Austral. & New Zealand J. Statist. **41**(3), 1999, 337–352

MODEL SELECTION FOR A FAMILY OF DISCRETE FAILURE TIME DISTRIBUTIONS

BERNHARD KLAR[1]

Universität Karlsruhe

Summary

In recent years, a large number of new discrete distributions have appeared in the literature. However, flexible discrete models which, at the same time, allow for easy statistical inference, are still an exception. This paper makes a detailed analysis of a family of discrete failure time distributions which meets both requirements. It examines the maximum likelihood estimation of the unknown parameters and presents a goodness-of-fit test for this model. The test is used for the selection of an appropriate model for datasets of frequencies of the duration of atmospheric circulation patterns.

Key words: model selection; goodness-of-fit test; discrete distribution; failure time model; maximum likelihood estimation.

1. Introduction

This paper examines a family of discrete failure time distributions and its application to meteorological datasets. Although discrete time models occur quite naturally in many applications, for example in reliability theory or biology, there are only a few results on discrete life time distributions in the literature (apart from the geometric distribution that is characterized by a constant hazard rate). Besides a brief discussion in Kalbfleisch & Prentice (1980) about discrete failure distributions that can be obtained from continuous distributions such as the Weibull, discrete failure models are examined, for instance, in Salvia & Bollinger (1982), Xekalaki (1983), Padgett & Spurrier (1985). Their models have one or two parameters and allow constant, increasing or decreasing failure rates. The datasets considered below show that these models are not flexible enough in some cases. Moreover, they give no probability theory for the estimators, so their statistical inference has no rigorous foundation.

Other increasing failure rate distributions are the Poisson and the binomial distribution, whereas the negative binomial distribution has constant, increasing or decreasing failure rate depending on the parameter (see e.g. Johnson, Kotz & Kemp, 1992). However, these distributions are not typically used as failure time models because there are no simple expressions for the hazard function. Again, these distributions are not versatile enough for our purposes.

Adams & Watson (1989) introduced a very flexible parametric discrete time failure model. Because the number of parameters can be chosen arbitrarily, the model allows for a variety of hazard function and probability mass function shapes (as Figs 1–8 below show). The model is convenient for statistical inference. There are reasonably simple forms for the survivor function and the probability mass function because they are finite products of the hazard function (a property that is not shared by continuous models).

Received July 1998; Revised October 1998; accepted November 1998.

[1] Institut für Mathematische Stochastik, Universität Karlsruhe, Englerstr. 2, D-76128 Karlsruhe, Germany. email: Bernhard.Klar@math.uni-karlsruhe.de

Acknowledgments. The author thanks Norbert Henze for drawing attention to the work of Adams and Watson and for many constructive comments on earlier drafts of this paper.

We used this model to fit datasets of frequencies of the duration of atmospheric circulation patterns that are a basic part of a space–time model for daily rainfall (Bárdossy & Plate, 1992). In their work, Bárdossy & Plate used the generalized Poisson distribution (Consul, 1989) to model the overdispersion relative to the Poisson distribution. However, a careful analysis (Henze & Klar, 1995) showed that this assumption may not be justified in many cases. Although it is certainly not possible to give a theoretical justification for a particular distribution, the development of an atmospheric circulation pattern suggests the application of a lifetime distribution.

The paper is organized as follows. Section 2 describes the discrete failure model and some of its properties. We point out that, in its original definition, the model does not always generate a proper probability distribution. In specific applications to survival analysis (for example, the time to recurrence of symptoms that might never occur), improper distributions may be appropriate. However, we wish to apply the model in a situation where the variables must be finite, so we need to modify the model in order that a proper distribution results. Section 3 takes a closer look at the estimation of the parameters; in particular, it establishes the uniqueness of the maximum likelihood estimator (MLE). Furthermore, it proves, rigorously, the assertion of Adams & Watson (1989), that the asymptotic distribution of the estimator is multivariate normal. In Section 4 we briefly describe a goodness-of-fit test for discrete distributions and use it as a model selection procedure for frequencies of the duration of circulation pattern. In our opinion, it is preferable to the selection procedure used by Adams and Watson.

2. The Discrete Failure Time Model

Adams & Watson (1989) introduced the following parametric discrete failure time model. Let G and g denote the distribution function and the probability density function of a continuous symmetric distribution, i.e.

$$g(x) = g(-x), \quad G(x) = 1 - G(-x) \qquad (x \in \mathbb{R}).$$

Then by means of a (low order) polynomial

$$\xi(t) = \theta_0 + \theta_1 t + \cdots + \theta_m t^m, \tag{1}$$

where $\theta = (\theta_0, \theta_1, \ldots, \theta_m)' \in \Theta = \mathbb{R}^{m+1}$, the hazard mass function h is defined by

$$h(t) = \Pr(T = t \mid T \geq t) = G\big(\xi(t)\big) \qquad (t \in \mathrm{N}_0),$$

where N_0 denotes the non-negative integers and T is a non-negative, integer-valued random variable representing failure time. From the well-known formulas

$$h(t) = \frac{p(t)}{S(t)} \qquad \text{and} \qquad S(t) = \prod_{s=0}^{t-1} \big(1 - h(s)\big), \tag{2}$$

where p denotes the probability mass function (pmf) and $S(t) = \Pr(T \geq t)$ is the probability of survival until time t, it follows that

$$S(t) = \prod_{s=0}^{t-1} G\big(-\xi(s)\big) \qquad \text{and} \qquad p(t) = G\big(\xi(t)\big) \prod_{s=0}^{t-1} G\big(-\xi(s)\big). \tag{3}$$

Throughout the following, G is the logistic function

$$G(x) = \frac{1}{1+e^{-x}} \ . \tag{4}$$

This seems to be the most appropriate choice, as already mentioned by Adams & Watson (1989). Although other possible choices of G, such as the normal distribution function, result in models of comparable smoothness and flexibility, statistical inference is more involved.

For $m \geq 1$, $S(\infty) = \lim_{t\to\infty} S(t) = 0$ only if $\theta_m > 0$, whereas $S(\infty) \in (0,1)$ if $\theta_m < 0$ (note that the infinite product $\prod_{k=0}^{\infty}(1-a_k)$, $a_k \in (0,1)$, converges to a positive number if, and only if, $\sum_{k=0}^{\infty} a_k$ converges). Hence it is necessary that the highest order coefficient is positive to ensure that a proper pmf is defined by (3). One way to overcome this drawback is to restrict the parameter space, but this raises difficulties in the estimation and testing procedure. Moreover, the possible shapes of the model would be restricted in an undesirable way.

Normalizing the pmf with $\big(1-S(\infty)\big)^{-1}$ is not appropriate because then the normalized hazard function

$$h'(t) = \frac{p(t)}{S(t)-S(\infty)} = \frac{G\big(\xi(t)\big)}{1-\prod_{s=k}^{\infty} G\big(-\xi(s)\big)}$$

would not have a simple form any longer.

