

Ryszard Magiera

**MODELE I METODY
STATYSTYKI
MATEMATYCZNEJ**

**Część I
ROZKŁADY I SYMULACJA
STOCHASTYCZNA**

Wydanie III rozszerzone



Oficyna Wydawnicza GiS
Wrocław 2018

ryszard.magiera@pwr.edu.pl

Autor projektu okładki *Ryszard Magiera*.

Copyright © 2002, 2005, 2018 by Ryszard Magiera

All rights reserved. No part of this book may be translated or reproduced in any form without written permission of the Copyright owner.

Printed in Poland

Skład komputerowy książki w systemie $\text{\LaTeX}$ wykonał autor. Do sporządzenia wykresów wykorzystano procedury pakietu komputerowego *Mathematica* 11.2 (licencja L4953-2950).

ISBN 978-83-62780-56-3

Wydanie III, Wrocław 2018
Oficyna Wydawnicza GiS, s.c.
Druk i oprawa: Drukarnia I-BiS sp. z o.o., A.Bieroński, P.Bieroński s.j.

Przedmowa

Książka jest pierwszą częścią dwutomowego kompendium z zakresu statystyki matematycznej pod tytułem *Modele i metody statystyki matematycznej*. Zawiera podstawowe definicje, twierdzenia i fakty z teorii prawdopodobieństwa, tworząc podstawową bazę do opisu modeli stochastycznych i formułowania metod wnioskowania statystyki matematycznej opisanych w drugim tomie – *Modele i metody statystyki matematycznej. Część II. Wnioskowanie statystyczne* [28].

Dużą część książki stanowią definicje i podstawowe własności rozkładów prawdopodobieństwa mających zastosowania w analizach statystycznych w różnych dziedzinach nauki i techniki. Podano wiele twierdzeń dotyczących zbieżności zmiennych losowych i rozkładów. Twierdzenia te oraz przedstawiony obszerny zbiór twierdzeń dotyczących nierówności stochastycznych wykorzystywane są w analizie stochastycznej i konstrukcji narzędzi wnioskowania statystycznego. Opisano metodę Monte Carlo obliczania całek oraz algorytmy Metropolisa-Hastingsa i Gibbsa generowania zmiennych losowych tworzących stacjonarny łańcuch Markowa. Teoretyczne podstawy metody Monte Carlo opartej na łańcuchu Markowa zostały podane w tematach dotyczących łańcucha Markowa i jego własności. Przedstawiono podstawowe metody symulacji zmiennych losowych. W opisach większości tematów podano przykłady ilustrujące przedstawione pojęcia, twierdzenia i metody. Podano przykłady wykorzystania języka programowania pakietu komputerowego *Mathematica*.

Poszczególne wpisy ułożone są w kolejności alfabetycznej. Dla uzyskania spójności tematów dotyczących *Metody Monte Carlo* oraz *Symulacji zmiennej losowej*, do wiodącej nazwy tematu wprowadzono dodatkowo znacznik cyfrowy.

W celu ułatwienia korzystania z literatury w języku angielskim oraz z angielskich wersji komputerowych pakietów statystycznych, polskie nazwy ważnych pojęć uzupełniono terminologią angielską. W tym celu został również dołączony skorowidz terminów angielskich.

Rozkłady i symulacja stochastyczna stanowiły Rozdział I książki *Modele i metody statystyki matematycznej* wydanej w 2002 r. Po pewnych uzupełnieniach rozdział ten został opracowany w oddzielnym tomie jako Wydanie II książki *Modele i metody statystyki matematycznej. Część I. Rozkłady i symulacja stochastyczna*, która ukazała się w 2005 r. Obecne wydanie tej książki zostało znacznie zmienione i rozszerzone w porównaniu do poprzednich wydań. Poprawiono i uzupełniono tekst wielu tematów oraz dodano nowe tematy, m.in. *Rozkład funkcji wektora losowego*, *Twierdzenie Bayesa*, *Rozwinięcie Edgewortha*, *Rozkład Frecheta*, *Rozkład Landau'a*, *Rozkład Lévy'ego*, *Rozkład wartości ekstremalnej maksymalnej, uogólniony*, *Rozkład wartości ekstremalnej minimalnej, uogólniony*, *Rozkłady ciężko ogonowe*.

W końcowej części książki umieszczono tablice zawierające oznaczenia i spis podstawowych charakterystyk rozkładów najczęściej rozpatrywanych w praktyce. W spisie literatury zostały umieszczone pozycje, z których korzystałem przy opracowywaniu tematów zawartych w niniejszej książce.

Wśród licznej i różnorodnej literatury poświęconej teorii prawdopodobieństwa i statystyce matematycznej, niniejsza książka stanowić może dodatkową pomoc, spełniając jednocześnie dodatkowe funkcje wynikające z zastosowanej formy prezentacji tematów.

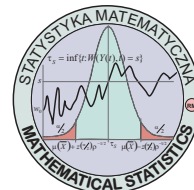
Książka adresowana jest przede wszystkim do środowiska akademickiego, w szczególności do studentów matematyki, statystyki, ekonometrii, fizyki i niektórych dziedzin inżynierskich. Zastosowany sposób prezentacji tematów, jak również przejrzysty i precyzyjny opis tematów, mają na celu zwiększenie jej użyteczności dla szerszego kręgu odbiorców - przedstawicieli różnych dyscyplin nauki i techniki.

Wrocław, wrzesień 2018

Ryszard Magiera

Spis treści

Przedmowa	v
Spis oznaczeń	viii
Spis tabel	xiv
Spis wykresów	xv
1 Rozkłady i symulacja stochastyczna	1
Tablice rozkładów	226
Literatura	234
Skorowidz terminów angielskich	239



Rozkłady i symulacja stochastyczna

Distributions and Stochastic Simulation

Algorytm Gibbsa $\triangleright$ metoda Monte Carlo (2) oparta na łańcuchu Markowa (2)

Algorytm Metropolisa-Hastingsa $\triangleright$ metoda Monte Carlo (2) oparta na łańcuchu Markowa (1)

Centralne twierdzenie graniczne (CTG) [central limit theorem (CLT)]

CTG odgrywa podstawową rolę w asymptotycznej teorii wnioskowań statystycznych.

Niech (X_j) , $j \in \mathbb{N}$, będzie ciągiem niezależnych zmiennych losowych o rozkładach określonych przez dystrybuanty $F_{X_j}(x) = P(X_j \leq x)$, dla których istnieją skończone wartości oczekiwane $E(X_j) = \mu_j$ i wariancje $\text{Var}(X_j) = \sigma_j^2$, przy czym $B_n^2 := \sum_{j=1}^n \sigma_j^2 > 0$ (tzn. co najmniej jedna ze zmiennych $X_1, \dots, X_n$ ma niezdegenerowany rozkład). Sumę unormowaną [normed sum] zmiennych losowych $X_1, \dots, X_n$ nazywamy zmienną losową

$$Y_n = \frac{\sum_{j=1}^n (X_j - \mu_j)}{B_n},$$

dla której $E(Y_n) = 0$ i $\text{Var}(Y_n) = 1$ dla każdego $n \in \mathbb{N}$.

Niech $F_{Y_n}(y)$ oznacza dystrybuantę sumy unormowanej Y_n , a

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y \exp\left(-\frac{t^2}{2}\right) dt$$

– dystrybuantę $\triangleright$ rozkładu normalnego $\mathcal{N}(0, 1)$. Przy założeniu skończonych wariancji σ_k^2 , $k = 1, 2, \dots$, CTG podaje warunki, przy których

$$\lim_{n \rightarrow \infty} F_{Y_n}(y) = \Phi(y) \quad (1.1)$$

Centralne twierdzenie graniczne

dla każdego $y \in \mathbb{R}$ (z Twierdzenia 1.66 (Pólya) wynika, że zbieżność ta jest jednostajna względem $y \in \mathbb{R}$). Zbieżność (1.1) oznacza zbieżność według rozkładu ($\triangleright$ *rodzaje zbieżności zmiennych losowych*) do zmiennej losowej o rozkładzie $\mathcal{N}(0, 1)$ i zapisuje się ją symbolicznie $Y_n \xrightarrow{D} \mathcal{N}(0, 1)$. Równoważnie oznacza to, że zmienna losowa $\sum_{j=1}^n X_j$ ma $\triangleright$ *rozkład asymptotycznie normalny* $\mathcal{AN}\left(\sum_{j=1}^n \mu_j, B_n^2\right)$, a zmienna losowa $\bar{X} := \frac{1}{n} \sum_{j=1}^n X_j$ (często używana w statystyce) ma rozkład asymptotycznie normalny $\mathcal{AN}\left(\frac{1}{n} \sum_{j=1}^n \mu_j, B_n^2/n^2\right)$.

TWIERDZENIE 1.1 (Lindeberga-Lévy’ego) [Lindeberg-Lévy theorem]

Jeżeli (X_j) , $j \in \mathbb{N}$, jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie z wartością oczekiwaną $\mu = E(X_j)$ i skończoną wariancją $\sigma^2 = \text{Var}(X_j)$, to dla dystrybuant $F_{Y_n}(y)$ sum unormowanych

$$Y_n = \frac{\sum_{j=1}^n X_j - n\mu}{\sigma\sqrt{n}}$$

zachodzi zbieżność (1.1). $\square$

Ważnym przypadkiem szczególnym twierdzenia Lindeberga-Lévy’ego jest $\triangleright$ *twierdzenie Moivre’a-Laplace’a*.

