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_Texas.pdf

Presented at CombinaTexas 2021
February 20th, 2021

Outline

- Background: Irwin-Hall Limits, maj on $SYT(\lambda)$
- Mahonian statistics and forests
- DUSTPAN distributions
- DUSTPAN limits

Background: Irwin-Hall Limits

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

Notation Write

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

for standardized r.v.

The Mth Irwin-Hall distribution is

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

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

Background: Irwin-Hall Limits

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 by Mann-Whitney, Chen-Wang-Wang, Canfield-Janson-Zeilberger, and Diaconis.

$T =$

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

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

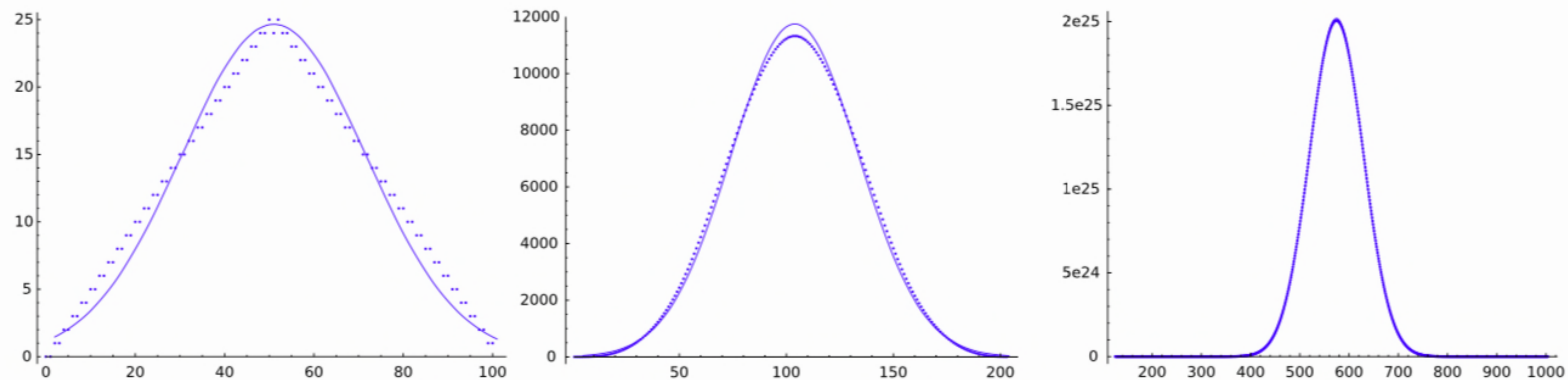
$\mathrm{des}(T) = |\mathrm{Des}(T)| = 5$

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

Background: Irwin-Hall Limits

Def Each integer partition λ has a random variable $X_\lambda[\mathrm{maj}]$ (uniformly sampled).

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

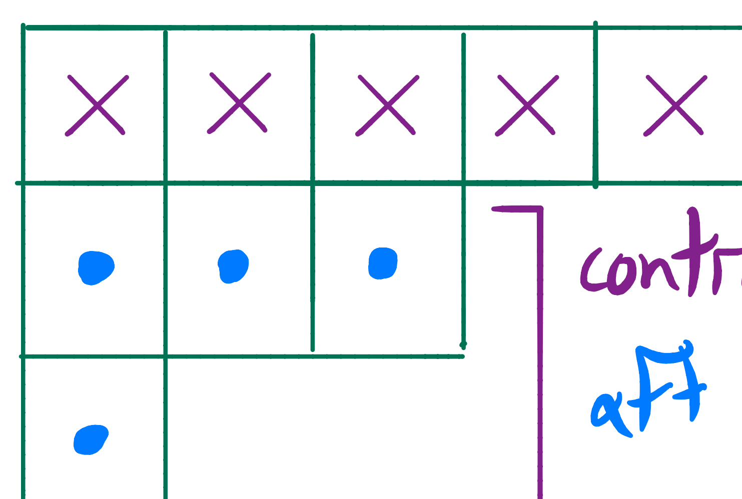
- $\chi_{(50,2)}[\mathrm{maj}]$ "looks like" $IH_2 = \mathcal{U}[0,1] + \mathcal{U}[0,1]$
- $\chi_{(8,8,7,6,5,5,5,2,2)}[\mathrm{maj}]$ "looks like" $\mathcal{N}(0,1)$!

