

## ZEROS OF RANDOM TROPICAL POLYNOMIALS, RANDOM POLYGONS AND STICK-BREAKING

FRANCOIS BACCELLI AND NGOC MAI TRAN

ABSTRACT. For  $i = 0, 1, \dots, n$ , let  $C_i$  be independent and identically distributed random variables with distribution  $F$  with support  $(0, \infty)$ . The number of zeros of the random tropical polynomials  $\mathcal{T}f_n(x) = \min_{i=1, \dots, n} (C_i + ix)$  is also the number of faces of the lower convex hull of the  $n + 1$  random points  $(i, C_i)$  in  $\mathbb{R}^2$ . We show that this number,  $Z_n$ , satisfies a central limit theorem when  $F$  has polynomial decay near 0. Specifically, if  $F$  near 0 behaves like a  $\text{gamma}(a, 1)$  distribution for some  $a > 0$ , then  $Z_n$  has the same asymptotics as the number of renewals on the interval  $[0, \log(n)/a]$  of a renewal process with inter-arrival distribution  $-\log(\text{Beta}(a, 2))$ . Our proof draws on connections between random partitions, renewal theory and random polytopes. In particular, we obtain generalizations and simple proofs of the central limit theorem for the number of vertices of the convex hull of  $n$  uniform random points in a square. Our work leads to many open problems in stochastic tropical geometry, the study of functionals and intersections of random tropical varieties.

### 1. INTRODUCTION

Consider the tropical min-plus algebra  $(\mathbb{R}, \odot, \oplus)$ ,  $a \odot b = a + b$ ,  $a \oplus b = \min(a, b)$ . A tropical polynomial  $\mathcal{T}f : \mathbb{R} \rightarrow \mathbb{R}$  of degree  $n$  has the general form

$$(1) \quad \mathcal{T}f(x) = \bigoplus_{i=0}^n (C_i \odot x^i) = \min_{i=0, \dots, n} (C_i + ix),$$

for coefficients  $C_i \in \mathbb{R}$ . The zeros of  $\mathcal{T}f$  are points in  $\mathbb{R}$  where the minimum in (1) is achieved at least twice. When the coefficients  $C_i$  are random, the zeros of  $\mathcal{T}f$  form a collection of random points in  $\mathbb{R}$ . A natural model of randomness is one where the  $C_i$ 's are independent and identically distributed (i.i.d.) according to some distribution  $F$ . In this paper, we derive the asymptotic distribution for the number of zeros as  $n \rightarrow \infty$ , under various distributions  $F$ .

**Theorem 1.** *Let  $F$  be a continuous distribution, supported on  $(0, \infty)$ . Assume  $F(y) \sim Cy^a + o(y^a)$  as  $y \rightarrow 0$  for some constants  $C, a > 0$ . Let  $Z_n$  be the number of zeros of  $\mathcal{T}f$  in (1). Then as  $n \rightarrow \infty$ ,*

$$(2) \quad \frac{Z_n - \frac{2a+2}{2a+1} \log(n)}{\sqrt{\frac{2a(a+1)(2a^2+2a+1)}{(2a+1)^3} \log(n)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

---

Received by the editors April 1, 2014 and, in revised form, August 11, 2014 and September 20, 2014.

2010 *Mathematics Subject Classification.* Primary 11S05, 60C05, 60D05, 14T05.

This work was supported by an award from the Simons Foundation (#197982 to The University of Texas at Austin). The authors would like to thank an anonymous referee for the careful reading and helpful suggestions.

The classical analogue of  $Z_n$  is the number of real zeros of a classical polynomial  $f(x) = \sum_{i=0}^n C_i x^i$  with random coefficients  $C_i \in \mathbb{R}$  or  $\mathbb{C}$ . This problem has an extensive literature, ranging from works in the mid-twentieth century [22, 23, 26] to a very recent paper by Tao and Vu [35], who proved a local universality phenomenon. Roughly speaking, this asserts that the asymptotic behavior of the zeros of  $f$  as  $n \rightarrow \infty$ , appropriately normalized, should become independent of the choice of the common distribution of the  $C_i$ 's. Our main result is a version of this statement for tropical polynomials.

Our setup provides a natural solution to counting zeros of polynomials over fields with valuations such as the  $p$ -adic numbers or Puiseux series, where methods developed for  $\mathbb{R}$  and  $\mathbb{C}$  do not easily apply. Such a polynomial  $f$  comes with a tropicalized version  $\mathcal{T}f$ . By the Fundamental Theorem of Tropical Algebraic Geometry, the zeros of  $\mathcal{T}f$  correspond to valuations of the zeros of  $f$  (see, for example, [27]). In particular,  $f$  and  $\mathcal{T}f$  have the same number of zeros. Studying combinatorial properties of classical varieties via tropicalization is the heart of tropical algebraic geometry. While zeros of random polynomials in fields with valuations have been studied [15], to the best of our knowledge this is the first result in the tropical setting.

**1.1. Connections to random polytopes.** Consider the polytope that is the convex hull of a finite set of points  $(i, C_i) \in \mathbb{R}^2$ . The *lower convex hull*  $\mathcal{C}(i, C_i) \subset \mathbb{R}^2$  of these points are faces of the polytope with support vectors of the form  $(1, \alpha)$  for some  $\alpha \in \mathbb{R}$ . Let  $|\mathcal{C}(i, C_i)|$  be the number of such faces. As we shall review in Lemma 3,

$$Z_n = |\mathcal{C}(i, C_i)|.$$

This connects random tropical polynomials with random polytopes. More explicitly, suppose  $F$  is *Uniform*(0, 1). If we replace  $i = 0, 1, \dots, n$  by  $n+1$  i.i.d. uniform points  $U_i$  on  $(0, 1)$ , then  $(U_i, C_i)$  are  $n+1$  uniform points on  $(0, 1)^2$ , and their convex hull is a random polytope. Statistics of such random polytopes have been studied extensively; see [32] for a recent review. For the convex hull of  $n$  uniform points in a square, Groeneboom [19] derived the central limit theorem for its number of vertices. It follows from his proof that the number of lower faces  $|\mathcal{C}(U_i, C_i)|$  satisfies

$$(3) \quad \frac{\mathbb{E}(|\mathcal{C}(U_i, C_i)|) - \frac{4}{3} \log(n)}{\sqrt{\frac{20}{27} \log(n)}} \xrightarrow{d} \mathcal{N}(0, 1),$$

and this is precisely Theorem 1 for the case  $a = 1$ . In fact, our proof of Theorem 1 is based on an extension of Groeneboom's results. Recently, Groeneboom [20] re-derived this result using a simpler argument with very similar ideas. The key idea of our proof, Lemma 11, appeared as Corollary 2 in [20]. However, he did not make the connection to stick-breaking or renewal theory explicit, nor generalize to the case where the random points are approximated by a non-homogeneous PPP.

**1.2. Connections to stick-breaking.** For  $C_i$  the  $i$ -th value of a random walk with exchangeable increments, the lower convex hull  $\mathcal{C}(i, C_i)$  is also known as the *greatest convex minorant* of the walk  $(C_i, i = 0, 1, \dots, n)$ . Various authors have studied greatest convex minorants (or concave majorants) of random walks [1], Brownian motion [2, 18], Lévy processes and other settings [9, 30]. A classical result by Andersen [5] states that if almost surely no two subsets of the increments have the same arithmetic mean, then  $|\mathcal{C}(i, C_i)|$  is distributed as the number of cycles in

a uniformly distributed random permutation of the set  $\{1, \dots, n\}$ . Its asymptotics in this case is

$$\frac{|\mathcal{C}(i, C_i)| - \log(n)}{\sqrt{\log(n)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Paraphrased, this is a very strong *universality result* on the zeros of tropical polynomials generated randomly under this setting. In fact, an even stronger result holds: the partition of  $n$  generated by the lengths of the faces in  $\mathcal{C}(i, C_i)$  has the same distribution as the partition of  $n$  generated by the cycles of a uniform random permutation. As  $n \rightarrow \infty$ , the lengths of these cycles converge in distribution to the lengths of sticks obtained from a *uniform stick-breaking process* (cf. Section 3).

In Section 3, we show that when  $F$  is  $Exp(1)$ , conditioned on  $C_n = 0$ , the partition of  $n$  generated by the lengths of the faces in  $\mathcal{C}(i, C_i)$  is a partially exchangeable partition. The limiting distribution of cycle lengths converge to the partition lengths of a  $Beta(2, 1)$  stick-breaking process. To our knowledge, the  $Beta(2, 1)$  scheme has not been considered in the literature. The classical two-parameter family contains the  $Beta(1, \theta)$  stick-breaking scheme [29, §3], [10]. A fundamental difference is exchangeability: in the  $Beta(1, \theta)$  case, one obtains an exchangeable partition of  $n$ , whereas our partition is only partially exchangeable. However, like in the classical case, the  $Beta(2, 1)$ , and in general, the  $Beta(2, a)$  stick-breaking scheme for  $a > 0$  enjoy the Polya's urn connection. In particular, it is a version of the Bernoulli sieve of Gneden et al. [16, 17]. For  $F = Exp(1)$ ,  $a = 1$ , and the number of cycles in a  $Beta(2, 1)$  partition is precisely the number of zeros of  $\mathcal{T}f$  conditioned on  $C_n = 0$ . In this case,  $G$  of Proposition 2 is just the Lebesgue measure, and we are back to the setting of Groeneboom [19]. By conditioning on the minimum index of the  $C_i$ 's, one obtains an elementary stick-breaking proof of the results of Groeneboom [19] (cf. Section 3).

**1.3. Proof overview.** To prove (3), Groeneboom showed that the points of  $(U_i, C_i)$  can be replaced by the points in  $(0, 1)^2$  of a homogeneous Poisson point process (PPP) with rate  $n$ . He then studied their lower convex hull by a pure jump Markov process on the vertices, indexed by the slopes of their support vectors.

Our first step is a generalization of this approach to a class of non-homogeneous PPP on  $\mathbb{R}_+^2$ . More precisely, let  $F$  be the distribution function in the hypothesis of Theorem 1. Consider the PPP with intensity measure  $n\lambda \otimes G$ , where  $\lambda$  is the Lebesgue measure and  $G$  is the measure on  $\mathbb{R}_+$  such that, for  $x \geq 0$ ,

$$(4) \quad G([0, x]) = -\ln(1 - F(x)).$$

The generalization in question is Proposition 2 below. We use the following definition:

For a sequence of random variables  $(X_n, n \geq 1)$  and a deterministic sequence  $(b_n, n \geq 1)$ , we say  $X_n = \mathcal{O}_P(b_n)$  if for all sequences  $(c_n, n \geq 1)$  satisfying  $c_n \rightarrow \infty$  as  $n \rightarrow \infty$ ,

$$\mathbb{P}(|X_n| \geq b_n c_n) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

**Proposition 2.** *Let  $\lambda$  be the Lebesgue measure,  $F$  be the distribution in Theorem 1, and  $\Phi_n$  be the restriction to  $\mathbb{R}_+^2$  of the Poisson point process on  $\mathbb{R} \times \mathbb{R}_+$  with intensity measure  $n\lambda \otimes G$ . For points of  $\Phi_n$  in  $(0, 1) \times (0, F^{-1}(1 - e^{-1}))$ , let  $|\mathcal{C}(\Phi_n)|$  be the number of lower faces in their convex hull. Then*

$$|\mathcal{C}(\Phi_n)| - (J_n + J'_n) = \mathcal{O}_P(1),$$

where  $J_n$  and  $J'_n$  are independent random variables, each distributed as the number of renewals on  $[0, \log(n)/a]$  of a delayed renewal process with  $-\log(\text{Beta}(a, 2))$  as the inter-arrival distribution. Consequently, as  $n \rightarrow \infty$ ,

$$\frac{|\mathcal{C}(\Phi_n)| - \frac{2a+2}{2a+1} \log(n)}{\sqrt{\frac{2a(a+1)(2a^2+2a+1)}{(2a+1)^3} \log(n)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Theorem 1 is a discrete version of this setup. Indeed, let  $\lambda_n$  be a discrete measure on  $\mathbb{R}$  which puts mass 1 at every point  $i/n$  for  $i = 0, 1, \dots$ , and 0 elsewhere. On each half line  $\{i/n\} \times (0, \infty)$ , run an independent PPP with intensity measure  $G$ . By design, the first jump of this process is distributed as  $F$ . But only the first point can possibly contribute to the lower convex hull. Thus the number of tropical zeros  $Z_n$  equals  $|\mathcal{C}(\tilde{\Phi}_n)|$ , where  $\tilde{\Phi}_n$  is a PPP with intensity measure  $\lambda_n \times G$ . A direct coupling between  $\Phi_n$  and  $\tilde{\Phi}_n$  (cf. Proposition 21) proves that

$$|\mathcal{C}(\Phi_n)| - |\mathcal{C}(\tilde{\Phi}_n)| = \mathcal{O}_P(1),$$

and Theorem 1 follows.

