# Chapter 3
# The Disaster Statistics for Various Natural Disasters

**Abstract** The application of statistical technique exposed in Chap. 2 to some concrete catalogs of natural processes and related losses are presented: global catalog of seismic moments; catalog of peak ground acceleration at five sites in Japan; catalog of victims of earthquakes-tsunamis, Japan; catalog of victims of floods, USA; catalog of economic losses from floods, USA; catalog of victims from tornadoes, USA; catalog of economic losses from hurricanes, USA. The Kolmogorov test is used as a powerful statistical tool for testing hypotheses on distribution under study. A modification of the Kolmogorov test is presented in the case of presence of estimated parameters in the hypothetical distribution function. Main statistical results are summarized in Tables 3.1, 4.1, 4.2 and 4.3.

**Keywords** Extreme value theory, EVT · Generalized Pareto distribution, GPD · "Peak-over-threshold" method · Intensity of point process · Non-stationarity of point process · Annual data

## 3.1 Earthquakes (Energy, Ground Acceleration)

### *3.1.1 The Global Harvard Catalog of Scalar Seismic Moments*

We use the method described above for the Harvard catalog of seismic moments within the period from January 1, 1976 to October 31, 2012. We restrict the depth of epicenters to $h \leq 70$ km and magnitudes to $m_W \geq 6.25$ (or seismic moments $\geq 2.985 \times 10^{25}$ dyne-cm). Note that this time interval contains two recent gigantic earthquakes: December 26, 2004, $m_W = 9.0$ (Sumatra) and March 11, 2011, $m_W = 9.1$ (Japan). To eliminate aftershocks from the catalog, the space–time window suggested in Knopoff et al. (1982) was used. Scalar seismic moments $M_s$ were converted into moment magnitude $m_W$ by the relation

V. F. Pisarenko and M. V. Rodkin, *Statistical Analysis of Natural Disasters and Related Losses*, SpringerBriefs in Earth Sciences,
DOI: 10.1007/978-3-319-01454-8_3, © The Author(s) 2014

$$m_W = (2/3) \cdot (\log_{10}(M_S) - 16.1).$$

The number of main shocks that has been left after aftershock elimination was $n = 1073$.

In order to test the Harvard catalog for stationarity we plotted in Fig. 3.1 the intensity (main events). We see that the seismic flow can be considered as fairly stationary. The general view of event flow is shown on Fig. 3.2. It confirms our conclusion about stationarity, but still there is a slightly increased density of events in the upper right corner of Fig. 3.2. We have checked this suspicion and plotted on Fig. 3.3 smoothed by 15-year window intensity of events with $m_W \geq 8.0$. We see that indeed there is some distinct increasing of such events in the last 15–20 years. We shall recall this fact later as we estimate quantiles $Q_q(\tau)$.

As we mentioned above, time moments $t_i$ of the stationary Poisson process are uniformly distributed on interval $[0; T]$ for any fixed sample size $n$. On Fig. 3.4 we compare the empirical DF of normalized time moments $s_i$ $\hat{F}_n(s)$ with uniform DF (diagonal of the square, $F(s) = s$). The normalization to the unite interval was effectuated by $s = (t - t_1)/(t_2 - t_1)$; $t_1$, $t_2$ are start and end of the catalog. The standard Kolmogorov test gives the Kolmogorov distance $D_n = \sqrt{n}$ max $|\hat{F}_n(s) - F(s)| = 1.127$ which corresponds to $p$ value $= 0.16$ (probability to exceed $D_n$ under condition that $\hat{F}_n(s)$ was generated by theoretical DF $F(s)$). Since this $p$-value ($p$–$v$) is more than 0.1, we formally have ground to accept the hypothesis of stationarity of $t_i$, although the $p$–$v$ is close to the boundary of rejection 0.1.

Now we apply to the Harvard catalog the statistical analysis exposed above (GPD fitting). The graph of the sample tail $1 - F(x)$ is shown on Fig. 3.5. We see that the tail decays rather slowly, and the recent gigantic earthquakes form a deviation from the general run. We take a grid of thresholds$h_j$ for magnitudes $m_W$ and fit both GPD

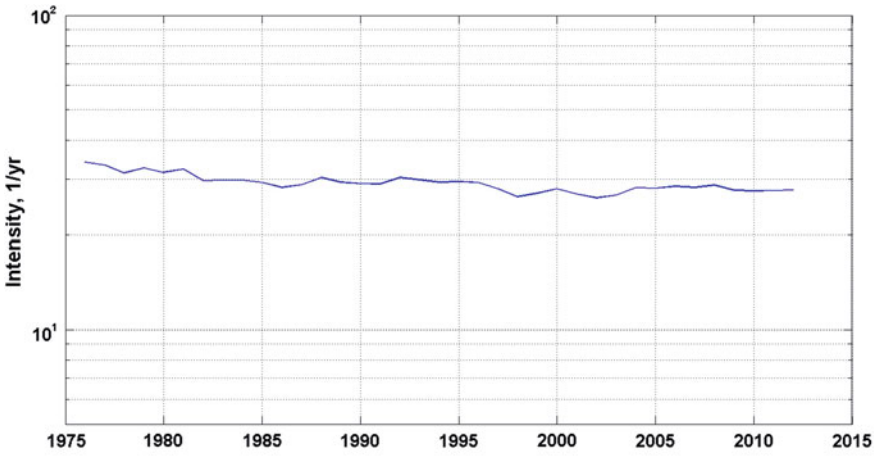


**Fig. 3.1** Intensity (main events, $m_W \geq 6.25$) of the Harvard catalog smoothed by 10-year time window

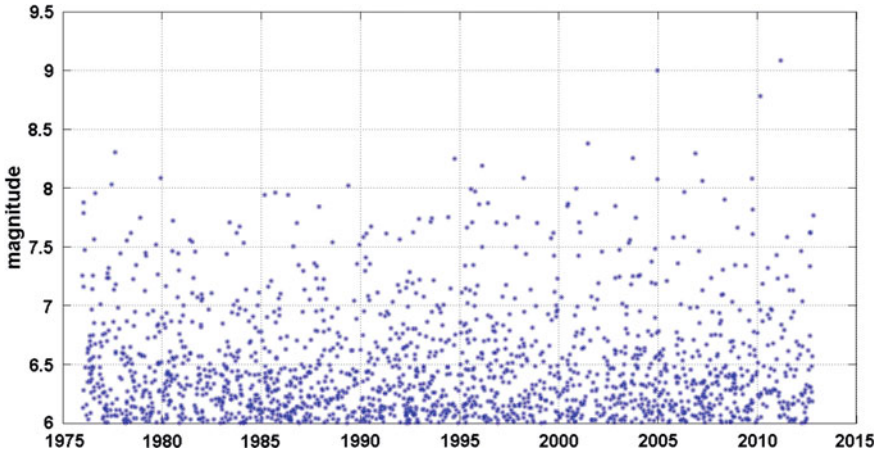


**Fig. 3.2** The time–magnitude diagram of the Harvard catalog (main events, $m_W \geq 6.0$)

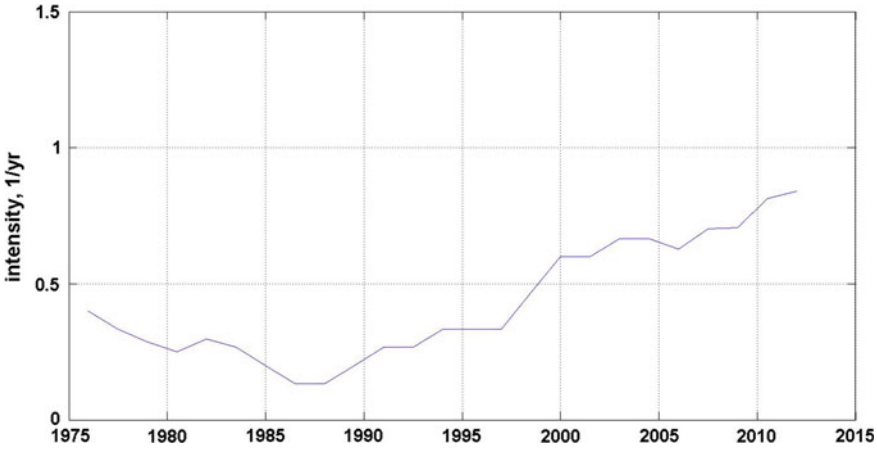


**Fig. 3.3** Intensity (main events, $m_W \geq 8.0$) of the Harvard catalog smoothed by 15-year time window

and exponential distributions (ED) for each $h_j$. The goodness of fit was tested by the Kolmogorov test. The results are shown on Fig. 3.7. We see that minimum KD-distance is reached by GPD at $h = 6.8$ ($KD = 0.609$). We have estimated by the simulation method the goodness of this fit with parameters corresponding to this threshold ($\hat{\xi} = -0.163 \pm 0.046$; $\hat{s} = 0.540 \pm 0.055$) and found out that $p$–$v$= 0.15. So, we took the threshold $h = 6.8$ and corresponding estimates. On Fig. 3.6 the extreme part of the tail used for parameter estimation is shown along with fitted GPD-curve. We can remark that the three largest earthquakes

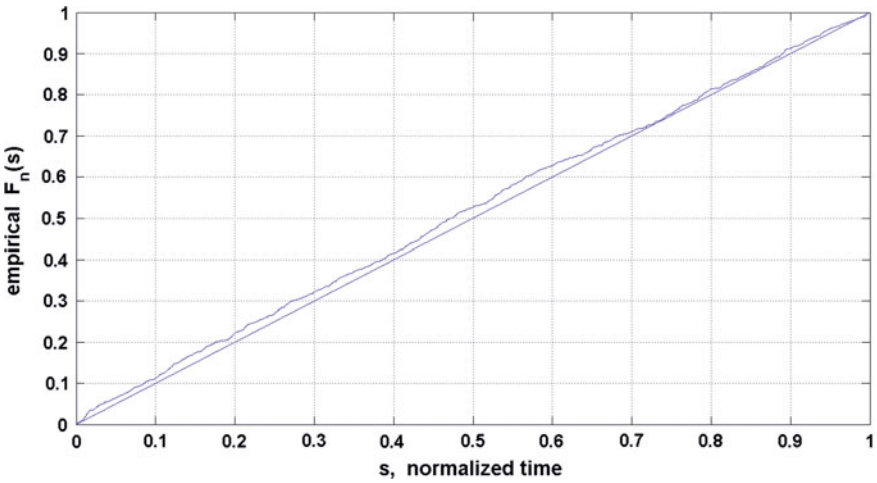


**Fig. 3.4** Empirical DF of normalized time moments of events of the Harvard catalog (main events, $m_W \geq 6.25$)

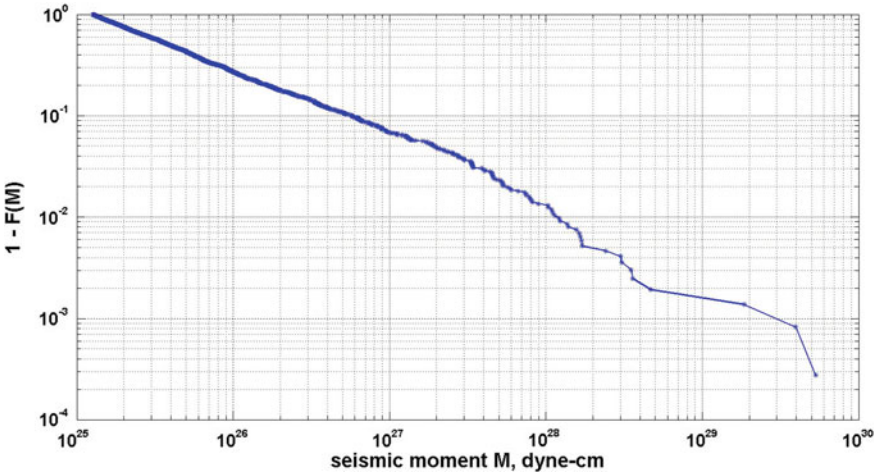


**Fig. 3.5** Sample tail $1 - F(x)$ of the Harvard catalog: main events, seismic moments, $M_s \geq 1.26 \times 10^{25}$ dyne-cm ($m_W \geq 6.0$)

occurred in the last decade ($m_W = 8.8,\ 27.2.2010;\ m_W = 9.0, 26.12.2004;\ m_W = 9.1, 11.3.2011$) made the approximating GPD-curve to deviate from observations in the range $8.0 \leq m_W \leq 8.5$ (recall Figs. 3.2 and 3.3). One can say that this compromise is chosen in accordance with statistical rules used in our procedure of fitting and corresponds to the best goodness-of-fit possible in this situation.

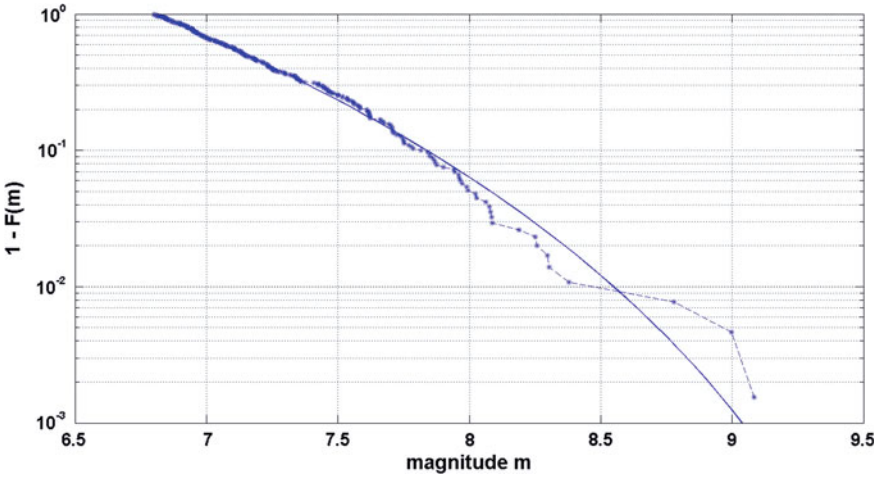


**Fig. 3.6** Harvard catalog, main events, moment magnitudes. The extreme tail $1 - F(x)$ and approximating GPD-tail: $h = 6.8$; $\xi = -0.163$; $s = 0.540$; $n = 324$

The number of events exceeding this threshold was $n = 324$, which is sufficient for a reliable statistical estimation. Parameter $M_{max}$ (right end-point of GPD) and 95 % quantile for future 10 years $Q_{0.95}(10)$ are:

$$M_{max} = h - \hat{s}/\hat{\xi} = 10.11;\ Q_{0.95}(10) = 9.13.$$

On Fig. 3.7 we see that besides of the best fitting point $h = 6.8$ there is one more point $h = 7.5$ whose KD is close to the best one. Let us compare parameters

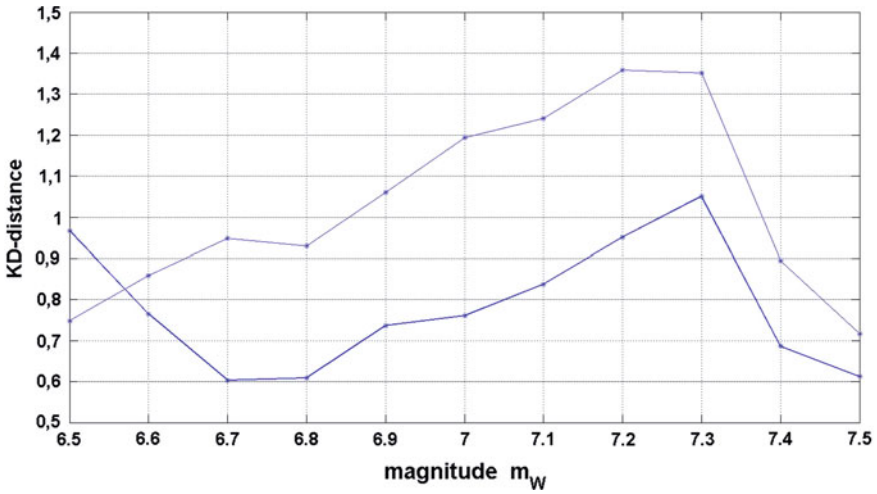


**Fig. 3.7** Harvard catalog, main events. The KD-distances for a grid of magnitude-thresholds $h_j$. *Thick line*—GPD-fitting; *thin line*—ED-fitting

of these two thresholds and demonstrate stability of the quantiles as compared with traditional parameter $M_{max}$. We have for threshold $h = 7.5$ following estimates:

$$\hat{\xi} = -0.0597 \pm 0.104;\ \hat{s} = 0.347 \pm 0.074;\ KD = 0.612;\ n = 82.$$

$$M_{max} = h - \hat{s}/\hat{\xi} = 13.31;\ Q_{0.95}(10) = 9.27.$$

We see that quantile $Q_{0.95}(10)$ has increased (due to random errors) insignificantly by 0.14, whereas $M_{max}$ has grown by 3.2! This example demonstrate clearly the instability of $M_{max}$ as compared with $Q_q(\tau)$.

Figure 3.8 shows the GPD-quantiles for three different confidence levels $q = 0.90; 0.95; 0.99$ with parameters corresponding to threshold $h = 6.8$. It is interesting to compare these quantiles with corresponding quantiles estimated on time interval (1976–2006), that were published in the book (Pisarenko and Rodkin 2010), not containing the great Japan earthquake March 11, 2011. Figure 3.9 shows such comparison for 95 %-quantiles. We see that the later estimates (1976–2012) provide slightly larger quantiles. This is a reaction on the great Japan earthquake March 11, 2011. In contrast with the moderate change of quantile values, change in corresponding $M_{max}$ values is much stronger that demonstrates again a weak stability of this parameter.

Note that even quantiles estimated for time interval before the occurrence of the great Tohoku earthquake (March 11, 2011, M9) show a quite high probability of occurrence of M9+ earthquakes. Besides, such events were not registered in Japan before both during the instrumental period of measurements and even among the historical earthquakes since 599 year (Usami 1979, 2002; Utsu 1979, 2002). Moreover, taking into account the considerable fragmentation of the lithosphere in

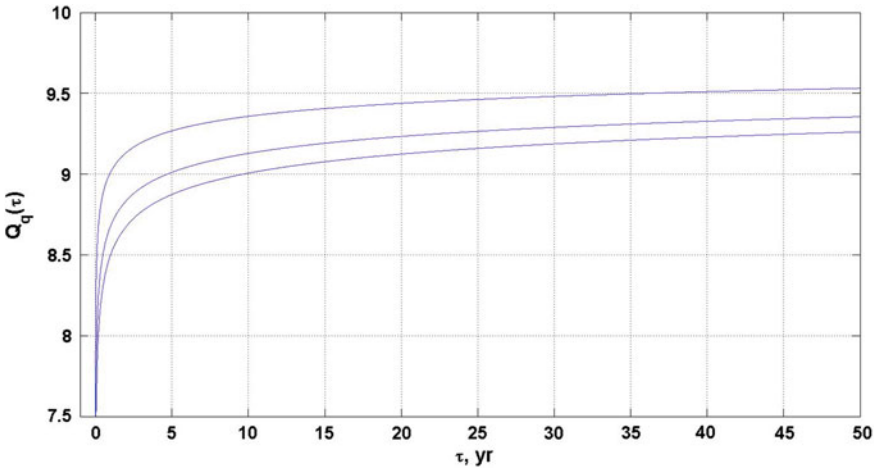


**Fig. 3.8** Harvard catalog, main events. GPD-quantiles $Q_q(\tau)$ of seismic magnitudes for three different confidence levels $q = 0.90$ (*lower curve*); 0.95 (*middle curve*); 0.99 (*upper curve*) with parameters corresponding to the threshold $h = 6.8$

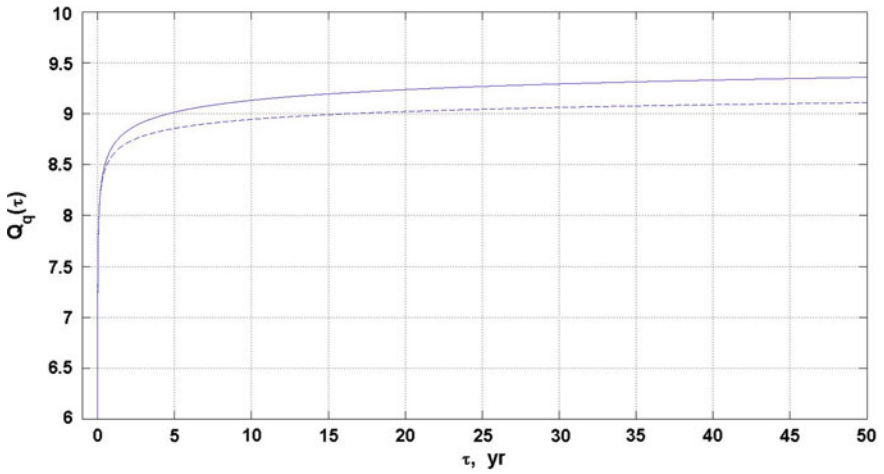


**Fig. 3.9** Harvard catalog, main events. 95 % GPD-quantiles $Q_q(\tau)$ derived from data 1976–2012 (*thin line*) and quantiles $Q_q(\tau)$ for period 1977–2006 (*dotted line*)

the Japan region and the absence of extended (approaching 1000 km long) unified segments of the Benioff zone, the very possibility of occurrence of the M9+ earthquakes in this region was negated by many seismologists. The occurrence of the Great Tohoku 2011 earthquake has confirmed however that our statistical estimate of possibility of occurrence of such earthquakes Pisarenko et al. (2010); Pisarenko and Rodkin (2013) was quite correct.

