VALIDATING SEQUENCE ANALYSIS TYPOLOGIES USING BOOTSTRAPPING

MATTHIAS STUDER

A u t h o r

Studer, M.

A b s t r a c t

In this article, we propose a methodology for the validation of sequence analysis typologies based on bootstraps. The method works by comparing the quality of the obtained clustering with the quality obtained by clustering similar but unstructured data. Several models are proposed in order to test the different structuring aspects of the sequences that are important in life course research, namely sequencing, timing and duration. The framework is extended to multichannel sequence analysis by looking at the structure stemming from the association between the channels.

A u t h o r s' a f f i l i a t i o n s

LIVES Center and Institute of Demographics and Socioeconomics, Faculty of Social Sciences, University of Geneva

C o r r e s p o n d e n c e t o

matthias.studer@unige.ch

Validating Sequence Analysis Typologies Using Bootstrapping

Matthias Studer*

LIVES Center and Institute of Demographics and Socioeconomics, Faculty of Social Sciences, University of Geneva

Abstract

In this article, we propose a methodology for the validation of sequence analysis typologies based on bootstraps. The method works by comparing the quality of the obtained clustering with the quality obtained by clustering *similar* but *unstructured* data. Several models are proposed in order to test the different structuring aspects of the sequences that are important in life course research, namely sequencing, timing and duration. The framework is extended to multichannel sequence analysis by looking at the structure stemming from the association between the channels.

1 Introduction

Since its introduction in the social sciences by Abbott and Forrest (1986), sequence analysis (SA) has become an increasingly popular tool to study trajectories. It has recurrently been identified as one of the most promising methods for life course research (e.g. Shanahan, 2000; Mayer, 2009; Liefbroer and Toulemon, 2010; Brzinsky-Fay, 2014). The main strength of SA, often called optimal matching, is to provide a holistic view of processes described as a sequence, i.e. a succession of states (Abbott, 1995).

Most of the time, the method is used in conjunction with cluster analysis in order to create a typology of the trajectories. The aim is to identify recurrent patterns in the sequences or, in other words, typical successions of states through which the trajectories run. The individual sequences are distinguished from one another by a multitude of small differences. The construction of a typology of

---

*Electronic address: `matthias.studer@unige.ch`.

sequences aims to ignore these small differences so as to identify types of trajectories. Ideally, these types should be homogeneous and distinct from one another.

The success of the SA methodological framework can be explained by several typical uses and interpretations of SA typologies. Some authors such as Levy et al. (2006) engage in "structural" interpretations of the typology resulting from SA, following an *institutional* approach (Bernardi et al., 2019). They assume typology *reveals* regularities that results from the most important socio-structural forces. In a way, the typology is interpreted as an indicator of the key social constraints. Legal, economic or social constraints might coerce trajectories into few different possible types, and hinder some theoretically possible trajectories that are in fact almost never observed. For instance, the absence of childcare service and gender norms might bend the professional trajectories of women to "at home" or "part-time employment" patterns. Temporal dependencies between different stages of the trajectories might also result in regularities in observed trajectories. Such patterns might be frequent because some steps are mandatory to attain some given professional position for instance.

As noted by Abbott and Hrycak (1990), typical patterns might be widely known in the population of interest. Actors might even use these typical patterns as a model to build their own trajectories, because they anticipate their own future. This interpretation is close to the image of the "trodden trail" developed in the life-course theory (Shanahan, 2000; Brückner and Mayer, 2005). Some paths become well known and are followed by some individuals because they were previously followed by many others, resulting in types of trajectories being more and more followed.

Finally, other authors make a more descriptive use of SA typologies. They mainly see the types as a convenient way to describe, measure or operationalize the whole diversity of observed trajectories. This typically allows including the complex concept of trajectories into further analysis. For instance, we might be interested to measure how previous trajectories influence health at old age. In this case, SA allows reducing the complexity of trajectories into a few types and use the simplification in subsequent analysis. This is generally justified by considering that the description of the social world requires a certain degree of simplification.

One of the key criticisms against sequence analysis relates to the lack of a validation procedure for the resulting typology. As raised by Levine (2000) already (but also Abbott and Tsay, 2000; Abbott, 2000; Warren et al., 2015, to name a few), cluster analysis always produces a typology, which might be relevant... or not.[1]

---

[1]This criticism was recently brought to light again by Warren et al. (2015). Using simulations, they showed that inference about the number of groups or the exact cluster membership (which sequences belong to which cluster) should be done with caution. Even if their results should be extended to consider sequences made of more than two states (i.e. the typical SA case), their study highlights the need for proper validation tools to adapt the interpretation of the results.

The cluster analysis procedure works by reducing the information into a few types of sequences. However, this reduction might be abusive or we might have failed to identify distinct types of trajectories. However, each of the above-mentioned uses and interpretations make the implicit assumptions that the typology reveals a significant structure of the observed trajectories. The "institutional" approach interprets the typology as the structure of the data. The "descriptive" approach implicitly assumes that the trajectories can be fully described by the typology as soon as it is used in subsequent analysis, which is most often the case. Both interpretations therefore need a proper validation procedure that would allow us to assess the relevance of the obtained typology.

Cluster quality indices (CQI) are the most commonly used tool to validate a typology in SA (see Studer, 2013, for a review). Their aim is to measure the quality of the clustering from a statistical point of view, generally by combining indicators of the within-cluster homogeneity and the between-cluster separation. These indices can be used to compare the results of different clustering algorithms, and to guide the choice of the number of groups. However, they have three main weaknesses.

First, they cannot be computed for the one-cluster solution (i.e. no clustering). It is therefore impossible to check whether it could have been better to avoid any clustering. CQI therefore give no indications on the statistical relevance of a typology.

Second, these indices lack clear interpretation thresholds. Their values are only indicative and meaningful when compared to another typology built using the same dataset and distance matrix. It is therefore impossible to know whether the identified structure is strong or weak, except for the extreme cases which are theoretically defined.

Finally, their behavior when the number of types varies is unknown. For instance, Milligan and Cooper (1985) showed that the HC index features among the best index to recover the true underlying number of groups, but that it tends to show a slight decrease when the number of groups rises. Therefore, small variations of this index should not be interpreted. But what are small variations? The same applies to the average silhouette width which tends to favor the two groups' solution and to disadvantage higher numbers of groups. Without well-defined interpretation thresholds, the results cannot be validated, and we cannot be confident that a significant structure was found in the data.

The lack of a statistical procedure for validation can probably be explained by the fact that the usual statistical machinery cannot be used for that purpose. Indeed, in the cluster analysis phase, the groups are built to be as different as possible. Even in the absence of structure in the data, the groups would therefore be significantly different from one another compared to the usual independence case.

In other words, we lack a well-defined "independence" case to test the significance of the typology as we do in most statistical tests.

Second, SA builds a typology by comparing sequences, without assuming any model on how the data were generated. This is a strength, because it might catch patterns resulting from complex constraints that would have been caught only by very complex model. At the same time, it is a weakness. Without a model, we cannot check whether this model can be safely generalized to the whole population, as generally done in statistical tests. In other words, we also lack a "full" model that would allow us to judge the quality of the clustering.

In this paper, we propose a new validation procedure for SA. This procedure is based on the bootstrap methods for cluster homogeneity, a framework for cluster validation recently introduced in the cluster analysis literature (e.g. Gordon, 1999; Hennig et al., 2015). It provides a null model to sequence analysis allowing overcoming the limitation identified above. As pointed out by Hennig (2015), this framework requires to be adapted to the field of application in order to provide meaningful results. We discuss here several adaptations of it for SA and the study of the life course. We conclude by highlighting the added value of each of these adaptations and show how it can be used and interpreted in upcoming research making use of SA.

2   Sample Issue

Before presenting the method, we introduce here a sample issue that will serve as an illustration throughout this article. We are interested in the construction of professional and family trajectories in Switzerland following the work of Levy et al. (2006). We rely on data from the biographical retrospective survey conducted by the Swiss Household Panel[2] in 2002. We focus on the demographically and professionally dense period between 20 and 45 years old. We retain all cases without missing data, that is 1237 trajectories. The occupational trajectory is measured yearly and allows distinguishing between the following states: full-time work, part-time work, non-working and education. Figure 1 presents the chronogram of these trajectories.

From the previous work of Levy et al. (2006), we know that, while men's trajectories are relatively homogeneous and exhibit three main phases, namely "education," "full-time work" and "retirement," those of women are much more varied. Their average curve of working rates has a camel shape with a decrease in working rates when children are young and a recovery thereafter. This average curve results, however, from very distinct types of trajectories. Some women stop
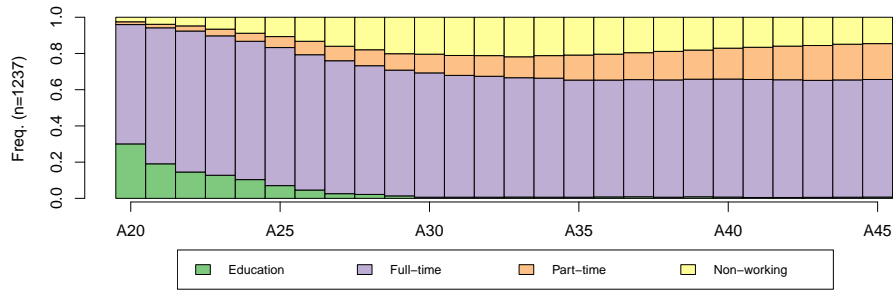
---

[2]http://www.swisspanel.ch

Figure 1: Chronogram of professional trajectories.

working or reduce their working rates while others return to work afterward. In addition, some women go back and forth between work and at home activity. We can therefore expect a strong clustering structure in the professional trajectory, distinguishing either two types, one for men and one for women, or more if the different types of women trajectories should be characterized.