Dla zmiennych losowych o różnych rozkładach zachodzi następujące twierdzenie graniczne, w którym sformułowane warunki są dość proste do sprawdzenia.

TWIERDZENIE 1.2 (Lapunowa) [Liapunov’s theorem] Jeżeli dla ciągu niezależnych zmiennych losowych (X_j) , $j \in \mathbb{N}$, istnieje liczba $\delta > 0$, taka że

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^{2+\delta}} \sum_{j=1}^n E(|X_j - \mu_j|^{2+\delta}) = 0, \quad (1.2)$$

to dla dystrybuant $F_{Y_n}(y)$ sum unormowanych

$$Y_n = \frac{\sum_{j=1}^n (X_j - \mu_j)}{B_n}$$

zachodzi zbieżność (1.1). $\square$

Warunek (1.2) nazywa się *warunkiem Lapunowa* [Liapunov’s condition].

PRZYKŁAD 1.3 Niech $X_1, X_2, \dots$, będą niezależnymi zmiennymi losowymi, przy czym $X_j \sim \mathcal{B}(1, p_j)$. Załóżmy, że $B_n^2 = \sum_{j=1}^n \text{Var}(X_j) = \sum_{j=1}^n p_j(1 - p_j) \rightarrow \infty$, gdy $n \rightarrow \infty$. Mamy $E(X_j) = p_j$ oraz

$$E(|X_j - p_j|^3) = (1 - p_j)^3 p_j + p_j^3 (1 - p_j)^3 \leq 2p_j(1 - p_j).$$

Zatem

$$\sum_{j=1}^n E(|X_j - p_j|^3) \leq 2 \sum_{j=1}^n p_j(1 - p_j) = 2B_n^2,$$

tzn. warunek Lapunowa (1.2) spełniony jest dla $\delta = 1$. Na podstawie Twierdzenia 1.2,

$$\frac{1}{B_n} \sum_{j=1}^n (X_j - p_j) \xrightarrow{D} \mathcal{N}(0, 1).$$

□

Najbardziej ogólnym warunkiem, przy którym zachodzi zbieżność (1.1) dla ciągu (X_j) , $j \in \mathbb{N}$, niezależnych zmiennych losowych mających skończoną wariancję jest *warunek Lindeberga*

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{j=1}^n \int_{|x - \mu_j| > \varepsilon B_n} (x - \mu_j)^2 dF_{X_j}(x) = 0 \quad \forall \varepsilon > 0. \quad (1.3)$$

Warunek Lindeberga (1.3) wynika z warunku Lapunowa (1.2).

TWIERDZENIE 1.4 (Lindeberga-Fellera) [Lindeberg-Feller theorem]
Niech (X_j) , $j \in \mathbb{N}$, będzie ciągiem niezależnych zmiennych losowych, dla których $0 < B_n^2 = \sum_{j=1}^n \sigma_j^2 < \infty$, $n \in \mathbb{N}$. Warunkiem koniecznym i wystarczającym na to, aby spełniony był *warunek jednostajnej asymptotycznej zaniedbywalności* [**uniformly asymptotic negligibility condition**]

$$\lim_{n \rightarrow \infty} \frac{\max_{1 \leq j \leq n} \sigma_j^2}{B_n^2} = 0 \quad (1.4)$$

oraz zachodziła zbieżność (1.1) dystrybuant F_{Y_n} sum unormowanych

$$Y_n = \frac{\sum_{j=1}^n (X_j - \mu_j)}{B_n},$$

jest spełnienie warunku Lindeberga (1.3). □

Warunek Lindeberga jest spełniony, jeśli rozpatrywane zmienne losowe mają jednakowy rozkład o skończonej wariancji. Warunek (1.4) pociąga za sobą warunek $B_n^2 \rightarrow \infty$. Z Twierdzenia 1.4 wynika zatem, że jeśli spełniony jest warunek Lindeberga, to $B_n^2 \rightarrow \infty$.

Twierdzenia przedstawione powyżej dotyczą asymptotycznej normalności sum $\sum_{j=1}^n X_j$ utworzonych z elementów $X_1, \dots, X_n$ pojedynczego ciągu zmiennych losowych (X_j) , $j \in \mathbb{N}$. Ogólniej można rozważać dwuwymiarową tablicę zmiennych losowych:

$$\begin{array}{cccc} X_{11}, & X_{12}, & \dots & X_{1k_1}; \\ X_{21}, & X_{22}, & \dots & X_{2k_2}; \\ \dots & \dots & \dots & \dots \\ X_{n1}, & X_{n2}, & \dots & X_{nk_n}; \\ \dots & \dots & \dots & \dots \end{array}$$

[double sequence of random variables]. Mówimy również, że mamy do czynienia z ciągiem serii zmiennych losowych; pierwszy indeks wskazuje numer serii, drugi – numer zmiennej w danej serii. W serii (wierszu) o numerze n występuje k_n zmiennych losowych X_{nj} , $j = 1, \dots, k_n$. W przypadku $k_n = n$ rozpatrywana tablica sprowadza się do tablicy trójkątnej. Zakłada się, że $k_n \rightarrow \infty$, gdy $n \rightarrow \infty$. Ponadto zakłada się niezależność zmiennych losowych w każdej serii (czyli niezależność wewnątrz wierszy), tzn. dla każdego $n \in \mathbb{N}$ zmienne losowe $X_{n1}, X_{n2}, \dots, X_{nk_n}$ są niezależne.

Oznaczmy $\mu_{nj} = E(X_{nj})$ oraz $B_n^2 = \sum_{j=1}^{k_n} \text{Var}(X_{nj})$.

TWIERDZENIE 1.5 Niech $(X_{nj}, j = 1, \dots, k_n; n = 1, 2, \dots)$ będzie dwuwymiarową tablicą niezależnych wewnątrz wierszy zmiennych losowych, dla których $0 < B_n^2 < \infty$, $n = 1, 2, \dots$. Warunkiem koniecznym i wystarczającym na to, aby spełniony był warunek jednostajnej asymptotycznej zaniedbywalności

$$\lim_{n \rightarrow \infty} \max_{1 \leq j \leq k_n} P(|X_{nj} - \mu_{nj}| > \varepsilon B_n) = 0 \quad \forall \varepsilon > 0$$

oraz zachodziła zbieżność

$$\frac{1}{B_n} \sum_{j=1}^{k_n} (X_{nj} - \mu_{nj}) \xrightarrow{D} \mathcal{N}(0, 1),$$

jest spełnienie warunku

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{j=1}^{k_n} E[(X_{nj} - \mu_{nj})^2 \mathbf{1}_{\{|X_{nj} - \mu_{nj}| > \varepsilon B_n\}}] = 0 \quad \forall \varepsilon > 0. \quad (1.5)$$

□

Innymi słowy, przy założeniach Twierdzenia 1.5,

$$\sum_{j=1}^{k_n} X_{nj} \sim \mathcal{AN}\left(\sum_{j=1}^{k_n} \mu_{nj}, B_n^2\right).$$

Warunek (1.5) jest wersją warunku Lindeberga (1.3). Wynika on z następującej wersji warunku Lapunowa

$$\sum_{j=1}^{k_n} |X_{nj} - \mu_{nj}|^{2+\delta} = o(B_n^{2+\delta})$$

dla pewnego $\delta > 0$.

Twierdzenie Lindeberga-Fellera jest szczególnym przypadkiem Twierdzenia 1.5.

Następujące twierdzenie jest wnioskiem z Twierdzenia 1.5.

TWIERDZENIE 1.6 (wielowymiarowe CTG) [multivariate CLT] Niech $\mathbf{X}_i = (X_{i1}, \dots, X_{ik})^T$, $i = 1, 2, \dots$, będą niezależnymi wektorami losowymi

Centralne twierdzenie graniczne dla losowo indeksowanych sum

o jednakowym rozkładzie, takim że $E\|\mathbf{X}_1\|^2 < \infty$. Niech $E(\mathbf{X}_1) = \boldsymbol{\mu}$ oraz $\boldsymbol{\Sigma} = \text{Var}(\mathbf{X}_1)$. Wtedy

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^n \mathbf{X}_i - n\boldsymbol{\mu} \right) \xrightarrow{D} \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma}).$$

□

Przy założeniach Twierdzenia 1.6, dla każdej ciągłej funkcji $h : \mathbb{R}^k \mapsto \mathbb{R}$ zachodzi zbieżność

$$h \left(\frac{\sum_{i=1}^n \mathbf{X}_i - n\boldsymbol{\mu}}{\sqrt{n}} \right) \xrightarrow{D} h(\mathbf{Z}),$$

gdzie $\mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma})$.

Z Twierdzenia 1.6 wynika następujący ważny wniosek.

