DUSTPAN distributions as limit laws for Mahonian statistics on forests

Joshua P. Swanson
University of California, San Diego (UCSD)

Based on joint work with
Sara Billey and *Matjaž Konvalinka*

Papers: [BS20], [BKS20] / arXiv: 2010.12701, 1905.00975
Slides: http://www.math.ucsd.edu/~jswanson/talks/2021_MSU.pdf

Presented at Michigan State University
Combinatorics and Graph TheorySeminar
January 27th, 2021

Outline

- Background: Irwin-Hall Limits
- Hook Length Formulas
- DUSTPAN Distributions
- DUSTPAN Limits
- Further Directions

Background: Irwin-Hall Limits

Def Let $U[a,b]$ denote a uniform continuous random variable supported on $[a,b]$.

The Mth Irwin-Hall distribution is

$$\mathcal{IH}_M = \underbrace{U[0,1] + \cdots + U[0,1]}_{M \text{ i.i.d. r.v.'s}}$$

Notation Write

$$X^* = \frac{X-\mu}{\sigma}$$

for standardized r.v.

Ex By CLT, $\mathcal{IH}_M^* \Rightarrow \mathcal{N}(0,1)$, i.e. $\forall t \in \mathbb{R}$, $\lim_{n\to\infty} \mathbb{P}[\mathcal{IH}_M^* \leq t] = \mathbb{P}[\mathcal{N}(0,1) \leq t]$

Background: Irwin-Hall Limits

$\boxed{\text{Rem}}$ Our motivation in [BKS20] was to study the distribution of the major index on standard Young tableaux, generalizing earlier work on the major index of permutations and words.

$T =$

1	$\underline{2}$	$\underline{4}$	$\underline{7}$	$\underline{9}$	12
3	6	$\underline{10}$			
5	8	11			

$\in SYT\underbrace{(6,3,3)}_{\text{an integer partition}}$ has

$\mathrm{Des}(T) = \{2, 4, 7, 9, 10\}$

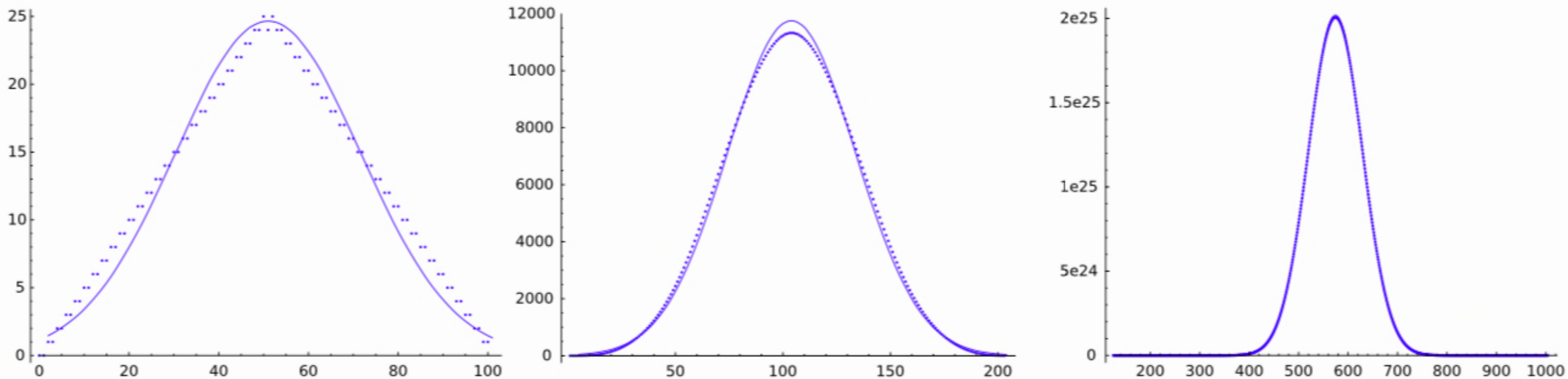
$\mathrm{des}(T) = |\mathrm{Des}(T)| = 5$