The co-residence trajectories are presented in Figure 2 and are coded using seven states, living with: two biological parents, one parent, alone, a partner, a partner and at least one child, being a single parent, and other situations. We anticipate less structured co-residence trajectories, but still centered around a few types. Previous studies have shown that family trajectories of these cohorts were highly standardized Levy et al. (2006) regarding the timing and the sequencing of the states in the trajectories.
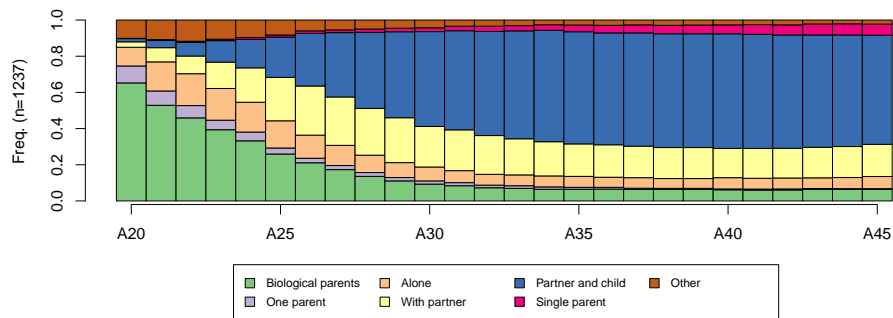

Figure 2: Chronogram of co-residence trajectories.

Following Gauthier et al. (2010), we are also interested to build a joint typology of the two sets of trajectories. The relative absence of childcare service in Switzerland leads us to expect a strongly intertwined professional and family trajectories of women. On the contrary, the strong male-breadwinner norms in Switzerland should result in a relatively low joint structure in the two sets of trajectories for men.

For all subsequent analyses, we use the Ward clustering algorithm and the

optimal matching distance with constant costs. This distance measure is known to take duration and sequencing into account while measuring dissimilarities between sequences (Studer and Ritschard, 2016). However, the proposed procedure might be used with any clustering algorithm and distance measure. After having presented our sample issues, we now turn to the presentation of the proposed methodology.

3   Bootstrap Test for Cluster Homogeneity

Recently, Hennig and Liao (2010) and Hennig and Lin (2015) proposed a new method that aims to provide interpretation thresholds for cluster quality indices (CQI). Building on the work of Gordon (1999), their generic framework is based on bootstraps to compare the quality of the obtained typology to the one we would obtain by clustering *similar* but *unstructured* data. In other words, we aim to measure to which extent the quality of the obtained typology surpass the one we would obtain for data with no clustering structure. If it is the case, we can then be confident that our typology reveals a "true" clustering structure. This comparison also provides a baseline value for interpreting the values of the CQI and a description of their behavior with varying number of groups.

As noted by Hennig (2015), this generic framework needs to be adapted to each field of study in order to be based on relevant criteria. This adaptation requirement is therefore a strength that guarantees the meaningfulness of the results for a specific field. In this paper, we propose several adaptations of this framework to sequence analysis. Each of these adaptations provides a distinct information on the quality of the obtained clustering.

3.1   The Generic Framework

As mentioned above, the general idea is to compare the quality of the obtained clustering with the quality obtained by clustering *similar* but *unstructured* data. If the quality of our clustering fall in the range of what we usually observe for unstructured data, we can conclude that the structure found with our clustering is weak. On the contrary, if our clustering quality is much higher, we can be confident and assess the relevance of our typology.

We rely on bootstrapping to estimate the CQI values of unstructured data. More specifically, the bootstrap procedure works by repeating n times the following operations:

- Generate *similar* but *unstructured* data using a "null" model.
- Cluster the generated data.
- Compute the value of the CQI of interest.

The result of this bootstrap procedure is n CQI values that we obtained by clustering unstructured data. We can then compare these n values to the one of our clustering.

In the usual statistical reasoning, we can think of these procedures as a way to estimate the null distribution of the CQI, where the null model refers to the absence of any clustering structure in the data. This strategy allows to overcome the lack of an independence model in cluster analysis that we identified above. Following the usual statistical reasoning, we use this information to derive a test-like framework to assess the clustering structure of the obtained typology.

This framework is very generic and requires to specify two important points. First, we need a null model to generate similar but unstructured data. We therefore need to define what similar and unstructured mean in the context of sequence analysis. As noted by Hennig (2015), there is no universal definition of the absence of structure, this should be defined according to the aims of the analysis and it depends on the use and interpretation of the results. Second, we need to define how we measure the quality of a clustering. We do so by choosing a CQI. However, different CQI exists, each emphasizing a different aspect of clustering structure.

The need to define these two points lead to the same question: what is structure? This is a sociological one, not a statistical one. In this paper, we propose to answer it based on the life-course paradigm.

3.2   Choosing a Cluster Quality Index

Several CQI are available. Each of them measures a slightly different aspect of the statistical quality of a given typology (Hennig, 2017). In order to make a choice among them, we need to decide which aspects of the structure we are mostly interested in.

First, we need to clarify the kind of latent structure we are interested in. Let us take the simple sociological example of Hennig and Liao (2010) to illustrate this aspect. Suppose we would like to create a typology of economic resource using income (x-axis) and wealth (y-axis) represented in Figure 3.

In this example, we might identify three groups that are represented using different colors and symbols. While cluster B and C seem to be well clustered, we can discuss the clustering of cluster A. If we are interested in a latent underlying structure measuring a *relationship* between income and wealth, cluster A is well defined. However, if we are interested in different average combination of income and wealth, this cluster is badly defined, as it regroups very different social realities. In the latter case, the four groups solution would be better.

Transposing this discussion to SA, we should decide whether we are mostly interested in difference in the average trajectory, or in the transition rates along these trajectories. The aim of SA is to identify frequently *observed* trajectories.
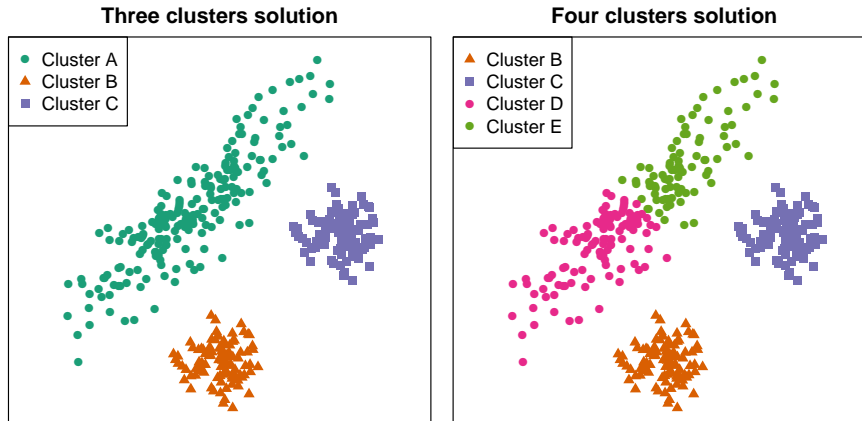
Figure 3: Scatterplot of the distribution of income on the x-axis, and wealth on the y-axis, for the three and four clusters solutions. Colors and symbols distinguish the groups.

In this sense, we are mostly interested in difference in location or average. On the contrary, latent Markov models would be more suited to capture differences in transition rates (see Piccarreta and Studer, 2019, for a review).[3]

When building a typology focusing on different locations between clusters, two statistical aspects could be of interest: between-cluster separation and within-cluster homogeneity. Figure 4 proposes a schematic graphical representation of them. Between-cluster separation refers to the extent to which types are different from one another. Within-cluster homogeneity refers to the homogeneity of each type, or in other words, the extent to which each type regroups a single trajectory. Most of the time, these two aspects are strongly interrelated: a well-separated clustering solution is often homogeneous as well.

Separation is an important aspect when giving a structural interpretation of the resulting typology. In order to look for underlying structural reasons for observing different types of trajectories, we need to observe types that are different enough, or in other words, well separated. Separation is also important when using the typology in subsequent analysis such as multinomial regression, even in a descriptive perspective. Having many sequences in between two types might create or hinder statistical relationships with other covariates of interest in this case (see Studer, 2013, for a detailed discussion).

