The Hidden Markov Model:

Practical guidelines regarding the sufficient length of a time series and the required degree of heterogeneity of the hidden states.

B.J. Boverhof

Supervised by E. Aarts

Utrecht University

The Hidden Markov Model: Practical guidelines

From the discipline of psychology to ecology and from the field of bioinformatics to economy: the Hidden Markov models (HMMs) provides a convenient framework for analyzing time series behavior. (Rabiner, 1989) (Visser, 2011) (Zucchini, MacDonald & Langrock, 2016). One analyzing time series data with the HMM does not have to rely on averaging cross-sectional data, but can rather model and study a process in its entirety (Visser, 2011). The HMM is flexible, stemming amongst other things from the fact that the observed sequence is not required to adhere to a single distribution: the sequence is rather allowed to be of multimodal nature (Zucchini et al., 2016) (Visser, 2011). Consequently, a time series adhering to many different distributions simultaneously can be modelled with the HMM, rendering the framework particularly flexible, and hereby suitable, for a wide range of applications within a wide range of different disciplines. These distinctive disciplines include for instance the field of speech recognition, in which the HMM is utilized as signaling model (Rabiner, 1989), the field of bioinformatics in which the HMM is utilized to analyze genome sequences (Eddy, 1998), the field of econometrics and finance in which the HMM is utilized to predict the volatility of stock markets (Mamon & Elliot, 2007), the field of ecology in which the HMM is utilized to model wildlife movement patterns (Langrock et al., 2012) and the field of behavioral neurosciences in which the HMM is utilized to model the behavior of mice (Aarts et al., 2015).

Despite the scientific potency of the HMM, still insufficient knowledge is available regarding the circumstances under which the model is able to perform properly, that is: able to derive accurate estimates. The present research will endeavor to explore the ability of the model to derive accurate estimates, given variations in (1) the length of a time series and (2) the degree to which there is a clear distinction between states. The latter will be endeavored by means of a simulation study, in which control can be exerted over the specification of the latter two variables. Ultimately, the aim is to define guidelines that contribute to model furnishing choices given variations of the aforementioned variables. More specifically, the aim is to define guidelines with respect to the required length of a time series and the required degree of heterogeneity, in order to be able to derive accurate estimates with the Hidden Markov model.

Case introduction

For clarification purposes, I will now introduce a straightforward and oversimplified case, with the objective of enhancing the understanding of the HMM. Please note that this example is fictional, solely adopted for clarification purposes. It is (arguably) common knowledge throughout the world that the footwear of choice for the Dutch are wooden shoes. Let us assume this signature product is of such vital importance that the Dutch economy thrives on the disposal of it. We assume that the observed wooden shoe sale at a particular moment in time is governed by the state of the Dutch economy at that same moment. Such a relationship seems plausible in regards of basic economic principles, for in times of high prosperity the Dutch people would possess more recourses allowing them to purchase additional pairs of wooden shoes (possibly a different pair for each occasion), hereby increasing the observed wooden shoe sale. Hence, we assume the degree of economic prosperity at a particular moment in time to govern the disposal of wooden shoes, which is for the sake of simplicity categorized as either being low, average or high. Additionally, for simplicity we define the level of economic prosperity in a binary manner, either as being prosperous or unprosperous.

Besides the relationship between state (economic situation) and observed behavior (*disposal of wooden shoes*), we furthermore presume a second dependency, namely one between the states at different moments in time. We presume that the state of the economy at a particular moment in time is dependent upon the state of economy at the preceding moment in time. Theoretically such a relationship would seem plausible: for it is more likely that an economically prosperous moment in time is followed up by another prosperous moment. A state, such as the state of economy, is often a stable and rigid phenomenon that does not change every other moment. The aforementioned relationships are schematically depicted in Figure 1, wherein $S_{1:T}$ denote the states at moments *1* to T and $X_{1:T}$ denote the observed behavior at moments *1* to T.

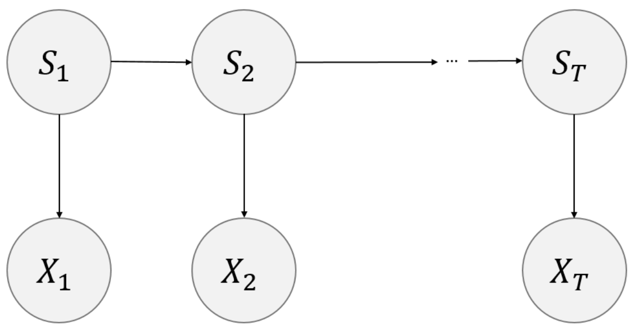

Figure 1: Dependency graph of the Hidden Markov Model.

The three components of the Hidden Markov Model

As was theoretically elaborated upon within the previous section, the exemplarily case is designed such that the series of observed behavior (disposal of wooden shoes) is governed by the unobserved underlying state (economic situation). This relationship is of probable nature, for the wooden shoe sale might be dependent on alternative factors than merely the economic situation, such as for example the popularity of alternative footwear. The probabilistic relationship between state and observed behavior is one out of the three components of the HMM*,* governed by the conditional distribution matrix A*.* The latter probabilities describe that, given residing in a state ($S = \text{prosperous}$ or $S = \text{unprosperous}$), what are the probabilities to observe behavior $X = \text{low disposal}$, $X = \text{medium disposal}$ or $X = \text{high disposal}$:

$$A = \begin{pmatrix} 0.60 & 0.35 & 0.05 \\ 0.06 & 0.30 & 0.64 \end{pmatrix}.$$

From A it can be discerned that an unprosperous moment in time is accompanied with a relatively high probability ($p = .60$) to observe a low disposal of wooden shoes, whereas a prosperous year is accompanied with a relatively high probability ($p = 0.64$) to observe a high disposal. It is important to note that the states of the HMM are by definition unobserved, or hidden, in the sense that they are not and cannot be directly observed. Suppose it would have been possible to measure the sequence of states directly, or infer these with determination from the observed sequence, the application of the HMM would render redundant (Visser, 2011). Concluding this section, we defined the observed behavior of the Hidden Markov model to be dependent upon the hidden state in a probable manner, denoted by Visser (2011) as the distribution function:

$$f(X_T|S_T = i), \text{ or } f_i(X_T).$$

Additionally, the states of the HMM are dependent upon their preceding state: they adhere to a Markovian process (Rabiner, 1989) (Visser, 2011). In regards of the aforementioned exemplary case, this would imply that the economic situation at time T is dependent upon the economic situation at time T_{-1}. Equally so, this process is probabilistic in nature, and provides the second out of the three components of the HMM*,* described by the transition distribution matrix B (Rabiner, 1989). B contains the probabilities that, given the state ($S = \text{prosperous}$ or $S = \text{unprosperous}$) at moment T_{-1}, what is the probability to remain in that state, or transition to the other state at time T:

$$B = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix},$$

denoted by Visser (2011) as:

$$P(S_T|S_1, \ldots, S_{T-1}) = P(S_T|S_{T-1}).$$

It becomes apparent from this expression that the state S_T is solely dependent upon the preceding state S_{T-1}. Restricting to this assumption is, however, not paramount, for the flexible nature of the HMM allows for the definition of so-called higher order models, wherein S_T may be dependent upon multiple preceding states (Visser, 2011).

At last, the first state of the sequence S_1 is specified a priori by the researcher and is known as the initial probability (Rabiner, 1989). The initial distribution C provides the probability of the initial state of the sequence to be either one state, or the other ($S = \text{prosperous or}$ $S = \text{unprosperous}$):

$$C = \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix}.$$

Guidelines concerning the furnishing of the Hidden Markov model

Despite the wide application and vast body of extensions build upon the base model, still little is known about the circumstances under which the base model is able to provide accurate estimates. In fact, elemental guidelines concerning model furnishing are until today non-existent within the literature. The objective of this paper is, therefore, to explore and provide such guidelines. With these guidelines I endeavor to provide researchers with a sense of insight concerning practical choices to be made with respect to model furnishing, specifically concerning (1) the length of the time series, (2) the degree to which a clear distinction between the states is required to exist and (3) potential interdependencies between the latter two.

It is expected that the length of a time series is positively related towards the accuracy of the model estimations: to a certain extent more data usually invokes more accurate estimations. The extension of data is expected, however, to lose its significance at a certain threshold, equally so as is the case with sample sizes in conventional statistics (Toepoel, 2016). I will endeavor to define a framework with respect to the length of a time series that is required in order to generate accurate estimates, by inquiring:

1. *What is the required length of a time series in order to achieve accurate estimates when conducting the Hidden Markov model?*

I expect the second parameter, the degree to which a clear distinction between the states exists, to be positively related towards the accuracy of the model estimation. Whenever the observed patterns of behavior are closely resembling one and another, one could expect the model to struggle more in the allocation of the correct state sequence, hereby negatively influencing the accuracy of the model over a number of repetitions. Vis-à-vis, one could expect the model to derive more accurate estimates whenever the states are clearly distinguishable. I will endeavor to gain insight in the extent to which (the absence of) a clear distinction between the states impacts the accuracy of the model, by inquiring:

2. *What is the required degree of distinctiveness between the states of the Hidden Markov model for achieving accurate estimates?*

It is important to realize that the answers on both research questions might be interdependent. For example, one could expect that in situations wherein there is an unclear distinction between states, a relatively long time series might be beneficial in order to derive accurate estimates. On the other hand, for a situation wherein there is a clear distinction between the states, a shorter time series might suffice. It might be important to acknowledge the existence of such potential interdependencies, for which I will also attempt to gain insight into model accuracy in relation to one of the delineated variables, given that the other is not kept constant, but rather is specified to be expectantly disadvantageous as opposed to expectantly advantageous with regards to providing accurate estimations. The latter will be endeavored by inquiring:

3. *What is the required length of the time series of a Hidden Markov Model in order to achieve accurate estimates given the distinctiveness between states is either unclear or clear?*

and:

4. *What is the required degree of distinctiveness between the states of the Hidden Markov model in order to achieve accurate estimates given the length of the time series is either short or long?*

Simulations

Simulation design

I will attempt to answer the proposed research questions by means of a simulation study. A number of different scenarios will be delineated, each varying in their specification in the length of their observed data sequence T, and degree to which a clear distinction between states λ, exists. On behalf of each scenario, a data sequence will be simulated accordingly the specification of both T and λ belonging to that particular scenario. Subsequently, the model will be applied on each data sequence, replicated either 250 or 500 times. Estimations of these replications will be averaged, resulting in a single estimate for each parameter within each simulation scenario. Accuracy will be assessed by contrasting these estimated parameters with their true counterparts, endeavoured by three different measures. Ultimately these measures will be contrasted across the different simulations, with which I will attempt to answer the research questions as put forward in the preceding section. T will be specified to vary alongside a range of five values on a spectrum ranging from short to long. λ will also vary alongside three values, ranging from unclear to clear.

Table 1: Simulation scenarios.

		T: The length of the observed time series (X_T)				
		Short	Short- medium	Medium	Medium - Long	Long
λ: Degree to which there is a clear distinction between states	Unclear	Simulation 1	Simulation 2	Simulation 3	Simulation 4	Simulation 5
	Moderately - Clear	Simulation 6	Simulation 7	Simulation 8	Simulation 9	Simulation 10
	Clear	Simulation 11	Simulation 12	Simulation 13	Simulation 14	Simulation 15

RQ1, concerning the specification of T in relation the accuracy of the estimated model, will be addressed by keeping constant (that is moderately clear) λ, consequently comparing the accuracy of the estimated model for the simulations 6, 7, 8, 9 and 10. *RQ*2, exploring the specification of λ in relation to the accuracy of the estimated model will be addressed in an equal manner, that is by keeping constant T on medium, consequently comparing scenarios 3, 8 and 13. *RQ*3, exploring the benefit of T given λ is unclear as opposed to clear, will be addressed by comparing the simulations 1,

2, 3, 4, 5 and 11, 12, 13, 14, 15. Ultimately, *RQ*4 exploring the benefit of λ given T is either short or long, will be assessed by comparing the simulations 1, 6, 11, 2, 7, 12 and 4, 9, 14, 5, 10, 15.

