Maximum Likelihood Estimators in a Statistical Model of Natural Catastrophe Claims with Stochastic Growth

Kukush, Alexander
National Taras Shevchenko University of Kyiv, Department of Mechanics & Mathematics
Volodymyrska st. 64
01601 Kyiv, Ukraine
E-mail: Alexander_kukush@univ.kiev.ua

1. Introduction

The results of this paper are joint with Prof. D. Pfeifer (Germany) and Dr. Yu. Chernikov (Ukraine). We present an approach to the investigation of natural catastrophe claims (e.g., with respect to wind storm losses). The so-called *Nevzorov's record model*, see Nevzorov (1988), is studied. We suppose that the yearly catastrophe claims are realizations of an independent sequence $\{X_i, i \geq 1\}$ of random variables (r.v.) with support $R^+ := [\,0, \infty\,]$ and continuous cumulative distribution function (cdf) $\{F_i, i \geq 1\}$, such that

$$F_i = F^{\gamma(i)}, \text{ with } \gamma(i) = \gamma^{i-1}, \gamma \geq 1. \tag{1}$$

Here F is a fixed continuous cdf with $F(0) = 0$. If $\gamma > 1$ then the X_i 's are stochastically increasing, and in a border case $\gamma = 1$ the sequence is i.i.d.

Within the *semi-parametric* approach the cdf F is unspecified. And within the parametric approach we assume that F belongs to the parametric class of Fréchet distributions (one of the extreme-value distribution classes) with unknown shape and scale parameters; our parametric model provides that a possible trend in the data is of exponential type. D. Pfeifer (1997) proposed to use a combination of parametric and semi-parametric methods to investigate catastrophe claims in the presence of trend. In the first step, the type of trend is analyzed using the number of record values in the times series of claims data, and in the second step, a maximum-likelihood estimator (MLE) is constructed from the data taking into account what type of trend has been detected before. In order to check the validity of the model assumptions, the estimates for the trend parameter obtained from both steps are compared. In Kukush et al. (2004) the two approaches are compared for U.S. hurricane events from 1949 to 1992 and for Japanese typhoon events from 1977 to 1991, and implications for insurance applications are considered.

Here we review some asymptotic properties of the MLE in Nevzorov's record model and corresponding goodness-of-fit (GOF) tests. The asymptotic results were announced in Kukush (1999) and then proven in Kukush et al. (2004). The GOF tests were constructed in Kukush and Chernikov (2001), and in Kukush and Chernikov (2002) it was shown that in both semi-parametric and parametric models the MLE is asymptotically efficient in the sense of Hajék bound.

The paper is organized as follows. Section 2 introduces Nevzorov's record model. Section 3 states the consistency, asymptotic normality, and efficiency of the semi-parametric MLE, and the semi-parametric GOF test is constructed. Section 4 gives similar results for the parametric model, and Section 5 reports the comparison of semi-parametric and parametric approaches in data analysis.

The proofs of the theorems can be found in Kukush et al. (2004) and Kukush & Chernikov (2001), (2002).

2. Nevzorov's record model

In the model (1) define record indicators as $I_1 = 1;\ I_n = 1,$ if $X_n > \max\{X_1,..., X_{n-1}\},$ and $I_n = 0,$ otherwise, for $n \geq 2.$ The record indicators are independent r.v. with

$$p_n(\gamma) := P_\gamma(I_n = 1) = \frac{1}{1+\gamma^{-1}+...+\gamma^{-n+1}} \ .$$

The unknown parameter $\gamma,$ see (1), is called *trend parameter*. Given the observations of record indicators till moment $n \geq 2,$ the log-likelihood function for $\gamma \geq 1$ is equal to

$$L_n(\gamma) := \sum_{i=2}^{n} I_i \log p_i(\gamma) + \sum_{i=2}^{n} (1 - I_i) \log(1 - p_i(\gamma)). \qquad (2)$$

The semi-parametric MLE $\hat{\gamma} = \hat{\gamma}_n$ is defined as a measurable function of the observed record indicators for which $\hat{\gamma} \in \arg\max\{L_n(\gamma), \gamma \geq 1\}.$

