Teoria zdarzeń ekstremalnych (Extreme Value Theory, EVT) czyli teoria katastrof

Krzysztof Czarnowski

1 sierpnia, 2016

(Wykorzystałem wiedzę i materiały udostępnione przez Joannę Czarnowską)

Gdzie zastosowania?

- Katastrofy naturalne
  Na przykład ocena ryzyka wystąpienia ekstremalnych opadów i związanego z tym ryzyka powodzi lub choćby przeciążenia systemów burzowych i kanalizacyjnych.

Gdzie zastosowania?

- Katastrofy naturalne
  Na przykład ocena ryzyka wystąpienia ekstremalnych opadów i związanego z tym ryzyka powodzi lub choćby przeciążenia systemów burzowych i kanalizacyjnych.
- Ubezpieczenia
  Szacowanie ryzyka wystąpienia szczególnie wysokich roszczeń i szacowanie poziomu wymaganych rezerw. Temat jest aktualny w związku z wdrażaniem dyrektywy Solvency II.

Gdzie zastosowania?

- Katastrofy naturalne
  Na przykład ocena ryzyka wystąpienia ekstremalnych opadów i związanego z tym ryzyka powodzi lub choćby przeciążenia systemów burzowych i kanalizacyjnych.
- Ubezpieczenia
  Szacowanie ryzyka wystąpienia szczególnie wysokich roszczeń i szacowanie poziomu wymaganych rezerw. Temat jest aktualny w związku z wdrażaniem dyrektywy Solvency II.
- Inwestycje
  Zarządzanie ryzykiem strat.

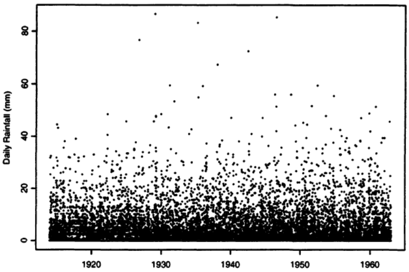
Gdzie zastosowania?

- Katastrofy naturalne
  Na przykład ocena ryzyka wystąpienia ekstremalnych opadów i związanego z tym ryzyka powodzi lub choćby przeciążenia systemów burzowych i kanalizacyjnych.
- Ubezpieczenia
  Szacowanie ryzyka wystąpienia szczególnie wysokich roszczeń i szacowanie poziomu wymaganych rezerw. Temat jest aktualny w związku z wdrażaniem dyrektywy Solvency II.
- Inwestycje
  Zarządzanie ryzykiem strat.
- Logistyka
  Szacowanie poziomu zapasów. (Tak naprawdę zdałem sobie sprawę z tego słuchając jednego z referatów w czasie *Innovation@Amazon*. Amazon to bardzo ciekawy przypadek!)
- ...

Problem

Przykład z książki S. Coles, s. 9 — opady dzienne w południowo wschodniej Anglii w latach 1914-62.

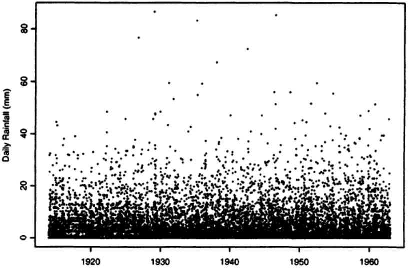
Estymacja prawdopodobieństwa wystąpienia ekstremalnych wartości czyli estymacja ogona rozkładu jest trudna.



Problem

Przykład z książki S. Coles, s. 9 — opady dzienne w południowo wschodniej Anglii w latach 1914-62.

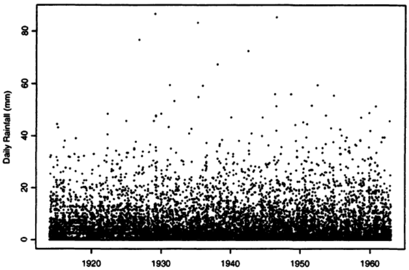
Estymacja prawdopodobieństwa wystąpienia ekstremalnych wartości czyli estymacja ogona rozkładu jest trudna.



