


Computing Scan Statistic p Values Using Importance Sampling, with Applications to Genetics and Medical Image Analysis
Author(s): Daniel Q. Naiman and Carey E. Priebe
Source: *Journal of Computational and Graphical Statistics*, Vol. 10, No. 2 (Jun., 2001), pp. 296-328
Published by: American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of America
Stable URL: http://www.jstor.org/stable/1391013
Accessed: 09/12/2010 11:56

Computing Scan Statistic p Values Using Importance Sampling, With Applications to Genetics and Medical Image Analysis

Daniel Q. NAIMAN and Carey E. PRIEBE

We present an importance sampling method for deciding, based on an observed random field, if a scan statistic provides significant evidence of increased activity in some localized region of time or space. Our method allows consideration of scan statistics based simultaneously on multiple scan geometries. Our approach yields an unbiased p value estimate whose variance is typically smaller than that of the naive hit-or-miss Monte Carlo technique when the p value is small. Furthermore, our p value estimate is often accurate for critical values that are not far enough in the tails of the null distribution to allow for accurate approximations via extreme value theory. The importance sampling approach unifies the analysis of various random field models, from (spatial) point processes to Gaussian random fields. For a scan statistic M, the method produces a p value of the form $P[M \geq \tau] = B\rho$, where B is the Bonferroni upper bound and the correction factor ρ measures the conservativeness of this upper bound. We present the application of our importance sampling estimator to multinomial sequences (molecular genetics), spatial point processes (digital mammography), and Gaussian random fields (PET scan brain imagery).

Key Words: Multiple testing; Random fields; Simultaneous inference; Spatial point processes.

1. INTRODUCTION

Consider a random field $Y = \{Y_x, x \in X\}$ observed at various locations $x \in X$, where X is an arbitrary domain. Our goal is to perform a test of homogeneity for the field versus the alternative that there is a localized subregion of X of nonhomogeneity. The location of the nonhomogeneity, which can be interpreted as a signal, is assumed to be unknown. In addition, its geometry (size and shape) may be taken to be either known or unknown. Examples include detecting a signal of unknown location and geometry in a Gaussian random field (GRF) and detecting a cluster of unknown location and geometry in a point process or a multinomial sequence.

Daniel Q. Naiman is Professor, and Carey E. Priebe is Associate Professor, Department of Mathematical Sciences, The Johns Hopkins University, Charles and 34th Streets, Baltimore, MD 21218 (E-mail: daniel.naiman@jhu.edu *and* cep@jhu.edu).

Journal of Computational and Graphical Statistics, Volume 10, Number 2, Pages 296–328

1.1 SCAN ANALYSIS

An intuitive approach to testing these hypotheses involves the partitioning of the region X into disjoint subregions. For cluster detection in spatial point processes this dates to the quadrat counts of Fisher, Thornton, and Mackenzie (1922); see Diggle (1983). Absent prior knowledge of the location and geometry of potential nonhomogeneities, this approach can have poor power characteristics.

A standard generalization of quadrats involves consideration of overlapping scan regions with fixed geometry. At each location x in the region X, introduce a *scan region* $R(x)$ about x, and define a scan process

$$\Psi = \{\Psi_x(Y), x \in X'\},$$

where $X' = \{x \in X \ : \ R(x) \subseteq X\}$, and where

$$\Psi_x(Y) = h(\{Y_x, x \in R(x)\}).$$

Here, h denotes a function of observations in a local neighborhood about x. Usually, h is a normalized count or average. We will refer to $\Psi_x(Y)$ as a *locality statistic.*

If the domain X forms a unit d-cube, then a standard choice for the scan regions are d-dimensional cubes

$$R(x) = \{y : x_j - w \leq y_j \leq x_j + w, \quad j = 1, \ldots, d\}.$$

where $w > 0$ is fixed.

For detecting clustering in point processes, when Y_x denotes a number points appearing at x, a common choice for the locality statistic is the number of events in the scan region

$$\Psi_x(Y) = \sum_{y \in R(x)} |Y_y|.$$

For detecting signals in Gaussian random fields the locality statistic will be a normalized average of the field Y in the scan region.

The *scan statistic*

$$M = \max_x \Psi_x(Y)$$

is defined as the maximum locality statistic over all scan regions.

Analysis of the univariate scan process ($d = 1$) has been considered by many authors, including Naus (1965), Cressie (1977, 1980), and Loader (1991). For a few simple random field models exact p values are available; many applications require approximations to the p value. The generalization to spatial scan statistics was considered by Naus (1965), Adler (1984), Loader (1991), and Chen and Glaz (1996). As noted by Cressie (1993), exact results for $d = 2$ have proved elusive; approximations to the p value based on extreme value theory are in general all that is available.

Although conventional scan analysis is superior to quadrat analysis, as the former circumvents the difficulty associated with unknown location, prior knowledge of the geometry of the nonhomogeneity is necessary. For applications in which such prior knowledge is unavailable, a satisfactory scan approach must therefore involve variable geometry. Here, the

scan regions are allowed to vary in shape or size, so for each location we have a collection of scan regions

$$R(x, w), x \in W(x),$$

where $W(x)$ is a set of scan window geometries (possibly depending on x) so that the locality statistics are given by

$$\Psi_{x,w}(Y) = h(\{Y_x, x \in R(x, w)\}),$$

for $x \in X$ and $w \in W(x)$, and the scan process

$$\Psi = \{\Psi_{x,w}(Y), x \in X, w \in W(x)\},$$

is indexed by location and geometry. The problem of choosing an optimal range of window geometries is driven by application considerations and is beyond the scope of this article. Consideration of the problem when the geometry of the anomaly is unknown has been previously considered in Loader (1991), Kulldorf (1997), Alm (1997), and Priebe, Olson, and Healy (1997).

1.2 IMPORTANCE SAMPLING

Our approach involves simple importance sampling techniques for estimating p values for hypothesis tests based on scan statistics. Expositions of importance sampling can be found in Fishman (1996, sec. 4.1) and Ross (1990, sec. 8.5). The procedure we employ for approximating the $p-$value is based on improving the naive hit-or-miss Monte Carlo simulation. In the naive approach, random fields are generated randomly according to an appropriate null distribution, and the rejection frequency is calculated. A single data realization may be quite extensive, and subsequent calculation of the test statistic can also involve extensive computation. Consequently, the number of Monte Carlo replications must be small and thus the variance of the Monte Carlo estimate of the p value can be too large to be practical when the true p value is small.

Our method can be viewed as providing a correction to the bound given by the Bonferroni (1936 a,b) method (see also Worsley 1982, 1985; Naiman and Wynn 1992). Our approach builds on the union counting procedure introduced by Frigessi and Versillis (1984); see also the "harmonic mean formula" appearing in Aldous (1989, p. 8). For an extensive discussion of how importance sampling leads to improved estimates in this context see Fishman (1996, sec. 4.1). Naiman and Wynn (1997) described a more general class of importance sampling algorithms based on sharpened higher-depth inclusion-exclusion inequalities.

Our method yields an unbiased p value estimate whose variance is typically smaller than that of the naive hit-or-miss Monte Carlo technique when the p value is small. Furthermore, our p value is often accurate for critical values that are not far enough in the tails of the null distribution to allow for accurate approximations via extreme value theory.

1.3 OUTLINE

Section 2 presents the main result, an importance sampling algorithm for estimating the p values for hypothesis tests based on scan statistics. Sections 3, 4, and 5 investigate the

applicability of the methodology for various random field scenarios and present specific examples indicating the utility of the approach. For each application, the general methodology of Section 2 is appropriately adapted and extended. Section 3 considers multinomial sequences and presents the example of detecting subsequences representing distinctive charge configurations in DNA or protein sequences. Section 4 considers inference for spatial point processes and presents the example of detecting clusters of microcalcifications in digital mammography. Section 5 considers detecting signals of unknown geometry and location in Gaussian random fields and presents the example of performing inference on the existence of regionally specific effects in PET scan brain imagery.

2. MAIN RESULT

The basic idea behind the method is described as follows. Let $Y = \{Y_x, x \in X\}$ be a random field whose null distribution is known. This field is assumed to come from observations collected in space or time, or some combination of both. Assume we test a hypothesis by determining if the scan statistic

$$M = \max_{1 \le i \le N} \Psi_i(Y)$$

exceeds some threshold, where Ψ_i are given real-valued functions indexed by i. As in Section 1.1 we refer to each $\Psi_i(Y)$ as a *locality statistic.* Here N is the number of locality statistics we consider in forming a scan statistic. Generalizations to cases of continuously indexed locality statistics are straightforward and are presented in the examples which follow.

As described in Section 1, in the applications we have in mind the problem is to decide if there is significant evidence of increased activity near some location or site in time or space. An appropriate null hypothesis might say that the observations were produced by the same mechanism that produces noise, or that outcomes occur according to a prespecified rate that applies over an entire region. The individual statistics $\Psi_i(Y)$ might be used to measure the average of a random field or the number of occurences of a spatial point process, in some local region; that is, near an unspecified site. Since this location is a priori unknown, varying the index i results in varying the measurement over a finite number of possible locations.

In addition, there may be little a priori information as to the size and shape of the neighborhoods in which to *search* for increased activity. We use the generic term *scan geometry* to refer to properties of the statistics Ψ_i that have to do with sizes or shapes of neighborhoods of points in which activity is measured. Thus, the indices i over which we define the Ψ_i serve a dual purpose in that they can indicate location as well as scan geometry.

Our goal is to evaluate

$$P_{H_0}[M \ge \tau], \tag{2.1}$$

for given τ (the observed value of M). Since all of the probabilities appearing below are evaluated under the null hypothesis, we will dispense with the subscript H_0 and denote all probabilities using simply P.

For the problems described below, the *Bonferroni* upper bound

$$B = \sum_{i=1}^{N} P[\Psi_i(Y) \geq \tau] \tag{2.2}$$

is easy to evaluate, but can be too conservative to be practical. A correction to this bound is obtained by writing the p value as follows. The probability (2.1) can be expressed as

$$\begin{aligned} P\left[\bigcup_{i=1}^{N}\{\Psi_i(Y) \geq \tau\}\right] &= \int I_{\bigcup_{i=1}^{N}\{\Psi_i(Y)\geq\tau\}}\, dP \\ &= \int \frac{I_{\bigcup_{i=1}^{N}\{\Psi_i(Y)\geq\tau\}}}{\sum_{j=1}^{N} I_{\{\Psi_j(Y)\geq\tau\}}} \sum_{i=1}^{N} I_{\{\Psi_i(Y)\geq\tau\}}\, dP \\ &= B\sum_{i=1}^{N} q_i \int \frac{1}{g(Y)} \frac{I_{\{\Psi_i(Y)\geq\tau\}}}{P[\Psi_i(Y) \geq \tau]}\, dP, \end{aligned}$$

where $g(Y) = \sum_{i=1}^{N} I_{\{\Psi_i(Y)\geq\tau\}}$, the number of indices i for which $\Psi_i(Y)$ exceeds τ, and

$$q_i = \frac{P[\Psi_i(Y) \geq \tau]}{\sum_{j=1}^{N} P[\Psi_j(Y) \geq \tau]},$$

so that the q_i define a probability distribution on $\{1, 2, \ldots, N\}$.

We conclude that the p value takes the form

$$P[M \geq \tau] = B\rho, \tag{2.3}$$

where

$$\rho = \sum_{i=1}^{N} q_i \int \frac{1}{g(Y)} \frac{I_{\{\Psi_i(Y)\geq\tau\}}}{P[\Psi_i(Y) \geq \tau]}\, dP, \tag{2.4}$$

and the correction factor ρ can be interpreted as the expected value of the random variable $\hat{\rho}_p$ generated in each iteration p of the following Monte Carlo experiment:

Importance Sampling Algorithm

For k from 1 to n do (independently)

Step 1. Generate a random index $J_k \in \{1, \ldots, N\}$ according to the probabilities q_i.
Step 2. Generate $\tilde{Y}_k$ from the conditional distribution of Y given $\Psi_{J_k}(Y) \geq \tau$.
Step 3. Count $g_k = g(\tilde{Y}_k)$ the number of locality statistics Ψ_i for which $\Psi_i(\tilde{Y}_k) \geq \tau$, and take $\hat{\rho}_k = 1/g_k$.