3. Asymptotic properties of semi-parametric MLE and goodness-of-fit test

Theorem 1: *Let* $\gamma \geq 1.$ *Then the MLE* $\hat{\gamma}$ *is strongly consistent.*

Theorem 2: *Let* $\gamma > 1.$ *Then the MLE* $\hat{\gamma}$ *is asymptotically normal, namely the normalized estimator* $\sqrt{n}(\hat{\gamma}_n - \gamma)$ *converges in distribution to a normal law with mean* 0 *and variance* $\sigma_\infty^2 = \gamma^2(\gamma - 1).$

Now, we state the asymptotic efficiency of the estimator in the sense of Hajék bound; see Ibragimov & Has'minskii (1981) for general theory of the asymptotic efficiency. Introduce the class $W_{e,2}$ of bell-shaped loss functions. Those real-valued functions satisfy the following conditions:

(a) $w(u) \geq 0, u \in R; w(0) = 0,\ w$ is continuous at $u = 0$ and is not identical $0.$

(b) w is even function.

(c) w is non-decreasing for $u \geq 0.$

(d) The growth of w as $u \to +\infty$ is slower than any one of the functions $\exp(\varepsilon u^2), \varepsilon > 0.$

Denote by ξ a standard Gaussian r.v.

Theorem 3: *Let* $\gamma_0 > 1.$

(a) For any real-valued function w *that is bounded, Borel measurable and continuous a.e. with respect to Lebesgue measure, it holds*

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \sup_{\gamma:|\gamma-\gamma_0|<\delta} E_\gamma w\left(\sqrt{\frac{n}{\gamma_0 - 1}} \times \frac{\hat{\gamma}_n - \gamma}{\gamma_0}\right) = Ew(\xi).$$

(b) For any family γ_n^* *of estimators of* γ*, based on observations of the record indicators and for any loss function* $w \in W_{e,2}$*, the inequality holds:*

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \sup_{\gamma:|\gamma-\gamma_0|<\delta} E_\gamma w\left(\sqrt{\frac{n}{\gamma_0 - 1}} \times \frac{\gamma_n^* - \gamma}{\gamma_0}\right) \geq Ew(\xi).$$

The latter inequality gives a lower bound for the loss of arbitrary normalized estimator. Theorem 3 shows that the MLE has asymptotically the smallest possible averaged loss.

Now, we pass to the GOF test. Let γ_0 be the true value of the trend parameter. One can show that with probability 1 the normalized log-likelihood function (2), $Q_n(\gamma) := L_n(\gamma)/n,$ converges uniformly on each interval $[a,b] \subset (1, +\infty)$ to the functional $Q_\infty(\gamma, \gamma_0)$ such that for $\gamma > 1, \gamma_0 > 1,$

$$Q_\infty(\gamma, \gamma_0) = (1 - \gamma_0^{-1}) \log(\gamma - 1) - \log \gamma.$$

Theorem 4: *In semi-parametric model with* $\gamma_0 > 1$

$$T_n := \sqrt{n}(Q_n(\hat{\gamma}_n) - Q_\infty(\hat{\gamma}_n, \hat{\gamma}_n)) \to^d N(0, \sigma^2(\gamma_0)),$$

with $\sigma^2(\gamma_0) = \dfrac{2l_0(l_0\gamma_0^2 - 2l_0\gamma_0 + l_0 + 1)}{\gamma_0^4}, \quad l_0 = (\log(\gamma_0 - 1))^2.$

The GOF test for semi-parametric model is based on the next corollary.

