

## THE APPROXIMATE DISTRIBUTION OF THE MAXIMUM OF A SMOOTHED POISSON RANDOM FIELD

Daniel Rabinowitz and D. Siegmund

*Columbia University and Stanford University*

*Abstract:* The distribution of the maximum score statistic for detecting a signal of known shape, but unknown amplitude, location, and scale is discussed when the underlying noise process is a homogeneous Poisson process. The approximation is based on an exponential change of measure to evaluate asymptotically the expected number of local maxima of a random field.

*Key words and phrases:* Maximum of a random field, Poisson process, signal detection.

### 1. Introduction

Let  $dW(x)$  denote Gaussian white noise in  $r$  dimensional Euclidean space, and assume  $f \geq 0$  is twice continuously differentiable and satisfies

$$\int f^2(x)dx = 1. \quad (1.1)$$

Under certain conditions the random variable

$$X_{\max} = \max X(t, \sigma), \quad (1.2)$$

where

$$X(t, \sigma) = \sigma^{-r/2} \int f[\sigma^{-1}(x - t)]dW(x), \quad (1.3)$$

is the likelihood ratio statistic for detecting a signal of known shape  $f$ , unknown amplitude, location  $t$ , and scale  $\sigma$  (cf. Siegmund and Worsley (1995)). The maximization is over  $t$  in a region  $C$  of possible locations in  $r$  dimensional Euclidean space and over  $\sigma$  in an interval of possible scale parameters:  $[\sigma_0, \sigma_1]$ .

The three applications cited by Siegmund and Worsley (1995) as motivating their work are (i) searching the celestial sphere for a point source of muons against the background of cosmic radiation (Giller (1994)), (ii) searching PET maps for signals (Worsley et al. (1992); Worsley (1995)), and (iii) searching for hot spots of disease incidence (Rabinowitz (1994)). These all involve underlying point processes, so the assumed normality of  $dW$ , hence (1.3), implicitly invokes a central limit approximation. Indeed, the classical papers of Rice (1944, 1945)

use a similar normal approximation for shot noise in vacuum tubes. Siegmund and Worsley (1995) conjecture that their approximation to the distribution of (1.2) is quite accurate under the assumed normality of (1.3). However, one might reasonably question that assumed normality, especially since extremely small tail probabilities for (1.3) will be involved in obtaining a small type one error probability for (1.2).

The purpose of this paper is to develop an approximation for the distribution of (1.2) for  $X$  defined by

$$X(t, \sigma) = \sigma^{-r/2} \int f[\sigma^{-1}(x - t)][dN_x - \lambda dx] / \lambda^{1/2}, \quad (1.4)$$

where  $N$  is a Poisson random field of rate  $\lambda$  in  $r$  dimensional Euclidean space. In the first instance the maximum is over  $t$  with  $\sigma$  fixed; in the second the maximum is over  $t$  and  $\sigma$ . Although the assumption of an underlying homogeneous Poisson field is very special, close examination of this case gives a good sense of the accuracy of a normal approximation and the tradeoff between computational effort and improved accuracy one can expect to find by analyzing the underlying point process.

The paper is organized as follows. In Section 2 we derive (1.4) as a score statistic for a particular hypothesis testing problem. In Section 3 we assume  $\sigma$  is fixed and sketch our approach to approximating the distribution of  $\max_t X(t, \sigma)$ , which involves a combination of exponential tilting and Gaussian process techniques. We consider several different approximations: one based on large deviation scaling and others based on “intermediate deviation” scaling, which are substantially easier to evaluate. In Section 4 we consider simultaneous maximization over  $t$  and  $\sigma$ , which gives us an idea of the difficulties involved in using our approximation in more general settings. In particular a non-homogeneous Poisson field leads to similar complications.

The distribution of (1.4) is quite different when  $f$  is the indicator of a rectangle. See Loader (1991).

## 2. A Class of Statistical Problems

Let  $f \geq 0$  satisfy (1.1). Assume that  $N$  is a nonhomogeneous Poisson process with intensity at the point  $x$  equal to

$$\lambda \exp\{\xi \sigma^{-r/2} f[(x - t)/\sigma]\}. \quad (2.1)$$

We shall assume  $\lambda$  is known, while the parameters  $\xi \geq 0$ ,  $\sigma > 0$ ,  $t \in C$  are in general all unknown. To avoid questions of edge effects, we assume that  $N(x)$

is observed throughout the  $r$  dimensional Euclidean space. Up to an additive constant the log likelihood function is

$$\xi \int \sigma^{-r/2} f[(x-t)/\sigma] dN_x - \lambda \int \{\exp[\xi \sigma^{-r/2} f[(x-t)/\sigma] - 1] dx. \quad (2.2)$$

Differentiation with respect to  $\xi$  gives the efficient score

$$\int \sigma^{-r/2} f[(x-t)/\sigma] dN_x - \lambda \int \sigma^{-r/2} f[(x-t)/\sigma] \exp\{\xi \sigma^{-r/2} f[(x-t)/\sigma]\} dx. \quad (2.3)$$

If  $t, \sigma$  were both known, we would obtain the score statistic for testing  $H_0 : \xi = 0$  by setting  $\xi = 0$  in (2.3) and standardizing the result to have unit variance under  $H_0$ . This yields (1.4). When  $t$  or the pair  $t, \sigma$  are unknown, we can test  $\xi = 0$  by maximizing (1.4) over those unknown parameters.

