Version of 21.7.01

The number of two dimensional maxima

A. D. Barbour[1] and A. Xia

Universität Zürich and University of Melbourne

Abstract. Let n points be placed uniformly at random in a subset A of the plane. A point is said to be maximal in the configuration if no other point is larger in both coordinates. We show that, for large n and for many sets A, the number of maximal points is approximately normally distributed. The argument uses Stein's method, and is also applicable in higher dimensions.

Keywords: Maximal points, Stein's method, Record values, Johnson–Mehl process.

AMS 2000 Subject Classifications: 60D05, 60F05, 60G70.

[1] Partially supported by Schweizerischer Nationalfondsprojekte 20-50686.97 and 20-61753.00

1. Introduction

Let n points $X_1, \ldots, X_n$ be placed uniformly at random in a convex set C in the plane. We use $\geq$ to denote the usual partial order, so that $u = (u^{(1)}, u^{(2)}) \geq v = (v^{(1)}, v^{(2)})$ if and only if $u^{(1)} \geq v^{(1)}$ and $u^{(2)} \geq v^{(2)}$. Let W be the number of indices i, $1 \leq i \leq n$, for which X_i is maximal, in the sense that $X_j \geq X_i$ for *no* $j \neq i$. The distribution of W is important in a number of applications, but in general it has been relatively little studied. The main contribution at this level of generality is that of Devroye (1993), who considered in some detail the asymptotics of the mean and variance of W for large n, and also proved that $W/\mathbb{E}W \xrightarrow{P} 1$.

There are two special cases in which much more is known. If $C = [0,1]^2$, then W has the same distribution as the number of record values in a sequence of n independent and identically distributed observations from a continuous distribution, which has been extensively investigated (Rényi, 1962). This can be seen by relabelling the points in decreasing order of their first coordinate, so that $X_i^{(1)} \geq X_j^{(1)}$ for all $1 \leq i \leq j \leq n$; then X_1 is a.s. always maximal, and X_i is maximal if and only if $X_i^{(2)} > X_j^{(2)}$ for all $1 \leq j < i$.

Results are also known for $C = T$, where T is the triangle $[0,1]^2 \cap \{(x,y) : x+y \leq 1\}$, and as a consequence also for polygons. If a Poisson number of points with mean n is placed uniformly at random in T, then W is asymptotically equivalent to the number of germs in a strip of a Johnson–Mehl growth process. This process, studied in particular in Holst, Quine and Robinson (1996) and in Chiu and Quine (1997), is defined in terms of a Poisson process Z on $\mathbb{R} \times \mathbb{R}_+$; its germs are the point process Ξ of 'uncovered' points of Z, where (w,t) is uncovered if the set $A_{(w,t)} := \{(y,s) \in \mathbb{R} \times \mathbb{R}_+ : |y-w| \leq t-s\}$ contains no points of Z. To establish this equivalence, first expand T by a factor of $\sqrt{n}$ to $T_n := [0,\sqrt{n}]^2 \cap \{(x,y) : x+y \leq \sqrt{n}\}$, and then change coordinates to (w,t) given by $w = x-y$, $t = \sqrt{n}-x-y$, so that the correspondingly transformed Poisson points are now the points of a unit intensity Poisson process on the triangle T_n' with vertices $(-\sqrt{n},0)$, $(\sqrt{n},0)$ and $(0,\sqrt{n})$. Points which are maximal according to the original definition are now uncovered in the Johnson–Mehl sense. The Johnson–Mehl process actually lives on the whole of $[-\sqrt{n},\sqrt{n}] \times [0,\infty)$, but an easy calculation such as that in Lemma 2.6 shows that the expected number of uncovered points in $[-\sqrt{n},\sqrt{n}] \times [0,\infty)$ that are not in T_n is of order $O(1)$, whereas both $\mathbb{E}W$ and $\operatorname{Var} W$ are of order $O(\sqrt{n})$.

A normal limit law for $(W - \mathbb{E}W)/\sqrt{\operatorname{Var} W}$ is proved by two different methods in the above papers, and the latter paper also has a generalization to higher dimensions. The asymptotic equivalence of the models with a random (Poisson) number of points and with a fixed number of points can be very simply shown: see Lemma 3.1 below. Bai *et al.* (2000)

also prove the normal limit law for W when $C = T$, with the number of points fixed, but by an ungainly method. In addition, they give very detailed asymptotics for the mean and variance when C is a polygon, and prove a number of intriguing Poisson approximations for special shapes of C which lead to relatively few maximal points. Some related variance considerations are to be found in Bai *et al.* (1998).

In the Johnson–Mehl processes considered in Chiu and Quine (1997), the Poisson process is allowed to have an intensity which depends on the t–coordinate. Generalizing in another direction, Baryshnikov (2000) allows $A_{(w,t)}$ to be of the form $\{(y,s) \in \mathbb{R} \times \mathbb{R}_+ : |y-w| \le f(t-s)\}$ for a wide range of functions f. We shall not pursue either of these variants, though our methods would be effective here, too; instead, we concentrate on the effect of moving from triangles and polygons to other shapes. We treat the case of a rather general C, not necessarily convex, allowing nearly as much generality for the boundary as Devroye (1993) does. We prove that the standardized variable $(W - \mathbb{E}W)/\sqrt{\operatorname{Var} W}$ typically has a normal limit, and give a rate of convergence with respect to the bounded Wasserstein distance: for probability measures P and Q on $\mathbb{R}$,

$$d_{BW}(P,Q) := \sup_{f \in F_{BL}} \left| \int f\,dP - \int f\,dQ \right|,$$

where

$$F_{BL} := \{f : \mathbb{R} \to \mathbb{R},\ |f(x) - f(y)| \le \min(1, |x-y|) \text{ for all } x, y \in \mathbb{R}\}.$$

Our argument is based on Stein's method, and is quite different from those referred to above. Both Chiu and Quine (1997) and Baryshnikov (2000) assume stationarity of the underlying Poisson process with respect to the first coordinate; this assumption, when translated back into our setting, would preclude the treatment of shapes other than triangles or polygons. Our method of proof also carries over easily to higher dimensions, though the geometrical considerations become more complicated there; we sketch an argument under fairly restrictive conditions on C. Curiously, asymptotic upper bounds for moments and related quantities are relatively easy to establish, but, because of the normalization, a lower bound of the correct order for the variance is also needed — as is generally the case when using Lyapounov bounds to prove convergence. This seems to be quite delicate in higher dimensions, and even in two dimensions for the Johnson–Mehl models studied in Chiu and Quine (1997), who assume as an explicit condition that such a lower bound exists; the lower bound is proved in their setting in two dimensions in Chiu and Lee (2000). We give a treatment in an arbitrary number of dimensions in Section 4.

2. The standard setting

Let f be a continuous decreasing function with $f(0) = 1$ and $f(1) = 0$, and let $A := \{(x, y) : 0 \le x \le 1, 0 \le y \le f(x)\}$ denote the part of the first quadrant lying below the curve $y = f(x)$. Let Z be a Poisson process on A with constant intensity λ, and define $\Xi(d\boldsymbol{\alpha}) := Z(d\boldsymbol{\alpha})I[Z(A_{\boldsymbol{\alpha}}) = 0]$, where, for $\boldsymbol{\alpha} = (x, y) \in A$,

$$A_{\boldsymbol{\alpha}} := \{(w, z) : x \le w \le f^{-1}(y), y \le z \le f(w)\} \setminus \{\boldsymbol{\alpha}\}. \tag{2.1}$$

Then Ξ is the process of maximal points of Z, having mean measure μ given by

$$\mu(d\boldsymbol{\alpha}) := \mathbb{E}\Xi(d\boldsymbol{\alpha}) = \lambda \exp\{-\lambda |A_{\boldsymbol{\alpha}}|\}\, d\boldsymbol{\alpha}, \tag{2.2}$$

where $|\cdot|$ denotes area when applied to subsets of the plane. Our aim in this section is to find the approximate distribution of $\Xi(A)$, for sufficiently well behaved functions f.

We first restrict attention to functions f satisfying the following conditions.

Assumptions F$'$:

F1$'$ The function f is continuous and decreasing on $[0, 1]$, and a.e. continuously differentiable on $(0, 1)$, with $f(0) = 1$ and $f(1) = 0$.

F2$'$ For some $c_1, c_2 > 0$, some $0 \le \tau < 2$ and for all $0 \le x \le 1$,

$$k_1(x) := \inf_{x/2 \le t \le 1} |f'(t)| \ge c_1 x^{\tau}; \qquad \sup_{0 \le t \le 1} |f'(t)| \le c_2. \tag{2.3}$$

F3$'$ For some $c_0 > 0$,

$$c_0 |f'(x)| \le \tilde{k}_1(x) := \inf_{x/2 \le t \le x} |f'(t)|, \qquad 0 < x \le 1. \tag{2.4}$$

These assumptions allow $|f'(x)|$ to approach zero as $x \to 0$, but ensure that f does not behave too irregularly in doing so, and that the behaviour at these points is more or less a power law with moderate exponent. If $|f'(x)| \sim e^{-1/x}$ as $x \to 0$, then Assumptions F2$'$ and F3$'$ are violated, and Assumption F3$'$ is violated, even though Assumption F2$'$ is not, if $|f'(x)| = 1/r!$ for $1/(r+1)! \le x < 1/r!$, for all large enough r. These assumptions are weakened to allow a symmetric treatment with $|f'(x)| \to \infty$ as $x \to 1$, following Corollary 2.5. The choices $f(0) = 1$ and $f(1) = 0$ are arbitrary standardizations; linearly changing scale on the x– or y–axis merely leads to a corresponding change in the value of λ.

The first step is to establish the asymptotic form of the mean under Assumptions F$'$, a result already to be found in Devroye (1993); we give a proof, since it illustrates our

subsequent arguments. We shall frequently use the following bounds for the function $c(k,\psi) := \min\{c_2, \tilde{c}(k,\psi)\}$, defined for $\psi, k \geq 0$, where

$$\tilde{c}(k,\psi) := \begin{cases} 8/(3e\psi^2), & \text{if } \psi^2 \leq 8/(3k); \\ k\exp\{-3\psi^2 k/8\}, & \text{if } 8/(3k) < \psi^2; \end{cases} \tag{2.5}$$

for this function c,

$$\begin{aligned} &\int_0^\infty c(k,\psi)\,d\psi \leq 2\sqrt{\frac{8c_2}{3e}}; \quad \int_0^\infty \psi c(k,\psi)\,d\psi \leq (8/3e)\{1 + (1/2)\log^+(ec_2/k)\}; \\ &\int_0^\infty \psi^2 c(k,\psi)\,d\psi \leq \{4 + e\sqrt{\pi}\}\frac{2}{3e}\sqrt{\frac{8}{3k}}, \end{aligned} \tag{2.6}$$

and both $c(k,\psi)$ and $\tilde{c}(k,\psi)$ are decreasing in k for each fixed ψ.

Lemma 2.1. *Under Assumptions F',*

$$\mu(A) = \lambda \int_0^1 dx \int_0^{f(x)} dy \exp\{-\lambda|A_{(x,y)}|\} \sim m\lambda^{1/2},$$

where $m := \sqrt{\pi/2}\int_0^1 |f'(x)|^{1/2}dx$. If f' is uniformly Lipschitz, more can be said: if $|f'(0)| > 0$, then $\mu(A) = m\lambda^{1/2} + O(1)$, and if $f'(0) = 0$ and $\tau = 1$ then $\mu(A) = m\lambda^{1/2} + O(\log\lambda)$.