In practical applications, however, one observes that even if θ_m is negative $S(\infty)$ is nearly zero. For instance, for $m = 2$ and $\theta_0 = -14.8, \theta_1 = 0.0736, \theta_2 = -0.000122$ (see Adams & Watson, 1989 Example 3.4) we have $S(\infty) \approx 0.02$. Hence for theoretical considerations we modify the polynomial in (1) and define

$$\xi(t) = \theta_0 + \theta_1 t + \cdots + \theta_m t^m + \varepsilon t^{m+1} \qquad (\varepsilon > 0),$$

where ε is an arbitrary but fixed positive constant. From a practical point of view, the last term should not affect estimation of the parameter vector θ; hence ε should be chosen to be very small (or even omitted) in a computer implementation.

From the definition of the hazard function it is obvious that h is increasing if and only if the polynomial ξ is increasing. For the special case $m = 0$ the distribution has constant failure rate, i.e. it is a geometric distribution. With regard to the shape of the pmf we have the following result.

Theorem 2.1. *For* $t \in \mathrm{N}_0$,

$$p(t+1) - p(t) < 0 \Longleftrightarrow e^{-\xi(t+1)} - e^{-\xi(t)} + 1 > 0. \tag{5}$$

Hence p *is decreasing if* ξ *is decreasing or if* $0 \leq \xi(t) \leq \xi(t+1)$. *In the case* $m = 1$ *the pmf is unimodal with the mode at the point*

$$t_0 = \max\left\{0, \left[\frac{\ln(1-e^{-\alpha_1}) - \alpha_0}{\alpha_1}\right] + 1\right\},$$

where $[x]$ *denotes the greatest integer not exceeding* x.

Proof. From (2) we obtain

$$p(t+1) = \frac{h(t+1)}{h(t)}\big(1-h(t)\big)p(t),$$

which yields

$$p(t+1) - p(t) = \frac{p(t)}{h(t)}\big[h(t+1)\big(1-h(t)\big) - h(t)\big].$$

Because $p(t)$ and $h(t)$ are positive, we get

$$p(t+1) - p(t) < 0 \Longleftrightarrow h(t+1)\big(1-h(t)\big) - h(t) < 0,$$

from which (5) follows. The result for $m = 1$ is an immediate consequence.

A drawback of the introduced model is that there are no simple expressions for the survivor function, probability mass function, moments, etc. For simulation purposes, however, this seems not to be a severe restriction of the applicability of the model. Moments can be computed from

$$\mathrm{E}(T^p) = \sum_{t=1}^{\infty} \big(t^p - (t-1)^p\big)\, S(t), \tag{6}$$

which follows directly from the well-known formula

$$\mathrm{E}\big(|X|^p\big) = \int_0^\infty p\, t^{p-1} \mathrm{Pr}\big(|X| > x\big)\, \lambda(dx),$$

where λ denotes Lebesgue measure. Using (6) we also obtain the existence of all moments, i.e.

$$\mathrm{E}(T^p) < \infty \qquad (p \in \mathrm{N}). \tag{7}$$

3. Maximum Likelihood Estimation

Let f_t denote the observed frequency of the value t, $t \in \mathrm{N}_0$, in a sample $(x_1, \ldots, x_n)$ of size n and define $n_t = \sum_{s \geq t} f_s$. Here, $x_1, \ldots, x_n$ are the realized values of random variables $X_1, \ldots, X_n$ which are independent and identically distributed (iid) according to p. From the definition of the likelihood function

$$L(\theta) = \prod_{i=1}^{n} p(x_i) = \prod_{t=0}^{\infty} p(t)^{f_t}$$

we obtain

$$L(\theta) = \prod_{t=0}^{\infty} \big[h(t)^{f_t}\big(1-h(t)\big)^{n_t - f_t}\big] = \prod_{t=0}^{\infty} \big[G\big(\xi(t)\big)^{f_t}\, G\big(-\xi(t)\big)^{n_t - f_t}\big].$$

Hence, the log-likelihood function $\mathcal{L}(\theta) = \log L(\theta)$ is given by

$$\begin{aligned} \mathcal{L}(\theta) &= \sum_{t=0}^{\infty} \big[f_t \log G\big(\xi(t)\big) + (n_t - f_t) \log G\big(-\xi(t)\big)\big] \\ &= \sum_{t=0}^{\infty} \big[f_t\, \xi(t) - n_t \log\big(1 + e^{\xi(t)}\big)\big]. \end{aligned} \tag{8}$$

Differentiation yields the score vector

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_k} = \sum_{t=0}^{\infty} t^k \big[f_t - n_t\, G\big(\xi(t)\big)\big] \qquad (k = 0, \dots, m),$$

and the elements I_{jk} of the observed information matrix $\mathbf{I}(\theta)$

$$I_{jk} = -\frac{\partial \mathcal{L}(\theta)}{\partial \theta_j \partial \theta_k} = \sum_{t=0}^{\infty} \big[t^{j+k} n_t\, G\big(\xi(t)\big)\, G\big(-\xi(t)\big)\big] \qquad (j, k = 0, \dots, m).$$

Lemma 3.1. *Let* $k_0 = \max\{t \in \mathrm{N}_0 : f_t > 0\}$, *i.e.* k_0 *is the largest value which occurs in the sample. Then, for the determinant of the observed information matrix, we have*

$$\det \mathbf{I}(\theta) = 0 \quad \textit{if } m > k_0,$$

and

$$\det \mathbf{I}(\theta) > 0 \quad \textit{if } 0 \le m \le k_0.$$

Furthermore, $\mathbf{I}(\theta)$ *is positive definite in the second case.*

Proof. Putting

$$c(t) = n_t\; G\big(\xi(t)\big)\; G\big(-\xi(t)\big), \tag{9}$$

we have

$$I_{jk} = \sum_{t=0}^{k_0} t^{j+k} c(t) \qquad (j, k = 0, \dots, m).$$

Note that $c(t) > 0$ for $0 \le t \le k_0$. $\mathbf{I}(\theta)$ is the usual product of the $(m+1) \times (k_0+1)$ matrix $\mathbf{A} = (a_{ij})$, where $a_{ij} = c(j)\, j^i$ and the $(k_0+1) \times (m+1)$ matrix $\mathbf{B} = (b_{ij})$, where $b_{jk} = j^k$. Now, if $m > k_0$, we can define a square matrix $\bar{\mathbf{A}}$ by adding $(m - k_0)$ zero columns to $\mathbf{A}$ and another square matrix $\bar{\mathbf{B}}$ by adding $(m - k_0)$ zero rows to $\mathbf{B}$. This does not affect the result of the multiplication, i.e. we still have $\mathbf{I}(\theta) = \bar{\mathbf{A}}\bar{\mathbf{B}}$. By means of the multiplication rule for determinants, it immediately follows that

$$\det \mathbf{I}(\theta) \;=\; \det \bar{\mathbf{A}} \, \det \bar{\mathbf{B}} \;=\; 0.$$