Model specification

Each simulation will inhibit an observed sequence of a categorical variable X, varying accordingly five different categories, observable in a fixed number of 3 states. As was elaborated on in the preceding section, the length of the observed time series T will vary across the simulations accordingly to being short, short to medium, medium, medium to long and long. Values for these variations will be specified as $T = 900, T = 1200, T = 1800, T = 2700$ and $T = 3600$ respectively. The degree to which a clear distinction between the states exists λ, can't be specified as input aforetime the simulations in a direct fashion such as T, and consequently calls for an indirect approach. This will, therefore, be conducted by manipulating the conditional distribution matrix A. The simulations wherein a clear distinction between the states exists will be defined such as a situation in which each category of X reflects a high probability to be observable in one or two states, and less so in the remaining states:

$$A_{clear} = \begin{pmatrix} 0.72 & 0.07 & 0.07 & 0.07 & 0.07 \\ 0.04 & 0.44 & 0.44 & 0.04 & 0.04 \\ 0.04 & 0.04 & 0.04 & 0.44 & 0.44 \end{pmatrix}.$$

Simulations with a moderately clear distinction between states will be defined such as a situation in which each category of X is moderately likely to be observable in one or two states, but this distinction is less profound as compared with the clear distinction:

$$A_{moderately-clear} = \begin{pmatrix} 0.56 & 0.11 & 0.11 & 0.11 & 0.11 \\ 0.10 & 0.36 & 0.36 & 0.09 & 0.09 \\ 0.09 & 0.10 & 0.09 & 0.36 & 0.36 \end{pmatrix}.$$

At last, an unclear distinction between states implies a situation where in each category of X is still observable in one or two states, however this distinction renders increasingly blurred:

$$A_{unclear} = \begin{pmatrix} 0.40 & 0.15 & 0.15 & 0.15 & 0.15 \\ 0.13 & 0.30 & 0.30 & 0.13 & 0.14 \\ 0.14 & 0.13 & 0.13 & 0.30 & 0.30 \end{pmatrix}.$$

The transition matrix B will be held constant across all simulations, and is specified as:

$$B = \begin{pmatrix} 0.80 & 0.10 & 0.10 \\ 0.05 & 0.80 & 0.15 \\ 0.00 & 0.10 & 0.90 \end{pmatrix}.$$

The initial distribution will also be held constant across all simulations, specified as:

$$C = \begin{pmatrix} 0.33 \\ 0.33 \\ 0.34 \end{pmatrix}.$$

The application of the HMM inhibits a trajectory embodying two different facets: (1) deriving the probabilities of the states at each moment in time T, summarized in the conditional and transition distribution matrices, and (2) optimize these probability matrices by iterating multiple times. The most probable state sequence will be derived by means of the Forward probability algorithm, estimating the most probable state sequences at times time T_1, T_2, …, T_{-1}, governing the observed sequence at that the same moments in times. This is a product of both the conditional and transition probabilities (Rabiner, 1989). After determining the most probable state sequence, a Gibbs sampling algorithm will sample states in a backwards manner and update the probabilities contained in the conditional and transition distribution matrices accordingly (Rydén, 2008) (Scott, 2002) (Please refer to Scott (2002) for a more extensive evaluation of the Gibbs sampling procedure for the HMM). For a schematic depiction of the estimation procedure I redirect the reader to figure 20 in appendix 5. The adoption of such a Bayesian estimation approach, that is by employing a Gibbs sampling algorithm, implies by definition the specification of a prior. A Dirichlet distribution will be allocated to each row of both matrices, in which the probabilities will be allocated commensurately, that is in a non-informative manner. Such a procedure is conventional practice in case of the specification of a non-informative prior (Rydén, 2008). The reason for adopting a non-informative prior resides in the fact that little research such as the present has preceded, for which the specification of an uninformative prior is most appropriate.

The previously delineated procedure of forward estimation and backwards sampling will be iterated a total of 4000 times per simulation, after which each parameter estimation will be derived by averaging the estimations of each iteration, resulting in a single estimation of all parameters per simulation. A burn-in period of 1000 will be specified, for the result of the first 1000 iterations might be deceptive due to a state of un-convergence, hereby producing non-meaningful estimates. The number of conducted replications will differ per scenario due to reasons of preservation of computational power, for the expectantly easier scenarios are likely to be more consistent in their estimations, consequently requiring less replications. All scenarios inhibiting sequential length of $T = 900$ and $T = 1200$ will embody a total of 500 replications, whereas the remainder scenarios inhibiting a length *of* $T = 1800$, $T = 2700$ and $T = 3600$ will be replicated 250 times.

Simulation results

Model evaluation measures

Three different measures will be consulted in order to evaluate model accuracy. Accuracy of the conditional and transition distributions will be evaluated and visualized separately. The first adopted measure is the bias*,* evaluating the difference between an estimated parameter and its true counterpart, calculated as:

$$\frac{1}{Z} \times \sum\nolimits_{z\,=\,1}^{Z} (estimated\ parameter\ value_z - true\ value).$$

The second adopted evaluation measure is the Root Mean Square of Approximation (RMSEA). The RMSEA of a parameter resembles the bias closely, differing in the fact that it also includes variance. The RMSEA is calculated as following:

$$\sqrt{\frac{1}{Z} \times \sum\nolimits_{z\,=\,1}^{Z} (estimated\ parameter\ value_z - true\ value)^2}.$$

Two variations on the Bias and RMSEA measures will be consulted, the first of which is simply the absolute measure, reflecting the difference between estimated and true probability in terms of absolute probability. Additionally, the absolute variant of both biases and RMSEA will be supplemented by a relative measure. The relative bias and RMSEA of a parameter are standardized, achieved by dividing the absolute measure by its true counterpart, resulting in a measure describing the proportional miss-estimation of a parameter in contrast to its true counterpart, aiming at an Enhancement of the level playing field with respect to the comparability of simulations that include differently sized true parameter values. Parameter *S3* to *S1* contained in the transition distribution matrix is excluded due to a misspecification aforetime the simulations.

The third and final evaluation measure is the coverage of the 95% credibility interval, reflecting the proportion of replications wherein the true parameter value fell within the 95% estimated credibility interval.

The occurrence of convergence has been assessed for several iterations of several models at random and was observed to be positive.

Bias

Depicted in figure 2 are the relative and absolute biases per parameter of the conditional distribution. Measures for each variation in clearness of the distinction between states are depicted in a separate plot. For an alternative visualization, that is in a different plot for each variation in length, grouped on the distinction between states, I redirect the reader to figure 7 in appendix 1.

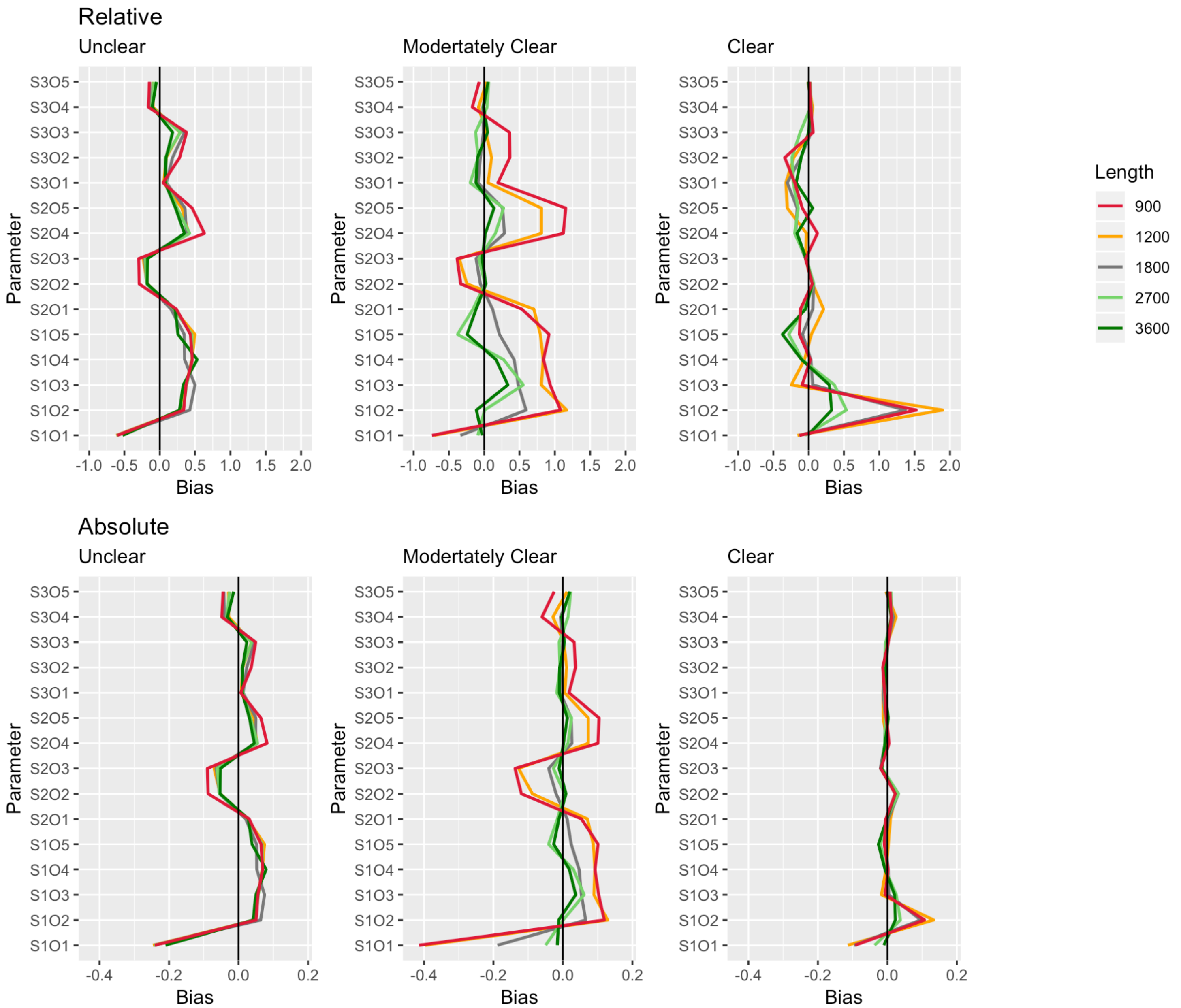

Figure 2: Biases of the Conditional Distribution: Grouping on clearness of the distinction between states. Values on the x-axis of the top row plots are biases in terms of absolute probability. Values on the x-axis of the bottom row plots are biases standardized proportional to the size of their true value

Depicted in figure 3 are the relative and absolute biases per parameter of the transition distribution. Measures for each variation in clearness of the distinction between states are depicted in a separate plot. For an alternative visualization of the same data, that is in a different plot for each varying length I redirect the reader to figure 8 in appendix 1.

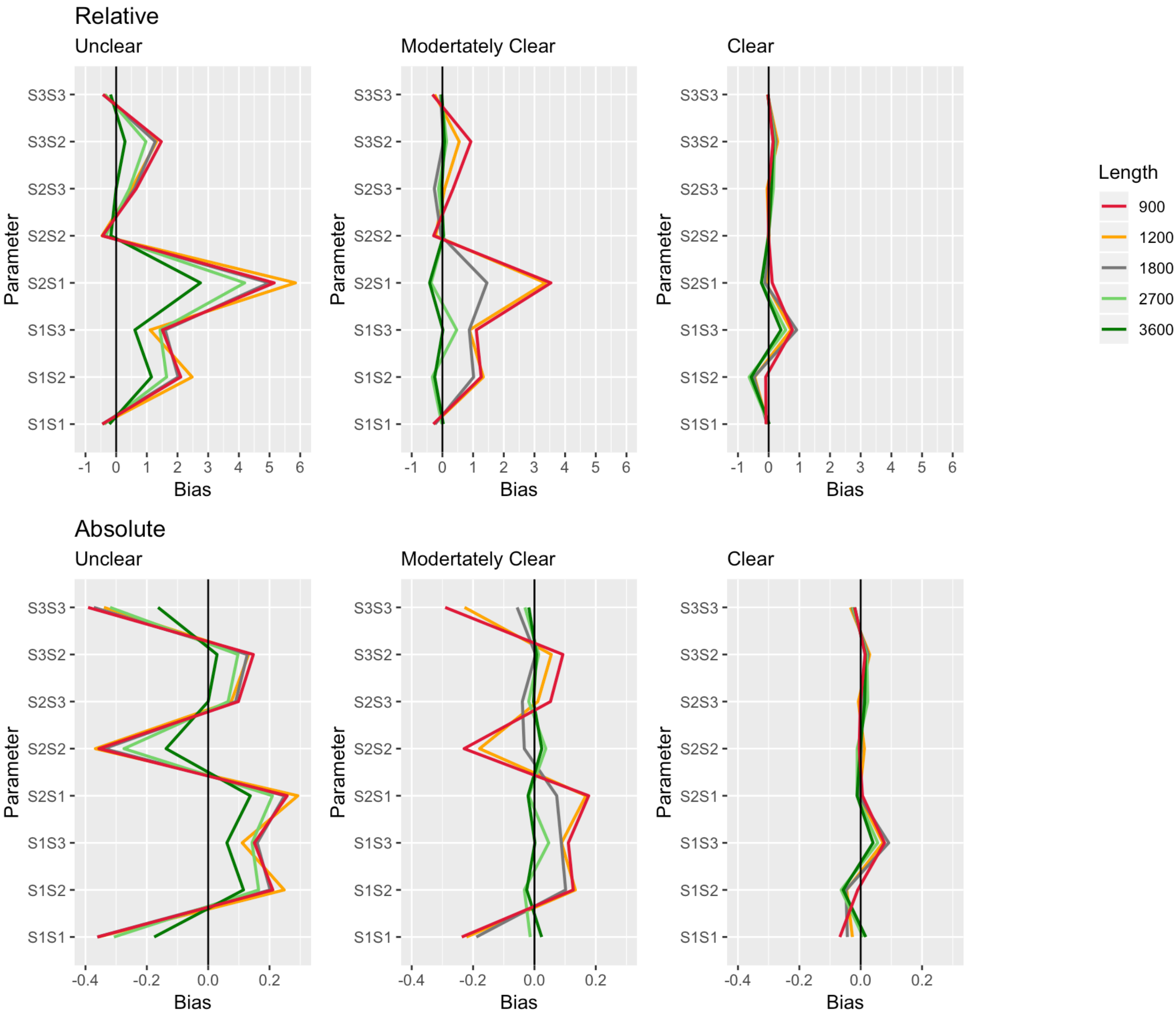

Figure 3: Bias Transition Distribution: Grouping clearness of the distinction between states. Values on the x-axis of the top row plots are biases in terms of absolute probability. Values on the x-axis of the bottom row plots are biases standardized proportional to the size of their true value.