$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k) \vdash n$

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

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

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



contribute to aft

Rem

FindStat: St001214

OEIS: A338621

Background: Irwin-Hall Limits

Thm [BKS20, Thm. 1.7]

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

(i) $\mathrm{aft}(\lambda^{(N)}) \to \infty$; or

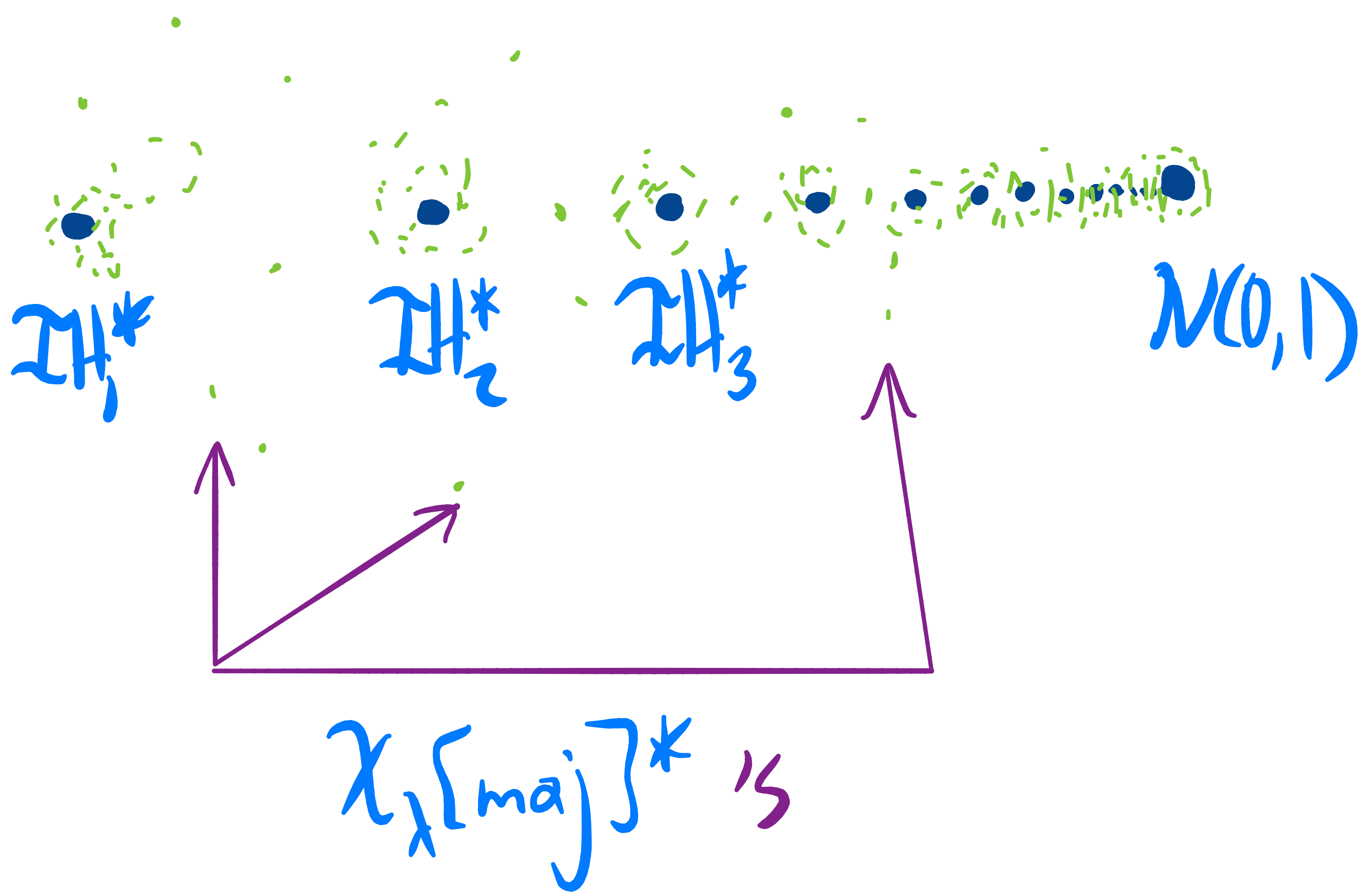
(ii) $|\lambda^{(N)}| \to \infty$ and $\mathrm{aft}(\lambda^{(N)}) \to M < \infty$; or