End do
Return $\hat{\rho} = \frac{1}{n}\sum_{k=1}^{n} \hat{\rho}_k$.

Of course, in order for this algorithm to be made practical, it is necessary to be able to efficiently generate J according to the given marginal distribution and $\tilde{Y}$ according to the required conditional distribution. For the three random field scenarios presented in Sections 3, 4, and 5 this is indeed practical.

Since the $\tilde{Y}$ is defined conditionally on $\Psi_J(\tilde{Y}) \geq \tau$, the number of locality statistics for which $\Psi_i(\tilde{Y}) \geq \tau$, is guaranteed to be at least one, so we always have $g(\tilde{Y}) \geq 1$. This leads to the conclusion that $\rho \leq 1$, which is to be expected because the Bonferroni procedure is known to be conservative. Interestingly, ρ measures the conservativeness of the Bonferroni bound and the source of this conservativeness becomes clear. If, conditionally given that a locality statistic yields an extreme value, none of the other locality statistics are likely to, then ρ is close to 1 and the Bonferroni bound is sharp. On the other hand, if there is a tendency for many locality statistics to yield extreme values conditionally given that locality statistic Ψ_J does, then this will be reflected in a larger value of g on average, and hence a smaller value of ρ, and we can improve on the Bonferroni bound. In the extreme case when, having conditioned on one of them being extreme, every locality statistic is extreme, we find that ρ reduces to the factor $1/N$, so that if all of the probabilities $P[\Psi_i(Y) \geq \tau]$ are the same, the corrected p value B/N becomes this common value.

2.1 ALGORITHMS FOR p VALUE BOUNDS

Since the indices i correspond to locations as well as scan geometries, the number of indices N can make the last step of the importance sampling algorithm (calculation of the number of locality statistics that exceed τ) unwieldy. For example, in a two-dimensional image that is 500×500 pixels, with 10 choices of scan geometries, N would be 2.5 million. Instead of estimating the exact p value, a less ambitious and more computationally feasible approach is to try to estimate a conservative (i.e., upper) bound for the p value which is tighter than B.

A general approach to achieving this goal is to use the same sampling procedure (Steps 1 and 2) in the foregoing importance sampling algorithm, but replace evaluation of g_k by an upper bound on g_k. Since g_k counts the number of threshold exceedances $\Psi_i(\tilde{Y}_k) \geq \tau$ by the locality statistics, we obtain an upper bound on $1/g_k$ by restricting this count to some subset of the set of all locality statistics. Since we are conditioning on threshold exceedance for a particular locality statistic Ψ_{J_k}, it seems natural to restrict our count only to those locality statistics Ψ_i that are *near* Ψ_{J_k}. Since for small p values the probability of two isolated exceedances is small, this can lead to approximations to sharp upper bounds for p values whose computation is far more tractable than the exact p value approximations.

3. DETECTING DISTINCTIVE FEATURES IN DNA OR PROTEIN SEQUENCES

In many molecular biology applications, the random field of interest forms a sequence of letters $Y = \{Y_i, i = 1, \ldots, L\}$ from an alphabet $\{a_1, \ldots, a_r\}$ of known size r. This sequence is studied and apparently distinctive features are uncovered. A problem is to decide whether the features discovered are due to chance, so it is critical to determine how

frequently an observed feature arises when the sequence is generated using a given random mechanism. Since many of the applications we envision for the methodology above come from molecular genetics, we give a brief description of some key concepts from genetics.

3.1 A BRIEF OVERVIEW OF MOLECULAR GENETICS

Several examples of alphabets arise in the study of DNA and protein sequences. We present a brief overview of some of the more basic ideas and terminology needed from molecular genetics (see Watson et al. 1987; Karlin and Altschul 1990). DNA forms a double-stranded molecule consisting of a sequence of *base pairs of nucleotides.* The nucleotides in DNA are the purines—adenine and guanine—and the pyrimidines—cytosine and thymine. Thus, we can view any portion of DNA as a sequence of letters from a *nucleotide alphabet* of size $r = 2$ or 4. RNA has a similar description except that the pyramidine thymine is replaced by the pyramidine uracil. Messenger RNA (mRNA) serves as a template in the synthesis of *protein.* The bases in mRNA form 3-tuples of bases called *codons.* There are 61 possible codons that can arise (out of the 64 possible 3 letter sequences of 4 bases). Therefore, portions of mRNA can be viewed as formed using a *codon alphabet* of 61 letters. Finally, mRNA forms a template in the synthesis of proteins.

A *protein* is polymer composed of a large number of *amino acids,* linked together by *peptide bonds.* Such chains are referred to as *polypeptides.* There are 20 different amino acids that can be found in proteins, so proteins could be described using an *amino acid alphabet* of size 20. Of the 20 amino acids found in proteins two of these, aspartic acid and glutamic acid, are acidic and carry a net charge of -1, three of them—lysine, arginine, and histidine—are basic and carry a net charge of $+1$. The remaining 15 are neutral and carry a net charge of 0. This leads to a *charge alphabet* $\{-1, 0, +1\}$ of size 3.

3.2 SCORING METHODS FOR DISTINCTIVE SEGMENTS

To identify distinctive segments in a sequence, one general approach (Karlin and Altschul 1990) is to associate a *score* $s(a_j)$, $j = 1, \ldots, r$ with each letter in the alphabet, and define the score associated with the segment at location i of width $2w + 1$ to be a normalized sum of scores

$$\Psi_{i,2w+1}(Y) = \frac{1}{c(w)} \sum_{j=i-w}^{i+w} s(Y_j), \quad i = w+1, \ldots, L-w,$$

where $c(w)$ denotes any normalization constant depending on the window width w. The sequence is then searched at each location using different window sizes and a *maximum scoring segment* is found. This segment is declared to be significant if the maximum score

$$M = \max_{i,w} \Psi_{i,2w+1}(Y)$$

exceeds some threshold.

There are various possible criteria for choosing the normalization constants $c(w)$. One idea (see Section 4.1) is to choose $c(w)$ so that the individual false detection probabilities,

the ones appearing in the Bonferroni sum (2.2), are roughly the same. The case of $c(w) = 1$ is used in the example of Section 3.3.

For the case of a fixed window size $2w + 1$, Karlin, Dembo, and Kawabata (1990) gave an asymptotic approximation to the distribution for M as the length of the sequence L tends to infinity, for the case when the $\{Y_i, i = 1, \ldots, L\}$ forms a sequence of iid random variables. Karlin and Dembo (1992) gave the limit distribution in the more general setting in which the Y_i are generated from a Markov chain model and in addition, the window size $2w + 1$ is allowed to vary.

The importance sampling algorithm described in Section 2 leads to an approximation for $P[M \geq \tau]$ when $\{Y_i,\ i = 1, \ldots, L\}$ are iid according to some probability distribution on the alphabet $\{a_1, \ldots, a_r\}$. The important extension of the importance sampling method to situations in which the Y_i are dependent, for example, the Markov chain case, is the focus of current research.

Steps 1 and 2 are carried out as follows. First, we introduce some notation. Let $U_{t,2w+1}$ denote the set of letter sequences $\{y_i, i = 1, \ldots, 2w + 1\}$ of length $2w + 1$ for which

$$\frac{1}{c(w)} \sum_{j=1}^{2w+1} s(y_j) = t,$$

for each window size $2w+1$, and for each $t \geq \tau$. Let $Y_{1:2w+1}$ denote $\{Y_i, i = 1, \ldots, 2w+1\}$.

To carry out Steps 1 and 2, we do the following:

Step i. Generate an exceedance value $T \geq \tau$ according to the distribution p_T where

$$p_T(t) = \frac{\sum_w (L - 2w) \sum_{y \in U_{t,2w+1}} P[Y_{1:2w+1} = y]}{\sum_{t \geq \tau} \sum_w (L - 2w) \sum_{y \in U_{t,2w+1}} P[Y_{1:2w+1} = y]},$$

for $t \geq \tau$, where the probabilities are calculated from the alphabet distribution.

Step ii. Conditionally, given $T = t$ generate W according to the distribution

$$p_{W|T}(w|t) = \frac{(L - 2w) \sum_{y \in U_{t,2w+1}} P[Y_{1:2w+1} = y]}{p_T(t)}.$$

Step iii. Conditionally, given T and W generate an index I in $\{W + 1, \ldots, L - W\}$ uniformly.

Step iv. Conditionally, given T, W, and I generate $(\tilde{Y}_{I-W}, \ldots, \tilde{Y}_{I+W})$ from $U_{T,2W+1}$ uniformly, and take the remaining $Y_i, i \notin \{I - W, \ldots, I + W\}$ to be iid and distributed according to the alphabet distribution.

In general, the start-up costs associated with these steps can be substantial. On the other hand, this algorithm is quite feasible when applied to the problem described in the next section. Feasibility of the algorithm under more complex distributional assumptions—for example, Markov assumptions—requires further study. This will be the subject of future investigation.

3.3 EXAMPLE: TESTS FOR DISTINCTIVE CHARGE CLUSTERS IN PROTEINS

To illustrate the method in some real examples, we consider the sequences of charges in proteins. Here, each amino acid Y_i in the protein is given a score of $-1, 0$ or $+1$, to represent its associated charge $s(Y_i)$. Following Karlin, Blaisdell, Mocarski, and Brendel (1989) we define the net charge in a window of width $2w + 1$ at a location i to

$$\Psi_{i,w}(Y) = \sum_{j=i-w}^{i+w} s(Y_j).$$

We then search the sequence using varying window sizes w [Karlin et al. (1989) suggested it is appropriate to use $30 \leq w \leq 60$] at various locations and charge clusters are determined to be significant if the maximum (or minimum) net charge

$$M = \max_{30 \leq w \leq 60} \max_{w+1 \leq i \leq L-w} \Psi_{i,w}(Y)$$

exceeds some threshold.

Karlin et al. (1989) provided an approximation to the p value of this test, treating w as fixed. Karwe and Naus (1997) gave a formula for the exact p value based on a fixed window size, assuming the charges $s(Y_i)$ form an iid sequence with frequencies f_0, f_- and f_+ given by the empirical charge frequencies for the entire sequence. Karlin et al. (1989) identified significant charge configurations for a collection of 20 sequences taken from Epstein–Barr virus polypeptides, and for each of these they gave an approximate p value. Karwe and Naus (1997) provided a table of exact p values for these 20 examples and compared these to the approximations.

We are motivated by two goals. One is the desire to see, for this same set of examples, how close the importance sampling approximation is to the exact p value, when a fixed window size is assumed. Second, we wish to determine the degree to which the p value is affected by accounting for variability in the window width. The results are given in Table 1. For the case when a fixed window size is assumed, we have found that for 19 of the 20 examples, when we used a Monte Carlo sample size of 1,000 the relative error for the importance sampling p value is always less than 0.1. Observe that for one entry, BRRF2, there is not close agreement between our approximate p value and the p value reported by Karwe and Naus (1997). We believe that their reported value is in error in this case.