- Standardowe metody parametryczne (MLE, metoda momentów) dobrze wyestymują ciało rozkładu, ogon zwykle bardzo słabo (mało danych, duża dysproporcja). I trzeba założyć jakiś kształt rozkladu!

Problem

Przykład z książki S. Coles, s. 9 — opady dzienne w południowo wschodniej Anglii w latach 1914-62.
Estymacja prawdopodobieństwa wystąpienia ekstremalnych wartości czyli estymacja ogona rozkładu jest trudna.



- Standardowe metody parametryczne (MLE, metoda momentów) dobrze wyestymują ciało rozkładu, ogon zwykle bardzo słabo (mało danych, duża dysproporcja). I trzeba założyć jakiś kształt rozkladu!
- Empiryczna estymacja kwantyli ma taki problem, że prawdopodobieństwo wystąpienia wartości większych niż zaobserwowane maksimum wynosi zero. To nie jest dobrze!

No to inaczej...

- Szukamy innego podejścia bardziej powiązanego z ogonem bazowego rozkładu, możliwie bez założeń o szczegółowym kształcie.

No to inaczej. . .

- Szukamy innego podejścia bardziej powiązanego z ogonem bazowego rozkładu, możliwie bez założeń o szczegółowym kształcie.
- Oszacujemy prawdopodobieństwo katastrofy w ciągu pewnego okresu czasu, na przykład wystąpienia wyjątkowo dużego opadu deszczu w ciągu roku.

No to inaczej. . .

- Szukamy innego podejścia bardziej powiązanego z ogonem bazowego rozkładu, możliwie bez założeń o szczegółowym kształcie.
- Oszacujemy prawdopodobieństwo katastrofy w ciągu pewnego okresu czasu, na przykład wystąpienia wyjątkowo dużego opadu deszczu w ciągu roku.
- Katastrofa wystąpi gdy w ciągu tego czasu pewna wielkość przynajmniej raz przekroczy krytyczny poziom.

No to inaczej...

- Szukamy innego podejścia bardziej powiązanego z ogonem bazowego rozkładu, możliwie bez założeń o szczegółowym kształcie.
- Oszacujemy prawdopodobieństwo katastrofy w ciągu pewnego okresu czasu, na przykład wystąpienia wyjątkowo dużego opadu deszczu w ciągu roku.
- Katastrofa wystąpi gdy w ciągu tego czasu pewna wielkość przynajmniej raz przekroczy krytyczny poziom.
- $X_1, X_2, \ldots X_n$ — niezależne zmienne losowe o jednakowym rozkładzie $F_X(x) = P(X \leqslant x)$ (dystrybuanta, ang. *cumulative distribution function*)

No to inaczej. . .

- Szukamy innego podejścia bardziej powiązanego z ogonem bazowego rozkładu, możliwie bez założeń o szczegółowym kształcie.
- Oszacujemy prawdopodobieństwo katastrofy w ciągu pewnego okresu czasu, na przykład wystąpienia wyjątkowo dużego opadu deszczu w ciągu roku.
- Katastrofa wystąpi gdy w ciągu tego czasu pewna wielkość przynajmniej raz przekroczy krytyczny poziom.
- $X_1, X_2, \ldots X_n$ — niezależne zmienne losowe o jednakowym rozkładzie $F_X(x) = P(X \leqslant x)$ (dystrybuanta, ang. *cumulative distribution function*)
- Na przykład: X_i — opad dzienny, $n = 365$.

No to inaczej. . .

- Szukamy innego podejścia bardziej powiązanego z ogonem bazowego rozkładu, możliwie bez założeń o szczegółowym kształcie.
- Oszacujemy prawdopodobieństwo katastrofy w ciągu pewnego okresu czasu, na przykład wystąpienia wyjątkowo dużego opadu deszczu w ciągu roku.
- Katastrofa wystąpi gdy w ciągu tego czasu pewna wielkość przynajmniej raz przekroczy krytyczny poziom.
- $X_1, X_2, \ldots\ X_n$ — niezależne zmienne losowe o jednakowym rozkładzie $F_X(x) = P(X \leqslant x)$ (dystrybuanta, ang. *cumulative distribution function*)
- Na przykład: X_i — opad dzienny, $n = 365$.
- $M_n = \max(X_1, X_2, \ldots\ X_n)$

No to inaczej...

