Statistica Sinica **23** (2013), 1657-1672
doi:http://dx.doi.org/10.5705/ss.2012.100s

TAIL APPROXIMATION IN MODELS THAT INVOLVE LONG RANGE DEPENDENCE: THE DISTRIBUTION OF OVERFLOWS

Benjamin Yakir

The Hebrew University of Jerusalem

Abstract: Long range dependence in stationary processes of increments corresponds to the situations where the variance of cumulative sums is dominated by the accumulation of the covariances between increments. The Hurst parameter, the exponent of the standard deviation of the sum as a function of the number of increments involved, is a characteristic of long range dependence. Models of long range dependence, models that involve an Hurst parameter $0.5 < H < 1$, are frequently used to model the incoming workload in computer networks and communication.

Consider a Gaussian arrival process with long range dependence, a buffer, and a departure process bounded by the bandwidth. This paper present an analytical approximations of the probability of a buffer overflow within a given time interval. The analysis uses and demonstrates a measure-transformation technique.

Key words and phrases: Cusum, likelihood-ratio identity, long-range dependence, maxima of a random field, overshoot correction.

1. Introduction

Joint work with David Siegmund has shaped my academic career. I am proud to be able to participate in the celebration of his achievements and contribute to this special issue in his honor. This paper demonstrates a technique that David and I have developed to analyze the tail distribution of the maxima of a random field. This technique involves a representation that clarifies the relations between global variability of the field and local fluctuations, including the effect of the "Overshoot" correction. In order not to repeat our previous works, this technique is demonstrated in the context of a model that involves long-range dependence between increments of the random field.

Specifically, we consider the distribution of overflows in a fluid or communication model. The components of the model are a random input process, a buffer, and a bounded output process. The concern is the distribution of buffer overflows. One can formulate the model in terms of the increments of the input process X_t, the bandwidth of the output process r, and the accumulation in the buffer of inputs waiting to be transmitted S_t. The accumulation in the buffer, in discrete time, is a reflective process that satisfies the recursion

$$S_{t+1} = (S_t + X_{t+1} - r)^+ .$$

Consider the time interval $(0, m]$. An overflows takes place over this interval if the event $A = \{\max_{0<t\leq m} S_t \geq x\}$ occurs, where x is the size of the buffer. The event under consideration is closely related to the Cusum statistic that is used in monitoring processes, for example in change-point detection. Indeed, when $S_0 = 0$ the event may be written in a form

$$A = \left\{ \max_{0\leq s<t\leq m} \sum_{i=s+1}^{t} (X_i - r) \geq x \right\} ,$$

that involves maximization over the entire collection of partial sums. This event corresponds to the maximal Cusum statistic exceeding the threshold x over the given interval.

Our goal in this paper is to demonstrate a measure-transformation technique for the development of an analytical approximations to the probability of an overflow. Emphasis is given to the details specific to the application of the technique. The application of other techniques that are required in order to complete the analysis of the problem will only be briefly mentioned.

The characteristic feature of the distribution of overflows in many fluid and communication applications is the long-range dependence between the increments of the input process. This long-range dependence is expressed in the fact that the variance of a sum of random elements is dominated by the contribution of covariances between elements that are far from each other. Consequently, the variance is not proportional to the number of elements involved in the sum but, instead, diverges to infinity like the number of elements to the power $2H$, for some $1/2 < H < 1$. The parameter H is called the Hurst parameter and is a characterization of the long-range dependence.

In the current paper we consider the case where the increments of the input process form a stationary Gaussian process. The covariance structure between the increments of the input process reflect a long-range dependence with Hurst parameter H; an example is the process of increments of a fractional Brownian motion.

The results in this paper may be compared to the results of Heath, Resnick, and Samorodnitsky (1997). They considered a similar model with a different input process. Specifically, they dealt with the situation where the input process is a single On-Off process, with heavy tailed On period, or a sum of finitely many such processes. The input process in the model we consider can be viewed as a limit case that involves input composed of a sum of infinitely many independent On-Off processes (Taqqu, Willinger, and Sherman (1997)). A more direct analysis of a model that involves the sum of a large number of On-Off processes is presented in Chapter 9 of Yakir (2013).

Results parallel to the results of this paper were obtained in Piterbarg (2001). He considered as input the fractional Brownian motion in continuous time and produced an expression for the tail of the distribution that involves the square of the Pickands constant. Below we remark on the connection between our representation and the representation that uses the Pickands constant.

We give a formal statement of the result in terms of the centered process of input increments and in terms of the drift μ, which is the difference between r and the expectation of an input increment. Misusing the notation somewhat, we denote henceforth the process of centered input increments by X.

Assume that the centered process X is stationary with a variance-covariance matrix Σ. Given an interval $t = (t_1, t_2]$, denote by $Y(t) = \sum_{i=t_1+1}^{t_2} X_i$ the partial sum associated with that interval. Observe that

$$\mathrm{Var}\,(Y(t)) = 1_t' \Sigma 1_t := g(|t|) \ ,$$

where 1_t is the indicator of the interval t and $|t| = 1_t' 1_t$ is the number of points in the interval. We assume that, as $t \to \infty$, $g(|t|) = v^2 |t|^{2H} + o\big(|t|^{4H-2}\big)$ for some $0.5 < H < 1$ and some constant v^2. Moreover, we also assume that the derivative of the function g satisfies $g'(|t|) = 2Hv^2|t|^{2H-1} + o\big(|t|^{2H-1}\big)$.

