A generalization of the Erdős-Turán law for the order of random permutation

Alexander Iksanov, Kiev, Ukraine

joint work with Alexander Gnedin (Utrecht, Netherlands) and Alexander Marynych (Kiev, Ukraine)

AofA'11, Będlewo, Poland, June 12-18, 2011

Overview

- basic notions
    - cycles of permutations
    - order of permutation
    - Ewens' permutations
    - classical Erdős-Turán law

Overview

- basic notions
  - cycles of permutations
  - order of permutation
  - Ewens' permutations
  - classical Erdős-Turán law
- permutations derived from stick-breaking
  - basic construction
  - a Markov chain representation
  - the infinite occupancy scheme

Overview

- basic notions
  - cycles of permutations
  - order of permutation
  - Ewens' permutations
  - classical Erdős-Turán law
- permutations derived from stick-breaking
  - basic construction
  - a Markov chain representation
  - the infinite occupancy scheme
- extension of the Erdős-Turán law
  - main result
  - scheme of the proof
  - weak convergence of the product of cycle lengths
  - asymptotics of the cycle partition

Basic notions:
Cycles of permutations

Let S_n be the set of permutations of $[n] := \{1, \ldots, n\}$.

Basic notions:
Cycles of permutations

Let S_n be the set of permutations of $[n] := \{1, \dots, n\}$.

$$\pi \in S_n \quad \longrightarrow \quad \pi = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ \pi_1 & \pi_2 & \pi_3 & \dots & \pi_n \end{pmatrix}$$

Basic notions:
Cycles of permutations

Let S_n be the set of permutations of $[n] := \{1, \ldots, n\}$.

$$\pi \in S_n \quad \to \quad \pi = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ \pi_1 & \pi_2 & \pi_3 & \ldots & \pi_n \end{pmatrix}$$

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 9 & 1 & 7 & 4 & 3 & 2 & 5 & 8 & 10 & 6 \end{pmatrix}$$

$$\downarrow$$

$\pi = (1\,9\,10\,6\,2)(3\,7\,5)(4)(8)$ – standard cycle notation

Basic notions:
Cycles of permutations

Let S_n be the set of permutations of $[n] := \{1, \ldots, n\}$.

$$\pi \in S_n \quad \rightarrow \quad \pi = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ \pi_1 & \pi_2 & \pi_3 & \ldots & \pi_n \end{pmatrix}$$

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 9 & 1 & 7 & 4 & 3 & 2 & 5 & 8 & 10 & 6 \end{pmatrix}$$

$$\downarrow$$

$\pi = (1\,9\,10\,6\,2)(3\,7\,5)(4)(8)$ – standard cycle notation

$(2, 0, 1, 0, 1, 0, 0, 0, 0, 0)$ cycle partition

The rth coordinate equals the number of cycles of length r.

Basic notions:
cycles of permutations

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 9 & 1 & 7 & 4 & 3 & 2 & 5 & 8 & 10 & 6 \end{pmatrix}$$

$$\downarrow$$

$\pi = (1\,9\,10\,6\,2)(3\,7\,5)(4)(8)$ – standard cycle notation

$(2, 0, 1, 0, 1, 0, 0, 0, 0, 0)$ cycle partition

The rth coordinate equals the number of cycles of length r.

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 9 & 1 & 7 & 4 & 3 & 2 & 5 & 8 & 10 & 6 \end{pmatrix}$$

$$\downarrow$$

$\pi = (7\,5\,3)(4)(6\,2\,1\,9\,10)(8)$ "non-standard" cycle notation

Basic notions:
Order of permutation

Equipped with the juxtaposition $\circ$ of permutations as the group operation S_n becomes the symmetric group on $[n]$.

Basic notions:
Order of permutation

Equipped with the juxtaposition $\circ$ of permutations as the group operation S_n becomes the symmetric group on $[n]$.

The order $O(\pi)$ of permutation $\pi \in S_n$ is the smallest $k \in \mathbb{N}$ such that $\underbrace{\pi \circ \ldots \circ \pi}_{k \text{ times}} = \text{Id}$.

Basic notions:
Order of permutation

Equipped with the juxtaposition $\circ$ of permutations as the group operation S_n becomes the symmetric group on $[n]$.

The order $O(\pi)$ of permutation $\pi \in S_n$ is the smallest $k \in \mathbb{N}$ such that $\underbrace{\pi \circ \ldots \circ \pi}_{k \text{ times}} = \text{Id}$.

Simpler, $O(\pi)$ equals the least common multiple of the cycle lengths of π.

Basic notions:
Order of permutation

Equipped with the juxtaposition $\circ$ of permutations as the group operation S_n becomes the symmetric group on $[n]$.

The order $O(\pi)$ of permutation $\pi \in S_n$ is the smallest $k \in \mathbb{N}$ such that $\underbrace{\pi \circ \ldots \circ \pi}_{k \text{ times}} = \mathrm{Id}$.

Simpler, $O(\pi)$ equals the least common multiple of the cycle lengths of π.

$$\pi = (1\,9\,6\,2)(3\,7\,5)(4\,8) \quad \longrightarrow \quad O(\pi) = 12$$

Basic notions:
Ewens' permutations

A random permutation Π_n of $[n]$ is a random object with values in S_n.

Basic notions:
Ewens' permutations

A random permutation Π_n of $[n]$ is a random object with values in S_n.

