Boundaries of the generalized Pascal triangles and larger graded graphs

Alexander Gnedin

Utrecht University

The general combinatorial framework

Weighted Bratteli diagram is an infinite directed graph with a vertex set $T = \cup_{n \geq 0} T_n$ such that

- vertices are organized in *finite* levels T_n, with a sole root vertex $\varnothing$ comprising T_0,
- each edge connects a vertex on level n with a vertex on level $n+1$, for $n = 0, 1, \ldots$,
- each vertex has at least one follower and at least one predecessor (except $\varnothing$)
- each edge is endowed with a positive weight (or one considers multiple edges if the weights are integer).

The boundary problem

For a *standard* (directed) finite path $(s_0, s_1, s_2, \dots, s_n) \in T_0 \times \cdots \times T_n$, the *weight* is the product of weights of edges $s_j \to s_{j+1}$ along the path.

Let $\mathcal{P}$ be the set of probability laws for transient Markov chains $S = (S_n, n \geq 0)$ on T with the properties:

- S_n assumes values in T_n, in particular $S_0 = \varnothing$,
- Gibbs condition (Vershik-Kerov's centrality) holds: conditionally given $S_n = s_n$ for some $s_n \in T_n$, each path $S_0 = s_0, \dots, S_{n-1} = s_{n-1}, S_n = s_n$ has probability proportional to its weight.

The set $\mathcal{P}$ is an infinite Choquet simplex, i.e. a compact convex set with the property of uniqueness of barycentric representation of a generic element as a mixture of extremes.

The *boundary problem* for T asks one to describe (as explicitly as possible) the set of extreme elements $\mathrm{ext}\mathcal{P}$ of $\mathcal{P}$.

Each $P \in \mathrm{ext}\mathcal{P}$ corresponds to an ergodic Markov chain (S_n), with trivial tail σ-algebra.

The Gibbs condition fixes transition probabilities for the time-reversed Markov chain $(\ldots, S_1, S_0)$.

Each $P \in \mathcal{P}$ is uniquely determined by a nonnegative function ϕ on T which is normalized by the condition $\phi(\varnothing) = 0$ and satisfies a backward recursion

$$\phi(s) = \sum_{s' \in T_{n+1}:\ s \to s'} \mathrm{weight}(s \to s')\phi(s'), \qquad s \in T_n, \quad n \geq 0.$$

The relation between P and ϕ is

$$P(S_0 = s_0, \ldots, S_n = s_n) = (\mathrm{weight\ of\ } s_0, \ldots, s_n) \times \phi(s_n).$$

Some roots from 60's-70's

- boundary theory for Polya's urn models: Blackwell, Kendall
- AF-algebras, dimension groups: Bratteli, Elliot, Effros et al
- characters of the infinite symmetric group $\mathfrak{S}_\infty$: Vershik and Kerov
- exchangeability: Diaconis and Freedman, Kingman

Examples of ‘big’ boundaries

1. The Young lattice $\mathbb{Y}$, with $\mathbb{Y}_n$ being the set of partitions of integer n, and simple edges (appending a box to the Young diagram). The extreme boundary is parameterizable by

$$\{(\alpha_j),(\beta_j) : \alpha_1 \geq \alpha_2 \cdots \geq 0,\ \beta_1 \geq \beta_2 \cdots \geq 0, \ \ \sum_j (\alpha_j + \beta_j) \leq 1\}.$$

   (Vershik and Kerov)

2. A composition poset: vertices on level n are compositions of integer n (like $(3,5,3,2)$). All edges are simple, except when a series of 1's is increased (e.g. $(2,1,1,2) \to (2,1,1,1,2)$). Each element of the extreme boundary can be identified with some closed subset of $[0,1]$.

Further relatives of $\mathbb{Y}$ with rich boundaries were studied by Vershik, Kerov, Okounkov, Olshanski, Kingman, G, Pitman, ...

Generalized Pascal triangles

The generalized Pascal graph T has weights

- $\ell_{n,k}$ for $(n,k) \to (n+1,k)$ (transition '0'),
- $r_{n,k}$ for $(n,k) \to (n+1,k+1)$ (transition '1')

A path is encoded into 0-1 sequence.