If $m \le k_0$, we can apply a rule for computing the determinant of a matrix which is the product of two rectangular matrices (see e.g. Kowalewski, 1948 p. 66). This yields

$$\begin{aligned} \det \mathbf{I}(\theta) &= \sum_{i_0 < \cdots < i_m} \det \left[\begin{pmatrix} c(i_0) & \cdots & c(i_m) \\ i_0 c(i_0) & \cdots & i_m c(i_m) \\ \vdots & & \vdots \\ i_0^m c(i_0) & \cdots & i_m^m c(i_m) \end{pmatrix} \begin{pmatrix} 1 & i_0 & \cdots & i_0^m \\ 1 & i_1 & \cdots & i_1^m \\ \vdots & \vdots & & \vdots \\ 1 & i_m & \cdots & i_m^m \end{pmatrix} \right] \\ &= \sum_{i_0 < \cdots < i_m} c(i_0) \cdots c(i_m) \, \det \mathbf{V}(i_0, \dots, i_m)' \, \det \mathbf{V}(i_0, \dots, i_m), \end{aligned}$$

where $i_0, \dots, i_m \in \{0, \dots, k_0\}$ and $\mathbf{V}(i_0, \dots, i_m)$ denotes the Vandermonde matrix of $i_0, \dots, i_m$. From $\det \mathbf{V}(i_0, \dots, i_m) = \prod_{k=0}^{m} \prod_{j=0}^{k-1} (i_k - i_j)^2$ we obtain

$$\det \mathbf{I}(\theta) = \sum_{i_0 < \cdots < i_m} \prod_{k=0}^{m} c(i_k) \prod_{k=0}^{m} \prod_{j=0}^{k-1} (i_k - i_j)^2 \;>\; 0. \tag{10}$$

Since this holds for arbitrary $m \le k_0$, the same is valid for all sub-determinants of $\mathbf{I}(\theta)$. Hence the last assertion of the lemma follows.

Corollary 3.2. *If $m \le k_0$, the inverse of the observed information matrix is given by*

$$\left[\mathbf{I}(\theta)\right]^{-1}_{ij} = \frac{(-1)^{i+j}}{\det \mathbf{I}(\theta)} \sum_{i_0<\cdots<i_{m-1}} s^{i_0,\ldots,i_{m-1}}_{m-i} s^{i_0,\ldots,i_{m-1}}_{m-j} \prod_{k=0}^{m-1} c(i_k) \prod_{k=0}^{m-1}\prod_{j=0}^{k-1} (i_k - i_j)^2,$$

where $c(t)$ is defined in (9), $\det \mathbf{I}(\theta)$ *is given in (10) and $s_i^{t_1,\ldots,t_m}$ are the elementary symmetric functions of $t_1,\ldots,t_m$, defined by $s_0^{t_1,\ldots,t_m} = 1$ and*

$$s_i^{t_1,\ldots,t_m} = \sum_{\nu_1<\cdots<\nu_i} t_{\nu_1}\cdots t_{\nu_i} \qquad (1 \le i \le m).$$

Proof. Let $\mathbf{I}^{(ij)}$ denote the square matrix which is obtained from $\mathbf{I}(\theta)$ by deleting the ith row and the jth column. Let $\mathbf{A}^{(i)}$ $(\mathbf{B}^{(j)})$ denote the matrix which is obtained from the matrix $\mathbf{A}$ $(\mathbf{B})$ in the proof of Lemma 3.1 by deleting the ith row (jth column). We then have $\mathbf{I}^{(ij)} = \mathbf{A}^{(i)}\,\mathbf{B}^{(j)}$, and proceeding in the same way as in the proof of Lemma 3.1, we obtain

$$\det \mathbf{I}^{(ij)} = \sum_{i_0<\cdots<i_{m-1}} \begin{vmatrix} c(i_0) & \cdots & c(i_{m-1}) \\ \vdots & & \vdots \\ i_0^{i-1}c(i_0) & \cdots & i_m^{i-1}c(i_{m-1}) \\ i_0^{i+1}c(i_0) & \cdots & i_m^{i+1}c(i_{m-1}) \\ \vdots & & \vdots \\ i_0^{m}c(i_0) & \cdots & i_m^{m}c(i_{m-1}) \end{vmatrix} \begin{vmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ i_0^{j-1} & \cdots & i_{m-1}^{j-1} \\ i_0^{j+1} & \cdots & i_{m-1}^{j+1} \\ \vdots & & \vdots \\ i_0^{m} & \cdots & i_{m-1}^{m} \end{vmatrix},$$

$$= \sum_{i_0<\cdots<i_{m-1}} \prod_{k=0}^{m-1} c(i_k) \;\; s^{i_0,\ldots,i_{m-1}}_{m-i}\; s^{i_0,\ldots,i_{m-1}}_{m-j} \left(\det \mathbf{V}(i_0,\ldots,i_{m-1})\right)^2,$$

where $i_0,\ldots,i_{m-1} \in \{0,\ldots,k_0\}$ and the $s_i^{t_1,\ldots,t_m}$ are the elementary symmetric functions of $t_1,\ldots,t_m$ as defined above. Observing that

$$\left[\mathbf{I}(\theta)\right]^{-1}_{ij} = (-1)^{i+j}\,\frac{\det \mathbf{I}^{(ij)}}{\det \mathbf{I}(\theta)},$$

the assertion follows.

Theorem 3.3. *Suppose that $f_t > 0$ for at least $m+2$ integers. Then*

(a) *the log-likelihood function $\mathcal{L}(\theta)$ is a strictly concave function of θ,*
(b) *there is a unique maximum likelihood estimate $\hat{\theta}^n \in \Theta$,*
(c) *$\mathcal{L}(\theta)$ has no other maxima or minima or other stationary points in Θ.*

Proof. Since the matrix of the second derivatives of $\mathcal{L}(\theta)$, which is the negative of the observed information matrix $\mathbf{I}(\theta)$, is negative definite by Lemma 3.1, $\mathcal{L}(\theta)$ is strictly concave and has at most one maximum point.

Let the sequence $(\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(k)}) \in \Theta$, converge to the boundary $\partial\Theta$ of Θ, i.e. $\lim_{k\to\infty} \|\theta^{(k)}\| = \infty$. We prove that, under the above assumptions, $\lim_{k\to\infty} \mathcal{L}(\theta^{(k)}) = -\infty$ holds.

Since $f_t > 0$ for at least $m+2$ integers, there are at least $m+1$ integers $t_0, \ldots, t_m$ such that $f_{t_i} > 0$ and $n_{t_i} - f_{t_i} > 0$. Assuming the existence of a constant $c > 0$ such that $\mathcal{L}(\theta^{(k)}) \geq -c$ for each k, there is a vector $u = (u_0, \ldots, u_m)$ such that for $i = 0, \ldots, m$, and for each k

$$-u_i \leq \theta_0^{(k)} + \cdots + \theta_m^{(k)} t_i^m + \varepsilon t_i^{m+1} \leq u_i$$

(otherwise (8) shows that $\mathcal{L}(\theta^{(k)})$ would tend to $-\infty$). This is equivalent to

$$\|\mathbf{V}(t_0, \ldots, t_m)\, \theta^{(k)}\| \leq \|u'\|$$

for some vector u', where $\mathbf{V}(t_0, \ldots, t_m)$ is the Vandermonde matrix of $t_0, \ldots, t_m$. Since $\mathbf{V}(t_0, \ldots, t_m)$ is regular, this implies

$$\|\theta^{(k)}\| \leq \|\mathbf{V}(t_0, \ldots, t_m)^{-1} u'\|$$

for each k, a contradiction. Hence the theorem is proved.

Remarks.

