Utrecht University

Netherlands Forensic Institute
Ministry of Justice and Security

Evidential value of duct tape comparison

A stochastic model based on loopbreaking patterns

Author:
Camille Diana van Dijk

Daily supervisor:
Prof. Dr. Marjan Sjerps (NFI)
Project supervisor:
Dr. Sjoerd Dirksen (UU)
Second reader:
Dr. Cristian Spitoni (UU)

MASTER'S THESIS

MATHEMATICAL SCIENCES

GRADUATE SCHOOL OF NATURAL SCIENCES

Utrecht University

Netherlands Forensic Institute

May 2021

Acknowledgements

It has been a long and sometimes hard process writing this thesis in times of the corona pandemic. I am very proud of what I have accomplished, but I would never have made it this far without the help of many people. I am very grateful for all of my friends, family and colleagues that supported me in any possible way during this process. In particular, I would like to thank the following people.

I want to thank my daily supervisor Marjan Sjerps, who has guided me through this project. Her support, ideas and critique during our weekly meetings were very helpful. Further, the forensic duct tape experts from the NFI, Koen Herlaar and Richard Visser, have kept me motivated by pointing out the real value of this research and helped me stay on the right track by stating the practical aspects of this problem. I also would like to thank the colleagues of the statistics team at the NFI, Leen, Ivo and Peter, for really making me part of the team and giving interesting feedback on my presentations. Furthermore, I am grateful for Sjoerd Dirksen for being my project supervisor and taking the time to give me extensive feedback on my thesis.

Furthermore, I am very thankful for Brian Janssen, who has learned me to code in Python and stood by my side during every challenging moment of this process, by being a loving boyfriend. Last but not least, I am grateful for my parents who always had my back and have supported me during my whole life.

Thank you all, without you I could not have accomplished this.

Abstract

Pieces of duct tape are retrieved from many crime scenes. They are a great source for forensic traces. The evidential value of these traces depend on the position of the piece of duct tape with respect to the roll and potential other pieces of duct tape. This can be determined based on findings that two pieces of duct tape used to be connected to each other. Forensic experts at the NFI proposed a new, additional method of examining this, which uses loopbreaking patterns. We construct a likelihood ratio-system which evaluates the likelihood ratios of these loopbreaking patterns. This LR-system consists of dynamic Bayesian networks, which assumes that the loopbreaking patterns comply to the Markov property. A small data set was available for the training and testing of this system. The validation of this system is evaluated in terms of accuracy, discriminating power and calibration. The results of the validation seem promising. For future work we recommend to validate the system on data which represent real forensic cases.

Contents

1. Introduction

Duct tape is found at many crime scenes. It is used to tie people up, make explosives and to package drugs. Duct tape is adhesive and is therefore a great source for forensic findings, such as hairs, other human biological traces, fingerprints and fibers. As a consequence, duct tape can play an important role in solving a criminal case. By comparing ends of duct tape one can investigate if two pieces of duct tape used to be directly connected to each other. This can be useful to link multiple crime scenes to each other, or to link a roll of duct tape to a crime scene. The forensic traces found on a piece of duct tape can be very valuable, but they are less valuable when they are found at the end of a piece that was on the outer end of the roll. The trace could have gotten there when a person was using the duct tape for another purpose, before the crime was committed. It is therefore not only important to determine whether a piece of duct tape belonged to a certain roll, but also its location on this roll.

1.1. Thesis overview

R. Visser and K. Herlaar, the forensic experts on duct tape at the Netherlands Forensic Institute (NFI), have come up with a new, additional method to evaluate the evidential strength of combinations of two torn pieces of duct tape, which is based on *loopbreaking patterns*. These are the patterns that arise when the vertical yarns (warp yarns) in duct tape break during the tearing process. These warp yarns loop into themselves, therefore the ends of the warp yarns on both ends of the tear can be either an open loop, a closed loop or complex. Also the horizontal position of these yarns with respect to the vertical yarns is taken into consideration. A. van Someren (NFI) performed an experiment, where she obtained the loopbreaking patterns of 136 pieces of torn duct tape, which she evaluated in her bachelor thesis [25]. The loopbreaking patterns and the available data will be discussed in depth in chapter 2.

In forensic science it is common to use the likelihood ratio (LR) to describe the strength of a piece of evidence. The aim of this research is to determine the LR of the loopbreaking patterns of two pieces of duct tape. Let E_1 and E_2 be the loopbreaking patterns of two pieces of duct tape. The hypotheses that we will consider for this evidence are

H_p: the two pieces of duct tape used to be directly connected;

H_d: the two pieces of duct tape have never been directly connected.

The likelihood ratio of E_1 and E_2 with respect to H_p relative to H_d can then be expressed as

$$LR = \frac{\mathbb{P}(E_1, E_2|H_p, I)}{\mathbb{P}(E_1|H_d, I)\mathbb{P}(E_2|H_d, I)}, \tag{1.1}$$

where I represents relevant background information.

In this thesis we will present an LR-system which evaluates the LR-values of the loopbreaking patterns of pieces of duct tape. This system is based on three dynamic Bayesian networks, which each evaluate one of the three probabilities in Equation 1.1. The main assumption for these dynamic Bayesian networks is that the loopbreaking patterns comply to the first order Markov

property. We use the available data and expert knowledge to establish the conditional probability distributions, needed for the dynamic Bayesian networks. We will perform logistic regression on the data and use the maximum likelihood estimation to obtain estimates for these probabilities. These mathematical concepts will be discussed in chapter 3. The dynamic Bayesian networks all consist of three building blocks: the loops, the horizontal positions and the observation errors. We will disregard cross-dependencies between the nodes, to keep the parameter space restricted. Otherwise the data set would be too small to provide reliable estimates for the parameters. We will use the software `HUGIN EXPERT` [9] to implement them. The construction of the dynamic Bayesian networks will be described in chapter 4.

The performance of this LR-system will be evaluated by dividing the available data into a training set and a test set. Additionally, extra data is simulated and the data in the test set is altered. The performance will be expressed terms of loss in accuracy, loss in discriminating power and loss in calibration, which is measured using empirical cross-entropy plots (ECE-plots) and log-likelihood-ratio costs (Cllr). In chapter 5 these methods and their results are discussed.

1.2. Literature

The process of forensic comparison of ends of duct tape pieces consists of two steps. The first step is studying and comparing the physical and chemical characteristics of the pieces of duct tape. Smith [24] describes the structure of duct tape and the forensic analysis of these characteristics is described. Mehltretter and Bradley [14] studied the quality of discrimination of different brands of duct tape based on these characteristics. LaPorte and Weimer [10] examined the variability of physical characteristics within rolls of duct tape. Mehltretter et al. [15] also compared the variability of characteristics within rolls and between rolls from the same jumbo roll in. Even though these studies consider duct tapes purchased in North America, they provide insight in the method of comparing characteristics. Furthermore, the forensic experts at the NFI agree that they are able to differentiate between the three main types of duct tape on the Dutch market based on these characteristics.

When it is not possible to exclude the pieces of duct tape based on the first step, because the observed characteristics are similar, the second step will be performed. In this step the physical fits of the ends of the pieces of duct tape are compared. In 2006 Bradley et al. [3] studied the validity and error rates of duct tape end matching. They found that 92% of the hand torn pieces were matched correctly, while the rest of the pieces were reported as inconclusive. A couple of years later, in 2011, Tulleners and Braun [26] carried out one of the first studies in which they provided a quantitative assessment of the quality of a physical fit for duct tape pieces. They assigned a match percentage to tape edges by measuring the match area lengths and dividing it by the total width of the tape. This same method was used in the study by McCabe et al. in 2013 [12], where they further evaluated the accuracy and error rate of duct tape end matching. Brooks et al. wrote a review article on physical fits of all types of materials [4].

Last year, Prusinowski et al. [20] proposed a new method to quantify the quality of an end match for duct tape. They used the scrim yarns to divide the duct tape ends into clearly definable scrim bins and they provided a similarity score for two duct tape ends based on the matching fraction of these scrim bins. With this method it should be easier to assess the end matching for stretched out tapes, since the bins are still clearly defined. The results of this study were promising, an accuracy of 84.9 to 99% and no false positives. However, on the high-quality tape they reported a false negative rate of 21.4%. This is also one of the first studies that uses the likelihood ratio framework for evaluating the value of evidence of a duct tape end

match, using score based likelihood ratios. This is a very important addition, since European forensic practitioners are currently required to use the likelihood ratio framework when writing an evaluative report, as stated in the guideline from the European Network of Forensic Science Institutes (ENFSI) [31].

Ristenpart et al. [22] developed an algorithm for digital images to calculate the distance between two duct tape ends, expressed in the sum of square residuals. This was a completely different way of solving this problem, the analysis was done automatically. Unfortunately, the error rates of this method were much higher than for human evaluation.

Furthermore, Wieten et al. created a Bayesian network to assist forensic examiners in the interpretation of evidence found on duct tape at the activity level, where one is interested in how a trace got onto the tape [30].

As of now, there is no published literature evaluating the loopbreaking patterns in torn pieces of duct tape.

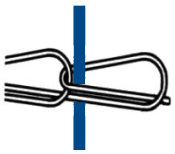
2. Forensic research on duct tape

2.1. Structure

Based on [14, 24, 25] the structure of duct tape can be described as follows. Duct tape consists of three basic layers: a polymeric backing, an adhesive, and in between them is reinforcement fabric, called the *scrim*. The scrim can be either loose weaves or knits with weft insertion. In this thesis we will only consider scrims with *weft insertion*. This type of scrim consists of warp yarns and weft yarns (also called fill yarns). *Warp yarns* are the yarns that are laid in the length direction of the tape, and the *weft yarns* are laid in the width of the tape. The warp yarns are thus very long compared to the weft yarns. Usually a roll duct tape is 5 cm wide, containing about 36 warp yarns over the whole length. The warp yarns loop into themselves many times (see Figure 2.1a). At these points the weft yarns are inserted into the loop of the warp yarn (see Figure 2.1b). In Figure 2.1c we see part of the whole pattern of a scrim with weft insertion.

During the examination of a piece of duct tape, a forensic expert can distinguish the side closest to the roll from the side that was furthest from the roll, based on the direction of the loops of the warp yarns. We will denote these as the *X-side* and the *Y-side* respectively, see figure 2.2.

(a) Illustration of part of a warp yarn looping into itself.



(b) Illustration of place of insertion of a weft yarn (blue) into warp yarn at the looping point.

(c) Part of the scrim under a microscope, with six warp yarns (horizontal) and two weft yarns (vertical) [14].

Figure 2.1.: Scrim with weft insertion pattern.

2.2. Comparing two pieces

In forensic science it is a common question: "Do these two pieces belong together?", for any type of material. In the case of duct tape, this question can be divided into three separate types of questions:

(i) *Common source:*
"Do the two pieces of tape come from the same roll?"

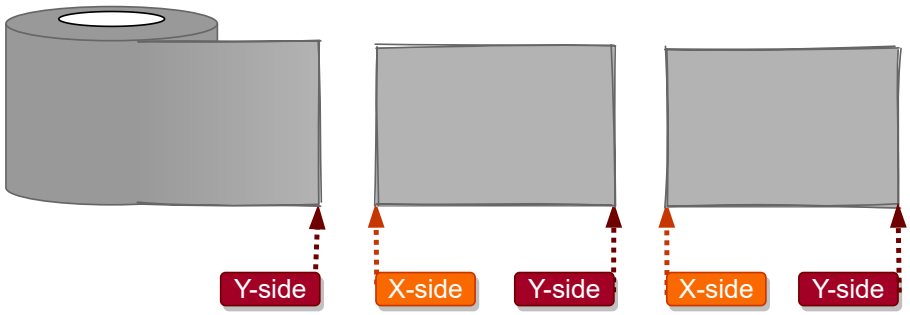
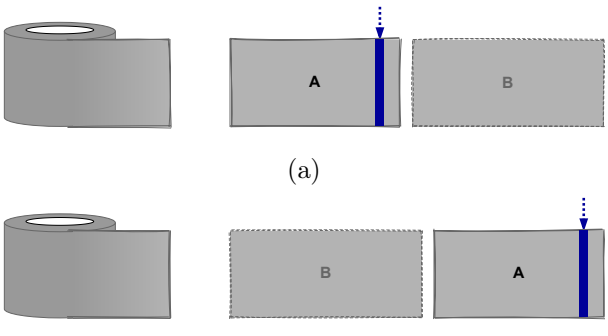


Figure 2.2.: Schematic view of roll of duct tape and two pieces with X-side and Y-side denoted.

(ii) *Specific source:*
"Does this piece/do these two pieces come from this roll?"

(iii) *Trace location:*
"Was this side of this piece of duct tape directly connected to that side of that piece of duct tape?"

The first question can be of interest when two pieces of duct tape are found at different places and there is no roll available. If the two pieces of tape came from the same roll, then it is likely that this roll was in both places. The second type of question can be of concern when one wants to find a link between the location or owner of a roll and the pieces of tape found at a crime scene. The third question is usually the most interesting. In the case that there is a sample of DNA or fibre found on one of the pieces of duct tape, the location of this sample is then quite important. If the sample location is at the end of the roll, then the defendant can argue that they used the roll and left the sample, but someone else used the roll later to commit the crime. If the sample is closer to the start of the roll, then it is less likely that the sample was left there before the crime was committed. For example, assume that there is a piece of tape, A, with a trace at the end that is farthest from the roll, and there is another piece of tape, B. If we know that piece A was connected to piece B, such that piece B is the furthest from the roll (see Figure 2.3a), then the distance between the trace location and the end of the roll is larger, than when piece B was connected to the other side of A (see Figure 2.3b). The interpretation of traces found on tape is further discussed in [30].



(a)

(b)

Figure 2.3.: Two pieces of duct tape, piece A with a trace location indicated (blue), and their respective locations with respect to the roll for two possible placements.

Example: Court case

An example of a court case where the forensic analysis of duct tape played an important role is a case in Limburg, The Netherlands, in 2015 [27]. Four pieces of duct tape were retrieved from the crime scene, at both ends of one piece DNA evidence was found. This DNA was used to identify the suspect. The defence argued that these DNA traces were not related to the crime and that they must have gotten there indirectly. Forensic experts at the NFI researched the four pieces of duct tape and came to the following conclusion

> *"To my firm belief, the 4 pieces of tape with SIN AAEY6500NL, AAEY6501NL and AAEY6502NL originally formed one whole piece in the order as depicted in figure 1."* [a].

See Figure 2.4. Furthermore, about the position of the roll the report states

> *"The findings with respect to the direction of the tape (orientation of the loops), similarities with duct tape from manufacturer Supertape and the knowledge of the duct tape production process, are much more probable when the missing roll was on side C of the assembled unit of tape, than when the missing roll was on side F of the entire composition of tape."* [b].

Based on these statements and the length of the entire composition of tape, the court decides that the two DNA traces must be related to the crime. Mainly based on these forensic findings the suspect is found guilty.

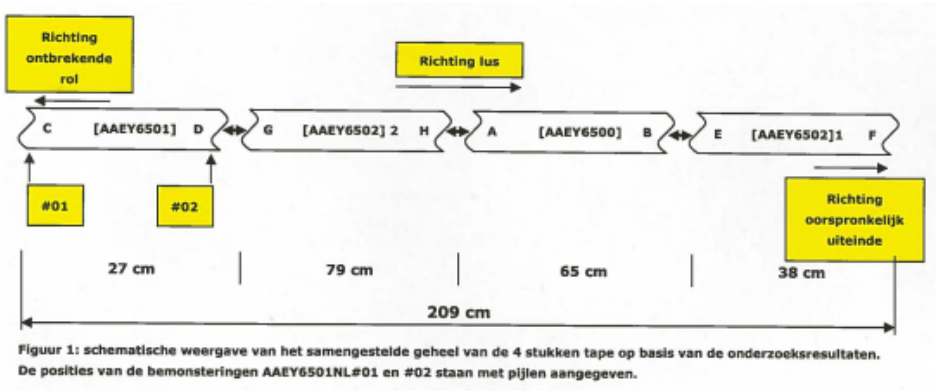


Figure 2.4.: Schematic representation of the formation of the 4 pieces of duct tape, based on research results. The positions of the biological traces #01 and #02 are indicated with arrows. The location of the missing roll is indicated to be on the left side of the formation. Retrieved from [27].

This court case illustrates the importance of finding *how* pieces of duct tape used to be connected to each other. Suppose it was mistakenly found that the order of the four pieces was not as depicted in Figure 2.4, but instead, say the first piece (AAEY6501) was located at the other end of the formation, such that ends C and F used to be connected instead of ends D and G. In that situation, it would have been more likely that the DNA trace was not related to the crime, than it was in the original situation. It is thus of great importance that the comparison of the ends of duct tape pieces is done very carefully.

[a] Translated from the original statement *"Naar mijn stellige overtuiging hebben de 4 stukken tape met SIN AAEY6500NL, AAEY6501NL en AAEY6502NL oorspronkelijk één geheel gevormd in de volgorde zoals schematisch weergegeven in figuur 1."* [27].

[b] Translated from the original dutch statement *"De bevindingen m.b.t. de richting van de tape (ligging van de lussen), overeenkomsten met duct tape van fabrikant Supertape en de kennis van de duct tape productie, zijn veel waarschijnlijker wanneer de ontbrekende rol tape aan de zijde C heeft gezeten van het samengestelde geheel tape, dan wanneer de ontbrekende rol tape aan zijde F heeft gezeten van het samengestelde geheel tape."* [27]

To answer all these types of questions, forensic researchers will examine and compare several characteristics of the found pieces of duct tape. They will first study if the pieces are of the same type of duct tape. One of the identifying characteristics is the type of scrim. There are currently three main types of duct tape with weft yarn insertion scrim, believed to be produced by three main brands. They have distinct physical characteristics, such as color and structure of the polymeric backing. Second, if it is believed that the two pieces are of the same type, the forensic examiners can study other physical and chemical characteristics of the different layers over the whole of the two pieces. Characteristics that can be of interest are the chemical compound of the adhesive and the regularity of the scrim. If these characteristics do not correspond in the two pieces of duct tape, then the answer to any of the three questions above will be "no". Third, one can examine if there is a *direct physical fit* between the pieces of duct tape. A direct physical fit is defined as two pieces that exhibit a sufficient number of individual characteristics that fit together based on the fractured edges [4]. This is especially of interest when considering the third question. To evaluate this, the structure of the polymeric backing at the edges and the alignment of the scrim can be analyzed. Additionally, in the case of duct tape with weft-insertion scrim, one can study the endings of the warp yarns of the scrim. If the two pieces of duct tape used to form one whole piece, then the warp yarns must all have been broken a certain way, this creates a *loopbreaking pattern*. In the next section we will describe this in further detail.

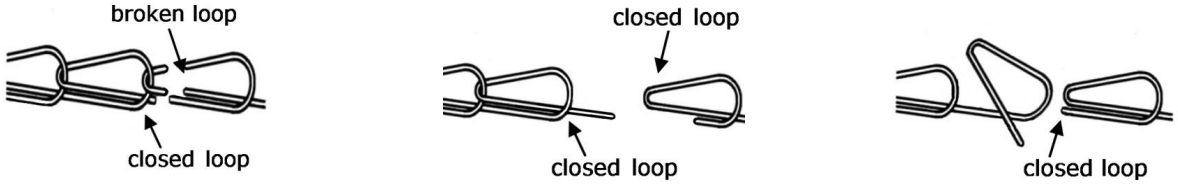
In this thesis we are interested in the evidential value of these loopbreaking patterns. However, the loopbreaking patterns are only studied if the two pieces could not have been excluded based on the other characteristics, since the examination of the loopbreaking patterns is time consuming. Therefore, in this thesis we will assume that the two pieces under consideration have the same physical and chemical characteristics, and we focus on the comparison of loopbreaking patterns.

2.3. Loopbreaking patterns

When duct tape is being torn along the weft yarns each warp yarn usually breaks at the point where it is looped into itself and the weft yarn is inserted, as this is supposed to be its weakest point, see e.g. Figure 2.7. At this point there are two loops intertwined (see Figure 2.1). Hence, at least one of these loops breaks and in some cases both of them break, see Figure 2.5. As a result at each yarn height a loop can be open (broken) or closed (whole). Moreover, in some cases the warp yarn does not break at the point where it loops into itself, which causes a *complex break*. In [25], three types of complex breaks are distinguished: ('7') a double break of one loop, such that part of the top of that loop still remains at the other loop; ('9') a prolonged loop end, some fibers have been released during the tearing process; ('3') a pulled out loop. They are also depicted in Figure 2.6. We will, however not discriminate between the different types of complex breaks, and therefore we will not discuss these types any further. It is important to remark though, that it often happens that forensic examiners misinterpret a complex break for a basic break at first sight. For simplicity, we will still indicate the end of a warp yarn as a *loop*,

(a) Closed-Open (b) Open-Closed (c) Open-Open

Figure 2.5.: Possible basic loopbreaks of the warp yarn. [25]



(a) Code '7' (b) Code '9' (c) Code '3'

Figure 2.6.: Different types of complex loopbreaks. [25]

even though it is not always a loop at the end.

Further, it is possible that there are missing warp yarns in the scrims. This can occur in production, e.g. a warp yarn is cut off at the top or bottom, or during the tearing, e.g. a warp yarn is pulled out. In some cases such a missing yarn is observable, because it left an imprint in the adhesive, or the cut off yarn is present somewhere along the length of the piece of tape. Furthermore, when two pieces are being compared the forensic examiners will study the vertical alignment of the scrims. If the scrims do not align at all, the patterns are not considered. However, if they do align, but there is a difference in the number of warp yarns this will be investigated. The spacing between the yarns in the scrim is usually very regular, so there must be space for missing yarns. When a missing warp yarn is observed, this is incorporated in the loopbreaking pattern.

In this way, we say that each loop can be in one of four states; *open*, *closed*, *complex* or *missing*. We will denote the state of each loop at height n as

$$X_n = \begin{cases} -1, & \text{if loop } n \text{ is missing,} \\ 0, & \text{if loop } n \text{ is whole (or closed),} \\ 1, & \text{if loop } n \text{ is broken (or open),} \\ 2, & \text{if loop } n \text{ is complex.} \end{cases}$$

The order of these loops is defined by the tearing direction. We consider two tearing directions as defined in section A.1. If the tearing direction is unknown, use tearing direction T.

Next to the state of the loops we can also observe the horizontal position. In some cases the forces on the tape are such that the breaking point of the next warp yarn is not aligned with the breaking point of the previous warp yarn, but it is at the next point where the warp yarn loops into itself (on the left or right). Then we see a *horizontal jump* at the edge. In this way one can observe the *horizontal position* of the end of each of the warp yarns. Notice that the breaking points are usually at the point of the weft yarn insertion. Hence, we can express the horizontal position in terms of the weft yarns. In the case of a complex break, the end of a yarn is not exactly at the point where the warp yarn loops into itself. We will denote the horizontal position of these warp yarns at the position of the first loop that is still whole. This has two reasons. First of all, we want to keep the horizontal position a discrete number. Second of all, this is how the observations, of the main data we used, were documented in [25]. In the case that the warp yarn is missing, the state of the corresponding horizontal position is omitted.

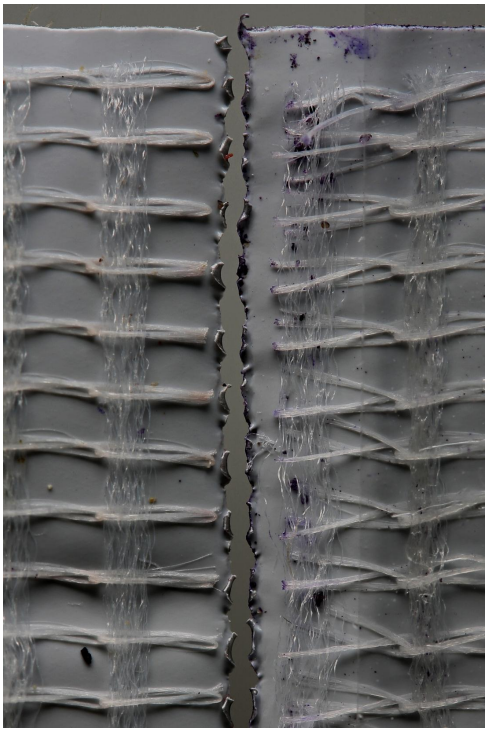


Figure 2.7.: Photo of the yarns of part of the edges of two pieces of torn duct tape under a microscope. The horizontal yarns are warp yarns and the vertical yarns are the weft yarns. [25]

We can record these positions with respect to the position of the first loop. Let us say the horizontal position of the loop at height n is denoted by ξ_n, then we declare $\xi_0 = 0$ and in general for $n > 0$

$$\xi_n = \begin{cases} -l, & \text{if the position of the } n\text{th loop is } l \text{ weft yarns to the left of the first loop;} \\ 0, & \text{if the position of the } n\text{th loop is at the same weft yarn as the first loop;} \\ +l, & \text{if the position of the } n\text{th loop is } l \text{ weft yarns to the right of the first loop.} \end{cases} \tag{2.1}$$

In the case that X_n is missing then the entry of ξ_n is omitted. If X_0 is missing, then we set the first present loop to be the reference point for the horizontal position and set its position to 0.

Example: Recording horizontal positions in loopbreaking patterns
In figure 2.8 we see a simplified illustration of an example of the edges of two pieces of torn duct tape, which have multiple horizontal jumps. The horizontal positions of the consecutive warp yarns on the X-side are in this case: $(\xi_n)_{n=0}^{8} = (0, 0, -1, -1, -2, -1, -1, -1, -1)$.
Notice that the horizontal positions on the Y-side are the same in this case.