**1.4. Scope and organization.** We hope to kindle the interest of researchers from both stochastic and tropical geometry. However, this paper has a rather narrow scope of counting the number of zeros for a natural model of random tropical polynomials. This is perhaps the simplest linear functional of the simplest tropical variety. A review of both fields with rigorous definition of stochastic tropical geometry is better suited for subsequent work.

Our paper is organized as follows. In Section 2 we prove the connection between zeros of  $\mathcal{Tf}$  and the lower convex hull of the points  $(i, C_i)$ . While this is a special case of the well-known connection between regular subdivision of Newton polytopes and tropical varieties, we provide an explicit proof via the Legendre transform for self-containment. Sections 3 to 5 treat the case  $F = \text{Exp}(1)$  via two different proofs: discrete stick-breaking and Poisson coupling. These proofs and their interactions provide intuition for the proof of the main case. Section 6 proves Proposition 2, and Section 7 proves Theorem 1 via a coupling argument. We summarize the paper and discuss open problems in Section 8.

*Notation.* Refer to Figure 1 for an illustration. For a set of points  $(x_i, y_i)$ , let  $\mathcal{C}(x_i, y_i)$  denote their lower convex hull. Let  $|\mathcal{C}(x_i, y_i)|$  denote the number of faces. For an underlying  $\mathcal{C}(x_i, y_i)$ , we list its vertices  $(V_i, i \geq 0)$  in increasing  $x$ -coordinate. For a vertex  $V$  of  $\mathcal{C}(x_i, y_i)$ , let  $x(V)$  denote its  $x$ -coordinate,  $y(V)$  its  $y$ -coordinate, and  $i(V) \in \{0, \dots, n\}$  its index. If  $0 \leq x_i \leq 1$ , let  $\Pi(\mathcal{C}(x_i, y_i))$  denote the partition of  $[0, 1]$  ordered by appearance, obtained by the points  $x(V_i)$  and  $\{0, 1\}$ . Let  $\Pi_n(\mathcal{C}(x_i, y_i))$  denote the partition of  $n$  ordered by appearance, defined by the sequence  $(i(V_j) - i(V_{j-1}), j = 1, \dots, |\mathcal{C}(x_i, y_i)|)$ .

We often split  $\mathcal{C}(x_i, y_i)$  into the ‘lower left’  $\mathcal{C}^+(x_i, y_i)$  and ‘lower right’  $\mathcal{C}^-(x_i, y_i)$  convex hulls; the first consists of faces with support vectors  $(1, \alpha)$  for  $\alpha > 0$ , and the second consists of those with  $\alpha < 0$ . Let  $|\mathcal{C}^+(x_i, y_i)|$  and  $|\mathcal{C}^-(x_i, y_i)|$  be the corresponding number of faces. For an underlying  $\mathcal{C}^+(x_i, y_i)$  or  $\mathcal{C}^-(x_i, y_i)$ , let  $(V_i^\downarrow, i \geq 0)$  be the vertices listed in decreasing  $y$ -coordinate, and let  $(V_i^\uparrow, i \geq 0)$  be the same set of vertices listed in increasing  $y$ -coordinate. For points in a fixed rectangle of some point process  $\Phi$ , we denote their lower convex hull by  $\mathcal{C}(\Phi)$ . Analogous quantities such as  $\mathcal{C}^+(\Phi)$ ,  $\Pi(\mathcal{C}(\Phi))$  follow the same naming convention.

For  $\alpha \in (0, \infty)$ , let  $L(\alpha)$  be the line orthogonal to the vector  $(1, \alpha)$  and which supports  $\mathcal{C}^+(\Phi)$ . Let  $L_x(\alpha)$  and  $L_y(\alpha)$  be its  $x$  and  $y$ -intercepts, respectively. Let  $(x(\alpha), y(\alpha))$  be the vertex of  $\mathcal{C}^+(\Phi)$  supported by  $L(\alpha)$ . If there are two or more such vertices, take the one with minimum  $y$ -coordinate.

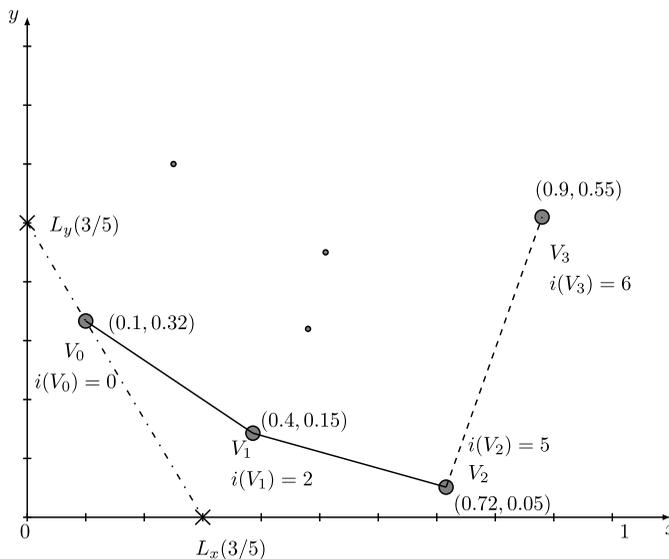


FIGURE 1. Notation illustration. Solid line is  $\mathcal{C}^+(x_i, y_i)$ , dashed line is  $\mathcal{C}^-(x_i, y_i)$ , dash-dotted line is  $L(3/5)$ , which supports  $V_0$ . Here  $\Pi_n(\mathcal{C}(x_i, y_i)) = (2, 3, 1)$  and  $\Pi(\mathcal{C}(x_i, y_i)) = (0.1, 0.3, 0.32, 0.18, 0.1)$ .

## 2. BACKGROUND

We now derive the connection between the zeros of our tropical polynomial  $\mathcal{T}f$  in (1) and the lower convex hull  $\mathcal{C}(i, C_i)$ . This is a special case of a classical result in tropical algebraic geometry, see [27, §2], which has been rediscovered several times across different literature [4, 37]. Write

$$g(x) = -\mathcal{T}f(x) = \max_{i=0, \dots, n} (-C_i) + (-i)x.$$

Then  $g$  is a convex, piecewise linear function. Its Legendre transform  $\hat{g}$  is also convex and piecewise linear, given by

$$\hat{g}(w) = \sup_x (wx - g(x)).$$

For  $w \in [-n, 0]$ ,  $\hat{g}(w)$  is finite, and  $-\hat{g}(w)$  is the  $y$ -intercept of the tangent to the graph of  $g$  with slope  $w$ . Since  $g$  is piecewise linear, the tangent line to  $g$  only changes at its angular points (the points of discontinuity of its slope). From [4, 37], we have that

$$\hat{g}(w) = \inf \left\{ \sum_i \omega_i C_i : \sum_i (-i)\omega_i = w, \omega_i \geq 0, \sum_i \omega_i = 1 \right\}.$$

Thus, over  $[-n, 0]$ , the graph of  $\hat{g}$  is precisely  $\mathcal{C}(-i, C_i)$ , the lower faces of the convex hull of the set of points  $\{(-i, C_i) : i = 0, \dots, n\}$ . The angular points of  $g$ , which are

the zeros of  $\mathcal{T}f$ , are hence in bijective correspondence to the slopes of the faces of  $\hat{g}$ . There is also a clear bijection between the faces of  $\mathcal{C}(-i, C_i)$  and those of  $\mathcal{C}(i, C_i)$ . We summarize these observations below.

**Lemma 3.** *For  $\mathcal{T}f$  in (1), there is a bijection between the zeros of  $\mathcal{T}f$  and the faces of  $\mathcal{C}(i, C_i)$ . The multiplicity of a zero is the lattice length of the corresponding face. In particular, the number of zeros of  $\mathcal{T}f$ , counting multiplicity, is the number of faces of  $\mathcal{C}(i, C_i)$ , counting their lattice lengths.*

In light of Lemma 3, it may be more natural to work with the *max-plus* tropical algebra, where the polynomials are convex functions. Indeed, the max-plus algebra found applications in many areas [3, 8, 14, 36]. We chose to work with min-plus, following the convention of tropical algebraic geometry [27].

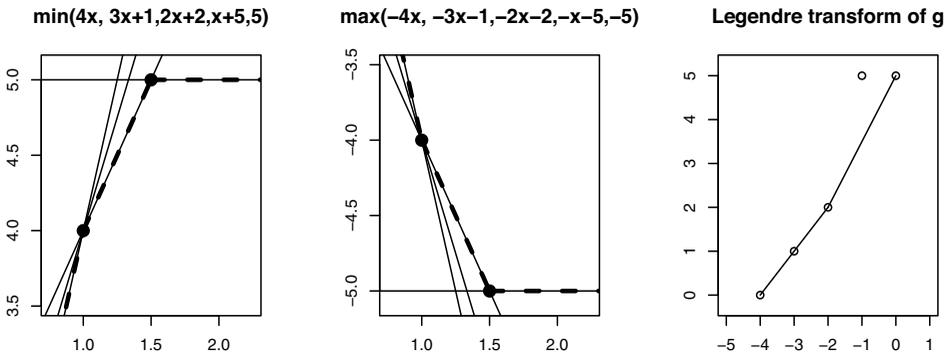


FIGURE 2. Graph of  $\mathcal{T}f$ ,  $g$  and the Legendre transform of  $g$ . The dotted thick lines are the actual graph of  $\mathcal{T}f$  and  $g$ , the continuous lines are the graph of the individual terms.

**Example 4.** Consider  $\mathcal{T}f(x) = 5 \oplus 5x \oplus 2x^2 \oplus 1x^3 \oplus x^4$ . Figure 2 shows the graph of  $\mathcal{T}f$ ,  $g$  and the Legendre transform of  $g$ . In this example,  $\mathcal{T}f$  has a double zero at  $x = 1$  and another zero at  $x = 3/2$ . The Legendre transform of  $g$  shows a face of lattice length 2 with slope 1, corresponding to the double zero at  $x = 1$ , and a face of lattice length 1 with slope  $3/2$ , corresponding to the zero at  $x = 3/2$ . The projection of the Legendre transform of  $g$  onto the  $x$ -axis creates a subdivision of the line  $-[0, 4]$ , which we identify with the line  $[0, 4]$ . In this case the partition  $\Pi_4(\mathcal{C}(i, C_i))$  is  $(2, 1, 1)$ .