(iii) the distribution of $\chi_{\lambda^{(N)}}[\mathrm{maj}]^*$ is eventually constant.

The limit law is $\mathcal{N}(0,1)$ 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:



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

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

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

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

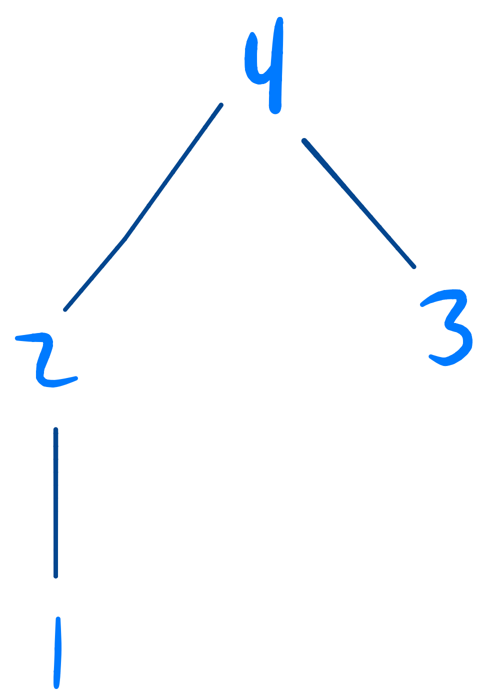
Forests

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

$(P,w) =$ 

$\mathcal{L}(P,w) = \{1234, 1324, 3124\} \Rightarrow \mathrm{inv}\text{'s} = \{0, 1, 2\}$

Forests

Q| What does the moduli space of forest distributions

$$M_{Forest} = \{ \chi_p[inv]^* \}$$

look like?

Q| What are the limit points of M_{Forest}?

DUSTPAN Distributions

Def Let $\ell_2 = \{ t = (t_1, t_2, \dots) : t_i \geq 0,\ \|t\|_2 < \infty \}$ be sequence space

(where $\|t\|_2 = \left(\sum_i t_i^2\right)^{\frac{1}{2}}$).

Let $\tilde{\ell}_2 = \{ (t_1, t_2, \dots) \in \ell_2 : t_1 \geq t_2 \geq \dots \geq 0 \}$ be a space of countable multisets.

Def The generalized uniform sum random variable for $t \in \tilde{\ell}_2$ is

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

requires $\|t\|_2 < \infty$ by Kolmogorov

DUSTPAN Distributions

Def A DUSTPAN distribution is

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

This is a distribution associated to a uniform sum for t plus a normal distribution.