On the other side, a strong structural interpretation of the typology also requires a high within-cluster homogeneity. If strong social constraints are at work, we

---

[3]The reasoning presented here probably partly explain the results for sequences analysis presented in (Warren et al., 2015). In their simulation study, latent classes were defined by difference in transition rates, not observed sequences.
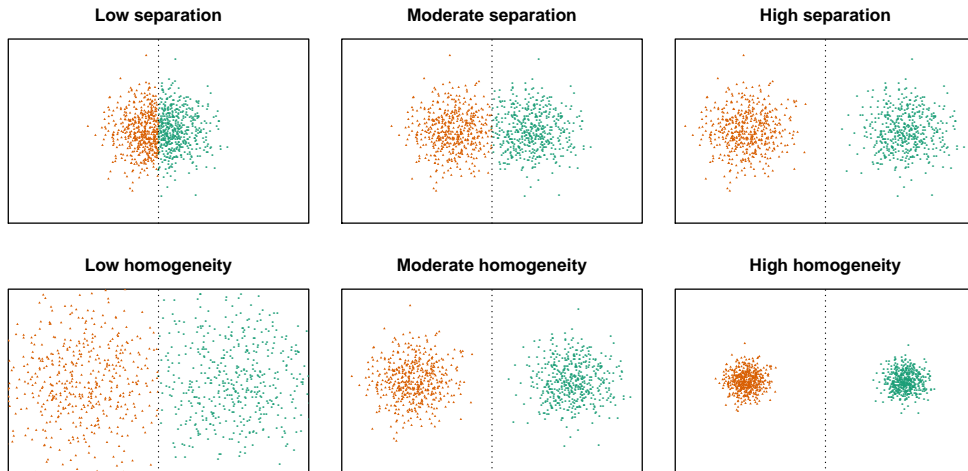
Figure 4: Examples of different levels of separation and homogeneity.

should expect homogeneous sequences within each type. The same applies for the "trodden trail" interpretation (Shanahan, 2000). Homogeneity is also important when the typology is used in subsequent analysis, because in this case, all sequences grouped into the same type will be assumed to be equal (see Studer, 2013, for a detailed discussion).[4]

To conclude, we are generally interested in both aspects. In this paper, we will therefore focus on the average silhouette width (ASW) which combines both aspects in a single index (Kaufman and Rousseeuw, 1990). The ASW measures the quality of the clustering by relating for each sequence the distances to the center of its own cluster (capturing homogeneity) to the distance to the closest other type (capturing separation). Although the authors provided some interpretation thresholds of the ASW values, these values are only indicative and their validity for sequence analysis is unknown. Furthermore, in many cases, the ASW index tends to favor the two group solutions (Hennig and Liao, 2010).

There are other indices that might be of interest in some applications. The Pseudo-R^2 aims to measure information reduction, by computing the share of the variability of the trajectories explained by the clustering (Studer et al., 2011). This index is usually avoided because of it lacks interpretation thresholds, even if it has a clear interpretation. Since the procedure developed here provides such threshold values, the index can be used. The same applies to the HC index which computes the gap between the best theoretically possible clustering and the obtained one (Hubert and Levin, 1976). It was shown to feature among the best indices to retrieve

---

[4]If we use a categorical variable to represent the typology, then all sequences in the same type have the same value on this variable.

the correct number of groups, even if small variations should not be interpreted (Milligan and Cooper, 1985). The proposed methodology provides a way to identify ignorable variations.

As noted by Hennig (2017), many CQI are available, each index measuring a slightly different aspects of cluster quality. It is often useful to combine several of them in order to describe more precisely the kind of structure found in the data. However, in this paper, we focus on the *ASW* to keep the presentation as simple as possible and to present the generic methodology.

3.3 Null Model Requirements

Aside from the choice of a CQI, we also need to define a null model. As a recall, the aim of this model is to provide typical value of the CQI when the data should not be clustered. We do so by generating data with the null model before clustering them and computing the associated CQI.

The null model should generate *similar*, but *unstructured* data (Hennig and Liao, 2010). By *similar*, we mean that the model should reproduce any non-informative structure of the data. In most applications, there is some structured information that we don't want to interpret as a clustering structure. This information should therefore be reproduced by the model.

However, we should also generate *unstructured* data. The data produced by the null model shouldn't convey any information that we consider as a clustering structure.

These two aspects require us to define the kind of structure that we are looking for. This is a sociological question not a statistical one. We propose here to answer it based on the life-course paradigm in order to provide a generic framework for sequence analysis. This requirement has been one of the key critics against this framework (see the comments on Hennig and Liao, 2010). On the contrary, we see it as one of its main strengths as it guarantees the interpretability and usefulness of the results in specific applications (Hennig, 2015).

We focus here our discussion on the single-channel case. The extension of the framework to the multichannel case is postponed to the next section.

3.3.1 Similar Data

Our null model should reproduce any aspect of the data structure we are not particularly interested in. Generally speaking, there are three of such aspects in most applications of sequence analysis.

First, some transitions might be impossible. This is generally known, and we are usually not interested to uncover this information with a typology. For instance, when studying civil status trajectories, we know that there will be no transition

between divorced and single. These impossible transitions results in regularities in the sequences, that should not be identified as "structure" as it is non-informative. It should therefore be reproduced by our null model.

Second, trajectories are generally organized by spells (Elzinga and Studer, 2015). When studying professional or family trajectories, we typically observe sequences made of a few different spells, some of them lasting for several times unit. For instance, the sequence "Education/4–Full-time/10–Inactive/6" is made of three spells. This is a very strong regularity. However, we are not particularly interested in it, because we already know it. Therefore, our null model should also generate data organized by spells.

Finally, we are often not interested in the fact that some states or spells are more common than others. Indeed, even if there was only one type of sequences in the data and therefore no clustering structure, we might expect different frequencies for each spell or state. For this reason, we don't consider spell or state frequency as a structuring aspect of the sequences. Here again, this means that our null models should reproduce, at least to some extent, the state or spell distributions of the data.

We identified three structuring aspects of sequence data that should be reproduced by our null model when using single-channel SA. These are impossible transitions, spell organization of the data and state/spell frequency. We now turn to the discussion of the structuring aspect of sequences data that we would like to capture.

3.3.2 Unstructured Data

Generally speaking, there are two broad cases for unstructured data, also called homogeneous clustering, that should lead us to avoid clustering our data. These cases are schematically represented in Figure 5. First, there might be only one type of trajectory in the data. As such, any clustering would create unneeded distinctions and should be avoided. This case is quite uncommon in social sciences, where individuals tend to follow diverse trajectories. Furthermore, many visualization tools are available in Sequence Analysis that could be used to identify it.

Second, the trajectories might be homogeneously distributed in the sequence spaces as represented on the right-hand side of Figure 5. In this case, several reasons should lead us to avoid clustering the data. First, the clusters are not homogeneous nor well separated, and, as discussed in section 3.2, this might raise concerns in most uses and interpretation of SA typology. Second, the type assigned to a given observation is dubious. In Figure 5, blue, red or green clustering are equally well defined, but each of these regrouping might have a different interpretation. The interpretation of the type of trajectory for the green point is therefore dubious as it depends on the choice between blue, red or green clustering. The same reasoning
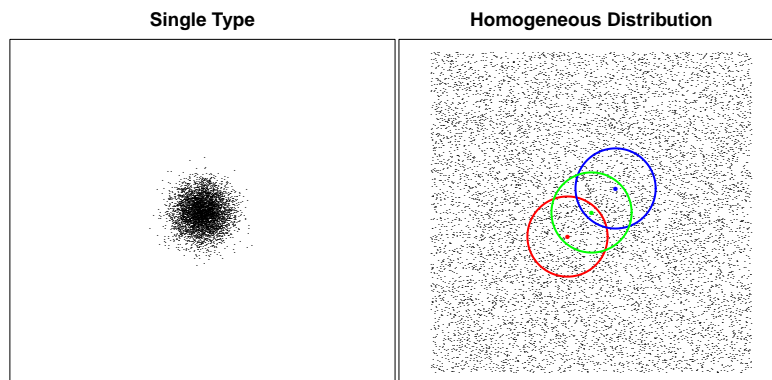
11

Figure 5: Schematic representation of homogeneous clustering cases.

could be done for any observations. Finally, in most clustering algorithms, the choice among these possibilities is not robust in this case. Partitioning around medoids (PAM) (Kaufman and Rousseeuw, 1990) relies on the overall layout and therefore ultimately on the borders of the sample. The bottom-up approach of many hierarchical clustering such as Ward is influenced by small variation on pairwise distances. In all cases, the results might depend on sampling, because the cluster is not defined according to a strong regularity of the type itself. In this article, we focus our null model on homogeneously distributed observation for two reasons. First, this case is trickier to identify using standard descriptive and visualization tools. Second, it is quite common.

We therefore want our null model to generate homogeneous data, but on which aspects? We base here our reasoning on the life-course framework. Extending the work of Settersten and Mayer (1997) and Billari et al. (2000), Studer and Ritschard (2016) identify three regularities of interest when using sequence analysis in life-course research.

The *sequencing* of the states shows the path taken by individuals. It is often thought to be of key importance, as many social norms or structural constraints relate to the ordering of the stages in a trajectory. For instance, having a child before or after marriage is not interpreted in the same way in many countries (Hogan, 1978). The sequencing also captures the dynamic of the trajectory. Observing unemployment before or after employment reveals opposite professional integration dynamics, even if the same states are observed. A recurrent ordering of the stage might also result from mandatory steps that are required to access to some specific positions.

A strong sequencing structure is found in the data, when we only observe a few, but recurrent, orderings of the states. In this case, a typology is an efficient way to summarize the main orderings. The homogeneous clustering is therefore found

when all paths are likely, as a sequence typology will probably regroup different ordering in the same types, and many sequences might lie in between types of sequences.