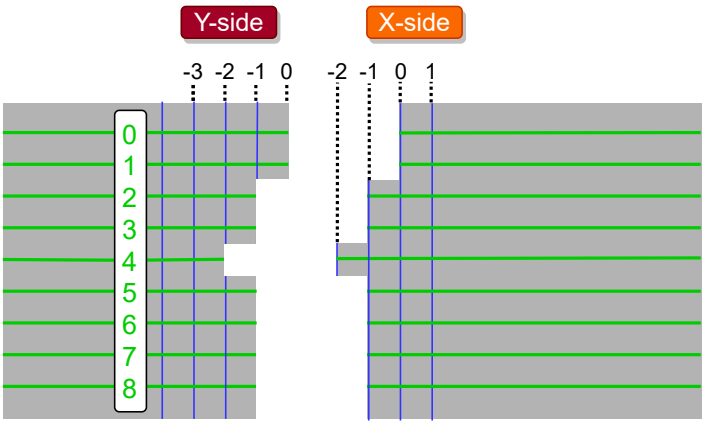


Figure 2.8.: Simplified illustration of example of tearing pattern of two pieces of duct tape with jumps in the horizontal positions. The 9 horizontal green lines represent the warp yarns and the vertical blue lines represent the weft yarns. The green numbers on the left denote the warp yarn height and the numbers on top denote the different horizontal positions based on the weft yarns.

Definition 2.1 (Loopbreaking pattern). *Let us consider one side of a torn piece of duct tape with m warp yarns and tearing direction d. Let the pattern order be defined by the tearing direction, such that the yarn at height 0 is the first yarn in the tearing direction. Let X_n denote the state of the loop at height n, such that $X_n \in \{$open, closed, complex, missing$\}$, for all $n \in \{0, 1, 2, \dots m-1\}$. Further, let for $i := \min\{i \leq m : X_i \neq \text{missing}\}$ $\xi_i = 0$ and for $n \in \{i, i+1, \dots m-1\}$ let ξ_n denote the horizontal position of the loop at height n in terms of weft yarns with respect to the position of the loop at height i, such that $\xi_n \in \mathbb{Z}$, except when X_n is missing. The* loopbreaking pattern *of this side of the piece of duct tape is then given by the sequence $(X_n, \xi_n)_{n=0}^{m-1}$.*

We can compare the loop breaking pattern of the X-side of one piece of duct tape with that of the Y-side of another piece of duct tape. If the two sides used to be directly connected, we would expect that in many cases across from any closed loop there is an open loop and the horizontal positions on both sides are the same. Thus, if there are two closed loops at the same height, or if the horizontal positions of the two sides are not the same at a certain height, in theory, we could conclude that it is not possible that these two sides used to be directly connected. However, we need to take the positions of the complex breaks and the possibility of observation errors into account. This will be further discussed in chapter 4.

2.4. Available data

The data that is used in this thesis is kindly provided by A. van Someren from NFI, received via personal communication. The experiment in which this data was obtained is described by van Someren in [25]. This data includes the loop breaking patterns of both sides of 136 pieces of torn duct tape (i.e., 272 patterns), of which there are 127 known matching pairs. These were torn by 3 selected persons and include 3 selected brands of tape, see table 2.1. Notice that, if a person tore k pieces of tape from a roll, then there are k patterns for the X-side and k patterns for the Y-side. Further, the Y-side of the first piece does not have a complementary X-side (it

		Tape			
		Tesa	Pattex	Supertape	Total
Person	A	19	19	12	50
	B	12	12	12	36
	C	19	19	12	50
	Total	50	50	36	136

Table 2.1.: Overview of available data. The numbers represent the pieces of duct tape per group.

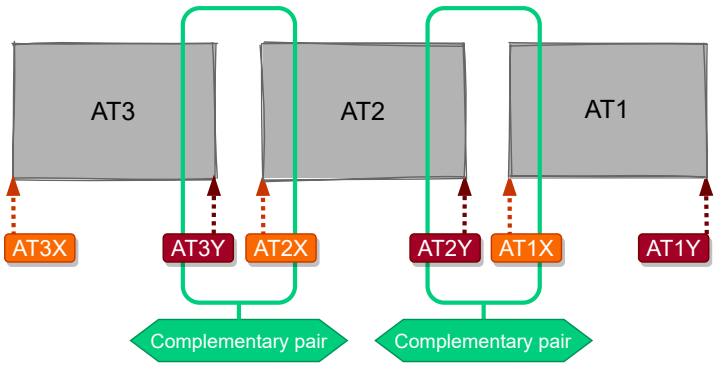


Figure 2.9.: Schematic example of encoding of the edges of the first three pieces of tape of brand Tesa, torn by person A.

does exist, but the loopbreaking pattern is not established), this holds as well for the X-side of the last piece of tape. Therefore, for this person with this tape, there are $k-1$ complementary pairs, each consisting of $k-1$ X-sides and $k-1$ Y-sides, see figure 2.9.

All the persons had a consistent tearing direction, for all the pieces. Persons A and B used the same tearing direction, T, and person C used tearing direction B. Additional information about the selection of the persons and some of their characteristics, such as their dominant hand and tearing method, is available in [25]. More pieces of torn duct tape acquired in this research are stored at the NFI (in total from 5 different persons), but their loop breaking patterns have not been determined yet. This is because the process of examining the loops is time consuming.

The available data for each end of duct tape consists of a complete loopbreaking pattern. As mentioned, in the data the three different complex breaks are distinguished. Moreover, when a cut off warp yarn at the bottom or top was observed, or when a pulled out warp yarn was observed it was recorded .

This data gives us some basic information about the loop tearing patterns. It gives us some insight in the way we could model the loopbreaking patterns and which factors might be relevant. At first sight, there is no apparent pattern in the loop breaking patterns. We do notice however that it does not occur often that the loops on both sides are broken. Furthermore, in some cases there is a long sequence of only closed or only open loops. The forensic experts at the NFI have a theory which could give an explanation for this situation. They assume this happens when the person gets hold of the weft yarn at the beginning of the tearing process, which tears all loops on one side open and on the other side they remain closed. At some point this weft yarn might break, from that moment on the loops break as they would normally do. It is assumed that this happens more often when the person places their fingers very close to each other and/or uses their nails. If we want to use this in our models, we would need more information and thus more data.

Furthermore, from the data and conversations with the forensic experts we found that there are many factors that could affect the loop breaking patterns, but it is still not clear which factors have the most effect. Some examples of such factors are: dominant hand, hand holding the roll, direction of tearing, placement of fingers, type of duct tape and homogeneity of scrim. Notice that most of these factors are unknown in a criminal case. In most cases it is not certain which person tore the tape and even if we would know that, we do not know how they tore the tape. Therefore, it is probably redundant to evaluate the effect of these separate factors. Rather we would like to obtain a data set which contains data that is from an experiment with a is sufficiently large sample size, such that it is a good representation of the whole population.

3. Mathematical methods

Before we can build a model to determine the evidential value of loopbreaking patterns in duct tape, we need to discuss some basic mathematical concepts.

3.1. Likelihood ratio

In a criminal case many different types of evidence can be found. It is essential to evaluate this evidence correctly. Eventually we want to know how this affects our belief that the suspect is guilty of committing the crime. For this we always consider two hypotheses: the *prosecutor's hypothesis* H_p, which is associated with the suspect being guilty and the alternative *defense hypothesis* H_d, which is associated with the suspect being innocent. We can express our *prior belief* in favour of H_p with respect to H_d in odds form as:

$$\frac{\mathbb{P}(H_p)}{\mathbb{P}(H_d)},$$

the *prior odds*. Our *posterior belief* in favour of H_p with respect to H_d, given the evidence E, can be expressed in odds form as

$$\frac{\mathbb{P}(H_p|E)}{\mathbb{P}(H_d|E)},$$

the *posterior odds*. These probabilities are always evaluated given certain background information I. For readability, we will omit the notation of this background information, for now. Using Bayes' theorem we can convert prior odds into posterior odds, as described in [1].

Theorem 3.1.1 (Bayes' theorem, odds form).

$$\begin{array}{ccccc} \dfrac{\mathbb{P}(H_p|E)}{\mathbb{P}(H_d|E)} & = & \dfrac{\mathbb{P}(E|H_p)}{\mathbb{P}(E|H_d)} & \times & \dfrac{\mathbb{P}(H_p)}{\mathbb{P}(H_d)} \\ \textit{posterior odds} & = & \textit{likelihood ratio} & \times & \textit{prior odds} \end{array} \tag{3.1}$$

The factor with which we multiply the prior odds to obtain the posterior odds is called the *likelihood ratio*. This is a measure of the evidential value of the evidence E.

Definition 3.1 (Likelihood ratio (LR)). *Let E denote the evidence, H_p the prosecutor's hypothesis and H_d the defense hypothesis. The* likelihood ratio *(LR) of the evidence E with respect to H_p relative to H_d is defined as*

$$LR_{H_p,H_d}(E) := \frac{\mathbb{P}(E|H_p)}{\mathbb{P}(E|H_d)}.$$

When no ambiguity is expected the notation can be shortened to $LR(E)$ or even LR.

In forensic science it has become common practice to use the likelihood ratio to express the evidential strength of the evidence. The European Network of Forensic Science Institutes (ENFSI) even requires forensic practitioners to report the evidential value in terms of a likelihood ratio, when submitting an evaluative report [31]. One is usually interested in the posterior odds for a case, but these depend on the prior odds. The prior odds are usually outside the domain of expert knowledge. Therefore it is common to only report the likelihood ratio for evidence rather than the posterior odds. This should give enough information for a judge or jury to update the personal prior odds of the hypotheses to the posterior odds.

3.1.1. Combining evidence

One can combine different types of evidence to generate a combined LR. Consider two types of evidence E_1 and E_2. The odds form of Bayes' theorem with $E = E_1 \cap E_2$ has the following form

$$\frac{\mathbb{P}(H_p|E_1, E_2)}{\mathbb{P}(H_d|E_1, E_2)} = \frac{\mathbb{P}(E_1, E_2|H_p)}{\mathbb{P}(E_1, E_2|H_d)} \times \frac{\mathbb{P}(H_p)}{\mathbb{P}(H_d)}.$$

Hence, the combined LR of the two pieces of evidence is

$$LR(E_1, E_2) = \frac{\mathbb{P}(E_1, E_2|H_p)}{\mathbb{P}(E_1, E_2|H_d)}.$$

When the two types of evidence are conditionally independent given either one of the hypotheses, then this can be done fairly straightforward

$$\begin{aligned} LR(E_1, E_2) &= \frac{\mathbb{P}(E_1, E_2|H_p)}{\mathbb{P}(E_1, E_2|H_d)} \\ &= \frac{\mathbb{P}(E_1|H_p)\mathbb{P}(E_2|H_p)}{\mathbb{P}(E_1|H_d)\mathbb{P}(E_2|H_d)} \\ &= LR(E_1)LR(E_2). \end{aligned}$$

However, when the two types of evidence E_1, E_2 are not conditionally independent, the LR of the combined evidence cannot be expressed in terms of the $LR(E_1)$ and $LR(E_2)$. We have

$$\begin{aligned} LR(E_1, E_2) &= \frac{\mathbb{P}(E_1, E_2|H_p)}{\mathbb{P}(E_1, E_2|H_d)} \\ &= LR(E_1|E_2)LR(E_2), \end{aligned}$$

where we write

$$LR(E_1|E_2) = \frac{\mathbb{P}(E_1|E_2, H_p)\mathbb{P}(E_2|H_p)}{\mathbb{P}(E_1|E_2, H_2)\mathbb{P}(E_2|H_d)}.$$

3.2. Markov Chain

A *Markov Chain* (MC) is a stochastic process, where the state of the process at a certain time n is only dependent on the state of the process at one time-step earlier. For example, if we assume that the weather is a Markov Chain, then the probability that it will rain today only depends on the weather of yesterday. Let us give a more formal definition based on [23, 7].

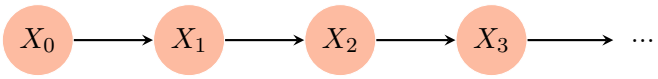


Figure 3.1.: Graph representing a Markov chain, each node represents the state of the process at a certain time.

Definition 3.2 (Discrete time Markov chain). *Let $(X_n)_{n\in\mathbb{N}_0}$ be a stochastic process and let M be a finite set of states, such that $\mathbb{P}(X_n \in M) = 1$ for all $n \in \mathbb{N}_0$. The stochastic process $(X_n)_{n\in\mathbb{N}_0}$ is called a* Markov Chain (MC) *if the* Markov property *holds, i.e., for all $n \geq 1$ it holds that*

$$\mathbb{P}(X_n|X_{n-1}, X_{n-2}, \ldots, X_0) = \mathbb{P}(X_n|X_{n-1}). \tag{3.2}$$

An MC is called homogeneous *if $\mathbb{P}(X_n = i|X_{n-1} = j)$ does not depend on n. In this case we denote the* transition probability *by $P_{ij} := \mathbb{P}(X_n = j|X_{n-1} = i)$ for any $i, j \in M$, such that $\sum_{j\in M} P_{ij} = 1$ for all $i \in M$. With $\mathbf{P} = (P_{ij})_{i,j\in M}$ we denote the corresponding* transition matrix. *The* initial probabilities *are denoted by $\pi_i := \mathbb{P}(X_0 = i)$ for all $i \in M$.*

One can determine the probability of observing a certain sequence $(i_0, i_1, \ldots, i_N)$ using

$$\begin{aligned}\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_N = i_N) =& \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1|X_0 = i_0)\mathbb{P}(X_2 = i_2|X_1 = i_1)\cdots \\ & \mathbb{P}(X_N = i_N|X_{N-1} = i_{N-1}) \\ =& \pi_{i_0} P_{i_0 i_1} P_{i_1 i_2} \cdots P_{i_{N-1} i_N}.\end{aligned}$$

Graphical representation An MC can be represented as a directed graph, as depicted in figure 3.1. Here each node describes the random variable at a certain time step. Each node can be in any of the possible states. The arrows represent the dependencies of the variables. In the case of a first-order MC, the (first-order) Markov property holds, so there are only arrows from X_n to X_{n+1} for each $n \geq 1$. In the case of a second-order MC there would be additional arrows from X_n and X_{n+1} to X_{n+2} for each $n \geq 1$.

3.3. Bayesian Network

A Bayesian Network (BN) is a graphical probabilistic model. In such a model we can describe uncertainties both quantitatively and qualitatively. First, we will note some basic notions from graph theory. A BN consists of *nodes* and *directed edges*. We denote the set of all nodes that have a directed edge towards a certain node X, i.e., the *parents* of X, by $\text{pa}\{X\}$. Similarly, the set of all nodes to which there is directed edge from X, the children of X, is denoted by $\text{ch}\{X\}$. Further, when there is a directed path from a node X to a node Y, then X is called an *ancestor* of Y and on the other hand Y is a *descendant* of X. We denote by $\text{desc}\{X\}$ and $\text{anc}\{X\}$ the sets of all ancestors and descendants of a node X respectively. We will now give a formal definition.

Definition 3.3 (Bayesian Network). *Let $G = (\mathbf{X}, E)$ be a directed acyclic graph (DAG). The $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ denote N random variables or nodes, which all have a finite set of mutually exclusive states. Further, E is a set of ordered pairs of nodes (X_i, X_j), which represent directed edges, characterizing the conditional dependencies between these variables. Let $\mathbb{P}(X_i|\text{pa}\{X_i\})$ be the conditional probability distributions, also called the* conditional probability

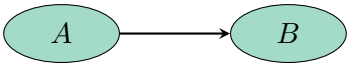


Figure 3.2.: Graph representing a BN with two nodes A and B, where the state of node B is dependent on the state of node A.

tables *(CPTs), that determines the probability distributions of each variable, X_i, conditionally on all of its parents. We call* $\left(G, (\mathbb{P}(X_i|\text{pa}\{X_i\}))_{i=1}^N\right)$ *a* Bayesian Network *(BN) if the following property holds. For any random variable X and any set of random variables $\boldsymbol{Y}$, which does not contain any descendants of X,it holds that*

$$\mathbb{P}\left(X|\text{pa}\{X\}, (\boldsymbol{Y})\right) = \mathbb{P}(X|\text{pa}\{X\}). \tag{3.3}$$

This is also known as the Markov Property for Bayesian Networks.

One of the most important properties of a BN, which follows from the Markov property mentioned above, is that the joint probabilities are fully defined by the conditional probability tables, as

$$\mathbb{P}(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{N} \mathbb{P}(X_i|\text{pa}\{X_i\}).$$

Hence, a BN consists of three ingredients

- a finite collection of random variables (nodes) and their corresponding finite state spaces;
- a set of directed edges between two nodes, which, together with the nodes, form a DAG;
- for each node X_i with parents $\text{pa}\{X_i\}$ a CPT, given by $\mathbb{P}(X_i|\text{pa}\{X_i\})$.

In Figure 3.2 we see an example of the graph of a BN with two nodes A and B. The arrow denotes that the state of B depends on the state of A. In Table 3.1 and Table 3.2 we see the corresponding CPTs for the nodes A and B, in the case that both nodes have two states.

A	a	$\neg a$
	$\mathbb{P}(A = a)$	$\mathbb{P}(A = \neg a)$

Table 3.1.: CPT for a node A with no parents. Where A has only two states.

A	a	$\neg a$
B		
b	$\mathbb{P}(B = b\|A = a)$	$\mathbb{P}(B = b\|A = \neg a)$
$\neg b$	$\mathbb{P}(B = \neg b\|A = a)$	$\mathbb{P}(B = \neg b\|A = \neg a)$

Table 3.2.: CPT for a node B with one parent, A. Both A and B have only two states.

3.3.1. Dynamic Bayesian Network

A dynamic Bayesian network is a Bayesian network extended with time slices. Each random variable X in the BN can now change over time, so its state at time n can be denoted by X_n. In this way a dynamic Bayesian network (DBN) can represent a discrete-time stochastic process [19].

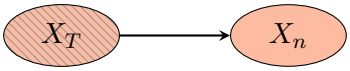


Figure 3.3.: Graph representing a DBN with one node, where the hatched node X_T represents the state of the random variable X at time $n-1$ and the plain node X_n represents the state of the random variable X at time n.

Definition 3.4 (Dynamic Bayesian Network). *A DBN is a pair $(B_0, B_{n|n-1})$, where B_0 represents a BN which describes the initial probabilities for the random variables $\boldsymbol{X_0}$ and $B_{n|n-1}$ represents a two-slice temporal Bayesian network which is a DAG that contains nodes of random variables in the current time (second slice) and nodes that represent the states of these random variables one time step back (first slice), together with the conditional probabilities tables $\mathbb{P}(X_n|pa\{X_n\})$ for all nodes in the second slice of the network.*

Remark that the nodes in the first time slice of $B_{n|n-1}$ are not provided with any probabilities. Moreover parents of nodes $X_{i,n}$ are allowed to be either in the same time slice or in the time slice prior to it. This directly implies that the first-order Markov property holds for the DBN. The Markov property for Bayesian networks also holds here, such that $\mathbb{P}(\boldsymbol{X_n}|\boldsymbol{X_{n-1}}) = \prod_{i=1}^{N} \mathbb{P}(X_{i,n}|\text{pa}\{X_{i,n}\})$, where $\boldsymbol{X_n}$ represents the whole set of random variables in the network at time n and $X_{i,n}$ is random variable i at time n [19].

Note that it is also possible to create a DBN with more time slices to make it comply to a higher order Markov property, but we will not consider these. The simplest DBN with one random variable, is thus a Markov Chain, which can be represented as in Figure 3.1. It can also be graphically represented as a DBN as in Figure 3.3. Here we can see both $B_{n|n-1}$ and B_0. The node X_T represents both X_0 and X_{n-1}. If node X_n is not yet evaluated, at time 0, then X_T represents X_0 and its distribution is described by the CPT associated with it in B_0. Otherwise the node X_T will copy the distribution from X_n in the former time slice. The arrow represents the dependency of X_n on X_{n-1}. The distribution of X_n is thus conditional on X_{n-1} and it is described by the CPT in $B_{n|n-1}$. This is the representation that we will use for the model. Notice that this representation must always be accompanied by the associated CPTs of the DBN.

This gives us the opportunity to use the appealing properties of a Bayesian network to represent a stochastic process for which the first-order Markov property holds.

3.4. Logistic Regression

The parameters we need to estimate for the current model are all (conditional) probabilities. Therefore we are interested in values of the form $\mathbb{P}(Y = j|\mathbf{x})$, where Y represents the output value and $\mathbf{x}$ the input values. This can be estimated using logistic regression.

Logistic regression is used when one is interested in the effect of certain covariates on the outcome of an experiment. In the case of duct tape we could think of: "What is the effect of the nth loop being open on the state of the $n+1$th loop?". Usually, in logistic regression, one is interested in a dichotomous outcome variable $Y \in \{0, 1\}$. However, in our case the number of states of the different nodes are at least three. Therefore we will consider *multinomial logistic regression* as described in chapter 8 of [8].

Suppose we have a data set consisting of points $(y_i, \boldsymbol{x_i})$, where $\boldsymbol{x_i}$ is a vector representing the values of the input variables and $y_i \in \{0, 1, 2, \ldots, K\}$ is the corresponding value of the output variable. In our case the input variables are categorical, therefore we will transform them, using

dummy variables, this is explained in Appendix B. Now we can fit a logistic function on these datapoints. We have for each $j \in \{0, 1, 2, \ldots, n\}$

$$\pi_j(\boldsymbol{x}) := \mathbb{P}(Y = j|\boldsymbol{x}) = \frac{e^{g_j(\boldsymbol{x})}}{\sum_{k=0}^{n} e^{g_k(\boldsymbol{x})}}, \tag{3.4}$$

here $g_0(\boldsymbol{x}) = 0$ and for $0 < k \leq n$ the logit functions are represented by

$$g_k(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta_k}. \tag{3.5}$$

Hence to estimate π_j we need to estimate the parameter vectors $\boldsymbol{\beta_k}$. It is common practice to use maximum likelihood estimation for this. We can show that for any l and j

$$\pi_j(\boldsymbol{x} = \boldsymbol{l}, \hat{\boldsymbol{\beta}}_{MLE}) = \frac{N_{jl}}{N_{\cdot l}}, \tag{3.6}$$

where $\boldsymbol{x} = \boldsymbol{l}$ denotes that the input variable has category l, N_{jl} denotes the total number of observations where $\boldsymbol{x} = \boldsymbol{l}$ and $Y = j$, and $N_{\cdot l} := \sum_{j=0}^{n} N_{jl}$. Thus, in this way we see that the conditional probabilities found using the MLE, are simply given by the fraction of the observations where the output is j and the input has category l of all of the observations with input category l. The proof of this fact is given in Appendix B.

4. Model construction

In this chapter we will describe the formalization and the construction of the model that we will use to determine the likelihood ratio of loopbreaking patterns of duct tape pieces.

4.1. Formulation of likelihood ratio

In this thesis we are interested in the likelihood ratio of two pieces of torn duct tape to have been directly attached to one another, relative to not have been directly attached to one another, based on the loop breaking patterns of these pieces. In this section we will formalize this.

When considering two pieces of duct tape, one can be interested in one of two propositions.

- *Specific comparison:* the X-side of piece 1 used to be attached to the Y-side of piece 2;
- *General comparison:* the two pieces of duct tape used to be attached to one another.

(Recall that the X-side and Y-side of the tapes are defined as described in Figure 2.2 and that two X-sides or two Y-sides cannot have been attached to one another.) In the first point we only consider one side of both pieces of tape, in the second point we consider both sides. In section 2.2 we discussed the different cases where these propositions could be applied.

Let us first consider the *specific comparison* proposition. We will define the two pieces as evidence as:

> E_X : the loop breaking pattern of the X-side of a torn piece of duct tape, piece 1, that starts at the first present loop, with direction d;
>
> E_Y : the loop breaking pattern of the Y-side of *another* torn piece of duct tape, piece 2, that starts at the first present loop, with direction d.

The direction d represents the direction that was used for the order of the pattern. This should correspond to the tearing direction if it is known. Any missing yarns at the top of the loopbreaking patterns are dismissed and the whole pattern is shifted such that X_0 and Y_0 now represent the state of the first present loop at the X-side and the Y-side respectively.

The hypotheses we will consider in this case are

> H_p : The X- and Y-side under consideration used to be directly attached to one another as one piece of duct tape, before being torn.
>
> H_d : The X- and Y-side under consideration have never been directly attached to one another. Either they used to be directly connected on the opposite sides of both pieces, or they came from the same roll and there was one or more pieces of tape in between them, or they came from different rolls of duct tape.

Furthermore, we will consider the background information I, this defines the context in which we consider the evidence. We assume that the pieces of duct tape have been examined by the experts in advance and therefore only consider the loop breaking pattern when the pieces of duct tape exhibit the same physical and chemical structure. Hence properties of the two pieces of duct tape that are included in the background information are: type of scrim (must be weft insertion), width of pieces, color of polymeric backing, sides that are torn (at least the sides under consideration), vertical alignment of the scrims, tearing directions of sides under consideration, number of observed missing yarns at the top of sides under consideration. In general if one of these properties do not correspond between the two pieces of duct tape under consideration, then one does not need to consider the loopbreaking patterns. An example for the background information is

> I : The 2 pieces of duct tape are both 5 cm wide, have a weft insertion scrim, a grey polymeric backing of the same structure, are torn on the sides under consideration and the scrims are vertically aligned. The tearing direction of both patterns is T. The X-side has one missing yarn at the top. The positions of any observed missing warp yarns on each side.