$\mathrm{maj}(T) = 2+4+7+9+10 = 32$

Background: Irwin-Hall Limits

Def Each integer partition λ has a random variable $X_\lambda[\mathrm{maj}]$

Ex Distributions of $X_\lambda[\mathrm{maj}]$:



(a) $\lambda = (50, 2)$, $\mathrm{aft}(\lambda) = 2$ (b) $\lambda = (50, 3, 1)$, $\mathrm{aft}(\lambda) = 4$ (c) $\lambda = (8, 8, 7, 6, 5, 5, 5, 2, 2)$, $\mathrm{aft}(\lambda) = 39$

Fig. 1. Plots of $\#\{T \in \mathrm{SYT}(\lambda) : \mathrm{maj}(T) = k\}$ as a function of k for three partitions λ, overlaid with scaled Gaussian approximations using the same mean and variance.

Background: Irwin-Hall Limits

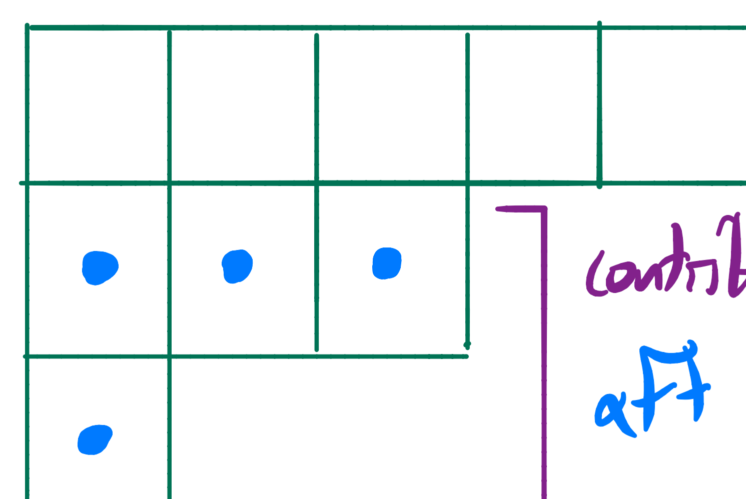
Note • $\chi_{(50,2)}[\mathrm{maj}]$ "looks like" $IH_2 = U[0,1] + U[0,1]$

• $\chi_{(8,8,7,6,5,5,5,2,2)}[\mathrm{maj}]$ "looks like" $N(0,1)$!

Def Given a partition $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k) \vdash n$, let

$$\mathrm{aft}(\lambda) = n - \max\{\lambda_1, k\}.$$

Ex $\lambda = (5,3,1) \Rightarrow \mathrm{aft}(\lambda) = 4$



contribute to aft

Rem On FindStat as St001214; "soon" on OEIS as A338621

Background: Irwin–Hall Limits

Thm (Billey–Konvalinka–S. [BKS20, Thm. 1.7])

