

# Random permutations (and beyond)

**Permutation Patterns 2012**

Jennie Hansen (Heriot-Watt)

Joint work with **Jerzy Jaworski** (Adam Mickiewicz University, Poznań, Poland)

Supported by

**RANDOMAPP** No. 236845 Marie Curie Intra-European Fellowship  
within the 7th European Community Framework Programme

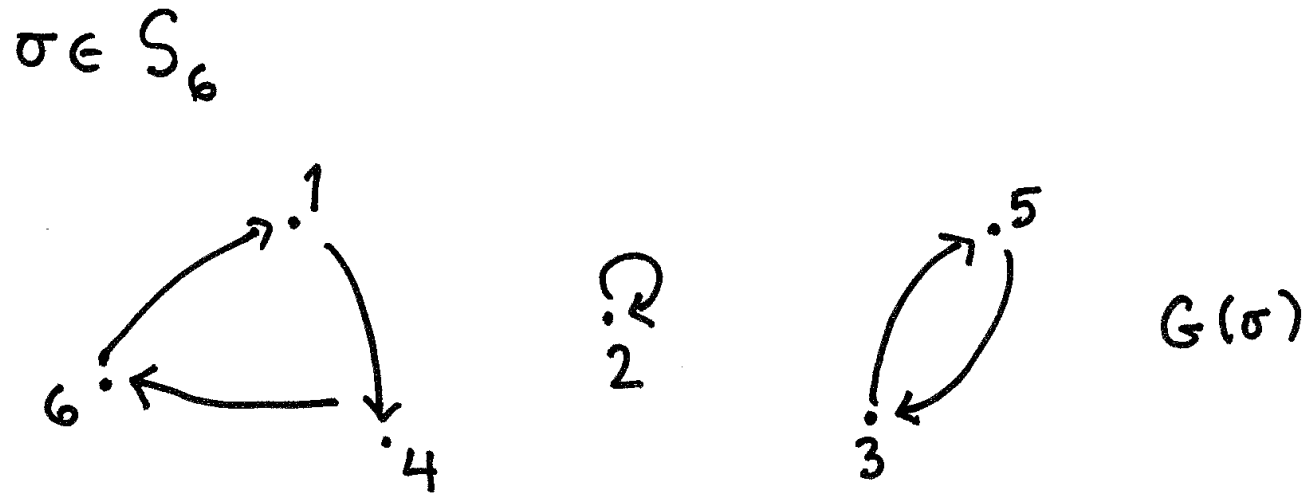
“In this talk we view uniform random permutations as part of a continuum of random mapping models and we investigate the component structure of the random mappings in this continuum as the mappings become (in some sense) more like permutations. ”

## Notation:

- $[n] = \{1, 2, \dots, n\}$
- $\mathcal{M}_n =$  the set of all mappings  $f : [n] \rightarrow [n]$
- $S_n =$  the set of all permutations  $\sigma : [n] \rightarrow [n]$
- For  $f \in \mathcal{M}_n$ ,  $G(f)$  denotes the directed  $n$  labelled vertices which 'represents' the mapping  $f$  - i.e. there is a directed edge  $i \rightarrow j$  in  $G(f)$  iff  $f(i) = j$ .  
**NB** The directed graph  $G(f)$  is characterised by the feature that every vertex has out-degree 1.

If  $\sigma \in S_n$ , the  $G(\sigma)$  is the directed graph that represents the cycle structure of  $\sigma$ :

