EKONOMETRIA ECONOMETRICS 3(41) • 2013
ISSN 1507-3866

**Łukasz Kuźmiński**
Wroclaw University of Economics

# THE APPLICATIONS OF THE KERNEL DENSITIES TO MODELING THE GENERALIZED PARETO DISTRIBUTIONS

**Abstract:** In this paper we present the tools used in the modeling of distributions with fat tails in the theory of extreme values. We present three tools: the sample distribution function, the histograms for grouped data and the kernel densities. The latter is described in detail. The presented examples show the application of the kernel densities to modeling the generalized Pareto distributions.

**Keywords:** Kernel densities, extreme values, generalized Pareto distributions.

## 1. Introduction

Extreme values because of their essential influence – in most cases negative – on many fields of life and science, have been are an object of interest not only for the scientists and the researches of the many fields for a long time. The negative influence of the described values of the definite characteristic is observed in fields such as economics and financial markets, metrology and hydrology, and insurance. It isnot without reason these three fields are mentioned. The recently intensifying economic crisis, which has influence on the qualitative changes in the financial time series and the violent and strong changes of atmospheric conditions, which are the cause of many meteorological and hydrological disasters, both in our country and worldwide, are the cause of the increasing interest in the extreme values theory.

This interest is focused on the way in which one may protect oneself from the negative influence of the extremely high and low definite financial characteristics, or the meteorological and hydrological characteristics, which are the direct causes of the appearance of the phenomena described above.

In order of protecting oneself from the negative influence of the extreme values we need to know the models, which will describe accurately the character and behavior of the observed variables. Up to that point, if only to model diligently the extreme values of the examined variables, we need to determine the theoretical distribution of the population from which the examined data come, or alternatively we can approximate the distribution of the extreme values for the examined variables on the basis of the examined samples.

In this work we present the tool with which one can identify the distribution of the continuous variables on the basis of the observations from the sample. This tool is called the kernel density. The application of this tool is presented using the example of the family of the generalized Pareto distributions, which are also used in the modeling of the extreme values.

## 2. The basic notions and symbols in the theory of extreme values

Suppose that $X_1, X_2, ...$ are the sequence of the identical independent random variables (iid), and in other words, the variables have the shared distribution function $F(x)$.

As $M_n$ we mean the random variable which is maximum form $n$ the random variables, namely

$$M_n = \max\left(X_1, X_2, ..., X_n\right).$$

Due to the fact that the considered theory for the distributions $M_n$ has an asymptotic character, properties in particular are satisfied when $n \to \infty$. From a practical point of view, it is stressed that we should consider the relatively long sequences of the random variables in the destination of receiving solid results. All the results for the maxima one can move into the simplest manner on the minima using the relation

$$m_n = \min\left(X_1, X_2, ..., X_n\right) = -\max\left(-X_1, -X_2, ..., -X_n\right).$$

In this situation, the distribution function can be presented in the simplest manner with the following formula,

$$P\left\{M_n \le x\right\} = P\left\{X_1 \le x, X_2 \le x, ..., X_n \le x\right\} = F^n(x),$$

where $F(x)$ means the shared distribution function of the variables $X_i$ ($i = 1, 2, \dots, n$), appropriately the distribution function of the random variable $m_n$ is given with the formula

$$F_{(1)}(x) = P\left\{X_{(1)} \le x\right\} = 1 - P\left\{X_{(1)} > x\right\} = 1 - P\left\{all\, X_i > x\right\} = 1 - \left[1 - F(x)\right]^n.$$

Additionally, we define the exceedance distribution function. Let $x_i$ be governed by a distribution function $F$ and let threshold $u$ be smaller than the right endpoint defined as $\omega(F) = \sup\left\{x : F(x) < 1\right\}$. We speak of high threshold $u$, if $u$ is close to the right endpoint $\omega(F)$. In that case, $p = 1 - F(u)$ is small and the number $k$ of exceedances may be regarded as the Poisson random variable. Subsequently, we deal with the magnitudes (sizes) of the exceedances. Exceedances occur conditioned on the event that an observation is larger than the threshold $u$. The pertaining conditional distribution function $F^{[u]}$ is called exceedance distribution function at $u$. If $X$ denotes a random variable with distribution function $F$, then

$$F^{[u]}(x) = P\left(X \le x \middle| X > u\right) = P\{X \le x, X > u\} / P\{X > u\} =$$
$$= \frac{F(x) - F(u)}{1 - F(u)}, \qquad x \ge u.$$

One should keep in mind that the left endpoint $\alpha\left(F^{[u]}\right) = \inf\left\{x:\ F^{[u]}(x) > 0\right\}$ of $F^{[u]}$ is equal to $u$. The generalized Pareto distribution functions (GP) will be fitted to exceedance distribution functions (df) $F^{[u]}$ in the next section.

