Tilted Euler characteristic densities for Central Limit random fields, with application to 'bubbles'

NICHOLAS CHAMANDY   KEITH J. WORSLEY   JONATHAN TAYLOR

FRÉDÉRIC GOSSELIN *

August, 2007

Abstract

Local increases in the mean of a random field are detected (conservatively) by thresholding a field of test statistics at a level u chosen to control the tail probability or p-value of its maximum. This p-value is approximated by the expected Euler characteristic (EC) of the excursion set of the test statistic field above u, denoted $\mathbb{E}\varphi(A_u)$. Under isotropy one can use the expansion $\mathbb{E}\varphi(A_u) = \sum_k \mathcal{V}_k \rho_k(u)$, where $\mathcal{V}_k$ is an intrinsic volume of the parameter space and ρ_k is an EC density of the field. EC densities are available for a number of processes; mainly those constructed from (multivariate) Gaussian fields via smooth functions. Using saddlepoint methods, we derive an expansion for $\rho_k(u)$ for fields which are only approximately Gaussian, but for which higher-order cumulants are available. We focus on linear combinations of n independent non-Gaussian fields, whence a Central Limit theorem is in force. The threshold u is allowed to grow with the sample size n, in which case our expression has a smaller relative asymptotic error than the Gaussian EC density. Several illustrative examples including an application to 'bubbles' data accompany the theory.

Abbreviated title: Euler characteristics of Central Limit fields.

1 Introduction

Data which can be modeled as realizations of a smooth random field are increasingly abundant, owing to advances in scanning technology and computer storage. Traditionally, such data have arisen in the fields of oceanography, astronomy and medical imaging (see [44] for an easy introduction). More recently (e.g. [18] or the January 2005 front cover of *Nature* [3]), the 'bubbles' paradigm in the cognitive sciences has provided new and interesting applications. The difficult inference step is most often assessing the significance of the test statistic field arising from such data, as the multiple comparisons problem precludes 'pixel-wise' significance testing. Commonly, the supremum of the statistic image is compared to an estimate of its upper p^{th} quantile under the global null hypothesis of a zero mean function (equivalently the statistic field is *thresholded* at this quantile to identify regions of activation). A leading approach to approximating these quantiles uses the expected Euler characteristic (EC) of the excursion of the field above a threshold ([2], [42]), as described below. However, a common thread in these analyses is the assumption that the random field data are Gaussian; in some applications this assumption is known to be flawed.

Practitioners of voxel-based morphometry in neuroimaging have highlighted the problem [31, 41], but have stopped short of proposing a theoretical solution. In the context of galaxy density random fields, Matsubara [25, 26] obtained approximations to the expected EC in three dimensions using Edgeworth series (though he never made the jump to a saddlepoint expansion). Catelan and coauthors [10] had previously used a similar approach (also in an astrophysics context) to get at the mean cluster size of excursions for non-Gaussian fields. Rabinowitz & Siegmund [29] used saddlepoint methods to approximate the tail probability for the supremum of a particular non-

Gaussian field, a smoothed Poisson point process. Although they supplied a first-order version of the expected EC and published heuristic arguments, their work originated some of the central lines of reasoning of the current article (indeed, since the paper is cited often we henceforth refer to it as RS). We shall formulate a general approximation to the expected EC for asymptotically Gaussian fields with known (decaying) cumulants, and compare its relative error to that under a full Gaussian assumption. In particular, our approximation is found to outperform the Gaussian expected EC in a particular 'large threshold' asymptotic setting. Throughout, we underscore the connection between our results and those of RS.

For concreteness, we consider processes of the form

$$Z(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i(t), \tag{1.1}$$

where the W_i's are i.i.d. smooth random fields on $T \subset \mathbb{R}^N$ with zero mean, constant variance σ^2, and finite higher cumulants. Such a $Z(t)$ is a special case of what we call a Central Limit random field (CLRF), and usually should be thought of as a test statistic derived from data. We call $A_u = A_u(Z, T) = \{t \in T; \ Z(t) \geq u\}$ the *excursion set* of Z above u. We denote the EC, an integer-valued set functional, by $\varphi(\cdot)$. For a working definition of the EC see [38], and for a more rigourous account the reader is referred to Part II of Adler & Taylor [2]. As the latter volume is the second major source of methodology used in this article, it will be referenced as AT. Very loosely (ignoring what happens on the boundary), in 2D the EC counts #'blobs' − #'holes' in a set, while in 3D it counts #'blobs' − #'tunnels' + #'hollows' [44].

In some applications the expected EC of A_u as a function of u is itself of central interest. For example, cosmologists compare the observed EC (or 1−EC, sometimes called the

3

genus) from galaxy survey data with the expected EC from various hypothesized models of the large-scale structure of the universe (see [21] and the illustration in Figure 1). In other applications, including neuroimaging and the bubbles problem treated in Section 8, interest centres around using this quantity as an approximation to the *p*-value:

$$\mathbb{E}\varphi(A_u) \approx \mathbb{P}\left\{\sup_{t \in T} Z(t) \geq u\right\} \tag{1.2}$$

for large *u*. The assumption that (1.2)—the so-called 'EC heuristic'—is adequate for statistical inference is no longer a mere heuristic for Gaussian fields; see [37] or AT, Chapter 14. See also Section 4.4 of the present paper.

To motivate the study of what are known as *EC densities*, we have the following well-established result.

Theorem 1 ([43], AT). *Let $Z(t)$ be any isotropic random field on $T \subset \mathbb{R}^N$, a locally convex finite union of convex bodies. Then the expected EC of its excursion set at the level $u \in \mathbb{R}$ can be written*

$$\mathbb{E}\varphi(A_u) = \sum_{k=0}^{N} \mathcal{V}_k(T)\rho_k(u), \tag{1.3}$$

where $\mathcal{V}_k(T)$ is the kth intrinsic volume of the search region and $\rho_k(u)$ is a function called the kth EC density of the field.

Proof. The expected EC $\mathbb{E}\varphi(A_u(Z, \cdot))$ is a real-valued additive functional acting on T;
$\mathbb{E}\varphi(A_u(Z, T \cup V)) = \mathbb{E}\varphi(A_u(Z, T) \cup A_u(Z, V)) = \mathbb{E}\varphi(A_u(Z, T)) + \mathbb{E}\varphi(A_u(Z, V))$
$- \mathbb{E}\varphi(A_u(Z, T) \cap A_u(Z, V)) = \mathbb{E}\varphi(A_u(Z, T)) + \mathbb{E}\varphi(A_u(Z, V)) - \mathbb{E}\varphi(A_u(Z, T \cap V))$. If the process is isotropic then $\mathbb{E}\varphi(A_u(Z, \cdot))$ is also translation- and rotation-invariant. Finally, the local convexity of T guarantees that the functional is continuous (for example in the Hausdorff metric). The result then follows from Hadwiger's seminal theorem

[20, 32], which states that any such functional is in the linear span of the intrinsic volumes, and hence admits an expansion like (1.3).   □

Remark If the field is nonisotropic but derived from a (multivariate) Gaussian field by a smooth function, then on sufficiently regular parameter spaces (1.3) holds with $\mathcal{V}_k(T)$ replaced by the kth Lipschitz-Killing curvature $\mathcal{L}_k(T)$. The latter differential geometric functionals depend both on the space T and on the covariance structure of the field itself; see AT Chapter 10.

1.1   Intrinsic volumes

The intrinsic volumes of a set T are a generalization of its volume to lower dimensional measures. There are several ways of describing these functionals; here we give an implicit definition, which is valid for any locally convex set T and sufficiently small radius r. Let $|\cdot|$ denote Lebesgue measure, $B(0, r)$ be the ball of radius r centred at 0, and $T \oplus B(0, r) = \{x + y : x \in T, y \in B(0, r)\}$ be the tube of radius r around T. Write $\omega_k = |B_k(0, 1)| = \pi^{k/2}/\Gamma(k/2 + 1)$ for the Lebesgue measure of the unit ball in $\mathbb{R}^k$. Then

$$|T \oplus B(0, r)| = \sum_{k=0}^{N} \omega_{N-k} r^{N-k} \mathcal{V}_k(T). \tag{1.4}$$

Thus, the intrinsic volumes are related to the coefficients in the Steiner-Weyl volume of tubes expansion. For example, in $\mathbb{R}^3$ we have $\mathcal{V}_0(T) = \varphi(T)$, $\mathcal{V}_1(T) = 2 \times$caliper diameter$(T)$, $\mathcal{V}_2(T) = 1/2 \times$surface area(T), and $\mathcal{V}_3(T) =$ volume(T). Remarkably, EC densities for all 'Gaussian-related' processes were shown to have a similar characterization in [35].
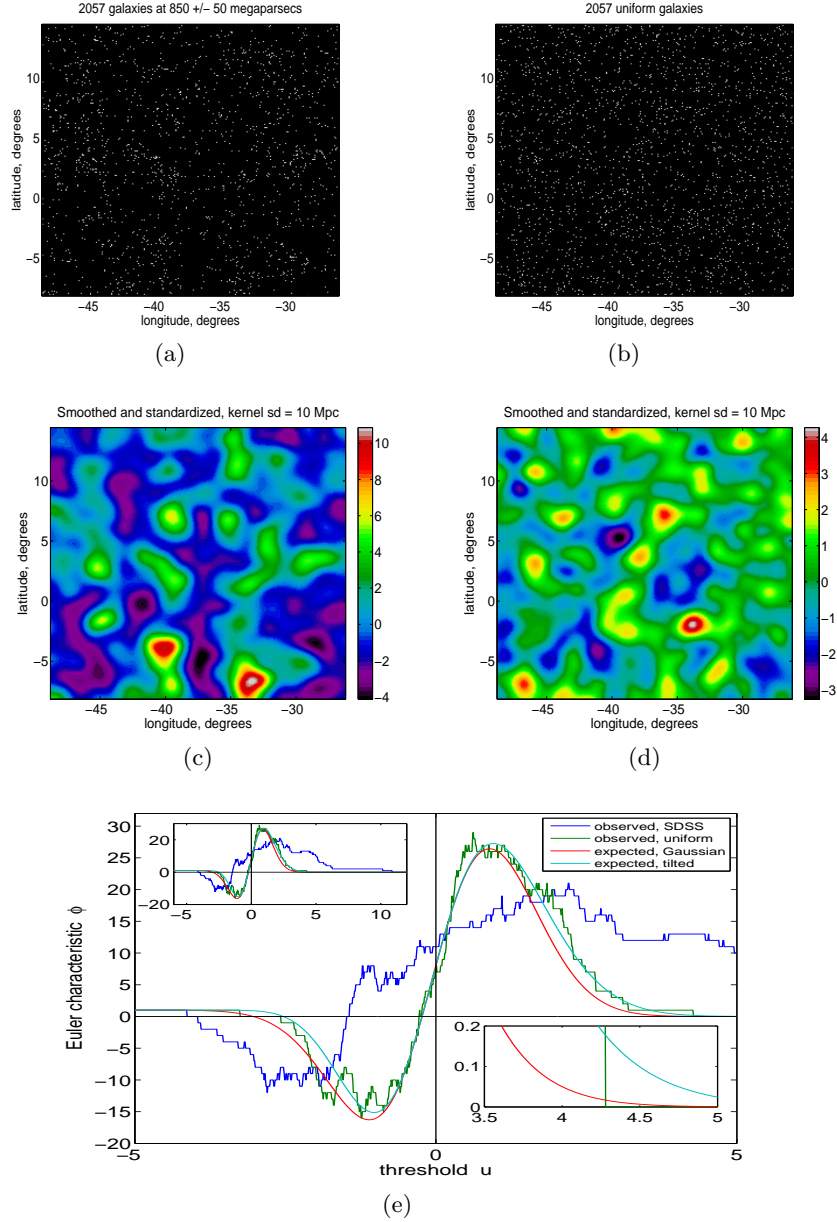
Figure 1: Galaxies are not uniformly distributed in the universe, but tend to congregate in clusters, strings or even sheets. Cosmologists have used the EC of regions of high galaxy density to characterize this large-scale structure ([21] used data from the first release of the Sloan Digital Sky Survey SDSS). In 2D the EC counts #'blobs' − #'holes' in a set. We examine luminous red galaxies (LRG's—the most distant objects in the SDSS) from the latest (5th) data release [28]. In (a) we show LRG's from a $22.5°$ square patch at 850 ± 50 megaparsecs (Mpc) with roughly the same galaxy density as in [21] (our sample is part of a sphere, rather than cone-shaped, so we avoid the luminosity correction). Panel (b) displays the same number of 'galaxies' simulated uniformly according to a Poisson process. In (c)-(d) the LRG's are smoothed with a Gaussian kernel of s.d. 10 Mpc, and standardized to have zero mean and unit variance under the Poisson model. In (e) we compare the *observed* EC of each image at a continuum of thresholds with the *expected* EC under the Poisson model computed using a Gaussian approximation and the 'tilted' correction derived in Sections 4-5. The observed EC for the SDSS image deviates from the others, reinforcing the non-uniformity of LRG's. The Gaussian and tilted expected EC's differ primarily in the tails; in the range where the expected EC approximates a p-value (bottom-right zoom) this correction is dramatic. In particular the tilted curve majorizes the Gaussian one since the LRG distribution has positive cumulants. The observed EC for the (asymptotically Gaussian) Poisson data conforms well with both curves. Note that all curves approach 1 as $u \to -\infty$ and 0 as $u \to \infty$ (top-left zoom).

1.2 Computing EC densities

For a matrix A or vector a we introduce the notation $A|_k = (A_{ij})_{1 \leq i,j \leq k}$ and $a|_k = (a_i)_{1 \leq i \leq k}$. When $k = 0$ this denotes, by convention, an 'empty' matrix or vector—*i.e.* a term or factor which disappears from the expression. We use I_p (or just I when the dimension is unambiguous) to denote the $p \times p$ identity matrix. When applied to a random field, ' $\dot{}$ ' represents spatial differentiation with respect to t, and we shall refer to the individual components of $\dot{Z}$ and $\ddot{Z}$, respectively, by Z_i and Z_{ij}. There are several useful formulations for computing the EC densities of stationary fields [43]. In the derivation of Theorem 4 we use

$$\rho_k(u) = \mathbb{E}\left\{ Z_k \mathbb{1}_{[Z_k \geq 0]} \det\left(-\ddot{Z}|_{k-1}\right) \Big| Z = u, \dot{Z}|_{k-1} = 0 \right\} f_{0,k-1}(u, 0) \qquad (1.5)$$

for $k = 1, \ldots, N$, where $f_{0,k-1}(\cdot)$ denotes the joint density of $\left(Z, \dot{Z}|'_{k-1}\right)$. Indeed, this was the original method used by Adler to compute the expected EC, and corresponds to using the Nth coordinate function $(t_1, t_2, \cdots, t_N) \mapsto t_N$ (instead of the field itself) as the function in Morse's Theorem (see [1], Chapter 4). The zeroth EC density is just the univariate tail probability, $\rho_0(u) \triangleq \mathbb{P}\{Z(t) \geq u\}$. It has a clear interpretation as the lowest order approximation to the p-value (1.2).

1.3 A glimpse ahead

EC densities are available in diverse parameter spaces for a number of processes [8, 9, 34, 36, 39, 42]; mainly those derived from (multivariate) Gaussian fields via smooth functions. Let $H_m(x) = m! \sum_{j=0}^{\lfloor m/2 \rfloor} \frac{(-1)^j x^{m-2j}}{2^j j! (m-2j)!}$ be the mth Hermite polynomial. For a mean zero Gaussian field with variance σ^2, the densities take the form (c.f. AT, Chapter

11)

$$\rho_k^\gamma(u) \triangleq (2\pi)^{-(k+1)/2} \sigma^{-k} \exp\left\{-\frac{u^2}{2\sigma^2}\right\} H_{k-1}\left(\frac{u}{\sigma}\right). \tag{1.6}$$

(Note that γ is used to denote the Gaussian measure, following the convention of AT.) Our main result (Theorem 4) will be to show that for a stationary CLRF like (1.1), and assuming that the threshold u is of the order $n^{1/2-\alpha}$ for some $1/4 < \alpha \le 1/2$, modifying only the exponential portion of (1.6) leads to an approximation to $\rho_k(u)$ whose relative error behaves like $n^{1/2-2\alpha}$. Adding an additional factor related to the mixed skewness terms $\mathbb{E}[Z^2 Z_i]$ can yield an additional improvement. Moreover, we shall show that this error is asymptotically strictly better than that of (1.6), exponentially so over the range $1/4 < \alpha \le 1/3$.

The remainder of the paper is organized as follows. In Section 2 we use saddlepoint methods to approximate the joint density of $(Z, \dot{Z}', \ddot{Z}')$ (suitably transformed), leading to a mixed tilted-direct Edgeworth expansion. In Section 3 we consider the growth rate of u and compute certain moments of the tilted distribution which shall be used in subsequent derivations. We derive the approximation to $\rho_k(u)$, by methods reminiscent of [1] and [42], in Section 4. Section 5 offers an analysis of the relative error of the tilted densities compared to that of a standard Gaussian approximation, and we comment on the contribution of RS in Section 6. In Section 7 we illustrate the method by treating a field—the χ_n^2 process—whose densities are known exactly, and Section 8 applies the theory to the test statistic from a bubbles experiment. We also present a simulation under the latter framework which examines the accuracy of our corrected version of (1.3) as a p-value approximation. In the interest of brevity, we have omitted several proofs and technical developments, all of which can be found in an online version of the present article (URL), or in the doctoral thesis [11].

8

2 A mixed expansion

In order to avoid having EC densities which depend on t in Theorem 4 below, we shall assume that Z is strictly stationary. For practical purposes (*i.e.* to apply Theorem 1 in the computation of $\mathbb{E}\varphi(A_u)$) the stronger requirement that Z be strictly isotropic shall be in force. In that case, purely for convenience, Z will be assumed to have unit second spectral moment, *viz.* $\mathbb{E}[Z_i Z_j] = \delta_{ij}$. The latter assumption can easily be relaxed to allow $\mathsf{Var}[\dot{Z}(t)] \equiv \lambda I$, which merely modifies (1.3) by multiplying the kth term by $\lambda^{k/2}$. Strict isotropy can, on very simple parameter spaces, be weakened to stationarity (with a similar correction to (1.3)); see Section 4.3 or AT, Chapter 11. If $\xi \geq 0$, the notation $x_n = O\left(n^{-\xi}\right)$ will indicate that for some $0 < C < \infty$ and all sufficiently large n, $|x_n| \leq C n^{-\xi}$. A similar notation is used for matrices, with the understanding that the relation holds componentwise. The notation $a_n \sim b_n$ signifies that $a_n/b_n \to 1$ as $n \to \infty$.

Fix $k \geq 2$ and put

$$M = M_k \triangleq \sigma \ddot{Z}|_{k-1} + \frac{Z}{\sigma} I_{k-1}. \tag{2.1}$$

The most important property of this shifted $(k-1) \times (k-1)$ Hessian is that it is uncorrelated with the field: $\mathbb{E}\{M_{ij}Z\} \equiv 0$. Let $\lambda_{ijlm} \triangleq \mathbb{E}\{Z_{ij}Z_{lm}\}$ (a symmetric, $O(1)$ function) denote a fourth-order spectral moment of the process. It is easily shown that

$$\lambda_{ijlm}^M \triangleq \mathbb{E}\{M_{ij}M_{lm}\} = \sigma^2 \lambda_{ijlm} - \delta_{ij}\delta_{lm}.$$

We shall write M to denote both the matrix and the corresponding random vector $\text{vec}(M)$, with the context eliminating confusion (similarly for $\ddot{Z}$). In view of this, let $\mathbf{Z} = \mathbf{Z}_k^n \triangleq (Z, \dot{Z}|_k', M')'$ have density $f(\mathbf{z})$, to which we shall apply the asymptotic

expansion. Define the exponential family

$$f_\theta(\mathbf{z}) = f_\theta(z, \dot{z}, m) = \exp\left\{\theta\mathbf{z} - K(\theta)\right\}f(z, \dot{z}, m), \qquad (2.2)$$

where $\theta = (\theta_0, \ldots, \theta_{d_k-1}) \in \mathbb{R}^{d_k}$. The function $K(\cdot)$ is of course the cumulant generating function (cgf) of $\mathbf{Z}$,

$$K(\theta) = \sum_{\nu=1}^{\infty} \sum_{\substack{\gamma \in \mathbb{N}^{d_k} \\ |\gamma| = \nu}} \frac{1}{\gamma!} k_\nu\left([\mathbf{Z}]^\gamma\right)\theta^\gamma, \qquad (2.3)$$

where if $\gamma \in \mathbb{N}^d$ we define $[Y]^\gamma = [\overbrace{Y_1, \ldots, Y_1}^{\gamma_1}, \ldots, \overbrace{Y_d, \ldots, Y_d}^{\gamma_d}]$, $(\theta_1, \ldots, \theta_d)^\gamma = \theta_1^{\gamma_1} \cdots \theta_d^{\gamma_d}$, $|\gamma| = \sum_{i=1}^d \gamma_i$, and $\gamma! = \prod_{i=1}^d \gamma_i!$. We have written k_ν for a νth order cumulant.

Tilting in the first argument and inverting (2.2) yields

$$f(z, \dot{z}, m) = \exp\left\{K(\theta_0, 0') - \theta_0 z\right\}f_{(\theta_0, 0')}(z, \dot{z}, m). \qquad (2.4)$$

The relationship (2.4) holds for all values of θ_0, in particular that value $\theta_0 = \hat{\theta}_u$ which makes the Edgeworth expansion to $f_{(\theta_0, 0')}(\mathbf{z})$ most accurate at the point $\mathbf{z} = (u, 0', 0')'$— namely that for which

$$u = \mathbb{E}_{(\hat{\theta}_u, 0')}Z = \left.\frac{\partial K(\theta_0, 0')}{\partial \theta_0}\right|_{\theta_0 = \hat{\theta}_u} \qquad (2.5)$$

holds. It is trivial to see that this requirement is satisfied if and only if $\hat{\theta}_u$ is the formal maximum likelihood estimate of θ_0 under the model $f_{(\theta_0, 0')}(\mathbf{z})$ when $Z = u$ is observed.

Finally, we let $\mu_/$ and $\Sigma_/$ denote the mean and covariance matrix under the density $f_{(\hat{\theta}_u, 0')}(\mathbf{z})$. Then (c.f. [5], p. 187) we have

$$\mu_/ = \begin{pmatrix} u \\ \nabla_{2 \cdots d_k} K(\theta)|_{\theta = (\hat{\theta}_u, 0')} \end{pmatrix} \quad \text{and} \quad \Sigma_/ = \nabla\nabla' K(\theta)|_{\theta = (\hat{\theta}_u, 0')},$$

where $\nabla_{2\cdots d_k}$ has been used to denote $(\partial/\partial\theta_2, \ldots, \partial/\partial\theta_{d_k})'$. The approximation to be exploited will eventually take the form

$$f(z, \dot{z}, m) = \exp\left\{K(\hat{\theta}_u, 0') - \hat{\theta}_u z\right\} \phi_{d_k}\left((z, \dot{z}', m')'; \mu_/, \Sigma_/\right)$$

$$\times \left\{1 + Q_{k,3}\left((z, \dot{z}', m')' - \mu_/\right)\right\} + e_{k,n}, \tag{2.6}$$

$$\triangleq \widehat{f}_k(z, \dot{z}, m)\left\{1 + Q_{k,3}\left((z, \dot{z}', m')' - \mu_/\right)\right\} + e_{k,n},$$

where $Q_{k,3}$ is a degree 3 polynomial containing mixed cumulants of $\left(Z, \dot{Z}|_k, M\right)$ of order 2 and 3, while $e_{k,n}$ denotes terms of smaller asymptotic order. It turns out that $Q_{k,3}$ has coefficients (including constant term) which range up to $O\left(n^{-1/2}\right)$. Expansion (2.6) is called 'mixed tilted-direct' because exponential tilting is applied to the first argument but not to the others.

3   Preliminary calculations

We argued heuristically that the expected EC will provide a good approximation to the p-value (1.2) for large thresholds u; we therefore allow the possibility that $u = u_n \to \infty$ as $n \to \infty$. Let $u = O\left(n^{1/2-\alpha}\right)$ for some $0 \le \alpha \le 1/2$. From (2.5) we have

$$u_n = K_Z'(\hat{\theta}_u)$$

$$= \sigma^2\hat{\theta}_u + \frac{1}{2}k_3(Z)\hat{\theta}_u^2 + \frac{1}{3!}k_4(Z)\hat{\theta}_u^3 + \cdots, \tag{3.1}$$

where $K_Z(\cdot)$ denotes the marginal cgf of Z. When θ is a univariate argument, we shall use ' $'$ ' or a superscript (e.g. ' $^{(3)}$ ') to denote (multiple) differentiation with respect to θ. Under some additional assumptions (3.1) is a proper asymptotic expansion, in the sense

that the error in the remainder is of smaller order than the previous term. We shall use the following lemma repeatedly, and often implicitly, in proving our main results.