Let T be a collection of time intervals. Of primary interest is the event of an overflow associated with the collection T. Denote this event by

$$A = \Big\{ \max_{t \in T} [Y(t) - \mu |t|] \geq x \Big\} \ . \tag{1.1}$$

The main result in this paper corresponds to the situation where the value of x is large, with the cardinality of the set T polynomial in x and T is composed of a grid of intervals. Specifically, let $\delta > 0$ be given. Take

$$T = \Big\{ t = (t_1, t_2] : t_1 = \lfloor j_1 \delta x^{2-1/H} \rfloor, t_2 = \lfloor j_2 \delta x^{2-1/H} \rfloor, 0 \leq j_1 \leq j_2 \leq \frac{m}{\delta x^{2-1/H}} \Big\} \ .$$

The requirement that the cardinality of T is polynomial in x implies that m is polynomial in x. The statement of the main result involves a bounded functional of the standard fractional Brownian motion $\{W_t : -\infty < t < \infty\}$:

$$\lambda_{\mathrm{SPRT}} = \mathrm{E}\Big[\frac{\max_i e^{\sigma W_{\delta i} - 0.5\sigma^2 |\delta i|^{2H}}}{\sum_i e^{\sigma W_{\delta i} - 0.5\sigma^2 |\delta i|^{2H}}} \Big] \ , \tag{1.2}$$

for $\sigma^2 = \mu^{4H} / \{v^2 H^{4H} (1-H)^{2-4H}\}$.

Theorem 1. *Let Y be a centered Gaussian process with stationary increments and let $\mu > 0$ be given. Assume that* $\mathrm{Var}\,(Y(t)) = v^2|t|^{2H} + o\big(|t|^{4H-2}\big)$ *and* $\partial\mathrm{Var}\,(Y(t))/\partial|t| = 2Hv^2|t|^{2H-1} + o\big(|t|^{2H-1}\big)$. *Let A be the event at (1.1) for a grid T that depends on a fixed $\delta > 0$. Then, for m, a high enough power of x,*

$$\frac{x^{5-2/H-2H}}{m}e^{\frac{x^{2(1-H)}\mu^{2H}}{2v^2H^{2H}(1-H)^{2(1-H)}}}\mathrm{P}(A) \underset{x\to\infty}{\longrightarrow} \left[\frac{\lambda_{SPRT}}{\delta}\right]^2 \frac{v^2H^{2H+1/2}}{\mu^{2H+1}(1-H)^{2H-1/2}}\ , \tag{1.3}$$

where λ_{SPRT} is at (1.2).

Examination of the statement of the theorem reveals that, since m is polynomial in x and the leading term in the approximation is converging exponentially fast to zero, the probability of the event A converges to zero. Larger values of m can be dealt with via a Poisson approximation. A Poisson statement can be formulated either through the concept of ϵ-upcrossing in the spirit of Chapter 13 of Leadbetter, Lindgren, and Rootzen (1983). Alternatively, one can divide total time to subintervals of length m and use the method of Arratia, Goldstein, and Gordon (1989) in order to show that the limit distribution of the number of intervals with an overflow is asymptotically Poisson.

The ratio $\lambda_{\mathrm{SPRT}}/(\delta\sigma^{1/H}2^{-1/(2H)})$ converges, as $\delta \to 0$, to a finite functional of the standard fractional Brownian motion defined over the real line. This function produces an alternative representation of the classical Pickands constant. This constant is typically expressed as the limit of the maximal exponentiated process: $\lim_{T\to\infty}(1/T)\mathrm{E}\big[\max_{0\leq t\leq T}\exp\{\sqrt{2}W_t - t^{2H}\}\big]$. The numerical evaluation of the classical representation is notoriously difficult; the alternative representation is a numerically much more stable method of evaluation (See Dieker and Yakir (2012)).

In the rest of the paper we provide a proof of Theorem 1. Before turning to the details of the proof it may be useful to have an outline. Broadly speaking, the analysis is composed of four steps.

A crude approximation. In Theorem 1 the cardinality of the index set T is polynomially large and the marginal probabilities are exponentially small. If the statistics that are maximized in the definition of the set A are not too correlated, in a sense that is made clearer below, then a first order approximation of the probability is obtained in terms of the sum of the marginal probabilities. This first order approximation can be used in order to identify the value(s) in T that maximize the marginal probabilities and the part of the index set that makes a non-negligible contribution to the probability. The maximizer value(s) may be used in order to identify the large deviation rate of the probability (the term in the exponent) and the region that contain non-negligible marginal probabilities may serve for a first localization of the computation. We consider the crude approximation in Section 2.

A measure transformation. The idea behind this step is to represent the problem of computing the probability of a maximum as a sum of expectations; each of the elements in the sum is approximated in a subsequent step and the approximations are summed up in the last step in order to produce the final approximation. The representation as a sum of expectations is obtained via a measure transformation likelihood-ratio identity. Each random variable in the event A is associated with a likelihood ratio and a weighted sum of all likelihood ratios is used in the identity. The measure transformation is carried out in Section 3.

An approximation. Each of the expectations in the representation is a function of a "global" term and a "local" random field. The former is associated with the specific log-likelihood ratio and the latter corresponds to the difference between other log-likelihood ratios and the specific one. The approximation identifies the asymptotic distribution of the local field and the asymptotic independence between it and the global term. The limit is given in terms of a bounded functional of the limit of the local field and in terms of the density of the global element. In particular, this analysis of the local field sheds light on the statement that the index set is not too dense. The local field should be such that the resulting functional is strictly positive. The approximation is discussed in Section 4.

Integration. In the final step the expectations in the sum are replaced by their approximations. The sum may be approximated using integration in order to produce the final approximation of the probability. This step is carried out in Section 5.