At the end of this section we introduce another closely related approach of extracting extreme values from the data taking the $k$ largest values $x_{n-k+1:n} \le \cdots \le x_{n:n}$ of the $x_i$., where the number $k$ is predetermined. Notice that $x_{n:n}$ is maximum, i.e. the execution of the random variables described by the formula. Within this approach, the ($k$ + 1)th largest observation $x_{n-k:n}$ may be regarded as a random threshold (see [David, Nagaraja 2003; Leadbetter et al. 1983; Thomas, Reiss 2007]).

## 3. Generalized Pareto distributions

In this section we present the family of generalized distributions of Pareto. The standard generalized distribution functions (dfs) Pareto (GP) $W_{i,\alpha}$ and $W_\gamma$ are adequate parametric dfs for exceedances, cf. also? The densities are denoted similarly by $\omega_{i,\alpha}$ and $\omega_\gamma$.

First we introduce the representation for GP df within three submodels corresponding to that for the three types dfs extremal values (EV) (see e.g. [Galambos 1978; Kuźmiński 2012; Leadbetter et al. 1983])

Exponential (GP0): $W_0(x) = 1 - e^{-x}$, $x \ge 0$

Pareto (GP1), $0:$ $W_{1,\alpha}(x) = 1 - x^{-\alpha}$, $x \ge 1$,

Beta (GP2), $\alpha < 0:$ $W_{2,\alpha}(x) = 1 - (-x)^{-\alpha}$, $-1 \le x \le 0$.

Of course, the exponential df $W_0$ is equal to zero for $x < 0$; the Pareto dfs $W_{1,\alpha}$ are equal to zero for $x < 1$; the Beta dfs $W_{2,\alpha}$ are equal to zero for $x < 0$ and equal to 1 for $x > 0$.

Note that $W_{2,-1}$ is the uniform df on the interval [–1, 0]. One should be aware that the dsf $W_{2,\alpha}$ constitute a subclass of the usual family of beta dfs. Subsequently, when we speak of beta dfs only dfs $W_{2,\alpha}$ are addressed. Warning, our parameterization for beta dfs differs from the standard one used in the statistical literature, where beta dfs with positive shape parameters are taken.

Below, we present the densities functions for dfs GP

Exponential (GP0): $w_0(x) = e^{-x}$, $x \geq 0$

Pareto (GP1), $\alpha > 0$: $w_{1,\alpha}(x) = \alpha x^{-(1+\alpha)}$, $x \geq 1$,

Beta (GP2), $\alpha < 0$: $w_{2,\alpha}(x) = |\alpha|(-x)^{-(1+\alpha)}$, $-1 \leq x \leq 0$.

The Pareto and exponential densities are decreasing on their supports. This property is shared by beta densities with shape parameter $\alpha < -1$. For $\alpha = -1$ one gets the uniform density on [–1, 0] as mentioned above. Finally, the beta densities with shape parameter $\alpha > -1$ are increasing, having a pole at zero.

One must add the location and scale parameters $\mu$ and $\sigma > 0$ in order to obtain the full statistical families GP dfs. Notice that the left endpoint of the Pareto df $W_{1,\alpha,\mu,\sigma}(x) = W_{1,\alpha}\left((x-\mu)/\sigma\right)$ is equal to $\mu + \sigma$ (see [Johnson, Kotz 1970; Thomas, Reiss 2007]).

## 4. Kernel densities

In this section, we catch a glimpse of the real world in the condensed form of data. Our primary aim is to fit generalized Pareto (GP) distributions, which were introduced in the foregoing section by means of limit theorems, to the data.

We describe visualization techniques, such as the sample distribution function, histograms for grouped data, and finally, kernel densities.

The sample df $\hat{F}_n(x)$ at *x* for series of univariate date $x_1, \ldots, x_n$ is the relative number of the $x_i$ that are smaller or equal to *x*. Thus,

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i \leq n} I\left(x_i \leq x\right),$$

where the indicator function is defined by $I\left(y \leq x\right) = 1$ if $y \leq 0$ and 0, elsewhere; furthermore, the summation runs over $i = 1, \ldots, n.$ Sample dfs are particularly useful for representing samples of e smaller size.