WNIOSEK 1.7 Niech spełnione będą założenia Twierdzenia 1.6. Jeżeli macierz $\boldsymbol{\Sigma}$ jest dodatnio określona, to

$$n(\bar{\mathbf{X}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}) \xrightarrow{D} \chi^2(k)$$

(zbieżność do zmiennej losowej o $\triangleright$ rozkładzie $\chi^2(k)$), gdzie $\bar{\mathbf{X}} = (1/n) \sum_{i=1}^n \mathbf{X}_i$. □

Centralne twierdzenie graniczne dla losowo indeksowanych sum [central limit theorem (CLT) for randomly indexed sums, random sum central limit theorem] Jeżeli (X_i) , $i \in \mathbb{N}$, jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie z wartością oczekiwaną μ i wariancją $\sigma^2 > 0$, to klasyczne $\triangleright$ *centralne twierdzenie graniczne* orzeka, że ciąg sum unormowanych $(S_n - n\mu)/(\sigma\sqrt{n})$, gdzie $S_n = \sum_{i=1}^n X_i$, jest zbieżny według rozkładu do zmiennej losowej o rozkładzie $\mathcal{N}(0, 1)$. Następujące twierdzenie podtrzymuje tę tezę w przypadku, gdy n zostaje zastąpione zmienną losową N_t o wartościach całkowitych, jeżeli ciąg (N_t/t) ($t \rightarrow \infty$) zbieżny jest według prawdopodobieństwa do pewnej skończonej stałej dodatniej.

TWIERDZENIE 1.8 (Anscombe’a) [Anscombe’s theorem] Niech (X_n) , $n \in \mathbb{N}$, będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie z wartością oczekiwaną μ i wariancją $\sigma^2 > 0$. Ponadto, niech (N_t) , $t \geq 0$, będzie ciągiem zmiennych losowych przyjmujących dodatnie wartości całkowite i takim, że $\frac{N_t}{t} \xrightarrow{P} c$, gdy $t \rightarrow \infty$, gdzie $0 < c < \infty$. Wówczas $\frac{S_{N_t} - N_t\mu}{\sigma\sqrt{N_t}} \xrightarrow{D} \mathcal{N}(0, 1)$ oraz $\frac{S_{N_t} - N_t\mu}{\sigma\sqrt{tc}} \xrightarrow{D} \mathcal{N}(0, 1)$, gdy $t \rightarrow \infty$. □

Twierdzenie 1.8 przedstawia jedną z wielu wersji centralnego twierdzenia granicznego dla sum o losowej liczbie składników (dla losowo indeksowanych sum) i nazywa się również **twierdzeniem Anscombe’a-Rényi’ego** [Anscombe-Rényi theorem].

Chwila zatrzymania

Twierdzenie 1.9 Niech (X_n) , $n \in \mathbb{N}$, będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie z wartością oczekiwaną $\mu > 0$ i wariancją $\sigma^2 > 0$. Ponadto, niech

$$N_t = N_t(\alpha) := \inf\{n \geq 1 : S_n > tn^\alpha\}, \quad \alpha \in (-\infty, 1), c > 0.$$

Wówczas

$$\frac{\mu(1-\alpha)[N_t(\alpha) - (t/\mu)^{1/(1-\alpha)}]}{\sigma(t/\mu)^{1/2(1-\alpha)}} \xrightarrow{D} \mathcal{N}(0, 1), \text{ gdy } t \rightarrow \infty.$$

□

WNIOSEK 1.10 Przy założeniach Twierdzenia 1.9, dla $\alpha \in (-\infty, 1)$,

$$\frac{\max_{1 \leq j \leq n} (S_j/j^\alpha) - \mu n^{1-\alpha}}{\sigma n^{(1/2)-\alpha}} \xrightarrow{D} \mathcal{N}(0, 1).$$

□

WNIOSEK 1.11 Niech (X_n) , $n \in \mathbb{N}$, będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie z wartością oczekiwaną $\mu > 0$ i wariancją $\sigma^2 > 0$. Ponadto, niech

$$N_t := \sup\{n \geq 1 : S_n \leq t\}, \quad t \geq 0.$$

Wówczas

$$\frac{N_t - t/\mu}{\sigma t^{1/2} \mu^{-3/2}} \xrightarrow{D} \mathcal{N}(0, 1), \text{ gdy } t \rightarrow \infty.$$

□

Chwila zatrzymania $\triangleright$ *martyngał*

Czas życia $\triangleright$ *funkcja intensywności awarii*, $\triangleright$ *rozkłady czasu życia; klasyfikacja*, $\triangleright$ *model losowego obcinania*

Eksces [excess, excess kurtosis] *Ekscesem* lub *współczynnikiem ekscesu* [coefficient of excess] zmiennej losowej X (rozkładu określonego przez dystrybuantę $F(x)$) nazywamy liczbę

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\kappa_4}{\kappa_2^2},$$

gdzie μ_2 i μ_4 oznaczają momenty centralne, odpowiednio rzędu drugiego i czwartego, rozkładu zmiennej losowej X , natomiast κ_2 i κ_4 są kumulantami, odpowiednio rzędu drugiego i czwartego, tego rozkładu ($\triangleright$ *funkcja tworząca kumulanty*). Jak widać z określenia, eksces jest bezpośrednio związany z $\triangleright$ *kurtozą* β_2 . Mianowicie, $\gamma_2 = \beta_2 - 3$. Ponieważ współczynnik γ_2 dla rozkładu normalnego równa się zero, może on służyć jako miara rozbieżności (związanej ze stopniem koncentracji w otoczeniu wartości modalnej) danego rozkładu od rozkładu normalnego o tej samej wartości oczekiwanej i wariancji. Wartość

dodatnia tego współczynnika wskazuje na zwiększoną koncentrację w stosunku do rozkładu normalnego. Inaczej mówiąc, współczynnik γ_2 może służyć jako miara porównawcza stopnia względnego wznoszenia krzywej gęstości danego rozkładu w otoczeniu wartości modalnej, w odniesieniu do krzywej gęstości rozkładu normalnego o tej samej wartości oczekiwanej i wariancji. Dlatego eksces nazywany jest również **współczynnikiem spłaszczenia**.

Entropia zmiennej losowej [entropy of a random variable] Niech X będzie zmienną losową o rozkładzie określonym przez dystrybuantę $F(x)$. *Entropią zmiennej losowej* X nazywamy wartość

$$H(f) = -E(\log_2 f(X)) = -\int_{-\infty}^{\infty} \log_2 f(x) dF(x)$$

$$= \begin{cases} -\int_{-\infty}^{\infty} f(x) \log_2 f(x) dx, & \text{gdy } X \text{ jest typu ciągłego o gęstości } f(x), \\ -\sum_x f(x) \log_2 f(x), & \text{gdy } X \text{ jest typu dyskretnego,} \end{cases}$$

$$f(x) = P(X = x)$$

(o ile całka istnieje lub szereg jest zbieżny). Zakłada się, że $0 \cdot \log_2 0 = 0$. W klasie wszystkich rozkładów ciągłych o wartości oczekiwanej 0 i wariancji 1, rozkład normalny $\mathcal{N}(0, 1)$ ma największą entropię. Pojęcie entropii pojawia się przede wszystkim w teorii informacji.

Entropię zmiennej losowej X definiuje się również przez $H(f) = -E(\log f(X))$, tzn. w oparciu o logarytm naturalny.

Fraktyl $\triangleright$ *kwantyl*

Funkcja beta [beta function] *Funkcja beta* $(\alpha, \beta) \mapsto B(\alpha, \beta)$ określona jest przez całkę Eulera pierwszego rodzaju

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx, \quad \alpha > 0, \beta > 0.$$

Inne użyteczne przedstawienie analityczne funkcji beta to

$$B(\alpha, \beta) = \int_0^{\infty} \frac{y^{\alpha-1}}{(1+y)^{\alpha+\beta}} dy.$$

Funkcja beta posiada następujące własności:

$$\begin{aligned} B(\alpha, \beta) &= B(\beta, \alpha); \\ B(\alpha, \beta + 1) &= \frac{\beta}{\alpha} B(\alpha + 1, \beta) = \frac{\beta}{\alpha + \beta} B(\alpha, \beta), \quad \alpha > 0, \beta > 0; \\ B(m, n) &= \frac{(n-1)!(m-1)!}{(m+n-1)!}, \quad m, n \in \mathbb{N}; \\ B(\alpha, 1-\alpha) &= \frac{\pi}{\sin \alpha\pi}, \quad 0 < \alpha < 1; \\ B(1/2, 1/2) &= \pi. \end{aligned} \tag{1.6}$$

Funkcja beta, niekompletna

Zachodzi następujący związek między funkcją beta a $\triangleright$ *funkcją gamma*:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

Funkcja

$$B_x(\alpha, \beta) = \int_0^x u^{\alpha-1}(1-u)^{\beta-1}du, \quad 0 \leq x \leq 1,$$

nazywa się **niekompletną funkcją beta** [**incomplete beta function**].

Oczywiście $B_1(\alpha, \beta) = B(\alpha, \beta)$.

Funkcję

$$I_x(\alpha, \beta) = \frac{B_x(\alpha, \beta)}{B(\alpha, \beta)} \quad 0 \leq x \leq 1,$$

nazywamy *regularyzowaną niekompletną funkcją beta* [**regularized incomplete beta function**]. Nazywa się ją również *ilorazem niekompletnej funkcji beta* [**incomplete beta function ratio**].

