

Estimating the fibre length distribution in fibre reinforced polymers

Jan Erik Niedermeyer

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern zur
Verleihung des akademischen Grades Doktor der Naturwissenschaften (Doctor
rerum naturalium, Dr. rer. nat.) genehmigte Dissertation.

D 386

Erstgutachterin:	Claudia Redenbach Technische Universität Kaiserslautern
Zweitgutachterin:	Aila Särkkä Chalmers University of Technology and University of Gothenburg
Datum der Disputation:	23.09.2019

Acknowledgements

One must imagine Sisyphus happy.

-Albert Camus

Even though I understand at least some of the clever meanings and implications of this quote, I would like to add something. Sisyphus would have been even happier, if he was sometimes allowed to push a new stone. I, for one, am happy that I am done with this stone.

This work was funded by the BMBF project ANiS and TU Kaiserslautern.

I am grateful to Markus Kronenberger from the Fraunhofer ITWM Image Processing Department, for providing me with segmented endpoints.

I further acknowledge the support by Prof Claudia Redenbach for providing me with the topic and guidance.

On a more personal note I wish to thank all those that helped me through this.

First, this is again my boss Claudia Redenbach, who cut me a lot of slack where it was needed and believed (or at least hoped) that I could steer this ship home, even when I did not.

I, of course, thank the whole statistics group the TU Kaiserslautern and the Image Processing Department at ITWM, for lots of good discussion, coffee and companionship.

Then I want to thank my friends for moral support, listening to my complaints and eating loads of burgers. These are, in no particular order, Markus, Dennis, Jan, Laura, Jean-Pierre, Andre, Kerstin and Konstantin.

I thank my mother, who put up with my more erratic behaviour, especially during the dark days, for support and her believing in me.

I want to thank Maela, for seeing more in me than I do, for her love and her compassion.











And I want to thank the cats, who stopped me from working too much by sitting on my laptop on multiple occasion.

Contents

1. Introduction	9
1.1. Fibre reinforced polymers	9
1.2. Ashing	10
1.3. CT images	10
1.4. Problem	11
1.5. State of the art in length estimation	12
2. Mathematical Introduction	15
2.1. Basic Results	15
2.1.1. Probability theory	15
2.1.2. Random Variables	16
2.1.3. Expectation	17
2.1.4. Convergence	18
2.1.5. Integration and differentiation	18
2.2. Statistical Inference	20
2.2.1. Maximum likelihood	20
2.2.2. Expectation Maximisation algorithm	27
2.3. Stochastic Geometry	30
2.3.1. Point Processes	30
2.3.2. Cluster Processes	33
2.3.3. Germ-grain Processes	34
2.4. Polar and spherical angles and their distributions	36
2.5. Fibre Models	39
2.5.1. Germ Model	39
2.5.2. Boolean Model	39
2.5.3. RSA Fibres	40
3. Estimation from fully segmented fibres	43
3.1. Problem	43
3.2. Horvitz-Thompson type estimator	44
3.3. EM-Estimator	46
3.3.1. Reweighed Densities	46
3.3.2. expectation maximization (EM)-Algorithm	57
3.3.3. Distribution of $L^\pi \tilde{L}^\pi$	59
3.3.4. Distribution of $\tilde{L}^\pi L^\pi$	60
3.3.5. Properties of the estimator	72
3.3.6. Variance of the estimator	84

3.4.	Simulation Study	88
3.4.1.	Geometrical Setting	89
3.4.2.	HT estimator	89
3.4.3.	Associated point rule	92
3.4.4.	Plus-sampling	94
3.4.5.	Influence of the orientation distribution on the estimator	99
3.4.6.	Alternative implementations	99
3.4.7.	HT weight or kernel density	100
4.	Estimation from fibre endpoints	103
4.1.	Fibre endpoint process	103
4.2.	Summary Statistics	103
4.3.	Estimators	104
4.3.1.	Non parametric estimation	104
4.3.2.	Minimum Contrast estimator	107
5.	Comparison and Applications	111
5.1.	Interacting Fibres	111
5.2.	Parametric estimation	111
5.2.1.	Full Fibres	111
5.2.2.	Fibre Endpoints	112
5.2.3.	Application to real data	113
5.3.	Conclusion	115
A.	Simulation Study Results	117
A.1.	HT estimator	118
A.2.	Associated point rule	120
A.3.	Plus Sampling	122
B.	CV	125

List of Figures

1.1.	Injection moulding machine	10
1.2.	A sketch of a μ CT. The source emits X-Ray beams through the turning sample, the X-Ray beams are then detected by the detector. . .	11
1.3.	Here we see two slices from FRPs. On the left we see a FRP with 15% mass of glass fibres and on the right with 50%.	12
2.1.	An orientation with u with the corresponding angle φ	36
2.2.	An orientation with u with the corresponding angle φ and ϑ	37
2.3.	Axial distribution with $\beta = 0.1$	38
2.4.	Isotropic distribution with $\beta = 1.0$	38
2.5.	Girdle distribution with $\beta = 10.0$	38
2.6.	Boolean and RSA fibres with a density of $1.99e-5$ which leads to volume content of 10% fibres for the RSA case, with a Schladitz β distribution with $\beta = 0.1$ and log normal length distribution with $\mu = 4.6002$ and $\sigma = 0.0998$. The radius is 4. The view is on the $x - y$ plane. The window size is $[0, 500]^3$ for the plus sampling a window of size $[0, 800]^3$ was used. The Boolean model is on the left, the RSA process is on the right.	42
3.1.	An example for a 2d line segment process.	43
3.2.	Visualisation of the observation window W	44
3.3.	Visualisation of the area $\Pi_-(l, \varphi)$ as  , the window W is given as  . Following this sampling rule the example fibre is sampled and fully visible in W	45
3.4.	Visualisation of the area $\Pi_a(l, \varphi)$ as  , the window W is given as  . Following this sampling rule the example fibre is sampled but only the black part will be visible in the window W and the fibre will be censored.	48
3.5.	Visualisation of the area $\Pi_+(l, \varphi)$ as  , the window W is given as  . Following this sampling rule the example fibre is sampled but only the black part will be visible in the window W and the fibre will be censored.	49
3.6.	Visualisation of the area p_1 as  , the window W is given as  . . .	49
3.7.	Visualisation of the area p_2 as  , the window W is given as  . . .	50
3.8.	The EM algorithm uses the visible part of the fibre inside the window and "adds" some length to it to get an estimate of the real length . . .	58

3.9. Visualisation of the area $a(l, \tilde{l}, \varphi)$ as --- , the window W is given as --- . In this case we have $\tilde{l} = \frac{l}{3}$	61
3.10. Visualisation of the area $a(l, \tilde{l}, u)$ as --- , the window W is given as --- . For this case $\tilde{l} = \frac{l}{3}$	65
3.11. Visualisation of the triangle T from $a(l, \tilde{l}, u)$ as --- , the window W is given as --- and $a(l, \tilde{l}, u)$ as --- . For this case $\tilde{l} = \frac{l}{3}$	65
3.12. Visualisation of the parallelogram P_1 from $a(l, \tilde{l}, u)$ as --- , the window W is given as --- and $a(l, \tilde{l}, u)$ as --- . For this case $\tilde{l} = \frac{l}{3}$	66
3.13. Visualisation of the parallelogram p_1 from $a(l, \tilde{l}, u)$ as --- , the window W is given as --- and $a(l, \tilde{l}, u)$ as --- . A fibre with length l has been placed on the border of $a(l, \tilde{l}, u)$. The midpoints of the fibre is marked by \circ For this case $\tilde{l} = \frac{l}{3}$	67
3.14. Visualisation of the area $a(l, \tilde{l}, u)$ as --- , the window W is given as --- . For this case $\tilde{l} = \frac{2l}{3}$	68
3.15. Visualisation of the area inner rectangle as --- , the window W is given as --- and. For this case $\tilde{l} = \frac{2l}{3}$	68
3.16. Visualisation of the inner rectangle as --- , the window W is given as --- and. For this case $\tilde{l} = \frac{2l}{3}$	69
4.1. Comparison of the isotropic estimator on the left and the anisotropic estimator on the right for window size $W_l = 400$. The mean of the estimator is solid, the true value is dashed and an estimated 95% region is dotted. Note the different scale in the figures.	106
4.2. Comparison of the isotropic estimator on the left and the anisotropic estimator on the right for window size $W_l = 600$. The mean of the estimator is solid, the true value is dashed and an estimated 95% region is dotted.	107
5.1. Comparison of the anisotropic length estimator on the left for window size $W_l = 400$ and on the right for window size $W_l = 600$ for RSA fibres. The mean of the estimator is solid, the true value is dashed and an estimated 95% region is dotted. The RSA fibres had a volume fraction of 10%	113
5.2. Comparison of the anisotropic length estimator on the left for window size $W_l = 400$ and on the right for window size $W_l = 600$ for RSA fibres. The mean of the estimator is solid, the true value is dashed and an estimated 95% region is dotted. The RSA fibres had a volume fraction of 10%	114
5.3. Anisotropic length estimator for segmented end points provided by Markus Kronenberger	115

1. Introduction

1.1. Fibre reinforced polymers

The information presented here about polymers and fibre reinforcement are cited from [2] and [11]. Fibre reinforced polymers are one of the newest and modern materials that borrowed their idea from one of the oldest, wood. In wood, cellulose fibres are embedded into a matrix made of lignin. The matrix is light and the cellulose fibres are strong. The result is a material that is able to hold up a whole tree without being crushed by its own weight. We use the same idea and reinforce a mechanically weak but light polymer matrix with stiff fibres.

So the goal for fibre reinforced polymers (FRP) is to get a low density material with high mechanical strength and stiffness. For instance polyamide(PA-6) has a modulus of elasticity in tension of $1200 \frac{\text{N}}{\text{mm}^2}$ and a density of $1.14 \frac{\text{g}}{\text{cm}^3}$, by adding 30% in mass of short glass fibres we get an elasticity in tension of $5500 \frac{\text{N}}{\text{mm}^2}$ while only raising the density to $1.36 \frac{\text{g}}{\text{cm}^3}$.

Possible polymers used are all kinds of polyamides, polypropylene etc. The fibres can be made of carbon, aramid or glass. While glass fibres are most abundantly used, carbon fibres lead to a stronger but more expensive material.

To produce parts made of FRP, there are many possible methods available. Examples are

- Hand-lay up: The fibres are manually placed in a mould and the polymer is poured in afterwards
- Wet moulding: The fibres are placed in a mould, the polymer is applied and then the mould closes under pressure
- injection moulding: The fibres and the polymer are premixed, molten and pressed into a form

Here we will especially look into the injection moulding process. A sketch of an injection moulding machine can be seen in Figure 1.1. For the injection moulding the FRP are bought in a granulate form, which consists of fibres already in cased in the polymer matrix. This granulate is put into a hopper which feeds into a screw extruder. In this extruder the granulate is molten down. During the melting the fibres are put under a large amount of stress, both thermal and mechanical. The FRP mass is then pushed through a nozzle into the mould. In this nozzle the fibres are put under large stress again. The stress endured during the melting in the screw and the nozzle lead to fibres breaking.

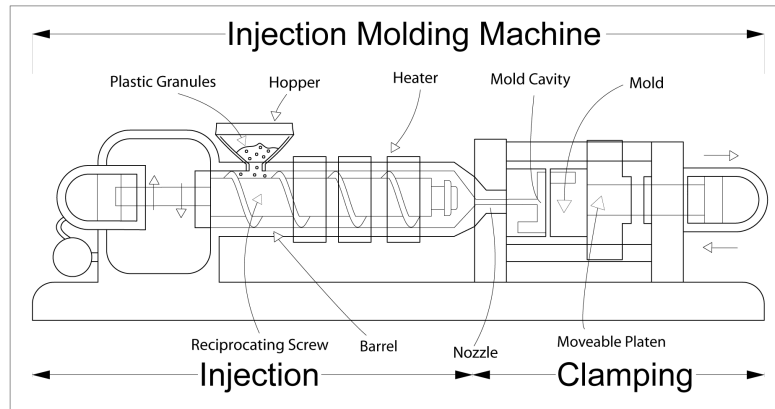


Figure 1.1.: Injection moulding machine^a

^aImage by User:Brockey, CC BY 3.0, <https://commons.wikimedia.org/w/index.php?curid=21365163>

Besides the used materials the mechanical properties of FRP are governed by two geometrical properties. The first is the orientation, FRP have a higher Young's modulus in direction with the fibres. The second is the length of the fibres, longer fibres lead to stronger materials. That means that stress endured by the fibres and the resulting fibre breaking will lead to worse mechanical performance.

1.2. Ashing

One standard method to investigate FRP is ashing. For the ashing process a sample of the FRP is placed in an muffle furnace and heated such that the polymer burns away, while the fibres do not burn and are left behind.

By weighing the sample before and after the burning we can get an estimate of the mass fraction of the polymer and the fibres. Furthermore these fibres can be put into a liquid solution and their length can then be measured by using a microscope.

With this method it is possible to get an estimation of the fibre length distribution. The problem with this method is, that it might affect the lengths of the fibres. During the burning fibres might break due to thermal strain. When the fibres are brought into solution and get divided up more of them might break. The fibre length distribution estimated from an ashed sample is most likely biased toward shorter fibres. Furthermore, during the ashing all information about the geometry, especially the orientation distribution of the fibres is lost.

1.3. CT images

To get a better understanding of the 3d geometry inside of different materials μ CT is used. They work by projecting an X-ray beam through the sample onto a detector

while the sample is turning, as seen in Figure 1.2. The data collected is then used

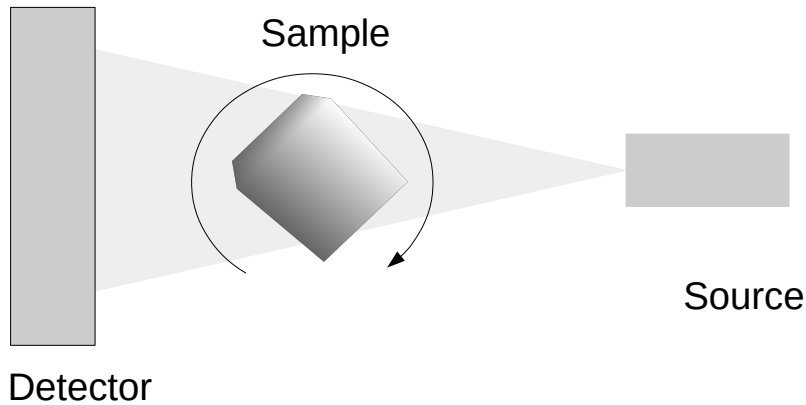


Figure 1.2.: A sketch of a μ CT. The source emits X-Ray beams through the turning sample, the X-Ray beams are then detected by the detector.

to reconstruct a 3d image of the sample using the inverse Radon transformation.

The Radon transformation is known since 1917 in [16]. Further work focusing on human tissue was done by Allen M. Cormack from 1957 to 1963. The first working prototype was manufactured in 1969 by Godfrey Hounsfield. CTs were first used in the medical field but were refined to also work on μ m scales. The result is the μ CT.

1.4. Problem

By using a μ CT we get a 3d image of a sample of FRP with intact orientational data. We therefore observe these fibres in a bounded window W . We want to use this to estimate the fibre length distribution. The estimation of the orientation distribution from CT data has been studied in [3].

If we wish to estimate the length distribution, there are two main problems we have to deal with. The first problem is censoring. If we look at the image we will directly see that fibres will be cut off at the edge of the window. This censoring problem is usually discussed in survival analysis. We will use the expectation maximization (EM) algorithm to obtain a maximum likelihood (ML) estimator for the fibre length distribution.

The second is a sampling problem. If we look at all fibres visible in the window we will note that, if these fibres were observed completely, longer fibres will be overrepresented in the sample since they have a higher probability to cut the window. This sampling bias would lead to an over estimation of the length. We will propose methods based on sampling rules and reweighing of the distribution to deal with this bias.

The method using the EM algorithm will only work if we can fully segment the fibres. This full segmentation is problematic, since the length of the fibres is big

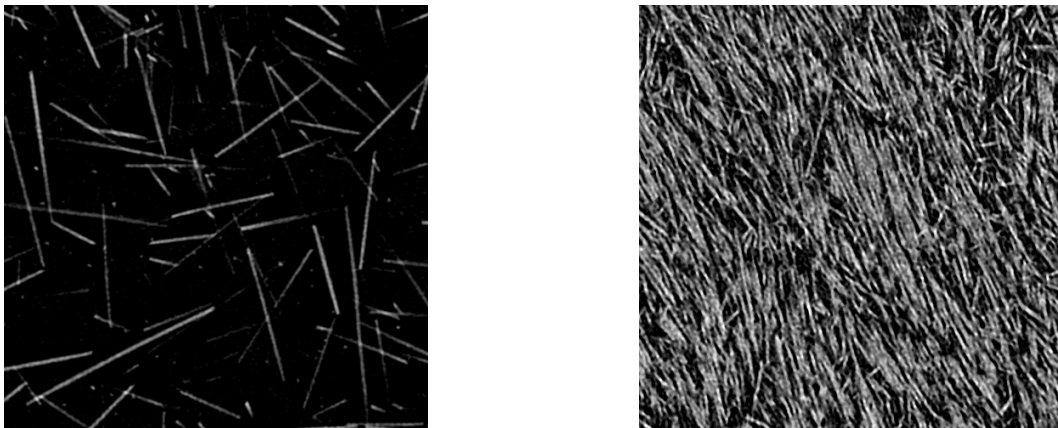


Figure 1.3.: Here we see two slices from FRPs. On the left we see a FRP with 15% mass of glass fibres and on the right with 50%.

compared to their radius. This leads to a problem of proper resolution. If we resolve the full length of the fibres, the resolution will be low and fibres will be hard to separate. If we have a good resolution of single fibres, the most if not all fibres will be censored. In Figure 1.3 we see two samples with different fibre content. We see that even a manual fibre segmentation would be hard for the low fibre content and impossible for the sample with the high content.

If we follow the approach given in [10], we do not segment the fibres but the fibre endpoints using local Gaussian curvature. These endpoints will then be interpreted as an endpoint process. We may then use point process statistics to estimate the length distribution.

1.5. State of the art in length estimation

Currently there are multiple works done on the fibre lengths estimation. In [14] an algorithm based on the the constrained path opening is presented to estimate the fibre length distribution. The constrained path opening is a morphological filter. Our work extends the work done in [21]. They investigated the fibre lengths distribution in wood cores. For this they had to deal with sampling bias and censoring as well. They solved the censoring problem by using an EM algorithm. They proofed consistency, asymptotic normality and convergence for their case in [22].

In their model they assumed that all fibres had the same orientation which essentially turned it into a 2d problem. We will extend this to random orientations. Furthermore they dealt with the sampling bias by first estimating the parameters of their distribution and then reweighing this distribution. We will directly reweigh the densities and then estimate the parameters. We will proof the consistency and asymptotic normality of the obtained estimators.

For the end point estimation we will look at an approach in [12]. They modelled

the endpoint process and found a closed formula for Ripley's- K function, which depends on the length. We will investigate this estimator for its viability and also propose an alternative minimum contrast estimator.

2. Mathematical Introduction

In this chapter we will give an overview over the mathematical theory used in this thesis. Namely these are basic stochastic, analysis regarding differentiation and integration, maximum likelihood (ML) and expectation maximization (EM) estimation. We also introduce point processes and some summary statistics. At the end we introduce random distributions on the sphere and two fibre models.

2.1. Basic Results

2.1.1. Probability theory

Definition 2.1. A σ -algebra \mathcal{A} is a class of subsets of a set Ω that satisfies

- $\emptyset \in \mathcal{A}$
- If $A_1, A_2, \dots \in \mathcal{A}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
- If $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

with A^c being the complement of a set A .

The sets in \mathcal{A} are called measurable, the pair (Ω, \mathcal{A}) is called a measurable space.

Definition 2.2. Consider a measurable space (Ω, \mathcal{A}) . A function $\mu : \mathcal{A} \rightarrow \mathbb{R}_{\geq 0}$ is called a measure if

- For all $A \in \mathcal{A} : \mu(A) \geq 0$
- $\mu(\emptyset) = 0$
- For all countable collections $A_1, A_2, \dots \in \mathcal{A}$ pairwise disjoint

$$\mu \left(\bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mu(A_i)$$

Definition 2.3. Suppose (Ω, \mathcal{A}) and (Ω', \mathcal{A}') are measurable spaces. A function

$$X : \Omega \rightarrow \Omega'$$

is called measurable if $X^{-1}(A') \in \mathcal{A}$ for all $A' \in \mathcal{A}'$.

Definition 2.4. Consider a measurable space (Ω, \mathcal{A}) . The measure P is a probability measure if $P(\Omega) = 1$.

With this we can define the probability space as a triple (Ω, \mathcal{A}, P) . We call Ω the sample space, the set of all possible outcomes of a random experiment. Elements $\omega \in \Omega$ are called events. We call sets $A \in \mathcal{A}$ events. We say $P(A)$ is the probability for the event $A \in \mathcal{A}$.

Definition 2.5. Suppose $A, B \in \mathcal{A}$ and $P(B) > 0$. The conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

The measure $P(\cdot|B)$ is a probability measure as well.

Theorem 2.1 (Bayes' Theorem). Suppose $A, B \in \mathcal{A}$ and $P(B) > 0$. Then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

2.1.2. Random Variables

Definition 2.6. Suppose (Ω, \mathcal{A}, P) is a probability space and (Ω', \mathcal{A}') a measurable space. A random variable X is a measurable mapping

$$X : \Omega \rightarrow \Omega'.$$

This introduces a measure P_X on Ω' as

$$P_X(X \in A') = P(X^{-1}(A'))$$

If $\Omega' = \mathbb{R}$ we speak of a real valued random variable.

Definition 2.7. We define

$$F_X(x) = P(X \leq x)$$

as the cumulative distribution function (CDF).

Definition 2.8. A real valued random variable X is called continuous if there exists a function f_X such that $f_X(x) \geq 0$,

$$\int_{-\infty}^{\infty} f_X(x) \, dx = 1$$

and

$$P(a < X < b) = \int_a^b f_X(x) \, dx.$$

We call f_X the probability density function (PDF) of X .

These definitions can be directly expanded to the bivariate case as well. If we have two real random variables X, Y their joint PDF is given by $f_{X,Y}(x, y)$ with the same properties.

The corresponding marginal densities are given by

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) \, dy \quad \text{and} \quad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) \, dx.$$

Definition 2.9. The conditional probability density function is given by

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

if $f_Y(y) > 0$. We further get

$$P(X \in A|Y = y) = \int_A f_{X|Y}(x|y) \, dx.$$

We can get the density of X by marginalising

$$f_X(x) = \int f_{X|Y}(x|y) f_Y(y) \, dy.$$

Furthermore, Bayes theorem can be expanded to densities as well via

$$f_{X|Y}(x, y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}.$$

2.1.3. Expectation

Definition 2.10. The mean or expected value of a random variable X is defined as

$$\mathbb{E}[X] = \int x \, dF_X(x) = \int x f_X(x) \, dx$$

with the last equality holding only if X is continuous with density f_X

The variance of X is defined as

$$\text{var}(X) = \mathbb{E}[X - \mathbb{E}[X]]^2$$

Definition 2.11. The conditional expectation of X given $Y = y$ is

$$\mathbb{E}[X|Y = y] = \int x f_{X|Y}(x|y) \, dx$$

Note that the conditional expectation $\mathbb{E}[X|Y]$ itself is a random variable depending on Y .

2.1.4. Convergence

Definition 2.12. Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables.

- X_n converges to X in p -th mean, $X_n \xrightarrow{L^p} X$, if

$$\mathbb{E} [|X_n - X|^p] \xrightarrow{n \rightarrow \infty} 0.$$

- X_n converges to X in probability, $X_n \xrightarrow{P} X$, if for all $\epsilon > 0$

$$P(|X_n - X| \geq \epsilon) \xrightarrow{n \rightarrow \infty} 0.$$

- X_n converges to X almost surely, $X_n \xrightarrow{a.s.} X$, if

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

- Assume that X, X_1, X_2, \dots have the distribution functions F, F_1, F_2, \dots . X_n converges to X in distribution, $X_n \xrightarrow{d} X$, if

$$F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$$

for all points of continuity of F .

2.1.5. Integration and differentiation

In this subsection we will give the basic results about switching the order of differentiation and integration. These results are cited from Chapter 7 of [18].

Definition 2.13. Let $f(x; \theta)$ with $x \in \mathbb{R}$ and $\theta \in \Theta \subset \mathbb{R}^d$. For all $\theta \in \Theta$ let $f(x; \theta)$ be integrable in x over the interval $[a, b]$ with $-\infty \leq a < b \leq \infty$. Then

$$F(\theta) = \int_a^b f(x; \theta) \, dx$$

is the parameter integral depending on θ .

Theorem 2.2. If $-\infty < a < b < \infty$ and $f(x; \theta)$ is continuous in $(a, b) \times \Theta$ then the parameter integral is continuous in θ .

If $-\infty \leq a < b \leq \infty$, if we have a majorant $g_0(x)$ such that

$$|f(x; \theta)| \leq g_0(x) \text{ and } \int_a^b g_0(x) \, dx < \infty$$

then the parameter integral is continuous in θ .

Definition 2.14. Let $f(x; \theta)$ with $x \in \mathbb{R}$ and $\theta \in \Theta \subset \mathbb{R}^d$ be integrable in x and n times continuously differentiable in θ and $-\infty \leq a < b \leq \infty$. We say that we can switch differentiation and integration for the parameter n times if

$$\frac{\partial^n}{\partial^{i_1} \theta_1 \dots \partial^{i_d} \theta_d} \int_a^b f(x; \theta) \, dx = \int_a^b \frac{\partial^n}{\partial^{i_1} \theta_1 \dots \partial^{i_d} \theta_d} f(x; \theta) \, dx$$

holds for all combinations of $i_j \in \mathbb{N}^0$ and $1 \leq j \leq d$ with $i_1 + \dots + i_d \leq n$.

Theorem 2.3. Let $f(x; \theta)$ with $x \in \mathbb{R}$ and $\theta \in \Theta \subset \mathbb{R}^d$ be integrable in x and n times continuously differentiable in θ .

If the integration bounds are bounded, i.e. $-\infty < a < b < \infty$, we can switch the differentiation and integration for the parameter θ n times.

If one or both of the integral bounds are unbounded we can use a majorant criterion.

Theorem 2.4. Let $f(x; \theta)$ with $x \in \mathbb{R}$ and $\theta \in \Theta \subset \mathbb{R}^d$ be integrable in x and n times continuous differentiable in θ . If we have majorants $g_{i_1, \dots, i_d}(x)$ such that

$$\left| \frac{\partial^n}{\partial^{i_1} \theta_1 \dots \partial^{i_d} \theta_d} f(x; \theta) \right| \leq g_{i_1, \dots, i_d}(x) \text{ and } \int_a^b g_{i_1, \dots, i_d}(x) \, dx < \infty$$

for all combinations of $i_j \in \mathbb{N}^0$ and $1 \leq j \leq d$ with $i_1 + \dots + i_d \leq n$, then we can switch differentiation and integration for the parameter θ n times for $f(x; \theta)$.

Theorem 2.5 (Leibniz integral rule). Let $f(x; \theta)$ with $x \in \mathbb{R}$ and $\theta \in \Theta \subset \mathbb{R}^d$ be integrable in x and continuously differentiable in θ . Let $a(\theta)$ and $b(\theta)$ be continuously differentiable in θ and $a(\theta), b(\theta) < \infty$ for all θ , then

$$\frac{\partial}{\partial \theta_i} \left(\int_{a(\theta)}^{b(\theta)} f(x; \theta) \, dx \right) = f(b(\theta), \theta) \frac{\partial}{\partial \theta_i} b(\theta) - f(a(\theta), \theta) \frac{\partial}{\partial \theta_i} a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial \theta_i} f(x; \theta) \, dx.$$

Theorem 2.6. Let $f(x; \theta)$ with $x \in \mathbb{R}$ and $\theta \in \Theta \subset \mathbb{R}^d$ be integrable in x and continuous differentiable in θ . Let $a(\theta)$ be continuously differentiable in θ . Is furthermore $g \in L^1(\mathbb{R})$ a majorant with

$$|f(x; \theta)| \leq g(x)$$

then

$$\frac{\partial}{\partial \theta_i} \lim_{b \rightarrow \infty} \int_{a(\theta)}^b f(x; \theta) \, dx = \lim_{b \rightarrow \infty} \frac{\partial}{\partial \theta_i} \int_{a(\theta)}^b f(x; \theta) \, dx$$

holds.

2.2. Statistical Inference

We will cite the basic results for ML estimators. We will also give the definition of the EM algorithm and basic convergence results.

2.2.1. Maximum likelihood

Suppose we have $n \in \mathbb{N}$ random variables X_1, \dots, X_n that are independent and identically distributed (iid) with $X_1 \sim F_X(x; \theta)$, with a parameter vector $\theta \in \mathbb{R}^d$. To estimate the distribution F_X we need to find an estimator $\hat{\theta}$. If we further assume that $F_X(x; \theta)$ is a continuous distribution with density $f(x; \theta)$, one classical method to find such an estimator is to use the maximum likelihood method.

Definition 2.15. Given X_1, \dots, X_n iid with PDF $f(x; \theta)$. The likelihood function is defined by

$$\mathcal{L}(\theta|X) = \prod_{i=1}^n f(X_i; \theta).$$

The log-likelihood function is defined by $l(\theta|X) = \log(\mathcal{L}(\theta|X))$.

The ML estimator is obtained by

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(\theta|X) = \arg \max_{\theta} l(\theta|X).$$

Definition 2.16. Given a random variable X with PDF $f(x; \theta)$ and parameter vector $\theta \in \mathbb{R}^d$ with θ_i being the i th component of the vector θ . The Fisher information matrix is defined as

$$\mathcal{I}(X; \theta) = \mathbb{E} \left[\nabla_{\theta} \ln f(x; \theta) (\nabla_{\theta} \ln f(x; \theta))^T \right],$$

where $\nabla_{\theta} \ln f(x; \theta)$ is the gradient in θ . If we can exchange the integration and differentiation with respect to θ_i we can write

$$\mathcal{I}(X; \theta) = \mathbb{E} [-\mathcal{H}(\ln f(X; \theta))]$$

with $\mathcal{H}(\ln f(X; \theta))$ being the Hessian matrix of $\ln f(X; \theta)$.

The Fisher information matrix is linked to the variance of an estimator.

Theorem 2.7 (Cramér-Rao inequality). Let X_1, \dots, X_n be iid with PDF $f(x; \theta_0)$. with $\theta \in \Theta \subset \mathbb{R}^d$ with true parameter $\theta_0 \in \Theta$. Let $\hat{\theta}$ be an estimator and $\psi(\theta) = \mathbb{E}_{\theta_0}(\hat{\theta}) - \theta_0$ is the bias of this estimator. The Cramér-Rao bound is

$$\text{cov}_{\theta}(\hat{\theta}) \geq \nabla_{\theta} \psi(\theta) \mathcal{I}(X; \theta)^{-1} \nabla_{\theta} \psi(\theta)^T,$$

where $\nabla_{\theta} \psi(\theta)$ is the Jacobian matrix of $\psi(\theta)$.

If the estimator is unbiased the variance of the estimator is bounded from below by

$$\text{var}(\hat{\theta}_i) = \text{cov}_{\theta}(\hat{\theta})_{i,i} \geq (\mathcal{I}(X; \theta)^{-1})_{i,i}. \quad (2.1)$$

That means that an unbiased estimator cannot be arbitrarily good in terms of its variance.

Theorem 2.8. *Let X_1, \dots, X_n be iid with PDF $f(x; \theta)$ with $\theta \in \Theta \subset \mathbb{R}^d$, Θ compact with true parameter $\theta_0 \in \Theta$ and $\hat{\theta}_n$ the ML estimate. If these regularity conditions*

R-1 *The PDF are distinct, i.e. $\theta \neq \theta' \Rightarrow f(\cdot; \theta) \neq f(\cdot; \theta')$ almost everywhere.*

R-2 *The PDF have common support for all $\theta \in \Theta$*

R-3 *There exists an open subset $\Theta_0 \subset \Theta$ such that $\theta_0 \in \Theta_0$ and all third partial derivatives of $f(x; \theta)$ exist for all $\theta \in \Theta_0$*

R-4 *There exist majorants like in Theorem 2.4 for all first and second derivatives of f for $\theta \in \Theta_0$*

R-5 *For all $\theta \in \Theta_0$, $\mathcal{I}(X, \theta)$ is positive definite*

R-6 *There exist functions $M_{i_1, \dots, i_d}(x)$ such that*

$$\left| \frac{\partial^3}{\partial^{i_1} \theta_1 \dots \partial^{i_d} \theta_d} \ln f(x; \theta) \right| \leq M_{i_1, \dots, i_d}(x), \forall \theta \in \Theta_0 \text{ and } i_j \in \mathbb{N}^0, \sum_{j=1}^d i_j = 3$$

and

$$\mathbb{E}_{\theta_0}(M_{i_1, \dots, i_d}(X)) < \infty, \text{ for } i_j \in \mathbb{N}^0, \sum_{j=1}^d i_j = 3$$

hold, then

1. *the likelihood equation*

$$\nabla_{\theta} l(\theta|X) = 0$$

has a solution $\hat{\theta}_n$ such that

$$\hat{\theta}_n \xrightarrow{\text{P}} \theta_0$$

2. *For any sequence that satisfies 1*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\text{d}} N(0, \mathcal{I}^{-1})$$

A proof of this result can be found in Chapter 6 of [6].

Remark

The condition **R-4** implies that we can switch the order of differentiation and integration for the parameter 2 times. This implies that the equations

$$\begin{aligned} \mathbb{E}_\theta \left[\frac{\partial}{\partial \theta_j} \ln f(x; \theta) \right] &= 0, \quad \text{for } j = 1, \dots, d \\ \mathcal{I}_{j,k}(X; \theta) &= \mathbb{E}_\theta \left[-\frac{\partial^2}{\partial \theta_j \partial \theta_k} \ln f(x; \theta) \right] \quad \text{for } j, k = 1, \dots, d \end{aligned} \tag{2.2}$$

hold, because of

$$\begin{aligned} 1 &= \int_{\mathbb{R}} f(x; \theta) \, dx \\ \Rightarrow 0 &= \frac{\partial}{\partial \theta_i} \int_{\mathbb{R}} f(x; \theta) \, dx \\ &= \int_{\mathbb{R}} \frac{\partial}{\partial \theta_i} f(x; \theta) \, dx \\ &= \int_{\mathbb{R}} \frac{\frac{\partial}{\partial \theta_i} f(x; \theta)}{f(x; \theta)} f(x; \theta) \, dx \\ &= \int_{\mathbb{R}} \frac{\partial}{\partial \theta_i} \ln f(x; \theta) f(x; \theta) \, dx \\ &= \mathbb{E} \left(\frac{\partial}{\partial \theta_i} \ln f(X; \theta) \right). \end{aligned}$$

If we take another derivative we get

$$\begin{aligned} 0 &= \frac{\partial^2}{\partial \theta_i \partial \theta_j} \int_{\mathbb{R}} f(x; \theta) \, dx \\ &= \int_{\mathbb{R}} \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x; \theta) \, dx \\ &= \int_{\mathbb{R}} \frac{\partial}{\partial \theta_j} \left(\frac{\partial}{\partial \theta_i} \ln f(x; \theta) f(x; \theta) \right) \, dx \\ &= \int_{\mathbb{R}} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x; \theta) f(x; \theta) \, dx + \int_{\mathbb{R}} \frac{\partial}{\partial \theta_i} \ln f(x; \theta) \frac{\partial}{\partial \theta_j} \ln f(x; \theta) f(x; \theta) \, dx \\ &= \mathbb{E} \left(\ln \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x; \theta) \right) + \mathbb{E} \left(\frac{\partial}{\partial \theta_i} \ln f(x; \theta) \frac{\partial}{\partial \theta_j} \ln f(x; \theta) \right) \\ &\Rightarrow \mathcal{I}(X; \theta) = \mathbb{E} [-\mathcal{H}(\ln(f(X; \theta)))] . \end{aligned}$$

Note, for the theorem of consistency and the asymptotic normality to hold, it is enough to demand that we may switch the order of differentiation and integration two times. We demand the existence of majorants in **R-4** for ease in later proofs.

The asymptotic normality directly implies, that the estimator is asymptotically unbiased and that its variance asymptotically reaches the Cramér-Rao bound given in equation (2.1). Therefore we wish to estimate the Fisher information matrix to get a bound for the variance of the estimator.

Theorem 2.9. *Let X_1, \dots, X_n be iid random variables with PDF $f(x; \theta)$ with $\theta \in \Theta$ with $\Theta \subset \mathbb{R}^d$ compact with true parameter $\theta_0 \in \Theta$ and $\hat{\theta}$ the ML estimator. Furthermore let the conditions of Theorem 2.8 be fulfilled, then*

$$\hat{\mathcal{I}}(X; \hat{\theta}) = -\frac{1}{n} \sum_{i=1}^n \mathcal{H}(\ln f(X_i; \hat{\theta}))$$

is a consistent estimator of the Fisher information matrix and

$$\hat{\mathcal{I}}^{-1}(X; \hat{\theta}) \xrightarrow{P} \mathcal{I}^{-1}(X; \theta)$$

In later examples we will model the fibre length as a log-normal distribution. We will therefore apply Theorem 2.8 to a log-normal distribution to show the consistency and asymptotic normality of the ML estimator.