### 3. STICK-BREAKING PROOF FOR THE EXPONENTIAL CASE

We now derive a proof of Theorem 1 for the case  $F = Exp(1)$  by considering the partition  $\Pi_n(\mathcal{C}(i, C_i))$ . Computations in this case are significantly simpler, and they give insights into the proof of the general case. Furthermore, this setup connects our results with those in the literature, as discussed in Section 5. Here  $a = 1$ , and (2) reads

$$\frac{Z_n - \frac{4}{3} \log(n)}{\sqrt{20/27} \log(n)} \xrightarrow{d} \mathcal{N}(0, 1).$$

Recall that  $V_0^\uparrow$  is the vertex of  $\mathcal{C}(i, C_i)$  with minimal  $C_i$ . Since  $F$  is continuous, its index  $i(V_0^\uparrow)$  is a.s. unique. It is uniformly distributed on  $\{0, \dots, n\}$ . Conditioned on  $i(V_0^\uparrow) = k$ , the random variables  $C_i - C_k$  are i.i.d. *Exp*(1). The partitions  $\Pi_k(\mathcal{C}^+(i, C_i))$  and  $\Pi_{n-k}(\mathcal{C}^-(i, C_i))$  are distributed as independent partitions  $\Pi_k(\mathcal{C}(i, C_i))$  and  $\Pi_{n-k}(\mathcal{C}(i, C_i))$  conditioned on the last  $C_i$  to be zero. Thus it is sufficient to derive the distribution of  $Z_n$  conditioned on the event  $\{C_0 = 0\}$ , in which case  $\mathcal{C}(i, C_i) = \mathcal{C}^-(i, C_i)$ .

**Proposition 5.** *For  $n = 1, 2, \dots$ , let  $p_n$  be the joint distribution function of  $\Pi_n(\mathcal{C}^-(i, C_i))$ . Then for all  $m \in \{1, 2, \dots, n\}$  and  $x_1, \dots, x_m \in \mathbb{N}$  with  $\sum_{i=1}^m x_i = n$ ,*

$$(5) \quad p_n(x_1, \dots, x_m) = \prod_{i=1}^m \frac{x_i}{\binom{n-s_i+1}{2}},$$

where  $s_i = \sum_{j < i} x_j$ ,  $s_1 = 0$ .

*Proof.* In this case, for any vertex  $V$ ,  $x(V) = i(V)$ . Thus the partition  $\Pi_n(\mathcal{C}^-(i, C_i))$  identifies with the sequence  $(X_j, j = 1, \dots, |\mathcal{C}^-(i, C_i)|)$  where  $X_j = x(V_j) - x(V_{j-1})$ . Equation (5) is the probability of the event  $\{X_j = x_j, j = 1, \dots, m = |\mathcal{C}^-(i, C_i)|\}$ .

Note that  $x(V_1^\uparrow)$  is distributed as the argmin of  $\{C_i/i, i = 1, \dots, n\}$ , where the  $C_i$ 's are i.i.d. *Exp*(1). Thus

$$(6) \quad \mathbb{P}(X_1 = x_1) = \frac{x_1}{\binom{n+1}{2}}.$$

Conditioned on  $x(V_1) = k$ , for  $i = k+1, \dots, n$ ,  $C_i \stackrel{d}{=} \frac{i}{k}C_k + \epsilon_i$  where  $\epsilon_i$  are i.i.d. *Exp*(1). Thus,  $x(V_2^\uparrow)$  is distributed as the argmin of

$$\frac{C_i - C_k}{i - k} = \frac{C_k}{k} + \frac{\epsilon_i}{i - k},$$

$i = k+1, \dots, n$ , which is the integer that achieves the minimum of  $\frac{\epsilon_i}{i - k}$ . So conditioned on  $x(V_1^\uparrow) = k$ ,  $x(V_2^\uparrow)$  is distributed as

$$\mathbb{P}(x(V_2) = k') = \frac{k' - k}{\binom{n-k+1}{2}}.$$

Now,  $X_2 = x(V_2) - x(V_1)$ . Thus conditioned on  $X_1 = x_1$ ,  $X_2$  is distributed as

$$\mathbb{P}(X_2 = x_2) = \frac{x_2}{\binom{n-k+1}{2}} = \frac{x_2}{\binom{n-s_2+1}{2}}.$$

Repeating the argument for  $j = 3, 4, \dots$  completes the proof.  $\square$

If we just keep track of the number of components of  $\Pi_n(\mathcal{C}^-(i, C_i))$ , then the recursion argument of Proposition 5 translates into the following recurrence relation, which appears as [11, Theorem 1]. We expand on this connection in Section 5.

**Corollary 6.** *Let  $p_k^n$  be the probability that the partition  $\Pi_n(\mathcal{C}^-(i, C_i))$  has  $k$  components. Then*

$$(7) \quad p_k^n = \frac{1}{\binom{n+1}{2}} \sum_{j=k-1}^{n-1} (n-j)p_{k-1}^j.$$

The recursion in Proposition 5 has a Polya’s urn scheme interpretation. Recall the basic setup: start with  $w$  white balls and  $b$  black balls. At each step, remove a random ball from the urn and replace it with two balls of the same color. Repeat this process  $n$  times, thus adding in total  $n$  balls to the urn. Let  $W_n$  denote the number of white balls newly added. Then for  $x \in \{0, 1, \dots, n\}$ ,

$$\mathbb{P}(W_n = x) = \binom{n}{x} \frac{(w+x-1)!(b+n-x-1)!(w+b-1)!}{(w-1)!(b-1)!(w+b+n-1)!}.$$

In particular, for  $w = 2, b = 1$ , then

$$(8) \quad \mathbb{P}(W_n = x) = \frac{x+1}{\binom{n+2}{2}}.$$

Thus, for each  $n$ ,  $X_1 - 1$  is distributed as the number of newly added white balls after  $n - 1$  steps of the Polya’s urn scheme starting with *two* white balls and one black ball. Conditioned on  $X_1 = x_1$ ,  $X_2 - 1$  is distributed as the number of newly added white balls after  $n - x_1 - 1$  steps of the same urn scheme, and so on. From the sequential description of Polya’s urn, the sequence of partition functions  $(p_n, n \geq 1)$  can be generated via the following variation of the *Chinese Restaurant Process* (CRP). Introduced by Dubins and Pitman [29, §3], this is a model for consistent random permutations. Start with an initially empty restaurant with an unlimited number of tables numbered  $1, 2, \dots$ , each capable of seating an unlimited number of customers. Customers numbered  $1, 2, \dots$  arrive one by one and choose their seats according to the following rules.

---

**Algorithm 1** *Beta*(2, 1) Chinese Restaurant Process

---

- 1: Customer 1 sits at table 1.
  - 2: Suppose after  $n$  customers, tables  $1, 2, \dots, k$  are occupied with  $x_1, \dots, x_k$  customers,  $\sum_i x_i = n$ . The  $(n + 1)$ -st customer joins according to the following rules:
    - He joins table 1 with probability  $\frac{x_1 + 1}{n + 2}$ .
    - Sequentially for  $i = 2, \dots, m$ , conditioned on not joining the previous tables  $1 \leq j \leq i - 1$ , he joins table  $i$  with probability  $\frac{x_i + 1}{n - s_j + 2}$ , where  $s_j = \sum_{j < i} x_i$ .
    - Conditioned on not joining any of the previous tables, he forms a new table.
- 

**Example 7** ( $n = 3$ ). The tree in Figure 3 computes the distribution of the random partition of 3 customers.

**Lemma 8.** *The partition of  $n$  generated by the above CRP equals  $\Pi_n(\mathcal{C}^-(i, C_i))$ . In other words, for  $n = 1, 2, \dots$ , the probability that tables  $1, 2, \dots, m$  have sizes  $x_1, \dots, x_m$  is exactly  $p_n(x_1, \dots, x_m)$  in (5).*

*Proof.* The claim is evident from the Polya’s urn construction. Let us keep track of  $N_1$ , the size of the first table in the CRP. After  $n$  steps,  $N_1 - 1$  is distributed as the number of newly added white balls in the (2, 1)-Polya’s urn scheme after  $n - 1$  steps. Thus,  $N_1 \stackrel{d}{=} X_1$ . Similarly, let  $N_2$  be the size of the second table. By construction,  $N_2$  is distributed as the number of white balls in the (2, 1)-Polya’s

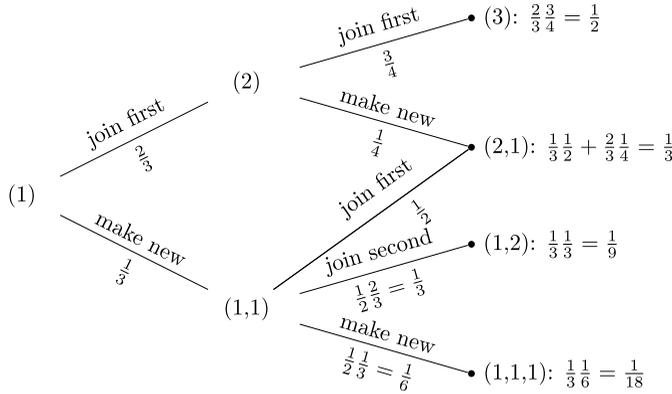


FIGURE 3

urn scheme after  $n - N_1 - 1$  steps. Therefore,  $N_2 \stackrel{d}{=} X_2$ . Repeating this argument proves the claim.  $\square$

As  $n \rightarrow \infty$ , the sequence of relative frequencies  $(X_i/n, i \geq 1)$  converges in distribution to the continuous *Beta(2, 1) stick-breaking sequence*

$$(9) \quad (P_1, P_2, \dots) = (B_1, \bar{B}_1 B_2, \bar{B}_1 \bar{B}_2 B_3, \dots),$$

where  $B_i$  are i.i.d. *Beta(2, 1)* random variables, and  $\bar{B}_i = 1 - B_i$  for  $i \geq 1$ . This sequence defines a distribution on the relative frequencies of a random partition  $\Pi$  of  $\mathbb{N}$ . Let  $p : \mathbb{N}^* = \bigcup_{k=1}^\infty \mathbb{N}^k \rightarrow [0, 1]$  be the partition function of  $\Pi$ .

**Lemma 9.** *Let  $\Pi'_n$  be the restriction of  $\Pi$  to  $[n]$ . Then  $\Pi_n(\mathcal{C}^-(i, C_i)) \stackrel{d}{=} \Pi'_n$ .*

*Proof.* By [28], the probability that  $\Pi'_n$  equals any specific partition  $(x_1, \dots, x_m)$  of  $[n]$ , in order of appearance, is given by

$$(10) \quad p_n(x_1, \dots, x_m) = \mathbb{E} \left( \prod_{i=1}^m P_i^{x_i-1} \prod_{i=1}^{m-1} (1 - \sum_{j=1}^i P_j) \right).$$

By a direct computation, we find that (10) equals (5). This proves the lemma.  $\square$

We call  $\Pi_n(\mathcal{C}^-(i, C_i))$  the *discrete Beta(2, 1) stick-breaking scheme*. It follows from [28, Theorem 6] that there is another CRP representation for  $\Pi_n(\mathcal{C}^-(i, C_i))$ , this time conditioned on the limiting sequence  $(P_1, P_2, \dots)$ . Specifically, given  $(P_1, P_2, \dots)$ , and given that the partition  $\Pi_n(\mathcal{C}^-(i, C_i))$  has sizes  $(x_1, \dots, x_m)$ ,  $\Pi_{n+1}(\mathcal{C}^-(i, C_i))$  is an extension of  $\Pi_n(\mathcal{C}^-(i, C_i))$  in which the  $(n + 1)$ -st customer does the following:

- Joins table  $i$  with probability  $P_i, 1 \leq i \leq m$ .
- Joins a new table with probability  $1 - \sum_{j=1}^m P_j$ .