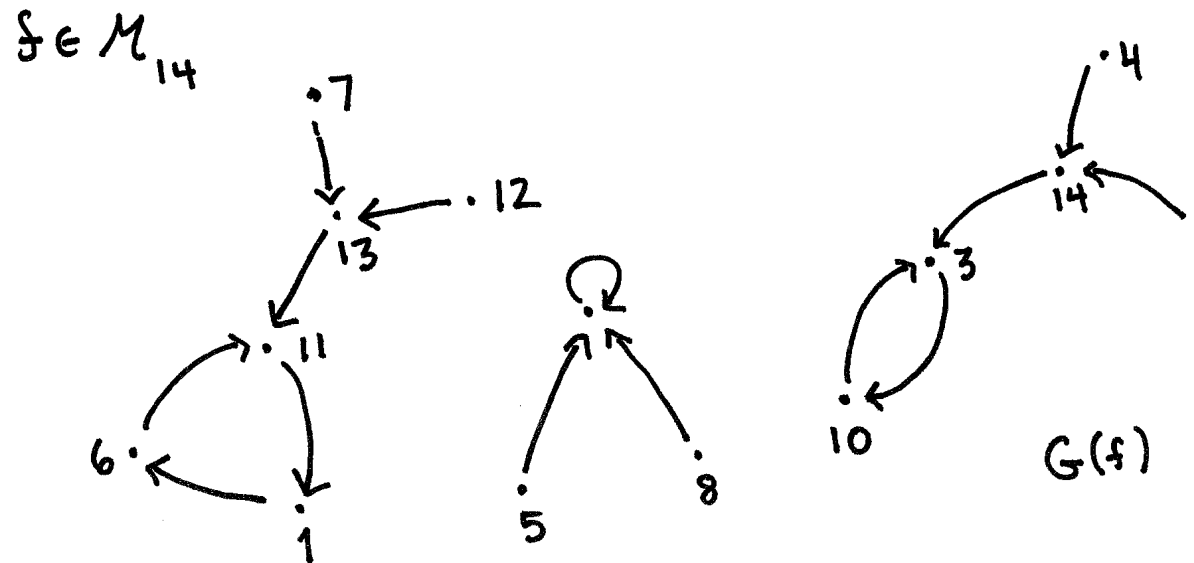
**Example**



**Important:** Every vertex in  $G(\sigma)$  has in-degree 1 (as well as out-degree 1).

If  $f \in \mathcal{M}_n \setminus S_n$ , then  $G(f)$  consists of directed cycles with directed trees attached to the cycles:

**Example**



If  $T$  is a *random* element of  $\mathcal{M}_n$ , then  $G(T)$  is a *random directed graph* on  $n$  labelled vertices and we can ask ‘statistical’ questions about the distribution of structural variables associated with  $G(T)$ , e.g.

- number of connected components in  $G(T)$
- size of a ‘typical’ component in  $G(T)$
- number of cyclical vertices in  $G(T)$
- size of the largest tree in  $G(T)$ , etc.

## Two important examples

$\sigma_n$  - uniformly distributed random element of  $S_n$

$T_n$  - uniformly distributed random element of  $\mathcal{M}_n$

**Main idea:**  $\sigma_n$  and  $T_n$  can be viewed as random mappings that are at either end of a *continuum* of random mapping models.



## Model for random mappings with bounded in-degrees

Suppose that  $m \geq 1$  is fixed and suppose that we have an urn which contains  $m$  balls labelled  $i$  for  $i = 1, 2, \dots, n$ .

Select  $n$  balls from the urn, one by one and without replacement, from the urn and define the random mapping  $T_{n,m}$  by setting

$$T_{n,m}(i) = j \quad \text{iff} \quad i^{\text{th}} \text{ removed from the urn is labelled } j.$$

### Note:

- The parameter  $m$  is an upper bound on the in-degree of the vertices in  $G(T_{n,m})$  and when  $m = 1$ , we have  $T_{n,1} = \sigma_n$ .
- If  $n$  is fixed, then  $T_{n,m}$  'converges' to  $T_n$  as  $m \rightarrow \infty$ .

**Main point:** We are interested in comparing the structure of  $G(T_{n,m})$  to the structure of  $G(\sigma_n)$  and  $G(T_n)$ .

**Example: Size of a ‘typical’ component in  $G(T_{n,m})$**

For any  $f \in \mathcal{M}_n$ , let  $C_1(T) =$  the size of the component in  $G(f)$  that contains the vertex labelled ‘1’.

**Uniform random permutations:** For  $1 \leq k \leq n$ ,

$$\mathbf{P}\{C_1(\sigma_n) = k\} = \frac{1}{n}$$

and for  $0 < a < b < 1$ ,

$$\lim_{n \rightarrow \infty} \mathbf{P}\left\{a \leq \frac{C_1(\sigma_n)}{n} \leq b\right\} = \int_a^b 1 dx.$$

**Uniform random mappings:** For  $1 \leq k \leq n$

$$\mathbf{P}\{C_1(T_n) = k\} = \frac{(n)_k (n-k)^{n-k}}{n^{n+1}} \sum_{j=0}^{k-1} \frac{(k)^j}{j!}$$

and for  $0 < a < b < 1$ ,

$$\lim_{n \rightarrow \infty} \mathbf{P}\left\{a \leq \frac{C_1(T_n)}{n} \leq b\right\} = \int_a^b \frac{1}{2\sqrt{1-x}} dx.$$

**Remark:** For both uniform random permutations and uniform random mappings, the limiting densities are Beta densities of the form  $f_\theta(x) = \theta(1-x)^{\theta-1}$  for  $0 < x < 1$ .

