Tail probabilities for the null distribution of scanning statistics

David Siegmund
Department of Statistics, Stanford University, Stanford, CA 94305

Benjamin Yakir
Department of Statistics, Hebrew University, Jerusalem, Israel

August 19, 1999

SUMMARY

This paper is concerned with statistics that scan a multidmensional spatial region to detect a signal against a noisy background. The background is modeled as independent observations from an exponential family of distributions with a known "null" value of the natural parameter, while the signal is given by independent observations from the same exponential family, but with a different value of the parameter on a particular subregion of the spatial domain. The main result is an extension to multidimensional time of the method of Pollak and Yakir (1998), which relies on a change of measure motivated by change-point analysis, to evaluate approximately the null distribution of the likelihood ratio statistic. Both large deviation and Poisson approximations are obtained.

Running Title: Tail probabilities for scanning statistics

Key Words and Phrases: scan statistic, change-point, likelihood ratio

1 Introduction.

Maxima of random fields arise in various scientific contexts. Our interest is motivated especially by statistical problems of searching a region for a deterministic signal against a noisy background. Examples are found in Levin and Kline (1985), who are concerned with transient increases in the rate of spontaneous abortions in epidemiological data, in Giller (1994), who discusses a search of the celestial sphere for an anomalously large astronomical point source of muons, in Karlin, Dembo and Kawabata (1990), who are concerned with searching the sequence of amino acids in a protein to find segments of anomalously large electrical charge or degree of hydrophobicity, in Rabinowitz (1994), who is interested in "hot spots" of disease incidence in a geographically defined region.

Although many methods have been developed to deal with one dimensional indexing sets, e.g., Pickands (1969), Siegmund (1985), Woodroofe (1976, 1982), the number of methods that has proved useful in higher dimensions is comparatively small. The first of these chronologically is Qualls and Watanabe's (1973) and Bickel and Rosenblatt's (1973) multidimensional extension of Pickands' (1969) method. These authors studied continuous parameter Gaussian processes, and their approximation involves a difficult to evaluate constant. Hogan and Siegmund (1986) adapted the method to discrete parameter processes and showed that one can find easily computable expressions for the constant for a large number of fields that behave locally as sums of independent one dimensional random walks. Siegmund (1988, 1992) extended to higher dimensions the method of Woodroofe (1976, 1982) and enlarged the number of examples for which explicit results have been obtained. Aldous (1989) obtained similar results (in their continuous index set versions) from his Poisson clumping heuristic.

The purpose of this paper is to extend the method recently introduced by Pollak and Yakir (1998) from one dimensional to multidimensional indexing sets. This method starts from a likelihood ratio identity motivated by ideas related to change-point problems. Since it is based on an exact representation of the required probability, one can more easily "see" the answer than with the methods mentioned above, which produce the answer only as the result of a substantial amount of computation. The representation is valid for either discrete or continuous indexing sets.

To motivate our results, we consider the following class of statistical problems. A finite subset $\mathcal{I}$ of the standard d-dimensional lattice (usually $d = 1, 2$, or 3) indexes independent random variables $X_{\mathbf{u}}, \mathbf{u} \in \mathcal{I}$. Over most of the region $\mathcal{I}$ the $X_{\mathbf{u}}$ have a "null" distribution, say standard normal or Bernoulli with known $p = p_0$, perhaps 1/2. Over a relatively small subset A of $\mathcal{I}$, which may be empty, the distribution of the $X_{\mathbf{u}}$ belongs to the same parametric family but has a different value of the parameter, say normal with mean $\mu > 0$ and variance 1, or Bernoulli with $p > p_0$. Our goal is to test whether indeed A is empty, in which case $X_{\mathbf{u}}$ has the null distribution for all $\mathbf{u} \in \mathcal{I}$.

Assume for a moment that A, if it is non-empty, is known. The likelihood ratio test statistic for a general multidimensional exponential family of distributions can be conveniently expressed as follows. We denote the log likelihood for a single observation from the exponential family by $\exp[\langle\theta, x\rangle - \psi(\theta)]dF(x)$. Without loss of generality we assume that the null value of θ is $\theta = 0$, and that $\psi(0) = 0$, $\dot{\psi}(0) = 0$. Let n_A denote the cardinality of A, $S_A = \Sigma_{\mathbf{u}\in A} X_{\mathbf{u}}$, and $\bar{X}_A = S_A/n_A$. Let $\varphi(x) = \sup_\theta[\langle\theta, x\rangle - \psi(\theta)]$. The likelihood ratio statistic is $n_A\varphi(\bar{X}_A)$.

Usually we will not know the location, size, or shape of A; and we propose to use as our test statistic the maximum of $n_A\varphi(\bar{X}_A)$ over a suitable collection of candidate sets. To this end let $\mathcal{J}$ be a collection of subsets j of $\mathcal{I}$. For each $j \in \mathcal{J}$ let $S_j = \Sigma_{\mathbf{u} \in j} X_{\mathbf{u}}$, and denote by n_j the cardinality of j. Also let $\bar{X}_j = S_j/n_j$. Consider the "scan" statistic

$$\max_{j \in \mathcal{J}} n_j \varphi(\bar{X}_j). \tag{1}$$

The probability under the null distribution that (1) exceeds a threshold a is the p-value of this statistic.

The following special case is typical and will be considered in detail below. Suppose $\mathcal{I}$ is the $m \times m$ square in the positive quadrant of the plane with one vertex at the origin. Suppose also that A, if it is non-empty, is a rectangle with its sides parallel to the coordinate axes. Then (1) is the likelihood ratio statistic when $\mathcal{J}$ denotes all sub-rectangles of $\mathcal{I}$, indexed in some convenient way, say by their lower left hand corner, length and width. A simpler statistic arises if one regards the dimensions of the rectangle A as known, so $\mathcal{J}$ consists simply of translations of a rectangle of fixed length and width. For the particular case of normal $X_{\mathbf{u}}$, $\varphi(x) = ||x||^2/2$ and (1) becomes $\max_{j \in \mathcal{J}} ||S_j||^2/2n_j$. The approximate p-value of this statistic when $\mathcal{I}$ is one dimensional has been given by Siegmund and Venkatraman (1995).

This paper is organized as follows. In Section 2 we introduce a fundamental likelihood ratio identity and indicate heuristically how it allows us to obtain a tail approximation to the desired probability under large deviation scaling. The approximation involves a constant that in general is quite complicated, but simplifies in special cases. In Section 3 we discuss some examples and an alternative formulation involving a Poisson approximation, which requires a substantially more intricate proof. Technical lemmas are given in Section 4, and a heuristic discussion of the case of multidimensional θ is given in Section 5. For completeness we give a slight generalization of an argument of Hogan and Siegmund (1986) in one appendix and a useful algebraic identity in another.

2 Likelihood ratio identity and a basic approximation.

We continue with the notation and assumptions of the preceding section. In particular we assume to simplify the exposition that $\mathcal{I}$ is the $m \times m$ square described there, although that plays no essential role in what follows. We also assume until further notice that the exponential family is one dimensional and that the alternatives to the null value of $\theta = 0$ are positive. Let $a > 0$. Assume that $\mathcal{J}$ consists of rectangles indexed by $\mathbf{u} \in \mathcal{I}$ and having sides of length r_i $i = 1, 2$. There are necessarily some technical assumptions relating the values of m, a, and the r_i, about which we will have more to say later. Convenient assumptions for this section are that the r_i are uniformly bounded below and above by multiples of a and $m < a^c$ for some $c > 1$. These assumptions will be weakened in Section 4. Since the function taking θ into $\eta(\theta) = \theta\dot{\psi}(\theta) - \psi(\theta)$ is non-negative and convex, for each j the equation

$$\eta(\theta_j) = a/n_j \tag{2}$$

has at most one positive solution, which we assume exists, at least for all sufficiently large n_j. Putting $\sigma_0^2 = \ddot{\psi}(0)$, we see that $\theta_j \sim (2a/\sigma_0^2 n_j)^{1/2} = O(a^{-1/2})$ uniformly in j as $a \to \infty$.

Define the probability P_j to be such that $X_{\mathbf{u}}, \mathbf{u} \in j$ have parameter value θ_j while otherwise $X_{\mathbf{u}}$ has the null parameter value 0. Then the log likelihood of P_j relative to the null probability P is

$$\ell_j = \theta_j S_j - n_j \psi(\theta_j),$$

which under P_j has expectation equal to $n_j\eta(\theta_j) = a$. It is readily shown that

$$\left\{\max_{j\in\mathcal{J}} n_j\varphi(\bar{X}_j^+) \geq a\right\} = \left\{\max_{j\in\mathcal{J}} \ell_j \geq a\right\}. \tag{3}$$

Let $\mathrm{Q} = \Sigma_{j\in\mathcal{J}}\mathrm{P}_j$. The likelihood ratio of Q relative to P is $\Sigma_{j\in\mathcal{J}}\exp(\ell_j)$ and hence

$$\mathrm{P}\left\{\max_{j\in\mathcal{J}} n_j\varphi(\bar{X}_j^+) \geq a\right\} = \Sigma_j\mathrm{E}_j\left[1/\Sigma_k\exp(\ell_k); \max_{j\in\mathcal{J}} \ell_j \geq a\right]. \tag{4}$$

It follows by elementary algebra that the term on the right hand side of (4) indexed by j can be rewritten as

$$\exp(-a)\mathrm{E}_j\left[\frac{\exp(\max \ell_k)}{\Sigma_k\exp(\ell_k)}\exp[-(\ell_j - a + \max_k(\ell_k - \ell_j)]; \ell_j - a + \max_k(\ell_k - \ell_j) \geq 0\right] \tag{5}$$

where the summation and the max's extend over $k \in \mathcal{J}$.

