Computational Statistics and Data Analysis 78 (2014) 21–32

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Statistical inference for the geometric distribution based on δ-records

Raúl Gouet [a], F. Javier López [c], Lina P. Maldonado [b], Gerardo Sanz [c,*]

[a] Dpto. Ingeniería Matemática y Centro de Modelamiento Matemático (UMI 2807, CNRS), Universidad de Chile, Chile

[b] Dpto. Métodos Estadísticos, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain

[c] Dpto. Métodos Estadísticos and BIFI, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain

ARTICLE INFO

Article history:
Received 1 August 2013
Received in revised form 2 April 2014
Accepted 4 April 2014
Available online 12 April 2014

Keywords:
δ-records
Geometric distribution
Maximum likelihood estimation
Bayes estimation
Empirical Bayes
Maximum likelihood prediction
Bayes prediction

ABSTRACT

New inferential procedures for the geometric distribution, based on δ-records, are developed. Maximum likelihood and Bayesian approaches for parameter estimation and prediction of future records are considered. The performance of the estimators is compared with those based solely on record-breaking data by means of Monte Carlo simulations, concluding that the use of δ-records is clearly advantageous. An example using real data is also discussed.

1. Introduction

Records are quite ubiquitous and interesting objects per se. They are naturally encountered in sports but also in climatology, seismology, finance, insurance, etc. Their mathematical theory has been under development for several decades and reached maturity, as can be seen in Ahsanullah (1995); Arnold et al. (1998) and Nevzorov (2001). In parallel but somewhat later, the theory of statistical inference based on record-breaking data began to develop and rapidly attained a good level of sophistication; see Gulati and Padgett (2003). A good reason for exploring record-based inference procedures is that record-breaking data are readily available in many situations. Consider, for example, the standard experimental setup of destructive stress-testing, where the data consists of (lower) records only; see Glick (1978) for an account of this issue. This strategy provides valuable information for estimating population quantiles, say, at a fraction of the measurement costs of classical sampling.

The concept of δ-record was introduced in Gouet et al. (2007) as a natural generalization of classical records, simple enough to allow for rigorous analysis of mathematical properties, such as the asymptotic behaviour of their counting process; see also Gouet et al. (2012). Loosely speaking, a δ-record is an observation which is either a record or falls short of being one (see below for definitions) and, as such, it is not surprising that we consider them as candidates for upgrading the so-called inference from record-breaking data. In fact, the destructive stress-testing setup mentioned above needs only a

* Correspondence to: Dpto. Métodos Estadísticos, Facultad de Ciencias, Universidad de Zaragoza, C/ Pedro Cerbuna, 12, 50009 Zaragoza, Spain. Tel.: +34 976762860; fax: +34 976762860.

E-mail addresses: rgouet@dim.uchile.cl (R. Gouet), javier.lopez@unizar.es (F.J. López), lmguaje@unizar.es (L.P. Maldonado), gerardo.sanz@unizar.es (G. Sanz).

http://dx.doi.org/10.1016/j.csda.2014.04.002
0167-9473/

minor adjustment to allow for the measurement of δ-record data. Of course, the total cost will increase as more items are destroyed but the type of additional data obtained, close to records, is likely to enhance the performance of procedures originally designed to accept only records as input.

A first example showing the usefulness of δ-records can be found in Gouet et al. (2012). Preliminary results are encouraging but much remains to be done, both in terms of theory and applications. This paper is meant to contribute in that direction, in the context of an important discrete model such as the geometric distribution.

Let $(X_n)_{n\geq 1}$ be a random sequence, $M_n = \max\{X_1, \ldots, X_n\}$, for $n \geq 1$, and $M_0 = -\infty$. Observation X_n is a record if it is greater than all previous observations, that is, if $X_n > M_{n-1}$. On the other hand, X_n is a δ-record if $X_n > M_{n-1} + \delta$, where δ is a fixed real parameter. If $\delta > 0$ every δ-record is a record but obviously not all records are δ-records. So, in this case, δ-records are even scarcer than ordinary records and clearly not adequate to replace records in inference procedures. The opposite situation is observed if $\delta < 0$ because every record is a δ-record and additionally, non-record observations within distance $-\delta$ of the current record are also δ-records. In this case δ-records correspond to records together with near-records, as defined in Balakrishnan et al. (2005). For discrete distributions, δ-records with $\delta = -1$ correspond to weak records, which have been largely analysed in the literature; see Castaño Martínez et al. (2013), Gouet et al. (2008) and Hashorva and Stepanov (2012) for recent developments in the study of weak records. Since we observe more δ-records than records we can expect better performance of δ-record-based inference than inference based only on records. Accordingly, Gouet et al. (2012) consider maximum likelihood estimation for continuous distributions, in particular exponential and Weibull distributions, showing that δ-records-based estimators outperform those based only on records. Also, López-Blázquez and Salamanca-Miño (2013) analyse, in Section 7, properties of maximum likelihood (ML) estimation based on δ-records in the exponential distribution.

The literature on record-based statistical inference for discrete models is rather scarce, when compared to that for continuous ones. Interesting references are Stepanov et al. (2003), dealing with the Fisher information contained in records, and Doostparast and Ahmadi (2006), on the statistical analysis of the geometric distribution, both from Bayesian and non-Bayesian viewpoints.

The aim of this paper is to assess the usefulness of δ-records by following a path similar to that of Doostparast and Ahmadi (2006), namely developing new (δ-record-based) point and interval estimators of the parameter and predictors of future records, for the geometric distribution, in both frequentist and Bayesian inferential frameworks. Monte Carlo simulations clearly show superior performance of procedures using δ-records, in all instances considered.

The paper is organized as follows. In Section 2 we introduce the probabilistic framework, define δ-records and related random variables and then focus on the estimation of parameter p. We consider maximum likelihood (in Section 2.2), Bayes and empirical Bayes estimation (in Section 2.3). The prediction of future records is addressed in Section 3, with maximum likelihood framework in Section 3.1 and Bayesian in Section 3.2. Conclusions and some ideas for future work are presented in Section 4.

2. Parameter estimation

2.1. Probabilistic framework and definitions

We consider a sequence $(X_n)_{n\geq 1}$ of independent, geometrically distributed random variables, with parameter $p \in (0, 1)$. That is, $P(k) := P[X_1 = k] = pq^{k-1}$, for $k = 1, 2, \ldots$, with $q := 1 - p$. Let $(M_n)_{n\geq 1}$ denote the sequence of partial maxima, with $M_n = \max\{X_1, \ldots, X_n\}$, for $n \geq 1$, $M_0 = -\infty$ and let $(R_n)_{n\geq 1}$ be the sequence of record values, obtained from partial maxima by discarding repetitions. Record times L_n, $n \geq 1$, are defined as $L_1 = 1$ and $L_n = \min\{m > L_{n-1}; X_m > X_{L_{n-1}}\}$, $n \geq 2$. Then, clearly $R_n = X_{L_n}$, for $n \geq 1$.

For δ-records we consider a fixed, negative-integer-valued parameter δ, because the random variables X_n are integer valued. However, it is convenient to include 0 as possible value of δ in order to see records as particular case of δ-records. Observe that in this context weak records, satisfying $X_n \geq M_{n-1}$, are δ-records with $\delta = -1$.