A random permutation Π_n is called θ-biased random permutation or Ewens' permutation if

$$\mathbb{P}\{\Pi_n = \pi\} = \theta^{|\pi|}/(\theta)_n, \quad \theta > 0, \ \pi \in S_n,$$

where $|\pi|$ denotes the number of cycles of π, and $(\theta)_n := \Gamma(\theta + n)/\Gamma(\theta)$.

Basic notions:
Ewens' permutations

A random permutation Π_n of $[n]$ is a random object with values in S_n.

A random permutation Π_n is called θ-biased random permutation or Ewens' permutation if

$$\mathbb{P}\{\Pi_n = \pi\} = \theta^{|\pi|}/(\theta)_n, \quad \theta > 0, \quad \pi \in S_n,$$

where $|\pi|$ denotes the number of cycles of π, and
$(\theta)_n := \Gamma(\theta + n)/\Gamma(\theta)$.

The Ewens sampling formula:

$$\mathbb{P}\{(K_{n,1}, \ldots, K_{n,n}) = (k_1, \ldots, k_n)\}$$
$$= \quad 1_{\{\sum_{j=1}^n jk_j = n\}} \frac{n!}{\theta(n)} \prod_{j=1}^{n} \left(\frac{\theta}{j}\right)^{k_j} \frac{1}{k_j!}, \quad k_j = 0, 1, \ldots, n,$$

where $K_{n,r}$ denotes the number of cycles of length r.

Basic notions:
Ewens' permutations

A random permutation Π_n of $[n]$ is a random object with values in S_n.

A random permutation Π_n is called θ-biased random permutation or Ewens' permutation if

$$\mathbb{P}\{\Pi_n = \pi\} = \theta^{|\pi|}/(\theta)_n, \qquad \theta > 0, \quad \pi \in S_n,$$

where $|\pi|$ denotes the number of cycles of π, and $(\theta)_n := \Gamma(\theta + n)/\Gamma(\theta)$.

The Ewens sampling formula:

$$\begin{aligned} &\mathbb{P}\{(K_{n,1}, \ldots, K_{n,n}) = (k_1, \ldots, k_n)\} \\ = \quad & 1_{\{\sum_{j=1}^n jk_j = n\}} \frac{n!}{\theta(n)} \prod_{j=1}^n \left(\frac{\theta}{j}\right)^{k_j} \frac{1}{k_j!}, \quad k_j = 0, 1, \ldots, n, \end{aligned}$$

where $K_{n,r}$ denotes the number of cycles of length r.

The instance $\theta = 1$ corresponds to the uniform random permutation (arises in the Hat-Check-Problem).

Basic notions:
Classical Erdős- Turán law

Let O_n denote the order of random permutation Π_n, i.e., the least common multiple of the different cycle lengths of Π_n.

- Erdős & Turán (1967): for the uniform random permutation

$$\frac{\log O_n - (1/2)\log^2 n}{\sqrt{(1/3)\log^3 n}} \overset{d}{\to} \mathcal{N}(0,1), \qquad n \to \infty.$$

Nowadays this relation is referred to as the Erdős - Turán law.

Basic notions:
Classical Erdős- Turán law

Let O_n denote the order of random permutation Π_n, i.e., the least common multiple of the different cycle lengths of Π_n.

- Erdős & Turán (1967): for the uniform random permutation

$$\frac{\log O_n - (1/2)\log^2 n}{\sqrt{(1/3)\log^3 n}} \overset{d}{\to} \mathcal{N}(0,1), \qquad n \to \infty.$$

  Nowadays this relation is referred to as the Erdős - Turán law.

- Arratia & Tavaré (1992): for the Ewens' permutation

$$\frac{\log O_n - (\theta/2)\log^2 n}{\sqrt{(\theta/3)\log^3 n}} \overset{d}{\to} \mathcal{N}(0,1), \qquad n \to \infty.$$

Basic notions:
Classical Erdős- Turán law

Let O_n denote the order of random permutation Π_n, i.e., the least common multiple of the different cycle lengths of Π_n.

- Erdős & Turán (1967): for the uniform random permutation

$$\frac{\log O_n - (1/2)\log^2 n}{\sqrt{(1/3)\log^3 n}} \xrightarrow{d} \mathcal{N}(0,1), \qquad n \to \infty.$$

  Nowadays this relation is referred to as the Erdős - Turán law.

- Arratia & Tavaré (1992): for the Ewens' permutation

$$\frac{\log O_n - (\theta/2)\log^2 n}{\sqrt{(\theta/3)\log^3 n}} \xrightarrow{d} \mathcal{N}(0,1), \qquad n \to \infty.$$

  The proof uses Feller coupling and asymptotic independence of cycle lengths.

Permutations derived from stick-breaking:
Basic construction

Let W be a random variable, called stick-breaking factor, with values in $(0, 1)$.

Permutations derived from stick-breaking:
Basic construction

Let W be a random variable, called stick-breaking factor, with values in $(0,1)$.

Let Q be a multiplicative random walk (also known as stick-breaking process) with points

$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent replicas of W.

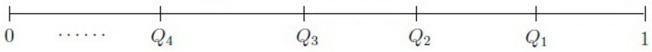
Permutations derived from stick-breaking:
Basic construction

Let W be a random variable, called stick-breaking factor, with values in $(0, 1)$.

Let Q be a multiplicative random walk (also known as stick-breaking process) with points

$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent replicas of W.



Permutations derived from stick-breaking:
Basic construction

Let W be a random variable, called stick-breaking factor, with values in $(0, 1)$.