### *3.1.2 Estimation of Maximum Peak Ground Acceleration*

Although the main seismic parameters like *b*-slope, seismic activity rate and $M_{max}$ can be of considerable interest, estimation of peak ground acceleration, $A_{max}$, is of more practical importance in designing structures and in seismic risk assessment. The earthquake hazard has been estimated in a variety of ways (Lamarre et al. 1992; Kijko and Sellevoll 1989, 1992; Campbell 1981; Cornell 1968). The characterization of the seismic hazard at a fixed site is usually done through the probability of non-exceeding various levels of ground acceleration in a certain number of years, i.e. through the probability distribution function of maximum peak acceleration, for a given time period *T*. An equivalent, but perhaps more convenient characteristic of seismic hazard is furnished by the quantiles of this distribution function (we recall, that quantile is inverse function of the distribution function). Seismic hazard analysis involves several unknown parameters and relations: seismic activity rate $\lambda$, parameters of magnitude-frequency law, attenuation model (for ground acceleration), source model, soil characteristics, a model for earthquake sequence. Thus, it is necessary to estimate these parameters and establish step by step the needed

relations. Statistical and modeling uncertainties should be introduced at each of these steps. We are going to apply the statistical method exposed in Section II to evaluation of quantiles of distribution of peak ground acceleration.

We are interested in analysis of peak ground acceleration (PGA) $A_{max}$ at a particular site. Suppose this site is located at epicentral distance $R$ from source of earthquake of magnitude $m$. In the seismic hazard analysis there are a number of model relations giving approximate value of $A_{max}$ as some function of ($R$, $m$), see e.g. Cornell (1968). In most cases these relations have the following general form:

$$\log_{10}(A_{max}) = a + b \cdot m - c \cdot \log(R + d), \tag{3.1}$$

where *a, b, c, d* are some non-negative coefficients, $m$ is magnitude, $R$ is epicentral distance. Numerous modifications of (3.1) are used, but all of them keep the general property: monotone increase with $m$ and monotone decrease with $R$.

Let us consider flow of earthquakes in some space-magnitude window registered at a certain point. We denote magnitudes and epicentral distances to a fixed point as $(m_1, R_1), (m_2, R_2), \ldots, (m_n, R_n), \ldots$. We suppose that this flow is a stationary random process. Then any relation of type (3.1) or any arbitrary function $\Phi(m, R)$ will provide a stationary random process $\Phi(m_1, R_1)$, $\Phi(m_2, R_2), \ldots, \Phi(m_n, R_n) \ldots$. Thus, if we apply the relation (3.1) to the series $(m_1, R_1), (m_2, R_2), \ldots, (m_n, R_n), \ldots$. we can consider resulting sequence as a stationary random process. We call it *estimated acceleration*. The expression (3.1) differs from the true peak ground acceleration by a random term $\varepsilon$. We discuss this random error below. Our aim is to study statistical characteristics of the estimated acceleration with the statistical technique exposed in Chap. 2.

Detailed studies showed that the relation (3.1) is in some contradiction with empirical data in the near-field zone. In Mahdavian et al. (2005); Aptikaev (2009); Graizer and Kalkan (2011); Steinberg et al. (1993) it was shown that peak ground acceleration (PGA) practically does not depend on magnitude in a vicinity of the earthquake fault zone but depends on the type of the focal mechanism. The size of this zone $D$ usually varies from a few km to 10 km depending on the magnitude and can be well scaled according to the empirical law

$$\partial \log D / \partial m \cong 0.34. \tag{3.2}$$

To meet the near-field zone data we shall use the Aptikaev's relation (Lamarre et al. 1992) where the near-field effects are taken into account:

$$\log_{10}(A_{max}) = \begin{cases} 2.76; & \rho \le 1; \\ 2.76 - 0.55 \cdot \log_{10}(\rho); & 1 \le \rho \le 10; \\ 3.50 - 1.29 \cdot \log_{10}(\rho); & 10 \le \rho; \end{cases} \tag{3.3}$$

where

$A_{max}$ is estimated peak ground acceleration in cm/s$^2$;

$\rho = R \cdot 10^{-0.325(m-5)}$ magnitude scaled distance;

$R$ is epicentral distance in km;

$m$ is magnitude.

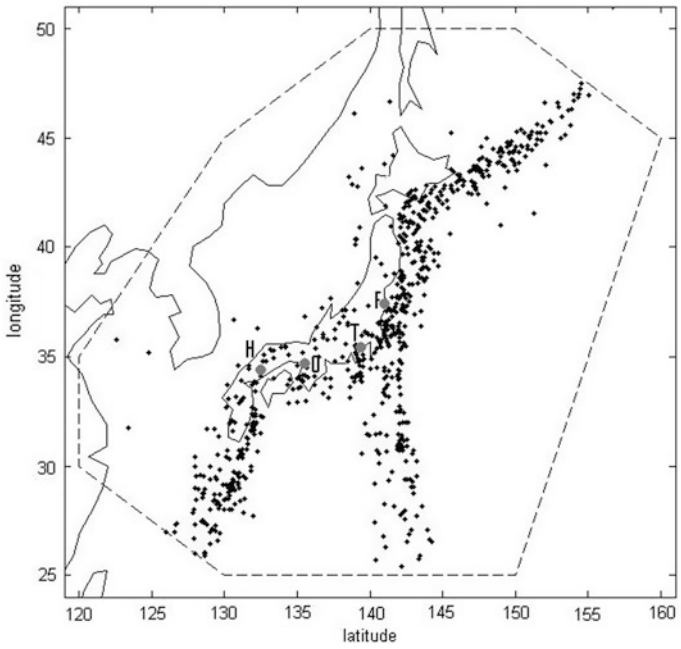


**Fig. 3.10** Japan earthquakes, 1900–2005. Tokyo, Hiroshima, Osaka, and Fukushima (atomic power station Fukushima Daiichi) are marked as 'T', 'H', 'O', and 'F'

Thus, relations (3.1) and (3.3) give logarithm of estimated peak ground acceleration $\log_{10}(A_{max})$ from an earthquake with magnitude $m$ at a site with epicentral distance $R$. The results obtained by means of both (3.3) and (3.1) are compared and discussed in Pisarenko and Rodkin (2013). In this paper we have derived quantiles of distribution of $\log_{10}(A_{max})$ for 4 points on the territory of Japan islands: Tokyo, Hiroshima, Osaka, and Fukushima (atomic power station Fukushima Daiichi). These points are shown on Fig. 3.10 where they are marked as 'T', 'H', 'O', and 'F'. We have used the earthquake catalog of the Japanese Meteorological Agency (JMA) over time period 1900–2005.

***Tokyo,* $\lambda$ = 35.41; $\varphi$ = 139.36;**

We have applied the statistical technique exposed in Chap. 2 to the estimated accelerations calculated from equations (3.3). The Kolmogorov distance $KD$ was used for to choose the most appropriate threshold value $h$ providing the best fitting of $GPD$ to the data:

$$KD = n_h^{1/2} \max |GPD_h(x|\hat{\xi}, \hat{s}) - F_{n_h}(x)|, \tag{3.4}$$

where $F_{n_h}(x)$ is sample stepwise distribution function generated by observations $(x_1 \leq \ldots \leq x_{n_h})$ exceeding threshold $h$:

$$F_{n_h}(x) = \begin{cases} 0; & x \le x_1; \\ r/n_h; & x_r < x \le x_{r+1}; \quad 1 < r < n_h; \\ 1; & x > x_{n_h}. \end{cases}$$

Since we use a theoretical GPD with parameters fitted to the data, we cannot use the standard Kolmogorov distribution tables to find the significance level of the observed *KD*. Instead, in order to determine the significance level of a given *KD*-distance (3.4), we used a numerically calculated distribution of *KD*-distances measured in a simulation procedure with 10,000 GPD-samples and parameters individually fitted to each sample. This method was suggested in Stephens (1974) for the Gaussian and the exponential distributions. We use the Kolmogorov distance to test the GPD distribution fitted to estimated accelerations.

The histogram of estimated accelerations calculated in accordance with (3.3) is shown on Fig. 3.11 (upper figure). We see that a monotone decreasing starts somewhere near $\log_{10}(A_{max}) \sim -0.2$. Since the theoretical GPD-density monotonically decreases, we have to restrict our analysis by thresholds $h > -0.2$. The *KD*-distance as function of $h$ is shown on Fig. 3.11, lower figure. We see that the lowest $KD = 0.582$ (the best fitting) corresponds to the threshold $h = 0.5$. Its significance level ($p$–value) equals to 0.72, so that the sample can be considered as belonging to GPD distribution (the testing would reject this distribution in the case of very small $p$–values, say, $p < 0.10$). For this threshold there are $n_h = 279$ observations exceeding this threshold. We got following estimates of unknown parameters:

$$\hat{\xi} = -0.23 \pm 0.05; \quad \hat{s} = 0.52 \pm 0.06; \quad Q_{0.90}(30) = 2.3. \tag{3.5}$$

Maximum of logarithmic estimated accelerations was 2.76 (it cannot be more because of restriction of estimated acceleration as it is given in equation (3.3)).

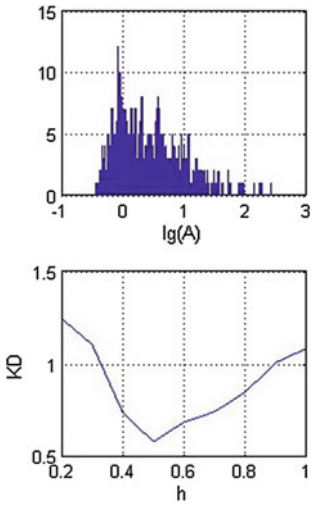


**Fig. 3.11** Tokio. The histogram of estimated accelerations (3.3) (*upper figure*). The *KD*-distance as function of $h$ (*lower figure*)

***Hiroshima, λ = 34.39; φ = 132.46***;

The histogram of estimated accelerations calculated in accordance with (3.3) is shown on Fig. 3.12, upper figure. We see that a monotone decreasing starts somewhere near $\log_{10}(A_{max}) \sim 0.1$. We restrict our analysis by thresholds $h$ in the interval (0.1; 0.9). The *KD*-distance as function of $h$ is shown on Fig. 3.12, lower figure. We see that there are 3 thresholds $h = 0.6; 0.7; 0.8$ with *KD* close to 0.6. We prefer to take $h = 0.6$ since sample size for this threshold ($n_h = 118$) is larger than others ($n_h = 95$; $n_h = 74$). Its significance level equals to 0.593, so that the sample can be considered as belonging to GPD distribution. We got following estimates of unknown parameters:

$$\hat{\xi} = -0.21 \pm 0.07; \quad \hat{s} = 0.46 \pm 0.08; \quad Q_{0.90}(30) = 2.1. \tag{3.6}$$

Maximum of logarithmic estimated accelerations was 2.76 (it cannot be more because of restriction of estimated acceleration in Eq. (3.3)).

***Osaka, λ = 34.69; φ = 135.50***;

The histogram of estimated accelerations calculated in accordance with (3.3) is shown on Fig. 3.13, upper figure. We see that a monotone decreasing starts somewhere near $\log_{10}(A_{max}) \sim 0.1$.We restrict our analysis by thresholds $h$ in the interval (0.1; 0.6). The *KD*-distance as function of $h$ is shown on Fig. 3.13, lower figure. We see that the best fitting corresponds to the threshold $h = 0.1$ with $KD = 0.62$. Its $p$–value equals to 0.63, so that the sample can be considered as belonging to GPD distribution. We got the following estimates of unknown parameters:

$$\hat{\xi} = -0.19 \pm 0.04; \quad \hat{s} = 0.50 \pm 0.04; \quad Q_{0.90}(30) = 2.1. \tag{3.7}$$

Maximum of logarithmic estimated acceleration was 2.76 (as above it cannot be more because of restriction of estimated acceleration in equation (3.3)).

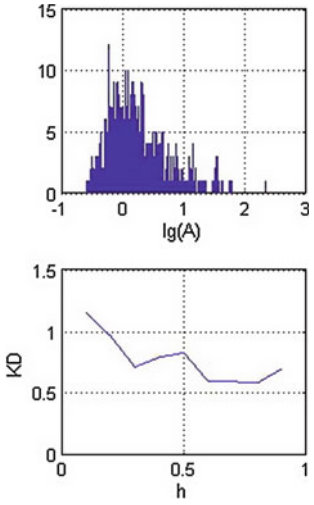


**Fig. 3.12** Hiroshima. The histogram of estimated accelerations (3.3) (*upper figure*). The *KD*-distance as function of $h$ (*lower figure*)

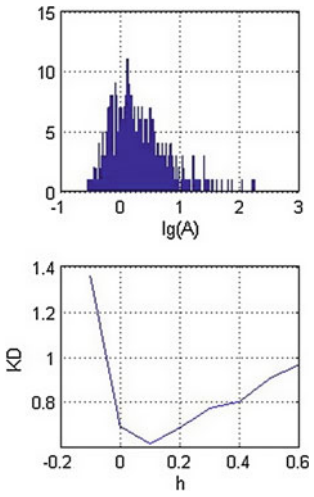


**Fig. 3.13** Osaka. The histogram of estimated accelerations (3.3) (*upper figure*). The *KD*-distance as function of $h$ (*lower figure*)

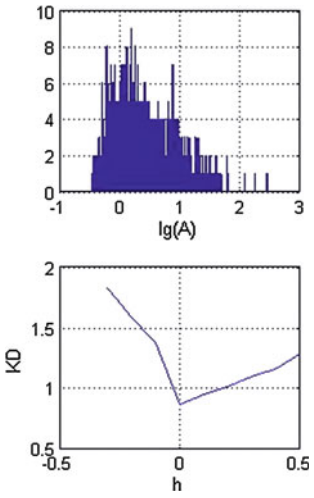


**Fig. 3.14** Fukushima. The histogram of estimated accelerations (3.3) (*upper figure*). The *KD*-distance as function of $h$ (*lower figure*)

***Fukushima Daiichi,* $\lambda$ = 37.4214*;* $\varphi$ = 141.0325**;

The histogram of estimated accelerations calculated in accordance with (3.3) is shown on Fig. 3.14, upper figure. We see that a monotone decreasing starts somewhere near $\log_{10}(A_{max}) \sim 0.2$.We restrict our analysis by thresholds $h$ in the interval (0.0; 0.5). The *KD*-distance as function of $h$ is shown on Fig. 3.14, lower figure. We see that the best fitting corresponds to the threshold $h = 0$ with $KD = 0.87$. Its $p$-value equals to 0.13, so that the sample still can be considered as belonging to GPD distribution. We got following estimates of unknown parameters:

**Table 3.1** Statistical estimates of *GPD* parameters fitted to the estimated acceleration data

| | $h$ | $n_h$ | $\xi$ | $h - s/\xi$ | $Q_{0.90}$ (30) | $p$-value of *GPD* |
|---|---|---|---|---|---|---|
| Tokyo | 0.5 | 279 | $-0.23 \pm 0.05$ | 2.76 | 2.3 | 0.72 |
| Hiroshima | 0.6 | 118 | $-0.21 \pm 0.07$ | 2.79 | 2.1 | 0.69 |
| Osaka | 0.1 | 462 | $-0.19 \pm 0.04$ | 2.73 | 2.1 | 0.63 |
| Fukushima | 0 | 544 | $-0.29 \pm 0.03$ | 2.62 | 2.3 | 0.13 |

$$\hat{\xi} = -0.29 \pm 0.03; \quad \hat{s} = 0.76 \pm 0.06; \quad Q_{0.90}(30) = 2.3. \tag{3.8}$$

Maximum of logarithmic estimated accelerations was 2.63.

The results of estimation of the parameters of GPD distribution fitted to the estimated acceleration data for the mentioned four sites are summarized in Table 3.1. In the fifth column the maximum possible values of the estimated acceleration are shown calculated by formula $\log(A_{max}) = h - s/\xi$. $Q_{0.90}(30)$ is the quantile of level $q = 0.90$ for the maximum estimated acceleration in a future time interval of 30 years.

### 3.1.3 The Accounting for Inaccuracy of the Estimated Acceleration

The estimated acceleration (3.3) differs from the true acceleration by a random value $\varepsilon$. We assume that

$$\varepsilon = \varepsilon_1 + \varepsilon_2,$$

where $\varepsilon_1$, $\varepsilon_2$ are independent random errors; $\varepsilon_1$ refers to inaccuracy of the used relations (3.3) and $\varepsilon_2$ characterizes the influence of the seismic source mechanism on the ground acceleration. In accordance with Aptikaev (2009) the random error of the relation (3.3) has standard deviation std $(\varepsilon_1) = 0.18$. The distribution of $\varepsilon_1$ is not critical: we compared on several artificial examples the Gauss distribution and the uniform distribution and found no essential differences in the estimates of quantiles $Q_q(\tau)$. So, we accept for $\varepsilon_1$ the Gauss distribution. In order to evaluate std$(\varepsilon_2)$. we suppose that all sources in Japan territory can be classified into three types with following relative frequencies:

$$\begin{aligned} \text{normal fault} &\sim 15\ \%; \\ \text{strike} - \text{slip} &\sim 20\ \%; \\ \text{inverse fault (thrust)} &\sim 65\ \%. \end{aligned} \tag{3.9}$$

These relative frequencies are taken from the regional earthquakes focal mechanism data Zlobin and Polets (2012), and they reflect the predominance of the compression tectonic forces in the Japan region. Following Aptikaev (2009) we assume further that these source types produce correspondingly in the epicentral zone the following mean peak ground accelerations (PGA):

$$\begin{aligned} \log_{10}(A) &= 2.65; \\ \log_{10}(A) &= 2.76; \\ \log_{10}(A) &= 2.95. \end{aligned} \tag{3.10}$$

We have used here the mean PGA value $\log_{10}(A) = 2.76$ valid for a totality of earthquakes with different types of focal mechanisms instead of that typical of strike-slip events and equal to 2.80 Aptikaev (2009). It gives us a possibility to use a simpler scheme of taking into account the difference of mean PGA values in the cases of different focal mechanisms. If we knew the source mechanism for each earthquake in our catalog we could take this information into account, but since it is unknown for the JMA catalog (at least for the first half of this catalog), we have to model the influence of the source mechanism by an additional random term $\varepsilon_2$. The mean value of random variable taking values (3.10) with probabilities (3.9) is $\sim 2.76$ and standard deviation is 0.15. Thus, we can accept that std of $\varepsilon_2$ is 0.15. We suppose that the distribution of $\varepsilon_2$ is the Gaussian as well. Then the error $(\varepsilon_1 + \varepsilon_2)$ has standard deviation 0.23. Thus, we can assume that the maximum estimated acceleration analyzed in the previous section differs from the true maximum ground acceleration by a random Gaussian error with zero mean and $std = 0.23$. So, we have to take into account the influence of this random error on the quantile $Q_q(\tau)$. We have done it by a simulation procedure, adding a random Gaussian rv with $std = 0.234$ to the GPD-random variable with estimated parameters (see Table 3.1) and repeating this operation 10,000 times. Figure 3.15 shows the quantiles $Q_q(\tau)$. both with error term $\varepsilon = (\varepsilon_1 + \varepsilon_2)$ (heavy curves) and without it (light curve) for all four points under analysis. We see that the accounting for errors is practically reduced to an increase of the undisturbed quantile by one std of the error.

The quantile $Q_q(\tau)$ is a final result of our statistical technique. It is a robust and meaningful characteristic of the seismic hazard.

It should be remarked that the estimation of the peak ground acceleration depends in a large extent on the used relation of type (3.1) or (3.3) connecting the log-acceleration with magnitude and distance. We have used the relation (3.3) due to Aptikaev (2009) that does not take into account the regional or local features connected with soil properties. Of course, relation taking into account regional and geological pecularities of the site would be preferable. So, our results exposed above might be considered as preliminary estimation of real acceleration and illustration of our statistical method for this problem.

## 3.2 Earthquakes, Tsunami (Victims)

In this section we shall use revised and extended version of the catalog of earthquake victims in Japan composed by Utsu (1979, 2002). The catalog includes as well victims of tsunami, caused by earthquakes, and covers time period 1900–2012. It contains two types of data: fatalities and the injured.