Let $D = (D_{n,k}, (n,k) \in T)$ denote the *dimension function* (partition function, statistical sum...), So $D_{n,k}$ = sum of weights of paths from $(0,0)$ to (n,k), and satisfies the forward recursion

$$D_{n,k} = r_{n-1,k-1} D_{n-1,k-1} + \ell_{n-1,k} D_{n-1,k}, \qquad 0 \le k \le n. \tag{1}$$

'Named' triangles (Pascal, Stirling, Euler, Lah, etc) derive their names from the numbers $D_{n,k}$.

$\mathcal{P}$ can be identified with the convex set of nonnegative solutions to the backward recursion

$$\phi_{n,k} = \ell_{n,k}\phi_{n+1,k} + r_{n,k}\phi_{n+1,k+1}, \qquad 0 \le k \le n;\ \phi(0,0) = 1. \qquad (2)$$

Note: each solution to (2) can be recovered from either of sequences $(\phi_{n,n},\ n \ge 0)$ or $(\phi_{n,0},\ n \ge 0)$ by iterated (weighted) differencing, thus the boundary problem reduces to finding all such sequences with nonnegative differences of any order.

For the related random walk S we have

$$P_\phi(S_0 = s_0, \dots, S_n = (n,k)) = (\text{weight of path}) \ \times \phi_{n,k},$$

so the marginal distributions of S are

$$P_\phi(S_n = (n,k)) = D_{n,k}\phi_{n,k}.$$

The backward transition probabilities do not depend on ϕ

$$P_\phi(S_{n-1} = (n-1,j) | S_n = (n,k)) = \frac{D_{n-1,j}}{D_{nk}}(\ell_{n-1,j}\delta_{jk} + r_{n-1,j}\delta_{j,k-1}).$$

Martin kernel

The Martin kernel is the ratio

$$\frac{D_{n,k}^{\nu,\varkappa}}{D_{\nu,\varkappa}} \qquad 0 \le \varkappa \le \nu,\ 0 \le n \le \nu, \tag{3}$$

where the extended dimension $D_{n,k}^{\nu,\varkappa}$ = sum of weights of paths from (n,k) to $(\nu,\varkappa)$ (so $D_{n,k} = D_{0,0}^{n,k}$).

- The *Martin boundary* $\partial_M \mathcal{P}$ is the set of weak limits of elementary measures corresponding to functions

$$\phi_{n,k}^{\nu,\varkappa} := \frac{D_{n,k}^{\nu,\varkappa}}{D_{\nu,\varkappa}} \tag{4}$$

  where $\nu \to \infty$ and $\varkappa = \varkappa(\nu)$.
- The *sequential boundary* $\partial_\downarrow \mathcal{P}$ if defined by taking limits in (4) *along infinite paths*.

We have $\mathrm{ext}\mathcal{P} \subset \partial_\downarrow \mathcal{P} \subset \partial_M \mathcal{P}$, but the inclusions may be strict (though in pathological examples).

The Pascal graph: de Finetti's theorem

For the Pascal graph $\ell_{n,k} = r_{n,k} = 1$.
cn The ratios

$$\phi_{n,k}^{\nu,\varkappa} = \frac{\binom{\nu-n}{\varkappa-k}}{\binom{\nu}{n}}$$

converge iff for $\varkappa = \varkappa(\nu)$ there exists

$$p := \lim_{\nu\to\infty} \frac{\varkappa}{\nu}$$

the asymptotic frequency of 1's. Since the graph describes branching of orbits of $\mathfrak{S}_\infty = \cup_n \mathfrak{S}_n$ acting on $\{0,1\}^\infty$, de Finetti's theorem follows: every extreme exchangeable 0-1 sequence is a Bernoulli(p) sequence.

Differencing $(\phi_{n,n}, n \geq 0)$ shows that the sequence must be completely monotone $\Rightarrow$ de Finetti's theorem is equivalent to 'Hausdorff's moments problem' on $[0,1]$.