Lemma 2. *Let $Z = n^{-1/2} \sum_i W_i$ with W_i centred, i.i.d. random variables with variance σ^2 such that $\mathbb{E}|W_i|^{s+1} < \infty$ for some $s \geq 1$. Suppose that $K_W^{(s+1)}$ is continuous on a closed interval $[-\epsilon_s, \epsilon_s]$ with $\epsilon_s > 0$. Let $|\theta| \leq cn^{1/2-\beta}$ for some finite c and $\beta > 0$. Then there exists a constant $c(s)$ depending only on s such that for all large n,*

$$\left| K_Z'(\theta) - \sigma^2\theta - k_3(Z)\frac{\theta^2}{2} - \cdots - k_s(Z)\frac{\theta^{s-1}}{(s-1)!} \right| \leq c(s)n^{1/2-s\beta}. \tag{3.2}$$

In particular, if $u \sim cn^{1/2-\alpha}$ for some $0 < \alpha \leq 1/2$ and $c > 0$, with Z be as described above and $s = 2$, then

$$\hat{\theta}_u = \hat{\theta}_{u,n} = \frac{u}{\sigma^2}\left[1 + O\left(n^{-\alpha}\right)\right] \tag{3.3}$$

as $n \to \infty$. If $\alpha = 1/2$ then u is (asymptotically) constant and $\hat{\theta}_u = u/\sigma^2 + O\left(n^{-1/2}\right)$.

3.1   The tilted moments

Let us now be more explicit about the 'tilted' parameters $\mu_/$ and $\Sigma_/$, computed by differentiating (2.3) with respect to the appropriate argument(s) and then setting $\theta = (\hat{\theta}_u, 0')$. To simplify notation we partition as follows:

$$\mu_/ = (u, \dot{\mu}', \ddot{\mu}')'; \qquad \Sigma_/ = \begin{pmatrix} \tau^2 & \dot{\sigma}' & \ddot{\sigma}' \\ \dot{\sigma} & \dot{\Sigma} & \ddot{\Sigma}' \\ \ddot{\sigma} & \ddot{\Sigma} & \ddot{\Sigma} \end{pmatrix}.$$

We alert the reader that the dot notation in the above parameters does not indicate differentiation, but was designed to emphasize that portion of the process $Y(t)$ with

12

which a particular mean or covariance is associated. In $\Sigma_/$, the diagonal entries in the partition are just the tilted variances, respectively, of Z, $\dot{Z}|_k$ and M. The off-diagonal entries are the corresponding covariances, and so $\dot{\sigma}$ is $k \times 1$, $\ddot{\sigma}$ is $k(k-1)/2 \times 1$, and $\dddot{\Sigma}$ is $k(k-1)/2 \times k$. We have, for example,

$$\dot{\mu}_i = \sum_{\nu=3}^{\infty} \frac{1}{(\nu-1)!} k_\nu(Z, \ldots, Z, Z_i) \hat{\theta}_u^{\nu-1}, \tag{3.4}$$

$$\dddot{\Sigma}_{i,jl} = \sum_{\nu=3}^{\infty} \frac{1}{(\nu-2)!} k_\nu(Z, \ldots, Z, Z_i, M_{jl}) \hat{\theta}_u^{\nu-2},$$

and many more such expressions. When $\alpha > 1/4$ it is easily deduced that:

$$\dot{\mu}_i = O\left(n^{1/2-2\alpha}\right), \quad \ddot{\mu}_{ij} = O\left(n^{1/2-2\alpha}\right);$$

$$\tau_u^2 \overset{\Delta}{=} \tau^2 = \sigma^2 + O\left(n^{-\alpha}\right), \quad \dot{\sigma}_i = O\left(n^{-\alpha}\right), \quad \ddot{\sigma}_{ij} = O\left(n^{-\alpha}\right); \tag{3.5}$$

$$\dot{\Sigma}_{ij} = \delta_{ij} + O\left(n^{-\alpha}\right), \quad \ddot{\Sigma}_{ij,lm} = \lambda_{ijlm}^M + O\left(n^{-\alpha}\right), \quad \dddot{\Sigma}_{i,jl} = O\left(n^{-\alpha}\right).$$

3.2 Some tilted conditional moments

We shall also need the following conditional moments (assuming $\alpha > 1/4$) under the tilted Normal density $\phi_k(\,\cdot\,) \overset{\Delta}{=} \phi_{d_k}(\,\cdot\,; \mu_/, \Sigma_/)$, the proofs of which are straightforward:

$$\nu \overset{\Delta}{=} \mathbb{E}_{\phi_k}\left\{ Z_k \,\middle|\, Z = u, \dot{Z}|_{k-1} = 0 \right\} = O\left(n^{1/2-2\alpha}\right); \tag{3.6}$$

$$\eta^2 \overset{\Delta}{=} \mathsf{Var}_{\phi_k}\left\{ Z_k \,\middle|\, Z = u, \dot{Z}|_{k-1} = 0 \right\} = 1 + O\left(n^{-\alpha}\right); \tag{3.7}$$

$$\mathbb{E}_{\phi_k}\left\{ M_{ij} \,\middle|\, Z = u, \dot{Z}|_{k-1} = 0, Z_k = x \right\} = \ddot{\mu}_{ij} + c_{ij}x + O\left(n^{1/2-3\alpha}\right); \tag{3.8}$$

$$\mathbb{E}_{\phi_k}\left\{ M_{ij}M_{lm} \,\middle|\, Z = u, \dot{Z}|_{k-1} = 0, Z_k = x \right\} = \lambda_{ijlm}^M + c_{ijlm}^0 + c_{ijlm}^1 x + c_{ijlm}^2 x^2; \tag{3.9}$$

where $c_{ij} = O\left(n^{-\alpha}\right)$ and

$$
\begin{aligned}
c_{ijlm}^0 &= O\left(n^{\max(-\alpha, 1-4\alpha)}\right), \\
c_{ijlm}^1 &= O\left(n^{1/2-3\alpha}\right), \\
c_{ijlm}^2 &= O\left(n^{-2\alpha}\right).
\end{aligned}
\tag{3.10}
$$

The punchline to (3.8) and (3.9) is that under the Gaussian approximation to the conjugate density $f_{(\hat{\theta}_u, 0')}$, and conditional on the event $\left\{Z = u, \dot{Z}|_{k-1} = 0, Z_k = x\right\}$, we have $M \overset{\mathscr{D}}{=} \Delta + \Xi_x$, where Δ is a centred Gaussian matrix with covariances satisfying (4.1) below, and Ξ_x is a small, non-random (given x) perturbation.

4 Euler characteristic densities

We shall now proceed to use the characterization (1.5) in order to derive our corrected EC density.

4.1 The case $\alpha > \frac{1}{4}$

The following analogue of AT's Lemma 11.5.2 can be proved in a similar fashion if one is mindful of the various error terms. In what follows, $|\cdot|$ is the determinant. We also add a subscript n to all EC density expressions, since we are interested in their asymptotic behaviour.

Lemma 3. *Let $\frac{1}{4} < \alpha \leq \frac{1}{2}$, let $0 \leq x$ be a $O(1)$ non-random scalar and ε a $O(1)$ symmetric function from $\{1, \ldots, k\}^4$ into $\mathbb{R}$. Suppose that Δ is a centred, (jointly) Gaussian*

$k \times k$ matrix whose covariances satisfy

$$\mathbb{E}\left(\Delta_{ij}\Delta_{lm}\right) = \varepsilon(i,j,l,m) - \delta_{ij}\delta_{lm} + s_{ijlm}(x) \tag{4.1}$$

where $s_{ijlm}(x)$ is a quadratic polynomial with coefficients as in (3.10). Let $b = O\left(n^{1/2-\alpha}\right)$ be a non-random scalar, and let $\Xi^0 = O\left(n^{1/2-2\alpha}\right)$, $\Xi^1 = O\left(n^{-\alpha}\right)$, and $\Xi_x = \Xi^0 + x\,\Xi^1$ be non-random matrices. Then

$$\mathbb{E}|\Delta - bI + \Xi_x| = (-1)^k H_k(b)\left\{1 + r_k(x)\right\}, \tag{4.2}$$

where r_k is a degree k polynomial whose largest (constant) coefficient is $a_0 = O\left(n^{-\alpha}\right)$ as $n \to \infty$.

Theorem 4. *Let Z be a strictly stationary CLRF as in (1.1), with covariance function belonging to $C^4(\mathbb{R}^N)$ and $\mathbb{E}[\dot{Z}\dot{Z}'] = I$. Let $u \sim cn^{1/2-\alpha}$ as $n \to \infty$ for some constants $0 < c < \infty$ and $\frac{1}{4} < \alpha \le \frac{1}{2}$, and suppose that $\hat{\theta}_u$ is a solution to*

$$K_Z'(\theta) = u, \tag{4.3}$$

where $K_Z(\cdot)$ is the cgf of $Z(t)$. Let $\tau_u^2 = K_Z''(\hat{\theta}_u)$ be the tilted variance of $Z(t)$ at $\hat{\theta}_u$, and $I_u(\theta) = u\theta - K_Z(\theta)$. Then subject to regularity conditions, the EC densities of Z for $k \ge 1$ are given by

$$\rho_{k,n}(u) = \widehat{\rho}_{k,n}(u) \times \left\{1 + \epsilon_n\right\}, \tag{4.4}$$

where

$$\widehat{\rho}_{k,n}(u) = (2\pi)^{-(k+1)/2}\tau_u^{-k}\exp\left\{-I_u(\hat{\theta}_u)\right\} H_{k-1}(\tau_u\hat{\theta}_u) \tag{4.5}$$

and as $n \to \infty$ the relative error satisfies $\epsilon_n = O\left(n^{1/2-2\alpha}\right)$.

Moreover, an approximation with relative error $O\left(n^{\max(-\alpha,1-4\alpha)}\right)$ is given by

$$\widehat{\rho}_{k,n}(u) \times \left(1 + \mathbb{E}[Z^2 Z_k]\sqrt{\frac{\pi}{2}}\frac{\hat{\theta}_u^2}{2}\right), \tag{4.6}$$

and (if $k \geq 2$) one with relative error $O\left(n^{\max\{-\alpha,3/2-6\alpha\}}\right)$ by

$$\widehat{\rho}_{k,n}(u) \times \left(1 + \mathbb{E}[Z^2 Z_k]\sqrt{\frac{\pi}{2}}\frac{\hat{\theta}_u^2}{2}\right)\exp\left\{-\frac{\hat{\theta}_u^4}{8}\sum_{i=1}^{k-1}\mathbb{E}[Z^2 Z_i]^2\right\}. \tag{4.7}$$

Remark The 'regularity conditions' mentioned above are just those on the random field $(Z, \dot{Z}|_k', \ddot{Z}|_{k-1}')$ required to apply the Expectation Meta-Theorem (Thm 11.1.1 in AT), those on the density $f(z, \dot{z}, m)$ assuring the correctness of the Edgeworth expansion, plus a mild assumption about the moments of $(W, \dot{W}, \ddot{W})$. These conditions are academic in that they can rarely be checked in practice, but are considerably weaker than the Gaussian assumption, which is currently the only alternative. See the Appendix for further details.

Proof (Leading term). For extra details the reader is referred to the Appendix, where we rigourously bound the error terms and consider the contribution to the integral (1.5) of the polynomial $Q_{k,3}(\cdot)$. It is seen there that the additional terms are of relative size at most $O\left(n^{-\alpha}\right)$, and are therefore negligible. The crux of the derivation, however, is in working out the leading term, which we do now. Plugging the saddlepoint approximation into (1.5) yields

$$\rho_{k,n}(u) \doteq \int_{x=0}^{\infty}\int_{m\in\mathbb{R}^{D_{k-1}}} x\det\left(-\frac{m - u/\sigma I}{\sigma}\right)\widehat{f}_k(u, 0', x, m)\, dm\, dx \tag{4.8}$$

$$= \exp\left\{-I_u(\hat{\theta}_u)\right\}\left(\frac{-1}{\sigma}\right)^{k-1}\int_{x=0}^{\infty} x\phi_k(x|u, 0')$$

$$\times \int_{\mathbb{R}^{D_{k-1}}}\det(m - u/\sigma I)\,\phi_k(m|u, 0', x)\, dm\, dx\,\phi_{0,k-1}(u, 0'), \tag{4.9}$$

where $D_j = j(j+1)/2$ and the various ϕ's denote the obvious Gaussian densities derived from $\phi_{d_k}(\cdot; \mu_/, \Sigma_/)$ by conditioning or marginalizing. Evaluating the inner expectation via (3.8)-(3.9) and Lemma 3, after noting that $\tau_u \hat{\theta}_u - u/\sigma = O(n^{-\alpha})$ has smaller order than Ξ^0, gives

$$\exp\left\{-I_u(\hat{\theta}_u)\right\} \sigma^{-(k-1)} H_{k-1}(\tau_u \hat{\theta}_u) \mathbb{E}\left\{Y(1 + r_{k-1}(Y)) \mathbb{1}_{[Y>0]}\right\} \phi_{0,k-1}(u, 0'), \qquad (4.10)$$

where $Y \sim \mathcal{N}(\nu, \eta^2)$ as in (3.6) and (3.7). Recall that $\sigma = \tau_u(1 + O(n^{-\alpha}))$, and consider the last factor in (4.10). We have

$$\phi_{0,k-1}(u, 0') = (2\pi)^{-k/2} \det(V_k)^{-1/2} \exp\left\{-\frac{1}{2}\dot{\mu}|'_{k-1}\dot{\Sigma}|^{-1}_{k-1}\dot{\mu}|_{k-1}\right\}, \qquad (4.11)$$

where

$$V_k \triangleq \begin{pmatrix} \tau_u^2 & \dot{\sigma}|'_{k-1} \\ \dot{\sigma}|_{k-1} & \dot{\Sigma}|_{k-1} \end{pmatrix}. \qquad (4.12)$$

So $\det(V_k) = \tau_u^2(1 + O(n^{-\alpha}))$, and the exponential portion of (4.11) behaves like

$$\exp\left\{O\left(n^{1/2-2\alpha}\right)' \left[I + O\left(n^{-\alpha}\right)\right] O\left(n^{1/2-2\alpha}\right)\right\} = \exp\left\{O\left(n^{1-4\alpha}\right)\right\} = 1 + O\left(n^{1-4\alpha}\right).$$

It is also standard fare (e.g. [1], Lemma 5.3.3) that

$$\begin{aligned}
\mathbb{E}\left\{Y\mathbb{1}_{[Y>0]}\right\} &= \frac{\eta}{\sqrt{2\pi}} \exp\left(-\frac{\nu^2}{2\eta^2}\right) + \nu\Phi\left(\frac{\nu}{\eta}\right) \\
&= \frac{\eta}{\sqrt{2\pi}}\left[1 - \frac{\nu^2}{2\eta^2} + O\left(n^{2-8\alpha}\right)\right] + \frac{\nu}{2} + \frac{\nu^2}{\sqrt{2\pi}\eta} + O\left(n^{3/2-6\alpha}\right) \\
&= \frac{1}{\sqrt{2\pi}} + \frac{\nu}{2} + O\left(n^{\max(-\alpha, 3/2-6\alpha)}\right) \qquad (4.13) \\
&= (2\pi)^{-1/2} + O\left(n^{1/2-2\alpha}\right). \qquad (4.14)
\end{aligned}$$

Let $p_k(y) = yr_{k-1}(y)$ and $Y_0 = (Y - \nu)/\eta$. Write $p_k(y) = \sum_{l=1}^{k} a_{l-1}y^l$, with the a_i's at most $O(n^{-\alpha})$. Then

$$
\begin{aligned}
p_k(Y) &= \sum_{l=1}^{k} a_{l-1}(\eta Y_0 + \nu)^l \\
&= \sum_{j=0}^{k} \left[\sum_{l=\max(1,j)}^{k} a_{l-1} \binom{l}{j} \eta^j \nu^{l-j} \right] Y_0^j \\
&= \sum_{j=0}^{k} c_j Y_0^j,
\end{aligned} \tag{4.15}
$$

where the largest coefficient c_j in (4.15) can be loosely bounded by $O(n^{-\alpha})$. The absolute moments of Y_0 are all finite, so that

$$
\left| \mathbb{E} p_k(Y) \mathbb{1}_{[Y>0]} \right| \leq \sum_{j=0}^{k} |c_j| \mathbb{E} |Y_0|^j = O(n^{-\alpha}),
$$

whereby

$$
\mathbb{E} \left\{ Y(1 + r_{k-1}(Y)) \mathbb{1}_{[Y>0]} \right\} = (2\pi)^{-1/2} + O(n^{1/2-2\alpha}).
$$

Putting the pieces together, (4.10) simplifies to

$$
e^{-I_u(\hat{\theta}_u)} \tau_u^{-k} (2\pi)^{-(k+1)/2} H_{k-1}(\tau_u \hat{\theta}_u) \left(1 + O(n^{1/2-2\alpha})\right) \left(1 + O(n^{-\alpha})\right) \left(1 + O(n^{1-4\alpha})\right)
$$
$$
= \widehat{\rho}_{k,n}(u) \times \left(1 + O(n^{1/2-2\alpha})\right).
$$

Towards (4.6), consider one additional term in (4.13)-(4.14):

$$
\begin{aligned}
(2\pi)^{-1/2} + \nu/2 + O\left(n^{\max\{-\alpha,3/2-6\alpha\}}\right) &= (2\pi)^{-1/2} + \dot{\mu}_k/2 + O\left(n^{\max\{-\alpha,3/2-6\alpha\}}\right) \\
&= (2\pi)^{-1/2} + k_3(Z,Z,Z_k)\hat{\theta}_u^2/4 + O\left(n^{\max\{-\alpha,3/2-6\alpha\}}\right) \\
&= (2\pi)^{-1/2} \left\{ 1 + \mathbb{E}[Z^2 Z_k] \sqrt{\frac{\pi}{2}} \frac{\hat{\theta}_u^2}{2} \right\} \left(1 + O\left(n^{\max\{-\alpha,3/2-6\alpha\}}\right)\right).
\end{aligned}
$$

As for (4.7), we have

$$-\frac{1}{2}\dot{\mu}|'_{k-1}\dot{\Sigma}|^{-1}_{k-1}\dot{\mu}|_{k-1} = -\frac{1}{2}\dot{\mu}|'_{k-1}\left(I + O\left(n^{-\alpha}\right)\right)\dot{\mu}|_{k-1}$$

$$= -\frac{1}{2}\sum_{i\leq k-1}\left(\mathbb{E}[Z^2 Z_i]\hat{\theta}_u^2/2 + O\left(n^{1/2-3\alpha}\right)\right)^2 + \dot{\mu}|'_{k-1}O\left(n^{-\alpha}\right)\dot{\mu}|_{k-1}$$

$$= -\frac{\hat{\theta}_u^4}{8}\sum_{i\leq k-1}\mathbb{E}[Z^2 Z_i]^2 + O\left(n^{1-5\alpha}\right).$$

Finally, note that

$$\max\{1/2 - 3\alpha, 1 - 5\alpha\} < -\alpha \leq \max\{-\alpha, 3/2 - 6\alpha\} \leq \max\{-\alpha, 1 - 4\alpha\}.$$

$\square$

It is worth emphasizing that if the *mixed* (third-order) cumulants $\mathbb{E}[Z^2 Z_i]$ of Z with its first derivatives vanish, then the approximation is improved regardless of the behaviour of the *marginal* cumulants of Z. Note that if $1/3 \leq \alpha \leq 1/2$ then (4.7) is no better than (4.6), since $-\alpha \geq 1 - 4\alpha$. An important special case of Theorem 4 is when $\alpha = 1/2$, so that u is a constant. In that 'small-threshold' setting each of the approximations (4.5)-(4.7) admits a relative error of the order $n^{-1/2}$.

4.2 What about $\alpha \leq \frac{1}{4}$?

Consider the case $\alpha \leq \frac{1}{4}$, in which the threshold u grows somewhat more quickly relative to n. Here we expect the EC heuristic to perform well, and the Gaussian approximation to perform poorly. But what of the tilting procedure? From (3.5) we have that when $\alpha < 1/4$, the tilted mean derivatives also grow with n: $\dot{\mu}_i$, $\ddot{\mu}_{ij} \to \infty$. Lemma 3 no longer holds in this scenario since the 'remainder' terms now dominate (4.1) and thus

19

the special covariance structure which leads to Hermite polynomials is lost (the problem is exacerbated if $\alpha < 0$).

A reasonable conjecture in this setting is that the tilted EC densities will provide a much better approximation to the true ones than will those derived from a Gaussian assumption (see the next section). Less obvious is whether the tilted densities themselves have an acceptable relative error—indeed, Theorem 4 seems to suggest the contrary. While we have yet to work out the details in general, empirical evidence in at least two cases ((see Sections 6 and 7) supports using the tilted densities.

4.3   Brief comment on Rabinowitz & Siegmund

In the language and notation used here, the Rabinowitz-Siegmund heuristic can be phrased

$$\mathbb{P}\left\{\sup_t Z(t) \geq u\right\} \approx 1 - \exp\left\{-\mathbb{E}\left\{M_u(Z)\right\}\right\} \approx \mathbb{E}\left\{M_u(Z)\right\}, \qquad (4.16)$$

where $M_u(Z)$ is the number of local maxima of the field Z above u. In RS, the authors approximated this expectation by

$$\mathbb{E}\left\{M_u(Z)\right\} \doteq \exp\left\{-I_u(\hat{\theta}_u)\right\} \frac{|T||\widehat{\Lambda}|}{(2\pi)^{(k+1)/2}\tau_u}\hat{\theta}_u^{N-1}, \qquad (4.17)$$

where $\widehat{\Lambda} = \mathsf{Var}_{\hat{\theta}_u}[\dot{Z}]$ is the tilted variance of the full derivative process (matrix of second spectral moments), having removed the local isotropy condition. Note that $\hat{\theta}_u^{N-1}/\tau_u$ is the leading term in the polynomial $\tau_u^{-N}H_{N-1}(\tau_u\hat{\theta}_u)$, so that (modulo the spectral determinant) (4.17) is equivalent to the leading term of our approximation.   Naïvely

20

extending (4.17) for, say, a stationary process on a rectangle, one might obtain

$$\mathbb{E}\varphi(A_u) \doteq \sum_{k=0}^{N} \sum_{J \in \mathcal{O}_k} |J|_k |\widehat{\Lambda}_J|^{1/2} \widehat{\rho}_{k,n}(u), \qquad (4.18)$$

where now $\widehat{\Lambda}_J$ denotes the tilted variance of the derivatives restricted to facet J of the rectangle, $\mathcal{O}_k$ is the set of k-dimensional facets touching the origin (c.f. AT, Chapter 11), and $|\cdot|_k$ is k-dimensional Lebesgue measure. (See Section 5.2 for a definition of $\widehat{\rho}_{0,n}$.) For a general isotropic process, where $\widehat{\Lambda} = \widehat{\lambda}I$, (4.18) would become

$$\mathbb{E}\varphi(A_u) \doteq \sum_{k=0}^{N} \widehat{\lambda}^{k/2} \mathcal{V}_k(T) \widehat{\rho}_{k,n}(u). \qquad (4.19)$$

It is readily seen that the relative error between the approximations (4.18)-(4.19) and our own versions (with the untilted λ or Λ_J replacing $\widehat{\lambda}$ or $\widehat{\Lambda}_J$) is of order $O\left(n^{-\alpha}\right)$. Indeed, $\lambda_{jl} = \mathbb{E}\left\{Z_j(t)Z_l(t)\right\} = O(1)$, and

$$\widehat{\lambda}_{jl} = \sum_{\nu=2}^{\infty} k_\nu(Z, \ldots, Z, Z_j, Z_l) \frac{\widehat{\theta}_u^{\nu-2}}{(\nu-2)!} = \lambda_{jl}\left(1 + O\left(n^{-\alpha}\right)\right). \qquad (4.20)$$

Hence $|\widehat{\Lambda}|^{1/2} = |\Lambda|^{1/2}\left(1 + O\left(n^{-\alpha}\right)\right)$, and similarly for $\widehat{\Lambda}_J$. This implies a $O\left(n^{-\alpha}\right)$ discrepancy for each $k \geq 1$ term in (4.18) and (4.19). Since $-\alpha \leq 1/2 - 2\alpha$, it is unclear from the theory whether one should use the tilted or untilted second spectral moment(s) in computing $\mathbb{E}\varphi(A_u)$. Similarly, it is seen in Section 5 and the proof of Theorem 4 that using the Gaussian argument u/σ in the place of $\tau_u\widehat{\theta}_u$ in the Hermite polynomials leaves the relative error unchanged from (4.4). Intuition suggests that $\tau_u\widehat{\theta}_u$ may give more accurate results in finite samples. A geometric argument also supports its use; see [13]. The $k = 0$ term is unaffected by these changes, and can be improved by the methods to be discussed in Section 5.2.

21

4.4 Some remarks about validity

It has been shown that for a Gaussian process the relative error of the expected EC as an approximation to the p-value is exponential as $u \to \infty$, in the sense that the first term of their difference is *exponentially smaller* than the zeroth term in expansion (1.3), which behaves like $u^{-1}e^{-u^2/(2\sigma^2)}$ [37]. The authors of [38] (and an anonymous reviewer) rightly pointed out that this result is, for the time being, a purely Gaussian one: the exponential error likely does not hold even for a t random field. Nevertheless, much of the development in [37] makes no use of Gaussianity. We argue here that for CLRF's, the expected EC is exponentially close to the expected number of local maxima above u, which is obviously an upper bound for the p-value.