- Szukamy innego podejścia bardziej powiązanego z ogonem bazowego rozkładu, możliwie bez założeń o szczegółowym kształcie.
- Oszacujemy prawdopodobieństwo katastrofy w ciągu pewnego okresu czasu, na przykład wystąpienia wyjątkowo dużego opadu deszczu w ciągu roku.
- Katastrofa wystąpi gdy w ciągu tego czasu pewna wielkość przynajmniej raz przekroczy krytyczny poziom.
- X_1, X_2, ... X_n — niezależne zmienne losowe o jednakowym rozkładzie $F_X(x) = P(X \leqslant x)$ (dystrybuanta, ang. *cumulative distribution function*)
- Na przykład: X_i — opad dzienny, $n = 365$.
- $M_n = \max(X_1, X_2, \ldots X_n)$
- Interesuje nas estymacja dystrybuanty $F_{M_n}(x) = P(M_n \leqslant x)$

No to inaczej...

- Szukamy innego podejścia bardziej powiązanego z ogonem bazowego rozkładu, możliwie bez założeń o szczegółowym kształcie.
- Oszacujemy prawdopodobieństwo katastrofy w ciągu pewnego okresu czasu, na przykład wystąpienia wyjątkowo dużego opadu deszczu w ciągu roku.
- Katastrofa wystąpi gdy w ciągu tego czasu pewna wielkość przynajmniej raz przekroczy krytyczny poziom.
- X_1, X_2, ... X_n — niezależne zmienne losowe o jednakowym rozkładzie $F_X(x) = P(X \leqslant x)$ (dystrybuanta, ang. *cumulative distribution function*)
- Na przykład: X_i — opad dzienny, $n = 365$.
- $M_n = \max(X_1, X_2, \ldots X_n)$
- Interesuje nas estymacja dystrybuanty $F_{M_n}(x) = P(M_n \leqslant x)$
- Jak to się zachowuje gdy n rośnie? Jakiś rozkład graniczny?

Przecież to banalne!

- $\max(X_1, X_2, \ldots\ X_n) \leqslant x$ wtedy i tylko wtedy, gdy $X_1 \leqslant x$ i $X_2 \leqslant x$ i tak dalej, i w końcu $X_n \leqslant x$

Przecież to banalne!

- $\max(X_1, X_2, \ldots\ X_n) \leqslant x$ wtedy i tylko wtedy, gdy $X_1 \leqslant x$ i $X_2 \leqslant x$ i tak dalej, i w końcu $X_n \leqslant x$
- czyli $F_{M_n}(x) = F_X(x) \cdot F_X(x) \cdots F_X(x) = (F_X(x))^n$ (Założyliśmy że wszystkie X_i mają ten sam rozkład i przede wszystkim niezależność! Czyli prawdopodobieństwa się mnożą.)

Przecież to banalne!

- $\max(X_1, X_2, \ldots\ X_n) \leqslant x$ wtedy i tylko wtedy, gdy $X_1 \leqslant x$ i $X_2 \leqslant x$ i tak dalej, i w końcu $X_n \leqslant x$
- czyli $F_{M_n}(x) = F_X(x) \cdot F_X(x) \cdots F_X(x) = (F_X(x))^n$ (Założyliśmy że wszystkie X_i mają ten sam rozkład i przede wszystkim niezależność! Czyli prawdopodobieństwa się mnożą.)
- W takim razie, jeśli $F_X(x) < 1$ dla $x < x_0$ i $F_X(x) = 1$ dla $x \geqslant x_0$, to

$$\lim_{n\to\infty} F_{M_n}(x) = \begin{cases} 0 & \text{dla } x < x_0 \\ 1 & \text{dla } x \geqslant x_0. \end{cases}$$

Przecież to banalne!