Regularyzowana niekompletna funkcja beta równa jest dystrybucie zmiennej losowej o $\triangleright$ *rozkładzie beta* $\mathcal{Be}(\alpha, \beta)$ oraz związana jest z dystrybucją zmiennej losowej o $\triangleright$ *rozkładzie dwumianowym* i z dystrybucją zmiennej losowej o $\triangleright$ *rozkładzie dwumianowym, ujemnym*.

Niektóre własności regularyzowanej niekompletnej funkcji beta:

$$\begin{aligned} I_0(\alpha, \beta) &= 0, \quad I_1(\alpha, \beta) = 1, \\ I_x(\alpha, 1) &= x^\alpha, \quad I_x(1, \beta) = 1 - (1-x)^\beta, \\ I_x(\alpha, \beta) &= 1 - I_{1-x}(\beta, \alpha), \\ I_x(\alpha + 1, \beta) &= I_x(\alpha, \beta) - \frac{x^\alpha(1-x)^\beta}{\alpha B(\alpha, \beta)}, \\ I_x(\alpha, \beta + 1) &= I_x(\alpha, \beta) + \frac{x^\alpha(1-x)^\beta}{\beta B(\alpha, \beta)}, \\ I_x(m, n - m + 1) &= \sum_{j=m}^n \binom{n}{j} x^j (1-x)^{n-j}, \quad m, n \in \mathbb{N}, x \in [0, 1), \quad (1.7) \\ I_x(m, n) &= (1-x)^n \sum_{j=m}^{\infty} \binom{n+j-1}{j} x^j, \quad m, n \in \mathbb{N}, x \in [0, 1). \end{aligned}$$

W programie komputerowym **Mathematica** wartości $I_x(\alpha, \beta)$ podaje funkcja `BetaRegularized[x, α, β]`.

Funkcja beta, niekompletna $\triangleright$ *funkcja beta*

Funkcja błędu Gaussa [**Gauss error function**] *Funkcją błędu Gaussa* lub *funkcją błędu* [**error function**] nazywamy funkcję

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x \exp(-t^2) dt = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt.$$

Jeżeli X jest zmienną losową o $\triangleright$ rozkładzie normalnym $\mathcal{N}(0, \sigma^2)$, to prawdopodobieństwo tego, że jej wartości znajdują się w przedziale $[-x, x]$ równa się $\operatorname{erf}(x/\sigma\sqrt{2})$. Definicja funkcji błędu $\operatorname{erf}(x)$ opisuje ten fakt w przypadku, gdy $\sigma^2 = 1/2$. Nazwa tej funkcji pochodzi stąd, że w wielu rozpatrywanych modelach statystycznych, probabilistycznych i teorii pomiarów zmienną losową X można interpretować jako błąd, tzn. różnicę $X = \hat{Y} - Y$ między zmienną losową Y a jej nieobciążonym oszacowaniem ($\triangleright$ estymatorem nieobciążonym) $\hat{Y}$.

Funkcję $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ nazywamy *uzupełniającą funkcją błędu Gaussa* [complimentary Gauss error function].

Funkcja charakterystyczna [characteristic function] *Funkcją charakterystyczną* zmiennej losowej X o rozkładzie określonym przez dystrybuantę $F(x) = P(X \leq x)$ nazywamy funkcję $\phi : \mathbb{R} \mapsto \mathbb{C}$ określoną wzorem

$$\begin{aligned} \phi(t) &= E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} dF(x) \\ &= \begin{cases} \int_{-\infty}^{\infty} e^{itx} f(x) dx, & \text{gdy } X \text{ jest typu ciągłego o gęstości } f(x), \\ \sum_k e^{itx_k} p_k, & \text{gdy } X \text{ jest typu dyskretnego, } p_k = P(X = x_k) \end{cases} \end{aligned}$$

(i oznacza jednostkę urojoną, tzn. $i = \sqrt{-1}$).

Podstawowe własności funkcji charakterystycznej ϕ :

- 1) $\phi(0) = 1$;
- 2) $|\phi(t)| \leq 1, t \in \mathbb{R}$;
- 3) $\phi(t)$ jest jednostajnie ciągła względem $t \in \mathbb{R}$;
- 4) $\phi(t) = \overline{\phi(-t)}$;
- 5) $\phi(t)$ jest dodatnio określona, tzn. dla każdego $n \in \mathbb{N}$, dla dowolnych liczb $z_1, \dots, z_n \in \mathbb{C}$ i dowolnych liczb $t_1, \dots, t_n \in \mathbb{R}$,

$$\sum_{j=1}^n \sum_{k=1}^n \phi(t_j - t_k) z_j \overline{z_k} \geq 0;$$

- 6) jeżeli zmienne losowe $X_1, \dots, X_n$ są niezależne i $S_n = X_1 + \dots + X_n$, to

$$\phi_{S_n}(t) = \prod_{j=1}^n \phi_{X_j}(t),$$

gdzie $\phi_Y(t)$ oznacza funkcję charakterystyczną zmiennej losowej Y ;

Funkcja charakterystyczna

7) jeżeli $Y = aX + b$, gdzie a i b są stałymi, to

$$\phi_Y(t) = e^{itb} \phi_X(at);$$

8) jeżeli $E(|X|^n) < \infty$ dla pewnego $n \in \mathbb{N}$, to dla każdego $k = 0, 1, \dots, n$,

$$E(X^k) = (-i)^k \left. \frac{d^k \phi(t)}{dt^k} \right|_{t=0}.$$

Na podstawie definicji, danemu rozkładowi przyporządkowana jest w sposób jednoznaczny funkcja charakterystyczna. Z poniższego twierdzenia, zwanego *wzorem Lévy'ego na odwrócenie*, wynika, że funkcja charakterystyczna wyznacza rozkład jednoznacznie.

TWIERDZENIE 1.12 (*wzór Lévy'ego na odwrócenie*) [**Lévy inversion formula**] Jeżeli $\phi(t)$ jest funkcją charakterystyczną zmiennej losowej X o dystrybucji $F(x)$, to dla $-\infty < a < b < \infty$,

$$\begin{aligned} \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt \\ = P(a < X < b) + \frac{1}{2} [P(X = a) + P(X = b)]. \end{aligned} \quad (1.8)$$

□

Jeżeli a, b są punktami ciągłości dystrybucji $F(x)$, to prawa strona wzoru (1.8) redukuje się do $F(b) - F(a)$.

WNIOSEK 1.13 Między dystrybucją na $\mathbb{R}$ a jej funkcją charakterystyczną istnieje wzajemnie jednoznaczne odwzorowanie. □

Na mocy Wniosku 1.13, przy rozwiązywaniu problemów probabilistycznych można używać przemienne dystrybucji i ich funkcji charakterystycznych.

WNIOSEK 1.14 Jeżeli $\int_{-\infty}^{\infty} |\phi(t)| dt < \infty$, to dla $-\infty < a < b < \infty$,

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt &= P(a < X < b) + \frac{1}{2} [P(X = a) + P(X = b)] \\ &= F(b) - F(a), \end{aligned}$$

gdzie F jest absolutnie ciągłą dystrybucją o ograniczonej, ciągłej gęstości

$$f(x) = F'(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt. \quad (1.9)$$

□

Mówimy, że ciąg dystrybucji $(F_n(x))$ jest słabo zbieżny do dystrybucji $F(x)$, jeżeli zbieżność $F_n(x) \rightarrow F(x)$ zachodzi w każdym punkcie ciągłości dystrybucji granicznej $F(x)$ ($\triangleright$ *rodzaje zbieżności zmiennych losowych*).

TWIERDZENIE 1.15 (*Lévy'ego o ciągłości*) [**Lévy continuity theorem**] Ciąg dystrybucji $(F_n(x))$ jest słabo zbieżny do dystrybucji $F(x)$ wtedy

i tylko wtedy, gdy ciąg ich funkcji charakterystycznych $(\phi_n(t))$ jest zbieżny do ciągłej funkcji granicznej $\phi(t)$. Funkcja $\phi(t)$ jest przy tym funkcją charakterystyczną granicznej dystrybuanty $F(x)$ i zbieżność $\phi_n(t) \rightarrow \phi(t)$ jest jednostajna w każdym skończonym przedziale. $\square$

Poniższe twierdzenie podaje warunki konieczne i dostateczne na to, aby funkcja $\phi(t)$ argumentu rzeczywistego była funkcją charakterystyczną.

TWIERDZENIE 1.16 (Bochnera) [Bochner's theorem] Funkcja $\phi : \mathbb{R} \mapsto \mathbb{C}$ jest funkcją charakterystyczną pewnego rozkładu prawdopodobieństwa wtedy i tylko wtedy, gdy jest ciągła, dodatnio określona (tzn. ma własność 5) oraz spełnia równość $\phi(0) = 1$. $\square$

Niech $\mathbf{X} = (X_1, \dots, X_k)^T$ będzie wektorem losowym o dystrybuancie $F(\mathbf{x}) = P(X_1 \leq x_1, \dots, X_k \leq x_k)$, $\mathbf{x} = (x_1, \dots, x_k)^T \in \mathbb{R}^k$.