Now that these events are properly defined, we can formulate the LR for this first case

$$LR_{H_p,H_d}(E_X, E_Y) := \frac{\mathbb{P}(E_X, E_Y|H_p, I)}{\mathbb{P}(E_X, E_Y|H_d, I)}.$$

Assuming that the two patterns are independent given H_d and I, we can rewrite this as

$$LR_{H_p,H_d}(E_X, E_Y) = \frac{\mathbb{P}(E_X, E_Y|H_p, I)}{\mathbb{P}(E_X|H_d, I)\mathbb{P}(E_Y|H_d, I)}. \tag{4.1}$$

In Appendix C the case when the tearing directions are unknown is discussed. Omitting the notation of the background information, the LR yields in that case

$$\begin{aligned} &LR(E_X, E_Y|\mathcal{D}_X, \mathcal{D}_Y \in \{T, B\}) \\ &= 2 \cdot \frac{\mathbb{P}(E_{X,T}, E_{Y,T}|H_p) + \mathbb{P}(E_{X,B}, E_{Y,B}|H_p)}{\mathbb{P}(E_{X,T}, E_{Y,T}|H_d) + \mathbb{P}(E_{X,T}, E_{Y,B}|H_d) + \mathbb{P}(E_{X,B}, E_{Y,T}|H_d) + \mathbb{P}(E_{X,B}, E_{Y,B}|H_d)}, \end{aligned} \tag{4.2}$$

where $\mathcal{D}_X, \mathcal{D}_Y$ denote the random variables representing the tearing directions for the X-side and the Y-side respectively, T, B are the possible tearing directions, and for $i \in \{X, Y\}$ and $d \in \{T, B\}$ $E_{i,d}$ represents the loopbreaking pattern E_i with tearing direction d.

For the *general comparison* proposition the likelihood ratio can be expressed in terms of the likelihood ratio for the specific comparison. This is described in Appendix D.

For the evaluation of loopbreaking patterns that contain missing warp yarns, we need to provide a separate method for evaluating tapes that have missing warp yarns at the start of their loopbreaking patterns. These pose a problem, since we set the horizontal position of the first yarn at zero, which is our reference point for both sides. In Appendix E we discuss our solution thoroughly. In short, the idea is that we split up the patterns in the parts that are overlapping and the parts that are not overlapping. Each of these parts is shifted, such that they all start at height zero again. We assume that the probabilities of these shifted parts is equal to the unshifted parts. This is based on the assumption that the missing yarns do not provide any information. We evaluate the probabilities of all these shifted parts separately and assume that

the transition probabilities between the parts are negligible. In this way we can set the reference positions at the start of the patterns and we keep the most informative part of the patterns, namely the overlapping part, intact.

In Appendix E we also discuss how to deal with loopbreaking patterns of two sides where the vertical alignment is uncertain. This is the case for our data under H_d. This can also be relevant when only partial patterns are retrieved from the crime scene, or pieces are partially damaged.

4.2. Structure of dynamical Bayesian network

To estimate the value of the LR for loopbreaking patterns, we will create three models. Each of the models will estimate one of the probabilities in the likelihood ratio Equation 4.1. So we will have two models that model the behaviour of the loopbreaking patterns independent of one another, in order to estimate $\mathbb{P}(E_X|H_d)$ and $\mathbb{P}(E_Y|H_d)$ respectively. We call these models the *one-side models* and they will always have the same representation, but the associated probabilities may differ. Next to that we will construct a *two-side model* to estimate $\mathbb{P}(E_X, E_Y|H_p)$, this will model the behaviour of the two loopbreaking patterns, assuming that they used to be directly attached.

For the models we will assume, for simplicity, that the two loopbreaking patterns under consideration are of the form $E_X = (\boldsymbol{X_n})_{n=0}^{m-1}$ and $E_Y = (\boldsymbol{Y_n})_{n=0}^{m-1}$, where m is the total number of warp yarns in both sides. For the cases where this number is not the same in both sides, we apply the methods as described in Appendix E.

The main assumption for our model is that the loopbreaking patterns of duct tape comply to the first-order Markov property. In the sense that the state of a loop is only dependent on the state of the loop prior to it and of the loop opposite to it. We will describe this as a DBN and use the structure of the DBN to describe the dependencies between the nodes. We will not consider a model with dependencies between all the nodes, since there is not enough data available to estimate the corresponding probabilities. We used `HUGIN EXPERT` [9] to construct and evaluate these models.

The models that we construct consist of three main building blocks: the loops, the horizontal positions and the observation errors. In the following subsections we will describe the build-up of the models, each time adding one of the building blocks to both the one-side models and the two-side model. We will only give the corresponding CPTs of the final models, these are discussed in section 4.3. There, one can also find Table 4.1 in which we present an overview of the nodes from the final models with their corresponding mathematical notation, their parents, their states and the construction of their conditional probability tables.

4.2.1. Loops

First, we only consider the *state of the loops* of the loopbreaking patterns. Thus the evidence that we consider is $E_X = (X_n)_{n=0}^{m-1}$ and $E_Y = (Y_n)_{n=0}^{m-1}$ in general.

One-side models We model the behaviour of the loops as Markov Chains. So we assume for the random variables X_n and Y_n that they comply to the Markov property, i.e.,

$$\mathbb{P}(X_{n+1}|X_n, X_{n-1}, \ldots, X_0) = \mathbb{P}(X_{n+1}|X_n), \tag{4.3}$$

$$\mathbb{P}(Y_{n+1}|Y_n, Y_{n-1}, \ldots, Y_0) = \mathbb{P}(Y_{n+1}|Y_n), \tag{4.4}$$

for all $0 \leq n \leq m-2$. Recall from section 2.3 that each loop can be in one of four states; open, closed, complex or missing. However, when a yarn is missing, its actual state could have been

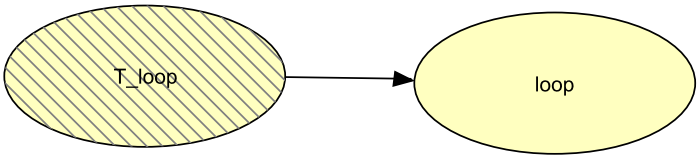


Figure 4.1.: Graph of DBN for the one-side model for loops only. The hatched yellow node represents the state of the loop at height $n-1$ and the plain yellow node represents its state at height n.

any of the three other states. Therefore, we will dismiss the state missing and we let for all $0 \leq n \leq m-1$

$$X_n = \begin{cases} 0, & \text{if loop } n \text{ is whole (or closed),} \\ 1, & \text{if loop } n \text{ is broken (or open),} \\ 2, & \text{if loop } n \text{ is complex.} \end{cases}$$

If X_n was observed as missing, none of the states should be initiated. The states of Y_n are defined analogous. Hence, the corresponding state space for both MCs is $M_1 = \{0, 1, 2\}$.

We can express this as a dynamic Bayesian network as described in section 3.2, see Figure 4.1. In this DBN both nodes have the state space M_1. The CPT of the first node *T_loop* is given by the initial probabilities from the MCs and the CPT of the second node is given by the transition matrices of the MCs.

Two-side model For the two-side model we have as evidence the states of the loops from the paired loopbreaking patterns, i.e., $(E_X, E_Y) = (X_n, Y_n)_{n=0}^{m-1}$. Recall that we are aiming to estimate $\mathbb{P}(E_X, E_Y|H_p)$ here, thus we presume that the X-side and the Y-side under consideration used to be directly connected. Again, we will model this as an DBN, using the assumption of the Markov property for the combined loop states, thus

$$\mathbb{P}(X_n, Y_n|X_{n-1}, Y_{n-1}, X_{n-2}, Y_{n-2}, \ldots, X_0, Y_0) = \mathbb{P}(X_n, Y_n|X_{n-1}, Y_{n-1}). \tag{4.5}$$

The corresponding state space for this DBN is given by $M_2 = \{(x, y) : x, y \in M_1\}$ or

$$M_2 := \begin{Bmatrix} (0,0), & (1,0), & (2,0), \\ (0,1), & (1,1), & (2,1), \\ (0,2), & (1,2), & (2,2) \end{Bmatrix}$$

where 0, 1 and 2 still denote the states closed, open and complex respectively. It is possible to go from each of these states to any of the other states. To describe the DBN, let us first consider a BN, which describes the behaviour of the loops on both sides at a given height. We can create different nodes for the X-side and the Y-side in one model, each of these nodes have state space M_1. Notice that we can write

$$\mathbb{P}(E_X, E_Y|H_p) = \mathbb{P}(E_X|E_Y, H_p)\mathbb{P}(E_Y|H_p) = \mathbb{P}(E_Y|E_X, H_p)\mathbb{P}(E_X|H_p).$$

Let us consider the middle option. This can be modeled as a directed edge from the node for the Y-side towards the node for the X-side. Let us insert the information for the loop on the Y-side being in a certain state, this gives us $\mathbb{P}(Y_n = y_n|H_p)$. If we then include the information for

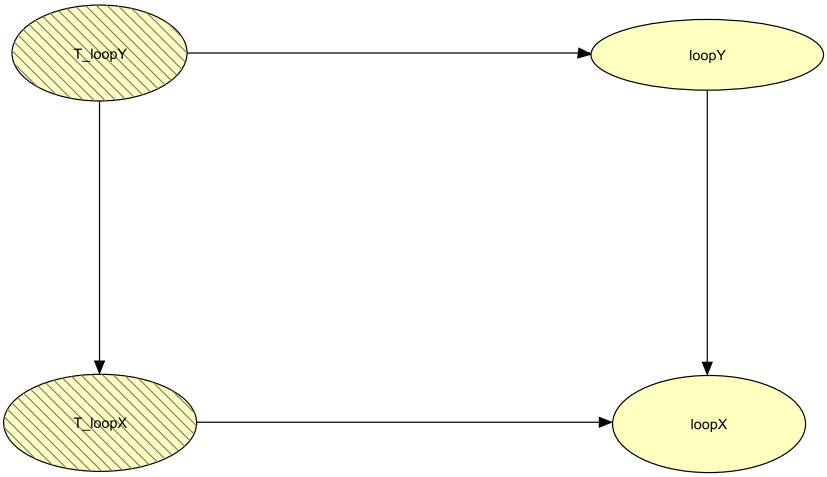


Figure 4.2.: Graph of DBN for the two-side model for loops only.

the loop on the X-side to be in a certain state, we get $\mathbb{P}(X_n = x_n|Y_n = y_n, H_p)\mathbb{P}(Y_n = y_n|H_p)$. Now, we can include the dependency on the state of the former loop, by converting this BN into a DBN. This is done by setting this BN as the second slice of the DBN and adding, so-called, *temporal clones* of both nodes to the first slice of the DBN. We assume that the state of both loops is dependent on the state of its former loop on the same side. However, we assume that the cross-dependence, their dependence on the state of the former loop on the opposite side, is negligible. In other words,

$$\begin{aligned}\mathbb{P}(X_n, Y_n|X_{n-1}, Y_{n-1}, H_p) &= \mathbb{P}(X_n|Y_n, X_{n-1}, Y_{n-1}, H_p)\mathbb{P}(Y_n|X_{n-1}, Y_{n-1}, H_p) &(4.6)\\ &\approx \mathbb{P}(X_n|Y_n, X_{n-1}, H_p)\mathbb{P}(Y_n|Y_{n-1}, H_p) &(4.7)\end{aligned}$$

The DBN corresponding to this analogy is displayed in Figure 4.2.

4.2.2. Horizontal positions

We want to accompany the former models with the horizontal position of each loop. For this we consider the entire loopbreaking pattern as the evidence. Hence, we have $E_X = (\mathbf{X_n})_{n=0}^{m-1} := (X_n, \xi_n)_{n=0}^{m-1}$ where X_n still represents the state of the loop of the nth warp yarn and ξ_n represents the horizontal position of the nth warp yarn, as described in section 2.3. Similarly, we have $E_Y = (\mathbf{Y_n})_{n=0}^{m-1} := (Y_n, \gamma_n)_{n=0}^{m-1}$. By definition, $\xi_0 = \gamma_0 = 0$ and, for $1 \leq n \leq m-1$,

$$\xi_n = \begin{cases} -l, & \text{if the position of the } n\text{th loop is } l \text{ weft yarns to the left of the first loop;} \\ 0, & \text{if the position of the } n\text{th loop is at the same weft yarn as the first loop;} \\ +l, & \text{if the position of the } n\text{th loop is } l \text{ weft yarns to the right of the first loop.} \end{cases} \quad (4.8)$$

The states of γ_n are defined analogous. We can also express the horizontal position in terms of the former horizontal position and the horizontal jump. A *horizontal jump* at height n is the difference between the horizontal positions of the loop at height $n-1$ and the loop at height n, we denote this by μ_n and ν_n for the X-side and the Y-side respectively. In other words, for all $1 \leq n \leq m-1$ we have

$$\mu_n := \xi_n - \xi_{n-1}, \quad (4.9)$$

$$\nu_n := \gamma_n - \gamma_{n-1}. \quad (4.10)$$

Notice that $\xi_n, \gamma_n \in \mathbb{Z}$. However, for the practical implementation we will restrict the state space of the horizontal positions to $\{-4^+, -3, -2, \ldots, +2, +3, +4^+\}$, where -4^+ represents all horizontal positions that are -4 or lower, and $+4^+$ represents all horizontal positions that are $+4$ or higher. We denote the random variables for the horizontal positions restricted to this state space by $\bar{\xi}_n$ and $\bar{\gamma}_n$. Furthermore, in the model, we will only consider the direction of the horizontal jumps. Thus for the horizontal jump we consider the state space of $\operatorname{sign}\mu_n$ and $\operatorname{sign}\nu_n$, which is $\{-1, 0, +1\}$. The random variables that represent the jumps restricted to this state space, so the directions of the jumps, are denoted by $\bar{\mu}_n$ and $\bar{\nu}_n$. These two simplifications are based on the data. In the available data it never occurs that the horizontal position differs four or more from the initial horizontal position. Further, in the data it only rarely happens that the jumps are larger than 1, and only when either one of those loops itself was missing or when one of the opposite loops was missing.

One-side models For the one-side models, we already have the model for only loops, see Figure 4.1. Now, the evidence is accompanied with the horizontal position. By the assumed Markov property we have

$$\mathbb{P}(X_n, \xi_n | X_{n-1}, \xi_{n-1}, X_{n-2}, \xi_{n-2}, \ldots, X_0, \xi_0, H_d) = \mathbb{P}(X_n, \xi_n | X_{n-1}, \xi_{n-1}, H_d). \tag{4.11}$$

Furthermore, let $(x_i, a_i)_{i=0}^{m-1}$ be any realization of the states of the loops and the horizontal positions of the X-side, then for any $1 \leq n \leq m-1$ it holds that

$$\begin{aligned}
&\mathbb{P}(X_n = x_n, \xi_n = a_n | X_{n-1} = x_{n-1}, \xi_{n-1} = a_{n-1}, H_d) \\
&\qquad = \mathbb{P}(\xi_n = a_n | X_n = x_n, X_{n-1} = x_{n-1}, \xi_{n-1} = a_{n-1}, H_d) \\
&\qquad\qquad \cdot \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, \xi_{n-1} = a_{n-1}, H_d) \\
&\qquad = \mathbb{P}(\mu_n = a_n - a_{n-1} | X_n = x_n, X_{n-1} = x_{n-1}, \xi_{n-1} = a_{n-1}, H_d) \\
&\qquad\qquad \cdot \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, \xi_{n-1} = a_{n-1}, H_d).
\end{aligned}$$

Here we will assume that horizontal jump is dependent on the state of the corresponding loop and former loop, but that we can neglect the dependency that the loop state has on the previous horizontal position. Furthermore, we assume that the distribution of the horizontal jump is not affected by the former horizontal position. Under these assumptions it holds that

$$\mathbb{P}(X_n, \xi_n = a_{n-1} + b_n | X_{n-1}, \xi_{n-1} = a_{n-1}, H_d) = \mathbb{P}(\mu_n = b_n | X_n, X_{n-1}, H_d) \mathbb{P}(X_n | X_{n-1}, H_d). \tag{4.12}$$

In this way it would suffice to only consider the former loop, current loop and the horizontal jump in our model. However, we do want to keep track of the current position, this will be needed for the two-side model. Hence, we will include the former position and the current position in the model, where the current position will be determined by the horizontal jump and the former position, using $\xi_n = \xi_{n-1} + \mu_n$. However, we do not consider ξ_n and μ_n, but the random variables with the restricted state spaces $\bar{\xi}_n$ and $\bar{\mu}_n$, so there is still some uncertainty in the state of the current horizontal position given the states of these nodes.

This same analogy holds for the loopbreaking patterns of the Y-side, where (Y_n, γ_n) denote the loop state and the horizontal position of the loop at height n.

The DBN corresponding to this reasoning is depicted in Figure 4.3.

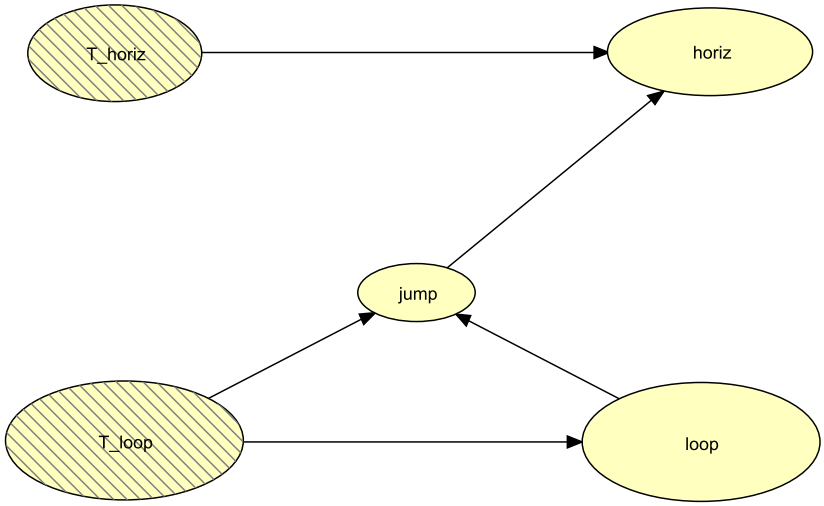


Figure 4.3.: Graph of DBN for the one-side model with loops and horizontal positions.

Two-side model For the two-side model we will apply the same reasoning as for the one-side model. Additionally, we will again assume no cross-dependencies. Further, since we are under H_p, we assume that the horizontal position of the X-side is completely determined by the horizontal position on the Y-side, such that $\bar{\xi}_n = \bar{\gamma}_n$. The DBN corresponding to this reasoning is shown in Figure 4.4.

4.2.3. Observation errors

We want to complete our models by taking the observation errors into account. This can be accomplished by using the method of a hidden Markov chain, however in our opinion it is important to be able to clearly view and easily change the distribution of the observation errors. Therefore, we will replace each node in the models described before, with three new nodes: the *real node* (blue), representing the real state of the node; the *observation node* (yellow), representing the observed state of that node; and an *observational error node* (green), representing the observation error that might have occurred in the observation. This will be done with all nodes with exception of the jump node and the node representing the former horizontal position. These nodes are not affected by observation errors directly. The jump is not something we observe, but is determined by the former and current horizontal position. The first horizontal position is determined to be zero and from then on it will be determined by the former and current horizontal position.

The observed state of the nodes is determined by their real state and the observation error. The state spaces of the observation nodes is equal to the state spaces of their corresponding real nodes. The observation errors on the horizontal positions, $\varepsilon_{H_n,n}$, have an associated state space of $\{-2,-1,0,+1,+2\}$. These states correspond directly to the number of warp yarns with which the position was misjudged. Thus we assume implicitly that the observation errors that occur are never larger than two horizontal positions (i.e., two weft yarns). For the observation errors on the loops, $\varepsilon_{L,n}$, the assigned state space is {yes, no}. These states just denote if there was an observation error or not. In this way we have that

$$\mathbb{P}(X_n^{(o)}|X_n^{(r)}, \varepsilon_{L,n} = \text{no}) = \begin{cases} 1, & \text{if } X_n^{(o)} = X_n^{(r)} \\ 0, & \text{otherwise.} \end{cases}$$

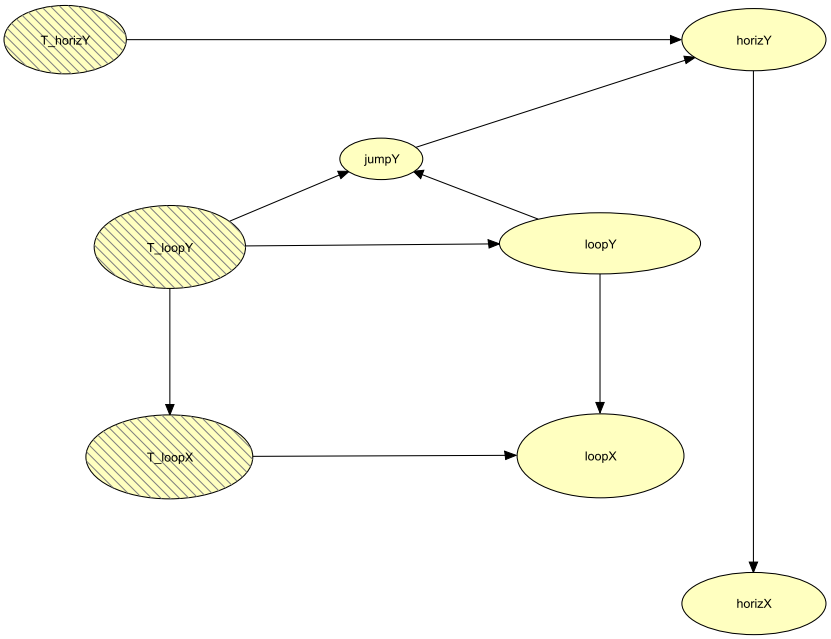


Figure 4.4.: Graph of DBN for two-side model with loops and horizontal positions.

and

$$\mathbb{P}(X_n^{(o)}|X_n^{(r)}, \varepsilon_{L,n} = \text{yes}) = \begin{cases} 0, & \text{if } X_n^{(o)} = X_n^{(r)} \\ \mathbb{P}(X_n^{(o)}|X_n^{(r)}, \varepsilon_{L,n} = \text{yes}), & \text{otherwise.} \end{cases}$$

Notice that an observation error at the first horizontal position, $\varepsilon_{H_0,0}$, should be dealt with extra carefully. The horizontal position of the first loop is used as a reference point. If an error occurs here, the horizontal positions of all the following loops are incorrect. This needs to be taken into consideration, therefore we have two nodes associated to this: `T_hor_first_errX` and `hor_first_errX`, which represent the states of $\varepsilon_{H_0,0}$ and $\varepsilon_{H_0,n}$ respectively. The first has a CPT assigned to it which reflects the probability distribution for these errors, and the second one takes over this distribution. So that for all the following nodes this distributions remains the same and can be taken into consideration for the observed horizontal position. So that $\xi_n^{(o)} = \xi_n^{(r)} + \varepsilon_{H_n,n} - \varepsilon_{H_0,n}$, where $\xi_n^{(o)}$ and $\xi_n^{(r)}$ represent the observed and real state of the horizontal position at height n, $\varepsilon_{H_n,n}$ represents the observation error for the horizontal position at height n and $\varepsilon_{H_0,n}$ represents the observation error for the horizontal position at height 0. However, we still need to be careful for the edge cases, since we only record $\bar{\xi}_n^{(r)}$, which is restricted to the determined state space $\{-4, \ldots + 4\}$. So when $\bar{\xi}_n^{(r)} = \pm 4$ then the value of $\bar{\xi}_n^{(o)}$ is not fully deterministic. It is dependent on the distribution for $\mathbb{P}(\xi_n|\bar{\xi}_n)$.

One-side model For the horizontal positions we assume that the observation error is dependent of the loop state on both sides. As we know, the horizontal positions of complex breaks are not always complementary, so we take this into account with the observation errors. Notice that this is not exactly an observation error, but we can model it as if it is one, since the horizontal position is not clear in those cases. Therefore, in the model for one-side we will also include the real nodes of the loops of the opposite side. We do not enter any observations into these nodes, but their distributions are defined by the loops on the current side. We can use these distributions to get a better estimation of the observation error for the horizontal position. The

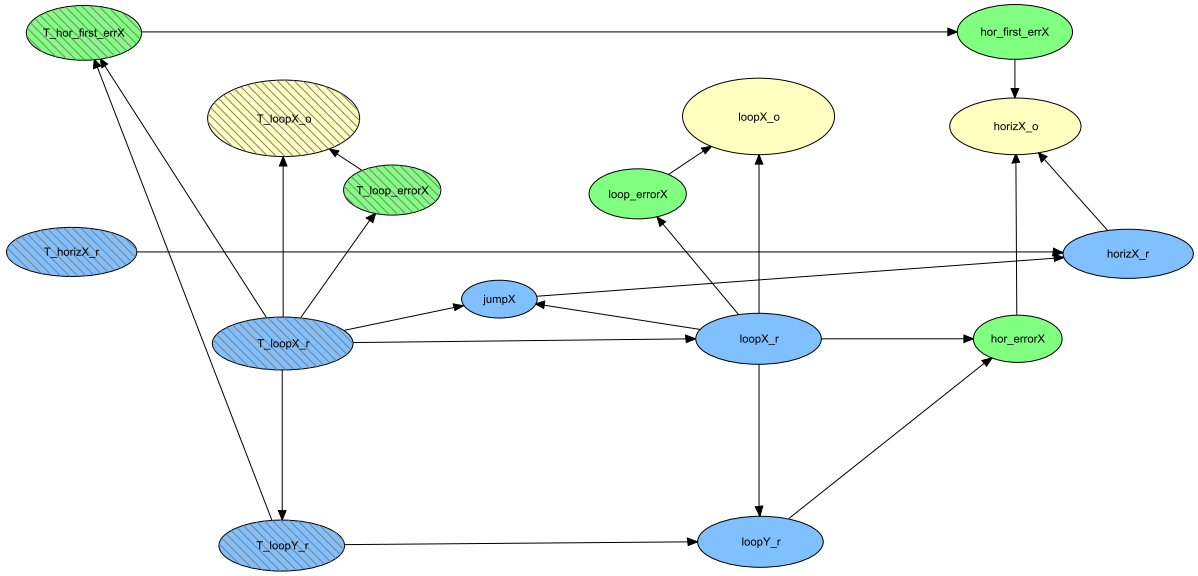


Figure 4.5.: Graph of DBN for one-side model with loops, horizontal positions and observation errors. The blue nodes represent the unobserved (hidden) "real" states, the yellow nodes represent the observed states and the green nodes represent the observation errors.

corresponding graph of the DBN for the one-side model for the X-side is depicted in Figure 4.5. The basis of this graph is given by the structure of the blue (real) nodes, which is retrieved from the DBN of the one-side model with loops and horizontal positions, see Figure 4.3, except that the nodes representing the real states of the loops of the opposite side are added.