Theorem 2.10. *Let X_1, \dots, X_N be iid random variables with a log normal distribution. The density is given by*

$$f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2x}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right).$$

with parameters $\theta = (\mu, \sigma)$. Let $\Theta \subset \mathbb{R}^2$ compact be the parameter space. Let $\Theta_0 \subset \Theta$ open with the true parameters $\theta_0 = (\mu_0, \sigma_0) \in \Theta_0$. Then the ML estimator $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ is consistent and asymptotically normal.

Proof. **R-1** Follows by direct calculations

R-2 The support of $f_X(x; \mu, \sigma)$ is always $(0, \infty)$.

R-3 The existence of the open subset Θ_0 follows by assumption. $f_X(x; \mu, \sigma)$ is arbitrarily smooth in the parameters, therefore the third derivatives exist.

R-4 We choose $\mu_{max} = \max\{|\mu| \mid \mu \in \Theta\}$ and σ_{min} and σ_{max} similarly. To construct the majorants we look at the first derivatives and get

$$\begin{aligned} \frac{\partial}{\partial \mu} f_X(x; \mu, \sigma) &= \frac{\ln(x) - \mu}{\sqrt{2\pi\sigma^3x}} \exp\left(\frac{-(\ln(x) - \mu)^2}{2\sigma^2}\right) \\ &= \frac{\ln(x) - \mu}{\sigma^2} f_X(x; \mu, \sigma) \\ &\leq \frac{|\ln(x)| + |\mu|}{\sigma^2} f_X(x; \mu, \sigma) \\ &\leq \frac{|\ln(x)| + \mu_{max}}{\sigma_{min}^2} f_X(x; \mu, \sigma) \end{aligned} \tag{2.3}$$

for the derivative in μ and

$$\begin{aligned}
\frac{\partial}{\partial \sigma} f_X(x; \mu, \sigma) &= \frac{(\ln(x) - \mu)^2 - \sigma^2}{\sqrt{2\pi\sigma^4 x}} \exp\left(\frac{-(\ln(x) - \mu)^2}{2\sigma^2}\right) \\
&= \frac{\ln(x)^2 - 2\ln(x)\mu + \mu^2 - \sigma^2}{\sigma^3} f_X(x; \mu, \sigma) \\
&\leq \frac{\ln(x)^2 + 2|\ln(x)||\mu| + \mu^2 + \sigma^2}{\sigma^3} f_X(x; \mu, \sigma) \\
&\leq \left(\frac{\ln(x)^2 + 2|\ln(x)|\mu_{max} + \mu_{max}^2}{\sigma_{min}^3} + \frac{1}{\sigma_{min}} \right) f_X(x; \mu, \sigma)
\end{aligned} \tag{2.4}$$

for the derivative in σ .

We see that if we can find a majorant for $|\ln(x)|^k f_X(x; \mu, \sigma)$, $k = 0, 1, 2$ we have the majorants for the first derivatives. To find a majorant for the density $f_X(x; \mu, \sigma)$ we start to look at the exponent of the density

$$\begin{aligned}
\frac{-(\ln(x) - \mu)^2}{2\sigma^2} &= \frac{-\ln(x)^2 + 2\mu\ln(x) - \mu^2}{2\sigma^2} \\
&\leq \frac{-\ln(x)^2 + 2\mu_{max}|\ln(x)|}{2\sigma^2} \\
&= \frac{-(|\ln(x)| - \mu_{max})^2 + \mu_{max}^2}{2\sigma^2} \\
&= \frac{-(|\ln(x)| - \mu_{max})^2}{2\sigma^2} + \frac{\mu_{max}^2}{2\sigma^2} \\
&\leq \frac{-(|\ln(x)| - \mu_{max})^2}{2\sigma_{max}^2} + \frac{\mu_{max}^2}{2\sigma_{min}^2}.
\end{aligned}$$

We get

$$\begin{aligned}
f_X(x; \mu, \sigma) &\leq \frac{1}{\sqrt{2\pi\sigma_{min}^2 x}} \exp\left(\frac{-(|\ln(x)| - \mu_{max})^2}{2\sigma_{max}^2} + \frac{\mu_{max}^2}{2\sigma_{min}^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma_{min}^2 x}} \exp\left(\frac{-(|\ln(x)| - \mu_{max})^2}{2\sigma_{max}^2}\right) \exp\left(\frac{\mu_{max}^2}{2\sigma_{min}^2}\right) \\
&= \frac{\exp\left(\frac{\mu_{max}^2}{2\sigma_{min}^2}\right) \sqrt{\sigma_{max}^2}}{\sqrt{\sigma_{min}^2}} \frac{1}{\sqrt{2\pi\sigma_{max}^2 x}} \exp\left(\frac{-(|\ln(x)| - \mu_{max})^2}{2\sigma_{max}^2}\right) \\
&= c \frac{1}{\sqrt{2\pi\sigma_{max}^2 x}} \exp\left(\frac{-(|\ln(x)| - \mu_{max})^2}{2\sigma_{max}^2}\right) =: g_0(x)
\end{aligned} \tag{2.5}$$

with $c = \frac{\exp\left(\frac{\mu_{max}^2}{2\sigma_{min}^2}\right) \sqrt{\sigma_{max}^2}}{\sqrt{\sigma_{min}^2}}$. We choose

$$g_k(x) = |\ln(x)|^k g_0(x) \tag{2.6}$$

as a majorant for $|\ln(x)|^k f_X(x; \mu, \sigma)$. It is L^1 since

$$\begin{aligned}
\int_{-\infty}^{\infty} |\ln(x)|^k g_0(x) \, dx &= \int_0^{\infty} |\ln(x)|^k c \frac{1}{\sqrt{2\pi\sigma_{max}^2}x} \exp\left(\frac{-(|\ln(x)| - \mu_{max})^2}{2\sigma_{max}^2}\right) \, dx \\
&= \int_{-\infty}^{\infty} |z|^k c \frac{1}{\sqrt{2\pi\sigma_{max}^2}} \exp\left(\frac{-(|z| - \mu_{max})^2}{2\sigma_{max}^2}\right) \, dz \\
&= 2c \int_0^{\infty} z^k \frac{1}{\sqrt{2\pi\sigma_{max}^2}} \exp\left(\frac{-(z - \mu_{max})^2}{2\sigma_{max}^2}\right) \, dz < \infty
\end{aligned}$$

where we substituted $z = \ln(x)$. We note that $\frac{1}{\sqrt{2\pi\sigma_{max}^2}} \exp\left(\frac{-(z - \mu_{max})^2}{2\sigma_{max}^2}\right)$ is the density of a normal distribution, for which all moments exist. Therefore the last inequality holds.

For the second derivative in μ we get

$$\begin{aligned}
\frac{\partial^2}{\partial \mu^2} f_X(x; \mu, \sigma) &= \frac{\ln(x)^2 - 2\mu \ln(x) + \mu^2 - \sigma^2}{\sigma^4} f_X(x; \mu, \sigma) \\
&\leq \frac{\ln(x)^2 + 2\mu_{max} |\ln(x)| + \mu_{max}^2 + \sigma_{max}^2}{\sigma_{min}^4} f_X(x; \mu, \sigma),
\end{aligned} \tag{2.7}$$

for σ we have

$$\begin{aligned}
\frac{\partial^2}{\partial \sigma^2} f_X(x; \mu, \sigma) &= \left(\frac{\ln(x)^4 - 4\mu \ln(x)^3}{\sigma^6} \right. \\
&\quad + \frac{(6\mu^2 - 5\sigma^2) \ln(x)^2 + (10\mu\sigma^2 - 4\mu^3) \ln(x)}{\sigma^6} \\
&\quad \left. + \frac{\mu^4 - 5\mu^2\sigma^2 + 4\sigma^4}{\sigma^6} \right) f_X(x; \mu, \sigma) \\
&\leq \left(\frac{\ln(x)^4 + 4\mu_{max} |\ln(x)|^3}{\sigma_{min}^6} \right. \\
&\quad + \frac{(6\mu_{max}^2 + 5\sigma_{max}^2) \ln(x)^2}{\sigma_{min}^6} \\
&\quad + \frac{(10\mu_{max}\sigma_{max}^2 + 4\mu_{max}^3) |\ln(x)|}{\sigma_{min}^6} \\
&\quad \left. + \frac{\mu_{max}^4 + 5\mu_{max}^2\sigma_{max}^2 + 4\sigma_{max}^4}{\sigma_{min}^6} \right) f_X(x; \mu, \sigma).
\end{aligned} \tag{2.8}$$

and the mixed second derivative is

$$\begin{aligned}
\frac{\partial^2}{\partial\mu\partial\sigma}f_X(x;\mu,\sigma) &= \frac{\ln(x)^3 - 3\mu\ln(x)^2 + (3\mu^2 - 3\sigma^2)\ln(x) - \mu^3 + 3\mu\sigma^2}{\sigma^5} \\
&f_X(x;\mu,\sigma) \\
&\leq \left(\frac{\ln(x)^3 + 3\mu_{max}\ln(x)^2}{\sigma_{min}^5} \right. \\
&\quad \left. + \frac{(3\mu_{max}^2 + 3\sigma_{max}^2)|\ln(x)| + \mu_{max}^3 + 3\mu_{max}\sigma_{max}^2}{\sigma_{min}^5} \right) f_X(x;\mu,\sigma).
\end{aligned} \tag{2.9}$$

We see that we can use the majorants $g_k(x)$ for $|\ln(x)|^k f_X(x;\mu,\sigma)$ derived previously and get the desired majorants for the second derivatives.

R-5 If we calculate the Fisher information we get

$$I(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}.$$

This is a diagonal matrix where all entries are positive. Therefore the eigenvalues are positive as well and the matrix is positive definite.

R-6 For this we look at

$$\begin{aligned}
\frac{\partial^3}{\partial\mu^3} \ln f_X(x;\mu,\sigma) &= 0 \\
\frac{\partial^3}{\partial\mu^2\partial\sigma} \ln f_X(x;\mu,\sigma) &= \frac{2}{\sigma^3} \\
\frac{\partial^3}{\partial\mu\partial\sigma^2} \ln f_X(x;\mu,\sigma) &= \frac{6(\ln(x) - \mu)}{\sigma^4} \\
\frac{\partial^3}{\partial\sigma^3} \ln f_X(x;\mu,\sigma) &= \frac{-2\sigma^2 + 12(\ln(x) - \mu)^2}{\sigma^5} \\
&= \frac{-2\sigma^2 + 12\ln(x)^2 - 24\ln(x)\mu + \mu^2}{\sigma^5}
\end{aligned}$$

We can construct

$$\begin{aligned}
M_{30}(x) &= 0, \\
M_{21}(x) &= \frac{2}{\sigma_{min}^3}, \\
M_{12}(x) &= \frac{6(|\ln(x)| + \mu_{max})}{\sigma_{min}^4} \text{ and} \\
M_{03}(x) &= \frac{2\sigma_{max}^2 + 12\ln(x)^2 + 24|\ln(x)|\mu_{max} + \mu_{max}^2}{\sigma_{min}^5}
\end{aligned}$$

where μ_{max}, σ_{max} and σ_{min} are defined like in the previous part of the proof.

All of these have finite mean and are majorants for their respective derivative.

We can therefore conclude, that the ML estimator for the log-normal distribution is consistent and asymptotically normal. \square

Remark

1. If we look at the ML estimators for the log-normal distribution we get

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln(X_i) \text{ and } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\ln(X_i) - \hat{\mu})^2,$$

since the X_i are log-normally distributed, the $\ln(X_i)$ are normally distributed with parameters μ, σ . We see that the consistency and asymptotic normality of these estimators follow directly by the law of large numbers and the central limit theorem. We did the full proof of all properties of Theorem 2.8 since in later examples we will model the fibre length with a log-normal distribution and we will use the regularity conditions from Theorem 2.8.

2. In the above theorem we demanded that the parameter space Θ is compact. In practical applications this is no problem, because will usually be able to bound the parameter space. For the fibres for instance we could say that the expected length should be bigger than a hydrogen atom and smaller than the diameter of the observable universe. If we do something similar for the variance, we get a bounded and closed parameter space for the parameters μ and σ .
3. We further demand the existence of the open subset $\Theta_0 \subset \Theta$ with $\theta_0 \in \Theta_0$. In \mathbb{R}^2 this means that the parameters θ_0 are not located on the boundary of Θ . If we choose Θ big enough, this will not happen.

2.2.2. Expectation Maximisation algorithm

In some instances we may not observe all the interesting data, some of it might be missing or latent. To get good estimations directly might be quiet cumbersome in these situations. In these cases the application of the EM algorithm might be good way to still get a ML estimator.

Suppose we observe data X with a distribution that results in a complicated log-likelihood function. Instead of working with this log-likelihood function we augment the observed data with latent data Y , such that the resulting log-likelihood $l(\theta|X, Y)$ is easy to maximise.

This idea was described by [1], with a proof of convergence into a local maximum. The proof for global convergence given in the paper has an error which was fixed in [24].

For the basic idea, assume we observe $X = (X_1, \dots, X_n)$ with PDF $f(x; \theta)$. The resulting likelihood is $\mathcal{L}(\theta|X) = f(X; \theta)$. Now let us assume that something of the data is missing, i.e. the X could be censored. We augment the data X with

Y , such that X, Y have the joint density $f(X, Y; \theta)$. This Y is used to add the missing information and is chosen in a way to make $f(X, Y; \theta)$ simple. If we take the conditional density from Definition 2.9 we get

$$\begin{aligned} f(Y|X; \theta) &= \frac{f(X, Y; \theta)}{f(X; \theta)} \\ \Rightarrow \ln f(Y|X; \theta) &= \ln(f(X, Y; \theta)) - \ln(f(X; \theta)) \end{aligned} \quad (2.10)$$

which is the basic identity of the EM algorithm. We use this to transform the likelihood function

$$l(\theta|X) = l(\theta|X, Y) - l(\theta|[Y|X])$$

where $l(\theta|X, Y)$ is the complete-data log-likelihood. For any θ_0 we can take the conditional expectation with respect to $Y|X$ and get

$$l(\theta|X) = \mathbb{E}_{\theta_0} [l(\theta|X, Y) - \mathbb{E}_{\theta_0} [l(\theta|[Y|X])|X]]. \quad (2.11)$$

The first part in this equality is called the expected complete-data likelihood

$$Q(\theta|\theta_0) = \mathbb{E}_{\theta_0} [l(\theta|Y, X)|X],$$

which we maximise via an iteration. For this we choose a starting value θ_0 . We compute

$$\theta_j = \arg \max_{\theta} Q(\theta|\theta_{j-1})$$

in each step until convergence is reached. The estimator is given after n steps as $\hat{\theta}_{EM} = \theta_n$.

Algorithm 1 The EM algorithm

Choose θ_0, ϵ
 $j \leftarrow 1$
repeat
 Compute $Q(\theta|\theta_{j-1}) = \mathbb{E}_{\theta_{j-1}} [l(\theta|Y, X)|X]$ {E-Step}
 Maximise $\theta_j = \arg \max_{\theta} Q(\theta|\theta_{j-1})$ {M-Step}
 $j \leftarrow j + 1$
until $\|\theta_j - \theta_{j-1}\| < \epsilon$ {convergence}
return θ_j

If we look at Algorithm 1 we can clearly see that we have to do two steps in each iteration. The E(xpectation) step calculates the mean of the complete-data log likelihood, while the M(aximisation) step maximises this mean for the parameter.

Theorem 2.11. *The EM sequence $\hat{\theta}_j$ defined by algorithm 1 satisfies*

$$l(\theta_{j+1}|X) \geq l(\theta_j|X)$$

with equality holding if and only if $Q(\theta_{j+1}|\theta_j) = Q(\theta_j|\theta_j)$.

This theorem assures that the likelihood will increase in each step and is shown by [1]. An easily verifiable convergence result can be found in [17].

Theorem 2.12. *If the expected complete-data log-likelihood $Q(\theta; \theta_0)$ is continuous in both θ and θ_0 , then every limit point $\hat{\theta}_{EM}$ of an EM sequence θ_i is a stationary point of $\mathcal{L}(\theta|X)$ and $\mathcal{L}(\theta_i|X)$ converges monotonically to $\mathcal{L}(\hat{\theta}_{EM}|X)$.*

For an EM sequence θ_i we now know, that the likelihood increases and that it converges to a stationary point of the likelihood function. This stationary point might only be a local maximum or saddle point. In practice it may be necessary to start the EM algorithm for multiple start points to assure that a global maximum is reached.

Global convergence is studied in [24]. The paper states that if the EM algorithm does not get trapped in a stationary point, that is not a local maximum of the likelihood function, it converges to a local maximum of the likelihood function. If furthermore the likelihood function is uni modal with a unique stationary point and the gradient $\nabla_{\theta_0} Q(\theta|\theta_0)$ is continuous in θ and θ_0 then the EM algorithm converges to the unique maximizer of the likelihood function. These conditions are hard to verify, especially the uni modality of the underlying likelihood function cannot be shown in many cases.

To estimate the variance of the EM estimator we need to calculate the Fisher information matrix. An overview of different methods for this is given in [23, 74f]. We will use the so called Louis method. From Equation (2.10) we conclude

$$-\mathcal{H}(\ln f(X; \theta)) = -\mathcal{H}(\ln f(X, Y; \theta)) + \mathcal{H}(\ln f(Y|X; \theta)).$$

or in terms of likelihood we get

$$-\mathcal{H}(l(\theta|X)) = -\mathcal{H}(l(\theta|X, Y)) + \mathcal{H}(l(\theta|[X|Y])).$$

We now take the mean with respect to $Y|X; \theta_0$ on each side and get

$$-\mathcal{H}(l(X, \theta)) = \mathbb{E}_{\theta_0} [-\mathcal{H}(l(\theta|X, Y))|X] + \mathbb{E}_{\theta_0} [\mathcal{H}(l(\theta|[Y|X]))|X] \quad (2.12)$$

The Louis method states that

$$\mathbb{E}_{\theta_0} [\mathcal{H}(l(\theta|[Y|X]))|X] = -\text{cov}_{Y|X; \theta_0} \nabla_{\theta} l(\theta|X, Y) \quad (2.13)$$

where

$$\begin{aligned} \text{cov}_{Y|X; \theta_0} \nabla_{\theta} l(\theta|X, Y) &= E_{\theta_0} [(\nabla_{\theta} l(\theta|X, Y) - E_{\theta_0} [\nabla_{\theta} l(X, Y; \theta)|X]) \\ &\quad (\nabla_{\theta} l(X, Y; \theta) - E_{\theta_0} [\nabla_{\theta} l(\theta|X, Y)|X])^T |X]. \end{aligned}$$

If we plug the covariance into Equation (2.12) we get

$$\begin{aligned} -\mathcal{H}(l(X; \theta)) &= \mathbb{E}_{\theta_0} [-\mathcal{H}(l(\theta|X, Y))|X] - \text{cov}_{Y|X; \theta_0} \nabla_{\theta} l(\theta|X, Y) \\ &= -\mathcal{H}(Q(\theta|\theta_0)) - \text{cov}_{Y|X; \theta_0} \nabla_{\theta} l(\theta|X, Y). \end{aligned} \quad (2.14)$$

If we have a sample X and an EM estimator $\hat{\theta}_{EM}$ we can estimate the Fisher information with

$$\hat{\mathcal{I}}(X, \hat{\theta}_{EM}) = -\mathcal{H}(Q(\hat{\theta}_{EM}|\hat{\theta}_{EM})) - \text{cov}_{Y|X; \hat{\theta}_{EM}} \nabla_{\theta} l(\hat{\theta}_{EM}|X, Y). \quad (2.15)$$

We will use the EM algorithm here to deal with the censoring of the fibres.

2.3. Stochastic Geometry

Here we will give a quick overview of stochastic geometry and spatial statistics used in the following chapters. We will start with the definition of a point process and the Poisson point process and give an overview of the used summary statistics. Furthermore, we will give a definition of a particle process. The results here are cited from [15].

2.3.1. Point Processes

Let E be a locally compact space with a countable base and Borel σ -Algebra \mathcal{B} . Let \mathcal{C} be the set of all compact sets in \mathcal{B}

A measure μ on (E, \mathcal{B}) is called locally finite if $\mu(C) < \infty$, for all $C \in \mathcal{C}$. The set

$$N = \{\mu \mid \text{locally finite}, \mu(B) \in \mathbb{N}_0 \cup \{\infty\} \text{ for all } B \in \mathcal{B}\}$$

is the set of all locally finite counting measures. A counting measure φ is called simple if $\varphi(\{x\}) \leq 1$ for all $x \in E$.

Definition 2.17 (Point Process). Let (Ω, \mathcal{A}, P) be a probability space. A point process is a measurable mapping

$$\Phi : (\Omega, \mathcal{A}, P) \rightarrow (N, \mathcal{N})$$

The distribution of the point process is the image measure $P_{\Phi} := \Phi(P)$. The σ -algebra \mathcal{N} is defined as the smallest σ -algebra on N to make all mappings $\Phi \mapsto \Phi(B)$ measurable, for all bounded Borel sets B .

A realisation of such a point process is a counting measure. A point process is called simple if its realisations are almost surely simple.

Definition 2.18. Let $E = \mathbb{R}^d$. A point process Φ is called stationary if $\Phi \stackrel{d}{=} \Phi + x$ for all $x \in \mathbb{R}^d$. It is called isotropic if $\Phi \stackrel{d}{=} R\Phi$ for all rotations $R \in \text{SO}_d$.

Definition 2.19. The intensity measure Λ of a point process Φ is

$$\Lambda(B) = \mathbb{E}[\Phi(B)]$$

for all Borel sets B . So $\Lambda(B)$ is the mean number of points in B .

If the point process is stationary then the intensity measure is translation invariant as well. This implies that

$$\Lambda(B) = \lambda \nu(B)$$

for some constant $0 < \lambda < \infty$ and $\nu(B)$ the Lebesgue measure of B .

To estimate the intensity for an isotropic and stationary point process Φ observed in a window W we use

$$\hat{\lambda} = \frac{\Phi(W)}{\nu(W)}. \quad (2.16)$$

Definition 2.20. A simple point process Φ on \mathbb{R}^d with intensity measure Λ is a Poisson point process if

- for all $A \in \mathcal{B}$ with $\Lambda(A) < \infty$, the random variable $\Phi(A)$ has a Poisson distribution with parameter $\mathbb{E}[\Phi(A)] = \Lambda(A)$
- for pairwise disjoint Borel sets $A_1, \dots, A_n \in \mathcal{B}, n \in \mathbb{N}$, the random variables $\Phi(A_1), \dots, \Phi(A_n)$ are independent.

Remark

For a simple point process Φ we write

- $x \in \Phi$ means that the point x belongs to the point process Φ
- $\Phi(B) = n$ means that the set B contains n points of Φ
- Since Φ is a collection of points x_1, x_2, \dots we may also write $\Phi := \{x_n\}$

Definition 2.21. A Polish space M is a separable completely metrizable topological space. Its σ -algebra \mathcal{M} is induced by the topology, i.e. \mathcal{M} is the smallest σ algebra that contains all open subsets of M .

Definition 2.22 (Marked Point Process). Let M be a polish space with its σ -algebra \mathcal{M} . A point process Φ on $\mathbb{R}^d \times M$ is called a marked point process. M is called the mark space. We will write $\Phi = [\{x_n; m_n\}]$, where $\{x_n\}$ is a point process in \mathbb{R}^d and m_n is the mark corresponding to its respective point.

Besides the distribution P_Φ of a stationary point process there is also the Palm distribution P_Φ^0 . The Palm distribution can be interpreted as a point process conditioned that 0 is part of the point process. For a simple stationary point process Φ and a Borel set B we can write

$$P_\Phi^0(S) = \frac{\mathbb{E} \sum_{x \in \Phi} 1_B(x) 1_S(\Phi - x)}{\mathbb{E} \sum_{x \in \Phi} 1_S(x)}.$$

With the Palm distribution we can define the reduced second moment measure

$$\lambda \mathcal{K}(B) = \mathbb{E}^0 [\Phi(B \setminus \{0\})] \quad (2.17)$$

for a stationary point process with intensity λ .

If we set $B = B(0, r)$ we get Ripley's K-function

$$K(r) = \mathcal{K}(B(0, r)).$$

By just looking at the K function all anisotropic behaviour of point process is not visible. We therefore also introduce a version which takes direction into account as well.

For $u \in S^{d-1}$ we can define

$$S(u, \varphi) = \left\{ x \in S^{d-1}, \text{acos}(\langle x, u \rangle) < \varphi \text{ or } \text{acos}(\langle x, -u \rangle) < \varphi \right\}$$

the set of all points on S^{d-1} that have an angle smaller than φ to u or $-u$, which we use to define

$$B(r_1, r_2, u, \varphi) = \left\{ x \in \mathbb{R}^d, r_1 < \|x\| < r_2 \text{ and } \frac{x}{\|x\|} \in S(u, \varphi) \right\},$$

which is a part of an annulus, with all points with a norm between r_1 and r_2 and with angle not greater than φ to u . If we plug $B(0, r, u, \varphi)$ into the reduced second moment measure we have the directed K-function

$$K(r, u, \varphi) = \mathcal{K}(B(0, r, u, \varphi)). \quad (2.18)$$

This function has a preferred direction u and only looks at points that are in a cone with an angle φ around this direction. This can be used to investigate anisotropic point processes.

For a stationary Poisson point process we get

$$\mathcal{K}(B) = \nu(B) \text{ and therefore } K(r) = b_d r^d,$$

where b_d is the volume of the d dimensional unit sphere. For easier interpretation

$$L(r) = \sqrt[d]{\frac{K(r)}{b_d}}$$

is used. This leads to $L(r) = r$ in the stationary Poisson case, deviations from this can be analysed easier. Additionally this transformation stabilises the variance in the estimation of $K(r)$.

For the estimation, we assume to have a stationary point process X with intensity λ in a bounded window W . We use

$$\widehat{\kappa(B)} = \sum_{x_1, x_2 \in W}^{\neq} \frac{1_B(x_2 - x_1)}{\nu(W_{x_1} \cap W_{x_2})}$$

as an unbiased estimator for $\lambda^2 \mathcal{K}(B)$ for a bounded subset B . In this estimator W_x is the window W shifted by the vector x . The reweighing by $\nu(W_{x_1} \cap W_{x_2})$ deals with the bias induced by working with a bounded window W . Since the window is bounded we will see more points with a small distance to each other, for those $\nu(W_{x_1} \cap W_{x_2})$ will be large and their influence is weight down. If the distance between x_1 and x_2 is large, $\nu(W_{x_1} \cap W_{x_2})$ will be small and the impact on the estimator is weight up.

For a point process Φ we use

$$\hat{\lambda}^2 = \frac{\Phi(W) (\Phi(W) - 1)}{\nu(W)^2}$$

to estimate λ^2 . This estimator is unbiased for a stationary Poisson point process. We use

$$\widehat{\mathcal{K}(B)} = \frac{\widehat{\kappa(B)}}{\hat{\lambda}^2} \quad (2.19)$$

as an estimator for the reduced second moment measure. This estimator is ratio unbiased. To get an estimator for Ripley's K -function $\hat{K}(r)$ we use (2.19) and set $B = B(0, r)$.

Remark

Summary statistics like Ripley's K -function are used to find spatial correlation and interaction in a point process. An estimated K -function is usually compared to the K -function of a Poisson point process, since for a Poisson point process case we have total spatial randomness and no point interaction.

If we have a point process where the K function has regions where it is larger than in the Poisson case we say we have attraction between points in these regions. A classical example for this is clustering.

If the K function is smaller than for the Poisson case, we say we have repulsion between points. A classical example for this are hardcore processes. In these each point has a radius r where no other points are located.

2.3.2. Cluster Processes

Let Ψ be a point process. In a cluster process each point $x \in \Psi$ is replaced by a cluster N_x . These clusters are finite point processes. The union of all clusters

$$\Phi = \bigcup_{x \in \Psi} N_x$$

is a cluster process. Note that $x \in N_x$ is not necessarily true.

If the underlying process Ψ is a homogeneous Poisson point process with intensity λ and the clusters are independent from each other we get a so called Neymann-Scott process. It is stationary and if the clusters are isotropic, it is isotropic as well. Its intensity is

$$\lambda_c = \lambda \bar{c}$$

where \bar{c} is the mean number of points in the clusters. For the Palm distribution we get

$$P_0 = P * c_0$$

where c_0 is the Palm distribution of the typical cluster N_0 . Assuming that p_n is the probability that N_0 has exactly n points and $F(r)$ is the CDF of the distance between the points in N_0 by [15, p172] we get

$$\lambda \mathcal{K}(B_r) = b_d r^d + \frac{1}{\lambda \bar{c}} \sum_{n=2}^{\infty} p_n n(n-1) F(r) \quad (2.20)$$

as a closed formula for the reduced second moment measure.

2.3.3. Germ-grain Processes

In this part we will discuss germ-grain processes. It is a marked point process where the mark space \mathcal{K} is a subspace of all compact subsets of \mathbb{R}^d . We will use germ-grain processes to model fibre systems.

We will write the points of the underlying point process with small letters x_i and the marks as K_i . The points of the resulting germ-grain process is written $X_i = K_i + x_i$. Where the typical grain $K_0 \in \mathcal{K}$ is distributed with distribution Q .

Remark

- *Note that the points x_i of the underlying point process are often not observable.*
- *If the underlying point process is stationary or isotropic the germ-grain process is stationary or isotropic*

A germ-grain process is always observed in a bounded window $W \subset \mathbb{R}^d$. We therefore only see a subset of our whole process X . This subset is given by a sampling rule I with $I(X_i) = 1$ if the grain is sampled and $I(X_i) = 0$ if not. The following about sampling bias is cited from Chapter 2 of [9].

Classical sampling rules are

- **plus-sampling:** We keep every object X_i with $X_i \cap W \neq \emptyset$.
- **minus-sampling:** We keep every object X_i with $X_i \subset W$.
- **associated point rule:** We choose a function a that selects a unique point for each object X_i , we keep the object if $a(X_i) \in W$

The quality of these sampling rules can be discussed in the terms of sampling bias.

Theorem 2.13. *Let X_i be a stationary germ-grain process in \mathbb{R}^d with germ intensity λ and compact grains K_i with typical grain K_0 , which is Q distributed. Let $I(X)$ be a measurable sampling rule.*

Then for any translation-invariant, measurable $f : \mathcal{K} \rightarrow \mathbb{R}$

$$\mathbb{E} \left[\sum_i I(X_i) f(X_i) \right] = \lambda \mathbb{E}^0 [f(K_0) \pi(K_0)]$$

with

$$\pi(K) = \int_{\mathbb{R}^d} I(K+x) \, dx$$

for every K . $\pi(K)$ is the volume of the set of all vectors x such that $K+x$ would be included in the sample.

The objects sampled by a sampling rule are reweighed by

$$\frac{\mathbb{E} [\sum_i f(X_i) I(X_i)]}{\mathbb{E} [\sum_i I(X_i)]} = \frac{\mathbb{E}^0 [f(K_0) \pi(K_0)]}{\mathbb{E}^0 [\pi(K_0)]} \quad (2.21)$$

for translation invariant f . This reweighing is only unbiased if $\pi(K_0)$ is independent of K_0 . Also note that the right hand side of (2.21) is the expectation of f under the π weighted distribution with respect to Q . We will call π the bias factor.

Theorem 2.14. *The bias factors for the three sampling rules are*

- For plus-sampling $\pi_+(K_0) = |W \oplus K_0|$.
- For minus-sampling $\pi_-(K_0) = |W \ominus K_0|$.
- For the associated-point rule $\pi_a(K_0) = |W|$.

So only the associated point rule is unbiased. Plus-sampling prefers bigger objects, while minus-sampling leans towards smaller objects.

If we have an unbiased sampling rule, the distribution of the sample matches the original distribution and we can do direct estimation from the sample. If the distribution Q is parametric, we could use ML estimation to estimate the parameters. For the other two we have to use the bias factor to weight the original distribution.

Theorem 2.15 (Horvitz-Thompson Estimator). *Let X be a stationary germ-grain model in \mathbb{R}^d with germ intensity λ and compact grains K_i with typical grain K_0 and distribution Q and. Let $I(X_i)$ be a measurable sampling rule with bias factor $\pi(K)$.*

Then for any translation-invariant, measurable $f : \mathcal{K} \rightarrow \mathbb{R}$

$$\frac{\mathbb{E} \left[\sum_i \frac{f(X_i) I(X_i)}{\pi(X_i)} \right]}{\mathbb{E} \left[\sum_i \frac{I(X_i)}{\pi(X_i)} \right]} = \mathbb{E}^0 [f(K_0)]$$

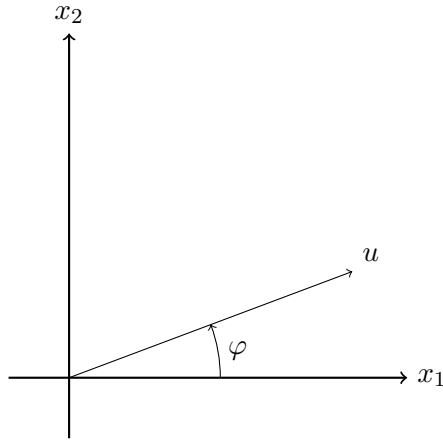


Figure 2.1.: An orientation with u with the corresponding angle φ .

therefore

$$\frac{\sum_i \frac{f(X_i)I(X_i)}{\pi(X_i)}}{\sum_i \frac{I(X_i)}{\pi(X_i)}}$$

is a ratio unbiased estimator for $\mathbb{E}^0 [f(K_0)]$

Another way is to reweigh the complete distribution. Let us say the typical grain K_0 has the distribution Q with the density $f_Q(K; \theta)$ and parameters $\theta \in \mathbb{R}^d$. The biased sample we see in our window then has the distribution Q_I with density $f_{Q_I}(K; \theta) = \frac{\pi(K)}{\mathbb{E}_\theta^0[\pi(K_0)]} f_Q(K; \theta)$. We can use this reweighed density to get an estimator $\hat{\theta}$ for instance with the ML method.

We now have the tools to deal with the sampling bias of the fibres. To estimate the fibre length we might either use Horvitz Thompson (HT) type estimators or we reweighed their density to get an ML estimator.

2.4. Polar and spherical angles and their distributions

To model the fibres we need to look into their orientation $u \in S^{d-1}$. In the 2d case we will identify this direction with an angle $\varphi \in [0, 2\pi]$ with the parametrisation

$$u(\varphi) = \begin{pmatrix} \cos(\varphi) \\ \sin(\varphi) \end{pmatrix}.$$

In this parametrisation the angle φ is between the direction vector of the fibre and the x -axis, see Figure 2.1.

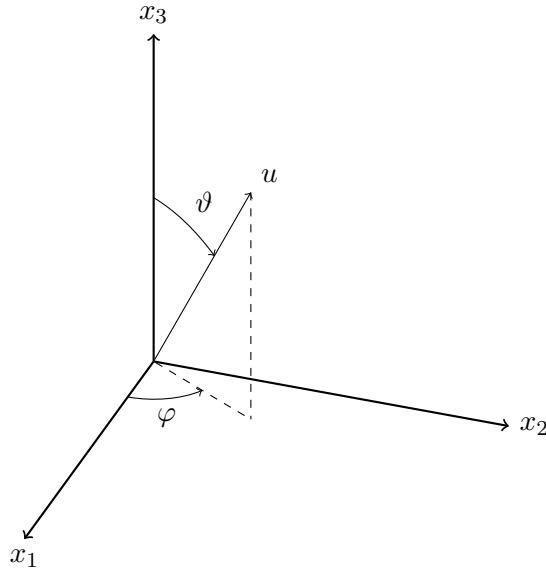


Figure 2.2.: An orientation with u with the corresponding angle φ and ϑ .

In 3d we will use the angles $\varphi \in [0, 2\pi]$ and $\vartheta \in [0, \pi]$ and the parametrisation

$$u(\varphi, \vartheta) = \begin{pmatrix} \sin(\vartheta) \cos(\varphi) \\ \sin(\vartheta) \sin(\varphi) \\ \cos(\vartheta) \end{pmatrix}.$$

In this parametrisation the angle φ is the angle between the orthogonal projection of the direction vector to the $x_1 - x_2$ plane and the x_1 -axis. θ is the angle between the vector and the x_3 -axis, see Figure 2.2.

We will use

$$\begin{aligned} c_1 &= |\cos(\varphi)| & c_1 &= |\sin(\vartheta) \cos(\varphi)| \\ c_2 &= |\sin(\varphi)| & c_2 &= |\sin(\vartheta) \sin(\varphi)| \\ & & c_3 &= |\cos(\vartheta)| \end{aligned}$$

in 2 and 3 dimensions as short hands.

Finding distributions for these angles can be seen as distributions on the unit sphere S^{d-1} . Some known distribution are Mises-Fisher, Watson and angular central Gaussian (ACG) which can be found in [13]. We will look here at the ACG distribution.

