MI
Metody Identyfikacji

wykład #1

1. *Wstęp*

Kwestie formalne

- Wykład (55 pkt.):
  - sprawdzian zaliczeniowy (2 godz.) –25 maja 2020
  - sprawdzian poprawkowy (2 godz.) –8 czerwca 2020
- Projekt (nieograniczona ilość punktów):
  - Zadanie: dane przemysłowe
  - obrona (konieczna): 1 czerwca 2020
  - Ocena za projekt składa się z trzech składowych:
    - Prowadzący wykład: 0-15 pkt.
    - Prowadzący projekt: 0-20 pkt.
    - Każdy Zespół studencki ma do rozdysponowania po 10 punktów pomiędzy inne Zespoły

Kwestie formalne

- Warunki zaliczenia przedmiotu:
  - **wykład:** ponad 27 pkt.
  - **projekt:** dostarczenie w terminie (1 czerwca 2020, godz. 16:15) sprawozdania
- Skala ocen:
  - <50.5 = 2
  - 50.5-60 = 3
  - 60.5-70 = 3.5
  - 70.5-80 = 4
  - 80.5-90 = 4.5
  - 90.5-10 = 5

Literatura

- Zdzisław Bubnicki: *Identyfikacja obiektów sterowania*, PWN, 1974.
- Michael S. Mahoney: Historical Perspectives on Models and Modeling
- Lenart Ljung: *System Identification: Theory for Users*
- Chris Bissell and Chris Dillon (Eds.): *Ways of Thinking, Ways of Seeing Mathematical and Other Modelling in Engineering and Technology*, Springer, 2012.
- Rolf Isermann and Marco Münchhof: *Identification of Dynamic Systems*

Pośrednie wprowadzenie

1. Nate Silver: *The signal and the noise. The art and science of prediction*
2. Nassim Nicholas Taleb: *Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications*

pośrednie wprowadzenie

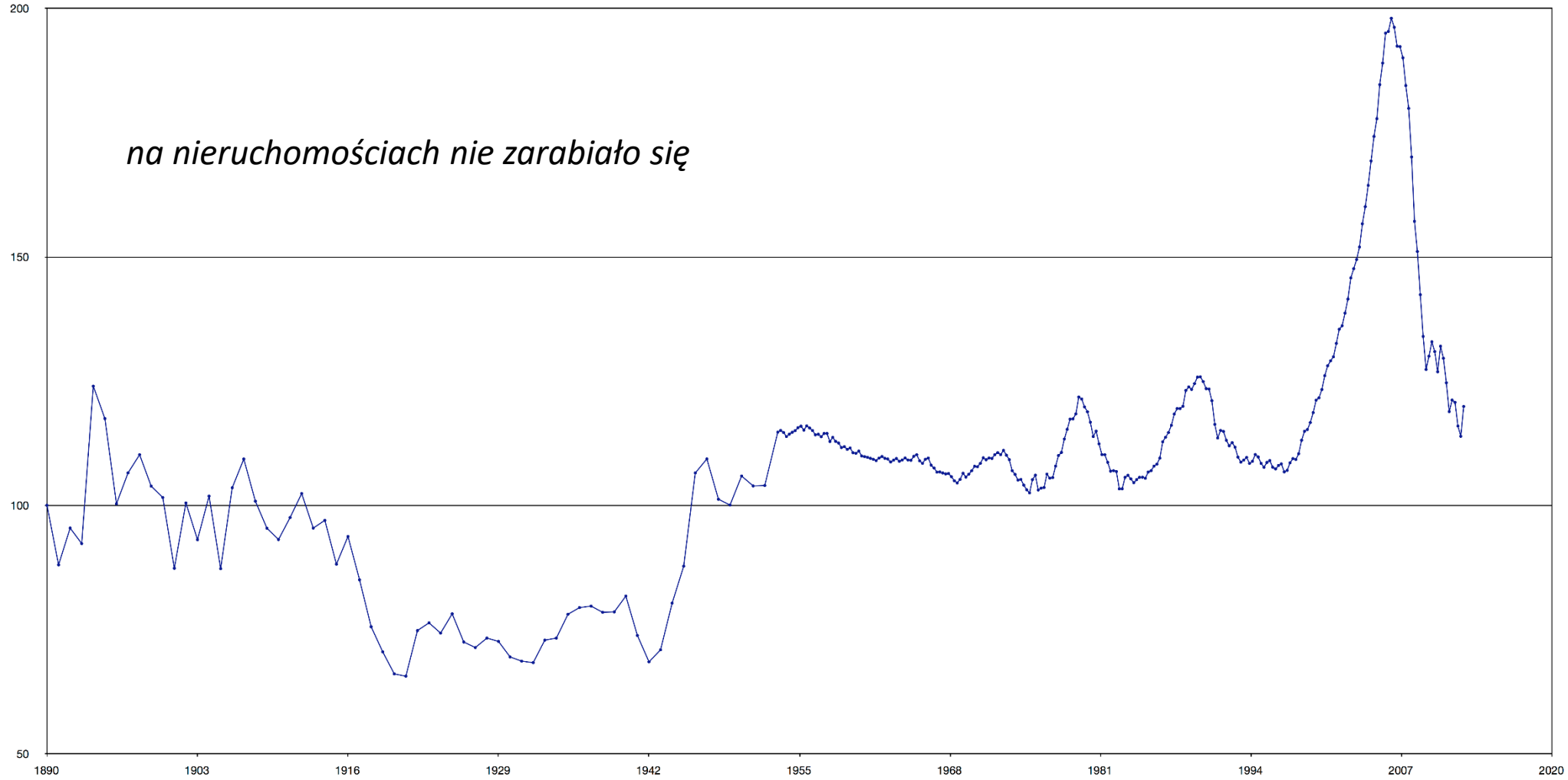
Nate Silver: *The signal and the noise. The art and science of prediction*.

- A Catastrophic Failure of Prediction

We focus on those signals that tell a story about world as we would like it to be, not how it really is. We ignore the risks that are hardest to measure, even when they pose the greatest threats to our well-being.

  - Bańka na rynku nieruchomości
  - Agencje ratingowe dawały najwyższą ocenę AAA dla tych inwestycji
    - Zależność od siebie poszczególnych kredytów hipotecznych – poszczególne narzędzia i zdarzenia nie były od siebie niezależne
    - Mnóstwo powiązanych instrumentów pochodnych
    - Wszyscy wiedzieli, nikt nic nie zrobił – nikomu nie zależało na jakości prognoz – agencje zarabiały na ilości prognoz
  - Różnica pomiędzy *ryzykiem* a *niepewnością*.
    - **Ryzyko**: coś co możesz wyrazić za pomocą jakieś ceny (Frank Knight, 1921)
    - **Niepewność:** niemierzalne