2. A Crude Approximation

In the crude approximation step we consider only the marginal probabilities. The first task is to identify the large deviation rate of the access accumulation. Consider the case where the cardinality of the collection T is small in comparison to the reciprocal of the marginal probability of an overflow. In such a case the large deviation rate is determined by the largest marginal probability $\mathrm{P}(Y(t) - \mu|t| \geq x)$, where the maximum is taken over all $t \in T$. In particular, if T is dense enough, the rate may be approximated by the maximum of the marginal probabilities over the continuum of intervals t.

The marginal probability in the normal case is

$$\mathrm{P}(Y(t)-\mu|t|\geq x)=\bar{\Phi}\Big(\frac{x+\mu|t|}{[g(|t|)]^{1/2}}\Big)\sim\bar{\Phi}\Big(\frac{x|t|^{-H}+\mu|t|^{1-H}}{v}\Big)\;,$$

where $\bar{\Phi}$ is the survival function of the standard normal distribution. The marginal probability is maximized when the function

$$x|t|^{-H}+\mu|t|^{1-H} \tag{2.1}$$

is minimized. Taking a derivative with respect to $|t|$ we get that the first derivative is zero when $|\hat{t}|=Hx/[(1-H)\mu]$, the derivative is negative to the left of this solution and positive to the right of it. Therefore, the given value of $|\hat{t}|$ is indeed the minimum. As a result, the maximum value of the marginal probability is

$$\begin{aligned}\max_{t\in T}\mathrm{P}(Y(t)-\mu|t|\geq x)&\sim\bar{\Phi}\Big(\frac{x|\hat{t}|^{-H}+\mu|\hat{t}|^{1-H}}{v}\Big)\\&=\bar{\Phi}\Big(\frac{x^{1-H}\mu^{H}}{vH^{H}(1-H)^{1-H}}\Big)\approx\exp\Big\{-\frac{x^{2(1-H)}\mu(H)}{(1-H)}\Big\}\;,\end{aligned}$$

where

$$\mu(H)=\frac{\mu^{2H}}{2v^{2}H^{2H}(1-H)^{1-2H}}. \tag{2.2}$$

It follows that if the cardinality of T is polynomial in x then

$$x^{-2(1-H)}\log\mathrm{P}\big(\max_{t\in T}\{Y(t)-\mu|t|\}\geq x\big)\underset{x\to\infty}{\longrightarrow}-\frac{\mu(H)}{1-H} \tag{2.3}$$

since the probability on the left-hand side is bounded from below by the maximal marginal probability and bounded from above by the same probability times the cardinality of T. (When T is continuous an approximation by a discrete subset is required; in the Gaussian setting this can be carried out using Fernique's Inequality.)

It is interesting to compare the given rate to the rate obtained in the case $H=1/2$ that is associated with short-range dependence. Then we get that the exponent of x is equal to $2(1-H)=1$, and the constant that multiplies x is equal to $-\mu(H)/(1-H)=-2\mu/v^{2}$, which corresponds to the result that one obtains for the case of independent increments.

In order to identify intervals that make a non-negligible contribution one can examine the marginal probabilities. The intervals of length $Hx/[(1-H)\mu]$ form a ridge in the surface of marginal survival functions. A first order approximation of the probability of interest is the integral of the marginal probabilities; this integral is dominated by the probabilities in the vicinity of the ridge. The first localization step involves the identification of the relevant neighborhood about the ridge, eliminating the need to consider intervals that make a negligible contribution to the probability of an overflow.

We examine the rate of decrease of the marginal probabilities as a function of the deviation of the length of the interval from the maximizing length. The ratio between a marginal probability and the maximal marginal probability is dominated by the difference between the terms that go into the exponent. The first derivative of the function (2.1) that was used in order to find the maximizing interval length vanishes at the maximizer. The second derivative, evaluated at the maximizer, is

$$\frac{H(H+1)x}{|\hat{t}|^{H+2}} - \frac{(1-H)H\mu}{|\hat{t}|^{H+1}} = \frac{[\mu(1-H)]^{H+2}}{[Hx]^{H+1}} ,$$

and a Taylor expansion produces

$$\frac{x|t|^{-H} + \mu|t|^{1-H}}{v} \approx \frac{x^{1-H}\mu^H}{vH^H(1-H)^{1-H}} + \frac{(|t|-|\hat{t}|)^2}{2}\frac{[\mu(1-H)]^{H+2}}{v[Hx]^{H+1}} .$$

Taking the square of the approximation, divided by 2, the leading terms give

$$\frac{(x|t|^{-H} + \mu|t|^{1-H})^2}{2v^2} \approx \frac{x^{2(1-H)}\mu(H)}{(1-H)} + \Big(\frac{|t|-|\hat{t}|}{x^H}\Big)^2 \frac{\mu^{2(H+1)}(1-H)^{2H+1}}{2v^2H^{2H+1}} .$$

It follows that the only intervals that contribute non-negligibly to the approximation are those with lengths in the range $|\hat{t}| \pm O(x^H)$. Henceforth, we restrict T to such intervals.

When independent increments are involved, intervals of length $x/\mu \pm O\big(x^{1/2}\big)$ are of interest. This agrees with the current analysis when $H = 1/2$.

3. A Measure Transformation

The next step uses a likelihood-ratio identity that substitutes the computation of the probability of a rare event by a computation of an expectation of an appropriate function of likelihood ratios over an event that is much more likely to occur. This substitution is produced via a replacement of the underlying probability measure by an alternative one.