In other words, conditioned on  $(P_i, i \geq 1)$ , for fixed  $n \geq 1, X_1 \stackrel{d}{=} \text{Binomial}(n, P_1)$ . Conditioned on  $X_1, X_2 \stackrel{d}{=} \text{Binomial}(n - X_1, P_2)$ , and so on. When  $P_i = \prod_{j < i} \bar{B}_j B_i$  for i.i.d.  $B_j \sim B$  for some random variable  $B$  as in our case, this is also known as the *Bernoulli sieve* model, a recursive allocation of  $n$  balls in infinitely many boxes  $j = 1, 2, \dots$ . Here  $X_j$  is the number of balls in the  $j$ -th box (if a box  $j$  is not

discovered, then  $X_j = 0$ ). Gnedin [16] and Gnedin et al. [17] studied the various functionals of this model, including  $|\Pi_n(\mathcal{C}^-(i, C_i))|$ , the number of boxes occupied by at least one ball. Through methods from renewal theory, they showed that  $|\Pi_n(\mathcal{C}^-(i, C_i))|$  has the same asymptotics as the number of renewals on the interval  $[0, \log(n)]$  of a renewal process whose inter-arrival time is distributed as  $-\log(B)$ . Specifically, for  $\mu = \mathbb{E}(-\log(B))$  and  $\sigma^2 = \text{Var}(-\log(B))$ ,

$$(11) \quad \frac{|\Pi_n(\mathcal{C}^-(i, C_i))| - \mu^{-1} \log(n)}{\sqrt{\sigma^2 \mu^{-3} \log(n)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

In our case,  $B \stackrel{d}{=} \text{Beta}(2, 1)$ , thus  $\mu = \frac{3}{2}, \sigma^2 = \frac{5}{4}$ . Substituting into (11) yields

$$(12) \quad \frac{|\Pi_n(\mathcal{C}^-(i, C_i))| - \frac{2}{3} \log(n)}{\sqrt{\frac{10}{27} \log(n)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

We can now complete the proof of Theorem 1 for the case  $F = \text{Exp}(1)$ .

*Proof of Theorem 1 when  $F$  is  $\text{Exp}(1)$ .* As previously argued, we have

$$Z_n = |\Pi_{U_n}(\mathcal{C}^-(i, C_i))| + |\Pi'_{n-U_n}(\mathcal{C}^-(i, C_i))|$$

where  $U_n$  is the discrete uniform distribution on  $\{0, 1, \dots, n\}$ , and conditioned on  $U_n = u$ ,  $|\Pi_u(\mathcal{C}^-(i, C_i))|$  and  $|\Pi'_{n-u}(\mathcal{C}^-(i, C_i))|$  are independent, with asymptotics given in (12). Thus  $Z_n$ , appropriately scaled, also converges to the standard normal distribution. It remains to find  $\mathbb{E}(Z_n)$  and  $\text{Var}(Z_n)$ . Couple  $U_n$  with a continuous uniform  $U \stackrel{d}{=} \text{Uniform}(0, 1)$  in the obvious way. Then  $|U_n/n - U| < \frac{1}{n}$  a.s. Therefore,

$$\begin{aligned} \mathbb{E}(Z_n|U = u) &= \frac{2}{3} (\log(nu) + \log(n(1-u))) + O(1) \\ &= \frac{4}{3} \log(n) + \frac{2}{3} (\log(u) + \log(1-u)) + O(1). \end{aligned}$$

Therefore,  $\mathbb{E}(Z_n) = \frac{4}{3} \log(n) + O(1)$ . Similarly,

$$\begin{aligned} \text{Var}(Z_n|U = u) &= \frac{10}{27} (\log(nu) + \log(n(1-u))) + O(1) \\ &= \frac{20}{27} \log(n) + \frac{10}{27} (\log(u) + \log(1-u)) + O(1). \end{aligned}$$

By the above calculation,  $\mathbb{E}(\text{Var}(Z_n|U)) = \frac{20}{27} \log(n) + O(1)$ , and

$$\text{Var}(\mathbb{E}(Z_n|U)) = \text{Var}\left(\frac{2}{3}(\log(U) + \log(1-U))\right) = O(1).$$

Thus,  $\text{Var}(Z_n) = \frac{20}{27} \log(n) + O(1)$ . The central limit theorem for  $Z_n$  follows by the Continuous Mapping Theorem (see, for example, [13, §2]).  $\square$

For any  $a > 0$ , we can define a discrete  $\text{Beta}(2, a)$  stick-breaking process. One could ask whether there exists some  $F$  such that  $\Pi_n$  a partition from such a process. In the hindsight of the proof of Proposition 2, such a process could arise when  $F$  is the  $\text{gamma}(a, b)$  distribution for some  $b > 0$ . However, Proposition 5 relies on the lack of memory of exponentials to obtain the recursion. We do not know how to generalize this to other distributions in the gamma family.

## 4. POISSON PROOF FOR THE EXPONENTIAL CASE

We now prove Proposition 2 when  $F$  is  $Exp(1)$ . Here  $G$  is the Lebesgue measure on  $(0, \infty)$ ,  $\Phi_n$  is a homogeneous Poisson point process on the first quadrant with rate  $n$ , and the rectangle in consideration is the square  $(0, 1)^2$ . As discussed in the introduction, in this case, Proposition 2 is a version of Groeneboom's result [19].

We now give an elementary proof of this result using stick-breaking. The key idea of this proof, Lemma 11 below, appeared as Corollary 2 in a later paper of Groeneboom [20]. However, he did not make the connection to stick-breaking and renewal theorems explicit. Our proof clarifies the connection between the discrete and continuous setup, a key idea behind our main theorems. We discuss this in Section 5, together with connections to the results of Buchta [11, 12] and Groeneboom [19, 20].

We first consider the lower convex hull  $\mathcal{C}^+(\Phi_n)$  only. At the end of Section 1 we introduced the process  $(x(\alpha), y(\alpha))$ ,  $\alpha \in (0, \infty)$  of vertices of  $\mathcal{C}^+(\Phi_n)$  indexed by the slope of their support vectors. Note that  $(y(\alpha), \alpha \in (0, \infty))$  is a pure-jump process indexed by increasing  $\alpha$ . The sequence of vertices of this process is precisely the sequence  $(V_i^\downarrow, i \geq 0)$  of vertices of  $\mathcal{C}^+(\Phi_n)$  ordered by decreasing  $y$ -values. The number of jumps is precisely  $|\mathcal{C}^+(\Phi_n)|$ . Groeneboom [19] proved the following.

**Lemma 10** (by [19]). *The process  $((x(\alpha), y(\alpha)), \alpha \in (0, \infty))$  is a pure-jump Markov process indexed by  $\alpha$ . Its number of jumps on  $(0, \infty)$ ,  $|\mathcal{C}^+(\Phi_n)|$ , is a.s. finite.*

The infinitesimal generator for the process  $(x(\cdot), y(\cdot))$  is rather complicated; see [19]. Remarkably, the transition rule in the  $y$ -coordinate is very simple. Groeneboom discovered this fact in his more recent paper [20].

**Lemma 11** (by [19]). *Let  $(B_i, i \geq 1)$  be an infinite sequence of i.i.d.  $Beta(1, 2)$  random variables, independent of  $y(V_0^\downarrow)$ . The transition rule in  $y(\cdot)$  is as follows. Given that  $V_i^\downarrow = (x^*, y^*)$  for some  $i \geq 0$ , the process  $y(\cdot)$  either terminates with probability  $\exp(-n \frac{(1-x^*)y^*}{2})$ , in which case  $i = |\mathcal{C}^+(\Phi_n)|$ , or it jumps to a uniform random point  $y(V_{i+1}^\downarrow)$ ,  $0 < y(V_{i+1}^\downarrow) < y^*$ , where*

$$y(V_{i+1}^\downarrow) \stackrel{d}{=} y^* B_{i+1}.$$

*In other words,  $y(V_0^\downarrow) \stackrel{d}{=} Uniform(0, 1)$  and  $y(V_i) = y(V_0) \prod_{j=1}^i B_j$  for all  $i \leq |\mathcal{C}^+(\Phi_n)|$ .*

*Proof.* The proof is by induction on  $i$ . Note that  $y(V_0^\downarrow)$  is distributed as a  $Uniform(0, 1)$ . Almost surely,  $y(V_0^\downarrow) = y(\alpha_0)$  for some  $\alpha_0 > 0$ . As  $\alpha$  increases, the event that  $y(\cdot)$  jumps from  $y(V_0^\downarrow)$  to  $y(V_1^\downarrow)$  in  $(\alpha, \alpha + d\alpha)$  is the event that there is at least one point in the shaded triangle in Figure 4.

Conditioned on  $V_1^\downarrow$  being in this triangle, it is uniformly distributed. In particular,  $\mathbb{P}(y(V_1^\downarrow) \in dy)$  is proportional to the length of the line highlighted, which is exactly proportional to  $y(V_0^\downarrow) - y$ . Note that this is independent of  $\alpha$ , as long as  $\alpha > 0$ . Thus  $B_1 \stackrel{d}{=} Beta(1, 2)$ , and is independent of  $y(V_0^\downarrow)$ . Repeating this argument shows that  $B_i$  as defined is distributed as  $Beta(2, 1)$ , independent of  $y(V_{i-1}^\downarrow)$  and thus of other  $B_j$ 's, as long as  $\alpha > 0$ . The event of termination at some

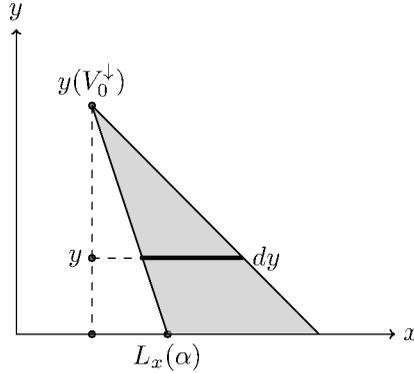


FIGURE 4. The process  $y(\cdot)$  has a jump in  $d\alpha$  if there is at least one point in the triangle. Conditioned on its existence, this point is uniformly distributed.

point  $(x^*, y^*)$  is exactly when the point process  $\Phi_n$  has no points in the rectangle with vertices  $(x^*, y^*)$ ,  $(1, y^*)$ ,  $(0, y^*)$  and  $(1, 0)$ . The probability of this event is  $\exp(-n(1 - x^*)y^*/2)$ .  $\square$

Lemma 11 explains the appearance of  $Beta(1, 2)$ :  $y(V_0^\downarrow) - y(V_1^\downarrow)$  is a *size-biased* pick from the  $Uniform(0, y(V_0^\downarrow))$  distribution. On the phenomenon of size-bias in random combinatorial structures; see [7, 31]. In fact, the discrete  $Beta(1, 2)$  appearing in Proposition 5 is also a size-bias phenomenon. We make the connection between Lemma 11 and Proposition 5 explicit in Section 5.

Excluding the stopping rule,  $y(\cdot)$  is a discrete time Markov chain with some internal filtration. At each step, it terminates with some stopping rule that depends on the bigger filtration generated by the Markov process  $(x(\cdot), y(\cdot))$ . To avoid dealing with the bigger process, we couple  $y(\cdot)$  with the Markov chain  $(y'_i, i \geq 0)$ , by taking  $y'_0 = y(V_0^\downarrow)$  and  $y'_i = y'_0 \prod_{j=1}^i B_j$ , for the same sequence  $(B_i, i \geq 1)$  of i.i.d.  $Beta(2, 1)$  random variables appearing in Lemma 11. For any  $t > 0$ , let  $J'(t)$  be the first time  $i$  when  $y'_i < e^{-t}$ . Clearly  $J'(t)$  is a stopping time for the Markov chain  $(y'_i, i \geq 0)$ . In fact, it is the number of renewals on  $[0, t]$  of a delayed renewal process with i.i.d. inter-arrival distribution  $-\log(Beta(1, 2))$ , and first point distributed as  $-\log(Uniform(0, 1))$ . As  $t \rightarrow \infty$ , by the strong law of large numbers for delayed renewal process [21, Theorem 2.5.1],

$$(13) \quad \frac{J'(t)}{t} \xrightarrow{\text{a.s.}} \frac{1}{\mu},$$

where  $\mu = \mathbb{E}(-\log(Beta(1, 2))) = \frac{3}{2}$ .

The point of the coupling is that  $J'(\log(n))$  (defined w.r.t.  $(y'_i, i \geq 0)$ ) and  $|\mathcal{C}^+(\Phi_n)|$  (defined w.r.t. the initial process) are close with high probability. Indeed, these two quantities are closed as long as  $y(V_0^\uparrow) = y(V_{|\mathcal{C}^+(\Phi_n)|}^\downarrow)$  is close to  $\frac{1}{n}$ .

