $A_n$  są parami rozłączne to

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

W tym celu musimy rozszerzyć definicję tak aby lewa strona powyższej równości miała sens, czyli chcemy mieć możliwość obliczenia prawdopodobieństwa według naszego schematu

$$P(A) = \int_0^1 \mathbb{I}_A(x) 3x^2 dx$$

ale dla ogólniejszych zbiorów  $A$ .

Na przykład, niech  $A = W = \mathbb{Q} \cap [0, 1]$ ,  $W = \{w_n\}$  (oznaczenie  $W$  wprowadzamy aby pamiętać, że chodzi o zbiór liczb wymiernych, zbiór liczb wymiernych jest przeliczalny więc możemy go ustawić w ciąg) czyli mamy

$$W = \bigcup_{n=1}^{\infty} [w_n, w_n].$$

Własność przeliczalnej addytywności daje

$$\begin{aligned} P(W) &= P\left(\bigcup_{n=1}^{\infty} [w_n, w_n]\right) \\ &= \sum_{n=1}^{\infty} P([w_n, w_n]) \\ &= 0 \end{aligned}$$

gdyż dla każdego  $n$

$$P([w_n, w_n]) = \int_{w_n}^{w_n} 3x^2 dx = 0.$$

To się zgadza z interpretacją całki jako pola bo jeśli przedziałem jest punkt, to prostokąt nad nim redukuje się do odcinka, a odcinek ma pole zerowe.

W związku z tym wzór

$$P(A) = \int_0^1 \mathbb{I}_A(x) 3x^2 dx$$

dla  $W = \mathbb{Q} \cap [0, 1]$  powinien dać

$$P(W) = 0.$$

Niestety, prawa strona tego wzoru nie ma sensu gdyż, jak to zaraz pokażemy, funkcja  $\mathbb{I}_W(x) 3x^2$  nie jest całkowalna w sensie Riemanna.

Weźmy podział odcinka  $[0, 1]$ ,  $0 = c_1 < c_2 < \dots < c_n = 1$ , gdzie punkty pośrednie  $c_i$  są wymierne (aby funkcja była całkowalna, aproksymacje dla dowolnych podziałów powinny dać te same wyniki, więc mamy prawo tak wybrać punkty pośrednie). Budujemy sumy aproksymacyjne i dla sum górnych łatwo widać, że ponieważ funkcja jest rosnąca to przyjmuje największą wartość w prawym końcu przedziału który analizujemy

$$\max_{x \in [c_i, c_{i+1}]} \mathbb{I}_W(x) 3x^2 = 3c_{i+1}^2$$

co daje

$$S_n = \sum_{i=1}^{n-1} \max_{x \in [c_i, c_{i+1}]} (\mathbb{I}_W(x) 3x^2) (c_{i+1} - c_i) \rightarrow 1$$

czyli mamy w granicy pole 1. Wprowadzenie funkcji charakterystycznej zbioru  $W$  nic nie zmienia bo sumy aproksymacyjne są takie same.

Przechodzimy do sum dolnych. W każdym przedziale mamy liczby niewymierne, a dla nich nasza funkcja ma wartość zero, stąd minimum na każdym takim przedziale będzie zero

$$\min_{x \in [c_i, c_{i+1}]} \mathbb{I}_W(x) 3x^2 = 0$$

a więc

$$s_n = \sum_{i=1}^{n-1} \min_{x \in [c_i, c_{i+1}]} (\mathbb{I}_W(x) 3x^2) (c_{i+1} - c_i) = 0.$$

Suma zer to zero, mamy więc ciąg sum dolnych zbieżny do zera bo jest to ciąg stały. Jest to inna granica niż dla sum górnych.

Pokazaliśmy, że nasza funkcja nie jest całkowalna w sensie Riemanna. Aby zrealizować plan zdefiniowania prawdopodobieństwa w duchu pola pod wykresem spełniającego własność przeliczalnej addytywności, musimy zmodyfikować definicję całki.

## 2.3 Całka Lebesgue'a

Podstawowa idea konstrukcji całki jest podobna jak poprzednio, a kluczowa różnica to dzielenie na części osi pionowej zamiast poziomej. Wydaje się to mało istotne ale jednak różnica jest fundamentalna. Weźmy więc liczby  $d_i$  stanowiące podział przedziału wartości funkcji  $f$

$$\min_{x \in [a, b]} f(x) = d_1 < d_2 < \dots < d_n > \max_{x \in [a, b]} f(x),$$

i rozważamy przedziały  $[d_i, d_{i+1})$  (tu jesteśmy trochę bardziej staranni i unikamy przecinających się części). Najprościej podzielić cały przedział na pół, potem znów na pół i tak dalej. Czyli mamy kolejne fazy zagęszczania podziału na etapie  $k$  mając

$$d_i = \frac{i}{2^k} : i = 1, \dots, k2^k.$$

Mamy zagwarantowane to, że średnica podziału zmierza do zera. Teraz budujemy sumy aproksymujące pole definiując

$$s_n(x) = \sum_{i=1}^n d_i \mathbb{I}_{f^{-1}([d_i, d_{i+1}))}(x)$$

gdzie

$$\mathbb{I}_A(x) = \begin{cases} 1 & \text{gd } x \in A \\ 0 & \text{gd } x \notin A \end{cases}$$

to funkcja charakterystyczna zbioru  $A$ , a

$$f^{-1}([d_i, d_{i+1})) = \{x : f(x) \in [d_i, d_{i+1}))\}$$

Te zbiory ułożone są na osi poziomej ale to nie muszą być przedziały. Jeśli  $f$  jest ciągła to będą to na pewno sumy przedziałów, czyli zbiory borelowskie. To co chcemy zrobić to obliczyć pola zbiorów aproksymujących pole pod wykresem, analogicznych do prostokątów, co wymaga zmierzenia długości podstawy. Podstawa jest zbiorem borelowskim więc możemy użyć do tego celu miary Lebesgue'a kładąc

$$\sum_{i=1}^n d_i m(f^{-1}([d_i, d_{i+1})))$$

Funkcja  $s_n$  to przykład tzw. funkcji prostej, czyli sumy funkcji charakterystycznych rozłącznych przedziałów pomnożonych przez pewne liczby. Powyższa liczba to całka Lebesgue'a funkcji  $s_n$  co zapisujemy tak

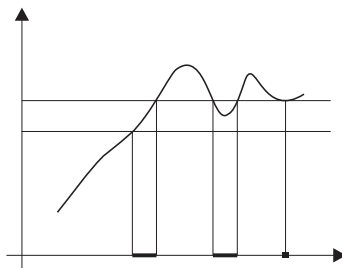
$$\int_{\Omega} s_n dm = \sum_{i=1}^n d_i m(f^{-1}([d_i, d_{i+1})))$$

i ilustrujemy na Rysunku 2.1.

Aby ta wielkość miała sens, musimy umieć zmierzyć zbiory  $f^{-1}([d_i, d_{i+1}))$  miarą Lebesgue'a, więc nakładamy na  $f$  warunek, który zawarty jest w następującej definicji.

## Definicja 2.2

Funkcja  $f$  jest mierzalna jeśli dla każdego borelowskiego zbioru  $B$ , zbiór  $f^{-1}(B)$  jest borelowski.



**Rysunek 2.1** Inna aproksymacja pola pod wykresem.

Na przykład, funkcja charakterystyczna zbioru liczb wymiernych, z którą mieliśmy kłopoty przy całce Riemanna jest mierzalna gdyż przeciwobrazy  $\mathbb{I}_{\mathbb{Q}}^{-1}(B)$  to albo  $\mathbb{Q}$  albo jego uzupełnienie, a są to zbiory borelowskie.

Można pokazać, że wystarczy wymagać aby przeciwobrazy przedziałów były zbiorami borelowskimi. Oczywiście funkcja ciągła jest mierzalna.

Teraz zagęszczamy podział  $\{d_i\}$  i jeśli ciąg całek funkcji  $s_n$  jest zbieżny (tak nie zawsze musi być) to mówimy że funkcja  $f$  jest całkowalna i piszemy

$$\int_{[a,b]} f dm = \lim \int_{[a,b]} s_n dm.$$

Pojawiają się drobne ale istotne problemy: czy istotny jest sposób w jaki przedział wartości został podzielony na kawałki? Czy ma on wpływ na granicę? Wygodniej jest sformułować definicję nieco inaczej przeprowadzając konstrukcję w trzech etapach. Ustalamy zbiór borelowski  $B$  po którym będziemy całkować funkcję  $f$ .

**Etap 1.** Funkcje proste.

Funkcja  $s$  jest prosta jeśli

$$s(x) = \sum s_i \mathbb{I}_{B_i}(x)$$

gdzie zbiory borelowskie  $B_1, \dots, B_n$  są parami rozłączne i  $\bigcup B_i = B$ . Jej całka po zbiorze  $B$  to

$$\int_B s dm = \sum_{i=1}^n s_i m(B_i).$$

**Etap 2.** Funkcje mierzalne nieujemne.

Dla mierzalnej  $f \geq 0$

$$\int_B f dm = \sup \left\{ \int_B s dm : s \leq f \right\}.$$

czyli bierzemy funkcje proste (schodkowe) znajdujące się pod wykresem funkcji  $f$ , obliczamy pola i bierzemy supremum czyli przechodzimy do granicy. Jeśli

powyższa liczba jest skończona to mówmy że  $f$  jest **całkowalna** w sensie Lebesgue'a.

Można pokazać, że dla takich funkcji konstrukcja przez specjalne funkcje  $s_n$  tak jak poprzednio daje dobry wynik: Ciąg całek z funkcji  $s_n$  jest zbieżny do całki z  $f$  jeśli  $f$  jest całkowalna, niezależnie od wyboru podziału  $\{d_n\}$ .

Powyższa definicja może być zastosowana tylko do funkcji nieujemnych. Do celów budowy miar probabilistycznych jest to wystarczające, ale dla innych zastosowań będziemy potrzebować całki funkcji dowolnej. Rozszerzenie definicji jest bardzo proste poprzez zapis dowolnej funkcji przy użyciu funkcji nieujemnych.

**Etap 3.** Funkcje mierzalne.

Kładziemy

$$f = f^+ - f^-, \quad f^+ = \max\{0, f\}, \quad f^- = \max\{0, -f\}$$

oraz mówimy, że  $f$  jest całkowalna jeśli całkowalne są nieujemne funkcje  $f^+, f^-$  i wtedy z definicji

$$\int_B f dm = \int_B f^+ dm - \int_B f^- dm.$$

Dla rozróżnienia obu całek może być wygodnie stosowanie następujących oznaczeń

$$\begin{aligned} \int_a^b f(x) dx & \text{ (całka Riemanna)} \\ \int_{[a,b]} f dm & \text{ (całka Lebesgue'a)} \end{aligned}$$

Niemniej dla szerokiej klasy funkcji to rozróżnienie nie jest istotne.

### Twierdzenie 2.3

Dla funkcji ciągłych obie całki istnieją i są równe.

Jako wniosek mamy możliwość stosowanie klasycznego rachunku całkowego do obliczania całek Lebesgue'a z funkcji ciągłych.

## 2.4 Własności całki

Przyjmujemy niektóre podstawowe własności bez dowodów. Pierwsza mówi o całkowaniu po podzbiorach

**Twierdzenie 2.4**

Jeśli  $A \subset B$  i  $f$  jest całkowalna po  $B$  to jest całkowalna po  $A$  i

$$\int_A f dm = \int_B \mathbb{I}_A f dm.$$

Całka wywodzi się z pola pod wykresem, a pole sumy to suma pól więc nie zaskakuje kolejna własność.

**Twierdzenie 2.5**

Całka jest operacją liniową, to znaczy dla dowolnych całkowalnych  $f, g$  oraz liczb  $a, b$  mamy

$$\int_B (af + bg) dm = a \int_B f dm + b \int_B g dm.$$

Najważniejsza przewaga całki Lebesgue'a nad całką Riemanna to poniższe twierdzenia pokazujące że całka granicy to granica całek.

Wnioskiem z tego twierdzenia jest zakończona addytywność całki jako funkcji zbioru po którym całkujemy.

**Twierdzenie 2.6**

Jeśli  $A \cap B = \emptyset$  to

$$\int_{A \cup B} f dm = \int_A f dm + \int_B f dm.$$

**Dowód**

Rozważmy funkcje  $\mathbb{I}_A, \mathbb{I}_B$  i dzięki temu że zbiory są rozłączne

$$\mathbb{I}_{A \cup B} = \mathbb{I}_A + \mathbb{I}_B$$

więc również

$$\mathbb{I}_{A \cup B} f = \mathbb{I}_A f + \mathbb{I}_B f$$

z liniowości

$$\int \mathbb{I}_{A \cup B} f dm = \int \mathbb{I}_A f dm + \int \mathbb{I}_B f dm$$

ale lewa strona to  $\int_{A \cup B} f dm$  a prawa to  $\int_A f dm + \int_B f dm$ . □

Stosując indukcję możemy rozszerzyć własność addytywności do dowolnej skończonej rodziny zbiorów parami rozłącznych.

Addytywność całki względem zbioru po którym całkujemy pozwala stosować techniki rachunku całkowego dla funkcji które są przedziałami ciągle. Wtedy liczymy całkę dla każdego przedziału osobno i wyniki dodajemy.

### Twierdzenie 2.7

(zbieżność monotoniczna) Jeśli  $f_n, f$  są całkowalne,  $f_n(x) \rightarrow f(x)$  oraz ciąg  $f_n(x)$  jest niemalejący dla każdego  $x$  ( $f_n(x) \leq f_{n+1}(x)$ ) to

$$\int_B f_n dm \rightarrow \int_B f dm.$$

### Twierdzenie 2.8

(zbieżność zdominowana) Jeśli  $f_n, f$  są całkowalne,  $f_n(x) \rightarrow f(x)$  oraz istnieje całkowalna nieujemna funkcja  $g$  taka, że  $|f_n(x)| \leq g(x)$  dla każdego  $x$  i  $n$  to

$$\int_B f_n dm \rightarrow \int_B f dm.$$

Dla nas najważniejszym wnioskiem z tych twierdzeń to możliwość budowy miar probabilistycznych. Niech  $\Omega \subset \mathbb{R}$  będzie zbiorem borelowskim, niech  $f : \Omega \rightarrow [0, \infty)$  będzie całkowalna z tym, że wymagamy aby

$$\int_{\Omega} f dm = 1.$$

Wtedy dla borelowskich  $B \subset \Omega$  kładziemy

$$P(B) = \int_B f dm$$

czyli

$$P(B) = \int_{\Omega} \mathbb{I}_B f dm.$$

### Twierdzenie 2.9

$P$  jest miarą probabilistyczną na  $\Omega$  z sigma ciałem zbiorów borelowskich.



### Dowód

Dowód jest prostym zastosowaniem powyższych twierdzeń więc go podajemy choć może być pominięty. Niech  $A = \bigcup A_n$  gdzie borelowskie zbiory  $A_n$  są parami rozłączne. Chcemy pokazać, że

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Niech  $B_n = \bigcup_{k=1}^n A_k$  i wtedy  $\mathbb{I}_A = \lim \mathbb{I}_{B_n}$ ,  $\mathbb{I}_{B_n} \leq \mathbb{I}_A$ . Funkcja  $g = f\mathbb{I}_A$  jest całkowna, dominuje ciąg  $f_n = f\mathbb{I}_{B_n}$  to znaczy

$$f\mathbb{I}_{B_n} \leq f\mathbb{I}_A,$$

więc możemy skorzystać z twierdzenia o zbieżności zdominowanej

$$\begin{aligned} P(A) &= \int_A f dm = \int_{\Omega} \mathbb{I}_A f dm \quad \text{z definicji} \\ &= \lim_{n \rightarrow \infty} \int_{\Omega} \mathbb{I}_{B_n} f dm \quad \text{z twierdzenia o zbieżności} \\ &= \lim_{n \rightarrow \infty} \int_{B_n} f dm = \lim_{n \rightarrow \infty} P(B_n) \quad \text{z definicji} \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n P(A_k) \quad \text{ze skończonej addytywności} \\ &= \sum_{k=1}^{\infty} P(A_k). \end{aligned}$$

□

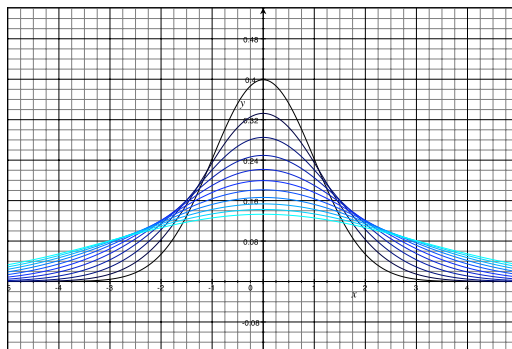
Kończymy ten fragment definicją, która jest umotywowana twierdzeniami o zbieżności. Chodzi o rozszerzenie całki na cały zbiór liczb rzeczywistych. Umiemy całkować po przedziałach, więc naturalna jest następująca konstrukcja. Niech  $B_n = [-n, n]$  i załóżmy, że  $f \geq 0$  jest całkowna po każdym z  $B_n$ . Wtedy definiujemy

$$\int_{\mathbb{R}} f dm = \lim_{n \rightarrow \infty} \int_{[-n, n]} f dm,$$

o ile ta granica istnieje.

## 2.5 Gęstość

Całka daje nam do ręki bardzo wygodne narzędzie do budowania przestrzeni probabilistycznych.



### Definicja 2.3

Całkowalną funkcję  $f$  określoną na borelowskim zbiorze  $\Omega$  nazywamy gęstością jeśli

1.  $f(x) \geq 0$ .
2.  $\int_{\Omega} f dm = 1$ .

Gęstość definiuje miarę probabilistyczną na zbiorach borelowskich wzorem

$$P(B) = \int_B f dm.$$

Niech  $\Omega = [a, b]$ . Gęstość jednostajna to znana nam już funkcja

$$f(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x).$$

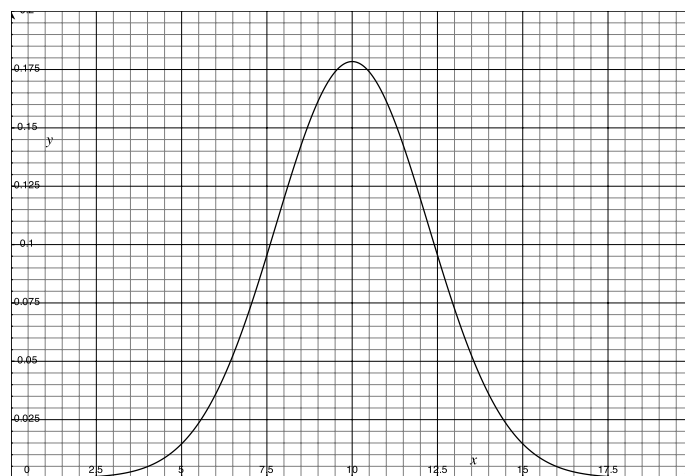
Najważniejszy przykład to gęstość normalna, lub gaussowska dana wzorem

$$n(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

gdzie  $\mu \in \mathbb{R}$ ,  $\sigma \in (0, \infty)$  są dwiema zadanymi liczbami (później poznamy ich interpretację).

Na Rysunku 2.5 mamy wykresy funkcji  $n$  dla  $\mu = 0$  i  $\sigma$  przebiegającego od 1 (kolor czarny) do 3 (kolor niebieski).

Naszukujemy wykres dla  $\mu = 10$ ,  $\sigma = 2,236$  – Rysunek 2.2. Jak zobaczymy później, ten wybór jest związany z dyskutowanym wcześniej rozkładem cen dla 20 kroków w górę i w dół o 5% z prawdopodobieństwem 0.5 (jak na Rysunku 1.1).



**Rysunek 2.2** Gęstość normalna.

Z punktu widzenia modelowania cen niedogodność jest związana z tym, że dopuszczone są ujemne ceny, choć jak można zgadnąć z wykresu gęstości prawdopodobieństwa ich wystąpienia

$$P((-\infty, 0)) = \int_{(-\infty, 0)} n(x) dm(x)$$

będzie małe (później je obliczymy). Klasyczną modyfikacją gęstości normalnej jest tzw. gęstość lognormalna dana wzorem

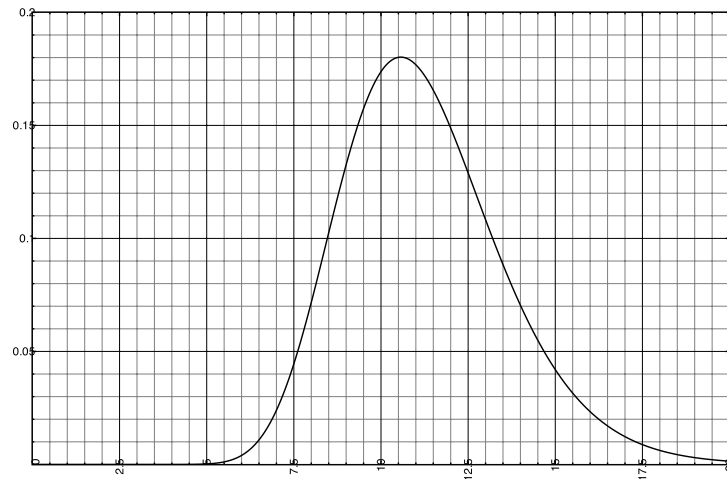
$$l(x) = \frac{1}{x\eta\sqrt{2\pi}} e^{-\frac{(\ln x - \kappa)^2}{2\eta^2}} \quad \text{dla } x > 0, \quad l(x) = 0 \text{ gdy } x \leq 0.$$

Wykres dla  $\kappa = \ln 10$ ,  $\eta = 0.223747$  jest na Rysunku 2.3; parametry są dostosowane do przykładu na Rysunku 1.2. Ta gęstość jest powszechnie stosowana jako model cen akcji i będzie przez nas analizowana w przyszłości.

## 2.6 Dystrybuanta

Ważnym narzędziem do budowy i opisu przestrzeni probabilistycznych, które już poznaliśmy, jest dystrybuanta, czyli funkcja  $F : \mathbb{R} \rightarrow [0, 1]$  dana wzorem

$$F(x) = P((-\infty, x]).$$



**Rysunek 2.3** Gęstość lognormalna.

Jeśli prawdopodobieństwo  $P$  ma gęstość  $f$  to

$$F(x) = \int_{-\infty}^x f(y) dy.$$

Jeśli ta gęstość jest ciągła to

$$F'(x) = f(x).$$

### Przykład 2.3

Jeśli

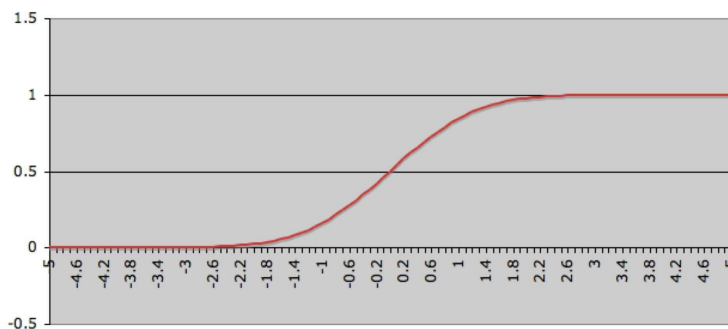
$$f(x) = \mathbb{I}_{[0,1]}(x)$$

to mamy funkcję nieróżniczkowalną, ale różniczkowalną przedziałami

$$F(x) = \begin{cases} 0 & \text{gdy } x < 0 \\ x & \text{gdy } 0 \leq x < 1 \\ 1 & \text{gdy } x \geq 1. \end{cases}$$

Dla gęstości normalnej

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-m)^2} dy$$



Dla tak zwanej standardowej gęstości normalnej co odpowiada wyborowi parametrów  $m = 0$ ,  $\sigma = 1$  oznaczamy

$$N(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

a jej wykres przedstawiamy na Rysunku 2.6

Poprzedni dysktowaliśmy postać dystrybuanty dla miar będących miarami Diraca, ich kombinacjami, jak i mieszaniną miary Diraca i miary Lebesgue'a. Teraz dodamy do tego przykład gdzie łączymy miarę Diraca i miarę posiadającą gęstość. Przypuśćmy, że chcemy zbudować model dla cen akcji firmy która jest narażona na bankructwo. Jak już powiedzieliśmy, naturalnym wyborem jest gęstość lognormalna ale chcemy dopuścić możliwość, że z dodatnim prawdopodobieństwem cena będzie zero. Przypuśćmy, że prawdopodobieństwo bankructwa wynosi  $p > 0$ . Oznaczmy przez  $P_l$  miarę z gęstością lognormalną. Kładziemy

$$P = p\delta_0 + (1 - p)P_l$$

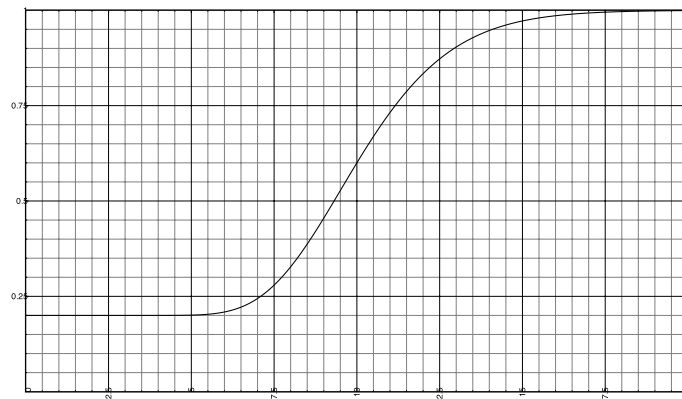
czyli na zbiorach

$$P(A) = p\delta_0(A) + (1 - p) \int_A l(x) dx.$$

W szczególności,  $P(\{0\}) = p$ . Wykres dystrybuanty ma skok w punkcie 0 i jest przedstawiony na Rysunku 2.4 (tylko wykres dla  $x \geq 0$ , dla  $x < 0$  mamy zero).

Na koniec rozdziału szkicujemy możliwość wykorzystania dystrybuanty do budowy całek przez wykorzystanie tzw całki Stieltjesa.

Przypominamy, że dystrybuanta definiuje miarę wzorem  $P_F((x, y]) = F(y) - F(x)$ . Przy konstrukcji całki stosujemy aproksymacje w duchu całki Riemanna and do obliczania pól prostokątów zastąpimy długość podstawy przez miarę  $P_F$  tej podstawy.



**Rysunek 2.4** Dystrybuanta miary z możliwością bankructwa.

Bierzemy funkcję  $h : [a, b] \rightarrow \mathbb{R}$  i dzielimy przedział na zagęszczające się części  $a = c_1 < c_2 < \dots < c_n = b$ . Definiujemy sumy górne

$$S_n = \sum \max_{x \in (c_i, c_{i+1}]} h(x) [F(c_{i+1}) - F(c_i)]$$

i dolne

$$s_n = \sum \min_{x \in (c_i, c_{i+1}]} h(x) [F(c_{i+1}) - F(c_i)].$$

Funkcja  $h$  jest całkowalna w sensie Riemanna-Stieltjesa jeśli

$$\lim s_n = \lim S_n = \int_a^b h(x) dF(x).$$

Rozważmy dwa szczególne przypadki które pozwalają zrozumieć związek tej całki z rozważanymi poprzedni.

Jeśli funkcja  $F$  jest różniczkowalna a  $F'$  jest ciągła, to  $F(c_{i+1}) - F(c_i) \approx F'(x)(c_{i+1} - c_i]$  z twierdzenia o wartości średniej. Wstawiając to do sum górnych i dolnych mamy

$$S_n \approx \sum \max_{x \in (c_i, c_{i+1}]} h(x) F'(x) [(c_{i+1} - c_i)]$$

$$s_n \approx \sum \min_{x \in (c_i, c_{i+1}]} h(x) F'(x) [(c_{i+1} - c_i)]$$

czyli dostajemy sumy górne i dolne dla całki Riemanna funkcji  $hF'$ . Można pokazać, że jeśli  $h$  jest na przykład ciągła to

$$S_n, s_n \rightarrow \int_a^b h(x) F'(x) dx$$

czyli

$$\int_a^b h(x) dF(x) = \int_a^b h(x) F'(x) dx$$

Pamiętamy, że pochodna  $f$  dystrybuanty to gęstość prawdopodobieństwa co pozwala przepisać ostatnią całkę jako  $\int_a^b h(x) f(x) dx$ .

Jeśli  $F$  jest postaci

$$F(x) = \begin{cases} 0 & \text{gdy } x < d \\ 1 & \text{gdy } x \geq d \end{cases}$$

to sumy aproksymacyjne zredukują się do jednego składnika,

$$\begin{aligned} \sum \max_{x \in (c_i, c_{i+1}]} h(x) [F(c_{i+1}) - F(c_i)] &= \max_{x \in (c_i, c_{i+1}]} h(x) [F(c_{i+1}) - F(c_i)] \\ &= \max_{x \in (c_i, c_{i+1}]} h(x) \end{aligned}$$

$$\begin{aligned} \sum \min_{x \in (c_i, c_{i+1}]} h(x) [F(c_{i+1}) - F(c_i)] &= \min_{x \in (c_i, c_{i+1}]} h(x) [F(c_{i+1}) - F(c_i)] \\ &= \min_{x \in (c_i, c_{i+1}]} h(x) \end{aligned}$$

dla którego mamy  $d \in (c_i, c_{i+1}]$  gdyż  $F(c_{i+1}) - F(c_i)$  jest jeden, a dla pozostałych składników  $F$  jest stała, więc jej przyrosty będą zerami. Zagęszczając podziały, zakładając że  $h$  jest ciągła, dostajemy w granicy

$$\begin{aligned} \max_{x \in (c_i, c_{i+1}]} h(x) &\rightarrow h(d) \\ \min_{x \in (c_i, c_{i+1}]} h(x) &\rightarrow h(d) \end{aligned}$$

więc

$$\int h(x) dF(x) = h(d)$$

W sytuacji gdy  $F$  ma skok wielkości  $p$  to mamy  $F(c_{i+1}) - F(c_i) = p$  i w granicy dostajemy  $ph(d)$ .





# 3

## Zmienne losowe

### 3.1 Definicja zmiennej losowej

Jedno z głównych zadań matematyki finansowej to wycena instrumentów pochodnych. Są to walory finansowe wyznaczone przez losowe wypłaty w przyszłości, wypłaty uzależnione od wartości innych walorów, tzw. walorów bazowych. Niech tym walorem bazowym będzie akcja, oraz niech wypłata instrumentu pochodnego będzie zadana w chwili  $T$  jako pewna funkcja ceny akcji. Czyli rozważamy wielkość postaci  $f(S(T))$ , na przykład  $\max\{S(T) - K, 0\}$ , co definiuje opcję kupna. Jeśli mamy model przyszłych cen akcji które tworzą zbiór  $\Omega$  to wypłata instrumentu pochodnego jest funkcją określoną na zbiorze  $\Omega$ , na przykład dla opcji kupna mamy  $\omega \rightarrow \max\{\omega - K, 0\}$ . Takie funkcje nazywamy zmiennymi losowymi. Dla zmiennych losowych będziemy używać liter  $X, Y, Z$  w ogólności, a dla instrumentów pochodnych  $H$

$$H : \Omega \rightarrow \mathbb{R}$$

a wartość zmiennej losowej odpowiadającą wynikowi  $\omega$  będziemy na ogół oznaczać przez  $H^\omega$ .

#### Uwaga 3.1

Wybór zbioru  $\Omega$  jest do pewnego stopnia dowolny. Sama cena akcji może być traktowana jako zmienna losowa. Pokażemy niebawem, że równoważne podejście do przyjęcia  $(0, \infty)$  jako zbioru możliwych cen o rozkładzie lognormalnym

jest przyjęcie  $\Omega = R$ , zadanie na tym zbiorze gęstości normalnej i potraktowanie ceny jako funkcji  $S(T) = e^\omega$ . Wtedy opcja kupna będzie zmienną losową postaci  $\max\{e^\omega - K, 0\}$ .

Dla zmiennej losowej  $H$  kluczowe problem to odpowiedź na pytanie: jakie jest prawdopodobieństwo, że wartości zmiennej losowej są w danym przedziale, czyli chodzi nam o obliczenie następującej wielkości

$$P(\{\omega : H^\omega \in [a, b]\}).$$

Aby to było możliwe, zbiór  $\{\omega : H^\omega \in [a, b]\}$  który zapisujemy jako  $H^{-1}([a, b])$  musi należeć do dziedziny funkcji  $P$ , którą jest sigma ciało  $\mathcal{F}$ . Nakładamy więc stosowny warunek.

### Definicja 3.1

Funkcja  $X : \Omega \rightarrow R$  jest zmienną losową jeśli dla każdego przedziału  $[a, b]$

$$X^{-1}([a, b]) \in \mathcal{F}$$

gdzie  $\mathcal{F}$  jest ustalonym sigma ciałem na zbiorze  $\Omega$ . Mówimy wtedy, że  $X$  jest mierzalna, lub  $\mathcal{F}$ -mierzalna.

Jak zobaczymy niebawem, w pewnych sytuacjach będą uzasadnione różne możliwości wyboru sigma ciała i tym samym pojęcie zmiennej losowej stanie się względne. W sytuacji najprostszej, gdy  $\Omega$  to zbiór skończony a  $\mathcal{F}$  to rodzina wszystkich jego podzbiorów, każda funkcja określona na  $\Omega$  jest zmienną losową.

Najprostszy przykład polega na rozważeniu najmniejszego możliwego sigma ciała, czyli  $\mathcal{F} = \{\emptyset, \Omega\}$  (wszystkie warunki nałożone na sigma ciało są oczywiście spełnione) i wtedy zmienne losowe są to funkcje stałe. Jeśli  $X^\omega = c$  dla każdego  $\omega$ , to dla przedziałów zawierających  $c$  mamy  $X^{-1}([a, b]) = \Omega$  a jeśli  $a \notin [a, b]$  to  $X^{-1}([a, b]) = \emptyset$ . Jasne jest, że jedynie funkcje stałe są mierzalne względem  $\mathcal{F} = \{\emptyset, \Omega\}$ .

## 3.2 Rozkład zmiennej losowej

Jak już odnotowaliśmy, wybór przestrzeni probabilistycznej jest do pewnego stopnia dowolny. To co jest najważniejsze, to prawdopodobieństwa z jakim zmienne losowe przyjmują konkretne wartości. Założenie o mierzalności pozwala nam zdefiniować nową miarę prawdopodobieństwa ułożoną na zbiorze wartości zmiennej losowej

### Definicja 3.2

Rozkład zmiennej losowej  $X$  to funkcja  $P_X$  określona na sigma ciele  $\mathcal{B}$  zbiorów borelowskich w  $\mathbb{R}$  wzorem

$$P_X(B) = P(X^{-1}(B)).$$

To jest dobrze określone pojęcie bo z definicji mierzalności,  $X^{-1}(B) \in \mathcal{F}$  więc prawa strona jest zdefiniowana.

### Twierdzenie 3.1

$P_X$  jest miarą prawdopodobieństwa na  $\mathbb{R}$ .

### Dowód

Dowód jest bardzo prosty. Po pierwsze  $P_X(\emptyset) = P(\emptyset) = 0$ ,  $P_X(\mathbb{R}) = P(\Omega) = 1$  a przeliczalna addytywność wynika z własności operacji brania przeciwobrazu zbioru i przeliczalnej addytywności  $P$

$$\begin{aligned} P_X\left(\bigcup B_n\right) &= P\left(X^{-1}\left(\bigcup B_n\right)\right) \quad \text{z definicji funkcji } P_X \\ &= P\left(\bigcup X^{-1}(B_n)\right) \quad \text{własność przeciwobrazu} \\ &= \sum P\left(X^{-1}(B_n)\right) \quad P \text{ jest miarą} \\ &= \sum P_X(B_n) \quad \text{z definicji } P_X. \end{aligned}$$

□

Najprostszy przykład to zmienna losowa o rozkładzie dyskretnym. Taka zmienna losowa ma skończoną lub przeliczalną liczbę wartości  $x_n$  i przyjmuje je z dodatnimi prawdopodobieństwami:

$$P(\{\omega : X^\omega = x_n\}) = p_n > 0, \quad n = 1, 2, 3, \dots$$

Wprowadźmy jeszcze oznaczenie na dystrybuantę:

$$F_X(x) = P_X((-\infty, x])$$

która ma własności odnotowane poprzednio, gdyż  $P_X$  jest miarą probabilistyczną.