The *timing* of the states or transitions, i.e. when an individual is in each state or experience a transition, is also an important regularity in trajectories. Many studies have emphasized the pregnant role of age norms over the life-course (Widmer et al., 2003). Going even further, Lesnard (2010) claims that the social interpretation of a state depends on when it is observed. For instance, being unemployed at 25 or 60 years old has different causes and implications that should be studied on their own. A timing structure is found whenever some states or transition are specific to some age range.

Finally, the *duration*, i.e. the time spent in each state or the duration of spells, is also of interest in many applications. The pattern of time spent in unemployment is a key indicator in many studies of professional integration. Spell duration also capture the relative timing, sometimes called the *spacing* (Settersten and Mayer, 1997), between the main transitions, such as the time spent between marriage and first child. A duration structure is found when a spell of a given type typically last for a specific duration; for instance, when individuals tend to stay either for a long or short time in a given type of spell, such as living alone.

These three aspects are strongly related one to the others. For instance, the first and last state of a sequence partially defines the timing and the sequencing of a trajectory. Very long time spent in a state usually implies a simpler sequencing. It is therefore impossible to isolate each aspect from the others. However, in many applications, understanding specific differences stemming from each aspect is of interest as it might lead to more precise interpretation of the results (Studer and Ritschard, 2016).

Our aim is to capture the structure of the data arising in these three aspects. Our null model should therefore not reproduce it. Going even further, it should generate data showing a timing, sequencing or duration as unstructured as possible. However, it might also be of interest to measure to which extent the data are structured according to each of these aspects taken separately.

3.4 Conclusion

The generic framework aims to provide baseline interpretation of the CQI by computing the value we would obtain by clustering *similar, but unstructured data*. We can then have a new insight on the structure found by our own typology. As we have seen, this generic framework implies choosing a relevant CQI and a null model to generate the data. In this section, we decided to focus on the *ASW* to measure the quality of the clustering as separation and homogeneity are generally of interest. However, the procedure works for other indices as well.

13

We presented several properties that should be fulfilled by the null model. In life course research, we are interested to capture the structure stemming from sequencing, timing and duration, either globally or in the specific structure in each aspect.At the same time, we would like to avoid taking into account the spell organization of sequences, the state or spell frequencies, nor the non-occurrence of impossible transitions. Finally, the null model depends on the aim of the analysis. In the next section, we present several null models, each following a different goal and providing a different information.

4 Null Models for Sequence Analysis

We present here several null models for single-channel sequence analysis. We start by discussing three spell-based null models, which aims to reproduce the spell organization of sequences, before presenting a state independence model, generating sequence position by position, and a transition-rate based model. All these models are complementary as each of them provides a specific information on the kind of structure revealed (or not) by the created typology.

4.1 Spell-Based Null Models

Sequences in the social sciences are generally organized by spells and can be represented as such (Elzinga and Studer, 2015). Let us take an example. The professional trajectory of a woman working full time (F) for 5 years before working part-time (P) for 20 years is usually represented as a sequence of 25 positions such as

$$\underbrace{FF\ldots F}_{5}\underbrace{PP\ldots P\ldots P}_{20} \tag{1}$$

However, it can also be represented as a "spell sequence" in the form $(F, 5)(P, 20)$. In the latter case, the sequence is described by a state vector $x = (F, P)$ and an attached duration vector $t = (5, 20)$. Such representation emphasizes two dimensions of the sequences: the sequencing of the state (the state vector) and the time spent in each state (the duration vector). However, contrary to the usual representation of sequences, the timing dimension is less pregnant.

Using this spell representation, we can create sequence data by generating two vectors, one for the states and another for the durations, emphasizing sequencing and duration. By doing so, we reproduce one of the strongest regularities in sequences as used in the social sciences, its organization by spells.

We propose here three models, each bringing a specific information on the quality of the obtained typology. While the "randomized sequencing" measures

the added value of the typology when compared to the situation where all paths are likely, the "randomized duration" measures the captured structure in the data that results from duration in the time spent in each state. Finally, the first model presented, called "combined randomization", combine the two dimensions.

4.1.1 Combined Randomization

Following the above-presented spell representation of sequences, we create two vectors for each sequence. The state vector is made by sampling states among all spells. In order to avoid impossible transitions, we use the following procedure. We start by randomly sampling the first state. The next states are sampled among possible states given the previous one. In all cases, the sampling is weighted by the relative frequencies of state among all the overall possible spells. The sequencing is therefore as random as possible, although spell frequency is kept.

The duration vector is created by sampling duration among all the observed spells. Since we do not attach duration to the state, the durations are completely random, even if the durations are those typically observed in the data. If our sequences are generally made of long spells, this will be reproduced by the null model. Finally, the sequences are cut to match the observed length in the data.

This null model therefore reproduces the spell structure of sequences, impossible transition, spell state's frequencies and durations, while randomizing sequencing, duration and therefore timing as well. Figure 6 presents sequences generated using this null model for the professional career data. While the chronogram shows the sequences generated among all bootstraps, the index plot represents only one bootstrap to avoid overplotting. As expected, timing, duration and sequencing are not reproduced but the spell organization of sequences is. State frequency at each time point is not reproduced because durations are not attached to states, but the full-time (violet) state is still the most frequent.
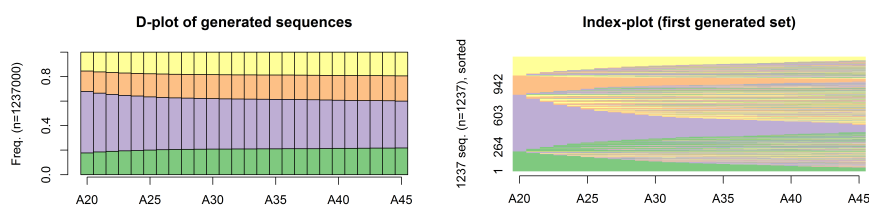

Figure 6: Chronogram and index plot of some sequences generated by the combined randomization null model.

As a recall, the aim of such a procedure is to generate sequences, cluster them and compute the CQI. These CQI values are then recorded and used to provide

a baseline value for the CQI of our own clustering. Figure 7 provides a density plot of this null distribution for clustering between 2 and 10 clusters using the Ward clustering algorithm and optimal distance with constant costs. The clustering algorithm and distance measures should be the same as in our original analysis. The gray area under the curve represents a 90%-observed interval of these values. The values obtained in the original data are represented using vertical red line labeled with the number of groups.
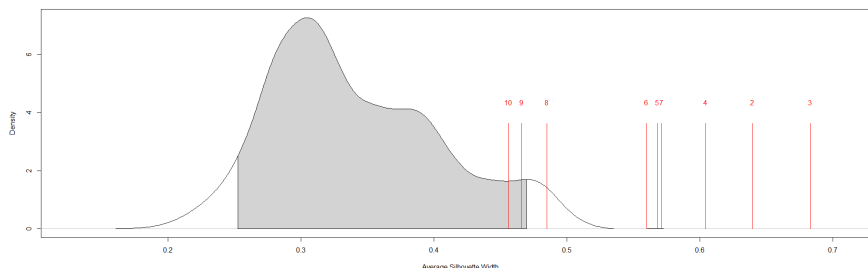

Figure 7: Distribution of the null values of the Average Silhouette Width for two to ten clusters. The observed values of the ASW are represented with vertical red bars.

The central 90% ASW values lie in the interval $[0.25; 0.47]$. This roughly equals the indicative inconclusive interval $[0.25; 0.5]$ proposed by Kaufman and Rousseeuw (1990). However, this is application-specific. For the co-residence trajectory, the bootstrapped 90%-interval is $[0.17; 0.34]$ which is below these standard values. This illustrates that it is not possible to define generic CQI value thresholds and therefore the need for the proposed framework.

By construction, any ASW value in the 90% interval $[0.25; 0.47]$ is comparable to a situation where we don't have any clustering structure. We can therefore exclude the solutions in 9 or 10 groups. On the contrary, the ASW values obtained with two to seven groups are well outside this range. We can therefore consider that they reveal a true structure in the data. Going even further, Hennig and Liao (2010) propose a test-like interpretation of the results. We can reject the assumption of our null model based on our data of homogeneous sequencing and spell durations. Our typology therefore reveals a timing, duration or sequencing structure if we consider two to seven groups.

We can also look at the null-CQI according to the number of groups. Following the presentation proposed by Hennig and Lin (2015), the left-hand side of Figure 8 presents the ASW of our typology for a varying number of groups using a solid black line. The 1000 bootstrapped CQI values are represented using individual gray lines, while an 90% interval is computed and represented with a gray polygon. The

16

right-and side of the figure represents the ASW values of our typology standardized using the mean and standard deviation of the bootstrapped values for a given number of groups. The aim of the standardization is to account for the changing behavior of the CQI measure and clustering algorithm for different number of clusters. A high standardized value can then be interpreted as a high relative gain in the structure found for a given number of groups. According to this analysis, the two or three groups solutions seems particularly well suited. If we want to favor a bigger number of groups, the seven cluster solution seems to be a good compromise.