The background noise process, here a homogeneous Poisson field of rate  $\lambda$ , may be observed for a long enough time over a large enough region that it is reasonable to regard  $\lambda$  as known. The astronomical example mentioned above appears to be in this category. There are undoubtedly also applications where it is more reasonable to regard  $\lambda$  as unknown or to want to model the background noise by a completely different process. To deal with the comparatively simple problem of unknown  $\lambda$ , we can (i) assume that the random field is observed over a bounded region  $\tilde{C}$ , which contains  $C$ , and then some if we are to avoid edge effects, (ii) replace the known value of  $\lambda$  in (1.4) by the estimated value under the null hypothesis  $\xi = 0$ , to wit  $\hat{\lambda}_0 = N(\tilde{C})/\text{measure}(\tilde{C})$ , and (iii) compute all probabilities conditionally, given  $N(\tilde{C})$ . See, for example, Loader (1991). The issue of an appropriate model for the noise process and estimation of nuisance parameters of the model is tied to the specific application and sample size, and will not be discussed here.

### 3. Known $\sigma$

In this section we assume  $\sigma$  is known and without loss of generality set  $\sigma = 1$ . To simplify the notation we write  $X_t$  instead of  $X(t, 1)$ , for the random field defined in (1.4), so  $X_{\max} = \max_{t \in C} X_t$ . We write  $E_\xi$  and  $P_\xi$  to denote expectation and probability as a function of the parameter  $\xi$ . Dependence on the parameter  $t$  is suppressed in this notation. We shall also assume that  $f$  is symmetric in each of its arguments, smooth, and vanishes rapidly at infinity. Although the symmetry condition is not strictly necessary, it leads to quite substantial simplifications in the form of the final approximation. Our principal example is the Gaussian kernel

$$f(x) = \pi^{-r/4} \exp(-\|x\|^2/2). \quad (3.1)$$

Let  $Z_t = \lambda^{1/2} X_t$ ,  $\psi(\theta) = \log E_0[\exp(\theta Z_t)]$ , and note the likelihood ratio

$$dP_\xi/dP_0 = \exp[\xi Z_t - \psi(\xi)]. \quad (3.2)$$

Straightforward calculations yield

$$\psi(\theta) = \lambda \int \{\exp[\theta f(x)] - 1 - \theta f(x)\} dx, \quad (3.3)$$

$$\psi'(\theta) = E_\theta(Z_t) = \lambda \int f(x) \{\exp[\theta f(x)] - 1\} dx, \quad (3.4)$$

and

$$\psi''(\theta) = \text{Var}_\theta(Z_t) = \lambda \int f^2(x) \exp[\theta f(x)] dx. \quad (3.5)$$

We shall write  $D$  with coordinates  $D_i$  to denote differentiation. Thus the gradient of  $f$  is  $Df = (D_1 f, \dots, D_r f)^T$  and similarly its Hessian is  $D^2 f = (D_{i,j}^2 f)$ . By the assumed symmetry of  $f$ ,

$$E_\theta(DZ_t) = -\lambda \int Df(x) \{\exp[\theta f(x)] - 1\} dx = 0, \quad (3.6)$$

and hence

$$\text{Cov}_\theta(D_i Z_t, D_j Z_t) = \delta_{i,j} \lambda \int [D_i f(x)]^2 \exp[\theta f(x)] dx, \quad (3.7)$$

where  $\delta_{i,j}$  is 1 or 0 according as  $i = j$  or  $i \neq j$ . Also

$$E_\theta(D^2 Z_t) = -\theta \text{Cov}_\theta(DZ_t), \quad E_\theta(D^2 Z_t DZ_t) = 0, \quad (3.8)$$

$$\text{Cov}_\theta(Z_t, D^2 Z_t) = \lambda \int f(D^2 f) \exp(\theta f) dx, \quad \text{and} \quad \text{Cov}_\theta(Z_t, DZ_t) = 0. \quad (3.9)$$

With this notation we can formulate our basic result.

**Theorem.** *Assume  $b \rightarrow \infty, \lambda \rightarrow \infty$ , in such a way that  $b/\lambda^{1/2} \rightarrow c$ , where  $c$  is a positive finite constant. Let  $\theta_0$  denote the unique positive value of  $\theta$  satisfying*

$$\psi'(\theta_0) = b\lambda^{1/2}. \quad (3.10)$$

*Also let  $I(\theta) = \theta\psi'(\theta) - \psi(\theta)$ , and let  $|C|$  denote the Lebesgue measure of  $C$ . Then*

$$P_0\{X_{\max} \geq b\} \approx 1 - \exp[-E_0(M_C)], \quad (3.11)$$

*where  $M_C$  is the number of local maxima of the field  $X_t, t \in C$  of height  $\geq b$ , given approximately for large  $b$  by*

$$E_0\{M_C\} \sim \theta_0^{r-1} \exp[-I(\theta_0)] (2\pi)^{-(r+1)/2} |C| \{[\Pi_i \text{Var}_{\theta_0}(D_i Z_t)] / \text{Var}_{\theta_0}(Z_t)\}^{1/2}. \quad (3.12)$$