Proof. We begin by writing

$$|A_{(x,y)}| = \int_x^{f^{-1}(y)} du \int_y^{f(u)} dv.$$

Defining ψ by $y = f(x + \psi\lambda^{-1/2})$, $0 \leq \psi \leq (1-x)\lambda^{1/2}$, and changing variables in the integral to ϕ and σ given by $v = f(x + \phi\lambda^{-1/2})$ and $\sigma = (u-x)\lambda^{1/2}$, it follows that

$$\lambda|A_{(x,y)}| = \int_0^\psi d\sigma \int_\sigma^\psi d\phi\,|f'(x + \phi\lambda^{-1/2})| = \int_0^\psi \phi|f'(x + \phi\lambda^{-1/2})|\,d\phi. \tag{2.7}$$

Thus we can write

$$\begin{aligned} \mu(A) &= \lambda^{1/2}\int_0^1 dx \int_0^{(1-x)\lambda^{1/2}} d\psi \Big|f'\Big(x + \frac{\psi}{\sqrt{\lambda}}\Big)\Big| \exp\left\{-\int_0^\psi \phi\Big|f'\Big(x + \frac{\phi}{\sqrt{\lambda}}\Big)\Big|d\phi\right\} \\ &= \lambda^{1/2}\int_0^1 dx \int_0^\infty d\psi\, g_\lambda^{(0)}(x,\psi), \end{aligned}$$

say, where

$$g^{(0)}(x,\psi) := \lim_{\lambda\to\infty} g_\lambda^{(0)}(x,\psi) = |f'(x)|e^{-\frac{1}{2}\psi^2|f'(x)|} \quad \text{a.e.,} \tag{2.8}$$

because f' is a.e. continuous, and hence satisfies

$$\int_0^1 dx \int_0^\infty d\psi\, g^{(0)}(x,\psi) = m. \tag{2.9}$$

Now, from Assumption F3$'$, it follows that

$$\begin{aligned} 0 \le g_\lambda^{(0)}(x,\psi) &\le \left| f'\Big(x + \frac{\psi}{\sqrt{\lambda}}\Big)\right| e^{-3\psi^2 \tilde{k}_1(x+\psi\lambda^{-1/2})/8} \\ &\le c_0^{-1} \tilde{k}_1(x+\psi\lambda^{-1/2}) e^{-3\psi^2 \tilde{k}_1(x+\psi\lambda^{-1/2})/8}, \end{aligned}$$

where $\tilde{k}_1(x) = \inf_{x/2 \le t \le x} |f'(t)| \ge k_1(x)$; hence, for all $\lambda > 0$, we have

$$0 \le c_0 g_\lambda^{(0)}(x,\psi) \le c(k_1(x),\psi). \tag{2.10}$$

Since $\int_0^\infty c(k,\psi)d\psi$ is uniformly bounded in k by (2.6), the first part follows from (2.8) and (2.9) by dominated convergence.

For the last part, we have $|f'(x) - f'(t)| \le K|x-t|$ for all $x, t \in [0,1]$. Then it is immediate that

$$\begin{aligned} &\lambda^{1/2} \left| \left| f'\Big(x + \frac{\psi}{\sqrt{\lambda}}\Big)\right| \exp\left\{ -\int_0^\psi \phi \left| f'\Big(x + \frac{\phi}{\sqrt{\lambda}}\Big)\right| d\phi \right\} - |f'(x)| \exp\{-\tfrac{1}{2}\psi^2 |f'(x)|\} \right| \\ &\quad \le K\psi \exp\{-\tfrac{1}{2}\psi^2 k_1(x)\}\{1 + \tfrac{1}{3}\psi^2 |f'(x)|\}. \end{aligned}$$

Also, for any $T, k > 0$,

$$\begin{aligned} &\int_0^T \psi e^{-\frac{1}{2}\psi^2 k}\, d\psi = k^{-1}(1 - e^{-\frac{1}{2}T^2 k}) \le k^{-1} \wedge \tfrac{1}{2}T^2; \\ &\int_0^T k\psi^3 e^{-\frac{1}{2}\psi^2 k}\, d\psi = \Big[-\psi^2 e^{-\frac{1}{2}\psi^2 k}\Big]_0^T + 2\int_0^T \psi e^{-\frac{1}{2}\psi^2 k}\, d\psi \le 2(k^{-1} \wedge \tfrac{1}{2}T^2). \end{aligned}$$

Now, if f' is uniformly Lipschitz and $|f'(0)| > 0$, one can take $\tau = 0$ in Assumption F2$'$, and then $|f'(x)| \le c_3 k_1(x)$ for some $c_3 < \infty$; if $f'(0) = 0$ and $\tau = 1$, then it is also the case that $|f'(x)| \le c_3 k_1(x)$ for some $c_3 < \infty$. Hence

$$\begin{aligned} |\mu(A) - m\lambda^{1/2}| &\le \int_0^1 dx \int_0^{(1-x)\lambda^{1/2}} d\psi\, K\psi \exp\{-\tfrac{1}{2}\psi^2 k_1(x)\}\{1 + \tfrac{1}{3}\psi^2 k_1(x) c_3\} \\ &\qquad + \int_0^1 dx \int_{(1-x)\lambda^{1/2}}^\infty |f'(x)| \exp\{-\tfrac{1}{2}\psi^2 |f'(x)|\}\, d\psi \\ &= O\left(\int_0^1 \left(\{k_1(x)^{-1} \wedge \lambda(1-x)^2\} + \sqrt{c_2} \right) dx \right) \\ &= \begin{cases} O(1) & \text{if } |f'(0)| > 0; \\ O(\log\lambda) & \text{if } f'(0) = 0 \text{ and } \tau = 1, \end{cases} \end{aligned}$$

completing the proof of the lemma.

Remark. In special cases, the integral for $\mu(A)$ can of course be explicitly calculated, as for the triangle, when

$$\mu(A) = \sqrt{\pi\lambda/2} - 1 + O(\lambda^{-1}e^{-\lambda/2});$$

see also Quine and Robinson (1990).

The next result concerns the asymptotic form of the variance. In order to state it, we need some further notation. For $\boldsymbol{\alpha} = (x, y)$, we set

$$N_{\boldsymbol{\alpha}} := \{\boldsymbol{\beta} \in A : A_{\boldsymbol{\beta}} \cap A_{\boldsymbol{\alpha}} \neq \emptyset\}; \quad N^0_{\boldsymbol{\alpha}} := N_{\boldsymbol{\alpha}} \setminus \{\boldsymbol{\alpha}\};$$
$$N^U_{\boldsymbol{\alpha}} := N^0_{\boldsymbol{\alpha}} \cap [0, x] \times [y, 1]; \quad N^L_{\boldsymbol{\alpha}} := N^0_{\boldsymbol{\alpha}} \cap [x, 1] \times [0, y]$$

[see Diagrams 1 and 2]. The importance of $N_{\boldsymbol{\alpha}}$ is that $\Xi|_{A\setminus N_{\boldsymbol{\alpha}}}$ is 'independent' of $\Xi(d\boldsymbol{\alpha})$, in the sense that

$$\mathcal{L}^{\boldsymbol{\alpha}}(\Xi|_{A\setminus N_{\boldsymbol{\alpha}}}) = \mathcal{L}(\Xi|_{A\setminus N_{\boldsymbol{\alpha}}}), \tag{2.11}$$

where $\mathcal{L}^{\boldsymbol{\alpha}}$ denotes the Palm distribution of Ξ at $\boldsymbol{\alpha}$; note also that $\boldsymbol{\beta} \in N_{\boldsymbol{\alpha}}$ if and only if $\boldsymbol{\alpha} \in N_{\boldsymbol{\beta}}$.

Lemma 2.2. *Under Assumptions F′,*

$$\begin{aligned}\mathrm{Var}\,(\Xi(A)) &= \mu(A) + 2\int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in A} 1_{N^L_{\boldsymbol{\alpha}}}(\boldsymbol{\beta})\mathbb{E}\{\Xi(d\boldsymbol{\alpha})\Xi(d\boldsymbol{\beta})\} - \int_{\boldsymbol{\alpha}\in A}\mu(N_{\boldsymbol{\alpha}})\mu(d\boldsymbol{\alpha}) \\ &\sim (2\log_e 2 - 1)m\lambda^{1/2}.\end{aligned}$$

Remark. If f' is Lipschitz and $|f'(0)| > 0$, the error in the asymptotic approximation is of order $O(1)$; horrendous calculations show that the error is of order $O(\log\lambda)$ if f' is Lipschitz, $f'(0) = 0$ and $\tau = 1$.

Proof. It is immediate that

$$\begin{aligned}\mathrm{Var}\,(\Xi(A)) &= \mathbb{E}\{(\Xi(A) - \mu(A))^2\} = \mathbb{E}\left\{\int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in A}\Xi'(d\boldsymbol{\alpha})\Xi'(d\boldsymbol{\beta})\right\} \\ &= \mathbb{E}\left\{\int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in A} 1_{N_{\boldsymbol{\alpha}}}(\boldsymbol{\beta})(\Xi(d\boldsymbol{\alpha}) - \mu(d\boldsymbol{\alpha}))(\Xi(d\boldsymbol{\beta}) - \mu(d\boldsymbol{\beta}))\right\},\end{aligned}$$

because, if $\boldsymbol{\beta} \notin N_{\boldsymbol{\alpha}}$, the point processes $\Xi|_{A_{\boldsymbol{\alpha}}}$ and $\Xi|_{A_{\boldsymbol{\beta}}}$ are independent. Now, on the set $(\{\boldsymbol{\beta} \geq \boldsymbol{\alpha}\} \cup \{\boldsymbol{\beta} \leq \boldsymbol{\alpha}\}) \setminus \{\boldsymbol{\beta} = \boldsymbol{\alpha}\}$, $\Xi(d\boldsymbol{\alpha})\Xi(d\boldsymbol{\beta})$ is the zero measure, by the definition of maximal points. Hence the measure $\Xi(d\boldsymbol{\alpha})\Xi(d\boldsymbol{\beta})$ only contributes to $\mathrm{Var}\,(\Xi(A))$ when $\boldsymbol{\alpha} = \boldsymbol{\beta}$, or when $\boldsymbol{\beta} \in N^L_{\boldsymbol{\alpha}}$, or when $\boldsymbol{\alpha} \in N^L_{\boldsymbol{\beta}}$, giving the positive parts of the variance formula. The negative part is immediate.

For the asymptotics, Lemma 2.1 gives $\mu(A) \sim m\lambda^{1/2}$. Taking the negative part next, observe that, for $\boldsymbol{\alpha} = (x, y)$,

$$\mu(N_{\boldsymbol{\alpha}}) = \lambda \left\{ \int_0^x dw \int_0^{f(x)} dz + \int_x^{x+\psi\lambda^{-1/2}} dw \int_0^{f(w)} dz \right\} \exp\{-\lambda |A_{(w,z)}|\}, \tag{2.12}$$

where, as in Lemma 2.1, $f(x + \psi\lambda^{-1/2}) = y$. Defining ρ by $z = f(w + \rho\lambda^{-1/2})$, we have

$$\lambda |A_{(w,z)}| = \int_0^\rho \phi \Big| f'\Big(w + \frac{\phi}{\sqrt{\lambda}}\Big) \Big| \, d\phi$$

as in (2.7). In the first double integral in (2.12), take $\zeta = (x - w)\lambda^{1/2}$, giving

$$\begin{aligned} &\lambda \int_0^x dw \int_0^{f(x)} dz \exp\{-\lambda |A_{(w,z)}|\} \\ &\quad = \int_0^{x\lambda^{1/2}} d\zeta \int_\zeta^{(1-x)\lambda^{1/2}+\zeta} d\rho \, \Big| f'\Big(x + \frac{(\rho - \zeta)}{\sqrt{\lambda}}\Big) \Big| \exp\left\{ - \int_0^\rho \phi \Big| f'\Big(x + \frac{(\phi - \zeta)}{\sqrt{\lambda}}\Big) \Big| \, d\phi \right\} \\ &\quad =: \int_0^\infty d\zeta \int_\zeta^\infty d\rho \, g_\lambda^{(1)}(x, \rho, \zeta); \end{aligned} \tag{2.13}$$

in the second double integral, take $\zeta = (w - x)\lambda^{1/2}$, giving

$$\begin{aligned} &\lambda \int_x^{x+\psi\lambda^{-1/2}} dw \int_0^{f(w)} dz \exp\{-\lambda |A_{(w,z)}|\} \\ &\quad = \int_0^\psi d\zeta \int_0^{(1-x)\lambda^{1/2}-\zeta} d\rho \, \Big| f'\Big(x + \frac{(\rho + \zeta)}{\sqrt{\lambda}}\Big) \Big| \exp\left\{ - \int_0^\rho \phi \Big| f'\Big(x + \frac{(\phi + \zeta)}{\sqrt{\lambda}}\Big) \Big| \, d\phi \right\} \\ &\quad =: \int_0^\psi d\zeta \int_0^\infty d\rho \, g_\lambda^{(2)}(x, \rho, \zeta). \end{aligned} \tag{2.14}$$

Since

$$\lim_{\lambda\to\infty} g_\lambda^{(1)}(x, \rho, \zeta) = \lim_{\lambda\to\infty} g_\lambda^{(2)}(x, \rho, \zeta) = g^{(0)}(x, \rho) \quad \text{a.e.}$$

and

$$\begin{aligned} &\int_0^1 dx \int_0^\infty d\psi \, g^{(0)}(x, \psi) \left\{ \int_0^\infty d\zeta \int_\zeta^\infty d\rho + \int_0^\psi d\zeta \int_0^\infty d\rho \right\} g^{(0)}(x, \rho) \\ &\quad = \int_0^1 dx \int_0^\infty d\psi \int_0^\infty d\rho \, (\psi + \rho) g^{(0)}(x, \psi) g^{(0)}(x, \rho) = 2m, \end{aligned}$$

it follows that

$$\int_{\boldsymbol{\alpha} \in A} \mu(N_{\boldsymbol{\alpha}}) \mu(d\boldsymbol{\alpha}) \sim 2m\lambda^{1/2} \tag{2.15}$$

if both $g_\lambda^{(0)}(x,\psi)g_\lambda^{(1)}(x,\rho,\zeta)$ and $g_\lambda^{(0)}(x,\psi)g_\lambda^{(2)}(x,\rho,\zeta)$ are bounded uniformly in λ by integrable functions.