Assuming that the parameter space is a manifold without boundary, one can write the difference (before taking expectations) as

$$|\varphi(A_u) - M_u| \leq \# \left\{ t \in T : \ Z(t) \geq u, \ \dot{Z}(t) = 0, \ \ddot{Z} \not\prec 0 \right\} \tag{4.21}$$

$$= \# \left\{ t \in T : \ Z(t) \geq u, \ \dot{Z}(t) = 0, \ \left(\ddot{Z}(t) + Z(t)I \right) \not\prec Z(t)I \right\} \tag{4.22}$$

$$\leq \# \left\{ t \in T : \ Z(t) \geq u, \ \dot{Z}(t) = 0, \ \zeta_1 \left(\ddot{Z}(t) + Z(t)I \right) > u \right\}, \tag{4.23}$$

where $\zeta_1(\cdot)$ extracts the largest eigenvalue. We have again assumed $\mathsf{Var}\dot{Z}(t) = I$, purely for convenience. It is not hard to argue (for example using AT's Expectation Meta-Theorem in combination with an Edgeworth expansion) that for a stationary CLRF the expectation of (4.23) behaves to first order like

$$|T|e^{-u^2/(2\sigma^2)}e^{-u^2/(2\sigma_c^2)},$$

where $\sigma_c^2 \triangleq \sup_{\|v\|=1} \mathsf{Var}\left[v'(\ddot{Z} + ZI)v \right]$ is the so-called critical variance. The key fact is

that $\ddot{Z} + ZI$, $\dot{Z}$ and Z are uncorrelated, and indeed *almost* independent. When T has boundary, (4.21) must account for points where A_u intersects ∂T; the argument is more finicky but not fundamentally different. In the interest of brevity we leave the details to a future article.

Since $p \leq \mathbb{E}M_u \approx \mathbb{E}\varphi(A_u)$, clearly using the expected EC is conservative (at least asymptotically). Determining conditions under which the p-value approximation is *sharp* will require modification of the purely Gaussian portions of [37]. Numerical evidence from a simulation (see Section 7) is reassuring.

5   Comparison with the Gaussian approximation

We shall use the notation

$$\epsilon_n[c_n, d_n] = \beta \quad \Leftrightarrow \quad c_n = d_n \times \left(1 + O(n^\beta)\right) \tag{5.24}$$

as $n \to \infty$ to describe relative error for $\beta \leq 0$. We say that the error is *sharp*, and write $\epsilon_n[c_n, d_n] \overset{s}{=} \beta$, if (5.24) holds, but fails if β is replaced by $\beta - \epsilon$ for any $\epsilon > 0$. It is important to determine whether there are values of α for which our derived expression provides a better approximation to $\rho_{k,n}(u)$ than does the Gaussian EC density. We examine only the approximation $\widehat{\rho}_{k,n}(u)$; similar considerations for the refined formulae (4.6) and (4.7) are straightforward and left to the reader.

5.1 EC densities with $k \geq 1$

In Theorem 4 we derived a 'tilted' order $O\left(n^{1/2-2\alpha}\right)$ (in relative error) approximation to $\rho_{k,n}(u)$ for $k \geq 1$, denoted by $\widehat{\rho}_{k,n}(u)$ and given in (4.5). Comparing this approximation to the Gaussian EC density $\rho_k^\gamma = \rho_{k,n}^\gamma$ from (1.6), we have (under the regularity conditions of Theorem 4):

Corollary 5. *Let* $u \sim cn^{1/2-\alpha}$ *for some* $0 < c < \infty$. *If* $\frac{1}{3} \leq \alpha \leq \frac{1}{2}$ *and* $k_3(Z) \neq 0$ *then for all* $k \geq 1$,

$$\epsilon_n[\rho_{k,n}^\gamma(u), \rho_{k,n}(u)] \overset{s}{=} 1 - 3\alpha. \tag{5.25}$$

In particular,

$$\epsilon_n[\rho_{k,n}^\gamma(u), \rho_{k,n}(u)] - \epsilon_n[\widehat{\rho}_{k,n}(u), \rho_{k,n}(u)] \geq \frac{1}{2} - \alpha \geq 0. \tag{5.26}$$

Moreover, if $\frac{1}{4} < \alpha < \frac{1}{3}$ *and* $k_3(Z) \neq 0$ *then the error of the Gaussian densities is exponential, in the sense that*

$$\rho_{k,n}^\gamma(u) = \rho_{k,n}(u) \times \exp(b_n) \tag{5.27}$$

for $k \geq 1$ *and some sequence* b_n *with* $|b_n| \to \infty$.

Remark In the proof of Corollary 5 (see Appendix A.5), it is easily seen that the sequence referred to in (5.27) can be written $b_n = k_3(Z)\hat{\theta}_u^3/6 + O\left(n^{1-4\alpha}\right)$, where the skewness may be positive or negative.

5.2 The $k = 0$ case

Until now we have only considered EC densities for $k \geq 1$. The zeroth density is generally defined by

$$\rho_{0,n}(u) \triangleq \mathbb{P}\left\{Z(t) \geq u\right\}, \tag{5.28}$$

the univariate tail probability above the level u. Since nothing in the derivations of Theorem 4 or Corollary 5 explicitly requires $u > 0$, the results mentioned thus far are valid if we weaken the threshold condition to $u \sim cn^{1/2-\alpha}$, $|c| < \infty$. (Their use is limited, however, by the facts that (a) when $|u|$ is small the Gaussian approximation is adequate and (b) when u is large and negative the zeroth density dwarfs the others.) With the exception of the Lugananni-Rice formula (5.32), the discussion in this subsection *does* assume $u \geq 0$, since it pertains to saddlepoint-type approximations to (5.28). If one wishes to approximate $\rho_{0,n}(u)$ for $u < 0$ (as is the case in Figure 1), then the expressions can be modified by making use of the symmetry in their derivations and of the fact that $\rho_{0,n}(u) = 1 - \mathbb{P}\left\{Z(t) < u\right\}$. Such generalizations are left to the reader. It should be noted that when a p-value approximation is sought and hence u is large, the error in $\rho_{0,n}(u)$ typically has a negligible impact compared to those in the higher EC densities. Approximating tail probabilities via saddlepoint methods is far from a new problem; the principal purpose of this subsection is merely to enumerate some useful formulae.

Let $\Psi(\cdot) \triangleq 1 - \Phi(\cdot)$ denote the standard Normal survival function. Given Theorem 4, one who is inspired by AT's definition $H_{-1}(x) = \sqrt{2\pi}\Psi(x)e^{x^2/2}$ might naïvely estimate (5.28) by

$$\begin{aligned}
\widehat{\rho}_{0,n}(u) &\triangleq \exp\left\{-I_u(\hat{\theta}_u)\right\}(2\pi)^{-1/2}H_{-1}(\tau_u\hat{\theta}_u) \\
&= \exp\left\{\tau_u^2\hat{\theta}_u^2/2 - I_u(\hat{\theta}_u)\right\}\Psi(\tau_u\hat{\theta}_u).
\end{aligned}$$

Perhaps surprisingly, this is what Robinson [30] called the 'first' saddlepoint approximation, and turns out to be a reasonable estimate if $u > 0$. The zeroth Gaussian density is

$$\rho_{0,n}^{\gamma}(u) = \Psi\left(\frac{u}{\sigma}\right) = \frac{\sigma e^{-u^2/(2\sigma^2)}}{u\sqrt{2\pi}}\left(1 + O\left(n^{2\alpha-1}\right)\right).$$

Under typical Central Limit Theorem conditions (e.g. [17] Theorem 2.3), one has to first order

$$\rho_{0,n}(u) \doteq \rho_{0,n}^{\gamma}(u) + \frac{k_3(Z)}{6\sigma^3}\left(\frac{u^2}{\sigma^2} - 1\right)\phi(u/\sigma), \tag{5.29}$$

which implies (sharply, assuming non-zero skewness) that

$$\rho_{0,n}(u) = \rho_{0,n}^{\gamma}(u)\left(1 + O\left(n^{1-3\alpha}\right)\right).$$

Daniels [16] gave two differently-derived formulae. The first can be written

$$\rho_{0,n}(u) \doteq \exp\left\{\tau_u^2\hat{\theta}_u^2/2 - I_u(\hat{\theta}_u)\right\}$$
$$\times\left[\Psi(\tau_u\hat{\theta}_u)\left(1 - \frac{K_Z^{(3)}(\hat{\theta}_u)\tau_u^3\hat{\theta}_u^3}{6}\right) + \phi(\tau_u\hat{\theta}_u)\frac{K_Z^{(3)}(\hat{\theta}_u)}{6}\left(\tau_u^2\hat{\theta}_u^2 - 1\right)\right] \tag{5.30}$$

and corresponds to the so-called 'second' saddlepoint approximation [30]. Let $f_Z(\cdot)$ denote the marginal density of $Z(t)$. Upon using the large x estimate $\Psi(x) = \phi(x)/x \times \left(1 + O\left(x^{-2}\right)\right)$, and under the regularity condition (c.f. [30])

$$\sup_z\left|\Phi\left(\frac{z-u}{\tau_u}\right) - \int_{-\infty}^{z}e^{\hat{\theta}_u y - K_Z(\hat{\theta}_u)}f_Z(y)\,dy\right| = O\left(n^{-1/2}\right), \tag{5.31}$$

(5.30) leads to

$$\rho_{0,n}(u) = \widehat{\rho}_{0,n}(u)\left(1 + O\left(n^{-\alpha}\right)\right).$$

26

The second approach is the Lugannani-Rice formula [24]

$$\rho_{0,n}(u) = \Psi(\omega_u) + \phi(\omega_u)\left[\left(\frac{1}{\zeta_u} - \frac{1}{\omega_u}\right) + \frac{1}{\zeta_u}\left(\frac{K_Z^{(4)}(\hat{\theta}_u)}{8} - \frac{5K_Z^{(3)}(\hat{\theta}_u)^2}{24}\right)\right.$$
$$\left.- \frac{K_Z^{(3)}(\hat{\theta}_u)}{2\zeta_u^2} + \left(\frac{1}{\omega_u^3} - \frac{1}{\zeta_u^3}\right) + \cdots\right], \tag{5.32}$$

where $\omega_u = \text{sgn}[\hat{\theta}_u]\sqrt{2I_u(\hat{\theta}_u)}$ and $\zeta_u = \tau_u\hat{\theta}_u$. Note that this approximation is valid for large negative as well as positive u, and consequently the first two terms in (5.32) will be used in the astrophysics example of Figure 1. One can reduce (5.32) to

$$\rho_{0,n}(u) = \Psi(\omega_u)\left(1 + O\left(n^{-\alpha}\right)\right) = \Psi\left(\text{sgn}[\hat{\theta}_u]\sqrt{2I_u(\hat{\theta}_u)}\right)\left(1 + O\left(n^{-\alpha}\right)\right) \tag{5.33}$$
$$\triangleq \bar{\rho}_{0,n}(u)\left(1 + O\left(n^{-\alpha}\right)\right).$$

Filling in a few details gives the following:

Proposition 6. *Let $\frac{1}{4} < \alpha \leq \frac{1}{2}$ and suppose that (5.31) holds. Then*

$$\epsilon_n[\hat{\rho}_{0,n}(u), \rho_{0,n}(u)] = \epsilon_n[\bar{\rho}_{0,n}(u), \rho_{0,n}(u)] = -\alpha.$$

Also, if the field Z is as in (1.1), $k_3(Z) \neq 0$ and $\frac{1}{3} \leq \alpha \leq \frac{1}{2}$ then

$$\epsilon_n[\rho_{0,n}^{\gamma}(u), \rho_{0,n}(u)] \overset{s}{=} 1 - 3\alpha.$$

In particular,

$$\epsilon_n[\rho_{0,n}^{\gamma}(u), \rho_{0,n}(u)] > \epsilon_n[\hat{\rho}_{0,n}(u), \rho_{0,n}(u)] \quad \forall\, \alpha \in [1/3, 1/2]\,.$$

When $\alpha < 1/3$ the error in the zeroth Gaussian density diverges to infinity.

So once again the tilted densities perform better than the Gaussian one. However, $\bar{\rho}_{0,n}$ and $\widehat{\rho}_{0,n}$ can both be improved upon if one is prepared to numerically integrate the pointwise saddlepoint expansion of the density of z,

$$f(z) = \frac{1}{\sqrt{2\pi}\tau_z} \exp\left\{-I_z(\hat{\theta}_z)\right\} \left\{1 + \left[\frac{1}{8}\mu_4(z) - \frac{5}{24}\mu_3(z)^2\right] + \varepsilon_n(z)\right\}. \tag{5.34}$$

In (5.34) $\varepsilon_n(x)$ is a $O\left(n^{-2}\right)$ (for fixed z) function of z, and

$$\mu_3(z) = \tau_z^{-3} K_Z^{(3)}(\hat{\theta}_z); \qquad \mu_4(z) = \tau_z^{-4} K_Z^{(4)}(\hat{\theta}_z)$$

are the standardized skewness and kurtosis of Z under the tilted (conjugate) density. The square-bracketed term in (5.34) is of order $O\left(n^{-1}\right)$ for each fixed z. Hence

$$\rho_{0,n}(u) = \mathbb{P}\left\{Z(t) \geq u\right\}$$
$$\doteq \int_{z=u}^{\infty} \widehat{f}_n(z)\,dx + \int_{z=u}^{\infty} \widehat{f}_n(z)\left(\frac{1}{8}\mu_4(z) - \frac{5}{24}\mu_3(z)^2\right)\,dz, \tag{5.35}$$

where $\widehat{f}_n(z) \triangleq \exp\left\{-I_z(\hat{\theta}_z)\right\}\Big/\left[\sqrt{2\pi}\tau_z\right]$ is assumed integrable. So provided

$$\sup_{z \geq u}\left|\frac{1}{8}\mu_4(z) - \frac{5}{24}\mu_3(z)^2 + \varepsilon_n(z)\right| = o\left(n^{-\alpha}\right) \tag{5.36}$$

as $n \to \infty$, the first term in (5.35) will give an approximation with smaller relative error than $\widehat{\rho}_{k,n}(u)$ or $\bar{\rho}_{k,n}(u)$ (assuming numerical integration does not compound the error). The tradeoff is an increase in computation, although solving the saddlepoint equation everywhere can be avoided by a judicious change of variables ([17], Section 6.2). It is well known (*e.g.* [17], Remark 3.2) that gains can be achieved when $\widehat{f}_n(z)$ is normalized to become a density. Let $\widehat{\bar{f}}_n(z) = \widehat{f}_n(z)/\|\widehat{f}_n\|_1$; a good approximation to $\rho_{0,n}(u)$ is given

by

$$\widehat{\rho}_{0,n}(u) \triangleq \int_{z=u}^{\infty} \widehat{f}_n(z)\, dz.$$

For p-value calculation the order $O\left(n^{-\alpha}\right)$ approximations $\widehat{\rho}_{0,n}$ and $\overline{\rho}_{0,n}$ will often be sufficiently accurate. If the field is not too smooth and/or the search region is large in volume, higher terms begin to dominate (1.3) and the Gaussian density $\rho_{0,n}^{\gamma}$ may in fact be adequate. For a more complete account of tail probability approximations, see the references already mentioned in this subsection.

Corollary 5 and Proposition 6 suggest that, asymptotically, tilting the EC densities is worthwhile provided u grows with n. In statistical applications, "$u \to \infty$" corresponds to "large u", which is a crucial if unstated assumption when thresholding random field data via the expected EC heuristic. However, the tilted densities are not guaranteed to be any more accurate than the Gaussian ones in the small threshold scenario $\alpha = 1/2$, at least as $n \to \infty$. The performance of the approximation in finite samples is another matter altogether. These issues and others will be addressed in Sections 6 and 7.

6   Example: a scaled χ_n^2 random field

It is instructive to consider a random field for which both the tilted and exact EC densities are available explicitly; the χ_n^2 random field is one such case. Let $X_1, X_2, \ldots, X_n$ be i.i.d. centred, unit variance Gaussian random fields on T and put $Y(t) = \sum_{i=1}^{n} X_i(t)^2$. It has been established (by two distinct techniques [35, 42]) that the EC densities of Y are

given by

$$\rho_{k,n}^{\chi}(u) = \frac{(k-1)! u^{(n-k)/2} e^{-u/2}}{(2\pi)^{k/2} \Gamma\left(\frac{n}{2}\right) 2^{n/2-1}} \sum_{j=0}^{\lfloor \frac{k-1}{2} \rfloor} \sum_{l=0}^{k-1-2j} \binom{n-1}{k-1-2j-l} \frac{(-1)^{k-1}(-u)^{j+l}}{l!\, j!\, 2^j} \mathbb{1}_{\{n \geq k-2j-l\}}$$

(6.1)

for $k \geq 1$.

We normalize by forming $Z(t) = n^{-1/2}(Y(t) - n)$, so that Z is centred, approximately Gaussian, and $\sigma^2 = \mathsf{Var}[Z(t)] = 2$. Note that the χ^2 field is 'twice as rough' as each X_i; if the component Gaussian fields are isotropic with second spectral moment λ_x, then (see [1], Chapter 7) $\mathsf{Var}(\dot{Z}) = n^{-1}\mathsf{Var}(\dot{Y}) = 4\lambda_x I_N$. To insure that $\lambda = 1$ for the normalized process we therefore assume that the X_i's are isotropic with $\Lambda_X \stackrel{\Delta}{=} \mathsf{Var}(\dot{X}_i) = I_N/4$. With that covariance structure and for a square parameter space T, we have

$$\mathbb{E}\varphi\left(\{t:\; Z(t) \geq u\}\right) = \mathbb{E}\varphi\left(\{t:\; Y(t) \geq \sqrt{n}u + n\}\right)$$
$$= \sum_k |I_k/4|^{1/2} \rho_{k,n}^{\chi}(\sqrt{n}u + n) \mathcal{V}_k(T),$$
$$= \sum_k 2^{-k} \rho_{k,n}^{\chi}(\sqrt{n}u + n) \mathcal{V}_k(T).$$

Thus the proper comparison is between $\widehat{\rho}_{k,n}(u)$ and $2^{-k}\rho_{k,n}^{\chi}(\sqrt{n}u + n)$ for $k = 0, \ldots, N$ (this is a minor notational discrepancy; (6.1) is in fact $2^k \rho_{k,n}^{\chi}(u)$ by the conventions of this thesis). We have:

$$K_Z(\theta) = -\sqrt{n}\theta - \frac{n}{2}\log\left(1 - 2\frac{\theta}{\sqrt{n}}\right);$$

$$K_Z'(\theta) = \frac{2\theta}{1 - 2\theta/\sqrt{n}};$$

$$K_Z''(\theta) = 2\left(1 - 2\frac{\theta}{\sqrt{n}}\right)^{-2}.$$

30

It follows that:

$$\hat{\theta}_u = \frac{\sqrt{n}u}{2(u + \sqrt{n})}; \ I_u(\hat{\theta}_u) = \frac{\sqrt{n}u}{2} - \frac{n}{2}\log\left(1 + \frac{u}{\sqrt{n}}\right); \ \tau_u^2 = 2\left(1 + \frac{u}{\sqrt{n}}\right)^2; \ \tau_u\hat{\theta}_u = \frac{u}{\sqrt{2}}.$$

Note that the Hermite portion of the tilted density is identical to that of the Gaussian density in this case. Plugging these into (4.5) we have, for $k \geq 1$,

$$\widehat{\rho}_{k,n}(u) = \frac{e^{-u\sqrt{n}/2}\left(1 + u/\sqrt{n}\right)^{n/2-k}}{(2\pi)^{(k+1)/2}2^{k/2}}H_{k-1}(u/\sqrt{2}). \tag{6.2}$$

Collecting all of the constants in (6.1), again for $k \geq 1$,

$$\rho_{k,n}(u) = 2^{-k}\rho_{k,n}^{\chi}(\sqrt{n}u + n)$$
$$= \frac{e^{-u\sqrt{n}/2}\left(1 + u/\sqrt{n}\right)^{n/2-k}}{(2\pi)^{(k+1)/2}2^{k/2}}R_{k-1,n}(u), \tag{6.3}$$

where (assuming $n \geq N$) $R_{k,n}$ is the following degree k polynomial:

$$R_{k,n}(x) = 2^{-(k-2)/2}k!\frac{\sqrt{2\pi}e^{-n/2}}{\Gamma(n/2)}\left(\frac{n}{2}\right)^{(n+1)/2}$$
$$\times \sum_{j=0}^{\lfloor\frac{k}{2}\rfloor}\sum_{l=0}^{k-2j}\binom{n-1}{k-2j-l}\frac{(-1)^{k+j+l}n^{j+l-k/2-1}}{l!\,j!\,2^j}\left(1 + \frac{x}{\sqrt{n}}\right)^{j+l+(k+1)/2}.$$

Also, note that

$$\frac{\sqrt{2\pi}e^{-n/2}}{\Gamma(n/2)}\left(\frac{n}{2}\right)^{(n+1)/2} \sim \frac{\Gamma(n/2+1)}{\Gamma(n/2)} = \frac{n}{2}$$

as $n \to \infty$. The corresponding Gaussian densities are

$$\rho_{k,n}^{\gamma}(u) = \frac{e^{-u^2/4}}{(2\pi)^{(k+1)/2}2^{k/2}}H_{k-1}(u/\sqrt{2}), \quad k \geq 1. \tag{6.4}$$

31

For the $k = 0$ EC density, we have

$$\rho_{0,n}(u) = \rho_{0,n}^{\chi}(\sqrt{n}u + n) = \mathbb{P}\left\{Y \geq \sqrt{n}u + n\right\} = \int_{\sqrt{n}u+n}^{\infty} g_n(y)dy,$$

where $Y \overset{\mathscr{D}}{=} Y(t) \sim \chi_n^2$ and $g_n(y) = \Gamma(n/2)^{-1}2^{-n/2}y^{n/2-1}e^{-y/2}$ is the density of Y. It is a well-known fact [15] that the first-order saddlepoint approximation $\hat{g}_n$ to g_n is exact up to a constant, so that

$$g_n(y) \equiv \hat{\hat{g}}_n(y). \tag{6.5}$$

To obtain a similar result for $\rho_{0,n}(u)$ and $\hat{\rho}_{0,n}(u)$ we shall use the following lemma, which is easily proved in Appendix A.6.

Lemma 7. *Let* $Z = \sigma_n^{-1}(Y - \mu_n)$ *be a random variable with density* $f_n(z) = \sigma_n g_n(\sigma_n z + \mu_n)$, *where* $g_n(y)$ *is the density of* Y. *Then* $\forall\, z \in \mathbb{R}$,

$$\hat{f}_n(z) = \sigma_n \hat{g}_n(\sigma_n z + \mu_n).$$

Let $f_n(z)$ denote the density of $Z(t)$. Putting $\sigma_n = \sqrt{n}$ and $\mu_n = n$ in the lemma and using (6.5), we have

$$
\begin{aligned}
\rho_{0,n}^{\chi}(\sqrt{n}u + n) &= \int_{\sqrt{n}u+n}^{\infty} \hat{\hat{g}}_n(y)\, dy \\
&= n^{-1/2} \int_{\sqrt{n}u+n}^{\infty} \hat{\hat{f}}_n(n^{-1/2}(y - n))\, dy \\
&= \int_{u}^{\infty} \hat{\hat{f}}_n(z)\, dz \\
&= \hat{\rho}_{0,n}(u).
\end{aligned}
$$

So the integrated, *scaled* tilted zeroth density is exact in this case, and this is a rare instance where normalization of the saddlepoint approximation does not increase com-

putation. The ordinary tilted and Gaussian $k = 0$ densities are given, respectively, by

$$\widehat{\rho}_{0,n}(u) = e^{(u^2 - 2\sqrt{n}u)/4} \left(1 + \frac{u}{\sqrt{n}}\right)^{n/2} \Psi\left(\frac{u}{\sqrt{2}}\right);$$

$$\rho_{0,n}^{\gamma}(u) = \Psi\left(\frac{u}{\sqrt{2}}\right).$$

As a term-by-term comparison of (6.2) and (6.3) is onerous, we compare $H_k(u/\sqrt{2})$ and $R_{k,n}(u)$ for $k = 0, 1, 2$, the cases of interest in a 2D or 3D problem. We have:

$$H_0(u/\sqrt{2}) = 1, \qquad R_{0,n}(u) = c_n\sqrt{1 + \frac{u}{\sqrt{n}}} = 1 + O\left(n^{-\alpha}\right);$$

$$H_1(u/\sqrt{2}) = \frac{u}{\sqrt{2}}, \qquad R_{1,n}(u) = c_n\frac{u}{\sqrt{2}}\left(1 + \frac{u}{\sqrt{n}} + \frac{1}{u\sqrt{n}} + \frac{1}{n}\right) = \frac{u}{\sqrt{2}}\left(1 + O\left(n^{-\alpha}\right)\right);$$

$$H_2(u/\sqrt{2}) = \frac{u^2}{2} - 1,$$

$$R_{2,n}(u) = c_n\left(\frac{u^2}{2} - 1 + \frac{u}{2\sqrt{n}} + \frac{1}{n}\right)\left(1 + \frac{u}{\sqrt{n}}\right)^{3/2} = \left(\frac{u^2}{2} - 1\right)\left(1 + O\left(n^{-\alpha}\right)\right);$$

where $c_n = \sqrt{2\pi}e^{-n/2}\Gamma(n/2)^{-1}(n/2)^{(n-1)/2} = 1 + O\left(n^{-1}\right)$. Hence, the tilted EC densities for $k = 1, 2, 3$ have a relative error $O\left(n^{-\alpha}\right)$. Note that the Gaussianity of the X_i's in this example leads to $\mathbb{E}[Z^2 Z_j] = 0$, which implies that for all $k \geq 1$, a relative error *no worse* than $O\left(n^{\max\{-\alpha, 3/2 - 6\alpha\}}\right)$ must be achieved by $\widehat{\rho}_{k,n}$—see (4.7). It is easily verified that $I_u(\hat{\theta}_u) = u^2/4 - u^3/(6\sqrt{n}) + O\left(n^{1-4\alpha}\right)$, so that $\epsilon_n[\rho_{k,n}^{\gamma}(u), \rho_{k,n}(u)] \stackrel{s}{=} 1 - 3\alpha$ when $\alpha \geq 1/3$.