- $\max(X_1, X_2, \ldots X_n) \leqslant x$ wtedy i tylko wtedy, gdy $X_1 \leqslant x$ i $X_2 \leqslant x$ i tak dalej, i w końcu $X_n \leqslant x$
- czyli $F_{M_n}(x) = F_X(x) \cdot F_X(x) \cdots F_X(x) = (F_X(x))^n$ (Założyliśmy że wszystkie X_i mają ten sam rozkład i przede wszystkim niezależność! Czyli prawdopodobieństwa się mnożą.)
- W takim razie, jeśli $F_X(x) < 1$ dla $x < x_0$ i $F_X(x) = 1$ dla $x \geqslant x_0$, to

$$\lim_{n\to\infty} F_{M_n}(x) = \begin{cases} 0 & \text{dla } x < x_0 \\ 1 & \text{dla } x \geqslant x_0. \end{cases}$$

- Ups! Zginęliśmy w osobliwości! (Jeśli rozkład X nie jest ograniczony i stosownego x_0 nie ma, to jest jeszcze gorzej — osobliwość ucieka do nieskończoności.)
- … czy to jest cały czas ten sam problem?

Normalizacja durniu! Normalizacja!

- Gdzieś to już widzieliśmy...
- Jeśli $S_n = X_1 + X_2 + \cdots + X_n$, to S_n też zwykle „uciekają do nieskończoności”.

Normalizacja durniu! Normalizacja!

- Gdzieś to już widzieliśmy...
- Jeśli $S_n = X_1 + X_2 + \cdots + X_n$, to S_n też zwykle „uciekają do nieskończoności”.
- Sytuację ratuje normalizacja!

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

(μ i σ^2 — średnia i wariancja rozkładu X)

Normalizacja durniu! Normalizacja!

- Gdzieś to już widzieliśmy...
- Jeśli $S_n = X_1 + X_2 + \cdots + X_n$, to S_n też zwykle „uciekają do nieskończoności".
- Sytuację ratuje normalizacja!

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

(μ i σ^2 — średnia i wariancja rozkładu X)
- Wówczas ciąg S_n^* ma rozkład graniczny $N(0,1)$.

Normalizacja durniu! Normalizacja!

- Gdzieś to już widzieliśmy...
- Jeśli $S_n = X_1 + X_2 + \cdots + X_n$, to S_n też zwykle „uciekają do nieskończoności".
- Sytuację ratuje normalizacja!

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

(μ i σ^2 — średnia i wariancja rozkładu X)

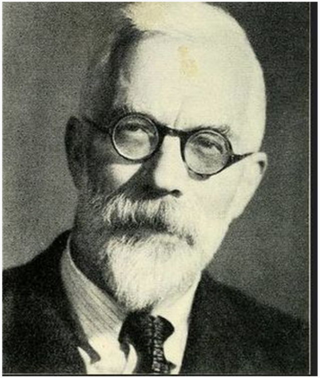
- Wówczas ciąg S_n^* ma rozkład graniczny $N(0, 1)$.
- W przypadku maksimów normalizacja nie jest tak oczywista jak dla sum...

Fisher-Tipett (1928), Gnedenko (1948)

Maksimum z próbki niezależnych zmiennych losowych po odpowiedniej normalizacji może zbiegać według rozkładu tylko do jednego z trzech rozkładów: Weibulla, Gumbela lub Frecheta.

Leonard Tipett, 1902–1985
angielski statystyk

Boris Gniedenko, 1912-1995
radziecki matematyk



Ronald Fisher, 1890-1962
angielski statystyk i biolog

Twierdzenie

Jeśli istnieją ciągi liczb a_n i b_n takie, że $a_n > 0$ oraz znormalizowany ciąg maksimów

$$M_n^* = \frac{M_n - b_n}{a_n}$$

jest zbieżny (według rozkładu), czyli istnieje niezdegenerowana dystrybuanta $H(x)$ taka, że

$$\lim_{n\to\infty} P\left(M_n^* \leqslant x\right) = H(x),$$

to H należy do jednej z trzech rodzin dystrybuant:

Weibulla, Gumbela lub Frecheta

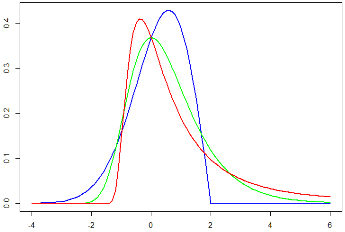
GEV, Generalized Extreme Value distribution

Rozkłady Weibulla, Gumbela i Frecheta można zebrać w **uogólniony rozkład wartości ekstremalych**

$$H_0(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \quad \text{dla} \quad \xi = 0,$$

oraz

$$H_\xi(x) = \exp\left(-\left(1+\xi\cdot\frac{x-\mu}{\sigma}\right)_+^{-1/\xi}\right) \quad \text{dla} \quad \xi \neq 0$$



$\xi < 0$ — rozkład Weibulla

$\xi = 0$ — rozkład Gumbela

$\xi > 0$ — rozkład Frecheta

- Pozornie konieczność normalizacji maksimów jest problemem. . .