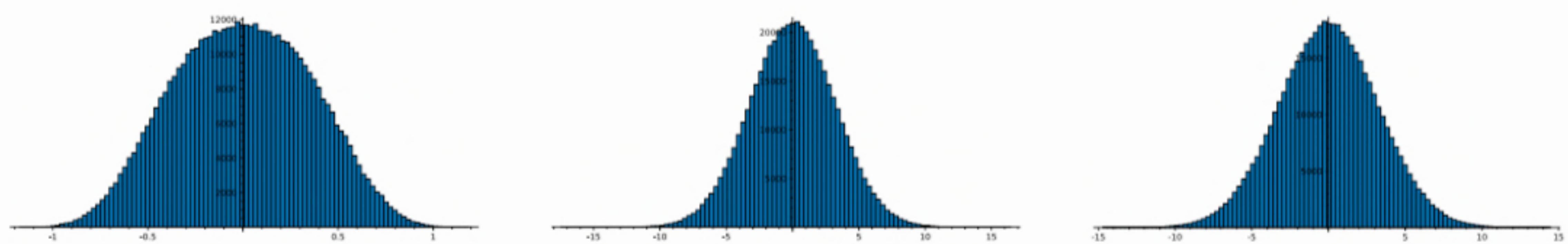


FIGURE 5. Histograms of the distributions $\mathcal{S}_{\mathbf{t}}$, $\mathcal{N}(0,\sigma)$, and $\mathcal{S}_{\mathbf{t}} + \mathcal{N}(0,\sigma)$ with $\mathbf{t} = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8})$ and $\sigma \approx 3.22$.

DUSTPAN Distributions

Def The moduli space of standardized DUSTPAN distributions is

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

Lévy metric

The standardized DUSTPAN parameter space is

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

metric space characterized by pointwise convergence

DUSTPAN Distributions

Thm [BS20] The map

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

where

$$\Phi(f) = S_f + \mathcal{N}\left(0, \sqrt{1 - \frac{\|f\|_2^2}{12}}\right)$$

is a homeomorphism.

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

Rem Thus M_{DUST} is "canonical": closure of $\{\mathcal{U}[a,b]^*\}$ under (standardized) sums and limits!

DUSTPAN Limits

Let $n = |P|$ and let $r = \mathrm{rank}(P)$ be the maximal length of a chain in P.

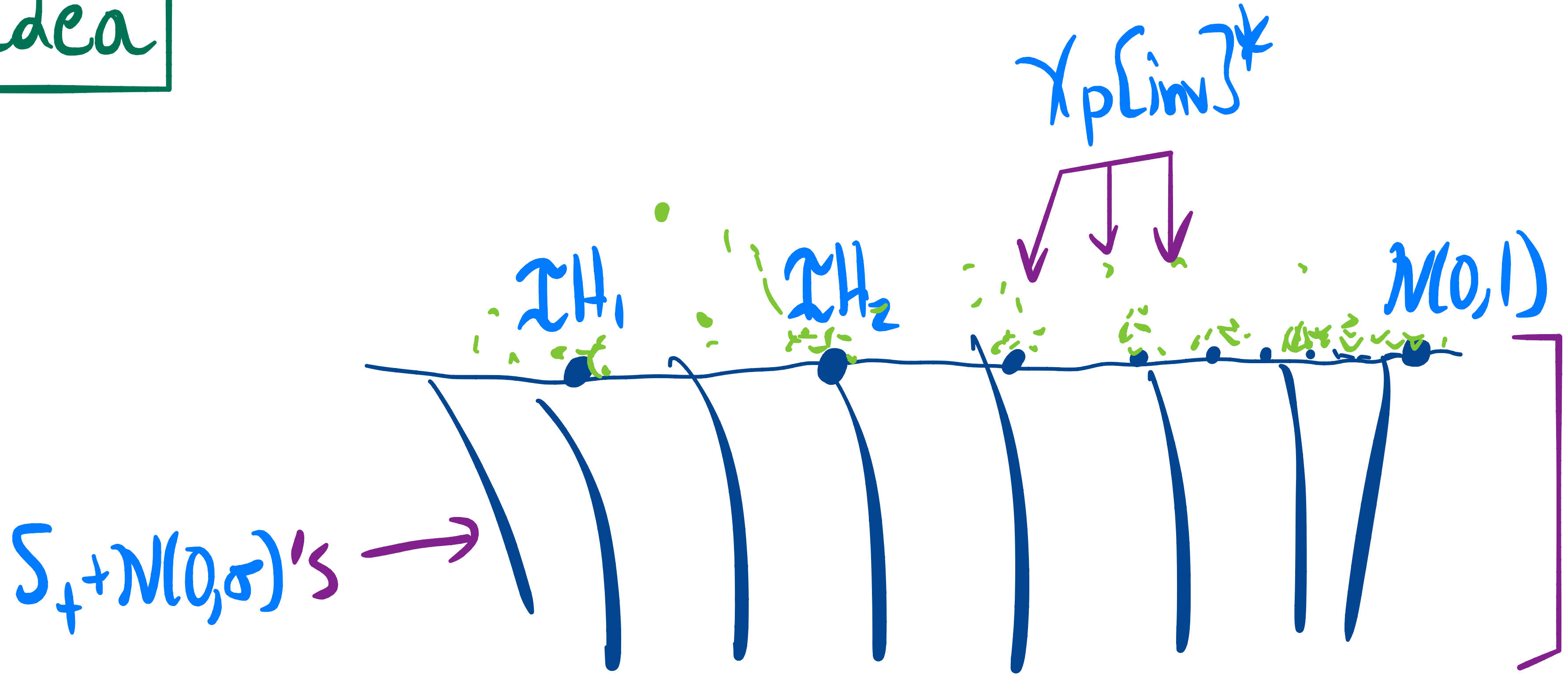
$\boxed{\text{Thm}}$ [BS20, Thm. 1.13] Let P be an infinite sequence of standardized trees with $n - r = o(n^{1/2})$. Then $X_P[\mathrm{inv}]^*$ converges in distribution if and only if the elevation multisets $\hat{e}$ associated to P converge pointwise to some $t \in \tilde{\ell}_2$.

