COMMUN. STATIST.—THEORY METH., 24(7), 1875–1896 (1995)

BOOTSTRAP BASED GOODNESS OF FIT TESTS FOR THE GENERALIZED POISSON MODEL

Norbert Henze[1] and Bernhard Klar[2]

[1] *Institut für Mathematische Stochastik, Universität Karlsruhe, Englerstr. 2, 76128 Karlsruhe*

[2] *Institut für Wissenschaftliches Rechnen und Mathematische Modellbildung, Universität Karlsruhe, Engesserstr. 6, 76128 Karlsruhe*

Key Words and Phrases: Generalized Poisson distribution; goodness of fit test; empirical distribution function; weak convergence; parametric bootstrap; atmospheric circulation patterns

ABSTRACT

Due to its versatile nature, the Generalized Poisson distribution (*GPD*) of Consul and Jain (1973) has been an object of sustained interest. However, apart from the classical χ^2-test with its inherent problems, there is a paucity of genuine goodness of fit tests for checking the *GPD* model on the

AMS 1991 Subject classifications: Primary 62G10, Secondary 62G30

basis of given data. In this paper we study empirical distribution function based tests for the *GPD* model. A key tool is a weak convergence result for an estimated (discrete) empirical process, regarded as a random element in some suitable sequence space. A parametric bootstrap version of the procedure is shown to maintain a desired level of significance very closely even for small sample sizes. The test is applied to data sets of frequencies of the duration of atmospheric circulation patterns.

1 Introduction

The Poisson distribution provides an adequate model for counts arising from random phenomena with intrinsic "ideal spatial randomness". It depends on a single parameter which is the mean as well as the variance.

Since this principle of ideal spatial randomness is not very natural in many situations, there have been several ideas to generalize the Poisson distribution (for a survey of classical approaches see e.g. Haight, 1967, Chapter 3).

Consul and Jain (1973) suggested an interesting new generalized Poisson distribution (henceforth called *GPD*) with two parameters which, due to its versatile nature, has been an object of sustained interest. The definition of the *GPD* model is based on the fact that

$$e^a = \sum_{k=0}^{\infty} \frac{a(a+kb)^{k-1}}{k!} e^{-kb}$$

($a > 0$, $b < 1$, $|b| < e^{b-1}$; see Jensen, 1902).
A random variable X is said to have a (untruncated) *GPD* (λ, ξ) distribution if

$$P(X = k) = \frac{\lambda(\lambda + \xi k)^{k-1}}{k!} e^{-\lambda - \xi k}, \qquad k = 0, 1, 2, \ldots \tag{1.1}$$

where $\lambda > 0$, $0 \leq \xi < 1$. Obviously, the distribution of X is Poisson with parameter λ if $\xi = 0$. Since

$$E(X) = \frac{\lambda}{1-\xi}, \qquad \mathrm{Var}(X) = \frac{\lambda}{(1-\xi)^3} \tag{1.2}$$

(Consul and Jain, 1973), the variance of the untruncated *GPD* distribution is always larger than or equal to the mean

To enhance the flexibility of the *GPD* model so as to include distributions where the variance is smaller than the mean, Consul and Jain proposed to admit negative values of ξ by putting

$$k_0 = k_0(\lambda,\xi) = \max\{k \geq 0 : \lambda + \xi k > 0\},$$
$$S(\lambda,\xi) = \sum_{k=0}^{k_0} \frac{\lambda}{k!}(\lambda+\xi k)^{k-1} e^{-\lambda-\xi k}$$

This leads to the right-truncated *GPD*(λ,ξ) distribution having probability mass function

$$P(X=k) = \begin{cases} \frac{1}{S(\lambda,\xi)} \frac{\lambda(\lambda+\xi k)^{k-1}}{k!} e^{-\lambda-\xi k}, & 0 \leq k \leq k_0 \\ 0, & k > k_0 \end{cases} \quad (1.3)$$

Note that, formally, $k_0 = \infty$ and $S(\lambda,\xi) = 1$ if $0 \leq \xi < 1$ which implies that (1.1) may be regarded as a special case of (1.3). However, the expressions (1.2) as well as simple formulae for skewness and kurtosis and important properties like *GPD*$(\lambda_1,\xi) *$ *GPD*$(\lambda_2,\xi) =$ *GPD*$(\lambda_1+\lambda_2,\xi)$ for convolutions (Consul, 1989) are only valid for the untruncated *GPD*(λ,ξ) model (1.1) (see, e.g., Johnson, Kotz and Kemp, 1992, p. 396 ff.) This fact has apparently not been emphasized enough, and thus it is not surprising that (1.2) is tacitly assumed to hold also for negative values of ξ (see, e.g., the recent paper of Alzaid and Al-Osh, 1993)

Since the truncated *GPD* distribution, as a descriptive model, provides an excellent fit for data in numerous applications (see e.g., Janardan and Schaeffer, 1977), Consul and Shoukri (1985) made a detailed error analysis of the effect of the multiplication factor $S(\lambda,\xi)^{-1}$ in (1.3). Their conclusion is that for most practical applications of the *GPD* model, $S(\lambda,\xi)$ is very close to 1 and thus becomes unnecessary. In particular, this holds if the truncation point $k_0(\lambda,\xi)$ is at least 4

Because of its attractiveness for applications (see Consul, 1989, Ch. 5), there is a vital interest in testing the goodness of fit of the *GPD* model for given data. In this respect, a mere graphical check of the closeness of observed and fitted frequencies is often misleading, since *no eye observation of such diagrams, however experienced, is really capable of discriminating whether or not the observations differ from the expectation by more than we would expect from the circumstances of random sampling. (R. A. Fisher, 1925, p. 35).*

Consul (1989) recommends to perform the classical χ^2–test in order to assess the goodness of fit (GOF) of observed and fitted frequencies and provides a FORTRAN program to compute χ^2 values. However, a careless use of this test may result in erroneous decisions for the following reasons.

