# Patterns in Random Permutations

Chaim Even-Zohar[1]

[1] *University of California, Davis, USA, chaim@math.ucdavis.edu*

The density at which fixed patterns occur in large permutations has received much attention in Combinatorics. Pattern densities give rise to extremal questions, and play a role in the construction of limiting objects for permutations, and in permutation property testing. The case where some patterns are avoided is studied extensively.

We report on the study of pattern densities in *random* permutations. Our work extends the discussion by Janson, Nakamura and Zeilberger in Section 4 of [1]. In particular, we address the question in its closing paragraph, on the emerging general structure. To this end, we analyze the distribution of pattern densities using representations of the symmetric group.

This viewpoint of pattern densities provides a unified framework for several measures from non-parametric statistics, such as Kendall's $\tau$, Spearman's $\rho$ and some two-sample independence tests. It is also related to the spectral analysis of statistical data on nonabelian groups, as introduced by Diaconis [2].

We present some definitions before stating the main questions and results. Let $\pi \in S_n$ and let $k \leq n$. Consider all $\binom{n}{k}$ restrictions of $\pi$ to $k$ entries $\pi_{a_1} \dots \pi_{a_k}$ where $a_1 < a_2 < \dots < a_k$. The relative ordering of such $k$ values induces a *pattern* $\sigma \in S_k$. For example, the restriction of $\pi = \underline{4}\,\underline{1}\,2\,\underline{5}\,3$ to the marked entries induces the 3-pattern $\sigma = 2\,1\,3$.

Let the *density* of $\sigma \in S_k$ be $P_\sigma(\pi) := N_\sigma(\pi)/\binom{n}{k}$, where $N_\sigma(\pi)$ is the number of times $\sigma$ occurs as a $k$-pattern in $\pi$. The *k-profile* of $\pi$ is the $k!$-dimensional vector of all $k$-pattern densities $\mathbf{P}_k(\pi) := (\,P_\sigma(\pi)\,)_{\sigma \in S_k}$. When $\pi \in S_n$ is sampled uniformly at random, we denote its $k$-profile by $\mathbf{P}_{kn}$.

A first observation is that $\mathbf{P}_{kn} \to \mathbf{U}_k := \left(\frac{1}{k!}, \dots, \frac{1}{k!}\right)$ in probability as $n \to \infty$. It is hence interesting to understand how the $k$-profile deviates from this limit. What is the order of magnitude of $(\mathbf{P}_{kn} - \mathbf{U}_k)$ as $n$ grows? What directions in the $k!$-dimensional space are typical of this vector? Does it have a natural decomposition into lower-dimensional components? What is the shape of the distribution when properly normalized?

It turns out that linear representations of $S_k$ provide some answers to these questions. Recall that each simple representation $R^\lambda$ corresponds to an integer partition $k = \lambda_1 + \dots + \lambda_\ell$ where $\lambda_1$ is the largest. Consider the subspace spanned by the matrix elements $(R^\lambda_{ij}(\sigma))_{\sigma \in S_k}$ viewed as $k!$-dimensional vectors.

Orthogonal projections on these subspaces provide an initial decomposition of the $k$-profile. We show that the component that corresponds to $R^{\lambda}$ has order of magnitude $n^{(\lambda_1-k)/2}$ asymptotically as $n$ grows. One can use this decomposition to *normalize* the distribution of the profile, multiplying the different components by the appropriate powers of $n$.

We also show that components of different orders are asymptotically uncorrelated, in the sense that the cross-covariance matrix of the two normalized vectors converges to zero. This indicates that representations of the symmetric group may also help to *diagonalize* the profile's distribution.

Indeed, for $k \leq 6$ we found specific unitary matrix representations of $S_k$, whose matrix elements diagonalize the normalized distribution of the $k$-profile. This means that its covariance matrix, with respect to that basis, converges to a diagonal with positive entries. We hope to extend this result to every $k$ in future work.

The above results were discovered by computer exploration. Our starting point was the interpolation of the profile's covariance matrix, symbolically as rational functions of $n$. This allowed us to extract several leading coefficients that determined the asymptotic behavior, and to look at their diagonal forms.

The full analysis and verification of the cases $k = 3, 4, 5, 6$ were undertaken by explicit computation of appropriate unitary representations, that seem to have interesting properties by their own.

# References

[1] S. Janson, B. Nakamura and D. Zeilberger, *On the asymptotic statistics of the number of occurrences of multiple permutation patterns*, J. Comb. **6**, pp. 117-143 (2015).

[2] P. Diaconis, *Group representations in probability and statistics*, Inst. Math. Stat. Hayward CA (1988).