Now $c_0 g_\lambda^{(0)}(x,\psi) \le c(k_1(x),\psi)$ from (2.10). Then, for $\rho \ge \zeta$ and $\zeta \le x\lambda^{1/2}$,

$$\int_0^\rho \phi\Big|f'\Big(x+\frac{(\phi-\zeta)}{\sqrt{\lambda}}\Big)\Big|\,d\phi \ge \left\{\int_{(\rho+\zeta)/2}^\rho + \int_{\zeta/2}^\zeta\right\}\phi\Big|f'\Big(x+\frac{(\phi-\zeta)}{\sqrt{\lambda}}\Big)\Big|\,d\phi$$
$$\ge \tfrac{3}{8}\{(\rho-\zeta)^2\tilde{k}_1(x+(\rho-\zeta)\lambda^{-1/2})+\zeta^2\tilde{k}_1(x)\},$$

and thus

$$\begin{aligned} g_\lambda^{(1)}(x,\rho,\zeta) &\le c_0^{-1}\tilde{k}_1(x+(\rho-\zeta)\lambda^{-1/2})\exp[-\tfrac{3}{8}\{(\rho-\zeta)^2\tilde{k}_1(x+(\rho-\zeta)\lambda^{-1/2})+\zeta^2\tilde{k}_1(x)\}] \\ &\le c_0^{-1}c(k_1(x),\rho-\zeta)e^{-3\zeta^2 k_1(x)/8}. \end{aligned} \tag{2.16}$$

But now

$$\int_0^1 dx\int_0^\infty d\psi\, c(k_1(x),\psi)\int_0^\infty d\zeta\int_\zeta^\infty d\rho\, c(k_1(x),\rho-\zeta)e^{-3\zeta^2 k_1(x)/8}$$
$$\le \int_0^1 dx\left\{\int_0^\infty d\psi\, c(k_1(x),\psi)\right\}^2\sqrt{\frac{2\pi}{k_1(x)}},$$

finite by (2.6) and Assumption F2′. Also, for $\rho,\zeta \ge 0$,

$$\begin{aligned} g_\lambda^{(2)}(x,\rho,\zeta) &\le c_0^{-1}\tilde{k}_1(x+(\rho+\zeta)\lambda^{-1/2})\exp\{-\tfrac{3}{8}\rho^2\tilde{k}_1(x+(\rho+\zeta)\lambda^{-1/2})\} \\ &\le c_0^{-1}\tilde{c}\left(\tilde{k}_1(x+(\rho+\zeta)\lambda^{-1/2}),\rho\right) \le c_0^{-1}\tilde{c}(k_1(x),\rho), \end{aligned}$$

since $\tilde{c}(k,\rho)$ is decreasing in k for ρ fixed; thus also, using Assumption F2′,

$$g_\lambda^{(2)}(x,\rho,\zeta) \le c_0^{-1}c(k_1(x),\rho). \tag{2.17}$$

Then

$$\int_0^1 dx\int_0^\infty d\psi\, c(k_1(x),\psi)\int_0^\psi d\zeta\int_0^\infty d\rho\, c(k_1(x),\rho)$$
$$\le \int_0^1 dx\left(\int_0^\infty \psi c(k_1(x),\psi)\,d\psi\right)\left(\int_0^\infty c(k_1(x),\rho)\,d\rho\right),$$

again finite by (2.6) and Assumption F2′. Hence, by dominated convergence, (2.15) follows.

The remaining term in the variance can be expressed as $2\int_{\boldsymbol{\alpha}\in A}\mu(d\boldsymbol{\alpha})\mathbb{E}^{\boldsymbol{\alpha}}(\Xi(N_{\boldsymbol{\alpha}}^L))$, where $\mathbb{E}^{\boldsymbol{\alpha}}$ denotes expectation with respect to the Palm measure of Ξ at $\boldsymbol{\alpha}$. Now, with

$\boldsymbol{\alpha} = (x, f(x+\psi\lambda^{-1/2}))$ as usual,

$$\begin{aligned}
\mathbb{E}^{\boldsymbol{\alpha}}(\Xi(N^L_{\boldsymbol{\alpha}})) &= \lambda \int_x^{x+\psi\lambda^{-1/2}} dw \int_0^y dz \exp\{-\lambda |A_{(w,z)} \setminus A_{\boldsymbol{\alpha}}|\} \\
&= \int_0^\psi d\zeta \int_{\psi-\zeta}^{(1-x)\lambda^{1/2}-\zeta} d\rho \left| f'\Big(x + \frac{(\rho+\zeta)}{\sqrt{\lambda}}\Big)\right| \\
&\qquad \exp\left\{ -\int_0^{\rho+\zeta-\psi} (\phi+\psi-\zeta) \left| f'\Big(x+\frac{(\phi+\psi)}{\sqrt{\lambda}}\Big)\right| d\phi \right\} \\
&=: \int_0^\psi d\zeta \int_{\psi-\zeta}^\infty d\rho\, g^{(3)}_\lambda(x,\rho,\zeta,\psi),
\end{aligned} \tag{2.18}$$

and

$$\begin{aligned}
g^{(3)}(x,\rho,\zeta,\psi) :&= \lim_{\lambda\to\infty} g^{(3)}_\lambda(x,\rho,\zeta,\psi) \\
&= |f'(x)| \exp\{-|f'(x)|[\tfrac{1}{2}(\rho+\zeta-\psi)^2 + (\psi-\zeta)(\rho+\zeta-\psi)]\} \quad \text{a.e.}
\end{aligned}$$

Then

$$\begin{aligned}
&\int_0^1 dx \int_0^\infty d\psi\, g^{(0)}(x,\psi) \int_0^\psi d\zeta \int_{\psi-\zeta}^\infty d\rho\, g^{(3)}(x,\rho,\zeta,\psi) \\
&= \int_0^1 dx \int_0^\infty d\psi \int_0^\psi d\chi \int_0^\infty d\theta\, |f'(x)|^2 \exp\{-\tfrac{1}{2}|f'(x)|(\psi^2+\theta^2+2\chi\theta)\} \\
&= \int_0^1 dx \int_0^\infty d\psi \int_0^\infty d\theta\, |f'(x)| \exp\{-\tfrac{1}{2}|f'(x)|(\psi^2+\theta^2)\}\theta^{-1}(1-e^{-\psi\theta|f'(x)|}) \\
&= \int_0^1 |f'(x)|^{1/2} dx \left(\int_0^\infty ds \int_0^\infty dt\, t^{-1}(1-e^{-st}) e^{-\frac{1}{2}(s^2+t^2)} \right) \\
&= m \log_e 2,
\end{aligned}$$

the last line after expanding $(1-e^{-st})$ as a power series. Hence

$$2\int_{\boldsymbol{\alpha}\in A} \mu(d\boldsymbol{\alpha}) \mathbb{E}^{\boldsymbol{\alpha}}(\Xi(N^L_{\boldsymbol{\alpha}})) \sim 2\log_e 2\, m\lambda^{1/2}, \tag{2.19}$$

provided that $g^{(0)}_\lambda(x,\psi) g^{(3)}_\lambda(x,\rho,\zeta,\psi)$ is bounded uniformly in λ by an integrable function.

Now $c_0 g^{(0)}_\lambda(x,\psi) \le c(k_1(x),\psi)$ from (2.10), as before, and, in $\rho \ge \psi-\zeta$ and $0 \le \zeta \le \psi$,

$$\begin{aligned}
g^{(3)}_\lambda(x,\rho,\zeta,\psi) &\le \left| f'\Big(x+\frac{(\rho+\zeta)}{\sqrt{\lambda}}\Big)\right| \exp\left\{ -\int_{(\rho+\zeta-\psi)/2}^{\rho+\zeta-\psi} \phi \tilde{k}_1(x+(\rho+\zeta)\lambda^{-1/2})\, d\phi \right\} \\
&\le c_0^{-1} \tilde{k}_1(x+(\rho+\zeta)\lambda^{-1/2}) \exp\{-\tfrac{3}{8}(\rho+\zeta-\psi)^2 \tilde{k}_1(x+(\rho+\zeta)\lambda^{-1/2})\} \\
&\le c_0^{-1} c(k_1(x), \rho+\zeta-\psi),
\end{aligned} \tag{2.20}$$

much as for (2.17). But

$$\int_0^1 dx \int_0^\infty d\psi\, c(k_1(x),\psi) \int_0^\psi d\zeta \int_{\psi-\zeta}^\infty d\rho\, c(k_1(x),\rho+\zeta-\psi)$$
$$= \int_0^1 dx \left(\int_0^\infty \psi c(k_1(x),\psi)\, d\psi\right)\left(\int_0^\infty c(k_1(x),\theta)\, d\theta\right),$$

finite once more by (2.6); (2.19) follows, and the proof is complete. Note also that, from (2.18), (2.20) and (2.6),

$$\mathbb{E}^{\boldsymbol{\alpha}}(\Xi(N_{\boldsymbol{\alpha}}^L)) \le 2c_0^{-1}\psi\sqrt{\frac{8c_2}{3e}}. \tag{2.21}$$

In order to prove a normal approximation, some quantities involving third moments must also be bounded. With τ as in Assumption F2$'$, define

$$\varepsilon(\lambda) := \begin{cases} \lambda^{-1/4}, & \text{if } 0 \le \tau < 1; \\ \lambda^{-1/4}\log\lambda, & \text{if } \tau = 1; \\ \lambda^{-\frac{3}{4}(2-\tau)/(2+\tau)}, & \text{if } 1 < \tau < 2. \end{cases} \tag{2.22}$$

Lemma 2.3. *Under Assumptions F$'$,*

$$(1)\quad \lambda^{-3/4}\int_{\boldsymbol{\alpha}\in A} \mu(N_{\boldsymbol{\alpha}})^2\mu(d\boldsymbol{\alpha}) = O(\varepsilon(\lambda));$$
$$(2)\quad \lambda^{-3/4}\int_{\boldsymbol{\alpha}\in A} \mathbb{E}\{\Xi(N_{\boldsymbol{\alpha}})^2\}\mu(d\boldsymbol{\alpha}) = O(\varepsilon(\lambda));$$
$$(3)\quad \lambda^{-3/4}\int_{\boldsymbol{\alpha}\in A} \mathbb{E}\{\Xi(N_{\boldsymbol{\alpha}})^2\Xi(d\boldsymbol{\alpha})\} = O(\lambda^{-1/4}).$$