The data $x_1, \ldots, x_n$ ordered from the smallest to the largest are denoted by

$$x_{1:n} \leq \cdots \leq x_{n:n}.$$

We have $\hat{F}_n(x_{i:n}) = i/n$ if $x_{i:n}$ is not a multiple point. Notice that $\hat{F}_n$ is constant between consecutive ordered values. The ordered values can be recaptured from the sample df and, thus, there is a one – to – one correspondence between the sample df and the ordered values.

Occasionally we write $\hat{F}_n(\mathbf{x}; x)$ in place of $\hat{F}_n(x)$ to indicate the dependence on the vector data $\mathbf{x} = (x_1, \ldots, x_n)$. We will primarily deal with situations where each of the $x_i$ is generated under a common df *F* and the sample df $\hat{F}_n$ is approximately equal to *F*. This relation will be briefly written as

$$\hat{F}_n(x) \approx F(x).$$

Clearly, one result from that $\hat{F}_n(x)$ is an estimator of $F(x)$.

Below we have the graphs Pareto df and the sample df, which was made for the set of Pareto date for the sample $n$ = 100.

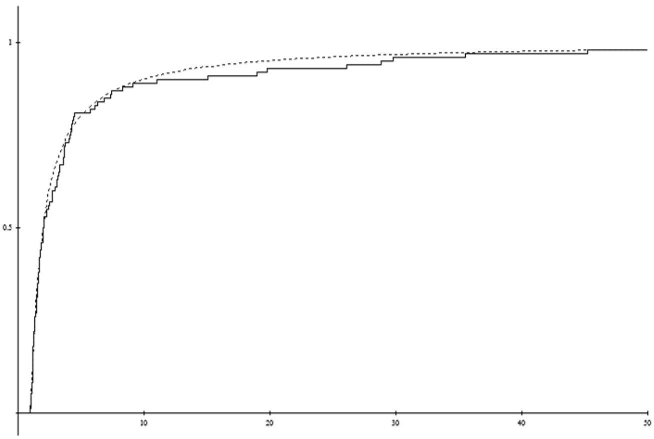


**Figure 1.** Pareto df (dotted) and sample df of Pareto data set with 100 points

Source: own study.

To a larger extent, our statistical arguments are based on the relation between the sample df $\hat{F}_n(x)$ and the underlying df $F$ so that, apparently, the statistical prerequisites for understanding this text book are of an elementary nature.

Now we present linearly sample distributions. If the underlying df is continuous, then it is plausible to estimate this df by means of a continuous sample df. Such a df can be constructed by linearly interpolating the sample df $\hat{F}_n(x)$ in over intervals $(t_j,\ t_{j+1}]$, where the $t_j < t_{j+1}$ constitute a grid on the real line. One gets the continuous sample df

$$F_n(x) = \hat{F}_n(t_j) + \frac{x - t_j}{t_{j+1} - t_j}\left(\hat{F}_n(t_{j+1}) - \hat{F}_n(t_j)\right) =$$

$$= \hat{F}_n(t_j) + \frac{(x - t_j)n_j}{t_{j+1} - t_j}, \qquad \mathit{for}\ \ t_j < x \leq t_{j+1},$$

where $n_j$ is the frequency of data $x_1, \ldots, x_n$ in the interval $(t_j,\ t_{j+1}]$. Thus, the sample df $F_n$ only depends on the data in a grouped form (see [Cleveland 1993; Simnoff 1996]). The application of this tool will be present in the next section in an example.

The sample df based on grouped data is piecewise continuously differentiable and, therefore, it has a density in the form of a histogram. The next concept can be regarded as a modification of the histograms for grouped data.

Let $n_j$ be the frequency of data in the interval $(t_j,\ t_{j+1}]$. Taking the derivative of the preceding sample df $F_n$ based on grouped data as given, one gets the probability density

$$f_n(x) = \frac{n(j)}{n(t_{j+1} - t_j)}, \qquad t_j < x \le t_{j+1}.$$

It is very natural to visualize frequencies by means of such a histogram. It is apparent that this histogram is an appropriate estimate of the density $f$ of $F$. The histogram may also be addressed as sample density.

Practitioners use histograms because of their simplicity in representing data, even if the data are given in a continuous form. One disadvantage of a histogram is that one must choose the location of the grid.