Running on a 166 Mhz Pentium PC using the Windows 95 operating system, using a C program written by one of authors and compiled using the Watcom C Compiler version 10.6, the computation time for each of the 20 p value approximations was about 1 second

To study the effect of using multiple window sizes, we approximate the p value for each of the 20 significant clusters as if the search for significant clusters had been restricted to all windows of size $w - u$ through $w + u$, where w denotes the size of cluster found, and u denotes a nonnegative integer. Thus, $2u + 1$ gives the number of different sizes for windows scanned. In particular, the case of $u = 0$ corresponds to treating the window size as fixed at w. We calculated the approximate p value for u ranging from 0 through 10. In every case, the plot of the approximate $\log_{10}(p \text{ value})$ appears close to linear in u, especially for small values of u. In Table 1 we report the slope of the simple linear least squares regression line fitted to these data. This slope could be used to determine the first-order effect of not

Table 1. Approximate p Values for Charge Configurations in Epstein–Barr Virus Polypeptides

Random field				Charge configuration		Fixed w			Variable w regression	
						Exact	Approx	rel		
ORF	L	f_+ (%)	f_- (%)	τ	w	p value	p value	error	$\hat{\beta}$	MAD
BYRF1	512	7.4	10.2	13	32	2.6×10^{-4}	2.54×10^{-4}	−0.02	0.11	0.03
BYRF1	512	7.4	10.2	13	35	1.1×10^{-5}	1.03×10^{-5}	−0.06	0.098	0.02
BPLF11	3149	11.1	11.1	16	35	3.5×10^{-5}	3.46×10^{-5}	−0.01	0.13	0.03
BMLF1	459	11.8	16.3	16	38	9.8×10^{-6}	1.01×10^{-5}	0.03	0.10	0.02
BLLF1	907	6.7	6.6	8	33	4.6×10^{-2}	4.67×10^{-2}	0.02	0.048	0.02
BERF1	839	10.5	11.0	12	30	9.2×10^{-4}	9.30×10^{-4}	0.01	0.095	0.02
BERF1	839	10.5	11.0	11	30	7.1×10^{-3}	7.45×10^{-3}	0.05	0.081	0.02
BERF2B	840	10.2	12.6	15	39	8.8×10^{-4}	8.46×10^{-4}	−0.04	0.092	0.01
BERF2B	840	10.2	12.6	16	30	2.7×10^{-7}	2.71×10^{-7}	0.002	0.17	0.04
BERF4	872	9.6	12.6	12	36	1.1×10^{-3}	1.15×10^{-3}	0.04	0.075	0.02
BZLF1	200	6.5	8.0	12	43	5.5×10^{-4}	5.56×10^{-4}	0.01	0.068	0.02
BRRF2	537	12.1	10.4	21	55	1.8×10^{-7}	3.55×10^{-6}	18.72	0.29	0.353
BKRF1	641	7.5	10.9	13	38	3.0×10^{-3}	2.93×10^{-3}	−0.02	0.084	0.02
BKRF1	641	7.5	10.9	18	54	1.1×10^{-4}	1.01×10^{-4}	−0.09	0.085	0.01
BKRF1	641	7.5	10.9	10	34	1.9×10^{-3}	1.94×10^{-3}	0.02	0.063	0.02
BKRF1	641	7.5	10.9	17	41	3.1×10^{-8}	3.17×10^{-8}	0.02	0.13	0.02
BYRF4	226	17.3	12.4	30	58	4.1×10^{-9}	3.98×10^{-9}	−0.03	0.16	0.02
BBRF3	405	5.7	9.4	11	31	2.4×10^{-3}	2.34×10^{-3}	−0.03	0.088	0.02
BXLF1	607	11.4	15.2	13	40	1.2×10^{-3}	1.19×10^{-3}	−0.009	0.064	0.02
BNLF1	386	13.2	7.5	16	45	1.3×10^{-3}	1.41×10^{-3}	0.08	0.084	0.02

NOTE: The first four columns describe properties of a particular portion of a protein sequence from the Epstein–Barr polypeptide. Column 1 gives an identifier called the *open reading frame*, column 2 gives the number of amino acids in the sequence, and columns 3 and 4 provide the frequencies of positive and negative charged proteins in the sequence. These columns appear in Table 1 of Karwe and Naus (1997) and come from Table 1 of Karlin et al. (1989) where a complete description is provided. Columns 5 and 6 are determined from these prior tables and describe properties of a particular cluster of significant net charge. Column 5 gives the net charge τ for the cluster and column 6 gives the number of amino acids in the cluster. Column 7 gives the exact p value from Karwe and Naus (1997), column 8 gives our p value approximation and column 9 gives its relative error. Columns 10 and 11 give the results of fitting a regression line to the approximate $\log_{10}(p \text{ value})$ versus maximum window size w data; column 10 gives the slope of this line and column 11 gives the mean absolute deviation for the fit (in $\log_{10}$ scale).

knowing which window size to use and specifying a range, as opposed to being fortunate enough to have made an optimal a priori guess of the window size to use.

We summarize the results in Table 1. For illustrative purposes, one of these plots is provided for the first entry in Table 1 (BYRF1) in Figure 1.

We illustrate how the regression fit can be interpreted for this case. A similar discussion can be given for each of the other 19 significant charge clusters found. In the case of BYRF1, the least squares fitting line, which fits the data well, has a slope of $.11 = \log_{10}(1.28)$. Since the exact p value is given by 2.6×10^{-4} we conclude that the p value for the multiple window case is approximately

$$2.6 \times 10^{-4} \times 1.28^{u}$$

for small values of u (in the range $0 \leq u \leq 10$.) The nonlinearity of the plot in Figure 1 suggests that the simplistic approach of using a linear fit has definite limitations.

The seventh entry of Table 1, the second entry for which the ORF is BERF1 provides an example of how accounting for the effect of multiple window sizes results in changing a

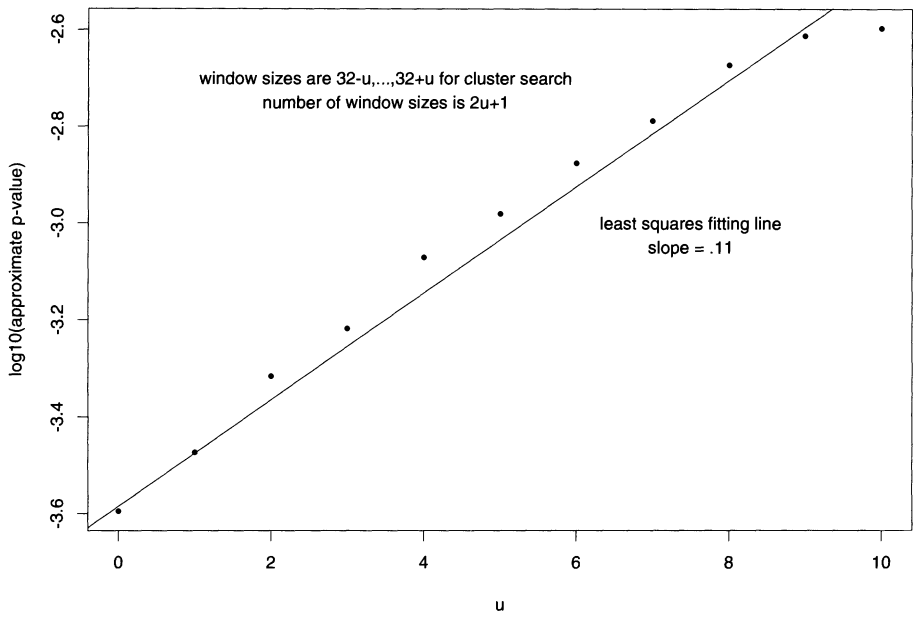


Figure 1. P Value Versus Number of Window Sizes. Open reading frame = BYRF1.

significant p value (0.007) to an insignificant one. If we assume that window sizes from 15 - 45 (i.e., $u = 15$) are searched for significant charge clusters, then importance sampling gives an estimated p value of 0.071.

4. SPATIAL POINT PROCESSES

In this Section we consider point processes in the unit square $[0, 1]^2$. For concreteness, the development in Sections 4.1–4.5 focusses on *Poisson* point processes, but the methods described here are easily modified to handle other index sets and probability models. In particular, we consider a binomial process in a grid in Section 4.6.2.

Let us denote the observed process by $Y = \{Y_{u,v}, (u, v) \in [0, 1]^2\}$, where $Y_{u,v}$ takes the value 0 except at finitely many locations (u, v) where the value is 1. Let the set of these locations be denoted by $\{(U_1, V_1), \ldots, (U_N, V_N)\}$. For a null hypothesis, we assume that Y is a homogeneous Poisson point process, so that $N \sim \mathcal{P}(\lambda)$ and conditionally given N the (U_i, V_i) are iid and uniformly distributed in $[0, 1]^2$.

4.1 SCAN STATISTICS BASED ON A FINITE NUMBER OF WINDOW SIZES

There are a number of possible procedures for testing the null hypothesis against the hypothesis of a particular type of *nonhomogeneity* or *clustering* characterized by a single small subregion of increased activity. In this Section, we consider square windows of width w centered at (u, v) :

$$A_{w,u,v} = \{(u', v') \; : \; |u' - u| \leq w/2, |v' - v| \leq w/2\}.$$

We require that a window be contained in the unit square which amounts to the constraint that u and v lie in the interval $I_w = [\frac{1}{2}w, 1 - \frac{1}{2}w]$.

In order to detect clustering of the point process Y, we fix a finite set of window sizes $\{w_1, \ldots, w_m\}$ and compute the number of occurences in each window

$$N(A_{w,u,v}) = \sum_{(u',v') \in A_{w,u,v}} Y_{u',v'}.$$

(The case when we restrict the window size to lie in an interval reduces to this one, as shown in Section 4.5.) We then reject the null hypothesis if there are an extreme number of occurences in some window. Thus, we compute an *individual p* value based on the Poisson (λw^2) distribution function

$$1 - F_{\mathcal{P}(\lambda w^2)}(N(A_{w,u,v}) - 1),$$

for each scan window. This is the probability that the Poisson random variable for the particular window is at least as large as the actual number of occurences in the window. We then reject H_0 if one of these p values is sufficiently small; in other words, if the quantity

$$\Gamma = \min_{w \in \{w_1, \ldots, w_m\}} \min_{u,v \in I_w} 1 - F_{\mathcal{P}(\lambda w^2)}(N(A_{w,u,v}) - 1)$$

is sufficiently small, say $\Gamma \leq C$.

To relate this to the framework introduced in Section 2, we define the locality statistic

$$\Psi_{w,u,v}(Y) = N(A_{w,u,v}) - c_w,$$

where

$$c_w = F^{-1}_{\mathcal{P}(\lambda w^2)}(1 - C) + 1.$$

Here we use the standard definition of the inverse distribution function found in Serfling (1980) for example. Using an elementary property of the inverse distribution (Serfling 1980, Lemma iii, p. 3), it follows that $1 - F_{\mathcal{P}(\lambda w^2)}(N(A_{w,u,v}) - 1) \leq C$ if and only if $N(A_{w,u,v}) \geq c_w$, so the test is then equivalent to the one that rejects H_0 if $M \geq 0$ where M denotes the scan statistic

$$M = \sup_{w \in \{w_1, \ldots, w_m\}} \sup_{u,v \in I_w} \Psi_{w,u,v}(Y).$$

4.2 EXCEEDANCE PROBABILITY VIA IMPORTANCE SAMPLING

The situation just described does not fit exactly into the framework of Section 2 because here the scan statistic involves the supremum of a continuum of locality statistics. In addition, the interpretation of the importance sampling approach as providing a correction to the Bonferroni bound no longer makes sense because the Bonferroni bound is $+\infty$. However, it is possible to derive a version of the importance sampling algorithm in this case. The essential idea is to replace the index set for the windows, which is a square, by a lattice which is finite but arbitrarily. Since the exceedance probability in the continuum case can

be approximated by the exceedance probability in the finite lattice case, and expressions (2.3) and (2.4) hold for lattice case, we can use a simple limiting argument making use of Fubini's theorem (see Rudin 1974) to give analogous expressions to (2.3) and (2.4) for the continuum case. Letting

$$p_w = P[\Psi_{w,u,v} \geq 0], \text{ for } u, v \in I_w,$$

we define

$$B^* = \sum_{w \in \{w_1, \ldots, w_m\}} (1-w)^2 p_w,$$

$$q_w = p_w (1-w)^2 / B^*,$$

$$g^*(Y) = \sum_{w \in \{w_1, \ldots, w_m\}} \text{Area}\left(\{(u,v) \in I_w \times I_w \ : \ \Psi_{w,u,v}(Y) \geq 0\}\right),$$

and

$$\eta(w,u,v) = \int \frac{1}{g^*(Y)} \frac{I_{\{\Psi_{w,u,v}(Y) \geq 0\}}}{P[\Psi_{w,u,v}(Y) \geq 0]} dP.$$

It then follows that

$$P[M \geq 0] = B^* \rho^*, \tag{4.1}$$

where

$$\rho^* = \sum_{w \in \{w_1, \ldots, w_m\}} q_w \int_{u=w/2}^{1-w/2} \int_{v=w/2}^{1-w/2} \eta(w,u,v) dv du / (1-w)^2.$$

This expression translates into the following algorithm.