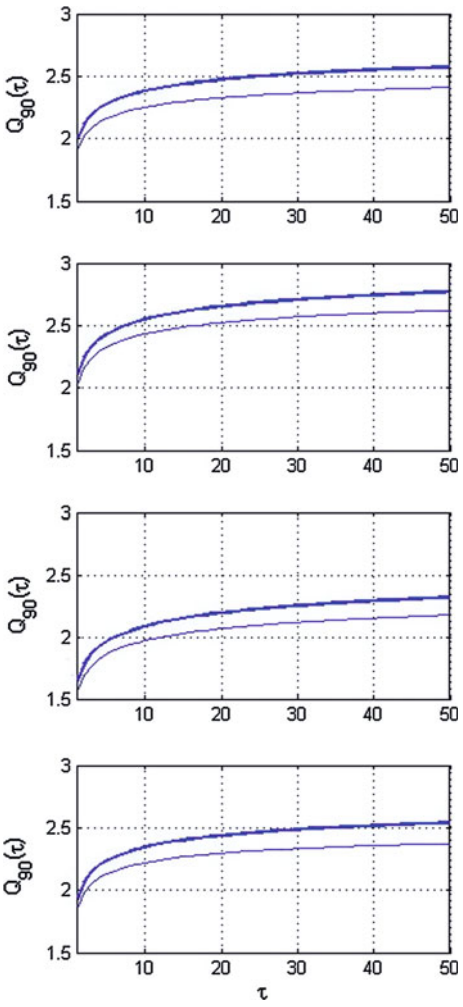


**Fig. 3.15** GPD-quantiles $Q_{0.90}(\tau)$ of estimated acceleration both with error term $\varepsilon = (\varepsilon_1 + \varepsilon_2)$ (*heavy curves*) and without it (*light curves*) for all four points under analysis. From *top* to *bottom*: Tokyo, Hiroshima, Osaka, Fukushima

**(a)** ***Fatalities***

The catalog contains fatality data of 73 earthquakes. Figure 3.16 shows general view of these data. The maximum fatality (142,807) occurred at 1923, Tokio. Such high number can be explained by gigantic fires that struck old houses and structures existed at that time. So, if one is interested only in possibilities of similar disasters in the future this observation, perhaps, could be eliminated from the sample, but we left it nevertheless, counting for the robustness of our method. Looking at Fig. 3.16 we see that the intensity of events visually slightly decreases with time, although this effect is not very strong. Perhaps, this decrease can be explained by more safe modern structure and more effective preventive measures. It should be noted that this effect competes with the expected increase in a number of fatalities because of the natural growth of population.

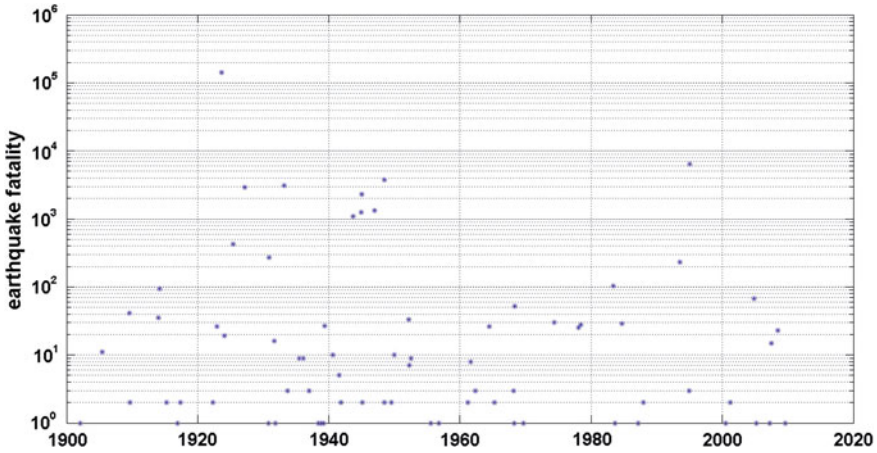


**Fig. 3.16** The fatality events of the catalog of earthquake victims (Japan) composed by Utsu, 1900–2012

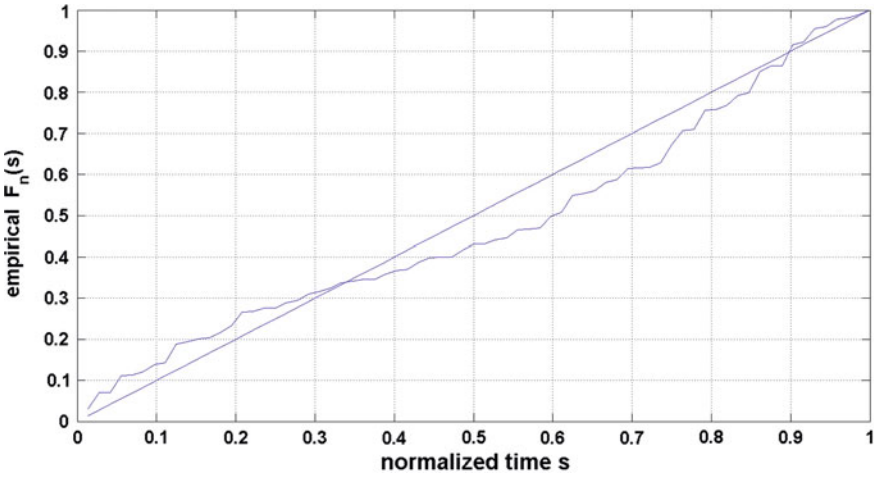


**Fig. 3.17** Catalog of earthquake victims, Japan, 1900–2012. The empirical distribution function $F_n(s)$) of normalized occurrence times $s_j$ compared with the uniform distribution function (*diagonal line*)

As it was mentioned above for the stationary Poisson process time moments $t_i$ are uniformly distributed on interval $[0;\ T]$. On Fig. 3.17 we compare the empirical DF of normalized time moments $s_i$ $\hat{F}_n(s)$ with uniform DF ($F(s) = s$; $0 \le s \le 1$). The standard Kolmogorov test gives the Kolmogorov distance $D_n = \sqrt{n}\ max\ |\ \hat{F}_n(s) - F(s)| = 1.004$ which corresponds to $p$-value $= 0.23$ (probability to exceed $D_n$ under condition that $\hat{F}_n(s)$ was generated by theoretical

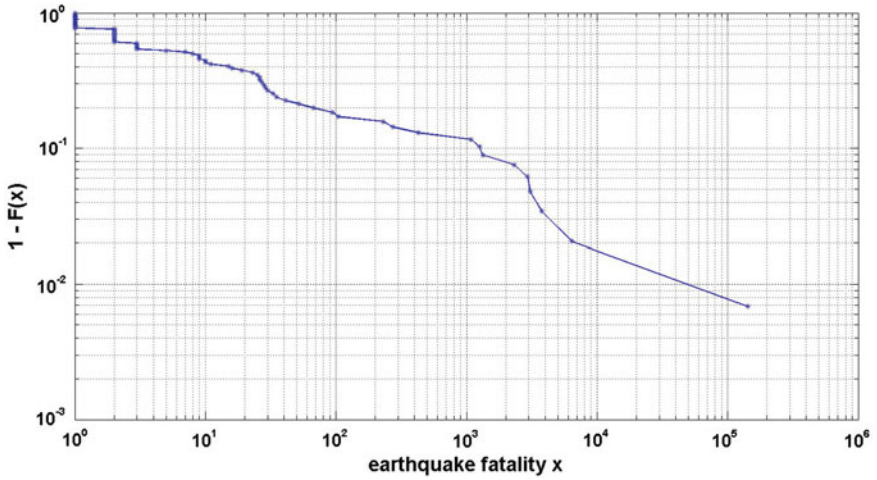


**Fig. 3.18** Catalog of earthquake victims, Japan, 1900–2012. The tail graph of fatalities

DF $F(s)$. Since this $p$-value ($p$–$v$) is not sufficiently small (more than 0.1), we can accept the hypothesis of stationarity of $t_i$.

Now we consider the fatalities figures. The tail graph is shown on Fig. 3.18. We see that on the whole the tail $1- F(x)$ has power-like behavior with appreciable irregularities. So, it is reasonable to take logarithms and to analyze $\log(x)$. The concentration of victims in the extreme range of the catalog is very high: it turns out that 10 % of the most disastrous events are responsible for 98 % of the total number of perished. Even if we exclude maximum fatality case (142, 807 occurred at 1923, Tokio) 10 % of the most disastrous events would make up 83 % of the total number of perished.

On Fig. 3.20 the cumulative sums of log-fatalities are shown. In spite of some local fluctuations the trend on the whole looks like a linear function. There are some deviations, but they have no definite tendency. So, we accept the hypothesis of stationarity of this catalog.

Now we apply to our catalog the statistical analysis exposed above. We take a grid of thresholds $h_j$ for $\log(x)$ and fit both GPD and ED for each $h_j$. The goodness of fit is measured by the Kolmogorov distance. The results are shown on Fig. 3.21. We see that minimum KD-distance 0.665 is reached for GPD-approach at $h = 0.45$. The GPD fitting includes estimation of two parameters ($\xi$, s), whereas the standard Kolmogorov testing assumes no unknown parameter. For this reason we cannot use directly the standard $p$–$v$ from tables of the Kolmogorov distribution. In order to calculate $p$–$v$ we used the simulation method as we mentioned above. The best fit of GPD-distribution for $h = 0.45$ ($\xi = -0.260 \pm 0.111$; $s = 1.657 \pm 0.430$) provides $p$–$v = 0.433$, which allows to accept the GPD. Figure 3.19 shows the extreme part of the tail used for parameter estimation along

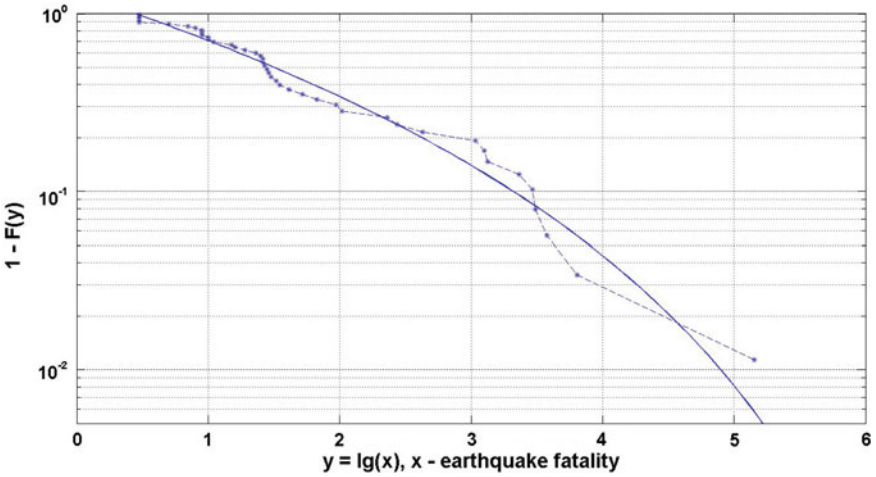


**Fig. 3.19** Catalog of earthquake victims, Japan, 1900–2012. The extreme tail $1 - F(y)$ ($y = \log(x)$, $x$—number of dead) and approximating GPD-tail: $h = 0.45$; $\xi = -0.260$; $s = 1.657$; $n = 44$

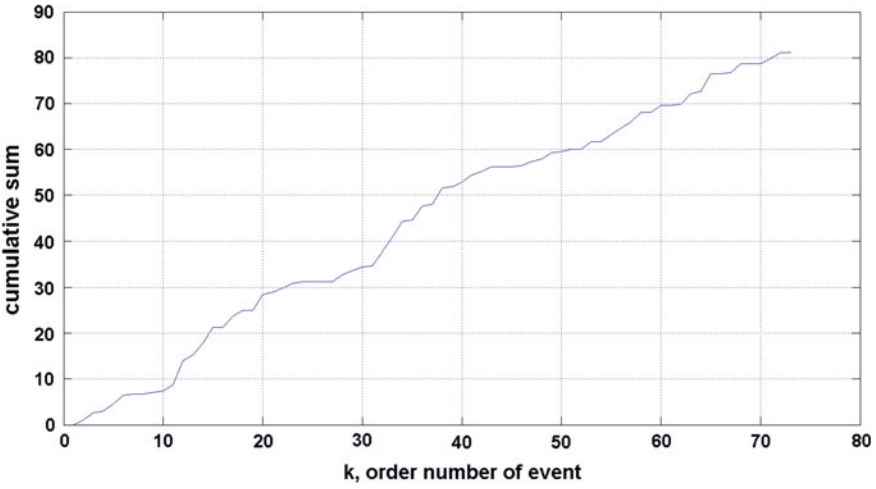


**Fig. 3.20** Catalog of earthquake victims, Japan, 1900–2012. The cumulative sums of log-fatalities

with fitted GPD-curve. We see that observations exhibit some irregular oscillations around the approximating GPD-curve, but on the whole the fitting is satisfactory.

Now we are able to calculate the quantiles $Q_q(\tau)$ which are the final goal of our estimation. Figure 3.22 shows the GPD-quantiles for 3 different confidence levels $q = 0.90$; 0.95; 0.99. It should be remarked that for very small $\tau$ the quantiles

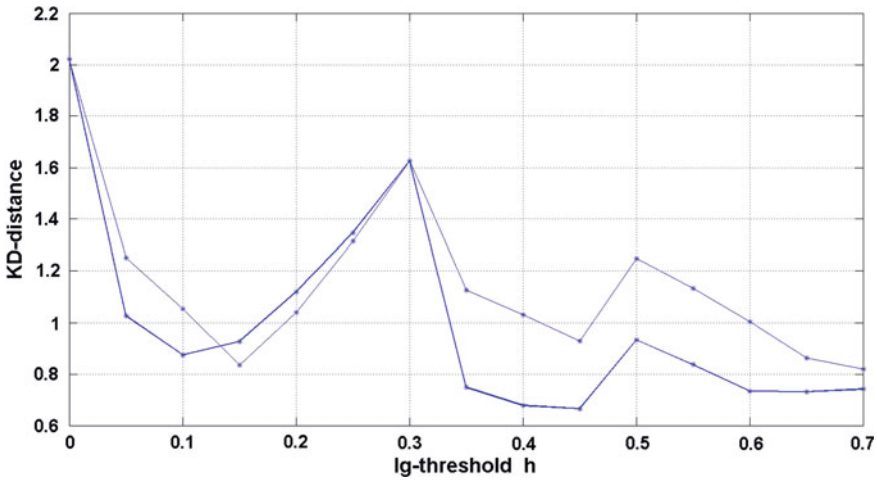


**Fig. 3.21** Catalog of earthquake victims, Japan, 1900–2012. The KD-distances for a grid of log-thresholds $h_j$. *Thick line*—GPD-fitting; *thin line*—ED-fitting

takes zero values and have small jumps equal to $exp(-\lambda\tau)$ which is the probability that there is no event on the interval $[0; \tau]$.

It is interesting to calculate "maximum possible size", i.e. the rightmost limit $M_{max}$ of the GPD for $\log(x)$:

$M_{max} = h - s/\xi = 6.83$ (this corresponds to 6,800,000 fatalities).

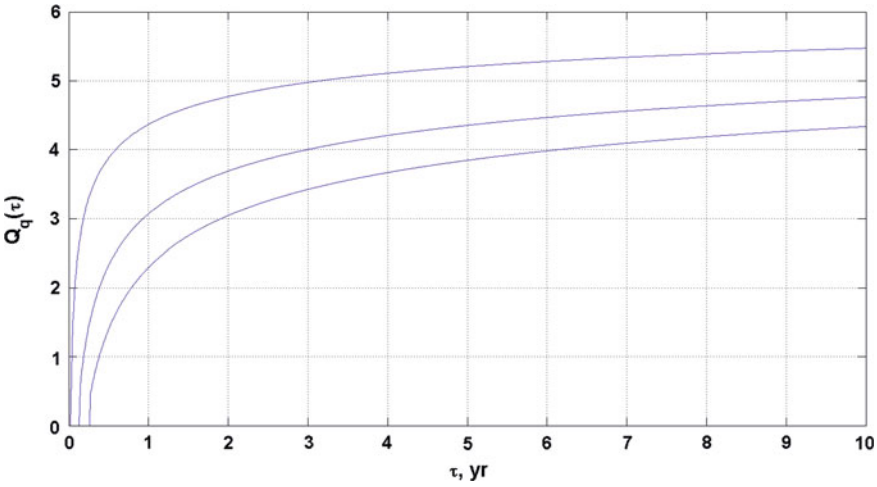


**Fig. 3.22** Catalog of earthquake victims, Japan, 1900–2012. The GPD-quantiles for three different confidence levels $q = 0.90$ (*lower curve*); 0.95 (*middle curve*); 0.99 (*upper curve*)

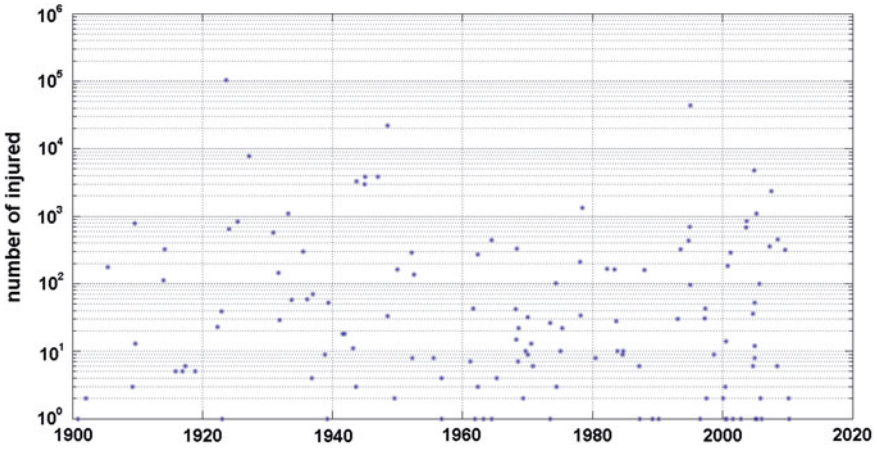


**Fig. 3.23** The injured by earthquakes. Catalog of earthquake victims (Japan), 1900–2012

Such gigantic figure appears to be unreal, it hardly can be used as a useful statistical characteristic of real fatalities and has little practical value. In the same time the quantile $Q_{0.95}(10) = 5.66$ (460,000) looks quite realistic. This comparison shows once more stability of the quantiles $Q_q(\tau)$ with respect to the "maximum possible size" parameter.

**(b)** ***The injured***

The catalog contains numbers of injured in 131 earthquakes. Figure 3.23 shows these numbers in time. The maximum number (103,733) occurred at 1923, Tokio. As we mentioned above such high number of dead and injured can be explained by gigantic fires. Looking at Fig. 3.23 we see that the intensity of events visually slightly increases with time, although this effect is not very strong. Perhaps, this increase can be explained by more attentive registration of events with minor number of injured.

On Fig. 3.24 we compare the empirical DF of normalized time moments $s_i$ $\hat{F}_n(s)$ with uniform DF. We see that there is a certain down deviation of the empirical DF from the diagonal, which testifies that the visual effect of an intensity increase is real. The Kolmogorov distance $D_n = 2.048$ corresponds to $p$–$v = 0.0005$ which makes us to reject the hypothesis of stationarity of the catalog. Looking at Fig. 3.23 we can suspect that the non-stationarity is caused by weaker events. In order to check this suspicion we tried several lower thresholds. The resulting intensities smoothed by 15-year time window are shown on Fig. 3.25. We see that the evident non-stationary intensity of the original catalog ($h = 0$) decreases with growing $h$ and practically vanishes at $h = 400$. This threshold seems high (it is left only 23 observations above this threshold), but, fortunately, just this threshold provides the best GPD-fitting, as we shall see below. The empirical DF of normalized event times for $h = 400$ is shown on Fig. 3.26.