Proof. From (2.13), (2.14), (2.16) and (2.17), and with $\boldsymbol{\alpha} = (x,y)$ and $y = f(x+\psi\lambda^{-1/2})$,

$$\begin{aligned}\mu(N_{\boldsymbol{\alpha}}) &\le c_0^{-1}\left\{\int_0^{x\lambda^{1/2}} d\zeta\, e^{-3\zeta^2 k_1(x)/8} + \psi\right\}\int_0^\infty c(k_1(x),\rho)\, d\rho \\ &= O(\psi + (x\lambda^{1/2})\wedge x^{-\tau/2}),\end{aligned} \tag{2.23}$$

using Assumption F2$'$. Hence, for Part (1), we simply use (2.6) to show that

$$\int_0^1 dx \int_0^\infty d\psi\, c(k_1(x),\psi)\{\psi + (x\lambda^{1/2})\wedge x^{-\tau/2}\}^2$$
$$= O\left(\int_0^1 dx\, \{x^{-\tau/2} + (x^2\lambda)\wedge x^{-\tau}\}\right),$$

in which the former integral is finite for all $\tau < 2$ and the latter for all $\tau < 1$. If $1 \le \tau < 2$, split the latter integral at $x = \lambda^{-1/(2+\tau)}$ to get

$$\int_0^1 dx\,(x^2\lambda) \wedge x^{-\tau} = O(\lambda^{(\tau-1)/(\tau+2)}), \quad 1 < \tau < 2;$$
$$\int_0^1 dx\,(x^2\lambda) \wedge x^{-1} = O(\log \lambda).$$

For Part (2), write

$$\int_{\boldsymbol{\alpha}\in A} \mathbb{E}\{\Xi(N_{\boldsymbol{\alpha}})^2\}\mu(d\boldsymbol{\alpha}) = \int_{\boldsymbol{\alpha}\in A} \mu(d\boldsymbol{\alpha}) \int_{\boldsymbol{\beta}\in N_{\boldsymbol{\alpha}}} \int_{\boldsymbol{\gamma}\in N_{\boldsymbol{\alpha}}} \mathbb{E}\{\Xi(d\boldsymbol{\beta})\Xi(d\boldsymbol{\gamma})\}.$$

That part of the integral arising when $\boldsymbol{\gamma} \notin N_{\boldsymbol{\beta}}$ is bounded by Part (1), since in this case $\mathbb{E}\{\Xi(d\boldsymbol{\beta})\Xi(d\boldsymbol{\gamma})\} = \mu(d\boldsymbol{\beta})\mu(d\boldsymbol{\gamma})$. If $\boldsymbol{\beta} \ne \boldsymbol{\gamma}$ and either $\boldsymbol{\beta} \ge \boldsymbol{\gamma}$ or $\boldsymbol{\beta} \le \boldsymbol{\gamma}$, there is no contribution, by the definition of maximal points; that from $\boldsymbol{\beta} = \boldsymbol{\gamma}$ is of order $O(\lambda^{1/2})$, as in (2.15). For the remainder, it is enough to bound $\int_{\boldsymbol{\beta}\in A} \mu(d\boldsymbol{\beta}) \int_{\boldsymbol{\alpha}\in N_{\boldsymbol{\beta}}} \mu(d\boldsymbol{\alpha})\mathbb{E}^{\boldsymbol{\beta}}(\Xi(N^L_{\boldsymbol{\beta}}))$. Taking $\boldsymbol{\beta} = (x, y)$ with $y = f(x + \psi\lambda^{-1/2})$, and using (2.21) and (2.23), it follows that

$$\begin{aligned}
&\int_{\boldsymbol{\beta}\in A} \mu(d\boldsymbol{\beta}) \int_{\boldsymbol{\alpha}\in N_{\boldsymbol{\beta}}} \mu(d\boldsymbol{\alpha})\mathbb{E}^{\boldsymbol{\beta}}(\Xi(N^L_{\boldsymbol{\beta}})) \\
&\quad = O\left(\lambda^{1/2} \int_0^1 dx \int_0^\infty d\psi\, c(k_1(x), \psi)\{\psi + x^{-\tau/2}\}\psi\right) \\
&\quad = O\left(\lambda^{1/2} \int_0^1 dx\, x^{-\tau/2}(1 + \log(1/x))\right) = O(\lambda^{1/2}),
\end{aligned}$$

completing the proof of Part (2).

For Part (3), write

$$\int_{\boldsymbol{\alpha}\in A} \mathbb{E}\{\Xi(N_{\boldsymbol{\alpha}})^2\Xi(d\boldsymbol{\alpha})\} = \int_{\boldsymbol{\alpha}\in A} \int_{\boldsymbol{\beta}\in N_{\boldsymbol{\alpha}}} \int_{\boldsymbol{\gamma}\in N_{\boldsymbol{\alpha}}} \mathbb{E}\{\Xi(d\boldsymbol{\alpha})\Xi(d\boldsymbol{\beta})\Xi(d\boldsymbol{\gamma})\}.$$

Once again, there is no contribution to the integral if any one of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ or $\boldsymbol{\gamma}$ dominates another, and the cases where two or three of them are equal make contributions bounded by Lemma 2.1 and (2.19). The remaining contribution is at most

$$\begin{aligned}
&6 \int_{\boldsymbol{\alpha}\in A} \int_{\boldsymbol{\beta}\in N^L_{\boldsymbol{\alpha}}} \int_{\boldsymbol{\gamma}\in N^L_{\boldsymbol{\beta}}} \mathbb{E}\{\Xi(d\boldsymbol{\alpha})\Xi(d\boldsymbol{\beta})\Xi(d\boldsymbol{\gamma})\} \\
&\quad = 6 \int_{\boldsymbol{\alpha}\in A} \int_{\boldsymbol{\beta}\in N^L_{\boldsymbol{\alpha}}} \int_{\boldsymbol{\gamma}\in N^L_{\boldsymbol{\beta}}} \mu(d\boldsymbol{\alpha})\mathbb{E}^{\boldsymbol{\alpha}}(\Xi(d\boldsymbol{\beta}))\mathbb{E}^{\boldsymbol{\beta}}(\Xi(d\boldsymbol{\gamma})),
\end{aligned}$$

this last because, with this arrangement of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, $\Xi|_{N^L_{\boldsymbol{\beta}}}$ is independent of $\Xi|_{N^U_{\boldsymbol{\beta}}}$ and $\boldsymbol{\alpha} \in N^U_{\boldsymbol{\beta}}$. Using (2.18), (2.20) and (2.21), this gives an order of

$$\begin{aligned}
&O\left(\lambda^{1/2}\int_0^1 dx \int_0^\infty d\psi\, c(k_1(x),\psi)\int_0^\psi d\zeta \int_{\psi-\zeta}^\infty d\rho\, \zeta c(k_1(x),\rho+\zeta-\psi)\right)\\
&= O\left(\lambda^{1/2}\int_0^1 dx \int_0^\infty d\psi\, \psi^2 c(k_1(x),\psi)\right)\\
&= O\left(\lambda^{1/2}\int_0^1 x^{-\tau/2}\, dx\right) = O(\lambda^{1/2}),
\end{aligned}$$

completing the proof.

After these preparations, we can prove the main theorem of the section.

Theorem 2.4. *Write* $\sigma^2 := \mathrm{Var}\,(\Xi(A))$ *and* $W_\lambda := \sigma^{-1}\Xi'(A)$. *Then, under Assumptions* F',

$$d_{BW}(\mathcal{L}(W_\lambda), \mathcal{N}(0,1)) = O(\varepsilon(\lambda)),$$

where $\varepsilon(\lambda)$ *is defined in (2.22).*

Proof. Using Stein's method as in Barbour, Karoński and Ruciński (1989, Section 2), it is enough to show that if $g : \mathbb{R} \to \mathbb{R}$ is bounded and has two bounded derivatives, then

$$|\mathbb{E}\{g'(W_\lambda) - W_\lambda g(W_\lambda)\}| = O(\|g''\|\varepsilon(\lambda)). \tag{2.24}$$

Now

$$W_\lambda g(W_\lambda) = \sigma^{-1}\int_{\boldsymbol{\alpha}\in A} \Xi'(d\boldsymbol{\alpha}) g(\sigma^{-1}[\Xi\{\boldsymbol{\alpha}\} + \Xi'(N^0_{\boldsymbol{\alpha}}) + \Xi'(A\setminus N_{\boldsymbol{\alpha}})]),$$

and hence, by Taylor's expansion,

$$\begin{aligned}
&\Big|\mathbb{E}(W_\lambda g(W_\lambda)) \qquad\qquad (2.25)\\
&- \frac{1}{\sigma}\mathbb{E}\left(\int_{\boldsymbol{\alpha}\in A} \Xi'(d\boldsymbol{\alpha})\left\{g(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}})) + \sigma^{-1}(\Xi\{\boldsymbol{\alpha}\} + \Xi'(N^0_{\boldsymbol{\alpha}}))g'(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}}))\right\}\right)\Big|\\
&\quad \le \tfrac{1}{2}\sigma^{-3}\|g''\|\mathbb{E}\left(\int_{\boldsymbol{\alpha}\in A}(\Xi(d\boldsymbol{\alpha}) + \mu(d\boldsymbol{\alpha}))(\Xi\{\boldsymbol{\alpha}\} + \Xi'(N^0_{\boldsymbol{\alpha}}))^2\right)\\
&\quad \le \tfrac{1}{2}\sigma^{-3}\|g''\|\mathbb{E}\left(\int_{\boldsymbol{\alpha}\in A}(\Xi(d\boldsymbol{\alpha}) + \mu(d\boldsymbol{\alpha}))(\Xi(N_{\boldsymbol{\alpha}}) + \mu(N_{\boldsymbol{\alpha}}))^2\right)\\
&\quad \le \sigma^{-3}\|g''\|\mathbb{E}\left(\int_{\boldsymbol{\alpha}\in A}(\Xi(d\boldsymbol{\alpha}) + \mu(d\boldsymbol{\alpha}))(\Xi(N_{\boldsymbol{\alpha}})^2 + \mu(N_{\boldsymbol{\alpha}})^2)\right)\\
&\quad =: \|g''\|\eta_1. \qquad\qquad (2.26)
\end{aligned}$$

Now, by (2.11),

$$\int_{\boldsymbol{\alpha}\in A} \mathbb{E}\{\Xi(d\boldsymbol{\alpha})g(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}}))\} = \int_{\boldsymbol{\alpha}\in A} \mu(d\boldsymbol{\alpha})\mathbb{E}g(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}})),$$

making $\mathbb{E}\left(\int_{\boldsymbol{\alpha}\in A}\Xi'(d\boldsymbol{\alpha})g(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}}))\right) = 0$. Then, since $\Xi\{\boldsymbol{\alpha}\}g'(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}})) = 0$ μ–a.e., we have

$$\int_{\boldsymbol{\alpha}\in A} \Xi'(d\boldsymbol{\alpha})\Xi\{\boldsymbol{\alpha}\}g'(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}})) = \int_{\boldsymbol{\alpha}\in A} \Xi(d\boldsymbol{\alpha})g'(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}})),\ \mu \text{ a.e.},$$

and thus, again by (2.11),

$$\sigma^{-2}\mathbb{E}\left(\int_{\boldsymbol{\alpha}\in A} \Xi'(d\boldsymbol{\alpha})\Xi\{\boldsymbol{\alpha}\}g'(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}}))\right) = \sigma^{-2}\int_{\boldsymbol{\alpha}\in A} \mu(d\boldsymbol{\alpha})\mathbb{E}g'(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}})). \tag{2.27}$$