Recall that X_1 is a δ-record by convention and that X_n is a δ-record if $X_n > M_{n-1} + \delta$, $n \geq 2$. We say that X_n is a δ-record associated to the mth record R_m if $X_n > M_{n-1} + \delta = R_m + \delta$ and $L_m \leq n < L_{m+1}$. The number of δ-records associated to R_m, not counting R_m itself, is denoted by S_m and the vector of δ-records associated to R_m (excluding R_m) by $(Y_m^1, \ldots, Y_m^{S_m})$. Observe that $0 \leq S_m < L_{m+1} - L_m$ and also that $R_m + \delta < Y_m^j \leq R_m$, for $j = 1, \ldots, S_m$.

Proposition 2.1. *Let $n \geq 1$ and $\delta \leq -1$, then,*

(i) *conditional on $R_i = r_i$, S_i has geometric distribution (starting at 0) with success parameter $p_i := \bar{F}(r_i)/\bar{F}(r_i + \delta)$, for $i = 1, \ldots, n$, where $\bar{F}(x) := 1 - F(x)$ is the survival function of X_1.*

(ii) *Conditional on $R_i = r_i$, $S_i = s_i$, the random variables $Y_i^1, \ldots, Y_i^{S_i}$ are independent, with common (conditional) probability mass function*

$$P_i(k) := \frac{P(k)}{\bar{F}(r_i + \delta) - \bar{F}(r_i)}, \quad k = r_i + \delta + 1, \ldots, r_i.$$

Proof. (i) Suppose $R_i = r_i$, then S_i counts the number of independent observations X_j, with $j > L_i$, such that $X_j > r_i + \delta$ until one observation is greater than r_i (success). Clearly, the probability of success is $P[X_j > r_i | X_j > r_i + \delta] = p_i$ (given $R_i = r_i$).

(ii) Given that among $X_{L_i+1}, \ldots, X_{L_{i+1}-1}$ there are s_i observations X_j with value in $(r_i + \delta, r_i]$, each one takes the value k with probability $P[X_j = k | X_j \in (r_i + \delta, r_i]] = P_i(k)$. □

Corollary 2.2. *Let $n \geq -\delta \geq 1$, then*

(i) *the random variables S_i, $i = -\delta, \ldots, n$, are independent and geometrically distributed (starting at 0), with success parameter $q^{-\delta}$ and*

(ii) *conditional on $S_1, \ldots, S_n$, the random variables $Z_i^j := R_i - Y_i^j$, $i = -\delta, \ldots, n; j = 1, \ldots, S_i$, are independent with probability-mass function $P[Z_i^j = k] = \frac{p}{1-q^{-\delta}} q^{-\delta-1-k}$, for $0 \leq k < -\delta$, and expectation $E[Z_i^j] = -\left(\frac{1}{p} + \frac{\delta q^{\delta}}{q^{\delta}-1}\right)$.*

Proof. (i) If $i \geq -\delta$ then $R_i \geq -\delta$ because the random variables are integer valued, with $R_1 = X_1 \geq 1$. So $\bar{F}(r_i + \delta) = q^{r_i+\delta}$ and $p_i = \bar{F}(r_i)/\bar{F}(r_i + \delta) = q^{-\delta}$.

(ii) Same argument as in part (i). □

We present below the likelihood function of the sample $\mathbf{T} := (\mathbf{R}, \mathbf{S}, \mathbf{Y})$, where $\mathbf{R} = (R_1, \ldots, R_n)$, $\mathbf{S} = (S_1, \ldots, S_n)$ and $\mathbf{Y} = (Y_1^1, \ldots, Y_1^{S_1}, \ldots, Y_n^1, \ldots, Y_n^{S_n})$.

Proposition 2.3. *The likelihood function of* $\mathbf{T}$ *is given by*

$$\mathcal{L}(\mathbf{t}, p) = \mathcal{L}(p) = \prod_{i=1}^{n} \left(\frac{q^{r_i-1} p}{q^{(r_i+\delta)^+ (s_i+1)}} \prod_{j=1}^{s_i} q^{y_i^j - 1} p \right) q^{r_n}, \tag{1}$$

with $\mathbf{t} = (\mathbf{r}, \mathbf{s}, \mathbf{y})$, $0 < r_1 < \cdots < r_n$, $s_i \geq 0$ and $r_i + \delta < y_i^j \leq r_i$, for $j = 1, \ldots, s_i$, $i = 1, \ldots, n$, and $(r_i + \delta)^+ = \max\{r_i + \delta, 0\}$.

Proof. From Proposition 2.1 and knowing that, conditional on R_i, the number and values of its associated δ-records are independent of the values of R_j, $j > i$, and of the number and values of their associated δ-records, the likelihood function can be written as

$$\mathcal{L}(\mathbf{t}, p) = \bar{F}(r_n) \prod_{i=1}^{n} \frac{P(r_i)}{\bar{F}(r_i + \delta)^{s_i+1}} \prod_{j=1}^{s_i} P(y_i^j),$$

and (1) follows. □

Observe that the sample corresponding to the likelihood function (1) consists of the n first record values $R_1, \ldots, R_n$ and, for each record value R_i, the number S_i and the values $Y_i^1, \ldots, Y_i^{S_i}$ of δ-records associated to R_i. Note that $R_1, Y_1^1, \ldots, Y_1^{S_1}, \ldots, R_n, Y_n^1, \ldots, Y_n^{S_n}$ are (all) the δ-records observed before but not including record R_{n+1}.

Throughout the paper, random variables are denoted by uppercase letters and random vectors, by uppercase boldfaced letters. We introduce below two frequently used random variables. Let

$$A = \sum_{i=1}^{n} S_i + n \quad \text{and} \quad B = \sum_{i=1}^{n} R_i + \sum_{i=1}^{n} \sum_{j=1}^{S_i} Y_i^j - \sum_{i=1}^{n} (S_i + 1)(R_i + \delta)^+ + R_n - A. \tag{2}$$

Let us also define the corresponding lowercase versions of A and B, to be used as arguments in probability functions.

$$a = \sum_{i=1}^{n} s_i + n \quad \text{and} \quad b = \sum_{i=1}^{n} r_i + \sum_{i=1}^{n} \sum_{j=1}^{s_i} y_i^j - \sum_{i=1}^{n} (s_i + 1)(r_i + \delta)^+ + r_n - a. \tag{3}$$

2.2. Maximum likelihood estimation

In order to derive the maximum likelihood estimator (MLE) of p from the sample $(\mathbf{R}, \mathbf{S}, \mathbf{Y})$, we take logarithms in (4) and differentiate with respect to p, yielding the MLE of p

$$\widehat{p}_{\delta} = \frac{A}{A + B}, \tag{4}$$

with A, B defined in (2).