The analysis of (5) proceeds via several approximations, valid asymptotically as $a \to \infty$. For the technical steps to justify these approximations see Section 4. The first approximation is the replacement of the summation and max's over $k \in \mathcal{J}$ by a smaller set of indices $\mathcal{J}(j,t)$ that are close to j in the sense that the distance of each edge of k from the corresponding edge of j is no more than $t = c\log a$ or some other function that grows slowly with a (cf. Lemma 2 in Section 4). Then within this range of k, we can replace ℓ_k by $\tilde{\ell}_k = \theta_j S_k - n_k\psi(\theta_j)$ (Lemma 4). (Obviously $\tilde{\ell}_k$ also depends on j, although this is suppressed in the notation.) Then ℓ_j is replaced by $\tilde{\ell}_h$, where h is the intersection of all $k \in \mathcal{J}(j,t)$ (cf. Lemmas 6, 7 and display (21)). The fraction in the expectation in (5) can be rewritten as

$$\frac{\exp(\max \tilde{\ell}_k - \ell_j)}{\Sigma_k\exp(\tilde{\ell}_k - \ell_j)}, \tag{6}$$

which is easily seen to be independent of $\tilde{\ell}_h$. Finally, the approximation for $\max_k(\ell_k - \ell_j)$, namely $\max_k(\tilde{\ell}_k - \tilde{\ell}_h)$, is also independent of $\tilde{\ell}_h$ and hence can be shown to be negligible. It follows that the expectation in (5) is approximately the product

$$\mathrm{E}_j\left[\frac{\exp(\max \tilde{\ell}_k - \ell_j)}{\Sigma_k\exp(\tilde{\ell}_k - \ell_j)}\right] \times \mathrm{E}_j\left[\exp\{-(\tilde{\ell}_h - a)\}; \tilde{\ell}_h - a \geq 0\right]. \tag{7}$$

Recalling that $a = n_j\eta(\theta_j) = \mathrm{E}_j(\ell_j) = n_j n_h^{-1}\mathrm{E}_j(\tilde{\ell}_h)$, so $\mathrm{E}_j(\tilde{\ell}_h) = a + O(t)$, we see from a local central limit theorem if, for example, the $X_{\mathbf{u}}$ have a density function, that the second expectation in (7)

$$\sim 1/\theta_j[2\pi n_j\ddot{\psi}(\theta_j)]^{1/2}$$

(cf. Lemma 9), so from (5) we find that (4)

$$\sim [4\pi a]^{-1/2}\exp(-a)\Sigma_j\mathrm{E}_j\left[\frac{\exp(\max \ell_k)}{\Sigma_k\exp(\ell_k)}\right]. \tag{8}$$

We now turn to evaluation of the final expectation in (8), or equivalently evaluation of

$$\Sigma_j\mathrm{E}_j\left[\frac{\exp(\max \ell_k - \ell_j)}{\Sigma_k\exp(\ell_k - \ell_j)}\right]. \tag{9}$$

¿From the preceding argument we see that the summation and max can be restricted to the relatively small set $\mathcal{J}(j,t)$ described above, while ℓ_k can be replaced by $\tilde{\ell}_k$. Let $m_0 = 2t$, so there are asymptotically m_0^4 rectangles in $\mathcal{J}(j,t)$. The term in (9) subscripted by j is approximately equal to

$$\mathrm{E}_i\left[\frac{\exp(\max_k \ell_k - \ell_i)}{\Sigma_k \exp(\ell_k - \ell_i)}\right]$$

for all $i \in \mathcal{J}(j,t)$, except for a relatively small number of indices i near the boundaries of $\mathcal{J}(j,t)$. Hence the preceding display

$$\sim m_0^{-4}\Sigma_i \mathrm{E}_i\left[\frac{\exp(\max_k \ell_k - \ell_j)}{\Sigma_k \exp(\ell_k - \ell_j)}\right], \tag{10}$$

where the indices i and k run over $\mathcal{J}(j,t)$. Let $\mathrm{Q} = \Sigma_i \mathrm{P}_i$, so $dQ/dP_j = \Sigma_k \exp(\ell_k - \ell_j)$. Then (10)

$$= m_0^{-4}\mathrm{E}_j \exp\{\max_k(\ell_k - \ell_j)\},$$

which up to a factor of $1 \pm \epsilon$, where ϵ can be arbitrarily small, is bounded above and below by

$$m_0^{-4}\mathrm{E}_h \exp\{\max_k(\tilde{\ell}_k - \tilde{\ell}_h)\}$$

$$= m_0^{-4}\mathrm{E}_h \exp\{\max_k[\theta_j(S_k - S_h) - (n_k - n_h)\psi(\theta_j)]\}. \tag{11}$$

The random field in the last expression consists of a sum of four independent one dimensional random fields that arise from the enlargement of h in each direction. A slight generalization of an argument of Hogan and Siegmund (1986) shows that (11)

$$\sim \Pi_{\delta=1}^2 [r_\delta \eta(\theta_j)\nu(r_\delta^{1/2}\sigma_0\theta_j)]^2, \tag{12}$$

where r_1 and r_2 are the lengths of the sides of the rectangle j, $\sigma_0^2 = \ddot{\psi}(0) = \mathrm{Var}(X_u)$, and the function $\nu(\cdot)$ is defined by Siegmund (1985, p. 82), where there is also a simple approximation for small x.

We now substitute (12) into (9); and using the fact that $\eta(\theta_j) = a/n_j$, we approximate the multiple sum by a multiple integral. This shows that (9)

$$\sim m^2 a^2 \left(\int_{(2a/m)^{1/2}}^{\infty} (x - 2a/mx)\nu^2(x)dx\right)^2.$$

Substitution of this result into (8) yields our final approximation:

$$\mathrm{P}\left\{\max_{j\in\mathcal{J}} n_j \varphi(\bar{X}_j^+) \geq a\right\} \sim$$
$$m^2 a^{3/2}\exp(-a)[4\pi]^{-1/2}\left(\int_{(2a/m)^{1/2}}^{\infty} (x - 2a/mx)\nu^2(x)dx\right)^2. \tag{13}$$

See Siegmund and Venkatraman (1995) for a version of this result for a one dimensional search involving normally distributed observations.

3 Examples and discussion.

Although we have given the preceding argument for the case that $\mathcal{J}$ consists of rectangles of variable width, with minor variations the approximation (8) is valid for very general $\mathcal{J}$. However, an appropriate strategy for evaluating (9) will depend on more specific assumptions. If $\mathcal{J}$ consists of translations of a fixed set, e.g., a circle or a rectangle, then except for possible edge effects the terms in (9) are all equal, so it is necessary to calculate only one of them, which might be accomplished by simulation if other methods fail. If the search sets consist of rectangles of fixed dimensions, say r_1, r_2, so $n_j = n = r_1 r_2$ and likewise $\theta_j = \theta$ is constant in j, then the increments $\ell_k - \ell_j$ assume the particularly simple form $\theta(S_k - S_j)$. Instead of the right hand side of (11) we obtain the much simpler expression

$$m_0^{-2}\mathrm{E}_j \exp\{\max_k \theta(S_k - S_j)\}.$$

The random field $S_k - S_j$ is approximately the sum of two independent two sided random walks corresponding to shifting j to the right or left and shifting it up or down. The increment of the random walk corresponding to a unit shift to the right or left is distributed as the the sum of two independent random variables, the first having the distribution of a sum of r_2 independent variables with parameter 0 and the second having the distribution of the *negative* of a sum of r_2 independent variables with parameter θ. In place of (12) we get

$$\Pi_{\delta=1}^2 [2r_\delta \eta(\theta)\nu\{(2r_\delta)^{1/2}\sigma_0\theta].$$

Since this expression does not depend on j, the final approximation becomes

$$2\pi^{-1/2} m^2 a^{3/2} (r_1 r_2)^{-1} \exp(-a) \Pi_{\delta=1}^2 \nu\{2(a/r_\delta)^{1/2}\}. \tag{14}$$

For circular search regions no such simple evaluation seems possible, although for Gaussian fields one can use Slepian's inequality and inscribed and circumscribed squares to obtain upper and lower approximations.

The specific form of the approximation in Section 2 is a consequence of the asymptotic normalization introduced above. It has the advantage of being relatively simple to evaluate, since there are easily computed, good approximations for the function ν (Siegmund 1985). However, there are alternative asymptotic formulations leading to approximations that depend more heavily on the underlying distribution. Indeed, even the formulation of Section 2 leads to more complicated approximations when the indexing set is one-dimensional, since then the increments $S_k - S_j$ need not contain a large number of terms, hence need not be approximately normally distributed.

To consider one other possibility, suppose that the $X_{\mathbf{u}}$ are infinitely divisible and that in principle one might observe the process S_j over a *continuous* set of rectangles j. A specific case of interest is a Poisson random field. (Gaussian fields are irrelevant to these considerations; one obtains the same approximation regardless of the normalization.) Since in practice we make observations at a discrete set of points, assume that the possible distributions of $X_{\mathbf{u}}$ have cumulant generating function $\Delta\psi(\theta)$, where Δ is a small parameter that reflects the size of the pixel $\mathbf{u}$ in the indexing field. For example, the pixels may be squares of area Δ. The functional equation defining θ_j becomes $n_j \Delta \eta(\theta_j) = a$. If Δ is assumed proportional to a^{-1}, then for rectangles having sides proportional to a, θ_j is bounded away from 0; and the increments $S_k - S_j$ for k close to j will not be asymptotically normal but will involve the parent class of infinitely divisible distributions. A consequence is that the function corresponding to ν above will depend on the underlying distribution and may be substantially more difficult to evaluate. For rectangles of fixed dimensions, only a few evaluations are necessary, but for rectangular scanning sets of variable size the additional numerical computation can be onerous. If Δ is of smaller order than a^{-1}, the approximation will be the same as for a continuous scan. See Loader (1991) and Tu (1997) for analyses of Poisson random fields using this normalization in conjuncton with the method of Siegmund (1988).