Funkcja charakterystyczna wektora losowego $\mathbf{X}$ określona jest wzorem

$$\phi(\mathbf{t}) = E\left(e^{i\mathbf{t}^T \mathbf{X}}\right) = \int_{\mathbb{R}^k} e^{i \sum_{j=1}^k t_j X_j} dF(\mathbf{x}), \quad \mathbf{t} = (t_1, \dots, t_k)^T \in \mathbb{R}^k.$$

Jeżeli istnieją tzw. *mieszane momenty* $E\left(X_1^{j_1} X_2^{j_2} \dots X_k^{j_k}\right)$ wektora losowego $\mathbf{X}$, to zachodzi wzór

$$E\left(X_1^{j_1} X_2^{j_2} \dots X_k^{j_k}\right) = (-i)^j \left. \frac{\partial^j \phi(\mathbf{t})}{\partial t_1^{j_1} \partial t_2^{j_2} \dots \partial t_k^{j_k}} \right|_{\mathbf{t}=\mathbf{0}},$$

gdzie $j_1, j_2, \dots, j_k \in \overline{\mathbb{N}}$; $j = j_1 + j_2 + \dots + j_k$.

Niech $\mathbf{Y}$ będzie k -wymiarowym wektorem losowym niezależnym od $\mathbf{X}$. Wtedy

$$\phi_{\mathbf{X}+\mathbf{Y}}(\mathbf{t}) = \phi_{\mathbf{X}}(\mathbf{t})\phi_{\mathbf{Y}}(\mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^k.$$

Niech $\mathbf{Y} = \mathbf{C}^T \mathbf{X} + \mathbf{c}$, gdzie $\mathbf{C}$ jest macierzą wymiaru $k \times m$, a $\mathbf{c} \in \mathbb{R}^m$. Wtedy

$$\phi_{\mathbf{Y}}(\mathbf{u}) = e^{i\mathbf{u}^T \mathbf{c}} \phi_{\mathbf{X}}(\mathbf{C}\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^m.$$

Funkcję charakterystyczną $\phi(\mathbf{t})$ można otrzymać z $\triangleright$ *funkcji tworzącej momenty* $\psi(\mathbf{t})$ zamieniając $\mathbf{t}$ w funkcji $\psi(\mathbf{t})$ na $i\mathbf{t}$, jeżeli $\psi(\mathbf{t})$ jest skończona dla pewnego $\mathbf{t} \neq \mathbf{0}$.

Funkcja digamma $\triangleright$ *funkcja gamma*

Funkcja gamma [gamma function] *Funkcja gamma* $s \mapsto \Gamma(s)$ określona jest poprzez całkę Eulera drugiego rodzaju

$$\Gamma(s) = \int_0^\infty u^{s-1} e^{-u} du, \quad s > 0.$$

Funkcja gamma

Funkcję gamma nazywa się również *funkcją gamma Eulera*.

Funkcja gamma posiada następujące własności:

$$\begin{aligned}\Gamma(1) &= 1, \quad \Gamma(1/2) = \sqrt{\pi}; \\ \Gamma(s+1) &= s\Gamma(s), \quad \Gamma(n+1) = n!, \quad n \in \mathbb{N}; \\ \Gamma(s+1/2) &= \sqrt{\pi} \frac{\Gamma(2s)}{\Gamma(s)} 2^{1-2s}; \\ \Gamma(2n+1) &= \frac{2^{2n}}{\sqrt{\pi}} \Gamma(n+1/2) \Gamma(n+1), \quad n \in \mathbb{N}; \\ \frac{\Gamma(s)}{a^s} &= \int_0^\infty u^{s-1} e^{-au} du, \quad a = b + ic \in \mathbb{C}, \quad b > 0.\end{aligned}$$

W statystyce używa się również dwóch postaci **niekompletnej funkcji gamma** [**incomplete gamma function**]: funkcji określonej wzorem

$$\gamma(s, x) = \int_0^x u^{s-1} \exp(-u) du, \quad x > 0, \quad (1.10)$$

nazywanej *dolną niekompletną funkcją gamma* [**lower incomplete gamma function**] i jej “dopełnienia”

$$\Gamma_s(x) = \int_x^\infty u^{s-1} \exp(-u) du,$$

zwanego *górną niekompletną funkcją gamma* [**upper incomplete gamma function**]. W programie komputerowym *Mathematica* funkcja $\Gamma_s(x)$ nazywa się *niekompletną funkcją gamma* [**incomplete gamma function**] i wywoływana jest przez `Gamma[s, x]`.

Dystrybuanty niektórych rozkładów wyrażają się poprzez tzw. *regularyzowane niekompletne funkcje gamma* [**regularized incomplete gamma functions**]

$$P(s, x) = \frac{\gamma(s, x)}{\Gamma(s)} \quad \text{ i } \quad Q(s, x) = \frac{\Gamma_s(x)}{\Gamma(s)}.$$

Oczywiście,

$$\begin{aligned}\Gamma_s(0) &= \Gamma(s), \\ \Gamma_s(x) + \gamma(s, x) &= \Gamma(s), \\ Q(s, x) &= 1 - P(s, x).\end{aligned}$$

Ponadto

$$\begin{aligned}\Gamma_1(x) &= e^{-x}, \quad \gamma(1, x) = 1 - e^{-x} \\ \Gamma_s(x) &= (s-1)! e^{-x} \sum_{k=0}^{s-1} \frac{x^k}{k!} \quad \text{ dla } s \in \mathbb{N}.\end{aligned} \quad (1.11)$$

Funkcja $P(s, x/\lambda)$, $\lambda > 0$, jest równa dystrybucie zmiennej losowej o $\triangleright$ *rozkładzie gamma* $\mathcal{G}(s, \lambda)$, natomiast dla $x = 0, 1, 2, \dots$, funkcja $Q(x+1, \lambda)$ równa się dystrybucie zmiennej losowej o $\triangleright$ *rozkładzie Poissona* $\mathcal{P}(\lambda)$.

W pogramie **Mathematica** wartości funkcji $Q(s, x)$ podaje funkcja `GammaRegularized[s, x]`.

Pochodna logarytmiczna funkcji gamma, tzn.

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

nazywa się **funkcją digamma** [**digamma function**].

Funkcja gamma, niekompletna $\triangleright$ *funkcja gamma*

Funkcja hazardu $\triangleright$ *funkcja intensywności awarii*

Funkcja hazardu, skumulowana $\triangleright$ *funkcja intensywności awarii*

Funkcja intensywności awarii [**failure rate function**] Niech T będzie nieujemną zmienną losową, określającą np. czas pracy pewnego elementu. Taką zmienną losową nazywa się **czasem życia** [**lifetime**] lub **czasem przeżycia** [**survival time**]. Niech $F(t)$ oznacza dystrybucję zmiennej losowej T . Funkcję

$$\overline{F}(t) = P(T > t) = 1 - F(t)$$

nazywamy **funkcją niezawodności** [**reliability function**] lub **funkcją przeżycia** [**survival function**]. **Niezawodność** [**reliability**] elementu w chwili t jest to prawdopodobieństwo pracy tego elementu przez okres dłuższy niż t .

Załóżmy, że T jest zmienną losową typu ciągłego o gęstości $f(t)$. *Funkcja intensywności awarii* określona jest wzorem

$$\rho(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{\overline{F}(t)} = -\frac{\overline{F}'(t)}{\overline{F}(t)}. \quad (1.12)$$

Funkcja intensywności awarii nazywana jest również **funkcją hazardu** [**hazard function**] lub **stopą hazardu** [**hazard rate**].

Funkcja intensywności awarii w punkcie t jest gęstością rozkładu prawdopodobieństwa czasu pracy elementu pod warunkiem, że czas pracy elementu jest dłuższy niż t :

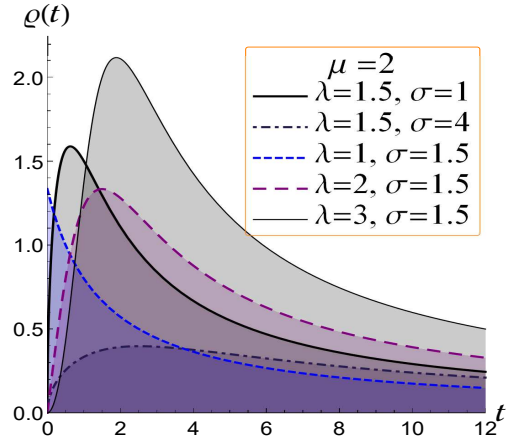
$$\begin{aligned} \rho(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t P(T > t)} = \lim_{\Delta t \rightarrow 0} \frac{f(t)\Delta t + o(\Delta t)}{\Delta t [1 - F(t)]} \\ &= \lim_{\Delta t \rightarrow 0} \left[\rho(t) + \frac{o(\Delta t)}{\Delta t} \right]. \end{aligned}$$

Dla $\triangleright$ *rozkładu wykładniczego* $\mathcal{E}(\lambda)$ funkcja intensywności awarii jest stała, $\rho(t) = 1/\lambda$, a dla $\triangleright$ *rozkładu Weibulla* $We(\alpha, \beta)$ funkcja intensywności awarii

Funkcja intensywności awarii

jest postaci $\rho(t) = \alpha\beta^{-\alpha}t^{\alpha-1}$ (patrz Tabela 1.1). Na Rysunku 1.1 przedstawiono wykresy funkcji hazardu dla rozkładu Burra typu XII ($\triangleright$ rozkłady Burra).