Boundaries of Pascal pyramids (action of permutations on $\{1,\dots,d\}^\infty$) and de Finetti's theorem for $\mathbb{R}^\infty$ derive from this basic result.

q-exchangeability, the q-Pascal graph

Kerov '85, G and Olshanski '09

Two problems:

1. for $q > 0$, describe quasi-invariant under $\mathfrak{S}_\infty$ measures on $\{0,1\}^\infty$, with cocycle

$$q^{-c(\sigma,w)}, \quad \sigma \in \mathfrak{S}_\infty,\ w = w_1 w_2 \cdots \in \{0,1\}^\infty$$

   $c(w,\sigma) :=$
   $\lim_n \#(\text{inversions in } w_1 \dots w_n) - \#(\text{inversions in } (\sigma w)_1 \dots (\sigma w)_n)$,

2. for q a power of prime number, describe (all distributions for) random subspaces of $\mathbb{F}_q^\infty$ invariant under $GL(\infty, \mathbb{F}_q)$,

are reduced to the boundary problem for the q-Pascal triangle, which has the weights

$$\ell_{nk} = 1, r_{nk} = q^{n-k}, \qquad (n,k) \in T.$$

For the q-Pascal graph the Martin kernel is

$$\binom{\nu - n}{\varkappa - k}_q \Big/ \binom{\nu}{n}_q,$$

where the q-binomial coefficients are defined via q-integers $[n]_q := (1 - q^n)/(1 - q)$. For $q < 1$ this converges only if $\varkappa(\nu) = m$ for large ν or $\varkappa \to \infty$, so the boundary $\mathrm{ext}\mathcal{P} = \partial_M \mathcal{P}$ is discrete. Explicitly, the limits are

$$\phi_{nk}(m) = \frac{q^{(m-k)(n-k)}(1-q)\cdots(1-q^m)}{(1-q)\cdots(1-q^{m-k})}, \qquad k \le m$$

A q-analogue of Bernoulli coin-tossing is the q-shuffle: a word $1\ldots1000\ldots$ (m 1's) is re-arranged by iterated choices of a symbol in the ξth position, where ξ has geomeric distribution.

Similar q-analogues of de Finetti's theorem hold for sequences on larger ordered alphabets (G and Olshanski '09).

Parameterization of the boundary

G and Pitman observed:

The sequential boundary $\partial_{\downarrow}\mathcal{P}$ is homeomorphic to a closed subset of $[0,1]$ by the virtue of function

$$P \mapsto P(S_1 = (1,1)) = \phi_{1,1}\ell_{0,1}$$

(probability of the first '1').

Proof: S_n under distribution $\phi^{\nu,\varkappa}$ is strictly stochastically larger than S_n under distribution $\phi^{\nu,\varkappa'}$ provided that $\varkappa' > \varkappa$.

This is helpful to identify the topology of the boundary.

The case of discrete boundary

Theorem

Suppose for $m = 0, 1, \ldots$ there are distributions $P_m \in \mathcal{P}$ such that $P_m(S_n = (n, m)) \to 1$ as $n \to \infty$, then each P_m is extreme (and satisfies $S_n = (n, m)$ for large n a.s.).

If above that $P_m(S_1 = (1, 1)) \to 1$ as $m \to \infty$ then P_m converge to the trivial distribution P_∞ with $S_n = (n, n)$ P_∞-a.s., and in this case

$$\mathrm{ext}\mathcal{P} = \partial_M \mathcal{P} = \{P_0, P_1, \ldots, P_\infty\}.$$