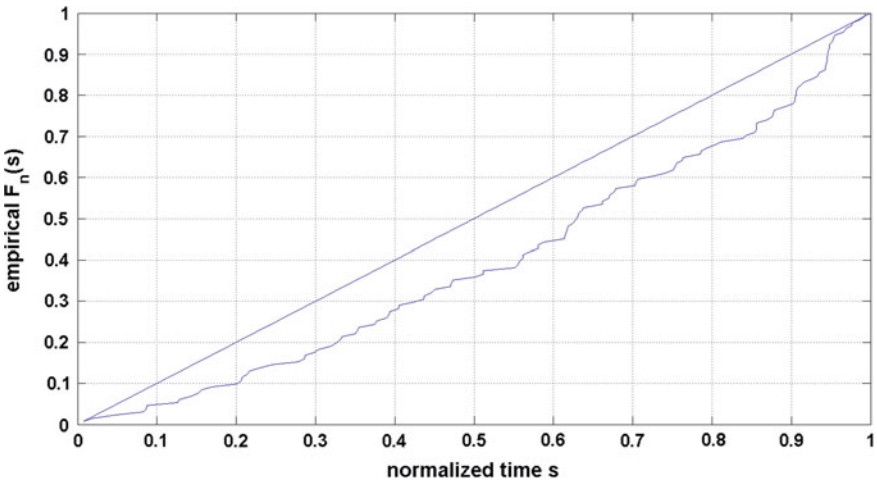


**Fig. 3.24** The injured by earthquakes. Catalog of earthquake victims (Japan), 1900–2012. The empirical distribution function $F_n(t)$ of normalized occurrence times $t_j$ compared with the uniform distribution function (*diagonal line*)

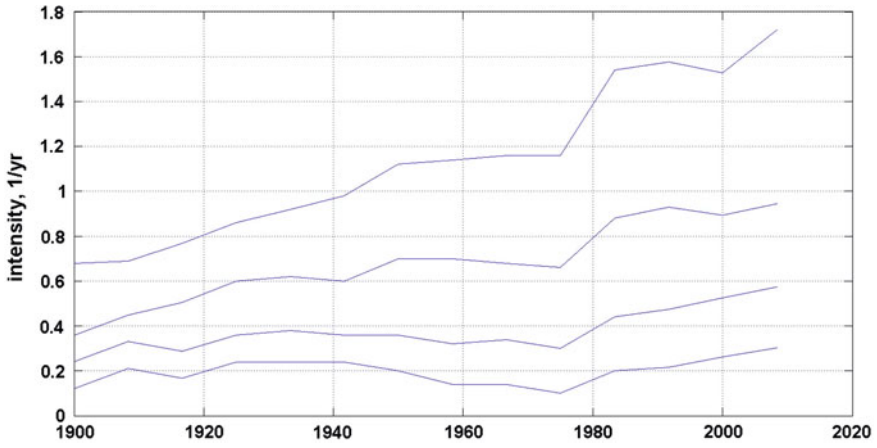


**Fig. 3.25** The injured by earthquakes. Catalog of earthquake victims (Japan), 1900–2012. Intensities for four thresholds from bottom to top: $h_1 = 1$ ($n_1 = 131$); $h_2 = 10$ ($n_2 = 78$); $h_3 = 100$ ($n_3 = 44$); $h_4 = 400$ ($n_4 = 23$), smoothed by 15-year time window

The Kolmogorov distance KD = 1.21 which corresponds to $p$–$v = 0.11$. This value is on the border of acceptance, but still more than 0.10 and we can accept with some reservation the hypothesis of the stationarity of the event times for $h = 400$. Now we check for stationarity the distribution of injured. Figure 3.27 shows the cumulative sums of $\log(x_j)$, $x_j$—numbers of injured. We see that

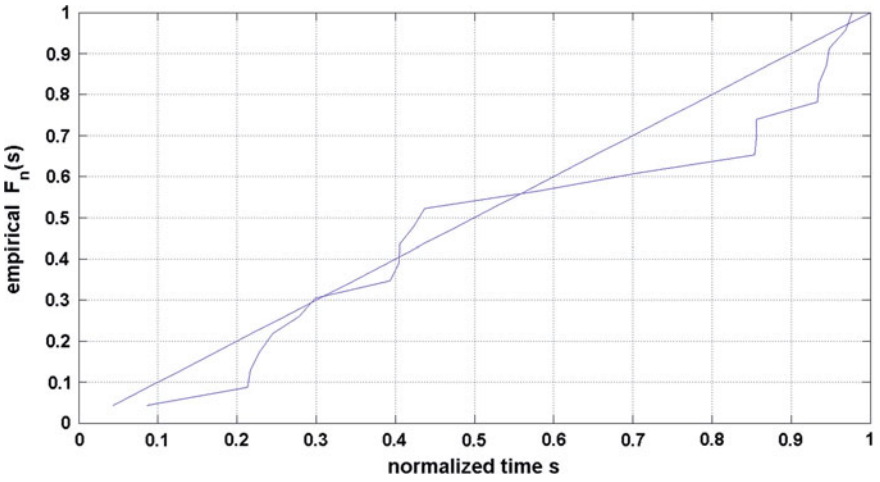


**Fig. 3.26** The injured by earthquakes. Catalog of earthquake victims (Japan), 1900–2012. The empirical DF of normalized event times for $h = 400$. The empirical distribution function $F_n$(s) of normalized occurrence times $s_j$ compared with the uniform distribution function (*diagonal line*)


**Fig. 3.27** The injured by earthquakes. Catalog of earthquake victims (Japan), 1900–2012. The cumulative sums of $\log(x_k)$, $x_k$—numbers of injured

deviations from a straight line are not significant, and we can accept the hypothesis of the stationarity of distribution of injured. So, we accept the hypothesis of the stationarity of this catalog.

The tail graph of injured is shown on Fig. 3.28. We see that on the whole the tail 1– F($x$) has power-like behavior with possible increasing inclination. So, it is reasonable to take logarithms and to analyze $\log(x)$. Figure 3.30 shows the

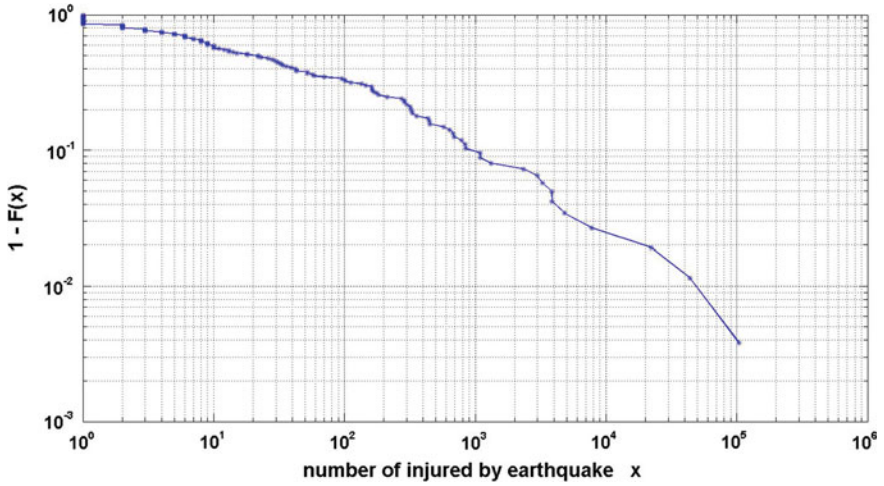


**Fig. 3.28** The injured by earthquakes. Catalog of earthquake victims (Japan), 1900–2012. The tail graph 1–$F(x)$

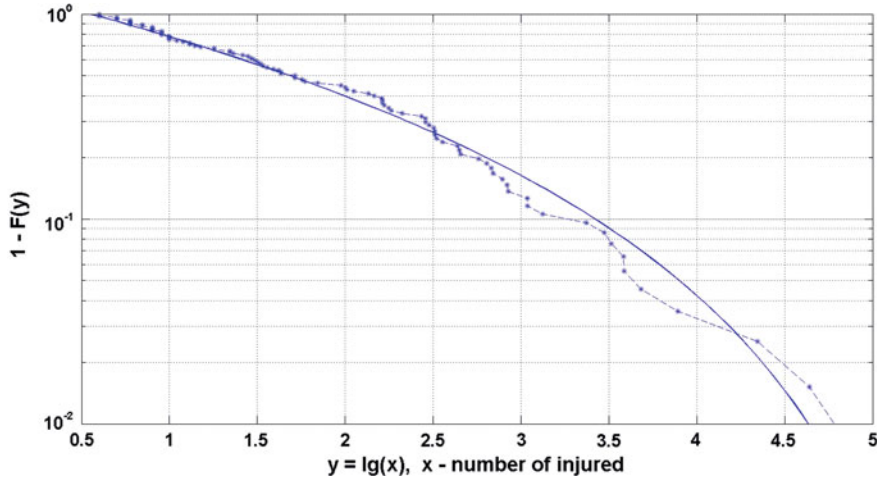


**Fig. 3.29** The injured by earthquakes. Catalog of earthquake victims (Japan), 1900–2012. The extreme tail 1–$F(y)$ ($y = \log(x)$, $x$—number of injured by earthquake) and approximating GPD-tail: $h = 0.55$; $\xi = -0.374$; $s = 1.860$; $n = 99$

contribution of the $p$-fraction of the largest events to the total sum. We see that 10 % of the most disastrous events are responsible for 94 % of the total number of injured. Such sample can be characterized as sample with a strong concentration. Figure 3.29 shows the extreme part of the tail used for parameter estimation along with fitted GPD-curve. We see that the GPD-approximation is on the whole satisfactory in spite of some oscillations.

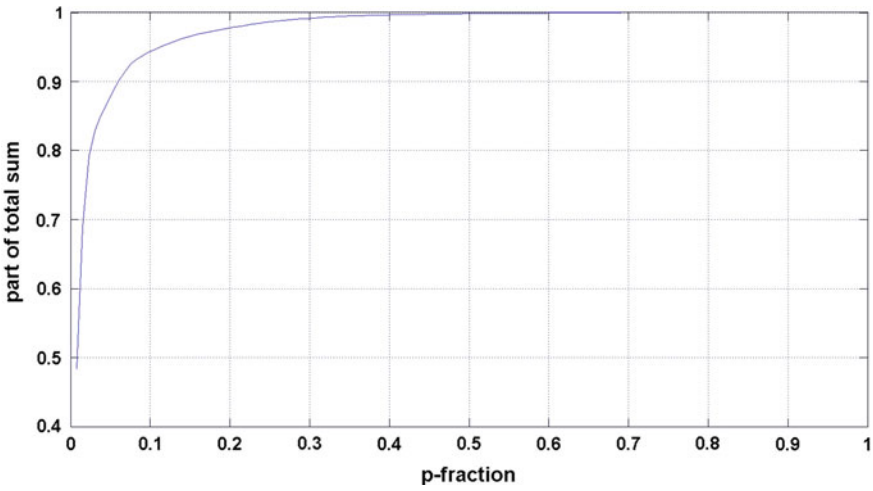


**Fig. 3.30** The injured by earthquakes. Catalog of earthquake victims (Japan), 1900–2012. The contribution of the *p*-fraction of the most deadly events to the total death toll

The Kolmogorov test for a grid of thresholds is shown on Fig. 3.31. The best fit of GPD-distribution under $h = 0.55$ ($\xi = -0.374 \pm 0.063$; $s = 1.86 \pm 0.30$) provides $p$–$v = 0.69$, which supports GPD-distribution.

Now we are able to calculate the quantiles $Q_q(\tau)$. Figure 3.32 shows the GPD-quantiles for 3 different confidence levels $q = 0.90$; 0.95; 0.99. Again we see jumps for very small $\tau$ corresponding to absence of events on time interval $[0; \tau.]$

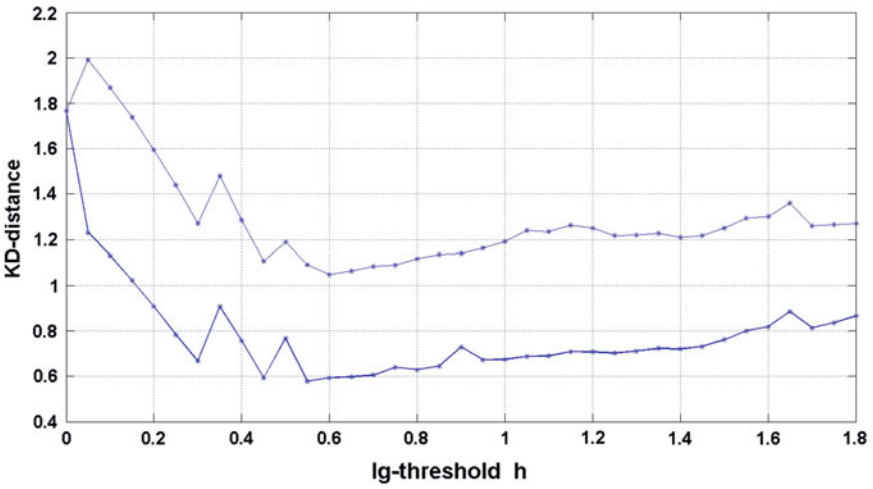


**Fig. 3.31** The injured by earthquakes. Catalog of earthquake victims (Japan), 1900–2012. The KD-distances for a grid of log-thresholds $h_j$. *Thick line*—GPD-fitting; *thin line*—ED-fitting

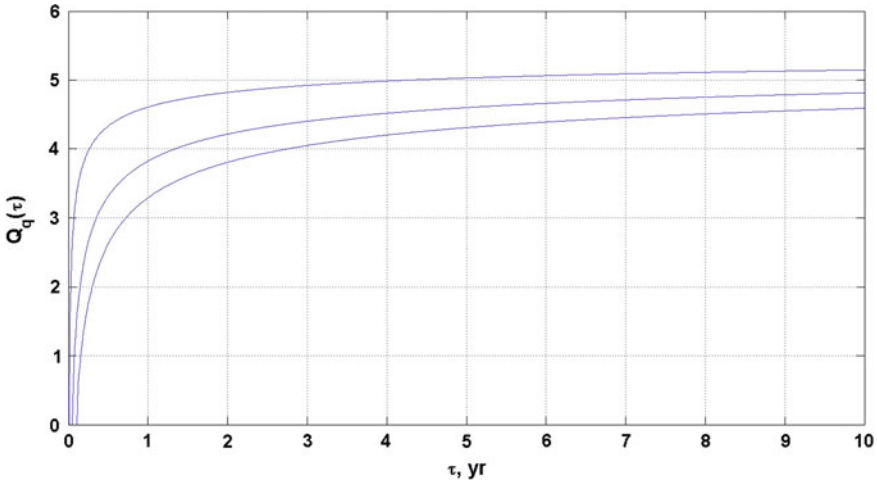


**Fig. 3.32** The injured by earthquakes. Catalog of earthquake victims (Japan), 1900–2012. The GPD-quantiles for three different confidence levels: $q = 0.90$ (*lower curve*); 0.95 (*middle curve*); 0.99 (*upper curve*)

The "maximum possible size" (rightmost limit $M_{max}$ of the GPD) is:

$$M_{max} = h - s/\xi = 5.52 \text{ (this corresponds to 330,000 injured persons)}$$

This is much more than the quantile $Q_{0.95}(10) = 4.81$ (65,000).

## 3.3 Floods (Victims, Overall Economic Losses)

### *3.3.1 Cautions on the Accuracy of the Flood Damage Data*

Flood damage estimates are reported in many different ways, and are subject to a wide variety of errors. Estimates come from federal, state, or county level government officials. Some inaccuracies and mistakes in the data are inevitable, damages are often underreported. Besides, different definitions of term "flood" are used. One of the most critical discrepancies of these data occurs with storm surge related flooding caused by tropical cyclones. Coastal flooding caused by storm surge is not counted in the figures of the flood damage data used below. The record season of 2005, with hurricanes Katrina and Rita, were undoubtedly enormous flooding events. However, the damages associated with hurricane Katrina were largely due to storm surge, and not fresh water flooding (associated to rainfall). Therefore, the annual figure of $51B for water year 2005, although much higher than any other year, does not account for most of the flooding produced by Katrina. On the other hand, the damages from hurricanes that we shall analyze below,

include the hurricane Katrina with $108B damage, which contributed in $141B of annual US hurricane damage, 2005.

In this section we shall use the data of the International Disaster Database (www.emdat.be/). The data consist of three types of flood losses in USA, 1900-2011: fatalities, numbers of affected by flood and estimated economic losses from flood.

**(a)** ***Fatalities***

The catalog contains fatality data of 99 floods. Figure 3.33 shows general view of these data. Looking at Fig. 3.33 we see that the intensity of events sharply increases since 1995. Perhaps, this increase can be explained by more careful registration of victims in later times. On Fig. 3.34 the intensity of events is shown for 1995–2011. We see a stable behavior of the event flow. So, we shall use below data from 1995 onwords.

On Fig. 3.35 we compare the empirical DF of normalized time moments $s_i$. $\hat{F}_n(s)$ with uniform DF. The deviations from diagonal are small. The Kolmogorov distance $KD = 0.633$ corresponds to $p$–value $= 0.82$. Since this $p$-value ($p$–$v$) is considerably more than 0.1, we can accept the hypothesis of stationarity of $t_i$.

Now we consider the fatalities figures. The tail graph is shown on Fig. 3.36. The tail decrease goes rather gradually at the middle range (which is typical of the Pareto distribution), but accelerates at $x > 15$, approaching the exponential tail behavior. Thus, it is not clear, whether the log-transformation of data is appropriate? KD-distances, characterizing goodness of fit are shown on Fig. 3.38 for a grid of thresholds. We see that the original data have been fitted by ED much better ($h = 10^{0.45} = 2.5$; $n = 41$; $KD = 0.836$; $p$–$v = 0.22$) than logarithms ($h = 10^{0.7} = 5$; $n = 32$; $KD = 1.515$; $p$–$v = 0.0004$). Thus, we use the original data for further processing. Figure 3.39 shows the contribution of the $p$-fraction of

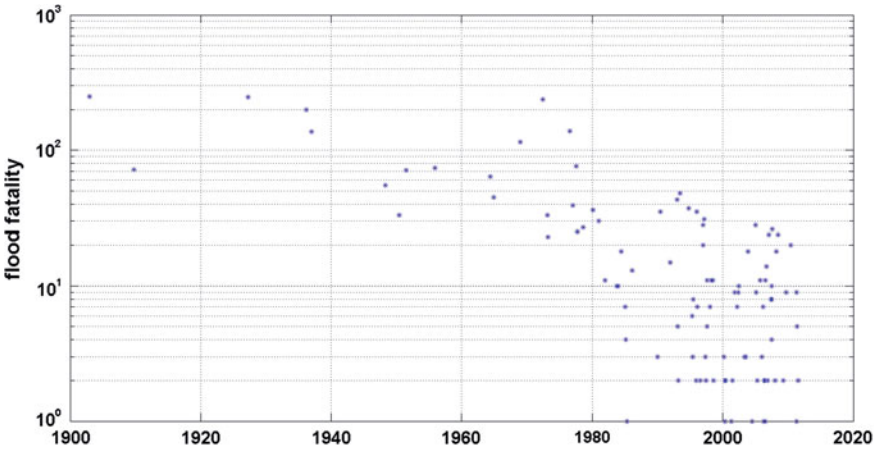


**Fig. 3.33** The fatality events of the catalog of flood victims, USA, 1900–2011

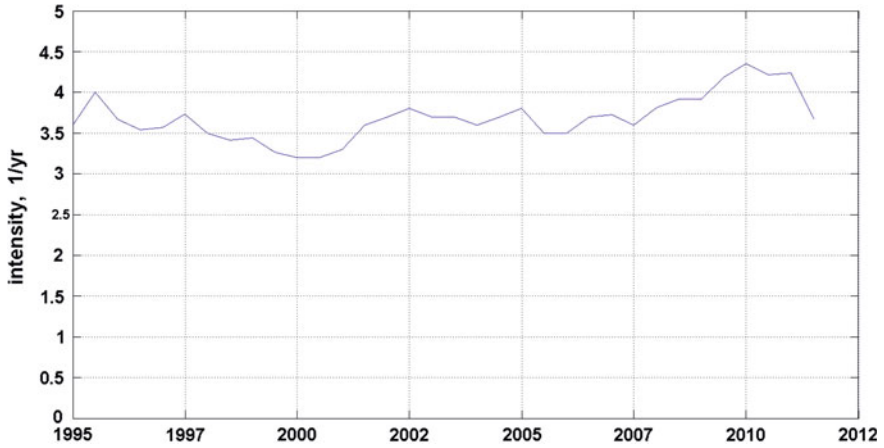


**Fig. 3.34** The intensity of floods with fatalities, USA, 1995–2011, smoothed by 5-year time window

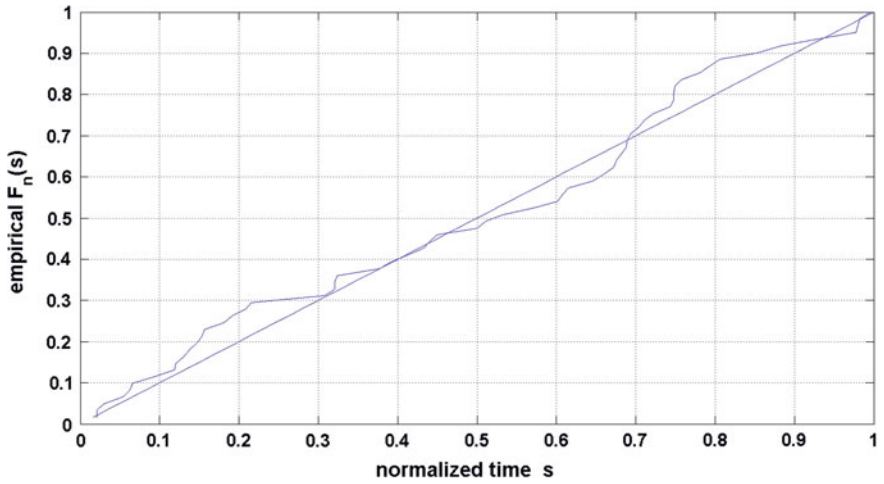


**Fig. 3.35** Flood fatalities, USA, 1995–2011. The empirical DF of normalized event times $s_i$ $\hat{F}_n(s)$

the most deadly events to the whole death toll. We see that 10 % of the most deadly events are responsible for 55 % of the death toll. Such a sample can be characterized as a sample with a weak concentration. Figure 3.37 shows the extreme part of the tail used for parameter estimation along with fitted exponential curve. We see that behavior of the extreme part of sample tail is rather unstable, and an exponential curve gives the best possible approximation in this complicated situation. The $p$-value 0.22 is not too high, but still it gives the ground to accept the ED.