Interpretacja funkcji intensywności awarii $\rho(t)$ jest następująca: jeżeli w chwili $t_0 = 0$ zostanie podłączony element do pracy i do chwili t nie nastąpiła awaria, to prawdopodobieństwo, że ulegnie on awarii w przedziale $(t, t + \Delta t]$ można oszacować przez $\rho(t)\Delta t$.



Rys. 1.1: Wykresy funkcji hazardu rozkładu Burra typu XII.

W programie komputerowym **Mathematica** funkcja hazardu $\rho(t)$ rozkładu, oznaczonego przez D , określona jest przez funkcję postaci `HazardFunction[D, t]`. Na przykład, dla funkcji hazardu rozkładu Burra typu XII, tzn. $Bu^{(3)}(\mu, \lambda, \sigma)$, $D = \text{SinghMaddalaDistribution}[\mu, \lambda, \sigma]$ ($\triangleright$ rozkłady Burra).

Tab. 1.1: Funkcje hazardu niektórych rozkładów ciągłych.

	Rozkład	Funkcja hazardu $\rho(t)$
Funkcja $\gamma(x, \alpha)$ występująca we wzorze dla funkcji hazardu rozkładu gamma w Tabeli 1.1 oznacza zmodyfikowaną niekompletną funkcję gamma ($\triangleright$ funkcja gamma).	Wykładniczy $\mathcal{E}(\lambda)$	$1/\lambda$
	Weibulla $We(\alpha, \beta)$	$\alpha\beta^{-\alpha}t^{\alpha-1}$
	Gamma $\mathcal{G}(\alpha, \lambda)$	$\frac{t^{\alpha-1} \exp(-t/\lambda)}{\lambda^\alpha \Gamma(\alpha) [1 - \gamma(t/\lambda, \alpha)]}$
	Gompertza $Gom(\lambda, \alpha)$	$\alpha\lambda \exp(\lambda t)$
	Burra $Bu^{(3)}(\mu, \lambda, \sigma)$ (typu XII)	$\frac{\lambda\mu(t/\sigma)^{\lambda-1}}{\sigma[1 + (t/\sigma)^\lambda]}$

Przy założeniu, że $\overline{F}(t) > 0$, ze wzoru (1.12) otrzymujemy

$$\overline{F}(t) = \exp\left(-\int_0^t \rho(u)du\right), \quad t \geq 0.$$

A więc całkowalna funkcja $\rho(t)$ jest funkcją intensywności awarii wtedy i tylko

wtedy, gdy

$$\rho(t) \geq 0, \quad t \geq 0, \quad \int_0^\infty \rho(t)dt = \infty. \quad (1.13)$$

Funkcja $H(t) = \int_0^t \rho(u)du = -\ln[\bar{F}(t)]$ nazywa się **skumulowaną funkcją hazardu** [cumulative hazard function]. Znając postać funkcji przeżycia $\bar{F}(t)$, wartość oczekiwaną i wariancję zmiennej losowej T można obliczyć ze wzorów

$$E(T) = \int_0^\infty t f(t)dt = \int_0^\infty \bar{F}(t)dt,$$

$$\text{Var}(T) = 2 \int_0^\infty t \bar{F}(t)dt - \left[\int_0^\infty \bar{F}(t)dt \right]^2.$$

Funkcja niezawodności $\triangleright$ *funkcja intensywności awarii*

Funkcja odwrotna do dystrybuanty [inverse distribution function] Niech F oznacza dystrybuantę zmiennej losowej X : $F(x) = P(X \leq x)$. Funkcję odwrotną do dystrybuanty F definiuje się w sposób następujący:

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad \text{gdzie } 0 < u < 1.$$

TWIERDZENIE 1.17 Jeżeli F jest ciągłą dystrybuantą, to zmienna losowa $U = F(X)$ ma rozkład jednostajny $\mathcal{U}(0, 1)$. $\square$

TWIERDZENIE 1.18 Jeżeli U jest zmienną losową o rozkładzie $\mathcal{U}(0, 1)$, to zmienna losowa $X = F^{-1}(U) = \inf\{x : F(x) \geq U\}$ ma rozkład o dystrybuancie F . $\square$

Jeżeli F jest funkcją ciągłą i rosnącą, to F^{-1} jest funkcją odwrotną do F w zwykłym sensie. Tabela 1.2 zawiera przykłady rozkładów ciągłych, dla których istnieje analityczna postać funkcji odwrotnej do dystrybuanty. Postać funkcji odwrotnej do dystrybuanty wykorzystywana jest w metodzie przekształcenia odwrotnego $\triangleright$ *symulacji zmiennej losowej*.

Tab. 1.2: Funkcje odwrotne do dystrybuant niektórych rozkładów ciągłych.

Rozkład	$F^{-1}(u)$
Beta $\mathcal{B}e(\alpha, 1)$	$u^{1/\alpha}$
Beta $\mathcal{B}e(1, \beta)$	$1 - (1 - u)^{1/\beta}$
Burra $\mathcal{B}u^{(2)}(\lambda, \mu)$	$\left(\frac{\mu u}{1 - u}\right)^{1/\lambda}$
Cauchy'ego $\mathcal{C}(\alpha, \lambda)$	$\alpha + \lambda \tan\left(\pi\left(u - \frac{1}{2}\right)\right)$
Wykładniczy $\mathcal{E}(\lambda)$	$-\lambda \ln(1 - u)$
Logistyczny $\mathcal{LG}(\mu, \sigma)$	$\mu + \frac{\sqrt{3}}{\pi} \sigma \ln\left(\frac{u}{1 - u}\right)$
Pareto $\mathcal{P}a(x_0, \alpha)$	$\frac{x_0}{(1 - u)^{1/\alpha}}$
Jednostajny $\mathcal{U}(a, b)$	$a + (b - a)u$
Weibulla $\mathcal{W}e(\alpha, \beta)$	$\beta[-\ln(1 - u)]^{1/\alpha}$

Rozkład wartości ekstremalnej maksymalnej, uogólniony

▷ Funkcja charakterystyczna wyraża się wzorem

$$\phi(t) = e^{\iota\mu t}\Gamma(1 - \iota\sigma t)$$

(Γ oznacza ▷ funkcję gamma), a półniezmienniki (▷ funkcja tworząca kumulanty) dane są przez $\kappa_k = (k-1)!\sigma^k\zeta(k)$, $k = 2, 3, \dots$

W programie komputerowym **Mathematica** rozkład $\mathcal{EV}(\mu, \sigma)$ określony jest przez funkcję `ExtremeValueDistribution`[μ, σ].

Rozkład $\mathcal{EV}(\mu, \sigma)$ jest szczególnym przypadkiem ▷ rozkładu wartości ekstremalnej, uogólnionego i ▷ rozkładu wartości ekstremalnej maksymalnej, uogólnionego, mianowicie

$$\mathcal{EV}(\mu, \sigma) \equiv \mathcal{GEV}(\mu, \sigma, 0) \equiv \text{MaxStab}(\mu, \sigma, 0).$$

Ponadto zachodzą następujące związki:

jeżeli $X \sim \mathcal{EV}(\mu, \sigma)$, to $Y = -X \sim \text{Gum}(-\mu, \sigma) \equiv \text{MinStab}(-\mu, \sigma, 0)$;

jeżeli $X \sim \mathcal{EV}(\mu, \sigma)$, to $Y = \exp(-X) \sim \text{We}(1/\sigma, \exp(-\mu))$; w szczególności,

jeżeli $X \sim \mathcal{EV}(\mu, 1)$, to $Y = \exp(-X) \sim \mathcal{E}(\exp(-\mu))$.

Rozkład $\mathcal{EV}(\mu, \sigma)$ nazywany jest też **rozkładem podwójnie wykładniczym** [**double exponential distribution**] lub ▷ **rozkładem Gumbela** (zobacz też ▷ **rozkłady wartości ekstremalnej**).

Rozkład $\mathcal{EV}(\mu, \sigma)$ znajduje zastosowania m.in. w teorii niezawodności oraz przy oszacowaniach

Rozkład wartości ekstremalnej maksymalnej, uogólniony [**generalized maximum extreme value distribution**] Zmienna losowa X ma *uogólniony rozkład wartości ekstremalnej maksymalnej* z parametrami μ, σ i ξ ($\mu, \xi \in \mathbb{R}, \sigma > 0$), jeżeli jej gęstość wyraża się wzorem

$$\begin{aligned} f(x) &= f(x; \mu, \sigma, \xi) \\ &= \begin{cases} \frac{1}{\sigma} \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi-1} \exp \left[- \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} \right] \mathbf{1}_{[\mu - \sigma/\xi, \infty)}(x), & \xi > 0, \\ \frac{1}{\sigma} \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi-1} \exp \left[- \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} \right] \mathbf{1}_{(-\infty, \mu - \sigma/\xi]}(x), & \xi < 0, \\ \frac{1}{\sigma} \exp \left[- \exp \left(- \frac{x - \mu}{\sigma} \right) - \frac{x - \mu}{\sigma} \right], & x \in \mathbb{R}, \quad \xi = 0. \end{cases} \end{aligned}$$

Rozkład ten oznaczamy przez $\text{MaxStab}(\mu, \sigma, \xi)$. Nośnikiem tego rozkładu dla $\xi \neq 0$ jest zbiór $\{x \in \mathbb{R} : \xi x \geq \xi\mu + \sigma\}$.