Estymacja

- Pozornie konieczność normalizacji maksimów jest problemem…
- … ale dla danego n stałe normalizacyjne a_n i b_n można wchłonąć w μ i σ, które i tak trzeba estymować!

Estymacja

- Pozornie konieczność normalizacji maksimów jest problemem...
- ... ale dla danego n stałe normalizacyjne a_n i b_n można wchłonąć w μ i σ, które i tak trzeba estymować!
- Zatem jeśli tylko M_n^* ma rozkład zbliżony do GEV, to rozkład M_n też możemy przybliżać przy pomocy GEV tylko o innych parametrach (parametr kształtu ξ jest ten sam!).

Estymacja

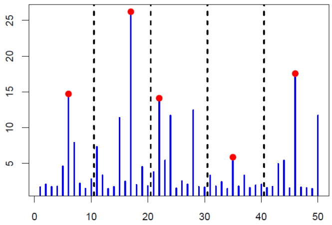
- Pozornie konieczność normalizacji maksimów jest problemem. . .
- . . . ale dla danego n stałe normalizacyjne a_n i b_n można wchłonąć w μ i σ, które i tak trzeba estymować!
- Zatem jeśli tylko M_n^* ma rozkład zbliżony do GEV, to rozkład M_n też możemy przybliżać przy pomocy GEV tylko o innych parametrach (parametr kształtu ξ jest ten sam!).

  (Tyle tylko, że w przypadku M_n^* to jest ten sam rozklad, coraz lepszy gdy n rośnie, a w przypadku M_n rozkład „pływa". Ale to zwykle nie przeszkadza.)

Metoda maximów blokowych

- Dane dzielimy na m niezachodzących na siebie bloków, każdy o równej długości n – na przykład dane miesięczne czy roczne.
- Z każdego bloku wybieramy wartość największą: $M_{n1}, M_{n2}, \ldots M_{nm}$.
- Na podstawie uzyskanych m niezależnych obserwacji, estymujemy parametry ξ, μ i σ rozkładu GEV, na przykład metodą największej wiarygodności.

S. Coles — 3.3.2, s. 55
Maximum Likelihood Estimation (MLE, metoda największej wiarygodności)



Kwantyle ekstremalne

- Po dopasowaniu GEV zwykle najbardziej interesujące jest obliczenie stosownego kwantyla z_α rzędu $\alpha = 1 - p$, zwykle dla małych wartosci p.

Kwantyle ekstremalne

- Po dopasowaniu GEV zwykle najbardziej interesujące jest obliczenie stosownego kwantyla z_α rzędu $\alpha = 1 - p$, zwykle dla małych wartosci p.
- Odwrócenie dystrybuanty GEV daje wzór:

$$z_\alpha = \begin{cases} \mu - \frac{\sigma}{\xi}\left(1 - (-\ln\alpha)^{-\xi}\right) & \text{dla } \xi \neq 0 \\ \mu - \sigma \ln(-\ln\alpha) & \text{dla } \xi = 0. \end{cases}$$

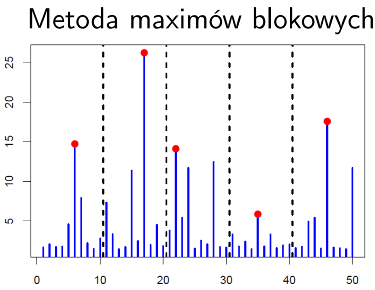
Kwantyle ekstremalne

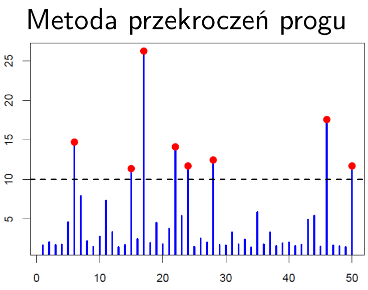
- Po dopasowaniu GEV zwykle najbardziej interesujące jest obliczenie stosownego kwantyla z_α rzędu $\alpha = 1 - p$, zwykle dla małych wartosci p.
- Odwrócenie dystrybuanty GEV daje wzór:

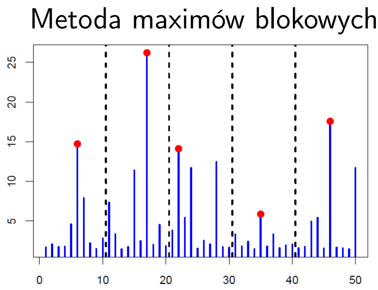
$$z_\alpha = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - (-\ln \alpha)^{-\xi}) & \text{dla } \xi \neq 0 \\ \mu - \sigma \ln(-\ln \alpha) & \text{dla } \xi = 0. \end{cases}$$

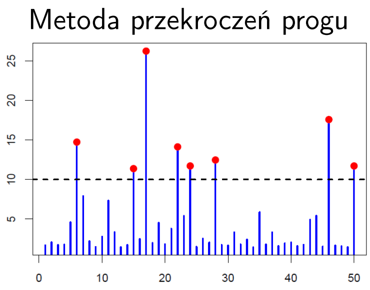
- Interpretacja jest następująca: na przyklad dla $p = 5\%$ w kontekście opadów, wartość odpowiedniego kwantyla $z_{95\%}$ to taki poziom, który będzie przekraczany średnio co 20 lat.

Estymacja ogona inaczej — Peaks Over Threshold (POT)









Dla zmiennej losowej X o dystrybuancie F i ustalonego progu $u > 0$ definiujemy *nadwyżkę* jako $Y = X - u$. Dystrybuanta *nadwyżki* dana jest wzorem

$$F_u(y) = P(X - u \leqslant y | X > u)$$

Uogólniony rozkład Pareto (GPD)

Twierdzenie udowodnione niezależnie przez Balkema i de Haana (1974) oraz Pickandsa (1975).

Uogólniony rozkład Pareto (GPD)

Twierdzenie udowodnione niezależnie przez Balkema i de Haana (1974) oraz Pickandsa (1975).

Twierdzenie
Dystrybuanty maximów $M_n = \max(X_1, \dots, X_n)$ są zbieżne do dystrybuanty GEV z parametrem kształtu ξ wtedy i tylko wtedy, gdy

$$\lim_{u\to\infty} \sup_x |F_u(x) - G_{\xi,\beta(u)}(x)| = 0.$$

dla pewnej funkcji β zależnej od u, gdzie $G_{\xi,\beta}$ jest dystrybuantą uogólnionego rozkladu Pareto:

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \frac{\xi \cdot x}{\beta})^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp(-\frac{x}{\beta}) & \xi = 0 \end{cases}$$

Metoda POT — co z tym progiem?

- Za niski próg może dawać za dużo danych pochodzących nie z ogona rozkładu, co zwiększy obciążenia estymatorów.

Metoda POT — co z tym progiem?

- Za niski próg może dawać za dużo danych pochodzących nie z ogona rozkładu, co zwiększy obciążenia estymatorów.
- Wybór zbyt dużego progu spowoduje, że pozostanie nam za mało danych i wpłynie na wariancję.

Metoda POT — co z tym progiem?

- Za niski próg może dawać za dużo danych pochodzących nie z ogona rozkładu, co zwiększy obciążenia estymatorów.
- Wybór zbyt dużego progu spowoduje, że pozostanie nam za mało danych i wpłynie na wariancję.
- Wielu autorów sugeruje taki wybór progu, aby liczba pozostałych danych była nie większa niż 10-15%. Często wybierany jest próg tak aby liczba ta była rzędu 5-10%.

Metoda POT — co z tym progiem?