Firstly, Consul (1989, p. 235ff.) incorporates the option of using maximum likelihood (ML) as well as moment (M) estimates for λ and ξ. Now, the χ^2–test statistic with moment fitted frequencies does not have a limiting χ^2 distribution under the *GPD* model (for a corrected version of the χ^2–test in this case see Mirvaliev, 1987).

Secondly, ML estimates for λ and ξ are computed **before** an eventual combination of classes with low observed frequencies is performed. This also has an effect on the limiting null distribution of the χ^2 test statistic (which, of course, is a χ^2 distribution if ML estimation is done **after** combining classes).

Apart from these problems, sample sizes occuring in applications are often not large enough to justify a χ^2 approximation to the null distribution of the test statistic.

Concerning other methods in testing the GOF for families of discrete distributions, there have been recently various suggestions to use the empirical generating function in testing the GOF for the Poisson model (see, e.g., Baringhaus and Henze (1992), Nakamura and Pérez–Abreu (1993), and Rueda et al. (1991)). However, these approaches do not carry over to our more difficult testing problem since there is no explicit form for the generating function of the *GPD*(λ, ξ) distribution.

To overcome these deficiencies, we suggest the use of classical GOF test statistics like those of Kolmogorov–Smirnov or Cramér - von Mises in order to assess the validity of the *GPD* model. Of course, these statistics are well known in testing the GOF for a **continuous distribution** (see e.g. D'Agostino and Stephens, 1986).

The motivation to consider this problem stems from work of Bárdossy and Plate (1992) where the *GPD* distribution is incorporated in a space–time model for daily rainfall to describe the duration of atmospheric circulation patterns.

The paper is organized as follows. In Section 2 we specify the setup and give the mathematical derivations. In Section 3 we present the results of a small power study and apply the tests to observed frequencies of the duration of circulation patterns. We would like to thank András Bárdossy for making these data available to us.

2 Main results

On a common probability space $(\Omega, \mathcal{A}, P)$, let $X, X_1, \ldots, X_n, \ldots$ be a sequence of independent and identically distributed random variables taking nonnegative integer values. Putting

$$\Theta = \{\vartheta = (\lambda, \xi) \,:\, 0 < \lambda < \infty,\ 0 < \xi < 1\},$$

and writing P^X for the unknown distribution of X, the problem is to test, on the basis of $X_1, \ldots, X_n$, the hypothesis

$$H_0 \,:\, P^X \in \mathit{GPDU}$$

where $\mathit{GPDU} = \{\mathit{GPD}(\lambda, \xi) : (\lambda, \xi) \in \Theta\}$ denotes the class of all untruncated GPD distributions.

It should be remarked at the outset that a restriction to untruncated *GPD* distributions (just as Mirvaliev did) is necessary in order to make "the mathematics work". However, in the spirit of the discussion given in Section 1, negative values of ξ do not deter the tests from "working in practice".

In what follows, let $F(t) = P(X \le t)$ denote the distribution function (df) of X, and write

$$F(t, \vartheta) = \sum_{k=0}^{[t]} \frac{\lambda(\lambda + \xi k)^{k-1}}{k!} \cdot e^{-\lambda - \xi k}, \qquad t \ge 0, \tag{2.1}$$

for the df of $\mathit{GPD}(\lambda, \xi)$, where $\vartheta = (\lambda, \xi)$ and $[t]$ is the largest nonnegative integer not exceeding t. Furthermore, writing $\mathbf{1}\{\cdot\}$ for the indicator function, let

$$F_n(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\{X_j \le t\}$$

be the empirical distribution function (edf) of $X_1, \ldots, X_n$. It is natural to measure the GOF of the class *GPDU* for the data $X_1, \ldots, X_n$ by choosing a "distance" between $F_n(\cdot)$ and $F(\cdot, \tilde{\vartheta}_n)$, where $\tilde{\vartheta}_n = (\tilde{\lambda}_n, \tilde{\xi}_n)$ is some "good" estimator for ϑ based on $X_1, \ldots, X_n$. In what follows, we advocate the use of the M-estimator $\hat{\vartheta}_n = (\hat{\lambda}_n, \hat{\xi}_n)$ which, in view of (1.2), takes the simple form

$$\hat{\lambda}_n = \left(\frac{\bar{X}_n^3}{\hat{\sigma}_n^2}\right)^{\frac{1}{2}}, \qquad \hat{\xi}_n = 1 - \left(\frac{\bar{X}_n}{\hat{\sigma}_n^2}\right)^{\frac{1}{2}}, \tag{2.2}$$

where $\bar{X}_n = n^{-1}\sum_{j=1}^{n} X_j$, $\hat{\sigma}_n^2 = n^{-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2$. If $\hat{\sigma}_n^2 = 0$ we put $\hat{\lambda}_n = \hat{\xi}_n = 0$.

Note that, although the validity of H_0 implies that $\mu < \sigma^2$, where, for short, $\mu = E(X)$ and $\sigma^2 = \mathrm{Var}(X)$, $\bar{X}_n$ may be larger than $\hat{\sigma}_n^2$ which entails a negative value for $\hat{\xi}_n$. In this case $F(\cdot, \hat{\vartheta}_n)$ is defined to be the df of the truncated *GPD* distribution with parameter $\hat{\vartheta}_n$. However, by the strong law of large numbers, we have

$$\hat{\xi}_n \underset{n\to\infty}{\longrightarrow} 1 - \left(\frac{\mu}{\sigma^2}\right)^{\frac{1}{2}} = \xi \qquad P_\vartheta\text{-almost surely,}$$

where ξ is positive by definition. Here and in what follows we use the notation P_ϑ to denote probabilities computed under H_0 when $\vartheta = (\lambda, \xi)$ is the "true" parameter value.

Of course, at least in principle, other methods of estimation for ϑ like e.g. maximum likelihood are possible (however, see Consul and Shoukri, 1984, for theoretical and practical problems with ML estimation in this context).