Corollary 5: *Under the conditions of Theorem 4,* $V_n := T_n / \sigma(\hat{\gamma}_n) \to^d N(0,1).$

4. The three-parametric model

Now, we assume that the cdf F_i for the yearly claims are of the form

$$F_i(x) = \exp(-\gamma^{i-1}(Ax)^{-\alpha}),\ n = 1,2,...,\ x > 0.$$

Here $(A,\alpha,\gamma) \in \Theta := (0,+\infty)\times(0,+\infty)\times[1,+\infty)$ are parameters of interest. Then the log-likelihood function for the observed data set $X_1,..., X_n$ is given by

$$L(A,\alpha,\gamma) := \frac{n(n-1)}{2}\log\gamma - (\alpha+1)\sum_{i=1}^{n}\log X_i - \sum_{i=1}^{n}\gamma^{i-1}(AX_i)^{-\alpha} + n\log(\alpha A^{-\alpha}). \qquad (3)$$

Define the joint MLE of the parameters of interest as a measurable vector function $(\hat{A},\hat{\alpha},\hat{\gamma})$ of $X_1,..., X_n$ for which

$$(\hat{A},\hat{\alpha},\hat{\gamma}) \in \arg\max L(A,\alpha,\gamma),$$

where maximum is taken over Θ. One can show that that the maximum here is attained eventually, that is for all $n \geq n_0(\omega)$, a.s.

Theorem 6: *Let* $\gamma \geq 1$. *Then the joint MLE is strongly consistent, moreover*

$$\hat{A} \to A,\ \hat{\alpha} \to \alpha,\ n(\hat{\gamma} - \gamma) \to 0, \quad \textit{as } n \to \infty, \quad \textit{a.s.}$$

Thus, in the three-parametric model, the trend parameter γ is better estimable than other two parameters.

Theorem 7: *If* $\gamma > 1$ *then the joint MLE is asymptotically normal, namely the normalized estimator*

$$\sqrt{n}\,(R_n^t T R_n)^{1/2}(\hat{A} - A,\ \hat{\alpha} - \alpha,\ n(\log\hat{\gamma} - \log\gamma))^t$$

converges in distribution to a normal law with mean 0 *and a unit covariance matrix, where the entries of the right triangular matrix* R_n *and symmetric matrix* T *are:*

$$R_n^{11} = \frac{\alpha}{A}, R_n^{12} = R_n^{13} = 0, R_n^{22} = -\frac{1}{\alpha}, R_n^{23} = n \times \frac{\log \gamma}{\alpha}, R_n^{33} = -1,$$
$$T^{11} = 1, T^{12} = \tau, T^{13} = \frac{1}{2}, T^{22} = \frac{\pi^2}{6} + \tau^2, T^{23} = \frac{\tau}{2}, T^{33} = \frac{1}{3}.$$

Here $\tau = 1 - \gamma_e$, γ_e *stands for Euler's constant,* $\gamma_e \approx 0.5772.$

Theorems 6 and 7 are applied to forecast claims. The observed yearly claims can be represented as

$$X_i = f_i Z_i^{1/\alpha}, \quad f_i = f_i(A, \alpha, \gamma) = A^{-1}(\gamma^{1/\alpha})^{i-1}, \quad i = 1, 2, ... \tag{4}$$

Then the transformed observations $Z_1, Z_2, ...$ form an i.i.d. sequence with standard Fréchet distribution $F(x) = \exp(-x^{-1}), x > 0.$ We interpret the trend as a growth of the median of X_i, and relation (4) clearly shows that it is a trend of an exponential type. The forecast of claims for the year number $k > n$ will be

$$\hat{X}_k = f_k(\hat{A}, \hat{\alpha}, \hat{\gamma}) \times med(Z_1^{1/t})|_{t=\hat{\alpha}} .$$

And Theorem 7 makes it possible to construct a confidence interval for the forecast based on the confidence region for the true values of A, α, γ.

In Kukush & Chernikov (2002), (2001) theorems analogous to Theorems 3 and 4 are proven for the joint MLE in the three-parameter model. Therefore, in this model the MLE is asymptotically efficient in the sense of Hajék bound, and a GOF test can be constructed based on the log-likelihood function (3) and the MLE.

5. Analysis of U.S. and Japanese data

Here we briefly report numerical results from Pfeifer (1997) concerning the following two data sets:

(a) yearly claims in Million U.S. $ from U.S. hurricane events from 1949 to 1992,

(b) yearly claims in 1000 JYen from Japanese typhoon events from 1977 to 1991.