The original distribution P of the process of increments X is Gaussian with mean 0 and covariance matrix Σ. For a given interval t and for some scalar θ we consider an alternative distribution P_t that keeps the covariance structure unchanged, but replaces the mean vector by the vector $\theta\Sigma 1_t$. The log-likelihood ratio between the new and the old distributions is

$$\ell_t = \theta 1_t' X - \frac{\theta^2}{2} 1_t'\Sigma 1_t = \theta Y(t) - \frac{\theta^2}{2} g(|t|) .$$

In order to select the value of θ, note that $\mathrm{E}_t[Y(t)] = \theta g(|t|)$. For the maximizer $|\hat{t}| = Hx/[(1-H)\mu]$ we have that the expected value is $x + \mu|\hat{t}|$ iff

$$\theta = \frac{x + \mu|\hat{t}|}{g(|\hat{t}|)} = \frac{x^{1-2H}\mu^{2H}}{v^2 H^{2H}(1-H)^{1-2H}} = 2\mu(H)x^{1-2H} \ .$$

In the sequel we use this value of θ, it depends on x but is independent of t.

The event A can be rewritten in terms of the given log-likelihood ratios as

$$A = \Big\{ \max_{t\in T}[\ell_t + \frac{\theta^2 g(|t|)}{2} - \theta\mu|t|] \geq \theta x \Big\} = \Big\{ \max_{t\in T}[\ell_t + k(|t|)] \geq \theta x \Big\} \ ,$$

where

$$k(|t|) = \frac{\theta^2 g(|t|)}{2} - \theta\mu|t| \ .$$

A different representation of the probability of the event may be obtained via a measure transformation. The likelihood ratio that we use for this transformation is $\sum_{t\in T}\exp\{\ell_t + k(|t|)\}$ divided by the cardinality of T. The resulting likelihood ratio gives

$$\begin{aligned} \mathrm{P}(A) &= \frac{1}{|T|}\sum_{t\in T} e^{k(|t|)}\mathrm{E}_t\Big[\frac{1}{(1/|T|)\sum_{s\in T} e^{\ell_s + k(|s|)}}; A\Big] \\ &= \sum_{t\in T} e^{k(|t|)}\mathrm{E}_t\Big[\frac{e^{-(\ell_t + k(|t|))}}{\sum_{s\in T} e^{\ell_s - \ell_t + k(|s|) - k(|t|)}}; A\Big] \qquad (3.1) \\ &= \sum_{t\in T} e^{k(|t|)-\theta x}\mathrm{E}_t\Big[\frac{M_t}{S_t}e^{-(\ell_t + k(|t|) - \theta x + m_t)}; \ell_t + k(|t|) - \theta x + m_t \geq 0\Big] \ , \end{aligned}$$

for

$$S_t = \sum_{s\in T} e^{\ell_s - \ell_t + k(|s|) - k(|t|)} \ , \quad M_t = \max_{s\in T} e^{\ell_s - \ell_t + k(|s|) - k(|t|)}$$

and $m_t = \log M_t$.

The probability of interest is now represented as a weighted sum of expectations, the sum extending over all intervals t that are included in the set T. The weights are of the form $\exp\{k(|t|) - \theta x\}$. The term that is multiplied by a weight is an expectation computed under the tilted distribution associated with the interval t. The expectation involves a functional of the random variable $\tilde{\ell} = \ell_t + k(|t|) - \theta x$, which we call *the global term*, and of the random field $\{\ell_s - \ell_t + k(|s|) - k(|t|) : s \in T\}$, which we call *the local field*.

The subsequent step in the analysis is carried out via a local approximation of each of the expectations in the sum. The final formula for the asymptotic evaluation of the probability results from the summation of the local approximations. We take a closer look at the global terms and the local random fields:

An element of the local process takes the form $w_t(s) = \ell_s - \ell_t + k(|s|) - k(|t|)$. From the definition of the terms involved we get that

$$w_t(s) = \theta[1_s - 1_t]'X - (|s| - |t|)\theta\mu \ .$$

The distribution of this local process is determined by the expectation, the variance, and the covariance between the elements of the process, all computed under the tilted P_t-distribution.

The P_t-expectation of an element is

$$\mathrm{E}_t[w_t(s)] = \theta^2[1_s - 1_t]'\Sigma 1_t - (|s| - |t|)\theta\mu.$$

The evaluation of this expectation calls for the investigation, as a function in s, of the term $[1_s - 1_t]'\Sigma 1_t$. However,

$$[1_s - 1_t]'\Sigma 1_t = \frac{1}{2}\big\{1_s'\Sigma 1_s - 1_t'\Sigma 1_t - (1_s - 1_t)'\Sigma(1_s - 1_t)\big\} \ .$$

Consequently,

$$\mathrm{E}_t[w_t(s)] = -\frac{\theta^2}{2}(1_s - 1_t)'\Sigma(1_s - 1_t) + \frac{\theta^2}{2}\big[g(|s|) - g(|t|) - (|s| - |t|)\theta\mu \ .$$

Moreover, up to a $o\big(x^{2H-1}\big)$ term,

$$\begin{aligned}
\frac{\theta}{2}[g(|s|) - g(|t|)] &= \frac{v^2|t|^{2H-1}\theta}{2}|t|\Big[\Big(\frac{|s|}{|t|}\Big)^{2H} - 1\Big] \\
&= \Big(1 - \frac{|t| - |\hat{t}|}{|\hat{t}|}\Big)^{2H-1} \times \frac{\mu|t|}{2H}\Big[\Big(1 + \frac{|s| - |t|}{|t|}\Big)^{2H} - 1\Big] \\
&= \Big(1 - \frac{|t| - |\hat{t}|}{|\hat{t}|}\Big)^{2H-1} \times \Big(1 + \frac{|\tilde{s}| - |t|}{|t|}\Big)^{2H-1} \times (|s| - |t|)\mu \\
&= (|s| - |t|)\mu + O\big(\big||t| - |s|\big|x^{-(1-H)}\big) \ .
\end{aligned}$$

This last follows from the fact that length $|\tilde{s}|$ belongs to the interval of lengths that are between $|s|$ and $|t|$ and all the lengths are $|\hat{t}| \pm O(x^H)$.

