J. Appl. Prob. **34**, 818–822 (1997)
Printed in Israel

LETTERS TO THE EDITOR

Dear Editor,

The extremal index in 10 *seconds*

Introduction. In a recent paper, Smith [3] introduces a method to calculate the extremal index of a stationary Harris chain $\{X_n\}$. Loosely speaking, a stationary sequence $\{X_n\}$ with marginal distribution F has extremal index θ if

$$\boldsymbol{P}\{\max(X_1,\cdots, X_n) \leqq u_n\} - (F(u_n))^{n\theta} \to 0,$$

as $n \to \infty$, for sequences u_n with $F^n(u_n) = c \in (0, 1)$. The main assumption of Smith is that the transition density q of the Harris chain satisfies

$$\lim_{u \to \infty} q(u, u+x) = h(x),$$

for some limiting function h, with $h(x) \geqq 0$ and $\int h(x)dx \leqq 1$.

In this letter we show that there is a simple way to compute the extremal index. The numerical method given here is adapted from the Wiener–Hopf algorithm developed by Grübel [2], designed to calculate the distribution of the stationary waiting time of a stable $G/G/1$ queue.

Implementation. From (2.6)–(2.8) of [3],

$$\theta = \int_{-\infty}^{0} e^x \boldsymbol{P}\{S_1 < x, S_2 < x, \cdots \mid S_0 = 0\} dx, \tag{1}$$

where $S_0 = 0$, S_1, $S_2, \cdots$ is a random walk with stepsize density h. In order to facilitate the use of Grübel's algorithm, consider the random walk $S'_k = -S_k$, with density g, $g(x) = h(-x)$. To avoid trivial cases, assume that

$$\int xg(x)dx > 0. \tag{2}$$

As h may be defective, with missing mass transferred to $-\infty$, g may be defective with mass at ∞, in which case the expectation in (2) is infinite.

Define

$$M = \inf_{k \geqq 0} \{S'_k \mid S'_0 = 0\};$$

condition (2) implies that M is finite. Its distribution can be computed with Grübel's algorithm. From (1), we obtain

$$\theta = \int_0^\infty e^{-x} \boldsymbol{P}\left\{\inf_{k \geqq 1} S'_k > x \mid S'_0 = 0\right\} dx$$

(3)
$$= \int_0^\infty e^{-x} \int_x^\infty \boldsymbol{P}\{M > x-y\} g(y) dy\, dx$$

$$= \int_0^\infty e^{-x} \int_{-\infty}^\infty \boldsymbol{P}\{M > x-y\} g(y) dy\, dx;$$

the last equality follows from $\boldsymbol{P}\{M>0\}=0$.

Note that θ can be expressed as the probability of an event: let Z, Y, and M be independent random variables, with Z exponentially distributed with mean 1, M as defined, and Y a copy of the stepsize of the random walk $S'_0, S'_1, \cdots$. Then

(4)
$$\theta = \boldsymbol{P}\{Y + M - Z > 0\}.$$

This leads to the following algorithm for the computation of θ. Steps 3–6 below are steps (iii)–(viii) in Grübel's algorithm; for details we refer to [2]. Note that steps 1 and 2 differ from the first two steps in Grübel's algorithm: for the $G/G/1$ case the stepsize distribution first has to be computed as the difference of two independent random variables representing an interarrival time and a service time.

Step 1. Discretize the distribution of the stepsize Y. For a large positive integer m the distribution of Y is approximated by the vector p of length $2m$ with

$$p(k) = \boldsymbol{P}\{(k-\tfrac{1}{2})h < Y \leqq (k+\tfrac{1}{2})h\}, \qquad k = -m, -m+1, \cdots, m-1.$$

The gridsize h should be as small as possible, whereas m should be chosen so that $(-mh, mh)$ gives a fair coverage of the range of both Y and Z. For computational efficiency it is advised to take m equal to a power of 2.

Step 2. Calculate the discrete Fourier transform (fft) fp on $2m$ points of the vector p:

$$fp(k) = \sum_{n=-m}^{m-1} p(n) e^{2\pi i k n/2m}, \qquad k = 0, \cdots, 2m-1.$$

For the non-defective case we need the fft of the tailvector r,

$$r(k) = \boldsymbol{P}\{Y > h(k+\tfrac{1}{2})\}, \qquad k = 0, 1, \cdots, m-1,$$
$$= -\boldsymbol{P}\{Y \leqq h(k+\tfrac{1}{2})\}, \qquad k = -m, \cdots, -1,$$

given by

$$fr(k) = \sum_{n=-m}^{m-1} r(n) e^{2\pi i k n/2m}, \qquad k = 0, \cdots, 2m-1.$$

Step 3. Calculate $fs = -\log(fr)$, where $x \to \log x$ denotes the complex logarithm.

Step 4. Calculate the inverse Fourier transform of *fs*:

$$s(k)=\frac{1}{2m}\sum_{n=-m}^{m-1} fs(n)\mathrm{e}^{-2\pi ikn/2m}, \qquad k=-m,-m+1,\cdots,m-1,$$

(in shorthand $s=\mathrm{ifft}(fs)$), and define

$$\begin{aligned} sm(k)&=s(k), \qquad k\leqq 0,\\ &=0, \qquad k>0. \end{aligned}$$

The vector sm is an approximation to the harmonic renewal function of the descending ladder height H^-, corresponding to the random walk generated by Y.