Consulting figure 2, relative biases of the conditional distribution are observed to be ranging from a minimum of approximately -0.75 to a maximum of approximately 2.0, wherein -0.75 reflects an underestimation of 75%, and 2.0 reflects an overestimation of 200% as opposed to the true parameter value. The window of error is wider for estimations of the transition distribution, ranging from a minimum underestimation of around 50% to an overestimation of maximum 600%. Absolute biases of the conditional distribution range from a minimum underestimation of 0.4 to a maximum overestimation of around 0.16. Again, the bias window for the transition distribution is more expansive, ranging from an underestimation of approximately 0.4 to an overestimation of about 0.25. The general observable tendency is that of models encompassing a clear distinction between states to produce substantially less biased estimates. This tendency is clearly distinguishable, both for the absolute and relative variations, equally visible in the conditional and transition distribution. Biases of the models with a clear distinction between states have rendered substantially small as compared with models of moderately clear and unclear distinction, with under- and over – estimations ranging from -0.05 to 0.05 for all but two parameters. A considerable reduction of bias for the models with clear distinction, as opposed to models with moderately clear and unclear distinction is equally observable when consulting estimations of the transition distribution. Absolute biases have shrunken to range from approximately -0.08 to 0.08. A comparison of the simulations with different lengths resembles another clearly distinguishable tendency: the models of lengths 2700 to 3600, coloured by shades of green, produced almost exclusively the least biased estimates. This tendency is particularly profound for models that encompass a moderately clear distinction between states, with observable absolute biases of both the conditional and transition distribution approximately halve the sizes as compared with the models of length 900 and 1200.

By inspecting relative biases of the conditional distribution more closely, it becomes apparent that models of lengths 900 and 1200, including a moderately-clear distinction, produced severely biased results. Somewhat to the contrary, these biases range even higher than their counterparts of the same length with unclear distinction. Overestimations of the conditional distribution range from 75% to 100% for a considerable number of parameters. Equally so, absolute biases of the latter models are of higher magnitude as compared with their counterparts with unclear distinction. Why is it the models of lengths 900 and 1200 with a moderately clear distinction between states produce higher biases, as compared with their counterparts including an unclear distinction? In order to explain this, I redirect the reader to figure 11, 12, 13 and 14 in appendix 2, displaying the estimates per parameter for all four models. From these it becomes apparent that estimations of the 4 models have something in common: all wrongly estimate numerous parameters with great uncertainty, the last of which is illustrated by the wide spread of variance. This is especially profound for the observations in state 1 and 2. Figure 15 in appendix 2 contrasts averaged estimations of simulation 1, reflecting the model with unclear distinction and length 900, with the estimations of simulation 6, reflecting the model with moderately clear distinction and length 900, both contrasted contrary to their true values. From this it becomes apparent that estimation of, especially probabilities of states 1 and 2, of both models display approximately equal estimates, and consequently wrongly estimate the latter parameters in terms of equal uncertainty. It is important to realize, however, that by definition of the bias measures the model with moderately clear distinction gets punished more severely for the latter miss estimation. The reason for this resides in the fact that the moderately clear model includes more extreme, that is either very small or very high, true parameter values. To even further illustrate this claim, take the example of parameter S1O5. Both the unclear and moderately clear models estimated this parameter to reflect a probability of approximately $P = 0.22$. The true value of this parameter for the unclear model equals $P = 0.15.$, resulting in an absolute bias of $P = 0.07$. The true value of parameter S105 for the moderately clear model equals $P = 0.11$, which would result in an absolute bias of $P = 0.11$. Hence, in a situation of equal uncertain estimations, models including more extreme true parameter values are punished more severely in terms of higher biases. Consulting of the standardized variations does not omit this problem, for relative biases of $P = 0.46$ and $P = 1$ are observed respectively.

Despite all models including a clear distinction portraying low biases in general, parameter S1O2 of the conditional distribution inhibits disproportionally high relative biases for the models of length 900, 1200 and 1800, ranging from an overestimation of approximately 150% to approximately 200%. Estimations for the conditional distributions per parameter of the models with a clear distinction and length 900, 1200 and 1800 can be consulted as figure 16, 17 and 18 respectively, in appendix 3. It should be noted that such overestimations might be inflated due to the procedure of normalisation. Parameters inhibiting an outstandingly small true probability tend display huge relative biases, for they are punished more severely as compared with a parameter inhibiting a true probability of higher magnitude. Therefore, interpretation of relative biases should always be contrasted with their absolute counterpart, from which doing so it becomes apparent that biases of parameter S1O2 are substantially reduced, however, still persist, indicating a systematic overestimation of solely parameter S1O2 within the clear models of length 900, 1200 and 1800. The same tendency is able to explain the disproportionally high relative bias of parameter S2 to S1 of the transition distribution, with a true parameter value as low as 0.05.

A final conspicuous observation can be made when comparing the magnitude of biases for the conditional distribution with those of the transition distribution. The window of error for both the relative and absolute bias of the transition distribution range substantially wider in contrast with the window of error displayed for the conditional distribution. Again, this can be explained by taking notice of the true parameter values. True values of the transition distribution are of substantially more extreme magnitude, as compared with those of the conditional distribution, producing biases of more voluminous magnitude due to the nature of the bias measure.

RMSEA

Depicted in figure 4 on the top row are the relative RMSEA measures for the conditional and transition distribution respectively, supplemented by the absolute RMSEA on the bottom row. All have been grouped on the degree to which a clear distinction between the states exists. The RMSEA measures of al parameters within each simulation have been averaged, resulting in a single measure per simulation. For an alternative visualization of the same observations, that is when not grouping on clearness of the distinction between states, but rather on the length of the time series, I redirect the reader to figure 9 in appendix 1.

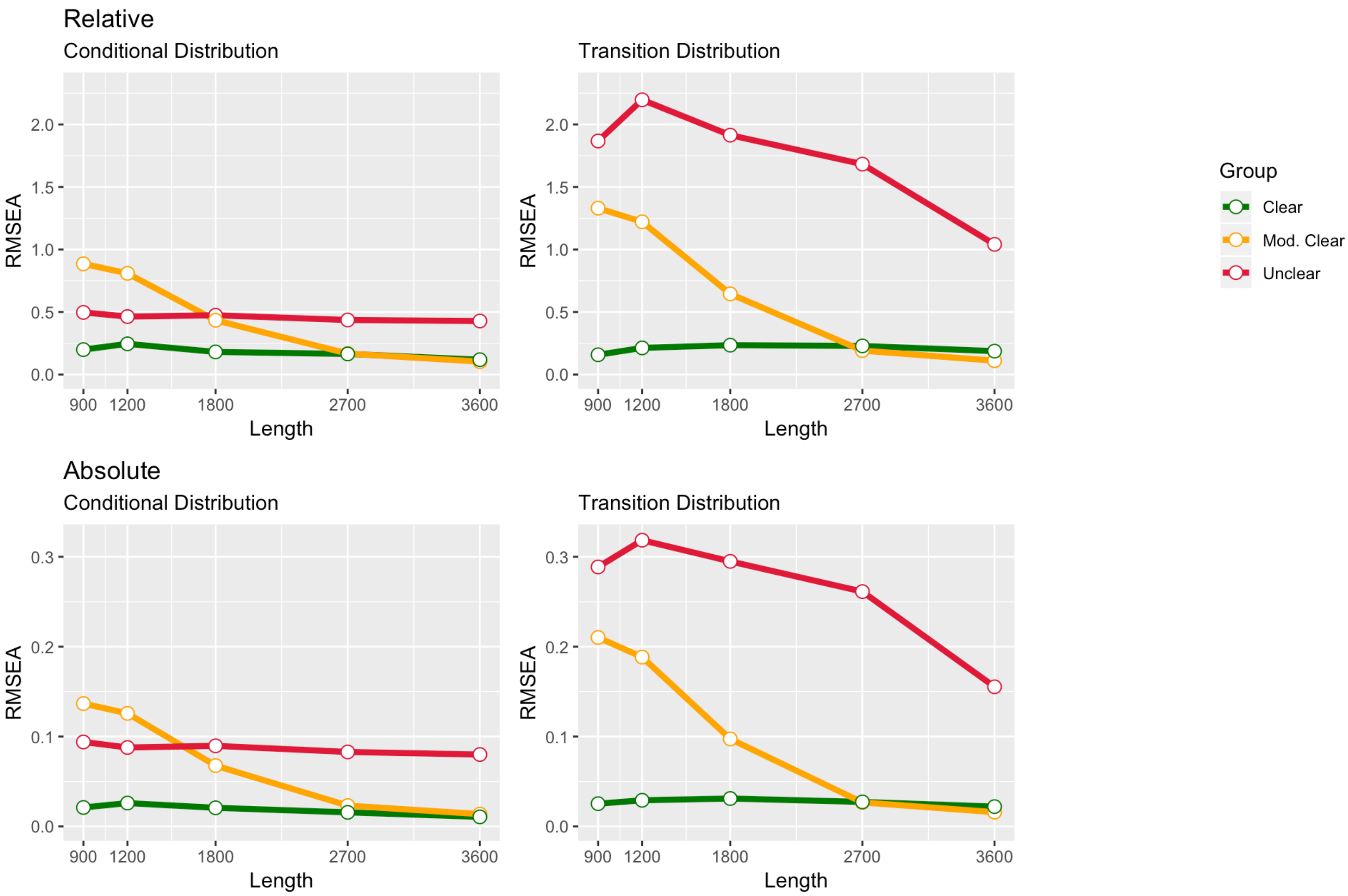

Figure 4: Relative and absolute Root Mean Square Error of Approximation for the conditional and transition distribution. Grouping on degree of distinctiveness between the states.

From figure 4 can be discerned that the models with clear distinction between states are barely benefitting from a longer sequence, with observable absolute RMSEA measures ranging under the threshold of approximately 0.03 for both the conditional and transition distribution estimates, regardless of length. This is in harmony with the biases we have observed in the preceding section: models with clear distinction were able to produce low biases, regardless of length. This is equally so visualized by figures 16, 17 and 18 in appendix 3, depicting the parameter estimates for a selection of models with clear distinction. All estimations of the latter models display very modest bias, in addition with exceptionally high certainty, the last of which is reflected by the low variance.

Models with a moderately-clear distinction do, on the contrary, benefit from a longer sequence, with a stark decline in relative and absolute RMSEA from length 900 to 2700. Again, this benefit is more substantive for estimations of the transition distribution, with an observable averaged absolute RMSEA of 0.22 at length 900 decreased to approximately 0.03 at length 1800 and onwards. A similar, however less profound, tendency is observable for the conditional distribution, with an averaged absolute RMSEA of approximately 0.14 at length 900 decreasing to around 0.03 at length 1800 and onwards.

At last, RMSEA measures of the conditional distribution for the models inhibiting an unclear distinction between states are observed to be of higher magnitude. Little accuracy is gained by incrementing length, with an averaged absolute RMSEA of approximately 0.09 for the model or length 900, in contrast with an averaged RMSEA of approximately 0.08 for the model with length 3600. On the contrary, much benefit is gained from incrementing length with respect to estimations of the transition distribution. Absolute averaged RMSEA measures are observed to have decreased from approximately 0.29 in the model of length 900, to around 0.16 for the model of length 3600. Despite this benefit, a RMSEA of considerable size still persists, which is in consensus with bias measures observed in preceding sections.

Ultimately, it becomes apparent that RMSEA measures display estimations of the transition distribution to be substantially less accurate in contrast to those of the conditional distribution. This tendency was equallo so distinguished when observing biases in the previous section, with biases of the transition distribution of considerably higher magnitude.

Coverage

Depicted in figure 5 are the coverage measures for both the conditional and transition distribution, grouped on the degree to which a clear distinction between states exists. The coverage measures of all parameters per simulation have been averaged, such that each simulation is represented by a single measure. For an alternative visualization of the same observations, that is when not grouping on clearness of the distinction between states, but rather on the length of the time series, I redirect the reader to figure 10 in appendix 1.