Definition 2.23. Let X be a random variable on S^{d-1} for $d \in \mathbb{N}$. Let $\Sigma \in \mathbb{R}^{d \times d}$ be a symmetric and positive definite matrix. X is said to have a d -variate ACG distribution if its density is given by

$$f_X(x; \Sigma) = \alpha_d^{-1} (\det \Sigma)^{-\frac{1}{2}} (x^T \Sigma^{-1} x)^{-\frac{d}{2}}$$

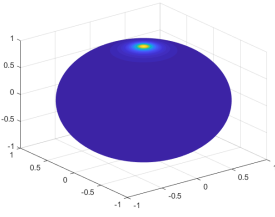


Figure 2.3.: Axial distribution with $\beta = 0.1$

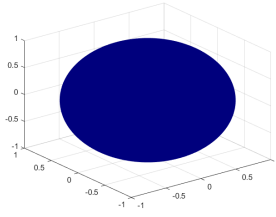


Figure 2.4.: Isotropic distribution with $\beta = 1.0$

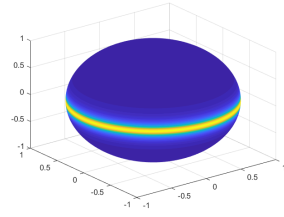


Figure 2.5.: Girdle distribution with $\beta = 10.0$

where α_d is the surface area of S^{d-1} .

This distribution is invariant under rescaling of the Σ , i.e. for a constant $c \in \mathbb{R}$ $c\Sigma$ and Σ will lead to the same distribution.

The ACG distribution is related to a d -variate normal distribution. If X is d -variate normally distributed with mean 0 and covariance matrix Σ then $\frac{X}{\|X\|}$ has a d -variate ACG distribution with parameter Σ . With this we also get a direct sampling method, we sample from a multivariate normal distribution and then normalise the sampled vector.

This distribution family is particularly suited for the modelling in fibres since it is point symmetric. Since fibres usually have no main direction this is a good property to have.

Using a full ACG is cumbersome, since estimating a full covariance matrix is hard. Even if we assume we know the three main directions and only have to estimate a diagonal covariance matrix we still have to estimate d parameters.

One particular distribution that has been used to model the orientation of fibres in [19] is the Schladitz- β distribution with density

$$f_U(\varphi, \theta; \beta) = \frac{1}{4\pi} \frac{\beta \sin(\theta)}{(1 + (\beta^2 - 1) \cos^2(\theta))^{\frac{3}{2}}}, \theta \in [0, \pi), \varphi \in [0, 2\pi).$$

This distribution has only one parameter β . The β in this distribution controls the type and degree of anisotropy. For $\beta = 1$ we get a uniform distribution on the sphere like in Figure 2.4. For $\beta < 1$ we get an axial distribution like in Figure 2.3, where the resulting vectors tend to be aligned with the z -axis. For smaller β we get a bigger degree of alignment. For $\beta > 1$ we get a girdle distribution like in Figure 2.5, where the vectors tend to lie around the equator of the sphere. This distribution was introduced in [19] to model the orientation distribution in a non-woven fibrous material. Estimation methods for the β parameter were studied in [4]. The Schladitz- β distribution is a special case of an ACG distribution with a diagonal covariance matrix $\Sigma = \text{diag}\left(1, 1, \frac{1}{\beta^2}\right)$.

2.5. Fibre Models

In this section we will give an overview over the used fibre models. These are namely a Boolean and a random sequential adsorption (RSA) model. For this we will first give a model for the typical fibre. Then we give two models for the position of these fibres in space. We will model the fibres as a germ-grain model. Where the grains are cylinders or line segments. For the underlying point process we will use a Poisson point process and an RSA process. In the Poisson case fibres might overlap, while the RSA process introduces non overlapping fibres.

2.5.1. Germ Model

One possible model for fibres is the cylinder. A cylinder cyl is characterised by its length $l > 0$, a radius $r > 0$ and an orientation $u \in S^{d-1}$. We will only use cylinders with the midpoint at 0. For a cylinder with a length l , radius r and orientation u we write $\text{cyl}(l, r, u)$. With this we can identify the set of all cylinders

$$S(\text{cyl}) = \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times S^{d-1}.$$

For the grain distribution we model the length L as a real positive random variable with finite mean. The orientation U is modelled as a random variable on S^{d-1} . We will assume that the length and orientation are independent. We assume the radius r to be constant. We will therefore only write $\text{cyl}(l, u)$.

For $L_s \subset \mathbb{R}_{>0}$, $U_s \subset S^{d-1}$ and $S_s(\text{cyl}) = \{\text{cyl}(l, u) \in S(\text{cyl}) | l \in L_s, u \in U_s\}$ we can write

$$\begin{aligned} P(K_0 \in S_s(\text{cyl})) &= P((L, U) \in L_s \times U_s) \\ &= P(L \in L_s)P(U \in U_s) \end{aligned}$$

for the typical cylinder K_0 . With this we can model the individual fibre. If we speak about the position of these we take the midpoint of these cylinders.

Since fibres are usually long compared to their radius we will ignore the cylinder radius in later modelling. This leads to line segments lin , which are also characterised by their length $l > 0$ and orientation $u \in S^{d-1}$. Line segments are cylinders with a radius $r = 0$.

2.5.2. Boolean Model

The first approach for a fibre process is to go for a Boolean model. We therefore model the position of the cylinders as stationary Poisson point process with an intensity λ . This leads to a non interacting fibre model, i.e. the fibres may overlap.

Remark

To simulate the effects of cylinders that enter the window from outside, we propose to use plus-sampling and to simulate in a bigger window W_+ . In the end we only take those parts of the cylinders that are visible in the window W .

We do not advocate the use of periodic boundaries . This will lead to a loss of the independence of the visible objects.

Algorithm 2 Simulating a non interacting fibre system

Choose intensity λ , a window W , and distributions P_L, P_U
 Sample n from a Poisson distribution with parameter $\lambda \cdot \nu(W)$
 Sample n points in W
 Sample n objects and set as marks for the points
return Collection of marked points

The upside of the Boolean model is the simple implementation and fast simulation. Furthermore a lot of closed formulas exist for the Boolean model, for instance the volume fraction is known to be

$$p = 1 - \exp(-\lambda\pi r^2 \mathbb{E}(L)).$$

The downside is that non interacting fibres are not a realistic model. A realisation of this process can be seen in Figure 2.6 on the left..

2.5.3. RSA Fibres

A way to introduce interaction between the fibres, is to use an RSA approach. In this approach we choose a window W and place one object in this window. We then propose another object and check if it intersects with the first one. If it does we choose a new position, otherwise we keep the new object. We repeat this until one of the stopping condition from list 1 is met. Note that option 3 is a fall-back option for the other two. If the algorithm failed too many times to add in a new object the process will stop. Usually this is achieved by giving a maximum number of retries to position the cylinder.

List 1 (Stopping conditions)

1. Stop when n objects have been added
2. Stop when a desired volume fraction has been reached
3. Stop when the window is full, i.e. the proposed object cannot be added to the window-

Remark

- In the RSA case we propose using plus-sampling in window W_+ as well. Additionally, we propose to use periodic edge treatment on the enlarged window, i.e. a cylinder leaving an image on the left, enters it again on the right. Otherwise we will lose stationarity, since the RSA algorithm will then put longer cylinders at the edge of the window W_+ .

- *Keeping the sampled length and orientation is mandatory. If we re-sample a new length with the new position we will bias our data towards short fibres or suitable orientations.*
- *Instead of working with cylinders with flat tops, we check with capsules instead. Since the radius compared to the length is small the error we make is small and the check for the interaction with other objects is faster.*

Algorithm 3 Simulating a non interacting fibre system

Choose a stop criterion from list 1
 Choose a window W , and distributions P_L, P_U
repeat
 Sample a point in W and set point for candidate c
 Sample a cylinder with a length from P_L and an orientation from P_U and set as mark in candidate c
 while c intersects with existing objects **do**
 Sample a new point in W and set for c
 end while
 Add c to collection of accepted points
until Stopping condition is met
return Collection of marked points

The RSA process leads to a non overlapping system of random fibres. It takes exponentially longer the more fibres have to be added. The maximum achievable volume fraction depends strongly on the orientation distribution and the variance and mean of the length distribution. The achievable volume fraction rises the more the fibres are directed in one direction. It also rises with a small variance and mean of the length.

If we have an RSA cylinder process with intensity λ and constant radius r we can calculate the volume fraction as

$$p = \lambda \mathbb{E} [\nu(\text{cyl}(L, U))] = \lambda \pi r^2 \mathbb{E} [L].$$

A realisation of a RSA fibre process can be seen in Figure 2.6 on the right, the fibres do not overlap.

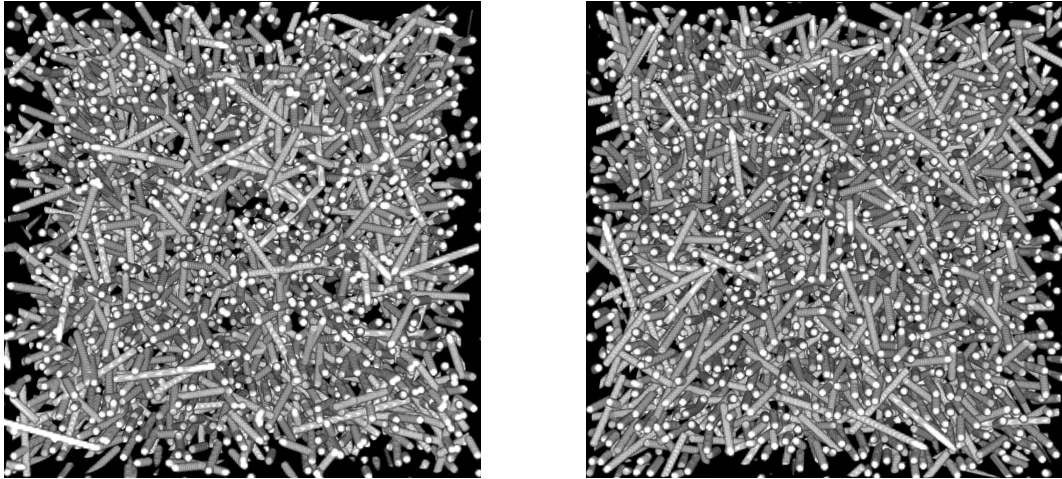


Figure 2.6.: Boolean and RSA fibres with a density of $1.99\text{e}-5$ which leads to volume content of 10% fibres for the RSA case, with a Schladitz β distribution with $\beta = 0.1$ and log normal length distribution with $\mu = 4.6002$ and $\sigma = 0.0998$. The radius is 4. The view is on the $x - y$ plane. The window size is $[0, 500]^3$ for the plus sampling a window of size $[0, 800]^3$ was used. The Boolean model is on the left, the RSA process is on the right.

3. Estimation from fully segmented fibres

In this chapter we will investigate methods to estimate the fibre length distribution using fully segmented fibres.

3.1. Problem

We model the fibre process as a germ grain process. The typical grain is a line segment lin with random length $L > 0$ and orientation $U \in S^{d-1}$. We observe this process in a bounded window $W \subset \mathbb{R}^d$, as seen for 2d in Figure 3.1. We assume that the length L has a parametric distribution with parameters $\theta \in \mathbb{R}^p$ and density $f_L(l; \theta)$.

Segmenting the fibres in this image gives us an i.i.d. sample (\tilde{L}_i^{*+}, Z_i) with $i = 1, \dots, n$ where $Z_i = 1$ means that the fibre is censored and $Z_i = 0$ means it is not.

If we do estimation with this sample we see two problems. The first one is censoring, a lot of the fibres are cut off at the edge of the window. The second one is sampling bias, since the sample of all fibres cutting the window is biased. The underlying sampling rule is plus sampling. We will propose different methods to deal with the sampling bias and censoring based on the EM algorithm and prove their consistency and asymptotic normality.

In the following chapter we will look at examples and give closed formulas for the models used. In these examples we assume the observation windows seen in figure 3.2 has a rectangular or cuboid form. In 2d this leads to $W = [0, W_1] \times [0, W_2]$ and in 3d to $W = [0, W_1] \times [0, W_2] \times [0, W_3]$ where W_1, W_2 and W_3 are the side lengths of the cuboid window in the respective dimension.

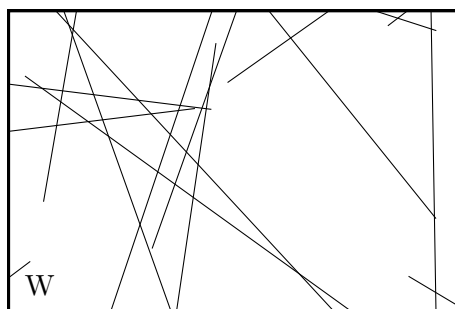


Figure 3.1.: An example for a 2d line segment process.

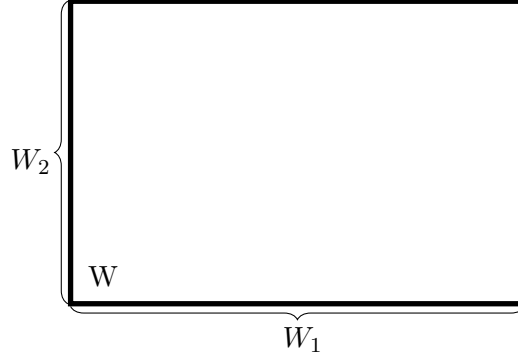


Figure 3.2.: Visualisation of the observation window W .

In this cuboid window we see a cylinder process with a log normally distributed length L with density

$$f_L(l; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}l} \exp\left(-\frac{(\ln(l) - \mu)^2}{2\sigma^2}\right). \quad (3.1)$$

with $\mathbb{E}[\ln(L)] = \mu$ and $\text{var}(\ln(L)) = \sigma^2$. For the orientation distribution on S^2 we use the Schladitz- β distribution with density

$$f_U(\varphi, \theta; \beta) = \frac{1}{4\pi} \frac{\beta \sin(\theta)}{(1 + (\beta^2 - 1) \cos^2(\theta))^{\frac{3}{2}}}, \theta \in [0, \pi), \varphi \in [0, 2\pi)$$

Remark

In the following we will deal with sampling rules I and their the bias factor

$$\pi(l, u) = \int_{\mathbb{R}^d} I(\text{lin}(l, u) + x) \, dx$$

for a line segment $\text{lin}(l, u)$. The bias factor can usually be interpreted as an area of a geometric object. We will use $\pi(l, u)$ for this area and $\Pi(l, u)$ for the geometric object, such that

$$\pi(\text{lin}(l, u)) = |\Pi(\text{lin}(l, u))|.$$

3.2. Horvitz-Thompson type estimator

One classical method to do estimation with a particle process is to use a Horvitz Thompson type estimator. The original idea in [7] was used to deal with sampling bias in the case of drawing from a finite population.

For this approach we will ignore all fibres that are censored. Therefore only fibres that are fully visible in the window W will be used. This leads to an uncensored i.i.d.

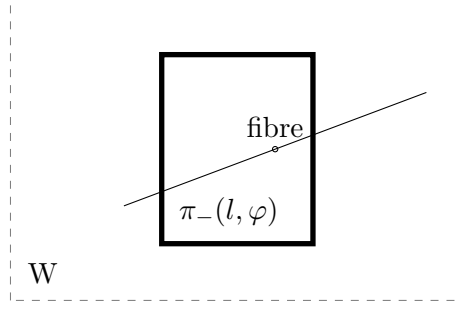


Figure 3.3.: Visualisation of the area $\Pi_-(l, \varphi)$ as \blacksquare , the window W is given as \cdots . Following this sampling rule the example fibre is sampled and fully visible in W .

sample L_i^- with $i = 1, \dots, n$. The sample is biased since using only the uncensored fibres is minus-sampling.

To deal with the minus-sampling we use the bias factor

$$\pi_-(L, U) = |W \ominus \text{lin}(L, U)|,$$

since we assume the direction distribution to be known, we marginalise and get

$$\pi_-(L) = \int_{S^{d-1}} |W \ominus \text{lin}(L, U)| f_U(u) \, du$$

as a weight. With this we get

$$\hat{M}_j = \frac{\sum_{i=1}^n \frac{L_i^j}{\pi_-(L_i)}}{\sum_{i=1}^n \frac{1}{\pi_-(L_i)}}$$

as a ratio unbiased estimator of the j th moment of L .

One problem with this approach is, that we loose a lot of data, especially when the the fibres are long compared to the window size.

HT-estimator 1:

Here we will give closed formulas for the bias factor π_- in our example setting. A visualisation of Π_- can be seen in Figure 3.3. In [20, p247f] we see that that the bias factor can be simplified to

$$\pi_-(l, u) = |W \ominus \text{box}(\text{lin}(l, u))|$$

where box is the axis aligned bounding box of the fibre. The bounding box is given by the cuboid with a diagonal of length l and diagonal orientation

u . Since W is a cuboid, this leads to the closed form

$$\pi_{-}(l, \varphi) = (W_1 - c_1 l)(W_2 - c_2 l)$$

in 2d and

$$\pi_{-}(l, \varphi, \theta) = (W_1 - c_1 l)(W_2 - c_2 l)(W_3 - c_3 l)$$

in 3d.

We marginalise this weight

$$\pi_{-}(l) = \int_{S^{d-1}} |W \ominus \text{lin}(l, u)| f_U(u) \, du,$$

to get the bias factor for the length.

3.3. EM-Estimator

In this section we will propose an estimator that will deal with the censoring by using an EM-algorithm. For this we assume that we use a sampling rule with the bias factor π . The resulting biased sample is L_i^π with $i = 1, \dots, n$ of fibres. We will reweigh the density of L to get a model for L^π and deal with the introduced sampling bias.

Since the sample is also censored we get $(\tilde{L}_i^{*\pi}, Z_i)$ with $i = 1, \dots, n$ of censored lengths and $Z_i = 1$ indicating that the lengths is censored and $Z_i = 0$ indicating it is not. We then use the EM-algorithm to deal with the censoring.

3.3.1. Reweighed Densities

We use the bias factor to reweigh the density of true length distribution f_L and get

$$f_{L^\pi}(l; \theta) = \frac{\pi(l)}{\mathbb{E}_\theta[\pi(L)]} f_L(l; \theta)$$

the density of the biased L^π .

Associated Point Rule 1: Density

Here we will discuss an associated point rule to get an unbiased sub-sample.

We first have to choose a unique and observable associated point.

Definition 3.1. Let W be cuboid window. Let $\text{lin}(l, u)$ be a line segment with $l > 0$ and orientation $u \in S^{d-1}$. The two endpoints of this line segments are given as x and y . We define the associated point as

$$a(\text{lin}(l, u)) := \begin{cases} x, & \text{if } x_1 < y_1 \\ y, & \text{if } x_1 > y_1 \end{cases}.$$

If the first components x_1 and y_1 are the same we compare the next components x_2 and y_2 and so. The case that all components are equal is impossible, since then $x = y$ and the line segment is a point.

We use this point to define the associated point rule

$$I_a(\text{lin}(l, u)) = \begin{cases} 1, & \text{if } a(\text{lin}(l, u)) \in W \\ 0, & \text{otherwise} \end{cases}.$$

If the point $a(\text{lin}(l, u))$ is inside of the window W we keep the fibre in our sample. This point is chosen because we can always decide if a line segment is sampled, even in our censored sample.

Assume we have a line segment $\text{lin}(l, u)$ which is censored, with the endpoints x and y . Since it is censored one of the endpoints will be on the border of the window, assume this endpoint is y . If $a(\text{lin}(l, u)) = x$ we will keep line segment, if $a(\text{lin}(l, u)) = y$ we do not keep it. Since the line segment is straight the true associated point of the uncensored line segment will be outside of the window W .

The canonic choice for the associated point, the midpoint, cannot be observed due to censoring and therefore cannot be used.

We know that the associated point rule is unbiased and that the bias factor is $\pi_a(l, u) = |W|$. A visualisation of the area $\Pi_a(l, u)$ can be seen in Figure 3.4. We therefore get

$$f_L(l; \theta) = \frac{|W|}{\mathbb{E}_\theta[|W|]} f_L(l; \theta) = f_L(l; \theta),$$

as the density of the sample if the associated point rule is used.

Remark

If the used sampling rule is unbiased, the bias factor π is constant and the reweighed density is equal to the old density.

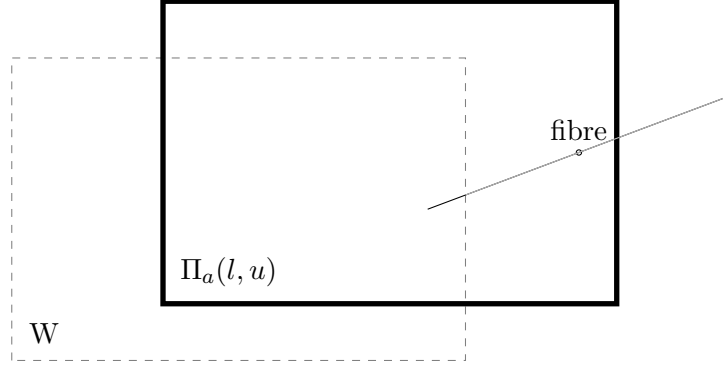


Figure 3.4.: Visualisation of the area $\Pi_a(l, \varphi)$ as **—**, the window W is given as **---**. Following this sampling rule the example fibre is sampled but only the black part will be visible in the window W and the fibre will be censored.

Plus Sampling 1: Density

If we use all fibres visible in the window the sampling rule is plus-sampling. The general formula for the bias factor for a line segment is

$$\pi_+(l, u) = |W \oplus \text{lin}(l, u)|.$$

In Figure 3.5 we see a visualisation of $\Pi_+(l, \varphi)$ for a given angle and length. The area can be split into three parts, the window W and a parallelogram on each side of the window. The parallelograms on the top and bottom have the same area as well as the parallelograms at the sides. We will denote these parallelograms p_1 and p_2 . We conclude

$$\pi_+(l, \varphi) = |W| + 2|p_1| + 2|p_2|.$$

In Figure 3.6 we see a visualisation of the upper parallelogram p_1 . The area is given by

$$|p_1| = W_1 \cdot h$$

where h is the height of the parallelogram. We know that the leg of the parallelogram has a length of $\frac{l}{2}$ and an angle of φ , using basic trigonometry we conclude

$$h = \sin(\varphi) \cdot \frac{l}{2}$$

and therefore

$$|p_1| = W_1 \cdot \sin(\varphi) \cdot \frac{l}{2} = W_1 \cdot c_2 \frac{l}{2}.$$

In Figure 3.7 we see a visualisation of the parallelogram p_2 on the side of

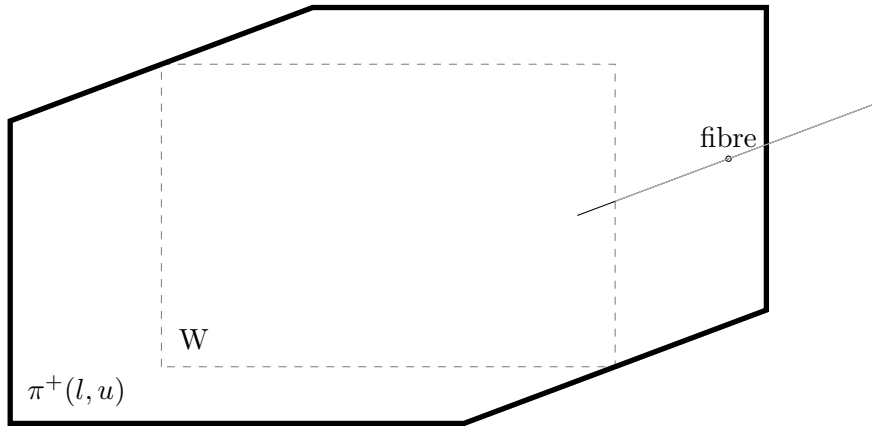


Figure 3.5.: Visualisation of the area $\Pi_+(l, \varphi)$ as **—**, the window W is given as **---**. Following this sampling rule the example fibre is sampled but only the black part will be visible in the window W and the fibre will be censored.

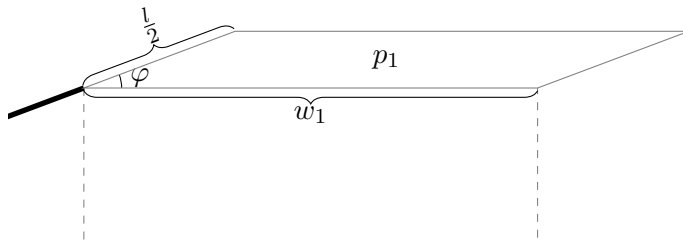


Figure 3.6.: Visualisation of the area p_1 as **—**, the window W is given as **---**.

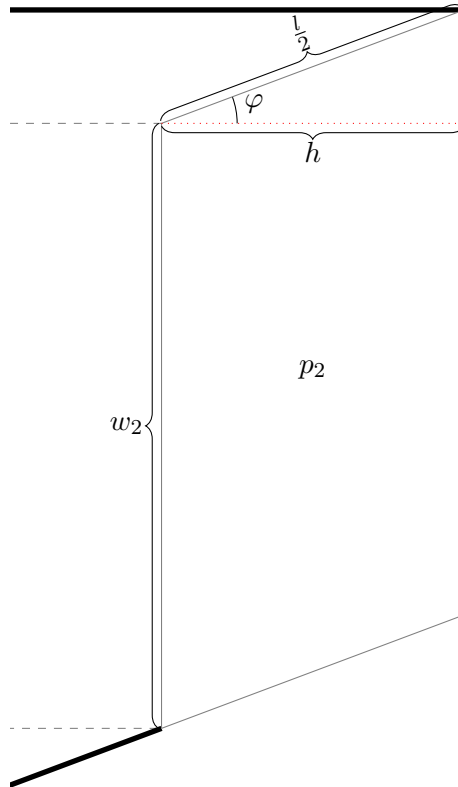


Figure 3.7.: Visualisation of the area p_2 as —, the window W is given as ----.

the window W . In this case the height is given by

$$h = \cos(\varphi) \cdot \frac{l}{2}$$

and the base has the length W_2 . We conclude

$$|p_2| = W_2 \cdot \cos(\varphi) \cdot \frac{l}{2} = W_2 \cdot c_2 \cdot \frac{l}{2}.$$

We get

$$\boxed{\pi_+(l, \varphi) = |W| + 2|p_1| + 2|p_2| = |W| + W_1 c_2 l + W_2 c_1 l}$$

for the complete bias factor.

In 3d the parallelograms on the sides of W turn into parallelepipeds denoted by $p_{i,j}$ with $i, j = 1, 2, 3$ and $i \neq j$. The parallelepiped $p_{i,j}$ has the base on the W_i, W_j face of the window. To calculate each of the six volumes we have to calculate the height of the parallelepipeds.

We look at the face with the side lengths W_1 and W_2 the height of the corresponding parallelepiped is then given by the z component of the direction vector: $|\cos(\vartheta)|$ times the length $\frac{l}{2}$. The volume of the parallelepiped is then given by

$$|p_{1,2}| = W_1 W_2 \cdot \cos(\vartheta) \frac{l}{2} = W_1 W_2 \cdot c_3 \frac{l}{2}.$$

Analogue computations lead to the volumes of the other parallelepipeds

$$\begin{aligned} |p_{1,3}| &= W_1 W_3 \cdot \sin(\vartheta) \sin(\varphi) \frac{l}{2} = W_1 W_3 \cdot c_2 \frac{l}{2} \\ |p_{2,3}| &= W_2 W_3 \cdot \sin(\vartheta) \cos(\varphi) \frac{l}{2} = W_2 W_3 \cdot c_1 \frac{l}{2}. \end{aligned}$$

In conclusion we get

$$\boxed{\begin{aligned} \pi_+(l, \varphi, \theta) &= |W| + 2|p_{1,2}| + 2|p_{2,3}| + 2|p_{1,3}| \\ &= |W| + W_1 W_2 \cdot c_3 l + W_1 W_3 \cdot c_2 l \\ &\quad + W_2 W_3 \cdot c_1 l \end{aligned}}$$

in 3 dimension.

We can now use the reweighed density

$$f_{L^+}(l) = \frac{\pi_+(l)}{\mathbb{E}[\pi_+(L)]} f_L(l; \theta)$$

with

$$\pi_+(l) = \int_{S^{d-1}} \pi_+(l, u) f_U(u) \, du$$

to deal with the sampling bias. We note that $\pi_+(l, u)$ and $\pi_+(l)$ are linear in l . For some $\xi \in \mathbb{R}$ we may write

$$\pi_+(l) = \xi l + |W|.$$

In conclusion we get

$$f_{L^+}(l) = \frac{\xi l + |W|}{\xi \exp\left(\mu + \frac{\sigma^2}{2}\right) + |W|} f_L(l; \theta) \quad (3.2)$$

as a reweighed density in the case of plus-sampling.

We now have a density for the biased sample L_i^π . We can use this density to do maximum likelihood estimation to receive an unbiased estimator for the parameter θ . We will now prove that if the density f_L fulfils the properties of theorem 2.8 and f_{L^π} fulfils some mild additional conditions, the resulting ML estimator is asymptotically normal and consistent.

Theorem 3.1. *Assume we have a line segment process observed in a bounded window W with random length L with density $f_L(l, \theta)$. The sampled line segments are chosen according to a sampling rule I with bias factor $\pi(l)$. This results in a sample L_i^π with $i = 1, \dots, n$ and an reweighed density*

$$f_{L^\pi}(l, \theta) = \frac{\pi(l)}{\mathbb{E}[\pi(L)]} f_L(l, \theta).$$

let

$$\hat{\theta} = \arg \max_{\theta} \sum_{i=1}^n \ln(f_{L^\pi}(l, \theta))$$

be the ML estimator.

If $f_L(l; \theta)$ fulfils the properties of theorem 2.8 and

R-7 $\mathbb{E}[\pi(L)]$ is three times continuous differentiable in an open subset $\Theta_0 \subset \Theta$.

R-8 There are functions $g_i(l)$ and $g_{ij}(l)$ with

$$\begin{aligned} |\pi(l) \frac{\partial}{\partial \theta_i} f(l; \theta)| &\leq g_i(l) \\ |\pi(l) \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(l; \theta)| &\leq g_{ij}(l) \end{aligned}$$

and $g_i, g_{ij} \in L^1$.

R-9 The Hessian matrix $\mathcal{H}(\ln(\mathbb{E}[\pi(L)]))$ is positive semi definite

Then the ML estimator $\hat{\theta}$ is consistent and asymptotically normal distributed with

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(L^\pi)^{-1}).$$

Proof. We have to prove the conditions **R-1** to **R-6** from theorem 2.8.

R-1 To prove that the density is distinct we assume that $\theta \neq \theta'$ but $f_{L^\pi}(l; \theta) = f_{L^\pi}(l; \theta')$. We get

$$\begin{aligned} \frac{\pi(l)}{\mathbb{E}_\theta[L]} f_L(l; \theta) &= \frac{\pi(l)}{\mathbb{E}_{\theta'}[L]} f_L(l; \theta') \\ \Rightarrow \frac{1}{\mathbb{E}_\theta[L]} f_L(l; \theta) &= \frac{1}{\mathbb{E}_{\theta'}[L]} f_L(l; \theta'). \end{aligned}$$

Since $f_L(l; \theta) \neq f_L(l; \theta')$ by assumption, we conclude $\mathbb{E}_\theta[L] \neq \mathbb{E}_{\theta'}[L]$ has to hold for above equation to be true. We further get

$$f_L(l; \theta) = \frac{\mathbb{E}_\theta[L]}{\mathbb{E}_{\theta'}[L]} f_L(l; \theta').$$

If we integrate both sides we get have

$$\begin{aligned} 1 &= \frac{\mathbb{E}_\theta[L]}{\mathbb{E}_{\theta'}[L]} \\ \Rightarrow \mathbb{E}_\theta[L] &= \mathbb{E}_{\theta'}[L] \end{aligned}$$

which is a contradiction. We conclude that the reweighed density is distinct.

R-2 $f_{L^\pi}(l, \theta)$ has common support for all θ since it has the same support as $f_L(l; \theta)$, which has common support for all θ .

R-3 Since $f_L(l; \theta)$ fulfils **R-3** we have an open subset Θ_0 with $\theta_0 \in \Theta_0$. The existence of all third derivatives follows since they exist for $f_L(l; \theta)$ and **R-7**.

R-4 If we calculate the first derivative of the reweighed density we get

$$\begin{aligned} \left| \frac{\partial}{\partial \theta_i} f_{L^\pi}(l, \theta) \right| &= \left| \frac{1}{\mathbb{E}[\pi(L)]} \pi(l) \frac{\partial}{\partial \theta_i} f_L(l; \theta) - \frac{\frac{\partial \mathbb{E}[\pi(L)]}{\partial \theta_i}}{\mathbb{E}[\pi(L)]^2} \pi(l) f_L(l; \theta) \right| \\ &\leq \frac{1}{|\mathbb{E}[\pi(L)]|} \left| \pi(l) \frac{\partial}{\partial \theta_i} f_L(l; \theta) \right| + \left| \frac{\frac{\partial \mathbb{E}[\pi(L)]}{\partial \theta_i}}{\mathbb{E}[\pi(L)]^2} \right| |\pi(l) f_L(l; \theta)| \end{aligned}$$

using condition **R-8** yields

$$\left| \frac{\partial}{\partial \theta_i} f_{L^\pi}(l, \theta) \right| \leq \frac{1}{|\mathbb{E}[\pi(L)]|} g_i(l) + \left| \frac{\frac{\partial \mathbb{E}[\pi(L)]}{\partial \theta_i}}{\mathbb{E}[\pi(L)]^2} \right| g_0(l)$$

and since condition **R-7** holds $\frac{1}{|\mathbb{E}[\pi(L)]|}$ and $\left| \frac{\frac{\partial \mathbb{E}[\pi(L)]}{\partial \theta_i}}{\mathbb{E}[\pi(L)]^2} \right|$ are continuous and take a unique maximum on the parameter space. Therefore we can construct majorants for all first derivatives.

We get

$$\begin{aligned} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_{L^\pi}(l, \theta) \right| &= \left| \frac{1}{\mathbb{E}[\pi(L)]} \pi(l) \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_L(l; \theta) - \frac{\frac{\partial \mathbb{E}[\pi(L)]}{\partial \theta_j}}{\mathbb{E}[\pi(L)]^2} \pi(l) f(l; \theta) \right. \\ &\quad \left. - \frac{\partial}{\partial \theta_j} \left(\frac{\frac{\partial \mathbb{E}[\pi(L)]}{\partial \theta_i}}{\mathbb{E}[\pi(L)]^2} \right) \pi(l) f_L(l; \theta) - \frac{\frac{\partial \mathbb{E}[\pi(L)]}{\partial \theta_i}}{\mathbb{E}[\pi(L)]^2} \pi(l) \frac{\partial}{\partial \theta_j} f(l; \theta) \right| \\ &\leq \left| \frac{1}{\mathbb{E}[\pi(L)]} \right| g_{ij}(l) + \left| \frac{\frac{\partial \mathbb{E}[\pi(L)]}{\partial \theta_j}}{\mathbb{E}[\pi(L)]^2} \right| g_0(l) \\ &\quad + \left| \frac{\partial}{\partial \theta_j} \left(\frac{\frac{\partial \mathbb{E}[\pi(L)]}{\partial \theta_i}}{\mathbb{E}[\pi(L)]^2} \right) \right| g_0(l) + \left| \frac{\frac{\partial \mathbb{E}[\pi(L)]}{\partial \theta_i}}{\mathbb{E}[\pi(L)]^2} \right| g_j(l) \end{aligned}$$

for the second derivatives since condition **R-8** holds. We can therefore construct majorants for the first and second derivatives and may switch the order of integration of differentiation and integration two times.

R-5 The fisher information matrix for the reweighed density is

$$\begin{aligned} \mathcal{I}(L^\pi) &= -\mathbb{E}[\mathcal{H}(\ln(\pi(l))) - \mathcal{H}(\ln(\mathbb{E}[\pi(L)])) + \mathcal{H}(\ln(f_L(l; \theta)))] \\ &= \mathbb{E}[\mathcal{H}(\ln(\mathbb{E}(\pi(L))))] - \mathbb{E}[\mathcal{H}(\ln(f_L(l; \theta)))] \\ &= \mathcal{H}(\ln(\mathbb{E}(\pi(L)))) - \mathbb{E}[\mathcal{H}(\ln(f_L(l; \theta)))] . \end{aligned}$$

It is positive definite since $-\mathbb{E}[\mathcal{H}(\ln(f_L(l; \theta)))]$ is positive definite since $f_L(l; \theta)$ fulfils condition **R-5** and $\mathcal{H}(\ln(\mathbb{E}(\pi(L))))$ is positive semi definite by condition **R-9**. Since the sum of a positive definite matrix and a positive semi definite matrix is positive definite, the fisher information of the reweighed density is positive definite.

R-6 By assumption we have functions $M_{ijk}(l)$ such that

$$\left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln f_L(l; \theta) \right| \leq M_{ijk}(l)$$

with $\mathbb{E}_{\theta_0} [M_{ijk}(L)] < \infty$ for $i, j, k = 1, \dots, d$.

For the reweighed density we get

$$\begin{aligned} \left| \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} \ln f_{L^\pi}(l; \theta) \right| &= \left| \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} \ln \pi(l) - \ln \mathbb{E}(\mathbb{E}(\pi(L))) + \ln f_{L^\pi}(l; \theta) \right| \\ &\leq \left| \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} \mathbb{E}_\theta(\pi(L)) \right| + \left| \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} \ln f_L(l; \theta) \right| \\ &\leq \left| \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} \mathbb{E}_\theta(\pi(L)) \right| + M_{ijk}(l). \end{aligned}$$

The first term in this formula does not depend on θ and is furthermore a continuous function in θ by assumption **R-7**. Since the parameter space is compact there exists a unique maximum m_{ijk} such that

$$\left| \frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k} \ln f_{L^\pi}(l; \theta) \right| \leq m_{ijk} + M_{ijk}(l) =: \tilde{M}_{ijk}(l)$$

Therefore the ML estimator in regards to the reweighed density f_{L^π} is consistent and asymptotically normal. \square

Associated Point Rule 2: Consistency and asymptotic normality

In our example for the associated point rule we have seen, that the density of the sample is the density of the log-normal distribution. We have shown that ML estimators based on this density are consistent and asymptotically normal in chapter 2. We conclude that the estimator based on the associated point rule will be consistent and asymptotically normal.