In the case of discrete data, a sample histogram is given by

$$p_n(j) = n(j)/n,$$

where $n(j)$ is number of data $x_1, \ldots, x_n$ equal to the integer $j$. As an analogy, we have

$$p_n(j) \approx P\{j\},$$

$P$ which is the underlying discrete distribution (under which the $x_i$ were generated). Note that discrete values $x_i$ – ordered according to their magnitudes – can be recaptured from the histogram. In view of the elementary character of this tool we do not present the examples.

The last and fundamental tool which we present is the kernel density. Starting with continuous data $x_1, \ldots, x_n$, the histogram for grouped data may be constructed in the following manner. Replace each point $x_i$ in the interval $(t_j,\ t_{j+1}]$ by the constant function

$$g(x, x_i) = \frac{1}{n(t_{j+1} - t_j)}, \qquad t_j < x \le t_{j+1}$$

with weight $1/n$. In summing up the single terms $g(x, x_i)$, one gets the histogram for grouped data in the representation $f_n(x) = \sum_{i \le n} g(x,\ x_i)$. If continuous data are given, the choice of the grid is crucial for the performance of the histogram.

We represent an alternative construction of a sample density. In contrast to it, replace $x_i$ by the function

$$g_b(x, x_i) = \frac{1}{nb} k\left(\frac{x - x_i}{b}\right),$$

where $k$ is a function (kernel) such that $\int k(y)dy = 1$ and $b > 0$ is the chosen bandwidth. If $k \ge 0$, then $k\big((x - x_i)/b\big)/b$ may be regarded as a probability density with location and scale parameters $x_i$ and $b > 0$. The function $g_b(\cdot, x_i)$ again possesses the weight $1/n$.

In summing up the single terms, one gets the kernel density

$$f_{n,b}(x) = \sum_{i \le n} g_b(x, x_i) = \frac{1}{nb} \sum_{i \le n} k\left(\frac{x - x_i}{b}\right)$$

which is a probability density if $k \ge 0$.

A very important matter in creating of the kernel density is the choice of the form of function $k(x)$ – the grid of the histogram. Below we present the general forms of the kernel functions:

$$k_{0.75}(x) = 0.75(1 - x^2)I(-1 \le x \le 1),$$

$$k_{0.5}(x) = 0.5 \times I(-1 \le x \le 1),$$

$$k_{0.125}(x) = 0{,}125(9 - 15x^2)I(-1 \le x \le 1),$$

$$k_{0.469}(x) = 0{,}46875(3 - 10x^2 + 7x^4)I(-1 \le x \le 1),$$

these two last kernels satisfy the additional condition $\int x^2 k(x)dx = 0$.

In an analogy to the choice of the grid the histogram – particularly of the bandwidth – the choice of an appropriate bandwidth $b$ is crucial for the performance of the kernel density.

If the bandwidth $b$ is small, which is related to a small scale parameter, then one can still recognize the terms $g_b(x, x_i)$ representing the single data. If $b$ is large, then an oversmoothing of the data may prevent the detection of certain clues in the data (see [Cleveland 1993]).

In the practical applications, in the first step we establish the bandwidth automatically. An automatic bandwidth selection is provided by cross–validation (see e.g. [Marron 1988] or [Simnoff 1996]). For finite sample sizes, the automatic choice of the bandwidth must be regarded as the first crude choice of the bandwidth. It is useful to vary bandwidth around the automatically selected parameter; e.g. decrease the bandwidth until the graph of the kernel density becomes bumpy.

If it is known that none of the observations is below or respectively, above a specific threshold – e.g. life spans are non negative or exceedances over a certain threshold $t$ exceed $t$ – the foregoing smoothing of data should not result in shifting weight below or above such thresholds. In this situation we take bandwidths that vary with the location.

If one realizes that there is a mode at a boundary point – as in the case of exponential density at zero – one should employ less smoothing around this point.

## 5. Examples of the practical applications

In this section, we present the applications of the kernel densities for fitting the generalized Pareto distributions to the data coming from the definite population and empirical data which come from the unidentified population.

In Figure 2 the exponential distribution with three different kernel densities are presented. The kernel densities are based on the sample $n = 100$, which comes from the population of the exponential distribution.