Akt 1: ceny nieruchomości

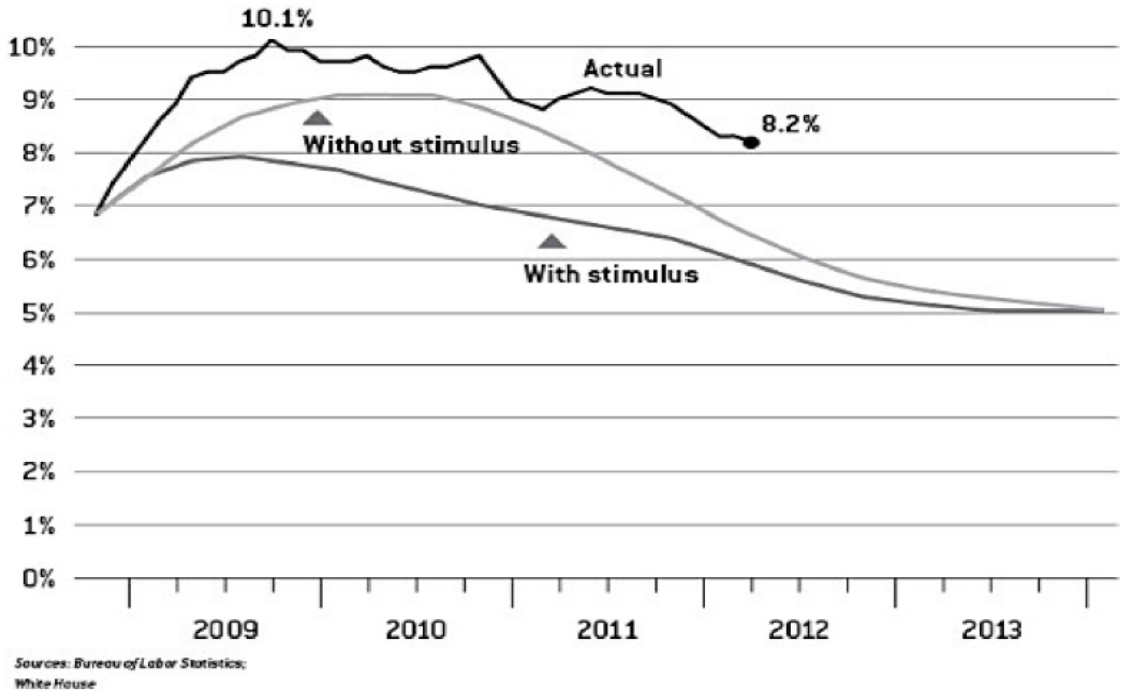


na nieruchomościach nie zarabiało się

Ciąg dalszy

- Akt 2: dźwignia
  - Zwykły Amerykanin miał *65%* dobrobytu związanego z nieruchomością, natomiast oszczędności malały
  - Dla każdego ***1$*** zainwestowanego w hipoteki istniało co najmniej ***50$*** w instrumentach pochodnych na giełdzie
  - dodatnie sprzężenie zwrotne
    - podaż-popyt → sprzężenie ujemne, stabilizuje rynek
    - strach-pożądanie → sprzężenie dodatnie, destabilizacja

- Akt 3: Bezrobocie
  - Kryzys finansowy generuje bezrobocie



Sytuacja jakiej nie było wcześniej: *out of sample*

Model jest tylko przybliżeniem rzeczywistości – musimy znać jago zakres stosowalności

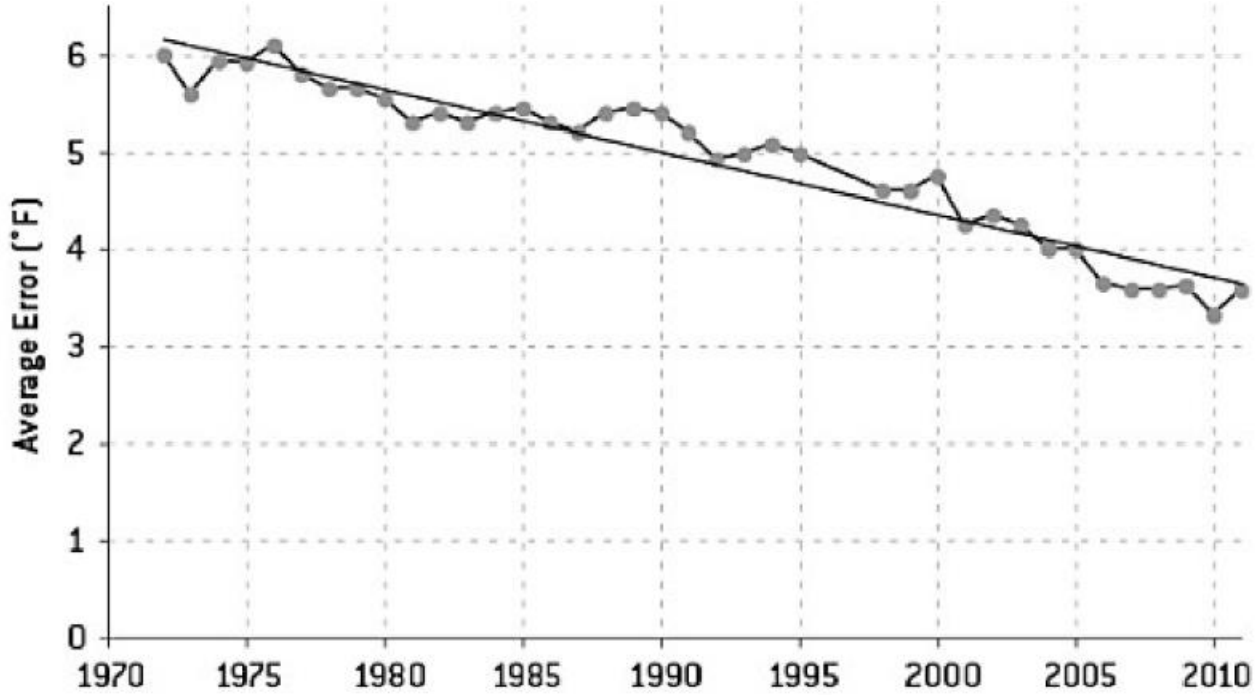
Are you Smarter than a Television Pundit?

- *Mądre głowy* w telewizji zawsze mylą się w swoich prognozach
  - 15% zdarzeń, które nie miały szans na zdarzenie, wystąpiło
  - 25% zdarzeń, które były absolutnie pewne, nigdy nie wydarzyło się
- Hedgehogs and foxes
  The fox knows many little things, but the hedgehog knows one big thing. Archilochus
- Błąd w myśleniu: *dobra prognoza nie ulega zmianie*

- Wykorzystujmy każdą nadarzającą się informację, również tą jakościową – podejścia hybrydowe

For Years You've Been Telling us that Rain is Green

- Prognozy pogody są bardzo dobre, wraz za zwiększającą się mocą obliczeniową następuje poprawa prognoz – podejście podziału dziedziny w 3D
- Ekspert modyfikuje modele i prognozy – lepiej widzimy wzorce w szumie
- Kalibracja
- Własności dobrej prognozy:
  - Dokładność
  - Rzetelność
  - Wartość ekonomiczna



Desperately Seeking Signal

- W przeciwieństwie do pogody nie jesteśmy w stanie przewidywać trzęsień ziemi
  - nie ma świętego Grala
- Różnica pomiędzy prognozą a przewidywaniem
- *Power Law* – prawo potęgowe
- Poszukiwanie sygnału w szumie
- Overfitting
- Ekstrapolacja