Plus Sampling 2: Consistency and asymptotic normality

Here we will prove that the estimator based on the reweighed density for the plus sampled line segments are consistent and asymptotically normal.

Theorem 3.2. *Let*

$$f_{L^+}(l; \theta) = \frac{\xi l + |W|}{\xi \exp\left(\mu + \frac{\sigma^2}{2}\right) + |W|} f_L(l; \theta)$$

for some $\xi \in \mathbb{R}$ be the reweighed density for the plus-sampling case. Then the conditions of theorem 3.1 are fulfilled and the resulting ML estimator is consistent and asymptotically normal.

Proof. R-7 We have

$$\mathbb{E}(\pi^+(L)) = \xi \mathbb{E}(L) + |W|.$$

The mean of a log normal distribution is given by $\mathbb{E}(L) = \exp(\mu + \frac{\sigma^2}{2})$. This function is smooth and therefore all third derivatives exist.

R-8 To construct the majorants for the derivatives we look at

$$\begin{aligned} |\pi(l) \frac{\partial}{\partial \mu} f_L(l; \mu, \sigma)| &= \left| (\xi l + |W|) \frac{\ln(l) - \mu}{\sigma^2} f_L(l; \mu, \sigma) \right| \\ &= \left| \frac{\xi l \ln(l) - \xi l \mu + |W| \ln(l) - \|W\| \mu}{\sigma^2} f_L(l; \mu, \sigma) \right| \\ &\leq \frac{\xi l |\ln(l)| + \xi l |\mu| + |W| |\ln(l)| + \|W\| \mu}{|\sigma^2|} f_L(l; \mu, \sigma) \end{aligned}$$

similarly to the proof for the log normal distribution we can put in the maximum or minimum allowed values for μ and σ . The only part left to majorize is

$$l |\ln(l)|^k f_L(l; \mu, \sigma) \leq l |\ln(l)|^k g_0(l)$$

with $g_0(l)$ is from Equation (2.5). Similarly to the log normal case we get

$$\begin{aligned} \int_{-\infty}^{\infty} l |\ln(l)|^k g_0(l) \, dl &\propto \int_0^{\infty} \exp(z) z^k \frac{1}{\sqrt{2\pi\sigma_{max}^2}} \exp\left(-\frac{(z - \mu_{max})^2}{2\sigma_{max}^2}\right) \\ &\quad + \int_0^{\infty} \exp(-z) z^k \frac{1}{\sqrt{2\pi\sigma_{max}^2}} \exp\left(-\frac{(z - \mu_{max})^2}{2\sigma_{max}^2}\right) \end{aligned}$$

calculating these integral using Matlab Symbolic Toolbox we see that they are finite for $k = 1, 2, 3, 4$.

Constructing the majorants for the other derivatives can be done in a similar fashion.

R-9 We look at the Hessian

$$\begin{aligned} \mathcal{H}(\ln(\mathbb{E}(\pi(L)))) &= \mathcal{H}(\ln(\xi \exp(\mu + \frac{\sigma^2}{2}) + |W|)) \\ &= \begin{pmatrix} \frac{\xi |W| \exp(\mu + \frac{\sigma^2}{2})}{(\exp(\mu + \frac{\sigma^2}{2}) + |W|)^2} & \frac{\xi |W| \sigma \exp(\mu + \frac{\sigma^2}{2})}{(\xi \exp(\mu + \frac{\sigma^2}{2}) + |W|)^2} \\ \frac{\xi |W| \sigma \exp(\mu + \frac{\sigma^2}{2})}{(\xi \exp(\mu + \frac{\sigma^2}{2}) + |W|)^2} & \frac{\xi \exp(\mu + \frac{\sigma^2}{2}) (\xi \exp(\mu + \frac{\sigma^2}{2}) + |W|) + |W| \sigma^2}{(\exp(\mu + \frac{\sigma^2}{2}) + |W|)^2} \end{pmatrix} \end{aligned}$$

Calculating the first principle minors yields

$$\begin{aligned}\mathcal{H}(\ln(\mathbb{E}(\pi(L))))_{1,1} &= \frac{c|W| \exp(\mu + \frac{\sigma^2}{2})}{\left(\exp(\mu + \frac{\sigma^2}{2}) + |E|\right)^2} > 0 \\ \det \mathcal{H}(\ln(\mathbb{E}(\pi(L)))) &= \frac{c^2|W| \exp(2\mu + \sigma^2)}{\left(c \exp\left(\mu + \frac{\sigma^2}{2}\right) + |W|\right)^3} > 0\end{aligned}$$

both of them are bigger than 0 and the Silvester criterion yields that the matrix is positive definite. \square

3.3.2. EM-Algorithm

Here we look at an i.i.d. sample $(\tilde{L}_i^{\pi*}, Z_i)$ with $i = 1, \dots, n$ where $Z_i = 1$ means that the fibre is censored and $Z_i = 0$ means it is not. They are sampled with sampling rule I with bias factor $\pi(l)$. We use the reweighing described in 3.3.1 to get a density

$$f_{L^\pi}(l; \theta) = \frac{\pi(l)}{\mathbb{E}[\pi(l)]} f_L(l; \theta)$$

for the distribution of L^π .

We group the sample $\tilde{L}_i^{\pi*}$ in an uncensored group L_i^π $i = 1, \dots, n_1$ and a censored group \tilde{L}_i^π $i = 1, \dots, n_2$ with $n = n_1 + n_2$. We use the EM-algorithm from algorithm 1 to get an ML estimator of the parameter θ .

We get

$$\begin{aligned}Q(\theta|\theta_0) &= \mathbb{E}_{\theta_0} \left[l(\theta|L^\pi, \tilde{L}^{\pi*}) \right] \\ &= \sum_{i=1}^n \mathbb{E}_{\theta_0} \left[\log f_L(L_i^\pi; \theta) | \tilde{L}_i^{\pi*} \right] \\ &= \sum_{i=1}^{n_1} \log f_{L^\pi}(L_i^\pi; \theta) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} \left[\log f_{L^\pi}(L_i^\pi; \theta) | \tilde{L}_i^\pi \right].\end{aligned}\tag{3.3}$$

We can interpret the summands $\mathbb{E}_{\theta_0}[\log f_{L^\pi}(L_i^\pi) | \tilde{L}_i^\pi]$ as the value of $\log f_{L^\pi}(L_i^\pi)$ for true length of a fibre given the censored length, as seen in Figure 3.8. Since for the uncensored fibres the true length is already known we can drop the conditional mean for the uncensored fibres.

To maximise $Q(\theta|\theta_0)$ we need to calculate the conditional mean $\mathbb{E}_{\theta_0}(\log f_{L^\pi}(L_i^\pi) | \tilde{L}_i^\pi)$ in each step. We will therefore need the distribution of the true length L^π given some censored length \tilde{L}^π .

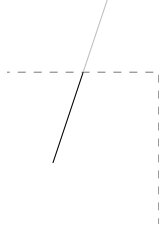


Figure 3.8.: The EM algorithm uses the visible part of the fibre inside the window and "adds" some length to it to get an estimate of the real length

Associated Point Rule 3: EM Algorithm

With the associated point rule we have a sample \tilde{L}_i^* $i = 1, \dots, n$ with L_i $i = 1, \dots, n_1$ of uncensored and \tilde{L}_i $i = 1, \dots, n_2$ of censored lengths and $n = n_1 + n_2$ samples overall.

To get an EM estimator we use Equation (3.3). We have to maximise

$$\begin{aligned}
 Q(\theta|\theta_0) &= \mathbb{E}_{\theta_0}[l(\theta|L)|\tilde{L}] \\
 &= \sum_{i=1}^{n_1} \ln f_L(L_i; \theta) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0}[\ln f_L(L_i; \theta)|\tilde{L}_i] \\
 &= \sum_{i=1}^{n_1} -\ln(L_i) - \ln(\sqrt{2\pi}\sigma) + \left(-\frac{(\ln(L_i) - \mu)^2}{2\sigma^2} \right) \\
 &+ \sum_{i=1}^{n_2} -\mathbb{E}_{\theta_0}[\ln(L_i)|\tilde{L}_i] + \ln(\sqrt{2\pi}\sigma) + \left(-\frac{\mathbb{E}_{\theta_0}[(\ln(L_i) - \mu)^2|\tilde{L}_i]}{2\sigma^2} \right)
 \end{aligned}$$

and overall

$$\begin{aligned}
 Q(\theta|\theta_0) &= \sum_{i=1}^{n_1} -\ln(L_i) - \ln(\sqrt{2\pi}\sigma) + \left(-\frac{(\ln(L_i) - \mu)^2}{2\sigma^2} \right) \\
 &+ \sum_{i=1}^{n_2} -\mathbb{E}_{\theta_0}[\ln(L_i)|\tilde{L}_i] + \ln(\sqrt{2\pi}\sigma) \\
 &+ \left(-\frac{\mathbb{E}_{\theta_0}[\ln(L_i)^2|\tilde{L}_i] - \mu \mathbb{E}_{\theta_0}[\ln(L_i)|\tilde{L}_i] + \mu^2}{2\sigma^2} \right)
 \end{aligned} \tag{3.4}$$

for μ and σ using $\theta = (\mu, \sigma)$ and $\theta_0 = (\mu_0, \sigma_0)$.

Plus Sampling 3: EM Algorithm

For this estimator we use all visible fibres and have a sample \tilde{L}_i^{*+} for

$i = 1, \dots, n$. We assume we see n_1 uncensored fibres with length L_i^+ and n_2 censored fibres with visible length \tilde{L}_i^+ , with $n = n_1 + n_2$.

We use the reweighed density from Equation (3.2). Where the bias factor

$$\pi_+(l) = \xi l + |W|.$$

is linear in l . We take the reweighed density and use it for an EM estimator. For this we have to iterate with the function

$$\begin{aligned} Q(\theta|\theta_0) &= \sum_{i=1}^{n_1} \ln f_{L^+}(L_i^+; \theta) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} \left[\ln f_{L^+}(L_i^+; \theta) | \tilde{L}_i^+ \right] \\ &= \sum_{i=1}^{n_1} \ln(\xi L_i^+ + |W|) - \ln(\xi e^{\mu + \frac{\sigma^2}{2}} + |W|) + \ln(f(L_i^+; \theta)) + \\ &\quad \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} \left[\ln(\xi L_i^+ + |W|) | \tilde{L}_i^+ \right] - \ln(\xi e^{\mu + \frac{\sigma^2}{2}} + |W|) \\ &\quad + \mathbb{E}_{\theta_0} \left[\ln(f_L(L_i^+; \theta)) | \tilde{L}_i^+ \right] \end{aligned} \quad (3.5)$$

with $\theta = (\mu, \sigma)$ and $\theta^0 = (\mu^0, \sigma^0)$. We plug in the log normal density $f_L(l; \mu, \sigma)$ and get

$$\begin{aligned} Q(\theta|\theta_0) &= -n_1 \ln(\xi e^{\mu + \frac{\sigma^2}{2}} + |W|) - n_1 \ln(\sqrt{2\pi}\sigma) \\ &\quad + \sum_{i=1}^{n_1} \ln(\xi l_i + |W|) - \ln(L_i) + \left(-\frac{(\ln(L_i) - \mu)^2}{2\sigma^2} \right) \\ &\quad + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} \left[\ln(\xi L_i^+ + |W|) | \tilde{L}_i^+ \right] - \mathbb{E}_{\theta_0} \left[\ln(L_i^+) | \tilde{L}_i^+ \right] \\ &\quad + \left(-\frac{\mathbb{E}_{\theta_0} \left[\ln(L_i^+)^2 | \tilde{L}_i^+ \right] - \mu \mathbb{E}_{\theta_0} \left[\ln(L_i^+) | \tilde{L}_i^+ \right] + \mu^2}{2\sigma^2} \right) \end{aligned} \quad (3.6)$$

as the function we iterate for the EM estimation.

3.3.3. Distribution of $L^\pi | \tilde{L}^\pi$

We want to deduce the distribution for the true length L^π given a censored length $\tilde{L}^\pi = \tilde{l}$. By following the Bayesian theorem we get

$$f_{L^\pi | \tilde{L}^\pi}(\tilde{l}; \theta) = \frac{f_{\tilde{L}^\pi | L^\pi}(\tilde{l} | l) f_{L^\pi}(l; \theta)}{f_{\tilde{L}^\pi}(\tilde{l}; \theta)}. \quad (3.7)$$

for the density. The density $f_{L^\pi}(l; \theta)$ is already known by model assumption. We can get

$$f_{\tilde{L}^\pi}(\tilde{l}) = \int_{\mathbb{R}} f_{\tilde{L}^\pi|L^\pi}(\tilde{l}|l) f_{L^\pi}(l; \theta) dl$$

by marginalisation if $f_{\tilde{L}^\pi|L^\pi}(\tilde{l}|l)$ is known. We therefore need the distribution of the censored length \tilde{L}^π given the true length $L^\pi = l$.

3.3.4. Distribution of $\tilde{L}^\pi|L^\pi$

We wish to get the probability $P(\tilde{L}^\pi < \tilde{l}|L^\pi)$. For this we will investigate the areas where a fibre might be located to be completely visible and then we will look into areas where it is censored and the visible length is less then \tilde{l} .

Assume we have the sampling rule I with bias factor π . We investigate a line segment lin with length l and orientation u in a window W . We first look at the area where the fibre has to be located to be visible in the window W . This is depended on the sampling rule and is the bias factor

$$w(l, u) = \pi(\text{lin}(l, u)).$$

For the fibre to be completely visible it has to be located in

$$v(l, u) = \pi_-(\text{lin}(l, u)) = |W \ominus \text{lin}(l, u)|,$$

since this is minus sampling. We know that a fibre cannot be located in this area, since it is censored. Therefore a censored line segment with length l and orientation u has to be located in the area $w(l, u) - v(l, u)$.

A line segment with true length l and a visible length of more then \tilde{l} will be in an area of

$$a(l, \tilde{l}, u) = |\pi(\text{cyl}(l, u)) \ominus \text{box}(\text{cyl}(\tilde{l}, u))|. \quad (3.8)$$

We can therefore conclude that

$$P(\tilde{L}^\pi \leq \tilde{l}|L^\pi = l, U = u) = \frac{w(l, u) - a(l, \tilde{l}, u)}{w(l, u) - v(l, u)}. \quad (3.9)$$

with the density

$$f_{\tilde{L}^\pi|L, U}(\tilde{l}|l, u) = \frac{-\frac{\partial}{\partial \tilde{l}} a(l, \tilde{l}, u)}{w(l, u) - v(l, u)}. \quad (3.10)$$

We marginalise the orientation to get

$$P(\tilde{L}^\pi \leq \tilde{l}|L^\pi = l) = \int_{S^{d-1}} P(\tilde{L}^\pi \leq \tilde{l}|L^\pi = l, U = u) f_U(u) du$$

$$f_{\tilde{L}^\pi|L}(\tilde{l}|l) = \int_{S^{d-1}} f_{\tilde{L}^\pi|L, U}(\tilde{l}|l, u) f_U(u) du.$$

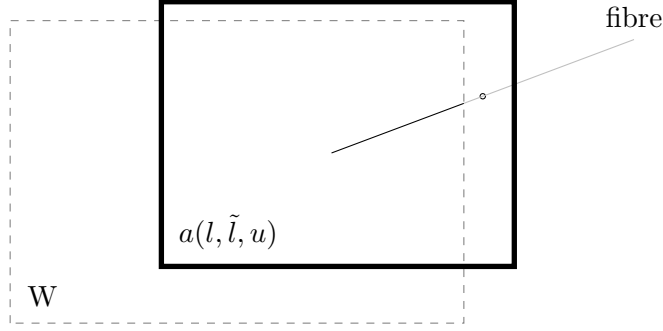


Figure 3.9.: Visualisation of the area $a(l, \tilde{l}, \varphi)$ as \blacksquare , the window W is given as \cdots .
In this case we have $\tilde{l} = \frac{l}{3}$.

With this we can calculate the density

$$f_{L^\pi | \tilde{L}^\pi}(l | \tilde{l}; \theta) = \frac{f_{\tilde{L}^\pi | L^\pi}(\tilde{l} | l) f_{L^\pi}(l; \theta)}{f_{\tilde{L}^\pi}(\tilde{l}; \theta)}.$$

which we will in turn use to calculate the conditional means

$$\mathbb{E}_{\theta_0}[\log f_{L^\pi}(L^\pi; \theta) | \tilde{L}^\pi = \tilde{l}] = \int_{\mathbb{R}} \log f_{L^\pi}(l; \theta) f_{L^\pi | \tilde{L}^\pi}(l | \tilde{l}; \theta_0) dl$$

for the EM algorithm.

Associated Point Rule 4: Distribution of $\tilde{L} | L$

Here we will give the formula for $w(l, u)$, $v(l, u)$ and $a(l, \tilde{l}, u)$ for the associated point rule.

For the bias factor we have

$$w(l, u) = |W|$$

and

$$v(l, u) = \pi_-(\text{cyl}(l, u)) = \begin{cases} (W_1 - c_1 l)(W_2 - c_2 l)(W_3 - c_3 l) & \text{if } W_i - c_i l > 0 \forall i \\ 0 & \text{else} \end{cases}$$

directly. So we have to deduce a formula for $a(l, \tilde{l}, u)$ as seen in Figure 3.9, the area where a cylinder of length l and orientation u might be located such that we see at least a cylinder of length \tilde{l} in the window.

For the area $a(l, \tilde{l}, u)$ from Equation (3.8) we see that in this case we get

$$a(l, \tilde{l}, u) = |W \ominus \text{lin}(\tilde{l}, u)|.$$

this is just like the formula for $\pi_-(\tilde{l}, u)$ and we get

$$\begin{aligned} a(l, \tilde{l}, \varphi) &= (W_1 - c_1 \tilde{l})(W_2 - c_2 \tilde{l}) \\ a(l, \tilde{l}, \varphi, \vartheta) &= (W_1 - c_1 \tilde{l})(W_2 - c_2 \tilde{l})(W_3 - c_3 \tilde{l}) \end{aligned}$$

in 2 and 3 dimensions where $(W_i - c_i \tilde{l}) \geq 0$ for $i = 1, 2, 3$.

We now have the formulae for $w(l, u)$, $v(l, u)$ and $a(l, \tilde{l}, u)$. In conclusion we have

$$f_{\tilde{L}|L,U}(\tilde{l}|l, u) = \frac{c_1(W_2 - c_2 \tilde{l})(W_3 - c_3 \tilde{l}) + (W_1 - c_1 \tilde{l})c_2(W_3 - c_3 \tilde{l}) + (W_1 - c_1 \tilde{l})(W_2 - c_2 \tilde{l})c_3}{|W| - (W_1 - c_1 l)(W_2 - c_2 l)(W_3 - c_3 l)}$$

the density of $\tilde{L}|L, U$. It is a polynomial of order 2 in \tilde{l} . Note that this also holds for the marginalised density

$$f_{\tilde{L}^\pi|L}(\tilde{l}|l) = \int_{S^2} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, u) f_U(u) \, du.$$

Remark

A simple sanity check for the correctness of these formulae is to look at the limiting cases of $a(l, \tilde{l}, u)$ for \tilde{l} . The case of $\tilde{l} = 0$ means that we at least see something of the fibre, so we expect $a(l, 0, u) = \pi_a(l, u)$ which is true since

$$a(l, 0, \varphi) = (W_1 - c_1 0)(W_2 - c_2 0) = W_1 W_2 = \pi_a(l, u).$$

The case of $\tilde{l} = l$ means that we see the whole fibre, we expect $a(l, l, u) = \pi_-(l, u)$ which is true since

$$a(l, l, \varphi) = (W_1 - c_1 l)(W_2 - c_2 l) = \pi_-(l, u).$$

Remark

The condition $(w_i - c_i \tilde{l}) \geq 0$ for $i = 1, 2, 3$ is not only a condition for \tilde{l} it also bounds the allowed orientation.

If we look at the case $i = 3$ we get

$$\begin{aligned} W_3 - |\cos(\vartheta)|\tilde{l} &\geq 0 \\ |\cos(\vartheta)| &\leq \frac{W_3}{\tilde{l}} \\ \Rightarrow \cos^{-1}\left(\frac{W_3}{\tilde{l}}\right) &\leq \vartheta \leq \pi - \cos^{-1}\left(\frac{W_3}{\tilde{l}}\right) \end{aligned} \quad (3.11)$$

if $\frac{W_3}{\tilde{l}} \leq 1$ and no constrictions otherwise. If we look at $i = 1, 2$ we have

$$\begin{aligned} W_1 - |\sin(\vartheta) \cos(\varphi)|\tilde{l} &\geq 0 \\ W_2 - |\sin(\vartheta) \sin(\varphi)|\tilde{l} &\geq 0 \\ \Rightarrow \frac{W_1}{\tilde{l}} &\geq |\sin(\vartheta) \cos(\varphi)| \\ \frac{W_2}{\tilde{l}} &\geq |\sin(\vartheta) \sin(\varphi)| \end{aligned}$$

now we square both inequalities and sum them up, we get

$$\begin{aligned} \frac{W_1^2 + W_2^2}{\tilde{l}^2} &\geq \sin(\vartheta)^2(\cos(\varphi)^2 + \sin(\varphi)^2) \\ \frac{\sqrt{W_1^2 + W_2^2}}{\tilde{l}} &\geq |\sin(\vartheta)|. \end{aligned}$$

Note that $\sqrt{W_1^2 + W_2^2}$ is the diagonal in the ground plane of the window. If this diagonal is longer than \tilde{l} we have no further condition on ϑ . If \tilde{l} is bigger than this diagonal we get the conditions

$$\begin{aligned} \vartheta &\leq \sin^{-1} \left(\frac{\sqrt{W_1^2 + W_2^2}}{\tilde{l}} \right) \text{ or} \\ \vartheta &\geq \pi - \sin^{-1} \left(\frac{\sqrt{W_1^2 + W_2^2}}{\tilde{l}} \right) \end{aligned}$$

If we take these conditions and the one from Equation (3.11), we have the full conditions on ϑ as

$$\begin{aligned} c_{\vartheta,l,1}(\tilde{l}) &:= \cos^{-1} \left(\frac{W_3}{\tilde{l}} \right) \leq \vartheta \leq \sin^{-1} \left(\frac{\sqrt{W_1^2 + W_2^2}}{\tilde{l}} \right) =: c_{\theta,u,1}(\tilde{l}) \text{ or} \\ c_{\vartheta,l,2}(\tilde{l}) &:= \pi - \cos^{-1} \left(\frac{W_3}{\tilde{l}} \right) \geq \vartheta \geq \pi - \sin^{-1} \left(\frac{\sqrt{W_1^2 + W_2^2}}{\tilde{l}} \right) =: c_{\theta,2,u}(\tilde{l}) \end{aligned} \quad (3.12)$$

For φ we look at $i = 1, 2$ again and get

$$\begin{aligned} W_1 - |\sin(\vartheta) \cos(\varphi)|\tilde{l} &\geq 0 \\ \frac{W_1}{\sin(\vartheta)} &\geq \cos(\varphi) \\ \Rightarrow \varphi &\geq \cos^{-1} \left(\frac{W_1}{\sin(\vartheta)} \right) \end{aligned}$$

and

$$\begin{aligned}
W_2 - \sin(\vartheta) \sin(\varphi) \tilde{l} &\geq 0 \\
\frac{W_2}{\sin(\vartheta) \tilde{l}} &\geq \sin(\varphi) \\
\Rightarrow \varphi &\leq \sin^{-1} \left(\frac{W_2}{\sin(\vartheta) \tilde{l}} \right)
\end{aligned}$$

gives a condition for each quadrant

$$\begin{aligned}
c_{\varphi,l,1}(\tilde{l}, \vartheta) &:= \cos^{-1} \left(\frac{W_1}{\sin(\vartheta) \tilde{l}} \right) \leq \varphi \leq \sin^{-1} \left(\frac{W_2}{\sin(\vartheta) \tilde{l}} \right) =: c_{\varphi,u,1}(\tilde{l}, \vartheta) \text{ or} \\
c_{\varphi,l,2}(\tilde{l}, \vartheta) &:= \pi - \sin^{-1} \left(\frac{W_2}{\sin(\vartheta) \tilde{l}} \right) \leq \varphi \leq \pi - \cos^{-1} \left(\frac{W_1}{\sin(\vartheta) \tilde{l}} \right) =: c_{\varphi,u,2}(\tilde{l}, \vartheta) \text{ or} \\
c_{\varphi,l,3}(\tilde{l}, \vartheta) &:= \pi + \cos^{-1} \left(\frac{W_1}{\sin(\vartheta) \tilde{l}} \right) \leq \varphi \leq \pi + \sin^{-1} \left(\frac{W_2}{\sin(\vartheta) \tilde{l}} \right) =: c_{\varphi,u,3}(\tilde{l}, \vartheta) \text{ or} \\
c_{\varphi,l,4}(\tilde{l}, \vartheta) &:= 2\pi - \sin^{-1} \left(\frac{W_2}{\sin(\vartheta) \tilde{l}} \right) \leq \varphi \leq 2\pi - \cos^{-1} \left(\frac{W_1}{\sin(\vartheta) \tilde{l}} \right) =: c_{\varphi,u,4}(\tilde{l}, \vartheta)
\end{aligned} \tag{3.13}$$

Plus Sampling 4: Distribution of $\tilde{L}^+ | L^+$

We need formulae for the areas $w(l, u)$, $v(l, u)$ and $a(l, \tilde{l}, u)$. Since $w(l, u) = \pi_+(l, u)$ we already know this formula. The formula for $v(l, u) = \pi_-(l, u)$ is also known. The area $a(l, \tilde{l}, u)$ is the area where the midpoint of a fibre with length l and orientation u might be located such that at least a fibre of length \tilde{l} is visible in the window W in regards to the plus sampling rule.

For this we have to look at two cases. The first case is $\tilde{l} < \frac{l}{2}$ in Figure 3.10. We see that we have two parallelograms P_1 at the top and bottom and P_2 at the two sides of the window W . Furthermore we see that that two triangles T that belong to the observation window are outside of $a(l, \tilde{l}, u)$. We conclude that

$$a(l, \tilde{l}, \varphi) = |W| - 2|T| + 2|P_1| + 2|P_2|.$$

To calculate the area of the triangle we look at Figure 3.11. First we notice that a fibre with length l and orientation u placed on the border of $a(l, \tilde{l}, u)$ is, by construction of a visible with a length of \tilde{l} in the observation window. Therefore the part of the border of $a(l, \tilde{l}, u)$ that crosses through the window has to have a length of \tilde{l} , this part of the border is also the hypotenuse of the triangle T . Since we know the orientation of the fibre we

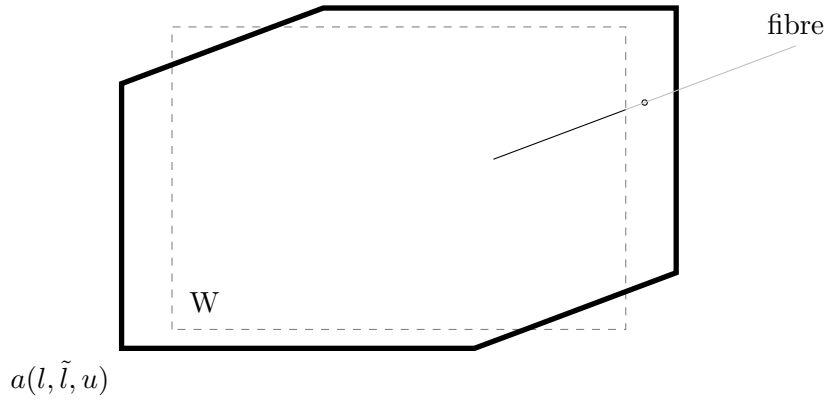


Figure 3.10.: Visualisation of the area $a(l, \tilde{l}, u)$ as --- , the window W is given as --- .
For this case $\tilde{l} = \frac{l}{3}$.

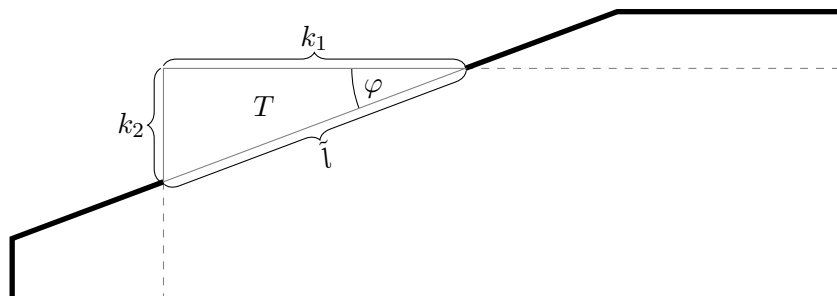


Figure 3.11.: Visualisation of the triangle T from $a(l, \tilde{l}, u)$ as --- , the window W is given as --- and $a(l, \tilde{l}, u)$ as --- . For this case $\tilde{l} = \frac{l}{3}$.

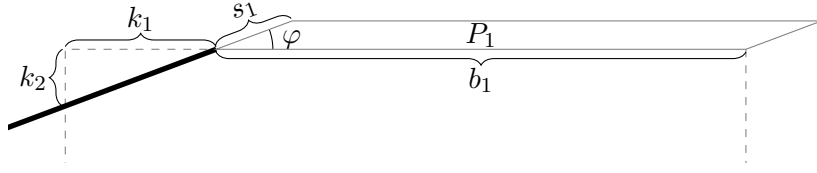


Figure 3.12.: Visualisation of the parallelogram P_1 from $a(l, \tilde{l}, u)$ as --- , the window W is given as --- and $a(l, \tilde{l}, u)$ as --- . For this case $\tilde{l} = \frac{l}{3}$.

also know the angles in the triangle. We can therefore calculate the catheti as

$$k_1 = |\cos(\varphi)|\tilde{l} = c_1\tilde{l}k_2 = |\sin(\varphi)|\tilde{l} = c_2\tilde{l}$$

and the overall area of the triangle is

$$|T| = \frac{1}{2}k_1k_2 = \frac{1}{2}c_1c_2\tilde{l}^2.$$

To calculate the area of P_1 we look at Figure 3.12. We see directly that the base of the parallelogram is

$$b_1 = W_1 - k_1 = W_1 - c_1\tilde{l}.$$

The length of the leg s_1 is not directly apparent. We place a fibre on the border in Figure 3.13. Therefore the part of the fibre that extends into $a(l, \tilde{l}, u)$ has the length $\frac{l}{2}$ and the part of the fibre that is inside W has, by construction of $a(l, \tilde{l}, u)$, the length \tilde{l} , therefore

$$s_1 = \frac{l}{2} - \tilde{l}.$$

In the end we conclude that

$$\begin{aligned} |P_1| &= |\sin(\varphi)|s_1 \cdot b_1 \\ &= c_2\left(\frac{l}{2} - \tilde{l}\right) \cdot (W_1 - c_1\tilde{l}). \end{aligned}$$

The other parallelogram on the side of W can be constructed the same way as

$$|P_2| = c_1\left(\frac{l}{2} - \tilde{l}\right) \cdot (W_2 - c_2\tilde{l}).$$

We conclude that

$$\begin{aligned} a(l, \tilde{l}, \varphi) &= |W| - 2|T| + 2|P_1| + 2|P_2| \\ &= W_1W_2 - c_1c_2\tilde{l}^2 + 2c_2\left(\frac{l}{2} - \tilde{l}\right) \cdot (W_1 - c_1\tilde{l}) + 2c_1\left(\frac{l}{2} - \tilde{l}\right) \cdot (W_2 - c_2\tilde{l}) \end{aligned}$$

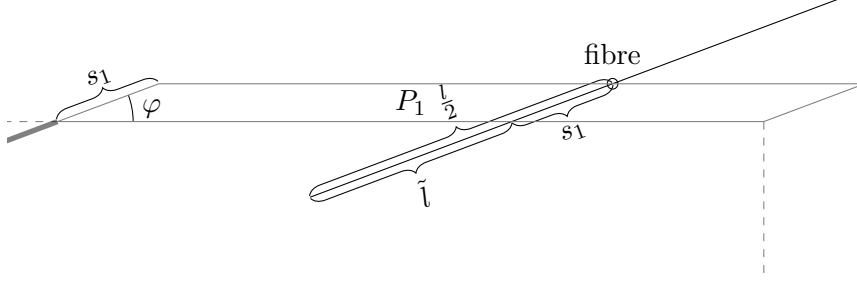


Figure 3.13.: Visualisation of the parallelogram p_1 from $a(l, \tilde{l}, u)$ as --- , the window W is given as --- and $a(l, \tilde{l}, u)$ as --- . A fibre with length l has been placed on the border of $a(l, \tilde{l}, u)$. The midpoints of the fibre is marked by \circ For this case $\tilde{l} = \frac{l}{3}$.

for the case that $\tilde{l} < \frac{l}{2}$.

In 3 dimensions the parallelograms turn into parallelepipeds, as they did for $\pi_+(l, u)$. The triangles T , turn into prisms T_i for $i = 1, 2, 3$. For $i = 1$ it has a triangular base with area $\frac{1}{2}c_2c_3\tilde{l}^2$ and the height W_1 we get

$$|T_1| = \frac{1}{2}c_2c_3\tilde{l}^2W_1$$

for its volume. The volumes of $T_i, i = 2, 3$ are constructed similarly. In conclusion we get

$$\begin{aligned} a(l, \tilde{l}, \varphi, \theta) = & W_1W_2W_3 - c_1c_2\tilde{l}^2W_3 - c_1c_3\tilde{l}^2W_2 - c_2c_3\tilde{l}^2W_1 \\ & + 2c_3\left(\frac{l}{2} - \tilde{l}\right) \cdot (W_1 - c_1\tilde{l}) \cdot (W_2 - c_2\tilde{l}) \\ & + 2c_2\left(\frac{l}{2} - \tilde{l}\right) \cdot (W_1 - c_1\tilde{l}) \cdot (W_3 - c_3\tilde{l}) \\ & + 2c_1\left(\frac{l}{2} - \tilde{l}\right) \cdot (W_2 - c_2\tilde{l}) \cdot (W_3 - c_3\tilde{l}) \end{aligned} \quad (3.14)$$

for the 3d case.

For the case $\tilde{l} > \frac{l}{2}$ we will start with the 2d case again, an example can be seen in Figure 3.14. We see that $a(l, \tilde{l}, u)$ resembles a rectangle with two triangles missing. We will first calculate the area of the inner rectangle and we will then subtract the area of the triangles.

In Figure 3.15 we placed a fibre on the upper edge of $a(l, \tilde{l}, u)$. We see that the distance to the border of W is

$$d_2 = \sin(\varphi)h_2.$$

To calculate h_2 we notice that the full visible length is \tilde{l} by construction. The part below the midpoint of the fibre is $\frac{l}{2}$ long, therefore

$$h_2 = \tilde{l} - \frac{l}{2}$$

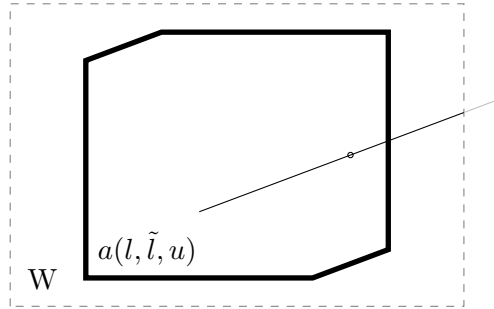


Figure 3.14.: Visualisation of the area $a(l, \tilde{l}, u)$ as **—**, the window W is given as **----**.
 For this case $\tilde{l} = \frac{2l}{3}$.

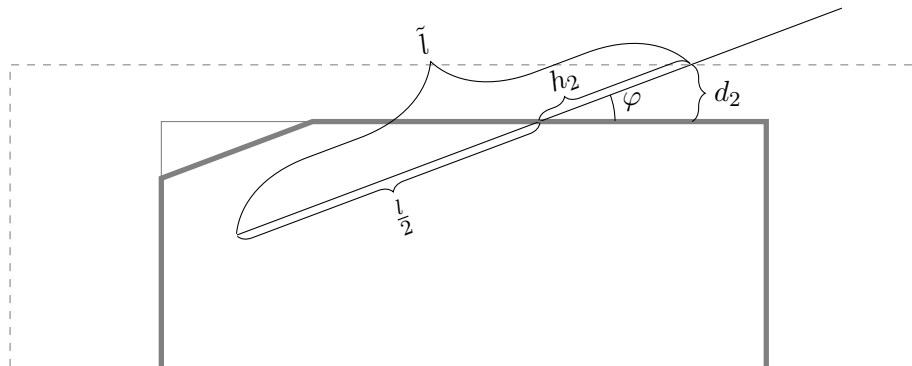


Figure 3.15.: Visualisation of the area inner rectangle as **—**, the window W is given as **----** and. For this case $\tilde{l} = \frac{2l}{3}$.