Observe that, for $\delta = 0$, formula (4) simplifies to $\widehat{p}_0 = n/R_n$, which is the well-known estimator of p based on records only. On the other hand, for $\delta = -1$ (weak records), we have $(R_i + \delta)^+ = R_i - 1$ since the observations take positive integer values. Also, $Y_i^j = R_i$, for $j = 1, \ldots, S_i$ and so, the denominator in (4) simplifies to $n + \sum_{i=1}^{n} S_i + R_n = A + R_n$. Finally we have

$$\widehat{p}_{-1} = \frac{A}{A + R_n}. \tag{5}$$

Table 1
EMSE of the MLE of parameter p in the geometric distribution from 10^4 simulation runs, with $p = 0.5$ and 0.7.

p	n	EMSE($\widehat{p}_\delta$)		
		$\delta = 0$	$\delta = -1$	$\delta = -3$
0.5	5	3.14E−02	1.10E−02	4.07E−03
	10	1.36E−02	5.81E−03	1.82E−03
0.7	5	2.81E−02	8.79E−03	1.66E−03
	10	1.48E−02	4.58E−03	5.40E−04

Note that in this case A coincides with the number of weak-records. It is interesting to see that $\widehat{p}_0$ only uses the last piece of information contained in $(R_1, \ldots, R_n)$ but $\widehat{p}_{-1}$ relies on extra information provided by the number of ties for the current maximum ($\sum_{i=1}^n S_i$), although ignoring the actual values of $R_1, \ldots, R_{n-1}$.

In order to assess the performance of the MLE of p, we carried out 10^4 simulation runs for several values of p, n and δ and the respective estimated mean square errors (EMSE) of $\widehat{p}_\delta$ were computed. Results are presented in Table 1. A quick inspection of Table 1 shows a steep decrease of the EMSE as $|\delta|$ increases. In particular, it can be seen that, for $p = 0.5$ and the three values of n, the EMSE decreases by more than 50% when passing from records ($\delta = 0$) to weak records ($\delta = -1$). The situation is even better for $p = 0.7$, with a 70% decrease of the EMSE. It is interesting to observe that the extra information contained in $\sum_{i=1}^n S_i$ has a significant impact on the EMSE. Also note that the EMSE, for $n = 10$, $p = 0.5$ and $\delta = 0$ is 0.0136 while, for $n = 5$, $p = 0.5$ and $\delta = -1$, we have 0.011. This means that we make slightly better inferences about p with just 5 records and their weak-records than with 10 records alone. This is interesting in terms of cost because in order to have twice as many records, from 5 to 10, we must increase the number of observations by a factor of roughly 150.

We now study some analytical properties of $\widehat{p}_\delta$. It is known that the set of record values $\{R_n; n \geq 1\}$ from the geometric distribution behaves as a Bernoulli point process, with constant success probability p. This implies that R_n is distributed as the sum of n independent and identically distributed (iid) geometric random variables, with parameter p; see Arnold et al. (1998) and Nevzorov (2001). So, we immediately obtain from the strong law of large numbers that $n/R_n \to p$ a.s., showing the well-known fact that $\widehat{p}_0$ is strongly consistent. In the context of δ-records we show that $\widehat{p}_\delta$ with $\delta \leq -1$ is also strongly consistent.

Proposition 2.4. *Let $\delta \leq -1$ and $\widehat{p}_\delta$ be defined by (4). Then $\widehat{p}_\delta \to p$ a.s., as $n \to \infty$. Moreover, $\widehat{p}_\delta$ is asymptotically unbiased.*

Proof. Observe that since $R_i > -\delta$ for $i > -\delta$, and so $(R_i + \delta)^+ = R_i + \delta$, consistency of $\widehat{p}_\delta$ is equivalent to that of

$$\frac{A}{\sum_{i=1}^n R_i + \sum_{i=1}^n \sum_{j=1}^{S_i} Y_i^j - \sum_{i=1}^n (S_i + 1)(R_i + \delta) + R_n} = \frac{A}{-\sum_{i=1}^n \sum_{j=1}^{S_i} Z_i^j - \delta A + R_n}, \tag{6}$$

where $Z_i^j = R_i - Y_i^j$. From Corollary 2.2 and the law of large numbers, we have $A \sim n(1 + q^\delta(1 - q^{-\delta})) = nq^\delta$ (the notation $a_n \sim b_n$ stands for $a_n/b_n \to 1$ a.s). Also, from the representation of R_n as sum of iid geometric random variables, we have $R_n \sim np^{-1}$. Finally, from Corollary 2.2, Wald's identity and the law of large numbers, we obtain

$$\sum_{i=1}^n \sum_{j=1}^{S_i} Z_i^j \sim -n(q^\delta - 1)\left(\frac{1}{p} + \frac{\delta q^\delta}{q^\delta - 1}\right) = -n\left(\frac{q^\delta - 1}{p} + \delta q^\delta\right).$$

Collecting partial results we have that the r.h.s. of (6) converges a.s. to

$$\frac{q^\delta}{\left(\frac{q^\delta - 1}{p} + \delta q^\delta\right) - \delta q^\delta + 1/p} = p, \tag{7}$$

which proves the first assertion. The second assertion follows from (7) and the dominated convergence theorem, since $\widehat{p}_\delta \leq 1$. □

A non-asymptotic general result for the expectation of $\widehat{p}_\delta$ seems intractable. However, for illustrative purposes, in the case $\delta = -1$ (weak records) we derive an "almost closed-form" expression.

Proposition 2.5. *Let $\widehat{p}_{-1}$ be defined by (5). Then*

$$e(p) := E[\widehat{p}_{-1}] = \sum_{k,l=n}^{\infty} \binom{k-1}{n-1} \binom{l-1}{n-1} \frac{k}{k+l} p^k q^l. \tag{8}$$

Proof. From (5) we have

$$\widehat{p}_{-1} = \frac{\sum_{i=1}^{n}(S_i + 1)}{\sum_{i=1}^{n}(S_i + 1) + R_n}.$$

On the other hand, from Corollary 2.2, $A = \sum_{i=1}^{n}(S_i + 1)$ is distributed as the sum of iid geometric random variables, with parameter q, starting at 1. Hence, A has a Pascal (negative binomial) distribution with $P[A = k] = \binom{k-1}{n-1} q^n p^{k-n}, \ k \geq n$.

Moreover, R_n is independent of A and distributed as the sum of iid geometric random variables, with parameter p, starting at 1. Hence, $P[R_n = l] = \binom{l-1}{n-1} p^n q^{l-n}, \ l \geq n$ and formula (8) follows. □

Proposition 2.5 tells that $\widehat{p}_{-1}$ is biased. It can be shown that $e(p)$ increases with p and has a fixed point at $p = 1/2$. A numerical analysis reveals also that $e(p) > p$, for $p \in (0, 1/2)$, and therefore, $e(p) < p$, for $p \in (1/2, 1)$, because of the identity $e(p) + e(1 - p) = 1$.

Other properties of $\widehat{p}_{\delta}$, such as admissibility, could be considered. This is done in Bagrezaei et al. (2012) in the context of record-based methods. It would be interesting to carry out a similar analysis here but this may be arduous, due to the complex form of $\widehat{p}_{\delta}$.

2.3. Bayesian framework

Bayesian inference is a convenient method to be used with record-breaking data. Indeed, given that records are so scarce, prior information is welcome. There are a number of papers on Bayesian inference using records, mostly for continuous distributions; see chapter 5 of Arnold et al. (1998) and, more recently, the works by Jaheen (2003); Raqab et al. (2007); Soliman et al. (2006) for continuous distributions and Doostparast and Ahmadi (2006) for the geometric distribution.