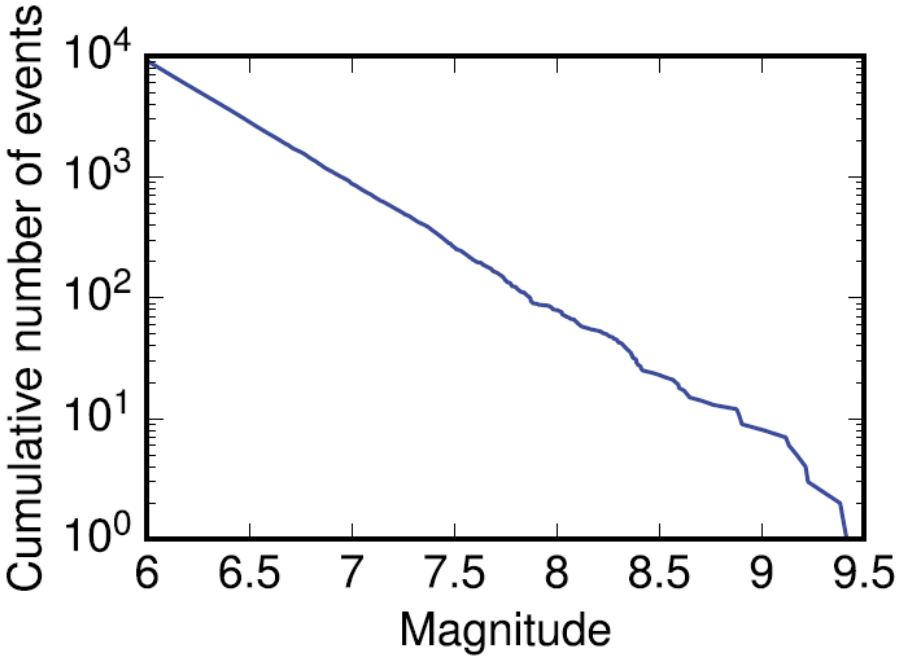
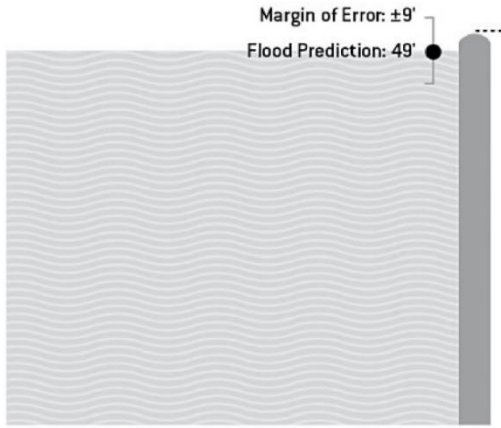
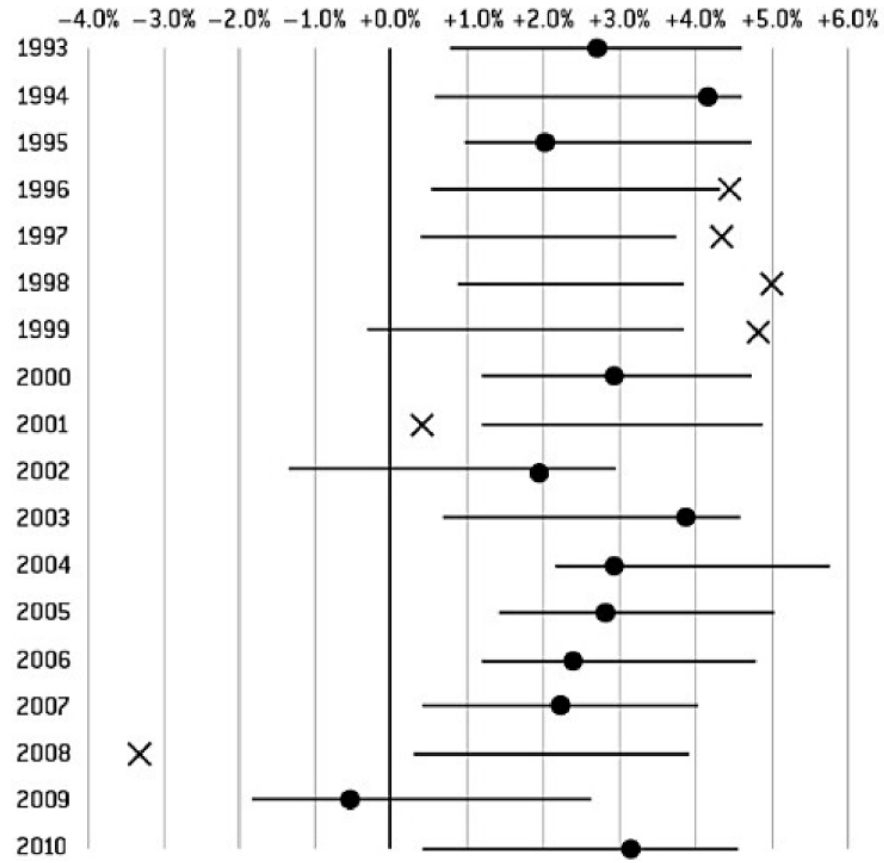


Figure 3. Magnitude-frequency distribution used in the synthetic data sets. Events follow a Gutenberg-Richter distribution with $b = 1$ and a minimum magnitude of $M_{min} = 6$. While this differs from the empirical value of $b = 1.26$ observed for the PAGER/PDE catalog, using $b = 1$ makes determining the number of events at a given magnitude level more straightforward. Alternative magnitude-frequency distributions with $b = 0.8$, $b = 1.2$, and $M_{min} = 5$ are also considered for the ETAS simulations.

How to Drown in Three Feet of Water

- Przekleństwo statystyki
- Prognozy PKB – dokładność
- Mnóstwo danych
- Mnóstwo korelacji
- Brak zależności przyczynowo - skutkowej
- Jak powinna być struktura modelu
- Prognozy zagregowane są lepsze niż pojedyncze





Role models

- Prognozowanie rozprzestrzeniania chorób / pandemii
- Niebezpieczeństwo ekstrapolacji
- Prognozy samospełniające się
  - Głosowanie na kandydata wygrywającego
  - Prognozowanie modnego koloru na przyszły rok
  - Choroby
- Prognozy samo-kasujące się
  - Wszyscy używają jednego systemu nawigacji w celu ominięcia korka
- Paradoks HIV w San Francisco 1990-2000
- Modele agentowe
  - *All models are wrong, but some models are useful*
  - *The best model of a cat is a cat.*

Less and Less and Less Wrong

to err and err and err again, but less and less and less, Piet Hein

- prawdopodobieństwem *a priori*
  - Dotyczy zdarzenia, które jeszcze nie nastąpiło, a my pytamy o prawdopodobieństwo zajścia jakiegoś zdarzenia przed jego wykonaniem, czyli *a priori*.
  - Liczymy to prawdopodobieństwo zgodnie ze wzorem na prawdopodobieństwo całkowite
  - Mówiąc krótko, znamy wszystkie możliwe przyczyny, a nie znamy skutku (rezultatu).
  - Ronald Aymer Fisher (próbkowanie)
- Prawdopodobieństwem *a posteriori*
  - W tym przypadku doświadczenie już zostało wykonane, znamy jego wynik. Nas interesuje prawdopodobieństwo tego, że dana przesłanka przyczyniła się do danego skutku.
  - To jest sytuacja *a posteriori*, znamy skutek, a pytamy o prawdopodobieństwo, że to właśnie ta określona przyczyna spowodowała skutek.
  - Thomas Bayes

pośrednie wprowadzenie

Nassim Nicholas Taleb: *Statistical Consequences of Fat Tails*

- *The main idea behind the Incerto project is that while there is a lot of uncertainty and opacity about the world, and an incompleteness of information and understanding, there is little, if any, uncertainty about what actions should be taken based on such an incompleteness, in any given situation.*
- *Complication without insight: the clarity of mind of many professionals using statistics and data science without an understanding of the core concepts, what it is fundamentally about.*