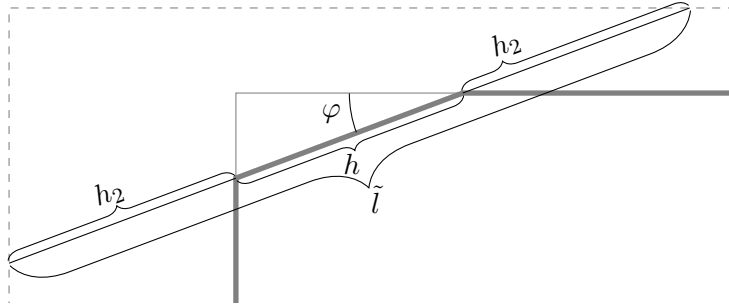


Figure 3.16.: Visualisation of the inner rectangle as —, the window W is given as --- and. For this case $\tilde{l} = \frac{2l}{3}$.

and

$$d_2 = c_2 \left(\tilde{l} - \frac{l}{2} \right).$$

With the same reasoning we get that the distance to the sides of W as

$$d_1 = c_1 \left(\tilde{l} - \frac{l}{2} \right).$$

We now look at the triangle in Figure 3.16. To calculate the hypotenuse h we place a line segment somewhere on the hypotenuse. The visible length of this fibre is again \tilde{l} , the distances h_2 have been calculated in the previous step. We get

$$h = \tilde{l} - 2h_2 = l - \tilde{l}.$$

The area of this triangle is therefore

$$|T| = \frac{1}{2} c_1 c_2 h^2 = \frac{1}{2} c_1 c_2 (l - \tilde{l})^2.$$

In conclusion we get

$$\begin{aligned} a(l, \tilde{l}, u) &= (W_1 - 2d_1)(W_2 - 2d_2) - 2|T| \\ &= (W_1 - 2c_1 \left(\tilde{l} - \frac{l}{2} \right))(W_2 - 2c_2 \left(\tilde{l} - \frac{l}{2} \right)) - c_1 c_2 (l - \tilde{l})^2. \end{aligned}$$

In 3 dimension we get similarly

$$\begin{aligned}
a(l, \tilde{l}, u) = & \left(W_1 - 2c_1 \left(\tilde{l} - \frac{l}{2} \right) \right) \left(W_2 - 2c_2 \left(\tilde{l} - \frac{l}{2} \right) \right) \left(W_3 - 2c_3 \left(\tilde{l} - \frac{l}{2} \right) \right) \\
& - c_1 c_2 \left(l - \tilde{l} \right)^2 \left(W_3 - 2c_3 \left(\tilde{l} - \frac{l}{2} \right) \right) \\
& - c_1 c_3 \left(l - \tilde{l} \right)^2 \left(W_2 - 2c_2 \left(\tilde{l} - \frac{l}{2} \right) \right) \\
& - c_2 c_3 \left(l - \tilde{l} \right)^2 \left(W_1 - 2c_1 \left(\tilde{l} - \frac{l}{2} \right) \right)
\end{aligned}
\tag{3.15}$$

The full density is therefore given by

$$\begin{aligned}
f_{\tilde{L}^+ | L^+, \varphi, \vartheta} &= \frac{-\frac{\partial}{\partial \tilde{l}} a(l, \tilde{l}, \varphi, \vartheta)}{\pi^+(l)} \\
&= \begin{cases} \frac{2(W_1 - c_1 \tilde{l})(W_2 - c_2 \tilde{l})c_3 + 2(W_1 - c_1 \tilde{l})c_2(W_3 - c_3 \tilde{l}) + 2c_1(W_2 - c_2 \tilde{l})(W_3 - c_3 \tilde{l})}{(W_1 + c_1 l)(W_2 + c_2 l)(W_3 + c_3 l)} + \frac{4c_1 c_2 (W_3 - c_3 \tilde{l})(\frac{l}{2} - \tilde{l}) + 4c_1 (W_2 - c_2 \tilde{l})c_3(\frac{l}{2} - \tilde{l}) + 4(W_1 - c_1 \tilde{l})c_2 c_3(\frac{l}{2} - \tilde{l})}{(W_1 + c_1 l)(W_2 + c_2 l)(W_3 + c_3 l)} & \tilde{l} < \frac{l}{2} \\ \frac{+2W_1 c_2 c_3 \tilde{l} + 2c_1 W_2 c_3 \tilde{l} + 2W_1 c_2 c_3 \tilde{l}}{2(W_1 - 2c_1(\tilde{l} - \frac{l}{2}))(W_2 - 2c_2(\tilde{l} - \frac{l}{2}))c_3 + 2(W_1 - 2c_1(\tilde{l} - \frac{l}{2}))c_2(W_3 - 2c_3(\tilde{l} - \frac{l}{2}))} & \\ \frac{2c_1(W_2 - 2c_2(\tilde{l} - \frac{l}{2}))(W_3 - 2c_3(\tilde{l} - \frac{l}{2}))}{(W_1 + c_1 l)(W_2 + c_2 l)(W_3 + c_3 l)} & \tilde{l} > \frac{l}{2} \end{cases}
\end{aligned}$$

Remark

Using these we can get conditions on the angles ϑ and φ again. We first look at the case $\tilde{l} < \frac{l}{2}$ and Equation (3.14). To make sure the the sides all have positive lengths we demand

$$W_i - c_i \tilde{l} \geq 0$$

for $i = 1, 2, 3$. These are the same conditions that lead to $c_{\theta, j}(\tilde{l}), c_{\varphi, k}(\tilde{l}, \theta)$ in Equations (3.12) and (3.13) for the associated point rule case.

For $\tilde{l} > \frac{l}{2}$ we look at Equation (3.15) and get

$$W_i - 2c_i \left(\tilde{l} - \frac{l}{2} \right) \geq 0$$

for $i = 1, 2, 3$ as conditions for the orientations ϑ and φ . By doing the same calculations as we did to get Equations (3.12) and (3.13) for the associated

point rule we get

$$\begin{aligned}
c_{\vartheta,l,1}(\tilde{l}, l) &:= \cos^{-1} \left(\frac{W_3}{2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \leq \vartheta \leq \sin^{-1} \left(\frac{\sqrt{W_1^2 + W_2^2}}{2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \\
&=: c_{\vartheta,u,1}(\tilde{l}, l) \\
c_{\vartheta,u,2}(\tilde{l}, l) &:= \pi - \cos^{-1} \left(\frac{W_3}{2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \leq \vartheta \leq \pi - \sin^{-1} \left(\frac{\sqrt{W_1^2 + W_2^2}}{2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \\
&=: c_{\vartheta,l,2}(\tilde{l}, l)
\end{aligned}$$

as conditions for ϑ and the conditions

$$\begin{aligned}
c_{\varphi,l,1}(\tilde{l}, l, \vartheta) &:= \cos^{-1} \left(\frac{W_1}{\sin(\vartheta) 2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \leq \varphi \\
&\leq \sin^{-1} \left(\frac{W_2}{\sin(\vartheta) 2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \\
&=: c_{\varphi,u,1}(\tilde{l}, l, \vartheta) \\
c_{\varphi,l,2}(\tilde{l}, l, \vartheta) &:= \pi - \sin^{-1} \left(\frac{W_2}{\sin(\vartheta) 2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \leq \varphi \\
&\leq \pi - \cos^{-1} \left(\frac{W_1}{\sin(\vartheta) 2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \\
&=: c_{\varphi,u,2}(\tilde{l}, l, \vartheta)
\end{aligned}$$

$$\begin{aligned}
c_{\varphi,l,3}(\tilde{l}, l, \vartheta) &:= \pi + \cos^{-1} \left(\frac{W_1}{\sin(\vartheta)2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \leq \varphi \\
&\leq \pi + \sin^{-1} \left(\frac{W_2}{\sin(\vartheta)2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \\
&:= c_{\varphi,u,3}(\tilde{l}, l, \vartheta) \\
c_{\varphi,l,4}(\tilde{l}, l, \vartheta) &:= 2\pi - \sin^{-1} \left(\frac{W_2}{\sin(\vartheta)2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \leq \varphi \\
&\leq 2\pi - \cos^{-1} \left(\frac{W_1}{\sin(\vartheta)2 \left(\tilde{l} - \frac{l}{2} \right)} \right) \\
&:= c_{\varphi,u,4}(\tilde{l}, l, \vartheta)
\end{aligned}$$

for φ .

3.3.5. Properties of the estimator

If the EM described in Equation (3.4) converges we will obtain the ML estimator for the parameters θ . The ML estimation is in respect to the sample of censored fibres $(\tilde{L}_i^{\pi^*}, Z_i)$.

The density for \tilde{L}^{π^*} is

$$f_{\tilde{L}^{\pi^*}}(\tilde{l}; \theta) = \left(\int_{\tilde{l}}^{\infty} f_{\tilde{L}^{\pi}|L^{\pi}}(\tilde{l}, l) f_{L^{\pi}}(l; \theta) dl \right) Z_i + (1 - Z_i) \cdot f_{L^{\pi}}(\tilde{l}; \theta). \quad (3.16)$$

Note that this density has a bounded support. For a window W we say that W_{max} is longest straight line that can be drawn inside of W . It is impossible to observe a lengths that is greater than this, therefore the density $f_{\tilde{L}^{\pi^*}}(\tilde{l}; \theta)$ has the support $(0, W_{max}]$

We have to prove the properties **R-1** to **R-6** from theorem 2.8 to proof the asymptotic normality of the resulting EM estimator.

Theorem 3.3. *Assume we have a line segment process in W where the length is distributed with density $f_L(l; \theta)$ and the orientation is distributed with density f_U . We observe this process in a bounded window W with sampling rule I and bias factor π . Let W_{max} be the longest straight line in W . We observe $\tilde{L}_i^{\pi^*}, Z_i$ for $i = 1, \dots, n$ with a density from Equation (3.16).*

*Assume that $f_L(l; \theta)$ fulfils the properties **R-1** to **R-6** from theorem 2.8 and $f_{L^{\pi}}(l; \theta)$ fulfils the properties **R-7** to **R-9** from theorem 3.1. Furthermore we need*

R-10 For all $\tilde{l} \in (0, W_{max}]$ we have a $c_{max}(\tilde{l})$ such that for all $l \in [\tilde{l}, \infty)$ we have $f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) \leq c_{max}(\tilde{l}) < \infty$

R-11 The function

$$f_{L^\pi}(\tilde{l}, \theta) = \left(\int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) f_{L^\pi}(l, \theta) dl \right) \leq c < \infty$$

is bounded for $\tilde{l} \in (0, W_{max}]$ and some $c \in \mathbb{R}$.

R-12 For all $\tilde{l} \in (0, W_{max}]$, $n = 1, 2, 3$ and $i_1 + \dots + i_d = n$ we have a constant c

$$\frac{\left| \frac{\partial^d}{\partial^{i_1} \theta_1 \dots \partial^{i_d} \theta_d} f_{\tilde{L}^\pi}(\tilde{l}) \right|}{f_{\tilde{L}^\pi}(\tilde{l})} = \frac{\left| \frac{\partial^d}{\partial^{i_1} \theta_1 \dots \partial^{i_d} \theta_d} \left(\int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) f_{L^\pi}(l) dl \right) \right|}{\int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) f_{L^\pi}(l) dl} \leq c < \infty$$

R-13 The fisher information matrix $\mathcal{I}(\tilde{L})$ is positive definite.

then the ML estimator $\hat{\theta}$ using the density $f_{\tilde{L}^{\pi*}}(\tilde{l}; \theta)$ is consistent and asymptotically normal.

Proof. We have to proof that the regularity conditions from theorem 2.8.

R-1 We take two parameter vectors θ and θ' . We now look at $Z = 0$, the case for an uncensored fibre and investigate

$$f_{L^\pi}(l, \theta) = f_{L^\pi}(l, \theta'),$$

since f_{L^π} is distinct by assumption this can only be true if $\theta = \theta'$. For the other case we look at

$$\begin{aligned} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) f_{L^\pi}(l, \theta) dl &= \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) f_{L^\pi}(l, \theta') dl \\ \Rightarrow \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) (f_{L^\pi}(l, \theta) - f_{L^\pi}(l, \theta')) dl &= 0. \end{aligned}$$

Since this has to be true for all \tilde{l} it follows that the integrand has to be 0 and we get

$$\begin{aligned} f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) (f_{L^\pi}(l, \theta) - f_{L^\pi}(l, \theta')) &= 0 \\ \Rightarrow (f_{L^\pi}(l, \theta) - f_{L^\pi}(l, \theta')) &= 0 \end{aligned}$$

which can only be true for $\theta = \theta'$ since f_{L^π} is distinct. Therefore $f_{\tilde{L}^\pi}(\tilde{l}, \theta)$ is distinct.

R-2 The support of $f_{\tilde{L}^{\pi*}}(\tilde{l}, \theta)$ is $(0, W_{max}]$ where W_{max} is the longest straight line that can be drawn in the window W . This does not depend on the parameter θ .

R-3 To proof that the first three derivatives of $f_{\tilde{L}^{\pi*}}(\tilde{l})$ exist, we look at

$$\begin{aligned} \frac{\partial^n}{\partial^{j_1}\theta_1 \dots \partial^{j_d}\theta_d} f_{\tilde{L}^{\pi*}}(\tilde{l}) &= Z \frac{\partial^n}{\partial^{j_1}\theta_1 \dots \partial^{j_d}\theta_d} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) f_{L^\pi}(l, \theta) dl \\ &\quad + (1 - Z) \frac{\partial^n}{\partial^{j_1}\theta_1 \dots \partial^{j_d}\theta_d} f_{L^\pi}(l, \theta). \end{aligned}$$

We see that if we could switch the differentiation and the integral in this equation the existence of the third derivatives follows since $\frac{\partial^n}{\partial^{j_1}\theta_1 \dots \partial^{j_d}\theta_d} f_{L^\pi}(l, \theta)$ exists by assumption. To show that we can switch differentiation and integration 3 times for $f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) f_{L^\pi}(l, \theta)$ we have to find majorants. The majorants are given by

$$\begin{aligned} \left| \frac{\partial^n}{\partial^{j_1}\theta_1 \dots \partial^{j_d}\theta_d} f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) f_{L^\pi}(l, \theta) \right| &\stackrel{\mathbf{R-10}}{\leq} c_{max}(\tilde{l}) \left| \frac{\partial^n}{\partial^{j_1}\theta_1 \dots \partial^{j_d}\theta_d} f_{L^\pi}(l, \theta) \right| \\ &\stackrel{\mathbf{R-4}}{\leq} c_{max}(\tilde{l}) g_{i_1, \dots, i_d}(l) \end{aligned}$$

with $1 \leq n \leq 3$ and $i_1 + \dots + i_d = n$. Where $g_{i_1, \dots, i_d}(l)$ is a majorant for $\frac{\partial^n}{\partial^{j_1}\theta_1 \dots \partial^{j_d}\theta_d} f_{L^\pi}(l, \theta)$ such that

$$\left| \frac{\partial^n}{\partial^{j_1}\theta_1 \dots \partial^{j_d}\theta_d} f_{L^\pi}(l, \theta) \right| \leq g_{i_1, \dots, i_d}(l).$$

g_{i_1, \dots, i_d} exist since $f_{L^\pi}(l)$ fulfils assumptions **R-4** and **R-11**.

R-4 Since the function is bounded on $(0, W_{max}]$ by c , we set the needed majorants as

$$g(\tilde{l}) = \begin{cases} c, & \tilde{l} \in (0, W_{max}] \\ 0, & \text{else} \end{cases}$$

and get the needed result. where the last step is valid since $g_{i_1, \dots, i_d} \in \mathcal{L}^1$. We know that $f_{\tilde{L}^\pi}(\tilde{l}; \theta)$ is bounded by a constant. We set $\tilde{g}_{i_1, \dots, i_d}(l) = \tilde{c}$.

R-5 The fisher information is positive definite by assumption.

R-6 For this property we investigate the third derivatives of $\ln f_{\tilde{L}^\pi}(\tilde{l})$. If we look at the case $Z = 0$ we get

$$\left| \frac{\partial^3}{\partial \theta_j \theta_k \theta_l} \ln f_{\tilde{L}^\pi}(\tilde{l}, \theta) \right| = \left| \frac{\partial^3}{\partial \theta_j \theta_k \theta_l} \ln f_{L^\pi}(\tilde{l}, \theta) \right|$$

since $f_{L^\pi}(l, \theta)$ fulfils condition **R-6** the majorant M_{i_1, \dots, i_d} exist and

$$\left| \frac{\partial^3}{\partial \theta_j \theta_k \theta_l} \ln f_{\tilde{L}^\pi}(\tilde{l}, \theta) \right| < M_{i_1, \dots, i_d}(\tilde{l}) \text{ and } \mathbb{E}(M_{i_1, \dots, i_d}(\tilde{L})) < \infty$$

exists as well. For $Z = 1$ we look at

$$\begin{aligned} \left| \frac{\partial^3}{\partial \theta_j \theta_k \theta_l} \ln f_{\tilde{L}^\pi}(\tilde{l}, \theta) \right| &\leq \left| \frac{\frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} f_{\tilde{L}^\pi}(\tilde{l}, \theta)}{f_{\tilde{L}^\pi}(\tilde{l}, \theta)} \right| \\ &+ \left| \frac{\frac{\partial}{\partial \theta_j} f_{\tilde{L}^\pi}(\tilde{l}, \theta) \frac{\partial^2}{\partial \theta_k \partial \theta_l} f_{\tilde{L}^\pi}(\tilde{l}, \theta)}{f_{\tilde{L}^\pi}^2(\tilde{l}, \theta)} \right| \\ &+ \left| \frac{\frac{\partial}{\partial \theta_k} f_{\tilde{L}^\pi}(\tilde{l}, \theta) \frac{\partial^2}{\partial \theta_j \partial \theta_l} f_{\tilde{L}^\pi}(\tilde{l}, \theta)}{f_{\tilde{L}^\pi}^2(\tilde{l}, \theta)} \right| \\ &+ \left| \frac{\frac{\partial}{\partial \theta_l} f_{\tilde{L}^\pi}(\tilde{l}, \theta) \frac{\partial^2}{\partial \theta_j \partial \theta_k} f_{\tilde{L}^\pi}(\tilde{l}, \theta)}{f_{\tilde{L}^\pi}^2(\tilde{l}, \theta)} \right| \\ &+ 2 \left| \frac{\frac{\partial}{\partial \theta_j} f_{\tilde{L}^\pi}(\tilde{l}, \theta) \frac{\partial}{\partial \theta_k} f_{\tilde{L}^\pi}(\tilde{l}, \theta) \frac{\partial}{\partial \theta_l} f_{\tilde{L}^\pi}(\tilde{l}, \theta)}{f_{\tilde{L}^\pi}^3(\tilde{l}, \theta)} \right|. \end{aligned} \tag{3.17}$$

We can split each of the summands into factors of the form

$$\begin{aligned} &\left| \frac{\frac{\partial^d}{\partial^{i_1} \theta_1 \dots \partial^{i_d} \theta_d} f_{\tilde{L}^\pi}(\tilde{l}, \theta)}{f_{\tilde{L}^\pi}(\tilde{l}, \theta)} \right| = \\ &\left| \frac{\frac{\partial^d}{\partial^{i_1} \theta_1 \dots \partial^{i_d} \theta_d} \left(\int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi | L^\pi}(\tilde{l}, l) f_{L^\pi}(l, \theta) dl \right)}{\int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi | L^\pi}(\tilde{l}, l) f_{L^\pi}(l, \theta) dl} \right| \end{aligned}$$

for $d = 1, 2, 3$ and $i_0 + \dots + i_d = d$. The existence of these fractions is given by assumption **R-12**. We can therefore conclude that

$$\left| \frac{\partial^3}{\partial \theta_j \theta_k \theta_l} \ln f_{\tilde{L}^\pi}(\tilde{l}) \right| \leq c$$

where c is constant. We choose $M_{i_1, \dots, i_d}(\tilde{l}) = c$, the mean of this is

$$\mathbb{E}(M_{i_1, \dots, i_d}(\tilde{L})) = \mathbb{E}(c) = c$$

and therefore finite.

We choose

$$\tilde{M}_{i_1, \dots, i_d}(\tilde{l}) = Zc + (1 - Z)M_{i_1, \dots, i_d}(\tilde{l})$$

as the majorant.

Theorem 3.4. *If $f_{\tilde{L}^\pi|L^\pi,U}(\tilde{l}, l, u)$ has $c(\tilde{l}, u)$ for all $u \in S^{d-1}$ and $\tilde{l} \in (0, W_{max}]$ such that*

$$f_{\tilde{L}^\pi|L^\pi,U}(\tilde{l}, l, u) < c(\tilde{l}, u)$$

for all $l \in [\tilde{l}, \infty)$, then

$$f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l) < \mathbb{E}(c(\tilde{l}, U)).$$

Proof. Follows directly by the monotony of the mean. □

This means that when we deal with the boundedness of the marginalised density $f_{\tilde{L}^\pi|L^\pi}(\tilde{l}, l)$ we can look at the easier density $f_{\tilde{L}^\pi|L^\pi,U}(\tilde{l}, l, u)$.

Associated Point Rule 5: Consistency and asymptotic normality and convergence

Theorem 3.5. *Assume we have line segment process with a log-normal length distribution with parameters $\theta_0 = (\mu_0, \sigma_0) \in \Theta$ with Θ compact, which is observed in a rectangular window W with the associated point rule and orientation distribution with density $f_U(u)$. This leads to a censored sample $(\tilde{L}_i, Z_i), i = 1, \dots, n$ with density*

$$f_{\tilde{L}}(\tilde{l}; \theta) = Z \sum_{i=1}^2 \sum_{j=1}^4 \int_{c_{\vartheta,i,l}(\tilde{l})}^{c_{\vartheta,i,u}(\tilde{l})} \int_{c_{\varphi,l,j}(\tilde{l},\vartheta)}^{c_{\varphi,u,j}(\tilde{l},\vartheta)} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_L(l; \mu, \sigma) dl f_U(\theta, \varphi) d\varphi d\theta \\ + (1 - Z) f_L(\tilde{l}; \mu, \sigma).$$

If we furthermore assume that fisher information is positive definite, then the resulting maximum likelihood estimator is consistent and asymptotically normal.

Proof. We need to verify the conditions of theorem 3.3. We know that the log normal distribution fulfils the conditions of theorem 2.8.

We first look at

$$f_{\tilde{L}|L,U}(\tilde{l}|l, u) = \frac{-\frac{\partial}{\partial \tilde{l}} a(l, \tilde{l}, u)}{|W| - v(l)}$$

and note that the denominator is always positive and the numerator only depends on \tilde{l} and not l . Therefore to investigate the behaviour in \tilde{l} we only have to look at the numerator. We take its derivative in \tilde{l} and get

$$-\frac{\partial^2}{\partial \tilde{l}^2} a(l, \tilde{l}, u) = -2c_1 c_2 (W_3 - c_3 \tilde{l}) - 2c_1 (W_2 - c_2 \tilde{l}) c_3 - 2(W_1 - c_1 \tilde{l}) c_2 c_3$$

and since $W_1 - c_1\tilde{l} > 0, W_2 - c_2\tilde{l} > 0$ and $W_3 - c_3\tilde{l} > 0$ this is always negative . Therefore $f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, u)$ will be decreasing in \tilde{l} for all l and u .

Furthermore we see that the denominator is increasing with l . We therefore conclude that

$$c_{max}(\tilde{l}) = f_{\tilde{L}|L,U}(\tilde{l}|\tilde{l}, u).$$

To investigate the boundedness on $\tilde{l} \in (0, W_{max}]$ we only have to look at

$$\begin{aligned} \lim_{\tilde{l} \rightarrow 0} f_{\tilde{L}}(\tilde{l}; \theta) &= \lim_{\tilde{l} \rightarrow 0} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_L(l; \mu, \sigma) dl f_U(\vartheta, \varphi) d \\ &= \int_0^{\infty} f_{\tilde{L}^\pi|L,U}(0|l, \vartheta, \varphi) f_L(l; \mu, \sigma) dl \end{aligned}$$

due to continuity of the function. If we look at the integrand, that

$$\lim_{\tilde{l} \rightarrow 0} f_{\tilde{L}^\pi|L,U}(0|l, \vartheta, \varphi) = \infty$$

and the integral might be unbounded. We see that

$$\lim_{\tilde{l} \rightarrow 0} f_{\tilde{L}^\pi|L,U}(0|l, \vartheta, \varphi) f_L(l; \mu, \sigma) = 0$$

therefore the integrand is bounded and the whole integral is bounded. Therefore $f_{\tilde{L}}(\tilde{l}; \theta)$ is continuous and bounded on $(0, W_{max}]$.

For condition **R-12** we first note that if this property holds for each of the summands in the density $f_{\tilde{L}}$ it holds for the whole density. We only look at one summand and have to show that

$$\frac{\frac{\partial^d}{\partial^{i_1} \mu \partial^{i_2} \sigma} \left(\int_{c_{\vartheta,u}(\tilde{l})}^{c_{\varphi,u}(\tilde{l},\vartheta)} \int_{c_{\varphi,l}(\tilde{l},\vartheta)}^{\infty} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_L(l; \mu, \sigma) dl f_U(\vartheta, \varphi) d\varphi d\vartheta \right)}{\sum_{i=1}^2 \sum_{j=1}^4 \int_{c_{\vartheta,l,i}(\tilde{l})}^{c_{\vartheta,u,i}(\tilde{l})} \int_{c_{\varphi,l,j}(\tilde{l},\vartheta)}^{c_{\varphi,u,j}(\tilde{l},\vartheta)} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_L(l; \mu, \sigma) dl f_U(\vartheta, \varphi) d\varphi d\vartheta} < \infty.$$

where the angle boundaries of the numerator is one of the possible combination.

We note that for $\tilde{l} \in (0, W_{max})$ $f_{\tilde{L}|L}(\tilde{l}, l) > 0$ and therefore the fraction exists. The problem is at the limit of $\tilde{l} = W_{max}$. For $\tilde{l} = D$ the integration boundaries $c_{\vartheta,l}(W_{max}) = c_{\vartheta,u}(W_{max})$ and therefore $c_{\varphi,l}(W_{max}, c_{\vartheta,l}(W_{max})) = c_{\varphi,u}(W_{max}, c_{\vartheta,l}(W_{max}))$, we are integrating over a null set and the integral is 0 for numerator and denominator . We therefore use the rule of l'Hospital

and compute for the numerator

$$\begin{aligned}
& \frac{\partial}{\partial \tilde{l}} \int_{c_{\vartheta,i}(\tilde{l})}^{c_{\vartheta,u}(\tilde{l})} \int_{c_{\varphi,i}(\tilde{l},\vartheta)}^{c_{\varphi,u}(\tilde{l},\vartheta)} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_L(\tilde{l}; \mu, \sigma) dl f_U(\vartheta, \varphi) d\varphi d\vartheta = \\
& \frac{\partial c_{\vartheta,u}(\tilde{l})}{\partial \tilde{l}} \int_{c_{\varphi,i}(\tilde{l}, c_{\vartheta,u}(\tilde{l}))}^{c_{\varphi,2}(\tilde{l}, c_{\vartheta,u}(\tilde{l}))} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_L(l; \mu, \sigma) dl f_U(c_{\vartheta,2}(\tilde{l}), \varphi) d\varphi \\
& - \frac{\partial c_{\varphi,l,j}(\tilde{l})}{\partial \tilde{l}} \int_{c_{\varphi,1}(\tilde{l}, c_{\varphi,l,j}(\tilde{l}))}^{c_{\varphi,2}(\tilde{l}, c_{\varphi,l,j}(\tilde{l}))} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_L(l; \mu, \sigma) dl f_U(c_{\varphi,l,j}(\tilde{l}), \varphi) d\varphi \\
& + \int_{c_{\varphi,l,j}(\tilde{l})}^{c_{\vartheta,2}(\tilde{l})} \frac{\partial}{\partial \tilde{l}} \int_{c_{\varphi,1}(\tilde{l}, \vartheta)}^{c_{\varphi,2}(\tilde{l}, \vartheta)} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_L(l; \mu, \sigma) dl f_U(\vartheta, \varphi) d\varphi d\vartheta
\end{aligned}$$

using Leibniz integral rule. This still converges to zero and we will still integrate over a null set in the limit. If we apply the rule of l'Hospital again and ignore all integrals over angles we get summands of the form

$$\begin{aligned}
& \frac{\partial c_{\vartheta,i}(\tilde{l})}{\partial \tilde{l}} \frac{\partial c_{\varphi,j}(\tilde{l}, c_{\vartheta,i}(\tilde{l}))}{\partial \tilde{l}} f_U(c_{\vartheta,i}(\tilde{l}), c_{\varphi,j}(\tilde{l}, c_{\vartheta,i}(\tilde{l}))) \\
& \frac{\partial}{\partial \tilde{l}} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, c_{\vartheta,i}(\tilde{l}), c_{\varphi,j}(\tilde{l}, c_{\vartheta,i}(\tilde{l}))) f_L(l; \mu, \sigma) dl
\end{aligned}$$

where

$$\frac{\partial c_{\vartheta,i}(\tilde{l})}{\partial \tilde{l}} \frac{\partial c_{\varphi,j}(\tilde{l}, c_{\vartheta,i}(\tilde{l}))}{\partial \tilde{l}} f_U(c_{\vartheta,i}(\tilde{l}), c_{\varphi,j}(\tilde{l}, c_{\vartheta,i}(\tilde{l})))$$

converges to a constant for $\tilde{l} \rightarrow W_{max}$. We therefore have to investigate the limit behaviour of

$$\begin{aligned}
\frac{\partial}{\partial \tilde{l}} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \sigma) f_L(l; \mu, \sigma) dl &= \int_{\tilde{l}}^{\infty} \frac{\partial}{\partial \tilde{l}} f_{\tilde{L}|L,U}(\tilde{l}, l, \vartheta, \varphi) f_L(l, \mu, \sigma) dl \\
&- f_{\tilde{L}|L,U}(\tilde{l}, \tilde{l}, \vartheta, \varphi) f_L(\tilde{l})
\end{aligned}$$

which still converges to 0 since $\tilde{l} = W_{max}$ is a double root of $f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \sigma)$.

We therefore differentiate again and get

$$\begin{aligned} \frac{\partial^2}{\partial \tilde{l}^2} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \sigma) f_L(l; \mu, \sigma) dl &= \int_{\tilde{l}}^{\infty} \frac{\partial^2}{\partial \tilde{l}^2} f_{\tilde{L}|L,U}(\tilde{l}, l, \vartheta, \varphi) f_L(l, \mu, \sigma) dl \\ &- \frac{\partial}{\partial \tilde{l}} f_{\tilde{L}|L,U}(\tilde{l}, \tilde{l}, \vartheta, \varphi) f_L(\tilde{l}) - \frac{\partial}{\partial \tilde{l}} \left(f_{\tilde{L}|L,U}(\tilde{l}, \tilde{l}, \vartheta, \varphi) f_L(\tilde{l}) \right) \end{aligned} \quad (3.18)$$

we see that

$$\frac{\partial^2}{\partial \tilde{l}^2} f_{\tilde{L}|L,U}(\tilde{l}, l, \vartheta, \varphi) = \frac{-6c_1 c_2 c_3}{|W|} \neq 0$$

for $\tilde{l} \rightarrow W_{max}$. We see that after differentiating the denominator 4 times we have a limit that is constant and neither 0 or ∞ for $\tilde{l} \rightarrow W_{max}$. If we do the same for the nominator, we will get similar results. We therefore conclude, that

$$\frac{\frac{\partial^d}{\partial \tilde{l}^d} \int_{c_{\vartheta,l}(\tilde{l})}^{c_{\vartheta,u}(\tilde{l})} \int_{c_{\varphi,l}(\tilde{l},\vartheta)}^{c_{\varphi,u}(\tilde{l},\vartheta)} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_L(l; \mu, \sigma) dl f_U(\vartheta, \varphi) d\varphi d\vartheta}{\sum_{i=1}^2 \sum_{j=1}^4 \int_{c_{\vartheta,l,i}(\tilde{l})}^{c_{\vartheta,u,i}(\tilde{l})} \int_{c_{\varphi,l,j}(\tilde{l},\vartheta)}^{c_{\varphi,u,j}(\tilde{l},\vartheta)} \int_{\tilde{l}}^{\infty} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_L(l; \mu, \sigma) dl f_U(\vartheta, \varphi) d\varphi d\vartheta} = c < \infty$$

for a constant $c \in \mathbb{R}$ and all $\tilde{l} \in (0, W_{max}]$. \square

Therefore the ML estimator is consistent and asymptotically normal. To investigate the convergence of the EM algorithm we use theorem 2.12.

Theorem 3.6. *The function $Q(\theta|\theta_0)$ from Equation (3.4) is continuous in θ and θ_0 and therefore the EM algorithm converges to a stationary point of the likelihood of $f_{\tilde{L}^*}$.*

Proof. We look at the function

$$\begin{aligned} Q(\theta|\theta_0) &= \sum_{i=1}^{n_1} -\ln(L_i) - \ln(\sqrt{2\pi}\sigma) + \left(-\frac{(\ln(L_i) - \mu)^2}{2\sigma^2} \right) \\ &+ \sum_{i=1}^{n_2} -\mathbb{E}_{\theta_0} [\ln(L_i)|\tilde{L}_i] + \ln(\sqrt{2\pi}\sigma) \\ &+ \left(-\frac{\mathbb{E}_{\theta_0} [\ln(L_i)^2|\tilde{L}_i] - \mu \mathbb{E}_{\theta_0} [\ln(L_i)|\tilde{L}_i] + \mu^2}{2\sigma^2} \right) \end{aligned}$$

since $\sigma > 0$ this is continuous for all σ and also for μ . To proof continuity in μ_0, σ_0 we have to look at

$$\begin{aligned}\mathbb{E}_{\theta_0} \left[\ln(L)^k | \tilde{L} \right] &= \int_0^{\infty} \ln(l)^k f_{L|\tilde{L}}(l|\tilde{L}, \mu_0, \sigma_0) \, dl \\ &= \int_0^{\infty} \ln(l)^k \frac{f_{\tilde{L}|L}(\tilde{L}, l) f_L(l; \mu_0, \sigma_0)}{f_{\tilde{L}}(\tilde{L}, \mu_0, \sigma_0)} \, dl \\ &= \frac{1}{f_{\tilde{L}}(\tilde{L}, \mu_0, \sigma_0)} \int_0^{\infty} \ln(l)^k f_{\tilde{L}|L}(\tilde{L}, l) f_L(l; \mu_0, \sigma_0) \, dl\end{aligned}$$

and show that it is continuous in θ_0 . To show the continuity of the integral we need a majorant for

$$\begin{aligned}|\ln(l)^k f_{\tilde{L}|L}(\tilde{L}, l) f_L(l; \mu_0, \sigma_0)| &\leq c_{max}(\tilde{L}) |\ln(x)^k| f_L(l; \mu_0, \sigma_0) \\ &\leq c_{max}(\tilde{L}) g_k(l)\end{aligned}$$

where the majorant $g_k(l)$ is given in (2.6), since $f_L(l, \mu, \sigma)$ is the density of a log-normal distribution. $c_{max}(\tilde{L})$ exists since $f_{\tilde{L}|L}$ fulfils property **R-10**. To show that

$$f_{\tilde{L}}(\tilde{l}, \mu_0, \sigma_0) = \int_{\tilde{l}}^{\infty} f_{\tilde{L}|L}(\tilde{l}, l) f_L(l, \mu_0, \sigma_0) \, dl$$

is continuous we have to find a majorant. Using again the existence of c_{max} and the existence of the majorant for the log-normal density we get

$$\begin{aligned}|f_{\tilde{L}|L}(\tilde{L}, l) f_L(l, \mu_0, \sigma_0)| &\leq c_{max}(\tilde{L}) |f_L(l, \mu_0, \sigma_0)| \\ &\leq c_{max}(\tilde{l}) g_0(l).\end{aligned}$$

Therefore $Q(\theta|\theta_0)$ is continuous in θ and θ_0 and the EM algorithm will converge to a stationary point of the likelihood function. \square

Therefore the EM algorithm will converge to a stationary point of the likelihood function of the censored fibres \tilde{L} .

Plus Sampling 5: Consistency and asymptotic normality and convergence

Theorem 3.7. *Assume we have line segment process with a log-normal length distribution with parameters $\theta_0 = (\mu_0, \sigma_0)$, which is observed in a rectangular window W with plus sampling as the used sampling rule and orientation distribution with density $f_U(u)$. This leads to a censored sample $(\tilde{L}_i^{*+}, Z_i), i = 1, \dots, n$ with density*

$$f_{\tilde{L}^{*+}}(\tilde{l}; \theta) = Z \sum_{i=1}^2 \sum_{j=1}^4 \int_{\tilde{l}}^{\infty} \int_{c_{\vartheta, l, i}(\tilde{l}, l)}^{c_{\vartheta, u, i}(\tilde{l}, l)} \int_{c_{\varphi, l, j}(\tilde{l}, l, \vartheta)}^{c_{\varphi, u, j}(\tilde{l}, l, \vartheta)} f_{\tilde{L}|L^+, U}(\tilde{l}|l, \vartheta, \varphi) f_{L^+}(l; \theta) dl f_U(\vartheta, \varphi) d\varphi d\vartheta dl + (1 - Z) f_{L^+}(\tilde{l}; \theta).$$

If the resulting fisher information matrix is positive definite, then the resulting maximum likelihood estimator is consistent and asymptotically normal.