Powyższe definicje pozwalają nam uwypuklić względność wyboru zbioru  $\Omega$ . Powiedzmy, że  $X : \Omega \rightarrow \mathbb{R}$ , a na  $\Omega$  mamy sigma ciało  $\mathcal{F}$  i miarę  $P$ . Definiujemy

nową przestrzeń probabilistyczną (tak zwaną kanoniczną) kładąc

$$\Omega' = \mathbb{R}$$

$$\mathcal{F}' = \mathcal{B}$$

$$P' = P_X$$

Na  $\Omega'$  definiujemy nową zmienną losową dość trywialnym sposobem

$$(X')^\omega = \omega.$$

Wtedy jest oczywiste, że rozkład tej nowej zmiennej losowej jest taki sam jak rozkład zmiennej  $X$

$$P_{X'} = P_X$$

gdyż

$$\begin{aligned} P_{X'}(B) &= P'(X'^{-1}(B)) = P'(B) \quad \text{bo } X' \text{ jest identycznością} \\ &= P_X(B) \quad \text{z definicji } P'. \end{aligned}$$

### 3.3 Gęstość

Znane już nam pojęcie gęstości miary dopasowujemy do rozkładu zmiennej losowej.

#### Definicja 3.3

Funkcja  $f_X$  jest gęstością zmiennej losowej  $X$  jeśli

$$P_X(B) = \int_B f_X dm.$$

Innymi słowy, gęstość zmiennej losowej to gęstość miary będącej jej rozkładem.

Zmienna losowa o rozkładzie dyskretnym nie ma gęstości gdyż dla dowolnej funkcji  $f$

$$\int_{\{x_n\}} f dm = 0$$

ale  $P_X(\{x_n\}) = p_n > 0$ .

Najprostszy przykład gęstości to znana nam już funkcja

$$f(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x)$$

i jeśli ona jest gęstością zmiennej losowej  $X$  to mówimy, że  $X$  ma rozkład jednostajny w przedziale  $[a, b]$ . Konkretny przykład takiej zmiennej losowej może być skonstruowany na  $\Omega = [0, 1]$  z miarą Lebesgue'a wzorem

$$X^\omega = a + (b - a)x$$

ale wzór

$$X^\omega = b - (b - a)x$$

również definiuje zmienną losową w rozkładzie jednostajnym na  $[a, b]$ . Oba wzory dają funkcje liniowe przekształcające  $[0, 1]$  na  $[a, b]$ . Nie jest trudno znaleźć więcej wzorów dających ten sam efekt.

Najważniejszy przykład to gęstość normalna: zmienna losowa  $X$  ma gęstość normalną jeśli

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

dla pewnych liczb  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ , czyli

$$P_X([a, b]) = P(a \leq X \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

(w praktyce wystarczy ograniczyć się do przedziałów, a całka Lebesgue'a sprowadza się wtedy do całki Riemanna). W takim przypadku piszemy krótko że  $X$  ma rozkład  $N(\mu, \sigma)$ . Dystrybuenta zmiennej o rozkładzie normalnym z  $\mu = 0$  i  $\sigma = 1$ , czyli o tzw. standardowym rozkładzie normalnym, jest oznaczona przez

$$N(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Ta funkcja jest dostępna w Excelu (jak i w każdym języku programowania) i daje na natychmiastową odpowiedź na pytanie: jakie jest prawdopodobieństwo że  $X$  należy do pewnego przedziału: jeśli  $X$  ma rozkład  $N(0, 1)$  to

$$P(X \in [a, b]) = N(b) - N(a).$$

Zauważmy, że  $P(X = c) = 0$  gdyż całka po zbiorze  $\{c\}$  jest zero. W związku z tym prawdopodobieństwa te będą mieć takie same wartości dla wszystkich przedziałów o końcach  $a, b$ , czyli na przykład  $P_X([a, b]) = P_X((a, b))$ . Rachunek różniczkowy mówi nam że funkcja  $N$  ma pochodną, ta pochodna to gęstość

$$N'(x) = f_X(x).$$

To samo zachodzi dla dowolnego rozkładu normalnego, niekoniecznie standardowego

$$F'_X = f_X.$$

Ponieważ funkcja  $f_X$  jest zawsze dodatnia,  $F_X$  jest silnie rosnąca w związku z tym posiada funkcję odwrotną  $F_X^{-1}$ . Dla rozkładu standardowego funkcja  $N^{-1}$  jest również dostępna w Excelu i odpowiada nam na istotne pytanie: mając dane prawdopodobieństwo  $p$  wyznaczyć  $a$  takie, że  $P(X \leq a) = p$ . Pokażemy teraz jak praktycznie skonstruować zmienną losową o rozkładzie normalnym. Będzie ona dana wzorem, jako funkcja określona na przedziale  $[0, 1]$ . Nie jest to jedyna możliwość, a inne, które poznamy później, są bardziej efektywne numerycznie i pozwalają na symulacje komputerowe.

Rozważamy więc  $\Omega = [0, 1]$ , z sigma ciałem  $\mathcal{F}$  składającym się ze zbiorów borelowskich, z miarą Lebesgue'a  $P$ .

### Twierdzenie 3.2

$N^{-1}$  ma standardowy rozkład normalny.

#### Dowód

Liczymy dystrybuantę zmiennej losowej  $N^{-1}$

$$\begin{aligned} F_{N^{-1}}(a) &= P(\{\omega \in [0, 1] : (N^{-1})^\omega \leq a\}) \quad \text{definicja dystrybuanaty} \\ &= P(\{\omega \in [0, 1] : N((N^{-1})^\omega) \leq N(a)\}) \quad \text{obkładając obie strony funkcją } N \\ &= P(\{\omega \in [0, 1] : \omega \leq N(a)\}) \quad \text{z definicji funkcji odwrotnej} \\ &= P([0, N(a)]) \quad \text{zbiór pod } P \text{ to przedział} \\ &= N(a) \quad \text{gdyż } P \text{ jest miarą Lebesgue'a.} \end{aligned}$$

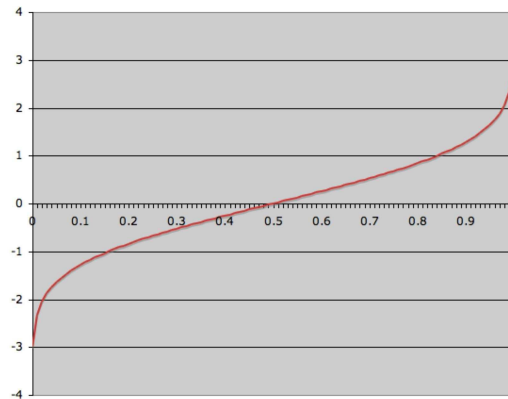
□

Na Rysunku 3.3 przedstawiony jest wykres zmiennej losowej  $N^{-1} : [0, 1] \rightarrow \mathbb{R}$ .

Tą konstrukcję możemy przeprowadzić dla dowolnej zmiennej losowej której dystrybuenta jest funkcją ściśle rosnącą a więc odwracalną. Wtedy

$$Y = F_X^{-1}$$

rozważana na przestrzeni  $\Omega = [0, 1]$  z miarą Lebesgue'a ma taki sam rozkład jak  $X$ . Aby to zobaczyć powatrazamy powyższy dowód w ogólneuch oznaczeniach:



Wystarczy zauważyć, że  $X$  i  $Y$  mają tę samą dystrybuantę

$$\begin{aligned}
 F_Y(a) &= P(\{\omega : Y^\omega \leq a\}) && \text{definicja } F_Y \\
 &= P(\{\omega : F_X^{-1}(\omega) \leq a\}) && \text{definicja } Y \\
 &= P(\{\omega : \omega \leq F_X(a)\}) \\
 &= P([0, F_X(a)]) \\
 &= F_X(a) && \text{gdyż } P = m.
 \end{aligned}$$

Zaletą tego punktu widzenia jest następująca: pakiety komputerowe czy języki programowania, mają wbudowane tzw. generatory liczb losowych i przy ich użyciu można symulować liczby losowe z przedziału  $[0, 1]$  z rozkładem jednostajnym, czyli z  $P = m$ . Znając dystrybuantę  $F_X$ , co ma na przykład miejsce dla standardowego rozkładu normalnego, możemy wtedy symulować wartości zmiennej  $Y$  a na podstawie powyższego faktu, symulujemy w ten sposób zmienną  $X$ .

Inną zaletą takiego wyboru przestrzeni probabilistycznej to możliwość szkicowania wykresów zmiennych losowych. Kolejną zaletą będzie pokazana niebawem, gdy będziemy całkować zmienne losowe. Przy takim podejściu będą to całki po przedziale  $[0, 1]$ , co jest dogodne pojęciowo oraz numerycznie.

Przypomnijmy jeszcze pojęcie gęstości lognormalnej w kontekście zmiennych losowych co pozwoli nam lepiej zrozumieć występujące tam parametry.

### Definicja 3.4

Mówimy, że  $Y$  ma rozkład lognormalny gdy  $X = \ln Y$  ma rozkład normalny.

Wyznamy gęstość rozkładu zmiennej  $Y$  :

$$\begin{aligned} P(a \leq Y \leq b) &= P(\ln a \leq \ln Y \leq \ln b) \quad \text{gdyż } X = \ln Y \\ &= \int_{\ln a}^{\ln b} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad \text{bo } X \text{ ma rozkład normalny.} \end{aligned}$$

Analiza matematyczna daje nam następujący wzór

$$\int_a^b f(g(y))g'(y)dy = \int_{g(a)}^{g(b)} f(x)dx,$$

który dla  $g(y) = \ln y$  daje ( $g'(y) = \frac{1}{y}$ ,  $f(x) = e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ )

$$\int_{\ln a}^{\ln b} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_a^b \frac{1}{y} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} dy.$$

Korzystamy z tego do innego zapisania całki

$$\int_{\ln a}^{\ln b} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_a^b \frac{1}{y\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} dy.$$

To daje postać gęstości zmiennej  $Y$

$$f_Y(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}}.$$

### 3.4 Przykłady finansowe

Rozpatrzmy konkretny przykład: niech  $X$  ma rozkład  $N(0, 1)$  i zdefiniujmy

$$Y = \mu T + \sigma\sqrt{T}X.$$

Wtedy  $Y$  ma rozkład  $N(\mu T, \sigma\sqrt{T})$  co widać z następujących przekształceń

$$\begin{aligned} P(Y \leq a) &= P(\mu T + \sigma\sqrt{T}X \leq a) \\ &= P(X \leq \frac{a - \mu T}{\sigma\sqrt{T}}) \\ &= \int_{-\infty}^{\frac{a - \mu T}{\sigma\sqrt{T}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad \text{bo } X \text{ ma rozkład } N(0, 1) \\ &= \int_{-\infty}^{a - \mu T} \frac{1}{\sigma\sqrt{T}\sqrt{2\pi}} e^{-\frac{y^2}{2T\sigma^2}} dy \quad \text{po zmianie zmiennych } y = x\sigma\sqrt{T} \\ &= \int_{-\infty}^a \frac{1}{\sigma\sqrt{T}\sqrt{2\pi}} e^{-\frac{(z - \mu T)^2}{2T\sigma^2}} dz \quad \text{podstawiając } z = y + \mu T. \end{aligned}$$



Przyjmując ceny akcji postaci

$$S(T) = S(0)e^{\mu T + \sigma\sqrt{T}X}.$$

widzimy, że mają one rozkład lognormalny.

Możemy wypisać wzór na gęstość

$$f_{S(T)}(x) = \frac{1}{x\sqrt{2\pi T}\sigma^2} e^{-\frac{(\ln x - \ln S(0) - \mu T)^2}{2T\sigma^2}}$$

co powstaje z ogólnego wzoru, który przypominamy

$$f_Y(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

zastępując  $\mu$  przez  $\ln S(0) + \mu T$ , oraz  $\sigma$  przez  $\sigma\sqrt{T}$ .

Tak jak powiedzieliśmy, możemy rozpatrywać  $X$  na  $\Omega = [0, 1]$ . Stosując funkcję `LOS()` w Excelu generujemy miare probabilistyczną na  $[0, 1]$ , a podstawiając uzyskiwane liczby w miejsce  $X$  otrzymujemy symulację cen akcji. Możemy też symulować wykładnik  $\mu T + \sigma\sqrt{T}X$  jeśli dysponujemy odpowiednim generatorem liczb losowych pozwalającym od razu symulować liczby z zadany rozkładem normalnym.

Rozwiążemy jeszcze problem wyznaczania rozkładu prostego instrumentu pochodnego, mianowicie opcji kupna o wypłacie  $H = \max\{S(T) - K, 0\}$ . Po pierwsze odnotujemy, że  $P(H < 0) = 0$  bo wypłata jest nieujemna.. Dla  $a > 0$

$$\begin{aligned} & \{\omega : (S(T)^\omega - L)^+ \leq a\} \\ &= \{\omega : S(T)^\omega - L \leq 0\} \cup \{\omega : S(T)^\omega - L > 0 \text{ oraz } S(T)^\omega - L \leq a\} \\ &= \{\omega : S(T)^\omega \leq L + a\} \end{aligned}$$

więc

$$\begin{aligned} P(H \leq a) &= P((S(T) - L)^+ \leq a) \\ &= P(S(T) - L \leq a) \\ &= P(S(0)e^{\mu T + \sigma\sqrt{T}X} \leq L + a) \\ &= P(X \leq \frac{\ln \frac{L+a}{S(0)} - \mu T}{\sigma\sqrt{T}}) \\ &= N(d(a)) \end{aligned}$$

gdzie

$$d(a) = \frac{\ln \frac{L+a}{S(0)} - \mu T}{\sigma\sqrt{T}}.$$

Otrzymujemy więc

$$F_H(a) = \begin{cases} 0 & \text{gdy } a < 0, \\ N(d(a)) & \text{gdy } a \geq 0. \end{cases}$$

Dystrybuanta ma skok więc nie ma gęstości. Rozkład zmiennej  $H$  jest kombinacją miary Diraca i miary  $Q$  która posiada gęstość

$$P_H = \delta_0 N(d(0)) + Q$$

Tą gęstość możemy wyliczyć różniczkując dystrybuantę dla  $a > 0$

$$\begin{aligned} F'_H(a) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\ln \frac{K+a}{S(0)} - \mu T)^2}{2\sigma^2 T}\right) \frac{1}{\sigma\sqrt{T}} \frac{1}{K+a} \\ &= f_{S(T)}(K+a) \end{aligned}$$

porównując z gęstością zmiennej  $S(T)$ .

## 3.5 Wartość oczekiwana i wariancja

### 3.5.1 Przypadek dyskretny

Wróćmy do prostych sytuacji prawdopodobieństw wyznaczonych przez skończone (lub przeliczalne) zbiory dopuszczalnych wartości.

Rozkład zmiennej losowej daje nam informację o umiejscowieniu na osi liczbowej wartości zmiennych losowych. Niekiedy pełna informacja jest mało przydatna, gdyż jest zbyt szczegółowa, a ze względów praktycznych i poglądowych, na przykład do porównań, wolelibyśmy informację uproszczoną. Wprowadzimy pewne liczbowe opisy zmiennych losowych opisujących je syntetycznie a pierwsza jest oparta na idei "środka" właściwie rozumianego.

Wróćmy do cen akcji w czasie dyskretnym gdzie  $S(4)$  przy zmianach o 5% w jednym kroku miało pięć możliwych wartości postaci  $10 \times 1.05^n \times 0.95^{4-n}$ , czyli

$$S(4) \in \{8.1450625, 9.0024375, 9.9500625, 10.9974375, 12.1550625\}.$$

Wartość środkowa to 9.9500625 (tzw. mediana) ale jest jasne że informacja o zmiennej losowej podające tylko medianę nie jest dostatecznie wyczerpująca choćby dlatego, że ignoruje prawdopodobieństwa. Wprowadzamy pojęcie średniej ważonej, gdzie wagami są poszczególne prawdopodobieństwa przyjmowania

konkretnych wartości

$$\begin{aligned} & 8.1450625 \times 0.0625 + 9.0024375 \times 0.25 + 9.9500625 \times 0.375 \\ & + 10.9974375 \times 0.25 + 12.1550625 \times 0.0625 \\ & = 10. \end{aligned}$$

Jeśli zmiany byłyby o 8% w górę a 5% w dół w jednym kroku, to otrzymalibyśmy w przybliżeniu 10.6136.

Ogólnie, przypuśćmy, że zmienna losowa  $X$  przyjmuje wartości  $x_i$  (zbiór skończony lub ciąg) z prawdopodobieństwami  $p_i$

$$p_i = P(\{\omega : X^\omega = x_i\}) = P(X = x_i)$$

### Definicja 3.5

Wartość średnia, lub oczekiwana zmiennej losowej  $X$  to liczba

$$\mathbb{E}(X) = \sum x_i p_i$$

ze stosownym zakresem sumowania, skończonym lub nie.

W szczególnym przypadku gdy  $\Omega$  ma  $N$  elementów a rozkład jest jednostajny mamy  $p_i = \frac{1}{N}$  i wartość oczekiwana to po prostu średnia arytmetyczna

$$\mathbb{E}(X) = \frac{1}{N} \sum_{i=1}^N x_i.$$

### Uwaga 3.2

Terminologia: "wartość oczekiwana" nie jest zbyt fortunna. Przypuśćmy że rzucamy kostką do gry a zmienna losowa to liczba kropek. Wtedy  $E(X) = 3.5$  i jasne jest że nazywanie tego wartością oczekiwaną nie jest dobrym pomysłem, bo nigdy się nie doczekamy takiego wyniku. Niemniej, ten termin jest uświęcony i trudno z tym walczyć.

Jeśli zmiany w jednym kroku wynoszą  $\pm 8\%$  zamiast  $\pm 5\%$  to wartość oczekiwana będzie taka sama. Pojawia się konieczność wprowadzenia liczbowej charakterystyki stopnia rozproszenia wartości. W tym celu obliczamy średnią wa-

zoną kwadratu odległości wyniku od wartości oczekiwanej:

$$\begin{aligned} & (8.1450625 - 10)^2 \times 0.0625 + (9.0024375 - 10)^2 \times 0.25 \\ & + (9.9500625 - 10)^2 \times 0.375 + (10.9974375 - 10)^2 \times 0.25 \\ & + (12.1550625 - 10)^2 \times 0.0625 \\ & \approx 1.003756. \end{aligned}$$

Ta liczba to wariancja zmiennej losowej. Gdy zmiany w jednym kroku będą  $\pm 8\%$  to otrzymamy 2,584681. Ogólne, przy rozkładzie dyskretnym

$$\text{Var}(X) = \sum (x_i - \mathbb{E}(X))^2 p_i$$

co daje motywację ogólnej definicji

### Definicja 3.6

Wariancję zmiennej losowej  $X$  nazywamy liczbę

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2],$$

a jej pierwiastek to odchylenie standardowe oznaczane przez

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

Jako przykład rozważmy najprostszą zmienną losową, mianowicie dla  $\Omega = \{U, D\}$  z prawdopodobieństwami  $p_U, p_D$ ,  $p_U + p_D = 1$  mamy zmienną losową  $X$  o wartościach  $X^U, X^D$ . jej wartość oczekiwana to

$$\mathbb{E}(X) = p_U X^U + p_D X^D.$$

Liczmy wariancję:

$$\begin{aligned} \text{Var}(X) &= (X^U - \mathbb{E}(X))^2 p_U + (X^D - \mathbb{E}(X))^2 p_D \quad \text{z definicji} \\ &= (X^U - p_U X^U - p_D X^D)^2 p_U + (X^D - p_U X^U - p_D X^D)^2 p_D \\ &= (p_D X^U - p_D X^D)^2 p_U + (p_U X^D - p_U X^U)^2 p_D \\ &\quad \text{bo } 1 - p_U = p_D, \quad 1 - p_D = p_U \\ &= (X^U - X^D)^2 (p_D^2 p_U + p_U^2 p_D) \\ &= (X^U - X^D)^2 p_U p_D (p_D + p_U) \\ &= (X^U - X^D)^2 p_U p_D \quad \text{bo } p_U + p_D = 1 \end{aligned}$$

Dla odchylenia standardowego mamy wzór

$$\sigma_X = |X^U - X^D| \sqrt{p_U p_D}.$$

### 3.5.2 Własności

Nie jest trudno sprawdzić, że dla zmiennych losowych o rozkładzie dyskretnym zachodzą następujące związki

$$\begin{aligned}\mathbb{E}(aX) &= a\mathbb{E}(X), \\ \mathbb{E}(X + Y) &= \mathbb{E}(X) + \mathbb{E}(Y),\end{aligned}$$

i dalej jako wniosek mamy następujący wzór

#### Twierdzenie 3.3

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

#### Dowód

Poniższy dowód opiera się na powyższych własnościach (liniowość  $\mathbb{E}$ ) i nie wykorzystuje tego że mamy do czynienia z rozkładem dyskretnym (bo tylko dla takiego mamy definicję wartości oczekiwanej). będzie mógł być więc wykorzystany w innych sytuacjach. Zaczynamy od definicji wariancji

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}(X))^2] \\ &= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \quad \text{podnosimy różnicę do kwadratu} \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2\mathbb{E}(1) \quad \text{liniowość } \mathbb{E}, \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 \quad \text{bo } \mathbb{E}(1) = 1 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2.\end{aligned}$$

□

#### Twierdzenie 3.4

$$\begin{aligned}\text{Var}(aX) &= a^2\text{Var}(X) \\ \sigma_{aX} &= |a|\sigma_X\end{aligned}$$

#### Dowód

Wariancja to operacja oparta na kwadracie stąd nie dziwi dowodzony fakt jak

i to że wynika on wprost z liniowości wartości oczekiwanej:

$$\begin{aligned}\mathrm{Var}(aX) &= \mathbb{E}((aX)^2) - \mathbb{E}(aX)^2 && \text{poprzednie twierdzenie} \\ &= a^2(\mathbb{E}(X^2) - \mathbb{E}(X)^2) && \text{liniowość } \mathbb{E}, \\ &= a^2\mathrm{Var}(X).\end{aligned}$$

Wzór dla odchylenia standardowego wynika już natychmiast:

$$\sigma_{aX} = \sqrt{\mathrm{Var}(aX)} = \sqrt{a^2\mathrm{Var}(X)} = |a| \sqrt{\mathrm{Var}(X)} = |a|\sigma_X.$$

□

Wariancja mierzy rozproszenie, a to nie zmienia się jeśli przesunąć wartości o stałą, stąd poniższa własność.

### Twierdzenie 3.5

$$\mathrm{Var}(X + a) = \mathrm{Var}(X)$$

#### Dowód

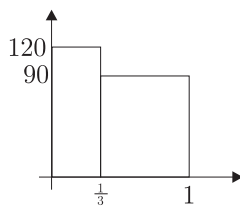
Korzystając z powyższego wzoru na wariancję do dowodu potrzeba tylko odrobina elementarnej algebry

$$\begin{aligned}\mathrm{Var}(X + a) &= \mathbb{E}((X + a)^2) - (\mathbb{E}(X + a))^2 \\ &= \mathbb{E}(X^2 + 2aX + a^2) - \mathbb{E}(X)^2 - 2a\mathbb{E}(X) - a^2 && \text{wzór na kwadrat sumy} \\ &= \mathbb{E}(X^2) + 2a\mathbb{E}(X) + a^2 - \mathbb{E}(X)^2 - 2a\mathbb{E}(X) - a^2 && \text{liniowość } \mathbb{E} \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= \mathrm{Var}(X)\end{aligned}$$

□

Odnotujmy wnosiek, że dla liczenia wariancji można założyć, że  $\mathbb{E}(X) = 0$ . Mianowicie biorąc  $X' = X - \mathbb{E}(X)$ ,

mamy  $\mathbb{E}(X') = 0$ , a wariancje obu tych zmiennych są takie same  $\mathrm{Var}(X') = \mathrm{Var}(X)$ .



### 3.5.3 Przypadek ogólny

Zaczynamy od prostego przykładu motywującego definicję. Niech  $X$  będzie określona na  $\Omega = [0, 1]$  wzorem

$$X^\omega = \begin{cases} 120 & \text{dla } \omega \in [0, \frac{1}{3}], \\ 90 & \text{dla } \omega \in (\frac{1}{3}, 1]. \end{cases}$$

Rozkład  $X$  jest dyskretny,

$$\begin{aligned} x_1 &= 120, & p_1 &= \frac{1}{3}, \\ x_2 &= 90, & p_2 &= \frac{2}{3} \end{aligned}$$

i możemy obliczyć wartość oczekiwaną

$$\mathbb{E}(X) = x_1 p_1 + x_2 p_2 = \frac{1}{3} \times 120 + \frac{2}{3} \times 90$$

Możemy spojrzeć na to obliczenie inaczej: jest to pole pod wykresem pewnej funkcji, jak na Rysunku 3.5.3, czyli całka z tej funkcji. Motywuje to definicję

#### Definicja 3.7

Dla  $X$  określonej na  $\Omega = [0, 1]$  jej wartość oczekiwana to całka

$$\mathbb{E}(X) = \int_{[0,1]} X dm.$$

Wtedy mamy wzór na wariancję

$$\text{Var}(X) = \int_{[0,1]} (X - \mathbb{E}(X))^2 dm = \int_{[0,1]} (X^\omega - \mathbb{E}(X))^2 dm(\omega).$$

Ta definicja może być dalej uogólniona, przez zastąpienie przedziału  $[0, 1]$  przez dowolny zbiór  $\Omega$  ale w tym celu potrzebujemy rozszerzenia pojęcia całki.

Definiujemy całkę po  $\Omega$  względem  $P$  z funkcji  $X : \Omega \rightarrow \mathbb{R}$  mierzalnej powtarzając schemat definicji całki Lebesgue'a:

**1 Krok.** Całka funkcji prostej  $s = \sum_{i=1}^n s_i \mathbb{I}_{A_i}$ , gdzie  $A_i \in \mathcal{F}$  są parami rozłączne to

$$\int_{\Omega} s dP = \sum_{i=1}^n s_i P(A_i).$$

**2. Krok.** Całka funkcji nieujemnej  $X \geq 0$

$$\int_{\Omega} X dP = \sup \left\{ \int_{\Omega} s dP : s \leq X \right\}.$$

**3. Krok.** Całka dowolnej funkcji  $X$  opiera się na rozbiciu jej na różnicę funkcji nieujemnych,  $X = X^+ - X^-$ ,

$$\int_{\Omega} X dP = \int_{\Omega} X^+ dP - \int_{\Omega} X^- dP.$$

Oczywiście, taka całka nie musi istnieć. Jeśli istnieje to mówimy, że zmienna losowa jest całkowalna i definiujemy

### Definicja 3.8

Wartość oczekiwana  $X$  dana jest wzorem

$$\mathbb{E}(X) = \int_{\Omega} X dP.$$

### Uwaga 3.3

Odnajdujemy fakt, że całka jest operacją liniową więc wartość oczekiwana jest również liniowa. Oznacza to że udowodnione poprzednio własności wariancji zachodzą również w sytuacji ogólnej bo do dowodów wykorzystywaliśmy tylko liniowość.

### 3.5.4 Obliczanie

Powyższy schemat definicji całki możemy też na przykład zastosować do całkowania funkcji  $g : \mathbb{R} \rightarrow \mathbb{R}$  względem miary  $P_X$  oznaczając całkę przez

$$\int_{\mathbb{R}} g(x) dP_X(x).$$

Mamy związek między różnymi całkami.



### Twierdzenie 3.6

Jeśli  $g : \mathbb{R} \rightarrow \mathbb{R}$  jest całkowalna względem  $P_X$  a  $g \circ X$  jest całkowalna względem  $P$ , to

$$\int_{\Omega} g(X(\omega)) dP(\omega) = \int_{\mathbb{R}} g(x) dP_X(x)$$

i ta wielkość to  $\mathbb{E}(g(X))$ .

### Dowód

Weźmy szczególny przypadek gdzie  $g(x) = \mathbb{I}_B(x)$  dla zbioru borelowskiego  $B$  i sprawdzenie równości jest proste:

$$\begin{aligned} \int_{\Omega} \mathbb{I}_B(X(\omega)) dP(\omega) &= \int_{\Omega} \mathbb{I}_{\{X(\omega) \in B\}}(\omega) dP(\omega) \\ &= P(\{\omega : X(\omega) \in B\}) \quad \text{definicja całki funkcji charakterystycznej} \\ &= P_X(B) \quad \text{definicja miary } P_X \\ &= \int_{\mathbb{R}} \mathbb{I}_B(x) dP_X(x) \quad \text{definicja całki.} \end{aligned}$$

Ta równość rozciąga się na funkcje schodkowe z liniowości, potem na nieujemne  $g$  będące granicami funkcji schodkowych  $s_n \nearrow g$  a każda funkcja mierzalna może być tak zapisana:

$$\begin{aligned} \int_{\mathbb{R}} g(x) dP_X(x) &= \lim \int_{\mathbb{R}} s_n(x) dP_X(x) \\ &= \lim \int_{\Omega} s_n(X(\omega)) dP \\ &= \int_{\Omega} g(X(\omega)) dP \end{aligned}$$

ponieważ  $s_n \circ X \nearrow g \circ X$  i stosuje się twierdzenie o zbieżności monotonicznej.  $\square$

Ważne są szczególne przypadki. Jeśli  $g(x) = x$  to mamy

$$\mathbb{E}(X) = \int_{\mathbb{R}} x dP_X(x)$$

a dla  $g(x) = (x - \mathbb{E}(X))^2$

$$\text{Var}(X) = \int_{\mathbb{R}} (x - \mathbb{E}(X))^2 dP_X(x).$$

Jeśli zmienna losowa ma gęstość to można ją wykorzystać dla obliczenia wartości oczekiwanej.

### Twierdzenie 3.7

Jeśli dla  $g : \mathbb{R} \rightarrow \mathbb{R}$  poniższe całki mają sens to są równe:

$$\int_{\mathbb{R}} g dP_X = \int_{\mathbb{R}} g f_X dm.$$

### Dowód

Bierzemy jak poprzednio  $g = \mathbb{I}_B$  z  $B \subset \mathbb{R}$  i wtedy

$$\text{lewa strona} = P_X(B)$$

$$\text{prawa strona} = \int_B f_X dm$$

są równe z definicji gęstości. Potem korzystamy z liniowości i przejścia granicznego tak jak poprzednio.  $\square$

Czyli, dla  $X$  posiadającej gęstość mamy

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dm(x)$$

oraz

$$\text{Var}(X) = \int_{\mathbb{R}} (x - \mathbb{E}(X))^2 f_X(x) dm(x).$$

Oczywiście w praktyce te całki sprowadzą się do całek Riemanna jeśli gęstości będą ciągle lub kawałkami ciągle.

### Twierdzenie 3.8

Jeśli  $X$  ma rozkład normalny  $N(\mu, \sigma)$  to

$$\mathbb{E}(X) = \mu.$$

### Dowód

Założmy, że  $\mu = 0$ ,  $\sigma = 1$ . Wtedy

$$\mathbb{E}(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}x^2} dx = -\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \Big|_{-\infty}^{+\infty} = 0$$

bo  $(e^{-\frac{1}{2}x^2})' = -xe^{-\frac{1}{2}x^2}$ . Dla dowolnego  $\mu$

$$\begin{aligned}\mathbb{E}(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{1}{2}(x-\mu)^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (y+\mu)e^{-\frac{1}{2}y^2} dy \quad \text{zmiana zmiennej } y = x - \mu \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ye^{-\frac{1}{2}y^2} dy + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu e^{-\frac{1}{2}y^2} dy = \mu\end{aligned}$$

Udowodniliśmy tezę dla  $\sigma = 1$  i teraz bierzemy dowolne  $\sigma > 0$ :

$$\begin{aligned}\mathbb{E}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu)e^{-\frac{1}{2}y^2} dy \quad \text{zmiana zmiennej } y = \frac{1}{\sigma}(x - \mu), \quad dy = \frac{1}{\sigma} dx \\ &= \sigma \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ye^{-\frac{1}{2}y^2} dy + \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy \\ &= \mu\end{aligned}$$

□

### Twierdzenie 3.9

Jeśli  $X$  ma rozkład normalny  $N(\mu, \sigma)$  to

$$\text{Var}(X) = \sigma^2.$$

### Dowód

Niech  $\mu = 0$  co jak wiemy możemy założyć bez szkody dla ogólności. Wariancję możemy obliczyć wykorzystując gęstość

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

Liczymy całkę nieoznaczoną

$$\begin{aligned}\int x^2 e^{-\frac{1}{2\sigma^2}x^2} dx &= \int u(x)v'(x) dx \quad \text{gdzie } u(x) = x, v'(x) = xe^{-\frac{1}{2\sigma^2}x^2} \\ &\quad u'(x) = 1, \quad v(x) = -\sigma^2 e^{-\frac{1}{2\sigma^2}x^2} \\ &= u(x)v(x) - \int u'(x)v(x) dx \quad \text{całkowanie przez części} \\ &= -x\sigma^2 e^{-\frac{1}{2\sigma^2}x^2} + \int \sigma^2 e^{-\frac{1}{2\sigma^2}x^2} dx\end{aligned}$$

Teraz

$$\begin{aligned}\text{Var}(X) &= -x \frac{1}{\sigma\sqrt{2\pi}} \sigma^2 e^{-\frac{1}{2\sigma^2}x^2} \Big|_{-\infty}^{+\infty} + \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma^2 e^{-\frac{1}{2\sigma^2}x^2} dx \\ &= 0 + \sigma^2 \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}x^2} dx \\ &= \sigma^2 \quad \text{bo} \quad \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}x^2} dx = 1.\end{aligned}$$

□

### Twierdzenie 3.10

Jeśli  $X \sim N(0, \sigma)$  to

$$\mathbb{E}(e^X) = e^{\frac{1}{2}\sigma^2}.$$

### Dowód

Dowód opiera się na rachunku całkowym:

$$\begin{aligned}\mathbb{E}(e^X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^x e^{-\frac{1}{2\sigma^2}x^2} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{x - \frac{1}{2\sigma^2}x^2 - \frac{1}{2}\sigma^2} e^{\frac{1}{2}\sigma^2} dx \quad \text{dodając i odejmując } \frac{1}{2}\sigma^2 \\ &= e^{\frac{1}{2}\sigma^2} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{1}{\sigma\sqrt{2}}x - \frac{1}{\sqrt{2}}\sigma\right)^2} dx \quad \text{ze wzoru } (a-b)^2 = a^2 - 2ab + b^2 \\ &= e^{\frac{1}{2}\sigma^2} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{1}{\sigma\sqrt{2}}y\right)^2} dy \quad \text{zmiana zmiennej } \frac{1}{\sigma\sqrt{2}}y = \frac{1}{\sigma\sqrt{2}}x - \frac{1}{\sqrt{2}}\sigma \\ &= e^{\frac{1}{2}\sigma^2}\end{aligned}$$

□

### Twierdzenie 3.11

Jeśli  $X \sim N(\mu, \sigma)$ , to

$$\mathbb{E}(e^X) = e^{\mu + \frac{1}{2}\sigma^2}.$$

### Dowód

Kładziemy  $Y = X - \mu$  i łatwo pokazać, że  $Y \sim N(0, \sigma)$ . Korzystając z poprzedniego twierdzenia mamy

$$\mathbb{E}(e^X) = \mathbb{E}(e^{Y+\mu}) = \mathbb{E}(e^Y e^\mu) = e^\mu \mathbb{E}(e^Y) = e^\mu e^{\frac{1}{2}\sigma^2} = e^{\mu + \frac{1}{2}\sigma^2}.$$

Zakończymy przykładem zmiennej losowej o mieszanym rozkładzie.  $\square$

### Przykład 3.1

Przypuśćmy, że  $P_X = p\delta_{x_0} + (1-p)P_1$  gdzie  $P_1$  ma gęstość  $f_1$ . Wtedy

$$\mathbb{E}(X) = px_0 + (1-p) \int_{\mathbb{R}} xf_1(x)dx$$

To wynika z ogólnej własności całki którą przyjmujemy bez dowodu: całka względem kombinacji liniowej miar to kombinacja liniowa całek.



# 4 Zależności

## 4.1 Niezależność

Do tej pory rozważaliśmy modele pojedynczych walorów. Nie ulega wątpliwości, że jest to niewystarczające. Wprowadzając wiele walorów pojawia się problem opisu ich wzajemnych relacji. Inny aspekt tego problemu to zależność między cenami jednego waloru, ale w różnych chwilach.

Zaczynamy prezentację stosownych narzędzi od prostego przykładu. Zakładamy, że mamy dwie akcje, których ceny w przyszłym momencie mogą przyjąć 5 wartości, czyli zbiór  $\Omega$  jest pięcio-elementowy:  $\Omega = \{\omega_1, \dots, \omega_5\}$ , ceny i ich prawdopodobieństwa są następujące

$$\begin{aligned} p_1 &= 0.1, & S_1^{\omega_1}(T) &= 56, & S_2^{\omega_1}(T) &= 28, \\ p_2 &= 0.2, & S_1^{\omega_2}(T) &= 54, & S_2^{\omega_2}(T) &= 26, \\ p_3 &= 0.4, & S_1^{\omega_3}(T) &= 50, & S_2^{\omega_3}(T) &= 22, \\ p_4 &= 0.2, & S_1^{\omega_4}(T) &= 47, & S_2^{\omega_4}(T) &= 24, \\ p_5 &= 0.1, & S_1^{\omega_5}(T) &= 46, & S_2^{\omega_5}(T) &= 20, \end{aligned}$$

Zakładamy, że ceny wyjściowe to  $S_1(0) = 50$ ,  $S_2(0) = 23$ .