The behaviour of these approximations for $k \leq 3$ is explored further in Figure 2. In each column we consider a different asymptotic regime: the most general one in which $\alpha = 1/3$ and both n and u grow; the classical one in which $\alpha = 1/2$ so that u is fixed while n grows; and the less common one in which n is fixed while u grows (corresponding

to $\alpha = -\infty$). The latter regime is meaningful for statistical applications since the sample size (n) is large but finite and the relative error in $\mathbb{E}\varphi(A_u)$ as a p-value approximation is conjectured to fall off exponentially as $u \to \infty$ (see Section 4.4). The theory of Sections 4.1 and 5 does not directly apply to this scenario. A fourth regime, where $k \to \infty$, might also be of interest; empirical evidence for the χ^2_n field suggests that the tilted densities perform poorly under that scheme.

For $k = 0, 1$ in this example, the tilted densities offer a clear improvement over the Gaussian approximation even for small n and/or u. For $k = 2, 3$ a larger sample size or threshold is needed before the difference emerges, but it is still unequivocal when $u \to \infty$. When $\alpha = 1/2$ for $k = 2$ or 3, the tilted and Gaussian approximations perform similarly, as predicted by Corollary 5. In the third regime where $u \to \infty$ while n remains fixed, the Gaussian densities are quickly diverging from the correct densities; however, they are more accurate than the tilted densities for $k = 2, 3$ in the range $2 \leq u \leq 4$. Note that the chosen sample size $n = 500$ was quite large in that regime. When $k = 0$ the naïve tilted density $\widehat{\rho}_{0,n}(u)$ appears to outperform the Lugannani-Rice estimate's leading term $\overline{\rho}_{0,n}(u)$, but this difference is small and may be specific to the χ^2_n field. The Gaussian assumption generally leads to *under*estimation of the expected EC in this example. In other words, using a Gaussian assumption for statistical thresholding of χ^2_n fields is anti-conservative.

7 Application: subtracted 'bubbles' images

A moment's thought reveals that the form (1.1) taken for the Z process is far more restrictive than is necessary. Indeed, the crucial property of $\mathbf{Z} = (Z, \dot{Z}', \ddot{Z}')$ appearing
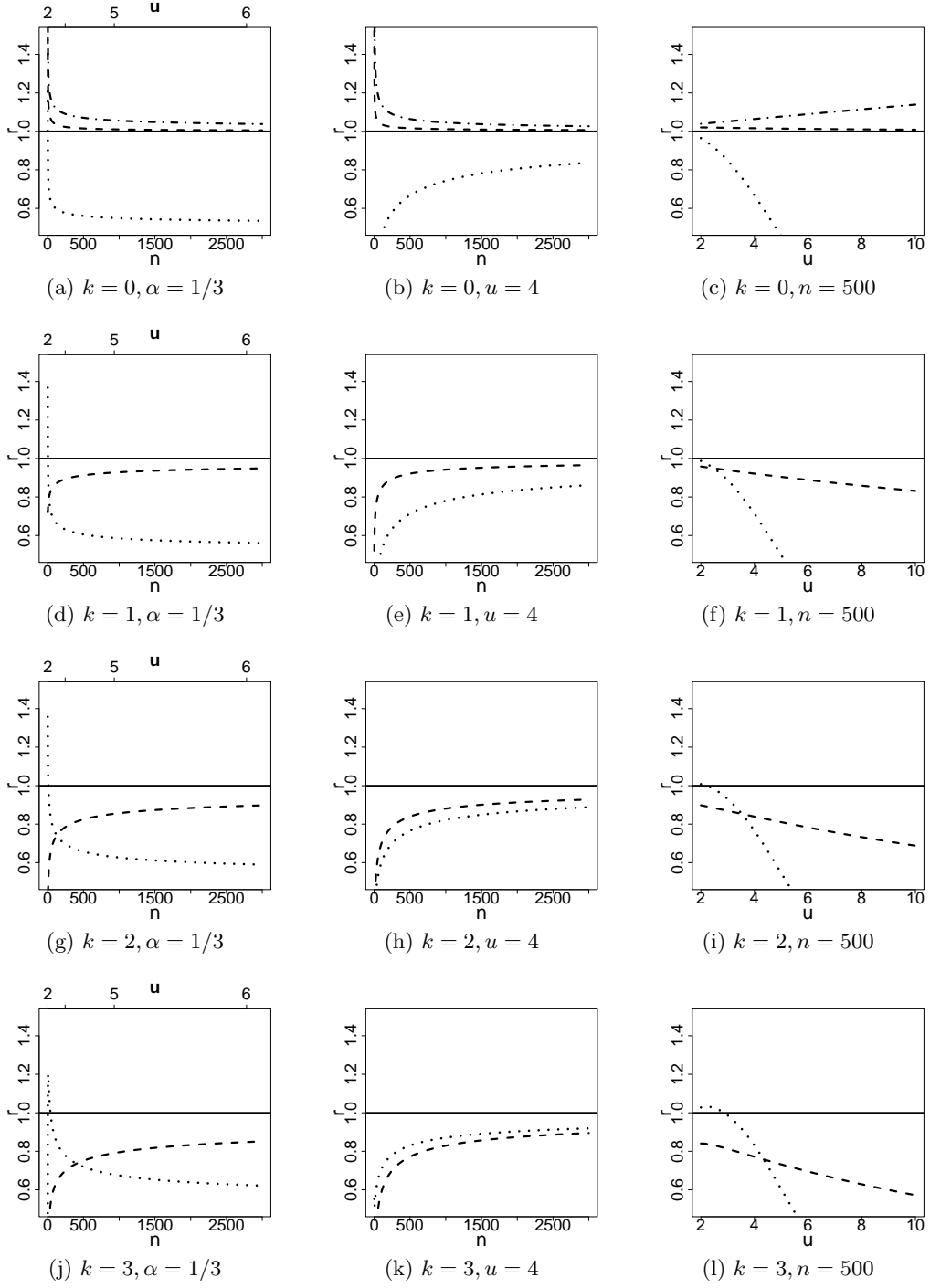
Figure 2: EC densities for the χ_n^2 random field: exact ($\rho_{k,n}$, '——'); tilted ($\widehat{\rho}_{k,n}$, '- -'); and Gaussian ($\rho_{k,n}^\gamma$, '· · ·'). Ratios r of approximate to exact densities are plotted for three regimes: $u \sim cn^{1/6} \to \infty$ ($\alpha = 1/3$, 1$^{\text{st}}$ col.); u fixed and $n \to \infty$ ($\alpha = 1/2$, 2$^{\text{nd}}$ col.); n fixed and $u \to \infty$ ($\alpha = -\infty$, 3$^{\text{rd}}$ col.). For $k = 0$, $\rho_{0,n} \equiv \widehat{\rho}_{0,n}$, and '-··-' denotes $\overline{\rho}_{0,n}$.

in the proof of Theorem 4 is the rate of decay of its (joint) cumulants; *viz.*

$$\mathbb{E}[Z(t)] \equiv 0, \quad \mathsf{Var}[Z(t)] \equiv \sigma^2; \tag{7.1}$$

and

$$k_\nu([\mathbf{Z}(t)]^\gamma) = O\left(n^{-(\nu-2)/2}\right), \quad \gamma \in \mathbb{N}^{(N+1)(N+2)/2}, \ |\gamma| = \nu, \ \nu \geq 2. \tag{7.2}$$

One might therefore expect Theorem 4 to apply to a variety of stationary random fields satisfying (7.1) and (7.2) in addition to the regularity conditions in the Appendix. The following is a slightly more general construction than (1.1): let $\{W_i(t)\}_i$ be independent but perhaps nonidentically-distributed random fields with finite cumulants of all orders. Assume that the random field of interest takes the form

$$Z(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n a_i W_i(t) \tag{7.3}$$

for finite constants $(a_i)_{i=1}^\infty$. This framework covers the case of a two-sample problem (e.g. a difference in means between two groups), as well as the following application.

The bubbles methodology of Gosselin & Schyns has been described in detail elsewhere [18]. Typically, the experimenters' goal is to determine which parts of an image are most important in a visual discrimination task. The data analyzed later in this section come from [3]. On each trial of the experiment, a subject was shown a 256^2 image of a face that was either fearful or happy. The images were masked apart from a random number of localised regions or 'bubbles' that revealed only selected parts of the face; see Figure 3. The subject was then asked whether the partially revealed face was fearful or happy, and the trial was repeated about 3,000 times on each of 10 subjects.

Each masked image was generated as follows. First a range of scales was chosen, $\sigma_{Bj} =$
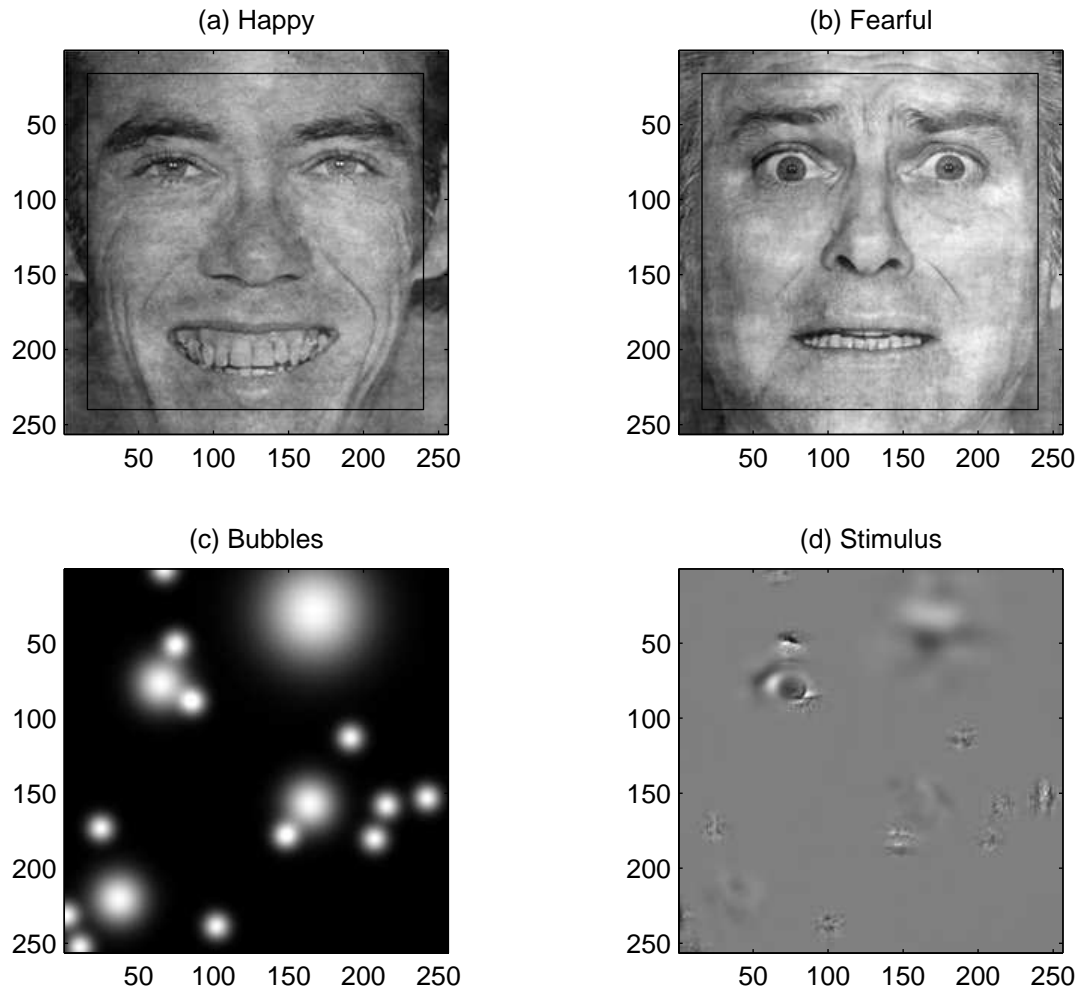
Figure 3: Bubbles experiment. The subject is asked to discriminate between the happy (a) and fearful (b) faces on presentation of the stimulus (d) which is one of the two faces, here the fearful face, partially revealed by random 'bubbles' (c). The 227^2 search region T is inside the black frame in (a) and (b).

3×2^j, $j = 1, \ldots, 5$. Then the original image $\mathcal{I}_0$ (Figure 3(a) or (b)) was smoothed by an isotropic Gaussian filter with standard deviation $\sigma_{Bj}/2$ to produce images $\mathcal{I}_j$. The smoothed images were differenced to produce images $D_j = \mathcal{I}_{j-1} - \mathcal{I}_j$ that reveal image features at five different scales. Differenced image D_j was then multiplied by a mask consisting of the sum of a random number of isotropic Gaussian 'bubbles', each with standard deviation σ_{Bj}. The bubble centres were chosen at random from the 256^2 pixels, *i.e.* according to a Poisson process. The number of bubbles for each scale was a multinomial random variable with probabilities inversely proportional to bubble area ($\propto \sigma_{Bj}^{-2}$), so that equal areas were revealed by each of the five bubble sizes, on average. The sum of all the bubbles over all five scales is shown in Figure 3(c) for one of the trials. The sum of all the bubbles times the differenced images is shown in Figure 3(d). On the basis of Figure 3(d) the subject must decide if the face is happy or fearful (fearful is the correct answer in this case). The total number of bubbles was chosen dynamically to maintain a roughly 0.75 probability of correctly identifying the face.

7.1 Test statistic, cumulants and tilted densities

For the present analysis we consider a single-subject run. We assume that the grid is square, and that the bubble standard deviation σ_B (or equivalently the bubble full width at half max $F = \sqrt{8 \log 2} \sigma_B$) is fixed, as is the number m of bubbles per image. We denote by p_C and $p_I = 1 - p_C$ the proportions of correctly and incorrectly classified images. Let Y_{ip} denote the number of bubbles in image i which are centred at pixel p, with $i = 1, \ldots, n$ and $p = 1, \ldots, P$. Then $Y_{ip} \sim \mathcal{P}(m/P)$, and the Y_{ip} are approximately independent (exactly when they come from different images). The test statistic field for detecting significant regions is formed by signing and summing the bubbles, with sign depending on whether the image was correctly or incorrectly classified. Thus, the

unnormalized test statistic at pixel t is given by

$$\widetilde{Z}(t) \triangleq n^{-1/2} \sum_{i=1}^{n} \sum_{p=1}^{P} a_i b_p(t) Y_{ip}, \tag{7.4}$$

where

$$a_i = \begin{cases} 1/p_C & \text{image } i \text{ correct} \\ \\ -1/p_I & \text{image } i \text{ incorrect}, \end{cases}$$

and

$$b_p(t) = \exp\left\{-\frac{\|p-t\|^2}{2\sigma_B^2}\right\}.$$

The statistic is normalized by forming

$$Z(t) = \widetilde{Z}(t) \Big/ \sqrt{\mathsf{Var}\widetilde{Z}(t)}\,.$$

Inference can be either conditional on the observed vector a or unconditional, in which case the a_i are i.i.d. scaled Rademacher random variables and p_C is taken as the (assumed known) *population* proportion. One may apply the theory of the present article under either analysis; we have chosen the conditional approach since it simplifies cumulant calculations. Thus p_C and p_I are observed proportions.

The most natural global null hypothesis is

$$H_0: \quad a_i \text{ is independent of } (Y_{ip})_{p=1}^{P} \ \forall\, i \le n;$$

which implies

$$H_0': \quad \mathbb{E}\left\{Z(t)|\ a\right\} = 0 \ \forall\, t \in T;$$

since $\mathbb{E}_{H_0}\left\{Z(t)|\ a\right\} \propto \sum_p b_p(t) \sum_i a_i \mathbb{E}_{H_0}\left\{Y_{ip}|\ a\right\} = \frac{m}{P}\left(\sum_p b_p(t)\right)\left(\sum_i a_i\right) = 0$. Ignoring

mild boundary effects, note that the null distribution of $Z(t)$ is unchanged if we translate or rotate the coordinate system; thus $Z(t)$ is (strictly) isotropic. By (7.4), Z is of the form $n^{-1/2} \sum_i a_i W_i$ where the W_i's are i.i.d. random fields. The distribution of $W_i(t) = \sum_p b_p(t) Y_{ip}$ is very non-Gaussian, with a high probability of being nearly zero (if there are no bubble centres near t), and a small probability of being large. The distributions become more Gaussian as m or F increase.

An important practical note is that most users of this methodology choose to 'clamp' the summed bubbles at a maximum height equal to the peak height of a single bubble. In particular, two bubbles centred at the same pixel, when added, have the same peak height as a single bubble. This is how the mask is actually created, and it is felt that this is how the stimulus is perceived. Clamping effectively replaces (7.4) with $n^{-1/2} \sum_{i=1}^{n} a_i \min \left\{ \sum_{p=1}^{P} b_p(t) Y_{ip}, 1 \right\}$. Doing so not only complicates cumulant derivations, but more importantly degrades the 'smoothness' of the random field. When the number of bubbles is small compared to the number of pixels and their width is not too large (as in the data analyzed below), the loss of smoothness due to clamping is negligible. Our choice of an unclamped analysis is always valid, but perhaps less powerful at detecting facial features if indeed the probability of response is related to the clamped rather than the unclamped bubble mask.

Without any extra effort we generalize the cumulant computations to N dimensions, although currently in bubbles experiments $N = 2$. It can be shown that the jth null conditional cumulant of $Z(t)$ is approximately

$$\kappa_j = \left(\frac{2^{j/2}}{j} \right)^{N/2} \frac{p_C^{1-j} + (-1)^j p_I^{1-j}}{(p_C^{-1} + p_I^{-1})^{j/2}} \left\{ \left(\frac{4 \log 2}{\pi} \right)^{N/2} \frac{1}{B} \right\}^{j/2-1}, \qquad (7.5)$$

which only depends on p_C and the total number $B \triangleq nm \big/ \left(P F^{-N} \right)$ of bubbles per

40

N-dimensional 'resel' (volume element). These have been used to compute the tilted variance of Z,

$$\tau_u^2 = \sum_{j=0}^{\infty} \kappa_{j+2} \frac{\hat{\theta}_u^j}{j!}.$$

In practice (for typical values of p_C and B) $j \leq 20$ seems to be enough; beyond this the cumulants are extremely close to zero. The tilting parameter $\hat{\theta}_u$ is found by solving $K_Z'(\theta) = u$ with a Newton-Raphson algorithm started from the Gaussian solution $\theta = u$; five iterations appear to suffice in most cases. An approach like this one was advocated by RS.

One can also show that the second spectral moment of Z under H_0 is given (approximately) by

$$\mathsf{Var}\left\{\dot{Z}(t)\,\middle|\,a\right\} \approx \frac{1}{2\sigma_B^2} I_N \triangleq \lambda I_N. \tag{7.6}$$

The p-value approximation when $N = 2$ is thus

$$\mathbb{P}\left\{\sup_t Z(t) \geq u\right\} \approx \mathbb{E}\varphi(A_u) \doteq \widehat{\rho}_0(u) + 2(P\lambda)^{1/2}\widehat{\rho}_1(u) + P\lambda\widehat{\rho}_2(u), \tag{7.7}$$

where the EC densities are

$$\widehat{\rho}_0(u) = \exp\left\{\tau_u^2\hat{\theta}_u^2/2 - I_u(\hat{\theta}_u)\right\}\Psi\left(\tau_u\hat{\theta}_u\right),$$

$$\widehat{\rho}_1(u) = \exp\left\{-I_u(\hat{\theta}_u)\right\}(2\pi\tau_u)^{-1}, \tag{7.8}$$

$$\widehat{\rho}_2(u) = \exp\left\{-I_u(\hat{\theta}_u)\right\}(2\pi)^{-3/2}\tau_u^{-1}\hat{\theta}_u.$$

By contrast, the Gaussian theory EC densities are

$$\rho_0^\gamma(u) = \Psi(u),$$
$$\rho_1^\gamma(u) = \exp\left\{-u^2/2\right\}(2\pi)^{-1},$$
$$\rho_2^\gamma(u) = \exp\left\{-u^2/2\right\}(2\pi)^{-3/2}u;$$

(7.9)

with p-value approximation

$$\rho_0^\gamma(u) + 2(P\lambda)^{1/2}\rho_1^\gamma(u) + P\lambda\rho_2^\gamma(u).$$

(7.10)

When P is large, the third terms in (7.7) and (7.10) dominate. Typical values for the parameters are $P = 256^2$, $m = 16.5$, $F = 14.1$, $n = 3000$, $p_C = 0.75$; setting $u = 3.965$ (the $p = 0.05$ threshold using (7.7)) then gives

$$\widehat\rho_0(u) = 1.4 \times 10^{-5}, \quad 2(P\lambda)^{1/2}\widehat\rho_1(u) = 0.00171, \quad P\lambda\widehat\rho_2(u) = 0.0484.$$

In view of this, the ratio of the tilted to Gaussian p-value approximations,

$$r \approx \frac{\widehat\rho_2(u)}{\rho_2^\gamma(u)} = \exp\left\{\frac{u^2}{2} - I_u(\hat\theta_u)\right\}\frac{\hat\theta_u}{\tau_u u},$$

(7.11)

only depends on u, B and p_C. We also consider the approximation suggested by extension (4.18) of RS, namely

$$\widehat\rho_0(u) + 2\left(P\widehat\lambda_u\right)^{1/2}\widehat\rho_1(u) + P\widehat\lambda_u\widehat\rho_2(u),$$

(7.12)

where $\widehat\lambda_u$ denotes the *tilted* second spectral moment of Z. It can be demonstrated that $\widehat\lambda_u \approx \tau_u^2\lambda$. To compare this approximation to the Gaussian theory, one can again form

42

the p-value ratio

$$r_{\mathrm{RS}} \approx \frac{\widehat{\lambda}_u \widehat{\rho}_2(u)}{\lambda \rho_2^\gamma(u)} \approx \tau_u^2 r. \tag{7.13}$$

These ratios are plotted for $u = 3.5$ and a range of B and p_C in Figure 4. Note that in
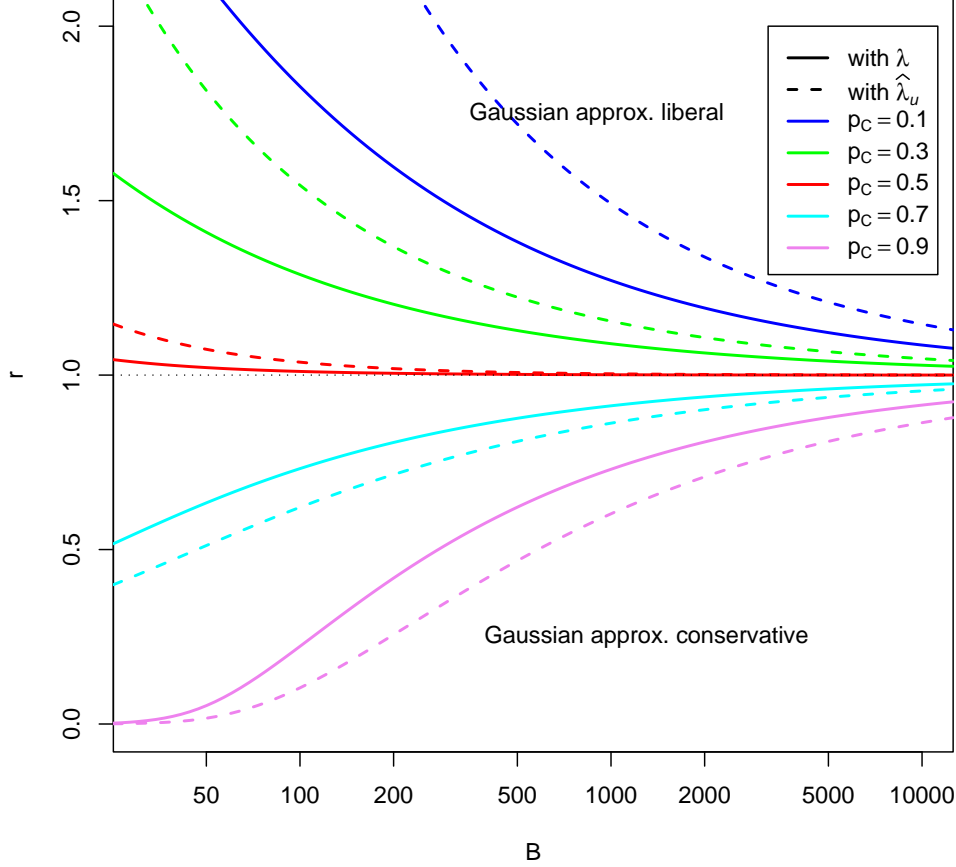

Figure 4: Ratios of tilted to Gaussian-approximated expected Euler Characteristic of the excursion of a test statistic from a typical bubbles experiment above the threshold $u = 3.5$, using both the untilted (λ) and the tilted ($\widehat{\lambda}_u$) second spectral moment. The ratios are plotted for a reasonable range of B (the total number of bubbles per resel) and for several values of p_C (the proportion of correctly classified images). When $p_C = 1/2$ (red curve), the skewness of Z is zero, and the saddlepoint correction is less pronounced. Assuming that the tilted expressions are closer to the true p-value, $r > 1$ signifies that the Gaussian approximation is anti-conservative, while $r < 1$ means that the Gaussian procedure is conservative. Note that $p_C \approx 0.75$ is commonly targeted by the originators of the bubbles paradigm, making the Gaussian theory valid though less powerful than the tilted inference.

this application the approximation (7.7) deviates *less* extremely from the Gaussian one

than does the heuristic of RS. This can be understood by noting that, when $\alpha = 1/2$,

$$\tau_u^2 = 1 + \kappa_3 \hat{\theta}_u + O\left(n^{-1}\right), \tag{7.14}$$

where $\hat{\theta}_u \approx u > 0$ and $\operatorname{sgn}\{\kappa_3\} = \operatorname{sgn}\left\{\frac{1}{2} - p_C\right\}$. At the same time, $r > 1 \Leftrightarrow p_C \leq 1/2$. The $O\left(n^{-1}\right)$ term in (7.14) can be positive or negative but is of smaller order than the skewness term. It is also worth noting that unlike in the χ_n^2 example and the fields considered by RS, the Gaussian approximation can be conservative (*i.e.* lead to larger p-values) in bubbles experiments when typical parameter values ($p_C > 1/2$) are used.