In the following we develop a Bayesian estimation procedure for p, based on δ-records. As in Doostparast and Ahmadi (2006), we take the beta distribution, with parameters α, β, as prior of p. From the likelihood function (1), we readily obtain the posterior distribution $\pi(p|\mathbf{t})$, which is also beta, with parameters $\alpha + a$ and $\beta + b$.

Under the squared error loss function, the Bayes estimator of p, say $\widehat{p}_{\pi,\delta}$, is given by the posterior mean, namely

$$\widehat{p}_{\pi,\delta} = \frac{A + \alpha}{A + B + \alpha + \beta}. \tag{9}$$

Of course, for $\delta = 0$ we recover the estimator given in Doostparast and Ahmadi (2006), while for $\delta = -1$ (weak records) we obtain

$$\widehat{p}_{\pi,-1} = \frac{A + \alpha}{A + R_n + \alpha + \beta},$$

where A coincides with the number of weak records.

As in the previous section we evaluate the performance of $\widehat{p}_{\pi,\delta}$ for $\delta = 0, -1, -3$. We proceed as follows: for fixed (α, β) we simulate 10^4 values of p from the prior Beta(α, β); for each value of p we simulate a random sample of n records and their associated δ-records, for different values of δ, and compute $\widehat{p}_{\pi,\delta}$. The EMSE is obtained as the average of $(\widehat{p}_{\pi,\delta} - p)^2$ along the 10^4 simulation runs. The results, displayed in columns 1–3 of Table 2, reveal a pattern similar to the one observed for the MLE in Table 1. That is, δ-records also bring about a significant reduction of the quadratic error in the Bayesian framework. Furthermore, using the same data we compute $\widehat{p}_{\delta}$ (MLE) and its EMSE, which are displayed in columns 4–6 of Table 2. This can serve as a basis for the comparison of non-Bayesian and Bayesian estimates, similar to what is done in Soliman et al. (2006) for the Weibull distribution, using record values. As seen in Table 2, the EMSE of $\widehat{p}_{\delta}$ is greater than that of $\widehat{p}_{\pi,\delta}$, for all values of α and β considered, showing that the use of correct prior information is beneficial in the estimation of p.

Credible intervals for p are obtained from the posterior distribution. The extremes l, u of a $100(1-\alpha)\%$ credible interval are such that $\int_l^u \pi(p|\mathbf{t})dp = 1 - \alpha$. We numerically compute shortest (HPD) intervals for p. We use the interval width as a measure of performance and rely again on simulations to compute the average widths of intervals, for several values of δ, α, β and n. Results are displayed in Table 3 and, as for point estimation, a clear improvement is observed when the information of δ-records is incorporated in the construction of the intervals.

We also consider empirical Bayes (EB) estimation of p, which uses the Bayesian paradigm but does not have an explicit form of the prior. Instead, a number of previous (past) samples are available and used to estimate the prior. When the prior is known up to a k-dimensional (hyper)parameter, the procedure is termed parametric empirical Bayes; see Maritz and Lwin (1995).

Doostparast and Ahmadi (2006) use a beta prior, with unknown hyperparameters α and β, which they estimate by the method of moments, given m past samples of the first n records. We also use the beta prior here, for the sake of comparability, but our m past samples are made of δ-records. Furthermore, another important difference is that we adopt restricted

Table 2
EMSE of $\widehat{p}_{\pi,\delta}$ and $\widehat{p}_{\delta}$ for different values of δ, α, β and n, from 10 000 simulation runs.

(α,β)	n	EMSE($\widehat{p}_{\pi,\delta}$)			EMSE($\widehat{p}_{\delta}$)		
		$\delta=0$	$\delta=-1$	$\delta=-3$	$\delta=0$	$\delta=-1$	$\delta=-3$
(3, 3)	5	1.34E−02	7.68E−03	2.89E−03	2.57E−02	8.92E−03	3.19E−03
	10	8.16E−03	4.29E−03	1.34E−03	1.16E−02	4.58E−03	1.47E−03
(4, 5)	5	1.15E−02	7.36E−03	3.30E−03	2.57E−02	9.36E−03	3.71E−03
	10	7.28E−03	4.23E−03	1.62E−03	1.14E−02	4.96E−03	1.70E−03
(4, 1)	5	1.03E−02	4.71E−03	8.71E−04	1.75E−02	5.26E−03	9.62E−04
	10	7.26E−03	2.73E−03	4.06E−04	8.91E−03	2.46E−03	3.61E−03
(1, 4)	5	6.14E−03	4.18E−03	2.29E−03	1.16E−02	5.04E−03	2.49E−03
	10	3.32E−03	2.22E−03	1.12E−03	4.85E−03	2.56E−03	1.14E−03

Table 3
Average width of Bayesian credible intervals for parameter p.

(α,β)	n	Average width 95% CI		
		$\delta=0$	$\delta=-1$	$\delta=-3$
(3, 3)	5	0.438	0.335	0.197
	10	0.344	0.250	0.136
(4, 5)	5	0.402	0.328	0.214
	10	0.322	0.251	0.151
(4, 1)	5	0.363	0.217	0.079
	10	0.294	0.158	0.049
(1, 4)	5	0.251	0.217	0.167
	10	0.188	0.158	0.117

maximum likelihood estimation instead of the method of moments; see pp. 40-41 of Maritz and Lwin (1995). This strategy, also known as Type II maximum likelihood, is a well-established empirical Bayes methodology.

We first compute the marginal probability function of a sample $\mathbf{t}$ (see Section 2.1 for definitions) as

$$m(\mathbf{t}|\alpha,\beta) = \int_0^1 \mathcal{L}(\mathbf{t},p)\pi(p)dp = \frac{1}{\mathbf{B}(\alpha,\beta)}\int_0^1 p^{(a+\alpha-1)}(1-p)^{(b+\beta-1)}dp = \frac{\mathbf{B}(a+\alpha,b+\beta)}{\mathbf{B}(\alpha,\beta)},$$

where $\mathbf{B}(\alpha,\beta)$ denotes the beta function.

Next, suppose we have m past samples $\mathbf{t}^{(1)},\ldots,\mathbf{t}^{(m)}$. Then, assuming independence, the joint marginal likelihood of the past samples is given by

$$m(\mathbf{t}^{(1)},\ldots,\mathbf{t}^{(m)}|\alpha,\beta) = \prod_{k=1}^{m}\frac{\mathbf{B}\left(a^{(k)}+\alpha,b^{(k)}+\beta\right)}{\mathbf{B}(\alpha,\beta)}, \tag{10}$$

where $a^{(k)}, b^{(k)}$ are as defined in (3), for the kth past sample. Finally, $\hat{\alpha}, \hat{\beta}$ are defined as solutions of the problem $\max_{\alpha,\beta>0} m(\mathbf{t}^{(1)},\ldots,\mathbf{t}^{(m)}|\alpha,\beta)$. However, depending on the data, unrestricted maximization is not always possible, as (10) may increase indefinitely with α and β. For illustration, suppose $m=2$, $a^{(1)}=1$, $b^{(1)}=1$, $a^{(2)}=2$ and $b^{(2)}=1$, then (10) takes the form

$$F(\alpha,\beta) := \frac{\alpha^2(\alpha+1)\beta^2}{(\alpha+\beta)^2(\alpha+\beta+1)^2(\alpha+\beta+2)}, \quad \alpha,\beta>0. \tag{11}$$

Letting $\alpha+\beta=t$ above, we obtain

$$G(\alpha,t) := \frac{\alpha^2(\alpha+1)(t-\alpha)^2}{t^2(t+1)^2(t+2)}, \quad 0<\alpha<t.$$

It is easy to show that, for fixed $t>0$, G reaches its maximum at $\alpha^*(t) = (3t-4+\sqrt{(3t-4)^2+40t})/10$. That is,

$$H(t) := G(\alpha^*(t),t) = \max_{\alpha} G(\alpha,t), \quad t>0.$$

Finally, it can be shown that H is increasing and converges to $H(\infty)=108/3125=\sup_{\alpha,\beta}F(\alpha,\beta)$. So, the function in (11) tends to its supremum when $\alpha,\beta\to\infty$.