Average Silhouette Width (ASW)
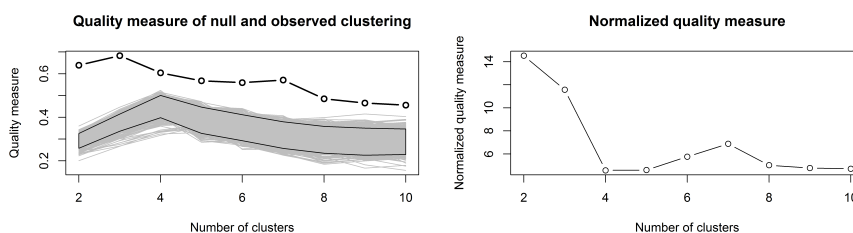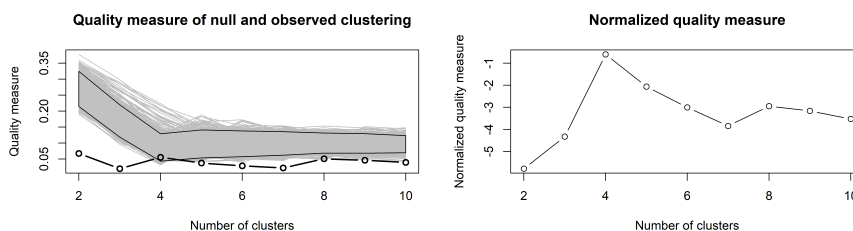

Hubert's C index (HC)

Figure 8: Combined Randomization: Observed and bootstrapped values of ASW and HC values for a varying number of clusters. The right-hand-side plots show the evolution of the standardized cluster quality indices.

Figure 8 also describes the evolution of CQI with a varying number of groups when no clustering should be found in the data. In this case, we tend to observe the highest ASW values with four groups corresponding to the number of states in the alphabet.

This visualization is of special interest when using a CQI that cannot be directly compared for different number of groups such as the Pseudo-R^2 or the Hubert's C index. As a recall, the HC index, which should be minimized, featured among the best one to select the correct number of groups according to Milligan and Cooper (1985), *if small decreases are not taken into account.* The proposed procedure allows taking the expected decrease into account. The bootstrapped and observed

17

value of the *HC* index are presented in Figure 8 as well and would lead us to consider the same number of groups.

4.1.2 Randomized Sequencing

The previously introduced null model is very generic as timing, duration and sequencing are randomized. We propose here to focus on the sequencing only and to measure the structure arising from the ordering of the state. We achieved this by randomly generating state vector, using the same procedure as for the above-presented "combined" approach. However, the durations are sampled among the spell of a given state. Here again, the sequences are then cut to match the observed length in the data. With such a strategy, the sequencing is completely randomized but the time spent in each state is kept, i.e. remain as in the original data. By comparing our clustering to the one obtained with randomized sequencing, we therefore measure the difference relating to structured sequencing in the data.

Sequences Generated by the Randomized Sequencing Null Model
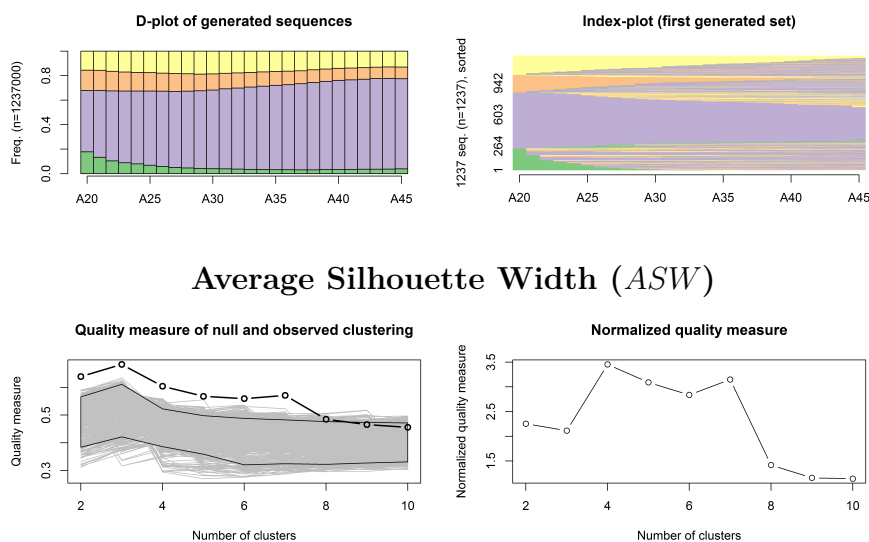

Average Silhouette Width (*ASW*)

Figure 9: Randomized Sequencing Simulation Results. The top plots present a chronogram and index plot of sequences generated by the randomized sequencing null model, while the bottom ones presents bootstrapped and observed *ASW* values for a varying number of clusters.

Figure 9 presents a chronogram of the sequences generated among all bootstraps and an index plot of one of the generated datasets. The generated sequences are likely, in the sense that they could have been observed. The frequencies of each

18

state are reproduced, but the timing is not. As intended, we observe some unlikely orderings such as sequences ending by an education spell.

Figure 9 also presents bootstrapped and observed ASW per number of clusters. The observed ASW are higher than the null values with fewer than height groups, meaning that our typology revealed a structure state's ordering in the sequences. The highest "relative" gain is found for the four and seven cluster solutions if we look at the normalized ASW. Further interpretations can be drawn by comparing these results with the combined randomization ones, which was favoring two or three groups solutions. Since it is not the case for "randomized sequencing", it means that the structure identified for two or three groups mostly relate to the time spent in each state.

4.1.3 Randomized Duration

Instead of focusing on sequencing, we can focus on duration. We generate sequences by keeping the same observed sequencing but randomizing the associated duration in each state. We do so by randomly drawing proportion of the trajectory spent in each state using the uniform distribution. While the strategy allows measuring the structure arising from duration in sequences made of more than one spell, the sequences with only one spell are fully reproduced. Indeed, the proportion of the time spent when there is only one spell is necessary 100%.

Figure 10 presents a chronogram and an index plot of some sequences generated with this null model. The sequences could have been observed, but we observe sequences with a long time spent in education, which is unlikely. As expected, sequences in full-time employment are fully reproduced and will therefore be counted as unstructured data.

As before, Figure 10 presents the observed and bootstrapped ASW for our typology. We observe quite a high value of the bootstrapped CQI for the solutions in four clusters, i.e. one cluster for each possible state in the sequences. This is probably because a clustering solution according to the dominant state in each sequence tends to be quite a good one. According to this null model, there is no added value of going beyond eight groups. Confirming our previous interpretation, we observe a high relative gain (i.e. standardized value) for two and three clusters solutions which should therefore be primarily linked to the time spent in each state.

4.2 Position-Based Null Models

Sequences can also be generated position by position using transition rates. Starting from a given first state, the sequences are then built position by position by looking at the previous state. The solution is appealing as it is possible to reproduce many

Sequences Generated by the Randomized Duration Null Model
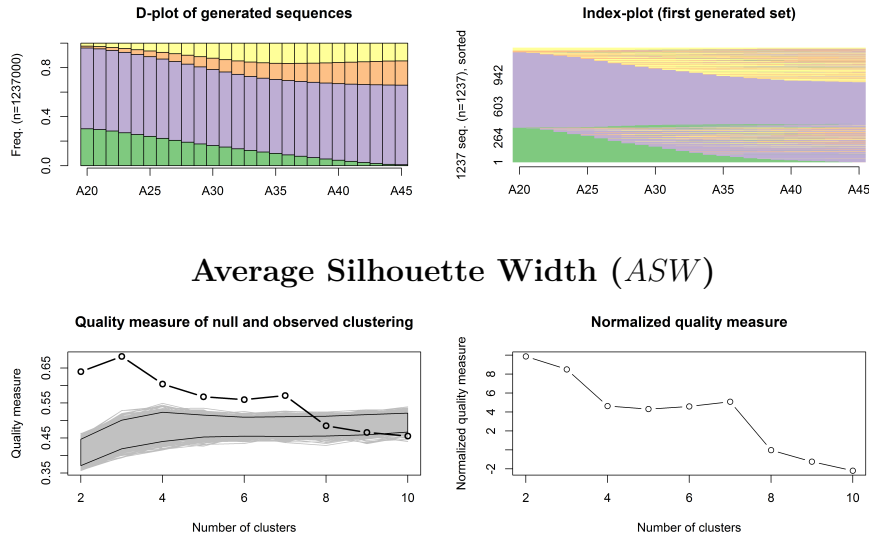

Average Silhouette Width (ASW)

Figure 10: Randomized Duration Simulation Results. The top plots present a chronogram and index plot of sequences generated by the randomized sequencing null models, while the bottom ones presents bootstrapped and observed ASW values for a varying number of clusters.

timing, sequencing and durations depending on the transition matrix. However, two objections can be made.

First, as we already argued, we are not interested in difference in the relationships between variables such as transition rates, but in difference in the average observed sequence.