**Lemma 12.** *Let  $\delta(n) = |\mathcal{C}^+(\Phi_n)| - J'(\log(n))$ . Then  $\delta(n) = \mathcal{O}_P(1)$ .*

*Proof.* We have to prove that for all functions  $\ell : \mathbb{R} \rightarrow \mathbb{R}_+$  growing to infinity arbitrarily slowly,

$$\mathbb{P} \left\{ J'(\log(n)) - \ell(n) \leq |\mathcal{C}^+(\Phi_n)| \leq J'(\log(n)) + \ell(n) \right\} \rightarrow 1,$$

when  $n \rightarrow \infty$ . By (13) and the strong law of large numbers, it is sufficient to show that for all  $\ell$  as above,

$$(14) \quad \mathbb{P} \left\{ \log(n) - \ell(n) \leq -\log(y(V_0^\uparrow)) \leq \log(n) + \ell(n) \right\} \rightarrow 1,$$

when  $n \rightarrow \infty$ . We now prove (14) by using the fact that  $y(V_0^\uparrow)$  is the minimum of the  $y$ -coordinates of the points of  $\Phi_n$  in  $(0, 1)^2$ . For this, we divide the region  $(0, 1) \times (0, \infty)$  into  $n$  vertical strips  $(i/n, (i+1)/n) \times (0, \infty)$  for  $i$  in  $\{0, \dots, n-1\}$ . In each strip, the  $y$ -coordinate of the points of  $\Phi_n$  form a homogeneous Poisson point process on  $(0, \infty)$ . In particular, in the  $i$ -th strip, the first  $y$  jump, denoted by  $\epsilon_i$ , is distributed as a standard exponential. Let  $Y' := \min(\epsilon_1, \dots, \epsilon_n) \stackrel{d}{=} \text{Exp}(n)$ . Then, conditioned on the event  $Y' < 1$ , which happens with probability  $1 - e^{-n}$ ,  $Y' = y(V_0^\uparrow)$ . Now,

$$\mathbb{P} \left( \frac{1}{n\ell(n)} \leq Y' \leq \frac{\ell(n)}{n} \right) = e^{-1/\ell(n)} - e^{-\ell(n)}.$$

Thus the probability of the event in (14) is at least

$$(1 - e^{-n})(e^{-1/\ell(n)} - e^{-\ell(n)}),$$

which concludes the proof.  $\square$

*Proof of Proposition 2 when  $F$  is exponential.* Write

$$|\mathcal{C}(\Phi_n)| = |\mathcal{C}^+(\Phi_n)| + |\mathcal{C}^-(\Phi_n)| = J'(\log(n)) + J''(\log(n)) + \delta^+(n) + \delta^-(n).$$

Here  $J'$  and  $J''$  are stopping times of independent versions of the Markov chain  $y'$  previously defined, and the terms  $\delta^+(n)$ ,  $\delta^-(n)$  are error terms. Conditioned on the minimum  $y$ -coordinate  $y(V_0^\uparrow)$ , the pair  $|\mathcal{C}^+(\Phi_n)|$  and  $|\mathcal{C}^-(\Phi_n)|$ , hence the pair  $\delta^+(n)$  and  $\delta^-(n)$ , are independent. By (14),  $\delta^+(n)$  and  $\delta^-(n)$  are both  $\mathcal{O}_P(1)$ ; therefore, so is their sum. So

$$|\mathcal{C}(\Phi_n)| = J'(\log(n)) + J''(\log(n)) + \mathcal{O}_P(1).$$

The asymptotics of  $J'(\log(n))$  are the same as that of  $|\Pi_n(\mathcal{C}^-(i, C_i))|$  in (12). Thus Proposition 2 follows by the Continuous Mapping Theorem.  $\square$

## 5. CONNECTIONS BETWEEN THE DISCRETE AND CONTINUOUS SETUP

For general distribution  $F$ , we present a coupling argument in Lemma 12 to compare Theorem 1 and Proposition 2. When  $F = \text{Exp}(1)$ ,  $G = \lambda$ , the coupling can be described even more explicitly, and in fact, it results in an equality in distribution.

**Lemma 13.** *For  $i = 0, 1, \dots, n$ , let  $U_i$  be i.i.d. Uniform(0, 1), independent of the  $C_i$ 's. Then*

$$(15) \quad \Pi_n(\mathcal{C}(U_i, C_i)) \stackrel{d}{=} \Pi_n(\mathcal{C}(i, C_i)),$$

and conditioned on  $\Phi_n$  having  $n + 1$  points in  $(0, 1)^2$ ,

$$(16) \quad \Pi(\mathcal{C}(U_i, C_i)) \stackrel{d}{=} \Pi(\mathcal{C}(\Phi_n)).$$

*Proof.* Equation (15) follows from repeating the proof of Proposition 5 for the points  $(U_i, C_i)$  instead of  $(i, C_i)$ . For the second statement, note that conditioned on  $\Phi_n$  having  $n + 1$  points in  $(0, 1)^2$ , these points are distributed as  $(U_i, \frac{C_i}{\sum_i C_i + \epsilon})$  for some independent  $\epsilon \stackrel{d}{=} \text{Exp}(1)$ . In other words, these points are a version of the points  $(U_i, C_i)$ , rescaled by some random amount in the  $y$ -axis. But such rescaling does not change the partition of  $[0, 1]$ , thus we have (16).  $\square$

**Lower convex hull of points in a triangle.** Buchta [11] considered  $|\mathcal{C}(\Delta)|$ , the number of faces in the lower convex hull of  $(0, 1)$ ,  $(1, 0)$  and  $n$  points distributed uniformly at random in the triangle  $\Delta$  with vertices  $(0, 1)$ ,  $(0, 0)$  and  $(1, 0)$ . Theorem 1 in [11] derives a recurrence relation for  $\mathbb{P}(|\mathcal{C}(\Delta)| = k)$ , which is exactly our  $p_k^n$  in (7). In the light of Lemmas 11 and 13, this connection is clear.

**Corollary 14.** *The partition of  $[0, 1]$  induced by projecting the lower faces of  $\mathcal{C}(\Delta)$  onto the  $x$ -axis is precisely the partition of  $[0, 1]$  obtained by projecting faces of  $\mathcal{C}^+(\Phi_n)$  onto  $[0, y(V_0^\downarrow)]$  in the  $y$ -axis, and rescaled by  $y(V_0^\downarrow)$ .*

Lemma 13 shows that Groeneboom's result, stated in the form of (3), follows directly from the asymptotics of the  $Beta(2, 1)$  stick-breaking in Section 3. This was the spirit of Buchta's approach [11, 12]. In particular, he derived  $\mathbb{E}(|\mathcal{C}(\Delta)|)$  and  $\text{Var}(|\mathcal{C}(\Delta)|)$  exactly for finite  $n$  using (7). He then generalized this approach to derive the exact analogues for results in Groeneboom [19], including distribution of the number of vertices and the area outside the convex hull of a uniform sample from the interior of a convex polygon with  $r$  vertices.

## 6. LOWER CONVEX HULL OF AN INHOMOGENEOUS POISSON POINT PROCESS

We now prove Proposition 2 in the general case. Our strategy is to generalize the proof given in the previous section. Consider the vertices  $(V_i^\downarrow, i \geq 0)$  of  $\mathcal{C}^+(\Phi_n)$  ordered in decreasing  $y$ -coordinate. The analogue of Lemma 11 is the following.

**Lemma 15.** *For each  $i \geq 1$ , let  $B_i = \frac{y(V_i^\downarrow)}{y(V_{i-1}^\downarrow)}$ . For  $s > 0$ , define the distribution  $I_s$  on  $[0, 1]$  via its cdf (also denoted  $I_s$ ):*

$$I_s(b) := \mathbb{P}(B_i \leq b \mid y(V_{i-1}^\downarrow) = s).$$

*Then with  $F$  defined in the statement of Proposition 2 and  $G$  defined in (4),*

$$(17) \quad I_s(b) = \frac{(1-b)G([0, sb]) + \int_0^b G([0, st])dt}{\int_0^1 G([0, st])dt}, \quad 0 \leq b \leq 1.$$

*In particular, as  $s \rightarrow 0$ ,  $I_s \rightarrow I_0 := Beta(a, 2)$ .*

*Proof.* Consider Figure 4. Conditioned on  $y(V_{i-1}^\downarrow) = s$  and on the fact that the shaded triangle contains a point, the probability that  $y(V_i^\downarrow)$  belongs to  $dy$  at  $y = sb$  is proportional to the area of the thin boldface rectangle of the figure w.r.t. the  $\lambda \times G$  measure, that is, proportional to  $(1-b)dG([0, sb])$ . This in turn implies the above formula for  $I_s$ , the cumulative distribution function of  $B_i$  given  $y(V_{i-1}^\downarrow) = s$ . Now, as  $s \rightarrow 0$ ,  $G([0, sb]) \rightarrow C(sb)^a$  uniformly over all  $b \in [0, 1]$ . Thus  $\frac{d}{db}G([0, sb]) \rightarrow Cas^ab^{a-1}$ . The term  $s^a$  cancels, leaving  $I_s(db) \rightarrow I_0(db)$  at all points  $b \in [0, 1]$ , where  $I_0(db) \propto b^{a-1}(1-b)$ . Thus,  $I_0 = Beta(a, 2)$ . Since  $I_0$  is continuous everywhere,  $I_s \rightarrow I_0$  as distributions. (See, for example, [13, §2]).  $\square$

Define  $S_0 := -\log(y(V_0^\downarrow))$  and for  $i \geq 1$ ,

$$S_i := -\log(y(V_i^\downarrow)) = -\log(y(V_0^\downarrow)) + \sum_{j=1}^i -\log(B_j).$$

Then  $(S_i, i = 0, 1, \dots)$  is a state-dependent random walk. At the  $i$ -th step, conditioned on  $S_i = t$ ,  $S_{i+1} - S_i$  is an independent random variable with distribution  $-\log(I_{\exp(-t)})$  for the family of distributions  $I_{(\cdot)}$  in Lemma 15. Alternatively,  $(S_i)$  can be viewed as a Markov renewal process, or a Markov modulated random walk. In these settings, the inter-arrival times (or the jumps of the walk) are driven by a Markov process, in this case,  $(S_i)$  itself. Since the jumps are a.s. positive, the Markov chain is transient. To the best of our knowledge, Korshunov [24, 25] is one of the only authors who proved a central limit theorem for transient Markov renewal processes. Specialized to our case, this result [24, Theorem 5] is as follows.