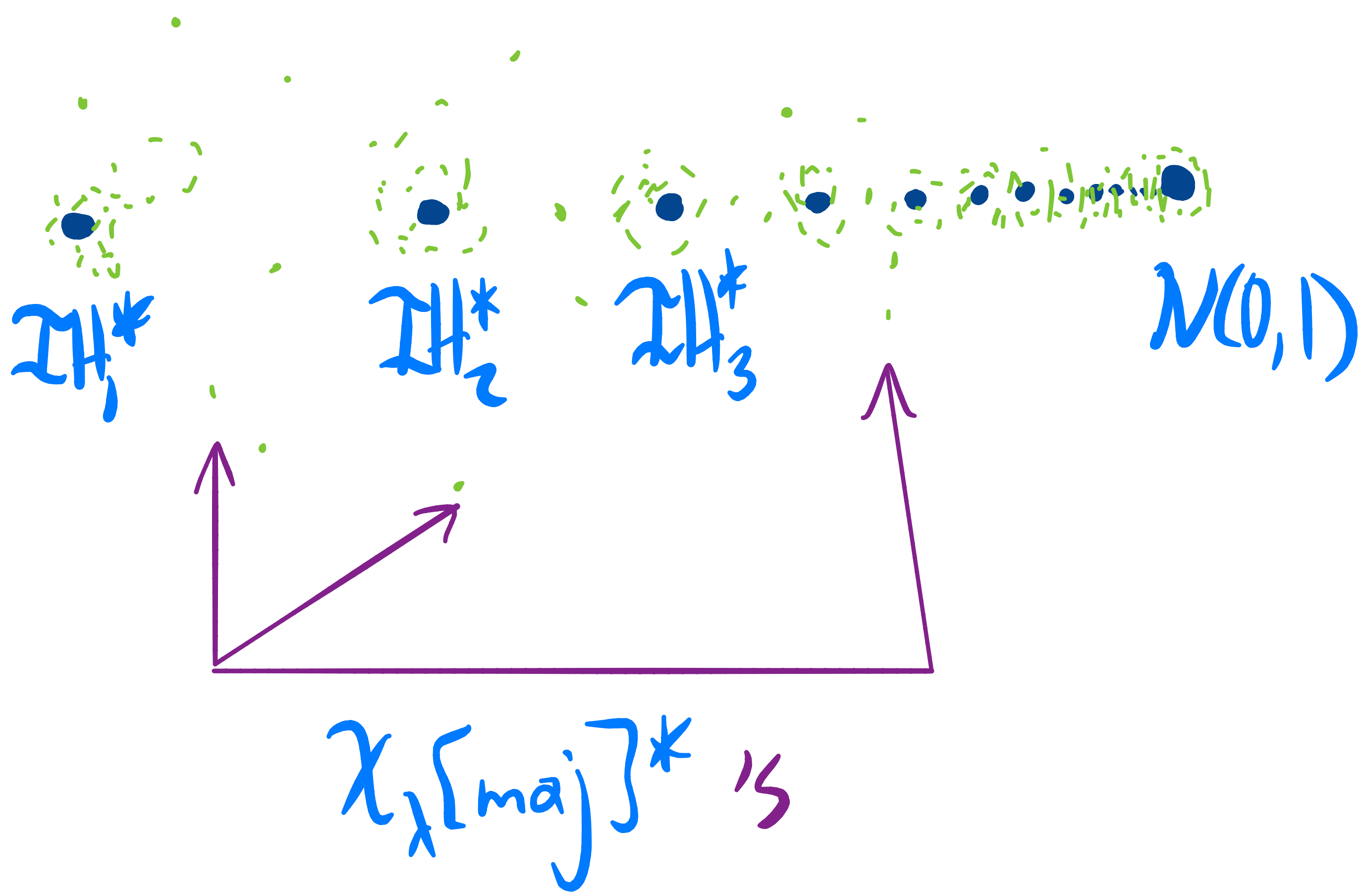
Suppose $\lambda^{(1)}, \lambda^{(2)}, \ldots$ is a sequence of partitions. Then $\mathcal{X}_{\lambda^{(1)}}[\mathrm{maj}]^*, \mathcal{X}_{\lambda^{(2)}}[\mathrm{maj}]^*, \ldots$ converges in distribution if and only if

- (i) $\mathrm{aft}(\lambda^{(N)}) \to \infty$; or
- (ii) $|\lambda^{(N)}| \to \infty$ and $\mathrm{aft}(\lambda^{(N)}) \to M < \infty$; or
- (iii) the distribution of $\mathcal{X}_{\lambda^{(N)}}[\mathrm{maj}]^*$ is eventually constant.

The limit law is $\mathcal{N}$ in case (i), $\mathcal{IH}_M^*$ in case (ii), and discrete in case (iii).