There are two possible interpretations of the approximation in (3.11). (i)  $C$  remains fixed or  $|C|$  increases so slowly that the right hand side of (3.12) converges to zero. In that case the approximation denotes asymptotic equivalence, i.e., the ratio of the two sides tends to one. (ii) The value of  $|C|$  increases sufficiently rapidly as a function of  $b$  that the right hand side of (3.12) is bounded away from zero and infinity. In that case the approximation in (3.11) indicates that the difference of the two sides tends to zero. We give an informal discussion of (3.12) below, which appears to contain the ingredients of a rigorous proof. We shall not attempt a proof of (3.11), although presumably one can be given along established lines, e.g., Siegmund (1988) and Kim and Siegmund (1988). See Aldous (1989), Rabinowitz (1994) or Siegmund and Worsley (1995) for heuristic discussions of this approximation for Gaussian processes, and Adler (1981) for a more rigorous treatment and additional references.

The expected number of local maxima of height at least  $b$  of the random field  $X_t, t \in C$  involves the integral over  $C$  of a quantity calculated from the joint probability (density) that  $X_t \geq b, DZ_t = 0$ , and  $D^2Z_t = z$ . To obtain an approximation to this probability and hence (3.12) we employ an exponential change of measure and a local central limit theorem. Before presenting some details of the calculation we examine (3.12) more closely.

Although it is unnecessary to utilize the exponential family  $P_\theta$  in the Gaussian case, i.e., when  $X_t$  is defined by (1.3) with  $\sigma = 1$ , we can, nevertheless use this approach. We obtain  $\psi(\theta) = \lambda\theta^2/2 = I(\theta)$ , so (3.10) becomes the linear equation  $\theta_0 = b/\lambda^{1/2}$ , and (3.11)-(3.12) agrees with a well known result (e.g., Adler (1981), p. 160).

When  $C$  is small compared to its boundary, we can add a boundary correction to the right hand side of (3.12), which will contain  $\theta_0$  to one less power and for a Gaussian process will agree with the corresponding boundary correction in Siegmund and Worsley (1995). However, in the analysis that produces (3.12) we incur an error in our application of the local central limit theorem, which appears to be of the same order of magnitude as the boundary correction. Although one can reasonably expect that including a boundary correction will lead to a better approximation, there is no mathematically rigorous guarantee that this will occur, even asymptotically. For the numerical examples given in this section, we have chosen  $C$  large compared to the length of its boundary, so the boundary correction is negligible.

The right hand side of (3.12), which in principle involves the integral of a function of  $t$  over  $C$ , has in this case simplified quite considerably because of the stationarity of  $X_t$ . The constant on the right hand side, although written as if it is a function of  $t$ , in fact does not depend on  $t$ . When we do not have this stationarity, the solution of equation (3.10) depends on the value of  $t$ , and an

expression like the right hand side of (3.12) (except for the factor  $|C|$ ) must be integrated numerically over  $C$ . This can make the evaluation of the approximation very complicated in other contexts. A nonstationary example is presented in Section 4.

Assuming (as will be demonstrated numerically below) that (3.11)-(3.12) is often substantially more accurate than a Gaussian process approximation, and especially in view of the numerical difficulties mentioned in the preceding paragraph if one wants to implement this approximation more generally, one naturally asks if there is a simpler approximation that achieves most of the gain in accuracy of (3.12). One possibility, which takes into account the skewness of the Poisson distribution but ignores other features, is to replace  $b/\lambda^{1/2} \rightarrow c$  by  $b/\lambda^{1/4} \rightarrow 0$  (cf. Feller (1972), p. 552). By (3.4) and (3.10)  $\theta_0 \rightarrow 0$ , so

$$\Pi_i \text{Var}_{\theta_0}(D_i Z_t) / \text{Var}_{\theta_0}(Z_t) \sim \theta_0^r \lambda^{(r-1)/2} \frac{\Pi_i \int (D_i f)^2 (1 + \theta_0 f + O(\theta_0^2)) dx}{(1 + \theta_0 \int f^3 + O(\theta_0^2))}. \quad (3.13)$$

To analyse the exponential factor in (3.12) we use the expansion  $I(\theta) = \frac{1}{2}\lambda\theta^2[1 + 2c_3\theta/3 + O(\theta^2)]$ , where  $c_3 = \int f^3 dx$ , and use for  $\theta_0$  the solution of the quadratic equation obtained by the corresponding expansion of (3.10):

$$b\lambda^{1/2} = \lambda(\theta_0 + \frac{1}{2}c_3\theta_0^2). \quad (3.14)$$

Hence we obtain

**Corollary.** *If  $b/\lambda^{1/4} \rightarrow 0$ , then*

$$E_0(M_C) \sim |C| \frac{(\theta_0 \lambda^{1/2})^{r-1}}{(2\pi)^{(r+1)/2}} \exp[-\lambda(\theta_0^2/2 + c_3\theta_0^3/3)] \left\{ \frac{\Pi_i \int (D_i f)^2 (1 + \theta_0 f) dx}{1 + \theta_0 \int f^3 dx} \right\}^{1/2}, \quad (3.15)$$

where  $\theta_0$  is the positive root of (3.14).