Często mamy do dyspozycji częściową informację, na przykład, przypuśćmy, że cena akcji  $S_2$  wzrosła i oznaczmy to zdarzenie przez  $B$ , czyli  $B = \{\omega_1, \omega_2, \omega_4\}$ ,  $P(B) = 0.5$ . Ponieważ przestrzeń probabilistyczna jest zbudowana na zbiorze wszystkich dopuszczalnych wyników czy scenariuszy, powinniśmy teraz rozważać zbiór  $B$  jako nowe  $\Omega$ . Badając teraz jakieś inne zdarzenie  $A$ , jedyne co odgrywa rolę to część zdarzenia  $A$  która zawiera się w zdarzeniu

$B$  czyli możemy zastąpić  $A$  przez część wspólną  $A \cap B$  aby ocenić prawdopodobieństwo zajścia  $A$ . Liczbę  $P(A \cap B)$  dzielimy jednak przez  $P(B)$  tak aby prawdopodobieństwo zdarzenia  $B$  było jeden, co odpowiada jego interpretacji jako zbioru wszystkich dopuszczalnych scenariuszy. Otrzymana liczba to tzw. prawdopodobieństwo warunkowe zdarzenia  $A$  przy warunku  $B$  którą oznaczamy w następujący sposób

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

Aby to pojęcie miało sens musimy oczywiście założyć, że  $P(B) \neq 0$ .

Jest naturalne powiedzieć, że zdarzenia  $A$  i  $B$  są niezależne jeśli zajście jednego nie ma wpływu na prawdopodobieństwo drugiego, czyli gdy

$$P(A|B) = P(A)$$

lub, symetrycznie, gdy

$$P(B|A) = P(B).$$

To prowadzi do następującego związku, który jest często przyjmowany jako definicja niezależności

$$P(A \cap B) = P(A)P(B).$$

Oczywiście jeśli  $P(A \cap B) = P(A)P(B)$  to  $P(A|B) = P(A)$  i na odwrót.

Na przykład zdarzenia

$A$  – ” $S_1(T)$  przyjmuje wartości ekstremalne”

$B$  – ” $S_2(T) < S_2(0)$ ”

są niezależne, podczas gdy

$A = \{S_1(T) \geq S_1(0)\}$

$B = \{S_2(T) < S_2(0)\}$

nie są niezależne.

Z punktu widzenia zmiennych losowych niezależność powinna oznaczać, że informacja dostarczana przez obserwacje jednej nie ma wpływu na drugą.

#### Definicja 4.1

Zmienne losowe  $X$  i  $Y$  są niezależne gdy zdarzenia

$$A = \{\omega : X(\omega) \in [a, b]\}$$

$$B = \{\omega : Y(\omega) \in [c, d]\}$$

są niezależne dla wszelkich  $a, b, c, d \in \mathbb{R}$ .



Niezależność zdarzeń  $A, B$  pociąga za sobą następujący związek

$$\begin{aligned} P((X, Y) \in [a, b] \times [c, d]) &= P(X \in [a, b] \text{ oraz } Y \in [c, d]) \\ &= P(\{X \in [a, b]\} \cap \{Y \in [c, d]\}) \\ &= P(\{X \in [a, b]\})P(\{Y \in [c, d]\}) \\ &= P_X([a, b])P_Y([c, d]). \end{aligned}$$

W naszym pierwszym przykładzie zmienne losowe  $S_1(T)$  i  $S_2(T)$  nie są niezależne. Nie jest łatwo zmodyfikować przykład aby uzyskać niezależność co wskazuje że jest to bardzo mocna i specjalna własność. Poniższy przykład jest bardzo specjalny i jak zobaczymy później podpada pod pewien uniwersalny schemat gwarantujący niezależność i w istocie jedyny który to daje.

Niech  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ ,  $X = S_1(T)$ ,  $Y = S_2(T)$ , gdzie  $S_1(0) = 50$ ,  $S_2(0) = 23$ ,

$$\begin{aligned} p_1 &= 0.25, & S_1^{\omega_1}(T) &= 56, & S_2^{\omega_1}(T) &= 22, \\ p_2 &= 0.25, & S_1^{\omega_2}(T) &= 56, & S_2^{\omega_2}(T) &= 26, \\ p_3 &= 0.25, & S_1^{\omega_3}(T) &= 47, & S_2^{\omega_3}(T) &= 22, \\ p_4 &= 0.25, & S_1^{\omega_4}(T) &= 47, & S_2^{\omega_4}(T) &= 26, \end{aligned}$$

Sprawdzamy warunek dla zbiorów takich jak w definicji, ale widać że wystarczy się ograniczyć do zbiorów jedno lub dwuelementowych opartych na powyższych wartościach. Dla dwuelementowych zbiorów  $\{47, 56\}$  czy  $\{22, 26\}$  warunek jest oczywisty bo ich prawdopodobieństwa wynoszą jeden. Weźmy  $A = \{S_1 = 56\}$ ,  $P(A) = 0.5$ ,  $B = \{S_2 = 26\}$ ,  $P(B) = 0.5$  a wtedy  $A \cap B = \{\omega_2\}$  i  $P(A \cap B) = 0.25 = P(A)P(B)$ . Można sprawdzić wszystkie warianty i okaże się że warunek niezależności jest zawsze spełniony.

Wprowadzimy teraz precyzyjnie mechanizm losowych zmian cen akcji, który intuicyjnie przedstawiliśmy na początku, jako motywację pierwszych modeli. Mianowicie chodzi o zmiany cen w kolejnych krokach, addytywne, lub multiplikatywne, ale zakładamy że zmiany te będą niezależne.

Jeśli wyobrazić sobie, że te zmiany są wyznaczone kolejnymi rzutami monety (ograniczmy się na razie do dwóch kroków) intuicyjnie jest jasne że możemy rzucić jednocześnie dwie symetryczne monety i że wyniki tych rzutów są niezależne. Tą intuicję postaramy się sformalizować i natępnie sprawdzić, że formalny warunek niezależności jak zdefiniowany powyżej, jest spełniony.

Dwie monety to dwie przestrzenie probabilistyczne  $\Omega_1 = \{U_1, D_1\}$ ,  $\Omega_2 = \{U_2, D_2\}$  z prawdopodobieństwami  $P_1, P_2$  wyznaczonymi przez dodatnie liczby

$p_1, q_1, p_2, q_2, p_1 + q_1 = 1, p_2 + q_2 = 1$ , wszystkie równe 0,5 dla monet symetrycznych. Wprowadzamy zmienne losowe opisujące zmiany cen

$$\begin{aligned} X_1 &: \Omega_1 \rightarrow \mathbb{R} \\ X_2 &: \Omega_2 \rightarrow \mathbb{R} \end{aligned}$$

$$\begin{aligned} X_1^{U_1} &= 5\%, & X_1^{D_1} &= -5\%, \\ X_2^{U_2} &= 5\%, & X_2^{D_2} &= -5\%. \end{aligned}$$

Cena po dwóch krokach zależy od wyników obu rzutów a z drugiej strony powinna być zmienną losową określoną na pojedynczej przestrzeni. W tym celu kładziemy

$$\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}.$$

Najlepiej zapisać też zbiór w formie macierzy gdzie dla uproszczenia opuszczamy indeksy rozróżniające elementy przestrzeni co jest usprawiedliwione tym, że ta przynależność jest wyznaczona przez kolejność występowania symboli

$$\Omega_1 \times \Omega_2 = \left\{ \begin{array}{cc} UU & UD \\ DU & DD \end{array} \right\}$$

gdzie pierwsze elementy odpowiadają ruchom wyznaczonym przez pierwszą monetę a drugie elementy przez drugą

$$\begin{aligned} Y_1^{UU} &= 5\%, & Y_1^{UD} &= 5\%, & Y_1^{DU} &= -5\%, & Y_1^{DD} &= -5\%, \\ Y_2^{UU} &= 5\%, & Y_2^{UD} &= -5\%, & Y_2^{DU} &= 5\%, & Y_2^{DD} &= -5\%. \end{aligned}$$

Potrzebujemy teraz prawdopodobieństwa na podzbiorach zbioru  $\Omega_1 \times \Omega_2$  a w tym celu wystarczy podać jego wartości na zbiorach jednoelementowych. Dla niezależności, mając na względzie definicję, kładziemy

$$(P_1 \times P_2)(\{(\omega_1, \omega_2)\}) = P_1(\{\omega_1\})P_2(\{\omega_2\}).$$

Oczywiste jest, że rozkłady  $X_1$  i  $Y_1$  jak i rozkłady  $X_2$  i  $Y_2$  są takie same.

Wybór  $\Omega_1 = \Omega_2 = [0, 1]$ , z  $P_1$  i  $P_2$  będącymi miarą Lebesgue'a na zbiorach borelowskich daje możliwość budowy bardziej wyrafinowanych modeli (na przykład w postaci rozkładu normalnego, ale również pozwala modelować rzuty monet).

Naturalne jest zdefiniować zbiór  $\Omega = \Omega_1 \times \Omega_2$  jak poprzednio, a tym razem możemy go zwizualizować jako kwadrat o bokach  $[0, 1]$ , a prawdopodobieństwo na prostokątach zdefiniować tak

$$(P_1 \times P_2)([a, b] \times [c, d]) = P_1([a, b])P_2([c, d]).$$

Teraz wprowadzamy sigma ciało na  $\Omega$  biorąc najmniejsze sigma ciało zawierające wszystkie prostokąty. Możemy przyjąć bez dowodu, że można na tym sigma ciełe zdefiniować miarę która na prostokątach zgadza się z powyższym wzorem.

Mając dwie zmienne losowe  $X_i : \Omega_i \rightarrow \mathbb{R}$ ,  $i = 1, 2$ , konstruujemy ich wersje  $Y_i$  (tak aby rozkłady  $X_1, X_2$  były takie same jak rozkłady  $Y_1, Y_2$ , odpowiednio) zdefiniowane na wspólnej przestrzeni

$$Y_i : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R},$$

wzorami

$$\begin{aligned} Y_1^{(\omega_1, \omega_2)} &= X_1^{\omega_1}, \\ Y_2^{(\omega_1, \omega_2)} &= X_2^{\omega_2}. \end{aligned}$$

Wygodnie będzie wprowadzić stosowne oznaczenia i najlepiej zrobić to w sytuacji ogólnej. Niech więc  $(\Omega_1, \mathcal{F}_1, P_1)$   $(\Omega_2, \mathcal{F}_2, P_2)$  będą dowolnymi przestrzeniami probabilistycznymi oraz założymy, że

$$\begin{aligned} X_1 &: \Omega_1 \rightarrow \mathbb{R}, \\ X_2 &: \Omega_2 \rightarrow \mathbb{R} \end{aligned}$$

Biorąc

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

jak powyżej oznaczamy przez  $\mathcal{F}_1 \times \mathcal{F}_2$  najmniejsze sigma ciało zawierające prostokąty  $A_1 \times A_2$ ,  $A_1 \in \mathcal{F}_1$ ,  $A_2 \in \mathcal{F}_2$ . Prawdopodobieństwo które oznaczamy  $P_1 \times P_2$  jest zdefiniowane na prostokątach

$$(P_1 \times P_2)(A_1 \times A_2) = P_1(A_1)P_2(A_2)$$

a następnie rozszerzone na  $\mathcal{F}_1 \times \mathcal{F}_2$ . Przyjmujemy bez dowodu

#### Twierdzenie 4.1

$P_1 \times P_2$  można rozszerzyć z prostokątów na  $\mathcal{F}_1 \times \mathcal{F}_2$  tak, że rozszerzenie jest przeliczalnie addytywne.

Musimy sprawdzić, że rozkłady  $Y_i$  i  $X_i$  są takie same. To wymaga odrobiny uwagi, gdyż zmienne te są określone na innych przestrzeniach. Weźmy  $B \subset \mathbb{R}$ ,

$$\begin{aligned} (P_1 \times P_2)(Y_1 \in B) &= (P_1 \times P_2)(\{(\omega_1, \omega_2) : X_1(\omega_1) \in B\}) \\ &= (P_1 \times P_2)(\{\omega_1 : X_1(\omega_1) \in B\} \times \Omega_2) \\ &= P_1(X_1 \in B). \end{aligned}$$

### Twierdzenie 4.2

Zmienne losowe  $Y_1, Y_2$  zdefiniowane tak jak powyżej na  $\Omega = \Omega_1 \times \Omega_2$  wyposażonym w sigma ciało  $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$  oraz prawdopodobieństwo  $P = P_1 \times P_2$  są niezależne.

#### Dowód

Weźmy  $A = \{\omega : Y_1(\omega) \in [a, b]\}$ ,  $B = \{\omega : Y_2(\omega) \in [c, d]\}$ , a wynik dostajemy po teorio-mnogościowych łatwych przekształceniach

$$\begin{aligned}
 P(A \cap B) &= P(Y_1 \in [a, b] \text{ oraz } Y_2 \in [c, d]) \\
 &= P((Y_1, Y_2) \in [a, b] \times [c, d]) \\
 &= P((X_1 \in [a, b]) \times (X_2 \in [c, d])) \\
 &= P_1(X_1 \in [a, b])P_2(X_2 \in [c, d]) \\
 &= P_1(X_1 \in [a, b])P_2(\Omega_2)P_1(\Omega_1)P_2(X_2 \in [c, d]) \\
 &= P((X_1 \in [a, b]) \times \Omega_1)P(\Omega_1 \times (X_2 \in [c, d])) \\
 &= P(Y_1 \in [a, b])P(Y_2 \in [c, d]) \\
 &= P(A)P(B).
 \end{aligned}$$

□

Główny cel tych wszystkich rozważań, to konkluzja że niezależność zmiennych losowych oznacza w istocie fakt, że zmienne te są zdefiniowane na różnych przestrzeniach, co jest bardzo wygodne we wszelkich dowodach i konstrukcjach. Oznacza to bowiem, że jedna zmienna losowa zależy od  $\omega_1$  a druga od  $\omega_2$ , co jak zobaczymy ułatwia na przykład całkowanie.

### Twierdzenie 4.3

Zmienne losowe  $Z_i : \Omega \rightarrow \mathbb{R}$ ,  $i = 1, 2$ , są niezależne wtedy i tylko wtedy gdy istnieją przestrzenie probabilistyczne  $\Omega_i$ , zmienne losowe  $X_i : \Omega_i \rightarrow \mathbb{R}$  takie że  $Z_i$  mają takie same rozkłady jak zmienne  $Y_i$  zdefiniowane powyżej na  $\Omega_1 \times \Omega_2$ .

#### Dowód

Założmy niezależność  $Z_i$  i weźmy  $\Omega_1 = \Omega_2 = \mathbb{R}$  ze zbiorami borelowskimi i prawdopodobieństwami  $P_1 = P_{Z_1}$ ,  $P_2 = P_{Z_2}$ . Zdefiniujmy  $X_i^\omega = \omega$  i mamy

$$P_{X_i}(B) = P_i(X_i \in B) = P_{Z_i}(x \in B) = P_{Z_i}(B)$$

co wystarcza gdyż  $P_{X_i} = P_{Y_i}$ . Odwrotna implikacja jest oczywista na podstawie poprzedniego twierdzenia.  $\square$

W ogólnym przypadku, mając dwie zmienne losowe  $X_1, X_2$  definiujemy miarę na zbiorach borelowskich (najmniejsze sigma ciało zawierające prostokąty zbudowane z przedziałów) w  $\mathbb{R} \times \mathbb{R}$  wzorem

$$P_{(X_1, X_2)}(B) = P((X_1, X_2) \in B), \quad B \subset \mathbb{R} \times \mathbb{R} \text{ Borel},$$

i nazywamy ją rozkładem łącznym pary  $(X_1, X_2)$ . Teoria całki może być odpowiednio rozszerzona i w szczególności mamy

$$\mathbb{E}(g(X_1, X_2)) = \int g(x_1, x_2) dP_{(X_1, X_2)}(x_1, x_2).$$

Zmienne niezależne mają rozkład łączny szczególnej postaci

$$P_{(X_1, X_2)}([a, b] \times [c, d]) = P_{X_1}([a, b])P_{X_2}([c, d]).$$

Mamy teraz jedno z najważniejszych zastosowań powyższych rozważań.

#### Twierdzenie 4.4

Jeśli  $X, Y$  są niezależne to  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$  (pod warunkiem, że wszystkie wartości oczekiwane występujące w tym związku są skończone).

#### Dowód

Wiemy, że możemy założyć że zmienne losowe są określone na  $\Omega = \Omega_1 \times \Omega_2$  oraz  $X$  zależy tylko od  $\omega_1 \in \Omega_1$  a  $Y$  od  $\omega_2 \in \Omega_2$ . Liniowość całki jest wykorzystana poniżej: wielkość nie zawierająca zmiennej całkowania jest względem tego całkowania stałą i może być wyciągnięta przed znak całki:

$$\begin{aligned} \mathbb{E}(XY) &= \int_{\Omega_1} \int_{\Omega_2} X(\omega_1)Y(\omega_2) dP_2(\omega_2) dP_1(\omega_1) = \\ &= \int_{\Omega_1} X(\omega_1) \left( \int_{\Omega_2} Y(\omega_2) dP_2(\omega_2) \right) dP_1(\omega_1) \\ &= \int_{\Omega_2} Y(\omega_2) dP_2(\omega_2) \int_{\Omega_1} X(\omega_1) dP_1(\omega_1) \\ &= \mathbb{E}(X)\mathbb{E}(Y). \end{aligned}$$

$\square$

Dla zmiennej losowej  $X$  definiujemy sigma ciało generowane przez  $X$  wzorem

$$\mathcal{F}_X = \{X^{-1}(B) : B \in \mathcal{B}\}.$$

Łatwo pokazać, że jest to istotnie sigma ciało gdyż zbiory borelowskie tworzą sigma ciało, a przeciwobrazy zbiorów z sigma ciała spełniają własności sigma ciała. Nasz cel to przeformułowania pojęcia niezależności w języku sigma ciał generowanych przez zmienne losowe. Najpierw rozszerzamy pojęcie niezależnych zdarzeń na sigma ciała.

#### Definicja 4.2

Sigma ciała  $\mathcal{F}_1$  i  $\mathcal{F}_2$  są niezależne jeśli dla dowolnych  $A_1 \in \mathcal{F}_1$ ,  $A_2 \in \mathcal{F}_2$ , mamy  $P(A_1 \cap A_2) = P(A_1)P(A_2)$ .

#### Twierdzenie 4.5

Zmienne losowe  $X$  i  $Y$  są niezależne wtedy i tylko wtedy gdy generowane sigma ciała  $\mathcal{F}_X$  i  $\mathcal{F}_Y$  są niezależne.

#### Dowód

To wynika natychmiast z definicji przeciwobrazu  $X^{-1}(B) = \{\omega : X^\omega \in B\}$  i definicji niezależności.  $\square$

#### Definicja 4.3

Mówimy, że funkcja  $f : R \rightarrow R$  jest mierzalna w sensie Borela (lub borelowska) jeśli dla dowolnych  $B \in \mathcal{B}$  mamy  $f^{-1}(B) \in \mathcal{B}$ .

#### Twierdzenie 4.6

Jeśli  $X, Y$  są niezależne to  $f(X)$  i  $g(Y)$  są niezależne dla dowolnych funkcji borelowskich  $f, g : \mathbb{R} \rightarrow \mathbb{R}$ .

#### Dowód

Sigma ciało generowane przez  $f(X)$  jest na ogół mniejsze niż sigma ciało generowane przez  $X$ : jeśli  $A \in \mathcal{F}_{f(X)}$  to

$$A = (f \circ X)^{-1}(B) = X^{-1}(f^{-1}(B)) \in \mathcal{F}_X$$

gdyż  $f^{-1}(B) \in \mathcal{B}$ . Pokazaliśmy że  $\mathcal{F}_{f(X)} \subset \mathcal{F}_X$  i to samo zachodzi dla  $Y$ . Jeśli  $\mathcal{F}_X, \mathcal{F}_Y$  są niezależne to bezpośrednio na podstawie definicji jest to również prawdą dla mniejszych sigma ciał  $\mathcal{F}_{f(X)}, \mathcal{F}_{f(Y)}$ .  $\square$

Powyższe rozważania dotyczące niezależności mogą być rozszerzone na dowolną skończoną liczbę zmiennych losowych.

## 4.2 Zależność

Na ogół jednak zmienne losowe niezależne nie są. Wprowadzimy podstawową miarę ich zależności, mianowicie kowariancję, która dla dwóch zmiennych losowych  $X, Y$  zdefiniowana jest wzorem (zakładamy, że występujące w nim wartości oczekiwane są skończone)

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\ &= \mathbb{E}[XY - \mathbb{E}(X)Y - X\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)] \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \end{aligned}$$

gdzie wykorzystaliśmy liniowość wartości oczekiwanej. Jeśli  $X, Y$  są niezależne, to  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$  i wtedy  $\text{Cov}(X, Y) = 0$ .

Łatwo pokazać podstawowe własności kowariancji wynikające z liniowości wartości oczekiwanej

$$\begin{aligned} \text{Cov}(aX, Y) &= a\text{Cov}(X, Y), \\ \text{Cov}(W + X, Y) &= \text{Cov}(W, Y) + \text{Cov}(X, Y). \end{aligned}$$

Warto odnotować relację (dowód jest prostym ćwiczeniem wykorzystującym wzór  $(a + b)^2 = a^2 + 2ab + b^2$ )

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

Dowód poniższej nierówności jest trochę bardziej skomplikowany

$$|\text{Cov}(X, Y)| \leq \sigma_X \sigma_Y$$

a ta nierówność w szczególności gwarantuje, że współczynnik korelacji zdefiniowany wzorem

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

jest liczbą z przedziału  $[-1, 1]$ . W przypadku niezależności  $\text{Cov}(X, Y) = 0$  więc również  $\rho_{XY} = 0$ .

Rozważmy teraz dowolną skończoną liczbę zmiennych losowych  $X_1, \dots, X_n$  które można traktować jak pojedynczą ale o wartościach w przestrzeni wektorów  $\mathbb{R}^n$  czyli tak zwany wektor losowy:

$$\mathbf{X} : \Omega \rightarrow \mathbb{R}^n, \quad \mathbf{X}^\omega = (X_1^\omega, \dots, X_n^\omega).$$

Jego rozkład to miara na  $\mathbb{R}^n$  dana przez

$$P_{\mathbf{X}}(B) = P(\{\omega : (X_1^\omega, \dots, X_n^\omega) \in B\}).$$

Rozkład wektora wyznacza rozkłady jego składowych

$$P_{X_i}(B) = P_{\mathbf{X}}(\mathbb{R} \times \dots \times \mathbb{R} \times B \times \mathbb{R} \times \dots \times \mathbb{R}), \quad B \subset \mathbb{R}.$$

#### Przykład 4.1

Pokażemy, że różne wektory mogą mieć takie same rozkłady dla składowych. Niech  $n = 2$ , a wektor  $X$  niech przyjmuje wartości  $\{(1, 1), (1, 2), (2, 1), (2, 2)\}$  każdą z prawdopodobieństwem  $\frac{1}{4}$ . Niech wektor  $Y$  przyjmuje wartości  $\{(1, 1), (2, 2)\}$  każdą z prawdopodobieństwem  $\frac{1}{2}$ . Składowe obu wektorów mają takie same rozkłady gdyż przyjmują one wartości 1, 2, każda z prawdopodobieństwem  $\frac{1}{2}$ .

#### Definicja 4.4

Mówimy, że  $X_1, \dots, X_n$  są niezależne jeśli dla każdego  $k \leq n$ , i wszelkich  $B_1, \dots, B_k$  zachodzi

$$\begin{aligned} P((X_1 \in B_1) \cap (X_2 \in B_2) \cap \dots \cap (X_k \in B_k)) \\ = P(X_1 \in B_1)P(X_2 \in B_2) \cdots P(X_k \in B_k). \end{aligned}$$

Ten warunek, który jest naturalnym rozszerzeniem warunku dla dwóch zmiennych losowych, musi być narzucony dla każdego  $k$ . Można pokazać przykłady że zachodzi on dla  $k = 3$ , a nie zachodzi dla  $k = 2$  i na odwrót.

Niezależność  $X_1, \dots, X_n$  oznacza, że rozkład wektora ma specjalną postać

$$P_{\mathbf{X}}([a_1, b_1] \times \dots \times [a_n, b_n]) = P_{X_1}([a_1, b_1]) \cdots P_{X_n}([a_n, b_n]).$$

Wartość oczekiwana wektora to wektor wartości oczekiwanych składowych

$$\mathbb{E}(X) = (\mathbb{E}X_1, \dots, \mathbb{E}X_n)$$

a kowariancje tworzą macierz symetryczną rozmiaru  $n \times n$

$$\begin{aligned} C &= [\text{Cov}(X_i, X_j)]_{i,j=1}^n \\ \text{Cov}(X_i, X_j) &= \text{Cov}(X_j, X_i) \end{aligned}$$



z wariancjami na przekątnej:  $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$ .

Rozkład wektora może mieć gęstość  $f_{\mathbf{X}}$ , co oznacza, że

$$P_{\mathbf{X}}(B) = \int_B f_{\mathbf{X}}(x) dx$$

a gęstość jest funkcją zdefiniowaną  $\mathbb{R}^n$ .

Ograniczmy się do wektorów dwuwymiarowych. Rozkład wektora to miara

$$P_{(X_1, X_2)}(B) = P((X_1, X_2) \in B), \quad B \subset \mathbb{R} \times \mathbb{R}$$

Przypominamy, że

$$\rho_{X_1 X_2} = \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}}$$

więc

$$c_{12} = \rho_{X_1 X_2} \sigma_{X_1} \sigma_{X_2} = \rho \sigma_1 \sigma_2$$

Macierz kowariancji przyjmuje postać

$$C = \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}$$

i łatwo policzyć jej wyznacznik

$$\det C = \sigma_1 \sigma_2 \sqrt{1 - \rho^2}$$

Gęstość wektora to funkcja  $f_{(X_1, X_2)}(x_1, x_2) \geq 0$  taka, że

$$P_{(X_1, X_2)}(B) = \int \int_B f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2$$

Specjalny przypadek to gęstość normalna dana wzorem

$$f(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^2 \det C}} \exp\left\{-\frac{1}{2} \sum_{i,j=1}^2 d_{ij}(x_i - \mu_i)(x_j - \mu_j)\right\}$$

gdzie:  $\mu_1, \mu_2 \in \mathbb{R}$ ,  $C = [c_{ij}]$  jest dowolną macierzą symetryczną, odwracalną i dodatnio określona (co oznacza, że  $\sum_{i,j=1}^2 c_{ij} x_i x_j \geq 0$  dla dowolnych  $x_1, x_2 \in \mathbb{R}$ ),  $\det C$  to wyznacznik macierzy  $C$ , jest on różny od zera, a  $D = [d_{ij}] = C^{-1}$  oznacza macierz odwrotną do  $C$ .

Można pokazać, że dla rozkładu normalnego

$$\mu_1 = \mathbb{E}(X_1), \quad \mu_2 = \mathbb{E}(X_2)$$

$$c_{ij} = \text{Cov}(X_i, X_j).$$

Podstawiając wzór na wyznacznik macierzy  $C$ , przy założeniu, że  $\rho \neq 1$ ,  $\rho \neq -1$  mamy

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left( \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right) \right\}$$

Jeśli założymy, że  $X_1, X_2$  są niezależne to

$$C = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

Rozkład normalny ma wtedy postać

$$\begin{aligned} f_{(X_1, X_2)}(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp \left\{ -\frac{1}{2} \left( \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right) \right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left\{ -\frac{1}{2} \frac{(x_1 - \mu_1)^2}{\sigma_1^2} \right\} \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left\{ -\frac{1}{2} \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right\} \\ &= f_{X_1}(x_1)f_{X_2}(x_2) \end{aligned}$$

Okazuje się, że ten przykład ilustruje ogólną własność.

#### Twierdzenie 4.7

Jeśli  $f_{(X_1, X_2)}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$  to  $X_1, X_2$  są niezależne.

#### Dowód

Sprawdzamy warunek definicji niezależności

$$\begin{aligned} &P_{(X_1, X_2)}([a, b] \times [c, d]) \\ &= P((X_1, X_2) \in [a, b] \times [c, d]) \\ &= \int \int_{[a, b] \times [c, d]} f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 \\ &= \int_{[a, b]} \int_{[c, d]} f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 \quad \text{korzystamy z postaci gęstości} \\ &= \int_{[a, b]} f_{X_2}(x_2) \left( \int_{[c, d]} f_{X_1}(x_1) dx_1 \right) dx_2 \quad \text{liniowość całki} \\ &= \int_{[a, b]} f_{X_1}(x_1) dx_1 \int_{[c, d]} f_{X_2}(x_2) dx_2 \quad \text{liniowość całki} \\ &= P(X_1 \in [a, b]) P(X_2 \in [c, d]) \\ &= P_{X_1}([a, b]) P_{X_2}([c, d]) \end{aligned}$$

W trzecim kroku funkcję nie zawierającą zmiennej całkowania możemy wyciągnąć przed znak całki. To dotyczy całki po  $x_1$  a tą funkcją jest  $f_{X_2}(x_2)$ . Korzystamy z liniowości kolejny raz wyciągając stałą  $\int_{[c,d]} f_{X_1}(x_1)dx_1$  przed całkę względem  $x_2$ . Definicja gęstości kończy dowód.  $\square$

Ogólnie, dla przypadku  $n$  wymiarowego mamy analogiczne fakty.

Gęstość normalna ma postać

$$f(x_1, \dots, x_n) = \frac{1}{\sqrt{2\pi \det C}} \exp\left\{-\frac{1}{2} \sum d_{ij}(x_i - \mu_i)(x_j - \mu_j)\right\}$$

gdzie  $(\mu_1, \dots, \mu_n)$  to wektor liczb rzeczywistych, a  $D = [d_{ij}] = C^{-1}$  jest macierzą odwrotną do pewnej macierzy  $C = [c_{ij}]$ , o której zakładamy, że jest symetryczna ( $c_{ij} = c_{ji}$ ), odwracalna i dodatnio określona co oznacza, że  $\sum_{i,j=1}^n c_{ij}x_i x_j \geq 0$  dla dowolnych  $(x_1, \dots, x_n) \in \mathbb{R}^n$ .

Wtedy składowe mają jednowymiarowe rozkłady normalne.

Podobne rachunki jak dla jednego wymiaru pokazują, że

$$\mathbb{E}(\mathbf{X}) = (\mu_1, \dots, \mu_n)$$

$$\text{Cov}(X_i, X_j) = c_{ij}.$$

Dla niezależności składowych wektora o rozkładzie normalnym wystarcza że elementy macierzy kowariancji zerują się poza przekątną. Poprzedzimy to twierdzeniem będącym ogólną wersją twierdzenia pokazanego dla wymiaru 2.

#### Twierdzenie 4.8

Jeśli  $f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = f_{X_1}(x_1) \times \dots \times f_{X_n}(x_n)$  to  $X_1, \dots, X_n$  są niezależne.

#### Corollary 4.9

Jeśli wektor  $(X_1, \dots, X_n)$  ma rozkład normalny  $\text{Cov}(X_i, X_j) = 0$  dla  $i \neq j$  to  $X_1, \dots, X_n$  są niezależne.

#### Dowód

Macierz kowariancji ma zera poza przekątną i to samo zachodzi dla macierzy odwrotnej  $d_{ij} = 0$   $i \neq j$ . Na przekątnej elementy macierzy odwrotnej to odwrotności elementów macierzy  $C$  to znaczy  $d_{ii} = \frac{1}{c_{ii}} = \frac{1}{\sigma_i^2}$ . Wyznacznik takiej

macierzy to iloczyn elementów na przekątnej. Te fakty z algebry pozwalają zauważyć, że gęstość wielowymiarowa jest iloczynem gęstości jednowymiarowych

$$\begin{aligned} f_X(x_1, \dots, x_n) &= \frac{1}{\sqrt{2\pi\sigma_1^2 \cdots \sigma_n^2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} (x_i - \mu_i)^2\right\} \\ &= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2\sigma_1^2} (x_1 - \mu_1)^2\right\} \cdots \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left\{-\frac{1}{2\sigma_n^2} (x_n - \mu_n)^2\right\}. \end{aligned}$$

Z poprzedniego twierdzenia otrzymujemy żądany wniosek.  $\square$

# 5

## Warunkowa wartość oczekiwana

### 5.1 Model dwumianowy

Model dwumianowy to jeden z najważniejszych modeli w matematyce finansowej i pojawia się w wielu kursach. Tu skoncentrujemy się na starannej konstrukcji naturalnej przestrzeni probabilistycznej co pozwoli nam zdefiniować na przykładzie tego modelu ważne i trudne pojęcie, mianowicie warunkową wartość oczekiwaną. Ten model to idealny przykład gdyż pozwala zrozumieć to pojęcie w najprostszym możliwym przypadku, a jednocześnie pozwala zilustrować różne subtelności z tym związane.

Definiujemy (przypominamy) konstrukcję pierwszego kroku:  $S(0)$  jest dane, a przyszła cena to zmienna losowa

$$S(T) : \Omega \rightarrow [0, +\infty)$$

przyjmująca tylko dwie wartości

$$S(T) = \begin{cases} S(0)(1 + U) & \text{z prawdopodobieństwem } p, \\ S(0)(1 + D) & \text{z prawdopodobieństwem } 1 - p, \end{cases}$$

gdzie  $-1 < D < U$ ,  $p \in (0, 1)$ . Wygodnie jest wprowadzić pomocniczą zmienną losową reprezentującą stopę zwrotu

$$K = \begin{cases} U & \text{z prawdopodobieństwem } p, \\ D & \text{z prawdopodobieństwem } 1 - p, \end{cases}$$

gdyż wtedy możemy krótko zapisać

$$S(T) = S(0)(1 + K).$$

Dla jednego kroku najlepszy wybór zbioru  $\Omega$  to  $\Omega = \{U, D\}$  wyposażony w prawdopodobieństwo  $P$  wyznaczone przez liczbę  $p$ :  $P(\{U\}) = p$ ,  $P(\{D\}) = 1 - p$ , i wtedy  $K(\omega) = \omega$ .

Model dla wielu kroków konstruujemy zakładając, że zwroty w poszczególnych krokach są zmiennymi losowymi niezależnymi o takim samym rozkładzie.

