EJASA:DSS – Electronic Journal of Applied Statistical Analysis: Decision Support Systems and Services Evaluation
http://siba-ese.unisalento.it/index.php/ejasa_dss/index
e-ISSN: 2037-3627
DOI: 10.1285/i2037-3627v4n1p9

A goodness-of-fit test for maximum order statistics from discrete distributions

By Silvia Facchinetti, Silvia Angela Osmetti

Published: 28 December 2013

Electronic Journal of Applied Statistical Analysis: Decision Support Systems and Services Evaluation
Vol. 04, Issue 01, 2013, 9-21
DOI: 10.1285/i2037-3627v4n1p9

A goodness-of-fit test for maximum order statistics from discrete distributions

Silvia Facchinetti *and Silvia Angela Osmetti

Università Cattolica del Sacro Cuore, Dipartimento di Scienze statistiche, Milano, Italy

Published: 28 December 2013

In economic, financial and environmental sciences studies the extreme value theory is used for the evaluation of several complex occurring phenomena, e.g., risk management theory, natural calamities, meteorology and pollution studies. When the observed values are discrete, like count measurements, the discrete extreme value distributions should be applied. In this paper we propose a procedure to evaluate the goodness of fit of extreme values from discrete distributions. In particular we modify the classic statistic of the Kolmogorov-Smirnov goodness of fit test for continuous distribution function. This modification is necessary since the assumption of the Kolmogorov-Smirnov test is the continuity of the distribution specified under the null hypothesis. The distribution of the proposed test is given. The exact critical values of the test statistic are tabulated for extreme values from some specific discrete distributions. An application in environmental science is presented.

keywords: Kolmogorov-Smirnov Goodness of Fit Test, Order statistics, Discrete Distributions.

1 Introduction

One of the most important goodness-of-fit test was suggested by Kolmogorov and Smirnov (see Kolmogorov (1933) and Smirnov (1939)).
Let's consider the following hypotheses:

*Corresponding Author: silvia.facchinetti@unicatt.it

ISSN: 2037-3627
http://siba-ese.unisalento.it/index.php/ejasadss/index

$$\begin{cases} H_0 : F(x) = F_0(x) & \text{for every } x \\ H_1 : F(x) \neq F_0(x) & \text{for some } x \end{cases} \tag{1}$$

where $F(x)$ is the true cumulative distribution function.
The test is based on a comparison between the cumulative distribution function $F_0(x)$ and the empirical one $S_n(x)$:

$$D_n = \sup_{-\infty < x < \infty} |S_n(x) - F_0(x)|. \tag{2}$$

In particular, let X be a continuous random variable with distribution function $F(x)$, and let $(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ be an ordered simple random sample of size n from X. The empirical distribution function is a step function with jumps occurring at the sample values defined as follows:

$$S_n(x) = \begin{cases} 0 & \text{for } x < x_{(1)} \\ k/n & \text{for } x_{(k)} \leq x < x_{(k+1)} \text{ with } k = 1, 2, \ldots, n-1. \\ 1 & \text{for } x \geq x_{(n)} \end{cases} \tag{3}$$

The critical region of size α to reject the null hypothesis in (1) is defined by:

$$R = \left\{ D_n : D_n > \frac{d_\alpha}{\sqrt{n}} \right\}$$

where d_α depends only on α. For a continuous random variable X, the Kolmogorov-Smirnov test is distribution-free and its exact and approximate critical values have been calculated. If the continuity assumption of X is not satisfied, the probability distribution of D_n depends on $F_0(x)$ and the test is not distribution-free. Several authors have studied the Kolmogorov-Smirnov test for discrete random variables (Kolmogorov (1941), Noether (1963), Conover (1972), Pettitt and Stephens (1977)).
In this paper we propose a modification of the Kolmogorov-Smirnov test that can be applied to discrete extreme values. Many of the most significant events in several areas are extreme events or rare events, in economics and finance, in medicine and epidemiology, in meteorology and natural science. In economics and finance, some applications of extreme value theory are financial strategy of risk management, Value at Risk and credit risk (Embrechts et al. (1997), Dahan and Mendelson (2001), Barro (1998)). In natural science and in epidemiology the extreme or rare events, as natural disasters and the epidemics, occur frequently but they are considered for great important (Frei and Schar (1998)). The methodology for modeling continuous extreme values or rare events is well established. When the observed values are discrete, like count measurements, we apply the discrete extreme value distributions. The binomial and Poisson distributions are generally used to model the occurrence and frequency of these events (Falk et al. (2010)). In this paper we consider the discrete extreme value distributions because of the lack of exhaustive literature about these distributions compared with the continuous ones (Balakrishnan and Rao (1998)).