For k from 1 to n do (independently)

Step 1. Generate a random window size W according to the probabilities q_w.
Step 2. Generate a random location (U, V) uniform in the set $I_W \times I_W$.
Step 3. Generate $\tilde{Y}_k$ from the conditional distribution of Y given $\Psi_{W,U,V}(Y) \geq 0$.
Step 4. Compute the quantity $g^*(Y)$ and take $\hat{\rho}_k = 1/g^*(Y)$.

End do
Return $B^* \hat{\rho}$ where $\hat{\rho} = \frac{1}{n} \sum_{k=1}^{n} \hat{\rho}_k$.

Step 3 is trivial to carry out. We first generate $\Psi_{W,U,V}(Y)$, conditioned on $\Psi_{W,U,V}(Y) \geq 0$, which amounts to generating a random variable $N' = N(A_{W,U,V}) \sim \mathcal{P}(\lambda W^2)$ conditioned on $N(A_{W,U,V}) \geq c_W$. Then we generate $(U_i', V_i'), i = 1, \ldots, N'$ iid uniform in

$A_{W,U,V}$. Finally, we generate $N'' \sim \mathcal{P}(\lambda * (1 - W^2))$ and $(U_i'', V_i''), i = 1, \ldots, N''$ iid uniform in $[0,1]^2 \backslash A_{W,U,V}$. Then we define Y by taking $N = N' + N''$ and

$$\{(U_i, V_i), i = 1, \ldots, N\} = \{(U_i', V_i'), i = 1, \ldots, N'\} \cup \{(U_i'', V_i''), i = 1, \ldots, N''\}.$$

s There are only finitely many Poisson tail distributions involved so instead of generating from a given one on the fly, it is possible to store what is needed to generate from them in data structures during the setup step.

4.3 CALCULATING THE AREA OF THE THRESHOLD EXCEEDANCE SET

The computation in Step 4 requires the area of the set

$$\{(u, v) \ : \ N(A_{w,u,v}) \geq c_w\} \tag{4.2}$$

for each $w \in \{w_1, \ldots, w_m\}$. To compute this area for fixed window size w, order the distinct points $U_i \pm \frac{1}{2}w$, $i = 1, \ldots, N$ which lie in the interval I_w, and let the ordered sequence of points U_i obtained in Step 3 of the current iteration be denoted by $U_{(i)}$ $i = 1, \ldots, N_u - 1$. Also define $U_{(0)} = \frac{1}{2}w$ and $U_{(N_u)} = 1 - \frac{1}{2}w$, so that

$$\frac{1}{2}w = U_{(0)} < U_{(1)} < \cdots < U_{(N_u)} = 1 - \frac{1}{2}w.$$

Do the same for the $V_i \pm \frac{1}{2}w$, $i = 1, \ldots, N$ to give an ordered sequence $V_{(i)}, i = 1, \ldots, N_v$ with

$$\frac{1}{2}w = V_{(0)} < V_{(1)} < \cdots < V_{(N_v)} = 1 - \frac{1}{2}w.$$

Observe that the set

$$\{i \in \{1, \ldots, N\} \ : \ (U_i, V_i) \in A_{w,u,v}\},$$

and thus $N(A_{w,u,v})$, is unchanged as we vary the center (u, v) over the (open) rectangle

$$(U_{(i)}, U_{(i+1)}) \times (V_{(j)}, V_{(j+1)}).$$

Consequently, the area in (4.2) is given by

$$\sum_{(i,j) \in \mathcal{F}_w} (U_{(i+1)} - U_{(i)})(V_{(j+1)} - V_{(j)}), \tag{4.3}$$

where

$$\mathcal{F}_w = \left\{(i, j) \in \{1, \ldots, N_u\} \times \{1, \ldots, N_v\} \ : \ N(A_{w,\tilde{U}_i,\tilde{V}_j}) \geq c_w\right\}, \tag{4.4}$$

$$\tilde{U}_i = \frac{1}{2}(U_{(i)} + U_{(i+1)}),$$

and

$$\tilde{V}_j = \frac{1}{2}(V_{(j)} + V_{(j+1)}).$$

Note that

$$N(A_{w,\tilde{U}_i,\tilde{V}_j}) = \left|\left\{r \,:\, \tilde{U}_i - w \le U_r \le \tilde{U}_i + w, \quad \text{and} \quad \tilde{V}_j - w \le V_r \le \tilde{V}_j + w\right\}\right|.$$

In our implementation of the importance sampling algorithm, instead of checking every possible pair of indices (i, j) for inclusion in $\mathcal{F}_w$ we proceed as follows. For each index i we compute the set

$$\mathcal{F}_w^{u,i} = \left\{r \,:\, \tilde{U}_i - w \le U_r \le \tilde{U}_i + w\right\}$$

and we store in a list all of the $\mathcal{F}_w^{u,i}$ for which $|\mathcal{F}_w^{u,i}| \ge c_w$. Similarly, for each index j we compute

$$\mathcal{F}_w^{v,j} = \left\{r \,:\, \tilde{U}_j - w \le U_r \le \tilde{U}_j + w\right\}$$

and store in a list all of the $\mathcal{F}_w^{v,j}$ for which $|\mathcal{F}_w^{v,j}| \ge c_w$. Finally, the sum (4.3) is calculated using the fact that

$$\mathcal{F}_w = \left\{(i,j) \,:\, |\mathcal{F}_w^{u,i} \cap \mathcal{F}_w^{v,j}| \ge c_w\right\}$$

and in the right side only the stored $\mathcal{F}_w^{u,i}$ and $\mathcal{F}_w^{v,j}$ need be considered.

4.4 Bounds for the Exceedance Probability

As described in Section 2.1, it is possible to obtain an upper bound for $P[M \ge \tau]$ by replacing the function g^* in (4.1) and in Step 4 by a lower bound whose calculation requires considerably less computational effort. We have implemented the following simple idea for bounding $g^*(Y)$ below.

Choose any $F > 0$. Assume W, U, and V are obtained from Steps 1 and 2 of the algorithm. In the calculation of $g^*(Y)$, instead of using the points $U_i \pm \frac{1}{2}w$, $i = 1, \ldots, N$ which lie in the interval I_w, to define the $U_{(i)}$ we use only the ordered $U_i \pm \frac{1}{2}w$, $i = 1, \ldots, N$ which lie in the interval $U \pm \frac{1}{2}FW$, and we take $U_{(0)} = U - \frac{1}{2}FW$ and $U_{(N_u)} = U + \frac{1}{2}FW$. We define the $V_{(j)}$ analogously.

4.5 Scan Statistics Based on an Interval of Window Sizes

If we restrict the window size w to lie in an interval, say $w \in [w_L, w_H]$, then we show in this section that the resulting scan statistic reduces to one in which there are finitely many window sizes and so this case is handled by the methods described above. As above, we compute an individual p value for each choice of $w \in [w_L, w_H]$ and for each $u, v \in I_w$, and we reject H_0 for sufficiently small values of

$$\Gamma = \inf_{w \in [w_L, w_H]} \inf_{u,v \in I_w} 1 - F_{\mathcal{P}(\lambda w^2)}(N(A_{w,u,v})),$$

say $\Gamma \le C$. In other words, based on the locality statistic

$$\Psi_{w,u,v}(Y) = N(A_{w,u,v}) - c_w,$$

where

$$c_w = F^{-1}_{\mathcal{P}(\lambda w^2)}(1 - C),$$

the test is equivalent to the one that rejects H_0 if $M \geq 0$ where M denotes the scan statistic

$$M = \sup_{w \in [w_L, w_H]} \sup_{u,v \in I_w} \Psi_{w,u,v}(Y).$$

The calculation of M is simplified by noting that the constant c_w changes at only finitely many window widths in the interval $[c_L, c_W]$. Let

$$w_L = w_0 < w_1 < w_2 < \cdots < w_{m-1} < w_m = w_H,$$

where the w_i satisfy

$$F_{\mathcal{P}(\lambda w_i^2)}(c_{w_i}) = 1 - C, \quad i = 1, \ldots, m-1.$$

Then

$$c_w = c_{w_i} \quad \text{for} \quad w_{i-1} < w \leq w_i.$$

If $w < w'$, then any window of width w is contained in some window of width w'. Thus,

$$\sup_{u,v \in I_w} N(A_{w,u,v}) \leq \sup_{u,v \in I_{w'}} N(A_{w,u,v}).$$

If, in addition, $c_w = c_{w'}$ then we have

$$\sup_{u,v \in I_w} \Psi_{w,u,v} = \sup_{u,v \in I_{w'}} \Psi_{w',u,v}.$$

Since the critical constants c_w change only at the w_i we conclude that

$$M = \sup_{w \in \{w_0, \ldots, w_m\}} \sup_{u,v \in I_{w_i}} \Psi_{w,u,v}.$$

4.6 POINT PROCESS EXAMPLES

4.6.1 Monte Carlo Simulation Experiment

We now present the results of a Monte Carlo simulation experiment designed to compare the importance sampling algorithm with naive hit-or-miss sampling in terms of relative efficiency,

$$R = \frac{\text{hit-or-miss computation time}}{\text{importance computation time}},$$

where the sample sizes for the two procedures are determined so that they give equal width confidence intervals for the target p value.

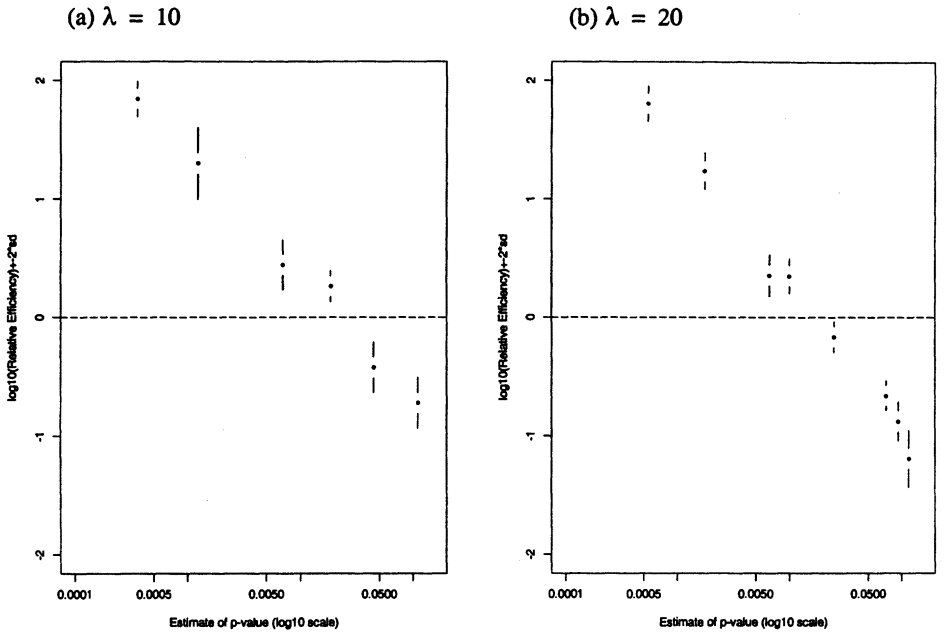


Figure 2. Comparison of Importance Sampling and Naive Estimation for the Spatial Poisson Case.

We consider a point process defined on the unit square and compare p value estimates obtained from the importance sampling algorithm and naive sampling for a test of the Poisson ($\mathcal{P}(\lambda)$) null hypothesis against the general alternative. We consider simultaneously five square scan window sizes w in $\{0.1, 0.125, 0.15, 0.175, 0.2\}$ representing a search for a cluster of between 1% and 4% of the overall process domain.

The simulation consists of specifying a value for the scan statistic M and running both Monte Carlo estimators to determine their per-trial variance and per-trial computational costs. For small p values the additional cost incurred in the importance sampling algorithm due to the requirement of generating samples from the appropriate conditional distribution will be compensated for by a smaller per-trial variance, yielding a relative efficiency greatly than unity and suggesting the superiority of the importance sampling approach.

