Report on the thesis
Asymptotic properties of Robinson–Schensted–Knuth algorithm and jeu de taquin
by Łukasz Maślanka

Prof. dr hab. Paweł Hitczenko
Department of Mathematics, Drexel University

1. Topics

The topics covered in the thesis concern analysis of certain features of randomly growing (in a sense described below) Young tableaux. While Young tableaux traditionally have been of interest in algebraic combinatorics and representation theory of symmetric group, their connections to other branches of mathematics, including probability theory, have been growing recently. One, very well–known and classical by now example, concerns the length of the longest increasing sequence in a random permutation, which corresponds to the length of the bottom row (in the convention of this thesis) of Young tableau in the model considered here. Another, more recent and not fully explored in my view, connections are with exclusion processes, an area of probability theory related to randomly growing surfaces that is motivated by theoretical physics and has been experiencing a spectacular growth in the past decade or so. Some of those connections are mentioned in the last chapter of this thesis. All of this is to say that the topic of the thesis lies at the interface of several areas of contemporary mathematics and is of current relevance and interest to researchers working in these areas.

2. Content

Original results are presented in Chapters 2, 3 and 4 of the thesis. They are preceded by Chapter 1 of introductory nature that provides some background and sets the stage for the subsequent developments. The results of Chapters 2–4 are based on the following works, respectively (the first lists 'Marciniak' twice at the beginning of Chapter 2, which I believe is an error):

> M. Marciniak, Ł. Maślanka and P. Śniady *Poisson limit theorems for the Robinson–Schensted correspondence and for the multi-line Hammersley process.*
>
> M. Marciniak, Ł. Maślanka, P. Śniady *Poisson limit of bumping routes in the Robinson-Schensted correspondence.* Probability Theory and Related Fields **181** (2021), no. 4, 1053–1103.
>
> Ł. Maślanka and P. Śniady *Second class particles and limit shapes of evacuation and sliding paths for random tableaux.*

As far as I have been able to determine, the first and the third item have not been published yet but, as is stated in the thesis, have been available on `arxiv.org`. In addition, a 12 page extended abstract of the latter manuscript has been published (under a different title) in the proceedings of the 32nd International Conference on Formal Power Series and Algebraic Combinatorics (the precise reference is: Ł. Maślanka, P. Śniady, *Limit shapes of evacuation and jeu de taquin paths in random square tableaux.* Sém. Lothar. Combin. **84B** (2020), Art. 8, 12 pp.).

While there are no solo papers by the candidate in this list, his contributions to each of them are clearly described in the candidate's and his PhD advisor's (Piotr Śniady) separate statements and based on them, these contributions are significant.

The thesis is concerned with the analysis of dynamics of various parameters of randomly growing Young tableaux. By randomly growing it is meant, as is customary in this context, applying deterministic insertion algorithm, namely Robinson–Schensted–Knuth (RSK) algorithm to random input. More specifically, one generates an infinite sequence (w_k) of independent, identically distributed (iid) random variables with uniform distribution on $[0,1]$ and repeatedly applies Schensted insertion to the elements of this sequence. There is no point in giving details here as they are explained in Section 1.2 of the thesis. Suffices to say that exectuing the insertion on the first n elements of the sequence (w_k) gives a pair of Young tableaux each having n boxes and the same shape and that an insertion of any element results in either extending one of the existing rows by one or adding a new row at the top of the tableaux, subject to preserving the requirements of Young tableau. These requirements are that boxes of a tableau are arranged in rows of non–increasing length (measured in number of boxes) stacked upon one another and left–aligned (something that is missing in the description given in Section 1.1!). The first n boxes are filled with the entries $w_1, \ldots, w_n$ according to the rules of Schensted insertion (these rules imply that the entries end up being increasing within rows and columns as we move in a positive direction along the axes of standard coordinate system). This is referred to as the insertion tableau while a companion tableau, called the recording tableau has the same shape (i.e number and lengths of rows) and entries $1, \ldots, n$ reflecing a moment in the process when a given box was created. (All of this is explained and accompanied by examples and pictures in Chapter 1.) Applying RSK to iid uniform input $(w_k)_{k \le n}$ results in random Young tableau having n boxes. The resulting probability measure on the set of such tableaux (considered as a discrete probability space) is referred to as the Plancherel measure.