The graphical data displayed in logarithmic scale showed that the assumption of an exponential trend in the data was reasonable. A stochastic search procedure was performed to compute MLEs based on maximization of log-likelihood functions (2) and (3). The trend parameters were estimated from the three approaches:

(a) *semi-parametric* as described in Section 2,

(b) *joint maximum likelihood* as described in Section 4,

(c) *least squares* – from the graphical analysis; here $\hat{\gamma} = \exp(\hat{\alpha}\,\hat{m})$ where $\hat{m}$ is the estimated slope for the regression line in logarithmic scale, and $\hat{\alpha}$ is the estimator of α from the joint MLE.

For the *U.S. data*, all the three approaches give nearly the same estimator for γ:

$$\hat{\gamma} \approx 1.15; \quad 1.10; \quad 1.11.$$

For the U.S., $\hat{\alpha} \approx 1.06$, and the prediction line in logarithmic scale looks reasonable.

The situation is not so clear for the *Japan data,* where respectively,

$$\hat{\gamma} \approx 1.81; \quad 1.30; \quad 1.34.$$

The shape parameter is estimated as $\hat{\alpha} \approx 0.91$. Thus, the Japan data have steeper trend compared with the U.S. data. For Japan, the prediction line looks unreasonable because of poor fitting of the three-parametric model.

Next, in Kukush et al. (2004) asymptotic 95 percent confidence regions were constructed based on Theorems 2 and 7. Below we present those numerical results.

U.S. data:

(a) Semiparametric case. $\gamma \in (1.0184, 1.2814)$.

(b) The three-parametric case. Projections of the confidence ellipsoid are:

$$A \in (0.0729, 0.1679); \quad \alpha \in (0.8858, 1.2492); \quad \gamma \in (1.0857, 1.1188).$$

Japan data:

(a) Semiparametric case. $\gamma \in (0.9856, 2.6341)$.

(b) The three-parametric case. Projections of the confidence ellipsoid are:

$$A \in (0.0003, 0.0029); \alpha \in (0.6236, 1.1953); \gamma \in (1.2147, 1.3814).$$

Thus, for U.S. data the confidence region is quite small, and for Japan data the region is larger.

In Kukush et al. (2004) the efficiency of the proposed methods was illustrated via simulation. 1000 series of n random Fréchet distributed values were simulated with parameters (A, α, γ) similar to one estimated on the U.S. and Japan data and with realistic n. For those parameters the statistics were computed corresponding to Theorems 2 and 7, and we checked how many simulated data fall into the 95% confidence region. The simulations showed that the proposed methods can be applied even for small sample size, that insurance and re-insurance companies deal with, though the empirical coverage probability is often a bit less than 0.95 for 95% regions.

REFERENCES

Ibragimov, I.A. and Has'minskii, R.Z. (1981), *Statistical Estimation, Asymptotic Theory*, Springer, New York – Berlin.

Kukush, A. (1999), On maximum likelihood estimator in a statistical model of natural catastrophe claims, *Theory of Stochastic Process* **5(21)**, 64-70.

Kukush, A.G. and Chernikov, Y.V. (2001), Goodness-of-fit test in Nevzorov's model, *Theory of Stochastic Process* **7(23)**, 203-214.

Kukush, A.G. and Chernikov, Yu.V. (2002), Efficiency of maximum likelihood estimators in Nevzorov's record model, in *Proceedings of the Ukrainian Mathematical Congress-2001*, Kyiv, August, 2001, Probability Theory and Mathematical Statistics, Section 9.

Kukush, A., Chernikov, Yu., and Pfeifer, D. (2004), Maximum likelihood estimators in a statistical model of natural catastrophe claims with trend, *Extremes* **7**, 309-337.

Nevzorov, V.B. (1988), Records, *Theory of Probability and Applications* **32**, 201-228.

Pfeifer, D. (1997), A statistical model to analyse natural catastrophe claims by mean of record values, in *Proceedings of the 28 International ASTIN Colloquium*, Cairns, Australia 10-12 August, 1997, The Institute of Actuaries of Australia.