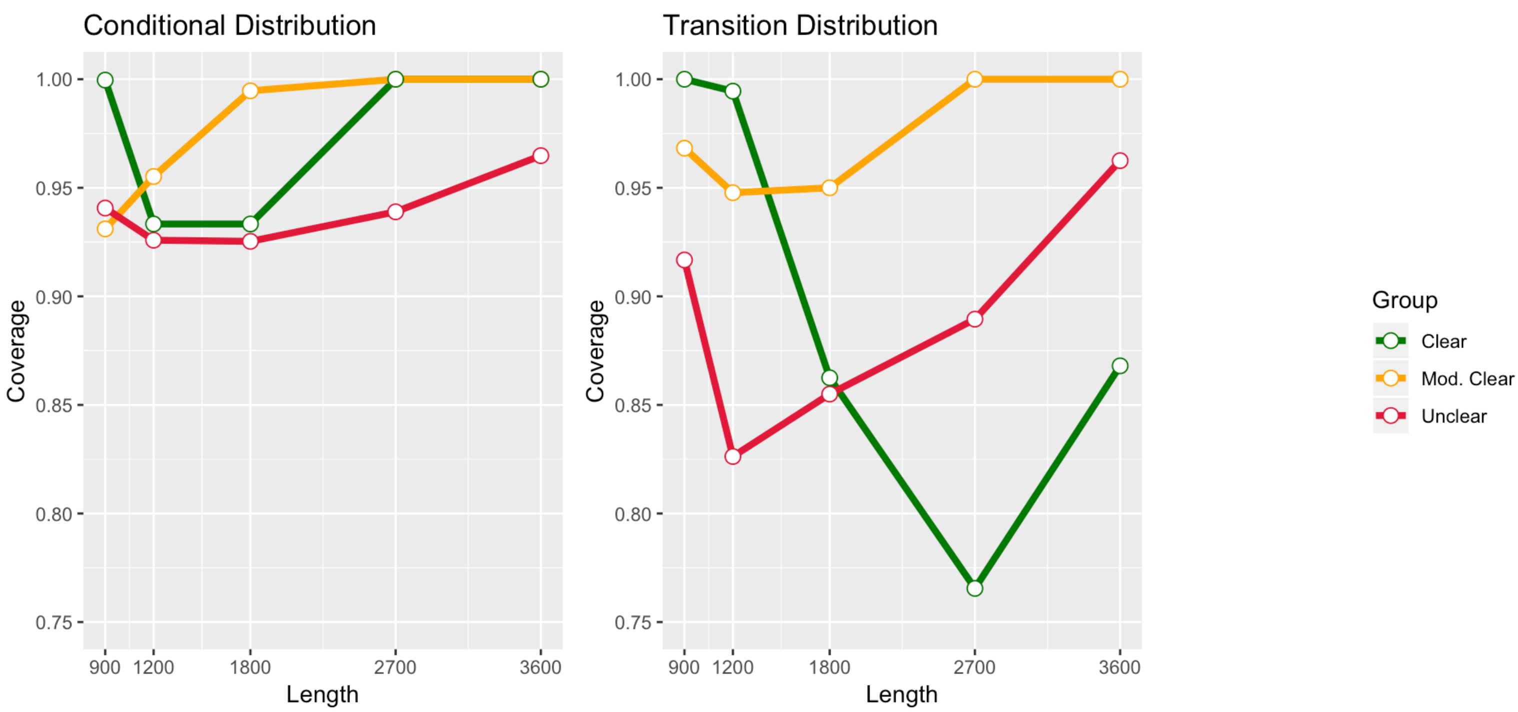

Figure 5: Averaged coverage of estimations within the credibility interval for the conditional distribution and transition distribution. Grouping on degree of distinctiveness between the states

First and foremost, coverages of the conditional distribution are all observed to be above the respectable threshold of 92%. A longer sequence is observed to generally increase coverage. Not merely the coverage measures of the clear models are of respectable size, coverage of the models with unclear distinction display approximately equally respectable results. Additionally, coverages estimates of the transition distribution range decently high for a selection of models with unclear distinction, as well as all models with moderately clear distinction. This is, however, not in harmony with results we have observed in preceding sections. For example, models of all lengths with unclear distinction between states have produced biases and RMSEA measures of substantive proportions, both salient for estimations of the conditional, as well as estimations of the transition distribution. The same holds true for the models of shorter length and moderately clear distinction between states. These disparate observations call for further exploration, which will be done so by means of figure 6, displaying the averaged sizes of the credibility interval per simulation.

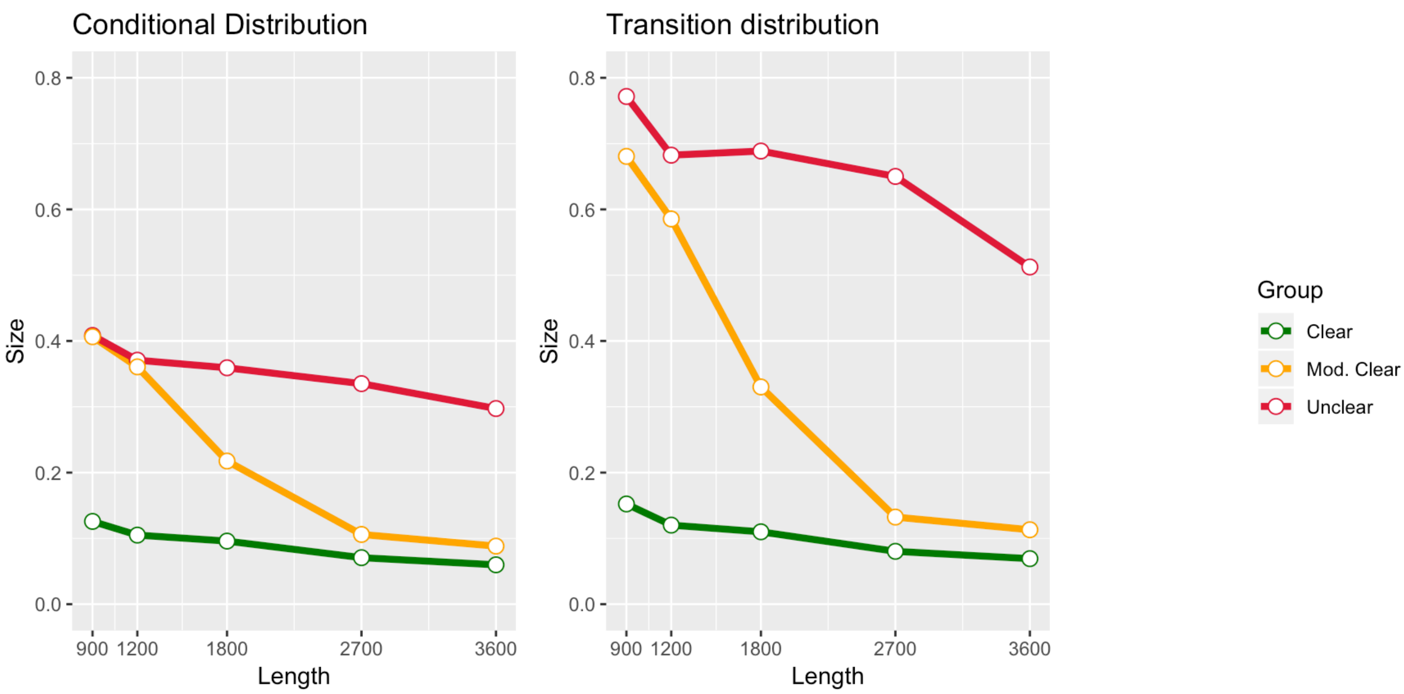

Figure 6: Averaged size of the credibility interval for the conditional and transition distribution

As can be discerned, especially the models with unclear distinction display credibility intervals of enormous proportion, with those of the transition distribution displaying sizes of around 0.7 for the shorter models. Credibility intervals of such spectacular size are also observable for the models of lengths 900 and 1200 with moderately clear distinction. Despite the sizes of the credibility intervals for estimations of the conditional distribution being of lower magnitude as compared with those of the transition distribution, still intervals of spectacular sizes are observable. Naturally, such vast credibility intervals would translate into high coverage measures: even so for estimations that encompass severe over- or under –estimation, which could still easily fall within intervals of such vast volume. Coverages measures for the models including such credibility intervals are therefore not a valid indication of model accuracy.

Coverages measures of the models with clear distinction also reflect somewhat conspicuous results. The model with a clear distinction and length 900 displays 100% coverage, dramatically dropping to approximately 76% for the model of length 2700 for estimates of the transition distribution. Another drop, however somewhat less dramatic, is observable when consulting coverage measures of the conditional distribution. The model of length 900 displays a 100% coverage, which has decreased to approximately 92% for the models of length 1200 and 1800.

Explanation for these findings also resides in the sizes of the credibility intervals, as depicted in figure 6. As can be discerned, sizes of the credibility intervals for models with a clear distinction between states are observed to shrink as length is incremented, starting at a size of approximately 0.18 for the model of length 900, and reduced to a size of approximately 0.14 for the model of length 3600. Is the detriment in coverage due to results getting increasingly less accurate, or could it rather be explained by the reduction in sizes of the credibility intervals? Figure 19 in appendix 4 provides an answer: it can be discerned that the averaged estimations of all parameters, of all models with clear distinction, are approximately the same. Additionally, all models closely resemble their true values. The observable drop in coverages for models with a clear distinction and increasingly longer sequences are therefore not an indication for increasingly less accurate estimates, but rather have to be a result of the decreasing sizes of the credibility intervals. The coverage measures as depicted in figure 5 should therefore be interpreted with caution, and always in full consideration of the sizes of the credibility intervals as depicted in figure 6.

Discussion

The present study has been conducted with the objective of exploring the ability of the Hidden Markov Model to derive accurate estimates, given a variation in circumstances. These circumstances encompass variations in (1) the length of the time series and variations in (2) the degree to which a clear distinction between states exists. Additionally, interdependencies between the later variations in relation with model accuracy have been explored. Furthermore, the case on which the HMM was applied was of relatively complex nature, encompassing a situation with 5 variations on the observed behaviour and 3 different states. This was endeavoured by means of a simulation study, wherein a total of 15 distinct data sequences have been simulated, each unique in its synthesis of the two aforementioned circumstantial variations. The HMM was applied on each of these data sequences, after which the accuracy has been evaluated by contrasting the resulting probability matrices with their true counterparts. Three different measures have been consulted, allowing for a cross-comparison across simulations. Four research questions have been defined in section 1, on which answers will be provided in the succeeding section.

The first research question inquired the following: *What is the required length of a time series in order to achieve accurate estimates when conducting the Hidden Markov model?* As was set forth in section 2, this question will be answered by contrasting the models varying in the length of their observed time series, while keeping constant (that is: moderately clear) the distinction between states. It has been observed that model of length 2700 was able to minimize both bias and RMSEA to a substantial degree, and no additional accuracy was accumulated by incrementing length to 3600. Hence, the required length of a time series for a model with 5 observations and 3 states in order to achieve accurate estimates equals a length of 2700.

The second research question inquired the following: *What is the required degree of distinctiveness between the states of the Hidden Markov model for achieving accurate estimates?* As was set forth in section 2, this question will be answered by contrasting the models with varying distinction between states, while keeping constant (that is: average) the length of the observed time series. It has been observed that the accuracy improved the most substantial amount for the model with a moderately clear distinction in contrast with the model inhibiting an unclear distinction. Accuracy is enhanced slightly more when distinction between states render clear, however the accuracy of the model with moderately clear distinction can already be deemed of a sufficient magnitude. Subsequently, the required degree of distinctiveness between states in order to achieve accurate estimates for a model with 5 observations and 3 states is moderately clear (reflected by the conditional and transition distribution matrices provided in section 2).

The third research question inquired the following: *What is the required length of the time series of a Hidden Markov Model in order to achieve accurate estimates, given the distinctiveness between states is either unclear or clear?*
It was observed that models inhibiting an unclear distinction between states benefit from a longer time series, however inaccuracy remained existent. The model encompassing the longest time series was still not able to derive accurate estimates, however increasingly accumulation in accuracy was gained as length was incremented. Hence, the required length of a time series given an unclear distinction between states is undefinable by means of the contemporary results, and calls for further research in which length is incremented even further than 3600. Additionally, a length of 900 for models encompassing a clear distinction between states was proven to be sufficient with respect to producing accurate estimates, for which the extent to which length could be decreased further remains undefined by the present research. Further research including models with a length shorter than 900 might therefore prove beneficial.

The fourth and final research question inquired the following: *What is the required degree of distinctiveness between the states of the Hidden Markov model for achieving accurate estimates given the length of the time series is either short or long?*
It was observed that models with a short length, that is length 900 and 1200, produced severely inaccurate estimates given both an unclear and moderately clear distinction between states. The models of the latter lengths and clear distinction, however, were able to minimize error, consequently deriving accurate estimates. The models of higher length, that is 2700 and 3600, were observed to display substantial accuracy given a moderately clear or clear distinction, as opposed to models of the latter length that inhibit an unclear distinction. Subsequently, we conclude that the required degree of distinctiveness between the states for a model of 5 observations and 3 states that inhibits a shorter length is clear, and the required degree of distinctiveness between the states for models of higher lengths is moderately clear.

Additional takeaways from the present research concern the three adopted evaluation measures. It was observed that the construction of the bias measure is dependent upon the size, or more particularly the extremity, of the true values. Since bias is also included in the RMSEA measure, the same argument upholds for the latter. By nature of these measures, parameters that inhibit more extreme true values are punished more severely, than would these true values be of less extreme size. For this reason, comparisons across simulations, but also the comparison of unequally sized parameters within simulations, should be conducted with care. Additionally, the coverage measure has not been proven of significant use in the present research, due to extremely sized 95%

credibility intervals. Due to this dependency, coverage measures should always be contrasted with the sizes of the credibility interval.

It should be noted that explored and defined guidelines in the present research are exclusively salient for the furnishing of a Hidden Markov Model including 5 observations and 3 states. The extent to which factors such as the number of varying observations or the number of states are related to model accuracy is unknown, however respective guidelines might differ for the latter models. Therefore, generalization of the delineated guidelines to models of different structure is not advised. The extent to which factors such as the number of varying observations or the number of states are related to model accuracy remains until the present unexplored, and could not be included in the contemporary research due to the limitations of both restricted time and computing power. This consequently calls for further research implications, exploring the relationships between the numbers of varying observations and the number of states with regards to model accuracy.

Literature

Aarts, E. (2016). Beyond the average: Choosing and improving statistical methods to optimize inference from complex neuroscience data.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics (Oxford, England)*, *14*(9), 755-763.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., & Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, *93*(11), 2336-2342.

Mamon, R. S., & Elliott, R. J. (Eds.). (2007). *Hidden Markov models in finance* (Vol. 460). New York: Springer.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257-286.

Rydén, T. (2008). EM versus Markov chain Monte Carlo for estimation of hidden Markov models: A computational perspective. *Bayesian Analysis*, *3*(4), 659-688.

Scott, S. L. (2002). Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association*, *97*(457), 337-351.

Toepoel, V. (2015). *Doing surveys online*. Sage.

Visser, I. (2011). Seven things to remember about hidden Markov models: A tutorial on Markovian models for time series. *Journal of Mathematical Psychology*, *55*(6), 403-415.

Zucchini, W., MacDonald, I. L., & Langrock, R. (2016). *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CR.