This DBN is the final model that we will use to evaluate the probabilities under H_d.

Two-side model For the two-side model there are no additional changes other than described above. The graph of the corresponding DBN is shown in Figure 4.6. The basis of this graph is again given by the structure of the blue (real) nodes, which is retrieved from the DBN of the two-side model with loops and horizontal positions, see Figure 4.4.

This DBN is the final model that we will use to evaluate the probabilities under H_p.

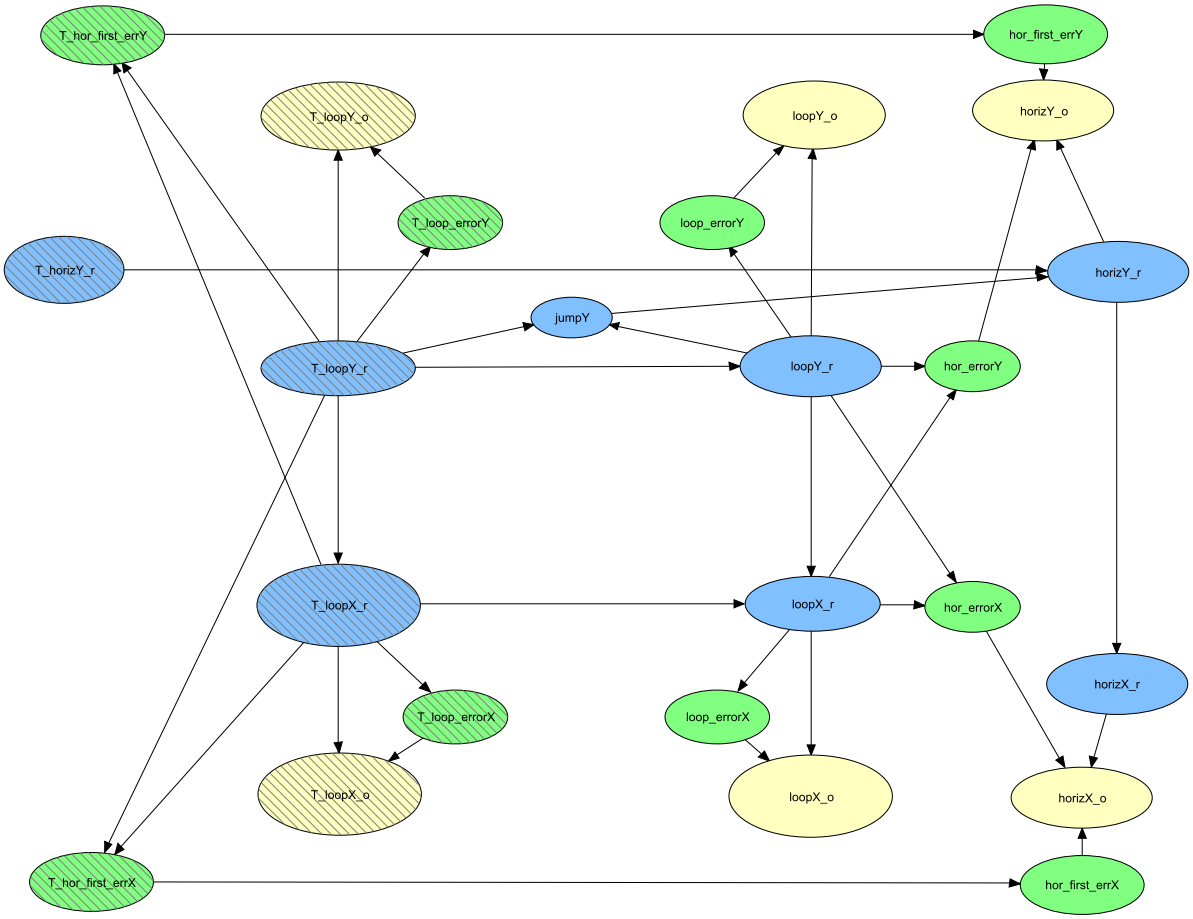


Figure 4.6.: Graph of DBN for two-side model with loops, horizontal positions and observation errors. The blue nodes represent the unobserved (hidden) "real" states, the yellow nodes represent the observed states and the green nodes represent the observation errors.

4.3. Conditional probability tables

We have to assign a conditional probability table (CPT) to each node in the models we constructed before. We will do this for the final models. In Table 4.1 we give an overview of the different nodes and their associated CPTs. In Appendix F one can find the values of these CPTs.

Recall that there are some nodes that have (mostly) deterministic CPTs, we call these model induced CPTs. Other CPTs are based on the data. We will use a subset of the available data set as training data for these CPTs, this will be further discussed in chapter 5. The probabilities for these CPTs, $\pi_{ij} = \mathbb{P}(Y = j|x = i)$, are estimated using the MLE estimate of the logistic regression, i.e. by their fractions, as described in section 3.4. Notice, that for this we assume that the categories for x are independent. However, since our data set is small in some cases there were no observations for certain combinations of events. Except for the cases which represent two closed loops opposite of each other, we have no indication that these probabilities should be set to zero. Hence, for these cases we expect that their probabilities are small, but not zero. To overcome this, we smoothed the estimates for these events, as suggested in [2, 17]. We used a Dirichlet prior distribution with equal-valued parameters, such that their estimated posterior probabilities are given by

$$\hat{\pi}_{ij} = \frac{N_{jl} + 1}{N_{\cdot l} + K + 1},$$

where $K+1$ represents the number of categories that the output Y has. This is called a *(Laplace) smoothed estimate*, a *maximum a posteriori (MAP) estimate* or a *Bayes' estimate* [2, 17].

For the estimation of the CPTs for the nodes `T_loopX_r` and `T_loopY_r`, we will not restrict to the observations of the first yarns, but instead, we use the observations of all yarns in the patterns. Due to the shifting of patterns with missing yarns at the top, these nodes do not necessarily only represent the state of the first yarns, as discussed in section E.2.

Unfortunately at this point we do not have any data about the observation errors. Therefore, we cannot make an estimation about the values in the CPTs for the error nodes and the observation nodes, based on the data. For now we will set these values based on the expert knowledge of the forensic examiners. Note that for the observation nodes it holds that given that there was no error, then the observation will be equal to the real value. For the loop states we will assume that the probability of observing a closed loop while it was actually open and vice versa is very small. So given that the real loop state is open or closed and there was an observation error, then it is much more likely to observe a complex loop than to observe a closed or open loop respectively.

Node	**Parents**	**State space**	**Probabilities**
`T_loop_r` $(X_0^{(r)}, Y_0^{(r)})$		Closed; Complex; Open	Estimate from data.
`loop_r` $(X_n^{(r)}, Y_n^{(r)})$	`T_loop_r`	Closed; Complex; Open	Estimate from data.
`jump` $(\bar{\mu}_n, \bar{\nu}_n)$	`T_loop_r`, `loop_r`	$-1, 0, +1$	Estimate from data.
`T_horiz_r` $(\bar{\xi}_0^{(r)}, \bar{\gamma}_0^{(r)})$		$-4, -3, \ldots, 3, 4$	Deterministic, such that $\mathbb{P}(\bar{\xi}_0^{(r)} = 0) = 1$.
`horiz_r` $(\bar{\xi}_n^{(r)}, \bar{\gamma}_n^{(r)})$ (not for X-side in two side model)	`T_horiz_r`, `jump`	$-4, -3, \ldots, 3, 4$	In principle deterministic $\xi_n^{(r)} := \xi_{n-1}^{(r)} + \delta_n$, but due to limited state spaces we need estimates for $\mathbb{P}(\xi_n \vert \bar{\xi}_n)$ and $\mathbb{P}(\delta_n \vert \bar{\delta}_n)$.
`horizX_r` $(\bar{\xi}_n^{(r)})$ (only for X-side in two side model)	`horizY_r`	$-4, -3, \ldots, 3, 4$	Deterministic, such that $\xi_n := \gamma_n$.
`T_loop_r'` $(Y_0^{(r)}, X_0^{(r)})$	`T_loop_r`	Closed; Complex; Open	Estimate from data.
`loop_r'` $(Y_n^{(r)}, X_n^{(r)})$	`T_loop_r'`, `loop_r`	Closed; Complex; Open	Estimate from data.
`T_hor_first_err` $(\varepsilon_{H_0,0})$	`T_loop_r`, `T_loop_r'`	$-2, -1, 0, 1, 2$	Calculated guess based on expert knowledge. Identical to probabilities for `hor_error`.
`hor_first_err` $(\varepsilon_{H_0,n})$	`T_hor_first_err`	$-2, -1, 0, 1, 2$	Deterministic $\varepsilon_{H_0,n} := \varepsilon_{H_0,n-1} = \varepsilon_{H_0,0}$.
`hor_error` $(\varepsilon_{H,n})$	`loop_r`, `loop_r'`	$-2, -1, 0, 1, 2$	Calculated guess based on expert knowledge. Identical to probabilities for `T_hor_first_err`.
`horiz_o` $(\bar{\xi}_n^{(o)}, \bar{\gamma}_n^{(o)})$	`horiz_r`, `hor_error`, `hor_first_err`	$-4, -3, \ldots, 3, 4$	In principle deterministic $\bar{\xi}_n^{(o)} := \bar{\xi}_n^{(r)} - \varepsilon_{H_0,n} + \varepsilon_{H,n}$, but again due to limited state spaces we need estimates for $\mathbb{P}(\xi_n \vert \bar{\xi}_n)$.
`T_loop_error`$(\varepsilon_{L,n})$	`T_loop_r`	yes; no	Calculated guess based on expert knowledge. Identical to probabilities for `loop_error`.
`loop_error` $(\varepsilon_{L,n})$	`loop_r`	yes; no	Calculated guess based on expert knowledge. Identical to probabilities for `T_loop_error`.
`T_loop_o` $(X_0^{(o)}, Y_0^{(o)})$	`T_loop_r`, `T_loop_error`	Closed; Complex; Open	In the case that $\varepsilon_{L,n} =$ no, then deterministic, such that $X_0^{(o)} := X_0^{(r)}$. For the cases where $\varepsilon_{L,n} =$ yes, we make a calculated guess based on expert knowledge. Identical to probabilities for `loop_o`.
`loop_o` $(X_n^{(o)}, Y_n^{(o)})$	`loop_r`, `loop_error`	Closed; Complex; Open	In the case that $\varepsilon_{L,n} =$ no, then deterministic, such that $X_n^{(o)} := X_n^{(r)}$. For the cases where $\varepsilon_{L,n} =$ yes, we make a calculated guess based on expert knowledge. Identical to probabilities for `T_loop_o`.

Table 4.1.: Overview of all the nodes in the models. With `<node>` we denote the node of the side under consideration (for the two side model this is the Y-side), with `<node>'` we denote the node of the opposite side (for the two side model this is always the X-side). After the name of the node in the DBNs, the associated notation of the node in mathematical representations is given in brackets, with first the notation for the X side and second for the Y side, if relevant.

5. Validation

The model proposed in chapter 4 provides LR-values for the loopbreaking patterns of two pieces of duct tape, but are these LR-values valid? The model must be valid, in order to be applicable in forensic cases. In this section we will evaluate the performance of that model. First we discuss the methods that are used for this purpose. Second we will discuss the results of these validation methods for our proposed model. However, this is merely performed as a proof-of-concept, since the available data set is small and might not be a good representation of the entire population.

5.1. Validation methods

There is not yet an established standard protocol to validate likelihood ratio systems. With that in mind, Meuwly et al. [16] wrote a guideline for the validation of methods that calculate LRs. They consider both feature-based and score-based methods. Our model could be seen as a feature-based method, we form an LR directly from the features (i.e., the loopbreaking patterns), without first converting them to scores. They state that the validation process should consist of theoretical validation and empirical validation. The theoretical framework of our model has been discussed intensively, so we will focus here on the empirical validation. They explain that we first need to define the *performance characteristics*, the *performance metrics* and the *performance criteria*. Performance characteristics are characteristics of the method that should be evaluated, which are measured by corresponding *performance metrics*. The *performance criteria* are criteria for which outcomes of the performance metrics the method is deemed valid. The latter should be based on the current baseline method.

Meeuwly et al. state that the validation process should contain two stages: the *method development stage* and the *validation stage*. For the fist stage, a designated database is required, which should be split into a training database and a test database. These databases should be non-overlapping, independent and representative, and they should contain the ground truth labels of the data. The ground truth of a datapoint is the proposition that is true for that datapoint, i.e., H_p or H_d. The model parameters are obtained using the training database. The test database is used to evaluate the performance of the LR-system on this unseen data. The LR-values of the datapoints in the test data are calculated using the trained model and the performance metrics for this set of LRs are evaluated. If the criteria are not yet met, the model should be adjusted to increase its performance. If the criteria are met, then one can continue to the validation stage. Here another database, the validation database, should be used. This database should contain data that is representative for realistic casework. For the method to be used in casework, it should, of course, also perform well on this database and meet the performance criteria.

In our case, unfortunately we do not have data about the current baseline method, which is the evaluation by the forensic experts. Therefore, we will not define the performance criteria. Furthermore, we do not have access to a separate database which could be used as a validation database. Instead we will split our database into a training database and a test database, which are non-overlapping, but not independent and we will give a proof-of-concept for the evaluation of the proposed LR-system, this will be discussed in section 5.2. In future work these results could

be compared to the current baseline method and the method could be applied to a validation database.

5.1.1. Performance characteristics

Validity or performance of a system is almost always associated with the *accuracy* of the system [18, 21, 32]. Even the first sentence of the Wikipedia page on validation (in statistics) reads: *"Validity is the extent to which a concept, conclusion or measurement is well-founded and likely corresponds accurately to the real world."* [28]. In general, we say a system is valid when its outcomes correspond to the actual values. The *accuracy* of a system is defined to be the closeness of a given magnitude to its true value [32]. However, as mentioned by Leegwater et al. [11], the application of this definition to the likelihood ratio framework is not straightforward. There are no true LR-values available, so it is not possible to compare the LR-values produced by the model with the true LR-values. Moreover, the notion of a true LR-value is still subject of discussion. We will apply the notion of *accuracy* as proposed by Meuwly et al. [16], as the extent the LR-value supports the ground truth.

The accuracy of a set of LRs can be decomposed into two components: the *discriminating power* and the *calibration*. The *discriminating power* is defined as the degree to which the LR-values are discriminating based on their corresponding ground truths [16]. This can be visualised as the relative overlap of the histogram for the LR-values given H_p and the histogram for the LR-values given H_d. The degree of *calibration* of a set of LRs is defined as the extent to which the LRs are exactly as ensured by the data [16]. In other words, the likelihood ratio of an LR-value should be equal to that value ("the LR of the LR is the LR"). This is a theoretical property of the likelihood ratio, as proven in [13], and therefore it is a desirable property of the set of LRs. This definition is based on the notion of calibration of probability assessments. As explained in [21], suppose there are N probability assessments made for some random variable Z_i for $1 \leq i \leq N$, which can take one of two values each time z_p or z_d. This set of probability assessments is then called *well-calibrated*, if the fraction of probability assessments $\mathbb{P}(Z_i = z_p|K) = q \pm \delta$ that are correct is q, for some small $\delta > 0$.

5.1.2. Performance metrics

There are several metrics which can be used to quantify and compare the performance characteristics defined above. The metrics that we will use are based on *strictly proper scoring rules* (SPSR). A *scoring rule* is a function, $S(P, x)$, which value describes a loss (or cost) assigned to a probability distribution given to a certain unknown variable, P, depending on the true value of this variable, x [6]. A scoring rule is called *striclty proper* if for the best possible probability distribution Q and for all $P \neq Q$ then $S(P, Q) > S(Q, Q)$ and $S(P, Q) = S(Q, Q)$ if and only if $P = Q$, where $S(P, Q)$ denotes the expected values of $S(P, \cdot)$ under Q, as described in [6].

First, we consider a measure for the loss of accuracy: the *empirical cross-entropy (ECE)*. As described in [32] the ECE is a logarithmic strictly proper scoring rule. It is defined as

$$ECE := -\frac{\mathbb{P}(H_p|I)}{N_p} \sum_{i:H_p \text{ true}} \log_2 \left(\mathbb{P}(H_p|E_i, I)\right) - \frac{\mathbb{P}(H_d|I)}{N_d} \sum_{j:H_d \text{ true}} \log_2 \left(\mathbb{P}(H_d|E_j, I)\right),$$

where E_i represents the evidence of the ith observation in the test set, and N_p and N_d denote the total number of observations where H_p or respectively H_d are true. We can rewrite this in

terms of LRs and prior odds in favour of H_p ($O(H_p|I) := \frac{\mathbb{P}(H_p|I)}{\mathbb{P}(H_d|I)}$) as

$$ECE = \frac{\mathbb{P}(H_p|I)}{N_p} \sum_{i:H_p \text{ true}} \log_2 \left(1 + \frac{1}{LR_i \cdot O(H_p|I)}\right) + \frac{\mathbb{P}(H_d|I)}{N_d} \sum_{j:H_d \text{ true}} \log_2 \left(1 + LR_j \cdot O(H_p|I)\right). \quad (5.1)$$

We can evaluate the ECE on a range of prior odds. The ECE is a strictly proper scoring rule, hence it measures the loss in accuracy of the LR-system, by penalizing LR-values that do not support the corresponding ground truth and the magnitude of this penalty increases with the level of support for the other hypothesis.

Now to find a measure for the discriminating power and the calibration, recall that accuracy can be split into discriminating power and calibration. So we want to find which part of the loss in accuracy is due to losses in discriminating power and which part is due to losses in calibration. As proposed in [5] this can be determined by applying the *pool adjacent violators (PAV) algorithm* to the LRs in the test set. This is an algorithm that performs isotonic regression by identifying the step-wise constant, isotonic (i.e., non-decreasing) function that best fits the data based on a mean-squared error criterion [33]. The basics of applying the PAV algorithm to an LR-system is described by Brümmer and du Preez in [5] as follows: (1) Sort the LR-values in ascending order; (2) Replace each LR-value with a posterior probability based on their ground truth, one if the ground truth is H_p and zero if the ground truth is H_d; (3) Apply the PAV algorithm, which iteratively pools adjacent values that violate the monotonicity constraint of the isotonic function and replace all values in that region by the mean over that region; (4) Apply Bayes' rule to obtain the (log) LRs from the posteriors, where the prior odds are the proportion of H_p ground truths in the set; (5) Revoke the sorting done at step 1 to recover the (log) LR-values corresponding to the original LR-values. We call the set of LRs obtained with this method *PAV LRs*. These PAV LRs are optimally calibrated, i.e., they do not have any loss in calibration. Hence, all of the loss in accuracy is due to loss in discriminating power. Thus a measure for the discriminating power is the ECE of the PAV LRs, also denoted by ECE^{min}. Subsequently a measure for the calibration is the difference between the ECE and ECE^{min}. To represent all of these metrics, one can make an ECE plot, which consists of three curves: one representing the ECE-values of the original LRs, one representing the ECE-values of the PAV LRs and one representing neutral reference values for a range of prior odds. The neutral reference values are generated by always giving an LR of 1, which does not support any of the two propositions. This reference curve can be seen as a floor of performance.

Another performance metric, which has been increasing in popularity is the *log-likelihood-ratio cost* (*Cllr*) [18]. It was introduced by Brümmer and du Preez [5]. Cllr is also a logarithmic strictly scoring rule, and it also measures the loss in accuracy, it is defined as

$$Cllr = \frac{1}{2}\left(\frac{1}{N_p} \sum_{i:H_p \text{ true}} \log_2 \left(1 + \frac{1}{LR_i}\right) + \frac{1}{N_d} \sum_{j:H_d \text{ true}} \log_2 \left(1 + LR_j\right)\right). \quad (5.2)$$

As opposed to the ECE, the Cllr does not depend on prior odds. It can be interpreted as the averaged cost of a decision over all the prior probabilities and costs involved [16]. Despite their difference in interpretation, the Cllr can be expressed in terms of the ECE. Namely Cllr equals ECE where $O(H_p|I) = 1$. Notice that for a neutral set of LRs, i.e., all LR-values are equal to one, the Cllr is equal to one. A Cllr smaller than one, gives thus an indication of the degree to

which the system performs better than the neutral set. In a similar fashion as for the ECE, we can decompose the Cllr, such that $Cllr = Cllr^{min} + Cllr^{cal}$, where $Cllr^{min}$ represents the loss in discriminating power and $Cllr^{cal}$ represents the loss in calibration. Note that again $Cllr^{min}$ equals ECE^{min} where $O(H_p|I) = 1$.

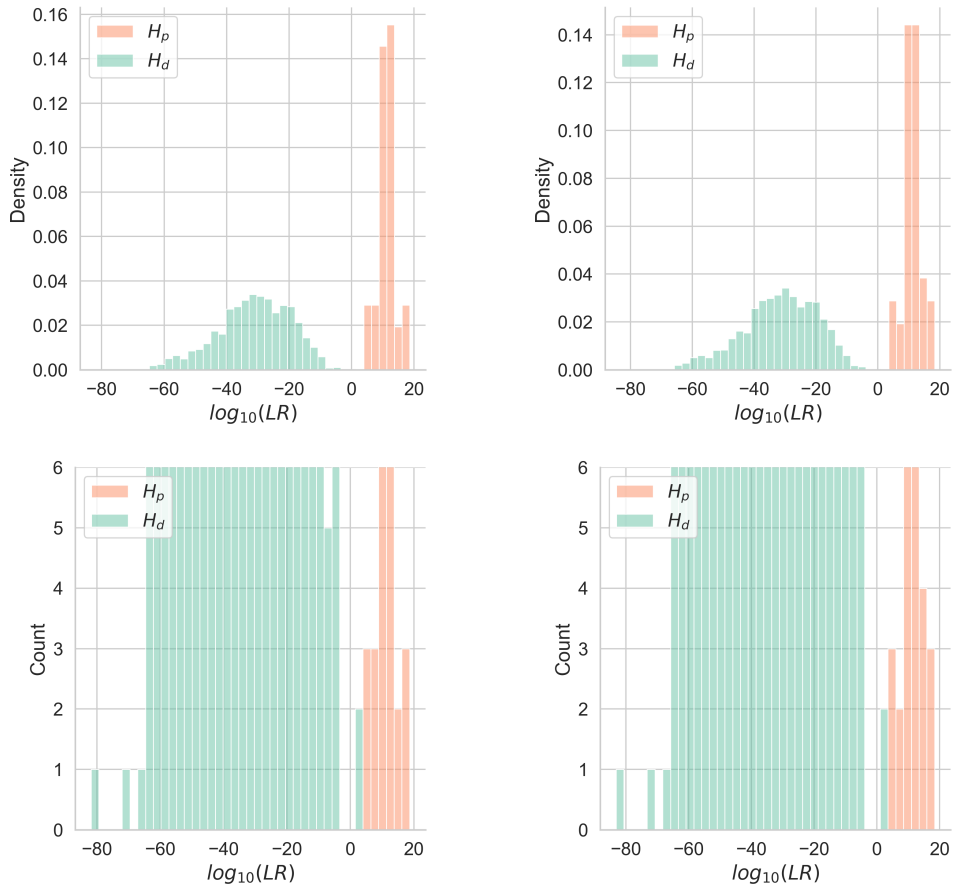
5.1.3. Data sets

To evaluate the performance of the system we used the available data, which we already discussed in section 2.4. Recall that person C applied a different tearing direction than persons A and B. This implies that in practice the patterns of person C could have never been attached to the patterns of person A and B, since the first loops of these patterns are on opposite sides of the tape, see Appendix A. However, since the data set is small, we do actually combine the patterns of person C with those of person A and B, in such a way that we assume that the patterns of person C were actually observed in the reversed direction. In other words, the first loops remain the ones that were first torn, but we pretend that they were observed at the top instead of the bottom of the tape.

We split the available data, $\mathrm{D_{all}}$, into a test set, $\mathrm{D_{test}}$, and training set, $\mathrm{D_{train}}$. The 127 known matching pairs in $\mathrm{D_{all}}$ are split randomly over $\mathrm{D_{test}}$ (33%) and $\mathrm{D_{train}}$ (67%). The other 18 patterns, which do not have a complementary pattern in the data set, were added to $\mathrm{D_{test}}$, since they cannot be used for the training of the two-side model. So the test set consists of 51 patterns for both the X-side and the Y-side. Thus in total there are $51^2 = 2601$ combinations of an X-side and an Y-side (or data points) possible in the $\mathrm{D_{test}}$. The number of combinations for which H_p is true is 42. For the other 2559 possible combinations, H_d is true. For these combinations of patterns, the vertical alignments are unknown. To overcome this, we assume that the patterns can only vertically align if the number of observed warp yarns (including the observed missing yarns) differ at most one. In the case that the number of observed warp yarns are equal, we assume that either no missing yarns remained unobserved or there were unobserved missing yarns on both sides, but those were at the same vertical positions on both sides. So in this case only one possible vertical alignment is considered. In the case that the number of observed warp yarns differ exactly by one, we assume that there exists one additional (unobserved) missing warp yarn at the top or bottom of the pattern of the side with the least amount of observed warp yarns. Besides that we assume again that either no other missing yarns remained unobserved or there were other unobserved missing yarns, but those were unobserved on both sides and at the same vertical position. Thus in this case two possible vertical alignments are considered, this is evaluated as described in Appendix E. For consistency these assumptions regarding the vertical alignments are also applied to the combinations of patterns in $\mathrm{D_{test}}$ for which H_p is true, even though the vertical alignment of these combinations were described in the data.