**Theorem 16** (Korshunov's Central Limit Theorem for  $S_i$  [24]). *For  $s \geq 0$ , let  $\xi_s$  be a random variable with distribution  $I_s$ . Let  $\mu = \mathbb{E}(-\log(\xi_0))$ ,  $\sigma^2 = \text{Var}(-\log(\xi_0))$ . Suppose that*

$$(18) \quad \exists \delta \in (0, 1) \text{ such that } \{(\log(\xi_s))^2, 0 < s < \delta\} \text{ is uniformly integrable,}$$

$$(19) \quad \mathbb{E}(-\log(\xi_s)) = \mu + o\left(\frac{1}{\sqrt{-\log(s)}}\right),$$

$$(20) \quad \text{Var}(-\log(\xi_s)) \rightarrow \sigma^2 \text{ as } s \rightarrow 0.$$

Then as  $s \rightarrow 0$ ,  $t = -\log(s) \rightarrow \infty$ , and

$$(21) \quad S_t/t \xrightarrow{a.s.} \mu, \quad \text{and} \quad \frac{S_t - t\mu}{\sqrt{t\sigma^2}} \xrightarrow{d} \mathcal{N}(0, 1).$$

**Lemma 17.** *The distribution  $I_s$  satisfies (18), (19) and (20). In particular, suppose  $(t_n)$  is an increasing sequence of indices, such that  $t_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Define  $s_n = e^{-t_n}$ . Then as  $n \rightarrow \infty$ ,  $s_n \rightarrow 0$ , and (21) holds for the sequence  $(t_n)$ .*

*Proof.* By the assumptions of Proposition 2, as  $s \rightarrow 0$ ,  $G([0, s]) = Cs^a + \epsilon(s)$ , where  $\epsilon(s) = o(s^a)$ . Thus there exists some small  $S$  such that for all  $s < S$ , for all  $b \in [0, 1]$ ,  $|\epsilon(sb)| < 2|\epsilon(s)|$ . Bound  $I_s(b)$  in (17) by

$$\frac{1 - 2|\epsilon(s)|}{1 + 2|\epsilon(s)|} I_0(b) \leq I_s(b) \leq \frac{1 + 2|\epsilon(s)|}{1 - 2|\epsilon(s)|} I_0(b),$$

where  $I_0(b) = (a+1)b^a - ab^{a+1}$ . For some small  $S'$ , for all  $s < S'$ ,  $|\epsilon(s)| < 1/4$ . So

$$(22) \quad |I_0(b) - I_s(b)| \leq 8|\epsilon(s)|,$$

where this bound holds for all  $s < \min(S, S')$ . Note that  $-\log(\xi_0)$  is a.s. positive and has finite third moments for all  $a > 0$ . For any  $r > 0$ , write

$$\mathbb{E}((-\log(\xi_s))^r) = \int_0^\infty \mathbb{P}((-\log(\xi_s))^r > x) dx = \int_0^\infty \mathbb{P}(\xi_s > \exp(-x^{1/r})) dx.$$

Thus for all  $s < \min(S, S')$ , (22) implies  $|\mathbb{E}(-\log(\xi_s)^3) - \mathbb{E}(-\log(\xi_0)^3)| < (1+8\epsilon(s))$ , so the family of squared jumps  $\{\log(\xi_s)^2, s < \min(S, S')\}$  is uniformly integrable.

Similarly, we have convergence of the expectation and variance. The error in the expectation is bounded by

$$|\mathbb{E}(-\log(\xi_s)) - \mu| < 8|\epsilon(s)|.$$

For  $a > 0$ ,  $s^a = o\left(\frac{1}{\sqrt{-\log(s)}}\right)$ . Since  $\epsilon(s) = o(s^a)$  by assumption,  $\epsilon(s) = o\left(\frac{1}{\sqrt{-\log(s)}}\right)$ . □

**Proposition 18.** *Let  $J(t)$  be the number of jumps in  $[0, t]$  of the random walk  $(S_i)$ . Then as  $t \rightarrow \infty$ ,*

$$(23) \quad \frac{J(t)}{t} \xrightarrow{\text{a.s.}} \frac{1}{\mu}, \quad \text{and} \quad \frac{J(t) - \mu^{-1}t}{\sqrt{\sigma^2\mu^{-3}t}} \xrightarrow{d} \mathcal{N}(0, 1),$$

where  $\mu = \frac{1}{a} + \frac{1}{a+1}$ ,  $\sigma^2 = \frac{1}{a^2} + \frac{1}{(a+1)^2}$ .

*Proof.* This result follows from (21) by standard techniques of renewal theory. Our treatment follows that of Gut [21, §2.5]. Let  $\nu(t)$  be the first time in which  $S_i > t$ . Then  $\nu(t) = J(t) + 1$  a.s. By definition,

$$S_{J(t)} \leq t < S_{\nu(t)}.$$

Therefore,

$$\frac{S_{J(t)}}{J(t)} \leq \frac{t}{J(t)} < \frac{S_{\nu(t)}}{\nu(t)} \cdot \frac{J(t) + 1}{J(t)}.$$

Since the walk has a.s. positive increments,  $J(t) \rightarrow \infty$  as  $t \rightarrow \infty$  a.s. But  $S_t/t \xrightarrow{\text{a.s.}} \mu$  by (21), therefore, by [21, Theorem 2.1],

$$\frac{S_{J(t)}}{J(t)} \xrightarrow{\text{a.s.}} \mu, \quad \text{and} \quad \frac{S_{\nu(t)}}{\nu(t)} \cdot \frac{J(t) + 1}{J(t)} \xrightarrow{\text{a.s.}} \mu.$$

Thus  $\frac{t}{J(t)} \xrightarrow{\text{a.s.}} \mu$ , and  $\frac{J(t)}{t} \xrightarrow{\text{a.s.}} \mu$  by the Continuous Mapping Theorem. Similarly,

$$\frac{S_{J(t)} - J(t)\mu}{\sigma\sqrt{t/\mu}} \leq \frac{t - J(t)\mu}{\sigma\sqrt{t/\mu}} \leq \frac{S_{\nu(t)} - \nu(t)\mu}{\sigma\sqrt{t/\mu}} + \frac{\mu}{\sigma}\sqrt{\frac{\mu}{t}}.$$

As argued above,  $J(t) \rightarrow \frac{t}{\mu}$  a.s., and therefore so does  $\nu(t)$ . By Anscombe’s theorem [21, §1.3], [6], the sequences  $(S_{J(t)})$  and  $(S_{\nu(t)})$  satisfy the same central limit theorem as that of  $S_t$ . Therefore, as  $t \rightarrow \infty$ ,

$$\frac{t - J(t)\mu}{\sigma\sqrt{t/\mu}} = -\frac{J(t) - \mu^{-1}t}{\sqrt{\sigma^2\mu^{-3}t}} \xrightarrow{d} \mathcal{N}(0, 1).$$

□

Recall that  $|\mathcal{C}^+(\Phi_n)| = J[-\log(y(V_0^\uparrow)) + \log(y(V_0^\downarrow))]$ . Since  $F(x) \sim Cx^a$  for  $x$  near 0,  $y(V_0^\uparrow) \sim F^{-1}(\frac{1}{n}) \sim \frac{1}{C}n^{-1/a}$ , so we expect  $-\log(y(V_0^\uparrow)) \sim \frac{1}{a}\log(n)$  and  $|\mathcal{C}^+(\Phi_n)| \approx J(\frac{1}{a}\log(n))$ . Formally, we have the following analogue of Lemma 12.

**Lemma 19.** *Let  $\delta(n) = Z_n^+ - J(\frac{1}{a}\log(n))$ . Then  $\delta(n) = \mathcal{O}_P(1)$ .*

*Proof.* The proof is along the same lines as Lemma 12. By (23), it is sufficient to show that for a function  $\ell$  tending to infinity arbitrarily slowly, with high probability,

$$(24) \quad \frac{1}{a} \log(n) - \ell(n) \leq -\log(y(V_0^\uparrow)) \leq \frac{1}{a} \log(n) + \ell(n).$$

Again, we divide the region  $(0, 1) \times (0, \infty)$  into  $n$  vertical strips  $(i/n, (i+1)/n) \times (0, \infty)$  for  $i = 0, \dots, n-1$ . In each strip, the  $y$ -coordinates of the points of  $\Phi_n$  form an inhomogeneous Poisson point process on  $(0, \infty)$  with intensity measure  $G$ . In particular, for the  $i$ -th strip, the first  $y$  jump  $\epsilon_i$  has distribution  $F$ . Define  $Y' := \min(\epsilon_1, \dots, \epsilon_n)$ . Then  $Y' \stackrel{d}{=} F^{-1}(B)$  where  $B \stackrel{d}{=} \text{Beta}(1, n)$ . Now,

$$\mathbb{P}(B \leq \frac{\ell(n)}{n}) = 1 - \left(1 - \frac{\ell(n)}{n}\right)^n \rightarrow 1$$

as  $n \rightarrow \infty$ . Conditioned on the event  $B \leq \frac{\ell(n)}{n}$ , for large  $n$ ,

$$F^{-1}(B) \leq \frac{2}{C} B^{1/a} \leq \frac{2}{C} \frac{\ell(n)^{1/a}}{n^{1/a}}.$$

Therefore, for large  $n$ ,

$$\mathbb{P}\{-\log(Y') \geq \frac{1}{a} \log(n) - \frac{1}{a} \log(\ell(n)) - \log\left(\frac{2}{C}\right)\} = 1 - \left(1 - \frac{\ell(n)}{n}\right)^n.$$

Similarly,

$$\mathbb{P}\left(B \leq \frac{1}{n\ell(n)}\right) = 1 - \left(1 - \frac{1}{n\ell(n)}\right)^n \rightarrow 0$$

as  $n \rightarrow \infty$ . By the same argument, for large  $n$ ,

$$\mathbb{P}\{-\log(Y') \leq \frac{1}{a} \log(n) + \frac{1}{a} \log(\ell(n)) + \log\left(\frac{2}{C}\right)\} = 1 - \left(1 - \frac{1}{\ell(n)n}\right)^n.$$

Therefore, (24) holds with high probability for  $Y'$  in place of  $y(V_0^\uparrow)$ . But  $Y' = y(V_0^\uparrow)$  conditioned on the event  $Y' < G^{-1}(1)$ , with  $G^{-1}(1)$  the real number such that  $G([0, G^{-1}(1)]) = 1$ , namely  $G^{-1}(1) = F^{-1}(1 - e^{-1})$ . This is precisely the event that there is at least one point of  $\Phi_n$  in the square  $(0, 1) \times (0, G^{-1}(1))$ , so this event happens with probability  $1 - e^{-n}$ . Thus, (24) happens with probability at least

$$(1 - e^{-n}) \left(1 - \left(1 - \frac{1}{\ell(n)n}\right)^n\right) \rightarrow 1.$$

□

*Proof of Proposition 2.* The argument is similar to that given in Section 4 for the case  $F$  is exponential. In summary, Proposition 18 establishes the central limit theorem for  $J(\frac{1}{a} \log(n))$ . The number of faces to the left and right of  $V_0^\uparrow$ ,  $|\mathcal{C}^+(\Phi_n)|$  and  $|\mathcal{C}^-(\Phi_n)|$ , are independent conditioned on  $y(V_0^\uparrow)$ . Lemma 19 states  $|\mathcal{C}^+(\Phi_n)|$  and  $|\mathcal{C}^-(\Phi_n)|$  are simultaneously well-approximated by two independent copies of  $J(\frac{1}{a} \log(n))$ . Thus  $N(\Phi_n)$  is distributed like their sum, and this concludes the proof. □

For  $\alpha \in (0, \infty)$ , recall that  $L(\alpha)$  is the line orthogonal to  $(1, \alpha)$  which supports  $\mathcal{C}^+(\Pi_n)$ , and that  $L_x(\alpha)$  is its  $x$ -intercept. Note that  $L_x(\cdot)$  is a pure-jump process indexed by  $\alpha$ . Let  $L_x^i, i = 1, \dots, |\mathcal{C}^+(\Phi)|$  be the sequence of values of  $L_x(\cdot)$  ordered in increasing value. Define  $L_x^0 = x(V_0^\downarrow)$ . Consider the triangles by the consecutive

lines and the  $x$ -axis as in Figure 5. For  $i = 1, 2, \dots, |\mathcal{C}^+(\Phi)|$ , let  $T_i$  be the  $i$ -th triangle, and  $area(T_i)$  denote its area with respect to the measure  $\lambda \otimes G$ .

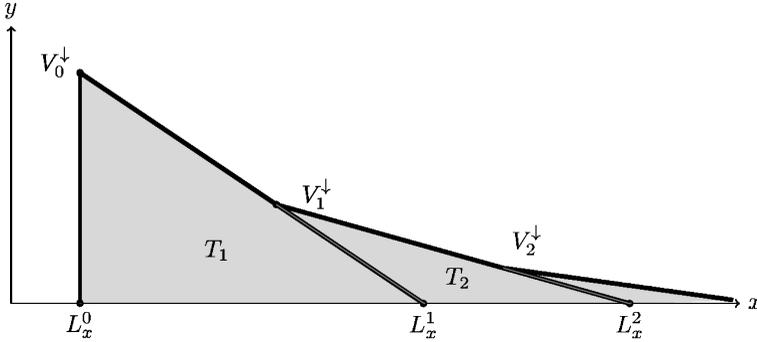


FIGURE 5. Divide up the region between  $\mathcal{C}^+(\Phi_n)$  and the  $x$ -axis into triangles based on the jumps of  $L_x(\cdot)$ .