One concludes that

$$\mathrm{E}_t[w_t(s)] = -\frac{\theta^2}{2}(1_s - 1_t)'\Sigma(1_s - 1_t) + O\big(\big||t| - |s|\big|x^{-H}\big) + o(1) \ . \tag{3.2}$$

Comparing the first term in the expectation to the $O(\cdot)$ term, the latter, in the worst case scenario, is of the order of the square root of the former.

The variance is

$$\mathrm{Var}_t[w_t(s)] = \theta^2[1_s - 1_t]'\Sigma[1_s - 1_t] \ , \tag{3.3}$$

which is essentially equal to the absolute value of twice the expectation, and the covariance is

$$\mathrm{Cov}\,_t[w_t(r), w_t(s)] = \theta^2[1_r - 1_t]'\Sigma[1_s - 1_t]\ . \tag{3.4}$$

For a given $t \in T$, The global term corresponds to the random variable $\tilde{\ell} = \ell_t + k(|t|) - \theta x$. The P_t-distribution of this term is normal. The expectation is $\mathrm{E}_t(\tilde{\ell}) = \theta^2 g(|t|) - \theta\mu|t| - \theta x$; it vanishes for all t such that $|t| = |\hat{t}|$ and is of order $O(x^{1-H})$ for lengths in the relevant range.

The variance is $\mathrm{Var}\,_t(\tilde{\ell}) = \theta^2 g(|t|)$, so the standard deviation is

$$\theta[g(|t|)]^{1/2} = x^{1-H} v(\frac{t}{x})^H 2\mu(H) + o(1) = x^{1-H}\frac{\mu^H}{vH^H(1-H)^{1-H}} + o(1)\ . \tag{3.5}$$

4. A Second Localization

In this step we replace each of the expectations in the sum (3.1) by an approximation. The approximation is of the form of a product between a density and an expectation of the ratio of two terms. The density corresponds to the density of the global term evaluated at zero. The two terms that appear in the ratio are the asymptotic versions of the sum element S_t and the maximum element M_t that appear in (3.1).

To proceed, we cite a result that states sufficient conditions for the approximation to hold and check that these conditions are met here. Theorem 5.1 of Siegmund, Yakir, and Zhang (2010) considers a triangular array in which κ is the primary index. It denotes by $\tilde{\ell}_\kappa$ the global element, a random variable that is closely related to the global log-likelihood ratio, and considers M_κ and S_κ. a pair of random variables, measurable with respect to the collection of observations that form a local field.

The theorem has the limit, as $\kappa \to \infty$, of the term

$$\kappa^{1/2}\mathrm{E}\big[(M_\kappa/S_\kappa)\exp[-(\tilde{\ell}_\kappa + \log M_\kappa)]; \tilde{\ell}_\kappa + \log M_\kappa \geq 0\big]\ .$$

The approximation that leads to the limit results from replacement of the quantities M_κ and S_κ by local versions that depend on only a finite number of elements from the local field and are almost independent of $\tilde{\ell}_\kappa$. A local approximation is applied to $\tilde{\ell}_\kappa$, and further approximations are applied to the localized version of M_κ/S_κ. Together these yield the desired limit.

We consider an appropriately selected sequence of σ-fields $\hat{\mathcal{F}}_\kappa$, $\kappa = 1, 2, \ldots$. Let $\hat{M}_\kappa$ and $\hat{S}_\kappa$ be approximations of M_κ and S_κ, respectively, which are measurable with respect to $\hat{\mathcal{F}}_\kappa$. Given $\epsilon > 0$ and for all large κ, we list a set of conditions that imply the second localization step that is summarized in Theorem 2:

Condition 1. Both $0 < M_\kappa \leq S_\kappa < \infty$ and $0 < \hat{M}_\kappa \leq \hat{S}_\kappa < \infty$ hold with probability one.

Condition 2. $\mathrm{E}\big[\hat{M}_\kappa/\hat{S}_\kappa\big]$ converges to $\mathrm{E}\big[\hat{\mathcal{M}}/\hat{\mathcal{S}}\big]$ and $\big|\mathrm{E}\big[\hat{\mathcal{M}}/\hat{\mathcal{S}}\big] - \mathrm{E}[\mathcal{M}/\mathcal{S}]\big| \leq \epsilon_2$, for an appropriate $\mathrm{E}[\mathcal{M}/\mathcal{S}]$.

Condition 3. There exist $\mu \in \mathrm{R}$ and $\sigma \in \mathrm{R}^+$ such that for every $0 < \epsilon_3, \delta$, for any event $E \in \hat{\mathcal{F}}_\kappa$ having boundary measure 0, and for all large enough κ,

$$\sup_{|v| \leq \epsilon\kappa^{1/2}} \Big|\kappa^{1/2}\mathrm{P}\big(\tilde{\ell}_\kappa \in v + (0,\delta], E\big) - \frac{\delta}{\sigma}\phi\Big(\frac{\mu}{\sigma}\Big)\mathrm{P}(E)\Big| \leq \epsilon_3 \ .$$

Condition 4. For an event A let $\nu_\kappa(A|\hat{\mathcal{F}}_\kappa) = \sup\big\{\mathrm{P}(A|\hat{\mathcal{F}}_\kappa, \tilde{\ell}_\kappa) : |\tilde{\ell}_k| \leq \epsilon\kappa^{1/2}\big\}$, where the sup is taken over the indicated values of $\tilde{\ell}_\kappa$. For any $\epsilon_4 > 0$ and for all large enough κ,

$$\mathrm{E}\big[\nu_\kappa\big(\{|\log M_\kappa - \log \hat{M}_\kappa| > \epsilon\} \cup \{|\hat{S}_\kappa/S_\kappa - 1| > \epsilon\}\big|\hat{\mathcal{F}}_\kappa\big)\big] < \epsilon_4 \ . \qquad (4.1)$$

Condition 5. The probabilities $\mathrm{P}\big(|\log M_\kappa| > \epsilon\kappa^{1/2}\big)$, $\mathrm{P}\big(|\log \hat{M}_\kappa| > \epsilon\kappa^{1/2}\big)$, and $\mathrm{P}\big(\log M_\kappa - \log \hat{M}_\kappa < -\epsilon\big)$ are all $o(\kappa^{-1/2})$.