Dla dwóch kroków wprowadzamy chwile czasowe:  $0, T, 2T$  i upraszczające oznaczenia. Piszemy  $S(T) = S(1)$ ,  $S(2T) = S(2)$  zaznaczając tylko numer kroku i pomijając jego długość. Dla niezależności bierzemy dwie przestrzenie  $\Omega_1 = \{U_1, D_1\}$ ,  $\Omega_2 = \{U_2, D_2\}$  i kładziemy

$$\Omega = \Omega_1 \times \Omega_2 = \{(U_1, U_2), (U_1, D_2), (D_1, U_2), (D_1, D_2)\}$$

pisząc w skrócie

$$\Omega = \{UU, UD, DU, DD\}.$$

Miara jest dana wzorami

$$P(UU) = p^2, \quad P(UD) = P(DU) = p(1 - p), \quad P(DD) = (1 - p)^2$$

a zwroty

$$\begin{aligned} K_1^{UU} &= K_1^{UD} = U, & K_1^{DU} &= K_1^{DD} = D, \\ K_2^{UU} &= K_2^{DU} = U, & K_2^{UD} &= K_2^{DD} = D. \end{aligned}$$

To daje postać ceny

$$S(2) = S(0)(1 + K_1)(1 + K_2)$$

Dla trzech kroków mamy chwile czasowe  $0, T, 2T, 3T$  i ceny  $S(0), S(1), S(2), S(3)$ . Zwroty oznaczamy

$$K_n = \begin{cases} U & \text{z prawdopodobieństwem } p, \\ D & \text{z prawdopodobieństwem } 1 - p, \end{cases}$$

$n = 1, 2, 3$ . Biorąc  $\Omega_n = \{U, D\}$  definiujemy  $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$  która zawiera osiem elementów, które będziemy nazywać ścieżkami

$$\Omega = \{UUU, UUD, UDU, UDD, DUU, DUD, DDU, DDD\}.$$

Zgodnie z postulowaną niezależnością wprowadzamy prawdopodobieństwo zgodnie ze schematem: prawdopodobieństwo iloczynu zdarzeń to iloczyn prawdopodobieństw, czyli

$$\begin{aligned} P(UUU) &= p^3, \\ P(UDU) &= P(UUD) = P(DUU) = p^2(1 - p), \\ P(UDD) &= P(DUD) = P(DDU) = p(1 - p)^2, \\ P(DDD) &= (1 - p)^3 \end{aligned}$$

Wartość zmiennej losowej  $K_n$  jest  $n$ -tym elementem ciągu  $\omega \in \Omega$  więc na przykład  $K_2^{UDU} = D$ . Ceny zbudowane są w oparciu o schemat jednego kroku

$$S(n+1) = S(n)(1 + K_n)$$

dla  $n = 0, 1, 2$ . Dokładniej,

$$S(1) = S(0)(1 + K_1)$$

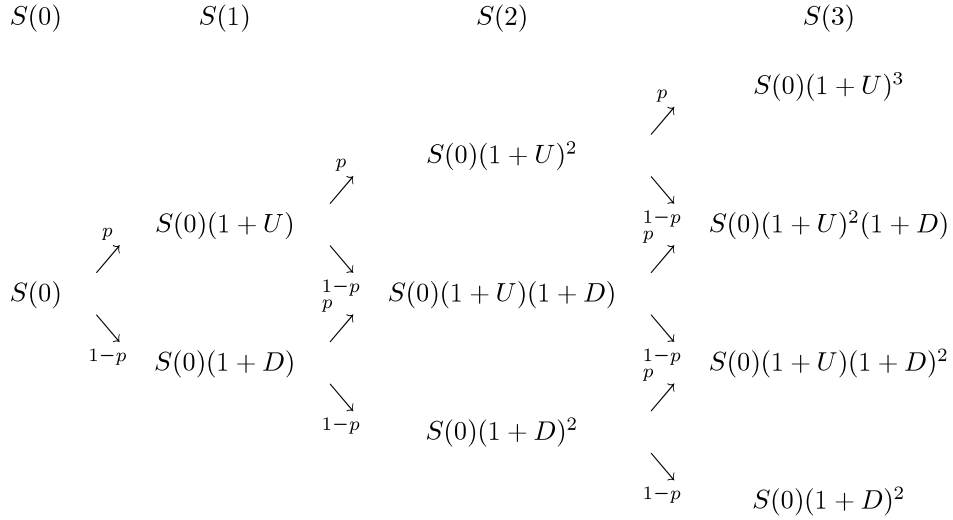
$$S(2) = S(1)(1 + K_2) = S(0)(1 + K_1)(1 + K_2)$$

$$S(3) = S(2)(1 + K_3) = S(0)(1 + K_1)(1 + K_2)(1 + K_3)$$

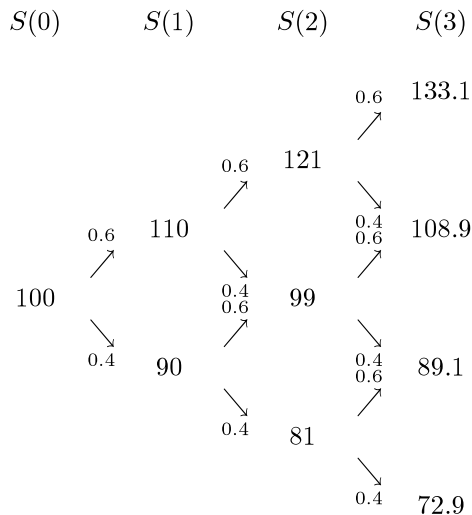
a rozwijając to dla ścieżek

$$S(3) = S(0) \times \begin{cases} (1+U) \times \begin{cases} (1+U) \times \begin{cases} (1+U) \\ (1+D) \end{cases} \\ (1+D) \times \begin{cases} (1+U) \\ (1+D) \end{cases} \end{cases} \\ (1+D) \times \begin{cases} (1+U) \times \begin{cases} (1+U) \\ (1+D) \end{cases} \\ (1+D) \times \begin{cases} (1+U) \\ (1+D) \end{cases} \end{cases} \end{cases}$$

Mamy mniej wartości niż ścieżek. Na przykład są tylko 3 wartości w chwili 2 (a 4 ścieżki) oraz 3 wartości w chwili 3 (8 ścieżek). Wyłania się drzewo postaci:



Dla  $S(0) = 100$ ,  $U = 0.1$ ,  $D = -0.1$ ,  $p = 0.6$ , mamy przykład numeryczny



Liczba  $X$  ruchów w górę to zmienna losowa o następującym rozkładzie

$$\begin{aligned}
 P(X = 3) &= p^3, \\
 P(X = 2) &= 3p^2(1 - p), \\
 P(X = 1) &= 3p(1 - p)^2, \\
 P(X = 0) &= (1 - p)^3.
 \end{aligned}$$

Ogólnie, jeśli mamy  $n$  kroków to wartościami  $X$  są liczby  $k = 0, 1, 2, \dots, n$  z prawdopodobieństwami

$$P(X = k) = \frac{n!}{k!(n - k)!} p^k (1 - p)^{n - k}$$

i wtedy tradycyjnie mówimy że rozkład jest dwumianowy, stąd nazwa modelu.

Obliczmy teraz wartość oczekiwaną ceny w chwili 3

$$\begin{aligned}
 \mathbb{E}(S(3)) &= 131.1 \times P(\{UUU\}) + 108.9 \times P(\{UUD, UDU, DUU\}) \\
 &\quad + 89.1 \times P(\{UDD, DUD, DDU\}) + 72.9 \times P(\{DDD\}) \\
 &= 131.1 \times 0.6^3 + 108.9 \times 3 \times 0.6^2 \times 0.4 \\
 &\quad + 89.1 \times 3 \times 0.6 \times 0.4^2 + 72.9 \times 0.4^3 \\
 &\approx 106.12.
 \end{aligned}$$

Będziemy się przyglądać cenie w chwili 3 z punktu widzenia informacji gromadzonej w poszczególnych krokach.



## 5.2 Podziały i informacja

### 1. Sytuacja po pierwszym kroku.

**1-U.** Przypuśćmy, że cena poszła do góry w pierwszym kroku. To oznacza, że dopuszczalne są tylko ścieżki zaczynające się od  $U$  i oznaczymy ich zbiór natępująco

$$B_U = \{UUU, UUD, UDU, UDD\}.$$

Ten zbiór gra teraz rolę przestrzeni probabilistycznej. Wyposażamy go w prawdopodobieństwo: dla  $A \subseteq B_U$  kładziemy

$$P_U(A) = \frac{P(A)}{P(B_U)}.$$

Oczywiście  $P_U(B_U) = 1$ . Ponieważ  $A \subseteq B_U$ ,

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

więc prawdopodobieństwo  $P_U$  jest niczym innym jak prawdopodobieństwem warunkowym  $P(\cdot|B_U)$  rozważanym dla podzbiorów zdarzenia  $B_U$ .

Teraz obliczamy wartość oczekiwaną zmiennej losowej  $S(3)$  w tej nowej przestrzeni. Przyjmuje ona trzy wartości i mamy

$$131.1 \times 0.6^2 + 108.9 \times 2 \times 0.6 \times 0.4 + 89.1 \times 0.4^2 \approx 114.44.$$

Wprowadzamy dwa alternatywne oznaczenia dla tej wartości oczekiwanej: pierwsze jest zgodne z oznaczeniem prawdopodobieństwa warunkowego i precyzuje wyraźnie warunek, a drugie jest bardziej intuicyjne a zdarzenie które jest warunkiem jest zapisane skrótowo

$$\mathbb{E}(S(3)|B_U) = \mathbb{E}(S(3)|S(1) = 110) \approx 114.44.$$

**1-D.** Przypuśćmy, że cena w pierwszym kroku spadła. Wprowadzamy podobnie zbiór ścieżek odrzucając ścieżki wykluczone, czyli zaczynające się od  $U$ , a pozostawiając ścieżki które wciąż wchodzą w rachubę

$$B_D = \{DUU, DUD, DDU, DDD\}$$

i dla  $A \subseteq B_D$  rozważamy prawdopodobieństwo warunkowe

$$P_D(A) = P(A|B_D).$$

Łatwo policzyć, że wartość oczekiwana ceny  $S(3)$  w tej przestrzeni wynosi

$$\mathbb{E}(S(3)|B_D) = \mathbb{E}(S(3)|S(1) = 90) \approx 93.64.$$

Dwa rozpatrywane przypadki rozbijają nam zbiór  $\Omega$  na dwa rozłączne podzbiory. Mamy do czynienia z tak zwanym podziałem (lub partycją) zbioru na składowe  $\Omega = B_U \cup B_D$ , co motywuje następującą definicję

### Definicja 5.1

Rodzina zdarzeń  $\{B_i\}$  jest podziałem zbioru  $\Omega$  jeśli  $B_i \cap B_j = \emptyset$  dla  $i \neq j$  oraz

$$\Omega = \bigcup B_i.$$

Związek pojęcia podziału z informacją jest w tym kontekście następujący. Jesteśmy umiejscowieni dziś, czyli nie mamy żadnej informacji o przyszłości. Rozważając podział  $\Omega = B_U \cup B_D$  wprowadzamy częściową informację ale prowadzimy analizę typu: "co by było gdyby" rozpatrując przypadki typu: wydarzyło się jedno ze zdarzeń danego podziału. Konsekwencją tego, że dziś nic nie wiemy jest konieczność rozpatrzenia wszystkich przypadków. Informacja jest częściowa gdyż znając  $S(1)$  nie znamy  $\omega \in \Omega$  ale wiemy do którego zbioru podziału należy  $\omega$ : do  $B_U$  czy do  $B_D$

Powyższe dwa przypadki możemy zebrać razem pisząc

$$\mathbb{E}(S(3)|S(1)) = \begin{cases} 114.44 & \text{jeśli } S(1) = 110, \\ 93.64 & \text{jeśli } S(1) = 90. \end{cases}$$

Zdefiniowaliśmy w ten sposób nową zmienną losową przyjmującą dwie wartości. Na każdym z dwóch elementów podziału funkcja ta jest stała. Zapiszmy to jeszcze nieco inaczej

$$\mathbb{E}(S(3)|S(1))^\omega = \begin{cases} 114.44 & \text{jeśli } \omega \in B_U, \\ 93.64 & \text{jeśli } \omega \in B_D. \end{cases}$$

na ogół używamy pojedynczych liter na oznaczenie zmiennych losowych. Tu mamy  $X = \mathbb{E}(S(3)|S(1))$  czyli zapis skomplikowany, ale z kolei zawarte są w nim wszystkie istotne szczegóły.

### 2. Sytuacja po drugim kroku.

Przypuśćmy, że znamy ruchy cen w pierwszych dwóch krokach. Są cztery przypadki, które mogą być zakodowane w podziale zbioru  $\Omega$  na cztery składowe:

$$\begin{aligned} B_{UU} &= \{UUU, UUD\}, \\ B_{UD} &= \{UDU, UDD\}, \\ B_{DU} &= \{DUU, DUD\}, \\ B_{DD} &= \{DDU, DDD\}. \end{aligned}$$

Te zbiory stają się przestrzeniami probabilistycznymi gdy wyposażymy je w prawdopodobieństwa,

$$P_{UU}(A) = P(A|B_{UU})$$

$$P_{UD}(A) = P(A|B_{UD})$$

$$P_{DU}(A) = P(A|B_{DU})$$

$$P_{DD}(A) = P(A|B_{DD})$$

dla stosownych  $A$ .

Tak jak poprzednio, obliczamy wartość oczekiwaną ceny  $S(3)$  w każdej z tych przestrzeni dostając następujące wartości

$$\mathbb{E}(S(3)|B_{UU}) = 122.22,$$

$$\mathbb{E}(S(3)|B_{UD}) = 100.98,$$

$$\mathbb{E}(S(3)|B_{DU}) = 100.98,$$

$$\mathbb{E}(S(3)|B_{DD}) = 82.62.$$

Widzimy, że mamy tylko trzy wartości. W środkowych przypadkach ceny są takie same co możemy skomentować tak: zależą one tylko od wartości akcji w chwili 2 czyli od wartości zmiennej losowej  $S(2)$  a nie od ścieżki na której ta wartość została osiągnięta, co odpowiada faktowi  $S^{UD}(2) = S^{DU}(2)$ . Aby to odzwierciedlić stosujemy oznaczenie  $\mathbb{E}(S(3)|S(2))$  pisząc

$$\mathbb{E}(S(3)|S(2)) = \begin{cases} 123.42, & \text{gdy } S(2) = 121, \\ 100.98, & \text{gdy } S(2) = 99, \\ 82.62, & \text{gdy } S(2) = 81. \end{cases}$$

Zauważmy, że podział generowany przez wartości ceny  $S(2)$  jest inny niż podział generowany przez kombinacje dwóch pierwszych ruchów. Podział związany z  $S(2)$  składa się z tylko trzech zbiorów:  $B_{UU}$ ,  $B_{UD} \cup B_{DU}$ ,  $B_{DD}$ . Fakt że  $\mathbb{E}(S(3)|B_{UD}) = \mathbb{E}(S(3)|B_{DU})$  daje równoważność obu metod: oparcie się na ścieżkach lub wartościach. Jest to specyfika drzewa dwumianowego o jednakowych zwrotach i tak nie musi być ogólnym przypadkiem co zobaczymy później.

Odnajdujemy jeszcze ważny fakt: wartości ceny  $S(2)$  są nieistotne gdyż nie występują w obliczeniach wartości oczekiwanej. Jedyne co odgrywa rolę to do którego elementu podziału należy rozważana ścieżka  $\omega$  co oddaje inny zapis

$$\mathbb{E}(S(3)|S(2))^\omega = \begin{cases} 123.42, & \text{gdy } \omega = UU, \\ 100.98, & \text{gdy } \omega \in \{UD, DU\}, \\ 82.62, & \text{gdy } \omega = DD. \end{cases}$$

**2. Sytuacja po trzecim kroku.** Znamy wszystkie ceny  $S(1), S(2), S(3)$ , i znamy odpowiadające im  $\omega$ , Podział zbioru  $\Omega$  jest na zbiory jednoelementowe.

Jeśli przestrzeń probabilistyczna jest jednoelementowa,  $\Omega = \{\omega\}$ , to dla  $A \subset \Omega$ , albo  $A = \emptyset$  albo  $A = \Omega$ , więc  $P(A|\{\omega\}) = 0$  lub  $P(A|\{\omega\}) = 1$ . Oznacza to, że

$$\mathbb{E}(S(3)|\{\omega\}) = S(3)^\omega$$

czyli w innym zapisie

$$\mathbb{E}(S(3)|S(3)) = S(3).$$

**Podsumowanie:** Zbierając razem powyższe wyniki mamy

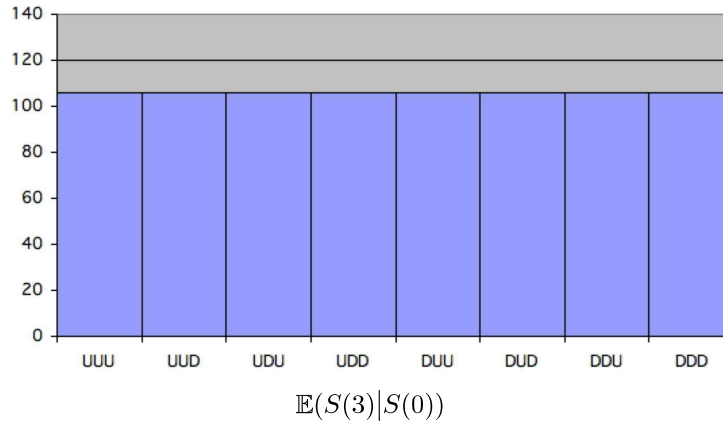
$$\mathbb{E}(S(3)|S(0))^\omega = \mathbb{E}(S(3)) \text{ dla każdego } \omega \in \Omega$$

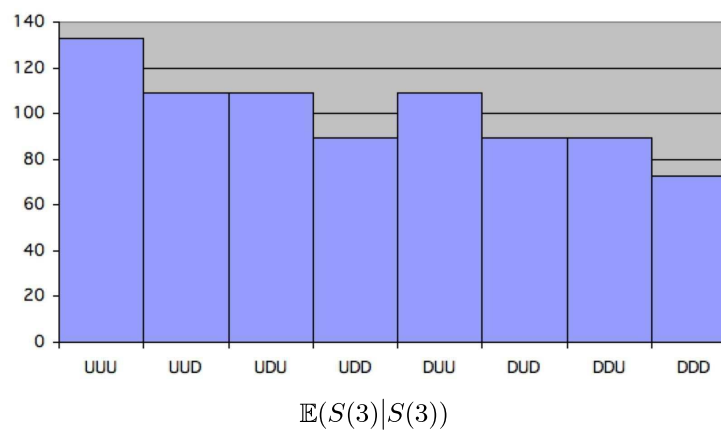
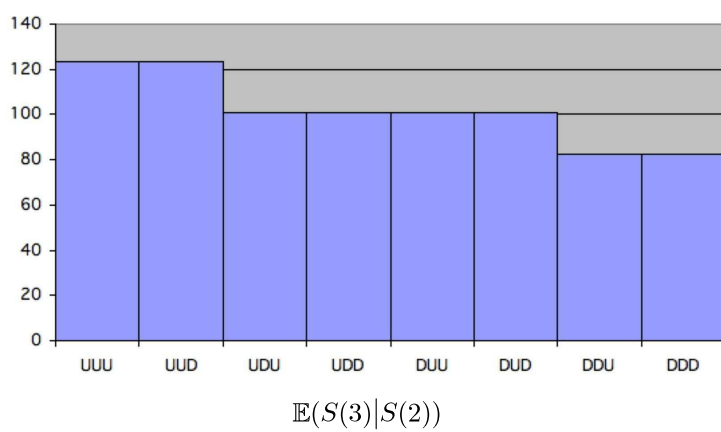
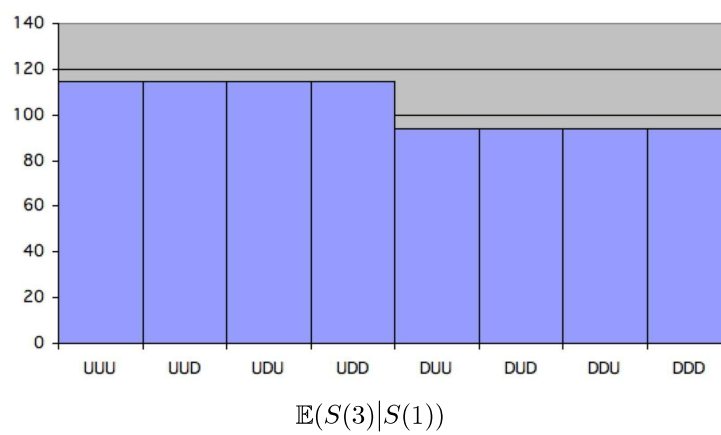
$$\mathbb{E}(S(3)|S(1))^\omega = \begin{cases} 114.44 & \text{jeśli } \omega \in \{UUU, UUD, UDU, UDD\}, \\ 93.64 & \text{jeśli } \omega \in \{UUU, UUD, UDU, UDD\}. \end{cases}$$

$$\mathbb{E}(S(3)|S(2))^\omega = \begin{cases} 123.42, & \text{gdzie } \omega \in \{UUU, UUD\}, \\ 100.98, & \text{gdzie } \omega \in \{UDU, UDD, DUU, DUD\}, \\ 82.62, & \text{gdzie } \omega \in \{DDU, DDD\}. \end{cases}$$

$$\mathbb{E}(S(3)|S(3))^\omega = \begin{cases} 133.1, & \text{gdzie } \omega \in \{UUU\}, \\ 108.9, & \text{gdzie } \omega \in \{UUU, UDU, DUU\}, \\ 89.1, & \text{gdzie } \omega \in \{UDD, DUD, DDU\}, \\ 72.9 & \text{gdzie } \omega \in \{DDD\}. \end{cases}$$

Ewolucja naszej wiedzy na temat  $S(3)$  jest zilustrowana na poniższych rysunkach przedstawiających kolejno wykresy zmiennych losowych  $\mathbb{E}(S(3)|S(0))$ ,  $\mathbb{E}(S(3)|S(1))$ ,  $\mathbb{E}(S(3)|S(2))$ ,  $\mathbb{E}(S(3)|S(3))$ .





### 5.3 Warunkowa wartość oczekiwana – definicja

Powyższe rozważania motywują następującą definicję

#### Definicja 5.2

Dla dowolnej zmiennej losowej  $X$  i zdarzenia  $B$  takiego, że  $P(B) \neq 0$  warunkowa wartość oczekiwana zmiennej  $X$  pod warunkiem  $B$  jest zdefiniowana wzorem

$$\mathbb{E}(X|B) = \sum_{\omega \in \Omega} X^\omega P(\{\omega\}|B).$$

Niech  $\mathbb{I}_B$  oznacza funkcję charakterystyczną zbioru  $B$  ( $\mathbb{I}_B^\omega = 1$  gdy  $\omega \in B$  a zero w przeciwnym przypadku). Mamy

$$\begin{aligned} \sum_{\omega \in \Omega} X^\omega P(\{\omega\}|B) &= \sum_{\omega \in \Omega} X^\omega \frac{P(\{\omega\} \cap B)}{P(B)} \quad \text{definicja } P(A|B) \\ &= \frac{1}{P(B)} \sum_{\omega \in B} X^\omega P(\{\omega\}) \quad \text{gdy } \omega \notin B, \text{ to } \{\omega\} \cap B = \emptyset \\ &= \frac{1}{P(B)} \sum_{\omega \in \Omega} X^\omega \mathbb{I}_B^\omega P(\{\omega\}) \\ &= \frac{1}{P(B)} \mathbb{E}(\mathbb{I}_B X). \end{aligned}$$

Wprowadziliśmy pod koniec funkcję  $\mathbb{I}_B$  aby mieć sumę po wszystkich  $\omega$  w związku z czym mamy elegancki wzór

$$\mathbb{E}(X|B) = \frac{1}{P(B)} \mathbb{E}(\mathbb{I}_B X) \in \mathbb{R}. \quad (5.1)$$

#### Definicja 5.3

Dla danego podziału  $P = \{B_i\}$  zbioru  $\Omega$  powyższa definicja zastosowana do każdego  $B_i$  osobno daje funkcję zdefiniowaną na całym  $\Omega$ , stałą na zbiorach  $B_i$  i oznaczamy ją następująco

$$\mathbb{E}(X|\mathcal{P}) : \Omega \rightarrow \mathbb{R}$$

a jej wartości to

$$B_i \ni \omega \rightarrow \mathbb{E}(X|B_i).$$

Nazywamy ją warunkową wartością oczekiwaną zmiennej  $X$  względem podziału  $P = \{B_i\}$ .

Stosując wzór (5.1) dla  $\omega \in B_i$  dostajemy

$$\mathbb{E}(X|\mathcal{P})^\omega = \frac{1}{P(B_i)}\mathbb{E}(\mathbb{I}_{B_i}X),$$

i sklejając te wzory w jeden mamy

$$\mathbb{E}(X|\mathcal{P}) = \sum_i \mathbb{I}_{B_i} \frac{1}{P(B_i)}\mathbb{E}(\mathbb{I}_{B_i}X). \quad (5.2)$$

Obie strony to zmienne losowe. Z prawej strony losowość tkwi tylko w jednym miejscu, mianowicie w funkcjach charakterystycznych zbiorów  $B_i$

$$\mathbb{E}(X|\mathcal{P})^\omega = \sum_i \mathbb{I}_{B_i}^\omega \frac{1}{P(B_i)}\mathbb{E}(\mathbb{I}_{B_i}X).$$

Powyższe własności motywują następującą definicję.

#### Definicja 5.4

Mówimy, że zmienna losowa  $X$  jest mierzalna względem podziału  $P$  jeśli jest stała na jego elementach.

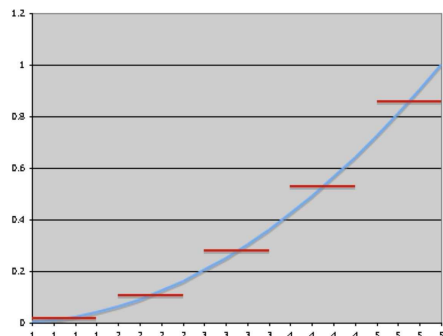
Definicja warunkowej wartości oczekiwanej gwarantuje, że zmienna losowa  $\mathbb{E}(X|\mathcal{P})$  jest mierzalna względem  $\mathcal{P}$ .

#### Przykład 5.1

Niech  $\Omega = [0, 1)$ ,  $X^\omega = \omega^2$ ,  $B_n = [\frac{n-1}{5}, \frac{n}{5})$ ,  $n = 1, 2, 3, 4, 5$ . Dla  $\omega \in [\frac{n-1}{5}, \frac{n}{5})$  mamy

$$\mathbb{E}(X|\mathcal{P})^\omega = 5 \int_{\frac{n-1}{5}}^{\frac{n}{5}} x^2 dx$$

a wykres otrzymanej zmiennej losowej jest pokazany na poniższym rysunku, gdzie widać, że zrealizowana jest idea przybliżenia zmiennej  $X$  funkcjami mierzalnymi względem danego podziału.



Jako szczególny przypadek powyższej definicji otrzymujemy definicję, gdzie warunkiem jest inna zmienna losowa (czyli przypadki to zdarzenia, że przyjmuje pewne wartości).

$Y$  zmienna losowa o rozkładzie dyskretnym,  $\{y_n\}$  wartości  $Y$

### Definicja 5.5

Jeśli  $X$  jest dowolną zmienną losową, a  $Y$  ma rozkład dyskretny z wartościami  $y_1, y_2, \dots$ , to warunkowa wartość oczekiwana  $E(X|Y)$  jest zdefiniowana jako warunkowa wartość oczekiwana  $X$  względem podziału  $P_Y = \{B_n\}$  gdzie  $B_n = \{\omega : Y^\omega = y_n\}$ , czyli

$$\mathbb{E}(X|Y) = \mathbb{E}(X|\mathcal{P}_Y).$$

Mówimy, że podział  $B_n = \{\omega : Y^\omega = y_n\}$  jest generowany przez zmienną losową  $Y$ . Dla  $\omega \in B_n$ ,  $Y^\omega = y_n$  więc

$$\begin{aligned} \mathbb{E}(X|Y)^\omega &= \mathbb{E}(X|Y = y_n) \quad \text{dla } \omega \text{ takich, że } Y^\omega = y_n \\ &= \frac{1}{P(Y = y_n)} \mathbb{E}(\mathbb{I}_{\{Y=y_n\}} X). \end{aligned}$$

### Przykład 5.2

Niech  $\Omega = [0, 1]$ ,  $X^\omega = \omega^2$ ,  $Y^\omega = [5\omega]$  gdzie  $[x]$  to część całkowita liczby  $x$  czyli

$$\begin{aligned} Y^\omega &= 0 \text{ dla } \omega \in [0, \frac{1}{5}), \\ Y^\omega &= 1 \text{ dla } \omega \in [\frac{1}{5}, \frac{2}{5}) \text{ itd.} \end{aligned}$$

Jako warunkową wartość oczekiwana dostajemy zmienną losową znaną z poprzedniego przykładu co wynika z tego że  $X$  jest taka sama oraz podział gene-



rowany przez  $Y$  pokrywa się z rozpatrywanym powyżej:

$$\mathbb{E}(X|Y)^\omega = 5 \int_{\frac{n-1}{5}}^{\frac{n}{5}} x^2 dx \text{ dla } \omega \in [\frac{n-1}{5}, \frac{n}{5}).$$

Zauważmy, że jeśli  $Y$  jest stałe na jakimś zbiorze to  $\mathbb{E}(X|Y)$  jest również stałe na tym samym zbiorze. Wartości zmiennej losowej  $\mathbb{E}(X|Y)$  zależą tylko od zdarzeń będących warunkami a nie od konkretnych wartości przyjmowanych przez  $Y$ . Jako konsekwencja, mamy następujący fakt: jeśli dwie zmienne losowe  $Y$  i  $W$  definiują ten sam podział to warunkowe wartości oczekiwane zmiennej  $X$  są takie same

$$\mathbb{E}(X|Y) = \mathbb{E}(X|W).$$

Jeśli  $V$  i  $Z$  wyznaczają różne podziały to oczywiście na ogół

$$\mathbb{E}(X|V) \neq \mathbb{E}(X|Z).$$

Dla jasności dobrze jest prześledzić najprostszy możliwy przykład który to ilustruje na drzewach dwumianowych.

### Przykład 5.3

Będziemy rozważać ceny kilku akcji. Ceny pierwszej  $S_1(n)$  są skonstruowane biorąc niezależne zwroty  $K_1 = \pm 20\%$ ,  $K_2 = \pm 15\%$ ,  $K_3 = \pm 10\%$  i jej wartości widzimy poniżej (zakładając  $S(0) = 100$ )

$\omega$	$S_1(1)$	$S_1(2)$	$S_1(3)$
$UUU$	120	138	151.8
$UUD$	120	138	124.2
$UDU$	120	102	112.2
$UDD$	120	102	91.8
$DUU$	80	92	101.2
$DUD$	80	92	82.8
$DDU$	80	68	74.8
$DDD$	80	68	61.2

Teraz konstruujemy akcje  $S_2$ ,  $S_3$  które dostaczą nam warunków do badania oczekiwań względem  $S_1(3)$  więc wystarczy je zbudować dla dwóch kroków. Bierzemy zwroty  $\pm 15\%$ ,  $\pm 10\%$  dla  $S_2$  oraz  $\pm 8\%$ ,  $\pm 12\%$  dla  $S_3$  i kładziemy  $X = S_1(3)$ .  $Y = S_2(2)$ ,  $W = S_3(2)$ . Niech prawdopodobieństwo ruchu w górę wynosi  $p = 0.6$  w każdym kroku i obliczmy warunkowe wartości oczekiwane  $E(X|Y)$ ,  $E(X|W)$ . One się pokrywają mimo że wartości  $Y$  i  $W$  są inne, ale podziały przez nie wyznaczone są te same, mianowicie jest to podział

$\{UUU, UUD\}, \{UDU, UDD\}, \{DUU, DUD\}, \{DDU, DDD\}$ .

$\omega$	$P$	$X$	$Y$	$\mathbb{E}(X Y)$	$W$	$\mathbb{E}(X W)$
$UUU$	0.216	151.8	126.5	140.8	120.96	140.8
$UUD$	0.144	124.2	126.5	140.8	120.96	140.8
$UDU$	0.144	112.2	103.5	104	95.04	104
$UDD$	0.096	91.8	103.5	104	95.04	104
$DUU$	0.144	101.2	93.5	93.84	103.04	93.84
$DUD$	0.096	82.8	93.5	93.84	103.04	93.84
$DDU$	0.096	74.8	76.5	69.36	80.96	69.36
$DDD$	0.064	61.2	76.5	69.36	80.96	69.36

Teraz weźmy  $S_4$  ze zwrotami  $K_1 = 0$ ,  $K_2 = \pm 12\%$  oraz  $S_5$  ze zwrotami  $K_1 = \pm 12\%$ ,  $K_2 = 0$ . Ich wartości w chwili 2 są takie same ale wyznaczają inne podziały, więc otrzymujemy różne warunkowe wartości oczekiwane zmiennej  $X$  względem  $V = S_4(2)$  oraz  $Z = S_5(2)$

$\omega$	$P$	$X$	$V$	$\mathbb{E}(X V)$	$Z$	$\mathbb{E}(X Z)$
$UUU$	0.216	151.8	112	122	112	126.1
$UUD$	0.144	124.2	112	122	112	126.1
$UDU$	0.144	112.2	88	90.17	112	126.1
$UDD$	0.096	91.8	88	90.17	112	126.1
$DUU$	0.144	101.2	112	122	88	84.05
$DUD$	0.096	82.8	112	122	88	84.05
$DDU$	0.096	74.8	88	90.17	88	84.05
$DDD$	0.064	61.2	88	90.17	88	84.05

## 5.4 Własności warunkowej wartości oczekiwanej

Najpierw pokażemy, że dla dowolnych  $X_1, X_2, Y$  i liczb  $a, b \in \mathbb{R}$  mamy liniowość

$$\mathbb{E}(aX_1 + bX_2|Y) = a\mathbb{E}(X_1|Y) + b\mathbb{E}(X_2|Y)$$

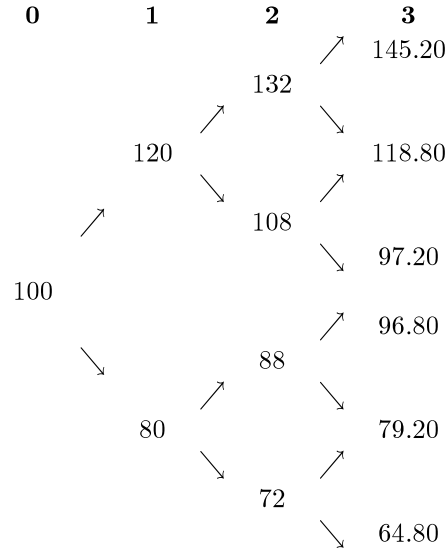
Ustalmy  $\omega$  takie, że  $Y^\omega = y$ . Przypominamy definicję:

$$\mathbb{E}(X|Y) = 1 \frac{1}{P(Y = y_n)} \mathbb{E}(\mathbb{I}_{\{Y=y_n\}} X)$$

Liczymy lewą stronę

$$\begin{aligned}
 \mathbb{E}(aX_1 + bX_2|Y) &= \frac{1}{P(Y = y_n)} \mathbb{E}(\mathbb{I}_{\{Y=y_n\}}(aX_1 + bX_2)) \quad \text{z definicji } \mathbb{E}(X|Y) \\
 &= \frac{1}{P(Y = y_n)} [a\mathbb{E}(\mathbb{I}_{\{Y=y_n\}}X_1) + b\mathbb{E}(\mathbb{I}_{\{Y=y_n\}}X_2)] \\
 &= a \frac{1}{P(Y = y_n)} \mathbb{E}(\mathbb{I}_{\{Y=y_n\}}X_1) + b \frac{1}{P(Y = y_n)} \mathbb{E}(\mathbb{I}_{\{Y=y_n\}}X_2) \\
 &= a\mathbb{E}(X_1|Y) + b\mathbb{E}(X_2|Y).
 \end{aligned}$$

Jako wiodący przykład ilustrujący inne własności weźmy następujące drzewo cen gdzie zwroty w kolejnych krokach są równe  $\pm 20\%$  w pierwszym oraz  $\pm 10\%$  w pozostałych, a prawdopodobieństwa wzrostu ceny są  $p = 0.75$  w każdym przypadku (widzimy, że ceny się nie sklejają w chwilach 2 i 3 co miałyby miejsce gdyby zwroty były takie same):



Warunkowa wartość oczekiwana  $\mathbb{E}(S(3)|S(2))$  przyjmuje 4 wartości a podział generowany przez cenę  $S(2)$  zgadza się z podziałem wyznaczonym przez pierwsze dwa ruchy cen. Konkretnie, mamy

$$\begin{aligned}
 \mathbb{E}(S(3)|B_{UU}) &= 145.20 \times p + 118.80 \times (1 - p) = 138.60, \\
 \mathbb{E}(S(3)|B_{UD}) &= 118.80 \times p + 97.20 \times (1 - p) = 113.40, \\
 \mathbb{E}(S(3)|B_{DU}) &= 96.80 \times p + 79.20 \times (1 - p) = 92.40, \\
 \mathbb{E}(S(3)|B_{DD}) &= 79.20 \times p + 64.80 \times (1 - p) = 75.60.
 \end{aligned}$$

Policzmy wartość oczekiwaną zmiennej losowej  $\mathbb{E}(S(3)|S(2))$  :

$$\begin{aligned}\mathbb{E}(\mathbb{E}(S(3)|S(2))) &= 138.60 \times p^2 + 113.40 \times p(1-p) \\ &\quad + 92.40 \times (1-p)p + 75.40 \times (1-p)^2 \\ &= 121.275\end{aligned}$$

Niezależnie, policzmy wartość oczekiwaną ceny  $S(3)$

$$\begin{aligned}\mathbb{E}(S(3)) &= 145.20 \times p^3 + 118.80 \times 2 \times p^2(1-p) + 97.20 \times p(1-p)^2 \\ &\quad + 96.80 \times (1-p)p^2 + 79.20 \times 2 \times (1-p)^2p + 64.80 \times (1-p)^3 \\ &= 121.275\end{aligned}$$

i widzimy że otrzymaliśmy te same wyniki:  $\mathbb{E}(\mathbb{E}(S(3)|S(2))) = \mathbb{E}(S(3))$ . To jest tylko ilustracja ogólnej własności

### Twierdzenie 5.1

Wartość oczekiwana warunkowej wartości oczekiwanej jest taka sama jak wartość oczekiwana

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X). \quad (5.3)$$

### Dowód

Oznaczamy  $B_n = \{\omega : Y(\omega) = y_n\}$ , i mamy

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X|Y)) &= \sum \mathbb{E}(X|B_n)P(B_n) \text{ (definicja wartości oczekiwanej)} \\ &= \sum_n \frac{1}{P(B_n)} \mathbb{E}(X\mathbb{I}_{B_n})P(B_n) \\ &\quad \text{(definicja warunkowej wartości oczekiwanej)} \\ &= \sum \mathbb{E}(X\mathbb{I}_{B_n}) \\ &= \mathbb{E}(X \sum \mathbb{I}_{B_n}) \text{ (addytywność wartości oczekiwanej)} \\ &= \mathbb{E}(X) \text{ (} B_n \text{ jest podziałem więc } \sum \mathbb{I}_{B_n} = 1\text{)}\end{aligned}$$

co kończy dowód. □

Własność (5.3) wymaga uogólnienia, gdyż wartość oczekiwana nie jest dostatecznie interesującym pojęciem w aspekcie zastosowań. Zaczniemy od przykładu gdzie  $S(1)$  i  $S(2)$  są jak podane powyżej. Podział wyznaczony przez  $S(1)$  to dwa zbiory  $B_U$ ,  $B_D$ , a podział dany przez  $S(2)$  to cztery zbiory  $B_{UU}$ ,  $B_{UD}$ ,  $B_{DU}$ ,  $B_{DD}$ . Odnotujemy, że  $B_U = B_{UU} \cup B_{UD}$ ,  $B_D = B_{DU} \cup B_{DD}$  i w takim

przypadku mówimy, że podział  $\mathcal{P}_2$  jest drobniejszy niż  $\mathcal{P}_1$ . W języku zmiennych losowych oznacza to, że  $S(2)$  niesie więcej informacji niż  $S(1)$ , czyli, że znając wartości  $S(2)$  znamy wartości  $S(1)$ .