The paper is organized as follows. The next section explains the characteristics of order statistic from discrete random variable, with particular reference to the the maximum order statistic. In section 3 the Kolmogorov-Smirnov test applied to discrete extreme random variables with bounded domain is proposed, the test statistic and its distribution are presented. In subsection 3.1 we extended the proposed procedure to distributions with unbounded domain. Section 4 presents the exact critical values of the proposed test statistic for two particular discrete random variables: extreme values from the uniform-discrete distribution and the binomial one. Finally, in section 5, we apply our proposal to empirical data in environmental science field for an illustrative purpose.

2 Distribution of the order statistic from discrete random variables

Let X be a discrete random variable that assumes the values x_i for $i = 1, 2, \ldots, k$ with cumulative distribution function $F(x_i)$ and probability distribution function $p(x_i)$. Let $\underline{X}_{(\cdot)} = (X_{(1)}, \ldots, X_{(r)}, \ldots, X_{(n)})$ be the random vector of order statistics obtained by a random sample with replacement of size n from X. The generic order statistic $X_{(r)}$ for $r = 1, 2, \ldots, n$ is a discrete random variable that assumes values in the finite set $S_X = \{x_i : i = 1, 2, \ldots, k\}$ with cumulative distribution function:

$$F_r(x) = \begin{cases} 0 & \text{if } x < x_1 \\ \sum_{i=1}^{h} p_r(x_i) & \text{if } x_h \leq x < x_{h+1} \qquad h = 1, 2, \ldots, k-1 \\ 1 & \text{if } x \geq x_k \end{cases} \tag{4}$$

where $p_r(x_i) = F_r(x_i) - F_r(x_{i-1})$ $x_i \in S_X$ is the probability distribution function of the order statistic $X_{(r)}$ calculated in x_i. The expression (4) may also appear as a function of the cumulative distribution function $F(x)$ of the random variable X:

$$F_r(x) = \sum_{j=r}^{n} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}. \tag{5}$$

This distribution function can be associated with an incomplete Beta function, defined as follows:

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x u^{a-1}(1-u)^{b-1} du \qquad 0 \leq x \leq 1$$

where $B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$, with $a, b > 0$, is the beta function.
Thus the expression (5) can be written:

$$F_r(x) = \frac{1}{B(r, n-r+1)} \int_0^{F(x)} u^{r-1}(1-u)^{n-r} du = I_{F(x)}(r, n-r+1). \tag{6}$$

It is well known that the probability integral transformations $F_r(X_r)$ of each order statistic is distribution-free. In particular, (5) assumes the form of a survival function

of a binomial random variable with parameters $(n, F(x))$ evaluated in r.
For $r = n$ we consider the maximum order statistic $X_{(n)}$, with cumulative distribution function and probability distribution function respectively given by:

$$F_n(x) = [F(x)]^n \tag{7}$$

$$p_n(x_i) = F_n(x_i) - F_n(x_{i-1}) = [F(x_i)]^n - [F(x_{i-1})]^n \qquad x_i \in S_X. \tag{8}$$

3 The methodological proposal: the goodness of fit test for discrete extreme values

Let $X_{(n)}$ be the maximum order statistic from a discrete random variable X with bounded domain and let $(x_1^*, \ldots, x_s^*, \ldots, x_m^*)$ be an ordered random sample with replacement of size m from $X_{(n)}$ with values not necessarily distinct. Let $y_1, \ldots, y_j, \ldots, y_q$ be the different ordered values observed in the sample, with $x_1 \le x_1^* = y_1 < ... < y_q = x_m^* \le x_k$, and let r_j be the number of observations equal to y_j, for $j = 1, \ldots, q$.
The empirical cumulative distribution function of the maximum order statistic is:

$$S_{(n)m}(x) = \begin{cases} 0 & \text{if } x < y_1 \\ \frac{1}{m}\sum_{j=1}^{h} r_j & \text{if } y_h \le x < y_{h+1} \qquad h = 1, 2, ..., q-1 \\ 1 & \text{if } x \le y_q \end{cases} \tag{9}$$