Let Q be a multiplicative random walk (also known as stick-breaking process) with points

$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent replicas of W.

Let $U_1, \ldots, U_n$ be a sample from the uniform $[0, 1]$ distribution, independent of Q.

Permutations derived from stick-breaking:
Basic construction

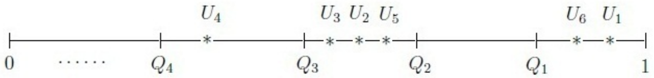
Let W be a random variable, called stick-breaking factor, with values in $(0, 1)$.

Let Q be a multiplicative random walk (also known as stick-breaking process) with points

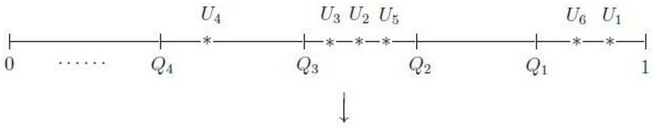
$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent replicas of W.

Let $U_1, \ldots, U_n$ be a sample from the uniform $[0, 1]$ distribution, independent of Q.

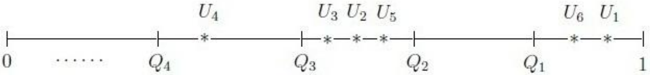


Permutations derived from stick-breaking:
Basic construction



(4) (3 2 5) (6 1)

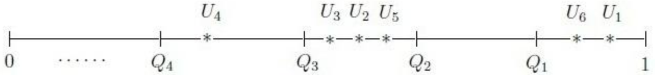
Permutations derived from stick-breaking:
Basic construction



$$(4)\ (3\ 2\ 5)\ (6\ 1)$$

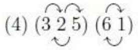
$$(4)\ (3\ 2\ 5)\ (6\ 1) \longrightarrow (1\ 6)\ (2\ 5\ 3)\ (4)$$ Standard Cycle Notation

Permutations derived from stick-breaking:
Basic construction

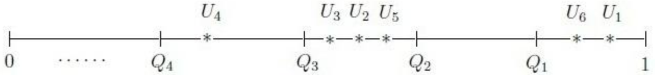


$$(4)\ (3\ 2\ 5)\ (6\ 1)$$

$$(4)\ (3\ 2\ 5)\ (6\ 1) \longrightarrow (1\ 6)\ (2\ 5\ 3)\ (4)$$ Standard Cycle Notation

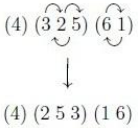


Permutations derived from stick-breaking:
Basic construction

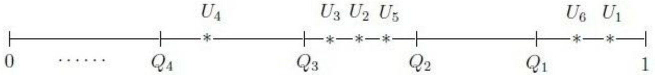


$$(4)\ (3\ 2\ 5)\ (6\ 1)$$

$$(4)\ (3\ 2\ 5)\ (6\ 1) \longrightarrow (1\ 6)\ (2\ 5\ 3)\ (4)$$ Standard Cycle Notation



Permutations derived from stick-breaking:
Basic construction



$$(4)\ (3\ 2\ 5)\ (6\ 1)$$

$$(4)\ (3\ 2\ 5)\ (6\ 1) \longrightarrow (1\ 6)\ (2\ 5\ 3)\ (4)$$ Standard Cycle Notation

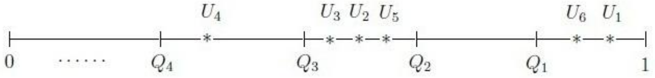
$$(4)\ (3\ 2\ 5)\ (6\ 1)$$

$$\downarrow$$

$$(4)\ (2\ 5\ 3)\ (1\ 6)$$

$$(4)\ (2\ 5\ 3)\ (1\ 6)$$

Permutations derived from stick-breaking:
Basic construction



$$(4)\ (3\ 2\ 5)\ (6\ 1)$$

$$(4)\ (3\ 2\ 5)\ (6\ 1) \longrightarrow (1\ 6)\ (2\ 5\ 3)\ (4)$$

$$(4)\ (3\ 2\ 5)\ (6\ 1)$$

$$\downarrow$$

$$(4)\ (2\ 5\ 3)\ (1\ 6)$$

$$(4)\ (2\ 5\ 3)\ (1\ 6)$$

$$\downarrow$$

$$(1\ 6)\ (2\ 5\ 3)\ (4)$$ Standard Cycle Notation

Permutations derived from stick-breaking:

Basic construction

- Regeneration : For $m \in \{1, \ldots, n-1\}$, conditionally given the last cycle of Π_n has length m, the cycle partition of Π_n with the last cycle deleted has the same distribution as the cycle partition of Π_{n-m}.

Permutations derived from stick-breaking:
Basic construction

- Regeneration : For $m \in \{1, \dots, n-1\}$, conditionally given the last cycle of Π_n has length m, the cycle partition of Π_n with the last cycle deleted has the same distribution as the cycle partition of Π_{n-m}.

- Coherence : Permutations Π_n's are defined consistently for all values of n. Passing from Π_{n+1} to Π_n amounts to removing integer $n+1$ from a cycle.

Permutations derived from stick-breaking:
Basic construction

- Regeneration : For $m \in \{1, \dots, n-1\}$, conditionally given the last cycle of Π_n has length m, the cycle partition of Π_n with the last cycle deleted has the same distribution as the cycle partition of Π_{n-m}.