### Uwaga 5.1

Tak nie jest przy jednakowych zwrotach gdzie znając  $S(2)$  dla wartości środkowej nie wiemy co się stało w chwili jeden gdyż  $S^{UD}(2) = S^{DU}(2)$  – w obecnym przykładzie te wartości są inne. Wtedy aby informacja w chwili dwa była pełna trzeba zmodyfikować podział w chwili 2 biorąc zbiory na których zarówno  $S(1)$  jak i  $S(2)$  są stałe.

### Definicja 5.6

Mając dwa podziały  $P_1, P_2$ , mówimy że  $P_2$  jest drobniejszy jeśli każdy element  $P_1$  można zbudować z elementów podziału  $P_2$ .

Powyżej obliczyliśmy wartości zmiennej losowej  $\mathbb{E}(S(3)|S(2))$ . Teraz policzymy jej warunkową wartość oczekiwaną względem  $S(1)$ , a wynik porównamy z warunkową wartością oczekiwaną zmiennej  $S(2)$  względem  $S(1)$ .

Przypomnijmy, że

$$\mathbb{E}(S(3)|S(2)) = \begin{cases} 138.60 & \text{gdy } S(2) = 132 \\ 113.40 & \text{gdy } S(2) = 108 \\ 92.40 & \text{gdy } S(2) = 88 \\ 75.60 & \text{gdy } S(2) = 72 \end{cases}$$

Na  $B_U$  mamy

$$\mathbb{E}(\mathbb{E}(S(3)|S(2))|S(1) = 120) = 138.60 \times p + 113.40 \times (1 - p)$$

podczas gdy na  $B_D$

$$\mathbb{E}(\mathbb{E}(S(3)|S(2))|S(1) = 80) = 92.40 \times p + 75.60 \times (1 - p)$$

Teraz liczymy  $\mathbb{E}(S(3)|S(1))$ . Na  $B_U$  mamy

$$\begin{aligned} \mathbb{E}(S(3)|S(1) = 120) &= 145.20 \times p^2 + 118.80 \times 2p(1 - p) + 97.20 \times (1 - p)^2 \\ &= p(145.20 \times p + 118.80 \times (1 - p)) \\ &\quad + (1 - p)(118.80 \times p + 97.20 \times (1 - p)) \\ &= 138.60 \times p + 113.40 \times (1 - p), \end{aligned}$$

a na  $B_D$

$$\begin{aligned}\mathbb{E}(S(3)|S(1) = 80) &= 96.80 \times p^2 + 79.20 \times 2p(1-p) + 64.80 \times (1-p)^2 \\ &= p(96.80 \times p + 79.20 \times (1-p)) \\ &\quad + (1-p)(79.20 \times p + 64.80 \times (1-p)) \\ &= 92.40 \times p + 75.60 \times (1-p).\end{aligned}$$

Widzimy, że

$$\mathbb{E}(\mathbb{E}(S(3)|S(2))|S(1)) = \mathbb{E}(S(3)|S(1))$$

co ilustruje ogólną własność sformułowaną poniżej.

### Twierdzenie 5.2

Niech  $Y, Z$  mają rozkład dyskretny i niech podział wyznaczony przez  $Y$  będzie drobniejszy niż podział wyznaczony przez  $Z$ . Wtedy

$$\mathbb{E}(\mathbb{E}(X|Y)|Z) = \mathbb{E}(X|Z). \quad (5.4)$$

Ta własność nosi nazwę: iterowana wartość oczekiwana.

### Dowód

Idea jest taka jak w poprzedzającym przykładzie. Stosujemy poprzednie twierdzenie do każdego ze ziorów  $\{Z = z_i\}$ . Niech  $C_j = \{Z = z_j\}$ ,  $B_i = \{Y = y_i\}$ , ale ten podział jest drobniejszy, więc  $C_j = \bigcup_{i \in I_j} B_i$ . Dla  $\omega \in C_j$

$$\begin{aligned}\mathbb{E}(\mathbb{E}(X|Y)|Z)^\omega &= \frac{1}{P(C_j)} \mathbb{E}(\mathbb{E}(X|Y) \mathbb{I}_{C_j}) \quad \text{na podstawie (5.1)} \\ &= \frac{1}{P(C_j)} \mathbb{E}\left(\sum_{i \in I_j} \mathbb{I}_{B_i} \frac{1}{P(B_i)} \mathbb{E}(\mathbb{I}_{B_i} X) \mathbb{I}_{C_j}\right) \quad \text{na podstawie (5.2)} \\ &= \frac{1}{P(C_j)} \mathbb{E}\left(\sum_{i \in I_j} \mathbb{I}_{B_i} \frac{1}{P(B_i)} \mathbb{E}(\mathbb{I}_{B_i} X)\right) \quad \text{gdyż } B_i \subset C_j \text{ a więc } \mathbb{I}_{B_i} \mathbb{I}_{C_j} = \mathbb{I}_{B_i} \\ &= \frac{1}{P(C_j)} \sum_{i \in I_j} \frac{1}{P(B_i)} \mathbb{E}(\mathbb{I}_{B_i} X) \mathbb{E}(\mathbb{I}_{B_i}) \quad \text{z liniowości i } \mathbb{E}(\mathbb{I}_{B_i}) = P(B_i) \\ &= \frac{1}{P(C_j)} \sum_i \frac{1}{P(B_i)} \mathbb{E}(\mathbb{I}_{B_i} X) P(B_i) = \frac{1}{P(C_j)} \sum_i \mathbb{E}(\mathbb{I}_{B_i} X) \\ &= \frac{1}{P(C_j)} \mathbb{E}(\mathbb{I}_{C_j} X) \quad \text{gdyż } \mathbb{I}_{C_j} = \sum \mathbb{I}_{B_i} \\ &= \mathbb{E}(X|Z)^\omega\end{aligned}$$

co kończy dowód.  $\square$

Aby zilustrować kolejną własność rozważmy  $\mathbb{E}(S(1)S(3)|S(2))$  i dla przykładu skoncentrujmy się na wyznaczeniu  $\mathbb{E}(S(1)S(3)|S(2) = 132)$ .

Niech  $B = \{S(2) = 132\}$

$$\begin{aligned}\mathbb{E}(S(1)S(3)|B) &= \frac{1}{P(B)}\mathbb{E}(S(1)S(3)\mathbb{I}_B) \\ &= \frac{1}{p^2} (120 \times 145.2 \times p^3 + 120 \times 118.8 \times p^2(1-p)) \\ &= 120 \times \frac{1}{p^2} (145.2 \times p^3 + 118.8 \times p^2(1-p)) \\ &= 120 \times \frac{1}{P(B)}\mathbb{E}(S(3)\mathbb{I}_B) \\ &= S(1) \times \mathbb{E}(S(1)|B).\end{aligned}$$

Rachunek sprowadził się do obliczenia sumy, gdzie zakres sumowania to  $\omega \in B_{UU}$ . Ale dla obu takich ścieżek  $S(1)$  jest stałe. Można je więc wyciągnąć przed sumę i to daje po przeanalizowaniu każdego przypadku możliwość wyciągnięcia  $S(1)$  przed wartość oczekiwaną czyli

$$\mathbb{E}(S(1)S(3)|S(2)) = S(1)\mathbb{E}(S(3)|S(2)).$$

### Twierdzenie 5.3

Założmy, że podział wyznaczony przez  $Y$  jest drobniejszy niż wyznaczony przez  $Z$ . Wtedy

$$\mathbb{E}(ZX|Y) = Z\mathbb{E}(X|Y). \quad (5.5)$$

### Dowód

Ustalmy składową  $B$  podziału wyznaczonego przez  $Y$  i zauważmy, że  $Z$  jest stała na  $B$  przyjmując wartość którą oznaczmy przez  $z$ . Wtedy

$$\begin{aligned}\mathbb{E}(ZX|B) &= \sum_{\omega \in B} Z(\omega)X(\omega)P(\omega|B) \\ &= \sum_{\omega \in B} zX(\omega)P(\omega|B) \\ &= z \sum_{\omega \in B} X(\omega)P(\omega|B) \\ &= z\mathbb{E}(X|B)\end{aligned}$$

czyli mamy tezę na zbiorze  $B$  ale możemy otrzymane wzory skleić co kończy dowód.  $\square$

Intuicja jest następująca: jeśli znamy wartości  $Y$  to również znamy wartości  $Z$ , możemy je traktować jak stałe i wyciągać przed sumę jako wspólny czynnik. Ważny szczególny przypadek to  $X \equiv 1$ , gdzie dostajemy

$$\mathbb{E}(Z|Y) = Z.$$

Ostatnia własność ma związek z niezależnością rozumianą jako fakt, że prawdopodobieństwo warunkowe nie zależy od tego czy warunek zaszedł czy nie czyli  $P(A|B) = P(A)$  dla niezależnych zdarzeń. Mamy następującą wersję tego faktu dla warunkowych wartości oczekiwanych

#### Twierdzenie 5.4

Jeśli  $Y$  a  $Z$  są niezależne to

$$\mathbb{E}(Y|Z) = \mathbb{E}(Y) \quad \text{oraz} \quad \mathbb{E}(Z|Y) = \mathbb{E}(Z).$$

#### Dowód

Weźmy składową  $B$  podziału generowanego przez  $Z$ . Niech  $Y = \mathbb{I}_C$

$$\begin{aligned} \mathbb{E}(Y|Z) &= \frac{1}{P(B)} \mathbb{E}(\mathbb{I}_B Y) && \text{definicja} \\ &= \frac{1}{P(B)} \mathbb{E}(\mathbb{I}_B \mathbb{I}_C) && \text{postać zmiennej } Y \\ &= \frac{1}{P(B)} \mathbb{E}(\mathbb{I}_{B \cap C}) \\ &= \frac{1}{P(B)} P(B \cap C) \\ &= \frac{1}{P(B)} P(B)P(C) && \text{niezależność} \\ &= P(C) = \mathbb{E}(\mathbb{I}_C) \\ &= \mathbb{E}(Y) \end{aligned}$$

z takim samym dowodem dla drugiego wzoru. □

### 5.5 Definicja w przypadku ogólnym

Zacznijmy od obserwacji, że podział może być rozszerzony do sigma ciała poprzez dodanie zbioru pustego i wszelkich możliwych sum składowych podziału

$$\mathcal{F} = \{\emptyset\} \cup \left\{ \bigcup B_j : \{B_1, B_2, \dots\} \subset \mathcal{P} \right\}.$$



Na odwrót, w pewnych sytuacjach, ta procedura może być odwrócona. Możemy to na pewno zrobić dla skończonych zbiorów  $\Omega$ . Mając wtedy sigma ciało (w takim przypadku słowo "sigma" jest przesadą gdyż nie mamy w istocie do czynienia z przeliczalnymi operacjami a tylko skończonymi, więc wystarczyłoby mówić "ciało") możemy zbudować podział biorąc pod uwagę najmniejsze niepuste elementy ciała  $\mathcal{F}$ , tak zwane atomy, co jest intuicyjnym i trafnym terminem. Budując tak podział  $\mathcal{P}$  a następnie rekonstruując z niego ciało tak jak to opisaliśmy powyżej, dostaniemy wyjściową rodzinę  $\mathcal{F}$ .

W przypadku gdy  $\Omega$  jest zbiorem nieskończonym, ta równoważność na ogół nie zachodzi. W przypadku  $\Omega = [0, 1]$  atomy to zbiory jednoelementowe. Nie można zbudować z nich zbiorów borelowskich biorąc przeliczalne sumy. Ponadto nie nadają się one do przeprowadzenia powyższych konstrukcji warunkowych wartości oczekiwanych, gdyż warunek że  $P(B_i) \neq 0$  nie jest spełniony. W takich sytuacjach czeka nas jeszcze pewien wysiłek aby zdefiniować warunkową wartość oczekiwaną i podziały do tego celu się nie przydadzą, czego szkoda, gdyż jest to pojęcie bardziej intuicyjne niż sigma ciało.

Przypominamy, że w przypadku dyskretnym warunkowa wartość oczekiwana zmiennej losowej  $X$  jest zdefiniowana względem podziału zbioru  $\Omega$  na składowe  $\mathcal{P} = \{B_i\}$  jako zmienna losowa stała na każdym  $B = B_i$  o wartości

$$\mathbb{E}(X|\mathcal{P}) = \frac{\mathbb{E}(\mathbb{I}_B X)}{P(B)}.$$

W ogólnej sytuacji, jak już wspomnieliśmy, ta konstrukcja nie może być zastosowana gdyż podział może być nieuchronnie na zbiory o prawdopodobieństwie zero. Szukając odpowiedniej wersji przekształcamy powyższy wzór mnożąc go obustronnie przez  $P(B)$ . Lewa strona to liczba  $P(B)\mathbb{E}(X|\mathcal{P})$  która jest taka sama na każdym  $\omega \in B$  czyli możemy tą wielkość zapisać inaczej, jako całkę z funkcji stałej na  $B$

$$P(B)\mathbb{E}(X|\mathcal{P}) = \int_B \mathbb{E}(X|\mathcal{P}) dP.$$

Dalej, całka po  $B$  może być zapisana jako całka po całym  $\Omega$  po wprowadzeniu funkcji charakterystycznej zbioru  $B$

$$\int_B \mathbb{E}(X|\mathcal{P}) dP = \int_{\Omega} \mathbb{I}_B \mathbb{E}(X|\mathcal{P}) dP$$

a to jest po prostu wartością oczekiwaną zmiennej losowej  $\mathbb{I}_B \mathbb{E}(X|\mathcal{P})$ . Otrzymujemy więc związek

$$\mathbb{E}(\mathbb{I}_B \mathbb{E}(X|\mathcal{P})) = \mathbb{E}(\mathbb{I}_B X)$$

dla każdej składowej podziału. Jeśli weźmiemy sigma ciało  $\mathcal{G}$  generowane przez podział to ta równość zachodzi dla każdego  $B \in \mathcal{G}$  gdyż generowanie polega na

sumowaniu mnogościowym składowych i wzór zachowuje się dzięki liniowości wartości oczekiwanej. Przejście od podziałów do sigma ciał jest kluczowe, gdyż w ogólnym przypadku powyższy warunek ma sens i może być przyjęty jako definicja.

### Definicja 5.7

Warunkowa wartość oczekiwana całkowalnej zmiennej losowej  $X$  określonej na przestrzeni  $(\Omega, F, P)$ , względem sigma ciała  $G \subset F$  to zmienna losowa oznaczona przez  $E(X|G)$  spełniająca dwa warunki

1.  $E(X|G)$  jest  $G$ -mierzalna (czyli przeciwobrazy zbiorów borelowskich są w  $G$ )
2. dla każdego  $B \in \mathcal{G}$

$$\mathbb{E}(\mathbb{I}_B E(X|G)) = \mathbb{E}(\mathbb{I}_B X).$$

Pierwszy warunek odpowiada temu, że w przypadku dyskretnym zakładaliśmy że warunkowa wartość oczekiwana jest stała na składowych podziału.

Odnotujmy, że dowiedzione powyżej własności zachodzą w przypadku ogólnym a wyrafinowanych czytelników zapraszamy do analizy ich dowodów

## 5.6 Własności\*

Ten fragment jest zaawansowany. Dowiedzimy teraz własności znanych z przypadku dyskretnego dla warunkowej wartości oczekiwanej zdefiniowanej ogólnie. Ciekawe jest to że dowody w przypadku ogólnym są krótsze i prostsze niż w przypadku szczególnym, ale wymagają one przygotowania matematycznego

Przypomnijmy, że  $E(X|G)$  jest  $G$ -mierzalna i dla każdego  $B \in \mathcal{G}$

$$\mathbb{E}(\mathbb{I}_B E(X|G)) = \mathbb{E}(\mathbb{I}_B X).$$

Specjalny przypadek zmiennej  $X$  całkowalnej z kwadratem zasługuje na uwagę, gdyż definicja nabiera specjalnego znaczenia. Bierzemy  $Y$  mierzalną względem sigma ciała  $\mathcal{G}$  i przybliżamy ją funkcjami schodkowymi  $s = \sum s_i \mathbb{I}_{B_i}$ ,  $B_i \in \mathcal{G}$  a powyższa równość daje

$$\mathbb{E}(s E(X|G)) = \mathbb{E}(sX)$$

co na podstawie stosownego twierdzenia o zbieżności pociąga

$$\mathbb{E}(Y E(X|G)) = \mathbb{E}(YX).$$

Przerzucając wszystko na lewo i korzystając z liniowości wartości oczekiwanej mamy

$$\mathbb{E}(Y\mathbb{E}(X|\mathcal{G})) - \mathbb{E}(YX) = \mathbb{E}(Y[\mathbb{E}(X|\mathcal{G}) - X]) = 0.$$

Wyrażenie  $\mathbb{E}(YZ)$  definiuje iloczyn skalarny w przestrzeni zmiennych losowych całkowalnych z kwadratem więc w tym sensie  $Y$  są  $\mathbb{E}(X|\mathcal{G}) - X$  prostopadłe. Oznaczając

$$\mathcal{M} = \{Y : Y \text{ jest } \mathcal{G}\text{-mierzalna}\}$$

widzimy, że  $\mathbb{E}(X|\mathcal{G})$  jest rzutem ortogonalnym  $X$  na  $\mathcal{M}$ . Innymi słowy, wśród  $\mathcal{G}$ -mierzalnych zmiennych losowych, zmienna  $\mathbb{E}(X|\mathcal{G})$  jest najbliższa do  $X$  gdzie odległość między  $X$  i  $Z$  jest wyznaczona przez nasz iloczyn skalarny, czy jest dana przez  $\mathbb{E}((X - Z)^2)$ .

Przechodzimy do dowodów własności.

### Twierdzenie 5.5

Jeśli  $\mathcal{G} \subseteq \mathcal{H}$  to  $\mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})$ ,

#### Dowód

Weźmy  $A \in \mathcal{G}$ . Musimy pokazać, że

$$\mathbb{E}(\mathbb{I}_A \mathbb{E}(X|\mathcal{H})) = \mathbb{E}(\mathbb{I}_A \mathbb{E}(X|\mathcal{G}))$$

Z definicji warunkowej wartości oczekiwanej względem  $\mathcal{H}$  ( $A \in \mathcal{H}$  gdyż  $\mathcal{G} \subset \mathcal{H}$ ) mamy

$$\text{lewa strona} = \mathbb{E}(\mathbb{I}_A \mathbb{E}(X|\mathcal{H})) = \mathbb{E}(\mathbb{I}_A X).$$

Z definicji warunkowej wartości oczekiwanej względem  $\mathcal{G}$

$$\text{prawa strona} = \mathbb{E}(\mathbb{I}_A \mathbb{E}(X|\mathcal{G})) = \mathbb{E}(\mathbb{I}_A X)$$

co kończy dowód. □

### Corollary 5.6

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X).$$

#### Dowód

$\mathbb{I}_\Omega = 1$ ,  $\Omega \in \mathcal{G}$  więc stosując definicję warunkowej wartości oczekiwanej względem  $\mathcal{G}$  dostajemy

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(\mathbb{I}_\Omega \mathbb{E}(X|\mathcal{G})) = \mathbb{E}(\mathbb{I}_\Omega X) = \mathbb{E}(X).$$

□

**Twierdzenie 5.7**

Jeśli  $Z$  jest niezależna od  $\mathcal{G}$ , wtedy  $\mathbb{E}(Z|\mathcal{G}) = \mathbb{E}(Z)$ ,

**Dowód**

Weźmy  $A \in \mathcal{G}$ . Mamy pokazać, że

$$\mathbb{E}(\mathbb{I}_A Z) = \mathbb{E}(\mathbb{I}_A \mathbb{E}(Z)).$$

Niech  $Z = \mathbb{I}_B$ ,  $B \in \mathcal{G}$ . Zauważmy, że  $\mathbb{E}(\mathbb{I}_C) = P(C)$ . Wtedy

$$\begin{aligned} \text{lewa strona} &= \mathbb{E}(\mathbb{I}_A \mathbb{I}_B) = \mathbb{E}(\mathbb{I}_{A \cap B}) = P(A \cap B) \\ \text{prawa strona} &= \mathbb{E}(\mathbb{I}_A \mathbb{E}(\mathbb{I}_B)) = \mathbb{E}(\mathbb{I}_A P(B)) = P(B) \mathbb{E}(\mathbb{I}_A) \\ &= P(A)P(B) \quad \text{z liniowości,} \end{aligned}$$

więc założona niezależność dowodzi tezy w tym przypadku. Dalej z liniowości dostajemy ją dla funkcji prostych. Dla dowolnej  $Z$  bierzemy ciąg  $Z_n \nearrow Z$ , funkcji prostych  $Z_n$  i przejście graniczne kończy dowód.  $\square$

**Twierdzenie 5.8**

Jeśli  $Z$  jest  $\mathcal{G}$ -mierzalna, to  $\mathbb{E}(ZX|\mathcal{G}) = Z\mathbb{E}(X|\mathcal{G})$ ,

**Dowód**

Weźmy  $A \in \mathcal{G}$ . Mamy pokazać, że

$$\mathbb{E}(\mathbb{I}_A ZX) = \mathbb{E}(\mathbb{I}_A Z\mathbb{E}(X|\mathcal{G}))$$

Niech  $Z = \mathbb{I}_B$ ,  $B \in \mathcal{G}$

$$\begin{aligned} \text{lewa strona} &= \mathbb{E}(\mathbb{I}_A \mathbb{I}_B X) = \mathbb{E}(\mathbb{I}_{A \cap B} X) \\ \text{prawa strona} &= \mathbb{E}(\mathbb{I}_A \mathbb{I}_B \mathbb{E}(X|\mathcal{G})) = \mathbb{E}(\mathbb{I}_{A \cap B} \mathbb{E}(X|\mathcal{G})) \\ &= \mathbb{E}(\mathbb{I}_{A \cap B} X) \quad , \end{aligned}$$

gdzie w ostatnim przejściu skorzystaliśmy z definicji warunkowej wartości oczekiwanej co jest usprawiedliwione gdyż  $A \cap B \in \mathcal{G}$ . Dalej z liniowości dostajemy rezultat dla funkcji prostych i przejście graniczne kończy dowód.  $\square$

**Corollary 5.9**

Jeśli  $Z$  jest  $\mathcal{G}$ -mierzalna to  $\mathbb{E}(Z|\mathcal{G}) = Z$ .

### Dowód

Weźmy  $X = 1$  w poprzednim twierdzeniu. □

Nowy fakt to przydatna nierówność która jest uogólnieniem znanego faktu, że kwadrat całki to mniej niż całka z kwadratu.

### Twierdzenie 5.10

Jesli  $X$  jest całkowalna z kwadratem to mamy nierówność Jensena  $[\mathbb{E}(X|\mathcal{G})]^2 \leq \mathbb{E}(X^2|\mathcal{G})$ .

### Dowód

Wykorzystujemy pewien fakt z analizy:

$$x^2 = \sup_{a,b} \{ax + b : ax + b \leq x^2\}.$$

Weźmy dowolne  $a, b$  takie, że  $ax + b \leq x^2$  i z liniowości mamy

$$a\mathbb{E}(X|\mathcal{G}) + b = \mathbb{E}(aX + b|\mathcal{G}) \leq \mathbb{E}(X^2|\mathcal{G}).$$

Bierzemy supremum z lewej strony. Nierówność jest zachowana a po lewej stronie dostajemy  $[\mathbb{E}(X|\mathcal{G})]^2$ . □



## Część II

### Procesy stochastyczne





# 6

## Martyngały

### 6.1 Filtracja

Zacznijmy od przykładu opartym na modelu dwumianowym z trzema krokami o jednakowych zwrotach  $K_1 = \pm 10\%$ ,  $K_2 = \pm 10\%$ ,  $K_3 = \pm 10\%$ . Naturalna przestrzeń probabilistyczna to

$$\Omega = \{UUU, UUD, UDU, UDD, DUU, DUD, DDU, DDD\}.$$

i w niej wprowadzamy podziały wyznaczone przez kolejne zmienne losowe  $S(n)$

$$\begin{aligned}\mathcal{P}_{S(1)} &= \{\{UUU, UUD, UDU, UDD\}, \\ &\quad \{DUU, DUD, DDU, DDD\}\} \\ \mathcal{P}_{S(2)} &= \{\{UUU, UUD\}, \\ &\quad \{UDU, UDD, DUU, DUD\}, \\ &\quad \{DDU, DDD\}\} \\ \mathcal{P}_{S(3)} &= \{\{UUU\}, \\ &\quad \{UUD, UDU, DUU\}, \\ &\quad \{UDD, DUD, DDU\}, \\ &\quad \{DDD\}\}\end{aligned}$$

Ilustrację graficzną przedstawiamy na Rysunku 6.1. Widzimy pewną niezgodność gdyż jeśli podział reprezentuje informację, to informacja w chwili 2 powinna obejmować wiedzę o obu chwilach przeszłych, 1 i 2, a nie tylko 2. W

$\omega$	S(1)	S(2)	S(3)
UUU	110	121	133.1
UUD	110	121	108.9
UDU	110	99	108.9
UDD	110	99	89.1
DUU	90	99	108.9
DUD	90	99	89.1
DDU	90	81	89.1
DDD	90	81	72.9

Rysunek 6.1

związku z tym zbiory z  $\mathcal{P}_1$  dzielimy zgodnie z  $S(2)$

$$\begin{aligned}
 \{UUU, UUD, UDU, UDD\} &= \{\omega : S(2) = 121\} \cup \{\omega : S(2) = 99\} \\
 &= \{UUU, UUD\} \cup \{UDU, UDD\} \\
 \{DUU, DUD, DDU, DDD\} &= \{\omega : S(2) = 99\} \cup \{\omega : S(2) = 81\} \\
 &= \{DUU, DUD\} \cup \{DDU, DDD\}
 \end{aligned}$$

Dostaliśmy podział drobniejszy niż  $\mathcal{P}_1$

$$\begin{aligned}
 \mathcal{P}_2 = \{ & \{UUU, UUD\}, \\
 & \{UDU, UDD\}, \\
 & \{DUU, DUD\}, \\
 & \{DDU, DDD\} \} = \{B_{UU}, B_{UD}, B_{DU}, B_{DD}\}
 \end{aligned}$$

Podobnie postępujemy w następnym kroku

$$\mathcal{P}_3 = \{\{UUU\}, \{UUD\}, \{UDU\}, \{UDD\}, \{DUU\}, \{DUD\}, \{DDU\}, \{DDD\}\}$$

Otrzymujemy w ten sposób ciąg podziałów gdzie następny jest drobniejszy niż poprzedni i nazywamy go filtracją

### Definicja 6.1

Filtracja to ciąg podziałów  $P_k$  zbioru  $\Omega$  taki, że  $P_{k+1}$  jest drobniejszy niż  $P_k$ .

Ogólnie, niech  $X$  będzie zmienną losową przyjmującą wartości  $\{x_1, \dots, x_k\}$  a podział generowany przez  $X$  to rodzina zbiorów  $\mathcal{P}_X = \{B_i\}_{i=1}^k$  danych przez

$$B_i = \{\omega : X^\omega = x_i\}.$$

Idea polega na rozłożeniu zbioru  $\Omega$  na fragmenty na których  $X$  jest stała. Weźmy teraz drugą zmienną losową  $Y$  przyjmującą wartości  $\{y_1, \dots, y_l\}$ . Każdy

element  $B_i$  podziału  $\mathcal{P}_X$  rozkładamy na fragmenty na których stała jest  $Y$

$$B_i = \bigcup_{j=1}^l B_{ij}, \quad B_{ij} = \{\omega \in B_i : Y^\omega = y_j\}.$$

Otrzymany podział oznaczamy  $\mathcal{P}_{XY} = \{B_{ij}\}$  i jest on oczywiście drobniejszy niż  $\mathcal{P}_X$ . Możemy tak postępować dalej, zagęszczając podziały w oparciu o kolejne zmienne losowe.

### Definicja 6.2

Niech  $X_n$  będzie ciągiem zmiennych losowych a podziały wyznaczone przez  $X_1, \dots, X_k$  oznaczamy przez  $\mathcal{P}_k$  i nazywamy je filtracją generowaną przez ciąg  $X_n$

### Przykład 6.1

Rozważmy przykład drzewa dwumianowego o identycznych zwrotach w 4 krokach. Zbiór ścieżek ma postać

$$\begin{aligned} \Omega = \{ & UUUU, UUUD, UUDU, UUDD \\ & UDUU, UDUD, UDDU, UDDD \\ & DUUU, DUUD, DUDU, DUDD \\ & DDUU, DDUD, DDDU, DDDD \} \end{aligned}$$

a filtracja generowana przez  $S(0), (1), S(2), S(3), S(4)$  jest postaci

$$\begin{aligned}
 \mathcal{P}_0 &= \{\Omega\} \\
 \mathcal{P}_1 &= \{\{UUUU, UUUD, UUDU, UUDD, UDUU, UDUD, UDDU, UDDD\} \\
 &\quad \{DUUU, DUUD, DUDU, DUDD, DDUU, DDUD, DDDU, DDDD\}\} \\
 &= \{B_U, B_D\} \\
 \mathcal{P}_2 &= \{\{UUUU, UUUD, UUDU, UUDD\}, \\
 &\quad \{UDUU, UDUD, UDDU, UDDD\}, \\
 &\quad \{DUUU, DUUD, DUDU, DUDD\}, \\
 &\quad \{DDUU, DDUD, DDDU, DDDD\}\} \\
 &= \{B_{UU}, B_{UD}, B_{DU}, B_{DD}\}, \\
 \mathcal{P}_3 &= \{\{UUUU, UUUD\}, \\
 &\quad \{UUDU, UUDD\}, \\
 &\quad \{UDUU, UDUD\}, \\
 &\quad \{UDDU, UDDD\}, \\
 &\quad \{DUUU, DUUD\}, \\
 &\quad \{DUDU, DUDD\}, \\
 &\quad \{DDUU, DDUD\}, \\
 &\quad \{DDDU, DDDD\}\} \\
 &= \{B_{UUU}, B_{UUD}, B_{UDU}, B_{UDD}, B_{DUU}, B_{DUD}, B_{DDU}, B_{DDD}\},
 \end{aligned}$$

a ostatni podział  $P_4$  ma 16 jedno-elementowych zbiorów  $\{\omega\}$ .

Przedyskutujmy teraz dokładniej wspomniany już wcześniej związek podziałów i informacji. Powyższy przykład jest dogodnym polem do dyskusji. mamy w nim ograniczenie czasowe do czterech kroków. Jest to umowna granica poza którą nasz świat nie wychodzi. Osiągnąwszy ją, wiemy wszystko gdyż znamy całą historię, czyli wiemy która z 16 dostępnych ścieżek będzie wybrana.

Dziś jednak nie wiemy nic ale krok po kroku, nasza wiedza będzie rosła. Po jednym kroku będziemy w stanie stwierdzić że nieznana ewolucja cen należy do  $B_U$  lub  $B_D$ . Dalsze rozumowanie prowadzimy na zasadzie "co by było gdyby". Rozpatrujemy każdą z tych dwóch możliwości jako hipotetyczną rzeczywistość i patrzymy na dalszą przyszłość z tej perspektywy. Po każdym kroku rośnie liczba tych hipotetycznych możliwości, ale za to każda staje się węższa, mniejszy jest zakres niepewności, a większa nasza wiedza o lokalizacji końcowego wyniku, czyli całej ścieżki. Wszystkie te przyszłe hipotetyczne stany to części składowe odpowiedniego podziału.

Filtracja  $P_0, P_1, P_2, \dots$  reprezentuje więc zmieniającą się strukturę napływających danych. Przeszłość nie jest zapomniana co związane jest z warunkiem że  $P_{n+1}$  jest drobniejszy niż  $P_n$  (czyli,  $P_{n+1}$  zawiera więcej informacji niż  $P_n$ ).

Wracamy do przykładu numerycznego cen akcji w trzech krokach z jednakowymi zwrotami  $\pm 10\%$ . Rysunek 6.2 przedstawia dogęszone podziały będące filtracją generowaną przez  $S(n)$

$\omega$	$S(1)$	$S(2)$	$S(3)$
UUU	110	121	133.1
UUD	110	121	108.9
UDU	110	99	108.9
UDD	110	99	89.1
DUU	90	99	108.9
DUD	90	99	89.1
DDU	90	81	89.1
DDD	90	81	72.9

**Rysunek 6.2**

Rozważmy ciąg zmiennych losowych danych znanym wzorem (wyплаты opcji kupna)

$$H(n) = \max\{S(n) - 90, 0\}, \quad n = 1, 2, 3.$$

Mamy na przykład

$$H(2) = \begin{cases} 31 & \text{na } B_{UU} \\ 9 & \text{na } B_{UD} \cup B_{DU} \\ 0 & \text{na } B_{DD} \end{cases}$$

i widzimy, że zmienna losowa  $H(2)$  jest  $\mathcal{P}_2$  mierzalna, nie jest  $\mathcal{P}_1$  mierzalna. Ogólnie, łatwo sprawdzić, że dla każdego  $n = 1, 2, 3$  zmienna losowa  $H(n)$  jest  $\mathcal{P}_n$  mierzalna. Formułujemy tę własność ogólnie.

### Definicja 6.3

Jeśli  $X(n)$  jest  $P_n$  mierzalna to ciąg  $X(n)$  jest adaptowany.

Wexmy inny przykład, który można interpretować jako dywidendę wypłacaną jako procent ceny z poprzedniego okresu. Jest to typowy przykład bo często dywidendy są ogłaszane i znane z wyprzedzeniem. Niech

$$D(n) = 10\% \times S(n-1), \quad n = 1, 2, 3$$

wtedy na przykład

$$D(2) = \begin{cases} 11 & \text{na } B_{UU} \cup B_{DD} \\ 9 & \text{na } B_{DU} \cup B_{DD} \end{cases}$$

i widać, że  $D(2)$  jest  $P_1$  mierzalna, a ogólnie  $D(n)$  jest  $P_{n-1}$  mierzalna. To jest przykład który występuje w zastosowaniach i zasługuje na nazwę.

### Definicja 6.4

Jeśli  $X(n)$  jest  $\mathcal{P}_{n-1}$  mierzalna to ciąg  $X(n)$  jest przewidywalny.

Inny przykład reprezentuje premię dla doradcy wypłaconą za trafny wybór akcji i zależy ona od liczby wzrostów cen

$$G(2) = \begin{cases} 100 & \text{gdy } \omega = UUU \\ 50 & \text{gdy } \omega \in \{UUD, UDU, DUU\} \\ 0 & \text{w pozostałych przypadkach} \end{cases}$$

W tym przykładzie  $G(2)$  nie jest  $\mathcal{P}_2$  mierzalny - na  $B_{UU}$  przyjmuje on dwie wartości. Ta sytuacja jest nierealistyczna, gdyż nie można wypłacić w chwili 2 kwoty opartej na przyszłych wydarzeniach. Ciąg nie jest adaptowany i to go w pewnym sensie dyskwalifikuje.