Proof R-10 If we look at the density we have

$$f_{\tilde{L}^\pi|L, U}(\tilde{l}|l, \vartheta, \varphi) = \frac{\frac{\partial}{\partial \tilde{l}} a(l, \tilde{l}, \vartheta, \varphi)}{w(l, \vartheta, \varphi) - v(l, \vartheta, \varphi)}.$$

We first look at the denominator for $v(l) > 0$ we get

$$\begin{aligned} w(l, \vartheta, \varphi) - v(l, \vartheta, \varphi) &= |W| + (W_1 W_2 c_3 l) + (W_1 c_2 l W_3) + (c_1 l W_2 W_3) \\ &\quad - (W_1 - c_1 l)(W_2 - c_2 l)(W_3 - c_3 l) \\ &\geq |W| + (W_1 W_2 c_3 \tilde{l}) + (W_1 c_2 \tilde{l} W_3) + (c_1 \tilde{l} W_2 W_3) \\ &\quad - (W_1 - c_1 \tilde{l})(W_2 - c_2 \tilde{l})(W_3 - c_3 \tilde{l}) \end{aligned}$$

since $l > \tilde{l}$. The numerator itself is continuous in \tilde{l} and therefore bounded for all l and \tilde{l} . We now look at the limit case

$$\lim_{l \rightarrow \infty} \frac{\frac{\partial}{\partial \tilde{l}} a(l, \tilde{l}, \vartheta, \varphi)}{w(l, \vartheta, \varphi)} = \frac{2(W_1 - c_1 \tilde{l})c_2 c_3 + 2c_1(W_2 - c_2 \tilde{l})c_3 + 2c_1 c_2(W_3 - c_3 \tilde{l})}{W_1 W_2 c_3 + W_1 c_2 W_3 + c_1 W_2 W_3}$$

we note that we only need to look at the case for $\tilde{l} < \frac{l}{2}$ since $l \rightarrow \infty$. Also note that $v(l, \vartheta, \varphi) = 0$ for l large enough. We see that the limit in l is bounded. Since the function is continuous we conclude that it is bounded for all $l \in [\tilde{l}, \infty)$. We therefore conclude, that there exists a $c_{max}(\tilde{l})$ such that $f_{\tilde{L}^\pi|L, U}(\tilde{l}|l, \vartheta, \varphi) < c_{max}(\tilde{l})$.

R-11 Like for the associated point the boundedness of the integral is no problem for $\tilde{l} > 0$. We will look at the case for $\tilde{l} \rightarrow 0$

$$\begin{aligned} f_{\tilde{L}^\pi|L,U}(0|l, \vartheta, \varphi) f_{L^+}(l; \mu, \sigma) &= \frac{-\frac{\partial}{\partial l} a(l, 0, u)}{w(l, u) - v(l, u)} \frac{\xi l + |W|}{\xi \mathbb{E}(L) + |W|} f_L(l; \mu, \sigma) \\ &= \frac{1}{\xi \mathbb{E}(L) + |W|} \left(\frac{-\xi \frac{\partial}{\partial l} a(l, 0, u)}{w(l, u) - v(l, u)} f_L(l; \mu, \sigma) \right. \\ &\quad \left. - \frac{|W| \frac{\partial}{\partial l} a(l, 0, u)}{w(l, u) - v(l, u)} f_L(l; \mu, \sigma) \right) \end{aligned}$$

We note that there is a possible singularity for $l \rightarrow 0$. We get

$$\lim_{l \rightarrow 0} w(l, u) - v(l, u) = 0$$

but since the density $\lim_{l \rightarrow 0} f_L(l; \mu, \sigma) = 0$ and converges faster than $w(l, u) - v(l, u)$ we get

$$\lim_{l \rightarrow 0} \frac{1}{w(l, u) - v(l, u)} f_L(l; \mu, \sigma) = 0.$$

Therefore the integral

$$\int_{\tilde{l}}^{\infty} f_{\tilde{L}|L}(\tilde{l}, l) f_{L^+}(l, \mu, \sigma) dl$$

is bounded on $\tilde{l} \in (0, W_{max}]$.

R-12 For condition **R-12** we have to show that

$$\frac{\frac{\partial^d}{\partial \tilde{l}^1 \mu \partial \vartheta^2 \sigma} \left(\int_{\tilde{l}}^{\infty} \int_{c_{\vartheta,2}(\tilde{l}, l)}^{c_{\varphi,2}(\tilde{l}, l, \vartheta)} \int_{c_{\varphi,1}(\tilde{l}, l, \vartheta)}^{c_{\varphi,1}(\tilde{l}, l)} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_{L^+}(l; \mu, \sigma) f_U(\vartheta, \varphi) d\varphi d\vartheta dl \right)}{\int_{\tilde{l}}^{\infty} \int_{c_{\vartheta,1}(\tilde{l}, l)}^{c_{\varphi,2}(\tilde{l}, l)} \int_{c_{\varphi,1}(\tilde{l}, l, \vartheta)}^{c_{\varphi,2}(\tilde{l}, l, \vartheta)} f_{\tilde{L}^\pi|L,U}(\tilde{l}|l, \vartheta, \varphi) f_{L^+}(l; \mu, \sigma) f_U(\vartheta, \varphi) d\varphi d\vartheta dl} < \infty.$$

we first note that $f_{\tilde{L}^+|L,U}(\tilde{l}|l, \vartheta, \varphi) > 0$ for all $\tilde{l} \in (0, W_{max}]$ and $l \in [\tilde{l}, \infty)$. Problematic is, that in the limit case $\tilde{l} \rightarrow W_{max}$ we will integrate over null sets for the angles and both nominator and denominator will be 0. We deal with this the same way as in the associated point rule case, we apply the rule of l'Hospital multiple times using the Leibniz integral rule until all integrals over null sets are gone and since $f_{\tilde{L}^+|L,U}(\tilde{l}|l, \vartheta, \varphi) \neq 0$ for all $\tilde{l} \in [0, W_{max}]$ and $l \in [\tilde{l}, \infty)$ we know that this fractions exists and are bounded by a constant $c < \infty$ for all $\tilde{l} \in [0, W_{max}]$ and $l \in [\tilde{l}, \infty)$.

R-13 The fisher information is positive definite by assumption. \square

Theorem 3.8. *The function $Q(\theta|\theta_0)$ from Equation (3.6) is continuous in θ and θ_0 and therefore the EM algorithm converges to a stationary point of the likelihood of $f_{\tilde{L}}$.*

Proof. We look at the function from Equation(3.6)

$$\begin{aligned} Q(\theta|\theta^0) &= -n \ln(\xi e^{\mu + \frac{\sigma^2}{2}} + |W|) - n \ln(\sqrt{2\pi}\sigma) \\ &+ \sum_{i=1}^{n_1} \ln(\xi L_i + |W|) - \ln(L_i^+) + \left(-\frac{(\ln(L_i^+) - \mu)^2}{2\sigma^2} \right) \\ &+ \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} \left[\ln(\xi L_i^+ + |W|) | \tilde{L}_i^+ \right] - \mathbb{E}_{\theta_0} \left[\ln(L_i^+) | \tilde{L}_i^+ \right] \\ &+ \left(-\frac{\mathbb{E}_{\theta_0} \left[\ln(L_i^+)^2 | \tilde{L}_i^+ \right] - \mu \mathbb{E}_{\theta_0} \left[\ln(L_i^+) | \tilde{L}_i^+ \right] + \mu^2}{2\sigma^2} \right) \end{aligned}$$

since $\sigma > 0$ we see that this is continuous in μ and σ .

To show the continuity in μ_0, σ_0 we look at

$$\begin{aligned} \mathbb{E}_{\theta_0} \left(\ln(L^+)^k | \tilde{L}^+ \right) &= \int_0^{\infty} \ln(l)^k f_{L^+ | \tilde{L}^+}(l | \tilde{L}^+, \mu_0, \sigma_0) dl \\ &= \int_0^{\infty} \ln(l)^k \frac{f_{\tilde{L}^+ | L^+}(\tilde{l} | l) f_{L^+}(l, \mu_0, \sigma_0)}{f_{\tilde{L}^+ *}(\tilde{l}, \mu_0, \sigma_0)} dl \\ &= \frac{1}{f_{\tilde{L}^+ *}(\tilde{l}, \mu_0, \sigma_0)} \int_0^{\infty} \ln(l)^k f_{\tilde{L}^+ | L^+}(\tilde{l}, l) f_{L^+}(l, \mu_0, \sigma_0) dl \end{aligned}$$

To show the continuity of the integral we need a majorant for

$$\begin{aligned} |\ln(l)^k f_{\tilde{L}^+ | L^+}(\tilde{l}, l) f_{L^+}(l, \mu_0, \sigma_0)| &\leq c_{max}(\tilde{l}) |\ln(x)^k| f_{L^+}(l, \mu_0, \sigma_0) \\ &\leq c_{max}(\tilde{l}) g_k(l) \end{aligned}$$

where the majorant $g_k(l)$ is the known majorant for the reweighed density case from proof of theorem 3.2 and $c_{max}(\tilde{l})$ exists.

To show that

$$f_{\tilde{L}^+ *}(\tilde{l}, \mu_0, \sigma_0) = \int_{\tilde{l}}^{\infty} f_{\tilde{L}^+ | L^+}(\tilde{l}, l) f_{L^+}(l, \mu_0, \sigma_0) dl$$

is continuous we have to find a majorant

$$\begin{aligned} |f_{\tilde{L}^+|L^+}(\tilde{l}, l) f_{L^+}(l, \mu_0, \sigma_0)| &\leq c_{max} |f_{L^+}(l, \mu_0, \sigma_0)| \\ &\leq c_{max} g_0(l). \end{aligned}$$

where $g_0(l)$ is the known majorant for the reweighed case. \square

Remark

In our proofs for the consistency and asymptotic normality of our estimators we assumed that the resulting fisher information matrix is positive definite. The problem is, that direct computation by using the density $f_{\tilde{L}^{\pi^}}(\tilde{l})$ we get terms that are analytically not tractable.*

Whether the fisher information matrix is positive definite remains an open question.

3.3.6. Variance of the estimator

To estimate the mean square error of the EM estimator we use the Louis method.

The first part of the Louis method is the Hessian of the complete data likelihood. We calculate the Hessian of Equation (3.3) and get

$$\begin{aligned} \mathcal{H}(Q(\theta|\theta_0)) &= -n \mathcal{H}(\ln(\mathbb{E}[\pi(L)])) + \sum_{i=1}^{n_1} \mathcal{H}(\ln(f(L_i^\pi, \theta))) \\ &\quad + \sum_{i=1}^{n_2} \mathcal{H} \mathbb{E}_{\theta_0} \left[(\ln(f(L_i^\pi, \theta))) | \tilde{L}_i^\pi \right] \end{aligned}$$

For the covariance we need the gradient of the log-likelihood function and get

$$\begin{aligned} \nabla l(\theta|L) &= \nabla \sum_{i=1}^n \ln(\pi(l)) - \ln \mathbb{E}[\pi(L)] + \ln f_L(L_i^\pi, \theta) \\ &= -n \nabla \ln \mathbb{E}[\pi(L)] + \sum_{i=1}^n \nabla \ln f_{L^\pi}(L_i^\pi, \theta) \end{aligned}$$

Calculating the covariance yields

$$\begin{aligned} \text{cov}_{L^\pi | \tilde{L}; \hat{\theta}_0}(\nabla l(\theta|L^\pi)) &= \text{cov}_{L^\pi | \tilde{L}; \hat{\theta}_0} \left(-n \nabla \ln \mathbb{E}[\pi(L)] + \sum_{i=1}^n \nabla \ln f_{L^\pi}(L_i^\pi, \theta) \right) \\ &= \text{cov}_{L^\pi | \tilde{L}; \hat{\theta}_0} \left(\sum_{i=1}^n \nabla \ln f_{L^\pi}(L_i^\pi, \theta) \right) \end{aligned}$$

Associated Point Rule 6:

The EM iteration leads to an estimator $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$, to estimate the variance we need the fisher information. To estimate it we use Equation (2.14). We therefore need all first two derivatives of $l(\theta|L)$, we get

$$\begin{aligned}
\frac{\partial l(\theta|L)}{\partial \mu} &= \frac{\sum_{i=1}^n \ln(L_i) - n\mu}{\sigma^2} \\
\frac{\partial l(\theta|L)}{\partial \sigma} &= \frac{\sum_{i=1}^n (\ln(L_i) - \mu)^2 - n\sigma^2}{\sigma^3} \\
&= \frac{\sum_{i=1}^n (\ln(L_i)^2 - \ln(L_i)\mu) + n\mu^2 - n\sigma^2}{\sigma^3} \\
\frac{\partial^2 l(\theta|L)}{\partial \mu^2} &= \frac{-n}{\sigma^2} \\
\frac{\partial^2 l(\theta|L)}{\partial \mu \partial \sigma} &= \frac{-2 \left(\sum_{i=1}^n \ln(L_i) - n\mu \right)}{\sigma^3} \\
\frac{\partial^2 l(\theta|L)}{\partial \sigma^2} &= \frac{-3 \left(\sum_{i=1}^n (\ln(L_i) - \mu)^2 + n\sigma^2 \right)}{\sigma^4} \\
&= \frac{-3 \left(\sum_{i=1}^n (\ln(L_i)^2 - 2\mu \ln(L_i)) + n\mu^2 + n\sigma^2 \right)}{\sigma^4}.
\end{aligned}$$

With this we get the components of $\mathcal{H}(Q(\theta|\theta_0))$ as

$$\begin{aligned}
\frac{\partial^2 Q(\theta|\theta_0)}{\partial \mu^2} &= \frac{-n}{\sigma^2} \\
\frac{\partial^2 Q(\theta|\theta_0)}{\partial \mu \partial \sigma} &= \frac{-2 \left(\sum_{i=1}^{n_1} \ln(L_i) + \sum_{i=1}^{n_2} \mathbb{E} \left[\ln(L_i) | \tilde{L}_i \right] - n\mu \right)}{\sigma^2} \\
\frac{\partial^2 Q(\theta|\theta_0)}{\partial \sigma^2} &= \frac{-3 \left(\sum_{i=1}^{n_1} (\ln(L_i)^2 - 2\mu \ln(L_i)) \right)}{\sigma^4} \\
&\quad + \frac{-3 \left(\sum_{i=1}^{n_2} \left(\mathbb{E}_{\theta_0} \left[\ln(L_i)^2 | \tilde{L}_i \right] - 2\mu \mathbb{E}_{\theta_0} \left[\ln(L_i) | \tilde{L}_i \right] \right) + n\mu^2 + n\sigma^2 \right)}{\sigma^4}.
\end{aligned} \tag{3.19}$$

The first diagonal component of $\text{cov}(\nabla_{\theta} l(\theta|L))$ is

$$\begin{aligned} \text{var}\left(\frac{\partial l(\theta|L)}{\partial \mu}\right) &= \text{var}\frac{\sum_{i=1}^n \ln(L_i) - n\mu}{\sigma^2} \\ &= \frac{1}{\sigma^4} \sum_{i=1}^n \text{var}(\ln(L_i)) \\ &= \frac{1}{\sigma^4} \left(\sum_{i=1}^{n_1} \ln(L_i)^2 - \sum_{i=1}^{n_1} (\ln(L_i))^2 \right. \\ &\quad \left. + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i)^2 | \tilde{L}_i] - \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i) | \tilde{L}_i]^2 \right) \end{aligned}$$

We note that the influence of the uncensored fibres vanishes since the variance is in respect to the given censored length and we get

$$\text{var}\left(\frac{\partial l(\theta|L)}{\partial \mu}\right) = \frac{1}{\sigma^4} \left(\sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i)^2 | \tilde{L}_i] - \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i) | \tilde{L}_i]^2 \right). \quad (3.20)$$

The second diagonal component is

$$\begin{aligned} \text{var}\left(\frac{\partial l(\theta|L)}{\partial \sigma}\right) &= \text{var}\left(\frac{\sum_{i=1}^n (\ln(L_i)^2 - \ln(L_i)\mu) + n\mu^2 - n\sigma^2}{\sigma^3}\right) \\ &= \frac{1}{\sigma^6} \text{var}\left(\sum_{i=1}^n (\ln(L_i)^2 - \ln(L_i)\mu)\right) \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n (\text{var}(\ln(L_i)^2 - \ln(L_i)\mu)) \end{aligned}$$

The influence of the uncensored fibres vanishes as well and we get

$$\begin{aligned} \text{var}\left(\frac{\partial l(\theta|L)}{\partial \sigma}\right) &= \frac{1}{\sigma^6} \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} \left[\left(\ln(L_i)^2 - \ln(L_i)\mu - \mathbb{E}_{\theta_0} (\ln(L_i)^2 - \ln(L_i)\mu | \tilde{L}_i) \right)^2 | \tilde{L}_i \right] \\ &= \frac{1}{\sigma^6} \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} \left[\ln(L_i)^4 | \tilde{L}_i \right] + \mu^2 \mathbb{E}_{\theta_0} \left[\ln(L_i)^2 | \tilde{L}_i \right] \\ &\quad + \left(\mathbb{E}_{\theta_0} \left[\ln(L_i)^2 | \tilde{L}_i \right] \right)^2 + \mu^2 \left(\mathbb{E}_{\theta_0} \left[\ln(L_i) | \tilde{L}_i \right] \right)^2 \\ &\quad - 2\mu \mathbb{E}_{\theta_0} \left[\ln(L_i)^3 | \tilde{L}_i \right] - 2 \left(\mathbb{E}_{\theta_0} \left[\ln(L_i)^2 | \tilde{L}_i \right] \right)^2 \\ &\quad + 2\mu \mathbb{E}_{\theta_0} \left(\ln(L_i)^2 | \tilde{L}_i \right) \mathbb{E}_{\theta_0} \left[\ln(L_i) | \tilde{L}_i \right] \\ &\quad - 2\mu^2 \left(\mathbb{E}_{\theta_0} \left[\ln(L_i) | \tilde{L}_i \right] \right)^2. \end{aligned} \quad (3.21)$$

The last component is given by

$$\begin{aligned} \text{cov} \left(\frac{\partial l(\theta|L)}{\partial \mu}, \frac{\partial l(\theta|L)}{\partial \sigma} \right) &= \text{cov} \left(\frac{\sum_{i=1}^n \ln(L_i) - n\mu}{\sigma^2}, \right. \\ &\quad \left. \frac{\sum_{i=1}^n (\ln(L_i)^2 - \ln(L_i)\mu) + n\mu^2 - n\sigma^2}{\sigma^3} \right) \\ &= \frac{1}{\sigma^5} \text{cov} \left(\sum_{i=1}^n \ln(L_i), \sum_{i=1}^n (\ln(L_i)^2 - \ln(L_i)\mu) \right) \end{aligned}$$

using the bilinearity of the covariance we get

$$\text{cov} \left(\frac{\partial l(\theta|L)}{\partial \mu}, \frac{\partial l(\theta|L)}{\partial \sigma} \right) = \frac{1}{\sigma^5} \sum_{i=1}^n \sum_{j=1}^n \text{cov} (\ln(L_i), \ln(L_j)^2) - \mu \text{cov} (\ln(L_i), \ln(L_j))$$

since the L_i are independent we have $\text{cov} (\ln(L_i), \ln(L_j)^2) = 0$ and $\text{cov} (\ln(L_i), \ln(L_j)) = 0$ for $i \neq j$. This leads to

$$\begin{aligned} \text{cov} \left(\frac{\partial l(\theta|L)}{\partial \mu}, \frac{\partial l(\theta|L)}{\partial \sigma} \right) &= \frac{1}{\sigma^5} \sum_{i=1}^{n_2} \text{cov} (\ln(L_i), \ln(L_i)^2) - \mu \text{cov} (\ln(L_i), \ln(L_i)) \\ &= \frac{1}{\sigma^5} \sum_{i=1}^{n_2} \mathbb{E}[\ln(L_i)^3 | \tilde{L}_i] - \mathbb{E}[\ln(L_i) | \tilde{L}_i] \mathbb{E}[\ln(L_i)^2 | \tilde{L}_i] \\ &\quad - \mu \left(\mathbb{E}[\ln(L_i)^2 | \tilde{L}_i] - \mathbb{E}[\ln(L_i) | \tilde{L}_i]^2 \right) \end{aligned} \tag{3.22}$$

We can calculate the observed fisher information for an EM estimator θ_{EM} by using the Louis method and get

$$\mathcal{I}(\theta_{EM}, \tilde{L}) = -\mathcal{H}(Q(\theta_{EM} | \theta_{EM})) - \text{cov} (\nabla_{\theta} l(\theta_{EM} | L)) \tag{3.23}$$

we can estimate the covariance of the estimator by

$$\text{cov}(\theta_{EM}) = \mathcal{I}(\theta_{EM}, \tilde{L})^{-1}.$$

Plus Sampling 6:

The EM estimator leads to an estimator $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$. we use the Louis method to estimate the observed fisher information by

$$\hat{\mathcal{I}}(\tilde{L}^+, \hat{\theta}) = -\mathcal{H}(Q(\hat{\theta}|\hat{\theta})) - \text{cov}_{L^+|\tilde{L}^+, \hat{\theta}} \nabla_{\theta} l(\tilde{L}^+, L^+; \hat{\theta})$$

we look at

$$\text{cov}_{L^+|\tilde{L}^+, \hat{\theta}} \nabla_{\theta} l(\tilde{L}^+, L^+; \hat{\theta}) = \text{cov}_{L^+|\tilde{L}^+, \hat{\theta}} \sum_{i=1}^n \ln(f_{L^+}(L_i^+; \mu, \sigma))$$

which leads to the same formulae as in the associated point rule.

We further see

$$\mathcal{H}(Q(\theta|\theta_0)) = -n \mathcal{H} \ln(\mathbb{E}(\pi(L))) + \sum_{i=1}^n \mathcal{H} \mathbb{E}_{\theta_0} \left[\ln f_L(L_i; \mu, \sigma) | \tilde{L}_i \right]$$

and note that

$$\sum_{i=1}^n \mathbb{E}_{\theta_0} \left[\ln f_L(L_i; \mu, \sigma) | \tilde{L}_i \right]$$

is also known from the associated point rule. We get

$$\mathcal{H} \ln(\mathbb{E}(\pi(L^+))) = \frac{xi|W| \exp(\mu + \frac{\sigma^2}{2})}{\left(\xi \exp(\mu + \frac{\sigma^2}{2}) + |W| \right)} \begin{pmatrix} 1 \\ \sigma \frac{\sigma}{\xi \exp(\mu + \frac{\sigma^2}{2}) + |W|} \end{pmatrix}$$

for the last unknown part. We can now estimate the complete fisher information with

$$\begin{aligned} \hat{\mathcal{I}}(\tilde{L}, \hat{\theta}) &= n \mathcal{H} \ln(\mathbb{E}(\pi(L^+))) - \sum_{i=1}^n \mathbb{E} \left[f_L(L_i; \mu, \sigma) | \tilde{L}_i \right] \\ &\quad + \text{cov}_{L^+|\tilde{L}^+, \hat{\theta}} \sum_{i=1}^n \mathbb{E}(\ln(f(l; \mu, \sigma))). \end{aligned} \tag{3.24}$$

Note that, even though a lot of the formula look similar as in the associated point rule case, the conditional means are in regards to \tilde{L}^+ and not \tilde{L} .

3.4. Simulation Study

In this section we will do a simulation study for the proposed EM estimator. In this study we will investigate the EM estimator described before using the two sampling

$\mathbb{E}[L]$	$\text{std}(L)$	μ	σ
100	10	4.6002	0.0998
100	30	4.5621	0.2946
100	50	4.4936	0.4772

Table 3.1.: Parameters of the length distribution.

rules. The associated point rule which leads to an unbiased sample but ignores some of the available data and plus sampling which uses all the data but leaves us with biased sample. We will use a reweighed density to deal with this problem.

We will directly compare these methods to the HT style estimators.

We will investigate the influence of window size and parameters of the length distribution.

3.4.1. Geometrical Setting

We will investigate a boolean fibre process in a die of size $[0, W]^3$. We choose three different sizes $W = 100, 200, 300$ to simulate different degrees of censoring. In a smaller window more fibres will be censored. We will do this with a sample size of $n = 20, 50, 100, 500$, chosen with the appropriate sampling rule.

For the length L we choose a log normal distribution with three sets of parameters seen in table 3.1 to see the influence of different variances on the estimator. The parameters were chosen to get round numbers in the mean and standard deviation. For the orientation distribution we look at a bipolar distribution with $\beta = 0.1$. We make 100 realisations for each case and calculate the estimators $\hat{\theta}$. We then calculate their sample mean $\mathbb{E} \hat{\theta}$ and sample variance $\text{var}(\hat{\theta})$. We investigate the estimators in regards to bias and consistency in regards to the variance of the sample and the window size.

All iterative methods were started at $\mu_0 = \mu + 0.2$ and $\sigma_0 = \sigma + 0.2$.

3.4.2. HT estimator

In this section we have a sample of l_i^- for $i = 1, \dots, n$ uncensored lengths. To estimate the parameters we use an HT estimator with the ideas introduced in Section 3.2. The weights are

$$\begin{aligned} \pi_-(L) &= \int_{S^{d-1}} |W \ominus \text{lin}(L, U)| f_U(u) \, du \\ &= \int_0^{2\pi} \int_0^\pi |W \ominus \text{lin}(L, U)| f_U(\varphi, \vartheta) \, d\vartheta \, d\varphi. \end{aligned}$$

The integral will be solved by using standard quadrature formulas. The intervals $[0, \pi]$ and $[0, 2\pi]$ were discretized using 50 steps.

Since estimating μ is estimating the first moment of $\ln(L)$ we can directly use the HT estimator and get

$$\hat{\mu}_{HT} = \frac{1}{n_{HT}} \sum_{i=1}^n \frac{\ln(l_i^-)}{\pi_-(l_i^-)}$$

with

$$n_{HT} = \sum_{i=1}^m \frac{1}{\pi_-(L_i)}$$

to normalise the sum. The estimator $\hat{\mu}_{HT}$ is ratio unbiased.

To estimate σ we have to estimate the standard deviation of $\ln(L)$. For a sample l_i $i = 1, \dots, n$ from a log-normal distribution, the standard ML estimator is

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (l_i - \hat{\mu})^2}$$

where $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n l_i$ is the ML estimator of μ . We now reweighed this sum to get

$$\hat{\sigma}_{HT1} = \sqrt{\frac{1}{n_{HT}} \sum_{i=1}^n \frac{(\ln(l_i^-) - \hat{\mu}_{HT})^2}{\pi_-(l_i^-)}}$$

as the HT estimator. We now assume that the weights $\frac{1}{\pi_-(l_i^-)}$ where independent of l_i^- . In this case a correction for the bias is given by

$$\hat{\sigma}_{HT2} = \sqrt{\frac{1}{n_{HT} - \frac{m_{HT}}{n_{HT}}} \sum_{i=1}^n \frac{(\ln(l_i^-) - \hat{\mu}_{HT})^2}{\pi_-(l_i^-)}}$$

with

$$m_{HT} = \sum_{i=1}^n \frac{1}{\pi_-(l_i^-)^2}$$

For the simulation study results for the parameter $\hat{\mu}_{HT}$ we will look at Table 3.2, where we see the estimation results for $\mathbb{E}L = 100$ and $\text{std}L = 30$. We see that for $W = 100$ the estimators seem to systematically underestimate the true parameter. This underestimation vanishes with a bigger window size. If we look at the estimation for results for $\text{std}L = 10$ we see similar results, except that the estimators have a lower variance overall and the underestimation is less severe. For $\text{std}L = 50$ we see the opposite effect. These results can be found in Tables A.1 and A.3.

	$E \hat{\mu}_{HT}$	$\text{var}(\hat{\mu}_{HT})$
$W = 100$		
$n = 20$	4.3842	0.0072
$n = 50$	4.4212	0.0079
$n = 100$	4.4462	0.0073
$n = 500$	4.4801	0.0034
$W = 200$		
$n = 20$	4.5432	0.0044
$n = 50$	4.5556	0.0018
$n = 100$	4.5562	0.0016
$n = 500$	4.5585	0.0004
$W = 300$		
$n = 20$	4.5803	0.0057
$n = 50$	4.5678	0.0022
$n = 100$	4.5668	0.0009
$n = 500$	4.5630	0.0002

Table 3.2.: Simulation study results for different window sizes W and sample sizes n for $\hat{\mu}_{HT}$, with true $\mu = 4.5621$ and $\sigma = 0.2936$

For the results regarding the estimators $\hat{\sigma}_{HT1}$ and $\hat{\sigma}_{HT2}$ we look at Table 3.3. We see that both tend to underestimate the true parameter as well. The estimator $\hat{\sigma}_{HT2}$ seems to have a smaller bias especially for smaller sample sizes but it has a higher variance. The difference in bias and variance vanishes for higher sample sizes. We conclude that $\hat{\sigma}_{HT2}$ should be used to estimate σ . These results are similar for $\text{std } L = 10$ in Table A.2 and $\text{std } L = 50$ in Table A.4, only the variance of the estimator rises and falls with the variance in the sample.

Remark

One might ask why the estimator $\hat{\sigma}_{HT1}$ was even investigated, since it is dominated by $\hat{\sigma}_{HT2}$. The reason is, that in the context of weighted statistics the estimator for the mean is simply the reweighed sum. Estimating the variance is not as straight forward.

Depending of what kind of weights one deals with, different estimators are preferred. In this part we are dealing with probability weights, in other cases we might deal with frequency weights, where different estimator would be better. That being said, implementing a general reweighed variance estimator needs to deal with multiple cases. Some software packages do not do that, and will use the simple estimator $\hat{\sigma}_{HT1}^2$ for all reweighed variances. A user of such software should keep this in mind.

	$E \hat{\sigma}_{HT1}$	$\text{var}(\hat{\sigma}_{HT1})$	$E \hat{\sigma}_{HT2}$	$\text{var}(\hat{\sigma}_{HT2})$
$W = 100$				
$n = 20$	0.1787	0.0006	0.1913	0.0011
$n = 50$	0.1937	0.0005	0.2058	0.0013
$n = 100$	0.2061	0.0005	0.2168	0.0014
$n = 500$	0.2276	0.0004	0.2325	0.0006
$W = 200$				
$n = 20$	0.2636	0.0017	0.2716	0.0018
$n = 50$	0.2757	0.0011	0.2791	0.0011
$n = 100$	0.2822	0.0006	0.2840	0.0006
$n = 500$	0.2890	0.0003	0.2893	0.0003
$W = 300$				
$n = 20$	0.2822	0.0020	0.2897	0.0022
$n = 50$	0.2897	0.0009	0.2930	0.0009
$n = 100$	0.2901	0.0005	0.2914	0.0005
$n = 500$	0.2946	0.0001	0.2949	0.0001

Table 3.3.: Simulation study results for different window sizes W and sample sizes n for $\hat{\sigma}_{HT1}$ and $\hat{\sigma}_{HT2}$, with true $\mu = 4.5621$ and $\sigma = 0.2936$

3.4.3. Associated point rule

In the setting for the associated point rule we have a censored sample \tilde{l}_i^* with $i = 1, \dots, n$. We assume we have l_i with $i = 1, \dots, n_1$ uncensored fibres and \tilde{l}_i with $i = 1, \dots, n_2$ censored fibres. For the EM iterations we have to iterate with Equation (3.4). For the maximisation in each step we calculate the gradient and get

$$\begin{aligned} \frac{\partial}{\partial \mu} Q(\theta | \theta_0, \tilde{l}^*) &= \sum_{i=1}^{n_1} \frac{2(\ln(l_i) - \mu)}{2\sigma^2} + \sum_{i=1}^{n_2} \frac{2(\mathbb{E}_{\theta_0}(\ln(L_i) | \tilde{L}_i = \tilde{l}_i) - \mu)}{2\sigma^2} \\ \frac{\partial}{\partial \sigma} Q(\theta | \theta_0, \tilde{l}^*) &= \sum_{i=1}^{n_1} \frac{-1}{\sigma} + \frac{(\ln(l_i) - \mu)^2}{\sigma^3} \\ &\quad + \sum_{i=1}^{n_2} \frac{-1}{\sigma} + \frac{\mathbb{E}_{\theta_0}(\ln(L_i)^2 | \tilde{L}_i = \tilde{l}_i) - \mu \mathbb{E}_{\theta_0}(\ln(L_i) | \tilde{L}_i = \tilde{l}_i) + \mu^2}{\sigma^3} \end{aligned}$$

and calculate the roots. We get

$$\begin{aligned} \mu &= \frac{1}{n_1 + n_2} \left(\sum_{i=1}^{n_1} \ln(l_i) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0}(\ln(L_i) | \tilde{L}_i = \tilde{l}_i) \right) \\ \sigma^2 &= \frac{\sum_{i=1}^{n_1} (\ln(l_i) - \mu)^2 + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0}(\ln(L_i)^2 | \tilde{L}_i = \tilde{l}_i) - \mu \mathbb{E}_{\theta_0}(\ln(L_i) | \tilde{L}_i = \tilde{l}_i) + \mu^2}{n_1 + n_2} \end{aligned}$$

in each iteration.

The conditional means

$$\mathbb{E}_{\theta_0} \left[\ln(L)^k | \tilde{L} = \tilde{l} \right] = \int_{\tilde{l}}^{\infty} \int_0^{2\pi} \int_0^{\pi} \ln(l)^k f_{L|\tilde{L},U}(l|\tilde{L}, \varphi, \vartheta, \theta_0) f_U(\varphi, \vartheta)$$

are calculated by using numeric quadrature. We first use $b = 10 \cdot W$ as an upper bound for L , we have

$$\mathbb{E}_{\theta_0} \left[\ln(L)^k | \tilde{L} = \tilde{l} \right] \approx \int_{\tilde{l}}^b \int_0^{2\pi} \int_0^{\pi} \ln(l)^k f_{L|\tilde{L},U}(l|\tilde{L}, \varphi, \vartheta) f_U(\varphi, \vartheta)$$

and approximated this integrals with the composite $\frac{3}{8}$ -rule. We discretize the angle intervals $[0, \pi]$ and $[0, 2\pi]$ with fifty steps. The length interval $[\tilde{l}, b]$ with 100 steps.

We do this iteration until we reach convergence. Convergence is reached if in the j th-step $|\hat{\mu}_j - \hat{\mu}_{j-1}| < 1e-6$ and $|\hat{\sigma}_j - \hat{\sigma}_{j-1}| < 1e-6$. This iteration leads to an estimator $\hat{\theta}_{EM}^a = (\hat{\mu}_{EM}^a, \hat{\sigma}_{EM}^a)$.

To estimate the fisher information we use

$$\mathcal{I}(\theta_{EM}^a, \tilde{L}) = -\mathcal{H}(Q(\theta_{EM}^a | \theta_{EM}^a)) - \text{cov}(\nabla_{\theta} l(\theta_{EM}^a | L)),$$

where the components of $\mathcal{H}(Q(\theta_{EM}^a | \theta_{EM}^a))$ can be found in Equation (3.19) and the components of $\text{cov}(\nabla_{\theta} l(\theta_{EM}^a | L))$ in Equations (3.20), (3.22) and (3.21). Here we will look at the estimators $\hat{\mathcal{I}}_{a,1,1}^{-1}$ and $\hat{\mathcal{I}}_{a,2,2}^{-1}$ which should estimate a lower bound for the variances of $\hat{\mu}_{EM}^a$ and $\hat{\sigma}_{EM}^a$.

For the simulation study results we will first look at Table 3.4, where we see the estimation results for $\mathbb{E}L = 100$ and $\text{std}L = 30$. We first note that the variance of the estimator is lower with a higher sample size, this fits about the consistency of the estimator. The variance is also lower for a higher window size. That means that with less censoring we will have a better estimation result. Lastly we see that with a larger sample size we will need fewer iterations to achieve convergence. The influence of the sample size on the number of iterations is lowered with a bigger window size.

If we compare these results to the ones seen in Table A.7 for $\mu = 4.5621$ and $\sigma = 0.2936$, we see the same results but note that the variance of the estimators is higher since the variance in the data is higher. This is consistent with the results seen in Table A.5 for $\mu = 4.6002$ and $\sigma = 0.0998$, where the variance of the estimator lower.

For the estimation of the inverse Fisher information we start with the case $\mathbb{E}L = 100$ and $\text{std}L = 30$ in Table 3.5. To quantify the quality of these estimators we have to compare them the estimated variances $\text{var} \hat{\mu}_a$ and $\text{var} \hat{\sigma}_a$ in Table 3.4. The variances of these estimators should be bounded from below by their respective component of the inverted Fisher information. If we look at the simulation study

	$E \hat{\mu}_a$	$\text{var}(\hat{\mu}_a)$	$E \hat{\sigma}_a$	$\text{var}(\hat{\sigma}_a)$	iterations
$W = 100$					
$n = 20$	4.4227	0.0193	0.2230	0.0150	84.3
$n = 50$	4.4163	0.0057	0.2514	0.0072	72.42
$n = 100$	4.4131	0.0033	0.2549	0.0029	65.37
$n = 500$	4.4179	0.0006	0.2601	0.0005	57.23
$W = 200$					
$n = 20$	4.4946	0.0107	0.2745	0.0032	18.23
$n = 50$	4.5007	0.0037	0.2844	0.0011	18.01
$n = 100$	4.5046	0.0019	0.2896	0.0006	16.83
$n = 500$	4.5049	0.0003	0.2951	0.0002	16.11
$W = 300$					
$n = 20$	4.5532	0.0066	0.2707	0.0021	11.36
$n = 50$	4.5426	0.0026	0.2766	0.0009	10.52
$n = 100$	4.5357	0.0012	0.2804	0.0005	9.88
$n = 500$	4.5328	0.0003	0.2833	9.30e-5	9.21

Table 3.4.: Simulation Study Results for the parameters for the associated point rule, with $\mu = 4.5621$ and $\sigma = 0.2936$

results we see that this is true for a sample size between $n = 200$ or $n = 500$. Before that the estimated value may still be used to get a general idea of the quality of the estimator.