Step 5. Apply the Fourier transform and then the transformation $y\to 1-\exp(-y)$ to obtain the Fourier transform of the (defective) probability mass function fh of the ladder height:

$$fh=1-\exp(-(\mathrm{fft}(sm))).$$

Step 6. Compute the Fourier transform of M from the Fourier transform of the ladder height by

$$fm=(1-fh(0))/(1-fh).$$

Step 7. On the same grid $-mh,\cdots,(m-1)h$ make a discretization of an exponentially distributed random variable Z, with mean 1, and calculate the discrete Fourier transform $fminz$ on $2m$ points of $-Z$. Compute the product $fm\cdot fr\cdot fminz$, this is an approximation to the Fourier transform of $Y+M-Z$.

Step 8. Apply the inverse Fourier transform and sum the probabilities corresponding to positive subscripts to get the extremal index θ. Adding half of the probability at zero generally improves the accuracy.

Note. In the defective case the vector of tail probabilities r is not needed and fs in step 3 can be computed directly from the defective vector p by $fs=-\log(1-fp)$.

Runtimes. We illustrate the method with Example 2 from Smith [3]. In this example $H(z)=\int_{-\infty}^{z} h(x)dx=(1+\mathrm{e}^{-rz})^{1/r-1}$, where $r>1$. For m we use powers of 2: $m=2^k$, $k=8,9,\cdots$; $h=15/m$ gives adequate coverage for r between 2 and 5.

Using a 386 20 MHz personal computer and 386-MATLAB, we obtained the values of θ, for $r=2$ shown in Table 1.

The approximation can be improved considerably by applying a simple extrapolation method. It is conjectured by Grübel in [2] that the approximation of M_h, with gridsize h, is of the form

$$\boldsymbol{P}\{M_h\leqq t\}=\boldsymbol{P}\{M\leqq t\}+ch+o(h),$$

for $h\to 0$. If this is true and the density g of the stepsize of the random walk S'_k is sufficiently smooth, then the same discretization error is present in θ. If we call θ_k the

TABLE 1

m	Computer time	θ	m	Computer time	θ
2^8	2 sec	0.32148	2^{12}	25 sec	0.32809
2^9	4 sec	0.32498	2^{13}	59 sec	0.32831
2^{10}	6 sec	0.32675	2^{14}	173 sec	0.32842
2^{11}	12 sec	0.32764			

approximation of θ, based on $m=2^k$, then $2 * \theta_{k+1} - \theta_k$ should have a discretization error of the order $o(h)$, since the discretization parameter h used to calculate $\boldsymbol{\theta}_{k+1}$ is half the value of the discretization parameter for θ_k. Embrechts *et al.* [1] give a rigorous mathematical derivation of this method, called Richardson's deferred approach to the limit, and apply it to compound distributions.

We found as a rule of thumb that the maximum of h^2 and 10 times the maximum of the missing probability masses can be used as an error upper bound, where h is the smallest of the two grid sizes used for the extrapolation.

The calculation with extrapolation for $m=2^8$ takes only 6 seconds of computing time. It can even be done on an 80286 or 8086 personal computer using PC-MATLAB (the computing time is then ≈ 30 seconds). Table 2 presents the values of θ for $r=2, 3, 4, 5$, which are also included in Smith [3]; as before, we took $h=15/m$.

TABLE 2

r	m	θ	m	θ
2	2^8	0.32848	2^{13}	0.32853
3	2^8	0.15794	2^{13}	0.15806
4	2^8	0.09218	2^{13}	0.09234
5	2^8	0.06024	2^{13}	0.06043

The value 0.0616 for $r=5$ given by Smith seems to be slightly off.

Remark. As noted by the referee, for $r \to 1$ the value of mh has to be increased considerably to obtain a fair coverage of the distribution of the stepsize, hence for r close to 1 the algorithm becomes unstable.

An alternative approximation for θ can be given through

$$\theta' := \lim_{u\to\infty} \boldsymbol{P}\{X_2 < u \mid X_1 > u\}.$$

It is easy to show that θ' is an upper bound for the extremal index θ: $\theta \leqq \theta' \leqq 1$. It depends on the length of the arrays that can be stored efficiently in the working memory at which point one should abandon the algorithm and switch to the approximation θ'. It seems that the algorithm can safely be used in the range $r \geqq 1.01$, provided that mh is chosen in such a way that the missing probability mass of Y and Z is small; this missing mass directly affects the error. For $r=1.01$, $mh=800$ and extrapolating on the values $m=2^{13}$ and $m=2^{14}$ we obtain: $\theta=0.98629$. In this case $\theta'=2^{1/r}-1=0.98632$.

References

[1] EMBRECHTS, P., GRÜBEL, R. AND PITTS, S. M. (1993) Some applications of the fast Fourier transform algorithm in insurance mathematics. *Statist. Neerlandica* **47**, 59–75.

[2] GRÜBEL, R.(1991) Algorithm AS 265: *G/G/1* via fast Fourier transform. *Appl. Statist.* **40**, 355–365.

[3] SMITH, R. L. (1992) The extremal index for a Markov chain. *J. Appl. Prob.* **29**, 37–45.

Yours sincerely,

GERARD HOOGHIEMSTRA
LUDOLF E. MEESTER

Vakgroep SSOR
Faculteit TWI
Mekelweg 4
2628 CD Delft
The Netherlands