Theorem 2. *If Conditions* 1–5 *hold, then*

$$\lim_{\kappa\to\infty} \kappa^{1/2}\mathrm{E}\big[(M_\kappa/S_\kappa)e^{-(\tilde{\ell}_\kappa + \log M_\kappa)}; \tilde{\ell}_\kappa + \log M_\kappa \geq 0\big] = \sigma^{-1}\phi\big(\frac{\mu}{\sigma}\big)\mathrm{E}[\frac{\mathcal{M}}{\mathcal{S}}] \ .$$

The proof of Theorem 2 can be found in Siegmund, Yakir, and Zhang (2010). In the rest of this section we validate each one of the conditions for the current problem.

For Conditions 1 one may fix $t = (\lfloor j_1\delta x^{2-1/H}\rfloor, \lfloor j_2\delta x^{2-1/H}\rfloor) \in T$. Consider a finite sub-collection of intervals $T_t \subset T$ given by

$$T_t = \{(s_1, s_2] : s_1 = \lfloor i_1\delta x^{2-1/H}\rfloor,\ s_2 = \lfloor i_2\delta x^{2-1/H}\rfloor,\ |i_1 - j_1| \vee |i_2 - j_2| \leq c\} \ ,$$

for some large enough but fixed c.

Take $M_\kappa = M_t$ and $S_\kappa = S_t$. Let $\hat{\mathcal{F}}_\kappa$ be the σ-field generated by the collection $\{\ell_s - \ell_t : s \in T_t\}$. Let

$$\hat{S}_\kappa = \sum_{s\in T_t} e^{\ell_s - \ell_t + k(|s|) - k(|t|)} \ , \quad \hat{M}_\kappa = \max_{s\in T_t} e^{\ell_s - \ell_t + k(|s|) - k(|t|)}.$$

Condition 1 follows from the fact that a sum of a sequence of non-negative random variables dominates the maximal element in the sequence.

The ratio $\hat{M}_\kappa/\hat{S}_\kappa$ is a bounded functional of the elements that produce $\hat{\mathcal{F}}_\kappa$. The expectations of these elements are given in (3.2), the variances are given in (3.3) and the covariance between two elements are presented in (3.4).

For Condition 2 consider the limit distribution of the elements in the local field. Start with the variance. The variance is a function of the difference $1_s - 1_t$, which indicates the symmetric difference between the interval s and the interval t (with either a negative or a positive sign). This symmetric difference is associated with two sub-intervals, one with s_1 and t_1 as endpoints and one with s_2 and t_2 as endpoints. In general, the bi-linear form $[1_s - 1_t]'\Sigma[1_s - 1_t]$ contains three elements, two are themselves bi-linear forms, one for each sub-interval, and a mixed term that involves the two intervals and is multiplied by 2. This mixed term is vanishing in the limit. This fact and the divergence properties of the function g combine to produce

$$\lim_{x\to\infty} \text{Var}_t[w_t(s)] = \sigma_\delta^2\big\{|i_1 - j_1|^{2H} + |i_2 - j_2|^{2H}\big\}\ ,$$

for $\sigma_\delta^2 = 4v^2[\mu(H)]^2\delta^{2H}$. A similar argument gives, for the covariance,

$$\lim_{x\to\infty} \text{Cov}_t[w_t(r), w_t(s)] = \sigma_\delta^2\big\{R(h_1 - j_1, i_1 - j_1) + R(h_2 - j_2, i_2 - j_2)\big\}\ ,$$

where

$$R(h,i) = \frac{1}{2}\big\{|h|^{2H} + |i|^{2H} - |h-i|^{2H}\big\}\ ,$$

multiplied by the sign of $(h \cdot i)$.

Here R is the covariance function of a self-similar two-sided process with increments of unit variance. It follows that the limiting covariance structure of the elements of the field $\hat{\mathcal{F}}_\kappa$ corresponds to a sum of two independent self-similar processes, one associated with the left endpoints of the interval about t_1 and the other with the right endpoints about t_2. The standard deviation of the increments of these processes is σ_δ.

Consider the limit of the expectations. Consult (3.2)where the first term converges to minus one half of the variance. The second term is bounded by a constant times $x^{2-1/H-H}$ and converges to zero since $2 - 1/H - H < 0$, for $0.5 < H < 1$. It follows that the limit expectation of a partial sum that involves $i - j$ increments in either of the self-similar processes is $-0.5\,\sigma_\delta^2|i-j|^{2H}$.

The term $\mathrm{E}\big[\hat{\mathcal{M}}/\hat{\mathcal{S}}\big]$ is the expectation of the ratio between the maximum and the sum of the exponentiated limit field. This limit field can be represented as a sum of two independent and identically distributed processes, one indexed by i_1 and the other by i_2. The result is that the ratio is equal to the product of two ratios, one for each process. A corollary of independence is that the expectation of the product of ratios is the product of the expectations, namely the square of a single expectation.