This procedure causes some combinations under H_d in $\mathrm{D_{test}}$ to be dismissed due to their difference in observed number of warp yarns. The number of combinations of patterns in $\mathrm{D_{test}}$ for which H_d is true, reduced to 2520. The combinations of patterns for which H_p is true do not differ more than one in observed length, thus the number of these combinations remains the same.

The training set consists of the remaining 85 X-sides and their respective complementary Y-sides. These were used to estimate the values of the CPTs for the nodes `T_loop_r`, `loop_r` and `jump` in the final models. Here we did use the known vertical alignment of the known matching pairs for the estimation of the CPTs in the two-side model. Notice that splitting the data in this manner, makes the two data sets not independent from one another, in the sense that in both sets there is data for every person-tape combination.



(a) Model with training data D_{train}.

(b) Model with training data D_{all}.

Figure 5.1.: Histograms of the $\log_{10}(LR)$-values where the test data is D_{test}. At the top are the density histograms and at the bottom the respective count histograms zoomed in.

5.2. Validation results

5.2.1. Test and training data

In Figure 5.1a we see the histograms of the resulting LRs when we use D_{train} as the training set and D_{test} as the test data set. Initially (Figure 5.1a top) it seems that the system is fully discriminating and that all the LRs under H_p are larger than 1 and all the LRs under H_d are smaller than 1. When we zoom in (Figure 5.1a bottom) we see that the system is indeed fully discriminating on this test set and that all the LRs under H_p are actually larger than 1. However, we also observe that there are two data points under H_d which have an LR larger than 1, i.e., misleading evidence. Since this data is entirely discriminating, all the loss in accuracy is caused by loss in calibration.

There exists a process, which improves the level of calibration, which is also called *calibration*. To avoid confusion, we will refer to this process only as the *calibration process*. This is a process of performing transformations on the set of LR-values, such that the loss in calibration decreases, which has been considered mainly for score-based LR-systems in speaker recognition, see e.g., [29]. Since this set of LRs is fully discriminated, this calibration process would just shift the

LRs, such that all of the LRs under H_d would become smaller than one, while keeping the set of LRs under H_p larger than one. However, we will not perform this method, since this could create overfitting on the test set. Moreover, we would lose the appealing property that the LR-system is explainable.

Furthermore, it is important to remark that the smallest LR that we find in this test set is approximately 10^{-80}, this is extremely small, however still within machine precision.

We also evaluated the LRs of the test set on the system where we used the whole data set $\mathrm{D_{all}}$ as training data, the corresponding CPTs are given in section F.2.1. This is usually not advisable, but it is interesting to compare the results with the model trained on only $\mathrm{D_{train}}$. The corresponding histograms are shown in Figure 5.1b. We see that there are no significant changes in the resulting distributions. This could imply that more data does not necessarily improve the performance of the model. Additionally, for these two models (with $\mathrm{D_{train}}$ as training data and $\mathrm{D_{all}}$ as training data) we evaluated the results for the training set $\mathrm{D_{all}}$. The resulting LRs are depicted in Figure 5.2. Here we observe again that the distributions do not significantly change when the amount of data in the training set is increased. We also notice that now the system is not fully discriminating, there are still no data points under H_p which have LRs smaller than 1, but there are quite some data points under H_d that are larger than 1. In Figure 5.3 the corresponding ECE-plots are depicted. Notice that the line for the PAV LRs is always zero when the system is fully discriminating on the test data. It is interesting to observe that the system with training data $\mathrm{D_{train}}$ gives very similar performance as the system that was trained on $\mathrm{D_{all}}$. Furthermore, as expected the discriminating power of the two systems is reduced when evaluating all the patterns in $\mathrm{D_{all}}$, but this is still limited.

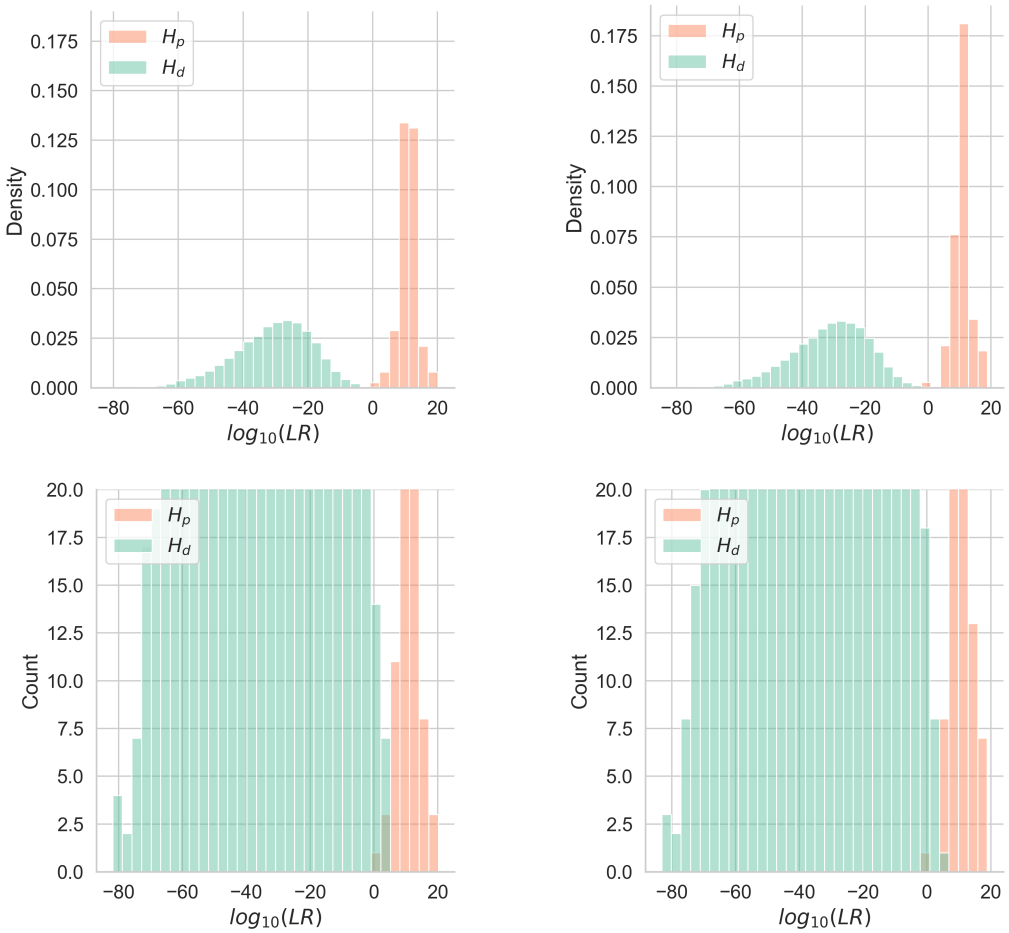
For all the following validation analysis we will only use $\mathrm{D_{train}}$ as the training data set.

5.2.2. Simulated test data

To evaluate the performance of the LR-system on a larger dataset, we simulated test data of loopbreaking patterns with 36 loops.

First of all, we simulated 1000 complementary pairs of loopbreaking patterns, which were based on the behaviour of the patterns in $\mathrm{D_{test}}$. These patterns contain some randomness. For each pair of patterns, one side was set as the basic side and the states of the loops were set independently from each other with a probability distribution based on the patterns in $\mathrm{D_{test}}$, the state of the horizontal position of the first loop was set to zero and at each height there was a probability, based on the patterns in $\mathrm{D_{test}}$, that a jump to the left or right occurred. The pattern opposite of the basic side was set to be complementary to the states of the basic side, however some randomness, based on the $\mathrm{D_{test}}$, was involved, such that open loops were possible to sit opposite of one another and complex loops could occur everywhere, also inconsistency in the horizontal positions were possible. In this way the simulated patterns do not depend on the training data. Moreover, the Markov property is not assumed for these patterns.

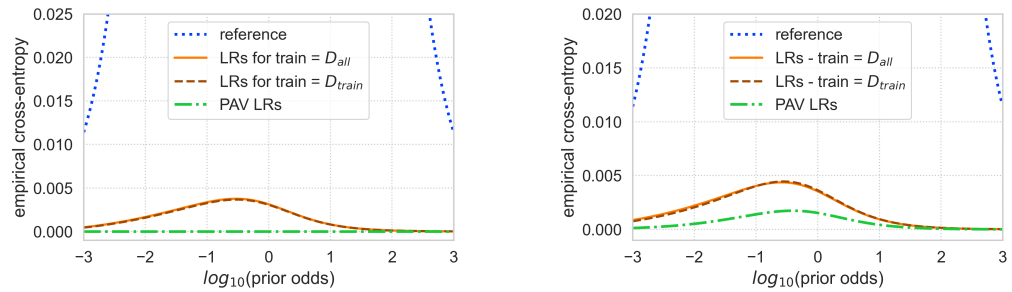
We combined each pattern from this set of simulated patterns with its opposite (H_p is true) and 50 other patterns from the set (H_d is true). Let us denote this set of combinations of simulated patterns by $\mathrm{S_{test}}$. In Figure 5.4 the LRs obtained from this set are shown in a histogram. Notice that, compared to the distribution of the LRs obtained from $\mathrm{D_{test}}$, the mean of the $\log_{10}(LR)$-values under H_d is shifted further to the left and the tail is even longer. For this set, the LR-system is able to completely discriminate between the combinations of patterns for which H_p is true, and the combinations for which H_d is true. Moreover, all of the LRs for combinations under H_p are larger than one, and all of the LRs for combinations under H_d are smaller than one. The performance of the system on this set of patterns is thus even better than



(a) Model with training data D_{train}.

(b) Model with training data D_{all}.

Figure 5.2.: Histograms of the $\log_{10}(LR)$-values where the test data is D_{all}. At the top are the density histograms and at the bottom the respective count histograms zoomed in.



(a) Test data is D_{test}.

(b) Test data is D_{all}.

Figure 5.3.: ECE-plots for the system trained on D_{all} and for the system trained on D_{train}, with testing data D_{test} and D_{all}.

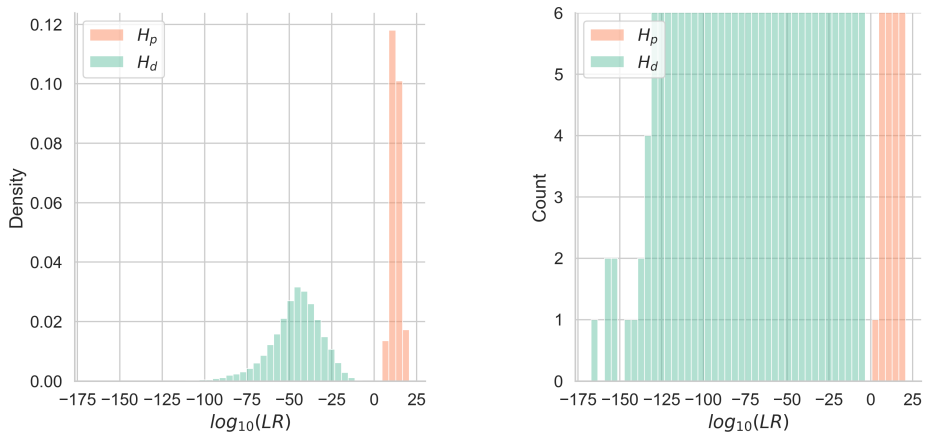


Figure 5.4.: Histograms of the $\log_{10}(LR)$-values for the patterns from $\mathrm{S_{test}}$. On the left is the density histogram and on the right is the count histogram zoomed in.

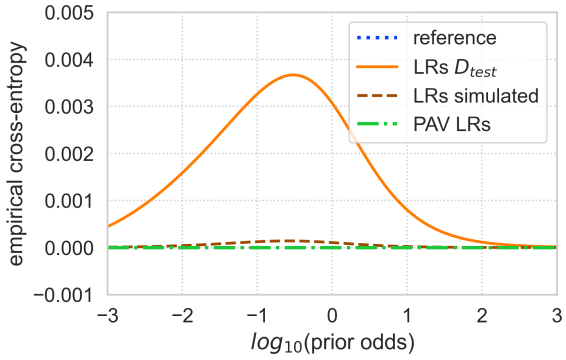


Figure 5.5.: ECE-plots of the LRs obtained from $\mathrm{D_{test}}$ (LRs D_{test}) and for the LRs obtained from $\mathrm{S_{test}}$(LRs simulated).

for the set $\mathrm{D_{test}}$. This can also be observed in Figure 5.5 and Table 5.1. A possible explanation for this high performance, is that the simulated patterns might contain less randomness than the original patterns. Additionally, it might be the case that the simulated patterns contain horizontal positions which are further from zero, and therefore are more differentiated under H_d.

Second of all, we simulated *plain patterns*, patterns which have little variety in the states of the loops and horizontal positions. The forensic experts noted that they observe these type of patterns sometimes and they are uncertain how to evaluate their evidential strength, since they do not seem to give much information. However these types of patterns did not occur in the data set. Therefore, we simulated them. We generated four *perfect plain* patterns for each side. These perfect plain patterns have no jumps and all the loops are in the same state, such that for each side two of the four perfect plain patterns consist of loops that are all closed (*perfect-open*) and the other two have loops that are all open (*perfect-closed*). Additionally, we generated 16 *semi-plain patterns* for both sides. The loops of these patterns are mostly all open or all closed, but each loop has a small probability to be complex. The horizontal positions of these patterns were established by setting the first to zero and at each height there was a probability to jump

	Patterns from D_{test}	Patterns from S_{test}
$Cllr$	0.003065	0.000105
$Cllr^{min}$	0	0
$Cllr^{cal}$	0.003065	0.000105

Table 5.1.: Cllr values for the patterns from D_{test} and S_{test}, rounded down.

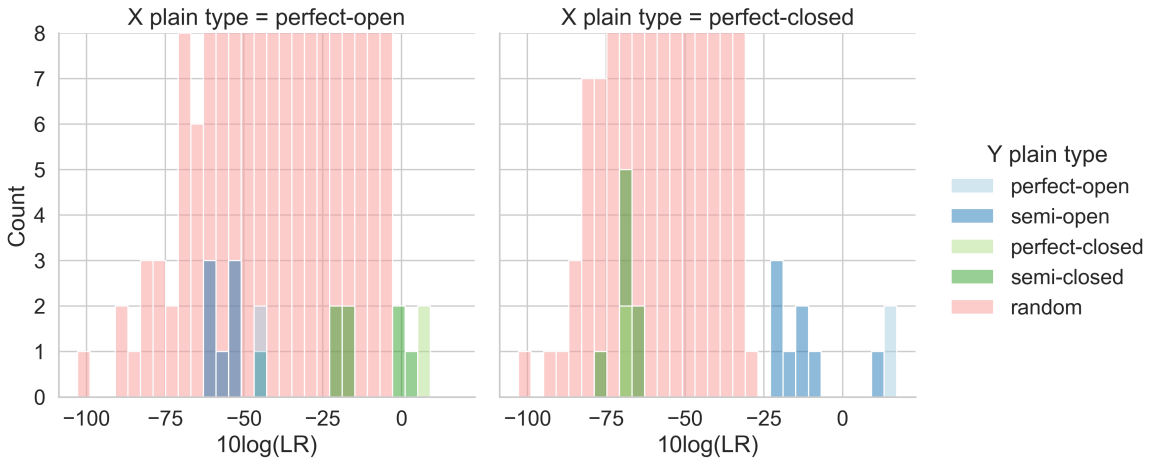


Figure 5.6.: Histograms for the $\log_{10}(LR)$-values obtained from the comparison of one perfect plain pattern for the X-side with different types of simulated patterns for the Y-side. Left: the pattern for the X-side contains only open loops; right: the pattern for the X-side contains only closed loops.

to the right or left. For each of these patterns we also generated a possible opposite side. So that in total there were 16 semi plain patterns for each side. The loops of the patterns of the opposite side were mostly complementary to the loops of the original sides. The horizontal positions as well. However, we applied some randomness to these states.

We combined one perfect-open pattern and one perfect-closed pattern for the X-side with all the simulated patterns for the Y-side, i.e., the 1000 patterns which had randomness based on D_{test}(random), the eight semi-plain patterns which had mostly open loops (semi-open), the eight semi-plain patterns which had mostly closed loops (semi-closed), the two perfect-open patterns and the two perfect-closed patterns. The resulting LR-values are depicted in Figure 5.6. Notice that the system was able to discriminate the random patterns from the perfect patterns. It also gave the highest LR-value for the complementary perfect pattern, which is what we would expect. Further, we observe that the system even discriminates quite well between the perfect patterns and their complementary semi-perfect patterns, only two of the semi-closed patterns on the Y-side got an LR-value of larger than 1 when combined with a perfect-open pattern on the X-side, and for the comparison with the perfect-closed pattern on the X-side only one semi-open pattern of the Y-side received an LR-value larger than one.

5.2.3. Altered patterns

To analyse the performance of the LR-system trained with D_{train}, even more, we made two types of alterations on the patterns in D_{test}. First, we consider the performance of the LR-system on

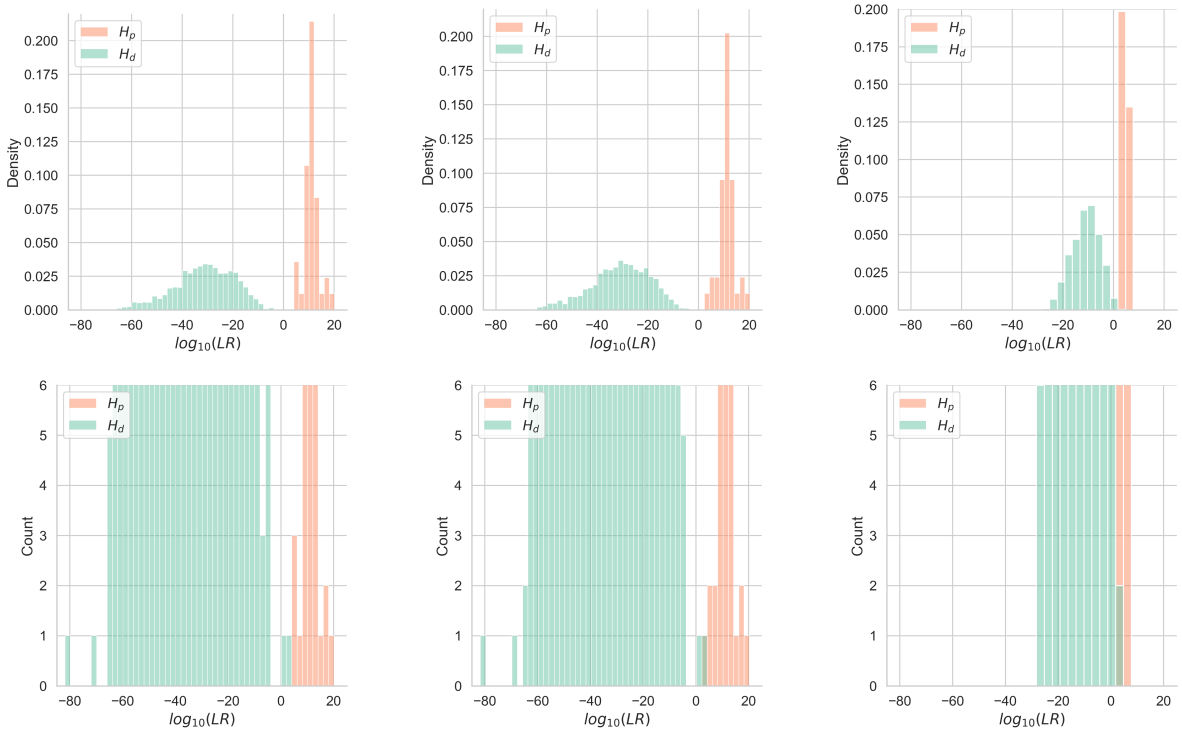
the patterns with reversed tearing directions. Second, we study the performance on shorter patterns (with the original tearing direction).

Reversed tearing direction For all the pieces of duct tape that are in D_{all}, it is known in which direction the person tore. The top of each loopbreaking pattern is set to the starting point of the tear. In real court cases however, the tearing direction is usually unknown. The forensic experts from the NFI suspect they can observe an indication of the tearing direction on the tape in most cases, but they have never been able to confirm this hypothesis, since the ground truth was unknown. Thus we will analyse the performance of our system on the set $D_{test,rev}$, which consists of all the patterns from D_{test}, but reversed, i.e., the first yarn becomes the yarn at the bottom of the tear and the last yarn becomes the yarn at the top of the tear. In Figure 5.7b the histograms corresponding to the LR-values for the D_{test} and $D_{test,rev}$ are shown. The system yields very similar distributions for the reversed patterns. It seems as if the system is not fully discriminating on $D_{test,rev}$, but this is in fact not true, the highest LR-value under H_d is still smaller than the lowest LR-value under H_p. The results of the performance metrics are given in the ECE-plot in Figure 5.8 and the corresponding Cllr-values are given in Table 5.2. We see that the performance of the system on the reversed patterns is almost the same as for the original patterns. For prior odds lower than $\sim 10^{-\frac{1}{2}}$ the system even performs a little bit better on the reversed patterns. While, based on the Cllrs, the loss in accuracy is a little higher for the LRs obtained from the reversed patterns. The observation that the performance of the system is very similar on both sets of LRs could be an indication that either the transition probabilities are similar for both directions, or the model does not highly rely on the transition probabilities, but more so on the theoretical framework that it is built on.

Shorter patterns In order to evaluate the performance of our model on shorter patterns, we adjusted the loopbreaking patterns in D_{test}, such that each pattern was cut off at height 15. In this way each pattern consisted of the loop states of about 16 yarns (or in some cases 17, if the other pattern starts with a missing yarn), instead of about 36 yarns. We denote this set of shortened patterns by $D_{test,short}$. The possible vertical alignments of these patterns remained the same. The resulting $\log_{10}(LR)$-values are presented in Figure 5.7c. It is clear that the distribution of the $\log_{10}(LRs)$ both under H_p and under H_d is more compact and closer to zero. This can be expected, since the patterns in $D_{test,short}$ contain less information, thus their evidential value should be smaller. Further, when looking closer we see that the system does not fully discriminate anymore on the short patterns. In Figure 5.9 the corresponding ECE-plots are depicted. Additionally, the Cllr-values are given in Table 5.2. We find that the loss in discriminating power for the shorter patterns is almost of the same size as the total loss in accuracy for the normal patterns in D_{test}. Moreover, the total loss in accuracy is quite large in comparison, which is mostly due to the loss in calibration (the difference between ECE of the LRs and the ECE of the PAV LRs). This loss in accuracy can be explained by the number of LR-values under H_d that are larger than 1. This is much higher in the set of LRs obtained from $D_{test,short}$, compared to the set of LRs obtained from D_{test}. However, the system still performs much better than the neutral reference.

5.2.4. Observation error probabilities

For our model we have based the probabilities for observation errors on expert knowledge, since we had no data about it. To evaluate the performance of the model for different levels of these probabilities, we evaluated the model for three different levels of these probabilities.



(a) Patterns from D_{test} (b) Patterns from $D_{test,rev}$ (c) Patterns from $D_{test,short}$

Figure 5.7.: Histograms of the $\log_{10}(LR)$ values for patterns from D_{test} and $D_{test,rev}$. At the top are the density histograms and at the bottom the respective count histograms zoomed in.

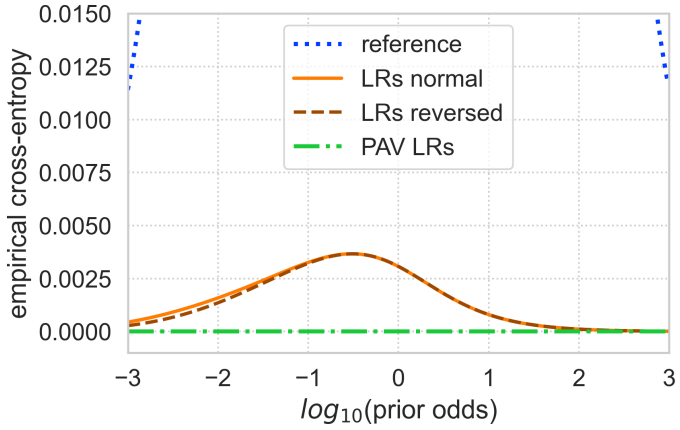
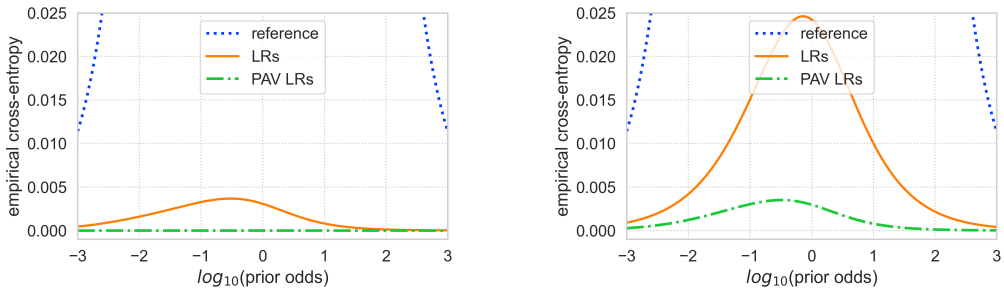


Figure 5.8.: ECE-plots for LRs obtained from D_{test} (LRs normal) and for LRs obtained from $D_{test,rev}$ (LRs reversed).



(a) Test data is D_{test}.

(b) Test data is $D_{test,short}$.