For the remaining term in (2.25), again by Taylor's expansion,

$$\begin{aligned}
&\sigma^{-2}\Bigg|\mathbb{E}\left(\int_{\boldsymbol{\alpha}\in A} \Xi'(d\boldsymbol{\alpha})\Xi'(N^0_{\boldsymbol{\alpha}})g'(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}}))\right)\\
&\qquad - \mathbb{E}\left(\int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in N^0_{\boldsymbol{\alpha}}} \Xi'(d\boldsymbol{\alpha})\Xi'(d\boldsymbol{\beta})g'(\sigma^{-1}\Xi'(A\setminus\{N_{\boldsymbol{\alpha}}\cup N_{\boldsymbol{\beta}}\}))\right)\Bigg|\\
&\le \frac{\|g''\|}{\sigma^3}\mathbb{E}\left(\int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in N^0_{\boldsymbol{\alpha}}} (\Xi(d\boldsymbol{\alpha})+\mu(d\boldsymbol{\alpha}))(\Xi(d\boldsymbol{\beta})+\mu(d\boldsymbol{\beta}))\{\Xi(N_{\boldsymbol{\beta}}\setminus N_{\boldsymbol{\alpha}})+\mu(N_{\boldsymbol{\beta}}\setminus N_{\boldsymbol{\alpha}})\}\right)\\
&\le \frac{\|g''\|}{\sigma^3}\mathbb{E}\left(\int_{\boldsymbol{\beta}\in A} (\Xi(N_{\boldsymbol{\beta}})+\mu(N_{\boldsymbol{\beta}}))^2(\Xi(d\boldsymbol{\beta})+\mu(d\boldsymbol{\beta}))\right) \le 2\|g''\|\eta_1,
\end{aligned} \tag{2.28}$$

and, by (2.11),

$$\begin{aligned}
&\sigma^{-2}\mathbb{E}\left(\int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in N^0_{\boldsymbol{\alpha}}} \Xi'(d\boldsymbol{\alpha})\Xi'(d\boldsymbol{\beta})g'(\sigma^{-1}\Xi'(A\setminus\{N_{\boldsymbol{\alpha}}\cup N_{\boldsymbol{\beta}}\}))\right)\\
&\qquad = \sigma^{-2}\int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in N^0_{\boldsymbol{\alpha}}} \mathbb{E}\{\Xi'(d\boldsymbol{\alpha})\Xi'(d\boldsymbol{\beta})\}\mathbb{E}g'(\sigma^{-1}\Xi'(A\setminus\{N_{\boldsymbol{\alpha}}\cup N_{\boldsymbol{\beta}}\})).
\end{aligned} \tag{2.29}$$

Finally, when considering (2.27) and (2.29), note that

$$\begin{aligned}
&|\mathbb{E}g'(\sigma^{-1}\Xi'(A\setminus N_{\boldsymbol{\alpha}})) - \mathbb{E}g'(W_\lambda)| \le 2\sigma^{-1}\|g''\|\mu(N_{\boldsymbol{\alpha}});\\
&|\mathbb{E}g'(\sigma^{-1}\Xi'(A\setminus\{N_{\boldsymbol{\alpha}}\cup N_{\boldsymbol{\beta}}\})) - \mathbb{E}g'(W_\lambda)| \le 2\sigma^{-1}\|g''\|\{\mu(N_{\boldsymbol{\alpha}})+\mu(N_{\boldsymbol{\beta}})\},
\end{aligned} \tag{2.30}$$

so that, combining (2.25) – (2.30), it follows that

$$\begin{aligned}
&\left|\mathbb{E}(W_\lambda g(W_\lambda)) - \sigma^{-2}\left(\mu(A) + \int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in N^0_{\boldsymbol{\alpha}}} \mathbb{E}\{\Xi'(d\boldsymbol{\alpha})\Xi'(d\boldsymbol{\beta})\}\right)\mathbb{E}g'(W_\lambda)\right|\\
&\qquad = |\mathbb{E}(W_\lambda g(W_\lambda)) - \mathbb{E}g'(W_\lambda)| \le (3\eta_1+\eta_2+\eta_3)\|g''\|,
\end{aligned} \tag{2.31}$$

where η_1 is as in (2.26), $\eta_2 := 2\sigma^{-3}\int_{\boldsymbol{\alpha}\in A}\mu(N_{\boldsymbol{\alpha}})\mu(d\boldsymbol{\alpha})$ and

$$\begin{aligned}\eta_3 &:= 2\sigma^{-3}\int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in N^0_{\boldsymbol{\alpha}}}\mathbb{E}\{(\Xi(d\boldsymbol{\alpha})+\mu(d\boldsymbol{\alpha}))(\Xi(d\boldsymbol{\beta})+\mu(d\boldsymbol{\beta}))\}(\mu(N_{\boldsymbol{\alpha}})+\mu(N_{\boldsymbol{\beta}}))\\ &\le 4\sigma^{-3}\int_{\boldsymbol{\alpha}\in A}\mathbb{E}\{(\Xi(d\boldsymbol{\alpha})+\mu(d\boldsymbol{\alpha}))(\Xi(N^0_{\boldsymbol{\alpha}})+\mu(N_{\boldsymbol{\alpha}}))\}\mu(N_{\boldsymbol{\alpha}})\le 6\eta_1.\end{aligned}$$

Thus (2.31) is precisely as required in (2.24), provided that $\eta_1+\eta_2 = O(\varepsilon(\lambda))$. However, under Assumptions F′, from Lemma 2.2, $\sigma^3 \asymp \lambda^{3/4}$; from Lemma 2.3, $\eta_1 = O(\varepsilon(\lambda))$, and $\eta_2 = O(\lambda^{-1/4})$ from (2.15). The theorem is proved.

Remark. The theorem actually exploits a dependence structure similar to the 'finite dependence' of Chen (1986), and it is therefore to be expected that the same order of approximation can also be proved for Kolmogorov distance, using exactly the same estimates and a variant of Chen's theorem. For the triangle, this would then be slightly better than the order given in Chiu and Quine (1997, (6.3)) for $d = 1$.

Corollary 2.5. *Under Assumptions F′, $\lambda^{-1/4}\Xi'(A) \xrightarrow{\mathcal{D}} \mathcal{N}(0,(2\log_e 2-1)m)$ as $\lambda\to\infty$, where, as before, $m = \sqrt{\pi/2}\int_0^1|f'(x)|^{1/2}\,dx$. If f' is uniformly Lipschitz and either $|f'(0)|>0$ or $f'(0)=0$ and $\tau=1$, then the same limit holds for $\lambda^{-1/4}(\Xi(A)-m\lambda^{1/2})$.*

Proof. Use Theorem 2.4 and Lemmas 2.1 and 2.2.

We can actually extend Theorem 2.4 and Corollary 2.5 to cover sets $A := \{(x,y):\ 0\le x\le 1,\ 0\le y\le f(x)\}$ that are defined using the wider class of functions f which satisfy the following assumptions; these allow $|f'(x)|$ to approach infinity as $x\to 1$, in much the same way as $|f'(x)|$ is allowed to approach zero as $x\to 0$.

Assumptions F:

F1 The function f is continuous and decreasing on $[0,1]$, and a.e. continuously differentiable on $(0,1)$, with $f(0)=1$ and $f(1)=0$.

F2 For some $c_1, c_2 > 0$, some $0\le\tau<2$ and for all $0\le x\le 1$

$$\inf_{x/2\le t\le 1/2}|f'(t)|\ge c_1x^\tau;\qquad \sup_{(1-x)/2\le 1-t\le 1/2}|f'(t)|\le 1/\{c_1(1-x)^\tau\};$$
$$\sup_{0\le t\le 1/2}|f'(t)|\le c_2;\qquad \inf_{0\le 1-t\le 1/2}|f'(t)|\ge 1/c_2.$$

F3 For some $c_0>0$,

$$|f'(x)|\le c_0^{-1}\inf_{x/2\le t\le x}|f'(t)|,\qquad 0<x\le 1/2,$$

and

$$|f'(x)| \geq c_0 \sup_{(1-x)/2 \leq 1-t \leq 1-x} |f'(t)|, \qquad 1/2 \leq x < 1.$$

To see this, take any $0 < X < 1$, and split A into three parts $A_X^U \cup B_X \cup A_X^L$, where

$$A_X^U := A \cap [0,1] \times (f(X),1]; \quad A_X^L := A \cap (X,1] \times [0,1]; \quad B_X := [0,X] \times [0,f(X)]$$

[see Diagram 3]; then, correspondingly,

$$\Xi(A) = \Xi(A_X^U) + \Xi(B_X) + \Xi(A_X^L).$$

However, the random variables $\Xi(A_X^U)$ and $\Xi(A_X^L)$ are independent, since the former is a function of the configuration of Z restricted to A_X^U, and the latter of Z restricted to A_X^L, and, after appropriate scaling and, in the case of $\Xi(A_X^L)$, exchanging the rôle of x and y, both A_X^U and A_X^L have boundaries satisfying Assumptions F$'$. Hence, from Theorem 2.4 and the definition of d_{BW}, it follows easily that

$$d_{BW}(\mathcal{L}(W'_\lambda), \mathcal{N}(0,1)) = O(\varepsilon(\lambda)),$$

where

$$W'_\lambda := \{\operatorname{Var}\Xi(A_X^U) + \operatorname{Var}\Xi(A_X^L)\}^{-1/2}\Xi'(A_X^U \cup A_X^L).$$

Thus, to replace Assumptions F$'$ by Assumptions F in Theorem 2.4 and Corollary 2.5, it is enough to show that

$$d_{BW}(\mathcal{L}(\lambda^{-1/4}\Xi'(B_X)), \Delta_0) = O(\lambda^{-1/4}) \to 0 \tag{2.32}$$

as $\lambda \to \infty$, where Δ_0 denotes the degenerate distribution at 0, and that

$$\{\operatorname{Var}\Xi(A)\}^{-1}\{\operatorname{Var}\Xi(A_X^U) + \operatorname{Var}\Xi(A_X^L)\} \to 1. \tag{2.33}$$

To prove these two statements, we need the following lemma, which uses the fact that, from Assumption F2, given any $0 < \delta < 1/2$, there exists a constant $c_\delta > 0$ such that

$$c_\delta < |f'(x)| < 1/c_\delta \text{ for all } \delta \leq x \leq 1-\delta. \tag{2.34}$$

Lemma 2.6. *Suppose that f satisfies Assumptions F. Then, for any $x \in (0,1)$, $\mu(B_x) \leq \pi/c_\delta$, where c_δ is as in (2.34) and $\delta = \delta(x) = \min\{\frac{1}{2}x, 1 - f^{-1}(\frac{1}{2}f(x))\}$.*

Proof. The triangles T_1 with vertices $\{(0,f(x)), (0, f(x) + \frac{1}{2}c_\delta x), (x, f(x))\}$ and T_2 with vertices $\{(x,0), (x + \frac{1}{2}c_\delta f(x), 0), (x, f(x))\}$ are both contained in A, because of (2.34). Hence, for any $(w,z) \in B_x$,

$$\begin{aligned} |A_{(w,z)}| &\geq |T_1 \cap [w,1] \times [0,1]| + |T_2 \cap [0,1] \times [z,1]| \\ &= \tfrac{1}{4}c_\delta\{(x-w)^2 + (f(x)-z)^2\}. \end{aligned}$$

Thus

$$\mu(B_x) = \lambda \int_0^x dw \int_0^{f(x)} dz \exp\{-\lambda |A_{(w,z)}|\}$$
$$\leq \lambda \int_0^\infty dr \int_0^\infty ds \exp\{-\tfrac{1}{4}\lambda c_\delta (r^2+s^2)\} = \pi/c_\delta.$$

In view of Lemma 2.6, and because $\Xi(B_X) \geq 0$, (2.32) is immediate; (2.33) follows, after some calculation, from the bound

$$|A_{(x,y)}| \geq \tfrac{1}{4} c_{\delta(X)} \{(X-x)^2 + (f(X)-y)^2\}, \quad (x,y) \in B_X,$$

established in the proof of Lemma 2.6, combined with the bound (2.23) on $\mu(N_{\boldsymbol{\alpha}})$ and the formula

$$\mathbb{E}^{\boldsymbol{\alpha}} \Xi(N^U_{\boldsymbol{\alpha}}) = \lambda \int_0^x dw \int_y^{f(x)} dz \, \exp\left\{-\lambda \left((x-w)(f(x)-z) + \int_w^x du \, (f(u)-f(x)) \right) \right\}.$$

Having established a normal approximation to the number of maxima, it is natural also to look for a functional form, which gives extra information about where the maxima occur; c.f. also Theorem 5.1 of Chiu and Quine (1997). To this end, we define a random element Y_λ of $D[0,1]$ by

$$Y_\lambda(x) := \lambda^{-1/4} \Xi'(A \cap \{[1-x,1] \times [0,1]\}), \quad 0 \leq x \leq 1, \tag{2.35}$$

an appropriately normalized version of the number of maxima with first coordinate at least $1-x$. This we shall approximate by a process Y, where $Y(x) := W(t(x))$, $0 \leq x \leq 1$, with W standard Brownian motion and

$$t(x) := (2\log_e 2 - 1)\sqrt{\pi/2} \int_{1-x}^1 |f'(t)|^{1/2} \, dt. \tag{2.36}$$

Convergence of the finite dimensional distributions under Assumptions F follows easily, using scaling, from Lemma 2.6, Corollary 2.5 and the argument above; we concentrate on tightness. We begin with a lemma.