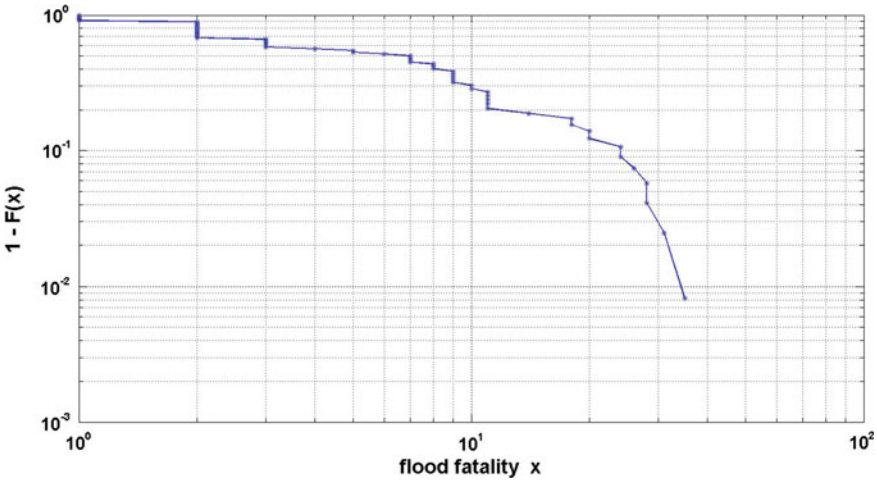


**Fig. 3.36** Flood fatalities, USA, 1995–2011. The tail graph 1–$F(x)$

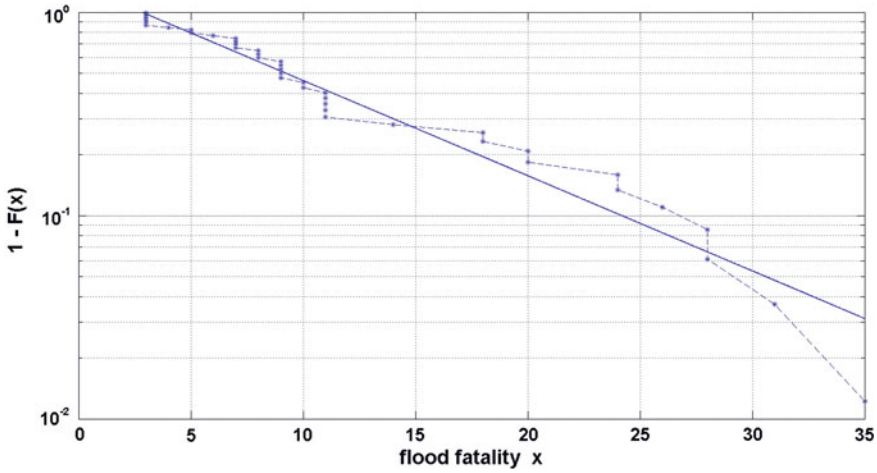


**Fig. 3.37** Flood fatalities, USA, 1995–2011. The extreme tail 1–$F(x)$, and approximating EXP-tail $\exp[-\alpha\cdot(x-h)]$; $h = 0.45$; $\alpha = 0.108$; $n = 41$

On Fig. 3.40 the cumulative sums of fatalities for 1995–2011 ($n = 61$) are shown. In spite of some local fluctuations the trend on the whole looks like a linear function. If there are some deviations they seem to be insignificant. So, we accept the hypothesis of the stationarity of this catalog. Now we are able to calculate the quantiles $Q_q(\tau)$, using ED fitting. Figure 3.41 shows the ED-quantiles for three different confidence levels $q = 0.90; 0.95; 0.99$.

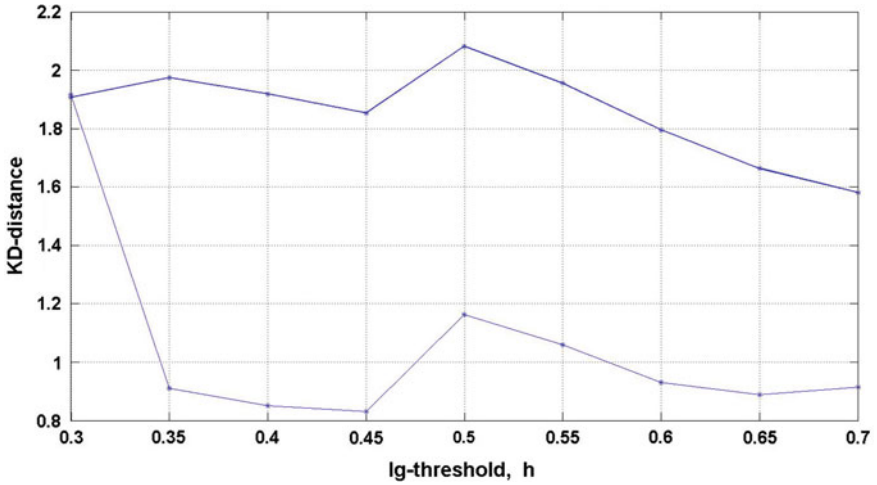


**Fig. 3.38** Flood fatalities, USA, 1995–2011. The KD-distances for a grid of log-thresholds $h_j$. *Thick line*—GPD-fitting; *thin line*—ED-fitting

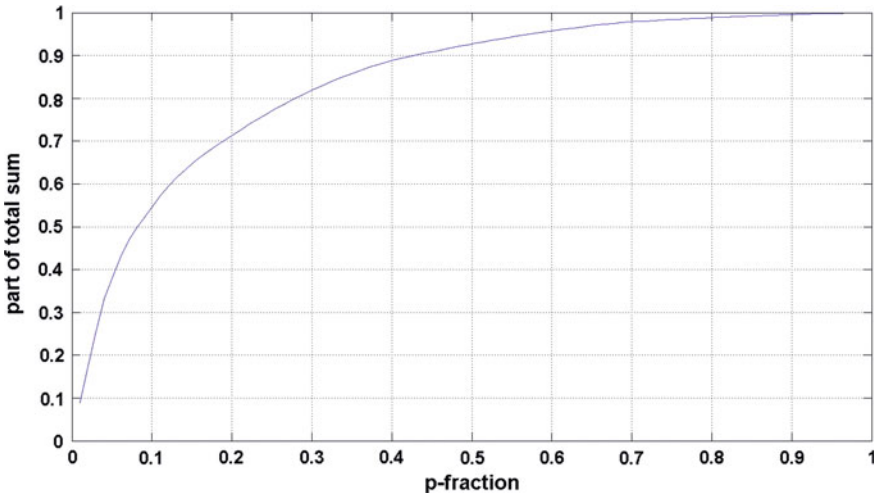


**Fig. 3.39** Flood fatalities, USA, 1995–2011. The contribution of the $p$-fraction of the most deadly events to the whole death toll

**(b)** ***Affected by floods***

Affected people, as defined in EM-DAT, are people who require immediate assistance during a period of emergency, including displaced or evacuated people. The catalog contains numbers of affected in 95 floods, 1970–2011. Figure 3.42 shows these numbers in time. The maximum number (11,000,148) occurred at

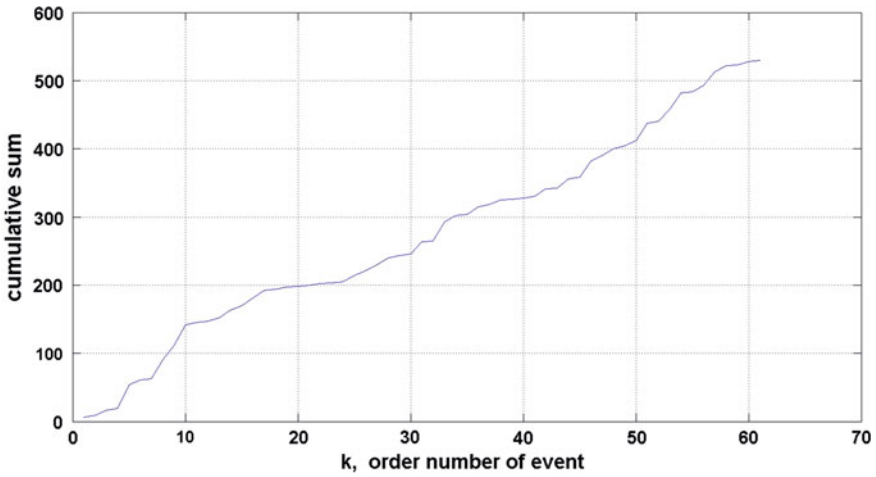


**Fig. 3.40** Flood fatalities, USA, 1995–2011. The cumulative sums of fatalities for 1995–2011 (n = 61) are shown

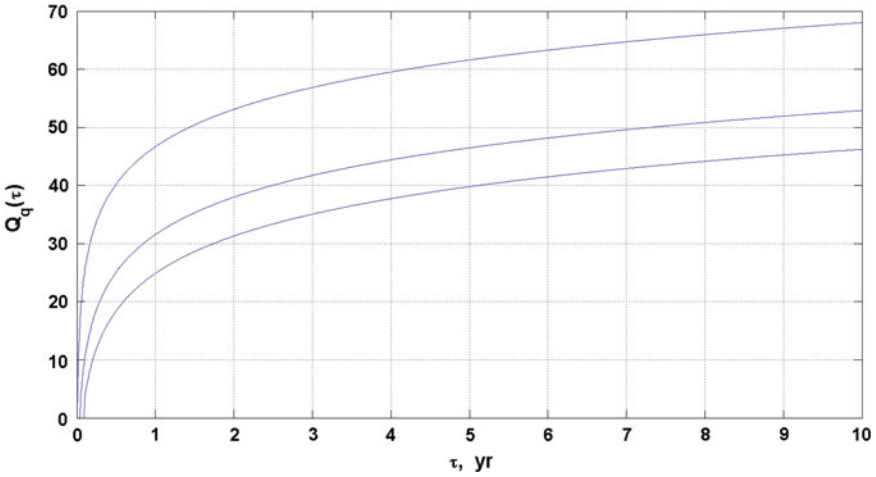


**Fig. 3.41** Flood fatalities, USA, 1995–2011. ED-quantiles for three different confidence levels: $q = 0.90$ (*lower curve*); 0.95 (*middle curve*); 0.99 (*upper curve*)

09.06.2008. Looking at Fig. 3.42 we see that the intensity of events sharply increases at 1995, which coincides with mentioned above behavior of flood fatalities. On Fig. 3.43 the intensity of events is shown for 1995–2011. We see a stable behavior of the event flow. So, we have used the data from 1995 onwards.

On Fig. 3.44 we compare the empirical DF of normalized time moments $s_i$. $\hat{F}_n(s)$ with uniform DF. The deviations from diagonal are relatively small. The

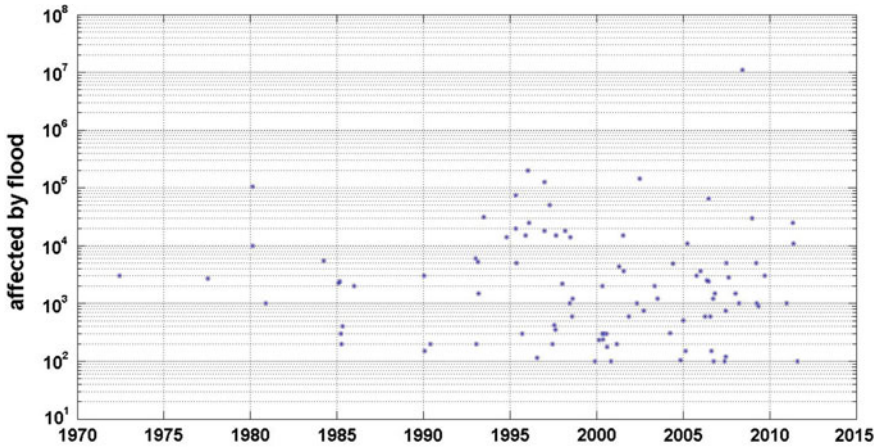


**Fig. 3.42** Affected in floods, USA, 1995–2011. Time–event size diagram

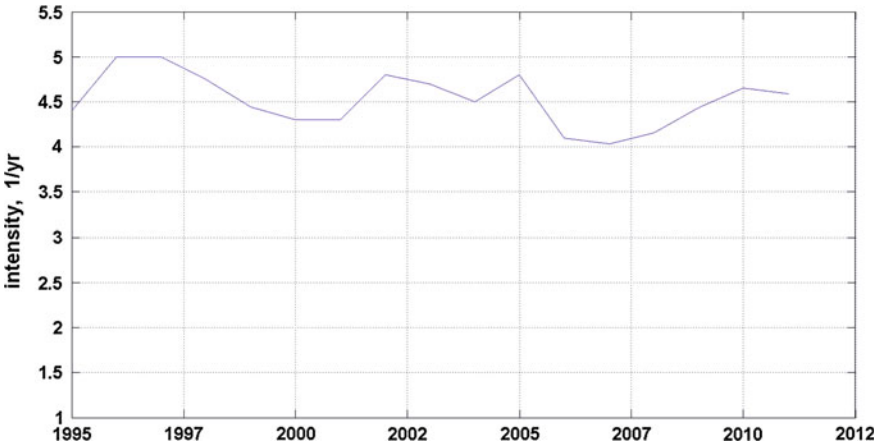


**Fig. 3.43** Affected in floods, USA, 1995–2011. The intensity of events for 1995–2011, smoothed by 10-year time window

Kolmogorov distance $D_n = 0.692$ corresponds to $p$-value = 0.72. Since this $p$-value ($p$–$v$) is more than 0.1, we can accept the hypothesis of stationarity of $t_i$.

Now we consider the figures of affected by floods. The tail graph is shown on Fig. 3.45. We see that on the whole the tail 1 – F($x$) has power-like behavior with one outlier (maximum event May 09, 2008 with 11,000,148 affected). So, it is reasonable to take logarithms and to analyze log($x$). Figure 3.47 shows the contribution of the $p$-fraction of the largest events to the total sum. We see that 10 % of the most disastrous events are responsible for 97.7 % of the total number of affected. This is a tail with "strong concentration".

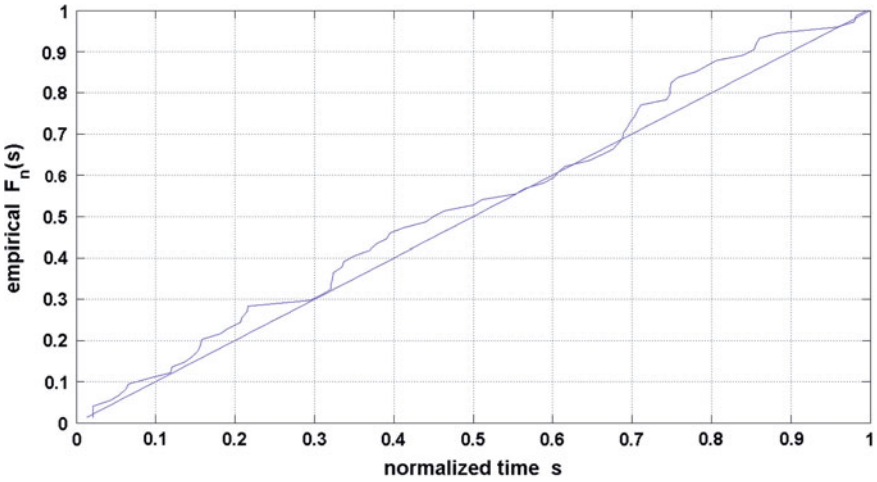


**Fig. 3.44** Affected in floods, USA, 1995–2011. The empirical DF $\hat{F}_n(s)$ of normalized event times $s_i$

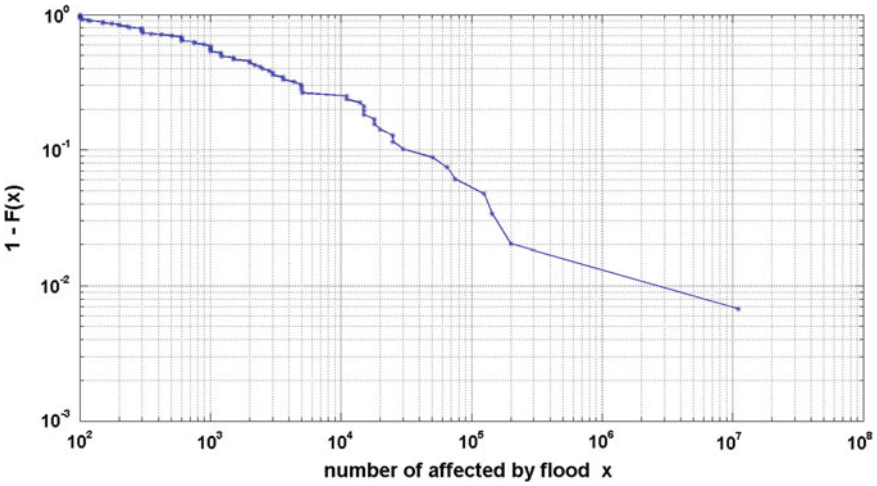


**Fig. 3.45** Affected in floods, USA, 1995–2011. The tail graph 1 – ($Fx$)

The Kolmogorov test for a grid of thresholds is shown on Fig. 3.48. We see that minimum KD-distance 0.567 is reached at $h = 2.7$ by GPD. Using the simulation method we got $p–v = 0.665$, which allows us to accept GPD-distribution ($h = 2.7$; $\xi = -0.182 \pm 0.113$; $s = 1.205 \pm 0.302$). Figure 3.46 shows the extreme part of the tail used for parameter estimation along with fitted GPD-curve. We see that the approximation is more or less satisfactory.

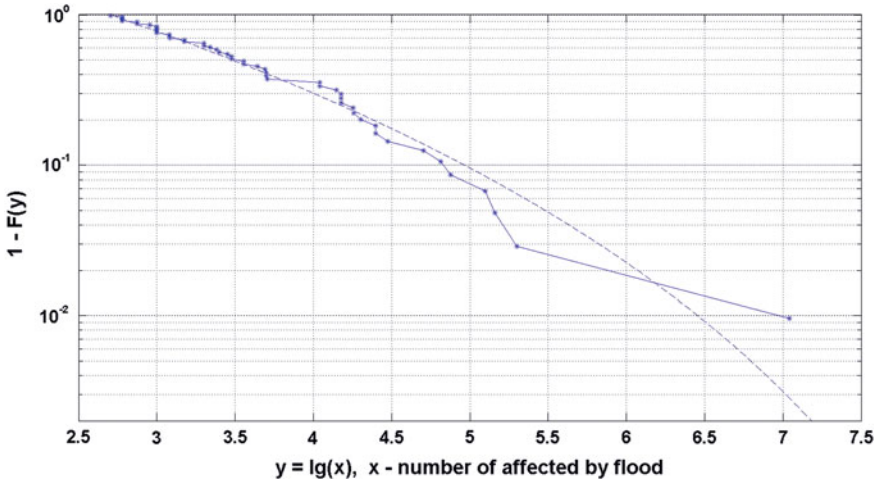


**Fig. 3.46** Affected in floods, USA, 1995–2011. The extreme tail $1 - F(y)$ ($y = \log(x)$, $x$—number of affected in flood) and approximating GPD-tail: $h = 2.7$; $\xi = -0.182$; $s = 1.205$; $n = 52$

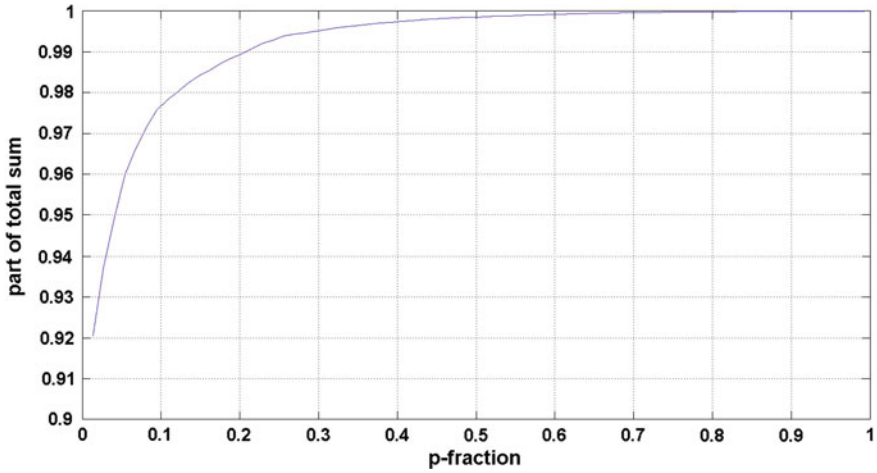


**Fig. 3.47** Affected in floods USA, 1995–2011. The contribution of the $p$-fraction of the most disastrous events to the total sum of affected

Now we are able to calculate the quantiles $Q_q(\tau)$. We use for these quantiles GPD-distribution as providing the best fitting. Figure 3.49 shows GPD-quantiles for three different confidence levels $q = 0.90; 0.95; 0.99$.

The "maximum possible size" (rightmost limit $M_{max}$ of the GPD) is:

$$M_{max} = h - s/\xi = 9.31(2 \cdot 10^9 \text{peoples}).$$

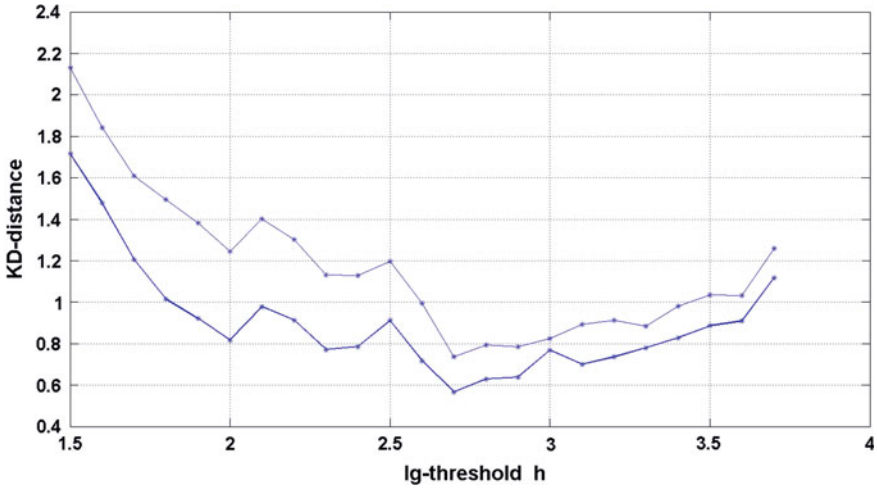


**Fig. 3.48** Affected in floods USA, 1995–2011. The KD for a grid of log-thresholds. *Thick line*—GPD-fitting; *thin line*—ED-fitting

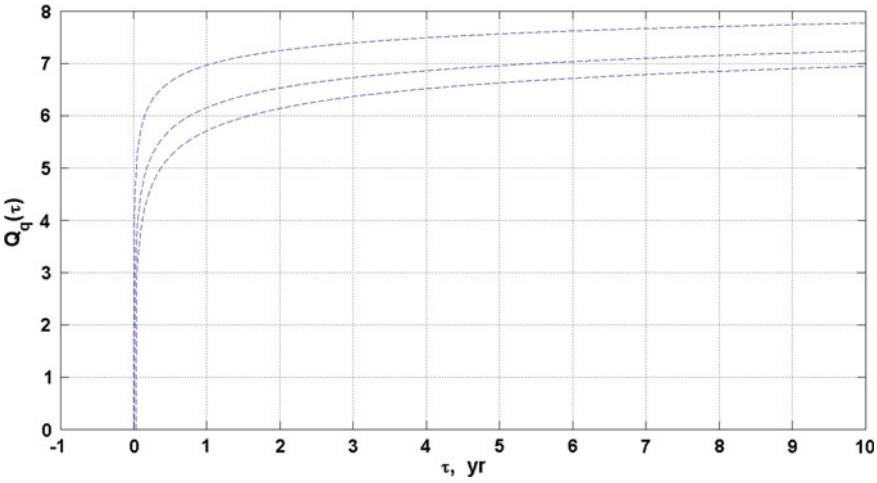


**Fig. 3.49** Affected in floods USA, 1995–2011. GPD-quantiles for three different confidence levels: $q = 0.90$ (*lower curve*); 0.95 (*middle curve*); 0.99 (*upper curve*)

Again, we can say that such gigantic figure hardly has a practical value. In the same time the quantile $Q_{0.95}(10) = 7.25$ ($17.7 \cdot 10^6$ peoples) looks quite realistic.