Figure 5.9.: ECE-plots for LRs obtained from D_{test} and for LRs obtained from $D_{test,short}$.

	LRs D_{test}	LRs $D_{test,rev}$	LRs $D_{test,short}$
$Cllr$	0.003065	0.003073	0.024243
$Cllr^{min}$	0	0	0.002961
$Cllr^{cal}$	0.003065	0.003073	0.021282

Table 5.2.: Cllr-values for the sets of LR-values obtained from the patterns in D_{test} $D_{test,rev}$ and $D_{test,short}$ respectively. The values are rounded down.

- *Medium P(err)* The original model with observation error probabilities as selected based on expert knowledge.
- *High P(err)* The model where the probability of an observation error is higher than in the original model.
- *Low P(err)* The model where the probability of an observation error is lower than in the original model.

The respective CPTs for these different levels of observation errors for the models are given in section F.2.2.

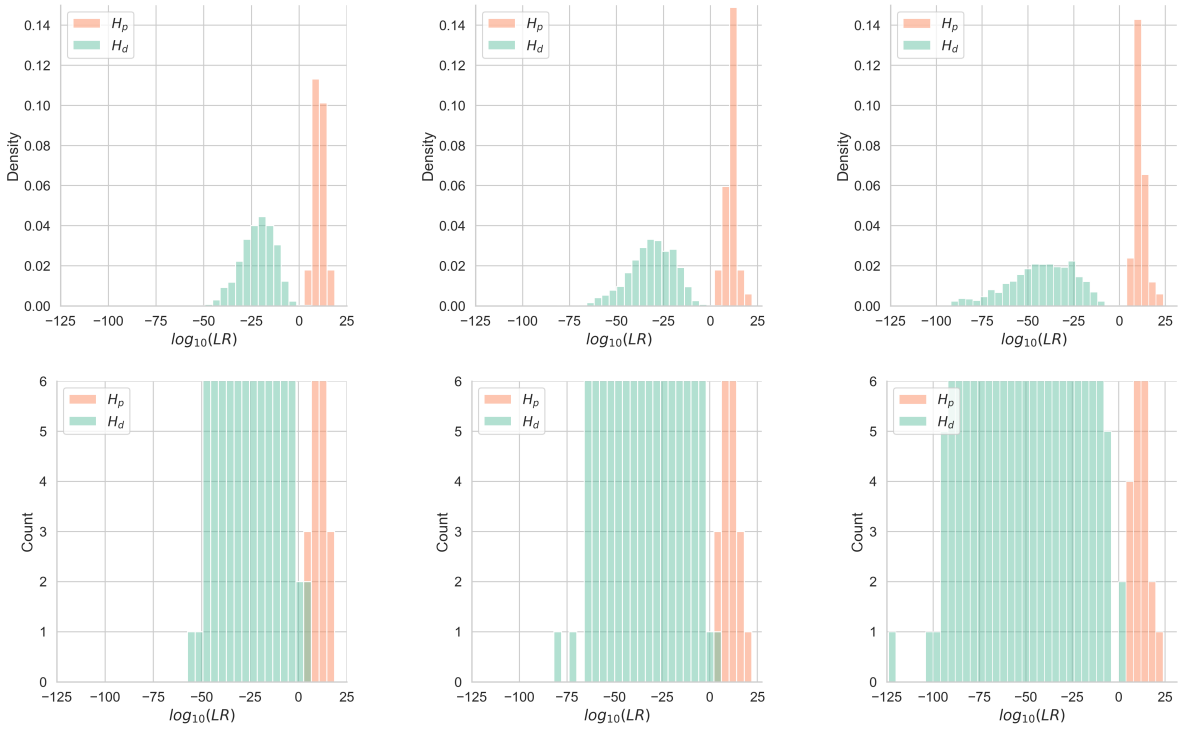
With these different levels of observation probabilities in the models, we obtain three different LR-systems. To analyse the effect of the difference in observation error probabilities on the performance, we calculate the LR-values for the patterns from D_{test} with each of the three systems. These LRs are depicted in the histograms in Figure 5.10. Notice that the LR-values under H_d become smaller for smaller values of the error probabilities. This could be expected, if two observed patterns do not match, then their evidential value is closer to one if there is more uncertainty about these observations. The distribution of the LR-values under H_p does not seem to change much. We would expect that if the error probabilities are set to be too high, then also the evidential value of two patterns under H_p would become smaller, since the probability that the observation of two loops being complementary is actually an error is so high. Further, if the error probabilities were set too low, then we would expect that some of the LR-values under H_p, would be closer to one or even less than one, because any observation errors made would have such a small probability of occurring. Therefore, it seems that the error probabilities picked are not too extreme to make the system perform significantly worse under H_p. Furthermore, we see that the three systems seem to discriminate fully between H_p and H_d. Although this cannot be said with certainty for the high and medium error probabilities

	High $\mathbb{P}$(err)	Medium $\mathbb{P}$(err)	Low $\mathbb{P}$(err)
$Cllr$	0.004573	0.003065	0.002146
$Cllr^{min}$	0	0	0
$Cllr^{cal}$	0.004573	0.003065	0.002146

Table 5.3.: Cllr-values for the LRs obtained from D_{test}, for the three different levels of observation error probabilities.

based on the histograms shown in Figure 5.10a and Figure 5.10b. However, for the medium error probabilities we have already seen that it is fully discriminating on the set D_{test}. After further investigation on the LR-values from the LR-system with the high error probabilities, we find that the highest LR given H_d is $\sim 10^{3.91}$, while the lowest LR given H_p is $\sim 10^{4.84}$. We also see that in all cases, even though the systems are fully discriminating, there are some LR-values under H_d which are larger than 1, i.e. misleading evidence. This implies that there is some loss in accuracy in all of these systems. In Figure 5.11 the corresponding Tippett plots are shown, which display the cumulative distributions of the LRs. Here we can make the same observations, except that the misleading evidence is such a small fraction of the test data under H_d, that it is not visible. The ECE-plots for the three LR-systems are presented in Figure 5.12. First we observe that in all cases the PAV LRs is 0 for all values of log_{10}(prior odds). This is the consequence of the fact that all the systems are fully discriminating. We see that the system LRs are in all cases a bit shifted to the left.We also note that this shift becomes smaller when the error probabilities become smaller and subsequently the distance between the PAV LRs and the system LRs becomes smaller. This implies that the loss in accuracy is smaller for lower error probabilities. This is also visible in Table 5.3, which denotes the respective Cllr-values for the three systems.

So we find that the system performs better for our test data if the error probabilities are lower. This could imply that these error probabilities should actually be set to these lower probabilities or even more lower. However, these values might differ greatly between different forensic examiners. Therefore, it is advised to perform experiments to better determine these error probabilities.



(a) High error probabilities (b) Medium error probabilities (c) Low error probabilities

Figure 5.10.: Histograms of the $\log_{10}(LR)$-values obtained from D_{test}, for different levels of observation error probabilities. At the top are the density histograms and at the bottom the respective count histograms zoomed in.

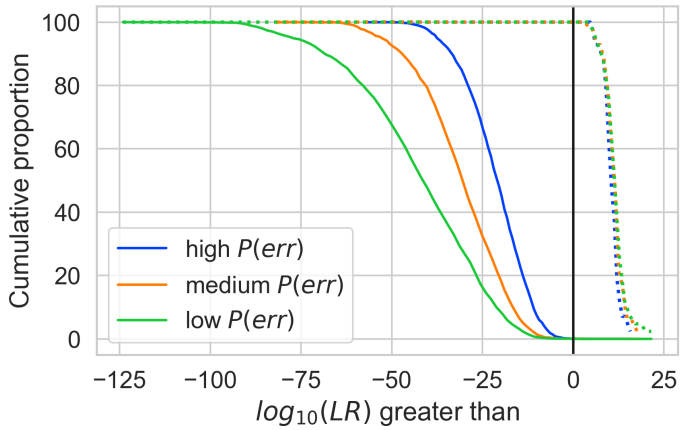


Figure 5.11.: Tippett plots for the LRs obtained from D_{test}, for high, medium and low observation error probabilities. The solid lines represent the LR-values under H_d and the dotted lines represent the LR-values under H_p.

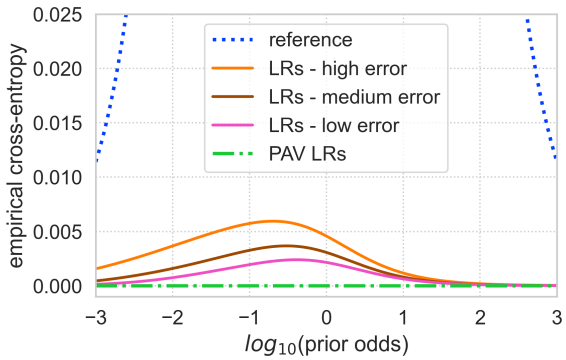


Figure 5.12.: ECE-plots for the LRs obtained from D_{test}, for the three different levels of observation error probabilities.

6. Discussion

In this thesis we proposed an LR-system to evaluate the evidential value of the new, additional method for duct tape comparisons, using loopbreaking patterns. The LR-system consists of three dynamical Bayesian networks, which provide estimates for the probabilities under H_p and under H_d for the X-side and the Y-side separately. These can be combined to obtain the LR. These probability estimates are based on a combination of theoretical properties, expert knowledge and data.

Based on the performance evaluation of the trained LR-system on the unseen test data as discussed in section 5.2, we find that the LR-system is very accurate and produces very strong evidence. The performance of the system is even better for the simulated data set S_{test}. The LR-system also seems robust under variations in the observation error probabilities and reversed tearing directions. However, for future work it would be even more interesting to also evaluate the performance of the LR-system on loopbreaking patterns with unknown tearing directions, by evaluating their LRs using Equation 4.2, since in most practical cases the tearing direction is unknown. The performance of the LR-system on shortened patterns is significantly reduced, but it seems still useful. It would be interesting to evaluate what the minimal length of a pattern should be to still be applicable for this system.

However, we need to be careful drawing conclusions based on the presented validation results, since we randomly split the data from the available dataset into the training and test set, these sets were not independent from one another. Moreover, we were not able to reach a validation decision due to the lack of a validation data set and validation criteria. The experimental conditions also were the same for all persons and are not representative for real forensic cases. This means that the system was probably overfit on these persons and conditions. However, the overfitting might be limited, since only part of the system is based on the data, and the theoretical framework seems to have a large impact on the system, but this needs to be evaluated further. Moreover, the evaluation of the validity of LR-systems based on the described performance characteristics (accuracy, discriminating power and calibration) is criticized by Meester and Slooten in [13]. They remark that these characteristics do not necessarily characterize a valid method. A valid method would produce 'true' LR-values. In some cases it is known that the 'true' LR actually supports the opposite of the ground truth and this should not make the method less valid. However, we could see these characteristics as the characteristics of a useful method, which might be what we are actually aiming for.

Additionally, the main assumption of our LR-system, that the loopbreaking patterns comply to the Markov property, has not been verified. This could be done by evaluating the results of a system that assumes the second order Markov property (i.e., that each loop state depends only on the loop states of the former two loops) and a system that assumes independence between all the loop states within a loopbreaking pattern, and compare their performance with the current system. However, a system based on the second order Markov property, becomes even larger and so does its parameter space. It does not seem feasible to make such a system with the current dataset. There are many simplifications in the models, e.g., dismissing cross-dependencies and neglecting the information of missing yarns. We deem these simplifications necessary, because of the small amount of data available. It is also unlikely that there will be massive amounts of

data available in the near future, since analyzing the loopbreaking patterns is a time-consuming process. Including more dependencies would lead to a substantial increase in the number of parameters. For example, in the two-side model for loops only, if we would include the cross-dependency for the current loop on the X-side on the former loop on the Y-side, then the number of parameters for the node representing the current loop on the X-side increases from $2 \cdot 3 \cdot 3 = 18$ to $2 \cdot 3 \cdot 3 \cdot 3 = 72$.

Our most important recommendation for future work is that the proposed LR-system should be validated with data that is representative for real forensic cases. Such a validation data set could be obtained, either by evaluating the system on old forensic cases, where the ground truth seems obvious or by performing additional experiments where the conditions are close to real forensic cases. The benefits of executing the former option, is that the LRs given by the forensic experts can be compared with the LRs given by this system, and the conditions are exactly as in a forensic case. Also the time investment is limited, since no additional loopbreaking patterns have to be established. The benefits of the latter is that the ground truths are certain and that the amount of data is theoretically limitless, as opposed to the data from old cases. For both options it is possible to add simulated data based on the retrieved loopbreaking patterns. However, the retrieved patterns should then still be a good representation for most forensic cases. Notice that the loopbreaking patterns in the current dataset were all obtained by *tearing* the duct tape. In forensic cases it is also possible that the tape is *pulled*. It is not clear what the effect of this is on the loopbreaking patterns, but if the model is to be used for those cases as well, these types of loopbreaking patterns should be included in the validation set. If this makes the model invalid, then even more data could be obtained to train the model also on these types of patterns. Further, before applying this model, it is important to remark that the observation error probabilities were obtained for specific forensic examiners. These values can vary highly between forensic examiners, therefore these should be adjusted before using this system in practice.

Bibliography

[1] C. Aitken and F. Taroni. *Statistics and the evaluation of evidence for forensic scientists*. 2nd ed. John Wiley & Sons, 2004. ISBN: 0-470-84367-5.

[2] R. T. Bailey. "Estimation from zero-failure data". In: *Risk Analysis* 17.3 (1997), pp. 375–380.

[3] M. J. Bradley et al. "A Validation Study for Duct Tape End Matches". In: *Journal of Forensic Sciences* 51.3 (2006), pp. 504–508. DOI: 10.1111/j.1556-4029.2006.00106.x.

[4] E. Brooks et al. "Forensic physical fits in the trace evidence discipline: A review". In: *Forensic Science International* 313 (2020), p. 110349. DOI: 10.1016/j.forsciint.2020.110349.

[5] N. Brümmer and J. du Preez. "Application-independent evaluation of speaker detection". In: *Computer Speech & Language* 20.2 (2006), pp. 230–275. DOI: 10.1016/j.csl.2005.08.001.

[6] T. Gneiting and A. E. Raftery. "Strictly proper scoring rules, prediction, and estimation". In: *Journal of the American Statistical Association* 102.477 (2007), pp. 359–378.

[7] M. Holmes-Cerfon. "Markov Chains (I)". In: *Applied Stochastic Analysis*. Lecture notes. 2019. Chap. 2. URL: https://cims.nyu.edu/~holmes/teaching/asa19/handout_Lecture2_2019.pdf.

[8] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant. *Applied logistic regression*. 3rd ed. John Wiley & Sons, 2013. ISBN: 9781118548356.

[9] *HUGIN EXPERT*. Bayesian Network software. Version 8.8. URL: www.hugin.com.

[10] K. LaPorte and R. Weimer. "Evaluation of Duct Tape Physical Characteristics: Part I - Within-Roll Variability". In: *Journal of the American Society of Trace Evidence Examiners* 7.1 (2017), pp. 15–34.

[11] A. J. Leegwater et al. "Performance study of a score-based likelihood ratio system for forensic fingermark comparison". In: *Journal of Forensic Sciences* 62.3 (2017), pp. 626–640. DOI: 10.1111/1556-4029.13339.

[12] K. R. McCabe et al. "A Quantitative Analysis of Torn and Cut Duct Tape Physical End Matching". In: *Journal of Forensic Sciences* 58.s1 (2013), S34–S42. DOI: 10.1111/1556-4029.12013.

[13] R. Meester and K. Slooten. *Probability and Forensic Evidence: Theory, Philosophy, and Applications*. Cambridge University Press, 2021.

[14] A. H. Mehltretter and M. J. Bradley. "Forensic analysis and discrimination of duct tapes". In: *Journal of the American Society of Trace Evidence Examiners* 3.1 (2012), pp. 2–20.

[15] A. H. Mehltretter et al. "Intra-roll and intra-jumbo roll variation of duct tapes". In: *Journal of the American Society of Trace Evidence Examiners* 6.1 (2015), pp. 21–41.

[16] D. Meuwly, D. Ramos, and R. Haraksim. "A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation". In: *Forensic Science International* 276 (2017), pp. 142–153. DOI: 10.1016/j.forsciint.2016.03.048.

[17] T. M. Mitchell. "Generative and discriminative classifiers: Naive Bayes and logistic regression". In: *Machine learning*. Draft of October 1, 2020. Additional chapter for second edition of book. Chap. 3. URL: http://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf.

[18] G. S. Morrison. "Measuring the validity and reliability of forensic likelihood-ratio systems". In: *Science & Justice* 51.3 (2011), pp. 91–98. DOI: 10.1016/j.scijus.2011.03.002.

[19] K. P. Murphy. "Dynamic bayesian networks: representation, inference and learning". PhD thesis. University of California, 2002.

[20] M. Prusinowski, E. Brooks, and T. Trejos. "Development and validation of a systematic approach for the quantitative assessment of the quality of duct tape physical fits". In: *Forensic Science International* 307 (2020), p. 110103. DOI: 10.1016/j.forsciint.2019.110103.

[21] D. Ramos and J. Gonzalez-Rodriguez. "Reliable support: measuring calibration of likelihood ratios". In: *Forensic Science International* 230.1-3 (2013), pp. 156–169.

[22] W. Ristenpart, F. Tulleners, and A. Alfter. *Quantitative Algorithm for the Digital Comparison of Torn and Cut Duct Tape*. University of California, 2017.

[23] S. M. Ross. "Markov Chains". In: *Introduction to Probability Models*. 12th ed. Academic Press, 2019. Chap. 4. DOI: 10.1016/c2017-0-01324-1.

[24] J. M. Smith. "Forensic Examination of Pressure Sensitive Tape". In: *Forensic Analysis on the Cutting Edge*. John Wiley & Sons, 2007. Chap. 12, pp. 291–332. DOI: 10.1002/9780470166932.ch12.

[25] A. van Someren. "De basis voor het valideren van het soucheonderzoek aan duct tape met lusverbindingen". Unpublished. Bachelor Thesis. Hogeschool van Amsterdam, Netherlands Forensic Institute, 2018.

[26] F. A. Tulleners and J. V. Braun. "The statistical evaluation of torn and cut duct tape physical end matching". In: *Final report to the National Institute of Justice, grant* 2009-DN (2011).

[27] *Uitspraak: ECLI:NL:RBLIM:2015:1588*. Rechtbank Limburg, Feb. 2015. URL: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBLIM:2015:1588 (visited on Mar. 3, 2021).

[28] *Validity (statistics)*. Wikipedia. URL: https://en.wikipedia.org/wiki/Validity_(statistics) (visited on May 4, 2021).

[29] D. A. van Leeuwen and N. Brümmer. "The distribution of calibrated likelihood-ratios in speaker recognition". In: *Proceedings of Interspeech*. Lyon, France, 2013.

[30] R. Wieten et al. "The interpretation of traces found on adhesive tapes". In: *Law, probability and risk* 14.4 (2015), pp. 305–322.

[31] S. M. Willis et al. *ENFSI guideline for evaluative reporting in forensic science*. Version 3.0. European Network of Forensic Science Institutes, 2015. URL: https://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf.

[32] G. Zadora et al. "Performance of likelihood ratio methods". In: *Statistical analysis in forensic science: evidential value of multivariate physicochemical data*. 1st ed. John Wiley & Sons, 2014. Chap. 6.

[33] B. Zadrozny and C. Elkan. "Transforming classifier scores into accurate multiclass probability estimates". In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (2002).

Appendix A.

Tearing directions

A.1. Notation of tearing directions

For this research we are interested in the order in which the yarns have been broken. For these tearing directions of the patterns for the X-side and the Y-side, $\mathcal{D}_X$ and $\mathcal{D}_Y$, we consider two possibilities:

- T the tearing started at the top and ended at the bottom;
- B the tearing started at the bottom and ended at the top.

Here the top and bottom can be defined arbitrarily, but we will define them as the top and bottom of the tape placed such that the X-side is on the left and the Y-side on the right, and the polymeric backing is to the front, see Figure A.1. Let us denote with $E_{X,T}$ and $E_{X,B}$ the pattern E_X such that the tearing direction is T and B respectively. For E_Y we denote this similarly by $E_{Y,T}$ and $E_{Y,B}$.

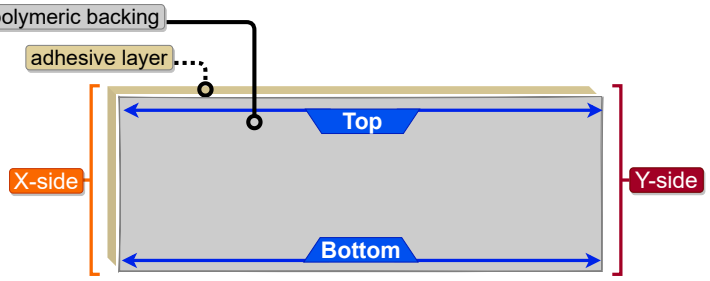


Figure A.1.: Schematic figure presenting the location of top and bottom on a piece of duct tape. The X-side is on the left, the Y-side on the right, the adhesive layer at the back and the polymeric backing is at the front.

Appendix B.

Maximum likelihood estimator for logistic regression

Let $(y_i, \tilde{x}_i)$ denote the values of the categorical output and corresponding categorical input variables for $1 \leq i \leq N$, such that $y_i \in \{0, 1, 2, \dots K\}$ and $\tilde{x}_i \in \{0, 1, 2, \dots, M\}$ are independent categories. Let us apply the method of multinomial logistic regression. As explained in [8], the input variables need to have some numerical interpretation, otherwise *dummy variables* should be used to denote the input variables as a vector. In this case we can denote the different categories using $M - 1$ dummy variables, x_m for $1 \leq m \leq M$, such that the category $l \in \{0, 1, 2, \dots M\}$ is represented as $x_m = 0$ for all $m \neq l$ and $x_l = 1$. Additionally, we set $x_{0,i} = 1$ for all i, this allows for the intercept. Accordingly, for any category l, we can denote the input variable $\tilde{x}_i = l$ as a vector $\boldsymbol{x_i} = \boldsymbol{l} := (1, e_1, e_2, \dots, e_M)^\top$ such that $e_l = 1$ and $e_m = 0$ for all $m \neq l$.

Then we define the logistic function as

$$\pi_j(\boldsymbol{x}) := \mathbb{P}(Y = j|\boldsymbol{x}) = \frac{e^{g_j(\boldsymbol{x})}}{\sum_{k=0}^{K} e^{g_k(\boldsymbol{x})}},$$

with $g_0(\boldsymbol{x}) := 0$ and for all $1 \leq k \leq K$

$$g_k(\boldsymbol{x}) := \boldsymbol{x}^\top \boldsymbol{\beta_k}$$

denote the logits for some unknown parameters $\boldsymbol{\beta_k}$.

Proposition. *Let $(y_i, \tilde{x}_i)$ denote the values of the categorical output and corresponding categorical input variables for $1 \leq i \leq N$, such that $y_i \in \{0, 1, 2, \dots K\}$ and $\tilde{x}_i \in \{0, 1, 2, \dots, M\}$ are independent categories. Let us denote with N_{jl} the number of datapoints $(y_i, \tilde{x}_i)$ such that $y_i = j$ and $\tilde{x}_i = l$, for any $j \in \{0, 1, \dots K\}$ and $l \in \{0, 1, \dots M\}$. Let $N_{\cdot l} := \sum_{j=0}^{K} N_{jl}$ for any $l \in \{0, 1, 2, \dots M\}$ and $N_{j\cdot} := \sum_{l=0}^{M} N_{jl}$ for any $j \in \{0, 1, 2, \dots K\}$.*

Then it holds for all $m \in \{0, 1, \dots, M\}$ and all $j \in \{0, 1, \dots, K\}$ that

$$\pi_j(\boldsymbol{x} = \boldsymbol{m}, \hat{\boldsymbol{\beta}}_{MLE}) = \frac{N_{jm}}{N_{\cdot m}}, \tag{B.1}$$

where $\hat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}$ denotes the maximum likelihood estimator of the parameters $\boldsymbol{\beta}$.

Proof. First we introduce some notation for the categorical output variable, as suggested in [8]. Let for any $0 \leq i \leq N$ and any category $j \in \{0, 1, 2, \dots, K\}$ $y_i = j$, we set $y_{ji} = 1$ and $y_{ki} = 0$ for all $k \neq j$. In this way we get that $\sum_{k=0}^{K} y_{ki} = 1$ for all i.

The maximum likelihood estimates (MLE) of the β_k are found at the maximum of the log-likehood function. Following [8], we get that the likelihood function is given by

$$\ell(\boldsymbol{\beta}) = \prod_{i=1}^{N} \prod_{j=0}^{K} \pi_j(\boldsymbol{x_i})^{y_{ji}},$$

and taking the logarithm, provides us with the log-likelihood function

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{N} -\log\left(\sum_{j=0}^{K} e^{g_j(\boldsymbol{x_i})}\right) + \sum_{j=1}^{K} y_{ji} g_j(\boldsymbol{x_i}),$$

using that $\sum_{j=0}^{K} y_{ji} = 1$ for all i. The $\hat{\boldsymbol{\beta}}_{MLE}$ can be found by taking the derivatives of the log-likelihood function with respect to β_{jm} for all $1 \leq j \leq K$ and $0 \leq m \leq M$ and setting these to zero. We find that

$$\frac{\partial \mathcal{L}(\boldsymbol{\beta})}{\partial \beta_{jm}} = \sum_{i=1}^{N} x_{mi} \cdot (y_{ji} - \pi_j(\boldsymbol{x}_i)). \tag{B.2}$$

We need to solve $\frac{\partial \mathcal{L}(\hat{\boldsymbol{\beta}}_{MLE})}{\partial \beta_{jm}} = 0$ for all j, m. We will address the cases that $m = 0$ and $m \neq 0$ separately.