Lemma 2.7. *Let $B^{x,\delta} := A \cap \{[x, x+\delta] \times [f(x+\delta), 1]\}$. Then, under Assumptions F′,*

$$\lambda^{-1/2} \operatorname{Var} \Xi(B^{x,\delta}) = O(\delta(1 + |\log \delta|)),$$

uniformly in $0 \leq x \leq 1$ and $0 < \delta \leq 1-x$.

Proof. We use the estimates in the proof of Lemma 2.2, but with f replaced by $f^{x,\delta}$, where

$$f^{x,\delta}(t) := \{f(x+t\delta) - f(x+\delta)\}/\{f(x) - f(x+\delta)\}.$$

For this function, writing Df in place of f', we have

$$\begin{aligned} &|Df^{x,\delta}(t)| \le c_0^{-1} \inf_{t/2\le u\le t} |Df^{x,\delta}(u)| =: c_0^{-1}\tilde{k}_1^{x,\delta}; \\ &\tilde{k}_1^{x,\delta}(t) \ge k_1^{x,\delta}(t) := \inf_{u\ge t/2} |Df^{x,\delta}(u)| \ge \frac{\delta k_1(\delta t)}{f(x)-f(x+\delta)}, \end{aligned} \tag{2.37}$$

and

$$c_2^{x,\delta} := \sup_{0\le t\le 1} |Df^{x,\delta}(t)| \le \frac{c_2\delta}{f(x)-f(x+\delta)}. \tag{2.38}$$

Then $\Xi(B^{x,\delta})$, when generated by a Poisson process Z of intensity λ, has the same distribution as $\Xi(A^{x,\delta})$, with $A^{x,\delta}$ defined as was A except for $f^{x,\delta}$ in place of f, when generated by a Poisson process of intensity $\lambda^{x,\delta} := \lambda\delta(f(x)-f(x+\delta))$. Thus, using (2.10) and (2.6), we have

$$\begin{aligned} \mu^{x,\delta}(A^{x,\delta}) &\le c_0^{-1}\lambda^{1/2}\sqrt{\delta(f(x)-f(x+\delta))}\int_0^1 dt\int_0^\infty d\psi\, c(k_1^{x,\delta}(t),\psi) \\ &\le c_0^{-1}\lambda^{1/2}\sqrt{\delta(f(x)-f(x+\delta))}\int_0^1 dt\, 2\sqrt{\frac{8c_2}{3e}}\sqrt{\frac{\delta}{f(x)-f(x+\delta)}} \\ &\le \frac{2\lambda^{1/2}\delta}{c_0}\sqrt{\frac{8c_2}{3e}}. \end{aligned}$$

This bounds the first positive term in the expression for $\operatorname{Var}\Xi(B^{x,\delta})$ which can be derived from Lemma 2.2. Arguing similarly, using (2.18), (2.20) and (2.6), the remaining positive term is of order

$$O\left(\lambda^{1/2}\delta\int_0^1 \left\{1+\log\{ec_2^{x,\delta}/k_1^{x,\delta}(t)\}\right\} dt\right) = O\left(\lambda^{1/2}\delta(1+|\log\delta|)\right),$$

uniformly in $0\le x\le 1$ and in $0<\delta\le 1-x$, where the last equality follows from (2.37) and (2.38). The lemma is proved.

Theorem 2.8. *If Assumptions F hold, then $Y_\lambda \xrightarrow{\mathcal{D}} Y$ in $D[0,1]$, where Y_λ and Y are as for (2.35) and (2.36).*

Proof. We just need to prove tightness; as before, it is enough to work under Assumptions F$'$. Since Y is a.s. continuous, it is enough, by Aldous (1978), Corollary 1, to show that, for any sequence of stopping times τ_λ taking only finitely many values in $[0,1]$, and for any sequence of reals $\delta_\lambda \to 0$, then

$$\mathbb{P}[|Y_\lambda((\tau_\lambda+\delta_\lambda)\wedge 1) - Y_\lambda(\tau_\lambda)| > \varepsilon] \to 0$$

for all $\varepsilon > 0$.

To do so, we simply bound

$$\mathbb{P}\left[|Y_\lambda((\tau_\lambda + \delta_\lambda) \wedge 1) - Y_\lambda(\tau_\lambda)| > \varepsilon \,\middle|\, \mathcal{F}_t \cap \{\tau_\lambda = t\}\right]$$

for any $t \in [0,1]$, where $\mathcal{F}_t := \sigma(Z|_{A \cap \{[1-t,1]\times[0,1]\}})$. On $\{\tau_\lambda = t\}$, we have

$$Y_\lambda((\tau_\lambda + \delta_\lambda) \wedge 1) - Y_\lambda(\tau_\lambda) = \lambda^{-1/4}\left\{\Xi'(B(\lambda,t)) + \Xi'([1-t-\delta_\lambda^t, 1-t) \times [0, f(1-t)))\right\},$$

where $B(\lambda,t) := B^{(1-t-\delta_\lambda^t),\delta_\lambda^t}$ and $\delta_\lambda^t := \delta_\lambda \wedge (1-t)$. Now, by record value theory (Rényi 1962, Theorem 2′, p. 461) and Jensen's inequality,

$$\begin{aligned}
&\mathbb{E}\left\{|\Xi'([1-t-\delta_\lambda^t, 1-t) \times [0, f(1-t)))| \,\middle|\, \mathcal{F}_t \cap \{\tau_\lambda = t\}\right\} \\
&\quad \le \mathbb{E}\left\{|\Xi'([1-t-\delta_\lambda^t, 1-t) \times [0, f(1-t)))| \,\middle|\, Z(A \cap \{[1-t,1]\times[0,1]\}) = 0\right\} \\
&\quad \le 2(\log\lambda + 1),
\end{aligned}$$

so that

$$\begin{aligned}
&\mathbb{P}\left[\lambda^{-1/4}|\Xi'([1-t-\delta_\lambda^t, 1-t) \times [0, f(1-t)))| > \varepsilon/2 \,\middle|\, \mathcal{F}_t \cap \{\tau_\lambda = t\}\right] \\
&\qquad \le 4\lambda^{-1/4}(\log\lambda + 1)/\varepsilon \to 0.
\end{aligned}$$

Then, by Chebyshev's inequality and Lemma 2.7, and since $\Xi'(B(\lambda,t))$ is independent of $\mathcal{F}_t$, we have

$$\begin{aligned}
\mathbb{P}[\lambda^{-1/4}|\Xi'(B(\lambda,t))| > \varepsilon/2 \,|\, \mathcal{F}_t \cap \{\tau_\lambda = t\}] &\le 4\varepsilon^{-2}\lambda^{-1/2}\mathrm{Var}\left(\Xi'(B(\lambda,t))\right) \\
&= O(\delta_\lambda(1 + |\log\delta_\lambda|) \to 0,
\end{aligned}$$

and the theorem follows.

3. Extensions

3.1 Fixed numbers of points

Suppose now that exactly n points are placed uniformly at random in A. Let $\widetilde{\Xi}$ denote the resulting process of maximal points, and set $\tilde{\mu}(d\boldsymbol{\alpha}) := \mathbb{E}\,\widetilde{\Xi}(d\boldsymbol{\alpha})$. Let μ denote the mean measure of the maximal point process Ξ in the Poisson model with $\lambda = \lambda(n) = n/|A|$, and define $\widetilde{\Xi}' := \widetilde{\Xi} - \mu$, $\widetilde{W}_n := \{\mathrm{Var}\,\Xi(A)\}^{-1/2}\,\widetilde{\Xi}'(A)$. The following lemma shows that the two models are asymptotically equivalent; see also Baryshnikov (2000).

Lemma 3.1. *Under Assumptions F, with* $\lambda(n) := n/|A|$, $d_{BW}(\mathcal{L}(\widetilde{W}_n), \mathcal{L}(W_{\lambda(n)})) = O(n^{-1/2}\sqrt{\log n})$.

Proof. Let $A^{(a)} := \{(x,y) : 0 \le x \le 1-a,\ 0 \le y \le 1-a,\ y+a \le f(x+a)\}$ denote the subset of points (x,y) of A such that the square of side a with lower left corner at (x,y) is contained in A; note that $|A \setminus A^{(a)}| \le 2a$. Then, since

$$\tilde{\mu}(d\boldsymbol{\alpha}) = \frac{n}{|A|}\left(1 - \frac{|A_{\boldsymbol{\alpha}}|}{|A|}\right)^{n-1} d\boldsymbol{\alpha},$$

it follows that

$$\tilde{\mu}(A^{(a)}) \le n(1 - a^2/|A|)^{n-1} \le 2ne^{-na^2/|A|}$$

if $a^2 \le |A|/2$. In particular, if $a = a(n) = 2(|A|/n)^{1/2}\sqrt{\log n}$ and $n/\log n \ge 8$, then $\tilde{\mu}(A^{(a)}) \le 2n^{-3}$, and the same is true for $\mu(A^{(a)})$. Thus, for this choice of a,

$$d_{TV}(\mathcal{L}(\widetilde{\Xi}(A \setminus A^{(a)})), \mathcal{L}(\widetilde{\Xi}(A))) \le 2n^{-3}; \quad d_{TV}(\mathcal{L}(\Xi(A \setminus A^{(a)})), \mathcal{L}(\Xi(A))) \le 2n^{-3}. \tag{3.1}$$

Now, conditional on the number $\widetilde{N}$ of the n original points which fall in $A \setminus A^{(a)}$ taking the value r, $\widetilde{\Xi}(A \setminus A^{(a)})$ is distributed as the number of maximal points among r points placed uniformly at random in $A \setminus A^{(a)}$; and the same is true for $\Xi(A \setminus A^{(a)})$, conditional on $N := Z(A \setminus A^{(a)}) = r$. Hence

$$\begin{aligned} d_{TV}(\mathcal{L}(\widetilde{\Xi}(A \setminus A^{(a)})), \mathcal{L}(\Xi(A \setminus A^{(a)}))) &\le d_{TV}(\mathcal{L}(\widetilde{N}), \mathcal{L}(N)) \\ &= d_{TV}(\mathrm{Bi}\,(n, |A \setminus A^{(a)}|/|A|), \mathrm{Po}\,(n|A \setminus A^{(a)}|/|A|)) \le 2a/|A| = 4(n|A|)^{-1/2}\sqrt{\log n}, \end{aligned} \tag{3.2}$$

the latter inequality as in Barbour, Holst and Janson (1992), (1.23). The lemma now follows from (3.1) and (3.2). Note also that one can replace $\mu(A)$ by $\tilde{\mu}(A)$ and $\mathrm{Var}\,\Xi(A)$ by $\mathrm{Var}\,\widetilde{\Xi}(A)$ in this lemma, since they agree to the necessary orders.

3.2 Convex sets

Let C be a bounded convex set in the plane. By translation and scaling, it can typically be assumed that the upper right boundary of C is a function $f : [0,1] \to [0,1]$ which is continuous and decreasing, and has $0 < |f'(x)| < \infty$ for all $x \in (0,1)$, possibly augmented by a horizontal segment $\{(x,1) : -a \le x \le 0\}$ and a vertical segment $\{(1,y);\ -b \le y \le 0\}$. Letting A be as before, it follows that $\lambda^{-1/4}\Xi'(A)$ is approximately normally distributed if f satisfies Assumptions F, where Ξ' denotes the centred process of maximal points of C derived from a Poisson process Z with uniform intensity λ on C. It is then a simple matter to check that $\mu(C \setminus A) = O(\log \lambda)$, so that $\lambda^{-1/4}\Xi'(C)$ has the same normal approximation.

An exception to the above occurs if the upper right boundary of C degenerates to a single point, possibly augmented by a horizontal and a vertical segment. The classic example is that of the square $C := [0,1] \times [0,1]$. Here, the number of maxima among n randomly distributed points is just the number of record values in n independent and identically distributed uniform trials, and is hence close in distribution to $\mathrm{Po}\,(\log n)$, so that $(\log n)^{-1/2}\Xi'(C)$ is asymptotically normal. The technique of Section 2 is of no use in proving such a result, since, for any $\boldsymbol{\alpha} \in C$, the neighbourhood $N_{\boldsymbol{\alpha}}$ is the whole of C. This indicates that the restrictions imposed by Assumptions F2 and F3 may well be artefacts of the method of proof.