with $r_j \in \mathbf{N}^+$ and $\sum_{j=1}^{q} r_j = m$. This is a function with $q \le k$ steps of height equal to $\frac{r_j}{m}$, for $j = 1, 2, ..., q$.
Let's now adapt the Kolmogorov-Smirnov test to evaluate the goodness of fit of the maximum discrete order statistics to a previously specified random variable. In this context the assumption of continuity is not satisfied, and the classic Kolmogorov-Smirnov test is no longer applicable.
The test is about the null hypothesis

$$H_0 : F_n(x) = F_0(x), \tag{10}$$

where $F_0(x)$ is the true cumulative distribution function of the maximum discrete order statistic. The test statistic is based on the differences between the true and the empirical cumulative distribution function in (9):

$$\Delta_h = F_n(x_h) - S_{(n)m}(x_h) = \sum_{i=1}^{h} p_n(x_i) - \frac{1}{m}\sum_{i=1}^{h} r_i \qquad \text{with} \qquad h = 1, 2, \ldots, k.$$

The calculations are formally extended to all $x_i \in S_X$ values, observed or not.
In our proposal the Kolmogorov-Smirnov statistic becomes:

$$\Delta_m = \max_{h=1,\ldots,k} |\Delta_h| = \max_{h=1,2,\ldots,k} |\ [F(x_h)]^n - S_{(n)m}(x_h)\ |. \tag{11}$$

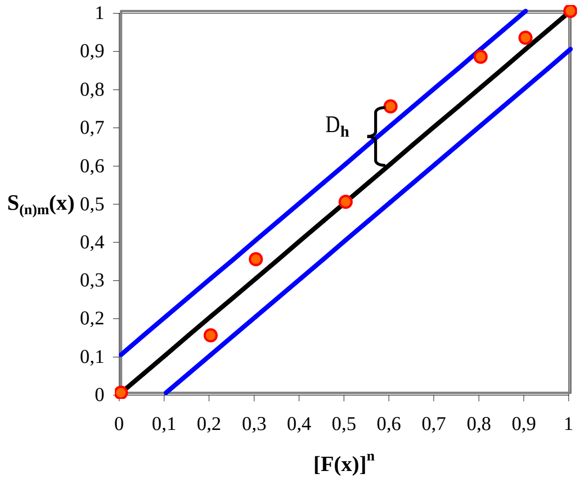

Figure 1: The empirical cumulative distribution function versus the true one.

The new statistic Δ_m is not distribution-free with respect to the classic Kolmogorov-Smirnov statistic for a continuous random variable. In fact the distribution of Δ_m depends on the sample size m, on the order parameter n and on the distribution of the discrete random variable X.
Since the distribution of the order statistic $X_{(r)}$ depends on X only through the cumulative distribution function $F(x)$, Δ_m in (11) depends only on $F(x)$, for fixed n and m. For a fixed specified size α, we don't reject the null hypothesis if $\Delta_m \leq d_\alpha$, or equivalently:

$$F_0(x_h) - d_\alpha \leq S_{(n)m}(x_h) \leq F_0(x_h) + d_\alpha \qquad \forall h = 1, 2, \dots, k \tag{12}$$

otherwise we reject the null hypothesis if $\exists h$ such that at least one difference Δ_h falls outside the region defined above:

$$S_m(x_h) < [F(x_h)]^n - d_\alpha \quad or \quad S_m(x_h) > [F(x_h)]^n + d_\alpha.$$

In order to apply the test and to define the critical values, we need to specify the distribution of the test statistic under the null hypothesis H_0:

$$F_{\Delta_m}(d) = P(\Delta_m \leq d | H_0). \tag{13}$$

This distribution is determined by considering all sets of points $(F_0(x_h), S_{(n)m}(x_h))$, for $h = 1, 2, \dots, k$ that meet the condition (12), i.e. those are included into the region defined in Figure 1.
The figure shows one possible trajectory of points and one of the differences Δ_h between the empirical cumulative distribution function and the true one. These differences are the vertical distances between the points and the bisector, where the two functions $F_0(x)$ and $S_{(n)m}(x)$ overlap.
For the case of extreme value statistic, once the distribution law of the random variable