**Q.** What can we say about the exact and asymptotic distribution of  $C_1(T_{n,m})$ ?

## Detour: Random mappings, $T_n^{\hat{D}}$ , with exchangeable in-degrees

**Notation:** For any sequence of non-negative integers  $d_1, d_2, \dots, d_n$  such that  $\sum_{i=1}^n d_i = n$ , let  $\mathcal{M}_n(d_1, d_2, \dots, d_n) =$  set of all mappings from  $[n] \rightarrow [n]$  with in-degree sequence  $(d_1, d_2, \dots, d_n)$ .

### Construction of $T_n^{\hat{D}}$ :

Start with a sequence  $\hat{D}_1, \hat{D}_2, \dots, \hat{D}_n$  of exchangeable, non-negative, integer-valued random variables such that  $\sum_{i=1}^n \hat{D}_i \equiv n$ , then the random mapping  $T_n^{\hat{D}}$  is constructed as follows:

- Suppose that  $\hat{D}_i = d_i$ , for  $i = 1, 2, \dots, n$ .
- *Given* that  $\hat{D}_i = d_i$ , for  $i = 1, 2, \dots, n$ , choose a mapping  $f$  uniformly from  $\mathcal{M}_n(d_1, d_2, \dots, d_n)$ .
- Set  $T_n^{\hat{D}} = f$ .

**Key Point:** There is a calculus based on the variables  $\hat{D}_1, \hat{D}_2, \dots, \hat{D}_n$  for computing the distribution of many variables associated with the structure of  $G(T_n^{\hat{D}})$ .

**Example:** For any  $f \in \mathcal{M}_n$ , let  $X_n(f) =$  the number of cyclical vertices in  $G(f)$ .

**Theorem**

$$\mathbf{P}\{X_n(T_n^{\hat{D}}) = k\} = \frac{k}{n-k} E((\hat{D}_1 - 1)\hat{D}_1\hat{D}_2 \cdots \hat{D}_k)$$

for  $1 \leq k \leq n - 1$ , and

$$\mathbf{P}\{X_n(T_n^{\hat{D}}) = n\} = E(\hat{D}_1\hat{D}_2 \cdots \hat{D}_n).$$

## Some (easy) consequences of the Theorem

1. Given that  $X_n(T_n^{\hat{D}}) = k$ ,  $T_n^{\hat{D}}$  restricted to the cyclical vertices of  $G(T_n^{\hat{D}})$  is a uniform random permutation. It follows that the number of components in  $G(T_n^{\hat{D}})$  equals the number of cycles in the 'induced' random permutation. So, using the Theorem and known results for uniform random permutations, we can obtain the distribution of the number of components in  $G(T_n^{\hat{D}})$ .

2.

$$\begin{aligned} \mathbf{P}\{G(T_n^{\hat{D}}) \text{ is connected}\} &= \mathbf{P}\{\text{induced permutation is cyclical}\} \\ &= \sum_{k=1}^n \frac{1}{k} \mathbf{P}\{X_n(T_n^{\hat{D}}) = k\} \end{aligned}$$

3. Using 2 above, we obtain (with a bit more work) the distribution of  $C_1(T_n^{\hat{D}})$

**Sketch of proof of Theorem:**

Let  $\mathcal{L}_n(f) = \{ \text{the cyclic vertices of } G(f) \}$ , where  $f \in \mathcal{M}_n$ .

**Step 1:**  $\mathbf{P}\{X_n(T_n^{\hat{D}}) = k\} = \binom{n}{k} \mathbf{P}\{\mathcal{L}_n(T_n^{\hat{D}}) = [k]\}$ .

**Step 2:** Use the Partition Theorem to compute  $\mathbf{P}\{\mathcal{L}_n(T_n^{\hat{D}}) = [k]\}$ .

To do this, we need to determine

$$\mathbf{P}\{\mathcal{L}_n(T_n^{\hat{D}}) = [k] \mid \hat{D}_i = d_i, 1 \leq i \leq n\} = \frac{|\{f \in \mathcal{M}(d_1, \dots, d_n) : \mathcal{L}_n(f) = [k]\}|}{|\mathcal{M}_n(d_1, d_2, \dots, d_n)|}.$$

**Note:**  $|\mathcal{M}_n(d_1, d_2, \dots, d_n)| = \binom{n}{d_1 \dots d_n}$

To determine  $|\{f \in \mathcal{M}(d_1, \dots, d_n) : \mathcal{L}_n(f) = [k]\}|$

we exploit a **bijection** (!) between  $\mathcal{M}_n(d_1, d_2, \dots, d_n)$  and

$\mathcal{S}_n(d_1, d_2, \dots, d_n) = \{\text{sequences of length } n \text{ with } d_i \text{ occurrences of each } i\}$ .

