

INV AND MAJ ASYMPTOTICS

JOSHUA P. SWANSON

ABSTRACT. These notes were for a lecture given in the University of Washington probability seminar on March 26th, 2018.

CONTENTS

1. Inversions	1
1.1. Kendall's τ Statistic	1
1.2. Generating Functions	2
2. Inversions and Major Index	3
2.1. Baxter–Zeilberger	3
2.2. Roselle and a Correction Factor	3
2.3. Proof Highlights	4
3. Tableaux Generalizations	5

1. INVERSIONS

1.1. Kendall's τ Statistic.

Question 1.1. Given a joint random variable (X, Y) , are X and Y independent?

Suppose $(x_1, y_1), \dots, (x_n, y_n)$ is a sequence of distinct observations, reordered so that $x_1 < \dots < x_n$. Say $y_{w_1} < \dots < y_{w_n}$. If X and Y are independent, then for each $i < j$, despite knowing $x_i < x_j$, both $y_i < y_j$ and $y_i > y_j$ are equally likely.

Definition 1.2. Let $w = w_1 \cdots w_n \in S_n$ be a permutation in the symmetric group S_n . The *inversion number* of w is

$$\text{inv}(w) := \#\{i < j : w_i > w_j\}.$$

Example 1.3. If $w = \text{id} = 1\ 2\ \cdots\ n$, then $\text{inv}(\text{id}) = 0$. If $w = w_0 = n\ (n-1)\ \cdots\ 1$, then $\text{inv}(w_0) = \binom{n}{2}$.

Note that $\text{inv}(w_1 \cdots w_n)$ is $\#\{i < j : y_i > y_j\}$ or, if not sorting,

$$\#\{i < j : x_i < x_j \text{ and } y_i > y_j, \text{ or } x_i > x_j \text{ and } y_i < y_j\}.$$

Definition 1.4 (Kendall). Set $\tau = 1 - \frac{2\text{inv}(w)}{\binom{n}{2}} \in [-1, 1]$.

If X, Y are independent, then, for example, $\mathbb{P}[\tau = 1] = \mathbb{P}[\text{inv} = 0] = 1/n!$ is extremely unlikely.

Question 1.5. *What is the distribution of inv on S_n , taken uniformly at random?*

Write $[i, j] := \{i, i + 1, \dots, j - 1, j\}$.

Lemma 1.6. *Let*

$$\begin{aligned} \text{inv}_i &: S_n \rightarrow [0, i - 1] \\ \text{inv}_i(w_1 \cdots w_n) &:= \#\{j : i < j \text{ and } w_i > w_j\}. \end{aligned}$$

The map

$$\begin{aligned} S_n &\rightarrow \prod_{i=1}^n [0, n - i] \\ w &\mapsto (\text{inv}_1(w), \dots, \text{inv}_n(w)) \end{aligned}$$

is a bijection.

Corollary 1.7. *inv is the sum of independent (though not identically distributed) discrete uniform random variables with variance $\sigma_n^2 \sim n^3/36$.*

Corollary 1.8. *As $n \rightarrow \infty$, inv , and hence τ , are approximately normally distributed. So, we can do a Z -test on τ to check if X, Y are probably not independent.*

1.2. Generating Functions. The following is a key bridge between combinatorics and probability.

Definition 1.9. Let W be a finite set and suppose $\text{stat}: W \rightarrow \mathbb{Z}_{\geq 0}$. The (ordinary) generating function of stat on W is

$$W^{\text{stat}}(q) := \sum_{w \in W} q^{\text{stat}(w)}.$$

The probability generating function of stat on W taken uniformly at random is then

$$\sum_{i \geq 0} \mathbb{P}[\text{stat} = i] q^i = \frac{W^{\text{stat}}(q)}{\#W}.$$

The characteristic function of stat is then

$$\phi_{\text{stat}}(t) := \frac{W^{\text{stat}}(e^{it})}{\#W}.$$

In this notation, the preceding lemma gives:

Corollary 1.10 (Netto). *We have*

$$S_n^{\text{inv}}(q) = [n]_q! := [n]_q [n - 1]_q \cdots [1]_q$$

where

$$[c]_q := 1 + q + \cdots + q^{c-1} = \frac{1 - q^c}{1 - q}.$$

An enormous amount of work in algebraic combinatorics has gone into finding “nice” expressions for a huge variety of generating functions. We’ll see several more examples in today’s talk.

2. INVERSIONS AND MAJOR INDEX

2.1. Baxter–Zeilberger.

Definition 2.1. The *descent set* of a permutation is given by

$$\begin{aligned} \text{Des}: S_n &\rightarrow 2^{[n]} \\ \text{Des}(w_1 \cdots w_n) &:= \{1 \leq i < n : w_i > w_{i+1}\}. \end{aligned}$$

The *major index* is

$$\text{maj}(w) := \sum_{i \in \text{Des}(w)} i.$$

Example 2.2. $\text{Des}(314652) = \{1, 4, 5\}$, $\text{maj}(314652) = 1 + 4 + 5 = 10$.