- Coherence : Permutations Π_n's are defined consistently for all values of n. Passing from Π_{n+1} to Π_n amounts to removing integer $n+1$ from a cycle.
- Exchangeability : Given the cycle partition $(K_{n,1}, \dots, K_{n,n})$ the distribution of Π_n is uniform.

Permutations derived from stick-breaking:
Basic construction

- Regeneration : For $m \in \{1, \dots, n-1\}$, conditionally given the last cycle of Π_n has length m, the cycle partition of Π_n with the last cycle deleted has the same distribution as the cycle partition of Π_{n-m}.

- Coherence : Permutations Π_n's are defined consistently for all values of n. Passing from Π_{n+1} to Π_n amounts to removing integer $n+1$ from a cycle.
- Exchangeability : Given the cycle partition $(K_{n,1}, \dots, K_{n,n})$ the distribution of Π_n is uniform.

Regeneration+exchangeability $\rightarrow$ given the last cycle of Π_n is of length m, a permutation resulting from deletion of the last cycle and re-labelling the remaining elements by the increasing bijection with $[n-m]$ is a distributional copy of Π_{n-m}.

Permutations derived from stick-breaking:
Basic construction

- Regeneration : For $m \in \{1, \ldots, n-1\}$, conditionally given the last cycle of Π_n has length m, the cycle partition of Π_n with the last cycle deleted has the same distribution as the cycle partition of Π_{n-m}.

- Coherence : Permutations Π_n's are defined consistently for all values of n. Passing from Π_{n+1} to Π_n amounts to removing integer $n+1$ from a cycle.
- Exchangeability : Given the cycle partition $(K_{n,1}, \ldots, K_{n,n})$ the distribution of Π_n is uniform.

Gnedin & Pitman (2005): each system of coherent random permutations $(\Pi_n)_{n \in \mathbb{N}}$ with the properties of exchangeability and regeneration uniquely corresponds to the closed range of a subordinator S.

Permutations derived from stick-breaking:
Basic construction

- Regeneration : For $m \in \{1, \dots, n-1\}$, conditionally given the last cycle of Π_n has length m, the cycle partition of Π_n with the last cycle deleted has the same distribution as the cycle partition of Π_{n-m}.
- Coherence : Permutations Π_n's are defined consistently for all values of n. Passing from Π_{n+1} to Π_n amounts to removing integer $n+1$ from a cycle.
- Exchangeability : Given the cycle partition $(K_{n,1}, \dots, K_{n,n})$ the distribution of Π_n is uniform.

Gnedin & Pitman (2005): each system of coherent random permutations $(\Pi_n)_{n\in\mathbb{N}}$ with the properties of exchangeability and regeneration uniquely corresponds to the closed range of a subordinator S.

For the permutations derived from stick-breaking:

- S is a compound Poisson process with jumps $(|\log W_k|)$;

Permutations derived from stick-breaking:
Basic construction

- Regeneration: For $m \in \{1, \dots, n-1\}$, conditionally given the last cycle of Π_n has length m, the cycle partition of Π_n with the last cycle deleted has the same distribution as the cycle partition of Π_{n-m}.
- Coherence: Permutations Π_n's are defined consistently for all values of n. Passing from Π_{n+1} to Π_n amounts to removing integer $n+1$ from a cycle.
- Exchangeability: Given the cycle partition $(K_{n,1}, \dots, K_{n,n})$ the distribution of Π_n is uniform.

Gnedin & Pitman (2005): each system of coherent random permutations $(\Pi_n)_{n \in \mathbb{N}}$ with the properties of exchangeability and regeneration uniquely corresponds to the closed range of a subordinator S.

For the permutations derived from stick-breaking:

- S is a compound Poisson process with jumps $(|\log W_k|)$;
- The length of the last cycle of Π_n is of order $\Theta(n)$, as n grows.

Permutations derived from steak-breaking:
A Markov chain representation

$K_{n,r}$ is the number of cycles of length r.

Consider a decreasing Markov chain on nonnegative integers with absorption at 0 and the decrement matrix

$$q(n,m) = \binom{n}{m} \frac{\mathbb{E}[W^{n-m}(1-W)^m]}{1-\mathbb{E}W^n}, \quad 1 \le m \le n,$$

specifying transition probabilities from n to $n-m$.

$K_{n,r}$ is the number of cycles of length r.

Consider a decreasing Markov chain on nonnegative integers with absorption at 0 and the decrement matrix

$$q(n,m) = \binom{n}{m} \frac{\mathbb{E}[W^{n-m}(1-W)^m]}{1-\mathbb{E}W^n}, \qquad 1 \le m \le n,$$

specifying transition probabilities from n to $n-m$.

For the Markov chain starting at n, $K_{n,r}$ is the number of decrements of size r on the path from n to 0.

Permutations derived from stick-breaking:
The infinite occupancy scheme

$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent copies of the stick-breaking factor W.

For $j \in \mathbb{N}$, the interval $(Q_j, Q_{j-1}]$ = box with index j and frequency

$$P_j := W_1 W_2 \cdots W_{j-1}(1 - W_j).$$

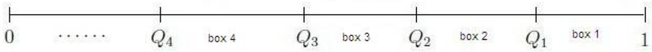
Permutations derived from stick-breaking:
The infinite occupancy scheme

$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent copies of the stick-breaking factor W.