Situations such as the one depicted above have been known for a long time in maximum likelihood estimation. In order to reduce this drawback it is often convenient to maximize over a constrained parameter space. Here we propose to restrict the set of hyperparameters α,β by imposing a constraint related to the method of moments. Observe that the expected value of the nth record is given by $E[R_n]=n\left(\frac{\alpha+\beta-1}{\alpha-1}\right)$; see (2.20) in Doostparast and Ahmadi (2006). Then, we replace the

Table 4
EMSE of $\widehat{p}_{\pi,\delta}$ using the empirical Bayes method.

(α, β)	n	EMSE($\widehat{p}_{\pi,\delta}$)			
		$\delta = 0$	$\delta = 0$	$\delta = -1$	$\delta = -3$
(3, 3)	5	2.44E−02	1.93E−02	9.75E−03	3.20E−03
	10	1.34E−02	1.12E−02	4.82E−03	1.51E−03
(4, 5)	5	2.09E−02	1.56E−02	9.04E−03	4.15E−03
	10	1.08E−02	9.37E−03	5.26E−03	1.78E−03
(4, 1)	5	2.02E−02	1.58E−02	5.39E−03	8.73E−04
	10	1.17E−02	9.61E−03	2.67E−03	3.69E−04
(1, 4)	5	2.50E−02	9.05E−03	5.00E−03	2.52E−03
	10	1.52E−02	4.23E−03	2.74E−03	1.20E−03

Table 5
Observed data in Section 2.4 (read from left to right).

1, 5, 1, 1, 1, 3, 2, 4, 3, 2, 3, 1, 1, 1, 3, 1, 3, 1, 6, 4, 1, 9, 2, 6, 2, 1, 3, 1, 3, 1, 1, 10, 2, 7, 1, 8, 1, 1, 2, 1, 1, 6, 1, 2, 1, 4,
1, 1, 1, 3, 5, 1, 1, 1, 1, 5, 2, 4, 5, 1, 2, 2, 1, 3, 1, 1, 1, 3, 1, 2, 1, 1, 1, 1, 1, 1, 5, 2, 2, 4, 6, 1, 3, 1, 1, 1, 1

Table 6
δ-record values for data set of Table 5.

δ	δ-record values
0	1, 5, 6, 9, 10
−1	1, 5, 6, 9, 10
−3	1, 5, 3, 4, 3, 3, 3, 3, 9, 10, 8

expectation by the average of the m past nth records; set

$$n\left(\frac{\alpha+\beta-1}{\alpha-1}\right) = \bar{R} := \frac{1}{m}\sum_{k=1}^{m} R_n^{(k)}, \tag{12}$$

and maximize (10) subject to (12), where $R_n^{(k)}$ denotes the nth record of the kth past sample. It is not clear if constraint (12) guarantees that a maximum always exists, although this has been the case in all of our simulations.

Once α and β are estimated from the past samples $\mathbf{t}^{(1)}, \ldots, \mathbf{t}^{(m)}$, $\widehat{p}_{\pi,\delta}$ is computed from (9) using the $(m+1)$th sample $\mathbf{t}^{(m+1)}$. For $m = 10$ and several combinations of values of α, β and n, we performed 10^3 simulation runs and results are shown in Table 4. The first column from the left (under the heading $\delta = 0$) displays the EMSE for estimations based on the method of moments and relying only on records, as in Doostparast and Ahmadi (2006). In the second one, estimations depend only on records as well, but we apply the maximum likelihood approach described above. The remaining columns are all related to δ-record data and the maximum likelihood estimation of α and β.

It is interesting to compare the first two columns of Table 4 and see that our restricted ML strategy, for estimating the hyperparameters, beats the method of moments used in Doostparast and Ahmadi (2006). This is presumably related to the lack of efficiency of the latter method. In the remaining columns we find further evidence supporting the use of δ-records.

2.4. Real data

We apply our methods to a real data set. The data shown in Table 5 are taken from Table III of Xie and Goh (1993). They correspond to the number of items observed until a defective item is detected and are modelled in Xie and Goh (1993) by the geometric distribution. There are 5 records in the sample and the values of δ-records observed, for $\delta = 0, -1, -3$, are shown in Table 6.

We estimate p using records and δ-records, and compare the estimations with those obtained using the whole sample of 87 observations. When dealing with real data, the sample may possibly not include all δ-records associated to the last observed record. In such case the expression of the likelihood has to be corrected by replacing $P[S_n = s_n]$ with $P[S_n \geq s_n]$. This is finally tantamount to replacing the rightmost term q^{r_n} in (1) with $q^{(r_n+\delta)^+}$. Of course, the estimators must be modified accordingly.

Table 7 shows the MLE and Bayes estimates of p using δ-records and using the complete data. For Bayesian estimation we take the uniform prior Beta(1, 1), as we have no previous information of p. We observe that the estimations with $\delta = -3$ are close to those obtained with the whole sample, a fact which is interesting since the former are based only on 11 δ-records instead of the 87 observations of the latter.

Table 7
Estimations of p in Section 2.4.

	$\delta = 0$	$\delta = -1$	$\delta = -3$	All data
$\widehat{p}_{\delta}$	0.500	0.357	0.448	0.420
$\widehat{p}_{\pi,\delta}$	0.500	0.375	0.452	0.423

3. Prediction

A very important question when analysing extreme data is to predict future high values. The problem of predicting future records has been investigated by many authors. For continuous distributions see Arnold et al. (1998) and references therein, Soliman et al. (2006) and Sultan et al. (2008). In the case of the geometric distribution, Doostparast and Ahmadi (2006) present both Bayesian and non-Bayesian prediction of future records. We follow their scheme and show how δ-records can be successfully incorporated to improve performance.

3.1. Maximum likelihood prediction

The technique of maximum likelihood prediction of future records was introduced by Basak and Balakrishnan (2003). It consists in defining the so-called predictive likelihood function, as the joint likelihood of the observed sample of records and the unobserved future records. Then this function is maximized, both in the parameter and in the future record values.

Our aim is to predict the mth record R_m assuming that δ-records have been observed up to the nth record, $(m > n)$. For this, we use the likelihood function (1) and the probability function of the mth record, conditional on the data $\mathbf{T}$, given by

$$P[R_m = z|\mathbf{T} = \mathbf{t}, p] = \binom{z - r_n - 1}{m - n - 1} p^{m-n}(1-p)^{z-r_n-m+n}, \tag{13}$$

for $z \geq r_n + m - n$.