Figure 2 shows log(Relative Efficiency) for this simulation experiment. Plots are presented for (2a) $\lambda = 10$ and (2b) $\lambda = 20$ for various values of the exceedance probability corresponding to various choices for the observed value of the scan statistic.

Investigation of Figure 2 indicates that the importance sampling algorithm outperforms naive hit-or-miss sampling at small p values, that the performance improvement can be quite dramatic, and that the performance improvement can be realized at practically important p values (the range of p values for which importance sampling is recommended is $(0, c)$ where $c > 0.01$).

Table 2. Comparison of Importance Sampling With Extreme Value Approximations

k	Import. Sampling	Naive Sampling	Rel. Effic.	Monte Carlo (C&G)	Product-type approx. (C&G)	Std. Poisson approx. (C&G)	Improved Poisson approx. (C&G)	Bonf.-type upper bd. (C&G)
15	0.2437 (.0040)	0.2446 (.0086)	4.52	0.2420	0.2830	0.4853	0.2954	0.4305
16	0.1060 (.0015)	0.1070 (.0062)	16.4	0.1068	0.1170	0.2077	0.1232	0.1613
17	0.0401 (.00051)	0.0408 (.00040)	59.8	0.0383	0.0423	0.0734	0.0446	0.0559
18	0.0138 (.00016)	0.0149 (.0024)	233.	0.0132	0.0138	0.0232	0.0146	0.0181
19	0.00438 (.000044)	0.00370 (.00121)	729.	0.0052	0.0042	0.0067	0.0044	0.0054

NOTE: The table presents the exceedance probability for an iid Binomial (5,0.05) process defined on a 25 × 25 grid using a 5 × 5 scan window. The Monte Carlo approaches (columns 1, 2, and 4) are each based on 10,000 trials. For columns 1 and 2 the number in parentheses is two times the standard error of the estimate. Columns 5–8 present the results of various extreme value estimators. Columns 4–8 are due to Chen and Glaz (1996).

4.6.2 Comparison of Importance Sampling with Extreme Value Approximations

As has been demonstrated, our method yields an unbiased p value estimate whose variance is typically smaller than that of the naive hit-or-miss Monte Carlo technique when the p value is small. Furthermore, as demonstrated in Table 2, our p value estimate is often accurate for critical values that are not far enough in the tails of the null distribution to allow for accurate approximations via extreme value theory.

Chen and Glaz (1996, tab. 3) presented a comparison of five approximations to the exceedence probability for a binomial process. The model consists of an 25 × 25 grid with the random variable at each grid location

$$Y_{u,v} \overset{\text{iid}}{\sim} \text{Binomial}(5, 0.05).$$

Scan windows $A_{u,v}$, $u = 1, \ldots, 21, v = 1, \ldots, 21$ forming 5 × 5 squares fitting inside the grid are considered, and the exceedence probability $P[M \geq k]$ for the scan statistic

$$M = \max_{u,v} N(A_{u,v})$$

is approximated by (1) naive simulation based on 10,000 trials, (2) Bonferroni-type inequality, (3) a standard Poisson approximation, (4) an improved Poisson approximation, and (5) a product-type approximation.

Table 2 presents a comparison of the performance of our importance sampling algorithm with these various competing approximations. An investigation of the table reveals that, as suggested by Chen and Glaz (1996), the product-type approximation and the improved Poisson approximation are reasonably accurate in the extreme tails of the null distribution (more accurate in fact than the Chen and Glaz Monte Carlo estimates). However, for practically important values of the exceedance probability not so far in the tails, such as

$0.04 \approx P[M \geq 17]$, these approximations are in error by more than ten standard deviations while at the same time the relative efficiency of importance sampling versus naive sampling approaches two orders of magnitude. Furthermore, these importance sampling estimates require only a matter of seconds in terms of their computational burden.

4.7 DIGITAL MAMMOGRAPHY EXAMPLE

The Monte Carlo simulation experiment presented in Section 4.6.1 was designed to have relevance to the problem of identifying cancer indicators in X-ray mammography.

Spiculated lesions, circumscribed masses, and clustered microcalcifications are among the important early indicators of malignant breast cancer, and early detection has been shown to improve survivability (Kopans 1998). Thus, the detection of clustered microcalcifications in screening mammograms is an area of considerable research activity in the computer-aided detection and diagnosis (CAD) community. The proceedings from the first four international workshops on digital mammography (Bowyer and Astley 1994; Gail, Astley, Dance, and Cairns 1994; Doi, Giger, Nishikawa, and Schmidt 1996; Karssemeijer, Thijssen, Hendriks, and van Erning 1998) include results from numerous efforts addressing this application.

We consider the map of detected candidate microcalcifications in a mammogram to be a realization of a point process and perform a test for homogeneity of the point process versus the alternative that the process has higher intensity in some subregion of the mammogram. Rejection of the null hypothesis of homogeneity identifies the region of higher intensity as a potential cluster of microcalcifications. Homogeneity may imply the "uniformly healthy tissue" case while a cluster warrants closer inspection.

To investigate the utility of spatial scan analysis in digital mammography, a mammogram is digitized and a microcalcification detector produces an observed point pattern of potential microcalcifications. (The point pattern may contain both true microcalcifications and false detections; furthermore, true microcalcifications that are present in the mammogram may be missed by the detector.) The example digitized mammogram presented in Figure 3 is from the Digital Database for Screening Mammography (DDSM) available from the University of South Florida and supported through a grant from the DOD Breast Cancer Research Program, U.S. Army Research and Material Command DAMD17-94-J-4015. The mammogram depicted in Figure 3 contains a cluster of microcalcifications (pathology: malignant) representing less than 4% of the image. A simple matched filter microcalcification detector (e.g., Castleman 1996; Jain 1989) produces a point pattern with 10 detections. The p value for this pattern, estimated via both importance and naive sampling with extraordinarily large sample sizes, is $p = 0.00085$. As the CAD application can be a time-critical one, a reasonable question is: How much computation is necessary to allow for a confident statement that the p value is less than 0.001?

To address this question, note that Figure 2(a) suggests that the relative efficiency for this problem is significantly greater than unity. (In fact, we estimate the relative efficiency to be 39.2.) In order to obtain an estimate of the p value with a standard deviation $\sigma_i <= 0.000075$, and thereby yielding $p + 2 * \sigma_i <= 0.001$, the importance sampling algorithm requires a computation time of 1 second. For the naive estimator to obtain the same accuracy ($\sigma_n = 0.000075$) requires a nearly 40-fold increase in computational effort, or 39.2 seconds

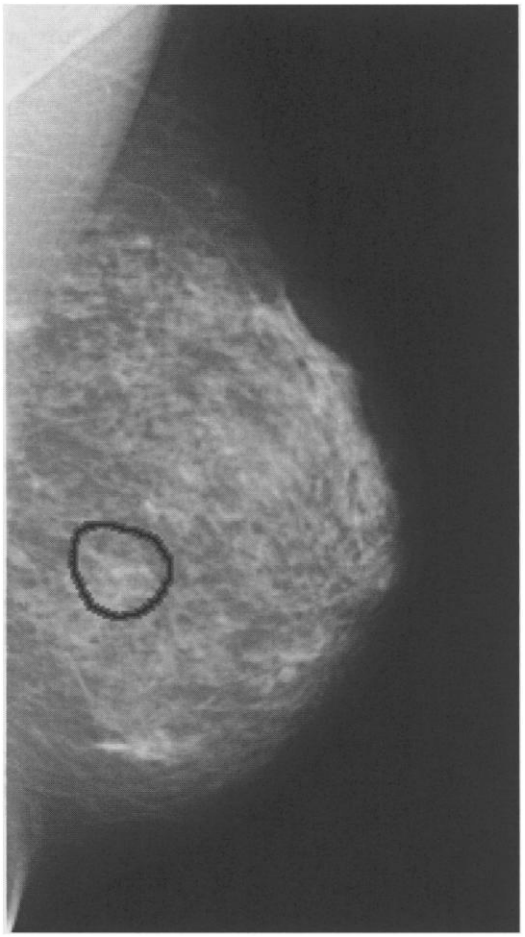


Figure 3. Digitized Mammogram for Point Process Analysis Example.

of computation time.

The reduced computational requirements of the importance sampling algorithm may play a part in the ultimate viability of a time-critical computer assisted diagnosis application such as screening mammography.

5. GAUSSIAN RANDOM FIELDS (GRFs)

We consider the problem of approximating the p value when testing for a signal with unknown location and (possibly) unknown scale in a stationary Gaussian random field. Siegmund and Worsley (1995) investigated the same problem using the differential geometric "volumes of tubes" approach. As they pointed out, differential geometry leads to expressions that are "related, but not equal to, the quantities of interest." We explain below why this is the case. On the other hand, the importance sampling approach gives an unbiased estimate of the exceedance probability whose variance we can control by adjusting

the number of iterations, and when the exceedance probability is small, we have a better method for quantifying the quality of the volume of tubes approximation than is provided by naive hit-or-miss sampling.

We now describe the setting in which the importance sampling procedure is used. Consider a Gaussian random field $Y = \{Y_x, x \in X\}$ indexed by some *finite* set X. Under the null hypothesis Y has mean 0 and some given covariance structure. We can represent this random field as obtained by forming linear combinations of a white noise sequence, so assume

$$Y_x = \sum_{j=1}^{h} a_{xj}\epsilon_j,$$

for some constants a_{xj}, where $\epsilon_j, j = 1, \ldots, h$ are iid $N(0,1)$ random variables.

Consider a finite collection of locality statistics $\{\Psi_i, i \in I\}$ where each Ψ_i is a linear function, say

$$\Psi_i(Y) = \sum_{x \in X} b_{ix} Y_x.$$

As a result, we have

$$\Psi_i(Y) = \sum_{j=1}^{h} c_{ij}\epsilon_j,$$

where

$$c_{ij} = \sum_{x \in X} b_{ix} a_{xj}.$$

The marginal null distribution of $\Psi_i(Y)$ is therefore $N(0, \sum_{j=1}^{h} c_{ij}^2)$.

A key to applying the importance sampling algorithm of Section 2 to approximating the distribution of $M = \max_{i \in I} \Psi_i(Y)$ is the ability to generate the random field Y conditionally given $\Psi_i(Y)$ for any particular choice of $i \in I$. This conditional distribution is described using the following elementary Lemma, which follows from the distribution of a multivariate normal random variable conditioned on a particular linear combination being fixed, and simple linear algebra.

Lemma 1. *Let $\epsilon_1, \ldots, \epsilon_h$ be iid $N(0,1)$ and suppose $c_1, \ldots, c_h$ are constants with $c_r \neq 0$ for some particular index r. Then conditionally given $\sum_{j=1}^{h} c_j\epsilon_j = t$, the $\epsilon_q, q \neq r$ are jointly distributed as*

$$\left(\frac{t}{\sum_{j=1}^{h} c_j^2} - \delta \left(\sum_{j \in \{1,\ldots,h\} \setminus \{r\}} c_j z_j \right) \right) c_q + z_q, \quad \text{for} \quad q \neq r,$$

where the z_j are iid $N(0,1)$, and

$$\delta = \frac{1}{\sum_{j \neq r} c_j^2} \left\{ 1 \pm \frac{c_r}{\sqrt{\sum_{j=1}^{h} c_j^2}} \right\}.$$

The lemma leads to an algorithm for generating $Y_x, x \in X$ conditionally given $\Psi_i(Y) = t$, the critical Step 2 of the importance sampling algorithm. Pick any index r such that $c_{ir} \neq 0$. Next, generate $z_q, q \neq r$ iid $N(0,1)$ and take

$$\epsilon_q = \left(\frac{t}{\sum_{j=1}^{h} c_{ij}^2} - \delta \left(\sum_{j \in \{1,\ldots,h\} \setminus \{r\}} c_{ij} z_j \right) \right) c_{iq} + z_q, \quad \text{for} \quad q \neq r,$$

and

$$\epsilon_r = \frac{t - \sum_{j \neq r} c_{ij} \epsilon_j}{c_{ir}}$$

(i.e., solve the equation $\sum_{j=1}^{h} c_{ij} \epsilon_j = t$ for ϵ_r). Finally, take

$$Y_x = \sum_{j=1}^{h} a_{xj} \epsilon_j \quad \text{for} \quad x \in X.$$

Remarks.