Background: Irwin-Hall Limits

Idea The moduli space of SYT distributions under (say) the Lévy metric:



$M_{SYT} = \{\mathcal{X}_\lambda[\mathrm{maj}]^*\}$

$M_{IH} = \{\mathcal{IH}_M^*\}$

$\Rightarrow \overline{M_{IH}} = M_{IH} \sqcup \{\mathcal{N}(0,1)\}$

Cor $\overline{M_{SYT}} = M_{SYT} \sqcup \underbrace{\overline{M_{IH}}}_{\text{the set of limit points}}$

Hook Length Formulas

Rem Proof in [BKS20, Thm. 1.7] relies on Stanley's q-hook length formula:

$$\sum_{T \in SYT(\lambda)} q^{maj(T)} = q^{r(\lambda)} \frac{[n]_q!}{\prod_{c \in \lambda} [h_c]_q}$$

ratio is key to cumulant formula! then method of moments

Q What other combinatorial statistics arise as quotients of q-integers? (Cyclotomic generating functions)

Recall $[n]_q = \frac{1-q^n}{1-q} = 1 + q + \cdots + q^{n-1}$

Hook Length Formulas

Thm The rank on semistandard tableaux of shape λ and entries $\leq m$ is

$$\sum_{T\in SSYT_{\leq m}(\lambda)} q^{\mathrm{rank}(T)} = \underbrace{q^{r(\lambda)}\prod_{u\in\lambda}\frac{[m+c_u]_q}{[h_u]_q}}_{\text{Stanley's q-hook-content formula}} = \underbrace{q^{r(\lambda)}\prod_{1\leq i<j\leq m}\frac{[\lambda_i-\lambda_j+j-i]_q}{[j-i]_q}}_{\text{q-Weyl dimension formula (type A)}}.$$

Thm The size on plane partitions in an $a\times b\times c$ box is

$$\left.\sum_{P\in PP(a\times b\times c)} q^{\mathrm{size}(P)} = \prod_{i=1}^{a}\prod_{j=1}^{b}\prod_{k=1}^{c}\frac{[i+j+k-1]_q}{[i+j+k-2]_q}\right] \text{MacMahon}$$

Hook Length Formulas

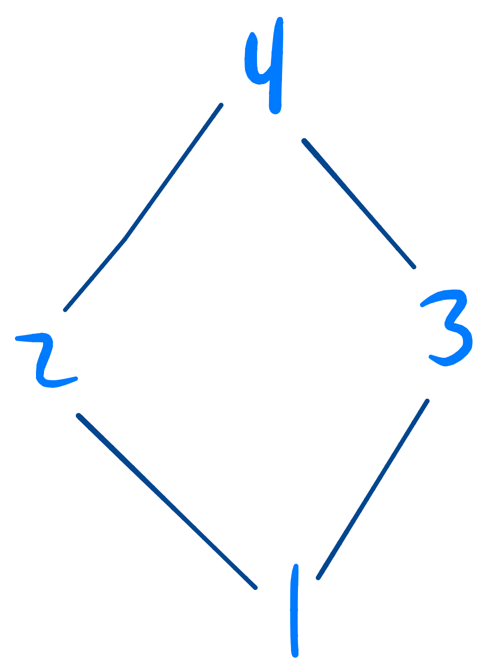
Def Let P be a **Forest** viewed as a poset with roots as maximal elements. Fix an order-preserving bijection $w: P \to [n]$.

Let $\mathcal{L}(P,w) = \{$linear extensions of P, viewed as permutations of $[n]$ via $w\}$.

The **inversion number** of $\pi \in S_n$ is $\mathrm{inv}(\pi) = \#\{(i,j) : 1 \le i < j \le n,\ \pi(i) > \pi(j)\}$.

Ex

$(P,w) =$



$\mathcal{L}(P,w) = \{1234, 1324\} \Rightarrow$ inv's $= \{0,1\}$

Hook Length Formulas

Thm (Stanley, Björner–Wachs)

$$\sum_{\pi \in \mathcal{L}(P,\omega)} q^{\mathrm{inv}(\pi)} = \frac{[n]_q!}{\prod_{u \in P} [h_u]_q}.$$

Q What does the moduli space of forest distributions

$$M_{Forest} = \{ \chi_P^{*}[\mathrm{maj}] \}$$

look like?