Then the predictive likelihood function is defined by

$$\mathcal{L}(z, p) := \mathcal{L}(z, p, \mathbf{t}) = P\left[R_m = z|\mathbf{T} = \mathbf{t}, p\right] \mathcal{L}(\mathbf{t}, p) = \binom{z' - 1}{m' - 1} p^{a+m'}(1-p)^{z'-m'+b}, \tag{14}$$

for $z \geq r_n + m - n$, where a, b are defined in (3) and $z' = z - r_n$, $m' = m - n$. So, we obtain the log predictive likelihood as

$$l(z', p) := \left(a + m'\right) \log p + \left(z' - m' + b\right) \log(1-p) + \log \binom{z' - 1}{m' - 1}.$$

For fixed z' we maximize l with respect to p by solving $\partial l/\partial p = 0$, thus obtaining $\tilde{p} = \frac{a+m'}{z'+a+b}$. Then we replace p by $\tilde{p}$ in l and maximize $l(z', \tilde{p})$ with respect to z', which takes integer values. So, maximizing (14) is equivalent to maximizing

$$\Phi(z') = \frac{(z' - m' + b)^{z'-m'+b}}{(z' + a + b)^{z'+a+b}} \frac{(z' - 1)!}{(z' - m')!}. \tag{15}$$

It is easy to see that, for $z' \geq m'$, $\Phi(z') \leq (z'+a+b)^{-(a+1)}$, hence $\Phi(z') < \Phi(m')$, for every $z' > t := \Phi(m')^{-1/(a+1)} - (a+b)$. Thus we simply evaluate Φ at $z' = m', \ldots, \lceil t \rceil$, where $\lceil . \rceil$ denotes the ceiling function.

In the particular case of $m = n + 1$ we have $m' = 1$ and so (15) takes the simple form

$$\Phi(z') = \frac{(z' - 1 + b)^{z'-1+b}}{(z' + a + b)^{z'+a+b}},$$

which is easily shown to be decreasing in z'. So Φ attains its maximum value at $z' = 1$ or, equivalently at $z = r_n + 1$. Thus, in the case $m = n + 1$, the predicted $(n + 1)$th record, denoted by $\tilde{R}_{n+1}$, and the corresponding estimation of p are given by

$$\tilde{R}_{n+1} = R_n + 1 \quad \text{and} \quad \tilde{p} = \frac{A + 1}{A + B + 1}.$$

Note however that, for $m > 1$, $R_n + m$ needs not be the maximum likelihood predictor of R_m.

We performed a simulation study to investigate the behaviour of the maximum likelihood predictor of future records, for several combinations of δ, n, m and p. Table 8 displays the EMSE of the predicted record $\tilde{R}_m$ and of $\tilde{p}$. Observe that no improvement in the prediction of the next record value $(m = n + 1)$ results from using δ-records, because $\tilde{R}_{n+1} = R_n + 1$, regardless of the value of δ. However up to a 40% reduction of the EMSE is achieved when predicting records further into the future. This suggests that the more distant the record (to be predicted) is, the more is gained by using δ-records.

From Table 8 we also see that, while the EMSE of $\tilde{p}$ seems to decrease to 0, as $|\delta|$ grows, this is not the case for $\tilde{R}_m$. The reason for this apparently strange behaviour is that the EMSE of any predictor based on the available information, say R^*_m,

Table 8
EMSE of the maximum likelihood predictor of future records and the associated estimator of p, from 10 000 simulation runs.

p	n	m	EMSE($\tilde{R}_m$)			EMSE($\tilde{p}$)		
			$\delta = 0$	$\delta = -1$	$\delta = -3$	$\delta = 0$	$\delta = -1$	$\delta = -3$
0.5	5	6	3.047	3.024	2.986	3.36E−02	1.05E−02	3.67E−03
	5	8	11.026	10.026	8.532	4.26E−02	1.14E−02	3.51E−03
	5	10	21.223	16.659	12.897	4.61E−02	1.33E−02	3.49E−03
	10	11	3.032	2.987	2.998	1.55E−02	5.65E−03	1.71E−03
	10	13	9.979	8.965	8.528	1.46E−02	6.11E−03	1.71E−03
	10	15	17.140	14.824	12.567	1.90E−02	6.41E−03	1.69E−03
0.7	5	6	0.805	0.802	0.803	2.67E−02	7.47E−03	1.51E−03
	5	8	3.218	3.189	3.157	3.01E−02	7.29E−03	1.66E−03
	5	10	6.344	4.822	4.121	3.34E−02	7.29E−03	1.65E−03
	10	11	0.806	0.804	0.794	1.44E−02	4.09E−03	5.36E−04
	10	13	3.578	3.499	3.356	9.32E−03	4.10E−03	5.96E−04
	10	15	5.589	4.356	4.243	1.79E−02	3.88E−03	5.99E−04

has a positive lower bound. Indeed, let $\mathcal{F}$ be the σ-algebra generated by $\{X_k;\ k \le L_{n+1}-1\}$ and suppose R_m^* is $\mathcal{F}$-measurable. That is, R_m^* may depend on all observations, with index less than the $(n+1)$th record time, and not only on $\mathbf{T}$. Then

$$\begin{aligned} E[(R_m - R_m^*)^2 \mid \mathcal{F}, p] &\ge \min_r E\left[(R_m - r)^2 \mid \mathcal{F}, p\right] \\ &\ge \min_r E\left[(R_m - r)^2 \mid R_n, p\right] \\ &= E\left[(R_m - E[R_m \mid R_n, p])^2 \mid R_n, p\right] \\ &= E\left[\left(R_m - \frac{m-n}{p}\right)^2 \Big| R_n, p\right] = (m-n)\frac{q}{p^2}, \end{aligned}$$

where the last line of the display above follows from the fact that R_m, conditional on R_n, is distributed as sum of $m-n$ independent geometric random variables, with parameter p, starting at 1. Finally, taking expectation on both sides above, we have

$$\text{EMSE}(R_m^*) := E\left[(R_m - R_m^*)^2\right] \ge (m-n)\frac{q}{p^2}.$$

So, for instance, letting $p = 0.5$, $n = 5$ and $m = 8$ (second line in Table 8), we see that the EMSE of $\tilde{R}_m$ cannot be lower than 6, even if we knew the actual p.

3.2. Bayesian approach

In this section we consider point and interval prediction of future records, from the Bayesian viewpoint. As in Section 2.3, we use quadratic errors and a beta prior for p, with hyperparameters α, β.

A Bayesian prediction of a future record R_m is obtained from the Bayes predictive density function. We use, as above, the expression for the conditional probability of $R_m = r_m$ given $\mathbf{T}$, shown in (13). Recall, from Section 2.3, that the posterior distribution $\pi(p|\mathbf{t})$ is beta, with parameters $\alpha + a$, $\beta + b$, where a, b are defined in (3). Then, using the notation of Section 3.1, the Bayes predictive density function of R_m given $\mathbf{T}$ is calculated as

$$\begin{aligned} P[R_m = z|\mathbf{T} = \mathbf{t}] &= \int_0^1 P[R_m = z|\mathbf{T} = \mathbf{t}, p]\pi(p|\mathbf{t})dp. \\ &= \int_0^1 \binom{z'-1}{m'-1} p^{m'}(1-p)^{z'-m'} \frac{p^{\alpha+a-1}(1-p)^{\beta+b-1}}{B(\alpha+a, \beta+b)} dp \\ &= \binom{z'-1}{m'-1} \frac{B\left(m'+\alpha+a, z'-m'+\beta+b\right)}{B(\alpha+a, \beta+b)}, \end{aligned}$$

for $z' \ge m'$. Thus, conditional on $\mathbf{T} = \mathbf{t}$, $R_m - r_n$ has a beta-Pascal distribution $BP(m', \alpha + a, \beta + b)$.