Under this model the thesis studies three distinct features of the evolution of Young tableaux. In Chapter 2, loosely, the following question is considered: fix $k \ge 1$, fix (for a moment) n and consider a Young tableau at the instance n. Now, consider the k dimensional stochastic process:

$$(X_t^{(n)}) := \left(\lambda_1(n+t\sqrt{n}) - \lambda_1(n), \lambda_2(n+t\sqrt{n}) - \lambda_2(n), \ldots, \lambda_k(n+t\sqrt{n}) - \lambda_k(n)\right),$$

where $\lambda_j(u)$ is the number of boxes in the jth row after, say, $\lfloor u \rfloor$ boxes have been inserted in the tableau. (This is just the increase in length in each of the bottom k rows in the time window $[n, n+t\sqrt{n}]$.)

Now, let $n \to \infty$. The main result of Chapter 2 (Theorem 2.2.1) asserts that *all finite dimensional distributions* of $(X_t^{(n)})$ converge in law to the corresponding finite dimensional distributions of the process

$$(N_1(t), \ldots N_k(t)),$$

where $N_j(t)$, $1 \le j \le k$, are k independent copies of the standard Poisson process. Several variants and related results are discussed as well.

This is certainly an interesting and significant contribution to theory, particularly given the fact, that the *total* lengths of the bottom rows is governed by a different distribution, so being able to identify the distribution of the bottom rows at their fringes is quite remarkable. While this result is perhaps not entirely unexpected as they have been indications of such behavior in work of Aldous and Diaconis ([AD95] in the reference list), it is still a solid step forward and, in particular, an extension in some sense of [AD95] who studied just the bottom row. I write 'in some sense' as it appears that one point is missed here. This is actually the main (and really the only) weakness of the thesis, so let me explain: the author too often *does not* make a clear distinction between the notion of convergence of finite dimensional distributions for stochastic processes and the notion of convergence in distribution for stochastic processes, which is unfortunate and occasionally may lead to confusion. For example, while the author explains in the statement of Theorem 2.1.1 that by 'convergence in distribution' he means 'convergence of finite dimensional distributions' this is absent in the statements of Corollaries 2.1.2, 2.1.3 (where, I believe, the author means the convergence of finite dimensional distributions). In particular, as written, Section 2.1.8.2 discussing connections to [AD95] is not convincing to me. Specifically, as far as I can tell, the argument of that section shows that the finitely dimensional distribution of (2.1.17) converge to those of the standard Poisson process while [AD95] proved the actual convergence in distribution. Perhaps it is a good moment to mention that the convergence in distribution for stochastic processes means the weak convergence of their distributions as measures on the function space containing sample paths of stochastic processes under consideration (typically taken to be the space of continuous functions with uniform topology for processes with continuous sample paths or the space D equipped with Skorokhod topology for cadlag processes). Then, the basic criterion for convergence in distribution for a sequence $(Y_t^{(n)})$ to a stochastic process (Y_t) is that

$$(Y_{t_1}^{(n)}, \dots Y_{t_m}^{(n)}) \xrightarrow{d} (Y_{t_1}, \dots Y_{t_m}), \quad m \geq 1, \quad t_1 < \dots t_m$$

(i.e. the final dimensional distributions converge) and that the sequence of distributions $(\mu_{Y^{(n)}})$ of $(Y^{(n)})$ (i.e. measures on the function space containing the sample paths) is tight. Thus, one would need to know that the distributions of (2.1.17) are tight on. a suitable space, which while probably true is never mentioned.

Similar inconsistencies occur in the subsequent chapter, e.g. Theorems 3.1.2 and 3.1.5 refer to convergence in distributions and only in subsequent remarks (3.1.3 and 3.1.6, respectively) it is explained that it is convergence of finite dimensional distributions that is meant. Formally, it is correct, but I wish the author would make the more careful distinction between the two notions. Or, better yet, addressed the issue of tightness; this would make the results of these two chapters stronger, more complete and more in line with what is typically done in probability.

Chapter 3 concerns the limiting shape of the bumping route when an element of moderate value is inserted in the tableau (of infinite size for the purpose of this section). (Roughly, when an element is inserted in a large tableau, it is likely to be put in place of an existing element in the first row; this element is moved ('bumped') to the second row and it may bump an element from that row, and so on until an addition of a new box at the end of one of the rows or creating a new row occurs). The collection of boxes whose entries have been changed is referred to as a bumping route and after suitable scaling converges to a deterministic curve in the plane. In particular, the author is interested in studying the process (Y_x) which is the number of the row at which the bumping route enters one of the x leftmost columns for the first time (by the properties of Schensted insertion the values are non–increasing in x). The main result of this chapter (Theorems 3.1.2, 3.1.5, and 3.1.7) is that suitably transformed process $(Y_x)_{x\geq 0}$ converges (in the sense of finite dimensional distributions) to the standard Poisson process.