DUSTPAN Distributions

Def The rank of P is the length of a maximal chain.

Thm ("Generic" case) Given a sequence of forests P,

$$\mathcal{X}_P[\mathrm{inv}]^* \Rightarrow \mathcal{N}(0,1)$$

if

$$|P| \to \infty \quad \text{and} \quad \limsup \frac{\mathrm{rank}(P)}{|P|} < 1.$$

Idea $\mathcal{N}(0,1)$

DUSTPAN Distributions

Let $n = |P|$ and $r = \text{rank}(P)$.

$\boxed{\text{Q}}$ Consider the set of rooted, unlabeled forests with n vertices, sampled uniformly at random. What is the expected value of the rank r? How does r compare to n as $n \to \infty$?

$\underline{\boxed{\text{Rem}}}$ Brantin-Flajolet proved $\mathbb{E}[r] \sim C\sqrt{n}$ for binary trees.

Typically $\mathbb{E}[r] \sim D\sqrt{n}$ for ordered/labeled variations.

(Certainly $\limsup \frac{r}{n} < 1$ is "typical"!)

DUSTPAN Distributions

Rem The "degenerate" case $n \sim r$ hides a world of behavior!!

Def Let $\ell_2 = \{(t_1, t_2, \dots) : t_i \geq 0\}$ and

$\tilde{\ell}_2 = \{(t_1, t_2, \dots) : t_1 \geq t_2 \geq \dots \geq 0\}$

be sequence spaces.

Let $t \in \tilde{\ell}_2$. The corresponding generalized uniform random variable is

$$S_t = \sum_i U\left[-\frac{t_i}{2}, \frac{t_i}{2}\right]$$

works for $\mathrm{supp}(t) > 0$ iff $\|t\|_2^2 = \sum_i t_i^2 < \infty$

DUSTPAN Distributions

Def A DUSTPAN Distribution is a distribution associated to a uniform sum for t plus a normal distribution

$$S_t + \mathcal{N}(0,\sigma^2) \qquad \text{where } t \in \tilde{\ell}_2,\ \sigma \in \mathbb{R}_{\geq 0}.$$

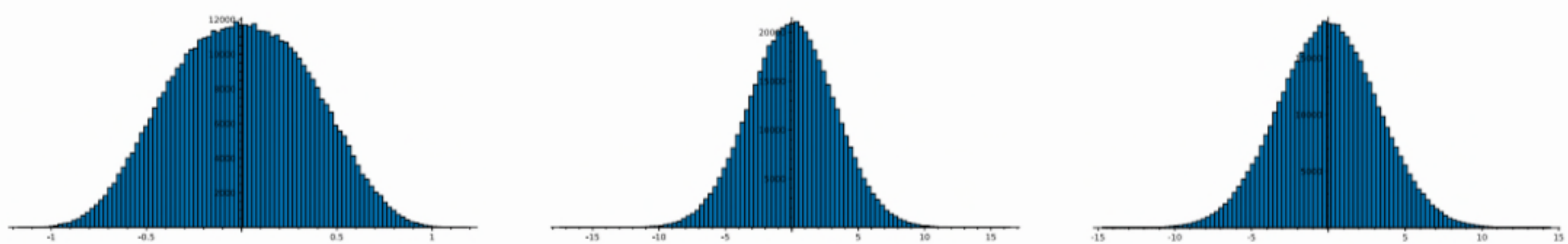


FIGURE 5. Histograms of the distributions $\mathcal{S}_{\mathbf{t}}$, $\mathcal{N}(0,\sigma)$, and $\mathcal{S}_{\mathbf{t}} + \mathcal{N}(0,\sigma)$ with $\mathbf{t} = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8})$ and $\sigma \approx 3.22$.