For testing H_0, we shall consider the Kolmogorov–Smirnov and Cramér–von Mises type statistics

$$K_n = \sup_{k\ge 0} \sqrt{n}\,\left|F_n(k) - F(k, \hat{\vartheta}_n)\right| \tag{2.3}$$

and

$$C_n = n\sum_{k=0}^{\infty}(F_n(k) - F(k, \hat{\vartheta}_n))^2[F(k, \hat{\vartheta}_n) - F(k-1, \hat{\vartheta}_n)], \tag{2.4}$$

respectively These are functionals of the *estimated (discrete) empirical process*

$$\mathcal{Z}_n = (Z_{n,k})_{k\geq 0}, \tag{2.5}$$

where $Z_{n,k} = \sqrt{n}\,(F_n(k) - F(k,\hat{\vartheta}_n))$.

Since $\lim_{k\to\infty} Z_{n,k} = 0$ almost surely, we may regard $\mathcal{Z}_n$ as a random element with values in the Banach space c_0 of all sequences $x = (x_k)_{k\geq 0}$ converging to zero, equipped with the norm $||x|| = \sup_{k\geq 0} |x_k|$

Henze (1994) studied the weak convergence of c_0–valued estimated empirical processes in a general setting, thus avoiding the sophisticated strong approximation methodology of Burke et. al. (1979) which, besides, does not cover a locally *uniform* convergence needed for an indispensible parametric bootstrap. In the following, we prove that the pertinent regularity conditions (see assumptions (A1), (A2) of Henze, 1994) are satisfied in the present situation if estimation of ϑ is done by the moment method

To this end, recall the definition of the moment estimator $\hat{\vartheta}_n = (\hat{\lambda}_n, \hat{\xi}_n)$ given in (2.2) The next result shows that the sequence $\hat{\vartheta}_n$ satisfies a standard regularity condition

Lemma 2.1:
Let $l(k,\vartheta) = (l_1(k,\vartheta), l_2(k,\vartheta))$, where

$$\begin{aligned} l_1(k,\vartheta) &= \frac{1-\xi}{2}\left\{3(k-\mu) - (1-\xi)^2[(k-\mu)^2 - \sigma^2]\right\} \\ l_2(k,\vartheta) &= \frac{1}{2(1-\xi)}\left\{\frac{\mu}{\sigma^4}[(k-\mu)^2 - \sigma^2] - \frac{k-\mu}{\sigma^2}\right\} \end{aligned}$$

($\mu = \lambda(1-\xi)^{-1}$, $\sigma^2 = \lambda(1-\xi)^{-3}$) Then, under P_ϑ, we have

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} l(X_j,\vartheta) + \varepsilon_n, \tag{2.6}$$

where $\varepsilon_n = (\varepsilon_{n,1}, \varepsilon_{n,2}) = o_{P_\vartheta}(1)$ as $n \to \infty$.

PROOF In order to get rid of square roots, note that

$$\sqrt{n}(\hat{\lambda}_n - \lambda) = \sqrt{n}(\hat{\lambda}_n^2 - \lambda^2)\cdot\frac{1}{2\lambda} - \frac{1}{\sqrt{n}}\left[\sqrt{n}(\hat{\lambda}_n - \lambda)\right]^2\cdot\frac{1}{2\lambda}. \tag{2.7}$$

We will prove that

$$\sqrt{n}(\hat{\lambda}_n^2 - \lambda^2) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} 2\lambda l_1(X_j, \vartheta) + \tilde{\varepsilon}_{n,1}, \tag{2.8}$$

where $\tilde{\varepsilon}_{n,1} = o_{P_\vartheta}(1)$ as $n \to \infty$. Since

$$\sqrt{n}(\hat{\lambda}_n^2 - \lambda^2) = \sqrt{n}(\hat{\lambda}_n - \lambda)(2\lambda + o_{P_\vartheta}(1)),$$

(2.8) would entail the tightness of $(\sqrt{n}(\hat{\lambda}_n - \lambda))_{n\geq 1}$, whence, by (2.7),

$$\sqrt{n}(\hat{\lambda}_n - \lambda) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} l_1(X_j, \vartheta) + o_{P_\vartheta}(1). \tag{2.9}$$

A straightforward calculation yields

$$\sqrt{n}(\hat{\lambda}_n^2 - \lambda^2) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \frac{3\mu^2}{\sigma^2}(X_j - \mu) - \frac{\mu^3}{\sigma^4}[(X_j - \mu)^2 - \sigma^2] \right\} + \tilde{\varepsilon}_{n,1},$$

where $\mu = \lambda(1-\xi)^{-1}$, $\sigma^2 = \lambda(1-\xi)^{-3}$ and

$$\begin{aligned} \tilde{\varepsilon}_{n,1} &= \frac{\sqrt{n}(\bar{X}_n - \mu)^3}{\hat{\sigma}_n^2} + 3\mu \frac{\sqrt{n}(\bar{X}_n - \mu)^2}{\hat{\sigma}_n^2} \\ &\quad + 3\mu^2 \sqrt{n}(\bar{X}_n - \mu)\left(\frac{1}{\hat{\sigma}_n^2} - \frac{1}{\sigma^2}\right) + \frac{\mu^3}{\sigma^2\hat{\sigma}^2}\sqrt{n}(\bar{X}_n - \mu)^2 \\ &\quad + \frac{\mu^3}{\sigma^2}\left(\frac{1}{\sigma^2} - \frac{1}{\hat{\sigma}_n^2}\right)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}[(X_j - \mu)^2 - \sigma^2] \end{aligned}$$