1. Note that the requirement of Theorem 3.3 is met with probability tending to 1 as the sample size tends to infinity.
2. In the condition, it is not possible to replace $m+2$ by $m+1$ as the number of parameters would suggest. For instance, let $m = 1$, $f_0, f_1 > 0$ and $f_t = 0, t > 1$; hence $n_0 - f_0 = f_1 > 0$ and $n_t - f_t = 0, t > 0$. Then

$$\mathcal{L}(\theta) = f_0 \log\big(G(\theta_0)\big) + f_1 \log\big(G(-\theta_0)\big) + f_1 \log\big(G(\theta_0 + \theta_1)\big)$$

has no stationary points.

In what follows we have to check some regularity conditions to obtain standard asymptotic results for the maximum likelihood estimator (MLE) $\hat{\theta}^n$. Write $p(\cdot, \theta)$ instead of $p(\cdot)$ to make the dependence on the parameter vector explicit. Throughout the following, assume that $X, X_1, \ldots, X_n$ are iid $p(\cdot, \theta), \theta \in \Theta$. The distributions P_θ corresponding to $p(\cdot, \theta)$ are identifiable, i.e. $\theta \neq \theta'$, $\theta, \theta' \in \Theta$ implies $P_\theta \neq P_{\theta'}$. Moreover, the distributions have common support for all $\theta \in \Theta$, and all third derivatives of $p(\cdot, \theta)$ with respect to θ exist for all $\theta \in \Theta$.

To prove that further standard regularity conditions in the context of maximum likelihood estimation hold, note that

$$\log p(x, \theta) = \log G\big(\xi(x)\big) + \sum_{s=0}^{x-1} \log G\big(-\xi(s)\big),$$

$$\frac{\partial \log p(x, \theta)}{\partial \theta_j} = x^j - \sum_{s=0}^{x} s^j G\big(\xi(s)\big), \tag{11}$$

$$\frac{\partial^2 \log p(x, \theta)}{\partial \theta_j \partial \theta_k} = \sum_{s=0}^{x} s^{j+k} G\big(\xi(s)\big) G\big(-\xi(s)\big), \tag{12}$$

$$\frac{\partial^3 \log p(x, \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} = \sum_{s=0}^{x} s^{i+j+k} G\big(\xi(s)\big) G\big(-\xi(s)\big)\Big(G\big(\xi(s)\big) - G\big(-\xi(s)\big)\Big), \tag{13}$$

where $i, j, k \in \{0, \dots, m\}$. By (11), $\mathrm{E}_\theta[\partial \log p(X,\theta)/\partial\theta_j] = 0$ for $j = 0, \dots, m$, since

$$\mathrm{E}_\theta\Big[\sum_{s=0}^{X} s^j G\big(\xi(s)\big)\Big] = \sum_{x=0}^{\infty}\Big[\sum_{s=0}^{x} s^j G\big(\xi(s)\big)\Big] p(x,\theta) = \sum_{s=0}^{\infty} s^j G\big(\xi(s)\big) S(s) = \mathrm{E}_\theta[X^j],$$

where $\mathrm{E}_\theta\left[X^j\right]$ was shown to be finite in (7). Next, we have (cf. (12))

$$\begin{aligned} -\mathrm{E}_\theta\Big[\frac{\partial^2 \log p(X,\theta)}{\partial\theta_j\partial\theta_k}\Big] &= \sum_{x=0}^{\infty}\sum_{s=0}^{x} s^{j+k} G\big(\xi(s)\big) G\big(-\xi(s)\big) p(x,\theta) \\ &= \sum_{s=0}^{\infty} s^{j+k} G\big(\xi(s)\big) G\big(-\xi(s)\big) S(s) \\ &= \sum_{s=0}^{\infty} s^{j+k} G\big(-\xi(s)\big) p(s,\theta) \\ &= \mathrm{E}_\theta\big[X^{j+k} G\big(-\xi(X)\big)\big] \qquad (j,k = 0,\dots,m), \end{aligned}$$

where the existence of the last expectation follows again from (7). The same result holds for the elements $I_{jk}(\theta)$ of the Fisher information matrix $\mathbf{I}_F(\theta)$, defined by

$$I_{jk}(\theta) = \mathrm{E}_\theta\Big[\Big(\frac{\partial}{\partial\theta_j}\log p(X,\theta)\Big)\Big(\frac{\partial}{\partial\theta_k}\log p(X,\theta)\Big)\Big].$$

Hence they satisfy

$$I_{jk}(\theta) = -\mathrm{E}_\theta\left[\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log p(X,\theta)\right].$$

For fixed $\theta \in \Theta$, define matrices $\mathbf{I}^r(\theta), r = 1, 2, \dots,$ with elements

$$I^r_{jk}(\theta) = \sum_{s=0}^{r} s^{j+k} G\big(-\xi(s)\big) p(s,\theta) \qquad (j,k = 0,\dots,m).$$

In the same way as in the proof of Lemma 3.1 it follows that

$$0 < \det \mathbf{I}^m(\theta) < \det \mathbf{I}^{m+1}(\theta) < \cdots .$$

Since $\lim_{r\to\infty} \mathbf{I}^r(\theta) = \mathbf{I}_F(\theta)$ entails $\lim_{r\to\infty} \det \mathbf{I}^r(\theta) = \det \mathbf{I}_F(\theta)$, we see that $\mathbf{I}_F(\theta)$ is positive definite.

To show the existence of functions $M_{ijk}(x)$ such that for all $i, j, k = 0, \dots, m$,

$$\sup_{\theta\in\Theta}\left|\frac{\partial^3 \log p(x,\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| \le M_{ijk}(x),$$

where $\mathrm{E}_\theta\big(M_{ijk}(X)\big) < \infty$, recall that $0 < G(\cdot) < 1$; therefore (13) implies

$$\sup_{\theta\in\Theta}\left|\frac{\partial^3 \log p(x,\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| \le \sum_{s=0}^{x} s^{i+j+k} = M_{ijk}(x).$$

To see that the expectation of $M_{ijk}(X)$ is finite note that

$$\mathrm{E}_\theta\Big[\sum_{s=0}^{X} s^{i+j+k}\Big] = \sum_{s=0}^{\infty} s^{i+j+k} S(s),$$

where the series on the right side converges.

Since the standard regularity conditions for maximum likelihood estimation are fulfilled, we have the following theorem; for a proof see Lehmann (1983 Theorem 6.4.1).

Theorem 3.4. *For the MLE $\hat{\theta}^n = \hat{\theta}^n(X_1, \dots, X_n)$ which exists and is unique under the assumption of Theorem 3.3, we have:*

(a) $\hat{\theta}^n$ *is consistent, i.e. for all $\epsilon > 0$ and $j = 0, \dots, m$*

$$\Pr(|\hat{\theta}_j^n - \theta_j| > \epsilon) \longrightarrow 0 \quad \text{as } n \to \infty;$$

(b) $\sqrt{n}(\hat{\theta}^n - \theta)$ *is asymptotically normal with mean zero and covariance matrix $[\mathbf{I}_F(\theta)]^{-1}$ and has the representation*

$$\sqrt{n}(\hat{\theta}^n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \ell(X_i, \theta) + \epsilon_n,$$

where ϵ_n converges in probability to zero as $n \to \infty$, and the vector ℓ is the product of the score vector times the inverse of the information matrix, i.e.

$$\ell_j(X_i, \theta) = \sum_{k=0}^{m} \left[\frac{\partial \log p(X_i, \theta)}{\partial \theta_k} \bigg|_{\theta} [\mathbf{I}_F(\theta)]_{jk}^{-1} \right] \qquad (j = 0, \dots, m);$$

(c) $\hat{\theta}_j^n$ *is asymptotically efficient, i.e.*

$$\sqrt{n}(\hat{\theta}_j^n - \theta_j) \xrightarrow{\mathrm{d}} \mathrm{N}(0, [\mathbf{I}_F(\theta)]_{jj}^{-1}) \qquad (j = 0, \dots, m),$$

where $\xrightarrow{\mathrm{d}}$ denotes convergence in law.