$X_{(n)}$ has been defined, we define the cumulative distribution function of Δ_m under H_0 as follow:

$$F_{\Delta_m}(d) = \sum_{r_h \in \mathcal{H}} P(\underline{r}) \tag{14}$$

where

$$P(\underline{r}) = \frac{m!}{\left(\prod_{h=1}^{k-1} r_h!\right)\left(n - \sum_{h=1}^{k-1} r_h\right)!} \prod_{h=1}^{k-1} [F_0(x_h) - F_0(x_{h-1})]^{r_h} (1 - F_0(x_{k-1}))^{m - \sum_{h=1}^{k-1} r_h} \tag{15}$$

is the probability of occurrence of the different possible trajectories $\underline{r} = \{r_1, r_2, ..., r_k\}$ that meet the condition (12). In particular $P(\underline{r})$ is a multinomial probability distribution function.
In (14) $\mathcal{H}$ is the set

$$\mathcal{H} = \left\{ \bigcup_{h=1}^{k-1} \{\max(0, \underline{r}_h) \leq \sum_{i=1}^{h} r_i \leq \min(m, \overline{r}_h)\} \right\}, \tag{16}$$

where $[\underline{r}_h/m, \overline{r}_h/m]$ is $\forall h$ the interval of values of the empirical cumulative distribution function $S_{(n)m}(x_h)$ such that the points $(F_0(x), S_{(n)m}(x_h))$ be in accordance with (12). In particular

$$\underline{r}_h = \lfloor m[F(x_h)]^n + d \rfloor \qquad \text{and} \qquad \overline{r}_h = \lfloor m[[F(x_h)]^n - d] \rfloor + 1. \tag{17}$$

The cumulative distribution function (14) allow us to calculate the exact critical values of the test. It should be noted that the distribution in (15) depends on the distribution function of the random variable X from which the order statistics derive, on the sample size n from which depends the order of the maximum order statistic and on the number of replications of the sample m. Therefore, for a fixed α, the critical values of the test statistic depend on the distribution law of the random variable from which the extreme values are derived.
The proposed test assumes that the null hypothesis is simple and the null hypothesis completely specifies the distribution of the population. The goodness of fit test procedure may be extend to a composite null hypothesis in which the distribution of the population belongs to some parametric family distribution:

$$H_0 : F_n(x) = F_0(x; \theta) : \theta \in \Theta.$$

The test statistics then becomes $\max_{h=1,...,k} |F_0(x_h, \widehat{\theta}) - S_{(n)m}(x_h)|$, where $\widehat{\theta}$ is an estimator of θ. The distribution of such test statistic is a conditional distribution respect to the $\widehat{\theta}$ value and it has the same functional form of the one defined in (14). The test statistics is again not distribution free.

3.1 Distribution with a countable number of jumps

We now extend the proposed procedure to discrete distributions with unbounded domain. For example the Poisson distribution is usually applied to model a rare event. For these distributions we can apply the proposed procedure, using the distribution defined in (14), with a modification of the set $\mathcal{H}$ in (16). In particular we define

$$\underline{r}_{k^*-1} = \lim_{k\to\infty} \underline{r}_{k-1}$$

and we suppose that $\exists\epsilon \simeq 0^+$ such that

$$\underline{r}_{k^*-1} \geq \lfloor m(F^* - \varepsilon) - d\rfloor + 1,$$

where $F^* = \lim_{h\to\infty} F(x_h) = 1$. Therefore, $\exists\epsilon \simeq 0^+$ such that the distribution of the test statistic under the null hypothesis becomes

$$F_{\Delta_m}(d) = \sum_{r_h \in \mathcal{H}^*} \left\{ \frac{m!}{\left(\prod_{h=1}^{k^*-1} r_h!\right)\left(n - \sum_{h=1}^{k^*-1} r_h\right)!} \prod_{h=1}^{k^*-1} [F_0(x_h) - F_0(x_{h-1})]^{r_h} (1 - F_0(x_{k^*-1}))^{r_h^*} \right\} \tag{18}$$

where $r_h^* = m - \sum_{h=1}^{k^*-1} r_h$ and

$$\mathcal{H}^* = \bigcup_{h=1}^{k^*-1} \{\max(0, \underline{r}_h) \leq \sum_{i=1}^{h} r_i \leq \min(m, \overline{r}_h)\}. \tag{19}$$

In particular for $h = (k^* - 1)$ we have $\underline{r}_{k^*-1} \leq \sum_{i=1}^{k^*-1} r_i \leq m$.
The problem is solved by using on the left side argument of the $\mathcal{H}^*$, analogously it can be solved by the use of the right side one.