Appendix 1

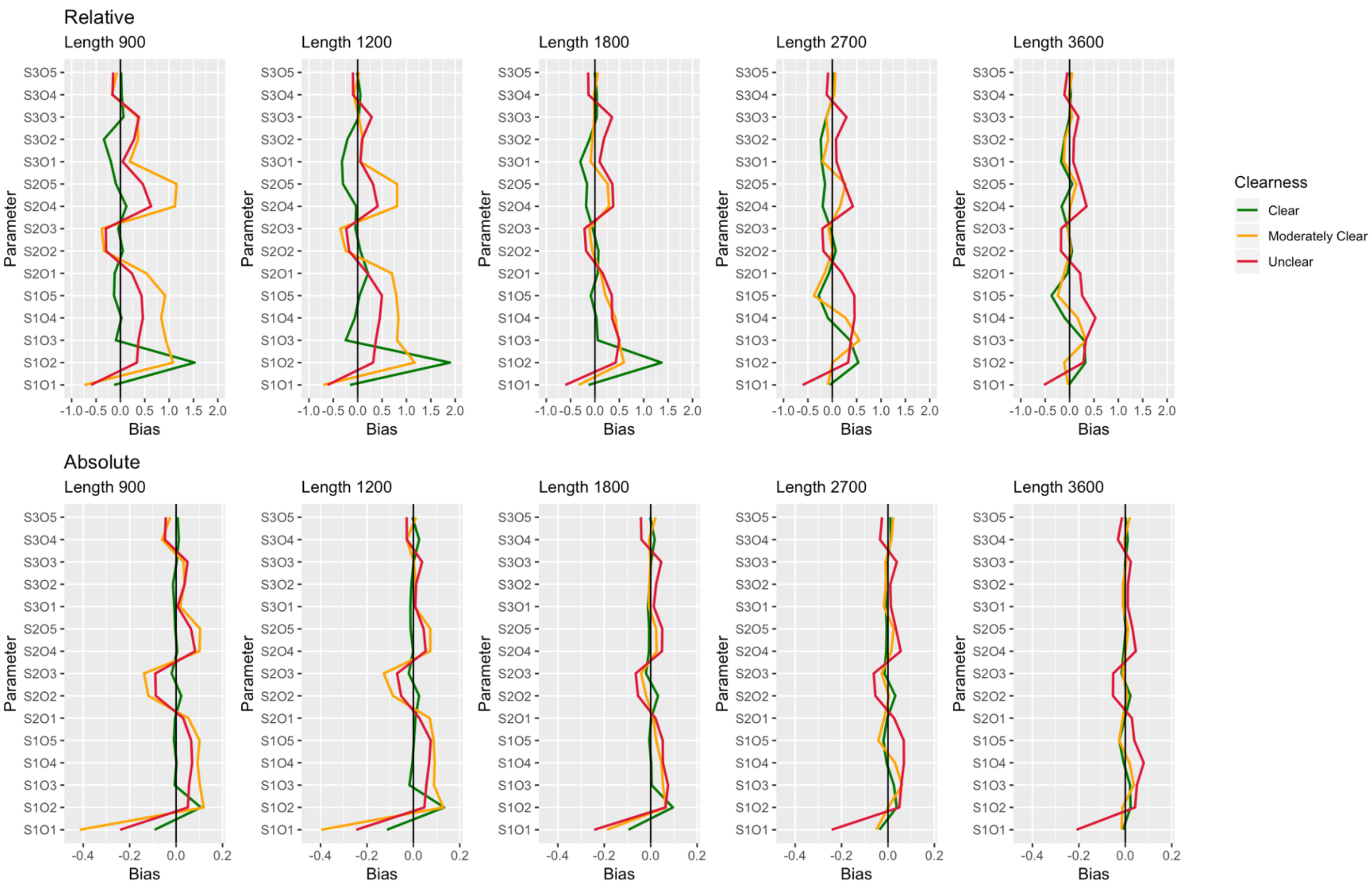

Figure 7: Bias Conditional Distribution: Grouping length of the time series. Values on the x-axis of the top row plots are biases in terms of absolute probability. Values on the x-axis of the bottom row plots are biases standardized proportional to the size of their true value

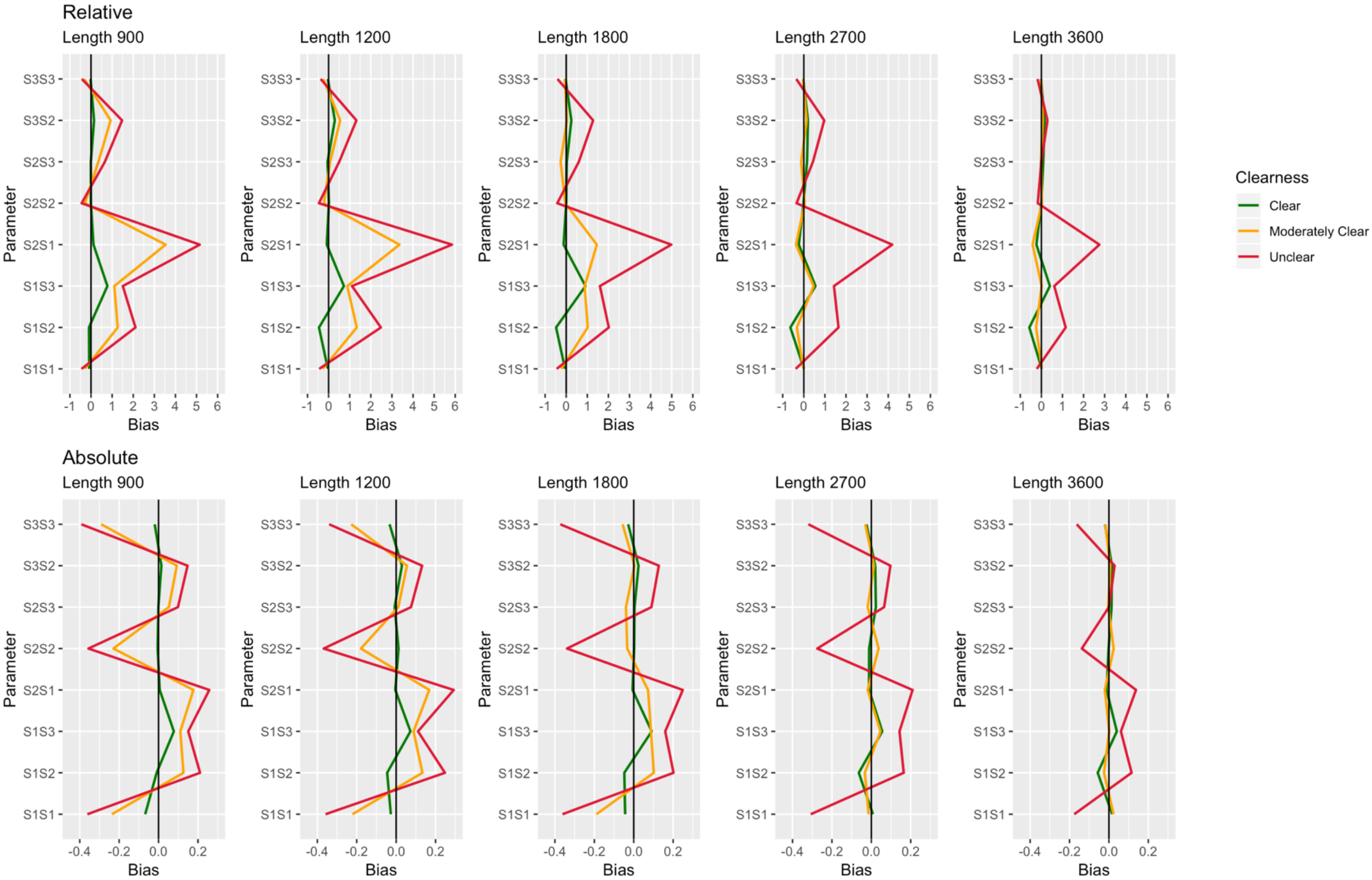

Figure 8: Bias Transition Distribution: Grouping length of the time series. Values on the x-axis of the top row plots are biases in terms of absolute probability. Values on the x-axis of the bottom row plots are biases standardized proportional to the size of their true value

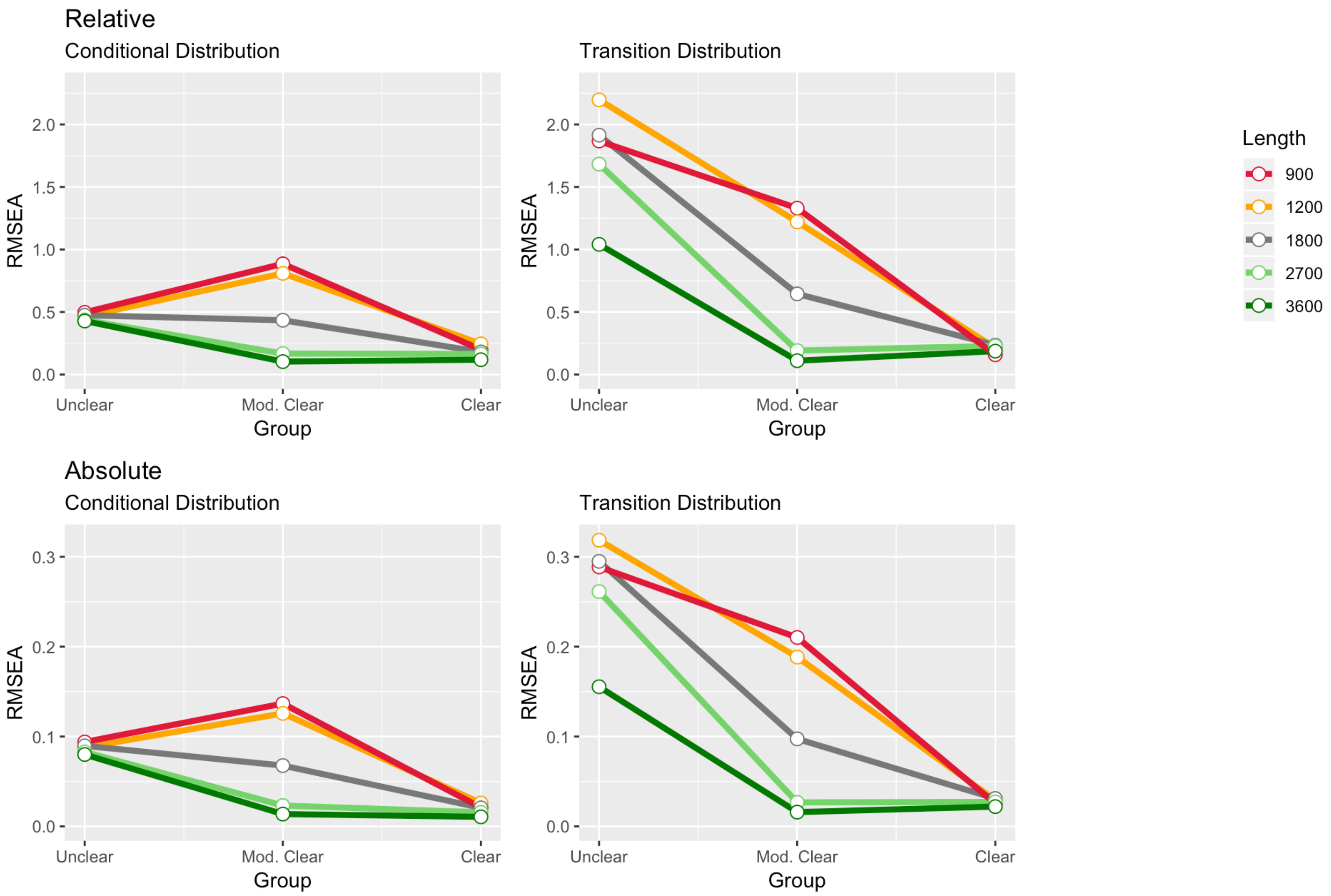

Figure 9: Relative and absolute Root Mean Square Error of Approximation for the conditional and transition distribution. Grouping on length of the time series

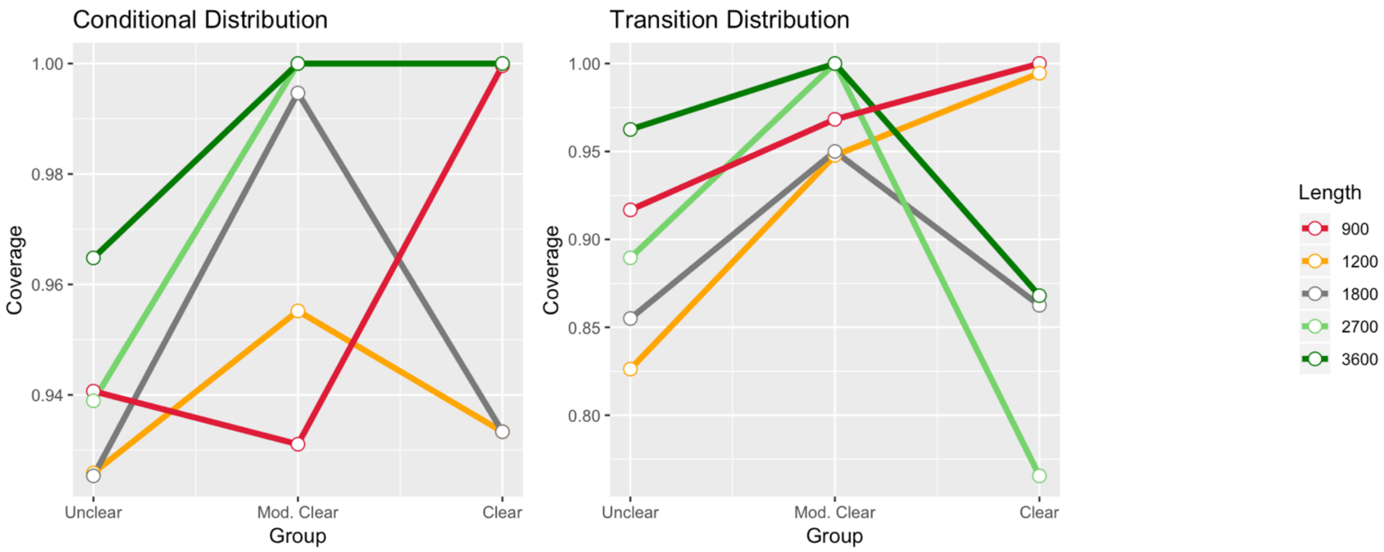

Figure 10: Averaged coverage of estimations within the credibility interval for the conditional distribution and transition distribution. Grouping on degree of distinctiveness between the states