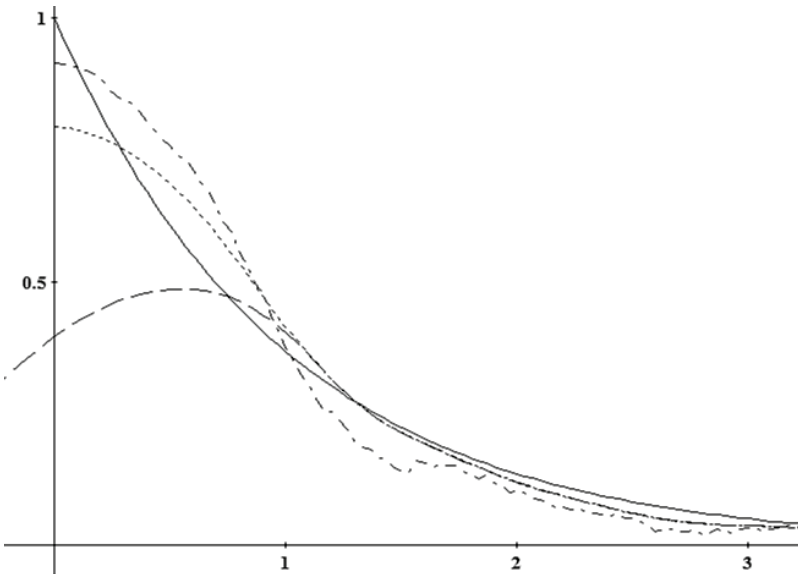


**Figure 2.** Exponential density (solid), kernel density with $k(x)$ function without bounded (dashed), kernel density with $k(x)$ function left bounded (dotted) and kernel density with $k(x)$ function left bounded (dashed – dotted)

Source: own study.

From Figure 2 one can see that for the left–side limited distribution one should apply the kernel density left bounded. Additionally, one can see that all three use the kernel densities very well, are fitted to the fat tail of the exponential distribution. Secondly, the tail of the left bounded kernel density with the $k(x)$ function is best fitted to the basic distribution. Also one see that the kernel density with the function is more smoothed than the one with the function. For all the kernel densities in Figure 2, the bandwidth was generated automatically.

In Figure 3 the Pareto density with $\alpha = 1.5$ and the kernel densities for the function with three different values of the bandwidth are presented. Clearly it is visible that the kernel density with the bandwidth equal to 0.5 is best fitted to the density Pareto. In this case also all the kernel density are well fitted in the tail of the density Pareto.

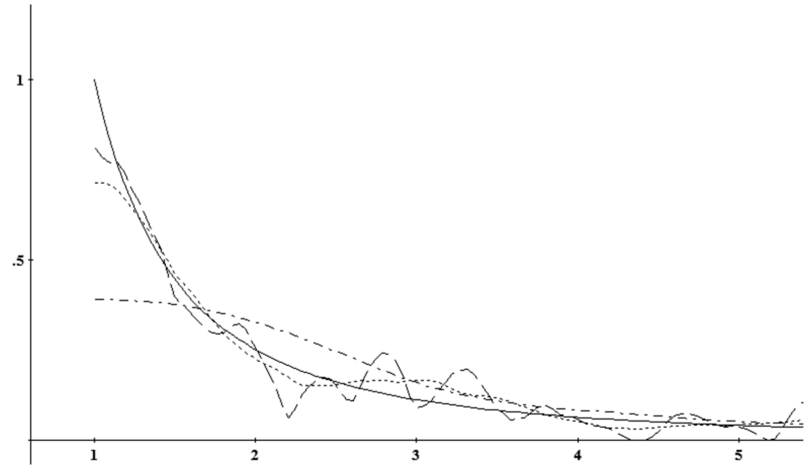


**Figure 3.** Pareto density with parameter alpha –1.5 (solid), kernel density with $k(x)$ function left bounded with bandwidth equal to 2 (dashed – dotted), equal to 0.5 (dotted) and equal to 0.2 (dashed)

Source: own study.