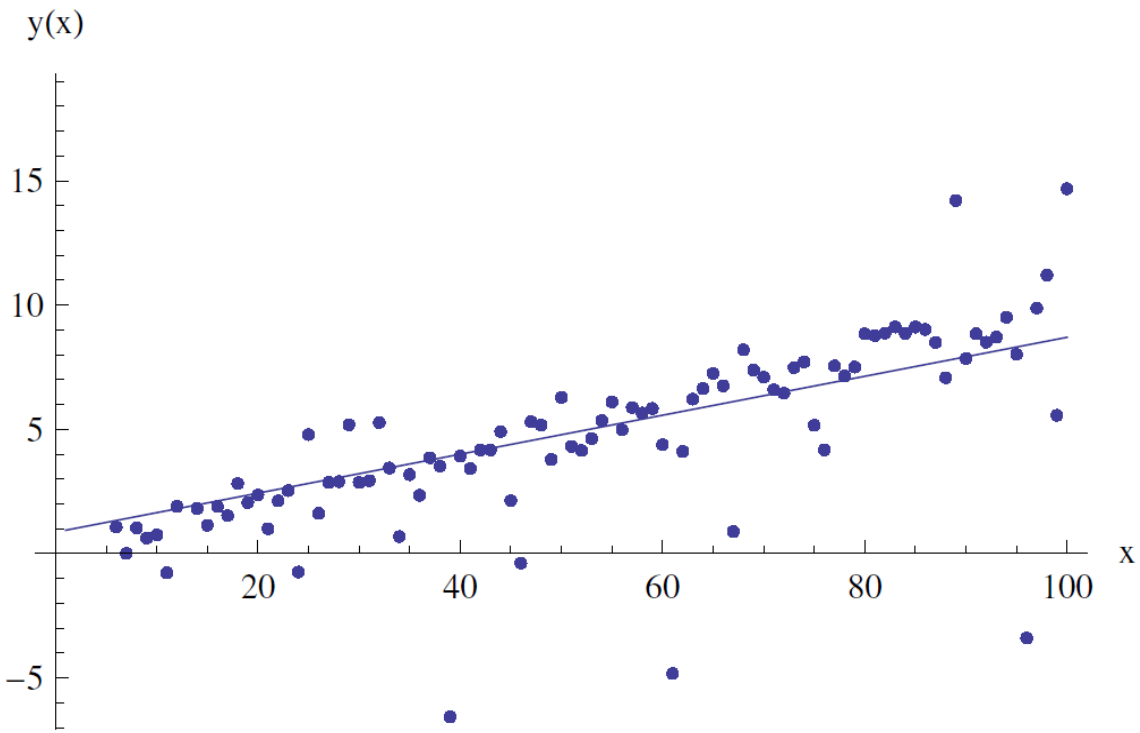
Pinker problem

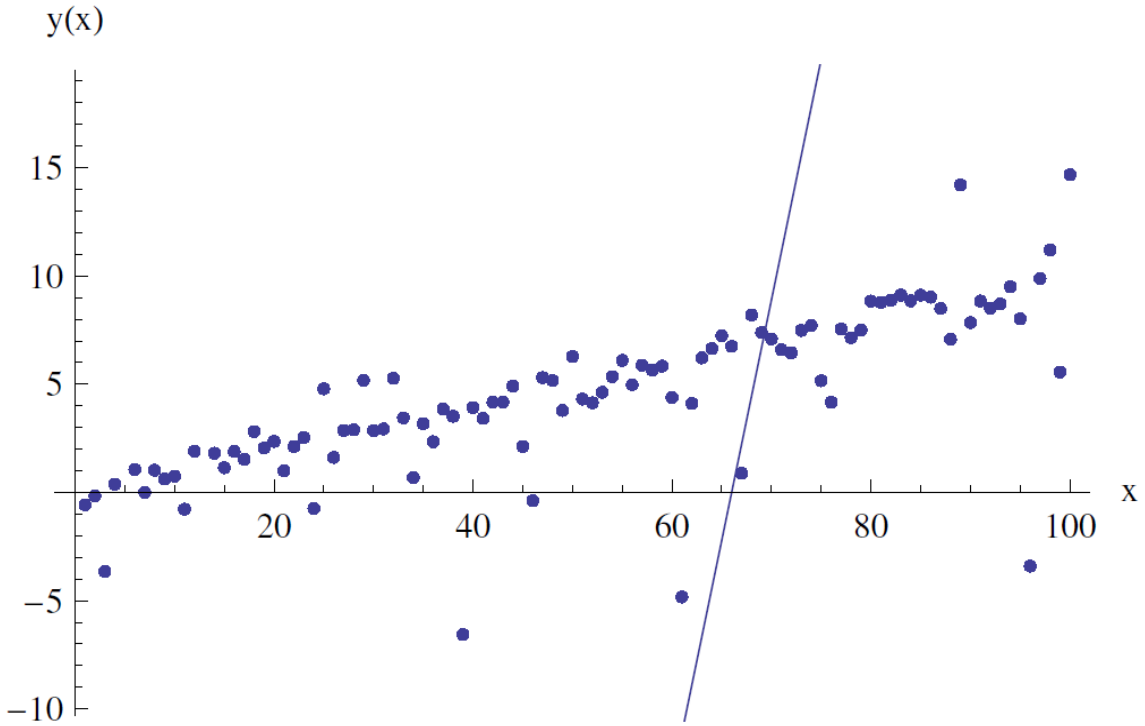
- Discussion of "evidence" that is not statistically significant, or use of metrics that are uninformative because they do not apply to the random variables under consideration – like for instance making inferences from the means and correlations for fat tailed variables. This is the result of:
  - the focus in statistical education on Gaussian or thin-tailed variables,
  - the absence of probabilistic knowledge combined with memorization of statistical terms,
  - complete cluelessness about dimensionality.
- Example of pseudo-empiricism: comparing death from terrorist actions or epidemics such as ebola (fat tailed) to falls from ladders (thin tailed).