Theorem 2.3 (MacMahon). *We have*

$$S_n^{\text{maj}}(q) = [n]_q!.$$

Consequently, inv and maj are equidistributed on S_n .

Definition 2.4. Let

$$X_n := \frac{\text{inv} - \mu_n}{\sigma_n} \quad \text{and} \quad Y_n := \frac{\text{maj} - \mu_n}{\sigma_n}.$$

Theorem 2.5 (Baxter–Zeilberger). *X_n, Y_n are jointly independently asymptotically normally distributed as $n \rightarrow \infty$.*

Baxter–Zeilberger’s proof used the method of (mixed, factorial) moments and combinatorial recursions. Romik asked whether a certain generating function expression of Roselle could instead be used. Zeilberger subsequently offered a \$300 reward for such a proof, which we’ll sketch next. The new proof has the benefit of providing significant intuition into why the result should be true.

2.2. Roselle and a Correction Factor.

Theorem 2.6 (Roselle). *Let*

$$H_n(p, q) := \sum_{w \in S_n} p^{\text{inv}(w)} q^{\text{maj}(w)} = S_n^{(\text{inv}, \text{maj})}(p, q).$$

Then

$$\sum_{n \geq 0} \frac{H_n(p, q)}{(p)_n (q)_n} z^n = \prod_{a, b \geq 0} \frac{1}{1 - p^a q^b z}$$

where $(p)_n := (1 - p)(1 - p^2) \cdots (1 - p^n)$.

Easy manipulations with Roselle’s formula give the following equivalent formulation:

$$(1) \quad \frac{H_n(p, q)}{n!} = \frac{[n]_p! [n]_q!}{n!^2} F_n(p, q)$$

where

$$F_n(p, q) := \frac{n! \cdot \{z^n\} \left(\prod_{a, b \geq 0} (1 - p^a q^b z)^{-1} \right)}{[(1 - p)(1 - q)]^{-n}}.$$

From Netto’s formula, MacMahon’s formula, and Lévy continuity, we have

$$\phi_{X_n}(s) = \phi_{Y_n}(s) = e^{-i\mu_n s / \sigma_n} \frac{[n]_{e^{is/\sigma_n}}!}{n!} \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} e^{-s^2/2}.$$

Also by Lévy, Baxter–Zeilberger’s conclusion is equivalent to

$$\lim_{n \rightarrow \infty} \phi_{(X_n, Y_n)}(s) = \frac{1}{2\pi} e^{-(s^2 + t^2)/2}.$$

Combining (1) and these observations, Baxter–Zeilberger’s result is equivalent to

$$\lim_{n \rightarrow \infty} F_n(e^{is/\sigma_n}, e^{it/\sigma_n}) = 1.$$

The “correction factor” F_n has a nice combinatorial interpretation.

- Note

$$[(1-p)(1-q)]^{-1} = (1+p+p^2+\cdots)(1+q+q^2+\cdots) = \sum_{(a,b) \in \mathbb{Z}_{\geq 0}^2} p^a q^b$$

is the generating function of pairs $(a, b) \in \mathbb{Z}_{\geq 0}^2$.

- Consequently, $[(1-p)(1-q)]^{-n}$ is the generating function for size- n lists of pairs in $\mathbb{Z}_{\geq 0}^2$.
- In a similar way, one may show that $\{z^n\} \left(\prod_{a,b \geq 0} (1 - p^a q^b z)^{-1} \right)$ is the generating function for size- n multisets from $\mathbb{Z}_{\geq 0}^2$.

That is,

$$F_n(p, q) = \frac{n! \cdot \text{g.f. of size-}n \text{ multisets from } \mathbb{Z}_{\geq 0}^2}{\text{g.f. of size-}n \text{ lists from } \mathbb{Z}_{\geq 0}^2}.$$

One may expect the numerator and denominator to agree “to first order”, which provides an intuitive explanation for Baxter–Zeilberger. The rest of the proof essentially makes this intuition precise.

2.3. Proof Highlights.

Definition 2.7. A *partition of n of length k* is a sequence $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ are in $\mathbb{Z}_{\geq 0}$ and $\lambda_1 + \dots + \lambda_k = n$. In this case we write $\lambda \vdash n$ and $\ell(\lambda) = k$.

The *Young diagram* of λ is (draw out).

Theorem 2.8 (S.). *We have*

$$F_n(p, q) = \sum_{d=0}^n [(1-p)(1-q)]^d \sum_{\substack{\mu \vdash n \\ \ell(\mu) = n-d}} \frac{c_\mu}{\prod_{i=1}^{n-d} [\mu_i]_p [\mu_i]_q}$$

for some explicit constants $c_\mu \in \mathbb{Z}$.