4 Exact calculation of the critical values

We apply the proposed test to the maximum order statistic from two particular discrete distributions: the uniform and the binomial. We need to calculate the probability distribution of the test statistic Δ_m from (14).

Discrete uniform random variable. Let X be a discrete uniform random variable that assumes values $(1, 2, \ldots, k)$ with probability $P(X = x) = 1/k\ \forall x = 1, 2, \ldots, k$. Let $X_{(n)}$ be the maximum order statistic from X, with cumulative distribution function

$$F_n(x) = \begin{cases} 0 & x < 1 \\ (h/k)^n & h \leq x < h+1 \quad h = 1, \ldots, k-1. \\ 1 & x \geq k \end{cases} \tag{20}$$

Table 1 and table 2 show the values of the cumulative distribution function of the test statistic for several values of the parameters k, n and the critical value d, for $m = 5$ and $m = 10$ sample sizes, respectively.

Table 1: Exact distribution of Δ_m statistic for a discrete uniform random variable with $m = 5$

n	k	d=0.1	d=0.2	d=0.3	d=0.4	d=0.5
2	5	0.05225	0.36791	0.70592	0.86363	0.99064
	10	0.00448	0.20190	0.56987	0.82122	0.94789
	15	0.00048	0.13698	0.51884	0.80253	0.93055
5	5	0.18981	0.64269	0.77911	0.95534	0.99239
	10	0.03812	0.39110	0.67587	0.87491	0.97889
	15	0.00525	0.25661	0.54541	0.88029	0.93655
10	5	0.32164	0.90753	0.98927	0.98953	0.99939
	10	0.10551	0.61781	0.81959	0.94146	0.99425
	15	0.03205	0.42394	0.69637	0.92304	0.95551
15	5	0.83603	0.98847	0.98847	0.99959	0.99959
	10	0.33942	0.61546	0.93126	0.99249	0.99249
	15	0.10440	0.60858	0.81795	0.93635	0.99342
20	5	0.94367	0.99870	0.99870	0.99998	0.99998
	10	0.32762	0.88496	0.98425	0.98515	0.99900
	15	0.30500	0.64632	0.88108	0.98295	0.98295

From table 1 and table 2 we note that the probability to accept the null hypothesis depends on the parameter k characterizing the distribution and on the rank n. Indeed, for a fixed value of n, the probability $1 - \alpha$ increases with an increase of the parameter k. Moreover, for fixed values of d and k this probability increase with n, the rank of the maximum order statistic. Finally, since the bandwidth in figure 1 increase with the increase of d, this probability has an increment with the critical value d. For $m = 10$ the probabilities $1 - \alpha$ are higher than the once obtained for $m = 5$.

Binomial random variable. Let now X be a binomial random variable with parameters l and p. Let $X_{(n)}$ be the maximum order statistic from X, with cumulative distribution

Table 2: Exact distribution of Δ_m statistic for a discrete uniform random variable with $m = 10$

n	k	d=0.1	d=0.2	d=0.3	d=0.4	d=0.5
2	5	0,18443	0,66327	0,91484	0,98680	0,99845
	10	0,05713	0,51864	0,86053	0,97416	0,99702
	15	0,02683	0,48484	0,85563	0,97297	0,99706
5	5	0,39470	0,79307	0,95928	0,99655	0,99966
	10	0,15824	0,63556	0,90670	0,98813	0,99929
	15	0,07114	0,60779	0,87777	0,97847	0,99760
10	5	0,59474	0,98361	0,99774	0,99978	0,99999
	10	0,31513	0,81117	0,95948	0,99531	0,99947
	15	0,18159	0,68950	0,92084	0,98588	0,99870
15	5	0,95383	0,99566	0,99973	0,99999	1,00000
	10	0,48524	0,86074	0,99241	0,99895	0,99990
	15	0,31954	0,80667	0,95730	0,99439	0,99937
20	5	0,99438	0,99983	1,00000	1,00000	1,00000
	10	0,61228	0,97491	0,99606	0,99956	0,99997
	15	0,48421	0,85616	0,97859	0,99627	0,99956

function

$$F_n(x) = \begin{cases} 0 & x < 0 \\ \left[\sum_{i=0}^{x} \begin{pmatrix} l \\ i \end{pmatrix} p^i (1-p)^{l-i}\right]^n & h \leq x < h+1 \quad h = 0, \ldots, l-1. \\ 1 & x \geq l \end{cases} \tag{21}$$

Table 3 and table 4 show the values of the cumulative distribution function of the test statistic for several values of the distribution parameters l, p and n and of the critical value d. We consider the sample sizes $m = 5$ and $m = 10$.