In that case, the limit distribution is

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

DUSTPAN Limits

Idea



M_{DUST} is a <u>large</u> 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 $M_{\varepsilon TREE} = \{\chi_p^* : P \in \varepsilon TREE\} \subset M_{Forest}$. Then

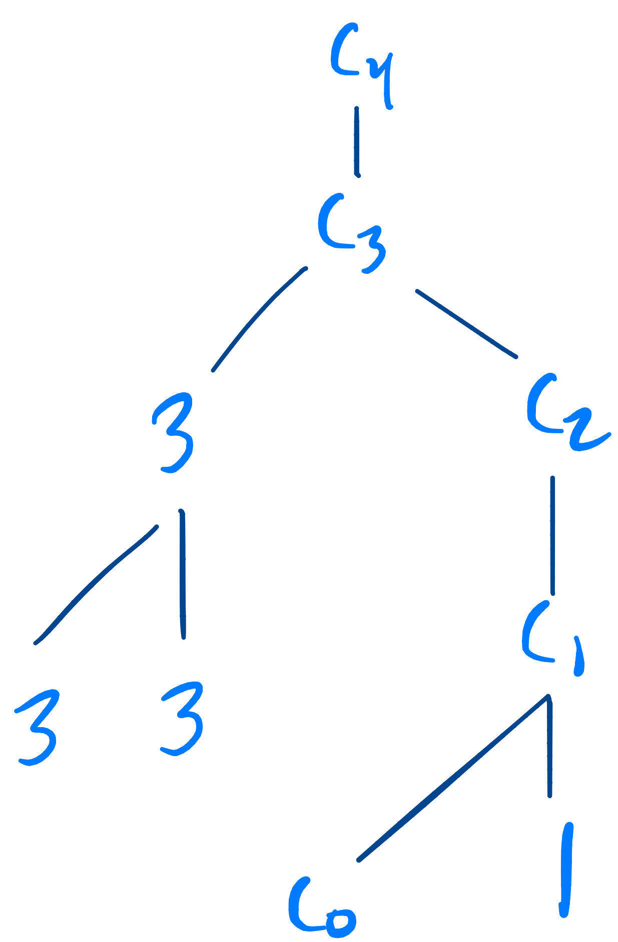
$$\overline{M_{\varepsilon TREE}} = M_{\varepsilon TREE} \sqcup \underbrace{M_{DUST}}_{\text{all limit points!}}.$$