Nazwa rozkładu $\text{MaxStab}(\mu, \sigma, \xi)$ związana jest z tym, że w teorii wartości ekstremalnej używany jest termin **rozkład max-stabilny** [**max-stable distribution**] do opisu sytuacji, które są ekstremalnie nieprawdopodobne (tzn. w których zbiory danych tworzone są przez zmienne o ekstremalnych odchyleniach od mediany). Rozkłady te służą do modelowania wielu zjawisk w takich dziedzinach jak m.in. finansach i ekonomii.

Dystrybuanta rozkładu $MaxStab(\mu, \sigma, \xi)$ jest postaci

$$F(x) = F(x; \mu, \sigma, \xi) = \begin{cases} \exp \left[- \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right], & x \geq \mu - \sigma/\xi, & \xi > 0, \\ \exp \left[- \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right], & x \leq \mu - \sigma/\xi, & \xi < 0, \\ \exp \left[- \exp \left(- \frac{x - \mu}{\sigma} \right) \right], & x \in \mathbb{R}, & \xi = 0, \\ 1, & x \geq \mu - \sigma/\xi, & \xi > 0, \\ 0, & x \leq \mu - \sigma/\xi, & \xi < 0. \end{cases}$$

Uogólniony rozkład wartości ekstremalnej maksymalnej jest jedynym rozkładem asymptotycznym zmiennej losowej $X_{n:n} = \max\{X_1, \dots, X_n\}$, gdzie $X_1, \dots, X_n$ są niezależnymi zmiennymi losowymi o tym samym $\triangleright$ rozkładzie normalnym lub $\triangleright$ rozkładzie Cauchy'ego lub $\triangleright$ rozkładzie beta. Innymi słowy jest to rozkład asymptotyczny n -tej (maksymalnej) $\bowtie$ statystyki pozycyjnej z $\bowtie$ próby losowej $(X_1, \dots, X_n)$.

Rozkład $MaxStab(\mu, \sigma, \xi)$ związany jest z innymi rozkładami wartości ekstremalnej, takimi jak $\triangleright$ rozkład Gumbela, $\triangleright$ rozkład wartości ekstremalnej, $\triangleright$ rozkład Weibulla. Zachodzi bezpośredni związek rozkładu $MaxStab(\mu, \sigma, \xi)$ z $\triangleright$ rozkładem wartości ekstremalnej minimalnej, uogólnionym, mianowicie, jeżeli $X \sim MaxStab(\mu, \sigma, \xi)$, to $Y = -X \sim MinStab(-\mu, \sigma, \xi) \equiv \mathcal{GGum}(-\mu, \sigma, -\xi)$.

Ponadto

$$MaxStab(\mu, \sigma, \xi) \equiv \mathcal{GEV}(\mu, \sigma, -\xi);$$

$$MaxStab(\mu, \sigma, 0) \equiv \mathcal{GEV}(\mu, \sigma, 0) \equiv \mathcal{EV}(\mu, \sigma);$$

$$\text{jeżeli } X \sim MaxStab(0, \sigma, -1/\alpha), \alpha > 0, \text{ to } Y = -\frac{1}{\alpha}X + \sigma \sim We(\alpha, \sigma).$$

W przypadku $\xi < 0$, dla $x \leq \mu - \sigma/\xi$ dystrybuantę można zapisać w postaci

$$\begin{aligned} F(x; \mu, \sigma, \xi) &= \exp \left[- \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right] = \exp \left\{ - \left[- \frac{x - (\mu - \sigma/\xi)}{-\sigma/\xi} \right]^{-1/\xi} \right\} \\ &= \exp \left[- \left(- \frac{x - \mu^*}{\sigma^*} \right)^\alpha \right], \end{aligned}$$

gdzie $\mu^* = \mu - \sigma/\xi$, $\sigma^* = -\sigma/\xi$, $\alpha = -1/\xi$. Wynika stąd, że rozkład wartości ekstremalnej typu III (wg klasyfikacji podanej w temacie $\triangleright$ rozkłady wartości ekstremalnej), określony przez dystrybuantę $G_3(x; \mu, \sigma, \alpha)$, $x \leq \mu$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\alpha > 0$, równoważny jest uogólnionemu rozkładowi wartości ekstremalnej maksymalnej $MaxStab(\mu - \sigma/\alpha, \sigma/\alpha, -1/\alpha)$.

Zauważmy, że jeżeli $X \sim G_3(x; \mu^*, \sigma^*, \alpha)$, to

$$\begin{aligned} Y = -\ln(\mu^* - X) &\sim G_3(x; \mu^* - \exp(-y), \sigma^*, \alpha) \\ &= \exp \left[-\exp \left(-\frac{y + \ln \sigma^*}{1/\alpha} \right) \right] \equiv G_1(y; -\ln \sigma^*, 1/\alpha) \\ &\sim \mathcal{EV}(-\ln(-\sigma/\xi), -\xi), \quad \xi < 0, \end{aligned}$$

gdzie $G_1(x; \mu, \sigma)$ oznacza dystrybucję rozkładu wartości ekstremalnej typu I ($\triangleright$ rozkłady wartości ekstremalnej). Ponieważ $\text{MaxStab}(\mu, \sigma, \xi) \sim G_3(x; \mu^*, \sigma^*, \alpha)$, wynika stąd następujący bezpośredni związek rozkładu $\text{MaxStab}(\mu, \sigma, \xi)$, dla $\xi < 0$, z rozkładem wartości ekstremalnej typu I: jeżeli $X \sim \text{MaxStab}(\mu, \sigma, \xi)$, to $Y = -\ln(\mu - \sigma/\xi - X) \sim \mathcal{EV}(-\ln(-\sigma/\xi), -\xi)$ dla $\xi < 0$.

Rozkład $\text{MaxStab}(\mu, \sigma, \xi)$ jest rozkładem jednomodalnym (tzn. z jedną wartością modalną równą maksimum globalnemu funkcji gęstości) i na przebieg funkcji gęstości mają wyraźny wpływ wszystkie trzy parametry tego rozkładu.

Wartość oczekiwana, wariancja i mediana rozkładu $\text{MaxStab}(\mu, \sigma, \xi)$ równe są, odpowiednio

$$E(X) = \begin{cases} \mu + \sigma C & \text{dla } \xi = 0, \\ \frac{\mu\xi - \sigma + \sigma\Gamma(1 - \xi)}{\xi} & \text{dla } \xi \neq 0 \text{ i } \xi < 1, \\ \text{nieokreślona} & \text{dla } \xi \geq 1, \end{cases}$$

(C oznacza stałą Eulera)

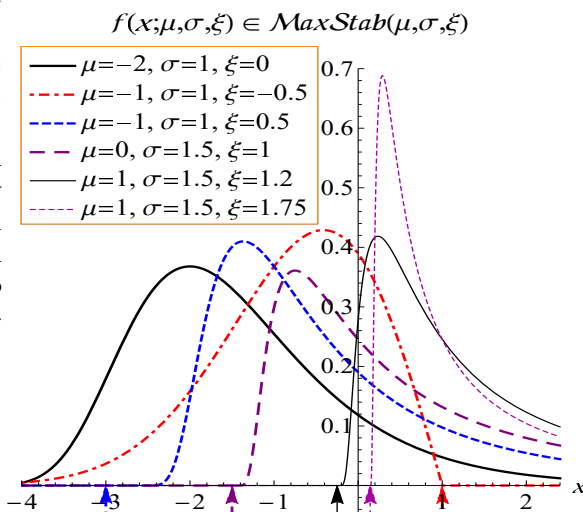
$$\text{Var}(X) = \begin{cases} \pi^2 \sigma^2 / 6 & \text{dla } \xi = 0, \\ \frac{\sigma^2 [\Gamma(1 - 2\xi) - \Gamma^2(1 - \xi)]}{\xi^2} & \text{dla } \xi \neq 0 \text{ i } 2\xi < 1, \\ \text{nieokreślona} & \text{dla } 2\xi \geq 1, \end{cases}$$

$$\text{me}(X) = \begin{cases} \mu - \sigma \ln(\ln 2) & \text{dla } \xi = 0, \\ \mu - \frac{\sigma [1 - (\ln 2)^{-\xi}]}{\xi} & \text{dla } \xi \neq 0. \end{cases}$$

Rozkład wartości ekstremalnej minimalnej, uogólniony

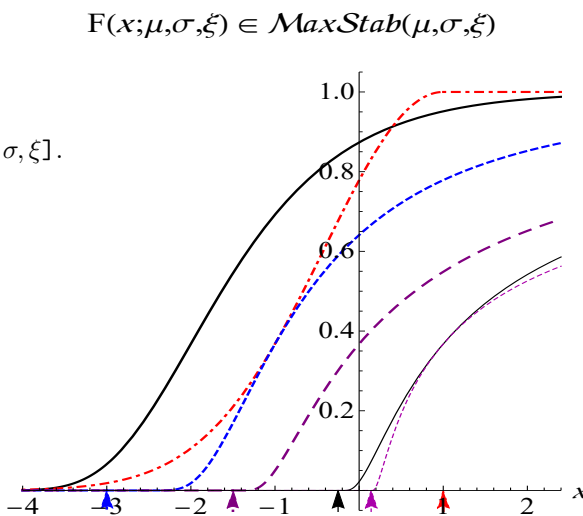
Ponieważ funkcja gęstości rozkładu jest postaci podwójnie wykładniczej, jej wykres przebiega bardziej ekspresyjnie w tym sensie, że jej piki są wyższe i tzw. ogony "cieńsze" w stosunku do innych rozkładów. Ogony rozkładu są cienkie w tym sensie, że funkcja gęstości maleje bardziej wykładniczo niż potęgowo dla dużych wartości x .