1. In the importance sampling algorithm we need to determine how many indices i are such that $\Psi_i(Y) \geq \tau$ for some threshold τ, and for this it is not necessary to calculate Y directly, since the *random field* $\Psi_i(Y), i \in I$ can be obtained using

$$\Psi_i(Y) = \sum_{j=1}^{h} c_{ij} \epsilon_j.$$

The constants c_{ij} can be determined as part of the setup for the algorithm.

2. If the locality statistic Ψ_i is the average of the random field Y at a small set of points, then generating the noise sequence $\epsilon_j, j = 1, \ldots, h$ conditionally given $\Psi_i(Y)$ involves modifying a white noise random field at a small set of points, where the number of sites for modification corresponds to the number of nonzero coefficients from among the c_{ij}, $j = 1, \ldots, h$.

3. For the special case when we scan the random field for a linear combination with a high z score, the c_{ij} are normalized so that $\sum_{j=1}^{h} c_{ij}^2 = 1$, for each $i \in I$. As a result, the probabilities

$$q_i = \frac{P[\Psi_i(Y) \geq t]}{\sum_{i \in I} P[\Psi_i(Y) \geq t]}$$

in the importance sampling algorithm (see Section 2) are all equal, so the index J generated in Step 1 of the algorithm is uniformly distributed in I.

In addition, we have

$$\delta = \frac{1}{1 \pm c_r}.$$

5.1 EXAMPLE: CONVOLUTIONS OF A GRFS ON A TOROIDAL GRID

An important example to consider is a GRF indexed by a toroidal grid, where the white noise and the locality statistics have the same index set, and both the random field and the locality statistics are obtained by convolution. Then much of the computational effort can be carried out efficiently using fast Fourier transforms (FFTs). The model is natural from a mathematical point of view, and appears, at first glance, to be unnatural from a practical point of view. However, if the GRF is indexed by a rectangular grid, and the locality statistics are not modified to take into account edge effects, then we can embed the rectangular grid in a larger toroidal grid, and modify the toroidal importance sampling algorithm in a simple way to handle this case, and thus retaining much of the efficiency gained from the use of FFTs.

The framework we consider is the following. The index set X forms a d-dimensional $m_1 \times \cdots \times m_d$ toroidal grid, $X = Z_1 \times \cdots \times Z_d$. This is the d-dimensional integer lattice with points $(x_1, \ldots, x_d)$ and $(x'_1, \ldots, x'_d)$ identified if $x_i \equiv x'_i \pmod{m_i}$ for $i = 1, \ldots, d$. Addition of elements of X is performed coordinatewise using modulus arithmetic. The random field Y is taken to be of the form

$$Y_x = \sum_{u \in X} a_u e_{x-u},$$

the convolution of a white noise random field $e_x, x \in X$ by a linear filter $\{a_x, x \in X\}$ (a collection of constants indexed by X). The locality statistics are also indexed by X and are obtained by convolving with another filter $\{b_x, x \in X\}$, so that

$$\Psi_v(Y) = \sum_{w \in X} b_w Y_{v-w},$$

and consequently

$$\Psi_v(Y) = \sum_{w \in X} c_w e_{v-w},$$

where the filter coefficients satisfy

$$c_w = \sum_{u \in X} b_w a_{u-w}.$$

The fact that we get the locality statistic process by convolving the noise with a filter allows us to create the conditional locality statistic field using fast Fourier transforms (FFTs). After computing the conditional noise field, we get the locality statistic field $\{\Psi_x(Y), x \in X\}$ by making a forward FFT of the noise field, coordinatewise multiplying by the FFT of the filter $\{c_x, x \in X\}$ (which need only be calculated once) and performing an inverse FFT.

5.2 EVALUATING THE "VOLUME OF TUBES" APPROACH

An alternative approach to the approximation of exceedance probabilities is to replace the discrete index set for the locality statistics by a continuous approximation, and use

techniques from differential geometry, in particular, the key Hotelling (1939) and Weyl (1939) *volume of tubes formula* and related developments. See Naiman (1986), Naiman (1990), Knowles and Siegmund (1989), Siegmund and Zhang (1993), Siegmund and Worsley (1995), Worsley (1995a), and Worsley (1995b). The resulting p values are multidimensional integrals giving close approximations to p values in the tails of the distribution.

Although importance sampling can give an improvement to hit-or-miss Monte Carlo sampling for approximating exceedance probabilities, it can be also be used to determine how well the tube approximations perform. We demonstrate this with a simple example. First, we give a brief review of the volume of tubes approach. An extensive overview of the volume of tubes approach is given by Adler (1998). Assume that our locality statistics are all normalized so that $\sum_{j=1}^{k} c_{ij}^2 = 1$ for $i \in I$. Then we can write the exceedance probability of interest as

$$P[M \geq t] = P\left[\max_{i \in I} \langle \gamma^{(i)}, E \rangle \geq t\right], \tag{5.1}$$

where $\gamma^{(i)}$ denotes the unit k vector with components c_{ij} and E is the random k vector with components ϵ_j, $j = 1, \ldots, k$. Write $E = RU$ where $R = \sqrt{\sum_{j=1}^{k} \epsilon_j^2}$ and $U = \frac{1}{R}E$, so that U is uniformly distributed in S^{k-1} the unit sphere in R^k, $R^2 \sim \chi_k^2$, and U and R are independent. Writing the probability of interest as an integral and making a change of variables leads to the expression

$$\int_{r=0}^{\infty} P\left[\max_{i \in I} \langle \gamma^{(i)}, U \rangle \geq t/r\right] f_{\chi_k^2}(r^2) 2r dr$$
$$= \int_{\theta=0}^{\pi/2} \mu\left[D(\Gamma, \theta)\right] f_{\chi_k^2}\left(\frac{t^2}{\cos^2\theta}\right) \frac{2t^2 \sin\theta}{\cos^3\theta} d\theta,$$

where

$$\Gamma = \left\{\gamma^{(i)} \,:\, i \in I\right\} \subseteq S^{k-1},$$

$D(\Gamma, \theta)$ denotes the so-called *tubular neighborhood* of Γ, of angular radius θ, which is the set of points in S^{k-1} whose angular distance from Γ is at most θ, and μ denotes the uniform measure in S^{k-1}.

If Γ is approximated by a submanifold Γ^* (possibly having a boundary) in S^{k-1} of some appropriate dimension, then we can use $\mu\left[D(\Gamma^*, \theta)\right]$ to approximate $\mu\left[D(\Gamma, \theta)\right]$. This idea was applied by McCann and Edwards (1996) in conjunction with the inequality in Naiman (1986) (see below) to compute critical values for certain multiple comparison procedures.

Fortunately, if Γ^* is sufficiently smooth and without *self-overlap,* and θ is sufficiently small Weyl and Hotelling's tube formulas and related results (see, e.g., Naiman 1990) give exact expressions for the latter measure, which take the form of integrals over Γ^* (and integrals over its boundary).

The key consequence of the assumptions that enables one to use differential geometry to determine the tube volume is that geodesic curves of length θ which emanate from the set Γ^* along normals to the set from different points do not intersect. In particular, if Γ^*

does not have a boundary, the tube $D(\Gamma^*, \theta)$ can be *coordinatized* in a natural way using the product space $\Gamma^* \times B_{k-1-\dim(\Gamma^*)}(\theta)$ where $B_p(\theta)$ denotes a p-dimensional ball of radius θ. The analogue of this statement when Γ^* does have a boundary was described by Naiman (1990).

For one- and two-dimensional sets Γ^*, expressions for the measure of the tube are rather simple to describe. If Γ^* forms a simple closed curve (the one-dimensional case) and the coordinatization condition mentioned above holds, then Hotelling's (1939) formula states that

$$\mu(\Gamma^*, \theta) = \frac{L}{2\pi} B_{\frac{N-2}{2},1}(\sin^2(\theta)), \tag{5.2}$$

the normalized length of the curve multiplied by the curve's cross-sectional area. Here $B_{\cdot,\cdot}(\cdot)$ denotes the beta distribution function.

Even if the coordinatization condition fails, this formula gives an approximation for the measure of the tube. For the case when Γ^* is a one-dimensional curve with endpoints, this expression is corrected by adding an additional term so that

$$\mu(\Gamma^*, \theta) = \frac{1}{2} B_{\frac{N-1}{2},\frac{1}{2}}(\sin^2(\theta)) + \frac{L}{2\pi} B_{\frac{N-2}{2},1}(\sin^2(\theta)). \tag{5.3}$$

On the other hand, Naiman (1986) (see also Johnstone and Siegmund 1989 and the references therein) showed that for a curve, either closed or not, even if the coordinatization condition fails, inclusion of the additional term always gives an upper bound for the measure of the tube. Knowles and Siegmund (1989) gave analogous expressions for the case of a two-dimensional submanifold with boundary. Sun (1993) gave an approach to bounding the exceedance probability for the case of higher-dimensional sets Γ^*.

Siegmund and Worsley (1995) introduced the volume of tubes approach to p value approximation for detecting a signal of unknown location and shape in a stationary Gaussian random field. Their starting point is a Gaussian random field on a continuous set, so that the set Γ is already a submanifold. They also derive the same approximations for exceedance probabilities using the alternative *expected Hadwiger characteristic of the excursion sets.*

Since Γ and I parameterize the space of locality statistics, and Γ^* can be viewed as doing this in an *approximate* sense, the volume of tubes approach and the importance sampling approach are similar in that both rely on p value expressions that can be viewed as integrals over *locality statistic space.* However, the differential geometric approach fails to correctly account for multiple overlap of the tube when it occurs. Even when the coordinatization condition holds for Γ^* there remains the issue for the differential geometric approach that Γ^* is only an approximation for Γ, when Γ is discrete. A program of research designed to further investigate the relative performance of the two approaches is currently underway. In the next section, we present a comparison for a simple case, when Γ^* is one-dimensional.

5.3 EXAMPLE: ASSESSING VOLUME OF TUBE APPROXIMATIONS FOR A GRF IN THE CIRCLE

We illustrate how importance sampling can be used to assess the volume of tubes approximation to exceedance probabilities by focusing on a simple but important special

case—approximating the distribution of the maximum of a stationary Gaussian random field indexed by a finite number N of equispaced points in the unit circle; that is, the one-dimensional case of a GRF in the toroidal grid. We take the random field to be the convolution of white noise with the filter having coefficients

$$c_x = C2^{-\left(\frac{|x|}{2\pi h}\right)^2},$$

where $|x|$ denotes the angular distance between x and 0 in the circle, and the constant C is chosen so that $\sum_x c_x^2 = 1$. Thus, the filter is define by a discretized Gaussian filter, and the quantity h gives the distance (in fraction of the circumference) at which the coefficients achieve half of their maximum value.

In this example, we take the locality statistics to be the observations Y_x themselves. We chose this simplified example because if we take a random field Y with Gaussian covariance function, and use locality statistics formed from Gaussian kernels, the resulting random field of locality statistics is approximately of this form. (This holds only "approximately," due to the fact that the circle, being compact, requires a truncated Gaussian kernel; the convolution of Gaussian kernels is not Gaussian in the circle.)

For our numerical example, we take the number of grid points N to be 64. In this context, the vectors $\gamma^{(i)}$ of (5.1) are all of the cyclic permutations of the coefficient vector $c = (c_0, \ldots, c_{N-1})$; that is, vectors of the form $c^{(i)} = (c_i, c_{i+1}, \ldots, c_{N-1}, c_0, \ldots, c_{i-1})$, for $i = 0, \ldots, N-1$. This set can be approximated by forming a piecewise great circular arc Γ^* by connecting successive points $c^{(0)}, \ldots, c^{(N-1)}, c^{(0)}$ to form a closed loop whose length is

$$L = N \cos^{-1}(\langle c^{(0)}, c^{(1)} \rangle).$$

In fact, a *circle* of approximately the same length is obtained when we take

$$\Gamma^* = \left\{ \gamma^{(x)}, 0 \le x \le 2\pi \right\},$$

where $\gamma^{(x)}$ has coordinates

$$\gamma_i^{(x)} = C2^{\frac{-|x-2\pi i/N|}{2\pi}}, \quad i = 0, \ldots, N-1.$$

Figure 4 graphs three approximations to the exceedance probability $P[\max_x Y_x \ge t]$ as we vary the half-width half-max for the filter coefficients, and for $t = 2, 3$, and 4. For the importance sampling approximation we used a sample size of 100,000 that yields a negligible standard error.