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:



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

Rem Normalize via

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

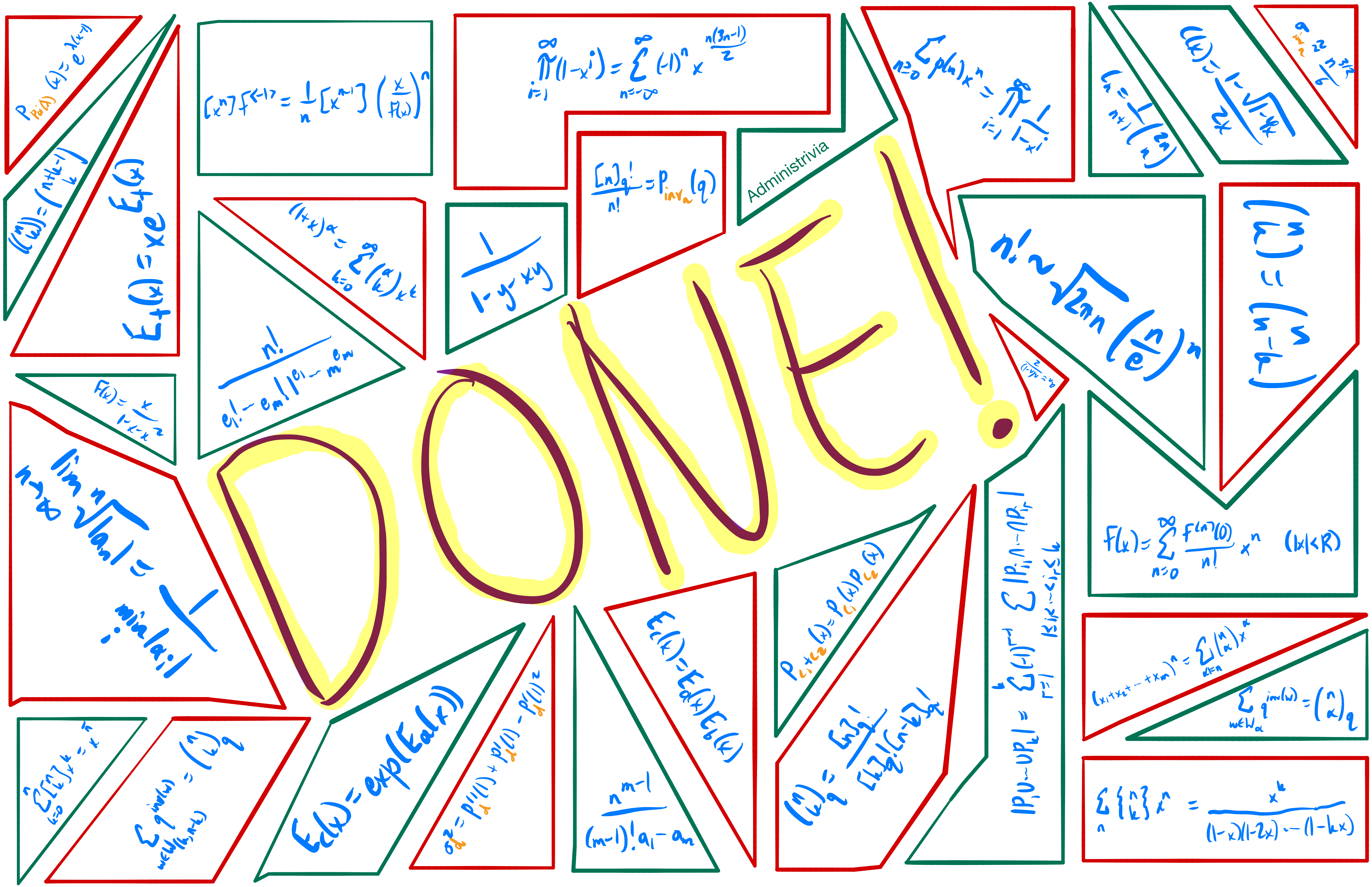
Further Directions

Rem Have related results for

- inv on forests when $n-r=\omega(n^{11/12})$ (generic case)
- rank on $SSYT_{\leq m}(\lambda)$: 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} where CGF means cyclotomic g.f.'s, i.e. q-integer quotients??
- More applications of DUSTPAN's?


DONE!
Administrivia
$P_{Poi(\lambda)}(x) = e^{\lambda(x-1)}$
$[x^n] f^{\langle -1 \rangle} = \frac{1}{n}[x^{n-1}]\left(\frac{x}{f(x)}\right)^n$
$\prod_{i=1}^{\infty}(1-x^i) = \sum_{n=-\infty}^{\infty}(-1)^n x^{\frac{n(3n-1)}{2}}$
$\sum_{n=0} p(n)x^n = \prod_{i=1} \frac{1}{1-x^i}$
$C(x) = \frac{1-\sqrt{1-4x}}{2x}$
$\left(\binom{n}{k}\right) = \binom{n+k-1}{k}$
$E_+(x) = xe^{E_+(x)}$
$(1+x)^\alpha = \sum_{k=0}^{\infty}\binom{\alpha}{k}x^k$
$\frac{1}{1-y-xy}$
$\frac{[n]_q!}{n!} = P_{inv_n}(q)$
$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$
$\frac{n!}{e_1!\cdots e_m!\, 1^{e_1}\cdots m^{e_m}}$
$F(x) = \frac{x}{1-x-x^2}$
$\lim_{n\to\infty} \sqrt[n]{|a_n|} = \frac{1}{\min_i |a_i|}$
$F(x) = \sum_{n=0}^{\infty}\frac{F^{(n)}(0)}{n!}x^n \quad (|x|<R)$