**(c)** ***Estimated damages caused by floods in USA***

The catalog contains estimated economic loss data of 78 floods in USA, 1900–2011 (indexed to 2011) in millions of USA $. Figure 3.50 shows general

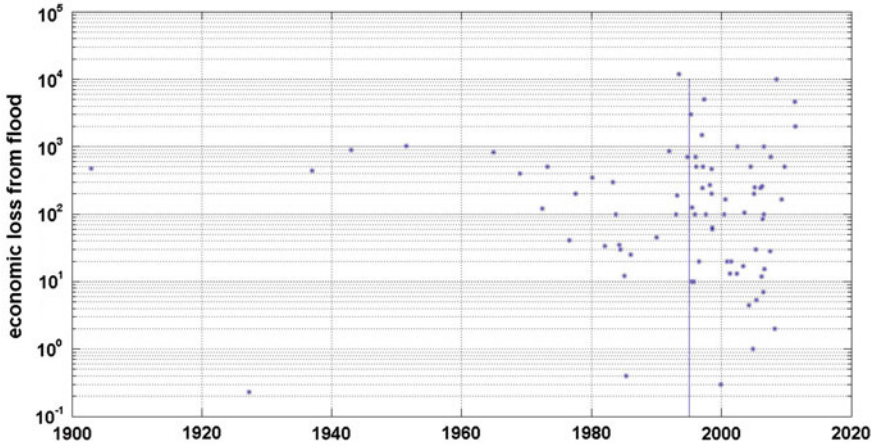


**Fig. 3.50** Estimated economic losses from floods, USA, 1995–2011, in $10^6$ $ (adjusted to 2011). Time–event size diagram

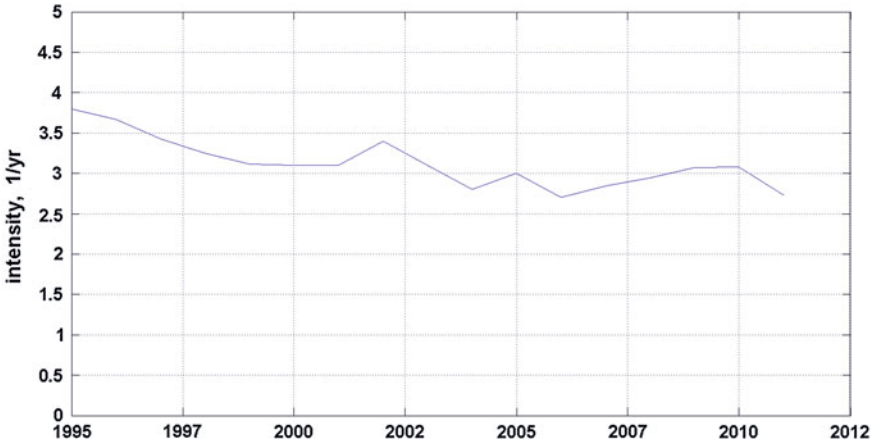


**Fig. 3.51** Estimated economic losses from floods, USA, 1995–2011, in $10^6$ $. The intensity of events, smoothed by 10-year time window

view of these data. Looking at Fig. 3.50 we see that the intensity of events sharply increases since 1995 (shown by vertical line on the Figure), which is consistent with behavior of data on fatalities and affected. Again, we shall analyze data only from 1995 onwards.

On Fig. 3.51 the intensity of events is shown for 1995–2011. We see a stable behavior of the event flow with a weak decrease to the end of interval.

On Fig. 3.52 we compare the empirical DF of normalized time moments $s_i$. $\hat{F}_n(s)$ with uniform DF. The Kolmogorov distance $D_n = 1.062$ corresponds to

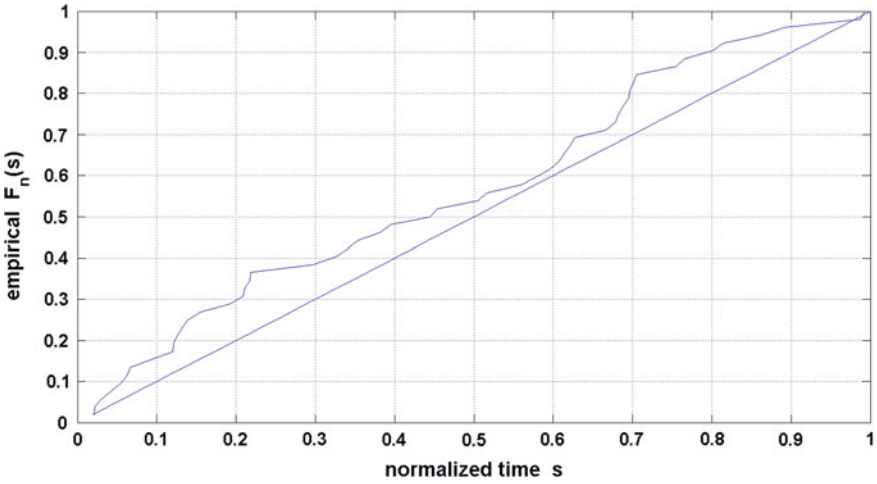


**Fig. 3.52** Estimated economic losses from floods, USA, 1995–2011, in $10^6$ \$. The empirical DF $\hat{F}_n(s)$ of normalized time moments $s$

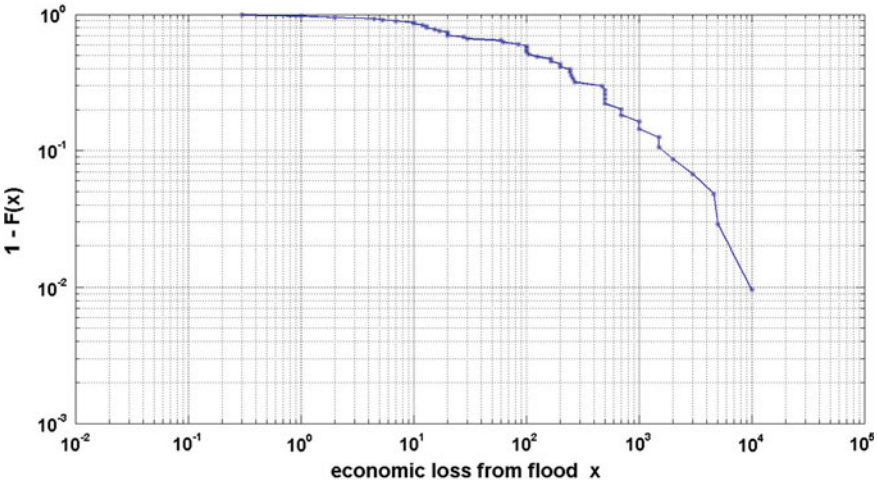


**Fig. 3.53** Estimated economic losses from floods, USA, 1995–2011, in $10^6$ \$. The tail graph $1 - F(x)$

$p$-value $= 0.21$. Since this $p$-value is more than 0.1, we can accept the hypothesis of stationarity of $t_i$.

Now we consider the loss figures. The tail graph is shown on Fig. 3.53. The tail decrease goes rather gradually which is typical for the Pareto distribution. So, we take logarithms for further analysis. KD-distances, characterizing goodness of fit

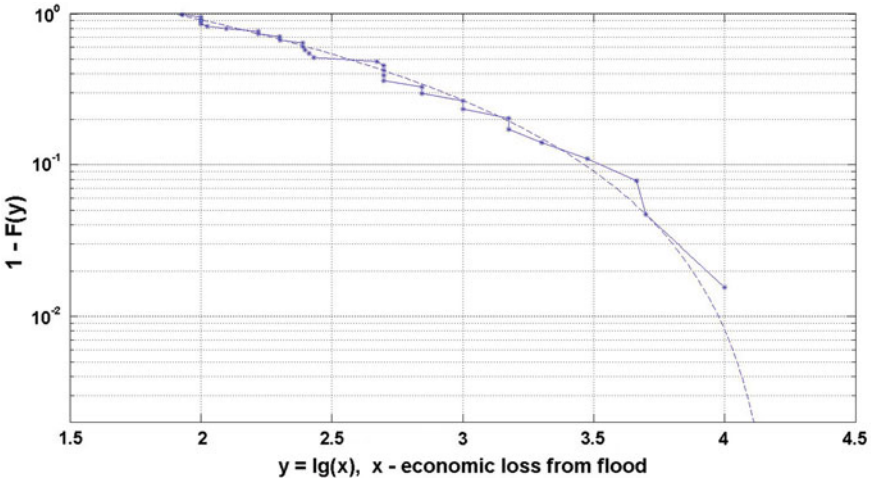


**Fig. 3.54** Estimated economic losses from floods, USA, 1995–2011, in $10^6$ \$. The extreme tail $1 - F(y)$ ($y = \log(x)$, $x$—loss in $10^6$ USD) and approximating GPD-tail: $h = 1.9$; $\xi = -0.486$; $s = 1.129$; $n = 32$

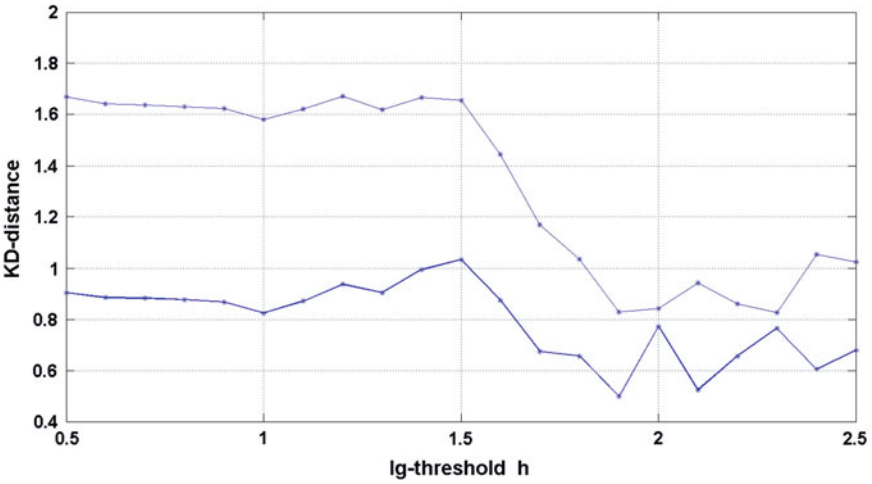


**Fig. 3.55** Estimated economic losses from floods, USA, 1995–2011, in $10^6$ \$. KD-distances for a grid of log-thresholds. *Thick line*—GPD-fitting; *thin line*—ED-fitting

are shown on Fig. 3.55 for a grid of thresholds. The threshold $h = 1.9$ provides the best fitting for GPD ($KD = 0.501$; $\xi = -0.486 \pm 0.091$; $s = 1.129 \pm 0.286$).

Figure 3.56, shows the contribution of the $p$-fraction of the most costly events to the whole sum of losses. We see that 10 % of the most deadly events are responsible for 68 % of the total loss. Such sample can be characterized as sample

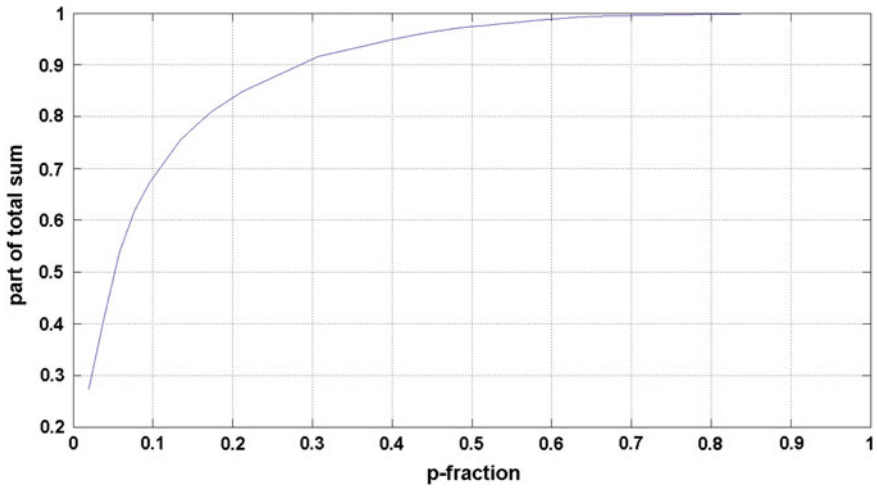


**Fig. 3.56** Estimated economic losses from floods, USA, 1995–2011, in $10^6$ \$. The contribution of the $p$-fraction of the most costly events to the total sum of all losses

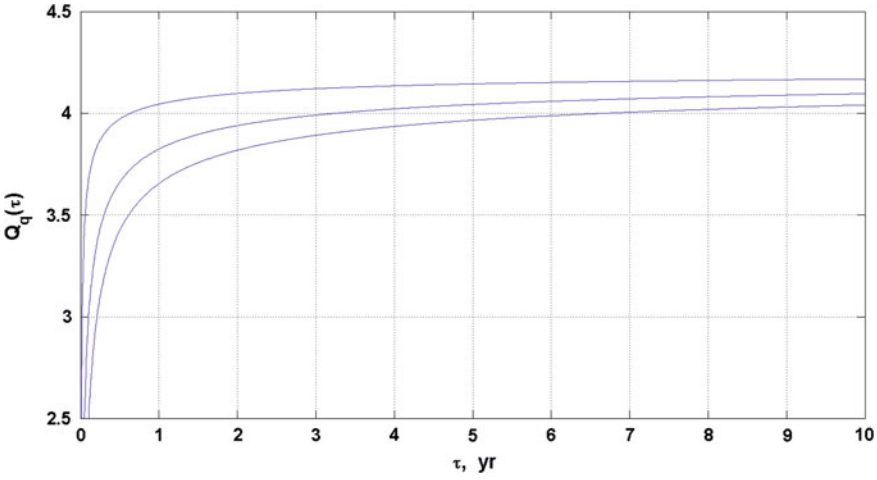


**Fig. 3.57** Estimated economic losses from floods, USA, 1995–2011, in $10^6$ \$. GPD-quantiles for three different confidence levels: $q = 0.90$ (*lower curve*); 0.95 (*middle curve*); 0.99 (*upper curve*)

with a "moderate concentration". Figure 3.54 shows the extreme part of the tail used for parameter estimation along with fitted GPD-curve. We see that the approximation is quite satisfactory.

Now we calculate the quantiles $Q_q(\tau)$, using GPD fitting. Figure 3.57 shows the GPD-quantiles for three different confidence levels $q = 0.90$; 0.95; 0.99.

## 3.4 Tornadoes (Fatalities)

In this section we shall use the data of tornado fatalities displayed in the Internet (http://en.wikipedia.org/wiki/List_of_North_American_tornadoes_and_tornado_outbreaks#1900.E2.80.931919 ). The data consist of fatality numbers of 249 tornadoes, USA, 1900–2012. Figure 3.58 shows general view of these data. Looking at Fig. 3.58 we see that the intensity of events depends on their values. On Fig. 3.59 the intensity of events is shown for two thresholds: $h_1 = 1$ ($n_1 = 249$) and $h_2 = 10$ ($n_2 = 143$). We see that the threshold $h_2$ provides event flow with rather stable intensity whereas there are $n = 143$ events exceeding this threshold.

On Fig. 3.60 we compare the empirical distribution functions $F_{n_1}^{(1)}(x)$, $F_{n_2}^{(2)}(x)$ of normalized time moments corresponding to these two thresholds with uniform DF. The deviations of $F_{n_1}^{(1)}(x)$ from diagonal are large (corresponding $KD = 3.69$; $p–v = 3 \cdot 10^{-12}$), whereas the deviations of $F_{n_2}^{(2)}(x)$ are much smaller (corresponding $KD = 1.08$; $p–v = 0.20$). Since the last $p–v$ is more than 0.1 we can accept the hypothesis of stationarity of time moments $t_i$ for events exceeding $h_2 = 10$.

Now we check stationarity of the fatalities $x_j$ exceeding $h_2 = 10$. Figure 3.61 shows the cumulative sums of $\log_{10}(x_j)$. Somewhere near $t_j = 60$ (corresponding to May 09, 1953) we can distinguish a small but clear decrease of slope. In order to clarify the situation we plotted two sample DF: $G_{m_1}^{(1)}(x)$ and $G_{m2}^{(2)}(x)$, relating to $t \leq$ May 09, 1953, ($m_1 = 60$) and to $t >$ May 09, 1953 ($m_2 = 83$) correspondingly, see Fig. 3.62. We see that distribution functions differ quite definitely, in particular in the middle range. The Kolmogorov–Smirnov distance (KSD)

$$KSD = \sqrt{\frac{m_1 m_2}{m_1 + m_2}} \max \left| G_{m_1}^{(1)}(x) - G_{m_2}^{(2)}(x) \right|$$

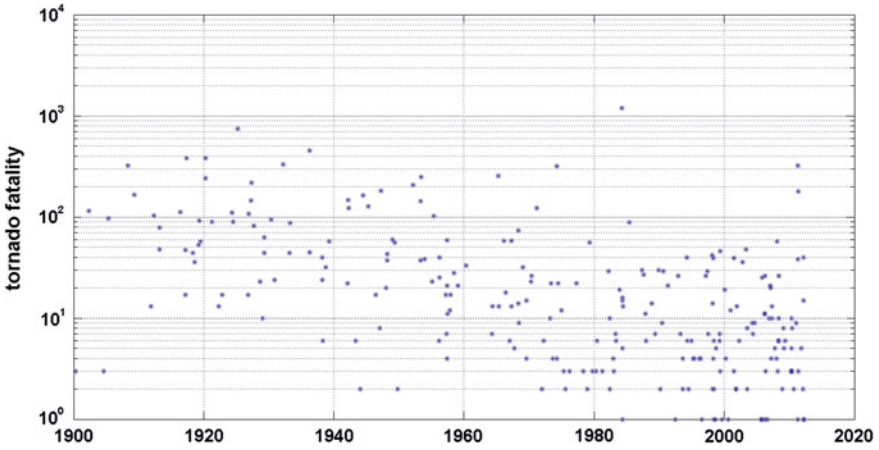


**Fig. 3.58** Tornado fatalities, USA, 1900–2012. Time–event size diagram

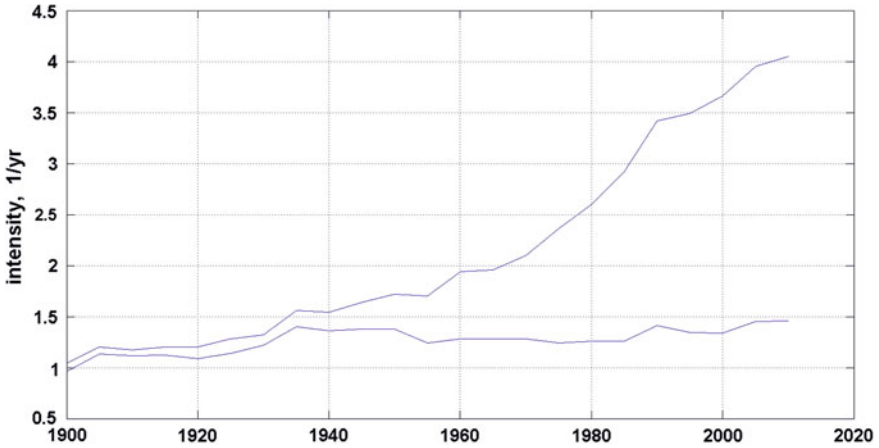


**Fig. 3.59** Tornado fatalities, USA, 1900–2012. The intensities smoothed by 40-year time window for two lower thresholds: $h_1 = 1$ ($n = 249$), *upper curve;* $h_2 = 10$ ($n = 143$), *lower curve*

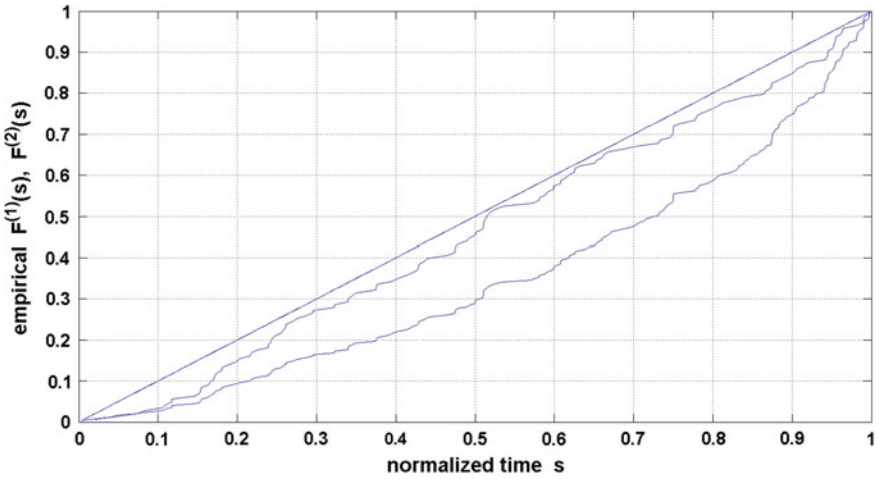


**Fig. 3.60** Tornado fatalities, USA, 1900–2012. Two empirical distribution functions $F_{n_1}^{(1)}(x)$, $F_{n_2}^{(2)}(x)$ of normalized time moments corresponding to thresholds $h_1 = 1$ ($n = 249$), *lower curve;* $h_2 = 10$ ($n = 143$), *upper curve*

equals to 3.02 which corresponds to $p$–$v = 1 \cdot 10^{-8}$. So, we have to use for the further analysis only 83 data observed after May 09, 1953.