Appendix 2

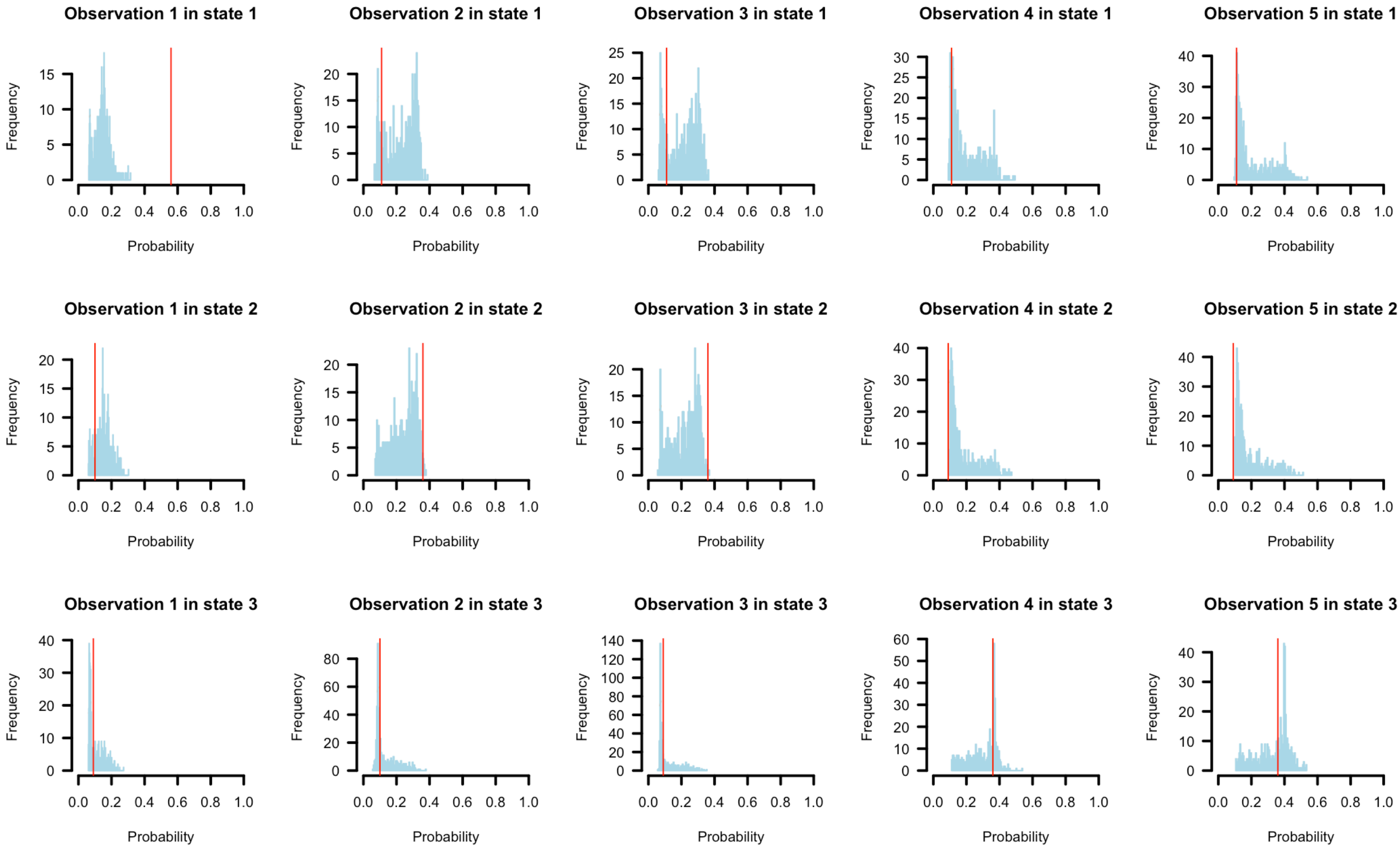

Figure 11: Parameter estimates of the conditional distribution (displayed by the blue distribution) as opposed to its true value (displayed by the red line) for scenario 6 – moderately clear distinction and length 900

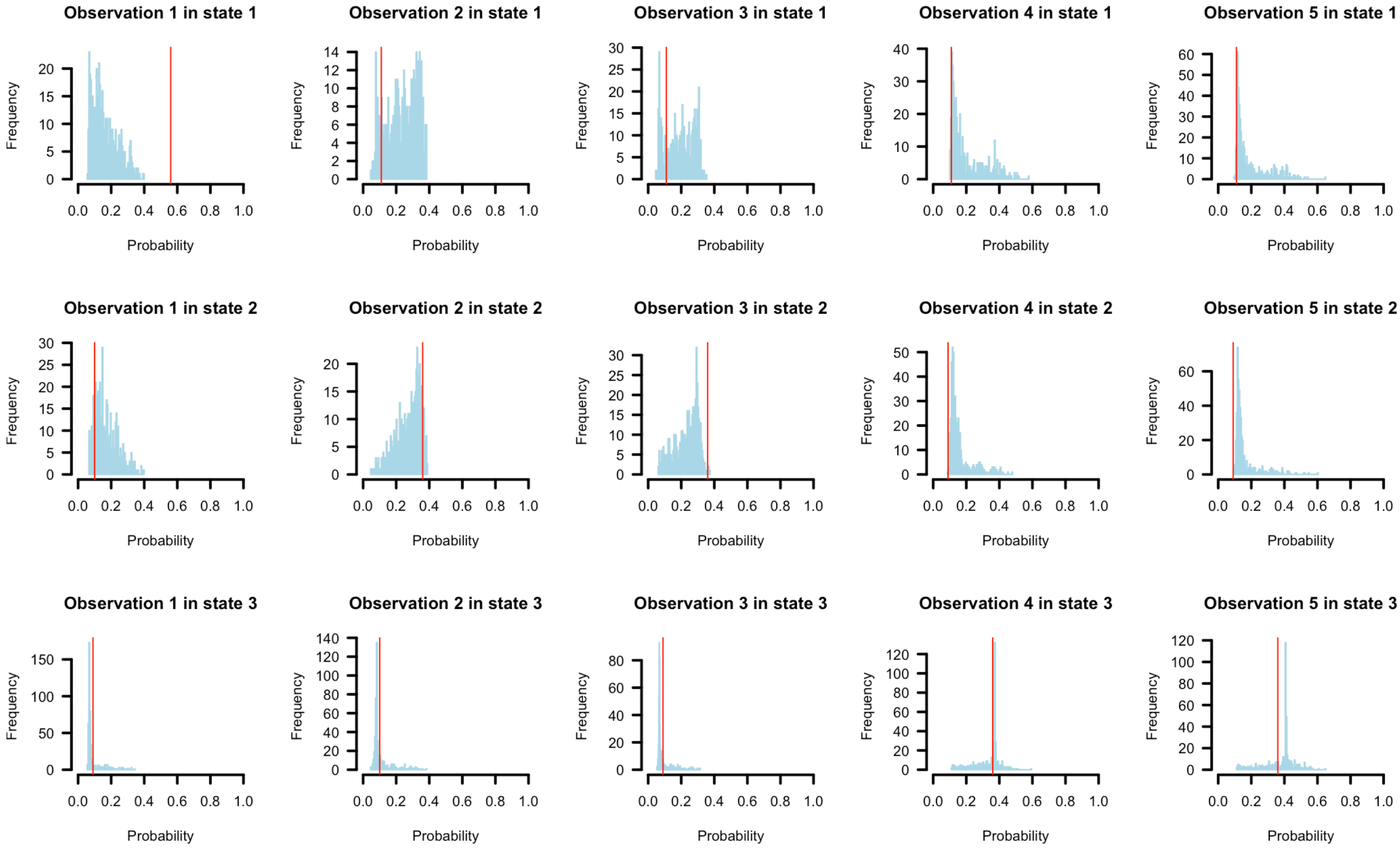

Figure 12: Parameter estimates of the conditional distribution (displayed by the blue distribution) as opposed to its true value (displayed by the red line) for scenario 7 – moderately clear distinction and length 1200

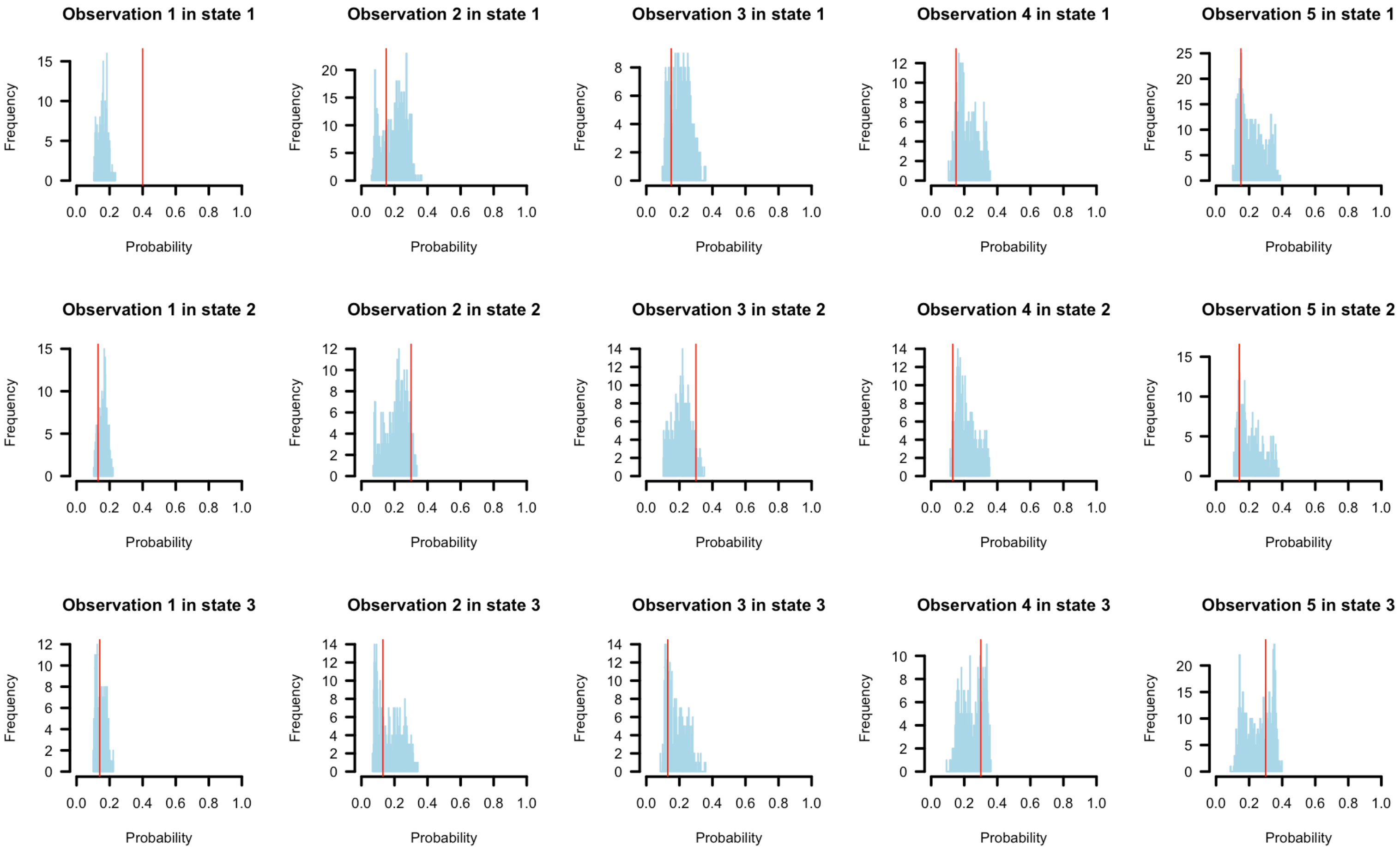

Figure 13: Parameter estimates of the conditional distribution (displayed by the blue distribution) as opposed to its true value (displayed by the red line) for scenario 1 – unclear distinction and length 900