DUSTPAN Distributions

Def The moduli space of standardized DUSTPAN distributions is

$$M_{DUST} = \left\{ S_t + N(0,\sigma^2) : \frac{\|t\|_2^2}{12} + \sigma^2 = 1 \right\}.$$

The standardized DUSTPAN parameter space is

$$P_{DUST} = \left\{ t \in \tilde{\ell}_2 : \|t\|_2^2 \le 12 \right\}.$$

metric space under pointwise convergence

DUSTPAN Distributions

Thm The map

$$\Phi : P_{DUST} \to M_{DUST}$$

where

$$\Phi(t) = S_t + \mathcal{N}(0, \sigma^2) \quad \text{where } \sigma = \sqrt{1 - \|t\|_2^2/12}$$

is a homeomorphism.

Cor M_{DUST} is closed under convergence in distribution!

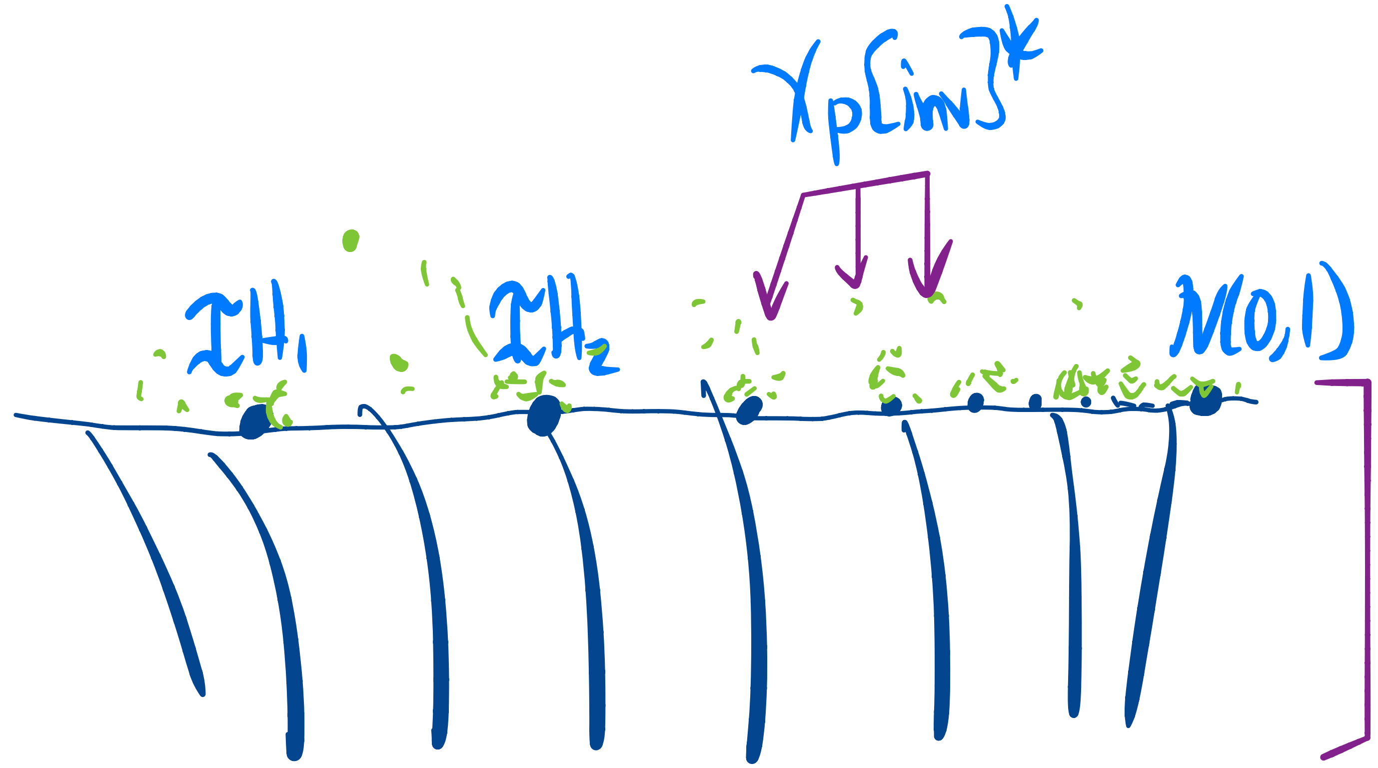
DUSTPAN Limits

Thm [BS20, Thm. 1.13] Let P be an infinite sequence of standardized trees with $n-r=o(n^{1/2})$. Then $X_P^*[\mathrm{inv}]$ converges in distribution if and only if the elevation multisets $\hat{e}$ associated to P converge pointwise to some $t \in \tilde{\ell}_2$.