Cienkie a grube ogony

- In Mediocristan, when a sample under consideration gets large, no single observation can really modify the statistical properties.
  - the probability of sampling higher than *X* twice in a row is greater than sampling higher than 2*X* once
- In Extremistan, the tails (the rare events) play a disproportionately large role in determining the properties.
  - the probability of sampling higher than 2X once is greater than the probability of sampling higher than X twice in a row.

regresja





obserwacje

1. *The law of large numbers, when it works, works too slowly in the real world.*
2. *The mean of the distribution will rarely correspond to the sample mean; it will have a persistent small sample effect (downward or upward) particularly when the distribution is skewed (or one-tailed).*
3. *Metrics such as standard deviation and variance are not useable.*
4. *Practically every single economic variable and financial security is thick tailed. Of the 40,000 securities examined, not one appeared to be thin-tailed. This is the main source of failure in finance and economics.*
5. *Practically any paper in economics using covariance matrices is suspicious.*
6. *Linear least-square regression doesn't work (failure of the Gauss-Markov theorem).*
7. *Maximum likelihood methods can work well for some parameters of the distribution (good news).*

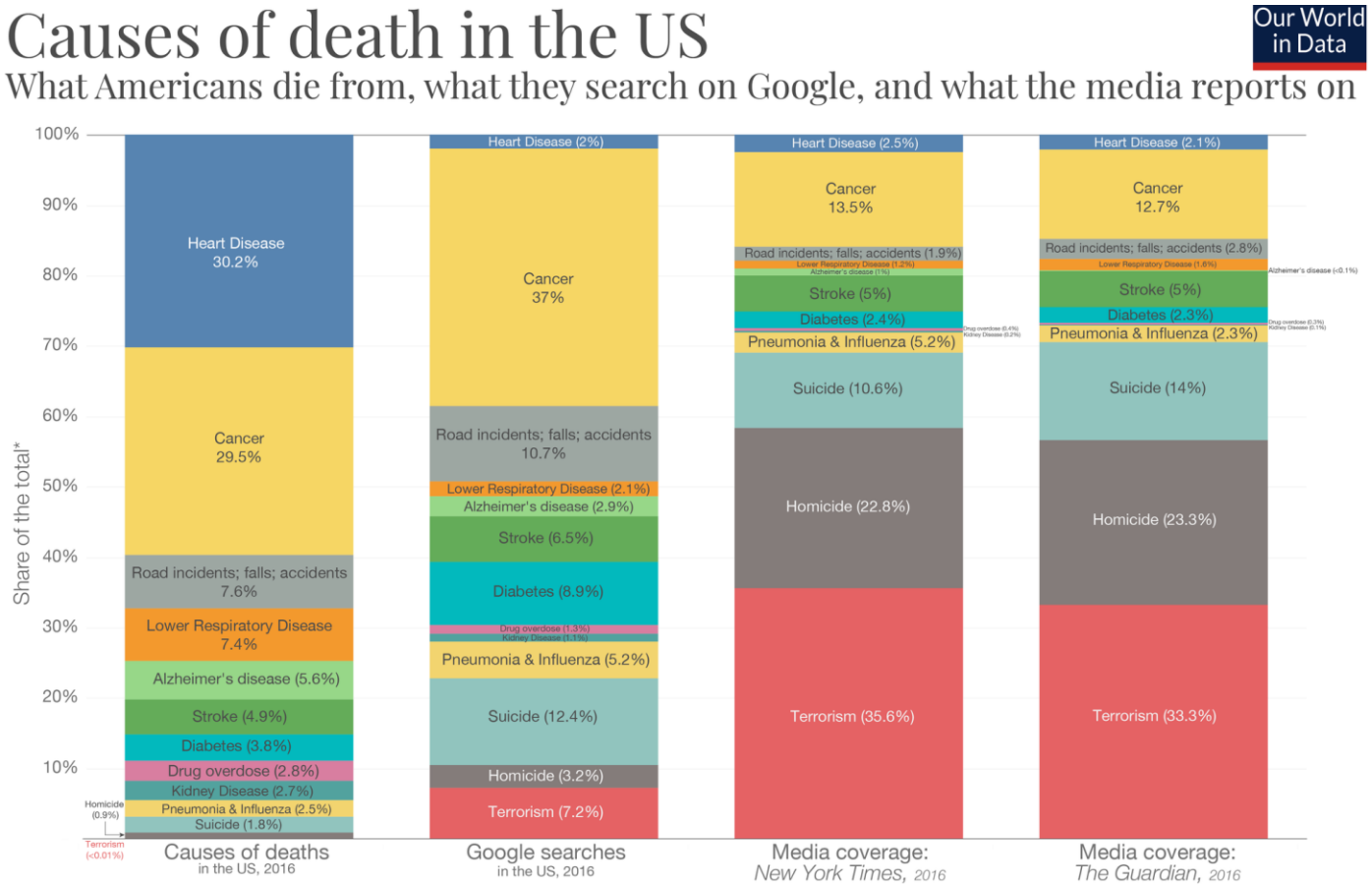
obserwacje, cd.

8. *The gap between disconfirmatory and confirmatory empiricism is wider than in situations covered by common statistics i.e., the difference between absence of evidence and evidence of absence becomes larger.*
9. *The method of moments (MoM) fails to work. Higher moments are uninformative or do not exist.*
10. *There is no such thing as a typical large deviation*

Epistemologia

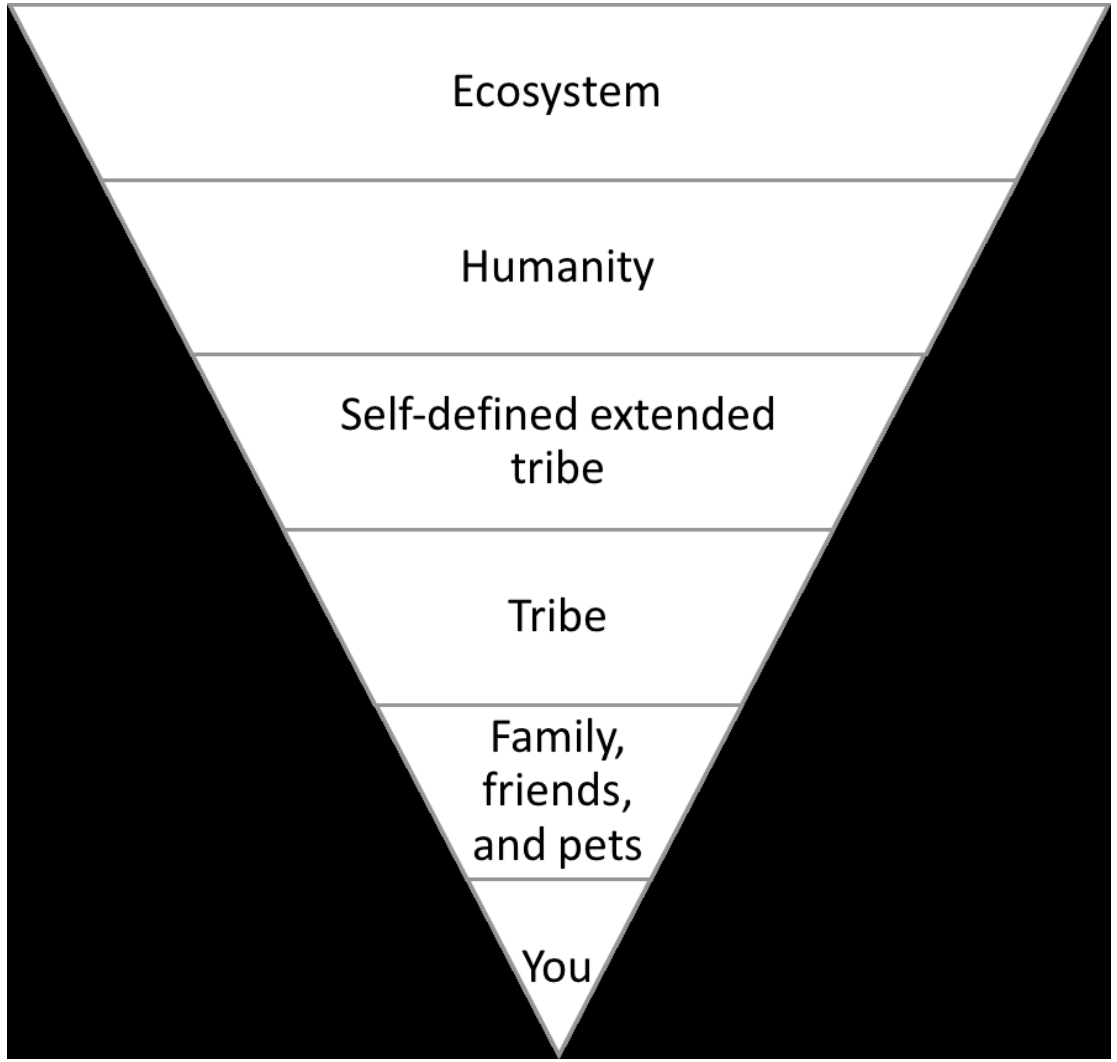
- We do not observe probability distributions, just realizations.
- A probability distribution cannot tell you if the realization belongs to it.
- You need a meta-probability distribution to discuss tail events (i.e., the conditional probability for the variable to belong to a certain distributions vs. others).