The tail graph is shown on Fig. 3.63. The tail is typical of the Pareto distribution. So, we use logarithmically transformed data. KD-distances, characterizing goodness of fit are shown on Fig. 3.65 for a grid of thresholds. We see that the original data

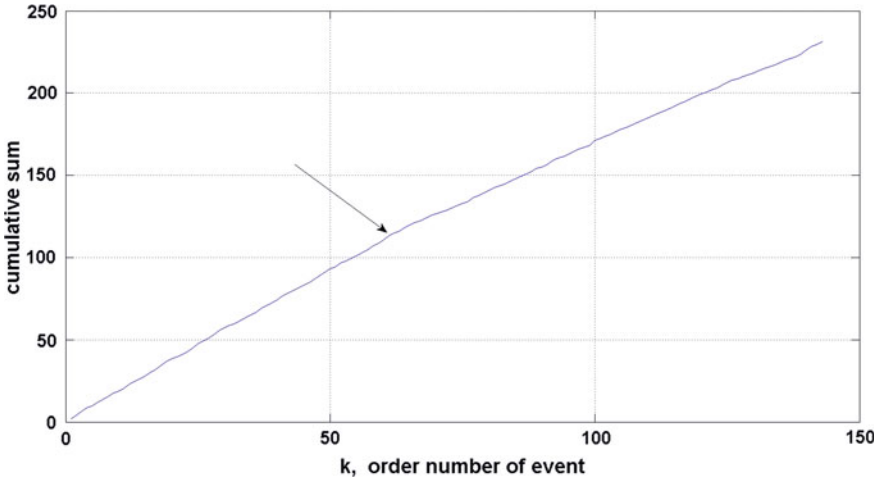


**Fig. 3.61** Tornado fatalities, USA, 1900–2012. The cumulative sums of $\log_{10}(x_k)$, $x_k$—number of dead. The arrow indicates slope break (at time 09.05.1953)

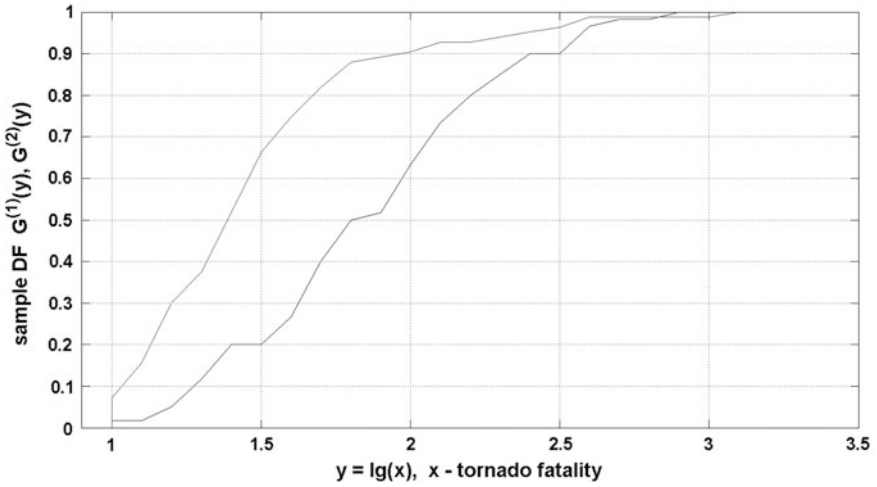


**Fig. 3.62** Tornado fatalities, USA, 1900–2012. Two sample DF: $G^{(1)}_{m_1}(x)$ and $G^{(2)}_{m2}(x)$, relating to $t \leq$ May 09, 1953, ($m_1 = 60$), *right curve*, and to $t >$ May 09, 1953 ($m_2 = 83$), *left curve*

have been fitted by both distributions (ED and GPD) almost identically for thresholds $h = 1.3 - 1.35$. The form parameter of GPD for these thresholds was found to be practically zero ($\sim 10^{-8}$). Thus, GPD and ED in this case are practically identical. The best fit corresponds to the threshold $h = 1.3$; $n = 53$. The parameter of exponential distribution $\alpha = 2.75 \pm 0.38$. As we remarked above, the exponential distribution for $\log(x)$ means the Pareto distribution for original $x$ with parameter

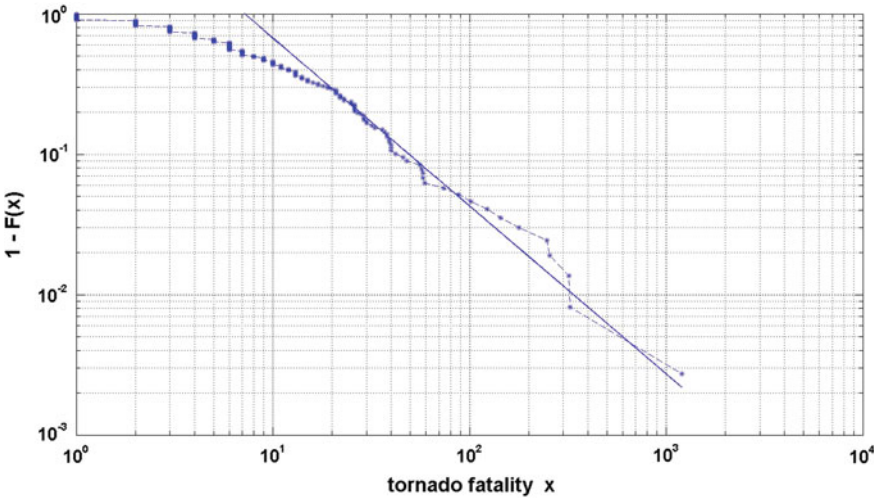


**Fig. 3.63** Tornado fatalities, USA, 1953–2012, threshold $h \geq 10$, $n = 83$. The tail function $1 - F(x)$. Approximating line is the same as on Fig. 3.64

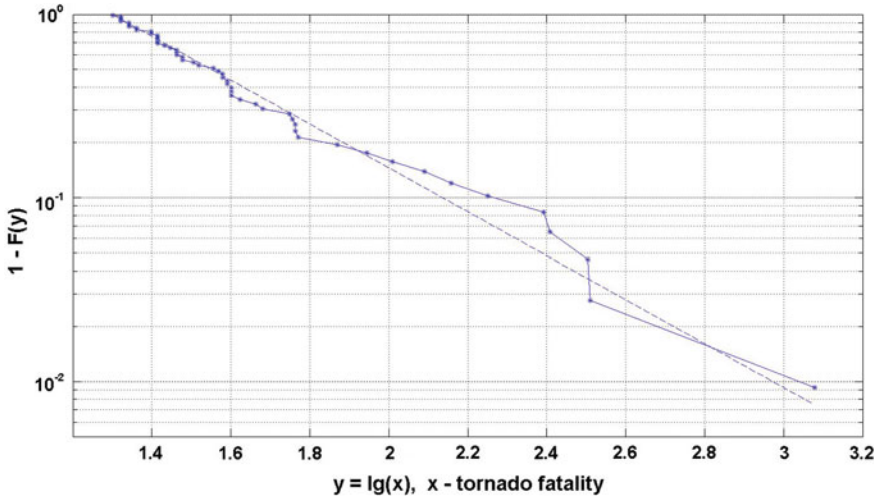


**Fig. 3.64** Tornado fatalities, USA, 1953–2012. The extreme tail $1 - F(y)$ ($y = \log(x)$, $x$—number of dead) and approximating EXP-tail $exp[-\alpha \cdot (y - h)]$; $h = 1.3$; $\alpha = 2.754$; $n = 53$

$\beta = \alpha/\log(10) = 1.19 \pm 0.17$. We see that the tail of tornado victims is close to a heavy one (expectation is finite, but variance is infinite). Taking into account that the confidence interval for parameter $\beta$

$$(1.19 - 0.17;\ 1.19 + 0.17)$$

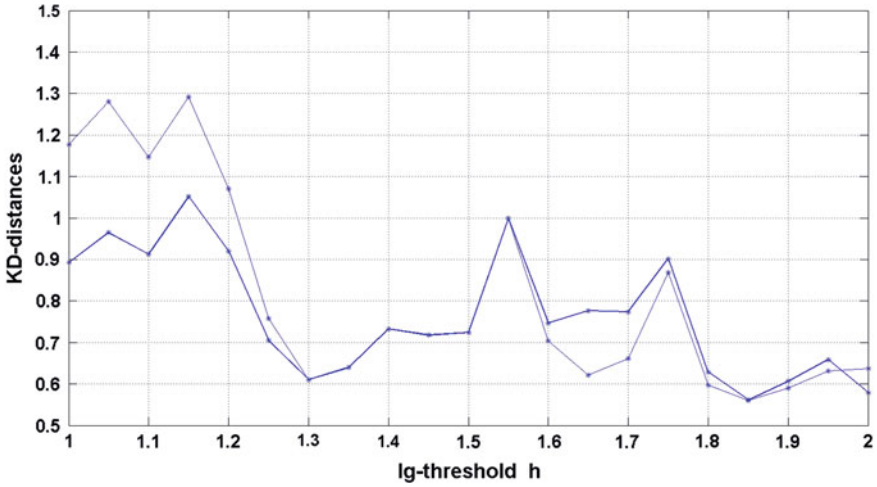


**Fig. 3.65** Tornado fatalities, USA, 1953–2012. KD-distances for a grid of log-thresholds. *Thick line*—GPD-fitting; *thin line*—ED-fitting

contains the true parameter value only with probability 84 % (it is valid for the Gauss distribution), one cannot exclude possibility that the true $\beta < 1$, i.e. the tail is really heavy.

Figure 3.66 shows the contribution of the $p$-fraction of the most deadly events to the whole death toll. We see that 10 % of the most deadly events are responsible

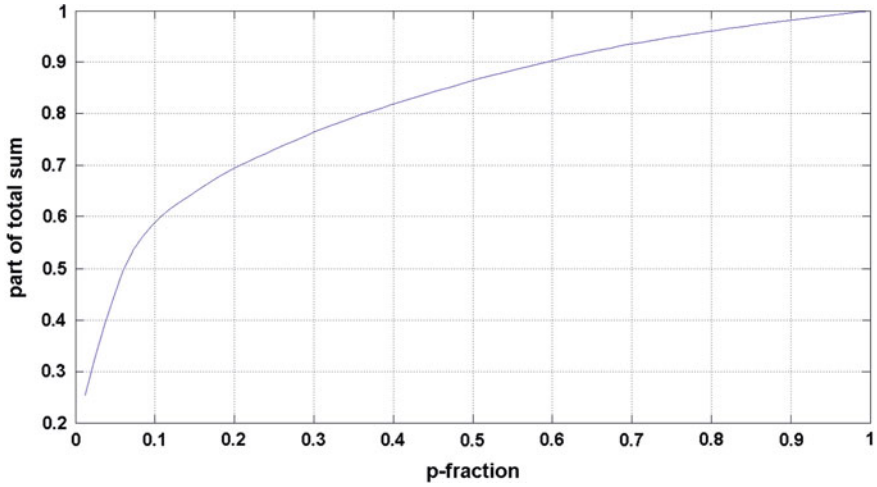


**Fig. 3.66** Tornado fatalities, USA, 1953–2012, threshold $h \geq 10$. The contribution of the $p$-fraction of the most deadly events to the total death toll

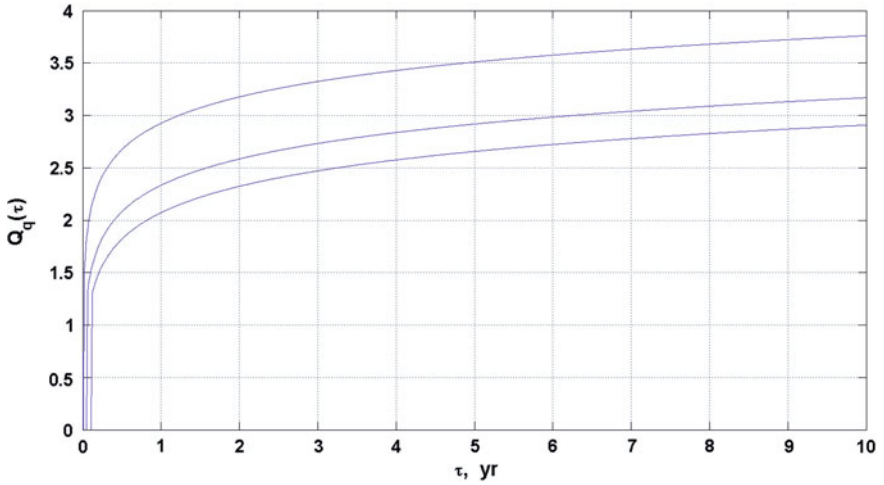


**Fig. 3.67** Tornado fatalities, USA, 1953–2012, threshold $h \geq 10$. The ED-quantiles for three different confidence levels: $q = 0.90$ (*lower curve*); 0.95 (*middle curve*); 0.99 (*upper curve*)

for 60 % of the death toll. Such sample can be characterized as sample with a "weak concentration". Figure 3.64 shows the extreme part of the tail used for parameter estimation along with fitted ED-curve. We see that the approximation is more or less satisfactory.

Now we are able to calculate the quantiles $Q_q(\tau)$. Figure 3.67 shows the ED-quantiles for 3 different confidence levels $q = 0.90; 0.95; 0.99$.

## 3.5 Annual Economic Losses from Floods, USA

In this section we analyze the annual flood damage data compiled by the US National Weather Service (www.flooddamagedata.org). We take the time period 1903–2011. The time series of flood damages in log-scale is shown on Fig. 3.68. The damage is measured in $\$10^9$ (adjusted to 2011). We see that behavior of the time series noticeably changes its character somewhere near 1940. This conclusion is supported by the graph of cumulative sums of log-damage shown on Fig. 3.69. A more or less stable trend is established only after 1940. Thus, we took for further analysis the time period 1940–2011. The sample tail $1 - F(x)$ of these data is shown on Fig. 3.70. We see rather moderate power-like decreasing (straight line at the extreme range). Figure 3.72 shows the contribution of the $p$-fraction of the most costly years to the total damage. We see that 10 % of the most costly years are responsible only for 40 % of the total damage. Such sample can be characterized as sample with a "weak concentration". Figure 3.71 shows the extreme part of the tail used for parameter estimation along with fitted GPD-curve. We see that the approximation is quite satisfactory.

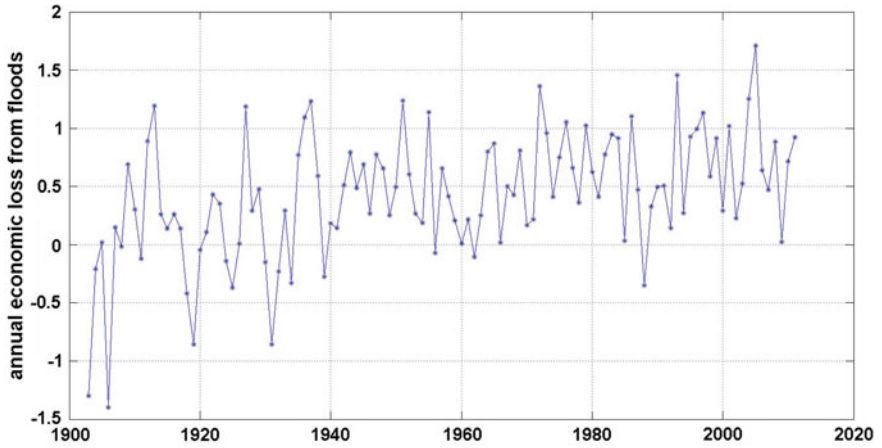


**Fig. 3.68** The time series of annual flood damages in log-scale, USA, 1903–2011. The damage is in $\$10^9$ (adjusted to 2011)

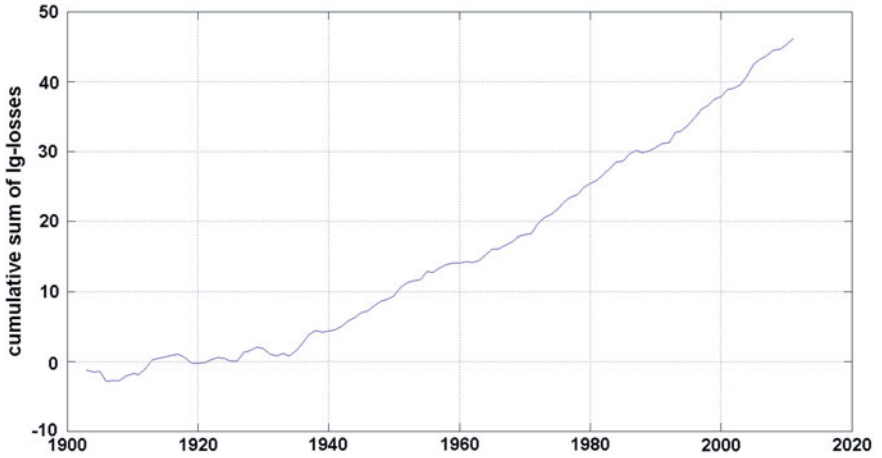


**Fig. 3.69** Annual flood damages in the USA 1903–2011. The cumulative sums of log-damages

Now we apply GPD-fitting to peaks over thresholds $h$. The resulting KD-distances characterizing goodness-of-fit are shown on Fig. 3.73. The best GPD-fitting occurs at $h = 0.4$ ($n = 48$; $\xi = -0.354 \pm 0.093$; $s = 0.541 \pm 0.126$; $p$–$v = 0.90$). The $Q$-quantiles in log-scale given by Eq. (2.44) are shown on Fig. 3.74.