3.1. Numerical Computation of the Maximum Likelihood Estimator

In order to find the MLE $\hat{\theta}^n$ one has to maximize $\mathcal{L}(\theta)$ or one has to find the root of the MLE equations $\nabla_\theta \mathcal{L}(\theta) = 0$. In both cases usually some variant of the Newton algorithm is applied. Since the θ_j in general decrease in j, the problem is badly scaled. This is particularly the case if one uses the following step-up procedure for the selection of the degree of the polynomial $\xi(\cdot)$ (cf. Adams & Watson, 1989 Sect. 3.3). The composite null hypothesis $\theta \in \Theta_0 = \{\theta \in \Theta : \theta_m = 0\}$ within the parametric model can be tested using the likelihood ratio statistic

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)}{\sup_{\theta \in \Theta} \mathcal{L}(\theta)}$$

or the statistic $\lambda = -2 \log \Lambda$, which is asymptotically distributed as χ_1^2 under the above regularity conditions. The degree of the polynomial is successively increased until two consecutive terms are non-significant. The degree is then chosen to be the last value for which a significant result has been obtained. It is obvious that in the last two steps of this procedure the highest coefficients of the polynomial are nearly zero.

For this reason it is necessary to get good starting points for the search algorithm. As a first trial one could equate the hazard function $h(t) = G(\xi(t))$ with the empirical hazard $\bar{h}_t = f_t/n_t$ for $t = 0, \dots, m$. This leads to the first approximation $\bar{\theta}^n = \mathbf{V}(0, \dots, m)^{-1} \bar{\xi}$ where $\bar{\xi}_t = \log\left(f_t/(n_t - f_t)\right)$, $t = 0, \dots, m$. However, this procedure gives poor results in most cases.

Therefore, $h(t)$ and $\bar{h}_t$ should be equated for all t with $n_t > f_t > 0$. This results in an over-determined system of equations for the polynomial $\bar{\xi}$ of degree $m+1$ for which a least squares solution can be obtained. Evaluating the polynomial at $m+1$ points leads, in the same way as above, to an approximation to $\hat{\theta}^n$ which is very satisfactory in our experience.

Even more important is an appropriate rescaling of the problem in such a way that all components of the solution vector are of the same order of magnitude. We suggest rescaling with $\theta_j' = 10^j \theta_j$ $(j = 0, \ldots, m)$, which permits the numerical solution of all optimization problems in the next section without difficulties. Without rescaling, we found that the optimization algorithm (we used algorithm E04LBF of the NAG library) needed several thousand iteration steps for a satisfactory solution or even terminated without result if the degree of the polynomial was 5 or more.

4. Goodness-of-fit Testing and Data Analysis

Let $X, X_1, \ldots, X_n, \ldots$ be a sequence of iid random variables taking non-negative integer values. The problem is to test the hypothesis H_0 that the unknown distribution of X belongs to the class of discrete failure time distributions with pmf $p(\cdot, \theta)$, $(\theta \in \Theta)$.

For this purpose, we use the Cramér–von Mises statistic for discrete data

$$C_n = n \sum_{k=0}^{\infty} \left[F_n(k) - F(k, \hat{\theta}^n)\right]^2 p(k, \hat{\theta}^n),$$

where $F(\cdot, \theta)$ is the distribution function corresponding to the distribution with pmf $p(\cdot, \theta)$, and $F_n(\cdot)$ is the empirical distribution function of $X_1, \ldots, X_n$, i.e.

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

From Henze (1996 Corollary 3.4), it follows that a limiting null distribution of C_n exists. Note that the representation of $\sqrt{n}(\hat{\theta}^n - \theta)$ in Theorem 3.4(b) together with the foregoing regularity conditions imply Henze (1996 assumption A1) (for details, see Durbin, 1973 Sect. 4); further Henze (1996 assumption A2) can be seen to hold. Since the null distribution of C_n, $H_{n,\theta}(t) = P_\theta(C_n \le t)$ depends on θ, it can be estimated by the following Monte Carlo procedure.

Given $X_1, \ldots, X_n$, compute $\hat{\theta}^n$. Then compute $C_{j,n}^* = C_n(X_{j1}^*, \ldots, X_{jn}^*)$, $1 \le j \le b$, where, conditionally on $X_1, \ldots, X_n$, the random variables $X_{j1}^*, \ldots, X_{jn}^*$, $1 \le j \le b$, are iid with pmf $p(\cdot, \hat{\theta}^n)$. Denoting by $c_{n,b}^*(\alpha)$ the $(1-\alpha)$-quantile of $H_{n,b}^*$, where

$$H_{n,b}^*(t) = \frac{1}{b} \sum_{j=1}^{b} \mathbf{1}\{C_{j,n}^* \le t\}$$

is the empirical distribution function of $C_{1,n}^*, \ldots, C_{b,n}^*$, the hypothesis H_0 is rejected at level α if C_n exceeds $c_{n,b}^*(\alpha)$.

Since $\hat{\theta}^n$ converges in probability to θ, Henze (1996 Theorem 3.6) shows that this bootstrap version of the Cramér–von Mises test has asymptotic level α.

The next result shows that the test based on C_n is consistent against each alternative distribution with finite moment of order $m+2$ having a support of at least $m+2$ points.

The latter condition is a natural requirement that ensures the existence of a unique MLE with probability 1 as $n \to \infty$ (cf. Theorem 3.3).

All quantities computed under the alternative distribution are given an index A, e.g. we write $p_A(\cdot)$ for the pmf of the distribution.

Theorem 4.1. *Let $X_1, X_2, \ldots$ be a sequence of iid random variables from any alternative distribution with finite moment of order $m+2$ and a support $\{x \in \mathrm{N}_0 : p_A(x) > 0\}$ which consists of at least $m+2$ points. Then the power of the test tends to 1 if the number of Monte Carlo samples tends to infinity as $n \to \infty$, i.e. we have*

$$\Pr(C_n > c^*_{n,b}(\alpha)) \to 1 \quad \textit{as} \quad n, b \to \infty.$$

Proof. The assumption on the support implies $0 < h_A(t) < 1$ for at least $m+1$ points. By that it can be seen that

$$\inf_{\theta \in \Theta} \sup_{t \in \mathrm{N}_0} |F_A(t) - F(t, \theta)| > 0,$$

where F_A denotes the distribution function of the alternative distribution. The consistency of the test now follows from Henze (1996 Remark 3.7), provided $\hat{\theta}^n$ converges in probability to some $\theta \in \Theta$. To this end, consider the function

$$g(\theta) = \mathrm{E}_A[\log p(X, \theta)] = \sum_{t=0}^{\infty} \big[p_A(t) \log G\big(\xi(t)\big) + S_A(t+1) \log G\big(-\xi(t)\big)\big].$$

The definitions of G and ξ together with the inequality

$$\sum_{t=0}^{\infty} S_A(t+1) t^{m+1} \le \mathrm{E}_A(X^{m+2}) < \infty$$

show that $g(\theta)$ is finite.