We see that the tube approximation (5.2) performs well provided h is not too large or too small. For large values of h the self-overlap of the tube becomes substantial as the approximating curve Γ^* collapses to a point. Then the tube approximation gives an underestimate of the exceedance probability and the true value is somewhere between the approximation (5.2) and the upper bound (5.3).

Figure 5 gives the relative efficiency (defined in Section 4.6.1) for importance sampling versus hit-or-miss sampling for each of the simulations used to produce Figure 4. We see that for even the smallest threshold considered ($t = 2$) the relative efficiency exceeds 1

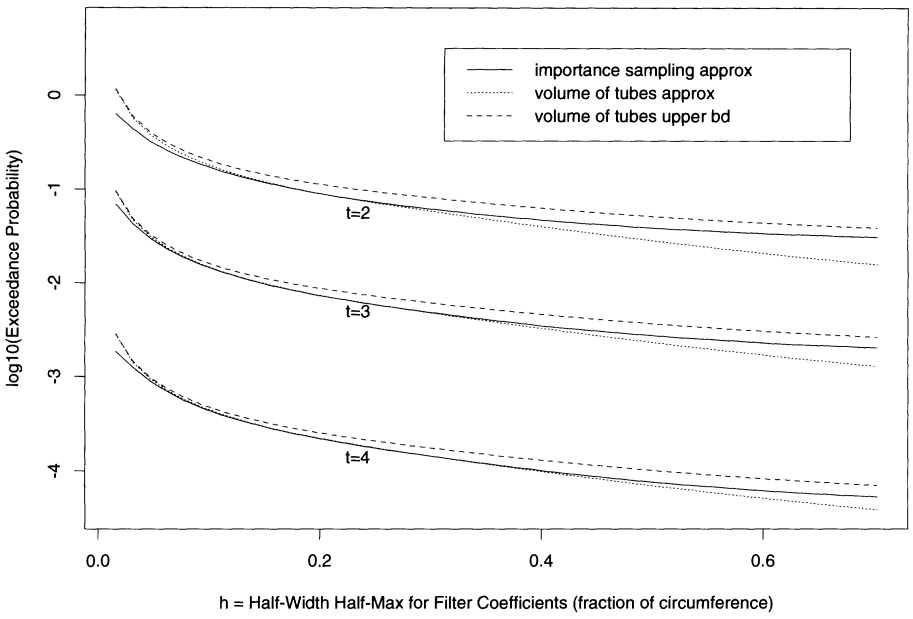


Figure 4. Comparison of Exceedance Probability Approximations Gaussian Random Field in the Circle.

throughout the range of values of h, and exceeds 10 when $h \geq .15$ For $t = 3$ the relative efficiency is about 100 and for $t = 4$ it is about 10,000. Thus, as we have seen in the point process examples the improvement of importance sampling over hit-or-miss sampling can be quite substantial.

Knowles and Siegmund (1989) used naive Monte Carlo simulation to assess the volume of tubes approximation for a different example, with a sample size of 10,000. Importance sampling provides a more accurate method for determining true exceedance probabilities for their example.

5.4 PET GRF EXPERIMENT

We now present the results of an illustrative example designed to suggest the utility of the importance sampling algorithm in the GRF setting and the applicability of our algorithm to positron emission tomography (PET) scan brain volume analysis.

PET indirectly measures regional cerebral blood flow (rCBF); see, for example, Cho, Jones, and Singh (1993). One application of PET involves collecting volumetric brain image data from each of a number of subjects during two different states, A and B, and subtracting these to produce a contrast image $C = B - A$. Local regions of high intensity in the contrast image C are considered to be regions with increased blood flow associated with state B relative to state A. This in turn is considered to be indicative of increased neural activity representing regionally specific effects attributable to state B. Assessing the significance of such regions is a stage in the attempt to understand the workings of the brain.

For a particular tone recognition task under investigation, six scans each of two condi-

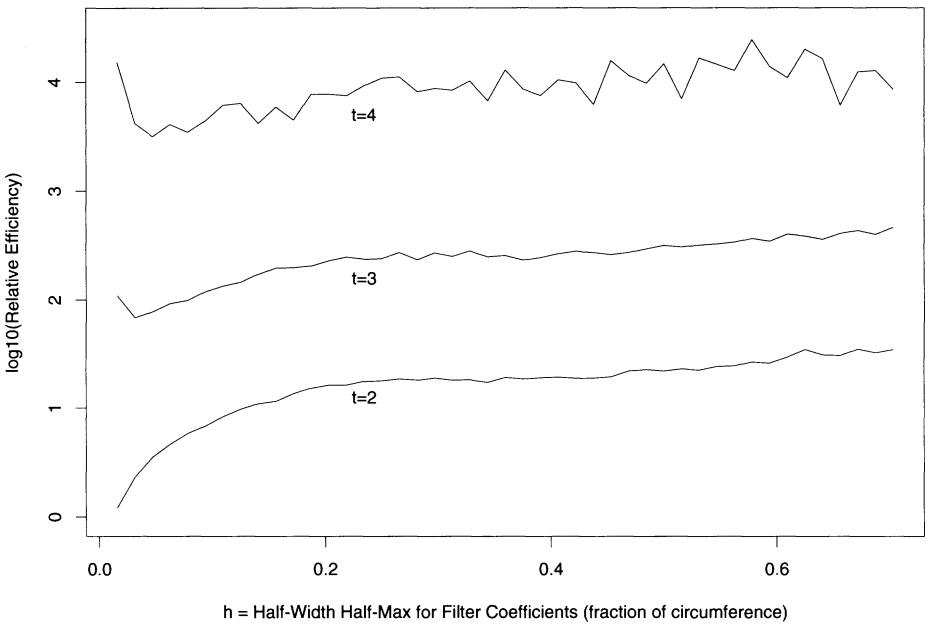


Figure 5. Relative Efficiency for Importance Versus Hit-or-Miss Sampling GRF in the Circle.

tions,

$$\text{Tone Recognition Decision} + \text{Sensory Motor Control},$$

and

$$\text{Sensory Motor Control Alone},$$

are obtained for each of 12 normal subjects. The contrast,

$$C = \text{Tone Recognition Decision} + \text{Sensory Motor Control} - \text{Sensory Motor Control Alone},$$

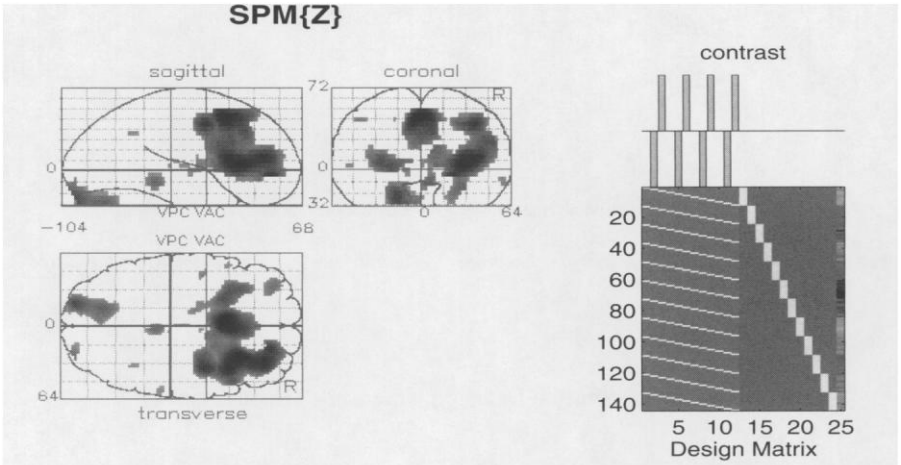
is considered; this application involves the identification of localized neural regions involved in the tone recognition decision.

A common and valuable approach to investigating such data is the statistical parametric mapping (SPM) approach of Friston et al. (1995). To facilitate intersubject pooling the software package SPM95, available from The Wellcome Department of Cognitive Neurology at University College London, performs anatomical registration of the dataset to the standard Talairach & Tournoux space (Talairach and Tournoux 1988) and subsequently performs an ANCOVA normalization routine. The size of the volumetric images in Talairach & Tournoux space is 147,030 voxels representing a $65 \times 87 \times 26$ voxel cube with voxel spacings of 2mm $\times$ 2mm $\times$ 4mm. The overall contrast volumetric image is obtained from pooling the normalized and registered data from the 12 subjects. After registration and normalization, this subtractive application allows for a null hypothesis of *no change* to be modeled as a GRF (or more accurately a large-df t-field) henceforth to be denoted by z.

Figure 6 presents the results of an SPM95 analysis of the tone recognition data. The field is considered to be a GRF whose spatial covariance function is Gaussian with a resolution (the full width half maximum (FWHM)) of [17.9, 20.0, 19.3] mm, so that the correlation between a pair of points whose distance apart is (d_x, d_y, d_z)mm is given by

$$2^{-\left\{(d_x/(17.9/2))^2+(d_y/(20.0/2))^2+(d_z/(19.3/2))^2\right\}}.$$

The actual search volume considered in the analysis consists of a subset of 47,s428 voxels in the 65 × 87 × 26 voxel Talairach & Tournoux cube; these are the voxels considered to



Regional effects: /ego/nma/auditory/normal

region	size {k}	P(n max > k)	Z	P(Z max > u)	(Uncorrected)	{x,y,z mm}
1	1072	0.014	6.07	0.000	(0.000)	−2 12 48
2	2473	0.000	5.71	0.000	(0.000)	28 22 4
			5.45	0.000	(0.000)	38 14 8
			5.35	0.000	(0.000)	34 38 12
3	345	0.263	4.66	0.005	(0.000)	−18 −94 −24
			3.79	0.130	(0.000)	−14 −74 −28
4	413	0.194	4.13	0.042	(0.000)	−24 12 4
			4.13	0.043	(0.000)	−30 18 8
			3.29	0.461	(0.000)	−12 12 8
5	65	0.850	3.66	0.193	(0.000)	4 −36 −12
6	45	0.901	3.42	0.350	(0.000)	26 −94 −16
7	73	0.829	3.12	0.616	(0.001)	−32 46 12
8	10	0.973	2.81	0.860	(0.002)	−44 −4 44
9	7	0.978	2.69	0.917	(0.004)	34 −54 32
10	13	0.967	2.59	0.951	(0.005)	58 −34 −12
11	6	0.979	2.51	0.969	(0.006)	−52 14 4
12	6	0.979	2.43	0.982	(0.008)	0 −18 4
13	1	0.988	2.37	0.987	(0.009)	−20 −8 44
14	1	0.988	2.35	0.988	(0.009)	6 −28 4
15	4	0.983	2.35	0.989	(0.009)	4 −6 4

Threshold = 2.33; Volume [S] = 47428 voxels; df = 120
FWHM = [17.9 20.0 19.3] mm (i.e. 110 RESELS)

SPM analysis – date: 22–Oct–97 user: nma

Figure 6. SPM Analysis for PET Scan Brain Image Gaussian Random Field Analysis Example.

actually represent locations in the brain. Excursion regions in the realization are defined as contiguous regions with z values above a threshold $z_t = 2.33$. For example, for our realization the significance of a three-dimensional excursion region around Talairach & Tournoux coordinates $(-24, 12, 4)$ in what is referred to as the insula on the edge of the putamen consisting of $n_{\text{region}} = 413$ voxels and with a maximum z value of $z_{\text{max}} = 4.13$ is in question (region 4 in Figure 6). To assess the significance of this excursion region, SPM95 reports a p value of $p_z = 0.042$ based on the maximum z score and $p_n = 0.194$ based on the size of the region. Both of these p values are based on extreme value theory. (See Friston et al. 1995, pp. 195–196; see also Adler 1981, Friston et al. 1991, Worsley et al. 1992, Friston et al. 1994.)