Denote by λ_{SPRT} the expectation of such a ratio applied to an infinite double ended process,

$$\lambda_{\text{SPRT}} = \mathrm{E}\Big[\frac{\max_i e^{\sigma_\delta W_i - 0.5\sigma_\delta^2 |i|^{2H}}}{\sum_i e^{\sigma_\delta W_i - 0.5\sigma_\delta^2 |i|^{2H}}}\Big] ,$$

for $\sigma_\delta^2 = 4v^2[\mu(H)]^2\delta^{2H}$ and for $\{W_i : -\infty < i < \infty\}$ a standard fractional Brownian motion. This term is the term that appears in (1.2). Notice that this term is the limit, as $c \to \infty$, of parallel terms that are defined by restricting the maximization and summation to a range of radius c about the origin.

The term $\mathrm{E}\big[\mathcal{M}/\mathcal{S}\big] = [\lambda_{\text{SPRT}}]^2$ corresponds to the limit of $\mathrm{E}\big[\hat{\mathcal{M}}/\hat{\mathcal{S}}\big]$, as $c \to \infty$, for c the quantity that bounds the number of terms the produce the local random field.

For Condition 3 take $\tilde{\ell}_\kappa = \tilde{\ell} = \ell_t + k(|t|) - \theta x$. This random variable has a Normal distribution. The expectation is of the order of magnitude x^{1-H} and the standard deviation is given in (3.5). Let E be an event that is defined in terms of the local process. Clearly,

$$\mathrm{P}\big(\tilde{\ell}_\kappa \in v + (0, \delta], E\big) = \int_v^{v+\delta} \mathrm{P}\big(E\big|\tilde{\ell}_\kappa = u\big)\mathrm{P}(\tilde{\ell}_\kappa = u)du .$$

Let $\kappa = x^{2(1-H)}$. From the definition of the expectation and the standard deviation of $\tilde{\ell}$ we get that $\sqrt{\kappa}\mathrm{P}(\tilde{\ell} = u)$ is

$$\Big[vH^H \frac{(1-H)^{1-H}}{\mu^H}\Big]\phi\Big(\frac{\mathrm{E}_t[\tilde{\ell}]}{[\mathrm{Var}_t(\tilde{\ell})]^{1/2}}\Big) .$$

Consequently, in order to show that Condition 3 holds with

$$\sigma = \frac{\mu^H}{vH^H(1-H)^{1-H}} , \tag{4.2}$$

$$\frac{\mu}{\sigma} = \frac{\mathrm{E}_t[\tilde{\ell}]}{[\mathrm{Var}_t(\tilde{\ell})]^{1/2}} , \tag{4.3}$$

it is sufficient to demonstrate that the conditional expectation vector and joint covariance matrix of the the elements of the local field $\hat{\mathcal{F}}_\kappa$, given $\tilde{\ell} = u$, converges to the marginal expectation vector and marginal joint covariance matrix. This convergence should hold uniformly in u, for $|u| \le \epsilon\kappa^{1/2}$.

Consider the conditional distribution of the vector $\{w_t(s)\}$, given $\tilde{\ell} = u$.

Here

$$\begin{aligned}
\mathrm{Cov}_t[w_t(s), \tilde{\ell}] &= \theta^2[1_s - 1_t]'\Sigma 1_t \ , \\
\mathrm{E}_t[w_t(s)|\tilde{\ell} = u] &= \mathrm{E}_t[w_t(s)] - \theta^2[1_s - 1_t]'\Sigma 1_t \frac{\mathrm{E}_t(\tilde{\ell}) - u}{\mathrm{Var}_t(\tilde{\ell})} \ , \\
\mathrm{Cov}_t[w_t(r), w_t(s)|\tilde{\ell} = u] &= \mathrm{Cov}_t[w_t(r), w_t(s)] \\
&\quad - \frac{\theta^4}{\mathrm{Var}_t(\tilde{\ell})}([1_s - 1_t]'\Sigma 1_t)([1_r - 1_t]'\Sigma 1_t) \ .
\end{aligned}$$

Clearly, the local random field is asymptotically independent of the event $\{\tilde{\ell} = u\}$ provided that $\theta^2[1_s - 1_t]'\Sigma 1_t/[\mathrm{Var}_t(\tilde{\ell})]^{1/2} \to 0$, for all s in the index set of the local field. However, the numerator is $O(x^{3-2H-1/H})$ and the denominator is $\Theta(x^{1-H})$. Condition 3 then follows from the fact that $2 - H - 1/H < 0$.

The conditional density of the local random field, conditional on $\tilde{\ell} = u$, is bounded away from zero uniformly in $|u| \le \epsilon\kappa^{1/2}$ and over an event of probability larger than $1 - \epsilon_4$. Therefore, in order to prove Condition 4 it is sufficient to show that:

$$\sup_{|u| \le \epsilon\kappa^{1/2}} \mathrm{P}_t\big(w_t(s) > f_1(|t \bigtriangleup s|)\big|\tilde{\ell} = u\big) \le \exp\{-f_2(|t \bigtriangleup s|)\} \ ,$$

for $t \bigtriangleup s$ the symmetric difference between the intervals and for some diverging functions f_j such that $\sum_{i=1}^{\infty} e^{f_j(i)} < \infty$, $j = 1, 2$. However, if we take

$$f_1(|t \bigtriangleup s|) = -0.5(1 - \epsilon)\mathrm{E}_t[w_t(s)] \ , \quad f_2(|t \bigtriangleup s|) = (1 - \epsilon)\frac{(\mathrm{E}_t[w_t(s)])^2}{8\mathrm{Var}_t[w_t(s)]}$$

then the result follows.