A still different formulation is appropriate when the baseline value of θ, taken here to be zero, is unknown. The likelihood ratio statistic would be

$$\max_j \{n_j\varphi(S_j/n_j) + (|\mathcal{I}| - n_j)\varphi[(S_\mathcal{I} - S_j)/(|\mathcal{I}| - n_j)] - |\mathcal{I}|\varphi(S_\mathcal{I}/|\mathcal{I}|)\},$$

and to obtain a p-value that is free of unknown nuisance parameters we would evaluate the exceedance probability conditionally given the value of $S_\mathcal{I}$. The probability P_j can be defined as follows. Let P denote conditional probability given that $S_\mathcal{I} = |\mathcal{I}|\xi_0$. Define parameters ξ_1 as solutions of the equation

$$\max_j \{n_j\varphi(\xi_1) + (|\mathcal{I}| - n_j)\varphi[(|\mathcal{I}|\xi_0 - n_j\xi_1)/(|\mathcal{I}| - n_j)] - |\mathcal{I}|\varphi(\xi_0)\} = a.$$

For each of the two values of ξ_1, say $\xi_1 > \xi_0$, the probability P_j is defined as the conditional probability given $S_\mathcal{I} = |\mathcal{I}|\xi_0$ that is in the exponential family generated by P and gives S_j the mean value $n_j\xi_1$. For the special case of unit variance normally distributed X_u, so $\varphi(x) = x^2/2$, simple algebra shows that $\xi_1 = \xi_0 + b[(1 - n_j/|\mathcal{I}|)/n_j]^{1/2}$, where $a = b^2/2$, and

$$\ell_j = b[S_j - n_j\xi_0]/[n_j(1 - n_j/|\mathcal{I}|)]^{1/2} - b^2/2.$$

There are technical assumptions in our discussion relating the size of the rectangles $j \in \mathcal{J}$, the threshold a and the size of the search region defined by m. Although these assumptions may not be restrictive in applications, which typically involve a fixed value of m and search regions that we choose, there are nevertheless mathematical questions about the importance of the assumtions. The analysis indicated above applies to the case of large deviations, i.e., m is small enough that (14) or the right hand side of (13) converges to 0. In the case that m is proportional to a the requirement that the sides of the rectangles be bounded by ca poses no restriction; but if m is of larger order of magnitude, it is natural to ask if we can remove the assumed upper bounds on the size of the rectangles, so the scanning sets can take up a positive fraction of the search region.

A similar issue arises if m is so large that (14) or the right hand side of (13) converges to a positive limit, say λ. Then one asks if a Poisson approximation holds, i.e., the corresponding probability converges to $1 - \exp(-\lambda)$. This is easily shown to be true in the case of (14), where the scanning sets are of fixed dimensions and there obviously are no "long range" dependencies. The case of (13) is substantially more delicate if one also asks, as seems natural in this case, whether the condition that the lengths of the sides of the rectangles have upper bounds of ca can also be dropped. Then long range dependencies might conceivably be important, and the Poisson parameter is *not* proportional to the product of the number of rectangles and the probability that an arbitrary rectangle exceeds the threshold. In the special case of one search dimension and Gaussian $X_\mathbf{u}$ a Poisson approximation was given by Siegmund and Venkatraman (1995). In Section 4 we prove such a Poisson approximation for the random field of Section 2.

It is also possible to remove the assumed lower bound on the lengths of the sides of the scanning rectangles, as we show in Section 4.

4 A more precise treatment.

In this section we make more precise the argument leading to (7) and (8) in Section 2. It will be apparent that the argument is quite general up to the application of a local limit theorem, where one must deal with the specific distribution of X_u.

Given two points $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$, we say that $\mathbf{x} \le \mathbf{y}$ if $x_1 \le y_1$ and $x_2 \le y_2$. With each pair of points in the grid, $\mathbf{x}$ and $\mathbf{y}$, such that $\mathbf{x} \le \mathbf{y}$, a rectangle of grid points can be associated. The points in the rectangle are all points $\mathbf{u}$ such that $\mathbf{x} < \mathbf{u} \le \mathbf{y}$. Given a collection of rectangles $\mathcal{J}$, denote a particular member by j. Thus, $j = (\mathbf{x}, \mathbf{y}]$, for some $\mathbf{x}, \mathbf{y} \in \mathcal{I}$. Let, also, $r_1 = y_1 - x_1$ and $r_2 = y_2 - x_2$ and define $n_j = |(\mathbf{x}, \mathbf{y}]| = r_1 \times r_2$ — the cardinality of the rectangle $(\mathbf{x}, \mathbf{y}]$. We investigate the case where the collection $\mathcal{J}$ contains all rectangles with $\epsilon a \le r_1 \le ca$ and $\epsilon a \le r_2 \le ca$, for some $0 < \epsilon < c < \infty$. These bounds on the lengths of the sides of the rectangles will be removed in the arguments following the statement of Theorem 1. Initially we also assume that $m = O(a^c)$ for some $c > 1$.

Throughout this section we will introduce various constants. The exact values of these constants do not effect the final result. All that is needed is that they are positive but small (in which case they will be denoted by ϵ) or that they are large (in which case they will be denoted by c). Hence, for example, two c's appearing in the same proof may correspond, as a matter of fact, to two different numbers.

Lemma 1

$$\mathrm{P}\left(\max_{i\in\mathcal{J}} \ell_i \ge a\right) = \sum_{j\in\mathcal{J}} \mathrm{E}_j\left[\frac{1}{\sum_{i\in\mathcal{J}} \exp\{\ell_i\}}; \max_{i\in\mathcal{J}} \ell_i \ge a\right]$$

Proof: This is just a formal restatement of (4), which was proved in Section 2. ∎

Confine attention now to a given rectangle j. We will prove that

$$a^{1/2}e^a\mathrm{E}_j\left[\frac{1}{\sum_{i\in\mathcal{J}} \exp\{\ell_i\}}; \max_{i\in\mathcal{J}} \ell_i \ge a\right] =$$
$$\mathrm{E}_j\left[\frac{\max_{i\in\mathcal{J}} \exp\{\ell_i\}}{\sum_{i\in\mathcal{J}} \exp\{\ell_i\}} a^{1/2}\exp\left\{-\left(\max_{i\in\mathcal{J}} \ell_i - a\right)\right\}; \max_{i\in\mathcal{J}} \ell_i \ge a\right] \tag{15}$$

can be approximated, when a is large, by a constant. The constant may depend on j, but by virtue of the assumption that $r_i \le ca$ for $i = 1, 2$ it is bounded away from 0. (This is apparent from the explicit evaluation given in the preceding section.) In addition the approximation is uniformly accurate for all rectangles $j \in \mathcal{J}$.

The proof will proceed in two steps. In the first step it will be shown that the term in (15) can be replaced with a similar term, for which the maximization and summation is with respect to a smaller set of rectangles — the rectangles in the vicinity of j. In the second step this term will be approximated by a constant.

Define, for $t = c \log a$, a neighborhood of $j = (\mathbf{x}, \mathbf{y}]$ by

$$\mathcal{J}(j,t) = \{(\mathbf{u}, \mathbf{v}] : |u_i - x_i| < t, |v_i - y_i| < t, i = 1, 2\}.$$

Lemma 2 *Let $\epsilon > 0$ be given. Then, uniformly in $j \in \mathcal{J}$,*

$$a^{1/2} e^a \mathrm{E}_j \left[\frac{1}{\sum_{i \in \mathcal{J}} \exp\{\ell_i\}}; \max_{i \in \mathcal{J}} \ell_i \geq a \right] \leq$$
$$a^{1/2} e^a \mathrm{E}_j \left[\frac{1}{\sum_{i \in \mathcal{J}(j,t)} \exp\{\ell_i\}}; \max_{i \in \mathcal{J}(j,t)} \ell_i \geq a \right] + \epsilon \tag{16}$$

and

$$a^{1/2} e^a \mathrm{E}_j \left[\frac{1}{\sum_{i \in \mathcal{J}} \exp\{\ell_i\}}; \max_{i \in \mathcal{J}} \ell_i \geq a \right] \geq$$
$$\frac{1}{1+\epsilon} a^{1/2} e^a \mathrm{E}_j \left[\frac{1}{\sum_{i \in \mathcal{J}(j,t)} \exp\{\ell_i\}}; \max_{i \in \mathcal{J}(j,t)} \ell_i \geq a \right] - \epsilon, \tag{17}$$

provided that a is large enough.