In (3.15) the Gaussian process approximation has been corrected for skewness of the Poisson distribution. If we relax the growth condition on  $b$  to  $b/\lambda^{3/10} \rightarrow 0$ , we obtain an approximation that also corrects for kurtosis. We expand  $I(\theta)$  and the ratio of variances on the right hand side of (3.12) one more term and use in (3.10) one more term in the expansion of  $\psi'(\theta)$ . The equation (3.14) is replaced by a cubic which is easily solved in two or three steps of a Newton-Raphson iteration starting from  $b/\lambda^{1/2}$ . We omit the details of this calculation.

Figure 1 gives a numerical example. The set  $C$  is a disc in the plane of radius  $R = 10$ , so  $r = 2$  and  $|C| = \pi R^2$ ; and  $\lambda = 10$ . The Figure displays (i) a Gaussian process approximation to  $P\{X_{\max} \geq b\}$ , (ii) the large deviation approximation (3.11)-(3.12), (iii) the simpler approximation from (3.14) and (3.15), and (iv) the

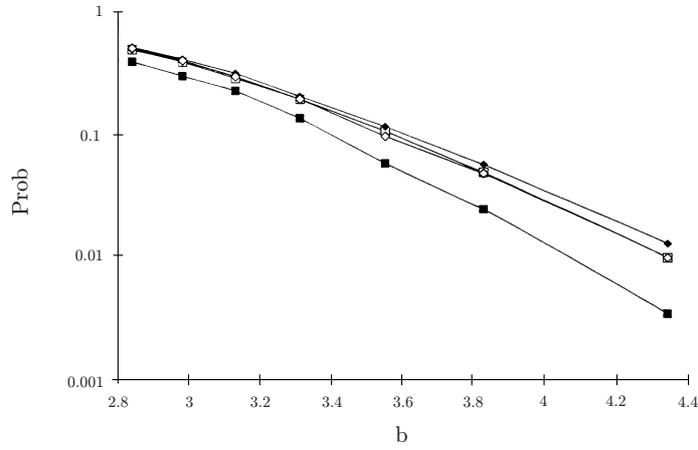


Figure 1. Comparison of approximations of the tail probability for the maximum of the random field (1.4) (with  $\sigma = 1$ ): (i) Gaussian process approximation  $\blacksquare$ , (ii) first order corrected approximation based on (3.14)-(3.15)  $\square$ , (iii) large deviation approximation based on (3.10)-(3.12)  $\blacklozenge$ , and (iv) Monte Carlo simulation based on 2500 repetitions  $\diamond$ .

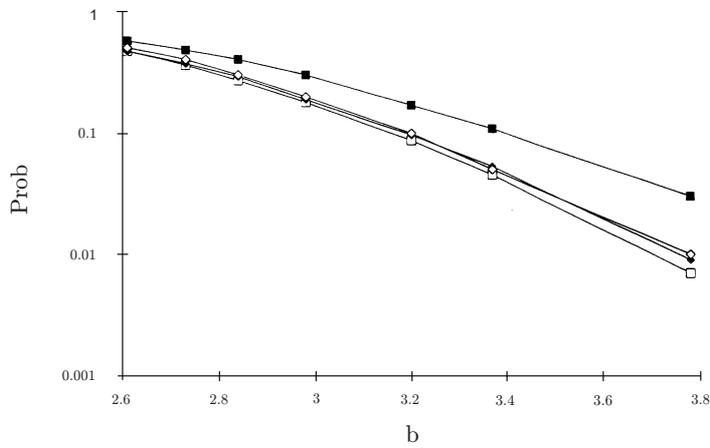


Figure 2. Comparison of the approximations of the tail probability for the minimum of the random field (1.4) (with  $\sigma = 1$ ): (i) Gaussian process approximation  $\blacksquare$ , (ii) first order corrected approximation based on (3.14)-(3.15)  $\square$ , (iii) large deviation approximation based on (3.10)-(3.12)  $\blacklozenge$ , and (iv) Monte Carlo simulation based on 2500 repetitions  $\diamond$ .

results of a Monte Carlo experiment based on 2500 realizations of the random field (1.4) with  $\sigma = 1$ . The numerical results indicate that the Gaussian process approximations are always too small (anti-conservative), often considerably so. The putatively more accurate approximations are all about equally accurate. Thus in practice, (3.15), which is much easier to evaluate than (3.12) and provides only slightly different values, might reasonably be the preferred method. If one looks in more detail at the numerical results, one sees some favorable cancellation in the various constituent parts of (3.15). The next order approximation, obtained by expanding (3.10) and (3.12) to one more term in  $\theta$ , gives numerical results in slightly better agreement with (3.12), with the constituent factors in substantially better agreement. Figure 2 gives similar numerical results for the minimum over  $t$  of the random field (1.4), or equivalently the maximum of the negative of the random field. In this case the Gaussian process approximation substantially overestimates the true probability, while the other three approximations are in close agreement with each other and the Monte Carlo simulations.

The terms of order  $\theta_0$  in the right hand factor of (3.15) need not be included for the indicated asymptotic result to be valid. Omitting them leads to a slightly larger, but still numerically insignificant discrepancy between the simplified approximation and the large deviation approximation.