Example 2.9. Let $q \rightarrow 1$ in the above. Only the $d = 0$ term survives, which includes only $\mu = (1, \dots, 1) = (1^n)$, giving $F_n(p, 1) = c_{(1^n)} = 1$ and $H_n(p, 1) = [n]_p!$, recovering Netto’s result. Likewise we recover MacMahon’s result.

The proof reinterprets the above list and multiset generating functions and performs a “change of basis” using Möbius inversion on the set partition lattice. Ultimately, we arrive at the following.

Theorem 2.10 (S.). *We have*

$$\lim_{n \rightarrow \infty} |F_n(e^{is/\sigma_n}, e^{it/\sigma_n}) - 1| \leq \lim_{n \rightarrow \infty} 3 \sum_{d=1}^n |st|^d \frac{n^{2d}}{\sigma_n^{2d}} = 0.$$

The proof uses the explicit expression for c_μ along with some estimates related to Stirling numbers of the first kind. Baxter–Zeilberger follows immediately.

3. TABLEAUX GENERALIZATIONS

This section involves joint work with Sara Billey and Matjaž Konvalinka.

Descent sets are an extremely important notion in algebraic combinatorics. The following objects are fundamental to the representation theory of symmetric groups and also have a natural notion of “descent set.”

Definition 3.1. Given a partition λ , a *standard Young tableaux* of shape λ is a filling of the cells of the Young diagram of λ with labels $1, 2, \dots, n$ (each used once) which increase along rows and columns. Write $\text{SYT}(\lambda)$ for the set of standard Young tableaux of shape λ .

Example 3.2. Let $\lambda = (4, 4, 2)$ and $T \in \text{SYT}(\lambda)$ where

$$T = \begin{array}{|c|c|c|c|} \hline 1 & 2 & 5 & 6 \\ \hline 3 & 7 & 8 & 10 \\ \hline 4 & 9 & & \\ \hline \end{array}$$

Definition 3.3. The *descent set* of $T \in \text{SYT}(\lambda)$ is the set of all i where $i + 1$ is in a lower row of T than i . The *major index* of T is again

$$\text{maj}(T) := \sum_{i \in \text{Des}(T)} i.$$

Example 3.4. T above has $\text{Des}(T) = \{2, 3, 6, 8\}$ and $\text{maj}(T) = 2 + 3 + 6 + 8 = 19$.

$\text{SYT}(\lambda)^{\text{maj}}(q)$, that is, the probability generating function of maj on $\text{SYT}(\lambda)$, has surprisingly deep connections to the representation and invariant theory of complex reflection groups. We may naturally ask:

Question 3.5. *What does the distribution of maj on $\text{SYT}(\lambda)$ look like?*

For S_n , as $n \rightarrow \infty$, up to rescaling the answer is a normal distribution. An immediate difficulty here is that there is no one sense in which “ $\lambda \rightarrow \infty$.” Nonetheless, we completely classify the possible (continuous) limiting distributions as follows. Write $X^* := \frac{X - \mu}{\sigma}$.

Definition 3.6. Define a statistic on partitions $\lambda \vdash n$,

$$\text{aft}(\lambda) := n - \max\{\text{first row length, first column length}\}.$$

Example 3.7. (Draw it out.)

Theorem 3.8. *Let $\lambda^{(1)}, \lambda^{(2)}, \dots$ be a sequence of partitions. Then $(X_{\lambda^{(N)}}[\text{maj}]^*)$ converges in distribution if and only if*

- (i) $\text{aft}(\lambda^{(N)}) \rightarrow \infty$; or
- (ii) $|\lambda^{(N)}| \rightarrow \infty$ and $\text{aft}(\lambda^{(N)})$ is eventually constant; or
- (iii) the distribution of $X_{\lambda^{(N)}}^*[\text{maj}]$ is eventually constant.

The limit law is $\mathcal{N}(0, 1)$ in case (i), Σ_M^* in case (ii), and discrete in case (iii). Here Σ_M is the “uniform-sum” or Irwin–Hall distribution.

Example 3.9. Let $\lambda^{(N)} = (N, N)$, so that $\text{aft}(\lambda^{(N)}) = 2N - N = N \rightarrow \infty$ and we are in case (i), asymptotic normality. This recovers a result of Chen–Wang–Wang on coefficients of q -Catalan numbers.

The proof uses Stanley’s q -hook length formula, cumulants, and some combinatorial estimates involving hooks.

Remark 3.10. We have a more general result using certain “skew partitions” which includes the permutation case (where aft is $n - 1 \rightarrow \infty$, as needed) and which generalizes work of Canfield–Janson–Zeilberger and others. We’ve also completely classified the “internal zeros” of $\text{SYT}(\lambda)^{\text{maj}}(q)$ using a combinatorial argument which mutates the descent sets of tableaux in a very controlled manner. This latter result strengthens an earlier result of (S.) answering a conjecture of Sundaram. That earlier result was proven by using representation theory to give a local limit theorem for the statistic maj modulo n on $\text{SYT}(\lambda)$ with a uniform limit.