3.4.4. Plus-sampling

In this setting we will investigate the EM estimator if we use plus sampling and a reweighed density. We have a censored sample \tilde{l}_i^{*+} with $i = 1, \dots, n$. We assume we have l with $i = 1, \dots, n_1$ uncensored lengths and \tilde{l} with $i = 1, \dots, n_2$ censored lengths with $n = n_1 + n_2$. For the EM algorithm we have to maximise Equation (3.5) iteratively. For this we calculate the gradient

$$\begin{aligned} \frac{\partial}{\partial \mu} Q(\theta|\theta_0) &= \frac{-n\xi e^{\mu + \frac{\sigma^2}{2}}}{\xi e^{\mu + \frac{\sigma^2}{2}} + |W|} - \frac{\mu n}{\sigma^2} + \\ &\quad \frac{\sum_{i=1}^{n_1} \ln(l_i) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i)|\tilde{L}_i = \tilde{l}_i]}{\sigma^2} \\ \frac{\partial Q(\theta|\theta_0)}{\partial \sigma} &= \frac{-n\xi \sigma e^{\mu + \frac{\sigma^2}{2}}}{\xi e^{\mu + \frac{\sigma^2}{2}} + |W|} - \frac{1}{\sigma} + \frac{\sum_{i=1}^{n_1} \ln(l_i)^2 + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i)^2|\tilde{L}_i = \tilde{l}_i]}{\sigma^3} - \\ &\quad \frac{2\mu \left(\sum_{i=1}^{n_1} \ln(l_i) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i)|\tilde{L}_i = \tilde{l}_i] \right)}{\sigma^3} + \frac{n\mu^2}{\sigma^3}. \end{aligned}$$

	$E\mathcal{I}_a^{-1}{}_{1,1}$	$\text{var}(\mathcal{I}_a^{-1}{}_{1,1})$	$E\mathcal{I}_a^{-1}{}_{2,2}$	$\text{var}(\mathcal{I}_a^{-1}{}_{2,2})$
$W = 100$				
$n = 20$	0.0224	0.0019	0.01643	0.0010
$n = 50$	0.0066	3.45e-5	0.0062	4.93e-5
$n = 100$	0.0030	3.21e-6	0.0027	4.17e-6
$n = 500$	0.0006	1.67e-8	0.0005	1.86e-8
$W = 200$				
$n = 20$	0.0071	8.67e-6	0.0038	2.35e-6
$n = 50$	0.0029	6.07e-7	0.0017	1.79e-7
$n = 100$	0.0015	7.60e-8	0.0008	2.32e-8
$n = 500$	0.0003	6.99e-10	0.0002	2.02e-10
$W = 300$				
$n = 20$	0.0052	3.50e-6	0.0026	1.08e-6
$n = 50$	0.0021	2.14e-7	0.0010	6.04e-8
$n = 100$	0.0011	2.78e-8	0.0005	8.17e-9
$n = 500$	0.0002	2.34e-10	0.0001	6.42e-11

Table 3.5.: Simulation Study Results for the fisher information for the associated point rule, with $\mu = 4.5621$ and $\sigma = 0.2936$

We wish to find the roots of this gradient. We can reformulate this to finding the roots of

$$g_I(\mu, \sigma) := -\xi e^{\mu + \frac{\sigma^2}{2}} \sigma^2 - \left(\xi e^{\mu + \frac{\sigma^2}{2}} + |W| \right) \mu + \left(\xi e^{\mu + \frac{\sigma^2}{2}} + |W| \right) \frac{\sum_{i=1}^{n_1} \ln(l_i) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i) | \tilde{L}_i = \tilde{l}_i]}{n}$$

and

$$g_{II}(\mu, \sigma) := -\xi e^{\mu + \frac{\sigma^2}{2}} \sigma^4 - \left(\xi e^{\mu + \frac{\sigma^2}{2}} + |W| \right) \sigma^2 + \left(\xi e^{\mu + \frac{\sigma^2}{2}} + |W| \right) \frac{\sum_{i=1}^{n_1} \ln(l_i)^2 + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i)^2 | \tilde{L}_i = \tilde{l}_i]}{n} - \left(\xi e^{\mu + \frac{\sigma^2}{2}} + |W| \right) \frac{2\mu \left(\sum_{i=1}^{n_1} \ln(l_i) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i)^2 | \tilde{l}_i = \tilde{l}_i] \right)}{n} \left(\xi e^{\mu + \frac{\sigma^2}{2}} + |W| \right) \mu^2.$$

To calculate the roots we will use a newton type method. Therefore we have to calculate the derivatives and obtain

$$\begin{aligned}\frac{\partial g_I}{\partial \mu}(\mu, \sigma) &= -\xi \sigma^2 e^{\mu + \frac{\sigma^2}{2}} + \xi e^{\mu + \frac{\sigma^2}{2}} \frac{\sum_{i=1}^{n_1} \ln(l_i) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L) | \tilde{L} = \tilde{l}_i]}{n} \\ &\quad - \xi e^{\mu + \frac{\sigma^2}{2}} (\mu + 1) - |W| \\ \frac{\partial g_I}{\partial \sigma}(\mu, \sigma) &= -\xi \sigma (\sigma^2 + 2) e^{\mu + \frac{\sigma^2}{2}} + \xi \sigma e^{\mu + \frac{\sigma^2}{2}} \frac{\sum_{i=1}^{n_1} \ln(l_i) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L) | \tilde{L} = \tilde{l}_i]}{n} \\ &\quad - \xi e^{\mu + \frac{\sigma^2}{2}} \mu \sigma\end{aligned}$$

for g_I and

$$\begin{aligned}\frac{\partial g_{II}}{\partial \mu}(\mu, \sigma) &= -\xi \sigma^2 (\sigma^2 + 1) e^{\mu + \frac{\sigma^2}{2}} + c \mu (\mu + 2) e^{\mu + \frac{\sigma^2}{2}} + 2 \mu |W| \\ &\quad + \xi e^{\mu + \frac{\sigma^2}{2}} \frac{\sum_{i=1}^{n_1} \ln(l_i)^2 + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i)^2 | \tilde{L}_i = \tilde{l}_i]}{n} \\ &\quad - 2 \left(\xi e^{\mu + \frac{\sigma^2}{2}} (\mu + 1) + |W| \right) \frac{\sum_{i=1}^{n_1} \ln(l_i) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(L_i) | \tilde{L}_i = \tilde{l}_i]}{n} \\ \frac{\partial g_{II}}{\partial \sigma}(\mu, \sigma) &= -\xi \sigma^3 (\sigma^2 + 4) e^{\mu \frac{\sigma^2}{2}} - \xi \sigma (\sigma^2 + 2) e^{\mu \frac{\sigma^2}{2}} \\ &\quad + \xi \sigma e^{\mu \frac{\sigma^2}{2}} \frac{\sum_{i=1}^{n_1} \ln(l_i)^2 + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(l_i)^2 | \tilde{L}_i = \tilde{l}_i]}{n} \\ &\quad - 2 \xi \mu \sigma e^{\mu \frac{\sigma^2}{2}} \frac{\sum_{i=1}^{n_1} \ln(l_i) + \sum_{i=1}^{n_2} \mathbb{E}_{\theta_0} [\ln(l_i) | \tilde{L}_i = \tilde{l}_i]}{n} + \xi \mu^2 \sigma e^{\mu \frac{\sigma^2}{2}} - 2 \sigma |W|\end{aligned}$$

for g_{II} . With this we have the Newton iteration

$$\begin{pmatrix} \mu_{n+1} \\ \sigma_{n+1} \end{pmatrix} = \begin{pmatrix} \mu_n \\ \sigma_n \end{pmatrix} - \begin{pmatrix} \frac{\partial g_I}{\partial \mu}(\mu_n, \sigma_n) & \frac{\partial g_I}{\partial \sigma}(\mu_n, \sigma_n) \\ \frac{\partial g_{II}}{\partial \mu}(\mu_n, \sigma_n) & \frac{\partial g_{II}}{\partial \sigma}(\mu_n, \sigma_n) \end{pmatrix}^{-1} \begin{pmatrix} g_I(\mu_n, \sigma_n) \\ g_{II}(\mu_n, \sigma_n) \end{pmatrix}$$

which stops if the update

$$\left\| \begin{pmatrix} \frac{\partial g_I}{\partial \mu}(\mu_n, \sigma_n) & \frac{\partial g_I}{\partial \sigma}(\mu_n, \sigma_n) \\ \frac{\partial g_{II}}{\partial \mu}(\mu_n, \sigma_n) & \frac{\partial g_{II}}{\partial \sigma}(\mu_n, \sigma_n) \end{pmatrix}^{-1} \begin{pmatrix} g_I(\mu_n, \sigma_n) \\ g_{II}(\mu_n, \sigma_n) \end{pmatrix} \right\| < 10^{-6}.$$

This Newton iteration yields a new set of parameters in each step of the EM Algorithm.

The conditional means

$$\mathbb{E}_{\theta_0} \left[\ln(L)^k | \tilde{L} = \tilde{l} \right] = \int_{\tilde{l}}^{\infty} \int_0^{2\pi} \int_0^{\pi} \ln(l)^k f_{L|\tilde{L},U}(l|\tilde{L}, \varphi, \vartheta, \theta_0) f_U(\varphi, \vartheta)$$

calculated by using numeric quadrature like for the associated point rule. We first use $b = 10 \cdot W$ as an upper bound for L , we have

$$\mathbb{E}_{\theta_0} \left[\ln(L)^k | \tilde{L} = \tilde{l} \right] \approx \int_{\tilde{L}}^b \int_0^{2\pi} \int_0^{\pi} \ln(l)^k f_{L|\tilde{L},U}(l|\tilde{L}, \varphi, \vartheta) f_U(\varphi, \vartheta)$$

and approximated this integrals with the composite $\frac{3}{8}$ -rule. We discretize the angle intervals $[0, \pi]$ and $[0, 2\pi]$ with fifty steps. The length interval $[\tilde{l}, b]$ with 100 steps.

Convergence is reached if in the j th-step $|\hat{\mu}_j - \hat{\mu}_{j-1}| < 1e-6$ and $|\hat{\sigma}_j - \hat{\sigma}_{j-1}| < 1e-6$. This iteration leads to an estimator $\hat{\theta}_{EM}^+ = (\hat{\mu}_{EM}^+, \hat{\sigma}_{EM}^+)$.

For the variance estimation we use the Louis method and directly implement the formula in (3.24).

The simulation study results are very similar to the associated point rule. We will first look at Table 3.6, where we see the estimation results for $\mathbb{E}L = 100$ and $\text{std}L = 30$. The variance of the estimator is lower with a higher sample size, fitting the consistency of the estimator. The variance is also lower for a higher window size. That means that with less censoring we will have a better estimation result. Lastly we see that with a larger window size we will need fewer iterations to achieve convergence. The influence of the sample size on the number of iterations is lower than the sample size. We conclude that more censoring leads to more iterations being necessary to achieve convergence.

If we compare these results to the ones seen in Table A.11 for $\mu = 4.5621$ and $\sigma = 0.2936$, we see the same results but note that the variance of the estimators is higher since the variance in the data is higher. This is consistent with the results seen in Table A.5 for $\mu = 4.6002$ and $\sigma = 0.0998$, where the variance of the estimator lower.

For the estimation of the inverse Fisher information we start with the case $\mathbb{E}L = 100$ and $\text{std}L = 30$ in Table 3.7. To quantify the quality of these estimators we have to compare them the estimated variances $\text{var} \hat{\mu}_+$ and $\text{var} \hat{\sigma}_+$ in Table 3.6. The variances of these estimators should be bounded from below by their respective component of the inverted Fisher information. If we look at the simulation study results we see that this is true for a sample size between $n = 200$ or $n = 500$. Before that the estimated value may still be used to get a general idea of the quality of the estimator. Looking at Tables A.10 and A.12 we see the similar results for the other parameter sets.

	$E \hat{\mu}_+$	$\text{var}(\hat{\mu}_+)$	$E \hat{\sigma}_+$	$\text{var}(\hat{\sigma}_+)$	iterations
$W = 100$					
$n = 20$	4.4532	0.0291	0.2112	0.0184	92.52
$n = 50$	4.4641	0.0088	0.2471	0.0051	87.19
$n = 100$	4.4591	0.0034	0.2689	0.0026	72.75
$n = 500$	4.4575	0.0007	0.2800	0.0005	68.99
$W = 200$					
$n = 20$	4.4755	0.01264	0.2779	0.0064	32.12
$n = 50$	4.4833	0.0073	0.3038	0.0019	29.64
$n = 100$	4.4810	0.0025	0.3073	0.0008	27.69
$n = 500$	4.4790	0.0005	0.3157	0.0002	25.24
$W = 300$					
$n = 20$	4.5402	0.0073	0.2679	0.0031	16.47
$n = 50$	4.5302	0.0028	0.2700	0.0009	14.07
$n = 100$	4.5222	0.0014	0.2752	0.0004	13.41
$n = 500$	4.5175	0.0003	0.2792	0.0001	12.43

Table 3.6.: Simulation Study Results for the parameters for the plus sampling rule, with $\mu = 4.5621$ and $\sigma = 0.2936$

	$E \hat{\mathcal{I}}_{+ 1,1}^{-1}$	$\text{var}(\hat{\mathcal{I}}_{+ 1,1}^{-1})$	$E \hat{\mathcal{I}}_{+ 2,2}^{-1}$	$\text{var}(\hat{\mathcal{I}}_{+ 2,2}^{-1})$
$W = 100$				
$n = 20$	0.0175	0.0005	0.0115	0.0002
$n = 50$	0.0073	1.56e-5	0.0058	1.43e-5
$n = 100$	0.0039	1.84e-6	0.0030	1.99e-6
$n = 500$	0.0008	1.36e-8	0.0006	1.55e-8
$W = 200$				
$n = 20$	0.0105	4.12e-5	0.0059	1.97e-5
$n = 50$	0.0049	3.16e-6	0.0027	1.23e-6
$n = 100$	0.0024	3.17e-7	0.0013	9.05e-8
$n = 500$	0.0005	2.59e-9	0.0003	7.19e-10
$W = 300$				
$n = 20$	0.0062	6.90e-6	0.0030	2.25e-6
$n = 50$	0.0024	2.99e-7	0.0011	9.29e-8
$n = 100$	0.0012	3.83e-8	0.0006	1.15e-8
$n = 500$	0.0002	3.11e-10	0.0001	1e-10

Table 3.7.: Simulation Study Results for the fisher information for the associated point rule, with $\mu = 4.5621$ and $\sigma = 0.2936$

3.4.5. Influence of the orientation distribution on the estimator

Up to this point we have assumed that the orientation distribution is known perfectly. That means that the bias factor, which depends on the orientation distribution, can be computed exactly and the true distribution is known. To evaluate the robustness of the estimators in regards to a wrong β we used the same data as above but used $\beta = 10$ for the estimation. Note that this is an extreme choice, for a fibre sample with a strong main direction such a strong miss estimation would be noticed. We did the simulation study for the case $\mathbb{E}[L] = 100$ and $\text{std}(L) = 30$. For the EM estimators we see the results in Table 3.8. We see that both of them underestimate μ and overestimate σ . But they still are close to the true value and the variance is falling. The estimators seem to be robust under the misspecified β . In Table 3.9 we see the same results for the HT type estimators. They show a robust behaviour as well. The underestimation of σ is a little stronger. Otherwise the results are comparable to the correct specified case.

n	$\mathbb{E} \hat{\mu}_a$	$\text{var}(\hat{\mu}_a)$	$\mathbb{E} \hat{\sigma}_a$	$\text{var}(\hat{\sigma}_a)$	$\mathbb{E} \hat{\mu}_+$	$\text{var}(\hat{\mu}_+)$	$\mathbb{E} \hat{\sigma}_+$	$\text{var}(\hat{\sigma}_+)$
20	4.4963	0.01074	0.2746	0.0032	4.4800	0.0128	0.2792	0.0065
50	4.5025	0.0037	0.2846	0.0011	4.4886	0.0074	0.3048	0.0020
100	4.5064	0.0019	0.2899	0.0006	4.4860	0.0025	0.3082	0.0009
500	4.5066	0.0003	0.2953	0.0002	4.4839	0.0005	0.3167	0.0002

Table 3.8.: Estimation results with a misspecified orientation distribution for the EM estimators, with window size $W = 200$ and true parameters $\mu = 4.5621$ and $\sigma = 0.2936$

n	$\mathbb{E} \hat{\mu}_{HT}$	$\text{var}(\hat{\mu}_{HT})$	$\mathbb{E} \hat{\sigma}_{HT2}$	$\text{var}(\hat{\sigma}_{HT2})$
20	4.5429	0.0042	0.2701	0.0017
50	4.5545	0.0017	0.2771	0.0010
100	4.5544	0.0014	0.2814	0.0005
500	4.5554	0.0002	0.2855	0.0002

Table 3.9.: Estimation with a misspecified orientation distribution for the HT estimators, with window size $W = 200$ and true parameters $\mu = 4.5621$ and $\sigma = 0.2936$

3.4.6. Alternative implementations

An alternative to this method is to sample from the distribution of $L|\tilde{L}$ and use the Monte Carlo method to estimate the conditional means. The result would be a so called Monte Carlo EM. We did not follow this approach because in the plus-sampling case sampling from the distribution is computationally expensive. Also we would have to make more EM iterations then strictly necessary, since in the MCEM we have to take the mean of the last iterations to get a stable estimator. An

implantation using a slice sampler and 1000 samples to estimate the means showed to be significantly slower than the approach using numeric quadrature without leading to better estimators.

3.4.7. HT weight or kernel density

In this part we wish to return to the setting of a Horvitz Thompson estimator. We will ignore all censored lengths and have a biased sample L_i^- with $i = 1, \dots, n$. The true length L has the density $f(l)$. We will formulate a non-parametric estimator using a kernel density estimator for this sample.

Assume K is a kernel and h is bandwidth. For a given sample L_1, \dots, L_n the kernel density estimator is then defined as

$$\hat{f}_h(l) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{l - L_i}{h}\right).$$

If we plug in the biased L_i^- with $i = 1, \dots, n$ directly we will have a biased estimator.

Instead we look at the weighted sample $(L_i^-, \pi^-(L_i^-))$ and follow the approach given in [5]. This approach leads to a weighted kernel density estimator

$$\hat{f}_{HT,h}(l) = \frac{\sum_{i=1}^n K\left(\frac{l - L_i^-}{h}\right) / \pi^-(L_i^-)}{h \sum_{i=1}^{n_1} 1 / \pi^-(L_i^-)}.$$

The consistency and asymptotic normality were proven in [5] as well.

To investigate the quality of this estimator we made a small simulation study with a line segment study $E(L) = 100$ and $std(L) = 30$ and a window size of $W = 200$. We look at 100 realisations of this process with $n = 20, 50, 100, 500$ samples. We compute the kernel density estimator using Matlab and the built in `ksdensity` function. The bandwidth is automatically chosen by Matlab. We then compute $|\hat{f}_{HT,h} - f_L|_2^2$ for each realisation and estimate the MSE. The results can be seen in Table 3.10, we see that the error goes down with a higher sample size.

$W = 200$	$E \ \hat{f}_{HT,h} - f\ _2^2$
$n = 20$	0.001027
$n = 50$	0.000499
$n = 100$	0.000311
$n = 500$	0.000097

Table 3.10.: Simulation Study Results for the weighted kernel density estimator, where the true density is a log-normal density with parameters $\mu = 4.5621$ and $\sigma = 0.2936$

4. Estimation from fibre endpoints

The methods on fully segmented fibres described in the previous chapter show great performance in simulation studies. The problem is, that a full fibre segmentation is not possible, at least not for interesting window size to fibre length ratios.

A method to segment the fibre endpoints using Gaussian curvature was introduced in [10]. In [12] a model based on a Neymann Scott cluster process for a fibre endpoint process was introduced. They also developed an estimator for the length distribution based on this model and the reduced second moment measure.

We will investigate this estimator and introduce a new estimator based on minimum contrast estimation.

4.1. Fibre endpoint process

We will model the fibre endpoint as a cluster process. The germs of each cluster are the midpoints of the fibres. We model these midpoints as a Poisson point process \tilde{X} with intensity $\tilde{\gamma}$. We then add two points to each germ as the endpoints. These points have a random distance $L > 0$ and a random orientation $U \in S^{d-1}$. Note that this model is closely related to the line segment model already discussed. If you take a line segment process and take just the endpoints of each segment we will have the proposed endpoint process here.

The resulting end point process has an intensity of $\gamma = 2\tilde{\gamma}$. It is only isotropic if the orientation distribution is uniform on the sphere.

4.2. Summary Statistics

To investigate the endpoint process we will derive closed formulas for the reduced second moment measure, Ripley's K and the pair correlation function.

These results are cited from [12].

Theorem 4.1. *Let X be a fibre endpoint process with intensity $\gamma > 0$. Let the random variables $L > 0$ be the length of the fibres and $U \in S^2$ be the orientation of the fibres with the random angles $\varphi \in [0, 2\pi)$, $\vartheta \in [0, \pi)$. Let $I_L \subset \mathbb{R}^+$, $I_\varphi \subset [0, 2\pi)$ and $I_\vartheta \subset [0, \pi)$. Consider sets $B(I_L, I_\varphi, I_\vartheta) \subset \mathbb{R}^3$ of the form*

$$B(I_L, I_\vartheta, I_\varphi) = \left\{ l \cdot (\sin(\vartheta) \cos(\varphi), \sin(\vartheta) \sin(\varphi), \cos(\vartheta))^T \mid l \in I_L, \vartheta \in I_\vartheta, \varphi \in I_\varphi \right\}.$$

if $I_\vartheta = [0, \pi)$ and $I_\varphi = [0, 2\pi)$ we use

$$B(I_L) = B(I_L, I_\vartheta, I_\varphi)$$

Then following by Equation (2.20) we get

$$\mathbb{E}^0(X(B(I_L, I_\vartheta, I_\varphi))) = \gamma \nu(B(I_L, I_\vartheta, I_\varphi)) + P(L \in I_L, \varphi \in I_\varphi, \vartheta \in I_\vartheta)$$

since $p_n = 0$ for all $n \neq 2$ and $p_2 = 1$. If the orientation and the length are independent we get

$$\mathbb{E}^0(X(B(I_L, I_\vartheta, I_\varphi))) = \gamma \nu(B(I_L, I_\vartheta, I_\varphi)) + P(L \in I_L)P(\varphi \in I_\varphi, \vartheta \in I_\vartheta)$$

By using the reduced second moment measure from Equation (2.17) we get

$$\gamma \mathcal{K}(B(I_L, I_\vartheta, I_\varphi)) = \gamma \nu(B(I_L, I_\vartheta, I_\varphi)) + P(L \in I_L)P(\varphi \in I_\varphi, \vartheta \in I_\vartheta) \quad (4.1)$$

and for $I_L = [0, r]$, $I_\vartheta = [0, \pi)$ and $I_\varphi = [0, 2\pi)$ we have

$$K(r) = \mathcal{K}(B(I_L))$$

we get

$$\gamma K(r) = \gamma \nu(B(I_L)) + P(L < r) \quad (4.2)$$

as a model for Ripley's- K function or also

$$\gamma K(r, u, \varphi) = \gamma \nu(B(0, r_1, u, \varphi)) + P(L < r)P(U \in S(u, \varphi)) \quad (4.3)$$

for the directed K function.

4.3. Estimators

We will first discuss the non parametric estimator introduced in [12]. We will then introduce a parametric minimum contrast estimator.

4.3.1. Non parametric estimation

Assume we have an endpoint process with intensity γ

The non parametric estimators work by taking Equation (4.1) and plugging in concrete intervals I_L, I_ϑ and I_φ . Since we are interested in the length distribution of the fibres we transform the equation and get

$$P(L \in I_L) = \frac{\gamma \mathcal{K}(B(I_L, I_\vartheta, I_\varphi)) - \gamma \nu(B(I_L, I_\vartheta, I_\varphi))}{P(\varphi \in I_\varphi, \vartheta \in I_\vartheta)}. \quad (4.4)$$

Isotropic estimator

The first idea is to take $I_L = [l_1, l_2]$ as an interval in \mathbb{R}^+ . We set $I_\vartheta = [0, \pi]$ and $I_\varphi = [0, 2\pi)$. That means we look at the whole sphere. The resulting model is

$$\begin{aligned} P(L \in [l_1, l_2]) &= \gamma \mathcal{K}(B(I_L)) - \gamma \frac{4}{3} \pi (l_2^3 - l_1^3) \\ &= \frac{\gamma^2 \mathcal{K}(B(I_L))}{\gamma} - \gamma \frac{4}{3} \pi (l_2^3 - l_1^3), \end{aligned}$$

since $P(\varphi \in [0, 2\pi), \vartheta \in [0, \pi]) = 1$. To get an estimator for $P(L \in [l_1, l_2])$ we plug in the estimator $\hat{\kappa}$ for $\gamma^2 \mathcal{K}(B(I_L, I_\vartheta, I_\varphi))$ from Equation (2.19) and the intensity estimator from Equation (2.16) to get

$$\hat{p}_{I_L, iso} = \hat{\gamma} \hat{\kappa}(B(I_L, [0, \pi], [0, 2\pi))) - \hat{\gamma} \frac{4}{3} \pi (l_2^3 - l_1^3) \quad (4.5)$$

as an estimator of $P(L \in [l_1, l_2])$.

Anisotropic estimator

We assumed that the orientation distribution is known in the previous chapter. We do the same here. Let us assume we have ascending $\vartheta_0, \dots, \vartheta_n$ with $\vartheta_0 = 0$ and $\vartheta_n = \pi$ and ascending $\varphi_0, \dots, \varphi_m$ with $\varphi_0 = 0$, $\varphi_m = 2\pi$. With this we have a partition of the angles by $I_{\vartheta_i} = [\vartheta_{i-1}, \vartheta_i]$ and $I_{\varphi_j} = [\varphi_{j-1}, \varphi_j]$ for $i = 1, \dots, n$ and $j = 1, \dots, m$. If we apply this to Equation (4.4) and plug in the same estimators as for Equation (4.5) we get

$$\hat{p}_{I_L, i, j} = \frac{\hat{\gamma} \hat{\kappa}(B(I_L, I_{\vartheta_i}, I_{\varphi_j})) - \gamma \frac{2}{3} (\varphi_j - \varphi_{j-1}) (\cos(\vartheta_{i-1}) - \cos(\vartheta_i)) (l_2^3 - l_1^3)}{P(\varphi \in I_{\varphi_j}, \vartheta \in I_{\vartheta_i})}$$

$n \cdot m$ estimators for $P(L \in [l_1, l_2])$.

In [12] it was proposed to use a weighted sum

$$\hat{p}_{I_L, aniso} = \sum_{i=1, j=1}^{n, m} \frac{\hat{p}_{I_L, i, j}}{w_{i, j}}$$

where the weight is given by

$$w_{i, j} = \frac{\text{var}(\hat{p}_{I_L, i, j})}{\sum_{k=1, l=1}^{n, m} \hat{p}_{I_L, k, l}}.$$

This approach is optimal if the $\hat{p}_{I_L, i, j}$ are independent. The variance was approximated by using a Poisson point process and numerical approximation.

Simulation Study

To compare the isotropic estimator and the anisotropic estimator we made a simulation study. For this we simulated a $3d$ process of line segments with a log-normal length distribution with parameters $\mu = 4.6002$ and $\sigma = 0.0998$ and a Schladitz β distributed orientation with $\beta = 0.1$ in rectangular window W with the die lengths $W_l = 400$ and 600 . The intensity of the fibre process is $1.9895e-5$.

For the estimation we divide the interval $[0, 150]$ into $I_{L,j} = [(j-1)5, j5]$ for $j = 1, \dots, 30$. We estimated $\hat{p}_{I_{L,j},iso}$ and $\hat{p}_{I_{L,j},aniso}$ on each interval.

For the anisotropic estimator we have to discretise the angles as well. For $\varphi \in [0, 2\pi]$ we follow the suggestion given in [12] and always work on the full interval, i.e. $\varphi_0 = 0$ and $\varphi_1 = 2\pi$. For $\vartheta \in [0, \pi]$ we want to achieve an equisize partition of the sphere. We will first discretise the interval $[0, 1]$ with n points $u_i = 1 - \frac{i}{n}$ for $i = 0, \dots, n$. We get a discretization of $[0, \frac{\pi}{2}]$ by setting $\vartheta_i = \text{acos}(u_i)$ with $i = 1, \dots, n$. We do not discretize the lower part of the sphere due to the point symmetry of the orientation distribution.

We made 100 realisations of these processes and calculated the estimators. We then estimated point wise 95% confidence bands by estimating the 2.5% and 95% quantiles for both estimators.

For a window size with $W_l = 400$ we see the results for both estimators in Figure 4.1. We directly see both estimators have a higher variance the further to the right we move on the length axis. Of both estimators the anisotropic estimator has a lower variance. If we look at the same estimation for $W_l = 600$ Figure 4.2 we see that the variance is smaller than for $W_l = 400$. This is not surprising since the estimator for the reduced second moment measure is better for bigger window sizes. We conclude to get a meaningful estimation, the window should have at least 4 times the size of the lengths of interest.

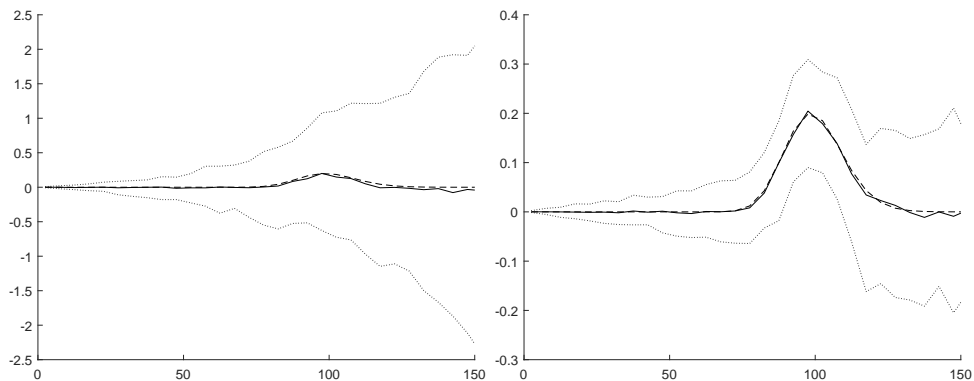


Figure 4.1.: Comparison of the isotropic estimator on the left and the anisotropic estimator on the right for window size $W_l = 400$. The mean of the estimator is solid, the true value is dashed and an estimated 95% region is dotted. Note the different scale in the figures.

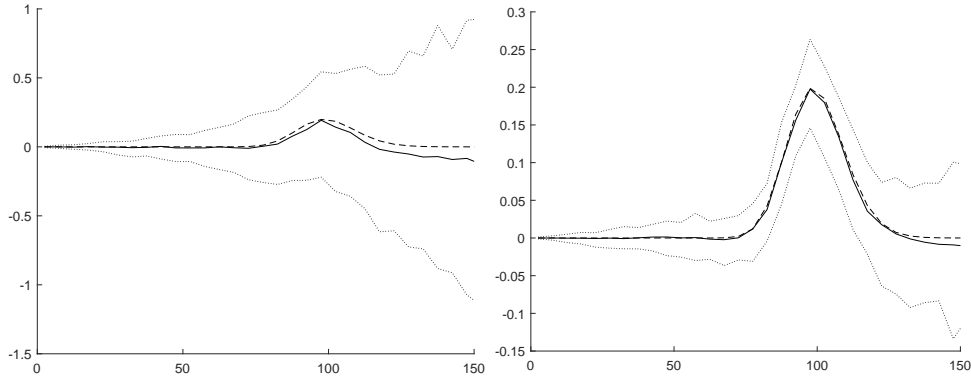


Figure 4.2.: Comparison of the isotropic estimator on the left and the anisotropic estimator on the right for window size $W_l = 600$. The mean of the estimator is solid, the true value is dashed and an estimated 95% region is dotted.

4.3.2. Minimum Contrast estimator

Here we will introduce estimators based on a minimum contrast approach.

Definition 4.1. Let Φ be a simple point process observed in a window W depending on parameters $\theta \in \mathbb{R}^d$. Let $T(r, \theta)$ be a known statistic of this point process. Let $\hat{T}(r)$ be an estimator of this statistic. The minimum contrast estimator is defined as

$$\hat{\theta} = \arg \max_{\theta} \int_{r_1}^{r_2} |T(r, \theta) - \hat{T}(r)|^q dr$$

where $q \in \mathbb{N}$ and $r_1 < r_2$.

In this definition $d(T, \hat{T}) := \int_{r_1}^{r_2} |T(r, \theta) - \hat{T}(r)|^q dr$ is the contrast function. A usual choice is $q = 2$ which leads to distance in the L^2 norm.

We wish to use the minimum contrast estimator to estimate the parameters of the fibre length distribution. For this we assume that the length L has a parametric distribution with parameters $\theta \in \mathbb{R}^d$. We choose the model for the directed Ripley's- K from Equation (4.3) as our summary statistic and get

$$\hat{\theta} = \arg \max_{\theta} \int_{r_1}^{r_2} \left| \sqrt{K(r, u, \varphi, \theta)} - \sqrt{\frac{\hat{K}(r, u, \varphi)}{\hat{\gamma}^2}} \right|^2 dr$$

as the minimum contrast estimator. We use \sqrt{K} in this approach, to control for the variance in the estimator as described in [8].

Here we will especially investigate for fibres with a strong main direction $u \in S^{d-1}$.

The bounds of the integration r_1 and r_2 are usually chosen on an ad-hoc basis. Here we will present a method to chose r_1 and r_2 we first look at the Equation (2.18). We note that the first part of the right hand side $\nu(B(0, r, u, \varphi))$ is the formula for the directed K function of a Poisson point process. All of our length estimator therefore measure the defiance of the estimated K function and the analytically known K function for the Poisson point process. Therefore if $\hat{K}(r) \approx \nu(B(0, r, u, \varphi))$, we cannot detect any influence of the length distribution and the process looks like a Poisson point process. We therefore propose to choose r_1 and r_2 in a way that we see the maximum influence of the length distribution on $K(r, u, \theta)$. Since the length distribution is unknown, we propose to look at Poisson point processes instead. We simulate n_p Poisson point processes with the estimated intensity $\hat{\gamma}$ to estimate a $(1 - p)$ upper confidence band $\hat{K}_{1-p}(r)$. We choose r_1 as the smallest r where $\hat{K}(r, u, \varphi) > \hat{K}_{1-p}(r)$ and r_2 as the biggest r with $\hat{K}(r, u, \varphi) > \hat{K}_{1-p}(r)$. Our reasoning is, that we will get the parts of \hat{K} that are significantly different from a Poisson point process and will therefore find the largest influence of the length distribution.

Simulation Study

In this example we look at a 3d line segment process in a window $W = [0, W_s]^3$ with side lengths $W_s = 400$ and 600 . The midpoint process is a Poisson point process. As a length distribution we choose a log-normal distribution with parameters $\mu = 4.6002$ and $\sigma = 0.0998$. For the orientation distribution we chose a Schladitz β -distribution with $\beta = 0.1$. The intensity of the fibre process is $1.9895e-5$. In the smaller window we expect to see about 1254 fibres and 4232 fibres in the larger window. We ran 100 simulations for each window size. We get the model

$$\gamma K(r, u, \varphi) = \gamma \nu(B(0, r, u, \varphi)) + P_\theta(L < r)P(U \in S(u, \varphi))$$

where P_θ is the CDF of the log-normal distribution. The function we wish to maximise is

$$D(\theta) = \int_{r_1}^{r_2} \left(\sqrt{\frac{\hat{\gamma} \nu(B(0, r, u, \varphi)) + P_\theta(L < r)P(U \in S(u, \varphi))}{\hat{\gamma}^2}} - \sqrt{\frac{\hat{K}(r, u, \varphi)}{\hat{\gamma}^2}} \right)^2 dr.$$

To calculate the maximum we use a Newton method. We have to calculate the gradient and the Hessian matrix. Straight forward calculations yields

$$\begin{aligned} \frac{\partial}{\partial \theta_i} D(\theta) &= \int_{r_1}^{r_2} -2 \frac{\sqrt{\hat{K}(r, u, \varphi)}}{\hat{\gamma}^2} \frac{\frac{\partial}{\partial \theta_i} P_\theta(L < r)P(u \in S(u, \varphi))}{\sqrt{\hat{\gamma} \nu(B(0, r_1, u, \varphi)) + P_\theta(L < r)P(u \in S(u, \varphi))}} \\ &\quad + 2 \frac{\partial}{\partial \theta_i} P_\theta(L < r)P(u \in S(u, \varphi)) \nu(B(0, r, u, \varphi)) \\ &\quad + 2 P_\theta(L < r) \frac{\partial}{\partial \theta_i} P_\theta(L < r)P(U \in S(u, \varphi))^2 dr \end{aligned}$$

for $\theta_i = \mu$ or $\theta_i = \sigma$ for the first derivatives. We further know

$$\frac{\partial}{\partial \theta_i} P_\theta(L \leq r) = \int_0^r \frac{\partial}{\partial \theta_i} f_L(l; \theta) dl$$

where f_L is the density of a log-normal distribution. The first derivatives are known from Equations (2.3) and (2.4). For the Hessian we look at the second derivatives and get

$$\begin{aligned} \frac{\partial^2}{\partial \theta_i \partial \theta_j} D(\theta) &= \int_{r_1}^{r_2} \frac{\sqrt{\hat{K}(r, u, \varphi)}}{\hat{\gamma}^2} \left(\frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} P_\theta(L < r) P(U \in S(u, \varphi))}{\sqrt{\hat{\gamma} \nu(B(0, r, u, \varphi)) + P_\theta(L < r) P(U \in S(u, \varphi))}} \right. \\ &\quad \left. - \frac{\frac{\partial}{\partial \theta_i} P_\theta(L < r) \frac{\partial}{\partial \theta_j} P_\theta(L < r) P(U \in S(u, \varphi))}{2(\hat{\gamma} \nu(B(0, r, u, \varphi)) + P_\theta(L < r) P(U \in S(u, \varphi)))^{\frac{3}{2}}} \right) \\ &\quad + 2 \frac{\partial^2}{\partial \theta_i \partial \theta_j} P_\theta(L < r) P(U \in S(u, \varphi)) \hat{\gamma} \nu(B(0, r, u, \varphi)) \\ &\quad + 2 P(U \in S(u, \varphi))^2 \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} P_\theta(L < r) + \frac{\partial}{\partial \theta_i} P_\theta(L < r) \frac{\partial}{\partial \theta_j} P_\theta(L < r) \right) \end{aligned}$$

for $\theta_i = \mu, \text{ or } \sigma$ and $\theta_j = \mu$ or σ .