Second, a transition matrix is not necessarily a good candidate for a null model, as it might reproduce many clustering structure in the data. This can be exemplified using the following transition matrix.

$$\begin{bmatrix} & A & B & C \\ A & 0 & .5 & .5 \\ B & 0 & 1 & 0 \\ C & 0 & 0 & 1 \end{bmatrix}$$

And further suppose that all sequences start in the state A. With such transition matrix, we would end up with two different observed sequences. A sequence with the pattern $AB\ldots B$ and the other one with $AC\ldots A$. As a result, the observed sequences are strongly structured in two groups even if we used a single transition matrix. We therefore need to choose a suitable transition matrix. Hennig and Lin (2015) use time-varying observed transition rates. However, with such a number of

20

parameters, we are able to almost fully reproduce the original data, including its potential clustering structure. We tested this option and, as in Hennig and Lin (2015), the results are along this discussion, with observed CQI always lying in the confidence interval of bootstrapped CQI values.[5]

The question is therefore which transition matrix to use. We investigate here two possibilities, each corresponding to a model-based approach. First we discuss the state independence model, which is strongly related to the underlying assumptions of Latent Class Analysis (Piccarreta and Studer, 2019). Second, we discuss the use of homogeneous transition rates, an assumption underlying some (Hidden) Markov Models (Piccarreta and Studer, 2019).

4.2.1 State Independence

In the state independence model, the state at a given position is independent of the previous one. In order to respect the basic properties of the data, the state distribution at each time point is respected. Technically, this can be simply achieved by using a time-dependent transition matrix, where the distribution of the states are independent of the previous state. This null model is roughly equivalent to the conditional state independence assumption that underlines latent class analysis.

Figure 11 presents a chronogram and an index plot of some sequences generated with this null model. While the chronogram seems plausible with time-varying state distribution perfectly reproduced, the index-plot presents many sequences we would never observe in the social sciences, with far too many transitions. The spell organization of the observed sequences is therefore not reproduced by the model.

As a result, the bootstrapped CQI values are low and present a decreasing trend as shown in Figure 11. With this null model, we always conclude that we found a significant structure in the data. However this is most probably related to the spell structure of sequences in the social sciences, that is not respected here. We are therefore not recommending it for test-like interpretation. However, it might be used to guide the choice of the number of groups. The standardized CQI indicates a solution in seven or ten groups.

4.2.2 First Order Markov Null Model

In the first order Markov null model, we use the transition rates computed over the whole sequences to generate sequence position by position. This is often called a Markov assumption where the state only depends on the previous one. The first state of the sequence is randomly selected using the observed proportion of each state at time one.

---

[5]The results are available on request.

Sequences Generated by the State Independence Null Model
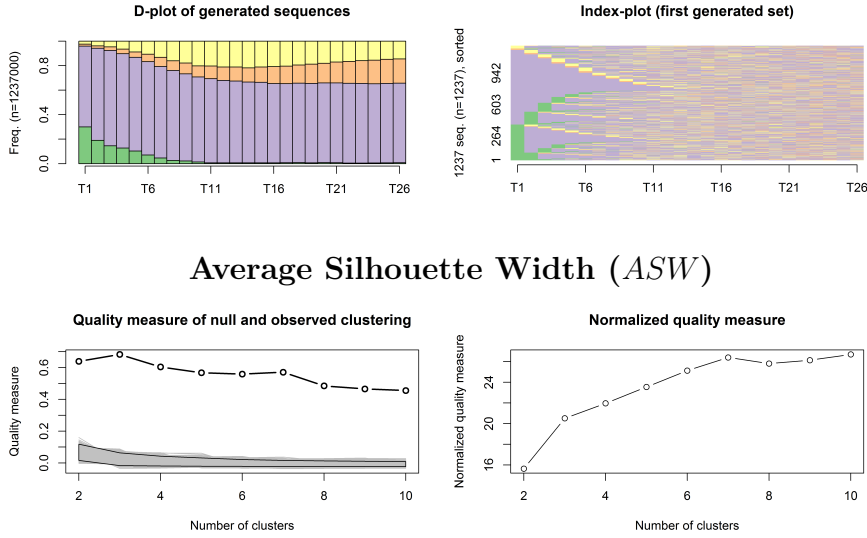

Average Silhouette Width (ASW)

Figure 11: State Independence Simulation Results. The top plots present a chronogram and index plot of sequences generated by the randomized sequencing null models, while the bottom ones presents bootstrapped and observed ASW values for a varying number of clusters.

Within the life-course paradigm, the notion of timing can either refer to the states within the process as discussed earlier, or to the timing of the transitions between the states. By using a first-order Markov null model, we assume that transition rates are non-time-varying. Therefore, in our framework, a strong structure would be found if the timing of transitions differ among clusters.

Figure 12 presents a chronogram of all the generated sequences and an index plot of one of the generated datasets. The generated sequences are quite similar to the observed ones, except that we observe less transition out of full-time employment at the beginning and more in the end of the sequences.

As a result of this similarity, the observed ASW is not much higher than the bootstrapped ones as shown in Figure 12. Here, the highest standardized value is found for the seven cluster solution. This leads us to think that the seven-cluster solution has captured a significant structured timing in the data.

4.3   Conclusion

We introduced several null models aiming to provide a baseline interpretation of CQI values. These models distinguish one another by the kind of structure that they reproduce or not. The interpretation of the results depends on the similarities

Sequences Generated by the First Order Markov Null Model
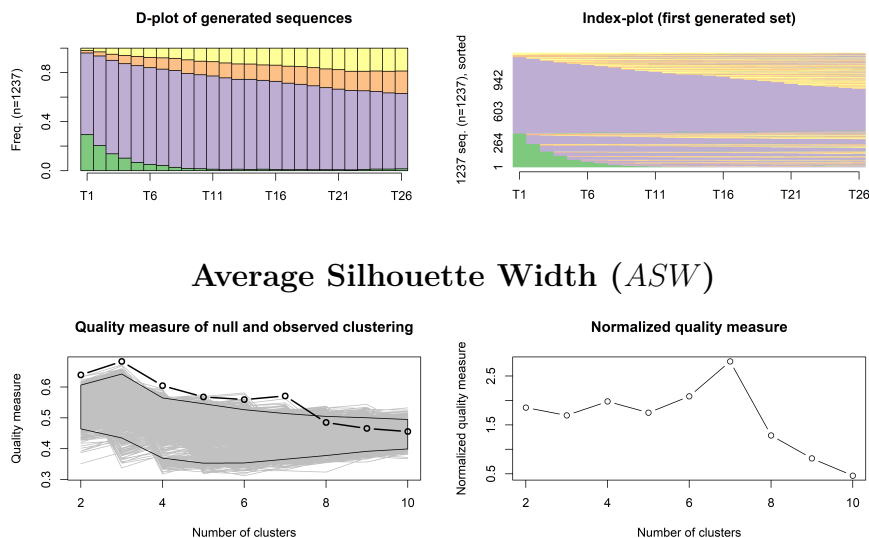

Average Silhouette Width (*ASW*)

Figure 12: First Order Markov Null Model Simulation Results. The top plots present a chronogram and index plot of sequences generated by the randomized sequencing null models, while the bottom ones presents bootstrapped and observed *ASW* values for a varying number of clusters.

and differences between the observed and generated sequences. In this section, we take a broader perspective and discuss the added value of each null model and its typical use.

Table 1 presents some indicators of the dissimilarity between the data generated by a null model and the observed ones. We do so by computing the Manhattan distance between, on the one side, a distribution or a set of indicators of the original data, and, on the other side, the same indicators computed on the data generated by a null model. These dissimilarities are computed on the overall state frequencies (across all-time points) and average spell duration (in each state separately) to measure differences in durations. The timing dimension is captured by measuring difference in the time-varying state distribution and time-varying transition rates. Finally, the differences in sequencing are measured through the use of transition rates among sequence of spell, ignoring duration.

The value of these indicators cannot be interpreted as they depend on the scale of each aspect. We standardized them by dividing by the value of the "combined randomization" null model. The choice of this model is justified by the fact that it is the most random, while still keeping the spell structure of sequences.

As expected, the "combined randomization" model presents the most important dissimilarities in all measured aspects. The only exception is found with the "state

Table 1: Dissimilarities between original and simulated professional and co-residence sequences for different null models. A greater value and darker gray scale relate to a higher dissimilarity on the specified aspect.

	Combined		Sequencing		Duration		State indep.		FOMM	
	Coh	Prof	Coh	Prof	Coh	Prof	Coh	Prof	Coh	Prof
Overall state distribution	1.00	1.00	0.12	0.10	0.94	0.36	0.00	0.00	0.20	0.10
Average spell duration	1.00	1.00	0.30	0.12	1.24	0.66	2.40	1.99	0.25	0.23
Time-varying state distribution	1.00	1.00	0.55	0.42	0.76	0.34	0.00	0.00	0.24	0.17
Time-varying transition rate	1.00	1.00	0.88	0.74	0.84	0.87	7.45	6.25	0.70	0.64
Spell transition rates	1.00	1.00	1.00	1.00	0.00	0.00	0.90	0.89	0.01	0.00

FOMM stands for "First Order Markov Model".

independence" model that produces very unlikely sequences with very short spells and, for the same reason, very different time-varying transition rates.

As intended, the "sequencing" null model mainly differs from the original data in the spell transition rates (and therefore the sequencing of the state), and is quite similar regarding spell duration. This null model also randomizes, although to a lesser extent, the timing of the state and transitions.