We can first consider the Bayes point predictor $\tilde{R}_m^B$ of r_m $(m > n)$, calculated as the posterior mean $E[R_m|\mathbf{T}] = R_n + E[R_m - R_n|\mathbf{T}]$. That is,

$$\tilde{R}_m^B = R_n + (m-n)\frac{A + B + \alpha + \beta - 1}{A + \alpha - 1}. \tag{16}$$

Table 9
EMSE of Bayes, MAP and ML predictors of future records, with a Beta(α, β) prior, from 10 000 simulation runs.

(α, β)	n	m	EMSE($\tilde{R}_m^B$)			EMSE($\tilde{R}_m^{MAP}$)			EMSE($\tilde{R}_m$)		
			$\delta = 0$	$\delta = -1$	$\delta = -3$	$\delta = 0$	$\delta = -1$	$\delta = -3$	$\delta = 0$	$\delta = -1$	$\delta = -3$
(3, 3)	5	6	7.341	7.252	6.505	11.050	11.973	10.661	11.050	11.973	10.661
	5	8	31.230	29.621	28.406	44.987	44.416	40.560	43.463	43.251	39.225
	5	10	62.452	58.862	52.689	85.895	84.712	71.497	105.894	99.041	86.210
	10	11	8.206	8.157	6.963	13.171	13.405	12.270	13.171	13.405	12.270
	10	13	27.453	27.132	25.051	39.250	39.023	35.030	44.853	42.059	36.654
	10	15	52.832	51.864	47.213	71.302	70.791	62.625	96.478	92.335	85.467
(4, 1)	5	6	0.693	0.671	0.656	0.906	0.921	0.914	0.906	0.921	0.914
	5	8	2.858	2.469	2.241	5.008	3.841	3.257	3.885	3.153	2.831
	5	10	5.575	4.566	3.873	9.120	6.724	5.120	7.352	5.522	5.894
	10	11	0.714	0.625	0.612	0.988	0.925	0.919	0.988	0.925	0.919
	10	13	2.502	2.370	2.106	4.014	3.426	3.383	3.415	3.243	3.193
	10	15	4.837	4.083	3.676	7.187	5.762	5.125	6.501	5.868	6.486

Table 10
Average width (AW) and coverage probability (CP) of Bayesian prediction bounds based on 10 000 simulation runs.

n	m	Beta(3, 3)						Beta(4, 1)					
		$\delta = 0$		$\delta = -1$		$\delta = -3$		$\delta = 0$		$\delta = -1$		$\delta = -3$	
		AW	CP	AW	CP	AW	CP	AW	CP	AW	CP	AW	CP
5	6	5.821	0.961	5.533	0.965	5.479	0.960	1.638	0.972	1.601	0.977	1.538	0.977
5	8	12.675	0.956	12.125	0.960	11.767	0.965	3.646	0.969	3.343	0.972	3.176	0.974
5	10	18.859	0.952	17.636	0.960	16.507	0.957	5.620	0.969	4.830	0.967	4.392	0.976
10	11	5.563	0.959	5.499	0.963	5.431	0.967	1.714	0.976	1.597	0.978	1.531	0.975
10	13	12.005	0.961	11.489	0.961	11.245	0.962	3.348	0.961	3.237	0.970	3.175	0.974
10	15	17.441	0.959	16.591	0.958	16.016	0.962	5.062	0.967	4.503	0.969	4.339	0.974

We may also consider the MAP (maximum a posteriori) predictor, which is simply the posterior mode. An inspection of the beta-Pascal probability mass function, with parameters l, s, t, reveals that the mode is located at 1 if $l = 1$ or at $\left\lceil \frac{(l-1)(s+t)}{s+1} \right\rceil$, if $l > 1$. Therefore, if $m = n + 1$, then $m' = 1$ and the MAP predictor is given by $\tilde{R}_m^{MAP} = R_n + 1$. Otherwise, if $m > n + 1$, then $m' > 1$ and so, the MAP predictor is

$$\tilde{R}_m^{MAP} = R_n + \left\lceil \frac{(m - n - 1)\,(A + B + \alpha + \beta)}{A + \alpha + 1} \right\rceil .$$

Table 9 shows the EMSE of $\tilde{R}_m^B$, $\tilde{R}_m^{MAP}$ and $\tilde{R}_m$ (ML predictor) for several combinations of values of α, β, n, m and δ. Observe that, for $m = n+1$, $\tilde{R}_m^{MAP}$ depends only on R_n and so, there is no gain by using δ-records. However, $\tilde{R}_m^B$ does take advantage of δ-records and some reduction of the EMSE is seen as $|\delta|$ grows. On the other hand, for $m > n + 1$, the decrease of the EMSE of both predictors, as $|\delta|$ grows, is even more notorious. As in Table 2 we also include the EMSE of the predictions when ML prediction is used.

Observe that several entries of Table 9 are quite large. This phenomenon can be explained in terms of a lower bound for the EMSE, as we did in Section 3.1. Recall that the EMSE of any $\mathcal{F}$-measurable predictor is bounded (below) by $(n-m)(1-p)/p^2$ which, after integration with respect to the beta prior, yields

$$\int_0^1 (m-n)\frac{p^{\alpha-3}(1-p)^{\beta}}{\mathbf{B}(\alpha, \beta)}dp = \frac{(m-n)\mathbf{B}(\alpha-2, \beta+1)}{\mathbf{B}(\alpha, \beta)} = \frac{(m-n)\beta(\alpha+\beta-1)}{(\alpha-1)(\alpha-2)}, \tag{17}$$

for $\alpha > 2$. In the particular case $\alpha = \beta = 3$, $n = 5$, $m = 8$ and $\delta = -3$, (17) yields 45/2, a value not too far off from the corresponding entry (28.406) of Table 9. A closer agreement can be seen in the case $\alpha = 4$, $\beta = 1$, $n = 5$, $m = 8$ and $\delta = -3$.

Now we establish prediction intervals for future records. From the Bayes predictive density function of R_m given $\mathbf{T}$, we compute the shortest interval of values with probability at least $1 - \alpha$. In Table 10 we show the average width of credible intervals for R_m, for several values of α, β, n, m and δ. We also include empirical coverage probabilities. As it can be expected, they are above 0.95 since the endpoints of the intervals take integer values.

3.3. Real data

We consider the data presented in Section 2.4 that we use now for predicting future record values. In Table 11 we give the predictions based on records and δ-records and on the whole sample.

As in Section 2.4 we observe an overall good agreement of predictions based on δ-records ($\delta = -3$) with those based on the whole sample.