7.2 Simulation

A simulation was devised to compare (7.7), (7.10) and (7.12) with the true tail probability (1.2) of $\sup_t Z(t)$ above a threshold under H_0. In each of 10^5 Monte Carlo iterations, we generated 3000 bubbles images on a $P = 208 \times 208$ square with corresponding classifications, using $p_C = 0.75$ and bubble width $F = 24$. For each iteration we then calculated a sequence of test statistic fields $Z_n(t)$ for $n = 100, 200, \ldots, 3000$. Since kernel smoothing with Fourier methods was used to speed up computation, the data were generated first on a larger 256×256 square before discarding the outer edges to eliminate periodicity. The number of bubbles per image was fixed so that an average of $m = 20$ would fall within the search region; this has a negligible effect on the cumulants (7.5). The mixed skewness terms $\mathbb{E}[Z(t)^2 Z_i(t)]$ can be shown to be (approximately) proportional to $\sum_p (t_i - p_i) b_p(t)^3$, and hence are very close to zero for pixels far from the boundary (recall that they were exactly zero in the χ^2 example). We therefore expect the tilted approximations to be within the order $O\left(n^{-\alpha}\right)$ of the true p-value. Figure 5 displays the simulation results under the three asymptotic regimes. In addition to using both λ and
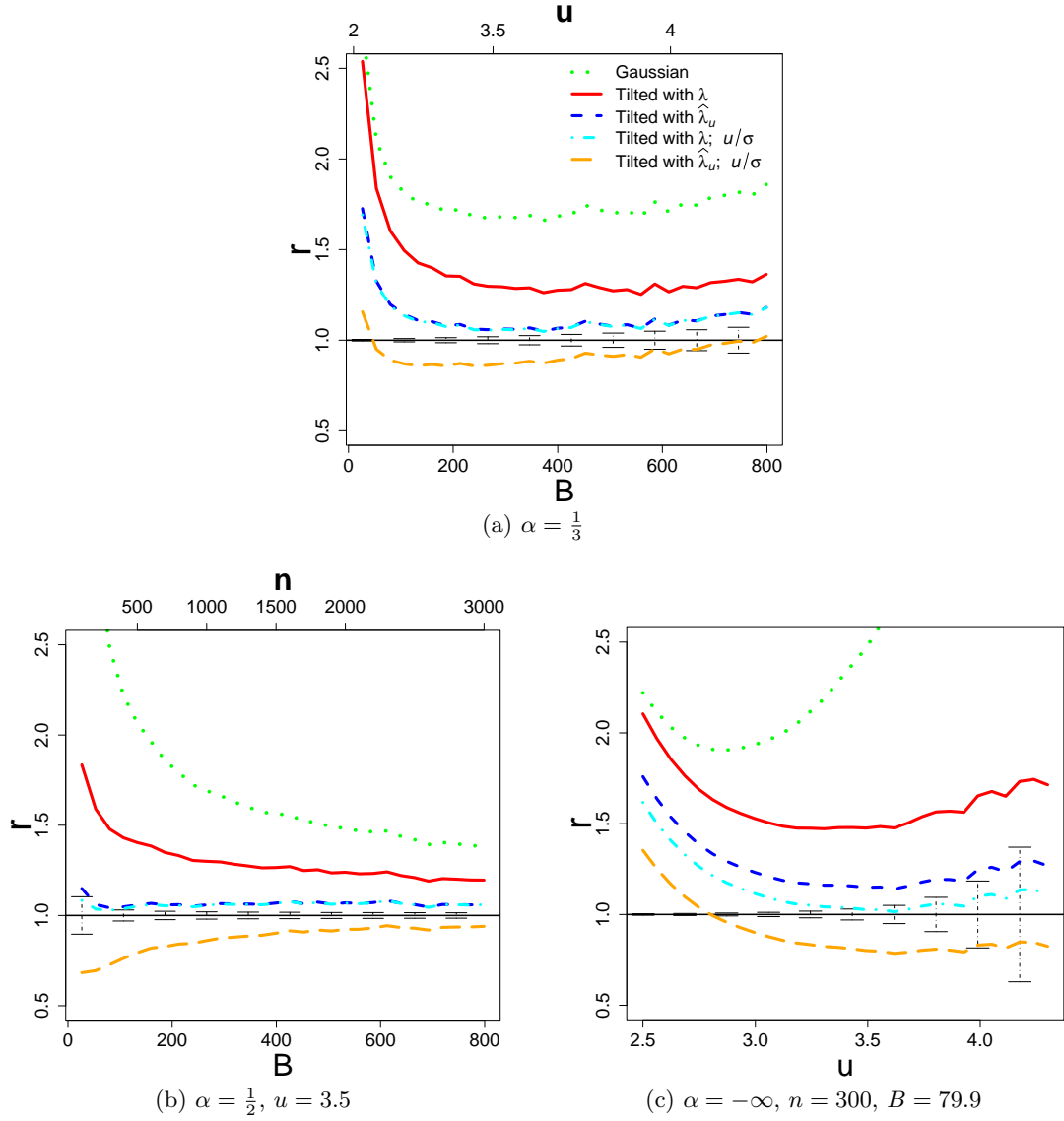
44

Figure 5: Ratios of expected EC approximations to true (empirical) p-values for simulated bubbles data, under the three asymptotic regimes of Section 7. Bubbles parameter values were $P = 208^2, F = 24, m = 20, p_C = 0.75$, and n ranging from 100 to 3000. The ratio scale was chosen so that absolute deviations from the line $r = 1$ correspond to relative error. The Gaussian expected EC is given by (7.10). The tilted approximations use either the untilted (λ) or tilted ($\widehat{\lambda}_u$) second spectral moment, and either the tilted ($\tau_u \hat{\theta}_u$) or untilted (u/σ) Hermite argument. Monte Carlo error bars (± 1 standard deviation) are for the Gaussian curve ('$\cdots$'), but are attached to the line $r = 1$ to aid interpretation. They are computed pointwise, using $\text{Var}(\widehat{p}/p_{emp}) \approx 10^{-5} \widehat{p}^2 (1 - p_{emp})/p_{emp}^3$, where p_{emp} is an empirical p-value and $\widehat{p}$ one of the EC approximations. Error bars for the other curves are slightly narrower. The threshold $u = 3.5$ was chosen in (b) in order to have empirical probabilities near 0.05. The sample size $n = 300$ was chosen in (c) in order to have the total bubbles/resel close to the value $B = 63.6$ observed in the data. In all cases the tilted approximations outperform the Gaussian one, including in (c) where n is fixed, u grows and all curves appear to diverge from unity.

$\widehat{\lambda}_u$, we also tried u/σ in the place of $\tau_u\hat{\theta}_u$ as the argument in the Hermite polynomials. This resulted in third and fourth variations on the tilted p-value approximation. In all instances the tilted expected EC's were closer to the observed p-value than the Gaussian expected EC. In particular, the expansion which uses the tilted spectral moment $\widehat{\lambda}_u$ and the Gaussian Hermite argument u/σ ('$--$', orange) performed best, although differences between tilted approximations were small. The observed directional trend between the four tilted expressions is likely specific to this application. We note that while feasible for parameter spaces of this size, deriving p-values by simulation for bubbles-type data is time-consuming, and quickly becomes impracticable as the dimension N increases beyond 2. The benefits of having a closed-form p-value approximation are evident.

7.3 Data analysis

We analyzed a portion of the data from [3] described at the beginning of this section (control group). One subject (#9) was presented with $n = 3072$ images of size 256×256 pixels, which he/she classified as either happy or fearful, at a rate of $p_C = 0.76$. The search region was trimmed to the inner $P = 227^2$ pixels to avoid Fourier edge effects. In agreement with the theory in this article we have limited our analysis to a single bubble width, $F = 18.8$ pixels or $\sigma_B = 8$ (the subject was in fact shown bubbles of width $F = 28.3$; *i.e.* $j = 2$, $\sigma_{Bj} = 12$; but we chose a narrower filter when smoothing the test statistic in order to totally eliminate the 'clamping' issue). The mean number of bubbles per image at this scale and within the search region was $m = 3.0$. Our chosen single bubble width was the second-narrowest class ($j = 2$); an example of the three bubbles is clearly visible in Figure 3(c). These parameter values correspond to $B = 63.6$ bubbles/resel. Thresholded images are displayed in Figure 6. There is strong signal in the right eye region for this subject. The global maximum of $Z(t)$ exceeded

5.8, and hence all thresholding methods identified that peak. At the $p = 0.05$ level the Gaussian threshold picked up only the right eye, while the tilted threshold found additional activation in the upper lip. At $p = 0.1$ significance the tilted expected EC identified a third signal in the left eye which went undetected by the Gaussian theory. Note that the Gaussian approximation is conservative in this case since $Z(t)$ exhibits negative skewness $(p_C > 1/2)$.

8   Discussion and future work

We have provided a rigourous theoretical justification for the use of saddlepoint methods in thresholding asymptotically-Gaussian random fields, first proposed by Rabinowitz & Siegmund [29]. In particular we have extended their 'expected number of maxima' heuristic to an expected Euler Characteristic approximation for locally isotropic Central Limit-type fields, and derived relative errors for the EC densities when the threshold is allowed to grow. We have also compared these formulae to a Gaussian approximation, both analytically and in numerical and real data examples involving a χ_n^2 field and a linear combination of Poisson fields.

As did RS (who only considered a Poisson process) vis-à-vis expected local maxima, we found substantial differences between the Gaussian and the various tilted approximations (which were uniformly more accurate in a particular case—see Fig. 2) to the expected EC. It was shown in Section 5 that in a large-threshold setting the saddlepoint correction reduces asymptotic relative error. In contrast to the examples chosen by RS, we have demonstrated via simulation that the Gaussian approximation can be conservative in some instances rather than always underestimating the p-value; see Section 8 and Figs. 4-5. We have shown that the important correction for non-Gaussianity occurs in the

(a) $Z(t)$ (b) $u = 2$

(c) $u = 3.92$ (d) $u = 3.76$
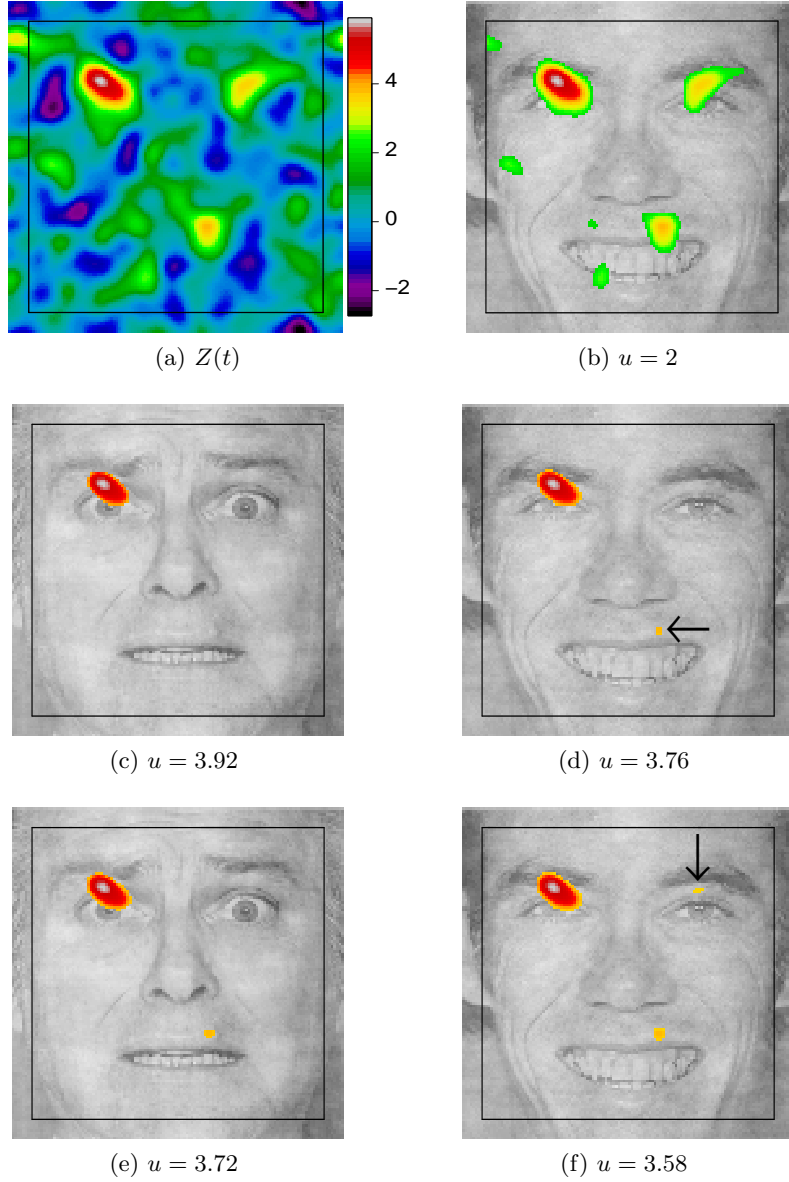
(e) $u = 3.72$ (f) $u = 3.58$

Figure 6: Thresholded bubbles data for subject # 9, with the search region framed. Test statistic (a), with excursions above: (b) $u = 2$ (Gaussian *uncorrected* $p \approx 0.05$); (c) $u = 3.92$ (Gaussian $p \approx 0.05$); (d) $u = 3.76$ (tilted $p \approx 0.05$); (e) $u = 3.72$ (Gaussian $p \approx 0.1$); (f) $u = 3.58$ (tilted $p \approx 0.1$). P-values approximated by $\mathbb{E}EC$ under the Gaussian (7.10) and tilted theory with argument $\tau_u \hat{\theta}_u$ and spectral moment $\widehat{\lambda}_u$; see (7.12). Inference using (7.7) is similar and omitted. There is strong signal in the right eye region. At $p = 0.05$, the Gaussian theory identifies *only* the right eye (c); tilting finds activation on the upper lip (d). At $p = 0.1$, tilting detects signal above the left eye (f); the Gaussian approximation does not (e). Tilting *reduces* thresholds because $p_C > 1/2$, so $Z(t)$ exhibits *negative* skewness.

exponential portion of the EC densities. Additional corrections to the variance of the first derivative field and to the argument of the Hermite polynomial are of smaller order and may not be needed unless the expected EC over the full range of u is required. The examples considered indicate that tilting the spectral moment has a greater impact than tilting the Hermite argument. Our results support RS's recommendation to employ a tilted expansion when the additional computation is feasible; we add that the benefit is greatest if the field has nonzero skewness and/or if $u \to \infty$ (is large) along with n.

In addition to the bubbles test statistic field presented above, we have begun to apply the theory of this article to lesion density maps from neuroimaging [12], which can be non-Gaussian in a similar fashion. Handling this and further applications will require generalizing these results to certain nonisotropic and nonstationary CLRF's; this work is already underway. As discussed in Section 4.4, it would also be desirable to rigourously extend the proof in [37] of the EC heuristic to the asymptotically Gaussian case. In a forthcoming paper we shall explore a geometric interpretation to the tilted EC densities in the spirit of [35]. This will tie into similar investigations of 'CL-related' processes, including asymptotically χ^2, t and F random fields.

A   Appendix

A.1   Tilted moment calculations

Proof of (3.6) *and* (3.7). It is elementary that

$$\nu = \dot{\mu}_k + \begin{pmatrix} \dot{\sigma}_k & \dot{\Sigma}_{1,k} & \cdots & \dot{\Sigma}_{k-1,k} \end{pmatrix} V_k^{-1} \begin{pmatrix} 0 \\ -\dot{\mu}|_{k-1} \end{pmatrix},$$

and

$$\eta^2 = \dot{\Sigma}_{k,k} - \begin{pmatrix} \dot{\sigma}_k & \dot{\Sigma}_{1,k} & \cdots & \dot{\Sigma}_{k-1,k} \end{pmatrix} V_k^{-1} \begin{pmatrix} \dot{\sigma}_k \\ \dot{\Sigma}_{1,k} \\ \vdots \\ \dot{\Sigma}_{k-1,k} \end{pmatrix},$$

where V_k is as in (4.12). Letting $W_k = \dot{\Sigma}|_{k-1} - \frac{1}{\tau^2}\dot{\sigma}|_{k-1}\dot{\sigma}|'_{k-1}$, one has

$$V_k^{-1} = \frac{1}{\tau^2} \begin{pmatrix} 1 + \frac{1}{\tau^2}\dot{\sigma}|'_{k-1}W_k^{-1}\dot{\sigma}|_{k-1} & -\dot{\sigma}|'_{k-1}W_k^{-1} \\ -W_k^{-1}\dot{\sigma}|_{k-1} & \tau^2 W_k^{-1} \end{pmatrix}$$

$$= \frac{1}{\tau^2} \begin{pmatrix} 1 + O\left(n^{-2\alpha}\right) & O\left(n^{-\alpha}\right) \\ O\left(n^{-\alpha}\right) & \tau^2\left(I + O\left(n^{-\alpha}\right)\right) \end{pmatrix}. \tag{A.1}$$

Using (3.5), we have

$$\nu = \dot{\mu}_k + O\left(n^{-2\alpha}\right) = O\left(n^{1/2-2\alpha}\right)$$

and

$$\eta^2 = \dot{\Sigma}_{k,k} + O\left(n^{-2\alpha}\right) = 1 + O\left(n^{-\alpha}\right).$$

$\square$

Proof of (3.8) *and* (3.9). Write

$$\mathbb{E}_{\phi_k}\left\{ M_{ij} \left| Z = u, \dot{Z}|_{k-1} = 0, Z_k = x \right. \right\} = \ddot{\mu}_{ij} + V_{ij}V_{k+1}^{-1}v,$$

where

$$V_{ij} = \begin{pmatrix} \ddot{\sigma}_{ij} & \dddot{\Sigma}_{ij,1} \cdots \dddot{\Sigma}_{ij,k} \end{pmatrix}; \quad v = \begin{pmatrix} 0 & -\dot{\mu}|'_{k-1} & x - \dot{\mu}_k \end{pmatrix}'.$$

Using (A.1) together with (3.5) establishes (3.8). Towards (3.9), write

$$\mathsf{Cov}_{\phi_k}\left\{M_{ij}, M_{lm}\,\middle|\,Z = u, \dot{Z}|_{k-1} = 0, Z_k = x\right\}$$

$$= \ddot{\Sigma}_{ij,lm} - V_{ij}V_{k+1}^{-1}V_{lm}'$$

$$= \ddot{\Sigma}_{ij,lm} + O\left(n^{-2\alpha}\right) \tag{A.2}$$

$$= \mathbb{E}_{\phi_k}\left\{M_{ij}M_{lm}\,\middle|\,Z = u, \dot{Z}|_{k-1} = 0, Z_k = x\right\}$$

$$- \mathbb{E}_{\phi_k}\left\{M_{ij}\,\middle|\,Z = u, \dot{Z}|_{k-1} = 0, Z_k = x\right\}\mathbb{E}_{\phi_k}\left\{M_{lm}\,\middle|\,Z = u, \dot{Z}|_{k-1} = 0, Z_k = x\right\}.$$

So

$$\mathbb{E}_{\phi_k}\left\{M_{ij}M_{lm}\,\middle|\,Z = u, \dot{Z}|_{k-1} = 0, Z_k = x\right\}$$

$$= \ddot{\Sigma}_{ij,lm} + O\left(n^{-2\alpha}\right) + \left(\ddot{\mu}_{ij} + c_{ij}x + O\left(n^{1/2-3\alpha}\right)\right)\left(\ddot{\mu}_{lm} + c_{lm}x + O\left(n^{1/2-3\alpha}\right)\right),$$

$$= \ddot{\Sigma}_{ij,lm} + O\left(n^{\max\{-2\alpha,1-5\alpha\}}\right) + \ddot{\mu}_{ij}\ddot{\mu}_{lm} + \left(\ddot{\mu}_{ij}c_{lm} + \ddot{\mu}_{lm}c_{ij}\right)x + c_{ij}c_{lm}x^2,$$

with constants as in (3.8). Noting that $1 - 4\alpha > \max\{-2\alpha, 1 - 5\alpha\}$ and $\ddot{\Sigma}_{ijlm} = \lambda_{ijlm}^M + O\left(n^{-\alpha}\right)$ completes the proof. $\qquad\square$

A.2 Proof of Lemma 2

By Taylor's Theorem, if n is large enough so that $|\theta/\sqrt{n}| \le \epsilon_s$, then

$$K_Z'(\theta) - \sigma^2\theta - k_3(Z)\frac{\theta^2}{2} - \cdots - k_s(Z)\frac{\theta^{s-1}}{(s-1)!} = \frac{\theta^s}{s!}K_Z^{(s+1)}(\theta^\star) \tag{A.3}$$

for some $\theta^\star \in (0, \theta)$. Note that $K_Z(\theta) = nK_W(\theta/\sqrt{n})$, and thus $K_Z^{(s+1)}(\theta^\star) = n^{1-(s+1)/2} \times K_W^{(s+1)}(\theta^\star/\sqrt{n})$. Then since $\theta^s n^{1-(s+1)/2} \leq cn^{1/2-s\beta}$, (3.2) holds with

$$c(s) = c \sup_{[-\epsilon_s, \epsilon_s]} K_W^{(s+1)}/s!.$$

We now consider the second statement. Note that $\hat{\theta}_u$ is unique under general conditions [14]. Moreover,

$$cn^{-\alpha} \sim n^{-1/2}u = n^{-1/2}K_Z'(\hat{\theta}_u) = K_W'(n^{-1/2}\hat{\theta}_u).$$

It follows (on letting n grow) that $\hat{\theta}_u = O\left(n^{1/2}\right)$, since the derivative of a convex function is nondecreasing. Without loss of generality, let n be such that $\hat{\theta}_u \leq \sqrt{n}$. Suppose that $\hat{\theta}_u > cn^{1/2-\beta}$ for some $\beta > 0$. Then for large n, by the integral form of the Taylor remainder,

$$
\begin{aligned}
n^{\beta-\alpha} > u\hat{\theta}_u^{-1} = K_Z'(\hat{\theta}_u)/\hat{\theta}_u &= \int_{v=0}^{1} K_Z''(v\hat{\theta}_u)\, dv \\
&= \int_{v=0}^{1} K_W''(v\hat{\theta}_u/\sqrt{n})\, dv \\
&= \sqrt{n}\hat{\theta}_u^{-1} \int_{v=0}^{\hat{\theta}_u/\sqrt{n}} K_W''(v)\, dv \\
&\geq \inf_{v\in[0,1]} K_W''(v).
\end{aligned}
$$

This implies that $\beta \geq \alpha$, since otherwise $\inf_{v\in[0,1]} K_W''(v) = 0$, in violation of the convexity of K_W (by assumption K_W'' is continuous). A similar conclusion is reached if we assume $-\hat{\theta}_u < -cn^{1/2-\beta}$; thus in particular $|\hat{\theta}_u| \leq cn^{1/2-\alpha}$. Applying (3.2) with $s = 2$, $\theta = \hat{\theta}_u$, and $\beta = \alpha$ finishes the proof. $\qquad\square$

A.3 Proof of Lemma 3

First we need

Lemma 8. *Let everything be as in Lemma 3. Then*

$$\mathbb{E}\,|\Delta| = \begin{cases} 0 & k = 2m-1 \\[2mm] \frac{(-1)^m (2m)!}{m!\,2^m} + p_{2m}(x) & k = 2m \end{cases}$$

as $n \to \infty$, where m is any positive integer and $p_{2m}(x)$ is a degree $2m$ polynomial with largest (constant) term of order $O\left(n^{\max\{-\alpha,\,1-4\alpha\}}\right)$.