**Remark.** The simplified approximation based on the quadratic equation (3.14) involves some technical problems, since that equation has two roots, whereas (3.10) has only one. Since we are interested in the positive root, which in fact is larger than  $b/\lambda^{1/2}$ , the correct root is easily recognized in practice. If we were interested in the *minimum* of the random field, the problem would be potentially more complicated. For small  $\lambda$  the quadratic equation has no real roots, so the simplified approximation (3.15) does not exist, although the second simplified approximation based on a cubic equation does have the desired root. This problem is in principle worse in the following section where the simplified approximation involves simultaneous solution of two quadratic equations, hence possibly four roots. In practice it is easy to identify the appropriate roots by their proximity to the roots of the linear equations corresponding to the Gaussian case, which serve as starting points for iterative solutions.

To derive (3.12) we start from the standard representation of  $E_0(M_C)$  as the integral over  $C$  of

$$E_0\{\|D^2 Z_t\|; Z_t \geq b\lambda^{1/2}, DZ_t = 0, D^2 Z_t < 0\},$$

where the double vertical bars indicate the absolute value of the determinant and the inequality  $D^2 Z_t < 0$  indicates negative definiteness of the matrix. By (3.2)

and (3.10) this equals

$$\exp[-I(\theta_0)]E_{\theta_0}\{\|D^2Z_t\|\exp[-\theta_0(Z_t - b\lambda^{1/2})]; Z_t \geq b\lambda^{1/2}, DZ_t = 0, D^2Z_t < 0\}. \tag{3.16}$$

The expectation in (3.16) can be expressed as

$$\int_0^\infty e^{-\theta_0 y} E_{\theta_0}\{\|D^2Z_t\|; D^2Z_t < 0 | Z_t = b\lambda^{1/2} + y, DZ_t = 0\} P_{\theta_0}\{Z_t \in b\lambda^{1/2} + dy, DZ_t = 0\}. \tag{3.17}$$

Given  $Z_t = E_{\theta_0}(Z_t) + y, DZ_t = 0$ , one can see from the moment relations (3.6)-(3.9) that  $D^2Z_t$  is approximately normally distributed with mean  $E_{\theta_0}(D^2Z_t) + O(y)$  and variance  $O(\lambda)$ . By virtue of the exponential in (3.17) it is only necessary to integrate over  $y$  in a large bounded interval, so the term  $O(y)$  is negligible compared to  $E_{\theta_0}(D^2Z_t)$ , which is of order  $\lambda$ . Moreover, since the conditional standard deviation is of order  $\lambda^{1/2}$ , the variability in  $D^2Z_t$  can be ignored for the purpose of evaluating the conditional expectation in (3.17), which, therefore, by (3.7) and (3.8)

$$\sim \|E_{\theta_0}(D^2Z_t)\| = \theta_0^r \|E_{\theta_0}(DZ_t D^T Z_t)\| = \theta_0^r \Pi_i \text{Var}_{\theta_0}(D_i Z_t). \tag{3.18}$$

By (3.7), (3.9) and a local limit theorem, for bounded  $y$

$$P_{\theta_0}\{Z_t \in E_{\theta_0}(Z_t) + dy, DZ_t = 0\} \sim (2\pi)^{-(r+1)/2} [\text{Var}_{\theta_0}(Z_t) \|E_{\theta_0}[(DZ_t)(D^T Z_t)]\|]^{-1/2} dy.$$

Substituting this and (3.18) into (3.17) and performing the indicated integration yield (3.12).

**4. Unknown  $\sigma$**

The case of unknown  $\sigma$  is treated similarly, but the methods sketched above are substantially trickier to implement. In particular we must imbed  $P_0$  in a two dimensional exponential family and replace (3.10) by two simultaneous equations. The quantity corresponding to (3.16) depends on the point in the random field where it is evaluated and hence must be integrated numerically over the set indexing the field—in this case over the range of  $\sigma$ . The approximation corresponding to (3.14) involves replacement of two transcendental equations by two quadratic equations ((4.14) and (4.15) below) and hence requires substantially less computation. It is easy to see from the results of Siegmund and Worsley (1995) that a boundary correction at the minimum value of  $\sigma$  is necessary to obtain a reasonable approximation in this case. The reason is that small values of  $\sigma$  lead to large fluctuations in the random field (1.3) or (1.4). Indeed, for a

broad range of thresholds a reasonably good approximation is obtained by considering  $\sigma$  to be fixed at its minimum value and maximizing over  $t$  alone. (This cannot be true without some qualification because the boundary term with  $\sigma$  fixed is smaller by an order of magnitude when  $b \rightarrow \infty$  than the term involving the maximum over  $\sigma$ ; but numerically in the range of  $b$  of primary interest, say  $2.5 \leq b \leq 5$ , it works out that an approximation that is wrong in principle can be reasonably accurate in practice. The situation is completely different at the maximum value of  $\sigma$ , which plays practically no role in the approximation. The approximation approaches a limit as the maximum value becomes infinite, and for most numerical purposes one can just use that limiting value.) The primary contribution of the following discussion may be to indicate the methodology required and the difficulties that will be encountered in other more complex problems. In particular, a non-homogeneous Poisson field with intensity  $\lambda = \lambda(x)$  leads to similar calculations.

