Number of inversions in multiset permutations

Janosch Ortmann, Zhaleh Safikhani, Petr Smirnov, Ian Smith

December 2018

1 Introduction

The concordance index is a non-parametric metric for comparison between two orderings on a set. In practice, this metric can be used to determine whether a candidate biomarker is informative of a clinical outcome, like sensitivity to anti-cancer drugs.

The best way to think about permuting elements with ties between them is as follows: Let the set of distinct elements to be permuted be denoted $E = \{e_1, ..., e_n\}$ and let $a_j \in \mathbb{Z}_{>0}$ denote the multiplicity of element e_j, that is how often it appears. Thus there are $\alpha := \sum_{j=1}^{n} \alpha_j$ elements in total. We denote by $M = \{e_1^{\alpha_1}, ..., e_n^{\alpha_n}\}$ the multiset containing all elements (with ties).

2 Exact expressions

2.1 Explicit formula

In [?] we have the following result for *sets*, that is $\alpha_j = 1$ for all j (and hence $\alpha = n$): let $I_n(k)$ denote the number of inversions of S with k inversions then

$$\Phi_n(x) := \sum_{j=1}^{\binom{n}{2}} I_n(x) x^k = \prod_{j=1}^{n} \sum_{k=1}^{j-1} x^k. \tag{1}$$

It turns out that an analogous result can be obtained for multisets. The original reference is to a paper from 1915 – see [Mac15] in [?] – but it's easier to read in modern notation. Let $\mathrm{inv}(\sigma)$ denote the number of inversions of a permutation of the multiset (set with ties) S. In [?] the *distribution* of inv is shown to be

$$D_M(x) = \sum_{\sigma \in S_M} x^{\mathrm{inv}(\sigma)} = \begin{bmatrix} \alpha \\ \alpha_1 ... \alpha_n \end{bmatrix}_x = \frac{\alpha!_x}{\alpha_1!_x..\alpha_n!_x} \tag{2}$$

with the *q-factorial* being defined by

$$m!_x = \prod_{k=1}^{r} \left(1 + x + ... + x^{k-1}\right) \tag{3}$$

(The expression on the right-hand side of (**??**) is also called the *q-multinomial coefficient*. Observe that, by splitting the sum over S_M according to the number of inversions,

$$D_M(x) = \sum_{k \geq 0} \sum_{\substack{\sigma \in S_M \\ \mathrm{inv}(\sigma)=k}} x^{\mathrm{inv}\,\sigma} = \sum_{k \geq 0} \sum_{\substack{\sigma \in S_M \\ \mathrm{inv}(\sigma)=k}} x^k = \sum_{k \geq 0} I_M(k) x^k \tag{4}$$

where $I_M(k)$ denotes the number of permutations of the multiset M with k inversions. Thus, (**??**) is the exact multiset analogue of (**??**).

2.2 Recursive formula

The paper [?] also has a recursion formula, expressing $I_n(k)$ in terms of the $I_{n-1}(j)$: in terms of the generating function this reads

$$\Phi_n(x) = \left(\sum_{k=0}^{n-1} x^k\right) \Phi_{n-1}(x). \tag{5}$$

The proof proceeds by looking at permutations of the first $n-1$ elements and then inserting the last element at all possible position. By keeping track of how many extra inversions this insertion introduces, we arrive at (??).

This argument extends rather well to the case with ties: let M be the multiset as described in the introduction and denote by M^- the set obtained from M by removing one occurrence of e_n. That is, if $M = e_1^{\alpha_1}, .., e_n^{\alpha_n}$, then

$$M^- = e_1^{\alpha_1}, e_2^{\alpha_2}, \ldots, e_{n-1}^{\alpha_{n-1}}, e_n^{\alpha_n - 1}, \tag{6}$$

and in particular if $|M| = n$ then $|M^-| = n-1$. We can give the following analogue of (??):

$$D_M(x) = \begin{bmatrix} \alpha \\ \alpha_n \end{bmatrix}_x D_{M-}(x) = \frac{\alpha!_x}{(\alpha - \alpha_n)!_x \alpha_n!_x} D_{M-}(x), \tag{7}$$

with $m!_x$ defined in (??).

3 Gaussian approximation

In [?], asymptotics are also discussed. It seems like there is also a Gaussian approximation result for the inversions in the multiset case, see [?].

References

[1] Conger M and Viswanath D (2006), Permutations of Multisets. arXiv:math/0508242

[2] Margolius BH (2001), Permutations with Inversions. *Journal of Integer Sequences* **4**, Article 01.2.4

[3] Wilson AT (2015), An extension of Macmahon's equidistribution theorem to ordered multiset partitions. arXiv: 1508.06261.