The "duration" null model keep the sequencing of the original data, but randomize duration and timing. It should be noted that duration differences are more important for the co-residence than the professional trajectories. This is linked to the share of sequences that stay in a single state that are unaltered by the null model. Theses kinds of sequences are much more frequent in the professional sequences, leading to fewer differences according to this dimension. For this reason, this null model should be primarily used when such kinds of sequences are infrequent.

As explained, the "state independence" model doesn't reproduce the spell structure of the sequences, leading to huge, but unlikely differences in the spell duration distribution and the transition rates. This null model should therefore be used with caution.

The "First-Order Markov Model (FOMM)" produce the most similar sequences on every aspect except on the time-varying transition rates. For this reason, it is a reasonable null model when we are interested in the timing structure of the transitions within the sequences.

While null models randomizing duration favor the two or three cluster solutions for the professional trajectories, those randomizing timing and sequencing (i.e. sequencing and FOMM) would lead to the choice of the seven cluster solution. In all cases, a test-like interpretation of the results lead to the conclusion that a significant timing, duration or sequencing structure was found in the data. We can therefore be confident in the interpretation of our typology in seven groups for instance.
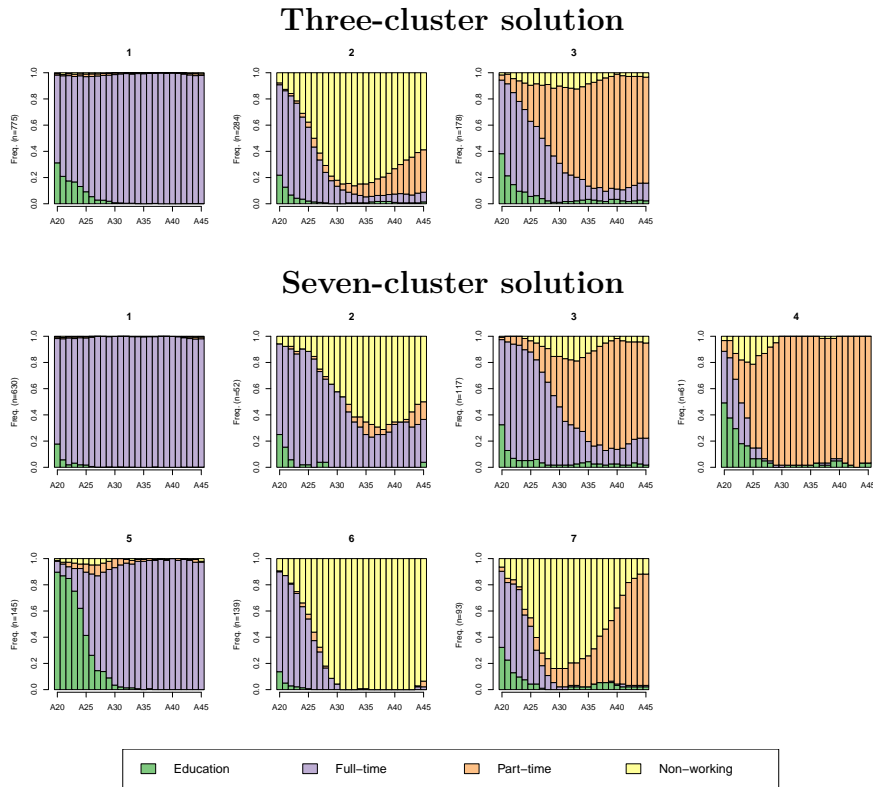
Figure 13: Final Typologies of Professional Trajectories in Switzerland

This interpretation can be confirmed by looking at Figure 13, where the three and seven cluster solutions are presented. The three clusters solution mainly distinguishes the patterns according to the overall time spent in the three most frequent states: full-time, part-time and non-working. The seven cluster solution makes further distinctions according to the timing (cluster 3 vs 4 or cluster 2 vs 6), but also the sequencing (cluster 6 vs 7 for instance).

In our sense, the choice between these solutions is no longer a statistical one. They were both validated by the proposed procedure as were all cluster solutions with less than height groups. The choice should be made according to the application and the theories that are being tested. Is it worth having seven categories representing the professional trajectories? Should we distinguish according to the timing and sequencing, or mainly to the time spent in each state? The aim of the validation procedure is to avoid drawing wrong conclusion about the structure of the data, not to replace the theory of the domain under study.

5 Extension to the Multichannel Case

Multichannel sequence analysis (MCSA) refers to the joint clustering of parallel sequences, such as professional and family trajectories (Pollock, 2007; Gauthier et al., 2007; Piccarreta, 2017). The aim is generally to build a single typology for the two (or more) sets of sequences.

Generally speaking, two different uses of MCSA can be distinguished (Studer, 2015). First, we might have several indicators of the same underlying trajectory. For instance, the professional career might have been measured with several indicators such as hierarchical position, income level and occupational rate. In this case, the goal is to build a typology of professional trajectories. Although we technically use parallel sequences, we still aim to study a single career. For this reason, the previously introduced null models are perfectly suitable. Indeed, we need a null model for the underlying unique dimension: the professional trajectory in our example.

The second and most frequent use of MCSA is to build a joint typology of trajectories of two or more life domains. The aim is generally to understand how the two trajectories are intertwined. For instance, in our sample application, we are interested in building a joint typology of family and professional trajectories in Switzerland. The relative absence of childcare system and the strong male-breadwinner norms in Switzerland lead us to think that the two types of trajectories are strongly linked, at least for women.

As point out by Piccarreta (2017), such a use of MCSA is only meaningful if the two life domains are associated. The MCSA joint typology summarizes the association, if there is one... or not. Let us take a small example to illustrate this. Suppose that we are interested in the joint clustering of family and professional trajectories, and that there are two types F_1, F_2 of family trajectories and three types $P_1, P_2, P3$ of professional careers. If family and professional careers are independent, we should identify six types in our joint typology: F_1/P_1, F_1/P_2, F_1/P_3, F_2/P_1, F_2/P_2 and F_2/P_3. Such a typology is generally unclear because the family trajectory in F_1/P_1 and F_1/P_2 would typically differ by small variation and it will be hard to know whether it should be interpreted or not.

On the contrary, if there is a perfect association between professional and family trajectories, a joint typology is very useful and an efficient way to summarize the association. Using our previous example, we could have only three types to summarize each combined trajectory F_1/P_1, F_2/P_2 and F_2/P_3. We could therefore conclude that F_1 generally goes in pairs with P_1 while F_2 is generally associated with either P_2 or P_3.

MCSA should therefore only be used when the parallel trajectories are intertwined. Based on this reasoning, we extend our framework to validate an MCSA typology. In this case, we want our null model to reproduce any structure in each

channel taken separately. Indeed, the goal is to capture the structure stemming from the *association* between the channels, not the structure of each channel taken separately. On the contrary, in our null model there should be no association between the channels.

This can be achieved using permutation tests. These tests work as follows. The first channel is kept as it appears in the data, while the observations of the second channel are randomly permuted. By doing so, any association between the two channels is removed. The procedure can be easily adapted to more than two channels by randomly permuting any additional channels. Our null model therefore keep the structure of each channel taken separately while randomizing the association between channels.
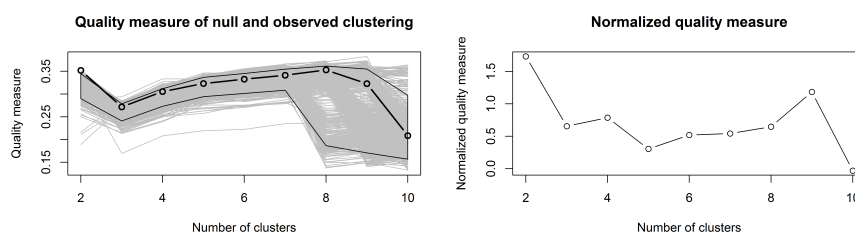

Figure 14: Multichannel observed and bootstrapped *ASW*-value for a varying number of groups and evolution of the standardized *ASW*.

Figure 14 presents the values of the CQI for the joint clustering of family and professional trajectories for our whole sample. As before, the gray lines show the values of the 1000 random permutations in the data. The observed CQI values always lies within the observed confidence interval. We can therefore conclude that our joint clustering do not capture more structure than a situation without any association between the channels. It might be better to consider creating two separate typologies. In all cases, the bootstrapped 90% interval is quite low, as compared to the values proposed by Kaufman and Rousseeuw (1990). This can be explained by the high heterogeneity of the data as soon as several channels are taken into account. This should be the case in most MCSA applications.

Figure 15 presents the same analysis, but focusing only on women. We get a different picture. Indeed, we observe a better clustering quality for the two, six, seven and height cluster solutions. The difference in conclusions for men and women can be explained by the patterns of men. Most men follow the same full-time professional pattern, but different types of family trajectories. For men, there is therefore almost no association between channels, they are independent. However, there is an association for women and we found a significant common structure with our typology.
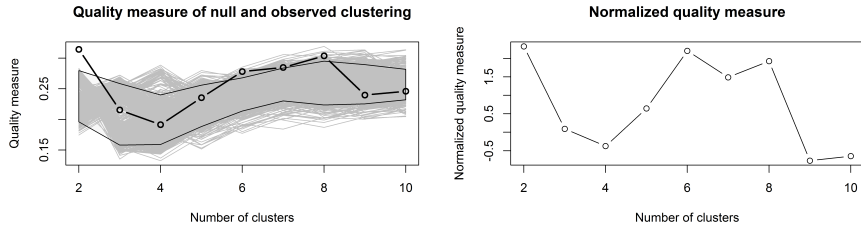
27

Figure 15: Women only: observed and bootstrapped *ASW*-value for a varying number of groups and evolution of the standardized *ASW*.