In that case, the limit distribution is $S_t + \mathcal{N}(0,\sigma^2)$ where $\frac{\|t\|_2^2}{12} + \sigma^2 = 1$.

DUSTPAN Limits

Idea



M_{DUST} is a large collection of limit points!

Cor For any fixed $\varepsilon > 0$, let εTREE be the set of standardized trees for which $n - r < n^{\frac{1}{2} - \varepsilon}$. Let $M_{\varepsilon TREE} = \{X_p^* : p \in \varepsilon TREE\} \subset M_{Forest}$. Then

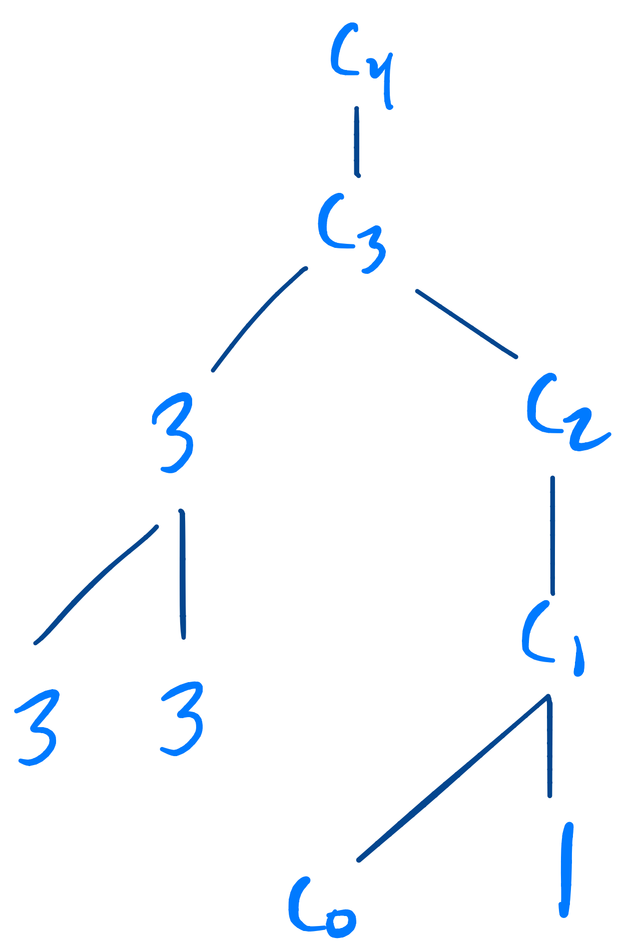
$$\overline{M_{\varepsilon TREE}} = M_{\varepsilon TREE} \sqcup \underbrace{M_{DUST}}_{\text{all limit points!}}.$$

DUSTPAN Limits

"Standardized tree"? "Elevation multiset"?

Def A rooted tree is standardized if its root has at least two children.

Def The elevation multiset of P relative to a fixed maximal chain is:



$$e = \{1,3,3,3\} = (3,3,3,1,0,0,\dots)$$

Rem Normalize via

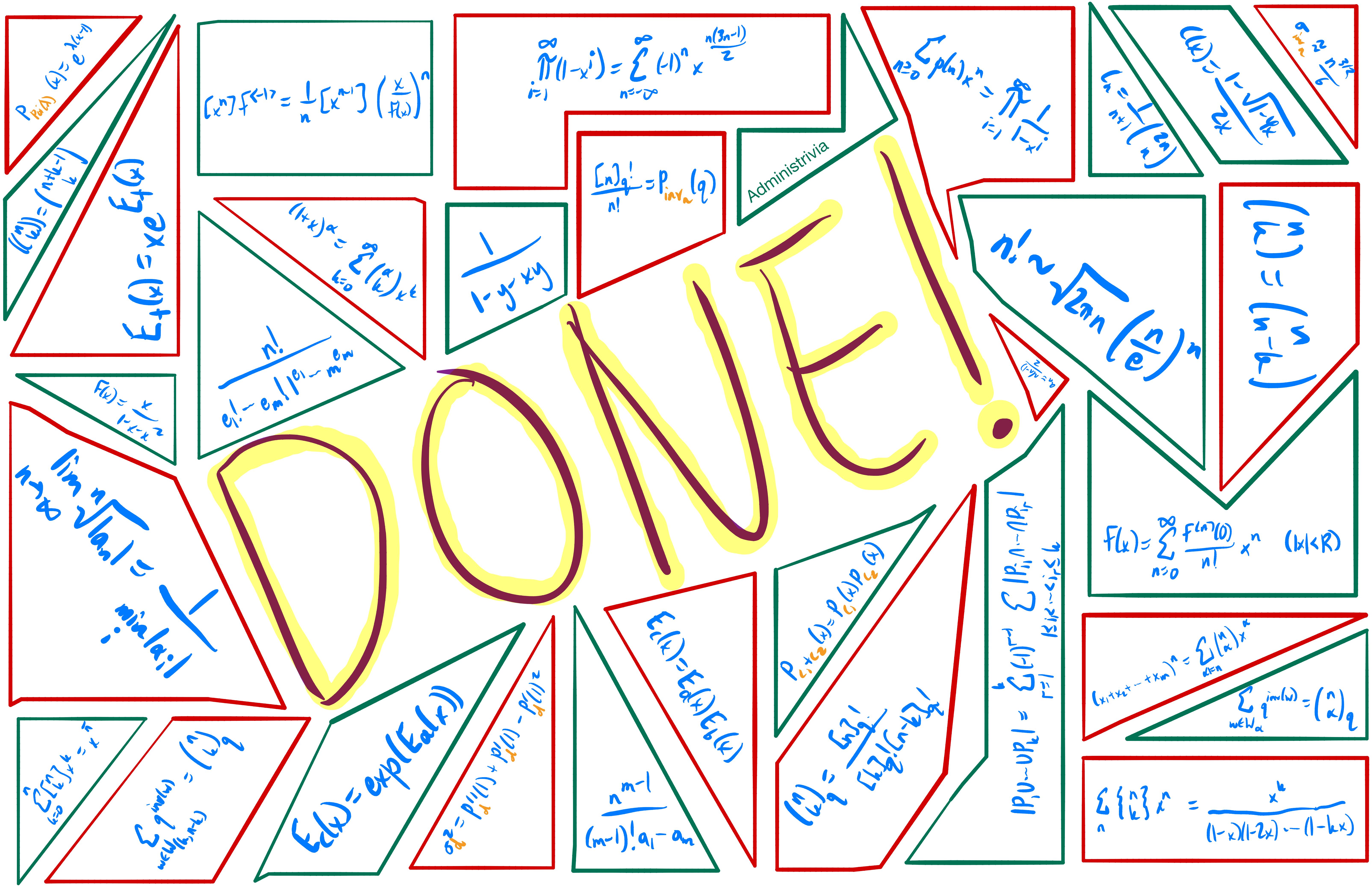
$$\hat{e} = \frac{\sqrt{12}\cdot e}{\|e\|_2}$$

Further Directions

Rem Have related results for...

- rank on $SSYT_{S_m}(\lambda)$: complicated subset of $M_{DUSTPAN}$
- size on $PP_{a\times b\times c}$: only $M_{\mathcal{I}H}$

Q
- What about between $n-r=\omega(n^{11/12})$ and $n-r=o(n^{1/2})$ regimes for M_{Forest}?
- What about M_{CGF}??
- More applications of DUSTPAN's?


DONE!
Administrivia