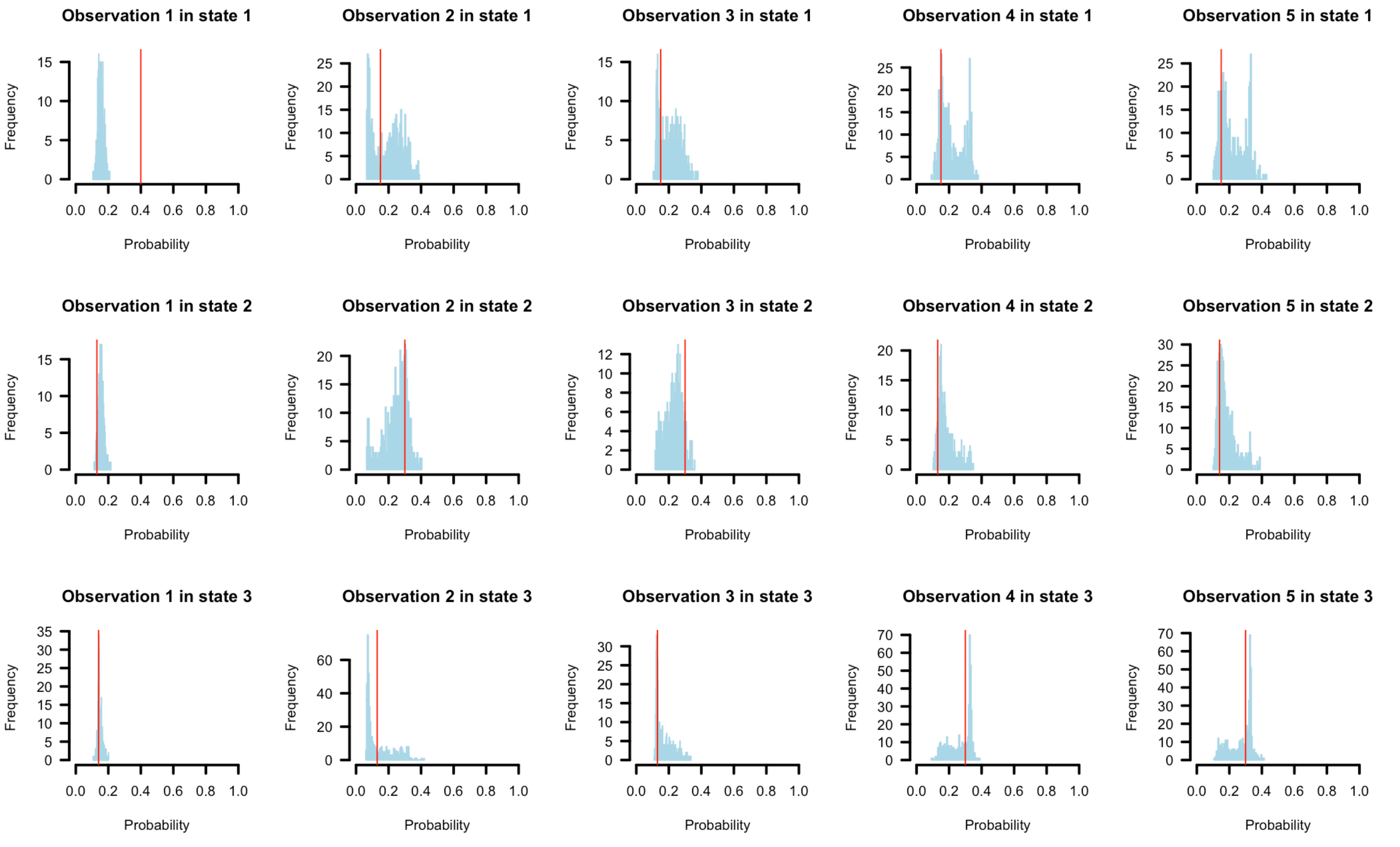

Figure 14: Parameter estimates of the conditional distribution (displayed by the blue distribution) as opposed to its true value (displayed by the red line) for scenario 2 – unclear distinction and length 900

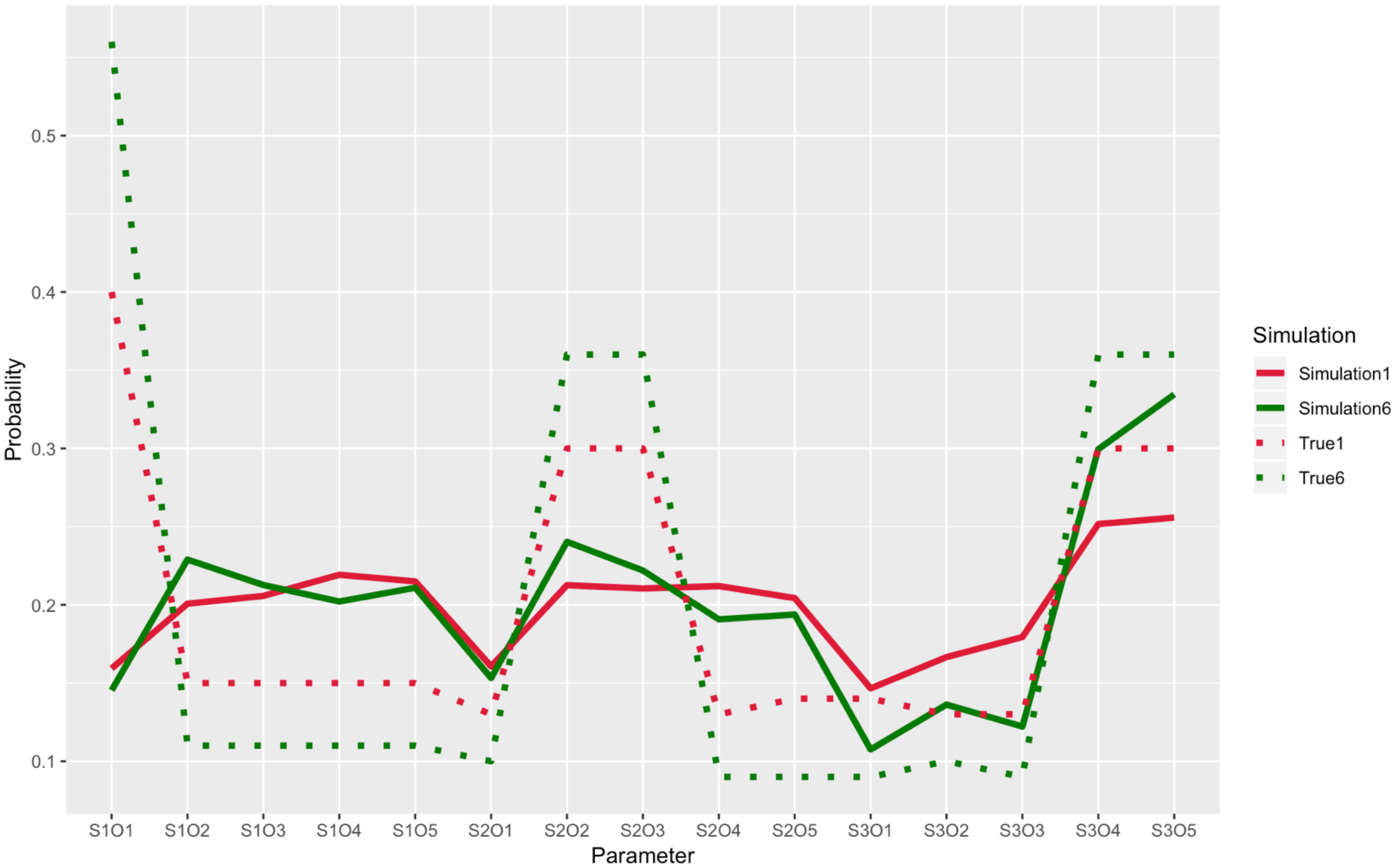

Figure 15: Parameter estimates of simulation 1&6 contrasted with their true values. Simulation 1 includes an unclear distinction between states and length 900. Simulation 2 includes a moderately clear distinction between states and length 900.

Appendix 3

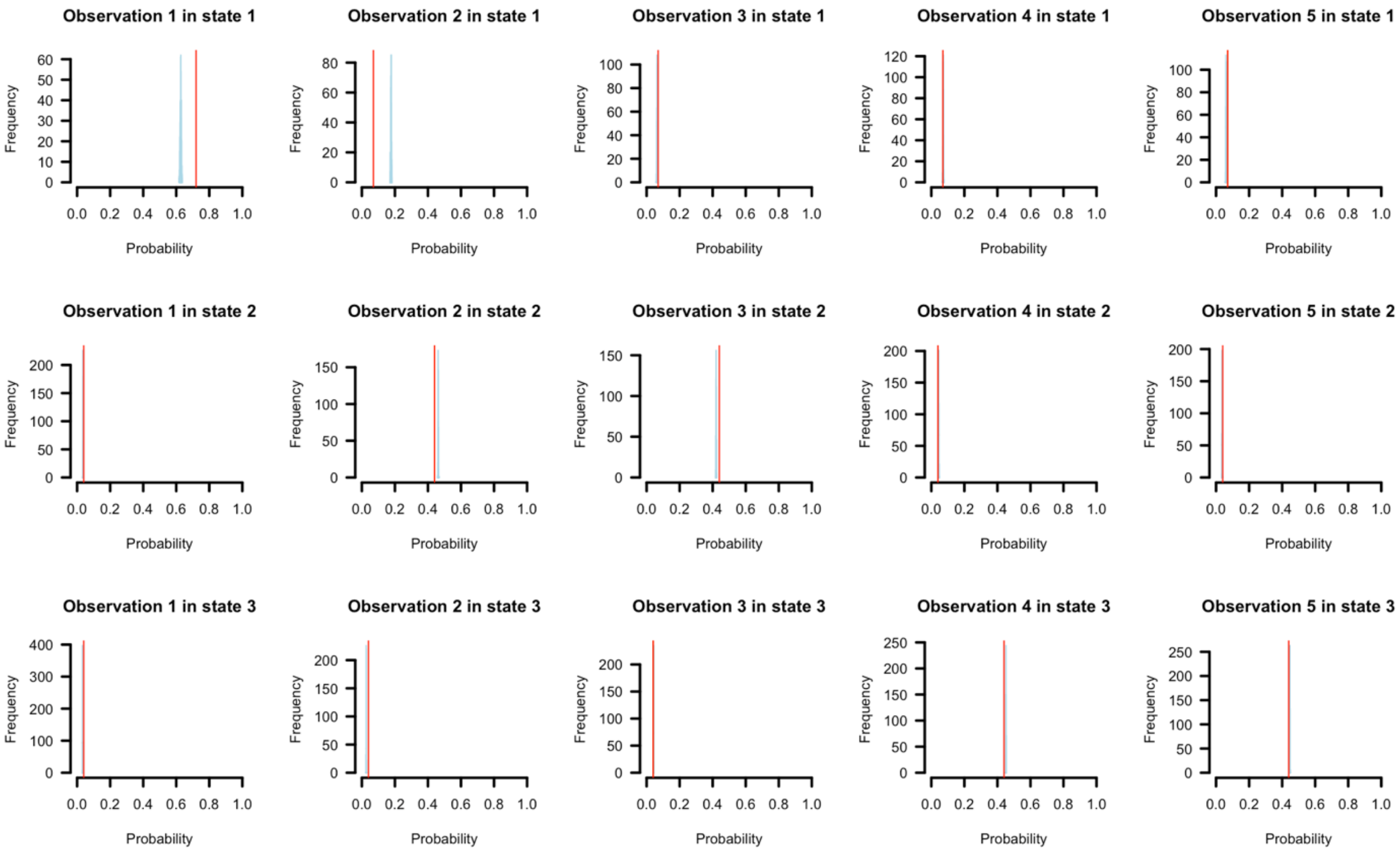

Figure 16: Parameter estimates of the conditional distribution (displayed by the blue distribution) as opposed to its true value (displayed by the red line) for scenario 11 – clear distinction and length 900

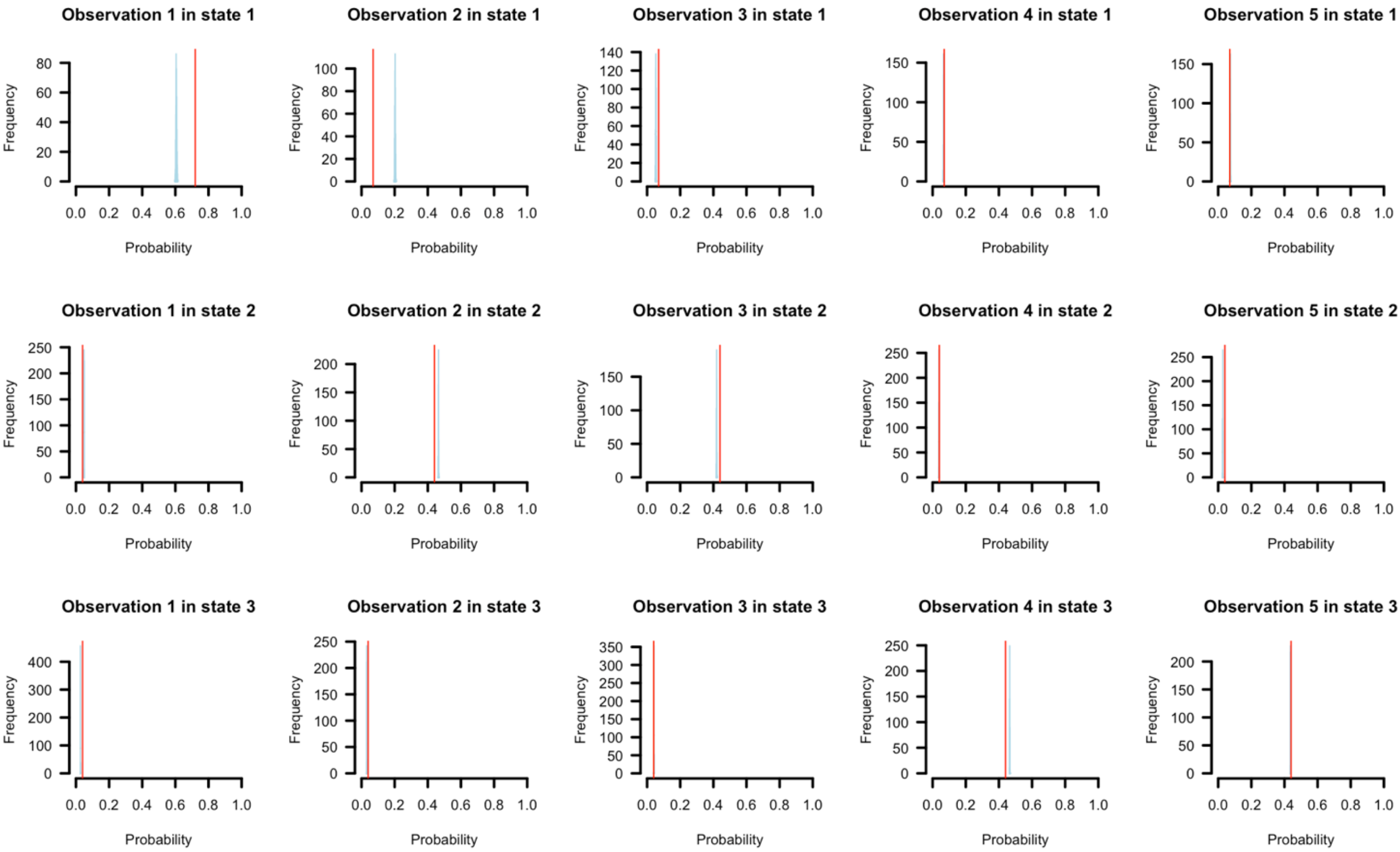

Figure 17: Parameter estimates of the conditional distribution (displayed by the blue distribution) as opposed to its true value (displayed by the red line) for scenario 12 – clear distinction and length 1200