Proof: On the one hand, since the random variable in (15) is bounded by $a^{1/2}$,

$$\begin{aligned}
& a^{1/2} e^a \mathrm{E}_j \left[\frac{1}{\sum_{i \in \mathcal{J}} \exp\{\ell_i\}}; \max_{i \in \mathcal{J}} \ell_i \geq a \right] \\
\leq \;& a^{1/2} e^a \mathrm{E}_j \left[\frac{1}{\sum_{i \in \mathcal{J}} \exp\{\ell_i\}}; \max_{i \in \mathcal{J}(j,t)} \ell_i \geq a \right] \\
& + a^{1/2} \mathrm{P}_j \left(\max_{i \in \mathcal{J}} \ell_i > \max_{i \in \mathcal{J}(j,t)} \ell_i \right) \\
\leq \;& a^{1/2} e^a \mathrm{E}_j \left[\frac{1}{\sum_{i \in \mathcal{J}(j,t)} \exp\{\ell_i\}}; \max_{i \in \mathcal{J}(j,t)} \ell_i \geq a \right] \\
& + a^{1/2} \mathrm{P}_j \left(\max_{i \notin \mathcal{J}(j,t)} \ell_j > 0 \right).
\end{aligned}$$

On the other hand

$$\begin{aligned}
& a^{1/2} e^a \mathrm{E}_j \left[\frac{1}{\sum_{i \in \mathcal{J}} \exp\{\ell_i\}}; \max_{i \in \mathcal{J}} \ell_i \geq a \right] \\
\geq \;& a^{1/2} e^a \mathrm{E}_j \left[\frac{1}{\sum_{i \in \mathcal{J}} \exp\{\ell_i\}}; \max_{i \in \mathcal{J}(j,t)} \ell_i \geq a \right] \\
\geq \;& \frac{1}{1+\epsilon} a^{1/2} e^a \mathrm{E}_j \left[\frac{1}{\sum_{i \in \mathcal{J}(j,t)} \exp\{\ell_i\}}; \max_{i \in \mathcal{J}(j,t)} \ell_i \geq a \right] \\
& - a^{1/2} \mathrm{P}_j \left(\sum_{i \notin \mathcal{J}(j,t)} \exp\{\ell_i - \ell_j\} \geq \epsilon \right).
\end{aligned}$$

The proof now follows from Lemma 3 below and the assumption that m increases at most algebraically with a so $|\mathcal{J}| = O(a^c)$ for some $c > 0$. ∎

Lemma 3 *Let*

$$\kappa = \kappa(j,i) = \kappa((\mathbf{x},\mathbf{y}],(\mathbf{u},\mathbf{v}]) = |((\mathbf{x},\mathbf{y}] \setminus (\mathbf{u},\mathbf{v}]) \cup ((\mathbf{u},\mathbf{v}] \setminus (\mathbf{x},\mathbf{y}])|$$

be the number of points in the symmetric difference between the rectangles $j = (\mathbf{x},\mathbf{y}]$ *and* $i = (\mathbf{u},\mathbf{v}]$. *Then*

$$\mathrm{P}_j\left(\ell_i - \ell_j \geq -\epsilon\kappa/a\right) \leq \exp\left\{-\epsilon\kappa/a\right\},$$

for some positive ϵ *and for all* $i \in \mathcal{J}$.

Proof: By an exponential Markov inequality

$$\mathrm{P}_j\left\{\ell_i - \ell_j \geq -\epsilon\kappa/a\right\} \leq \exp(\epsilon\kappa/2a)\mathrm{E}\exp[(\ell_i + \ell_j)/2].$$

We now write the sums involved in $\ell_i + \ell_j$ as sums over the disjoint sets $i \setminus j$, $j \setminus i$ and $i \cap j$ to evaluate this expectation in terms of the function ψ. The convexity of ψ implies that we get an upper bound if we replace $\psi[(\theta_i + \theta_j)/2]$ by $[\psi(\theta_i) + \psi(\theta_j)]/2$. The asymptotic relations $\psi(\theta_i) \sim \sigma_0^2\theta_i^2/2$ and $\theta_i \sim (2a/\sigma_0^2 n_i)^{1/2}$ now allow us to complete the proof. ∎

Define, for all $i \in \mathcal{J}(j,t)$,

$$\tilde{\ell}_i = \tilde{\ell}_i(j) = \theta_j S_i - n_i\psi(\theta_j).$$

In the next lemma we claim that ℓ_i can be replaced by $\tilde{\ell}_i$.

Lemma 4 *Let* $\epsilon > 0$ *be given. Then, uniformly in* $j \in \mathcal{J}$,

$$a^{1/2}e^a\mathrm{E}_j\left[\frac{1}{\sum_{i\in\mathcal{J}(j,t)}\exp\{\ell_i\}};\ \max_{i\in\mathcal{J}(j,t)}\ell_i \geq a\right] \leq$$
$$\frac{1+\epsilon}{1-\epsilon}a^{1/2}e^a\mathrm{E}_j\left[\frac{1}{\sum_{i\in\mathcal{J}(j,t)}\exp\{\tilde{\ell}_i\}};\ \max_{i\in\mathcal{J}(j,t)}\tilde{\ell}_i \geq a-\epsilon\right] + \epsilon \tag{18}$$

and

$$a^{1/2}e^a\mathrm{E}_j\left[\frac{1}{\sum_{i\in\mathcal{J}(j,t)}\exp\{\ell_i\}};\ \max_{i\in\mathcal{J}(j,t)}\ell_i \geq a\right] \geq$$
$$\frac{1-\epsilon}{1+\epsilon}a^{1/2}e^a\mathrm{E}_j\left[\frac{1}{\sum_{i\in J(j,t)}\exp\{\tilde{\ell}_i\}};\ \max_{i\in J(j,t)}\tilde{\ell}_i \geq a+\epsilon\right] - \epsilon, \tag{19}$$

provided that a is large enough.

Proof: It is sufficient to prove that

$$\mathrm{P}_j(\max\nolimits_{i\in\mathcal{J}(j,t)}|\tilde{\ell}_i - \ell_i| > \epsilon) \leq \epsilon/a^{1/2},$$

which is established in the following lemma. ∎

Lemma 5 $\mathrm{P}_j(\max_{i\in\mathcal{J}(j,t)}|\tilde{\ell}_i - \ell_i| > \epsilon) \leq \epsilon/a^{1/2}$.

Proof: Note that

$$P_j(\max_{i\in\mathcal{J}(j,t)} |\tilde{\ell}_i - \ell_i| > \epsilon) \le \sum_{i\in\mathcal{J}(j,t)} P_j(|\tilde{\ell}_i - \ell_i| > \epsilon)$$

and

$$\ell_i - \tilde{\ell}_i = n_i[(\theta_i - \theta_j)\bar{X}_i - (\psi(\theta_i) - \psi(\theta_j)].$$

¿From $\dot{\eta}(\theta) = \theta\ddot{\psi}(\theta)$ and the assumed lower bound $n_i \ge \epsilon^2 a^2$, it follows that

$$\theta_i - \theta_j \sim \frac{a}{\theta_j\ddot{\psi}(\theta_j)}\left(\frac{1}{n_i} - \frac{1}{n_j}\right) = O(t/a^{3/2}).$$

Chebyshev's inequality and the approximation

$$n_i[(\theta_i - \theta_j)\dot{\psi}(\theta_j) - (\psi(\theta_i) - \psi(\theta_j))] \sim -n_i\ddot{\psi}(\theta_j)(\theta_i - \theta_j)^2/2 = O(t^2/a)$$

can be used to establish the proof. ▮

One can represent the leading term in Lemma 2 as (cf. (15)

$$a^{1/2}e^a\mathrm{E}_j\left[\frac{1}{\sum_{i\in\mathcal{J}(j,t)}\exp\{\tilde{\ell}_i\}};\max_{i\in\mathcal{J}(j,t)}\tilde{\ell}_i \ge a\right] = \tag{20}$$

$$\mathrm{E}_j\left[\frac{\max_{i\in\mathcal{J}(j,t)}\exp\{\tilde{\ell}_i\}}{\sum_{i\in\mathcal{J}(j,t)}\exp\{\tilde{\ell}_i\}}a^{1/2}\exp\{-(\max_{i\in\mathcal{J}(j,t)}\tilde{\ell}_i - a)\};\max_{i\in\mathcal{J}(j,t)}\tilde{\ell}_i \ge a\right]$$

Preparing for the second step of showing that this term can be approximated by a constant, the next two lemmas demonstrate that the event $\{\max_{i\in\mathcal{J}(j,t)}\tilde{\ell}_i \ge a\}$ can be intersected with two events. The first of the two is the event $\{\max_{i\in\mathcal{J}(j,t)}\tilde{\ell}_i \le a + \log a\}$; the second is given following Lemma 6.

Lemma 6 *For any $a > 1$,*

$$a^{1/2}e^a\mathrm{E}_j\left[\frac{1}{\sum_{i\in\mathcal{J}(j,t)}\exp\{\tilde{\ell}_i\}};\max_{i\in\mathcal{J}(j,t)}\tilde{\ell}_i \ge a + \log a\right] \le 1/a^{1/2}$$

Proof: This inequality is true for the random variables, hence *a fortiori* for the expectation. ▮

Now let $\mathbf{x}_t = (x_1 + t, x_2 + t)$, $\mathbf{y}_t = (y_1 - t, y_2 - t)$ and define $h = (\mathbf{x}_t, \mathbf{y}_t]$. Note that $h = \cap_{i\in\mathcal{J}(j,t)}(\mathbf{u},\mathbf{v}]$. The second of the two aforementioned events is $\{\max_{i\in\mathcal{J}(j,t)}\tilde{\ell}_i - \tilde{\ell}_h \le \epsilon a^{1/2}\}$

Lemma 7 *Let $\epsilon > 0$ be given. Then*

$$a^{1/2}P_j\left(\max_{i\in\mathcal{J}(j,t)}\tilde{\ell}_i - \tilde{\ell}_h \ge \epsilon a^{1/2}\right) \le \epsilon.$$

Proof: For any $i \in \mathcal{J}(j,t)$, by the martingale property of a sequence of likelihood ratios and the Markov inequality

$$P_j\left(\tilde{\ell}_i - \tilde{\ell}_j \ge \epsilon a^{1/2}/2\right) \le \exp\{-\epsilon a^{1/2}/2\},$$

so

$$P_j\left(\max_{i\in\mathcal{J}(i,t)}\tilde{\ell}_i - \tilde{\ell}_j \ge \epsilon a^{1/2}/2\right) \le (2t)^4\exp\{-\epsilon a^{1/2}/2\}.$$

Hence it suffices to show that the P_j-probability of the event $\{\tilde{\ell}_j - \tilde{\ell}_h \ge \epsilon a^{1/2}/2\}$ is bounded, when a is large, by $\epsilon/(2a^{1/2})$. This is the content of Lemma 8 below.

Lemma 8 $\mathrm{P}_j\left(\tilde{\ell}_j - \tilde{\ell}_h \geq \epsilon a^{1/2}/2\right) < \epsilon/(2a^{1/2})$.