Let  $X$  be defined by (1.4). It will be helpful to introduce  $s = (\log \sigma)$  and put

$$Y(t, s) = \int f[e^{-s}(x-t)](dN_x - \lambda dx), \quad Z(t, s) = e^{-rs/2}Y(t, s).$$

For ease of exposition we shall take  $r = 1$ ;  $D_2$  denotes partial differentiation with respect to  $s$ , while  $D_1$  denotes partial differentiation with respect to  $x$  or  $t$ . Also put  $f_1(x) = xD_1f(x)$ .

Let  $\psi(\theta_1, \theta_2; s) = \log E_0\{\exp[\theta_1 Y(t, s) + \theta_2 D_2 Y(t, s)]\}$ . A simple calculation gives

$$\psi(\theta_1, \theta_2; s) = \lambda e^s \int \{\exp[\theta_1 f - \theta_2 f_1] - 1 - (\theta_1 f - \theta_2 f_1)\} dx. \quad (4.1)$$

Since  $D_2 Z(t, s) = -Z(t, s)/2 + e^{-s/2} D_2 Y(t, s)$ , for  $\tilde{\psi}(\xi_1, \xi_2; s) = \log E_0\{\exp[\xi_1 Z(t, s) + \xi_2 D_2 Z(t, s)]\}$  we obtain

$$\tilde{\psi}(\xi_1, \xi_2; s) = \psi(\theta_1, \theta_2; s), \quad (4.2)$$

where  $\theta_1 = e^{-s/2}(\xi_1 - \xi_2/2)$ ,  $\theta_2 = e^{-s/2}\xi_2$ . Let the probability  $P_\xi$  be defined by the likelihood ratio

$$dP_\xi/dP_0 = \exp[\xi_1 Z(t, s) + \xi_2 D_2 Z(t, s) - \tilde{\psi}(\xi_1, \xi_2; s)]. \quad (4.3)$$

We shall be interested in the particular values  $(\xi_1, \xi_2)$  (for which we continue to use the same notation rather than introducing another subscript as in Section 3) such that

$$E_\xi[Z(t, s)] = b\lambda^{1/2}, \quad E_\xi[D_2 Z(t, s)] = 0.$$

Straightforward calculation of these expectations allows us to express the equations analytically as

$$be^{-s/2}/\lambda^{1/2} = \int f\{\exp[\theta_1 f - \theta_2 f_1] - 1\} dx, \quad (4.4)$$

and

$$-be^{-s/2}/2\lambda^{1/2} = \int f_1 \{\exp[\theta_1 f - \theta_2 f_1] - 1\} dx. \quad (4.5)$$

Let  $[s_0, s_1]$  denote the range over which  $s$  can vary and let  $\tilde{C} = C \times [s_0, s_1]$ . For an approximation analogous to (3.11) we now want to use in the exponent  $E_0(M_{\tilde{C}})$ , the expected number of local maxima of  $Z(t, s)$ ,  $(t, s) \in \tilde{C}$  of height  $\geq b\lambda^{1/2}$ . For large  $b$  the reasoning leading to (3.16) in conjunction with (4.3), (4.4) and (4.5) show that  $E_0(M_{\tilde{C}})$  is given by the integral over  $\tilde{C}$  of

$$\exp[-I(\xi)] E_\xi \{ \|D^2 Z\| \exp[-\xi_1(Z - b\lambda^{1/2})]; Z \geq b\lambda^{1/2}, DZ = 0, D^2 Z < 0 \}, \quad (4.6)$$

where  $I(\xi) = \xi_1 E_\xi(Z) - \tilde{\psi}(\xi_1, \xi_2; s)$  and the dependence of  $Z$  on the point  $(t, s)$  has been suppressed.

One can evaluate the expectation in (4.6) in terms of the following moments. We let  $f_1(x) = x f'(x)$  as above, and  $f_2(x) = x^2 f''(x)$ . Then

$$\text{Var}_\xi(Z) = \lambda \int f^2 \exp(\theta_1 f - \theta_2 f_1) dx; \quad (4.7)$$

$$\text{Var}_\xi(D_1 Z) = \lambda e^{-2s} \int (D_1 f)^2 \exp(\theta_1 f - \theta_2 f_1) dx; \quad (4.8)$$

$$\text{Var}_\xi(D_2 Z) = \lambda \int (f/2 + f_1)^2 \exp(\theta_1 f - \theta_2 f_1) dx; \quad (4.9)$$

$$\text{Cov}_\xi(Z, D_2 Z) = -\lambda \int (f^2/2 + f_1 f) \exp(\theta_1 f - \theta_2 f_1) dx; \quad (4.10)$$

$$\text{Cov}_\xi(Z, D_1 Z) = 0, \quad \text{Cov}_\xi(D_1 Z, D_2 Z) = 0; \quad (4.11)$$

$$E_\xi(D_1^2 Z) = \lambda e^{-3s/2} \int D_1^2 f [\exp(\theta_1 f - \theta_2 f_1) - 1] dx, \quad E_\xi(D_{12}^2 Z) = 0; \quad (4.12)$$

$$E_\xi(D_2^2 Z) = \lambda e^{s/2} \int (3f_1/2 + f_2) [\exp(\theta_1 f - \theta_2 f_1) - 1] dx. \quad (4.13)$$