For $j \in \mathbb{N}$, the interval $(Q_j, Q_{j-1}]$ = box with index j and frequency

$$P_j := W_1 W_2 \cdots W_{j-1}(1 - W_j).$$



Permutations derived from stick-breaking:
The infinite occupancy scheme

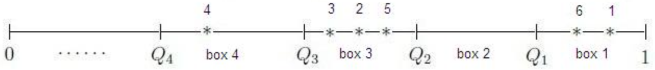
$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent copies of the stick-breaking factor W.

For $j \in \mathbb{N}$, the interval $(Q_j, Q_{j-1}]$ = box with index j and frequency

$$P_j := W_1 W_2 \cdots W_{j-1}(1 - W_j).$$

Given the frequencies, balls $1, 2, \dots$ are thrown independently so that each ball hits box j with probability P_j.



Permutations derived from stick-breaking:
The infinite occupancy scheme

$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

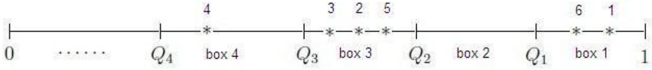
where (W_i) are independent copies of the stick-breaking factor W.

For $j \in \mathbb{N}$, the interval $(Q_j, Q_{j-1}]$ = box with index j and frequency

$$P_j := W_1 W_2 \cdots W_{j-1}(1 - W_j).$$

Given the frequencies, balls $1, 2, \dots$ are thrown independently so that each ball hits box j with probability P_j.

Then $K_{n,r}$ is the number of boxes occupied by exactly r out of the first n balls.



Permutations derived from stick-breaking:
The infinite occupancy scheme

$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent copies of the stick-breaking factor W.

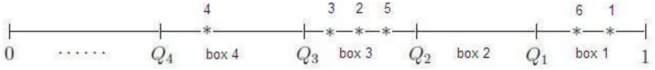
For $j \in \mathbb{N}$, the interval $(Q_j, Q_{j-1}] =$ box with index j and frequency

$$P_j := W_1 W_2 \cdots W_{j-1}(1 - W_j).$$

Given the frequencies, balls $1, 2, \ldots$ are thrown independently so that each ball hits box j with probability P_j.

Then $K_{n,r}$ is the number of boxes occupied by exactly r out of the first n balls.

$$K_{6,4} = 0, \quad K_{6,3} = 1$$



Permutations derived from stick-breaking:
The infinite occupancy scheme

For $j \in \mathbb{N}$, the interval $(Q_j, Q_{j-1}]$ = box with index j and frequency

$$P_j := W_1 W_2 \cdots W_{j-1}(1 - W_j).$$

Given the frequencies, balls $1, 2, \ldots$ are thrown independently so that each ball hits box j with probability P_j.

Then $K_{n,r}$ is the number of boxes occupied by exactly r out of the first n balls.

This 'balls-in-boxes' scheme is called the Bernoulli sieve.

Permutations derived from stick-breaking:
The infinite occupancy scheme

For $j \in \mathbb{N}$, the interval $(Q_j, Q_{j-1}]$ = box with index j and frequency

$$P_j := W_1 W_2 \cdots W_{j-1}(1 - W_j).$$

Given the frequencies, balls $1, 2, \ldots$ are thrown independently so that each ball hits box j with probability P_j.

Then $K_{n,r}$ is the number of boxes occupied by exactly r out of the first n balls.

This 'balls-in-boxes' scheme is called the Bernoulli sieve.

The model was investigated by

- Gnedin (2004),
- Gnedin, I. & Roesler (2008),
- Gnedin, I., Roesler & Negadajlov (2009),
- Gnedin, I. & Marynych (2010a,b),
- I. (2011+).

Extension of the Erdős-Turán law:
Main result

Let O_n denote the order of random permutation Π_n derived from stick-breaking of $[0,1]$, i.e.,

$$O_n = \text{l.c.m.}\{r \in [n] : K_{n,r} > 0\},$$

where $K_{n,r}$ is the number of cycles of Π_n with length r.

Theorem

Suppose the stick-breaking factor W has density f which satisfies

$$\sup_{x\in[0,1]} x^\alpha (1-x)^\alpha f(x) < \infty,$$

for some $\alpha \in [0,1)$. Then $\sigma^2 := \text{Var}\,(\log W) < \infty$ and

$$\frac{\log O_n - (2\mu)^{-1}\log^2 n}{\sqrt{(3\mu^3)^{-1}\sigma^2 \log^3 n}} \overset{d}{\to} \mathcal{N}(0,1), \quad n \to \infty,$$

where $\mu := \mathbb{E}|\log W|$.

Extension of the Erdős-Turán law:
Main result

Theorem

Suppose the stick-breaking factor W has density f which satisfies

$$\sup_{x\in[0,1]} x^\alpha(1-x)^\alpha f(x) < \infty,$$

for some $\alpha \in [0,1)$. Then $\sigma^2 := \mathrm{Var}(\log W) < \infty$ and

$$\frac{\log O_n - (2\mu)^{-1}\log^2 n}{\sqrt{(3\mu^3)^{-1}\sigma^2 \log^3 n}} \overset{d}{\to} \mathcal{N}(0,1), \quad n\to\infty,$$

where $\mu := \mathbb{E}|\log W|$.

