Referee report on Ph.D. dissertation of Michał D. Lemańczyk, M. Sc.

Recurrence of stochastic processes in some concentration of measure and entropy problems

The thesis of Michał D. Lemańczyk, M. Sc. deals with some properties of stationary random sequences. It consists of essentially two parts. The first one concerns the dynamical system $(S, \mathscr{X}, \mu)$ induced by a random sequence on the respective path space. Here $\mathscr{X}$ is a subspace of the product of infinitely many copies of a finite state space $\mathcal{X}$, the shift transformation S satisfies $S^{-1}(\mathscr{X}) \subset \mathscr{X}$ and a probability measure μ, defined by the law of the random sequence, is invariant under S. The topics of interest cover: the formulas for topological pressure of the shift transformation, the Gibbs property of the invariant measure, explicit formulas for the maxima of a certain functional involving the Kolmogorov-Sinai entropy on a proscribed set of the laws, the entropy of the multiplicative convolution of the laws of two independent, finitely valued processes and some properties of the Mirsky measures corresponding to $\mathscr{B}$-free systems.

The second part of the dissertation deals with the concentration inequalities for stationary random sequences. It generalizes the Bernstein concentration inequalities, formulated classically for sequences of i.i.d. random variables, to random sequnces with a finite dependence range (the so called m-dependent processes) and geometrically ergodic Markov chains. These inequalities are derived under various assumptions concerning the integrability of the random variables.

Brief description of the results.

Results about subshifts over the space $\{0,1\}^{\mathbb{Z}}$. The first set of results, formulated in Sections 2.1.3, concerns the question of the Gibbs property (see definition given in (2.1.1)) of measures invariant under a subshift over the space $\{0,1\}^{\mathbb{Z}}$. Theorem 2.1.2 gives a sufficient condition, in terms of the topological entropy and density of ones, for the multiplicative convolution of the Bernoulli measure with an ergodic shift-invariant measure to not have the Gibbs property, see (2.1.1). As an application, see Corollary 2.1.4, it is shown that if the Mirsky measure associated with a $\mathcal{B}$-free system is non-periodic, then any measure of maximal entropy does not have the Gibbs property. The proofs of these results are contained in Section 4.2.1 and 4.2.2 and use the technical results shown in Section 4.1.

Other results formulated in this section, see Theorems 2.1.8-9, 2.1.11 and 2.1.16 give some explicit formulas for calculating the topological pressure od subshifts on $\{0,1\}^{\mathbb{Z}}$ under various assumptions made on the potential φ. Results of Section 2.2 deal with the question of the entropy of the multiplicative convolution of two random sequences. One of these sequences $\mathbf{X}$ is modulated by a $\{0,1\}$-valued signal $\mathbf{Y}$. The author then quantifies the decrease of entropy for the multiplicative convolution, see Theorem 2.2.8 (and Theorem 3.2.12). As a consequence it is possible to conclude the result about the strong decrease of entropy for the convolution with a weak Bernoulli sequence, see Corollary 2.2.10. The proofs of the results discussed above are shown in Chapter 3 of the thesis. They rely on using the strong Markov property and ergodicity of the random sequence $\mathbf{Y}$ in calculation of the relative entropy $H(\mathbf{X}|\mathbf{Y})$, see (3.1.1). These, highly non-trivial, calculations are quite technical and done very skillfully by the author throughout the chapter.

Variational principles for the functional defining topological pressure. Section 2.2.3 is devoted to formulating the results on the maximum of the functional appearing in the definition of the topological pressure, over a prescribed set of probability measures (usually induced by the multiplicative convolution of two independend random sequences), see (2.2.10). The respective results are formulated in Theorems 2.2.14, 2.2.18. They are proven in Section 3.2.4. In particular, it turns out, see Theorem 3.2.21, that if the potential φ depends only on a single coordinate and one of the sequences in the multiplicative convolution is of null entropy, the maximum of the functional can be calculated explicitly. Its maximizer must be of the form of the convolution of a null entropy sequence with an independent random sequence whose law is given by a Gibbs measure. This result is further generalized in Theorems 3.2.32 and 3.2.34.

The results on concentration inequalities. This is the part of the thesis with most probabilistic flavor. It formulates concentration inequalities for some classes of Markov chains and m-dependent random sequences. They generalize the well known Bernstein concentration inequalities for i.i.d. sequences to some classes of stationary random sequences of dependent random variables. The results are formulated in Section 2.3 and shown throughout Chapter 5.

Theorem 5.3.5 concerns the case of bounded centered random variables, whose conditional expectations satisfy assumptions 1-4. These assumptions allow to make a reduction to the case of stationary sequence of k-dependent random variables (see (5.3.7)), which in turn allows to use the classical Bernstein concentration inequality for independent variables. The results for the suprema of sums of random variables with a finite exponential norm are formulated in Section 5.3.2, see Theorem 5.3.10 and crucial in this section Lemma 5.3.12. As an application to Markov stationary chains the author obtains Corollary 5.3.14. The results for geometrically ergodic Markov chains, see Theorems 6.2.1, 6.2.3 and 6.2.4, are concluded by using the splitting (also, I think, sometimes called coupling) technique of Nummelin, Athreya-Ney and applying the proven result for 1-block independent chains.