3.3 Higher dimensions

The method of Section 2 relies on two features of the problem. First, almost all maxima occur close to the boundary $(x, f(x))$; within a distance $O(\lambda^{-1/2}(\log\lambda)^{1/2})$, as in the proof of Lemma 3.1. Secondly, for most pairs of points 'close' to the boundary, the events of being maximal are independent, in the sense that, for $\boldsymbol{\alpha}$ 'near' to the boundary, $\mu(N_{\boldsymbol{\alpha}})$ and $\mathbb{E}^{\boldsymbol{\alpha}}\Xi(N^0_{\boldsymbol{\alpha}})$ are typically only of order $O(1)$. Most of the additional technical detail in Section 2 comes from allowing f to become almost horizontal near 0 or nearly vertical near 1.

In higher dimensions, similar considerations apply. As illustration, we take the Poisson model, with A the unit simplex $\{(x_1,\ldots,x_m) \subset \mathbb{R}^m_+ : \sum_{j=1}^m x_j \le 1\}$. Defining $r(\boldsymbol{\alpha}) := 1 - \sum_{j=1}^m x_j$ when $\boldsymbol{\alpha} = (x_1,\ldots,x_m)$, it follows that $\mu(d\boldsymbol{\alpha}) = \lambda\exp\{-\lambda r^m/m!\}d\boldsymbol{\alpha}$, so that, with probability at least $1-\lambda^{-2}/m!$, all the maxima have $r(\boldsymbol{\alpha}) \le \{3\lambda^{-1}m!\log\lambda\}^{1/m}$, and $\mu(A) \asymp \lambda^{1-1/m}$.

To assess the magnitude of the variance, write $A_{\boldsymbol{\alpha}} := \{\boldsymbol{\beta} \in A : \ \boldsymbol{\beta} \ge \boldsymbol{\alpha}\}$ and $N_{\boldsymbol{\alpha}} = \{\boldsymbol{\beta} \in A : \ A_{\boldsymbol{\beta}} \cap A_{\boldsymbol{\alpha}} \ne \emptyset\}$ as before. Then it also follows from above that

$$\mu(N_{\boldsymbol{\alpha}}) = O\left(\int_0^\infty ds\,\lambda(r+s)^{m-1}e^{-\lambda s^m/m!}\right) = O(1 + (\lambda^{1/m}r)^{m-1}),$$

where $r = r(\boldsymbol{\alpha})$, implying that $\int_{\boldsymbol{\alpha}\in A}\mu(d\boldsymbol{\alpha})\mu(N_{\boldsymbol{\alpha}}) = O(\lambda^{1-1/m})$. It thus remains to consider $\int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in N_{\boldsymbol{\alpha}}}\mathbb{E}\{\Xi(d\boldsymbol{\alpha})\Xi(d\boldsymbol{\beta})\}$. Now, for any $B \subset A$, it is immediate that $\Xi(B) \le Z(B)$, where Z denotes the underlying Poisson process. Hence, letting $R_{\boldsymbol{\alpha}} := \{\boldsymbol{\beta} \in A : r(\boldsymbol{\beta}) \le r(\boldsymbol{\alpha})\}$, it follows that

$$\mathbb{E}^{\boldsymbol{\alpha}}\{\Xi(N^0_{\boldsymbol{\alpha}} \cap R_{\boldsymbol{\alpha}})\} \le \mathbb{E}Z(N^0_{\boldsymbol{\alpha}} \cap R_{\boldsymbol{\alpha}}) = \lambda|N_{\boldsymbol{\alpha}} \cap R_{\boldsymbol{\alpha}}|,$$

and $|N_{\boldsymbol{\alpha}} \cap R_{\boldsymbol{\alpha}}| = O(r^m(\boldsymbol{\alpha}))$. Thus

$$\begin{aligned}
&\int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in N_{\boldsymbol{\alpha}}} \mathbb{E}\{\Xi(d\boldsymbol{\alpha})\Xi(d\boldsymbol{\beta})\} \\
&\quad \le \int_{\boldsymbol{\alpha}\in A} \lambda\mu(d\boldsymbol{\alpha})|N_{\boldsymbol{\alpha}} \cap R_{\boldsymbol{\alpha}}|\, d\boldsymbol{\alpha} + \int_{\boldsymbol{\beta}\in A} \lambda\mu(d\boldsymbol{\beta})|N_{\boldsymbol{\beta}} \cap R_{\boldsymbol{\beta}}|\, d\boldsymbol{\beta} + \int_{\boldsymbol{\alpha}\in A} \mu(d\boldsymbol{\alpha}) \\
&\quad = O\left(\int_0^\infty \lambda e^{-\lambda r^m/m!}\lambda r^m\, dr + \lambda^{1-1/m}\right) = O(\lambda^{1-1/m}),
\end{aligned}$$

and hence $\sigma^2 := \operatorname{Var}\Xi(A) = O(\lambda^{1-1/m})$.

For the Stein argument, one also requires third moment estimates analogous to those of Lemma 2.3. These can be derived in similar fashion, and are also all of order $O(\lambda^{1-1/m})$. For instance, in treating $\int_{\boldsymbol{\alpha}\in A} \mathbb{E}\{\Xi(d\boldsymbol{\alpha})\Xi^2(N^0_{\boldsymbol{\alpha}})\}$, note that

$$\begin{aligned}
&\int_{\boldsymbol{\alpha}\in A}\int_{\boldsymbol{\beta}\in N^0_{\boldsymbol{\alpha}}}\int_{\boldsymbol{\gamma}\in N^0_{\boldsymbol{\alpha}}\cap N_{\boldsymbol{\beta}}{}^0} \mathbb{E}\{\Xi(d\boldsymbol{\alpha})\Xi(d\boldsymbol{\beta})\Xi(d\boldsymbol{\gamma})\} \\
&\quad = O\left(\int_{\boldsymbol{\alpha}\in A} \mu(d\boldsymbol{\alpha})\mathbb{E}\{Z^2(N^0_{\boldsymbol{\alpha}} \cap R_{\boldsymbol{\alpha}})\}\right) = O(\lambda^{1-1/m}),
\end{aligned}$$

and that

$$\begin{aligned}
&\int_{\boldsymbol{\beta}\in A}\int_{\boldsymbol{\alpha}\in N_{\boldsymbol{\beta}}{}^0\cap R_{\boldsymbol{\beta}}}\int_{\boldsymbol{\gamma}\in N^0_{\boldsymbol{\alpha}}\cap R_{\boldsymbol{\beta}}} \mathbb{E}\{\Xi(d\boldsymbol{\alpha})\Xi(d\boldsymbol{\beta})\Xi(d\boldsymbol{\gamma})\} \\
&\quad \le \int_{\boldsymbol{\beta}\in A} \mu(d\boldsymbol{\beta})\int_{\boldsymbol{\alpha}\in N_{\boldsymbol{\beta}}{}^0\cap R_{\boldsymbol{\beta}}}\int_{\boldsymbol{\gamma}\in N^0_{\boldsymbol{\alpha}}\cap R_{\boldsymbol{\beta}}} \mathbb{E}\{Z(d\boldsymbol{\alpha})Z(d\boldsymbol{\gamma})\} \\
&\quad = O\left(\int_{\boldsymbol{\beta}\in A} \mu(d\boldsymbol{\beta})(\lambda r^m(\boldsymbol{\beta}))^2\right) = O(\lambda^{1-1/m}).
\end{aligned}$$

Thus the error terms arising in the Stein argument are all of order $O(\sigma^{-3}\lambda^{1-1/m})$, giving a normal approximation with d_{BW}–error of order $O(\lambda^{-\frac{1}{2}(1-1/m)})$, provided that σ^2 is of *exact* order $O(\lambda^{1-1/m})$. This was no problem in the previous section, where the asymptotic formula for the variance could be explicitly evaluated. However, for Johnson–Mehl processes with time varying Poisson intensity, this is a problem even in two dimensions — see Chiu and Lee (2000) for a solution in this particular case — and in higher dimensions even more so. We therefore treat this problem in rather more detail in Section 4: see Theorem 4.3 in particular.

Functional generalizations of the normal limit theorem could also be considered. If some set indexed generalization of $D[0,1]$ were the function space of interest, tightness might be expected to cause the most trouble. One way of simplifying the problem would be to consider the semi–functional convergence discussed in Ivanoff and Merzbach (2000, Section 7.3).

Similar considerations apply to more general simplex–like sets $A \subset \mathbb{R}_+^m$ whose boundary, apart from the coordinate hyperplanes, is continuously differentiable and has outward normal with all components positive and uniformly bounded away from 0. Take, for instance, $A := \{x \in \mathbb{R}_+^m : f(x) \leq 1\}$, where $f : \mathbb{R}_+^m \to \mathbb{R}_+$ is an increasing function of each coordinate variable, is continuously differentiable, and satisfies

$$\theta_i^- \leq \partial f / \partial x_i \leq \theta_i^+ \quad \text{for all} \quad x \in \mathbb{R}_+^m, \tag{3.3}$$

for $\theta^-, \theta^+ \in \mathbb{R}_+^m$ such that $0 < \theta^- < \theta^+ < \infty$; here and subsequently, for elements $x, y \in \mathbb{R}^m$, we use $x < y$ to mean that $x_i < y_i$ for *all* $1 \leq i \leq m$.

Note that the total variation error incurred by replacing the Poisson model by the model with a fixed number n of points is larger in higher dimensions, being of order $\{n^{-1} \log n\}^{1/m}$. Note also that the case $C = [0,1]^m$, which is far from being covered by our arguments, is successfully treated using a coordinate transformation in Baryshnikov (2000).

4. A lower bound for the variance

To justify the exact asymptotic order of the variance in higher dimensions, we use an approach which can also be modified to apply to variants of our model, such as the Johnson–Mehl model with inhomogeneous Poisson germs. The idea is to replace the model which has A fixed and an increasing Poisson intensity λ by the equivalent version where the Poisson intensity remains fixed, but A expands at the corresponding rate, having volume of order $O(\lambda)$, and hence surface area of order $O(\lambda^{1-1/m})$. The maximal points then typically occur at a distance of order $O(1)$ from the upper surface of A, and the idea is to show that there is enough independence in the point process to make the variance additive over the upper surface. We do this essentially by conditioning on the realization of the Poisson process Z on a surface layer of finite width, and identifying conditionally independent components of the remaining Ξ–process of maximal points, lying in enclosed 'tents' next to the layer. The details are as follows.

First, we modify the definition of A at the end of Section 3.3 by assuming instead that

$$A_\lambda := \lambda^{1/m} A_1 = \{x \in \mathbb{R}_+^m : f(x\lambda^{-1/m}) \leq 1\}, \tag{4.1}$$

where $f : \mathbb{R}_+^m \to \mathbb{R}_+$ is as at (3.3), and we define

$$C(\theta^-, \theta^+) := \max_{1 \leq i \leq m} \theta_i^+ / \theta_i^-. \tag{4.2}$$

We shall also use the following notation, for $\theta > 0$ in $\mathbb{R}^m_+$:

$$H_K(\theta) := \left\{ x \in \mathbb{R}^m : \sum_{j=1}^m x_j/\theta_j \le K \right\}; \quad \Delta_K(\theta) := H_K \cap \mathbb{R}^m_+, \tag{4.3}$$

and we further define $I_1(y) := \{x \in \mathbb{R}^m : x \le y\}$. The argument now proceeds by way of two lemmas. The first shows in essence that, when the process Z has been realized on a layer $A_\lambda \cap B^c$ near the boundary, there are typically a number of isolated 'tents' left inside B, within each of which the process Ξ is determined solely by any remaining points of Z which may occur in them.

Lemma 4.1. *Take any $K \ge 2m$, $\psi > 0$ and $0 < \theta \in \mathbb{R}^m$. Let $L \subset \mathbb{R}^m$ be any closed subset containing $\{x : 0 \le x \le 2\theta\}$, $B \subset \mathbb{R}^m$ be any closed subset such that $B \cap \Delta_K(\theta) = \Delta_1(\theta)$. Suppose that Z is a Poisson process on $\mathcal{X}_K := H_K(\theta) \cap L$, and that Ξ is the associated process of maximal points; let $\mathcal{F}(F)$ denote the σ–algebra $\sigma(Z|_F)$ for any measurable $F \in \mathcal{X}_K$. Then there is an event $E \in \mathcal{F}(B^c)$ such that, on $\mathcal{F}(B^c) \cap E$,*

$$\Xi|_{I_1(\theta)\cap B} = \Xi|_{\Delta_1(\theta)} \quad \text{and is independent of} \quad \Xi|_{I_1^c(\theta)\cap B}, \tag{4.4}$$

and

$$\mathrm{Var}\,\{\Xi(\Delta_1(\theta)) \,|\, \mathcal{F}(B^c) \cap E\} \ge q(\theta, \psi) > 0, \tag{4.5}$$

where q is continuous in θ and ψ; furthermore,

$$\mathbb{P}[E] \ge p(K, \psi, \theta) > 0, \tag{4.6}$$

where p is continuous in K, ψ and θ.