$(x_1+x_2+\cdots+x_m)^n = \sum_{|\alpha|=n}\binom{n}{\alpha}x^\alpha$
$\sum_{w\in W_\alpha} q^{inv(w)} = \binom{n}{\alpha}_q$
$\sum_n \left\{ n \atop k \right\} x^n = \frac{x^k}{(1-x)(1-2x)\cdots(1-kx)}$
$E_c(x) = E_a(x)E_b(x)$
$P_{c_1+c_2}(x) = P_{c_1}(x)P_{c_2}(x)$
$|P_1\cup\cdots\cup P_k| = \sum_{r=1}^{k}(-1)^{r-1}\sum_{1\le i_1<\cdots<i_r\le k}|P_{i_1}\cap\cdots\cap P_{i_r}|$
$\binom{n}{k}_q = \frac{[n]_q!}{[k]_q![n-k]_q!}$
$\frac{n^{m-1}}{(m-1)!\,a_1\cdots a_m}$
$\sigma_d^2 = P_d''(1) + P_d'(1) - P_d'(1)^2$
$E_c(x) = \exp(E_a(x))$
$\sum_{w\in W(k,n-k)} q^{inv(w)} = \binom{n}{k}_q$
$\sum_{k=0}^{n}\left[n \atop k \right] x^k = x^{\bar{n}}$

Generic Forest Limit

Thm [BS20] ("Generic" case) Given a sequence of forests P where $n \to \infty$ and $\limsup \frac{r}{n} < 1$, we have

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

Idea

Expected Rank?

<u>Q</u> 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 $n \to \infty$?

<u>Rem</u> Brautin-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"!))

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

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

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

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

MacMahon