Table 11
ML and Bayes prediction of future records for data set in Section 2.4 using records, δ-records and all data (AD).

m	$\widetilde{R}_m$				$\widetilde{R}_m^B$				$\widetilde{R}_m^{MAP}$			
	$\delta=0$	$\delta=-1$	$\delta=-3$	AD	$\delta=0$	$\delta=-1$	$\delta=-3$	AD	$\delta=0$	$\delta=-1$	$\delta=-3$	AD
6	11	11	11	11	12.2	13	12.31	12.39	11	11	11	11
8	14	15	15	15	16.6	19	16.92	17.17	14	15	15	15
10	17	20	19	20	21	25	21.54	21.95	17	20	19	20

4. Conclusions and future work

In this paper we attempt to evaluate the impact of δ-records on various inferential procedures for the geometric distribution. For that purpose we consider the record dependent methods in Doostparast and Ahmadi (2006) as the baseline scenario. These methods are easily "upgraded" to accept δ-records as input and their performance is evaluated, mainly in terms of mean-square error, from simulated as well as real data.

Results show, in various degrees, better performance of methods incorporating δ-records, either under Bayesian or non-Bayesian frameworks, but this comparison has to be taken with caution because we are actually using more information. However, a promising fact is that a few records (and their δ-records) can outperform a larger number of records; see comments following Table 1. So, under particular experimental conditions, such as those in destructive stress testing, it may be more efficient to reduce the number of records to be observed in exchange for some δ-records.

The importance of the geometric distribution justifies a detailed study such as ours but the loss-of-memory property may tend to mask the impact of δ-records. It is therefore necessary to evaluate their usefulness in other popular discrete models but then, we cannot expect to have explicit formulas for estimators such as (4) or (16). Finally, there is also a need for theoretical results to complement the empirical knowledge obtained from simulations. In that direction, we believe that Proposition 2.4, on the consistency of $\widehat{p}_\delta$, is important. However, as the reader may notice, the simplicity of its proof is tied to the loss-of-memory and so, departures from the geometric model will bring about a more complex analysis.

Another important aspect to consider when using δ-records, is the sampling scheme. For instance, in this paper we fix the number n of records and observe all their associated δ-records (see Section 2.1 for definitions). However, it may be natural to observe some but not all δ-records associated to R_n, in which case the likelihood function must be adjusted. In both schemes above, the sample size is random but this can be easily changed by fixing the total number of δ-records. Of course, in this case, the number of records in the sample is random.

Last, we mention the problem of finding an "optimal" δ: i.e., if we fix δ before the experiment, how should we proceed? We have shown throughout the paper that, as $|\delta|$ gets larger, we obtain better estimations and predictions, since increasing $|\delta|$ increases the number of observations available for inference. However, in many cases, such as destructive testing, each observation has a cost and therefore a compromise between accuracy and cost must be found. The expected number of δ-records associated to the first n records is given by

$$\sum_{i=1}^{-\delta-1}\left(\left(\frac{p}{q}\right)^i\sum_{k=i}^{-\delta-1}\binom{k-1}{i-1}+\sum_{k=0}^{i-1}\binom{-\delta-1}{k}\left(\frac{p}{q}\right)^k\frac{1}{q}\right)+(n+\delta+1)q^\delta,$$

expression which can be approximated by nq^δ. Although the value of q is unknown, previous knowledge may serve as a guide to infer the number of δ-records to be observed. Another possibility is to consider different values of δ along the experiment, adapted to the successive record values. The likelihood in (1) can be modified consequently to yield estimations and predictions in this situation. Thus, when a small (big) number of δ-records are observed for the first records, a change in δ for the remaining records can be useful to increase accuracy (decrease cost).

Acknowledgements

We thank the referees and the Associate Editor for their useful comments and suggestions. Support is thankfully acknowledged from project PFB-03-CMM, FONDECYT grant 1120408 and MINECO project MTM2010-15972. The third author gratefully acknowledges support from a Gran Mariscal de Ayacucho Foundation grant (Venezuela) E-228-239-2007-1. The authors are members of the research group Modelos Estocásticos (DGA).

References

Ahsanullah, M., 1995. Record Statistics. Nova Science Publishers, Commack, New York.
Arnold, B., Balakrishnan, N., Nagaraja, H., 1998. Records. John Wiley & Sons, New York.
Bagrezaei, S., Nezad, D., Parsian, A., 2012. A note on further statistical analysis of geometric distribution. J. Appl. Probab. Stat. 2, 61–70.
Balakrishnan, N., Pakes, A., Stepanov, A., 2005. On the number and sum of near-records observations. Adv. Appl. Probab. 37, 765–780.
Basak, P., Balakrishnan, N., 2003. Maximum likelihood prediction of future record statistic. In: Lindquist, B.H., Doksun, K. (Eds.), Mathematical and Statistical Methods in Reliability. World Scientific Publishing, pp. 159–175.
Castaño Martínez, A., López-Blázquez, F., Salamanca-Miño, B., 2013. An additive property of weak records from geometric distributions. Metrika 76, 449–458.

Doostparast, M., Ahmadi, J., 2006. Statistical analysis for geometric distribution based on records. Comput. Math. Appl. 52, 905–916.
Glick, N., 1978. Breaking records and breaking boards. Amer. Math. Monthly 52, 905–916.
Gouet, R., López, F., Sanz, G., 2007. Asymptotic normality for the counting process of weak records and δ-records in discrete models. Bernoulli 13, 754–781.
Gouet, R., López, F., Sanz, G., 2008. Laws of large numbers for the number of weak records. Stat. Probab. Lett. 78, 2010–2017.
Gouet, R., López, F., Sanz, G., 2012. On δ-record observations: asymptotic rates for the counting process and elements of maximum likelihood estimation. Test 21, 188–214.
Gulati, D., Padgett, W., 2003. Parametric and Nonparametric Inference from Record-Breaking Data. In: Lecture Notes in Statistics, vol. 72. Springer-Verlag, New York.
Hashorva, E., Stepanov, A., 2012. Limit theorems for the spacings of weak records. Metrika 75, 307–320.
Jaheen, Z., 2003. A Bayesian analysis of record statistics from the Gompertz model. Appl. Math. Comput. 145, 307–320.
López-Blázquez, F., Salamanca-Miño, B., 2013. Distribution theory of δ-record values. case $\delta \leq 0$. Test 22, 715–738.
Maritz, J., Lwin, T., 1995. Empirical Bayes Methods, second ed. In: CRC Monographs on Statistics & Applied Probability, vol. 35. Chapman & Hall, London.
Nevzorov, V., 2001. Records: Mathematical Theory. In: Translations of Mathematical Monographs, vol. 194. American Mathematical Society, Providence R.I.
Raqab, M., Ahmadi, J., Doostparast, M., 2007. Statistical inference based on record data from Pareto model. Statistics 41, 105–118.
Soliman, A., Abd Ellah, A., Sultan, K., 2006. Comparison of estimates using record statistics from Weibull model: Bayesian and non-Bayesian approaches. Comput. Statist. Data Anal. 51, 2065–2077.
Stepanov, A., Balakrishnan, N., Hofmann, G., 2003. Exact distribution and Fisher information of weak record values. Stat. Probab. Lett. 64, 69–81.
Sultan, K., AL-Dayian, G., Mohammad, H., 2008. Estimation and prediction from gamma distribution based on record values. Comput. Statist. Data Anal. 52, 1430–1440.
Xie, M., Goh, T., 1993. Spc of a near zero-defect process subject to random shocks. Qual. Reliab. Eng. Int. 9, 89–93.