## 6.2 Definicja i przykłady

Przypominamy własność: jeśli  $\mathcal{P}_2$  jest drobniejszy niż  $\mathcal{P}_1$  to

$$\mathbb{E}(\mathbb{E}(X|\mathcal{P}_2)|\mathcal{P}_1) = \mathbb{E}(X|\mathcal{P}_1)$$

Niech:

$$Y(1) = \mathbb{E}(X|\mathcal{P}_1)$$

$$Y(2) = \mathbb{E}(X|\mathcal{P}_2)$$

wtedy (wstawiając  $Y$ ) mamy

$$\mathbb{E}(Y(2)|\mathcal{P}_1) = Y(1)$$

Ogólnie, jeśli  $\mathcal{P}_0, \mathcal{P}_1, \mathcal{P}_2, \dots$  jest filtracją a  $X$  zmienną losową to ciąg

$$Y(n) = \mathbb{E}(X|\mathcal{P}_n)$$

spełnia

$$\mathbb{E}(Y(n+1)|\mathcal{P}_n) = Y(n).$$

To motywuje jedno z najważniejszych pojęć.

### Definicja 6.5

Ciąg  $M(n)$  jest martyngałem względem filtracji  $\{\mathcal{P}_n\}$  jeśli

$$\mathbb{E}(M(n+1)|\mathcal{P}_n) = M(n).$$

W języku informacji niesionej przez podziały oznacza to, że patrząc z perspektywy chwili wcześniejszej  $n$  na przyszłe ceny, cena w tej wcześniejszej chwili jest taka jak oczekiwana przyszła cena.

Zilustrujmy to pojęcie rozwijając ideę z początku tego podrozdziału. Niech

$$\begin{aligned} X &= \max(S(3) - 90, 0) \\ S(0) &= 0, \\ K_1 &= \pm 10\%, \quad K_2 = \pm 15\%, \quad K_3 = \pm 20\% \end{aligned}$$

Kładziemy

$$Y(n) = \mathbb{E}(X | \mathcal{P}_n)$$

czyli patrzymy na wypłatę  $X$  z punktu widzenia chwil wcześniejszych.

Zakładamy, że  $p = \frac{1}{2}$ . Rozważmy na przykład:  $\omega \in B_{UU}$ , czyli  $\omega = UUU$  lub  $\omega = UUD$ ,

$$Y(2)^\omega = \mathbb{E}(X | \mathcal{P}_2)^\omega = \frac{1}{2}(61.8 + 11.2) = 36.5$$

Łatwo sprawdzić, że mamy sytuację jak na Rysunku a ponadto, że zachodzą równości (wynikające zresztą z własności warunkowej wartości oczekiwanej)

$$\begin{aligned} Y(0) &= \mathbb{E}(Y(1) | \mathcal{P}_0) \\ Y(1) &= \mathbb{E}(Y(2) | \mathcal{P}_1) \end{aligned}$$

i oznaczając  $X = Y(3)$  mamy

$$Y(n) = \mathbb{E}(Y(m) | \mathcal{P}_n)$$

dla  $n \leq m$ .

$\omega$	$Y(0)$	$Y(1)$	$Y(2)$	$X$
UUU	16.4	23.8	36.5	61.8
UUD	16.4	23.8	36.5	11.2
UDU	16.4	23.8	11.1	22.2
UDD	16.4	23.8	11.1	0
DUU	16.4	9	17.1	34.2
DUD	16.4	9	17.1	0
DDU	16.4	9	0.9	1.8
DDD	16.4	9	0.9	0

**Rysunek 6.3**

### Twierdzenie 6.1

Jeśli  $M$  jest martyngałem względem  $\mathcal{P}_n$  to

$$\mathbb{E}(M(m) | \mathcal{P}_n) = M(n) \quad \text{dla } n < m$$

### Dowód

Z definicji mamy

$$\begin{aligned} M(n) &= \mathbb{E}(M(n+1)|\mathcal{P}_n) \\ M(n+1) &= \mathbb{E}(M(n+2)|\mathcal{P}_{n+1}) \end{aligned} \quad (1)$$

Podstawiamy  $M(n+1)$  i dostajemy

$$\begin{aligned} M(n) &= \mathbb{E}(\mathbb{E}(M(n+2)|\mathcal{P}_{n+1})|\mathcal{P}_n) \\ &= \mathbb{E}(M(n+2)|\mathcal{P}_n) \end{aligned}$$

Pokazaliśmy tezę dla  $m = n + 2$ . Tak samo rozumiemy dla  $m = n + 3$  i dalej dla dowolnego  $m$  a formalnie widać, że możemy przeprowadzić dowód indukcyjny.  $\square$

### Twierdzenie 6.2

Jeśli  $M$  jest martyngałem względem  $\mathcal{P}_n$  to

$$\mathbb{E}(M(n+1) - M(n)|\mathcal{P}_n) = 0$$

### Dowód

Z definicji

$$\mathbb{E}(M(n+1)|\mathcal{P}_n) - M(n) = 0$$

Ponieważ  $M(n) = \mathbb{E}(\dots|\mathcal{P}_n)$ , zmienna  $M(n)$  jest stała na elementach  $\mathcal{P}_n$ , więc

$$\mathbb{E}(M(n)|\mathcal{P}_n) = M(n).$$

Podstawiamy  $M(n)$  do warunku definicji mamy

$$\mathbb{E}(M(n+1)|\mathcal{P}_n) - \mathbb{E}(M(n)|\mathcal{P}_n) = 0$$

a liniowość warunkowej wartości oczekiwanej daje

$$\mathbb{E}(M(n+1)|\mathcal{P}_n) - \mathbb{E}(M(n)|\mathcal{P}_n) = \mathbb{E}(M(n+1) - M(n)|\mathcal{P}_n)$$

co należało pokazać.  $\square$

Przypominamy, że jeśli  $\mathcal{P}_n$  jest drobniejszy niż  $\mathcal{P}_{M(n)}$  to mówimy że  $M(n)$  jest  $\mathcal{P}_n$ -mierzalna. Oznacza to że zbiór  $\{\omega : M^\omega(n) = a\}$  jest sumą pewnej liczby części z  $\mathcal{P}_n$ . Poniżej mamy wygodny do sprawdzania warunek na własność martyngału.



**Twierdzenie 6.3**

Jeśli  $M(n)$  jest mierzalny względem  $\mathcal{P}_n$  oraz  $\mathbb{E}(M(n+1) - M(n)|\mathcal{P}_n) = 0$  to  $M$  jest martynałem.

**Dowód**

Liniowość daje

$$\mathbb{E}(M(n+1) - M(n)|\mathcal{P}_n) = \mathbb{E}(M(n+1)|\mathcal{P}_n) - \mathbb{E}(M(n)|\mathcal{P}_n).$$

Z założenia:  $M(n)$  mierzalna względem  $\mathcal{P}_n$ , wynika że

$$\mathbb{E}(M(n)|\mathcal{P}_n) = M(n)$$

i podstawiając to do pierwszej równości mamy

$$\mathbb{E}(M(n+1) - M(n)|\mathcal{P}_n) = \mathbb{E}(M(n+1)|\mathcal{P}_n) - M(n).$$

Jej lewa strona równa się 0 (z założenia) więc

$$\mathbb{E}(M(n+1)|\mathcal{P}_n) - M(n) = 0$$

co kończy dowód.  $\square$

Wprowadźmy ważny przykład. Weźmy ciąg niezależnych zmiennych losowych

$$k_n = \begin{cases} 1 & \text{z prawdopodobieństwem } \frac{1}{2} \\ -1 & \text{z prawdopodobieństwem } \frac{1}{2} \end{cases} \quad n = 1, 2, \dots, N$$

i filtrację  $\mathcal{P}_n$  generowaną przez  $k_1, \dots, k_n$ . Ten ciąg reprezentuje ruchy punktu na osi liczbowej w wyniku rzutu monetą, w prawo lub w lewo o jednostkę. Pozycja zajmowana przez nasz punkt w chwili  $n$  jest ważną zmienną losową gdyż posłuży do skonstruowania procesu Wienera, który jest podstawą współczesnych modeli cen.

**Definicja 6.6**

Błądzenie przypadkowe to ciąg

$$\begin{aligned} w(0) &= 0 \\ w(n) &= k_1 + \dots + k_n \quad \text{dla } n = 1, 2, 3, \dots, N \end{aligned}$$

**Twierdzenie 6.4**

Ciąg  $w(n)$  jest martynałem.

### Dowód

Badamy warunek z definicji martyngału

$$\begin{aligned}
 \mathbb{E}(w(n+1)|\mathcal{P}_n) &= \mathbb{E}(k_1 + \dots + k_n + k_{n+1}|\mathcal{P}_n) \\
 &= \mathbb{E}(k_1|\mathcal{P}_n) + \dots + \mathbb{E}(k_n|\mathcal{P}_n) + \mathbb{E}(k_{n+1}|\mathcal{P}_n) \quad \text{liniowość} \\
 &= k_1 + \dots + k_n + \mathbb{E}(k_{n+1}) \\
 &\text{gdyż } k_i \text{ są } \mathcal{P}_n \text{ mierzalne a } k_{n+1} \text{ jest niezależny of } \mathcal{P}_n \\
 &= k_1 + \dots + k_n \quad \text{gdyż } \mathbb{E}(k_n) = 0 \\
 &= w(n)
 \end{aligned}$$

□

Powyższe pojęcie zasługuje na proste upgólnienie co pozwoli nam zobaczyć istotę rzeczy czyli kluczowe założenia gwarantujące własność martyngału.

Niech:

$$Y_k \quad k = 1, 2, 3, \dots$$

będą niezależnymi zmiennymi losowymi o takich samych rozkładach:  $P_{Y_k} = P_{Y_1}$  i niech filtracja  $\mathcal{P}_n$  będzie generowana przez  $Y_1, \dots, Y_n$ .

Niech

$$X(n) = \sum_{k=1}^n Y_k$$

### Twierdzenie 6.5

Jeżeli  $\mathbb{E}(Y_k) = 0$  to  $X(n)$  jest martyngałem.

### Dowód

Korzystamy z twierdzenia: jeśli  $\mathbb{E}(M(n+1) - M(n)|\mathcal{P}_n) = 0$  to  $M$  jest martyngałem:

$$\begin{aligned}
 \mathbb{E}(X(n+1) - X(n)|\mathcal{P}_n) &= \mathbb{E}(Y_{n+1}|\mathcal{P}_n) \\
 &= \mathbb{E}(Y_{n+1}) \\
 &= 0
 \end{aligned}$$

□

Poniższe twierdzenie daje warunek konieczny. Jeśli ciąg nie ma stałej wartości oczekiwanej to nie może być martyngałem.

### Twierdzenie 6.6

Jeśli  $M(n)$  jest martynałem względem  $\mathcal{P}_n$ , to

$$\mathbb{E}(M(0)) = \mathbb{E}(M(1)) = \mathbb{E}(M(2)) = \dots$$

### Dowód

Z definicji

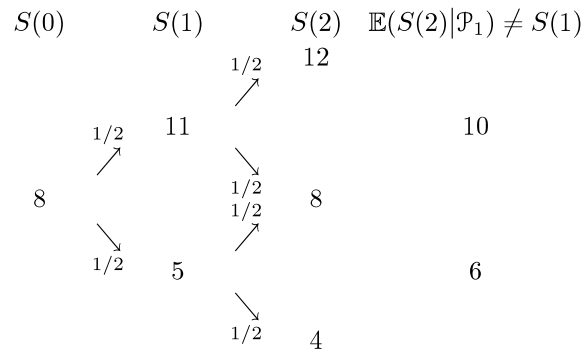
$$M(n) = \mathbb{E}(M(n+1)|\mathcal{P}_n)$$

Obkładając obie strony wartością oczekiwaną i korzystając z tego że wartość oczekiwana warunkowej wartości oczekiwanej redukuje się do wartości oczekiwanej mamy

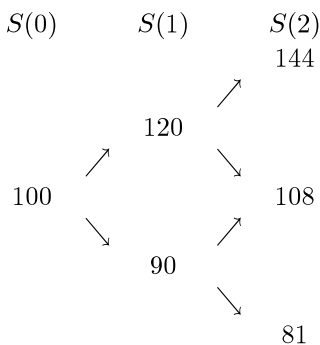
$$\mathbb{E}(M(n)) = \mathbb{E}(\mathbb{E}(M(n+1)|\mathcal{P}_n)) = \mathbb{E}(M(n+1)).$$

□

Stała wartość oczekiwana to warunek konieczny ale nie wystarczający jak pokazuje kolejny przykład. Mamy w nim  $\mathbb{E}(S(0)) = \mathbb{E}(S(1)) = \mathbb{E}(S(2)) = 8$  ale nie jest to martynał:



Weźmy teraz prosty przypadek drzewa dwumianowego z prawdopodobieństwami  $\frac{1}{2}$  i zwrotami 20%,  $-10\%$  w każdym z dwóch kroków



Bierzemy naturalną filtrację

$$\begin{aligned}\mathcal{P}_0 &= \{\Omega\} \\ \mathcal{P}_1 &= \{B_U, B_D\}\end{aligned}$$

i badamy, czy  $S(n)$  jest martyngałem. Łatwo sprawdzić, że

$$\begin{aligned}\mathbb{E}(S(2)|B_U) &= 126 \\ \mathbb{E}(S(2)|B_D) &= 94.5\end{aligned}$$

więc

$$\mathbb{E}(S(2)|\mathcal{P}_1) \neq S(1).$$

Sytuacja jest inna jeśli prawdopodobieństwo ruchu w górę jest  $p_* = \frac{1}{3}$  i wtedy własność martyngału zachodzi (do warunkowej wartości oczekiwanej względem  $p_*$  stosujemy oznaczenie  $\mathbb{E}_*$ ):

$$\begin{aligned}\mathbb{E}_*(S(2)|B_U) &= \frac{1}{\frac{1}{3}}\left(\frac{1}{3} \times \frac{1}{3} \times 144 + \frac{1}{3} \times \frac{2}{3} \times 108\right) = 120 \\ \mathbb{E}_*(S(2)|B_D) &= \frac{1}{\frac{2}{3}}\left(\frac{2}{3} \times \frac{1}{3} \times 99 + \frac{2}{3} \times \frac{2}{3} \times 81\right) = 90\end{aligned}$$

więc

$$\mathbb{E}_*(S(2)|\mathcal{P}_1) = S(1).$$

To jeszcze nie koniec gdyż pozostał jeden przypadek do sprawdzenia:

$$\mathbb{E}_*(S(1)|\mathcal{P}_0) = \frac{1}{3} \times 120 + \frac{2}{3} \times 90 = 100 = S(0).$$

Mamy więc przykład martyngału co prowokuje naturalne pytania: czy każdy ciąg zmiennych losowych może być martyngałem względem stosownego prawdopodobieństwa i jak znaleźć takie prawdopodobieństwo. Łatwo odpowiedzieć

na drugie pytanie dla drzewa dwumianowego. Tu o wszystkim decyduje pojedynczy krok, powtarzany wielokrotnie z rónnymi cenami wyjściowymi. Dla pierwszego kroku, aby ciąg  $S(0), S(1)$  był martynałem wystarczy aby

$$\mathbb{E}_*(S(1)) = S(0)$$

co oznacza, że

$$p^*S(0)(1+U) + (1-p^*)S(0)(1+D) = S(0).$$

To równanie możemy rozwiązać względem  $p^*$ ,

$$p^* = \frac{-D}{U-D}.$$

Warunki aby  $p_*$  dawało prawdopodobieństwo (czyli  $p_* \in (0,1)$ ) dają ograniczenia na zwroty

$$\begin{aligned} 0 < p_*, & \text{ gdy } D < 0, \\ p_* < 1, & \text{ gdy } U > 0. \end{aligned}$$

Problem "poprawiania" modelu aby uzyskać własność martynału jest istotny, własność ta odpowiada pożądanym zjawiskom. W związku z tym zobaczmy co się dzieje jeśli "popsuć" martynał.

Przypuśćmy że  $M(n)$  jest martynałem. Zdefiniujmy

$$\begin{aligned} S(0) &= M(0) \\ S(1) &= M(1) + 1 \\ S(2) &= M(2) + 2 \quad \text{etc.} \end{aligned}$$

Ciąg  $S(n) = M(n) + n$  nie jest martynałem bo ciąg wartości oczekiwanych

$$\begin{aligned} \mathbb{E}(S(n)) &= \mathbb{E}(M(n) + n) \\ &= \mathbb{E}(M(n)) + n \\ &= \mathbb{E}(M(0)) + n \end{aligned}$$

nie jest ciągiem stałym. Badając wartości warunkowe mamy

$$\begin{aligned} \mathbb{E}(S(n+1)|\mathcal{P}_n) &= \mathbb{E}(M(n+1) + n+1|\mathcal{P}_n) \\ &= M(n) + n+1 \quad \text{bo } \mathbb{E}(M(n+1)|\mathcal{P}_n) = M(n) \\ &= S(n) + 1 > S(n) \end{aligned}$$

co motywuje kolejną definicję.

### Definicja 6.7

Ciąg  $S(n)$  jest submartyngałem względem  $\mathcal{P}_n$  jeśli jest adaptowany oraz

$$\mathbb{E}(S(n+1)|\mathcal{P}_n) \geq S(n).$$

Przypuśćmy, że  $M(n)$  jest martyngałem. Zdefiniujmy

$$T(n) = M(n) - n$$

Ciąg  $T(n)$  nie jest martyngałem gdyż wartości oczekiwane maleją

$$\mathbb{E}(T(n)) = \mathbb{E}(M(n) - n) = \mathbb{E}(M(n)) - n = \mathbb{E}(M(0)) - n.$$

Dla warunkowych wartości oczekiwanych mamy analogiczną nierówność

$$\begin{aligned} \mathbb{E}(T(n+1)|\mathcal{P}_n) &= \mathbb{E}(M(n+1) - n - 1|\mathcal{P}_n) \\ &= M(n) - n - 1 \quad \text{bo } \mathbb{E}(M(n+1)|\mathcal{P}_n) = M(n) \\ &= T(n) - 1 < T(n) \end{aligned}$$

### Definicja 6.8

Ciąg  $T(n)$  jest supermartyngałem względem  $\mathcal{P}_n$  jeśli jest adaptowany oraz

$$\mathbb{E}(T(n+1)|\mathcal{P}_n) \leq T(n).$$

### Uwaga 6.1

Terminologia wywodzi się z innej gałęzi matematyki i nie jest zbyt fortunna, gdyż jeśli  $M$  reprezentuje posiadane pieniądze to nie jest wcale "super" że one średnio maleją. Stosowane są też polskie terminy: podmartyngał zamiast submartyngał (oraz nadmartyngał zamiast supermartyngał) ale jest jeszcze gorzej gdyż jak pokazuje przykład, podmartyngał jest nad martyngałem a nie pod nim.

Oczywiście martyngał jest submartyngałem i supermartyngałem gdyż mamy równość która implikuje słabą nierówność (w ostatnich definicjach mamy właśnie słabe nierówności, sytuacja byłaby inna gdyby były mocne).

Jeśli  $M$  jest submartyngałem i supermartyngałem to jest martyngałem gdyż mamy dwie przeciwne nierówności

$$\begin{aligned} \mathbb{E}(M(n+1)|\mathcal{P}_n) &\geq M(n) \\ \mathbb{E}(M(n+1)|\mathcal{P}_n) &\leq M(n) \end{aligned}$$

które pociągają równość

$$\mathbb{E}(M(n+1)|\mathcal{P}_n) = M(n).$$

### 6.3 Własności

Martyngał jest uważany za model gry równych szans. Wygrana, średnio, pozostaje bez zmian, czyli szanse obu stron w przypadku gry są wyrównane. To jest interpretacja dość istotna dla rynków finansowych, gdzie ważne jest aby nie było uprzywilejowania niektórych uczestników. Udowodniliśmy powyżej, że jeśli  $M$  jest martyngałem, to wartość oczekiwana  $\mathbb{E}(M(n))$  nie zależy od  $n$ . Poniższe własności nie są zaskakujące gdyż widzieliśmy je na przykładach.

#### Twierdzenie 6.7

Jeśli  $S(n)$  jest submartyngałem to

$$\mathbb{E}(S(0)) \leq \mathbb{E}(S(1)) \leq \mathbb{E}(S(2)) \leq \dots$$

#### Dowód

Z definicji submartyngału mamy

$$S(n) \leq \mathbb{E}(S(n+1)|\mathcal{P}_n).$$

Ogólnie, jeśli  $X \leq Y$  to  $\mathbb{E}(X) \leq \mathbb{E}(Y)$  więc obkładając obie strony nierówności wartością oczekiwaną mamy

$$\begin{aligned} \mathbb{E}(S(n)) &\leq \mathbb{E}(\mathbb{E}(S(n+1)|\mathcal{P}_n)) \\ &= \mathbb{E}(S(n+1)) \end{aligned}$$

gdyż  $\mathbb{E}(\mathbb{E}(X|\mathcal{P})) = \mathbb{E}(X)$ . □

#### Twierdzenie 6.8

Jeśli  $S(n)$  jest supermartyngałem, to

$$\mathbb{E}(S(0)) \geq \mathbb{E}(S(1)) \geq \mathbb{E}(S(2)) \geq \dots$$

#### Dowód

Powtarzamy poprzedni argument zamieniając symbol  $\leq$  na  $\geq$ . □

Submartyngał to gra korzystna dla gracza gdyż średnio jego stan posiadania się zwiększa, natomiast supermartyngał modeluje sytuację odwrotną. Ciekawa jest następna własność pokazująca czego brak submartyngałowi aby być martyngałem.

**Twierdzenie 6.9**

Submartyngał o stałej wartości oczekiwanej jest martyngałem.

**Dowód**

Zakładamy

$$\mathbb{E}(S(n)) = \mathbb{E}(S(1)).$$

Ponieważ  $S(n)$  jest submartyngałem, to z definicji

$$S(n) \leq \mathbb{E}(S(n+1)|\mathcal{P}_n)$$

czyli  $X = \mathbb{E}(S(n+1)|\mathcal{P}_n) - S(n) \geq 0$ . Liczymy wartość oczekiwaną zmiennej  $X$  :

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(S(n+1)|\mathcal{P}_n) - S(n)) \\ &= \mathbb{E}(\mathbb{E}(S(n+1)|\mathcal{P}_n)) - \mathbb{E}(S(n)) \quad \text{liniowość} \\ &= \mathbb{E}(S(n+1)) - \mathbb{E}(S(n)) \quad \text{gdyż } \mathbb{E}(\mathbb{E}(X|\mathcal{P})) = \mathbb{E}(X) \\ &= 0 \quad \text{bo } \mathbb{E}(S(n)) \text{ jest stałe} \end{aligned}$$

May nieujemną zmienną losową  $X$  o zerowej wartości oczekiwanej. Musi być więc  $X = 0$ , ściśle mówiąc,  $X$  może być niezerowa ale na zbiorze o prawdopodobieństwie zerowym.  $\square$

**Twierdzenie 6.10**

Supermartyngał o stałej wartości oczekiwanej jest martyngałem.

**Dowód**

Argument jest identyczny jak poprzednio ale inna definicja zmienne  $X$  :

$$X = S(n) - \mathbb{E}(S(n+1)|\mathcal{P}_n) \geq 0,$$

$\mathbb{E}(X) = 0$ , więc  $X = 0$  (prawie wszędzie).  $\square$

Widzieliśmy już przykład jak manipulując prawdopodobieństwem można uzyskać własność martyngału. Teraz na temat "poprawiania" ciągów aby zrobić z nich martyngały udowodnimy twierdzenie pochodzące od Dooba i Meyera, które staje się intuicyjnie jasne w kontekście naszych przykładów gdzie psuliśmy martyngał dodając stałe i dostając submartyngał. jeśli więc mamy submartyngał, gdzie wartość oczekiwana rośnie, to logiczne jest że trzeba coś odjąć aby dostać martyngał.



Typowym przykłądem do którego to twierdzenie się stosuje jest następujący przypadek. Jeśli  $M(n)$  jest martynałem to  $M^2(n)$  nim nie jest ale na podstawie nierówności Jensena, która mówi że:  $\mathbb{E}(X^2|\mathcal{P}) \geq (\mathbb{E}(X|\mathcal{P}))^2$  mamy własność submartynału

$$\mathbb{E}(M^2(n+1)|\mathcal{P}_n) \geq (\mathbb{E}(M(n+1)|\mathcal{P}_n))^2 = M^2(n)$$

i w zastosowaniach poprawianie tego submartynału staje się konieczne.

### Twierdzenie 6.11

Jeśli  $S(n)$  jest submartynałem to istnieją jedyne

- $M(n)$  martynał
- $A(n)$  proces o własnościach
  1.  $A(0) = 0$
  2. rosnący:  $A(n) \leq A(n+1)$
  3. przewidywalny:  $A(n)$  jest  $\mathcal{P}_{n-1}$  mierzalne

takie, że

$$S(n) - A(n) = M(n).$$

Twierdzenie to nazywane jest "rozkład Doobe-Meyera" gdyż możemy zapisać

$$S(n) = M(n) + A(n)$$

czyli rozkładamy submartynał na sumę martynału i procesu przewidywalnego i rosnącego. Jednoznaczność jest zaskakująca gdyż rozkład liczby na sumę nie jest jednoznaczny. Ta jednoznaczność przypomina rozkład wektora na płaszczyźnie na liniową kombinację składowych. Występuje tu jednoznaczność a kluczem jest liniowa niezależność lich składowych.

Dowód nie jest trudny, jest efektwny (jednoznaczność) ale na pewno jest nadobowiązkowy.

### Dowód

Stosujemy heurzę czyli jak gdyby zakładamy, że mamy poszukiwane  $M$  i  $A$  a warunki które pozwolą je wyznaczyć konkretnie wynikną z analizy tezy. Nasz cel to pokazanie

$$\begin{aligned} S(n) - A(n) &= \mathbb{E}(S(n+1) - A(n+1)|\mathcal{P}_n) \\ &= \mathbb{E}(S(n+1)|\mathcal{P}_n) - \mathbb{E}(A(n+1)|\mathcal{P}_n) \quad \text{liniowość} \\ &= \mathbb{E}(S(n+1)|\mathcal{P}_n) - A(n+1) \quad \text{gdyż } A(n+1) \text{ jest } \mathcal{P}_n \text{ mierzalny} \end{aligned}$$

Definiujemy  $A$  zgodnie z powyższą równością (potem zdefiniować  $m$  będzie już łatwo, będzie to stosowna różnica)

$$\begin{aligned} A(n+1) &= \mathbb{E}(S(n+1)|\mathcal{P}_n) - S(n) + A(n), \\ A(0) &= 0 \end{aligned}$$

Własności  $A(n)$  w kontekście tezy twierdzenia są następujące

1.  $A(0) = 0$  z definicji
2. jest rosnący gdyż

$$A(n+1) - A(n) = \mathbb{E}(S(n+1)|\mathcal{P}_n) - S(n) \geq 0$$

bo  $S$  jest submartyngałem.

3. jest przewidywalny: Z definicji widac ze  $A(n+1)$  jest zdefiniowany z użyciem zmiennych losowych związanych z chwilą  $n$ . Są one  $\mathcal{P}_n$  mierzalne, a więc  $A(n+1)$  też odziedziczy tę własność.

Definiujemy  $M$

$$M(n) = S(n) - A(n).$$

Pytanie: czy  $M(n)$  jest martyngałem? Sprawdzamy definicję

$$\begin{aligned} \mathbb{E}(M(n+1)|\mathcal{P}_n) &= \mathbb{E}(S(n+1) - A(n+1)|\mathcal{P}_n) \\ &= \mathbb{E}(S(n+1)|\mathcal{P}_n) - \mathbb{E}(A(n+1)|\mathcal{P}_n) \quad \text{liniowość. Wstawiamy } A(n+1) : \\ &= \mathbb{E}(S(n+1)|\mathcal{P}_n) - \mathbb{E}(\mathbb{E}(S(n+1)|\mathcal{P}_n) + S(n) - A(n)|\mathcal{P}_n) \\ &= \mathbb{E}(S(n+1)|\mathcal{P}_n) - \mathbb{E}(S(n+1)|\mathcal{P}_n) + S(n) - A(n) \\ &= M(n) \quad \text{z definicji } M(n) \end{aligned}$$

Ostatnie pytanie, to czy rozkład  $S(n) = M(n) + A(n)$  jest jednoznaczny? Aby to pokazać, przypuśćmy dla dowodu nie wprost, że istnieje inny sposób zapisania  $S$  w postaci sumy martyngału i procesu  $A$  o znanych własnościach:

$$S(n) = M'(n) + A'(n)$$

Mamy więc

$$M'(n) + A'(n) = M(n) + A(n)$$

a po przekształceniach

$$M'(n) - M(n) = A(n) - A'(n).$$

Niech  $n = 0$ ,  $A(0) = A'(0)$ , gdyż oba są zerami, więc  $M'(0) = M(0)$ .

Niech  $n = 1$ . Mamy

$$M'(1) - M(1) = A(1) - A'(1)$$

co obkładamy obustronnie  $\mathbb{E}(\cdot|\mathcal{P}_0)$  czyli w istocie wartością oczekiwaną

$$\mathbb{E}(M'(1) - M(1)|\mathcal{P}_0) = \mathbb{E}(A(1) - A'(1)|\mathcal{P}_0).$$

Z liniowości mamy

$$\mathbb{E}(M'(1)|\mathcal{P}_0) - \mathbb{E}(M(1)|\mathcal{P}_0) = \mathbb{E}(A(1)|\mathcal{P}_0) - \mathbb{E}(A'(1)|\mathcal{P}_0)$$

Z lewej strony korzystamy z własności martynału a z prawej z przewidywalności ciągu  $A$  i dostajemy

$$M'(0) - M(0) = A(1) - A'(1)$$

Ale wiemy już że  $M(0) = M'(0)$  więc  $A(1) = A'(1)$ , co z kolei daje  $M'(1) = M(1)$  bo

$$M'(1) - M(1) = A(1) - A'(1)$$

Niech  $n = 2$ . Powtarzamy to rozumowanie, również i dla dalszych  $n$ , lub stosujemy indukcję.  $\square$



## Ciągi zmiennych losowych

### 7.1 Zbieżność ciągów zmiennych losowych

Skończone ciągi zmiennych losowych mogą nie wystarczyć do budowy dostatecznie wyrafinowanych modeli. Dyskretny czas oznacza, że zmiany cen zachodzą co pewien z góry ustalony przedział czasowy. To jest niepraktyczne ograniczenie bo niekiedy te zmiany zachodzą bardzo często. Pojawia się konieczność zagęszczania siatki momentów czasowych i prowadzi to do badania przejścia granicznego gdzie długość kroku zmierza do zera i pojawia się model w czasie ciągłym którego zaletą jest brak ograniczeń na chwile zmian.

W naturalny sposób pojawia się kwestia badania zbieżności ciągów zmiennych losowych. Dla ciągów liczb jest to sprawa prosta. Zmienne losowe są to jednak funkcje i ich zbieżność ma różne formy. Załóżmy, że mamy

$$\begin{aligned} X_n : \Omega &\rightarrow \mathbb{R}, \quad n = 1, 2, 3, \dots \\ X : \Omega &\rightarrow \mathbb{R}. \end{aligned}$$

Pytamy co to znaczy, że

$$X_n \rightarrow X?$$

Najprostsza zbieżność to zbieżność wartości  $X_n^\omega$  dla wszystkich  $\omega \in \Omega$ . Nie jest to jednak praktyczne gdyż zdarzenia o prawdopodobieństwie zero muszą być zignorowane.

### Przykład 7.1

Weźmy  $\Omega = [0, 1]$ ,  $P = m$  oraz ciąg dany wzorem

$$\begin{aligned} X_n^\omega &= \omega^n, \\ X^\omega &= 0, \quad \omega \in [0, 1]. \end{aligned}$$

Jeśli  $\omega \in [0, 1)$ , to ciąg  $\omega^n$  jest zbieżny do zera, natomiast dla  $\omega = 1$ ,  $\omega^n = 1 \not\rightarrow 0$  czyli

$$X_n \rightarrow X \quad \text{na } [0, 1), \quad P([0, 1)) = 1$$

ale

$$X_n \not\rightarrow X \quad \text{na } \{1\}, \quad P(\{1\}) = 0$$

Prowadzi to do następującej definicji.

### Definicja 7.1

Mówimy że  $X_n \rightarrow X$  prawie wszędzie (w skrócie p.w.) jeśli  $X_n^\omega \rightarrow X^\omega$  dla wszystkich  $\omega \in \Omega' \subset \Omega$  dla pewnego zdarzenia  $\Omega'$  gdzie  $P(\Omega \setminus \Omega') = 0$ , czyli  $P(\Omega') = 1$ .

W skrócie piszemy

$$P(X_n \rightarrow X) = 1.$$

Jest oczywiste, że zbieżność prawie wszędzie zależy od wyboru prawdopodobieństwa

### Przykład 7.2

Firma zaciąga kredyt 1mln. zabezpieczony aktywami. Bank dostaje

$$\begin{cases} 1 & \text{jeśli aktywa są więcej warte niż 1} \\ \omega & \text{jeśli aktywa są warte } \omega \leq 1 \end{cases}$$

Przypuśćmy, że prawdopodobieństwo bankructwa wynosi 10%. Przyjmujemy uproszczony model gdzie  $\omega \in \Omega = [0, 1]$  to wypłata dla banku,  $P = \frac{9}{10}\delta_1 + \frac{1}{10}m$  czyli

$$\begin{aligned} P(\{1\}) &= \frac{9}{10}, \\ P([0, \omega]) &= \frac{1}{10}\omega. \end{aligned}$$

Ciąg  $X_n^\omega = \omega^n$  nie zmierza do  $X^\omega = 0$  prawie wszędzie bo

$$P(X_n \not\rightarrow X) = \frac{9}{10}$$

**Przykład 7.3**

Bierzemy  $\Omega = [0, 1]$ ,  $P = m$ ,

$$X_n^\omega = n\mathbb{I}_{[0, \frac{1}{n}]}(\omega).$$

Dla  $\omega = 0$  mamy rozbieżność

$$X_n^0 = n\mathbb{I}_{[0, \frac{1}{n}]}(0) = n \rightarrow \infty.$$

Dla  $\omega > 0$  mamy zbieżność gdyż

$$X_n^\omega = \begin{cases} n & \text{gdy } \omega \leq \frac{1}{n} \\ 0 & \text{gdy } \omega > \frac{1}{n}. \end{cases}$$

Na przykład, dla  $\omega = \frac{1}{3}$ ,

$$X_1^\omega = 1$$

$$X_2^\omega = 2$$

$$X_3^\omega = 3$$

$$X_4^\omega = 0 \text{ bo } \frac{1}{3} > \frac{1}{4}$$

$$X_n^\omega = 0 \text{ dla pozostałych } n$$

Ciąg zmierza więc do  $X = 0$  prawie wszędzie

$$P(X_n \rightarrow 0) = P((0, 1]) = 1.$$

Pojawia się naturalne pytanie, czy zbieżność prawie wszędzie pociąga zbieżność wartości oczekiwanych czy wariancji. Tak być nie musi, jak pokaże poniższy przykład, więc pojawiają się warianty pojęcia zbieżności ciągu zmiennych losowych.

**Przykład 7.4**

Analizujemy sytuację opisaną w poprzednim przykładzie:  $\Omega = [0, 1]$ ,  $P = m$ ,  $X_n^\omega = n\mathbb{I}_{[0, \frac{1}{n}]}(\omega)$ ,  $X^\omega = 0$  i wiemy, że  $P(X_n \rightarrow 0) = 1$ . Liczymy wartość oczekiwaną

$$\begin{aligned} \mathbb{E}(X_n) &= \int_0^1 n\mathbb{I}_{[0, \frac{1}{n}]}(x)dx \\ &= n \int_0^1 \mathbb{I}_{[0, \frac{1}{n}]}(x)dx \\ &= nP([0, \frac{1}{n}]) = 1 \end{aligned}$$

czyli nie mamy zbieżności  $\mathbb{E}(X_n)$  do  $\mathbb{E}(X) = 0$ .

Pewna modyfikacja tego przykładu da nam jednak zbieżność wartości oczekiwanych.