- Za niski próg może dawać za dużo danych pochodzących nie z ogona rozkładu, co zwiększy obciążenia estymatorów.
- Wybór zbyt dużego progu spowoduje, że pozostanie nam za mało danych i wpłynie na wariancję.
- Wielu autorów sugeruje taki wybór progu, aby liczba pozostałych danych była nie większa niż 10-15%. Często wybierany jest próg tak aby liczba ta była rzędu 5-10%.

Pomocny może być ciekawy fakt. Jeśli przez $e(u)$ oznaczyć wartość średnią z przekroczeń progu:

$$e(u) = E(\underbrace{X - u}_{Y} | X > u),$$

to jeśli dla pewnego u rozkład nadwyżek dobrze aproksymuje się GPD, to dla progów $v > u$, $e(v)$ jest linową funkcją v.

A dalej jest nieskończoność. . .

- Mocno upraszczaliśmy! Co na przyklad w sytuacji, gdy założenie o niezależności jest trudne do utrzymania? Są dostępne uogólnienia:
  - procesy stacjonarne
  - pewne procesy niestacjonarne, model sezonowy
  - procesy punktowe

A dalej jest nieskończoność. . .

- Mocno upraszczaliśmy! Co na przyklad w sytuacji, gdy założenie o niezależności jest trudne do utrzymania? Są dostępne uogólnienia:
  - procesy stacjonarne
  - pewne procesy niestacjonarne, model sezonowy
  - procesy punktowe
- Ekstrema wektorów losowych — włączenie modelowania wzajemnych zależności. Także kopuły ekstremalne (książka QRM).

A dalej jest nieskończoność. . .

- Mocno upraszczaliśmy! Co na przyklad w sytuacji, gdy założenie o niezależności jest trudne do utrzymania? Są dostępne uogólnienia:
  - procesy stacjonarne
  - pewne procesy niestacjonarne, model sezonowy
  - procesy punktowe
- Ekstrema wektorów losowych — włączenie modelowania wzajemnych zależności. Także kopuły ekstremalne (książka QRM).
- Metody bayesowskie.

A dalej jest nieskończoność. . .

- Mocno upraszczaliśmy! Co na przyklad w sytuacji, gdy założenie o niezależności jest trudne do utrzymania? Są dostępne uogólnienia:
  - procesy stacjonarne
  - pewne procesy niestacjonarne, model sezonowy
  - procesy punktowe
- Ekstrema wektorów losowych — włączenie modelowania wzajemnych zależności. Także kopuły ekstremalne (książka QRM).
- Metody bayesowskie.
- Mało danych? Bootstrap? (To trochę moja fantazja, ale chyba widziałem gdzieś już takie prace.)

A dalej jest nieskończoność. . .

- Mocno upraszczaliśmy! Co na przyklad w sytuacji, gdy założenie o niezależności jest trudne do utrzymania? Są dostępne uogólnienia:
  - procesy stacjonarne
  - pewne procesy niestacjonarne, model sezonowy
  - procesy punktowe
- Ekstrema wektorów losowych — włączenie modelowania wzajemnych zależności. Także kopuły ekstremalne (książka QRM).
- Metody bayesowskie.
- Mało danych? Bootstrap? (To trochę moja fantazja, ale chyba widziałem gdzieś już takie prace.)
- . . .

Literatura

- https://en.wikipedia.org/wiki/Extreme_value_theory
- Coles S. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London 2001
  Dobrze napisana, wyważone proporcje teorii i praktycznych przykładów z rożnych dziedzin.
- Alexander J. McNeil, Rudiger Frey, Paul Embrechts, *Quantitative Risk Management. Concepts, Techniques and Tools*. Princeton University Press 2005
  Biblia w dziedzinie zarządzania ryzykiem w finansach. Jest tam rozdział o EVT.
- Embrechts P., Klüppelberg C. and Mikosch T. *Modelling extremal events for insurance and finance*. Springer Berlin 1997
  Teoretyczne, czyli są tam wszystkie dowody, chyba…
- Ciekawy artykuł, który zwrócił moją uwagę:
  Coles S., Pericchi L., *Aniticipating Catastrophies Through Extreme Value Modeling*, Journal of the Royal Statistical Society: Series C (Applied Statistics), Vol. 52.4 2003