By Slutzky's Lemma and the Central Limit Theorem, we have $\tilde{\varepsilon}_{n,1} = o_{P_\vartheta}(1)$. Using the fact that $\mu^2/\sigma^2 = \lambda(1-\xi)$, and $\mu^3/\sigma^4 = \lambda(1-\xi)^3$, (2.8) follows. The assertion

$$\sqrt{n}(\hat{\xi}_n - \xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} l_2(X_j, \vartheta) + \varepsilon_{n,2}, \qquad \varepsilon_{n,2} = o_{P_\vartheta}(1) \tag{2.10}$$

is proved similarly. Here, square roots may be avoided by putting $\tilde{\xi} = 1 - \xi$, $\xi_n^* = 1 - \hat{\xi}_n$. Then, starting from (2.7) with $\hat{\lambda}_n$ replaced by ξ_n^* and λ replaced by $\tilde{\xi}$, and noting that

$$\sqrt{n}((\xi_n^*)^2 - \tilde{\xi}^2) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \frac{X_j - \mu}{\sigma^2} - \frac{\mu}{\sigma^4}[(X_j - \mu)^2 - \sigma^2] \right\} + \tilde{\varepsilon}_{n,2},$$

where

$$\begin{aligned}\tilde{\varepsilon}_{n2} &= -\frac{1}{\sigma^2\hat{\sigma}_n^2}(\bar{X}_n-\mu)\sqrt{n}(\hat{\sigma}_n^2-\sigma^2)-\frac{\mu}{\sigma^2\hat{\sigma}_n^2}\sqrt{n}(\bar{X}_n-\mu)^2\\ &\quad -\frac{\mu}{\sigma^2}\left(\frac{1}{\hat{\sigma}_n^2}-\frac{1}{\sigma^2}\right)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}[(X_j-\mu)^2-\sigma^2]\\ &= o_{P_\vartheta}(1),\end{aligned}$$

(2.10) follows. Combining (2.10) with (2.9) yields (2.6). ■

By straightforward algebra it follows that $E_\vartheta[l_j(X,\vartheta)]=0\quad(j=1,2)$. Furthermore, $D(\vartheta)=E_\vartheta[l(X,\vartheta)'l(X,\vartheta)]$ defines a finite nonnegative matrix that depends continuously on ϑ. Here and in what follows, w' denotes the transpose of a row vector w, and E_ϑ means expectation under P_ϑ. In view of Lemma 2.1, we obtain that assumption (A1) of Henze (1994) holds.

Denoting by $\nabla_\vartheta F(k,\vartheta^*)=\left(\frac{\partial}{\partial\lambda}F(k,\vartheta^*),\frac{\partial}{\partial\xi}F(k,\vartheta^*)\right)$ the vector of partial derivatives of $F(k,\vartheta)$ defined in (2.1) with respect to λ and ξ, evaluated at ϑ^*, and writing $\|\cdot\|_2$ for the Euclidean norm in $\mathbb{R}^2$, the next result shows that assumption (A2) of Henze (1994) holds. The proof is straightforward and therefore omitted.

Lemma 2.2:
For fixed $k\geq 0$, $\nabla_\vartheta F(k,\vartheta)$ is a continuous function of ϑ. Moreover,

$$\lim_{k\to\infty}\sup_{\vartheta^*\in\mathcal{U}(\vartheta)}\|\nabla_\vartheta F(k,\vartheta^*)\|_2=0,$$

where $\mathcal{U}(\vartheta)$ is a sufficiently small neighborhood of ϑ.

By Theorem 3.1 of Henze (1994), there is a centered Gaussian sequence $\mathcal{W}=(W_k)_{k\geq 0}$ in c_0 such that, under P_ϑ, the estimated discrete empirical process $\mathcal{Z}_n$ defined in (2.5) converges weakly to $\mathcal{W}$ in the space c_0. The covariance function of $\mathcal{W}$ depends on ϑ, but there is little point in recording the algebraic details.

As a consequence (see Corollary 3.4 of Henze (1994)), we obtain that the limit behavior of the test statistics K_n and C_n defined in (2.3), (2.4) under P_ϑ is given by

$$K_n \xrightarrow{\mathcal{D}} \sup_{k\geq 0} |W_k|,$$

$$C_n \xrightarrow{\mathcal{D}} \sum_{k=0}^{\infty} W_k^2 \ (F(k,\vartheta) - F(k-1,\vartheta)).$$

Moreover, the modified Cramér – von Mises statistic

$$C_n^* = n \sum_{k=0}^{\infty} (F_n(k) - F(k,\hat{\vartheta}_n))^2 [F_n(k) - F_n(k-1)] \tag{2.11}$$

which may also be used for testing H_0 has the same limiting null distribution as C_n.

To compute the statistics K_n, C_n and C_n^* in practice, note that

$$K_n = \sqrt{n} \max_{0\leq k\leq M} \left| F_n(k) - F(k,\hat{\vartheta}_n) \right|$$

and

$$C_n^* = n \sum_{k=0}^{M} (F_n(k) - F(k,\hat{\vartheta}_n))^2 (F_n(k) - F_n(k-1)),$$

where $M = \max_{1\leq j\leq n} X_j$. The infinite series representing C_n may be truncated at some suffiently large value $l \geq M$ since

$$\sum_{k=l+1}^{\infty} \left(F_n(k) - F(k,\hat{\vartheta}_n) \right)^2 [F(k,\hat{\vartheta}_n) - F(k-1,\hat{\vartheta}_n)] \leq [1 - F(l,\hat{\vartheta}_n)]^3.$$

Observe that, under H_0, both the finite sample and the asymptotic distributions of K_n, C_n and C_n^* depend on the unknown "true" value of ϑ. To perform a GOF test for H_0 based on K_n, C_n or C_n^*, we suggest a *parametric bootstrap*, i. e., estimating the critical value from the data $X_1, \ldots, X_n$. To be precise, let $T_n = T_n(X_1, \ldots, X_n)$ denote any of the test statistics K_n, C_n or C_n^*, and let $H_{n,\vartheta}(t) = P_\vartheta(T_n \leq t)$ be the df of the null distribution of T_n when ϑ is the "true" parameter value. Then a natural critical value for T_n would be the $(1-\alpha)$–quantile of $H_{n,\hat{\vartheta}_n}$. Since the latter is difficult to calculate, it will be estimated by the following Monte Carlo procedure which requires the generation of pseudo random numbers from a GPD distribution.