**Back to**  $T_{n,m}$

Suppose that  $D_1^m, D_2^m, \dots, D_n^m$  are i.i.d.  $Bin(m, p)$  random variables.

Let  $\hat{D}_1^m, \hat{D}_2^m, \dots, \hat{D}_n^m$  be a sequence of random variables with distribution given by

$$\mathbf{P}\{\hat{D}_1^m = d_1, \dots, \hat{D}_n^m = d_n\} = \mathbf{P}\left\{D_1^m = d_1, \dots, D_n^m = d_n \mid \sum_{i=1}^n D_i^m = n\right\}$$

Then

$$T_{n,m} \stackrel{d}{\sim} T_n^{\hat{D}^{(m)}}$$



Using the calculus for random mappings with exchangeable in-degrees, we obtain:

**Theorem 1**

For  $0 < a < b < \infty$  and  $1 < m < \infty$ ,

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ a \leq \frac{X_n(T_{n,m})}{\sqrt{n}} \leq b \right\} = \int_a^b \left( \frac{m-1}{m} \right) x \exp \left( \frac{-(m-1)x^2}{2m} \right) dx$$

**Theorem 2**

For  $0 < c < d < 1$  and  $1 < m < \infty$

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ c \leq \frac{C_1(T_{n,m})}{n} \leq d \right\} = \int_c^d \frac{1}{2\sqrt{1-x}} dx.$$

**Q.** How much more do we have to restrict the in-degrees of a random mapping in order to see ‘transition’ in the component structure of a random mapping?

## Transition model

1. Start with  $n$  blue balls, numbered  $1, 2, \dots, n$  and  $n$  red balls, numbered  $1, 2, \dots, n$ .
2. Remove  $r$  red balls at random, where  $0 \leq r \leq n$  and let  $a = n - r =$  number of red balls remaining in the urn.
3. Remove  $n$  balls, one at a time, at random and without replacement, from the urn and define the random mapping  $\hat{T}_{n,a}$  by

$$\hat{T}_{n,a}(i) = j \quad \text{iff the } i^{\text{th}} \text{ ball removed from the urn is numbered } j.$$

**Note:**  $\hat{T}_{n,0} = \sigma_n$  and  $\hat{T}_{n,n} = T_{n,2}$ .

So the models  $\hat{T}_{n,1}, \hat{T}_{n,2}, \dots, \hat{T}_{n,n-1}$  lie 'between'  $\sigma_n$  and  $T_{n,2}$ .

**Goal:** To understand the (asymptotic) structure of  $\hat{T}_{n,a}$  as a function of  $a$  (and of  $n$ ).

## Good news

The models  $\hat{T}_{n,a}$  are *also* examples of random mappings with exchangeable in-degrees.

## Not so good news

The joint distribution of the in-degree variables  $\hat{D}_{1,n}^a, \hat{D}_{2,n}^a, \dots, \hat{D}_{n,n}^a$  is not so easy to work with - but it is still possible to get exact and asymptotic results.

### Theorem 3

Suppose that  $a \rightarrow \infty$  as  $n \rightarrow \infty$ , then for any  $0 < \alpha < \beta < \infty$

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \alpha \leq \frac{\sqrt{a}}{n+a} X_n(\hat{T}_{n,a}) \leq \beta \right\} = \int_{\alpha}^{\beta} 2x \exp(-x^2) dx.$$

If  $a \geq 1$  is fixed as  $n \rightarrow \infty$ , then for  $0 < \alpha < \beta < 1$ ,

$$\mathbf{P} \left\{ \alpha \leq \frac{X_n(\hat{T}_{n,a})}{n} \leq \beta \right\} = \int_{\alpha}^{\beta} 2ax(1-x^2)^{a-1} dx.$$

**Theorem 4** Suppose that  $a \rightarrow \infty$  as  $n \rightarrow \infty$ , then for any  $0 < \alpha < \beta < 1$

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \alpha \leq \frac{C_1(\hat{T}_{n,a})}{n} \leq \beta \right\} = \int_{\alpha}^{\beta} \frac{1}{2\sqrt{1-x}} dx.$$

If  $a \geq 1$  is fixed as  $n \rightarrow \infty$ , then for  $0 < \alpha < \beta < 1$ ,

$$\mathbf{P} \left\{ \alpha \leq \frac{C_1(\hat{T}_{n,a})}{n} \leq \beta \right\} = \int_{\alpha}^{\beta} \sum_{b=0}^a \binom{2b}{b}^{-1} \binom{a}{b}^2 (2x)^{2b} (1-x)^{2a-2b} dx.$$

**Discussion ...**

Thank you!