Proof: Note that $h \subset j$. Hence,

$$\tilde{\ell}_j - \tilde{\ell}_h = \sum_{i \in j \backslash h} [\theta_j X_i - \psi(\theta_j)].$$

Now,

$$\sum_{i \in j \backslash h} [\theta_j X_i - \psi(\theta_j)] = \theta_j \sum_{i \in j \backslash h} [X_i - \dot{\psi}(\theta_j)] + \frac{(n_j - n_h)a}{n_j},$$

since $\theta_j \dot{\psi}(\theta_j) - \psi(\theta_j) = \eta(\theta_j) = a/n_j$. The Markov inequality can be used to establish the proof. ▮

Lemmas 6 and 7 can be summarized by saying that the term

$$a^{1/2} e^a \mathrm{E}_j \left[\frac{1}{\sum_{i \in \mathcal{J}(j,t)} \exp\{\tilde{\ell}_i\}}; \max_{i \in \mathcal{J}(j,t)} \tilde{\ell}_i \geq a\right]$$

can be approximated, up to a $\mathrm{o}(1)$ term, by yet another representation:

$$\mathrm{E}_j \left[\frac{\max_{i \in \mathcal{J}(j,t)} \exp\{\tilde{\ell}_i\}}{\sum_{i \in \mathcal{J}(j,t)} \exp\{\tilde{\ell}_i\}} \times a^{1/2} \exp\left\{-\left(\tilde{\ell}_h + \max_{i \in \mathcal{J}(j,t)} \tilde{\ell}_i - \tilde{\ell}_h - a\right)\right\}; A_1 \cap A_2\right], \quad (21)$$

where

$$\begin{aligned} A_1 &= \left\{a \leq \tilde{\ell}_h + \max_{i \in \mathcal{J}(j,t)} \tilde{\ell}_i - \tilde{\ell}_h \leq a + \log a\right\} \\ A_2 &= \left\{0 \leq \max_{i \in \mathcal{J}(j,t)} \tilde{\ell}_i - \tilde{\ell}_h \leq \epsilon a^{1/2}\right\}. \end{aligned}$$

The main ingredient in the second step is achieved in the following lemma, where we compute the conditional expectation of $\exp\left\{-\left(\tilde{\ell}_h + \max_{i \in \mathcal{J}(j,t)} \tilde{\ell}_i - \tilde{\ell}_h - a\right)\right\}$ given

$$\frac{\max_{i \in \mathcal{J}(j,t)} \exp\{\tilde{\ell}_i\}}{\sum_{i \in \mathcal{J}(j,t)} \exp\{\tilde{\ell}_i\}} = \frac{\max_{i \in \mathcal{J}(j,t)} \exp\{\tilde{\ell}_i - \tilde{\ell}_h\}}{\sum_{i \in \mathcal{J}(j,t)} \exp\{\tilde{\ell}_i - \tilde{\ell}_h\}} = y$$

and

$$\max_{i \in \mathcal{J}(j,t)} \{\tilde{\ell}_i - \tilde{\ell}_h\} = z.$$

Note that by independence the conditional distribution of S_h, hence $\tilde{\ell}_h$, is the same as the unconditional distribution.

Lemma 9 *Define* $\tilde{a} = a - z$, *for* $0 \leq z \leq \epsilon a^{1/2}$. *Let* $\sigma_0^2 = \ddot{\psi}(0)$ *denote the null variance of* $X_{\mathbf{u}}$, *and assume that the* θ_j *distribution of* S_h *satisfies a local limit theorem. Then, for large* a,

$$\left|\mathrm{E}_j \left[a^{1/2} \exp\left\{-\left(\tilde{\ell}_h - \tilde{a}\right)\right\}; \tilde{a} \leq \tilde{\ell}_h \leq \tilde{a} + \log a\right] - (1/2\pi\sigma_0^2)^{1/2}\right| \leq \epsilon,$$

where $\tilde{\ell}_h = \theta_j S_h - n_h \psi(\theta_j)$. *This approximation holds uniformly in* z.

Proof: To consider one specific case, suppose the distribution of X_u is lattice with span 1. The argument is virtually the same if the distribution has an integrable characteristic function, so its density obeys a local central limit theorem. A Taylor expansion of $\eta(\theta)$ around $\theta = 0$ can be used to show that

$$\left|\theta_j/(2a/n_j)^{1/2} - \sigma_0^{-1}\right| \leq \epsilon,$$

provided that a is large. The probability $\mathrm{P}_j(S_h = s)$ is approximated by a normal density. In particular, up to a factor of $1 \pm \epsilon$ the approximation is $1/(2\pi n_h \sigma_0^2)^{1/2}$, for all s in a neighborhood of the mean of S_h of radius $\epsilon n_h^{1/2}$, so

$$\left|\frac{a^{1/2}\mathrm{P}_j(S_h = s)}{(a/n_h)^{1/2}} - \left(\frac{1}{2\pi\sigma_0^2}\right)^{1/2}\right| \leq \epsilon.$$

It follows that

$$\begin{aligned} &\mathrm{E}_j\left[a^{1/2}\exp\left\{-\left(\tilde{\ell}_h - \tilde{a}\right)\right\}; \tilde{a} \leq \tilde{\ell}_h \leq \tilde{a} + \log a\right] \\ &\leq \; (1/2\pi\sigma_0^2)^{1/2}(1+\epsilon)(a/n_h)^{1/2}\sum_{s=1}^{\infty}\exp\{-(1-\epsilon)(2a/n_h\sigma_0^2)^{1/2}s\}, \end{aligned}$$

and

$$\begin{aligned} &\mathrm{E}_j\left[a^{1/2}\exp\left\{-\left(\tilde{\ell}_h - \tilde{a}\right)\right\}; \tilde{a} \leq \tilde{\ell}_h \leq \tilde{a} + \log a\right] \\ &\geq \; (1/2\pi\sigma_0^2)^{1/2}(1-\epsilon)(a/n_h)^{1/2}\sum_{s=2}^{\infty}\exp\{-(1+\epsilon)(2a/n_h\sigma_0^2)^{1/2}s\}. \end{aligned}$$

∎

Lemmas 1 to 9 can be summarized in the following theorem.

Theorem 1 *For the scanning statistic computed over* $\mathcal{J}$ *(all rectangles* $(\mathbf{x}, \mathbf{y}]$ *which satisfy* $\epsilon a \leq y_i - x_i \leq ca$, *i=1,2) the two:*

$$a^{1/2}e^a\mathrm{P}\left(\max_{j\in\mathcal{J}} \ell_j \geq a\right) \tag{22}$$

and

$$\left(\frac{1}{4\pi}\right)^{1/2}\sum_{j\in\mathcal{J}}\mathrm{E}_j\left[\frac{\max_{i\in\mathcal{J}}\exp\{\ell_i\}}{\sum_{i\in\mathcal{J}}\exp\{\ell_i\}}\right] \tag{23}$$

are asymptotically equivalent.

We now remove the technical conditions that $\epsilon a < y_i - x_i < ca$ for $i = 1, 2$. We first consider the comparatively simple lower bound. We decompose $\mathcal{J}_0 = \{j : \min(r_1, r_2) < \epsilon a, \max(r_1, r_2) < ca\}$ as follows. Let c' diverge slowly to ∞ with a and set $\mathcal{J}_{01} = \{j : n_j > c'a,\ r_1 < \epsilon a,\ \epsilon a \leq r_2 < ca\}$, $\mathcal{J}_{02} = \{j : n_j > c'a,\ \epsilon a \leq r_1 < ca,\ r_2 < \epsilon a\}$, $\mathcal{J}_{03} = \{j : n_j > c'a, r_1 < \epsilon a, r_2 < \epsilon a\}$ and $\mathcal{J}_{04} = \{j : n_j \leq c'a\}$. With regard to $\mathcal{J}_{04}$, observe that the number of rectangles with $n_j < c'a$ and maximum dimension less than ca is no more than $m^2c'a\log(c^2a/c') = o[m^2a^{3/2}]$. Since $\mathrm{P}\{\ell_j > a\} \leq \exp(-a)$, the simple

Bonferroni bound suffices to show that these rectangles make a negligible contribution to the total. Also

$$P(\max_{j\in\mathcal{J}_{03}} \ell_j \geq a) \leq \Sigma_{j\in\mathcal{J}_{03}} P(\ell_j \geq a).$$

There are at most $\epsilon^2 m^2 a^2$ terms in this sum, each of which is (uniformly) $O[a^{-1/2}\exp(-a)]$ by a local limit theorem. Finally, writing $j = j_1 \times j_2$, where j_1 and j_2 are the projections of j on the respective coordinate axes, we have

$$P(\max_{j\in\mathcal{J}_{01}} \ell_j \geq a) = P(\max_{j\in\mathcal{J}_{02}} \ell_j \geq a) \leq \epsilon maP(\max_{\{j_1:\epsilon a\leq r_1<ca\}} \ell_{j_1\times(0,r_2]} \geq a) \sim \epsilon m^2 a^{3/2} e^{-a} K,$$

for some K — the "one-dimensional" constant, which can derived in the same way we derive the "two-dimensional" constant in the argument give above.

Remark. The preceding argument works in dimensions two and more, but if $\mathcal{I}$ is one dimensional, a more delicate argument involving the probabilities of very large deviations is required. We do not discuss that case.

To remove the condition that $y_i - x_i < ca,\ i = 1, 2$, we consider first the simpler case that $\mathcal{I}$ is one dimensional. Now let $\mathcal{J}_0 = \{j \in \mathcal{J} : n_j \leq ca\}$, and set $\mathcal{J}_1 = \{j \in \mathcal{J} : n_j > ca\}$. The arguments given above apply to the maximum taken over $\mathcal{J}_0$, so it suffices to show that the probability of the maximum over $\mathcal{J}_1$ is negligible.