potoczność



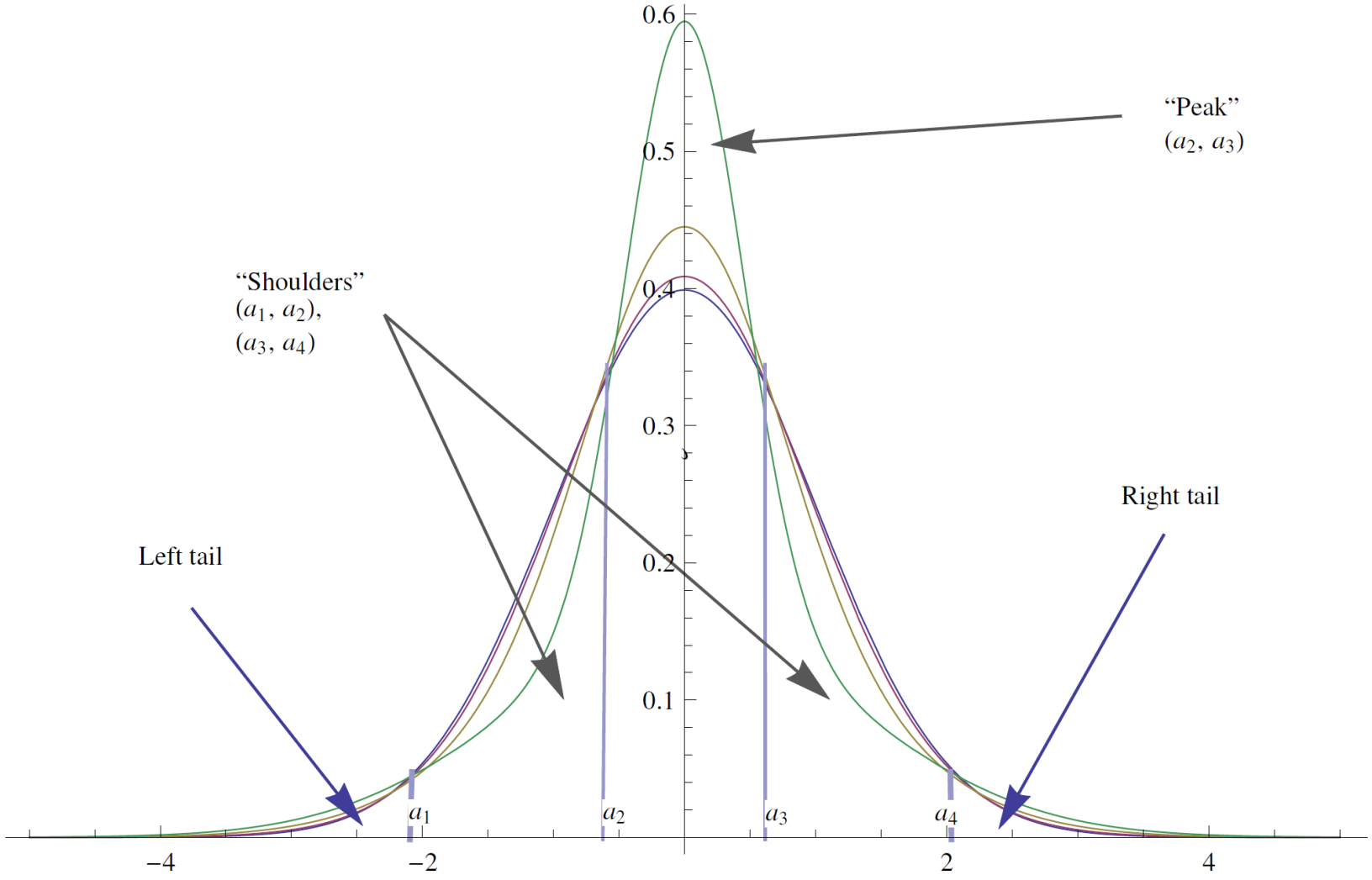
Bill Gates's Naive (Non-Statistical) Empiricism: the founder of Microsoft is promoting and financing the development of the above graph, yet at the same time claiming that the climate is causing an existential risk, not realizing that his arguments conflict since existential risks are necessarily absent in past data. Furthermore, a closer reading of the graphs shows that cancer, heart disease, and Alzheimer, being ailments of age, do not require the attention on the part of young adults and middle-aged people something terrorism and epidemics warrant.

Another logical flaw is that terrorism is precisely low because of the attention it commands. Relax your vigilance and it may go out of control. The same applies to homicide: fears lead to safety.