The case of continuous boundary

Theorem

Suppose there is a sequence of positive constants $\{c(n); n = 0, 1, \ldots\}$ *with* $c(n) \to \infty$, *and for each* $t \in [0, \infty]$ *there is a distribution* $P_t \in \mathcal{P}$ *which satisfies*

$$S_n \sim c(n)t \quad P_t - \text{a.s.}.$$

Suppose the mapping $t \mapsto P_t$ *is a continuous injection from* $[0, \infty]$ *to* $\mathcal{P}$, *with* 0 *and* ∞ *corresponding to the trivial Markov chains (*$S_n \equiv 0$ *respectively* $S_n \equiv n$ a.s.*). Then a path* $\{\varkappa(\nu); \nu = 0, 1, \ldots\}$ *induces a limit if and only if* $\varkappa(\nu)/c(\nu) \to t$ *for some* $t \in [0, \infty]$, *in which case the limit is* P_t. *Moreover,*

$$\text{ext}\mathcal{P} = \partial_M \mathcal{P} = \{P_t, t \in [0, \infty]\}.$$

Stirling triangles

A parametric class of triangles (G and Pitman '05), a subclass of Stirling triangles introduced by Kerov.

The generalized Stirling triangle has $r_{nk} = 1$ and $\ell_{nk} = (n+1) - \alpha(k+1)$ for $-\infty < \alpha < 1$.
For $\alpha = -\infty$ take $\ell_{nk} = k + 1$.
The generalised Stirling numbers $D_{nk} = \begin{bmatrix} n+1 \\ k+1 \end{bmatrix}_\alpha$
are the connection coefficients in

$$(z)_n = \sum_{k=1}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_\alpha \alpha^n (z/\alpha)_n,$$

(where $z_n = z(z+1)\dots(z+n-1)$). These are the Stirling numbers of the second kind for $\alpha = -\infty$, and signless Stirling numbers of the first kind $\alpha = 0$.

The interest to Stirling triangles is motivated by applications to

- Kingman's partition structures
- Bayesian nonparametric inference, species sampling
- urn models
- excursions of Brownian motion and recurrent Bessel processes
- random permutations and Ewens' sampling formula of population genetics

A phase transition for Stirling triangles

- For $-\infty \le \alpha < 0$ the extreme boundary is discrete. For $m = 1, 2, \dots$

$$\phi_{n,0}(m) = \frac{1}{(m|\alpha| + 1)_n} \quad \text{for} \quad -\infty < \alpha < 0,$$

and

$$\phi_{n,0}(m) = \frac{1}{m^n} \quad \text{for} \quad \alpha = -\infty.$$

- For $\alpha = 0$ the extreme boundary becomes continuous,

$$\phi_{n,0}(t) = \frac{1}{(t+1)_n} \qquad t \in [0, \infty].$$

- For $0 < \alpha < 1$ the boundary is also continuous, and

$$\phi_{n,0}(t) = \frac{t\alpha}{(1-\alpha)_n \Gamma(1-\alpha) g_\alpha(t)} \int_0^1 y^{n-\alpha}(1-y)^{-1-\alpha} g_\alpha(t(1-y)^{-\alpha}) \mathrm{d}y$$

where g_α is related to the stable density (with Laplace transform $e^{-\lambda^\alpha}$) via

$$f_\alpha(y) = \alpha y^{-1-\alpha} g_\alpha(y^{-\alpha}).$$

S_n has the same law as the number of blocks in a partition of $\{1, \ldots, n\}$ induced by a Bessel process conditioned on the value of local time t, and $S_n \sim t n^\alpha$ a.s.

The Eulerian triangle

$\ell_{nk} = k+1,\ r_{nk} = n-k+1$ the dimension is the Eulerian number $D_{n,k} = \left\langle {n+1 \atop k} \right\rangle$ (which enumerates permutations with a given number of descents).
The extreme elements of $\mathcal{P}$ are given by

$$\phi_{n,k}(m) = \frac{1}{(n+1)!} \prod_{i=-k}^{n-k} \left(1 + \frac{i}{m}\right)$$

with $m \in \mathbb{Z} \cup \{\infty\}$.
(G and Olshanski '06)

As P runs over $\mathcal{P}$, the collection of probabilities $P(S_1 = (1,1))$ is symmetric about $1/2$, with $1/2$ being the only accumulation point.
This $1/2$-case corresponds to the uniform distributions on $\mathfrak{S}_n$'s, and to the unique fully suported ergodic measure on the graph.