Applicability : Theorem applies if, for instance,

- W has a bounded density;
- W has the beta (a,b) density, $a,b>0$:

$$f(x) = \frac{1}{\mathrm{B}(a,b)} x^{a-1}(1-x)^{b-1} 1_{(0,1)}(x).$$

Extension of the Erdős-Turán law:
Main result

Theorem

Suppose the stick-breaking factor W has density f which satisfies

$$\sup_{x\in[0,1]} x^{\alpha}(1-x)^{\alpha} f(x) < \infty,$$

for some $\alpha \in [0,1)$. Then $\sigma^2 := \mathrm{Var}\,(\log W) < \infty$ and

$$\frac{\log O_n - (2\mu)^{-1}\log^2 n}{\sqrt{(3\mu^3)^{-1}\sigma^2 \log^3 n}} \xrightarrow{d} \mathcal{N}(0,1), \quad n \to \infty,$$

where $\mu := \mathbb{E}|\log W|$.

Applicability : Theorem applies if, for instance,

- W has a bounded density;
- W has the beta (a, b) density, $a, b > 0$:

$$f(x) = \frac{1}{\mathrm{B}(a,b)} x^{a-1}(1-x)^{b-1} 1_{(0,1)}(x).$$

! If $W \stackrel{d}{=} \mathrm{beta}(\theta, 1)$, $\theta > 0$ we recover the Erdős-Turán law for Ewens' permutations.

Extension of the Erdős-Turán law:
Scheme of the proof

$$O_n = \text{l.c.m.}\{r \in [n] : K_{n,r} > 0\} = \text{l.c.m.}\{Z_{n,k} : k \le K_n\},$$

where $Z_{n,k}$ is the length of the kth cycle of Π_n, and $K_n := \sum_{r=1}^{n} K_{n,r}$ is the total number of cycles of Π_n.

STEP 1: Show that $\log T_n$ yields a good approximation to $\log O_n$, where

$$T_n := \prod_{r=1}^{n} r^{K_{n,r}} = \prod_{k=1}^{K_n} Z_{n,k},$$

is the product of cycle lengths of Π_n.

Extension of the Erdős-Turán law:
Scheme of the proof

$$O_n = \text{l.c.m.}\{r \in [n] : K_{n,r} > 0\} = \text{l.c.m.}\{Z_{n,k} : k \leq K_n\},$$

where $Z_{n,k}$ is the length of the kth cycle of Π_n, and $K_n := \sum_{r=1}^{n} K_{n,r}$ is the total number of cycles of Π_n.

STEP 1: Show that $\log T_n$ yields a good approximation to $\log O_n$, where

$$T_n := \prod_{r=1}^{n} r^{K_{n,r}} = \prod_{k=1}^{K_n} Z_{n,k},$$

is the product of cycle lengths of Π_n.

STEP 2: Investigate the weak convergence of $\log T_n$.

Extension of the Erdős-Turán law:
Scheme of the proof

$$O_n = \text{l.c.m.}\{r \in [n] : K_{n,r} > 0\} = \text{l.c.m.}\{Z_{n,k} : k \le K_n\},$$

where $Z_{n,k}$ is the length of the kth cycle of Π_n, and $K_n := \sum_{r=1}^n K_{n,r}$ is the total number of cycles of Π_n.

STEP 1: Show that $\log T_n$ yields a good approximation to $\log O_n$, where

$$T_n := \prod_{r=1}^{n} r^{K_{n,r}} = \prod_{k=1}^{K_n} Z_{n,k},$$

is the product of cycle lengths of Π_n.

STEP 2: Investigate the weak convergence of $\log T_n$.

Functional $\sum_r K_{n,r} h(r)$ is called a separable statistic. The term was coined by Medvedev (1977).

! The variables $\log T_n$, $K_{n,r}$, K_n are instances of separable statistics.

Extension of the Erdős-Turán law:
Scheme of the proof

$$O_n = \text{l.c.m.}\{r \in [n] : K_{n,r} > 0\} = \text{l.c.m.}\{Z_{n,k} : k \le K_n\},$$

where $Z_{n,k}$ is the length of the kth cycle of Π_n, and $K_n := \sum_{r=1}^{n} K_{n,r}$ is the total number of cycles of Π_n.

STEP 1: Show that $\log T_n$ yields a good approximation to $\log O_n$, where

$$T_n := \prod_{r=1}^{n} r^{K_{n,r}} = \prod_{k=1}^{K_n} Z_{n,k},$$

is the product of cycle lengths of Π_n.

STEP 2: Investigate the weak convergence of $\log T_n$.

Functional $\sum_r K_{n,r} h(r)$ is called a separable statistic. The term was coined by Medvedev (1977).

! The variables $\log T_n$, $K_{n,r}$, K_n are instances of separable statistics.

For Ewens' permutations quite general separable statistics were studied in a series of papers by Babu and Manstavičius, for instance, Babu and Manstavičius (2002), Manstavičius (2009).

Extension of the Erdős-Turán law:

Weak convergence of the product of cycle lengths

$\mu = \mathbb{E}|\log W|$, $\sigma^2 = \mathrm{Var}\,(\log W)$

Theorem (a) If $\sigma^2 < \infty$ then, with

$$b_n = \mu^{-1}\left(2^{-1}\log^2 n - \int_0^{\log n}\int_0^z \mathbb{P}\{|\log(1-W)| > x\}\mathrm{d}x\mathrm{d}z\right) \quad (1)$$

and $a_n = ((3\mu^3)^{-1}\sigma^2 \log^3 n)^{1/2}$, the limiting distribution of $(\log T_n - b_n)/a_n$ is standard normal.