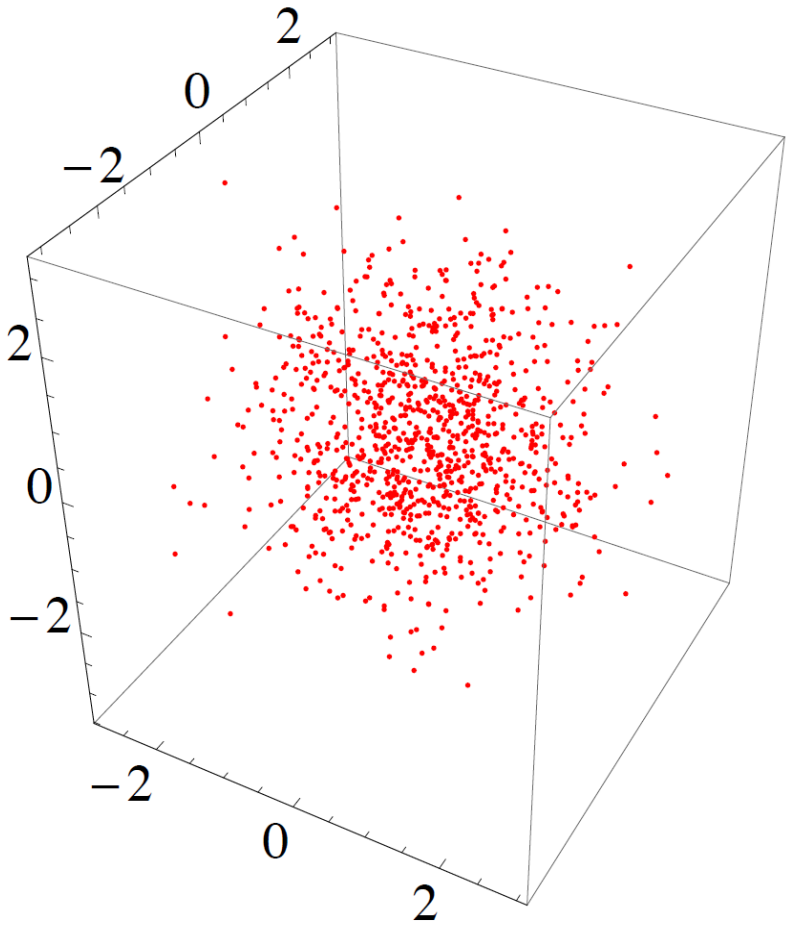


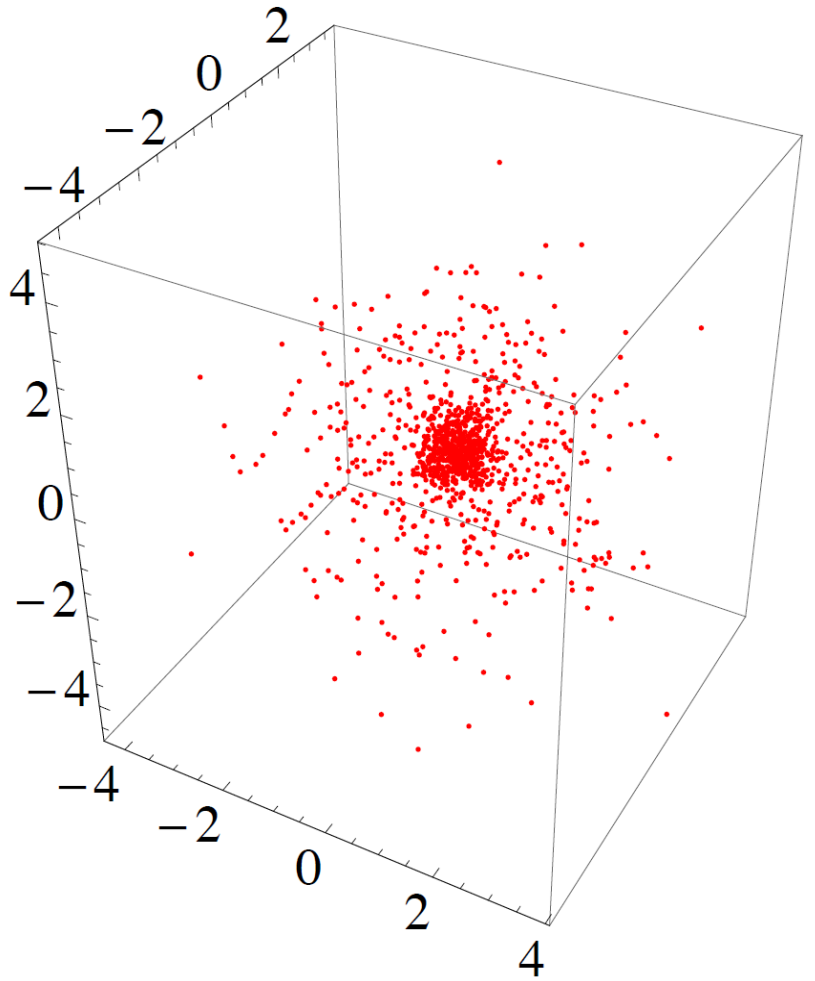
A hierarchy for survival. Higher entities have a longer life expectancy, hence tail risk matters more for these. Lower entities such as you and I are renewable.

Gdzie zaczyna się ogon?



Ogony a wymiary

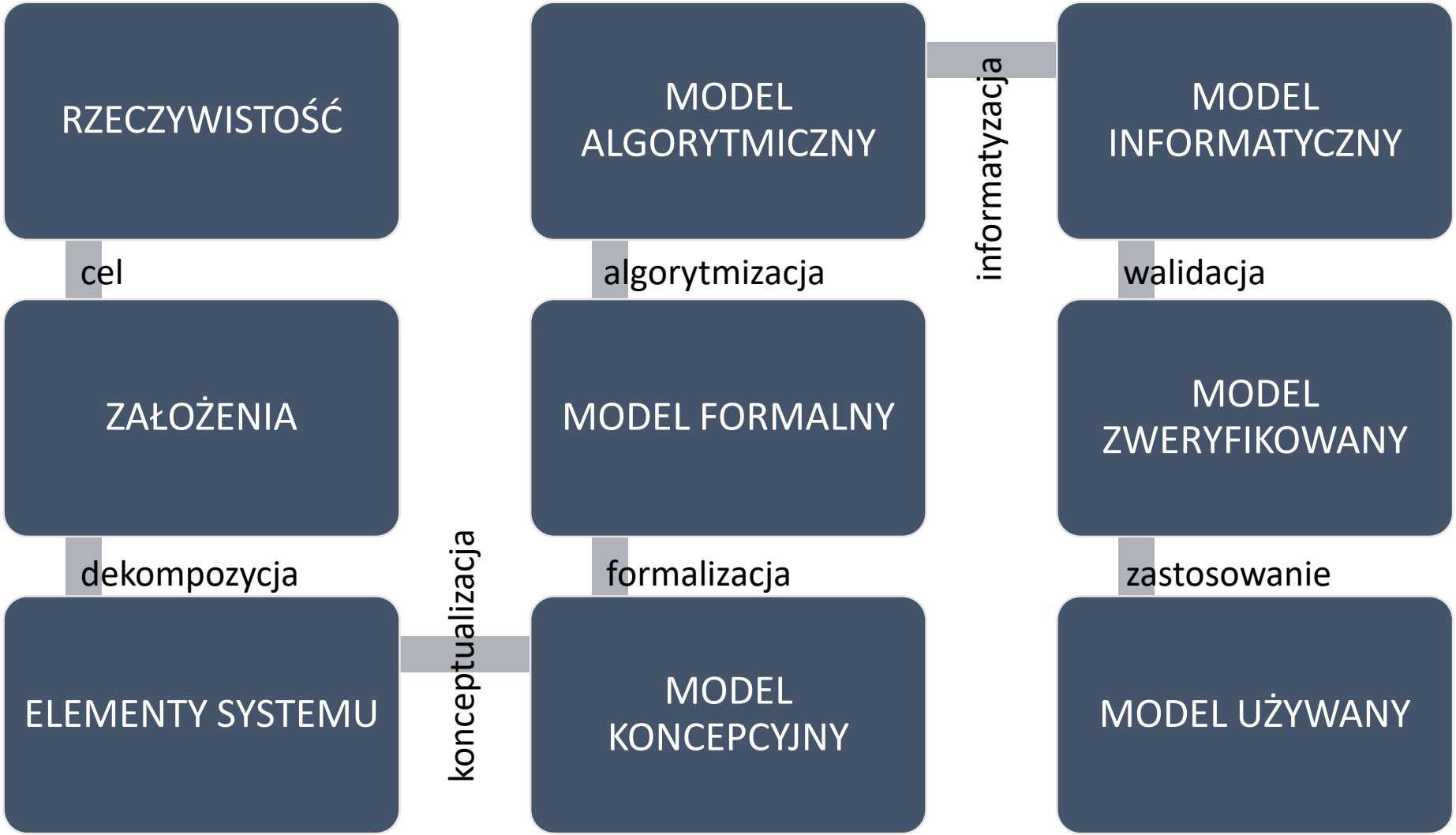




Inne kwestie

- Jak wiele danych potrzeba do analizy?
- Wartości ekstremalne, outliers, ogony

Proces modelowania



Identyfikacja

- IDENTYFIKACJA: wygenerowanie z istniejącego fragmentu rzeczywistości elementów niezbędnych do zbudowania modelu z punktu widzenia potrzeby, możliwości, warunków i ograniczeń itp.
- PYTANIA:
  - Jakie elementy tworzą system?
  - Jakie relacje tworzą strukturę?
  - Jakie elementy mi relacje są istotne ze względu na cel?
  - Jak struktura systemu wpływa na funkcję systemu?
  - Jak otoczenie może wpływać na zmianę struktury?
- PROBLEMY:
  - Funkcje i procesy realizowane w systemie
  - Zachowanie się systemu w danych warunkach
  - Organizacja realizacji procesów w systemie
  - Uzyskanie pożądanego przebiegu procesów
  - Związki cech opisujących system z badaną właściwością