Again we know

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} P_\theta(L \leq r) = \int_0^r \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_L(l; \theta) dl$$

with f_L the density of a log-normal distribution, where the second derivatives can be found in Equations (2.7), (2.9) and (2.8).

We now have the complete gradient $\nabla D(\theta)$ and Hessian $\mathcal{H} D(\theta)$ we can now make a Newton iteration with

$$\theta_{i+1} = \theta_i + \mathcal{H}(D(\theta_i))^{-1} \nabla D(\theta_i).$$

For the estimation the main direction is $u = (0, 0, 1)^T$ and a the cone angle used for the directed K function is $\varphi = 0.3176$. With this φ we cover 20% of the sphere around the main direction u . To estimate the integration boundaries we chose the distance $p = 0.05$. The results can be seen in Table 4.1. We see that the estimation for μ is at least close to the true parameter, while the results for the estimation of σ show a significant underestimation for both window sizes. Furthermore we see that in some cases we see no convergence to any parameter. At least for a log-normal distribution the use of a minimum contrast estimator can not be recommended.

	$E(\hat{\mu})$	$\text{var}(\hat{\mu})$	$E(\hat{\sigma})$	$\text{var}(\hat{\sigma})$	no convergence
$W = 400$	4.6854	0.5770	0.0296	0.0078	7
$W = 600$	5.2061	0.4510	0.0115	0.0025	1

Table 4.1.: Simulation Study Results for the minimum contrast estimator, with true values $\mu = 4.6002$ and $\sigma = 0.0998$

5. Comparison and Applications

In this chapter we want to compare the presented estimators and discuss possible application.

5.1. Interacting Fibres

To directly compare the estimators, we simulated RSA fibres processes. These introduce non overlapping fibres, which is more realistic compared to the non interacting systems used in the estimation before.

For the RSA we will model the fibres as cylinders with a random length $L > 0$ and a random orientation $U \in S^{d-1}$. We choose a constant radius of $r = 4$.

For the length we chose a log-normal distribution with parameters $\mu = 4.6002$ and $\sigma = 0.0998$ and a Schladitz β -distribution with $\beta = 0.1$ for the orientation. We simulated in a window $W = [0, 800]^3$ with periodic boundary conditions. We stopped the RSA simulation once we reached a fibre volume of 10%. We made 30 realisations with these parameters. All simulations discussed in this chapter are the result of this RSA process.

5.2. Parametric estimation

5.2.1. Full Fibres

Here we will investigate the HT and EM estimators discussed in Chapter 3.

For the estimation from complete fibres we take a smaller sub-window from the simulations, we will look into centred sub-windows of size $W = 100^3, 200^3$ and 300^3 . In $W = 100^3$ we expect to see 20 fibres, in 200^3 we expect to see around 160 fibres and in $W = 300^3$ we expect to see 537 fibres.

The results can be seen in Table 5.1. We first note that for $W = 100$ the HT estimator could not be computed, since usually there were no uncut fibres left. For bigger window sizes the HT estimator gives good estimates of the parameters. The estimators obtained using an unbiased associated point rule lead to an overestimation of the parameter σ that grows with the window size. The estimator based on the plus sampling rule leads to an even larger overestimation. For small window sizes this overestimation vanishes.

	$E \hat{\mu}$	$\text{var } \hat{\mu}$	$E \hat{\sigma}$	$\text{var } \hat{\sigma}$
$W = 100$				
HT	-	-	-	-
associated point	4.5432	0.0114	0.0908	0.0022
plus sampling	4.5818	0.0109	0.1080	0.0029
$W = 200$				
HT	4.6001	0.0001	0.0974	6.76e-5
associated point	4.5875	0.0002	0.1192	6.34e-5
plus sampling	4.5752	0.0003	0.1355	7.21e-5
$W = 300$				
HT	4.5996	2.64e-5	0.0998	2.26e-5
associated point	4.5892	6.30e-5	0.1447	3.21e-5
plus sampling	4.5662	9.74e-5	0.1730	2.83e-5

Table 5.1.: Simulation Study for HT and EM estimators with true parameters $\mu = 4.6002$ and $\sigma = 0.0998$.

5.2.2. Fibre Endpoints

Here we will discuss the estimators based on fibre endpoints. We will start with the non parametric estimator. Here we will only look at the anisotropic estimator, since the isotropic estimator has shown to have significantly worse results. For the parametrization we used the same parameter as in the model with no interaction, we discretised the interval $[0, \pi]$ with 50 steps on the sphere. The length interval was discretised with steps of 5.

The results in Figure 5.1 show that the estimator systematically underestimates the true probabilities, especially for short lengths. This is due to the interaction in the fibre process that leads to a repulsion of the underlying endpoint process of the fibre. This directly leads to a negative correlation between the endpoints, which in turn leads to a K function that is smaller than the K function in the Poisson case. How strong this repulsion is directly connected to the volume fraction. If we compare the window sizes $W_l = 400$ and $W_l = 600$ we see that the variance is higher for smaller window sizes.

To investigate the repulsion in the endpoints we made a small simulation study with an RSA process with a volume fraction of 2.5% fibres. In Figure 5.2 we see that the underestimation for small lengths is significantly smaller. This is explained by the lower degree of interaction in lower volume fractions, i.e. the process with a lower volume fraction is closer to a Boolean model.

In application on real data we expect to see this underestimation as well.

We now evaluate the minimum contrast estimator. We did a simulation study with the same parameters as for the process with no influence. For $W_l = 400$ the minimum contrast estimator never converged. For $W_l = 600$, the results in Table 5.2 show the same results as for the Boolean model. Problematic is, that for RSA fibres the minimum contrast estimator only converges for 23 of the 30 simulations.

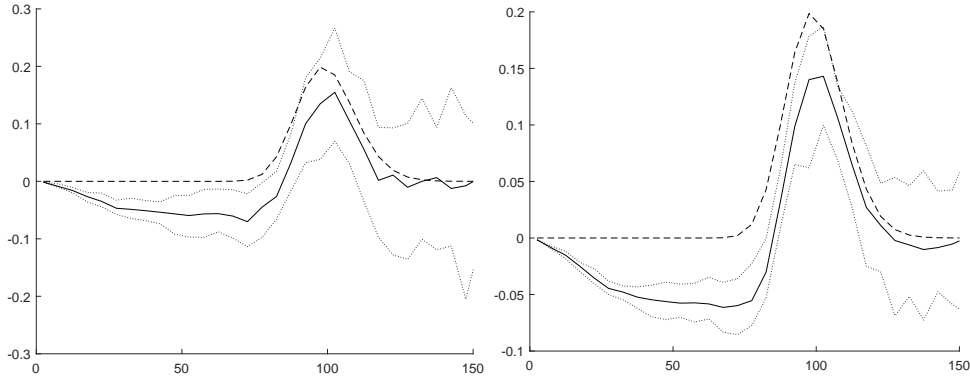


Figure 5.1.: Comparison of the anisotropic length estimator on the left for window size $W_l = 400$ and on the right for window size $W_l = 600$ for RSA fibres. The mean of the estimator is solid, the true value is dashed and an estimated 95% region is dotted. The RSA fibres had a volume fraction of 10%

The problem is that for the RSA fibres, the estimated K function lies beneath the K function for the Poisson case. Since we measure how far above the estimated K function lies above the Poisson K function the minimum contrast estimator cannot be used.

	$E(\hat{\mu})$	$\text{var}(\hat{\mu})$	$E(\hat{\sigma})$	$\text{var}(\hat{\sigma})$	no convergence
$W = 600$	5.3024	0.0679	0.0255	0.0012	23

Table 5.2.: Simulation Study Results for the minimum contrast estimator on RSA fibres, with true values $\mu = 4.6002$ and $\sigma = 0.0998$

All of these estimation where done using the model that uses a Poisson midpoint process. One might ask why no Monte Carlo simulation for the minimum Contrast estimation was done. In this approach we do not know a closed formula for the K function but instead use a Monte Carlo approach to estimate it. We could then do a grid search or follow the nelder mead simplex approach to find an optimal parameter set. In this case this cannot be done, since simulating one RSA sample of fibres takes around 24 hours to complete. To do a full optimisation with this method we would need to simulate hundreds or thousands of these, this is not simply not practically possible.

5.2.3. Application to real data

The methods that work with fully segmented fibres could not be done, since a full fibre segmentation is still not possible.

An application on real fibre endpoints was tried. A dataset of endpoints was provided by M. Kronenberger from the Fraunhofer ITWM. They were segmented

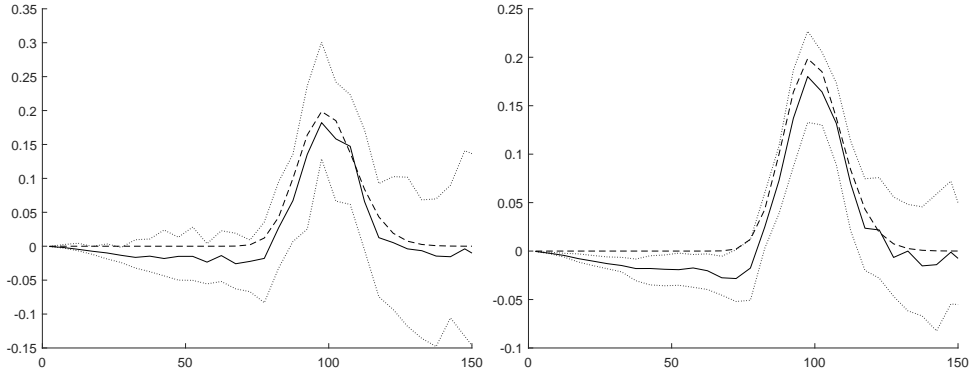


Figure 5.2.: Comparison of the anisotropic length estimator on the left for window size $W_l = 400$ and on the right for window size $W_l = 600$ for RSA fibres. The mean of the estimator is solid, the true value is dashed and an estimated 95% region is dotted. The RSA fibres had a volume fraction of 10%

using local Gaussian curvature as described in [10]. The result was a point process with 6172 points in a window $W = [0, 350] \times [0, 1000] \times [0, 900]$ with a resulting intensity of $\delta = \frac{6172}{350 \cdot 1000 \cdot 900} = 1.96e-5$. For the orientation distribution a Schladitz- β distribution was used with an estimated $\hat{\beta} = 0.102$ with a main direction $u = (-0.0130315, 0.00281079, -0.999911)$.

Using this data set we tried to estimate the parameters of a log-normal and a normal length distribution using the minimum contrast approach. But the minimum contrast estimator did not converge.

For the non parametric estimator presented we chose a length interval of $[0, 90]$ since $0.25 \cdot 350 = 87.5 \approx 90$, the maximum length where direct estimation of Ripley's K function is deemed feasible. The results can be seen in Figure 5.3. With this estimator we can conclude, that 22% of all fibres are shorter than 90 pixels.

On the one hand these results look good. On the other we do not see the expected results of the interacting fibres close to the origin, where the simulation study suggests that it should be smaller than zero. Therefore something is pulling the K function up. We see two possible sources for this effect.

The first is fibre bending. When a fibre is not perfectly straight the distance between the endpoints is less than the full length of the fibre. Therefore the endpoints are closer together and we would see this attraction in the K function. But since fibres usually only bend small amounts we do not expect this effect to be large.

The second effect is miss segmentation of the fibre endpoints. The main problem here is false positives, i.e. points that are segmented without being endpoints. These endpoints will be located on the fibre structure and are therefore correlated to the true endpoints. This may add an attraction to the points that would lead to a higher K function.

From those two effects we expect miss segmentation to have a larger influence.

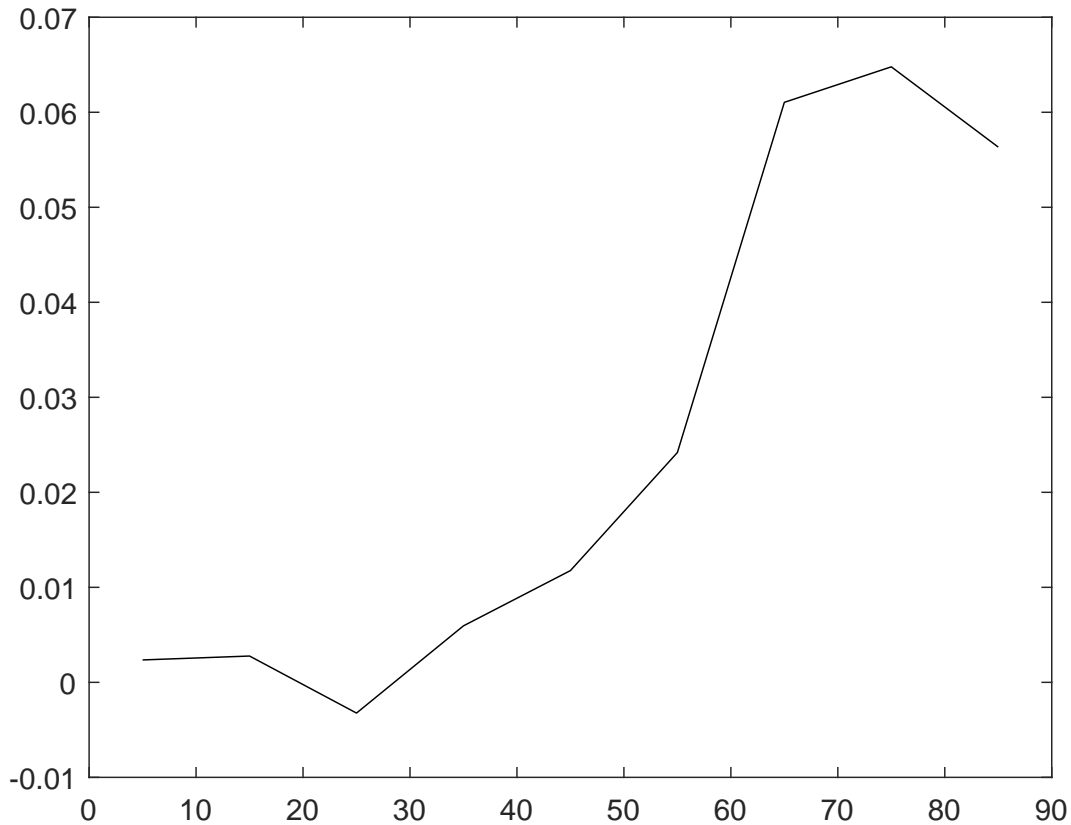


Figure 5.3.: Anisotropic length estimator for segmented end points provided by Markus Kronenberger

In the end we cannot be sure if the estimated probabilities in Figure 5.3 are correct or not but the missing dip at the beginning is at least cause for concern.

5.3. Conclusion

We presented multiple estimators to estimate the fibre length distribution in FRP. For fully segmented fibres we presented methods based on the EM algorithm in union with a reweighed distribution, first to deal with the censoring, the second to deal with sampling bias. We have shown that these estimators are consistent and asymptotically normal distributed under some mild conditions. Furthermore we presented methods to estimate the Fisher information matrix for these estimators. These estimators performed well in simulation studies, both for the parameters of a log-normal distribution as well as for the estimation of the Cramer-Rao bound. They have proven to be robust under a misspecified orientation distribution and also for non overlapping fibre systems.

Since a full fibre segmentation is currently not possible for interesting fibre volumes

we investigated the estimator based on fibre endpoints presented in [12]. In simulation studies we have seen, that the anisotropic estimator outperforms the isotropic estimator. We have seen that these kind of estimator seems to underestimate the true length distribution for RSA fibres.

We used the model for the Ripley's K function to present a minimum contrast estimator. The performance of this estimator in simulation studies was bad, it was not able to estimate the parameters of a log-normal distribution and sometimes did not converge at all. This behaviour gets even worse for the RSA fibres.

We applied the endpoint estimators to a real data set provided by Markus Kronenberger from Fraunhofer ITWM. The anisotropic estimator showed a possibly true estimation result, but did not show the underestimation that we expected from the interaction of the fibres. We think that this might be due to possible miss segmentation of endpoints. We could not get an estimate using minimum contrast estimator.

In conclusion, the use of the EM estimator for full segmented fibres seems to be a feasible method to get a consistent and asymptotically normal estimate of the parameters of the length distribution. For the estimation from endpoints we think that only the anisotropic estimator in [12] could possibly be used, if we keep in mind that we expect an underestimation of the result. The minimum contrast estimator should not be used at all.

A. Simulation Study Results

A.1. HT estimator

	$E \hat{\mu}_{HT}$	$\text{var}(\hat{\mu}_{HT})$
$W = 100$		
$n = 20$	4.5797	0.0016
$n = 50$	4.5877	0.0010
$n = 100$	4.5908	0.0005
$n = 500$	4.5982	0.0002
$W = 200$		
$n = 20$	4.5958	0.0005
$n = 50$	4.5958	0.0002
$n = 100$	4.5980	$9.93e-5$
$n = 500$	4.5998	$1.92e-5$
$W = 300$		
$n = 20$	4.5992	0.0004
$n = 50$	4.5988	0.0002
$n = 100$	4.6004	0.0001
$n = 500$	4.5998	$2.20e-5$

Table A.1.: Simulation study results for different window sizes W and sample sizes n for $\hat{\mu}_{HT}$, with true $\mu = 4.6002$ and $\sigma = 0.0998$

	$E \hat{\sigma}_{HT1}$	$\text{var}(\hat{\sigma}_{HT1})$	$E \hat{\sigma}_{HT2}$	$\text{var}(\hat{\sigma}_{HT2})$
$W = 100$				
$n = 20$	0.0770	0.0003	0.0840	0.0005
$n = 50$	0.0850	0.0002	0.0884	0.0003
$n = 100$	0.0893	0.0001	0.0910	0.0002
$n = 500$	0.0964	0.0001	0.0970	0.0001
$W = 200$				
$n = 20$	0.0956	0.0002	0.0981	0.0003
$n = 50$	0.0973	$9.6e-5$	0.0982	$9.85e-5$
$n = 100$	0.0981	$5.74e-5$	0.0986	$5.80e-5$
$n = 500$	0.0997	$1.28e-5$	0.0998	$1.2865e-5$
$W = 300$				
$n = 20$	0.0942	0.0003	0.0965	0.0003
$n = 50$	0.0977	$8.91e-5$	0.0986	$9.0243e-5$
$n = 100$	0.0994	$4.21e-5$	0.0999	$4.29e-5$
$n = 500$	0.0996	$1.18e-5$	0.0997	$1.18e-5$

Table A.2.: Simulation study results for different window sizes W and sample sizes n for $\hat{\sigma}_{HT1}$ and $\hat{\sigma}_{HT2}$, with true $\mu = 4.6002$ and $\sigma = 0.0998$

$W = 100$	$E \hat{\mu}_{HT}$	$\text{var}(\hat{\mu}_{HT})$
$n = 20$	4.2239	0.0281
$n = 50$	4.2457	0.0204
$n = 100$	4.2791	0.0234
$n = 500$	4.3056	0.0094
$W = 200$		
$n = 20$	4.4328	0.0129
$n = 50$	4.4475	0.0063
$n = 100$	4.4463	0.0035
$n = 500$	4.4709	0.0040
$W = 300$		
$n = 20$	4.4670	0.0157
$n = 50$	4.4724	0.0076
$n = 100$	4.4761	0.0034
$n = 500$	4.4870	0.0006

Table A.3.: Simulation study results for different window sizes W and sample sizes n for $\hat{\mu}_{HT}$, with true $\mu = 4.4936$ and $\sigma = 0.4724$

$W = 100$	$E \hat{\sigma}_{HT1}$	$\text{var}(\hat{\sigma}_{HT1})$	$E \hat{\sigma}_{HT2}$	$\text{var}(\hat{\sigma}_{HT2})$
$n = 20$	0.2862	0.0022	0.3193	0.0040
$n = 50$	0.3054	0.0008	0.3268	0.0023
$n = 100$	0.3164	0.0005	0.3368	0.0021
$n = 500$	0.3403	0.0006	0.3483	0.0012
$W = 200$				
$n = 20$	0.3960	0.0043	0.4093	0.0049
$n = 50$	0.4224	0.0021	0.4288	0.0022
$n = 100$	0.4265	0.0012	0.4301	0.0012
$n = 500$	0.4386	0.0015	0.4401	0.0017
$W = 300$				
$n = 20$	0.4328	0.0056	0.4448	0.0062
$n = 50$	0.4500	0.0032	0.45477	0.0034
$n = 100$	0.4602	0.0019	0.4625	0.0019
$n = 500$	0.4622	0.0003	0.4629	0.0003

Table A.4.: Simulation study results for different window sizes W and sample sizes n for $\hat{\sigma}_{HT1}$ and $\hat{\sigma}_{HT2}$, with true $\mu = 4.4936$ and $\sigma = 0.4724$

A.2. Associated point rule

	$E \hat{\mu}_a$	$\text{var}(\hat{\mu}_a)$	$E \hat{\sigma}_a$	$\text{var}(\hat{\sigma}_a)$	iterations
$W = 100$					
$n = 20$	4.5844	0.0175	0.1207	0.0129	131.67
$n = 50$	4.5827	0.0065	0.1210	0.0021	132.68
$n = 100$	4.5858	0.0023	0.1240	0.0010	143.1
$n = 500$	4.5905	0.0004	0.1279	0.0001	124.01
$W = 200$					
$n = 20$	4.5828	0.0014	0.1129	0.0005	21.29
$n = 50$	4.5866	0.0006	0.1189	0.0002	21.19
$n = 100$	4.5866	0.0003	0.1190	8.92e-6	20.04
$n = 500$	4.5877	6.48e-5	0.1202	2.16e-5	18.67
$W = 300$					
$n = 20$	4.5929	0.0014	0.1379	0.0007	14.85
$n = 50$	4.5914	0.0006	0.1210	0.0021	14.85
$n = 100$	4.5894	0.0003	0.1427	0.0001	14.13
$n = 500$	4.5905	9.00e-5	0.1447	3.08e-5	14.55

Table A.5.: Simulation Study Results for the parameters for the associated point rule, with $\mu = 4.6002$ and $\sigma = 0.0998$

	$E \hat{\mathcal{I}}_{a,1,1}^{-1}$	$\text{var}(\hat{\mathcal{I}}_{a,1,1}^{-1})$	$E \hat{\mathcal{I}}_{a,2,2}^{-1}$	$\text{var}(\hat{\mathcal{I}}_{a,2,2}^{-1})$
$W = 100$				
$n = 20$	0.0419	0.0379	0.0008	0.0105
$n = 50$	0.0060	0.0004	0.0033	0.0002
$n = 100$	0.0022	7.14e-6	0.0015	1.45e-5
$n = 500$	0.0004	8.75e-9	0.0002	5.45e-19
$W = 200$				
$n = 20$	0.0016	2.44e-6	0.0006	3.76e-7
$n = 50$	0.0006	8.15e-8	0.0003	1.03e-8
$n = 100$	0.0003	7.85e-9	0.0001	1.22e-9
$n = 500$	6.15e-5	7.64e-11	2.44e-5	1.11e-11
$W = 300$				
$n = 20$	0.0019	2.46e-6	0.0007	2.27e-7
$n = 50$	0.0007	1.40e-7	0.0003	1.74e-8
$n = 100$	0.0003	1.09e-8	0.0001	8.94e-10
$n = 500$	7.43e-5	1.18e-10	2.67e-5	8.59e-12

Table A.6.: Simulation Study Results for the fisher information for the associated point rule, with $\mu = 4.6002$ and $\sigma = 0.0998$

	$E \hat{\mu}_a$	$\text{var}(\hat{\mu}_a)$	$E \hat{\sigma}_a$	$\text{var}(\hat{\sigma}_a)$	iterations
$W = 100$					
$n = 20$	4.2614	0.0480	0.3809	0.0231	57.69
$n = 50$	4.2770	0.0152	0.4084	0.0105	48.97
$n = 100$	4.2832	0.0049	0.4226	0.0052	45.92
$n = 500$	4.2864	0.0009	0.4348	0.0011	44.42
$W = 200$					
$n = 20$	4.4009	0.0179	0.4200	0.0105	15.36
$n = 50$	4.4018	0.0057	0.4338	0.0034	13.75
$n = 100$	4.4016	0.0031	0.4419	0.0017	13.36
$n = 500$	4.3974	0.0006	0.4399	0.0003	13.12
$W = 300$					
$n = 20$	4.4298	0.0135	0.4191	0.0082	10.51
$n = 50$	4.4405	0.0064	0.4353	0.0029	9.89
$n = 100$	4.4414	0.0030	0.4448	0.0016	9.62
$n = 500$	4.4462	0.0005	0.4452	0.0003	8.96

Table A.7.: Simulation Study Results for the parameters for the associated point rule, with $\mu = 4.4936$ and $\sigma = 0.4724$

	$E \hat{\mathcal{I}}_{a^{-1},1,1}^{-1}$	$\text{var}(\hat{\mathcal{I}}_{a^{-1},1,1}^{-1})$	$E \hat{\mathcal{I}}_{a^{-1},2,2}^{-1}$	$\text{var}(\hat{\mathcal{I}}_{a^{-1},2,2}^{-1})$
$W = 100$				
$n = 20$	0.0308	0.0015	0.0241	0.0010
$n = 50$	0.0113	6.45e-5	0.0093	3.94e-5
$n = 100$	0.0054	4.42e-6	0.0047	3.14e-6
$n = 500$	0.0011	3.58e-8	0.00010	2.72e-8
$W = 200$				
$n = 20$	0.0143	4.26e-5	0.0087	2.37e-5
$n = 50$	0.0056	1.96e-6	0.0033	1.01e-6
$n = 100$	0.0020	2.73e-7	0.0017	1.47e-7
$n = 500$	0.000	1.88e-9	0.0003	9.23e-10
$W = 300$				
$n = 20$	0.0118	2.82e-5	0.0067	1.10e-5
$n = 50$	0.0049	1.39e-6	0.0027	5.52e-7
$n = 100$	0.0025	2.06e-7	0.0014	7.90e-8
$n = 500$	0.0005	1.70e-9	0.0003	6.56e-10

Table A.8.: Simulation Study Results for the fisher information for the associated point rule, with $\mu = 4.4936$ and $\sigma = 0.4724$

A.3. Plus Sampling

	$E \hat{\mu}^+$	$\text{var}(\hat{\mu}^+)$	$E \hat{\sigma}^+$	$\text{var}(\hat{\sigma}^+)$	iterations
$W = 100$					
$n = 20$	4.5261	0.0227	0.1280	0.0085	116.66
$n = 50$	4.5598	0.0087	0.1296	0.0024	141.98
$n = 100$	4.5870	0.0050	0.1272	0.0019	164.61
$n = 500$	4.6023	0.0013	0.1242	0.0005	253.52
$W = 200$					
$n = 20$	4.5620	0.0058	0.1408	0.0023	37.74
$n = 50$	4.5752	0.0012	0.1313	0.0004	40.16
$n = 100$	4.5759	0.0006	0.1330	0.0002	36.06
$n = 500$	4.5776	0.0001	0.1352	4.87e-5	34.37
$W = 300$					
$n = 20$	4.5592	0.0042	0.1702	0.0010	21.56
$n = 50$	4.5677	0.0014	0.1704	0.0003	22.22
$n = 100$	4.5663	0.0008	0.1724	0.0002	21.4
$n = 500$	4.5639	0.0002	0.1734	5.28e-5	20.96

Table A.9.: Simulation Study Results for the parameters for the plus sampling rule, with $\mu = 4.6002$ and $\sigma = 0.0998$

	$E \hat{\mathcal{I}}_{+1,1}^{-1}$	$\text{var}(\hat{\mathcal{I}}_{+1,1}^{-1})$	$E \hat{\mathcal{I}}_{+2,2}^{-1}$	$\text{var}(\hat{\mathcal{I}}_{+2,2}^{-1})$
$W = 100$				
$n = 20$	0.0126	0.0004	0.0091	0.0009
$n = 50$	0.0047	1.80e-5	0.0023	6.43e-6
$n = 100$	0.0026	8.29e-6	0.0009	1.50e-6
$n = 500$	0.0007	3.77e-7	0.0002	1.02e-7
$W = 200$				
$n = 20$	0.0042	2.48e-5	0.0016	5.42e-6
$n = 50$	0.0014	1.91e-6	0.0006	6.73e-7
$n = 100$	0.0006	7.44e-8	0.0002	1.20e-8
$n = 500$	0.0001	6.61e-10	5.01e-5	1.41e-10
$W = 300$				
$n = 20$	0.0037	5.31e-6	0.0011	6.40e-7
$n = 50$	0.0015	3.01e-7	0.0005	2.50e-8
$n = 100$	0.0007	5.34e-8	0.0002	2.16e-9
$n = 500$	0.0001	5.07e-10	4.37e-5	2.05e-11

Table A.10.: Simulation Study Results for the fisher information for the associated point rule, with $\mu = 4.6002$ and $\sigma = 0.0998$

	$E \hat{\mu}_+$	$\text{var}(\hat{\mu}_+)$	$E \hat{\sigma}_+$	$\text{var}(\hat{\sigma}_+)$	iterations
$W = 100$					
$n = 20$	4.3453	0.0446	0.3606	0.0350	67.29
$n = 50$	4.3381	0.0210	0.4029	0.0116	64.8
$n = 100$	4.3340	0.00972	0.4284	0.0053	52.87
$n = 500$	4.3380	0.0016	0.4431	0.0010	48.1
$W = 200$					
$n = 20$	4.3986	0.0195	0.4307	0.0106	22.54
$n = 50$	4.3952	0.0078	0.4479	0.0032	18.99
$n = 100$	4.3931	0.0025	0.4549	0.0012	18.35
$n = 500$	4.3903	0.0007	0.4577	0.0004	18.23
$W = 300$					
$n = 20$	4.4158	0.0161	0.4081	0.0078	14.96
$n = 50$	4.4320	0.0060	0.4241	0.0029	13.52
$n = 100$	4.4381	0.0035	0.4342	0.0018	13.3
$n = 500$	4.4429	0.0006	0.4361	0.0005	12.2

Table A.11.: Simulation Study Results for the parameters for the plus sampling rule, with $\mu = 4.4936$ and $\sigma = 0.4724$

	$E \hat{\mathcal{I}}_{+ 1,1}^{-1}$	$\text{var}(\hat{\mathcal{I}}_{+ 1,1}^{-1})$	$E \hat{\mathcal{I}}_{+ 2,2}^{-1}$	$\text{var}(\hat{\mathcal{I}}_{+ 2,2}^{-1})$
$W = 100$				
$n = 20$	0.0292	0.0005	0.0183	0.0002
$n = 50$	0.0141	0.0001	0.0100	6.25e-5
$n = 100$	0.0070	5.93e-6	0.0052	3.00e-6
$n = 500$	0.0015	4.80e-9	0.0011	2.55e-8
$W = 200$				
$n = 20$	0.0181	5.03e-5	0.0103	2.34e-5
$n = 50$	0.0072	2.49e-6	0.0041	1.17e-6
$n = 100$	0.0036	2.60e-7	0.0020	1.08e-7
$n = 500$	0.0007	2.87e-9	0.0004	1.21e-9
$W = 300$				
$n = 20$	0.0129	4.07e-5	0.0074	1.98e-5
$n = 50$	0.0053	1.84e-6	0.0030	6.53e-7
$n = 100$	0.0027	2.90e-7	0.0016	1.05e-7
$n = 500$	0.0005	2.90e-9	0.0003	1.13e-9

Table A.12.: Simulation Study Results for the fisher information for the associated point rule, with $\mu = 4.4936$ and $\sigma = 0.4724$

B. CV

Akademischer Werdegang

- 06.2006 Abitur am Immanuel Kant Gymnasium
in Bad Oeynhausen
- 10.2006-06.2012 Studium der Technomathematik
an der Universität Bremen
- 06.2012 Abschluss als Diplom Technomathematiker
an der Universität Bremen
- 11.2013-heute Wissenschaftlicher Mitarbeiter und Doktorand
im Fachbereich Mathematik an der TU Kaiserslautern

Academic CV

- 06.2006 Abitur at the Immanuel Kant Gymnasium
in Bad Oeynhausen
- 10.2006-06.2012 Studis of technomathematics
at University of Bremen
- 06.2012 Diploma in technomathematics
at University of Bremen
- 11.2013-heute Scientific Assistant and phd student
in the mathematics department at TU Kaiserslautern

Bibliography

- [1] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.
- [2] H. Domininghaus, P. Elsner, P. Eyerer, and T. Hirth. *Dmininghaus - Kunststoff: Eigenschaften und Anwendungen*. VDI-Buch. Springer Berlin Heidelberg, 2012.
- [3] J. Franke, C. Redenbach, and N. Zhang. On a mixture model for directional data on the sphere. *Scandinavian Journal of Statistics*, 43(1):139–155, 2016.
- [4] J. Franke, C. Redenbach, and N. Zhang. On a mixture model for directional data on the sphere. *Scandinavian Journal of Statistics*, 43(1):139–155, 2016.
- [5] A. Guillamón, J. Navarro, and J.M. Ruiz. Kernel density estimation using weighted data. *Communications in Statistics - Theory and Methods*, 27(9):2123–2135, 1998.
- [6] R.V. Hogg, J.W. McKean, and A.T. Craig. *Introduction to Mathematical Statistics*. Always learning. Pearson, 2013.
- [7] D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.
- [8] J. Illian, P.A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. Statistics in Practice. Wiley, 2008.
- [9] W.S. Kendall and M.N.M. van Lieshout. *Stochastic Geometry: Likelihood and Computation*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1998.
- [10] M. Kronenber, O. Wirjadi, and K. Schladitz. Local curvature for 3d-characterization of fiber-reinforced materials. In *Proceedings of the 14th International Congress for Stereology and Image Analysis*, July 2015.
- [11] A.V.K.I.V. Ku. *Handbuch Faserverbundkunststoffe/Composites: Grundlagen, Verarbeitung, Anwendungen*. Springer Fachmedien Wiesbaden, 2014.
- [12] M. Kuhlmann and C. Redenbach. Estimation of fibre length distributions from fibre endpoints. *Scandinavian Journal of Statistics*, 42(4):1010–1022, 2015.

- [13] K.V. Mardia and P.E. Jupp. *Directional Statistics*. Wiley Series in Probability and Statistics. Wiley, 2009.
- [14] A. Miettinen, C. L. Luengo Hendriks, G. Chinga-Carrasco, E. Kristofer Gamstedt, and Markku Kataja. A non-destructive x-ray microtomography approach for measuring fibre length in short-fibre composites. *Composites Science and Technology*, 72(15):1901–1908, 2012.
- [15] S. Nok Chiu S., D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic geometry and its applications*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, 2013.
- [16] J. Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Akad. Wiss.*, 69:262–277, 1917.
- [17] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2005.
- [18] W. Rudin. *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1976.
- [19] K. Schladitz, S. Peters, D. Reinel-Bitzer, A. Wiegmann, and J. Ohser. Design of acoustic trim based on geometric modeling and flow simulation for non-woven. *Computational Materials Science*, 38(1):56 – 66, 2006.
- [20] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA, 1983.
- [21] I. Svensson, S. Sjöstedt-De Luna, and L. Bondesson. Estimation of wood fibre length distributions from censored data through an em algorithm. *Scandinavian Journal of Statistics*, 33(3):503–522, 2006.
- [22] I. Svensson and S. Sjöstedt de Luna. Asymptotic properties of a stochastic em algorithm for mixtures with censored data. *Journal of Statistical Planning and Inference*, 140(1):111 – 127, 2010.
- [23] M. A. Tanner. *Tools for statistical inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer, 1996.
- [24] C. F. Jeff Wu. On the convergence properties of the em algorithm. *Ann. Statist.*, 11(1):95–103, 03 1983.