Given $X_1, \ldots, X_n$, first compute $\hat{\vartheta}_n = \hat{\vartheta}_n(X_1, \ldots, X_n) = (\hat{\lambda}_n, \hat{\xi}_n)$. Then, conditionally on $X_1, \ldots, X_n$, let $X_{j1}^*, \ldots, X_{jn}^*$, $1 \leq j \leq k_n$, be independent and identically distributed random variables with the $GPD(\hat{\lambda}_n, \hat{\xi}_n)$ distribution and compute $T^*_{jn} = T_n(X^*_{j1}, \ldots, X^*_{jn})$, $1 \le j \le k_n$. Note that, to compute T^*_{jn}, parameter estimation has to be done for each j separately. Writing

$$H^*_n(t) = \frac{1}{k_n} \sum_{j=1}^{k_n} 1\{T^*_{jn} \le t\} \qquad (2.12)$$

for the empirical df of $T^*_{1,n}, \ldots, T^*_{k_n n}$, the $(1-\alpha)$–quantile $c^*_{n\alpha} = (H^*_n)^{-1}(1-\alpha)$ of H^*_n is given by

$$c^*_{n,\alpha} = \begin{cases} T^*_{[k_n(1-\alpha)]:k_n}, & \text{if } k_n(1-\alpha) \text{ is an integer} \\ T^*_{[k_n(1-\alpha)]+1:k_n}, & \text{otherwise,} \end{cases} \qquad (2.13)$$

where $T^*_{1:k_n} \le T^*_{2:k_n} \le \ldots \le T^*_{k_n:k_n}$ are the order statistics of $T^*_{1n}, \ldots, T^*_{k_n,n}$. The hypothesis H_0 is rejected at level α if T_n exceeds $c^*_{n,\alpha}$.

Since $\lim_{n\to\infty} \hat{\vartheta}_n = \vartheta$ P_ϑ–almost surely, Theorem 3.6 of Henze (1994) yields

$$\lim P_\vartheta(T_n > c^*_{n\alpha}) = \alpha \quad \text{as } n, k_n \to \infty.$$

This shows that the parametric bootstrap versions of the tests based on K_n, C_n or C^*_n have asymptotic level α.

The consistency of the bootstrap tests based on K_n, C_n or C^*_n against any fixed nonnegative integer–valued distribution F having finite positive variance larger than the mean follows from Remark 3.7 of Henze (1994) since

$$\inf_{\vartheta\in\Theta} \sup_{k\ge 0} \left| F(k) - F(k, \vartheta) \right| > 0$$

provided that the distribution function F does not belong to the class *GPDU*.

3 Simulations and data analysis

To gain some insight into the actual level of the bootstrap test based on K_n or C_n, a Monte Carlo experiment was performed for sample sizes $n = 15, 25, 50$ and the nominal levels of significance $\alpha = 0.1$ and $\alpha = 0.05$

The bootstrap sample size k_n was taken to be $\max(n, [1/\alpha])$ and, as a slight amendment of (2.12), the critical value was taken to be

$$\bar{c}_{n,\alpha} = T^*_{\alpha_n:k_n} + (1-\gamma_n)\left(T^*_{\alpha_n+1:k_n} - T^*_{\alpha_n:k_n}\right), \tag{3.1}$$

where $\alpha_n = k_n - [\alpha(k_n+1)]$, $\gamma_n = \alpha(k_n+1) - [\alpha(k_n+1)]$ (see also Baringhaus and Henze, 1992).

Each entry in Table 3.1 is the *estimated actual level* (*percentage of* rejections of H_0) of the Kolmogorov–Smirnov test based on 5000 Monte Carlo samples for the nominal level $\alpha = 0.1$ and a wide range of values for $\vartheta = (\lambda, \xi)$. The results show that the actual level is indeed very close to the nominal level even for a sample of size 15. Although not covered by our theoretical derivations, we included parameter values from the "Poisson line" (i.e., $\xi = 0$). In practice, a Poisson distribution, representing a member on the boundary of the *GPDU* model, should not be rejected, and this is clearly supported by the simulations. The results of Table 3.2 for the Cramér – von Mises test are completely similar. Since the impression of the close agreement between actual and nominal level of significance is the same for $\alpha = 0.05$, it will not be reproduced here. Moreover, the behavior of the modified Cramér– von Mises statistic is very similar to that of C_n.

We conducted a small simulation study for sample sizes $n = 25$ and $n = 50$ in order to assess the power of several GOF tests for the *GPD* model. Tables 3.3 and 3.4 show estimated powers at the 0.05 level of significance. Each number represents the percentage of 5000 Monte Carlo replications declared to be significant by the different tests. In all cases, parameter estimation was done by the method of moments.

The following procedures were compared:

(i) The tests K_n, C_n and C_n^* as described above with $k_n = n$ and the critical value given by (3.1).