Clearly, $M_\kappa \ge \hat{M}_\kappa \ge 1$. Consequently, for Condition 5 it is sufficient to prove that

$$\mathrm{P}_t\big(\log M_\kappa > \epsilon\sqrt{\kappa}\big) = \mathrm{P}_t\big(M_\kappa > e^{\epsilon\sqrt{\kappa}}\big) \le \sum_{s \in T} \mathrm{P}_t\big(w_t(s) \ge \epsilon\sqrt{\kappa}\big) \ .$$

The probabilities in the sum are of the form of a normal survival function, computed at the point $(\epsilon x^{1-H} - \mathrm{E}_t[w_t(s)])/(\mathrm{Var}_t[w_t(s)])^{1/2}$. Consequently, the probabilities converge to zero at a rate faster than polynomial. If m is not diverging too fast we get that the sum of all probabilities is still $o\big(\kappa^{-1/2}\big)$.

5. Integration

The final step in the approximation of the probability of an overflow results from replacing the expectations in the representation (3.1) by their approximations that are produced in the second localization. The resulting sum is then approximated by an integral.

We identify the exponent that is produced from the combination of the weights and the normal density in the second localization limit. The exponent of the weights are of the form $k(|t|) - \theta x$. The second approximation contributes to the exponent the square of the expectation of the global term $\tilde{\ell}$, divided by twice its variance. Consequently, the exponent following the second localization is of the form

$$\begin{aligned} k(|t|) - \theta x - \frac{[\mathrm{E}_t(\tilde{\ell})]^2}{2\mathrm{Var}\,_t(\tilde{\ell})} &= k(|t|) - \theta x - \frac{(\theta^2 g(|t|)/2 + k(|t|) - \theta x)^2}{2\theta^2 g(t)} \\ &= -\frac{(\theta^2 g(|t|)/2 - k(|t|) + \theta x)^2}{2\theta^2 g(t)} \\ &= -\frac{(\mu|t| + x)^2}{2g(t)} \approx -\frac{(\mu|t|^{1-H} + x|t|^{-H})^2}{2v^2} \,. \end{aligned}$$

The discussion that led to the first localization may be applied to this term. Introducing this to (3.1) gives

$$\begin{aligned} \mathrm{P}(A) &= \sum_{t\in T} e^{k(|t|)-\theta x} \mathrm{E}_t\Big[\frac{M_t}{S_t} e^{-(\ell_t + k(|t|) - \theta x + m_t)}; \ell_t + k(|t|) - \theta x + m_t \geq 0\Big] \\ &\approx e^{-\frac{x^{2(1-H)}\mu(H)}{(1-H)}} [\lambda_{\text{SPRT}}]^2 \frac{1}{\sqrt{2\pi} x^{1-H}\sigma} \sum_{t\in T} e^{-\left(\frac{|t|-|\hat{t}|}{x^H}\right)^2 \frac{\mu^{2(H+1)}(1-H)^{2H+1}}{2v^2 H^{2H+1}}} \\ &\approx e^{-\frac{x^{2(1-H)}\mu(H)}{(1-H)}} [\lambda_{\text{SPRT}}]^2 \frac{\tilde{\sigma}}{x^{1-H}\sigma} \sum_{i=1}^{m/[\delta x^{2-1/H}]} \sum_{|j-\hat{j}|\leq c/\tilde{\sigma}} \frac{\exp\left\{-\frac{(j-\hat{j})^2}{2\tilde{\sigma}^2}\right\}}{\sqrt{2\pi}\tilde{\sigma}} \\ &\approx \frac{m e^{-\frac{x^{2(1-H)}}{\mu(H)}(1-H)}}{x^{5-2/H-2H}} \Big[\frac{\lambda_{\text{SPRT}}}{\delta}\Big]^2 \frac{v^2 H^{2H+1/2}}{\mu^{2H+1}(1-H)^{2H-1/2}} \,, \end{aligned}$$

where c is large constant that depends on the definition of T after the first localization step,

$$\frac{1}{\tilde{\sigma}} = \delta x^{2-1/H-H} \frac{\mu^{H+1}(1-H)^{H+1/2}}{vH^{H+1/2}}$$

and σ is given in (4.2). The last approximation is appropriate for large values of c.

If $H = 1/2$ the power of x is zero and the last constant is of the form $v^2/[2\mu^2]$.

Acknowledgement

This work has been supported by grant No. 325/09 of the Israeli Science Foundation. We would like to thank Ton Dieker for his helpful comments.

References

Arratia, R., Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.* **21**, 9-25.

Dieker, A. B. and Yakir, B. (2012). On asymptotic constants in the theory of Gaussian processes. (arXiv:1206.5840).

Heath, D., Resnick, S. and Samorodnitsky, G. (1997). Patterns of buffer overflow in a class of queues with long memory in the input stream. *Ann. Appl. Probab.* **7**, 1021-1057.

Leadbetter, M. R., Lindgren, G. and Rootzen, H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New York.

Piterbarg, V. I. (2001). Large deviations of a storage process with fractional brownian motion as input. *Extremes* **4**, 147-164.

Siegmund, D. O., Yakir, B. and Zhang, N. R. (2010). Tail approximations for maxima of random fields by likelihood ratio transformations. *Sequential Analysis* **29**, 245-262.

Taqqu, M. S., Willinger, W. and Sherman, R. (1997). Proof of a fundamental result in self-similar traffic modeling. *Computer Comm. Rev.* **27**, 5-23

Yakir, B. (2013). *Extremes in Random Fields: A Theory and Its Applications.* John Wiley and Sons, UK (To appear).

Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel.

E-mail: msby@mscc.huji.ac

(Received March 2012; accepted March 2013)