**Corollary 20.** For  $i = 1, 2, \dots, R_n^+$ ,  $area(T_i)$  are *i.i.d.*  $Exp(n)$ , independent of the vertices  $V_i^\downarrow$  of  $\mathcal{C}^+(\Phi_n)$ .

*Proof.* Let  $S = L_x^i - L_x^{i-1}$  be the  $i$ -th increment of the process  $L_x(\cdot)$ . Conditioned on  $V_{i-1}^\downarrow$  and  $L_x^{i-1}$ , let  $T_S$  denote the triangle with vertices  $V_{i-1}^\downarrow$ ,  $(L_x^{i-1}, 0)$  and  $(L_x^{i-1} + S, 0)$ . Now,

$$area(T_S) = S \int_0^{y(V_0^\downarrow)} (1 - y/y(V_0^\downarrow)) dG([0, y]) = Sf(y(V_0^\downarrow)),$$

where  $f(y(V_0^\downarrow))$  is the above integral, which is independent of  $S$ . We have

$$\mathbb{P}(S > s) = \exp(-n (area(T_S))) = \exp(-nsf(y(V_0^\downarrow))).$$

This implies

$$\begin{aligned} \mathbb{P}(area(T_S) > t) &= \mathbb{P}\left(S > \frac{t}{f(y(V_0^\downarrow))}\right) \\ &= \exp\left(-n \frac{t}{f(y(V_0^\downarrow))} f(y(V_0^\downarrow))\right) = \exp(-nt). \end{aligned}$$

Therefore,  $area(T_S)$ , independent of  $V_{i-1}^\downarrow$ ,  $L_x^{i-1}$  and  $S$ , is distributed as an exponential random variable with rate  $n$ . □

Groeneboom [20] proved this when  $G$  is the Lebesgue measure. He used it to derive the asymptotics for the sum  $A_n = \sum_i area(T_i)$  jointly with  $|\mathcal{C}^+(\Phi_n)|$ . Since we approximate  $|\mathcal{C}^+(\Phi_n)|$  by the number of renewals in a fixed interval (cf. Lemma 19), the asymptotic normality for  $A_n$  easily follows. For general  $G$ , let us compute

the expectation and variance of  $A_n$  for large  $n$ . We have

$$\mathbb{E}(A_n) = \mathbb{E}(\mathbb{E}(A_n | J(\log(n)/a))) = \mathbb{E}(J(\log(n)/a)) = \frac{a+1}{2a+1} \log(n),$$

$$\text{Var}(A_n) = \mathbb{E}(J(\log(n)/a)) + \text{Var}(J(\log(n)/a)) = \frac{6a^3 + 8a^2 + 4a + 1}{(2a+1)^3} \log(n).$$

For  $a = 1$ , this reduces to  $\mathbb{E}(A_n) = \frac{2}{3} \log(n)$ ,  $\text{Var}(A_n) = \frac{28}{27} \log(n)$ , as shown in [12, 20]. See [32] for a historical review and summary of recent developments on asymptotics of  $A_n$  in higher dimensions.

Finally, we note that Proposition 2 is stated with the Poisson point process being restricted to the rectangle  $(0, 1) \times (0, G^{-1}(1))$ . If we widen this rectangle to  $(0, 1) \times (0, 2G^{-1}(1))$ , the probability of points in  $(0, 1) \times (G^{-1}(1), 2G^{-1}(1))$  being a vertex of the lower convex hull is clearly very small. Thus Proposition 2 also holds for the lower convex hull of points from the infinite strip  $(0, 1) \times (0, \infty)$ . We chose to state it for the rectangle to make the role of  $n$  clear: on the rectangle  $(0, 1) \times (0, G^{-1}(1))$ , we have  $Poisson(n)$  points. This makes conditioning arguments such as that in Lemma 13 a little more convenient.

## 7. PROOF OF THE MAIN THEOREM

The discrete case corresponds to a PPP on  $(0, 1) \times (0, \infty)$  with intensity measure  $\lambda_n \times G, M$  where  $\lambda_n$  is a discrete measure on  $\mathbb{R}$  which puts mass 1 at every point  $i/n$  for  $i = 0, 1, \dots$ , and 0 elsewhere. Clearly  $\frac{1}{n} \lambda_n$  converges weakly to  $\lambda$  restricted to  $[0, 1]$ , thus we expect the discrete and continuous cases to have the same asymptotics. We make this rigorous through a direct coupling.

Divide  $(0, 1) \times (0, \infty)$  into  $n$  vertical strips  $(i/n, (i+1)/n) \times (0, \infty)$ ,  $i = 0, \dots, n-1$ . Form the new point process  $\tilde{\Phi}_n$  from  $\Phi_n$  as follows: for each point  $(X_i, Y_i)$  in the  $i$ -th strip in  $\Phi_n$ , place a point  $(i/n, Y_i)$  in  $\tilde{\Phi}_n$ . This produces an a.s. bijection  $\psi : \Phi_n \rightarrow \tilde{\Phi}_n$ , such that a point  $(X_i, Y_i)$  of  $\Phi_n$  and its image  $\psi((X_i, Y_i))$  have equal  $y$ -coordinates, and differ by at most  $\frac{1}{2n}$  in their  $x$ -coordinates. We use this coupling to show the following, which implies that Theorem 1 is equivalent to Proposition 2.

### Proposition 21.

$$|\mathcal{C}^+(\Phi_n)| - |\mathcal{C}^+(\tilde{\Phi}_n)| = \mathcal{O}_P(1).$$

*Proof.* Recall that  $L(\alpha)$  is the line supporting  $\mathcal{C}^+(\Phi_n)$  with slope  $\alpha$ . Define

$$L^{1/n}(\alpha) = \{(x, y) \in \mathbb{R}^2 : \exists \epsilon \in (0, 1/n] \text{ such that } (x - \epsilon, y) \in L(\alpha)\}.$$

Let  $(\tilde{x}(\alpha), \tilde{y}(\alpha))$  be the vertex supported by the vector  $(1, \alpha)$  in  $\tilde{\Phi}_n$ . Then almost surely,  $(\tilde{x}(\alpha), \tilde{y}(\alpha)) = \psi(V)$  for some point  $V$  of  $\Phi_n$  lying in  $L^{1/n}(\alpha)$  (see Figure 6). Let us condition on the vertices of  $\mathcal{C}^+(\Phi_n)$  and the values  $L_x^i$ ,  $i = 0, 1, \dots, |\mathcal{C}^+(\Phi_n)| - 1$ . Define  $Y^* = \min\{L_y(\alpha_0), G^{-1}(1)\}$ . Extend  $\Phi_n$  to include the point  $(Y^*, 0)$ , and consider the lower convex hull, as in Figure 7. Let  $P^{1/n}$  denote the collection of points not in this convex hull, but whose  $x$ -coordinate is at most  $\frac{1}{2n}$  away from this convex hull. That is,

$$P^{1/n} = \{(x, y) \in \mathbb{R}^2 : \exists \epsilon \in (0, 1/n] \text{ such that } (x - \epsilon, y) \in \mathcal{C}^+(\Phi_n \cup (0, Y^*))\}.$$

By definition,  $\Phi_n$  has no point below  $\mathcal{C}^+(\Phi_n \cup (0, Y^*))$ . Thus  $|\mathcal{C}^+(\tilde{\Phi}_n)|$  is at most the number of points of  $\Phi_n$  in  $P^{1/n} \cup \mathcal{C}^+(\Phi_n \cup (0, Y^*))$ .

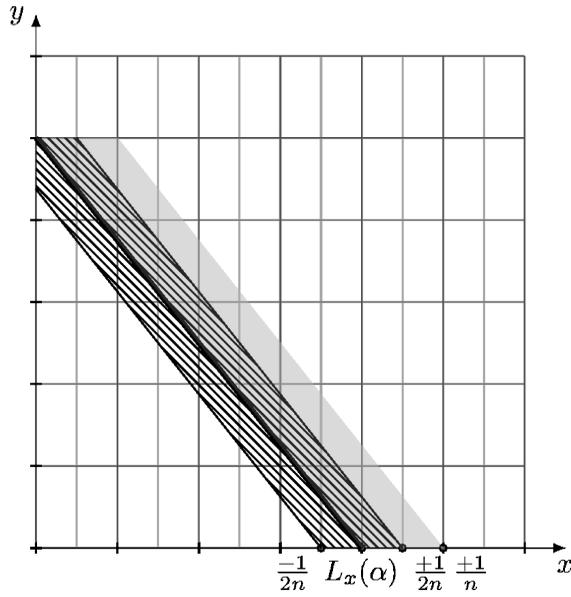


FIGURE 6. The thick line is  $L(\alpha)$ , the gray region is  $L^{1/n}(\alpha)$ . The vertex  $(\tilde{x}(\alpha), \tilde{y}(\alpha))$  of  $\tilde{\Phi}_n$  has to lie in the stripped region. Thus  $(\tilde{x}(\alpha), \tilde{y}(\alpha)) = \psi(V)$  for some point  $V$  of  $\Phi_n$  in  $L^{1/n}(\alpha)$ .

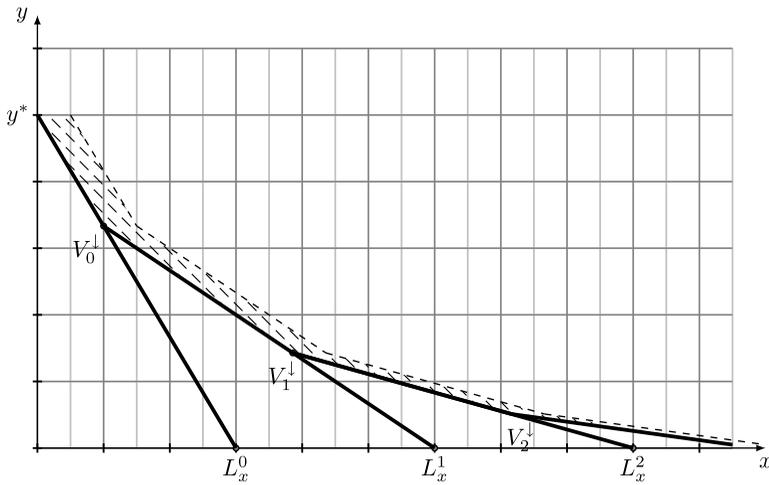


FIGURE 7. Conditioned on the vertices of  $\mathcal{C}^+(\Phi_n)$ ,  $Y^*$  and  $L_x^i$ ,  $|\mathcal{C}^+(\tilde{\Phi}_n)|$  is at most the number of points of  $\Phi_n$  lying within  $x$ -distance  $1/n$  of the lower convex hull. This is precisely the region shaded, which consists of parallelograms of width  $1/n$ .