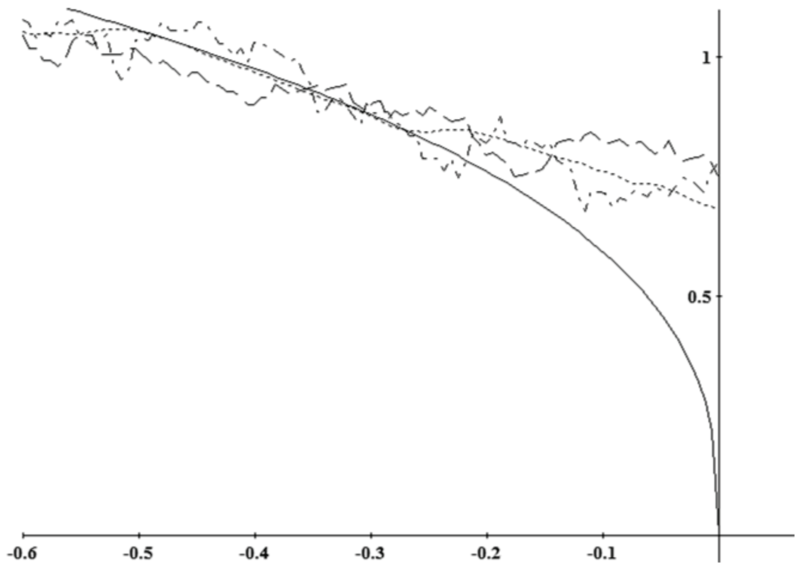


**Figure 4.** Beta density with parameter alpha –1.36 (solid), kernel densities with function and $b$ = 0.26 (dotted), with function and $b$ = 0.36 (dashed) and with function and $b$ = 0.38

Source: own study.

Figure 4 presents the density beta with $\alpha = -1.36$ and the kernel densities with the functions, and with three different bandwidths. In this case, the best fitted is the kernel density with function and $b$ = 0.26.

## 6. Final remarks

In this paper we considered three different tools to the visualization of the empirical distribution. The first two tools, that is the histogram for grouped data and the sample distribution function, were presented in less detail than the kernel densities. Additionally, in the paper the family generalized Pareto distributions were presented. The applications of the kernel densities were shown in examples of the generalized distributions and the samples coming from the generalized Pareto populations. In the examples we presented the use of the kernel densities for the different functions $k(x)$ and the bandwidths.

A fact worthy of notice is that the kernel density is a very good tool to approximate the distribution with fat tails. Namely, that the kernel density is useful in the modeling distributions of the extreme values.

## Literature

Coles S., *An Introduction to Statistical Modeling of Extreme Values,* Springer, London 2001.
Cleveland W.S., *Visualizing Data*, Hobart Press, New Jersey 1993.
Czekała M., *Statystyki pozycyjne w modelowaniu ekonometrycznym. Wybrane problemy,* Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2001.
David H.A., Nagaraja H.N., *Order Statistics,* John Wiley & Sons, Inc., 2003.
De Haan L., *Sample extremes: An elementary introduction,* Statist. Neerlandica 1976, 30, pp. 161-172.
Galambos J., *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York 1978.
Johnson N.L., Kotz S., *Distributions in Statistics: Continous Univariate Distributions -1,* Houghton Mifflin, Boston 1970.
Kuźmiński Ł., *Statystyki pozycyjne w prognozach ostrzegawczych*, [w:] *Zastosowanie metod ilościowych w ekonomii i zarządzaniu*, S. Forlicz (red.), CeDeWu, Warszawa 2012.
Magiera R., *Modele i metody statystyki matematycznej,* Oficyna Wydawnicza GiS, Wrocław 2002.
Marron J.S., *Automatic smoothing parameter selection: A survey*, "Empirical Economics" 1988, 13.
Leadbetter R., Lindgren G., Rootzen H., *Extremes and Related Properties of Random Sequences and Processes,* Springer – Verlag, New York, Heidelberg 1983.
Loynes, R.M., *Extreme values in uniformly mixing stationary stochastic processes,* "Ann. Math. Soc." 1965, 18, pp. 308-314.
Simnoff J.S., *Smoothing Methods in Statistics,* Springer, New York 1996.
Thomas M., Reiss R., *Statistical Analysis of Extreme Value with Applications to Insurance, Finance, Hydrology and Other Fields,* Birkhauser, Berlin 2007.

## ZASTOSOWANIA FUNKCJI JĄDRA GĘSTOŚCI DO MODELOWANIA UOGÓLNIONYCH ROZKŁADÓW PARETO

**Streszczenie:** W artykule prezentujemy narzędzia wykorzystywane w modelowaniu rozkładów charakteryzujących się ciężkimi ogonami. Prezentujemy trzy narzędzia; są nimi: dystrybuanta empiryczna, histogram dla pogrupowanych danych i funkcja jądra gęstości. Ostatnie z wymienionych narzędzi opisano w pracy bardzo szczegółowo. Przedstawione przykłady pokazują zastosowanie funkcji jądra gęstości do modelowania uogólnionych rozkładów Pareto.

**Słowa kluczowe:** funkcje jądra gęstości, wartości ekstremalne, uogólnione rozkłady Pareto.