Identyfikacja

- KONCEPTUALIZACJA: transformacja zbioru założeń otrzymanych w procesie identyfikacji do postaci modelu określającego relacje między zidentyfikowanymi elementami i ustalającymi atrybuty niezbędne do opisu systemu.
- FORMALIZACJA: budowa modelu formalnego (matematycznego).
- ALGORYTMIZACJA: przekształcenie modelu matematycznego w postać numeryczną (algorytm, schemat blokowy).
- INFORMATYZACJA: budowa modelu komputerowego (programu komputerowego).
- WERYFIKACJA: konfrontacja oszacowań otrzymanych w modelu z oszacowaniami otrzymanymi w warunkach naturalnych (rzeczywistych).
- ADAPTACJA: ustalenie zakresu, warunków zastosowania, możliwości posługiwania się modelem (opracowanie instrukcji posługiwania się modelem).

Identyfikacja

- Identyfikacja zajmuje się wyznaczaniem modeli matematycznych obiektów.
  - Model matematyczny definiuje w sposób ścisły zachowanie się obiektu w określonych warunkach.
  - Warunki te są określone przez wejścia i wyjścia obiektu w chwili obecnej i w przeszłości.
  - Model jest pewnym przybliżeniem rzeczywistego obiektu, idealizacją z ograniczającymi założeniami (ograniczona ilość wejść i wyjść, liniowość).
  - Stosuje się różne postacie modeli w zależności od przeznaczenia tworzonego modelu i struktury identyfikowanego obiektu.
  - Identyfikacja jest przeprowadzana na podstawie informacji pomiarowej o wielkościach wejściowych i wyjściowych obiektu.
  - W odniesieniu do sygnałów również stosuje się często idealizację, przybliżając rzeczywiste sygnały ich idealnymi odpowiednikami opisywanymi matematycznie.
  - Wstępnym etapem identyfikacji jest określenie charakteru obiektu na podstawie rejestracji jego sygnałów (statyczny czy dynamiczny, jeśli dynamiczny to jakiego rzędu, o jakiej dynamice).

Klasyfikacja modeli

• Liniowe	⟷	• Nieliniowe
• Niezbyt dokładne (jakościowe)	⟷	• Precyzyjne (ilościowe)
• Przyczynowo - skutkowe	⟷	• Pozostałe
• Statyczne	⟷	• Dynamiczne
• Deterministyczne	⟷	• Stochastyczne
• O stałych skupionych	⟷	• Rozproszone (o stałych rozłożonych)
• Stacjonarne	⟷	• Niestacjonarne
• z czasem dyskretnym	⟷	• z czasem ciągłym
• Fizykochemiczne	⟷	• Empiryczne
• SISO	analiza wymiarowa ⟷	• MIMO

Identyfikacja (jako proces)

• Nieinwazyjna (nie zaburza)	⟷	• Inwazyjna
• strukturalna	⟷	• parametryczna
• ilościowa	⟷	• jakościowa
• statyczna	⟷	• dynamiczna
• jednorazowa	⟷	• powtarzalna, okresowa, rekurencyjna
• *a priori*	⟷	• *a posteriori*

Identyfikacja

- Procedura identyfikacji jest procedurą iteracyjną:
  - Planowanie pomiarów w celu uzyskania danych. Często uzyskanie danych wystarczających do identyfikacji jest możliwe tylko na drodze odpowiedniego pobudzania wejść układu, a takie oddziaływanie nie zawsze jest możliwe w praktyce.
  - Pozyskanie i weryfikacja danych obiektowych
  - Wybór struktury modelu, tzn. określenie postaci i rzędu równań, opóźnień itp.
    → Identyfikacja strukturalna
  - Estymację parametrów modelu. Metoda estymacji zależy od przyjętego modelu i tego, czy obliczenia są prowadzone w trybie on- czy off-line.
    → Identyfikacja parametryczna
  - Sprawdzenie (walidacja) poprawności modelu. Do sprawdzenia najlepiej wykorzystać dane pomiarowe inne niż wykorzystane do estymacji parametrów. W razie negatywnego wyniku testu procedurę powtarza się.

Sztuka prognozowania

- **Przewidywanie** - wnioskowanie o zdarzeniach nie znanych na podstawie zdarzeń znanych – coś się może wydarzyć:
  - racjonalne (logiczny proces)
    - zdroworozsądkowe (wnioskowanie oparte na doświadczeniu),
    - naukowe (wnioskowanie z wykorzystaniem reguł nauki),
  - nieracjonalne (wróżby i proroctwa).
- **Prognozowanie** (predykcja) - racjonalne, naukowe przewidywanie przyszłych zdarzeń (zazwyczaj wnioskowanie oparte na danych czasowych lub przekrojowych) – co się wydarzy i kiedy.
  - sformułowany z wykorzystaniem dorobku nauki
  - odnoszący się do określonej przyszłości
  - weryfikowalny empirycznie w przyszłości
  - wartość logiczna sądu nieznana w momencie jego formułowania
  - niepewny, ale akceptowany
  - sąd oznajmujący, a nie warunkowy
- Co prognozujemy/przewidujemy?

Cele prognozowania

- Preparacyjna
- Aktywizująca
- Informacyjna

Klasyfikacja prognoz

- Perspektywa czasowa
  - Prognozy krótkookresowe
  - Prognozy średniookresowe
  - Prognozy długookresowe
- Cel
  - Prognoza badawcza (w tym ostrzegawcza)
  - Prognoza realistyczna
- Skutek
  - Prognoza samorealizująca się
  - Prognoza samounicestwiająca się

Błąd prognozy

- Ex ante
  - Błąd modelu na danych historycznych, np. błąd średniokwadratowy
- Ex post
  - Trafność prognozy (jej jakość, skuteczność, powtarzalność, stosowalność)

Zagrożenia

- chętnych do dzielenia się swoimi wizjami nie brakuje - dostajemy to czego chcieliśmy i stanowczo zbyt wiele opieramy na tych prognozach
- skłonność do mylenia pewności siebie z biegłością w temacie
- zwykle nie jesteśmy w stanie ocenić wartości prognoz
- tworzenie prognoz tak nieprecyzyjnych, rozmytych i nieokreślonych w czasie, że wręcz niemożliwych do zweryfikowania
- nikt nie jest nieomylny

- wartości odstające - *outliers*