First, let $m \in \{1, 2, \ldots, M\}$ and $j \in \{0, 1, \ldots, K\}$. Note that $x_{mi} = 0$ for all i such that $\boldsymbol{x}_i \neq \boldsymbol{m}$, while $x_{mi} = 1$ for all i such that $\boldsymbol{x}_i = \boldsymbol{m}$. Applying this to Equation B.2, we find

$$\frac{\partial \mathcal{L}(\boldsymbol{\beta})}{\partial \beta_{jm}} = \sum_{i:\boldsymbol{x}_i = \boldsymbol{m}} 1 \cdot (y_{ji} - \pi_j(\boldsymbol{x}_i)).$$

We can group the data based on their output value, this yields

$$\frac{\partial \mathcal{L}(\boldsymbol{\beta})}{\partial \beta_{jm}} = \sum_{k=0}^{K} N_{km} \cdot (\boldsymbol{y_{kj}} - \pi_j(\boldsymbol{x} = \boldsymbol{m})),$$

where $\boldsymbol{y_{kj}}$ represents the jth element of the vector representing $y = k$, i.e., $\boldsymbol{y_{kj}} = 1$ if $k = j$ and otherwise $\boldsymbol{y_{kj}} = 0$. Thus it follows that

$$\frac{\partial \mathcal{L}(\boldsymbol{\beta})}{\partial \beta_{jm}} = N_{jm} \cdot (1 - \pi_j(\boldsymbol{x} = \boldsymbol{m})) - \pi_j(\boldsymbol{x} = \boldsymbol{m}) \cdot \sum_{k \neq j} N_{km}.$$

Setting this to zero and further rewriting yields

$$\begin{aligned} N_{jm} - \pi_j(\boldsymbol{x} = \boldsymbol{m}) \cdot N_{jm} - \pi_j(\boldsymbol{m}) \cdot \sum_{k \neq j} N_{km} &= 0 \\ \pi_j(\boldsymbol{x} = \boldsymbol{m}) \cdot \sum_{k=0}^{K} N_{km} &= N_{jm} \\ \pi_j(\boldsymbol{x} = \boldsymbol{m}) &= \frac{N_{jm}}{\sum_{k=0}^{K} N_{km}} = \frac{N_{jm}}{N_{\cdot m}} \end{aligned}$$

So, in the case that $m \neq 0$ Equation B.1 holds.

Now, suppose $m = 0$ and $j \in \{0, 1, \ldots, K\}$. By definition it holds that $x_{0i} = 1$ for all i. Thus Equation B.2 comes down to

$$\frac{\partial \mathcal{L}(\boldsymbol{\beta})}{\partial \beta_{jm}} = \sum_{i=1}^{N} y_{ji} - \pi_j(\boldsymbol{x_i})).$$

Similarly as above, we can group on the different combinations of categories for the input and output, we get

$$\begin{aligned}\frac{\partial \mathcal{L}(\boldsymbol{\beta})}{\partial \beta_{jm}} &= \sum_{l=0}^{M} N_{jl} \cdot (1 - \pi_j(\boldsymbol{x} = \boldsymbol{l})) - \sum_{k \neq j} \sum_{l=0}^{M} N_{kl} \cdot \pi_j(\boldsymbol{x} = \boldsymbol{l}) \\ &= \sum_{l=0}^{M} N_{jl} - \sum_{k=0}^{K} \sum_{l=0}^{M} N_{kl} \cdot \pi_j(\boldsymbol{x} = \boldsymbol{l}).\end{aligned}$$

Setting this to zero, changing the order of summation and using that $\pi_j(\boldsymbol{x} = \boldsymbol{l}) = \frac{N_{jl}}{N_{\cdot l}}$ holds for all $l > 0$, gives

$$\begin{aligned} N_{j\cdot} - \sum_{l=0}^{M} \sum_{k=0}^{K} N_{kl} \cdot \pi_j(\boldsymbol{x} = \boldsymbol{l}) &= 0 \\ \sum_{l=0}^{M} N_{\cdot l} \cdot \pi_j(\boldsymbol{x} = \boldsymbol{l}) &= N_{j\cdot} \\ N_{\cdot 0} \cdot \pi_j(\boldsymbol{x} = \mathbf{0}) + \sum_{l=1}^{M} N_{\cdot l} \cdot \frac{N_{jl}}{N_{\cdot l}} &= N_{j\cdot} \\ N_{\cdot 0} \cdot \pi_j(\boldsymbol{x} = \mathbf{0}) + N_{j\cdot} - N_{j0} &= N_{j\cdot} \\ \pi_j(\boldsymbol{x} = \mathbf{0}) &= \frac{N_{j0}}{N_{\cdot 0}}. \end{aligned}$$

Hence, Equation B.1 also holds for $m = 0$. □

Appendix C.

Likelihood ratio for unknown tearing directions

Suppose both tearing directions are unknown. For the probability under H_p we know that the tearing directions must be the same, so

$$\begin{aligned}\mathbb{P}(E_X, E_Y|H_p) &= \sum_{d_x,d_y\in D} \mathbb{P}(E_X, E_Y|\mathcal{D}_X = d_x, \mathcal{D}_Y = d_y, H_p)\mathbb{P}(\mathcal{D}_X = d_x, \mathcal{D}_Y = d_y|H_p)\\ &= \sum_{d\in D} \mathbb{P}(E_X, E_Y|\mathcal{D}_X = d, \mathcal{D}_Y = d, H_p)\mathbb{P}(\mathcal{D}_X = d, \mathcal{D}_Y = d|H_p).\end{aligned}$$

Here D denotes the set of possible tearing directions, $\{T, B\}$. For the probability under H_d we get

$$\begin{aligned}&\mathbb{P}(E_X, E_Y|H_d)\\ &\quad= \sum_{d_x,d_y\in D} \mathbb{P}(E_X, E_Y|\mathcal{D}_X = d_x, \mathcal{D}_Y = d_y, H_d)\mathbb{P}(\mathcal{D}_X = d_x, \mathcal{D}_Y = d_y|H_d)\\ &\quad= \sum_{d_x,d_y\in D} \mathbb{P}(E_X|\mathcal{D}_X = d_x, H_d)\mathbb{P}(\mathcal{D}_X = d_x|H_d)\mathbb{P}(E_Y|\mathcal{D}_Y = d_y, H_d)\mathbb{P}(\mathcal{D}_Y = d_y|H_d).\end{aligned}$$

Here we use again the assumption that the two patterns are independent given H_d and also that their tearing directions are independent given H_d.

Generally we assume that there is no information at all about the tearing directions, hence we assume that their distributions are uniform. However, this can be adjusted for special cases. We have

$$\begin{aligned}\mathbb{P}(E_X, E_Y|H_p) &= \frac{1}{2}\mathbb{P}(E_X, E_Y|\mathcal{D}_X = T, \mathcal{D}_Y = T, H_p) + \frac{1}{2}\mathbb{P}(E_X, E_Y|\mathcal{D}_X = B, \mathcal{D}_Y = B, H_p)\\ &= \frac{1}{2}\big(\mathbb{P}(E_{X,T}, E_{Y,T}|H_p) + \mathbb{P}(E_{X,B}, E_{Y,B}|H_p)\big)\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}(E_X, E_Y|H_d) &= \frac{1}{4}\sum_{d_x,d_y\in D} \mathbb{P}(E_X|\mathcal{D}_X = d_x, H_d)\mathbb{P}(E_Y|\mathcal{D}_Y = d_y, H_d)\\ &= \frac{1}{4}\Big(\mathbb{P}(E_{X,T}|H_d)\mathbb{P}(E_{Y,T}|H_d) + \mathbb{P}(E_{X,T}|H_d)\mathbb{P}(E_{Y,B}|H_d)\\ &\qquad + \mathbb{P}(E_{X,B}|H_d)\mathbb{P}(E_{Y,T}|H_d) + \mathbb{P}(E_{X,B}|H_d)\mathbb{P}(E_{Y,B}|H_d)\Big)\end{aligned}$$

Thus the LR in terms of the patterns with different directions, yields

$$\begin{aligned}&LR(E_X, E_Y|\mathcal{D}_X \in D, \mathcal{D}_Y \in D)\\ &\quad= 2\cdot\frac{\mathbb{P}(E_{X,T}, E_{Y,T}|H_p) + \mathbb{P}(E_{X,B}, E_{Y,B}|H_p)}{\mathbb{P}(E_{X,T}, E_{Y,T}|H_d) + \mathbb{P}(E_{X,T}, E_{Y,B}|H_d) + \mathbb{P}(E_{X,B}, E_{Y,T}|H_d) + \mathbb{P}(E_{X,B}, E_{Y,B}|H_d)},\end{aligned}$$

where $\mathbb{P}(E_{X,d_x}, E_{Y,d_y}|H_d) = \mathbb{P}(E_{X,d_X}|H_d)\mathbb{P}(E_{Y,d_y}|H_d)$.

This same methodology can be applied when the tearing direction of only one side is known. In that case there is only one possible direction for the other side under H_p and two under H_d.

Appendix D.

Likelihood ratio for general comparison

Now for the case of the second proposition: the two pieces of duct tape used to be attached to one another. For this we will setup distinct definitions for the evidence, hypotheses and background information.

We now consider the following two pieces of evidence:

E_1 : the loop breaking patterns of the X- and Y-side of a torn piece of duct tape, piece 1;

E_2 : the loop breaking pattern of the X- and Y-side of another torn piece of duct tape, piece 2.

Furthermore, we will discriminate between the loop breaking patterns of the X- and Y-side of the two pieces of duct tape by denoting them as E_{1X}, E_{1Y}, E_{2X} and E_{2Y} respectively.

The hypotheses we consider for this case are

H'_p : Duct tape pieces 1 and 2 were directly connected as one piece of duct tape, before being torn.

H'_d : Duct tape pieces 1 and 2 have never been directly connected. Either they came from the same roll and there was a piece of tape in between them, or they came from different rolls of duct tape.

Both these hypotheses consists of two parts, there are two ways for two pieces of duct tape to be connected or not. We define

H'_{pXY} : The X-side of piece 1 used to be directly connected to the Y-side of piece 2.

H'_{pYX} : The Y-side of piece 1 used to be directly connected to the X-side of piece 2.

We see that $H'_p = H'_{pXY} \cup H'_{pYX}$. We assume that it is not possible for the pieces of tape to form a connected loop, so we have $H'_{pXY} \cap H'_{pYX} = \emptyset$. Similar for H'_d, we define

H'_{dXY} : The X-side of piece 1 was never directly connected to the Y-side of piece 2. Either they used to be directly connected on the opposite sides of both pieces, either they came from the same roll and there was one or more pieces of tape in between them, or they came from different rolls of duct tape.

H'_{dYX} : The Y-side of piece 1 was never directly connected to the X-side of piece 2. Either they used to be directly connected on the opposite sides of both pieces, either they came from the same roll and there was one or more pieces of tape in between them, or they came from different rolls of duct tape.

Notice that H'_{dXY} and H'_{dYX} are equal to the complements of H'_{pXY} and H'_{pYX} respectively. Further, we have that $H'_d = H'_{dXY} \cap H'_{dYX}$.

For the background information we make the same assumptions as before. There is one additional requirement, which is that both sides of the pieces of duct tape are torn, i.e.:

> I' : The 2 pieces of duct tape have the same physical and chemical appearances. The 2 pieces are both 5 cm wide, have a weft insertion scrim, a grey polymeric backing of the same structure, are torn on *both* sides and the warp yarns are vertically aligned. The positions of any observed missing warp yarns on each side.

From this we see that the likelihood ratio is given as

$$LR'(E_1, E_2) := LR_{H'_p, H'_d}(E_1, E_2|I') = \frac{\mathbb{P}(E_1, E_2|H'_p, I')}{\mathbb{P}(E_1, E_2|H'_d, I')}. \tag{D.1}$$

With some work we can rewrite this LR into two separate terms. For this we notice that due to symmetry it holds that $\mathbb{P}(H'_{pXY}|H'_p, I') = \mathbb{P}(H'_{pYX}|H'_p, I') = \frac{1}{2}$. Furthermore, we use $\mathbb{P}(E_1, E_2|H'_p, I') = \mathbb{P}(E_1, E_2|H'_{pXY}, I')\mathbb{P}(H'_{pXY}|H'_p, I') + \mathbb{P}(E_1, E_2|H'_{pYX}, I')\mathbb{P}(H'_{pYX}|H'_p, I')$ and we find

$$LR'(E_1, E_2|I') = \frac{1}{2}\frac{\mathbb{P}(E_1, E_2|H'_{pXY}, I')}{\mathbb{P}(E_1, E_2|H'_d, I')} + \frac{1}{2}\frac{\mathbb{P}(E_1, E_2|H'_{pYX}, I')}{\mathbb{P}(E_1, E_2|H'_d, I')}. \tag{D.2}$$

Each of these terms can be further rewritten such that we only need to take one side of each piece of duct tape into consideration, i.e.,

$$LR'(E_1, E_2|I') = \frac{1}{2}\frac{\mathbb{P}(E_{1X}, E_{2Y}|H'_{pXY}, I')}{\mathbb{P}(E_{1X}, E_{2Y}|H'_{dXY}, I')} + \frac{1}{2}\frac{\mathbb{P}(E_{1Y}, E_{2X}|H'_{pYX}, I')}{\mathbb{P}(E_{1Y}, E_{2X}|H'_{dYX}, I')}. \tag{D.3}$$

Here we used that $H'_{pXY} \subset H'_{dYX}$ and $H'_{pYX} \subset H'_{dXY}$ and we assumed that $(H'_{pXY} \perp\!\!\!\perp E_{1Y}, E_{2X})|H'_{dYX}$ and similarly $(H'_{pYX} \perp\!\!\!\perp E_{1X}, E_{2Y})|H'_{dXY}$. In (D.3) we recognize the LR as we defined for specific comparison, twice. The first one only considers the X-side of piece 1 and the Y-side of piece 2. The second one only considers the Y-side of piece 1 and the X-side of piece 2. Notice that

$$\mathbb{P}(E_{iX}, E_{jY}|H'_{pXY}, I') = \mathbb{P}(E_{iX}, E_{jY}|H_p, I),$$
$$\mathbb{P}(E_{iX}, E_{jY}|H'_{dXY}, I') = \mathbb{P}(E_{iX}, E_{jY}|H_d, I),$$

for $(i, j) \in \{(1, 2), (2, 1)\}$. Where H_p, H_d and I are as described above for specific comparison. Hence we see that

$$LR'(E_1, E_2|I') = \frac{1}{2}LR_{H_p,H_d}(E_{1X}, E_{2Y}|I) + \frac{1}{2}LR_{H_p,H_d}(E_{2X}, E_{1Y}|I). \tag{D.4}$$

Now that we have shown that we can express the LR for the general comparison in LRs for specific comparison, we will focus on the calculation of the LR for specific comparison. In other words we are interested in the value of

$$LR_{H_p,H_d}(E_X, E_Y) = \frac{\mathbb{P}(E_X, E_Y|H_p, I)}{\mathbb{P}(E_X|H_d, I)\mathbb{P}(E_Y|H_d, I)}. \tag{D.5}$$

To find the probabilities of these LRs we will formulate a model. From now on when we are talking about *the LR* we are referring to this last one.

Appendix E.

Likelihood ratio for patterns with missing yarns

E.1. Missing yarns at middle or bottom of pattern

Let us consider two loopbreaking patterns $(\boldsymbol{X_n})_{n=0}^m$ and $(\boldsymbol{Y_n})_{n=0}^m$ of the same length $(m+1)$, such that only the pattern for the X-side contains one missing yarn at height k, i.e., $X_k = -1$, for some $1 \leq k \leq m$. The LR for these two patterns is given by

$$LR(E_X, E_Y) = \frac{\mathbb{P}((\boldsymbol{X_n}, \boldsymbol{Y_n})_{n=0}^m | H_p)}{\mathbb{P}((\boldsymbol{X_n})_{n=0}^m | H_d)\mathbb{P}((\boldsymbol{Y_n})_{n=0}^m | H_d)}.$$

We assume that the first order Markov property applies to these patterns. The Y-side does not contain any missing yarns, so the expression for $\mathbb{P}((\boldsymbol{Y_n})_{n=0}^m | H_d)$ remains unchanged. To deal with the missing yarn in the pattern for the X-side, we just sum over all the possible states of this missing yarn. Hence we get for the probability given H_d

$$\mathbb{P}((\boldsymbol{X_n})_{n=0}^m | H_d) = \mathbb{P}((\boldsymbol{X_n})_{n=0}^{k-1} | H_d) \sum_{\boldsymbol{i} \in M} \mathbb{P}((\boldsymbol{X_n})_{n=k+1}^m | \boldsymbol{X_k} = \boldsymbol{i}, H_d)\mathbb{P}((\boldsymbol{X_k} = \boldsymbol{i} | \boldsymbol{X_{k-1}}, H_d),$$

where M is the state space of $\boldsymbol{X_n}$, which does not include the state missing. Notice that if the missing yarn is at the bottom of the pattern, i.e., $k = m$, then we just get the probability for a shorter pattern, since

$$\begin{aligned} \mathbb{P}((\boldsymbol{X_n})_{n=0}^m | H_d) &= \mathbb{P}((\boldsymbol{X_n})_{n=0}^{m-1} | H_d) \sum_{\boldsymbol{i} \in M} \mathbb{P}((\boldsymbol{X_m} = \boldsymbol{i} | \boldsymbol{X_{m-1}}, H_d) \\ &= \mathbb{P}((\boldsymbol{X_n})_{n=0}^{m-1} | H_d) \cdot 1. \end{aligned} \tag{E.1}$$

The probability of the two patterns given H_p, for any $1 \leq k \leq m$, is given by

$$\begin{aligned} &\mathbb{P}((\boldsymbol{X_n})_{n=0}^m, (\boldsymbol{Y_n})_{n=0}^m | H_p) = \mathbb{P}((\boldsymbol{X_n})_{n=0}^{k-1}, (\boldsymbol{Y_n})_{n=0}^{k-1} | H_p) \\ &\quad \cdot \sum_{\boldsymbol{i} \in M} \mathbb{P}(\boldsymbol{Y_k}, \boldsymbol{X_k} = \boldsymbol{i} | \boldsymbol{X_{k-1}}, \boldsymbol{Y_{k-1}}, H_p) \cdot \mathbb{P}((\boldsymbol{X_n})_{n=k+1}^m, (\boldsymbol{Y_n})_{n=k+1}^m | \boldsymbol{Y_k}, \boldsymbol{X_k} = \boldsymbol{i}, H_p). \end{aligned}$$

This gives us the expression for the LR. This method can be expanded to any number of missing yarns in the two patterns, that are not at the top of the patterns.

E.2. Missing yarns at top of pattern

Any missing yarns at the top of the loopbreaking patterns will change the vertical alignment of the two patterns. Furthermore, the horizontal position of each yarn is defined based on the

horizontal position of the first yarn. Therefore, the missing yarns at the top of the pattern should be evaluated with care.

Consider two loopbreaking patterns $(\boldsymbol{X_n})_{n\geq 0} = (\boldsymbol{x_n})_{n=0}^{m_x}$ and $(\boldsymbol{Y_n})_{n\geq 0} = (\boldsymbol{y_n})_{n=0}^{m_y}$, recall $\boldsymbol{X_n} = (X_n, \xi_n)$ and $\boldsymbol{Y_n} = (Y_n, \gamma_n)$, where $\boldsymbol{x_0}$ and $\boldsymbol{y_0}$ are the respective realizations of the states of the first present warp yarn. Further, $m_x + 1$ and $m_y + 1$ denote the lengths of the patterns without the missing yarns at the top. Since these patterns do not include any missing yarns at the top of the pattern, the $\boldsymbol{X_n}$ does not necessarily represent the warp yarn at height n. When comparing two sides given H_p, we want to compare the states of the loops that used to be connected to each other. Therefore, we want to incorporate the vertical alignment of these patterns. In that case we will consider the shifted sequences, represented by the random variables $\boldsymbol{\mathcal{X}_n}$ and $\boldsymbol{\mathcal{Y}_n}$. Let S_X and S_Y denote the random variables representing the starting height of the loopbreaking patterns of the X-side and Y-side respectively. Suppose $S_X = s_x$, i.e., there are s_x missing yarns above the first present yarn, then we get $(\boldsymbol{\mathcal{X}_n})_{n\geq s_x} = (\boldsymbol{x_{n-s_x}})_{n\geq s_x}^{m_x+s_x}$. We assume that the possible values of S_X and S_Y are known and contained in I. In general, a certain combination of starting heights is only contained in I if the total number of warp yarns (present and missing) on both sides are equal in that case. If that does not hold initially, it must be possible to have additional unobserved missing yarns at the bottom of the shortest pattern. This does not change the probability of finding this pattern given H_d, by Equation E.1.

One possible vertical alignment Let us first consider the case that the vertical alignment of the two patterns is known and thus the number of missing yarns at the top is known, i.e., I contains only one combination of S_X and S_Y, say $S_X = s_x$ and $S_Y = s_y$. Let us evaluate the probability given H_d. For the X-side we have

$$\begin{aligned}\mathbb{P}(E_X|S_X = s_x, H_d) &= \mathbb{P}\left((\boldsymbol{X_n})_{n\geq 0} = (\boldsymbol{x_n})_{n=0}^{m_x}|S_X = s_x, H_d\right)\\ &= \mathbb{P}\left((\boldsymbol{\mathcal{X}_n})_{n\geq s_x} = (\boldsymbol{x_{n-s_x}})_{n=s_x}^{m_x+s_x}|S_X = s_x, H_d\right).\end{aligned}$$

In our model we will use the horizontal position of the first yarn as a reference point for the following yarns. Therefore we always set its position to zero, even if this yarn is missing. However, in that case its position is unknown, and then we do not know the respective horizontal positions of the yarns that are present. It is thus not suitable to use a pattern starting with any missing yarns. Therefore we neglect the information that is provided by the missing warp yarns and we approximate the probability

$$\mathbb{P}\left((\boldsymbol{\mathcal{X}_n})_{n\geq s_x} = (\boldsymbol{x_{n-s_x}})_{n=s_x}^{m_x+s_x}|S_X = s_x, H_d\right) \approx \mathbb{P}\left((\boldsymbol{\mathcal{X}_n})_{n\geq 0} = (\boldsymbol{x_n})_{n=0}^{m_x}|S_X = 0, H_d\right). \tag{E.2}$$

We will omit the notation of the end of the sequence and that of the starting height if it equals zero. Thus we simply write

$$\mathbb{P}(E_X|S_X = s_x, H_d) \approx \mathbb{P}\left((\boldsymbol{\mathcal{X}_n})_{n\geq 0} = (\boldsymbol{x_n})_{n\geq 0}|H_d\right). \tag{E.3}$$

Note that for the Y-side the reasoning is analogous.