We shall at the same time consider the possibility that m is so large that the right hand side of (13), or the analogous quantity when $\mathcal{I}$ is one dimensional, namely $ma^{1/2}e^{-a}$, converges to a limit λ and show that a Poisson approximation applies, i.e., the probability (4) converges to $1 - \exp(-K\lambda)$. Given the previous results in the case that m is of order a^c, the general blocking argument of Arratia, Goldstein and Gordon (1989) or a direct decomposition into almost independent blocks of size ca along the lines of Venkatraman and Siegmund (1995) shows that a Poisson approximation holds as claimed when the maximization is restricted to $\mathcal{J}_0$. Hence, as above, it suffices to show that maximizing over $\mathcal{J}_1$ produces a negligible probability.

The following lemma will be useful.

Lemma 10 *Let $X_1, X_2, \ldots, X_m$ be independent and identically distributed with* $\mathrm{E}(X_1) = 0$ *from a distribution that can be imbedded in an exponential family. Define $S_n = \sum_{u=1}^n X_u$. Then*

$$\mathrm{E}\exp\left\{\max_{1\leq n\leq m}[\theta S_n - n\psi(\theta)]\right\} \leq \left[1 + \frac{\theta}{\lambda-\theta}\right]\exp\left\{m\theta\left(\frac{\psi(\lambda)}{\lambda} - \frac{\psi(\theta)}{\theta}\right)\right\}$$

for all $0 < \theta < \lambda$.

Proof: Note that

$$\begin{aligned}
&\mathrm{E}\exp\left\{\max_{1\leq n\leq m}[\theta S_n - n\psi(\theta)]\right\} = \\
&\quad \mathrm{E}\exp\left\{\max_{1\leq n\leq m}\left[\frac{\theta}{\lambda}(\lambda S_n - n\psi(\lambda)) + n\theta\left(\frac{\psi(\lambda)}{\lambda} - \frac{\psi(\theta)}{\theta}\right)\right]\right\} \\
&\leq \exp\left\{m\theta\left(\frac{\psi(\lambda)}{\lambda} - \frac{\psi(\theta)}{\theta}\right)\right\}\mathrm{E}\exp\left\{(\theta/\lambda)\max_{1\leq n\leq m}(\lambda S_n - n\psi(\lambda))\right\}
\end{aligned}$$

Now, by Doob's inequality,

$$\bar{F}(x) = \mathrm{P}_0\left(\max_{1\leq n\leq m}[\lambda S_n - n\psi(\lambda)] \geq x\right) \leq e^{-x}.$$

Integration by parts yields

$$\begin{aligned}\mathrm{E}\exp\left\{\frac{\theta}{\lambda}\max_{1\le n\le m}(\lambda S_n - n\psi(\lambda))\right\} &= \int_0^\infty e^{\frac{\theta}{\lambda}x} d(-\bar{F}(x)) \\ = 1+\frac{\theta}{\lambda}\int_0^\infty e^{\frac{\theta}{\lambda}x}\bar{F}(x)dx &\le 1+\frac{\theta}{\lambda-\theta},\end{aligned}$$

which completes the proof.

Let $\epsilon > 0$. We define below a subset $\tilde{\mathcal{J}}_1 \subset \mathcal{J}_1$ such that for every $i \in \mathcal{J}_1$ there exists $j \in \tilde{\mathcal{J}}_1$ such that the cardinality of the symmetric difference between the two is no more than $\epsilon n_j/a$. We then show that

$$\mathrm{P}\left(\max_{j\in\tilde{\mathcal{J}}_1}\ell_j \ge a\right) \le |\tilde{\mathcal{J}}_1| a^{-1/2}e^{-a}. \tag{24}$$

The set $\tilde{\mathcal{J}}_1$ can be constructed in the following way. Let $m_s = ca(1+\epsilon/a)^s$, so $m_{s+1} = m_s(1+\epsilon/a)$, for $s = 1, 2, \cdots S = \min\{s : m_s \ge m\}$. The set $\tilde{\mathcal{J}}_1$ consists of all translations of $(1, m_s]$ by $t\epsilon m_s/a$, for all $0 \le t \le (am)/(\epsilon m_s)$ and all s. It is easy to see that

$$|\tilde{\mathcal{J}}_1| \le \sum_{s=1}^{S}\frac{am}{\epsilon ca(1+\epsilon/a)^s} < \frac{am}{\epsilon^2 c}. \tag{25}$$

The inequalities (24) and (25) together yield the desired result by choosing c so large that the right hand side of (24) is less than ϵ.

We now turn to the proof of (24). For any interval $j \in \mathcal{J}_1$, which by an abuse of notation we denote $(j, j+n_j]$, let $\tilde{n}$ be $\max\{m_s : m_s \le n_j\}$ and let $\tilde{j} = \max\{t\epsilon\tilde{n}/a : t\epsilon\tilde{n}/a \le j\}$. It follows that $(\tilde{j}, \tilde{j}+\tilde{n}]$ belongs to $\tilde{\mathcal{J}}_1$ and yet the cardinality of the symmetric difference is no more than $2\epsilon(1+\epsilon)\tilde{n}/a$. Note, also, that $\tilde{n} \le n_j$. As a consequence we see that for each $j \in \tilde{\mathcal{J}}_1$ there exists a subset $\mathcal{J}(j) \subset \mathcal{J}_1$ such that $\cup_{j\in\tilde{\mathcal{J}}_1}\mathcal{J}(j) = \mathcal{J}_1$ and the following hold: for each $i \in \mathcal{J}(j)$, $n_i \ge n_j$; the size of the symmetric difference of i and j is no more than $\epsilon n_j/a$; and the left endpoint of i is to the right of the left endpoint of j.

Obviously

$$\mathrm{P}\left(\max_{j\in\mathcal{J}_1}\ell_j \ge a\right) \le \sum_{j\in\tilde{\mathcal{J}}_1}\mathrm{P}\left(\max_{i\in\mathcal{J}(j)}\ell_i \ge a\right).$$

For a given $j \in \tilde{\mathcal{J}}_1$ let h denote the intersection of all $i \in \mathcal{J}(j)$ and consider the collection of likelihood ratios $\{\tilde{\ell}_i : i \in \mathcal{J}(j)\}$, where $\tilde{\ell}_i = \tilde{\ell}_i(h) = \theta_h S_i - n_i\psi(\theta_h)$. Note that on the event $\{\theta_i S_i - n_i\psi(\theta_i) \ge a\}$

$$\begin{aligned}\theta_h S_i - n_i\psi(\theta_h) &\ge \theta_i S_i - n_i\psi(\theta_i) + n_i[\psi(\theta_i) - \psi(\theta_h)] \\ &\ge a - \delta,\end{aligned}$$

since $\theta_h \ge \theta_i$ yet $n_i \le (1+2\epsilon/a)n_h$. Hence, $\{\max_{i\in\mathcal{J}(j)}\ell_i \ge a\} \subset \{\max_{i\in\mathcal{J}(j)}\tilde{\ell}_i \ge a-\delta\}$. Observe that $\max_{i\in\mathcal{J}(j)}\tilde{\ell}_i \le \ell_h + M_1 + M_2$, where where M_1 and M_2 are maxima of partial sums associated with the increments $\tilde{\ell}_i - \ell_h$ at the left and right endpoints of h, respectively. In particular M_1 and M_2 are independent of ℓ_h and of each other. Hence, putting $\tilde{a} = a - \delta$, we see that

$$\mathrm{P}\left(\max_{i\in\mathcal{J}(j)}\ell_i \ge a\right) \le \mathrm{P}\left(\ell_h + M_1 + M_2 \ge \tilde{a}\right),$$

which by a change of measure equals

$$e^{-\tilde{a}}\mathrm{E}_h\left[\exp(M_1+M_2)\exp[-(\ell_h+M_1+M_2-\tilde{a})];\ell_h+M_1+M_2\geq\tilde{a}\right].$$

By conditioning on M_1+M_2 and using arguments parallel to those in Lemmas 6–9, we see that uniformly on $\{M_1+M_2<a^{1/2}\}$,

$$\mathrm{E}_h\left[\exp[-(\ell_h+M_1+M_2-\tilde{a})];\ell_h+M_1+M_2\geq\tilde{a}|M_1+M_2\right]\leq ca^{-1/2}$$

for some constant c and all large enough a. Now take $\theta=\theta_h$, $\lambda=2\theta_h$ in Lemma 10, and expand $\psi(\lambda)$ about θ_h. Recalling that $\eta(\theta_h)=a/n_h$, we see that $\mathrm{E}_h[\exp(M_1+M_2)]$ is bounded. These approximations holds uniformly in $j\in\tilde{\mathcal{J}}_1$, which proves (24).

We now consider the more complicated case of a two dimensional random field. The relation of m and a appropriate for a Poisson approximation is $m^2a^{3/2}\exp(-a)\to\lambda$ for some $0<\lambda<\infty$. We begin with two lemmas.

Lemma 11 *Let $\{X_{\mathbf{u}}:\mathbf{u}\in\mathcal{I}\}$ be independent and identically distributed with mean 0 from a distribution that can be imbedded in an exponential family. Define $S_j=\sum_{\mathbf{u}\in j}X_{\mathbf{u}}$ and $n_j=|j|$ for $j\in\mathcal{J}^*$. Let $m_1=max\{|j|:j\in\mathcal{J}^*\}$ and $m_2=|\mathcal{J}^*|$. Then*

$$\mathrm{E}\exp\left\{\max_{j\in\mathcal{J}^*}[\theta S_j-n_j\psi(\theta)]\right\}\leq\left[1+\frac{m_2\theta}{\lambda-\theta}\right]\exp\left\{m_1\theta\left(\frac{\psi(\lambda)}{\lambda}-\frac{\psi(\theta)}{\theta}\right)\right\}$$

for all $0<\theta<\lambda$.