Although the evaluation of (4.6) is straightforward in principle, since (4.4) and (4.5) depend on  $s$ , the solution of those equations, substitution into (4.6) and integration with respect to  $s$  is onerous. In view of the numerical accuracy of the simplified approximation based on (3.14) in Section 3, we consider only that case in detail here. Since  $\int f^k f_1 = -(k+1)^{-1} \int f^{k+1}$ , a Taylor series expansion of equations (4.4) and (4.5) up to the quadratic terms yields

$$be^{-s/2}/\lambda^{1/2} = \theta_1 + \theta_2/2 + \theta_1^2 \int f^3/2 + \theta_1 \theta_2 \int f^3/3 + \theta_2^2 \int f f_1^2/2 \quad (4.14)$$

and

$$be^{-s/2}/\lambda^{1/2} = \theta_1 + 2\theta_2 \int f_1^2 + \theta_1^2 \int f^3/3 + 2\theta_1 \theta_2 \int f f_1^2 - \theta_2^2 \int f_1^3. \quad (4.15)$$

The expansion of  $I(\xi)$  up to the cubic terms gives

$$\frac{1}{2}\lambda e^s \left\{ \theta_1^2 + \theta_1 \theta_2 + \left( \frac{1}{2} - \int f_1^2 \right) \theta_2^2 + \frac{2}{3} \int f^3 \theta_1^3 + \frac{5}{6} \int f^3 \theta_1^2 \theta_2 + \int f^3 \theta_1 \theta_2^2 / 3 + \int (f f_1^2 / 2 - f_1^3 / 3) \theta_2^3 \right\}. \quad (4.16)$$

The expectation in (4.6) is easily evaluated by expanding the appropriate moments in (4.7)-(4.13) with  $\xi_1 \sim b/\lambda^{1/2} \rightarrow 0$ ,  $\xi_2 \sim 0$ . The leading term is easily seen to equal

$$\exp[-\hat{I}(\xi)] \xi_1 \lambda^{1/2} (2\pi)^{-3/2} e^{-s} \left\{ \int (f')^2 \left( \int f_1^2 - 1/4 \right) \right\}^{1/2}, \quad (4.17)$$

where  $\hat{I}(\xi)$  is the cubic approximation to  $I(\xi)$  given in (4.16).

If we drop the quadratic terms in (4.14) and (4.15), the solution of the resulting linear equations is  $\xi_1 = b/\lambda^{1/2}$ ,  $\xi_2 = 0$ , as expected. Using these as initial values we have solved (4.14) and (4.15) numerically by using three steps of a Newton-Raphson iteration, and substituted the solution into (4.16), then integrated (4.17)  $\times |C|$  with respect to  $s$  to give an approximation to  $P\{X_{\max} \geq b\}$ . For this field the greatest fluctuations occur when  $\sigma$  equals its minimum value, so a boundary correction at that minimum value appears necessary before one can hope to have a reasonable overall approximation. The expression obtained by Siegmund and Worsley (1995) for the Gaussian case illustrates this phenomenon. (For small  $C$  one might want to add other boundary corrections as well.) The boundary term involves an expectation similar to (4.6), but modified as follows: (i)  $s$  is set equal to its minimum value,  $s_0$ ; (ii) the conditions  $DZ = 0$ ,  $D^2Z < 0$  in (4.6) are replaced by  $D_1Z = 0$ ,  $D_1^2Z < 0$ ,  $D_2Z < 0$ , so that we count local maxima in  $t$  that have not already been counted as local maxima in  $t, s$  for  $s > s_0$ ; (iii) the exponential becomes  $\exp[-\xi_1(Z - b\lambda^{1/2}) - \xi_2 D_2Z]$ , since we no longer have  $D_2Z = 0$  (cf. (4.3)). It is easy to see from (4.14) and (4.15) that  $\xi_2 = O(\xi_1^2)$  and hence the boundary term, to be added to the preceding approximation, is

$$\sim |C| \exp[-\hat{I}(\xi)] (4\pi)^{-1} e^{-s_0} \left[ \int (f')^2 dx \right]^{1/2}. \quad (4.18)$$

In the Gaussian case this boundary correction is just one-half the leading term of the approximation to the tail probability for  $\max_{t \in C} Z(t, s_0)$ , as one easily sees geometrically from the representation of that probability in terms of a tube volume. In principle, because of the two dimensional exponential family inbedding, the situation is different here; but in the range of our numerical examples the parameter  $\xi_2$  is close to zero, so numerically (4.18) is roughly equal to one-half of (3.15).

We have simulated 400 realizations of this field for  $0 \leq t \leq 50$ ,  $0 \leq s \leq 2$ , with  $\lambda = 10$ . The results, displayed in Figure 3, show in agreement with Figure 1 that the approximation developed here is more accurate than the Gaussian process approximation, which consistently gives values smaller than the simulated probability.