Let us consider $\mathbb{P}(E_X, E_Y|H_p)$. Here we do need to take both vertical alignments into account. Further, the total number of yarns (missing and present) of both patterns should be equal, otherwise the patterns do not align, i.e., $S_X + m_x = S_Y + m_y$. Let us assume w.o.l.o.g. that $S_X \geq S_Y$. First consider the simple case that $S_X = S_Y$, we get

$$\begin{aligned}&\mathbb{P}(E_X, E_Y|S_X = S_Y = s, H_p)\\ &\qquad= \mathbb{P}\left((\boldsymbol{X_n})_{n\geq 0} = (\boldsymbol{x_n})_{n\geq 0}, (\boldsymbol{Y_n})_{n\geq 0} = (\boldsymbol{y_n})_{n\geq 0}|S_X = S_Y = s, H_p\right)\\ &\qquad= \mathbb{P}\left((\boldsymbol{\mathcal{X}_n})_{n\geq s} = (\boldsymbol{x_{n-s}})_{n\geq s}, (\boldsymbol{\mathcal{Y}_n})_{n\geq s} = (\boldsymbol{y_{n-s}})_{n\geq s}|S_X = S_Y = s, H_p\right)\\ &\qquad\approx \mathbb{P}\left((\boldsymbol{\mathcal{X}_n}, \boldsymbol{\mathcal{Y}_n})_{n\geq 0} = (\boldsymbol{x_n}, \boldsymbol{y_n})_{n\geq 0}|S_X = S_Y = 0, H_p\right).\end{aligned}$$

This last approximation is based on the assumption that we can disregard the missing yarns that are missing on both sides. Using this and Equation E.3 we can now give an expression for the corresponding likelihood ratio

$$LR_{H_p,H_d}(E_X, E_Y|S_X = S_Y = s) \approx \frac{\mathbb{P}\left((\boldsymbol{\mathcal{X}_n}, \boldsymbol{\mathcal{Y}_n})_{n\geq 0} = (\boldsymbol{x_n}, \boldsymbol{y_n})_{n\geq 0}|H_p\right)}{\mathbb{P}\left((\boldsymbol{\mathcal{X}_n})_{n\geq 0} = (\boldsymbol{x_n})_{n\geq 0}|H_d\right) \cdot \mathbb{P}\left((\boldsymbol{\mathcal{Y}_n})_{n\geq 0} = (\boldsymbol{y_n})_{n\geq 0}|H_d\right)}. \tag{E.4}$$

For the case where $S_X > S_Y$, i.e., at the start of the pattern of the Y-side there are no yarns on the X-side. We can split the patterns into two parts, the first part that has only observations on the Y-side and the second part starting at the point that there are observations on both sides. We can write

$$\begin{aligned}
&\mathbb{P}(E_X, E_Y|S_X = s_x, S_Y = s_y < s_x, H_p) \\
&\quad = \mathbb{P}\left((\boldsymbol{X_n})_{n\geq 0} = (\boldsymbol{x_n})_{n\geq 0}, (\boldsymbol{Y_n})_{n\geq 0} = (\boldsymbol{y_n})_{n\geq 0}|S_X = s_x, S_Y = s_y < s_x, H_p\right) \\
&\quad = \mathbb{P}\left((\boldsymbol{\mathcal{X}_n})_{n\geq s_x} = (\boldsymbol{x_{n-s_x}})_{n\geq s_x}, (\boldsymbol{\mathcal{Y}_n})_{n\geq s_y} = (\boldsymbol{y_{n-s_y}})_{n\geq s_y}|S_X = s_x, S_Y = s_y < s_x, H_p\right) \\
&\quad = \mathbb{P}\left((\boldsymbol{\mathcal{X}_n}, \boldsymbol{\mathcal{Y}_n})_{n\geq s_x} = (\boldsymbol{x_{n-s_x}}, \boldsymbol{y_{n-s_y}})_{n\geq s_x}, (\boldsymbol{\mathcal{Y}_n})_{n=s_y}^{s_x-1} = (\boldsymbol{y_{n-s_y}})_{n=s_y}^{s_x-1}|S_X = s_x, S_Y = s_y < s_x, H_p\right) \\
&\quad = \mathbb{P}\left((\boldsymbol{\mathcal{X}_n}, \boldsymbol{\mathcal{Y}_n})_{n\geq s_x} = (\boldsymbol{x_{n-s_x}}, \boldsymbol{y_{n-s_y}})_{n\geq s_x}|\boldsymbol{\mathcal{Y}_{s_x-1}} = \boldsymbol{y_{s_x-1}}, S_X = s_x, S_Y = s_y < s_x, H_p\right) \\
&\qquad \cdot \mathbb{P}\left((\boldsymbol{\mathcal{Y}_n})_{n=s_y}^{s_x-1} = (\boldsymbol{y_{n-s_y}})_{n=s_y}^{s_x-1}|S_Y = s_y < s_x, H_p\right).
\end{aligned}$$

For this last equality we apply the assumption of the Markov property for the patterns. Note that only the states of the yarns at height s_x depend on the state of the yarn at height $s_x - 1$. We assume that the effect of the yarn at height $s_x - 1$ on the yarns at height s_x is negligible. This is based on the idea that the part(s) of the patterns that have present yarns on both sides provides the most information. In this way we get

$$\begin{aligned}
&\mathbb{P}(E_X, E_Y|S_X = s_x, S_Y = s_y < s_x, H_p) \\
&\qquad \approx \mathbb{P}\left((\boldsymbol{\mathcal{X}_n}, \boldsymbol{\mathcal{Y}_n})_{n\geq s_x} = (\boldsymbol{x_{n-s_x}}, \boldsymbol{y_{n-s_y}})_{n\geq s_x}|S_X = s_x, S_Y = s_y < s_x, H_p\right) \\
&\qquad\quad \cdot \mathbb{P}\left((\boldsymbol{\mathcal{Y}_n})_{n=s_y}^{s_x-1} = (\boldsymbol{y_{n-s_y}})_{n=s_y}^{s_x-1}|S_Y = s_y < s_x, H_p\right)
\end{aligned}$$

Notice that for the X-side the horizontal positions are all respective to the horizontal position $\xi_{s_x} = 0$, however the horizontal positions of the Y-side are defined with respect to γ_{s_y}. When the two patterns have the same starting height, then these reference points align. However that is not the case and we want to adjust for this. Let us define for any $s \geq 0$

$$\sigma_s : (y_n, \gamma_n) \mapsto (y_n, \gamma_n - \gamma_s)$$

for all $n \geq 0$. This gives us a horizontally shifted pattern. We are only concerned about the horizontal positions of the yarns respective to its former and following yarns, hence for any $s \geq 0$ we state that

$$\mathbb{P}\left((\boldsymbol{\mathcal{Y}_n})_{n\geq 0} = (\boldsymbol{y_n})_{n\geq 0}\right) = \mathbb{P}\left((\boldsymbol{\mathcal{Y}_n})_{n\geq 0} = (\sigma_s(\boldsymbol{y_n}))_{n\geq 0}\right).$$

For our model we assume that the horizontal position of the reference points are aligned. Therefore, we want the reference points to be at the same height and set both to zero. So we are interested in

$$\begin{aligned}
&\mathbb{P}(E_X, E_Y|S_X = s_x, S_Y = s_y < s_x, H_p) \\
&\qquad \approx \mathbb{P}\left((\boldsymbol{\mathcal{X}_n}, \boldsymbol{\mathcal{Y}_n})_{n\geq 0} = (\boldsymbol{x_n}, \sigma_{s_x-s_y}(\boldsymbol{y_{n+s_x-s_y}}))_{n\geq 0}|S_X = 0, S_Y = 0, H_p\right) \\
&\qquad\quad \cdot \mathbb{P}\left((\boldsymbol{\mathcal{Y}_n})_{n=0}^{s_x-s_y-1} = (\boldsymbol{y_n})_{n=0}^{s_x-s_y-1}|S_Y = 0, H_p\right).
\end{aligned}$$

We assume that the probability of finding $(\boldsymbol{Y_n})_{n=0}^{s_x-s_y-1} = (\boldsymbol{y_n})_{n=0}^{s_x-s_y-1}$ given that it used to be connected to the X-side is equal to the probability of finding it given that it used to be connected to any other piece of duct tape, since we do not have any information about the yarns opposite of these first yarns. Hence, we have

$$\mathbb{P}\left((\boldsymbol{\mathcal{Y}_n})_{n=0}^{s_x-s_y-1} = (\boldsymbol{y_n})_{n=0}^{s_x-s_y-1} | H_p\right) = \mathbb{P}\left((\boldsymbol{\mathcal{Y}_n})_{n=0}^{s_x-s_y-1} = (\boldsymbol{y_n})_{n=0}^{s_x-s_y-1} | H_d\right).$$

For these parts we can thus either use the model for two sides (given H_p) or the model for the single sides (given H_d). Now we are finally able to give a general expression for the LR

$$\begin{aligned} &LR_{H_p,H_d}(E_X, E_Y | S_X = s_x, S_Y = s_y < s_x, \hat{s} = s_x - s_y) \\ &\qquad \approx \frac{\mathbb{P}\left((\boldsymbol{\mathcal{Y}_n})_{n\geq 0} = (\boldsymbol{y_n})_{n=0}^{\hat{s}-1} | H_d\right) \cdot \mathbb{P}\left((\boldsymbol{\mathcal{X}_n}, \boldsymbol{\mathcal{Y}_n})_{n\geq 0} = (\boldsymbol{x_n}, \sigma_{\hat{s}}(\boldsymbol{y_{n+\hat{s}}}))_{n\geq 0} | H_p\right)}{\mathbb{P}\left((\boldsymbol{\mathcal{X}_n})_{n\geq 0} = (\boldsymbol{x_n})_{n\geq 0} | H_d\right) \cdot \mathbb{P}\left((\boldsymbol{\mathcal{Y}_n})_{n\geq 0} = (\boldsymbol{y_n})_{n\geq 0} | H_d\right)}. \end{aligned} \tag{E.5}$$

Multiple possible vertical alignments In some cases the vertical alignment of the two pieces of duct tape is not certain. For our data we will consider at most 2 possible alignments, i.e., either the X-side has two possible starting heights or the Y-side has two possible starting heights. In general we let the set of possible starting heights for the X-side and the Y-side respectively be denoted by $\mathcal{S}_X$ and $\mathcal{S}_Y$. Recall that two patterns can only align if the total number of warp yarns is equal. However when the starting height is changed on one side, the total number of warp yarns on that side is changed as well. This causes the total number of warp yarns on the two sides to be different. We overcome this difference by adding unobserved missing warp yarns at the bottom of the pattern with the least amount of total warp yarns. If this is impossible for some

Let us denote the set of possible combinations of the starting height by $\mathcal{S}_{XY} := \{(s_x, s_y) : s_x \in \mathcal{S}_X, s_y \in \mathcal{S}_Y\}$. Notice that

$$\mathbb{P}(E_X, E_Y | S_X = s_{x1}, S_Y = s_{y1}, H_p) = \mathbb{P}(E_X, E_Y | S_X = s_{x2}, S_Y = s_{y2}, H_p)$$

if $s_{x1} - s_{y1} = s_{x2} - s_{y2}$. This also obviously holds under H_d, by Equation E.2. So we can consider the set

$$\mathcal{S}'_{XY} := \{\varphi(s_x, s_y) : (s_x, s_y) \in \mathcal{S}_{XY}\},$$

where

$$\varphi : (s_x, s_y) \mapsto (s_x - \min\{s_x, s_y\}, s_y - \min\{s_x, s_y\}).$$

We denote the *weight* of any $(s'_x, s'_y) \in \mathcal{S}'_{XY}$ by $\omega_{\mathcal{S}_{XY}}(s'_x, s'_y)$, which is defined as

$$\omega_{\mathcal{S}_{XY}}(s'_x, s'_y) := \frac{|\{(s_x, s_y) \in \mathcal{S}_{XY} : \varphi(s_x, s_y) = (s'_x, s'_y)\}|}{|\mathcal{S}_{XY}|}.$$

For all $(s'_x, s'_y) \in \mathcal{S}'_{XY}$ it holds that either $s'_x = 0$ or $s'_y = 0$. Therefore, we can also look at the disjoint sets

$$\begin{aligned} \mathcal{S}^*_X &:= \{s'_x : (s'_x, 0) \in \mathcal{S}'_{XY}\}, \\ \mathcal{S}^*_Y &:= \{s'_y : (0, s'_Y) \in \mathcal{S}'_{XY},\ s'_y \neq 0\}. \end{aligned} \tag{E.6}$$

Note that $\mathcal{S}_X^* \cup \mathcal{S}_Y^* = \mathcal{S}'_{XY}$. We see that

$$\begin{aligned}
\mathbb{P}(E_X, E_Y|\mathcal{S}_{XY}, H_p) = &\sum_{s_x \in \mathcal{S}_X^*} \mathbb{P}\left(E_X, E_Y|S_X - S_Y = s_x, \mathcal{S}_{XY}, H_p\right) \cdot \mathbb{P}\left(S_X - S_Y = s_x|\mathcal{S}_{XY}, H_p\right) \\
&+ \sum_{s_y \in \mathcal{S}_Y^*} \mathbb{P}\left(E_X, E_Y|S_Y - S_X = s_y, \mathcal{S}_{XY}, H_p\right) \cdot \mathbb{P}\left(S_Y - S_X = s_y|\mathcal{S}_{XY}, H_p\right) \\
= &\sum_{s_x \in \mathcal{S}_X^*} \mathbb{P}\left(E_X, E_Y|S_X = s_x, \mathcal{S}_{XY}, H_p\right) \cdot \omega_{\mathcal{S}_{XY}}(s_x, 0) \\
&+ \sum_{s_y \in \mathcal{S}_Y^*} \mathbb{P}\left(E_X, E_Y|S_Y = s_y, \mathcal{S}_{XY}, H_p\right) \cdot \omega_{\mathcal{S}_{XY}}(0, s_y).
\end{aligned}$$

Here we assume that there is no other information about the starting positions and therefore we give each combination of starting positions the same probability. This can be adjusted for different situations where there is prior belief about the possible vertical alignments.

Appendix F.

Assignment of conditional probability tables

We will provide the conditional probailitiy tables for the final dynamic Bayesian networks as described in section 4.2. We will refer to the two one-side models as dbnX and dbnY for the X-side and Y-side respectively, and we will refer to the two-side model as dbnXY.

F.1. Standard conditional probability tables

Let us consider the CPTs for the standard LR-system, which is trained on the set D_{train}(see chapter 5).

F.1.1. Model induced conditional probability tables

Here we present the CPTs that are induced by the structure of the model.

The CPTs for the `horiz_o` nodes are too large to present in a table format. They are defined by

$$\mathbb{P}(\bar{\xi}_n^{(o)} = a_{o,n} | \bar{\xi}_n^{(r)} = a_{r,n}, \varepsilon_{X,H_0,n} = e_{H_0,n}, \varepsilon_{X,H,n} = e_{H,n}) = \begin{cases} 1, & \text{if } a_{o,n} = a_{r,n} - e_{H_0,n} + e_{H,n}; \\ 0, & \text{otherwise.} \end{cases}$$

Here $\bar{\xi}_n^{(o)}$ represents the node `horizX_o`, $\bar{\xi}_n^{(r)}$ represents the node `horizX_r`, $\varepsilon_{X,H_0,n}$ represents the node `hor_first_errX`and $\varepsilon_{X,H,n}$ represents the node `hor_errorX`. The CPT for `horizY_o` is defined equivalently.

Table F.1.: CPT for `T_horizX_r` and `T_horizY_r` for all of the models.

T_horiz_r	
-4	0
-3	0
-2	0
-1	0
0	1
1	0
2	0
3	0
4	0

Table F.2.: CPT for `hor_first_errX` and `hor_first_errY` for all of the models.

`T_hor_first_err`	-2	-1	0	1	2
`hor_first_err`					
-2	1	0	0	0	0
-1	0	1	0	0	0
0	0	0	1	0	0
1	0	0	0	1	0
2	0	0	0	0	1

Table F.3.: CPT for `horizX_r` in dbnXY only.

`horizY_r`	-4	-3	-2	-1	0	1	2	3	4
`horizX_r`									
-4	1	0	0	0	0	0	0	0	0
-3	0	1	0	0	0	0	0	0	0
-2	0	0	1	0	0	0	0	0	0
-1	0	0	0	1	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	1	0	0	0
2	0	0	0	0	0	0	1	0	0
3	0	0	0	0	0	0	0	1	0
4	0	0	0	0	0	0	0	0	1

Table F.4.: CPT for `loop_o` and `T_loop_o` for all models.

`loop_error`	yes			no		
`loop_r`	Closed	Complex	Open	Closed	Complex	Open
`loop_o`						
Closed	0.0	0.5	0.1	1	0	0
Complex	0.9	0.0	0.9	0	1	0
Open	0.1	0.5	0.0	0	0	1

Table F.5.: CPT for `horizX_r` and `horizY_r` for all models.

`T_horiz_r`	-4			-3			-2			-1			0			1			2			3			4		
`jump`	-1	0	1	-1	0	1	-1	0	1	-1	0	1	-1	0	1	-1	0	1	-1	0	1	-1	0	1	-1	0	1
`horiz_r`																											
-4	1	1	0.09	1	0	0	0.1	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-3	0	0	0.819	0	1	0	0.9	0	0	0.09	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-2	0	0	0.082	0	0	0.9	0	1	0	0.9	0	0	0.09	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0
-1	0	0	0.009	0	0	0.09	0	0	0.9	0	1	0	0.9	0	0	0.09	0	0	0.01	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0.01	0	0	0.09	0	0	0.9	0	1	0	0.9	0	0	0.09	0	0	0.01	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0.01	0	0	0.09	0	0	0.9	0	1	0	0.9	0	0	0.09	0	0	0.009	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0	0.09	0	0	0.9	0	1	0	0.9	0	0	0.082	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0	0.09	0	0	0.9	0	1	0	0.819	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0	0.1	0	0	1	0.09	1	1

F.1.2. Conditional probability tables based on expert knowledge

The CPTs for the nodes with observation errors `T_hor_first_err`, `hor_error`, `T_loop_error` and `loop_error` were based on expert knowledge. Their values are represented in the tables below.

Table F.6.: CPT for `loop_error` and `T_loop_error` corresponding to the medium level of observation error probabilities, for all models.

`loop_r`	Closed	Complex	Open
`loop_error`			
yes	0.005	0.1	0.005
no	0.995	0.9	0.995

Table F.7.: CPT for `hor_error` and `T_hor_first_err` corresponding to medium error probabilities for all models.

`loop_r'`	Closed			Complex			Open		
`loop_r`	Closed	Complex	Open	Closed	Complex	Open	Closed	Complex	Open
`hor_error`									
-2	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025
-1	0.475	0.475	0.475	0.475	0.475	0.475	0.475	0.475	0.475
0	99	1	99	99	3	99	99	1	99
1	0.475	0.475	0.475	0.475	0.475	0.475	0.475	0.475	0.475
2	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025

F.1.3. Conditional probability tables estimated from data

The CPTs for the nodes `T_loop_r`, `loop_r` and `jump` in all three models are fully based on the training data. The probabilities are estimated as described in section 4.3. Here we represent their values using D_{train} as the training data. The probabilities denoted in the tables below are rounded to three decimal points.

CPTs for dbnXY and dbnY The CPTs given below are the CPTs used for both dbnXY and dbnY, for the nodes `jumpY`, `T_loopY_r`, `loopY_r`, `T_loopX_r` and `loopX_r`.

Table F.8.: CPT for `jumpY` for dbnXY and dbnY. Training data is D_{train}.

`T_loopY_r`	Closed			Complex			Open		
`loopY_r`	Closed	Complex	Open	Closed	Complex	Open	Closed	Complex	Open
`jumpY`									
-1.0	0.017	0.429	0.015	0.077	0.032	0.100	0.057	0.786	0.022
0.0	0.968	0.543	0.956	0.615	0.871	0.200	0.933	0.143	0.961
1.0	0.014	0.029	0.029	0.308	0.097	0.700	0.010	0.071	0.017

Table F.9.: CPT for `loopY_r` for dbnXY and dbnY. Training data is D_{train}.

`T_loopY_r`	Closed	Complex	Open
`loopY_r`			
Closed	0.706	0.506	0.546
Complex	0.019	0.403	0.012
Open	0.276	0.091	0.442

Table F.12.: CPT for `loopX_r` for dbnXY and dbnY. Training data is D_{train}.

`loopY_r`	Closed			Complex			Open		
`T_loopX_r`	Closed	Complex	Open	Closed	Complex	Open	Closed	Complex	Open
`loopX_r`									
Closed	0	0	0	0.375	0.095	0.233	0.972	0.714	0.953
Complex	0.002	0.120	0.005	0.125	0.857	0.326	0.008	0.214	0.008
Open	0.998	0.880	0.995	0.500	0.048	0.442	0.020	0.071	0.039

Table F.10.: CPT for `T_loopY_r` for dbnXY and dbnY. Training data is D_{train}.

`T_loopY_r`	
Closed	0.650
Complex	0.026
Open	0.324

Table F.11.: CPT for `T_loopX_r` for dbnXY and dbnY. Training data is D_{train}.

`T_loopY_r`	Closed	Complex	Open
`T_loopX_r`			
Closed	0	0.231	0.958
Complex	0.007	0.423	0.009
Open	0.993	0.346	0.032

CPTs for dbnX The CPTs given below are the CPTs used for both dbnX for the nodes `jumpX`, `T_loopX_r`, `loopX_r`, `T_loopY_r` and `loopY_r`.

Table F.13.: CPT for `jumpX` for dbnX. Training data is D_{train}.

`T_loopX_r`	Closed			Complex			Open		
`loopX_r`	Closed	Complex	Open	Closed	Complex	Open	Closed	Complex	Open
`jumpX`									
-1.0	0.013	0.100	0.074	0.154	0.040	0.720	0.012	0.036	0.022
0.0	0.977	0.700	0.922	0.769	0.920	0.240	0.956	0.393	0.958
1.0	0.010	0.200	0.004	0.077	0.040	0.040	0.033	0.571	0.019

Table F.14.: CPT for `T_loopX_r` for dbnX. Training data is D_{train}.

`T_loop_r`	
Closed	0.317
Complex	0.018
Open	0.665

Table F.15.: CPT for `loopX_r` for dbnX. Training data is D_{train}.

`T_loopX_r`	Closed	Complex	Open
`loopX_r`			
Closed	0.426	0.185	0.269
Complex	0.008	0.407	0.013
Open	0.566	0.407	0.718

Table F.16.: CPT for `T_loopY_r` for dbnX. Training data is D_{train}.

`T_loopX_r`	Closed	Complex	Open
`T_loopY_r`			
Closed	0	0.236	0.971
Complex	0.019	0.600	0.014
Open	0.981	0.164	0.016

Table F.17.: CPT for `loopY_r` for dbnX. Training data is D_{train}.

`loopX_r`	Closed			Complex			Open		
`T_loopY_r`	Closed	Complex	Open	Closed	Complex	Open	Closed	Complex	Open
`loopY_r`									
Closed	0	0	0	0.357	0.053	0.111	0.981	0.741	0.966
Complex	0.018	0.300	0.013	0.536	0.895	0.222	0.008	0.241	0.008
Open	0.982	0.700	0.987	0.107	0.053	0.667	0.011	0.019	0.027

F.2. Alternative conditional probability tables

F.2.1. Training set is D_{all}

To train the model on the set D_{all}, the CPTs based on the data are replaced by the following tables.

Table F.18.: CPT for `T_loopY_r` for dbnXY and dbnY. Training data is D_{all}.

`T_loopY_r`	
Closed	0.638
Complex	0.021
Open	0.342

Table F.19.: CPT for `T_loopX_r` for dbnXY and dbnY. Training data is D_{all}.

`T_loopY_r`	Closed	Complex	Open
`T_loopX_r`			
Closed	0.000	0.261	0.961
Complex	0.006	0.391	0.008
Open	0.994	0.348	0.031

Table F.20.: CPT for `loopY_r` for dbnXY and dbnY. Training data is D_{all}.

`T_loopY_r`	Closed	Complex	Open
`loopY_r`			
Closed	0.697	0.505	0.537
Complex	0.014	0.371	0.011
Open	0.288	0.124	0.452

Table F.21.: CPT for `loopX_r` for dbnXY and dbnY. Training data is D_{all}.

`loopY_r`	Closed			Complex			Open		
`T_loopX_r`	Closed	Complex	Open	Closed	Complex	Open	Closed	Complex	Open
`loopX_r`									
Closed	0.000	0.000	0.000	0.409	0.091	0.260	0.973	0.800	0.957
Complex	0.001	0.100	0.006	0.136	0.864	0.300	0.008	0.133	0.005
Open	0.999	0.900	0.994	0.455	0.045	0.440	0.019	0.067	0.039

Table F.22.: CPT for `jumpY` for dbnXY and dbnY. Training data is D_{all}.

`T_loopY_r`	Closed			Complex			Open		
`loopY_r`	Closed	Complex	Open	Closed	Complex	Open	Closed	Complex	Open
`jumpY`									
-1.0	0.019	0.442	0.013	0.061	0.028	0.067	0.048	0.810	0.024
0.0	0.968	0.512	0.964	0.633	0.861	0.200	0.946	0.143	0.958
1.0	0.014	0.047	0.023	0.306	0.111	0.733	0.006	0.048	0.018

Table F.23.: CPT for `T_loopX_r` for dbnX. Training data is D_{all}.

`T_loopX_r`	
Closed	0.329
Complex	0.015
Open	0.655

Table F.24.: CPT for `T_loopY_r` for dbnX. Training data is D_{all}.

`T_loopX_r`	Closed	Complex	Open
`T_loopY_r`			
Closed	0.000	0.262	0.973
Complex	0.016	0.554	0.011
Open	0.984	0.185	0.016

Table F.25.: CPT for `loopX_r` for dbnX. Training data is D_{all}.

`T_loopX_r`	Closed	Complex	Open
`loopX_r`			
Closed	0.436	0.197	0.279
Complex	0.009	0.324	0.011
Open	0.555	0.479	0.710

Table F.26.: CPT for `loopY_r` for dbnX. Training data is D_{all}.

`loopX_r`	Closed			Complex			Open		
`T_loopY_r`	Closed	Complex	Open	Closed	Complex	Open	Closed	Complex	Open
`loopY_r`									
Closed	0.000	0.000	0.000	0.424	0.053	0.077	0.981	0.742	0.971
Complex	0.016	0.308	0.011	0.485	0.895	0.308	0.007	0.242	0.005
Open	0.984	0.692	0.989	0.091	0.053	0.615	0.012	0.016	0.024

Table F.27.: CPT for `jumpX` for dbnX. Training data is D_{all}.

`T_loopX_r`	Closed			Complex			Open		
`loopX_r`	Closed	Complex	Open	Closed	Complex	Open	Closed	Complex	Open
`jumpX`									
-1.0	0.014	0.071	0.065	0.235	0.038	0.676	0.011	0.029	0.023
0.0	0.974	0.714	0.931	0.706	0.923	0.297	0.958	0.353	0.959
1.0	0.012	0.214	0.004	0.059	0.038	0.027	0.032	0.618	0.018

F.2.2. Different observation error probabilities

We consider low, medium and high observation error probabilities. The CPTs for `loop_error` and `hor_error` corresponding to the medium level are given above. These tables can be replaced by the following tables for low or high observation error probabilities.

Table F.28.: CPT for `hor_error` and `T_hor_first_err` corresponding to low error probabilities for all models.

`loop_r'`	Closed			Complex			Open		
`loop_r`	Closed	Complex	Open	Closed	Complex	Open	Closed	Complex	Open
`hor_error`									
-2	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
-1	0.495	0.495	0.495	0.495	0.495	0.495	0.495	0.495	0.495
0	999	1	999	999	3	999	999	1	999
1	0.495	0.495	0.495	0.495	0.495	0.495	0.495	0.495	0.495
2	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005

Table F.29.: CPT for `loop_error` and `T_loop_error` corresponding to low error probabilities for all models.

`loop_r`	Closed	Complex	Open
`loop_error`			
yes	0.001	0.01	0.001
no	0.999	0.99	0.999

Table F.30.: CPT for `hor_error` and `T_hor_first_err` corresponding to high error probabilities for all models.

`loop_r'`	Closed			Complex			Open		
loop_r	Closed	Complex	Open	Closed	Complex	Open	Closed	Complex	Open
`hor_error`									
-2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
-1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0	9	1.5	9	5.667	4	5.667	9	1.500	9
1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Table F.31.: CPT for `loop_error` and `T_loop_error` corresponding to high error probabilities for all models.

`loop_r`	Closed	Complex	Open
`loop_error`			
yes	0.05	0.2	0.05
no	0.95	0.8	0.95