Proof: We use the same argument as in the proof of Lemma 10 with the trivial additional observation that

$$\bar{F}(x)=\mathrm{P}_0\left(\max_{j\in\mathcal{J}^*}[\lambda S_j-n_j\psi(\lambda)]\geq x\right)\leq m_2e^{-x}.$$

Lemma 12 *Let $\{X_{\mathbf{u}}:\mathbf{u}\in\mathcal{I}\}$ be as in the previous lemma. Define $S_j=\sum_{\mathbf{u}\in j}X_{\mathbf{u}}$ and $n_j=|j|$ for $j\in\mathcal{J}^*$. Let $m_1=\max\{|j|:j\in\mathcal{J}^*\}$ and $m_2=|\mathcal{J}^*|$. Then*

$$\mathrm{P}\left(\max_{j\in\mathcal{J}}[\theta S_j-n_j\psi(\theta)]\geq\delta\right)\leq m_2\exp\left\{m_1\psi(\lambda)-\lambda\delta/\theta\right\}$$

for all $0<\theta<\lambda$.

Proof: Observe that

$$\mathrm{P}\left(\max_{j\in\mathcal{J}}[\theta S_j-n_j\psi(\theta)]\geq\delta\right)\leq\sum_{j\in\mathcal{J}}\mathrm{P}(\lambda S_j\geq\lambda\delta/\theta).$$

and apply an exponential Markov inequality.

Let us divide the index set $\mathcal{J}$ into five disjoint subsets:

$$\begin{aligned}\mathcal{J}_0&=\{r_1\leq ca,r_2\leq ca\}\\\mathcal{J}_1&=\{r_1<\epsilon a,r_2>ca\}\\\mathcal{J}_2&=\{r_1>ca,r_2<\epsilon a\}\\\mathcal{J}_3&=\{\epsilon a\leq r_1\leq ca,r_2>ca\}\\\mathcal{J}_4&=\{r_1>ca,\epsilon a\leq r_2\leq ca\}\\\mathcal{J}_5&=\{r_1>ca,r_2>ca\}\end{aligned}$$

We will show that the statistic calculated by maximizing over $\mathcal{J}$ is well approximated by the statistic obtained by maximizing only over $\mathcal{J}_0$, for which it suffices to show that the probabilities associated with the maxima over $\mathcal{J}_1, \ldots, \mathcal{J}_5$ are small compared to $m^2 a^{3/2} \exp(-a)$. Consider approximations of the last four subsets along the lines of the corresponding subsets in the one dimensional case discussed above. In these definitions s and t are integers that run from 1 to an appropriate S and T defined as above for the one dimensional case. Let

$$\begin{aligned}
\tilde{\mathcal{J}}_1 &= \{x_1 \leq m, r_1 < \epsilon a, r_2 = ca(1+\epsilon/a)^s, x_2 = tr_2\epsilon/a\} \\
\tilde{\mathcal{J}}_2 &= \{r_1 = ca(1+\epsilon/a)^s, x_1 = tr_1\epsilon/a, x_2 \leq m, r_2 < \epsilon a\} \\
\tilde{\mathcal{J}}_3 &= \{r_1 = \epsilon a(1+\epsilon/a)^s, x_1 = tr_1\epsilon/a, r_2 = ca(1+\epsilon/a)^s, x_2 = tr_2\epsilon/a\} \\
\tilde{\mathcal{J}}_4 &= \{r_1 = ca(1+\epsilon/a)^s, x_1 = tr_1\epsilon/a, r_2 = \epsilon a(1+\epsilon/a)^s, x_2 = tr_2\epsilon/a\} \\
\tilde{\mathcal{J}}_5 &= \{r_1 = ca(1+\epsilon/a)^s, x_1 = tr_1\epsilon/a, r_2 = ca(1+\epsilon/a)^s, x_2 = tr_2\epsilon/a\}.
\end{aligned}$$

Note that $|\tilde{\mathcal{J}}_1| = |\tilde{\mathcal{J}}_2| < m^2a^2/(\epsilon c)$, $|\tilde{\mathcal{J}}_3| = |\tilde{\mathcal{J}}_4| < m^2a^2/(\epsilon^5 c)$ and $|\tilde{\mathcal{J}}_5| < m^2a^2/(\epsilon^2c^2)$. It follows that the claim can be proved, provided we can show that $a^{1/2}e^a \mathrm{P}_0(\max_{i\in\mathcal{J}(j)} \ell_i \geq a)$ is bounded (uniformly in $j \in \tilde{\mathcal{J}}_1 \cup \cdots \cup \tilde{\mathcal{J}}_5$).

Consider, first, the sets $\tilde{\mathcal{J}}_1$ and $\tilde{\mathcal{J}}_2$. We can take the index sets $\mathcal{J}(j)$ to be linear in these cases, so the argument given above can be applied.

Regarding the sets $\tilde{\mathcal{J}}_3$, $\tilde{\mathcal{J}}_4$ and $\tilde{\mathcal{J}}_5$, one can bound the maximum over the two-dimensional set $\mathcal{J}(j)$ by ℓ_h plus the sum of three random variables. The first random variable, M_1, is the maximum of partial sums related to observations in the rectangles with one edge equal to those edges of h that are parallel to the x-axis. The second random variable, M_2, is associated with edges of h parallel to the y-axis. The third random variable, M_3, is associated with partial sums of random variables in the rectangles with corners touching h. Note that the number of observations in each of these rectangles is at most n_h/a^2.

The random variables M_1 and M_2 are treated by an application of Lemma 10, as above. For M_3 we treat differently the case $n = n_j \leq c_1a^3$, were c_1 is large enough to assure that $1/(\epsilon^2 c_1)$ belongs to the natural parameter space, and the case $n > c_1a^3$. In the former case the expectation of $\mathrm{E}_0\exp\{M_1 + M_2 + M_3\} = \prod_{i=1}^3 \mathrm{E}_0\exp\{M_i\}$ is bounded. The bound follows from the bound on $\mathrm{E}_0\exp\{M_1\}$ and $\mathrm{E}_0\exp\{M_2\}$ derived above and from a bound on $\mathrm{E}_0\exp\{M_3\}$ derived below. In the latter case the probability of the event under investigation is bounded by the sum of the probability of the event $\{\ell_j + M_1 + M_2 \geq a - 4\}$ and the probability of the event $\{M_3 \geq 4\}$. The first probability is bounded with the aid of the moment generating function of M_1 and M_2 as before. The probability of the second event is shown below to be negligible.

In order to show that $\mathrm{E}_0\exp\{M_3\}$ is bounded when $n \leq c_1a^3$, we apply Lemma 2 with $m_1 = m_2 \sim n/a^2$ and $\theta \sim (2a/n)^{1/2}$. It follows that $m_i\theta = O[(n/a^3)^{1/2}]$, which is bounded by assumption. Choosing any λ in the interior of the natural parameter space in Lemma 11 would establish the needed result.

Regarding the probability of the event $\{M_3 > 4\}$ when $n > a^3$ we can use Lemma 12 with $\lambda \sim [a/m_1]^{1/2} = [a^3/n]^{1/2}$ and $\delta = 4$. It can be shown that $4\lambda/\theta \sim 2 \cdot 2^{1/2}a$, whereas $m_1\psi(\lambda) \sim a/2$. The last claim follows since $m_2 = o(e^a)$.

5 Multidimensional exponential families.

Now assume that the $X_{\mathbf{u}}$ have a distribution belonging to a multidimensional exponential family, which we write as $\exp[\langle\theta, x\rangle - \psi(\theta)]dF(x)$, where as above we assume ψ has been standardized so that $\psi(0) = 0, \dot\psi(0) = 0$. Let $\eta(\theta) = \langle \theta, \dot\psi(\theta)\rangle - \psi(\theta)$. In this section we indicate heuristically the modifications appropriate to generalize our earlier approximations. For simplicity we assume that θ is two-dimensional.

¿From the convexity of η it follows that the equation (2) has as its solution a convex curve $\theta_j = \theta_j(\omega)$, parameterized by the angular coordinate ω of the point θ. A very useful result in the calculations to follow is obtained by differentiating (2) with respect to ω to obtain

$$(D\theta_j)'\ddot\psi(\theta_j)\theta_j = 0, \tag{26}$$

where D denotes differentiation with respect to ω and prime denotes transpose.

For any interval $j \in \mathcal{J}$ and any ω let the probability $\mathrm{P}_{j,\omega}$ be defined by the likelihood ratio

$$\ell_{j,\omega} = [\langle \theta_j, S_j\rangle - n_j\psi(\theta_j)].$$

Similarly, let P_j be defined by

$$\ell_j = \log\{(2\pi)^{-1}\int_0^{2\pi} \exp[\ell_{j,\omega}]d\omega\}. \tag{27}$$

Let $\hat\theta_j = \theta_j(\hat\omega)$ be the maximum likelihood estimator of θ restricted to the curve θ_j. We have the equivalence (cf. (3))

$$\{\max_j n_j\varphi(\bar X_j) \geq a\} = \{\max_j[\langle\hat\theta_j, S_j\rangle - n_j\psi(\hat\theta_j)] \geq a\}. \tag{28}$$

Generalizing (4) and (5) we have the representation

$$2\pi e^a \mathrm{P}\left\{\max_{j\in\mathcal{J}} n_j\varphi(\bar X_j) \geq a\right\} = 2\pi e^a \Sigma_j \mathrm{E}_j\left[1/\Sigma_k \exp(\ell_k); \max_k n_k\varphi(\bar X_k) \geq a\right] =$$
$$\Sigma_j \int_0^{2\pi} \mathrm{E}_{j,\omega}\left\{\frac{\exp(\max_k \ell_k)}{\Sigma_k \exp(\ell_k)} \exp\{-[\ell_j + \max_k(\ell_k - \ell_j) - a]\}; \max_k n_k\varphi(\bar X_k) \geq a\right\} d\omega. \tag{29}$$