As in Lemma 3.1 (or the proof of the positive definiteness of $\mathbf{I}_F(\theta)$ on p.334, if the distribution has infinite support), it can be seen that the matrix of second derivatives $(g_{jk})_{j,k=0,\ldots,m}$, where

$$g_{jk} = -\sum_{t=0}^{\infty} t^{j+k} S_A(t) G\big(\xi(t)\big) G\big(-\xi(t)\big) \qquad (j, k = 0, \ldots, m),$$

is negative definite (note that, since $S_A(t) > 0$ for at least $m+1$ points t, the assumption of Lemma 3.1 is fulfilled). Hence, by the same reasoning as in the proof of Theorem 3.3, we obtain that $g(\theta)$ is a strictly concave function with a unique maximum point $\bar{\theta}$. Defining

$$g_n(\theta) = \frac{1}{n} \mathcal{L}\big(\theta, (X_i)_1^n\big) = \sum_{t=0}^{\infty} \Big(\frac{f_t}{n} \log G\big(\xi(t)\big) + \frac{n_t - f_t}{n} \log G\big(-\xi(t)\big)\Big),$$

we have $g_n(\theta) \to g(\theta)$ for all $\theta \in \Theta$ almost surely by Glivenko–Cantelli. This implies the almost sure convergence of the MLE $\hat{\theta}^n$ to $\bar{\theta}$.

TABLE 1

Absolute frequencies of the durations of atmospheric circulation patterns 1, 2, 8 and 10

		t																
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	> 15
ACP 1	sp	0	2	10	6	5	2											
	su	1	9	18	18	8	8	4	3	3								
	au	3	6	15	8	9	3	0	4	2	0	0	0	1				
	wi	1	3	4	8	1	7	4	1	0	1	1	0	0	0	0	1	
ACP 2	sp	1	12	15	11	7	10	7	3	0	1	3	1	0	1	0	1	
	su	5	9	21	17	18	10	9	2	4	3	1	0	1	1	0	0	1
	au	3	6	19	18	14	11	8	8	6	4	1	0	2	1	0	1	
	wi	2	10	16	22	15	15	11	7	4	4	1	4	1	1			
ACP 8	sp	2	7	16	12	6	1	0	2	1								
	su	1	4	14	9	6	6	1	2	0	1	0	1					
	au	4	9	11	10	2	1	4	1	0	1							
	wi	2	5	20	6	7	5	4	2	0	0	0	0	0	1			
ACP 10	sp	5	7	16	15	4	4	1	3	2	1	0	0	0	1			
	su	2	9	21	17	11	3	6	2	0	4	1	0	1				
	au	2	10	23	15	10	9	5	2	1	0	1	0	0	0	1	0	1
	wi	4	18	18	15	11	10	2	2									

As mentioned in Section 1, the discrete time failure model under consideration may be incorporated into a space–time model for daily precipitation. A crucial part of this model is a set $\{a_1, \ldots, a_k\}$ of k possible atmospheric circulation patterns (ACPs) following a semi-Markov process. The random duration of a_i is described by a discrete distribution with parameters depending only on i and the season. Since the Generalized Poisson distribution, which was used hitherto for this purpose, turned out to be of limited flexibility, we examined the adequacy of the discrete failure model for fitting the duration of the ACPs. We applied the above-mentioned goodness-of-fit test to various datasets consisting of observed durations of circulation patterns in Central Europe for the period 1951–1989. The ACPs are classified according to the scheme of the German Weather service which distinguishes between 29 different circulation patterns, numbered from 1 to 29.

Table 1 shows the absolute frequencies of the duration of ACPs 1, 2, 8 and 10. In each case, the ACPs are analysed separately for the seasons spring (sp), summer (su), autumn (au) and winter (wi). Note that $t = 0$ corresponds to a duration of one day, etc. The results of the goodness-of-fit tests for ACPs 1 and 2 and for ACPs 8 and 10 are given in Table 2 and Table 3, respectively. There, m is the degree of the polynomial $\xi(\cdot)$, n is the sample size, and $\hat\theta_0, \ldots, \hat\theta_m$ are the MLEs for $\theta_0, \ldots, \theta_m$. The bootstrap sample size b was always taken to be 499 in order to have 500 samples altogether (including the original one). The entry p_C gives the position

$$p_C = 1 + \sum_{j=1}^{b} \mathbf{1}\{C^*_{jn} < C_n\}$$

of C_n within the set of the bootstrap sample values. The hypothesis H_0, which implies that the model is appropriate, is rejected at level α if $p_C > [b(1-\alpha)] + 1$.

The main conclusion is that the proposed model is flexible enough to fit the data in almost all cases. In general, four parameters are necessary to yield a satisfactory approximation. The larger p-values for $m = 4$ show that an additional free parameter yields no further improvement.

TABLE 2

Test results for atmospheric circulation patterns 1 and 2

		ACP 1				ACP 2			
		sp $n=25$	su $n=72$	au $n=51$	wi $n=32$	sp $n=73$	su $n=102$	au $n=102$	wi $n=113$
$m=1$	$\hat\theta_0$	−3.33	−2.32	−1.78	−2.05	−1.89	−1.76	−2.25	−2.35
	$\hat\theta_1$	1.20	0.46	0.23	0.17	0.15	0.12	0.21	0.25
	p_C	428	500	498	498	498	500	499	500
$m=2$	$\hat\theta_0$	−4.40	−2.94	−2.27	−2.73	−2.34	−2.35	−2.78	−2.89
	$\hat\theta_1$	2.36	1.02	0.64	0.56	0.45	0.45	0.53	0.59
	$\hat\theta_2$	−0.25	−0.08	−0.05	−0.03	−0.03	−0.03	−0.03	−0.03
	p_C	424	472	480	395	466	491	473	430
$m=3$	$\hat\theta_0$	−9.83	−4.20	−2.83	−3.51	−2.97	−2.92	−3.36	−3.68
	$\hat\theta_1$	11.01	2.89	1.41	1.30	1.15	0.97	1.11	1.43
	$\hat\theta_2$	−4.14	−0.72	−0.25	−0.18	−0.18	−0.12	−0.15	−0.23
	$\hat\theta_3$	0.516	0.058	0.014	0.008	0.008	0.004	0.007	0.012
	p_C	268	54	290	441	422	110	345	75
$m=4$	$\hat\theta_0$	−9.22	−4.13	−2.90	−3.10	−3.20	−2.91	−3.52	−3.71
	$\hat\theta_1$	9.73	2.73	1.57	0.65	1.52	0.95	1.34	1.48
	$\hat\theta_2$	−3.25	−0.63	−0.33	0.05	−0.32	−0.11	−0.23	−0.25
	$\hat\theta_3$	0.263	0.038	0.026	−0.020	0.025	0.003	0.017	0.015
	$\hat\theta_4$	0.0251	0.0010	−0.0006	0.0010	−0.0006	0.0000	−0.0004	−0.0001
	p_C	296	104	316	461	442	170	301	118