Aside from mentioned earlier issue of tightness, this is to me, aesthetically and mathematically a pleasing result and a good introduction to developments in Chapter 4 that are equally appealing and technically much more involved. The results of Chapter 3 extend earlier work of Dan Romik and the candidate's PhD advisor.

The last chapter concerns the shape of the so-called sliding path after one execution of *jeu de taquin* and also of the evacuation path. *Jeu de taquin* refers to the following action on a Young tableau: remove the smallest element from the tableau (necessarily in its bottom left corner), consider the smaller of its north and east neighbor and 'slide' it into the emptied box. Repeat the procedure until a box on the north-west boundary of the tableau is reached and, once empty, remove it. Evacuation path refers to the path taken by the largest element in the tableau when *jeu de taquin* is repeated until this element is removed (obviously the last one).

The main results here (Theorem 4.2.4 and 4.2.3, respectively) that in a square Young tableau after suitable normalization, the sliding path or evacuation path follows a (random) curve from the family of predetermined curves. Extensions to tableaux of (mildly) more general shapes are discussed. As an interesting and valuable addition connections to totally asymmetric exclusion processes (TASEP) are discussed (see Section 4.11 and Theorem 4.1.1).

This chapter is, in my view, the highlight of the thesis. The results are substantial, technically involved, and seem to have required serious development of new machinery. While the candidate's PhD advisor worked with Dan Romik on similar problems in different setting, the methods do not seem transferable and a whole new set of tools needed to be developed. Ability to use these results to draw conclusion about TASEP highlights the relevance of this work. I agree with the PhD advisor's statement that Chapter 4 by itself would be a solid doctoral thesis.

3. Assessment and comments

As indicated in the above comments, presented research is of high quality, impact and relevance. Overall, the candidate demonstrated very good mastery of current methods used in this area. The contribution to advancing the field is substantial. It is worth noting that the results of Chapter 3 were published in an elite journal in probability theory. While the other two papers on which this thesis is based have not been published yet, a presentation based on content of Chapter 4 was accepted as a contributed talk at FPSAC 2020 and the extended abstract was published in the proceedings of that conference (FPSAC is very restrictive and has one of the lowest acceptance rates among conferences in mathematics). This is a good testimony to the quality of work presented here. My somewhat critical comments stated above and below (I am not done yet) are of secondary nature and should not be taken as a criticism of the scientific content of this thesis.

Let me now pass to comments on presentation. The thesis is based on three manuscripts and clearly the author could have taken more care when merging them together and could have showed more attention to detail while doing so. As it is, there are repetitions (including definitions, sections and their titles (for example Sec. 3.1.1 is almost identical to Sec.2.1.1 and Sec. 3.1.2 is verbatim Sec. 2.1.2; similarly, Sec 3.2.2 is essentially Sec. 2.2.2 and Sec. 3.2.1 is contained in Sec 2.2.1).

Consistent references to 'this paper' (throughout Chapters 2–4 of the thesis) and to 'the second named author' (Chapter 4) should have been avoided. Likewise, references to colors in figures in a thesis that is printed in black and white is not very helpful (this might be easily dismissed as the pdf version has full colors, but perhaps it would be wise to alert the readers of the hard copy to that). As for the (very few) mathematical inaccuracies aside from that mentioned above, as a probabilist, I wish to point out that there is more to random variable than just being a function '$X : \Omega \to V$' for 'some set V' (top of p. 10); where is the probability measure, what are the sigma fields, how about measurability? Even in the discrete case, it should be explained what is meant. Finally, in the proof of Lemma 2.2.6, Lemma 2.2.4 by itself does not give the upper bound in (2.2.11) (unless one *knows* that the limit exists), but gives that $\limsup_{n\to\infty} \sqrt{n} s_K^{(n)} \le K + 1$ which is good enough when combined the rest of the argument.

The thesis contains an unusually small number of misprints or typographical errors and I refrain from trying to list them here.

4. Conclusion

In my opinion the volume, the quality, and the relevance of the results presented in the thesis constitute a significant contribution to the understanding of dynamics of Young tableaux under the Plancherel measure.

In my view the thesis clearly satisfies all the customary and legal requirements set forth by the Polish law. I therefore recommend that the dissertation is accepted and that the PhD defense of Mr. Łukasz Maślanka advances to its next phase.

Paweł Hitczenko

Filadelfia, June 29, 2022