Proof. This is merely an extension of AT's Lemma 11.5.2. The straightforward modification of their proof is left to the reader. □

Expanding the determinant (c.f. AT equation (11.5.13)), we have

$$|\Delta - bI| = (-1)^k \left\{ b^k + \sum_{j=1}^{k} (-1)^j S_j(\Delta) b^{k-j} \right\},$$

where $S_j(\Delta)$ is the sum of the $j \times j$ principal minors of Δ. From Lemma 8,

$$\begin{aligned}
\mathbb{E}|\Delta - bI| &= (-1)^k \left\{ b^k + \sum_{m=1}^{\lfloor k/2 \rfloor} \left[\binom{k}{2m} \frac{(-1)^m (2m)!}{m!\,2^m} + q_{2m}(x) \right] b^{k-2m} \right\} \\
&= (-1)^k \left\{ H_k(b) + \sum_{m=1}^{\lfloor k/2 \rfloor} q_{2m}(x) b^{k-2m} \right\} \\
&= (-1)^k H_k(b) \left\{ 1 + s_k(x) \right\}, \tag{A.4}
\end{aligned}$$

where $q_{2m}(x)$ is a sum of $\binom{k}{2m}$ of the p_{2m}'s, and $s_k(x)$ is a degree $2\lfloor k/2 \rfloor$ polynomial

with largest (constant) term of order $O\left(n^{\max\{\alpha-1,-2\alpha\}}\right)$. Put $A = \Delta - bI$. Using a basic linear algebra result (c.f. [19] Theorem 2.9), we can write

$$\det(A + \Xi_x) = \det(A) + \sum_{\substack{A^\star \in \mathcal{A}^\star(l,m) \\ m \geq 0, \, l \geq \max\{1,m\}}} \det(A^\star), \tag{A.5}$$

where $\mathcal{A}^\star(l,m)$ is the collection of $k \times k$ matrices composed from m columns of $x\,\Xi^1$, $l - m$ columns of Ξ^0, and $k - l$ columns of A. If we name the columns of these matrices by writing $A = (a_1 \, a_2 \cdots a_k)$ and $\Xi^i = (\xi^i_1 \, \xi^i_2 \cdots \xi^i_k)$, for $i = 0, 1$, then up to a multiple, $q^\star$ say, of -1, the determinants in the sum in (A.5) are of the form

$$x^m \det\left(a^\star_1 \, a^\star_2 \cdots a^\star_{k-l} \, \xi^{\star 0}_1 \cdots \xi^{\star 0}_{l-m} \, \xi^{\star 1}_1 \cdots \xi^{\star 1}_m\right),$$

where the * denotes that the original column indices have been permuted (e.g. $a^\star_1 = a_{\rho(1)}$ for some permutation ρ). So taking the convention $a^\star_i = (a^\star_{1,i} \cdots a^\star_{k,i})'$, etc., and using $(-1)^\pi$ as shorthand for the sign of a permutation π, the summation is

$$\sum_{m=0}^{k} x^m \sum_{l=\max(1,m)}^{k} \sum_{A^\star \in \mathcal{A}^\star(l,m)} (-1)^{q^\star}$$

$$\times \sum_{\pi \in S_k} (-1)^\pi a^\star_{\pi(1),1} \cdots a^\star_{\pi(k-l),k-l} \, \xi^{\star 0}_{\pi(k-l+1),1} \cdots \xi^{\star 0}_{\pi(k-m),l-m} \, \xi^{\star 1}_{\pi(k-m+1),1} \cdots \xi^{\star 1}_{\pi(k),m}$$

$$= \sum_{m=0}^{k} x^m \sum_{l=\max(1,m)}^{k} \sum_{A^\star \in \mathcal{A}^\star(l,m)} \sum_{\pi \in S_k} c^\star_{lm}(\pi) a^\star_{\pi(1)1} \cdots a^\star_{\pi(k-l),k-l}.$$

for some constants $c^\star_{lm}(\pi) = O\left(n^{(l-m)(1/2-2\alpha)-m\alpha}\right)$. Consider the innermost summand. Ignoring the constant, it is a product of length $k - l$ with some factors of the form $(\Delta_{ii} - b)$, and others of the form Δ_{ij}. Multiplying this out and taking its expectation, with the help of Lemma 11.5.1 in AT, yields a polynomial in x of degree $2\lfloor(k-l)/2\rfloor$ with largest (constant) coefficient $O\left(n^{d_\pi(1/2-\alpha)}\right)$, where $d_\pi \leq k - l$. Note that there exist π

for which $d_\pi = k - l$. Therefore, for some (l, m)-dependent, $O(1)$, degree $2\lfloor (k-l)/2 \rfloor$ polynomials g_{klm}, we have

$$
\mathbb{E} \det \left(A + \Xi_x \right) = \mathbb{E} \det(A)
$$

$$
+ \sum_{m=0}^{k} x^m \sum_{l=\max(1,m)}^{k} O \left(n^{(k-l)(1/2-\alpha)+(l-m)(1/2-2\alpha)-m\alpha} \right) g_{klm}(x)
$$

$$
= \mathbb{E} \det(A) + \sum_{m=0}^{k} x^m O \left(n^{(k-m)(1/2-\alpha)} \right) \sum_{l=\max(1,m)}^{k} O \left(n^{-l\alpha} \right) g_{klm}(x)
$$

$$
= \mathbb{E} \det(A) + O \left(n^{k(1/2-\alpha)} \right) \sum_{m=0}^{k} c_m x^m, \tag{A.6}
$$

where each of the c_m's is at most $O\left(n^{-\alpha} \right)$ [in particular $c_0 = O\left(n^{-\alpha} \right)$]. So the second term in (A.6) behaves like $H_k(b) \times p_k(x)$, where $p_k(x)$ is a degree k polynomial with largest (constant) coefficient $O\left(n^{-\alpha} \right)$. Combining this with (A.4) finishes the proof. $\square$

A.4   Technical considerations for Theorem 4

A.4.1   Regularity conditions

We assume that the random vectors $\mathbf{W}^i = (W^i, \dot{W}^i|_k', \text{vec}[\sigma \ddot{W}^i|_{k-1} + W^i I / \sigma]')'$ are non-lattice. (While saddlepoint expansions have been established for lattice distributions, the resulting random fields could scarcely be imagined to be regular in the manner described below.) There are simple conditions which validate the tilted expansion (2.6) of the joint density $f(z, \dot{z}, m)$ of $\left(Z(t), \dot{Z}|_k(t)', M(t)' \right)$ (which exists, and is bounded and continuous, under said conditions). For example, it is sufficient that: (i) $\mathbf{W}^i$ come from an exponential family indexed by $\theta \in \Theta$ an open, convex subset of $\mathbb{R}^{d_k}$ (which appears tacitly in (2.2)); and (ii) $\mathbf{W}^i$ have characteristic function ψ_θ for which $|\psi_\theta(s)|^\nu$

is integrable for some $\nu = \nu(\theta) \geq 1$. For the latter, c.f. [5], Theorem 6.5, and their condition (c'). An extension of Theorem 4 to the case where the W_i's are nonidentically distributed could also be conceived; for regularity conditions the reader is referred to Theorem 6.6 of the same monograph.

The exponential family assumption guarantees that the $\mathbf{W}^i$ have finite joint cumulants of all orders, but to bound the error we shall assume in particular that for every $\nu \leq 4$ the cgf of $\mathbf{W}^i$ belongs to $C^\nu(V)$ for some neighbourhood V containing 0. We shall also require that for some finite constants $H > 0, (b_i)_{i\geq 2}$ and all $s \geq 2$,

$$\mathbb{E}\|\mathbf{W}^i\|^s \leq \frac{s!}{2}b_i^2 H^{s-2}. \tag{A.7}$$

As for the validity of (1.5), conditions are laid out in [1] and AT as follows.

A $T \in \mathbb{R}^N$ is compact and ∂T has zero Lebesgue measure.

B Z is almost surely *suitably regular* on T at the level u, *viz.*:

1. Z has continuous partial derivatives up to second order in an open neighbourhood of T;

2. there is no $t \in T$ such that $Z(t) - u = Z_1(t) = \cdots = Z_k(t) = 0$;

3. there is no $t \in \partial T$ and permutation ρ of $\{1, \ldots, k\}$ such that

$$Z(t) - u = Z_{\rho_1}(t) = \cdots = Z_{\rho_{k-1}}(t) = 0;$$

4. there is no $t \in T$ and permutation ρ of $\{1, \ldots, k\}$ such that

$$Z(t) - u = Z_{\rho_1}(t) = \cdots = Z_{\rho_{k-1}}(t) = \det\left(Z_{\rho_i\rho_j}(t)\right)_{i,j\leq k-1} = 0.$$

C The second-order derivatives of Z have finite variances.

D The joint density of $(Z, \dot{Z}|_k', \ddot{Z}|_{k-1}')$ is continuous in each argument.

E The conditional density of $(Z, \dot{Z}|_{k-1}')$ given $(Z_k, \ddot{Z}|_{k-1}')$ is bounded above.

F The moduli of continuity ω_i and ω_{ij} of the various derivatives satisfy

$$\mathbb{P}\left\{\max_{i,j \leq k}[\omega_i(\eta), \omega_{ij}(\eta)] > \epsilon\right\} = o\left(\eta^k\right)$$

as $\eta \to 0$ for every $\epsilon > 0$.

Note that under the oft-made Gaussian assumption, many of these conditions are trivially satisfied.

A.4.2 Bounding of the error term

Throughout this section, let $\widehat{f}(z, \dot{z}, m; \theta)$ denote the first order saddlepoint approximation to the density of $(Z, \dot{Z}|_k, \text{vec} M')'$ tilted at $\theta \in \mathbb{R}^{d_k}$, and $f = f_k$ the true density. The following lemma is not difficult, but since it is used several times in the arguments below we state it formally. We use superscript notation for derivatives of cgf's.

Lemma 9. *Let $\mathbf{W} \overset{\mathscr{D}}{=} \mathbf{W}^i$. Under the conditions of Theorem 4, there exist $\epsilon > 0$ and $C < \infty$ such that for every $1 \leq \nu \leq 4$,*

$$\max_{|\gamma|=\nu} \sup_{\|\theta\|<\epsilon} |K_{\mathbf{W}}^\gamma(\theta)| \leq C. \tag{A.8}$$

Proof. Since all joint cumulants of $\mathbf{W}^1$ of order ν are finite, we have $|K_{\mathbf{W}}^\gamma(0)| \leq K_\nu$ for all γ, $|\gamma| = \nu$ and some constant K_ν. The result then follows from the assumption that

$K_{\mathbf{W}}^{\gamma}$ is continuous in a neighbourhood of 0. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

We have already shown in Section 4 that the relative difference between $\widehat{\rho}_{k,n}(u)$ and the approximation

$$\int_{x=0}^{\infty} \int_{\mathbb{R}^{D_{k-1}}} x \det\left(-\frac{m-u/\sigma I}{\sigma}\right) \widehat{f}(u, 0', x, m; (\hat{\theta}_u, 0')) \, dm \, dx \qquad (\text{A.9})$$

is of the stated order as $n \to \infty$. Thus it remains to bound the remainder

$$\begin{aligned}
E_n &\triangleq \int_{x=0}^{\infty} \int_{\mathbb{R}^{D_{k-1}}} x \det\left(-\frac{m-u/\sigma I}{\sigma}\right) \left[f(u, 0', x, m) - \widehat{f}(u, 0', x, m; (\hat{\theta}_u, 0'))\right] \, dm \, dx \\
&= E_n^{(1)} + E_n^{(2)},
\end{aligned}$$

where

$$E_n^{(1)} \triangleq \iint_{\|(x,m)\| \leq u} \{\cdots\} \, dm \, dx, \quad E_n^{(2)} \triangleq \iint_{\|(x,m)\| > u} \{\cdots\} \, dm \, dx.$$

We give detailed proof only for $k = 1$; for general k the argument is completely analogous, but with a considerable investment in notation. When $k = 1$ the derivative is one-dimensional and the integral over m drops out of the expression (we mention what happens for general k at the end). We shall assume for simplicity (but no loss of generality) that $\sigma = 1$. First consider $E_n^{(2)}$:

$$\begin{aligned}
\left|E_n^{(2)}\right| &\leq \int_{x>u} x f(u, x) \, dx + (2\pi)^{-1} \tau_u^{-1} e^{-I_u(\hat{\theta}_u)} \int_{x>u} x \phi\left(x; \nu, \eta^2\right) \, dx \\
&\leq \int_{x>u} x f(u, x) \, dx + C_1 e^{-I_u(\hat{\theta}_u) - u^2/2}
\end{aligned} \qquad (\text{A.10})$$

for some constant C_1. Clearly the second term in (A.10) is smaller than $O\left(n^{-q}\right) \widehat{\rho}_{k,n}(u)$ for any $q > 0$. As for the first term, we can employ a multivariate Bernstein-type inequality [6, 45]. Let $V = \dot{Z} \mathbb{1}_{\{Z=u, \dot{Z}>u\}}$ and $V_\epsilon = \dot{Z} \mathbb{1}_{B_1(\epsilon)}(Z - u) \mathbb{1}_{\{\dot{Z}>u\}}$ for $\epsilon > 0$.

Then by dominated convergence,

$$\int_{x>u} x f(u,x)\,dx = \mathbb{E}_f V$$

$$= \lim_{\epsilon \to 0} \mathbb{E}_f V_\epsilon$$

$$= \lim_{\epsilon \to 0} \int_{v=0}^{\infty} \mathbb{P}_f\{V_\epsilon \geq v\}\,dv$$

$$= u \lim_{\epsilon \to 0} \mathbb{P}_f\{|Z - u| \leq \epsilon, \dot{Z} > u\} \tag{A.11}$$

$$+ \lim_{\epsilon \to 0} \int_{v=u}^{\infty} \mathbb{P}_f\{|Z - u| \leq \epsilon, \dot{Z} > v\}\,dv$$

$$\leq u \lim_{\epsilon \to 0} \mathbb{P}_f\left\{|Z + \dot{Z}| > 2u - \epsilon\right\}$$

$$+ \lim_{\epsilon \to 0} \int_{v=u}^{\infty} \mathbb{P}_f\left\{|Z + \dot{Z}| > u + v - \epsilon\right\}\,dv.$$

We can now use inequality (1.3) of [45] to bound the last line of (A.11). Putting $\xi_i = W_i + \dot{W}_i$, $b_i = \sqrt{2}$, $B_n = \sqrt{2n}$, that result gives:

$$\mathbb{P}_f\left\{|Z + \dot{Z}| > 2u - \epsilon\right\} \leq 2\exp\{-(2u - \epsilon)^2/4\}\left(1 + H(2u - \epsilon)/\sqrt{2n}\right)^{-1}$$

$$\xrightarrow{\epsilon \to 0} 2e^{-u^2}\left(1 + \sqrt{2}Hn^{-1/2}u\right)^{-1}$$

$$= 2e^{-u^2}\left(1 + O\left(n^{-\alpha}\right)\right);$$

$$\int_{v=u}^{\infty} \mathbb{P}_f\left\{|Z + \dot{Z}| > u + v - \epsilon\right\}\,dv$$

$$\leq \int_{v=u}^{\infty} 2\exp\{-(v + u - \epsilon)^2/4\}\left(1 + H(v + u - \epsilon)/\sqrt{2n}\right)^{-1}\,dv$$

$$\leq \int_{v=u}^{\infty} 2\exp\{-(v + u - \epsilon)^2/4\}\,dv$$

$$\xrightarrow{\epsilon \to 0} \int_{w=2u}^{\infty} 2\exp\{-w^2/4\}\,dw$$

$$\leq \int_{w=2u}^{\infty} 2w\exp\{-w^2/4\}\,dw = 4e^{-u^2}.$$

Therefore, $E_n^{(2)} \sim C_2 u e^{-u^2}$ for some constant C_2, which is asymptotically smaller than $n^{-q} \widehat{\rho}_{k,n}(u)$ for any q. Next, write

$$E_n^{(1)} = E_n^{(1,1)} + E_n^{(1,2)},$$

with

$$E_n^{(1,1)} \triangleq \int_{x=0}^{u} x \left[f(u,x) - \widehat{f}(u,x; \hat{\theta}_{u,x}) \right] dx;$$

$$E_n^{(1,2)} \triangleq \int_{x=0}^{u} x \left[\widehat{f}(u,x; \hat{\theta}_{u,x}) - \widehat{f}(u,x; (\hat{\theta}_u, 0)) \right] dx.$$

Here $\hat{\theta}_{u,x}$ denotes the 'correct' tilting parameter for the point (u,x), *i.e.*

$$\nabla K_{Z,\dot{Z}}(\theta)|_{\theta = \hat{\theta}_{u,x}} = (u,x)'. \tag{A.12}$$

The justification for splitting $E_n^{(1)}$ in this way is that by well-known results, the difference of densities in the first integrand has *relative* order n^{-1} over the domain of integration. More precisely, the corollary in the Appendix of [4] (see also [22]) says that

$$\sup_{z,\dot{z}} \widehat{f}(z,\dot{z}; \hat{\theta}_{z,\dot{z}})^{-1} \left[f(z,\dot{z}) - \widehat{f}(z,\dot{z}; \hat{\theta}_{z,\dot{z}}) \right] = O\left(n^{-1}\right), \tag{A.13}$$

where the supremum is over any domain for which $n^{-1/2} \hat{\theta}_{z,\dot{z}}$ remains in an (arbitrary) compact set. [Note that their result is framed in terms of $T = \sqrt{n}(Z, \dot{Z})$, but an easy transformation of the densities and parameters—in particular noting that if $t = \sqrt{n}(u,x)$ then (tilting under their respective distributions) $\hat{\theta}_t = n^{-1/2} \hat{\theta}_{u,x}$—recovers (A.13).] Since $x \leq u \sim n^{1/2-\alpha}$ on the domain of integration, a bivariate version of Lemma 2 (see also below) guarantees that $n^{-1/2} \|\hat{\theta}_{u,x}\| = O\left(n^{-\alpha}\right)$ is bounded for large n. It follows that $E_n^{(1,1)}$ behaves like $n^{-1} \rho_{k,n}(u)$ as $n \to \infty$.

Finally, consider $E_n^{(1,2)}$. Write $E_n^{(1,2)} = E_n^{(1,2,1)} + E_n^{(1,2,2)}$, where:

$$E_n^{(1,2,1)} \triangleq \int_{x=0}^u \frac{x \exp\{-I_u(\hat\theta_u)\} e^{-(x-\nu)^2/(2\eta^2)}}{2\pi |\Sigma_{u,x}|^{1/2}} \left(e^{J_n(x)} - 1 \right) dx;$$

$$E_n^{(1,2,2)} \triangleq \int_{x=0}^u (2\pi)^{-1} x \exp\{-I_u(\hat\theta_u)\} e^{-(x-\nu)^2/(2\eta^2)} \left(|\Sigma_{u,x}|^{-1/2} - (\tau_u \eta)^{-1} \right) dx;$$

$$J_n(x) \triangleq K_{Z,\dot Z}(\hat\theta_{u,x}) - K_Z(\hat\theta_u) - [\hat\theta'_{u,x}(u,x)' - u\hat\theta_u] + (x-\nu)^2/(2\eta^2);$$

$$\Sigma_{u,x} \triangleq \nabla\nabla' K_{Z,\dot Z}(\theta)|_{\theta=\hat\theta_{u,x}}.$$

Note that

$$K_{Z,\dot Z}(\hat\theta_{u,x}) - K_Z(\hat\theta_u)$$

$$= K_{Z,\dot Z}(\hat\theta_{u,x}) - K_{Z,\dot Z}(\hat\theta_u, 0)$$

$$= \left(\hat\theta_{u,x} - (\hat\theta_u,0)'\right)' \begin{pmatrix} u \\ \dot\mu \end{pmatrix} + \frac{1}{2}\left(\hat\theta_{u,x} - (\hat\theta_u,0)'\right)' \Sigma_{/} \left(\hat\theta_{u,x} - (\hat\theta_u,0)'\right)$$

$$+ \frac{1}{6}\left\{\left(\hat\theta_{(u),x} - \hat\theta_u\right)\frac{\partial}{\partial\theta} + \hat\theta_{u,(x)}\frac{\partial}{\partial\dot\theta}\right\}^3 K_{Z,\dot Z}(\theta,\dot\theta)\Bigg|_{(\theta,\dot\theta)=\theta^\star}, \quad (A.14)$$

where $\hat\theta_{u,x} = \left(\hat\theta_{(u),x}, \hat\theta_{u,(x)}\right)'$ and $\theta^\star$ is on the line joining $(\hat\theta_u, 0)$ to $\hat\theta_{u,x}$. With the obvious superscript notation for mixed partial derivatives of $K_{Z,\dot Z}$, the last term in (A.14) can be expanded as

$$\frac{1}{6}\left\{(\hat\theta_{(u),x} - \hat\theta_u)^3 K_{Z,\dot Z}^{(3,0)}(\theta^\star) + 3(\hat\theta_{(u),x} - \hat\theta_u)^2 \hat\theta_{u,(x)} K_{Z,\dot Z}^{(2,1)}(\theta^\star)\right.$$

$$\left. + 3(\hat\theta_{(u),x} - \hat\theta_u)\hat\theta_{u,(x)}^2 K_{Z,\dot Z}^{(1,2)}(\theta^\star) + \hat\theta_{u,(x)}^3 K_{Z,\dot Z}^{(0,3)}(\theta^\star)\right\}. \quad (A.15)$$

The following result will be helpful.

Lemma 10. *Suppose that $0 \le x \le u \sim cn^{1/2-\alpha}$ for some $c < \infty$. Let $\dot\kappa_3 = k_3(Z,Z,\dot Z)$,*

61

$\dddot{\kappa}_3 = k_3(Z, \dot{Z}, \ddot{Z})$ and $\dddot{\kappa}_3 = k_3(\dot{Z})$. Then there exist constants $c_1, c_2, c_3 < \infty$ and a quadratic polynomial $q(x)$ (independent of n) such that for all large n,

(a) $|\hat{\theta}_{(u),x} - u| \leq u c_1 n^{-\alpha}$

(b) $|\hat{\theta}_{u,(x)} - x + \dot{\kappa}_3 u^2/2 + \ddot{\kappa}_3 ux + \dddot{\kappa}_3 x^2/2| \leq c_2 n^{1/2 - 3\alpha}$

(c) $\left|\hat{\theta}_{(u),x} - \hat{\theta}_u\right| \leq q(x) n^{-\alpha}$

(d) $||\Sigma_{u,x}| - 1| \leq c_3 n^{-\alpha}$.