### Przykład 7.5

Jak poprzednio:  $\Omega = [0, 1]$ ,  $P = m$  ale inaczej definiujemy ciąg

$$\begin{aligned} X_n^\omega &= \sqrt{n} \mathbb{I}_{[0, \frac{1}{n}]}(\omega) \\ X^\omega &= 0 \end{aligned}$$

Wtedy łatwo widać, że

$$P(X_n \rightarrow 0) = 1$$

choćby dlatego, że wartości nowego ciągu są mniejsze. Liczymy wartość oczekiwaną

$$\begin{aligned} \mathbb{E}(X_n) &= \int_0^1 \sqrt{n} \mathbb{I}_{[0, \frac{1}{n}]}(x) dx \\ &= \sqrt{n} \int_0^1 \mathbb{I}_{[0, \frac{1}{n}]}(x) dx \\ &= \sqrt{n} P([0, \frac{1}{n}]) \\ &= \frac{1}{\sqrt{n}} \rightarrow 0 \end{aligned}$$

czyli mamy zbieżność

$$\mathbb{E}(X_n) \rightarrow 0 = \mathbb{E}(X).$$

### Definicja 7.2

Wprowadzamy oznaczenie na całkowalne zmienne losowe:

$$L^1(\Omega) = \{Y : \Omega \rightarrow \mathbb{R} \text{ takie że } \mathbb{E}(|Y|) < \infty\}$$

Przypominamy że dowolną zmienną losową możemy zapisać jako różnicę zmiennych nieujemnych  $Y = Y^+ - Y^-$  gdzie  $Y^+ = \max\{Y, 0\} \geq 0$ ,  $Y^- = \max\{-Y, 0\} \geq 0$  i wartość oczekiwana (czyli całka) dana jest wzorem

$$\mathbb{E}(Y) = \mathbb{E}(Y^+) - \mathbb{E}(Y^-)$$

o ile  $\mathbb{E}(Y^+) < \infty$  i  $\mathbb{E}(Y^-) < \infty$ . Wtedy również istnieje całka z wartości bezwzględnej gdyż  $|Y| = Y^+ + Y^-$ .

Na odwrót, jeśli  $\mathbb{E}(|Y|) < \infty$  to  $\mathbb{E}(Y) < \infty$  bo  $\mathbb{E}(Y^+) \leq \mathbb{E}(|Y|)$  i  $\mathbb{E}(Y^-) \leq \mathbb{E}(|Y|)$ .



**Przykład 7.6**

Dla dowolnego  $\Omega$  zakładamy że  $X$  jest ograniczona:

$$-L \leq X^\omega \leq L$$

gdzie  $L$  jest pewna stałą. wtedy

$$\mathbb{E}(|X|) \leq \mathbb{E}(L) = L < \infty$$

czyli

$$X \in L^1(\Omega).$$

**Przykład 7.7**

Niech  $\Omega = [0, 1]$ ,  $P = m$

$$Y(x) = \begin{cases} \frac{1}{x} & \text{dla } x \in (0, 1] \\ 0 & \text{dla } x = 0 \end{cases}$$

Wtedy

$$\mathbb{E}(Y) = \int_0^1 \frac{1}{x} dx = \ln x \Big|_0^1 = \infty$$

czyli

$$Y \notin L^1([0, 1])$$

Natomiast dla

$$Z(x) = \begin{cases} \frac{1}{\sqrt{x}} & \text{dla } x \in (0, 1] \\ 0 & \text{dla } x = 0 \end{cases}$$

mamy

$$\mathbb{E}(Z) = \int_0^1 \frac{1}{\sqrt{x}} dx = \int_0^1 x^{-\frac{1}{2}} dx = 2x^{\frac{1}{2}} \Big|_0^1 = 2$$

czyli

$$Z \in L^1([0, 1]).$$

**Definicja 7.3**

Jeśli  $X_n, X \in L^1(\Omega)$  wtedy mówimy, że  $X_n \rightarrow X$  w  $L^1(\Omega)$  jeśli

$$\mathbb{E}(|X_n - X|) \rightarrow 0.$$

Istota tej definicji to badanie wartości oczekiwanych, ale nie stawiamy warunku  $\mathbb{E}(X_n) \rightarrow \mathbb{E}(X)$ , który nie ma większego sensu jak pokazuje przykład.

### Przykład 7.8

Niech  $\Omega = [0, 1]$ ,  $X = 0$ ,  $X_n = I_{[0, \frac{1}{n})} - I_{[\frac{1}{n}, 1]}$  (z interpretacją: wygrywam złotówkę gdy wypadnie orzeł a przegrywam gdy wypadnie reszka) i wtedy wartości oczekiwane wszystkich tych zmiennych losowych wynoszą zero, ale trudno się zgodzić z twierdzeniem, że  $X_n$  zmierzają do zera. Badając warunek zbieżności w  $L^1$  mamy oczywiście

$$\mathbb{E}(|X_n - X|) = \mathbb{E}(|X_n|) = \mathbb{E}(I_{[0,1]}) = 1 \not\rightarrow 0.$$

Mamy znane twierdzenie o zbieżności zdominowanej które pomaga nam odowodnić rozważaną zbieżność.

### Twierdzenie 7.1

Jeśli  $|X_n| \leq Y$  dla  $Y \in L^1(\Omega)$  oraz  $X_n \rightarrow X$  p.w. to  $X_n \rightarrow X$  w  $L^1(\Omega)$ .

### Przykład 7.9

Dla  $\Omega = [0, 1]$ ,  $P = m$  kładziemy

$$Y(x) = \begin{cases} \frac{1}{\sqrt{x}} & \text{dla } x \in (0, 1] \\ 0 & \text{dla } x = 0 \end{cases}$$

i wiemy już, że  $Y \in L^1([0, 1])$ . Dla ciągu  $X_n$  ta funkcja posłuży jako ograniczenie

$$X_n(x) = \sqrt{n} I_{[0, \frac{1}{n}]}(x) \leq \frac{1}{\sqrt{x}}$$

więc biorąc funkcje graniczną  $X = 0$  mamy

$$\mathbb{E}(|X_n - X|) \rightarrow 0$$

Wariancja wymaga całkowania kwadratu, stąd naturalna jest następująca wersja zbieżności:

### Definicja 7.4

Wprowadzamy oznaczenie na zmienne losowe całkowalne z kwadratem:

$$L^2(\Omega) = \{Y : \Omega \rightarrow \mathbb{R} \text{ takie że } \mathbb{E}(Y^2) < \infty\}.$$

**Przykład 7.10**

Bierzemy nieco mniejszy zbiór aby rozpatrywane funkcje były dobrze określone:  $\Omega = (0, 1]$ ,  $P = m$  i definiujemy

$$Y(x) = \begin{cases} \frac{1}{\sqrt{x}} & \text{dla } x \in (0, 1] \end{cases}$$

Wiemy, że  $\mathbb{E}(Y) < \infty$ , czyli  $Y \in L^1((0, 1])$ . Biorąc kwadrat mamy

$$Y^2(x) = \begin{cases} \frac{1}{x} & \text{dla } x \in (0, 1] \end{cases}$$

czyli  $\mathbb{E}(Y^2) = \infty$ , co pociąga  $Y \notin L^2((0, 1])$ .

Mamy wniosek, że

$$L^1(\Omega) \not\subset L^2(\Omega)$$

Zachodzi natomiast inkluzja odwrotna

$$L^2(\Omega) \subset L^1(\Omega)$$

gdyż

$$2X \leq 1 + X^2 \quad \text{bo } a^2 - 2a + 1 = (a - 1)^2 \geq 0$$

co daje oszacowanie

$$\mathbb{E}(X) = \frac{1}{2}\mathbb{E}(2X) \leq \frac{1}{2}\mathbb{E}(1 + X^2) = \frac{1}{2} + \frac{1}{2}\mathbb{E}(X^2) < \infty.$$

Tu istotne jest że  $P(\Omega)$  jest skończone bo dzięki temu funkcja stała jest całkowalna.

**Definicja 7.5**

Jeśli  $X_n, X \in L^2$  wtedy mówimy, że  $X_n \rightarrow X$  w  $L^2$  jeśli

$$\mathbb{E}([X_n - X]^2) \rightarrow 0.$$

Ta zbieżność jest mocniejsza niż zbieżność w  $L^1$  na podstawie wspomnianej wcześniej nierówności Jensena

$$[\mathbb{E}(|X_n - X|)]^2 \leq \mathbb{E}([X_n - X]^2).$$

### Przykład 7.11

Pokażemy, że zbieżność w  $L^1$  czy w  $L^2$  nie implikuje zbieżności prawie wszędzie co jest zaskakujące, gdyż ktoś mógłby sądzić, że jeśli  $E(X_n^2) \rightarrow 0$ , że wtedy również ciąg  $X_n$  musi zmierzać do zera. Tradycyjnie bierzemy  $\Omega = [0, 1]$  natomiast konstrukcja ciągu jest wyrafinowana,

$$\begin{aligned} X_1 &= \mathbb{I}_{[0,1]}, \\ X_2 &= \mathbb{I}_{[0, \frac{1}{2})}, X_3 = \mathbb{I}_{[\frac{1}{2}, 1)}, \\ X_4 &= \mathbb{I}_{[0, \frac{1}{4})}, X_5 = \mathbb{I}_{[\frac{1}{4}, \frac{1}{2})}, X_6 = \mathbb{I}_{[\frac{1}{2}, \frac{3}{4})}, X_7 = \mathbb{I}_{[\frac{3}{4}, 1)}, \\ X_8 &= \mathbb{I}_{[0, \frac{1}{8})}, X_9 = \mathbb{I}_{[\frac{1}{8}, \frac{1}{4})} \text{ i tak dalej.} \end{aligned}$$

Bierzemy  $X^\omega = 0$ , i mamy  $X_n \rightarrow X$  w  $L^1$  oraz w  $L^2$  gdyż

$$\mathbb{E}(X_n) = \mathbb{E}(X_n^2) \rightarrow 0.$$

Biorąc jednak dowolny punkt  $\omega$  znajdziemy nieskończenie wiele indeksów  $n$  takich że  $X_n^\omega = 1$  czyli ciąg liczb  $X_n^\omega$  nie zmierza do zera

$$P(X_n \rightarrow 0) = 0$$

(ten ciąg zbudowany jest z liczb 0 i 1 a taki ciąg zmierza do zera tylko wtedy jeśli od pewnego momentu jest równy 0 czyli zawiera skończoną liczbę jedynek).

Dla naszych celów powyższe rodzaje zbieżności są za mocne, trudno jest je pokazać w ważnych konkretnych przypadkach. Następny typ zbieżności jest mniej wymagający.

### Definicja 7.6

Mówimy, że  $X_n \rightarrow X$  według prawdopodobieństwa jeśli dla każdego  $\varepsilon > 0$

$$P(|X_n - X| > \varepsilon) \rightarrow 0.$$

### Przykład 7.12

Rozważamy tradycyjnie:  $\Omega = [0, 1]$ ,  $P = m$ ,  $X = 0$  i znany ciąg

$$\begin{aligned} X_1 &= \mathbb{I}_{[0,1]}, \\ X_2 &= \mathbb{I}_{[0, \frac{1}{2})}, X_3 = \mathbb{I}_{[\frac{1}{2}, 1)}, \\ X_4 &= \mathbb{I}_{[0, \frac{1}{4})}, X_5 = \mathbb{I}_{[\frac{1}{4}, \frac{1}{2})}, X_6 = \mathbb{I}_{[\frac{1}{2}, \frac{3}{4})}, X_7 = \mathbb{I}_{[\frac{3}{4}, 1)}, \\ X_8 &= \mathbb{I}_{[0, \frac{1}{8})}, X_9 = \mathbb{I}_{[\frac{1}{8}, \frac{1}{4})} \text{ i tak dalej.} \end{aligned}$$

Dla  $\varepsilon \geq 1$ ,  $P(|X_n| > \varepsilon) = 0$  dla każdego  $n$ .

Dla  $\varepsilon < 1$ ,  $|X_n| > \varepsilon$  wtedy i tylko wtedy gdy  $X_n = 1$ ,

$$P(X_n = 1) = \text{długość przedziału na którym } X_n = 1$$

Te długości przedziałów maleją do zera. Czyli  $X_n \rightarrow X$  według prawdopodobieństwa.

Ten przykład ilustruje następane twierdzenie które przyjmujemy bez dowodu jak również pokazuje, że zbieżność według prawdopodobieństwa nie musi oznaczać zbieżności w  $L^1$  czy  $L^2$  (do tego można też wykorzystać wcześniejszy przykład  $X_n^\omega = n\mathbb{I}_{[0, \frac{1}{n}]}$  gdyż tu  $P(|X_n| > \varepsilon) \leq \frac{1}{n}$ ).

### Twierdzenie 7.2

Jeśli  $X_n \rightarrow X$  prawie wszędzie to  $X_n \rightarrow X$  według prawdopodobieństwa.

Zbieżność w  $L^1$  pociąga zbieżność według prawdopodobieństwa (jak również zbieżność w  $L^2$ , która jest mocniejsza niż zbieżność w  $L^1$ ). Dowód tego jest prosty i pouczający a stosuje ważną nierówność Czebyszewa.

### Twierdzenie 7.3 (Nierówność Czebyszewa)

Jeśli  $Y \geq 0$ ,  $\varepsilon > 0$ , to

$$P(Y \geq \varepsilon) \leq \frac{\mathbb{E}(Y)}{\varepsilon}.$$

### Dowód

To wynika natychmiast z własności całki. Oznaczmy  $A = \{\omega : Y^\omega \geq \varepsilon\}$  i wtedy

$$\begin{aligned} \mathbb{E}(Y) &= \int_{\Omega} Y dP \quad \text{z definicji wartości oczekiwanej} \\ &\geq \int_A Y dP \quad \text{całka po mniejszym zbiorze jest mniejsza} \\ &\geq \int_A \varepsilon dP \quad \text{gdyż na } A \text{ mamy } Y^\omega \geq \varepsilon \\ &= P(A)\varepsilon, \end{aligned}$$

□

### Twierdzenie 7.4

Zbieżność w  $L^1$  implikuje zbieżność według prawdopodobieństwa.

### Dowód

Niech  $Y_n = |X_n - X|$  i załóżmy, że  $\mathbb{E}(|X_n - X|) \rightarrow 0$ . Dla dowolnego  $\varepsilon > 0$  z nierówności Czebyszewa mamy

$$P(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}(|X_n - X|)}{\varepsilon} \rightarrow 0.$$

□

Wracając do ostatniego przykładu, gdzie nie mamy zbieżności prawie wszędzie (a nawet gorzej, mamy rozbieżność wszędzie) zauważmy co się stanie jeśli ograniczymy się do pewnego wybranego podciągu: weźmy wyrazy o numerach 1, 2, 4, 8, 16... i mamy kolejno

$X_1 = \mathbb{I}_{[0,1]}$ ,  $X_2 = \mathbb{I}_{[0,\frac{1}{2}]}$ ,  $X_4 = \mathbb{I}_{[0,\frac{1}{4}]}$ ,  $X_8 = \mathbb{I}_{[0,\frac{1}{8}]}$ ,  $X_{16} = \mathbb{I}_{[0,\frac{1}{16}]}$  i tak dalej

i widać, że

$$P(X_{2^k} \rightarrow 0) = 1$$

Jest to ilustracja ogólnego faktu, który będzie wykorzystany w kluczowej konstrukcji dla budowy modelu Blacka-Scholesa.

### Twierdzenie 7.5

Jeśli  $X_n \rightarrow X$  według prawdopodobieństwa to dla pewnego podciągu mamy zbieżność do  $X$  prawie wszędzie.

## 7.2 Centralne twierdzenie graniczne

Jak pamiętamy, wykres prawdopodobieństw rozkładu dwumianowego przypomina wykres gęstości rozkładu normalnego. Dla rozkładu dwumianowego mamy pionowe słupki o wysokości odpowiadającej prawdopodobieństwom. Można sobie wyobrazić (lub zbudować stosowną wersję wykresu), że mając więcej kroków słupki będą tak gęste że utworzą wykres funkcji - to nastąpi gdy grubość słupka będzie taka jak szerokość punktu na ekranie komputera. Wtedy zobaczylibyśmy, że wykres byłby identyczny jak wykres gęstości rozkładu normalnego. (Stosowne rysunki znajdziemy poniżej). Oznacza to, że mamy do czynienia ze zbieżnością rozkładów prawdopodobieństw. Ta zbieżność jest opisana przez twierdzenie uważane za najważniejsze w rachunku prawdopodobieństwa.

Wypowiedź twierdzenia poprzedzamy wprowadzeniem pewnych oznaczeń. Weźmy ciąg niezależnych zmiennych losowych  $X_k$  o tym samym rozkładzie, że

skończoną wartością oczekiwaną

$$\mu = \mathbb{E}(X_k) < \infty$$

i wariancją

$$\sigma^2 = \text{Var}(X_k) < \infty$$

Rozważmy sumy

$$S_n = X_1 + \cdots + X_n$$

i policzmy ich wartości oczekiwane:

$$\mathbb{E}(S_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = n\mu$$

oraz wariancje (korzystamy z niezależności która daje wariancję sumy równą sumie wariancji)

$$\text{Var}(S_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n) = n\sigma^2.$$

Widzimy, że gdy  $n \rightarrow \infty$  to nie ma wielkich szans na zbieżność. Konieczna jest tzw. normalizacja, która polega na przeskalowaniu zmiennych losowych  $S_n$  tak aby miały takie same wartości oczekiwane i wariancje. Kładziemy

$$T_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}}.$$

Wtedy

$$\mathbb{E}(T_n) = \frac{1}{\sqrt{\text{Var}(S_n)}} \mathbb{E}(S_n - \mathbb{E}(S_n)) = 0$$

a dla wariancji mamy

$$\begin{aligned} \text{Var}(T_n) &= \frac{1}{\text{Var}(S_n)} \text{Var}(S_n - \mathbb{E}(S_n)) \quad \text{gdyż } \text{Var}(aX) = a^2 \text{Var}(X) \\ &= 1 \quad \text{ponieważ } \text{Var}(X + c) = \text{Var}(X) \end{aligned}$$

To że wartości oczekiwane i wariancje są równe oczywiście samo w sobie nie gwarantuje zbieżności ale przy założonej niezależności mamy następujący fakt (centralne twierdzenie graniczne).

### Twierdzenie 7.6 (Lindeberg–Feller)

Jeśli  $X_n$  są niezależne, o tych samych rozkładach, ze skończonymi wartościami oczekiwanymi i wariancjami, to ciąg  $T_n$  jest zbieżny do zmiennej losowej  $Z$  o standardowym rozkładzie normalnym w następującym sensie

$$P(a < T_n < b) \rightarrow P(a < Z < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx.$$

Można pokazać, że zastępując zmienne  $X_n$  przez inne zmienne losowe  $X'_n$ , o takich samych rozkładach ale zdefiniowane na innym zbiorze wyników  $\Omega'$ , mamy zbieżność odpowiadających  $T'_n$  prawie wszędzie. W dodatku, jako tą nową przestrzeń można wziąć przedział  $[0, 1]$  z miarą Lebesgue'a.

Jako wniosek mamy następujący szczególny przypadek, który wykorzystujemy do badania drzew dwumianowych

### Twierdzenie 7.7 (de Moivre-Laplace)

Niech  $X_n$  będą niezależne i takie, że  $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$ . Wtedy

$$P(a < T_n < b) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx.$$

Przedstawimy teraz ilustrację graficzną tego twierdzenia. Mamy

$$T_n = \frac{S_n}{\sqrt{n}}$$

a wartości  $S_n$  to liczby  $\{-n, -n+2, \dots, n-2, n\}$  i jest ich  $n+1$ . Wprowadźmy pomocniczą zmienną losową, która pokazuje liczbę kroków w górę,  $G \in \{0, 1, 2, \dots, n\}$  z prawdopodobieństwami

$$P(G = i) = \binom{n}{i} 0,5^n$$

więc

$$P(S_n = -n + 2i) = P(G = i).$$

Wartości znormalizowanych zmiennych  $T_n$  w tej konkretnej sytuacji przyjmują wartości

$$\left\{-\frac{n}{\sqrt{n}}, -\frac{n}{\sqrt{n}} + \frac{2}{\sqrt{n}}, \dots, \frac{n}{\sqrt{n}} - \frac{2}{\sqrt{n}}, \frac{n}{\sqrt{n}}\right\}.$$

Niech na przykład  $n = 100$  wtedy mamy wartości

$$\{-10, -9.8, -9.6, \dots, -0.2, 0, 0.2, 0.4, \dots, 9.8, 10\}$$

Niech na przykład  $a = 0.1$ ,  $b = 0.3$ . Twierdzenie graniczne mówi, że w przybliżeniu

$$P(0.1 < T_n < 0.3) = P(T_n = 0.2) = \int_{0.1}^{0.3} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx.$$

Z lewej strony mamy liczbę

$$P(T_n = 0.2) = P(N = 51) = \binom{100}{51} 0,5^{100} = 0.0780287$$



Z prawej strony możemy użyć dystrybuanty standardowego rozkładu normalnego

$$\frac{1}{\sqrt{2\pi}} \int_{0.1}^{0.3} e^{-\frac{1}{2}x^2} dx = N(0.3) - N(0.1) = 0.0780836$$

Aby porównać wykres gęstości rozkładu normalnego z prawdopodobieństwami danymi przez rozkład dwumianowy zapisujemy lewą stronę równości

$$P(G = 51) = \int_{0.1}^{0.3} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

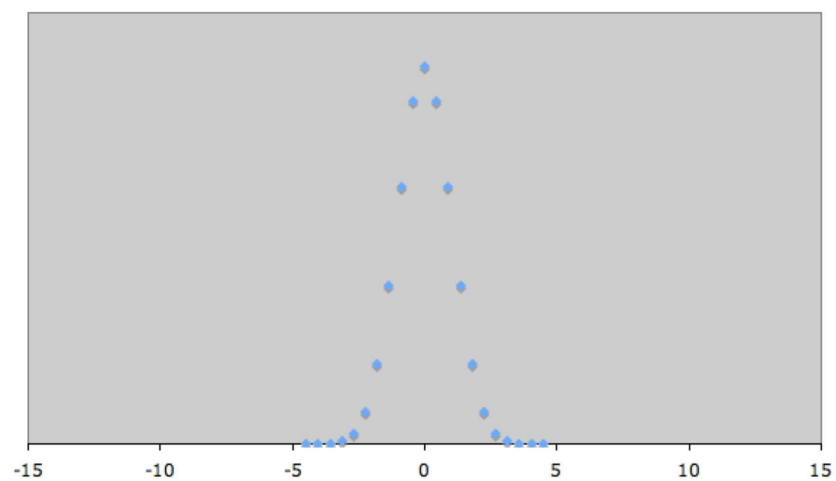
jako całkę po przedziale  $[0.1, 0.3]$  z funkcji stałej

$$\text{lewa strona} = \int_{0.1}^{0.3} \frac{P(G = 51)}{0.3 - 0.1} dx$$

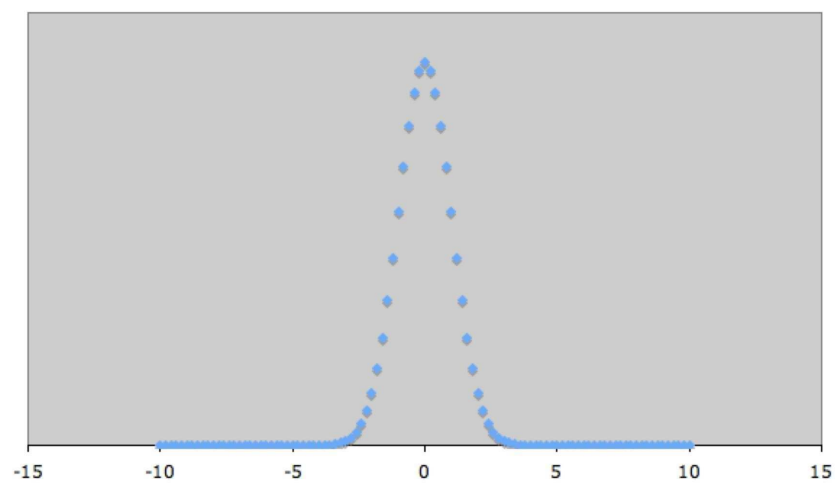
Porównujemy funkcje podcałkowe na obu stronach:

$$\begin{aligned} \text{lewa strona} &: \frac{P(G = 51)}{0.3 - 0.1} = 0.390143 \\ \text{prawa strona} &: \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}0.2^2} = 0.407001 \end{aligned}$$

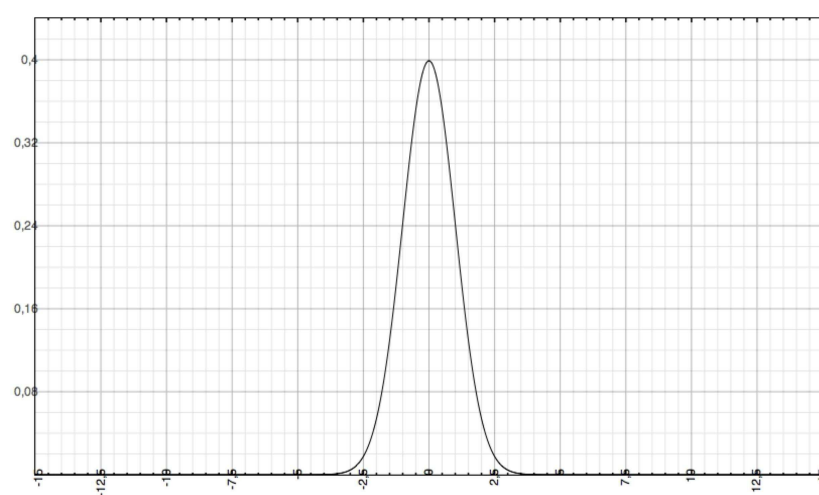
Ogólnie dzielimy przedział  $[-n, n]$  na odcinki długości  $\frac{2}{\sqrt{n}}$ , na każdym obliczamy  $P(G = i)$ , dzielimy przez długość odcinka i szkicujemy otrzymaną funkcję ograniczając się do pojedynczych kropek. Na Rysunkach 7.1 i 7.2 mamy wykresy dla  $n = 20$  i  $n = 100$  do porównania z wykresem gęstości standardowego rozkładu normalnego na Rysunku 7.3.



Rysunek 7.1  $n = 20$



Rysunek 7.2  $n = 100$



**Rysunek 7.3**      Gęstość standardowego rozkładu normalnego



# 8

## *Proces Wienera i Poissona*

### 8.1 Skalowane błędzenie przypadkowe

Naszym kolejnym celem jest konstrukcja procesu Wienera, który jest ogólnie przyjętym narzędziem do modelowania cen akcji, walut i innych walorów, których ceny podlegają losowym fluktuacjom. Z uwagi na wagę tego pojęcia podamy dwie metody konstrukcji. Trudność polega na tym, że proces Wienera jest złożonym obiektem który trudno opisać bez zaawansowanych narzędzi matematycznych. Ponieważ nimi nie dysponujemy, musimy poprzestać na pewnych aproksymacjach i przyjąć własności procesu Wienera na wiarę opierając się na wyrobionych intuicjach dla prostszych przypadków.

Pierwsza konstrukcja polega na zagęszczaniu drzewa dwumianowego w tym sensie, że długość kroku maleje do zera i w związku z tym w danym okresie czasu rośnie liczba kroków.

Wygodnie założyć, że mamy do dyspozycji nieskończony ciąg  $X_1, X_2, X_3, \dots$  niezależnych zmiennych losowych o tym samym rozkładzie postaci

$$X_n = \begin{cases} +1 & \text{z prawdopodobieństwem } 1/2, \\ -1 & \text{z prawdopodobieństwem } 1/2. \end{cases}$$

Wprawdzie zakładając skończony horyzont czasowy przy danej liczbie kroków potrzebujemy tylko skończonego takiego ciągu, niemniej liczba ta ulega zmianie przy zmniejszaniu długości kroku i lepiej mieć do dyspozycji ciąg nieskończony.

Pojawia się problem wyboru przestrzeni probabilistycznej. Aby móc skorzystać z nabytego doświadczenia wygodną przestrzenią podstawową jest  $\{D, U\}$

albo wręcz  $\{-1, +1\}$  dla pojedynczej zmiennej losowej, a następnie iloczyn kartezjański  $\{-1, +1\}^N$  dla niezależnych  $N$  zmiennych. Dla ciągu musimy wziąć iloczyn nieskończony  $\{-1, +1\}^{\mathbb{N}}$  czyli zbiór ciągów nieskończonych składających się z liczb  $-1, +1$ . Z uwagi na to że ten wybór powoduje komplikacje matematyczne związane z wprowadzeniem prawdopodobieństwa, wybierzmy inna przestrzeń probabilistyczną, mianowicie przedział  $[0, 1]$  z miarą Lebesgue'a. Musimy teraz na niej skonstruować zmienne losowe  $X_n : [0, 1] \rightarrow \mathbb{R}$ .

Kładziemy

$$\begin{aligned} X_1^\omega &= \begin{cases} +1 & \text{jeśli } \omega \in [0, \frac{1}{2}) \\ -1 & \text{w przeciwnym przypadku.} \end{cases} \\ X_2^\omega &= \begin{cases} +1 & \text{jeśli } \omega \in [0, \frac{1}{2^2}) \cup [\frac{2}{2^2}, \frac{3}{2^2}) \\ -1 & \text{w przeciwnym przypadku.} \end{cases} \\ &\vdots \\ X_n^\omega &= \begin{cases} +1 & \text{jeśli } \omega \in [0, \frac{1}{2^n}) \cup [\frac{2}{2^n}, \frac{3}{2^n}) \cup \dots \cup [\frac{k}{2^n}, \frac{k+1}{2^n}) \cup \dots \text{ (parzyste } k), \\ -1 & \text{w przeciwnym przypadku.} \end{cases} \end{aligned}$$

W ten sposób liczba wybrana losowo z przedziału  $[0, 1]$  z jednostajnym prawdopodobieństwem wyznacza cały ciąg. Oczywiście jest to idealizacja, gdyż tą liczbę znamy tylko z pewną dokładnością. Niemniej w praktyce możemy mieć tak długi ciąg jak potrzebujemy. Jeśli komputer daje nam 16 miejsc po przecinku, a potrzebujemy 32 to intuicja trafnie podpowiada, że wystarczy wylosować dwie liczby i zestawć ich cyfry po przecinku.

Żeby móc sprawdzić, że powyższe zmienne losowe są niezależne potrzebujemy stosownej wersji definicji niezależności zmiennych losowych dostosowanej do ciągów.

### Definicja 8.1

Zmienne losowe  $X_n$ ,  $n = 1, 2, \dots$  są niezależne jeśli dla każdego  $N$  zmienne  $X_1, \dots, X_N$  są niezależne.

Dowód niezależności dla dwóch elementów naszego ciągu jest bardzo prosty, choć nieco żmudny

Pokażemy, że

$$P((X_1 \in A) \cap (X_2 \in B)) = P(X_1 \in A) \times P(X_2 \in B)$$

Rozpatrujemy różne przypadki zbiorów  $A$  i  $B$ .

1.  $A = \{-1, 1\}$ ,  $P(X_1 \in A) = 1$ ,  $B$  dowolny i wtedy

$$P((X_1 \in A) \cap (X_2 \in B)) = P([0, 1] \cap (X_2 \in B)) = P(X_2 \in B)$$

co daje niezależność zdarzeń.

2.  $B = \{-1, 1\}$ ,  $A$  dowolny i rozumiemy analogicznie.
3.  $A, B$  są jednoelementowe i również łatwo widać niezależność

$$P((X_1 \in A) \cap (X_2 \in B)) = \frac{1}{4}$$

$$P(X_1 \in A) = \frac{1}{2}$$

$$P(X_2 \in B) = \frac{1}{2}$$

4.  $A = \emptyset$  lub  $B = \emptyset$  co jest najprostszy przypadkiem:

$$P((X_1 \in A) \cap (X_2 \in B)) = 0$$

Dowód dla wielu zmiennych będzie równie prosty ale niewyobrażalnie już żmudny więc go pomijamy.

Niech  $\mathcal{P}_n$  będzie podziałem generowanym przez  $X_1, \dots, X_n$ . Widać, że

$$\mathcal{P}_n = \{[0, \frac{1}{2^n}), [\frac{1}{2^n}, \frac{2}{2^n}), \dots, [\frac{k}{2^n}, \frac{k+1}{2^n}), \dots\}$$

a  $\mathcal{F}_n$  niech będzie sigma ciałem generowanym przez  $\mathcal{P}_n$ . Jeśli  $A \in \mathcal{P}_n$ , to dla wszystkich  $\omega \in A$  mamy identyczne wartości dla  $X_i^\omega$ ,  $i \leq n$ , czyli  $X_i$ ,  $i \leq n$ , są stałe na  $A \in \mathcal{P}_n$ .

Definiujemy nowy ciąg

$$Y_0 = 0$$

$$Y_n = \sum_{i=1}^n X_i.$$

i nazywamy go symetrycznym błądzeniem przypadkowym. Odnotujmy proste własności tego ciągu

### Twierdzenie 8.1

Przyrosty  $Y_k - Y_j, Y_m - Y_n$  są niezależne ( $j < k \leq n < m$ )

### Dowód

Z definicji ciągu  $Y_n$  mamy

$$Y_k - Y_j = \sum_{i=j+1}^k X_i,$$

$$Y_m - Y_n = \sum_{i=n+1}^m X_i$$

a wektory  $(X_{j+1}, \dots, X_k)$  i  $(X_{n+1}, \dots, X_m)$  są niezależne.  $\square$

### Twierdzenie 8.2

$$\mathbb{E}(Y_n) = 0.$$

#### Dowód

Mamy  $\mathbb{E}(X_i) = 0$ , a wartość oczekiwana sumy to suma wartości oczekiwanych.  $\square$

### Twierdzenie 8.3

$$\text{Var}(Y_n) = \sum_{i=1}^n \text{Var}(X_i) = n$$

#### Dowód

Jest oczywiste, że  $\text{Var}(X_i) = 1$  (bo  $X_i^2 = 1$ ), a wariancja sumy niezależnych zmiennych losowych to suma ich wariancji.  $\square$

### Twierdzenie 8.4

$Y_n$  jest martyngałem względem  $\mathcal{P}_n$ .

#### Dowód

Sprawdzamy definicję martyngału

$$\begin{aligned} \mathbb{E}(Y_{n+1}|\mathcal{P}_n) &= \mathbb{E}\left(\sum_{i=1}^{n+1} X_i|\mathcal{P}_n\right) \quad \text{z definicji } Y_{n+1} \\ &= \sum_{i=1}^{n+1} \mathbb{E}(X_i|\mathcal{P}_n) \quad \text{z liniowości.} \end{aligned}$$

Dla  $i \leq n$ , zmienne  $X_i$  są mierzalne względem  $\mathcal{P}_n$  więc  $\mathbb{E}(X_i|\mathcal{P}_n) = X_i$ .

Dla  $n+1$ , zmienna  $X_{n+1}$  jest niezależna od  $\mathcal{P}_n$ , więc  $\mathbb{E}(X_{n+1}|\mathcal{P}_n) = \mathbb{E}(X_{n+1})$ . Te fakty dają

$$\begin{aligned} \mathbb{E}(Y_{n+1}|\mathcal{P}_n) &= \sum_{i=1}^n X_i + \mathbb{E}(X_{n+1}) \\ &= Y_n \end{aligned}$$



gdyż  $\mathbb{E}(X_{n+1}) = 0$ . □

Dla ustalonego  $N$  kładziemy  $h = \frac{1}{N}$  i definiujemy losową funkcję

$$W_N^\omega : [0, \infty) \rightarrow \mathbb{R}$$

dla punktów postaci  $nh$  daną przez

$$W_N(0) = 0,$$

$$W_N(h) = \sqrt{h}X_1 = \begin{cases} \sqrt{h} \text{ z prawdopodobieństwem } \frac{1}{2} \\ -\sqrt{h} \text{ z prawdopodobieństwem } \frac{1}{2} \end{cases}$$

$$W_N(2h) = \sqrt{h}(X_1 + X_2) = \begin{cases} 2\sqrt{h} \text{ z prawdopodobieństwem } \frac{1}{4} \\ 0 \text{ z prawdopodobieństwem } \frac{1}{2} \\ -2\sqrt{h} \text{ z prawdopodobieństwem } \frac{1}{4} \end{cases}$$

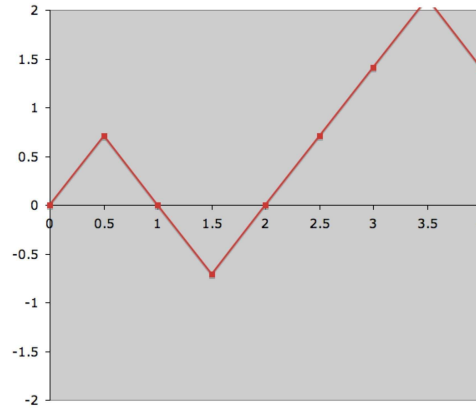
...

Ogólnie, kładziemy

$$W_N(nh) = \sqrt{h} \sum_{i=1}^n X_i = \sqrt{h}Y_n$$

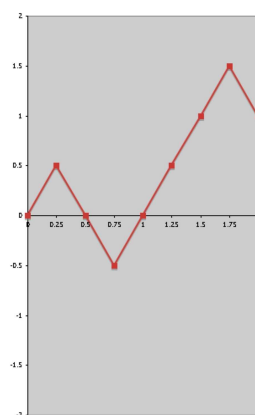
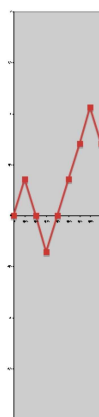
a poza punktami postaci  $nh$  uzupełniamy funkcję.