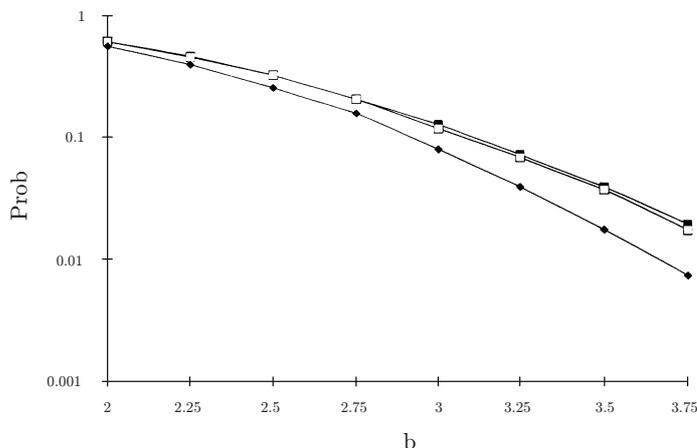


Figure 3. Comparison of approximations for the tail probability of the maximum of the random field (1.4) over  $0 < t < 50$ ,  $0 < s < 2$ : (i) Gaussian process approximation  $\blacklozenge$ , (ii) Monte Carlo simulation based on 400 repetitions  $\square$ , and (iii) approximation derived in Section 4  $\blacksquare$ .

## 5. Discussion

We have shown that it is possible at some increase in computing effort to improve substantially on the known Gaussian process approximations to the distribution of (1.2) when the underlying random field is a Poisson field. The same methods are applicable to other underlying point processes, provided the appropriate moment generating function can be evaluated. However, given the simplicity of the Gaussian process approximation, one may quite reasonably prefer to use it, provided it is not drastically misleading.

In the examples given above for the distribution of  $X_{\max}$ , the Gaussian approximation was about 1/3 to 1/2 the true probability when that probability was in the range 0.01 to 0.05. For larger values of  $\lambda$  the Gaussian process approximation will improve; for smaller values it will be worse, and there can be larger discrepancies between the basic approximation exemplified by (3.12) and the simplified approximations. For example, for the field of Figure 1, but with  $R = 10$ ,  $\lambda = 1$ ,  $b = 4.5$ , the approximation using (3.12) is 0.034, the first simplified approximation using (3.15) is 0.023, the second simplified approximation is 0.031. A Monte Carlo simulation with 400 repetitions gave 0.033, while the Gaussian approximation is a miserable 0.0018.

Another interesting case involves larger search regions, which can easily arise in higher dimensional problems, resulting in higher thresholds  $b$  to obtain smaller marginal tail probabilities, hence less accurate Gaussian approximations. Simulation over the relatively small regions used in our numerical examples is quite

time consuming, and could be almost prohibitive for larger regions. For the field of Figure 1 with  $R = 40$ , a value roughly appropriate for the astrophysical whole sky search (Giller, 1994), and with  $\lambda = 10$  (much smaller than for the whole sky search), for  $b = 4, 4.5$ , and 5 our approximations give 0.4, 0.1 and 0.02 compared to Gaussian process approximations of 0.2, 0.03 and 0.003, respectively. For the admittedly less interesting minimum, there are substantially greater discrepancies. Our approximations are 0.05, 0.003 and 0.0001, respectively.

### Acknowledgements

This research was supported in part by NSF Grant DMS-9403794.

### References

- Adler, R. J. (1981). *The Geometry of Random Fields*. John Wiley, New York.
- Aldous, D. (1988). *Probability Approximations via the Poisson Clumping Heuristic*, Springer-Verlag, New York.
- Belyaev, Yu. K. and Piterbarg, V. I. (1972). The asymptotic formula for the mean number of  $A$ -points of excursions of Gaussian fields above high levels (in Russian). In *Bursts of Random Fields*, 62-89. Moscow University Press, Moscow.
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications, Vol. II*, 2nd ed. John Wiley, New York.
- Giller, G. (1994). The construction and analysis of a whole-sky map using underground muons, Ph. D. Thesis, University of Oxford.
- Kim, H.-J. and Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika* **76**, 409-423.
- Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.
- Loader, C. R. (1991). Large-deviation approximations to the distribution of scan statistics, *Adv. in Appl. Probab.* **23**, 751-771.
- Rabinowitz, D. (1994). Detecting clusters in disease incidence. *Change-point Problems* (Edited by E. Carlstein, H.-G. Müller and D. Siegmund), 255-275. IMS, Hayward California.
- Rice, S. O. (1944,1945). Mathematical analysis of random noise, *Bell System Technical J.* **23**, **24**. Reprinted in Wax, N. (1954) *Selected Papers on Noise and Stochastic Processes*. Dover, New York.
- Siegmund, D. (1988). Approximate tail probabilities for the maxima of some random fields. *Ann. Probab.* **16**, 487-501.
- Siegmund, D. and Worsley, K. J. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Ann. Statist.* **23**, 608-639.
- Worsley, K. J., Evans, A. C., Marrett, S. and Neelin, P. (1992). A three dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism* **12**, 900-918.
- Worsley, K. J. (1995). Estimating the number of 'peaks' in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Ann. Statist.* **23**, 640-669.

Department of Statistics, Columbia University, New York, NY 10027, U.S.A.

Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.

(Received September 1995; accepted March 1996)