Proof. The proof of (a)-(b) essentially follows that of Lemma 2. Applying Taylor's theorem (about the origin) separately to each of the functions $d/d\theta K_{Z,\dot{Z}}(\theta, \dot{\theta})$ and $d/d\dot{\theta} K_{Z,\dot{Z}}(\theta, \dot{\theta})$ and using $\mathbb{E} Z\dot{Z} = 0$ gives:

$$u = \hat{\theta}_{(u),x} + \frac{\hat{\theta}_{(u),x}^2}{2\sqrt{n}} K_{W,\dot{W}}^{(3,0)}(\theta_1^\star) + \frac{\hat{\theta}_{u,(x)}^2}{2\sqrt{n}} K_{W,\dot{W}}^{(1,2)}(\theta_1^\star) + \frac{\hat{\theta}_{(u),x}\hat{\theta}_{u,(x)}}{\sqrt{n}} K_{W,\dot{W}}^{(2,1)}(\theta_1^\star); \qquad (A.16)$$

$$\begin{aligned}
x = \hat{\theta}_{u,(x)} &+ \hat{\theta}_{(u),x}^2 \dot{\kappa}_3/2 + \hat{\theta}_{u,(x)}^2 k_3(\dot{Z})/2 + \hat{\theta}_{(u),x}\hat{\theta}_{u,(x)}\ddot{\kappa}_3 \\
&+ \frac{1}{6n}\left[\hat{\theta}_{(u),x}^3 K_{W,\dot{W}}^{(3,1)}(\theta_2^\star) + 3\hat{\theta}_{(u),x}^2 \hat{\theta}_{u,(x)} K_{W,\dot{W}}^{(2,2)}(\theta_2^\star)\right. \\
&\left.+ 3\hat{\theta}_{(u),x}\hat{\theta}_{u,(x)}^2 K_{W,\dot{W}}^{(1,3)}(\theta_2^\star) + \hat{\theta}_{u,(x)}^3 K_{W,\dot{W}}^{(0,4)}(\theta_2^\star)\right],
\end{aligned} \qquad (A.17)$$

where $\theta_i^\star, i = 1, 2$ are on the line joining 0 and $n^{-1/2}\hat{\theta}_{u,x}$. An argument like the one in the proof of Lemma 2 shows that $\|\hat{\theta}_{u,x}\| = O\left(n^{1/2-\alpha}\right)$. Thus (A.16) can be written $u = \hat{\theta}_{(u),x} + O\left(n^{1/2-2\alpha}\right)$, yielding (a). Towards (b), (A.17) immediately gives $\hat{\theta}_{u,(x)} =$

$x(1 + O(n^{-\alpha}))$, and therefore

$$x = \hat{\theta}_{u,(x)} + u^2(1 + O(n^{-\alpha}))\dot{\kappa}_3/2 + x^2(1 + O(n^{-\alpha}))\ddot{\kappa}_3/2$$
$$+ ux(1 + O(n^{-\alpha}))\ddot{\kappa}_3 + O(n^{1/2-3\alpha}),$$

which implies (b). To prove (c), let $g(\theta, \dot{\theta}) = d/d\theta K_{Z,\dot{Z}}(\theta, \dot{\theta})$ and expand about $(\hat{\theta}_u, 0)$:

$$0 = g(\hat{\theta}_{u,x}) - g(\hat{\theta}_u, 0) = g^{(1,0)}(\hat{\theta}_u, 0)\left(\hat{\theta}_{(u),x} - \hat{\theta}_u\right) + g^{(0,1)}(\hat{\theta}_u, 0)\hat{\theta}_{u,(x)}$$
$$+ \frac{1}{2}\left(\hat{\theta}'_{u,x} - (\hat{\theta}_u, 0)\right)\begin{bmatrix} g^{(2,0)}(\theta^\star) & g^{(1,1)}(\theta^\star) \\ g^{(1,1)}(\theta^\star) & g^{(0,2)}(\theta^\star) \end{bmatrix}\left(\hat{\theta}_{u,x} - (\hat{\theta}_u, 0)'\right),$$

where $\theta^\star$ is on the line joining $(\hat{\theta}_u, 0)$ to $\hat{\theta}_{u,x}$. Note that $g^{(2,0)}(\theta^\star) = K_{Z,\dot{Z}}^{(3,0)}(\theta^\star) = n^{-1/2}K_{W,\dot{W}}^{(3,0)}(\theta^\star/\sqrt{n}) = O(n^{-1/2})$, and similarly for the other 2$^{\text{nd}}$ order derivatives of g. Therefore, using (a), (b) and Lemma 2 [in particular $|\hat{\theta}_{(u),x} - \hat{\theta}_u| = O(n^{1/2-2\alpha})$], we can bound the last term in the expansion by a quadratic in x whose largest coefficient behaves like $n^{-1/2}$. Call it $q_n^\star(x)$. Thus

$$\tau_u^2\left(\hat{\theta}_{(u),x} - \hat{\theta}_u\right) = -\dot{\sigma}\hat{\theta}_{u,(x)} - q_n^\star(x)$$
$$= -\dot{\sigma}\left[x - \dot{\kappa}_3 u^2/2 - \ddot{\kappa}_3 ux - \ddot{\kappa}_3 x^2/2 + O(n^{1/2-3\alpha})\right] - q_n^\star(x).$$

Using $\tau_u = 1 + O(n^{-\alpha})$ and $\dot{\sigma} = O(n^{-\alpha})$ gives result (c). Towards (d), let $d(\theta) =$

det $\nabla\nabla' K_{Z,\dot{Z}}(\theta)$. It is easily shown that

$$|\Sigma_{u,x}| = d(\hat{\theta}_{u,x})$$

$$= d(0) + \hat{\theta}'_{u,x}\nabla d(\theta^\star)$$

$$= 1 + n^{-1/2}\left\{ K^{(0,2)}_{W,\dot{W}}(\theta^\star)\left[\hat{\theta}_{(u),x}K^{(3,0)}_{W,\dot{W}}(\theta^\star) + \hat{\theta}_{u,(x)}K^{(2,1)}_{W,\dot{W}}(\theta^\star)\right] \right.$$

$$+ K^{(2,0)}_{W,\dot{W}}(\theta^\star)\left[\hat{\theta}_{(u),x}K^{(1,2)}_{W,\dot{W}}(\theta^\star) + \hat{\theta}_{u,(x)}K^{(0,3)}_{W,\dot{W}}(\theta^\star)\right]$$

$$\left. -2K^{(1,1)}_{W,\dot{W}}(\theta^\star)\left[\hat{\theta}_{(u),x}K^{(2,1)}_{W,\dot{W}}(\theta^\star) + \hat{\theta}_{u,(x)}K^{(1,2)}_{W,\dot{W}}(\theta^\star)\right] \right\}$$

where again $\theta^\star$ (different from the one in (A.15)) is on the line joining 0 and $n^{-1/2}\hat{\theta}_{u,x}$. Result (d) then follows by the boundedness of the second- and third-order derivatives of $K_{W,\dot{W}}$ on a compact set. $\qquad\square$

Replacing $\hat{\theta}_{(u),x}$ by $u + O\left(n^{1/2-3\alpha}\right)$ and $\hat{\theta}_{u,(x)}$ by $x - \dot{\kappa}_3 u^2/2 - \ddot{\kappa}_3 ux - \dddot{\kappa}_3 x^2/2 + O\left(n^{1/2-3\alpha}\right)$, (A.15) then simplifies to

$$a_n(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3;$$

$$a_0 = O\left(n^{1-6\alpha}\right), \quad a_1 = O\left(n^{1/2-4\alpha}\right), \quad a_2 = O\left(n^{-2\alpha}\right), \quad a_3 = O\left(n^{-1/2}\right).$$

Therefore

$$J_n(x) = u\left(\hat{\theta}_{(u),x} - \hat{\theta}_u\right) + \dot{\mu}\hat{\theta}_{u,(x)} + \frac{1}{2}\left(\hat{\theta}_{(u),x} - \hat{\theta}_u, \hat{\theta}_{u,(x)}\right)\begin{pmatrix} \tau_u^2 & \dot{\sigma} \\ \dot{\sigma} & \dot{\Sigma} \end{pmatrix}\begin{pmatrix} \hat{\theta}_{(u),x} - \hat{\theta}_u \\ \hat{\theta}_{u,(x)} \end{pmatrix}$$

$$+ a_n(x) - u\hat{\theta}_{(u),x} - x\hat{\theta}_{u,(x)} + u\hat{\theta}_u + \frac{1}{2\eta^2}(x^2 - 2\nu x + \nu^2)$$

$$= \dot{\mu}\hat{\theta}_{u,(x)} + \frac{1}{2}\left(\hat{\theta}_{(u),x} - \hat{\theta}_u\right)^2 + \hat{\theta}_{u,(x)}^2/2 + \dot{\sigma}\hat{\theta}_{u,(x)}\left(\hat{\theta}_{(u),x} - \hat{\theta}_u\right)$$

$$+ b_n(x) - x\hat{\theta}_{u,(x)} + x^2/2 - x\dot{\mu} + \dot{\mu}^2/2, \tag{A.18}$$

where now $b_n(x) = b_0 + b_1 x + b_2 x^2$ with $b_0 = O\left(n^{1-5\alpha}\right), b_1 = O\left(n^{1/2-3\alpha}\right), b_2 = O\left(n^{-\alpha}\right)$.

Note that we have used the facts that $\dot{\sigma} = O\left(n^{-\alpha}\right)$, $\eta = 1 + O\left(n^{-\alpha}\right)$, $\nu = \dot{\mu} + O\left(n^{-2\alpha}\right)$

and $x = O\left(n^{1/2-\alpha}\right)$ on the domain of integration. The terms of order $O\left(n^{1-4\alpha}\right)$ in

(A.18) require some care. From Lemma 10(b),

$$x - \hat{\theta}_{u,(x)} = \dot{\kappa}_3 u^2/2 + \ddot{\kappa}_3 ux + \dddot{\kappa}_3 x^2/2 + O\left(n^{1/2-3\alpha}\right)$$

$$\Rightarrow \frac{1}{2}(\hat{\theta}_{u,(x)} - x)^2 = \dot{\kappa}_3^2 u^4/8 + s_n(x), \tag{A.19}$$

where $s_n(x)$ is a quadratic whose largest coefficient is $O\left(n^{1/2-3\alpha}\right)$. Moreover, from (3.4)

and Lemma 2 we have

$$\dot{\mu}\left(\hat{\theta}_{u,(x)} - x\right) = -\dot{\kappa}_3^2 u^4/4 + t_n(x), \tag{A.20}$$

where $t_n(x)$ is another quadratic with $O\left(n^{1/2-3\alpha}\right)$ coefficients, as well as

$$\dot{\mu}^2/2 = \dot{\kappa}_3^2 \hat{\theta}_u^4/8 + O\left(n^{1-5\alpha}\right) = \dot{\kappa}_3^2 u^4/8 + O\left(n^{1-5\alpha}\right). \tag{A.21}$$

Finally, Lemma 10(c) guarantees that the terms

$$\left(\hat{\theta}_{(u),x} - \hat{\theta}_u\right)^2/2 + \dot{\sigma}\hat{\theta}_{u,(x)}\left(\hat{\theta}_{(u),x} - \hat{\theta}_u\right)$$

can be bounded by a cubic in x with $O\left(n^{1/2-3\alpha}\right)$ coefficients. Combining (A.18), (A.19),

(A.20) and (A.21) therefore gives (since $1/2 - 3\alpha < -\alpha$):

$$J_n(x) = J_{n,0} + J_{n,1}x + J_{n,2}x^2 + J_{n,3}x^3; \quad J_{n,r} = O\left(n^{-\alpha}\right), \quad r = 0, 1, 2, 3.$$

65

We now employ the estimate $|e^J - 1| \leq |J|e^J$, whereby (for large n)

$$
\left| E_n^{(1,2,1)} \right| \leq \exp\{-I_u(\hat{\theta}_u)\} \int_{x=0}^u \frac{x|J_n(x)|}{2\pi} \exp\left\{ J_n(x) - \frac{(x-\nu)^2}{2\eta^2} \right\} dx \times [1 + O\left(n^{-\alpha}\right)]
$$

$$
\leq 2 \exp\{-I_u(\hat{\theta}_u)\} \int_{x=0}^u \frac{|J_{n,0}|x + |J_{n,1}|x^2 + |J_{n,2}|x^3 + |J_{n,3}|x^4}{2\pi} \exp\{ax^2 + bx + c\} \, dx,
$$

where a, b, c are appropriately chosen, non-zero, bounded constants (with $-1 < a < -1/4$, say). Similarly, since $|\Sigma_{u,x}|^{-1/2}$ and $(\tau_u \eta)^{-1}$ can both be approximated by $1 + O\left(n^{-\alpha}\right)$, we have

$$
\left| E_n^{(1,2,2)} \right| \leq \exp\{-I_u(\hat{\theta}_u)\} \int_{x=0}^u x \exp\{a'x^2 + b'x\} \, dx \times O\left(n^{-\alpha}\right),
$$

with $a' < -1/4, |b'| < 1$, say. All the integrals in question are bounded (even when the upper limit is ∞), and therefore

$$
E_n^{(1,2)} = \exp\{-I_u(\hat{\theta}_u)\} \times O\left(n^{-\alpha}\right) = \hat{\rho}_{k,n}(u) \times O\left(n^{-\alpha}\right).
$$

This establishes (4.4), (4.6) and (4.7) when $k = 1$. For $k \geq 2$, the proof is virtually identical. Instead of the pair of random variables $(Z, \dot{Z})$ we have the pair of random vectors $(\mathbf{Z}, \dot{\mathbf{Z}}) \triangleq ([Z, \dot{Z}|'_{k-1}], [Z_k, M'])$. The 'correct' tilting parameter, that is the one obtained from tilting f at $(u, 0', x, m')$, can be denoted $\hat{\theta}_{u,0,x,m}$.

The error $E_n^{(2)}$ is not fundamentally changed, while we again split $E_n^{(1)}$:

$$
E_n^{(1,1)} \triangleq \int_{\|(x,m')\| \leq u} x \det\left(-[m - uI]\right) \left[f(u, 0', x, m') - \hat{f}(u, 0', x, m'; \hat{\theta}_{u,0,x,m}) \right] dm \, dx;
$$

66

$$E_n^{(1,2)} \triangleq \int_{\|(x,m')\| \le u} x \det\left(-[m - uI]\right)$$

$$\times \left[\widehat{f}(u, 0', x, m'; \hat{\theta}_{u,0,x,m}) - \widehat{f}(u, 0', x, m'; (\hat{\theta}_u, 0')) \right] dm\, dx$$

(with the usual convention of writing m for both the vector and the matrix). The treatment of $E_n^{(1,1)}$ is virtually unchanged. Observing that the determinant factor can be handled, for example, using

$$\left|\det(-[m - uI])\right| = |u|^{k-1} \left|\det(m/u - I)\right| \le |u|^{k-1} 2^{k-1},$$

it is not difficult to see that we may proceed as before for $E_n^{(1,2)}$ as well. The powers of $|u|$ which arise from bounding the determinant correspond to powers of u in $H_{k-1}(\tau_u \hat{\theta}_u)$. The additional integrals over $m \in \mathbb{R}^{k(k-1)/2}$ do not pose any problem. The only portion of the above proof requiring slight modification when $k \ge 2$ is the treatment of J_n, which now depends on both x and m. The following generalization of Lemma 10, proved in much the same way, gives us the necessary tools. We write $\hat{\theta}_{u,0,x,(m_{ij})}$ as shorthand for $\hat{\theta}_{u,0,x,m_{11},\ldots,(m_{ij}),\ldots,m_{k-1,k-1}}$, and (just in this section) write $\lambda_m \triangleq \lambda_{ij,ij}^M$ (c.f. Section 4).

Lemma 11. *Suppose that $0 \le x, (m_{ij})_{i \le j \le k-1} \le u \sim cn^{1/2-\alpha}$ for some $c < \infty$. Let $\mathbf{z} = (u, x, m')$, $\Sigma_{u,x,m} \triangleq \nabla\nabla' K_{Z,\dot{Z}|_k, M}(\theta)|_{\theta = \hat{\theta}_{u,0,x,m}}$,*

$$\dot{h}_3(u, x, m) \triangleq \sum \frac{1}{\gamma!} k_3\left([Z, Z_k, M]^{\gamma + \iota_2}\right) \mathbf{z}^{\gamma + \iota_2},$$

and

$$\ddot{h}_3^{ij}(u, x, m) \triangleq \sum \frac{1}{\gamma!} k_3\left([Z, Z_k, M]^{\gamma + e_{ij}}\right) \mathbf{z}^{\gamma + e_{ij}},$$

where e_{ij} is a row vector of all 0's except for a 1 in the position corresponding to m_{ij}. The sums are over all integral row vectors $\gamma \in \{0, 1, 2\}^{d_k - (k-1)}$ such that $|\gamma| = 2$. Other-

wise, notation is as in (2.3). Then there exist constants $c_1, c_3, c_4, c_6 < \infty$ and quadratic polynomials $q_2(x, m)$ and $q_5(x, m)$ (independent of n) such that for all large n,

(a) $|\hat{\theta}_{(u),0,x,m} - u| \le u c_1 n^{-\alpha}$

(b) $\|\hat{\theta}_{u,(0),x,m}\| \le q_2(x, m) n^{-\alpha}$

(c) $\left| \hat{\theta}_{u,0,(x),m} - x + \dot{h}_3(u, x, m) \right| \le c_3 n^{1/2 - 3\alpha}$

(d) $\left| \hat{\theta}_{u,0,x,(m_{ij})} - \lambda_m^{-1} m_{ij} + \lambda_m^{-1} \ddot{h}_3^{ij}(u, x, m) \right| \le c_4 n^{1/2 - 3\alpha} \quad \forall\, i \le j \le k - 1$

(e) $\left| \hat{\theta}_{(u),0,x,m} - \hat{\theta}_u \right| \le q_5(x, m) n^{-\alpha}$

(f) $\left| |\Sigma_{u,x,m}| - 1 \right| \le c_6 n^{-\alpha}.$

The analysis of J_n then proceeds as for $k = 1$, except that there are additional terms involving m and cumulants of M which must be bounded. To illustrate we consider $k = 2$, in which case m and M are scalar,

$$\dot{h}_3(u, x, m) = \left(\mathbb{E}[Z^2 Z_2] u^2 + \mathbb{E}[Z_2 M^2] m^2 + \mathbb{E}[Z_2^3] x^2 \right)/2$$
$$+ \mathbb{E}[Z Z_2 M] um + \mathbb{E}[Z Z_2^2] ux + \mathbb{E}[Z_2^2 M] mx,$$

and

$$\ddot{h}_3(u, x, m) \triangleq \ddot{h}_3^{1,1}(u, x, m) = \left(\mathbb{E}[Z^2 M] u^2 + \mathbb{E}[Z_2^2 M] x^2 + \mathbb{E}[M^3] m^2 \right)/2$$
$$+ \mathbb{E}[Z Z_2 M] ux + \mathbb{E}[Z M^2] um + \mathbb{E}[Z_2 M^2] xm.$$

Let $E(x)$ be defined by (3.8). We can use (A.2), Lemma 11 and a Taylor expansion of

$K_{Z,\dot{Z}|_k,M}$ to show that

$$J_n(x,m) = \left[K_{Z,\dot{Z}|_k,M}(\hat{\theta}_{u,0,x,m}) - K_Z(\hat{\theta}_u) \right] - (u,0,x,m)\hat{\theta}_{u,0,x,m} + u\hat{\theta}_u$$

$$+ (x-\nu)^2/(2\eta^2) + (m-E(x))^2/[2\ddot{\Sigma} + O\left(n^{-2\alpha}\right)]$$

$$= -(x - \dot{\mu}_2)\hat{\theta}_{u,0,(x),m} - (m - \ddot{\mu})\hat{\theta}_{u,0,x,(m)}$$

$$+ \frac{1}{2}\left\{ (\hat{\theta}_{(u),0,x,m} - \hat{\theta}_u)^2 \tau_u^2 + \hat{\theta}_{u,(0),x,m}^2 \dot{\Sigma}_{11} + \hat{\theta}_{u,0,(x),m}^2 \dot{\Sigma}_{22} + \hat{\theta}_{u,0,x,(m)}^2 \ddot{\Sigma} \right\}$$

$$+ (x-\nu)^2/(2\eta^2) + (m-E(x))^2/(2\ddot{\Sigma}) + r_n(x,m),$$

where $r_n(x,m)$ is a cubic having coefficients which are at most $O\left(n^{-\alpha}\right)$. Recall that $\ddot{\Sigma} = \lambda_m + O\left(n^{-\alpha}\right)$. The above can be simplified further:

$$J_n(x,m) = \frac{1}{2}\left(x - \hat{\theta}_{u,0,(x),m} \right)^2 + \frac{1}{2\lambda_m}\left(m - \lambda_m \hat{\theta}_{u,0,x,(m)} \right)^2$$

$$- \left(x - \hat{\theta}_{u,0,(x),m} \right)\dot{\mu}_2 - \left(m - \lambda_m \hat{\theta}_{u,0,x,(m)} \right)\frac{\ddot{\mu}}{\lambda_m} + \frac{1}{2}\left(\dot{\mu}_2^2 + \frac{\ddot{\mu}^2}{\lambda_m} \right) + s_n(x,m)$$

$$= \frac{1}{2}\left(\mathbb{E}[Z^2 Z_2]\frac{u^2}{2} \right)^2 + \frac{1}{2\lambda_m}\left(\mathbb{E}[Z^2 M]\frac{u^2}{2} \right)^2 - \mathbb{E}[Z^2 Z_2]\frac{u^2 \dot{\mu}_2}{2} - \mathbb{E}[Z^2 M]\frac{u^2 \ddot{\mu}}{2\lambda_m}$$

$$+ \frac{1}{2}\left(\dot{\mu}_2^2 + \frac{\ddot{\mu}^2}{\lambda_m} \right) + t_n(x,m),$$

where again all polynomial terms with $O\left(n^{-\alpha}\right)$ coefficients are absorbed into s_n and t_n. Now use $\dot{\mu}_2 = \mathbb{E}[Z^2 Z_2]u^2/2[1 + O\left(n^{-\alpha}\right)]$ and $\ddot{\mu} = \mathbb{E}[Z^2 M]u^2/2[1 + O\left(n^{-\alpha}\right)]$. It follows that on the domain of integration $\{\|(x,m)\| \leq u\}$, $J_n(x,m)$ can be bounded by a polynomial of degree 4 with coefficients at most $O\left(n^{-\alpha}\right)$. The arguments given for $k = 1$ then suffice to complete the proof. $\qquad \square$

A.4.3 Inversion of $\Sigma_/$

Although we have proved Theorem 4 by bounding the error in the approximation, it is instructive to examine the leading term in that error, *i.e.* the first term 'omitted' in the approximation (4.8). It turns out that we shall need to have a handle on $\Sigma_/^{-1}$. We proceed in a blockwise approach. Write

$$\Sigma_/ = \begin{pmatrix} A & B \\ B' & D \end{pmatrix}$$

with

$$A = \begin{pmatrix} \tau^2 & \dot{\sigma}' \\ \dot{\sigma} & \dot{\Sigma} \end{pmatrix}, \qquad B = \begin{pmatrix} \ddot{\sigma}' \\ \ddot{\Sigma}' \end{pmatrix}, \qquad D = \ddot{\Sigma}.$$

We have immediately that

$$\Omega_/ \triangleq \Sigma_/^{-1} = \begin{pmatrix} A^{-1} + FE^{-1}F' & -FE^{-1} \\ -E^{-1}F' & E^{-1} \end{pmatrix}$$

with $E = D - B'A^{-1}B$, $F = A^{-1}B$. By similar considerations,

$$A^{-1} = \frac{1}{\tau^2} \begin{pmatrix} 1 + \frac{1}{\tau^2}\dot{\sigma}'\left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2}\right)^{-1}\dot{\sigma} & -\dot{\sigma}'\left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2}\right)^{-1} \\ -\left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2}\right)^{-1}\dot{\sigma} & \tau^2\left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2}\right)^{-1} \end{pmatrix},$$

and therefore

$$F = \frac{1}{\tau^2} \begin{pmatrix} \left[1 + \frac{1}{\tau^2}\dot{\sigma}'\left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2}\right)^{-1}\dot{\sigma}\right]\ddot{\sigma}' - \dot{\sigma}'\left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2}\right)^{-1}\ddot{\Sigma}' \\ \left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2}\right)^{-1}\left(\tau^2\ddot{\Sigma}' - \dot{\sigma}\ddot{\sigma}'\right) \end{pmatrix} \triangleq \frac{1}{\tau^2} \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}.$$

Meanwhile,

$$E = \ddot{\Sigma} - \frac{1}{\tau^2} \left\{ \left[1 + \frac{1}{\tau^2} \dot{\sigma}' \left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2} \right)^{-1} \dot{\sigma} \right] \ddot{\sigma}\ddot{\sigma}' - \ddot{\sigma}\dot{\sigma}' \left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2} \right)^{-1} \ddot{\Sigma}' \right.$$
$$\left. + \ddot{\Sigma} \left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2} \right)^{-1} \left(\tau^2 \ddot{\Sigma}' - \dot{\sigma}\ddot{\sigma}' \right) \right\}.$$

We now partition the inverse as

$$\Omega_/ = \begin{pmatrix} \omega & \dot{\omega}' & \ddot{\omega}' \\ \dot{\omega} & \dot{\Omega} & \ddot{\Omega}' \\ \ddot{\omega} & \ddot{\Omega} & \ddot{\Omega} \end{pmatrix},$$

so that

$$\omega = \frac{1}{\tau^2} \left(1 + \frac{1}{\tau^2} \dot{\sigma}' \left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2} \right)^{-1} \dot{\sigma} \right) + \frac{1}{\tau^4} F_1 E^{-1} F_1',$$

$$\dot{\omega} = -\frac{1}{\tau^2} \left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2} \right)^{-1} \dot{\sigma} + \frac{1}{\tau^4} F_2 E^{-1} F_1',$$

$$\ddot{\omega} = -\frac{1}{\tau^2} E^{-1} F_1', \tag{A.22}$$

$$\ddot{\Omega} = -\frac{1}{\tau^2} E^{-1} F_2',$$

$$\dot{\Omega} = \left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2} \right)^{-1} + \frac{1}{\tau^4} F_2 E^{-1} F_2',$$

$$\ddot{\Omega} = E^{-1}.$$

Using (3.5) it also straightforward to see that

$$F_1 = O\left(n^{-\alpha} \right), \quad F_2 = O\left(n^{-\alpha} \right), \quad E = \ddot{\Sigma} + O\left(n^{-2\alpha} \right). \tag{A.23}$$

Moreover, since $\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2} = I + O\left(n^{-\alpha}\right)$, we also have $\left(\dot{\Sigma} - \frac{\dot{\sigma}\dot{\sigma}'}{\tau^2}\right)^{-1} = I + O\left(n^{-\alpha}\right)$. Plugging (A.23) into (A.22):

$$\omega = \frac{1}{\tau^2}\left\{1 + O\left(n^{-2\alpha}\right)\right\} = O(1), \quad \dot{\omega} = O\left(n^{-\alpha}\right), \quad \ddot{\omega} = O\left(n^{-\alpha}\right);$$

$$\dddot{\Omega} = O\left(n^{-\alpha}\right), \quad \dot{\Omega} = I + O\left(n^{-\alpha}\right), \quad \ddot{\Omega} = \ddot{\Sigma}^{-1}\left\{I + O\left(n^{-2\alpha}\right)\right\}.$$