Aside from providing baseline value for the CQI in MCSA, the results of the proposed extension also support a nice test-like interpretation. We can compare the MCSA typology to the case when no association is found in the data. We are confident that this is a useful addition to the MCSA methodology.

6   Conclusion

In this article, we proposed a methodology for the validation of sequence analysis typologies based on bootstraps. The method works by comparing the quality of the obtained clustering with the quality obtained by clustering *similar* but *unstructured* data. If the quality of our clustering fall in the range of what we usually observes for unstructured data, we can conclude that the structure found with our clustering is weak. On the contrary, if our clustering quality is much higher, we can be confident and assess the relevance of our typology.

In order to use the framework, we need a null model to generate *similar* but *unstructured* data. We proposed several null models aiming to test different structuring aspects of the sequences that are important in life course research, namely sequencing, timing and duration. We also extended the framework to multichannel sequence analysis by looking at the structure stemming from the association between the channels.

The proposed methodology is in our sense a great addition to sequence analysis. It is a first step toward an integrated approach for the evaluation and validation of sequence analysis typologies. The lack of a proper methodology featured among the long-lasting critics against sequence analysis.

The proposed methodology provides expected behavior of cluster quality indices for different kind of structure. This allows to give a better interpretation of the quality of the clustering and the kind of structure found in the data. It also supports a test-like interpretation. Finally, the proposition of several specific null model's forces us to think about the kind of structure we are looking for in the

28

data. We think that this is one of the key strengths of the generic framework proposed by (Hennig and Liao, 2010). This question should be answered based on social-science arguments, and will therefore always be application specific. However, the proposed null models provide some directions to think about it based on the life-course paradigm.

The proposed methodology focus on two aspects of the validation of a typology from a statistical perspective, namely within-cluster homogeneity and between cluster separation. However, other aspects are also of key importance. Han et al. (2017) stress that a good typology should reproduce known association with other key variables such as education or gender. In Hennig (2007) and Hennig (2008), the emphasis is put on the stability of the clustering across several samples. Finally, many articles stress the key importance of the interpretability and theoretical soundness of the results (see for instance Piccarreta and Studer, 2019). All of these aspects are important and should be taken into account. However, we still lack a well-defined integrated approach to do so. This is probably one of the most important challenge facing sequence analyses.

References

Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology 21*, 93–113.

Abbott, A. (2000). Reply to Levine and Wu. *Sociological Methods & Research 29*(1), 65–76.

Abbott, A. and J. Forrest (1986). Optimal matching methods for historical sequences. *Journal of Interdisciplinary History 16*, 471–494.

Abbott, A. and A. Hrycak (1990). Measuring resemblance in sequence data: An optimal matching analysis of musician's carrers. *American Journal of Sociolgy 96*(1), 144–185.

Abbott, A. and A. Tsay (2000). Sequence analysis and optimal matching methods in sociology, Review and prospect. *Sociological Methods and Research 29*(1), 3–33. (With discussion, pp 34-76).

Bernardi, L., J. Huinink, and R. A. Settersten (2019). The life course cube: A tool for studying lives. *Advances in Life Course Research 41*, 100258. Theoretical and Methodological Frontiers in Life Course Research.

Billari, F. C., J. Fürnkranz, and A. Prskawetz (2000). Timing, sequencing, and quantum of life course events: a machine learning approach. Working Paper 010, Max Planck Institute for Demographic Research, Rostock.

Brückner, H. and K. U. Mayer (2005). De-standardization of the life course: What it might mean? and if it means anything, whether it actually took place? *Advances in Life Course Research 9*, 27–53.

Brzinsky-Fay, C. (2014). The measurement of school-to-work transitions as processes. *European Societies 16*(2), 213–232.

Elzinga, C. H. and M. Studer (2015). Spell sequences, state proximities and distance metrics. *Sociological Methods and Research 44*(1), 3–47. published online 9 July 2014.

Gauthier, J.-A., E. D. Widmer, P. Bucher, and C. Notredame (2007). Multichannel sequence analysis applied to social science data. Manuscript, University of Lausanne. (Under review).

Gauthier, J.-A., E. D. Widmer, P. Bucher, and C. Notredame (2010). Multichannel sequence analysis applied to social science data. *Sociological Methodology 40*(1), 1–38.

Gordon, A. D. (1999). *Classification* (2nd Edition ed.), Volume 82 of *Monographs on Statitics and Applied Probability*. Chapman & Hall/CRC.

Han, Y., A. Liefbroer, and C. Elzinga (2017). Comparing methods of classifying life courses: sequence analysis and latent class analysis. *Longitudinal and Life Course Studies 8*(4).

Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis 52*, 258–271.

Hennig, C. (2008). Dissolution point and isolation robustness: robustness criteria for general cluster analysis methods. *Journal of Multivariate Analysis 99*, 1154–1176.

Hennig, C. (2015). What are the true clusters? *Pattern Recognition Letters 64*, 53–62.

Hennig, C. (2017). Cluster validation by measurement of clustering characteristics relevant to the user. *ArXiv e-prints*.

Hennig, C. (2017). Cluster validation by measurement of clustering characteristics relevant to the user. arxiv e-prints.

Hennig, C. and T. F. Liao (2010). Comparing latent class and dissimilarity based clustering for mixed type variables with application to social stratification. Technical report, Department of Statistical Science, UCL, Department of Sociology, University of Illinois.

Hennig, C. and C.-J. Lin (2015). Flexible parametric bootstrap for testing homogeneity against clustering and assessing the number of clusters. *Statistics and Computing 25*(4), 821–833.

Hennig, C., M. Meila, F. Murtagh, and R. Rocci (Eds.) (2015). *Handbook of Cluster Analysis*. Chapman and Hall/CRC.

Hogan, D. P. (1978). The variable order of events in the life course. *American Sociological Review 43*, 573–586.

Hubert, L. and J. Levin (1976). A general statistical framework for assesing categorical clustering in free recall. *Psychological Bulletin 83*, 1072–1080.

Kaufman, L. and P. J. Rousseeuw (1990). *Finding groups in data. an introduction to cluster analysis*. New York: Wiley.

Lesnard, L. (2010). Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns. *Sociological Methods and Research 38*, 389–419.

Levine, J. (2000). But what have you done for us lately. *Sociological Methods & Research 29 (1)*, pp. 35–40.

Levy, R., J.-A. Gauthier, and E. Widmer (2006). Entre contraintes institutionnelle et domestique : les parcours de vie masculins et féminins en Suisse. *Cahiers canadiens de sociologie 31*(4), 461–489.

Liefbroer, A. C. and L. Toulemon (2010). Demographic perpectives on the transition to adulthood: An introduction. *Advances in Life Course Research 15*, 53–58.

Mayer, K. U. (2009). New directions in life course research. *Annual Review of Sociology 35*, 413–433.

Milligan, G. and M. Cooper (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika 50*(2), 159–179.

Piccarreta, R. (2017). Joint sequence analysis: Association and clustering. *Sociological Methods & Research 46*(2), 252–287. Published ahead of print.

Piccarreta, R. and M. Studer (2019). Holistic analysis of the life course: Methodological challenges and new perspectives. *Advances in Life Course Research 41*, 100251. Theoretical and Methodological Frontiers in Life Course Research.

Pollock, G. (2007). Holistic trajectories: A study of combined employment, housing and family careers by using multiple-sequence analysis. *Journal of the Royal Statistical Society A 170*(1), 167–183.
English

Settersten, Richard A., J. and K. U. Mayer (1997). The measurement of age, age structuring, and the life course. *Annual Review of Sociology 23*, 233–261.

Shanahan, M. J. (2000). Pathways to adulthood in changing societies: Variability and mechanisms in life course perspective. *Annual Review of Sociology 26*, pp. 667–692.

Studer, M. (2013). WeightedCluster library manual: A practical guide to creating typologies of trajectories in the social sciences with R. LIVES Working Papers 24, NCCR LIVES, Switzerland.

Studer, M. (2015). Comment: On the use of globally interdependent multiple sequence analysis. *Sociological Methodology 45*(1), 81–88.

Studer, M. and G. Ritschard (2016). What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society, Series A 179*(2), 481–511.

Studer, M., G. Ritschard, A. Gabadinho, and N. S. Müller (2011). Discrepancy analysis of state sequences. *Sociological Methods and Research 40*(3), 471–510.

Warren, J. R., L. Luo, A. Halpern-Manners, J. M. Raymo, and A. Palloni (2015). Do different methods for modeling age-graded trajectories yield consistent and valid results? *American Journal of Sociology 120*(6), 1809–1856.

Widmer, E. D., R. Levy, A. Pollien, R. Hammer, and J.-A. Gauthier (2003). Between standardisation, individualisation and gendering: An analysis of personal life courses in Switzerland. *Swiss Journal of Sociology 29*(1), 35–65.