We observe that the probability to accept the null hypothesis depends on the parameters l and p characterizing the distribution and on the rank n like the results obtained the maximum order statistic from a discrete uniform random variable. Moreover the probability $1 - \alpha$ increase with the sample size m and with the critical value d.
In conclusion in both cases, we observe that the probability to accept the null hypothesis depends on the parameters that characterize the distributions in analysis, as well as the distribution function of the random variable X from which come the order statistics. Furthermore, the probability increases with the sample size m and the critical value d.

Table 3: Exact distribution of Δ_m statistic for a binomial random variable with $m = 5$

n	l	p	d=0.1	d=0.2	d=0.3	d=0.4	d=0.5
2	5	0.1	0.12445	0.52410	0.80925	0.91693	0.99220
		0.5	0.08399	0.44748	0.72847	0.93455	0.97768
	10	0.1	0.04575	0.47728	0.79239	0.90501	0.97781
		0.5	0.01748	0.36524	0.69152	0.86884	0.98619
5	5	0.1	0.18233	0.61561	0.78363	0.94447	0.99109
		0.5	0.11434	0.53030	0.80860	0.91846	0.99213
	10	0.1	0.08546	0.43504	0.73535	0.95709	0.98630
		0.5	0.06739	0.41264	0.70367	0.90237	0.96939
10	5	0.1	0.21000	0.57431	0.78053	0.92077	0.98079
		0.5	0.14810	0.59159	0.85571	0.96239	0.97747
	10	0.1	0.09375	0.55244	0.77223	0.92465	0.97198
		0.5	0.04552	0.47233	0.80311	0.89723	0.98262
15	5	0.1	0.13725	0.59650	0.84793	0.96164	0.97532
		0.5	0.27536	0.61307	0.82182	0.92738	0.99137
	10	0.1	0.11498	0.50636	0.79534	0.91264	0.99154
		0.5	0.09852	0.50694	0.78751	0.90425	0.98147
20	5	0.1	0.15584	0.61177	0.92328	0.92378	0.99251
		0.5	0.30044	0.61942	0.80385	0.93523	0.97702
	10	0.1	0.14155	0.42192	0.82784	0.97688	0.97688
		0.5	0.11052	0.48556	0.78012	0.90518	0.99022

5 Empirical evidence

We now apply of the proposed goodness of fit test in environmental sciences field: the PM10 (Particulate Matter) air pollution study.
A data set of daily average concentrations of PM10 in $\mu g/m^3$ in Sondrio (Italy) comes from ARPA is used. We consider the number of days per week for which the PM10 level are greater than the safety threshold observed in January from the year 2000 to the year 2009. The PM10 level recommended is 50 micro grams per a cubic meter of air ($50\mu g/m^3$) measured as a 24 hour running average. We consider the PM10 concentration only in the month of January because the winter shows the highest values with respect to summer months. The observations follow a binomial distribution with parameters l (number of weekly detection days in which the PM10 level is greater than the safety threshold) and p (probability to overcome the safety threshold). The value of p is estimated from daily historical observations. We consider the maximum of the $n = 4$ values observed in the

Table 4: Exact distribution of Δ_m statistic for a binomial random variable with $m = 10$