Proof. Consider the subset $D(u^{(1)}, \ldots, u^{(m)}) := \bigcup_{j=1}^m I_1(u^{(j)})$, where

$$u^{(j)} := 2(\theta - \theta_j \varepsilon^{(j)}), \quad 1 \le j \le m,$$

and $\varepsilon^{(j)}$ denotes the j-th coordinate vector. It is disjoint from the set $\{x > 0\} \subset \mathbb{R}^m$, and it contains every point $x \in I_1(\theta)$ having $x_j \le 0$ for at least one j. In particular, it does not contain the interior $\Delta_1^o(\theta)$ of $\Delta_1(\theta)$, but it contains all of the boundary of $\Delta_1(\theta)$ apart from the upper face $\left\{\sum_{j=1}^m x_j/\theta_j = 1\right\} \cap \Delta_1(\theta)$. By continuity, there exists an $\varepsilon > 0$ such that, for any $v^{(j)}$, $1 \le j \le m$, satisfying $|v^{(j)} - u^{(j)}| < \varepsilon$, we have

$$D(v^{(1)}, \ldots, v^{(m)}) \cap I_1(\theta) \cap H_1^o(\theta) = (I_1(\theta) \cap H_1^o(\theta)) \setminus \Delta^o(v^{(1)}, \ldots, v^{(m)}), \tag{4.7}$$

where $\Delta(v^{(1)},\dots,v^{(m)})$ is a simplex of volume at least $\frac{1}{2}|\Delta_1(\theta)|$, with one face on the hyperplane $\left\{\sum_{j=1}^m x_j/\theta_j = 1\right\}$ and the others contained in $D(v^{(1)},\dots,v^{(m)})$.

Now let

$$\Delta_K^*(\theta) := \left\{ x \geq 0 : \, 1 \leq \sum_{j=1}^m x_j/\theta_j \leq K \right\},$$

and let E be the event that $Z|_{\Delta_K^*(\theta)}$ consists of exactly m points $v^{(1)},\dots,v^{(m)}$ satisfying $|v^{(j)} - u^{(j)}| < \varepsilon$, $1 \leq j \leq m$, for ε as defined for (4.7). Note that the event E indeed has positive probability which is bounded below, uniformly in L, by a continuous function $p(K,\psi,\theta)$ of K, θ and ψ. Then, on E, it is immediate that (4.4) holds. For (4.5), observe that

$$\begin{aligned} \Xi(\Delta_1(\theta)) = 0 \quad &\text{if} \quad Z(\Delta(v^{(1)},\dots,v^{(m)})) = 0; \\ \Xi(\Delta_1(\theta)) \geq 1 \quad &\text{if} \quad Z(\Delta(v^{(1)},\dots,v^{(m)})) \geq 1, \end{aligned} \tag{4.5}$$

and hence

$$\begin{aligned} &\mathrm{Var}\left\{\Xi(\Delta_1(\theta)) \,|\, \mathcal{F}(B^c) \cap E\right\} \\ &\qquad \geq \min\{\mathbb{P}[Z(\Delta(v^{(1)},\dots,v^{(m)})) = 0], \mathbb{P}[Z(\Delta(v^{(1)},\dots,v^{(m)})) \geq 1]\} \\ &\qquad \geq q(\theta,\psi), \end{aligned} \tag{4.8}$$

where

$$q(\theta,\psi) := \min\left\{1 - \exp\{-\tfrac{1}{2}\psi|\Delta_1(\theta)|\}, \exp\{-\psi|\Delta_1(\theta)|\}\right\} > 0, \tag{4.9}$$

with $|\Delta_1(\theta)| = (1/m!)\prod_{j=1}^m \theta_j$, is a continuous function of ψ and θ. This proves the lemma.

The second lemma is used to give an upper bound for the set of points in A_λ which dominate a given point near its upper boundary.

Lemma 4.2. *Suppose that $g : \mathbb{R}_+^m \to \mathbb{R}$ is an increasing function of each coordinate variable, and that $0 < \theta_j^- \leq 1/(\partial g/\partial x_j) \leq \theta_j^+ < \infty$ for each $1 \leq j \leq m$. Let θ be such that $\theta^- \leq \theta \leq \theta^+$. Then if $K = 2mC^2$, where C is as defined in (4.2), it follows that*

$$\{x \in \mathbb{R}_+^m : \, g(x) \leq g(2\theta)\} \subset \Delta_K(\theta).$$

Proof. For any $x \in \mathbb{R}_+^m$, define $J := \{s : \, x_s/\theta_s \geq 2\}$. Then

$$\begin{aligned} g(x) - g(2\theta) &\geq \sum_{s\in J}\left(\frac{x_s}{\theta_s} - 2\right)\frac{\theta_s}{\theta_s^+} - \sum_{s\notin J}\left(2 - \frac{x_s}{\theta_s}\right)\frac{\theta_s}{\theta_s^-} \\ &\geq \sum_{s=1}^m \frac{1}{C}\left(\frac{x_s}{\theta_s} - 2\right) - 2|J|(C - 1/C) \\ &\geq C^{-1}\sum_{s=1}^m (x_s/\theta_s) - 2mC. \end{aligned}$$

Hence, if $\sum_{s=1}^m (x_s/\theta_s) > 2mC^2$, it follows that $g(x) > g(2\theta)$.

We are now able to prove the main theorem.

Theorem 4.3. *Let $Z = Z_\lambda$ be a Poisson process of unit rate defined on A_λ as in (4.1), and let $\Xi = \Xi_\lambda$ be the associated process of maximal points. Then* $\mathrm{Var}\,(\Xi(A_\lambda)) \geq c_4 \lambda^{1-1/m}$ *for some $c_4 > 0$.*

Proof. Fix any $\theta \in \mathbb{R}_+^m$ such that $\theta^- \leq \theta \leq \theta^+$, with θ_i^- and θ_i^+ as in (3.3). Pick as many points $x^{(i)}$ in the set $\{f(x\lambda^{-1/m}) = 1\}$ as possible such that, with $K = 2mC^2$ for C as in (4.2), the sets

$$\Delta_K^{(i)} := x^{(i)} - 2\theta + \Delta_K(\theta) \tag{4.10}$$

are all disjoint; let there be $M(\lambda)$ of them. For instance, if $k^* := \max\{k > 0 : \Delta_k(\theta) \subset A_1\}$, then $\Delta_{\lambda^{1/m}k^*} \subset A_\lambda$ has an upper face which can be covered by $\{1/(m-1)!\}s(s+1)\ldots(s+m-2)$ disjoint translates of the corresponding face of $\Delta_K(\theta)$, where

$$s = \lfloor k^* \lambda^{1/m}/K \rfloor - 1 \asymp \lambda^{1/m}.$$

Attach a copy of $\Delta_K(\theta)$ in each of these positions, and then translate them outwards until the points at position 2θ relative to their apexes lie on $\{x \in \mathbb{R}_+^m : f(x\lambda^{-1/m}) = 1\}$, still keeping them disjoint. Thus we can always take $M(\lambda) \asymp \lambda^{1-1/m}$. Note also that, by Lemma 4.2,

$$\{x : x^{(i)} - 2\theta \leq x \leq x^{(i)}\} \subset \{x \in A_\lambda : x \geq x^{(i)} - 2\theta\} \subset \Delta_K(\theta), \quad 1 \leq i \leq M(\lambda). \tag{4.11}$$

Now take any closed subset $B \subset A_\lambda$ such that

$$B \cap \Delta_K^{(i)} = x^{(i)} - 2\theta + \Delta_1(\theta) =: \Delta_1^{(i)}, \quad 1 \leq i \leq M(\lambda), \tag{4.12}$$

and observe that

$$\mathrm{Var}\,(\Xi(A_\lambda)) \geq \mathbb{E}\{\mathrm{Var}\,(\Xi(A_\lambda)) \,|\, \mathcal{F}(B^c)\}. \tag{4.13}$$

Since the events $E^{(i)}$, defined as E was earlier, but now for each of the $\Delta_K^{(i)}$, are all measurable with respect to $\mathcal{F}(B^c)$, and since each has probability at least $p(K, 1, \theta)$, it follows from (4.13) and the conditional independence of $\Xi(\Delta_1^{(i)})$ and $\Xi|_{I_1^c(x^{(i)}-\theta)\cap B}$ on $\mathcal{F}(B^c) \cap E^{(i)}$, guaranteed in (4.4), that

$$\mathrm{Var}\,(\Xi(A_\lambda)) \geq \sum_{i=1}^{M(\lambda)} \mathbb{P}[E^{(i)}] q(\theta, 1) \geq M(\lambda) p(K, 1, \theta) q(\theta, 1) \asymp \lambda^{1-1/m}.$$

This proves the theorem.

The above method of proof is very flexible. It can easily be adapted to Johnson–Mehl processes in $m-1$ dimensions, where the intensity of the Poisson process could be allowed to vary continuously in space as well as time; the excluded neighbourhoods in these models are cones rather than simplexes, but this makes no essential difference. It can also be used if only a part of the boundary of A_λ is as in (4.1), because the conditional independence proved in Lemma 4.1 does not depend on the form of the set L in any detail; hence functions f could be allowed in Theorem 4.3 which are somewhat more general than those given in (3.3), in that the partial derivatives could for instance approach zero or infinity in certain ways near the coordinate hyperplanes.

Acknowledgement. ADB thanks the Department of Mathematics and Statistics at Monash University for their warm hospitality, during the preparation of this paper.

References.

Aldous, D. (1978) Stopping times and tightness. *Ann. Probab.* **6**, 335–340.

Bai, Z.–D., Chao, C.–C., Hwang, H.–K. and Liang, W.–Q. (1998) On the variance of the number of maxima in random vectors and its applications. *Ann. Appl. Probab.* **8**, 886–895.

Bai, Z.–D., Hwang, H.–K., Liang, W.–Q. and Tsai, T.–H. (2000) Limit theorems for the number of maxima in random samples from planar regions. Tech. Rept C–2000–05, Inst. Stat. Science, Academia Sinica, Taiwan.

Barbour, A. D., Holst, L. and Janson, S. (1992) *Poisson Approximation.* Oxford University Press.

Barbour, A. D., Karoński, M. and Ruciński, A. (1989) A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47**, 125–145.

Baryshnikov, Y. (2000) *** *Prob. Theory Rel. Fields* **, ***–***.

Chen, L. H. Y. (1986) The rate of convergence in a central limit theorem for dependent random variables with arbitrary index set. Technical Report, IMA Minnesota.

Chiu, S. N. and Lee, H. Y. (2000) A regularity condition and strong limit theorems for linear birth–growth systems. *Math. Nachr.* (to appear)

Chiu, S. N. and Quine, M. P. (1997) Central limit theory for the number of seeds in a growth model in R^d with inhomogeneous Poisson arrivals. *Ann. Appl. Probab.* **7**, 802–814.

Devroye, L. (1993) Records, the maximal layer, and uniform distributions in monotone sets. *Comp. Math. Applics* **25**(5), 19–31.

Holst, L., Quine, M. P. and Robinson, J. (1996) A general stochastic model for nucleation and linear growth. *Ann. Appl. Probab.* **6**, 903–921.

Ivanoff, G. and Merzbach, E. (2000) *Set–indexed martingales.* Chapman and Hall, Boca Raton.

Quine, M. P. and Robinson, J. (1990) A linear random growth model. *J. Appl. Prob.* **27**, 499–509.

Rényi, A. (1962) On the extreme elements of observations. *MTA III. Oszt. Közl.* **12**, 105–121; *Selected papers of Alfréd Rényi,* Vol. 3, Ed. P. Turán, Akadémiai Kiadó, Budapest (1976), 50–66.