Remarks. The dissertation of M. Lemańczyk is very impressive. The author has masterfully applied ideas from various branches of mathematics: dynamical systems, with applications to number theory, information theory, theory of Markov chains and classical probability. I am impressed by high technical skills demonstrated by the Ph.D. candidate and his ability to conduct demanding, quite long and complicated arguments. I must also admit that I have some problems reading the thesis at some places. It is quite technical, full of complicated notation and fairly demanding on a reader. Few intuitions are presented and this does not facilitate the readership. I think providing some ideas, behind the technical details, could facilitate readership of this highly non-trivial thesis. For example, it would have been nice if the author had given some intuitions related to the notions he uses throughout the dissertation. E.g. in the case of the Gibbs property, expressed by the estimate (2.1.1), it would be helpful to get some explanation what the idea is behind this property and why it is important. It would further highlight the importance of Theorem 2.1.2. Overall I think that the author should take more into account limitations of an average reader to follow his presentation of the material.

Another general comment concerns the structure of the thesis. As far as I can see the connection between the results of Part I and Part II (the results of Sections 5 and 6) of the thesis is not that strong and I am wondering whether there has been a point in making the compilation of these parts. It seems to me that e.g. the results of Part I, or Part II by themselves would suffice for a good Ph.D. disseration. In the present form the thesis is 132 page long.

Some of my other detailed remarks are listed below.

- I have not found the definition of the Mirsky measure. It is a topic of much discussion, in the thesis.
- What is $M_{[0,n]}$ in the proof of Theorem 3.2.1?
- Can you give an example of m-dependent l-Markov chain that appears in Theorem 2.3.2?
- **Question.** How to interpret Theorem 3.2.12 in the case $Y_k \equiv 1$ for all k? It seems that in this case $r_1 = 1$, isn't it. I am not sure what the meaning of $X_{[1,1)}$ is in this case...
- The statement about the dependence of parameters K and τ on the transition probabilities $P(\cdot,\cdot)$ is not completely clear to me. Do you mean that these constants depend on $P(x,\cdot)$?
- Do you know that $\sigma^2_{Mrv} > 0$, in (2.3.12), when f is not null π a.s.?

  I think, in the case when G is bounded, one can show it in the following way: suppose that χ is the (unique) solution of the equation $(I - P)\chi(x) = f(x)$ satisfying $\int \chi d\pi = 0$, given by the Neumann series (convergent, due to geometric ergodicity)

  $$\chi(x) = \sum_{n=0}^{+\infty} P^n f(x).$$

  One can easily show, using (2.3.11) with G bounded, that χ is bounded. Then

  $$\sum_{i=0}^{n-1} f(X_i) = \mathcal{M}_n - \chi(X_n) + \chi(X_0),$$

  where

  $$\mathcal{M}_n := \sum_{i=0}^{n-1} [\chi(X_{i+1}) - P\chi(X_i)]$$

  is a martingale. Since (X_i) is stationary, the martingale increments are stationary and

  $$\sigma^2_{Mrv} \sim \frac{1}{n}\mathbb{E}_\pi \mathcal{M}_n^2 = \mathbb{E}_\pi[\chi(X_1) - P\chi(X_0)]^2$$
  $$= \int \Big(P\chi^2 - (P\chi)^2\Big) d\pi.$$

  If $\sigma^2_{Mrv} = 0$, this would imply that $\chi \equiv \mathrm{const}$, which, in turn yields, $\chi \equiv 0$, π a.s. Hence $f = 0$, π a.s.
- I am not sure I understand the definition of tautness. What is the meaning of $\delta(\mathcal{M}_\mathcal{B})$ in the last display on p. 21? The author defined $\delta(\cdot)$ only on the subsets of $\mathbb{Z}$, while uses is on the set of measures, if I am not mistaken.
- It is hard to tell the difference between the author results and the external results. E.g. are Theorems 3.2.1, or 3.2.21 known, or are they new? This remark concerns the entire thesis.
- In the Theorem 10.0.1 cited in l. 1, p. 90 about the existence and uniqueness of the invariant measure, for the uniqueness part, don't you need some assumption of irreducibility?
- I am not sure I understand the notation $\mathbb{P}_{X\in A}[B]$ used in the appendix, see e.g. p. 106. Where it has been introduced?
- In Theorem 6.2.1, it seem the author assumes, besides geometric ergodicity of the chain also the existence of a small set. Could you explain this?

Conclusion. Summarizing, I think that the Ph.D. thesis of M. Lemańczyk, M. Sc., fulfills, with excess, the requirements of a doctoral dissertation and thereby I am pleased to recommend the Ph. D. committee to support the motion to confer the degree of doctor of mathematical sciences upon M. Lemańczyk, M. Sc.

Tomasz Komorowski

Lublin, January 4-th, 2022 *r.*