n	l	p	d=0.1	d=0.2	d=0.3	d=0.4	d=0.5
2	5	0,1	0,29419	0,77605	0,94993	0,99358	0,99930
		0,5	0,23746	0,71170	0,94287	0,99244	0,99916
	10	0,1	0,19872	0,76039	0,94079	0,99152	0,99949
		0,5	0,12050	0,66100	0,91183	0,98697	0,99840
5	5	0,1	0,39009	0,79190	0,95725	0,99517	0,99949
		0,5	0,30561	0,78103	0,95113	0,99353	0,99929
	10	0,1	0,21832	0,69729	0,95213	0,99663	0,99969
		0,5	0,19842	0,66990	0,92120	0,98566	0,99434
10	5	0,1	0,38214	0,76518	0,94315	0,99264	0,99975
		0,5	0,33140	0,82748	0,96571	0,99360	0,99918
	10	0,1	0,28000	0,76374	0,93833	0,99011	0,99928
		0,5	0,19456	0,75502	0,93931	0,99178	0,99963
15	5	0,1	0,32657	0,82325	0,96220	0,99264	0,99902
		0,5	0,44592	0,79946	0,95017	0,99153	0,99894
	10	0,1	0,30253	0,75824	0,94510	0,99320	0,99925
		0,5	0,27096	0,75209	0,93955	0,99190	0,99964
20	5	0,1	0,18584	0,84965	0,96496	0,99768	0,99933
		0,5	0,46403	0,79345	0,94616	0,99179	0,99947
	10	0,1	0,27902	0,76174	0,97370	0,99579	0,99955
		0,5	0,28634	0,73510	0,93676	0,99170	0,99928

four weeks of January for the 10 considered years.
Our aim is to evaluate the distribution of the maximum of the observed values, thus we test the null hypothesis $H_0 : F_n(x) = F_0(x)$, where $F_0(x)$ is the cumulative distribution function of the maximum with rank $n = 4$ from the binomial random variable in (21) with parameters $l = 7$ and $p = 0.8107$. Considering a sample of size $m = 10$, we calculate the empirical cumulative distribution function and we obtain a value of the test statistic in equation (11) equal to 0.1512 and a p-value equal to 0.3711. For a specified $1-\alpha = 96\%$, we obtain a critical value $d_\alpha = 0.3$ (> 0.1512). Therefore we accept the null hypothesis that the observed values come from the maximum of the binomial random variable with parameters $l = 7$ and $p = 0.8107$.

6 Conclusions

In this paper we propose a procedure to evaluate the goodness of fit of an order statistic from discrete distributions. In particular, we modify the classic statistic of Kolmogorov-Smirnov test in order to apply it to extreme values from discrete distributions. We find the cumulative distribution function of the test statistic under the null hypothesis H_0 and we observe that the statistic is not distribution-free with respect to the continuous random variable case. Its distribution depends both on the one defined under H_0 as well as the sample size m. Moreover, as the order statistics are distribution-free, the distribution of the test statistic Δ_m depends on that of the random variable X from which the order statistics are derived. These results have been verified applying the procedure to extreme values from discrete uniform and binomial distributions. The exact values of the cumulative distribution function for finite sample sizes are given. This test is applicable in several areas (economic, financial and environmental sciences) and has a general validity in the theory of extreme values, particularly for those complex phenomena in which the values observed are discrete. In particular we have illustrated an example of the application of the proposed procedure to a study on the PM10 air pollution.

References

Balakrishnan, N. and Rao, C. (1998). *Order Statistics: Theory and Methods.* Handbook of Statistics, 16.

Barro, R. (1998). Rare disasters, asset prices, and welfare costs. *American Economic Review*, 99(1):243–264.

Conover, W. (1972). A kolmogorov goodness-of-fit test for discontinuous distributions. *Journal of the American Statistical Association*, 67:591–596.

Dahan, E. and Mendelson, H. (2001). An extreme-value model of concept testing. *Management Science*, 47:102–116.

Embrechts, P., Klppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer Verlag.

Falk, M., Hsler, J., and Reiss, R. (2010). *Laws of Small Numbers: Extreme and Rare Events.* Springer.

Frei, C. and Schar, C. (1998). A prepicitations climatology of the alps from high-resolution rain-gauge observations. *Int. J. Climatol.*, 18:873–900.

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4:83–91.

Kolmogorov, A. (1941). Confidence limits for an unknown distribution function. *The Annals of Mathematical Statistics*, 12(4):461–463.

Noether, G. (1963). Note on the kolmogorov statistic in the discrete case. *Metrika*, 7:115–116.

Pettitt, A. and Stephens, M. (1977). The kolmogorov-smirnov goodness of fit statistic with discrete and grouped data. *Technometrics*, 19:205–210.

Smirnov, V. (1939). Sur les écarts de la courbe de distribution empirique. *Recueil Mathematique*, 19:3–26.