TABLE 3

Test results for atmospheric circulation patterns 8 and 10

		ACP 8				ACP 10			
		sp $n=47$	su $n=45$	au $n=43$	wi $n=52$	sp $n=59$	su $n=77$	au $n=80$	wi $n=80$
$m=1$	$\hat\theta_0$	−1.92	−2.01	−1.56	−1.77	−1.59	−1.89	−1.64	−2.10
	$\hat\theta_1$	0.49	0.31	0.29	0.25	0.18	0.24	0.14	0.55
	p_C	500	500	492	500	500	500	500	382
$m=2$	$\hat\theta_0$	−2.76	−2.79	−1.89	−2.45	−2.03	−2.48	−2.35	−2.33
	$\hat\theta_1$	1.36	0.93	0.66	0.77	0.54	0.71	0.62	0.84
	$\hat\theta_2$	−0.14	−0.08	−0.06	−0.06	−0.04	−0.05	−0.04	−0.06
	p_C	327	455	450	499	497	500	494	375
$m=3$	$\hat\theta_0$	−3.74	−3.86	−2.32	−3.22	−2.55	−3.45	−3.26	−2.71
	$\hat\theta_1$	3.03	2.23	1.55	1.73	1.26	1.93	1.61	1.78
	$\hat\theta_2$	−0.77	−0.43	−0.39	−0.30	−0.23	−0.38	−0.26	−0.49
	$\hat\theta_3$	0.059	0.025	0.030	0.015	0.012	0.022	0.011	0.050
	p_C	222	309	236	492	403	223	396	204
$m=4$	$\hat\theta_0$	−3.02	−3.88	−2.37	−3.12	−2.60	−3.85	−3.32	−2.88
	$\hat\theta_1$	0.55	2.26	1.73	1.54	1.39	2.64	1.69	2.55
	$\hat\theta_2$	1.02	−0.44	−0.50	−0.22	−0.29	−0.70	−0.29	−1.12
	$\hat\theta_3$	−0.370	0.027	0.054	0.002	0.021	0.070	0.015	0.216
	$\hat\theta_4$	0.0319	−0.0001	−0.0015	0.0005	−0.0004	−0.0023	−0.0001	−0.0136
	p_C	161	339	332	488	426	63	416	110

TABLE 4

Results of the likelihood ratio test

	ACP 8, winter		ACP 10, autumn		ACP 10, winter	
m	Log-likelihood	λ_m	Log-likelihood	λ_m	Log-likelihood	λ_m
0	– 120.037		– 190.190		– 173.664	
1	– 113.411	13.25	– 184.472	11.44	– 152.582	42.16
2	– 108.013	10.80	– 175.384	18.18	– 151.871	1.42
3	– 104.605	6.82	– 166.158	18.45	– 150.137	3.47
4	– 104.554	0.10	– 166.137	0.04	– 149.641	0.99
5	– 101.767	5.57	– 165.391	1.49	– 149.218	0.85

At the 10% level and for $m = 3$, we reject H_0 only in one case, namely for the winter data of ACP 8. Looking at the observed frequencies of this dataset, the sharp peak for $t = 2$ (which corresponds to a duration of the ACP of 3 days) seems to cause the strong rejection of the failure model in this case. This presumption is confirmed by inspection of the observed and fitted frequencies. Increasing the degree of the polynomial yields $p_C = 496$ for $m = 5$ and $p_C = 472$ for $m = 6$. Hence seven parameters are necessary if the model is not to be rejected at the 5% level.

Here, we want to compare the results of the goodness-of-fit test with the step-up procedure described in Section 3.1. Table 4 shows the value of the log-likelihood function and the test statistic λ_m for three selected cases.

The results are in good agreement with the results of the goodness-of-fit test. For the winter data of ACP 8 we do not find any two successive terms which are non-significant at the 5% level up to $m = 5$. A further increase of the degree of the polynomial yields $\lambda_6 = 6.00, \lambda_7 = 0.66, \lambda_8 = 0.66$. Hence the step-up procedure suggests the use of seven parameters.

For the autumn data of ACP 10, the values of λ_m clearly indicate that 4 parameters are necessary. In the last case, the winter data of ACP 10, the results are somewhat ambiguous. At the 10% level, a linear polynomial is sufficient for ξ, whereas at the 5% level, $m = 3$ is necessary. This coincides with the goodness-of-fit test with estimated p-values of 0.76 for $m = 1$ and 0.41 for $m = 3$.

However, it should be clear that the likelihood ratio step-up procedure provides only a guideline about the selection of the number of parameters but says nothing about the actual fit of the data. Furthermore, since the maximum likelihood principle inevitably leads to choosing the highest possible dimension, one has to introduce a termination criterion which seems to be a little arbitrary. Instead of the step-up procedure described above one could also use Akaike's (1974) or Schwarz's (1978) selection rules.

In contrast to this, higher p-values in the case of a greater number of parameters clearly point out an overspecification of the model. In addition, the test indicates whether the model is compatible with the data. For these reasons, we favour the goodness-of-fit test as a model selection tool of the likelihood ratio procedure.

Since $\sqrt{n}(\hat{\theta}^n - \theta)$ tends in law to a normal distribution, $\sqrt{n}\big(\hat{\xi}(t) - \xi(t)\big)$ also tends to a normal distribution with asymptotic variance

$$\sigma_t^2 = (1, t, \dots, t^m)' \, [\mathbf{I}_F(\theta)]^{-1} \, (1, t, \dots, t^m).$$

Similarly, $\sqrt{n}\big(\hat{h}(t) - h(t)\big)$, where $\hat{h}(t) = G\big(\hat{\xi}(t)\big)$ is the estimated hazard function, has an asymptotic normal distribution with variance $g\big(\xi(t)\big)^2 \, \sigma_t^2$.