(ii) The χ^2–test. As it is common use, we approximated the distribution of this test statistic by a χ^2_{k-3} distribution, where k denotes the number of classes. As a criterion for cell selection we used the minimum expected frequencies (MEF) criterion. This is fairly objective, easy to implement, and frequently used in practice. In Tables 3.3 and 3.4

Table 3.1: Empirical level for the Kolmogorov–Smirnov test based on 5000 Monte Carlo replications, $\alpha = 0.1$

$\alpha = 0.1$		(λ, ξ)					
n	k_n	(1,0)	(3,0)	(5,0)	(10,0)	(25,0)	(50,0)
15	15	0.0800	0.0986	0.0932	0.0886	0.0902	0.0886
25	25	0.0996	0.1050	0.1018	0.1030	0.0998	0.1026
50	50	0.1044	0.1008	0.1096	0.1056	0.1080	0.1080

$\alpha = 0.1$		(λ, ξ)					
n	k_n	(1,0.25)	(3,0.25)	(5,0.25)	(10,0.25)	(25,0.25)	(50,0.25)
15	15	0.0990	0.0900	0.0942	0.0914	0.0930	0.0854
25	25	0.1022	0.1012	0.0968	0.0976	0.0942	0.0974
50	50	0.1040	0.0994	0.0952	0.0960	0.1000	0.1038

$\alpha = 0.1$		(λ, ξ)					
n	k_n	(1,0.5)	(3,0.5)	(5,0.5)	(10,0.5)	(25,0.5)	(50,0.5)
15	15	0.0976	0.0882	0.0904	0.0872	0.0918	0.0884
25	25	0.1136	0.0938	0.1012	0.0998	0.0978	0.1008
50	50	0.1078	0.0958	0.1110	0.1030	0.0956	0.1048

we included the results for MEF = 1 (denoted by $\chi^2_{n,1}$) and MEF = 3 (denoted by $\chi^2_{n,3}$). A hyphen indicates that the tests could not be carried out since, in many cases, the number of cells was less than 4.

(iii) The test of Mirvaliev mentioned in the introduction. The formal description of this procedure would take too much space and is therefore omitted. Under H_0, the test statistic, denoted by M_n, has an asym-

Table 3.2: Empirical level for the Cramér – von Mises test based on 5000 Monte Carlo replications, $\alpha = 0.1$

$\alpha = 0.1$		(λ, ξ)					
n	k_n	(1,0)	(3,0)	(5,0)	(10,0)	(25,0)	(50,0)
15	15	0.0820	0.0918	0.0894	0.0876	0.0930	0.0850
25	25	0.1034	0.9820	0.1024	0.0958	0.1012	0.1022
50	50	0.1100	0.1052	0.1130	0.1058	0.1036	0.1102

$\alpha = 0.1$		(λ, ξ)					
n	k_n	(1,0.25)	(3,0.25)	(5,0.25)	(10,0.25)	(25,0.25)	(50,0.25)
15	15	0.0906	0.0832	0.0882	0.0766	0.0836	0.0844
25	25	0.0986	0.0952	0.0892	0.0882	0.0942	0.0930
50	50	0.1042	0.0986	0.0962	0.0986	0.0962	0.0988

$\alpha = 0.1$		(λ, ξ)					
n	k_n	(1,0.5)	(3,0.5)	(5,0.5)	(10,0.5)	(25,0.5)	(50,0.5)
15	15	0.1098	0.0828	0.0870	0.0828	0.0916	0.0884
25	25	0.1142	0.0924	0.0906	0.0994	0.0954	0.0940
50	50	0.1074	0.0962	0.1066	0.0990	0.0956	0.1018

ptotic χ^2_{k-1} distribution. Since the results varied little for MEF = 1,2,3, only the case MEF = 2 was included in Tables 3.3 and 3.4.

The first three rows in the table show the actual level of the tests for different *GPD* distributions.

The alternative distributions selected in the study are the uniform distribution on $\{0, \ldots, 8\}$ and $\{0, \ldots, 20\}$, a mixture of Poisson distributions, two

Table 3.3: Percentage of 5000 Monte Carlo samples declared significant by various test for the *GPD* model; $\alpha = 0.05$; $n = 25$

	K_n	C_n	C_n^*	$\chi^2_{n,1}$	$\chi^2_{n,3}$	M_n
$gpd(1, 0.5)$	4.8	4.4	5.0	6.6	—	3.3
$gpd(5, 0.5)$	3.8	4.1	3.8	5.5	6.5	4.3
$gpd(5, 0.1)$	4.5	4.3	4.0	4.8	6.6	4.1
$U(0 \ldots 8)$	31	36	27	49	17	43
$U(0 \ldots 20)$	39	47	38	42	23	50
$\frac{1}{2}[P(2) + P(10)]$	40	47	44	34	41	27
$\frac{1}{2}[gpd(2, 0.2) + gpd(10, 0.2)]$	34	42	42	27	34	26
$\frac{1}{2}[gpd(2, 0.5) + gpd(10, 0.1)]$	31	37	32	28	27	29
$N_A(0.8, 1.5)$	18	21	21	—	—	10
$N_A(0.8, 3)$	65	61	71	50	—	38

Table 3.4: Percentage of 5000 Monte Carlo samples declared significant by various test for the *GPD* model; $\alpha = 0.05$; $n = 50$

	K_n	C_n	C_n^*	$\chi^2_{n,1}$	$\chi^2_{n,3}$	M_n
$gpd(1, 0.5)$	5.3	5.6	5.6	7.1	8.8	4.1
$gpd(5, 0.5)$	4.6	4.4	4.5	5.8	5.8	4.4
$gpd(5, 0.1)$	4.4	4.3	4.3	5.4	5.6	4.7
$U(0 \ldots 8)$	64	75	69	89	51	87
$U(0 \ldots 20)$	76	86	83	88	59	89
$\frac{1}{2}[P(2) + P(10)]$	77	86	85	69	62	63
$\frac{1}{2}[gpd(2, 0.2) + gpd(10, 0.2)]$	68	79	79	58	51	51
$\frac{1}{2}[gpd(2, 0.5) + gpd(10, 0.1)]$	65	74	71	62	52	61
$N_A(0.8, 1.5)$	36	41	42	32	—	21
$N_A(0.8, 3)$	94	92	95	83	92	78

mixtures of *GPD* distributions and two Neyman Type A distributions, where the first is L–shaped and the second is bimodal. For the definition of the $N_A(\lambda, \phi)$ distribution, see Johnson, Kotz and Kemp, 1992, p. 368.

It should be mentioned that many other standard distributions with variance greater than the mean are not easily detected by any of the tests since the *GPD* model is quite flexible. Especially, the *GPD* and the negative binomial distribution allow of nearly identical shapes.