A.4.4 Largest asymptotic coefficients

We first find the order of the coefficients of $Q_{k,3}(\mathbf{z} - \mu_/)$ in (2.6), where $\mathbf{z} = (u, 0', x, m')'$ so that

$$\mathbf{z} - \mu_/ = \begin{bmatrix} 0 \\ O\left(n^{1/2-2\alpha}\right) \\ x + O\left(n^{1/2-2\alpha}\right) \\ m + O\left(n^{1/2-2\alpha}\right) \end{bmatrix} \triangleq \begin{pmatrix} v \\ \dot{v} \\ \ddot{v} \end{pmatrix} = \mathbf{v}. \tag{A.24}$$

We shall use the usual subscript notation for the components of $\dot{v}$ and $\ddot{v}$. The form of the Edgeworth polynomial (c.f. [5], p.176) is

$$Q_{k,3}(\mathbf{v}) = Q_{k,3}(\mathbf{v}; \Omega_/) = \frac{1}{6}\sum_{|\gamma|=3}\widehat{k}_3\left([Z, \dot{Z}|_k, M]^\gamma\right)h_\gamma(\mathbf{v}; \Sigma_/), \tag{A.25}$$

where $\Omega_/ = \Sigma_/^{-1}$, $\gamma \in \{0, 1, 2, 3\}^{d_k}$ is an integral vector, $\widehat{k}_3(\cdot)$ is a tilted third order (mixed) cumulant—see (A.27) below—and $h_\gamma(\mathbf{v}; \Sigma_/)$ is the third covariant tensorial Hermite polynomial for γ with respect to the covariance $\Sigma_/$ evaluated at $\mathbf{v}$. Note that the sum is over all possible triples of random variables in the vector $(Z, \dot{Z}|_k', M')$ (with repetition). Define $\iota_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ to be a vector of 0's with a 1 in the jth position. For a generic covariance matrix Σ with inverse $\Sigma^{-1} = \left(\sigma^{ij}\right)_{i,j=1}^d$, and integral vector

$\gamma = \iota_i + \iota_j + \iota_k$ $(1 \le i \le j \le k \le d)$, the tensorial Hermite polynomial is given at $s = (s_1, \ldots, s_d)'$ by

$$h_\gamma(s; \Sigma) = \tilde{s}_i \tilde{s}_j \tilde{s}_k - \left(\sigma^{ij} \tilde{s}_k + \sigma^{ik} \tilde{s}_j + \sigma^{jk} \tilde{s}_i \right), \tag{A.26}$$

where

$$\tilde{s}_i = \sum_{j=1}^d \sigma^{ij} s_j.$$

It remains to specify the cumulants $\widehat{k}_3 \left([Z, \dot{Z}|_k, M]^\gamma \right)$ and the arguments $\tilde{v}$, $\tilde{v}_i$, and $\tilde{v}_{jk}$. Since we are interested in first-order calculations, we make do with the fact that

$$\widehat{k}_3 \left([Z, \dot{Z}|_k, M]^\gamma \right) = \left. \frac{\partial^3}{\partial \theta^\gamma} K_{Z, \dot{Z}|_k, M}(\theta) \right|_{\theta = (\hat{\theta}_u, 0')} \tag{A.27}$$

$$= k_3 \left([Z, \dot{Z}|_k, M]^\gamma \right) \left\{ 1 + O \left(n^{-\alpha} \right) \right\}.$$

$$= O \left(n^{-1/2} \right). \tag{A.28}$$

As for the second requirement, easy calculations give

$$\tilde{v} = \dot{\omega}_k x + \sum_{i,j \le k-1} \ddot{\omega}_{ij} m_{ij} + c,$$

$$\tilde{v}_i = \dot{\Omega}_{ik} x + \sum_{j,l \le k-1} \ddot{\Omega}_{i,jl} m_{jl} + c_i, \tag{A.29}$$

$$\tilde{v}_{ij} = \dddot{\Omega}_{k,ij} x + \sum_{l,p \le k-1} \ddot{\Omega}_{ij,lp} m_{lp} + c_{ij},$$

where $c = O \left(n^{1/2 - 3\alpha} \right)$, $c_i = O \left(n^{1/2 - 2\alpha} \right)$ and $c_{ij} = O \left(n^{1/2 - 2\alpha} \right)$ are constants. The size of the polynomial coefficients in each of these appears in Table 1. The ingredients are now in place to explicitly calculate each polynomial term. It turns out in fact that the largest coefficients in $Q_{k,3}$ are $O \left(n^{-2\alpha} \right)$ for the constant and quadratic terms, and

$O\left(n^{-1/2}\right)$ for the linear and cubic terms (including the mixed ones). The first term missing from approximation (4.8) is therefore a sum of integrals of the form

$$\int_{x=0}^{\infty} \int_{\mathbb{R}^{D_{k-1}}} c_{abcd}\, x^a m_{ij}^b m_{lp}^c m_{qs}^d \det\left(-\frac{m-u/\sigma I}{\sigma}\right) \widehat{f}_k(u,0',x,m)\, dm\, dx$$

$$= c_{abcd}\, (-1)^{k-1} \sigma^{-(k-1)} \exp\left\{-I_u(\hat{\theta}_u)\right\} \phi_{0,k-1}(u,0')$$

$$\times \mathbb{E}_{\phi_k}\left\{ Z_k^a \mathbb{1}_{[Z_k \geq 0]} M_{ij}^b M_{lp}^c M_{qs}^d \det\left(M - u/\sigma I\right) \Big| Z = u, \dot{Z}|_{k-1} = 0\right\},$$

where $1 \leq a, 0 \leq b, c, d \in \mathbb{Z}$, $a+b+c+d \leq 4$, and

$$c_{abcd} = \begin{cases} O\left(n^{-2\alpha}\right) & \text{if } a+b+c+d \text{ odd} \\ O\left(n^{-1/2}\right) & \text{if } a+b+c+d \text{ even.} \end{cases}$$

The constants c_{abcd} can be computed directly from (A.22), (A.26), and (A.29) if desired. Thus, computing the order $n^{-1/2}$ terms in the saddlepoint approximation to the EC density is reduced again to calculating the (conditional) expectation of a polynomial in jointly Gaussian random variables.

Table 1: Order of the coefficients in $\tilde{\mathbf{v}}$; $\square$ represents $O\left(n^{\square}\right)$

Variable	x	m_{lp}	Constant
$\tilde{v}$	$-\alpha$	$-\alpha$	$1/2 - 3\alpha$
$\tilde{v}_i$	0	$-\alpha$	$1/2 - 2\alpha$
$\tilde{v}_{jk}$	$-\alpha$	0	$1/2 - 2\alpha$

A.5 Proof of Corollary 5

Recall that

$$\tau_u^2 = \sigma^2 \left(1 + O\left(n^{-\alpha}\right)\right) \quad \text{and} \quad \hat{\theta}_u = \frac{u}{\sigma^2}\left(1 + O\left(n^{-\alpha}\right)\right),$$

whereby

$$\tau_u \hat{\theta}_u = \frac{u}{\sigma}\left(1 + O\left(n^{-\alpha}\right)\right).$$

It follows that

$$\epsilon_n[H_{k-1}(\tau_u \hat{\theta}_u), H_{k-1}(u/\sigma)] = -\alpha.$$

Also, $K_Z(\hat{\theta}_u) = \sigma^2 \hat{\theta}_u^2/2 + k_3(Z)\hat{\theta}_u^3/6 + O\left(n^{1-4\alpha}\right)$. From the saddlepoint equation (3.1) we have

$$u\hat{\theta}_u = \frac{u^2}{\sigma^2} - \frac{k_3(Z)u\hat{\theta}_u^2}{2\sigma^2} + O\left(n^{1-4\alpha}\right)$$

and

$$\frac{u}{\sigma^2} = \hat{\theta}_u \left(1 + \frac{k_3(Z)\hat{\theta}_u}{2\sigma^2} + O\left(n^{-2\alpha}\right)\right). \tag{A.30}$$

Squaring and inverting before dropping the $O\left(n^{1-4\alpha}\right)$ terms, (A.30) yields

$$\frac{\sigma^2 \hat{\theta}_u^2}{2} \doteq \frac{u^2}{2\sigma^2} - \frac{k_3(Z)u^2\hat{\theta}_u}{2\sigma^4}.$$

Therefore,

$$I_u(\hat{\theta}_u) = u\hat{\theta}_u - K_Z(\hat{\theta}_u)$$

$$= \frac{u^2}{\sigma^2} - \frac{k_3(Z)u\hat{\theta}_u^2}{2\sigma^2} - \left\{\frac{u^2}{2\sigma^2} - \frac{k_3(Z)u^2\hat{\theta}_u}{2\sigma^4} + \frac{k_3(Z)\hat{\theta}_u^3}{6}\right\} + O\left(n^{1-4\alpha}\right).$$

Since $u/\sigma^2 = \hat{\theta}_u[1 + O\left(n^{-\alpha}\right)]$, the last expression is just

$$\frac{u^2}{2\sigma^2} - \frac{k_3(Z)\hat{\theta}_u^3}{6} + O\left(n^{1-4\alpha}\right). \tag{A.31}$$

Note that the $O\left(n^{1-3\alpha}\right)$ term above does not disappear unless the skewness of the field is zero. So provided $\alpha > \frac{1}{3}$ and $k_3(Z) \neq 0$,

$$\epsilon_n\left[e^{-I_u(\hat{\theta}_u)}, e^{-u^2/(2\sigma^2)}\right] \overset{s}{=} 1 - 3\alpha;$$

whence $\epsilon_n[\hat{\rho}_{k,n}(u), \rho_{k,n}^{\gamma}(u)] = \max\{-\alpha, 1 - 3\alpha\} \overset{s}{=} 1 - 3\alpha$. Dropping the argument u and the subscripts, let

$$\frac{\rho^{\gamma} - \hat{\rho}}{\hat{\rho}} = c_n; \qquad \frac{\hat{\rho} - \rho}{\rho} = d_n,$$

where c_n is (sharply) $O\left(n^{1-3\alpha}\right)$ and d_n is (at most) $O\left(n^{1/2-2\alpha}\right)$. In particular $d_n = c_n \times o(1)$. Therefore

$$\begin{aligned}
\frac{\rho^{\gamma} - \rho}{\rho} &= \frac{\rho^{\gamma} - \hat{\rho}}{\hat{\rho}}\left(\frac{\hat{\rho}}{\rho}\right) + \frac{\hat{\rho} - \rho}{\rho} \\
&= c_n(1 + d_n) + d_n \\
&= c_n(1 + o(1)),
\end{aligned}$$

establishing (5.25). Observation (5.26) is immediate, and (5.27) follows from (A.31) with $b_n = k_3(Z)\hat{\theta}_u^3/6 + O\left(n^{1-4\alpha}\right)$. Note that $k_3(Z)$ may be positive or negative. $\qquad\square$

A.6 Proof of Lemma 7

Let $\hat{\theta}_W(t)$ be the solution to $K'_W(\theta) = t$ for $W = Y$ or Z. Then it is enough to show that

(a) $K_Z(\hat{\theta}_Z(z)) - \hat{\theta}_Z(z)z = K_Y(\hat{\theta}_Y(\sigma_n z + \mu_n)) - \hat{\theta}_Y(\sigma_n z + \mu_n)(\sigma_n z + \mu_n)$

(b) $K_Z''(\hat{\theta}_Z(z)) = K_Y''(\hat{\theta}_Y(\sigma_n z + \mu_n))/\sigma_n^2$

(c) $\|\widehat{f}_n\|_1 = \|\widehat{g}_n\|_1$.

We have

$$K_Z(\theta) = K_Y(\theta/\sigma_n) - (\mu_n/\sigma_n)\theta; \;\; K_Z'(\theta) = K_Y'(\theta/\sigma_n)/\sigma_n - \mu_n/\sigma_n;$$

$$K_Z''(\theta) = K_Y''(\theta/\sigma_n)/\sigma_n^2. \tag{A.32}$$

From the second of these relations it follows that $\hat{\theta}_Z(z) = \sigma_n \hat{\theta}_Y(\sigma_n z + \mu_n)$. Together with the first equality in (A.32) (and some rearrangement), this proves (a). Requirement (b) is also immediate using the third relation. But (a) and (b) imply that $\widehat{f}_n(z) \equiv \sigma_n \widehat{g}_n(\sigma_n z + \mu_n)$. So

$$\|\widehat{f}_n\|_1 = \int_{-\infty}^{\infty} \widehat{f}_n(z)\, dz = \sigma_n \int_{-\infty}^{\infty} \widehat{g}_n(\sigma_n z + \mu_n)\, dz = \int_{-\infty}^{\infty} \widehat{g}_n(y)\, dy = \|\widehat{g}_n\|_1.$$

$\square$

A.7   Calculations for the bubbles application

A.7.1   The conditional null cumulants of Z

Approximating sums by integrals (assuming σ is large) gives

$$\sum_p b_p(t)^j \approx \int \exp\left\{-j(x-t)'(x-t)/(2\sigma_B^2)\right\} dx = \left(\frac{2\pi\sigma_B^2}{j}\right)^{N/2}. \tag{A.33}$$

From H_0, (7.4) and the approximation $Y_{ip} \perp\!\!\!\perp Y_{iq}, p \neq q$ we have

$$k_j[\widetilde{Z}(t) \mid a] = n^{-j/2} \sum_{p=1}^{P} b_p(t)^j \sum_{i=1}^{n} a_i^j k_j(Y_{ip})$$

$$= n^{1-j/2} \frac{m}{P} \left(p_C^{1-j} + (-1)^j p_I^{1-j} \right) \sum_{p=1}^{P} b_p(t)^j.$$

$$\approx n^{1-j/2} \frac{m}{P} \left(p_C^{1-j} + (-1)^j p_I^{1-j} \right) \left(\frac{2\pi \sigma_B^2}{j} \right)^{N/2}.$$

Standardizing yields the approximate conditional null cumulants of $Z(t)$:

$$\kappa_j \triangleq k_j[\widetilde{Z} \mid a]/k_2[\widetilde{Z} \mid a]^{j/2} = \left(\frac{2^{j/2}}{j} \right)^{N/2} \frac{p_C^{1-j} + (-1)^j p_I^{1-j}}{\left(p_C^{-1} + p_I^{-1} \right)^{j/2}} \left(\frac{(2\pi \sigma_B^2)^{N/2} nm}{P} \right)^{1-j/2}.$$

Finally, substituting $B = nmF^N/P$ and $F = \sqrt{8 \log 2} \sigma_B$ gives (7.5).

A.7.2   Approximate second spectral moment

Here and in the following subsection all calculations are understood to be conditional on a, although we drop this conditioning from the notation. Let $Y_p = \sum_i a_i Y_{ip}$. The derivative random field is given by

$$\dot{Z}(t) = \left(n \mathsf{Var}[\widetilde{Z}(t)] \right)^{-1/2} \sum_{p=1}^{P} \dot{b}_p(t) Y_p. \tag{A.34}$$

Its variance, for t away from the boundary and assuming H_0, is

$$\text{Var}(\dot{Z}) \doteq n^{-1}\text{Var}[\widetilde{Z}(t)]^{-1}\sum_{p=1}^{P}\dot{b}_p\dot{b}_p'\text{Var}(Y_p)$$

$$= n^{-1}\left(n^{-1}\sum_{p=1}^{P}b_p^2\text{Var}(Y_p)\right)^{-1}\sum_{p=1}^{P}\dot{b}_p\dot{b}_p'\text{Var}(Y_p)$$

$$\approx \left(\int\exp\left\{-(x-t)'(x-t)/\sigma_B^2\right\}\,dx\right)^{-1}$$

$$\times\int\frac{(x-t)(x-t)'}{\sigma_B^4}\exp\left\{-\frac{1}{\sigma_B^2}(x-t)'(x-t)\right\}\,dx \qquad (A.35)$$

$$= \left(\pi\sigma_B^2\right)^{-N/2}\times\left(\pi\sigma_B^2\right)^{N/2}\frac{I_N}{2\sigma_B^2} = \frac{I_N}{2\sigma_B^2},$$

where in (A.35) we have approximated sums via integrals after substituting the Gaussian-shaped bubble. So under the null hypothesis, the isotropic random field $Z(t)$ has the second spectral moment $\lambda \approx 1/(2\sigma_B^2)$.

A.7.3 The tilted variance of $\dot{Z}$

Under the null hypothesis we have

$$\tau_u^2 = \text{Var}_{\hat{\theta}_u}(Z) \doteq [n\text{Var}(\widetilde{Z})]^{-1}\sum_p b_p^2\text{Var}_{\hat{\theta}_u}(Y_p).$$

Since the distribution of Y_p is (approximately) independent of p under the null, this gives

$$\text{Var}_{\hat{\theta}_u}(Y_p) \approx n\text{Var}(\widetilde{Z})\frac{\tau_u^2}{\sum_p b_p^2}.$$

Therefore,

$$
\begin{aligned}
\mathsf{Var}_{\hat{\theta}_u}(\dot{Z}) &= n^{-1}\mathsf{Var}(\widetilde{Z})^{-1} \sum_{p=1}^{P} \dot{b}_p \dot{b}_p' \mathsf{Var}_{\hat{\theta}_u}(Y_p) \\
&\approx \tau_u^2 \frac{\sum_p \dot{b}_p \dot{b}_p'}{\|b\|_2^2} \\
&= \tau_u^2 \frac{I}{2\sigma_B^2} \\
&= \tau_u^2 \lambda I.
\end{aligned}
\tag{A.36}
$$

Alternatively, one can show directly (dropping $O(1/n)$ terms and assuming m/P is small) that $\kappa_3 \doteq 2\sigma_B^2 \, k_3(Z, Z_j, Z_j)$; c.f. (4.20) and (7.14). For the Gaussian case $\tau_u^2 = 1$, so that (A.36) agrees with the untilted variance.

Acknowledgements

References

[1] ADLER, R. J. (1981). *The Geometry of Random Fields.* Wiley, New York.

[2] ADLER, R. J. AND TAYLOR, J. E. (2007). *Random Fields and Geometry.* Springer.

[3] ADOLPHS, R., GOSSELIN, F., BUCHANAN, T. W., TRANEL, D., SCHYNS, P., AND DAMASIO, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature 433*, 68–72.

[4] BARNDORFF-NIELSEN, O. AND COX, D. R. (1979). Edgeworth and Saddle-Point Approximations with Statistical Applications. *J. R. Statist. Soc. B 41*, 279–312.

[5] BARNDORFF-NIELSEN, O. E. AND COX, D. R. (1989). *Asymptotic Techniques for use in Statistics.* Monographs on Statistics and Applied Probability. Chapman & Hall, London.

[6] BERNSTEIN, S. N. (1946). *Probability Theory, 4th Edition (Russian).* M.-L. Gostechizdat, Ed., Moscow.

[7] BHATTACHARYA, R. N. AND RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

[8] CAO, J. AND WORSLEY, K. J. (1999). The detection of local shape changes via the geometry of Hotelling's t^2 fields. *Ann. Statist. 27,* 925–942.

[9] CARBONELL, F., GALAN, L., AND WORSLEY, K. J. (2005). The geometry of the Wilks's Lambda random field. Submitted.

[10] CATELAN, P., LUCCHIN, F., AND MATARRESE, S. (1988). Peak Number Density of Non-Gaussian Random-Fields. *Phys. Rev. Lett. 61,* 3, 267–270.

[11] CHAMANDY, N. (2007a). Inference for asymptotically Gaussian random fields. Ph.D. thesis, McGill University, Montréal.

[12] CHAMANDY, N. (2007b). Thresholding asymptotically Gaussian random fields using the tilted mean EC, with application to lesion density maps. *Proceedings of the 56th Session of the International Statistical Institute.* To appear.

[13] CHAMANDY, N., TAYLOR, J. E., AND WORSLEY, K. J. Euler characteristic densities for asymptotically Gaussian random fields: a geometric approach. In preparation.

[14] DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist. 25,* 631–650.

[15] DANIELS, H. E. (1980). Exact saddlepoint approximations. *Biometrika 67,* 1, 59–63.

[16] DANIELS, H. E. (1987). Tail Probability Approximations. *Internat. Statist. Review 55,* 1, 37–48.

[17] FIELD, C. AND RONCHETTI, E. (1990). *Small Sample Asymptotics.* Lecture Notes–Monograph Series, Vol. 13. Institute of Mathematical Statistics.

[18] GOSSELIN, F. AND SCHYNS, P. G. (2001). Bubbles, a technique to reveal the use of information in recognition. *Vision Research 41,* 2261–71.

[19] GRAGG, W. B. (1972). The Padé Table and its Relation to Certain Algorithms of Numerical Analysis. *SIAM review 14,* 1, 1–62.

[20] HADWIGER, H. (1957). *Vorlesüngen Über Inhalt, Oberfläche und Isoperimetrie.* Springer-Verlag, Berlin.

[21] Hoyle, F., Vogeley, M. S., Gott, J. R., Blanton, M., Tegmark, M., Weinberg, D., Bachall, N., Brinchman, J., and York, D. (2002). Two-Dimensional Topology of the SDSS. *Astrophys. J. 580*, 663–671.

[22] Jensen, J. L. (1988). Uniform Saddlepoint Approximations. *Adv. Appl. Probab. 20*, 3, 622–634.

[23] Loader, C. R. and Sun, J. (1997). Robustness of tube formulae. *J. Comp. Graph. Statist. 6*, 2, 242–250.

[24] Lugannani, R. and Rice, S. O. (1980). Saddle point approximations for the distribution of the sum of independent random variables. *Adv. Appl. Probab. 12*, 475–490.

[25] Matsubara, T. (1994). Analytic expression of the genus in weakly non-Gaussian field induced by gravity. *Astrophys. J. 434*, L43–L46.

[26] Matsubara, T. and Yokoyama, J. (1996). Genus statistics of the large-scale structure with non-Gaussian density fields. *Astrophys. J. 463*, 409–419.

[27] Nardi, Y., Siegmund, D., and Yakir, B. (2006). The Distribution of Maxima of Approximately Gaussian Random Fields. *Ann. Statist.* To appear.

[28] Padmanabhan, N., Schlegel, D. J., Seljak, U., and Makarov, A, *et. al.* (2006). The Clustering of Luminous Red Galaxies in the Sloan Digital Sky Survey Imaging Data. *Mon. Not. R. Astron. Soc.* Submitted, Available at `http://arxiv.org/abs/astro--ph/0605302`.

[29] Rabinowitz, D. and Siegmund, D. (1997). The approximate distribution of the maximum of a smoothed Poisson random field. *Statist. Sinica 7*, 167–180.

[30] Robinson, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *J. R. Statist. Soc. B 44*, 1, 91–101.

[31] Salmond, C. H., Ashburner, J., Vargha-Khadem, F., Connelly, A., Gadian, D. G., and Friston, K. J. (2002). Distributional Assumptions in Voxel-Based Morphometry. *Neuroimage 17*, 1027–1030.

[32] Schneider, R. (1993). *Convex bodies: the Brunn-Minkowski theory.* Volume 44 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge.

[33] Sun, J., Loader, C. R., and McCormick, W. P. (2000). Confidence bands in generalized linear models. *Ann. Statist. 28*, 2, 429–460.

[34] Taylor, J. E. (2003). Gaussian volumes of tubes, Euler characteristic densities and correlated conjunction. Proceedings of *Singular models and geometric methods in statistics*, 7–33. Institute of Statistical Mathematics.

[35] TAYLOR, J. E. (2006). A Gaussian kinematic formula. *Ann. Probab. 34*, 1, 122–158.

[36] TAYLOR, J. E. AND ADLER, R. J. (2003). Euler characteristics for Gaussian fields on manifolds. *Ann. Probab. 31*, 2, 533–563.

[37] TAYLOR, J. E., TAKEMURA, A., AND ADLER, R. J. (2005). Validity of the expected Euler characteristic heuristic. *Ann. Probab. 33*, 4, 1362–1396.

[38] TAYLOR, J. E. AND WORSLEY, K. J. (2006). Detecting sparse signal in random fields, with an application to brain mapping. *J. Am. Stat. Assoc.* To appear.

[39] TAYLOR, J. E. AND WORSLEY, K. J. (2007). Random fields of multivariate test statistics, with applications to shape analysis and fMRI. *Ann. Statist.* Submitted.

[40] TAYLOR, J. E., WORSLEY, K. J., AND GOSSELIN, F. Maxima of discretely sampled random fields, with an application to 'bubbles'. *Biometrika 94*, 1–18.

[41] VIVIANI, R., BESCHONER, P., EHRHARD, K., SCHMITZ, B., AND THÖNE, J. (2007). Non-normality and transformations of random fields, with an application to voxel-based morphometry. *Neuroimage 35*, 121–130.

[42] WORSLEY, K. J. (1994). Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Adv. Appl. Probab. 26*, 13–42.

[43] WORSLEY, K. J. (1995). Boundary corrections for the expected euler characteristic of excursion sets of random fields, with an application to astrophysics. *Adv. Appl. Probab. 27*, 943–959.

[44] WORSLEY, K. J. (1996). The geometry of random images. *Chance 9*, 1, 27–40.

[45] YURINSKIĬ, V. V. (1976). Exponential Inequalities for Sums of Random Vectors. *J. Multivariate Anal. 6*, 473–499.