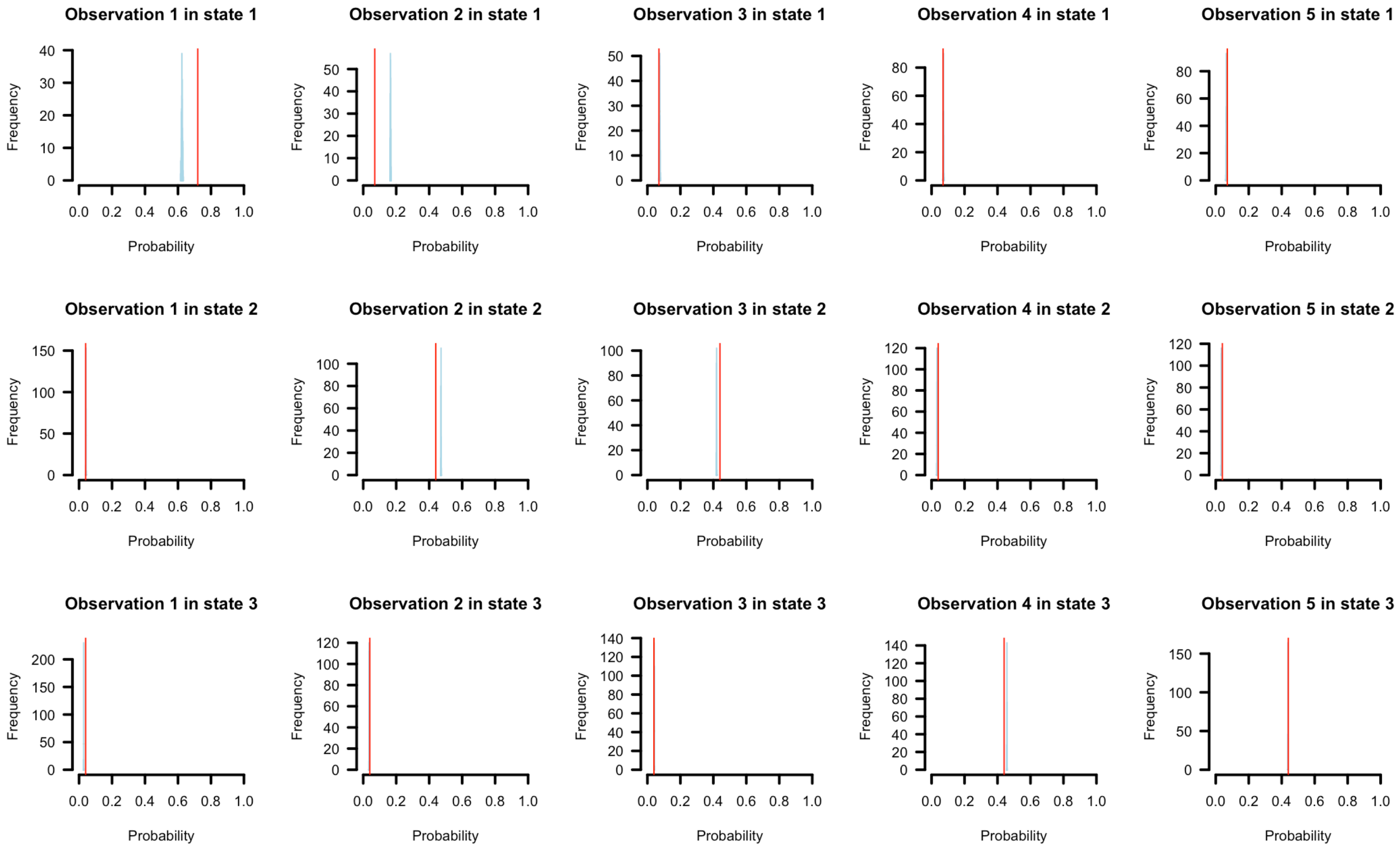

Figure 18: Parameter estimates of the conditional distribution (displayed by the blue distribution) as opposed to its true value (displayed by the red line) for scenario 13 – clear distinction and length 1800

Appendix 4

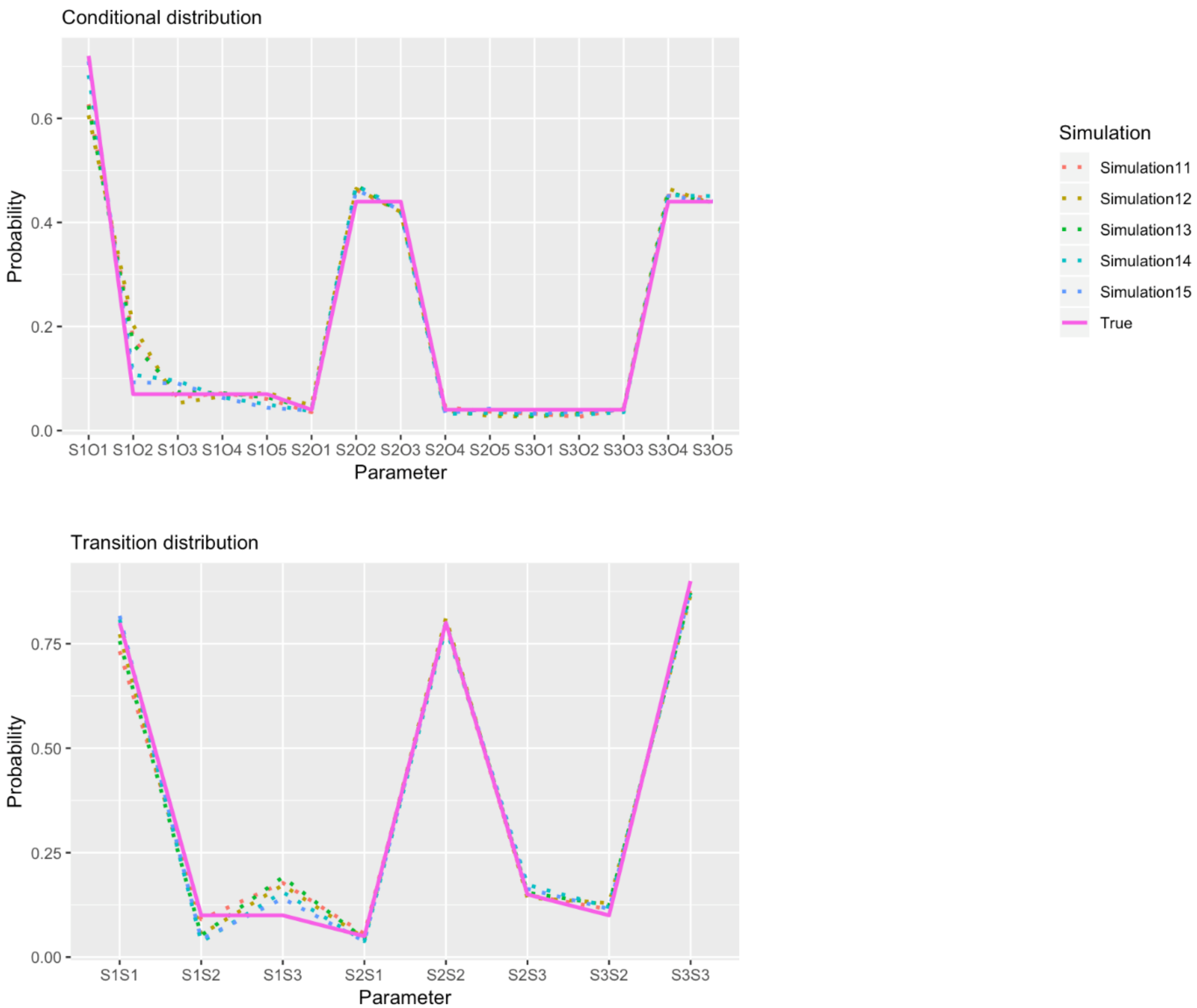

Figure 19: Parameter estimates of both the conditional and transition distribution for simulations 11 trough 15, contrasted with their true values. Simulations 11 through 15 all include a clear distinction between states, and lengths 900, 1200, 1800, 2700 and 3600 respectively.

Appendix 5

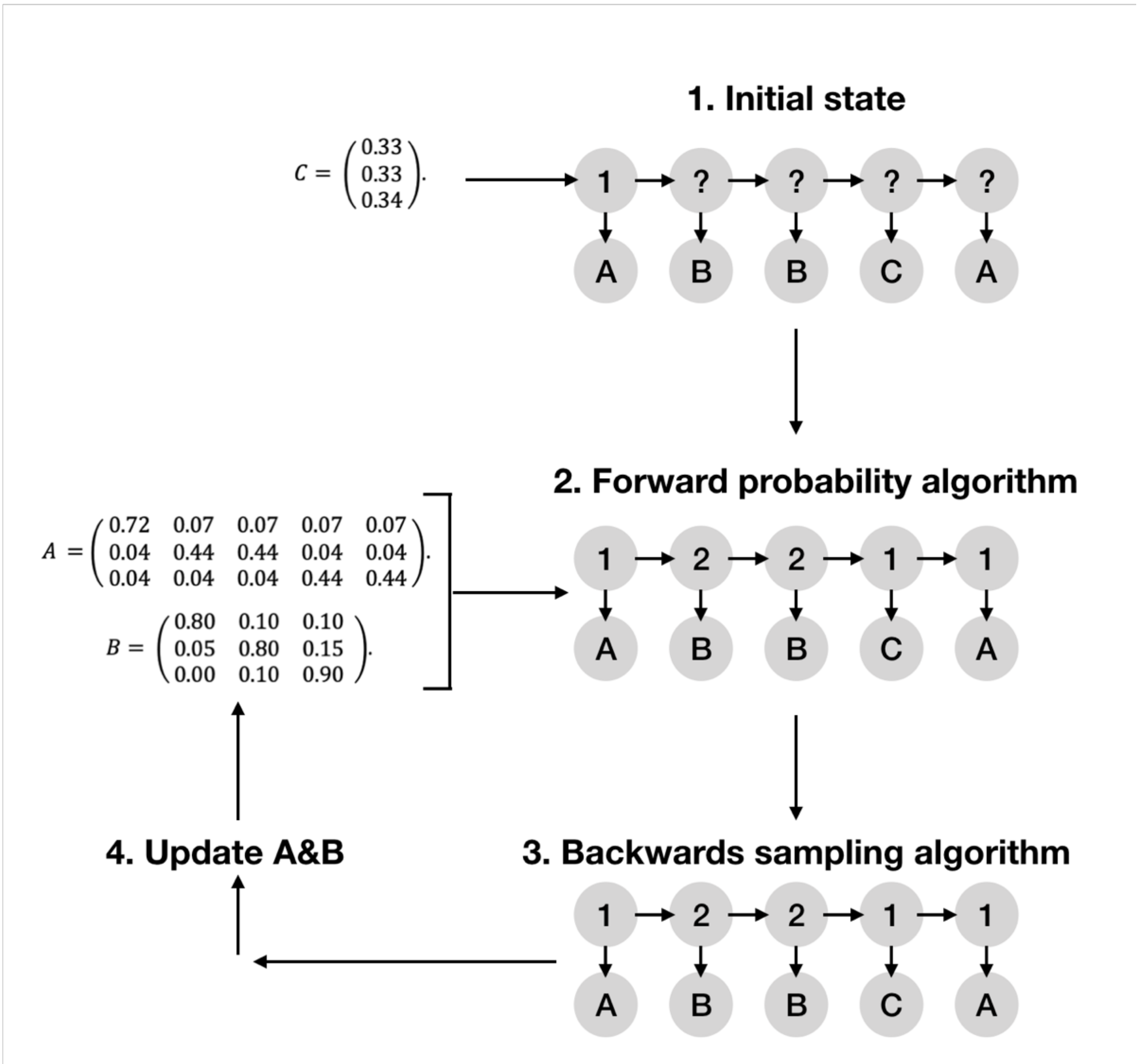

Figure 20: Semantical depiction of the employed Bayesian model estimation method.