Rys. 1.40: Wykresy gęstości uogólnionego rozkładu wartości ekstremalnej maksymalnej $MaxStab(\mu, \sigma, \xi)$.



W programie komputerowym **Mathematica** rozkład $MaxStab(\mu, \sigma, \xi)$ określony jest przez funkcję `MaxStableDistribution[ $\mu, \sigma, \xi$ ]`.

Rys. 1.41: Wykresy dystrybucyj uogólnionego rozkładu wartości ekstremalnej maksymalnej $MaxStab(\mu, \sigma, \xi)$.



Rozkład wartości ekstremalnej minimalnej, uogólniony [generalized minimum extreme value distribution] Rozkład ten znany jest również jako **rozkład Fishera-Tippeta** [Fisher-Tippet distribution]. Zmienna losowa X ma uogólniony rozkład wartości ekstremalnej minimalnej z parametrami

Rozkład wartości ekstremalnej minimalnej, uogólniony

μ, σ i ξ ($\mu, \xi \in \mathbb{R}, \sigma > 0$), jeżeli jej gęstość wyraża się wzorem

$$f(x) = f(x; \mu, \sigma, \xi) = \begin{cases} \frac{1}{\sigma} \left(1 - \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi-1} \exp \left[- \left(1 - \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} \right] \mathbf{1}_{(-\infty, \mu + \sigma/\xi]}(x), & \xi > 0, \\ \frac{1}{\sigma} \left(1 - \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi-1} \exp \left[- \left(1 - \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} \right] \mathbf{1}_{[\mu + \sigma/\xi, \infty)}(x), & \xi < 0, \\ \frac{1}{\sigma} \exp \left[- \exp \left(\frac{x - \mu}{\sigma} \right) + \frac{x - \mu}{\sigma} \right], & x \in \mathbb{R}, \quad \xi = 0. \end{cases}$$

Rozkład ten oznaczamy przez $\text{MinStab}(\mu, \sigma, \xi)$. Nośnikiem rozkładu $\text{MinStab}(\mu, \sigma, \xi)$ dla $\xi \neq 0$ jest zbiór $\{x \in \mathbb{R} : \xi x \leq \xi \mu + \sigma\}$.

Nazwa rozkładu $\text{MinStab}(\mu, \sigma, \xi)$ związana jest z tym, że w teorii wartości ekstremalnej używany jest termin **rozkład min-stabilny** [**min-stable distribution**] do opisu sytuacji, które są ekstremalnie nieprawdopodobne (tzn. w których zbiory danych tworzone są przez zmienne o ekstremalnych odchyleniach od mediany). Rozkłady takie służą do modelowania wielu zjawisk w takich dziedzinach jak m.in. finansach i ekonomii.

Dystrybuenta rozkładu $\text{MinStab}(\mu, \sigma, \xi)$ jest postaci

$$F(x) = F(x; \mu, \sigma, \xi) = \begin{cases} 1 - \exp \left[- \left(1 - \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} \right], & x \leq \mu + \sigma/\xi, \quad \xi > 0, \\ 1 - \exp \left[- \left(1 - \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} \right], & x \geq \mu + \sigma/\xi, \quad \xi < 0, \\ 1 - \exp \left[- \exp \left(\frac{x - \mu}{\sigma} \right) \right], & x \in \mathbb{R}, \quad \xi = 0, \\ 1, & x > \mu + \sigma/\xi, \quad \xi > 0, \\ 0, & x < \mu + \sigma/\xi, \quad \xi < 0. \end{cases}$$

Uogólniony rozkład wartości ekstremalnej minimalnej jest jedynym rozkładem asymptotycznym zmiennej losowej $X_{1:n} = \min\{X_1, \dots, X_n\}$, gdzie $X_1, \dots, X_n$ są niezależnymi zmiennymi losowymi o tym samym $\triangleright$ *rozkładzie normalnym* lub $\triangleright$ *rozkładzie Cauchy'ego* lub $\triangleright$ *rozkładzie beta*. Innymi słowy jest to rozkład asymptotyczny 1-szej (minimalnej) $\bowtie$ *statystyki pozycyjnej* z $\bowtie$ *próby losowej* $(X_1, \dots, X_n)$.

Rozkład $\text{MinStab}(\mu, \sigma, \xi)$ związany jest z innymi rozkładami wartości ekstremalnej, takimi jak $\triangleright$ *rozkład Gumbela*, $\triangleright$ *rozkład wartości ekstremalnej* (typu I), $\triangleright$ *rozkład Weibulla*. Związany jest bezpośrednio z uogólnionym rozkładem wartości ekstremalnej maksymalnej, mianowicie

jeżeli $X \sim \text{MinStab}(\mu, \sigma, \xi)$, to

$$Y = -X \sim \text{MaxStab}(-\mu, \sigma, \xi) \equiv \mathcal{GEV}(-\mu, \sigma, -\xi).$$

Ponadto

$$\text{MinStab}(\mu, \sigma, \xi) \equiv \mathcal{GGum}(\mu, \sigma, -\xi), \text{ w szczególności}$$

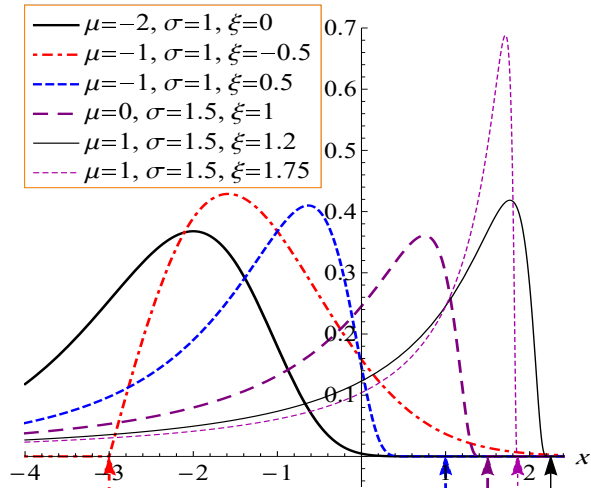
Rozkład wartości ekstremalnej minimalnej, uogólniony

$\text{MinStab}(\mu, \sigma, 0) \equiv \mathcal{GGum}(\mu, \sigma, 0) \equiv \mathcal{Gum}(\mu, \sigma);$

jeżeli $X \sim \text{MinStab}(0, \sigma, -1/\alpha)$, $\alpha > 0$, to $Y = \frac{1}{\alpha}X + \sigma \sim \text{We}(\alpha, \sigma)$.

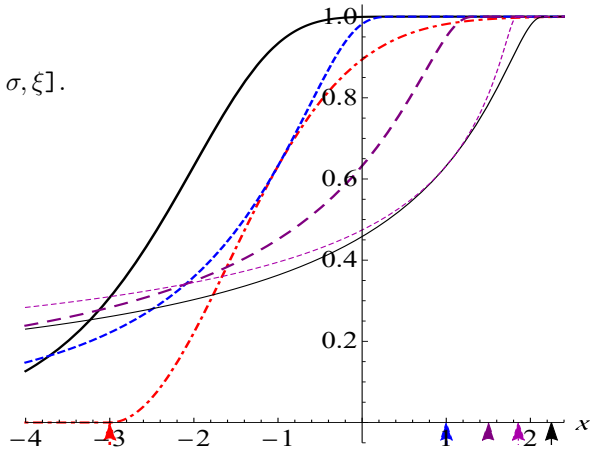
Rozkład $\text{MinStab}(\mu, \sigma, \xi)$ jest rozkładem jednomodalnym (tzn. z jedną wartością modalną równą maksimum globalnemu funkcji gęstości) i na przebieg funkcji gęstości mają wyraźny wpływ wszystkie trzy parametry tego rozkładu. Ponieważ funkcja gęstości rozkładu jest postaci podwójnie wykładniczej, jej wykres przebiega bardziej ekspresyjnie w tym sensie, że jej piki są wyższe i tzw. ogony "cieńsze" w stosunku do innych rozkładów. Ogony rozkładu są cienie w tym sensie, że funkcja gęstości maleje wykładniczo bardziej niż potęgowo dla dużych wartości x .

$$f(x; \mu, \sigma, \xi) \in \text{MinStab}(\mu, \sigma, \xi)$$



Rys. 1.42: Wykresy gęstości uogólnionego rozkładu wartości ekstremalnej minimalnej $\text{MinStab}(\mu, \sigma, \xi)$ (dla tych samych wartości parametrów jak dla rozkładu $\text{MaxStab}(\mu, \sigma, \xi)$ na Rysunku 1.40).

$$F(x; \mu, \sigma, \xi) \in \text{MinStab}(\mu, \sigma, \xi)$$



Rys. 1.43: Wykresy dystrybucyj uogólnionego rozkładu wartości ekstremalnej minimalnej $\text{MinStab}(\mu, \sigma, \xi)$.