(b) If $\sigma^2 = \infty$, and

$$\int_0^x y^2\, \mathbb{P}\{|\log W| \in \mathrm{d}y\} \ \sim \ \ell(x), \quad x \to \infty,$$

for some ℓ slowly varying at ∞, then, with b_n given by (1) and $a_n = (3\mu^3)^{-1/2} c_{[\log n]} \log n$, where (c_n) is any positive sequence satisfying $\lim_{n\to\infty} n\ell(c_n)/c_n^2 = 1$, the limiting distribution of $(\log T_n - b_n)/a_n$ is standard normal.

Extension of the Erdős-Turán law:
Weak convergence of the product of cycle lengths

$\mu = \mathbb{E}|\log W|$

(c) If

$$\mathbb{P}\{|\log W| > x\} \sim x^{-\alpha}\ell(x), \quad x \to \infty,$$

for some ℓ slowly varying at ∞ and $\alpha \in (1,2)$ then, with

$$b_n = \mu^{-1}\left(2^{-1}\log^2 n - \int_0^{\log n}\int_0^z \mathbb{P}\{|\log(1-W)| > x\}\mathrm{d}x\mathrm{d}z\right)$$

and

$$a_n = ((\alpha+1)\mu^{\alpha+1})^{-1/\alpha} c_{\lfloor \log n \rfloor} \log n,$$

where (c_n) is any positive sequence satisfying $\lim_{n\to\infty} n\ell(c_n)/c_n^\alpha = 1$, the limiting distribution of $(\log T_n - b_n)/a_n$ is the α-stable law with characteristic function

$$u \mapsto \exp\{-|u|^\alpha \Gamma(1-\alpha)(\cos(\pi\alpha/2) + i\sin(\pi\alpha/2)\,\mathrm{sgn}(u))\}, \ u \in \mathbb{R}.$$

Extension of the Erdős-Turán law:
Weak convergence of the product of cycle lengths

Theorem (a) If $\sigma^2 < \infty$ then, with

$$b_n = \mu^{-1}\left(2^{-1}\log^2 n - \int_0^{\log n}\int_0^z \mathbb{P}\{|\log(1-W)| > x\}\mathrm{d}x\mathrm{d}z\right) \qquad (1)$$

and $a_n = ((3\mu^3)^{-1}\sigma^2\log^3 n)^{1/2}$, the limiting distribution of $(\log T_n - b_n)/a_n$ is standard normal.

Extension of the Erdős-Turán law:
Weak convergence of the product of cycle lengths

Theorem (a) If $\sigma^2 < \infty$ then, with

$$b_n = \mu^{-1}\Big(2^{-1}\log^2 n - \int_0^{\log n}\int_0^z \mathbb{P}\{|\log(1-W)| > x\}\mathrm{d}x\mathrm{d}z\Big) \qquad (1)$$

and $a_n = ((3\mu^3)^{-1}\sigma^2\log^3 n)^{1/2}$, the limiting distribution of $(\log T_n - b_n)/a_n$ is standard normal.

Example 1. Assume that

$$W \overset{d}{=} \mathrm{beta}(\theta, 1), \quad \theta > 0.$$

Then $\mu = \theta^{-1}$, $\sigma^2 = \theta^{-2}$ and

$$\lim_{n\to\infty} \frac{\int_0^{\log n}\int_0^z \mathbb{P}\{|\log(1-W)| > x\}\mathrm{d}x\mathrm{d}z}{\log^{3/2} n} = 0,$$

since the numerator is $O(\log n)$. Application of the theorem yields

$$\frac{\log T_n - (\theta/2)\log^2 n}{\sqrt{(\theta/3)\log^3 n}} \overset{d}{\to} \mathcal{N}(0,1), \quad n\to\infty,$$

which agrees with Arratia & Tavaré (1992).

Extension of the Erdős-Turán law:
Weak convergence of the product of cycle lengths

Theorem (a) If $\sigma^2 < \infty$ then, with

$$b_n = \mu^{-1}\left(2^{-1}\log^2 n - \int_0^{\log n}\int_0^z \mathbb{P}\{|\log(1-W)| > x\}\mathrm{d}x\mathrm{d}z\right) \qquad (1)$$

and $a_n = ((3\mu^3)^{-1}\sigma^2 \log^3 n)^{1/2}$, the limiting distribution of $(\log T_n - b_n)/a_n$ is standard normal.

Example 2. Assume that, for some $\gamma \in (0, 1/2)$,

$$\mathbb{P}\{W > x\} = \frac{1}{1 + |\log(1-x)|^\gamma}, \quad x \in [0, 1).$$

Then $\mathbb{E}\log^2 W < \infty$ and

$$\mathbb{P}\{|\log(1-W)| > x\} \sim x^{-\gamma} \quad \text{as } x \to \infty,$$

and in this case,

$$a_n = \text{const}\log^{3/2} n$$

and

$$b_n = \mu^{-1}(2^{-1}\log^2 n - \text{const}\log^{2-\gamma} n + o(\log^{2-\gamma} n)).$$

!! Thus we see that the second term $b_n - (2\mu)^{-1}\log n$ of centering cannot be ignored.

Extension of the Erdős-Turán law:
Asymptotics of the cycle partition

Q–a multiplicative random walk with points

$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent replicas of W.

$U_1, \ldots, U_n$– a sample from the uniform [0, 1] distribution, independent of Q.

$K_{n,r}$– the number of boxes occupied by exactly r out of the first n balls.