Divide the region  $P^{1/n}$  into parallelograms of width  $1/n$ . The vertices of the  $i$ -th parallelogram are  $V_i^\downarrow, (x(V_i^\downarrow) + 1/n, y(V_i^\downarrow)), V_{i+1}^\downarrow$  and  $(x(V_{i+1}^\downarrow) + 1/n, y(V_{i+1}^\downarrow))$ . Conditioned on  $y(V_i^\downarrow) = y_i, Y^* = y^*$ , and conditioned on  $|\mathcal{C}^+(\Phi_n)| = r$ , the area of  $P^{1/n}$  under  $\lambda \otimes G$  is

$$G(y^*) - G(y_1) + \sum_{i=1}^{r-1} (G(y_i) - G(y_{i+1})) = G(y^*) - G(y_r) \leq 2$$

since  $y^*, y_r \leq G^{-1}(1)$ . By definition,  $\Phi_n$  has  $r$  vertices in  $\mathcal{C}^+(\Phi_n)$ . Therefore,

$$|\mathcal{C}^+(\tilde{\Phi}_n)| - r = |\mathcal{C}^+(\tilde{\Phi}_n)| - |\mathcal{C}^+(\Phi_n)| \leq \text{Poisson}(2).$$

Thus,  $|\mathcal{C}^+(\tilde{\Phi}_n)| - |\mathcal{C}^+(\Phi_n)| = \mathcal{O}_P(1)$  for all realizations of the point process  $\Phi_n$ .  $\square$

## 8. DISCUSSION

We considered random tropical polynomials  $\mathcal{T}f_n(x) = \min_{i=1, \dots, n} (C_i + ix)$  where the coefficients  $C_i$  are i.i.d. random variables with some c.d.f.  $F$  with support on  $(0, \infty)$ . We showed that  $Z_n$ , the number of zeros of  $\mathcal{T}f_n$  satisfies a central limit theorem under mild assumptions on the rate of decay of  $F$  near 0. Specifically, if  $F$  near 0 behaves like the  $\text{gamma}(a, 1)$  distribution for some  $a > 0$ , then  $Z_n$  has the same asymptotics as the number of points on the interval  $[0, \log(n)/a]$  of a renewal process with inter-arrival distribution  $-\log(\text{Beta}(2, a))$ . The proof techniques draw on connections between random partitions, renewal theory and random polytopes constructed from Poisson point processes. They lead to simpler proofs of the central limit theorem for the number of vertices of the convex hull of  $n$  uniform random points in a square.

The assumption that the support of  $F$  is  $[0, \infty)$  can easily be extended to the case with support on  $(c, \infty)$  for some constant  $c$ , provided the behavior of  $F$  near  $c$  is as above. This follows from the fact that the number of vertices of a polytope is invariant under translation and scaling by constants. It is crucial, however, that  $F$  be a continuous distribution. In particular, Theorem 1 does not hold for discrete distributions. Indeed, if  $F(0) = p > 0$ ,  $Z_n$  is at most the sum of two independent  $\text{Geometric}(p)$  random variables for all  $n$ , and certainly does not have a normal scaling.

This work is a first stab at *stochastic tropical geometry*, the study of linear functionals and intersections of random tropical varieties. These are common zeros of a collection of random tropical polynomials. In fields with valuations, they are precisely the tropicalization of random algebraic varieties. By considering these varieties at random, we gain insights into the global structure of tropical varieties and their preimages as a collection of sets. Unlike classical varieties, the tropical analogues are polyhedral in nature. Random tropical varieties are strongly connected with random polytopes, a rich branch of stochastic geometry [32–34]. This is a key ingredient in our proof of Theorem 1.

Our next steps will focus on random tropical polynomials in several variables and a system of random tropical polynomials. A tropical polynomial in  $m$  variable is a map  $\mathcal{T}f : \mathbb{R}^m \rightarrow \mathbb{R}$ , given by

$$\mathcal{T}f(x) = \min_{i \in I} (C_i + i \cdot x),$$

where  $C_i \in \mathbb{R}$ ,  $I \subset \mathbb{Z}^m$  is some indexing set, and  $\cdot$  is the usual inner product in  $\mathbb{R}^m$ . The convex hull of  $I$  is called the Newton polytope, and its subdivision by

$\mathcal{C}(i, C_i)$  is a regular subdivision, or in other words, weighted Delaunay triangulations [27, §2.3]. Thus, random tropical polynomials generate a type of random partition of subsets of  $\mathbb{Z}^m$ . It would be very interesting to understand this lattice partition. For example, if we consider a random tropical polynomial of degree  $n$  in  $m$  variables with i.i.d. coefficients  $C_i$ , as  $n \rightarrow \infty$ , is there a scaling limit for the number of cells of such partitions?

## REFERENCES

- [1] Josh Abramson and Jim Pitman, *Concave majorants of random walks and related Poisson processes*, *Combin. Probab. Comput.* **20** (2011), no. 5, 651–682, DOI 10.1017/S0963548311000307. MR2825583 (2012i:60106)
- [2] Josh Abramson, Jim Pitman, Nathan Ross, and Gerónimo Uribe Bravo, *Convex minorants of random walks and Lévy processes*, *Electron. Commun. Probab.* **16** (2011), 423–434, DOI 10.1214/ECP.v16-1648. MR2831081 (2012g:60147)
- [3] Marianne Akian, Ravindra Bapat, and Stéphane Gaubert, *Asymptotics of the Perron eigenvalue and eigenvector using max-algebra* (English, with English and French summaries), *C. R. Acad. Sci. Paris Sér. I Math.* **327** (1998), no. 11, 927–932, DOI 10.1016/S0764-4442(99)80137-2. MR1659185 (99k:15012)
- [4] M. Akian, R. Bapat, and S. Gaubert, *Min-plus methods in eigenvalue perturbation theory and generalised Lidskiĭ-Vishik-Ljusternik theorem*, arXiv preprint arXiv:math/0402090v3, 2004.
- [5] Erik Sparre Andersen, *On the fluctuations of sums of random variables. II*, *Math. Scand.* **2** (1954), 195–223. MR0068154 (16,839e)
- [6] F. J. Anscombe, *Large-sample theory of sequential estimation*, *Biometrika* **36** (1949), 455–458. MR0034006 (11,529d)
- [7] R. Arratia and L. Goldstein, *Size bias, sampling, the waiting time paradox, and infinite divisibility: when is the increment independent?*, arXiv preprint arXiv:1007.3910, 2010.
- [8] François Louis Baccelli, Guy Cohen, Geert Jan Olsder, and Jean-Pierre Quadrat, *Synchronization and linearity*, *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*, John Wiley & Sons, Ltd., Chichester, 1992. An algebra for discrete event systems. MR1204266 (94b:93001)
- [9] Jean Bertoin, *The convex minorant of the Cauchy process*, *Electron. Comm. Probab.* **5** (2000), 51–55 (electronic), DOI 10.1214/ECP.v5-1017. MR1747095 (2001i:60081)
- [10] Tamara Broderick, Michael I. Jordan, and Jim Pitman, *Beta processes, stick-breaking and power laws*, *Bayesian Anal.* **7** (2012), no. 2, 439–475, DOI 10.1214/12-BA715. MR2934958
- [11] Christian Buchta, *On the distribution of the number of vertices of a random polygon*, *Anz. Österreich. Akad. Wiss. Math.-Natur. Kl.* **139** (2003), 17–19 (2004). MR2135177 (2005k:52014)
- [12] Christian Buchta, *Exact formulae for variances of functionals of convex hulls*, *Adv. in Appl. Probab.* **45** (2013), no. 4, 917–924, DOI 10.1239/aap/1386857850. MR3161289
- [13] Rick Durrett, *Probability: theory and examples*, 4th ed., *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge, 2010. MR2722836 (2011e:60001)
- [14] L. Elsner and P. van den Driessche, *Max-algebra and pairwise comparison matrices*, *Linear Algebra Appl.* **385** (2004), 47–62, DOI 10.1016/S0024-3795(03)00476-2. MR2063346 (2005b:15015)
- [15] Steven N. Evans, *The expected number of zeros of a random system of  $p$ -adic polynomials*, *Electron. Comm. Probab.* **11** (2006), 278–290 (electronic), DOI 10.1214/ECP.v11-1230. MR2266718 (2007m:60013)
- [16] Alexander V. Gnedin, *The Bernoulli sieve*, *Bernoulli* **10** (2004), no. 1, 79–96, DOI 10.3150/bj/1077544604. MR2044594 (2004k:60018)
- [17] Alexander V. Gnedin, Alexander M. Iksanov, Pavlo Negadajlov, and Uwe Rösler, *The Bernoulli sieve revisited*, *Ann. Appl. Probab.* **19** (2009), no. 4, 1634–1655, DOI 10.1214/08-AAP592. MR2538083 (2011e:60042)
- [18] Piet Groeneboom, *The concave majorant of Brownian motion*, *Ann. Probab.* **11** (1983), no. 4, 1016–1027. MR714964 (85h:60119)

- [19] Piet Groeneboom, *Limit theorems for convex hulls*, Probab. Theory Related Fields **79** (1988), no. 3, 327–368, DOI 10.1007/BF00342231. MR959514 (89j:60024)
- [20] Piet Groeneboom, *Convex hulls of uniform samples from a convex polygon*, Adv. in Appl. Probab. **44** (2012), no. 2, 330–342, DOI 10.1239/aap/1339878714. MR2977398
- [21] Allan Gut, *Stopped random walks*, 2nd ed., Springer Series in Operations Research and Financial Engineering, Springer, New York, 2009. Limit theorems and applications. MR2489436 (2010f:60131)
- [22] M. Kac, *On the average number of real roots of a random algebraic equation*, Bull. Amer. Math. Soc. **49** (1943), 314–320. MR0007812 (4,196d)
- [23] M. Kac, *On the average number of real roots of a random algebraic equation. II*, Proc. London Math. Soc. (2) **50** (1949), 390–408. MR0030713 (11,40e)
- [24] D. A. Korshunov, *Limit theorems for general Markov chains* (Russian, with Russian summary), Sibirsk. Mat. Zh. **42** (2001), no. 2, 354–371, ii, DOI 10.1023/A:1004841130674; English transl., Siberian Math. J. **42** (2001), no. 2, 301–316. MR1833162 (2002c:60113)
- [25] Dmitry Korshunov, *The key renewal theorem for a transient Markov chain*, J. Theoret. Probab. **21** (2008), no. 1, 234–245, DOI 10.1007/s10959-007-0132-8. MR2384480 (2009a:60097)
- [26] J. E. Littlewood, *On the Number of Real Roots of a Random Algebraic Equation*, J. London Math. Soc. **S1-13**, no. 4, 288, DOI 10.1112/jlms/s1-13.4.288. MR1574980
- [27] D. Maclagan and B. Sturmfels, *Introduction to tropical geometry*, Book in preparation, 2014.
- [28] Jim Pitman, *Exchangeable and partially exchangeable random partitions*, Probab. Theory Related Fields **102** (1995), no. 2, 145–158, DOI 10.1007/BF01213386. MR1337249 (96e:60059)
- [29] J. Pitman, *Combinatorial stochastic processes*, Lecture Notes in Mathematics, vol. 1875, Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002; With a foreword by Jean Picard. MR2245368 (2008c:60001)
- [30] Jim Pitman and Gerónimo Uribe Bravo, *The convex minorant of a Lévy process*, Ann. Probab. **40** (2012), no. 4, 1636–1674, DOI 10.1214/11-AOP658. MR2978134
- [31] J. Pitman and N. M. Tran, *Size-biased permutation of a finite sequence with independent and identically distributed terms*, arXiv preprint arXiv:1206.2081v1, 2012.
- [32] Rolf Schneider, *Recent results on random polytopes*, Boll. Unione Mat. Ital. (9) **1** (2008), no. 1, 17–39. MR2387995 (2008m:52011)
- [33] Rolf Schneider and Wolfgang Weil, *Stochastic and integral geometry*, Probability and its Applications (New York), Springer-Verlag, Berlin, 2008. MR2455326 (2010g:60002)
- [34] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 1987. With a foreword by D. G. Kendall. MR895588 (88j:60034a)
- [35] T. Tao and V. Vu, *Local universality of zeroes of random polynomials*, arXiv preprint arXiv:1307.4357v2, 2013.
- [36] Ngoc Mai Tran, *Pairwise ranking: choice of method can produce arbitrarily different rank order*, Linear Algebra Appl. **438** (2013), no. 3, 1012–1024, DOI 10.1016/j.laa.2012.08.028. MR2997792
- [37] M. van Manen and D. Siersma, *Power diagrams and their applications*, arXiv preprint arXiv:math/0508037v2, 2005.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TEXAS, AUSTIN, TEXAS 78712

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TEXAS, AUSTIN, TEXAS 78712