To analyse (29), we begin with the linear Taylor series approximation

$$n_j\varphi(\bar X_j) - a = \langle\theta_j, S_j\rangle - n_j\psi(\theta_j) - a + O_p(1) = \langle\theta_j, S_j - n_j\dot\psi(\theta_j)\rangle + O_p(1). \tag{30}$$

By a Laplace expansion of (27) we obtain

$$e^{\ell_j} \sim \exp[\langle\hat\theta_j, S_j\rangle - n_j\psi(\hat\theta_j)]/[2\pi n_j(D\theta_j)'\ddot\psi(\theta_j)(D\theta_j)]^{1/2}. \tag{31}$$

Using a Taylor series approximations of $\dot\psi(\hat\theta_j)$ and $\ell_{j,\hat\omega}$ along with (26) we see that

$$\langle\hat\theta_j, S_j\rangle - n_j\psi(\hat\theta_j) - a = \langle\hat\theta_j, S_j - n_j\dot\psi(\hat\theta_j)\rangle = \langle\theta_j, S_j - n_j\dot\psi(\theta_j)\rangle + O_p(1). \tag{32}$$

Substitution of (30) and (31) into (29) and arguing as in the preceding sections suggests that the expectation in (29)

$$\sim \mathrm{E}_{j,\omega}\{\frac{\exp[\max_k(\ell_k - \ell_j)]}{\Sigma_k \exp(\ell_k - \ell_j)}\}\{\frac{(D\theta_j)'\ddot\psi(\theta_j)(D\theta_j)}{\theta_j'\ddot\psi(\theta_j)\theta_j}\}^{1/2}. \tag{33}$$

Calculations similar to those in the proof of Lemma 5, but more complicated, show that in the expectaton in (33) $\ell_k - \ell_j$ can be replaced asymptotically by $\langle\hat\theta_j, S_k - S_j\rangle - (n_k - n_j)\psi(\hat\theta_j)$, which in turn can be replaced by a similar expression with the true values θ_j in place of the estimators $\hat{\theta}_j$. Hence this expectation can be evaluated as above to yield the appropriate multidimensional version of (12) given in Appendix A (cf. (A1)).

We are now in a position to approximate the sum in (29) by an integral. Let $\Sigma_0 = \ddot{\psi}(0)$, and let $v = (\cos\omega, \sin\omega)'$, so $\theta = ||\theta|| v$. Substituting the expressions obtained in the preceding paragraph and using the algebraic relation

$$\det\{\ddot{\psi}(\theta_j)\}||\theta_j||^4 = (D\theta_j)'\ddot{\psi}(\theta_j)(D\theta_j)\theta_j'\ddot{\psi}(\theta_j)\theta_j \tag{34}$$

proved in Appendix B, we see from (2) that the right hand side of (29) is

$$\sim m^2 a^2 [\det(\Sigma_0)]^{1/2} \int_0^{2\pi} \{\int_0^{m/a} (1/x^2 - a/mx)\nu^2[(2v'\Sigma_0 v/x)^{1/2}]dx\}^2 (v'\Sigma_0 v)^{-1} d\omega. \tag{35}$$

In the special case that $m >> a$, this can be simplified to

$$m^2 a^2 [\det(\Sigma_0)]^{1/2} \int_0^{2\pi} \{\int_0^{\infty} y\nu^2(y)dy\}^2 (v'\Sigma_0 v)^{-3} d\omega.$$

APPENDIX A

For completeness, we give here, in the form needed for this paper, an argument of Hogan and Siegmund (1986), which is important for the evaluation of (11). Let $Y_1, \ldots, Y_n, \ldots$ be independent random variables with probability distribution from the exponential family $\exp[\langle\theta, x\rangle - \psi(\theta)]dF(x)$. We assume that distribution of Y_1 when $\theta = 0$ has been centered, so that $\dot{\psi}(0) = 0$; and we write P_θ to emphasize dependence of the probability on θ. Let $\eta(\theta) = \langle\theta, \dot{\psi}(\theta)\rangle - \psi(\theta)$ and $S_n = Y_1 + \cdots + Y_n$.

Assume that $r \to \infty, \theta \to 0$ in such a way that $r\psi(\theta)$ converges to a positive constant. Suppose also that $n \to \infty$, but slowly compared to r. Then following Hogan and Siegmund (1985), we shall show that

$$n^{-1} E_0\{\max_{0<k\leq n} [\theta S_{rk} - rk\psi(\theta)]\}$$

$$\sim r\eta(\theta)\nu[(r\theta'\Sigma_0\theta)^{1/2}], \tag{A1}$$

where $\Sigma_0 = \ddot{\psi}(0)$ and $\nu(\cdot)$ is the function defined by Siegmund (1985, p. 82).

To prove (A1), we begin by noting that by integration by parts and a standard likelihood ratio identity the left hand side of (A1)

$$\sim n^{-1} \int_0^\infty e^x P_0\{\max_{k\leq n}[\langle\theta, S_{rk}\rangle - rk\psi(\theta)] \geq x\}dx$$

$$\sim n^{-1} \int_0^\infty E_\theta\{\exp[-(\langle\theta, S_{r\tau}\rangle - r\tau\psi(\theta) - x)]; \tau \leq n\}dx, \tag{A2}$$

where

$$\tau = \tau_x = \min\{k : \langle\theta, S_{rk}\rangle - rk\psi(\theta) \geq x\}.$$

Under the probability P_θ the random variable $\langle\theta, S_{rk}\rangle - rk\psi(\theta)$ has expectation $rk\eta(\theta)$ and variance $rk\theta'\ddot{\psi}(\theta)\theta$, both of which converge (for fixed k) to positive constants. Hence by a simple law of large numbers argument $P_\theta(\tau \leq n)$ converges to 0 for $n < (1-\epsilon)x/r\eta(\theta)$

and to 1 for $n > (1+\epsilon)x/r\eta(\theta)$. It follows that (A2) is asymptotically bounded above and below by

$$(1 \pm \epsilon) r\eta(\theta) \lim_{x} \mathrm{E}_\theta\{\exp[-(\langle \theta, S_{r\tau}\rangle - r\tau\psi(\theta) - x)]\}. \qquad (A2)$$

The increments $\langle \theta, S_r\rangle - r\psi(\theta)$ are easily seen by the central limit theorem to be asymptotically normally distributed with mean $r\theta'\Sigma_0\theta/2$ and variance $r\theta'\Sigma_0\theta$, so the limit in (A2) is the same as it would be for Gaussian random walk, which is one way of defining the function ν (Siegmund 1985, Chapter 8).

APPENDIX B

Proof of (34). To simplify the notation let $\Sigma = \ddot{\psi}(\theta_j)$, $a = \theta_j$, and $b = D\theta_j$. In this notation equation (26) becomes $a'\Sigma b = 0$. It is easy to see by writing $\theta_j = ||\theta_j||(\cos\omega, \sin\omega)'$ and differentiating that $\det(a,b) = ||\theta_j||^2$. Hence

$$(a'\Sigma a)(b'\Sigma b) = \det[(a,b)'\Sigma(a,b)]$$

$$= \det(\Sigma)[\det(a,b)]^2 = \det(\Sigma)||\theta_j||^4,$$

which is equivalent to (34).

Acknowledgment

The research of the first author was supported by the National Science Foundation. That of the second author was supported by a grant from the Israeli Academy of Science.

REFERENCES

Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*, Springer-Verlag, New York.

Arratia, R., Goldstein, L. and Gordon L. (1989). Two moments suffice for Poisson approximation: The Chen-Stein method, *Ann. Probab.* **17**, 9-25.

Bickel, P. and Rosenblatt, M. (1973). Two-dimensional random fields, in *Multivariate Analysis, III* (P. K. Krishnaiah, editor), Academic Press, New York, 3-15.

Giller, G. (1994). The construction and analysis of a whole-sky map using underground muons, Ph. D. Thesis, University of Oxford.

Hogan, M.L. and Siegmund, D. (1986). Large deviations for the maxima of some random fields, *Advances in Appl. Math.* **7**, 2-22.

Karlin, S., Dembo, A. and Kawabata, T. (1990). Statistical composition of high-scoring segments from molecular sequences, *Ann. Statist.* **18**, 571-581.

Levin, B. and Kline, J. (1985). The cusum test of homogeneity with an application to spontaneous abortion methodology, *Statist. in Medicine* **4**, 469-488.

Loader, C.R. (1991). Large-deviation approximations to the distribution of scan statistics, *Advances in Appl. Probab.* **23**, 751-771.

Pickands, J. (1969). Upcrossing probabilities for stationary Gaussian processes *Trans. Amer. Math. Soc.* **145**, 51-73.

Qualls, C. and Watanabe, H. (1973). Asymptotic properties of Gaussian random fields, *Trans. Amer. Math. Soc.* **177**, 155-171.

Rabinowitz, D. (1994). Detecting clusters in disease incidence. *Change-point Problems* (E. Carlstein, H.-G. Müller, and D. Siegmund, eds.), IMS, Hayward California, 255-275.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlaq, New York-Heidelberg-Berlin.

Siegmund, D. (1988). Tail probabilities for the maxima of some random fields. *Ann. Probab.* **16**, 487-501.

Siegmund, D. (1992). Tail approximations for maxima of random fields. In *Probability Theory: Proceedings of The Singapore Probability Conference* (L. Chen et al., editors), Springer-Verlag, 147-158.

Siegmund, D. and Venkatraman, E.S. (1995) Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist.* **23**, 255-271.

Tu, I-Ping (1997). Theory and application of scan statistics, Stanford University dissertation.

Woodroofe, M. (1976). Frequentist properties of Bayesian sequential tests, *Biometrika* **63**, 101-110.

Woodroofe, M. (1982). *Nonlinear renewal theory in sequential analysis*, Society for Industrial and Applied Mathematics, Philadelphia.

Yakir, B. and Pollak, M. (1998). A new representation for a renewal-theoretic constant appearing in asymptotic approximations of large deviations, *Ann. Appl. Probab.* **8** 749-774.