Wykresy dla  $N = 2, 4, 8$  dla  $x \in [0, 8h]$  są przedstawione na Rysunkach 8.1, 8.2, 8.3.



**Rysunek 8.1**  $n = 2$

Mamy tu wykresy odpowiadające pierwszemu 8 krokom. Kroki są coraz krótsze, natomiast kąt nachylenia odcinków rośnie. Wynika to stąd, że liczba  $\sqrt{h}$  jest większa niż  $h$  dla  $h < 1$ , a w pojedynczym odcinku przy zmianie argumentu o  $h$  mamy zmianę wartości o  $\sqrt{h}$ .

Rysunek 8.2  $n = 4$ Rysunek 8.3  $n = 8$ 

### Definicja 8.2

Ciąg  $W_N(nh)$ ,  $n = 0, 1, 2, \dots$  nazywamy skalowanym błędzeniem przypadkowym.

Ponieważ różni się on od  $Y_n$  tylko o stałą, ma podobne własności jak  $Y_n$ .

### Twierdzenie 8.5

1.  $W_N(s) - W_N(r)$ ,  $W_N(u) - W_N(t)$  dla  $s < r \leq t < u$  będących wielokrotnościami  $h$ , są niezależne.

2.  $\mathbb{E}(W_N(nh)) = 0$ .
3.  $\text{Var}(W_N(nh)) = \frac{n}{N}$ .
4. Ciąg  $W_N(nh)$ ,  $n = 0, 1, 2, 3, \dots$  jest martyngałem

### Dowód

1. Zmienne  $Y_k - Y_j, Y_m - Y_n$  są niezależne i to się utrzyma po pomnożeniu przez stałą.

$$2. \mathbb{E}(W_N(nh)) = \mathbb{E}(\sqrt{h}Y_n) = 0 \text{ gdyż } \mathbb{E}(Y_n) = 0$$

$$3. \text{Var}(W_N(nh)) = h\text{Var}(Y_n) = hn = \frac{n}{N}$$

4.  $Y_n$  jest martyngałem więc

$$\begin{aligned} \mathbb{E}(W_N((n+1)h)|\mathcal{P}_n) &= \mathbb{E}(\sqrt{h}Y_{n+1}|\mathcal{P}_n) \\ &= \sqrt{h}\mathbb{E}(Y_{n+1}|\mathcal{P}_n) = \sqrt{h}Y_n. \end{aligned}$$

□

### Definicja 8.3

Definiujemy wahanie (lub wariację) ciągu  $\{W_N(nh)\}_n$  na przedziale  $[s, t]$ , gdzie  $s = \frac{m}{N}$ ,  $t = \frac{n}{N}$ , jako

$$V_{[s,t]}(W_N) = \sum_{i=m}^n |W_N((i+1)h) - W_N(ih)|.$$

### Twierdzenie 8.6

$$V_{[s,t]}(W_N) = \sqrt{N}(t - s)$$

### Dowód

Wartości bezwzględne kolejnych przyrostów są równe długości kroku. Musimy ją pomnożyć przez liczbę składników:

$$\begin{aligned} \sum_{i=m}^n |W_N((i+1)h) - W_N(ih)| &= \sqrt{h}(n - m) \\ &= \frac{N}{\sqrt{N}}(t - s) = \sqrt{N}(t - s) \quad \text{gdzie } h = \frac{1}{N} \end{aligned}$$

□

Jako wniosek mamy następujący fakt:

$$V_{[s,t]}(W_N) \rightarrow \infty \quad \text{gdy } N \rightarrow \infty.$$

### Definicja 8.4

Wahanie kwadratowe ciągu  $\{W_N(nh)\}_n$  na  $[s, t]$ ,  $s = \frac{m}{N}$ ,  $t = \frac{n}{N}$ , to liczba

$$V_{[s,t]}^2(W_N) = \sum_{i=m}^n (W_N((i+1)h) - W_N(ih))^2.$$

### Twierdzenie 8.7

$$V_{[s,t]}^2(W_N) = t - s.$$

### Dowód

Kwadraty przyrostów są równe jeden więc rachunek jest podobny do poprzedniego tylko z uwagi na kwadrat stała jest innego rzędu

$$\begin{aligned} \sum_{i=m}^n (W_N((i+1)h) - W_N(ih))^2 &= h(n-m) \\ &= \frac{n-m}{N} = t - s. \end{aligned}$$

□

W szczególności dla  $s = 0$  mamy

$$V_{[0,t]}^2(W_N) = t.$$

### Twierdzenie 8.8

Ciąg  $W_N^2(nh)$  jest submartyngałem i nie jest martyngałem.

### Dowód

Korzystamy z definicji ciągu  $W_N(nh)$  i mamy

$$W_N^2((n+1)h) = h \left( \sum_{i=1}^{n+1} X_i \right)^2 = h \sum_{i,j=1}^{n+1} X_i X_j.$$

Liczmy wartość warunkową lewej strony

$$\begin{aligned}\mathbb{E}(W_N^2((n+1)h)|\mathcal{P}_n) &= \mathbb{E}(h \sum_{i,j=1}^{n+1} X_i X_j | \mathcal{P}_n) \\ &= h \mathbb{E}(\sum_{i,j=1}^n X_i X_j + \sum_{i=1}^n X_i X_{n+1} + X_{n+1}^2 | \mathcal{P}_n) \\ &\text{rozbijamy sumę} \\ &= h \sum_{i,j=1}^n X_i X_j + h \sum_{i=1}^n X_i \mathbb{E}(X_{n+1}) + h \mathbb{E}(X_{n+1}^2 | \mathcal{P}_n)\end{aligned}$$

mierzalne zmienne losowe wyłączamy przed warunkową wartość oczekiwaną

$$\begin{aligned}&= h \left( \sum_{i=1}^n X_i \right)^2 + h \mathbb{E}(1 | \mathcal{P}_n) \quad \text{bo } \mathbb{E}(X_{n+1}) = 0 \text{ a } X_{n+1}^2 = 1 \\ &> W_N^2(nh) \quad \text{bo } \mathbb{E}(1 | \mathcal{P}_n) = 1.\end{aligned}$$

Dostaliśmy pożądaną nierówność i to silną, równość nie zachodzi, więc udowodniliśmy oba fakty.  $\square$

Okazuje się, że jeśli od  $W_N^2(nh)$  odjąć wahanie kadratowe, to dostajemy martyngał. To twierdzenie wiąże się z rozkładem Dooba-Meyera omawianym poprzednio, gdyż tam również odejmowaliśmy coś od submartyngału i w wyniku dostawaliśmy martyngał. Tutaj mamy specjalną postać ciągu  $A_n$  (zwanego kompensatorem) a mianowicie jest to ciąg  $nh$  który jest nielosowy.

Dowód jest wart przestudiowania gdyż stosuje się w nim klasyczne argumenty wykorzystywane w wielu sytuacjach, w istocie już je wszystkie znamy.

To twierdzenie jest kluczowym faktem związanym z tzw. wzorem Ito, który jest z kolei bazą słynnego wzoru Blacka-Scholesa.

### Twierdzenie 8.9

Ciąg  $W_N^2(t) - t$ ,  $t = nh$ , jest martyngałem.

#### Dowód

Weźmy  $s < t$ , oba wielokrotności liczby  $h$ ,  $s = kh$ ,  $t = mh$ ,  $k < m$ .

Wprowadźmy oznaczenie:  $\mathcal{P}_s = \mathcal{P}_k$  które ułatwi dowód (optycznie). Mamy pokazać, że

$$\mathbb{E}(W_N^2(t) - t | \mathcal{P}_s) = W_N^2(s) - s$$

Wiadomo, że:

$$\begin{aligned}\mathbb{E}(W_N^2(s)|\mathcal{P}_s) &= W_N^2(s) \\ \mathbb{E}(t|\mathcal{P}_s) &= t\end{aligned}$$

Mamy więc pokazać, że

$$\mathbb{E}(W_N^2(t) - W_N^2(s)|\mathcal{P}_s) = t - s \quad (*)$$

Prosta algebra (wzory skróconego mnożenia) pokazuje że

$$W_N^2(t) - W_N^2(s) = (W_N(t) - W_N(s))^2 + 2W_N(s)(W_N(t) - W_N(s))$$

gdyż prawa strona równa się

$$W_N^2(t) - 2W_N(t)W_N(s) + W_N^2(s) + 2W_N(s)W_N(t) - 2W_N^2(s)$$

Podstawiamy wyrażenie na  $W_N^2(t) - W_N^2(s)$  do (\*)

$$\begin{aligned}\mathbb{E}(W_N^2(t) - W_N^2(s)|\mathcal{P}_s) &= \mathbb{E}((W_N(t) - W_N(s))^2 + 2W_N(s)(W_N(t) - W_N(s))|\mathcal{P}_s) \\ &= \mathbb{E}((W_N(t) - W_N(s))^2|\mathcal{P}_s) \\ &\quad + \mathbb{E}(2W_N(s)(W_N(t) - W_N(s))|\mathcal{P}_s) \quad \text{z liniowości}\end{aligned}$$

Badamy pierwszy składnik. Niezależność  $W_N(t) - W_N(s)$  i  $\mathcal{P}_s$  daje

$$\mathbb{E}((W_N(t) - W_N(s))^2|\mathcal{P}_s) = \mathbb{E}((W_N(t) - W_N(s))^2) = t - s$$

czyli pierwszy składnik daje to co chcemy uzyskać.

W drugim składniku (który powinien więc być równy zero)  $W_N(s)$  jest mierzalne względem  $\mathcal{P}_s$  więc

$$\mathbb{E}(2W_N(s)(W_N(t) - W_N(s))|\mathcal{P}_s) = 2W_N(s)\mathbb{E}(W_N(t) - W_N(s)|\mathcal{P}_s)$$

Wiemy ponadto, że  $W_N$  jest martyngałem więc

$$\mathbb{E}(W_N(t) - W_N(s)|\mathcal{P}_s) = 0.$$

□

## 8.2 Definicja i własności procesu Wienera

Następny krok to przejście graniczne  $h \rightarrow 0$ . Badamy je dla ustalonej chwili  $t$  kładąc  $h = \frac{t}{n}$ , i rozważając ciąg skalowanych błędów przypadkowych  $W_n(t) = \sqrt{h} \sum_{i=1}^n X_i$ .

Centralne twierdzenie graniczne, jak wiemy, daje zbieżność

$$\lim_{n \rightarrow \infty} P(a < T_n < b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx$$

gdzie  $Y_n = X_1 + \dots + X_n$ ,

$$T_n = \frac{Y_n - \mathbb{E}(Y_n)}{\sqrt{\text{Var}(Y_n)}}.$$

Tutaj mamy

$$W_n(t) = \sqrt{h}Y_n = \sqrt{t} \frac{Y_n}{\sqrt{n}} = \sqrt{t}T_n$$

czyli dla  $c < d$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} P(c < W_n(t) < d) &= \lim_{n \rightarrow \infty} P\left(\frac{c}{\sqrt{t}} < T_n < \frac{d}{\sqrt{t}}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\frac{c}{\sqrt{t}}}^{\frac{d}{\sqrt{t}}} e^{-\frac{y^2}{2}} dy \quad \text{centralne twierdzenie graniczne} \\ &= \frac{1}{\sqrt{2\pi t}} \int_c^d e^{-\frac{z^2}{2t}} dz \end{aligned}$$

po zmianie zmiennych  $y = \frac{z}{\sqrt{t}}$ ,  $dy = \frac{1}{\sqrt{t}} dz$ ; a jeśli  $y = \frac{c}{\sqrt{t}}, \frac{d}{\sqrt{t}}$  to  $z = c, d$ , co daje granice całkowania.

Widzimy, że w granicy mamy rozkład normalny z wartością oczekiwaną zero i wariancją  $t$ . Otrzymaną w granicy zmienną losową oznaczamy przez  $W(t)$ .

Aby nabrać intuicji pokażemy wykres ciągu pozycji błędzenia przypadkowego dla  $N = 1000$ . Bierzymy  $h = \frac{1}{1000}$ , losujemy 1000 razy  $\omega \in [0, 1]$

$$X_n = \begin{cases} 1 & \text{gdy } \omega < 0.5 \\ -1 & \text{gdy } \omega \geq 0.5 \end{cases}, \quad n = 1, 2, \dots, 1000,$$

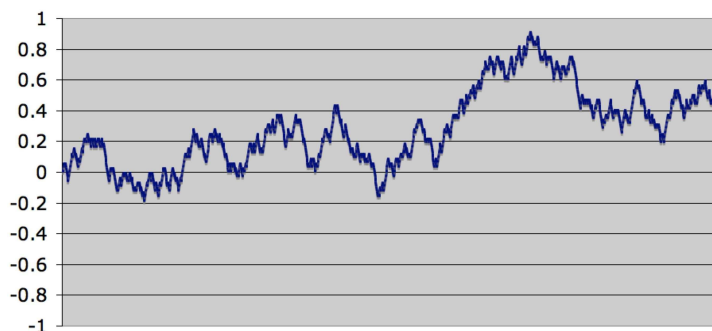
i kolejno dodajemy te wartości budując ciąg  $W_{1000}$

$$W_{1000}((n+1)h) = W_{1000}(nh) + \sqrt{h}X_n$$

a wykres widzimy na Rysunku 8.4.

Powyższe przejście graniczne przeprowadzamy dla każdego  $t \in [0, \infty)$  co daje proces stochastyczny, który jest funkcją dwóch zmiennych

$$W : \Omega \times [0, \infty) \rightarrow \mathbb{R}$$



Rysunek 8.4

gdzie dla każdego  $t \geq 0$  odwzorowanie  $W(t) : \Omega \rightarrow \mathbb{R}$  jest zmienna losową. Własności które pokazaliśmy dla błędzenia przypadkowego bierzemy jako postulaty dla otrzymanego procesu.

Powyższy argument graniczny nie pozwala na proste przeniesienie ich z  $W_N$  na  $W$ . Traktujemy te własności jako motywację definicji odkładając problem konstrukcji procesu  $W$ , która nie jest prosta. Zauważmy tylko, że powyższa zbieżność była badana dla poszczególnych chwil z osobna i nie mówi ona nam wiele o zbieżności całych ścieżek błędzenia, czyli zbieżności jednocznej dla wszystkich chwil.

### Definicja 8.5

Proces Wienera (proces ruchu Browna)  $W(t)$ ,  $t \geq 0$ , to proces o własnościach

1.  $W(0) = 0$ ,
2.  $W(t) - W(s)$  ma rozkład normalny z wartością średnią 0 i wariancją  $t - s$  czyli odchyleniem standardowym  $\sqrt{t - s}$ ,
3. Dla każdego  $m$  i wszystkich chwil  $0 \leq t_1 < t_2 < \dots < t_m$  przyrosty  $W(t_{n+1}) - W(t_n)$ ,  $n = 1, \dots, m - 1$  są niezależne.
4. Ścieżki, czyli funkcje  $t \mapsto W^\omega(t)$  są ciągłe dla prawie wszystkich  $\omega$ .

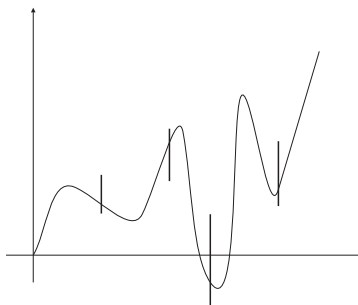
Problem istnienia takiego procesu jest nietrywialny i obejmuje kwestię konstrukcji stosownej przestrzeni probabilistycznej. Jak się domyślamy, może to być zrobione na wiele sposobów. Odnosimy skrótem jeden z nich, w pewnym sensie naturalny ale wymagający bardzo zaawansowanej matematyki. Bierzemy  $\Omega = \mathbb{R}^{[0, \infty)}$  czyli zbiór wszystkich funkcji określonych na  $[0, \infty)$  o wartościach rzeczywistych i kładziemy  $W^\omega(t) = \omega(t)$ . Sigma ciało  $\mathcal{F}$  jest generowane przez



tz. zbiory cylindryczne

$$\begin{aligned} C &= \{\omega : (\omega(t_1), \dots, \omega(t_n)) \in B = [a_1, b_1] \times \dots \times [a_n, b_n]\} \\ &= \{\omega : \omega(t_i) \in [a_i, b_i], i = 1, \dots, n\} \end{aligned}$$

dla wszelkich  $n$  i przedziałów  $[a_i, b_i]$ . Rysunek 8.5 pokazuje wykres funkcji która należy do cylindra z  $n = 4$  co poglądowo oznacza przejazd narciarza przez bramki slalomu.



**Rysunek 8.5** Funkcja należąca do zbioru cylindrycznego/

Prawdopodobieństwo jest określone na cylindrach tak, aby mieć rozkład normalny o żądanych parametrach

$$P(C) = \int_B p(t_1, 0, x_1) p(t_2 - t_1, x_1, x_2) \cdots p(t_n - t_{n-1}, x_{n-1}, x_n) dx_1 \dots dx_n$$

gdzie

$$p(t, x, y) = \frac{1}{\sqrt{(2\pi t)^n}} \exp\left\{-\frac{(x - y)^2}{2t}\right\}.$$

Rozszerzenie  $P$  do całego sigma ciała nie jest proste. Również pokazanie, że prawdopodobieństwo  $P$  jest skoncentrowane na funkcjach ciągłych (warunek 4 definicji) jest trudne.

Później pokażemy alternatywną konstrukcję która jest łatwiejsza matematycznie choć trochę trudniejsza poglądowo.

Mając proces  $W$  definiujemy rodzinę sigma ciał podobnie jak w przypadku dyskretnym:  $\mathcal{F}_t$  to najmniejsze sigma ciało takie, że wszystkie zmienne losowe  $W(s)$ ,  $s \leq t$  są względem niego mierzalne. Ta rodzina spełnia warunek:  $\mathcal{F}_s \subset \mathcal{F}_t$  jeśli  $s \leq t$  i nazywamy ją tak jak dla przypadku dyskretnego, filtracją generowaną przez  $W$ .

### Twierdzenie 8.10

$W(t)$  jest martyngałem względem  $\mathcal{F}_t$ .

### Dowód

Ten dowód wygląda prosto, jest powtórzeniem przejść znanych nam z przypadku dyskretnego, ale w istocie to co prezentujemy należy traktować jedynie jako ideę dyż nie pokazaliśmy koniecznych faktów w wymaganej postaci.

Przyrosty procesu Wienera są niezależne więc  $W(t) - W(s)$  i  $\mathcal{F}_s$  (generowane przez  $W(s) - W(0)$ ), są niezależne dla  $s < t$  więc

$$\mathbb{E}(W(t) - W(s) | \mathcal{F}_s) = \mathbb{E}(W(t) - W(s)) = 0$$

Z definicji  $W(s)$  jest  $\mathcal{F}_s$  mierzalne więc  $\mathbb{E}(W(s) | \mathcal{F}_s) = W(s)$  co pociąga

$$0 = \mathbb{E}(W(t) - W(s) | \mathcal{F}_s) = \mathbb{E}(W(t) | \mathcal{F}_s) - W(s).$$

□

Proces Wienera, tak jak skalowane błędzenie przypadkowe po podniesieniu do kwadratu jest submartyngelem i nie jest martyngelem. Tego nie omawiamy w szczegółach, ale koncentrujemy się na ważniejszym fakcie w kontekście zastosowań (do wzoru Ito).

### Twierdzenie 8.11

$W^2(t) - t$  jest martyngelem względem  $\mathcal{F}_t$

### Dowód

Jest to znów tylko idea gdyż wymaga własności warunkowych wartości oczekiwanych względem filtracji w czasie ciągłym których formalnie nie udowodniliśmy, wymagają one nieco pracy. Najpierw proste przekształcenie algebraiczne daje

$$W^2(t) - W^2(s) = (W(t) - W(s))^2 + 2W(s)(W(t) - W(s))$$

Podobnie jak poprzednio liczymy warunkową wartość oczekiwaną różnicy kwadratów a to co jest uderzające to że jest ona równa wartości oczekiwanej kwa-

dratu różnicy

$$\begin{aligned}
 & \mathbb{E}(W^2(t) - W^2(s) | \mathcal{F}_s) \\
 &= \mathbb{E}((W(t) - W(s))^2 + 2W(s)(W(t) - W(s)) | \mathcal{F}_s) \\
 &= \mathbb{E}((W(t) - W(s))^2 | \mathcal{F}_s) \\
 &\quad + \mathbb{E}(2W(s)(W(t) - W(s)) | \mathcal{F}_s) \quad \text{liniowość} \\
 &= \mathbb{E}((W(t) - W(s))^2 | \mathcal{F}_s) \\
 &\quad + 2W(s)\mathbb{E}(W(t) - W(s) | \mathcal{F}_s)
 \end{aligned}$$

mierzalne zmienne losowe wyłączamy przed warunkową wartość oczekiwaną

$$= \mathbb{E}((W(t) - W(s))^2) + \mathbb{E}(W(t) - W(s))$$

korzystając z niezależności która daje wartość oczekiwaną

$$= \mathbb{E}((W(t) - W(s))^2) \quad \text{bo } \mathbb{E}(W(t) - W(s)) = 0$$

$$= t - s$$

co wynika z definicji procesu Wienera, a konkretnie z faktu że przyrost ma taką właśnie wariancję  $\square$

Słynne twierdzenie Ley'ego mówi, że powyższe dwie własności w istocie charakteryzują proces Wienera.

### Twierdzenie 8.12 (Levy)

Jeśli  $M(t)$  jest martyngałem o ciągłych ścieżkach,  $M(0) = 0$  oraz  $M^2(t) - t$  jest martyngałem, to  $M$  jest procesem Wienera.

Poniższe twierdzenie ma zastosowanie w wycenie opcji.

### Twierdzenie 8.13

$\exp\{\sigma W(t) - \frac{1}{2}\sigma^2 t\}$  jest martyngałem.

### Dowód

(Idea.) Wiadomo, że jeśli  $Y$  ma rozkład normalny ze średnią 0 i wariancją  $\sigma^2$  to

$$\mathbb{E}(e^Y) = e^{\frac{1}{2}\sigma^2}.$$

Badamy definicję martyngału:

$$\begin{aligned} & \mathbb{E}(\exp\{\sigma W(t) - \frac{1}{2}\sigma^2 t\} | \mathcal{F}_s) \\ &= \mathbb{E}(\exp\{\sigma W(t) - \frac{1}{2}\sigma^2 t \pm \sigma W(s)\} | \mathcal{F}_s) \\ &= \exp\{\sigma W(s) - \frac{1}{2}\sigma^2 s\} \mathbb{E}(\exp\{\sigma(W(t) - W(s))\} | \mathcal{F}_s) \end{aligned}$$

wyłączając mierzalne zmienne losowe przed warunkową wartość oczekiwaną

$$\begin{aligned} &= \exp\{\sigma W(s) - \frac{1}{2}\sigma^2 s\} \mathbb{E}(\exp\{\sigma(W(t) - W(s))\}) \quad \text{niezależność} \\ &= \exp\{\sigma W(s) - \frac{1}{2}\sigma^2 s\} \exp\{\frac{1}{2}\sigma^2(t-s)\} \end{aligned}$$

gdź  $\sigma(W(t) - W(s))$  jest taki jak wspomniany na początku  $Y$

$$= \exp\{\sigma W(s) - \frac{1}{2}\sigma^2 s\}.$$

□

Ścieżki procesu Wienera są nieregularne na co wskazuje fakt, że skalowane błędzenia przypadkowe mają coraz stromsze odcinki w miarę zagęszczania kroków czasowych. Kąt nachylenia tych odcinków to pochodna, więc nie jest zaskakujące, że prawie wszystkie ścieżki procesu Wienera są nieróżniczkowalne i to w żadnym punkcie pochodna nie istnieje. Precyzyjny dowód jest trudny, a łatwo możemy tylko podać pewne częściowe wyjaśnienie: badamy iloraz różnicowy i okazuje się że jest on rozbieżny do nieskończoności w sensie  $L^2$

$$\mathbb{E}\left(\frac{W(t+h) - W(t)}{h}\right)^2 = \frac{1}{h^2} \mathbb{E}((W(t+h) - W(t))^2) = \frac{1}{h^2} h \rightarrow \infty.$$

Ta nieregularność manifestuje się również w pojęciu wahan. Jak pamiętamy, wahanie skalowanego błędzenia przypadkowego zmierzało do nieskończoności wraz z zagęszczaniem kroków. Dla procesu Wienera wahanie to jest zdefiniowane jako

$$\lim \sum |W(t_{i+1}) - W(t_i)|$$

gdzie mamy podział odcinka  $[0, t]$  na mniejsze  $0 = t_0 < t_1 < \dots < t_n = t$  i granica jest wzięta przy założeniu że te podziały się zagęszczają i  $\max\{t_{i+1} - t_i\} \rightarrow 0$ . Można pokazać, że ta granica jest nieskończona.

Wahanie kwadratowe dla procesu Wienera jest takie jak dla błędzenia przypadkowego, mianowicie definiujemy je jako

$$\lim \sum (X(t_{i+1}) - X(t_i))^2$$

przy zagęszczających się podziałach jak poprzednio i dostajemy w granicy  $t$ . Te dwie własności charakteryzują funkcje nieregularne. Jeśli funkcja jest różniczkowalna to ma wahanie kwadratowe zero a wahanie skończone.

### 8.3 Konstrukcja procesu Wienera

Pokażemy pewną konstrukcję procesu Wienera, alternatywną do skalowanego błędzenia przypadkowego. Jej zaleta polega na tym że zależność  $W$  od  $t$  i od  $\omega$  jest rozdzielona i łatwiej wskazać przestrzeń probabilistyczną.

Z punktu widzenia  $W(t)$  jako zmiennej losowej konstrukcja jest oparta na wzięciu ciągu zmiennych losowych  $G_n$  niezależnych o standardowym rozkładzie normalnym. Symulacja komputerowa skończonej liczby takich zmiennych losowych jest łatwa i będzie przedstawiona w innej części kursu. Opiera się ona na tym, że te zmienne losowe można zbudować na naszej specjalnej przestrzeni probabilistycznej:  $\Omega = [0, 1]$  z miarą Lebesgue'a.

Zależność od czasu jest wyznaczona przez następujący ciąg funkcji  $\psi_n(t)$  (tzw. falek). Ograniczamy się do  $t \in [0, 1]$ , a potem pokażemy jak rozszerzyć konstrukcję na dłuższe odcinki czasowe.

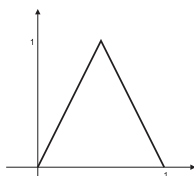
Najpierw dla  $n = 0$  kładziemy

$$\psi_0(t) = t.$$

Dalej definiujemy falekę podstawową, która wyznaczy kolejne

$$\psi_1(t) = \begin{cases} 2t & \text{dla } 0 \leq t < \frac{1}{2} \\ 2 - 2t & \text{dla } \frac{1}{2} \leq t < 1 \end{cases}$$

a jej wykres jest pokazany na Rysunku 8.6.



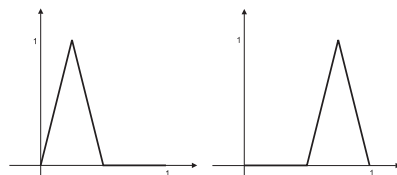
Rysunek 8.6

Ogólnie niech  $n = 2^j + k$  i bierzemy najpierw kolejne  $j$  a następnie  $k$  w zakresie  $0 \leq k < 2^j$  oraz definiujemy

$$\psi_n(t) = \psi_1(2^j t - k) \text{ dla } n = 2^j + k, \quad 0 \leq k < 2^j$$

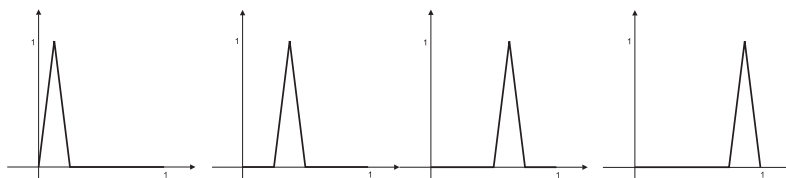
Czyli powyższa pierwsza funkcja odpowiada  $j = 0, k = 0, n = 1$ .

Następnie bierzemy  $j = 1$ , wtedy  $k = 0, 1, n = 2, 3$  i mamy wykresy na Rysunku 8.7



**Rysunek 8.7**

Dla  $j = 2, k = 0, 1, 2, 3, 4$ , a  $n = 4, 5, 6, 7$  co daje kolejne wykresy pokazane na Rysunku 8.8.



**Rysunek 8.8**

Ten schemat powtarza się dla dalszych  $n$ . mamy "wędrujące" fale o coraz mniejszej podstawie.

Można udowodnić, następujące twierdzenie. Dowód jest żmudny ale elementarny, dostępny dla osób bez wielkiego przygotowania matematycznego.

#### Twierdzenie 8.14

Niech  $G_n$  będą niezależne o standardowym rozkładzie normalnym, oraz weźmy ciąg

$$a_0 = 1$$

$$a_n = 2^{-(\frac{j}{2}+1)}$$

dla  $n = 2^j + k, k = 0, 1, \dots, 2^j - 1$ . Proces zdefiniowany wzorem

$$W^\omega(t) = \sum_{n=0}^{\infty} a_n \psi_n(t) G_n^\omega$$

jest procesem Wienera dla  $t \in [0, 1]$ .

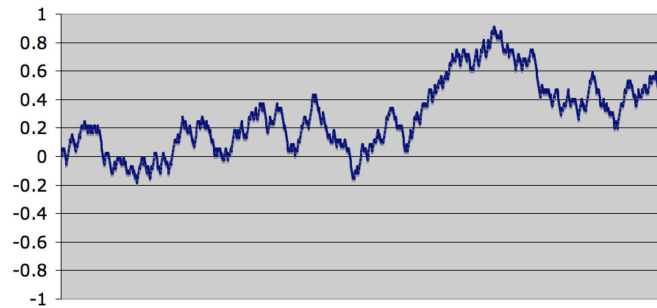
Mając proces dla  $t \in [0, 1]$  rozszerzamy go na przedział  $[0, 2]$ . Bierzemy dwie niezależne kopie powyższego procesu i oznaczamy je przez  $W_1, W_2$ . Wzór

$$W(t) = \begin{cases} W_1(t) & \text{dla } t \in [0, 1), \\ W_1(1) + W_2(t - 1) & \text{dla } t \in [1, 2) \end{cases}$$

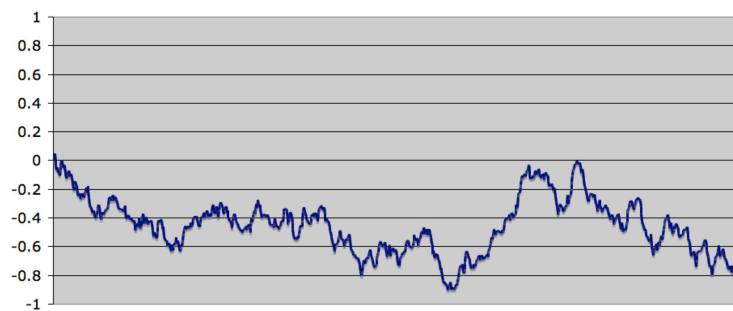
określa proces Wienera na  $[0, 2]$ . Stosując to wielokrotnie możemy uzyskać proces na przedziale o dowolnej długości.

Porównajmy jeszcze podany już wcześniej wykres błędzenia przypadkowego w 1000 krokach (Rysunek 8.9) z wykresem danym przez powyższą konstrukcję dla  $n = 1024$  (Rysunek 8.10). 261

Wykres powstały z falek jest bardziej urozmaicony. Jest on zbudowany z odcinków, tak jak przy błędzeniu, ale w falkach stromizna może wynika z mnożenia przez liczby o rozkładzie normalnym które mogą być bardzo duże (ale również małe).



Rysunek 8.9



Rysunek 8.10

## 8.4 Process Poissona

Proces Wienera ma ciągle trajektorie zgodnie z ostatnim punktem definicji. Oznacza to, że ewolucja w czasie wartości procesu Wienera przebiega bez skoków. Można się zastanowić, czy z punktu widzenia modelowania zjawisk finansowych jest to model dobry gdyż pojawia się obawa, że nie uchwycimy gwałtownych zmian zdarzających się podczas kryzysów.

W obronie procesu Wienera można powiedzieć, że dopuszcza on wielkie zmiany w krótkim okresie czasu. Jeśli ustalimy chwilę czasową  $t$  to zmienna losowa, którą często przyjmuje się jako model cen akcji, dana wzorem  $S(t) = S(0) \exp(\mu t + \sigma W(t))$  ma rozkład log-normalny. Łatwo odpowiedzieć na pytanie jakie jest prawdopodobieństwo, że cena spadnie na przykład o połowę, czyli

$$\begin{aligned} P(S(t) < \frac{1}{2}S(0)) &= P(S(0) \exp(\mu t + \sigma W(t)) < \frac{1}{2}S(0)) \\ &= P(\exp(\mu t + \sigma W(t)) < \frac{1}{2}) \\ &= P(\mu t + \sigma W(t) < \ln \frac{1}{2}) \\ &= P(\sigma W(t) < \ln \frac{1}{2} - \mu t) \\ &= P(X < \frac{1}{\sqrt{t}\sigma} (\ln \frac{1}{2} - \mu t)) \\ &= N(\frac{1}{\sqrt{t}\sigma} (\ln \frac{1}{2} - \mu t)), \end{aligned}$$

gdzie  $X$  ma standardowy rozkład normalny (wtedy  $\sigma\sqrt{t}X$  ma taki sam rozkład jak  $\sigma W(t)$  bo zgadzają się wartość oczekiwana i wariancja). Jest to bardzo mała liczba, na przykład dla  $t = \frac{1}{52}$  (tydzień),  $\mu = 10\%$ ,  $\sigma = 120\%$  (bardzo ryzykowna akcja), dostajemy 0.00147%. Przy analogicznym pytaniu dla okresu jednego dnia liczba będzie praktycznie zero.

Obrona ta poniosła porażkę, gdyż wiadomo że zdarzają się zmiany bardzo gwałtowne, a najbardziej spektakularne to przypadek bankructwa firmy typu Enron, gdzie bez ostrzeżenia następuje spadek skokowy, niekiedy do zera.

Pojawia się potrzeba dopuszczenia modelu w którym mogą zachodzić skoki cen. Taki prosty model opiszemy zaznaczając, że będzie to tylko pierwszy krok a dalsze konstrukcje są trudne matematycznie.

Zakładamy, że proces  $N(t)$  określony dla  $t \geq 0$  startuje z zera, czyli  $N(0) = 0$ . W losowej chwili  $\tau$  nastąpi skok do 1, gdzie zakładamy że  $\tau_1$  to zmienna losowa o znanej nam gęstości

$$f_{\tau_1}(t) = \begin{cases} \lambda e^{-\lambda t} & \text{dla } t \geq 0 \\ 0 & \text{dla } t < 0 \end{cases}$$



Dalej, zakładamy, że po skoku wszystko zaczyna się od początku, czyli nastąpi następny skok, do wartości 2, w losowej chwili  $\tau_2$  o takim rozkładzie jak  $\tau_1$ , ale niezależnej od  $\tau_1$  (czyli drugi skok będzie w chwili  $\tau_1 + \tau_2$ ). W wyniku otrzymujemy tak zwany proces Poissona

$$N(t) = \begin{cases} 0 & \text{gdy } 0 \leq t < \tau_1 \\ 1 & \text{gdy } \tau_1 \leq t < \tau_2 \\ 2 & \text{gdy } \tau_2 \leq t < \tau_3 \\ \text{itd.} \end{cases}$$

Wybór  $\omega \in \Omega$  (dla odpowiedniej przestrzeni której nie omawiamy) oznacza wybór liczb  $\tau_n^\omega$  czyli wybór trajektorii, procesu, czyli funkcji  $N^\omega(t)$ .

Z punktu widzenia modelowania witamy z zadowoleniem dopuszczenie możliwości skoków ale skoki te nie wydają się wystarczające do opisu na przykład bankructwa. Tak jest w istocie, ale jesteśmy już tylko mały krok od celu. Wystarczy zdefiniować tak zwany złożony proces Poissona, gdzie mamy do dyspozycji dodatkowo ciąg niezależnych zmiennych losowych  $Z_n$  i w każdej chwili skoku zmienna  $Z_n$  mówi o jego rozmiarze. Jeśli na przykład  $Z_n$  będą miały standardowy rozkład normalny to będą dopuszczone skoki zarówno dodatnie jak i ujemne. Taki proces jest stosowany do badania bankructwa ale to jest tematem specjalistycznego kursu.