The main conclusions that can be drawn from the simulation results are the following:

(1) For K_n, C_n, C_n^* and M_n, there is a good agreement between actual and nominal level of significance; in most cases, these tests are somewhat conservative. The χ^2 test is anticonservative; in some cases (see $\chi^2_{n,3}$ in the case $gpd(1, 0.5), n = 50$) the actual level is far above the nominal one.

(2) Among the empirical distribution function tests, C_n and C_n^* behave very similar. In most cases they perform better than K_n.

(3) The power of the χ^2 test depends strongly on the (chosen) criterion. In some cases, $\chi^2_{n,1}$ performs better than $\chi^2_{n,3}$; in other cases, the contrary holds.

(4) In most cases, C_n and C_n^* perform best. An exception is the case of the uniform distribution, where $\chi^2_{n,1}$ or M_n perform best, but $\chi^2_{n,3}$ has the lowest power.

(5) These conclusions are valid for $n = 25$ as well as for $n = 50$.

Based on these preliminary results, we clearly recommend the use of C_n or C_n^* as test statistics for the GOF of the *GPD* model if no special alternative has been established. Besides of being powerful, these statistics avoid the problem of cell selection inherent in the use of the χ^2–test and thus the danger of manipulation of P–values.

As already mentioned above, the motivation to study the problem of testing the goodness–of–fit for the *GPD* distribution originated from work of Bárdossy and Plate (1992) who developed a flexible space-time model for daily precipitation. A crucial constituent part of this model is a set $\{a_1, \ldots, a_k\}$ of k possible atmospheric circulation patterns following a Semi-Markov process. The random duration of a_i is described by a Generalized Poisson distribution (augmented by 1) with a parameter depending only on i and the season, but independent of all other random variables.

We applied our tests to data consisting of observed durations of circulation patterns in Central Europe for the period 1951–1989. The latter are classified according to the scheme of the German Weather Service which distinguishes between 29 different circulation patterns, for simplicity numbered from 1 to 29 in what follows.

For example, number 1 stands for "West anticyclonic" and number 2 for "West cyclonic". The "West" circulation patterns are characterized by a stationary high pressure area over the Azores and a low pressure area between Iceland and Scandinavia.

A small part of our results is given in Table 3.5 for the atmospheric circulation patterns no. 1,2,3,8,9 and 10, each analysed separately for the seasons spring (sp), summer (su), autumn (au) and winter (wi).

In each case, n is the sample size (observed cases of the pertinent atmospheric circulation pattern (acp) within the selected season), and $\hat{\lambda}, \hat{\xi}$ are the M-estimators for λ resp. ξ. The bootstrap sample size k_n was always taken to be 499 in order to have 500 samples altogether (including the original one).

To obtain an impression of the approximate p-level of an observed value of the test statistics, the entry m_K gives the position

$$m_K = 1 + \sum_{j=1}^{k_n} \mathbf{1}\{K_{j,n}^* < K_n\}$$

of K_n within the set of the bootstrap sample values. Recall that H_0 is rejected at level α for K_n if $m_K > [k_n(1-\alpha)] + 1$. In the same way,

$$m_C = 1 + \sum_{j=1}^{k_n} \mathbf{1}\{C_{j,n}^* < C_n\}$$

denotes the position of the Cramér – von Mises statistic C_n within the set $C_{1,n}^*, \ldots, C_{k_n,n}^*$ of bootstrap values for C_n.

Table 3.5: Test results for selected atmospheric circulation patterns

	acp 1				acp 2				acp 3			
	sp	su	au	wi	sp	su	au	wi	sp	su	au	wi
n	25	72	51	32	73	102	102	113	25	12	16	37
$\hat{\lambda}$	4.2	3.2	2.5	3.0	2.7	2.6	3.3	3.4	2.6	3.3	2.7	2.6
$\hat{\xi}$	-.50	.03	.23	.30	.33	.34	.27	.23	.03	.46	.27	.43
m_C	379	421	417	365	265	469	200	113	484	457	480	491
m_K	425	451	373	301	316	476	267	114	483	455	496	495

	acp 8				acp 9				acp 10			
	sp	su	au	wi	sp	su	au	wi	sp	su	au	wi
n	47	45	43	52	48	67	81	64	59	77	80	80
$\hat{\lambda}$	2.7	2.9	2.1	2.6	1.7	2.2	2.7	2.0	2.3	2.7	2.5	2.7
$\hat{\xi}$	.00	.14	.19	.19	.32	.10	.16	.30	.28	.22	.29	.01
m_C	474	435	315	489	215	499	234	204	490	491	497	197
m_K	423	427	405	487	171	497	202	216	484	455	497	141

It should be mentioned that all calculations were done on a workstation in FORTRAN 77. Whenever possible, we used standard routines of the NAG library. A program implementing the tests under discussion may be obtained upon request.

The results demonstrate that, roughly speaking, the Generalized Poisson distribution provides an adequate working model for the duration of atmospheric circulation patterns in slightly more than a half of the 24 cases considered. At the 10% level, rejection of H_0 is in 11 cases both for K_n and C_n. For some circulation patterns, e.g. no. 3 ("Southern West"), there