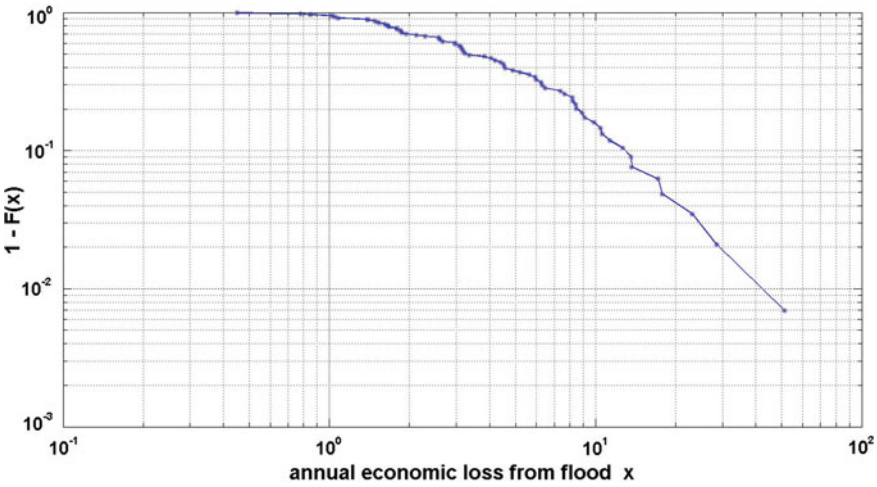


**Fig. 3.70** Annual flood damages in the USA 1940–2011. The sample tail $1 - F(x)$

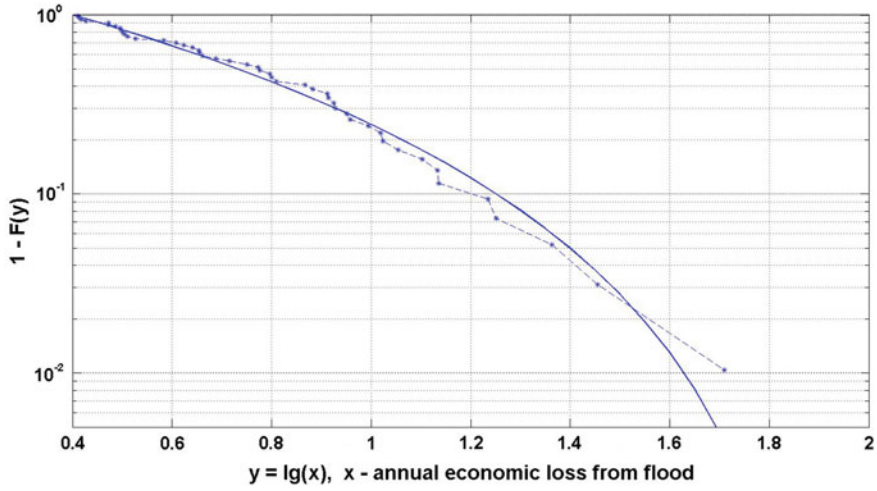


**Fig. 3.71** Annual flood damages in the USA 1940–2011. The extreme tail $1 - F(x)$ and approximating GPD-tail: $h = 0.4$; $\xi = -0.354$; $s = 0.541$; $n = 48$

## 3.6 Annual Economic Losses from Hurricanes, USA

In this section we analyze the annual hurricane damage data published in Blake and Gibney (2011). We take the time period 1940–2010. The time series of hurricane damages in log-scale is shown on Fig. 3.75. The damage is measured in

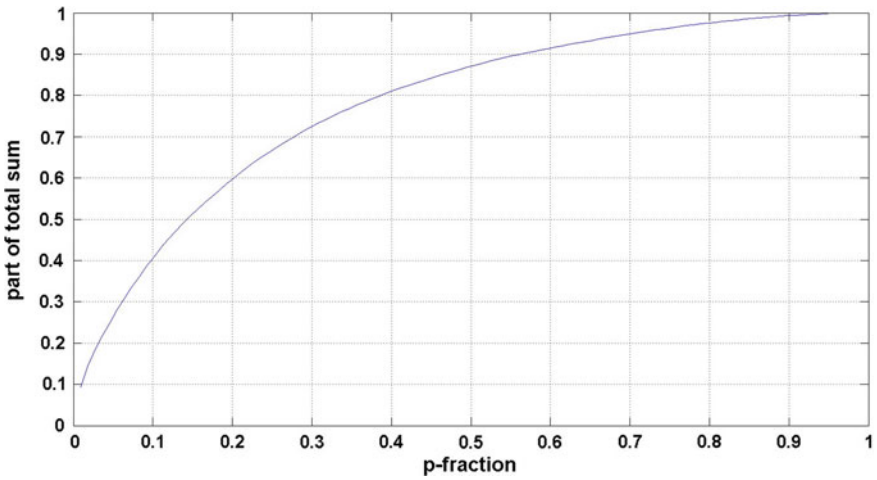


**Fig. 3.72** Annual flood damages in the USA 1940–2011. The contribution of the *p*-fraction of the costliest events to the total sum

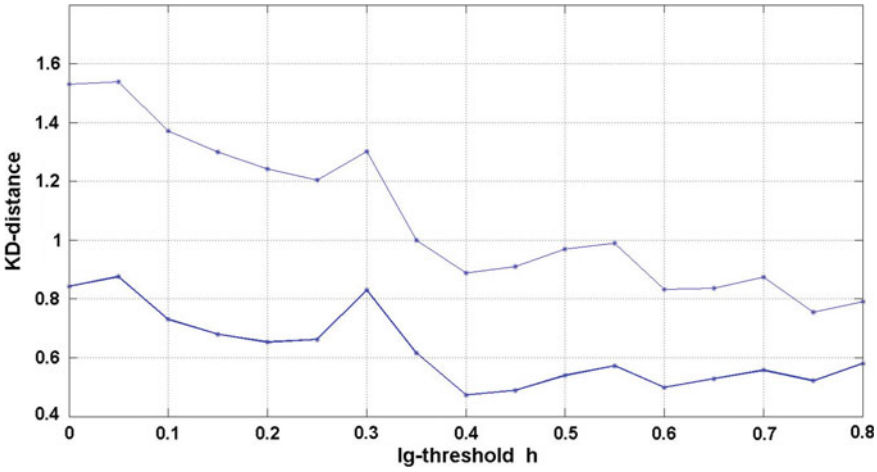


**Fig. 3.73** Annual flood damages in the USA 1940–2011. KD-distances for a grid of log-thresholds. *Thick line*—GPD-fitting; *thin line*—ED-fitting

$\$10^6$ (adjusted to 2010). We see that the time series does not exhibit non-stationarity. This conclusion is supported by the graph of cumulative sums of log-damage shown on Fig. 3.76. A more or less stable trend supports assumption of stationarity. The sample tail $1 - F(x)$ of these data is shown on Fig. 3.77. We see irregular decreasing with some fluctuations. Figure 3.79 shows the contribution of the *p*-fraction of the most costly years to the total damage. We see that 10 % of the

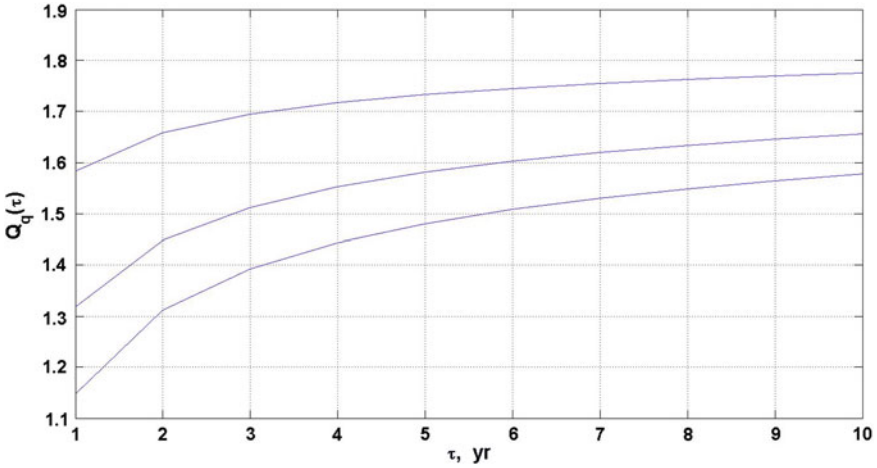


**Fig. 3.74** Annual flood damages in the USA 1940–2011. The GPD-quantiles for three different confidence levels: $q = 0.90$ (*lower curve*); 0.95 (*middle curve*); 0.99 (*upper curve*)

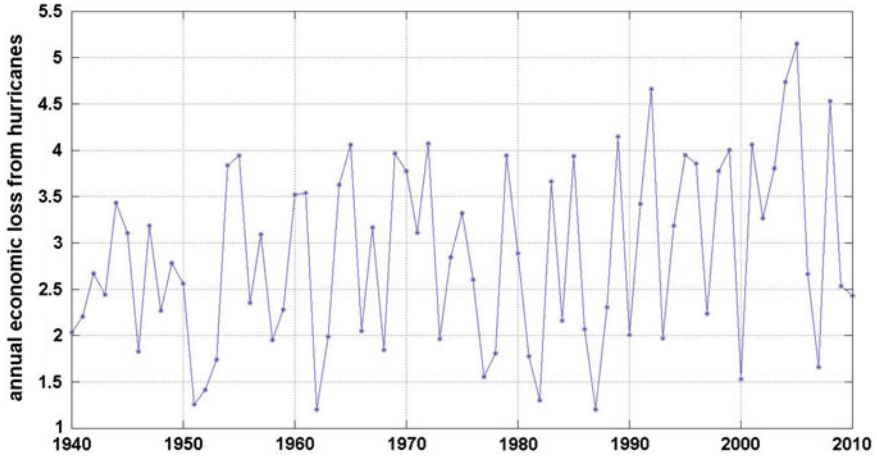


**Fig. 3.75** The time series of annual hurricane damages in log-scale, USA, 1940–2010. The damage is in $\$10^6$ (adjusted to 2010)

most costly years are responsible for 70 % of the total damage. Such sample can be characterized as sample with a "moderate concentration". Figure 3.78 shows the extreme part of the tail used for parameter estimation along with fitted GPD-curve. We see that the approximation is satisfactory although there are certain deviations.

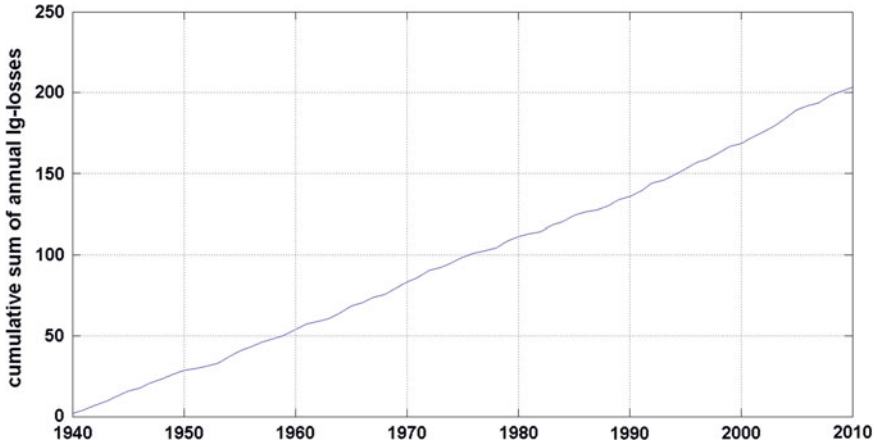


**Fig. 3.76** Annual hurricane damages in the USA 1940–2010. The cumulative sums of log-damages

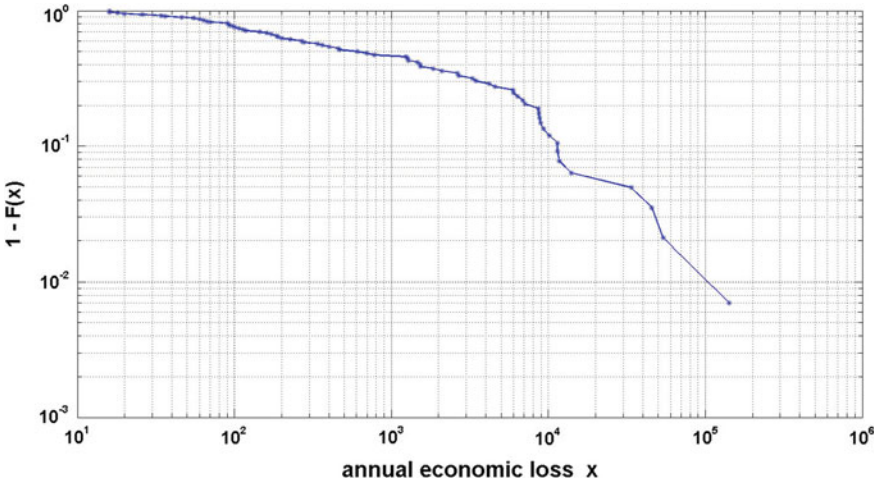


**Fig. 3.77** Annual hurricane damages in the USA 1940–2010. The sample tail $1 - F(x)$

Now we apply GPD-fitting to peaks over thresholds $h$. The resulting KD-distances characterizing goodness-of-fit are shown on Fig. 3.80. The best fitting occurs at $h = 1.5$ ($n = 64$; $\xi = -0.636 \pm 0.045$; $s = 2.368 \pm 0.352$; $p$–$v = 0.48$). The $Q$-quantiles in log-scale given by Eq. (2.44) are shown on Fig. 3.81.

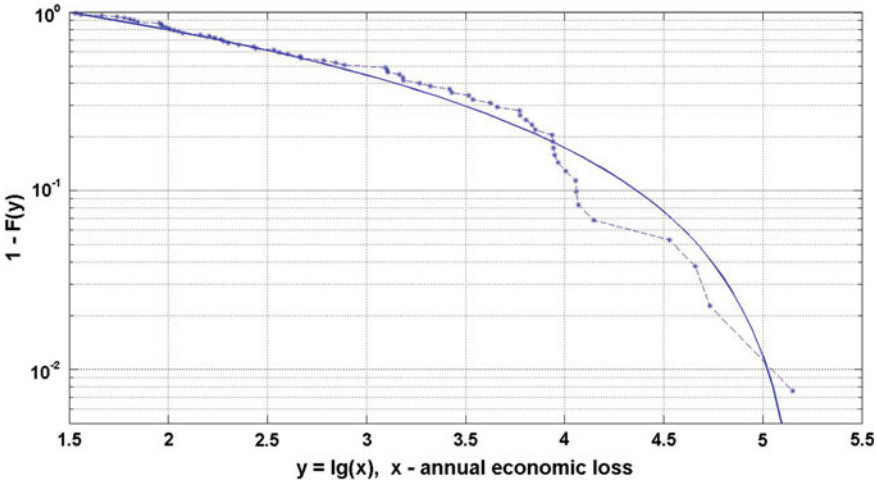


**Fig. 3.78** Annual hurricane damages in the USA 1940–2010. The extreme tail $1 - F(y)$ ($y = \log(x)$, $x$—annual hurricane damage) and approximating GPD-tail: $h = 1.5$; $\xi = -0.636$; $s = 2.368$; $n = 64$

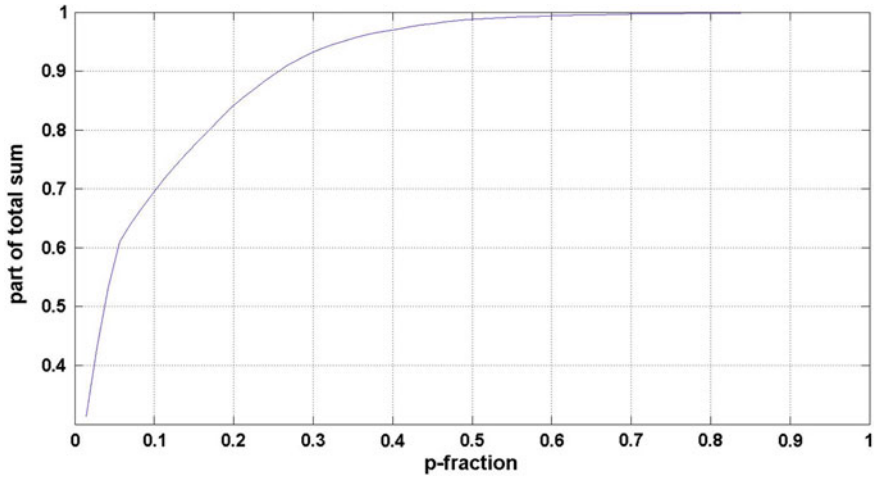


**Fig. 3.79** Annual hurricane damages in the USA 1940–2010. The contribution of the $p$-fraction of the costliest years to the total sum

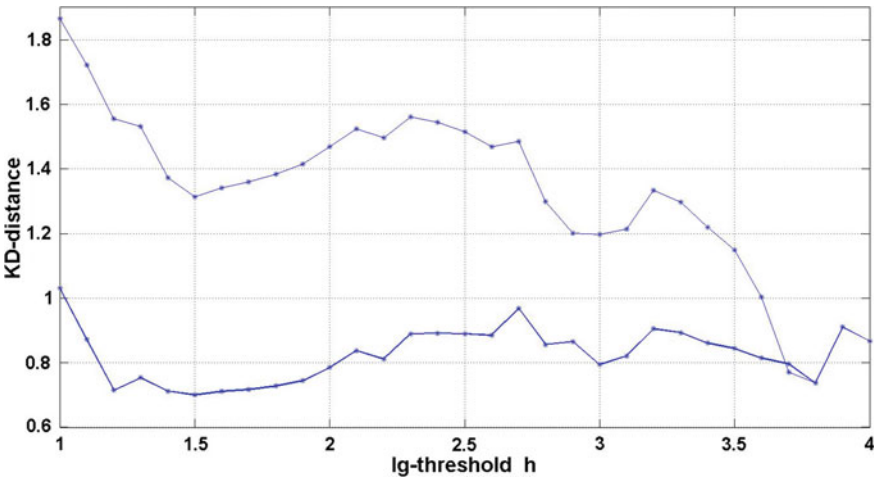


**Fig. 3.80** Annual hurricane damages in the USA 1940–2010. KD-distances for a grid of log-thresholds. *Thick line*—GPD-fitting; *thin line*—ED-fitting

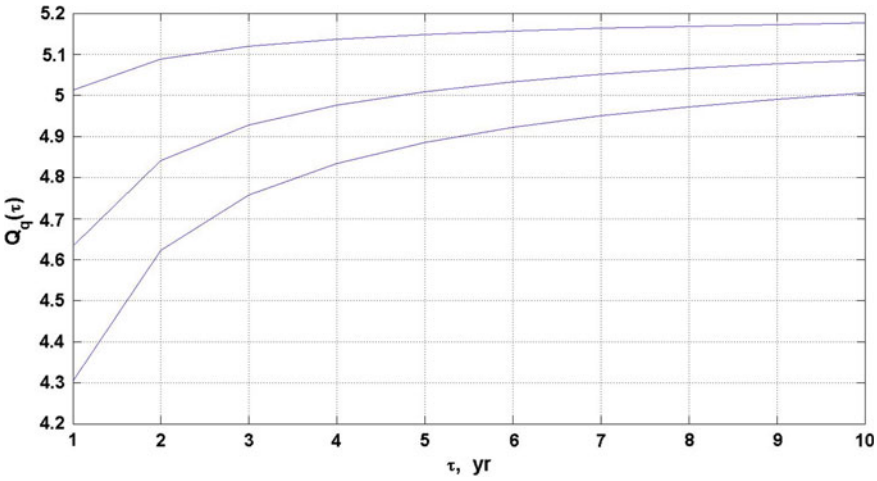


**Fig. 3.81** Annual hurricane damages in the USA 1940–2010. The GPD-quantiles for three different confidence levels: $q = 0.90$ (*lower curve*); 0.95 (*middle curve*); 0.99 (*upper curve*)

## References

Aptikaev FF (2009) Review of empirical scaling of strong ground motion for seismic hazard analysis. In: Trifunac MD (ed) Selected topics in earthquake engineering-from earthquake source to seismic design and hazard mitigation. Republika Srpska, pp 26–54

Blake B, Gibney E (2011) The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2010. NOOA Technical Memorandum NWS, NHS-6, August 2011, pp 1–47

Campbell KW (1981) Near-source attenuation of peak horizontal acceleration. Bull Seism Soc Am 71:2039–2070

Cornell CA (1968) Engineering seismic risk analysis. Bull Seism Soc Am 67:1173–1194

Graizer V, Kalkan E (2011) Modular filter-based approach to ground motion attenuation modeling. Seismol Res Lett 82:21–31

Kijko A, Sellevoll MA (1989) Estimation of earthquake hazard parameters from incomplete data files, Part I. Utilization of extreme and complete catalogues with different threshold magnitudes. Bull Seism Soc Am 79:645–654

Kijko A, Sellevoll MA (1992) Estimation of earthquake hazard parameters from incomplete data files, Part II. Incorporation of magnitude heterogeneity. Bull Seism Soc Am 82:120–134

Knopoff L, Kagan Y, Knopoff R (1982) b-values for foreshocks and aftershocks in real and simulated earthquake sequences. Bull Seism Soc Am 72:1663–1675

Lamarre M, Townshend B, Shah HC (1992) Application of the bootstrap method to qualify uncertainty in seismic hazard estimates. Bull Seism Soc Am 82:104–119

Mahdavian A, Aptikaev FF, Erteleva OO (2005) Ground motion parameters in seismically active zones of Iran. Izvestiya Phys Solid Earth 41(2):114–120

Pisarenko VF, Rodkin MV (2010) Heavy-tailed distributions in disaster analysis. Springer, Dordrecht-Heidelberg-London-New York

Pisarenko VF, Rodkin MV (2013) The new quantile approach: application to the seismic risk assessment. In: Rascobic B, Mrdja S (eds) Natural disasters: prevention, risk factors and management. NOVA Publishers, New York, pp 141–174

Pisarenko VF, Sornette D, Rodkin MV (2010) Distribution of maximum earthquake magnitudes in future time intervals: application to the seismicity of Japan (1923–2007). Earth Planets Space, 62:567–578

Steinberg VV, Saks MV, Aptikaev FF et al (1993) Methods of seismic ground motion estimation (handbook). In: Voprosy Inzhenernoi Seismologii, Issue 34. Moskow, Nauka. pp 5–97 (in Russian)

Stephens MA (1974) EDF statistics for goodness of fit and some comparisons. J Am Statist Ass 68(347):730–737

Usami T (1979) Study of historical earthquakes in Japan. Bull Earthquake Res Inst Univ Tokyo 54(3–4):399–439

Usami T (2002) Table of historical damaging earthquakes in Japan. In: Lee WHK, Kanamori H, Jennings PS, Kisslenger C (eds) International handbook of earthquake and engineering seismology (IASPEI), Part A, CD No. 1. Academic Press, San Diego

Utsu T (1979) Seismicity of Japan from 1885 through 1925—a new catalog of earthquakes of $M$ 6 felt in Japan and smaller earthquakes which caused damage in Japan. Bull Earthquake Res Inst Univ Tokyo 54(2):253–308

Utsu T (2002) A list of deadly earthquakes in the World: 1500–2000. In: Lee WK, Kanamori H, Jennings PC, Kisslinger C (eds) International handbook of earthquake and engineering seismology, Part A. Academic Press, San Diego, pp 691–717 (The revised and extended version is available at http://iisee.kenken.go.jp/utsu/index_eng.html)

Zlobin TK, Polets A. Yu (2012) Seismotectonics, deep structure and disastrous earthquakes in Kuril-Okhotsk region. LAP LAMBERT Academic Publishing GmbH & Co., KG Saarbrucken, pp 1–93 (In Russian)