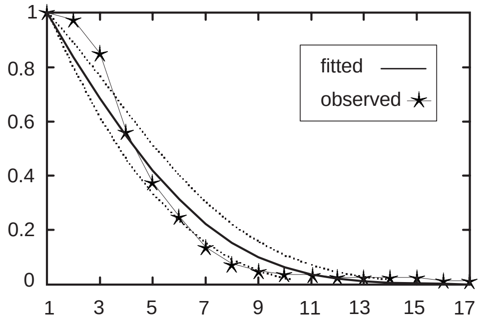

Fig. 1. Survivor function of autumn data for ACP 10, $m = 1$

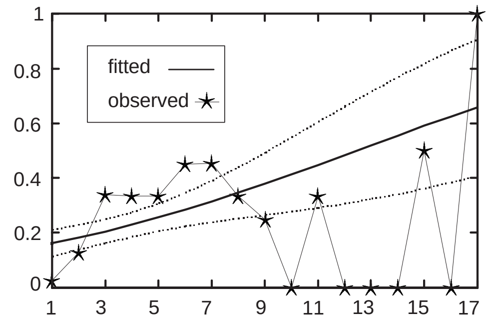

Fig. 2. Hazard function of autumn data for ACP 10, $m = 1$

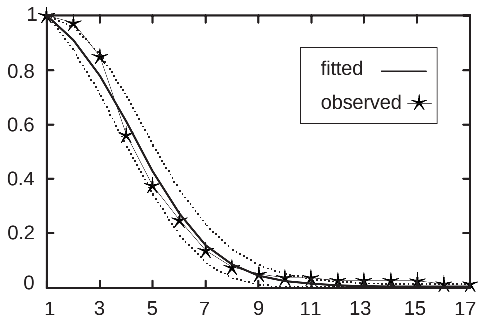

Fig. 3. Survivor function of autumn data for ACP 10, $m = 2$

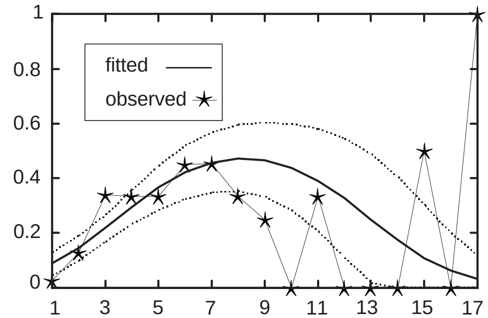

Fig. 4. Hazard function of autumn data for ACP 10, $m = 2$

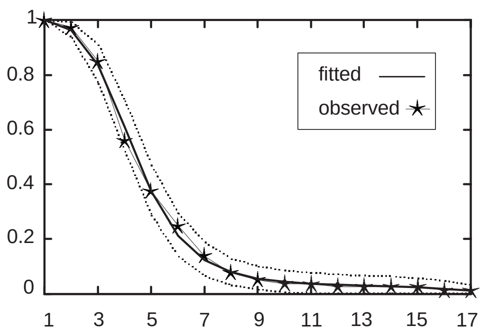

Fig. 5. Survivor function of autumn data for ACP 10, $m = 3$

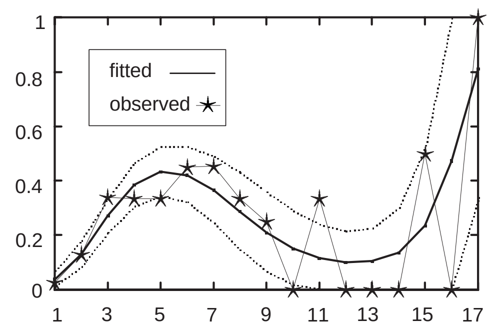

Fig. 6. Hazard function of autumn data for ACP 10, $m = 3$

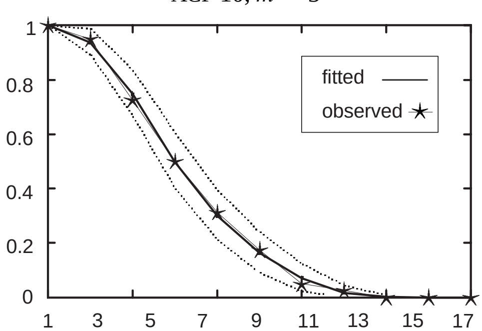

Fig. 7. Survivor function of winter data for ACP 10, $m = 3$

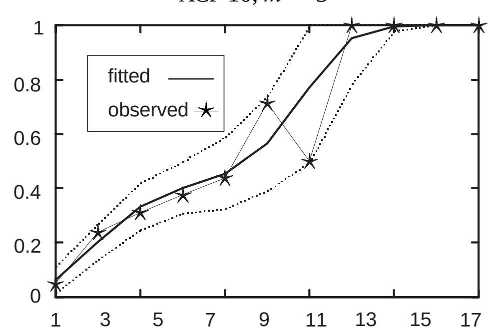

Fig. 8. Hazard function of winter data for ACP 10, $m = 3$

The Fisher information matrix $\mathbf{I}_F(\theta)$ can be estimated by $\mathbf{I}(\hat{\theta}^n)/n$ with the observed information matrix $\mathbf{I}(\theta)$, so an estimate for σ_t^2/n is

$$\hat{\sigma}_t^2 = (1, t, \ldots, t^m)' \, [\mathbf{I}(\hat{\theta}^n)]^{-1} \, (1, t, \ldots, t^m).$$

Therefore, a two-sided approximate confidence interval for $h(t)$ at the 95% level is given by

$$\hat{h}(t) \, \pm \, 1.96 \, g\big(\hat{\xi}(t)\big) \, \hat{\sigma}_t.$$

Similarly, an estimate of the survivor function is given by

$$\hat{S}(t) = \prod_{s=0}^{t-1} G\big(\hat{\xi}(s)\big).$$

Adams & Watson (1989 Sect. 3.2) give an approximation for the variance of $\sqrt{n}\big(\hat{S}(t) - S(t)\big)$.

Figures 1, 3, 5 and 7 show the estimated and observed survivor function together with the 95% confidence bands for the autumn data of ACP 10 $(m = 1, \ 2 \text{ and } 3)$ and for the winter data of ACP 10 $(m = 3)$. Figures 2, 4, 6 and 8 illustrate the estimated and observed hazard function for the same datasets. The observed values are marked with an asterisk, adjacent points connected by lines. For better visualization, the estimates and the confidence intervals are plotted as lines, but they are meaningful only at the discrete points $1, 2, 3, \ldots$. The figures are a confirmation of the results of the testing procedure.

References

ADAMS, G. & WATSON, R. (1989). A discrete time parametric model for the analysis of failure time data. *Austral. J. Statist.* **31**, 365–384.

AKAIKE, H. (1974). A new look at the statistical identification model. *IEEE Trans. Automat. Control* **19**, 716–723.

BÁRDOSSY, A. & PLATE, E.J. (1992). Space–time model for daily rainfall using atmospheric circulation patterns. *Water Resources Res.* **28**, 1247–1259.

CONSUL, P.C. (1989). *Generalized Poisson Distributions. Properties and Applications.* STATISTICS: Textbooks and Monographs, **99**. New York: Marcel Dekker.

DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279–290.

HENZE, N. (1996). Empirical-distribution-function goodness-of-fit tests for discrete models. *Canad. J. Statist.* **24**, 81–93.

HENZE, N. & KLAR, B. (1995). Bootstrap based goodness of fit tests for the Generalized Poisson Model. *Comm. Statist. A — Theory Methods* **24**, 1875–1896.

JOHNSON, N.L., KOTZ, S. & KEMP, A.W. (1992). *Univariate Discrete Distributions.* New York: J. Wiley & Sons.

KALBFLEISCH, J.D. & PRENTICE, R.L. (1980). *The Statistical Analysis of Failure Time Data.* New York: J. Wiley & Sons.

KOWALEWSKI, G. (1948). *Einführung in die Determinantentheorie.* New York: Chelsea Publishing Company.

LEHMANN, E.L. (1983). *Theory of Point Estimation.* New York: J. Wiley & Sons.

PADGETT, W.J. & SPURRIER, J.D. (1985). On discrete failure models. *IEEE Trans. Reliability* **R-34**, 253–256.

SALVIA, A.A. & BOLLINGER, R.C. (1982). On discrete hazard functions. *IEEE Trans. Reliability* **R-31**, 458–459.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461–464.

XEKALAKI, E. (1983). Hazard functions and life distributions in discrete time. *Comm. Statist. A — Theory Methods* **12**, 2503–2509.