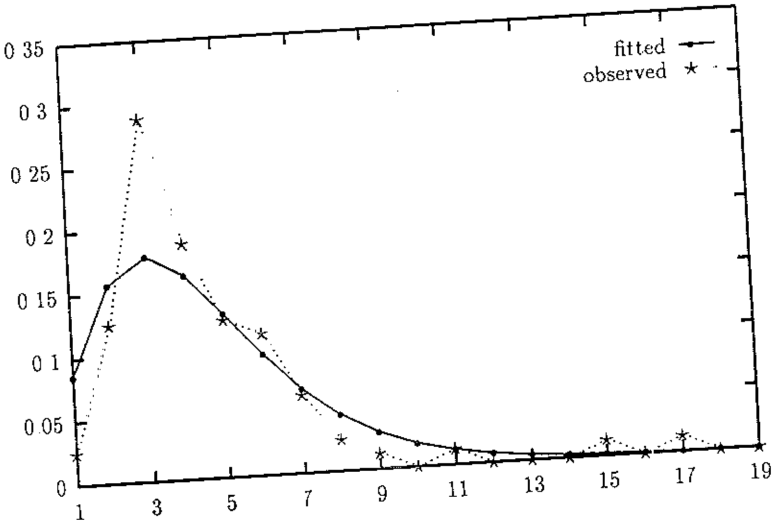

Figure 3 1: Fitted and observed frequencies of autumn data for acp 10

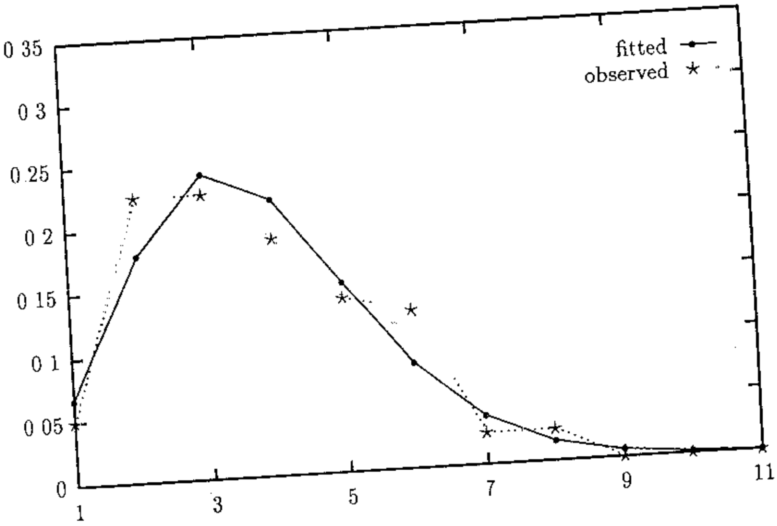

Figure 3 2: Fitted and observed frequencies of winter data for acp 10 is strong evidence against the *GPD* model so that alternatives should be envisaged.

It is interesting to note that, within a fixed circulation pattern, the goodness-of-fit for the *GPD* model may strongly depend on the season of the year. As an example, consider the acp no. 10 ("Central European ridge"). Here, the *GPD* model is strongly rejected for the "autumn data", but the "winter data" fit the *GPD* model very closely.

This fact is complemented by a visual inspection of polygons of fitted (solid lines) and observed (dashed lines) frequencies for the "autumn data" (Fig. 3.1) and "winter data" (Fig. 3.2), giving an indication of why the autumn data strongly reject the *GPD* model.

BIBLIOGRAPHY

Alzaid, A.A. and Al-Osh, M.A. (1993). "Some autoregressive moving average processes with generalized Poisson marginal distributions", *Ann. Inst. Statist. Math.* 45, 223–232.

Bárdossy, A., and Plate, E.J. (1992). "Space-Time model for daily rainfall using atmospheric circulation patterns", *Water Resources Res.* 28, 1247-1259.

Baringhaus, L. and Henze, N. (1992). "A goodness of fit test for the Poisson distribution based on the empirical generating function", *Statist. & Prob. Letters* 13, 269–274.

Burke, M.D., Csörgő, M., Csörgő, S. and Révész, P. (1979). "Strong approximations of the empirical process when parameters are estimated", *Ann. Probab.* 7, 790–810.

Consul, P.C. and Jain, G.C. (1973). "A Generalization of the Poisson Distribution", *Technometrics* 15, 791–799.

Consul, P.C. and Shoukri, M.M. (1984). "Maximum likelihood estimation for the generalized Poisson distribution", *Commun. Statist. A – Theor. Meth.* 13, 1533–1547.

Consul, P C and Shoukri, M M (1985). "The Generalized Poisson Distribution when the sample mean is larger than the sample variance", *Commun. Statist. B – Simul. Computa* 14, 667–681

Consul, P C (1989). *Generalized Poisson Distributions Properties and applications,* STATISTICS: textbooks and monographs, Vol. 99, Marcel Dekker, New York.

D'Agostino, R B. and Stephens, M.A. (eds.) (1986). *Goodness-of-fit techniques*, STATISTICS: textbooks and monographs, Vol 68, Marcel Dekker, New York.

Fisher, R.A. (1925) *Statistical methods for research workers* (14th ed 1970), Edinburgh: Oliver and Boyd

Haight, F.A. (1967). *Handbook of the Poisson Distribution*, J. Wiley & Sons, New York.

Henze, N. (1994). "Empirical distribution function goodness of fit tests for discrete models", (submitted)

Janardan, K G and Schaeffer, D.J (1977). "Models for the analysis of chromosomal aberrations in human leucocytes", *Biometrical Journal* 19, 599–612.

Jensen, J L W. (1902). "Sur identité d'Abel et pour d'autres formules analogues", *Acta Math* 26, 307–318.

Johnson, N L , Kotz, S. and Kemp, A W (1992) *Univariate discrete distributions* (2nd ed.), J. Wiley & Sons, New York

Mirvaliev, M (1987). "The χ^2–test for the Generalized Poisson distribution when parameters are estimated by the method of moments" (in Russian) *Akademija Nauk Uzbekskoj SSR (Taskent): UZSSR Fanlar Akademijasiming dokladlari Taskent*, 3–5

Nakamura, M and Pérez–Abreu, V (1993) "Use of an empirical generating function for testing a Poisson model", *Canad Journ Statist.* 21, 149–156.

Rueda, R , Pérez–Abreu, V. and O'Reilly, F (1991). "Goodness of fit for the Poisson distribution based on the empirical probability generating function", *Commun. Stat. A – Theor. Meth.* 20, 3093–3110

Received March, 1994; Revised March, 1995.