Extension of the Erdős-Turán law:
Asymptotics of the cycle partition

Q–a multiplicative random walk with points

$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent replicas of W.

$U_1, \ldots, U_n$– a sample from the uniform [0, 1] distribution, independent of Q.

$K_{n,r}$– the number of boxes occupied by exactly r out of the first n balls.

The cycle partition $(K_{n,1}, \ldots, K_{n,n})$ will remain the same when defined in terms of

$$nQ \quad \text{and} \quad \mathcal{U}_n := \{nU_{j,n} : \ 1 \le j \le n\},$$

where $U_{1,n} \le U_{2,n} \le \ldots \le U_{n,n}$ are the corresponding order statistics.

Extension of the Erdős-Turán law:

Asymptotics of the cycle partition

Q–a multiplicative random walk with points

$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent replicas of W.

$U_1, \ldots, U_n$– a sample from the uniform [0, 1] distribution, independent of Q.

$K_{n,r}$– the number of boxes occupied by exactly r out of the first n balls.

The cycle partition $(K_{n,1}, \ldots, K_{n,n})$ will remain the same when defined in terms of

$$nQ \quad \text{and} \quad \mathcal{U}_n := \{nU_{j,n} : \ 1 \le j \le n\},$$

where $U_{1,n} \le U_{2,n} \le \ldots \le U_{n,n}$ are the corresponding order statistics.

$\lim_{n\to\infty} \mathcal{U}_n = \mathcal{U}$ vaguely, where $\mathcal{U}$ is a Poisson process on $\mathbb{R}^+$ with unit intensity.

Extension of the Erdős-Turán law:
Asymptotics of the cycle partition

Q–a multiplicative random walk with points

$$Q_0 := 1, \quad Q_j := \prod_{i=1}^{j} W_i, \quad j \in \mathbb{N},$$

where (W_i) are independent replicas of W.

$U_1, \ldots, U_n$– a sample from the uniform [0, 1] distribution, independent of Q.

$K_{n,r}$– the number of boxes occupied by exactly r out of the first n balls.

The cycle partition $(K_{n,1}, \ldots, K_{n,n})$ will remain the same when defined in terms of

$$nQ \quad \text{and} \quad \mathcal{U}_n := \{nU_{j,n} : 1 \leq j \leq n\},$$

where $U_{1,n} \leq U_{2,n} \leq \ldots \leq U_{n,n}$ are the corresponding order statistics.

For $\mathcal{U}$ a Poisson process on $\mathbb{R}^+$ with unit intensity,

$$\lim_{n\to\infty} \mathcal{U}_n = \mathcal{U} \quad \text{vaguely.}$$

For $-\log \mathcal{Q}$ a stationary renewal process on $\mathbb{R}$,

$$\lim_{n\to\infty} nQ = \mathcal{Q} \quad \text{vaguely.}$$

Extension of the Erdős-Turán law:
Asymptotics of the cycle partition

For $\mathcal{U}$ a Poisson process on $\mathbb{R}^+$ with unit intensity,

$$\lim_{n\to\infty} \mathcal{U}_n = \mathcal{U} \text{ vaguely.}$$

For $-\log \mathcal{Q}$ a stationary renewal process on $\mathbb{R}$,

$$\lim_{n\to\infty} nQ = \mathcal{Q} \text{ vaguely.}$$

Limiting 'balls-in-boxes' scheme:

- 'boxes'–intervals induced by successive points of $\mathcal{Q}$
- 'balls'–points of $\mathcal{U}$

Extension of the Erdős-Turán law:
Asymptotics of the cycle partition

For $\mathcal{U}$ a Poisson process on $\mathbb{R}^+$ with unit intensity,

$$\lim_{n\to\infty} \mathcal{U}_n = \mathcal{U} \ \text{vaguely}.$$

For $-\log \mathcal{Q}$ a stationary renewal process on $\mathbb{R}$,

$$\lim_{n\to\infty} nQ = \mathcal{Q} \ \text{vaguely}.$$

Limiting 'balls-in-boxes' scheme:

- 'boxes'–intervals induced by successive points of $\mathcal{Q}$
- 'balls'–points of $\mathcal{U}$

In the limiting scheme, for $r \in \mathbb{N}$ denote by K_r the number of intervals that contain exactly r Poisson points.

Extension of the Erdős-Turán law:
Asymptotics of the cycle partition

For $\mathcal{U}$ a Poisson process on $\mathbb{R}^+$ with unit intensity,

$$\lim_{n\to\infty} \mathcal{U}_n = \mathcal{U} \text{ vaguely.}$$

For $-\log \mathcal{Q}$ a stationary renewal process on $\mathbb{R}$,

$$\lim_{n\to\infty} n\mathcal{Q} = \mathcal{Q} \text{ vaguely.}$$

Limiting 'balls-in-boxes' scheme:

- 'boxes'–intervals induced by successive points of $\mathcal{Q}$
- 'balls'–points of $\mathcal{U}$

In the limiting scheme, for $r \in \mathbb{N}$ denote by K_r the number of intervals that contain exactly r Poisson points.

Theorem (Gnedin, I. and Roesler (2008)). As $n \to \infty$, we have

$$(K_{n,1}, K_{n,2} \dots) \to_d (K_1, K_2, \dots),$$

along with

$$\mathbb{E}[K_{n,r}] \to \mathbb{E}[K_r] = \frac{1}{r\mathbb{E}|\log W|}.$$