We now present an importance sampling simulation designed to provide an alternative estimate of the the significance of the excursion region in question. It is perhaps noteworthy that the importance sampling approach takes into account both the size of the observed excursion region and its z scores, thus allowing for a unified analysis of the region's significance.

We proceed by embedding a $50 \times 64 \times 16$ grid (a search volume of 51,200 voxels) in the torus and (ignoring edge effects) considering balls with radii of the form $r \times r \times r/2$ voxels centered near the reported cluster coordinates $(-24, 12, 4)$. For each ball we compute the locality statistic Ψ_i to be the normalized sum of the z values in that ball. Using $r = 5.8$ (a ball of voxel radius $5.8 \times 5.8 \times 2.9$ yielding a region consisting of 409 voxels, analogous to the size of the SPM excursion region in question) yields an observed value of $M_{\text{obs}} = 3.420$ for the spatial scan test statistic. The p value estimate for this region is 0.24 ± 0.01, indicating that this region is not significant.

A smaller ball ($r = 3.9$; the size yielding the maximum locality statistic) yields a test statistic value of $M_{\text{obs}} = 4.424$ and a highly significant p value estimated to be 0.00875 ± 0.00036.

Setting aside the fact that we have considered for simplicity only ball-shaped regions, it is not surprising that our method leads to an indication that the cluster is smaller in spatial extent than reported by SPM. The observed values for the random field in question (the PET image) are, due to the definition of the excursion region and the spatial dependence, likely to be near the threshold of $z_t = 2.33$ near the edge of the excursion region. As such, an observed locality statistic for spatial scan analysis with a window geometry and location identical to that of the excursion region can be smaller than when using smaller scan windows centered at the field peak.

The application—a search for regionally specific effects of unknown spatial extent—requires simultaneous consideration of multiple window sizes. Toward this end we consider balls ranging in size from 1 voxel ($r = 0$) to a maximum cluster size of approximately 2,000 voxels ($r = 16$) in steps of size $\Delta r = 1$. This analysis yields a p value estimate of 0.03 ± 0.0025. The conclusion is, therefore, that the experiment *does* suggest a significant effect in the insula on the edge of the putamen, and that if attention is restricted to ball-shaped regions the significant effect is smaller than the 413 voxel excursion region reported by SPM. Furthermore, this example demonstrates the necessity of accounting for the simultaneous consideration of multiple window sizes, as the p value increases dramatically (from 0.00875 to 0.03) when the geometry being searched for is not considered to be known a priori.

ACKNOWLEDGMENTS

This work is partially supported by the Office of Naval Research Grant N00014-95-1-0777 and by the National Science Foundation Grant DMS-9504242. The authors thank Joseph Naus for providing information related to his own work and used in Section 3.3. We also thank Kevin Bowyer and Michael Heath for providing the digitized mammogram in Figure 3, and Henry Holcomb and Ning Ma for the PET brain scan data and for help in interpreting their analyses. Finally, we thank three anonymous reviewers and an associate editor for comments which greatly improved the readability of this article.

[Received March 1999. Revised August 1999.]

REFERENCES

Adler, R. J. (1981), *The Geometry of Random Fields*, New York: Wiley.

——— (1984), "The Supremum of a Particular Gaussian Field," *Annals of Probability*, 12, 436–444.

——— (1998), "On Excursion Sets, Tube Formulae, and Maxima of Random Fields," unpublished manuscript.

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

Alm, S. E. (1997). "On the Distributions of Scan Statistics of a Two-Dimensional Poisson Process," *Advances in Applied Probability*, 29, 1–18.

Bonferroni, C. E. (1936a), *Pubbl. 1st Sup. Sci. Econ. Communic. Firenze*, 8, 1–62.

——— (1936b), "Il Calcolo delle assicurazioni su gruupi di teste," in *Studii in onore del prof. S. O. Carboni*, Roma.

Bowyer, K. W., and Astley, S. (eds.) (1994), "State of the Art in Digital Mammographic Image Analysis: Proceedings of the 1st International Workshop on Digital Mammography," San Jose, Singapore: World Scientific.

Castleman, K. R. (1996), *Digital Image Processing*, Englewood Cliffs, NJ: Prentice-Hall.

Chen, J., and Glaz, J. (1996), "Two-Dimensional Discrete Scan Statistics," *Statistics and Probability Letters*, 31, 59–68.

Cho, Z.-H., Jones, J. P., and Singh, M. (1993), *Foundations of Medical Imaging*, New York: Wiley.

Cressie, N. (1977), "On Some Properties of the Scan Statistic on the Circle and the Line," *Journal of Applied Probability*, 14, 272–283.

——— (1980), "The Asymptotic Distribution of the Scan Statistic Under Uniformity," *Annals of Probability*, 8, 828–840.

——— (1993), *Statistics for Spatial Data*, New York: Wiley.

Diggle, P. J. (1983), *Statistical Analysis of Spatial Point Patterns*, New York: Academic Press.

Doi, K., Giger, M. L., Nishikawa, R. M., and Schmidt, R. A. (eds.) (1996), *Digital Mammography '96: Proceedings of the 3rd International Workshop on Digital Mammography*, Amsterdam: Elsevier.

Fisher, R. A., Thornton, H. G., and Mackenzie, W. A. (1922), "The Accuracy of the Plating Method of Estimating the Density of Bacterial Populations, with Particular Reference to the Use of Thornton's Agar Medium with Soil Samples," *Annals of Applied Biology*, 9, 325–359.

Fishman, G. (1996), *Monte Carlo: Concepts, Algorithms, and Applications*, New York: Springer-Verlag.

Frigessi, A., and Vercellis, C. (1984), "An Analysis of Monte Carlo Algorithms for Counting Problems," IAMI-84.2, Department of Mathematics, University of Milan.

Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. J. (1991), "Comparing Functional (PET) Images: The Assessment of Significant Change," *Journal of Cerebral Blood Flow and Metabolism*, 11, 690–699.

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., and Evans, A. C. (1994), "Statistical Parametric Maps in Functional Imaging: A General Linear Approach," *Human Brain Mapping*, 1, 214–220.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. J. (1995), "Statistical Parametric Maps in Functional Imaging: A General Linear Approach," *Human Brain Mapping*, 2, 189–210.

Gail, A. G., Astley, S. M., Dance, A. R., and Cairns, A. Y. (eds.) (1994), *Digital Mammography: Proceedings of the 2nd International Workshop on Digital Mammography*, Amsterdam: Elsevier.

Hotelling, H. (1939), "Tubes and Spheres in n-Spaces, and a Class of Statistical Problems," *American Journal of Mathematics*, 61, 440–460.

Jain, A.K. (1989), *Fundamentals of Digital Image Processing*, Englewood Cliffs, NJ: Prentice-Hall.

Johnstone, I., and Siegmund, D. (1989), "On Hotelling's Formula for the Volume of Tubes and Naiman's Inequality," *The Annals of Statistics*, 17, 184–194.

Karlin, S., and Altschul, S. F. (1990), "Methods for Assessing the Statistical Significance of Molecular Sequence Features by Using General Scoring Schemes," *Proceedings of the National Academy of Science*, 87, 2264–2268.

Karlin, S., Blaisdell, B. E., Mocarski, E. S., and Brendel, V. (1989), "A Method to Identify Dinstinctive Charge Configurations in Protein Sequences, with Application to Human Herpesvirus Polypeptides," *Journal of Molcecular Biology*, 205, 165–177.

Karlin, S., Dembo, A., and Kawabata, T. (1990), "Statistical Composition of High-Scoring Segments from Molecular Sequences," *The Annals of Statistics*, 18, 571–581.

Karlin, S., and Dembo, A. (1992), "Limit Distributions of Maximal Segmental Score Among Markov-Dependent Partial Sums," *Advances in Applied Probability*, 24, 113–140.

Karssemeijer, N., Thijssen, M., Hendriks, J., and van Erning, L. (eds.) (1998), *Digital Mammography '98: Proceedings of the 4th International Workshop on Digital Mammography*, Dordrecht, The Netherlands: Kluwer.

Karwe, V. V., and Naus, J. I. (1997), "New Recursive Methods for Scan Statistic Probabilities," *Computational Statistics and Data Analysis*, 23, 389–402.

Knowles, M., and Siegmund, D. (1989), "On Hotelling's Approach to Testing for a Nonlinear Parameter in Regression," *International Statistics Review*, 57, 205–220.

Kopans, D. B. (1998), *Breast Imaging* (2nd ed.), Philadelphia: Lippincott-Raven.

Kulldorff, M. (1997), "A Spatial Scan Statistic," *Communications in Statistics: Theory and Methods*, 26, 1481–1496.

Loader, C. R. (1991). "Large-Deviation Approximations to the Distribution of Scan Statistics," *Advances in Applied Probability*, 23, 751–771.

McCann, M., and Edwards, D. (1996), "A Path Length Inequality for the Multivariate t Distribution with Applications to Multiple Comparisons," *Journal of the American Statistical Association*, 91, 211–216.

Naiman, D. (1986), "Conservative Confidence Bands in Curvilinear Regression," *The Annals of Statistics*, 14, 896–906.

——— (1990), "Volumes of Tubular Neighborhoods of Spherical Polyhedra and Statistical Inference," *The Annals of Statistics*, 18, 685–716.

Naiman, D., and Wynn, H. P. (1992), "Inclusion-Exclusion-Bonferroni Identities and Inequalities for Discrete Tube-Like Problems via Euler Characteristics," *The Annals of Statistics*, 20, 43–76.

——— (1997), "Abstract Tubes, Improved Inclusion-Exclusion Identities and Inequalities, and Importance Sampling," *The Annals of Statistics*, 25, 1954–1983.

Naus, J. I. (1965), "Clustering of Random Points in Two Dimensions," *Biometrika*, 52, 263–267.

Priebe, C. E., Olson, T., and Healy, D. M. (1997), "A Spatial Scan Statistic for Stochastic Scan Partitions," *Journal of the American Statistical Association*, 92, 1476–1484.

Ross, S. H. (1990), *A Course in Simulation*, New York: MacMillan.

Rudin, W. (1974), *Real and Complex Analysis* (2nd ed.), New York: McGraw-Hill.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Siegmund, D., and Worsley, K. (1995), "Testing for a Signal with Unknown Location and Scale in a Stationary Gaussian Random Field," *The Annals of Statistics*, 23, 608–639.

Siegmund, D., and Zhang, H. (1993), "The Expected Number of Local Maxima of a Random Field and the Volume of Tubes," *The Annals of Statistics*, 21, 1948–1966.

Sun, J. (1993), "Tail Probabilities of the Maxima of Gaussian Random Fields," *The Annals of Probability*, 21, 34–71.

Talairach, J., and Tournoux, P. (1988), *Co-Planar Stereotaxic Atlas of the Human Brain*, Stuttgart: Thieme.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., and Weiner, A. A. (1987), *Molecular Biology of the Gene* (4th Ed.), Menlo-Park, CA: Benjamin/Cummings.

Weyl, H. (1939), "On the Volume of Tubes," *American Journal of Mathematics*, 61, 461–472.

Worsley, K. J. (1982), "An Improved Bonferroni Inequality," *Biometrika*, 69, 297–302.

——— (1985), "Bonferroni Wins Again," *The American Statistician*, 39, 235.

——— (1995a), "Boundary Corrections for the Expected Euler Characteristic of Excursion Sets of Random Fields, with an Application to Astrophysics," *Advances in Applied Probability*, 27, 943–959.

——— (1995b), "Estimating the Number of Peaks in a Random Field Using the Hadwiger Characteristic of Excursion Sets, with Applications to Medical Images," *The Annals of Statistics*, 23, 640–669.

Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. (1992), "A Three-Dimensional Statistical Analysis for rCBF Activation Studies in Human Brain," *Journal of Cerebral Blood Flow and Metabolism*, 12, 900–918.